

И.И.Елисеева, Н.А.Флуд, М.М.Юзбашев

ПРАКТИКУМ по общей теории статистики

**Под редакцией члена-корреспондента
Российской Академии наук
И.И.Елисеевой**

Допущено
Учебно-методическим объединением
по образованию в области статистики
в качестве практикума для студентов
высших учебных заведений,
обучающихся по специальности "Статистика"
и другим экономическим специальностям



**Москва
"Финансы и статистика"
2008**

УДК 311(076.5)

ББК 60.60я73

Е51

РЕЦЕНЗЕНТЫ:

Кафедра математической статистики и эконометрики

Московского государственного университета

экономики, статистики и информатики

(заведующий кафедрой – профессор В.С. Мхитарян);

Г.Л. Громыко,

доктор экономических наук, профессор кафедры статистики

МГУ им. М.В. Ломоносова

Елисеева И.И.

Е51 Практикум по общей теории статистики: учеб. пособие / И.И. Елисеева, Н.А. Флуд, М.М. Юзбашев; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2008. – 512 с.: ил.

ISBN 978-5-279-02595-4

Практикум предназначен в качестве сопровождения учебника И.И. Елисеевой, М.М. Юзбашева «Общая теория статистики» (М.: Финансы и статистика, 2006). Содержит краткие методические указания и решения типовых задач по всем разделам курса (включая системы регрессионных уравнений) с применением базовых пакетов прикладных программ Excel, Statistica, SP SS и др., а также контрольные задания, вопросы для повторения, методические рекомендации преподавателям и студентам, список применяемых формул, варианты контрольных работ и тестовых заданий для проведения письменного экзамена. В приложении даются статистико-математические таблицы, а также таблицы, содержащие статистические данные по странам мира, на базе которых составлены многие задания.

Для студентов, обучающихся по специальности «Статистика» и другим экономическим специальностям.

Е $\frac{0702000000-010}{010(01)-2008}$ 105 – 2007

УДК 311(076.5)

ББК 60.60я73

© Елисеева И.И., Флуд Н.А.,
Юзбашев М.М., 2008

ISBN 978-5-279-02595-4

© Издательство «Финансы и статистика», 2008

Предисловие

Предлагаемый Практикум является дополнением к учебнику «Общая теория статистики»*, в совокупности с которым составляет единый учебно-методический комплект по изучению указанной дисциплины.

«Общая теория статистики» — одна из ключевых дисциплин в подготовке экономистов. Содержание Практикума ориентировано на выработку практических навыков по сбору, обработке, анализу и интерпретации различных статистических данных. Приобретенные навыки реализации статистических методов будут полезны в самых различных сферах деятельности. Это старались учесть авторы Практикума при подборе материалов для задач и примеров. Задания Практикума построены прежде всего на данных государственной и международной статистики, отражающих самые разнообразные стороны экономической и социальной жизни. Многие использованные данные имеют самостоятельное познавательное значение и, безусловно, будут способствовать расширению кругозора студентов, знакомить с реальной величиной тех или иных статистических показателей. Ряд заданий из соображений упрощения иллюстрации того или иного статистического метода построены на условных данных. Задания, основанные на массовых статистических данных, предполагают использование компьютерной техники и соответствующих программных средств (Excel, Statistica, SPSS и др.).

Структура Практикума соответствует структуре учебника, что нашло отражение в названии тем. При изложении материала авторы стремились выдержать единый порядок. В начале каждой темы даются примеры решения типовых задач по теме, сопровождаемые кратким изложением теории. Далее предлагаются задания для самостоятельной работы студентов, которые можно использовать в ходе как аудиторной, так и внеаудиторной работы. Кроме того, приводятся методические указания для преподавателей по проведению практических занятий и орга-

* Елисеева И.И., Юзбашев М.М. Общая теория статистики: учебник / под ред. И.И. Елисеевой. — 5-е изд., перераб. и доп. — М.: Финансы и статистика, 2006.

низации внеаудиторной работы, контролю усвоения материала. В конце каждой темы дается список вопросов, который направлен на самоподготовку и самоконтроль студентов. По некоторым темам предлагаются варианты тестовых заданий и контрольных работ.

В приложении приведены статистико-математические таблицы, необходимые для решения задач. В отдельный блок выделены наиболее часто используемые статистические формулы.

В Практикуме получил отражение многолетний опыт организации учебного процесса по дисциплине «Общая теория статистики», накопленный на кафедре статистики и эконометрики Санкт-Петербургского государственного университета экономики и финансов. В соответствии с современными требованиями к учебному процессу больший акцент делается на самоподготовку и внеаудиторную работу студентов. Надеемся, что многовариантность решений задач, необходимость комплексного применения статистических методов, нацеленность на обобщение и интерпретацию полученных результатов статистического исследования будут способствовать формированию творческих навыков и аналитических способностей у будущих специалистов.

Большой набор методических материалов (задачи, тесты, контрольные вопросы и др.), различающихся по степени сложности, объему расчетов и трудозатрат в целом, позволит моделировать график учебного процесса и использовать Практикум в соответствии с конкретными условиями: при разном объеме и структуре часов аудиторных занятий, в зависимости от форм обучения и специальности студентов, наличия компьютерного обеспечения и т.д.

Понятие о статистике

1.1. Основные положения

В данной теме раскрывается сущность статистики как науки — ее предмет, метод, исходные понятия, категории.

Термин **«статистика»** употребляется в нескольких значениях:

- синоним слова **«данные»** (например, статистика заболеваемости, статистика урожайности, статистика правонарушений и т.д.);

- *отрасль знаний* о принципах и методах работы с числовыми данными, характеризующими массовые явления;

- *отрасль практической деятельности людей*, направленная на сбор, обработку, анализ и публикацию статистических данных.

Слово «статистика» (от лат. *status* — состояние, положение вещей) в научный обиход вошло в XVIII в. в значении «государствование», тогда же статистика была введена в университетское образование как самостоятельная научная дисциплина.

Корни статистики уходят в глубокую древность. Исторически развитие учета и статистики неразрывно связано с развитием государства. Задачи государственного управления требовали наличия данных о населении, его составе, доходах и имуществе. Организовывались переписи населения, составлялись земельные кадастры и описания государств. Так были заложены основы описательной статистики.

Позднее получили развитие разделы количественной статистики — с появлением теории вероятностей. Методы, основанные на законах теории вероятностей и математической статистики, нашли применение в страховании, теории рисков, аудите, контроле и анализе качества продукции, в изучении уровня жизни населения и других самых разнообразных социально-экономических исследованиях.

Статистика может быть определена как *собираание массовых данных, их обобщение, представление, анализ и интерпретация*. Статистика представляет особый метод, который используется в

различных сферах деятельности, в решении разнообразных задач.

Статистическая наука включает:

- дискрептивную (описательную) теорию статистики — изложение общих правил сбора и обработки массовых данных;
- теорию вероятностей — науку о свойствах генеральной совокупности бесконечно большого объема (так называемого универсума);
- математическую статистику, рассматривающую правила оценивания параметров и свойств генеральной совокупности по данным выборки;
- социально-экономическую статистику и статистику населения;
- финансовую статистику.

Все составляющие статистической науки взаимосвязаны.

Предметом статистического изучения выступают *совокупности — множества однокачественных, варьирующих явлений*.

В каждом отдельном явлении *необходимое* — то, что присуще всем явлениям данного вида, проявляется в единстве со *случайным*, индивидуальным, присущим лишь этому конкретному явлению. Так, например, в разных семьях на корзину продуктов питания расходуется различная сумма средств, составляющая ту или иную долю семейного бюджета. Это зависит от образа жизни семьи, ее привычек, предпочтений и, конечно, уровня благосостояния. Однако обобщение данных по большому числу семей показывает достаточно четкую закономерность: наличие обратной зависимости между долей расходов на питание и уровнем доходов. Чем выше уровень жизни — тем ниже доля расходов на питание (так называемый закон Энгеля*).

Закономерности, в которых необходимость неразрывно связана в каждом отдельном явлении со случайностью и лишь во множестве явлений проявляет себя как закон, называются *статистическими*. Статистическая закономерность обладает свойством устойчивости, т.е. стабильности и повторяемости при повторных наблюдениях. Свойство статистических закономерностей проявляется лишь в массе явлений, при обобщении данных по большому числу единиц. Связь между численностью исследуемых явлений и полнотой проявления изучаемой статистической закономерности называется *законом больших чисел*.

* Эрнст Энгель (1821–1896) — немецкий экономист, статистик.

Статистическая совокупность состоит из единиц совокупности.

Предел дробления объекта исследования (статистической совокупности), при котором сохраняются все свойства изучаемого процесса, называется *единицей совокупности*. Каждая единица совокупности представляет собой частный случай проявления изучаемой закономерности.

Единицы совокупности обладают определенными свойствами, которые принято называть *признаками*. Чем более однородна совокупность, тем больше общих признаков имеют ее единицы и тем меньше варьируют (различаются) их значения. Например, студенты имеют такие признаки, как возраст, семейное положение, национальность, курс обучения, специальность, форма обучения (дневная, вечерняя, заочная, экстернат), наличие практического опыта работы, уровень владения иностранными языками, уровень успеваемости и т.д. При этом, если мы будем рассматривать совокупность студентов, обучающихся в одной группе, часть перечисленных выше признаков не будет варьировать (форма обучения, курс, специальность) или будет варьировать незначительно (например, возраст), и такая совокупность может считаться достаточно однородной. Совокупность студентов вузов города уже менее однородна, поскольку меньше общих признаков имеют отдельные единицы и больше варьируют их значения.

Статистическая совокупность, как правило, включает в себя несколько *частных* совокупностей, представляющих особые типы явлений, модификации изучаемой закономерности.

Единицы разных совокупностей в рамках общего качества отличаются кругом признаков и их величинами.

Признаки различаются по характеру выражения, способу измерения, отношению ко времени и другими особенностями, которые влияют на приемы их статистического изучения. Все виды признаков отражены на рис. 1.1.

Описательные признаки выражаются словесно (форма собственности предприятия, вид используемого сырья, профессия работника и др.). Описательные признаки подразделяют на *номинальные*, которые нельзя упорядочить, ранжировать (национальность человека, отраслевая принадлежность предприятия и др.) и *порядковые*, которые можно ранжировать (тарифный разряд работника, балл успеваемости студента, рейтинги компаний и др.).



Рис. 1.1. Классификация признаков в статистике

Количественные признаки — такие, отдельные значения которых имеют числовое выражение (площадь территории региона, стоимость фондов предприятия, цена товара и т.д.). Чаще всего статистика оперирует с количественными признаками.

Первичные признаки характеризуют единицу совокупности в целом. Они могут быть измерены, сосчитаны, взвешены и существуют сами по себе независимо от их статистического изучения (численность жителей города, валовой сбор зерна, сумма страховых выплат).

Вторичные признаки получают расчетным путем через соотношение первичных признаков. Например, себестоимость единицы продукции — это результат деления производственных затрат на количество единиц выпущенной продукции, урожайность — отношение валового сбора к посевной площади, рентабельность капитала — отношение прибыли к стоимости капитала. Вторичные признаки являются продуктами человеческого сознания, результатами познания изучаемого объекта. Вторичный — не означает второстепенный. Термин определяет только путь по-

знания: сначала измеряются первичные признаки, а во вторую очередь на основе первичных рассчитываются значения вторичных признаков.

Прямые признаки — свойства, присущие тому объекту, который ими характеризуется.

Косвенные признаки — свойства, присущие не самому изучаемому объекту, а другим совокупностям, относящимся к объекту. Например, стоимость выпуска продукции — прямой признак предприятия, выпуск продукции на одного работника (выработка или производительность труда) — косвенный признак предприятия.

Альтернативные признаки — те, которые принимают только два значения (пол человека, место проживания (город—село), признаки обладания или необладания чем-то (квартира телефонизированная или нет)).

Дискретные признаки — имеют только целочисленные значения (число филиалов фирмы, количество коек в стационаре, число вкладов в банке).

Непрерывные признаки — способные принимать любые значения, как целые, так и дробные. К непрерывным относятся все вторичные признаки.

Моментные признаки — характеристики состояния, наличия чего-либо на определенный момент времени (площадь особо охраняемых природных территорий, число абонентов оператора сотовой связи, численность безработных в регионе).

Интервальные признаки — характеристики процесса за определенный промежуток времени: год, полугодие, квартал, месяц, сутки и т.д. (ввод в эксплуатацию жилья за квартал, число мигрантов за год, товарооборот фирмы за месяц).

Сущность статистической методологии заключается в комплексном использовании *статистического метода*, отражающего отдельные стадии статистического исследования — *сбор данных* (статистическое наблюдение), *их обобщение, представление, анализ и интерпретация*.

Статистические данные могут быть взяты из публикаций, а можно собрать новую информацию по изучаемой совокупности (провести специальное наблюдение).

Обобщение данных включает *группировку* — разграничение общей совокупности на группы однородных единиц и *сводку* — обобщение значений признаков в сводные статистические по-

казатели для каждой группы и совокупности в целом. Для представления информации широко используются таблицы и графики.

Процесс анализа охватывает все стадии исследования. Каждый шаг исследования завершается интерпретацией полученных результатов и выработкой рекомендаций для принятия решений.

1.2. Задания по теме

1. Какова история становления термина «статистика»? От какого слова он происходит? Кто и когда впервые ввел этот термин в научный оборот? Как и почему менялось содержание термина?

2. Сформулируйте определение статистики и дайте ему соответствующее обоснование.

3. Охарактеризуйте структуру статистической науки. Покажите взаимосвязь отдельных статистических дисциплин.

4. Назовите три основные черты статистической совокупности. Проиллюстрируйте свой ответ примерами конкретных совокупностей.

5. Укажите, какие совокупности можно выделить в финансовой сфере, в образовании.

6. Назовите совокупности, в которые входит жительница Санкт-Петербурга, работающая бухгалтером на машиностроительном предприятии и обучающаяся на заочном факультете СПбГУЭФ.

7. В какие совокупности входит кондитерское предприятие ООО «МАРС», расположенное по адресу: Московская обл., Ступинский район, г. Ступино-1?

8. В какие совокупности входит российская студентка, уехавшая на семестровое обучение в Потсдамский университет (Германия)?

9. Можно ли считать статистической закономерностью повышенный уровень заболеваемости острыми респираторными заболеваниями в весенне-осенний период?

10. Можно ли считать уменьшение численности населения России закономерностью, или это случайность?

11. Как знание закономерностей потребительского спроса используется в крупных универмагах при расположении товаров?

12. По результатам опроса студентов группы № 215 дневного отделения СПбГУЭФ установлено, что из 20 студентов курят 8 человек. Какую совокупность это характеризует:

- а) всех студентов Санкт-Петербурга;
- б) всех студентов СПбГУЭФ;
- в) только студентов этой группы;
- г) только студентов дневного отделения?

13. Какие закономерности потребительского спроса учитываются при размещении в городах сети булочных и универсальных магазинов?

14. Какие закономерности нужно учитывать при составлении расписания и маршрутов движения пригородного транспорта? Внутригородского транспорта?

15. Почему сети питания Fast Food обычно располагаются у станций метрополитена и на крупных транспортных узлах?

16. В табл. 1.1 перечислены некоторые признаки торгового предприятия. Определите, к каким видам признаков они относятся — проставьте значки в соответствующих графах, как показано в строке 1. Продолжите список признаков.

- 17.** Назовите, какими признаками можно охарактеризовать:
- 1-й вариант — страховую компанию;
 - 2-й вариант — автотранспортное предприятие;
 - 3-й вариант — учреждение связи.

Для каждого варианта укажите 8–10 признаков. К каким видам относятся названные вами признаки?

Для выполнения задания используйте макет таблицы, предложенный в задании 16.

18. Объясните, почему при сравнении производительности труда работников одного швейного цеха не учитывается такой признак, как уровень шума в помещении?

19. Можно ли считать связь между курением и раком легких примером статистической закономерности? Обоснуйте свой ответ.

Таблица 1.1

Признак	По характеру выражения			По способу измерения		По характеру вариации			По отношению ко времени	
	описательный		количественный	первичный	вторичный	альтернативный	дискретный	непрерывный	моментный	интервальный
	номинальный	порядковый								
Площадь торгового зала универмага			+	+				+	+	
Месячная сумма выручки										
Численность сотрудников на начало года										
Доля продавцов в общей численности сотрудников										
Наличие паркинга										
Средняя выручка на одного продавца за месяц										

Признак	По характеру выражения			По способу измерения		По характеру вариации			По отношению ко времени	
	описательный		количественный	первичный	вторичный	альтернативный	дискретный	непрерывный	моментный	интервальный
	номинальный	порядковый								
Доля отечественных товаров в общем объеме товарооборота										
Затраты на проведение рекламных акций за полугодие										
Рентабельность товарооборота										
Прибыль универмага за год										

1.3. Вопросы для самоконтроля

1. В каких значениях употребляется термин «статистика»?
2. Что является предметом статистики?
3. Что такое статистическая закономерность?
4. Дайте определение таким категориям, как статистическая совокупность, частная совокупность, единица совокупности.
5. Что определяет степень однородности статистической совокупности? Поясните ответ примерами.
6. По каким основаниям классифицируются признаки единиц совокупности? Дайте определения основным видам признаков. Приведите примеры.
7. В чем заключается сущность статистического метода?
8. Как вы понимаете выражение «универсальность статистического метода»?

1.4. Методические рекомендации преподавателям

Аудиторная работа. Форма организации и число практических занятий по теме определяются объемом часов аудиторных занятий. При их достаточном количестве целесообразно первое занятие по курсу организовать в виде семинара. На обсуждение могут быть вынесены вопросы:

1. История становления статистической науки.
2. Выдающиеся ученые-статистики.
3. Универсальность статистического метода.
4. Статистическая закономерность и роль закона больших чисел в статистике.

Участие в семинарах требует определенной подготовки студенческой аудитории. Поэтому необходимы предварительная самостоятельная внеаудиторная работа студентов, написание докладов, эссе, рефератов, конспектирование литературы по теме.

Рекомендуется подчеркнуть практическую значимость обсуждаемых вопросов. На примерах показать, как решается вопрос о единице и границах изучаемой совокупности в зависимости от цели и уровня исследования, к чему может привести неправильное определение границ исследуемой совокупности или единицы совокупности. Например, обсудить, каковы границы сово-

купности и что будет единицей совокупности, если изучается занятость и безработица в регионе, уровень жизни и потребления, вопросы эффективности затрат труда. Можно предложить назвать совокупности, которые могут быть выделены в высшем учебном заведении в зависимости от цели исследования. Какие частные совокупности включают рассматриваемые совокупности? Что будет являться единицей совокупности в каждом конкретном случае? Что определяет однородность совокупности? Какие существенные признаки единиц необходимо учесть в исследовании. Логическим продолжением дискуссии может быть обсуждение основных стадий статистического исследования. Важно показать их взаимосвязь и взаимообусловленность.

Существенная часть аудиторного времени должна быть уделена классификации признаков в статистике. Отработка навыков определения вида признаков будет продуктивной только в том случае, если студенты будут представлять, при решении каких задач эти навыки могут быть использованы. Так, умение различать первичные и вторичные признаки важно при выборе способа обобщения данных (первичные признаки обобщаются путем суммирования значений признаков по отдельным единицам совокупности, вторичные — с помощью расчета средних величин), характер вариации — при проведении группировок, деление на моментные и интервальные признаки учитывается при анализе динамики и т.д.

Самостоятельная внеаудиторная работа. Студенты изучают литературу по теме и конспектируют ее, пишут небольшие рефераты, эссе. В частности, могут быть предложены темы: «Основные этапы становления статистической науки», «Статистический учет в Древнем мире», «Вильям Петти — родоначальник статистики», «Роль политических арифметиков в формировании статистической науки», «Вклад немецкой школы государственоведения в развитие статистики», «Развитие традиций государственоведения в современной международной статистике», «История учетно-статистических работ в России», «Основные периоды развития советской статистики (плюсы и минусы)», «Особенности и задачи современного этапа развития статистики» и др. Можно предложить написание сочинений, посвященных вкладу в статистическую науку таких ученых, как А. Кетле, А.И. и А.А. Чупровы, Ю.Э. Янсон, А.А. Кауфман и др.

В качестве домашнего задания по теме может быть предложена индивидуальная работа «Признаки единиц совокупности и их виды». Результаты индивидуальной работы можно представить в форме таблицы по примеру задания 16. Каждому студенту предлагается определить признаки (не менее 10–15), характеризующие единицу конкретной совокупности, и определить вид признака. Примерный перечень совокупностей: учреждения здравоохранения, коммерческие банки, строительные компании, агентства по продаже недвижимости, автотранспортные предприятия, развлекательные учреждения, кинотеатры, дошкольные учреждения, авторемонтные мастерские, товарно-сырьевые биржи, средние образовательные учреждения, санаторно-курортные учреждения, фермерские хозяйства, туристические агентства, издательства, садово-огороднические товарищества, хлебозаводы.

Итоговый контроль может быть проведен в форме тестирования.

1.5. Методические указания студентам

Освоение теории. При подготовке необходимо ориентироваться на рекомендуемый список вопросов для самоконтроля. Особое внимание следует уделить подбору собственных примеров.

В результате изучения этой темы нужно знать:

- краткую историю статистики;
- что является предметом статистики;
- основные категории статистики: статистическая закономерность, совокупность и ее свойства, частная совокупность, единица совокупности, признаки единицы совокупности;
- сущность статистического метода;
- основные стадии статистического исследования, их задачи, особенности, взаимосвязь.

Практические навыки. Необходимо научиться определять вид конкретного признака по характеру выражения, способу измерения, характеру вариации, по отношению ко времени.

Организация статистики. Статистическое наблюдение

2.1. Основные положения

В теме 1 отмечалась потребность государства в статистических данных. Для их сбора, обработки и представления созданы специальные организации – статистические службы. В РФ статистическая деятельность осуществляется органами государственной статистики и ведомствами. Государственная статистика в России организована по административно-территориальному признаку. Главным органом государственной статистики выступает Федеральное агентство по статистике (Росстат). Ему подчиняются территориальные органы федеральной службы государственной статистики. Росстат подчинен премьер-министру РФ.

Статистическое наблюдение – это сбор массовых данных о явлениях социальной и экономической жизни. Суть наблюдения состоит в регистрации тех признаков единицы наблюдения, которые включены в программу наблюдения. Статистическое наблюдение составляет необходимую часть статистического метода, представляет собой первый (начальный) этап статистического исследования. Оно может проводиться государственной статистикой или выполняться исследователями, решающими поставленную научную задачу статистическими методами на основе специально собранных данных.

Статистическое наблюдение может быть организовано в форме отчетности – это сбор данных органами государственной статистики по утвержденным формам. Как правило, такого рода данные собираются со строгой периодичностью (месяц, квартал, полгода, год) и регулярно.

Статистическое наблюдение бывает и специально организованным – это может быть единовременное или текущее обследование, анкетирование. Специальное наблюдение – весьма дорогостоящая работа.

Виды статистического наблюдения:

а) по охвату наблюдением единиц совокупности:

- сплошное,
- несплошное —
 - монографическое,
 - основного массива,
 - цензовое,
 - выборочное,
 - квотное;

б) по времени наблюдения:

- постоянное (текущее),
- периодическое,
- единовременное;

в) по источнику сведений:

- непосредственное наблюдение,
- документальное,
- опрос —
 - виды опроса: саморегистрация, экспедиционный способ (ответы регистрирует счетчик), корреспондентский (по почте).

Программа наблюдения включает опознавательные признаки, вопросы, отвечающие цели исследования, и контрольные вопросы.

Вопросы программы наблюдения предполагают форму ответа:

- альтернативную,
- цифровую,
- многовариантную (с указанием выбрать один или несколько вариантов ответа). Варианты ответов называются *подсказом*.

Опрашиваемого обычно называют *респондентом*, того, кто опрашивает, — *интервьюером* (*регистратором, счетчиком*).

Вопросы программы наблюдения обычно приводятся в формуляре наблюдения. Однако все большее применение получает сбор данных в электронной форме (безбумажный документооборот).

Для обеспечения точности регистрации данных, единообразия понимания вопросов к формуляру наблюдения прилагается инструкция по его заполнению.

Продолжительность наблюдения зависит от того, насколько массовым оно является. Срок (период) наблюдения включает даты начала и окончания работ по сбору данных. Срок наблю-

дения устанавливается в зависимости от объекта и цели наблюдения. На период наблюдения влияет сложность программы наблюдения, число исполнителей (счетчиков-регистраторов), которых можно привлечь к работе по сбору данных.

Если объект наблюдения постоянно меняется (например, в населении страны происходят постоянные изменения — люди рождаются, вступают в брак, разводятся, женятся, меняют место жительства), то устанавливают **критический момент наблюдения** (например, при проведении переписи населения).

Местом наблюдения, в случае если данные собираются экспедиционным способом, может быть:

- место жительства;
- место работы;
- официально установленное место опроса;
- место опроса, установленное по договоренности регистратора с респондентом.

Важнейшая задача наблюдения — получение достоверных данных. Однако в ходе массового наблюдения обязательно возникают погрешности или, иначе говоря, **ошибки наблюдения**. Они могут возникнуть как ошибки регистрации — случайные ошибки, которые не имеют направленности, не тенденциозны. Такие ошибки не сказываются на значениях сводных показателей, поскольку они взаимопогашаются при обобщении единичных данных. Ошибки могут быть умышленными (преднамеренными). В этом случае они вносят определенную тенденциозность, которая не устраняется при обобщении единичных данных и искажает значения сводных показателей.

Для повышения достоверности и точности проводится **контроль собранных данных**. Он включает проверку полноты охвата единиц наблюдением, полноту регистрации признаков по каждой единице наблюдения; контроль достоверности и точности сведений по каждому из регистрируемых признаков: логический контроль — основан на логической увязке (постулируемых связях) признаков; счетный (арифметический) контроль — основан на количественных соотношениях между признаками.

2.2. Задания по теме

1. В анкете выборочного обследования населения по проблемам занятости предусмотрены следующие вопросы респондентам (см. анкету с. 20–23).

Выберите вариант ответа в отношении объекта наблюдения:

- а) все население;
- б) население в возрасте 16–54 лет (женщины),
16–59 лет (мужчины);
- в) работающие;
- г) население в возрасте 15–72 лет;
- д) население в трудоспособном возрасте без учета инвалидов без права работать.

АНКЕТА

выборочного обследования населения по проблемам занятости

СВЕДЕНИЯ О РЕСПОНДЕНТЕ

(заполняется счетчиком)

- 1. Сколько полных лет Вам исполнилось?
- 2. Ваше состояние в браке?
- 3. Укажите Ваше родственное отношение к респонденту, опрошенному первым.
- 4. Вы являетесь гражданином: России или другого государства?
- 5. Какое образование Вы имеете?

ТЕКУЩАЯ ЭКОНОМИЧЕСКАЯ АКТИВНОСТЬ

(вопросы относятся к обследуемой неделе с ____ по ____ 200_ г.)

- 6. Выполняли ли Вы какую-нибудь оплачиваемую деньгами или натурой работу или имели доходное занятие, даже если оно составляло хотя бы один час? (Да – Нет).
- 7. Возможно, Вы выполняли какую-нибудь неоплачиваемую работу на предприятии, принадлежащем кому-либо из Ваших родственников? (Да – Нет).
- 8. Была ли у Вас работа или доходное занятие, на которых Вы временно отсутствовали по различным причинам? (Да – Нет).

ОСНОВНАЯ РАБОТА

- 9. Какова нормальная (обычная) продолжительность рабочей недели в часах на Вашей основной работе?
- 10. Сколько часов, включая сверхурочные, Вы фактически отработали на Вашей основной работе?

11.1. Назовите главную причину, по которой Вы работали меньше часов по сравнению с нормальной продолжительностью рабочей недели.

11.2. Назовите главную причину, по которой Вы не работали на обследуемой неделе.

12. Ваша работа была: на предприятии, в учреждении, организации; в фермерском хозяйстве; предпринимательской деятельностью без образования юридического лица; по найму у физических лиц?

13. Вы работали: за заработную плату (вознаграждение) или за денежное довольствие; на собственном предприятии или в собственном деле; в качестве члена производственного кооператива (артели); в качестве помогающего без оплаты на предприятии, принадлежащем кому-либо из Ваших родственников?

14. Вы работали: без наемных работников; с наемными работниками на постоянной основе; привлекаемыми от случая к случаю?

15. Какое количество наемных работников Вы привлекали?

16. Была ли Ваша работа: постоянной на неопределенный срок; сезонной, временной, случайной; по контракту на определенный период; по договору на определенный объем работ?

17. Какова основная направленность деятельности предприятия, организации либо подразделения, на котором Вы работали, или собственного дела?

18. По какой профессии, в какой должности Вы работали на основной работе?

ДОПОЛНИТЕЛЬНАЯ РАБОТА

19. Была ли у Вас другая оплачиваемая работа или доходное занятие? (Да – Нет).

20.)
21.)
22.) (Повторяются вопросы 13–17 из раздела «Основная работа».)
23.)
24.)

25. Сколько примерно часов всего Вы отработали на второй работе и других дополнительных работах, если они были?

26. Искли Вы дополнительную работу или были готовы работать больше?

27. Вы хотели бы работать: больше часов на настоящей работе; иметь другую работу с большей продолжительностью; иметь работу в дополнение к настоящей?

ПОИСК РАБОТЫ, ЭКОНОМИЧЕСКАЯ НЕАКТИВНОСТЬ

28. Хотели бы Вы иметь оплачиваемую работу? (Да – Нет).

29. Если бы Вам предложили подходящую работу, смогли бы Вы приступить к ней немедленно? (Да – Нет).

30. Смогли бы Вы приступить к работе в течение двух недель? (Да – Нет).

31. По какой причине Вы не смогли бы приступить к работе в течение двух недель?

32. Ищете Вы работу или пытаетесь организовать собственное дело?

33. В течение четырех недель, предшествующих обследуемой неделе, какие шаги Вы предпринимали, чтобы найти работу или организовать собственное дело?

34. Назовите основную причину, по которой Вы не искали работу в последнее время.

35. Сколько приблизительно месяцев Вы, будучи незанятым, искали работу или пытались организовать собственное дело?

36. Вы хотели бы работать: по найму с полной рабочей неделей; по найму с неполной рабочей неделей; на собственном предприятии, в собственном деле; готовы к любой работе?

37. Какую работу Вы хотели бы иметь: постоянную на неопределенный срок; сезонную, временную, случайную; по контракту на определенный период; на определенный объем работ; любую работу?

38. По состоянию на обследуемую неделю были ли Вы зарегистрированы в службе занятости в качестве безработного? (Да – Нет).

39. Получали ли Вы пособие по безработице? (Да – Нет).

40. Имели ли Вы когда-либо оплачиваемую работу или доходное занятие, кроме любой случайной работы, и как давно Вы ее оставили?

41. По какой причине Вы оставили эту работу?

42.1. По какой профессии, в какой должности Вы работали на последнем месте работы?

42.2. Имеете ли Вы профессию или специальность и какова Ваша профессия, специальность?

43. К какой из категорий Вас можно отнести: студенты (учащиеся) дневной формы обучения; пенсионеры по возрасту, выслуге

лет, на льготных условиях; пенсионеры по инвалидности; получатели дохода от собственности; домашние хозяйства и другие родственники, отвечающие за уход за домом и детьми; другие.

2. Укажите, какие вопросы вышеприведенной анкеты:

- составляют опознавательные признаки респондента;
- предусматривают альтернативную форму ответа;
- являются открытыми;
- имеют подсказки;
- имеют логические связи между собой.

3. В отчетности за месяц организации «Весна» указана средняя списочная численность работников 110 человек, численность штатных сотрудников — 8 человек, отработано человеко-дней — 240, человеко-часов — 1920.

Укажите, какой из показателей вероятней всего является ошибочным.

4. Укажите, к какому виду статистического наблюдения относится проведенное Росстатом в 2006 г. пилотное обследование «Семья и рождаемость»:

- сплошное;
- монографическое;
- обследование основного массива;
- выборочное.

5. При проведении Всероссийской переписи населения 2002 г. (ВПН-2002) ответы на вопросы переписного листа записывались со слов опрашиваемых или на основе документов. Что в данном случае выступало источником сведений?

Какого рода ошибки возможны при заполнении переписных листов и каково их влияние на сводные итоги переписи?

6. Ежегодно свыше 45 тыс. домохозяйств в России ведут записи о своих доходах и расходах, их целевом назначении, территории, на которой они были произведены, — в рамках обследования бюджетов домашних хозяйств (ОБДХ). Укажите, к какому виду статистического наблюдения оно относится:

- сплошное единовременное;
- обследование основного массива;
- сплошное текущее;
- выборочное единовременное;
- выборочное текущее.

7. Источником официальных статистических сведений могут быть:

- только данные, которыми располагают органы государственной статистики;
- исследовательские группы (социологические институты и службы, организующие опросы населения);
- официальные органы статистики и исследовательские группы.

8. При каком наблюдении обследованию подвергается отобранная в определенном порядке часть единиц совокупности, а получаемые результаты распространяются на всю совокупность:

- монографическом;
- выборочном;
- основного массива.

9. В квартальной отчетности малого предприятия указан объем выполненных работ (услуг), равный 39 тыс. руб. В предыдущем квартале этот показатель в отчетности этого же малого предприятия составлял 350 тыс. руб. Считая данные последнего отчета ошибочными, укажите:

- какое наиболее вероятное значение данного показателя: 390, 395 или 439;
- к какому виду ошибок относится данная ошибка.

10. При проверке бланков транспортного обследования в крупном городе выявилось несоответствие в данных одного из респондентов, указавшего время, затраченное на передвижение к месту работы или обучения в минутах: трамваем — 7, метро — 25, пешком — 18; итого 55. Кроме того, было указано время на ожидание транспорта — 4; всего 62. Респондентом было указано число пересадок при пользовании транспортом — 1.

Укажите наиболее вероятные ошибки, дайте пояснения.

11. При проведении инвентаризации методом непосредственного наблюдения на складе организации оптовой торговли оказалось, что количество мужских костюмов черного цвета в наличии превышает на две единицы количество этого товара по данным бухгалтерии, а фактическое количество женских утепленных пальто оказалось на единицу меньше остатков пальто по данным бухгалтерии.

Каковы, по вашему мнению, наиболее вероятные версии расхождения инвентаризации с учетными данными в первом случае и во втором:

- ошибка в базе данных;
- нестача;
- наличие «левого» товара;
- другое.

12. Как проводится сбор данных при корреспондентском способе:

- специально подготовленные регистраторы на основе опроса заполняют переписные формуляры;
- статистические организации рассылают специально разработанные бланки и инструкции к их заполнению и получают заполненные бланки;
- работники организации, которая проводит опрос, раздают анкеты опрашиваемым лицам, инструктируют их, а затем собирают заполненные анкеты.

13. В качестве критического момента во Всероссийской переписи населения, проведенной 9–16 октября 2002 г., было принято 0 часов с 8 на 9 октября. Отметьте наличие брачного состояния, если в бланках переписи на вопрос: «Состоите ли в браке в настоящее время?» указано:

- брак расторгнут в 10-00 9 октября;
- брак зарегистрирован 8 октября в 16 ч;
- брак расторгнут в 17 ч 8 октября.

14. Инструментарий статистического наблюдения включает:

- опросные листы;
- предлагаемые варианты ответов;
- анкеты;
- макеты итоговых таблиц;
- инструкции по заполнению формуляров;
- схемы логического и счетного контроля.

15. Логический контроль основан на:

- логической взаимосвязи между признаками;
- жесткой связи между признаками, которая может быть проверена вычитанием, сложением, делением, умножением;
- сравнении с данными прошлого периода;
- сравнении с данными аналогичных единиц совокупности.

16. В супермаркете покупателям раздают анкеты и просят ответить на вопросы о работе его персонала. Такой способ сбора данных называется:

- экспедиционный;

- корреспондентский;
- саморегистрация.

17. В одном из переписных листов Всероссийской переписи населения, имевшей критическим моментом 0 часов с 8 на 9 октября 2002 г., были произведены следующие записи. Укажите, какие ответы не согласуются между собой:

- фамилия, имя, отчество — Сидоров Егор Иванович;
- пол — мужской;
- возраст — 42 года, родился 16 ноября 1960 г.;
- состоит ли в браке в настоящее время — да;
- национальность — русский;
- образование — высшее;
- место работы — завод «Волна»;
- занятие по месту работы — бухгалтер;
- общественная группа — рабочий.

18. Проведите арифметический контроль данных о готовой продукции и полуфабрикатах по трем предприятиям фирмы «Заря» за отчетный год, млн руб.

Номер предприятия	Стоимость готовой продукции	Стоимость полуфабрикатов	Всего
1	510	150	660
2	460	80	540
3	730	220	950
Итого	1600	450	2050

Назовите правильный ответ:

- ошибок нет;
- одна ошибка;
- две ошибки;
- три ошибки.

19. В супермаркете покупателям на входе раздают анкеты, в которых просят ответить на вопросы о качестве работы персонала. Можно ли сделать какие-то выводы на основе результатов такого опроса, если на выходе заполненные анкеты сдали лишь 5% посетителей?

20. В одном из переписных листов Всероссийской переписи населения при критическом моменте 0 часов с 8 на 9 октября 2002 г. были произведены следующие записи. Укажите, какие ответы не согласуются между собой:

- фамилия, имя, отчество — Михеева Елена Юрьевна;
- пол — женский;
- возраст — 42 года, родилась 18 июля 1980 г.;
- состоит ли в браке в настоящее время — да;
- национальность — русская;
- образование — высшее;
- место работы — ООО «Прибор»;
- занятие по месту работы — пиromетрист;
- общественная группа — служащая.

21. Проверьте следующие данные месячной выручки от обслуживания населения предприятиями связи города и дайте наиболее вероятное объяснение несоответствия между числами, руб.:

Всего выручка	255
В том числе от:	
продажи конвертов, марок, открыток и других товаров	162
подписки на периодические издания	103
продажи газет и журналов	65
почтовых отправлений (посылок, бандеролей, денежных переводов и др.)	25

22. К какому виду ошибок (систематическим или случайным) могут относиться ошибки в работе фасовочного аппарата и приводят ли они к искажению сводных показателей или взаимно погашаются при обобщении данных?

23. Выделите ошибки, возникающие на этапе сбора данных и на этапе обработки данных:

- дефекты в определении единицы и объекта наблюдения;
- неточности в классификации;
- нечеткая группировка данных;
- неправильная интерпретация характера связи между свойствами объектов;
- ошибочный выбор объектов наблюдения;
- преднамеренное (сознательное) искажение данных при опросах;
- неверное графическое представление, вводящее в заблуждение;
- игнорирование различий в степени вариации разных признаков;
- пропуски в данных или же дублирование данных;
- неправильное истолкование регистрируемого признака.

24. Независимым институтом социальной политики (НИСП) проведено обследование «Мужчины и женщины; семья и дети», которым было охвачено свыше четырех тысяч домохозяйств России. К какому типу наблюдения оно относится? Можно ли публиковать его результаты в официальных статистических сборниках?

25. Администрация города заказала группе социологов провести опрос относительно согласия горожан на строительство крупного офисного здания в центральной части города.

Укажите границы объекта наблюдения, единицу наблюдения, вид наблюдения, способ собирания и источник данных.

26. В период проведения конкурса инвестиционных проектов в крупном городе запланирован опрос потенциальных инвесторов.

Каким должно быть содержание вопросника? Должны ли найти в нем отражение потребности жителей города или же только самих инвесторов?

27. В ходе проведения ВПН-2002 счетчик посетил домохозяйство 12 октября и узнал, что 10 октября в этом домохозяйстве скончался дедушка, в день прихода счетчика родилась дочка, а ее отцу пришлось уехать в длительную командировку (на 3 месяца). Счетчик учел в переписном листе всех членов домохозяйства, включая дедушку и командировочного, но не учел новорожденную.

Правильно ли поступил счетчик?

28. При проведении ВПН-2002 счетчик потребовал документы для подтверждения ответов взрослых членов домохозяйства на вопросы об образовании, состоянии в браке, виде деятельности. Опрашиваемые отказались предоставлять документы.

Правильно ли они поступили?

29. Руководство торговой фирмой «Лента» приняло решение провести опрос посетителей в целях изучения состава контингента, выяснения доступности для посетителей приобретения определенных видов продовольственных и непродовольственных товаров, а также необходимости развития сопряженных услуг (упаковки и доставки товара, сборки мебели на дому, работы с производителями по рекламациям).

Укажите, к какому виду по времени, охвату и способу получения данных будет относиться данное наблюдение.

30. В целях определения нагрузки на полотно дороги на отрезках кольцевой автотрассы с наиболее интенсивным потоком грузовых автомобилей инспектору поручено произвести взвешивание 50 грузовиков. Какой способ наблюдения он применит:

- сплошное наблюдение;
- метод основного массива;
- выборочное обследование;
- непосредственное наблюдение;
- опрос;
- документальное наблюдение?

31. Согласно Указу Президента РФ «Об оценке эффективности деятельности органов исполнительной власти субъектов Российской Федерации» от 26.06.2007 г. № 825 одним из оценочных показателей является удовлетворенность населения деятельностью органов исполнительной власти субъектов Российской Федерации, в том числе их информационной открытостью (процент от числа опрошенных). Как бы вы предложили организовать опрос населения субъекта РФ для получения достоверной информации?

2.3. Вопросы для самоконтроля

1. При изучении вовлеченности студентов дневной формы обучения в региональный рынок труда какая совокупность выступает в качестве объекта наблюдения:

- а) работодатели;
- б) студенты дневной формы обучения;
- в) домохозяйства, в которых проживают студенты дневной формы обучения, работающие для заработка?

2. Какие признаки, включаемые в программу наблюдения, называются опознавательными? Приведите примеры.

3. Как вы понимаете утверждение о том, что постановка вопроса в программе наблюдения предполагает определенную форму ответа? (Приведите примеры). Насколько это утверждение истинно?

4. К какому виду и форме организации относится учет деканатом успеваемости студентов и посещаемости занятий?

5. Поясните, что представляет собой статистическая отчетность? Какие формы отчетности вам известны?

6. Как изменяется программа отчетности в зависимости от периодичности ее представления (годовая, полугодовая, квартальная, месячная)?

7. Что может служить источником информации при изучении планов проведения отпуска молодых семей с детьми-дошкольниками?

8. Какими принципами руководствуются при составлении программы специального наблюдения:

а) чем больше вопросов, тем лучше (хуже);

б) чем больше подсказов, тем лучше (хуже);

в) чем больше вопросов, предполагающих цифровой ответ, тем лучше (хуже);

г) вопросы должны быть независимыми, чтобы не дублировалась информация, или взаимосвязанными, чтобы была возможность контроля правильности ответов;

д) чем больше так называемых открытых вопросов, тем лучше (хуже)?

9. Что вы понимаете под «пилотным» обследованием и в каких случаях его рекомендуют проводить?

10. Из чего складываются затраты на проведение специального организованного статистического наблюдения?

2.4. Методические рекомендации преподавателям

Практические занятия по этой теме целесообразно ориентировать на выполнение какой-либо определенной работы, результаты которой востребованы конкретным пользователем. Такой работой может стать планирование и проведение опроса студентов относительно их самостоятельных занятий. Данная тема интересна возможностью сопоставить затраты времени и организации самостоятельных занятий с последующими результатами экзаменационной сессии, выявлением уровня взаимопомощи, места самостоятельных занятий, обращением за консультацией к преподавателям и т.д. В рамках такого проекта (или аналогичного, связанного с изучением какой-то стороны студенческой жизни — питания, отдыха, внеучебных занятий, приобретения второго профессионального образования и т.д.) можно обсудить все вопросы организации и проведения статистического наблюдения и провести такое наблюдение. Как пра-

вило, администрация вуза заинтересована в результатах обследования условий жизни студентов. То же относится и к обследованию условий жизни преподавателей вуза. Полезно дать студентам домашнее задание по составлению программы наблюдения с последующим коллективным обсуждением разработанных проектов.

При выборе формы организации питания в вузе целесообразно сочетать метод опроса с методом непосредственного наблюдения в студенческих и преподавательских столовых, буфетах, кафе.

Еще более эффективны занятия по данной теме в том случае, если силами кафедры (или межкафедральной группы) выполняется конкретный научно-исследовательский проект, в котором принимают участие студенты. В качестве таких работ можно назвать НИР, выполненные кафедрой статистики и эконометрики СПбГУЭФ: «Структура и функции семейных групп», «Бедность в России», «Семья после развода», «Население и милиция в крупном городе», «Проведение мониторинга хозяйственно-финансовых индикаторов малого бизнеса и индивидуальных предпринимателей в целях определения потенциально возможного размера годового дохода по видам деятельности, налогообложение которых целесообразно проводить по системе единого вмененного налога». Во всех этих работах студенты участвовали в проведении массовых опросов, контроле и разработке материалов, определении репрезентативности выборки. Прекрасной школой является участие студентов в крупных работах официальной статистики: перепись населения, обследование населения по проблемам занятости и т. д.

В изучении этой темы преподавателю важно показать место официальной статистики и специальных исследований, влияние пользователей на содержание программ статистического наблюдения, необходимость обеспечения достоверности и сопоставимости статистических данных, повышение роли опросных данных в информационном обеспечении принятия управленческих решений.

2.5. Методические указания студентам

В ходе изучения этой темы вы должны понять зависимость содержательности и качества статистических данных от этапа сбора данных. В целях выработки профессиональных на-

выков вам полезно самостоятельно составить программу наблюдения, причем самостоятельность не означает единоличность: программа наблюдения обычно составляется коллегиально. Участие в работе, связанной со сбором данных, позволяет приобрести коммуникативные навыки. Часто исследовательские группы и организации приглашают студентов для выполнения таких работ (в связи с предстоящими выборами или каким-то событием, выяснением потребительских предпочтений и т.д.). Рекомендуем вам не уклоняться от участия в таких работах, тем более что, как правило, за проведение опросов выплачивается вознаграждение. Подчеркнем важность выполнения инструкции, что позволяет в массовых работах, в которых участвуют десятки, а то и сотни людей, обеспечить единообразие данных. При проведении контроля собранных данных пользуйтесь схемами логической и арифметической увязки признаков. При определении репрезентативности собранных данных проведите их сопоставление с данными официальной статистики.

Участие в статистическом наблюдении позволяет выявить ваше умение организовать свою работу, следовать инструктивным указаниям, выполнять работы в срок. Все эти составляющие обеспечивают успех слаженной работы многих исполнителей статистического наблюдения.

Статистические показатели

3.1. Основные положения

В данной теме рассматриваются сущность статистического показателя, его атрибуты, классификация статистических показателей, их функции и принципы построения, вводится понятие системы статистических показателей.

Статистический показатель — это обобщающая характеристика какого-либо свойства совокупности, группы. Этим он отличается от характеристик индивидуальных свойств отдельных единиц совокупности, которые, как отмечалось, называются признаками. Например, наличие собственного автомобиля — это признак конкретного человека. Для характеристики уровня автомобилизации населения рассчитывается показатель «число автомобилей на 1000 человек населения». Так, на 1 января 2005 г. в РФ на 1000 человек населения приходилось 159 собственных легковых автомобилей.

Статистический показатель имеет несколько существенных признаков (рис. 3.1). Для нашего примера:

- качественная сторона показателя — наличие собственных легковых автомобилей;
- количественная сторона — 159 автомобилей на 1000 человек;
- территориальные границы — Российская Федерация;
- временные границы — 1 января 2005 г.

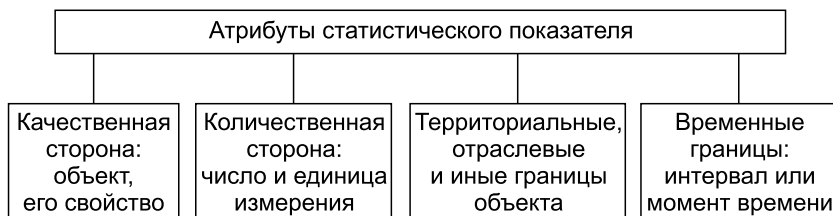


Рис. 3.1. Атрибуты статистического показателя

Статистические показатели разнообразны по своим задачам и способам построения.

По качественной стороне выделяют:

- показатели свойств конкретных объектов;
- показатели статистических свойств любых массовых явлений и процессов.

Рассмотрим особенности этих показателей.

Пример 3.1. В табл. 3.1 приведены данные, характеризующие стоимость основных фондов в РФ, их структуру и динамику.

Таблица 3.1

Основные фонды (на начало года; по полной учетной стоимости)

Год	Всего по отраслям экономики, млн руб.	Структура, %		К предыдущему году (в сопоставимых ценах), %		
		Отрасли, производящие товары	Отрасли, оказывающие рыночные и нерыночные услуги	Всего по отраслям экономики	Отрасли, производящие товары	Отрасли, оказывающие рыночные и нерыночные услуги
1995	5182040	50,3	49,7	99,8	99,0	101,6
2000	16605251	33,7	66,3	100,1	99,3	100,9
2001	20241428	30,8	69,2	100,4	99,6	101,1
2002	24430544	34,0	66,0	100,6	99,9	101,3
2003	30329106	29,1	70,9	100,7	99,7	101,5
2004	32501833	29,2	70,8	100,7	99,5	101,6
2005	34974998	29,4	70,6	100,9	99,7	101,9

Источники: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 336.

Показатели стоимости основных фондов (в млн руб.) относятся к группе показателей свойств конкретных объектов. Особенность показателей этой группы состоит в том, что их качественное содержание определяется конкретной предметной наукой. Экономическая теория, например, определяет основные фонды как производственные активы (здания, сооружения, машины, технические средства и т.д.), подлежащие использованию неоднократно или постоянно в течение длительного периода, но не менее одного года, для производства товаров и оказания услуг. Статистика вырабатывает методы учета и расчета количественной стороны этого показателя. Так, метод учета основных

фондов по полной учетной стоимости предполагает определение стоимости основных фондов как суммы учитываемых в бухгалтерских балансах организаций остаточной балансовой стоимости основных фондов и величины накопления износа, т.е. без учета утраты потребительских свойств основных фондов в процессе эксплуатации. Для учета основных фондов могут использоваться и другие виды стоимостной оценки (восстановительная, первоначальная, балансовая и т.д.).

Все остальные показатели, содержащиеся в табл. 3.1 (удельный вес основных фондов отдельных отраслей, показатели изменения стоимости по годам) по своей форме относятся к показателям статистических свойств. Они могут быть определены для любых массовых явлений и процессов. К таким показателям также относятся средние величины, показатели вариации, показатели связи, оценки надежности и точности статистических прогнозов и др. Статистика разрабатывает форму и методы расчета этих показателей, их свойства, сферу применения, особенности интерпретации, которые не зависят от конкретного содержания того или иного явления.

Еще одно основание для классификации статистических показателей — это способ отражения их количественной стороны. В этом отношении все показатели подразделяют на:

- абсолютные;
- относительные.

Абсолютные показатели — это показатели, отражающие либо суммарное число единиц, либо суммарное свойство объекта.

Пример 3.2. В табл. 3.2 приведены некоторые показатели деятельности сельскохозяйственных организаций в РФ.

Абсолютные показатели чаще всего выражаются именованными величинами в натуральных единицах измерения. *Натуральными* называют такие единицы измерения, которые выражают величину объектов, явлений в физических мерах веса, объема, длины, площади и т.д. Так, посевная площадь измеряется в гектарах, производство зерна — в тоннах, численность работников — в человеках и т.д.

В ряде случаев используют *условно-натуральные единицы*. Их применяют для определения суммарного объема признака разных единиц, обладающих одним потребительским назначением. Тогда одну из единиц принимают за эталон, а другие пересчитываются с помощью специальных коэффициентов в меру

Таблица 3.2

Основные показатели деятельности сельскохозяйственных организаций

Показатель	2000	2001	2002	2003	2004
Число сельскохозяйственных организаций (на конец года), тыс.	27,6	24,8	24,2	22,1	20,6
Среднегодовая численность работников, млн чел.	4,7	4,2	3,8	3,3	2,9
Продукция сельского хозяйства (в фактически действовавших ценах), млрд руб.	314,7	397,2	386,6	429,6	535,4
Число убыточных сельскохозяйственных организаций, тыс.	14,1	11,4	13,3	10,8	7,2
Посевная площадь, млн га	69,1	66,4	64,6	58,2	55,4
в том числе зерновых культур	37,8	37,5	36,6	31,3	30,9
Производство зерна, млн т	55,7	69,5	69,3	51,4	56,4

Источник: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 444.

этого эталона. Чтобы обобщить мощность двигателей, ее выражают в лошадиных силах. Для соизмерения разных видов топлива применяют условно-натуральную единицу — условное топливо (теплосодержание 1 кг условного топлива принято равным 29,3076 МДж). В какой-нибудь один продукт пересчитываются некоторые продукты питания для получения общего объема потребленных продуктов. Например, разные мясопродукты пересчитываются на мясо, молочные продукты — на молоко, хлебные продукты и макаронные изделия — на муку. Пересчет может осуществляться по усваиваемой доле белков, жиров, углеводов, калорийности продуктов.

В ряде случаев используют составные единицы измерения. Так, чтобы подсчитать общий объем работы транспорта, складывают либо тонно-километры перевезенных грузов (для грузового транспорта), либо пассажиро-километры (для пассажирского транспорта).

Широко используются в статистике *стоимостные* единицы измерения. Они позволяют соизмерить самые разнородные величины, за что получили название универсального соизмерителя. В стоимостных единицах определяется объем продукции предприятий и отраслей, валовой внутренний продукт, национальное богатство и т.д.

Для измерения объемов затрат труда, оценки эффективности использования трудовых ресурсов используются *трудовые* единицы измерения — человеко-час, человеко-день, человеко-смена.

Статистический метод не ограничивается характеристикой изолированных свойств объектов с помощью абсолютных показателей. Статистика находит их соотношения, характеризует изменения во времени, взаимосвязи между собой и с окружающей средой и т.д. Эти задачи решаются с помощью относительных показателей.

Относительные показатели — это показатели, полученные путем сравнения, сопоставления абсолютных или относительных показателей в пространстве (между объектами), во времени (по одному и тому же объекту) или сравнения показателей разных свойств изучаемого объекта.

Относительные показатели можно подразделить на следующие группы.

1. *Относительные показатели структуры* — рассчитываются как отношение части к целому. Эти показатели чаще называют долей, удельным весом и выражают в долях единицы или в процентах.

На основе данных табл. 3.2 можно вычислить:

- удельный вес убыточных организаций в общей численности сельскохозяйственных организаций. В 2000 г. эта величина составляла

$$\frac{14,1}{27,6} = 0,51, \text{ или } 51\%;$$

- удельный вес посевной площади, занятой зерновыми культурами, в общей посевной площади. В 2000 г. он составил

$$\frac{37,8}{69,1} = 0,55, \text{ или } 55\%.$$

К показателям структуры относятся также характеристики степени неравномерности долей (например, коэффициент неравномерности распределения).

2. *Относительные показатели динамики* — показатели, характеризующие изменения во времени. К относительным показателям динамики принадлежат темпы роста, темпы прироста, параметры уравнений трендов, коэффициенты колеблемости и устойчивости, индексы, обобщающие показатели структурных сдвигов. Большинство из них мы рассмотрим в следующих разделах курса. Остановимся на самых простых.

По данным табл. 3.2 охарактеризуем динамику посевных площадей в России за период 2000—2004 гг., рассчитав абсолютные приросты (разность абсолютных показателей за отдельные годы), темпы роста (отношение абсолютных показателей), темпы прироста. Можно провести расчет показателей динамики двумя способами:

1) размер посевных площадей каждого года сопоставить с размером посевных площадей в предыдущем году;

2) размер посевных площадей каждого года сопоставить с начальным уровнем динамического ряда, т.е. с 2000 г.

В зависимости от базы сравнения различают:

- относительные величины динамики с переменной базой сравнения — *цепные*;

- относительные величины динамики с постоянной базой сравнения — *базисные*.

Рассчитаем по нашим данным цепные и базисные показатели динамики размера посевных площадей.

Абсолютное изменение размера посевных площадей в 2001 г. по сравнению с 2000 г. составило: $66,4 - 69,1 = -2,7$ млн га. Если абсолютное изменение отрицательно, его называют *абсолютным сокращением*.

$$\text{Темп роста: } \frac{66,4}{69,1} \cdot 100\% = 96,1\%.$$

$$\text{Темп прироста: } 100\% - 96,1\% = 3,9\%.$$

Таким образом, посевные площади в РФ в 2001 г. по сравнению с 2000 г. сократились на 2,7 млн га, или на 3,9%.

Абсолютное изменение в 2002 г. по сравнению с 2001 г. составило: $64,6 - 66,4 = -1,8$ млн га; по сравнению с 2000 г.: $64,6 - 69,1 = -4,5$ млн га.

Темп роста в 2002 г. по сравнению с 2001 г. составил $(64,6 : 66,4) \cdot 100\% = 97,3\%$, по сравнению с 2000 г. — $(64,6 : 69,1) \cdot 100\% = 93,5\%$ и т.д. Результаты вычисления показателей динамики оформлены в табл. 3.3.

Таблица 3.3

Динамика посевных площадей в РФ

Показатель	2001	2002	2003	2004
Посевная площадь, млн га	66,4	64,6	58,2	55,4
Абсолютное сокращение, млн га:				
по сравнению с предыдущим годом	2,7	1,8	6,4	2,8
по сравнению с 2000 г.	2,7	4,5	10,9	13,7
Темп роста, в процентах				
к предыдущему году	96,1	97,3	90,1	95,2
к 2000 г.	96,1	93,5	84,2	80,2
Темп прироста, в процентах				
к предыдущему году	3,9	2,7	9,9	4,8
к 2000 г.	3,9	6,5	15,8	19,8

За период с 2000 по 2004 г. посевные площади в РФ сократились на 19,8%, что в абсолютном выражении составило 13,7 млн га. Наибольшее сокращение площадей за рассматриваемый период произошло в 2003 г.

Существуют особенности вычисления и интерпретации показателей динамики, вычисляемых на основе абсолютных показателей, имеющих стоимостное выражение. Недостатком этого рода единиц измерения является то, что с течением времени цены изменяются, поэтому стоимостные оценки, например объемов продукции, становятся несоизмеримыми. Показатели динамики, рассчитанные на основе фактически действовавших (текущих) цен, будут не только отражать изменение объемов производства, но и содержать инфляционную компоненту. Этот недостаток преодолевается статистикой либо путем применения *сопоставимых цен*, т.е. переоценки стоимостных показателей в цены одного периода, либо путем расчета наряду с *номинальными* темпами роста (темпы роста показателей в текущих ценах) *реальных* темпов роста, что достигается с помощью специального приема — «*дефлятирования*», т.е. деления номинальных темпов роста на соответствующий индекс цен (индекс-дефлятор).

Так, если мы рассчитаем темп роста продукции сельского хозяйства в фактически действовавших ценах в 2001 г. по сравнению с 2000 г., он составит

$$\frac{397,2}{314,7} \cdot 100\% = 126,2\%,$$

т.е. стоимость продукции сельского хозяйства в текущих ценах выросла на 26,2%. Однако эта величина отражает не только изменение объемов сельскохозяйственного производства в физических единицах измерения (рост производства зерна, овощей, картофеля в тоннах, скота и птицы в тоннах убойного веса, яиц в штуках и т.д.), но и изменение цен на сельхозпродукцию. Для исключения влияния ценового фактора необходимо разделить рассчитанный темп роста на индекс цен. В 2001 г. индекс цен на сельскохозяйственную продукцию составил 125,2%. Тогда реальный темп роста продукции сельского хозяйства составит

$$\frac{126,2}{125,2} \cdot 100\% = 100,8\%.$$

Этот показатель называют *индексом физического объема продукции*. Он показывает, что продукция сельского хозяйства в физических единицах измерения в 2001 г. по сравнению с 2000 г. выросла всего на 0,8%.

3. *Относительные показатели, характеризующие соотношения разных признаков одного и того же объекта между собой*, — обобщают вторичные признаки: выработка — отношение произведенной продукции к затратам труда; фондоотдача — отношение стоимости произведенной продукции к среднегодовой стоимости фондов; материалоемкость — отношение стоимости материальных затрат (без амортизации) к стоимости произведенной продукции и т.д.

Все показатели этой группы имеют двойные единицы измерения (руб. продукции на 1 руб. фондов, центнеров с 1 га и т.д.) и применяются для оценки эффективности или интенсивности использования тех или иных видов ресурсов.

По данным табл. 3.2 можно рассчитать:

- выработку продукции на одного занятого в сельском хозяйстве. В 2000 г. она составила

$$314,7 \text{ млрд руб.} : 4,7 \text{ млн чел.} = 66,958 \text{ тыс. руб./чел.};$$

- урожайность зерновых культур — сбор зерна с 1 га посевной площади. В 2000 г. урожайность составила

$$55,7 \text{ млн т} : 37,8 \text{ млн га} = 1,47 \text{ т/га} = 14,7 \text{ ц/га}.$$

Отметим, что показатель урожайности корректнее рассчитывать как производство зерна не на 1 га посевных площадей, а на 1 га убранных площадей (в силу разных причин эти площади могут существенно отличаться). Тем самым мы обеспечим большее соответствие абсолютных показателей и, следовательно, большую объективность относительного показателя. Для сравнения укажем, что по расчетам Росстата в 2000 г. урожайность зерновых составляла 15,6 ц с 1 га убранной площади.

4. *Относительные показатели сравнения разных объектов по одинаковым признакам* — применяются для межотраслевых, межрегиональных, международных сравнений. Главные условия их построения — обеспечение единства методологии, сопоставимость единиц измерения, временных границ.

Сравним урожайность зерновых культур в России и США. В США урожайность в 2000 г. составляла 58,1 ц с 1 га убранной площади. Тогда показатель сравнения будет рассчитан как $15,6 : 58,1 = 0,269$, или 26,9%, т.е. урожайность зерновых в России составляла 26,9% урожайности зерновых в США. Одна из причин такого различия объясняется существенными различиями в структуре выращиваемых зерновых культур.

5. Особый вид относительных показателей — *отношения фактически наблюдаемых величин признака к его плановым, нормативным, оптимальным или максимально возможным величинам*. К этой группе показателей относятся показатели планового задания, выполнения плана, выполнения норм выработки, норм расхода материалов и других ресурсов.

Относительная величина планового задания (ОВПЗ) — это отношение уровня, запланированного на текущий период ($y_{пл}$), к уровню, достигнутому в прошлом периоде (y_0):

$$ОВПЗ = \frac{y_{пл}}{y_0} \cdot 100\%.$$

Относительная величина выполнения плана (ОВВП) — отношение фактического уровня, достигнутого в текущем периоде (y_1), к запланированному уровню ($y_{пл}$):

$$ОВВП = \frac{y_1}{y_{пл}} \cdot 100\%.$$

Произведение ОВПЗ и ОВВП, выраженных в коэффициентах, дает *относительную величину динамики* (коэффициент роста) – отношение фактических уровней текущего периода (y_1) и прошлого периода (y_0):

$$\text{ОВД} = \text{ОВПЗ} \cdot \text{ОВВП}, \text{ или } \text{ОВД} = \frac{y_1}{y_0} \cdot 100\%.$$

Пример 3.3. Имеются данные о выпуске и отгрузке продукции машиностроительного предприятия за два полугодия (табл. 3.4).

Таблица 3.4

Данные о выпуске продукции

Показатель	1-е полугодие, фактически y_0	2-е полугодие	
		по бизнес- плану $y_{пл}$	фактически y_1
Выпуск продукции, млн руб.	500	550	540
в том числе отгружен- ная продукция	400	450	500

Определим ОВПЗ, ОВВП и ОВД по выпуску продукции.

ОВПЗ по выпуску продукции равна: $\frac{550}{500} = 1,1$, или 110%, т.е.

планом на 2-е полугодие предусматривался рост продукции по сравнению с 1-м полугодием на 10%.

ОВВП по выпуску продукции составила: $\frac{540}{550} = 0,982$, или 98,2%. Фактический выпуск продукции во 2-м полугодии составил 98,2% запланированного уровня, т.е. план был невыполнен на 1,8%.

ОВД или темп роста выпуска продукции во 2-м полугодии по сравнению с 1-м полугодием составила: $\frac{540}{500} = 1,08$, или 108%.

Эту величину можно было найти и другим способом: $1,1 \cdot 0,982 = 1,08$, или 108%.

6. *Относительные величины координации* – показатели, отражающие соотношение частей целого между собой. Поскольку

показатели дают некоторое представление о структуре совокупности, их нередко считают относительными величинами структуры. Однако их функция не характеристика структуры совокупности, а определение соотношений между ее отдельными частями. Относительные величины координации показывают, сколько единиц одной части целого приходится на 1, на 100, на 1000 и т.д. единиц другой части.

Пример 3.4. Имеются данные о численности населения в двух регионах (табл. 3.5).

Таблица 3.5

Численность населения на 1 января 2006 г., тыс. жителей

Регион	Всего	В том числе	
		мужчины	женщины
Ленинградская область	1644	756	888
Санкт-Петербург	4581	2055	2526

Рассчитаем относительные величины координации, характеризующие соотношение численности мужчин и женщин в двух регионах РФ – Ленинградской области и Санкт-Петербурге:

$$\frac{888}{756} \cdot 100 = 118, \text{ т.е. на начало 2006 г. в Ленинградской области на}$$

$$100 \text{ мужчин приходилось } 118 \text{ женщин; } \frac{2526}{2055} \cdot 100 = 123, \text{ следова-}$$

тельно, разрыв в численности населения по полу в Санкт-Петербурге на начало 2006 г. был еще более существенен: на 100 мужчин приходилось 123 женщины.

7. *Относительные величины интенсивности* – показатели, характеризующие степень распространения какого-либо явления в определенной сфере либо степень насыщенности среды данным явлением. Они рассчитываются как отношение явления, степень распространения которого изучается, к объему той среды, в которой происходит развитие этого явления.

К этой группе показателей относятся, например, показатель выбросов загрязняющих веществ в атмосферу на 1 км² территории, показатель числа больничных коек на 10000 жителей, число книг в библиотеках на 100 жителей, коэффициент мла-

денческой смертности, показывающий, сколько детей из 1000 родившихся не доживает до возраста одного года, и др.

8. *Относительные величины уровня экономического развития* — показатели, характеризующие результаты экономической деятельности (производства, накопления, потребления и т.д.) в расчете на душу населения. Например, производство зерна на душу населения, валовой региональный продукт на душу населения, валовой национальный доход на душу населения и т.д. Все эти показатели получены делением показателей объемов производства и других результатов деятельности на среднегодовую численность населения. Среднегодовая численность населения определяется обычно как полусумма численности населения на начало и конец года.

По отношению к характеризваемому свойству статистические показатели разделяют на *прямые* и *обратные*. Так, для оценки производительности труда используют либо показатель выработки продукции на единицу трудозатрат, либо показатель трудоемкости единицы продукции. Выработка — это прямой показатель производительности, поскольку с ростом выработки производительность труда растет. Трудоемкость — обратный показатель производительности: с ростом трудоемкости производительность падает.

Пример 3.5. Известны следующие данные по предприятию (табл. 3.6).

Таблица 3.6

Показатели деятельности предприятия

Показатель	Январь	Февраль
Выпуск продукции, тыс. изделий	540	700
Затраты труда, тыс. чел./ч	51,2	64,0

Найдем показатели выработки за каждый месяц:

- в январе: $\frac{540}{51,2} = 10,5$ изделия в час;
- в феврале: $\frac{700}{64,0} = 10,9$ изделия в час.

Показатели трудоемкости продукции составят:

- в январе: $\frac{51,2}{540} = 0,095$ ч на одно изделие;
- в феврале: $\frac{64,0}{700} = 0,091$ ч на одно изделие.

Рост показателя выработки и снижение показателя трудоемкости свидетельствуют о росте производительности труда на предприятии.

3.2. Задания по теме

1. Из статистического сборника «Российский статистический ежегодник» выпишите несколько абсолютных показателей, имеющих разные единицы измерения: натуральные (простые и составные), условно-натуральные, стоимостные.

2. С помощью каких абсолютных показателей характеризуется рынок труда? Из статистических сборников выпишите абсолютные величины, характеризующие рынок труда в вашем регионе и в России в целом за несколько последних лет. Оформите данные в виде таблицы. Какие относительные показатели могут быть дополнительно рассчитаны на основе имеющихся у вас данных? Рассчитайте их и напишите аналитическую записку о рынке труда в вашем регионе и в стране в целом, проведите сравнительный анализ. Укажите, какие виды относительных показателей вы использовали.

3. В чем отличие статистического показателя от статистического признака? Какими показателями можно измерить: успеваемость студентов группы, уровень загрязнения атмосферного воздуха в городе, качество выпускаемой продукции предприятий отрасли, жилищные условия горожан?

4. Используя данные табл.3.1, выполните следующие задания:
а) объясните, почему для оценки динамики стоимости фондов в таблице использованы показатели, вычисленные на основе сопоставимых цен;

б) определите для каждого года, сколько рублей стоимости основных фондов отраслей, оказывающих услуги, приходилось на 1 тыс. руб. стоимости основных фондов отраслей, производящих товары. К какой группе относительных показателей от-

носятся рассчитанные вами величины? Будут ли они сопоставимы в динамике;

в) напишите аналитическую записку о тенденциях изменения стоимости и отраслевой структуры основных фондов в РФ, используя дополнительно рассчитанные показатели.

5. Для характеристики уровня занятости населения предлагается использовать отношение численности занятого населения к численности:

а) всего населения;

б) населения в трудоспособном возрасте;

в) экономически активного населения;

г) населения в экономически активном возрасте (от 15 до 72 лет).

Какой из предложенных вариантов, на ваш взгляд, наиболее правильный? Обоснуйте свой выбор.

6. По данным табл. 3.4 рассчитайте относительные показатели планового задания, выполнения плана и динамики по отгруженной продукции. Покажите взаимосвязь вычисленных показателей.

7. Имеются данные о грузообороте транспорта общего пользования в России, млрд. т · км (табл. 3.7).

Таблица 3.7

Показатель	2000	2001	2002	2003	2004
Грузооборот транспорта общего пользования:					
железнодорожного	1373	1434	1510	1669	1902
автомобильного	153	160	167	178	182
морского	100	94	93	65	47
внутреннего водного	65	76	73	71	78
трубопроводного	745	797	896	1003	1116

И с т о ч н и к : Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 493.

Определите:

а) общий объем грузооборота за отдельные годы;

б) показатели структуры грузооборота для каждого года;

в) показатели динамики грузооборота отдельных видов транспорта (цепные и базисные).

Проанализируйте полученные результаты.

8. В табл. 3.8 приведены данные об общедоступных библиотеках Санкт-Петербурга и Ленинградской области на начало 2003 г. и 2006 г.

Таблица 3.8

Показатель	Санкт-Петербург		Ленинградская область	
	2003	2006	2003	2006
Число библиотек	342	286	500	470
Количество книг, млн экз.	52,8	52,6	8,7	8,1
Всего читателей, тыс. чел.	1283	1354	592	557

Кроме того, известно, что численность населения Санкт-Петербурга на начало 2003 г. составляла 4657 тыс. чел., 2006 г. — 4581 тыс. чел.; численность населения Ленинградской области на начало 2003 г. — 1667 тыс. чел., 2006 г. — 1644 тыс. чел.

Проанализируйте развитие библиотечного дела в указанных регионах, для чего дополните абсолютные показатели, приведенные в таблице, относительными показателями: число читателей на одну библиотеку, число читателей на 1000 жителей, число книг на одну библиотеку, число книг на 100 жителей. При написании аналитической записки используйте также показатели сравнения и динамики.

Подумайте, какие еще показатели вы бы предложили для характеристики библиотечного дела. Напишите ваши предложения.

9. Известны некоторые показатели деятельности товарных бирж в России в 2004 г.*:

- число бирж (на конец года) — 35;
- проведено торгов — 1,9 тыс.;
- заключено сделок — 28,5 тыс.;

* Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 555.

- среднесписочная численность работающих (без совместителей) и работников несписочного состава — 0,3 тыс. чел.;
- численность совместителей — 0,07 тыс. чел.;
- начислено на оплату труда работникам списочного состава (включая совместителей) — 26,6 млн руб.;
- платежи в бюджет — 13,2 млн руб., из них налог на прибыль — 1,5 млн руб.;
- биржевой оборот — 6346 млн руб.,
в том числе:
- потребительские товары — 5092 млн руб.;
- продукция производственно-технического назначения — 972 млн руб.;
- прочие — 282 млн руб.

Какие относительные показатели деятельности товарных бирж можно рассчитать на основе этих данных? Выполните вычисления и укажите, к каким видам относительных величин относятся рассчитанные вами показатели. Какие еще показатели вы бы включили в систему показателей для характеристики деятельности товарных бирж? Если вам не хватает данных, обратитесь к статистическим сборникам или используйте другие источники информации.

Напишите аналитическую записку по итогам вашей работы.

10. Подберите данные о производстве важнейших видов энергоресурсов в развитых странах мира. В каком случае эти показатели будут сопоставимы? Рассчитайте относительные показатели сравнения и уровня экономического развития.

Для выполнения задания воспользуйтесь статистическими сборниками, выпускаемыми Росстатом («Российский статистический ежегодник», «Россия и страны мира» и др.).

11. Сравните фактическое потребление продуктов питания на душу населения в России с нормативными значениями (табл. 3.9). Напишите краткие выводы.

Как вы думаете, почему нормы, закладываемые в расчет бюджета прожиточного минимума, отличаются от рациональных норм потребления?

12. Известны данные о выпуске плодово-ягодных консервов предприятием за два месяца (табл. 3.10).

Определите, как изменился общий объем выпуска консервной продукции предприятия за месяц.

Таблица 3.9

Продукты питания	Фактическое потребление на душу населения, кг/год		Нормы потребления на душу населения, кг/год	
	2001	2004	рациональные, разработанные АМН РФ	закладываемые в бюджет прожиточного минимума
Хлеб и хлебобродуцкты	120	119	110	130,8
Картофель	122	128	110	124,2
Мясо и мясопродукты	43	49	78	26,6

Таблица 3.10

Продукция	Масса нетто продукции в физической банке, г	Выпущено физических банок, шт.*	
		июль	август
Варенье вишневое	750	—	2400
Джем абрикосовый	400	2800	1040
Конфитюр клубничный	360	3500	1300
Конфитюр вишневый	360	1500	3000
* За условную банку консервов, учитываемых по массе, принимается банка массой 400 г.			

3.3. Вопросы для самоконтроля

1. Что такое статистический показатель? Чем он отличается от статистического признака?

2. Каковы основные атрибуты статистического показателя? Разберите их на конкретном примере.

3. Чем отличаются конкретные показатели природных, социально-экономических и технологических массовых явлений от показателей общих статистических свойств любых массовых явлений?

4. Какие показатели называют абсолютными? В каких единицах измерения они выражаются?

5. Для чего в статистике применяются условно-натуральные единицы измерения?

6. Почему стоимостные единицы измерения считаются универсальными?

7. Дайте определение относительных показателей и назовите их виды. Приведите примеры показателей каждой группы.

8. Почему нельзя провести анализ того или иного явления, используя только абсолютные показатели? Приведите конкретный пример.

9. Можете ли вы согласиться с утверждением, что «относительные показатели более информативны»? Обоснуйте свой ответ.

10. Опишите алгоритм расчета показателей структуры, интенсивности, координации, сравнения. Охарактеризуйте функции этих показателей.

11. Как связаны между собой относительные величины планового задания, выполнения плана и динамики?

12. В каких единицах измерения рассчитываются относительные показатели?

13. Почему при расчете относительных показателей на основе абсолютных, имеющих одинаковые единицы измерения, в одних случаях итоговую относительную величину выражают в коэффициентах, а в других – в процентах или промилле? Приведите примеры.

14. Почему важно использовать абсолютные и относительные показатели во взаимосвязи? Охарактеризуйте понятие системы статистических показателей. Приведите примеры жестко детерминированной и статистической взаимосвязи между показателями.

15. Какие функции выполняют системы показателей в управлении экономикой, в социальной жизни, в науке?

3.4. Методические рекомендации преподавателям

Аудиторная работа. Основное время на практических занятиях необходимо уделить вычислению отдельных видов относительных показателей. При подборе упражнений для практических занятий предпочтение должно быть отдано задачам, в которых рассматриваются системы статистических показателей. При этом показатели могут находиться как в жестко де-

терминированной, так и в статистической связи. Обсуждая конкретные примеры, необходимо подчеркивать, что свойства изучаемых статистических объектов (совокупностей, процессов) не изолированы, а связаны между собой. Поэтому и показатели этих свойств должны рассматриваться в системе. Кроме того, необходимо понимать, что любая система статистических показателей строится для решения конкретной аналитической задачи. В зависимости от той или иной ситуации система показателей может быть расширена или, наоборот, сужена. Могут выделяться основные и дополнительные показатели. Но не должно быть показателей, рассчитанных просто так, «на всякий случай». На конкретных примерах рекомендуется разобрать функции статистических показателей и их систем: познавательную-информационную, оценочную, прогностическую, рекламно-пропагандистскую.

Отдельного обсуждения заслуживают принципы построения относительных статистических показателей. Во-первых, всегда надо добиваться как можно большего соответствия по смыслу сравниваемых величин. Во-вторых, при построении относительного показателя сравниваемые величины могут различаться только одним атрибутом: или видом признака (при одинаковом объекте, периоде времени, плановом или фактическом характере показателей), или временем (при том же признаке, объекте и т.д.), или только фактическим, плановым или нормативным характером показателя (тот же объект, признак, время) и т.д. И наконец, необходимо знать возможные границы существования относительного показателя. Особая методика построения показателей необходима в тех случаях, когда сравниваемые величины имеют разные знаки (прибыль — убыток предприятия, прирост — убыль населения и т.п.) или одна из них имеет нулевое значение.

Решение задач по теме не должно сводиться только к расчету тех или иных показателей. Важно научить студентов правильно выбирать единицы измерения вычисляемых величин, грамотно интерпретировать полученные результаты. Заключительным этапом выполнения заданий должно быть написание кратких пояснений и содержательных выводов.

Самостоятельная внеаудиторная работа. Студентам может быть предложена конкретная тема исследования (например,

«Рождаемость и смертность в Российской Федерации», «Сравнительный анализ уровня жизни в отдельных странах», «Страховой рынок в России», «Миграция в России», «Роль фермерского хозяйства в сельскохозяйственном производстве», «Развитие высшего образования в России» и др.) и рекомендовано подобрать статистические данные по теме исследования из статистических сборников. Следующими этапами работы должны стать разработка и расчет системы показателей для оценки того или иного явления или процесса и написание итогового отчета. При оформлении результатов рекомендуется использовать таблицы и графики. Обязательным условием является интерпретация рассчитанных показателей и формулировка выводов. Поскольку работа достаточно трудоемка, ее можно организовать в малых группах, предложив студентам самостоятельно определить степень участия каждого.

Особое внимание необходимо уделить подбору тем. Они должны быть актуальны, иметь самостоятельную познавательную ценность, соответствовать уровню подготовки студентов и, самое главное, быть хорошо обеспечены информацией. Можно рекомендовать темы, ориентированные на использование данных региональной статистики и отражающие социально-экономические особенности конкретного региона.

Упрощенным вариантом самостоятельной работы может быть выполнение домашнего задания, состоящего из решения небольших задач на расчет различных относительных показателей и их интерпретацию.

Итоговый контроль может быть организован в форме тестирования либо задания письменно ответить на один из вопросов по теме. Например, «Почему анализ экономических или социальных явлений должен быть основан не на отдельных показателях, а на системе показателей?», «Какие функции выполняют статистические показатели?».

Если студенты в ходе самостоятельной внеаудиторной работы выполняли индивидуальные исследовательские задания, итоговый контроль можно не проводить. В этом случае целесообразнее уделить внимание обсуждению студенческих отчетов.

3.5. Методические указания студентам

Освоение теории. В этой теме раскрывается понятие статистического показателя, рассматриваются функции статистических показателей и принципы их построения. Знакомство с содержанием, формой, свойствами и сферой применения конкретных видов статистических показателей позволит вам в дальнейшем корректно применять статистические показатели в анализе социально-экономических явлений и процессов, лучше ориентироваться в потоке статистической информации.

Особое внимание обратите на принцип системного использования показателей. Необходимость рассмотрения исследуемого объекта во всех его связях и отношениях приводит к тому, что для получения целостной статистической характеристики изучаемых явлений применяют системы статистических показателей.

Изучая классификацию статистических показателей, обязательно подберите свои примеры различных видов показателей.

Практические навыки. В результате изучения темы необходимо научиться рассчитывать различные относительные показатели (структуры, сравнения, интенсивности, уровня экономического развития, координации, планового задания и выполнения плана) и давать им правильную интерпретацию.

При расчете показателей не забывайте о единицах измерения. Любой статистический показатель — это не абстрактное математическое число, а количественная характеристика конкретного социально-экономического явления или процесса, связанная к конкретному месту и времени. Будьте особенно внимательны при указании единиц измерения относительных величин. Если относительная величина получается в результате соотношения разноименных абсолютных показателей, единица измерения будет представлять собой сочетание наименований показателей, участвующих в расчете (например, урожайность измеряется в центнерах с гектара, водообеспеченность — в кубических метрах воды на душу населения, плотность населения — в человеках на квадратный километр и т.д.). При сравнении одноименных абсолютных показателей получают относительные показатели, выраженные в коэффициентах, процентах, промилле. Иногда прибегают к отношениям, выраженным на 10000 и 100000. Всегда старайтесь придать величине показателя удобный для восприятия вид, избегайте дробных чисел с большим числом знаков после запятой.

Тема **4**

Представление статистических данных: таблицы и графики

4.1. Основные положения

В данной теме описаны простейшие приемы построения сводных таблиц и графического представления статистических данных с помощью доступного всем табличного процессора Excel.

Пример 4.1. Составим таблицу, содержащую данные о распространении различных видов информационных коммуникаций в отдельных странах мира. Исходные данные приведены в табл. П1.2*. Для выполнения задания нам нужны данные о магистральных телефонных линиях, числе абонентов сотовой связи и числе пользователей сети Интернет в странах мира.

Любая статистическая таблица по своей структуре состоит из трех частей:

- заголовок таблицы;
- шапка таблицы;
- информационная часть.

В заголовке таблицы обычно указываются объект изучения, территория и время, к которым относятся данные. Шапкой таблицы называют систему заголовков граф таблицы. Заголовки граф содержат названия показателей (без сокращений, исключения составляют общепринятые сокращения, например ВВП, ВНД, ИРЧП и т.п.) и их единицы измерения. Последние могут указываться и в заголовке таблицы, если все показатели таблицы выражены в одних единицах измерения.

Разработаем макет электронной таблицы (табл. 4.1).

Сформируем заголовок и шапку таблицы на рабочем листе Excel. Установите курсор мыши в ячейку **A1** и введите с клави-

* Буква П обозначает, что таблица находится в приложении.

Таблица 4.1

**Распространение информационных коммуникаций
в странах мира, 2003 г.**

Страна	Магистральные телефонные линии на 1 тыс. чел.	Число абонентов сотовых сетей на 1 тыс. чел.	Число пользовате- лей сети Интернет на 1 тыс. чел.

атуры заголовков. Нажмите **Enter**. Введенный текст занял несколько ячеек **A1:D1**. Однако это расположение лишь на экране монитора, поскольку ширина введенного текста превышает ширину ячейки **A1**. Переведите курсор, например, в ячейку **C1** и убедитесь, что в строке состояния, где отображается содержимое ячейки, ничего не отражено.

В строке № 2 сформулируйте название граф таблицы, при этом текст заголовков должен быть помещен в одну ячейку. Для этого выделите строку № 2, установив на номере строки указатель мыши и щелкнув левой кнопкой. Активизируйте меню **Формат/Ячейки**, на вкладке **Выравнивание** поставьте флажок **Переносить по словам**. Нажмите **ОК**. Снимите выделение со строки, щелкнув левой кнопкой мыши по любой свободной ячейке рабочего листа. Введите в ячейки строки № 2 соответствующие тексты:

- в ячейку A2 – Страна;
- B2 – Магистральные телефонные линии (на 1 тыс. чел.);
- B3 – Число абонентов сотовых сетей (на 1 тыс. чел.);
- B4 – Число пользователей сети Интернет (на 1 тыс. чел.).

Большинство операций по оформлению таблицы сосредоточено в меню **Формат/Ячейки** на соответствующих вкладках. Кроме того, часто используемые операции (цвет шрифта, цвет заливки, выравнивание, границы и др.) выведены пиктограммами на панели инструментов **Форматирование**.

Для того чтобы оформить таблицу в соответствии с разработанным шаблоном, выполните следующие операции:

- увеличьте ширину ячеек и соответственно всего столбца (подведите указатель мыши к границе индексов столбцов, при появлении двунаправленной стрелки, удерживая левую кнопку мыши, расширьте столбец на необходимую величину);

- выделяя соответствующие ячейки, измените размер шрифта заголовка и названий граф таблицы (используйте шрифт Times New Roman размер 12);

- с помощью пиктограмм выравнивания текста установите для заголовка первой графы выравнивание **По левому краю**, для остальных граф – выравнивание **По центру**.

Для удобства последующей работы с таблицей рекомендуется выделить шапку таблицы цветным тоном (пиктограмма **Цвет заливки**).

После оформления шапки таблицы приступите к вводу исходных данных.

После ввода данных на заключительном этапе, выделяя соответствующие группы ячеек в шапке и информационной части таблицы, с помощью пункта меню **Формат/Ячейки** (закладка **Границы**) установите границы между ячейками в соответствии с правилами оформления таблиц. Напомним, что в таблице не должно быть лишних линий, только необходимые, отделяющие заголовок таблицы от заголовка ее граф, заголовки граф от цифровых данных. Иногда используют линию, отделяющую итоговую строку. Вертикальная разграфка может быть, а может и отсутствовать. Фрагмент окончательного вида электронной таблицы на экране представлен в табл. 4.2.

Таблица 4.2

Фрагмент рабочего листа Excel с исходными данными

	A	B	C	D
1	Распространение информационных коммуникаций в странах мира, 2003 г.			
2	Страна	Магистральные телефонные линии на 1 тыс. чел.	Число абонентов сотовых сетей на 1 тыс. чел.	Число пользова- телей сети Интернет на 1 тыс. чел.
3	Австралия	542	719	567
4	Австрия	481	879	465
5	Азербайджан	114	128	37
...				
146	Ямайка	169	680	228
147	Япония	472	679	483

По характеру построения такая таблица называется простой перечневой. В *подлежащем* простой перечневой таблицы дается перечень всех единиц изучаемой совокупности — отдельных стран мира (ячейки **A3:A147**), в *сказуемом* таблицы содержатся данные по каждой единице совокупности (ячейки **B3:D147**).

Поскольку число строк в построенной таблице достаточно велико, для удобства дальнейшей работы с таблицей Excel предлагает способ закрепления шапки таблицы. Активизируйте ячейку **A3**, в меню **Окно** выберите команду **Закрепить области**. Фиксируются области таблицы, которые располагаются слева и выше активной ячейки таблицы. Проверьте с помощью линейки прокрутки, что шапка таблицы все время находится на экране, а остальную часть таблицы можно просматривать. Аналогичные действия используют и тогда, когда таблица включает большое число столбцов. Например, чтобы на экране все время оставались не только шапка, но и подлежащее таблицы (в нашем примере — список стран), команду **Закрепить области** необходимо выполнить при активизации ячейки **B3**. Чтобы отменить фиксацию, в меню **Окно** выбирают команду **Снять закрепление областей**.

Если при просмотре составленной таблицы обнаружены ошибки, редактирование содержимого ячеек можно осуществить:

- набором в активной ячейке новой информации поверх ошибочной;
- нажатием клавиши **F2** в активной ячейке с последующей корректировкой содержимого;
- удалением ошибочной информации в активной ячейке клавишей **Delete**.

Дополнительное задание. Введите в построенную таблицу дополнительные данные о численности населения и об уровне доходов населения (см. табл. П1.2), для чего сформируйте две дополнительные графы E и F (табл. 4.3).

При вводе данных о численности населения обратите внимание, чтобы были соблюдены правила записи цифровых данных. В пределах одной графы данные должны иметь одну степень точности, при этом разряды чисел располагаются под разрядами, целая часть числа отделяется от дробной запятой. Отредактировать формат введенных данных в графе можно с помощью пункта меню **Формат/Ячейки**, на закладке **Число** в списке **Числовых форматов** выберите **Числовой**, установите **Число**

Таблица 4.3

Фрагмент рабочего листа Excel с исходными данными

	A	B	C	D	E	F
1	Распространение информационных коммуникаций в странах мира, 2003 г.					
2	Страна	Магистральные телефонные линии на 1 тыс. чел.	Число абонентов сотовых сетей на 1 тыс. чел.	Число пользовате- лей сети Интернет на 1 тыс. чел.	Уровень дохода населения	Числен- ность населе- ния, млн чел.
3	Австра- лия	542	719	567	Высокий	19,7
4	Австрия	481	879	465	Высокий	8,1
5	Азербай- джан	114	128	37	Средний	8,3
...						
146	Ямайка	169	680	228	Средний	2,6
147	Япония	472	679	483	Высокий	127,2

десятичных знаков 1 и флажок **Разделитель групп разрядов**. Задать формат ввода данных можно и перед началом ввода информации, активизировав столбец, в который будут вводиться данные.

Простые перечневые таблицы составляются, как правило, как вспомогательные, например как основа последующей группировки. Результаты группировки по одному признаку оформляются в *групповых таблицах*. Если группировка проводится по нескольким признакам, строятся *комбинационные таблицы*. В подлежащем групповой таблицы указывается перечень групп единиц совокупности, в сказуемом – число единиц в группах (абсолютное и (или) в процентах к итогу) и сводные показатели по группам.

Пример 4.2. На основе данных примера 4.1 осуществим группировку стран по уровню доходов населения и рассчитаем средние показатели обеспеченности магистральными телефонными линиями, сотовыми телефонами и Интернетом в каждой группе стран. В итоговой строке группировки укажем сводные показатели по всем странам мира. Макет итоговой групповой таблицы представлен в табл. 4.4.

Таблица 4.4

Распространение информационных коммуникаций в странах с разным уровнем доходов населения, 2003 г.

Уровень дохода населения	Число стран	Численность населения		Распространенность на 1 тыс. чел.		
		млн чел.	% к итогу	магистральных телефонных линий	абонентов сотовых линий	пользователей сети Интернет
Низкий						
Средний						
Высокий						
Весь мир						

Excel предоставляет возможность осуществлять группировки и формировать сводные таблицы разными способами. Например, с помощью мастера **Сводные таблицы** или команды **Консолидация** из меню **Данные** и др. Все зависит от характера решаемой задачи, параметров исходной информации и навыков пользователя. При решении нашего примера рассмотрим возможности для группировки, которые предоставляет использование команд **Сортировка** и **Итоги** из меню **Данные**.

Выделите блок ячеек **A3:F147** (шапка таблицы не принимает участия в сортировке). В меню **Данные** выберите команду **Сортировка** (при этом вся таблица выделится цветным фоном). В окне **Сортировать по** из выпадающего списка выберите **Уровень дохода населения**. Установите переключатель **По возрастанию**. Переключатель **Идентифицировать поля по** установите в положение **Подписям**. Щелкните по кнопке **Параметры**. Установите переключатель в положение **Строки диапазона**. Нажмите **ОК**. В окне **Сортировка диапазона** еще раз нажмите **ОК**. В результате мы получили новую таблицу, в которой страны расположены блоками: сначала страны с высоким уровнем доходов населения, потом — со средним и затем — с низким.

После выполнения сортировки можно использовать команду **Итоги** из меню **Данные**, которая добавляет строки промежуточных итогов для каждой отсортированной группы и создает общие итоги.

Выделите отсортированную таблицу — блок ячеек **A2:F147**. В меню **Данные** выберите команду **Итоги**. В диалоговом окне **Промежуточные итоги** в поле **При каждом изменении в** выберите

Уровень дохода населения, в поле **Операция** выберите **Количество**, в поле **Добавить итоги по** включите флажок **Страна** (проверьте, чтобы ничего другого выбрано не было). Установите переключатель **Итоги под данными**. Нажмите **ОК**.

Дополнительно рассчитаем численность населения в каждой группе стран. Вновь активизируйте в меню **Данные** команду **Итоги**. В диалоговом окне **Промежуточные итоги** в поле **При каждом изменении в** выберите **Уровень дохода населения**, в поле **Операция** выберите **Сумма**, в поле **Добавить итоги по** включите флажок **Численность населения** (проверьте, чтобы ничего другого выбрано не было). Нажмите **ОК**. Снимите выделение с таблицы и просмотрите результат на экране. В таблице появились дополнительные строки, где отражены промежуточные итоги по группам и общие итоги по совокупности. Кроме того, изменился вид экрана. В левой вертикальной части появились символы структуры документа — кнопки с номерами уровней **1,2,3** и **4**, кнопки **+** (плюс) и **—** (минус), позволяющие соответственно скрывать или раскрывать строки структурированной таблицы. Щелкните по кнопке уровня **3**. На экране скроются исходные данные и останутся лишь промежуточные итоги, необходимые нам для составления групповой таблицы (табл. 4.5).

Осталось сформировать электронный шаблон итоговой групповой таблицы и перенести в нее полученные результаты. Обратите внимание, что при переносе данных в итоговую таблицу необходимо копирование не формул и форматов ячеек, а только содержащихся в них значений. В этом случае обмен содержимого ячеек нужно осуществить следующим образом. Выделите ячейку или блок ячеек, содержимое которых надо скопировать. С помощью контекстного меню или значка пиктограммы **Копировать** содержимое ячейки поместите в буфер обмена. Далее активизируйте ячейку, в которую необходимо поместить информацию. Правой кнопкой мыши вызовите контекстное меню и, выбрав команду **Специальная вставка**, включите параметр **Значения**. Нажмите **ОК**. Для отмены плавающей рамки вокруг ячеек нажмите **Esc**.

Для удаления промежуточных итогов в таблице с исходными данными задайте команды **Данные, Итоги, Убрать все**. Сохраните созданный вами массив данных на отдельном листе, поскольку они будут использованы при выполнении заданий к этой и последующим темам Практикума.

Таблица 4.5

Итоги сортировки в Excel

	A	B	C	D	E	F
1	Распространение информационных коммуникаций в странах мира, 2003 г.					
2	Страна	Магистральные телефонные линии на 1 тыс. чел.	Число абонентов сотовых сетей на 1 тыс. чел.	Число поль- зователей сети Интернет на 1 тыс. чел.	Уровень дохода населения	Числен- ность наसे- ления, млн чел.
31					Высокий Итого	942,6
32	28				Высокий Количество	
84					Низкий Итого	2249,7
85	51				Низкий Количество	
152					Средний Итого	3001,6
153	66				Средний Количество	
154					Общий итог	6193,9
155	145				Общее количество	

Вычисление оставшихся показателей итоговой групповой таблицы – удельного веса численности населения отдельных стран и средние показатели обеспеченности населения отдельными средствами коммуникаций – выполните самостоятельно, используя стандартные процедуры обработки данных в Excel. Сравните полученные вами результаты с табл. 4.6.

Обратите внимание, что при вычислении средних показателей обеспеченности средствами связей на 1 тыс. чел. должны использоваться формулы средней арифметической взвешенной. Например, среднее число пользователей сети Интернет в каждой группе стран рассчитывается по формуле

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i},$$

где x_i – число пользователей сети Интернет на 1 тыс. чел. в i -й стране;

f_i – численность населения в i -й стране;

$x_i f_i$ – общее число пользователей сети Интернет в i -й стране.

Подробнее о правилах расчета средних величин см. тему 5 «Средние величины».

Оформите итоговую таблицу в соответствии с предлагаемым ниже образцом. Поскольку таблица рассчитана на основе заимствованных данных, под таблицей обязательно укажите источник информации (см. табл. 4.6).

Таблица 4.6

Итоговая таблица в Excel

	A	B	C	D	E	F	G
1	Распространение информационных коммуникаций в странах с разным уровнем доходов населения, 2003 г.						
2	Уровень дохода населения	Число стран	Численность населения		Распространенность на 1 тыс. чел.		
3			млн чел.	процент к итогу	магистральные телефонные линии	абонентов сотовых сетей	пользователей сети Интернет
4	Низкий	28	2249,7	36,3	30	24	13
5	Средний	66	3001,6	48,5	178	223	70
6	Высокий	28	942,6	15,2	566	707	472
7	Весь мир	145	6193,9	100,0	184	224	110
	Рассчитано по: Доклад о развитии человека, 2005. Международное сотрудничество на перепутье: помощь, торговля и безопасность в мире неравенства: пер. с англ. — М.: Весь мир, 2005. — С. 254–257, 284–287, 390.						

Как видно из табл. 4.6, практически половина населения планеты проживает в странах со средним уровнем дохода на душу населения. На одну тысячу человек в этих странах приходится 178 стационарных телефонов, 223 сотовых, 70 пользователей Интернет. Эти показатели достаточно близки к среднемировым. Самым распространенным видом коммуникаций в мире в 2003 г. была сотовая связь, к которой имел доступ каждый пятый житель планеты, а самым современным видом связи — Интернетом — пользовался каждый десятый. Обеспеченность различными формами коммуникаций существенно дифференцирована по груп-

пам стран. В странах с низким уровнем дохода на душу населения основным видом связи остаются магистральные телефонные линии, т.е. стационарные телефоны. Но даже этот вид связи доступен лишь 3% населения этих стран, еще меньший процент населения — 2,4 — пользуются сотовой связью, 1,3% населения имеют доступ в Интернет. Показатели обеспеченности в группе стран с низким уровнем дохода на душу населения существенно ниже среднемировых: по стационарным телефонам — в 6 раз, по сотовым и Интернету — в 10 раз. Еще больший контраст наблюдается, если проводить сравнение со странами с высокими доходами на душу населения. В этой группе стран обеспеченность стационарными телефонами выше в 19 раз, сотовыми — почти в 30, а Интернетом — в 36 раз выше, чем в первой группе. В странах с высоким уровнем дохода, на которые приходится 15,2% населения мира, практически половина населения имеет доступ в Интернет, и более 70% населения пользуются сотовой связью.

Самой эффективной формой представления данных с точки зрения восприятия являются статистические графики. Статистические таблицы дополняются графиками в том случае, если ставится цель подчеркнуть какую-то особенность данных, провести их сравнение.

Пример 4.3. Используя итоговые данные примера 4.2, постройте график, отражающий структуру населения мира по величине дохода на душу населения.

Наиболее часто для характеристики структуры изучаемой совокупности используется *секторная (круговая) диаграмма*. Чтобы построить ее с помощью Excel, выполните следующие шаги.

Шаг 1. Выделите блок ячеек **A4:A6** и, удерживая клавишу **Ctrl**, — блок ячеек **D4:D6** (см. табл. 4.6). Пиктограммой на стандартной панели управления вызовите **Мастер диаграмм**. В раскрывшемся диалоговом окне на вкладке **Стандартные** в списке **Тип** выберите **Круговая**, на вкладке **Вид** — вторую позицию в нижнем ряду (Объемный вид разрезанной круговой диаграммы). Щелкните **Далее**.

Шаг 2. На вкладке **Диапазон данных** переключатель **Ряды в** установите в положение **столбцах**. Убедитесь, что на вкладке **Ряд** в поле **Значения** стоят ссылки на ячейки **D4:D6**, а в поле **Подписи категорий** — ссылки на ячейки **A4:A6**. **Далее**.

Шаг 3. На вкладке **Заголовки** введите в окне **Название диаграммы** текст «Распределение населения мира по уровню дохода на душу населения, %». На вкладке **Подписи данных** установите флажок **Значения**. Далее.

Шаг 4. Установите переключатель **Поместить диаграмму на листе** в положение **Имеющемуся** и щелкните **Готово**.

Сравните построенную вами диаграмму с рис. 4.1 и при необходимости отредактируйте ее.

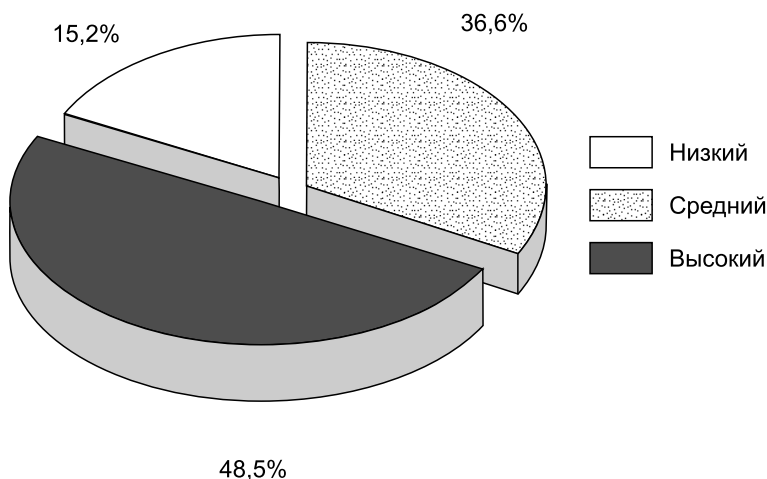


Рис. 4.1. Распределение населения мира по уровню дохода на душу населения

Для отображения структуры совокупности используют также *ленточные диаграммы*. Выбор типа и вида структурной диаграммы является субъективным и зависит от предпочтения пользователя. Считается, что проще воспринимать разницу в длине прямоугольников, чем углы секторов в круговых диаграммах, особенно если различия между долями не очень существенны или число выделяемых групп велико. Поэтому если приходится сравнивать структуры совокупностей, предпочтение часто отдается ленточным диаграммам (рис. 4.2). Однако если важно показать долю отдельной категории в общем «пироге», лучше использовать круговую диаграмму.

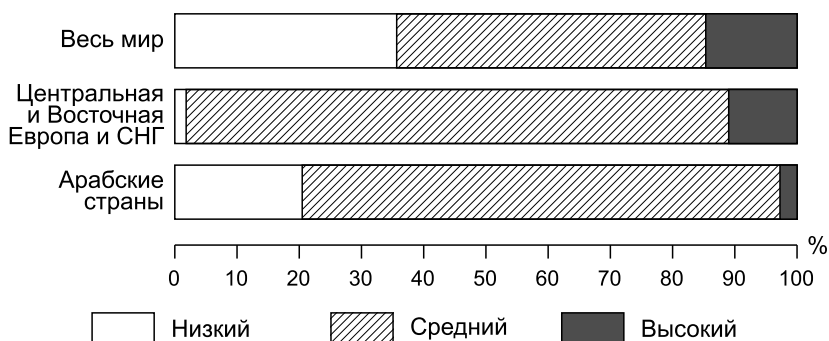


Рис. 4.2. Структура населения по уровню дохода в отдельных группах стран

Пример 4.4. На основе данных табл. 4.6 постройте график, отражающий различия в распространенности отдельных средств коммуникаций в странах с разным уровнем доходов населения.

Для иллюстрации соотношения отдельных переменных по единицам совокупности (группам единиц) чаще всего используют *столбиковые диаграммы*.

На рабочем листе, где построена итоговая таблица, выделите блок ячеек **A3:A6** и, удерживая клавишу **Ctrl**, блок ячеек **E3:G6** (см. табл. 4.6). На стандартной панели инструментов с помощью пиктограммы **Мастер Диаграмм** вызовите диалоговое окно **Тип диаграммы**. На вкладке **Стандартные** в опции **Тип** выберите **Гистограмма**, а в группе **Вид** — первую позицию во втором ряду (Объемный вариант обычной гистограммы). Щелкните **Далее**. В диалоговом окне **Источник данных диаграммы** на вкладке **Диапазон данных** в группе **Ряды в** установите опцию **столбцах** (заголовки строк автоматически становятся метками оси *X*, а заголовки столбцов — именами рядов данных, которые отображаются в качестве легенды диаграммы). Щелкните **Далее**. В диалоговом окне **Параметры диаграммы** на вкладке **Заголовки** введите в поле **Название диаграммы** «Распространение информационных коммуникаций в странах с разным уровнем дохода на душу населения, единиц на 1 тыс. чел.», в поле **Ось X** наберите «Уровень дохода населения». Щелкните **Далее**. В диалоговом окне **Размещение диаграммы** включите опцию **Имеющемуся** и щелкните кнопку **Готово**.

Сравните построенную вами диаграмму с рис. 4.3. Если вас не удовлетворяет вид построенной диаграммы, необходимо перейти в режим редактирования диаграммы (щелкните правой кнопкой мыши в любом свободном месте диаграммы). Можно изменить тип диаграммы, заголовок, подписи осей, добавить (убрать) данные и т.д. Можно вызвать режим редактирования отдельно для каждого элемента диаграммы: заголовок диаграммы, ось значений и ось категорий, легенда, линии сетки и т.д. Вызов диалогового окна редактирования может быть осуществлен разными способами, например поместив указатель мыши на один из элементов диаграммы и сделав двойной щелчок левой кнопкой мыши.

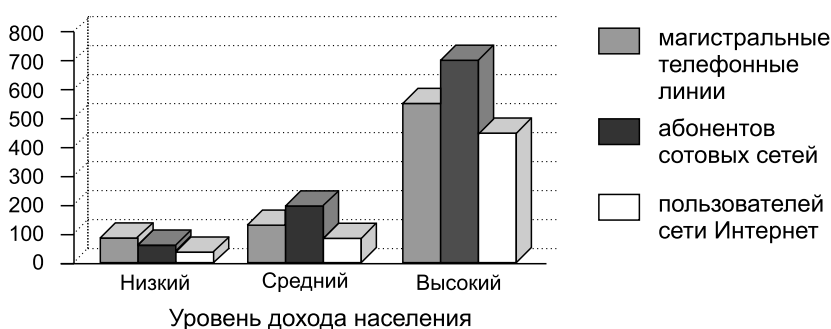


Рис. 4.3. Распространение информационных коммуникаций в странах с разным уровнем дохода на душу населения, единиц на 1 тыс. чел.

Построенный график подчеркивает существенные различия в обеспеченности населения различными средствами связи в странах с разными уровнями доходов населения.

Диаграмма может быть полезна и при изучении связи между переменными. Для этих целей применяют *диаграммы разброса* (в Excel они называются *точечные*, в статистической литературе обычно используется термин «*поле корреляции*»).

Пример 4.5. С помощью графического изображения исследуем зависимость между уровнями распространенности отдельных видов коммуникаций.

Постройте диаграмму разброса, отражающую зависимость между обеспеченностью населения магистральными телефонными линиями и числом пользователей Интернет. Воспользуйтесь массивом данных, созданным при выполнении предыдущего

задания (см. табл. 4.3). Выделите блок ячеек **B3:B147** и, удерживая клавишу **Ctrl**, блок ячеек **D3:D147**. Пиктограммой на стандартной панели управления вызовите **Мастер диаграмм**. В раскрывшемся диалоговом окне на вкладке **Стандартные** в списке **Тип** выберите **Точечная**. Щелкните **Далее**. На втором шаге на вкладке **Диапазон данных** переключатель **Ряды в** установите в положение **столбцах**. Щелкните **Далее**. На третьем шаге на вкладке **Заголовки** введите в окне **Название диаграммы** текст «Диаграмма разброса», в окне **Ось X** наберите «Число магистральных телефонных линий (на 1 тыс. чел.)», в окне **Ось Y** — «Число пользователей сети Интернет (на 1 тыс. чел.)». Щелкните **Далее**. На четвертом шаге установите переключатель **Поместить диаграмму на листе** в положение **имеющемся** и щелкните **Готово**. Удалите значок **Легенда** с поля построения диаграммы. Сравните построенную диаграмму с рис. 4.4 и при необходимости отредактируйте ее.



Рис. 4.4. Диаграмма разброса

Как видно на рис. 4.4, несмотря на большой разброс значений переменных, между ними существует прямая (положительная) зависимость — чем выше уровень телефонизации в стране, тем выше и число пользователей Интернет.

Самостоятельно постройте диаграмму разброса для переменных «Число абонентов сотовых сетей (на 1 тыс. чел.)» и «Число пользователей Интернет (на 1 тыс. чел.)» — рис. 4.5. График также

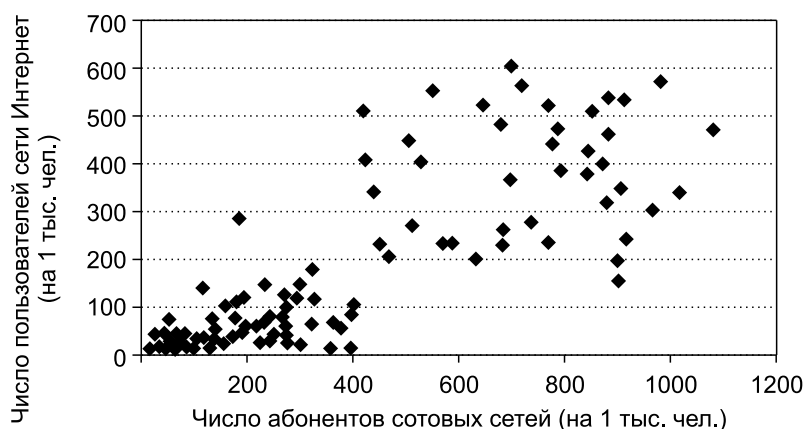


Рис. 4.5. Диаграмма разброса

свидетельствует о наличии некоторой прямой зависимости между переменными. Однако наблюдается больший разброс точек по полю, значит, связь между переменными менее тесная, чем в первом случае, особенно при больших значениях переменных. При низких значениях – связь очень тесная и в том и другом случае.

Пример 4.6. На основе табл. П1.3 сформируйте в формате Excel по предлагаемому образцу (табл. 4.7) массив данных, содержащий информацию о деятельности 62 банков Северо-Западного региона в первом полугодии 2006 г. Вычислите показатель рентабельности капитала для каждого банка и постройте вариационный ряд, характеризующий распределение банков по уровню рентабельности капитала. Постройте графики вариационного ряда: гистограмму, полигон и кумуляту распределения.

Рентабельность капитала рассчитывается по формуле

$$\text{Рентабельность капитала} = \frac{\text{Балансовая прибыль}}{\text{Собственный капитал}} \cdot 100\%.$$

Для расчета показателей рентабельности введите в ячейку **D3** формулу **=C3/V3*100** и скопируйте ее в остальные ячейки столбца. Для чисел, записываемых в ячейки этого столбца, установите числовой формат с одним десятичным знаком (выделите столбец **D**, **Формат/ Ячейки/ закладка Число/Числовые форматы – Числовой/ Число десятичных знаков – 1/ ОК**). Фрагмент

Таблица 4.7

Фрагмент рабочего листа Excel с исходными данными

	A	B	C	D
1	Банки Северо-Западного региона в первом полугодии 2006 г.			
2	Банк	Собственный капитал на 01.07.2006, тыс. руб.	Балансовая прибыль на 01.07.2006, тыс. руб.	Рентабельность капитала, %
3	Промстройбанк	12231433	2783466	
4	Банк «Санкт-Петербург»	4210065	528635	
5	Национальный банк «Траст»	3209107	148899	
63	Станкобанк	14009	503	
64	Промышленность и финансы	10510	457	

таблицы с результатами вычисления рентабельности капиталов банков представлен в табл. 4.8.

Таблица 4.8

Фрагмент рабочего листа Excel с исходными данными и результатами вычисления рентабельности

	A	B	C	D
1	Банки Северо-Западного региона в первом полугодии 2006 г.			
2	Банк	Собственный капитал на 01.07.2006, тыс. руб.	Балансовая прибыль на 01.07.2006, тыс. руб.	Рентабельность капитала, %
3	Промстройбанк	12231433	2783466	22,8
4	Банк «Санкт-Петербург»	4210065	528635	12,6
5	Национальный банк «Траст»	3209107	148899	4,6
63	Станкобанк	14009	503	3,6
64	Промышленность и финансы	10510	457	4,3

Просмотр полученных значений свидетельствует о существенном разбросе показателя рентабельности капитала: от 0,9% (Балт-соцкомбанк) до 22,8% (Промстройбанк). Кроме того, есть два нетипичных значения. Русский Региональный Банк по итогам работы в первом полугодии 2006 г. оказался убыточным, КИТ Финанс Инвестиционный Банк имел рентабельность более 60,0%.

Для анализа распределения банков по показателю рентабельности создадим сводную таблицу, в которой банки будут распределены по группам в зависимости от уровня рентабельности. Такой способ представления данных называют *распределением частот* или *вариационным рядом*. При построении распределения частот надо решить две задачи – выбрать количество групп и определить границы интервалов группирования. Количество групп (k) зависит от объема совокупности и обычно определяется по формуле $k = 1 + 1,44\ln n$, где n – объем совокупности. Совокупность анализируемых банков включает 61 единицу (исключим из дальнейшего анализа убыточный банк). Следовательно, число групп $k = 1 + 1,44\ln 61 = 6,9$. Поскольку число групп должно быть целым числом, будем формировать 7 групп. Величину интервала рассчитаем по формуле

$$i = \frac{x_{\max} - x_{\min}}{k},$$

где x_{\max} и x_{\min} – максимальное и минимальное значения признака в совокупности.

В нашем случае

$$i = \frac{x_{\max} - x_{\min}}{k} = \frac{22,8 - 0,9}{7} = 3,1\%.$$

Следовательно, будем формировать следующие группы по величине рентабельности, %: 0,9 – 4,0; 4,1 – 7,2; 7,3 – 10,4; 10,5 – 13,6; 13,7 – 16,8; 16,9 – 20,0; 20,1 и более (последний интервал оставим открытым, учитывая, что в совокупности есть банк с очень высоким уровнем рентабельности). Отметим, что при вычислении распределения частот необходимо так определить границы групп, чтобы они не пересекались. Перекрывание групп не допускается.

Для обозначения интервалов групп в Excel используется термин «карманы». «Карманы» представляют собой *упорядоченный массив верхних границ групп*, частоты которых необходимо определить.

Скопируйте значения уровней рентабельности банков (ячейки **D3:D64** из табл. 4.8) на новый рабочий лист (ячейки **A3:A64**). Удалите строку, содержащую нетипичное отрицательное значение переменной (строка 46). Введите в ячейки **B3:B9** значения верхних границ сформированных интервалов (табл. 4.9). Обратите внимание, что в «карманы» мы не вводим последний интервал, поскольку при создании распределения частот Excel автоматически создает группу «Еще», в которую попадают все единицы группируемой совокупности, имеющие значения больше, чем максимальное значение в указанном диапазоне карманов.

Таблица 4.9

**Фрагмент рабочего листа Excel с исходными данными
для вычислений распределения частот**

	A	B	C	D	E
1					
2	Рентабельность капитала, %	Карманы			
3	22,8	4,0			
4	12,6	7,2			
5	4,6	10,4			
6	60,1	13,6			
7	11,6	16,8			
8	7,1	20,0			
9	6,2				
62	3,6				
63	4,3				

Распределение частот строится с помощью процедуры создания гистограмм надстройки **Пакет анализа**. Для этого выполните команду **Сервис/ Анализ данных**, выберите из списка **Инструменты анализа**, расположенного в окне **Анализ данных**, пункт **Гистограмма** и щелкните по кнопке **ОК**. В диалоговом окне **Гистограмма** установите курсор в поле **Входной интервал**

и на рабочем листе выделите блок ячеек **A3:A63** (в окне автоматически появляются ссылки на эти ячейки \$A\$3:\$A\$63). Аналогичным образом введите в поле **Интервал карманов** ссылки на ячейки **B3:B8**. Установите переключатель **Параметры вывода** в положение **Выходной интервал** и в соответствующем окне укажите ссылку на ячейку, начиная с которой будут размещаться результаты группировки. Например, **D3**. Установите флажки **Интегральный процент** и **Вывод графика**. Нажмите **ОК**. На рабочем листе появится таблица, отражающая результаты вычисления распределения частот (табл. 4.10) и график (рис. 4.6).

Таблица 4.10

Результаты вычисления частот

Карман	Частота	Интегральный %
4,0	15	24,59%
7,2	16	50,82%
10,4	13	72,13%
13,6	8	85,25%
16,8	4	91,80%
20,0	2	95,08%
Еще	3	100,00%

ГИСТОГРАММА

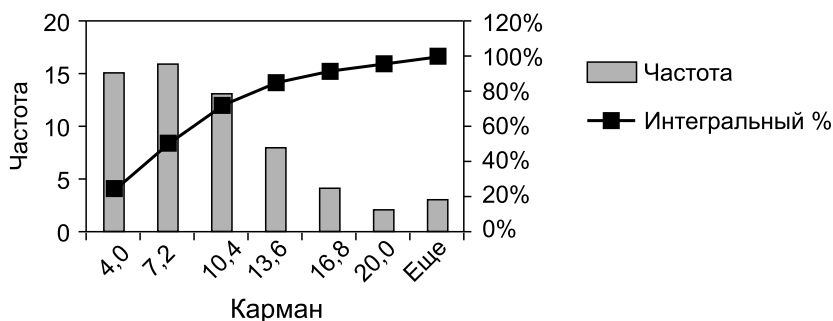


Рис. 4.6. Гистограмма

Чтобы улучшить внешний вид итоговой таблицы и сделать ее удобной для пользователя, скопируйте значения вычисления распределения частот в свободное место рабочего листа (при копировании пользуйтесь командой **Специальная вставка**) и от-

редактируйте внешний вид, заголовки, формат и содержание ячеек. Образец итоговой таблицы представлен в табл. 4.11. Обратите внимание, что в таблицу дополнительно включен столбец, содержащий данные об удельном весе банков каждой группы в общей численности совокупности, так называемое *распределение относительных частот (частостей)*, или *процентное распределение*. Работать с долями или процентами в некоторых случаях удобнее, чем с абсолютным числом единиц в каждой группе. В частности, процентное распределение позволяет сравнивать совокупности, имеющие разные объемы. Интегральные проценты или накопленные частоты позволяют получить дополнительную информацию о распределении. Так, например, по нашим данным (см. табл. 4.11) видно, что половина (50,8%) банков Северо-Западного региона в первом полугодии 2006 г. имела показатель рентабельности капитала менее 7,2%, и только у 4,9% (100,0–95,1) банков рентабельность капитала была выше 20,0%.

Таблица 4.11

**Распределение банков Северо-Западного региона
по уровню рентабельности собственного капитала**

Рентабельность капитала	Число банков	Удельный вес банков, %	Накопленные частоты, %
0,9 – 4,0	15	24,6	24,6
4,1 – 7,2	16	26,2	50,8
7,3 – 10,4	13	21,3	72,1
10,5 – 13,6	8	13,1	85,3
13,7 – 16,8	4	6,6	91,8
16,9 – 20,0	2	3,3	95,1
20,1 и более	3	4,9	100,0
Итого	61	100,0	–

График, построенный процедурой **Анализ данных /Гистограмма** (см. рис. 4.6), также неудобен для пользования и требует корректировки. Отметим основные недостатки графика: между столбцами гистограммы есть пробелы, заголовки и подписи осей содержат непонятные для пользователя тексты, максимальное значение на вспомогательной оси превышает 100%, не выдер-

жаны пропорции графика (наилучшим для восприятия считается соотношение размера по оси абсцисс к размеру по оси ординат 8 : 5, так называемое «золотое сечение»), имеются лишние линии.

Скопируйте график в свободное место рабочего листа и исправьте ошибки в соответствии с правилами оформления графиков вариационного ряда. Например, чтобы удалить пробелы между столбцами, щелкните правой кнопкой мыши на одном из столбцов гистограммы (появится подсказка «Ряд Частота»). Выберите в контекстном меню команду **Формат рядов данных**. В диалоговом окне **Формат ряда данных** на вкладке **Параметры** установите значение **Ширина зазора** равным нулю. Нажмите **ОК**. Чтобы изменить подписи осей или заголовок графика, щелкните по ним левой кнопкой (вокруг появится рамка), введите в строке формул новый текст и нажмите клавишу **Enter**. Самостоятельно откорректируйте и другие элементы графика. Правильные варианты оформления графиков вариационного ряда для нашего примера приведены на рис. 4.7, 4.8 и 4.9.

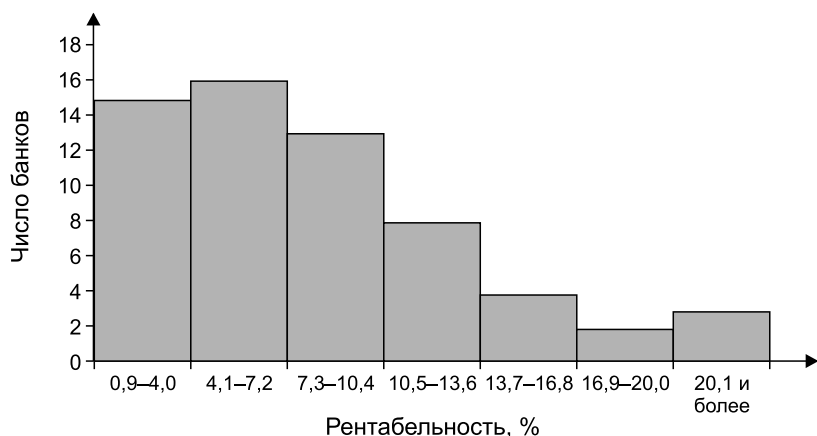


Рис. 4.7. Гистограмма распределения банков Северо-Западного региона по уровню рентабельности

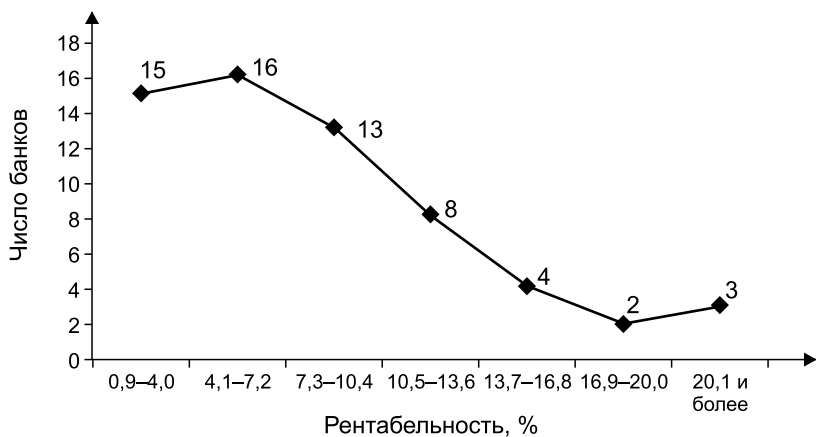


Рис. 4.8. Полигон распределения банков Северо-Западного региона по уровню рентабельности

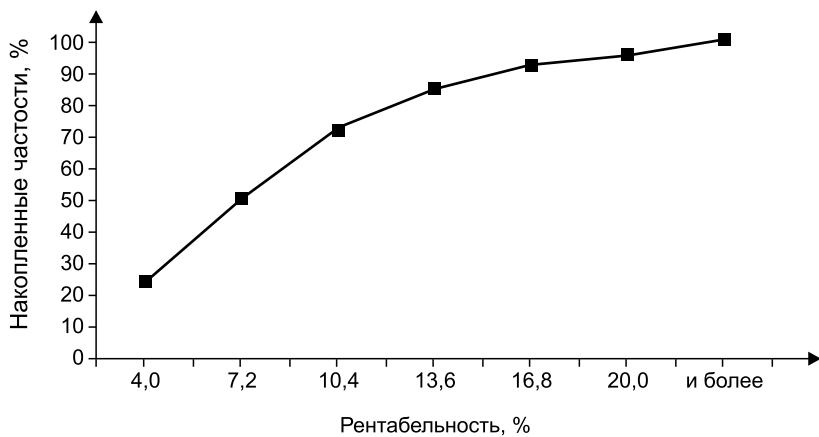


Рис. 4.9. Кумулята распределения банков Северо-Западного региона по уровню рентабельности

Пример 4.7. В табл. 4.12 приведен фрагмент рабочего листа Excel, на котором сформирован массив данных, характеризующий некоторые показатели деятельности компании «Роснефть».

Таблица 4.12

Фрагмент рабочего листа Excel с исходными данными

	A	B	C	D	E	F
1	Динамика финансовых и производственных показателей компании «Роснефть»					
2	Год	Добыча газа, млрд куб. м	Добыча нефти, млн т	Выручка, млрд долл.	Чистая прибыль, млрд долл.	Текущие обязательства, млрд долл.
3	1998	4,8	12,5	1,4	−0,1	2,4
4	1999	4,91	12,6	1,6	0,2	2,2
5	2000	5,6	13,11	2,5	0,5	2,3
6	2001	6,1	14,94	2,3	0,5	2,4
7	2002	6,4	16,1	2,7	0,3	3,0
8	2003	7,0	19,6	3,6	0,4	4,2
9	2004	9,4	21,6	5,3	0,8	20,1
10	2005	13,1	74,6	24,0	4,2	20,7
11	2006	15,65	78,97	30,1	3,2	20,6

Источник: Эксперт. – 2006. – № 37. – С. 188.

Постройте график, отражающий динамику добычи газа компанией «Роснефть». Дополните график линией тренда, наилучшим образом отражающего тенденцию динамики показателя. Отрадите на графике уравнение тренда и величину достоверности аппроксимации. Постройте графически прогноз уровня добычи газа на ближайшие два года.

Для графического отображения динамики добычи газа используйте *столбиковую диаграмму*.

Шаг 1. Для построения диаграммы вызовите **Мастер диаграмм** и на вкладке **Стандартные** выберите в списке диаграмм пункт **Гистограмма**, в списке **Вид** – первый вариант (Обычная гистограмма). Щелкните **Далее**.

Шаг 2. На вкладке **Диапазон данных** в окне редактирования **Диапазон** введите ссылки на ячейки **В3:В11** и переключатель **Ряды** в установите в положение **столбцах**. На вкладке **Ряды** в окне редактирования **Подписи оси X** введите ссылку на ячейки **А3:А11** и щелкните по кнопке **Далее**.

Шаг 3. Выполните следующие действия:

- на вкладке **Заголовки** введите в окне редактирования **Название диаграммы** заголовок графика «Динамика добычи газа компанией «Роснефть», млрд куб. м», в окне редактирования **Ось X (категорий)** – подпись «Годы»;
- на вкладке **Легенда** уберите флажок **Добавить легенду**;
- на вкладке **Линии сетки** снимите все флажки;
- на вкладке **Оси** оставьте только флажок **ось X (категорий)**;
- на вкладке **Подписи данных** установите флажок **Значения**.

Щелкните по кнопке **Далее**.

Шаг 4. Установите переключатель **Поместить диаграмму на листе** в положение **имеющемуся**, щелкните по кнопке **Готово**.

Для того чтобы убрать заливку и границу с области построения графика, щелкните по ней правой кнопкой и в контекстном меню выберите **Формат области построения**. В открывшемся диалоговом окне переведите переключатель **Рамка** в положение **невидимая**, а переключатель **Заливка** – в положение **прозрачная**. Нажмите **ОК**. Для того чтобы убрать лишние метки с оси X, щелкните по ней правой кнопкой мыши, выберите **Формат оси** и на вкладке **Вид** установите переключатели **Ось – обычная**, **Основные – нет**, **Промежуточные – нет**. Нажмите **ОК**.

Сравните полученный результат с рис. 4.10. Если необходимо, откорректируйте диаграмму. При построении графиков стремитесь к тому, чтобы они содержали как можно меньше излишеств и ненужных деталей, в то же время были достаточно информативны и понятны при чтении.

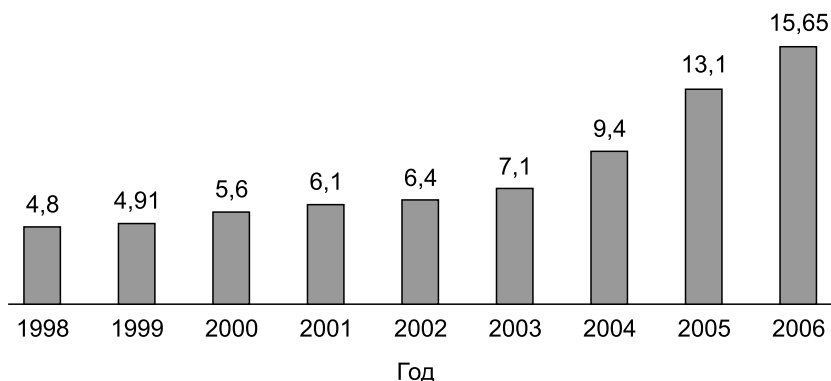


Рис. 4.10. Добыча газа компанией «Роснефть», млрд куб. м

Для того чтобы отразить на графике линию тренда, выделите ряд данных диаграммы, щелкнув по одному из столбцов диаграммы правой кнопкой. В открывшемся контекстном меню выберите пункт **Добавить линию тренда**. В диалоговом окне на вкладке **Тип** укажите тип линии, которая, на ваш взгляд, наилучшим образом аппроксимирует динамику показателя. Проверить, правильно ли выбрана линия тренда, можно по коэффициенту аппроксимации. Чем ближе к единице величина коэффициента, тем правильнее был выбран тип линии тренда. Для нашего примера выберите **Полиномиальная 4-й степени**. На вкладке **Параметры** в опции укажите **вперед на 2 периода** и установите флажки **Показывать уравнение на диаграмме** и **Поместить на диаграмму величину достоверности аппроксимации**. Нажмите **ОК**.

Сравните полученный график с рис. 4.11. Поскольку график преследует цель отразить не только динамику, но и прогноз развития, его оформление существенно отличается от предыдущего (см. рис. 4.10). Чтобы не перегружать график, мы убрали подписи значений, но одновременно ввели отображение оси *У*, шкала которой поможет лучше ориентироваться в фактических и прогнозных значениях. Поскольку коэффициент аппроксимации достаточно высок ($R^2 = 0,9905$), можно с большой степенью вероятности утверждать, что при сохранении сложившейся тенденции в ближайшие годы следует ожидать дальнейшего наращивания добычи газа. В 2008 г. она может превысить 20 млрд куб. м.

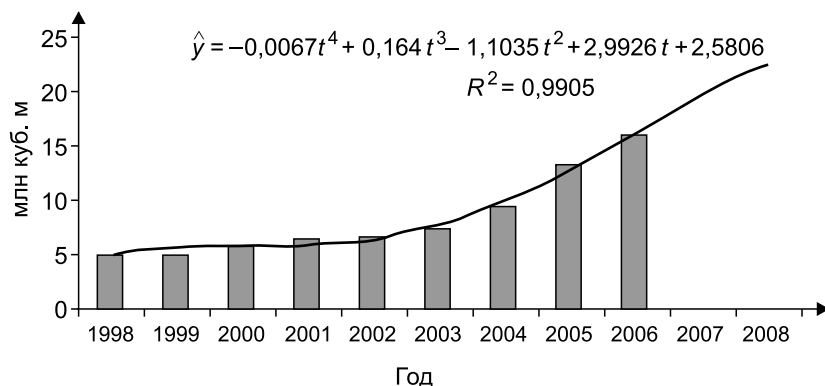


Рис. 4.11. Добыча газа компанией «Роснефть»

Пример 4.8. На основе данных предыдущего примера (см. табл. 4.12) постройте график, позволяющий сопоставить динамику основных финансовых и производственных показателей компании «Роснефть».

Чтобы при графическом изображении динамика двух и более показателей, имеющих разные единицы измерения и размерность, была сопоставимой, следует обеспечить их «единый старт». Для этого на графике отражают не значения уровней показателей, а базисные темпы роста показателей.

Базисные темпы роста рассчитываются как отношения уровня динамического ряда за текущий период к начальному уровню динамического ряда.

Построим шаблон таблицы для записи базисных темпов роста. Поскольку он будет практически дублировать таблицу с исходными данными, воспользуйтесь командами **Копировать** и **Вставить** (например, начиная с ячейки **A14**), затем очистите содержание таблицы и удалите строку, содержащую данные за 1998 г. Этот год придется исключить из анализа, поскольку темпы роста не рассчитываются для показателей, которые принимают и положительные, и отрицательные значения (в 1998 г. компания имела убыток в размере 100 млн долл.).

Далее введите в ячейки формулы для расчета базисных темпов роста за период:

- в ячейку B15 формулу $= B4/4,91 \cdot 100$;
- в ячейку C15 формулу $= C4/12,6 \cdot 100$;
- в ячейку D15 формулу $= D4/1,6 \cdot 100$;
- в ячейку E15 формулу $= E4/0,2 \cdot 100$;
- в ячейку F15 формулу $= F4/2,2 \cdot 100$.

Протаскиванием маркера заполнения скопируйте формулы расчета в остальные ячейки соответствующих столбцов. Результаты вычисления представлены в табл. 4.13.

Шаг 1. Вызовите **Мастер диаграмм** и в одноименном диалоговом окне на вкладке **Стандартные** выберите тип и вид диаграммы: **График с маркерами, помечающими точки данных**. Щелкните по кнопке **Далее**.

Шаг 2. На вкладке **Диапазон данных** в окне редактирования **Диапазон** введите ссылки на ячейки **B14:F22** и переключатель **Ряды** в установите в положение **столбцах**. На вкладке **Ряды** в окне редактирования **Подписи оси X** введите ссылку на ячейки **A15:A22** и щелкните по кнопке **Далее**.

Таблица 4.13

**Фрагмент рабочего листа Excel с результатами вычисления
базисных темпов роста**

	A	B	C	D	E	F
13	Базисные темпы роста финансовых и производственных показателей компании «Роснефть»					
14	Год	Добыча газа, млрд куб. м	Добыча нефти, млн т	Выручка, млрд долл.	Чистая прибыль, млрд долл.	Текущие обязатель- ства, млрд долл.
15	1999	100,0	100,0	100,0	100,0	100,0
16	2000	114,1	104,0	156,3	250,0	104,5
17	2001	124,2	118,6	143,8	250,0	109,1
18	2002	131,3	127,8	168,8	150,0	136,4
19	2003	144,6	155,6	225,0	200,0	190,9
20	2004	191,4	171,4	331,3	400,0	913,6
21	2005	266,8	592,1	1500,0	2100,0	940,9
22	2006	318,7	626,7	1881,3	1600,0	936,4

Шаг 3. На вкладке **Заголовки** введите в окне редактирования **Название диаграммы** заголовков графика «Темпы роста финансовых и производственных показателей компании «Роснефть» к уровню 1999 г.», в окне редактирования **Ось X (категорий)** – подпись «Годы», в окне редактирования **Ось Y(категорий)** – подпись «Проценты»; щелкните по кнопке **Далее**.

Шаг 4. Установите переключатель **Поместить диаграмму на листе** в положение **имеющемся**, щелкните по кнопке **Готово**.

Сравните полученную диаграмму с рис. 4.12 и при необходимости отредактируйте ее (измените шрифт заголовков и подписей, расположение легенды, тип и цвет линий, уберите лишние элементы и т.д.).

Пример 4.9. По данным табл. 4.14 постройте график, отражающий динамику инвестиций в основной капитал в РФ за два года.

Таблица 4.14

Фрагмент рабочего листа Excel с исходными данными

	A	B	C	D	E
1	Инвестиции в основной капитал, млрд. руб.				
2		1-й кв.	2-й кв.	3-й кв.	4-й кв.
3	2004	423,9	603,4	774,0	1003,5
4	2005	526,5	755,7	980,0	1271,8

Источник: Вопросы статистики. – 2006. – № 6. – С. 75.

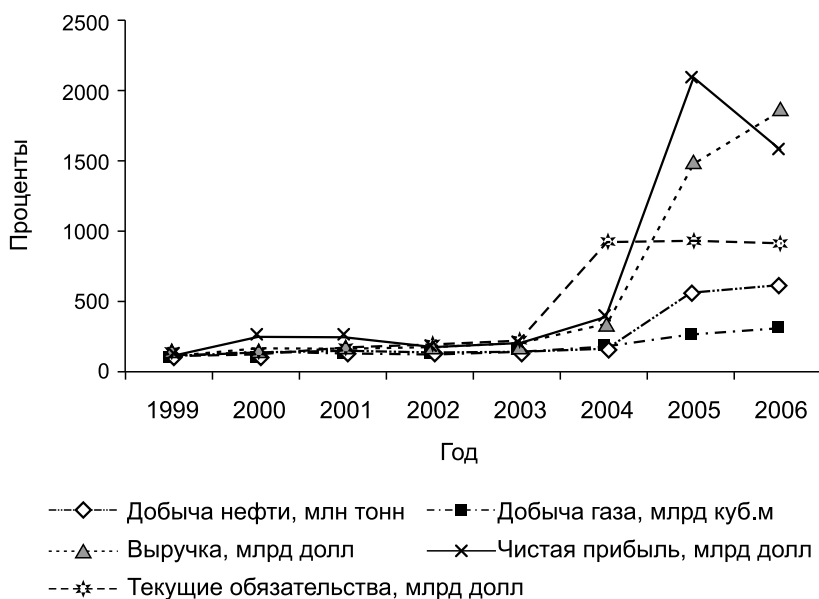


Рис. 4.12. Темпы роста финансовых и производственных показателей компании «Роснефть» к уровню 1999 г.

Для графического представления данных, в которых присутствуют так называемые сезонные колебания, можно использовать *радиальные диаграммы*.

Вызовите **Мастер диаграмм** и в одноименном диалоговом окне на вкладке **Стандартные** выберите тип диаграммы **Лепестковая**. Щелкните по кнопке **Далее**. На втором шаге на вкладке **Диапазон данных** в окне редактирования **Диапазон** введите ссылки на ячейки **A2:E4** и переключатель **Ряды** в установите в положение **строк**. Щелкните по кнопке **Далее**. На третьем шаге на вкладке **Заголовки** введите в окне редактирования **Название диаграммы** заголовок графика «Инвестиции в основной капитал в РФ, млрд руб.». Щелкните по кнопке **Далее**. На четвертом шаге установите переключатель **Поместить диаграмму на листе** в положение **имеющемся**, щелкните по кнопке **Готово**.

Сравните построенную диаграмму с рис. 4.13 и при необходимости отредактируйте ее.

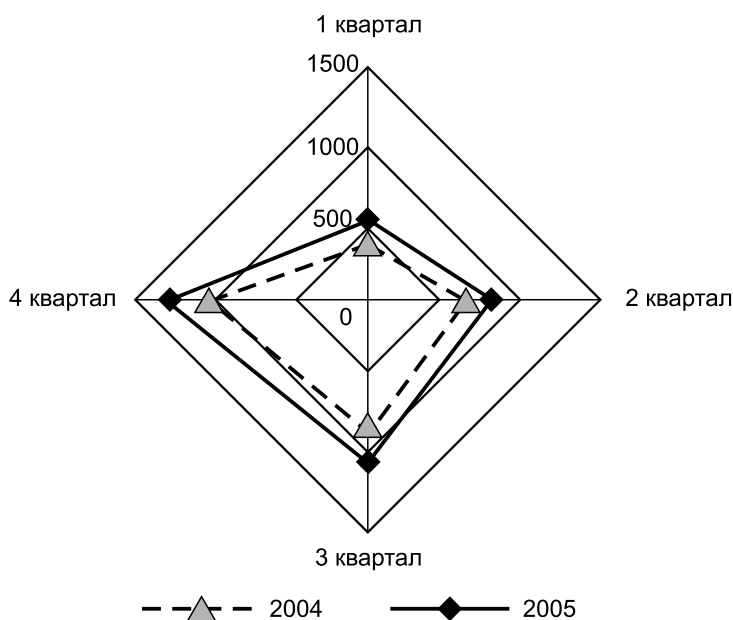


Рис. 4.13. Инвестиции в основной капитал в РФ, млрд руб.

На графике отчетливо виден рост инвестиций в основной капитал, при этом объем инвестиций очень неравномерно распределен по кварталам. Ежегодно наибольшие объемы инвестиций приходятся на 3-й и 4-й кварталы.

Пример 4.10. Имеются данные о внутрироссийской миграции по федеральным округам за два года (табл. 4.15).

Таблица 4.15

Внутрироссийская миграция

Федеральный округ	2001				2004	
	Прибыло, чел.	Выбыло, чел.	Среднегодовая численность населения, тыс. чел.	Прибыло, чел.	Выбыло, чел.	Среднегодовая численность населения, тыс. чел.
Центральный	450908	378335	38121,5	457010	373671	3763,5
Северо-Западный	201073	194479	14261,5	181557	176428	13777,0

Продолжение

Федеральный округ	2001				2004	
	Прибыло, чел.	Выбыло, чел.	Среднегодовая численность населения, тыс. чел.	Прибыло, чел.	Выбыло, чел.	Среднегодовая численность населения, тыс. чел.
Южный	306242	318135	22807,5	28541	295150	22835,5
Приволжский	447311	480830	31424,0	402839	426754	30806,0
Уральский	214213	210749	12444,5	203636	207042	12297,5
Сибирский	388607	414668	20255,5	354865	381325	19847,5
Дальневосточный	132230	163388	6787,5	115829	137907	6613,5

Составлено по: Российский статистический ежегодник. – 2005: Стат. сб. – М.: Росстат, 2006. – С. 82, 83, 130–133.

Построим график, отражающий различия в характере и интенсивности внутрироссийской миграции в разных округах РФ.

Учитывая то, что федеральные округа существенно отличаются по численности населения, для сравнения интенсивности миграции между округами необходимо перейти к относительному показателю миграции – коэффициенту механического прироста (убыли) населения ($K_{\text{мех}}$), который рассчитывается как отношение сальдо миграции к среднегодовой численности населения:

$$K_{\text{мех}} = \frac{\text{Прибыло} - \text{Выбыло}}{\text{Среднегодовая численность населения}} \cdot 1000.$$

Коэффициент механического прироста измеряется в *промилле* и показывает, на сколько процентов за счет миграции выросла (уменьшилась) численность населения территории на каждую тысячу человек населения.

Результаты вычисления коэффициентов миграции для отдельных округов, сформированные на рабочем листе Excel, представлены в табл. 4.16.

Таблица 4.16

**Внутрироссийская миграция по федеральным округам,
прирост «+» (убыль «-») на 1000 чел. населения**

	А	В	С
1		2001 г.	2004 г.
2	Центральный	+1,9	+2,2
3	Северо-Западный	+0,5	+0,4
4	Южный	-0,5	-0,6
5	Приволжский	-1,1	-0,8
6	Уральский	-0,3	-0,3
7	Сибирский	-1,3	-1,3
8	Дальневосточный	-4,6	-3,4

Как видно из таблицы, округа имеют как положительные, так и отрицательные значения коэффициентов миграции. Для графического изображения таких данных используют особый тип ленточных диаграмм. Построим диаграмму для нашего примера.

Шаг 1. Выделите блок ячеек **A1:C8**. Вызовите **Мастер диаграмм** и в одноименном диалоговом окне на вкладке **Стандартные** выберите тип диаграммы **Линейчатая**. Щелкните по кнопке **Далее**.

Шаг 2. На вкладке **Диапазон данных** переключатель **Ряды в** установите в положение **столбцах**. Щелкните по кнопке **Далее**.

Шаг 3. На вкладке **Заголовки** введите в окне редактирования **Название диаграммы** заголовок графика «Внутрироссийская миграция по федеральным округам, прирост (убыль) на 1000 человек населения». На вкладке **Линии сетки** уберите все линии, поскольку они будут только загромождать график. На вкладке **Подписи данных** в поле **Включить в подписи** установите флажок **значения**. Щелкните по кнопке **Далее**.

Шаг 4. Установите переключатель **Поместить диаграмму на листе** в положение **имеющемся**, щелкните по кнопке **Готово**.

Сравните построенный вами график с рис 4.14 и отредактируйте его. Чтобы названия округов располагались слева от графика (как на рис. 4.14) и оси не содержали лишних меток, правой кнопкой мыши выделите ось категорий и в контекстном меню выберите команду **Формат оси**. В открывшемся диалоговом окне на вкладке **Вид** установите переключатели: **Основные — нет**, **Промежуточные — нет**, **Метки делений — внизу**. Нажмите **ОК**. Выделив **Формат области построения**, сделайте невидимой рамку и уберите заливку (заливка — прозрачная).

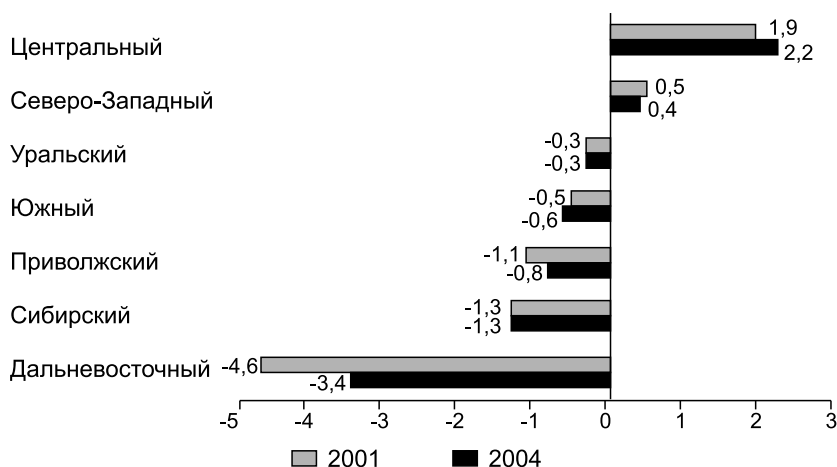


Рис. 4.14. Внутрироссийская миграция по федеральным округам, прирост «+» (убыль «-») на 1000 чел. населения

На рис. 4.14 видно, что только в двух округах РФ (Центральный и Северо-Западный) имеет место положительная внутрироссийская миграция. При этом в Центральном округе, где наблюдается самый высокий механический прирост населения за счет внутрироссийской миграции, оно продолжает расти. Численность населения большинства территорий продолжает сокращаться за счет внутренней миграции. Наиболее интенсивно происходит отток населения из Дальневосточного федерального округа, вместе с тем имеет место некоторое снижение интенсивности этого процесса.

Ленточные диаграммы, аналогичные приведенной на рис. 4.14, используют также для отображения отклонений значений признака у отдельных единиц совокупности от среднего значения признака по совокупности. Пример такого графика изображен на рис. 4.15 (данные условные).

Не следует перегружать рисунок элементами, не имеющими отношения к изображаемым данным: линии сетки, заливка области построения, декоративные элементы, излишние уточнения данных. Такие элементы относятся к *«графическому хламу»*.

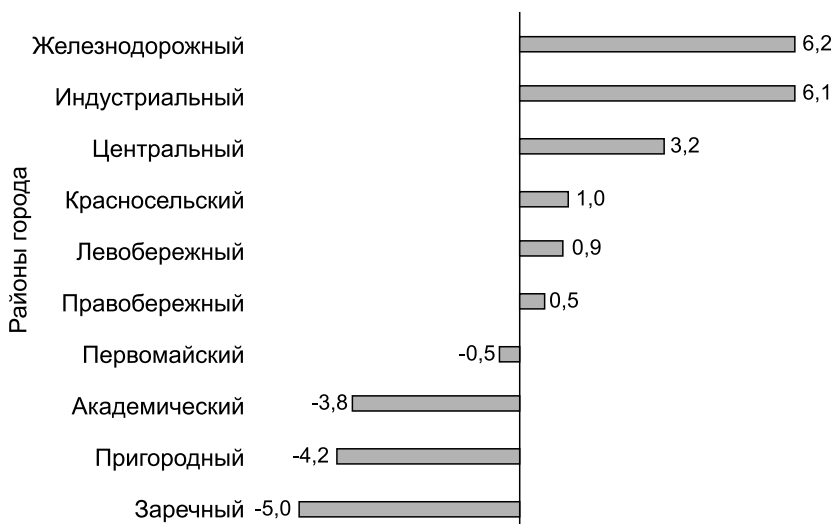


Рис. 4.15. Отклонение от средней величины выбросов вредных веществ в атмосферу по городу, кг на одного жителя

4.2. Задания по теме

1. Используя статистический сборник «Россия в цифрах», разработайте макет комбинационной таблицы для отражения распределения занятого населения РФ по двум признакам: размер среднемесячной заработной платы и вид деятельности.

2. Разработайте макет комбинационной таблицы для отражения распределения домохозяйств вашего региона по двум признакам: размер общей площади на одного человека, размер домохозяйства.

3. Разработайте макет комбинационной таблицы для отражения распределения оборудования предприятия по двум признакам: срок эксплуатации и процент износа.

4. В табл. П4.3 приведены данные о деятельности банков Северо-Западного региона в первом полугодии 2006 г. Разработайте систему макетов таблиц, позволяющую проанализировать банковскую деятельность в регионе.

5. Используя данные табл. П1.3, произведите группировку банков Северо-Западного региона:

- а) по величине прибыльности активов;
- б) по величине мультипликатора капитала.

Для каждого ряда распределения постройте гистограмму, полигон и кумуляту распределения.

П р и м е ч а н и е . Прибыльность активов банка рассчитывается как отношение прибыли банка к величине активов. Мультипликатором капитала называется отношение величины активов к собственному капиталу банка.

6. Из табл. П1.2 осуществите выборку 30 наиболее крупных по численности населения стран мира. Постройте перечневую таблицу, содержащую следующую информацию по странам: «Расходы на НИОКР (в % от ВВП)», «Работники, занятые в НИОКР (на 1 млн чел.)», «ВВП на душу населения, долл. США». Постройте диаграммы разброса, позволяющие определить наличие взаимосвязи между указанными признаками.

7. Используя данные о распространенности информационных коммуникаций в отдельных странах мира (табл. П1.2) и международную классификацию стран по региональной принадлежности (табл. П1.5 и П1.6), заполните комбинационную таблицу (табл. 4.17).

Сформулируйте заголовок таблицы. Проанализируйте полученные результаты. Напишите аналитическую записку.

8. Найдите в газетах или журналах диаграммы разных типов: линейные, плоскостные, фигурные и т.д. Правильно ли, на ваш взгляд, выбран тип диаграмм в том или ином случае? Соблюдаются ли основные правила построения и оформления диаграмм?

9. Найдите в периодической печати диаграмму, которая, на ваш взгляд, искажает реальность, поскольку построена с нарушениями принципов графического изображения данных.

10. В табл. 4.18 приведены данные, характеризующие уровни мировых цен на основные сырьевые товары.

Для отражения динамики уровня мировых цен на конкретный сырьевой товар по данным таблицы постройте:

- а) столбиковую диаграмму;
- б) линейный график.

Для выявления общей тенденции изменения уровня цен дополните линейный график линией тренда. Поместите на диаграмму уравнение тренда и величину достоверности аппроксимации. Постройте точечный прогноз уровня мировых цен на сентябрь 2008 г.

Таблица 4.17

Макет таблицы

Уровень доходов населения	Регион*	Число стран	Распространенность на 1 тыс. чел.		
			магистральных телефонных линий	абонентов сотовых линий	пользователей сети Интернет
Низкий	Ц и ВЕ; СНГ				
	АГ				
	ВА и ТР				
	Прочие страны				
Итого по группе					
Средний	Ц и ВЕ; СНГ				
	АГ				
	ВА и ТР				
	Прочие страны				
Итого по группе					
Высокий	Ц и ВЕ; СНГ				
	АГ				
	ВА и ТР				
	Прочие страны				
Итого по группе					
Весь мир по странам мира					
* Ц и ВЕ; СНГ – Центральная и Восточная Европа; Содружество Независимых Государств; АГ – Арабские государства; ВА и ТР – Восточная Азия и Тихоокеанский регион.					

Варианты заданий:

- 1-й вариант: нефть;
- 2-й вариант: натуральный газ;
- 3-й вариант: бензин;
- 4-й вариант: медь;
- 5-й вариант: алюминий;
- 6-й вариант: никель.

Таблица 4.18

Среднемесячные мировые цены в сентябре соответствующего года

Сырьевой товар	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Нефть (Брент), долл./барр.	22,9	18,12	13,1	22,6	32,15	26,18	28,26	27,1	42,68	61,7	62,1
Натуральный газ, долл./ 1 млн БТЕ	2,743	2,251	1,858	2,62	5,118	2,204	3,625	4,678	5,483	12,745	10,41
Бензин, долл./ галлон	0,648	0,579	0,421	0,683	0,941	0,757	0,797	0,832	1,26	1,958	1,596
Медь, долл./ т	1932,6	2032,7	1676,4	1876,8	2037,5	1452,9	1498,3	1816,4	2892,6	3858,0	7602
Алюминий, долл./т	1428,3	1544,6	1391,3	1493,6	1600,2	1342,6	1294,7	1415,0	1726,0	1840,0	2473
Никель, долл./т	7428,4	6523,6	4229,8	6932	8854,3	5040,9	6592,5	9996,1	13298	14228	30131

Источник: Российская экономика. — 2006. — № 12. — С. 15.

11. По данным предыдущего примера постройте график, позволяющий сопоставить динамику цен на основные сырьевые товары за период 1996–2006 гг. Проанализируйте полученные результаты.

12. В табл. 4.19 приведены данные, характеризующие развитие высшего образования в России.

13. Имеются данные о динамике внешней торговли стран ЕС с Россией (табл. 4.20).

Постройте график, отражающий динамику торговли стран ЕС с Россией. Подберите уравнения трендов, наилучшим образом отражающие тенденции стоимостного объема: а) экспорта; б) импорта. Проанализируйте полученные результаты.

С помощью разных видов графиков отразите особенности развития высшего образования в России за последнее десятилетие.

14. Представьте в форме таблицы следующую информацию: «На текущем санитарном надзоре в Ленинградской области находится 366 канализационных очистных сооружений (КОС). В 2004 г. к 1-й группе риска относилось 76 КОС, ко 2-й – 160, к 3-й – 130. В 2005 г. распределение по группам изменилось: число КОС, относимых к 1-й и 2-й группам, увеличилось соответственно до 78 и 165, 123 КОС имели 3-ю группу риска».

Постройте ленточную диаграмму, отражающую долю КОС разной степени риска за каждый год. Сделайте выводы об изменении состояния КОС. Напишите краткую аналитическую записку.

Примечания: **1.** Для построения диаграммы в Excel используйте тип диаграммы «Линейчатая», вид – «Нормированная линейчатая диаграмма. Отражает долю каждой категории в общей сумме». **2.** В основу распределения КОС по группам положены следующие критерии: 1-я группа – КОС находятся в исправном техническом состоянии, условия водоотведения находятся в соответствии с ПДС; 2-я группа – КОС находятся в исправном техническом состоянии, однако вследствие технологических нарушений условия водоотвода не соответствуют ПДС; 3-я группа – КОС находятся в неисправном состоянии, эффективность очистки сточных вод низкая, условия водоотведения не соответствуют ПДС.

15. В табл. 4.21 приведены данные о распределении безработных в РФ по возрастным группам, в процентах.

Таблица 4.19

Развитие высшего образования в России

Показатель	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Численность студентов в вузах на 10 тыс. населения	188	200	220	244	277	324	373	410	448	480	495
Число государственных вузов	569	573	578	580	590	607	621	655	654	662	655
Число негосударственных вузов	193	244	302	334	349	358	387	384	392	409	413
Численность студентов государственных вузов, тыс. чел.	2655,2	2802,4	3046,5	3347,2	3728,1	4270,8	4797,4	5228,7	5596,2	5860,1	5985,0
Численность студентов негосударственных вузов, тыс. чел.	135,5	162,5	201,8	250,7	344,9	470,6	629,5	718,8	859,5	1024,1	1079,0
Принято студентов в государственные вузы – всего, тыс. чел.	628,6	674,3	748,3	813,8	946,4	1140,3	1263,4	1299,9	1411,7	1384,5	1362,7
В том числе на отделения:											
очное	402,9	422,2	459,9	492,6	539,1	621,9	669,3	698,3	724,8	758,2	743,9
вечернее	39,2	41,4	46,7	52,8	58,5	69,1	74,8	72,2	69,7	68,7	65,3
заочное	186,4	210,0	240,7	284,6	344,8	429,8	493,8	495,4	520,0	535,9	527,4
экстернат	0,1	0,7	1,0	1,8	4,0	19,5	25,5	34,0	97,2	22,0	26,1

Продолжение

Показатель	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Принято студентов в негосударственные вузы — всего, тыс. чел.	52,4	54,9	66,3	81,1	112,6	152,2	198,2	204,0	231,7	274,6	267,7
В том числе на отделения:											
очное	20,2	23,6	28,4	39,7	49,9	65,6	76,4	76,4	79,0	83,1	84,0
вечернее	5,4	5,4	5,3	8,0	9,7	12,8	13,5	13,4	14,1	18,6	19,8
заочное	23,9	24,8	31,5	32,3	52,4	73,6	107,9	113,2	136,9	169,7	159,0
экстернат	2,9	1,1	1,1	1,1	0,6	0,2	0,4	1,0	1,7	3,2	4,9

Источники: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 265, 271.

Таблица 4.20

Динамика внешней торговли стран ЕС с Россией

Год	1995	1996	1997	1998	1999	2000	2002	2003	2004	2005	2006
Экспорт, млн евро	15326	19132	25529	21087	14772	22512	31370	34274	37086	45712	56398
Импорт, млн евро	17858	23392	27037	23172	25918	60918	62976	61908	67779	80476	106729

Источники: Вопросы статистики. — 2006. — № 11. — С. 54.

Таблица 4.21

Возраст, лет	До 20	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	Старше 60
1995	11,1	18,3	14,1	13,0	14,2	11,3	9,0	3,5	3,9	1,6
2004	10,4	17,7	12,3	10,6	10,9	12,1	11,6	8,7	3,2	2,5

Источник: Труд и занятость в России. 2006: Стат. сб. — М.: Росстат, 2006. — С. 112.

На одном графике постройте полигоны распределения безработных по возрасту в 1995 и 2004 гг. Какие выводы об особенностях распределения безработных по возрасту можно сделать на основе построенного графика? Какие изменения произошли в возрастной структуре безработных за рассматриваемый период? Напишите аналитическую записку.

16. Имеются данные о потребностях в работниках, заявленные организациями в органы государственной службы занятости, по отраслям экономики в 2004 г. (табл. 4.22).

Таблица 4.22

Отрасль	Потребность в работниках, тыс. чел.	
	всего	в том числе рабочих
Всего по экономике	788,3	576,2
В том числе:		
промышленность	158,4	121,6
сельское хозяйство	48,9	39,8
строительство	83,6	66,3
транспорт и связь	38,6	30,0
торговля и общественное питание	85,9	62,2
ЖКХ и бытовое обслуживание	61,5	49,0
другие отрасли	311,4	207,3

Источник: Труд и занятость в России. 2006: Стат. сб. — М.: Росстат, 2006. — С. 108.

Рассчитайте для каждой отрасли потребность в работниках прочих (кроме рабочих) специальностей и постройте ленточную диаграмму (в Excel используйте тип диаграммы «Линейчатая», вид «Линейчатая диаграмма с накоплением»), отражающую размеры и структуру потребности каждой отрасли в работниках разных специальностей.

Постройте секторные диаграммы для характеристики отраслевой структуры общей потребности в работниках (в долях) и в том числе потребности в работниках рабочих специальностей. Какие дополнительные выводы, которые были не очевидны при табличном представлении данных, позволяют сделать построенные графики?

17. В табл. 4.23 приведены данные о смертности населения в РФ в 2005 г. от причин, связанных с употреблением алкоголя.

Таблица 4.23

Причина смертности	Число умерших, чел.	Число умерших на 100 тыс. населения, чел.
Хронический алкоголизм	5217	3,6
Алкогольные психозы	968	0,7
Алкогольная болезнь печени	15385	10,8
Случайные отравления алкоголем	40877	28,6
Алкогольная кардиомиопатия	38069	26,6
Дегенерация нервной системы, вызванная алкоголем	3754	2,6
Хронический панкреатит алкогольной этиологии	387	0,27
Всего умерших от вышеперечисленных причин	104657	73,1

Источник: Вопросы статистики. — 2006. — № 11. — С. 22.

Выполните графическое представление указанных данных. Тип и вид графика выберите самостоятельно. Проанализируйте результаты.

18. Используя нижеприведенные данные о структуре денежных доходов населения (в процентах) в двух странах СНГ в 2004 г., постройте диаграммы, позволяющие более наглядно представить различия в основных источниках дохода населения этих двух стран (табл. 4.24).

Таблица 4.24

Страна	Источники денежных доходов домашних хозяйств				
	Трудовые доходы в целом	Оплата работающим по найму	Социальные выплаты	Поступления от собственности и продажи недвижимости	Поступления от продажи сельскохозяйственной продукции
Беларусь	62,5	59,2	23,4	2,0	2,2
Грузия	52,1	33,8	7,8	5,8	17,2

Источник: Вопросы статистики. — 2006. — № 2. — С. 70.

Проанализируйте построенные диаграммы. Напишите краткие выводы.

19. По данным табл. 4.12 постройте графики, отражающие динамику:

1-й вариант — добычи нефти компанией «Роснефть» (млрд куб. м);

2-й вариант — выручки компании «Роснефть» (млрд долл.);

3-й вариант — чистой прибыли компании «Роснефть» (млрд долл.);

4-й вариант — текущих обязательств компании «Роснефть» (млрд долл.).

На каждом графике покажите линию тренда. Постройте прогноз на 2008 г.

4.3. Вопросы для самоконтроля

1. Назовите способы представления статистических данных. Почему собранные данные необходимо представить в компактной форме?

2. В чем достоинства и недостатки той или иной формы представления данных?

3. В каких случаях самой эффективной является табличная форма представления данных, а в каких — графическая?

4. Назовите основные правила оформления статистических таблиц.

5. В каких случаях вместо цифровых данных в таблице ставится знак «—» (прочерк), знак «...» (троеточие), делается запись 0,0?

6. Какие задачи решаются с помощью перечневых, групповых и комбинационных таблиц?
7. На какие группы делятся графики по решаемым задачам?
8. Сформулируйте основные принципы графического представления данных.
9. Какие виды графиков применяются для отображения структуры совокупности, динамики явлений?
10. Можно ли на основе графика получить представление о взаимосвязи переменных?
11. Как вы понимаете выражение «графический хлам»?

4.4. Методические рекомендации преподавателям

Аудиторная работа. Занятия по теме целесообразно проводить в компьютерных аудиториях, чтобы на практике ознакомить студентов с возможностями, которые предоставляют электронные таблицы и графические пакеты для организации и представления информации. Целесообразно заранее подготовить базы данных с первичной информацией в формате Excel.

Таблицы и графики должны быть построены так, чтобы облегчить восприятие данных, их сравнение и анализ. Они не должны быть перегружены излишней информацией или графическими эффектами. Чтобы лучше разобраться с принципами оформления таблиц и графиков, можно рассмотреть несколько примеров, в которых эти принципы были нарушены.

Самостоятельная внеаудиторная работа. В качестве домашней работы студентам могут быть предложены задания по формированию в формате электронных таблиц Excel массивов статистических данных, их группированию и представлению в форме таблиц и графиков. Необходимо стремиться к тому, чтобы задания были подобраны с учетом возможности использования созданных массивов данных при изучении последующих тем курса.

Одним из вариантов внеаудиторной работы может быть подборка примеров графического представления данных из газет и журналов с последующим анализом корректности их построения и использования.

Итоговый контроль по теме целесообразно проводить в форме решения комплексной задачи. В качестве примера может служить задание 13.

4.5. Методические указания студентам

Без таблиц и графиков невозможно представить повседневную деятельность бухгалтера, статистика, менеджера, финансового аналитика и других специалистов. Навыки, полученные на основе изучения материалов данной темы, будут обязательно востребованы. Несомненно, широкое распространение электронных таблиц и графических пакетов значительно упрощают процедуры представления данных. Однако ответственность за качество таблиц и графиков полностью возлагается на вас, а не на компьютер. Если вы хотите, чтобы построенные вами таблица или график были понятны и удобны для пользования, не пренебрегайте правилами и принципами представления статистических данных.

В результате изучения темы вы должны приобрести навыки построения простых и сводных (групповых и комбинационных) таблиц с помощью различного программного обеспечения, прежде всего наиболее доступного на сегодняшний день табличного процессора Excel. А также вы должны научиться грамотно строить и оформлять основные виды графиков.

Тема **5**

Средние величины

5.1. Основные положения

Главное значение средних величин состоит в их *обобщающей* функции. Для того чтобы обобщить множество различных индивидуальных значений признака отдельных единиц совокупности, рассчитывается их **средняя величина**, характеризующая всю совокупность в целом.

Если средняя величина обобщает качественно однородные значения признака, то она является *типической* характеристикой признака в данной совокупности. Например, уровень заработной платы работников торговой фирмы, средняя выработка в бригаде токарей, средняя рентабельность хлебопекарных предприятий и т.д. *Системные средние* могут характеризовать как пространственные или объектные системы, существующие одновременно (государство, отрасль, регион, мир в целом и т.д.), так и динамические системы, протяженные во времени (год, десятилетие, сезон и т.п.). Примерами системных средних являются средняя урожайность зерновых, средняя себестоимость строительства одного квадратного метра жилья, средний уровень потребления молока и молочных продуктов на душу населения и т.д.

Для нахождения средних величин используются различные формы и виды средних.

Средняя арифметическая простая рассчитывается по формуле

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

где \bar{x} — средняя величина;

x_i — значение признака i -й единицы совокупности;

n — число единиц совокупности.

Средняя арифметическая простая используется для осреднения *первичных признаков по несгруппированным данным*.

Пример 5.1. В табл. 5.1 приведены данные по договорам страхования загородной недвижимости, заключенных за неделю страховым агентом с физическими лицами.

Таблица 5.1

Данные по страхованию загородной недвижимости

Номер договора	Страховая сумма, тыс. долл.	Страховой платеж, тыс. долл.
1	25	1,5
2	20	1,2
3	30	1,4
4	18	0,9
5	24	1,2
Итого	117	6,2

Все договоры были заключены на разные страховые суммы, и соответственно по каждому договору были уплачены различные суммы платежей. Для обобщения данных рассчитаем средние величины. Поскольку и страховая сумма, и страховой платеж являются первичными признаками, для вычисления используем формулу средней арифметической простой.

Средняя страховая сумма одного договора составила:

$$\bar{x} = \frac{25 + 20 + 30 + 18 + 24}{5} = \frac{117}{5} = 23,4 \text{ тыс. долл.}$$

Средний платеж по одному договору:

$$\bar{x} = \frac{1,5 + 1,2 + 1,4 + 0,9 + 1,2}{5} = \frac{6,2}{5} = 1,24 \text{ тыс. долл.}$$

Если число единиц совокупности велико, исходные данные для вычисления средней могут быть представлены в сгруппированном виде — в виде *ряда распределения*. В этом случае для нахождения средней величины первичного признака используется **средняя арифметическая взвешенная**. В качестве весов будет выступать число единиц совокупности, обладающих тем или иным значением признака.

Пример 5.2. В табл. 5.2 приведен ряд распределения сотрудников фирмы по стажу работы.

Таблица 5.2

Распределение сотрудников фирмы по стажу работы

Стаж работы в данной организации, лет	Число сотрудников, чел.	Число сотрудников, в процентах к итогу
3	9	6
4	12	8
5	60	40
6	12	8
7	21	14
8	36	24
Итого	150	100

Ряд распределения называют еще и *вариационным рядом*. В вариационном ряду принято выделять следующие элементы:

- *варианты* распределения — те значения, которые принимает признак. В нашем примере это стаж работы;
- *частоты* распределения — число единиц совокупности с данным значением признака. В табл. 5.2 это число сотрудников в каждой группе.

Если число сотрудников с тем или иным стажем выразить в процентах к итогу, принимая все число сотрудников (150 чел.) за 100%, то получим относительные показатели частот распределения, которые принято называть *частотами*.

Средняя величина признака в вариационном ряду рассчитывается по формуле *средней арифметической взвешенной*:

$$\bar{x} = \frac{\sum_{j=1}^k x_j f_j}{\sum_{j=1}^k f_j},$$

где k — число групп;

x_j — значение признака в j -й группе (варианты);

f_j — число единиц в j -й группе (частоты) или процент единиц j -й группы в общей численности совокупности (частоты).

Частоты распределения могут быть выражены в долях ($\sum_{j=1}^k f_j = 1$). В этом случае средняя арифметическая взвешенная будет иметь вид

$$\bar{x} = \sum_{j=1}^k x_j f_j.$$

Средний стаж сотрудников фирмы по данным табл. 5.2 можно вычислить, используя в качестве весов как абсолютные, так и относительные показатели частот распределения.

Если известно число работников каждой группы, средний стаж рассчитаем как

$$\bar{x} = \frac{3 \cdot 9 + 4 \cdot 12 + 5 \cdot 60 + 6 \cdot 12 + 7 \cdot 21 + 8 \cdot 36}{150} = 5,9 \text{ года.}$$

Эту же величину среднего стажа получим, взвешивая по показателю удельного веса, выраженному в процентах:

$$\bar{x} = \frac{3 \cdot 6 + 4 \cdot 8 + 5 \cdot 40 + 6 \cdot 8 + 7 \cdot 14 + 8 \cdot 24}{100} = 5,9 \text{ года,}$$

или в долях:

$$\bar{x} = 3 \cdot 0,06 + 4 \cdot 0,08 + 5 \cdot 0,40 + 6 \cdot 0,08 + 7 \cdot 0,14 + 8 \cdot 0,24 = 5,9 \text{ года.}$$

Отметим, что средняя величина может быть дробным числом, даже если индивидуальные значения признака выражены только целочисленными значениями.

Если при группировке значения осредняемого признака заданы интервалами, то при расчете средней арифметической величины в качестве значений признака в группах принимают середины интервалов, т.е. исходят из гипотезы о равномерном распределении единиц совокупности по интервалу значений признака.

Пример 5.3. В табл. 5.3 приведено распределение продовольственных магазинов одного из районов города N-ска по площади торгового зала.

Таблица 5.3

Распределение магазинов по площади торгового зала

Площадь торгового зала, м ²	60–80	80–100	100–120	120–140	140–160	160–180	180–200
Число магазинов	6	15	18	23	15	9	5

Необходимо рассчитать среднюю величину площади торговых залов продовольственных магазинов.

Заменим интервальные значения признака серединами интервалов. Середина первого интервала $x_1 = \frac{60 + 80}{2} = 70$, второго

интервала $x_2 = \frac{80 + 100}{2} = 90$ и т. д.

Расчет средней величины в интервальном вариационном ряду осуществляется по формуле

$$\bar{x} = \frac{\sum_{j=1}^k x'_j f_j}{\sum_{j=1}^k f_j},$$

где x'_j – середины интервалов в j -й группе.

Промежуточные вычисления удобно оформлять в табличной форме (табл. 5.4).

Таблица 5.4

Расчетная таблица

Площадь торгового зала, м ²	Число магазинов f_j	Середина интервалов x'_j	$x'_j f_j$
60–80	6	70	420
80–100	15	90	1350
100–120	18	110	1980
120–140	23	130	2990
140–160	15	150	2250
160–180	9	170	1530
180–200	5	190	950
Итого	91	—	11470

Средняя площадь торгового зала продовольственного магазина составляет:

$$\bar{x} = \frac{11470}{91} = 126 \text{ м}^2.$$

Таким образом, для решения задачи была использована средняя взвешенная.

Иногда интервалы в первой и последней группах ряда распределения бывают открытыми, т.е. известна соответственно только их верхняя или нижняя граница. В этом случае перед нахождением середины интервалов их надо «закрыть» – найти условную нижнюю (или верхнюю) границу. Минимальное и максимальное значения признака определяют экспертным путем исходя из сущности свойств признака и совокупности. Но чаще исходят из величины интервала соседних групп.

Пример 5.4. В табл. 5.5 отражены данные о возрасте студентов-заочников.

Таблица 5.5

**Распределение студентов
заочного факультета по возрасту**

Возраст, лет	До 25	25–30	30–35	35–40	40 и более
Число студентов	8	112	154	86	40

По данным табл. 5.5 минимальный возраст студентов-заочников можно считать равным 20 годам (25 лет – 5 лет, где 5 лет – величина второго интервала), а максимальный возраст – 45 лет (40 лет + 5 лет, где 5 лет – величина предпоследнего интервала). Дальнейший расчет средней будет производиться по схеме, изложенной в примере 5.3. Находим середины интервалов (x'_j), затем произведения середин интервалов на частоты ($x'_j f_j$). Промежуточные итоги заносим в табл. 5.6.

Таблица 5.6

Расчетная таблица

Возраст, лет	Число студентов f_j	Середина интервалов x'_j	$x'_j f_j$
До 25	8	22,5	180
25–30	112	27,5	3080
30–35	154	32,5	5005
35–40	86	37,5	3225
40 и более	40	42,5	1700
Итого	400	—	13190

Средний возраст студента заочного факультета составил:

$$\bar{x} = \frac{13190}{400} = 33 \text{ года.}$$

Рассмотрим особенности выбора формулы для осреднения вторичного признака. Напомним, что вторичный признак — это признак, полученный на основе соотношения двух первичных признаков. Это соотношение должно сохраняться и при расчете средней величины вторичного признака, поэтому в дальнейшем будем называть его *исходным соотношением средней (ИСС)*.

Пример 5.5. Имеются данные об оплате труда на предприятии за месяц (табл. 5.7).

Таблица 5.7

Данные об оплате труда отдельных категорий работников предприятия

Категория работников	Средняя заработная плата, тыс. руб. (З)	Среднесписочная численность, чел. (Ч)	Фонд оплаты труда, тыс. руб. (Ф)	Удельный вес в общей численности работников, % (d)
	1	2	3	4
Рабочие	18,5	150	2775	75,0
Руководители	25,7	10	257	5,0
Специалисты	16,2	25	405	12,5
Прочие	8,8	15	132	7,5

Для того чтобы обобщить данные о заработной плате работников отдельных категорий, необходимо рассчитать средний уровень заработной платы для предприятия в целом. Исходная информация для расчета средней заработной платы может быть разной:

1) известны средняя заработная плата и численность отдельных категорий работников (гр. 1 и гр. 2);

2) известны средняя заработная плата отдельных категорий работников и удельный вес работников этих категорий в общей численности работников предприятия (гр. 1 и гр. 4);

3) известны средняя заработная плата и фонд оплаты труда отдельных категорий работников (гр. 1 и гр. 3);

4) известны численность и фонд оплаты труда отдельных категорий работников (гр. 2 и гр. 3).

Для того чтобы правильно выбрать формулу для вычисления средней заработной платы, составим ИСС, отражающее сущность осредняемого признака, его логическое содержание. Исходное соотношение для нахождения средней заработной платы будет выглядеть следующим образом:

$$\begin{aligned} \text{Средняя заработная плата (З)} &= \\ &= \frac{\text{Фонд оплаты труда (Ф)}}{\text{Численность работников (Ч)}}. \end{aligned}$$

Зная два элемента из ИСС, всегда можно найти значение третьего.

Для нахождения среднего уровня заработной платы на предприятии необходимо найти частное от деления итоговых подсчетов по фондам оплаты труда и численности работников четырех выделенных категорий:

$$\bar{З} = \frac{\sum \Phi}{\sum Ч}.$$

Рассмотрим, как будет преобразовываться эта формула в зависимости от характера исходной информации.

В первом случае, когда имеются данные о средней заработной плате и численности работников отдельных категорий, в ИСС не известен числитель, т.е. величина фонда оплаты

труда. Для каждой категории работников фонд оплаты труда может быть определен как

$$\Phi = \sum z \cdot q.$$

Подставляя выражение для Φ в формулу в целях нахождения средней заработной платы, получим формулу *средней арифметической взвешенной*. Роль признака-веса выполняет численность работников:

$$\bar{z} = \frac{\sum z \cdot q}{\sum q} = \frac{\sum z \cdot q}{\sum q}.$$

Вычислим среднюю заработную плату для нашего примера:

$$\begin{aligned} \bar{z} &= \frac{\sum z \cdot q}{\sum q} = \frac{18,5 \cdot 150 + 25,7 \cdot 10 + 16,2 \cdot 25 + 8,8 \cdot 15}{150 + 10 + 25 + 15} = \\ &= 17,845 \text{ тыс. руб.} \end{aligned}$$

Средняя заработная плата на предприятии составляет 17,845 тыс. руб.

Рассмотрим второй случай, когда известны не абсолютные численности работников отдельных категорий, а их относительные величины — удельные веса (доли) в общей численности работников предприятия ($d = q/\sum q$). Формула средней арифметической взвешенной примет вид

$$\bar{z} = \sum z \cdot d,$$

если d — удельные веса, выраженные долями ($\sum d = 1$), или

$$\bar{z} = \frac{\sum z \cdot d}{\sum d},$$

если d — удельные веса в процентах ($\sum d = 100\%$).

Воспользуемся второй формулой:

$$\begin{aligned} \bar{z} &= \frac{\sum z \cdot d}{\sum d} = \frac{18,5 \cdot 75,0 + 25,7 \cdot 5,0 + 16,2 \cdot 12,5 + 8,8 \cdot 7,5}{75,0 + 5,0 + 12,5 + 7,5} = \\ &= 17,845 \text{ тыс. руб.} \end{aligned}$$

В третьем случае среднюю заработную плату на предприятии надо определить исходя из данных о средних заработных платах и фондах оплаты труда отдельных категорий работников. Теперь в *ИСС* не известен знаменатель — численность работников, но для отдельных категорий ее можно найти по формуле

$$\varphi = \frac{\Phi}{\bar{z}}.$$

Подставляя это выражение в формулу для расчета средней заработной платы, получим формулу *средней гармонической взвешенной*, где в качестве веса выступает фонд оплаты труда:

$$\bar{z} = \frac{\sum \Phi}{\sum \varphi} = \frac{\sum \Phi}{\sum \frac{\Phi}{\bar{z}}}.$$

Средняя заработная плата по предприятию, вычисленная по формуле средней гармонической, составит:

$$\bar{z} = \frac{\sum \Phi}{\sum \frac{\Phi}{\bar{z}}} = \frac{2775 + 257 + 405 + 132}{\frac{2775}{18,5} + \frac{257}{25,7} + \frac{405}{16,2} + \frac{132}{8,8}} = 17,845 \text{ тыс. руб.}$$

В четвертом случае, когда известны данные о численности и фондах оплаты труда отдельных категорий работников, т.е. известны и числитель, и знаменатель *ИСС*, расчет средней заработной платы по предприятию сводится к подстановке данных в исходную формулу:

$$\bar{z} = \frac{\sum \Phi}{\sum \varphi} = \frac{2775 + 257 + 405 + 132}{150 + 10 + 25 + 15} = 17,845 \text{ тыс. руб.}$$

Эту формулу расчета средней величины называют *агрегатной средней* или *неявной средней*, поскольку индивидуальные значения осредняемого признака не присутствуют в расчете в явном виде.

5.2. Задания по теме

1. В компании по продаже недвижимости выставлены на продажу три однокомнатные квартиры (табл. 5.8).

Таблица 5.8

Квартира	Общая площадь, м ²	Процент жилой площади	Цена за 1 м ² , долл.	Стоимость квартиры, тыс. долл.
1	45,5	55	1660	75,5
2	37,8	48	1635	61,8
3	32,4	56	1615	52,3

Обобщите данные о реализуемых однокомнатных квартирах, рассчитав:

- средний размер общей площади;
- средний процент жилой площади;
- средний размер жилой площади;
- среднюю цену 1 м²;
- среднюю стоимость однокомнатной квартиры.

2. Имеются данные о работе трех филиалов промышленного объединения (табл. 5.9).

Таблица 5.9

Филиал	Выпуск продукции, млн руб.		Процент выполнения бизнес-плана	Процент отгруженной продукции
	Бизнес-план	Фактически		
	1	2	3	4
1	25	25,5	102,0	78
2	15	14,7	98,0	95
3	38	40,1	105,5	89

1. Определите средний процент выполнения бизнес-плана в целом по объединению, используя показатели:

- граф 1 и 2;
- граф 2 и 3;
- граф 1 и 3.

2. Рассчитайте процент отгруженной продукции в среднем по объединению.

3. Имеются данные о реализации туристических путевок в Российской Федерации в 2004 г. (табл. 5.10).

Таблица 5.10

Направление туризма	Средняя цена путевки, тыс. руб.	Число проданных путевок		Общая стои- мость продан- ных путевок, млн руб.
		тыс. шт.	% к итогу	
По территории России	12,3	894	44	11003
По зарубежным странам	16,5	1140	56	18843

Вычислите среднюю цену одной туристической путевки всеми возможными способами.

4. Имеются данные о реализации картофеля на трех рынках города (табл. 5.11).

Таблица 5.11

Рынок	Сентябрь		Октябрь	
	Цена за 1 кг, руб.	Продано, т	Цена за 1 кг, руб.	Продано на сумму, тыс. руб.
1	10,0	50	11,0	517
2	9,5	45	10,5	420
3	8,5	60	9,5	475

Рассчитайте средние цены реализации картофеля в сентябре и октябре.

5. Выпуск продукции первого сорта двумя цехами завода за отчетный период характеризуется следующими данными (табл. 5.12).

Таблица 5.12

Цех	Бизнес-план		Фактически	
	Удельный вес продукции первого сорта, %	Стоимость продукции первого сорта, млн руб.	Удельный вес продукции первого сорта, %	Стоимость всей продукции, млн руб.
1	90	22,5	92	27,5
2	85	17	80	19,0

Определите:

а) запланированный и фактический удельный вес продукции первого сорта в целом по предприятию;

б) процент выполнения бизнес-плана по выпуску продукции первого сорта и по выпуску всей продукции.

6. Имеются данные по объектам офисной недвижимости (табл. 5.13).

Таблица 5.13

Офисный объект	Общая площадь здания, м ²	Процент арендуемых площадей	Средняя ставка аренды, долл. за 1 м ²	Средние эксплуатационные расходы, долл. за 1 м ²	Средний размер офиса, м ²
Бизнес-центр «Люкс»	120142	76	615	120	400
«Капитал»	40800	64	700	105	380
«Горизонт»	35150	88	695	110	250

Определите средние показатели рынка офисной недвижимости.

7. Продукция **А** производится в отрасли на двух предприятиях. В табл. 5.14 приведены данные, характеризующие производство на каждом предприятии за два последних года.

Таблица 5.14

Предприятие	Отчетный год		Базисный год	
	Произведено изделий, тыс. шт.	Себестоимость одного изделия, руб.	Произведено изделий, тыс. шт.	Затраты на производство, млн руб.
1	45	750	40	28
2	55	850	60	48

Рассчитайте среднеотраслевую себестоимость производства продукции **А** за каждый год.

8. Имеются данные о работе страховых агентов компании «НЕВА» за месяц (табл. 5.15).

Таблица 5.15

Ф.И.О.	Стаж работы в компании, лет	Собрано страховых платежей по договорам имущественного и личного страхования		Число договоров имущественного страхования	Процент договоров, заключенных с физическими лицами
		всего, тыс. руб.	в среднем на один договор, руб.		
Иванов А.Б.	1,5	52,5	3500	2	67
Клочков П.В.	2	85,1	3700	20	83
Митин К.А.	4	86,64	2280	30	72
Грибов А.Л.	2,5	75,62	3980	8	79

- Обобщите данные о работе страховых агентов, рассчитав:
- а) средний стаж работы;
 - б) среднее число договоров, заключенных одним агентом;
 - в) средний размер страхового платежа по одному договору страхования;
 - г) средний удельный вес договоров имущественного страхования;
 - д) средний удельный вес договоров, заключенных с юридическими лицами.

9. В табл. 5.16 приведены некоторые данные, характеризующие изменение численности населения, проживающего в районах Крайнего Севера, полученные по итогам переписей населения РФ в 1989 и 2002 гг.

Таблица 5.16

Район	Перепись населения 1989 г.		Перепись населения 2002 г.	
	Численность населения, тыс. чел.	Численность городского населения, тыс. чел.	Численность населения по сравнению с 1989 г., %	Удельный вес городского населения, %
Европейский Север	4775,3	3810,5	83,0	78,9
Азиатский Север	4998,2	3888,6	87,9	79,5

Определите:

- 1) средний процент городского населения для Северных районов РФ для каждого года;
- 2) на сколько процентов уменьшилась численность населения, проживающего в Северных районах РФ, за межпереписной период;
- 3) средний процент изменения численности:
 - а) городского населения Севера;
 - б) сельского населения Севера.

10. Известны данные об объеме кредиторской задолженности на предприятиях отрасли (табл. 5.17).

Таблица 5.17

Пред- приятие	Общая сумма кредиторской задолженности, тыс. руб.	Просроченная кредиторская задолженность, тыс. руб.	Удельный вес просроченной кредиторской задолженности, %
1	4500	225	5
2	2700	324	12
3	3000	300	10

Определите процент просроченной задолженности в среднем для отрасли всеми возможными способами.

11. Известны данные, характеризующие реализацию туристических путевок по странам Европы турфирмой «МИР» за месяц (табл. 5.18).

Таблица 5.18

Страна	Число про- данных путевок	Средняя цена путевки, руб.	Удельный вес недельных туров (8 дней/ 7 ночей) в общем числе проданных путевок, %	Процент «горящих» путевок в общем числе проданных
Австрия	84	18480	81	7,1
Венгрия	75	20650	67	9,3
Германия	54	19450	93	18,5

Определите средние показатели по указанным направлениям.

12. В табл. 5.19 приведены показатели, характеризующие жилищный фонд РФ в городской и сельской местности по состоянию на начало 2005 г.

Таблица 5.19

Жилищный фонд	Площадь жилищ- ного фонда, млн м ²		Процент площади, оборудованной			
	всего	в том числе в собствен- ности граждан	водо- про- водом	кана- лиза- цией	газом	горячим водо- снаб- жением
Городской	2115	1444	87	85	68	78
Сельский	802	659	42	33	75	20

Рассчитайте среднероссийские показатели:

а) процент жилой площади, находящейся в собственности граждан;

б) процент жилого фонда, оборудованного:

- водопроводом;
- канализацией;
- газом;
- горячим водоснабжением.

13. В табл. 5.20 приведены экономические показатели развития двух столичных регионов России за 2003 г.

Таблица 5.20

Регион	ВРП, млн руб.	ВРП на душу населе- ния, руб.	ВРП на 1 руб. основ- ных фондов, руб.	Инвести- ции в ос- новной капитал, в процен- тах к ВРП	Инвести- ции на 1 руб. основных фондов, руб.
Москва	2441426	234997	0,88	11,0	0,097
Московская область	447055	67537	0,39	23,5	0,092
Санкт- Петербург	435683	93891	0,53	25,6	0,137
Ленинградская область	132386	79583	0,34	37,5	0,128

Рассчитайте средние величины относительных экономических показателей:

а) для Москвы и Санкт-Петербурга;

б) для Москвы и Московской области;

в) для Санкт-Петербурга и Ленинградской области.

14. Известны данные о депозитах, привлеченных от физических лиц, в трех крупнейших по величине активов банках Северо-Западного региона по состоянию на 1 июля 2006 г. (табл. 5.21).

Таблица 5.21

Банк	Величина активов, тыс. руб.	Депозиты физических лиц, тыс. руб.	Депозиты физических лиц - нерезидентов, тыс. руб.	Доля валютных депозитов, %	Доля депозитов физических лиц в совокупных обязательствах, %
Промышленно-строительный банк	133469960	23727617	126033	27,9	20,0
Банк «Санкт-Петербург»	40890909	8516065	45378	36,9	23,9
Национальный банк «Траст»	40230228	6024484	135746	32,0	16,8

Определите в среднем для трех банков:

- а) долю депозитов физических лиц в совокупных обязательствах;
- б) долю валютных депозитов физических лиц;
- в) долю депозитов нерезидентов в общей сумме депозитов физических лиц.

15. В табл. 5.22 приведены некоторые показатели деятельности десяти крупнейших по размеру собственного капитала (на начало 2006 г.) банков Санкт-Петербурга.

Таблица 5.22

Банк	Собственный капитал на 01.01.2006 г., тыс. руб.	Изменение собственного капитала за 1-е полугодие 2006 г., %	Норма прибыли на капитал в 1-м полугодии 2006 г., %	Прибыльность активов в 1-м полугодии 2006 г., %
Промышленно-строительный банк	9991118	22,4	22,8	2,8
КИТ Финанс Инвестиционный банк	3155442	63,6	60,1	10,3
Россия	3130277	3,0	11,6	1,5
Банк «Санкт-Петербург»	3095878	36,0	12,5	1,4

Продолжение

Банк	Собственный капитал на 01.01.2006 г., тыс. руб.	Изменение собственного капитала за 1-е полугодие 2006 г., %	Норма прибыли на капитал в 1-м полугодии 2006 г., %	Прибыльность активов в 1-м полугодии 2006 г., %
Национальный банк «Траст»	2914956	10,1	4,6	0,4
Дрезднер банк	2587542	1,7	3,9	0,8
Международный банк Санкт-Петербурга	1505483	1,0	7,1	0,6
КАЛИОН				
РУСБАНК	1468061	3,0	11,0	1,1
Инкасбанк	1159986	7,4	5,9	0,7
Банк ВЕФК	1139447	1,6	6,2	0,7

1. Разделите банки на две группы:

а) банки, в которых размер собственного капитала за 1-е полугодие 2006 г. вырос менее чем на 10%;

б) банки, в которых размер собственного капитала за 1-е полугодие 2006 г. вырос более чем на 10%.

2. Для каждой группы вычислите:

а) среднюю норму прибыли на капитал;

б) среднюю прибыльность активов.

Примечание. Норма прибыли на капитал рассчитывается как отношение прибыли банка к размеру собственного капитала. Прибыльность активов – отношение прибыли банка к сумме активов.

16. В табл. 5.23 приведены данные по компаниям-лидерам в сфере телекоммуникаций.

Таблица 5.23

Компания	Объем реализации в 2005 г., млн руб.	Темп прироста объема реализации в 2005 г. по сравнению с 2004 г., %	Балансовая прибыль в 2005 г., млн руб.	Доля чистой прибыли в балансовой, %	Чистая прибыль на 1 руб. объема реализации, руб.
«Мобильные телесистемы»	141761,2	26,5	44242,7	72,0	0,225
«Вымпелком»	90842,5	49,2	23775,8	73,1	0,192
«Мегафон»	67564,2	58,4	15302,0	72,5	0,165

Вычислите в среднем для компаний:

- а) темп прироста объемов реализации;
- б) балансовую прибыль на 1 руб. объема реализации;
- в) чистую прибыль на 1 руб. объема реализации;
- г) долю чистой прибыли в балансовой.

Как взаимосвязаны между собой три последние средние?

17. Известны показатели социально-экономического развития основных нефтегазовых районов России в 2002 г. (табл. 5.24).

Таблица 5.24

Регион	Средне- довая чис- ленность населения, тыс. чел.	Объем промыш- ленного производ- ства на душу населения, тыс. руб.	ВРП на душу населения, тыс. руб.	Инве- стиции в основной капитал на душу населе- ния, тыс. руб.	Средне- месяч- ный до- ход, руб. на 1 чел.
Ямало- Ненецкий автономный округ	509	322,9	546,7	240,9	13388,1
Ханты- Мансийский автономный округ	1424	337,3	405,7	105,8	10883,0
Республика Коми	1117	56,6	83,8	20,1	5737,5
Томская область	1061	39,7	76,4	16,1	3896,3
Ненецкий автономный округ	44	261,6	363,9	269,9	7723,4
Республика Саха	983	91,1	117,0	24,5	6218,6

Рассчитайте среднедушевые показатели развития в целом для всех нефтегазовых районов Российской Федерации. К каким средним (типическим или системным) относятся вычисленные вами показатели?

Сравните полученные значения со среднероссийскими показателями. Результаты оформите в таблице (табл. 5.25).

Таблица 5.25

Показатели развития экономики в 2002 г. (на душу населения)

Показатель	Россия в целом	Нефтегазо- вые районы России	Нефтегазовые районы, в процен- тах к среднерос- сийскому уровню
Объем промышлен- ного производства, тыс. руб.	47,9		
ВРП, тыс. руб.	66,0		
Инвестиции в основ- ной капитал, тыс. руб.	12,3		
Среднемесячные дохо- ды населения, руб.	3947,2		

18. В табл. 5.26 приведены данные о деятельности крупней-
ших компаний в сфере розничной торговли за 2005 г.

Таблица 5.26

Компания	Выручка от про- дажи, млн руб.	Чистая при- быль, млн руб.	Рентабельность продаж, %
	1	2	3
«Магнит»	44632,3	1042,0	2,3
«Пятерочка»	38454,6	2580,0	6,7
«Перекресток»	28708,3	820,9	2,9
«Метро»	53834,4	2816,8	5,2

1. Определите средний уровень рентабельности продаж, если известны данные:

- а) граф 1 и 2;
- б) граф 1 и 3;
- в) граф 2 и 3.

2. Укажите, какие виды средних величин вы использовали.

П р и м е ч а н и е. Рентабельность продаж определяется отноше-
нием чистой прибыли к выручке от продаж.

19. Имеются данные об объемах деятельности пяти круп-
нейших банков (табл. 5.27) и пяти страховых компаний России
(табл. 5.28).

Таблица 5.27

Банк	Объем реализации в 2005 г., млн руб.	Темп прироста по сравнению с предыдущим годом, %
Сбербанк России	311892,1	45,0
Внешторгбанк (ВТБ)	72422,4	56,4
Газпромбанк	56098,3	59,6
Уралсиб	33963,7	80,7
Группа «Альфа-банк»	29777,8	23,8

Таблица 5.28

Страховая компания	Объем реализации в 2004 г., млн руб.	Темп прироста в 2005 г., % к 2004 г.
Российская государственная страховая компания («Росгосстрах»)	111731,2	42,1
«Уралсиб»	26277,4	39,0
Группа «Согаз»	17526,0	68,3
«Ингосстрах»	19364,1	43,9
Страховой дом ВСК	13760,1	42,8

1. Рассчитайте средние темпы прироста объемов реализации лидеров:

- а) банковской деятельности;
- б) страховой деятельности.

2. Сравните, в каком из финансовых секторов были достигнуты более высокие темпы прироста.

Примечание. Под термином «объем реализации» для банков понимается сумма процентных доходов (до вычета расходов), комиссионных доходов и непроцентных доходов за вычетом расходов по операциям с ценными бумагами и иностранной валютой; для страховых компаний – сумма подписанных страховых премий и доходов от инвестиций за вычетом премий, переданных в перестрахование.

20. В табл. 5.29 приведены некоторые характеристики бореальных лесов Европейского Севера.

Таблица 5.29

Регион	Лесопокрытая площадь, млн га	Площадь, покрытая хвойными лесами, в процентах от общей лесопокрытой площади	Общий запас древесины, млн м ³	Запас эксплуатационных лесов, млн м ³	Лесопокрытая площадь на одного жителя, га
Восточная часть европейских бореальных лесов					
Мурманская обл.	4,354	73,3	200	160	3,96
Республика Карелия	9,267	89,3	919	414	11,58
Архангельская обл.	19,851	86,0	2150	1634	12,41
Республика Коми	28,751	80,0	2837	2190	26,14
Западная часть европейских бореальных лесов					
Норвегия	8,9	68,5	621	—	1,78
Швеция	23,5	84,7	2900	—	2,47
Финляндия	20,0	80,0	1773	—	3,92

Определите:

- средний процент площади, покрытой хвойными лесами, от общей лесопокрытой площади;
- средний запас древесины на 1 га лесопокрытой площади;
- средний запас эксплуатационных лесов на одного жителя;
- средний запас эксплуатационных лесов на 1 га лесопокрытой площади;
- среднюю лесопокрытую площадь на одного жителя.

Задание выполняется по двум вариантам:

- восточная часть европейских бореальных лесов;
- европейские бореальные леса в целом.

21. Из путевого листа грузового автомобиля, принадлежащего автотранспортному цеху железобетонного комбината, известны данные о перевозках за день (табл. 5.30).

Таблица 5.30

Ездка	Грузополучатель	Пробег, км		Перевезено, т
		с грузом	без груза	
1	СУ-1	19	19	15
2	ДСК «ЛУЧ»	25	25	18
3	СУ-1	19	19	12
4	СУ-2	23	23	12

Рассчитайте:

- а) среднюю дальность ездки с грузом (среднее расстояние перевозки груза);
- б) среднюю дальность одной ездки;
- в) средний вес перевезенного груза за одну ездку;
- г) средний грузооборот, выполненный за одну ездку.

Примечание. Грузооборот – количество выполненных тонно-километров – определяется умножением массы перевезенных грузов на расстояние пробега с грузом.

22. По данным табл. 5.31 вычислите среднюю урожайность всех зерновых культур в целом за текущий и базисный периоды. Объясните причины изменения урожайности.

Таблица 5.31

Вид зерновых культур	Доля в посевной площади, %		Урожайность, ц/га	
	базисный период	текущий период	базисный период	текущий период
Пшеница озимая	30	20	40	42
Пшеница яровая	20	20	25	26
Ячмень	20	35	20	24
Гречиха	5	15	8	10
Кукуруза на зерно	25	10	60	62
Итого	100	100	—	—

23. Среди 400 крупнейших по объему реализации компаний России в 1-м полугодии 2006 г. наибольшую сумму чистой прибыли имел «Газпром» – 315933 млн. руб., одну из наименьших – Косогорский металлургический комбинат – около 200 тыс. руб.

Какую форму средних величин следует применить, чтобы найти среднюю сумму чистой прибыли. Вычислите эту среднюю величину.

5.3. Вопросы для самоконтроля

1. Какую функцию выполняют средние величины при раскрытии статистических закономерностей?

2. В чем заключается различие между типической средней величиной и системной средней? Приведите примеры типических и системных средних.

3. Для каких задач применяют простые средние, а для каких — взвешенные?

4. В каких случаях простая и взвешенная арифметические средние равны между собой?

5. В чем состоят причины различий простой и взвешенной средних?

6. Как рассчитывается средняя арифметическая величина на основе интервального ряда распределения?

7. Можно ли вычислить среднюю величину признака в совокупности, если неизвестно ни одно индивидуальное значение признака?

8. В чем состоит свойство мажорантности средних величин?

9. Сформулируйте основные свойства средней арифметической величины.

10. Докажите, что среднее отклонение индивидуальных значений признака от средней арифметической величины равно нулю.

11. Докажите, что сумма квадратов отклонений индивидуальных значений признака от средней арифметической меньше, чем от любого другого числа.

5.4. Методические рекомендации преподавателям

Аудиторная работа. Практическое занятие по теме целесообразно начать с примера на обобщение значений различных признаков (первичных и вторичных) по конкретной совокупности (по типу заданий № 1, 6, 8, 11 и т.п.). Для упро-

щения вычислений число единиц совокупности не должно быть большим. Главное внимание должно быть уделено сознательно-му выбору формулы расчета средней. При решении задач необходимо обсуждать со студентами, почему именно по этой, а не по другой формуле надо вычислять среднюю. Если студенты предлагают ошибочные варианты решения, целесообразно показать, к каким неправильным результатам может привести нарушение принципов выбора формы средней.

При осреднении вторичных признаков целесообразно показать, как меняется форма расчета средней в зависимости от характера исходной информации (см. задания **4, 7, 10, 18** и т.п.).

Решение каждого задания должно завершаться содержательной интерпретацией полученных результатов.

Самостоятельная внеаудиторная работа. Для отработки навыков расчета средних в качестве домашней работы студентам должен быть предложен набор задач на все виды и формы средних величин.

Итоговый контроль целесообразно провести в форме тестирования. В качестве тестов могут быть предложены как теоретические вопросы, так и краткие задачи.

5.5. Методические указания студентам

Освоение теории. Средняя величина — один из наиболее часто используемых статистических показателей. Прежде чем приступить к решению практических примеров, внимательно изучите теорию. Уясните такие понятия, как *однородность* и *вариация* массовых явлений. Почему одни средние можно считать *типическими* характеристиками совокупностей, а другие относятся к категории *системных* средних?

Наиболее часто вам придется вычислять среднюю величину по формуле средней арифметической. Обратите внимание, в каких случаях используется простая, а в каких — взвешенная форма средней? Обязательно запомните основные свойства средней арифметической величины. Знание математических свойств средней полезно как при ее использовании, так и при расчетах.

При вычислении средней величины вторичного признака особое внимание обращайтесь на социально-экономическое содержание и выяснение сущности осредняемого признака. Это

поможет вам правильно составить *исходное соотношение средней* (ИСС). При выборе формулы расчета средней величины вторичного признака рекомендуем придерживаться следующего правила:

- если известен знаменатель ИСС, но не известен числитель, то средняя величина вторичного признака определяется по формуле средней арифметической взвешенной;

- если известен числитель ИСС, но не известен знаменатель, то средняя вычисляется по формуле средней гармонической взвешенной;

- если известны и числитель, и знаменатель ИСС, то для нахождения средней величины используют неявную форму средней.

Это на первый взгляд формальное правило вытекает из требования конструировать формулу той или иной средней величины в соответствии с логикой показателя, его сущностью, социально-экономическим содержанием.

В результате изучения темы вы должны знать понятие средней величины, значение средних, виды и формы расчета.

Практические навыки. Нужно уметь вычислять средние величины первичных и вторичных признаков по негруппированным и сгруппированным данным.

Тема 6

Изучение вариации

6.1. Основные положения

Вариацией значений какого-либо признака в совокупности называется различие его значений у разных единиц данной совокупности в один и тот же период или момент времени.

Изучение вариации предполагает выполнение следующих основных этапов:

- построение вариационного ряда;
- расчет показателей центра распределения и структурных характеристик вариационного ряда;
- расчет показателей размера и интенсивности вариации;
- оценка вариационного ряда на асимметрию и эксцесс;
- графическое изображение вариационного ряда.

Построение вариационного ряда — упорядоченное распределение единиц совокупности по возрастающим или убывающим значениям признака и подсчет числа единиц с тем или иным значением признака. Различают три формы вариационного ряда: ранжированный, дискретный, интервальный.

Ранжированный ряд — перечень отдельных единиц совокупности в порядке возрастания (убывания) изучаемого признака.

Пример 6.1. В табл. 6.1 приведен ранжированный ряд, отражающий данные об объемах реализации десяти крупнейших компаний России в сфере информационных технологий в 2005 г.

Таблица 6.1

Крупнейшие компании России в сфере информационных технологий

Компания	Объем реализации, млрд руб.
Концерн «Ситроникс»	27,0
Национальная компьютерная корпорация	22,7
Группа компаний ЛАНИТ	17,2
Группа компаний IBS	16,3
«Техносервис А/С»	15,2

Продолжение

Компания	Объем реализации, млрд руб.
Группа компаний Vervysell	11,6
«НТ Компьютер»	9,7
«КРОК инкорпорейтед»	9,0
«Открытые технологии – 98»	6,9
«Микротест»	5,7

Рассчитаем основные характеристики данного вариационного ряда.

Среднее значение признака вычисляем по формуле средней арифметической простой:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

где x_i — значение признака i -й единицы совокупности;
 n — число единиц совокупности.

$$\begin{aligned}\bar{x} &= \frac{27,0 + 22,7 + 17,2 + 16,3 + 15,2 + 11,6 + 9,7 + 9,0 + 6,9 + 5,7}{10} = \\ &= \frac{141,3}{10} = 14,13 \text{ млрд руб.}\end{aligned}$$

Мода — значение признака, наиболее часто встречающееся в изучаемой совокупности. Поскольку в анализируемом ряду нет значений, которые встречались бы более одного раза, мода в данном ряду отсутствует.

Медиана — значение признака, делящее ранжированный вариационный ряд на две равные части таким образом, что половина единиц совокупности имеет значения признака меньше, чем медиана, а половина — больше, чем медиана. Для ранжированного ряда с нечетным числом элементов медианым является значение признака у единицы под номером $(n + 1) : 2$. Для ранжированного ряда с четным числом элементов медианой будет средняя арифметическая из значений признака у двух смежных единиц, расположенных в середине ряда. В рассматриваемом вариационном ряду, который состоит из 10 единиц,

медианой будет средняя из значений признака 5-й и 6-й единицы совокупности:

$$Me = \frac{15,2 + 11,6}{2} = 13,4 \text{ млрд руб.}$$

Следовательно, половина из рассматриваемых компаний имела объем реализации больше чем 13,4 млрд руб., а половина — меньше чем 13,4 млрд руб.

Для характеристики размера вариации рассчитывают абсолютные показатели вариации: размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, дисперсию. Все они, за исключением дисперсии, являются величинами именованными, т.е. имеют ту же единицу измерения, что и изучаемый признак. Дисперсия единиц измерения не имеет.

Размах вариации (размах колебаний) представляет собой разность между максимальным и минимальным значениями признака в совокупности:

$$R = x_{\max} - x_{\min}.$$

В нашем примере разница между максимальным и минимальными объемами реализации составляет: $27,0 - 5,7 = 21,3$ млрд руб.

Для несгруппированных данных:

- *среднее линейное отклонение*

$$\bar{a} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n};$$

- *среднее квадратическое отклонение*

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}};$$

- дисперсия

$$\sigma^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n}.$$

Для промежуточных вычислений построим расчетную таблицу (табл. 6.2).

Таблица 6.2

Расчетная таблица

№ п/п	Объем реализации, млрд руб. (x_i)	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$
1	27,0	12,87	165,6369
2	22,7	8,57	73,449
3	17,2	3,07	9,4249
4	16,3	2,17	4,7089
5	15,2	1,07	1,1449
6	11,6	2,53	6,4009
7	9,7	4,43	19,6249
8	9,0	5,13	26,3169
9	6,9	7,23	52,2729
10	5,7	8,43	71,0649
Итого	141,3	55,5	430,041

Среднее линейное отклонение составит:

$$\bar{a} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{55,5}{10} = 5,55 \text{ млрд руб.};$$

- дисперсия:

$$\sigma^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n} = \frac{430,041}{10} = 43,0041;$$

- среднее квадратическое отклонение:

$$\sigma = \sqrt{\sigma^2} = \sqrt{43,0041} = 6,557751 \approx 6,56 \text{ млрд руб.}$$

Это означает, что в среднем объем реализации в изучаемой совокупности компаний отклонялся от среднего значения этого признака на 5,55 млрд руб. по среднему линейному отклонению и на 6,56 млрд руб. по среднему квадратическому.

Расчет показателей вариации вручную достаточно трудоемок, особенно при больших объемах совокупностей. Стандартные статистические методы обработки данных включены в состав электронных таблиц, математических пакетов (Mathcad, Mathlab и др.), специализированных статистических пакетов (STATISTICA, SPSS, STATGRAPHICS и др.). Рассмотрим порядок расчета показателей вариации для нашего примера с помощью табличного процессора Microsoft Excel.

Сформируем таблицу с исходными данными на рабочем листе Microsoft Excel (табл. 6.3).

Таблица 6.3

Исходные данные

	А	В
1	Компания	Объем реализации, млрд руб.
2	Концерн «Ситроникс»	27,0
3	Национальная компьютерная корпорация	22,7
4	Группа компаний ЛАНИТ	17,2
5	Группа компаний IBS	16,3
6	«Техносервис А/С»	15,2
7	Группа компаний Verysell	11,6
8	«НТ Компьютер»	9,7
9	«КРОК инкорпорейтед»	9,0
10	«Открытые технологии – 98»	6,9
11	«Микротест»	5,7

Активизируем любую свободную ячейку, в которой мы хотим поместить результат вычисления того или иного показателя. Например, **В13**. Запускаем мастер функций командой **Функция** из меню **Вставка** или щелчком по кнопке вызова мастера функций на панели управления – f_x . В открывшемся диалоговом окне **Мастер функций – шаг 1 из 2** выбираем категорию **Статистические** и затем перемещаем строку выделения по списку функций до необходимой статистической функции. В нижней части окна будет представлена краткая информация о выделяемых функциях. В списке функций вы найдете все вычисленные для нашего примера функции:

СРЗНАЧ — рассчитывает среднюю арифметическую простую для заданных аргументов;

МОДА — отображает наиболее часто встречающиеся значения аргументов;

МЕДИАНА — рассчитывает медиану заданных аргументов;

СТАНДОТКЛОН — оценивает генеральное среднее квадратическое отклонение по выборке (если совокупность является генеральной, следует воспользоваться функцией **СТАНДОТКЛОНП**);

ДИСП — оценивает генеральную дисперсию по выборке (если совокупность является генеральной, следует воспользоваться функцией **ДИСПР**) и др.

Предположим, мы хотим рассчитать среднее квадратическое отклонение. После выбора функции щелкаем по кнопке **ОК** для перехода в следующее диалоговое окно **Аргументы функции**, в котором необходимо задать аргументы. Задать аргументы можно разными способами. Например, выделить блок ячеек с исходными данными, в нашем случае **B2:B11**. Для этого устанавливаем указатель мыши в ячейку **B2** и, удерживая левую кнопку мыши, протаскиваем указатель до ячейки **B11**. Убедитесь, что в окне **Аргументы функций** появилась запись **B2:B11**, а в строке формул отражена функция, по которой будет производиться вычисление — **=СТАНДОТКЛОНП(B2:B11)**. Щелкаем **ОК**. В ячейке **B13** появляется результат вычислений, равный 6,557751.

Возможен и другой порядок расчета. После формирования на рабочем листе Excel таблицы с исходными данными в меню **Сервис** выбираем команду **Анализ данных**. Появляется окно с одноименным названием, в котором представлен список реализованных в Excel методов статистической обработки данных. Выделяем указателем мыши режим **Описательная статистика** и щелкаем по кнопке **ОК**. В диалоговом окне режима **Описательная статистика** задаем параметры:

1. *Выходной интервал* — **B2:B11**.
2. *Группирование* — по столбцу.
3. *Выходной интервал* — указываем любую свободную ячейку. Она будет верхней левой ячейкой выходной таблицы с результатами вычислений.

4. *Итоговая статистика* — устанавливаем в активное состояние.

Щелкаем **ОК**.

Показатели, рассчитанные в данном режиме, представлены в табл. 6.4.

Таблица 6.4

Итоговая статистика

Столбец1	
Среднее	14,13
Стандартная ошибка	2,185917
Медиана	13,4
Мода	#Н/Д
Стандартное отклонение	6,912477
Дисперсия выборки	47,78233
Эксцесс	-0,33934
Асимметричность	0,667082
Интервал	21,3
Минимум	5,7
Максимум	27
Сумма	141,3
Счет	10

Обратите внимание, что показатели среднее квадратическое отклонение (стандартное отклонение) и дисперсия несколько отличаются от рассчитанных вручную. Это связано с тем, что в режиме **Описательная статистика** все функции реализуются по формулам для выборочных совокупностей.

Пример 6.2. Сотрудник отдела изучения потребительского спроса салона одежды и обуви в течение двух часов регистрировал число покупок, сделанных покупателями, посетившими салон в день сезонной распродажи. В табл. 6.5 представлены итоги регистрации.

Таблица 6.5

Итоги регистрации потребительского спроса

Номер покупателя	Число покупок	Номер покупателя	Число покупок	Номер покупателя	Число покупок
1	2	9	2	17	4
2	1	10	5	18	1
3	3	11	3	19	1
4	4	12	3	20	2
5	1	13	2	21	3
6	1	14	2	22	2
7	2	15	1	23	5
8	2	16	2	24	4

Число значений, которое принимает исследуемый признак – «число покупок», – невелико. Все покупатели сделали от одной до пяти покупок. В случае когда признак принимает ограниченное число значений и они выражены целочисленными значениями, для анализа вариации целесообразно представлять данные в виде *дискретных вариационных рядов* (табл. 6.6).

Таблица 6.6

Распределение покупателей по числу покупок

Число покупок	Количество покупателей
1	6
2	9
3	4
4	3
5	2
Итого	24

Расчет *средней величины признака* в вариационном ряду осуществляется по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_j f_j}{\sum f_j},$$

где x_j – варианты признака в j -й группе;
 f_j – частоты или частоты.

Для нашего примера:

$$\bar{x} = \frac{1 \cdot 6 + 2 \cdot 9 + 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 2}{24} = \frac{58}{24} = 2,4.$$

Каждый покупатель, посетивший салон, сделал в среднем 2,4 покупки.

Мода в дискретном ряду определяется без вычислений – как значение признака с наибольшей частотой (частотой). Из табл. 6.6 видно, что наиболее часто встречаются покупатели, которые делают две покупки. Следовательно, модальное значение равно двум покупкам.

Медианой в дискретном ряду считают значение признака в той группе, в которой накопленная частота (частость) превышает

половину численности совокупности. В рассматриваемом примере медианное значение будет равно двум покупкам (накопленная частота во второй группе — $6 + 9 = 15$ — превышает половину численности совокупности — 12).

Дальнейший анализ вариационного ряда включает расчет показателей размера и интенсивности вариации, асимметрии и эксцесса (см. пример 6.3).

Пример 6.3. В автопарке в течение дня эксплуатируются 367 автомобилей, суточный пробег которых колеблется от 105,1 до 239,5 км. Как видим, число единиц совокупности и число значений признака достаточно велико. В таких случаях анализ вариации удобнее проводить по сгруппированным данным, оформленным в виде интервальных вариационных рядов. Если вариация не очень сильная, т.е. совокупность достаточно однородная, целесообразно использовать *интервальный вариационный ряд с равными интервалами*.

Число групп устанавливают, придерживаясь формулы Стерджесса:

$$k = 1 + 3,32 \lg n = 1 + 1,44 \ln n$$

где k — число групп;

n — число единиц совокупности.

Для нашего примера

$$k = 1 + 1,44 \ln 367 = 8,5.$$

Поскольку число групп должно выражаться целым числом, будем формировать девять групп.

Величину интервалов рассчитывают по формуле

$$i = \frac{x_{\max} - x_{\min}}{k},$$

где x_{\max} и x_{\min} — максимальное и минимальное значения признака в совокупности.

В нашем случае

$$i = \frac{x_{\max} - x_{\min}}{k} = \frac{239,5 - 105,1}{9} = 14,9 \text{ км.}$$

Для построения вариационного ряда и его анализа по возможности лучше использовать округленные значения величины интервалов и его границ. Для нашего примера можно строить интервалы, равные 15 км.

В табл. 6.7 приведен готовый вариационный ряд, отражающий распределение автомобилей автотранспортного предприятия по величине суточного пробега.

Таблица 6.7

Распределение автомобилей по величине суточного пробега

Суточный пробег, км	105–120	120–135	135–150	150–165	165–180	180–195	195–210	210–225	225–240
Число автомобилей	25	43	112	77	41	20	25	15	9

Расчет *средней величины признака* в интервальном вариационном ряду осуществляется по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x'_j f_j}{\sum f_j},$$

где x'_j — середины интервалов в j -й группе;

f_j — частоты или частости j -й группы.

Для удобства вычислений построим расчетную таблицу, где будем отражать промежуточные вычисления.

Таблица 6.8

Расчетная таблица

Суточный пробег, км	Число автомобилей f_j	Середина интервала x'_j	$x'_j f_j$	Накопленные частоты S
105–120	25	112,5	2812,5	25
120–135	43	127,5	5482,5	68
135–150	112	142,5	15960,5	180
150–165	77	157,5	12127,5	257
165–180	41	172,5	7072,5	298
180–195	20	187,5	3750	318
195–210	25	202,5	5062,5	343
210–225	15	212,5	3187,5	358
225–240	9	232,5	2092,5	367
Итого	367	–	57547,5	–

Рассчитаем средний суточный пробег одного автомобиля:

$$x = \frac{\sum x'_j f_j}{\sum f_j} = \frac{57547,5}{367} = 156,8 \text{ км.}$$

В среднем пробег одного автомобиля в сутки составляет 156,8 км.

Вычисление моды в интервальном вариационном ряду весьма условно и осуществляется по формуле

$$M_o = x_{M_o} + i_{M_o} \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})},$$

где x_{M_o} — нижняя граница модального интервала;
 i_{M_o} — величина модального интервала;
 $f_{M_o}, f_{M_o-1}, f_{M_o+1}$ — частоты (частости) соответственно модального, до-модального и послемодального интервалов.

Модальный интервал — это интервал, имеющий наибольшую частоту (частость). В нашем примере это интервал от 135 до 150 км. В этой группе находятся 112 автомобилей, что больше, чем в любой другой.

Рассчитаем модальное значение суточного пробега:

$$M_o = 135 + 15 \cdot \frac{112 - 43}{(112 - 43) + (112 - 77)} = 144,9 \text{ км.}$$

Таким образом, наиболее часто встречающаяся величина суточного пробега (точечная мода) составляет 144,9 км.

Расчет модального значения для вариационного ряда с *неравными* интервалами осуществляется по формуле, аналогичной приведенной выше, только вместо показателей частот (частостей) используются показатели *плотности распределения*, которые обеспечивают сопоставимость неравных интервалов (см. табл. 6.13).

Медиана в интервальном ряду определяется по формуле

$$M_e = x_{M_e} + i_{M_e} \frac{\frac{1}{2} \sum f - S_{M_e-1}}{f_{M_e}},$$

где x_{Me} — нижняя граница медианного интервала;
 i_{Me} — величина медианного интервала;
 $\sum f$ — общая сумма частот (частостей) вариационного ряда;
 f_{Me} — частота (частость) медианного интервала;
 S_{Me-1} — сумма накопленных частот (частостей) в домедианном интервале.

Медианный интервал — это интервал, в котором находится единица совокупности с порядковым номером, равным $\sum f_j : 2$ (при четном числе единиц совокупности) или $(\sum f_j + 1) : 2$ (при нечетном)*. В нашем примере это будет интервал, в котором находится автомобиль, порядковый номер которого равен $(367 + 1) : 2 = 184$. Для определения медианного интервала рассчитаем накопленные частоты (гр. 5 табл. 6.8). В первом интервале находятся 25 автомобилей, в двух первых — 68, в трех — 180. Очевидно, что автомобиль с порядковым номером 184 попадает в четвертый интервал — от 150 до 165 км. Этот интервал и является медианным. Тогда

$$Me = 150 + 15 \cdot \frac{\frac{1}{2} \cdot 367 - 180}{77} = 150,7 \text{ км.}$$

Следовательно, половина автомобилей автопарка имеет точный пробег меньше чем 150,7 км, а половина — больше чем 150,7 км.

Среднее линейное и среднее квадратическое отклонения в вариационном ряду рассчитываются по формулам для сгруппированных данных:

$$a = \frac{\sum_{j=1}^k |x_j - \bar{x}| f_j}{\sum_{j=1}^k f_j};$$

* Если в вариационном ряду численность единиц каждой группы задана в виде относительных величин — частостей, медианным является первый интервал, в котором сумма накопленных частот превысит 50%.

$$\sigma = \sqrt{\frac{\sum_{j=1}^k (x_j - \bar{x})^2 f_j}{\sum_{j=1}^k f_j}},$$

где x_j – значение признака в j -й группе (для интервальных вариационных рядов – середина j -го интервала – x'_j);
 \bar{x} – средняя величина признака в совокупности;
 f_j – частота (частость) j -й группы;
 k – число групп.

Квадрат среднего квадратического отклонения дает величину *дисперсии* (σ^2). Для сгруппированных данных:

$$\sigma^2 = \frac{\sum_j^k (x_j - \bar{x})^2 f_j}{\sum_j^k f_j}.$$

Для определения среднего линейного, среднего квадратического отклонения и дисперсии составим расчетную таблицу (табл. 6.9).

Среднее линейное отклонение, дисперсия и среднее квадратическое отклонение соответственно составят:

$$a = \frac{\sum_{j=1}^k |x'_j - \bar{x}| f_j}{\sum_{j=1}^k f_j} = \frac{7939,9}{367} = 21,6 \text{ км};$$

$$\sigma^2 = \frac{\sum_{j=1}^k |x'_j - \bar{x}|^2 f_j}{\sum_{j=1}^k f_j} = \frac{288197,8}{367} = 785,3;$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{785,3} = 28,0 \text{ км}.$$

Таблица 6.9

Расчетная таблица

Суточный пробег, км	Число автомо- билей f_j	Середина интервала x'_j	$\left x'_j - \bar{x} \right $ $\bar{x}=156,8$	$\left x'_j - \bar{x} \right f_j$	$\left(x'_j - \bar{x} \right)^2 f_j$	$\left(x'_j - \bar{x} \right)^3 f_j$
105–120	25	112,5	44,3	1107,5	49062,25	–2173458
120–135	43	127,5	29,3	1259,9	36915,07	–1081612
135–150	112	142,5	14,3	1601,6	22902,88	–327511
150–165	77	157,5	0,7	53,9	37,73	26,411
165–180	41	172,5	15,7	643,7	10106,09	158665,6
180–195	20	187,5	30,7	614	18849,8	578688,9
195–210	25	202,5	45,7	1142,5	52212,25	2386100
210–225	15	212,5	55,7	835,5	46537,35	2592130
225–240	9	232,5	75,7	681,3	51574,41	3904183
Итого	367	—	—	7939,9	288197,8	6037214

Среднее линейное и среднее квадратическое отклонения показывают, на сколько километров в среднем величина суточного пробега отдельных автомобилей отличалась от среднего пробега по автопарку. По формуле среднего линейного отклонения это отличие составило $\pm 21,6$ км, по формуле среднего квадратического отклонения ± 28 км.

Для оценки *интенсивности вариации* рассчитывают *относительные показатели вариации*. Наиболее часто применяют *коэффициент вариации* (v), который представляет собой относительное квадратическое отклонение:

$$v = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Чем больше величина коэффициента вариации, тем больше разброс значений признака вокруг средней и выше неоднородность совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 30%. Средняя, рассчитанная для однородной совокупности, является ее типической характеристикой.

Для нашего примера коэффициент вариации составил:

$$v = \frac{28,0}{156,8} \cdot 100\% = 17,8\%,$$

что свидетельствует об однородности совокупности автомобилей автопарка по величине суточного пробега. Средний пробег 156,8 км можно считать типической характеристикой суточного пробега в автопарке.

Для характеристики формы распределения дается оценка асимметрии и эксцесса вариационного ряда.

Симметричным является распределение, в котором частоты двух вариантов, равностоящих в обе стороны от центра распределения, равны между собой. Если большая часть совокупности расположена левее центра, имеет место *левосторонняя асимметрия*, если правее — *правосторонняя*.

О симметричности эмпирического ряда распределения можно судить по соотношению характеристик центра распределения. В симметричном распределении средняя величина, медиана и мода равны между собой: $\bar{x} = Me = Mo$. Если $\bar{x} > Me > Mo$,

то имеет место правосторонняя асимметрия, соотношение $\bar{X} < Me < Mo$ характерно для *левосторонней асимметрии*.

Нашему примеру соответствует соотношение $\bar{X} > Me > Mo$ (156,8 км > 150,7 км > 144,9 км), характерное для правосторонней асимметрии. Наличие правосторонней асимметрии свидетельствует о том, что большая часть автомобилей имеет суточный пробег выше, чем его модальное значение 144,9 км.

Оценить асимметричность распределения можно также с помощью коэффициентов асимметрии.

Наиболее часто применяют коэффициент асимметрии, предложенный английским статистиком К. Пирсоном:

$$As_{\Pi} = \frac{\bar{x} - Mo}{\sigma}.$$

Если $As_{\Pi} < 0$, то это левосторонняя (отрицательная) асимметрия, при правосторонней (положительной) асимметрии $As_{\Pi} > 0$. Для нашего примера коэффициент асимметрии Пирсона составит:

$$As_{\Pi} = \frac{\bar{x} - Mo}{\sigma} = \frac{156,8 - 144,9}{28,0} = 0,425.$$

Моментный коэффициент асимметрии определяется по формуле

$$As = \frac{\mu_3}{\sigma^3},$$

где μ_3 — центральный момент третьего порядка.

Для сгруппированных данных он рассчитывается по формуле

$$\mu_3 = \frac{\sum_{j=1}^k (x_j - \bar{x})^3 f_j}{\sum_{j=1}^k f_j}.$$

Моментный коэффициент асимметрии для нашего примера будет равен (промежуточные вычисления см. табл. 6.9):

$$A_s = \frac{\mu_3}{\sigma^3} = \frac{6037214 : 367}{28,0^3} = 0,749.$$

Положительная величина коэффициента подтверждает вывод о правосторонней асимметрии.

Степень существенности асимметрии можно оценить с помощью средней квадратической ошибки коэффициента асимметрии, которая зависит от объема изучаемой совокупности и рассчитывается по формуле

$$\sigma_{A_s} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}},$$

где n — число единиц совокупности.

Если отношение $|A_s| : \sigma_{A_s} > 2$, асимметрия считается существенной, если $|A_s| : \sigma_{A_s} < 2$, то асимметрия признается несущественной, вызванной влиянием случайных обстоятельств. Для нашего примера $\sigma_{A_s} = 0,127$. Тогда $|A_s| : \sigma_{A_s} = 0,749 : 0,127 = 5,91$, следовательно, имеет место существенная асимметрия.

Под *эксцессом* понимают островершинность или плосковершинность распределения по сравнению с нормальным распределением при той же силе вариации. Эксцесс оценивается с помощью следующего показателя:

$$Ex = \frac{\mu_4}{\sigma^4} - 3,$$

где μ_4 — центральный момент четвертого порядка, для сгруппированных данных, рассчитываемый по формуле

$$\mu_4 = \frac{\sum_{j=1}^k (x_j - \bar{x})^4 f_j}{\sum_{j=1}^k f_j}.$$

Распределения более островершинные, чем нормальные, обладают положительным эксцессом ($Ex > 0$), более плосковершинные — отрицательным ($Ex < 0$). Положительный эксцесс свидетельствует о том, что в совокупности есть слабо варьирующее по данному признаку «ядро», а в плосковершинных распределениях такого «ядра» нет, и единицы рассеяны по всем значениям признака более равномерно.

Чтобы оценить существенность эксцесса распределения, рассчитывают среднюю квадратическую ошибку эксцесса:

$$\sigma_{Ex} = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}.$$

Если отношение $|Ex|/\sigma_{Ex} > 2$, то отклонение от нормального можно считать существенным.

Поскольку эксцесс определяется только для симметричных и умеренно асимметричных распределений, для нашего примера он не рассчитывается.

Графическое изображение вариационных рядов облегчает их анализ и наглядно отображает форму распределения. Для графического изображения вариационного ряда строят гистограмму, полигон и кумуляту распределения.

Гистограмма — столбиковая диаграмма (рис. 6.1), для построения которой на оси абсцисс откладывают отрезки, равные величине интервалов вариационного ряда. На отрезках строят прямоугольники, высота которых по оси ординат соответствует частотам (или частостям) (подробнее о графиках см. тему 4).

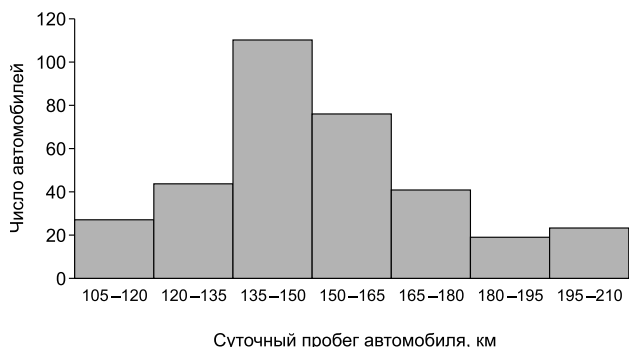


Рис. 6.1. Распределение автомобилей по величине суточного пробега

Пример 6.4. В табл. П 1.3 приведены данные о деятельности банков Северо-Западного региона в первом полугодии 2006 г. Требуется построить вариационный ряд, отражающий степень концентрации капитала в банковской сфере и позволяющий оценить дифференциацию банков по величине собственных капиталов. (Для упрощения примера ограничимся рассмотрением 50 крупнейших банков.)

Если следовать алгоритму, описанному в примере 6.3, и строить вариационный ряд с равными интервалами, получим результат, отраженный в табл. 6.10.

Таблица 6.10

**Распределение банков Северо-Западного региона по величине
собственного капитала на 1 июля 2006 г.**

Собственный капитал, млрд руб.	До 2	2–4	4–6	6–8	8–10	10 и более
Число банков	44	3	2	—	—	1

Как видим, большая часть банков попала в первую группу, а часть групп вообще оказалась незаполненной. Группировка с равными интервалами не отражает разнообразие значений признака в данной совокупности и, следовательно, не может решить поставленные выше задачи.

В тех случаях, когда вариация признака очень сильна (в нашем примере собственный капитал банков варьирует от 59 млн руб. (Банк АЖКБ, г. Архангельск) до 12231 млн руб. (Промстройбанк, г. Санкт-Петербург)), лучше использовать *равночастотные вариационные ряды*. Число групп в равночастотных рядах обычно принимают равным 4 или 10. Таким образом, в каждой группе содержится 25 или 10% единиц совокупности. Такие ряды называют соответственно *квартильными* и *децильными* распределениями.

В табл. 6.11 представлен вариационный ряд, отражающий распределение 50 банков Северо-Западного региона по величине собственного капитала по децильным группам.

Группировка по децильным группам наглядно показывает существенную концентрацию капитала в девятой и десятой группах банков. На 10 самых крупных банков приходится 74,7% (17,4 + + 57,3) совокупного капитала банков, а на долю остальных — только 25,3%. Пять банков, попавших в первую группу, представляют категорию 10% самых маленьких по величине собственного капитала банков. На их долю в сумме приходится менее 1% совокупной суммы капиталов банков.

Таблица 6.11

**Распределение банков Северо-Западного региона
по величине собственного капитала на 1 июля 2006 г.**

Собственный капитал, млн руб.	Число банков	Совокупный капитал в группе	
		млн руб.	%
59–75	5	331	0,7
83–115	5	477	1,0
121–161	5	712	1,5
161–204	5	905	1,8
205–276	5	1209	2,5
303–328	5	1546	3,2
395–624	5	2462	5,0
758–1193	5	4762	9,7
1246–2633	5	8521	17,4
3223–12231	5	28036	57,3
	50	48961	100

Верхняя граница первого и нижняя граница последнего интервалов равночастотного децильного распределения соответствуют фактическим значениям крайних *децилей* (в квартильном распределении – *квартилям*) распределения.

Децили – структурные характеристики вариационного ряда, делящие ряд на 10 равных частей. Первая дециль (D_1) характеризует максимальное значение признака у 10% единиц совокупности с самыми маленькими значениями признака, девятая дециль (D_9) – минимальное значение признака у 10% самых крупных по значению признака единиц совокупности. На соотношении децилей строится *коэффициент децильной дифференциации*:

$$K_D = \frac{D_9}{D_1}.$$

Для нашего примера коэффициент децильной дифференциации составит:

$$K_D = \frac{D_9}{D_1} = \frac{3223}{75} = 43.$$

Это означает, что минимальный размер собственного капитала банков, входящих в 10% самых крупных банков, превышает максимальный размер собственного капитала в группе 10%

самых маленьких банков в 43 раза, что говорит об очень существенной дифференциации банков Северо-Западного региона по величине собственного капитала.

Расчет децилей в интервальном ряду требует специальных вычислений.

Пример 6.5. В табл. 6.12 приведено распределение предприятий отрасли по балансовой прибыли. Необходимо оценить дифференциацию предприятий отрасли по величине прибыли с помощью коэффициента децильной дифференциации.

Таблица 6.12

Распределение предприятий по величине балансовой прибыли

Балансовая прибыль, млн руб.	Удельный вес предприятий, %	Накопленные проценты
До 10	9	9
10–20	10	19
20–30	7	26
30–40	15	41
40–50	23	64
50–60	18	82
60–70	11	93
70 и более	7	100
Итого	100	—

Расчет децилей в интервальном вариационном ряду носит условный характер и осуществляется по формулам

$$D_1 = x_{0_{D_1}} + i_{D_1} \frac{\frac{1}{10} \sum f - S_{D_{1-1}}}{f_{D_1}};$$

$$D_9 = x_{0_{D_9}} + i_{D_9} \frac{\frac{9}{10} \sum f - S_{D_{9-1}}}{f_{D_9}},$$

где $x_{0_{D_1}}, x_{0_{D_9}}$ — нижние границы интервалов, в которых находятся первая и девятая децили;

i_{D_1}, i_{D_9} — величины интервалов, где находятся первая и девятая децили;

$\sum f$ — общая сумма частот (частотностей);

S_{D1-1}, S_{D9-1} — суммы частот (частостей), накопленных в интервалах, предшествующих интервалам, в которых находятся первая и девятая децили;

f_{D_1}, f_{D_9} — частоты (частости) интервалов, содержащих первую и девятую децили.

Поскольку децили отсекают десятые части совокупности, по накопленным частостям определяем интервалы, куда попадают порядковые номера децилей: для первой децили — интервал, где находится вариант, отсекающий 10% единиц совокупности с наименьшими значениями признака; для девятой децили — интервал, содержащий вариант, отсекающий 90% единиц с наименьшими значениями, или, что то же самое, 10% единиц совокупности с наибольшими значениями признака.

В нашем примере в первой группе находится только 9% предприятий, в двух первых — 19%. Следовательно, вариант, отсекающий 10% единиц совокупности, находится во второй группе в интервале от 10 до 20 млн руб. Девятая дециль попадает в интервал от 60 до 70 млн руб. (в этом интервале находится вариант, отсекающий 90% предприятий с наименьшими доходами). Найдем величину соответствующих децилей:

$$D_1 = 10 + 10 \cdot \frac{\frac{1}{10} \cdot 100 - 9}{10} = 11 \text{ млн руб.}$$

Следовательно, максимальная величина балансовой прибыли у 10% наименее прибыльных предприятий отрасли составляет 11 млн руб.

$$D_9 = 60 + 10 \cdot \frac{\frac{9}{10} \cdot 100 - 82}{11} = 67,3 \text{ млн руб.}$$

Минимальная величина прибыли у 10% самых прибыльных предприятий — 67,3 млн руб.

Рассчитаем коэффициент децильной дифференциации:

$$K_D = \frac{D_9}{D_1} = \frac{67,3}{11} = 6,1.$$

Таким образом, минимальный размер балансовой прибыли у 10% наиболее прибыльных предприятий отрасли превышает максимальный размер прибыли 10% наименее прибыльных предприятий более чем в 6 раз.

Вычисление обобщающих показателей в вариационных рядах с неравными интервалами имеет особенности.

Пример 6.6. Представлены данные о распределении предприятий региона по удельному весу закупок по заказам, переданным по глобальным сетям (гр. 1 и 2 табл. 6.13). Определите модальное значение данного признака для предприятий региона.

Таблица 6.13

Распределение организаций региона по удельному весу закупок по заказам, переданным по глобальным сетям

Удельный вес закупок по заказам, переданным по глобальным сетям, %	Число предприятий	Величина интервала	Абсолютная плотность распределения
1	2	3	4
До 5	92	5	$92 : 5 = 18,4$
5–10	130	5	$130 : 5 = 26,0$
10–20	145	10	$145 : 10 = 14,5$
20–40	110	20	$110 : 20 = 5,5$
40–60	31	20	$31 : 20 = 1,55$
60–100	58	40	$58 : 40 = 1,45$
Итого	556	—	

Группировка с неравными интервалами в данном случае использована потому, что большинство предприятий имеет небольшой процент закупок по заказам через глобальные сети. Однако разная величина интервалов делает несопоставимой наполненность отдельных групп. В этом случае переходят к показателям плотности распределения. Показатели плотности распределения находятся либо как отношение частоты к величине интервала (*абсолютная плотность распределения*), либо как отношение частоты к величине интервала (*относительная плотность распределения*).

Для нашего примера рассчитаем абсолютные плотности распределения (гр. 4 табл. 6.13). Как следует из расчетов, наибольшую плотность распределения имеет второй интервал.

Вычисление моды в вариационном ряду с неравными интервалами осуществляется по формуле

$$M_o = x_{M_o} + i_{M_o} \frac{p_{M_o} - p_{M_o-1}}{(p_{M_o} - p_{M_o-1}) + (p_{M_o} - p_{M_o+1})},$$

где x_{M_o} — нижняя граница модального интервала;
 i_{M_o} — величина модального интервала;
 $p_{M_o}, p_{M_o-1}, p_{M_o+1}$ — плотности (абсолютные или относительные) модального, домодального и послемодального интервалов.

Модальный интервал в вариационном ряду с неравными интервалами — это интервал, имеющий наибольшую плотность распределения.

Значение моды для нашего примера равно:

$$M_o = 5 + 5 \cdot \frac{26,0 - 18,4}{(26,0 - 18,4) + (26,0 - 14,5)} = 7\%.$$

Следовательно, для предприятий регионов модальный (наиболее часто встречающийся) удельный вес закупок по заказам, переданным по глобальным сетям, составляет 7%.

В ряде случаев, например для определения объемов выборки, построения доверительных интервалов при проведении выборочных обследований, необходимо знать показатели вариации альтернативных признаков.

Пример 6.7. В табл. 6.14 приведены данные по результатам проверки органами Роспотребнадзора качества отечественных и импортных товаров, поступающих на потребительский рынок.

Таблица 6.14

**Качество отечественных и импортных товаров,
поступивших на потребительский рынок**

Товар	Забраковано и снижено в сортности, в процентах от общего количества проинспектированных товаров			
	2000 г.		2004 г.	
	отечественных товаров	импортных товаров	отечественных товаров	импортных товаров
Радиотовары	23,1	36,2	39,5	31,8
Стиральные машины	25,9	33,5	14,2	27,8
Холодильники	20,7	19,7	19,1	17,6

Источник: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 546.

Определим дисперсию и среднее квадратическое отклонение доли забракованных и сниженных в сортности радиотоваров отечественного и импортного производства в 2004 г.

Дисперсия альтернативного признака рассчитывается по формуле

$$\sigma^2 = p \cdot q,$$

где p — доля единиц, обладающих данным признаком (в нашем примере — доля забракованных и сниженных в сортности товаров от общего количества проинспектированных товаров);

q — доля единиц, не обладающих данным признаком.

Поскольку $p + q = 1$, формулу дисперсии можно записать и в следующем виде:

$$\sigma^2 = p \cdot (1 - p).$$

Максимальное значение дисперсии альтернативного признака равно 0,25; $0 \leq \sigma^2 \leq 0,25$.

Дисперсия доли забракованных и сниженных в сортности радиотоваров отечественного производства в 2004 г. составила:

$$\sigma^2 = p \cdot (1 - p) = 0,395 \cdot (1 - 0,395) = 0,395 \cdot 0,605 = 0,239.$$

Среднее квадратическое отклонение

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,239} = 0,489.$$

Дисперсия доли забракованных и сниженных в сортности радиотоваров импортного производства в 2004 г. составила:

$$\sigma^2 = p \cdot (1 - p) = 0,318 \cdot (1 - 0,318) = 0,318 \cdot 0,682 = 0,217.$$

Среднее квадратическое отклонение

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,217} = 0,466.$$

Таким образом, качество проинспектированных Роспотребнадзором и отечественных, и импортных радиотоваров отличалось достаточно высокой вариацией.

6.2. Задания по теме

1. В табл. 6.15 представлено распределение безработных (мужчин) по возрасту в Санкт-Петербурге в 2004 г.

Таблица 6.15

Возраст, лет	До 20	20–30	30–40	40–50	50–60	60 и более
Число безработных, %	3,1	25,6	27,2	19,9	16,9	7,3

Дайте характеристику данному ряду распределения, рассчитав все показатели вариации. Постройте гистограмму распределения.

2. Рассчитайте показатели размера и интенсивности вариации признака «число покупок» по данным табл. 6.6. Постройте полигон распределения. На основании графика сделайте вывод о форме распределения.

3. В течение летнего периода в местах рекреации населения систематически контролируется санитарно-гигиеническое состояние 80 водных объектов региона. Результаты замеров по состоянию на начало июля за два последних года представлены в табл. 6.16.

Таблица 6.16

Удельный вес проб воды, не отвечающих санитарно-гигиеническим нормам, %	Число водных объектов	
	01.07.2005 г.	01.07.2006 г.
До 5	7	12
5–10	17	20
10–15	19	19
15–20	22	14
20–25	9	10
25–30	6	5
Итого	80	80

1. Оцените изменение состояния водных объектов за рассматриваемый период. Рассчитайте показатели центра распределения: средний, модальный, медианный удельный вес проб воды, не отвечающих санитарно-гигиеническим нормам.

2. Определите показатели размера и интенсивности вариации. Как изменилась степень однородности водных объектов по санитарно-гигиеническому состоянию?

3. Постройте гистограммы распределения водных объектов по удельному весу проб воды, не отвечающих санитарно-гигиеническим нормам для каждого года.

4. Имеются данные о распределении фермерских хозяйств области за два года по уровню товарности сельскохозяйственно-го производства (табл. 6.17).

Таблица 6.17

Уровень товарности, %	Число фермерских хозяйств	
	1995 г.	2005 г.
До 30	24	17
30–50	35	24
50–70	39	49
70–90	27	95
90 и более	6	23
Итого	131	208

1. Заполните недостающие показатели в табл. 6.18 и дайте характеристику изменению уровня товарности фермерских хозяйств в области за рассматриваемый период.

Таблица 6.18

Год	$\bar{X}_{\text{ср}}$	Me	Mo	d	σ	ν	$As_{\text{П}}$
1995	53,3			19,3	22,6		
2005		72,9	77,8		21,6	32	

2. Изобразите на одном графике кумуляты распределения фермерских хозяйств области по уровню товарности для 1995 и 2005 гг.; определите графически удельный вес хозяйств с уровнем товарности более 80% в 1995 и 2005 гг.

П р и м е ч а н и е. Уровень товарности определяется отношением стоимости реализованной продукции к стоимости произведенной продукции.

5. Распределение населения в двух регионах по величине среднедушевого дохода характеризуется данными, представленными в табл. 6.19.

Таблица 6.19

Среднедушевой доход в среднем за месяц, тыс. руб. на 1 чел.	Число жителей, % к итогу	
	Регион А	Регион Б
До 2,0	5,8	5,6
2,0–4,0	9,7	5,3
4,0–6,0	13,0	16,1
6,0–8,0	19,1	35,2
8,0–10,0	22,2	15,5
10,0–12,0	19,2	14,7
12,0–14,0	8,5	4,2
Свыше 14,0	2,5	3,4

1. Рассчитайте средний, модальный и медианный среднедушевой доход.

2. Изобразите вариационный ряд графически. Найдите графически моду и медиану.

3. Определите размер и интенсивность вариации доходов населения. Оцените степень однородности распределения населения по величине среднедушевого дохода.

4. Измерьте дифференциацию населения по величине среднедушевых доходов на основе децильного коэффициента дифференциации.

5. Рассчитайте коэффициенты асимметрии и эксцесса распределения.

По результатам вычислений напишите аналитическую записку. Задание выполните по вариантам:

1-й вариант — регион А.

2-й вариант — регион Б.

6. Используя исходные данные задания 5, постройте гistogramмы и кумуляты распределения населения по величине среднедушевого дохода в регионе А и в регионе Б. Какие выводы об особенностях распределения населения по величине среднедушевого дохода можно сделать на основании построенных графиков? Определите графически значение модального и медианного среднедушевого дохода в каждом регионе. Напишите аналитическую записку.

7. В результате инвентаризации основных фондов на предприятии получена информация о возрастной структуре станочного оборудования (табл. 6.20).

Таблица 6.20

Годы ввода станков в эксплуатацию	Количество станков, % к итогу
До 1980	28,0
1980–1984	9,5
1985–1989	19,1
1990–1994	10,3
1995–1999	20,1
2000 и позднее	13,0

1. Рассчитайте:
 - а) средний возраст оборудования на предприятии;
 - б) модальный возраст оборудования на предприятии;
 - в) медианный возраст оборудования на предприятии.
2. Оцените однородность распределения оборудования по возрасту.
3. Постройте гистограмму и кумуляту распределения станочного оборудования по возрасту. Определите графически долю оборудования старше 15 лет.

8. В табл. 6.21 приведены данные о распределении по возрасту занятого и безработного населения РФ в 2004 г.

Таблица 6.21

Возраст, лет	Число лиц, % к итогу	
	занятые	безработные
До 20	1,9	10,4
20–29	22,4	30,0
30–39	24,4	21,5
40–49	30,0	23,7
50–59	17,4	11,9
60 и старше	3,9	2,5

Источник: Российский статистический ежегодник. 2005: Стат. сб. – М.: Росстат, 2006. – С. 153, 154.

Рассчитайте по каждой категории экономически активного населения систему показателей для анализа вариационных рядов (среднюю величину, моду, медиану, показатели размера и интенсивности вариации, коэффициент асимметрии). Напишите аналитическую записку о различиях возрастных характеристик занятых и безработных.

9. В табл. 6.22 приведены данные о распределении регионов РФ (сопоставимый круг регионов) по уровню официальной безработицы.

Таблица 6.22

1992		2004	
Уровень безработицы, в процентах к экономически активному населению	Число регионов	Уровень безработицы, в процентах к экономически активному населению	Число регионов
До 3	2	До 3	2
3–4	9	3–5	6
4–5	28	5–7	16
5–6	23	7–9	19
6–7	9	9–11	21
7–8	5	11–13	8
8 и более	4	13 и более	8
Итого	80	Итого	80

1. Проведите анализ рядов распределения. Рассчитайте для каждого года:

а) показатели центра распределения (средний, модальный и медианный уровни безработицы в регионах РФ);

б) среднее линейное и среднее квадратическое отклонения;

в) коэффициент вариации. Сделайте вывод об однородности регионов по уровню безработицы;

г) моментный и структурный коэффициенты асимметрии;

д) эксцесс распределения.

2. Оцените степень существенности асимметрии и эксцесса. Проинтерпретируйте полученные результаты. Проведите сравнительный анализ за два года. Постройте графики распределения.

3. Напишите аналитическую записку об изменении уровня безработицы в Российской Федерации.

10. Пассажирам трех коммерческих маршрутов предложено оценить удобство графика движения по пятибалльной шкале: 1 — «очень удобно», 5 — «крайне неудобно». В результате обследования получены результаты, представленные в табл. 6.23.

Таблица 6.23

Балл	Число опрошенных пассажиров		
	Маршрут № 1	Маршрут № 2	Маршрут № 3
1	—	10	9
2	16	18	20
3	30	30	31
4	24	20	10
5	30	2	—
Итого	100	80	70

Рассчитайте средний балл для каждого маршрута. Мнение пассажиров какого маршрута можно считать более однородным?

11. В течение месяца предприятие выпустило четыре партии изделия **А**. В табл. 6.24 даны результаты контроля качества изделий.

Таблица 6.24

Номер партии	1	2	3	4
Число изделий в партии	1000	1200	1500	1000
Число изделий, не отвечающих ТУ	60	72	90	30

1. Определите:

- а) процент брака в каждой партии и в среднем по всем партиям;
- б) дисперсию и среднее квадратическое отклонение доли брака в целом по четырем партиям;
- в) коэффициент вариации доли брака.

2. Можно ли считать средний процент брака типичным уровнем брака выпускаемой продукции?

12. По данным табл. 6.14 рассчитайте дисперсию и среднее квадратическое отклонение доли забракованных и сниженных в сортности бытовых электротоваров отечественного и импортного производства в 2000 и 2004 гг. Что вы можете сказать об изменении вариации этого признака?

Задание выполните по вариантам:

- 1-й вариант — стиральные машины;
- 2-й вариант — холодильники.

13. По итогам работы автотранспортного предприятия за прошлый год суточный пробег одного грузового автомобиля составил 125 км, при среднем квадратическом отклонении — 70 км,

среднее время вынужденного простоя — 1,5 ч в сутки, при среднем квадратическом отклонении — 1 ч, что свидетельствовало о неравномерном распределении нагрузки между единицами автотранспорта. Перед плановым отделом была поставлена задача пересмотреть графики и маршруты движения. Насколько успешно была решена эта задача, если после внедрения новой системы организации маршрутов распределение автомобилей по величине суточного пробега и времени простоя характеризовалось данными, представленными в табл. 6.25.

Таблица 6.25

Суточный пробег, км	До 100	100—125	125—150	150—175	175—200	200 и более	Итого
Число автомобилей	5	12	29	35	15	9	105

Продолжение

Время простоя, ч/сутки	До 0,5	0,5 — 1,0	1,0 — 1,5	1,5 — 2,0	2,0 и более	Итого
Число автомобилей	27	47	19	8	4	105

14. В табл. 6.26 представлены данные о распределении выручки двух оптовых торговых фирм за квартал.

Таблица 6.26

Выручка за один рабочий день, тыс. руб.	Число рабочих дней	
	«Омега»	«Гермес»
До 500	5	15
500—1000	14	20
1000—1500	28	23
1500—3000	20	4
3000—5000	10	15
Итого	77	77

Сравните стабильность выручки торговых фирм «Омега» и «Гермес», рассчитав для каждой фирмы среднюю выручку за день, среднее квадратическое отклонение выручки и коэффициент вариации выручки. Коммерческая деятельность какой фирмы имеет повышенный риск?

15. По данным табл. П1.3 постройте равночастотный вариационный ряд распределения 50 крупнейших банков региона по величине активов. Рассчитайте коэффициент децильной дифференциации. Сформулируйте выводы о концентрации активов в банковской сфере.

16. Из табл. П1.3 выберите банки, зарегистрированные в Санкт-Петербурге. На основе данных о размере капитала и балансовой прибыли для каждого банка рассчитайте показатель рентабельности собственного капитала. Постройте ранжированный ряд по этому признаку. Дайте характеристику вариации рентабельности капиталов банков Санкт-Петербурга. Напишите аналитическую записку.

17. Используя данные табл. П1.2, постройте ранжированный ряд 30 крупнейших (по численности населения) стран мира по признаку:

1-й вариант — магистральные телефонные линии (на 1 тыс. чел.);

2-й вариант — число абонентов сотовых сетей (на 1 тыс. чел.);

3-й вариант — число пользователей сети Интернет (на 1 тыс. чел.).

Рассчитайте основные характеристики построенного вариационного ряда.

18. Имеются данные о наиболее кассовых американских фильмах (табл. 6.27).

Таблица 6.27

№ п/п	Фильм	Валовой доход на апрель 2004 г., млн долл.
1	«Титаник»	1235
2	«Властелин колец: возвращение короля»	696
3	«Гарри Поттер и тайная комната»	651
4	«Гарри Поттер и философский камень»	604
5	«Властелин колец: две башни»	581
6	«Парк Юрского периода»	563
7	«Властелин колец: Братство кольца»	547
8	«В поисках Немо»	513
9	«День независимости»	505
10	«Звездные войны: Эпизод 1: Скрытая угроза»	491

Источник: Доклад о развитии человека. 2004. Культурная свобода в современном многообразном мире: пер с англ. — М.: Весь мир, 2004. — С. 117.

Проведите анализ данного ряда распределения. Рассчитайте средний, модальный и медианный доход от проката фильма, среднеквадратическое отклонение дохода, коэффициент вариации. Оцените однородность ряда распределения. По результатам вычислений напишите аналитическую записку.

19. Проведите анализ распределения банков Северо-Западного региона по уровню рентабельности собственного капитала в первом полугодии 2006 г. (табл. 6.28).

Таблица 6.28

Рентабельность капитала, %	Число банков	Удельный вес банков, %
0,9–4,0	15	24,6
4,1–7,2	16	26,2
7,3–10,4	13	21,3
10,5–13,6	8	13,1
13,7–16,8	4	6,6
16,9–20,0	2	3,3
20,1 и более	3	4,9
Итого	61	100,0

20. Известны данные (табл. 6.29) о распределении домашних хозяйств по площади жилищ, приходящейся в среднем на одного проживающего, в первой и пятой квинтильных группах населения по уровню дохода (по материалам выборочного обследования бюджетов домашних хозяйств в РФ в 2004 г.) (табл. 6.29).

Таблица 6.29

Площадь жилищ, приходящаяся в среднем на одного проживающего, м ²	Удельный вес домохозяйств, в процентах от общего числа домохозяйств соответствующей группы	
	первая группа населения с наименьшими доходами	пятая группа населения с наибольшими доходами
До 9,0	14,7	1,4
От 9,1 до 13,0	27,2	7,9
От 13,1 до 15,0	13,4	9,1
От 15,1 до 20,0	18,5	16,7
От 20,1 до 25,0	11,3	18,6
От 25,1 до 30,0	6,0	10,6
От 30,1 до 40,0	4,9	14,1
От 40,1 и более	4,0	21,6

Источник: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 235.

Проведите сравнительный анализ жилищных условий в группах населения с разным уровнем доходов. По результатам исследования напишите аналитическую записку.

21. Проведите анализ распределения игр чемпионата России по футболу 2006 г. по числу забитых за игру мячей (табл. 6.30).

Таблица 6.30

Забито мячей	0	1	2	3	4	5	6	7	8
Число игр	31	47	52	49	35	15	6	4	1

22. Проведите анализ распределения сельхозпредприятий региона по среднему надою молока на корову (табл. 6.31).

Таблица 6.31

Группа сельхозпредприятий по среднему надою на корову, ц/год	Число сельхозпредприятий
5–12	30
12–19	29
19–26	32
26–33	29
33–40	10
40–47	7
47–54	4
54–61	5
61–68	2
68–75	2

23. В табл. 6.32 приведена группировка крестьянских (фермерских) хозяйств РФ по размеру предоставленных им земельных участков (на конец 2004 г.). Проведите анализ распределения хозяйств по размеру земельных участков. Постройте графики распределения. Напишите аналитическую записку.

Таблица 6.32

Размер земельного участка, га	Число хозяйств	
	всего, тыс.	в процентах от общего числа хозяйств
До 3	49,9	19,1
4–5	24,2	9,3
6–10	35,0	13,4

Размер земельного участка, га	Число хозяйств	
	всего, тыс.	в процентах от общего числа хозяйств
11–20	37,2	14,2
21–50	45,4	17,3
51–70	15,2	5,8
71–100	14,1	5,4
101–200	18,0	6,9
Свыше 200	17,7	6,8
Хозяйства, которым земельный участок не предоставлялся	4,7	1,8
Итого по всем хозяйствам	261,4	100

Источник: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 446.

6.3. Вопросы для самоконтроля

1. Что такое вариация и в чем состоят ее причины?
2. От чего зависит число групп (градаций признака) при построении ряда распределения?
3. При каком характере распределения значений признака в совокупности предпочтительнее строить равноинтервальный ряд, а когда — равночастотный?
4. Какими показателями характеризуется центр распределения?
5. Какие показатели называют структурными характеристиками вариационного ряда?
6. В чем особенности вычисления моды в равноинтервальном и в равночастотном ряду распределения?
7. Как рассчитывается и что показывает децильный коэффициент дифференциации?
8. Как вычисляется среднее квадратическое отклонение от средней величины в дискретном и интервальном рядах распределения?
9. Что такое дисперсия? Как она рассчитывается для количественных и альтернативных признаков?

10. Как соотносятся между собой среднее линейное и среднее квадратическое отклонение? Объясните, почему?

11. Как следует интерпретировать разные значения коэффициента вариации?

12. Каковы основные свойства нормального распределения?

13. Какими показателями характеризуется степень отклонения формы ряда распределения от нормального?

6.4. Методические рекомендации преподавателям

Аудиторная работа. Задачи целесообразно решать в той последовательности, в какой они разобраны в примерах. Знакомя студентов с теми или иными показателями вариации, особое внимание необходимо уделять их значению и применению в практике социальных и экономических исследований.

Отдельного обсуждения требуют отношения между разными показателями вариации, их свойства, варианты вычисления разными способами. Обратите внимание на трактовку всех показателей с точки зрения однородности совокупности.

Решение задач должно сопровождаться интерпретацией вычисляемых показателей, написанием кратких выводов и пояснений, формат которых целесообразно обсудить в аудитории.

Для промежуточного контроля усвоения материала можно использовать короткие опросы в форме тестов или решения кроссворда (см. приложение 4).

Самостоятельная внеаудиторная работа. В качестве задания для самостоятельной работы можно дать комплексную задачу на построение вариационного ряда и его всесторонний анализ. В качестве исходного материала могут использоваться данные статистических ежегодников, отражающие распределение российских регионов по величине различных социально-экономических показателей, базы данных по коммерческим данным.

Итоговый контроль. В качестве итогового контроля по теме рекомендуется контрольная работа с решением задачи на измерение размера и интенсивности вариации, на характеристику формы распределения. При оценивании работ должен быть сделан акцент на наличие содержательной интерпретации всех расчетов.

6.5. Методические указания студентам

Освоение теории. Изучение вариации является одной из ключевых тем в курсе «Общая теория статистики». В теме раскрывается понятие вариации и описываются приемы ее статистического анализа. Несмотря на то что стандартные процедуры анализа вариации входят в состав электронных таблиц и специальных статистических пакетов обработки данных, чтобы лучше понимать сущность показателей вариации, обязательно сначала научиться вычислять их вручную на простых примерах.

Выучите формулы расчета основных показателей вариации — среднего квадратического отклонения, дисперсии, коэффициента вариации. Разберитесь в особенностях вычисления показателей вариации по сгруппированным и несгруппированным данным, в равноинтервальных и равночастотных вариационных рядах. Особое внимание уделяйте интерпретации рассчитываемых показателей.

В результате изучения темы необходимо знать:

- что такое вариация и в чем состоят ее причины;
- способы построения рядов распределения;
- какими показателями характеризуются центр распределения и структура распределения;
- абсолютные и относительные показатели вариации;
- показатели формы распределения.

Практические навыки. Необходимо уметь:

- строить ряды распределения;
- измерять размеры и интенсивность вариации, характеризовать форму распределения;
- грамотно интерпретировать полученные результаты;
- использовать программные средства статистической обработки данных для анализа рядов распределения.

Группировка

7.1. Основные положения

Группировка — это распределение единиц совокупности по группам, в которых различия между единицами, отнесенными к одной группе, меньше, чем между единицами, отнесенными к разным группам. Группировка обеспечивает однородность данных, их обобщение, представление в компактном обозримом виде, создает основу для последующей сводки и анализа данных.

Правила проведения группировки включают:

- определение группировочных признаков;
- определение значений, отделяющих одну группу от другой — интервалов группировки.

Виды группировок:

простая — на основе одного группировочного признака;

сложная (комбинационная) — на основе сочетаний (комбинаций) нескольких (двух-трех) группировочных признаков;

многомерная — на основе рассчитанного интегрального показателя, который называется многомерная средняя:

$$\bar{P}_i = \frac{\sum_{j=1}^k x_{ij} : \bar{x}_j}{k},$$

где x_{ij} — значение признака x_i i -й единицы совокупности;

$$\bar{x}_j — \text{среднее значение признака } x_j, \quad \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n};$$

k — число зарегистрированных признаков.

Группировки различаются по цели: типологическая, структурная, аналитическая.

Типологическая группировка служит для выделения социально-экономических типов.

Последовательность действий при проведении типологической группировки:

1) называются те типы явлений, которые могут быть выделены;

2) выбираются группировочные признаки, формирующие описания типов;

3) устанавливаются границы интервалов;

4) группировка оформляется в таблицу: на основе комбинации группировочных признаков единицы объединяются в намеченные типы и определяется численность каждого из них.

Пример 7.1. Имеются данные о производственном травматизме в России по отдельным видам экономической деятельности (табл. 7.1).

Таблица 7.1

Травматизм на производстве в 2004 г.

Вид экономической деятельности	Средне- списочная чис- ленность занятых, тыс. чел.	Численность пострадавших при несчастных случаях на произ- водстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом		Численность пострадавших при несчастных случаях со смертельным исходом	
		всего, тыс. чел.	на 1000 рабо- тающих, чел.	всего, тыс. чел.	на 1000 рабо- тающих, чел.
Всего	49130	87,8	1,7	3,29	0,067
В том числе:					
1. Сельское хозяй- ство, охота и лесное хозяйство	3454	19,9	5,8	0,73	0,213
2. Добыча полезных ископаемых	1032	5,5	5,7	0,34	0,312
3. Обрабатывающие производства	9920	32,3	3,9	0,76	0,098

Вид экономической деятельности	Средне- списоч- ная чис- ленность занятых, тыс. чел.	Численность пострадавших при несчастных случаях на произ- водстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом		Численность пострадавших при несчастных случаях со смертельным исходом	
		всего, тыс. чел.	на 1000 рабо- тающих, чел.	всего, тыс. чел.	на 1000 рабо- тающих, чел.
4. Производство и распределение электроэнергии, газа и воды	1859	3,3	1,9	0,18	0,108
5. Строительство	2865	7,1	4,4	0,53	0,333
6. Транспорт и связь	4182	8,5	2,4	0,41	0,114

И с т о ч н и к: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 728, 731.

Необходимо выделить типы экономической деятельности с высоким, средним и низким уровнем производственного травматизма и произвести группировку занятых в экономике по видам деятельности с разным уровнем травматизма.

Уровень травматизма определяется числом пострадавших при несчастных случаях на производстве с утратой трудоспособности в расчете на 1000 работающих. Те виды деятельности, в которых этот показатель ниже, чем в среднем по экономике, могут характеризоваться как виды деятельности с низким уровнем производственного травматизма. Ни один из шести видов экономической деятельности, указанных в таблице, не удовлетворяет этому условию. Можно предположить, что низкий уровень травматизма будет в прочих видах деятельности, не указанных в таблице. Чтобы проверить это, произведем следующие вычисления:

- среднесписочная численность занятых в прочих видах экономической деятельности: $49130 - (3454 + 1032 + 9920 + 1859 + 2865 + 4182) = 25818$ тыс. чел.;

- численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом в прочих видах экономической деятельности: $87,8 - (19,9 + 5,5 + 32,3 + 3,3 + 7,1 + 8,5) = 11,2$ тыс. чел.;

- численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом в расчете на 1000 работающих в прочих видах экономической деятельности: $(11,2 : 25818) \cdot 1000 = 0,4$ чел. на 1000 работающих.

Полученный результат существенно ниже, чем в среднем по экономике ($0,4 < 1,7$), что подтверждает наше предположение.

В трех видах деятельности: сельское хозяйство, охота и лесное хозяйство, добыча полезных ископаемых, строительство — уровень травматизма (соответственно 4,4, 5,5 и 5,8 чел. на 1000 работающих) существенно превышает средний уровень по экономике, следовательно, эти виды деятельности можно считать высокотравматичными.

Виды деятельности, в которых производственный травматизм превышает средний уровень по экономике не более чем в два раза, будем относить к видам деятельности со средним уровнем травматизма. Это производство и распределение электроэнергии, газа и воды, а также транспорт и связь.

В обрабатывающих производствах число пострадавших при несчастных случаях на производстве с утратой трудоспособности достаточно велико — 3,9 чел. на 1000 работающих, однако уровень травматизма со смертельным исходом здесь самый низкий из всех рассматриваемых видов деятельности, в связи с чем обрабатывающие производства тоже могут быть отнесены к видам деятельности со средним уровнем травматизма.

Для оформления промежуточных результатов группировки построим расчетную таблицу (табл. 7.2).

Таблица 7.2

Расчетная таблица

Тип производственного травматизма	Вид экономической деятельности	Средне-списочная численность занятых, тыс. чел.	Численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом, тыс. чел.	Численность пострадавших при несчастных случаях со смертельным исходом, тыс. чел.
Высокий	Сельское хозяйство, охота и лесное хозяйство; добыча полезных ископаемых; строительство	1032+ +2865+ +3454 = = 7351	19,9 + 5,5 + 7,1 = = 32,5	0,73 + 0,34 + + 0,53 = 1,6
Средний	Обрабатывающие производства; производство и распределение электроэнергии, газа и воды; транспорт и связь	4182+ +1859+ +9920 = = 15961	32,3 + 3,3 + 8,5 = = 44,1	0,76 + 0,18 + + 0,41 = 1,35
Низкий	Прочие виды деятельности	25818	11,2	3,29 – 1,6 – – 1,35 = 0,34

Окончательные итоги типологической группировки оформим в табл. 7.3. Для большей наглядности рассчитаем удельный вес занятых видами деятельности с разным уровнем производственного травматизма, а также показатели производственного травматизма в каждой группе в расчете на 1000 чел.

Итак, в 2004 г. в России более 7 млн чел., или 15,0% работающих, было занято видами деятельности, которые характеризуются высоким уровнем производственного травматизма. К ним относятся: сельское хозяйство, охота и лесное хозяйство, добыча полезных ископаемых, строительство. В 2004 г. в этих отраслях экономики число пострадавших при несчастных случаях на

Таблица 7.3

**Группировка занятых в экономике по видам деятельности
с разным уровнем производственного травматизма**

Тип производ- ственного травма- тизма	Среднесписочная чис- ленность занятых		Численность по- страдавших при несчастных случаях на производстве с утратой трудоспо- собности на один рабочий день и более и со смер- тельным исходом на 1000 работаю- щих, чел.	Численность пострадавших при несчастных случаях со смер- тельным исходом на 1000 работаю- щих, чел.
	тыс. чел.	в процентах от общей численно- сти занятых в экономи- ке		
Высокий	7351	15,0	4,4 $(\frac{32,5}{7351} \cdot 1000)$	0,217 $(\frac{1,6}{7351} \cdot 1000)$
Средний	15961	32,5	2,8 $(\frac{44,1}{15961} \cdot 1000)$	0,085 $(\frac{1,35}{15961} \cdot 1000)$
Низкий	25818	52,5	0,4 $(\frac{11,2}{25818} \cdot 1000)$	0,013 $(\frac{0,34}{25818} \cdot 1000)$

производстве с утратой трудоспособности составило 4,4 чел. на каждую тысячу работающих, что выше среднего показателя по экономике в 2,6 раза. Число пострадавших при несчастных случаях со смертельным исходом составило 1,6 тыс. чел., или 0,217, на каждую тысячу работающих. Этот показатель также более чем в 3 раза превышает средний уровень по экономике. Одна треть работающих была занята видами деятельности со средним уровнем производственного травматизма (обрабатывающие производства, производство и распределение электроэнергии, газа и воды, транспорт и связь), и лишь каждый второй – видами деятельности с низким уровнем производственного травматизма.

Структурная группировка характеризует структуру совокупности по какому-либо признаку. Структурная группировка

может быть построена как по качественному, так и по количественному признаку, в последнем случае ее называют вариационным рядом.

Для характеристики изменения структуры совокупности используют **обобщающие показатели структурных сдвигов**:

- *средний абсолютный показатель изменения структуры* —

$$S_a = \frac{\sum_{j=1}^m |d_{j1} - d_{j0}|}{m};$$

- *средний квадратический показатель изменения структуры* —

$$S_\sigma = \sqrt{\frac{\sum_{j=1}^m (d_{j1} - d_{j0})^2}{m}},$$

где m — число выделенных групп в совокупности;

d_{j1} — удельный вес j -й группы в общей численности совокупности в отчетном (текущем) периоде;

d_{j0} — удельный вес j -й группы в общей численности совокупности в прошлом (базисном) периоде.

Обобщающие показатели структурных сдвигов отражают, на сколько *процентных пунктов* в среднем изменился удельный вес отдельных структурных групп в общей численности совокупности в отчетном периоде по сравнению с базисным. При незначительных изменениях структуры совокупности эти показатели близки к нулю. Верхней границы показатели не имеют.

К обобщающим показателям структурных сдвигов относятся и *показатель степени интенсивности структурных сдвигов — индекс различий*:

$$I_{\text{разл}} = \frac{1}{2} \sum_{j=1}^m |d_{j1} - d_{j0}|.$$

В отличие от предыдущих показателей индекс различий имеет не только нижнюю, но и верхнюю границу: $0 \leq I_{\text{разл}} \leq 100$.

Обобщающие показатели структурных сдвигов используются также для установления различий структур двух совокупностей. В этом случае они трактуются как *обобщающие показатели структурных различий*.

Пример 7.2. В табл.7.4 приведены данные, характеризующие распределение студентов государственных и негосударственных вузов РФ по формам обучения.

Необходимо дать характеристику изменений, произошедших в структуре высшего образования по формам обучения. С помощью обобщающих показателей структурных сдвигов сравнить структурные изменения в распределении по формам обучения в государственных и негосударственных вузах. Рассчитать обобщающие показатели структурных различий в 1993/94 и 2004/05 учебных годах.

Таблица 7.4

**Распределение студентов высших учебных заведений
по формам обучения, %**

Формы обучения	Высшие учебные заведения			
	государственные		негосударственные	
	1993–1994 уч. год d_{j0}^G	2004–2005 уч. год d_{j1}	1993–1994 уч. год d_{j0}^H	2004–2005 уч. год d_{j1}^H
Дневная	63,9	53,6	52,7	28,3
Вечерняя	6,7	5,2	7,6	6,0
Заочная	29,4	38,9	31,0	64,8
Экстернат	0,0	2,3	8,7	0,9
Итого	100,0	100,0	100,0	100,0

Для ответа на поставленные вопросы воспользуемся средними показателями структурных сдвигов и различий. Промежуточные вычисления оформим в таблицу (табл. 7.5).

Таблица 7.5

Расчет обобщающих показателей структурных сдвигов и различий

Форма обучения	$ d_{j1}^{\Gamma} - d_{j0}^{\Gamma} $	$ d_{j1}^H - d_{j0}^H $	$ d_{j0}^{\Gamma} - d_{j0}^H $	$ d_{j1}^{\Gamma} - d_{j1}^H $
Дневная	10,3	24,4	11,2	25,3
Вечерняя	1,5	1,6	0,9	0,8
Заочная	9,5	33,8	1,6	25,9
Экстернат	2,3	7,8	8,7	1,4
$\sum d_{j1} - d_{j0} $	23,6	67,6	22,4	53,4
Средний абсолютный показатель структурных изменений (различий) — S_a , п.п.	23,6:4=5,9	67,6:4=16,9	22,4:4=5,6	53,4:4=13,4

Расчеты показывают, что государственная и негосударственная системы высшего образования имеют существенные структурные различия по формам обучения. Причем в последние годы эти различия усиливаются. Первые негосударственные вузы, появившиеся в России в 1993/94 учебном году, обычно организовывались «по образу и подобию» государственных высших учебных заведений. Основной формой обучения как в государственных, так и в негосударственных вузах была дневная форма обучения; около трети всех студентов обучалось на заочных отделениях, примерно одинаковый процент студентов обучался на вечерних отделениях государственных и негосударственных вузов. Основное различие касалось лишь обучения экстерном (в государственных вузах эта форма обучения, как правило, не практиковалась). Обобщающий показатель абсолютных различий в формах обучения составлял всего 5,6 п. п. (процентных пункта). Впоследствии, в 1993/94 учебном году, негосударственная система образования оказалась более гибкой в условиях быстро меняющейся ситуации в стране, она быстрее реагировала на расширение социальных запросов на образовательные услуги, наличие незаполненных «ниш» в системе высшего образования, спрос на ускоренное обучение и второе высшее образование и т.д. Если за рассматриваемый период структура государственной системы образования по формам обучения изменилась на 5,9 п. п., то негосударственной — на 16,9 п. п. В государственных вузах приоритетной формой обучения осталась дневная,

а негосударственные вузы переориентировались на широкое предоставление заочных образовательных услуг: в 2004/05 учебном году 64,8% студентов негосударственных вузов обучалось заочно. Обобщающий показатель абсолютных различий в формах обучения в государственных и негосударственных вузах вырос до 13,4 п. п.

Аналитическая группировка проводится для выявления взаимосвязи между двумя и более признаками, из которых один рассматривается как результат, другой (другие) — как фактор (факторы).

Рассмотрим пример построения и анализа однофакторной аналитической группировки.

Пример 7.3. В табл. П1.1, П1.2, П1.4–П1.6 приведены данные, позволяющие изучать связь между некоторыми показателями развития науки, инновационных процессов и результатами экономической деятельности в отдельных странах мира. Проанализируем, как влияет численность занятых научно-исследовательскими, опытными и конструкторскими разработками (НИОКР) — признак-фактор x , на размер валового национального дохода — признак-результат y .

Построим перечневую таблицу, содержащую исходные данные, проранжированные в порядке возрастания численности занятых в НИОКР на 1 млн чел. (табл. 7.6).

Таблица 7.6

Ранжирование стран

№ п/п	Страна	Работники, занятые в НИОКР (на 1 млн чел.)	ВНД на душу населения, долл.	№ п/п	Страна	Работники, занятые в НИОКР (на 1 млн чел.)	ВНД на душу населения, долл.
1	Сенегал	2	540	10	Конго	29	100
2	Оман	4	7830	11	Руанда	30	220
3	Ямайка	8	2980	12	Лесото	42	610
4	Нигерия	15	350	13	Замбия	47	380
5	Мадагаскар	15	290	14	Центрально-африканская Республика	47	260
6	Буркина-Фасо	17	300	15	Бангладеш	51	400
7	Бурунди	21	90	16	Непал	62	240
8	Уганда	25	250	17	Никарагуа	73	740
9	Сирия	29	1150	18	Кувейт	73	17960

Продолжение

№ п/п	Страна	Работники, занятые в НИОКР (на 1 млн чел.)	ВНД на душу населе- ния, долл.	№ п/п	Страна	Работники, занятые в НИОКР (на 1 млн чел.)	ВНД на душу населе- ния, долл.
19	Гондурас	74	970	51	Египет	493	1390
20	Колумбия	81	1810	52	Македония	500	1980
21	Парагвай	83	1110	53	Коста-Рика	533	4300
22	Эквадор	84	1830	54	Китай	633	1100
23	Пакистан	88	520	55	Таджики- стан	660	210
24	Панама	95	4060	56	Монголия	710	480
25	Того	102	310	57	Аргентина	715	3810
26	Гватемала	103	1910	58	Казахстан	744	1780
27	Боливия	118	900	59	Румыния	910	2260
28	Индия	120	540	60	Тунис	1013	2240
29	Индон е- зия	130	810	61	Италия	1156	21570
30	Сальва- дор	131	2340	62	Болгария	1158	2130
31	Филип- пины	156	1080	63	Азербай- джан	1248	820
32	Респу б- лика Молдова	171	590	64	Греция	1357	13230
33	Бенин	174	440	65	Чехия	1467	7150
34	ЮАР	192	2750	66	Польша	1469	5280
35	Шри- Ланка	197	930	67	Венгрия	1473	2580
36	Венесуэла	222	3490	68	Латвия	1476	4400
37	Перу	225	2140	69	Гонконг	1568	25860
38	Мексика	259	6230	70	Израиль	1570	16240
39	Вьетнам	274	480	71	Армения	1606	950
40	Гвинея	286	430	72	Словакия	1707	4940
41	Таиланд	289	2190	73	Португалия	1745	11800
42	Малайзия	294	3880	74	Украина	1749	970
43	Бразилия	324	2720	75	Узбекистан	1754	420
44	Турция	345	2800	76	Литва	1824	4500
45	Тринидад и Тобаго	347	7790	77	Беларусь	1870	1600
46	Маври- кий	360	4100	78	Хорватия	1920	5370
47	Уругвай	370	3820	79	Иордания	1977	1850
48	Киргизия	413	340	80	Испания	2036	1700
49	Чили	419	4360	81	Эстония	2253	5380
50	Иран	484	2010	82	Ирландия	2315	27010

Продолжение

№ п/п	Страна	Работники, занятые в НИОКР (на 1 млн чел.)	ВНД на душу насе- ле- ния, долл.	№ п/п	Страна	Работники, занятые в НИОКР (на 1 млн чел.)	ВНД на душу насе- ле- ния, долл.
83	Грузия	2317	770	94	Австралия	3446	21950
84	Австрия	2346	26810	95	Канада	3487	24470
85	Словения	2364	11920	96	Швейцария	3594	40680
86	Новая Зеландия	2593	15530	97	Сингапур	4352	21230
87	Велико- британия	2691	28320	98	Норвегия	4442	43300
88	Нидер- ланды	2826	26230	99	США	4526	37870
89	Респу- блика Ко- рея	2979	12030	100	Дания	4822	33570
90	Франция	3134	24730	101	Япония	5085	34180
91	Бельгия	3180	25760	102	Швеция	5171	28910
92	Германия	3222	25270	103	Финляндия	7461	27060
93	Россия	3415	2610				

Источник: Состояние окружающей среды. 2005: Статистический справочник Всемирного Банка: пер. с англ. — М.: Весь мир, 2005; Доклад о развитии человека 2005. Международное сотрудничество на перепутье: помощь, торговля и безопасность в мире неравенства: пер. с англ. — М.: Весь Мир, 2005.

Просмотр исходных данных показывает, что при одинаковой численности занятых в НИОКР в расчете на 1 млн чел. страны имеют очень разные размеры ВНД на душу населения. Однако в целом прослеживается закономерность: с ростом численности занятых в НИОКР растут и показатели ВНД. Для того чтобы установить связь между признаками, сгруппируем страны по признаку-фактору, затем для каждой группы рассчитаем среднее значение признака-результата.

Поскольку большинство стран мира характеризуются невысоким уровнем занятости в сфере НИОКР, построим аналитическую группировку с неравными интервалами (табл. 7.7).

Таблица 7.7

**Характеристика зависимости ВНД (на душу населения)
от числа работников, занятых в НИОКР (на 1 млн чел.) в странах мира**

Работники, занятые в НИОКР (на 1 млн чел.) x_j	Число стран n_j	Средний ВНД на душу населения, долл. \bar{y}_j
До 100	24	1875
100–1000	35	2191
1000–2000	20	6695
2000–3000	10	15570
3000–7500	14	27971
Итого	103	8122

Сопоставление изменений числа работников, занятых в НИОКР, с изменением среднего размера ВНД по группам стран явно свидетельствует о наличии связи между признаками. Поскольку с ростом признака-фактора признак-результат увеличивается, *связь* между признаками *прямая* (если бы с увеличением признака-фактора признак-результат уменьшался, связь была бы *обратной*).

Поскольку мы используем группировку с неравными интервалами, для того чтобы сделать вывод о *форме связи* (линейная или нелинейная), рассчитаем изменения признака-результата на единицу изменения признака-фактора. Такие величины называются показателями **силы связи**. Показатель силы связи отвечает на вопрос, как изменится признак-результат при изменении признака-фактора на одну единицу, и для каждой группы рассчитывается по формуле

$$b_{yx_j} = \frac{\bar{y}_j - \bar{y}_{j-1}}{x'_j - x'_{j-1}},$$

где b_{yx_j} — показатель силы связи в j -й группе;

\bar{y}_j, \bar{y}_{j-1} — средние значения признака-результата в j -й и в предыдущей $j - 1$ -й группах;

x'_j, x'_{j-1} — середины интервалов признака-фактора в j -й и в предыдущей $j - 1$ -й группах.

Если полученные показатели силы связи значительно отличаются по группам, делают вывод о наличии *нелинейной* связи, так как одинаковому приращению признака-фактора соответ-

ствуют разные приращения признака-результата. При близких групповых показателях силы связи можно говорить о *линейной* зависимости — одинаковому приращению признака-фактора соответствуют равные приращения признака-результата. В случае линейной связи рассчитывают показатель *средней силы связи*:

$$\bar{b}_{yx} = \frac{\bar{y}_m - \bar{y}_1}{x'_m - x'_1},$$

где m — число групп в совокупности;

\bar{y}_m, \bar{y}_1 — средние значения признака-результата соответственно в последней и 1-й группах;

x'_m, x'_1 — середины интервалов признака-фактора соответственно в последней и 1-й группах.

В случае прямой связи показатели силы связи имеют положительные значения, в случае обратной связи — отрицательные.

Рассчитаем показатели силы связи для нашего примера (табл. 7.8).

Таблица 7.8

Расчет показателей силы связи

Работники, занятые в НИОКР (на 1 млн чел.) x_j	Середин ы интервалов x'_j	$x'_j - x'_{j-1}$	Средний ВНД на душу насе- ления, долл. \bar{y}_j	Изменение ВНД на душу насе- ления $\bar{y}_j - \bar{y}_{j-1}$	Сила связи, долл./ чел. b_{yx_j}
До 100	50	—	1875	—	—
100—1000	550	500	2191	+316	+0,6
1000—2000	1500	950	6695	+4504	+4,7
2000—3000	2500	1000	15570	+8875	+8,9
3000—7500	5250	2750	27971	+12401	+4,5

Показатели силы связи существенно отличаются по группам, следовательно, между признаками существует нелинейная зависимость. Положительные значения показателей силы связи подтверждают прямой характер связи. Наибольшее значение показатель силы связи имеет в группе стран, где число занятых в НИОКР находится в пределах от 2 до 3 тыс. на 1 млн чел. населения. При увеличении занятых в НИОКР на 1 чел. на каждый

1 млн населения в этих странах можно ожидать приращение ВНД на душу населения на 8,9 долл. В странах, где сегодня наблюдаются самые высокие показатели занятости в научной сфере (последняя группа), эффект от дальнейшего наращивания занятости в этой сфере будет уже не столь существенным. Рост занятости на 1 чел. на каждый миллион населения приведет к росту ВНД на душу населения на 4,5 долл. Наименьший показатель силы связи в группе стран с самыми низкими показателями занятости в научной сфере — до 1 тыс. на 1 млн чел. Увеличение относительного показателя занятости в научной сфере на 1 чел. обеспечивает рост ВНД на душу населения всего на 0,6 долл.

По аналитической группировке можно определить еще одну характеристику связи — показатель **тесноты связи**. Его вычисление основано на *правиле разложения дисперсии*, согласно которому *общая дисперсия равна сумме межгрупповой дисперсии и средней из внутригрупповых дисперсий*:

$$\sigma_y^2 = \sigma_{\bar{y}x}^2 + \overline{\sigma_{yx}^2}.$$

Межгрупповая дисперсия характеризует колеблемость результата, вызванную признаком-фактором, поэтому ее также называют *факторной дисперсией* и определяют по формуле

$$\sigma_{\bar{y}x}^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 n_j}{\sum_1^m n_j},$$

где \bar{y} — среднее значение результативного признака в совокупности;
 \bar{y}_j — среднее значение результативного признака в j -й группе;
 n_j — число единиц в j -й группе.

Внутригрупповая дисперсия возникает за счет других факторов (не связанных с изучаемым). Внутригрупповая дисперсия называется *остаточной* (та колеблемость, которая осталась при закреплении изучаемого фактора x) и для каждой группы рассчитывается по формуле

$$\sigma_j^2 = \frac{\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{n_j},$$

где y_{ij} — значение результативного признака у i -й единицы в j -й группе.

Внутригрупповые дисперсии, рассчитанные для отдельных групп, осредняются в единый показатель внутригрупповой дисперсии:

$$\overline{\sigma_{yx}^2} = \frac{\sum_1^m \sigma_j^2 n_j}{\sum_1^m n_j}.$$

Отношение межгрупповой дисперсии к общей называется **коэффициентом детерминации**:

$$\eta^2 = \frac{\sigma_{yx}^2}{\sigma_y^2}.$$

Коэффициент детерминации показывает, какая часть вариации признака-результата вызвана вариацией признака-фактора. Квадратный корень из коэффициента детерминации — *эмпирическое корреляционное отношение*:

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}.$$

По значению эмпирического корреляционного отношения судят о *тесноте связи* между признаками. Обычно придерживаются следующей шкалы: если

- $\eta \leq 0,3$ — связь слабая;
- $0,3 < \eta \leq 0,5$ — связь заметная;
- $0,5 < \eta \leq 0,7$ — связь умеренно тесная;
- $0,7 < \eta \leq 0,9$ — связь тесная;
- $\eta > 0,9$ — связь очень тесная.

Определим тесноту связи для нашего примера. Промежуточные вычисления представлены в табл. 7.9.

Таблица 7.9

Расчет межгрупповой дисперсии

Работники, занятые в НИОКР (на 1 млн чел.) x_j	Число стран n_j	Средний ВНД на душу насе- ления, долл. \bar{y}_j	$\bar{y}_j - \bar{y}$ ($\bar{y} = 8122$)	$(\bar{y}_j - \bar{y})^2$	$(\bar{y}_j - \bar{y})^2 n_j$
До 100	24	1875	-6247	39025009	936600216
100–1000	35	2191	-5931	35176761	1231186635
1000–2000	20	6695	-1427	2036329	40726580
2000–3000	10	15570	7448	55472704	554727040
3000–7500	14	27971	19849	393982801	5515759214
Итого	103	8122	—	—	8278999685

Межгрупповая дисперсия для нашего примера составляет:

$$\sigma_{yx}^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 n_j}{\sum_1^m n_j} = \frac{8278999685}{103} = 80378637,7.$$

Общую дисперсию рассчитаем по несгруппированным данным (см. табл. 7.7) по формуле

$$\sigma_y^2 = \frac{\sum_i (y_i - \bar{y})^2}{n},$$

где y_i — индивидуальные значения признака-результата у i -й единицы совокупности;

n — число единиц совокупности.

Для нашего примера общая дисперсия составила 117728128,9. Тогда эмпирическое корреляционное отношение равно:

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}} = \sqrt{\frac{80378637,7}{117728128,9}} = \sqrt{0,683} = 0,826.$$

Полученное значение указывает на то, что связь между занятостью в НИОКР и показателями ВНД тесная; 68,3% вариации признака-результата определяется вариацией признака-фактора.

Для изучения влияния нескольких факторов на результат строится *многофакторная аналитическая группировка* как комбинационная группировка по признакам-факторам. Многофакторная аналитическая группировка позволяет уловить влияние факторов на результат с изменением условий (закреплением прочих факторов на разных уровнях). Однако метод многофакторной группировки имеет огромный минус — дробление совокупности, в результате чего выделяются подгруппы незаполненные или с малым числом единиц.

Полученные характеристики связи в таких случаях не могут считаться надежными. Поэтому многофакторную группировку следует применять лишь для совокупностей большого объема и при условии, что распределения признаков-факторов не являются крайне асимметричными. Подробнее об особенностях построения и анализа многофакторных группировок см. учебник И.И. Елисеевой, М.М. Юзбашева «Общая теория статистики»* (глава 6, разд. 6.2).

Изучая связи между признаками на основе аналитических группировок, следует помнить, что надежность выводов на их основе зависит от объема совокупности (n) и количества выделенных групп (m).

Пример 7.4. По исходным данным предыдущего примера (см. табл. 7.6) в целях изучения зависимости между численностью занятых в НИОКР и размерами ВНД построим варианты группировки с разным числом групп и выберем наилучший вариант.

Поскольку вариант группировки с пятью группами был рассмотрен выше, построим группировки, где число групп равно 3, 4, 6 и 7. Выделение большего числа групп нецелесообразно, так как это приведет к сильному дроблению совокупности. Результаты группировок представлены в табл. 7.10 и 7.11.

* Елисеева И.И., Юзбашев М.М. Общая теория статистики: учебник / под ред. И.И. Елисеевой. — 5-е изд., перераб. и доп. — М.: Финансы и статистика, 2006; далее — Учебник.

Таблица 7.10

**Зависимость ВНД (на душу населения) от числа работников,
занятых в НИОКР (на 1 млн чел.) в странах мира**

Работники, занятые в НИОКР (на 1 млн чел.) x_j	Число стран n_j	Средний ВНД на душу насе- ления, долл. \bar{y}_j	Работники, занятые в НИОКР (на 1 млн чел.) x_j	Число стран n_j	Средний ВНД на душу насе- ления, долл. \bar{y}_j
До 100	24	1875	До 100	24	1875
100–500	27	2251	100–1000	35	2191
500–1000	8	1990	1000–2000	20	6695
1000–2000	20	6695	2000–3000	10	15570
2000–3000	10	15570	3000–4000	7	23639
3000–4000	7	23639	4000–7500	7	32303
4000–7500	7	32303			
Итого	103	8122	Итого	103	8122

Таблица 7.11

**Зависимость ВНД (на душу населения) от числа работников,
занятых в НИОКР (на 1 млн чел.) в странах мира**

Работники, занятые в НИОКР (на 1 млн чел.) x_j	Число стран n_j	Средний ВНД на душу населения, долл. \bar{y}_j	Работники, занятые в НИОКР (на 1 млн чел.) x_j	Число стран n_j	Средний ВНД на душу населения, долл. \bar{y}_j
До 100	24	1875	До 1000	59	2062
100–1000	35	2191	1000–3000	30	9653
1000–2000	20	6695	3000–7500	14	27971
2000–7500	24	22804			
Итого	103	8122	Итого	103	8122

Как видим, все группировки свидетельствуют о наличии взаимосвязи между признаками. Какой из вариантов группировки лучше? Чтобы ответить на этот вопрос, для каждого варианта рассчитывают эмпирическое корреляционное отношение (или коэффициент детерминации). Лучшим считается тот вариант, в котором этот показатель будет выше. Однако окончательный вывод можно сделать на основе соотношения величины корреляционного значения (η_{yx}) с его случайным значением (η_0), которое зависит от числа групп и числа единиц совокупности:

$$\eta_0 = \sqrt{\frac{m-1}{n-1}}.$$

Группировка, для которой отношение $\eta_{yx} : \eta_0$ больше, признается лучшей (табл. 7.12).

Таблица 7.12

Фактические и случайные значения корреляционного отношения при разных вариантах группировки

Число групп	Корреляционное отношение η_{yx}	Случайное значение корреляционного отношения η_0	$\eta_{yx} : \eta_0$
7	0,839	0,238	3,52
6	0,839	0,221	3,80
5	0,826	0,198	4,17
4	0,780	0,171	4,56
3	0,799	0,140	5,70

Группировки, в которых число групп равно трем и четырем, отклоняются, поскольку имеют более низкие корреляционные отношения. Из других группировок наибольшее соотношение $\eta_{yx} : \eta_0$ имеет группировка с пятью группами. Следовательно, для нашего примера максимально раскрывает действие признака-фактора на признак-результат группировка с пятью группами, которая и была проанализирована выше.

Пример 7.5*. Нередко приходится проводить классификацию по многим признакам — *многомерную классификацию*. Проведем с помощью пакета прикладных программ SPSS классификацию банков, предоставляющих ипотечные программы кредитования в рублях в г. Санкт-Петербурге для покупки квартир только на вторичном рынке на 01.01.2007 г. (табл. 7.13).

В качестве классифицирующих признаков используем следующие характеристики ипотечных программ и некоторые требования к заемщикам ипотеки, выдвигаемые банками:

1) срок *min* — показывает, на какой минимальный срок в данном банке можно получить ипотеку: 1 — от 1 года до 3 лет, 2 — от 3 до 5 лет, 3 — от 5 лет;

* Подготовлен аспиранткой М.В. Бочениной.

Таблица 7.13
Характеристики ипотечных программ кредитования банков Санкт-Петербурга на 01.01.2007 г., руб.

Банк	Срок		Сумма		Взносы	Ставка min	Комиссия	Доход	Прописка	Платежи	Возраст М
	min	max	min	max							
Агропромкредит	1	3	2	3	2	2	2	1	2	1	2
Александровский	1	3	3	2	2	1	1	1	2	1	3
Альфа	3	1	3	3	3	3	2	1	2	1	2
Бинбанк	3	2	3	2	3	3	2	1	2	1	1
Башэкономбанк	1	3	1	1	3	1	2	2	2	1	3
Банк Москвы	2	3	2	3	1	2	2	1	1	1	2
ВЕФК	2	1	3	3	3	2	0	1	2	1	1
Викинг	2	3	3	2	3	1	2	2	3	1	3
Глобэкс	1	1	3	1	3	1	0	1	3	2	2
Инкасбанк	2	3	3	2	2	1	2	1	3	1	3
ИМПЭКСБАНК	3	3	3	1	1	2	2	1	1	1	2
Констанс-БАНК	1	3	3	2	2	1	2	1	1	1	2
МеждМосковский	1	2	2	3	3	1	0	2	0	1	2
НацРезервныйБ	3	3	2	3	3	3	2	1	1	1	2
НационалБанкТраст	2	3	3	1	2	1	2	1	1	1	2
Петро-Аэро-Банк	2	1	3	3	3	2	0	1	2	1	1
ПетерСоцКомБан	3	3	3	2	2	2	3	1	2	1	3
Российский Капитал	1	3	3	3	2	1	3	1	0	1	3
Русь-БАНК	1	3	3	2	2	1	3	1	2	1	3
Советский	1	3	3	2	2	2	2	1	2	1	2
Сосьете Жен Вос	1	1	3	3	3	3	3	1	0	1	2
Союз	1	2	3	3	3	3	3	1	1	1	1
Энергомашбанк	1	3	3	3	2	1	2	2	2	1	3
Юниаструм	2	3	2	2	2	1	2	1	2	1	3

2) срок t_{\max} — на какой максимальный срок банк выдает ипотеку: 1 — до 15 лет, 2 — до 20 лет, 3 — до 30 лет;

3) сумма \min — какую минимальную сумму можно получить в ипотеку в данном банке: 1 — от 50 тыс. руб., 2 — от 150 тыс. руб., 3 — от 300 тыс. руб.;

4) сумма t_{\max} — максимальная сумма ипотеки составляет: 1 — до 10 млн руб., 2 — до 17 млн руб., 3 — свыше 17 млн руб.;

5) взнос — первоначальный взнос на покупку квартиры в процентах от ее стоимости составляет: 1 — 0%, 2 — 10%, 3 — свыше 10%;

6) ставка \min — минимальная ставка по ипотечному кредитованию, которую можно получить в данном банке: 1 — от 11% годовых, 2 — от 12% годовых, 3 — от 13% годовых, 4 — плавающая ставка;

7) комиссия — сумма, которую берет банк одновременно за открытие и ведение ссудного счета по ипотеке, как правило, в процентах к сумме кредита: 0 — 0%, 1 — больше 0% — 0,5%, 2 — больше 0,5% до 1,5%, 3 — фиксированная сумма до 10000 руб.;

8) доход — форма подтверждения доходов заемщика, которая требуется банком: 1 — учитываются «серые» доходы, 2 — не учитываются «серые» доходы;

9) прописка — требование к заемщику по гражданству и виду регистрации в Санкт-Петербурге: 0 — гражданства и регистрации не требуется, 1 — постоянная регистрация в Санкт-Петербурге, 2 — гражданство РФ, 3 — временная регистрация в Санкт-Петербурге;

10) платежи — какими платежами можно возвращать ипотечный кредит: 1 — равные ежемесячные платежи, 2 — дифференциальные или равные ежемесячные платежи;

11) возраст M — максимально возможный возраст заемщика на конец срока ипотечного кредитования: 1 — 55 лет, 2 — 60 лет, 3 — 75 лет.

Применив процедуру кластерного анализа в программе SPSS, мы получили таблицу последовательности слияний (табл. 7.14) и дендрограмму (рис. 7.1).

В табл. 7.14 вторая колонка (Кластер объединен) содержит два столбца — с кластером 1 и с кластером 2, которые соответствуют номерам кластеров, объединяемым на данном этапе. После объединения кластеру присваивается номер, соответствующий номеру в столбце «с кластером» 1. Третья колонка (Коэф-

Таблица 7.14

Таблица последовательности слияния

Этап	Кластер объединен		Кoeffициент	Этап первого появления кластера		Следующий этап
	с кластером 1	с кластером 2		кластер 1	кластер 2	
1	7	16	0,000	0	0	13
2	12	20	2,843	0	0	6
3	10	19	3,921	0	0	4
4	2	10	4,117	0	3	9
5	21	22	4,670	0	0	18
6	12	15	4,734	2	0	10
7	3	4	5,140	0	0	13
8	1	6	5,232	0	0	15
9	2	24	5,484	4	0	10
10	2	12	6,793	9	6	14
11	8	23	7,007	0	0	19
12	11	17	8,471	0	0	14
13	3	7	9,957	7	1	17
14	2	11	10,880	10	12	15
15	1	2	11,939	8	14	16
16	1	18	12,747	15	0	19
17	3	14	14,875	13	0	18
18	3	21	15,464	17	5	21
19	1	8	16,615	16	11	21
20	5	13	23,302	0	0	22
21	1	3	23,518	19	18	22
22	1	5	30,601	21	20	23
23	1	9	49,809	22	0	0

фициент) содержит значения Евклидова расстояния между кластерами, которые объединяются на данном этапе. Следующая колонка (Этап первого появления кластера) показывает, на каком этапе до этого появлялись первый и второй из объединенных кластеров. Последняя колонка (Следующий этап) показывает, на каком этапе снова появится кластер, образованный на этом этапе.

Из табл. 7.14 видно, что на этапе 1 происходит объединение банков 7 и 16, расстояние между ними равно 0,000 (значит, условия предоставления ипотеки в этих двух банках практически совпадают). Ни один из этих банков не принадлежит какому-либо кластеру, о чем говорят нули в колонке (Этап первого по-

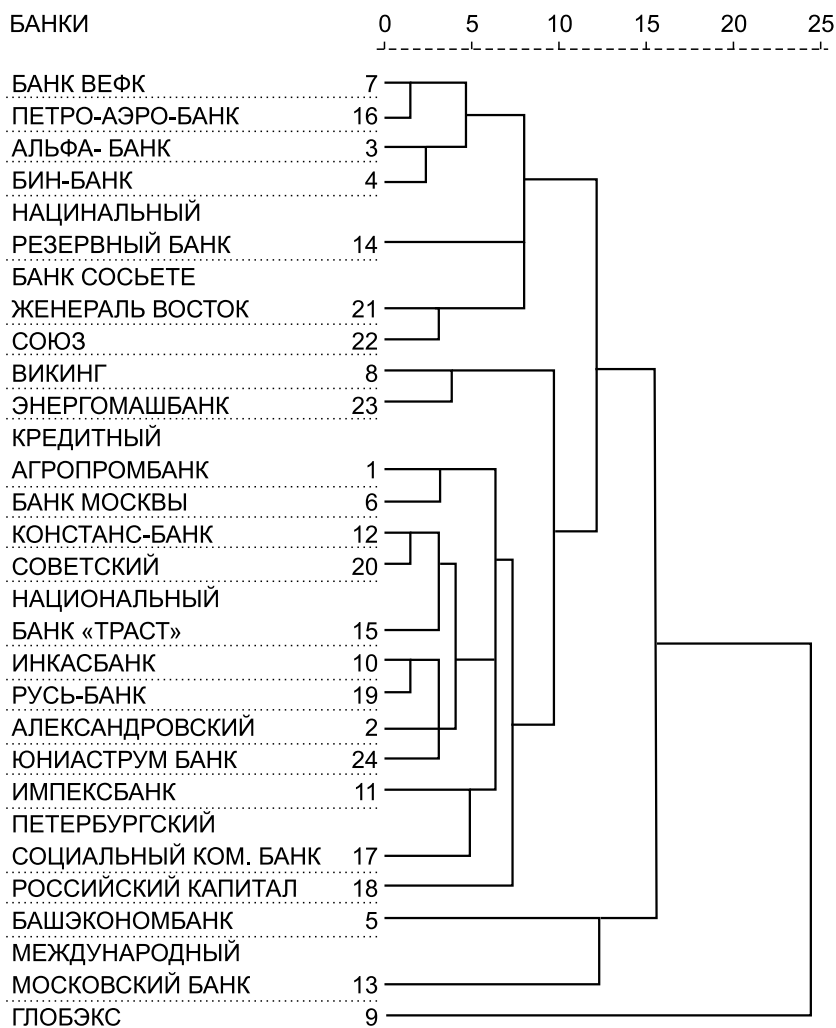


Рис. 7.1. Дендрограмма классификации банков, предоставляющих ипотеку для покупки жилья на вторичном рынке

явления кластера). Следующим этапом для данного кластера является этап 13 (это видно из последней колонки), на котором к нему присоединяется кластер 3, который появился на этапе 7. Расстояние между кластерами 7 и 3 равно 9,957. Следующий

этап 17, на котором к кластеру 3 присоединяется кластер 14, не принадлежащий какому-либо кластеру (расстояние — 14,875), затем этап 18, где соединяются кластер 21, образованный на этапе 5, и кластер 3 (расстояние — 15,464). Затем на этапе 21 к ним присоединяется кластер 1, образованный на этапе 19 (расстояние — 23,518). Далее на этапе 22 происходит объединение с кластером 5, образованным на этапе 20 (расстояние — 30,601) и последний этап 23 объединяет с кластером 9, не принадлежащим какому-либо кластеру (расстояние — 49,809) и т.д.

Таблица последовательности слияния позволяет оценить число кластеров. Для этого необходимо проследить динамику увеличения различий по шагам кластеризации. Оптимальному числу кластеров соответствует разность между количеством всех классифицируемых банков и порядковым номером шага, на котором был обнаружен заметный перепад различий между банками.

Как видно из табл. 7.14, заметное изменение различий наблюдается при переходе от 19-го к 20-му этапу кластеризации. На этом этапе оптимальное число кластеров равно $24 - 19 = 5$ или $24 - 20 = 4$, т.е. 5 или 4.

Дендрограмма (см. рис. 7.1) позволяет определить место любого банка на любом уровне кластеризации и показывает, каково расстояние между кластерами или банками на каждом уровне. Числа от 0 до 25 являются условной шкалой этих расстояний, где 0 соответствует наименьшему расстоянию на 1-м этапе, а 25 — наибольшему на последнем этапе.

Рассмотрим состав каждого из четырех кластеров.

Первый кластер (образованный на этапе 17) включает 7 банков, которые характеризуются большим первоначальным взносом (свыше 10% стоимости кредита); соглашаются учитывать доходы, которые заемщик не может подтвердить документально («серые»), для этого заемщик заполняет справку по форме банка; возврат кредита осуществляется равными ежемесячными платежами.

Второй кластер (образованный на этапе 19) состоит из 14 банков, характеризуется максимальным сроком кредитования (до 30 лет), преобладает минимальная сумма кредита (от 300000 руб.), первоначальный взнос составляет порядка 10% стоимости квартиры; возврат кредита осуществляется равными платежами.

Третий кластер состоит из одного Башэкономбанка, кредитует покупку квартир на срок от 1 года, первоначальный взнос

превышает 10%, минимальная ставка по кредиту составляет 11% годовых, не учитывает «серые» доходы, возврат кредита осуществляется равными ежемесячными платежами, комиссия за открытие и ведение ссудного счета в процентах к сумме кредита составляет около 1% единовременно, заемщик должен иметь гражданство РФ.

Четвертый кластер состоит из одного Международного Московского банка, кредитует покупку квартир на срок от 1 года, первоначальный взнос превышает 10%, минимальная ставка по кредиту составляет 11% годовых, не учитывает «серые» доходы, возврат кредита осуществляется равными ежемесячными платежами, комиссия за открытие и ведение ссудного счета отсутствует, гражданства РФ и регистрации заемщика не требуется.

Пятый кластер состоит из одного банка Глобэкс. Предоставляет ипотеку на срок от 1 года до 15 лет на сумму от 300 тыс. до 10 млн руб., первоначальный взнос превышает 10%, минимальная ставка по кредиту 11%, возврат кредита возможен дифференцированными платежами, комиссия за выдачу и обслуживание кредита отсутствует, максимально возможный возраст заемщика на конец срока ипотечного кредитования 60 лет.

В варианте классификации на четыре кластера: **третий** и **четвертый** кластеры предыдущего варианта сливаются в один кластер на этапе 20. Таким образом, **первый** кластер включает 7 банков, **второй** — 14 банков, **третий** — 2 банка, **четвертый** кластер — 1 банк.

7.2. Задания по теме

1. Приведите примеры признаков, которые могут быть положены в основание группировки предприятий региона на крупные, средние, мелкие в разных отраслях экономики.

2. Какие признаки, на ваш взгляд, могут быть использованы для группировки семей по уровню благосостояния. Разработайте макеты группировочных таблиц.

3. На российском соковом рынке лидирующие позиции занимают четыре основных производителя (табл. 7.15).

Таблица 7.15

Производитель	Год основания	Торговые марки соков
«Вимм-Билль-Данн Продукты питания»	1992	«Любимый сад», «J7», «Rio Grande», «100% Gold»
ЗАО «Мултон»	1995	«Добрый», «Дары лета», «Нико», «Rich»
ООО «Нидан»	1998	«Да», «Чемпион», «Моя семья»
ОАО «Лебедянский»	1967	«Фруктовый сад», «Я», «Тонус»

В табл. 7.16 приведены данные о доле рынка основных производителей соковой продукции по показателям объемов продаж, %.

Таблица 7.16

Производитель	2001	2002	2003	2004	2005
ОАО «Лебедянский»	12,9	18,2	20,6	20,9	22,1
ЗАО «Мултон»	22,1	22,2	21,1	20,2	19,7
ООО «Нидан»	11,7	13,8	14,6	15,7	16,7
«Вимм-Билль-Данн Продукты питания»	21,8	21,7	19,6	17,9	16,5
ООО «Троя-Ультра»	1,1	1,4	1,9	2,8	2,8
Другие	30,4	22,7	22,9	22,5	22,2
Итого	100,0	100,0	100,0	100,0	100,0
Примечание. В группу «Другие» вошли компании, доля которых не превышает 1% общего объема продаж.					

Источники: Невыдуманные истории: сборник кейсов по материалам газеты «Ведомости». Вып. 1. — СПб., 2006.

Таблица 7.17

Ценовой сегмент	Доля на российском рынке	Марка сока
Низкий (менее 24 руб./л)	33,7	«Чемпион», «Моя семья», «Дары лета»
Средний (24–38 руб./л)	36,9	«Добрый», «Фруктовый сад», «Rich»
Премиальный (более 38 руб./л)	29,4	«J7», «100% Gold»

Кроме того, имеется информация о доле отдельных ценовых сегментов на рынке соковой продукции (табл. 7.17).

Оцените степень концентрации на рынке соковой продукции. Проанализируйте изменения, которые произошли в структуре рынка соковой продукции за последние годы. Связаны ли эти изменения с тем, в каком ценовом сегменте работает та или иная компания? Напишите аналитическую записку с вашими выводами, для наглядного представления результатов используйте графики.

4. Известны данные о структуре занятых в экономике России и некоторых зарубежных стран по видам экономической деятельности, % (табл. 7.18).

С помощью обобщающих показателей структурных различий определите, экономики каких стран наиболее схожи по структуре занятых.

К распределению занятых в экономике какой из стран наиболее близко распределение по видам деятельности в России?

5. В табл. 7.19 приведены данные о структуре платных услуг населению в Санкт-Петербурге и Ленинградской области в 2005 г., %.

Объем платных услуг населению в Санкт-Петербурге и Ленинградской области составил в 2005 г. соответственно 126,4 и 14,8 млрд руб.

Среднегодовая численность населения в Санкт-Петербурге в 2005 г. — 4590,5 тыс. чел., в Ленинградской области — 1648,5 тыс. чел.

Рассчитайте обобщающие показатели структурных различий и сделайте вывод о существенности различий в структуре платных услуг в двух регионах. Постройте графики, отражающие структуру платных услуг в каждом регионе. Напишите аналитическую записку об объемах и структуре услуг в Санкт-Петербурге и Ленинградской области. Дополните свой анализ показателями размера платных услуг на душу населения (в целом и по видам).

6. В табл. 7.20 приведены показатели дошкольного образования в России (тыс. чел. на конец года).

Рассчитайте для каждого года удельный вес работников отдельных категорий в общей численности педагогических работников, а также показатели, характеризующие обеспеченность дошкольных учреждений педагогическими работниками разных категорий (в расчете на 1000 детей) и нагрузку детьми на одного педагогического работника. Оформите результаты в таблице. К какому виду группировок относится построенная вами таблица?

Таблица 7.18

Вид экономической деятельности	Россия, 2004	Вели- кобри- тания, 2003	Герма- ния, 2003	Поль- ша, 2003	США, 2003	Фин- ляндия, 2003	Фран- ция, 2003	Шве- ция, 2003
Всего	100	100	100	100	100	100	100	100
Сельское хозяйство, охота и лесное хозяйство	8,4	1,2	2,5	18,3	1,7*	5,0	4,3	2,1
Рыболовство, рыбоводство	0,3	0,1	0,0	0,1	...	0,0	0,1	0,0
Добыча полезных ископаемых	1,7	0,4	0,4	1,8	0,4	0,2	0,1	0,2
Обрабатывающие производства	18,8	14,7	22,8	19,0	12,3	18,6	17,2	16,3
Производство и распределение электроэнергии, газа и воды	2,8	0,7	0,8	1,8	0,9	0,8	0,9	0,6
Строительство	6,0	7,5	7,2	5,9	7,4	6,3	6,7	5,6
Оптовая и розничная торговля; ремонт авто- транспортных средств, мотоциклов, бытовых изделий и предметов личного пользования	15,9	15,6	14,0	14,4	15,0	12,0	13,4	12,4

Гостиницы и рестораны	1,9	4,3	3,4	1,7	6,5	3,2	3,2	2,8
Транспорт и связь	9,3	7,0	5,5	6,0	4,2	7,3	6,4	6,5
Операции с недвижимым имуществом, аренда и предоставление услуг	7,7	11,1	8,9	5,1	12,2	11,1	9,9	12,9
Государственное управление и обеспечение военной безопасности; обязательное социальное обеспечение	7,2	6,8	8,2	6,3	4,5	5,8	9,3	5,7
Образование	9,4	8,5	5,6	7,9	8,6	7,0	6,9	11,1
Здравоохранение и предоставление социальных услуг	7,4	11,5	11,1	6,2	11,9	14,5	11,4	16,2
Предоставление прочих коммунальных, социальных и персональных услуг**	3,2	9,9	9,1	5,4	14,5'	7,6	7,0	7,3
* Включая рыболовство и рыбоводство. ** Включая финансовые услуги.								

Источники: Российский статистический ежегодник. 2005: Стат. сб. — М.: Росстат, 2006. — С. 729.

Таблица 7.19

Вид услуги	Санкт-Петербург	Ленинградская область
Всего	100,0	100,0
В том числе:		
бытовые	6,1	8,8
транспортные	20,8	10,5
связи	23,9	13,8
жилищно-коммунальные и услуги гостиниц	23,4	49,8
системы образования	7,7	5,7
другие	18,1	11,4

Таблица 7.20

Показатель	1985	2004
Численность детей в дошкольных учреждениях	9179,6	4422,6
Педагогические работники – всего	800,2	619,4
В том числе:		
заведующие	73,4	45,3
воспитатели	661,2	440,2
музыкальные руководители	58,3	50,9
учителя-логопеды	6,2	26,6
учителя-дефектологи	1,1	4,6
прочие	0,0	51,8

И с т о ч н и к: Российский статистический ежегодник. 2005: Стат. сб. – М.: Росстат, 2006. – С. 245, 246.

С помощью обобщающих показателей структурных сдвигов охарактеризуйте изменения в структуре педагогических кадров дошкольных учреждений. Какие еще выводы можно сделать об изменениях, произошедших в системе дошкольного образования в России? Напишите аналитическую записку.

П р и м е ч а н и е. Число дошкольных учреждений в 1985 г. – 81,8 тыс., в 2004 г. – 47,2 тыс. Дополнительную информацию можно найти в статистических сборниках.

7. Имеются следующие данные (табл. 7.21 и 7.22).

Таблица 7.21

**Выпуск квалифицированных рабочих учреждениями
начального профессионального образования по профессиям, тыс. чел.**

Показатель	1995	2004
Выпущено квалифицированных рабочих – всего	840,6	708,0
В том числе по профессиям:		
промышленности	183,1	212,6
сельского хозяйства	126,0	81,9
строительства	114,2	85,9
транспорта	110,5	82,7
связи	4,8	4,6
торговли и общественного питания	93,4	107,1
сферы обслуживания	30,3	17,9
общим профессиям для отраслей экономики	167,0	61,9
другим	11,3	53,4

И с т о ч н и к : Российский статистический ежегодник. 2005: Стат. сб. – М.: Росстат, 2006. – С. 257.

Таблица 7.22

**Потребность в рабочих, заявленная организациями в органы
государственной службы занятости, по отраслям экономики, тыс. чел.**

Отрасль экономики	1995	2004
Всего	225,1	576,2
В том числе:		
промышленность	75,4	121,6
сельское хозяйство	9,9	39,8
строительство	39,9	66,3
транспорт и связь	23,0	30,0
торговля и общественное питание	11,4	62,2
жилищно-коммунальное хозяйство, бытовое обслуживание населения	12,1	49,0
другие отрасли	53,4	207,3

И с т о ч н и к : Российский статистический ежегодник. 2005: Стат. сб. – М.: Росстат, 2006. – С. 161.

1. Определите:

а) структуру выпуска квалифицированных рабочих учреждениями начального профессионального образования по профессиям для каждого года;

б) изменение структуры выпуска рабочих по профессиям;

в) структуру потребности в рабочих по отраслям экономики для каждого года;

г) изменение структуры потребности в рабочих по отраслям экономики;

д) насколько структура подготовки рабочих разных профессий соответствовала потребностям в рабочих в соответствующих отраслях в отдельные годы.

2. Проанализируйте полученные результаты.

8. В последние годы одним из популярных индикаторов устойчивого развития стал показатель «экологический след», который определяется как площадь территории, выраженная в глобальных гектарах (гга), необходимая для производства товаров и услуг в размере конечного потребления (человечеством, страной, регионом, отдельным человеком), размещения инфраструктуры, ассимиляции всевозможных отходов и абсорбции выбросов углекислого газа. Уже сегодня экологический след населения отдельных стран существенно превышает биологическую емкость (потенциал) их территорий.

В табл. 7.23 приведены показатели экологического следа и биологической емкости в 22 странах мира.

Таблица 7.23

Страна	Население, млн чел.	Площадь, тыс. кв. км	Экологический след, гга/чел.	Биологическая емкость, гга/чел.
Австралия	19,9	7682	7,6	14,6
Аргентина	36,8	2737	3,0	6,7
Бразилия	176,6	8459	2,4	6,0
Великобритания	59,3	241	5,3	1,6
Германия	82,5	349	4,7	1,7
Египет	67,6	995	1,5	0,8
Индия	1064,4	2973	0,8	0,7
Индонезия	214,7	1812	1,1	1,8
Италия	57,5	294	3,8	1,2

Продолжение

Страна	Население, млн чел.	Площадь, тыс. кв. км	Экологиче- ский след, гга/чел.	Биологиче- ская емкость, гга/ чел.
Канада	31,6	9221	8,8	14,2
Китай	1288,4	9327	1,5	1,0
Корея	47,9	99	3,3	0,7
Мексика	102,3	1909	2,5	1,7
Нидерланды	16,2	34	4,8	0,8
Пакистан	148,4	771	0,6	0,4
Россия	143,4	16889	4,5	4,8
США	290,8	9159	9,7	5,3
Таиланд	62,0	511	1,5	1,4
Филиппины	81,5	298	1,2	0,6
Франция	59,8	550	5,3	2,9
Швеция	9,0	412	6,7	7,3
Япония	127,6	365	4,8	0,7
Весь мир	6272,5	130331	2,3	1,9

Источник: Состояние окружающей среды. 2005. Статистический справочник Всемирного Банка: пер. с англ. — М.: Весь мир, 2005; WWF, Living Planet Report, 2002.

1. Проведите группировку стран по типу напряженности экологической ситуации, выделив три группы стран: с высокой, средней и низкой экологической напряженностью. В качестве группировочного признака используйте показатель экологического дефицита/запаса (экологический дефицит (–) или запас (+) = биологическая емкость – экологический след). К группе со средней экологической напряженностью отнесите страны, в которых абсолютный размер экологического дефицита/запаса не превышает 1,9 гга/чел., что соответствует среднемировому уровню биологической емкости на одного человека.

2. Для каждой группы стран рассчитайте численность населения, в том числе в процентах от общей численности населения обследуемых стран и от численности населения всего мира; площадь территории, в том числе в процентах от общей площади обследуемых стран и от общей площади стран мира; среднюю плотность населения; средний размер экологического следа.

3. Напишите краткие выводы. С чем, на ваш взгляд, связаны высокие значения экологического следа в отдельных странах мира и как можно добиться снижения антропогенного давления на окружающую среду?

9. На основе данных таблиц приложения 1 выполните группировки стран мира для изучения связи между признаками (табл. 7.24).

Таблица 7.24

Вариант	Признак-фактор	Признак-результат
1	Энергопотребление на душу населения, кг нефт. эквивал.	ВНД на душу населения, долл.
2	Энергопотребление на душу населения, кг нефт. эквивал.	Выбросы диоксида углерода на душу населения, м/т
3	ВВП на душу населения, ППС долл. США	Выбросы диоксида углерода на душу населения, м/т
4	Расходы на НИОКР, % от ВВП	ВНД на душу населения, долл.
5	ВВП на душу населения, ППС долл. США	Распространенность абонентов сотовых сетей на 1 тыс. чел.
6	ВНД на душу населения, долл.	Число пользователей сети Интернет на 1 тыс. чел.

По результатам группировки определите:

- а) направление связи;
- б) форму связи;
- в) силу связи;
- г) тесноту связи.

Проанализируйте полученные результаты и напишите выводы.

10. Используя данные табл. П1.1 и П1.2, постройте группировку стран мира по уровню автомобилизации населения, выделив пять групп: низкий, ниже среднего, средний, выше среднего, высокий. Рассчитайте численность населения, проживающего в странах с разным уровнем автомобилизации, в том числе в процентах от общей численности населения мира. Для каждой

группы стран определите средний размер ВНД на душу населения, средний уровень выбросов диоксида углерода на душу населения. По результатам группировки напишите аналитическую записку.

11. В табл. П1.3 приведены данные, характеризующие развитие банковской сферы в Северо-Западном регионе. Разработайте макет комбинационной таблицы, позволяющий отразить распределение банков региона по двум признакам: местоположение (выделите банки Санкт-Петербурга и прочие банки) и рентабельность капитала (ниже средней по совокупности, выше средней по совокупности). Для каждой группы определите следующие показатели:

а) средний размер собственного капитала, средний размер активов, среднюю прибыльность активов;

б) средний темп роста активов, средний темп роста прибыли, средний темп роста собственного капитала.

Заполните таблицу и проанализируйте полученные результаты.

12. Используя данные табл. П1.3, сгруппируйте банки Северо-Западного региона по трем группам в зависимости от изменения рейтинга: положение в рейтинге улучшилось, не изменилось, ухудшилось. Каждую группу охарактеризуйте показателями: темп роста активов, темп роста прибыли, темп роста собственного капитала, прибыльность активов, рентабельность капитала. Какие выводы можно сделать на основании данной группировки? Напишите аналитическую записку.

13. Постройте аналитическую группировку для изучения взаимосвязи двух признаков: размер собственного капитала банка (признак-фактор) и прибыльность активов (признак-результат). В качестве исходных данных используйте данные табл. П1.3. Проведите анализ результатов выполненной группировки.

14. В целях установления зависимости между маркой автомобиля и его страховой суммой провели 5%-ную выборку договоров страхования, заключенных за месяц (табл. 7.25).

Постройте аналитическую группировку, отражающую зависимость страховой суммы автомобиля от марки. Рассчитайте тесноту связи между признаками.

Таблица 7.25

№ п/п	Марка автомобиля	Страховая сумма, тыс. долл.	№ п/п	Марка автомобиля	Страховая сумма, тыс. долл.
1	Opel	7,2	16	Audi	11,8
2	Audi	12,3	17	Volkswagen	6,7
3	Ford	10,5	18	BA3	5,0
4	Volkswagen	8,8	19	Opel	6,0
5	BA3	3,8	20	Ford	12,3
6	Ford	12,0	21	Volkswagen	7,5
7	Opel	5,5	22	Opel	6,1
8	Volkswagen	7,9	23	BA3	4,0
9	BA3	4,2	24	BA3	4,5
10	Ford	8,7	25	Volkswagen	6,7
11	Volkswagen	9,5	26	Opel	5,5
12	Opel	4,9	27	Volkswagen	12,0
13	Opel	7,5	28	Audi	10,1
14	Volkswagen	10,0	29	Audi	11,8
15	Ford	8,0	30	Ford	11,5

15. Известны итоги промежуточного тестирования студентов группы (по 12-балльной шкале) и оценки экзамена по дисциплине «Статистика» (табл. 7.26).

Таблица 7.26

№ п/п	Ф.И.О. студента	Промежуточное тестирование			Экзаменационная оценка
		1	2	3	
1	Авдиенко К.И.	7	7	8	4
2	Александрова А.К.	5	5	4	3
3	Бершев Д.В.	9	9	9	5
4	Бируков В.Л.	10	11	10	4
5	Вишневецкий А.Г.	9	10	10	5
6	Голянова В.А.	2	3	6	3
7	Григорьева Р.Л.	11	8	10	4
8	Давыденко А.Д.	8	12	9	4
9	Донцова Е.Ю.	10	9	10	4

№ п/п	Ф.И.О. студента	Промежуточное тестирование			Экзаменационная оценка
		1	2	3	
10	Ефремова Т.Т.	9	6	4	3
11	Захаров С.Я.	10	11	12	5
12	Королев С.М.	8	9	12	5
13	Колупаев А.Н.	7	5	7	2
14	Кувшинова С.Б.	2	3	4	2
15	Лопатина А.Н.	8	9	6	4
16	Левченко Н.Е.	12	12	11	5
17	Лукина И.О.	11	9	9	5
18	Матвиенко О.З.	6	6	5	3
19	Мурашко Д.Г.	11	11	12	5
20	Павлюк Н.Ю.	8	10	5	4
21	Павлов Е.А.	8	10	9	4
22	Раевский Н.Н.	9	11	12	5
23	Романова О.П.	9	7	7	3
24	Рустамов К.Р.	8	9	10	4
25	Симакова Т.Ю.	5	4	7	4
26	Суслов Е.А.	5	6	5	3
27	Чирикин Е.Н.	8	9	7	3
28	Шапошников К.Г.	7	10	6	5
29	Яковлева М.Ю.	5	10	8	4
30	Ярошенко Д.И.	4	8	3	2

Для каждого студента определите средний балл промежуточного тестирования и проведите группировку студентов по уровню текущей успеваемости, выделив четыре группы студентов (средний балл промежуточного тестирования «до 6,0», «6,0 – 8,0», «8,0 – 10,0», «10,0 и более»). Для каждой группы рассчитайте средний балл экзаменационной оценки. К какому виду относится построенная вами группировка? Какие характеристики связи можно определить на ее основе?

Дайте характеристику направлению и формы связи между текущей успеваемостью и итоговой экзаменационной оценкой. Вычислите показатели силы и тесноты связи. Напишите краткие выводы.

7.3. Вопросы для самоконтроля

1. В чем состоит сущность метода группировки?
2. При соблюдении каких правил группировки сводные показатели для отдельных групп будут типичными и устойчивыми?
3. Какие группировки называют простыми? сложными?
4. В чем основной недостаток комбинационных группировок?
5. Поясните понятия «специализация группировочного признака», «специализация интервалов группировочного признака». Приведите примеры.
6. Как проверить правильность проведения группировки? В каких случаях схема группировки должна быть пересмотрена?
7. На какие виды принято подразделять группировки в зависимости от характера решаемых задач?
8. С помощью каких показателей изучается динамика структуры совокупности? Напишите их формулы.
9. Какие характеристики связи могут быть определены на основе аналитической группировки?
10. В чем заключается сущность метода стандартизации распределений в комбинационной таблице?
11. В чем состоит правило сложения дисперсий? Приведите формулы расчета межгрупповой, внутригрупповой и общей дисперсий.
12. В каких пределах принимает значения коэффициент детерминации? эмпирическое корреляционное отношение?
13. Назовите методы многомерной классификации.

7.4. Методические рекомендации преподавателям

Аудиторная работа. Поскольку построение группировок процесс достаточно трудоемкий, при выборе заданий для аудиторной работы необходимо прежде всего обращать внимание на методическое обеспечение работы студентов. Целесообразно заранее подготовить базы данных с первичной информацией в формате Excel. При отсутствии компьютерного обеспечения для построения типологических и аналитических группировок на занятиях целесообразно выбирать задачи, содержащие не более 20–30 единиц совокупности. Например, задачи на группировки

стран мира (задания 9, 10), можно выполнять по отдельным частным совокупностям (для стран отдельных регионов или наиболее крупным по численности населения). Задачи на анализ структурных группировок (задания 3, 4, 7) также можно разбить на варианты с последующим обсуждением и обобщением итогов.

Самостоятельная внеаудиторная работа может включать задания по типу заданий 9, 13 (аналитическая группировка), задания 7 (структурная группировка).

Итоговый контроль по теме может быть проведен в форме тестирования. Если студенты выполняли самостоятельную внеаудиторную работу, целесообразно организовать их защиту и обсуждение.

7.5. Методические указания студентам

Освоение теории. Группировка лежит в основе работы с собранной информацией. На основе группировки рассчитываются средние показатели по группам, появляется возможность их сравнения, анализа причин различий между группами, изучения взаимосвязей между признаками. При проведении группировок необходимо руководствоваться следующими правилами:

- состав групп должен быть качественно однородным, а различия между группами существенными;
- группы должны содержать достаточное число единиц, чтобы проявились типичные черты и закономерности, свойственные изучаемым явлениям. Выделенные группы не должны содержать по одной единице. Считается, что в группе должно быть не менее 5 единиц;
- нельзя выделять лишь две группы, так как между двумя группами обязательно будут различия, которые трудно объяснить. Минимальное количество групп должно быть равно трем.

В результате изучения темы вы должны:

- знать понятие группировки и понимать ее роль в анализе экономических и социальных явлений;

- знать классификацию группировок по видам.

Практические навыки. Необходимо научиться:

- выполнять типологические, структурные и аналитические группировки и правильно оформлять их в табличной форме;
- оценивать структурные изменения и различия;
- находить характеристики связи на основе аналитической группировки.

Выборочное наблюдение

8.1. Основные положения

Статистика подразделяется на описательную (дискриптивную) и выводную (индуктивную).

Задача описательной статистики — получение обобщающих характеристик изучаемой совокупности.

Задача выводной статистики — получение обобщающих характеристик по выборке в целях их последующего распространения на генеральную совокупность. **Генеральной** называется совокупность, из которой производится выборка.

Выводная статистика часто называется аналитической и базируется на вероятностном подходе.

Выборка из генеральной совокупности должна быть *репрезентативной*, т.е. адекватно представлять свойства генеральной совокупности. Репрезентативность выборки обеспечивается при условии объективности отбора единиц в выборку.

Отбор может быть случайный или механический (с определенным шагом отбора).

$$\text{Шаг отбора} = N : n,$$

где N — объем генеральной совокупности;

n — запланированный объем выборки.

Отбор единиц в выборку может быть повторным или бесповторным. Схема повторного отбора предполагает, что единица, однажды попавшая в выборку, участвует в дальнейшей процедуре отбора и имеет шанс повторно попасть в выборку. Как правило, пользуются схемой бесповторного отбора, т.е. единица имеет шанс только раз попасть в выборку.

Основные виды выборочного наблюдения различаются: а) единицей отбора; б) применением стратификации генеральной совокупности.

Основные виды выборки:

- 1) нестратифицированная простая выборка;
- 2) нестратифицированная серийная выборка;
- 3) стратифицированная простая выборка;
- 4) стратифицированная серийная выборка.

В первом и третьем случаях единица отбора совпадает с единицей наблюдения; во втором и четвертом — единица отбора представляет собой группу единиц наблюдения. В первом и втором случаях отбор производится из генеральной совокупности в целом; в третьем и четвертом — из предварительно выделенных страт (иногда используют термины: типы, районы, слои. Соответственно выборка будет называться типическая или районированная, или расслоенная).

По объему различают выборку большую (свыше 30 ед.) и малую (до 30 ед.). Генеральные характеристики называются *параметрами* и обозначаются греческими буквами. Например, μ — генеральная средняя, σ^2 — генеральная дисперсия, π — генеральная доля (относительная величина), ρ — генеральный коэффициент корреляции.

Выборочные характеристики называются *статистиками* (или *оценками параметров*) и обозначаются латинскими буквами. Например, \bar{x} — выборочная средняя, w — выборочная доля, s^2 — выборочная дисперсия, r — выборочный коэффициент корреляции. Применение выборочного метода связано с решением четырех типов задач:

- 1) расчет ошибки выборки и получение точечной и интервальной оценок генеральных параметров;
- 2) определение объема выборки, необходимого для обеспечения требуемой степени точности с заданной доверительной вероятностью;
- 3) определение вероятности того, что ошибка выборки не превысит допустимого предела;
- 4) сравнение двух выборочных характеристик в целях оценки статистической значимости их различий.

Решение задач первого, третьего и четвертого типов предполагает, что выборка проведена, тогда как задача второго типа решается до проведения выборки.

Ошибка выборки, или ошибка репрезентативности, — это разница между выборочной оценкой и генеральным параметром. Генеральные параметры не известны, но известно, что значения ошибки выборки распределяются по нормальному закону и образуют кривую выборочного распределения. Даже для ненормально распределенной генеральной совокупности в случае большой выборки ($n \geq 30$ ед.) выборочное распределение выборочных средних будет приблизительно нормальным.

Ошибка выборки представляется как центрированное и нормированное отклонение выборочной оценки от генерального параметра, т.е. как z -статистика:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}},$$

где $\sigma_{\bar{x}}$ — среднеквадратическое отклонение выборочной средней от генеральной средней (средняя ошибка выборочной средней).

Если формирование выборки производится случайно, то математическое ожидание выборочной средней будет не что иное, как генеральная средняя:

$$E(\bar{x}) = \mu,$$

т.е. средняя, полученная по выборочному распределению, будет *несмещенной* оценкой генеральной средней. Для нормально распределенной генеральной совокупности стандартная ошибка распределения выборочных средних определяется по формуле

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n} \left(\frac{N-n}{N-1} \right)}, \quad (8.1)$$

или

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N} \right)}$$

где σ_x^2 — генеральная дисперсия признака x .

Если генеральная совокупность велика по сравнению с выборкой ($n/N \leq 0,05$, или 5%), то

$$\sqrt{\frac{N-n}{N-1}} \approx 1.$$

Тогда стандартная ошибка выборочных средних определяется как:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_x^2}{n}} = \frac{\sigma_x}{\sqrt{n}}. \quad (8.2)$$

Стандартная ошибка выборочных средних — это *средняя ошибка выборки*. Формула (8.1) соответствует средней ошибке выборки для бесповторного отбора, формула (8.2) — для повторного отбора.

Интервальная оценка параметра предполагает расчет *предельной ошибки выборки*, которая связана с заданной доверительной вероятностью:

$$\Delta_{\bar{x}} = z \cdot \sigma_{\bar{x}}, \quad (8.3)$$

где z — коэффициент, связывающий стандартную ошибку выборочных средних с принятой доверительной вероятностью.

Значение z определяется по таблице нормального распределения Лапласа—Гаусса в зависимости от доверительной вероятности $F(z)$. Так, доверительной вероятности $F(z) = 0,95$ соответствует значение $z = 1,96$; $F(z) = 0,682$ соответствует $z = 1$; если $F(z) = 0,997$, то $z = 3$ и т.д.

Ничего не меняется, если оценивается генеральная доля, — только берется дисперсия альтернативного признака.

Средняя ошибка выборочной доли:

$$\sigma_w = \sqrt{\frac{\pi(1-\pi)}{n} \cdot \left(1 - \frac{n}{N}\right)} \text{ — для бесповторного отбора;} \quad (8.4)$$

$$\sigma_w = \sqrt{\frac{\pi(1-\pi)}{n}} \text{ — для повторного отбора.} \quad (8.5)$$

Предельная ошибка выборочной доли:

$$\Delta_w = z \cdot \sigma_p. \quad (8.6)$$

Поскольку генеральные дисперсии не известны, используют их выборочные оценки. Тогда средняя ошибка выборки (при бесповторном отборе):

- для выборочной средней

$$\sigma_{\bar{x}} = \sqrt{\frac{s_x^2}{n} \cdot \left(1 - \frac{n}{N}\right)} \quad (8.7)$$

- для выборочной доли

$$\sigma_w = \sqrt{\frac{p(1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)} \quad (8.8)$$

Пример 8.1. В результате выборочного обследования населения по проблемам занятости получено распределение тех, кто не имел работы и искал ее, пребывая в статусе безработного (табл. 8.1).

Таблица 8.1

Распределение безработных в выборке

Число месяцев безработицы	0–3	3–6	6–9	9–12	12 и более
Число безработных	20	50	15	10	5

Требуется с вероятностью 0,95 определить: а) среднюю продолжительность пребывания в статусе безработного; б) долю лиц, являющихся безработными один год и более.

Решение. Рассчитаем выборочную среднюю и дисперсию продолжительности безработицы (используем середины интервалов):

$$\begin{aligned} \bar{x} &= \frac{1,5 \cdot 20 + 4,5 \cdot 50 + 7,5 \cdot 15 + 10,5 \cdot 10 + 13,5 \cdot 5}{100} = \frac{367,5}{100} \approx 3,7 \text{ месяца;} \\ s^2 &= \frac{(1,5 - 3,7)^2 \cdot 20 + (4,5 - 3,7)^2 \cdot 50 + (7,5 - 3,7)^2 \cdot 15 + (10,5 - 3,7)^2 \cdot 10 + (13,5 - 3,7)^2 \cdot 5}{100} = \\ &= \frac{1288}{100} \approx 12,88. \end{aligned}$$

Значит, среднее отклонение от средней в выборке составляет $\sqrt{12,88} = \pm 3,6$ месяца.

Средняя ошибка выборки:

$$\sigma_{\bar{x}} = \frac{3,6}{\sqrt{100}} = \pm 0,36 \text{ месяца.}$$

Предельная ошибка выборки (с вероятностью 0,95):

$$\Delta_{\bar{x}} = 1,96 \cdot 0,36 = \pm 0,7 \text{ месяца.}$$

Тогда интервальная оценка генеральной средней будет равна

$$3,7 - 0,7 \leq \mu \leq 3,7 + 0,7;$$

$$3 \text{ месяца} \leq \mu \leq 4,4 \text{ месяца,}$$

т.е. с вероятностью 0,95 можно ожидать, что в генеральной совокупности (население в возрасте 15–72 лет) средняя продолжительность пребывания лиц в статусе безработного будет не менее 3 месяцев и не более 4,4 месяца.

Средняя ошибка доли лиц, не имеющих работы и ищущих ее 12 месяцев и более, составит:

$$\sigma_p = \sqrt{\frac{0,05 \cdot 0,95}{100}} = \frac{0,218}{10} = 0,0218.$$

Предельная ошибка доли составит (с вероятностью 0,95):

$$\Delta_p = 1,96 \cdot 0,0218 = \pm 0,043.$$

Отсюда интервальная оценка генеральной доли:

$$5\% - 4,3\% \leq \pi \leq 5\% + 4,3\%;$$

$$0,7\% \leq \pi \leq 9,3\%.$$

Очевидно, что полученный результат малоинформативен и дает слишком неопределенную оценку генеральной доли лиц, находящихся в статусе безработного один год и более.

В случае малой выборки распределение выборочных средних подчиняется распределению Стьюдента. Тогда вместо z в формулах фигурирует t -статистика. Значения статистик z и t отличаются для одних и тех же вероятностей (табл. 8.2).

Таблица 8.2

Значения z - и t -статистик

Вероятность	0,9	0,95	0,99
z -статистика	1,65	1,96	2,58
t -статистика ($df = 20$)	1,725	2,086	2,845

Распределение t -статистики зависит от числа степеней свободы (df). В табл. 8.2 приведены значения t -статистики при 20 степенях свободы. Видим, что при одних и тех же значениях доверительной вероятности значения t -статистики больше значений z -статистики. Это приводит к большей неопределенности интервальной оценки генерального параметра на основе малой выборки.

Применение других видов выборочного наблюдения сказывается на величине дисперсии в формуле (8.2) средней ошибки выборки. Оценки дисперсии основаны на правиле сложения дисперсий:

$$\sigma_{\text{общ}}^2 = \bar{\sigma}_{\text{внутригрупп}}^2 + \sigma_{\text{межгрупп}}^2.$$

Если выборка стратифицирована, то в формуле (8.2) берется средняя из внутригрупповых дисперсий; если серийная, то берется межсерийная дисперсия $\sigma_{\bar{y}_x}^2$. Кроме того, в формуле (8.2) при серийной выборке вместо n показывается число отобранных серий, r . Корректировка на бесповторность типа отбора производится путем введения сомножителя в подкоренное выражение формулы (8.2):

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_{\bar{y}_x}^2}{r} \cdot \left(1 - \frac{r}{R}\right)}, \quad (8.9)$$

где R — число серий в генеральной совокупности;
 r — число серий в выборке.

Так как выбор доверительной вероятности, или надежности доверительной оценки, допускает некоторый произвол, широкое распространение в практике получило правило «трех сигм», которое состоит в следующем.

Отклонение истинного значения изучаемой величины среднего арифметического значения из наблюдаемых величин не

превосходит трехкратной средней квадратической ошибки этого среднего значения.

Можно записать это правило, как

$$|a - \bar{x}| < 3\sigma / \sqrt{n}$$

или, если σ не известна, а имеется лишь ее выборочная оценка s , то

$$|a - \bar{x}| < 3s / \sqrt{n}.$$

Первая из этих оценок имеет надежность (независимо от числа наблюдений):

$$2\Phi(z = 3) = 0,9973 \approx 1 - 0,003.$$

Надежность второй оценки зависит от числа наблюдений, n .

Задача второго типа решается на стадии проектирования выборки.

Из формул (8.2) и (8.3) получим формулу для расчета требуемого объема выборки:

$$n = \frac{z^2 \sigma^2}{\Delta^2}. \quad (8.10)$$

Величина генеральной дисперсии, как правило, не известна. В этом случае ее определяют либо на основе пилотного обследования, либо на основе данных аналогичного обследования в прошлые годы, либо условно на основе свойств нормального распределения:

$$\sigma \cong \frac{1}{6}(x_{\max} - x_{\min}),$$

при асимметричном распределении

$$\sigma \cong \frac{1}{5}(x_{\max} - x_{\min})$$

или

$$\sigma \cong \frac{1}{3}\bar{x}.$$

В качестве оценки дисперсии доли, если нет никаких данных, принимают максимально возможную величину дисперсии:

$$\sigma^2 = \pi(1 - \pi) = 0,5 \cdot 0,5 = 0,25.$$

Значение предельно допустимой ошибки (Δ) и доверительной вероятности и соответственно величины z задаются при планировании выборки. При этом принимаются во внимание последствия ошибки выборки, прежде всего то, насколько они могут сказаться на жизни и здоровье людей. Нередко предельная ошибка выборки задается в относительной форме (как отношение к среднему значению). В этом случае дисперсия тоже берется по отношению к средней, т.е. в виде коэффициента вариации ν .

Пример 8.2. Фирма-оператор сотовой связи разрабатывает новый тарифный план, для чего необходимо определить среднюю продолжительность одного звонка внутри сети и долю звонков длительностью более 2 мин. Какое число соединений необходимо охватить выборочным наблюдением, чтобы предельная ошибка выборки не превышала 5% средней продолжительности одного звонка и 5% при определении доли звонков длительностью более 2 мин?

По данным прошлого обследования известно, что коэффициент вариации продолжительности разговоров внутри сети составляет 60%. Результаты гарантировать с вероятностью 0,954.

Решение.

$$n = \frac{z^2 \nu^2}{\Delta_{\text{отн}}^2}; \quad (8.11)$$

$$n = \frac{2^2 \cdot 0,6^2}{0,05^2} = \frac{1,44}{0,0025} = 576 \text{ соединений};$$

$$n = \frac{z^2 \sigma^2}{\Delta^2} = \frac{2^2 \cdot 0,25}{0,0025} = 400 \text{ соединений}.$$

Ответ: 576 соединений (принимается больший из рассчитанных объем выборки).

В данном случае общее число соединений — генеральная совокупность N , бесконечно большого объема. Если же объем N ограничен и отношение $(n : N) \geq 0,05$, т.е. объем выборки равен или превосходит 5% объема генеральной совокупности, то объем выборки рассчитывается по следующей формуле:

$$n = \frac{z^2 \sigma^2 N}{z^2 \sigma^2 + \Delta^2 N}. \quad (8.12)$$

Пример 8.3. Для определения средней цены говядины в розничной торговле крупного города предполагается организовать выборочную регистрацию цен. Известно, что цены на говядину колеблются от 80 до 250 руб. за 1 кг. Сколько торговых точек необходимо обследовать, чтобы с вероятностью 0,954 ошибка выборки при определении средней цены говядины не превышала 2 руб. за 1 кг. Известно, что в городе продажу мяса осуществляют 5000 торговых точек.

Решение.

Предположим, что распределение цен соответствует нормальному распределению, тогда

$$\sigma = \frac{1}{6}(x_{\max} - x_{\min}) = \frac{1}{6}(250 - 80) \approx 28 \text{ руб. } 33 \text{ коп.}$$

Вероятности $\Phi(z) = 0,954$ соответствует значение $z = 2$.

Рассчитаем численность выборки по формуле повторного отбора (без учета численности генеральной совокупности):

$$n = \frac{2^2 \cdot 28,33^2}{2^2} = 802,6 \cong 803 \text{ торговые точки.}$$

Поскольку доля планируемой выборки в генеральной совокупности составляет 16% ($803 : 5000$), т.е. больше 5%, переходим к расчету объема выборки по формуле (8.12) для бесповторного отбора:

$$n = \frac{2^2 \cdot 28,33^2 \cdot 5000}{2^2 \cdot 28,33^2 + 2^2 \cdot 5000} = \frac{16052000}{802,6 + 20000} \approx 772 \text{ торговые точки.}$$

Учет объема генеральной совокупности привел к несколько меньшей численности выборки, достаточной, чтобы обеспечить желаемую точность результатов с принятой вероятностью.

Решение задач третьего типа возникает в том случае, когда рассчитанная предельная ошибка выборки оказалась большей, чем принятая величина допустимой предельной ошибки.

Из формулы (8.3) находим величину z :

$$z = \frac{\Delta_{\bar{x}}}{\sigma_{\bar{x}}}. \quad (8.13)$$

Затем по таблице нормального распределения по величине z определяется вероятность $F(z)$.

Решение задач четвертого типа связано со сравнением результатов разных выборок. Детальнее эта задача будет рассмотрена в теме 9.

Пример 8.4. Опрос служащих на предприятиях пищевой промышленности показал, что среднемесячная заработная плата бухгалтеров на предприятиях данной отрасли варьирует от 15852 руб. в Краснодарском крае до 49675 руб. в Магаданской области. Средняя заработная плата в Москве бухгалтеров этой же отрасли составила 39020 руб., а в Санкт-Петербурге — 37183 руб. Проведена случайная выборка 40 бухгалтеров в Москве и 40 бухгалтеров в Санкт-Петербурге. Чему равна вероятность того, что выборочная средняя \bar{x}_1 в Москве превысит выборочную среднюю \bar{x}_2 в Санкт-Петербурге на 2500 руб. и больше при условии, что СКО в обоих генеральных распределениях можно считать приблизительно равными, т. е. $\sigma_1 = \sigma_2 = 5000$ руб.

Решение.

Разность выборочных средних составляет

$$\bar{x}_1 - \bar{x}_2 = 39020 \text{ руб.} - 37183 \text{ руб.} = 1837 \text{ руб.}$$

Средняя ошибка выборки разности двух выборочных средних определяется по формуле

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_{n_1}^2}{n_1} + \frac{\sigma_{n_2}^2}{n_2}}; \quad (8.14)$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{5000^2}{40} + \frac{5000^2}{40}} = 1118,03 \text{ руб.}$$

Распределение заработной платы в генеральной совокупности, выборочные средние и разности выборочных средних распределены нормально. Вычислим z -значение, соответствующее 2500 руб.:

$$z = \frac{2500 - 1837}{1118,03} = 0,59.$$

На рис. 8.1 площадь, соответствующая искомой вероятности, заштрихована.

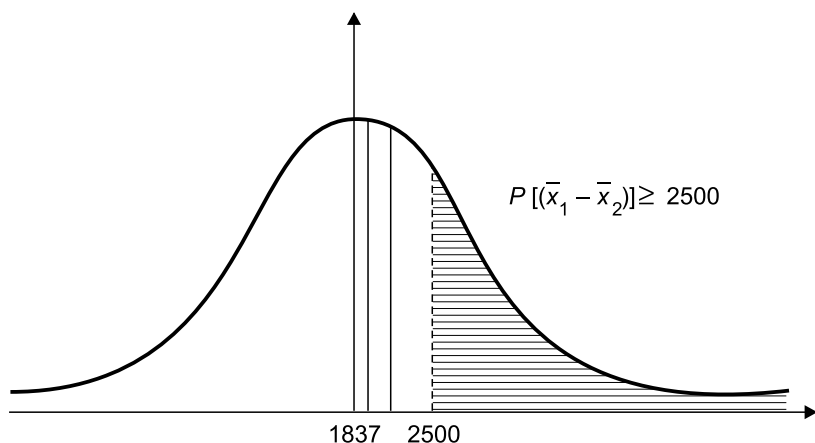


Рис. 8.1. Выборочное распределение разности выборочных средних

По таблице нормального распределения $\Phi(0,59) = 0,4448$. Половина этой площади кривой нормального распределения заштрихована прямой штриховкой. Тогда площадь, соответствующая искомой вероятности, равна:

$$P[(\bar{x}_1 - \bar{x}_2) \geq 2500] = 0,5 - \frac{1}{2} \cdot 0,4448 = 0,2776$$

где 0,5 – это половина площади интеграла вероятностей.

Таким образом, вероятность того, что средняя заработная плата бухгалтеров в Москве превышает среднюю заработную плату бухгалтеров в той же отрасли в Санкт-Петербурге на сумму 2500 руб. и более, составляет примерно 0,28.

8.2. Задания по теме

8.2.1. Расчет ошибок выборочных показателей

1. Автомат разливает жидкий шоколад в формы для получения шоколадных плиток. Обычно вес шоколада в формах соответствовал нормальному распределению со средней 100 г и со стандартным отклонением 2,5 г. В порядке контроля качества была взята случайная выборка из готовой продукции 15 плиток и произведено их взвешивание. Средний вес плитки в выборке оказался 99,5 г.

Определите доверительный интервал с вероятностью 95% для среднего значения веса шоколада, распределенного по формам.

2. Случайная выборка 800 домохозяек в центре города и их опрос, проведенный утром, показали, что 450 из них хотели, чтобы торговый центр города был пешеходной зоной.

Определите доверительные пределы доли всех домохозяек, кто бы придерживался того же мнения (с вероятностью 90%).

3. При обследовании бюджета времени студентов были учтены затраты времени на обед в столовой университета. Обследовано 60 студентов, отобранных в порядке случайной выборки. Из них: 5 студентов затратили на обед 21 мин, 10 студентов — 27, 30 студентов — 33, 12 студентов — 39, 3 студента — 45 мин.

Определить средние затраты времени на обед студентами университета (с вероятностью 0,97).

4. Абсолютная предельная ошибка выборки при определении среднего размера диаметра деталей составила 0,75 мм, что в процентах к средней составляет 10%-ную относительную ошибку.

Определить, в каких пределах находится средний размер диаметра деталей в генеральной совокупности.

5. На 100 предприятиях, отобранных в порядке механического отбора, обследованы затраты труда на производство одного из изделий. Было установлено, что средние затраты труда на это изделие составляют 120 чел.-ч при среднем квадратическом отклонении, равном 15 чел.-ч.

Определите с вероятностью 0,954 пределы средних затрат труда на одно изделие по всем предприятиям.

6. Из партии деталей в 4800 шт. отобрано методом случайного бесповторного отбора 800 деталей. Результаты выборки оказались следующие: средний диаметр отобранных деталей — 200 мм; среднее квадратическое отклонение — 7 мм.

Определите с вероятностью 0,954 границы, в которых можно ожидать средний диаметр детали во всей партии.

7. На ткацкой фабрике работает 400 ткачих. В порядке случайной, повторной выборки определена средняя дневная выработка 100 ткачих. В итоге этого обследования получены следующие данные (табл. 8.3).

Таблица 8.3

Дневная выработка, м	35–45	45–55	55–65	65–75
Количество ткачих	20	40	25	15

Определите с вероятностью 0,954 пределы значений дневной выработки для 400 ткачих.

8. При выборочном обследовании 2% изделий партии готовой продукции получены следующие данные о содержании влаги в образцах (табл. 8.4).

Таблица 8.4

Влажность, %	Число образцов, ед.
43–44	9
44–45	15
45–46	40
46–47	25
47–48	11
Итого	100

При условии, что к продукции высшего сорта отнесено 80 образцов, установить для всей партии:

а) с вероятностью 0,997 возможные пределы среднего процента влажности;

б) с вероятностью 0,954 возможные пределы процента продукции высшего сорта.

9. На металлургических заводах корпорации работают 100 формовщиков и 200 литейщиков.

Проведено районированное изучение среднемесячной заработной платы 60 рабочих с пропорциональным распределением их по специальностям в соответствии с удельным весом в общем количестве рабочих. Внутри каждой группы применялся метод повторного отбора. В результате получены следующие данные (табл. 8.5).

Таблица 8.5

Показатель	Формовщики	Литейщики
Месячная заработная плата, руб.	38000	41100
Среднее квадратическое отклонение по заработной плате, руб.	540	880

Определите предел ошибки репрезентативности при установлении средней заработной платы, гарантировав результат с вероятностью 0,954.

10. В АОЗТ имеется 500 коров. В порядке случайной повторной выборки в течение недели определяли средний дневной удой на одну корову. Обследовали 100 коров, которые по удою распределились следующим образом (табл. 8.6).

Таблица 8.6

Дневная удойность, кг	8–10	10–12	12–14	14–16
Количество коров	20	40	35	5

Определите среднюю ошибку репрезентативности при установлении дневной удойности коров.

11. Генеральная совокупность состоит из 60000 единиц, разбитых на 500 равных по объему серий. Для определения доли единиц, обладающих изучаемым признаком, произведена двухступенчатая комбинированная бесповторная выборка 50% серий и из каждой серии отобрано по 20% единиц. Общая дисперсия альтернативного признака оказалась равной 0,21, а межсерийная дисперсия – 0,09.

Определите с вероятностью 0,954 предельную ошибку выборки для доли.

12. Для определения доли женского труда в одной из отраслей обрабатывающей промышленности все предприятия отрасли были разбиты на три группы: крупные, средние и мелкие. По численности рабочих эти группы распределились следующим образом: крупные — 50%, средние — 40, мелкие — 10% от общего числа рабочих всей отрасли. В порядке бесповторной типической выборки был произведен отбор 10000 человек с пропорциональным распределением по группам, в результате которого доля мужчин оказалась равной в крупных предприятиях 60%, в средних — 50, в мелких — 45%.

Определите с вероятностью 0,997, в каких пределах заключена доля женщин среди всех рабочих данной отрасли.

13. На предприятии с числом работающих 2500 человек провели выборочный опрос о занятиях спортом в выходные дни. Из 500 опрошенных 100 человек регулярно занимаются спортом в выходные дни.

Какова численность работников предприятия, занимающихся спортом? Ответ дайте с вероятностью 0,90.

14. В случайной выборке 100 аспирантов-экономистов университетов Санкт-Петербурга 60 оказались уроженцами Санкт-Петербурга.

С вероятностью 0,950 определите долю петербуржцев среди аспирантов-экономистов.

15. При обследовании жилищ в городе был произведен механический отбор по списку всех домов с долей отбора, равной $1/50$. Из общего числа домов в выборке, равного 8491, нуждались в ремонте 627 домов.

Сколько домов в городе нуждается в ремонте? Ответ дайте с вероятностью 0,997.

16. Менеджеру универсама необходима информация о среднедневной потребности в кефире. Случайная выборка дала следующие результаты продаж (число проданных литровых тетрапакетов в день): 48, 59, 45, 62, 50, 68, 57, 80, 65, 58, 79, 69. Предположив, что это случайная выборка дневной потребности, постройте 90%-ный доверительный интервал для среднего числа пакетов с кефиром, которое надо завозить ежедневно в универсам.

17. Туристическая фирма, осваивающая новый вид экскурсионного обслуживания, провела опрос 120 случайно отобран-

ных потенциальных клиентов для выяснения того, интересен ли им новый вид услуг или нет. Результаты показали, что 28% опрошенных предпочли бы новый вид экскурсионного обслуживания старому. Постройте 95%-ный доверительный интервал для доли клиентов туристической компании, которые станут потребителями нового вида услуг.

18. На основе случайной выборки 500 постоянных курильщиков установлено, что они тратят в среднем на сигареты 310 руб. в неделю при стандартном отклонении $\pm 77,5$ руб. Укажите: а) точечную оценку недельных затрат на сигареты; б) доверительные границы недельных затрат на сигареты с вероятностью 95%.

8.2.2. Определение объема выборки

19. На заводах объединения выборочным методом был определен средний стаж работы 600 рабочих. Оказалось, что он равен 10 годам со средним квадратическим отклонением 3 года.

Сколько надо обследовать рабочих, чтобы ошибка при определении среднего стажа не превысила 1 год с вероятностью 0,90?

20. Имеются основания предполагать, что коэффициент вариации исследуемого признака равен 20%.

Определите объем собственно случайной повторной выборки, чтобы с вероятностью 0,954 можно было утверждать, что относительная ошибка выборки не превысит 2%.

21. Менеджер молодежного бара решил выяснить расходы посетителей на баночное пиво. На основе своего опыта он принял максимальную цену одной банки пива равной 70 руб., минимальную – 22 руб., среднее квадратическое отклонение от средних величины расхода равным ± 8 руб. Менеджер хочет быть уверенным в своих выводах на 90% с предельной ошибкой 5%. Сколько посетителей бара необходимо опросить?

22. Планируется провести выборочное обследование самозанятых торговцев в целях выявления продолжительности их пребывания в разъездах в течение года. Небольшое пилотное обследование показало, что они тратят на поездки, связанные с закупкой и реализацией товара, 150 дней в году при стандартном отклонении 14 дней. Если необходимо дать оценку гене-

ральной средней с ошибкой не более двух дней и с вероятностью 90%, то сколько торговцев необходимо опросить?

23. Из генеральной совокупности оплаченных фирмой за год 681 счета аудитору необходимо выбрать 10 счетов, имеющих трехзначную нумерацию. Перечислите, какие счета попадут в выборку, если аудитор решил использовать для отбора таблицу случайных чисел, начиная со строки 6 столбца 2 (см. приложение 2 — табл. П2.7).

8.2.3. Комплексные задачи

24. В порядке бесповторной 10%-ной пропорциональной выборки отобрано 100 единиц, и после их обследования получены следующие показатели (табл. 8.7).

Таблица 8.7

Типическая группа	Число единиц	Среднее значение	Доля признака	Дисперсия количественного признака
I	10	45	0,34	49
II	20	53	0,41	64
III	30	36	0,83	36
IV	40	69	0,10	81

Определите:

а) абсолютную ошибку типической бесповторной выборки для средней и для доли, гарантируя результаты с вероятностью 0,997;

б) какова была бы абсолютная ошибка выборки (с вероятностью 0,997) для средней и для доли, если бы была организована собственно случайная бесповторная выборка в 150 единиц;

в) сколько единиц нужно отобрать, чтобы с вероятностью 0,954 можно было утверждать, что абсолютная ошибка выборки для средней не превысит 0,5, если организована:

- типическая бесповоротная выборка;
- собственно случайная (простая) бесповторная выборка;

г) сколько единиц нужно отобрать из каждой типической группы при непропорциональном и оптимальном вариантах размещения общего объема выборки.

25. Генеральная совокупность состоит из 80 равновеликих по объему серий. В порядке бесповторной серийной выборки отобрано 10 серий. Получены следующие значения средней величины и доли в каждой из отобранных серий:

\bar{x}_j 45; 52; 56; 46; 52; 50; 51; 55; 45;

w_j 0,14; 0,12; 0,08; 0,11; 0,09; 0,10; 0,13; 0,11; 0,12; 0,10.

Определите:

а) абсолютную ошибку и относительную ошибку данной выборки для средней и для доли исследуемого признака, гарантируя результаты с вероятностью 0,954;

б) в каких пределах находится средняя и в каких пределах находится доля изучаемого признака в генеральной совокупности;

в) вероятность, с которой можно утверждать, что абсолютная ошибка данной выборки:

- для средней не превысит 2,5;
- для доли не превысит 0,18;

г) сколько серий при бесповторном отборе нужно отобрать, чтобы с вероятностью 0,997 можно было утверждать, что абсолютная ошибка выборки:

- для средней не превысит 3,0;
- для доли не превысит 0,025.

26. На машиностроительных заводах области заняты 3000 рабочих по четырем профессиям: формовщиков — 350, литейщиков — 700, штамповщиков — 950, токарей — 1000 человек. В порядке типической бесповторной выборки обследовано 5% рабочих с пропорциональным распределением их по указанным группам. На основе этого обследования получены следующие данные (табл. 8.8).

Таблица 8.8

Показатель	Формовщики	Литейщики	Штамповщики	Токари
Месячная заработная плата, руб.	28500	35600	24000	26500
Среднее квадратическое отклонение по заработной плате, руб.	380	460	650	980
Удельный вес рабочих со стажем работы свыше 10 лет, %	50	40	70	30

Определите:

а) абсолютную ошибку выборки при определении средней заработной платы и при определении доли рабочих со стажем работы свыше 10 лет, гарантируя результаты с вероятностью 0,954;

б) в каких пределах находятся средняя заработная плата всех рабочих машиностроительных заводов области и удельный вес рабочих со стажем свыше 10 лет;

в) сколько рабочих необходимо обследовать, чтобы определение ошибки (для средней и для доли) уменьшилось на 25%, гарантируя результаты с вероятностью 0,997.

27. На территории области в личной собственности граждан находится всего 3000 коров. В порядке собственно случайной бесповторной выборки обследовали 300 коров и получили следующие сведения: средний годовой удой 3200 кг, коэффициент вариации годового удоя 30%, удельный вес породистых коров в выборке 85%. Определите:

а) абсолютную ошибку выборки при определении среднего годового удоя и при определении доли породистых коров, гарантируя результаты выборки с вероятностью 0,954;

б) относительную ошибку выборки, предполагая, что способ отбора коров был повторным;

в) вероятность (при бесповторном отборе) того, что ошибка выборки при определении среднего удоя не превысит 12 кг;

г) сколько коров (при бесповторном отборе) нужно обследовать, чтобы абсолютная ошибка выборки при определении среднего удоя не превысила 15 кг и при определении доли породистых коров не превысила 0,02, гарантируя результаты выборки с вероятностью 0,954.

28. Организована двухфазная механическая выборка. Из генеральной совокупности в 10000 единиц на первой фазе отобрано 1000 единиц. Установлено, что дисперсия исследованного на первой фазе признака равна 500.

Определите:

а) абсолютную предельную ошибку на первой фазе выборки, гарантируя результат с вероятностью 0,997;

б) сколько единиц нужно отобрать на второй фазе выборки, чтобы относительная ошибка исследованного признака на этой фазе не превышала 10%, гарантируя результат с вероятностью 0,954.

Известно, что коэффициент вариации признака, исследуемого на второй фазе выборки, равен 0,5.

29. При социологическом обследовании бытовых условий 20% рабочих завода, отобранных механически по спискам рабочих в каждом цехе, были получены следующие распределения домохозяйств рабочих по обеспеченности жилой площадью (табл. 8.9).

Таблица 8.9

Жилплощадь на одного члена семьи, м ²	До 8	8–12	12–16	16–20
Механический цех	8	26	14	2
Сборочный цех	17	39	20	4

Определите средний размер жилплощади, приходящейся на одного человека в домохозяйствах рабочих завода (с вероятностью 0,940). Какова вероятность того, что ошибка при определении средней жилплощади не превысит 0,5 м²?

Какова вероятность того, что удельный вес рабочих, в семьях которых на 1 человека приходится свыше 12 м², по заводу не превысит 20%?

Является ли существенной разница между обеспеченностью жилой площадью семей рабочих двух цехов?

30. На электроламповом заводе контролером качества продукции в случайном порядке было взято для проверки 100 ламп.

Средняя продолжительность горения отобранных электроламп оказалась 2420 ч при $s = 61,03$ ч.

Какова продолжительность горения всех ламп в сдаваемой заводом партии? Ответ дайте с вероятностью 0,954.

У какой доли ламп срок службы окажется меньше установленного заказчиком лимита 2400 ч, если при проверке таких ламп оказалось 10?

31. Двое рабочих на одинаковых станках изготавливают одинаковые детали. Было взято случайным образом по 200 деталей, изготовленных каждым из них, и оказалось, что число бракованных деталей у первого рабочего составило 12, а у второго – 18.

Определите:

а) есть ли существенная разница в проценте допускаемого ими брака;

б) сколько нужно взять деталей, изготовленных каждым рабочим, чтобы обнаружить разницу в доле брака по крайней мере в 4%?

8.3. Вопросы для самоконтроля

1. Почему используется выборочное наблюдение?
2. Каково выборочное распределение выборочных средних?
3. Если генеральная совокупность не подчиняется нормальному распределению, то можно ли утверждать, что выборочное распределение выборочных средних будет нормальным?
4. Аудитору для проверки нужно отобрать каждый 20-й файл с документами из 1000, начиная с 5-го файла. Какие номера файлов подлежат аудиторской проверке и можно ли назвать такую выборку простой случайной выборкой?
5. Если в результате стратифицированной выборки оказалось, что представительность каждой страты не соответствует пропорциям в генеральной совокупности, то следует ли применять взвешивание в соответствии с генеральными пропорциями?
6. К какому типу выборки относится выборка, проведенная из каждой предварительно выделенной подгруппы генеральной совокупности?
7. Совпадают ли понятия объективности отбора и равной возможности для всех единиц генеральной совокупности попасть в выборку?
8. Можно ли утверждать, что оценка генеральной средней будет менее надежной при большом объеме выборки, нежели при малом за счет возрастания ошибок наблюдения?
9. Как называются значения доверительного интервала?
10. На основе чего принимаются решения относительно величины дисперсии при расчете объема выборки?
11. Как устанавливается степень доверия к результатам выборки?
12. От чего зависит величина максимально допустимой ошибки выборки и в решении какой задачи она используется?
13. Поясните, как вы понимаете «выборочное распределение средней»?
14. Дайте определение ошибки выборки и поясните, чем она вызывается?
15. В чем состоит правило «трех сигм» и как оно применяется?

8.4. Методические рекомендации преподавателям

Преподавание этой темы дает наиболее эффективные результаты, когда обеспечено сочетание аудиторных занятий в форме решения задач и тестирования студентов с выполнением студентами домашних заданий, организацией учебного выборочного обследования какой-либо стороны студенческой жизни, знакомством с практикой применения выборочного метода российской государственной статистикой.

Обратите внимание студентов на четыре основных причины использования выборки:

- уменьшение ошибок наблюдения и повышение точности данных;
- сокращение затрат на сбор и обработку данных;
- обеспечение возможности статистического подхода к контролю и анализу качества продукции, которые связаны с уничтожением (порчей) испытуемых образцов;
- возможность расчета ошибки выборки и распространения данных выборки на генеральную совокупность при объективности отбора.

Подчеркните, что сфера применения выборочного метода непрерывно расширяется.

Некоторые опасности кроются в том, что студенты могут путать предельную ошибку выборки (расчетную величину) с допустимым пределом ошибки выборки (величина которой задается экспертом). Следует подчеркнуть, что первая основана на результатах проведенной выборки, а вторая необходима на стадии планирования выборки. Важно объяснить понимание зависимости решения всех вопросов выборки от принятой степени доверия к ее результатам (т.е. от уровня доверительной вероятности).

Если позволяют часы аудиторных занятий, то весьма полезно было бы проведение контрольной работы, включающей как решение задач, так и тесты. При ограниченном времени рекомендуется проводить тестирование на каждом занятии в течение первых 15–20 мин.

По итогам выполнения домашних заданий рекомендуется провести разбор результатов с выявлением наиболее типичных ошибок. Рекомендуется использовать для домашних заданий

студентам комплексные задачи в сочетании с задачами других разделов.

8.5. Методические указания студентам

Приступая к изучению этой темы, подумайте, не приходилось ли вам в жизни сталкиваться с выборкой. Каждый раз, когда будете задаваться целью ответить на вопрос о формировании выборки, вы оказываетесь перед необходимостью определить границы генеральной совокупности. Например, если бы вы решили провести опрос об отношении общества к курению, то что понимать под «обществом»: взрослое население всего мира, России или конкретного населенного пункта? Итак, очевидно, что изучение этой темы связано с изучением тем 1, 2. Оно предполагает понимание и владение исходными понятиями, определениями, формулами.

Обратите особое внимание на понятие репрезентативности выборки и на факторы, обеспечивающие репрезентативность выборки. Базовым вопросом при изучении этой темы является объективность отбора единиц в выборку. Внимательно отнеситесь также к примеру использования таблицы случайных чисел для формирования выборки, приведенному в Учебнике (с. 220–221).

Для освоения выборочного метода рекомендуем самостоятельно прорешать типовые примеры, имеющиеся в Учебнике и в Практикуме, и сопоставить ваше решение с указанным. Обратите внимание и на то, что выборка — это еще одна сфера, где применяется правило сложения дисперсий.

Большое значение в освоении этой темы имеет умение правильно пользоваться статистико-математическими таблицами (таблицей интеграла вероятностей, распределением Стьюдента).

Проверка статистических гипотез

9.1. Основные положения

Генеральную характеристику получают либо на основе выборочного показателя с учетом его случайной ошибки:

$\bar{\mu} = \bar{x} \pm \Delta_{\bar{x}}$, либо в отношении свойств генеральной совокупности выдвигается некоторая гипотеза: о величине средней, дисперсии, характере распределения, тесноты связи. При проверке гипотезы проверяется согласованность эмпирических данных с гипотетическими. В основе проверки лежит вероятностное представление о величине случайной ошибки (Δ). Если расхождения между эмпирическими и теоретическими величинами при определенной вероятности не выходят за пределы случайных ошибок, гипотезу принимают (точнее — не отклоняют). При этом не делается никаких заключений о правильности самой гипотезы. Речь идет лишь о согласованности эмпирических и теоретических данных.

Практически оценка гипотезы сводится к расчету специальных показателей — критериев: z , t , χ^2 , F и к сопоставлению их с соответствующими максимально возможными значениями, рассчитанными с определенной вероятностью в условиях случайной выборки.

Проверка статистических гипотез складывается из следующих этапов.

1. Формулируется в виде статистической гипотезы задача исследования.
2. Выбирается статистическая характеристика гипотезы.
3. Анализируются возможные ошибочные решения и оцениваются последствия их.
4. Выбираются испытываемая и альтернативная гипотезы.
5. Задается уровень значимости α и определяется критическое значение статистической характеристики.

6. Вычисляется фактическое значение статистической характеристики, проверяется испытываемая гипотеза и в зависимости от результатов проверки отклоняется или принимается.

При этом безразлично, оцениваются ли гипотезы в отношении реальной генеральной совокупности или же в отношении гипотетической генеральной совокупности. Последнее открывает путь применению вероятностных оценок за пределами собственно выборки. Так, вероятностными оценками всегда дополняются результаты экспериментов. Эксперимент — это не выборочное наблюдение. Но при неизменных условиях он может быть повторен неограниченное число раз, поэтому результаты эксперимента могут рассматриваться как выборочные характеристики из бесконечно большой гипотетической генеральной совокупности.

Вероятностные критерии рекомендуется также использовать для оценки показателей, полученных по материалам сплошного наблюдения в совокупностях небольшого объема для оценки процессов внутри совокупности: связей, распределений. При малом числе единиц корреляционные показатели и характеристики распределения могут быть искажены действием случайных факторов. В связи с этим рекомендуется проверить, не вызвана ли установленная закономерность стечением случайных обстоятельств, насколько она характерна для того комплекса условий, в которых находится обследуемая совокупность. Изучаемая совокупность рассматривается как выборка из некоторой гипотетической генеральной совокупности, единицы которой находятся в условиях данной реальности. В этом плане наиболее эффективно использование вероятностных критериев для оценки существенности расхождений характеристик отдельных групп.

Гипотеза в статистике понимается как предположение о распределении случайных величин. Так, например, гипотезой является предположение, что некоторое распределение имеет среднее значение, равное 20.

Критерий проверки гипотезы определяет метод проверки, в результате которой выясняется, верна или неверна данная гипотеза, т.е. должна ли она «приниматься» или «отвергаться». Если отклонение от гипотезы мало и является случайным, гипотеза не отвергается. Если же это отклонение нельзя считать случайным и речь идет о так называемом *значимом отклонении*, то гипотеза отвергается.

Гипотеза, отклонения от которой приписываются случаю, называется *нулевой* и обозначается H_0 . Противоположная, или *альтернативная*, гипотеза обозначается H_1 .

Проверяются гипотезы с помощью функции критерия T . *Функция критерия* есть не что иное, как выборочная функция $T(X_1, \dots, X_n/H_0)$. Распределение $f(T/H_0)$ критерия в предположении правильности нулевой гипотезы H_0 обычно является известным, например, нормальным, или распределением χ^2 , или t -распределением, или F -распределением (так называемые *критериальные распределения*). Если обозначить α вероятность ошибки (уровень значимости), то границы T_1 и T_2 должны определяться так, чтобы удовлетворялись равенства

$$P(T \leq T_1/H_0) + P(T > T_2/H_0) = \alpha;$$

$$P(T_1/H_0 < T < T_2/H_0) = 1 - \alpha.$$

Вытекающее отсюда решающее правило состоит в том, что нулевая гипотеза отвергается, когда выборочное значение T оказывается вне границ T_1 и T_2 и, следовательно, попадает в так называемую критическую область. Гипотеза H_0 не отвергается, если значение T попадает в интервал между T_1 и T_2 , но и не принимается безоговорочно.

Виды ошибок при проверке гипотез представлены в табл. 9.1.

1. Ошибка первого рода. Нулевая гипотеза H_0 отвергается, хотя является верной. Вероятность совершения такой ошибки равна α .

Таблица 9.1

Виды ошибок при проверке гипотез

Решение по критерию		Истина	
		H_0 верна H_1 не верна	H_0 не верна H_1 верна
Принята гипотеза	H_0	Правильное решение $P = 1 - \alpha$	Ошибка второго рода $P = 1 - \beta$
	H_1	Ошибка первого рода $P = \alpha$	Правильное решение $P = \beta$

2. Ошибка второго рода. Нулевая гипотеза H_0 принимается, хотя является неверной. Вероятность совершения этой ошибки равна $1 - \beta$. Здесь β – вероятность того, что ошибка второго рода, так называемая *мощность критерия*, не будет допущена.

Пусть, например, нулевая гипотеза H_0 означает, что определенная партия товаров соответствует условиям поставки. Если эта партия отклоняется, хотя гипотеза соответствует истине, то совершается ошибка первого рода. В таком случае говорят о *риске производителя товаров*. Если же партия принимается, хотя она и не соответствует условиям поставки, то имеет место ошибка второго рода – так называемый *риск потребителя*.

Рассмотренные виды ошибок показаны на рис. 9.1.



Рис. 9.1. Виды ошибок при испытании гипотез

Виды критериев соответствуют виду гипотез. Выделяются гипотезы параметрические и непараметрические. *Параметрические* – это гипотезы о генеральных параметрах: средней величине (математическом ожидании), дисперсии, коэффициенте корреляции, параметрах уравнения регрессии. *Непараметрические* – это гипотезы о законе распределения генеральной совокупности, гипотезы о структурных характеристиках (например, о медиане и т.д.).

Параметрические гипотезы и соответственно критерии их проверки бывают односторонними и двусторонними, в зависимости от вида альтернативной гипотезы H_1 .

Односторонний критерий, если выдвигается следующая гипотеза:

$$H_0 : \mu = \mu_0 ;$$

$$H_1 : \mu > \mu_0 .$$

В этом случае критическое значение c выбирается так, чтобы удовлетворялось равенство $P(\bar{x})c | H_0 = \alpha$ (рис. 9.2):

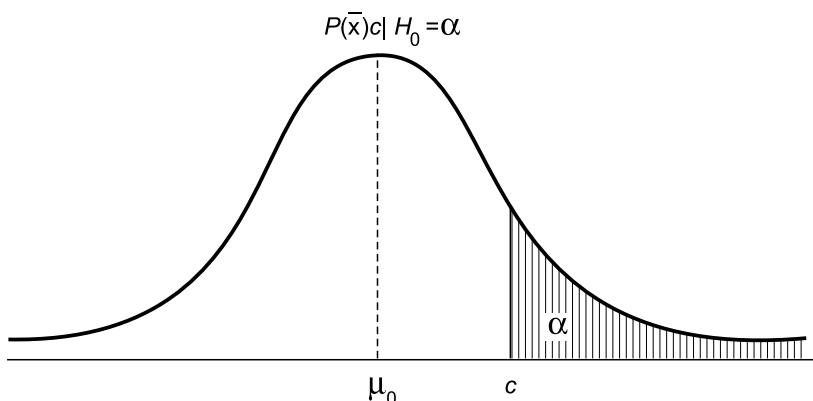


Рис. 9.2. Односторонний критерий

Двусторонний критерий, если выдвигаются нулевая и противоположная гипотезы:

$$H_0 : \mu = \mu_0 ;$$

$$H_1 : \mu \neq \mu_0 .$$

В этом случае критическое значение c (см. рис. 9.2) выбирается так, чтобы выполнялось равенство $P(|\bar{x} - \mu_0| > c | H_0 = \alpha_1 + \alpha_2 = \alpha$ (рис. 9.3).

Проверка статистических гипотез проводится применительно к одной и двум выборкам. В задаче с одной выборкой требуется ответить на вопрос: взята ли эта выборка из определенной заданной генеральной совокупности или нет? В задаче с двумя выборками нужно выяснить, взяты ли указанные не зависящие друг от друга выборки из одной и той же генеральной совокупности или нет.

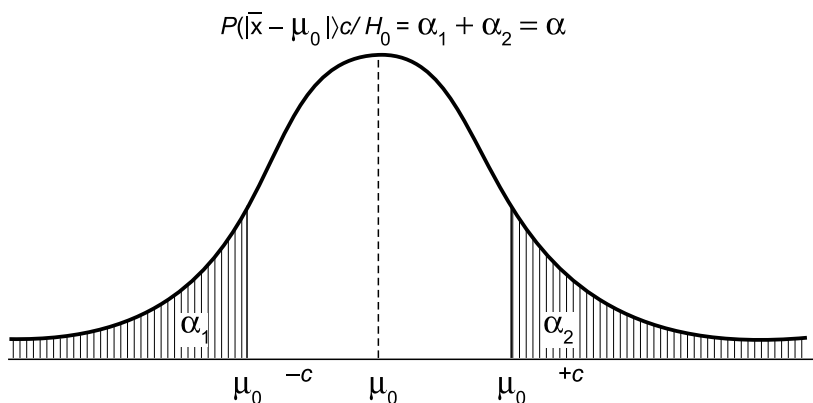


Рис. 9.3. Двусторонний критерий

Задача с одной выборкой

Пример 9.1. Односторонний критерий.

Пусть известно, что 40% студентов-экономистов имеют в личном пользовании карманные компьютеры ($\pi_0 = 0,4$). А по выборке объема $n = 25$ получена доля 46% ($p = 0,46$). В качестве нулевой гипотезы примем, что выявленное отклонение – чисто случайное:

$$H_0 : \pi = \pi_0;$$

$$H_1 : \pi > \pi_0 \text{ (односторонний критерий).}$$

Проверка гипотезы проводится в четыре шага.

Шаг 1: выбор уровня значимости α и определение теоретического критического значения критерия с помощью таблицы t -распределения (см. табл. П2.2). Прежде всего отметим, что при $n > 30$ критическая величина t имеет стандартное нормальное распределение. Для $\alpha = 0,025$ и при числе степеней свободы $df = n - 1 = 24$ из таблицы распределения Стьюдента получаем теоретическое критическое значение: $t = 2,064$.

Шаг 2: установление решающего правила. Гипотеза H_0 отвергается, когда эмпирическая (выборочная) величина t оказывается больше теоретического критического значения t .

Шаг 3: вычисление эмпирической (фактической) критической величины t . В одновыборочном случае t определяется по формуле

$$t = \frac{p - \pi_0}{\sqrt{\frac{pq}{n}}}.$$

По данным примера получаем:

$$t = \frac{0,46 - 0,4}{\sqrt{\frac{0,46 \cdot 0,54}{25}}} = 0,6.$$

Шаг 4: принятие решения.

$$t_{\text{факт}} < t_{\text{табл}};$$

$$0,6 < 2,06.$$

Гипотеза H_0 не отклоняется, так как различие между наблюдаемой и предполагаемой долями студентов, имеющих мини-компьютер, является чисто случайным.

Пример 9.2. Двусторонний критерий.

Пусть при нарезке металлических пластинок задана их номинальная длина 150 см ($\mu_0 = 150$). Выборка пластинок при $n = 9$ дала среднее значение 154 см ($\bar{x} = 154$). Стандартное отклонение равно 4 см:

$$H_0 : \mu = \mu_0;$$

$$H_1 : \mu \neq \mu_0.$$

Шаг 1: уровень значимости $\alpha = 0,01$; число степеней свободы $df = n - 1 = 8$.

В качестве теоретического критического значения из табл. t -распределения (см. приложение 2) получаем $t = 3,355$.

Шаг 2: гипотеза H_0 отвергается при $|t_{\text{факт}}| > t_{\text{табл}}$.

Шаг 3:

$$t_{\text{факт}} = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{154 - 150}{4} \sqrt{9} = 3.$$

Шаг 4: $|t| = 3 < t = 3,355$.

Отклонение случайное. Гипотеза H_0 не отвергается.

Задача с двумя выборками

Пример 9.3. При первом выборочном обследовании установлено, что доля водителей автомобилей, применяющих ремень безопасности, составляет 75% при $n = 20$ ($p_1 = 0,75$; $n_1 = 20$). Второе обследование дало значение 70% при объеме выборки $n = 32$. Существует ли между этими выборочными долями значимое различие или можно сказать, что обе выборки происходят из одной и той же генеральной совокупности?

$$H_0 : \pi_1 = \pi_2;$$

$$H_1 : \pi_1 \neq \pi_2.$$

Шаг 1:

$$\alpha = 0,05;$$

$$df = n_1 + n_2 - 2 = 20 + 32 - 2 = 50;$$

$$\text{отсюда } t_{\text{табл}} = 2,008.$$

Шаг 2: гипотеза H_0 отвергается, когда имеет место неравенство $|t_{\text{факт}}| > t_{\text{табл}}$.

Шаг 3:

$$t_{\text{факт}} = \frac{w_1 - w_2}{\sqrt{w(1-w) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

где $w(1-w)$ — объединенная дисперсия,

при этом w определяется по формуле

$$w = \frac{n_1 w_1 + n_2 w_2}{n_1 + n_2} = \frac{20 \cdot 0,75 + 32 \cdot 0,7}{52} \approx 0,7;$$

$$t = \frac{0,75 - 0,7}{\sqrt{(0,7 \cdot 0,3)(0,05 + 0,03)}} = \frac{0,05}{\sqrt{0,21 \cdot 0,08}} \approx 0,4.$$

Шаг 4:

$$|t| = 0,4 < t = 2,008.$$

Гипотеза H_0 не отвергается, так как различие выборочных долей чисто случайное. Выборки происходят из одной и той же генеральной совокупности.

Точно так же, как производится сравнение долей, производится сравнение средних. В случае *известных генеральных дисперсий* вычисляется значение критерия:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}.$$

Задавая желаемую вероятность вывода P (или вероятность ошибки $\alpha = 1 - P$), находим по таблице (см. приложение 2) соответствующее значение $t(P)$. Если $|t| > t(P)$, или иначе $|t_{\text{факт}}| > t_{\text{табл}}$, то расхождение средних можно считать неслучайным с надежностью вывода P . В противном случае при $t_{\text{факт}} < t_{\text{табл}}$ нет оснований считать расхождение значимым, т. е. оно является случайным.

Если дисперсии ошибок для обеих выборок одинаковы, $\sigma_1^2 = \sigma_2^2 = \sigma$, то формула критерия имеет вид

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{1/n_1 + 1/n_2}}.$$

Пример 9.4. Произведены две выборки: $n_1 = 25$, $n_2 = 40$. Средняя квадратическая ошибка $\sigma = 1,20$, средние значения: $\bar{x}_1 = 23,5$, $\bar{x}_2 = 22,8$.

Вычисляем

$$z = \frac{23,5 - 22,8}{1,20 \sqrt{1/25 + 1/40}} = \frac{0,7}{0,306} = 2,29.$$

По таблице интеграла вероятностей (табл. П2.1) находим $z_{\text{табл}} (\alpha = 0,01) = 2,576$. Так как $z_{\text{факт}} < z_{\text{табл}}$, то расхождение средних нельзя считать статистически значимым (с вероятностью 0,99). Но если принять надежность вывода $P = 0,95$, тогда $z_{\text{табл}} = 1,96$, а значит, $z_{\text{факт}} > z_{\text{табл}}$ и расхождение средних можно признать статистически значимым.

Сравнение средних *при неизвестной дисперсии* производится при предположении, что дисперсии в обеих совокупностях одинаковы. Если имеется выборка объема n_1 со средним значением \bar{x}_1 и дисперсией s_1^2 и выборка объема n_2 со средним значением

\bar{x}_2 и дисперсией s_2^2 , т.е. для решения вопроса о случайном или неслучайном расхождении средних значений вычисляется эмпирическое значение критерия:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{1/n_1 + 1/n_2}},$$

$$\text{где } s = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}}.$$

Задается желаемая вероятность вывода P (или уровень значимости α , т. е. вероятность ошибки первого рода) и находится табличное значение t , соответствующее заданному α и числу степеней свободы $df = n_1 + n_2 - 2$. Если $t_{\text{факт}} > t_{\text{табл}}$, то расхождение средних значений можно считать неслучайным (с вероятностью P). В противном случае нет оснований считать расхождение значимым.

Задача сравнения средних двух выборок может быть представлена как задача проверки гипотезы о равенстве двух средних нормально распределенных генеральных совокупностей:

$$H_0 : \mu_1 = \mu_2;$$

$$H_1 : \mu_1 \neq \mu_2.$$

Задавая вероятность вывода P (например, 0,95), находим симметричный интервал $(-t_{\alpha, df}, t_{\alpha, df})$. Статистика $t_{\text{факт}}$ может попасть в этот интервал случайно с вероятностью P . Значения t , лежащие вне этого интервала, образуют критическую область, вероятность попадания в которую составляет $1 - P = \alpha$.

Методы тестирования применяются и для проверки отличающихся наблюдений.

При получении результата наблюдения, резко отличающегося от всех других результатов, естественно возникает подозрение, что допущена грубая ошибка. При этом применяются различные критерии в зависимости от того, известна или нет средняя квадратическая ошибка измерений σ (предполагается, что все измерения производятся с одной и той же точностью и независимо друг от друга).

Метод исключения при известной σ . Обозначим «выскакивающее» значение через x_* , а все остальные результаты измерения через x_1, x_2, \dots, x_n , для которых среднее арифметическое значение определяется как

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Сравнивается абсолютная величина разности $x_* - \bar{x}$ с величиной $\sigma\sqrt{(n+1)/n}$:

$$t = \frac{|x_* - \bar{x}|}{\sigma\sqrt{(n+1)/n}}.$$

Затем определяется вероятность $1 - 2F(t)$ с помощью таблиц П2.2. Это вероятность того, что рассматриваемое отношение случайно примет значение, не меньшее, чем t , при условии, что значение x_* не содержит преднамеренной ошибки, т.е. что ошибка результата x_* только случайна. Если вычисленная вероятность окажется очень малой, то «выскакивающее» значение содержит неслучайную ошибку и его следует исключить из дальнейшей обработки данных.

Какую именно вероятность считать очень малой, зависит от конкретных условий решаемой задачи: если назначить слишком низкий уровень малых вероятностей, то грубые ошибки могут остаться, если же взять этот уровень неоправданно большим, то можно исключить результаты со случайными ошибками и тем самым необоснованно уменьшить объем выборки. Обычно принимают один из трех уровней малых вероятностей:

5%-ный уровень (исключаются данные, вероятность появления которых меньше 0,05);

1%-ный уровень (исключаются данные, вероятность появления которых меньше 0,01);

0,1%-ный уровень (исключаются данные, вероятность появления которых меньше 0,001).

При выбранном уровне α малых вероятностей «выскакивающее» значение x_* считают содержащим неслучайную ошибку, если для соответствующего отношения t вероятность $1 - 2F(t) < \alpha$. Чтобы подчеркнуть вероятностный характер этого заключения,

говорят, что значение x_* содержит неслучайную ошибку с надежностью вывода $P = 1 - \alpha$. Значение t , для которого $1 - 2F(t) = \alpha$ и, значит, $2F(t) = P$, называется критическим значением при надежности P . Так, если $\alpha = 0,01$ (1%-ный уровень), то $P = 0,99$, критическое значение $t = t(P) = 2,576$ (см. табл. П2.2) и, если $t_{\text{факт}}$ превзойдет это критическое значение, можно браковать «выскакивающее» значение x_* с надежностью вывода 0,99.

Пример 9.5. Пусть из 41 результата независимых измерений, произведенных со средней квадратической ошибкой $\sigma = 0,133$, обнаружено одно «выскакивающее» значение $x_* = 6,866$, причем среднее из остальных 40 результатов составляет $\bar{x} = 6,500$. Можно ли считать, что «выскакивающее» значение содержит неслучайную ошибку и его следует исключить из дальнейшей обработки данных?

Решение. Разность между «выскакивающим» значением и средним составляет $x_* - \bar{x} = 0,366$. Отсюда получаем

$$t = \frac{0,366}{0,133\sqrt{41/40}} = 2,72.$$

По таблице приложения 2 для $t = 2,72$ оцениваем вероятность $1 - 2F(t) = 0,0066 \leq 0,007$. Следовательно, с надежностью вывода $P > 0,993$ можно считать, что значение x_* содержит неслучайную ошибку, и необходимо исключить это значение из дальнейшей обработки результатов измерения.

Подчеркнем, что указанная методика применяется только тогда, когда величина средней квадратической ошибки σ точно известна заранее.

Метод исключения при неизвестной σ . Если величина σ заранее неизвестна, то она оценивается приближенно по результатам наблюдений, т. е. вместо нее используют эмпирическое стандартное отклонение:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

При этом абсолютную величину разности $|x_* - \bar{x}|$ между «выскакивающим» значением x_* и средним значением \bar{x} остальных результатов делят на стандартное отклонение:

$$t = |x_* - \bar{x}| / s.$$

Полученное отношение $t_{\text{факт}}$ сравнивают с критическими значениями $t_n(P)$ из табл. П2.2. Если при данном числе n приемлемых результатов $t_{\text{факт}}$ оказывается между двумя критическими значениями при надежности P_1 и P_2 ($P_2 > P_1$), то с надежностью вывода, большей P_1 , можно считать, что «высказывающее» значение содержит неслучайную ошибку, и нужно исключить его из дальнейшей обработки результатов.

Заметим, что если надежность вывода окажется недостаточной, то это свидетельствует не об отсутствии неслучайной ошибки, а лишь об отсутствии достаточных оснований для исключения «высказывающего» значения.

Пример 9.6. Пусть для n результатов независимых наблюдений некоторой величины среднее значение равно $\bar{x} = 6,500$, при стандартном отклонении $s = 0,133$ и пусть $(n + 1)$ -е наблюдение дало результат $x_* = 6,866$. Можно ли исключить этот результат из дальнейшей обработки?

Решение. Получаем $t = 0,366/0,133 = 2,75$. Если число приемлемых результатов $n = 40$, то полученное отношение превосходит критическое значение 2,742 при надежности $P = 0,99$, следовательно, значение x_* можно исключить с надежностью вывода, большей 0,99. Если же, например, число приемлемых результатов $n = 6$, то полученное отношение меньше критического значения 2,78 даже при надежности $P = 0,95$ и значение x_* исключать не следует.

Проверка нормальности распределения. Обычно исходные данные сгруппированы и представлены в виде вариационного ряда. При этом число единиц в каждом интервале (частота) должно быть не менее пяти ($f_j \geq 5$). Вычисляются вероятности P_j попадания в соответствующий интервал при нормальном законе распределения вероятностей:

$$P_j = F\left(\frac{x_j - \bar{x}}{s}\right) - F\left(\frac{x_{j-1} - \bar{x}}{s}\right),$$

где \bar{x} — среднее значение;

s — стандартное отклонение;

F — интеграл вероятностей (см. табл. П2.1).

Затем рассчитываются частоты, соответствующие нормальному распределению, $\hat{f}_j = n \cdot P_j$, где n — объем выборки.

Очевидно, что

$$\sum_{j=1}^k \hat{f}_j = \sum_{j=1}^k f_j = n.$$

Наконец, вычисляется значение критерия:

$$\chi_{\text{факт}}^2 = \sum_{j=1}^k \frac{(f_j - \mu_j)^2}{\mu_j},$$

где k – число интервалов (групп).

Затем $\chi_{\text{факт}}^2$ сравнивается с табличным (критическим) значением (см. табл. П2.4). Распределение $\chi_{\text{табл}}^2$ зависит от числа степеней свободы df :

$$df = k - p - 1,$$

где p – число параметров в формуле закона распределения, которые определяются по выборке и соответственно уменьшают число степеней свободы. Кроме того, снимается еще одна степень свободы, поскольку должно выполняться равенство сумм фактических и теоретических частот:

$$\sum_{j=1}^k f_j = \sum_{j=1}^k \mu_j.$$

В случае гипотезы о нормальном распределении $p = 2$ (так как рассчитываются выборочные \bar{X} и s).

Пример 9.7. Имеется распределение деталей по весу (табл. 9.2).

Таблица 9.2

Вес детали $x_{j-1} \div x_j$	Число деталей f_j	z_j	$F(z_j)$	P_j	\hat{f}_j	$\frac{(f_j - \mu_j)^2}{\mu_j}$
А	1	2	3	4	5	6
До 8,425	7	-1,614	-0,4467	0,0533	5,33	0,523
8,425–8,475	5	-1,220	-0,3888	0,0579	5,79	0,108
8,475–8,525	8	-0,827	-0,2959	0,0929	9,29	0,179
8,525–8,575	10	-0,433	-0,1676	0,1283	12,83	0,624
8,575–8,625	18	-0,039	-0,0156	0,1520	15,20	0,516

Продолжение

Вес детали $x_{j-1} \div x_j$	Число деталей f_j	z_j	$F(z_j)$	P_j	\hat{f}_j	$\frac{(f_j - \mu_j)^2}{\mu_j}$
А	1	2	3	4	5	6
8,625–8,675	17	0,354	0,1383	0,1539	15,39	0,168
8,675–8,725	12	0,748	0,2728	0,1345	13,45	0,157
8,725–8,775	9	1,142	0,3733	0,1005	10,05	0,110
8,775–8,825	7	1,536	0,4377	0,0644	6,44	0,048
8,825 и более	7	$+\infty$	0,5000	0,0623	6,23	0,095
Итого	100	—	—	1	100	2,528

$$\bar{x} = 8,63; \quad s = 0,127.$$

Во 2-м столбце вычислены $z_j = \frac{x_j - \bar{x}}{s}$ для правых границ

интервалов (x_j). Крайние интервалы приняты бесконечными. В 3-м столбце показаны соответствующие значения интеграла вероятностей (см. табл. П2.1). В 4-м столбце по значениям $F(z_j)$ вычислены разности $P_j = F(z_j) - F(z_{j-1})$. При вычислении P_1 принято $F(z_{j-1}) = F(-\infty) = -0,5$, точно так же при вычислении P_{10} принято $F(z_j) = F(\infty) = 0,5$. Суммируя значения, показанные в 6-м столбце, получаем статистику хи-квадрат: $\chi^2 = 2,528$.

Находим табличное значение критерия: $\chi_{\text{табл}}^2 = 14,07$ (при $\alpha = 0,05$, $df = 10 - 3 = 7$). $\chi_{\text{факт}}^2 < \chi_{\text{табл}}^2$, следовательно, нельзя отклонить гипотезу о нормальности распределения.

Приближенные методы проверки. В качестве приближенного метода проверки нормальности распределения применяют метод, связанный с оценками центральных моментов третьего (μ_3) и четвертого (μ_4) порядков. В случае нормального распределения $\mu_3 = 0$, $\mu_4 = 3\sigma^4$. Соответственно рассчитанные по выборке центральные моменты m_3 и m_4 должны быть $m_3 \approx 0$, $m_4 \approx 3s^4$. Приближенную оценку нормальности удобнее проводить на основе безразмерных характеристик:

$$\begin{aligned} \text{коэффициента асимметрии } As &= m_3/s^3, \\ \text{показателя эксцесса } Ex &= m_4/s^4 - 3. \end{aligned}$$

Обе эти характеристики должны быть малы в случае нормальности распределения. О малости As и Ex судят по сравнению их со среднеквадратическими ошибками:

$$\sigma_{As} = \sqrt{6(n-2)/(n+1)(n+3)};$$

$$\sigma_{Ex} = \sqrt{24n(n-2)(n-3)/(n+1)^2(n+3)(n+5)},$$

где n — число наблюдений (объем выборки).

Если хотя бы одна из указанных характеристик по абсолютной величине в два и более раз превосходит свою среднеквадратическую ошибку, то гипотеза о нормальности распределения должна быть подвергнута сомнению и проверена более тщательно (например, с помощью критерия хи-квадрат).

Пример 9.8. По данным примера 9.7: $s = 0,127$, интервал группировки $j_x = 0,05$, $m_3 = 0,05^3 \cdot (-2,4) = -0,0003$, $m_4 = 0,00067$.

$As = -0,0003 : 0,00205 = -0,146$; $Ex = 0,00067 : 0,00026 - 3 = 2,58 - 3 = -0,42$. Как As , так и Ex достаточно малы, причем статистически значимы.

Среднеквадратические ошибки этих показателей

$$\sigma_{As} = \sqrt{6 \cdot 98 / (101 \cdot 103)} = 0,24;$$

$$\sigma_{Ex} = \sqrt{24 \cdot 100 \cdot 98 \cdot 97 / (101^2 \cdot 103 \cdot 105)} = 0,46.$$

Значения показателей асимметрии и эксцесса превосходят их среднеквадратические ошибки, следовательно, нет оснований сомневаться в нормальности распределения.

Логарифмически нормальное распределение. Достаточно часто исходные данные не отвечают нормальному распределению, но их преобразование оказывается нормально распределенным. Чаще всего встречаются случаи, когда нормально распределены логарифмы наблюдаемых данных, т.е. $\ln x$. В этом случае говорят, что переменная x следует логарифмически нормальному распределению.

Обозначим $\ln x = y$, т.е. $x = e^y$.

Оценка среднего значения величины y :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Оценка среднего значения величины x :

$$e^{\bar{b}} \equiv e^{\bar{y}} = \sqrt[n]{\prod_{i=1}^n e^{y_i}} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}.$$

Доверительная оценка для центра распределения величины y имеет вид:

$$\bar{y} - t \frac{\sigma}{\sqrt{n}} < \beta < \bar{y} + t \frac{\sigma}{\sqrt{n}}.$$

Отсюда доверительный интервал для среднего значения величины x :

$$\sqrt[n]{\prod_{i=1}^n x_i} \cdot e^{-t\sigma/\sqrt{n}} < e^b < \sqrt[n]{\prod_{i=1}^n x_i} \cdot e^{t\sigma/\sqrt{n}}.$$

При больших n этот доверительный интервал имеет вид:

$$\sqrt[n]{\prod_{i=1}^n x_i} (1 - t \frac{\sigma}{\sqrt{n}}) < e^b < \sqrt[n]{\prod_{i=1}^n x_i} (1 + t \frac{\sigma}{\sqrt{n}}).$$

В этом и предыдущем неравенстве вместо \bar{X} использована медиана, $X_{\mu_e} \approx e^b$.

Дисперсионный анализ. Этот метод применяется, когда нужно сравнить более чем две средних величины. Существенность (или случайность) различий между ними устанавливается с помощью сравнения двух по-разному вычисленных дисперсий.

Выделяют *однофакторный* и *двухфакторный* (многофакторный) дисперсионный анализ. В первом случае группы образуются по одному признаку x и рассчитываются групповые средние \bar{y}_j , где j — номер группы.

Проверяемая гипотеза $H_0: \mu_1 = \mu_2 = \dots = \mu_m$;

альтернативная гипотеза $H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_m$.

В качестве оценок генеральных средних μ_j принимаются выборочные средние \bar{y}_j .

Вариация y , приписываемая переменной x , измеряется межгрупповой дисперсией $\sigma_{\bar{y}_j}^2$. Соответствующая сумма квадратов

отклонений имеет вид: $SS_{\text{факт}} = \sum_{(j)} (\bar{y}_j - \bar{y})^2 n_j$.

Вариация y , приписываемая другим переменным, не связанным с x , измеряется средней из внутригрупповых дисперсий $\overline{\sigma_y^2}$.

Соответствующая сумма квадратов отклонений

$$SS_{\text{ост}} = \sum_{(j)} \sum_{(i)} (y_{ij} - \bar{y}_j)^2.$$

Чтобы сравнить эти две суммы квадратов отклонений, их нужно привести к одной степени свободы. Число степеней свободы для вариации за счет фактора x (межгрупповая) равно $df_{\text{факт}} = m - 1$, т.е. числу групп без единицы. Одна степень свободы пропадает за счет связи групповых средних \bar{y}_j с общей средней \bar{y} :

$$\sum_{(j)} \bar{y}_j n_j = \bar{y} \cdot n.$$

Число степеней свободы для вариации за счет остальных факторов (не связанных с x) (внутригрупповая) равно $df_{\text{факт}} = n - m$, т.е. числу наблюдений минус число групп. Затем вычисляются:

$$s_{\text{факт}}^2 = \frac{SS_{\text{факт}}}{df_{\text{факт}}}; \quad s_{\text{ост}}^2 = \frac{SS_{\text{ост}}}{df_{\text{ост}}}.$$

После этого находится значение $F_{\text{критерия}} = \frac{s_1^2}{s_2^2}$, где $s_1^2 \geq s_2^2$.

В качестве s_1^2 обычно выступает $s_{\text{факт}}^2$, в качестве s_2^2 берется $s_{\text{ост}}^2$. Полученное значение F сравнивается с $F_{\text{табл}}(\alpha, df_1, df_2)$.

Если $F > F_{\text{табл}}$, то H_0 не принимается (отклоняется);
если $F < F_{\text{табл}}$, то H_0 принимается (не отклоняется).

Пример 9.9. По следующим данным нужно установить, влияет ли профиль инвестиционной компании на объем прибыли (табл. 9.3).

Таблица 9.3

Исходные данные

Профиль инвестиционной компании x	Число инвестиционных компаний n_j	Прибыль за год, млрд руб. y_{ij}	Средняя прибыль за год, млрд руб. \bar{y}_j	$(\bar{y}_j - \bar{y})^2 n_j$
1. Добыча, переработка и транспортировка полезных ископаемых	3	40, 28, 36	34,67	489,22
2. Строительство	4	19, 14, 21, 25	19,75	18,49
3. Обрабатывающая промышленность	3	12, 15, 9	12,00	294,03
Итого	10	219	21,90	801,74

Испытуемая гипотеза $H_0: \mu_1 = \mu_2 = \mu_3$;

альтернативная гипотеза $H_1: \mu_1 \neq \mu_2 \neq \mu_3$.

Сумма, показанная в последней графе табл. 9.3, представляет собой вариацию прибыли за счет профиля инвестиционной компании (межгрупповая колеблемость). Вариация за счет остальных факторов (остаточная) равна:

в группе 1 $(40 - 34,67)^2 + (28 - 34,67)^2 + (36 - 34,67)^2 = 74,67$;

в группе 2 $(19 - 19,75)^2 + (14 - 19,75)^2 + (21 - 19,75)^2 = 35,19$;

в группе 3 $(12 - 12)^2 + (15 - 12)^2 + (9 - 12)^2 = 18$.

В целом по трем группам остаточная вариация равна 127,86.

Составим таблицу дисперсионного анализа (табл. 9.4).

Таблица 9.4

Таблица дисперсионного анализа

Источник вариации	Сумма квадратов SS	Число степеней свободы df	$s^2 = \frac{SS}{df}$	$F = \frac{s_1^2}{s_2^2}$
Фактор x (межгрупповая)	801,74	$3 - 1 = 2$	400,87	—
Остальные факторы (внутригрупповые)	127,86	$10 - 3 = 7$	18,27	21,94
Общий	929,60	$10 - 1 = 9$	—	—

Табличное значение при $\alpha = 0,05$, $df_1 = 2$, $df_2 = 7$ составляет $F_{\text{табл}} = 4,74$. Таким образом, $F_{\text{факт}} > F_{\text{табл}}$, следовательно, H_0 отклоняется.

Двухфакторный дисперсионный анализ. В этом случае изучается вариация y под влиянием двух факторов, их взаимодействия и остаточной вариации (за счет прочих факторов).

Пример 9.10. По данным 18 инвестиционных компаний установить, влияет ли профиль компании и продолжительность существования на прибыль. Исходные данные представлены в табл. 9.5 и 9.6.

Таблица 9.5

Прибыль инвестиционных компаний

Продолжительность работы	Прибыль компании, млрд. руб.		
	А	Б	В
5 лет и более	16	20	9
	20	12	11
	17	13	10
До 5 лет	12	10	6
	8	10	9
	13	13	9

Таблица 9.6

Средние значения

	А	Б	В	Итого
а	$\bar{y}_{11} = \frac{53}{3}$	$\bar{y}_{12} = \frac{45}{3}$	$\bar{y}_{13} = \frac{30}{3}$	$\bar{y}_1 = \frac{128}{9}$
б	$\bar{y}_{21} = \frac{53}{3}$	$\bar{y}_{22} = \frac{38}{3}$	$\bar{y}_{23} = \frac{24}{3}$	$\bar{y}_2 = \frac{90}{9}$
Итого	$\bar{y}_1 = \frac{86}{6}$	$\bar{y}_2 = \frac{78}{6}$	$\bar{y}_3 = \frac{54}{6}$	$\bar{y} = \frac{218}{18}$

В этом случае проверяются три гипотезы:

- 1) $H_0 : \mu_1 = \mu_2$;
- 2) $H_0 : \mu_{.1} = \mu_{.2} = \mu_{.3}$;
- 3) $H_0 : \mu_{11} = \mu_{22} = \dots = \mu_{23}$.

Таблица дисперсионного анализа имеет следующий вид (табл. 9.7).

Таблица 9.7

Схема двухфакторного дисперсионного анализа

Источник вариации	SS	df
Фактор x_1	$\sum_{(i)} (\bar{y}_i - \bar{y})^2$	$m - 1$
Фактор x_2	$\sum_{(j)} (\bar{y}_j - \bar{y})^2$	$p - 1$
Взаимодействие $x_1 x_2$	$\sum_{(i)} \sum_{(j)} (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$	$(m - 1)(p - 1)$
Общий межгрупповой эффект	$\sum \sum (\bar{y}_{ij} - \bar{y})^2$	$mp - 1$
Внутригрупповой эффект	$\sum \sum (y_{ije} - \bar{y}_{ij})^2$	$mp(n_{ij} - 1)$
Общий	$\sum \sum \sum (y_{ije} - \bar{y})^2$ $(i)(j)(e)$	$mpn_{ij} - 1$
<p>Здесь m – число групп по фактору x_1; p – число групп по фактору x_2; n_{ij} – число единиц в подгруппе; \bar{y}_i – среднее значение y по группе с i-м значением переменной x_1; \bar{y}_j – среднее значение y по группе с j-м значением переменной x_2; \bar{y}_{ij} – среднее значение y в подгруппе с i-м значением x_1 и j-м значением x_2; y_{ije} – значение y для e-го наблюдения в подгруппе с i-м значением x_1 и j-м значением x_2; \bar{y} – среднее значение y по всем наблюдениям.</p>		

Результаты расчетов представлены в табл. 9.8.

Таблица 9.8

Таблица результатов

Источник вариации	SS	df	s^2	$F_{\text{факт}}$	$F_{\text{табл}}$ ($\alpha = 0,05$)
Фактор x_1	80,22	1	80,22	$F_1 = \frac{80,22}{6,22} = 12,9$	4,8
Фактор x_2	92,44	2	46,22	$F_2 = \frac{46,22}{6,22} = 7,4$	3,9
Взаимодействие x_1, x_2	16,45	2	8,22	$F_3 = 1,3$	3,9
Общий межгрупповой эффект	189,11	5	37,82	$F_4 = 6,1$	3,1
Внутригрупповой эффект	74,67	12	6,22	—	—
Общий	263,67	17	—	—	—

Сравнение $F_{\text{факт}}$ и $F_{\text{табл}}$ позволяет сделать выводы:

- 1) H_0 отклоняется;
- 2) H_0 отклоняется;
- 3) H_0 не отклоняется.

Все параметрические критерии основаны на предположении о нормальном генеральном распределении. Это предположение не всегда выполняется. Поэтому все большее применение находят *непараметрические критерии*, которые часто называют «свободными от распределения».

К непараметрическим критериям относятся *знаковые критерии*.

Пример 9.11. Рассмотрим данные о цене блокнотов для рисования (в руб.) в киосках книготорговых фирм А (x_i) и Б (y_i) (табл. 9.9).

Таблица 9.9

Исходные данные

x_i	y_i	Знак ($x_i - y_i$)
41,6	40,5	+
41,7	41,1	+
41,8	40,9	+
42,2	41,4	+
42,3	42,3	0

Продолжение

x_i	y_i	Знак $(x_i - y_i)$
41,2	41,7	—
40,9	41,8	—
41,3	41,1	+
41,5	40,7	+
41,7	41,2	+
41,8	41,4	+

Если обе выборки взяты из одного и того же распределения, то число знаков «плюс» и «минус» должно быть приблизительно равным. В нашем случае (при исключении совпавших значений) ожидаемое число минусов равно $0,5 \cdot n = 0,5 \cdot 10 = 5$.

Фактически число минусов оказалось равно двум. Какова вероятность такого результата, если ожидаемое значение равно пяти? Поскольку наблюдения независимы, то вероятность определяется биномиальным распределением при $n = 10$, $\pi = 0,5$, $x = 2$. Эта вероятность равна 0,0439, т.е. она достаточно мала. Значит, можно сделать вывод, что такое небольшое число минусов не случайно, а свидетельствует о том, что элементы x и y принадлежат разным генеральным совокупностям. Иначе говоря, цены в киосках фирмы А существенно выше цен в киосках фирмы Б.

Критерий знаков Вилкоксона для сравнения двух выборок.

Этот критерий часто называют *знаковым* критерием Вилкоксона. При его построении во внимание принимают разности $x_i - y_i = |d_i|$.

Пример 9.12. По данным предыдущего примера (см. табл. 9.9) вычислим абсолютные значения разности цены в 11 киосках (табл. 9.10).

Таблица 9.10

Расчет знакового критерия Вилкоксона

№ п/п	$d_i = x_i - y_i$	$ d_i $	Ранг	R^+	R^-
1	1,1	1,1	10	10	—
2	0,6	0,6	5	5	—
3	0,9	0,9	8,5	8,5	—
4	0,8	0,8	6,5	6,5	—
5	0	0	—	—	—
6	−0,5	0,5	3,5	—	3,5

Продолжение

№ п/п	$d_i = x_i - y_i$	$ d_i $	Ранг	R^+	R^-
7	-0,9	0,9	8,5	—	8,5
8	0,2	0,2	1	1	—
9	0,8	0,8	6,5	6,5	—
10	0,5	0,5	3,5	3,5	—
11	0,4	0,4	2	2	—
Итого	—	—	—	43	12

Минимальный ранг присваивается наименьшему значению $|d_i|$, отличному от нуля. Сумма рангов положительных разностей есть не что иное, как знаковый критерий Вилкоксона при сравнении двух выборок:

$$W = \sum R^+.$$

Табличное значение критерия для односторонней проверки на 10%-ном уровне значимости при $n = 10$ находится в интервале $[15, 40]$, так что $W_{\text{факт}} > W_{\text{табл}}$, и можно признать различия в ценах статистически значимыми. Но если принять $\alpha = 0,05$, то критические значения W -критерия находятся в интервале $[11, 44]$, т.е. $W_{\text{факт}}$ попадает в этот интервал и гипотеза о том, что обе выборки принадлежат одной и той же генеральной совокупности, не может быть отклонена.

9.2. Задания по теме

9.2.1. Испытание гипотезы о величине средней

1. Норма выработки бригады паркетчиков при настилке полов в жилых домах — 12,0 кв. м за смену. Для проверки обоснованности норм при строительстве интернет-кафе в течение пяти дней учитывали объем работ, выполняемый паркетчиками. Средняя сменная выработка составила 11,2 кв. м со средним квадратическим отклонением 2,6 кв. м. Следует ли изменять сменную норму выработки паркетчиков при работе на данном объекте? Ответ дайте на 1%-ном уровне значимости.

2. По данным сплошного учета рабочих-штамповщиков завода, средний процент выполнения норм выработки за июнь составил 102,6%. Хронометраж работы 10 штамповщиков 27 де-

кабря того же года показал следующее выполнение норм выработки, %: 93,9; 112,1; 104,7; 110,8; 102,0; 105,6; 107,1; 104,0; 101,9; 98,4. Можно ли с вероятностью 0,99 сделать вывод, что за полгода выполнение норм выработки у штамповщиков повысилось?

3. Средний часовой заработок одного рабочего по результатам случайной выборки 25 рабочих предприятия (общая численность рабочих 200 человек) составляет 126 руб., среднее квадратическое отклонение — 42 руб. Оцените с точностью до копейки средний часовой заработок всех 200 рабочих, используя 95%-ные доверительные пределы.

Значимо ли на 5%-ном уровне значимости отличие среднего часового заработка 126 руб. от заработка 150 руб./ч?

4. По материалам статистической отчетности, средняя годовая удойность одной коровы в районе составила 2308 литров. В АОЗТ «Маяк» при стаде в 218 голов удойность одной коровы за год составила 2512 литров при среднем квадратическом отклонении 916 литров. Можно ли считать АОЗТ «Маяк» передовым по продуктивности коров? Проверку сделайте на уровне значимости 0,01.

5. При выборочном обследовании молодых рабочих завода (100 девушек и 100 юношей) оказалось, что девушки в среднем затрачивают на занятия спортом в неделю 2,3 часа, а юноши 3,6 часа. Коэффициенты вариации равны соответственно 53 и 61%. Можно ли считать при $\alpha = 1\%$, что пол рабочего влияет на время, затрачиваемое на занятия спортом?

6. По материалам выборочного обследования рабочих завода получены данные о среднемесячной заработной плате 18 намотчиков — 12116 руб. и 32 регулировщиков — 16212 руб. Коэффициенты вариации равны соответственно 28 и 36%. Можно ли считать на 5%-ном уровне значимости, что специальность рабочих влияет на заработную плату?

7. В результате бюджетного обследования семей рабочих оказалось, что в 18 семьях с душевым доходом менее 3500 руб. в среднем потребляется за месяц 8 кг картофеля на одного члена семьи (коэффициент вариации 40%), а в 42 семьях с душевым доходом свыше 10000 руб. — 5 кг (коэффициент вариации 52%).

Можно ли считать на 5%-ном уровне значимости, что уровень душевого дохода влияет на потребление картофеля?

9.2.2. Испытание гипотезы о характере распределения

8. По данным годовой отчетности автотранспортного предприятия, автомобили по коэффициенту использования пробега распределились следующим образом (табл. 9.11).

Таблица 9.11

Коэффициент использования пробега	Число автомобилей
0,29–0,36	5
0,36–0,43	11
0,43–0,50	47
0,50–0,57	57
0,57–0,64	22
0,64–0,71	8
0,71–0,78	4
Итого	154

Можно ли считать, что различия в использовании пробега автомобилей носят случайный характер? Дать ответ на 5%-ном уровне значимости.

9. Имеются данные об оплате отработанного человеко-дня 32 доярок района Н-ской области, руб.:

665	549	733	416	680	425	438
560	645	577	485	550	400	418
400	425	421	435	443	453	466
462	481	469	479	489	492	469
532	554	562	670			

Установите, носят ли различия в уровне оплаты труда доярок случайный характер ($\alpha = 1\%$).

10. Выборочный учет сельскохозяйственных предприятий Северо-Западного федерального округа показал следующее распределение хозяйств по урожайности зерновых (табл. 9.12).

Таблица 9.12

Урожайность, ц/га	Число АОЗТ, % к итогу
6,0–7,2	4,2
7,2–8,4	8,7

Продолжение

Урожайность, ц/га	Число АОЗТ, % к итогу
8,4–9,6	11,1
9,6–10,8	34,2
10,8–12,0	19,6
12,0–13,2	14,0
13,2–14,4	8,2
Итого	100,0

Можно ли считать при $\alpha = 5\%$, что распределение АОЗТ по урожайности подчиняется нормальному закону?

11. При выборочном обследовании заработной платы рабочих, ИТР и служащих промышленного предприятия получены следующие данные (табл. 9.13).

Таблица 9.13

Месячный зарплаток, руб.	9000– 12000	12000– 15000	15000– 18000	18000– 21000	21000– 24000	24000– 27000
Рабочие	32	57	112	74	28	5
ИТР и служа- щие	—	23	73	44	10	7

По материалам статистической отчетности, средняя месячная заработная плата одного работника составила 12600 руб.

1. Можно ли на основании данных выборочного учета сделать выводы о различиях между заработной платой рабочих и служащих?

2. Являются ли распределения рабочих, ИТР и служащих и всех работников завода по заработной плате нормальными? Принять уровень значимости 0,05.

3. Нанесите на график фактические распределения по зарплате рабочих, ИТР и служащих и распределения, соответствующие нормальному закону.

12. Средний возраст жителей рабочего поселка, по данным Всероссийской переписи населения 2002 г., равнялся 31,6 года.

По материалам выборочного социологического обследования населения, проведенного в 2004 г., получены следующие данные о распределении населения по возрасту (табл. 9.14).

Таблица 9.14

Возраст, лет	Число обследованных лиц	Возраст, лет	Число обследованных лиц
0–5	16	35–40	13
5–10	20	40–45	15
10–15	24	45–50	10
15–20	16	50–55	8
20–25	15	55–60	11
25–30	14	60–70	14
30–35	18	70–80	6

1. Определите, произошли ли изменения в среднем возрасте населения за 2 года ($\alpha = 0,05$).

2. Проверьте, подчиняется ли распределение населения по возрасту нормальному закону ($\alpha = 0,05$).

3. Изобразите на графике фактическое и теоретическое распределения.

13. При выборочном изучении чтения художественной литературы получены данные о числе книг, прочитанных студентами разных курсов за октябрь (табл. 9.15).

Таблица 9.15

Число книг	Число обследованных студентов на курсах			
	I	II	III	IV
0,5	4	4	2	2
1	4	13	3	5
1,5	20	50	16	18
2	54	12	47	34
2,5	18	16	22	35
3	4	18	4	6
3,5 и более	2	7	6	—

1. Сопоставьте среднее число книг, прочитанных одним студентом младших (I и II) и старших (III и IV) курсов. Изменяется ли объем чтения при переходе от младших к старшим курсам?

2. Проверьте по каждому курсу и по всем обследованным студентам, подчиняется ли распределение студентов нормальному закону.

3. Можно ли считать характер распределения студентов одинаковым:

- на I и II курсах;
- на III и IV курсах.

4. Нанесите на график фактическое распределение студентов разных курсов, всей выборочной совокупности и нормальную кривую.

5. Существенно ли различие в варьировании объема чтения студентов младших и старших курсов?

Ответы дайте на 1%-ном и 5%-ном уровнях значимости.

14. При обследовании работающих на предприятиях сельскохозяйственного машиностроения получены следующие распределения рабочих по тарифным разрядам (табл. 9.16).

Таблица 9.16

Разряд		I	II	III	IV	V	VI
Возраст, лет	До 25	84	226	319	241	82	3
	25 и старше	48	204	139	120	51	7

Существенны ли различия в характере распределения рабочих разного возраста? Ответ дайте с вероятностью 0,95.

15. Известно, что распределение студентов по IQ подчиняется нормальному закону. Среднее значение IQ для группы студентов составило 100, стандартное отклонение равно 30.

Какова вероятность появления студентов с IQ 85 и меньше? 130 и выше?

9.2.3. Оценка существенности связи с помощью критерия χ^2

16. По данным выборочного обследования жителей крупного города требуется определить, есть ли зависимость между питанием в столовой и социальной группой жителей (табл. 9.17).

Таблица 9.17

Социальная группа	Питается в столовой	Не питается в столовой	Итого
Рабочие	21	42	63
Служащие	50	42	92
Итого	71	84	155

Дайте ответ в двух вариантах: на уровне значимости 5% и уровне значимости 1%. При установлении наличия связи измерьте ее тесноту с помощью коэффициента контингенции.

17. По данным обследования семей города с доверительной вероятностью 0,95 выявите влияние удаленности их проживания от рынка на состав покупаемых там продуктов (табл. 9.18).

Таблица 9.18

Состав покупок	Добираются до рынка		Итого
	пешком	транспортом	
Мясо, рыба и прочее	18	13	31
Только овощи, фрукты	70	24	94
Итого	88	37	125

При существенности связи измерьте ее тесноту с помощью коэффициента ассоциации.

18. На основе однодневного выборочного обследования 350 студентов университета 10 октября получены следующие данные, в процентах ко всем обследованным студентам (табл. 9.19).

Таблица 9.19

Место жительства	Место питания			Итого
	дома	в столовой института	в другой столовой	
Семья	20	25	5	50
Общешитие	10	20	—	30
Частная квартира	10	5	5	20
Итого	40	50	10	100

С помощью критерия χ^2 определите, существует ли зависимость между местом жительства и местом питания студентов (на уровне значимости 0,05). При наличии связи вычислите коэффициент взаимной сопряженности Чупрова.

19. По данным выборочного обследования абитуриентов технического университета оцените, имеется ли зависимость между возрастом абитуриентов и степенью знакомства с избираемой инженерной профессией (принять уровень значимости 0,01) (табл. 9.20).

Таблица 9.20

Возраст, лет	Степень знакомства с профессией			
	Хорошо	Приблизи- тельно	Совсем не знаю	Итого
16–17	86	574	57	717
18–19	55	166	15	236
20–21	21	51	5	77
22–24	12	16	2	30
25 и более				
Итого	174	807	79	1060

20. По данным выборочного обследования абитуриентов Технического университета (табл. 9.21) установите на 5%-ном уровне значимости, имеется ли связь между полом абитуриента и желаемым местом будущей работы?

Таблица 9.21

Пол	Где бы хотел работать				
	На произ- водстве	В НИИ, КБ	В аспи- рантуре	Не знаю	Итого
Мужчины	207	164	37	83	491
Женщины	144	162	10	72	388
Итого	351	326	47	155	879

21. По данным выборочного обследования абитуриентов Технического университета (табл. 9.22) выявите наличие влияния того, откуда абитуриенты пришли в вуз, на их жизненные планы.

Таблица 9.22

Откуда пришел в вуз	Жизненные планы		
	Работать на государствен- ном предприятии	Работать в коммерческой структуре	Поступить в аспирантуру
Из школы	106	293	126
С производства	186	127	46
Из армии	34	22	4
Итого	326	442	176

В случае существенности связи ($\alpha = 0,05$) дайте характеристику тесноты связи.

22. По данным обследования семейных групп крупного города (табл. 9.23) установите, оказывает ли влияние частота контактов семей разных поколений на использование детских учреждений.

Таблица 9.23

Посещение детских учреждений дошкольниками	Частота контактов			
	Проживают совместно	Встречаются часто	Встречаются редко	Итого
Посещают детские учреждения	15	52	75	142
Не посещают детские учреждения	45	82	25	152
Итого	60	134	100	294

Ответ дайте на 5%-ном уровне значимости.

23. По данным обследования 540 семейных групп крупного города (табл. 9.24) (семья старшего поколения (родительская семья) – семья сына или дочери) установите, влияет ли социальная группа детей на характер желаемого расселения с семьей родителей ($\alpha = 0,05$).

Таблица 9.24

Желаемое расселение с семьей родителей	Социально-профессиональная группа сына или дочери			
	Служащие на должностях, требующих высшего образования	Служащие на должностях, требующих среднего специального образования	Квалифицированные и малоквалифицированные рабочие и служащие без специального образования	Итого
Вместе с родителями	14,5	24,4	32,1	22,5
В квартирах на одной лестнице, в соседнем доме, в соседних домах	52,3	41,9	41,5	45,9

Желаемое расселение с семьей родителей	Социально-профессиональная группа сына или дочери			
	Служащие на должнос- тях, тре- бующих высшего образова- ния	Служащие на должнос- тях, тре- бующих среднего специаль- ного обра- зования	Квалифици- рованные и малоквали- фицирован- ные рабочие и служащие без специ- ального обра- зования	Итого
В одном микрорайоне (квартале), в одном районе	22,1	25,6	10,7	20,2
Нежелательно совме- стное или близкое проживание	11,1	8,1	15,7	11,4
Итого	100	100	100	100
В процентах к итогу	33,0	28,5	38,5	100

9.2.4. Проверка гипотез о показателях корреляции и регрессии

24. По данным автохозяйств области было получено следующее уравнение регрессии, характеризующее зависимость коэффициента использования грузоподъемности автомобилей от удельного веса самосвалов в общем парке машин: $\hat{y}_x = 0,789 + + 0,384x$.

Оцените статистическую значимость полученных характеристик связи, если число автохозяйств равно 154, средний коэффициент использования грузоподъемности машины 36%, средний удельный вес самосвалов по всем автохозяйствам – 0,21 при среднем квадратическом отклонении 0,10.

25. Зависимость между затратами на производство 1 т зерна и размером основных фондов в расчете на одного работника АОЗТ описывается следующим уравнением регрессии: $\hat{Y}_x = 61,44 - - 2,47x$. Принимая $\alpha = 0,05$, оцените статистическую значимость этого уравнения, учитывая, что:

$$\bar{y} = 524 \text{ руб./т};$$

$$\sigma_y = 161 \text{ руб.};$$

$$\bar{x} = 426 \text{ руб.}$$

$$\sigma_x = 107 \text{ руб.}$$

26. При изучении факторов урожайности пшеницы в АОЗТ Северо-Западного федерального округа получены следующие данные (табл. 9.25).

Таблица 9.25

Фактор урожайности	1	2	3	4	5	6	7
Удельный вес посевов семенных районированных сортов, %	—	10	62	7	94	85	19
Средняя урожайность, ц/га	7,0	8,0	9,6	7,1	12,4	10,2	8,3

Постройте уравнение регрессии и вычислите коэффициент парной корреляции.

Приняв 5%-ный уровень значимости, оцените статистическую значимость показателей силы и тесноты связи.

27. По десяти однородным предприятиям были получены следующие данные о средней выработке продукции на одного работника и электровооруженности труда (табл. 9.26).

Таблица 9.26

Электровооруженность труда на 1 работника, кВт·ч	6	5	6	8	3	6	7	8	5	4
Средняя месячная выработка на 1 работника, тыс. руб.	72	60	80	100	40	75	85	110	70	50

Охарактеризуйте зависимость признаков с помощью уравнения линейной регрессии; вычислите коэффициент корреляции. Приняв 5%-ный уровень значимости, оцените существенность показателей силы (b_{yx}) и тесноты связи (r_{yx}).

28. При изучении зависимости заработной платы рабочих механического цеха от тарифного разряда получено уравнение регрессии:

$$\hat{y}_x = 98,5 + 13,6 x.$$

Средняя зарплата одного рабочего равна 15250 руб. ($\sigma_y = 420$ руб.), средний тарифный разряд – 3,2 ($\sigma_x = 1,01$).

Оцените существенность показателей корреляции и регрессии и уравнения регрессии в целом на 1%-ном уровне значимости.

29. При анализе деятельности семи торфопредприятий выявлена зависимость объема добычи торфа от стоимости основных фондов:

$$\hat{y}_x = 0,858 + 0,132 x, \quad r_{yx} = 0,95.$$

Средняя добыча торфа на одном предприятии составляла 9,9 тыс. т со средним квадратическим отклонением 7,40 тыс. т. Средняя стоимость основных фондов – 4 млн руб. со средним квадратическим отклонением 1,29 млн руб.

Определите статистическую значимость показателей корреляции и регрессии и статистическую значимость уравнения регрессии в целом.

Ответы дайте при $\alpha = 1\%$ и $\alpha = 5\%$.

30. Зависимость размера цеховых расходов на машиностроительных заводах от размеров фонда заработной платы производственных рабочих выражается уравнением регрессии

$$\hat{y}_x = 14,43 + 1,32 x.$$

Фонд зарплаты производственных рабочих в среднем на один завод составляет 126 млн руб. со средним квадратическим отклонением 51 млн руб., средний уровень цеховых расходов – 53 млн руб. с коэффициентом вариации 46%. С вероятностью 0,95 оцените статистическую значимость показателей связи.

31. На основании выборочного учета работников двух АОЗТ построены следующие распределения рабочих по уровню образования (табл. 9.27).

Таблица 9.27

Время обучения, лет	До 8	8	9	10	11	12	13
Предприятие 1	12	18	24	36	64	17	29
Предприятие 2	9	13	16	21	47	9	17

Определите средний уровень образования рабочих каждого предприятия. Существенно ли различаются эти средние?

Можно ли считать характер распределения рабочих по уровню образования одинаковым на двух предприятиях?

Существенно ли различие в вариации уровня образования рабочих на двух предприятиях?

32. По следующим данным определите, значима ли зависимость живого веса коровы от числа отелов (табл. 9.28).

Таблица 9.28

Живой вес коров, кг	Число отелов					Всего
	1	2	3	4	5	
290–310	1	1	–	–	–	2
310–330	1	1	–	1	1	4
330–350	4	3	1	–	1	9
350–370	9	6	3	1	3	22
370–390	15	12	5	2	4	38
390–410	22	14	8	4	6	54
410–430	22	16	12	9	10	69
430–450	18	17	14	14	15	78
450–470	13	14	16	15	17	75
470–490	12	11	15	18	6	62
490–510	8	6	11	14	4	43
510–530	3	4	7	7	3	24
530–550	2	2	5	2	1	12
550–570	1	1	2	1	1	6
570–590	–	–	–	1	1	2
Всего	131	108	99	89	73	500

Можно ли считать на уровне значимости 0,05, что распределение коров по живому весу подчиняется нормальному закону?

9.2.5. Дисперсионный анализ

33. По данным двух районов Северо-Запада России оцените существенность влияния урожайности картофеля на производительность труда в хозяйствах (табл. 9.29).

Таблица 9.29

Группа хозяйств по урожайности картофеля, ц/га	Число хозяйств	Производительность труда, ц/чел./день
До 140	8	2,83
140–180	11	3,63
180 и более	8	4,31
Итого	27	3,60

Общая величина колеблемости производительности труда (сумма квадратов отклонений от средней по всем единицам) равна 23,89.

34. По материалам бюджетных обследований в городе выявлены следующие данные (табл. 9.30).

Таблица 9.30

Число человек в домохозяйстве	1	2	3	4	5 и более
Число обследованных семей	17	43	45	36	8
Месячный доход на одного человека, руб.	13000	12500	11700	10100	9600

Можно ли считать, что распределение семей по их размеру подчиняется нормальному закону?

Определите, существует ли зависимость между размером семьи и душевым доходом, если среднее квадратическое отклонение душевого дохода по совокупности составляет 370 руб.

35. В результате обработки материалов бюджетного обследования семей рабочих и служащих завода получены следующие данные (табл. 9.31).

Таблица 9.31

Доход на одного члена семьи, руб.	Число обследованных семей, % к итогу	Среднее потребление молока за месяц одним членом семьи, л
До 3000	7	12,0
3000–4000	11	12,7
4000–5000	34	10,6
5000–7000	36	10,0
7000 и более	10	9,7

Можно ли считать, что распределение семей работников завода по доходу подчиняется нормальному закону?

Влияет ли уровень душевого дохода семьи на потребление молока, если среднее квадратическое отклонение душевого потребления молока по обследованным семьям составляет 4,3 л?

Оценку дайте на 5%-ном уровне значимости.

36. По 25 рабочим механического цеха собраны данные (табл. 9.32) о прохождении ими технического обучения и уровне производительности труда (среднее количество деталей, произведенных каждым рабочим за смену).

Таблица 9.32

Группа рабочих	Число рабочих	Выработка рабочих, количество деталей
Не проходивших техобучения	5	6,9,5,12,10
Проходящие техобучение	8	7,14,13,10,15,18,13,8
Прошедшие обучение	12	10,18,15,17,23,12,25,14,16,9,20,12

Применив дисперсионный анализ, определите, влияет ли техобучение на повышение производительности труда ($\alpha = 0,01$).

37. По следующим данным (табл. 9.33) оцените значимость зависимости выпуска продукции от оснащенности предприятий основными фондами, принимая $\alpha = 0,05$. Измерьте тесноту связи.

Таблица 9.33

Основные фонды на начало периода, млн руб.	Число предприятий	Продукция в сопоставимых ценах, млн руб.
200–400	5	340; 240; 350; 410; 260
400–600	6	610; 450; 606; 850; 880; 10200
600–800	4	496; 1070; 1240; 1380
800–1000	3	2070; 1820; 1420
Итого	18	

38. По следующим данным (табл. 9.34) дайте характеристику связи указанных признаков.

Таблица 9.34

Расход кормовых единиц на одну корову в год, ц	Число коров	Суточный удой, кг
28–34	5	10,6; 11,2; 12,0; 10,8; 11,4
34–40	8	13,9; 12,4; 12,8; 14,2; 14,9; 13,1; 12,1; 13,8
40–46	6	15,2; 16,4; 16,0; 16,5; 15,0; 15,7
Итого	19	

Определите, значительно ли влияние уровня кормления на продуктивность коров.

39. По следующим данным (табл. 9.35) охарактеризуйте связь указанных признаков. Существенно ли влияние уровня кормления на живой вес коров?

Таблица 9.35

Расход кормовых единиц на одну корову в год, ц	Число коров	Живой вес коровы, ц
28–34	5	394, 326, 359, 338, 413
34–40	8	423, 442, 460, 397, 419, 393, 431, 446
40–46	7	469, 479, 492, 498, 509, 543, 502
Итого	20	

40. По данным табл. 9.36 определите, какой тип уравнения регрессии наиболее пригоден для описания данной зависимости: уравнение прямой или парабола 2-го порядка? Дайте ответ на уровне значимости 0,01.

Таблица 9.36

Товарооборот магазина, млн руб.	Производительность труда продавцов, тыс. руб.	Товарооборот магазина, млн руб.	Производительность труда продавцов, тыс. руб.
2,042	90	4,515	250
2,580	130	5,170	240
3,120	160	5,502	280
3,578	180	6,184	290
4,066	200	6,513	300

41. По материалам обследования в крупном городе владельцев индивидуальных автомашин получены следующие данные (табл. 9.37).

Таблица 9.37

№ п/п	Время пути от дома до гаража, мин	Количество дней использования автомашины за месяц	№ п/п	Время пути от дома до гаража, мин	Количество дней использования автомашины за месяц
1	20	16	8	3	18
2	3	30	9	20	9
3	15	12	10	10	28
4	10	16	11	3	25
5	45	4	12	15	8
6	20	7	13	30	5
7	10	28	14	40	6

Измерьте связь между учтенными признаками.

Постройте уравнение регрессии, вычислите коэффициент корреляции.

Оцените существенность показателей силы и тесноты связи.

Определите, какой тип уравнения регрессии — прямая или парабола второго порядка — более пригоден.

Сделайте выводы на уровне значимости $\alpha = 1\%$ и $\alpha = 5\%$.

42. Изучается зависимость материалоемкости продукции от размера предприятия по 10 однородным заводам (табл. 9.38).

Таблица 9.38

Завод	1	2	3	4	5	6	7	8	9	10
x	9,0	6,0	5,0	4,0	3,7	3,6	3,5	6,0	7,0	3,5
y	100	200	300	400	500	600	700	150	120	250
Здесь x — расход материала на единицу продукции, кг; y — выпуск продукции, тыс. ед.										

Проверьте гипотезу о том, что $y = f(x)$ представляет собой равностороннюю гиперболу.

9.3. Вопросы для самоконтроля

1. Дайте определение статистической гипотезы.
2. Какие виды гипотез формулируются?
3. Каковы правила проверки статистической гипотезы?
4. Что такое критическая область и область принятия решения?
5. Какие виды ошибок возможны при испытании гипотез?
6. В чем состоят особенности проверки односторонней и двусторонней гипотез?
7. Как проводятся испытания гипотез о равенстве двух средних:
 - а) если генеральная дисперсия известна;
 - б) если генеральная дисперсия неизвестна.
8. Как производится проверка гипотезы о распределении?
9. В чем состоит преимущество непараметрических критериев?
10. В чем особенности проверки гипотезы о коэффициенте корреляции?
11. Как производится проверка гипотезы о статистической значимости уравнения регрессии в целом?
12. Каково содержание таблицы дисперсионного анализа?
13. Как построить F -критерий, используя коэффициент множественной корреляции?
14. В чем состоят особенности одновыборочного и двухвыборочного критериев знаков Вилкоксона?
15. Какова область применения критерия суммы рангов Вилкоксона и критерия Краскала-Уоллиса?

9.4. Методические рекомендации преподавателям

На занятиях по этой теме рекомендуем опираться на знания студентов по дисциплине «Теория вероятностей и математическая статистика». Обратите внимание студентов на специфику односторонней и двусторонней проверки, на особенности использования статистико-математических таблиц. Имеет смысл подчеркнуть то, что зависимость распределения критерия χ^2 от числа степеней свободы приводит по сути не к одному, а к семейству распределений χ^2 -квадрат (при разном числе

степеней свободы). Рекомендуем обратить внимание на взаимосвязь статистико-математических критериев. Целесообразно подчеркнуть вероятностный характер статистического вывода. Вместе с тем мы считаем необходимым отметить, что, несмотря на несовершенство статистических методов проверки гипотез, использование критериев весьма полезно для осмысления результатов расчетов и принятия решений в отношении дальнейших действий. Важно показать, что методы статистического оценивания и проверки гипотез являются составной частью статистического анализа данных.

9.5. Методические указания студентам

Прежде всего вы должны понимать связь между проверкой статистических гипотез и оценкой генеральных характеристик. Обратите внимание на формулировку испытуемой и альтернативной гипотез. Важно разобраться в ошибках первого и второго рода, понять, в чем состоят особенности односторонней и двусторонней проверки. Немаловажно выяснение взаимосвязей между статистическими критериями. В процессе занятий по данной теме вы должны научиться пользоваться статистико-математическими таблицами, корректно записывать проверяемую и альтернативную гипотезы, а также принятое решение в отношении нулевой гипотезы (принимается или не принимается).

Корреляционно-регрессионный анализ и моделирование статистических связей

10.1. Основные положения

Корреляционный анализ — это метод изучения тесноты связи между признаками. Все показатели корреляции изменяются по абсолютной величине от 0 до 1, чем ближе к 1, тем теснее связь.

Регрессионный анализ — это представление связи между признаками в виде функции

$$y = f(x),$$

где y — результативный признак (зависимая переменная);
 x — факторный признак (независимая переменная).

Задачи корреляционно-регрессионного анализа сводятся к следующему.

1. Выбрать тип уравнения, наилучшим образом выражающего зависимость средних значений результативного признака y от одного признака x (парная связь) или от нескольких факторных признаков (x_1, x_2, \dots, x_k) — многофакторная или множественная связь.

2. Измерить степень влияния вариации факторного признака (или признаков) на вариацию результативного признака, определить меру тесноты связи.

3. Ранжировать факторы x_1, x_2, \dots, x_k по степени влияния их вариации на вариацию результативного признака y .

4. Выбрать такой комплекс факторных признаков, вариация которых достаточно полно и надежно объясняет вариацию результативного признака, т.е. построить регрессионную модель связи.

5. На основе регрессионной модели прогнозировать с достаточно высокой вероятностью возможные значения результативного признака при ожидаемых значениях факторных признаков.

10.1.1. Парная корреляция и регрессия

Построение уравнения парной регрессии включает следующие этапы.

1. Выбор типа уравнения связи на основе графического изображения — точечной диаграммы в декартовых координатах (поле корреляции).

Например, по данным на 01.01.2001 г. представим на диаграмме связь между ростом и весом мужчин в возрасте от 18 до 31 года, однородных по занятию: футболисты команды «Зенит» (рис. 10.1).

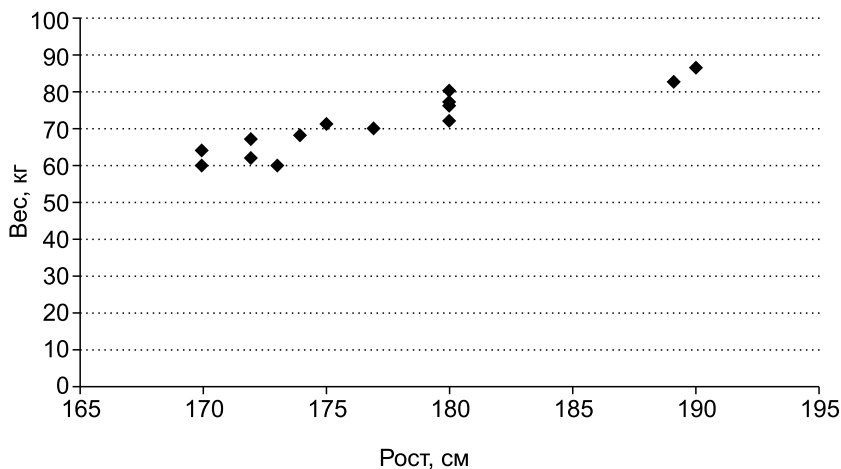


Рис. 10.1. Поле корреляции

График достаточно ясно указывает, что связь роста с весом в этой достаточно однородной группе мужчин имеет линейный характер, т.е. выражается уравнением прямой: $y = a + b \cdot x$.

2. Другой способ выявления формы связи между двумя признаками — построение аналитической группировки. Сгруппировав ту же совокупность мужчин по росту и выделив четыре группы, вычислим в каждой из групп средний вес, а затем, срав-

нивая последовательно группы, определим средний прирост веса на единицу прироста роста (кг на 1 см) (табл. 10.1). Если эти величины несущественно различаются между собой – связь можно считать линейной. Методика проверки существенности различия средних (дисперсионный анализ) излагается в теме 9.

Таблица 10.1

Дисперсионный анализ

Группа лиц по росту, см	Число лиц	Средний вес, кг	Средний рост, см	Прирост веса на 1 см роста к предыдущей группе, кг	Критерий различия средних	
					$F_{\text{факт}}$	$F_{\text{табл}}$
А	1	2	3	4	5	6
170–173	7	63,9	171,5	–		
174–175	9	69,4	175,5	1,38		
178–183	10	74,4	180,5	1,00	4,65	19,46
185–196	7	82,3	190,5	0,79		
Итого	33	72,5	179,3	1,06	–	–

Как видно из данных табл. 10.1 (графа 4), прирост веса на 1 см роста от группы к группе хотя и непостоянен, но различается на доли килограмма, в то время как в пределах одной и той же группы лиц по росту вариация веса достигает 6 кг. Дисперсия (взвешенная) веса между группами в расчете на 1 степень свободы (т. е. число групп минус 1) составила 1,47, а дисперсия тех же отношений различия веса к различиям роста внутри групп достигает 6,85 при 30 степенях свободы. Значит, большей в этом случае является внутригрупповая дисперсия, и тогда составляет ее отношение к межгрупповой, что дает $F_{\text{факт}} = 4,65$. Табличное значение критерия при уровне значимости 0,05, 30 степенях свободы для большей дисперсии и 2 степенях свободы для меньшей дисперсии составляет $F_{\text{табл}} = 19,46$. Следовательно, вероятность «нулевой гипотезы» о несущественности различия средних превышает $(1 - 0,05) = 0,95$, или 95%. Можно считать различия статистически несущественными, а значит, связь роста и веса в этой выборке линейная.

3. Третий прием выявления формы связи – перебор на ПЭВМ различных форм уравнений связи и выбор на основе сравнения

либо коэффициентов детерминации (в пользу большего), либо по остаточной дисперсии и средней ошибке аппроксимации (в пользу меньших). Следует помнить, что чем выше порядок параболы или чем больше число параметров уравнения, тем автоматически меньшей становится остаточная дисперсия. При параболе 2-го порядка она всегда меньше, чем при прямолинейной модели, и это не всегда означает, что парабола 2-го порядка лучше отражает форму связи. Только если снижение остаточной дисперсии при параболе является статистически существенным, можно предпочесть параболу 2-го порядка. Кривые высоких порядков вообще не рекомендуется использовать, особенно при небольших объемах выборки: их параметры будут ненадежны.

После выбора вида уравнения регрессии необходимо произвести вычисление параметров уравнения. Эта задача решается методом наименьших квадратов. Однако данный метод дает состоятельные и несмещенные оценки, если распределение совокупности по факторному и по результативному признакам нормальное. Поэтому сначала следует проверить, выполняется ли это условие: либо по показателям асимметрии и эксцесса распределения, либо по критерию хи-квадрат К. Пирсона. Программы статистического анализа для ПЭВМ дают возможность быстро получить оценки показателей асимметрии и эксцесса, их средних ошибок и значения критерия t -Стьюдента. Нужно рассчитать отношение показателей к их средним ошибкам, часто называемым «стандартизированными оценками асимметрии и эксцесса». Если значение критерия t -Стьюдента меньше табличных значений, можно считать, что распределения близки к нормальному. Если же асимметрия и эксцесс либо один из этих показателей существенно отличны от нуля, лучше провести «нормализацию» распределения, исключив из выборки единицы, значения признаков у которых отклоняются от средних более чем на два средних квадратичных отклонения.

После «нормализации» совокупности вычисляются параметры линейной связи $\hat{y} = a + b \cdot x$ по формулам

$$b = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2};$$

$$a = \bar{y} - b\bar{x}.$$

Коэффициент корреляции

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}.$$

Для экономии места приведем расчет не по всем 33 членам команды «Зенит», а по 16 членам — первым из случайного, например алфавитного списка (табл. 10.2).

Средние значения составили:

$$\bar{x} = 177,4 \text{ см};$$

$$\bar{y} = 71,31 \text{ кг}.$$

Для удобства последующих расчетов средние можно округлить до целых. На параметрах корреляции округление почти не отражается.

Получаем: $b = 659/535 = 1,23$; $a = 71,31 - 1,23 \cdot 177,44 = -146,9$.

Уравнение регрессии: $\hat{y} = 1,23x - 146,9$.

Коэффициент корреляции

$$r_{xy} = \frac{659}{\sqrt{535 \cdot 947}} = 0,926.$$

Параметр b означает, что в среднем с увеличением роста на 1 см вес футболистов возрастал на 1,23 кг. Значение коэффициента корреляции указывает на тесную связь между ростом и весом.

Рассчитав по уравнению регрессии значения веса при данном росте каждого лица, округлим их также до целых и запишем в графу 9 табл. 10.2. Отклонения фактического веса от расчетного ($y - \hat{y}$) представляют собой необъясненную вариацию веса. Возведя их в квадрат, можно вычислить корреляционное отношение по формуле

$$\eta = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}} = \sqrt{1 - \frac{146}{947}} = 0,920.$$

Полученное значение корреляционного отношения отличается от коэффициента корреляции только в третьем знаке в результате округлений. Квадрат этого показателя или коэффициента корреляции есть коэффициент детерминации, показы-

Таблица 10.2

Расчет параметров парной корреляции

Номер единицы	Рост, см x	Вес, кг y	Δx	Δy	$\Delta x \cdot \Delta y$	Δx^2	Δy^2	\hat{y}	$y - \hat{y}$	$(y - \hat{y})^2$
1	172	62	-5	-9	45	25	81	65	-3	9
2	170	59	-7	-12	84	49	144	62	-3	9
3	189	82	+12	+11	132	144	121	85	-3	9
4	170	64	-7	-7	49	49	49	62	2	4
5	177	70	0	-1	0	0	1	71	-1	1
6	172	67	-5	-4	20	25	16	65	2	4
7	177	70	0	-1	0	0	1	71	-1	1
8	190	86	+13	+15	195	169	225	87	-1	1
9	180	77	+3	+6	18	9	36	74	3	9
10	180	72	+3	+1	3	9	1	74	-2	4
11	175	71	-2	0	0	4	0	68	3	9
12	180	76	+3	+5	15	9	25	74	2	4
13	180	77	+3	+6	18	9	36	74	3	9
14	174	68	-3	-3	9	9	9	67	1	1
15	180	80	+3	+9	27	9	81	74	6	36
16	173	60	-4	-11	44	16	121	66	-6	36
Σ	2839	1141	----	----	659	535	947	1139		146

вающий, какая доля вариации веса связана с вариацией роста. Эта доля составляет 85%, средняя абсолютная ошибка аппроксимации составила:

$$\frac{\sum |y_i - \hat{y}_i|}{n} = \frac{42}{16} = 2,62 \text{ кг},$$

относительная ошибка (в процентах к средней) = $2,62/71,3 = 0,037$, или 3,7%.

Мы выполнили две первые задачи корреляционно-регрессионного анализа. Третья и четвертая задачи относятся к многофакторным системам, о которых сказано далее. Задача 5 – прогнозирование, требует расчета средней ошибки прогноза по линейному тренду. Эта ошибка зависит от объема выборки n , от величины остаточной (необъясненной) вариации y , и от того, насколько удалено значение прогнозируемого фактора $x_{\text{пр}}$ от средней величины \bar{x} . Имеем среднюю ошибку прогноза:

$$m_y = s_{\text{остат}} \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x^2}},$$

$S_{\text{остат}} = \sqrt{146/(16-2)}$, где $(16-2)$ – степени свободы остаточной вариации. $S_{\text{остат}} = 3,23$ кг. Если мы прогнозируем вес для футболиста ростом 165 см, то среднее ожидаемое значение его веса должно быть: $1,23 \cdot 165 - 146,9 = 56$ кг. С вероятностью 0,95 вес будет находиться в границах средней ожидаемой величины $\pm 2,14 m_y$, где 2,14 – это критерий Стьюдента для надежности 0,95 (т.е. значимости нулевой гипотезы 0,05) при $16 - 2 = 14$ степенях свободы.

Получаем границы веса:

$$56 \pm 2,14 \cdot 3,23 \sqrt{1 + \frac{1}{16} + \frac{(165 - 177,4)^2}{535}} = 56 \pm 8 \text{ кг},$$

или от 48 до 64 кг (с вероятностью 0,95).

Широкие границы прогноза связаны с малой величиной выборки и с желанием иметь высокую надежность прогноза.

Параболическая 2-го порядка зависимость характерна для такого влияния факторного признака на результативный, при котором до определенной величины фактора результативный

признак возрастает, а затем уменьшается, либо до определенной величины фактора уменьшается, а при дальнейшем росте фактора возрастает. В общем виде параболическая регрессия имеет вид: $y = a + bx + cx^2$. Такова, например, зависимость дохода физических лиц от их возраста.

Гиперболическая зависимость характерна для такого влияния факторного признака на результативный, при котором это влияние убывает по мере роста значений фактора. Например, при увеличении объема выпуска продукции на предприятии (переходе от мелкого производства к крупному, массовому) себестоимость единицы продукции сначала снижается резко, а с дальнейшим увеличением объема продукции — снижается все медленнее. Ведь часть затрат (затраты на основные материалы) не может снизиться — изделие требует определенного количества материала. Другая часть затрат (прежде всего накладные затраты, не зависящие от числа изделий) при делении на возрастающее число изделий снижается в расчете на единицу изделия по логарифмическому закону в уменьшающейся степени. Пусть, например, основные, неизменные затраты на одно изделие равны 100, а общая сумма накладных затрат равна 2000 денежных единиц. При выпуске 20 единиц продукции себестоимость составит: $100 + 2000/20 = 200$ единиц. При выпуске 40 единиц продукции: $100 + 2000/40 = 150$ единиц, т.е. на 50 единиц меньше. При дальнейшем таком же увеличении выпуска еще на 20 единиц, т.е. до 60, себестоимость составит: $100 + 2000/60 = 133$, или на 17 единиц меньше и т.д. Форма связи имеет вид: $y = a + b/x$, т.е. является гиперболой. Методика расчета параметров сводится к линеаризации путем замены факторного признака x на обратную величину: $1/x$. Система нормальных уравнений МНК для гиперболы после линеаризации та же, что и для прямой:

$$\begin{cases} na + b \sum \frac{1}{x_i} = \sum y_i; \\ a \sum \frac{1}{x_i} + b \sum \left(\frac{1}{x_i} \right)^2 = \sum \frac{y_i}{x_i}. \end{cases}$$

Другие, более сложные формы уравнений связи имеет смысл применять только с помощью статистических программ для ЭВМ.

10.1.2. Множественная корреляция и регрессия

Как правило, результативный признак зависит не от одного фактора, а от большого числа факторов. Однако не о всех факторах имеется надежная информация и не все факторы, по которым информация имеется, можно включить в анализ. Во-первых, ограничения на число включаемых факторов налагает объем совокупности (выборка). Поскольку связь не функциональная, а стохастическая, требуется для ее надежного измерения, чтобы число единиц совокупности хотя бы в 5–6 раз, а лучше — в десятки раз превосходило число факторов (объясняющих переменных). Во-вторых, не все факторы имеют распределение, хотя бы близкое к нормальному закону, и не всегда соблюдается постоянство вариации фактора и результативного признака по всей области их существования (независимость вариации y при всех значениях x), т.е. гомоскедастичность.

Предпочтительна линейная форма связи факторов с результативным признаком, при котором уравнение связи является уравнением плоскости (при двух факторах) или гиперплоскости (более двух факторов) вида

$$\hat{y} = a + \sum b_j \cdot x_j.$$

Явно нелинейно влияющие факторы следует преобразовать (провести линеаризацию связи x с результативным признаком). Каждый из коэффициентов уравнения b_j измеряет в абсолютном выражении отклонение результативного признака от его средней величины при отклонении фактора x_j от своей средней на единицу измерения фактора при постоянных, закрепленных на среднем уровне значениях остальных факторов, входящих в уравнение связи. Следовательно, в отличие от парной корреляции и регрессии коэффициенты многофакторной регрессии свободны от прямого влияния других, входящих в уравнение факторов, и могут быть названы *коэффициентами условно-чистой регрессии*.

В многофакторном корреляционно-регрессионном анализе главное внимание уделяется заданиям 3–5, указанным в начале раздела.

Чтобы ранжировать факторы по степени их влияния на вариацию результативного признака, нельзя использовать сами коэффициенты регрессии, выраженные именованными числами несоизмеримых признаков и зависящие от принятых единиц измерения. Для приведения коэффициентов в сравнимый вид применяются два способа их преобразования: в коэффициенты стандартизованной регрессии β_j или в коэффициенты эластичности e_j .

Бета-коэффициент вычисляется по формуле

$$\beta_j = b_j \cdot \frac{\sigma_{x_j}}{\sigma_y}$$

и измеряется в сигмах y , т.е. в отвлеченном числе, обычно меньшем единицы.

Коэффициент эластичности вычисляется по формуле

$$e_j = b_j \cdot \frac{\bar{x}_j}{\bar{y}}$$

и также выражается отвлеченным числом, интерпретируемым как среднее изменение результативного признака в процентах на 1% изменения фактора x_j при постоянстве остальных факторов.

Для измерения тесноты связи применяются различные формулы коэффициента множественной детерминации.

Коэффициент множественной детерминации $R^2_{yx_1 \dots x_k}$, измеряющий долю вариации результативного признака y за счет всего комплекса факторов, включенных в анализ, вычисляется по формуле

$$R^2_{yx_1 \dots x_k} = \frac{\sum_{(i)} (\hat{y}_i - \bar{y})^2}{\sum_{(i)} (y_i - \bar{y}_i)^2},$$

т.е. R^2 — отношение вариации y , объясненной за счет вариации факторов ко всей вариации результативного признака.

Иная формула:

$$R^2_{yx_1 \dots x_k} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2},$$

т.е. R^2 представлен как единица минус отношение необъясненной (остаточной) вариации y к его общей вариации.

Можно вычислить коэффициент множественной детерминации и через парные коэффициенты корреляции и бета-коэффициенты, как:

$$R^2_{yx_1 \dots x_k} = \sum_1^k r_{yx_j} \beta_j.$$

Для проверки надежности установления связи вариации результативного признака со всем комплексом факторов вычисляется его средняя ошибка m_{R^2} по формуле

$$m_{R^2} = \sqrt{\frac{R^2 (1 - R^2)}{n - k - 1}},$$

после чего рассчитывается t -критерий Стьюдента: $t = \frac{R^2}{m_{R^2}}$.

Полученное значение $t_{\text{факт}}$ сравнивается с табличным значением на уровне значимости 0,05 и при $df = n - k - 1$ степенях свободы.

Другой способ проверки надежности связи — дисперсионный анализ, включающий расчет критерия Фишера. Доказано, что мерой чистого влияния каждого из факторов на вариацию результативного признака служит квадрат бета-коэффициента.

Поэтому общее влияние комплекса факторов $R^2_{yx_1 \dots x_k}$ можно разложить на сумму бета-квадратов и системный эффект η^2 , измеряющий влияние взаимосвязей между факторами, их взаимодействия в системе. Системный эффект может быть как положительной величиной, так и отрицательной, в зависимости от того, преоб-

ладает ли прямое взаимодействие факторов (они как бы «помогают» друг другу) или обратное взаимодействие (они как бы «мешают» друг другу).

Для решения вопроса о целесообразности добавления фактора применяются коэффициенты частной корреляции (детерминации). При двухфакторной системе они вычисляются на основе парных коэффициентов корреляции по формуле

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}.$$

При большом числе факторов частный коэффициент корреляции может рассчитываться на основе коэффициентов детерминации. Например, для фактора x_4

$$r_{yx_4x_1x_2x_3}^2 = R_{yx_1x_2x_3}^2 - R_{yx_1x_2x_3}^2 / 1 - R_{yx_1x_2x_3}^2.$$

Покажем расчет с использованием цифровых значений: $0,80 - 0,65/1 - 0,65 = 0,43$.

Интерпретация частного коэффициента детерминации такова: при включении фактора x_4 в уравнение регрессии после факторов x_1, x_2, x_3 необъясненная ими вариация сокращается с 0,35 до 0,2, или на 43%. Видимо, включение фактора x_4 в модель целесообразно.

Рассмотрим пример анализа многофакторной корреляционно-регрессионной системы связей коэффициента рентабельности банков с факторами: x_1 — доля вложений в государственные ценные бумаги от актива, %; x_2 — отношение непроцентного дохода к процентному (в размах); x_3 — коэффициент полной ликвидности обязательств банка (табл. 10.3).

Таблица 10.3

Многофакторный анализ

Номер банка	y_i , %	x_{1i} , %	x_{2i} , число раз	x_{3i} , число раз	Расчет- ные зна- чения y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	13,5	24,0	2,5	1,27	8,1	5,4	29,16
2	25,5	51,0	4,5	1,97	20,0	5,5	30,25
3	1,2	10,4	2,5	2,15	7,6	-6,4	40,96
4	1,3	14,1	1,6	1,27	4,8	-3,5	12,25
5	4,5	4,7	0,3	1,34	1,9	2,6	6,76
6	2,7	15,8	0,5	0,97	3,8	-1,1	1,21
7	12,2	29,2	0,5	1,15	8,9	3,3	10,89
8	4,2	31,0	6,6	1,07	10,0	-5,8	33,64
9	4,4	13,5	1,0	1,08	3,7	0,7	0,49
10	2,1	2,2	0,6	1,36	1,2	0,9	0,81
11	7,5	50,3	2,1	1,11	15,6	8,1	65,61
12	14,4	28,3	7,2	1,18	9,8	4,6	21,16
13	11,4	30,4	1,2	1,10	9,1	2,3	5,29
Сумма	104,9	304,9	31,1	17,02	104,5	—	258,48
Сред- нее	8,07	23,47	2,39	1,31			
СКО	6,80	14,83	2,22	0,342			

Матрица коэффициентов корреляции каждого фактора с результатом и межфакторной корреляции представлена в табл. 10.4.

Таблица 10.4

Матрица коэффициентов корреляции

Признак	y	x_1	x_2	x_3
y	1	0,718	0,397	0,241
x_1		1	0,462	0,085
x_2			1	0,134
x_3				1

Как видим, факторов, коллинеарных между собой (с коэффициентом корреляции более 0,8), нет. Коэффициент множественной детерминации

$$R^2_{yx_1x_2x_3} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{n \cdot \sigma_y^2} = 1 - \frac{258,48}{13 \cdot 6,80^2} = 0,5700.$$

Скорректированный на потерю степеней свободы коэффициент детерминации составил: $R_{\text{корр}}^2 = 1 - (1 - R_{\text{некорр}}^2) \cdot n - 1 / n - k - 1 = 1 - (1 - 0,570) \cdot 12 / 9 = 0,427$.

Большое снижение скорректированного коэффициента детерминации в сравнении с первичным коэффициентом объясняется малым объемом выборки (что объяснимо учебным характером примера). Соответственно скорректированный коэффициент множественной корреляции составит $\sqrt{0,427} = 0,653$.

Получим следующее уравнение регрессии:

$$\hat{y} = -5,495 + 0,317x_1 + 0,154x_2 + 4,38x_3.$$

Критерий Стьюдента для коэффициентов регрессии составляет: $t_{b_1} = 2,8$; $t_{b_2} = 0,20$, $t_{b_3} = 0,98$. Таким образом, вполне надежно доказано влияние только первого фактора, влияние факторов x_2 и x_3 явно ненадежно при такой малой выборке. Влияние комплекса факторов определяется при помощи дисперсионного анализа (табл. 10.5).

Таблица 10.5

Дисперсионный анализ

Вариация y	Сумма квадратов	Число степеней свободы	На одну степень свободы	F -критерий	P
Факторная	341,58	3	113,86	—	—
Остаточная	258,48	9	28,72	3,96	0,047
Всего	600,06	12	—	—	—

Вероятность нулевой гипотезы об отсутствии влияния комплекса факторов на вариацию рентабельности банков мала, так что можно считать влияние доказанным.

Рассчитаем стандартизированные коэффициенты регрессии и коэффициенты эластичности:

$$\beta_1 = b_1 \cdot \frac{\sigma_{x1}}{\sigma_y} = 0,317 \frac{14,83}{6,80} = 0,6913;$$

$$\beta_1^2 = 0,4779;$$

$$\beta_2 = b_2 \cdot \frac{\sigma_{x2}}{\sigma_y} = 0,154 \frac{2,22}{6,80} = 0,050;$$

$$\beta_2^2 = 0,0025;$$

$$\beta_3 = b_3 \frac{\sigma_{x3}}{\sigma_y} = 4,38 \frac{0,342}{6,80} = 0,220;$$

$$\beta_3^2 = 0,485;$$

$$e_1 = b_1 \frac{\bar{x}_1}{\bar{y}} = 0,317 \frac{23,47}{8,07} = 0,922;$$

$$e_2 = b_2 \frac{\bar{x}_2}{\bar{y}} = 0,154 \frac{2,39}{8,07} = 0,046;$$

$$e_3 = b_3 \frac{\bar{x}_3}{\bar{y}} = 4,38 \frac{1,31}{8,07} = 0,711.$$

Судя как по бета-коэффициентам, так и по коэффициентам эластичности, наибольшее влияние оказывает вариация первого фактора, затем — третьего и самое слабое — вариация второго фактора. Для первого фактора и особенно для третьего коэффициенты эластичности значительно больше, чем бета-коэффициенты. Это связано с тем, что вариация этих факторов в относительном, сравнимом выражении значительно слабее, чем вариация рентабельности. Для второго фактора соотношение противоположное.

Рассчитаем показатель влияния системного эффекта факторов на вариацию рентабельности:

$$\eta = R^2 - \sum \beta_j^2 = 0,5700 - 0,4779 - 0,0025 - 0,0485 = 0,0411.$$

Подчеркнем, что в расчете необходимо использовать некорректированный коэффициент детерминации либо квадраты бета-коэффициентов корректировать в том же отношении, как и R^2 . В последнем случае величина системного эффекта, скорректированного на потерю степеней свободы, составит 0,0308.

Имеется слабый прямой системный эффект, т.е. факторы взаимодействуют в направлении повышения результативного признака (так как между факторами имеется прямая связь, хотя и слабая).

Частные коэффициенты корреляции составили:

для фактора x_1 $r_{yx_1 \cdot x_2 x_3} = 0,6846$;

для фактора x_2 $r_{yx_2 \cdot x_1 x_3} = 0,069$;

для фактора x_3 $r_{yx_3 \cdot x_1 x_2} = 0,3116$.

Исключив несущественный и ненадежный фактор x_2 , получаем уравнение регрессии

$$\hat{y} = -5,54 + 0,3276x_1 + 4,51x_3 ,$$

коэффициент детерминации $R^2_{yx_1x_3} = 0,5668$ и скорректированное значение $R^2 = 0,4802$. Как видим, при очень незначительной потере значения первоначального R^2 , при меньшем числе факторов, его скорректированное значение даже возросло с 0,427 до 0,480, или на 53 пункта. Критерий Фишера для двухфакторной модели возрос до 6,54, а вероятность нулевой гипотезы снизилась до 0,015.

Для прогнозирования ожидаемого значения рентабельности y по этому уравнению выбираем значения факторов x_1 и x_3 , находящиеся в границах размаха вариации и превышающие средние величины («прогрессивные»), например $x_1 = 40$, $x_3 = 1,8$. Подставляя эти значения в уравнение двухфакторной модели, имеем

$$\hat{y} = -5,54 + 0,3276 \cdot 40 + 4,51 \cdot 1,8 = 15,67\%$$

Средняя ошибка аппроксимации (прогноза) составила 5,1%. Из-за малого объема выборки нельзя получить прогноз в достаточно узких границах и с высокой надежностью. Если бы объем выборки был в 5 раз больше, то ошибка была бы прибли-

зительно в $\sqrt{5}$ раз меньше, т.е. 2,28%. Тогда с вероятностью 0,95 при t -Стюдента = 2,16 доверительные границы прогноза составили бы от 15,67% – 2,16 · 2,28% до 15,67% + 2,16 · 2,28%, или рентабельность банков можно ожидать в пределах от 10,75 до 20,59%.

10.2. Задания по теме

При выполнении заданий **1–10** необходимо сделать следующее:

- 1) обосновать выбор факторного и результативного признаков;
- 2) изобразить связь признаков на графике (поле корреляции);
- 3) обосновать выбор типа уравнения регрессии;
- 4) вычислить параметры уравнения регрессии и дать их интерпретацию;
- 5) проверить надежность параметров уравнения регрессии, оценив вероятность нулевой гипотезы;
- 6) вычислить коэффициенты детерминации и корреляции (либо корреляционное отношение); дать их интерпретацию;
- 7) рассчитать одно-два прогнозируемых значения результативного признака при заданном значении фактора.

1. Проведите анализ связи между ожидаемой продолжительностью жизни и душевым потреблением мяса в 20 странах (табл. 10.6).

Таблица 10.6

Страна	Продолжительность жизни, лет	Потребление мяса, кг/чел.
Австрия	77,0	95
Австралия	78,2	104
Белоруссия	68,0	59
Великобритания	77,2	72
Венгрия	75,7	59
Германия	78,2	86
Дания	75,7	98
Италия	78,2	79
Казахстан	67,6	50
Канада	79,0	98
Латвия	68,4	56
Нидерланды	77,0	88
Россия	66,6	46
Румыния	69,9	43
США	76,7	114
Украина	68,8	37
Финляндия	76,8	63
Франция	78,1	91
Чехия	73,9	70
Швейцария	78,6	56

Источник: Вопросы статистики. — 2000. — № 3.

2. Проведите анализ связи между надоем молока на одну сред-негодовую корову и себестоимостью молока по данным хозяйств одного района (табл. 10.7).

Таблица 10.7

Хозяйство	Надой молока, ц/гол.	Себестоимость, руб./ц
1	30	158
2	75	144
3	30	287
4	42	158
5	55	140
6	30	166
7	27	243
8	45	128
9	38	147
10	27	149
11	56	158
12	35	134
13	30	171
14	38	149

3. Проведите анализ связи между числом мячей, забитых фут-больными командами в ворота соперников, и числом получен-ных командами очков по итогам чемпионата России по футболу 1999 г. (табл. 10.8).

Таблица 10.8

Команда	Забито мячей	Получено очков
1	75	72
2	62	65
3	56	55
4	38	50
5	42	44
6	54	43
7	32	41
8	36	39
9	27	36
10	30	34
11	33	33
12	39	31
13	36	31
14	30	29
15	29	26
16	21	24

4. Проведите анализ между показателями деятельности банков России за I квартал 2000 г., 20%-ная выборка (табл. 10.9).

Таблица 10.9

Банк	Достаточность капитала	Текущая ликвидность
Газпромбанк	14	1,17
Росбанк	15	1,04
Башкредитбанк	22	0,79
Еврофинанс	38	0,72
Петрокоммерц	15	0,94
Возрождение	19	0,57
Конверсбанк	40	0,82
Петровский	10	0,61
Приобье	18	0,91
Красбанк	28	1,02
Пробизнесбанк	18	0,84
Мостстритэкономбанк	39	0,79
Евростат	25	0,78
Балтийский	13	0,61
Газбанк	17	0,72
Солидарность	11	0,92
Авангард	28	0,98
Тольяттихимбанк	15	1,26
РБИО	90	2,52
Сахадаймаондбанк	14	1,53

Источник: Известия. — 2000. — 18 марта.

5. Проведите анализ связи между внутренним валовым продуктом на душу населения и числом медалей, полученных спортсменами страны на Олимпийских играх 2000 г. (первая двадцатка страны) (табл. 10.10).

Таблица 10.10

Страна	ВВП/чел., долл.	Число медалей
США	25536	97
Россия	3962	88
Китай	2772	59
Австралия	19030	58
Германия	19295	57
Франция	19719	38
Италия	18676	34

Продолжение

Страна	ВВП/чел., долл.	Число медалей
Нидерланды	18997	25
Куба	2361	29
Великобритания	18241	28
Румыния	4243	26
Южная Корея	12602	28
Венгрия	6371	17
Польша	5529	14
Япония	21121	18
Болгария	4043	13
Греция	11165	13
Швеция	17645	12
Норвегия	21416	10
Эфиопия	637	8

6. Проведите анализ связи между уровнем ВВП на душу населения и средней ожидаемой продолжительностью жизни в 2000 г. (табл. 10.11).

Таблица 10.11

Страна	ВВП/чел., долл.	Продолжительность жизни, лет
Австрия	19970	77,0
Австралия	19030	78,2
Белоруссия	4009	68,0
Великобритания	18241	77,2
Венгрия	6371	70,9
Германия	19295	77,2
Дания	21078	75,7
Италия	18676	78,2
Казахстан	2330	67,6
Канада	20331	79,0
Латвия	3322	68,4
Нидерланды	18997	77,9
Россия	3962	66,6
Румыния	4243	69,9
США	25 536	76,7
Украина	2049	68,8
Финляндия	17239	76,8
Франция	19719	78,1
Чехия	9417	73,9
Швейцария	22 780	78,6

7. Проведите анализ между расходом домохозяйств США на питание и возрастом главы домохозяйства. Общая дисперсия расходов на питание равна 10,2 (табл. 10.12).

Таблица 10.12

Возраст главы домохозяйства, лет	Расход на питание, тыс. долл.
До 25	2,8
25–35	4,2
35–45	5,4
45–55	5,6
55–65	4,5
65 и старше	3,3
Итого	4,4

8. Проведите анализ связи между числом сотрудников аудиторских фирм России и выручкой фирм (табл. 10.13).

Таблица 10.13

Фирма	Число сотрудников, чел.	Выручка, млн руб.	Фирма	Число сотрудников, чел.	Выручка, млн руб.
1	287	77,1	16	52	9,1
2	217	73,5	17	22	9,0
3	309	35,6	18	26	6,7
4	155	29,5	19	57	6,2
5	267	22,5	20	71	5,7
6	110	21,6	21	55	4,6
7	39	20,7	22	54	4,4
8	66	19,3	23	23	4,1
9	153	16,9	24	49	4,1
10	116	14,6	25	20	4,1
11	91	14,3	26	52	4,1
12	73	12,6	27	27	3,9
13	77	12,2	28	38	3,5
14	68	12,0	29	65	3,4
15	46	10,3	30	35	3,3

Источник: Известия. — 1999. — 16 дек.

9. Проведите анализ связи между числом специалистов и выручкой аудиторских фирм за 2000 г. (табл. 10.14).

Т а б л и ц а 10.14

Фирма	Число специалистов, чел.	Выручка, млн. руб.
Аудитинформ	86	29,9
ИНЭК	68	25,0
Мариллон	38	24,2
Налоги России	51	22,7
ЦБА	35	22,7
Аудит-консалтинг РНЕ	60	20,5
Мауэр-Аудит	21	20,3
Донар	20	18,0
2 К Аудит-Деловая консультация	27	17,8
СТЕК	30	17,7
ВВД	72	17,6
Сиотт, Ригн и Флетчер аудит ЛТД	62	17,5
Пачолли	20	16,7
МББ-Аудит	34	16,2
«Финансы»	42	16,2

И с т о ч н и к : Известия. — 2001. — 22 марта.

10. Проведите анализ связи между размером капитала и прибылью 15 компаний США за 1995 г. (табл. 10.15).

Т а б л и ц а 10.15

Компания	Капитал, млрд долл.	Прибыль, млрд долл.
General Electric	106,4	5,92
American Telephone and Telegraph	101,5	4,71
Exxon Corporation	94,8	5,10
Coca-Cola	90,7	2,55
Merck	71,1	3,00
Philip Morris	70,9	4,72
Microsoft Corporation	58,8	1,45
Intel Corporation	57,5	2,29
Proctor and Gamble	55,6	2,64
International Business Machines	55,3	3,02

Продолжение

Компания	Капитал, млрд долл.	Прибыль, млрд долл.
Johnson & Johnson	52,8	2,00
Wall-Mart Stores	49,6	2,68
Hewlett Packard	47,4	1,60
Do Pont de Nemours	42,5	2,73
PepsiCo	41,4	1,78

Источник: Financial Times. – 1996. – Март.

При выполнении заданий **11 – 21** необходимо сделать следующее:

1) проверить близость распределения совокупности по всем признакам к нормальному, исключив, если нужно, резко отклоняющиеся единицы;

2) вычислить параметры уравнения множественной регрессии. Интерпретировать параметры уравнения;

3) вычислить стандартизированные коэффициенты регрессии и коэффициенты эластичности, интерпретировать их;

4) сделать выводы о степени надежности уравнения регрессии в целом и каждого из коэффициентов регрессии;

5) вычислить коэффициенты множественной детерминации и корреляции для комплекта факторов и сделать выводы по их значениям. Дать характеристику надежности связи по критерию Фишера;

6) разложить коэффициент множественной детерминации на составляющие, выделив меру «системного эффекта» и дать интерпретацию роли факторов и системного эффекта;

7) сделать прогноз ожидаемого значения результативного признака при допустимых значениях факторных признаков, используя показатель средней ошибки аппроксимации.

11. Проведите анализ зависимости валового регионального продукта (ВРП) от ряда факторов по регионам Урала и Сибири, 1999 г. (табл. 10.16).

Таблица 10.16

Регион	ВРП, млрд руб.	Среднего- довая чис- ленность занятых, тыс. чел.	Среднеме- сячная зарплата занятых, руб. / чел.	Среднегодо- вая стои- мость основ- ных фондов, млрд руб.
Республика Башкортостан	64,2	1666	1260	382
Удмуртская Республика	20,0	701	1183	145
Курганская обл.	9,7	421	1014	88
Оренбургская обл.	28,8	930	1273	201
Пермская обл.	55,6	1275	1719	277
Свердловская обл.	80,7	1953	1560	333
Республика Алтай	1,6	74	940	7
Алтайский край	22,4	1050	974	207
Кемеровская обл.	45,6	1249	1725	290
Новосибирская обл.	35,2	1001	1272	264
Омская обл.	29,5	937	1165	191
Томская обл.	21,2	446	1694	143
Тюменская обл.	201,2	1676	3885	904
Республика Бурятия	11,2	370	1385	85
Республика Тыва	1,8	108	1031	14
Республика Хакасия	8,2	228	1700	59
Красноярский край	71,6	1371	2048	396
Иркутская обл.	52,6	1100	1968	322
Читинская обл.	13,2	428	1670	94
Республика Саха (Якутия)	33,4	481	2994	177

12. Проведите анализ зависимости валового регионального продукта от ряда факторов по регионам Центра, Волго-Вятского, Центрально-Черноземного и Приволжского экономических районов, 1999 г. (табл. 10.17).

Таблица 10.17

Регион	ВРП, млрд руб.	Среднего- довая чис- ленность занятых, тыс. чел.	Среднеме- сячная зарплата занятых, руб. / чел.	Среднего- довая стоимость основных фондов млрд руб.
Тульская обл.	19,1	765	1081	150
Ярославская обл.	22,3	635	1306	168
Республика Марий Эл	6,6	314	896	64
Республика Мордовия	9,3	393	468	69
Чувашская Республика	12,1	554	851	104
Кировская обл.	16,9	682	1060	152
Нижегородская обл.	52,7	1643	1219	287
Белгородская обл.	19,6	607	1205	133
Воронежская обл.	23,9	989	967	207
Иркутская обл.	16,8	585	1067	129
Липецкая обл.	17,0	515	1245	123
Тамбовская обл.	10,5	476	880	110
Республика Калмыкия	1,7	117	817	22
Республика Татарстан	67,7	1597	1318	413
Астраханская обл.	10,8	398	1288	103
Волгоградская обл.	30,9	1082	1158	219
Пензенская обл.	11,1	634	915	116
Самарская обл.	72,7	1444	1662	374
Саратовская обл.	28,7	1164	988	220
Ульяновская обл.	16,5	594	1023	119

13. Проведите анализ зависимости валового регионального продукта от ряда факторов по регионам Севера и Центра России, 1999 г. (табл. 10.18).

Таблица 10.18

Регион	ВРП, млрд руб.	Среднего- довая чис- ленность занятых, тыс. чел.	Среднеме- сячная зарплата занятых, руб. / чел.	Среднегодо- вая стои- мость основ- ных фондов, млрд руб.
Республика Карелия	11,3	312	1847	81
Республика Коми	28,4	495	2437	180
Архангельская обл.	22,8	567	1774	170
Вологодская обл.	24,1	572	1709	124
Мурманская обл.	23,8	423	2689	142
Ленинградская обл.	21,2	671	1511	185

Продолжение

Регион	ВРП, млрд руб.	Среднего- довая чис- ленность занятых, тыс. чел.	Среднеме- сячная зарплата занятых, руб. / чел.	Среднегодо- вая стои- мость основ- ных фондов, млрд руб.
Санкт-Петербург	89,8	2330	1878	425
Новгородская обл.	9,4	303	1331	61
Псковская обл.	8,5	302	1028	69
Брянская обл.	11,9	546	850	120
Владимирская обл.	16,0	599	1095	115
Ивановская обл.	9,1	479	846	74
Калужская обл.	10,9	473	1100	95
Костромская обл.	8,9	325	1153	79
Московская обл.	100,6	2332	1622	489
г. Москва	362,5	5051	2875	1223
Орловская обл.	10,2	374	1039	55
Рязанская обл.	14,2	520	1037	107
Смоленская обл.	12,2	447	1148	113
Тверская обл.	17,7	631	1118	163

14. На основе приведенных данных вычислите стандартизированные коэффициенты регрессии, коэффициенты эластичности, множественной корреляции и детерминации, величину системного эффекта факторов.

1. Уравнение регрессии: $\hat{y} = -112 + 2,5x_1 + 0,85x_2 + 12,6x_3 + 0,055x_4$.

2. Средние значения признаков: $\bar{y} = 220$, $\bar{x}_1 = 42$, $\bar{x}_2 = 100$, $\bar{x}_3 = 7,9$, $\bar{x}_4 = 764$.

3. Средние квадратические отклонения: $\sigma_y = 48$, $\sigma_{x_1} = 10$, $\sigma_{x_2} = 19$, $\sigma_{x_3} = 2,2$, $\sigma_{x_4} = 97$.

4. Парные коэффициенты корреляции (табл. 10.19).

Таблица 10.19

	y	x_1	x_2	x_3	x_4
y	1	0,56	0,39	0,71	0,20
x_1		1	0,40	0,45	0,31
x_2			1	0,07	0,24
x_3				1	0,34
x_4					1

15. Проведите анализ зависимости выручки аудиторских фирм, млн руб. (y) от числа сотрудников (x_1) и числа постоянных клиентов (x_2). Данные за 1-е полугодие 1999 г. (табл. 10.20).

Таблица 10.20

Фирма	x_1	x_2	y
Росэкспертиза	287	184	77,1
ОБИ	217	67	73,5
Руфаудит	309	131	35,6
Гориславцев и К ^о	155	213	29,5
МКПЦН	267	500	22,5
МКД	110	35	21,6
Внешаудит	39	44	20,7
Арни	66	156	19,3
Финэсорт	153	86	16,9
Новгородаудит	116	87	14,6
Балт-аудит-эксперт	91	65	14,3
Мариллион	73	110	12,6
ПАКК	77	36	12,2
Эккаудитинг-сервис	68	49	12,0
Марка-аудит	46	52	10,3
Деловые консультации	52	9	9,1
Аудит-консалтинг	22	23	9,0
Порт-акдит	28	41	4,6
Финансы	57	150	4,4
ЦБА	71	86	4,1

16. В результате анализа связи результативного признака с тремя факторами по выборке объемом 20 единиц получены следующие данные (табл. 10.21).

Таблица 10.21

Выборка	$Y_{\text{факт}}$	$\hat{Y}_{\text{расчет}}$
1	47	44
2	49	50
3	54	54
4	54	59
5	57	61
6	67	66
7	61	55
8	70	80
9	75	75

Продолжение

Выборка	$U_{\text{факт}}$	$\hat{U}_{\text{расчет}}$
10	76	73
11	76	71
12	81	88
13	84	81
14	99	93
15	101	107
16	102	97
17	109	108
18	112	118
19	120	124
20	122	113

Вычислите корреляционное отношение, коэффициент детерминации, его скорректированное значение. Проверьте с помощью критерия Фишера надежно ли установлена связь.

17. По данным табл. 10.22 проведите анализ зависимости пассажирооборота (y , млрд пассажирокилометров в год) ряда стран Европы от длины сети железных дорог (x_1 , тыс. км), числа работников на 1 км сети (x_2 , чел./км) и числа пассажиров, проходящих за год на 1 работника железных дорог (x_3 , чел./чел.).

Таблица 10.22

Страна	y	x_1	x_2	x_3
Бельгия	7,0	3,4	11,8	3,6
Дания	5,0	2,2	5,0	13,0
Германия	59,6	38,4	6,2	5,6
Франция	61,6	31,8	5,5	4,3
Греция	1,9	2,5	4,7	1,1
Италия	49,5	16,0	7,6	3,9
Люксембург	0,3	0,3	10,3	3,9
Нидерланды	14,4	2,8	9,2	12,2
Норвегия	2,6	4,0	1,8	6,2
Австрия	8,1	5,7	9,7	3,3
Польша	19,9	22,3	10,2	1,4
Португалия	4,6	2,9	4,4	13,9
Швеция	6,5	10,2	1,2	8,6
Швейцария	12,4	2,9	10,9	1,4
Испания	16,0	12,3	3,0	10,9
Финляндия	3,4	5,9	2,4	3,5

18. Проведите анализ связей выручки аудиторских фирм России за 1-е полугодие 2000 г. (y) с числом филиалов (x_1), выручкой на 1 работника (x_2), числом постоянных клиентов (x_3) (табл. 10.23).

Таблица 10.23

№ п/п	Фирма	y , млн руб.	x_1 , ед.	x_2 , тыс. руб./чел.	x_3 , число клиентов
1	Юнион МС	167	6	292	107
2	ФБК	103	3	440	98
3	Современные бизнес-технологии	48	1	286	26
4	БДО Руфаудит Альянс	42	13	173	198
5	Топ-аудит-Порт-аудит	41	6	201	152
6	Русаудит-Дорнолф, Евсеев и К ^о	37	8	299	127
7	МКПЦН	33	8	136	298
8	Бизнес-аудит	31	8	199	47
9	ПАКК-Универсаудит	25	11	170	42
10	Балт-Аудит-Эксперт	24	2	250	80
11	МИД (PKF) Санкт-Петербург	22	0	189	63
12	Новгород-аудит	22	8	197	145
13	ЭНПИ-Консалт	22	8	230	45
14	ФИНН-Эскорт	20	7	118	42
15	АРНИ	16	5	240	114
16	Аваль	15	2	529	15
17	Бона-фиде финанс	15	5	331	23
18	Ал Руд	13	2	286	12
19	Мариллион	12	1	271	75
20	Аудит-консалтинг. Казань	10	2	112	115

19. По странам мира проведите анализ связи числа медалей, полученных спортсменами стран в Олимпиаде 2000 г. с факторами: x_1 — численность населения, млн. чел.; x_2 — ВВП на душу населения, тыс. долл.; x_3 — средняя продолжительность жизни, лет (табл. 10.24).

Таблица 10.24

№ п/п	Страна	y	x_1	x_2	x_3
1	Австрия	3	8	20,0	77,0
2	Австралия	58	18	18,0	78,2
3	Белоруссия	17	10	4,0	68,0
4	Великобритания	28	58	18,2	77,7
5	Венгрия	17	10	6,4	70,9
6	Германия	57	82	19,3	77,2
7	Дания	6	5	21,1	75,7
8	Италия	34	57	18,7	78,7
9	Казахстан	7	17	2,3	67,6
10	Канада	14	29	10,3	79,0
11	Латвия	3	2	3,3	68,4
12	Нидерланды	25	16	19,0	77,8
13	Россия	88	146	4,0	66,6
14	Румыния	26	23	4,2	69,9
15	США	97	264	25,5	76,7
16	Украина	23	50	2,0	68,8
17	Финляндия	4	5	17,2	76,8
18	Франция	38	58	19,7	78,1
19	Чехия	8	10	9,4	73,9
20	Швейцария	9	7	22,8	78,6

20. Проведите анализ связи рентабельности (убыточности) реализации картофеля по районам Ленинградской обл. в 1998 г. (y , %) с факторами: x_1 – валовой сбор, тыс.; x_2 – урожайность, ц/га; x_3 – средняя цена реализации, руб./ц, 1998 г. (табл. 10.25).

Таблица 10.25

№ п/п	Регион	y	x_1	x_2	x_3
1	Бокситогорский	–80	12	46	143
2	Волосовский	21	328	135	133
3	Волховский	–17	33	81	156
4	Всеволожский	18	62	117	170
5	Выборгский	–4	31	84	168
6	Гатчинский	13	220	152	181
7	Кингисеппский	55	88	147	162
8	Киришский	10	27	105	220
9	Кировский	12	26	87	174

Продолжение

№ п/п	Регион	y	x_1	x_2	x_3
10	Лодейнопольский	-43	7	60	153
11	Ломоносовский	-21	48	97	133
12	Лужский	15	153	103	152
13	Подпорожский	-51	1	59	190
14	Приозерский	2	16	85	176
15	Сланцевский	-18	15	79	140
16	Тихвинский	-27	48	84	146
17	Тосненский	35	121	131	177

21. Проведите анализ связи себестоимости картофеля с размером площади посадки и урожайностью по предприятиям Гатчинского района Ленинградской области, 1996 г. (табл. 10.26).

Т а б л и ц а 10.26

Предприятие	Площадь, га	Урожайность, ц/га	Себестоимость, руб./ц*
Красная Славянка	53	80	194
Гатчинское	257	68	113
Верево	130	112	109
Лесное	40	90	200
Суйда	180	138	75
Дружба	45	90	231
Нива	75	167	94
Орлинское	150	98	111
Красногвардейское	229	88	151
Рейзино	110	64	88
Рождественское	106	163	117
Память Ильича	120	65	76
Белогорка	25	74	159
Кобралово	5	52	478
Пламя	420	69	136
Большевик	70	65	141
* В деноминированных рублях.			

10.3. Вопросы для самоконтроля

1. В чем состоят различия между функциональной и стохастической связью? Приведите примеры связей разных видов.
2. Как понимать утверждение, что корреляционная связь является частным случаем стохастической связи?
3. Каковы предпосылки применения корреляционно-регрессионного анализа?
4. Что такое «ложная корреляция»? Почему и как следует избавляться от ложной корреляции?
5. Если вы хотите построить поле корреляции для изучения связи между успеваемостью в средней школе и в вузе по данным ста студентов, то что вы должны показать по оси абсцисс? по оси ординат?
6. Каковы пределы возможных значений коэффициента корреляции?
7. В каком случае связь теснее: при $r_{yx} = 0,72$ или $r_{yx} = -0,72$?
8. За счет чего возникают различия в значениях коэффициентов парной и частной корреляции?
9. Дайте определение терминов «корреляция», «регрессия».
10. В каком случае свободный член линейного уравнения регрессии содержательно интерпретируется?
11. Какова интерпретация коэффициентов регрессии в уравнении множественной регрессии в натуральном масштабе? в стандартизованном масштабе?
12. Почему коэффициент регрессии в линейном парном уравнении регрессии называют показателем наклона?
13. В чем состоит сходство и различие между коэффициентом парной, частной и множественной корреляции?
14. Как влияет форма уравнения регрессии на измерение тесноты связи между переменными?
15. Поясните применение правила сложения дисперсий в корреляционно-регрессионном анализе.
16. Каковы способы линеаризации уравнения регрессии?
17. В каких случаях для оценки статистической значимости коэффициента корреляции следует использовать z-преобразование Р. Фишера?
18. Каковы условия и возможные направления использования уравнения регрессии в социально-экономическом исследовании?

19. Каким образом можно ввести в уравнение регрессии объясняющие неколичественные переменные?

20. Если неколичественная переменная имеет три категории значений, то сколько параметров при ней будет в уравнении регрессии? Какова интерпретация таких параметров?

21. Если коэффициент корреляции равен 0,9, то чему равен коэффициент детерминации? Каков процент необъясненной (остаточной) вариации?

22. Почему использование метода наименьших квадратов (МНК) обеспечивает нахождение регрессии, наилучшим образом соответствующей исходным данным?

23. Если коэффициент корреляции — это мера тесноты линейной связи, то как измерить тесноту связи в случае нелинейной связи?

24. От чего зависит величина ошибки уравнения регрессии?

25. Как измеряется ошибка аппроксимации?

10.4. Методические рекомендации преподавателям

Практические занятия по данной теме рекомендует проводить в аудитории, оснащенной персональными компьютерами. Все расчеты и оценки статистической значимости полученных результатов позволяет выполнить специальный пакет прикладных программ, например SPSS. На уровне дескриптивной статистики можно работать в Excel.

Целесообразно заранее подготовить базу данных для решения общей задачи в аудитории и подготовить базы данных для индивидуальных или групповых домашних заданий студентам. Базы данных не должны включать большое число единиц (не более 50); число независимых переменных не должно быть более трех-четырёх. Следует подчеркнуть, что предпосылками выполнения корреляционно-регрессионного анализа выступают: измерение переменных на интервальной или порядковой шкале; однородность данных; нормальность распределений переменных.

Особое внимание следует уделить обоснованию выбора формы уравнения регрессии с позиций минимизации остаточной дисперсии результативного признака. Обязательно обращайтесь к графиче-

ческому представлению данных и уравнения регрессии. При решении общей задачи обратите внимание на интерпретацию параметров уравнения и требуйте от студентов пояснения смысла полученных результатов в задачах для самостоятельной работы.

Рекомендуется проведение контрольного тестирования студентов и выполнение одной-двух самостоятельных работ, включающих освещение проблем использования построенной регрессии.

Безусловно, интерес студентов к заданию повышается, если оно основано не на условных, а на реальных данных.

10.5. Методические указания студентам

Данная тема является одной из центральных. Ее усвоение обеспечит вам хорошие шансы успешного освоения в последующем курса «Эконометрика».

Вы должны хорошо понимать специфику задач измерения корреляции и построения уравнения регрессии: уравнение регрессии выражает зависимость y и x в среднем, при прочих равных условиях; измерение корреляции — это измерение степени рассеяния точек (наблюдений) на поле корреляции, их близости (или удаленности) от линии регрессии. Не забывайте, что корреляционные показатели измеряют степень согласованности изменений изучаемых переменных. Отсюда следует опасность измерения ложной корреляции.

Уделите особое внимание формулам, связывающим коэффициент корреляции и коэффициент регрессии.

Выполняя индивидуальное (или групповое) задание, подробно проанализируйте смысл полученных результатов, возможности использования построенной регрессии для решения задачи анализа, т.е. для выявления изменения результата за счет раздельного влияния учтенных факторов, а также для прогнозирования значений результата.

Системы регрессионных уравнений

11.1. Основные положения

Уравнение множественной регрессии описывает связь между зависимой переменной и независимыми переменными, улавливая ее через сопряженность изменения признаков. Сам механизм передачи влияния независимой переменной на зависимую не раскрывается. Построение системы регрессионных уравнений соответствует системе связей между всеми изучаемыми переменными. В рамках конкретной задачи признаки, подлежащие определению, называются *эндогенными* (или *внутренними*), а признаки, которые считаются заданными, — *экзогенными* (или *внешними*).

Если связь между признаками односторонняя, то система уравнений называется *рекуррентной* (или *рекурсивной*). Полезно представить всю систему связей между переменными в виде графа связей.

Граф связей строится как ориентированный граф, т.е. он содержит только упорядоченные пары вершин (дуги). Дуга не должна начинаться и кончаться в одной и той же вершине, т. е. граф не должен содержать *петель* или *контуров*. Контур — конечный путь, у которого начальная вершина совпадает с конечной. Элементарный контур — контур, проходящий через каждую из вершин не более одного раза. Гамильтонов контур — элементарный контур, проходящий через все вершины графа.

Последовательность ребер графа называется маршрутом или путем, соединяющим вершины y и x_1 (или x_1 и x_2 и т.д.). Путь является простым, если все его дуги различны. Между графом связей и системой уравнений должно быть взаимно однозначное соответствие.

Метод наименьших квадратов обеспечивает получение несмещенных оценок параметров, если корреляция между уточненными объясняющими переменными («ошибками») отсутствует.

Система уравнений, соответствующая структуре связей, называется **системой структурных уравнений**.

Уравнение, которое в правой части не содержит эндогенных переменных, называется **приведенным**.

Для однозначного перехода от коэффициентов приведенных уравнений к коэффициентам структурных уравнений требуется выполнение необходимого и достаточного условия точной идентификации*.

Самое простое выражение точной идентификации состоит в том, что в приведенном уравнении должно быть то же число параметров, что и в структурном. Условие идентификации можно сформулировать так: в правой части структурного уравнения должно отсутствовать столько же экзогенных переменных, сколько входит в нее эндогенных переменных.

Если в правую часть структурных уравнений входят все экзогенные переменные, имеющиеся в уравнениях других экзогенных переменных, то система не имеет решения и называется **неидентифицируемой**. Если в каждом из уравнений системы или в одном из них больше экзогенных переменных, чем эндогенных переменных в правой части уравнения, то такая система называется **сверхидентифицируемой**.

Оценка параметров идентифицируемой системы проводится косвенным методом наименьших квадратов (КМНК) или двойным методом наименьших квадратов (ДМНК), а оценка параметров сверхидентифицируемой системы проводится ДМНК.

Пример 11.1. Определите тип системы уравнений, опишите метод решения, решите систему и проверьте надежность решения с помощью статистических критериев. Исходные данные представлены в табл. 11.1. Соответствующий граф связей показан на рис. 11.1.

Таблица 11.1

№ п/п	x_1	x_2	y_1	y_2
1	10	8	20	32
2	15	12	16	36
3	20	6	18	33
4	25	10	24	38
5	30	14	22	50
6	35	15	23	50
7	40	12	24	48

* См.: Практикум по эконометрике / под ред. И.И. Елисеевой. – Изд. 2-е, перераб. и доп. – М.: Финансы и статистика, 2007. – С. 146–161.

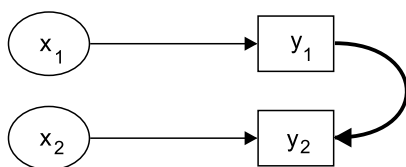


Рис. 11.1. Граф связей

Решение.

1. По виду графа связей определяем, что данная система является рекуррентной, так как вторая эндогенная переменная y_2 зависит от первой эндогенной y_1 , а первая от второй не зависит. Следовательно, вначале необходимо решить уравнение первой

эндогенной: $y_1 = a_1 + b_1 x_1$. Это обычное регрессионное уравнение, решаемое по описанной в теме 10 методике. Решив его,

имеем: $y_1 = 15,64 + 0,2143x_1$, оба члена уравнения надежны на уровне значимости $\alpha = 0,05$.

2. Чтобы решить второе уравнение: $y_2 = a_2 + b_{22}x_2 + c_{21}y_1$, в правую часть следует подставить не фактические значения эндогенной переменной y_1 , а ее выравненные, теоретические значения y_1^* , полученные подстановкой в первое уравнение значе-

ний x_1 для каждой единицы совокупности. Вектор y_1^* включает значения: 17,8; 18,9; 19,9; 21,0; 22,1; 23,1; 24,2. Используя эти значения и величины x_2 , решаем второе, множественное урав-

нение регрессии и получаем его решение: $y_2 = -14,85 + 1,285x_2 + 1,986y_1^*$. Критерий Стьюдента для коэффициента

при x_1 равен 3,66, для коэффициента при y_1^* он равен 4,08. Критерий F для уравнения в целом составил 39,9, вероятность нулевой гипотезы $P_0 = 0,0023$, коэффициент детерминации – 95,2%.

Пример 11.2. Имеются данные по 12 предприятиям (табл. 11.2).

Таблица 11.2

Номер предприятия	Энергово-ооруженность рабочих, кВт/чел., x_1	Тарифный разряд в среднем по предприятию x_2	Производительность труда, тыс. руб./чел. в месяц, y_1	Оплата труда, тыс. руб./чел. в месяц, y_2
1	12	4,7	24	1,8
2	12	4,9	18	2,1
3	15	4,8	26	3,4
4	17	4,6	21	1,6
5	18	3,9	33	1,7
6	21	4,7	31	3,3
7	21	4,7	28	3,8
8	22	4,9	37	3,6
9	22	5,1	34	4,2
10	25	5,0	37	4,0
11	27	4,4	33	2,3
12	28	4,7	38	4,1

Граф связей между переменными представлен на рис. 11.2.

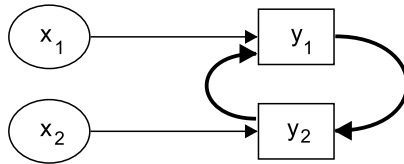


Рис. 11.2. Граф связей

По виду графа связей определяем, что система не является рекуррентной, так как каждая эндогенная переменная зависит от другой эндогенной переменной, структурные уравнения имеют вид:

$$y_1 = a_1 + b_{11}x_1 + c_{12}y_2;$$

$$y_2 = a_2 + b_{22}x_2 + c_{21}y_1.$$

В каждом уравнении в его правой части присутствует одна эндогенная переменная и отсутствует одна экзогенная из имеющихся в системе. Соблюдено условие точной идентификации

системы уравнений, систему следует решать косвенным методом наименьших квадратов, т.е. сначала нужно построить приведенные уравнения и решить их обычным МНК. Приведенные уравнения:

$$\begin{aligned} y_1 &= \alpha_1 + \delta_{11}x_1 + \delta_{12}x_2; \\ y_2 &= \alpha_2 + \delta_{21}x_1 + \delta_{22}x_2, \end{aligned}$$

где

$$\begin{aligned} \delta_{11} &= \frac{b_{11}}{1 - c_{12} \cdot c_{21}}; \quad \delta_{12} = \frac{b_{22} \cdot c_{12}}{1 - c_{12} \cdot c_{21}}; \\ \delta_{21} &= \frac{b_{11} \cdot c_{21}}{1 - c_{12} \cdot c_{21}}; \quad \delta_{22} = \frac{b_{22}}{1 - c_{12} \cdot c_{21}}; \\ \alpha_1 &= \frac{\alpha_1 + c_{12}a_2}{1 - c_{12} \cdot c_{21}}; \quad \alpha_2 = \frac{a_2 + c_{21}a_1}{1 - c_{12} \cdot c_{21}}. \end{aligned}$$

Решение.

1. Решаем приведенные уравнения и получаем

$$y_1 = 11,467 + 1,0388x_1 - 0,4773x_2.$$

Связь надежна, так как F -критерий равен 10,94 и вероятность нулевой гипотезы $P_0 = 0,0039$.

$$y_2 = -8,398 + 0,107x_1 + 1,97x_2.$$

Связь надежна, так как F -критерий равен 11,32 и вероятность нулевой гипотезы $P_0 = 0,0015$.

2. Теперь по коэффициентам приведенных уравнений рассчитываем коэффициенты структурных уравнений:

$$c_{21} = \frac{\delta_{21}}{\delta_{11}} = \frac{0,107}{1,039} = 0,103; \quad c_{12} = \frac{\delta_{12}}{\delta_{22}} = \frac{-0,4773}{1,97} = -0,2433;$$

$1 - c_{12}c_{21} = 1 + 0,02496 = 1,025$ — это знаменатель дробей;

$$b_{11} = 1,039 \cdot 1,025 = 1,065; \quad b_{22} = 1,97 \cdot 1,025 = 2,019.$$

3. Решаем систему двух уравнений относительно свободных членов:

$$a_1 - 0,2423a_2 = 11,467 \cdot 1,025 = 11,754$$

$$0,103a_1 + a_2 = -8,398 \cdot 1,025 = 8,602.$$

Решив эту систему, получаем: $a_1 = 9,44$; $a_2 = -9,58$.

Структурные уравнения принимают вид:

$$y_1 = 9,44 + 1,065x_1 - 0,2423y_2; \quad y_2 = -9,58 + 2,019x_2 + 0,103y_1.$$

Поскольку параметры структурных уравнений получены алгебраическим преобразованием из параметров приведенных уравнений, при точно идентифицированной системе нет необходимости заново проверять надежность этих параметров — надежность установления связи. Критерии для структурных уравнений равны критериям для приведенных уравнений. Решение закончено.

Пример 11.3. Исходные данные по 15 областям Центрального федерального округа представлены в табл. 11.3. Постройте и решите систему уравнений, описывающих валовой региональный продукт и розничный товарооборот за 2001 г.

Таблица 11.3

№ п/п	Область	Основные фонды в экономике, млрд руб., x_1	Продукция промышленности, млрд руб., x_2	Средний душевой доход, руб./чел. в месяц, x_3	ВРП, млрд руб., y_1	Розничный товарооборот, млрд руб., y_2
1	Белгородская	154	52	1819	44	21
2	Брянская	124	19	1482	26	14
3	Владимирская	136	46	1335	35	15
4	Ивановская	85	18	1022	18	10
5	Калужская	110	29	1566	26	12
6	Костромская	91	16	1453	18	8
7	Курская	138	30	1664	32	15
8	Липецкая	139	67	2184	48	19
9	Орловская	71	16	1804	26	12

Продолжение

№ п/п	Область	Основные фонды в эконо- мике, млрд руб., x_1	Продук- ция про- мыш- ленно- сти, млрд руб., x_2	Средний душевой доход, руб./чел. в месяц, x_3	ВРП, млрд руб., y_1	Рознич- ный товаро- оборот, млрд руб., y_2
10	Рязанская	123	29	1701	32	15
11	Смоленская	138	32	2124	30	19
12	Тамбовская	127	15	1950	26	17
13	Тверская	189	36	1583	39	18
14	Тульская	168	58	1824	44	19
15	Ярославская	200	59	2179	47	18

Исходя из экономического анализа построим граф связей (рис. 11.3).

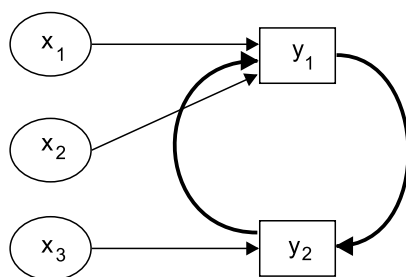


Рис. 11.3. Граф связей

Структурные уравнения, соответствующие графу связей:

$$\begin{aligned} y_1 &= a_1 + b_{11}x_1 + b_{12}x_2 + c_{12}y_2 \\ y_2 &= a_2 + b_{23}x_3 + c_{21}y_1 \end{aligned}$$

В первом уравнении в правой части присутствует одна эндогенная и отсутствует одна экзогенная переменная — это уравнение точно идентифицируемое. Но во втором уравнении в правой части присутствует одна эндогенная переменная, а отсутствуют две экзогенных! Условие точной идентификации нарушено. Приведенное уравнение будет содержать четыре параметра,

а структурное содержит три. Второе уравнение, а следовательно, и вся система в целом являются сверхидентифицируемыми и должны решаться с помощью двойного метода наименьших квадратов (ДМНК).

1. Вначале построим и решим приведенные уравнения:

$$\hat{y}_1 = a_1 + \delta_{11}x_1 + \delta_{12}x_2 + \delta_{13}x_3$$

$$\hat{y}_2 = a_2 + \delta_{21}x_1 + \delta_{22}x_2 + \delta_{23}x_3$$

В результате получим:

$$\hat{y}_1 = 1,671 + 0,0716x_1 + 0,376x_2 + 0,00494x_3;$$

$$\hat{y}_2 = 0,133 + 0,0488x_1 + 0,0406x_2 + 0,00434x_3.$$

Для первого уравнения F -критерий равен 59,7, для второго уравнения – $F = 11,8$. Оба уравнения надежны.

2. Вычислим по приведенным уравнениям расчетные значения эндогенных переменных. Используем два вектора величин \hat{y}_1 и \hat{y}_2 (табл. 11.4).

Таблица 11.4

	1	2	3	4	5	6	7	8
\hat{y}_1	41,2	25,0	35,3	19,6	28,2	21,6	31,0	47,1
\hat{y}_2	17,7	13,4	14,4	9,4	13,5	11,5	15,3	19,1

Продолжение

	9	10	11	12	13	14	15
\hat{y}_1	21,7	29,7	34,1	26,0	36,6	44,5	49,0
\hat{y}_2	12,1	14,7	17,4	15,4	17,7	18,6	21,8

3. Эти расчетные значения эндогенных переменных и экзогенные переменные, имеющиеся в структурных уравнениях, образуют исходные данные для расчета структурных уравнений, т.е. второго применения МНК. Получим решения структурных уравнений:

$$y_1 = 1,52 + 0,0160x_1 + 0,330x_2 + 1,137y_2;$$

$$y_2 = 1,728 + 0,00326x_3 + 0,2489y_1.$$

Первое уравнение точно идентифицировано, поэтому для него F -критерий остается прежним: $F_1 = 59,7$.

Второе уравнение сверхидентифицировано, для него следует заново определить критерий надежности: $F_2 = 15,2$, т.е. уравнение также надежно.

Прогнозировать ВРП можно и по приведенному уравнению, и по структурному; прогнозировать значения розничного товарооборота следует только по структурному уравнению. Решение закончено.

11.2. Задания по теме

1. Произведите идентификацию системы уравнений, опишите методику их решения и вычислите параметры уравнений. Исходные данные представлены в табл. 11.5. Соответствующий граф связей дан на рис. 11.5.

Таблица 11.5

	1	2	3	4	5	6	7	8	9
x_1	10	15	12	16	20	15	28	36	30
x_2	16	17	18	20	24	26	30	33	36
y_1	40	50	45	55	65	50	70	68	60
y_2	100	110	96	118	129	107	136	131	135

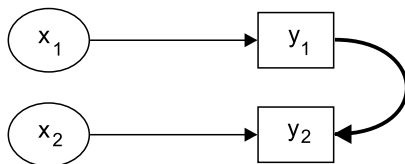


Рис. 11.4. Граф связей

2. Задание аналогично предыдущему. Исходные данные представлены в табл. 11.6.

Таблица 11.6

	1	2	3	4	5	6	7	8	9	10
x_1	8	16	24	32	36	44	52	60	68	76
x_2	17	13	23	19	35	27	39	23	35	49
y_1	100	110	95	124	138	124	141	149	156	148
y_2	80	82	100	110	145	129	143	126	155	156

Структурные уравнения: $y_1 = a_1 + b_{11}x_1 + c_{12}y_2$;

$$y_2 = a_2 + b_{23}x_2 + c_{21}y_1.$$

3. Структурные уравнения имеют вид:

$$y_1 = a_1 + b_{11}x_1 + b_{12}x_2 + c_{12}y_2;$$

$$y_2 = a_2 + b_{22}x_2 + b_{23}x_3 + c_{21}y_1.$$

При решении приведенных уравнений получено:

$$y_1 = 10 + 2x_1 + 6x_2 + 10x_2;$$

$$y_2 = -3 + 4x_1 + 2x_2 + 5x_3.$$

Определите параметры структурных уравнений.

4. Запишите структурные уравнения, проведите идентификацию системы, составьте приведенные уравнения и опишите методику решения системы. Соответствующий граф связей дан на рис. 11.5.

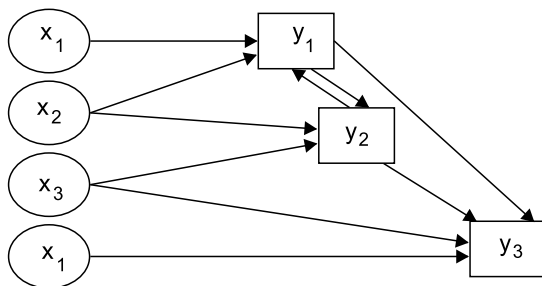


Рис. 11.5. Граф связей

5. Исходные данные 15 инвестиционных компаний России за 1-е полугодие 2002 г. представлены в табл. 11.7. Проведите идентификацию системы уравнений, опишите методику решения и решите эти уравнения:

$$y_1 = a_1 + b_{12}x_2 + b_{13}x_3 + c_{12}y_2;$$

$$y_2 = a_2 + b_{21}x_1 + b_{22}x_3 + c_{21}y_1.$$

Таблица 11.7

№ п/п	Компания	Число сотрудников, чел., x_1	Размер активов, млн руб., x_2	Доля собственного капитала, %, x_3	Прибыль, млн руб., y_1	Оборот на 1 сотрудника, млн руб., y_2
1	ООО АТОН	64	825	5,3	0,2	639
2	Ренессанс-капитал	32	1085	4,8	6,6	3793
3	ООО Компания БКС	94	340	15,9	3,2	1541
4	ЗАО Алор – Инвест	43	230	14,3	0,9	1327
5	Инвестброкцентр	46	3998	1,0	-0,1	7100
6	ВЭО открытие	62	78	19,1	2,2	1307
7	ЗАО Финанс-аналитик	120	337	8,6	2,0	429
8	ООО УК Резерв-Инвест	102	561	7,2	0,2	3466
9	ЗАО ИК Регион	29	1352	14,8	7,4	501
10	ООО ИК Велес-Капитал	48	329	9,7	3,0	865
11	ЗАО ИК РФИ	42	1403	6,5	7,3	584
12	ООО ИФК Метрополь	98	2413	21,4	1,2	364
13	ЗАО Вика	59	191	18,8	0,9	512
14	ОЛМА	109	225	15,1	0,4	151
15	ЗАО ИК Газинвест	22	53	50,0	-0,1	1087

6. Вычислите параметры структурных уравнений, объясните методы решения системы (рис. 11.6).

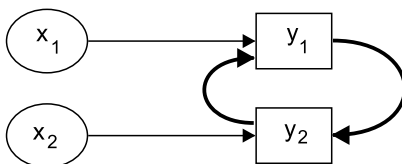


Рис. 11.6. Граф связей

В результате решения приведенных уравнений было получено:

$$y_1 = 30 + 4x_1 + 0,9x_2;$$

$$y_2 = 70 + 2x_1 + 3x_2.$$

7. По данным рис. 11.7 и табл. 11.8 идентифицируйте и решите систему уравнений. Опишите методику решения.

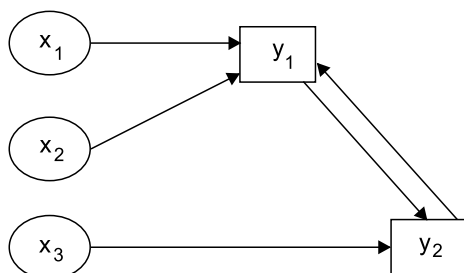


Рис. 11.7. Граф связей

Таблица 11.8

x_1	9	11	13	15	17	19	21	23	25	27
x_2	5	8	6	9	7	10	8	11	9	12
x_3	30	27	24	20	20	18	16	14	14	10
y_1	40	48	46	59	55	70	72	80	79	94
y_2	100	120	110	140	130	150	155	160	153	175

11.3. Вопросы для самоконтроля

1. Что представляет собой система взаимосвязанных регрессионных уравнений?

2. Какие признаки в системе уравнений играют роль эндогенных, а какие — роль экзогенных?

3. По какой причине решение системы взаимосвязанных уравнений обычным МНК может дать смещенные и несостоятельные оценки параметров связи?

4. Какая система уравнений является рекуррентной и каков метод ее решения?

5. Что такое «граф связей» признаков в системе и как его применить для составления структурных уравнений?
6. В чем состоит проблема идентификации системы уравнений?
7. По каким критериям можно произвести идентификацию системы уравнений?
8. Как преобразовать структурные уравнения в приведенные?
9. Из каких этапов (шагов) состоит косвенный метод наименьших квадратов?
10. Из каких этапов (шагов) состоит двойной метод наименьших квадратов?
11. Как следует поступить, если система уравнений оказалась неидентифицируемой?

11.4. Методические рекомендации преподавателям

Прежде всего отметим, что данная тема выходит за рамки курса статистики и принадлежит эконометрике. В рамках последней этот раздел излагается обычно под названием «Системы одновременных уравнений». Однако логика рассмотрения систем регрессионных уравнений свидетельствует о возможной подаче этого материала в контексте развития многофакторного регрессионного анализа.

Рекомендуем на практических занятиях разобрать общую задачу и найти численные значения оценок параметров, используя ППП «Econometric Views» (EViews). Обратите внимание студентов на то, что успешность решения системы регрессионных уравнений зависит от правильности представления связей между переменными. Поэтому преподавателю нужно выбрать пример, содержание которого доступно студенту.

Следует акцентировать внимание на верифицирующей роли графа связей: предостеречь студентов от появления *контуров* и т.д.

При построении системы структурных уравнений следует подчеркнуть ее изоморфизм графу связей. Рекомендуется ограничиться рекуррентной (рекурсивной) системой как в решении общей задачи, так и в индивидуальных (или групповых, на 4 — 5 человек) заданиях, выдаваемых для самостоятельной работы студентам.

Подчеркните деление переменных на эндогенные и экзогенные в рамках конкретной задачи, после чего постройте приведенные уравнения и обсудите выполнение условия точной идентификации уравнений.

В этой теме студенты впервые знакомятся с косвенным методом наименьших квадратов (КМНК). Пакет прикладных программ EViews позволяет найти оценки параметров этим методом. Можно ограничиться применением КМНК и не уделять внимания использованию двойного метода наименьших квадратов (ДМНК), который будет рассмотрен в курсе «Эконометрика». В любом случае старайтесь уделить как можно больше внимания интерпретации оценок параметров, добивайтесь от студентов понимания сущности результатов так, чтобы техническая сторона не уводила в сторону от экономической или социальной проблематики.

11.5. Методические указания студентам

Данная тема представляет собой раздел эконометрики, а это значит, что при ее изучении вы должны обладать междисциплинарными знаниями: иметь подготовку по микро- и макроэкономике, математике и статистике. Получив от преподавателя задание, постарайтесь тщательно разобраться в системе связей между изучаемыми переменными. С этой целью представьте все постулируемые связи графически, в виде ориентированного графа связей, в котором все дуги имеют направленность. Проверьте граф связей на наличие контуров, т. е. путей, которые начинаются и заканчиваются в одной и той же вершине (вершина графа соответствует переменной). Если выявится наличие контура, то следует внести коррективы в предположения о связях между переменными, их воздействие друг на друга. Граф связей помогает выделить эндогенные (внутренние для данной системы) переменные и экзогенные (внешние) переменные. На основе системы связей вы должны построить систему структурных уравнений.

Если связь между переменными односторонняя, то вы получите рекуррентную (рекурсивную) систему. Структурные уравнения следует преобразовать в приведенные, чтобы обеспечить несмещенность и состоятельность оценок параметров. В правой части приведенных уравнений должны быть только эк-

зогенные переменные, а в левой — по одной эндогенной переменной. Чтобы можно было перейти от параметров приведенного уравнения к параметрам структурного уравнения, необходимо выполнение условия точной идентификации уравнений. По этому вопросу полезно почитать учебник «Эконометрика» под ред. И.И. Елисеевой*. Решение задачи рекомендуем выполнить с использованием ППП EViews, который позволяет получить оценку параметров идентифицируемой системы уравнений косвенным методом наименьших квадратов (КМНК), а в случае сверхидентифицируемости применить двойной метод наименьших квадратов (ДМНК).

* Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. — 2-е изд., перераб. и доп. — М.: Финансы и статистика, 2005.

Статистический анализ неколичественных переменных

12.1. Основные положения

Неколичественные переменные — это переменные, измеренные на номинальной и порядковой шкалах. Измерение на номинальной шкале представляет указание градации признака x для i -го объекта. Измерение на порядковой шкале позволяет определить, в каком отношении по признаку x находится i -й объект с другими объектами — опережает, отстает и т.д. Нередко порядковые переменные задаются в виде шкалы Ликерта, включающей крайние позиции, промежуточную и нулевую (см. Учебник, с. 414).

Значимость неколичественной информации возрастает в связи:

- с необходимостью перевода числовых данных в нечисловые для сохранения коммерческой тайны;

- с построением рейтингов;
- с расширением социальных исследований (связанных с выявлением мотиваций, жизненных планов людей, потребительских предпочтений, оценкой качества товаров и услуг), содержащих вербальную информацию.

При изучении связи между двумя переменными выделяется случай анализа связи между бинарными (или дихотомическими) переменными. При этом таблица сопряженности имеет четыре клетки и называется четырехклеточной. Связь выявляется и измеряется на основе клеточных частот, n_{ij} — число единиц в клетке таблицы сопряженности с i -м значением переменной x и j -м значением переменной y .

Анализ *одномерного распределения* по неколичественной переменной можно провести с помощью показателя энтропии (неопределенности) распределения:

$$H(x) = -\sum_{(i)} P(x_i) \log_2 P(x_i), \quad (12.1)$$

где $P(x_i)$ – вероятность (частота) появления i -го значения переменной x .

Поскольку все $P(x_i) < 1$, то $\log_2 P(x_i) < 0$. Этим объясняется появление знака минус в формуле энтропии распределения. Значения $P(x_i) \log_2 P(x_i)$ табулированы (см. приложение 2).

Величина энтропии равна нулю, если все единицы принадлежат одной категории переменной x . Величина энтропии максимальна, если распределение единиц равномерно, т.е. шансы каждой категории x равновероятны. Таким образом,

$$0 \leq H(x) \leq H_{\max}. \quad (12.2)$$

Если $P(x_i) = \frac{1}{k}$ (где k – число категорий переменной x), то

$$H(x) = H_{\max}. \quad (12.3)$$

Пример 12.1. Рассмотрим данные о структуре экономики США (табл. 12.1).

Таблица 12.1

Структура экономики США (%)

Отрасль	1820 г. w_0	1992 г. w_1
Сельское хозяйство, добывающая промышленность, лесная, рыбная ловля	70	2,8
Обрабатывающая промышленность	15	23,2
Сфера услуг	15	74,0
Итого	100	100

Измерим энтропию распределения в 1820 и в 1992 гг.:

$$\begin{aligned} H(x) &= -(0,7) \log_2(0,7) + [-(0,15) \log_2(0,15) \cdot 2] = \\ &= 0,3602 + 0,4105 \cdot 2 = 1,1812; \end{aligned}$$

$$\begin{aligned} H(x) &= -(0,028) \log_2(0,028) + [-0,232 \log_2 0,232] + [-0,74 \log_2 0,74] = \\ &= 0,1444 + 0,4890 + 0,3215 = 0,9549, \end{aligned}$$

т.е. энтропия снизилась в 1992 г. по сравнению с 1820 г. на 19,2%.

Но этот результат мало о чем говорит, тогда как данные свидетельствуют о коренных изменениях: от доминирования добычи и сельского хозяйства к доминированию сферы услуг. При таких изменениях целесообразно воспользоваться другими показателями, а именно показателями структурных сдвигов. Рассчитаем коэффициент ранговой корреляции для измерения корреляции между рангами отраслей экономики в одном и другом периодах и другие показатели, характеризующие изменение структуры (табл. 12.2).

$$\rho = \frac{1 - 6\sum d_i^2}{k(k^2 - 1)},$$

где k — число категорий.

Получим

$$\rho = 1 - \frac{6 \cdot 6,5}{k \cdot (k^2 - 1)} = 1 - \frac{39}{3 \cdot 8} = -0,625;$$

$$\bar{d} = \frac{\sum |w_1 - w_o|}{k} = \frac{134,4}{3} = 44,8 \text{ nn};$$

$$\sigma = \sqrt{\frac{\sum (w_1 - w_o)^2}{k}} = \sqrt{\frac{8064,08}{3}} = 51,85 \text{ nn};$$

$$\sigma \frac{w_1}{w_o} = \sqrt{\frac{\sum \frac{(w_1 - w_o)^2}{w_o}}{w_o}} = \sqrt{301,06} = 17,65\%.$$

Разные показатели по-разному оценивают структурные изменения и в абсолютном выражении, и в относительном. В среднем удельный вес категории изменился на 17,65%.

Индекс структурных различий также свидетельствует о достаточно сильных изменениях ($0 \leq I_{\text{структ. разл.}} \leq 1$). Выражая этот индекс в процентах, получим:

$$I_{\text{структ. разл.}} = \frac{1}{2} \sum |w_1 - w_o| = \frac{1}{2} \cdot 134,4 = 67,2.$$

Таблица 12.2

Расчетная таблица

Отрасль	Ранги		$d_i^2 = (R_1 - R_0)$	$ w_1 - w_0 $	$(w_1 - w_0)^2$	$\frac{(w_1 - w_0)^2}{w_0}$
	R_0	R_1				
Сельское хозяйство, добывающая промышленность, лесная, рыбная ловля	1	3	4,0	67,2	4515,84	64,51
Обрабатывающая промышленность	2,5	2	0,25	8,2	67,24	4,48
Сфера услуг	2,5	1	2,25	59,0	3481,0	232,07
Итого:			6,50	134,4	8064,08	301,06

Для измерения связи между неколичественными переменными собранные данные представляются в виде таблицы сопряженности. При изучении парной связи таблица сопряженности — это двумерное распределение единиц совокупности по значениям признаков x и y . При изучении многофакторной связи таблица сопряженности — многомерное распределение.

Измерение связи между двумя дихотомическими переменными производится с помощью:

- коэффициента ассоциации

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}; \quad (12.4)$$

- коэффициента контингенции

$$\Phi = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{(n_{11} + n_{12})(n_{12} + n_{22})(n_{22} + n_{21})(n_{21} + n_{11})}}. \quad (12.5)$$

Оба показателя принимают значения в интервале $[0,1]$ по абсолютной величине от -1 до $+1$.

Пример 12.2. В результате опроса 100 посетителей кафе Макдональдс получены следующие данные (табл. 12.3).

Таблица 12.3

Зависимость частоты посещений от вкуса посетителей

Регулярность посещений	Нравится чизбургер	Не нравится чизбургер	Итого
Регулярно	38	15	53
Нерегулярно	7	40	47
Итого	45	55	100

$$Q = \frac{38 \cdot 40 - 15 \cdot 7}{38 \cdot 40 + 15 \cdot 7} = \frac{1415}{1625} = 0,871;$$

$$\Phi = \frac{38 \cdot 40 - 15 \cdot 7}{\sqrt{(38 + 15)(15 + 40)(7 + 40)(38 + 7)}} = \frac{1415}{2707} = 0,522.$$

Соотношение между этими показателями таково: $Q > \Phi$.

Для доказательства наличия статистически значимой связи по таблице сопряженности 2×2 используется тест χ^2 -квадрат с поправкой Ф. Йейтса:

$$\chi^2 = \sum_{(i)} \sum_{(j)} \frac{(|n_{ij} - \hat{n}_{ij}| - 0,5)^2}{\hat{n}_{ij}}, \quad (12.6)$$

где \hat{n}_{ij} — теоретическая клеточная частота в ячейке ij .

Пример 12.3. Нужно определить на 5%-ном уровне значимости, существенна ли связь между полом опрошиваемого и посещением кафе Макдональдс (табл. 12.4).

Таблица 12.4

Зависимость частоты посещений от пола посетителей

Регулярность посещений	Мужчины		Женщины		Итого
Регулярно	20	24	30	26	50
Нерегулярно	20	16	30	34	50
Итого	40		60		100

В таблице 2×2 достаточно вычислить теоретическую клеточную частоту лишь для одной ячейки, так как маргинальные частоты остаются теми же.

Найдем $n_{11} = \frac{40 \cdot 50}{100} = 20$, тогда $n_{12} = 30$, поскольку $(n_{11} + n_{12}) = n_1 = 50$. Так как $n_1 = 40$, то $n_{12} = 40$, а $n_{22} = 20$. Вычислим значение χ^2 -квадрат с поправкой Ф. Йейтса:

$$\chi^2 = \frac{(|24 - 20| - 0,5)^2}{20} + \frac{(|26 - 30| - 0,5)^2}{30} + \frac{(|16 - 20| - 0,5)^2}{20} + \frac{(|34 - 30| - 0,5)^2}{30} = 2,0316.$$

Полученное значение меньше критического (табличного):

$$\chi^2_{\alpha} = 3,841, \quad (\alpha = 0,05, \text{ df} = 1).$$

Следовательно, связь между регулярностью посещений кафе Макдональдс и полом не может считаться доказанной: $\chi^2_{\text{факт}} < \chi^2_{\text{табл}}$. В этом случае вычисление показателей тесноты связи, Q и Φ , может не производиться. Если не вычислять критерий хи-квадрат, а начать с расчета мер связи Q и Φ , то можно сделать неправильные выводы, так как $Q = 0,32$, а $\Phi = 0,16$. Эти значения отличны от нуля и возникает соблазн утверждать наличие умеренно тесной или слабой связи. Критерий хи-квадрат предостерегает от подобных выводов.

Если число категорий для переменных x и y больше двух, то строится таблица сопряженности произвольной размерности — с m строками и p столбцами. В клетках такой таблицы располагаются частоты совместного появления соответствующих категорий переменных x_i и y_j (n_{ij}). Анализ начинается с проверки гипотезы о независимости переменных x и y . Эта гипотеза проверяется с помощью критерия хи-квадрат:

$$\chi^2 = \sum_{(i)} \sum_{(j)} \frac{(n_{ij} - \mathring{n}_{ij})^2}{\mathring{n}_{ij}}. \quad (12.7)$$

Испытуемая гипотеза, $H_0 : n_{ij} = \mathring{n}_{ij}$, т. е. предполагается, что фактические клеточные частоты равны теоретическим, \mathring{n}_{ij} , которые вычисляются в предположении о независимости x и y :

$$\mathring{n}_{ij} = n \cdot \frac{n_i}{n} \cdot \frac{n_j}{n} = \frac{n_i \cdot n_j}{n}. \quad (12.8)$$

Для лучшего запоминания этой формулы укажем, что в числителе — произведение итогов по строке и столбцу, в знаменателе — объем выборки.

Обращаем внимание: теоретическая клеточная частота должна быть не менее 5 единиц, $\mathring{n}_{ij} \geq 5$. В этом состоит одно из условий применения критерия хи-квадрат. Распределение хи-квадрат зависит от числа степеней свободы df и принятого уровня значимости. Так как маргинальные частоты и объем выборки не изменяются, то $df = m \cdot p - m - p + 1 = (m - 1)(p - 1)$.

Если факт наличия связи установлен (с определенной вероятностью), то измеряется теснота связи с помощью коэффициентов взаимной сопряженности и других мер. Коэффициенты взаимной сопряженности включают статистику хи-квадрат:

- коэффициент взаимной сопряженности К. Пирсона

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{\varphi^2}{1 + \varphi^2}}; \quad (12.9)$$

- коэффициент взаимной сопряженности А. Чупрова

$$T = \sqrt{\frac{\chi^2 : n}{\sqrt{(m-1)(p-1)}}} = \sqrt{\frac{\varphi^2}{df^{1/2}}}; \quad (12.10)$$

- коэффициент взаимной сопряженности Г. Крамера

$$V = \sqrt{\frac{\chi^2 : n}{\min\{m-1, p-1\}}} = \sqrt{\frac{\varphi^2}{\min\{m-1, p-1\}}}, \quad (12.11)$$

где φ^2 — средняя квадратическая сопряженность;

m — число строк;

p — число столбцов таблицы.

Пример 12.4. По данным опроса 1248 российских мужчин, переживших хотя бы один развод, изучается связь между проживанием с родителями в период брака и возрастом респондента (табл. 12.5).

Таблица 12.5

Проживание с родителями – возраст мужчин

Проживание	Возраст, лет				Итого
	до 30	30–39	40–49	50 и более	
С родителями мужа	146 ¹²⁶	131 ¹³⁵	51 ⁵⁶	6 ¹⁷	334
С родителями жены	110 ¹⁰⁵	114 ¹¹³	40 ⁴⁷	15 ¹⁴	279
Попеременно	71 ⁵⁰	47 ⁵⁴	14 ²²	2 ⁷	134
Всегда жили одни	143 ¹⁸⁹	214 ²⁰³	103 ⁸⁴	41 ²⁶	501
Итого	470	506	208	64	1248

$$\begin{aligned}\chi^2 = & \frac{(146-126)^2}{126} + \frac{(131-135)^2}{135} + \frac{(51-56)^2}{156} + \frac{(6-17)^2}{17} + \frac{(110-105)^2}{105} + \\ & + \frac{(114-113)^2}{113} + \frac{(40-47)^2}{47} + \frac{(15-14)^2}{14} + \frac{(71-50)^2}{50} + \frac{(47-54)^2}{54} + \\ & + \frac{(14-22)^2}{22} + \frac{(2-7)^2}{7} + \frac{(143-189)^2}{189} + \frac{(214-203)^2}{203} + \frac{(103-84)^2}{84} + \\ & + \frac{(41-26)^2}{26} = 53,43.\end{aligned}$$

Число степеней свободы $df = (4 - 1)(4 - 1) = 9$; $\alpha = 0,05$.

Отсюда $\chi_{\text{табл}} = 16,92$. $\chi_{\text{факт}}^2 > \chi_{\text{табл}}^2$, следовательно, H_0 отклоняется: проживание с родителями и возраст респондента связаны между собой.

Можно оценить вклад отдельных категорий в общую величину χ^2 . Так, например, особенно большие различия между фактическими и теоретическими клеточными частотами для молодых (до 30 лет), кто всегда в браке жили без родителей: 189 против 143 и вклад этой категории в общую величину χ^2 -квadrat составляет

$\frac{(143-189)^2}{189} = 11,2$, т.е. почти 21% общего полученного

значения статистики χ^2 -квadrat. По данным табл. 12.5, $C = 0,203$, $T = 0,119$, $V = 0,119$. Так как таблица квадратная, то $T = V$. Факт наличия связи доказан, но связь слабая.

Меры связей λ -Гутмана.

Это семейство мер рассмотрено в Учебнике (с. 434–435).

Обратимся к примеру (см. табл. 12.5).

$$\lambda_b = \frac{(146 + 114 + 71 + 214) - 506}{1248 - 506} = \frac{548 - 506}{742} = 0,057;$$

$$\lambda_a = \frac{(146 + 214 + 103 + 41) - 501}{1248 - 501} = \frac{3}{747} = 0,004;$$

$$\lambda = \frac{42 + 3}{742 + 747} = 0,03.$$

Хотя связь между возрастом и проживанием с родителями доказана, но она весьма слабая. Меры λ подтверждают это: λ_b показывает, что вероятность правильно указать характер проживания с родителями повышается в зависимости от знания возраста мужчины на 5,7%; а λ_a показывает, что вероятность правильно указать возраст разведенного мужчины повышается на 0,4%, если известно то, как он проживал с родителями. Взаимосвязанный учет значений того и другого признака повышает правильность предсказания результатов только на 3%.

Поскольку меры λ_b и λ_a имеют тот недостаток, что они могут обратиться в нуль, если $\sum n_{ij(\max)} = n_{j(\max)}$ (по строке) или $\sum n_{ij(\max)} = n_{i(\max)}$ (по столбцу), то в этом случае можно использовать меры τ Гудмена – Краскала. Покажем расчет λ_b по данным табл. 12.5:

$$\tau_b = \frac{1}{1248(1248^2 - 470 \cdot 506^2 - 208^2 - 64^2)} \times$$

$$\times \left[\frac{(1248 \cdot 146 - 470 \cdot 334)^2}{334} + \frac{(1248 \cdot 131 - 506 \cdot 334)^2}{334} + \right.$$

$$+ \frac{(1248 \cdot 51 - 208 \cdot 334)^2}{334} + \frac{(1248 \cdot 6 - 64 \cdot 334)^2}{334} +$$

$$+ \frac{(1248 \cdot 110 - 470 \cdot 279)^2}{279} + \dots +$$

$$\left. + \frac{(1248 \cdot 103 - 208 \cdot 501)^2}{501} + \frac{(1248 \cdot 41 - 61 \cdot 501)^2}{501} \right] = 0,014.$$

Таким образом, правильность предсказания проживания с родителями при знании возраста мужчины повышается на 1,4%.

Измерение связи между порядковыми переменными. При измерении переменных на порядковой шкале для оценки тесноты связи между ними используются коэффициенты ранговой корреляции Ч. Спирмена и М. Кендалла. Коэффициент Спирмена:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)};$$

$$-1 \leq \rho \leq 1, \quad 0 \nless \rho \nless 1.$$

Пример 12.5. Пять фирм, производящих молочные продукты, проранжированы покупателями и продавцами (табл. 12.6).

Таблица 12.6

Фирма	Ранг покупателей R_x	Ранг продавцов R_y	d_i^2
А	1	2	1
В	2	1	1
С	3	5	4
Д	4	4	0
Е	5	3	4

$$\rho = 1 - \frac{6 \cdot 10}{5(25 - 1)} = 0,5.$$

Коэффициент ранговой корреляции М. Кендалла:

$$\tau = \frac{S}{\frac{1}{2}n(n-1)},$$

где $S = P - Q$ — фактическая сумма рангов;

$\frac{1}{2}n(n-1)$ — максимальная сумма рангов;

P — прямой порядок рангов (сумма рангов выше данного — по переменной y);

Q — обратный порядок рангов (сумма рангов ниже данного — по переменной y).

Произведем расчеты по данным примера 5.

$$S = (3 - 1) + (3 - 0) + (0 - 2) + (0 - 1) = 2.$$

Тогда

$$\tau = \frac{2}{\frac{1}{2} \cdot 5 \cdot 4} = 0,2.$$

Соотношение между коэффициентами ранговой корреляции таково: $\tau < \rho$. Для больших выборок выполняется следующее приближенное равенство:

$$\rho \approx \frac{3}{2} \tau.$$

Коэффициент конкордации. Для определения тесноты связи между несколькими порядковыми переменными применяется коэффициент конкордации W :

$$W = \frac{12S}{m^2(n^3 - n)},$$

где m — число переменных;

n — число наблюдений;

S — отклонение суммы квадратов рангов от среднего значения квадратов рангов.

Коэффициент конкордации принимает значение в интервале от -1 до $+1$.

Пример 12.6. По данным табл. 12.7 нужно определить степень согласованности мнений экспертов относительно финансовой устойчивости семи коммерческих банков.

Решение:

$$S = 1242,5 - \frac{84^2}{7} = 234,5; \quad W = \frac{12 \cdot 234,5}{3^2(7^3 - 7)} = \frac{2814}{3024} = 0,93.$$

Таким образом, степень согласованности оценок экспертов весьма высокая.

Значимость коэффициента конкордации проверяется на основе критерия χ^2 -квадрат:

$$\chi^2 = \frac{12S}{m \cdot n(n-1)}.$$

По данным примера $\chi^2_{\text{факт}} = \frac{12 \cdot 234,5}{3 \cdot 7(7-1)} = 22,33$. Табличное значение χ^2 (при $\alpha = 0,05$ и числе степеней свободы $df = n - 1 = 7 - 1 = 6$) составляет $\chi^2_{\text{табл}} = 12,59$. Следовательно, $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$,

Таблица 12.7

Исходные данные и расчет коэффициента конкордации

Номер банка	Экспертные оценки, баллов			Сумма рангов	Квадрат суммы рангов
	1-й эксперт R_1	2-й эксперт R_2	3-й эксперт R_3		
1	5	4	6	15	225
2	7	7	7	21	441
3	3	3	2	8	64
4	2	1,5	3	6,5	42,25
5	1	1,5	1	3,5	12,25
6	4	5	4	13	169
7	6	6	5	17	289
Итого	—	—	—	84	1242,5

т.е. вычисленное значение коэффициента конкордации статистически значимо и свидетельствует о тесной связи между оценками экспертов. Этот вывод не изменится и в том случае, если мы учтем наличие связанных рангов:

$$W = \frac{12S}{m^2(n^3 - n) - m \sum_{j=1}^m T_j},$$

где $T_j = \frac{1}{12} \sum_{j=1}^m (t_j^3 - t_j),$

t — количество связанных рангов по j -му столбцу.

Тогда проверка значимости осуществляется по формуле

$$\chi^2 = \frac{12S}{m \cdot n(n+1) - \frac{1}{n-1} \sum_{j=1}^m T_j}.$$

Критическое значение критерия определяется так же, как и в предыдущем случае.

12.2. Задания по теме

1. Распределение ответов менеджеров рекламных компаний на вопрос анкеты «С какими проблемами сталкивается Ваша организация при осуществлении рекламной деятельности?» представлено в табл. 12.8.

Таблица 12.8

Категория компании	Всего ответов	Ответы							
		1	2	3	4	5	6	7	8
Прибыльная	56	12	3	4	12	11	4	7	1
Убыточная	24	4	4	2	6	4	—	4	—
Частная	61	14	7	5	13	13	1	7	1
Смешанная	19	3	—	—	7	3	3	3	—

Расшифровка ответов:

1 — высокие налоги;

2 — высокая арендная плата;

- 3 – недостаток правовых и нормативных актов;
- 4 – высокая конкуренция на рынке рекламных услуг;
- 5 – низкий спрос на рекламные услуги у отечественных товаропроизводителей;
- 6 – захват рекламного рынка зарубежными рекламопроизводителями;
- 7 – недостаток квалифицированных кадров;
- 8 – другое.

Сравните энтропии распределений ответов рекламных предприятий разных категорий: а) прибыльных и убыточных; б) частных и смешанных. Влияет ли категория предприятия на оценку проблем в рекламном деле?

2. Получено распределение ответов на блок вопросов анкеты «Считаете ли Вы существенными по вашему месту жительства следующие экологические проблемы: качество питьевой воды; нехватка мест отдыха, парков, стадионов...» (табл. 12.9).

Таблица 12.9

Удовлетворяет ли качество питьевой воды	Нехватка мест отдыха		
	Да	Нет	Затруднились ответить
Да	150	296	48
Нет	10	53	11
Затруднились ответить	16	35	17

Проверьте гипотезу о связи. Вычислите асимметричные и симметричные меры связи Гутмана. Поясните, имеет ли смысл применение асимметричных мер в данном случае.

3. Имеется распределение ответов опрошенных клиентов об условиях оплаты и сроках выполнения заказов (табл. 12.10).

Таблица 12.10

Клиенты	Удовлетворяют условия оплаты	Сроки выполнения заказов	
		устраивают	не устраивают
Корпоративные	Да	322	33
	Нет	154	32
Частные	Да	270	45
	Нет	182	18

Рассчитайте парные, частные и множественный коэффициенты нормированной информации.

4. Проанализируйте следующие результаты опроса взрослого населения города N-ска (табл. 12.11).

Таблица 12.11

План изменить место жительства	Самооценка уровня доходов		
	Ниже средних	Средние	Выше средних
Планирует	19	51	24
Не планирует	241	501	52
Затрудняется с ответом	16	40	5

Проверьте, имеется ли связь между самооценкой уровня доходов и планами изменения места жительства. Если связь доказана, измерьте тесноту связи с помощью коэффициентов сопряженности.

5. По данным табл. 12.12 установите наличие или отсутствие связи и измерьте тесноту связи между прививкой и заболеваемостью вирусным гриппом.

Таблица 12.12

Прививка	Заболел	Не заболел	Итого
Сделана	10	140	150
Не сделана	40	10	50
Итого	50	150	200

Во сколько раз снижаются шансы заболеть для тех, кто сделал прививку, по сравнению с теми, кто не сделал прививки?

6. Соотношение типа клиента и качества обслуживания представлено в табл. 12.13.

Таблица 12.13

Тип клиента	Качество обслуживания		Итого
	Устраивает	Не устраивает	
Компания	150	18	168
Индивидуальный заказчик	352	113	465
Итого	502	131	633

Проверьте гипотезу об отсутствии связи. В случае ее отклонения измерьте связь с помощью коэффициента контингенции.

7. Измерьте корреляцию между экономической и социальной компонентами потенциала регионов Северо-Западного федерального округа (табл. 12.14).

Таблица 12.14

Регион	Ранг по социальной компоненте	Ранг по экономической компоненте
Республика Карелия	9	7
Республика Коми	2	4
Архангельская обл.	5	6
Вологодская обл.	4	2
Мурманская обл.	3	5
г. Санкт-Петербург	1	1
Ленинградская обл.	6	3
Новгородская обл.	8	8
Псковская обл.	10	10
Калининградская обл.	7	9

8. По данным опроса владельцев легковых автомобилей в крупном городе определите, влияет ли занятие предпринимательством владельца автомобиля на частоту противоправных действий, связанных с автомобилем (табл. 12.15).

Таблица 12.15

Группа населения	Совершался ли в текущем периоде угон автомобиля		
	Да	Нет	Итого
Предприниматели	6	63	69
Другие	19	526	545
Итого	25	589	614
	Совершалась ли кража личных вещей из автомобиля		
	Да	Нет	Итого
Предприниматели	26	42	68
Другие	156	395	551
Итого	182	437	619

9. Распределение ответов опрошенных клиентов об условиях оплаты и сроках выполнения заказов (табл. 12.16).

Таблица 12.16

Клиенты	Удовлетворяют условия оплаты	Сроки исполнения	
		устраивают	не устраивает
Корпоративные	Да	212	58
	Нет	112	12
Частные	Да	357	43
	Нет	92	8

Произведите декомпозицию таблицы сопряженности. Вычислите компонентные значения χ^2 -квadrat. Определите согласованность суммы компонентных значений χ^2 -квadrat. Установите вклад статистически значимых связей в исходное значение χ^2 -квadrat для исходной таблицы.

10. Проанализируйте связь между уровнем денежных расходов жителей малых городов области и лишениями, которые они испытывают в связи с этим (табл. 12.17).

Таблица 12.17

Уровень душевых денежных расходов	Испытывают лишения	
	связанные с питанием	связанные с бытовой техникой
До ПМ	47	5
1 – 1,5 ПМ	32	12
1,5 – 2 ПМ	24	37
2 – 3 ПМ	15	39
3 ПМ и выше	4	11

Примечание. ПМ – прожиточный минимум.

Вычислите асимметричную и симметричную меру связи Гудмена–Краскала.

11. По итогам выборов в Государственную Думу РФ (табл. 12.18) результаты голосования в регионах распределились следующим образом, %.

Таблица 12.18

Партия	Число голосовавших	
	Район А	Район В
Союз правых сил	8	4
Яблоко	16	10
ЛДПР	5	7
КПРФ	5	11
Справедливая Россия	24	28
Единая Россия	42	40

1. В каком районе выше степень неопределенности политических ориентаций взрослого населения?

2. Существенно или несущественно различается доля голосовавших за ЛДПР в районах А и В?

3. В каких пределах измеряется мера неопределенности?

4. Используя формулы асимметричной и симметричной теоретико-информационных мер связи, произведите измерение связи между принадлежностью к определенному району и результатами голосования.

12. Имеются данные о структуре экономики России, % (табл. 12.19).

Таблица 12.19

Год	Сельское хозяйство, рыбная ловля	Промышленность	Сфера услуг
1950	46	29	25
1992	17	36	47

Вычислите показатели структурных сдвигов. Обоснуйте выбор показателя, более эффективного для решения данной задачи.

13. В табл. 12.20 представлено распределение школьников по числу книг домашней библиотеки и образованию родителей.

Таблица 12.20

Образование родителей	Число книг в домашней библиотеке				
	до 50	50-100	100-500	500 и >	Итого
Ниже среднего	30	15	7	3	55
Среднее неполное	10	50	20	10	90
Среднее специальное	5	25	80	40	150
Высшее	3	12	25	70	110
Итого	48	102	132	123	405

Вычислите коэффициент ранговой корреляции Кендалла, используя формулу для сгруппированных данных.

14. Изучается влияние возраста рабочих на отношение к утверждению менеджеров, что заработная плата должна зависеть от производительности труда (табл. 12.21).

Таблица 12.21

Возраст, лет	18–20	20–25	25–30	30–40
Согласен	3	10	25	5
Не согласен	15	15	22	38

Произведите декомпозицию таблицы сопряженности. Вычислите компонентные значения χ^2 -квadrat. Определите согласованность суммы компонентных значений χ^2 -квadrat. Установите вклад случаев статистически значимой связи в исходное значение χ^2 -квadrat (χ^2 -квadrat для исходной таблицы).

15. Распределение опрошенных покупателей по мнению о фасовке и упаковке стирального порошка «Ариэль» (табл. 12.22).

Таблица 12.22

Пол	Устраивает фасовка		Устраивает упаковка
Мужчины	Да	270	28
	Нет	206	37
Женщины	Да	245	36
	Нет	207	27

Сравните результаты ответов мужчин и женщин и сделайте выводы о влиянии пола покупателя на удовлетворенность товаром.

12.3. Вопросы для самоконтроля

1. В чем состоят особенности измерения на номинальной шкале, на порядковой шкале?
2. Как и с какой целью рассчитывается отношение шансов?
3. Сформулируйте определение энтропии распределения.
4. Как определяется максимальная энтропия распределения?
5. В чем состоит особенность применения коэффициента Спирмена в изучении структурных сдвигов?
6. На какой мере связи базируется формула коэффициента Спирмена?

7. Каково соотношение коэффициентов ранговой корреляции М. Кендалла и Ч. Спирмена и при каком условии оно выполняется?

8. В чем состоит особенность проверки гипотезы о связи между двумя дихотомическими переменными?

9. В чем состоит общая основа коэффициентов взаимной сопряженности?

10. Какова интерпретация мер связи Гутмана и Гудмена–Краскала?

11. Какова особенность поведения коэффициентов взаимной сопряженности в интервале $[0,1]$?

12. С помощью каких показателей можно измерять изменение структуры?

13. В каких случаях для измерения структурных сдвигов целесообразно пользоваться индексом различий?

14. Дайте понятие связанных рангов и поясните, в чем состоит специфика расчета коэффициента конкордации.

15. На каком свойстве статистики χ^2 -квадрат основана декомпозиция таблицы сопряженности?

12.4. Методические рекомендации преподавателям

Изучая неколичественные переменные важно, чтобы студенты усвоили особенности измерения на номинальной и интервальной шкале; пояснить, почему в одном случае выполняется свойство транзитивности, а в другом — нет, в чем состоит свойство симметричности.

Говоря об использовании энтропии для оценки степени неопределенности вероятностного распределения, упомяните, что это понятие заимствовано социально-экономической статистикой из термодинамики. Тем самым появляется дополнительная возможность подчеркнуть, что статистика выполняет роль «моста», связывающего естественные и социальные науки. Весьма сложным является вопрос об особенностях поведения разных мер связи в интервале $[0,1]$. Это специальная математическая задача, и ее следует пояснить с точки зрения пользователей: важно понимать, что чувствительность коэффициента корреляции к изменению тесноты связи гораздо бо́льшая, нежели коэффици-

ентов взаимной сопряженности, а чувствительность последних выше, чем теоретико-информационных мер связей.

12.5. Методические указания студентам

Изучение этой темы интересно тем, что можно использовать много материала из повседневной практики в виде высказываний, утверждений, рейтингов (вузов, банков, артистов, фигуристов, шахматистов и др.), разного рода балльных оценок. Вроде бы такие данные не поддаются статистической обработке, но это только для непосвященного. Разобравшись в теоретическом материале, вы поймете, что, основываясь на частоте событий и их сочетаниях, можно провести статистический анализ, включая измерение связи и оценку изменения структуры.

Статистическое изучение динамики

13.1. Основные положения

Основой изучения динамики выступает ряд динамики или временной ряд.

Временной ряд — это таблица, в которой представлены значения показателя за последовательные периоды или на моменты времени. Временные ряды делятся на *моментные* и *интервальные*; построенные по абсолютным или по средним и относительным показателям. Уровни временного ряда должны быть сопоставимы друг с другом по методике расчета, по территории охвата, единицам измерения и счета. При изучении динамики необходимо выделять тенденцию и колеблемость:

$$y_t = \bar{y}_t \pm d_t,$$

где y_t — наблюдаемое значение показателя y в t -й период (момент) времени;

\bar{y}_t — его трендовое значение в t -й период (момент) времени;

d_t — отклонение от тренда, $d_t = y_t - \bar{y}_t$.

Если тренда нет, то колебания происходят относительно некоего постоянного уровня (среднего):

$$y_t = \bar{y} \pm d_t.$$

К показателям, характеризующим *тенденцию динамики*, относятся:

- абсолютное изменение уровней (абсолютный прирост):

$$\text{цепной } \Delta_t = y_t - y_{t-1},$$

$$\text{базисный } \Delta_t = y_t - y_0;$$

- ускорение $\Delta'_t = \Delta_t - \Delta_{t-1}$ (определяется на основе цепных абсолютных приростов);

- темп роста (относительный рост):

$$\text{цепной } k_t = \frac{y_t}{y_{t-1}},$$

$$\text{базисный } k_t = \frac{y_t}{y_0};$$

- темп прироста $(k_t - 100\%)$ или $(k_t - 1)$:
цепной,
базисный.

Средние показатели динамики:

- *средний уровень интервального ряда*

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n};$$

- *средний уровень моментного ряда*

$$\bar{y} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i},$$

где t_i — время, в течение которого сохранялось состояние y_i ;

- *средний абсолютный прирост*

$$\bar{\Delta} = \frac{\sum_{t=1}^n \Delta_t}{n} = \frac{y_n - y_0}{n},$$

где Δ_t — цепной абсолютный прирост;

- *средний темп динамики* (средний относительный рост)

$$\bar{k} = \sqrt[n]{\prod_{t=1}^n k_t} = \sqrt[n]{\frac{y_n}{y_0}},$$

где k_t — цепной темп динамики.

Очевидно, что цепные и базисные показатели связаны между собой.

Для выявления тренда проводится аналитическое выравнивание в соответствии с выбранным типом уравнения тренда

$$\hat{y} = f(t):$$

- линейный тренд $\hat{y} = a + bt$;
- параболический тренд $\hat{y} = a + bt + ct^2$;
- экспоненциальный тренд $\hat{y} = ak^t$;
- логарифмический тренд $\hat{y} = a + b \log t$;
- степенной тренд $\hat{y} = at^b$;
- гиперболический тренд $\hat{y} = a + \frac{b}{t}$;
- логистический тренд $\hat{y} = \frac{y_{\max} - y_{\min}}{e^{a+bt} + 1} + y_{\min}$,

где e — основание натуральных логарифмов.

Параметры соответствующей формы тренда находятся методом наименьших квадратов.

Виды колеблемости:

- пилообразная, или маятниковая (за счет сезонности, S);
- долгопериодическая циклическая (большой цикл, C);
- случайно распределенные во времени (случайные, E).

С учетом этого наблюдаемый уровень временного ряда может быть представлен как

$$y_t = T + S + C + E,$$

где T — trend; S — season; C — cycle; E — error,
или

$$y_t = T + D,$$

где $D = S + C + E$.

Для выявления больших циклов требуются весьма протяженные ряды динамики, включающие не одну сотню уровней, поэтому их обычно не выявляют.

Показатели колебаний:

- абсолютный

$$d_{ti} = |y_{ti} - \bar{y}_{ti}|;$$

- средний

$$\bar{d} = \frac{\sum_{(i)} |y_{ti} - \bar{y}_{ti}|}{n} = \frac{\sum d_{ti}}{n};$$

- среднеквадратический показатель колеблемости

$$S_d = \sqrt{\frac{\sum_{(i)} (y_{ti} - \bar{y}_{ti})^2}{n}};$$

- коэффициент колеблемости

$$k = \frac{\bar{d}}{y} \cdot 100\%;$$

- коэффициент устойчивости $k_s = 100\% - k$.

13.1.1. Разложение вариации уровней ряда динамики на составляющие элементы

Изучая ряд динамики какого-либо статистико-экономического показателя, вариацию уровней ряда можно разложить на составляющие элементы: изменение за счет тренда, регулярные, в частности сезонные колебания, и нерегулярные случайные колебания, т.е. не имеющие надежно установленной корреляции ни с трендом, ни с сезонностью. Проведем такой анализ на примере ряда динамики розничного товарооборота России за 2000–2002 гг. (табл. 13.1).

Таблица 13.1

Динамика розничного товарооборота, млрд руб.

Год, квартал	Уровень y_i за t_i	Номер периода t_i	z_1	z_2	z_3	Тренд	Модель	$U_{\text{сп}}$	$U_{\text{сп}}^2$	$U_{\text{сез}}$	$U_{\text{сез}}^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1999, IV	535	1	1	0	0	481	535	0	0	54	2916	-206	42436
2000, I	514	2	0	1	0	524	490	24	576	-34	1156	-227	51529
2000, II	539	3	0	0	1	567	538	1	1	-29	841	-202	40804
2000, III	594	4	0	0	0	611	601	-7	49	-10	100	-147	21609
2000, IV	689	5	1	0	0	654	708	-19	361	54	2916	-52	2704
2001, I	659	6	0	1	0	697	664	-5	25	-33	1089	-82	6724
2001, II	722	7	0	0	1	741	712	10	100	-29	841	-19	301
2001, III	776	8	0	0	0	784	774	2	4	-10	100	35	1225
2001, IV	883	9	1	0	0	827	882	1	1	55	3025	142	20164
2002, I	816	10	0	1	0	871	836	-20	400	-35	1225	75	5625
2002, II	874	11	0	0	1	914	885	-11	121	-29	841	133	17689
2002, III	952	12	0	0	0	957	947	5	25	-10	100	211	44521
2002, IV	1077	13	1	0	0	1001	1055	22	484	54	2916	336	112896
Итого	9630	—	—	—	—	9629	9627	—	2147	—	18066	—	368287

Средний уровень равен

$$\bar{y} = \frac{9630}{13} = 741.$$

Так как сезонность розничного товарооборота в основном состоит в значительном повышении в IV квартале, чтобы сезонность не искажала параметры тренда, необходимо, чтобы четвертые кварталы располагались равномерно по отношению к началу и концу ряда. Поэтому приходится начать ряд с уровня IV квартала 1999 г., а закончить IV кварталом 2002 г. Для отражения в модели ряда динамики сезонных колебаний применим аддитивную форму модели с тремя структурными переменными $z_1 = 1$ для четвертых кварталов; $z_2 = 1$ для первых кварталов; $z_3 = 1$ для вторых кварталов. Для учета тренда в модель включены номера периодов от начала ряда t_i . В результате расчета на компьютере получена следующая модель ряда с учетом тренда и сезонных колебаний:

$$\hat{y}_t = 427,3 + 43,34t + 64,36z_1 - 24,3z_2 - 19,0z_3.$$

Статистически надежно отличны от нуля следующие коэффициенты модели: свободный член, коэффициент при t и при z_1 ; недостаточно надежны коэффициенты при z_2 и z_3 .

Модель в целом имеет высокую надежность: F -критерий равен 312; средняя квадратическая ошибка по модели равна 17,3, что составляет 2,4% от среднего уровня. Коэффициент при t означает величину среднего квартального прироста розничного товарооборота; коэффициент при z_1 означает, что уровни товарооборота в четвертых кварталах в среднем на 64,36 млрд руб. выше, чем были бы в принятом за базу сравнения III квартале при том же значении номера периода t . Так как третьи кварталы имеют номера периодов на единицы меньше, чем четвертые кварталы, то уровни модели третьих кварталов ниже, чем четвертых кварталов тех же лет, еще на 43,34, а всего на $43,34 + 64,36 \approx 108$ млрд руб. ниже уровней четвертых кварталов, что и видим в табл. 13.1. Коэффициент при z_2 означает, что уровни товарооборота в первых кварталах соответственно на 24,3 млрд руб., а во вторых кварталах — при z_3 — на 19 млрд руб. ниже, чем уровни четвертых кварталов при том же номере периода t_i .

В последней графе табл. 13.1 показана общая вариация уровней $\sum_{i=1}^{13} (y_i - \bar{y})^2$, равная 368287. Сумма квадратов различий уровней

за счет тренда $\sum_{i=1}^{13} (y_i - \bar{y})^2$ равна 341861. Эта сумма может быть определена как произведение коэффициента при t в квадрате на сумму квадратов $(t_i - \bar{t})^2$, которая, как известно, равна $\frac{n^3 - n}{12}$. Для нашего примера $\frac{13^3 - 13}{12} = 182$. Имеем: $43,34^2 \times 182 = 341861$.

Сумма квадратов за счет сезонной колеблемости показана в графе таблицы, обозначенной $U_{\text{сез}}^2$, — 18066; сумма квадратов случайных колебаний приведена в графе $U_{\text{сл}}^2$ — 2147.

Суммируя три полученных составляющих ($341861 + 18066 + 2147$), получаем 362074, что несколько меньше суммы квадратов отклонений фактических уровней от среднего уровня (368287). Различие объясняется тем, что хотя случайная колеблемость и не имеет надежной корреляции с трендом и с сезонностью, но при ограниченной выборке (длине ряда динамики) ненадежная корреляция все же существует, и она может быть как прямой, так и обратной ввиду того, что сумма трех компонент, измеренных порознь, может быть как меньше, так и больше суммы неразделенной на компоненты вариации уровней.

Чтобы правильно измерить доли компонент вариации, необходимо скорректировать значения компонент, умножая каждую из них на отношение фактической суммы квадратов к сумме компонент, т.е. на величину $\frac{368287}{362074} = 1,01716$.

Произведем корректировку и определим вклад каждой компоненты:

- 1) за счет тренда: $341861 \cdot 1,01716 = 347727$, или 94,4%;
- 2) за счет сезонных колебаний: $18066 \cdot 1,01716 = 18376$, или 5%;

3) за счет случайных колебаний: $3147 \cdot 1,01716 = 2184$, или 0,6%

$$\frac{\quad}{\text{всего } 368287} = 100\%$$

Основная причина различия уровней розничного товарооборота в России за три года — это тренд, который формировался в основном за счет инфляции. Сезонность отвечает за 5% различий уровней, а роль случайных колебаний практически ничтожна.

Следует, однако, помнить, что сравнивать роль разных компонент вариации уровней можно только при равной длине рядов. С увеличением числа уровней n , как видно из вышеприведенной формулы, сумма квадратов отклонений за счет тренда возрастает пропорционально n^3 , а сумма квадратов отклонений за счет сезонности и случайных колебаний возрастает только пропорционально n . Поэтому при достаточно длинном ряде, даже при слабом тренде, сумма квадратов отклонений за счет тренда станет больше, чем сумма квадратов сколь угодно сильных колебаний.

13.1.2. Прогнозирование по тренду и сезонной компоненте

Расчет возможен для аддитивной и мультипликативной модели временного ряда.

Элементы аддитивной модели суммируются:

$$y_i = T + S + E.$$

Элементы мультипликативной модели перемножаются:

$$y_i = T \cdot S \cdot E.$$

Приблизительно равная сезонная вариация указывает на существование аддитивной модели; усиление сезонной вариации с возрастанием тренда указывает на существование мультипликативной модели.

Прогнозирование с учетом сезонной компоненты в аддитивной и мультипликативной моделях происходит примерно одинаково в такой последовательности:

1) выравнивание исходного временного ряда методом скользящей средней;

2) расчет значений сезонной компоненты;

- 3) устранение сезонной компоненты из исходных уровней ряда и получение десезонализированных данных;
- 4) расчет уравнения тренда по десезонализированным данным;
- 5) прогнозирование по тренду с учетом сезонной компоненты.

Пример 13.1. Прогнозирование на основе аддитивной модели.

Известно количество проданной продукции в течение 13 кварталов (табл. 13.2).

Таблица 13.2

Год	Квартал	Объем продаж, тыс. шт.
1-й	I	239
	II	201
	III	182
	IV	297
2-й	I	324
	II	278
	III	257
	IV	384
3-й	I	401
	II	360
	III	335
	IV	462
4-й	I	481

Как следует из диаграммы (рис. 13.1), возможен нарастающий тренд, содержащий сезонные колебания. Объемы продаж в зимний период значительно выше, чем в летний. Сезонная компонента практически не изменяется в течение трех лет. Объем продаж возрос, однако увеличения сезонных колебаний не произошло. Этот факт свидетельствует в пользу модели с аддитивной компонентой.

Расчет сезонной компоненты по аддитивной модели и расчет ее средних значений независимо от особенностей конкретного года приведены в табл. 13.3 и 13.4.

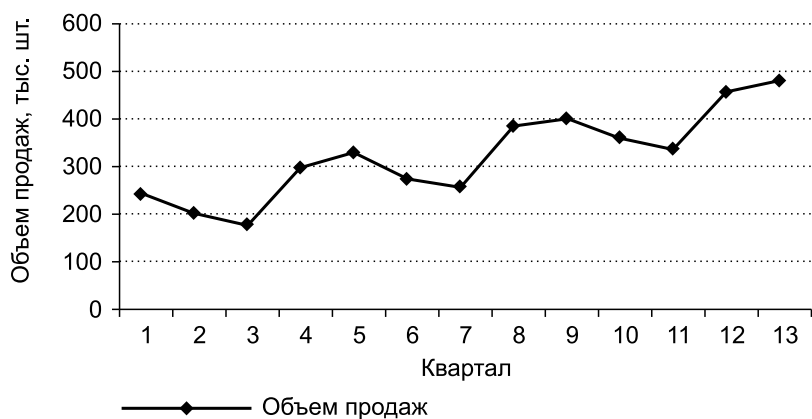


Рис. 13.1. Динамика объема продаж

Таблица 13.3

Расчет значений сезонной компоненты

Год	Квар- тал	Объем продаж, тыс. шт., Y_t	Скользя- щая сред- няя за че- тыре квартала	Центриро- ванная скользящая средняя S_t	Оценка сезонной компоненты $Y_t - T = S + E$
1	2	3	4	5	6 = (гр. 3 – гр. 5)
1-й	I	239	–		
	II	201	–		
			229,75		
	III	182		240,4	–58,4
			251		
	IV	297		260,6	36,4
			270,25		
2-й	I	324		279,6	44,4
			289		
	II	278		299,9	–21,9
			310,75		
	III	257		320,4	–63,4
			330		
	IV	384		340,3	43,8
			350,5		

Продолжение

Год	Квар- тал	Объем продаж, тыс. шт., Y_t	Скользя- щая сред- няя за че- тыре квартала	Центриро- ванная скользящая средняя S_t	Оценка сезонной компоненты $Y_t - T = S + E$
1	2	3	4	5	6 = (гр. 3 – гр. 5)
3-й	I	401		360,3	40,8
			370		
	II	360		379,8	-19,8
			389,5		
	III	335		399,5	-64,5
			409,5		
4-й	IV	462	–		
	I	481	–		

Таблица 13.4

Расчет средних значений сезонной компоненты

	Год	Квартал				
	1-й	I	II	III	IV	
				−58,4	36,4	
		2-й	44,4	−21,9	−63,4	43,8
		3-й	40,8	−19,8	−64,5	
Итого		85,2	−41,7	−186,3	80,2	
Среднее значение (оценка сезонной компоненты)		42,6	−20,8	−62,1	40,1	Сумма=−0,2
Скорректирован- ная сезонная компонента S		42,65	−20,55	−62,15	40,05	Сумма = 0

Для определения скорректированной сезонной компоненты сумму средних значений надо разделить на четыре квартала ($-0,2 : 4 = -0,05$). Затем из каждого среднего значения сезонной компоненты вычесть это число (например, $42,6 - (-0,05) = 42,65$).

Значения сезонной компоненты подтверждают выводы, сделанные на основе диаграммы (см. рис. 13.1). Объемы продаж за два зимних квартала превышают среднее трендовое значение на 43 и 40 тыс. шт., а объемы продаж за два летних периода ниже средних на 21 и 62 тыс. шт. соответственно.

Аналогичная процедура применима при определении сезонной вариации за любой промежуток времени. Если, например, в качестве сезонов выступают дни недели, то расчет скользящей средней производится не по четырем, а по семи точкам. Эта скользящая средняя представляет собой значение тренда в середине недели, т. е. в четверг; таким образом, необходимость в процедуре центрирования отпадает.

Устранение сезонной компоненты, т.е. построение десезонализированного ряда, показано в табл. 13.5.

Таблица 13.5

Устранение сезонной компоненты (десезонализация данных)

Год	Квартал	Объем продаж, тыс. шт., Y_t	Скорректированная сезонная компонента S	Десезонализированный объем продаж, тыс. шт. $Y_t - S = T + E$
1	2	3	4	5 = (гр. 3 – гр. 4)
1-й	I	239	42,65	196,35
	II	201	-20,55	221,55
	III	182	-62,15	244,15
	IV	297	40,05	256,95
2-й	I	324	42,65	281,35
	II	278	-20,55	298,55
	III	257	-62,15	319,15
	IV	384	40,05	343,95
3-й	I	401	42,65	358,35
	II	360	-20,55	380,55
	III	335	-62,15	397,15
	IV	462	40,05	421,95
4-й	I	481	42,65	438,35

На основе десезонализированного объема продаж (табл. 13.6) построим уравнение тренда (рис. 13.2 и табл. 13.7).

Таблица 13.6

Порядковый номер квартала	Десезонализированный объем продаж	Порядковый номер квартала	Десезонализированный объем продаж
1	196,35	8	343,95
2	221,55	9	358,35
3	244,15	10	380,55
4	256,95	11	397,15
5	281,35	12	421,95
6	298,55	13	438,35
7	319,15		

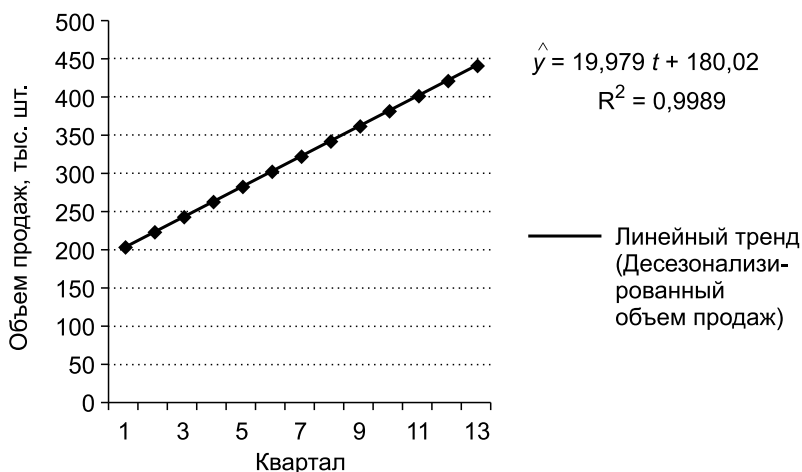


Рис. 13.2. Уравнение тренда объема продаж

Таблица 13.7

Прогнозирование по тренду с учетом сезонной компоненты

Год	Квартал	Трендовый прогноз	Скорректированная сезонная компонента S	Прогноз с учетом сезонной компоненты
1	2	3	4	5 = (гр. 3 + гр. 4)
5-й	I	438,35	42,65	481,00
	II	460,20	-20,55	439,65
	III	480,20	-62,15	418,05
	IV	500,20	40,05	540,25

Для расчета прогнозных значений объема продаж подставим в найденное уравнение регрессии порядковый номер квартала. Например, для нахождения объема продаж за II квартал четвертого года подставим в уравнение регрессии значение 14, за III квартал – 15, за IV – 16.

Пример 13.2. Прогнозирование на основе мультипликативной модели.

Известно количество проданной продукции в течение последних 13 кварталов (табл. 13.8).

Таблица 13.8

Год	Квартал	Объем продаж, тыс шт.
1-й	I	70
	II	66
	III	65
	IV	71
2-й	I	79
	II	66
	III	67
	IV	82
3-й	I	84
	II	69
	III	72
	IV	87
4-й	I	94

Объем продаж, как и в примере 13.1, подвержен сезонным колебаниям, и значения его в зимний период выше, чем в летний. Однако имеет место усиление сезонной вариации с возрастанием тренда (рис. 13.3). К таким данным следует применять мультипликативную модель (табл. 13.9).

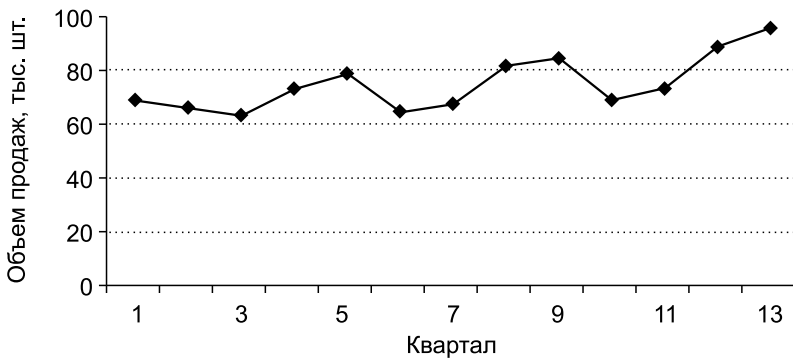


Рис. 13.3. Динамика объема продаж

Таблица 13.9

Расчет значений сезонной компоненты

Год	Квартал	Объем продаж, тыс шт., Y_t	Скользящая средняя за четыре квартала	Центрированная скользящая средняя S_t	Коэффициент сезонности $Y_t : T = S \cdot E$
1	2	3	4	5	6 = (гр. 3 : гр. 5)
1-й	I	70			
	II	66			
			68		
	III	65		69,13	0,940
			70,25		
	IV	71		70,25	1,011
			70,25		
2-й	I	79		70,50	1,121
			70,75		
	II	66		72,13	0,915
			73,5		
	III	67		74,13	0,904
			74,75		
	IV	82		75,13	1,092
			75,5		
3-й	I	84		76,13	1,103
			76,75		
	II	69		77,38	0,892
			78		
	III	72		79,25	0,909
			80,5		
	IV	87			
4-й	I	94			

Таблица 13.10

**Расчет средних значений
сезонной компоненты**

	Год	Квартал				
		I	II	III	IV	
	1-й			0,94	1,011	
	2-й	1,121	0,915	0,904	1,092	
	3-й	1,103	0,892	0,909		
Итого		2,224	1,807	2,753	2,103	
Среднее значение (оценка сезонной компоненты)		2,224:2= =1,112	1,807:2= =0,903	2,753:3= =0,918	2,103:2= =1,051	Сумма = 3,984
Скорректированная сезонная компонента <i>S</i>		1,116	0,907	0,922	1,055	Сумма = 4,0

Значения сезонных коэффициентов (табл. 13.10) получены на основе квартальных оценок по аналогии с алгоритмом, который применялся для аддитивной модели. Так как значения сезонной компоненты – это доли, а число сезонов равно четырем, необходимо, чтобы их сумма была равна четырем, а не нулю, как в предыдущем случае. (Если бы в исходных данных предполагалось семь сезонов в течение недели по одному дню каждый, то общая сумма значений сезонной компоненты должна была бы равняться семи.)

Скорректированная сезонная компонента получена путем деления среднего значения сезонной компоненты на коэффициент корректировки ($3,984 : 4 = 0,996$).

Как показывают оценки (табл. 13.11), в результате сезонных воздействий объемы продаж в I квартале увеличиваются на 11,6% соответствующего значения тренда (1,116); в IV квартале – на 5,5%. В двух других кварталах сезонные воздействия состоят в снижении объемов продаж (на 9,3 и 7,8%).

Таблица 13.11

Устранение сезонной компоненты (десезонализация данных)

Год	Квартал	Объем продаж, тыс шт., Y_t	Скорректированная сезонная компонента S	Десезонализированный объем продаж, тыс шт., $Y_t: S = T \cdot E$
1	2	3	4	5 = (гр. 3 : гр. 4)
1-й	I	70	1,116	62,7
	II	66	0,907	72,8
	III	65	0,922	70,5
	IV	71	1,055	67,3
2-й	I	79	1,116	70,8
	II	66	0,907	72,8
	III	67	0,922	72,7
	IV	82	1,055	77,7
3-й	I	84	1,116	75,3
	II	69	0,907	76,1
	III	72	0,922	78,1
	IV	87	1,055	82,5
4-й	I	94	1,116	84,2

На основе десезонализированного объема продаж (табл. 13.12) построим уравнение тренда и представим его на графике (рис. 13.4).

Таблица 13.12

Порядковый номер квартала	Десезонализированный объем продаж	Порядковый номер квартала	Десезонализированный объем продаж
1	62,72	8	77,73
2	72,77	9	75,27
3	70,50	10	76,07
4	67,30	11	78,09
5	70,79	12	82,46
6	72,77	13	84,23
7	72,67		

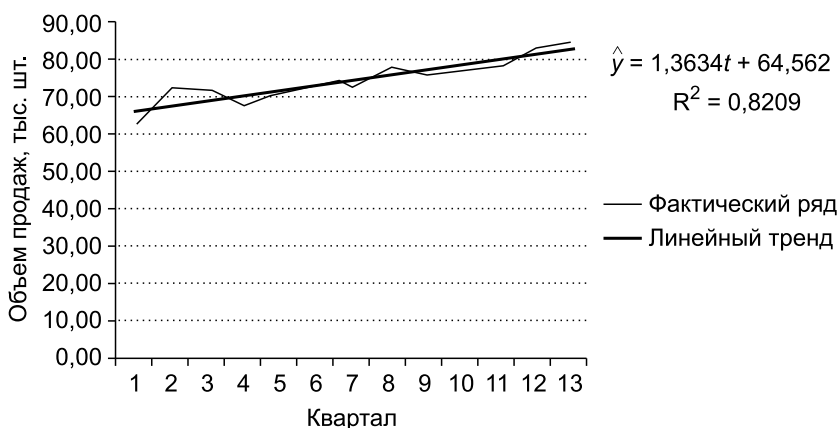


Рис. 13.4. Уравнение тренда

На основе уравнения тренда и коэффициентов сезонности рассчитаем прогнозное значение для следующего года (табл. 13.13).

Таблица 13.13

Прогнозирование по тренду с учетом сезонной компоненты

Год	Квартал	Трендовый прогноз	Скорректированная сезонная компонента S	Прогноз с учетом сезонной компоненты
1	2	3	4	$5 = (\text{гр. 3} \cdot \text{гр. 4})$
5-й	I	84,23	1,116	94,00
	II	83,64	0,907	75,86
	III	85	0,922	78,37
	IV	86,36	1,055	91,11

13.2. Задания по теме

1. На основе данных об объеме отгруженной продукции предприятия за девять месяцев с начала года построено уравнение тренда: $y = 8,19 + 1,11t$, где $t = 1, 2, 3, \dots$. Методом экстраполяции определите прогнозное значение объема отгруженной продукции на декабрь текущего года.

Варианты ответа:

- а) 11,52 млн руб.;
- б) 21,51 млн руб.;
- в) иное.

2. Сезонность реализации желудочных препаратов фармацевтической фирмой характеризуется следующими индексами сезонности, %: I квартал – 90,0; II квартал – 107,5; III квартал – 105,3; IV квартал – 97,2. Средний квартальный уровень реализации этих препаратов за три последних года составил 13 млн усл. ед. В предстоящем году предполагается увеличить средний квартальный уровень на 2 млн усл. ед. Определите плановый объем продаж на I квартал (млн усл. ед.).

Варианты ответа:

- а) 13,5;
- б) 11,7;
- в) 15,0;
- г) 14,8.

3. Имеется следующий ряд динамики производства продукции на предприятии, тыс. шт. (табл. 13.14).

Таблица 13.14

Год	2000	2001	2002	2003	2004	2005	2006
Производство, тыс. шт.	32,3	32,3	33,3	34,5	38,5	35,6	37,2

Определите средний уровень производства продукции; среднегодовой абсолютный прирост и среднегодовой темп роста и прироста. Выберите функцию тренда выпуска продукции и вычислите ее параметры. На основе уровня тренда дайте точечный и интервальный прогноз объема производства в 2008 г.

4. Месячный уровень инфляции – 1%. При сохранившейся тенденции за год инфляция составит, %:

- а) 12;
- б) более 12;
- в) менее 12.

5. Имущество предприятия составляло, млн руб.:

- на 01.01 – 20;
- на 01.04 – 30;
- на 01.07 – 32;
- на 01.10 – 35;
- на 01.01 следующего года – 40.

Назовите среднегодовую стоимость имущества, млн. руб.:

- а) 31,75;
- б) 31,40;
- в) 25,40.

6. Возрастная структура женщин в Санкт-Петербурге представлена в табл. 13.15.

Таблица 13.15

Год	Женщины, тыс. чел.	
	всего	в трудоспособном возрасте
1992	2720,4	1468,9
1993	2691,3	1435,5
1994	2654,4	1407,7
1995	2631,8	1391,9
1996	2613,6	1388,1
1997	2601,9	1391,6
1998	2586,1	1404,9
1999	2576,3	1425,8
2000	2559,1	1439,2
2001	2544,0	1443,6
2002	2529,1	1436,1
2003	2513,9	1427,8

Вычислите обобщающие показатели динамики и характеристики тренда. Постройте график динамики общей численности женщин и численности женщин в трудоспособном возрасте. Сделайте вывод о характере динамики возрастной структуры женщин в городе.

7. Имеются следующие данные об объеме выпущенной продукции А на заводе (табл. 13.16).

Таблица 13.16

Месяц	Выпуск продукции, тыс. ед.	Изменение по сравнению с предыдущим месяцем			
		Абсолютный прирост, тыс. ед.	Темп роста, %	Темп прироста, %	Абсолютное значение 1% прироста, тыс. ед.
Январь	800				
Февраль		170			
Март			90,0		
Апрель				+3,0	
Май					
Июнь		–20			9,0

Рассчитайте отсутствующие в таблице показатели. Определите средний выпуск за период, средний абсолютный прирост и среднемесячный темп динамики выпуска продукции.

8. По данным о курсе рубля и доллара США за 15 рабочих дней финансовой биржи (табл. 13.17) постройте уравнение тренда, измерьте колеблемость курса и сделайте прогноз возможного курса на 20-й день работы биржи с вероятностью 0,95.

Таблица 13.17

День работы биржи	Курс рубля за 1 долл.	День работы биржи	Курс рубля за 1 долл.
1	28,51	9	28,56
2	28,53	10	28,62
3	28,61	11	28,60
4	28,64	12	28,59
5	28,62	13	28,53
6	28,52	14	28,61
7	28,46	15	28,67
8	28,47		

9. По данным о производстве молока в России за 2001–2004 гг. (табл. 13.18) постройте аддитивную и мультипликативную модель сезонных колебаний и сделайте прогнозы: на III квартал 2005 г. и на I квартал 2006 г.

Таблица 13.18

Год, квартал	Производство, тыс. т	Год, квартал	Производство, тыс. т
2001, I	5885	2003, I	6369
2001, II	10776	2003, II	10537
2001, III	10439	2003, III	10418
2001, IV	5815	2003, IV	6050
2002, I	6248	2004, I	6147
2002, II	10828	2004, II	10050
2002, III	10336	2004, III	9910
2002, IV	6065	2004, IV	5885

10. По данным о средней за месяцы температуре воздуха в Санкт-Петербурге за 2000–2004 гг. (табл. 13.19) постройте аддитивную и мультипликативную модели сезонных колебаний, измерьте силу сезонных и случайных колебаний, сделайте прогно-

зы средней температуры на июль 2006 г. и на сентябрь 2007 г. с вероятностью 0,95.

Таблица 13.19

Месяц	2000	2001	2002	2003	2004
Январь	−3,7	−2,1	−3,6	−9,2	−7,8
Февраль	−1,7	−6,1	+0,1	−4,8	−5,0
Март	+0,2	−2,8	+1,2	−0,9	−0,2
Апрель	+8,7	+9,1	+6,5	+3,2	+5,2
Май	+11,5	+10,7	+12,7	+13,0	+11,2
Июнь	+16,2	+16,0	+17,0	+13,8	+14,7
Июль	+18,9	+22,2	+21,1	+21,9	+18,2
Август	+16,9	+17,0	+18,8	+17,0	+17,6
Сентябрь	+10,8	+12,8	+11,8	+12,6	+12,9
Октябрь	+8,9	+7,9	+2,0	+5,4	+6,5
Ноябрь	+2,8	−0,7	−2,0	+2,7	−0,1
Декабрь	−0,3	−8,4	−9,3	−0,8	−0,5

11. По данным о курсе рубля к доллару США и к евро (табл. 13.20):

- вычислите средний курс за 10 дней;
- определите показатели колебаний курсов;
- измерьте корреляцию между колебаниями курсов доллара и евро;
- дайте прогноз курсов за 3 дня вперед.

Таблица 13.20

Дата	Рублей за 1 долл.	Рублей за 1 евро
11.10.05	28,47	34,53
12.10.05	28,56	34,23
13.10.05	28,62	34,24
14.10.05	28,59	34,29
15.10.05	28,58	34,58
18.10.05	28,52	34,47
19.10.05	28,61	34,30
20.10.05	28,67	34,12
21.10.05	28,62	34,19
22.10.05	28,56	34,40

12. По данным о перевозках грузов на железных дорогах России за 2002–2004 гг. (табл. 13.21) определите, существен ли тренд, измерьте сезонные колебания и постройте модель динамики с учетом тренда и сезонности. Рассчитайте прогноз перевозки грузов на I квартал 2006 г.

Таблица 13.21

Год, квартал	Перевозка, млн т	Год, квартал	Перевозка, млн т
2002, I	248	2003, III	298
2002, II	270	2003, IV	302
2002, III	282	2004, I	289
2002, IV	284	2004, II	304
2003, I	269	2004, III	312
2003, II	291	2004, IV	315

13. По данным о среднегодовой температуре воздуха в Ленинграде-Петербурге за 1964–2004 гг. (табл. 13.22):

а) определите, имеется ли надежный тренд среднегодовой температуры;

б) если тренд существенный, то сделайте прогнозы среднегодовой температуры на 2010 и 2025 гг. с вероятностью 0,9.

14. По данным о производстве молока в России за 2001 – 2004 гг. (табл. 13.23) постройте аддитивную и мультипликативную модели сезонных колебаний и сделайте прогнозы на III квартал 2005 г. и на I квартал 2006 г.

15. По данным о средней температуре воздуха в Санкт-Петербурге за каждый месяц 2000–2004 гг. (табл. 13.24) постройте аддитивную и мультипликативную модели сезонных колебаний, изменений сезонных и случайных колебаний. Сделайте прогнозы средней температуры на июль 2006 г. и на сентябрь 2007 г. с вероятностью 0,95.

Таблица 13.22

Год	\bar{r}	Год	\bar{r}	Год	\bar{r}	Год	\bar{r}	Год	\bar{r}
1964	5,23	1973	5,09	1982	5,22	1991	6,24	2000	7,43
1965	4,30	1974	6,72	1983	6,10	1992	6,30	2001	6,30
1966	3,53	1975	6,44	1984	5,85	1993	5,25	2002	6,39
1967	5,58	1976	2,90	1985	3,62	1994	5,53	2003	6,16
1968	4,26	1977	4,69	1986	4,85	1995	6,46	2004	6,06
1969	3,77	1978	3,77	1987	3,22	1996	5,19		
1970	4,94	1979	4,81	1988	5,86	1997	5,74		
1971	4,78	1980	4,47	1989	7,44	1998	5,38		
1972	5,82	1981	5,32	1990	6,38	1999	6,77		

Таблица 13.23

Год, квартал	Молоко, тыс. т	Год, квартал	Молоко, тыс. т
2001, I	5885	2003, I	6369
2001, II	10776	2003, II	10537
2001, III	10439	2003, III	10418
2001, IV	5815	2003, IV	6050
2002, I	6248	2004, I	6147
2002, II	10828	2004, II	10050
2002, III	10336	2004, III	9910
2002, IV	6065	2004, IV	5885

Таблица 13.24

Месяц	2000	2001	2002	2003	2004
Январь	−3,7	−2,1	−3,6	−9,2	−7,8
Февраль	−1,7	−6,1	+0,1	−4,8	−5,0
Март	+0,2	−2,8	+1,2	−0,9	−0,2
Апрель	+8,7	+9,1	+6,5	+3,2	+5,2
Май	+11,5	+10,7	+18,7	+13,0	+11,2
Июнь	+16,2	+16,0	+17,0	+13,8	+14,7
Июль	+18,9	+22,2	+21,1	+21,9	+18,2
Август	+16,9	+17,0	+18,8	+17,0	+17,6
Сентябрь	+10,8	+12,8	+11,8	+12,6	+12,9
Октябрь	+8,9	+7,9	+2,0	+5,4	+6,5
Ноябрь	+2,8	−0,7	−2,0	+2,7	−0,1
Декабрь	−0,3	−8,4	−9,3	−0,8	−0,5

16. По данным о перевозках грузов на железных дорогах России за 2002–2004 гг. (табл. 13.25) определите, существен ли тренд. Измерьте сезонные колебания и постройте модель динамики и с учетом тренда и сезонности. Рассчитайте прогноз перевозки грузов на I квартал 2006 г.

Таблица 13.25

Год, квартал	Перевозка, млн т	Год, квартал	Перевозка, млн т
2002, I	248	2003, III	298
2002, II	270	2003, IV	302
2002, III	282	2004, I	289
2002, IV	284	2004, II	304
2003, I	269	2004, III	312
2003, II	291	2004, IV	315

13.3. Вопросы для самоконтроля

1. На основе чего выделяются виды рядов динамики?
2. Каковы порядок построения рядов динамики, проблемы интерполяции и построения системы рядов динамики?
3. В чем состоят особенности графического представления рядов динамики?
4. Назовите показатели динамики, поясните их аналитическое значение, взаимосвязь между ними.
5. Поясните значение и порядок расчета средних уровней и средних показателей динамики.
6. Каковы основные компоненты временного ряда?
7. Каковы методы выявления тенденции динамики?
8. В чем состоит сущность выравнивания по скользящей средней?
9. Как производятся выбор уравнения тренда и оценка параметров?
10. Как измеряются сезонные колебания в стационарных и нестационарных рядах динамики?
11. Назовите методы прогнозирования на основе временных рядов.

13.4. Методические рекомендации преподавателям

Интерес студентов к проблемам статистического изучения динамики во многом зависит от характера данных, по которым вы обсуждаете основные положения. Наибольшие возможности для прослеживания долгосрочных закономерностей дают сведения о численности населения мира. Затухающие темпы можно демонстрировать на примере российского экспорта или добычи нефти. В любом случае обсуждение динамики реальных данных весьма полезно для студентов, но требует большой подготовки преподавателя в соответствующей области экономики.

Рекомендуется давать распределенное задание. Например, разделить студентов на группы и каждой группе поручить анализ динамики основных макроэкономических показателей (ВРП на душу, уровень инфляции, инвестиции в процентах к ВРП и др.) по субъектам РФ в пределах федерального округа. По результатам расчетов студенты должны будут составить аналитический обзор о сравнительной динамике субъектов РФ и федеральных округов.

Анализ сезонности целесообразно проводить по месячным или квартальным данным о потребительских ценах на продовольственные товары, непродовольственные товары, услуги, а также о потреблении электроэнергии, воды населением. Большой интерес вызывает у студентов анализ сезонности смертности, рождаемости, брачности, разводимости. Все эти данные публикуются в официальных статистических сборниках, издаваемых территориальным комитетом по статистике.

13.5. Методические указания студентам

При изучении темы обратите внимание на характер данных: являются ли они условными или отражают реальную динамику. Если это фактические данные — уделите особое внимание интерпретации результатов анализа.

Расчет показателей динамики, а также параметров трендов целесообразно выполнять в Excel; аналитическое выравнивание может выполняться с помощью ППП Statgraphics, SPSS.

Индексы

14.1. Основные положения

14.1.1. Индивидуальные индексы

Индекс — это показатель сравнения двух состояний одного и того же явления.

При расчете индексов используют два вида данных: оцениваемые, которые принято называть *отчетные* и обозначать подстрочным значком «1», и данные, которые выступают в качестве базы сравнения, — *базисные*, обозначаемые значком «0».

Индекс, рассчитанный по отдельным единицам изучаемой совокупности, называется **индивидуальным** и обозначается символом i . Подстрочно дается значок, который указывает, для оценки какой величины построен индекс. Например, индивидуальный индекс цен обозначают i_p и рассчитывают как отношение цены конкретного товара в отчетном периоде (p_1) к цене этого товара в базисном периоде (p_0). Индивидуальные индексы цен показывают, как изменились цены на отдельные продукты, товары.

Любой индекс может быть выражен как коэффициент или в процентах. Индекс в виде коэффициента показывает, во сколько раз изменилось значение изучаемого признака в отчетном периоде по сравнению с базисным. Индекс больше 1 свидетельствует о росте признака, а меньше 1 — об уменьшении. Если индекс-коэффициент умножить на 100%, получим индекс, выраженный в процентах. Для того чтобы узнать, на сколько процентов вырос (уменьшился) изучаемый признак, из индекса, выраженного в процентах, необходимо отнять 100%.

Пример 14.1. Имеются данные о ценах, физическом объеме реализации* и выручке от продажи отдельных позиций ас-

* Под физическим объемом реализации понимают объем реализации в натуральных единицах измерения — штуках, килограммах, литрах, метрах и т.д.

ассортимента в магазине кондитерских изделий за два месяца. Для упрощения примера ограничим ассортимент тремя товарами (табл. 14.1).

Введены следующие обозначения для рассматриваемых признаков:

p — цена за единицу товара;

q — объем реализации товаров в натуральных единицах измерения (физический объем реализации);

$w = q \cdot p$ — выручка от реализации.

Требуется определить, как изменились цена, объем продаж и выручка по отдельным видам кондитерских изделий.

Динамика значений признаков по отдельным элементам изучаемой совокупности может быть оценена с помощью индивидуальных индексов цен, физического объема и выручки от реализации:

$$i_p = \frac{p_1}{p_0}, \quad i_q = \frac{q_1}{q_0}, \quad i_w = \frac{w_1}{w_0} = \frac{q_1 p_1}{q_0 p_0},$$

где p_1, q_1, w_1 — цена, физический объем продаж и выручка по отдельным видам товара в отчетном периоде;

p_0, q_0, w_0 — цена, физический объем продаж и выручка по отдельным видам товара в базисном периоде.

Индексы, выраженные в коэффициентах, образуют **систему взаимосвязанных индексов**, отражающую связь между рассматриваемыми признаками:

$$i_w = i_q i_p,$$

т.е. индекс выручки равен произведению индекса физического объема и индекса цен.

Рассчитаем индивидуальные индексы для нашего примера.

Индивидуальные индексы цен показывают, как изменились цены по каждой единице товара:

- пирожные $i_p^{\text{пир}} = \frac{p_1}{p_0} = \frac{15,5}{15,0} = 1,033$, или 103,3%;

- рулет $i_p^{\text{рул}} = \frac{p_1}{p_0} = \frac{88,0}{84,0} = 1,048$, или 104,8%;

Таблица 14.1

Данные о продажах кондитерских изделий

Кондитерское изделие	Единица измерения	Сентябрь			Октябрь		
		Цена за единицу товара, руб. (p_0)	Объем реализации, единиц (q_0)	Выручка от реализации, руб. (w_0)	Цена за единицу товара, руб. (p_1)	Объем реализации, ед. (q_1)	Выручка от реализации, руб. (w_1)
Пирожное «Корзиночка»	шт.	15,0	1520	22800	15,5	1450	22475
Рулет бисквитный	кг	84,0	180	15120	88,0	174	15312
Печенье миндальное	упаковка, 500 г	70,0	300	21000	70,0	310	21700

- печенье $i_p^{\text{печ}} = \frac{p_1}{p_0} = \frac{70,0}{70,0} = 1,0$, или 100,0%.

В октябре цены на пирожные выросли на 3,3% (103,3% – 100,0%), на рулет – на 4,8% (104,8% – 100,0%), цены на печенье не изменились.

Индивидуальные индексы объемов реализации характеризуют изменение объемов продаж в натуральном выражении по каждому виду кондитерских товаров:

- пирожные $i_q^{\text{пир}} = \frac{q_1}{q_0} = \frac{1450}{1520} = 0,954$, или 95,4%;

- рулет $i_q^{\text{рул}} = \frac{q_1}{q_0} = \frac{174}{180} = 0,967$, или 96,7%;

- печенье $i_q^{\text{печ}} = \frac{q_1}{q_0} = \frac{310}{300} = 1,033$, или 103,3%.

Количество проданных пирожных в октябре снизилось по сравнению с сентябрем на 4,6% (95,4% – 100,0%), объем продажи рулета сократился на 3,3% (96,7% – 100,0%), печенье продали больше на 3,3% (103,3% – 100,0%).

Индивидуальные индексы выручки отражают динамику средств, полученных от реализации отдельных видов товаров:

- пирожные $i_w^{\text{пир}} = \frac{w_1}{w_0} = \frac{22475}{22800} = 0,986$, или 98,6%;

- рулет $i_w^{\text{рул}} = \frac{w_1}{w_0} = \frac{15312}{15120} = 1,013$, или 101,3%;

- печенье $i_w^{\text{печ}} = \frac{w_1}{w_0} = \frac{21700}{21000} = 1,033$, или 103,3%.

В октябре по сравнению с сентябрем выручка от реализации пирожных снизилась на 1,4%, выручка от реализации рулета увеличилась на 1,3%, печенье – на 3,3%.

Индивидуальные индексы выручки можно было найти и через взаимосвязь индексов:

$$i_w = i_q i_p.$$

Для нашего примера:

$$i_w^{\text{пир}} = i_q^{\text{пир}} \cdot i_p^{\text{пир}} = 0,954 \cdot 1,033 = 0,986;$$

$$i_w^{\text{рул}} = i_q^{\text{рул}} \cdot i_p^{\text{рул}} = 0,967 \cdot 1,048 = 1,013;$$

$$i_w^{\text{печ}} = i_q^{\text{печ}} \cdot i_p^{\text{печ}} = 1,033 \cdot 1,0 = 1,033.$$

14.1.2. Сводные индексы в агрегатной форме

Для характеристики изменения обобщенных величин по всей совокупности строят **сводные (общие) индексы**. Обозначаются сводные индексы заглавной буквой *I*. Подстрочно указывается значок признака, изменение которого анализируется, например I_w — сводный индекс выручки. Сводные индексы могут быть построены в агрегатной форме или как средние из индивидуальных.

Агрегатные индексы считаются основной формой индекса. Они выполняют две функции: синтетическую и аналитическую. *Синтетическая* функция обеспечивается тем, что в одном индексе обобщаются (синтезируются) непосредственно несоизмеримые явления. Например, сводный индекс физического объема I_q должен характеризовать изменение объема реализации всех товаров. Непосредственно суммировать количество проданных единиц нельзя, поскольку товары имеют разные единицы измерения. Следовательно, данные по каждому товару должны браться с некоторым соизмерителем, который позволяет их агрегировать. Агрегатные индексы наряду с *индексируемым признаком* (признак, изменение которого изучается) содержат и *признак-вес*, который позволяет обобщить (соизмерить) разнородные элементы совокупности. Индексируемый признак при построении агрегатного индекса меняется — отчетный уровень сравнивается с базисным, а признак-вес берется на неизменном фиксированном уровне либо базисного периода (*по формуле Ласпейреса*), либо отчетного периода (*по формуле Пааше*). Отметим, что для одних и тех же данных величина индекса, рассчитанная по формуле Ласпейреса, как правило, превышает величину индекса по формуле Пааше (эффект Гершенкрона). *Индекс Фишера* пред-

ставляет среднюю оценку из этих двух индексов, за что и получил название «идеального» индекса.

Аналитическая функция индексов следует из взаимосвязи признаков. Например, если рассматривается система признаков $q \cdot p = w$ (физический объем \cdot цену = выручка), то индекс физического объема можно трактовать как индекс влияния физического объема продажи на сумму выручки от продажи, а индекс цен рассматривать как показатель влияния изменения цен на выручку. Это означает, что справедливы записи: $I_q = I_{w(q)}$ и $I_p = I_{w(p)}$.

Пример 14.2. По исходным данным примера 14.1 (см. табл. 14.1) рассчитайте сводные индексы цен, физического объема и выручки от реализации.

Вычисление агрегатов, необходимых для расчета сводных индексов, удобнее осуществлять в табличной форме. Построим расчетную таблицу (табл. 14.2).

Таблица 14.2

Расчетная таблица

Кондитерское изделие	Условная выручка от реализации, руб.	
	$q_1 \cdot p_0$	$q_0 \cdot p_1$
Пирожное «Корзиночка»	21750	23560
Рулет бисквитный	14616	15840
Печенье миндальное	21700	21000
Итого по всем товарам	58066	60400

Определим значения сводных индексов цен. Индекс, построенный по базисным весам — индекс Ласпейреса, равен:

$$I_p^L = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{15,5 \cdot 1520 + 88,0 \cdot 180 + 70,0 \cdot 300}{15,0 \cdot 1520 + 84,0 \cdot 180 + 70,0 \cdot 300} = \frac{60400}{58920} = 1,025,$$

или 102,5%.

Индекс по отчетным весам — индекс Пааше:

$$I_p^П = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{15,5 \cdot 1450 + 88,0 \cdot 174 + 70,0 \cdot 310}{15,0 \cdot 1450 + 84,0 \cdot 174 + 70,0 \cdot 310} = \frac{59487}{58066} = 1,024,$$

или 102,4%.

Индекс Фишера дает среднее геометрическое значение разновзвешенных индексов:

$$I_p^{\Phi} = \sqrt{I_p^{\text{Л}} \cdot I_p^{\text{П}}} = \sqrt{1,025 \cdot 1,024} = 1,0245, \text{ или } 102,45\%.$$

Таким образом, в октябре по сравнению с сентябрем наблюдался рост цен на кондитерские товары в среднем на 2,5% по формуле Ласпейреса, на 2,4% — по формуле Пааше и на 3,1% — по формуле Фишера.

Полученные результаты по сводным индексам цен можно интерпретировать и как изменение выручки за счет роста цен.

Например, индекс цен по схеме Ласпейреса ($I_p^{\text{Л}}$), т.е. построенный на весах базисного периода, показывает, что выручка от продажи кондитерских изделий базисного периода (q_0) с учетом изменения цен выросла на 2,5%.

Сводные индексы физического объема, рассчитанные по формулам Ласпейреса, Пааше, Фишера, составят:

$$I_q^{\text{Л}} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{58066}{58920} = 0,986, \text{ или } 98,6\%;$$

$$I_q^{\text{П}} = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{59487}{60400} = 0,985, \text{ или } 98,5\%;$$

$$I_q^{\Phi} = \sqrt{I_q^{\text{Л}} \cdot I_q^{\text{П}}} = \sqrt{0,986 \cdot 0,985} = 0,9855, \text{ или } 98,55\%.$$

Следовательно, в октябре по сравнению с сентябрем в целом наблюдалось снижение объема продаж кондитерских товаров в натуральном выражении: на 1,4% — по формуле Ласпейреса, на 1,5% — по формуле Пааше, на 1,45% — по формуле Фишера. Соответственно изменилась и выручка от реализации за счет сокращения физических объемов продаж.

Выручка от реализации является сопоставимым признаком, поэтому построение сводного индекса выручки не требует взвешивания. Сводный индекс товарооборота является *простым* и рассчитывается по формуле

$$I_w = \frac{\sum w_1}{\sum w_0}.$$

В отличие от аналитических индексов, которые выполняют две функции (синтетическую и аналитическую) и могут иметь двойную интерпретацию, простые индексы выполняют только одну функцию — они отвечают на вопрос, как изменилось изучаемое явление в отчетном периоде по сравнению с базисным.

Для нашего примера сводный индекс выручки составит:

$$I_w = \frac{\sum w_1}{\sum w_0} = \frac{59487}{58920} = 1,01, \text{ или } 101,0\%,$$

т.е. в октябре выручка от реализации кондитерских товаров была выше, чем в сентябре, на 1%.

Сводный индекс выручки может быть найден и через взаимосвязь индексов (*мультипликативная модель индексов*):

$$I_w = I_q^{\text{Л}} \cdot I_p^{\text{П}} \text{ или } I_w = I_q^{\text{П}} \cdot I_p^{\text{Л}}.$$

Для увязки индексов в систему веса в индексах первичного и вторичного признаков должны быть фиксированы на уровнях разных периодов. Отметим, что увязка взаимосвязанных индексов в систему выполняется лишь для сопоставимого круга элементов, т.е. при неизменном ассортименте реализации отдельных товаров в отчетном и базисном периодах.

Проверим правильность вычисления сводного индекса выручки для нашего примера через взаимосвязь индексов. Можно воспользоваться любой их двух моделей:

$$I_w = I_q^{\text{Л}} \cdot I_p^{\text{П}} = 0,986 \cdot 1,024 = 1,01, \text{ или } 101,0\%;$$

$$I_w = I_q^{\text{П}} \cdot I_p^{\text{Л}} = 0,985 \cdot 1,025 = 1,01, \text{ или } 101,0\%,$$

что подтверждает полученный ранее результат.

Участие каждого фактора в формировании общего изменения выручки в относительном выражении определяется по следующим формулам:

- прирост (уменьшение) выручки за счет изменения физического объема продаж

$$\% \Delta w_{(q)} = \frac{\Delta w_{(q)}}{w_0} = \frac{\sum q_1 p_0 - \sum q_0 p_0}{\sum q_0 p_0} = I_q - 1;$$

- прирост (уменьшение) выручки за счет изменения цен

$$\% \Delta w_{(p)} = \frac{\Delta w_{(p)}}{w_0} = \frac{\sum p_1 q_1 - \sum p_0 q_1}{\sum q_0 p_0} = I_w - I_q.$$

Совокупное влияние факторов в относительном выражении отражается следующей моделью:

$$\% \Delta w_{(q)} + \% \Delta w_{(p)} = \frac{\Delta w}{w_0} = \frac{\sum q_1 p_1 - \sum q_0 p_0}{\sum q_0 p_0} = I_w - 1.$$

Поскольку числитель и знаменатель сводных индексов имеют экономический смысл, в статистическом анализе нередко используют их разности. Так, например, разность числителя и знаменателя индекса выручки $\Delta w = \sum q_1 p_1 - \sum q_0 p_0$ характеризует абсолютный прирост (уменьшение) товарооборота в отчетном периоде по сравнению с базисным одновременно за счет:

- изменения физического объема продаж;
- изменения цен.

Измерить изолированное (элиминированное) влияние каждого из этих двух факторов можно через разность числителя и знаменателя соответствующих аналитических индексов.

Разность числителя и знаменателя индекса физического объема (по формуле Ласпейреса) показывает, как в абсолютном выражении изменилась выручка за счет роста (сокращения) физического объема продаж: $\Delta w_{(q)} = \sum q_1 p_0 - \sum q_0 p_0$.

Разность числителя и знаменателя индекса цен (по формуле Пааше) означает абсолютный прирост (уменьшение) выручки в результате роста (снижения) цен: $\Delta w_{(p)} = \sum p_1 q_1 - \sum p_0 q_1$.

Абсолютные изменения за счет отдельных факторов в сумме дают общее абсолютное изменение результативного признака:

$$\Delta w = \Delta w_{(q)} + \Delta w_{(p)}.$$

Эта же схема справедлива и для системы взаимосвязанных индексов, где индекс физического объема построен по отчетным весам (по формуле Пааше), а индекс цен — по базисным (по формуле Ласпейреса):

$$\Delta w_{(q)} = \sum q_1 p_1 - \sum q_0 p_1;$$

$$\Delta w_{(p)} = \sum p_1 q_0 - \sum p_0 q_0 ;$$

$$\Delta w = \Delta w_{(q)} + \Delta w_{(p)}.$$

Отметим, что в практическом использовании отдают предпочтение одному из вариантов построения взаимосвязанных агрегатных индексов. В отечественной статистике чаще придерживаются следующего правила выбора весов: индексы первичных признаков строятся на весах базисного периода, а индексы вторичных признаков – на весах отчетного периода. Следовательно, и разложение индексов в разностной форме чаще проводят по первым формулам.

Пример 14.3. По данным примера 14.2 определите, как в абсолютном выражении изменилась выручка от реализации кондитерских изделий в октябре по сравнению с сентябрем, в том числе абсолютное изменение выручки за счет изменения цен и за счет изменения физического объема продаж.

Абсолютное изменение выручки за счет изменения цен составило:

$$\Delta w_{(p)} = \sum p_1 q_1 - \sum p_0 q_1 = 59487 - 58066 = 1421 \text{ руб.}$$

Другими словами, за счет того, что произошел рост цен на кондитерские товары, дополнительно получено выручки на сумму 1421 руб.

Абсолютное изменение выручки за счет изменения физического объема продаж:

$$\Delta w_{(q)} = \sum q_1 p_0 - \sum q_0 p_0 = 58066 - 58920 = -854 \text{ руб.}$$

За счет сокращения физического объема продаж кондитерских изделий недополучено выручки на сумму 854 руб.

Совокупное изменение выручки за счет действия двух факторов можно найти двумя способами:

$$\Delta w = \sum q_1 p_1 - \sum q_0 p_0 = 58066 - 58920 = -854 \text{ руб.}$$

или

$$\Delta w = \Delta w_{(q)} + \Delta w_{(p)} = -854 + 1421 = 567 \text{ руб.}$$

При проведении статистического анализа может определяться также доля каждого фактора в формировании общего изменения результата:

- доля прироста (уменьшения) выручки за счет изменения физического объема продаж

$$d\Delta w_{(q)} = \frac{\Delta w_{(q)}}{\Delta w} = \frac{I_q - 1}{I_w - 1};$$

- доля прироста (уменьшения) выручки за счет изменения цен

$$d\Delta w_{(p)} = \frac{\Delta w_{(p)}}{\Delta w} = \frac{I_w - I_q}{I_w - 1}.$$

При этом $d\Delta w_{(q)} + d\Delta w_{(p)} = 1$, или 100%, если доли выражены в процентах.

Заметим, что оценка доли отдельных факторов в формировании результата проводится лишь в случае однонаправленного изменения признаков-факторов. Для нашего примера эти показатели не могут быть рассчитаны, поскольку факторы изменения выручки имели разную направленность: цены в среднем росли, а объем реализации снижался.

Если мультипликативная модель определяет результативный признак более чем двумя факторами, индексный анализ обычно проводят методом *цепных подстановок*. Признаки, стоящие слева от индексируемого признака, трактуются по отношению к нему как первичные и закрепляются на отчетном уровне, стоящие справа от индексируемого признака трактуются как вторичные и закрепляются на базисном уровне. Метод цепных подстановок требует соблюдения условия содержательной интерпретации при последовательном объединении признаков слева направо.

Пример 14.4. Известны данные по фирме, специализирующейся на пошиве рабочей одежды для медицинского персонала (табл. 14.3).

Таблица 14.3

Данные о производстве одежды для медицинского персонала

Вид изделия	Произведено изделий, тыс. шт.		Расход ткани на одно изделие, пог. м		Цена за 1 пог. м ткани, руб.	
	базисный период q_0	отчетный период q_1	базисный период l_0	отчетный период l_1	базисный период p_0	отчетный период p_1
Халат медицинский женский	4,0	4,2	2,2	2,1	120	125
Блузон медицинский мужской	1,5	1,4	1,6	1,5	120	133
Костюм операцион- ный, брючный	2,1	2,5	2,5	2,5	110	115

Определите, как изменились общие затраты на приобретение тканей, в том числе за счет изменения объемов производства, изменения расхода ткани вследствие использования новой схемы раскроя и за счет роста закупочных цен на ткань.

Затраты на ткань (C) определяются тремя факторами:

$$C = qmp,$$

где q — количество изделий;

m — расход ткани на одно изделие;

p — цена ткани.

Рассчитаем сводный индекс затрат на ткань:

$$\begin{aligned} I_c &= \frac{\sum q_1 m_1 p_1}{\sum q_0 m_0 p_0} = \frac{4,2 \cdot 2,1 \cdot 125 + 1,4 \cdot 1,5 \cdot 133 + 2,5 \cdot 2,5 \cdot 115}{4,0 \cdot 2,2 \cdot 120 + 1,5 \cdot 1,6 \cdot 120 + 2,1 \cdot 2,5 \cdot 110} = \\ &= \frac{2100,55}{1921,5} = 1,093, \text{ или } 109,3\%. \end{aligned}$$

В отчетном периоде затраты на ткань выросли на 9,3%, что в абсолютном выражении составило 179,05 тыс. руб. (2100,55 — 1921,5).

Изменение затрат вследствие изменения объемов производства характеризует индекс:

$$\begin{aligned} I_{c(q)} &= \frac{\sum q_1 m_0 p_0}{\sum q_0 m_0 p_0} = \frac{4,2 \cdot 2,2 \cdot 120 + 1,4 \cdot 1,6 \cdot 120 + 2,5 \cdot 2,5 \cdot 110}{4,0 \cdot 2,2 \cdot 120 + 1,5 \cdot 1,6 \cdot 120 + 2,1 \cdot 2,5 \cdot 110} = \\ &= \frac{2065,1}{1921,5} = 1,075, \text{ или } 107,5\%. \end{aligned}$$

Изменение объемов производства привело к увеличению затрат на ткань на 7,5%, или на 143,6 тыс. руб. (2065,1 — 1921,5).

Изменение затрат за счет использования новых схем раскроя покажет сводный индекс удельного расхода ткани:

$$\begin{aligned} I_{c(m)} &= \frac{\sum q_1 m_1 p_0}{\sum q_1 m_0 p_0} = \frac{4,2 \cdot 2,1 \cdot 120 + 1,4 \cdot 1,5 \cdot 120 + 2,5 \cdot 2,5 \cdot 110}{4,2 \cdot 2,2 \cdot 120 + 1,4 \cdot 1,6 \cdot 120 + 2,5 \cdot 2,5 \cdot 110} = \\ &= \frac{1997,9}{2065,1} = 0,967, \text{ или } 96,7\%. \end{aligned}$$

Уменьшение расхода ткани на одно изделие в среднем на 3,3% привело к соответствующему сокращению затрат на ткань. В абсолютном выражении экономия затрат на ткань за счет изменения удельных расходов составила 67,2 тыс. руб. (1997,9 – 2065,1).

Для оценки влияния изменения цен на затраты рассчитаем сводный индекс цен:

$$I_{c(p)} = \frac{\sum q_1 m_1 p_1}{\sum q_1 m_1 p_0} = \frac{2100,55}{1997,9} = 1,051, \text{ или } 105,1\%.$$

Увеличение закупочных цен на ткань в среднем на 5,1% привело к соответствующему росту затрат. Абсолютный перерасход за счет роста цен составил 102,65 тыс. руб. (2100,55 – 1997,9).

Все рассчитанные показатели увязываются в систему:

$$I_c = I_{c(q)} \cdot I_{c(m)} \cdot I_{c(p)};$$

$$1,093 = 1,075 \cdot 0,967 \cdot 1,051.$$

Между показателями абсолютных изменений также существует взаимосвязь:

$$\Delta_c = \Delta_{c(q)} + \Delta_{c(m)} + \Delta_{c(p)};$$

$$179,05 \text{ тыс. руб.} = 143,6 \text{ тыс. руб.} - 67,2 \text{ тыс. руб.} + 102,65 \text{ тыс. руб.}$$

14.1.3. Индексы средние из индивидуальных

Если информационная база не дает возможности провести индексный анализ в агрегатной форме, индексы могут быть построены в форме средних из индивидуальных.

Ниже приведены формулы некоторых средних индексов из индивидуальных.

Средний арифметический индекс физического объема

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{\sum i_q w_0}{\sum w_0} = \sum i_q d_{w_0},$$

где d_{w_0} — доля товарооборота отдельных видов продукции (товарных групп) в общем товарообороте базисного периода.

Средний гармонический индекс цен

$$I_p = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}} = \frac{\sum w_1}{\sum \frac{w_1}{i_p}} = \left(\sum \frac{d_{w_1}}{i_p} \right)^{-1},$$

где d_{w_1} — доля товарооборота отдельных видов продукции (товарных групп) в общем товарообороте отчетного периода.

Средний арифметический индекс цен

$$I_p = \frac{\sum i_p q_0 p_0}{\sum q_0 p_0} = \frac{\sum i_p w_0}{\sum w_0} = \sum i_p d_{w_0}.$$

Пример 14.5. Известны обобщенные данные о деятельности розничной торговли в Российской Федерации за два года (табл. 14.4).

Определите:

- как в целом изменился оборот розничной торговли в 2004 г. по сравнению с 2003 г.;
- как в среднем изменился физический объем розничного товарооборота;
- как в среднем изменились цены в розничной торговле;
- сумму перерасхода населением на покупку товаров за счет роста цен.

Динамику оборота розничной торговли отражает сводный индекс товарооборота:

$$I_w = \frac{\sum w_1}{\sum w_0} = \frac{5597,7}{4529,2} = 1,236, \text{ или } 123,6 \, \%.$$

Оборот розничной торговли в текущих ценах вырос на 23,6%.

Динамику физического объема розничного товарооборота можно оценить с помощью среднеарифметического индекса физического объема. Исходные данные позволяют рассчитать этот индекс двумя способами: используя в качестве веса а) абсолютные данные об обороте торговли по отдельным товарным группам и б) показатели структуры оборота розничной торговли:

Таблица 14.4

Деятельность розничной торговли в РФ

Группа товаров	Оборот розничной торговли				Индекс физического объема, 2004 г. к 2003 г. i_q
	2003		2004		
	трлн руб. w_0	% к итогу d_{w_0}	трлн руб. w_1	% к итогу d_{w_1}	
Продукты питания	1621,5	35,8	1987,1	35,5	110,4
Алкогольные напитки и пиво	471,0	10,4	569,3	10,2	110,4
Непродовольственные товары	2436,7	53,8	3041,0	54,3	114,4
Итого	4529,2	100,0	5597,7	100,0	—

$$\begin{aligned} \text{а) } I_q &= \frac{\sum i_q w_0}{\sum w_0} = \frac{1,104 \cdot 1621,5 + 1,104 \cdot 471,0 + 1,144 \cdot 2436,7}{1621,5 + 471,0 + 2436,7} = \\ &= \frac{5097,7}{4529,2} = 1,126, \text{ или } 112,6\%; \end{aligned}$$

$$\text{б) } I_q = \sum i_q d_{w_0} = 1,104 \cdot 0,358 + 1,104 \cdot 0,104 + 1,114 \cdot 0,538 = 1,126 \text{ или } 112,6\%.$$

Физический объем розничной торговли РФ в 2004 г. вырос по сравнению с 2003 г. в среднем на 12,6%.

Сводный индекс цен для розничной торговли в целом можно найти через взаимосвязь индексов:

$$I_P = I_w : I_q = 1,236 : 1,126 = 1,098, \text{ или } 109,8\%.$$

Цены в розничной торговле в 2004 г. по сравнению с предыдущим годом выросли в среднем на 9,8%.

Для того чтобы определить, какую сумму перерасхода имело население за счет роста цен, воспользуемся формулой

$$\begin{aligned} \Delta w_{(p)} &= \sum p_1 q_1 - \sum p_0 q_1 = \sum w_1 - \sum i_q w_0 = 5597,7 - 5097,7 = \\ &= 500 \text{ трлн руб.} \end{aligned}$$

За счет роста цен дополнительные расходы населения на покупку товаров в розничной торговле в 2004 г. составили 500 трлн руб.

14.1.4. Цепные и базисные индексы

На практике часто приходится анализировать изменение явления более чем за два периода. В этих случаях используют цепные и базисные индексы. **Цепные** индексы показывают изменения в текущем периоде по сравнению с предыдущим, **базисные** — изменения с начальным периодом анализируемого динамического ряда.

Цепные индивидуальные индексы:

$$i_{p_{1/0}} = \frac{p_1}{p_0}; \quad i_{p_{2/1}} = \frac{p_2}{p_1}; \quad i_{p_{3/2}} = \frac{p_3}{p_2}; \quad \dots \quad i_{p_{n/n-1}} = \frac{p_n}{p_{n-1}}.$$

Базисные индивидуальные индексы:

$$i_{p_{1/0}} = \frac{p_1}{p_0}; \quad i_{p_{2/0}} = \frac{p_2}{p_0}; \quad i_{p_{3/0}} = \frac{p_3}{p_0}; \quad \dots \quad i_{p_{n/0}} = \frac{p_n}{p_0}.$$

Между индивидуальными цепными и базисными индексами существует взаимосвязь — произведение цепных индексов равно последнему базисному индексу:

$$i_{p_{1/0}} \cdot i_{p_{2/1}} \cdot i_{p_{3/2}} \cdot \dots \cdot i_{p_{n/n-1}} = i_{p_{n/0}}.$$

Вычисление сводных индексов производится по нескольким схемам: с переменными и постоянными весами. Приведем варианты сводных индексов для анализа динамики цен.

Цепные индексы с переменными весами:

$$I_{p_{1/0}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_{p_{2/1}} = \frac{\sum p_2 q_2}{\sum p_1 q_2};$$

$$I_{p_{3/2}} = \frac{\sum p_3 q_3}{\sum p_2 q_3}; \quad \dots \quad I_{p_{n/n-1}} = \frac{\sum p_n q_n}{\sum p_{n-1} q_n}.$$

Цепные индексы с постоянными весами:

$$I_{p_{1/0}} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \quad I_{p_{2/1}} = \frac{\sum p_2 q_0}{\sum p_1 q_0};$$

$$I_{p_{3/2}} = \frac{\sum p_3 q_0}{\sum p_2 q_0}; \quad \dots \quad I_{p_{n/n-1}} = \frac{\sum p_n q_0}{\sum p_{n-1} q_0}.$$

Базисные индексы с переменными весами:

$$I_{p_{1/0}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; \quad I_{p_{2/0}} = \frac{\sum p_2 q_2}{\sum p_0 q_2};$$

$$I_{p_{3/0}} = \frac{\sum p_3 q_3}{\sum p_0 q_3}; \quad \dots \quad I_{p_{n/0}} = \frac{\sum p_n q_n}{\sum p_0 q_n}.$$

Базисные индексы с постоянными весами:

$$I_{p_{1/0}} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \quad I_{p_{2/0}} = \frac{\sum p_2 q_0}{\sum p_0 q_0};$$

$$I_{p_{3/0}} = \frac{\sum p_3 q_0}{\sum p_0 q_0}; \dots I_{p_{n/0}} = \frac{\sum p_n q_0}{\sum p_0 q_0}.$$

На практике выбор тех или иных индексов определяется характером решаемых задач. Так, для характеристики изменения цен в каждом периоде по сравнению с предыдущим без учета изменений в структуре проданных товаров применяют цепные индексы с постоянными весами, с учетом изменений в структуре — цепные индексы с переменными весами. Если необходимо оценить изменение цен по сравнению с начальным периодом без учета изменений в структуре проданных товаров, рассчитывают базисные индексы с постоянными весами, с учетом изменений в структуре — базисные индексы с переменными весами. Кроме того, рекомендуется придерживаться следующего правила: при построении индексов первичных признаков использовать постоянные веса, при построении индексов вторичных — переменные.

Между сводными цепными и базисными индексами также существует взаимосвязь: *произведение цепных индексов равно базисному*. Но выполняется это правило *только для индексов, построенных с постоянными весами*.

Пример 14.6. Имеются данные о реализации овощей на рынках города (табл. 14.5).

Таблица 14.5

Продажа овощей на рынках города

Овощная культура	Продано, тыс. т			Цена за 1 кг, руб.		
	Октябрь q_0	Ноябрь q_1	Декабрь q_2	Октябрь p_0	Ноябрь p_1	Декабрь p_2
Картофель	600	420	450	9	10	12
Лук	15	14	13	15	16	20

Вычислите сводные цепные и базисные индексы физического объема и цен.

Цепные индексы физического объема:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{420 \cdot 9 + 14 \cdot 15}{600 \cdot 9 + 15 \cdot 15} = \frac{3990}{5625} = 0,709, \text{ или } 70,9\%;$$

$$I_{q_{2/1}} = \frac{\sum q_2 p_0}{\sum q_1 p_0} = \frac{450 \cdot 9 + 13 \cdot 15}{420 \cdot 9 + 14 \cdot 15} = \frac{4245}{3990} = 1,064, \text{ или } 106,4\%.$$

В ноябре физический объем продажи овощей снизился по сравнению с октябрём на 29,1%, а в декабре по сравнению с ноябрём вырос на 6,4%.

Базисные индексы физического объема:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{3990}{5625} = 0,709, \text{ или } 70,9\%;$$

$$I_{q_{2/0}} = \frac{\sum q_2 p_0}{\sum q_0 p_0} = \frac{4245}{5625} = 0,755, \text{ или } 75,5\%.$$

В ноябре физический объем продажи овощей на рынках составил 70,9%, в декабре — 75,5% от продажи в октябре.

Поскольку индексы физического объема были построены по постоянным весам, между ними имеется взаимосвязь: произведение цепных индексов равно последнему базисному (незначительное несовпадение возникает за счет округления):

$$0,709 \cdot 1,064 = 0,754.$$

Цепные индексы цен:

$$I_{p_{1/0}} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{10 \cdot 420 + 16 \cdot 14}{9 \cdot 420 + 15 \cdot 14} = \frac{4424}{3990} = 1,109, \text{ или } 110,9\%;$$

$$I_{p_{2/1}} = \frac{\sum p_2 q_2}{\sum p_1 q_2} = \frac{12 \cdot 450 + 20 \cdot 13}{10 \cdot 450 + 16 \cdot 13} = \frac{5660}{4708} = 1,202, \text{ или } 120,2\%.$$

По сравнению с предыдущим месяцем цены на овощи в среднем выросли на 10,9 %, в декабре — на 20,2%.

Базисные индексы цен:

$$I_{p_{1/0}} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{4424}{3990} = 1,109, \text{ или } 110,9\%;$$

$$I_{p_{2/0}} = \frac{\sum p_2 q_2}{\sum p_0 q_2} = \frac{5660}{4245} = 1,333, \text{ или } 133,3\%.$$

По сравнению с октябрём цены на овощи в ноябре в среднем выросли на 10,9%, в декабре — на 33,3%.

Индексы цен построены с переменными весами. Для таких индексов взаимосвязь между цепными и базисными индексами отсутствует.

14.1.5. Индексный анализ взвешенной средней

Индексный метод применяется в статистике для изучения динамики средних величин и выявления факторов, влияющих на их динамику. Эти задачи решаются с помощью системы взаимосвязанных индексов переменного, постоянного состава и структурных сдвигов.

Индекс переменного состава представляет собой соотношение средних величин какого-либо признака в отчетном и базисном периодах:

$$I_{\text{перем}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}.$$

Как видно из формулы, индекс переменного состава характеризует изменение среднего уровня признака за счет влияния двух факторов:

1) изменения значений осредняемого признака (x) у отдельных единиц совокупности;

2) структурных изменений, под которыми понимается изменение доли отдельных единиц в общем объеме совокупности ($d = f / \sum f$).

Индекс постоянного (фиксированного) состава отражает изолированное действие первого фактора — среднюю величину изменения изучаемого признака у отдельных единиц совокупности и строится как отношение средних взвешенных величин с одними и теми же весами:

$$I_{\text{пост}} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}.$$

Индекс постоянного состава может быть рассчитан и в агрегатной форме:

$$I_{\text{пост}} = \frac{\sum x_1 f_1}{\sum x_0 f_1}.$$

Индекс структурных сдвигов характеризует влияние изменения структуры изучаемой совокупности на динамику среднего уровня признака:

$$I_{\text{стр}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}.$$

Индексы переменного, постоянного состава и структурных сдвигов увязываются в следующую систему:

$$I_{\text{перем}} = I_{\text{пост}} \cdot I_{\text{стр}}.$$

Если в индексах средних уровней в качестве весов используются удельные веса единиц совокупности в общей численности совокупности, т.е. показатели доли ($d = f/\Sigma f$), то система индексов может быть записана в следующем виде:

$$I_{\text{перем}} = \frac{\sum x_1 d_1}{\sum x_0 d_0}; \quad I_{\text{пост}} = \frac{\sum x_1 d_1}{\sum x_0 d_1}; \quad I_{\text{стр}} = \frac{\sum x_0 d_1}{\sum x_0 d_0}.$$

Может быть выполнено разностное разложение индексов переменного, постоянного состава и структурных сдвигов, отражающее абсолютное изменение среднего уровня вторичного признака за счет отдельных факторов. Общий абсолютный прирост (уменьшение) среднего уровня признака в целом по совокупности находится как разность числителя и знаменателя индекса переменного состава:

$$\Delta \bar{x} = \bar{x}_1 - \bar{x}_0 = \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} \quad \text{или} \quad \Delta \bar{x} = \sum x_1 d_1 - \sum x_0 d_0.$$

Абсолютный прирост (уменьшение) среднего уровня признака в целом по совокупности за счет изменения значений изучаемого признака у отдельных единиц совокупности и за счет структурных изменений рассчитывается соответственно как раз-

ность числителей и знаменателей индексов постоянного состава и структурных сдвигов:

$$\Delta \bar{x}_{(x)} = \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} \text{ или } \Delta \bar{x}_{(x)} = \sum x_1 d_1 - \sum x_0 d_1;$$

$$\Delta \bar{x}_{(d)} = \frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} \text{ или } \Delta \bar{x}_{(d)} = \sum x_0 d_1 - \sum x_0 d_0.$$

Абсолютные приросты связаны между собой:

$$\Delta \bar{x} = \Delta \bar{x}_{(x)} + \Delta \bar{x}_{(d)}.$$

Пример 14.7. Имеются данные о численности пенсионеров и размерах пенсии в Санкт-Петербурге за два года (табл. 14.6).

Таблица 14.6

Показатели пенсионного обеспечения в Санкт-Петербурге

Категория пенсионеров, получающих пенсии	Средний размер назначенной пенсии, руб.		Среднегодовая численность пенсионеров, тыс. чел.	
	2004 г. x_0	2005 г. x_1	2004 г. f_0	2005 г. f_1
По старости	2447,1	3043,4	985,1	1032,4
По инвалидности	1699,3	2017,2	172,6	136,1
По случаю потери кормильца	1081,9	1291,5	48,6	46,8
Итого	—	—	1206,3	1215,3

Определите:

- как изменились пенсии отдельных категорий пенсионеров;
- структуру численности пенсионеров за каждый год;
- как изменился средний размер пенсии, в том числе за счет изменения размера пенсий отдельных категорий и структурных изменений;
- изменение среднего размера пенсий в абсолютном выражении за счет: а) изменения размера пенсий отдельных категорий; б) структурных изменений; в) совокупного действия двух факторов.

Решение.

Оценку изменения пенсий отдельных категорий пенсионеров осуществим с помощью индивидуальных индексов:

- пенсии по старости $i_x = \frac{3043,4}{2447,1} = 1,244$, или 124,4%;
- пенсии по инвалидности $i_x = \frac{2017,2}{1699,3} = 1,187$, или 118,7%;
- пенсии по случаю потери кормильца $i_x = \frac{1291,5}{1081,9} = 1,194$, или 119,4%.

Итак, больше всего увеличились пенсии по старости — за год они выросли на 24,4%, в меньшей степени увеличились пенсии по случаю потери кормильца — на 19,4% и пенсии по инвалидности — на 18,7 %.

Рассчитаем удельный вес пенсионеров отдельных категорий по формуле $d = f/\Sigma f$, где f — численность пенсионеров той или иной категории; Σf — общая численность пенсионеров. Все промежуточные вычисления оформим в виде расчетной таблицы (табл. 14.7).

Таблица 14.7

Расчетная таблица

Категория пенсионеров, получающих пенсии	Изменение среднего размера пенсий i_x	Удельный вес пенсионеров, %		$x_0 d_0$	$x_1 d_1$	$x_0 d_1$
		2004 г. d_0	2005 г. d_1			
По старости	1,244	81,7	85,0	1999,3	2586,9	2080,0
По инвалидности	1,187	14,3	11,2	243,0	225,9	190,3
По случаю потери кормильца	1,194	4,0	3,8	43,3	49,1	41,1
Итого	—	100,0	100,0	2285,6	2861,9	2311,4

Как видно из вычислений, структура лиц, получающих пенсии, за год несколько изменилась. Увеличилась доля пенсионеров, получающих пенсии по старости, с 81,7 до 85,0%, или на 3,3 процентных пункта, уменьшились доли пенсионеров, полу-

чающих пенсии по инвалидности, на 3,1 процентных пункта (14,3% — 11,2%) и по случаю потери кормильца — на 0,2 процентных пункта (4,0% — 3,8%).

Изменение среднего размера пенсий характеризует индекс переменного состава. Он может быть рассчитан по двум формулам:

$$I_{\text{перем}} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}$$

или

$$I_{\text{перем}} = \frac{\sum x_1 d_1}{\sum x_0 d_0}.$$

Воспользуемся второй формулой. Все промежуточные расчеты приведены в табл. 14.7.

$$I_{\text{перем}} = \frac{\sum x_1 d_1}{\sum x_0 d_0} = \frac{2861,9}{2285,6} = 1,252, \text{ или } 125,2\%.$$

Средний размер пенсии в Санкт-Петербурге за год вырос с 2285,6 до 2861,9 руб., или на 25,2%, что больше, чем рост пенсий отдельных категорий пенсионеров (см. индивидуальные индексы пенсий). Это произошло потому, что на изменение среднего размера пенсий оказало влияние не только изменение размера пенсий отдельных категорий, но и структурные изменения в распределении пенсионеров по отдельным категориям. Для оценки изменения раздельного влияния каждого из факторов рассчитаем индексы постоянного состава и структурных сдвигов:

$$I_{\text{пост}} = \frac{\sum x_1 d_1}{\sum x_0 d_1} = \frac{2861,9}{2311,4} = 1,238, \text{ или } 123,8\%;$$

$$I_{\text{стр}} = \frac{\sum x_0 d_1}{\sum x_0 d_0} = \frac{2311,4}{2285,6} = 1,011, \text{ или } 101,1\%.$$

Таким образом, наиболее существенное влияние на изменение среднего размера пенсий оказал первый фактор. За счет роста пенсий отдельных категорий средний размер пенсий увеличился на 23,8%. Кроме того, структурные изменения, а именно рост

доли пенсионеров по старости, получающих более высокие пенсии, и уменьшение доли пенсионеров, у которых пенсии более низкие (по инвалидности и по случаю потери кормильца), обеспечили дополнительный рост средней пенсии на 1,1%.

Проверим увязку индексов в систему:

$$I_{\text{перем}} = I_{\text{пост}} \cdot I_{\text{стр}};$$

$$1,252 = 1,238 \cdot 1,011.$$

Изменение среднего размера пенсий в абсолютном выражении за счет изменения размера пенсий отдельных категорий составило:

$$\Delta \bar{x}_{(x)} = \sum x_1 d_1 - \sum x_0 d_1 = 2861,9 - 2311,4 = 550,5 \text{ руб.}$$

Изменение среднего размера пенсий в абсолютном выражении за счет структурных изменений:

$$\Delta \bar{x}_{(d)} = \sum x_0 d_1 - \sum x_0 d_0 = 2311,4 - 2285,6 = 23,8 \text{ руб.}$$

Изменение среднего размера пенсий за счет совокупного действия двух факторов составило:

$$\Delta \bar{x} = \Delta \bar{x}_{(x)} + \Delta \bar{x}_{(d)} = 550,5 + 23,8 = 574,3 \text{ руб.}$$

Пример 14.8. В табл. 14.8 приведены данные об экспорте Российской Федерацией сырой нефти.

Таблица 14.8

Экспорт сырой нефти РФ

Направление экспорта	Количество, млн т		Стоимость экспорта, млн долл. США	
	2003 г. q_0	2004 г. q_1	2003 г. w_0	2004 г. w_1
Страны СНГ	37,0	40,1	4944	7872
Страны вне СНГ	186,0	218,0	33872	50449
Всего	223,0	258,1	38816	58321

Определите:

- экспортные цены на нефть по отдельным направлениям экспорта и средние экспортные цены для каждого года;
- изменение средних экспортных цен;

• как за рассматриваемый период изменилась выручка от экспорта нефти в целом и в том числе за счет изменения отдельных факторов: объемов экспорта; экспортных цен на нефть; структуры экспорта.

Экспортные цены определяются как отношение соответствующей стоимости экспорта к физическому объему экспорта. Результаты расчета даны в табл. 14.9.

Таблица 14.9

Средние экспортные цены

Направление экспорта	2003 г.	2004 г.	2004 г., % к 2003 г.
	долл. за 1 т p_0	долл. за 1 т p_1	
По странам СНГ	4944:37,0=133,6	7872:40,1=196,3	146,9
По странам вне СНГ	33872:186,0=182,1	50449:218,0=231,4	127,1

Динамику средних экспортных цен на нефть отражает индекс переменного состава:

$$I_{\text{перем}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\sum u_1}{\sum q_1} : \frac{\sum u_0}{\sum q_0} = \frac{58321}{258,1} : \frac{38872}{223,0} = 226,0 : 174,1 = 1,298, \text{ или } 129,8\%.$$

В 2004 г. средние экспортные цены на нефть выросли по сравнению с 2003 г. на 29,8%.

Изменения экспортной выручки за счет изменения объема экспорта, экспортных цен и структуры экспорта составили соответственно:

$$\Delta w(q) = (\sum q_1 - \sum q_0) \cdot \bar{p}_0 = (258,1 - 223,0) \cdot 174,1 = 6110,9 \text{ млн долл.};$$

$$\Delta w(p) = \sum p_1 q_1 - \sum p_0 q_1 = 58321 - (133,6 \cdot 40,1 + 182,1 \cdot 218,0) = 13265,84 \text{ млн долл.};$$

$$\Delta w\left(\frac{q}{\sum q}\right) = \left(\frac{\sum p_0 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0}\right) \sum q_1 = \left(\frac{133,6 \cdot 40,1 + 182,1 \cdot 218,0}{258,1} - 174,1\right) \cdot 258,1 = 129,05 \text{ млн долл.}$$

Итак, как следует из расчетов, в 2004 г. по сравнению с 2003 г. большая часть прироста экспортной выручки — 13265,84 млн долл. — была получена в результате роста экспортных цен по всем направлениям экспорта. Рост физического объема экспорта увеличил выручку на 6110,91 млн долл. Незначительные изменения в структуре экспорта — увеличение доли экспорта в страны дальнего зарубежья с 83,4% ($186,0/223,0 \cdot 100\%$) до 84,4% ($218,0/258,1 \cdot 100\%$) и соответствующее снижение доли экспорта в страны СНГ с 16,6 до 15,5% — обеспечило прирост выручки на 132,1 млн долл. Общее изменение выручки от экспорта нефти составило:

$$\Delta w = \Delta w(q) + \Delta w(p) + \Delta w\left(\frac{q}{\sum q}\right) = 6110,91 + 13265,84 + 129,05 = \\ = 19505 \text{ млн долл.}$$

Проверка:

$\Delta w = \Sigma w_1 - \Sigma w_0 = 58321 - 38816 = 19505$ млн долл. (незначительное несовпадение результатов связано с округлениями в расчетах).

14.1.6. Взаимосвязи индексов

Взаимосвязи индексов существуют в нескольких аспектах:

- взаимосвязи индексов конкретных признаков;
- взаимосвязи индивидуальных и сводных индексов;
- взаимосвязи цепных и базисных индексов;
- взаимосвязи индексов переменного состава, постоянного состава и структурных сдвигов.

Знание взаимосвязей индексов позволяет решать различные аналитические задачи.

Пример 14.9. Фонд оплаты труда в учреждении за год вырос на 15%. За этот же период численность сотрудников сократилась на 7%. Как изменилась средняя заработная плата в учреждении?

Поскольку между индексами существует та же взаимосвязь, что и между изучаемыми признаками, индекс средней заработной платы может быть найден как отношение индекса фонда оплаты труда к индексу численности сотрудников:

$$1,15 : 0,93 = 1,198, \text{ или } 119,8\%.$$

Средняя заработная плата в учреждении за год выросла на 19,8%.

Пример 14.10. Известно, что цены за квадратный метр жилья на вторичном рынке недвижимости росли следующим образом: в январе — на 5,1%, феврале — на 4,3%, марте — на 3,2% по сравнению с предыдущим месяцем. Как изменились цены на недвижимость за I квартал?

Воспользуемся правилом взаимосвязи цепных и базисных индексов:

$$1,051 \cdot 1,043 \cdot 1,032 = 1,131, \text{ или } 113,1\%,$$

т.е. за квартал цены выросли на 13,1%.

Пример 14.11. Известны данные о динамике объема выпуска продукции промышленного предприятия за полугодие (табл. 14.10).

Таблица 14.10

Динамика выпуска продукции

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь
Объем выпуска, % к январю	100,0	105,3	114,2	108,1	105,2	108,9

Определите, как изменился объем выпуска продукции за последний месяц полугодия?

Решение.

Из правила взаимосвязи цепных и базисных индексов следует, что отношение базисного индекса последнего периода к базисному индексу предшествующего периода равно цепному индексу последнего периода. Следовательно, для ответа на поставленный вопрос необходимо базисный индекс за июнь разделить на базисный индекс за май:

$$1,089 : 1,052 = 1,035, \text{ или } 103,5\%.$$

В июне выпуск продукции вырос по сравнению с маем на 3,5%.

Пример 14.12. В отчетном периоде бензин подорожал на 8%, а дизельное топливо — на 4%. Как в среднем выросли цены на автомобильное топливо, если известно, что в структуре выручки от реализации топлива бензин занимает 65%?

Если на бензин приходится 65% выручки, следовательно, на дизельное топливо — 35% (100% — 65%). Зная структуру выручки, можно рассчитать сводный индекс цен как средневзвешенный из индивидуальных:

$$1,08 \cdot 0,65 + 1,04 \cdot 0,35 = 1,066, \text{ или } 106,6\%.$$

Таким образом, за отчетный период топливо в среднем подорожало на 6,6%.

Пример 14.13. В салоне мобильной связи средняя цена покупки одного мобильного телефона выросла на 10%, при этом модели отдельных телефонов подешевели в среднем на 3%. Что можно сказать об изменении структуры продаж?

Изменение средней цены покупки было рассчитано на основе индекса переменного состава, а среднее изменение цен характеризует индекс постоянного состава. Зная эти два индекса, можно определить индекс структурных сдвигов:

$$I_{\text{стр}} = I_{\text{пер}} : I_{\text{пост}} = 1,1 : 0,97 = 1,134, \text{ или } 113,4\%.$$

Значение индекса структурных сдвигов больше 1 (или больше 100%) свидетельствует об изменении структуры продаж в сторону увеличения доли более дорогих моделей телефонов.

Пример 14.14. Имеются данные (табл. 14.11) о динамике промышленного производства и среднесписочной численности работников организаций обрабатывающих производств на территории Ленинградской области.

Таблица 14.11

**Показатели деятельности обрабатывающих производств
в Ленинградской области, % к предыдущему месяцу**

Показатель	2003 г.	2004 г.	2005 г.
Индекс промышленного производства	113,0	124,1	107,5
Среднесписочная численность работников	101,3	99,6	102,2

Необходимо определить:

- как изменялась производительность труда в обрабатывающем секторе промышленности региона в отдельные годы, а также в целом за рассматриваемый период;

- доли прироста продукции обрабатывающего сектора, полученные за счет роста производительности труда и за счет увеличения численности работников в 2005 г.

Решение. Как известно, показатель производительности труда представляет собой отношение показателей объема продукции и трудозатрат, т.е. производительность труда на уровне обрабатывающих организаций региона (ν) можно определить так:

$$\nu = \frac{Q}{T},$$

где Q — объем производства в обрабатывающем секторе;
 T — среднесписочная численность работников.

Тогда

$$I_V = \frac{I_Q}{I_T}.$$

Найдем индексы производительности труда для каждого года:

$$I_{V03/02} = \frac{1,130}{1,013} = 1,115, \text{ или } 115,5\%;$$

$$I_{V04/03} = \frac{1,241}{0,996} = 1,246, \text{ или } 124,6\%;$$

$$I_{V05/04} = \frac{1,075}{1,022} = 1,052, \text{ или } 105,2\%.$$

Производительность труда в обрабатывающем секторе промышленности Ленинградской области в 2003 г. по сравнению с 2002 г. выросла на 15,5%, в 2004 г. по сравнению с 2003 г. — на 24,6%, в 2005 г. по сравнению с 2004 г. — на 5,2%.

Базисный индекс производительности труда найдем через произведение цепных индексов:

$$\begin{aligned} I_{V05/02} &= I_{V03/02} I_{V04/03} I_{V05/04} = 1,115 \cdot 1,246 \cdot 1,052 = \\ &= 1,462, \text{ или } 146,2\%. \end{aligned}$$

В 2005 г. по сравнению с 2002 г. выработка продукции на одного работающего в обрабатывающем секторе промышленности была выше на 46,2%.

Определим долю каждого фактора в формировании общего прироста промышленной продукции для 2005 г.:

- доля прироста продукции за счет роста производительности труда

$$d\Delta Q_{(V)} = \frac{\Delta Q_{(V)}}{\Delta Q} = \frac{I_Q - I_T}{I_Q - 1} = \frac{1,075 - 1,022}{1,075 - 1} = 0,707, \text{ или } 70,7\%;$$

- доля прироста продукции за счет увеличения числа работающих

$$d\Delta Q_{(T)} = \frac{\Delta Q_{(T)}}{\Delta Q} = \frac{I_T - 1}{I_Q - 1} = \frac{1,022 - 1}{1,075 - 1} = 0,293, \text{ или } 29,3\%.$$

Поскольку

$$d\Delta Q_{(V)} + d\Delta Q_{(T)} = 1,$$

то долю второго фактора можно было определить как $1 - 0,707 = 0,293$.

Таким образом, в 2005 г. за счет роста выработки на одного занятого получено 70,7% прироста промышленной продукции в обрабатывающем секторе региона, за счет увеличения числа работающих – 29,3% прироста.

14.2. Задания по теме

1. Имеются данные о реализации пиломатериалов деревообрабатывающим комбинатом (табл. 14.12) по кварталам.

Таблица 14.12

Продукция	Единица измерения	Квартал			
		I	II	III	IV
Отпускная цена, руб. за ед.					
Вагонка «евро»	м ²	140	135	150	165
Доска обрезная	м ³	3000	3300	3400	3500
Бревно оцилиндрованное с пазом	шт.	1100	1150	1200	1200

Продукция	Единица измерения	Квартал			
		I	II	III	IV
Отпущено продукции, ед.					
Вагонка «евро»	тыс.м ²	3,1	4,2	8,5	2,7
Доска обрезная	м ³	300	420	750	450
Бревно оцилиндрованное с пазом	шт.	392	820	1200	350

Постройте мультипликативную модель индексов, отражающую взаимосвязь стоимости реализованной продукции, физического объема реализации и цен. Выполните аддитивное разложение изменения стоимости реализованной продукции. На основе полученных результатов ответьте на вопрос: как изменилась стоимость реализованной продукции в целом и в том числе за счет изменения физических объемов реализации и за счет изменения цен (в относительном и абсолютном выражении)?

Анализ проводите для двух смежных периодов:

1-й вариант: I и II кварталы;

2-й вариант: II и III кварталы;

3-й вариант: III и IV кварталы;

4-й вариант: первое и второе полугодие.

2. На основе данных задания 1 рассчитайте на цепной и базисной основе:

1-й вариант: индивидуальные и общие индексы цен;

2-й вариант: индивидуальные и общие индексы физического объема реализации;

3-й вариант: индивидуальные и общие индексы стоимости реализованной продукции.

Покажите взаимосвязь цепных и базисных индексов.

3. ЗАО «Керамика» осуществляет производство и реализацию строительного кирпича разных марок. В табл. 14.13 приведены данные, характеризующие изменения некоторых показателей реализации продукции в 2006 г. по сравнению с 2005 г.

Таблица 14.13

Марка кирпича	Выручка от реализации			Прирост физического объема реализации, 2006 г. по сравнению с 2005 г., %	Отпускная цена в 2006 г., руб. за 1 тыс. шт.
	2006 г. к 2005 г., %	к общей выручке от реализации, %			
		2006 г.	2005 г.		
М-100	105	50	45	+1	7900
М-125	105	25	25	+5	8150
М-150	120	25	30	+15	8500

Определите:

- а) как в среднем подорожал кирпич, выпускаемый ЗАО «Керамика» и в том числе как изменились цены на отдельные марки кирпича;
- б) как в целом изменился физический объем реализации кирпича;
- в) на сколько процентов выросла выручка предприятия.

4. Имеются данные о реализации отделочных материалов складом-магазином за год (табл. 14.14).

Таблица 14.14

Отделочный материал	Единица измерения	Продано, тыс. ед.		Цена за ед., руб.	
		1-е полугодие	2-е полугодие	1-е полугодие	2-е полугодие
Плитка кафельная	м ²	168	210	300	350
Бордюр керамический	шт.	216	228	24	28
Плитка керамическая	шт.	1680	1500	50	55
Заделка для кафельной плитки	м	1920	1200	16	20

1. Постройте системы индивидуальных и сводных индексов, характеризующих изменение цен и объемов реализации (в натуральном и стоимостном выражении) по каждому наименованию реализованных материалов и в целом по данной товарной группе.

2. Определите, на сколько в абсолютном выражении увеличилась выручка от реализации отделочных материалов во втором полугодии:

- а) за счет роста цен;
- б) за счет изменения объемов реализации;
- в) в целом.

5. Имеются данные о реализации картофеля на трех рынках города.

Таблица 14.15

Рынок	Сентябрь		Октябрь	
	Цена за 1 кг, руб.	Продано, т	Цена за 1 кг, руб.	Выручка от продажи, тыс. руб.
1	10,0	50	11,0	517
2	9,5	45	10,5	420
3	8,5	60	9,5	475

1. Рассчитайте среднюю цену реализации картофеля в сентябре и октябре.

2. С помощью системы индексов переменного, постоянного состава и структурных сдвигов определите изменение средней цены на картофель, в том числе за счет роста цен на отдельных рынках и за счет изменения структуры реализации.

3. Определите изменение средней цены в абсолютном выражении за счет отдельных факторов:

- а) изменения собственно цен на рынках города;
- б) изменения в структуре реализации.

6. В табл. 14.16 приведены данные об изменениях численности занятых и заработной платы в трех основных отраслях экономики Санкт-Петербурга.

Таблица 14.16

Отрасль экономики	Численность занятых, тыс. чел.		Среднемесячная начисленная заработная плата в расчете на одного работника в 2004 г., руб.	Изменение реальной начисленной заработной платы, 2004 г., % к 1995 г.
	1995 г.	2004 г.		
Промышленность	596,2	483,6	8678	229,6
Строительство	251,4	271,5	7383	180,9
Оптовая и розничная торговля, общественное питание	348,0	483,6	4407	162,6

1. Рассчитайте среднюю заработную плату для трех основных отраслей экономики в 2004 г.

2. Как изменилась средняя реальная заработная плата в трех отраслях экономики за рассматриваемый период?

3. Какое влияние на изменение средней заработной платы оказало изменение оплаты труда в отдельных отраслях?

4. Охарактеризуйте изменения численности и структуры занятых за рассматриваемый период. Как эти изменения отразились на изменении средней реальной заработной платы?

5. Напишите аналитическую записку.

7. Имеются данные о производстве продукции **A** по двум предприятиям отрасли (табл. 14.17).

Таблица 14.17

Предприятие	1-е полугодие		2-е полугодие	
	Производство, тыс. шт.	Затраты на производство, млн руб.	Производство, тыс.шт.	Себестоимость 1 ед., руб.
№ 1	65	45,5	55	750
№ 2	35	28,0	45	850

1. Определите:

а) себестоимость продукции **A** на каждом предприятии в первом полугодии;

б) среднеотраслевую себестоимость продукции **A** за каждое полугодие;

в) изменение среднеотраслевой себестоимости и себестоимости на каждом предприятии.

2. Объясните, почему среднеотраслевая себестоимость увеличилась больше, чем себестоимость на каждом предприятии?

3. Как изменились затраты на производство продукции **A** в целом по отрасли (в абсолютном и относительном выражении), в том числе:

а) за счет сокращения физических объемов производства;

б) за счет роста себестоимости производимой продукции.

8. Производство продукции **A** осуществляется на двух предприятиях отрасли (табл. 14.18).

Таблица 14.18

Предприятие	1-е полугодие		2-е полугодие	
	Производство, тыс. шт.	Себестоимость 1 ед., руб.	Производство, тыс. шт.	Себестоимость 1 ед., руб.
1	40	700	50	710
2	60	800	50	810

1. Определите:

а) среднеотраслевую себестоимость продукции **А** за каждое полугодие;

б) как изменилась себестоимость производства продукции на каждом предприятии и в среднем по отрасли.

2. Почему при росте себестоимости на каждом предприятии, среднеотраслевая себестоимость осталась неизменной?

9. В табл. 14.19 приведены данные, характеризующие выращивание масличных культур в Российской Федерации за два года.

Таблица 14.19

Масличная культура	Валовой сбор, млн ц		Посевная площадь, тыс. га	
	2003 г.	2004 г.	2003 г.	2004 г.
Подсолнечник	49	48	5337	4848
Соя	3,93	5,55	586	571

1. Определите урожайность масличных культур в 2003 и 2004 гг.

2. С помощью системы сводных индексов охарактеризуйте изменение валового сбора масличных культур и в том числе изменения, вызванные:

а) сокращением посевных площадей;

б) ростом урожайности масличных культур.

3. Определите, как в абсолютном выражении изменился валовой сбор масличных культур в целом и за счет каждого из факторов.

10. Имеются данные о выращивании пшеницы в Российской Федерации (табл. 14.20).

Таблица 14.20

Пшеница	Валовой сбор, млн ц		Посевная площадь, тыс. га	
	2003 г.	2004 г.	2003 г.	2004 г.
Озимая	147	259	7412	8977
Яровая	194	195	14774	15052
Всего	341	454	22186	24029

1. Определите урожайность озимой и яровой пшеницы для каждого года.

2. Постройте сводные индексы, отражающие изменение валового сбора, площади посевных площадей и урожайности пшеницы.

3. Определите, как в абсолютном выражении изменился валовой сбор пшеницы в 2004 г. по сравнению с 2003 г., в том числе изменения, вызванные:

- а) ростом посевных площадей;
- б) ростом урожайности.

4. Найдите долю прироста валового сбора за счет каждого из факторов.

11. Используя данные табл. 14.1, определите, как в абсолютном и относительном выражении изменится выручка кондитерской в ноябре по сравнению с октябрём, если известно, что цены останутся неизменными, а количество реализации вырастет следующим образом: пирожных — на 2%, рулета — на 1,5%, печенья — на 1,3%.

12. Имеются данные о розничном товарообороте в регионе (табл. 14.21).

Таблица 14.21

Группа товаров	Розничный товарооборот, млн руб.		Изменение цен во 2-м полугодии, %
	1-е полугодие	2-е полугодие	
Продовольственные	21228	23596	+10,3
Непродовольственные	30984	35671	+7,6

1. Определите структуру розничного товарооборота в отчетном и базисном периодах.

2. Рассчитайте сводный индекс цен розничной торговли, используя:

а) абсолютные данные об объеме товарооборота;

б) показатели структуры товарооборота.

3. Вычислите индексы стоимостного объема товарооборота и физического объема товарооборота.

4. Как изменился физический объем товарооборота на душу населения в регионе, если численность населения за рассматриваемый период сократилась на 0,5%.

13. В табл. 14.22 представлены данные об объеме перевозок грузов в регионе по отдельным видам транспорта за год.

Таблица 14.22

Вид транспорта	Перевезено грузов		Индекс тарифов на грузовые перевозки, % к предыдущему году
	млн т	% к предыдущему году	
Железнодорожный	75,1	100,8	116,0
Автомобильный	10,5	106,5	105,9
Морской	2,1	81,7	106,5
Внутренний водный	5,8	101,3	102,3
Воздушный	0,4	83,6	113,0
Трубопроводный	63,1	107,3	101,3

Рассчитайте сводный индекс тарифов на грузовые перевозки по формулам Ласпейреса, Пааше, Фишера.

14. По материалам выборочных обследований домашних хозяйств структура потребительских расходов в регионе за два последних года изменилась (табл. 14.23).

Таблица 14.23

Направление расходов	% от общей суммы расходов	
	базисный год	отчетный год
Продовольственные товары	46,5	38,7
Непродовольственные товары	39,6	42,6
Услуги	13,9	18,7

При этом известно, что в регионе за рассматриваемый период цены на продовольственные товары выросли на 10,5%, на непродовольственные товары – на 7,54%, на платные услуги – на 15,3%.

Вычислите сводный индекс потребительских цен в регионе всеми возможными способами.

15. В табл. 14.24 представлены данные, характеризующие функционирование отдельных отраслей промышленности в регионе за два года, %.

Таблица 14.24

Отрасль промышленности	Доля в общем объеме производства		Индексы физического объема производства	Рост цен
	Базисный год	Отчетный год		
Электроэнергетика	11,5	12,2	104,7	18,2
Металлургия	6,7	4,2	123,7	4,6
Химическая и нефтехимическая	3,2	1,3	96,3	0,5
Машиностроение и металлообработка	29,3	27,6	113,6	5,2
Лесная, деревообрабатывающая и целлюлозно-бумажная	4,5	2,4	107,3	2,2
Промышленность строительных материалов	4,7	5,9	114,4	12,3
Легкая	6,9	6,0	91,3	1,3
Пищевая	30,1	38,4	107,5	6,5
Мукомольно-крупяная и комбикормовая	3,1	2,0	92,9	0,9

Определите сводный индекс промышленного производства (индекс физического объема) и сводный индекс цен производителей промышленной продукции.

16. Распределение работников машиностроительного предприятия по цехам характеризуется следующими данными: цех 1 – 160 чел., цех 2 – 200 чел., цех 3 – 75 чел. За отчетный период производительность труда в 1-м и 3-м цехах выросла соответственно на 3 и 8%, во 2-м цехе сократилась на 3%. Рассчитайте индекс производительности труда для предприятия в целом.

17. Имеются данные о численности работников и оплате труда на предприятии (табл. 14.25).

Таблица 14.25

Категория работников	Среднесписочная численность работников, чел.		Фонд оплаты труда, тыс. руб.	
	Базисный год	Отчетный год	Базисный год	Отчетный год
Рабочие	310	280	42780	48720
Руководители	35	35	12600	13860
Специалисты	50	80	13200	22080
Прочие служащие	10	5	1176	3600
В целом по предприятию	405	400	69756	88260

1. Рассчитайте среднемесячную заработную плату для каждой категории работников и по предприятию в целом.

2. Постройте систему индексов, отражающую изменение средней заработной платы на предприятии за счет отдельных факторов:

- а) роста оплаты труда отдельных категорий работников;
- б) изменения структуры численности работников предприятия.

3. Постройте мультипликативную и аддитивную модели, характеризующие изменение фонда оплаты труда на предприятии за счет:

- а) изменения заработной платы отдельных категорий работников;
- б) изменения численности работников предприятия;
- в) изменения структуры численности работников предприятия.

18. В табл. 14.26 приведены данные, характеризующие строительство жилья жилищно-строительными кооперативами и индивидуальными застройщиками в Санкт-Петербурге за два года.

Таблица 14.26

Застройщики	Построено квартир		Средний размер квартиры	
	1995 г.	2004 г.	1995 г.	2004 г.
Жилищно-строительные кооперативы	2574	246	61,7	91,8
Индивидуальные застройщики	128	344	205,4	229,9

1. Рассчитайте:

- а) средний размер построенной квартиры для каждого года;
- б) как изменились средние размеры квартир, построенные жилищно-строительными кооперативами и индивидуальными застройщиками?
- в) как изменился средний размер построенных квартир по городу в целом?
- г) индексы среднего размера квартир постоянного состава и структурных сдвигов.

2. По результатам вычислений напишите аналитическую записку.

19. Имеются данные об импорте табачного сырья в РФ (табл. 14.27).

Таблица 14.27

Направление импорта	Количество, тыс. т		Стоимость импорта, млн долл.	
	2000 г.	2004 г.	2000 г.	2004 г.
Из стран СНГ	85,7	26,2	115	29,8
Из стран вне СНГ	191	247	403	576
Итого	276,7	273,2	518	605,8

1. Определите:

- а) средние импортные цены на табачное сырье по отдельным направлениям импорта;
- б) как изменились цены за рассматриваемый период по отдельным направлениям и в среднем;
- в) структуру импорта табачного сырья для каждого года, как изменение структуры импорта отразилось на средних импортных ценах;
- г) как изменились затраты на импорт табачного сырья за счет изменения объемов импорта, структуры импорта и импортных цен?

2. По результатам расчетов напишите аналитическую записку.

20. Имеются данные об экспорте асбеста Российской Федерацией (табл. 14.28).

Таблица 14.28

Направление экспорта	Количество, тыс. т		Стоимость экспорта, млн долл.	
	2003 г.	2004 г.	2003 г.	2004 г.
Страны СНГ	119	161	19,2	27,2
Страны вне СНГ	348	414	52,5	66,8

1. Определите:

а) средние экспортные цены на асбест и их динамику;

б) изменение средней цены (в абсолютном и относительном выражении) по факторам: за счет изменения цен в среднем по направлениям экспорта и за счет изменений в структуре экспорта;

в) изменение выручки от экспорта асбеста (в абсолютном и относительном выражении);

г) изменение выручки от экспорта асбеста за счет следующих факторов:

- изменения средних экспортных цен;
- перераспределения структуры экспорта;
- изменения физического объема экспорта.

2. По результатам вычислений напишите аналитическую записку.

21. В табл. 14.29 приведены данные об экспорте Российской Федерацией железных руд и концентратов.

Таблица 14.29

Направление экспорта	Количество, тыс. т		Стоимость экспорта, млн долл.	
	2003 г.	2004 г.	2003 г.	2004 г.
Страны СНГ	7532	3165	91,9	73,3
Страны вне СНГ	9072	13824	177,0	437,0

Определите:

а) изменение общей выручки от экспорта железных руд и концентратов;

б) как в среднем за год изменились экспортные цены и как это отразилось на изменении выручки;

в) как изменился физический объем экспорта и как это сказалось на экспортной выручке;

г) какое влияние на выручку оказало изменение структуры экспорта железных руд и концентратов?

22. В табл. 14.30 приведены данные об экспорте кокса и полукокса Российской Федерацией.

Таблица 14.30

Направление экспорта	Количество, тыс. т		Стоимость экспорта, млн долл.	
	2003 г.	2004 г.	2003 г.	2004 г.
Страны СНГ	2041	1809	155	277
Страны вне СНГ	1496	2063	89,9	282

1. Рассчитайте средние экспортные цены для каждого года и экспортные цены по отдельным направлениям.

2. Оцените динамику средних экспортных цен в целом и по отдельным направлениям экспорта.

3. Разложите изменение средней экспортной цены (в абсолютном и относительном выражении) по факторам: за счет изменения цен в среднем по направлениям экспорта и за счет изменений в структуре экспорта.

4. Определите изменение выручки от экспорта кокса и полукокса.

5. Покажите изменение выручки от экспорта кокса и полукокса за счет следующих факторов:

- а) изменения средних экспортных цен;
- б) перераспределения структуры экспорта;
- в) изменения физического объема экспорта.

23. В табл. 14.31 приведены данные об экспорте труб из черных металлов Российской Федерацией.

Таблица 14.31

Направление экспорта	Количество, тыс. т		Стоимость, млн долл.	
	2003 г.	2004 г.	2003 г.	2004 г.
Страны СНГ	402	594	254	524
Страны вне СНГ	659	1000	254	478
Всего	1061	1594	508	1002

1. Определите, как изменился в целом физический объем экспорта труб из черных металлов? Как за счет этого фактора изменилась экспортная выручка?

2. Рассчитайте сводный индекс экспортной выручки.

3. Оцените изменение экспортных цен по отдельным направлениям экспорта и в среднем.

4. Как изменилась выручка от экспорта труб за рассматриваемый период в абсолютном выражении, в том числе за счет роста объемов поставок и за счет роста экспортных цен. Найдите долю прироста, полученную за счет каждого из факторов.

24. В табл. 14.32 приведены данные об экспорте фосфатов Российской Федерацией.

Таблица 14.32

Направление экспорта	Количество, тыс. т		Стоимость, млн долл.	
	2003 г.	2004 г.	2003 г.	2004 г.
Страны СНГ	263	339	19,4	25,8
Страны вне СНГ	3128	2743	121	127
Всего	3391	3082	140,4	152,8

1. Рассчитайте экспортные цены по отдельным направлениям экспорта. Какое направление экспорта фосфатов более выгодно с точки зрения экспортных цен?

2. Определите средние экспортные цены на фосфаты за каждый год.

3. Как изменились экспортные цены по отдельным направлениям экспорта и в целом.

4. Определите структуру экспорта фосфатов за каждый год.

5. Вычислите индексы средних экспортных цен постоянно-го состава и структурных сдвигов.

6. Интерпретируйте полученные результаты.

25. Используя данные предыдущей задачи, определите:

1. Индивидуальные и сводные индексы физического объема, экспортных цен и экспортной выручки.

2. Изменение выручки от экспорта фосфатов в абсолютном выражении.

3. Как на изменение выручки повлияли:
 - а) рост цен по отдельным направлениям экспорта;
 - б) изменение структуры экспорта;
 - в) изменение объемов экспорта.

14.3. Вопросы для самоконтроля

1. Дайте определение индекса. Какие задачи решаются с помощью индексов? Назовите главное условие применения индексного метода.

2. Что понимают под синтетической и аналитической функциями индексов?

3. Назовите виды индексов: по степени охвата единиц совокупности, по базе сравнения, способу построения, характеру весов, по выполняемым функциям.

4. Как вы понимаете выражение «система индексов»? Приведите примеры индексов, которые увязываются в систему?

5. Как происходит выбор весов при построении агрегатных индексов? По каким весам строится индекс, если он рассчитывается по формуле Ласпейреса? по формуле Пааше? Почему индекс Фишера называют «идеальным» индексом? В чем достоинства и недостатки указанных индексов?

6. Можно ли назвать агрегатный индекс «показателем центральной тенденции»? Обоснуйте свой ответ.

7. Как рассчитать сводный индекс цен, если известны индивидуальные индексы цен?

8. Дайте определения цепных и базисных индексов. В каких случаях существует взаимосвязь между цепными и базисными индексами?

9. Назовите основные свойства индексов.

10. Для решения каких задач строится система индексов переменного состава, постоянного состава и структурных сдвигов?

11. При изучении динамики среднего уровня производительности труда на предприятии индекс структурных сдвигов оказался равным 1,023. Поясните этот результат?

12. Приведите примеры использования индексного метода в практике государственной статистики.

14.4. Методические рекомендации преподавателям

Аудиторная работа. Перед решением задач по теме «Индексы» целесообразно кратко обсудить сущность и основные задачи индексного анализа, произвести разбор основных терминов, которые будут использоваться при решении задач (индивидуальные и сводные индексы, система индексов, индексируемый признак и признак-вес, простые и аналитические индексы и т.д.).

Знакомство с индексным методом целесообразно начать с решения комплексной задачи на расчет систем индивидуальных и сводных индексов взаимосвязанных признаков (например, цена, физический объем продаж и выручка или себестоимость, выпуск продукции и затраты) по совокупности, состоящей из двух-трех элементов (товаров). При построении сводных индексов основное внимание должно быть уделено правилам выбора весов. Целесообразно обсудить разные варианты взвешивания, их достоинства и недостатки, различия в итоговых результатах. Особо следует остановиться на соотношении значений индивидуальных и сводных индексов. Выработать навыки содержательной интерпретации промежуточных и итоговых результатов индексного анализа, написания аналитических записок.

Значение сводных индексов как характеристики центральной тенденции изменения признака наиболее наглядно демонстрируется на примере построения сводных индексов как средних из индивидуальных. При решении задач на расчет средних из индивидуальных индексов целесообразно показать, как меняются значение и интерпретация индекса в зависимости от выбора признака-веса.

Прежде чем приступать к решению задач на систему индексов переменного состава, постоянного состава, структурных сдвигов, необходимо повторить со студентами схему построения средних величин вторичных признаков на основе исходного соотношения средней. Это позволит в дальнейшем избежать ошибок в выборе признака-веса при конструировании формул вычисления индексов.

Если условие задачи предусматривает возможность вычисления индекса более чем одним способом, каждый последую-

щий способ расчета должен рассматриваться одновременно как способ проверки ранее полученных результатов.

Достаточно продуктивным является обсуждение небольших задач (тестов) на знание основных взаимосвязей индексов (между индексами конкретных признаков, индивидуальными и сводными индексами, цепными и базисными индексами, индексами переменного состава, постоянного состава и структурных сдвигов). Решение таких задач не только позволяет закрепить пройденный материал, но и демонстрирует широкие возможности и разнообразие сфер применения индексного метода.

Самостоятельная внеаудиторная работа. Предполагает решение задач на построение различных видов индексов для закрепления навыков, полученных на практических занятиях, а также освоение теоретических знаний по теме «Индексы», которые остались за рамками практических занятий. Например, на самостоятельное изучение могут быть вынесены вопросы: «Применение индексного метода в практике государственной статистики», «Порядок расчета индекса потребительских цен», «Индексы цен производителей», «Комплексное использование индексного и регрессионного анализа» и др.

Итоговый контроль, как правило, проводится в форме контрольной работы, состоящей из двух-трех задач на разные типы индексов.

14.5. Методические указания студентам

Освоение теории. Лучше запомнить формулы сводных агрегатных индексов вам поможет табл. 14.33.

Таблица 14.33

По схеме	Формула индекса	
	индекс физического объема и других первичных признаков	индекс цен и других вторичных индексов
Ласпейреса (по базисным весам)	$I_q^{\text{Л}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_p^{\text{Л}} = \frac{\sum p_1 q_0}{\sum p_0 q_0}$
Пааше (по отчетным весам)	$I_q^{\text{П}} = \frac{\sum q_1 p_1}{\sum q_0 p_1}$	$I_p^{\text{П}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}$

Увязываются между собой в систему индексы, расположенные на диагоналях таблицы:

$$I_w = I_q^{\Pi} \cdot I_p^{\Pi} \text{ или } I_w = I_q^{\Pi} \cdot I_p^{\Pi}.$$

Особое внимание уделяйте интерпретации вычисленных индексов. Если есть возможность вычислить индекс несколькими способами, обязательно воспользуйтесь этим для проверки результатов.

Табл. 14.34 поможет вам при формулировке выводов по индексам, характеризующим изменение взвешенных средних.

Таблица 14.34

Название индекса	Интерпретация индекса
Индекс переменного состава	Изменение среднего уровня признака в отчетном периоде по сравнению со средним уровнем признака в базисном периоде
Индекс постоянного состава	Среднее изменение осредняемого признака
Индекс структурных сдвигов	Изменение средней величины признака за счет признака-веса (структурных изменений)

В экономическом анализе индексы широко используются для оценки изменений средних взвешенных величин — средней заработной платы, себестоимости, урожайности, производительности, трудоемкости, рентабельности и т.д. Умение строить и интерпретировать индексы переменного, постоянного состава и структурных сдвигов пригодятся вам при изучении экономической, социальной и отраслевых статистик. Вы должны знать:

- сущность и значение индексного метода анализа;
- классификацию индексов по охвату единиц совокупности, форме построения, характеру решаемых задач.

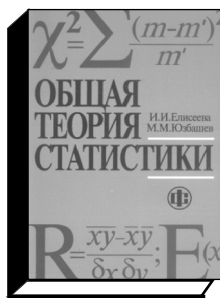
Практические навыки. Уметь использовать индексный метод для анализа социально-экономических явлений и процессов, грамотно интерпретировать полученные результаты.

Издательство
“ФИНАНСЫ И СТАТИСТИКА”
предлагает учебник

И.И.Елисеева, М.М.Юзбашев

**ОБЩАЯ
ТЕОРИЯ СТАТИСТИКИ**

656 с.



Рассматриваются вопросы организации статистики, ее особенности в нашей стране и других странах. Пятое издание (4-е изд. — 1999 г.) расширено изложением методов сбора данных с применением современных технологий, приведены методы многомерной классификации, раскрыта основополагающая роль вариационного анализа. Описание теоретических основ выборочного метода и методов проверки статистических гипотез сочетается с примерами их использования в исследованиях и в работе органов государственной статистики.

Особое внимание уделено статистическому анализу нечисловых переменных систем регрессионных уравнений, анализу временных рядов. Для студентов высших учебных заведений, обучающихся по направлению и специальности «Статистика».

**По вопросам приобретения литературы
обращайтесь, пожалуйста, в Издательство по адресу:**

101000, Москва, ул.Покровка, 7
(метро “Китай-город”, выход на ул.Маросейка)

Тел.: (495) 625-35-02, 623-80-42

Факс: (495) 623-18-68, 625-09-57

E-mail: mail@finstat.ru <http://www.finstat.ru>

При Издательстве работает киоск:

понедельник — четверг — с 9.00 до 19.30,

пятница — с 9.00 до 18.30

Тел.(495) 621-86-57

Система **“Книга — почтой”**

Стоимость пересылки почтовыми бандеролями —
25% от стоимости заказа

СТАТИСТИЧЕСКИЕ ДАННЫЕ

Таблица П.1

Макроэкономические показатели по странам мира в 2003 г.

Страна	ВВП на душу населения, долл. США	ВВП на душу населения, ППС долл. США	ВНД на душу населения, долл. США	Энергопотребление на душу населения, кг нефт. эквивалента	Выбросы диоксида углерода на душу населения, м/т	Легковые автомобили на 1000 чел.	Экспорт высоких технологий, % от промышленных товаров	Ожидаемая продолжительность жизни при рождении, лет
Австралия	26275	29632	21950	5732	18	...	14	80,2
Австрия	31289	30094	26810	3774	3774	494	13	78,9
Азербайджан	867	3617	820	1435	3,6	43	5	66,9
Албания	1933	4584	1740	617	0,9	43	1	73,7
Алжир	2090	6107	1930	985	2,9	...	2	71
Ангола	975	2344	740	672	0,5	40,7
Аргентина	3524	12106	3810	1543	3,9	140	9	74,3
Армения	918	3671	950	632	1,1	...	1	71,4
Бангладеш	376	1770	400	155	0,2	0	...	62,6
Беларусь	1770	6052	1600	2496	5,9	156	4	68,1
Бельгия	29096	28335	25760	5505	10	464	8	78,8
Бенин	517	1115	440	340	0,3	...	2	53,8
Болгария	2539	7731	2130	2417	5,3	287	4	72,1
Боливия	892	2587	900	499	1,3	...	8	63,9

Босния и Герцеговина	1684	5967	1530	1052	4,8	74,1
Ботсвана	4372	8714	3530	...	2,3	30	...	36,6
Бразилия	2788	7790	2720	1093	1,8	...	12	70,3
Буркина-Фасо	345	1174	300	...	0,1	...	2	47,4
Бурунди	83	648	90	...	0	...	22	43,5
Бутан	797	...	630	...	0,5	62,7
Великобритания	30253	27147	28320	3824	9,5	384	26	78,3
Венгрия	8169	14584	2580	2697	6,7	138	26	72,6
Венесуэла	3326	4919	3490	2141	6,5	...	4	72,8
Вьетнам	482	2490	480	530	0,7	...	2	70,4
Габон	4505	6397	3340	1209	2,8	54,6
Гаити	346	1742	400	251	0,2	51,5
Гамбия	278	1859	270	...	0,2	55,5
Гана	369	2238	320	411	0,3	...	3	56,7
Гватемала	209	4148	1910	615	0,9	1	7	67,1
Гвинея	459	2097	430	...	0,2	53,6
Гвинея-Бисау	160	711	140	...	0,2	44,6
Германия	29115	27756	25270	4198	9,5	516	16	78,7
Гондурас	1001	2665	970	504	0,7	51	...	67,6
Гонконг, Китай	22987	27179	25860	2412	5	57	13	81,5
Греция	15608	19954	13230	2637	8,2	254	12	78,2
Грузия	778	2588	770	494	1,2	56	24	70,5
Дания	39332	31465	33570	3675	8,4	360	20	77,1
Демократическая Республика Конго	107	...	100	299	0,1	43,1

Страна	ВВП на душу населения, долл. США	ВВП на душу населения, ППС долл. США	ВНД на душу населения, долл. США	Энергопотребление на душу населения, кг нефт. эквивалента	Выбросы диоксида углерода на душу населения, м/г	Легковые автомобили на 1000 чел.	Экспорт высоких технологий, % от промышленных товаров	Ожидаемая продолжительность жизни при рождении, лет
Доминиканская Республика	1893	6823	2130	948	3	...	1	67,1
Египет	1220	3950	1390	789	2,2	69,6
Замбия	417	877	380	639	0,2	...	2	37,4
Зимбабве	...	2443	...	751	1,2	...	3	37,2
Израиль	16481	20033	16240	3191	10	230	18	79,7
Индия	564	2892	540	513	1,1	6	5	63,1
Индонезия	970	3361	810	737	1,3	...	14	66,5
Иордания	1858	4320	1850	1036	3,2	...	2	71,2
Иран	2066	6995	2010	2044	4,9	...	2	70,2
Ирландия	38487	37738	27010	3894	11,1	349	34	77,7
Испания	20404	22391	17040	3215	7	441	7	79,5
Италия	25471	27119	21570	2994	7,4	542	8	80
Йемен	565	889	520	221	0,5	60,3
Казахстан	2000	6671	1780	3123	8,1	72	9	63,2
Камбоджа	315	2078	300	...	0	312	...	56
Камерун	776	2118	630	417	0,4	...	2	45,8
Канада	27079	30677	24470	7973	14,2	559	14	79,9
Кения	450	1037	400	489	0,3	8	4	47
Киргизия	378	1151	340	507	0,9	38	2	66,8

Китай	1100	5003	1100	960	2,2	7	27	71,5
Колумбия	1764	6702	1810	625	1,4	43	7	72,2
Конго	949	965	650	252	0,5	51,9
Коста-Рика	4352	9606	4300	904	1,4	...	45	78,1
Кот-д'Ивуар	816	1476	660	397	0,7	...	8	46
Куба	1262	2,8	16	29	77,2
Кувейт	17421	19047	17960	9503	21,9	...	1	76,8
Лаосская НДР	375	1759	340	...	0,1	54,5
Латвия	4771	10270	4400	1825	2,5	266	4	71,4
Лесото	635	2561	610	36,7
Ливан	4224	5074	4040	1209	3,5	...	2	71,9
Ливия	3433	10,9	73,4
Литва	5274	11702	4500	2476	3,4	346	5	72,2
Маврикий	4274	11287	410	...	2,4	86	5	72,1
Мавритания	384	1766	400	...	1,2	52,5
Малагаскар	324	809	290	...	0,1	55,3
Македония	2277	6794	1980	...	5,5	...	1	73,3
Малави	156	605	160	...	0,1	...	1	39,6
Малайзия	4187	9512	3880	2129	6,2	...	58	73
Мали	371	994	290	...	0,1	...	8	47,8
Марокко	1452	4004	1310	363	1,3	44	11	69,5
Мексика	6121	9168	6230	1560	4,3	107	21	74,9
Мозамбик	230	1117	210	436	0,1	...	3	41,9
Монголия	514	1850	480	...	3,1	28	...	63,9
Мьянма	258	0,2	60,1
Намбия	2120	6180	1930	599	1	38	3	48,6
Непал	237	1420	240	353	0,1	61,4
Нигер	232	835	200	...	0,1	...	3	44,3
Нигерия	428	1150	350	718	0,3	43,3

Страна	ВВП на душу населения, долл. США	ВВП на душу населения, ППС долл. США	ВНД на душу населения, долл. США	Энергопотребление на душу населения, кг нефт. эквивалента	Выбросы диоксида углерода на душу населения, м/г	Легковые автомобили на 1000 чел.	Экспорт высоких технологий, % от промышленных товаров	Ожидаемая продолжительность жизни при рождении, лет
Нидерланды	31532	29371	26230	4827	8,7	384	31	78,3
Никарагуа	745	3262	740	544	0,7	16	4	69,5
Новая Зеландия	19847	22582	15520	4573	8,3	613	10	79
Норвегия	48412	37670	43400	5843	11,1	417	19	79,3
ОАЭ	9609	18,1	...	2	77,9
Объединенная Республика Танзания	287	621	300	408	0,1	...	2	46
Оккупированные Палестинские Территории	1026	...	1110	72,4
Оман	...	13584	7830	4265	8,2	...	2	74
Пакистан	555	297	520	454	0,8	7	...	62,9
Панама	4319	6857	4060	1028	2,2	...	1	74,7
Папуа-Новая Гвинея	578	2619	500	...	0,5	...	39	55,1
Парагвай	1069	4684	1100	709	0,7	...	6	70,9
Перу	2231	5260	2140	450	1,1	30	2	69,8

Польша	5487	11379	5280	2333	7,8	259	3	74,3
Португалия	14161	18126	11800	2546	5,8	426	9	77,2
Республика Корея	12634	17971	12030	4272	9,1	205	32	76,9
Республика Молдова	463	1510	590	703	1,5	74	3	67,5
Россия	3018	9230	2610	4288	9,9	132	19	65,4
Руанда	195	1268	220	...	0,1	...	25	43,6
Румыния	2619	7277	2260	1696	3,8	144	4	71,3
Сальвадор	2277	4781	2340	670	1,1	30	1	70,7
Саудовская Аравия	9532	13226	9240	5775	18,1	71,6
Сенегал	634	1648	540	319	0,4	11	9	55,6
Сингапур	21492	24481	21230	6078	14,7	122	59	78,6
Сирийская Арабская Республика	1237	3576	1160	1063	3,3	9	1	73,2
Словакия (Словацкая Республика)	6033	13494	4940	3448	6,6	247	4	74
Словения	13909	19150	11920	3486	7,3	438	6	76,3
Судан	530	1910	460	483	0,2	...	7	56,3
США	37648	37562	37870	7943	19,8	481	31	77,3
Сьерра-Леоне	149	548	150	...	0,1	...	31	40,6
Таджикистан	246	1106	210	518	0,6	63,5
Таиланд	2305	7595	21090	1353	3,3	...	30	69,7
Того	362	1696	310	324	0,4	...	1	54,2
Тринидад и Тобаго	8007	10766	7790	7121	20,5	...	2	69,9
Тунис	2530	7161	2240	846	1,9	...	4	73,1

Страна	ВВП на душу населения, долл. США	ВВП на душу населения, ППС долл. США	ВНД на душу населения, долл. США	Энергопотребление на душу населения, кг нефт. эквивалента	Выбросы диоксида углерода на душу населения, м/т	Легковые автомобили на 1000 чел.	Экспорт высоких технологий, % от промышленных товаров	Ожидаемая продолжительность жизни при рождении, лет
Туркменистан	1275	5938	1120	3465	7,5	62,4
Турция	3399	6772	2800	1083	3,3	66	2	68,6
Уганда	249	1457	250	...	0,1	...	8	46,8
Узбекистан	389	1744	420	2047	4,8	66,5
Украина	1024	5491	970	2684	6,9	108	5	66,1
Уругвай	3308	8280	3820	747	1,6	...	2	75,3
Филиппины	989	4321	1080	525	1	9	74	70,2
Финляндия	31058	27619	27060	5852	10,3	419	24	78,4
Франция	29410	27677	24730	4470	6,2	491	19	79,4
Хорватия	6479	11080	5370	1852	4,5	280	12	74,9
Центрально-африканская Республика	309	1089	260	...	0,1	39,4
Чад	304	1210	240	...	0	43,6
Чехия	8794	16357	7150	4090	11,6	356	13	75,5

Чили	4591	10274	4360	1585	3,9	87	3	77,9
Швейцария	43553	30552	40680	3723	5,4	507	22	80,5
Швеция	33676	26750	28910	5718	5,3	452	15	80,1
Шри-Ланка	948	3778	930	430	0,6	11	1	73,9
Эквадор	2091	3641	1830	706	2	44	6	74,2
Эритрея	171	849	190	...	0,1	53,5
Эстония	6713	13539	5380	3324	11,7	296	13	71,2
Эфиопия	97	711	90	292	0,1	1	...	47,6
ЮАР	3489	10346	2750	2502	7,4	94	5	49
Ямайка	3083	4104	2980	1493	4,2	70,7
Япония	33713	27967	34180	4058	9,3	428	24	81,9

Источники: Доклад о развитии человека. 2005. Международное сотрудничество на перепутье: помощь, торговля и безопасность в мире неравенства: пер. с англ. — М.: Весь мир, 2005. — С. 272–275, 288–290, 296–299; Состояние окружающей среды. 2005: Статистический справочник Всемирного банка: пер. с англ. — М.: Весь мир, 2005. В таблицу включены страны с численностью населения более 1 млн чел., в которых проживает 98,1% населения мира.

Таблица П1.2
Распространение информационных технологий и НИОКР в странах мира в 2003 г.¹

№ п/п	Страна	Магистраль- ные теле- фонные линии на 1 тыс. чел.	Число обо- нентов сото- вых сетей на 1 тыс. чел.	Пользователи сети Интернет на 1 тыс. чел.	Расходы на НИОКР, % от ВВП	Работники, занятые в НИОКР, на 1 млн чел.	Уровень дохода на душу наследо- ния ²	Числен- ность населения, млн чел.
1	Австралия	542	719	567	1,5	3446	Высокий	19,7
2	Австрия	481	879	462	2,2	2346	Высокий	8,1
3	Азербайджан	114	128	37 ³	0,3	1248	Средний	8,3
4	Албания	83	358	10	Средний	3,1
5	Алжир	69	45	16 ³	Средний	31,9
6	Ангولا	7	9 ³	3 ³	Низкий	15,0
7	Аргентина	219 ³	178 ³	112 ³	0,4	715	Средний	38,0
8	Армения	148	30	37	0,3	1606	Средний	3,0
9	Бангладеш	5	10	2	...	51 ³	Низкий	136,6
10	Беларусь	311	113	141	0,6	1870	Средний	9,9
11	Бельгия	489	793	386	2,2	3180	Высокий	10,4
12	Бенин	9	34	10	...	174 ³	Низкий	7,9
13	Болгария	380	466	206	0,5	1158	Средний	7,8
14	Боливия	72	152	32 ³	0,3	118	Средний	8,8
15	Босния и Герцеговина	245	274	26 ³	Средний	3,9
16	Ботсвана	75	297	Средний	1,8
17	Бразилия	223	264	82 ³	1	324	Средний	181,4
18	Буркина- Фасо	5	19	4	0,2	17	Низкий	12,4
19	Бурунди	3	9	2	...	21 ³	Низкий	7,0
20	Бутан	34	11	20	Низкий	2,1

21	Великобритания	591 ³	841 ³	423 ³	1,9	2691	Высокий	59,3
22	Венгрия	349	769	232	1	1473	Средний	10,2
23	Венесуэла	111	273	60	0,4	222	Средний	25,8
24	Вьетнам	54	34	43	...	274 ³	Низкий	82,0
25	Габон	29	224	26	Средний	1,3
26	Гаити	17	38	18	Низкий	8,3
27	Гамбия	29 ³	75 ³	19 ³	Низкий	1,7
28	Гана	13	36	8 ³	Средний	21,2
29	Гватемала	77	165	33 ³	...	103 ³	Средний	12,0
30	Гвинея	3	14	5	...	286	Низкий	9,0
31	Гвинея-Бисау	8	1	15	Низкий	1,5
32	Германия	657	785	473	2,5	3222	Высокий	82,6
33	Гондурас	49	55	40	0,1	74	Средний	6,9
34	Гонконг, Китай (САР)	559	1079	472	0,6	1568	Высокий	6,9
35	Греция	454	902	150	0,6	1357	Высокий	11,1
36	Грузия	134	145	24	0,3	2317	Средний	4,6
37	Дания	669	883	541	2,5	4822	Высокий	5,4
38	Демократическая Республика Конго	1 ³	19	1 ³	Низкий	54,2
39	Доминиканская Республика	115	272	102	Средний	8,6
40	Египет	127	84	44	0,2	493 ³	Средний	71,3
41	Замбия	8	22	6	...	47	Низкий	11,3
42	Зимбабве	25 ³	30 ³	43 ³	Низкий	12,9
43	Израиль	458	961	301 ³	5,1	1570	Высокий	6,5

№ п/п	Страна	Магистраль- ные теле- фонные линии на 1 тыс. чел.	Число абоне- ментов сотовых сетей на 1 тыс. чел.	Пользователи сети Интернет на 1 тыс. чел.	Расходы на НИОКР, % от ВВП	Работники, занятые в НИОКР, на 1 млн чел.	Уровень дохода на душу населения ²	Численность населения, млн чел.
44	Индия	46	25	17	0,8	120	Низкий	1070,8
45	Индонезия	39	87	38	...	130 ³	Средний	217,4
46	Иордания	114	242	81	6,3 ³	1977	Средний	5,4
47	Иран	220	51	72	...	484	Средний	68,2
48	Ирландия	491	880	317	1,1	2315	Высокий	4,0
49	Испания	429	916	239	1	2036	Высокий	42,1
50	Италия	484	1018	337	1,1	1156	Высокий	58,0
51	Йемен	28 ³	35	5 ³	Низкий	19,7
52	Казахстан	141	64 ³	16 ³	0,3	744	Средний	14,9
53	Камбоджа	3	35	2	Низкий	13,5
54	Камерун	7 ³	66	4 ³	Низкий	15,7
55	Канада	651	419	513 ³	1,9	3487	Высокий	31,6
56	Кения	10	50	13 ³	Высокий	32,7
57	Киргизия	76	27	38	0,2	413	Низкий	5,1
58	Китай	209	215	63	1,2	633	Средний	1300,0
59	Колумбия	179	141	53	0,1	81	Средний	44,2
60	Конго	2	94	4	...	29	Низкий	3,8
61	Коста-Рика	278	181	288	0,4	533	Средний	4,2
62	Кот-д'Ивуар	14	77	14	Низкий	17,6
63	Куба	64	3	9	0,5	538	Средний	11,2
64	Кувейт	196	572	228	0,2	73	Высокий	2,5
65	Лаосская НДР	12	20	3	Низкий	5,7

66	Латвия	285	526	404	0,4	1476	Средний	2,3
67	Лесото	16	47	14	...	42	Низкий	1,8
68	Ливан	200	234	143	Средний	3,5
69	Ливия	136	23	29	...	361	Средний	5,6
70	Литва	239	630	202	0,7	1824	Средний	3,5
71	Маврикий	285	267	123	0,3	360 ³	Средний	1,2
72	Мавритания	14	127	4	Низкий	2,9
73	Мадагаскар	4	17	4	0,1	15	Низкий	17,6
74	Македония	252	372	60	0,3	500	Средний	2,0
75	Малави	8	13	3	Низкий	12,3
76	Малайзия	182	442	344	0,7	294	Средний	24,4
77	Мали	5 ³	23	3 ³	Низкий	12,7
78	Марокко	40	244	33	Средний	30,6
79	Мексика	160	295	120	0,4	259	Средний	104,3
80	Мозамбик	5 ³	23	3 ³	Низкий	19,1
81	Монголия	56	130	58	0,3	710	Низкий	2,6
82	Мьянма	7	1	1	Низкий	49,5
83	Намибия	66	116	34	Средний	2,0
84	Непал	16	2	3 ³	0,7	62	Низкий	26,1
85	Нигер	2 ³	6	1 ³	Низкий	13,1
86	Нигерия	7	26	6	...	15 ³	Низкий	125,9
87	Нидерланды	614	768	522	1,9	2826	Высокий	16,1
88	Никарагуа	37	85	17 ³	0,1	73	Низкий	5,3
89	Новая Зеландия	448	648	526	1,2	2593	Высокий	3,9
90	Норвегия	713	909	346	1,7	4442	Высокий	4,6
91	ОАЭ	281	736	275	Высокий	4,0
92	Объединенная Республика Танзания	4	25	7	Низкий	36,9

№ п/п	Страна	Магистраль- ные теле- фонные линии на 1 тыс. чел.	Число обо- нентов сото- вых сетей на 1 тыс. чел.	Пользователи сети Интернет на 1 тыс. чел.	Расходы на НИОКР, % от ВВП	Работники, занятые в НИОКР, на 1 млн чел.	Уровень дохода на душу наследо- вания ²	Числен- ность населения, млн чел.
93	Оккупиро- ванные Па- лестинские Территории	87	133	40	Средний	3,5
94	Оман	88	228	71 ³	...	4 ³	Средний	2,5
95	Пакистан	27	18	10 ³	0,2	88	Низкий	151,8
96	Панама	122	268	62	0,4	95	Средний	3,1
97	Папуа-Новая Гвинея	11 ³	3 ³	14 ³	Низкий	5,7
98	Парагвай	46	299	20	0,1	83	Средний	5,9
99	Перу	67	160	104	0,1	225	Средний	27,2
100	Польша	307	451	232	0,6	1469	Средний	38,6
101	Португалия	411	898	194 ³	0,9	1745	Высокий	10,4
102	Республика Корея	531	701	610	2,5	2979	Высокий	47,5
103	Республика Молдова	219	132	80	0,6 ³	171	Низкий	4,2
104	Россия	253	249	41 ³	1,2	3415	Средний	144,6
105	Руанда	3 ³	14	3 ³	...	30 ³	Низкий	8,8
106	Румыния	199	324	184	0,4	910	Средний	21,9
107	Сальвадор	113	173	83	47 ³	131	Средний	6,6
108	Саудовская Аравия	155	321	67	Средний	23,3

109	Сенегал	22	56	22	...	2	Низкий	11,1
110	Сингапур	450	852	509	2,2	4352	Высокий	4,2
111	Сирийская Арабская Республика	...	68	35	0,2	29	Средний	18,1
112	Словакия (Словацкая Республика)	241	684	256	0,6	1707	Средний	5,4
113	Словения	407	871	401	1,5	2364	Высокий	2,0
114	Судан	27	20	9	Низкий	34,9
115	США	624	546	556	2,7	4526	Высокий	292,6
116	Сьерра-Леоне	5 ³	14 ³	2 ³	Низкий	5,1
117	Таджикистан	37	7	1	...	660 ³	Низкий	6,4
118	Таиланд	105	394	11	0,2	289	Средний	63,1
119	Того	12	44	42	...	102 ³	Низкий	5,8
120	Тринидад и Тобаго	250 ³	399	106 ³	0,1	347	Средний	1,3
121	Тунис	118	197	64	0,6	1013	Средний	9,9
122	Туркменистан	77	2 ³	2 ³	Средний	4,7
123	Турция	268	394	85	0,7	345	Средний	71,3
124	Уганда	2	30	5	0,8	25	Низкий	26,9
125	Узбекистан	67	13	19	...	1754 ³	Низкий	25,8
126	Украина	233	136	18 ³	1,2	1749	Средний	47,5
127	Уругвай	280 ³	193 ³	119 ³	0,2	370	Средний	3,4
128	Филиппины	41	270	44 ³	...	156 ³	Средний	80,2
129	Финляндия	492	910	534	3,5	7461	Высокий	5,2
130	Франция	566	696	366	2,3	3134	Высокий	60,0
131	Хорватия	417 ³	584	232	1,1	1920	Средний	4,5

№ п/п	Страна	Магистраль- ные теле- фонные линии на 1 тыс. чел.	Число абон- ентов сото- вых сетей на 1 тыс. чел.	Пользователи сети Интернет на 1 тыс. чел.	Расходы на НИОКР, % от ВВП	Работники, занятые в НИОКР, на 1 млн чел.	Уровень дохода на душу населения ²	Числен- ность населения, млн чел.
132	Центрально- африканская Республика				...	47	Низкий	3,9
133	Чад	2 ³	10	1	Низкий	9,1
134	Чехия	2 ³	8	2 ³	Средний	10,2
135	Чили	360	965	308	1,2	1467	Средний	16,0
136	Швейцария	221	511	272	0,5	419	Высокий	7,2
137	Швеция	727	843	379	2,6	3594	Средний	9,0
138	Шри-Ланка	736 ³	980	573 ³	4,3	5171	Средний	20,4
139	Эквадор	49	73	13	0,2 ³	197	Средний	12,9
140	Эритрея	122	189	46	0,1	84	Низкий	4,1
141	Эстония	9	0	7	Средний	1,3
142	Эфиопия	341	777	444	0,7	2253	Низкий	73,8
143	ЮАР	6	1	1	Средний	46,9
144	Ямайка	107 ³	364	68 ³	0,7	192	Средний	2,6
145	Япония	169 ³	680	228 ³	...	8 ³	Высокий	127,7
		472	679	483	3,1	5085		

¹ Источники: Доклад о развитии человека. 2005. Международное сотрудничество на переломе: помощь, торговля и безопасность в мире неравенства: пер. с англ. — М.: Весь Мир, 2005. — С. 254–257, 284–287, 390. Доклад о развитии человека. 2004. Культурная свобода в современном многообразном мире: пер. с англ. — М.: Весь мир, 2004. — С. 202–205. В таблицу включены страны с численностью населения более 1 млн. чел., в которых проживает 98,1% населения мира.

² Классификация Всемирного банка (действительная на 1 июля 2004 г.) основывается на валовом национальном доходе (ВНД) на душу населения. Страны с высоким уровнем дохода — ВНД на душу населения в 2003 г. равен 9386 или более долл. США. Страны со средним уровнем дохода — ВНД на душу населения в 2003 г. равен 766–9385 долл. США. Страны с низким уровнем дохода — ВНД на душу населения в 2003 г. равен 765 или менее долл. США.

³ Данные за 2002 г.

Таблица П1.3

Банки Северо-Западного региона по величине активов в первом полугодии 2006 г.

Место в рейтинге	01.07.2006	01.01.2006	Город	Банк	Активы, тыс. руб.		Собственный капитал, тыс. руб.*	Место по количеству капитала на 01.07.2006	Балансовая прибыль, тыс. руб.	
					01.07.2006	01.01.2006			01.07.2006	01.01.2006 (приведенная к полугодию)
1	1		Санкт-Петербург	Промышленно-строительный банк	133 769 960	141 431 022	12 231 433	1	2 783 466	2 004 136
2	3		Санкт-Петербург	Банк «Санкт-Петербург»	40 890 909	31 166 365	4 210 065	3	528 635	583 519
3	2		Санкт-Петербург	Национальный банк «Траст»	40 230 228	38 057 445	3 209 107	5	148 889	288 783
4	4		Санкт-Петербург	Кит Финанс Инвестиционный банк	30 143 837	20 121 986	5 163 693	2	3 102 490	609 690
5	5		Санкт-Петербург	Россия	24 388 430	19 545 716	3 222 769	4	374 762	261 523
6	6		Санкт-Петербург	Международный банк Санкт-Петербурга	16 729 758	15 542 258	1 520 165	8	107 382	93 909

Место в рейтинге	Банк	Город	Активы, тыс. руб.		Собственный капитал, тыс. руб.*		Место по количеству капитала на 01.07.2006	Балансовая прибыль, тыс. руб.	
			01.07.2006	01.01.2006	01.07.2006	01.01.2006		01.07.2006	01.01.2005 (приведенная к полугодию)
7	Банк ВЕФК	Санкт-Петербург	14 880 891	14 421 529	1 611 144	887 381	7	100 196	129 389
8	КАЛИОН РУСБАНК	Санкт-Петербург	14 869 861	12 309 557	1 512 171	1 468 061	9	166 044	79 304
9	Дрезднер Банк	Санкт-Петербург	13 506 171	13 520 940	2 632 661	2 587 542	6	102 268	106 301
10	СЕВЕРГАЗ-БАНК	Вологда	10 672 768	9 984 745	782 914	728 165	14	137 763	79 593
11	Инкасбанк	Санкт-Петербург	10 590 456	9 277 544	1 245 600	1 159 986	10	73 075	72 270
12	Металлургический коммерческий банк	Череповец	10 440 331	9 070 822	1 193 036	1 047 986	11	170 323	92 008
13	Петро-Аэро-Банк	Санкт-Петербург	9 260 879	7 429 249	624 356	533 480	16	51 305	41 789

14	13	Таврический	Санкт-Петербург	8 803 426	9 066 928	1 157 624	1 139 477	12	89 818	33 893
15	15	БАЛТ-ИНВЕСТ-БАНК	Санкт-Петербург	7 420 214	6 473 511	869 684	857 238	13	38 730	40 311
16	16	Ухтабанк	Ухта	5 753 771	5 720 142	758 003	707 129	15	56 242	67239
17	17	Петербургский социальный коммерческий банк	Санкт-Петербург	3 633 065	3 660 208	568 802	518 540	17	72 791	57 939
18	19	Александровский	Санкт-Петербург	2 722 657	2 478 050	465 132	316 578	18	14 133	12 197
19	18	Рускобанк	Санкт-Петербург	2 545 940	2 533 076	263 374	244 101	27	21 684	23 597
20	20	Советский	Санкт-Петербург	2 429 350	2 312 761	276 055	264 243	26	15 704	9 080
21	25	ПЕТРО-ЭНЕРГО-БАНК	Санкт-Петербург	2 389 015	1 618 413	204 153	184 113	31	23 135	14 524
22	22	Северный Народный Банк	Сыктывкар	2 387 943	2 022 855	231 947	218 265	29	21 703	29 466
23	21	Инвестбанк	Калининград	2 272 936	2 047 784	304 461	290 218	23	14 788	16 495
24	23	Мончбанк	Мурманск	2 166 337	1 906 533	302 688	290 309	25	16 785	13 593

Место в рейтинге	Банк	Город	Активы, тыс. руб.		Собственный капитал, тыс. руб.*		Место по количеству капитала на 01.07.2006	Балансовая прибыль, тыс. руб.	
			01.07.2006	01.01.2006	01.07.2006	01.01.2006		01.07.2006	01.01.2005 (приведенная к полугодию)
25	Европейский	Калининград	2 115 081	1 769 573	160 569	144 770	35	34 714	27 748
26	БАРЕНЦБАНК	Мурманск	1 642 049	1 498 112	307 775	290 232	22	37 149	47 529
27	Промэнергобанк	Вологда	1 510 246	1 326 608	179 049	175 447	33	29 882	23 457
28	ЭКСИ-Банк	Санкт-Петербург	1 504 455	1 321 222	147 202	124 394	37	12 337	7 784
29	Сетевой Нефтяной Банк	Калининград	1 447 384	1 512 394	205 051	188 476	30	21 012	17 552
30	ВИКИНГ	Санкт-Петербург	1 424 006	1 499 507	394 539	397 330	20	10 988	14 141
31	Энерготрансбанк	Калининград	1 350 808	1 145 153	327 633	303 173	21	25 071	10 225

32	32	ЭНЕРГО- МАШБАНК	Санкт- Петер- бург	1 294 667	1 180 636	138 306	140 660	39	3 129	3 131
33	34	Новобанк	Вели- кий Новго- род	1 239 449	1 031 569	303 496	261 805	24	57 004	28 515
34	31	СЕВЗАП- ИНВЕСТ- ПРОМБАНК	Санкт- Петер- бург	1 051 706	1 189 140	408 655	409 597	19	12 137	7 109
35	36	Финансовый капитал	Санкт- Петер- бург	875 094	891 595	121 337	122 385	40	2 009	5 191
36	38	Выборг-Банк	Выборг	823 377	697 354	100 523	94 023	42	10 103	9 646
37	37	Сити Инвест Банк	Санкт- Петер- бург	777 413	735 145	160 566	157 670	36	4 289	4 466
38	35	Констанс- Банк	Санкт- Петер- бург	775 298	900 407	169 128	155 879	34	19 480	6 694
39	39	Воложанин	Воло- гда	731 438	690 919	143 682	114 586	38	23 750	15 505
40	45	Балткредито- банк	Кали- нин- град	707 622	369 273	233 432	46 520	28	2 925	2 448
41	43	Морской торгово- промышлен- ный банк	Санкт- Петер- бург	540 815	493 561	45 191	44 926	55	1 090	4 173

Место в рейтинге	Банк	Город	Активы, тыс. руб.		Собственный капитал, тыс. руб.*		Место по количеству капитала на 01.07.2006	Балансовая прибыль, тыс. руб.	
			01.07.2006	01.01.2006	01.07.2006	01.01.2006		01.07.2006	01.01.2005 (приведенная к полугодию)
42	Объединенный капитал	Санкт-Петербург	529 196	349 147	56 074	51 740	52	5 177	4 046
43	ГАНЗАКОМ-БАНК	Санкт-Петербург	505 031	497 426	44 394	40 905	56	4 143	3 263
44	Русский региональный банк	Псков	488 993	523 175	192 231	211 000	32	-18 690	4 712
45	Региональный кредитный банк	Калининград	480 847	298 215	83 301	40 940	45	3 830	1 633
46	МСКБ	Мурманск	468 261	357 816	75 454	72 657	46	5 444	6 109
47	Онего	Петрозаводск	446 429	496 573	114 972	111 499	41	6 269	5 498
48	Агрокредбанк	Череповец	408 346	414 366	55 411	49 385	53	7 452	7 024
49	Бумеранг	Череповец	408 015	340 489	88 515	85 979	44	5 527	7 844
50	ТЕТРА-ПОЛИС	Санкт-Петербург	408 015	253 241	65 702	60 397	48	5 917	5 296

51	49	Вологдабанк	Вологда	402 947	315 026	70 511	37 042	47	3 358	2 754
52	56	Воркута	Воркута	289 776	135 576	56 222	55 359	51	814	1 894
53	58	Петротрест-банк	Санкт-Петербург	254 927	109 298	30 925	30 249	58	4 212	1 753
54	53	Балтсоюзкомбанк	Калининград	237 533	195 123	97 959	97 587	43	890	336
55	52	АЖБ	Архангельск	236 934	217 038	59 397	57 581	50	2 458	1 013
56	55	Балтика	Калининград	199 157	179 925	60 275	60 143	49	1 044	798
57	54	Псковбанк	Псков	195 050	193 346	22 050	22 246	60	292	468
58	57	Морской коммерческий банк	Калининград	160 065	118 519	31 064	29 498	57	1 929	2 365
59	60	Муниципальный коммерческий банк	Калининград	143 391	94 789	53 343	54 275	54	808	310
60	59	Северный Кредит	Архангельск	120 576	105 985	25 685	25 153	59	1 481	2 198
61	61	Станкобанк	Рошино	42 463	81 253	14 009	13 295	61	503	279
62	62	Промышленность и финансы	Санкт-Петербург	33 961	27 530	10 510	10 493	62	457	211
* Без учета субординированных кредитов.										

Источник: Эксперт Северо-Запад. — 2006. — № 37. — С. 31–32.

Страны по сводным показателям дохода¹

Страны с высоким уровнем дохода (ВНД на душу населения в 2003 г. равен 9386 или более долл. США)	Страны со средним уровнем дохода (ВНД на душу населения в 2003 г. равен 766—9385 долл. США)	Страны с низким уровнем дохода (ВНД на душу населения в 2003 г. равен 765 или менее долл. США)
Австралия Австрия Андорра Багамы Бахрейн Бельгия Бруней Даруссалам Великобритания Германия Греция Гонконг (Китай, САР) Дания Израиль Ирландия Исландия Испания Италия Канада Катар Кипр Корея, Республика	Азербайджан Албания Алжир Антигуа и Барбуда Аргентина Армения Барбадос Белиз Беларусь, Республика Болгария Боливия Босния и Герцеговина Ботсвана Бразилия Вануату Венгрия Венесуэла, Боливарианская Рес- публика Габон	Ангола Афганистан Бангладеш Бенин Бутан Буркина-Фасо Бурунди Вьетнам Гаити Гамбия Гана Гвинея Гвинея-Бисау Замбия Зимбабве Индия Йемен Камбоджа Камерун Кения Киргизия
	Маршалловы Острова Мексика Микронезия, Федера- тивные Штаты Намибия Окупированные Палестинские Территории Оман Панама Парагвай Перу Польша Российская Федерация Румыния Сальвадор Самоа (Западное) Саудовская Аравия Свазиленд	Сьерра Леоне Таджикистан Танзания, Обье- диненная Респуб- лика Тимор — Лешти Того Уганда Узбекистан Центральная африканская Республика Чад Экваториальная Гвинея Эритрея Эфиопия (61 страна)

Кувейт	Гайана	Северные	КНДР
Люксембург	Гватемала	Марианские Острова	Коморские
Мальта	Гондурас	Сент-Винсент и Гренадины	Острова
Монако	Гренада	Сент-Китс и Невис	Конго
Нидерланды	Грузия	Сент-Люсия	Конго, Демократическая
Новая Зеландия	Джибути	Сейшельские Острова	Республика
Норвегия	Доминика	Сербия и Черногория	Кот-д'Ивуар
ОАЭ	Доминиканская Республика	Сирийская Арабская Республика	Лаосская НДР
Португалия	Египет, Арабская Республика	Словацкая Республика	Лесото
Сан-Марино	Индонезия	Суринам	Либерия
Сингапур	Иордания	Таиланд	Мавритания
Словения	Иран, Исламская Республика	Тонго	Малагаскар
США	Ирак	Тринидад и Тобаго	Малави
Финляндия	Кабо-Верде	Тунис	Мали
Франция	Казахстан	Туркменистан	Мозамбик
Швейцария	Кирибати	Турция	Молдова, Республика
Швеция	Китай	Украина	Монголия
Япония	Колумбия	Уругвай	Мьянма
(39 стран)	Коста-Рика	Фиджи	Непал
	Куба	Филиппины	Нигер
	Латвия	Хорватия	Нигерия
	Ливан	Чешская Республика	Никарагуа
	Ливийская Арабская Республика	Чили	Пакистан
	Джамаика	Шри-Ланка	Папуа-Новая Гвинея
	Литва	Эквадор	Руанда
	Маврикий	Эстония	Сан-Томе и Принсипи
	Македония, БЮР	ЮАР	Сенегал
	Малайзия	Ямайка	Соломоновы Острова
	Мальдивские Острова	(91 страна)	Сомали
	Марокко		Судан

¹ Классификация Всемирного банка (действительна на 1 июля 2004 г.) основывается на валовом национальном доходе (ВНД) на душу населения.

Страны по мировым сводным показателям

Развивающиеся страны					Центральная и Восточная Европа и Содружество Независимых Государств (СНГ)	Организация экономического сотрудничества и развития (ОЭСР)
Антигуа и Барбуда	Иран, Исламская Республика	Нигер	Филиппины	Азербайджан	Австралия	
Алжир	Йемен	Нигерия	Центрально-африканская Республика	Албания	Австрия	
Ангола	Кабо-Верде	Никарагуа	Чад	Армения	Бельгия	
Аргентина	Камбоджа	ОАЭ	Республика	Беларусь, Республика	Великобритания	
Афганистан	Катар	Окупируемые Палестинские Территории	Чили	Болгария	Венгрия	
Багамские Острова	Кения	Оман	Шри-Ланка	Босния и Герцеговина	Германия	
Бангладеш	Камерун	Пакистан	Экватор	Венгрия	Греция	
Барбадос	Кипр	Палау	Экваториальная Гвинея	Грузия	Дания	
Бахрейн	Кирибати	Панама	Эритрея	Казахстан	Ирландия	
Белиз	Китай	Папуа-Новая Гвинея	Эфиопия	Киргизия	Исландия	
Бенин	КНДР	Парагвай	ЮАР	Латвия	Испания	
Боливия	Колумбия	Перу	Ямайка	Литва	Италия	
Ботсвана	Коморские Острова	Руанда	(167 стран)	Македония, БЮР	Канада	
Бразилия	Конго	Сальвадор		Молдова,	Корея, Республика	
Бруней Даруссалам	Конго, Демократическая Республика	Самоа (Западное)		Республика	Люксембург	
Буркина Фасо	Корея, Республика	Сан-Томе и Принсипи		Польша	Мексика	
Бурунди	Коста-Рика	Саудовская Аравия		Российская Федерация	Нидерланды	
Бутан		Свазиленд		Румыния	Новая Зеландия	
Вануату					Норвегия	
Венесуэла, Боливарианская Республика						

Вьетнам Габон Гамбия Гана Гаити Гватемала Гайана Гвинея Гвинея-Бисау Гондурас Гонконг (Китай, САР) Джибути Доминика Доминиканская Республика Египет, Арабская Республика Замбия Индия Индонезия Иордания Ирак	Кот-д'Ивуар Куба Кувейт Лаосская НДР Лесото Либерия Ливан Ливийская Арабская Джамахирия Маврикий Мавритания Мадagasкар Малави Малайзия Мали Мальдивские Острова Марокко Маршалловы Острова Мексика Микронезия, Федеративные Штаты Мозамбик Монголия Мьянма Намибия Науру Непал	Сент-Винсент и Гренадины Сент-Китс и Невис Сент-Люсия Сейшельские Острова Сенегал Сингапур Сирийская Араб- ская Республика Соломоновы Острова Сомали Судан Суринам Сьерра-Леоне Танзания, Объединенная Республика Таиланд Тимор-Лешти Того Тонго Тринидад и Тобаго Тувалу Тунис Турция Уганда Уругвай Фиджи	Сербия и Черногория Словацкая Республика Словения Таджикистан Туркменистан Узбекистан Украина Хорватия Чешская Республика Чешская Республика Эстония (27 стран)	Польша Португалия Словацкая Республика США Турция Финляндия Франция Чешская Республика Швейцария Швеция Япония (30 стран)
---	--	--	--	--

Развивающиеся страны по мировым сводным показателям

Арабские государства	Восточная Азия и Тихоокеанский регион	Южная Азия	Латинская Америка и Карибский регион	Страны Африки к югу от Сахары
Алжир	Бруней	Афганистан	Антигуа и Барбуда	Ангола
Бахрейн	Даруссалам	Бангладеш	Аргентина	Бенин
Иордания	Вануату	Бутан	Багамские Острова	Ботсвана
Ирак	Вьетнам	Индия	Барбадос	Буркина-Фасо
Йемен	Гонконг (Китай)	Иран, Исламская Республика	Белиз	Бурунди
Джибути	Индонезия	Мальдивские Острова	Боливия	Габон
Египет, Арабская Республика	Камбоджа	Непал	Бразилия	Гамбия
Катар	Кирибати	Пакистан	Венесуэла, Боливарианская Республика	Гана
Кувейт	Китай	Шри-Ланка (9 стран)	Гаити	Гвинея
Ливан	КНДР		Гайана	Гвинея-Бисау
Ливийская Арабская Джамахирия	Корея, Республика		Гватемала	Замбия
Марокко	Лаосская НДР		Гондурас	Зимбабве
ОАЭ	Малайзия		Гренада	Камерун
Окупируемые Палестинские Территории	Маршалловы Острова		Доминика	Кабо-Верде
Оман	Микронезия, Федеративные Штаты	Южная Европа	Доминиканская Республика	Кения
Саудовская Аравия	Монголия	Кипр	Колумбия	Коморские Острова
Сирийская Арабская Республика	Маньяма	Турция (2 страны)	Коста-Рика	Конго
Сомали	Науру		Куба	Конго, Демократическая Республика
	Палау		Мексика	Кот-д'Ивуар
	Папуа-Новая Гвинея		Никарагуа	Лесото
	Самоа (Западное)		Панама	Либерия
				Малагаскар
				Маврикий

<p>Судан Тунис (20 стран)</p>	<p>Сингапур Соломоновы Острова Таиланд Тимор-Лешти Тонга Тувалу Фиджи Филиппины (28 стран)</p>		<p>Парагвай Перу Сальвадор Сент-Винсент и Гренадины Сент-Китс и Невис Сент-Люсия Суринам Тринидад и Тобаго Уругвай Чили Эквадор Ямайка (33 страны)</p>	<p>Мавритания Малави Мали Мозамбик Намбия Нигер Нигерия Руанда Сан-Томе и Принсипи Свазиленд Сейшельские Острова Сенегал Сьерра-Леоне Танзания, Объединенная Республика Того Уганда Центральноафрика нская Республика Чад Экваториальная Гвинея Эритрея Эфиопия ЮАР (45 стран)</p>
---------------------------------------	--	--	--	--

СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

Таблица П2.1

Значение интеграла вероятностей $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-z}^{+z} e^{-\frac{z^2}{2}} dz$

z	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0718
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3752
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4909	4971	5035	5098
0,7	5161	5223	5285	5346	5467	5497	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6626	6679	6729	6778
1,0	6817	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8788	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9089
1,7	9108	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9425	9438	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9608	9615	9625	9634
2,1	9643	9652	9660	9669	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9755	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9866	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9924	9926	9927	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	99730	99739	99747	99755	99763	99771	99779	99786	99793	99800
3,1	99807	99813	99819	99825	99831	99837	99842	99847	99853	99858
3,2	99863	99867	99872	99876	99880	99884	99888	99892	99896	99900
3,3	99903	3,6	99911	3,9	999904	4,4	9999892	5,0		99999943
3,4	99933	3,7	99937	4,0	999937	4,6	9999957	5,0		99999996
3,5	99953	3,8	99957	4,2	999973	4,8	9999984	6,0		999999998

Распределение Стьюдента

(значения t для $A_{\infty}^t = \varepsilon$ при df степеней свободы)

ε df	.95	.975	.99	.995	.9995
1	6.314	12.706	31.821	63.657	636.0619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.941
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.859
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.988	3.449	5.405
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

Значение F -критерия Фишера

df_2	df_1								
	1	2	3	4	5	6	7	8	
1	161	200	216	225	230	234	237	239	
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	
3	10,13	9,55	9,28	9,19	9,01	8,94	8,88	8,84	
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	
25	4,24	3,88	2,99	2,76	2,60	2,49	2,41	2,34	
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	

Примечание. df_1 — число степеней свободы для большей дисперсии; df_2 — число степеней свободы для меньшей дисперсии.

Таблица П2.3

при урвне значимости 0,05

$d.f._1$									
	9	10	11	12	14	16	20	30	∞
241	241	242	243	244	245	246	248	250	254
19,38	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,46	19,50
8,81	8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,53
6,00	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,63
4,78	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,36
4,10	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,67
3,68	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,23
3,39	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,08	2,93
3,18	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,71
3,02	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,54
2,90	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,40
2,80	2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,30
2,72	2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,21
2,65	2,65	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,13
2,59	2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,07
2,54	2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,01
2,50	2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,15	1,96
2,46	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,11	1,92
2,43	2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,07	1,88
2,40	2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,84
2,37	2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,81
2,35	2,35	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,78
2,32	2,32	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,76
2,30	2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,73
2,26	2,26	2,24	2,20	2,16	2,11	2,06	2,00	1,92	1,71
2,27	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,69
2,25	2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,67
2,24	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,65
2,22	2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,64
2,21	2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,62
2,12	2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,51
2,07	2,07	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,44
2,04	2,04	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,39
1,97	1,97	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,28
1,88	1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,46	1,00

Таблица П2.4

Значение χ^2 -критерия Пирсона при уровне значимости 0,10; 0,05; 0,01

<i>d.f.</i>	0,10	0,05	0,01	<i>d.f.</i>	0,10	0,05	0,01
1	2,71	3,84	6,63	21	29,62	32,67	38,93
2	4,61	5,99	9,21	22	30,81	33,92	40,29
3	6,25	7,81	11,34	23	32,01	35,17	41,64
4	7,78	9,49	13,28	24	33,20	36,42	42,98
5	9,24	11,07	15,09	25	34,38	37,65	44,31
6	10,64	12,59	16,81	26	35,56	38,89	45,64
7	12,02	14,07	18,48	27	36,74	40,11	46,96
8	13,36	15,51	20,09	28	37,92	41,34	48,28
9	14,68	16,92	21,67	29	39,09	42,56	49,59
10	15,99	18,31	23,21	30	40,26	43,77	50,89
11	17,28	19,68	24,72	40	51,80	55,76	63,69
12	18,55	21,03	26,22	50	63,17	67,50	76,15
13	19,81	22,36	27,69	60	74,40	79,08	88,38
14	21,06	23,68	29,14	70	85,53	90,53	100,42
15	22,31	25,00	30,58	80	96,58	101,88	112,33
16	23,54	26,30	32,00	90	107,56	113,14	124,12
17	24,77	27,59	33,41	100	118,50	124,34	135,81
18	25,99	28,87	34,81				
19	27,20	30,14	36,19				
20	28,41	31,14	37,57				

Таблица П2.5

**Критические значения коэффициентов корреляции для уровней
значимости 0,05; 0,01**

<i>d.f.</i>	$\alpha = 0,05$	$\alpha = 0,01$	<i>d.f.</i>	$\alpha = 0,05$	$\alpha = 0,01$
1	0,996917	0,9998766	17	0,4555	0,5751
2	0,950000	0,990000	18	0,4438	0,5614
3	0,8783	0,95873	19	0,4329	0,5487
4	0,8114	0,91720	20	0,4227	0,5368
5	0,7545	0,8745	25	0,3809	0,4869
6	0,7067	0,8343	30	0,3494	0,4487
7	0,6664	0,7977	35	0,3246	0,4182
8	0,6319	0,7646	40	0,3044	0,3932
9	0,6021	0,7348	45	0,2875	0,3721
10	0,5760	0,7079	50	0,2732	0,3541
11	0,5529	0,6835	60	0,2500	0,3248
12	0,5324	0,6614	70	0,2919	0,3017
13	0,5139	0,6411	80	0,2172	0,2830
14	0,4973	0,6226	90	0,2050	0,2673
15	0,4821	0,6055	100	0,1946	0,2540
16	0,4683	0,5897			

Для простой корреляции *d.f.* на 2 меньше, чем число пар вариантов; в случае частной корреляции необходимо также вычесть число исключаемых переменных.

Таблица П2.6

Z-преобразование. Значение величины *z* для значений *r*

<i>r</i>	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
0,5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
0,7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
0,8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Таблица случайных чисел

Ряд	Колонка							
	12345	67890	12345	67890	12345	67890	12345	67890
01	66194	28926	99547	16625	45515	67953	12108	57846
02	78240	43195	24837	32511	70880	22070	52622	61881
03	00833	88000	67299	68215	11274	55624	32991	17436
04	12111	86683	61270	58036	64192	90611	15145	01748
05	47189	99951	05755	03834	43782	90599	40282	51417
06	76396	72486	62423	27618	84184	78922	73561	52818
07	46409	17469	32483	09083	76175	19985	26309	91536
08	74626	22111	87286	46772	42243	68046	44250	42439
09	34450	81974	93723	49023	58432	67083	36876	93391
10	36327	72135	33005	28701	34710	49359	50693	89311
11	74185	77536	84825	09934	99103	09325	67389	45869
12	12296	41623	62873	37943	25584	09609	63360	47270
13	90822	60280	88925	99610	42772	60561	76873	04117
14	72121	79152	96591	90305	10189	79778	68016	13747
15	95268	41377	25684	08151	61816	58555	54305	86189
16	92603	09091	75884	93424	72586	88903	30061	14457
17	18813	90291	05275	01223	79607	95426	34900	09778
18	38840	26903	28624	67157	51986	42865	14508	49315
19	05959	33836	53758	16562	41081	38012	41230	20528
20	85141	21155	99212	32685	51403	31926	69813	58781
21	75047	59643	31074	38172	03718	32119	69506	67143
22	30752	95260	68032	62871	58781	34143	68790	69766
23	22986	82575	42187	62295	84295	30634	66562	31442
24	99439	86692	90348	66036	48399	73451	26698	39437
25	20389	93029	11881	71685	65452	89047	63669	02656
26	39249	05173	68256	36359	20250	68686	05947	09335
27	96777	33605	29481	20063	09398	01843	35139	61344
28	04860	32918	10798	50492	52655	33359	94713	28393
29	41613	42375	00403	03656	77580	87772	86877	57085
30	17930	00794	53836	53692	67135	98102	61912	11246
31	24649	31845	25736	75231	83808	98917	93829	99430
32	79899	34061	54308	59358	56462	58166	97302	86828
33	76801	49594	81002	30397	52728	15101	72070	33706

Ряд	Колонка							
	12345	67890	12345	67890	12345	67890	12345	67890
34	36239	63636	38140	65731	39788	06872	38971	53363
35	07392	64449	17886	63632	53995	17574	22247	62607
36	67133	04181	33874	98835	67453	59734	76381	63455
37	77759	31504	32832	70861	15152	29733	75371	39174
38	85992	72268	42920	20810	29361	51423	90306	73574
39	79553	75952	54116	65553	47139	60579	09165	85490
40	41101	17336	48951	53674	17880	45260	08575	49321
41	36191	17095	32123	91576	84221	78902	82010	30874
42	62329	63898	23268	74283	26091	68409	69704	82267
43	14751	13151	93115	01437	56945	89661	67680	79790
44	48462	59278	44185	29616	76537	19589	83139	28454
45	29435	88105	59651	44391	74588	55114	80834	85686
46	28340	29285	12965	14821	80425	16602	44653	70467
47	02167	58940	27149	80242	10587	79786	34959	75339
48	17864	00991	39557	54981	23588	81914	37609	13128
49	79675	80605	60059	35862	00254	36546	21545	78179
50	72335	82037	92003	34100	29879	46613	89720	13274
51	49280	88924	35779	00283	81163	07275	89863	02348
52	61870	41657	07468	08612	98083	97349	20775	45091
53	43898	65923	25078	86129	78496	97653	91550	08078
54	62993	93912	30454	84598	56095	20664	12872	64647
55	33850	58555	51438	85507	71865	79488	76783	31708
56	55336	71264	88472	04334	63919	36394	11095	92470
57	70543	29776	10087	10072	55980	64688	68239	20461
58	89382	93809	00796	95945	34101	81277	66090	88872
59	37818	72142	67140	50785	22380	16703	53362	44940
60	60430	22834	14130	96593	23298	56203	92671	15925
61	82975	66158	84731	19436	55790	69229	28661	13675
62	39087	71938	40355	54324	08401	26299	49420	59208
63	55700	24586	93247	32596	11865	63397	44251	43189
64	14756	23997	78643	75912	83832	32768	18928	57070
65	32166	53251	70654	92827	63491	04233	33825	69662
66	23236	73751	31888	81718	06546	83246	47651	04877
67	45794	26926	15130	82455	78305	55058	52551	47182

Ряд	Колонка							
	12345	67890	12345	67890	12345	67890	12345	67890
68	09893	20505	14225	68514	46427	56788	96297	78822
69	54382	74598	91499	14523	68479	27686	46162	83554
70	94750	89923	37089	20048	80336	94598	26940	36858
71	70297	34135	53140	33340	42050	82341	44104	82949
72	85157	47954	32979	26575	57600	40881	12250	73742
73	11100	02340	12860	74697	96644	89439	28707	25815
74	36871	50775	30592	57143	17381	68856	25853	35041
75	23913	48357	63308	16090	51690	54607	72407	55538
76	79348	36085	27973	65157	07456	22255	25626	57054
77	92074	54641	53673	54421	18130	60103	69593	49464
78	06873	21440	75593	41373	49502	17972	82578	16364
79	12478	37622	99659	31065	83613	69889	58869	29571
80	57175	55564	65411	42547	70457	03426	72937	83792
81	91616	11075	80103	07831	59309	13276	26710	73000
82	78025	73539	14621	39044	47450	03197	12787	47709
83	27587	67228	80145	10175	12822	86687	65530	49325
84	16690	20427	04251	64477	73709	73945	92396	68263
85	70183	58065	65489	31833	82093	16747	10386	59293
86	90730	35385	15679	99742	50866	78028	75573	67257
87	10934	93242	13431	24590	02770	48582	00906	58595
88	82462	30166	79613	47416	13389	80268	05085	96666
89	27463	10433	07606	16285	93699	60912	94532	95632
90	02979	52997	09079	92709	90110	47506	53693	49892
91	46888	69929	75233	52507	32097	37594	10067	67327
92	53638	83161	08289	12639	08141	12640	28437	09268
93	82433	61427	17239	89160	19666	08814	37841	12847
94	35766	31672	50082	22795	66948	65581	84393	15890
95	10853	42581	08792	13257	61973	24450	52351	16602
96	20341	27398	72906	63955	17276	10646	74692	48438
97	54458	90542	77563	51839	52901	53355	83281	19177
98	26337	66530	16687	35179	46560	00123	44546	79896
99	34314	23729	85264	05575	96855	23820	11091	79821
00	28603	10708	68933	34189	92166	15181	66628	58599

Таблица П2.8

Значения функции Пуассона: $\frac{\lambda^m}{m!} e^{-\lambda}$

$m \backslash \lambda$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,9048	0,8187	0,7408	0,6703	0,6055	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1638	0,2222	0,2681	0,3033	0,3293	0,3476	0,3596	0,3696
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,001	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4	—	—	0,0002	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5	—	—	—	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6	—	—	—	—	—	—	0,0001	0,0002	0,0003
$m \backslash \lambda$	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0055
3	0,0313	0,1804	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337
5	0,0081	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607
6	0,0006	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0,0001	0,0034	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1318
8	—	0,0009	0,0081	0,0298	0,0656	0,1033	0,1304	0,1396	0,1318
9	—	0,0002	0,0027	0,0132	0,0363	0,0688	0,1014	0,1241	0,0318
10	—	—	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1180
11	—	—	0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970
12	—	—	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728
13	—	—	—	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14	—	—	—	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15	—	—	—	—	0,0002	0,0009	0,0033	0,0090	0,0194
16	—	—	—	—	—	0,0003	0,0014	0,0045	0,0109
17	—	—	—	—	—	0,0001	0,0006	0,0021	0,0058
18	—	—	—	—	—	—	0,0002	0,0009	0,0029
19	—	—	—	—	—	—	0,0001	0,0004	0,0014
20	—	—	—	—	—	—	—	0,0002	0,0006
21	—	—	—	—	—	—	—	0,0001	0,0003
22	—	—	—	—	—	—	—	—	0,0001

Таблица П2.9

Критические значения D-критерия Колмогорова—Смирнова

Объем выборки	Уровень значимости, α				
	0,20	0,15	0,10	0,05	0,01
$n = 1$	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
Свыше 35	1,07 $\frac{1}{\sqrt{n}}$	1,14 $\frac{1}{\sqrt{n}}$	1,22 $\frac{1}{\sqrt{n}}$	1,36 $\frac{1}{\sqrt{n}}$	1,63 $\frac{1}{\sqrt{n}}$

Нижние и верхние значения критерия знаков Вилкоксона (W)

Двусторонняя проверка	$\alpha = 0,20$	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,02$	$\alpha = 0,01$
Односторон- няя проверка	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,025$	$\alpha = 0,01$	$\alpha = 0,005$
$n = 4$	1,9	0,10	0,10	0,10	0,10
5	3,12	1,14	0,15	0,15	0,15
6	4,17	3,18	1,20	0,21	0,21
7	6,22	4,24	3,25	1,27	0,28
8	9,27	6,30	4,32	2,34	1,35
9	11,34	9,36	6,39	4,41	2,43
10	15,40	11,44	9,46	6,49	4,51
11	18,48	14,52	11,55	8,58	6,60
12	22,56	18,60	14,64	10,68	8,70
13	27,64	22,69	18,73	13,78	10,81
14	32,73	26,79	22,83	16,89	13,92
15	37,83	31,89	26,94	20,100	16,104
16	43,93	36,100	30,106	24,112	20,116
17	49,104	42,111	35,118	28,125	24,129
18	56,115	48,123	41,130	33,138	28,143
19	63,127	54,136	47,143	38,152	33,157
20	70,140	61,149	53,157	44,166	38,172

Таблица П2.11

**Нижние и верхние критические значения критерия
суммы рангов Вилкоксона**

$\alpha = 0,025$ (односторонняя) или $\alpha = 0,05$ (двусторонняя)																	
$n_1 : 3$			4		5		6		7		8		9		10		
$n_2 : 3$	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33	
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44	
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56	
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70	
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83	
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98	
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114	
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131	
$\alpha = 0,05$ (односторонняя), $\alpha = 0,10$ (двусторонняя)																	
$n_1 : 3$			4		5		6		7		8		9		10		
$n_2 : 3$	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31	
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42	
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	4	
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67	
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80	
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95	
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111	
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127	

Примечание. $n_1 \leq n_2$.

Таблица П2.12

**Критические значения коэффициента ранговой корреляции Спирмена
для правосторонней проверки при α**

Объем выборки	0,10	0,05	0,025	0,01	0,005	0,001
$n = 4$	0,8000	0,8000				
5	0,7000	0,8000	0,9000	0,9000		
6	0,6000	0,7714	0,8286	0,8857	0,9429	
7	0,5357	0,6786	0,7450	0,8571	0,8929	0,9643
8	0,5000	0,6190	0,7143	0,8095	0,8571	0,9286
9	0,4667	0,5833	0,6833	0,7667	0,8167	0,9000
10	0,4424	0,5515	0,6364	0,7333	0,7818	0,8667
11	0,4182	0,5273	0,6091	0,7000	0,7455	0,8364
12	0,3986	0,4965	0,5804	0,6713	0,7273	0,8182
13	0,3791	0,4780	0,5549	0,6429	0,6978	0,7912
14	0,3626	0,4593	0,5341	0,6220	0,6747	0,7670
15	0,3500	0,4429	0,5179	0,6000	0,6536	0,7464
16	0,3382	0,4265	0,5000	0,5824	0,6324	0,7265
17	0,3260	0,4118	0,4853	0,5637	0,6152	0,7083
18	0,3148	0,3994	0,4716	0,5480	0,5975	0,6904
19	0,3070	0,3895	0,4579	0,5333	0,5825	0,6737
20	0,2977	0,3789	0,4451	0,5203	0,5684	0,6586
21	0,2909	0,3688	0,4351	0,5078	0,5545	0,6455
22	0,2829	0,3597	0,4241	0,4963	0,5426	0,6318
23	0,2767	0,3518	0,4150	0,4852	0,5306	0,6186
24	0,2704	0,3435	0,4061	0,4748	0,5200	0,6070
25	0,2646	0,3362	0,3977	0,4654	0,5100	0,5962
26	0,2588	0,3299	0,3894	0,4564	0,5002	0,5856
27	0,2540	0,3236	0,3822	0,4481	0,4915	0,5757
28	0,2490	0,3175	0,3749	0,4401	0,4828	0,5660
29	0,2443	0,3113	0,3685	0,4320	0,4744	0,5567
30	0,2400	0,3059	0,3620	0,4251	0,4665	0,5479

Таблица П2.13

Величины – $p\log_2 p$

p	0	1	2	3	4	5	6	7	8	9
0,00	—	0,0100	0,0179	0,0251	0,0319	0,0382	0,0443	0,0501	0,0557	0,0612
0,01	0,0664	0,0716	0,0766	0,0845	0,0862	0,0909	0,0955	0,0999	0,1043	0,1086
0,02	0,1129	0,1170	0,1211	0,1252	0,1291	0,0330	0,1369	0,1407	0,1444	0,1481
0,03	0,1518	0,1554	0,1589	0,1624	0,1659	0,1693	0,1727	0,1760	0,1793	0,1825
0,04	0,1858	0,1889	0,1921	0,1952	0,1983	0,2013	0,2043	0,2073	0,2103	0,2132
0,05	0,2161	0,2190	0,2218	0,2246	0,2274	0,2301	0,2329	0,2356	0,2383	0,2409
0,06	0,2435	0,2461	0,2487	0,2513	0,2538	0,2563	0,2588	0,2613	0,2637	0,2661
0,07	0,2686	0,2709	0,2733	0,2756	0,2780	0,2803	0,2826	0,2848	0,2871	0,2893
0,08	0,2915	0,2937	0,2959	0,2980	0,3002	0,3023	0,3044	0,3065	0,3086	0,3106
0,09	0,3127	0,3147	0,3167	0,3187	0,3207	0,3226	0,3246	0,3265	0,3284	0,3303
0,10	0,3322	0,3341	0,3359	0,3378	0,3398	0,3414	0,3432	0,3450	0,3468	0,3485
0,11	0,3503	0,3520	0,3537	0,3555	0,3571	0,3588	0,3605	0,3622	0,3638	0,3654
0,12	0,3671	0,3687	0,3703	0,3719	0,3734	0,3750	0,3766	0,3781	0,3796	0,3811
0,13	0,3826	0,3841	0,3856	0,3871	0,3886	0,3900	0,3915	0,3929	0,3943	0,3957
0,14	0,3971	0,3985	0,3999	0,4012	0,4026	0,4040	0,4053	0,4066	0,4079	0,4092
0,15	0,4105	0,4118	0,4131	0,4144	0,4156	0,4169	0,4181	0,4194	0,4206	0,4218
0,16	0,4230	0,4242	0,4254	0,4266	0,4277	0,4289	0,4301	0,4312	0,4323	0,4335
0,17	0,4346	0,4357	0,4368	0,4379	0,4390	0,4400	0,4411	0,4422	0,4432	0,4443
0,18	0,4453	0,4463	0,4474	0,4484	0,4494	0,4504	0,4514	0,4523	0,4533	0,4543
0,19	0,4552	0,4562	0,4571	0,4581	0,4590	0,4599	0,4608	0,4617	0,4626	0,4635
0,20	0,4644	0,4653	0,4661	0,4670	0,4678	0,4687	0,4695	0,4704	0,4712	0,4720
0,21	0,4728	0,4736	0,4744	0,4752	0,4760	0,4768	0,4776	0,4783	0,4791	0,4798
0,22	0,4806	0,4813	0,4820	0,4828	0,4835	0,4842	0,4849	0,4856	0,4863	0,4870
0,23	0,4877	0,4883	0,4890	0,4897	0,4903	0,4910	0,4916	0,4923	0,4949	0,4935
0,24	0,4941	0,4947	0,4954	0,4960	0,4966	0,4971	0,4977	0,4983	0,4989	0,4994
0,25	0,5000	0,5006	0,5011	0,5016	0,5022	0,5027	0,5032	0,5038	0,5043	0,5048
0,26	0,5053	0,5058	0,5063	0,5068	0,5072	0,5077	0,5082	0,5087	0,5091	0,5096
0,27	0,5100	0,5105	0,5109	0,5113	0,5118	0,5122	0,5126	0,5130	0,5134	0,5138
0,28	0,5142	0,5146	0,5150	0,5154	0,5158	0,5161	0,5165	0,5169	0,5172	0,5176
0,29	0,5179	0,5182	0,5186	0,5189	0,5192	0,5196	0,5199	0,5202	0,5205	0,5208
0,30	0,5211	0,5214	0,5217	0,5220	0,5222	0,5225	0,5228	0,5230	0,5233	0,5235

p	0	1	2	3	4	5	6	7	8	9
0,31	0,5238	0,5240	0,5243	0,5245	0,5247	0,5250	0,5252	0,5254	0,5256	0,5258
0,32	0,5260	0,5262	0,5264	0,5266	0,5268	0,5270	0,5272	0,5273	0,5275	0,5277
0,33	0,5278	0,5280	0,5281	0,5283	0,5284	0,5286	0,5287	0,5288	0,5289	0,5290
0,34	0,5292	0,5293	0,5294	0,5295	0,5296	0,5297	0,5298	0,5299	0,5299	0,5300
0,35	0,5301	0,5302	0,5302	0,5303	0,5304	0,5304	0,5305	0,5305	0,5305	0,5306
0,36	0,5306	0,5306	0,5307	0,5307	0,5307	0,5307	0,5307	0,5307	0,5307	0,5307
0,37	0,5307	0,5307	0,5307	0,5307	0,5307	0,5306	0,5306	0,5306	0,5305	0,5305
0,38	0,5304	0,5304	0,5303	0,5303	0,5302	0,5302	0,5301	0,5300	0,5300	0,5299
0,39	0,5298	0,5297	0,5296	0,5295	0,5294	0,5293	0,5292	0,5291	0,5290	0,5289
0,40	0,5288	0,5286	0,5285	0,5284	0,5283	0,5281	0,5280	0,5278	0,5277	0,5275
0,41	0,5274	0,5272	0,5271	0,5269	0,5267	0,5266	0,5264	0,5262	0,5260	0,5258
0,42	0,5256	0,5255	0,5253	0,5251	0,5249	0,5246	0,5244	0,5242	0,5240	0,5238
0,43	0,5236	0,5233	0,5231	0,5229	0,5226	0,5224	0,5222	0,5219	0,5217	0,5214
0,44	0,5211	0,5209	0,5206	0,5204	0,5201	0,5198	0,5195	0,5193	0,5190	0,5187
0,45	0,5184	0,5181	0,5178	0,5175	0,5172	0,5169	0,5166	0,5163	0,5160	0,5157
0,46	0,5153	0,5150	0,5147	0,5144	0,5140	0,5137	0,5133	0,5130	0,5127	0,5123
0,47	0,5120	0,5116	0,5112	0,5109	0,5105	0,5102	0,5098	0,5094	0,5090	0,5087
0,48	0,5083	0,5079	0,5075	0,5071	0,5067	0,5063	0,5059	0,5055	0,5051	0,5047
0,49	0,5043	0,5039	0,5034	0,5030	0,5026	0,5022	0,5017	0,5013	0,5009	0,5004
0,50	0,5000	0,4996	0,4991	0,4987	0,4982	0,4978	0,4973	0,4968	0,4964	0,4959
0,51	0,4954	0,4950	0,4945	0,4940	0,4935	0,4930	0,4926	0,4921	0,4916	0,4911
0,52	0,4906	0,4901	0,4896	0,4891	0,4886	0,4880	0,4875	0,4870	0,4865	0,4860
0,53	0,4854	0,4849	0,4844	0,4839	0,4833	0,4828	0,4822	0,4817	0,4811	0,4806
0,54	0,4800	0,4795	0,4789	0,4784	0,4778	0,4772	0,4767	0,4761	0,4755	0,4750
0,55	0,4744	0,4738	0,4732	0,4726	0,4720	0,4714	0,4708	0,4702	0,4697	0,4691
0,56	0,4684	0,4678	0,4672	0,4666	0,4660	0,4654	0,4648	0,4641	0,4635	0,4629
0,57	0,4623	0,4616	0,4610	0,4603	0,4597	0,4591	0,4584	0,4578	0,4671	0,4565
0,58	0,4558	0,4551	0,4545	0,4538	0,4532	0,4525	0,4518	0,4512	0,4505	0,4498
0,59	0,4491	0,4484	0,4477	0,4471	0,4464	0,4457	0,4450	0,4443	0,4436	0,4429
0,60	0,4422	0,4415	0,4408	0,4401	0,4393	0,4386	0,4379	0,4372	0,4365	0,4357
0,61	0,4350	0,4343	0,4335	0,4328	0,4321	0,4313	0,4306	0,4298	0,4291	0,4283
0,62	0,4276	0,4268	0,4261	0,4253	0,4246	0,4238	0,4230	0,4223	0,4215	0,4207

Продолжение

p	0	1	2	3	4	5	6	7	8	9
0,63	0,4199	0,4192	0,4184	0,4176	0,4168	0,4160	0,4153	0,4145	0,4137	0,4129
0,64	0,4121	0,4113	0,4105	0,4097	0,4089	0,4080	0,4072	0,4064	0,4056	0,4048
0,65	0,4040	0,4032	0,4023	0,4015	0,4007	0,3998	0,3990	0,3982	0,3973	0,3965
0,66	0,3957	0,3948	0,3940	0,3931	0,3922	0,3914	0,3905	0,3897	0,3888	0,3880
0,67	0,3871	0,3862	0,3854	0,3845	0,3836	0,3828	0,3819	0,3810	0,3801	0,3792
0,68	0,3784	0,3775	0,3766	0,3757	0,3748	0,3739	0,3730	0,3721	0,3712	0,3703
0,69	0,3694	0,3685	0,3676	0,3666	0,3657	0,3648	0,3639	0,3630	0,3621	0,3611
0,70	0,3602	0,3593	0,3585	0,3574	0,3565	0,3555	0,3546	0,3536	0,3527	0,3518
0,71	0,3508	0,3499	0,3489	0,3480	0,3470	0,3461	0,3451	0,3441	0,3432	0,3422
0,72	0,3412	0,3403	0,3393	0,3383	0,3373	0,3364	0,3354	0,3344	0,3334	0,3324
0,73	0,3314	0,3304	0,3295	0,3285	0,3275	0,3265	0,3255	0,3245	0,3235	0,3225
0,74	0,3215	0,3204	0,3194	0,3184	0,3174	0,3164	0,3154	0,3144	0,3133	0,3123
0,75	0,3113	0,3103	0,3092	0,3082	0,3071	0,3061	0,3051	0,3040	0,3030	0,3019
0,76	0,3009	0,2999	0,2988	0,2978	0,2967	0,2956	0,2946	0,2935	0,2925	0,2914
0,77	0,2903	0,2893	0,2882	0,2871	0,2861	0,2850	0,2839	0,2828	0,2818	0,2807
0,78	0,2796	0,2785	0,2774	0,2763	0,2753	0,2741	0,2731	0,2720	0,2709	0,2698
0,79	0,2687	0,2676	0,2664	0,2653	0,2642	0,2631	0,2620	0,2609	0,2598	0,2587
0,80	0,2575	0,2564	0,2553	0,2542	0,2531	0,2519	0,2508	0,2497	0,2485	0,2474
0,81	0,2462	0,2451	0,2440	0,2428	0,2417	0,2405	0,2394	0,2382	0,2371	0,2359
0,82	0,2348	0,2336	0,2324	0,2313	0,2301	0,2292	0,2278	0,2266	0,2255	0,2243
0,83	0,2231	0,2220	0,2208	0,2196	0,2184	0,2172	0,2160	0,2149	0,2137	0,2125
0,84	0,2113	0,2101	0,2089	0,2077	0,2065	0,2053	0,2041	0,2029	0,2017	0,2005
0,85	0,1993	0,1981	0,1969	0,1957	0,1944	0,1932	0,1920	0,1908	0,1896	0,1884
0,86	0,1871	0,1859	0,1847	0,1834	0,1822	0,1810	0,1797	0,1785	0,1773	0,1760
0,87	0,1748	0,1735	0,1723	0,1711	0,1698	0,1686	0,1673	0,1661	0,1648	0,1635
0,88	0,1623	0,1610	0,1598	0,1585	0,1572	0,1560	0,1547	0,1534	0,1522	0,1509
0,89	0,1496	0,1484	0,1471	0,1458	0,1445	0,1432	0,1419	0,1407	0,1394	0,1381
0,90	0,1368	0,1355	0,1342	0,1329	0,1316	0,1303	0,1290	0,1277	0,1264	0,1251

Продолжение

p	0	1	2	3	4	5	6	7	8	9
0,91	0,1238	0,1225	0,1212	0,1199	0,1186	0,1173	0,1159	0,1146	0,1133	0,1120
0,92	0,1107	0,1094	0,1080	0,1067	0,1054	0,1040	0,1027	0,1014	0,1000	0,0987
0,93	0,0974	0,0960	0,0947	0,0933	0,0920	0,0907	0,0893	0,0880	0,0866	0,0853
0,94	0,0839	0,0826	0,0812	0,0798	0,0785	0,0771	0,0758	0,0744	0,0730	0,1717
0,95	0,0703	0,0689	0,0676	0,0662	0,0648	0,0634	0,0621	0,0607	0,0593	0,0579
0,96	0,0565	0,0552	0,0538	0,0524	0,0510	0,0496	0,0482	0,0468	0,0454	0,0440
0,97	0,0426	0,0412	0,0398	0,0384	0,0370	0,0356	0,0342	0,0328	0,0314	0,0300
0,98	0,0286	0,0271	0,0257	0,0243	0,0230	0,0214	0,0201	0,0186	0,0172	0,0158
0,99	0,0140	0,0129	0,0115	0,0101	0,0086	0,0072	0,0058	0,0043	0,0029	0,0014

ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ФОРМУЛЫ

Средние величины. Вариация

Средняя арифметическая простая: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$.

Средняя арифметическая взвешенная: $\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$.

Различие между простой и взвешенной средней:

$$\bar{x}_z = \bar{x} + \frac{n\sigma_x\sigma_zr_{xz}}{nz} = \bar{x} + \sigma_x\nu_zr_{xz}.$$

Средняя квадратическая величина: $\bar{x}_{\text{кв}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$.

Средняя геометрическая: $x_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$.

Средняя гармоническая: $\bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$.

Средняя степенная: $\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^n x_i^k}{n}}$.

Правило мажорантности средних:

$$\bar{x}_{\text{гарм}} \leq \bar{x}_{\text{геом}} \leq \bar{x}_{\text{арифм}} \leq \bar{x}_{\text{квадр}} \leq \bar{x}_{\text{куб}}$$

Число групп в вариационном ряду:

$$k \approx 1 + 3,32 \cdot \lg n = 1 + 1,44 \ln n.$$

$$\text{Медиана: } Me = x_e + \frac{i}{f_{Me}} \left(\frac{n}{2} - f_{Me-1} \right),$$

где Me — медиана;

x_e — нижняя граница интервала, в котором находится медиана;

i — величина интервала;

n — число наблюдений;

f_{Me} — частота в медианном интервале;

f_{Me-1} — накопленная частота в интервале, предшествующем медианному.

Мода распределения:

$$Mo = x_0 + \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} \cdot i,$$

где x_0 — нижняя граница модального интервала;

f_{Mo} — частота в модальном интервале;

f_{Mo-1} — частота в предыдущем интервале;

f_{Mo+1} — частота в следующем интервале за модальным;

i — величина интервала.

Размах (амплитуда) вариации: $R = x_{\max} - x_{\min}$.

Среднее линейное отклонение (для несгруппированных дан-

$$\text{ных): } a = \sum_{i=1}^n |x_i - \bar{x}| : n.$$

Среднее линейное отклонение (для сгруппированных дан-

$$\text{ных): } a = \frac{\sum_{j=1}^n |x'_j - \bar{x}| f_j}{\sum_{j=1}^n f_j}.$$

Среднее квадратическое отклонение (для несгруппирован-

$$\text{ных данных): } \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}.$$

Среднее квадратическое отклонение (для сгруппированных

данных): $\sigma = \sqrt{\frac{\sum_{j=1}^n (x'_j - \bar{x})^2 f_j}{\sum_{j=1}^n f_j}}.$

Дисперсия (для несгруппированных данных): $\sigma^2 = \frac{\sum_{(i)} (x_i - \bar{x})^2}{n}.$

Дисперсия (для сгруппированных данных):

$$\sigma^2 = \frac{\sum_{(j)} (x'_j - \bar{x})^2 f_j}{\sum_{(j)} f_j}, \text{ или } \sigma^2 = \overline{x^2} - \bar{x}^2$$

Среднее квартильное расстояние:

$$q = \frac{(Q_3 - \text{Me}) + (\text{Me} - Q_1)}{2} = \frac{Q_3 - Q_1}{2}.$$

Относительные показатели вариации:

коэффициент осцилляции: $\rho = R : \bar{x}$;

относительное отклонение по модулю: $m = a : \bar{x}$;

коэффициент вариации: $v = \sigma : \bar{x}$;

относительное квартильное расстояние: $d = q : \bar{x}.$

Центральные моменты

Порядок момента	Формула	
	по несгруппирован- ным данным	по сгруппированным данным
Первый (μ_1)	$\frac{\sum_{(i)} (x_i - \bar{x})}{n}$	$\frac{\sum_{(j)} (x'_j - \bar{x}) f_j}{\sum_{(j)} f_j}$
Второй (μ_2)	$\frac{\sum_{(i)} (x_i - \bar{x})^2}{n}$	$\frac{\sum_{(j)} (x'_j - \bar{x})^2 f_j}{\sum_{(j)} f_j}$

Порядок момента	Формула	
	по несгруппированным данным	по сгруппированным данным
Третий (μ_3)	$\frac{\sum_{(i)} (x_i - \bar{x})^3}{n}$	$\frac{\sum_{(j)} (x'_j - \bar{x})^3 f_j}{\sum_{(j)} f_j}$
Четвертый (μ_4)	$\frac{\sum_{(i)} (x_i - \bar{x})^4}{n}$	$\frac{\sum_{(j)} (x'_j - \bar{x})^4 f_j}{\sum_{(j)} f_j}$

Коэффициент асимметрии: $As = \frac{\mu_3}{\sigma^3}$.

Коэффициент асимметрии Пирсона: $As_P = \frac{\bar{x} - Mo}{\sigma}$.

Экссес: $Ex = \frac{\mu_4}{\sigma^4} - 3$.

Показатель эксцесса распределения (для несгруппированных

данных): $Ex = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n\sigma^4} - 3$.

Показатель эксцесса распределения (для сгруппированных данных):

$$Ex = \frac{\sum_{j=1}^n (x'_j - \bar{x})^4 f_j}{\sigma^4 \sum_{j=1}^n f_j} - 3.$$

Группировка

Средняя сила связи: $b_{yx} = \frac{\bar{y}_m - \bar{y}_1}{x'_m - x'_1}$,

где \bar{y}_m, \bar{y}_1 – средние значения результативного признака в последней и первой группах соответственно;

x'_m, x'_1 – середины интервалов (или средние значения) факторного признака в последней и первой группах.

Средняя величина внутригрупповой дисперсии:

$$\bar{\sigma}_{yx}^2 = \frac{\sum_1^m \sigma_j^2 n_j}{\sum_1^m n_j}.$$

Правило сложения дисперсий: $\sigma_y^2 = \bar{\sigma}_{yx}^2 + \sigma_{\bar{y}_x}^2$.

Коэффициент детерминации:

$$\eta^2 = \frac{\sigma_{yx}^2}{\sigma_y^2}.$$

Эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}.$$

Евклидово расстояние:

$$d_{p,q} = \sqrt{\sum_{j=1}^k (x_{jp} - x_{jq})^2}.$$

Выборочное наблюдение

$$s^2 = \frac{n-1}{n} \cdot \sigma^2.$$

Средняя ошибка выборочной средней:

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n-1}} = \frac{s}{\sqrt{n-1}}.$$

$$\Delta_{\bar{x}} = z s_{\bar{x}} = z \frac{s}{\sqrt{n}}.$$

Средняя ошибка выборочной доли:

$$s_p = \sqrt{\frac{p(1-p)}{n}}.$$

Предельная ошибка выборочной доли с принятой доверительной вероятностью:

$$\Delta_p = z \cdot s_p = z \sqrt{\frac{p(1-p)}{n}}.$$

Формулы средней ошибки выборочной средней и выборочной относительной величины

Вид выборки	Средняя ошибка	
	выборочной средней	выборочной относительной величины (доли)
Повторная – отбор единицами	$\sqrt{\frac{s^2}{n}}$	$\sqrt{\frac{p(1-p)}{n}}$
Бесповторная – отбор единицами	$\sqrt{\frac{s^2}{n} \left(\frac{N-n}{N-1} \right)}$	$\sqrt{\frac{p(1-p)}{n} \left(\frac{N-n}{N-1} \right)}$
Серийная (нерайонированная)	$\sqrt{\frac{s_x^2}{r} \left(\frac{R-r}{R-1} \right)}$	$\sqrt{\frac{s_p^2}{r} \left(\frac{R-r}{R-1} \right)}$
Районированная – отбор единицами, бесповторная	$\sqrt{\frac{\bar{s}^2}{n} \left(\frac{N-n}{N-1} \right)}$	$\sqrt{\frac{\bar{s}_p^2}{n} \left(\frac{N-n}{N-1} \right)}$
Районированная – отбор сериями, бесповторная	$\sqrt{\frac{\bar{s}_x^2}{r} \left(\frac{R-r}{R-1} \right)}$	$\sqrt{\frac{\bar{s}_p^2}{r} \left(\frac{R-r}{R-1} \right)}$

Объем выборки (повторный отбор):

$$n = \frac{t^2 \sigma^2}{\Delta^2},$$

где Δ — допустимая погрешность, которая задается исследователем исходя из требуемой точности результатов проектируемой выборки;

z — табличная величина, соответствующая заданной доверительной вероятности $F(z)$, с которой будут гарантированы оценки генеральной совокупности по данным выборочного обследования;

σ^2 — генеральная дисперсия.

Объем выборки (бесповторный отбор): $n = \frac{N t^2 \sigma^2}{N \Delta^2 + z^2 \sigma^2}.$

Средняя возможная ошибка для малой выборки: $s_{\bar{x}} = \sqrt{\frac{s_x^2}{n-1}}.$

Гипотезы

Критерий χ^2 :
$$\chi^2 = \sum_{j=1}^k \frac{(f_j - \mu_j)^2}{\mu_j},$$

где k — число категорий ряда распределения;

f_j — частота эмпирического распределения;

μ_j — частота теоретического распределения;

j — номер категории.

$$\chi^2 = n \sum_{j=1}^k \frac{(p_j - \pi_j)^2}{\pi_j},$$

где p_j — частности эмпирического распределения;

π_j — вероятности теоретического распределения.

Критерий Колмогорова-Смирнова: $D = \max |F_i - E_i|,$

где F_i — наблюдаемая кумулятивная частность для i -го значения (или интервала);

E_i — ожидаемая кумулятивная частность для i -го значения (или интервала).

t -критерий:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Критерий Фишера: $F = \frac{s_1^2}{s_2^2}$.

Критерий Вилкоксона: $W = \sum R^+$.

Регрессия

Уравнение парной регрессии: $\hat{y} = a + bx$,

где \hat{y} — среднее значение результативного признака y при определенном значении факторного признака x ;

a — свободный член уравнения;

b — коэффициент регрессии, изменяющий среднее отношение отклонения результативного признака от его средней величины к отклонению факторного признака от его средней величины на единицу его измерения, — вариация y , приходящаяся на единицу вариации x .

Система нормальных уравнений: $na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$;

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i.$$

b -коэффициент регрессии: $b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$;

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}.$$

Параметры уравнения регрессии: $a = \frac{\Delta a}{\Delta}$; $b_j = \frac{\Delta b_j}{\Delta}$,

где Δ — определитель системы;

Δ_a — частный определитель, получаемый в результате замены коэффициентов при a свободными членами из правой части системы уравнений;

Δ_b — частный определитель, получаемый в результате замены коэффициентов при b свободными членами из правой части системы уравнений.

Коэффициент корреляции:

$$r_{yx} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}};$$

$$r_{yx} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y};$$

$$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}.$$

Мера тесноты связи: $K_{\text{Фехнера}} = \frac{C - H}{C + H}$.

Средняя ошибка оценки:

$$m_b = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 : (n - 2)}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

где \hat{y}_i — расчетные значения результативного признака для i -й единицы;

$n-2$ — число степеней свободы.

t -критерий Стьюдента: $t = \frac{b}{m_b} = \frac{r^2}{m_r}$.

Средняя квадратическая ошибка коэффициента детерминации:

$$m_r = \sqrt{\frac{r^2(1-r^2)}{n-2}}.$$

Средняя ошибка положения линии регрессии в генеральной совокупности при значении факторного признака:

$$m_{y(x_k)} = s_{y_{\text{ост}}} \cdot \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Средняя ошибка прогноза для индивидуального значения по правилу дисперсии суммы для независимых переменных:

$$m_{y(x_k)} = \sqrt{m_{y(x_k)}^2 + s_{y_{\text{ост}}}^2}.$$

Общий вид многофакторного уравнения регрессии:

$$\hat{y} = a + b_1 x_1 + \dots + b_k x_k = a + \sum_{j=1}^k b_j x_j.$$

Стандартизированный коэффициент регрессии:

$$\beta_j = b_j \frac{\sigma_{x_j}}{\sigma_y}.$$

Коэффициент эластичности: $e_j = b_j \frac{\bar{x}_j}{\bar{y}}$.

Коэффициент раздельной детерминации: $d_j^2 = r_{x_j y} \beta_j$.

Частный коэффициент детерминации:

$$r_{yx_m(x_1 \dots x_{m-1} x_{m+1} \dots x_k)}^2 = \frac{R_y^2 - R_{yx_1 \dots x_{m-1} x_{m+1} \dots x_k}^2}{1 - R_{yx_1 \dots x_{m-1} x_{m+1} \dots x_k}^2}.$$

Средняя ошибка коэффициента множественной корреляции:

$$m_R = \sqrt{\frac{1 - R^2}{n - k - 1}}.$$

Линейное уравнение регрессии с фиктивными переменными:

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4u_1 + b_5u_2$$

Статистический анализ неколичественных переменных

Коэффициент ассоциации Q : $Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}$,

где n_{11} — число единиц, имеющих значения x_1 и y_1 ;
 n_{22} — число единиц, имеющих значения x_2 и y_2 ;
 n_{12} — число единиц, имеющих значения x_1 и y_2 ;
 n_{21} — число единиц, имеющих значения x_2 и y_1 .

Коэффициент контингенции:

$$\Phi = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{(n_{11} + n_{12})(n_{22} + n_{21})(n_{11} + n_{21})(n_{12} + n_{22})}} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{11}n_{22}n_{12}n_{21}}}.$$

Коэффициент взаимной сопряженности Пирсона:

$$P = \sqrt{\frac{\phi^2}{1 + \phi^2}}.$$

Коэффициент взаимной сопряженности Чупрова:

$$T = \sqrt{\frac{\phi^2}{\sqrt{(m-1)(p-1)}}} = \sqrt{\frac{\frac{\chi^2}{n}}{[(m-1)(p-1)]^{1/2}}}.$$

Коэффициент взаимной сопряженности Крамера:

$$V = \sqrt{\frac{\phi^2}{\min\{m-1, p-1\}}}.$$

Коэффициент корреляции рангов:

$$\rho_{R_x R_y} = \frac{\sum_{i=1}^n (R_{x_i} - \bar{R}_x)(R_{y_i} - \bar{R}_y)}{\sqrt{\sum_{i=1}^n (R_{x_i} - \bar{R}_x)^2 \sum_{i=1}^n (R_{y_i} - \bar{R}_y)^2}}.$$

Коэффициент корреляции рангов (Спирмена):

$$\rho_{R_x R_y} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}.$$

$$t\text{-критерий Стьюдента: } t = \rho \sqrt{\frac{n-2}{1-\rho^2}}.$$

$$\text{Коэффициент корреляции рангов Кендэла: } \tau = \frac{S}{\frac{1}{2}n(n-1)},$$

где S — фактическая сумма рангов;

$\frac{1}{2}n(n-1)$ — максимальная сумма рангов.

$$\text{Коэффициент конкордации: } W = \frac{S}{\frac{1}{12}[m^2(n^3 - n)]},$$

где S — сумма квадратов отклонений суммы рангов каждого объекта от средней суммы рангов;

m — количество порядковых переменных;

n — объем выборки.

Динамика

Абсолютное изменение уровней (цепное): $\Delta_i = y_i - y_{i-1}$.

Абсолютное изменение уровней (базисное): $\Delta_i = y_i - y_0$.

$$\text{Цепной темп роста: } k_{i/i-1} = \frac{y_i}{y_{i-1}}.$$

$$\text{Базисный темп роста: } k_{i/0} = \frac{y_i}{y_0}.$$

Соотношение между цепными и базисными показателями:

$$\sum_{(i)} \Delta_{i(\text{цепн})} = \Delta_{i(\text{базисн})};$$

$$\prod_{i=1}^n k_{i(\text{цепн})} = k_{i(\text{баз})}.$$

Средний уровень интервального ряда динамики:

$$\bar{y} = \frac{\sum_{(i)}^n y_i}{n}.$$

Хронологическая средняя:

$$\bar{y}_{\text{хрон}} = \left(\frac{y_1}{2} + \sum_{i=2}^{n-1} y_i + \frac{y_n}{2} \right) : (n-1).$$

Средний уровень (точные даты):

$$\bar{y} = \frac{\sum_{(i)} y_i t_i}{\sum t_i}.$$

Средний абсолютный прирост:

$$\bar{\Delta} = \frac{\sum \Delta_i}{n} = \frac{y_n - y_0}{n}.$$

Средний темп изменения:

$$\bar{k} = \sqrt[n]{\prod_{i=1}^n k_i} = \sqrt[n]{\frac{y_n}{y_0}}.$$

Среднее линейное отклонение:

$$a(t) = \frac{\sum_{i=1}^n |y_i - \bar{y}_i|}{n - p}.$$

Среднее квадратическое отклонение:

$$s(t) = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - p}},$$

где n — число уровней;
 y_i — фактический уровень;
 \bar{y}_i — выравненный уровень, тренд;
 p — число параметров тренда.

Коэффициент автокорреляции отклонений от тренда:

$$r_u^{a_1} = \frac{\sum_{i=1}^{n-1} u_i u_{i+1}}{\frac{u_1^2}{2} + \sum_{i=2}^{n-1} u_i^2 + \frac{u_n^2}{2}}.$$

Разложение уровня временного ряда: $y_i = T + S + C + E$.

Коэффициент корреляции по отклонениям от трендов

$$r_{u_x u_y} = \frac{\sum_{(i)} u_{x_i} u_{y_i}}{\sqrt{\sum_{(i)} u_{x_i}^2 \sum_{(i)} u_{y_i}^2}}.$$

Индексы

$$I_p = \frac{\sum_{(i)} q_{0j} p_{1j}}{\sum_{(j)} q_{0j} p_{0j}} - \text{индекс Ласпейреса};$$

$$I_p = \frac{\sum_{(i)} q_{1j} p_{1j}}{\sum_{(j)} q_{1j} p_{0j}} - \text{индекс Пааше};$$

$$I_q = \frac{\sum_{(i)} q_{1j} p_{0j}}{\sum_{(j)} q_{0j} p_{0j}} - \text{индекс физического объема};$$

$$I_{\text{структуры}} = \frac{\sum_{(i)} q_{1i} p_{0i}}{\sum_{(i)} q_{1i}} : \frac{\sum_{(i)} q_{0i} p_{0i}}{\sum_{(i)} q_{0i}}.$$

КОНТРОЛЬНЫЕ РАБОТЫ ПО ОСНОВНЫМ ТЕМАМ КУРСА

Тема: Средние величины

В а р и а н т 1

1. Имеются данные о движении работников на предприятиях пищевой промышленности в регионе за отчетный год.

№ п/п	Списочная численность работников на начало года, чел.	Темп роста списочной численности работников за год, %	Коэффициент оборота по приему, %	Коэффициент восполнения работников	Доля рабочих среди выбывших работников, %
1	251	105,6	7,8	3,3	66,7
2	164	95,0	7,5	0,6	75,0
3	526	107,3	9,9	3,4	43,8

Вычислите средние показатели движения работников в отрасли: средний темп роста численности, средний коэффициент оборота по приему, средний коэффициент восполнения работников, среднюю долю рабочих среди выбывших работников.

Показатели движения персонала предприятия рассчитываются по формулам:

- среднесписочная численность работников за год:

$$\bar{T}_{\text{сп}} = \frac{\text{Списочная численность на начало года} + \text{Списочная численность на конец года}}{2};$$

- коэффициент оборота по приему:

$$K_{\text{об. по приему}} = \frac{\text{Число принятых}}{\bar{T}_{\text{сп}}} 100\%;$$

- коэффициент восполнения работников:

$$K_{\text{восп}} = \frac{\text{Число принятых работников}}{\text{Число выбывших работников}}.$$

2. Мебельная фабрика за II квартал изготовила учебных столов-комплектов в апреле 250 шт., мае — 280 шт., июне — 270 шт. Ежемесячно часть комплектов изготовлялась по заказам с предоплатой: в апреле — 210 шт., мае — 223 шт., июне — 235 шт. Цена одного комплекта составляла в апреле 2,950 тыс. руб., мае — 2,985 тыс. руб., июне — 2,995 тыс. руб.

Рассчитайте среднемесячный выпуск учебных столов-комплектов, среднюю цену одного комплекта, средний процент выпуска комплектов с предоплатой во II квартале.

Оформите исходные данные и решение в виде таблицы.

Вариант 2

1. Известны данные о движении, состоянии и использовании основных фондов в двух филиалах целлюлозно-бумажного комбината.

Филиал	Полная стоимость основных фондов на начало года, млн руб.	Коэффициент динамики стоимости фондов за год	Удельный вес активной части фондов на конец года, %	Коэффициент износа на конец года, %	Фондоемкость продукции, руб./ руб.
1	58,6	1,169	72,5	23,2	0,45
2	71,2	1,059	68,7	37,5	0,40

Вычислите средние показатели, характеризующие движение, состояние и использование основных фондов в целом по комбинату: средний коэффициент динамики, средний удельный вес активной части фондов, средний коэффициент износа, средние фондоемкость и фондоотдачу.

Для построения исходных соотношений средних используйте следующие формулы:

- коэффициент динамики стоимости фондов —

$$K_{\text{дин}} = \frac{\text{Полная стоимость фондов на конец года}}{\text{Полная стоимость фондов на начало года}};$$

- среднегодовая стоимость основных фондов —

$$\bar{\Phi} = \frac{\text{Стоимость фондов на начало года} + \text{Стоимость фондов на конец года}}{2};$$

- коэффициент износа на конец года:

$$K_{\text{изн}} = \frac{\text{Сумма износа на конец года}}{\text{Полная стоимость фондов на конец года}} 100\%;$$

- фондоемкость продукции:

$$f_e = \frac{\text{Среднегодовая стоимость основных фондов}}{\text{Стоимость выпуска продукции}}.$$

Фондоотдача является обратной величиной фондоемкости.

2. Численность учителей в трех школах муниципалитета на начало учебного года составила: школа № 57 — 54 чел., школа № 68 — 43 чел., школа № 32 — 48 чел. Средний педагогический стаж работы учителей по школам составил соответственно 15,7, 18,3 и 14,5 года. Количество учащихся на одного учителя в школе № 57 — 12,5 чел., школе № 68 — 15,6 чел., школе № 32 — 12,5 чел.

Рассчитайте среднее число учителей, среднее число учеников, среднюю нагрузку на одного учителя, средний стаж работы одного учителя в трех школах муниципалитета.

Оформите исходные данные и решение в виде таблицы.

Тема: Изучение вариации

Вариант 1

1. В лаборатории комбикормового завода произведена контрольная проверка пшеницы, поступившей для производства комбикормов, на содержание протеина. В результате получены следующие данные.

Содержание протеина, %	Число проб		
	I партия	II партия	III партия
6	9	—	2
7	14	11	18
8	20	22	26
9	4	10	4
10	3	7	—
Итого	50	50	50

Рассчитайте по каждой партии: модальное и среднее содержание протеина, среднее квадратическое отклонение, коэффициент вариации содержания протеина. Какая партия пшеницы является более однородной по содержанию протеина?

2. В ходе маркетингового исследования рынка услуг оператора сотовой связи получено следующее распределение продолжительности исходящих соединений внутри сети.

Продолжительность соединений, с	До 10	10–30	30–60	60–90	90–120	120 и более
Процент соединений	5,0	10,2	24,3	32,5	20,9	7,1

Определите: а) наиболее типичную (модальную); б) медианную; в) среднюю продолжительность внутрисетевых соединений. Сделайте выводы о форме распределения.

3. По данным задания 2 постройте кумуляту распределения внутрисетевых соединений по продолжительности. Определите графически долю соединений более 100 с.

В а р и а н т 2

1. На основе материалов выборочных обследований населения по проблемам занятости построены следующие распределения безработных региона по возрасту за два года (%).

Возраст, лет	До 20	20–29	30–39	40–49	50–59	60–72	Итого
Базисный год	14,9	33,3	21,6	14,9	8,8	6,5	100,0
Отчетный год	3,0	26,7	24,5	21,2	20,6	4,0	100,0

Рассчитайте для каждого года средний возраст безработных, среднее квадратическое отклонение, коэффициент вариации возраста безработных. Можно ли считать совокупность безработных однородной по возрасту в отчетном году? в базисном году?

Напишите краткие выводы об изменении возрастной структуры безработных в регионе.

2. На одном поле графика по данным задания 1 постройте полигоны распределения безработных по возрасту в базисном и отчетном годах.

Что вы можете сказать о форме распределения для каждого года?

3. Имеются следующие данные, характеризующие уровень заработной платы в отраслях А и Б (руб.).

Отрасль	Средняя зароботная плата	Децили распределения		
		первая	пятая	девятая
А	17310	5405	18560	28530
Б	19450	7580	17620	31110

Рассчитайте коэффициенты децильной дифференциации заработной платы в каждой отрасли. Дайте интерпретацию полученным результатам.

Для каждой отрасли укажите величину заработной платы, больше которой получает половина работников.

Можно ли на основании имеющихся данных сделать выводы о характере асимметрии распределения работников отраслей по уровню заработной платы? Обоснуйте свой ответ.

Кроссворд 1 (рис. П4.1)

По горизонтали:

1. График кумулятивного распределения.
2. Абсолютный показатель размера вариации.
3. Значения признака, делящие совокупность на четыре равные по числу единиц части.
4. Величина варьирующего признака, делящая совокупность на две равные части.
5. Соотношение форм средних, выраженное в виде неравенства

$$\bar{X}_{\text{гарм}} \leq \bar{X}_{\text{геом}} \leq \bar{X}_{\text{арифм}} \leq \bar{X}_{\text{квадр}} \leq \bar{X}_{\text{куб}}.$$

6. Одна из характеристик формы распределения.
7. Квадрат среднего квадратического отклонения.
8. Форма определения средней арифметической величины, при которой остаются неизвестными индивидуальные значения осредняемого признака.

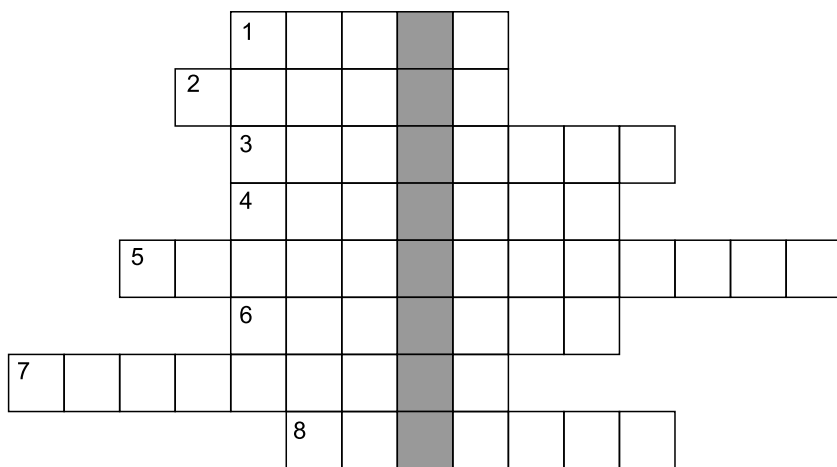


Рис. П4.1. Кроссворд 1

По вертикали ключевое слово: различие значений признака у разных единиц совокупности в один и тот же период или момент времени.

Кроссворд 2 (рис. П4.2)

По горизонтали:

1. Значения признака, делящие совокупность на десять равных по числу единиц частей.
2. Значение, которое принимает исследуемый признак в вариационном ряду.
3. Средняя, характеризующая пространственные или динамические системы.
4. Абсолютная разность между максимальным и минимальным значениями признака.
5. Характеристика формы распределения.
6. График, используемый для изображения вариационного ряда.
7. Коэффициент, характеризующий относительный размах вариации.
8. Английский статистик, предложивший формулу для расчета показателя асимметрии.
9. График вариационного ряда, построенного по накопленным частотам.

По вертикали ключевое слово: показатель вариации.

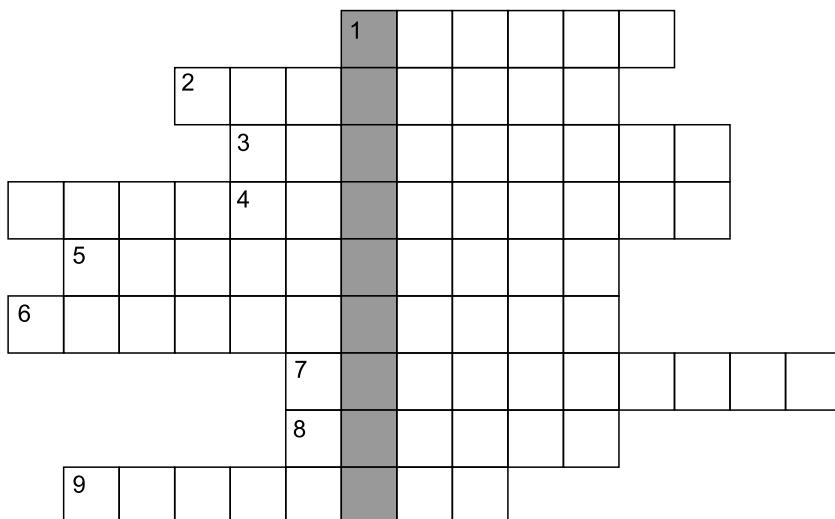


Рис. П4.2. Кроссворд 2

**Тема: Выборочное наблюдение.
Испытание статистических гипотез**

Вариант 1

1. В библиотеке изучаются сроки пользования читателями книгами. С этой целью из 20000 выданных книг в порядке бесповторной выборки обследовано 5% их числа. Установлено, что средний срок пользования книгой составил 18 дней при коэффициенте вариации 67%. Количество книг, находящихся в пользовании более установленного срока, оказалось равным 196.

С вероятностью 0,93 определите предельные значения среднего срока пользования книгой и количество книг, возвращенных с нарушением установленного срока пользования.

2. В результате социологического опроса 100 домохозяйств области получены следующие характеристики формы распределения по среднедушевому доходу в месяц:

- коэффициент асимметрии $(-0,547)$;
- показатель эксцесса $(-0,91)$.

Оцените, существенно ли отличается полученное распределение от нормального.

Вариант 2

1. Проводится выборочная проверка качества выпускаемых электроламп. Случайно отобраны 1000 ламп, что составляет 4% общего выпуска. Установлено, что средний срок горения лампы составляет 2300 ч, а коэффициент вариации срока горения – 30%.

С какой вероятностью можно утверждать, что средний срок горения всех изготовленных электроламп будет не менее 2256 и не более 2344 ч?

2. В целях определения средней дневной реализации продукции способом случайного отбора в городе отобрано 70 кафе. Средний размер реализации составил 90 тыс. руб. при среднем квадратическом отклонении 22,5 тыс. руб. Одновременно в другом городе в случайном порядке было обследовано 65 кафе, средний размер дневной реализации в которых составил 75 тыс. руб., а среднее квадратическое отклонение 28 тыс. руб.

Существенно ли расхождение в средней реализации в кафе двух городов? Дайте ответ на 5%-ном уровне значимости.

Тема: Корреляционно-регрессионный анализ

Вариант 1

1. Линейные коэффициенты корреляции между процентом выполнения норм выработки (y), квалификацией (x) и стажем работы рабочих предприятия (z) оказались равными: $r_{yx} = +0,587$; $r_{yz} = +0,348$; $r_{xz} = 0,108$.

Определите:

1) совокупный коэффициент корреляции между процентом выполнения норм выработки и двумя определяющими его факторами (квалификацией и стажем работы);

2) частные коэффициенты корреляции между:

а) процентом выполнения норм выработки и квалификацией рабочих;

б) процентом выполнения норм выработки и стажем работы рабочих;

3) совокупный и частный коэффициенты детерминации.
Сделайте выводы на основе полученных результатов.

2. Имеются следующие выборочные данные о денежных вкладах населения в коммерческие банки района.

Группа вкладчиков	Число вкладов	Средняя сумма вклада, тыс. руб.	Коэффициент вариации суммы вклада, %
Горожане	70	34	20
Сельские жители	30	16	30

Определите тесноту связи между средним размером вклада и группой вкладчиков.

Вариант 2

1. Изучается зависимость между размером собственного капитала банка и объемом депозитов, привлеченных от физических лиц, в коммерческих банках региона. Получены следующие данные.

Собственный капитал, млн руб.	Число банков	Средний объем депозитов в одном банке, млн руб.	Коэффициент вариации объема депозитов, %
50–100	6	368	48
100–500	16	652	46
500–1000	11	1850	26

Охарактеризуйте связь между размером собственного капитала банка и объемом депозитов, привлеченных от физических лиц, в коммерческих банках региона.

2. По 50 предприятиям отрасли изучается зависимость между рентабельностью производства (прибыль, % к стоимости основных и оборотных фондов), производительностью труда (x_1 — млн руб. на одного работника) и средним возрастом производственного оборудования (x_2 — лет). Получено следующее уравнение регрессии:

$$y = 6,9 + 0,67 \cdot x_1 - 0,3 \cdot x_2.$$

(0,1) (0,125)

В скобках приведены стандартные ошибки параметров уравнения.

$$R^2 = 0,98.$$

Задание:

Оцените статистическую значимость коэффициентов регрессии и уравнения регрессии в целом.

Дайте интерпретацию полученному уравнению регрессии и укажите, для решения каких задач оно может использоваться.

Тема: Статистическое изучение динамики

В а р и а н т 1

1. Имеются некоторые данные, характеризующие динамику ввода в действие общей площади жилых домов в Российской Федерации в 2000–2005 гг.

В приведенной таблице необходимо восстановить пропущенные данные. Рассчитайте средние показатели динамики: средний уровень ряда, среднее абсолютное изменение, средний темп роста и средний темп прироста. Сделайте выводы.

Год	Ввод в действие общей площади жилых домов, млн м ²	Цепные показатели динамики			
		Абсолютное изменение, млн м ²	Темп роста, %	Темп прироста, %	Абсолютное значение 1% прироста, млн м ²
2000	30,3	—	—	—	—
2001	...	1,4
2002	106,6
2003	7,7	...
2004
2005	...	2,6	0,41

2. Известны следующие данные о продаже товара по кварталам.

Квартал	1	2	3	4	5	6	7	8	9	10	11	12
Продано, млн шт.	2	3	5	3	4	5	8	5	6	8	10	7

Определите прогноз на I квартал четвертого года на основе среднего абсолютного прироста (с точностью до 0,01).

Рассчитайте средний индекс сезонности на II квартал четвертого года (в разгах с точностью до 0,001), если по исходным данным получено уравнение тренда $y = 1,86 + 0,56t$.

Вариант 2

1. Имеются некоторые данные, характеризующие динамику грузооборота транспорта общего пользования в Российской Федерации в 2000–2005 гг.

В приведенной таблице необходимо восстановить пропущенные данные. Рассчитайте средние показатели динамики: средний уровень ряда, среднее абсолютное изменение, средний темп роста и средний темп прироста. Сделайте выводы.

Год	Грузооборот транспорта общего пользования, трлн т·км	Цепные показатели динамики			
		Абсолютное изменение, трлн т·км	Темп роста, %	Темп прироста, %	Абсолютное значение 1% прироста, трлн т·км
2000	...	—	—	—	—
2001	102,9	...	0,035
2002	...	0,2
2003	7,9	...
2004
2005	4,5	...	102,3

2. На основе помесечных данных о потреблении электроэнергии в регионе (млн кВт·ч) за последние три года была построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за соответствующие месяцы приводятся ниже.

Январь	+25	Май	−32	Сентябрь	+2
Февраль	+10	Июнь	−38	Октябрь	+15
Март	+6	Июль	−25	Ноябрь	+27
Апрель	−4	Август	−18	Декабрь	?

Запишите модель в общем виде.

Определите значение сезонной компоненты за декабрь.

Тема: Индексы

В а р и а н т 1

1. Имеются следующие данные о расходах населения и индексах потребительских цен в регионе за два года.

Группа товара	Расходы населения, млрд руб.		Индексы цен, %
	Базисный год	Отчетный год	
Продовольственные товары	121,5	133,5	110,2
Алкогольные напитки	10,8	14,1	108,4
Непродовольственные товары	83,7	119,6	113,6
Платные услуги	54,0	84,8	120,4

Определите:

1) сводный индекс потребительских цен на товары и платные услуги;

2) на сколько изменилась стоимость набора товаров и услуг базисного периода за счет роста цен;

3) сводный индекс физического объема потребленных товаров и услуг;

4) как изменились реальные среднедушевые расходы населения в регионе, если численность населения за рассматриваемый период выросла на 8%;

5) сводный индекс покупательной способности денег.

2. Известны следующие данные о занятости и оплате труда на городском пассажирском транспорте.

Вид транспорта	Доля занятых, %		Среднемесячная заработная плата, тыс. руб.	
	Базисный год	Отчетный год	Базисный год	Отчетный год
Метрополитен	26,3	27,4	17,1	18,5
Трамвай	13,7	11,3	13,2	14,5
Троллейбус	16,3	14,8	13,0	15,1
Автобус	43,7	46,5	14,5	16,5

Рассчитайте:

1) средние начисленные заработные платы работников городского пассажирского транспорта в базисном и отчетном периодах;

2) индивидуальные индексы средней заработной платы по отдельным видам транспорта;

3) систему индексов средней заработной платы переменного состава, постоянного состава и структурных сдвигов;

4) как в абсолютном выражении изменилась средняя заработная плата на городском пассажирском транспорте в целом и в том числе за счет изменения уровня оплаты труда на отдельных видах транспорта и за счет изменений в структуре занятых по видам транспорта.

По результатам вычислений напишите краткие выводы.

Вариант 2

1. Имеются следующие данные о затратах предприятия на производство продукции.

Вид изделия	Затраты, млн. руб.		Прирост количества выпущенных изделий во 2-м полугодии по сравнению с 1-м, %
	1-е полугодие	2-е полугодие	
АС - 35	10,2	10,7	5
АК - 412	15,1	16,9	7
КМ -6	4,7	4,6	4

Определите:

1) сводный индекс физического объема продукции предприятия;

2) сводный индекс затрат на производство;

3) индивидуальные индексы себестоимости продукции;

4) сводный индекс себестоимости продукции;

5) изменение затрат в абсолютном выражении в целом и в том числе за счет роста объемов производства и за счет изменения себестоимости;

6) долю прироста затрат, полученную за счет каждого из факторов;

7) как изменилась рентабельность затрат на предприятии, если известно, что прибыль от реализации продукции во 2-м полугодии по сравнению с 1-м выросла на 8,5%.

2. Известны следующие данные о приобретении жилья населением на первичном рынке в городе N -ске.

Тип квартиры	Приобретено жилья, тыс. м ²		Средняя цена 1 м ² , долл.	
	Базисный год	Отчетный год	Базисный год	Отчетный год
Типовая	950	850	1100	1500
Улучшенной планировки	780	1100	1450	2100
Элитная	100	150	2690	4500

Рассчитайте:

1) средние цены 1 м² жилья на первичном рынке в отчетном и базисном периодах;

2) индексы изменения цен на жилье для отдельных типов квартир;

3) структуру приобретенного жилья по типу квартир в отчетном и базисном периодах;

4) систему индексов средних цен переменного состава, постоянного состава и структурных сдвигов;

5) как в абсолютном выражении изменилась средняя цена жилья на первичном рынке в целом и в том числе за счет изменения цен в отдельных ценовых сегментах и за счет изменения в структуре приобретенного жилья.

По результатам вычислений напишите краткие выводы.

ТЕСТЫ ДЛЯ ИТОГОВОГО КОНТРОЛЯ

Количество тестов: 30.

Время выполнения задания: 1 ч 30 мин (два академических часа).

Максимальная сумма баллов: 100.

Максимальная оценка по тестам, требующим вычисления, выставляется при приведении правильного решения. В ряде тестов возможно несколько правильных ответов.

Шкала итоговой оценки:

80 и более баллов — отлично;

60–79 баллов — хорошо;

40–59 баллов — удовлетворительно;

менее 40 баллов — неудовлетворительно.

В а р и а н т 1

1. Организуется выборочное обследование транспортной мобильности жителей пригорода крупного мегаполиса. Объектом наблюдения выступает совокупность:

а) населенных пунктов, находящихся в пригородной черте города;

б) транспортных предприятий, осуществляющих пригородные перевозки;

в) пассажиров, пользующихся пригородным транспортом;

г) жителей пригорода.

(max 2 балла)

2. Назовите вторичные признаки агентства по продаже недвижимости:

а) численность штатных сотрудников, чел;

б) численность внештатных сотрудников (совместителей), чел.;

в) число заключенных договоров купли-продажи, ед.;

г) средняя сумма договора, тыс. руб.;

д) площадь офиса, м²;

е) средний процент агентских вознаграждений.

(max 2 балла)

3. Бизнес-план развития торговой сети «Магнит» предусматривал рост торговых площадей в городе N-ске на 15%. Относительная величина выполнения плана по вводу торговых площадей составила 98%. На сколько процентов выросла за год торговая площадь сети «Магнит» в городе:

а) 12,7;

б) 14,7;

в) 14,4;

г) 13,0.

(max 3 балла)

4. При проверке качества электроламп выявлено, что на каждые 1000 ед. годной продукции в среднем приходится 85 бракованных изделий. Удельный вес годной продукции составляет (в %, с точностью до 0,1):

а) 92,2;

б) 85,0;

в) 91,5;

г) 92,9.

(max 3 балла)

5. Известны данные об остатках оборотных средств у предприятия:

1 апреля – 340 тыс. руб.;

1 мая – 315 тыс. руб.;

1 июня – 330 тыс. руб.;

1 июля – 290 тыс. руб.

Средний остаток оборотных средств за II квартал составил:

а) 321,0;

б) 328,33;

в) 318,75.

(max 4 балла)

6. Собственный капитал двух крупнейших банков в регионе равен соответственно 4,2 и 3,2 млрд руб. Балансовая прибыль, полученная за отчетный период, составила 1,4 и 0,5 млрд. руб. Средняя рентабельность собственного капитала составляет (в %, с точностью до 0,1):

а) 27,5;

б) 24,5;

в) 25,7.

(max 4 балла)

7. В результате модернизации оборудования выработка отдельных бригад выросла в среднем на 30%. Одновременно численность работников каждой бригады была сокращена на 10%. Средняя выработка по бригадам увеличилась:

- а) на 20%;
 - б) на 17%;
 - в) на 30%.
- (max 3 балла)

8. В двух курортных городах Индии — Бомбее (Мумбае) и Пури — вариация максимальных среднемесячных температур характеризуется среднеквадратическими отклонениями 1,3 и 1,7°C при коэффициентах вариации соответственно 4,1 и 5,7%. В каком городе максимальная среднегодовая температура выше:

- а) в Бомбее (Мумбае);
 - б) Пури;
 - в) определить нельзя.
- (max 4 балла)

9. Децильный коэффициент дифференциации доходов населения в регионе равен 9,5. Выберите правильную интерпретацию:

а) средний доход 10% населения, имеющего наибольшие доходы, превышает средний доход 10% населения с наименьшими доходами в 9,5 раза;

б) минимальный доход 10% населения, имеющего наибольшие доходы, превышает максимальный доход 10% населения с наименьшими доходами в 9,5 раза;

в) максимальный доход 10% населения, имеющего наибольшие доходы, превышает минимальный доход 10% населения с наименьшими доходами в 9,5 раза;

г) максимальные доходы населения в крайних децильных группах различаются в 9,5 раза.

(max 3 балла)

10. Индекс потребительских цен в регионе за 12 месяцев описывается уравнением $y = 101,3 + 0,6t + 0,1t^2$, где $\sum t = 0$. Среднемесячный абсолютный прирост потребительских цен в регионе составляет:

- а) 1,3 п.п.;
- б) 0,6 п.п.;

- в) 0,1 п.п.;
 - г) другое.
- (max 4 балла)

11. На конец 2004 г. в Москве на учете в лечебно-профилактических учреждениях с диагнозом наркомания состояло 21903 чел., что выше, чем на конец 2003 г., на 1,2%.

Определите численность больных наркоманией на конец 2000 г., если известно, что контингент больных наркоманией в Москве в 2003 г. был выше, чем в 2000 г., на 26,7%:

- а) определить нельзя;
 - б) 17125 чел.;
 - в) 17117 чел.;
 - г) 17082 чел.
- (max 4 балла)

12. Средняя цена мобильного телефона из числа проданных в пунктах «Евросеть» выросла за первое полугодие года на 18%, при этом цены на отдельные модели телефонов в среднем снизились на 8%.

Что можно сказать об изменении покупательского спроса на мобильные телефоны:

- а) увеличилась доля покупок более дорогих моделей;
 - б) увеличилась доля покупок более дешевых моделей;
 - в) определить нельзя.
- (max 4 балла)

13. Компания «Нева» занимается реализацией автомобильного бензина и дизельного топлива, при этом на долю автомобильного бензина приходится 80% общего объема продаж.

Как в среднем изменилась цена на топливо, реализуемое компанией, если известно, что бензин автомобильный подорожал на 4%, а дизельное топливо на 1,6%? Цена в среднем выросла на:

- а) 2,4%;
 - б) 3,5%;
 - в) 5,6%;
 - г) другое.
- (max 4 балла)

14. Укажите значение коэффициента ассоциации между числом детей в семье и образованием отца по данным следующей таблицы.

Образование отца	Число детей в семье	
	один	два и более
Среднее и неполное среднее	100	150
Высшее и незаконченное высшее	70	30

- а) 0,71;
 - б) 0,56;
 - в) $-0,33$;
 - г) другое.
- (max 3 балла)

15. Доля дисперсии зависимой переменной, которая объясняется знанием независимой переменной, указывается величиной коэффициента:

- а) корреляции;
 - б) сопряженности;
 - в) детерминации;
 - г) контингенции.
- (max 2 балла)

16. По данным о ежегодном числе регистраций новых автомобилей в США с 1951 по 1965 г. (тыс. шт.) построено уравнение тренда: $Y = 6,308 + 0,233t$.

Параметры уравнения получены при использовании метода условного нуля.

Укажите, какой год был принят за нулевой:

- а) 1950;
 - б) 1951;
 - в) 1958;
 - г) иное.
- (max 2 балла)

17. Свободный член уравнения тренда в тесте 16 означает:

- а) среднегодовое число зарегистрированных автомобилей;
- б) число зарегистрированных автомобилей в «нулевом году»;
- в) среднегодовой абсолютный прирост числа зарегистрированных автомобилей за рассматриваемый период.

(max 3 балла)

18. Коэффициент регрессии в уравнении тренда в тесте 16 означает:

- а) среднегодовой абсолютный прирост числа зарегистрированных автомобилей за рассматриваемый период;

б) среднегодовой прирост числа зарегистрированных автомобилей, считая от «нулевого года»;

в) среднегодовой темп роста числа зарегистрированных автомобилей.

(max 3 балла)

19. Как изменятся параметры уравнения тренда (тест 16), если отказаться от метода условного нуля и принять значения переменной «время» от 1 до 15:

1) параметр a :

а) не изменится;

б) увеличится;

в) уменьшится;

г) предсказать невозможно;

2) параметр b :

а) не изменится;

б) увеличится;

в) уменьшится;

г) предсказать невозможно.

(max 5 балла)

20. Дисперсия годовых расходов на приобретение продуктов питания составляет в регионе А 16800, в регионе Б — 24100.

В каком из регионов при равных объемах выборки ошибка выборки будет больше:

а) регион А;

б) регион Б;

в) данных для ответа на вопрос недостаточно.

(max 4 балла)

21. Изучается связь между сортом винограда и продолжительностью созревания. Получены следующие значения дисперсии: общая — 7240, межгрупповая — 5910, средняя из внутригрупповых — 1330. Можно ли утверждать, что связь между сортом винограда и продолжительностью созревания:

а) тесная;

б) умеренно тесная;

в) заметная;

г) слабая;

д) сделать вывод о тесноте связи нельзя.

(max 4 балла)

22. При анализе изменения структуры иностранных инвестиций в двух регионах получены следующие значения коэффициента корреляции рангов по видам инвестиций: регион А — 0,2, регион Б — 0,8 . В каком из регионов изменения структуры были более существенными:

- а) регион А;
 - б) регион Б;
 - в) данных недостаточно?
- (max 4 балла)

23. Если произведены все возможные выборки одного и того же объема из одной и той же генеральной совокупности, то средняя из этих выборочных средних будет:

- а) равна генеральной средней;
 - б) больше генеральной средней;
 - в) меньше генеральной средней;
 - г) иное.
- (max 3 балла)

24. Если построен 95%-ный доверительный интервал, то он означает, что в 95 случаях из 100 он включает:

- а) тот параметр, который оценивается;
 - б) ошибку выборки;
 - в) параметр, который оценивается, и ошибку выборки;
 - г) вероятность оценок.
- (max 3 балла)

25. Динамика числа родившихся в регионе характеризуется следующими данными.

Месяц	Индекс сезонности, %	Месяц	Индекс сезонности, %	Месяц	Индекс сезонности, %
Январь	—	Май	109	Сентябрь	94
Февраль	102	Июнь	108	Октябрь	91
Март	108	Июль	105	Ноябрь	84
Апрель	99	Август	100	Декабрь	78

Чему равен коэффициент сезонности для января:

- а) 72;
 - б) 112;
 - в) 122;
 - г) иное?
- (max 3 балла)

26. Как изменится коэффициент корреляции, определяющий зависимость между переменными y и x_1 при включении в анализ дополнительной переменной x_2 , если известно, что $r_{yx_1} = 0,897$, а $r_{x_1x_2} = -0,425$:

- а) увеличится;
 - б) уменьшится;
 - в) предсказать нельзя.
- (max 5 балла)

27. Текущие затраты на охрану окружающей среды на предприятии в сопоставимых ценах выросли за 6 месяцев с начала года на 18%. По какой формуле будет определен среднемесячный темп роста затрат?

- а) $18/6$;
- б) $1,18/6$;
- в) $\sqrt[6]{1,18}$;
- г) $\sqrt[6]{18}$.

(max 2 балла)

28. Коэффициент эластичности между ростом среднего размера пенсий и ростом среднемесячных затрат на лекарства составляет 0,66. Это значит, что с увеличением:

- а) среднего размера пенсии на 1% среднемесячные затраты на лекарства увеличиваются на 0,66%;
- б) среднего размера пенсии на 0,66% среднемесячные затраты на лекарства увеличиваются на 1%;
- в) среднего размера пенсии на 0,66% среднемесячные затраты на лекарства увеличиваются на 0,34%.

(max 4 балла)

29. Каково соотношение между наблюдаемыми и ожидаемыми частотами в таблице сопряженности?

- а) равны в каждой клетке;
- б) равны по итогам строк и столбцов;
- в) равны в целом по совокупности;
- г) равенство не достигается.

(max 3 балла)

30. Если таблица сопряженности для двух переменных имеет 5 строк и 6 столбцов, то число степеней свободы для статистики χ^2 -квадрат равно:

- а) 2;
 - б) 5;
 - в) 6;
 - г) 20;
 - д) 30.
- (max 3 балла)

Вариант 2

1. Студентам предложено по 5-балльной системе оценить обеспеченность вуза Интернет-ресурсами и мультимедийными средствами. Полученные распределения являются распределениями по:

- а) номинальным признакам;
- б) порядковым признакам;
- в) количественным признакам;
- г) дискретным признакам;
- д) описательным признакам.

(max 2 балла)

2. Известны данные о числе посещений пункта семейной медицины за пять дней с начала месяца:

- 1 апреля — 15 чел.;
- 2 апреля — 16 чел.;
- 3 апреля — 19 чел.;
- 4 апреля — 13 чел.;
- 5 апреля — 15 чел.

Укажите тип динамического ряда:

- а) моментный;
- б) интервальный;
- в) непрерывный;
- г) дискретный.

(max 3 балла)

3. Бизнес-план на отчетный период предусматривал снижение издержек производства на предприятии на 5%. По итогам отчетного периода достигнуто снижение издержек производства по сравнению с прошлым периодом на 6%. Относительная величина планового задания составила (в %, с точностью до 0,1);

- а) 1,0;
- б) 1,1;
- в) 101,1;

г) 1,2.

(max 4 балла)

4. В регионе на 1000 чел. в трудоспособном возрасте приходится 202 чел. моложе трудоспособного возраста и 374 чел. старше трудоспособного возраста. Удельный вес населения трудоспособного возраста в регионе составляет, %:

а) 42,4;

б) 46,9;

в) 57,6;

г) 63,5.

(max 4 балла)

5. Известны данные о числе выданных потребительских кредитов в пункте обслуживания физических лиц за неделю:

5 сентября – 12;

6 сентября – 18;

7 сентября – 23;

8 сентября – 17;

9 сентября – 19;

10 сентября – 21;

11 сентября – выходной день.

Среднее число кредитов, выданных за день, равно:

а) 18,3;

б) 15,7;

в) 14,8;

г) 18,7.

(max 3 балла)

6. В торговой сети города в июне продавался картофель урожая прошлого года по средней цене 15 руб. за 1 кг и картофель нового урожая в среднем по цене 27 руб. за 1 кг. Определите среднюю цену продажи картофеля, руб., в июне, если известно, что спрос на картофель нового урожая был в два раза выше, чем на картофель прошлого урожая:

а) 23,0;

б) 21,0;

в) 21,8;

г) 24,2.

(max 4 балла)

7. Между осредняемым признаком и признаком-весом существует прямая корреляция. В этом случае взвешенная средняя:

- а) равна простой;
 - б) больше простой;
 - в) меньше простой.
- (max 3 балла)

8. В прокурорском отчете об оперативности следствия имеются данные о количестве возбужденных ходатайств о продлении процессуальных сроков в одном из административных районов города.

- Всего возбуждено ходатайств - 213;
в том числе продлено:
- а) до 3 месяцев — 93;
 - б) от 3 до 6 месяцев — 81;
 - в) от 6 до 12 месяцев — 36;
 - г) свыше 12 месяцев.

Закончите фразу: «Половина процессуальных сроков продлевается на срок более чем»:

- а) 3,9 месяца;
 - б) 2,7 месяца;
 - в) 3,5 месяца;
 - г) иное.
- (max 4 балла)

9. Численность населения в регионе А составляет 1620 тыс. чел. и при этом в среднем за год сокращается на 24 тыс. чел. Численность населения региона Б составляет 1480 тыс. чел. и растет в среднем на 20 тыс. чел. в год.

Через сколько лет регионы сравняются по численности населения при условии, что демографические тенденции в регионах останутся неизменными:

- а) 3,2 года;
 - б) 4,0 года;
 - в) 5,1 года;
 - г) предсказать невозможно?
- (max 4 балла)

10. В регионе за отчетный период цены на продовольственные и непродовольственные товары в розничной торговле выросли на 9 и 11% соответственно.

Определите, как в среднем изменились цены розничной торговли, если известно, что на долю продовольственных товаров приходится 35% розничного товарооборота:

- а) выросли на 10,5% ;
 - б) выросли на 9,3 п.п.;
 - в) выросли на 10,3%;
 - г) предсказать невозможно.
- (max 4 балла)

11. Средняя цена 1 м² жилья на вторичном рынке выросла за год на 45%, при этом жилье на вторичном рынке по отдельным районам города в среднем подорожало на 41%. Выберите верное утверждение:

- а) увеличилась доля покупок жилья в более «дорогих» районах города;
 - б) увеличилась доля покупок жилья в более «дешевых» районах города;
 - в) сделать вывод об изменении структуры покупок нельзя.
- (max 4 балла)

12. Если ряд динамики характеризуется стабильным темпом роста, то для прогнозирования рекомендуется использовать:

- а) линейный тренд;
 - б) параболу 2-го порядка;
 - в) показательную функцию;
 - г) экспоненту;
 - д) логарифмическую параболу.
- (max 3 балла)

13. Продажа газированных напитков в торговой сети города характеризуется следующими данными.

Месяц	Индекс сезонности, %	Месяц	Индекс сезонности, %	Месяц	Индекс сезонности, %
Январь	78	Май	122	Сентябрь	88
Февраль	69	Июнь	143	Октябрь	83
Март	75	Июль	—	Ноябрь	72
Апрель	93	Август	141	Декабрь	78

Чему равен коэффициент сезонности для июля:

- а) 142;
 - б) 158;
 - в) 164;
 - г) другое?
- (max 2 балла)

14. Укажите формулу коэффициента детерминации:

а) $r^2 = 1 - \frac{\text{Необъясненная дисперсия}}{\text{Полная дисперсия}};$

б) $r^2 = 1 - \frac{\text{Полная дисперсия}}{\text{Необъясненная дисперсия}};$

в) $r^2 = \frac{\text{Необъясненная дисперсия}}{\text{Полная дисперсия}}.$

(max 2 балла)

15. Если объем выборки увеличить в 2 раза, то средняя ошибка выборки:

- а) не изменится;
- б) увеличится в 2 раза;
- в) уменьшится в 2 раза;
- г) иное.

(max 4 балла)

16. Изучается связь между типом телефонных переговоров (внутригородские, междугородние и международные) и продолжительностью телефонных разговоров. Получены следующие значения дисперсии: общая – 620, межгрупповая – 310, средняя из внутригрупповых – 310.

Можно ли утверждать, что связь между типом переговоров и продолжительностью разговора:

- а) тесная;
- б) умеренно тесная;
- в) заметная;
- г) слабая;
- д) сделать вывод о тесноте связи нельзя.

(max 4 балла)

17. Сколько децилей рассчитывается в ряду распределения:

- а) десять;
- б) девять;
- в) два;
- г) иное.

(max 2 балла)

18. По данным случайного опроса 400 избирателей в регионе А и 400 избирателей в регионе Б о намерении участвовать в выборах губернатора, в регионе А заявили 25%, а в регионе Б – 30% опрошенных.

Существенны или случайны выявленные различия в намерениях избирателей регионов А и Б?

- а) существенны;
 - б) случайны;
 - в) данных для ответа недостаточно.
- (max 4 балла)

19. В качестве критического момента во Всероссийской переписи населения, проведенной 9 октября 2002 г., было принято 0 ч с 8 на 9 октября.

Отметьте, кто был занесен в бланки переписи:

- а) ребенок, родившийся 8 октября в 15 ч 00 мин.;
- б) ребенок, родившийся 10 октября в 11 ч 00 мин.;
- в) ребенок, родившийся 16 октября в 8 ч 00 мин.;
- г) гражданин Х, умерший 16 октября в 12 ч 00 мин.;
- д) гражданин ХХ, умерший 8 октября в 22 ч 00 мин.

(max 3 балла)

20. Динамика поступлений иностранных инвестиций в экономику России за период 1995 г. и 2005 г. характеризуется уравнением тренда:

$$Y = 4185,6t - 5801,5.$$

Каков смысл полученного значения коэффициента при факторе времени t :

- а) ежегодно иностранные инвестиции снижаются на 1615,9 млн долл.;
 - б) ежегодно иностранные инвестиции возрастают на 4185,6 млн долл.;
 - в) среднегодовой прирост иностранных инвестиций составлял в 1995–2005 гг. 4185,6 млн долл.;
 - г) иностранные инвестиции в 1995 г. составляли 4185,6 млн долл.
- (max 3 балла)

21. Можно ли использовать уравнение тренда (тест 20) для прогноза поступления иностранных инвестиций в 2008 г., если коэффициент детерминации составил 0,7859?

- а) да;
- б) можно, но при большой ошибке прогноза;
- в) нельзя, поскольку начиная с 2000 г. объем иностранных инвестиций в экономику России увеличивался с ускорением;

г) нельзя из-за отрицательного свободного члена.
(max 3 балла).

22. Переход от уравнения линейного тренда, построенного по годовым данным, к уравнению, построенному по месячным данным, можно осуществить:

- а) делением параметров тренда на 12;
 - б) умножением параметров тренда на 12;
 - в) другим способом;
 - г) нельзя осуществить.
- (max 3 балла)

23. Динамика суммы прибыли (в млн руб.) за ряд лет описывается уравнением $Y = 25,7 \cdot 1,15^t$, где $t = 1, 2, \dots$. Ежегодно прибыль возрастает на:

- а) 25,7%;
 - б) 15%;
 - в) 15 млн руб.;
 - г) 1,15%;
 - д) 1,15 млн руб.
- (max 3 балла)

24. Если выборка большого объема извлечена из генеральной совокупности, имеющей отрицательный эксцесс, то распределение выборочных средних:

- а) нельзя предсказать заранее;
 - б) будет приближаться к нормальному распределению;
 - в) будет иметь положительный эксцесс;
 - г) не имеет ни одного из указанных свойств.
- (max 3 балла)

25. Средний возраст безработных в регионе 35,5 года, дисперсия возраста безработных 25. Можно ли считать совокупность безработных однородной по возрасту:

- а) да;
 - б) нет;
 - в) данных недостаточно?
- (max 3 балла)

26. Известны следующие значения парных коэффициентов корреляции:

$r_{yx_1} = 0,897$, $r_{yx_2} = 0,653$, $r_{yx_3} = 0,675$, $r_{x_1x_2} = 0,356$, $r_{x_2x_3} = 0,856$, $r_{x_1x_3} = 0,850$. Какие факторы должны быть включены в многофакторную регрессионную модель:

- а) x_1 и x_2 ;
 - б) x_1 и x_3 ;
 - в) x_2 и x_3 ;
 - г) x_1 , x_2 , x_3 ;
 - д) для отбора факторов данных недостаточно.
- (max 5 балла)

27. Взаимосвязь между уровнем дневной выработки и электровооруженностью труда на 10 предприятиях приборостроения в округе описывается уравнением $Y = 2,02 + 0,796t$. Известно, что средний уровень электровооруженности на предприятиях отрасли составляет 5,0 кВт на одного рабочего, средняя выработка 6,0 тыс. руб. на одного рабочего. Коэффициент эластичности между электровооруженностью и выработкой составляет:

- а) 1,2;
 - б) 0,83;
 - в) 0,66;
 - г) 0,11;
 - д) другое.
- (max 4 балла)

28. При определении тесноты связи двух переменных получены некоторые значения коэффициентов ассоциации и контингенции. Укажите заведомо неправильные результаты:

- а) $Q = 0,766$ и $\Phi = 0,456$;
 - б) $Q = 1,259$ и $\Phi = 0,486$;
 - в) $Q = 0,259$ и $\Phi = 0,894$;
 - г) $Q = 0,456$ и $\Phi = -0,236$.
- (max 3 балла)

29. Для проверки гипотезы о нормальности распределения числа отсутствующих студентов по дням недели имеются данные:

День недели	Число отсутствующих
Понедельник	12
Вторник	9
Среда	11
Четверг	10
Пятница	9
Суббота	9

- Число степеней свободы для статистики χ^2 -квадрат равно:
- а) 60;
 - б) 53;

в) 6;

г) 4;

д) 3.

(max 4 балла)

30. Следует ли при использовании χ^2 -квadrat проверять гипотезу о характере распределения в генеральной совокупности:

а) да;

б) нет;

в) проверяется гипотеза только о нормальности распределения;

г) иное.

(max 3 балла)

СОДЕРЖАНИЕ

Предисловие	3
Тема 1. ПОНЯТИЕ О СТАТИСТИКЕ	5
1.1. Основные положения	5
1.2. Задания по теме	10
1.3. Вопросы для самоконтроля	14
1.4. Методические рекомендации преподавателям	14
1.5. Методические указания студентам	16
Тема 2. ОРГАНИЗАЦИЯ СТАТИСТИКИ. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ	17
2.1. Основные положения	17
2.2. Задания по теме	19
2.3. Вопросы для самоконтроля	29
2.4. Методические рекомендации преподавателям	30
2.5. Методические указания студентам	31
Тема 3. СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ	33
3.1. Основные положения	33
3.2. Задания по теме	45
3.3. Вопросы для самоконтроля	49
3.4. Методические рекомендации преподавателям	50
3.5. Методические указания студентам	53
Тема 4. ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ: ТАБЛИЦЫ И ГРАФИКИ	54
4.1. Основные положения	54
4.2. Задания по теме	86
4.3. Вопросы для самоконтроля	95
4.4. Методические рекомендации преподавателям	96
4.5. Методические указания студентам	97

Тема 5. СРЕДНИЕ ВЕЛИЧИНЫ	98
5.1. Основные положения	98
5.2. Задания по теме	108
5.3. Вопросы для самоконтроля	121
5.4. Методические рекомендации преподавателям	121
5.5. Методические указания студентам	122
 Тема 6. ИЗУЧЕНИЕ ВАРИАЦИИ	 124
6.1. Основные положения	124
6.2. Задания по теме	149
6.3. Вопросы для самоконтроля	159
6.4. Методические рекомендации преподавателям	160
6.5. Методические указания студентам	161
 Тема 7. ГРУППИРОВКА	 162
7.1. Основные положения	162
7.2. Задания по теме	187
7.3. Вопросы для самоконтроля	200
7.4. Методические рекомендации преподавателям	200
7.5. Методические указания студентам	201
 Тема 8. ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ	 202
8.1. Основные положения	202
8.2. Задания по теме	214
8.3. Вопросы для самоконтроля	223
8.4. Методические рекомендации преподавателям	224
8.5. Методические указания студентам	225
 Тема 9. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ	 226
9.1. Основные положения	226
9.2. Задания по теме	249
9.3. Вопросы для самоконтроля	266
9.4. Методические рекомендации преподавателям	266
9.5. Методические указания студентам	267

Тема 10. КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ И МОДЕЛИРОВАНИЕ СТАТИСТИЧЕСКИХ СВЯЗЕЙ	268
10.1. Основные положения	268
10.2. Задания по теме	284
10.3. Вопросы для самоконтроля	299
10.4. Методические рекомендации преподавателям	300
10.5. Методические указания студентам	301
Тема 11. СИСТЕМЫ РЕГРЕССИОННЫХ УРАВНЕНИЙ ..	302
11.1. Основные положения	302
11.2. Задания по теме	310
11.3. Вопросы для самоконтроля	313
11.4. Методические рекомендации преподавателям	314
11.5. Методические указания студентам	315
Тема 12. СТАТИСТИЧЕСКИЙ АНАЛИЗ НЕКОЛИЧЕСТВЕННЫХ ПЕРЕМЕННЫХ	317
12.1. Основные положения	317
12.2. Задания по теме	330
12.3. Вопросы для самоконтроля	336
12.4. Методические рекомендации преподавателям	337
12.5. Методические указания студентам	338
Тема 13. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ	339
13.1. Основные положения	339
13.2. Задания по теме	356
13.3. Вопросы для самоконтроля	364
13.4. Методические рекомендации преподавателям	365
13.5. Методические указания студентам	365

Тема 14. ИНДЕКСЫ	366
14.1. Основные положения	366
14.2. Задания по теме	397
14.3. Вопросы для самоконтроля	411
14.4. Методические рекомендации преподавателям	412
14.5. Методические указания студентам	413
Приложения	416
<i>Приложение 1.</i> Статистические данные	416
<i>Приложение 2.</i> Статистико-математические таблицы	444
<i>Приложение 3.</i> Основные статистические формулы	462
<i>Приложение 4.</i> Контрольные работы по основным темам курса	476
<i>Приложение 5.</i> Тесты для итогового контроля	490

Учебное издание

**Елисеева Ирина Ильинична
Флуд Наталья Александровна
Юзбашев Михаил Михайлович**

**ПРАКТИКУМ
ПО ОБЩЕЙ
ТЕОРИИ СТАТИСТИКИ**

Заведующая редакцией *Л.А. Табакова*
Редактор *Е.А. Рыжова*
Младший редактор *Н.А. Федорова*
Художественный редактор *Ю.И. Артюхов*
Технический редактор *В.Ю. Фотиева*
Корректор *Г.Д. Кузнецова*
Компьютерная верстка *И.В. Витте*
Оформление художника *Н.М. Биксентеева*

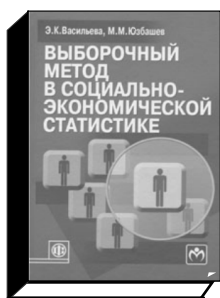
ИБ № 4492

Подписано в печать 25.01.2008
Формат 60х88¹/₁₆. Гарнитура «Таймс». Печать офсетная
Усл. п.л. 31,36. Тираж 3000 экз.
Заказ «С» 010

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефон (495) 625-35-02. Факс (495) 625-09-57
E-mail: mail@finstat.ru <http://www.finstat.ru>

ООО «Великолукская городская типография»
182100, Псковская область, г. Великие Луки,
ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: [zakaz @ veltip.ru](mailto:zakaz@veltip.ru)

Издательство
“ФИНАНСЫ И СТАТИСТИКА”
предлагает учебное пособие
Э.К. Васильева, М.М. Юзбашев
ВЫБОРОЧНЫЙ МЕТОД
В СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ
СТАТИСТИКЕ
256 с.



Представлено развернутое изложение комплекса теории и практики применения выборочного метода в социально-экономической статистике. Детально рассмотрены основные методологические подходы и принципы организации выборочного наблюдения. Раскрыты математические основы выборочного метода и показаны конкретные инструменты и технологии работы с выборочной информацией. Даны систематизированные материалы по основным направлениям современных выборочных обследований в системе Росстата.

Для студентов, аспирантов, преподавателей экономических вузов и факультетов, а также для широкого круга специалистов, осуществляющих формирование, переработку и использование социально-экономической информации.

**По вопросам приобретения литературы
обращайтесь, пожалуйста, в Издательство по адресу:**

101000, Москва, ул.Покровка, 7
(метро “Китай-город”, выход на ул.Маросейка)

Тел.: (495) 625-35-02, 623-80-42

Факс: (495) 623-18-68, 625-09-57

E-mail: mail@finstat.ru <http://www.finstat.ru>

При Издательстве работает киоск:

понедельник – четверг – с 9.00 до 19.30,

пятница – с 9.00 до 18.30

Тел.(495) 621-86-57

Система “Книга – почтой”

Стоимость пересылки почтовыми бандеролями –
25% от стоимости заказа