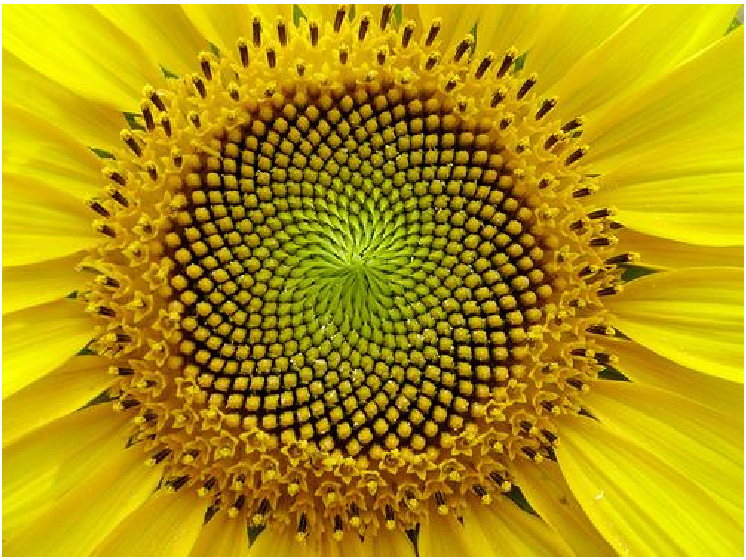
В.И. КРЮКОВ

ГЕНЕТИКА

Глава 9

Статистические методы изучения изменчивости



Орёл 2006

Учебное пособие разработано
на кафедре частной зоотехнии и биотехнологии ОрёлГАУ
д.б.н., проф. В.И.Крюковым.

УДК: 575

Электронная версия опубликована в 2012 году с некоторыми дополнениями и изменениями типографского издания:

Крюков В.И. **Генетика. Часть 5. Статистические методы изучения изменчивости**. Учебное пособие для сельскохозяйственных вузов. –Орёл: Изд-во ОрёлГАУ, 2006. –208 с. с илл.

Пособие рекомендовано Учебно-методическим объединением высших учебных заведений РФ по образованию в области зоотехнии и ветеринарии в качестве учебного пособия для студентов вузов обучающихся по специальности 110401 – Зоотехния и 111201– Ветеринария (Решение о присвоении **грифа № 06-523** от 26 мая 2006 г.)

Учебное пособие предназначено для самостоятельного изучения курсов «Ветеринарная генетика» и «Генетика и биометрия» студентами сельскохозяйственных ВУЗов, аспирантами для подготовки к сдаче кандидатского минимума по специальностям 03.00.15 – «Генетика» и 06.02.01 – «Разведение, селекция, генетика и воспроизводство сельскохозяйственных животных». Пособие может быть использовано преподавателями и студентами агрономических и биологических факультетов вузов. Оно будет полезно практическим работникам племенных хозяйств и селекционных центров, а также – преподавателям биологии средних школ и средних специальных заведений.

В учебном пособии рассмотрены основы статистического анализа изменчивости организмов. Описаны методы построения вариационных рядов, основные статистические показатели, законы статистического распределения, методы оценки достоверности различий статистических показателей, основы корреляционного, регрессионного и дисперсионного анализов.

Рецензенты:

Бакай А.В., д.с.-х.н., профессор, заведующий кафедрой генетики и разведения животных МГАИМиБ им. К.И.Скрябина;
Никульников В.С., к.с.-х.н., доцент кафедры почвоведения и прикладной биологи ОГУ;
Ляшук Р.Н., к.с-х.н., доцент, заведующий кафедрой частной зоотехнии ОрёлГАУ;
Шендаков А.И., к.с.-х.н., старший преподаватель кафедры частной зоотехнии ОрёлГАУ;

СОДЕРЖАНИЕ

9. СТАТИСТИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ ИЗМЕНЧИВОСТИ

Программные требования в курсе «Генетика и биометрия»

Понятие о биометрии и основных ее направлениях. Средний уровень варьирующего признака в выборочной или генеральной совокупности особей. Величины средних значений признака – средняя арифметическая, геометрическая, квадратическая, гармоническая, мода, медиана. Показатели, характеризующие степень изменчивости признака у животных.

Типы распределения варьирующих признаков (нормальное, биномиальное, асимметрическое, эксцессивное, трансгрессивное). Распределение Пуассона.

Определение статистических ошибок и достоверности разности между средними арифметическими двух выборок. Использование критерия «χ^2».

Основы корреляционного и регрессионного анализов.

Основы дисперсионного анализа.

Программные требования в курсе «Ветеринарная генетика»

Количественные и качественные признаки, особенности их изменчивости и методы изучения. Средняя арифметическая, средняя геометрическая, средняя гармоническая. Измерение степени изменчивости признака: лимиты, среднее квадратическое отклонение, варианса, коэффициент вариации. Понятие о статистических ошибках.

Типы распределения варьирующих признаков: биномиальное, нормальное. Понятие об асимметрии, эксцессе и трансгрессии. Понятие о редких событиях и формула Пуассона для учета этих событий.

Понятие о статистических ошибках. Определение достоверности разности между средними арифметическими двух выборок. Метод хи-квадрат для сравнения соответствия теоретического и фактического распределения.

Определение связи между количественными, качественными, количественными и качественными признаками. Коэффициент корреляции и методы его вычисления. Коэффициент регрессии.

Основы дисперсионного анализа. Показатель силы влияния.

Методические рекомендации по изучению темы

Математические методы анализа хозяйственной деятельности предприятия, результатов внедрения новых методов содержания, выращивания, селекции или лечения животных является необходимой частью работы современного специалиста. Научные исследования в современной ветеринарии и зоотехнии также невозможны без применения современных методов статистической обработки экспериментальных данных. В этом методическом пособии изложены основы математического анализа данных.

Приступая к изучению материала запомните, что изучает биометрия, что называется совокупностью и какими могут быть совокупности.

Предметом биометрического анализа являются признаки. Нужно запомнить, что называется биологическим признаком, какими эти признаки могут быть, что является источником варьирования признаков.

Первым этапом анализа данных является построение вариационного ряда. Уясните, что называется вариационным рядом, как он строится и как может быть представлен графически.

Кроме вариационного ряда выборочную совокупность характеризуют некоторые математические величины, которые называют статистическими показателями. Запомните какие статистические показатели используют для характеристики выборочной совокупности и изменчивости изучаемого признака в этой совокупности.

Для специалиста важно уметь вычислять основные статистические показатели. Поэтому уясните, как вычисляют среднюю арифметическую выборки прямым и непрямым способами.

В биометрии важным является понятие статистической ошибки – некоторого несовпадения величин полученных в результате анализа выборочной совокупности со статистическими величинами, характеризующими всю популяцию. Уясните, что называют ошибкой средней арифметической, ошибкой среднего квадратического отклонения и ошибкой коэффициента вариации.

Понимание статистического распределения вариант в вариационных рядах во многих случаях оказывается ключом к правильному выбору методов статистической обработки эмпирических данных. Поэтому изучение этого раздела представляет большое значение.

Приступая к изучению законов статистического распределения, уясните смысл понятий «статистическое распределение» и «вероятность». Обратите внимание на то, что случайные явления – это такая же объективная реальность, как и явления закономерные. Принципиальные различия между ними лишь в том, что при закономерном явлении какая-либо причина приводит объект лишь к одному конкретному результату, при случайном событии результат может быть различными (два и более вариантов).

Обратите внимание на то, что количественной характеристикой вероятности является относительная частота явления, обычно ранее установленная эмпирически, а также на то, что вероятность проявляется при условии большого числа наблюдений.

Очень важны для правильного анализа данных теорем сложения и теорема умножения вероятностей. Уясните их смысл. Обратите внимание на взаимосвязь эмпирических и теоретических вероятностей.

Математической статистике известно много различных типов теоретических вероятностей. В нашем курсе мы изучаем только некоторые из них.

Нормальное распределение – одно из наиболее часто встречающихся распределений вариант, отражающих изменчивость биологических объектов. Обратите внимание на графическое представление вариационной кривой, отражающей нормальный вариационный ряд. Запомните, чему равна площадь фигуры, образованной вариационной кривой нормального распределения и осью абсцисс. Запомните основные закономерности нормального распределения вариант и, прежде всего, правило «трёх сигм».

Запомните принцип анализа принадлежности варианты к данному вариационному ряду и условия исключения варианты из выборки при её резком отклонении от крайних границ вариационного ряда

В генетических исследованиях очень часто приходится встречаться с изменчивостью, характеризуемой биномиальным распределением. Уясните особенности этого распределения и обратите внимание на возможности его практического использования в статистическом анализе выборок.

Распределение Пуассона характеризует редко происходящие события с прерывистой изменчивостью. Обратите внимание на особенности графического представления распределения Пуассона. Уясните принципы использования распределения Пуассона для анализа эмпирических выборок.

При анализе эмпирических данных часто приходится сталкиваться с асимметричным и эксцессивным распределением вариант. Обратите внимание на вариационные кривые, характеризующие эти распределения. Отметьте причины, которые могут приводить к проявлению таких распределений.

В селекционной практике часто приходится сталкиваться с трансгрессивными рядами. Их детальный анализ позволяет вносить корректировки в процессы обора и подбора производителей. Обратите внимание на два свойства трансгрессивных рядов, метод вычисления коэффициента трансгрессии а также способы его использования в прикладном анализе выборочных совокупностей.

Достоверность статистических заключений – важная проблема теории математической статистики и биометрии. Для того чтобы разобраться в этой проблеме настоящее методическое пособие вначале предлагает вспомнить основные биометрические (статистические) понятия: выборочная и генеральная совокупности, ошибка средней арифметической, а также знакомит с некоторыми новыми понятиями: ошибкой среднего квадратического отклонения, ошибкой коэффициента вариации. законом больших чисел и др.

Для оценки достоверности различий между различными выборками очень важно руководствоваться законом больших чисел и стремиться выборки подопытных животных формировать таким образом, чтобы их объём был максимальным в пределах разумной численности.

Важным показателем в сравнении многих выборок является критерий достоверности t_d Стьюдента. Изучая этот критерий, обратите внимание на то, как связаны значения критерия t и доверительные вероятности с объёмами выборочных совокупностей. Запомните, что для доверительных вероятностей и соответствующих им уровней значимости составлены специальные справочные статистически таблицы, которыми следует пользоваться при проведении практических расчётов.

Статистические показатели вычисляются по данным проведённых эмпирических выборок. Поэтому статистические показатели выборочных совокупностей могут несколько отличаться от реальных параметров генеральной совокупности. Однако, специальные статистические приёмы позволяют определить доверительные границы параметров генеральной совокупности при избранном уровне значимости (доверительной вероятности). Уясните, как определяют доверительные границы для средней арифметической генеральной совокупности. Запомните, от чего зависят границы средней арифметической генеральной совокупности.

Приступая к сравнению двух выборок между собой или сравнивая эмпирические полученные распределения с теоретически ожидаемыми изначально исходят из нулевой гипотезы. Запомните, что согласно нулевой гипотезе между рассчитанными показателями двух выборок (и группами, на основе которых они получены) достоверного различия нет, т.е. что обе группы вместе составляют один и тот же однородный материал, одну генеральную совокупность. Если же сравнивается эмпирическое распределение с распределением теоретически ожидаемым, то нулевая гипотеза состоит в том, что между этими двумя распределениями нет статистически достоверных различий. Уясните пути принятия и отклонения нулевой гипотезы.

Одним из широко распространённых приёмов сравнения выборок является сравнение их средних арифметических с помощью критерия достоверности t_d. Запомните рабочую формулу для вычисления этого критерия и алгоритм его расчёта. Обратите внимание на то, что при расчёте числа степеней свободы из сумм двух объёмов выборочных совокупностей вычитается «2», а не единица.

В практической работе ветеринаров и зооинженеров очень часто приходится сравнивать выборки животных, характеризуемые качественными признаками, которыми обладают определённые доли животных в этих выборках. Уясните, чему равна средняя арифметическая качественных признаков в выборке, а также методы расчёта σ и *m* для качественных признаков. Запомните метод определения достоверности различий между выборочными долями и процентами животных, обладающих анализируемым признаком.

Часто животновода-аналитика может интересовать не только разность средних показателей выборочных совокупностей, но и сравнение самой изменчивости признака в двух выборках. Для этого проводят оценку достоверности различий между средними квадратическими отклонениями и вариансами двух выборок. Для этого служит специальный критерий – критерий F. Запомните формулу для расчёта этого критерия и способы определения по нему достоверности различий квадратических отклонений двух выборок.

Очень часто в животноводческой практике используется ещё один важный статистический критерий – критерий хи-квадрат (χ^2). Разработаны несколько различных способов для практического вычисления этого критерия. Использование того или иного способа (и соответствующей формулы) зависит от типа эмпирических данных. Запомните общую формулу вычисления критерия включающую в себя сумму дробей, каждая из которых получена от деления квадрата разности между эмпирическими ($p_{эмп}$) и теоретическими частотами ($p_{теор}$) на частоты теоретические ($p_{теор}$).

Очень важно избежать неправильного использования критерия хи-квадрат. Поэтому обратите внимание на ограничения в использовании этого критерия.

Уясните, как определяют соответствие эмпирически установленных распределений теоретически ожидаемым распределениям. При определении достоверности различий по критерию хи-квадрат эмпирически вычисленное значение критерия сравнивают с рассчитанной теоретической величиной по специальным статистическим таблицам. При сравнении обязательно учитывают выбранный уровень доверительной вероятности и число степеней свободы. По этой причине внимательно изучите, как определяют число степеней свободы при расчётах критерия хи-квадрат. При необходимости вернитесь к изучению понятий «доверительная вероятность» и «уровень значимости».

С помощью критерия хи-квадрат можно сравнивать и между собой две эмпирические выборки. Разберите способы вычисления критерия хи-квадрат в этом случае. Запомните алгоритм вычислений.

Для закрепления материала ответьте на контрольные вопросы (с. 110) и решите несколько задач, приведённых в конце раздела (с. 111).

Корреляционный анализ признаков является важным методом анализа в животноводстве и ветеринарии. Уясните и запомните базовые положения корреляционного анализа (с. 125-127).

Наиболее часто приходится сталкиваться с необходимостью выявления связей в малых выборках. Разберите и запомните алгоритм вычисления коэффициента корреляции для малой выборки (с.129).

В селекционной практике очень часто приходится анализировать большие массивы данных для выявления связи двух признаков. Уясните принцип построения корреляционной решётки и её заполнения соответствующими частотами вариант. Обратите внимание на методический приём вычисления суммы произведений $\Sigma p \cdot a_x \cdot a_y$. Запомните алгоритм вычисления коэффициента корреляции для большой выборки.

При проведении на производстве практических работ очень часто нет необходимости в очень точном расчёте коэффициента корреляции – достаточно определить лишь направление связи и дать её приблизительную оценку (слабая, умеренная, значительная). В этих случаях можно ограничиться анализом лишь распределения частот вариант по корреляционной решётке. Для того чтобы успешно выполнять на производстве такой экспресс-анализ внимательно изучите раздел «Определение величины и направления связи между признаками» (с. 140). Кроме того, ориентировочную оценку признаков, не прибегая к их точному измерению, можно выполнить с помощью полихорического показателя связи ρ. Изучите этот метод корреляционного анализа (с. 150). Обратите внимание на то, что ρ не выявляет направление связи, а только указывает на ее величину.

При анализе корреляционных связей важную роль играет коэффициент регрессии *b* и регрессионный анализ. Разберите смысл и принципы вычисления коэффициента регрессии для больших и малых выборок (с. 142, 144).

В животноводческой практике очень часто приходится анализировать взаимосвязь между качественными альтернативными признаками. Для проведения корреляционного анализа именно таких данных разработан специальный методический подход, излагаемый в разделе 1.9 (с.145). Уясните размещение данных в четырёхпольной корреляционной решётке и метод расчёта коэффициента r_a,

В животноводческой практике встречаются признаки животных, которые нельзя измерить ни точно, ни грубо. Поэтому их вариации обозначают порядковым местом (рангом). Для анализа таких данных разработан ранговый коэффициент связи по Спирмену. Разберите методику расчёта этого коэффициента и уясните, как правильно делать заключения по результатам этого корреляционного анализа (с.147).

Уясните и запомните методику расчёта генетического коэффициента корреляции, используемого для анализа наследования количественных признаков (с. 154).

Заключительный раздел этой главы посвящён дисперсионному анализу – статистическому методу исследований, который позволяет среди множества внешних и внутренних факторов, действующих на организмы, определять влияние отдельных факторов и оценивать их относительную роль в общей изменчивости признаков. Успешность усвоения материала этого раздела будет в значительной мере зависеть от владения вами основных понятий и терминов дисперсионного анализа. Поэтому следует запомнить определения всех новых терминов и понятий, приводимых по мере изложения материала. Обратите внимание что по числу анализируемых факторов дисперсионный анализ бывает одно-, двух-, и трёхфакторным. Запомните, какие три типа схем (статистических комплексов) различают при дисперсионном анализе, какие дисперсионные комплексы называют равномерными, пропорциональными и неравномерными. Обратите внимание на то, что нулевая гипотеза в дисперсионном анализе предполагает отсутствие влияния анализируемых факторов на изменчивость признаков. Разберите общую схему однофакторного дисперсионного комплекса и последовательность его анализа. Обратите внимание на рабочие формулы для вычисления дисперсий и варианс в дисперсионном комплексе – к этим формулам нужно будет обращаться при решении учебных задач. Очень информативным в дисперсионном анализе является показатель силы влияния η^2 (греческая буква «эта»). Этот показатель вычисляют двумя различными методами (Плохинского и Снедекора). Изучите эти методы оценки силы влияния. Достоверность влияния фактора оценивают с помощью критерия Фишера. Запомните, как вычисляют этот критерий и с чем необходимо его сравнивать для выявления статистически достоверного влияния фактора.

Детально разобранные конкретные примеры расчётов, приведённые в разделе 9.6.4, помогут вам научиться использованию метода дисперсионного анализа в практической работе.

В квадратных скобках (например, [3]) указаны цифры, соответствующие номеру использованных литературных источников, указанных в списке «Цитированная литература».

Пособие предназначено для подготовки студентов к семинарским занятиям и экзамену. По этой причине в тексте возможны небольшие повторы, нежелательные с позиций литературного редактора. Однако автор пособия сознательно пожертвовал стилистическим изяществом текста для того, чтобы достигнуть более ясного изложения материала и лучшего его запоминания.

Материал этого пособия будет дорабатываться и дополняться. Последняя его версия выложена на сайте www.labogen.ru

Автор будет признателен студентам и преподавателям за критические замечания о материале этого пособия, а также за конструктивные предложения по его совершенствованию. Направлять свои замечания, предложения можно по адресу: 302019, Орёл, ул. Генерала Родина, д.69, ОрёлГАУ, корп.1, ауд. 205, д.б.н., проф. Крюкову В.И.

ВВЕДЕНИЕ

Статистические методы исследования изменчивости признаков и свойств живых организмов изучает специальная наука – биометрия. Кратко её можно определить следующим образом: **биометрия** – наука о способах применения математических методов в биологии [1]. Точнее и полнее эту науку характеризуют следующим образом: **биометрия** – наука о статистическом анализе массовых явлений в биологии, т.е. таких явлений, в массе которых обнаруживаются закономерности, не выявляемые на единичных случаях наблюдений [3].

Предметом биометрии являются любые биологические объекты, если проводимые над ними наблюдения получают количественное выражение.

9.1. ПОСТРОЕНИЕ ВАРИАЦИОННЫХ РЯДОВ

9.1.1. Понятие о совокупности

Научные наблюдения проводятся не на одном биологическом объекте, а на *группах* таких *объектов*. Ими могут быть особи одного вида или породы, особи одного пола, особи одного возраста и т.п. Члены такой группы называются **единицами наблюдения**. Совокупность относительно однородных, но индивидуально различимых единиц, объединяемых по некоторым общим признакам для их совместного изучения, называют **статистической совокупностью** [3]. Та часть индивидуально различимых единиц, у которых мы можем изучить интересующий нас признак, представляют собой **выборочную совокупность**. В практической работе очень часто выборочную совокупность называют просто **выборкой**. Число этих единиц в выборочной совокупности называется **объёмом выборочной совокупности** и обозначается буквой n. Если объём выборочной совокупности n<30, то её называют **малой** выборочной совокупностью. Если объём выборочной совокупности n>30, то её называют **большой** выборочной совокупностью.

Совокупность может состоять из других, более частных совокупностей. Так, совокупность из всех животных *данной породы* распадается на частные совокупности – стада отдельных хозяйств. В пределах одного стада можно выделить ещё более частные совокупности, например всех потомков определённых быков.

Все индивидуально различимые единицы (биологические объекты) с исследуемым признаком, существующие в природе, называются **генеральной совокупностью**. Генеральная совокупность – это вся подлежащая изучению совокупность данных объектов. В пределе она рассматривается как состоящая из бесконечно большого количества отдельных единиц. Объём генеральной совокупности обозначают символом N и обычно принимают равным бесконечности (∞).

Выборочная совокупность должна как можно более точно отражать изучаемые свойства генеральной совокупности. Наиболее точно генеральную совокупность будет характеризовать случайная выборка, созданная без каких-либо субъективных мотивов специалиста. Для этого каждую единицу измерения из генеральной совокупности выбирают случайным образом. Выборка, сделанная именно таким, случайным образом называется **репрезентативной** (представительной).

9.1.2. Признаки, их свойства и классификация

Анализ биологических объектов проводится по определённым признакам. **Признак** – это характерная особенность строения или функционирования организма. Биологические признаки варьируют, но не всегда поддаются измерению. В связи с этим признаки разделяют на 2 класса: 1) признаки **качественные** и 2) признаки **количественные**. Два взаимоисключающих качественных признака называются альтернативными.

Признаки можно классифицировать иначе – на основе счёта (1, 2, 3, 4…) или на основе измерения (3 см, 3,5 см, 3,8 см, 4,2 см…). На этом основании биологические признаки делят на **мерные** (метрические) и **счётные** (меристические). Мерные признаки варьируют *непрерывно*; их величина может изменяться в определённых пределах, принимая целые и дробные значения. Счётные признаки варьируют *прерывисто* – и их значения всегда выражаются только целыми числами.

Значение или меру признака для конкретного члена выборочной совокупности называют **вариантой** и обозначают определённой буквой. Раньше в биометрической литературе варианты обозначали латинской буквой V с соответствующим индексом (V_1, V_2… V_i). В последнее время варианты чаще обозначают буквой x (x_1, x_2… x_n). Общее же обозначение любой варианты – x_i.

9.1.3. Источники варьирования признаков

Для большинства биологических объектов характерна сильная изменчивость присущих им признаков и свойств. Причиной варьирования исследуемого признака у различных биологических объектов могут быть:

1) генетические и онтогенетические особенности организма;

2) случайные причины, в том числе ошибки измерения (инструментальные, личные ошибки исследователя).

Различные методы статистического анализа позволяют проанализировать размах онтогенетической изменчивости признака, оценить наследственную составляющую в этой изменчивости, определить степень зависимости исследуемого признака от изменения действующих на организм факторов среды и представить в формализованном виде зависимость изменчивости признака от интенсивности воздействия фактора.

9.1.4. Вариационный ряд

Для анализа выборочной совокупности, очень часто нужно полученные данные сгруппировать для более наглядного представления. Особенно это необходимо в том случае если объём выборочной совокупности велик.

Если выборочная совокупность велика, то все значения разбивают на несколько групп, или **классов**. Разбиение выборочной совокупности на определённые классы называется *построением вариационного ряда*. **Вариационный ряд** – это упорядоченное изображение реально существующего распределения особей в группе по величине исследуемого признака. Другими словами **вариационный ряд** – это ряд ранжированных значений признака, в котором указана частота их встречаемости в данной совокупности. Графически вариационный ряд представляет собой *двойной ряд чисел*: 1-й ряд обозначает классы; 2-й ряд обозначает частоты различных вариант изучаемого признака. Указанный ряд *пар чисел* составляет **статистическое распределение** – распределение частот f_i по значениям x_i.

При составлении вариационных рядов следует учитывать тип признаков. При прерывистой изменчивости отдельные случаи прерывистых признаков (число голов, потомков и т. п.) сами по себе могут служить и классами. При непрерывной изменчивости, когда признаки могут быть выражены любым дробным числом (живая масса животных, промеры тела), имеют дело с самыми разнообразными величинами, которые редко повторяются. В этом случае используют разбиение вариационного ряда на классовые интервалы и определение частот вариант в каждом классе.

9.1.5. Построение вариационного ряда

Предположим нам необходимо проанализировать данные о содержании лейкоцитов в крови коров. При микроскопическом анализе крови коров были получены следующие данные (табл. 1).

Таблица 1

Количество лейкоцитов в крови 70 коров (тыс./мм^{3})

4,2	6,1	8,7	5,9	6,4	7,6	9,8
6,7	8,2	4,3	8,1	10,3	6,3	10,7
7,2	**11,4**	9,7	6,6	5,4	6,9	7,6
7,8	5,6	6,7	9,6	7,3	8,5	10,4
5,3	8,8	7,6	5,7	7,9	4,6	5,5
7,6	6,5	4,0	9,5	5,7	10,5	8,6
8,1	9,3	7,3	8,0	8,7	7,4	8,9
6,2	5,5	7,6	6,4	8,5	8,0	4,8
7,1	8,4	9,2	7,9	6,6	7,6	9,8
3,8	7,1	6,2	**3,2**	7,5	7,4	8,3

Для построения вариационного ряда выборочной совокупности нужно выполнить следующие процедуры.

1) Во всей выборочной совокупности (выборке) находят *минимальное* и *максимальное* значения. В нашем примере они выделены жирным шрифтом. Они равны min=3,2 и max=11,4.

2) Определяют *число классов* для вариационного ряда данной выборки. Этот параметр статистики рассчитывают по специальной формуле ($i=1+3{,}3\cdot\lg n$). Вычисления, выполненные по этой формуле, показывают, что число классов зависит от объёма выборочной совокупности следующим образом (табл. 2). Поэтому в практических расчётах достаточно пользоваться этой таблицей.

Таблица 2

Рекомендуемое число классов в вариационном ряде при различных объёмах выборочной совокупности

Число вариант в выборке	25-40	40-60	60-100	100-120
Число классов	5-6	6-8	7-10	8-12

3) Находят классный промежуток (k). **Классный промежуток** (т.е. размер класса) рассчитывают путём деления разности макси-

мального и минимального значений вариант на предполагаемое число классов. Предварительно максимальное и минимальное значения вариант рекомендуется *округлить*: минимальное значение округляют до ближайшей *меньшей* целой величины, а максимальное значение округляют до ближайшей *большей* целой величины (в нашем примере 3,2 округляют до 3,0, а 11,4 – до 12,0). При этом величина, на которую округляют значения вариант, не должны превышать величины k.

Объём нашей выборки равен 70 и, следовательно, число классов должно быть 7, 8, 9, или 10. Выберем любую из этих величин, например, 9. Тогда классный промежуток будет равен (12–3) **:** 9 = 1,0.

4) Определяют нижние и верхние границы каждого класса. Нижняя граница первого класса принимается равной 3,0. *Нижние границы* всех остальных классов должны быть на классовый интервал больше границы предыдущего класса. В нашем примере k=1,0 и, следовательно, нижние границы классов будут равны 3,0, 4,0, 5,0, 6,0, 7,0, 8,0, 9,0, 10,0, 11,0 и (вспомогательная величина) 12,0.

Верхние границы каждого класса должны быть на единицу точности измерения меньше нижней границы следующего класса. В нашем примере точность измерения равна 0,1. Тогда в первом классном интервале его верхняя граница будет равна 3,9 (4,0–0,1=3,9), во втором – 4,9 (5,0–0,1=4,9) и т.д. (табл. 3).

Нижняя граница вспомогательного, 10-го класса, нужна нам только для того чтобы по ней можно было определить верхнюю границу 9-го класса (11,9). Третью графу таблицы пока оставим свободной.

Таблица 3

Границы классов вариационного ряда

Номер класса по порядку	Нижние и верхние границы классов	Частоты вариант
1	3,0 – 3,9	
2	4,0 – 4,9	
3	5,0 – 5,9	
4	6,0 – 6,9	
5	7,0 –7,9	
6	8,0 – 8,9	
7	9,0 – 9,9	
8	10,0 – 10,9	
9	11,0 – 11,9	
10	*12,0*	

На правильное построение шкалы для классов надо обращать очень большое внимание. Границы классов должны быть намечены таким образом, чтобы одна и та же цифра *не повторялась* в двух классах. Если первый класс заканчивается величиной 3,9, то второй класс должен начинаться со следующей по порядку цифры – 4,0. Если бы классы были намечены следующим образом: 3,0-4,0; 4,0-5,0 5,0-6,0 и т.д., то всегда было бы сомнение, к какому классу отнести особь со значением, например, 4,0 или 5,0. Если же один класс будет охватывать значения вариант от 4,0 до 4,9 включительно, а другой – от 5,0 до 5,9 включительно, разноска вариант по намеченным классам не вызовет затруднений.

5) Разносят имеющиеся значения вариант по классам. Для удобства разноски каждую варианту, отнесённую в определённый класс, отмечают точкой или черточкой в виде следующих символов: единица изображается точкой (**.**); две единицы – двумя точками (**. .**); три единицы – тремя точками (**: .**); четыре единицы – четырьмя точками, поставленными по углам квадрата (**: :**); пять единиц – в виде точек и черточки ( ); шесть, семь, восемь единиц – соответственно фигурами ( ), ( ), ( ), девять и десять единиц – в виде «конвертов» ( ) и ( ) (табл. 4).

Таблица 4

Обозначения, используемые для последовательного обозначения частот вариант, распределяемых по классам.

1	2	3	4	5	6	7	8	9	10
•	:	:.	::						

Такое обозначение данных каждого члена совокупности, разнесённых в классы вариационного ряда, облегчает и ускоряет подсчет числа наблюдений в каждом классе. Если в какой либо класс попадает больше 10 вариант, рядом с первым «конвертом» последовательно изображают второй, третий и т.д.

После разноски всех вариант получают вариационный ряд, в котором частоты вариант указаны в символьном виде. Выполнив разноску всех вариант, символьные обозначения частот вариант в каждом классе заменяют на числовые (табл. 5, стр. 15).

Обратите внимание на то, что один из классов имеет наибольшее значение вариант анализируемого признака. Такой класс называют **модальным классом.**

Таблица 5

Результат распределения вариант по классам вариационного ряда

Номер класса по порядку	Нижние и верхние границы классов	Частоты вариант	
		символ	значение
1	3,0 – 3,9		2
2	4,0 – 4,9		5
3	5,0 – 5,9		8
4	6,0 – 6,9		12
5	7,0 –7,9		17
6	8,0 – 8,9		14
7	9,0 – 9,9		7
8	10,0 – 10,9		4
9	11,0 – 11,9		1

9.1.6. Графическое отображение распределения вариант по классам

Вариационный ряд можно отобразить графически. Существуют два способа графического отображения частот в классах.

Первый способ заключается в построении **гистограммы** – графика, на котором по оси абсцисс (*x*) отображены классы, а по оси ординат (*y*) – частоты вариант в соответствующем классе в виде столбиков, высота которых соответствует значениям частот (рис. 1).

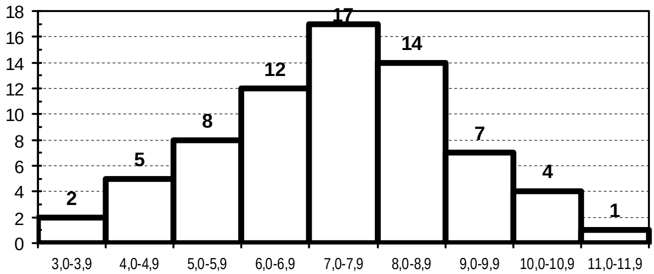


Рис. 1. Гистограмма распределения коров по содержанию в их крови лейкоцитов (тыс./мм^3).

Если середины верхушек столбиков обозначить точками, а затем соединить их линиями, получится график, называемый **вариационной кривой** или **полигоном распределения**.

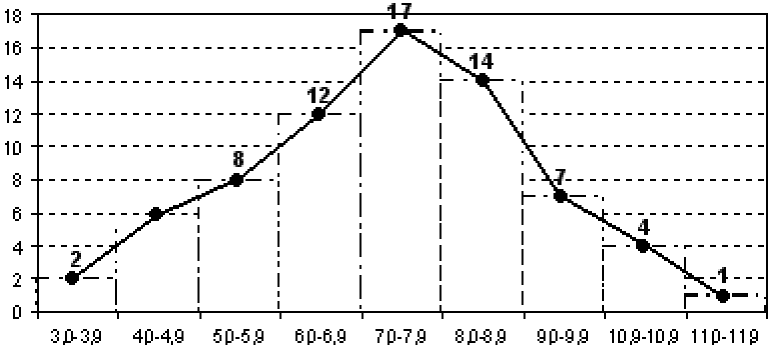


Рис. 2. Вариационная кривая частот коров с различным содержанием лейкоцитов в крови.

Анализ графиков показывает:

1) наибольшее число вариант расположено в центре вариационной кривой;

2) распределение вариант по частотам по обе стороны от вершины приблизительно симметрично:

3) число вариант в классах убывает к краям вариационного ряда.

Пример 2.

Допустим, что в результате взвешивания 25 кроликов были получены варианты, представленные в табл. 6, при этом они расположены в так называемом ранжированном виде, т.е. от меньших величин к большим.

Таблица 6

Живая масса 25 кроликов (кг)

3,2	4,5	5,2	5,6	6,0
3,8	4,7	5,2	5,7	6,3
4,1	4,9	5,3	5,8	6,4
4,3	5,0	5,3	5,8	6,7
4,3	5,1	5,4	5,9	7,3

Необходимо построить вариационный ряд для этих данных.

Весь этот ранжированный ряд надо разбить на определенное количество классов. Таблица 2 рекомендует для такого объёма выборочной совокупности 5-6 классов. Остановимся на 5 классах. Разница между наибольшим и наименьшим значением вариант в нашем примере 7,3–3,2=4,1 кг. Однако, в соответствии с рекомендациями округлим величину 3,2 до ближайшего целого и *меньшего* числа – до 3,0. Максимальную же варианту (7,3) также округлим до ближайшего целого и

Таблица 7

Два различных вариационных ряда, построенных по одним и тем же данным, приведённым в таблице 6.

Классы	Частоты
3,0-3,9	2
4,0-4,9	6
5,0-5,9	12
6,0-6,9	4
7,0-7,9	1

Классы	Частоты
3,0-3,4	1
3,5-3,9	1
4,0-4,4	3
4,5-4,9	3
5,0-5,4	7
5,5-5,9	5
6,0-6,4	3
6,5-6,9	1
7,0-7,4	1

большего числа – 8,0. Классовый интервал *k* в этом случае составит *k*=(8–3)**:**5=1,0. Разнеся все варианты по классам, мы получим вариационный ряд, указанный в левой части таблицы 7 и на рис. 3.

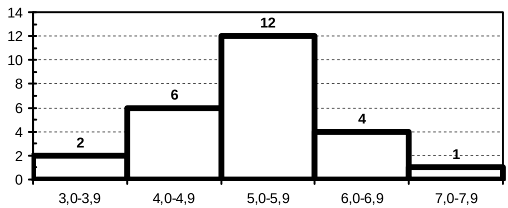


Рис. 3. Гистограмма распределения значений живой массы 25 кроликов в 5 классах вариационного ряда.

Следует обратить внимание на тот факт, что при необходимости вы можете для анализа выбрать число классов, отличающееся от рекомендованного. Например, если при анализе данных о весе кроликов принять за желательное количество классов не 5, а 8 или 9, то, уменьшив величину классового интервала до 0,5 кг, можно наметить следующие классы: 3,0-3,4 кг; 3,5-3,9; 4,0-4,4; 4,5-4,9; 5,0-5,4; 5,5-5,9; 6,0-6,4 кг и т. д. В этом случае получится вариационный ряд, представленный в правой части таблицы 7 и на рис. 4.

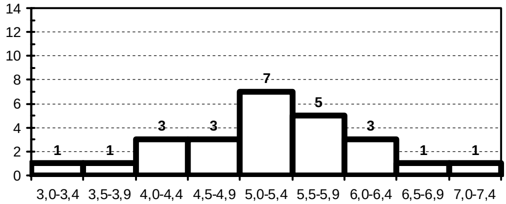


Рис. 4. Гистограмма распределения значений живой массы 25 кроликов в 9 классах вариационного ряда.

В некоторых случаях вариационная кривая может оказаться с двумя вершинами, например (рис. 5).

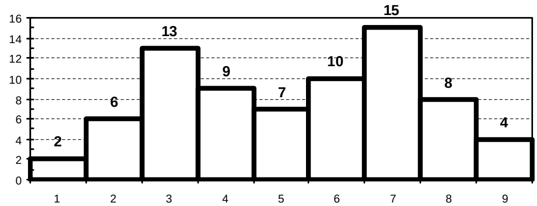


Рис 5. Пример гистограммы двухвершинного вариационного ряда.

Это происходит обычно в тех случаях, когда объём выборочной совокупности недостаточен для анализа. Устранить двухвершинность позволяет (правда, не всегда) выбор другого количества классов и соответственно – межклассового интервала.

В некоторых случаях вариационные кривые оказываются **асимметричными**, т.е. смещёнными по шкале абсцисс вправо, или влево. Обычно это происходит в тех случаях, когда условия факторы, влияющие на обследуемых животных, *не были однородными* (например, условия содержания обследуемых животных были неодинаковыми) или когда *выборка была неслучайной* (например, при тенденциозном подборе животных с худшими, или наоборот – лучшими показателями). Увеличение объёма выборки обычно позволяет устранить асимметрию вариационной кривой. Если же её асимметрия сохраняется и в выборке большого объёма, то это свидетельствует о существовании какого-то специфического воздействия, которое требует специального изучения.

9.2. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВЫБОРОЧНОЙ СОВОКУПНОСТИ

9.2.1. Статистические показатели, характеризующие совокупность

Статистические показатели – это важные абсолютные или относительные величины, характеризующие выборочную совокупность. Основными показателями являются следующие.

Средняя арифметическая ($\bar{x}$)

$$\bar{x} = \frac{\sum x_i}{n} \qquad (1)$$

где x_i – значения вариант, Σ– знак суммирования, n – общее число вариант. Средняя арифметическая показывает, какое значение признака наиболее характерно для данной совокупности.

Средняя геометрическая (G) – это такая средняя величина, которая выявляет средний прирост (или среднее уменьшение) какого-либо показателя за определённый период времени.

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} \qquad (2)$$

где x_n – варианты, n – число членов в выборке.

Средняя геометрическая должна применяться при определении фактического или планового показателя прироста поголовья стада, для выяснения прироста показателей продуктивности или живой массы, для установления снижения отхода или заболеваемости в стаде и т.п. Важно подчеркнуть, что для анализа указанных признаков нельзя пользоваться средней арифметической, так как она даст неправильное представление о темпах прироста или темпах снижения показателей, интересующих ветеринара или зоотехника.

Средняя геометрическая особенно пригодна для таких вариационных рядов, у которых имеется асимметричное распределение частот.

Средняя квадратическая (S) – параметр используемый для расчёта средних площадей, диаметров, радиусов и других подобных параметров биологических объектов. Примерами могут быть такие признаки, как диаметры мышечного волокна, диаметры семенных канальцев половых желёз самца, альвеолярных пузырьков молочной железы, диаметры клеток или их ядер и т.п.

Формула средней квадратической в общем виде такова:

$$S = \sqrt{\frac{\sum x_i^2}{n}} \qquad (3)$$

где x_i – варианты исследуемого признака, n – число членов в выборке.

Средняя гармоническая (H):

$$H = \frac{n}{\sum \frac{1}{x_i}} \qquad (4)$$

Средняя гармоническая используется для определения средних величин процессов, скорость которых меняется во времени, например выделение молока из вымени коровы.

Мода (Mo) – это наиболее часто встречающееся значение варианты. Мода может быть выражена как у качественных, так и количественных признаков. Например, у холмогорского скота черно-пестрая масть встречается чаще, чем красно-пестрая. Поэтому черно-пестрая масть будет модой. Модой является и четырехсосковость коров по сравнению с многососковостью, встречающейся у значительно меньшего числа самок этого вида.

Для количественных признаков модальным будет считаться та величина признака (например, веса или удоя), которым будет обладать большее число членов генеральной совокупности в случайной выборке. Например, если имеется вариационный ряд, составленный по жирномолочности коров стада (табл. 8), то модальным классом будет класс, в котором жирномолочность равна 3,8% жира, имеющимся у наибольшего числа животных (p =30).

Таблица 8

Вариационный ряд распределения коров по показателю процента жира в молоке

% жира	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
Число коров (p)	3	5	10	30	20	15	10	10	5	2

Величина модального варианта может или совпадать со средней арифметической ($\overline{X}$) или отличаться от её значения. В данном примере (табл. 8) мода равна 3,8%, а средняя арифметическая – 3,9%.

Встречаются вариационные ряды, которые имеют два модальных класса и, следовательно, два значения моды. Использование модального варианта особенно удобно при изучении качественных признаков.

Медиана (*Me*) – это варианта, расположенная в середине вариационного ряда и делящая его на две равные части. Одна половина объектов совокупности будет иметь значения варьирующего признака меньше, а другая половина объектов – больше медианы. Использование медианы особенно целесообразно при обработке качественных признаков.

9.2.2. Основные показатели вариации

Средняя арифметическая отражает среднее значение признака в совокупности, но не характеризует его изменчивость. Если в одном стаде средний удой коров $\bar{x}$ = 3500 кг, во втором $\bar{x}$ = 3600, то только по значениям $\bar{x}$ невозможно что-либо сказать об изменчивости удоя. Поэтому важно знать не только средние показатели, но и вариацию признака в выборочной совокупности (отражающей варьирование признака в совокупности генеральной).

Разнообразие признаков в совокупности характеризуют следующие основные статистические параметры:

– лимиты;
– среднее квадратическое отклонение;
– варианса;
– нормированное отклонение;
– коэффициент вариации.

Лимиты (*lim*) – это максимальное и минимальное значения признака в совокупности. Чем больше разность между максимальной (max) и минимальной (min) вариантой, тем, в общем, сильнее изменчивость признака.

Например, у здоровых коров лимиты содержания лейкоцитов в 1 мм^3 крови *lim* = 11,4–3,2. У коров больных некробактериозом *lim* = 13,6–3,6. Таким образом, этот признак сильнее варьирует у больных коров (разность в первом случае составляет 8,2, а во втором – 10 тыс./мм^3). Вместе с тем, даже при одинаковых лимитах изменчивость в сравниваемых выборках может различаться, потому, что лимиты не учитывают распределения отдельных вариант в выборочной совокупности.

Варианса (σ^2) – это сумма квадратов отклонений отдельных вариант от средней арифметической, деленная на число степеней свободы:

$$\sigma^2 = \frac{\sum(x_i - \overline{x})^2}{n-1} \qquad (5)$$

где x_i – i-тая варианта выборочной совокупности; $\overline{x}$ – средняя арифметическая выборочной совокупности; n –объём выборочной совокупности

Таким образом, исходным началом для вычисления вариансы и среднего квадратического отклонения является *сумма квадратов отклонений* от $\overline{x}$, или просто сумма квадратов. На сумме квадратов отклонений построены очень многие формулы, применяемые в различных разделах биологической статистики.

Среднее квадратическое отклонение (стандартное отклонение) является более точным параметром для характеристики изменчивости. Среднее квадратическое отклонение обозначается буквой σ (сигма). Среднее квадратическое отклонение можно получить из вариансы:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n-1}} \qquad (6)$$

Эта величина выражается в тех же единицах, что и $\overline{x}$ (кг, см, % и т. д.). Среднее квадратическое отклонение показывает, насколько в среднем каждая варианта отклоняется от средней арифметической, вычисленной для данной совокупности. Чем больше величина σ, тем выше изменчивость изучаемого признака.

Значение σ является *именованным* числом. Поэтому с помощью *среднего квадратического отклонения* можно сравнить изменчивость двух и более групп животных в отношении *только одинаковых* признаков, но нельзя сравнивать изменчивости признаков, выраженных в разных единицах измерения, например, *изменчивость* живой массы коров с *изменчивостью* числа лейкоцитов в их крови. Например, на племзаводе получены следующие сведения о первой лактации коров: удой $\overline{x}$ =4240 кг и σ=748 кг, жирность молока этих коров составляла $\overline{x}$ =3,8% и σ=0,17%. Очевидно, что невозможно сравнить сигму, выраженную в кг (σ=748 кг) с сигмой, выраженной в процентах (σ=0,17%). С помощью σ также нельзя сравнивать изменчивость признаков в двух биологически разнородных группах. Например, нельзя сравнивать живую массу животных в двух выборках, если в одной выборке она вычислена для взрослых коров, а во втором – для телят при рождении или телят в возрасте 1, 2, 3 месяцев. По этим причинам для сравнения изменчивости признаков, выраженных в *разных единицах*, применяют другой показатель – коэффициент вариации, который выражает изменчивость в *относительных* единицах (в %).

Правило трёх сигм. Все значения признака лежат от средней арифметической в пределах 3σ. Поэтому эта закономерность называется «правилом плюс-минус трех сигм». Для того, чтобы определить возможные крайние значения варьирующего признака нужно среднюю арифметическую увеличить и уменьшить на три сигмы. Это даст практически крайние значения признака.

Понятие о числе степеней свободы [6]. Величина (n–1), фигурирующая в знаменателях дробей в формулах вычисления σ^2 и σ, получила особое название – **число степеней свободы** (точнее, число степеней свободы *вариации*). Так как во многих разделах биометрии приходится пользоваться числом степеней свободы, то следует хорошо уяснить его значение.

Если известен ряд от x_i до x_n состоящий из n членов или наблюдений, то для него общей характеристикой является средняя арифметическая. Как может быть определено каждое *отдельное значение* ряда? Очевидно, это конкретное значение признака можно узнать, если известны а) средняя арифметическая и б) остальные наблюдения, т.е. n–1. Иначе говоря, определение одного значения в данной совокупности зависит от остальных значений. Так, например, если известно, что 2 кролика в сумме весят 6 кг, а один из них весит 2,5 кг, то вес второго уже точно определен весом первого, т.е. имеется лишь 1 степень свободы (2–1 = 1). Если 3 кролика весят 5 кг, то вес одного всегда точно определяется весом двух других, между которыми *уже возможна вариация*, т.е. в этом случае имеются 2 степени свободы вариации (3–1=2) и т. д.

Другой пример. От гетерозиготного растения гороха путём самоопыления получено 36 семян только 2 классов (желтые и зеленые). Известно, что 25 из них желтые. Если их отобрать, то все оставшиеся будут только зеленые, и их должно быть 11. Таким образом, из суммы двух слагаемых одно взято свободно (**одна** степень свободы), а второе определилось автоматически. Если в F_2 имеет место расщепление на три фенотипических класса (***AA* : *Aa* : *aa***), то после учета одного из них останется еще возможность выбора второго класса, а третий класс будет определен автоматически, т.е. при трех слагаемых два берутся свободно (**две** степени свободы), а третье имеет зависимое значение и т. д.

Таким образом, число степеней свободы при анализе классов расщепления всегда будет на единицу меньше числа последних т.е.

если n – число классов, то число степеней свободы (ν) будет равно (n–1).

В общем виде при численности членов совокупности n число степеней свободы $\nu = n–1$. Вот почему точнее вычислять σ и σ^2, пользуясь знаменателем n–1. При большом n разница между n и (n–1) мала. Поэтому она мало отразится на значении вариансы и сигмы. Но при малом n разница будет значительна. Так, если n = 6, а сумма квадратов отклонений равна 60, то средний квадрат отклонений от средней арифметической будет равен не 60:6=10, а 60:5=12. Поэтому надо разделить сумму квадратов отклонений на число степеней свободы, т.е. на n–1 = 5.

В некоторых типах статистических анализов число степеней свободы вычисляется более сложно.

Коэффициент вариации (C_v) показывает, какой процент от $\bar{x}$ составляет σ. Другими словами, коэффициент вариации показывает изменчивость признака в относительных величинах (в процентах):

$$C_v = \frac{\sigma}{\bar{x}} \cdot 100\% \qquad (7)$$

Чем больше значение C_v тем больше изменчивость признака у членов совокупности.

В селекционной работе очень важно знать изменчивость признака, ибо это дает правильное суждение о том, на сколько легко или трудно закрепить тот или иной признак, или насколько широко будет поле отбора животных с уклонением от средней в сторону +3σ.

Заключения о консерватизме наследственности у чистопородных животных и о расшатанной наследственности у помесей легче сделать, если проведено вычисление коэффициента вариации по ряду признаков у животных разных степеней инбридинга. Также удобно пользоваться этим коэффициентом для выявления пластичности признака к воздействиям внешней, среды. Например, удой и процент жира в молоке имеют разный коэффициент вариации; удой более изменчив, а процент жира – менее. Эти различия, выраженные через C_v и будут ориентировать животновода на то, что внешними факторами, например кормлением, легче изменить уровень *удоя* в стаде, чем уровень *жирномолочности*. Такой вывод, сделанный на основании статистического анализа изменчивости признаков, подтверждается и зоотехнической практикой.

В характеристике выборочной совокупности коэффициент вариации является *дополнительным* показателем и поэтому должен применяться *вместе* с основными параметрами $\bar{x}$ и σ.

Нормированное отклонение (t) – это отклонение конкретной варианты (или группы вариант) от средней арифметической, выраженной в сигмах. Это означает, что среднее квадратическое отклонение позволяет выразить в *относительных величинах* степень отклонения какого-либо члена совокупности от средней арифметической данного ряда.

Для вычисления t берут разность между значением данной варианты x_i и значением средней арифметической $\overline{X}$ и делят эту разность на среднее квадратическое отклонение σ. Полученное в результате такого деления выражение:

$$t = \frac{x_i - \overline{x}}{\sigma} \qquad (8)$$

и называют нормированным отклонением данного члена совокупности от средней арифметической в долях сигмы.

Чем больше значение нормированного отклонения, тем дальше от $\overline{X}$ отстоит данная варианта. Минимальное и максимальное значение вариант, как уже указывалось ранее, имеет нормированное отклонение, равное –3σ и +3σ.

Таблица 9

Использование нормированного отклонения для заключения о выздоровлении животного по показателям крови [4, с.61]

Показатели крови	Норма для здорового животного		Показатели крови больного животного в период лечения и их нормированное отклонение по декадам					
			1-я декада		2-я декада		3-я декада	
	$\overline{X}$	σ	x_i	t	x_i	t	x_i	t
% гемоглобина	80	5	60	–4σ	70	–2σ	80	0σ
Число лейкоцитов	6000	500	8000	+4σ	7000	+2σ	6500	–1σ
% белков в сыворотке крови	6,0	1,0	4,0	–2σ	5,0	–1σ	5,5	–0,5σ

Удобство использования нормированного отклонения заключается в том, что каждый член совокупности может быть комплексно охарактеризован по нескольким признакам в долях σ. Кроме того, нормированное отклонение позволяет сравнивать объекты из разных совокупностей по одноименному признаку. Например, можно сравнить жирномолочность коровы чистопородной и помесной, используя их нормированные отклонения.

Продемонстрируем сказанное на примерах из ветеринарной практики. Больное животное подвергалось лечению препаратом антибиотика. Для суждения о его выздоровлении был взяты три показателя: 1) процент гемоглобина, 2) число лейкоцитов и 3) процент общего белка в крови. Пробы крови брали 3 раза на протяжении месяца и сравнивали с уровнем этих показателей, типичных для животных при нормальном состоянии. Требуется выяснить ход выздоровления и установить было ли полученное выздоровление к концу месяца результатом применения антибиотика. Исходные данные и вычисление нормированного отклонения приведены в табл. 9.

Из таблицы 9 видно, что в начале болезни животное имело показатели крови, отклоняющиеся по проценту гемоглобина на –4σ, по лейкоцитам на +4σ и по общему белку сыворотки крови на –2σ. По мере лечения эти показатели сближались со средней у нормальных животных: к концу месяца достиг нормы процент гемоглобина; незначительно отклонялся (на –0,5σ) от нормы процент сывороточного белка, но еще оставалось повышенным число лейкоцитов (+1σ).

Нормированное отклонение удобно применять в зоотехнической работе при оценке производителей по качеству их потомства, когда показатели каждого потомка выражаются в относительных величинах, т.е. в долях σ по отношению к среднегрупповым показателям. Например, если дочери одного быка по удою отклоняются от среднего удоя сверстниц на +2σ; +1,5σ; +1,6σ ; +1,2σ, а дочери другого быка отклоняются от средней продуктивности сверстниц на +0,5σ; +0,25σ; +1,1σ; +1,5σ, то второй производитель будет менее ценным, чем первый.

Вычислим нормированное отклонение трех помесных и трех чистопородных коров по показателю их жирномолочности (табл. 10).

Сопоставление двух групп коров разных пород, приведенное в табл. 10, показывает, что по отношению к показателям *своей* группы помесные коровы были относительно лучше, чем чисто породные, если сравнивать их по средним показателям жирномолочности. В пределах каждой пары (1-я чистопородная – 1-я помесная; 2-я чистопородная – 2-я помесная и т.д.) коровы не отличались а) друг от друга и б) по абсолютной разнице их жирномолочности от среднего ее уровня. Так, для первой пары разность составляла –0,3% жира (3,2–3,5= –0,3% и 4,2–4,5= –0,3%). Для второй пары сравниваемых коров разность равна –0,1% жира (3,4–3,5= –0,1% и 4,4–4,5=–0,1%). Для третьей пары разность составляет +0,1% (3,6–3,5=0,1% и 4,6–4,5=0,1%). Но по величине нормированного отклонения t они имели разную величину изменчивости, что хорошо видно в соответствующих графах таблицы 10.

Таблица 10

Сопоставление жирномолочности помесных и чистопородных коров по величине нормированного отклонения t.

Группы коров	Средняя жирномолочность в каждой породе, %		Показатели конкретных коров					
			1-я корова		2-я корова		3-я корова	
	$\bar{x}$	σ	x_i	t	x_i	t	x_i	t
Чистопородные	3,5	0,1	3,2	–3,0σ	3,4	–1,0σ	3,6	+1,0σ
Помесные	4,5	0,2	4,2	–1,5σ	4,4	–0,5σ	4,6	+0,5σ

Нормированное отклонение t – очень важный показатель для выявления типа распределения членов совокупности по классам варьирующего признака. Типы распределения будут рассмотрены позже.

9.2.3. Вычисление средней арифметической

Средняя арифметическая является одним из основных статистических показателей. Она представляет собой как бы точку равновесия варьирующего признака, отклонения от которой вправо или влево взаимно уравновешиваются. Средняя арифметическая показывает, какую величину признака имели бы животные данной совокупности, если бы он был у всех одинакового значения.

Среднюю арифметическую для выборочной совокупности принято обозначать символом $\bar{x}$. В некоторых учебниках биометрии эту величину обозначают буквой M.

В зависимости от объёма выборочной совокупности среднюю арифметическую вычисляют двумя различными способами (прямым и непрямым). Принципиальные различия этих методов заключаются в разных способах определения достоверности получаемых статистических параметров, а также в том, что для большого числа наблюдений составляют вариационные ряды, для малого же числа наблюдений в биометрических расчетах непосредственно используют результаты измерений.

9.1.1. Прямой способ вычисления средней арифметической

Для малых выборок (n<30) $\bar{x}$ вычисляют *прямым методом* по формуле (1), указанной ранее:

$$\bar{x} = \frac{\sum x_i}{n}$$

Пример. Требуется определить среднюю живую массу семи дочерей курицы № 112, каждая из которых имеет массу (в кг): 1,8, 2,1, 1,7, 1,9, 1,8, 1,9, 2,0.

$\bar{x}$ =(1,8+2,1+1,7+1,9+1,8+1,9+2,0) **:** 7 = 13,2 **:** 7 = 1,89 ≈ 1,9 кг.

В данном примере сумма живых масс всех дочерей равна 13,2 кг, число дочерей –7, средняя живая масса – 1,9 кг.

9.2.5. Непрямой способ вычисления средней арифметической

По формуле (1) можно вычислить среднюю арифметическую для групп любой численности, если имеется вычислительная техника. Если ее нет, то при большом числе вариант применяют другой, *непрямой метод* вычисления средней арифметической. В этом методе используют построение вариационного ряда и используют *условную среднюю* по формуле:

$$\bar{x} = A + k \cdot \frac{\sum f \cdot a}{n} \qquad (9)$$

где A – условная средняя; k – величина классового промежутка; f – число вариант в классе; a – условное отклонение средних значений классов от выбранной условной средней; n – число наблюдений.

В основе этого метода лежит математическое положение, что сумма отклонений от средней арифметической всегда равна нулю. Если же сумма отклонений получена от *условной* средней, то она будет или больше, или меньше на величину, отличающую условную среднюю от средней арифметической.

Если сумма положительных и отрицательных отклонёний от A окажется равной нулю, то условная средняя A полностью совпадает с истинной средней арифметической $\bar{X}$. Если сумма всех отклонений окажется величиной положительной, значит, принятая условная средняя меньше истинной. Если же сумма всех отклонений будет величиной отрицательной, принятая условная средняя больше истинной. В обоих случаях для того, чтобы перейти от условной средней **A** к средней арифметической $\bar{X}$, надо внести в принятую величину A поправку с тем или иным знаком.

Алгоритм вычисления средней арифметической непрямым методом (методом условной средней) мы рассмотрим, используя данные о содержании лейкоцитов в крови 70 коров (табл.1, с. 16). Для вычисления средней арифметической строят такую же таблицу, как и при построении вариационного ряда, но в неё добавляют несколько дополнительных столбцов (табл. 11).

Вначале выполним первые 5 этапов построения вариационного ряда (с. 16-19) и заполним графы 1, 2 и 4 в таблице 11.

6) Вычислим средние значения каждого класса (w_i) вариационного ряда (графа 3 таблицы 11). Это можно сделать двумя способами. Способ первый: складываем минимальное и максимальное значения каждого класса и делим полученную сумму на два. Например, для первого класса среднее значение получим, (3,0+3,9)**:** 2=3,45 ≈3,5; для модального класса среднее значение получим, (7,0+7,9)**:** 2=7,45 ≈7,5. При желании можно проводить вычисления, не округляя средние значения классов. Способ второй: к значению нижней границы каждого класса прибавляем 1/2 величины классового интервала. В нашем примере

Таблица 11

Вспомогательная таблица для вычисления средней арифметической методом условной средней.

№№ классов п/п	Границы классов	Среднее значение класса, w_i	Частоты вариант f	a, отклонение w_i от A	$f·a$	$f·a^2$
1	2	3	4	5	6	7
1	3,0 – 3,9	3,5	2	–4	–8	32
2	4,0 – 4,9	4,5	5	–3	–15	45
3	5,0 – 5,9	5,5	8	–2	–16	32
4	6,0 – 6,9	6,5	12	–1	–12	12
5	7,0 –7,9	7,5=***A***	17	0	0	0
6	8,0 – 8,9	8,5	14	+1	+14	14
7	9,0 – 9,9	9,5	7	+2	+14	28
8	10,0 – 10,9	10,5	4	+3	+12	36
9	11,0 – 11,9	11,5	1	+4	+4	16
			n=70		Σ $f·a$ = –7	Σ $f·a^2$ = 215

классовый интервал равен 1,0; половина классового интервала будет равна 0,5. Прибавляем эту величину к нижней границе каждого интервала и получаем те значения, которые указаны в графе 3 таблицы 11.

7) Выбираем *условную среднюю* (A). Эта величина должна быть близкой к *истинной* средней арифметической выборки. Вероятнее всего, средняя арифметическая выборки будет находиться в *модальном классе*. Поэтому на практике в качестве условной средней обычно выбирают среднее значение модального класса. В нашем примере – это пятый класс по порядку со средним значением, равным 7,5.

8) Вычислим вспомогательные величины (*a*) – отклонение *среднего значения* каждого класса от условной средней. Вычисления выполняют по формуле

$$a = \frac{w_i - A}{k}. \qquad (10)$$

Начнём не с первого класса, а с модального (пятого по порядку). В этом классе $w_i = A = 7{,}5$. Подставив эти значения в формулу (10), получим $a = 0$. В шестом классе a окажется равным (8,5 – 7,5) **:** 1 = +1. Подставляя соответствующие значения w_i седьмого, восьмого и девятого классов в формулу получим значения +2, +3 и +4. Обратимся теперь к тем классам, средние значения которых меньше условной средней A. Подставляя средние значения первого-четвёртого классов в формулу (10), мы получим значения (–1), (–2), (–3), (–4). Такую картину распределения значений a, вы будете наблюдать во всех своих расчётах. Это позволяет для практических расчётов определять величину a как отклонение порядкового номера каждого класса от номера класса, в котором находится условная средняя [5].

9) Вычислим величину *произведений* $f{\cdot}a$, т.е. (частота каждого класса) × (отклонение среднего значения этого же класса). При вычислениях следите за правильностью знаков «+» и «–». Полученные величины вносим в соответствующие строки графы 6. После выполнения этих вычислений определим сумму произведений ($\Sigma f{\cdot}a$).

10) Подставим вычисленные величины в формулу (9, с.28):

$$\bar{x} = A + k \cdot \frac{\sum f \cdot a}{n}$$

$$\bar{x} = 7{,}5 + 1{,}0 \cdot \frac{-7}{70} = 7{,}5 + 1{,}0 \cdot (-0{,}1) = 7{,}5 - 0{,}1 = 7{,}4 \text{ тыс/мм}^3.$$

Таким образом, мы вычислили среднее арифметическое для нашей выборки.

Наряду с $\bar{x}$ очень важным статистическим показателем, характеризующим выборочную совокупность, является среднее квадратическое отклонение σ. При использовании непрямого метода вычисления $\bar{x}$ среднее квадратическое отклонение также приходится вычислять по более сложной формуле (11):

$$\sigma = \sqrt{\sigma^2} = k \cdot \sqrt{\frac{\sum f \cdot a^2}{n} - \left(\frac{\sum f \cdot a}{n}\right)^2} \qquad (11)$$

Обратите внимание на *первую* дробь в подкорневом выражении формулы (11). Её числитель представляет собой сумму произведений частот вариант в каждом классе (f) на *квадрат* соответствующих отклонений средних значений класса от условной средней (a^2). Для вычисления этой суммы необходимо найти произведения указанных величин в каждом классе (графа 7 в таблице 11), а затем найти их сумму. Во второй дроби подкорневого выражения находится уже вычисленная нами сумма произведений частоты вариант в классах на их условные отклонения. Обратите внимание на то, что эту сумму вначале необходимо разделить на n, и лишь потом (!) частное от деления возвести в квадрат:

$$\sigma = 1{,}0 \cdot \sqrt{\frac{215}{70} - \left(\frac{-7}{70}\right)^2} = 1{,}0 \cdot \sqrt{3{,}07 - (-0{,}1)^2} =$$

$$1{,}0 \cdot \sqrt{3{,}07 - 0{,}01} = 1{,}0 \cdot \sqrt{3{,}06} \approx 1{,}7 \text{ тыс./мм}^3.$$

Зная, среднюю арифметическую исследуемого признака и среднее квадратическое отклонение легко вычислить коэффициент вариации (формула 7):

$$C_v = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{1{,}7}{7{,}4} \cdot 100\% \approx 30\%.$$

Ниже показан ещё один пример вычисления средней арифметической методом условной средней для признака «живая масса» кур. В этом примере приводится уже построенный вариационный ряд (см. табл. 12 на с. 36).

Для упрощения расчетов за условную среднюю лучше принимать или *центральное значение* вариационного ряда, находящееся в его середине, или значение, включающее наибольшее число вариант. Так, в расчете средней живой массы кур (табл. 12) в качестве условной средней величины принято центральное значение класса – $(1{,}81+1{,}90):2 \approx$ $\approx 1{,}86$ кг, включающего наибольшее число кур. Расчёт среднего квадратического отклонения выполнен по формуле (11).

Таблица 12

Вычисление средней арифметической живой массы кур породы белый леггорн

Классы кур по живой массе (кг)	Среднее значение класса	Число кур (f)	Отклонение (a)	$f \cdot a$	$f \cdot a^2$
1,51–1,60	1,56	3	–3	–9	27
1,61–1,70	1,66	11	–2	–2	44
1,71–1,80	1,76	21	–1	–21	21
1,81–1,90	1,86=A	25	0	0	0
1,91–2,00	1,96	20	1	20	20
2,01–2,10	2,06	13	2	26	52
2,11–2,20	2,16	4	3	12	36
2,21–2,30	2,26	1	4	4	16
k=0,1 кг		n=98		Σfa=10	$\Sigma f \cdot a^2$= 216

$$A = \frac{1{,}81 + 1{,}90}{2} = 1{,}855 \approx 1{,}86$$

$$\bar{x} = 1{,}86 + 0{,}1 \cdot (10 : 98) = 1{,}86 + 0{,}1 \cdot 0{,}1 = 1{,}86 + 0{,}01 = 1{,}87 \text{ кг}$$

$$\sigma = 0{,}1 \cdot \sqrt{\frac{216}{98} - \left(\frac{10}{98}\right)^2} = 0{,}1 \cdot \sqrt{2{,}2 - 0{,}01} =$$

$$= 0{,}1\sqrt{2{,}19} = 0{,}1 \cdot 1{,}48 \approx 0{,}15 \text{ кг.}$$

$$C_V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{0{,}15}{1{,}87} \cdot 100\% = 0{,}08 \cdot 100\% = 8\%$$

Использование средней арифметической в селекционно-племенной работе позволяет характеризовать любое стадо, линию или семейство животных по уровню средней продуктивности или другому биологическому показателю. Важную роль играет средняя арифметическая в оценке производителей по качеству потомства. В этих случаях среднюю арифметическую селекционного признака по группе дочерей (сыновей) оцениваемого производителя сопоставляют со средней арифметической этого же признака по группе матерей, сверстниц, по стаду. Сопоставление средних арифметических позволяет вычислить разность между уровнями продуктивности групп и сделать вывод о качестве производителя на основании полученной разности. Кроме этого, средние арифметические, вычисленные по показателям родителей и потомства, могут быть использованы для выявления характера наследования количественных признаков, что позволяет изменять их в нужном направлении.

9.2.6. Вычисление ошибки средней арифметической

Вычисленная средняя арифметическая выборочной совокупности, как правило, не совпадает по значению со средней арифметической для целой популяции (т.е. генеральной совокупности). Это связано с тем, что средняя арифметическая для выборки рассчитана на основании изучения ограниченного числа объектов. При взятии нескольких выборок, даже из одной и той же популяции, получают величины, хотя и близкие по значению к ранее вычисленной средней арифметической, но все же другие. Поэтому возникает необходимость в вычислении **ошибки средней арифметической**, которая фактически является *стандартным отклонением* средней арифметической. В биометрической литературе эту величину обозначают буквой m (в некоторых пособиях по биометрии используют обозначение $s_{\overline{X}}$). Ошибку средней арифметической вычисляют по формуле (12)

$$m = \frac{\sigma}{\sqrt{n}} \qquad (12)$$

Величина ошибки средней арифметической зависит от размаха изменчивости признака и размера выборочной совокупности. Чем меньше размах изменчивости и чем больше объём выборки, тем меньше ошибка.

В нашем примере с числом лейкоцитов в крови коров ошибка равна

$$m = 1{,}7 : \sqrt{70} = 1{,}7 : 8{,}37 = 0{,}20.$$

Во втором примере с живой массой кур породы леггорн ошибка будет равна

$$m = 0{,}15 : \sqrt{98} = 0{,}15 : 9{,}90 = 0{,}02.$$

При необходимости могут быть рассчитаны ошибки среднего квадратического отклонения m_σ:

$$m_\sigma = \frac{\sigma}{\sqrt{2n}} \qquad (13)$$

и ошибка коэффициента вариации m_{Cv} [1, *с. 133*; 2, *с. 194*]

$$m_{C_v} = \frac{C_v}{\sqrt{2n}}. \quad (14)$$

Таким образом, использование статистических показателей при анализе выборочной совокупности позволяет точнее охарактеризовать изменчивость анализируемого признака. Однако точная характеристика изменчивости признака изучаемой выборки животных – не самоцель. Далее мы узнаем, что сравнение статистических показателей, вычисленных для разных выборок, позволяет сравнивать эти выборки и анализировать степень их сходства и различия. Такие сравнения очень важны в селекционной практике, при зоотехническом анализе различных рационов, сравнении эффективности типов и доз лекарственных препаратов и во многих других случаях практической деятельности специалистов.

СЛОВАРЬ НОВЫХ ТЕРМИНОВ

Биометрия – наука о способах применения математических методов в биологии; наука о статистическом анализе массовых явлений в биологии, т.е. таких явлений, в массе которых обнаруживаются закономерности, не выявляемые на единичных случаях наблюдений.

Варианса (σ^2) – статистический показатель, равный сумме квадратов отклонений отдельных вариант от средней арифметической, деленной на число степеней свободы.

Варианта – значение или мера признака для конкретного члена выборочной совокупности

Вариационная кривая (полигон распределения) – график, образованный ломаной линией, прямые отрезки которой соединяют точки, значениям частот вариант в соответствующем классе вариационного ряда.

Вариационный ряд – упорядоченное изображение реально существующего распределения особей в группе по величине исследуемого признака; ряд ранжированных значений признака, в котором указана частота встречаемости этих значений в данной совокупности.

Выборка – то же, что и *совокупность выборочная*.

Гистограмма – график, на котором по оси абсцисс (x) отображены классы, а по оси ординат (y) – частоты вариант в соответствующем классе в виде столбиков, высота которых соответствует значениям частот.

Единицы наблюдения – биологические объекты, группы которых изучают по одному или нескольким варьирующим признакам Единицами наблюдений могут быть особи одного вида или породы, особи одного пола, одного возраста и т.п.

Классный промежуток (размер класса) – величина, равная частному от деления разности максимального и минимального значений вариант на планируемое в вариационном ряду число классов.

Коэффициент вариации (C_v) статистический показатель, указывающий на то, какой процент от $\overline{x}$ составляет σ.

Лимиты – максимальное и минимальное значения признака в совокупности.

Медиана – варианта, расположенная в середине вариационного ряда и делящая его на две равные части.

Мода – наиболее часто встречающееся значение варианты в данной выборочной совокупности

Нормированное отклонение (t) – это отклонение конкретной варианты (или группы вариант) от средней арифметической, выраженной в сигмах.

Объём выборки – см.: *объём выборочной совокупности.*

Объём выборочной совокупности (объём выборки) – число единиц наблюдений (n) в выборочной совокупности. Если n <30 выборочную совокупность называют **малой**; если n >30 выборочную совокупность называют **большой**.

Признак – характерная особенность строения или функционирования организма.

Признак мерный (метрический) – признак, классифицируемый на основе измерения, варьирующий непрерывно и величина которого изменяется, принимая и целые, и дробные значения.

Признак счётный (меристический) – признак, классифицируемый на основе счёта, варьирующий прерывисто и значение которого выражаются только целыми числами.

Репрезентативная выборка (выборочная совокупность) – выборочная совокупность созданная случайным образом, имеющая большой объём, и поэтому достаточно точно отражающая изучаемые свойства генеральной совокупности.

Совокупность выборочная – та часть индивидуально различимых единиц, у которых мы можем изучить интересующий нас признак.

Совокупность генеральная – все индивидуально различимые единицы (биологические объекты) с исследуемым признаком, существующие в природе; вся подлежащая изучению совокупность данных объектов.

Совокупность статистическая – совокупность относительно однородных, но индивидуально различимых единиц, объединяемых по некоторым общим признакам для их совместного изучения.

Среднее квадратическое отклонение (σ, стандартное отклонение) – величина равная квадратному корню из значения вариансы.

Средняя арифметическая признака **–** математически вычисляемое значение признака, наиболее характерное для данной совокупности.

Средняя геометрическая – средняя величина, которая выявляет средний прирост (или среднее уменьшение) какого-либо показателя за определённый период времени.

ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

1. Что называется биометрией?

1.1. Что является предметом биометрии?

2. Понятие о совокупности.

2.1.Что называется статистической совокупностью?

2.2.Что называется выборочной совокупностью?

2.3.Что называется объёмом выборочной совокупности?

2.3.1. Какая выборочная совокупность называется малой?

2.3.2. Какая выборочная совокупность называется большой?

2.4.Что называется генеральной совокупностью?

3. Что называется признаком?

3.1.На какие два класса разделяют признаки?

3.2.Какие признаки называются мерными, какие – счётными?

3.3.Что называют вариантой?

3.4.Что является источником варьирования признаков?

4. Что называют вариационным рядом?

4.1.Что называют классом вариационного ряда?

4.2.Опишите алгоритм построения вариационного ряда.

4.3.Как графически отображается вариационных ряд?

4.4.Как распределяются варианты в вариационном ряду?

4.5.Что называется гистограммой?

4.6.Что называется полигоном распределения (вариационной кривой)?

4.7.Какой класс называется модальным?

4.8.Почему возникают двухвершинные (многовершинные) вариационные ряды?

4.9.Как можно устранить многовершинность вариационного ряда?

4.10. Почему возникают асимметричные вариационные ряды?

5.Какие статистические показатели характеризуют совокупность?

5.1.Что называется средней арифметической вариационного ряда?

5.1.1. Что показывает средняя арифметическая?

5.2.Что называется средней геометрической?
 5.2.1. Для чего используется средняя геометрическая?
5.3.Что называется средней квадратической?
 5.3.1. Для чего используется средняя квадратическая?
5.4.Что называется средней гармонической?
 5.4.1. Для чего используется средняя гармоническая?
5.5.Что называется модой?
5.6.Что называется медианой?

6. Какие показатели вариации являются основными?

6.1. Что называется лимитами?
6.2. Что называется варинсой?
6.3. Что называется средним квадратическим отклонением?
6.4. Почему нельзя среднее квадратическое отклонение использовать для сравнения варьирования различных признаков?
6.5. Сформулируйте правило трёх сигм.
6.6. Сформулируйте понятие степеней свободы.
6.7. Что называется коэффициентом вариации?
6.8. Что называется нормированным отклонением?

7. Как вычисляется средняя арифметическая выборочной совокупности?

7.1.Как вычисляется средняя арифметическая для малой выборочной совокупности?
7.2.Как вычисляется средняя арифметическая для большой выборочной совокупности?
7.3.Опишите алгоритм вычисления средней арифметической непрямым методом.
7.4.Что называется средним значением класса?
7.5.Как выбирают условную среднюю ***A***?
7.6.Как определяют условное отклонение ***a*** ?
7.7.Как вычисляется среднее квадратическое отклонение при расчёте $\overline{x}$ непрямым методом?
7.8.Что называется ошибкой средней арифметической?

ЗАДАЧИ

1. У 60 валахских овец была измерена длина правого уха (в см):

12	10	14	14	13	12	12	12	15	13
11	12	12	14	12	11	13	12	13	14
11	13	14	12	13	12	12	14	12	14
13	13	12	13	12	13	12	11	11	12
13	14	12	14	13	14	13	12	14	15
10	11	10	11	15	11	16	11	11	11

Составьте вариационный ряд и постройте полигон распределения. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

2. Ширина ушной раковины в стаде овец была следующей (в см):

5,8	6,2	6,3	6,1	6,1	5,7
6,5	6,0	6,1	5,8	6,3	6,2
6,2	5,4	5,9	6,0	5,7	5,9
6,1	6,3	6,2	6,5	6,2	6,7
6,2	5,7	6,1	5,7	5,9	6,0
5,7	5,9	6,1	5,9	6,0	6,1

Составьте вариационный ряд и начертите гистограмму. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

3. Количество яиц, отложенных волнистыми попугайчиками, в зоопарке было следующим (***n***=100):

4	5	4	5	5	4	5	4	3	5
6	4	4	4	5	5	3	5	5	4
6	2	3	4	5	5	5	5	5	5
5	6	4	6	2	5	5	3	5	5
6	4	5	5	5	5	5	5	5	5
4	6	7	6	3	5	5	6	5	5
4	2	4	4	6	2	6	5	4	5
5	5	4	5	4	6	5	4	7	5
6	6	4	4	4	6	5	4	3	5
5	5	5	5	4	3	7	6	4	4

Составьте вариационный ряд. Постройте полигон распределения. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

4. Представьте в виде вариационного ряда и графически данные о живой массе телят в конце одного научно-производственного эксприимента (в кг):

58	57	64	61	56	65	63	58	63
60	59	61	54	58	66	67	63	63
61	60	58	57	65	61	60	68	64
63	56	59	64	61	64	57	60	63
58	52	60	59	57	61	54	58	64
62	59	60	63	60	60	64	59	63
63	59	62	63	61	65	61	64	57
59	54	64	63	57	59	59	58	63
62	63	62	62	60	62	57	56	60
63	57	63	61	59	61	59	60	59

Составьте вариационный ряд. Постройте вариационную кривую. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

5. Было подсчитано число лучей в хвостовых плавниках камбалы:

53	51	52	55	56	49	51	52	53	56
54	53	52	53	51	55	53	55	53	54
51	51	56	54	54	53	54	54	55	53
52	55	53	53	56	53	52	56	52	52
56	55	50	54	49	54	54	55	54	55
52	51	55	52	55	54	51	54	53	54
54	56	54	55	53	53	56	55	54	53
55	52	53	52	51	55	53	54	51	50
53	54	55	52	55	52	53	50	53	52
58	57	57	58	56	57	56	58	57	57

Составьте вариационный ряд и начертите полигон распределения. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

6. При взвешивании новорождённых телят холмогорской породы были получены следующие данные (***n*** = 80; min=23; max=45)

27	28	36	32	27	32	31	26
32	33	34	26	35	23	27	35
28	35	30	36	37	35	31	31
35	36	35	28	34	32	33	27
36	32	36	33	24	28	26	25
32	34	28	26	35	36	32	35
34	26	30	36	34	30	39	33
32	32	34	26	32	28	28	26
39	33	37	38	33	37	31	37
43	30	45	37	32	35	40	30

Постройте вариационный ряд, разбив его на 7 классов. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

Постройте вариационный ряд, разбив его на 9 классов. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

Выясните – изменятся ли статистические показатели при изменении числа классов в вариационном ряду?

7. Длина тела у 77 экземпляров плотвы, отловленных в верховьях р. Орлик была следующей (в мм):

143	157	148	153	150	142	164	139	139	140
143	120	144	130	138	124	127	137	139	129
128	119	120	138	130	114	126	138	117	132
130	145	140	153	137	142	145	137	141	125
143	138	140	135	135	139	125	137	131	120
127	118	120	124	134	111	132	133	100	132
143	134	138	130	135	133	134	151	107	110
94	95	142	148	136	165	172			

Составьте вариационный ряд и гистограмму. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической

8. Изучен живой вес 60 телят холмогорских помесей при рождении:

27	32	32	31	32	28	37	35	26	28
32	39	34	30	37	26	27	40	35	37
28	43	26	35	45	26	35	32	32	35
35	28	32	36	32	36	37	33	28	31
36	33	33	28	23	26	34	32	36	27
32	39	36	38	24	32	30	31	35	33

Составьте вариационный ряд и изобразите его на графике. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

9. Масса тела студентов-зоотехников (***n*** = 80) 2-го курса составила:

80	75	75	85	78	85	80	77	83	85
88	94	95	86	80	73	78	90	95	90
80	75	83	70	78	83	75	78	86	81
62	77	75	73	80	80	74	73	82	72
80	90	80	78	60	65	75	72	64	67
74	80	68	75	76	65	70	78	75	83
85	70	88	73	56	75	70	73	68	66
65	68	75	78	63	68	68	70	60	56

Составьте вариационный ряд и начертите гистограмму. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

9а. Масса тела студенток-зоотехников (***n*** = 80) 2-го курса составила:

70	71	60	65	73	70	78	69	75	68
68	51	70	65	68	62	61	80	65	83
73	58	63	62	65	55	51	60	63	70
81	73	80	75	72	77	79	75	70	70
81	75	68	72	84	89	84	83	88	90
65	73	78	70	72	81	76	70	71	78
67	75	74	69	68	77	69	57	72	68
73	55	50	70	67	59	62	70	85	75

Составьте вариационный ряд и начертите гистограмму. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

10. Составьте вариационный ряд и изобразите его графически для следующих данных об удоях 204 коров за 300 дней лактации (в кг):

3586	2761	2825	3807	3858	2390
3453	2635	3752	2666	3331	3315
3165	2361	4055	2440	2763	3559
2148	2144	2856	2293	3246	2064
2358	2769	2849	3420	2833	2821
2580	3468	2035	3027	3177	3401
3115	2334	3339	2033	1850	4207
3847	2934	3676	4155	3306	3431
3465	2540	4288	3685	4708	2728
4052	3380	3154	4571	1426	2540
2340	2520	2855	2600	3711	3354
1360	2192	2690	3390	3350	3491
3445	3170	2271	2007	2107	4853
3458	1915	3090	1917	3382	4106
3651	4174	1274	2247	3859	4580
4659	4985	2132	3047	4582	1548
3021	4194	2654	3001	5190	2815
3148	3015	1785	2088	2026	2665
1450	2118	2936	4510	4216	2620
1501	2454	3287	1965	1563	2973
3904	3530	1951	2362	2729	3230
2923	2948	3428	2574	2581	3564
2838	2893	2461	4791	4011	4305
2955	3920	3205	2949	2559	5235
3528	3250	1474	2632	2108	4507
3666	3242	2715	2730	2748	2340
2093	3642	3736	3847	4080	3936
3373	2199	2468	2448	3293	2562
3758	2735	3363	3306	3511	3043
2981	3224	1480	1586	1953	4980
3073	3708	4167	4526	1600	4773
3009	3940	3510	3658	2326	2331
4901	3002	2934	3007	1687	1420
3684	3849	3211	3656	1966	2954

Составьте вариационный ряд и начертите полигон распределения. Вычислите среднюю арифметическую, среднее квадратическое отклонение, коэффициент вариации, ошибку средней арифметической.

9.3. ЗАКОНЫ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Изучая принципы построения вариационного ряда, мы познакомились с понятием статистического распределения. **Статистическое распределение** – это распределение частот f_i отдельных вариант по значениям этих вариант x_i или по классам вариационного ряда. Такие статистические распределения могут быть нескольких различных типов. Для каждого типа распределения характерны свои *особенности* и *закономерности*. Каждый тип распределения отражает причины, вызывающие варьирование признака, и может быть изображен графически. Ниже мы познакомимся с основными тремя типами статистических распределений: нормальным, биномиальным и распределением Пуассона, а также с часто встречающимися в практической работе асимметричным, эксцессивным и трансгрессивным распределениями.

9.3.1. Вероятность и её определение

Биометрия имеет дело не с единичными явлениями или объектами, а с их совокупностями. Отдельные члены совокупности, обычно в той или другой степени отличаются друг от друга, варьируют. Каждый из членов совокупности представляет собой реализацию определённого генотипа. Эта реализация осуществляется под влиянием многих определяющих причин. Однако этих причин может быть так много, что обнаружение их для каждого отдельного случая становится невозможным.

Каждое явление, рассмотренное отдельно, является случайным, но, взятые в массе, они обнаруживают определенные, так называемые *статистические закономерности.* Это позволяет предсказать результаты для массового явления в целом. В отношении же каждого единичного явления, каждого отдельного члена совокупности приходится говорить только об известной возможности, или вероятности, значения, которое они приобретают. Если зоотехник работает с крупным рогатым скотом холмогорской породы и знает, что средняя жирность молока коров этой породы около 3,5%, то для него ясно, что возможность, или вероятность, найти корову холмогорской породы с жирностью молока 3,5% очень велика. В то же время встретить корову этой породы с жирностью молока 4,5% маловероятно.

В приведённом выше примере оценка вероятности как возможности кажется очень ясной и понятной. Но так бывает далеко не всегда. Рассмотрим другой пример. Ветеринарный врач зверосовхоза применил для лечения заболевших какой-то болезнью лисиц новое лекарство. Выяснилось, что из лисиц, получивших лекарство, погибло только 4%. Среди не получавших лекарства лисиц отход был равен 13%. От чего же зависит разница в проценте отхода? От того ли, что в опытной группе применили новое лекарство, или, может быть, лекарство никакой роли не сыграло?

Когда нет точной уверенности в правильности того или другого суждения, часто употребляют слово «вероятно». Прибавкой к нему слов «очень» или «мало» выражают степень уверенности. Сторонник применения лекарства скажет: «Очень вероятно, что именно благодаря применению лекарства отход в опытной группе был меньше». Но человек, настроенный скептически, будет утверждать обратное: «Если бы и не давали лекарства, все равно в одной группе лисиц отход был бы больше, а в другой меньше в силу других причин, создавших разницу между группами».

Этот пример показывает, что биологу в его практике приходится очень часто встречаться с вероятностями, большими или малыми, и что ответить на многие вопросы можно, только зная некоторые, хотя бы самые элементарные положения теории статистики и лежащей в ее основе теории вероятности.

Что же такое вероятность? **Вероятность** – это возможность осуществления определенного события в некотором количестве случаев из общего числа возможных. Говоря иначе, вероятность – это степень уверенности в том, что событие произойдет.

Исходным в понятии вероятности является понятие равной возможности, на основе которого можно отделить необходимые явления от случайных. Так, если при осуществлении события ***A*** возможно только событие ***B***, то налицо необходимая связь явлений. Если же при ***A*** равновозможны и ***B*** и ***C***, мы имеем дело с проявлением возможности в виде случайного. *Случайное* – это такое же объективное явление, как и необходимое, и оно так же обусловлено различными причинами, как и необходимое, только характер причинности здесь иной, а именно: возможен не один, а два результата или более. Эти возможности и являются вероятностями. Процесс осуществления явления на основе известной его возможности, или вероятности, называется *вероятностным* или *стохастическим*. Теория вероятности изучает математические законы таких процессов.

Вероятность можно выразить математически по следующей формуле:

$$p = \frac{m}{n}$$

где m – число благоприятных случаев, а n – число всех возможных, или, правильнее, равновозможных, случаев.

Особи мужского и женского пола у очень многих видов животных рождаются примерно в равном количестве. Это значит, что на каждые 100 потомков в среднем должно родиться 50 самок и 50 самцов, отсюда вероятность рождения коровой телочки или бычка равна

p= 50 :100 = 0,5 или 50%.

Другой пример. Чтобы оценить вероятность рождения комолого теленка, надо знать количество рождавшихся ранее в данном стаде или породе комолых и рогатых животных. Так, если в данной породе за несколько последних лет обнаружено 110 комолых телят из общего количества 55000 родившихся, то вероятность рождения от коровы данной породы комолого теленка равна p = 110:55000 = 0,002. Это значит, что в среднем на каждые 1000 случаев приходится только 2 случая рождения комолых телят. На этом же примере легко понять и другую вероятность, как бы обратную величине p, что родится не комолый, а рогатый теленок. Последняя вероятность обозначается буквой q. Она выражается в данном случае величиной, равной 0,998. Алгебраическая сумма величин p и q равна 1, т.е. сумма вероятностей противоположных событий равна единице.

Из этих примеров вытекает *важный теоретический вывод*: **количественной характеристикой вероятности** того или иного явления, может быть относительная частота явления, установленная эмпирически на достаточно значительном фактическом материале. По этой эмпирически установленной частоте можно судить о возможности возникновения явления в будущем.

Если некоторое явление имеет вероятность p, то относительная частота его, обнаруживаемая в опыте или при наблюдении, будет близка к p. При этом она будет тем ближе к p, чем больше будет проведено опытов или наблюдений. Если заранее дается определенная вероятность, то по ней можно найти ожидаемую частоту явления, которая будет получена при проведении опытов. И наоборот, зная относительную частоту, можно найти приближенное значение вероятности, которое может служить для характеристики частоты данного явления в последующих опытах или наблюдениях. Но вероятность проявляется только при условии *большого числа наблюдений* или опытов.

Приведенные примеры показывают, что вероятности p и соответствующие им q могут иметь самые разные значения – от величин, близких к нулю или равных ему, до величин, близких к единице или равных ей. В нашем примере вероятность рождения комолого теленка очень мала. Однако существует немало событий, обладающих еще меньшей вероятностью. Наконец, если p=0, то на свершение данного события вообще нельзя рассчитывать.

Но могут быть события, вероятность которых, хотя и очень близка к нулю, но все же нулю не равняется. Практически можно утверждать, что вероятность обнаружения в стаде холмогорской породы коровы с жирностью молока 6,5% равна нулю, но возможен все же какой-то исключительный случай, когда в силу своеобразных физиологических условий холмогорская корова может дать молоко исключительно высокой жирности. Очевидно, вероятность подобного события будет выражена очень малой дробью.

Однако и события, обладающие очень малой вероятностью, осуществляются вполне закономерно, хотя они могут казаться невозможными. Маловероятные события при многократном повторении явления приобретают вполне устойчивую и определенную вероятность их осуществления, хотя бы такое событие происходило в одном случае из многих миллионов. Так, с точки зрения вероятности возникновение жизни на Земле представляется необычайно редким событием. Но каким бы невероятным ни казалось возникновение жизни или тех этапов, из которых складывалось возникновение жизни на Земле, времени для этого было достаточно, поэтому оно наверняка могло произойти хотя бы один раз, чего было уже достаточно для дальнейшего развития жизни.

По мере приближения величины p к единице событие становится все более достоверным. Если p=1, то событие бесспорно наступит. Оно вполне достоверно.

Оценка того, насколько мала должна быть вероятность, чтобы с ней можно было не считаться, в значительной мере зависит от степени важности события, о котором идет речь. Так, если вероятность воздействия нового удобрения на понижение урожая равна 0,05, это не должно помешать его применению, так как в 0,95 случая оно окажется полезным. Совсем другое дело, если оказывается, что новое лекарственное средство может с вероятностью, равной 0,05, принести не пользу, а вред организму больного. В этом случае его применение не может быть допущено.

Эти примеры иллюстрируют широко применяемый в опытах и наблюдениях методический приём, когда заранее намечают приемлемую величину (или уровень) вероятности и считают ее достаточной для доказательства получения того или иного эффекта.

9.3.1.1. Теоремы сложения и умножения вероятностей

Для понимания закономерностей случайной вариации важны две теоремы. Первая теорема – **теорема сложения вероятностей**: вероятность появления нескольких независимых и *несовместимых* событий равна сумме вероятностей этих событий. Эту теорему можно проиллюстрировать следующим элементарным примером. На клумбе растут 20 красных, 30 синих и 40 белых астр. Какова вероятность сорвать в темноте окрашенную астру? Она равна *сумме* вероятностей сорвать красную или синюю астру, т.е.

$$p = \frac{20}{90} + \frac{30}{90} = \frac{50}{90}$$

Вторая теорема – **теорема умножения вероятностей**: вероятность появления нескольких независимых, *но совместимых* друг с другом (или следующих друг за другом) событий равна произведению вероятностей этих событий.

Пример. Какова вероятность прохождения по лабиринту с шестью развилками и шестью тупиками? Очевидно, что на каждой развилке вероятности попасть или в тупик, или к следующей развилке одинаковы – по 1/2. Тогда при наличии шести развилок общая вероятность будет равна

$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2^6}=\frac{1}{64}.$$

9.3.1.2. Эмпирические и теоретические вероятности

В приведенных выше примерах мы рассчитывали так называемые *эмпирические вероятности*. Они применимы только к тем конкретным совокупностям, для которых они вычислены. Вероятность появления комолых телят относится к определенной изученной группе скота. Для породы, в которой очень много комолых животных, вероятность появления комолых телят окажется во много раз выше вычисленной. Для практики же очень важно судить не только об отдельных конкретных случаях, но и о всех возможных случаях этого рода.

Математическая теория, имея дело с отдельными, частными наблюдениями, выработала методы, позволяющие по результатам наблюдений судить о тех результатах, которые имели бы место, если бы изучалась не только данная совокупность осуществившихся случаев, но и теоретически мыслимая совокупность всех возможных случаев этого рода. Иначе говоря, по эмпирическим, опытным вероятностям, основанным на учете конкретных относительных частот ранее наблюдаемых явлений, можно судить о *теоретических вероятностях*.

Теоретическую совокупность, т.е. совокупность всех возможных наблюдений, в том числе и таких, которые практически не были осуществлены, называют **стохастической совокупностью**. Теория вероятностей дает возможность построить абстрактные совокупности, представляющие собой отображение реальных совокупностей. В таких абстрактных стохастических совокупностях, доступных точному математическому анализу, вероятности становятся теоретическими. Очевидно, в жизни мы встречаемся, как правило, с выборочными совокупностями, но по ним мы стремимся судить о генеральной или стохастической совокупности. Так, для изучения окуня данного озера нет надобности изучать всю его популяцию, т.е. генеральную совокупность, а достаточно взять выборочную совокупность в количестве 100, 200 или 1000 особей. По капле крови больного можно делать выводы о состоянии всей крови, данные об изменчивости нескольких десятков коров позволяют судить о всём стаде и т. д.

Если бы все особи популяции были сходны, то уже по одной особи можно было бы получить полную информацию о всей генеральной совокупности, всей популяции. Но в действительности существует очень большое разнообразие, как среди самих особей популяции, так и в отношении условий внешней среды, в которой они живут и развиваются. Поэтому и проводимые многократно выборки из генеральной совокупности никогда не будут одинаковыми.

Каковы закономерности вариации внутри каждой совокупности и каково взаимоотношение между разными типами совокупностей? Можно ли по статистическим показателям, полученным на основании изучения выборочной совокупности судить о статистических показателях генеральной совокупности? Иначе говоря, это вопрос о том, насколько достоверны статистические показатели, полученные по выборочной совокупности, чтобы можно было судить по ним о генеральной совокупности. Ниже мы будем пытаться ответить на все эти вопросы.

9.3.2. Нормальное распределение

Нормальное распределение – наиболее часто встречающийся тип распределения особей совокупности по классам вариационного ряда. Нормальное распределение играет важную роль при статистической обработке результатов наблюдений и поэтому часто используется для приближенного описания ошибок измерения.

Нормальное распределение можно изобразить в виде вариационной кривой. Ниже приведены вариационный ряд и кривая распределения коров по содержанию жира в молоке за лактацию.

Классы варьирующего признака	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0	4,1
Частоты (p)	3	5	10	12	20	20	20	12	10	5	3

Вариационную кривую получают следующим образом. По оси абсцисс (x) системы координат откладывают величины изучаемого признака от наименьшего значения к наибольшему. По вертикальной оси от каждой из этих точек проводят перпендикуляр, длина которого отражает частоту наблюдаемых вариант. Соединяют верхние точки построенных перпендикуляров. Получится вариационная кривая. Она покажет, что наибольшее число случаев относится к точке, близкой к средней величине варьирующего признака. В обе стороны от неё в направлении к наименьшей и наибольшей величине признака число случаев постепенно уменьшается. Концы ветвей не сливаются с осью абсцисс, а приближаются к ней в бесконечности (асимптотически). Теоретическая кривая нормального распределения при $n \to \infty$ имеет колоколообразный симметричный вид (рис. 6). Вершина нормальной кривой *(y_0)* определяется перпендикуляром, восстановленным из точки *x*, соответствующей величине средней арифметической данного признака. Величина y_0 соответствует наибольшей частоте *(p)* встречаемости тех особей популяции, у которых величина признака равна средней арифметической. В нормальном распределении точка $\bar{x}$ совпадает с величиной моды (*Mo*) и медианы (*Me*). Чем больше изменчивость признака, тем больше будет основание кривой.

Площадь, ограниченную осью абсцисс (*x*) и нормальной кривой принимают за 1 или за 100%. Она соответствует *общему числу* наблюдений выборки, которое также принимается за 100%.

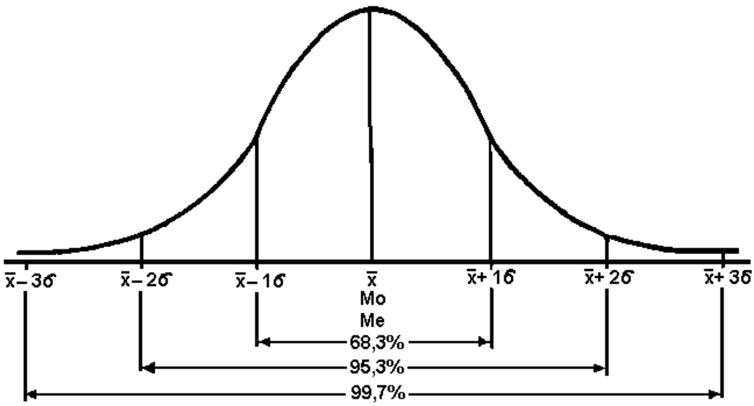


Рис.6. Кривая нормального распределения.

В нашем примере жирность молока (*x*) отмечена по горизонтальной оси (абсцисса), а частоты коров с соответствующей жирностью молока (*p*) – по вертикальной оси (ордината). В выборку включены 120 коров. Содержание жира в их молоке варьирует в пределах от x_{min} = 3,1 до x_{max} = 4,1%. Коровы распределены по 11 классам ряда соответственно их жирномолочности. Сумма всех частот (Σp) равна общему числу особей в выборке, т.е. объёму выборки *($\Sigma p = n$)*.

Нормальное распределение вариант в вариационном ряде характеризуют следующие закономерности:

1. Основное варьирование признака ограничивается лимитом, составляющим ±3σ от среднего значения признака ($\bar{x}$). В эти границы входит 99,7% всех членов совокупности. Эта закономерность была названа **правилом ±3σ**. За пределами ±3σ встречается только 0,3% особей с величиной признака больше +3σ или меньше –3σ.

В границах ±2σ находятся 95,5% значений вариант, а в пределы ±1σ попадают только 68,3% вариант.

2. Варьирование величины признака в границах ±3σ имеет особенность: можно установить теоретическую частоту (y) встречаемости особей, имеющих ту или иную величину признака x_i. Для этих целей используют уравнение нормальной кривой:

$$y_i = \frac{n}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \qquad (15)$$

где y_i – теоретическое число наблюдений для данной величины x;

σ – среднее квадратическое отклонение; π–постоянное число, 3,1416;

e– основание натуральных логарифмов, равное 2,71828;

(x– $\bar{x}$) –отклонение данного x от средней арифметической,

В приведённой формуле основание натуральных логарифмов возводится в степень $-\frac{(x-\bar{x})^2}{2\sigma^2}$. Поскольку выражение $\frac{x-\bar{x}}{\sigma}$ выражает нормированное отклонение t, то выражение $\frac{(x-\bar{x})^2}{2\sigma^2}$ может быть преобразовано следующим образом $\frac{(x-\bar{x})^2}{2\sigma^2} = \frac{t^2}{2}$. Тогда формула (15) может быть представлена как $y_i = \frac{n}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$.

Существуют специальные таблицы, в которых приведены результаты вычисления значений $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$, выражающую *первую функцию* нормального распределения. Её обозначают $f(t)$= 0,39894 ·2,71828[2].

Следовательно:

$$y_i = \frac{n}{\sigma} \cdot f(t) \text{ или } y_i = \frac{n \cdot k}{\sigma} \cdot f(t) \qquad (16)$$

где k – величина класса вариационного ряда.

Зная n и σ для какой-либо совокупности, можно определить величины теоретических частот и построить вариационный ряд для всех величин x, выражая отклонение x от $\bar{x}$ в нормированном отклонении $t=(x_i - \bar{x}):\sigma$, т.е. в долях σ (от 0 до 4). По специальной таблице узнают теоретическую частоту для любого значения x (табл. 13, с. 56).

Вторую функцию нормального распределения обозначают φ(t). Она показывает, какова площадь фигуры, образованной осью абсцисс, ординатами y_0 и y_1 и участком вариационной кривой между этими ординатами. Для практических расчётов важно то, что вторая функция нормального распределения соответствует числу (в %) особей совокупности, включённых в площадь фигуры, образованной отсеченной частью кривой. Существуют специальные статистические таблицы, в которых приведены величины φ(t) для любых значений $t_i=(x_i - \bar{x}):\sigma$, от 0 до 4σ (табл. 13, с. 56).

Третья функция нормального распределения $F(t)$ указывает, какова будет средняя величина признака в отсеченной части нормальной кривой. Вычисляют ее по формуле

$$F(t) = \frac{f(t)}{0{,}5 \pm \varphi(t)} \qquad (17)$$

Закономерности и три функции нормального распределения варьирующего признака используют при планировании селекционного процесса. Они позволяют определить:

1. теоретическую величину σ;
2. теоретические частоты, т.е. распределение животных в совокупности (стаде, породе) по данному признаку, если фактическое распределение их неизвестно, но исходные величины x и n даны;
3. количество животных (в процентах), которое может быть оставлено для племенного использования, если заданы те или иные уровни отбора (селекционного дифференциала);
4. средние величины признака животных, вошедших в группу при

заданном уровне отбора (при данном селекционном дифференциале);

5. границу отбора (селекционный дифференциал).

Таблица 13

Ординаты и площади нормальной кривой распределения, определяемые по первой и второй функциям (в сокращённом виде)

Нормированное отклонение в долях сигмы $t=\dfrac{x-\bar{x}}{\sigma}$	Первая функция $f(t)$, ордината y_i при $t=\dfrac{x-\bar{x}}{\sigma}$	Вторая функция φ(t), площадь между ординатами y_0 и y_i	Нормированное отклонение в долях сигмы $t=\dfrac{x-\bar{x}}{\sigma}$	Первая функция $f(t)$, ордината y_i при $t=\dfrac{x-\bar{x}}{\sigma}$	Вторая функция φ(t), площадь между ординатами y_0 и y_i
0,0	0,39894	0,00000	1,6	0,11092	0,44520
0,1	0,39695	0,03983	1,7	0,09405	0,45543
0,2	0,39104	0,07926	1,8	0,07895	0,46407
0,3	0,38139	0,11791	1,9	0,06562	0,47128
0,4	0,36827	0,15542	2,0	0,05399	0,47725
0,5	0,35207	0,19146	2,1	0,04398	0,48214
0,6	0,33322	0,22575	2,2	0,03547	0,48610
0,7	0,31225	0,25804	2,3	0,02833	0,48928
0,8	0,28969	0,28814	2,4	0,02239	0,49180
0,9	0,26609	0,31594	2,5	0,01753	0,43979
1,0	0,24197	0,34134	2,6	0,01358	0,49534
1,1	0,21785	0,36433	2,7	0,01042	0,49653
1,2	0,19419	0,38493	2,8	0,00792	0,49744
1,3	0,17137	0,40320	2,9	0,00595	0,49813
1,4	0,14973	0,41924	3,0	0,00443	0,49865
1,5	0,12952	0,43319	3,99	0,00014	0,49997

В табл. 13 приведены значения первой и второй функций нормального распределения, показывающие ординаты и площади кривой при разных значениях отклонения x от средней арифметической $\bar{x}$, выраженные в нормированном отклонении t.

Пример. В популяции коров айрширской породы численностью 5000 голов известен лимит жирномолочности (x_{min}=3,7 и x_{max}=4,5% при средней жирномолочности $\bar{x}$=4,05%). Требуется определить: 1) теоретическую величину σ; 2) распределение животных по этому признаку в популяции и 3) установить, какое количество коров можно выделить в племенную группу, если границей отбора считать величину, равную

+0,2σ [2, *с. 202*].

В данном случае лимит жирномолочности составит: x_{max}–x_{min}=4,5–3,7= =0,8%, Эти 0,8% приходятся на интервал от –3σ до +3σ (т.е. на интервал длиной 6σ). Тогда на 1σ приходится величина, равная *lim* : 6 = 0,8:6 = 0,13%. Запомните, что σ = 0,13% жирности молока – нам эту величину нужно будет использовать в дальнейших расчётах.

Чтобы решить поставленные задачи определяют размер класса (*k*=0,1%) и составляют вариационный ряд по жирномолочности. Вычисляют $\bar{x}$, затем вычисляют отклонения каждого класса от $\bar{x}$ (графа 2, табл. 14) и нормированное отклонение (графа 3, табл. 14)

Таблица 14

Вычисление теоретического вариационного ряда по жирномолочности коров

Класс по жирно-мо-лочности (x)	Откло-нение $x-\bar{x}$ = $=x-4{,}05$	Нормированное отклонение $t=\dfrac{x-\bar{x}}{\sigma}=\dfrac{x-4{,}05}{0{,}13}$	Значение $f(t)$ (по таблице)	Теоретические частоты $p_i=y_i$ $y_i=\dfrac{n\cdot k}{\sigma}\cdot f(t)=\dfrac{5000\cdot 0{,}1}{0{,}13}\cdot f(t)$	Округ-лённые часто-ты, p_m
3,7	–0,35	–0,35 : 0,13=–2,69	0,01071	41,2	41
3,8	–0,25	–0,25 : 0,13=–1,92	0,06316	242,6	243
3,9	–0,15	–0,15 : 0,13=–1,15	0,20594	792,0	792
4,0	–0,05	–0,05 : 0,13=–0,38	0,37115	1427,4	1427
4,1	+0,05	+0,05 : 0,13=+0,38	0,37115	1427,4	1427
4,2	+0,15	+0,15 : 0,13=+1,15	0,20594	792,0	792
4,3	+0,25	+0,25 : 0,13=+1,92	0,06316	242,2	243
4,4	+0,35	+0,35 : 0,13=+2,69	0,01071	41,2	41
4,5	+0,45	+0,45 : 0,13=+3,46	0,01000	3,85	4
					Σp_m=5010

Примечание. Величина *k* – классовый промежуток, равный 0,1%,

В табл. 14 представлен теоретический ряд распределения коров айрширской породы по содержанию жира в молоке для данного примера. По второй функции φ(t), используя данные справочной статистической таблицы (табл. 13) определяют, сколько коров можно отобрать из данной популяции в племенную группу, если взять границу отбора по жирномолочности +0,2σ и выше. По нашим данным, граница отбора будет следующей:

$$\bar{x}+0{,}2\sigma = 4{,}05+0{,}2\cdot 0{,}13 = 4{,}05+0{,}026 = 4{,}076\% \approx 4{,}08\%.$$

Эта граница отбора коров по жирномолочности означает отклонение от средней на t=0,2σ. В таблице 13 данной величине соответствует φ(t)=0,07926. Графически это выглядит так (рис. 7): площадь фигуры, образованной осью абсцисс (x) и вариационной кривой соответствет практически 100% животных или единице. Перпендикуляр, проведённый из точки $\bar{x}$, делит эту площадь на 2 части по 0,5 каждая. Из правой половины следует вычесть небольшую площадь (она заштрихована), соответствующую величине φ(t)=0,07926 (мы её нашли по таблице 13 при t=0,2σ); тогда в правой части графика не заштрихованная площадь будет равна разности 0,5–0,07926=0,42074, или ≈42,1% – это и есть то количество коров, которые должны быть включены в отбираемую группу за пределами уровня $\bar{x}$+0,2σ = 4,08% жира и более. Так как во всей популяции 5000 голов, то следует отобрать 42,1% коров (2105 голов) с жирномолочностью от 4,08% и более.

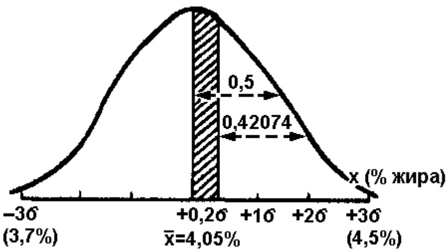


Рис. 7. Кривая нормального распределения при использовании второй функции φ(t) [2, с. 204].

Для определения средней жирномолочности 2105 коров *отобранной* группы используют формулу вычисления третьей функции нормального распределения (см. формулу 17). Используемые в этой формуле величины $f(t)$ и φ(t) находят по табл. 13: при t=0,2σ $f(t)$ =0,39104, а φ(t) =0,07926. Поскольку нас интересует площадь фигуры, *остающейся после отсечения* заштрихованной площади, то в знаменателе общей формулы сохраняем только знак «–».

$$F(t) = \frac{f(t)}{0,5 \pm \varphi(t)} = \frac{0,39104}{(0,5 - 07926)} = 0,93$$

Среднюю величину признака $\bar{x}_q$ *отсеченной* части вариационной кривой определяют следующим образом: $\bar{x}_q = \bar{x} + F_q(t) \cdot \sigma$ =4,05+0,93·0,13=4,20%.

Почему количественные признаки животных чаще всего имеют нормальное распределение? Это связано с тем, что в процессе своего развития они подвергаются воздействию большого количества внутренних и внешних факторов. Например, степень развития яйценоскости, живой массы кур или массы яиц определяется, с одной стороны, большим числом генов, присущих курам данной популяции, а с другой

– условиями внешней среды (кормлением, содержанием, микроклиматом и др.). Формирующийся признак в организме чаще всего встречается с *взаимодействием* наследственных и средовых факторов. В результате такого взаимодействия крайние отклонения, возникающие в результате действия только наследственных или только внешних условий, бывают редко. Средние или близкие к средним по своим показателям отклонения, получающиеся в результате взаимодействия *всех* факторов, встречаются значительно чаще. Поэтому изучаемые группы животных обычно имеют больше всего особей со средними или близкими к ним показателями. Чем дальше находится значение признака от средних показателей, тем таких животных меньше.

9.3.2.1. Нормированное отклонение

Нормальное распределение занимает важнейшее место в статистике вообще и в биометрии в частности, так как очень многие эмпирические распределения биологических признаков, характеризующихся непрерывной вариацией, приближаются к нормальному распределению.

Для изучения закономерностей вариации при нормальном распределении пользуются так называемым нормированным отклонением, которое обозначим буквой t. **Нормированное отклонение** – это отклонение той или другой варианты (или группы вариант) от средней арифметической, выраженное в сигмах

$$t = \frac{x_i - \bar{x}}{\sigma} \qquad (18)$$

Отсюда $x_i - \bar{x} = t \cdot \sigma$.

Каждая варианта характеризуется определенным значением *t*, указывающим ее положение в вариационном ряду или на кривой распределения. Так, если варианта № 26 имеет значение *t*= +1,5, это значит, что она располагается в правой части кривой на расстоянии в 1,5σ от $\bar{x}$. Если варианта № 38 имеет значение *t*= – 2,6, то она расположена в левой части кривой на расстоянии в 2,6·σ от $\bar{x}$ и т. д.

Вероятность любого отклонения от средней есть функция нормированного отклонения. Эта функция выражается сложной формулой, которая здесь не приводится. Используя её, статистики составили таблицу так называемого нормального интеграла вероятностей (табл. 15).

Таблица 15

Таблица вероятностей при нормальном распределении. Доли площади под нормальной кривой в пределах от $-t$ до $+t$ [6, с. 294].

t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1113	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	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6680	6729	6778
1,0	6827	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	8029
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	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1.8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9960	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9981	9986	9990	9993	9995	9997	9998	9999	9999

В табл. 15 первая колонка слева дает значения *t* с одним десятичным знаком, второй десятичный знак *t* представлен 10 столбцами, над которыми вверху указаны сотые доли *t* (цифры от 0 до 9). Тогда величине *t*=0,11 соответствует значение вероятности 0876 (в таблице

2-я строка, 2-я цифра), величине t=1,00 соответствует значение 6827 (11-я строка, 1-я цифра) и т. д. В целях упрощения таблицы в ней для вероятностей даны лишь десятичные знаки. Поэтому слева надо к ним присоединять ноль и запятую, т.е. число 0876 надо записать как 0,0876, число 6827– как 0,6827 и т.д.

Геометрически величины, находящиеся в табл. 15, являются долями площади нормальной кривой в границах от $-t$ до $+t$. *Эти доли в то же время выражают и вероятность*.

С помощью табл. 15 можно определить *вероятность нахождения вариант в заданных границах величины* $\pm t$, т.е. отклонения от средней арифметической, выраженного в сигмах. Например, вероятность того, что взятая наугад особь окажется из той части вариационного ряда, которая ограниченной справа и слева от $\bar{x}$ *одной сигмой*, т.е. $\pm 1t$ ($t = \pm 1\sigma$), равна 0,6827 (≈68%). Вероятность того, что случайно выбранная особь окажется из части вариационного ряда, ограниченной справа и слева от $\bar{x}$ *двумя сигмами*, т.е. $\pm 2t$, равна 0,9545 (≈95%). Наконец, вероятность того, что значение изучаемого признака любой случайным образом взятой особи будет находится в интервале, ограниченном тремя сигмами, т.е. в пределах $\pm 3t$, равна 0,9973 (≈99,7%).

9.3.2.2. Доверительная вероятность

В биометрии приняты три вероятности, которые используются во всех биологических, в том числе – зоотехнических и ветеринарных статистических анализах. Их выражают величинами 0,95, 0,99 и 0,999. Из табл. 15 можно установить, что с вероятностью 0,95 любая случайно взятая особь будет отклоняться от $\bar{x}$ не более чем на 1,96σ, или, иначе, с вероятностью 0,05 она будет за пределами ±1,96σ. С вероятностью же, равной 0,99, она будет отклоняться от $\bar{x}$ не более чем на ±2,58σ. Вероятность выхода этой варианты за пределы ±2,58σ равна 0,01. Наконец, с вероятностью, равной 0,999 любая, случайно выбранная особь будет отклоняться от $\bar{x}$ не более чем на 3,29σ, а вероятность попадания этой особи за пределы ±3,29σ будет равна всего 0,001.

Вероятность, выраженная величиной 0,99, достаточно велика. Поэтому в тех случаях, когда достигнута такая вероятность, можно с очень большой степенью уверенности делать вывод по поводу отнесения особи к той или иной группе, относительно результатов опыта и т.д. Но нередко можно пользоваться и более низким уровнем вероятности – 0,95. В этом случае отклонения от ожидаемого будут уже в 5% случаев (вероятность 0,05). Вероятности 0,95, 0,99, и 0,999 или 95%, 99% и 99,9% получили название доверительных вероятностей. **Доверительная вероятность** – это уровень вероятности, который считается достаточным для суждения о достоверности статистических показателей, получаемых на основании выборочных данных [3, с.288]

Вероятности, принятые как доверительные, в свою очередь определяют *доверительные границы* и *доверительный интервал* между ними. На них можно основывать оценку той или иной величины и те границы, в которых она может находиться при разных вероятностях.

Для различных вероятностей доверительные интервалы будут следующими:

Вероятности	Доверительные интервалы
0,95	–1,96σ ... +1,96σ
0,99	–2,58σ ... +2,58σ
0,999	–3,29σ ... +3,29σ

Вероятности можно обозначать как в долях единицы, так и в процентах, поэтому в последующем мы будем употреблять параллельно оба обозначения. Доверительную вероятность обозначают строчной буквой *p* [6, *с. 70* и *280*], однако могут встречаться обозначения и другими символами.

9.3.2.3. Уровень значимости

Определенным значениям вероятностей соответствуют так называемые *уровни значимости*. Вероятности p=0,95 (95%) соответствует уровень значимости 0,05 (5%). По отношению к закономерностям нормального распределения это означает, что выход за пределы принятых границ возможен в порядке случайности с вероятностью 0,05, т.е. в 5% случаев рискуют ошибиться в своих выводах.

При вероятности p=0,99 уровень значимости равен 0,01 (1%). Случайное отклонение возможно лишь с вероятностью 0,01, т.е. риск ошибиться в оценках составляет только 1% (1 случай на 100).

При вероятности p=0,999 уровень значимости равен 0,001 (0,1%). Случайное отклонение возможно лишь с вероятностью 0,001. Следовательно, риск ошибиться в оценках составляет только 0,1% (1 случай на 1000). Следовательно, *уровень значимости обозначает вероятность получения случайного отклонения от установленных с определенной вероятностью результатов*. С помощью уровня значимости можно установить, в каком проценте случаев (или с какой вероятностью) все же возможна ошибка в тех выводах, которые делаются на основе опыта, в оценке достоверности показателей или различий между какими-то величинами, полученными в опытах или при наблюдениях.

Обозначают уровень значимости прописной буквой Р с соответствующим подстрочным индексом (без указания нуля целых), например, P_{05}, P_{01}, P_{001} [6, *с. 71* и *280*]. Вместе с тем в литературе могут встречаться и другие обозначения уровня значимости.

При исследовании животных надо не только получить те или другие результаты, но и сделать выводы, поэтому очень важно, чтобы получаемые выводы имели достаточно высокую достоверность. Например, 5%-ный уровень значимости (0,05) указывает, что возможна в силу случайности ошибка в 5% случаев. В некоторых случаях можно удовлетвориться и таким результатом. Но если нужна большая доказательность результатов, то уровень значимости должен быть повышен до 1% (0,01). Чем цифра меньше, тем уровень значимости, а следовательно, и достоверность результатов выше. При уровне значимости 0,01 (1%) вывод не обоснован только в одном случае из 100. Такую значимость считают уже высокой и широко ею пользуются. Но бывают случаи, когда уровень значимости может быть еще выше – 0,001. Тогда вывод не обоснован только в одном случае из 1000. Уровень значимости – это та вероятность, которой решено пренебрегать в данном исследовании или явлении.

Таким образом, **уровни значимости** – это показатели, используемые для проверки статистических гипотез, связанные с такими значениями вероятности, при которых появление ожидаемых событий в данных условиях считается практически невозможным. Чем выше уровень значимости, тем меньше вероятность отвергнуть нулевую гипотезу [3, *с. 290*].

9.3.2.4. Исключение из совокупности резко отклоняющихся вариант

При статистической обработке материала часто возникает вопрос о том, принадлежат ли некоторые очень резко отклоняющиеся варианты (x_i) *к изучаемой* генеральной совокупности или они достоверно отклоняются от вариационного ряда выборочной совокупности и должны быть устранены, как представители *другой* генеральной совокупности. Иллюстрирует такую ситуацию следующий случай из практики. У 10 здоровых коров исследовано содержание лейкоцитов в крови. Минимальное (x_{min}) число лейкоцитов в крови составило 5900, а максимальное (x_{max}) – 10000. Средняя арифметическая ($\bar{x}$) по 10 коровам – 6000 лейкоцитов; σ = 100 лейкоцитов. Требуется определить, принадлежат ли *крайние* варианты к изучаемой генеральной совокупности [2, с.199].

Простейший способ решить вопрос о правомочности исключения резко отклоняющихся вариант (x_{min} или x_{max}) – использовать правило

±3σ (см.: с. 27). Для этого устанавливают теоретические границы варьирования, т.е находят теоретически ожидаемые отклонения крайних вариант от средней арифметической

$$x_{min} = \bar{x} - 3\sigma = 6000 - 3 \cdot 100 = 5700 \text{ лейкоцитов;}$$

$$x_{max} = \bar{x} + 3\sigma = 6000 + 3 \cdot 100 = 6300 \text{ лейкоцитов.}$$

Результаты расчётов показывают, что *эмпирический минимум* больше, чем теоретический ($x_{эмп}$=5900 > $x_{теор}$=5700). Следовательно, минимальная варианта лежит внутри границ $\bar{x}$ ±3σ и принадлежит генеральной совокупности, из которой сделана выборка 10 коров. Эмпирическая максимальная варианта резко превосходит теоретический максимум ($x_{эмп}$=10000 > $x_{теор}$=6300). Поэтому животное с таким уровнем лейкоцитов в крови из выборки *исключается*, как принадлежащее *другой* генеральной совокупности, которая характеризует животных, возможно, больных лейкозом, т.к. в крови у них содержится б**о**льшее количество лейкоцитов, чем у здоровых.

В некоторых случаях (когда известно, что исследуемый признак характеризуется больш**и**м коэффициентом вариации или когда аналитик хочет избежать ошибочных исключений крайних вариант) в качестве теоретических границ вариационного ряда можно брать ±4σ.

9.3.3. Биномиальное распределение

Статистические утверждения, полученные в генетических исследованиях, справедливы только с определенной вероятностью. Сумма вероятностей наступления данного события всегда равна 1, а отношение *a/n* называется **относительной частотой** этого **события**.

Значения дискретной переменной характеризуются всегда тем, что определенный признак *x* у изучаемой особи наблюдается с вероятностью, равной *p*, или не наблюдается, вероятность чего равна *q*. Если исследуют несколько групп животных, например *n*, то возможны следующие случаи: изучаемый признак *x* не наблюдается ни у одной особи (0 раз), у одной (1 раз), у двух (2 раза) и т. д., наконец, у всех исследованных животных (*n* раз). Каждый случай характеризуется определенной вероятностью.

Например, исследуют потомство 100 кур, от каждой из которых получено по 8 цыплят. Предположим, что вероятность рождения самца такова же, как и самки, т.е. равна 1/2. При этом может вывестись 0, 1,2, 3, 4, 5, 6, 7, 8 петушков, хотя 0 и 8 менее вероятно, чем 3 и 4. Определим вероятность, что *p* цыплят будут петушками, тогда $q = n - p$ цыплят будут курочками. Запишем все возможные случаи вывода курочек и петушков.

8♂	7♂	6♂	5♂	4♂	3♂	2♂	1♂	0♂
0♀	1♀	2♀	3♀	4♀	5♀	6♀	7♀	8♀
1/256	8/256	28/256	56/256	70/256	56/256	28/256	8/256	1/256

Подобное распределение называется *биноминальным*. Вероятности биноминального распределения получают путем разложения бинома

$$\left(\frac{1}{2}+\frac{1}{2}\right)^n$$

или для общих случаев $(p+q)^n$, где *p*–вероятность одного, а *q* – вероятность другого изучаемого признака; *n* – число событий (в данном примере число цыплят, полученных от каждой несушки).

Итак, если вероятности появления отдельных вариант выражаются величинами, соответствующими *коэффициентам разложения бинома Ньютона*, то такое распределение называется **биномиальным**.

Коэффициент биноминального распределения в приведенном примере получают из так называемого *треугольника Паскаля*, в котором каждый коэффициент представляет собой сумму двух лежащих над ним коэффициентов. При получении от каждой несушки 8 цыплят треугольник имеет следующий вид:

Степень (n)	Возможное распределение цыплят по полу																
0									1								
1								1		1							
2							1		2		1						
3						1		3		3		1					
4					1		4		6		4		1				
5				1		5		10		10		5		1			
6			1		6		15		20		15		6		1		
7		1		7		21		35		35		21		7		1	
8	1		8		28		56		70		56		28		8		1

Другое определение биномиального распределения [7]**:** биномиальным распределением называется *распределение вероятностей* дискретной *случайной величины X=m* принимающей значения 0, 1, 2, 3…*n* и задаваемой функцией вероятностей

$$P_{m,n} = C_n^m \cdot p^m \cdot q^{n-m}$$

где, $P_{m,n}$ – вероятность появления события *A* *m* раз в *n* независимых испытаниях, в каждом из которых событие *A* появляется с одной и той же вероятностью p и не появляется с вероятностью q=1–p; $C_n^m = \frac{n!}{m! \cdot (n-m)!}$ – число сочетаний из n по m.

Биномиальное распределение характеризует распределение признаков, варьирующих дискретно, прерывисто (например, число детенышей в помете, число самок и самцов в помёте, число заболевших особей, животных с доминантным признаком и т. п.).

При биномиальном распределении частоты (вероятности) появления альтернативных вариантов данного признака учитываются на независимых друг от друга субвыборках. Характеристикой биномиального распределения служат средняя арифметическая варьирующего признака

$$\bar{x} = \frac{\sum x \cdot p}{n} \qquad (19)$$

и среднее квадратическое отклонение

$$\sigma = \sqrt{k \cdot p \cdot q}\,, \qquad (20)$$

где n – объём выборки; k – число субвыборок; p и q – частоты появления каждого альтернативного признака; x – показатель числа классов альтернативного признака.

Пример 1. Изучить распределение кроликов по доминантной окраске шерсти в 25 помётах. В каждом помете – четыре крольчонка. Возможные варианты числа доминантно окрашенных крольчат будут следующими: нет крольчат с доминантным признаком (0); один крольчонок; два; три; четыре крольчонка с доминантным признаком. Вариационный ряд фактического распределения пометов по числу крольчат с доминантной окраской приведен в табл. 16.

Таблица 16

Вариационный ряд распределения пометов по альтернативному признаку

Число в помете крольчат с доминантным признаком (x)	0	1	2	3	4	(k=4)
Число пометов (p)	2	4	5	10	4	Σp=n=25
x• p	0	4	10	30	16	Σx• p =60

Среднее число доминантных крольчат на помет составит: $x = \Sigma x{\cdot}p/n$ = 60/25=2,4 головы. По формуле $\bar{x}=k{\cdot}p$ находят долю крольчат с доминантным признаком p = 2,40:4=0,60, а затем долю рецессивных крольчат $q = 1–p$ = 1–0,60 = 0,40. Определяют изменчивость окраски крольчат в пометах, исходя из формулы

$$\sigma = \sqrt{k \cdot p \cdot q} = \sqrt{4{\cdot}0{,}60{\cdot}0{,}40} = \sqrt{0{,}96} = 0{,}98.$$

Зная фактические частоты p и q, можно установить по биному Ньютона *теоретические* частоты распределения пометов по окраске крольчат. В данном примере это выразится так:

$$(p+q)^k = (0{,}60+0{,}40)^4.$$

По треугольнику Паскаля для k=4 находят коэффициенты, которые входят в многочлен бинома: 1, 4, 6, 1. Для рассматриваемого бинома в нашем примере возможны следующие вероятности:

$$(0{,}60+0{,}40)^4 = 1{\cdot}0{,}6^4+4{\cdot}0{,}6^3{\cdot}0{,}4+6{\cdot}0{,}6^2{\cdot}0{,}4^2+4{\cdot}0{,}6{\cdot}0{,}4^3+1{\cdot}0{,}4^4 =$$
$$=0{,}1296+4{\cdot}0{,}216{\cdot}0{,}4+6{\cdot}0{,}36{\cdot}0{,}16+4{\cdot}0{,}6{\cdot}0{,}064+0{,}0256=$$
$$=0{,}1296+0{,}3456+0{,}3456+0{,}15364+0{,}0256.$$

Умножая величины вероятности на общее число обследованных пометов, получают:

$P_{теор}=P{\cdot}n_{гол}$	Эмпирическое число помётов (p)
0,1296 ·25=3,24	2
0,3456 ·25=8,64	4
0,3456 ·25=8,64	5
0,1536 ·25=3,84	10
0,0256 ·25=0,64	4

Как в фактическом, так и в теоретическим распределении наблюдается «скошенность» вариационных кривых, обусловленная тем, что частоты доминантных и рецессивных признаков не равны друг другу (p–0,6; q=0,4), а число субвыборок не очень велико (n=25).

Пример 2. В хозяйстве изучено распределение семейств КРС по количеству больных туберкулезом коров (табл. 17). Каждое семейство состояло из 5 особей.

Таблица 17

Распределение семейств по количеству больных туберкулезом коров

Число больных (x)	5	4	3	2	1	0	(k=5)
Число семейств (p)	1	1	3	12	5	3	$\Sigma p = n$=25
$x{\cdot}p$	5	4	9	24	5	0	$\Sigma x{\cdot}p$ =47

Среднее количество туберкулезных коров в семействе $\overline{x} = \Sigma x{\cdot}p/n$ = 47:25 = 1,88. По формуле $\overline{x} = k{\cdot}p$ доля больных коров в семействе равна $p = \overline{x} : k$ = 1,88:5 = 0,38. Доля здоровых коров составит$q = 1 - p$ = 1–0,38 = 0,62. Находим среднее квадратическое отклонение

$\sigma = \sqrt{k \cdot p \cdot q} = \sqrt{5 \cdot 0{,}38 \cdot 0{,}62} = \sqrt{1{,}18} =$ 1,08. Так как данный вариационный ряд является рядом разложения бинома Ньютона (0,38 + 0,62)5 при n = 25, то можно вычислить теоретическую частоту распределения семейств по количеству больных туберкулезом коров. По треугольнику Паскаля коэффициенты бинома для k = 5 будут следующими: 1, 5, 10, 5, 1. Отсюда получаем следующие показатели вероятности:

(0,38 + 0,62)5 =

$=1{\cdot}0{,}38^5+5{\cdot}0{,}384{\cdot}0{,}62+10{\cdot}0{,}38^3{\cdot}0{,}622+10{\cdot}0{,}38^2{\cdot}0{,}62^3+5{\cdot}0{,}38{\cdot}0{,}62^4+$
$+1{\cdot}0{,}62^5 =$ 0,0079 + 0,0645 + 0,2106 + 0,3441 + 0,2808 + 0,0916.

Чтобы получить теоретическое распределение семейств, нужно умножить полученные показатели вероятности на суммарное число исследованных семейств (n=25):

$P_{теор}=P{\cdot}n_{гол}$	Эмпирическое число помётов (p)
0,0079 ·25=0,2	1
0,0645 ·25=1,6	1
0,2106 ·25=5,3	3
0,3441 ·25=8,6	12
0,2808 ·25=7,0	5
0,0916 ·25=2,3	3

Степень соответствия фактического распределения распределению теоретически ожидаемому можно определить методом хи-квадрат.

9.3.4. Распределение Пуассона

Распределение Пуассона относится к редко происходящим событиям с *дискретной* или *прерывистой, изменчивостью*. Оно имеет самостоятельное значение, хотя его можно рассматривать как *предельный* случай биноминального распределения. Если при *биноминальном* распределении значения p и q могут иметь близкие величины, то при *пуассоновом* распределении p очень мало, т.е. события осуществляются очень редко, а q приближается к единице. В отличие от нормального и биномиального распределений, для характеристики которых нужно знать $\bar{x}$ и σ, для распределения Пуассона достаточно одного параметра, а именно $\bar{x}$, так как показатель дисперсии σ^2 здесь совпадает с величиной $\bar{x}$. Именно по этому равенству $\bar{x}$ и σ легче всего определить, что данное распределение является пуассоновым.

Примером распределения Пуассона могут быть такие события как рождение уродов, появление мутаций, рождение троен и четвёрен у человека или крупного рогатого скота. Такие события могут быть подсчитаны и выражены только целыми числами, но регистрируются крайне редко, поэтому частота их появления очень мала. При графическом изображении *распределение Пуассона имеет резко асимметричную форму*. Но с увеличением $\bar{x}$ симметрия возрастает [6, с.65].

Появление редких событий учитывают в ряде субвыборок, проведенных в единицу времени (в единице объема, на единице площади), в которых редкое событие может отсутствовать (0) или проявиться 1, 2, 3 раза и т.д. Величины 0, 1, 2, 3 и т.д. образуют классы вариационного ряда редких событий, а число появления в каждом классе редкого события выражает частоты (p) или при большом числе субвыборок – теоретические частоты, соответствующие уровню вероятности.

Использование распределения Пуассона необходимо в генетическом анализе популяций животных, например, в связи с появлением мутантных особей. Особенно важно определить возможную частоту появления в популяции (стаде) мутантов, обладающих патологическими и аномальными свойствами. Носителей летальных мутаций необходимо устранять из популяции. Пользуясь распределением Пуассона, можно прогнозировать частоту появления нежелательных мутантов в стаде или породе и планировать предупредительные мероприятия селекционного характера. Для таких прогнозов используют уравнение Пуассона, показывающее теоретическую частоту (или вероятность) появления редкого события:

$$P_{x_i} = \frac{\bar{x}^{x_i}}{x_i!} e^{-\bar{x}} = \frac{\bar{x}^{x_i}}{x_i! \cdot 2{,}7183^{\bar{x}}}, \qquad (21)$$

где x_i – число появлений редко встречающегося события (мутация) в независимых повторных испытаниях; e – основание натуральных логарифмов; $\bar{x}$ – среднее число появления редкого события; ! – факториал $(1 \cdot 2 \cdot 3 \cdot \ldots \cdot n)$.

В специальных статистических таблицах (см.: табл. 19, с. 71) приводятся готовые данные о величине вероятности появления редкого события, если известны $\bar{x}$ и x_i, поэтому расчеты можно не проводить, а пользоваться этими таблицами.

Примером распределения Пуассона может служить появление в пометах гетерозиготных норок мутантных по окраске шерсти щенков [2]. Обследовано 50 пометов (субвыборок) с равным числом щенков – по 5 голов. Получены следующие результаты:

– в нулевом классе – 11 пометов (нет щенков с рецессивной окраской),

– в первом классе – 10 помётов (1 щенок с рецессивной окраской),

– во втором классе – 25 пометов (2 щенка с рецессивной окраской)

– в третьем классе – 2 помёта (3 щенка с рецессивной окраской)

– в четвёртом классе – 1 помёт (4 щенка с рецессивной окраской)

– в пятом классе – 1 помёт (все 5 щенков с рецессивной окраской)

Распределение пометов по частоте рождения мутантных щенков будет иметь следующий вид (табл. 18). Возможное появление рецессивных щенков составит в среднем: $\bar{x}_{рецес}$ = $\Sigma x{\cdot}p/2p$= 75/50 = 1,5 головы на один помет.

Таблица 18

Распределение помётов по частоте рождения мутантных щенков

Класс с числом рецессивных щенков *(x)*	5	4	3	2	1	0	Всего
Число пометов (субвыборок) – частоты (p)	1	1	2	25	10	11	$\Sigma p = n = 50$
$x{\cdot}p$	5	4	6	50	10	0	$\Sigma x{\cdot}p = 75$

Таблица 19

Значения вероятности $P_{x_i} = \dfrac{\bar{x}^{x_i}}{x_i!} e^{-\bar{x}}$ (в таблице ноль целых и запятая опущены) [1, с.140; 2, с.206; 3, с.269].

x_i, число редких событий	$\bar{x}$, среднее число редких событий,									
	0,1	0,2	0,3	0,4	0,5	0,6	0, 7	0,8	0,9	1, 0
0	9048	8187	7408	6703	6065	5488	4966	4493	4066	3679
1	0905	1637	2222	2681	3033	3293	3476	3595	3659	3679
2	0045	0164	0333	0536	0758	0988	1217	1438	1647	1839
3	0002	0011	0033	0072	0126	0198	0284	0383	0494	0613
4	0000	0001	0003	0007	0016	0030	0050	0077	0111	0153
5				0001	0002	0004	0007	0012	0020	0031
6							0001	0002	0003	0005
7										0001

Таблица 19 (продолжение)

x_i, число редких событий	$\overline{x}$, среднее число редких событий,									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	3329	3012	2725	2466	2231	2019	1827	1653	1496	1353
1	3662	3614	3543	3452	3347	3230	3106	2975	2842	2707
2	2014	2169	2303	2417	2510	2584	2640	2678	2700	2707
3	0738	0867	0998	1128	1255	1378	1496	1607	1710	1804
4	0203	0260	0324	0395	0471	0551	0636	0723	0812	0902
5	0045	0063	0084	0111	0141	0176	0216	0260	0309	0361
6	0008	0013	0018	0026	0035	0047	0061	0078	0098	0120
7	0001	0002	0003	0005	0008	0011	0015	0020	0027	0034
8			0001	0001	0001	0002	0003	0005	0006	0009
9							0001	0001	0001	0002

Таблица 19 (продолжение)

x_i, число редких событий	$\overline{x}$, среднее число редких событий,									
	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3,0
0	1225	1108	1003	0903	0821	0743	0672	0608	0550	0498
1	2572	2438	2306	2177	2052	1931	1815	1703	1596	1494
2	2700	2681	2652	2613	2565	2510	2450	2384	2314	2240
3	1890	1964	2083	2090	2138	2176	2205	2225	2234	2240
4	0992	1087	1169	1254	1336	1414	1488	1557	1622	1680
5	0417	0476	0538	0602	0668	0735	0804	0872	0941	1008
6	0146	0175	0206	0241	0278	0319	0362	0407	0455	0504
7	0044	0055	0068	0083	0099	0118	0140	0163	0188	0216
8	0012	0015	0020	0025	0031	0039	0047	0057	0068	0081
9	0003	0004	0005	0007	0009	0011	0014	0018	0022	0027
10	0001	0001	0001	0002	0002	0003	0004	0005	0006	0008

Таблица 19 (окнчание)

x_i, число редких событий	$\overline{x}$, среднее число редких событий,									
	3,5	4, 0	4,5	5,0	6,0	7, 0	8,0	9,0	10, 0	11,0
0	0302	0183	0111	0067	0025	0009	0003	0001	0000	0000
1	1057	0733	0500	0337	0149	0064	0027	0011	0005	0002
2	1850	1465	1125	0842	0446	0223	0107	0050	0023	0010
3	2153	1954	1687	1404	0892	0521	0286	0150	0076	0037
4	1888	1954	1898	1755	1339	0912	0573	0337	0189	0102
5	1327	1563	1708	1755	1606	1277	0916	0607	0378	0224
6	0771	1042	1281	1462	1606	1490	1221	0911	0631	0411
7	0386	0595	0824	1044	1377	1490	1396	1171	0901	0646
8	0169	0298	0463	0653	1033	1304	1396	1318	1126	0888
9	0066	0132	0232	0363	0688	1014	1241	1318	1251	1085

10	0023	0053	0104	0181	0413	0710	0993	1186	1251	1294
11	0007	0019	0043	0082	0225	0452	0722	0970	1137	1194
12	0002	0006	0016	0034	0113	0264	0481	0728	0948	1094
13	0001	0002	0006	0013	0052	0142	0296	0504	0729	0926
14	0000	0001	0002	0005	0022	0071	0169	0324	0521	0728

Чем больше величина $\bar{x}$, тем больше распределение Пуассона приближается к нормальному. Например, вероятность, с которой могут быть получены норки рецессивных окрасок, можно определить, или используя уравнение Пуассона или по табл. 7.

Определяют *вероятность* по таблице значений P_x для классов с появлением пяти, четырех, трех, двух, одного рецессивного щенка и их отсутствие при среднем числе рецессивных щенков $\bar{x}$ =1,5 головы. Подставляют значения в формулу Px исходные данные, а теоретические частоты появления мутантов (p) берут из стандартной таблицы. Вероятность рождения пяти мутантов в пометах с пятью щенками будет следующей:

$$P_{x_5} = \frac{1{,}5^5}{5!\cdot 2{,}7183^{1{,}5}} = \frac{1{,}5^5}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 2{,}7183^{1{,}5}} = 0{,}0141$$

Эту величину можно без расчета по формуле определить с помощью табл. 19, из которой находят, что при $\bar{x}$ =1,50 и при x_i=5 вероятность рождения пяти мутантов равна P_{x5}=0,0141. Аналогично, пользуясь табл. 19, находят теоретическую вероятность рождения рецессивных щенков в количестве четырех, трех, двух, одного на помет или их отсутствия. Таким образом: P_{x4}=0,0471; P_{x3}=0,1265; P_{x2}=0,2510; P_{x1}=0,3347: для пометов, в которых мутанты не обнаружены, вероятность события составит P_{x0}=0,2231.

Чем ближе будут друг к другу вероятности появления альтернативных событий (p – появление рецессивных, q – появление доминантных щенков), тем больше вариационный ряд Пуассона будет приближаться от скошенного к нормальному. Если нужно выразить этот ряд не в показателях теоретической вероятности P, а в абсолютных числах теоретических частот p, то полученные значения вероятностей умножают на соответствующее число субвыборок (n). В нашем примере n=50 (табл. 20).

Таблица 20

Фактические частоты появления щенков с рецессивным признаком [2, с.207]

Классы (по числу рецессивных щенков)	$P_{xi}{\cdot}n{=}P_{теор}$, голов	Фактические частоты, число голов
x_5	0,0141·50=0,7	1
x_4	0,0471·50= 2,38	1
x_3	0,1255·50=6,28	2
x_2	0,2510·50=12,55	25
x_1	0,3347·60=16,74	10
x_0	0,2231·50=11,2	11

Другой пример [1, с.140]. Если в популяции вероятность появления наследственного уродства p = 0,002, то можно определить вероятность появления 3-х, 2-х, 1-го и отсутствие уродов среди, например, 200 телят. Среднее число появления уродов в изучаемой совокупности $\bar{x}$ = 200·0,002 = 0,4 головы. Вероятность рождения трех уродов составит:

$$P_{x3}=\frac{0,4^3}{3!\cdot 2,7183^{0,4}}=\frac{0,064}{1\cdot 2\cdot 3\cdot 2,7183^{0,4}}=\frac{0,064}{6\cdot 1,491}=0,0072$$

Когда известны $\bar{x}$ и x_i, то по таблице значений вероятностей находим, что при $\bar{x}$ = 0,4 и x_i = 3 вероятность появления трех уродов P_3 = 0,0072. Вероятность появления двух уродов – P_2= 0,0536, одного – P_1= 0,2681 и отсутствие уродов – P_0 = 0,6703.

Если фактические частоты достоверно отличаются от теоретических, то и распределение при этом достоверно отличается от распределения Пуассона, характеризующего редко происходящие события. Следовательно, такое фактическое распределение отражает влияние не случайных причин, а систематически действующих факторов. Например, в стаде выявлено рождение уродливых телят, причем частота их появления достоверно выше того значения, которое характерно для распределения Пуассона. Поэтому следует сделать вывод, что рождение уродливых животных не случайно, а обусловлено патологической наследственностью производителя или вызвано действием на организм неблагоприятных факторов среды. В обоих случаях необходимо не считать такое явление случайным, а установить причину появления уродств в стаде. Использование распределения Пуассона позволяет сделать правильный вывод о причине нарушений и провести мероприятия, направленные на их предотвращение.

Для определения достоверности в расхождении фактических и теоретических частот используют метод χ^2 («хи-квадрат»).

9.3.5. Асимметричное и эксцессивное распределения

Наряду с нормальным распределением возможны асимметричные и эксцессивные распределения. При **асимметричном распределении** (распределение Максвелла) на графике получается кривая, скошенная вправо (рис. 8, А) или влево (рис. 8, Б) за счёт того, что происходит накопление частот в левой или правой части кривой. В случае графика, скошенного вправо, распределение будет **положительным**, или правосторонним. График, скошенный влево, называется **отрицательным**, или левосторонним.

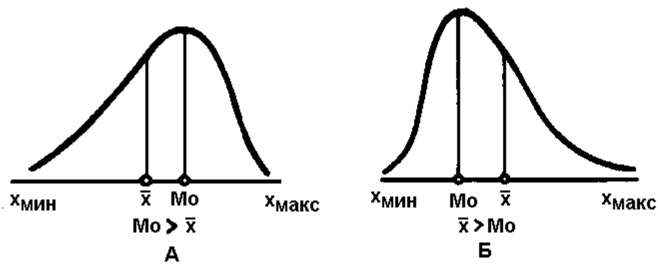


Рис. 8. Типы асимметричного распределения [2]:
А – отрицательная асимметрия; *Б* – положительная асимметрия.

Причины, вызывающие асимметричное распределение могут быть различными. Это может быть систематическое направленное действие каких либо внешних факторов, например, влияние каких-то экспериментально контролируемых факторов (уровень кормления, интенсивность отбора), изменяющих нормальное распределение и вызывающих асимметрию. Асимметрия может возникать как результат влияния случайных (неопределённых и неконтролируемых) факторов. Асимметрия может быть следствием неправильно сделанной выборки (не случайной, тенденциозной или же недостаточной по объёму). В этом случае нужно провести новый отбор особей из генеральной совокупности.

Эксцесс – это крайнее проявление чего-либо. В статистике **эксцессом** называют случаи чрезмерной крутовершинности или, наоборот, плосковершинности вариационных кривых, рассматриваемые как отклонения рядов распределения от нормальной кривой. **Эксцессивное** распределение характеризуется значительным накоплением частот в классах, близких по величине к среднему значению признака (положительный эксцесс). На графике это выражается крутовершинностью и уплощенностью ветвей кривой (рис. 9, А). При плосковершинности наблюдается снижение частот вариант в классах, расположенных в центральной части вариационного ряда (рис. 9, В) и увеличение частот вариант в классах, расположенных ближе к краям вариационного ряда.

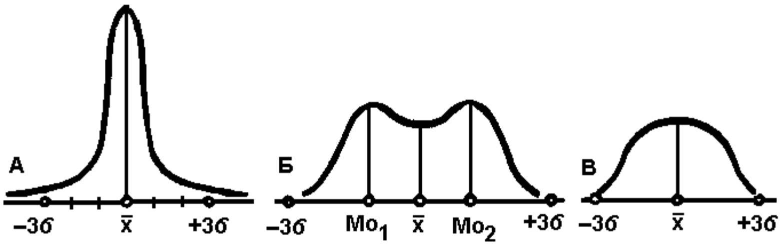


Рис. 9. Типы эксцессивного распределения [2]: А – положительный; Б – отрицательный: В – плосковершинный.

Двухвершинность указывает на то, что члены, входящие в состав выборки, неоднородны. Это отражает те или иные качественные сдвиги в состоянии варьирующего признака, вызванные влиянием на организм различных факторов. Так, появление двухвершинности в распределении овец по качеству шерсти может отражать наличие мутантных овец, у которых тонина шерсти отличается от показателя основного стада. Двухвершинность в распределении молочного скота по содержанию жира в молоке может явиться следствием того, что в стаде имеются помесные животные. Вместе с тем оба типа эксцесса могут возникнуть и в результате неправильно проведенной выборки, что недопустимо.

9.3.6. Трансгрессивное распределение

Трасгрессивными рядами называют такие вариационные ряды, которые отличаются друг от друга величиной средней арифметической и у которых крайние классы, лежащие около максимального класса первой кривой, служат крайними минимальными классами другой кривой, что создаёт в этих частях вариационных кривых их взаимное пересечение.

В соответствии с данным определением, при трансгрессивном распределении варьирующего признака несколько классов одного вариационного ряда в максимальных его величинах, являются в то же время классами минимальных значений другого вариационного ряда (рис. 10).

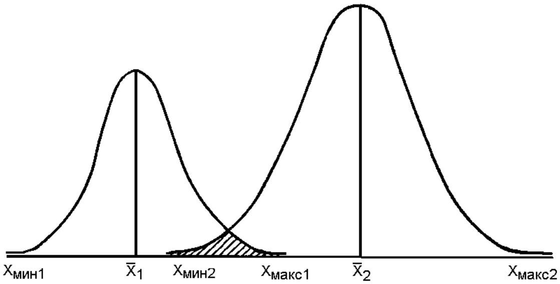


Рис. 10. Общий вид двух трансгрессивных вариационных рядов.

На графике вариационная кривая одного вариационного ряда частично накладывается на вариационную кривую другого ряда, образуя трансгрессирующую зону (заштрихована) с одинаковыми классами для части обеих кривых. Наличие трансгрессирующей зоны является *первым свойством* трансгрессии вариационных рядов. *Второе свойство* трансгрессивных рядов заключается в том, что их средние арифметические $\bar{x}_1$ и $\bar{x}_2$ статистически достоверно различаются между собой.

В зависимости от целей селекции может стать желательным уменьшение степени трансгрессии или ее увеличение, то есть сближение средних $\bar{x}_1$ с $\bar{x}_2$ и самих кривых. Для определения степени трансгрессии пользуются формулой коэффициента трансгрессии:

$$T = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2}, \tag{22}$$

где n_1 и n_2; – число членов каждой совокупности; p_1 и p_2 – доли частот, входящих в трансгрессию.

Величины долей, тех вариант (классов), которые входят в трансгрессию определяют с помощью *второй функции* нормальною распределения φ(t) (см. с. 55) по формулам: p_1=0,5±$\varphi(t_1)$ и p_2=0,5±$\varphi(t_2)$. В этих формулах t_1 и t_2 – нормированные отклонения варьирующего признака, ордината которого отсекает трансгрессирующую часть кривой,

Пример. Необходимо определить степень трансгрессии вариационного ряда помесных коров с вариационным рядом черно-пестрых коров по жирномолочности, если известны следующие показатели (см. табл. 21).

Таблица 21

Показатели жирномолочности чистопородных чёрно-пестрых и помесных коров.

Показатель	Черно-пестрые (1-я группа)	Помесные (2-я группа)
Средняя жирномолочность ($\bar{x}$)	3,0%	3,8%
Лимиты (*min* и *max*)	2,8-3,6%	3,3-4,2%
σ, %	0,1%	0,2%
Ошибки средних	0,05%	0,1%
n	500 голов	100 голов

На рис. 11 показаны вариационные кривые для 1-й и 2-й групп. Прежде необходимо доказать, что анализируемая трансгрессия соответствует второму свойству трансгрессирующих рядов, т.е. что средние арифметические обоих рядов статистически достоверно отличаются друг от друга. Для этого находим разность между средними арифметическими:

$$D = \bar{x}_2 - \bar{x}_1 = 3{,}8 - 3{,}0 = 0{,}8;$$

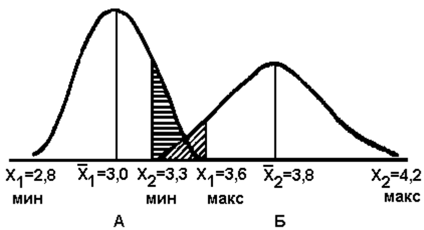
Зная ошибки средних арифметических, вычисляют m_D:

$$m_D = \sqrt{m^2 + m^2} = \sqrt{0{,}05^2 + 0{,}1^2} = 0{,}3\,.$$

Вычислив D и m_D, рассчитывают критерий достоверности t_D:

$$t_D = \frac{D}{m_D} = \frac{0{,}8}{0{,}3} = 2{,}7\,.$$

Рис. 11. Трансгрессивные кривые жирномолочности (для двух групп коров) [2];
А – животные с низкой жирностью молока; Б – помесные



Вычисляют t_1:

$$t_1 = \frac{x_{2\min} - \bar{x}_1}{\sigma_1} = \frac{3{,}3 - 3{,}0}{0{,}1} = \frac{0{,}3}{0{,}1} = 3$$

и по таблице 1 по вычисленному значению $t_1=3$ находят вторую функцию нормального распределения: $\varphi(t_1) = \varphi(3) = 0{,}49865$.

Аналогично вычисляют t_2 и определяют $\varphi(t_2)$:

$$t_2 = \frac{x_{1\max} - \bar{x}_2}{\sigma_2} = \frac{3{,}6 - 3{,}8}{0{,}2} = \frac{-0{,}2}{0{,}2} = -1$$

$$\varphi(t_2) = \varphi(-1) = 0{,}34134.$$

Далее определяют частоты p_1 и p_2, выражающие число членов совокупности, вошедших в отсеченную (заштрихованную) трансгрессивную часть обоих кривых. Эти величины можно выразить в виде разности между половиной площади кривой (0,5) и величиной $\varphi(t)$, которая определяет число наблюдений в части кривой в границах:

– от $\bar{x}_1$=3,0 до x_{2min}=3,3% в первой кривой и

– от $\bar{x}_2$=3,8 до x_{1max}=3,6% во второй кривой.

Следовательно, p_1–0,5–$\varphi(t)$=0,5–0,49865=0,00135, или 0,135% общего числа животных первой выборки, где n_1=500 голов. Тогда p_1= (0,135·500):100 = 0,675≈ 1.

Аналогичным образом находят величину p_2 по второй выборке: p_2=0,5 $\varphi(t)$=0,5–0,34134=0,15866, или ≈15,9%. Для второй группы помесных коров (n_2=100) это составляет 15,9 ≈ 16 голов.

Зная для первого и второго вариационного ряда, что n_1=500; p_1= =0,00135; n_2=100 и p_2=0,15866, находят коэффициент трансгрессии по формуле (8):

$$T = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2} = \frac{500 \cdot 0{,}00135 + 100 \cdot 0{,}15866}{500 + 100} = \frac{16{,}541}{600} = 0{,}0275$$

Эта величина, выраженная в процентах, свидетельствует, что 2,75% коров обеих выборок находятся в трансгрессирующей зоне, т.е. сходство с чёрно-пёстрыми коровами по величинам жирномолочности обнаруживают лишь 2,75% помесных животных. Отсюда следует заключение, что помесные коровы сильно отличаются от коров черно-пестрой породы по жирномолочности.

СЛОВАРЬ НОВЫХ ТЕРМИНОВ

Биномиальное распределение – такое распределение, при котором вероятности появления отдельных вариант выражаются величинами, соответствующими коэффициентам разложения бинома Ньютона.

Вероятность – возможность осуществления определенного события в некотором количестве случаев из общего числа возможных; **количественной характеристикой вероятности** какого-либо явления, может быть относительная частота явления, установленная эмпирически на достаточно значительном фактическом материале.

Доверительная вероятность – уровень вероятности, который считается достаточным для суждения о достоверности статистических показателей, получаемых на основании выборочных данных.

Нормированное отклонение – отклонение той или другой варианты (или группы вариант) от средней арифметической, выраженное в сигмах.

Статистическое распределение – распределение частот f_i отдельных вариант по значениям этих вариант x_i или по классам вариационного ряда.

Теорема сложения вероятностей: вероятность появления нескольких независимых и *несовместимых* событий равна сумме вероятностей этих событий.

Теорема умножения вероятностей: вероятность появления нескольких независимых, *но совместимых* друг с другом (или следующих друг за другом) событий равна произведению вероятностей этих событий.

Трасгрессивные ряды – такие вариационные ряды, которые отличаются друг от друга величиной средней арифметической, но у которых крайние классы, лежащие около максимального класса первой кривой, служат крайними минимальными классами другой кривой, что создаёт в этих частях вариационных кривых их взаимное пересечение.

Уровень значимости – величина, обозначающая вероятность получения случайного отклонения от установленных с определенной вероятностью результатов; показатель, используемый для проверки статистических гипотез и связанный с такими значениями вероятности, при которых появление ожидаемых событий в данных условиях считается практически невозможным.

ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

1. Понятие о статистическом распределении и вероятности.

1.1. Что называется статистическим распределением?
1.2. Что называется вероятностью?
1.3. Какое явление называется случайным?
1.4. Каковы причины возникновения случайного явления?
1.5. Чем случайное явление отличается от закономерного?
1.6. Какой процесс называется стохастическим?
1.7. Что является количественной характеристикой вероятности?
1.8. При каком условии проявляется вероятность некоторого события?
1.9.Сформулируйте и объясните теорему сложения вероятностей.
1.10. Сформулируйте и объясните теорему умножения вероятностей.

1.11. Объясните понятия «эмпирическая вероятность» и «теоретическая вероятность». Укажите взаимосвязь между ними.
1.12.Что называется стохастической совокупностью?

2. Понятие о нормальном распределении.

2.1. Охарактеризуйте вариационную кривую нормального распределения.
2.2. Укажите соотношение $\overline{X}$, моды и медианы в нормальном распределении.
2.3. Что отражает (чему равна) площадь фигуры, ограниченной осью абсцисс (x) и вариационной кривой нормального распределения?
2.4. Сформулируйте правило ±3σ.
2.5. Как можно определить частоту встречаемости особей, имеющих конкретную величину признака x_i (при нормальном распределении вариант)?
2.6. Как определяют вторую функцию нормального распределения φ(t)? Что позволяет определить эта функция в практическом анализе выборок?
2.7 Что позволяет определить третья функция нормального распределения F(t)?

3. Что называют нормированным отклонением?

4. Что называют доверительной вероятностью?

4.1.Сколько и какие уровни доверительной вероятности используют в биометрии?
4.2.Что означает каждый из этих уровней вероятности?

5. Что называют уровнем значимости?

5.1. Сколько и какие уровни значимости используют в биометрии?
5.2. Что означает каждый из этих уровней значимости?

6. На каких основаниях исключаются из вариационного ряда резко отклоняющиеся варианты?

7. Понятие о биномиальном распределении.

7.1. Как могут быть использованы закономерности биномиального распределения в практическом анализе выборок?

8. Общая характеристика распределения Пуассона.

8.1. Для каких практических целей можно использовать закономерности распределения Пуассона?

9. Какое распределение называется асимметричным?

9.1. Какие причины могут вызвать асимметрию?

10. Какое распределение называется эксцессивным?

10.1. Какие причины могут вызвать эксцесс
10.2.Какие причины могут вызвать двухвершинность?

11. Какое распределение называется трансгрессивным?

11.1. Укажите два свойства трансгрессивных вариационных рядов.
11.2. Как на практике можно использовать свойства трансгрессивных рядов?

ЗАДАЧИ

1. Какова вероятность, что в семье из 6 детей: а) все 6 будут девочками; б) все дети будут одного пола; в) что первые 5 детей будут девочками, а 6-й ребёнок – мальчиком?

2. Какова вероятность, что в семье, имеющей 7 детей: а) первые 3 будут девочками, а остальные 4 – мальчиками; б) что в семье будет, по крайней мере, одна девочка?

3. Какова вероятность, что при анализирующем скрещивании гороха, гетерозиготного по зеленой окраске горошин, выборка из 12 горошин будет: а) состоять только из зеленых горошин; б) только из желтых; в) содержать, по крайней мере, 1 желтую; с) содержать, по крайней мере, 1 зеленую?

4. В табуне лошадей гнедых было 250, а вороных – 150. Какова вероятность того, что одна из пойманных наудачу лошадей будет а) гнедой, б) вороной? Чему равна сумма этих двух вероятностей?

5. Какова вероятность, что в семье, имеющей 3 детей: а) пеервый будет мальчик, а остальные 2 – девочки; б) что последний будет мальчиком, а первые 2 – девочками; в) что будут 1 мальчик и 2 девочки (при любой последовательности рождения); г) что будут 2 мальчика и 1 девочка (при любой последовательности рождения)?

6. Предполагая, что частота рождения мальчиков 1/2, определите процент однояйцовых двоен по следующим данным: 235 615 пар было с 2 мальчиками, 220335– с 2 девочками; 265 291 пара состояла из 1 мальчика и 1 девочки.

7. При спаривании черных гетерозиготных по рыжей окраске коров с рыжим быком ожидается расщепление в потомстве 1:1. Какова вероятность, что все 6 телят от коров будут черными? Какова вероятность рождения двух черных телят (остальные рыжие)?

8. За период 6 лет было учтено 154 444 пары двоен. Из них было разнополых 58 382, только с мальчиками 49 425 и только с девочками 46 637.

Рассчитайте, какое количество пар двоен было однояйцовых. для простоты расчета примите, что соотношение полов точно 1:1. Какова вероятность рождения однояйцовых и двуяйцовых двоен?

9. По данным переписи населения в СССР в 1926 г. было 71043357 мужчин и 75984558 женщин.

Какова вероятность рождения особей мужского пола в населении?

10. 106 опоросов по 8 поросят в каждом распределились по числу самцов следующим образом:

Число самцов	1	2	3	4	5	6	7	8
Количество опоросов	5	9	22	25	26	14	4	1

Приняв, что в данном случае имеется биномиальное распределение, вычислите $\bar{x}$ и σ. С помощью $\bar{x}$ определите p и q. Попробуйте вычислить отдельные значения количества опоросов, развернув формулу $(p + q)^8$, при $n = 106$.

11. Проанализированы данные 10 000 семей с 4 детьми. Было установлено, что в 641 семье все дети были девочками, в 2625 семьях было по 3 девочки и 1 мальчику; в 3748 семьях – 2 девочки и 2 мальчика; в 2420 семьях в 1 девочка и 3 мальчика, в 566 семьях – все мальчики. Исходя из предположения о биномиальности распределения, вычислите вероятность рождения мальчиков и девочек.

12. Вероятность наступления некоторого события $p = 0{,}30$. Производится 2000 испытаний. В каком количестве случаев будет наблюдаться данное событие и каково его среднее квадратическое отклонение?

13. Среди 402 опоросов свиней джерсейской породы, в каждом из которых было 8 поросят, пометы распределялись следующим образом:

Количество самцов в помете	0	1	2	3	4	5	6	7	8
Количество опоросов	1	8	37	81	162	77	30	5	1

Определите $\bar{x}$ и σ обычным методом и по величине $\bar{x}$ определите p и q.

14. В 100 пробах, в каждой из которых находилось по 1200 зерен ржи, проверяли наличие двойных зародышей. Оказалось, что в некоторых пробах находили от 1 до 6 таких зародышей. Распределение найденных зерен с 2 зародышами по пробам было следующим:

Количество зерен с двумя зародышами	0	1	2	3	4	5	6
Число проб	6	24	32	18	9	6	5

Вычислите обычным путем среднюю арифметическую количества верен с 2 зародышами на пробу, а также варианcу данного ряда. К какому типу распределения следует отнести этот ряд? Какова вероятность нахождения зерен с 2 зародышами в общей популяции зерен ржи?

15. На 1000 мальчиков 13-летнего возраста было установлено, что 390 из них отклоняются от средней арифметической по росту (высоте тела) не более чем на 1,4 дюйма ($\overline{x}$ = 57,3 дюйма). Можно ли по этим данным определить примерную величину σ, если предусматривается нормальное распределение?

16. Известно, что группа коров, охватывающая 10 тысяч голов, имеет по удою за лактацию = 3200 кг и σ = 300 кг. Сколько в группе может быть коров, удои которых за лактацию превышают 4100 кг? В каких пределах колеблются удои преобладающей части группы (70%) коров?

17. Выловленная в пруду рыба имела среднюю живую массу 375 г и σ = 25 г. Какова вероятность, что вес дополнительно пойманных рыб будет:

а) в пределах от 325 до 425 г;
б) не более 400 г;
в) не более 425 г;
г) не менее 375 г;
д) не менее 350 г?

9.4. ОЦЕНКА ДОСТОВЕРНОСТИ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

9.4.1. Проблема достоверности в статистике.

Каждая совокупность может быть представлена в виде вариационного ряда. Для вариационного ряда можно определить статистические показатели, указывающие на наиболее типичный уровень проявления изучаемого признака, а также на степень вариации отдельных вариант (единиц наблюдения) вокруг этого уровня.

Большинство статистических показателей – именованные величины (например, средняя арифметическая, мода, медиана). Некоторые статистические показатели выражаются в процентах (коэффициент вариации) или же являются величинами неименованными (варианса, коэффициент асимметрии). Но так как все эти показатели являются *статистическими* величинами, т.е. основаны на изучении *массовых* явлений, возникает очень важный вопрос о том, насколько они достоверны.

9.4.2. Статистическая характеристика выборочных и генеральных совокупностей

Ранее мы выяснили, что **генеральная совокупность** – это вся подлежащая изучению совокупность данных объектов. В пределе она рассматривается как состоящая из бесконечно большого количества отдельных единиц ($N \to \infty$). Та часть объектов, которая подвергается исследованию, называется **выборочной совокупностью** или просто выборкой.

Оба типа совокупностей характеризуются одинаковыми закономерностями случайной вариации. Для их характеристики могут быть вычислены статистические показатели: средняя арифметическая и среднее квадратическое отклонение. Среднюю арифметическую мы обозначили символом $\bar{x}$. Теперь нам предстоит **различать среднюю арифметическую *выборочной* совокупности и среднюю арифметическую *генеральной* совокупности**. Условимся, что символ $\bar{x}$ теперь обозначает среднюю арифметическую *выборочной* совокупности. Среднюю арифметическую *генеральной* совокупности будем обозначать μ. (греческая буква мю). Определим, каково соотношение между $\bar{x}$ и μ?

Допустим, что для совокупности, состоящей из 168 коров симментальской породы, была получена средняя арифметическая глубины груди $\bar{x}$ =73,8 см. Группа в 168 коров представляют собой выборку из генеральной совокупности, охватывающей популяцию всех коров симментальской породы. Если бы мы взяли несколько выборок из популяции симментальской породы (например, из разных стад), то обнаружили бы, что средние арифметические $\overline{x}$ этих выборок будут различными.

Если выборок из различных стад достаточно большое количество, то можно обнаружить, очень важную особенность – их распределение близко к *нормальному*. Оказывается, что отдельные значения средних арифметических выборок ($\overline{x}$) варьируют вокруг средней арифметической генеральной совокупности μ. Вариация же выборочных средних $\overline{x}$ вокруг μ может быть измерена своим средним квадратическим отклонением σ. Эта сигма получила название *ошибки средней арифметической* или *средней ошибки*. Иногда ее называют также *стандартной ошибкой*. Именно она указывает на степень близости $\overline{x}$ и μ.

Ранее мы уже встречались с понятием ошибки средней арифметической (с. 32). Здесь она будет рассмотрена несколько подробнее.

9.4.2.1. Формула для ошибки средней арифметической

В предыдущем разделе (9.2.6.) мы кратко познакомились с ошибкой средней арифметической. Она была обозначена буквой ***m***. В некоторых руководствах по статистике используют другой символ средней ошибки – $s_{\overline{x}}$. Мы будем пользоваться первым обозначением.

Ошибку средней арифметической можно вычислить по формуле:

$$m = \frac{\sigma}{\sqrt{n}} \qquad (23)$$

Так как средние арифметические выборочных совокупностей варьируют вокруг средней арифметической генеральной совокупности (μ), то в формуле ошибки средней арифметической надо было бы взять среднее квадратическое отклонение (σ) *генеральной* совокупности. Чтобы получить его значение, пришлось бы изучить *всю* генеральную совокупность, что невозможно. Кроме того, тогда не было бы и надобности определять среднюю ошибку выборки.

Оказалось, что вместо среднего квадратического отклонения генеральной совокупности можно без большой погрешности взять для формулы (23) значение среднего квадратического отклонения *выборочной* совокупности. Разница между ними невелика, и при достаточно большом количестве наблюдений (практически несколько десятков) они будут равны друг другу.

В знаменателе формулы (23) под корнем *n* – объем выборочной совокупности. Это значит, что величина средней ошибки обратно пропорциональна численности выборочной совокупности.

В примере с глубиной груди у симментальских коров *n* = 168 и σ = 2,45. Отсюда средняя ошибка для средней арифметической глубины груди изученных 168 симментальских коров

$$m = \frac{\sigma}{\sqrt{n}} = \frac{2{,}45}{\sqrt{168}} = 0{,}17 \text{ см.}$$

9.4.2.2. Средняя ошибка – ошибка выборочности

Термин «ошибка» часто вводит в заблуждение начинающих аналитиков, которые предполагают, что она является результатом недостаточной аккуратности в работе. Это не так. Средняя ошибка – это статистическая ошибка. Она не имеет ничего общего с ошибкой точности. Статистические показатели для *выборочной* совокупности всегда имеют так называемые *ошибки выборочности* (их также называют *ошибками репрезентативности*). **Ошибка выборочности** – это средняя величина расхождения между средним значением изучаемого признака в выборке и генеральной совокупности. Из формулы (23) очевидно, что размер определяемой средней ошибки зависит от σ выборки и от её объема. Чем больше размеры выборки, тем меньше средняя ошибка.

Ветеринар и зооинженер почти всегда имеет дело с выборками – и при проведении опытов с животными, и при изучении отчётных материалов, полученных на производстве,– генеральные же совокупности остаются неизвестными. Поэтому нужно постоянно помнить о том риске, который сопутствует статистическим выводам. Часто эти выводы основываются на изучении *небольшого* материала, поэтому полученные в опытах или наблюдениях статистические показатели могут иметь значительные статистические ошибки. Из-за варьирования средних арифметических *нескольких выборок* вокруг средней арифметической *генеральной* совокупности один какой-либо опыт может дать результат, отклоняющийся от истинного на 2 или даже 3 значения ошибки выборочности. Но при значительном количестве опытов их результаты будут группироваться близко к центру распределения генеральной совокупности. Это дает возможность сделать правильное статистическое заключение.

9.4.2.3. Средние ошибки для σ и Cv

В некоторых случаях кроме ошибки средней арифметической могут понадобиться средние ошибки других биометрических показателей. Чаще всего бывают востребованы средние ошибки коэффициента вариации и среднего квадратического отклонения.

Среднюю ошибку для σ вычисляют по формуле:

$$m_{\sigma} = \frac{\sigma}{\sqrt{2n}} \tag{24}$$

Среднюю ошибку для *Cv* вычисляют по формуле:

$$m_{Cv} = \frac{Cv}{\sqrt{2n}} \tag{25}$$

Эти формулы следует применять только при статистической обработке больших выборок. При малых объёмах выборочной совокупности (n<30) применяются более сложные методы. Для знакомства с ними следует обратиться к специальным пособиям по биометрии.

Средняя ошибка и здесь дает возможность по такому же принципу, как для $\bar{x}$, определить доверительные границы для σ и *Cv*.

Допустим, что σ= 3,5, n = 200. Тогда $m_{\sigma} = \frac{3,5}{\sqrt{2 \cdot 200}}$=0,175. При уровне значимости P=0,01 t = 2,58. Доверительные границы для σ будут: 3,5–2,58•0,175 = 3,05 и 3,5+2,58•0,175=3,95. Это значит, что среднее квадратическое отклонение генеральной совокупности при уровне значимости 0,01 может находится между 3,05 и 3,95.

Формулы для вычисления *статистических ошибок* других биометрических показателей вы найдёте в учебнике [2, с.192], а также в специальных биометрических пособиях, список которых приведён в конце пособия.

Таким образом, с помощью средней ошибки можно установить, приняв определенную вероятность, возможные границы для колебаний средней арифметической, среднего квадратического отклонения и любых иных статистических показателей. Это дает возможность предсказать по выборочной совокупности свойства генеральной совокупности.

9.4.2.4. Закон больших чисел

Связь между статистическими показателями выборочных и генеральных совокупностей выражает так называемый *закон больших чисел*. Этот закон гласит, что чем больше число n некоторых случайных величин, тем их средняя арифметическая ближе к средней арифметической генеральной совокупности, тем меньше разница между $\bar{x}$ и μ.

9.4.2.5. Распределение $\bar{x}$ малых выборок

Иногда ветеринару и зооинженеру приходится обрабатывать данные о выборочных совокупностях, состоящих из очень ограниченного количества вариант или наблюдений. Когда выборки малы (n < 30), то возникает большое сомнение в возможности суждения по таким выборкам о генеральной совокупности. В значение нормированного отклонения $t=\frac{x_i-\bar{x}}{\sigma}$ может вкрасться значительная неточность.

Возникает вопрос о том, каковы в этих случаях закономерности распределения выборочных средних арифметических. Ответ на него в 1908 году дал английский математик Госсет, который публиковал свои научные работы под псевдонимом Стьюдент. Он нашёл распределение, которое зависит только от численности выборки. Поэтому изученное им распределение вероятностей получило название *t-распределения по Стьюденту*. Теоретическое обоснование закона распределения, открытого Стьюдентом, было дано Фишером. Очень важно, что t-распределение по Стьюденту может быть использовано при очень малых количествах вариант.

Величину t по Стьюденту-Фишеру вычисляют по формуле:

$$t=\frac{\bar{x}-\mu}{m} \qquad (26)$$

Легко видеть, что эта формула принципиально сходна с формулой нормированного отклонения $t=\frac{x_i-\bar{x}}{\sigma}$. Отличие формул в том, что в знаменателе находится не σ, а m. Величина же m вычисляется по формуле (23) – как частное от деления среднего квадратического отклонения выборочной совокупности на корень квадратный из численности той же совокупности:

$$m=\frac{\sigma}{\sqrt{n}}.$$

Чем меньше объём выборочной совокупности n, тем сильнее отличается распределение значений t от нормального. Поэтому вероятности расположения средних арифметических выборок в пределах определенных значений ± t значительно снижаются по сравнению с нормальным распределением. Другими словами, для достижения тех же вероятностей нужно взять значительно бóльшие интервалы $\bar{x} \pm t \cdot m$. Например, при n=5 вероятность 0,95 достигается лишь при t=±2,8, а вероятность 0,99 – при t =±4,6.

На рис. 12 представлены для сравнения две кривые: для нормального распределения при $n=\infty$ и для t-распределения при $n=5$. У нижней кривой края более растянуты вправо и влево. По 2,5% выборочных средних справа и слева отсекаются: в верхней кривой при $t=1{,}96$, в нижней – при $t = 2{,}78$. В обоих случаях вероятность – 0,95, а уровень значимости – 0,05.

В практической работе надо исходить из определенных уровней значимости, поэтому были составлены рабочие таблицы, по которым можно определять минимальное значение t, обязательно требующееся для данной вероятности (табл. 31, с. 119).

Таблица 31 построена на основе *заранее принятых* доверительных вероятностей и соответствующих им уровней значимости. Для упрощения в ней даны только 5 уровней значимости (0,1; 0,05; 0,02, 0,01 и 0,001), в полных таблицах обычно приводят и иные уровни значимости. Если, например, выборка включает только 10 наблюдений (число степеней свободы 9), а требуется по условиям опыта уровень значимости 0,01 (и доверительная вероятность 0,99), то величина t должна быть не менее 3,25. Уровню значимости 0,05 (и доверительной вероятности 0,95) удовлетворяет при $n = 9$ величина $t = 2{,}26$.

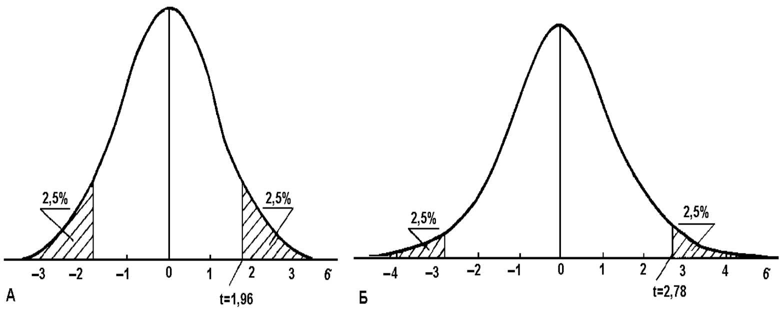


Рис. 12. Разные значения *t,* отсекающие по 2,5% площади справа и слева: *А* – под кривой нормального распределения (n=∞; t=1,96); Б – под кривой t-распределения по Стьюденту *(n* =5; t=2,78) [6, с.85].

По мере увеличения n t-распределение приближается к нормальному распределению. При $n>30$ разница между ними практически исчезает. Нижняя строка табл. 31, где $n=\infty$, связывает значения t по Стьюденту со значениями t для нормального распределения.

Исследования Стьюдента сыграли громадную роль, так как дали возможность работать с малыми выборочными совокупностями так же, как и с большими. При этом надо только учитывать различия в вероятностях для t в зависимости от размеров выборок.

9.4.2.6. Определение доверительного интервала для μ

Средние арифметические *выборок* колеблются вокруг средней арифметической *генеральной* совокупности μ. Поэтому по средним арифметическим выборок можно с некоторой вероятностью судить о μ. Для этого надо будет воспользоваться величиной ошибки m и нормированным отклонением t.

Можно установить с определенной вероятностью те границы, в которых находится средняя арифметическая генеральной совокупности μ, с помощью $\bar{x}$, m и t и формулы (27), а именно:

$$\bar{x} - t \cdot m \le \mu \le \bar{x} + t \cdot m \qquad (27)$$

или иначе,

$$\bar{x} - t \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + t \cdot \frac{\sigma}{\sqrt{n}} \qquad (27а)$$

Эти границы μ получили название **доверительных**. Интервал между максимумом и минимумом также называется **доверительным**.

Обычно заранее устанавливают ту или иную доверительную вероятность, с которой желают установить доверительные границы для μ, например p=0,95; p=0,99; p=0,999, что соответствует уровням значимости P=0,05; P=0,01; P=0,001. Чтобы указать, какой уровень значимости или вероятности принимается в данном случае, при букве t записывают показатель уровня значимости, например: t_{05} или t_{01}.

В примере с глубиной груди симментальского скота n = 168. Величина t_{05} (т.е. вероятность p=0,95, а уровень значимости P=0,05) по табл. 30 (см. с. 118) будет 1,96. Так как m=0,17 см, то доверительный интервал, в котором находится значение μ при уровне значимости 0,05, будет от 73,8–1,96· 0,17 = 73,5 до 73,8+1,96·0,17 = 74,2.

9.4.3. Нулевая гипотеза

Ошибки средней арифметической позволяют сравнивать между собой любые две группы организмов (например: две выборочные совокупности, взятые из двух разных стад одной породы; опытную и контрольную группы животных при постановке опытов) – и установить, насколько *статистически* достоверны различия между их статистическими показателями (средними арифметическими, вариансами и др.). Общие принципы сравнения основываются на анализе так называемой нулевой гипотезы.

Согласно **нулевой гипотезе**, первоначально принимается, что между данными показателями (и группами, на основе которых они получены) *достоверного различия нет*, т.е. что обе группы вместе составляют один и тот же однородный материал, одну совокупность.

Статистический анализ должен привести или *к отклонению* нулевой гипотезы, если будет доказана достоверность полученных различий, или к ее *принятию*, если достоверность различий не будет доказана, т.е. различия будут признаны случайными. Но так как все статистические показатели и различия между ними выполняются с определенными уровнями значимости *P*, то отбрасывание нулевой гипотезы должно быть связано с выбором определенного уровня значимости. Так, если признан необходимым уровень значимости P=0,01 и если вероятность достоверности различий между показателями не удовлетворяет этому условию, т.е. она ниже 0,99 (например, 0,97, 0,91, 0,88), то отвергать нулевую гипотезу не следует. Ее надо считать правильной, по крайней мере, до тех пор, пока новые данные не позволят ее опровергнуть, доказав, что существующие различия не являются случайными.

Конечно, и в том случае, когда нулевая гипотеза считается опровергнутой, какой-то шанс, что она в действительности верна, остается. При уровне значимости 0,01 этот шанс составляет 1 на 100, т.е. в 1 % случаев отбрасывание нулевой гипотезы может быть ошибочным. При P = 0,05 уверенность в правильности статистического вывода составляет лишь 95 случаев из 100, а в 5% случаев возможен неправильный вывод.

9.4.4. Оценка достоверности различий между средними арифметическими двух выборочных совокупностей

Очень часто в практической работе и исследованиях необходимо сравнить две *выборочные* совокупности, например: две группы морских свинок, подвергавшихся воздействию химических веществ или физических факторов, две группы коров, сравниваемые по удою и т.п. В этих случаях разница между средними арифметическими имеет свою статистическую ошибку, с которой ее можно сравнить и установить, достоверна эта разница или нет. Нулевая гипотеза в данном случае будет сводиться к тому, что две изучаемые выборочные совокупности происходят из одной и той же генеральной совокупности и разница между их средними арифметическими случайна, т.е. лежит в пределах ошибки выборочности.

Чтобы отвергнуть нулевую гипотезу, надо доказать, что разница между средними арифметическими *достоверна*, т.е. удовлетворяет требуемому уровню значимости.

Для определения достоверности разницы между средними арифметическими используют нормированное отклонение t, которое в этом случае обозначают буквами t_d и называют **критерием достоверности**. После некоторых математических преобразований (см.: [3], с.99; [6], с.93]) *рабочая формула* для расчёта достоверности различий средних арифметических двух выборочных совокупностей (критерия достоверности) приобретает следующий вид:

$$t_d = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{m_1^2 + m_2^2}} \ge t_{st}, \qquad (28)$$

где m_1 и m_2 – ошибки средних арифметических $\bar{x}_1$ и $\bar{x}_2$, соответственно. В таком виде эта формула используется в тех случаях, когда в обеих сравниваемых группах численность особей *больше* 30. Число степеней свободы при использовании этой формулы определяется следующим выражением:

$$\nu = n_1 + n_2 - 2 \qquad (29)$$

Вычисленное значение критерия достоверности (t_d) сравнивают (учитывая число степеней свободы и уровень доверительной вероятности) со стандартными значениям этого критерия t_{st}. Стандартные значения критерия достоверности рассчитаны Стьюдентом и приводятся в специальных справочных таблицах, имеющихся во всех учебниках биометрии и статистики. Если t_d больше табличной величины (t_{st}), то различия между сравниваемыми выборочными совокупностями статистически достоверны и нулевая гипотеза должна быть отвергнута. Если вычисленный критерий меньше табличного, то нулевую гипотезу отвергнуть нельзя.

Следует подчеркнуть следующее обстоятельство: так как при определении достоверности различий между средними арифметическими двух выборочных совокупностей расчётный критерий t_d сравнивают со стандартным критерием Стюдента (t_{st}), то на практике очень часто критерий достоверности называют **критерием Стьюдента**.

Допустим, что мы хотим сравнить по удою 2 группы коров, статистические показатели которых приведены ниже:

Номер выборки коров	n (голов)	$\bar{x}$ (кг)	m (кг)
№ 1	$n_1 = 50$	$\bar{x}_1 = 2100$	$m_1 = 120$
№ 2	$n_2 = 40$	$\bar{x}_2 = 2635$	$m_2 = 140$

Разница между средними удоями 2 групп составит

$$t_d = \frac{|\,2100 - 2635\,|}{\sqrt{120^2 + 140^2}} = \frac{535}{184} = 2{,}91\,.$$

По таблице нормального интеграла вероятности (табл. 30, с. 118) находим, что в этом случае вероятность достоверности (p) очень велика – 0,9964.

При отсутствии таблиц можно исходить из правила трех сигм: если разница превышает свою ошибку почти в три раза, она достоверна с вероятностью не менее 0,99. В практических расчётах при n>30, значение t=2,58 гарантирует достоверность различий с вероятностью 0,99.

9.4.5. Достоверность различий между двумя выборками животных, характеризуемых качественным признаком

До этого раздела все наши расчёты касались количественных признаков. Однако в животноводческой практике возникает необходимость в оценке животных и сравнении разных выборок по качественным признакам. Рассмотрим, как проводят эти оценки и сравнения.

9.4.5.1. Статистический анализ изменчивости качественных признаков

Средняя арифметическая качественных признаков отражает долю или процент особей, имеющих данный признак. Например, в одном хозяйстве из 1030 коров заболело лейкозом 28 голов, и 1002 остались здоровыми. Обозначим общую численность животных – n, численность первой группы (больные животные) – p_1, численность второй (здоровые животные) – p_0,. Доля больных (т.е. имеющих изучаемый признак) животных *(p)* определяют по формуле:

$$p = \frac{p_1}{n} = \frac{28}{1030} = 0{,}027 \text{ или } 2{,}7\%$$

Величина p соответствует средней арифметической ($\bar{x}$) при количественной изменчивости. Доля здоровых животных (q) составляет:

$$q = \frac{p_0}{n} = \frac{1002}{1030} = 0{,}973 \text{ или } 97{,}3\%$$

Среднее квадратическое отклонение при обработке подобных данных вычисляют по формуле:

$$\sigma = \sqrt{p \cdot q} \tag{30}$$

В нашем случае $\sigma = \sqrt{p \cdot q} = \sqrt{0{,}027 \cdot 0{,}973} = 0{,}163$ или 16,3%

Среднюю ошибку (m) частоты качественного признака, выраженную в долях единицы или в процентах, вычисляют по формуле

$$m = \sqrt{\frac{p \cdot (1 - p)}{n}} = \sqrt{\frac{p \cdot q}{n}} \tag{31}$$

Подставляем вычисленные нами величины в формулу:

$$m = \sqrt{\frac{0,027 \cdot 0,973}{1030}} = 0,005 \text{ или } 0,5\%.$$

Ошибка частот одинакова и для доли больных и доли здоровых животных: $p \pm m$=0,027+0,005 (или 2,7±0,5); $q \pm m$=0,973+0,005 (или 97,3±0,5%).

9.4.5.2. Определение достоверности различий между выборочными долями или процентами

На практике часто приходится определять – существуют ли статистически достоверные различия между частотами животных с определённым качественным признаком в двух разных группах.

В одном стаде из 82 дочерей быка № 1 заболели бруцеллезом 39 коров, а из 80 дочерей быка № 2 заболели 11 коров. Необходимо установить, различаются ли производители по восприимчивости дочерей к бруцеллезу. Для этого воспользуемся формулой (28) $t_d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{m_1^2 + m_2^2}}$, но вместо значений $\bar{x}$ подставим значения частот:

$$t_d = \frac{p_1 - p_2}{\sqrt{m_1^2 + m_2^2}} \qquad (32)$$

Определим долю больных дочерей: p_1=39**:**82=0,476 и p_2 =11**:**80=0,138. По формуле (31) вычислим m_1^2 и m_2^2.

$$m = \sqrt{\frac{0,476 \cdot (1 - 0,476)}{82}} = \sqrt{\frac{0,476 \cdot 0,524}{82}} = \sqrt{0,0030} = 0,055;$$

$$m = \sqrt{\frac{0,138 \cdot (1 - 0,138)}{80}} = \sqrt{\frac{0,138 \cdot 0,862}{80}} = \sqrt{0,0015} = 0,039.$$

$$t_d = \frac{0,476 - 0,138}{\sqrt{0,055^2 + 0,039^2}} = \frac{0,338}{\sqrt{0,0045}} = \frac{0,338}{0,067} = 14,9.$$

Стандартные значения критерия Стьюдента с учетом числа степеней свободы (формула 29) находим по табл. 31 (с. 119). Число степеней свободы равно $\nu = n_1 + n_2 - 2$ = 82+80–2 = 160. Такого значения ν в таблице нет, поэтому столь большую величину степеней свободы приравниваем ∞ и находим значения t_{st}, которые равны 1,96; 2,58; 3,29 для уровней значимости P_{05}, P_{01}, и P_{001}, соответственно. Так как вычисленная величина t_d = 14,9 больше t_{st} = 3,29 для третьего уровня вероятности, можно сделать вывод: разность между быками-производителями по частоте заболевания дочерей бруцеллезом достоверна с вероятностью p>0,999. Это значит, что дочери быка № 1 отличаются большей восприимчивостью к бруцеллезу, чем потомство производителя № 2.

9.4.5.3. Определение достоверности различий средних квадратических отклонений и варианс

Если сравниваемые группы численно достаточно велики, сравнение их изменчивости может быть проведено по тому же принципу, как и сравнение $\bar{x}$, т.е. с помощью показателя t. В данном случае

$$t = \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \qquad (33)$$

В силу ряда теоретических соображений, изложение которых выходит за рамки элементарного курса биометрии, значительно более точным методом для сравнения вариации и установления достоверности различий между сравниваемыми группами является так называемый **критерий F**, представляющий собой отношение варианс (средних квадратов):

$$F = \frac{\sigma_1^2}{\sigma_2^2} \qquad (34)$$

Обычно отношение варианс берут таким образом, чтобы в числителе была *большая* варианса.

Если обе вариансы σ_1^2 и σ_2^2 равны, тогда F=1. Очевидно, что нулевой гипотезой является признание равенства варианс. Если они не равны, то нужно доказать, что это неравенство достоверно. Значения F, являющиеся критериями для признания достоверности разницы между вариансами, приводятся в специальных таблицах, где учитываются а) разные объемы сравниваемых групп (вернее, разные числа степеней свободы этих групп) и б) различные уровни значимости. В несколько сокращенном виде они представлены в табл. 32 (с. 120-121 – для уровня значимости 0,05) и в табл. 33 (с. 122-123 – для уровня значимости 0,01).

Если полученная величина F больше табличного значения при принятом уровне значимости, различие между варианcами признается достоверным; если она меньше, то расхождение между варианcами может считаться несущественным, случайным, т.е. нулевая гипотеза остается не опровергнутой.

Во многих специальных разделах биометрии практическое значение F очень велико. Особенно важен критерий F в дисперсионном (или вариансном) анализе.

Можно привести и более простой пример использования критерия F. Допустим, нужно сравнить изменчивость по высоте в холке групп черно-пестрого и красно-пестрого скота. Для первой группы n_1 = 100 и σ^2_1 =16,32, для второго n_2= 42 и σ^2_2= 14,44. Тогда

$$F = \frac{16{,}32}{14{,}44} = 1{,}13$$

В таблицах 32 и 33 (с. 120, 122) в вертикальных столбцах нет цифры 100. Тогда надо взять $v_1 = \infty$. По горизонтали же можно взять v_2 = 40. Обратимся сначала к табл. 33. При уровне значимости 0,01 F должно быть больше 1,80. Этому уровню значимости полученное значение F явно не удовлетворяет. В таком случае, может быть, различие между варианcами σ^2_1 и σ^2_2 удовлетворяет уровню значимости 0,05? По таблице 32 F для $v_1 = \infty$ и v_2 = 40 равно 1,51. Фактическая величина F ниже и этой величины. Отсюда можно сделать вывод, что хотя черно-пестрый и красно-пестрый скот отличаются по масти, но их варианcы по высоте в холке достоверно не отличаются друг от друга. Вероятность различия между варианcами, как случайного, более 0,05. Нулевая гипотеза о равенстве варианс сохраняет свое значение и остается не опровергнутой.

9.4.6. Метод хи-квадрат (χ^2) и его применение

Сравнивать выборочные совокупности и определять особенности вариационных рядов можно различными статистическими методами. Одна группа методов основана на сравнении вариационных рядов путем сопоставления их параметров: $\bar{x}$, σ, *Mo*, *Me*, σ^2, *Cv*. Другая группа методов сравнения рядов основывается на сопоставлении частот (p_x) варьирующего признака одного ряда с частотами другого ряда (p_y).

В генетических исследованиях часто необходимо сравнить эмпирические частоты ($p_{эмп}$) с теоретически ожидаемыми ($p_{теор}$). Например, требуется определить степень соответствия эмпирических частот расщепления фенотипов (т.е. частот определённых экспериментально) во II поколении гибридного потомства от теоретически ожидаемых частот, обусловленных расщеплением по закону Менделя при дигибридном скрещивании на фенотипы 9 (A_B_) : 3 (A_bb) : 3 (aaB_) : 1 (aabb). Для сопоставления эмпирических и теоретических частот количественных и качественных признаков К. Пирсон (1900) предложил использовать критерий, получивший название критерий «хи-квадрат» (критерий χ^2), или *критерий соответствия*. Кроме того, критерий χ^2 также можно использовать для изучения связи между признаками (особенно имеющими качественные градации). Однако χ^2 указывает не на *степень* связи, а только на ее *наличие* или *отсутствие*.

Формула критерия χ^2 включает сумму дробей, каждая из которых получена от деления квадрата разности между эмпирическими ($p_{эмп}$) и теоретическими частотами ($p_{теор}$) на частоты теоретические ($p_{теор}$):

$$\chi^2 = \sum \frac{(p_{эмп} - p_{теор})^2}{p_{теор}} \qquad (35)$$

Величина χ^2 может принимать любые значения от 0 до $+\infty$. Если $\chi^2 = 0$, то наблюдается полное соответствие фактического расщепления теоретически ожидаемому.

Ограничения для использования метода «хи-квадрат»:

– критерий «хи-квадрат» нельзя применять, если частоты выражены процентами или долями, то есть *относительными* величинами.

– суммарный объем выборки (*n*) должен составлять не менее 20 (для проверки соответствия биномиальному распределению рекомендуют использовать выборку с объёмом не менее 50 вариант [1]);

– метод нельзя использовать если частота вариант в классах вариационного ряда *меньше* пяти (если же в крайних классах окажется менее пяти вариант, то их объединяют с частотами соседнего класса);

9.4.6.1. Определение числа степеней свободы при вычислении критерия χ^2 [2]

При использовании критерия χ^2 необходимо сравнивать вычисленное значение этого критерия с табличным значением величины χ^2. При этом обязательно нужно учитывать число степеней свободы (ν). неправильный выбор χ^2 может привести к неверному выбору табличного значения χ^2 и, следовательно, к неверному статистическому заключению.

Число степеней свободы при использовании метода χ^2 зависит от числа ограничивающих условий, ограничивающих свободу вариации. В частности:

– при нормальном распределении ограничивающими будут три условия: $\bar{x}$, σ и n, тогда $\nu=n-3$;

– при использовании распределения Пуассона будет два ограничивающих условия ($\bar{x}$ и n); тогда $\nu=n-2$;

– при четырех- или многопольных таблицах ν определяется как произведение числа строк (l_x) на число столбцов (l_y), взятых без единицы, т.е.:

$$\nu = (l_x-1)\cdot(l_y-1); \tag{36}$$

–при кодоминантных признаках (например, полиморфные системы белков) число степеней свободы определяется как число фенотипов минус число аллелей;

– при изучении расщепления особей в F_2 по фенотипу, число степеней свободы ν будет равно числу фенотипических классов минус единица. Так, при моногибридном скрещивании $\nu=(l_x-1)=2-1=1$; при дигибридном скрещивании получают четыре фенотипических класса, поэтому $\nu=4-l=3$.

Иногда требуется определить *общий* показатель χ^2 на основании *частных* величин χ^2 полученных в *отдельных* опытах или группах. В этом случае величина критерия χ^2 будет равна сумме частных значений:

$$\chi^2_{\text{общ}} = \chi^2_1 + \chi^2_2 + \ldots + \chi^2_n. \tag{37}$$

При сопоставлении частот вариационных рядов методом «хи-квадрат» исходят из **нулевой гипотезы** *(H_0)*. Согласно нулевой гипотезе между частотами сопоставляемых вариационных рядов нет достоверных различий. Следовательно, оба вариационных ряда являются выборками из одной генеральной совокупности. Если же частоты сравниваемых рядов достоверно различаются между собой, то ряды принадлежат разным генеральным совокупностям и нулевая гипотеза отвергается.

Для доказательства достоверности различий между частотами двух рядов надо сопоставить величину $\chi^2_{эмп}$, вычисленную по конкретным данным, с величиной χ^2, взятой из специальной таблицы ($\chi^2_{теор}$), с учетом числа степеней свободы ν (табл. 34 с. 124). Если окажется, что $\chi^2_{эмп} > \chi^2_{теор}$ при соответствующем уровне значимости, то нулевая гипотеза отвергается. Если же $\chi^2_{эмп} < \chi^2_{теор}$, то нулевая гипотеза остается в силе.

9.4.6.2. Определение достоверности различий между эмпирическими и теоретически ожидаемыми частотами вариант

Во многих случаях селекционной практики требуется определить соответствует ли полученное расщепление по фенотипам или генотипам теоретически ожидаемым частотам. Метод χ^2 позволяет проводить такие сравнения. При обработке эмпирического материала можно использовать два способа определения теоретических частот ($p_{теор}$).

Первый способ базируется на том, что *теоретическое соотношение частот известно* и на его основе можно установить число животных по классам варьирования.

Пример [2]. Необходимо установить соответствует ли теоретически ожидаемому расщеплению соотношение частот фенотипов в F_2 кроликов, полученных от скрещивания самок черных с нормальной шерстью (***AABB***) с пуховыми самцами-альбиносами (***aabb***). Доминантными признаками являются черная шерсть (***A***) и нормальная ее длина (***B***), а рецессивными – альбинизм (***a***) и пуховый тип шерсти (***b***).

В $\boldsymbol{F_1}$ всё потомство было с черной короткой шерстью (дигетерозиготным, ***AaBb***). При скрещивании дигетерозигот (***AaBb*** × ***AaBb***) получено 120 потомков, в том числе: 45 черных короткошерстных, 30 черных пуховых, 25 белых короткошерстных и 20 белых пуховых кроликов.

Теоретическое распределение по фенотипу особей, полученных во втором поколении при дигибридном скрещивании кроликов, по закону Менделя составляет 9:3:3:1. В соответствии с теоретически ожидаемым расщеплением можно ожидать следующее: черных короткошерстных – 9/16•120=67,7; черных пуховых – 3/16•120=22,5; белых короткошерстных – 3/16•120=22,5; белых пуховых – 1/16•120=7,5.

Занесём эти данные в таблицу (табл. 22) и выполним необходимые расчёты.

Таблица 22

Определение достоверности различий между эмирическими и теоретически ожидаемыми частотами расщепления по фенотипу у дигибридных кроликов

Частоты, голов	Фенотип кроликов				n
	Чёрные, коротко-шерстные	Чёрные длинно-шерстные	Белые, коротко-шерст-ные	Белые, длинно-шерстные	
$p_{эмп}$	45	30	25	20	120
$p_{теор}$	67,5	22,5	22,5	7,5	120
$p_{эмп}-p_{теор}$	–22,5	+7,5	+2,5	12,5	–
$(p_{эмп}-p_{теор})^2$	506,25	56,25	6,25	156,25	–
$\dfrac{\left(p_{эмп}-p_{теор}\right)^2}{p_{теор}}$	$\dfrac{506{,}25}{67{,}5}=$ =7,5	$\dfrac{56{,}25}{22{,}5}=$ =2,5	$\dfrac{6{,}25}{22{,}5}=$ =0,28	$\dfrac{156{,}25}{7{,}5}=$ =20,8	$\chi^2\approx$ ≈31,1

Табличное значение χ^2 (при числе степеней свободы ν =3 и P=0,01) составляет 11,3. Вычисленноеэмпирическое значение ≈31,1, т.е. значительно больше, чем $p_{теор}$ (11,3). Это доказывает достоверное различие между эмпирическими и теоретически ожидаемыми частотами фенотипов. Следовательно, полученная группа кроликов отклоняется по расщеплению фенотипов от закона Менделя при дигибридном скрещивании и свидетельствует о влиянии каких-то факторов, изменяющих тип расщепления по фенотипу в $\boldsymbol{F_2}$ помесей.

Пример. 25 семейств крупного рогатого скота были изучены на наличие в них животных, больных туберкулёзом. Данные приведены в табл. 23.

Таблица 23

Расчёт соответствия фактического распределения семейств по количеству больных туберкулезом коров теоретически ожидаемому (биномиальному) [1]

	Количество больных коров в семействе						
	5	4	3	2	1	0	
Фактическое	1	1	3	12	5	3	n=25
	Σ=5				Σ=8		
Теоретическое	0,2	1,6	5,3	8,6	7,0	2,3	n=25
	Σ=7,1				Σ=9,3		

Примечание: число классов (в связи с объединением крайних классов) l=3.

Определим, соответствует ли фактическое распределение семейств по количеству больных туберкулезом коров теоретически ожидаемому биноминальному.

При использовании χ^2 для оценки соответствия распределения вариант в вариационном ряду биномиальному или нормальному нужно, чтобы число вариант в *крайних* классах было не менее 5. Более точные результаты получаются также тогда, когда в выборке имеется 50 вариант и более. В связи с тем, что число вариант в крайних классах нашего вариационного ряда меньше 5, мы объединим три первых класса в один. Так же поступим с двумя последними классами. Подставляем данные таблицы в формулу (35) и определяем, что $\chi^2_{\text{эмпир}}=2{,}14$:

$$\chi^2 = \frac{(5-7{,}1)^2}{7{,}1} + \frac{(12-8{,}6)^2}{8{,}6} + \frac{(8-9{,}3)^2}{9{,}3} = 2{,}14$$

Так как имеется три фенотипических класса, то число степеней свободы $\nu=n–2=3–2=1$. Табличное значение χ^2 при $\nu=1$ и $p=0{,}95$ составляет 3,8. Поскольку фактическая величина χ^2 (2,14) меньше табличной, то различия между фактическим распределением и теоретически ожидаемым статистически недостоверны. Поэтому можно сделать заключение: фактическое распределение семейств по количеству больных туберкулезом коров соответствует теоретически ожидаемому (в данном случае биномиальному).

Второй способ определения достоверности различий между двумя выборочными совокупностями используется в тех случаях, когда *теоретические частоты* ($p_{теор}$) *не известны*, и их нужно определить.

В этом случае «теоретические частоты в многоклеточных таблицах (2-2, 3-2, 4-4 и др.) определяются путём получения пропорций фактических частот по градациям фактора к общему числу частот ($\Sigma p{:}n$) в выборке» [2, с 215]. Покажем на примере, как это делается.

Таблица 24

Эпирические данные о частоте мутантных и не мутантных культур у дрозофил в облучённой и контрольной группах

Группы дрозофил, l_x	Фактическое число культур дрозофилы, l_y		Всего
	образовавших мутации	не образовавших мутации	
Опытная	**p_1=50 (факт)**	**p_2=250 (факт)**	$p_1+p_2=300$
Контрольная	**p_3=10 (факт)**	**p_4=190 (факт)**	$p_3+p_4=200$
Всего	$p_1+p_3=60$	$p_2+p_4=440$	$n=500$

В опыте на дрозофиле изучали влияние γ-облучения родительского поколения на появление мутантов в потомстве и получили следующие данные: (табл. 24).

Для определения χ^2 необходимо знать теоретические частоты. Для этого находят суммы фактических частот по градациям факторов, полученные суммы перемножают и произведение делят на число частот во всей выборке ($\Sigma p_{факт}$ **:**n) (табл. 25).

Таблица 25

Таблица для расчёта теоретических частот

Группа дрозофил, l_x	Теоретическое число культур дрозофилы, l_y	
	образовавших мутации	не образовавших мутации
Опытная	$(p_1+p_2)\cdot(p_1+p_3):n$ (50+250)·(50+10)**:**500 $\boldsymbol{p_1}$ **=36 (теор)**	$(p_2+p_1)\cdot(p_2+p_4):n$ (250+50)·(250+190)**:**500 $\boldsymbol{p_2}$ **=264 (теор)**
Контрольная	$(p_3+p_1)\cdot(p_3+p_4):n$ (10+50)·(10+190)**:**500 $\boldsymbol{p_3}$ **=24 (теор)**	$(p_4+p_2)\cdot(p_4+p_3):n$ (50+250)·(50+10)**:**500 $\boldsymbol{p_4}$ **=176 (теор)**

Теперь мы знаем и фактические частоты, и теоретически ожидаемые. Их следует подставить в формулу (35):

$$\chi^2 = \sum \frac{(p_{эмп} - p_{теор})^2}{p_{теор}}.$$

$$\chi^2 = \frac{(50-36)^2}{36} + \frac{(250-264)^2}{264} + \frac{(10-24)^2}{24} + \frac{(190-176)^2}{176} =$$

$$=5{,}44+0{,}74+8{,}17+1{,}11 = 15{,}46$$

Число степеней свободы вычисляют по формуле $v=(l_x-1)\cdot(l_y-1)$ $=(2-1)\cdot(2-1)=1$. По табл. 34 (с. 124) находят теоретическую величину χ^2. При ν=1 и p=0,01 она составляет 6,63. Поскольку вычисленная величина χ^2 (15,46) больше табличной (6,63), можно сделать заключение, что γ-облучение родителей статистически достоверно увеличивает частоту мутаций в потомстве по сравнению с контрольной группой.

9.4.6.3. Использование критерия χ^2 для сравнения двух эмпирических распределений

Предположим, необходимо сравнить два вариационного ряда. Причём нужно не просто сравнить средние арифметические или какие-либо другие величины, характеризующие ряды, а сопоставить весь ход кривых распределения вариант в классах вариационного ряда. Такая задача возникает в тех случаях, когда нужно проверить однородность эмпирического материала (например, вы хотите объединить в одну выборку данные, полученные о величине взрослых животных в одном стаде, но данные были получены с интервалом в две недели). Если окажется, что две эмпирические совокупности распределены одинаково, то их можно будет считать выборками из одной и той же генеральной совокупности. Тогда их можно объединить в одну большую выборку, что приведёт к сужению доверительных интервалов для параметров. Такая проверка особенно важна в тех случаях, когда нужно объединить данные разных авторов.

Техника вычисления критерия хи-квадрат для установления разности в частотах двух сравниваемых эмпирических рядов сводится к тому, что в качестве частот неизвестной генеральной совокупности берут *полусуммы* частот по каждому классу сравниваемых рядов. Рассмотрим использование этого способа на примере.

Таблица 26

Живая масса кур двух разных пород

Середины классов по живому весу, кг	частоты кур первой породы p_1	частоты кур второй породы p_2	Теоретические частоты $p_3=(p_1+p_2):2$
1,2	5	9	(5+9) :2=7
1,5	6	10	(6+10) :2=8
1,8	8	16	(8+16) :2=12
2,1	10	20	(10+20) :2=15
2,4	20	16	(20+16) :2=18
2,7	14	14	(14+14) :2=14
3,0	17	5	(17+5) :2=11
3,3	15	5	(15+5) :2=10
3,6	5	5	(5+5) :2=5
	$\sum p_1=100$	$\sum p_2=100$	$\sum p_3=100$

Пусть требуется определить величину χ^2 для двух эмпирических рядов по живой массе кур двух разных пород [4, с.224]. Вторая порода кур была выведена из первой путём специальной системы скрещиваний и отбора. Данные о курах приведены в первых трёх графах табл. 26:

В 4-й графе табл. 5 вычисляют теоретические частоты. Для этого в каждом классе определяют полусумму частот обоих классов: $p_3=(p_1+p_2):2$. Затем составляют вспомогательную таблицу (табл. 27).

В первой графе таблицы 27 записывают последовательно частоты обоих рядов, во второй графе – усреднённые частоты из теоретического ряда, в третьей графе – разность между эмпирическими частотами и теоретическими. В четвёртой графе определяют частные значения χ^2 для каждой строчки, что осуществляется обычным путём из выражения: $(p_{эмп}–p_{теор})^2 : p_{теор}=\chi^2_{частн}$.

Таблица 27

Вычисление χ^2 при сравнении двух эмпирических распределений

Эмпирические частоты p_1 и p_2 в последовательной парной записи	Теоретические частоты (p_3)	Разница ($p_{эмп}–p_{теор}$)	$(p_{эмп}–p_{теор})^2$: $:p_{теор} = \chi^2_{частн}$
5	7	–2	$–2^2$:7=0,57
9	7	+2	$+2^2$:7=0,57
6	8	–2	$–2^2$:8=0,50
10	8	+2	$+2^2$:8=0,50
8	12	–4	$–4^2$:12=1,33
16	12	+4	$+4^2$:12=1,33
10	15	–5	$–5^2$:15=1,66
20	15	+5	$+5^2$:15=1,66
20	18	+2	$+2^2$:23=0,17
16	18	–2	$–2^2$:23=0,17
14	14	0	0^2:14=0
14	14	0	0^2:14=0
17	11	+6	$+6^2$:11=3,27
5	11	–6	$–6^2$:11=3,27
15	10	+5	$+5^2$:10=2,50
5	10	–5	$–5^2$:10=2,50
5	5	0	0^2:5=0
5	5	0	0^2:5=0
$\Sigma(p_1–p_2)$= 200	$\Sigma 2p_3$=200	–	$\Sigma\chi^2_{частн}= \chi^2_{общ}$ =20

В результате суммирования частных значений χ^2 каждой строчки получается общее значение χ^2 =20. Число степеней свободы в этом примере равно $v=n-1=9-1=8$. Находим по таблице (34 с. 124) значение χ^2 при $v=8$. Оно оказывается равным: а) 15,5 при P=0,95, б) 20,1 при P=0,99 и в) 26,1 при P=0,999. Сопоставив вычисленное значение χ^2 с табличным, можем сделать вывод, что различия двух эмпирических рядов статистически достоверны при P=0,95 и P=0,99. отсюда следует заключение что селекция существенно изменила живую массу кур новой породы по сравнению с исходной. Поэтому вторая порода может считаться качественно иной, и принадлежащей к другой генеральной совокупности.

Величину χ^2 для *четырехпольных* таблиц можно получить по упрощённой формуле без определения разности $(p_{\text{эмп}}-p_{\text{теор}})^2$. Эта формула имеет следующий вид [2, с.216]:

$$\chi^2 = \frac{(p_1 \cdot p_4 - p_2 \cdot p_3)^2 \cdot n}{(p_1 + p_2)\cdot(p_3 + p_4)\cdot(p_1 + p_3)\cdot(p_2 + p_4)} \qquad (38)$$

В приведенном выше примере (табл. 24) величина χ^2, рассчитанная по формуле (38), будет равна:

$$\chi^2 = \frac{(50\cdot 190 - 250\cdot 10)^2 \cdot 100}{300\cdot 200\cdot 60\cdot 440} = \frac{(9500-2500)^2\cdot 500}{1484000000} = \frac{24500}{1584} = 15{,}45$$

Как видите, получена такая же величина χ^2, как и при использовании ранее рекомендованной формулы.

Ещё один пример [1, с.143]. Предпринята попытка выяснить, связан ли антиген H^a H-системы групп крови свиней с синдромом злокачественной гипертермии. Свиньи могут быть гомозиготными (H^aH^a) и гетерозиготными (H^aH^-) по этому аллелю или же не иметь его, являясь гомозиготами по альтернативному аллелю (H^-H^-). Распределение свиней, чувствительных и нечувствительных к синдрому злокачественной гипертермии, по двум разным генотипическим классам дано в табл. 28.

Таблица 28

Синдром злокачественной гипертермии (MHS) у свиней с различными генотипами

Генотип	Чувствительность к гипертермии		Всего
	чувствительные	нечувствительные	
H^aH^a, H^aH^-	a=25	b=95	$a+b=120$
H^-H^-	c=15	d=175	$c+d=190$
Всего	$a+c=40$	$b+d=270$	$n=a+b+c+d=310$

По нулевой гипотезе этого исследования аллель H^a не связан с синдромом злокачественной гипертермии. Рассчитываем эмпирическое значение критерия χ^2:

$$\chi^2 = \frac{(ad - bc)^2 \cdot n}{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)} \tag{38а}$$

$$\chi^2 = \frac{(25\cdot 175 - 95\cdot 15)^2 \cdot 310}{(25+95)\cdot(15+175)\cdot(25+15)\cdot(95+175)} = 11$$

По формуле (36) находим число степеней свободы

$$v = (2 – 1)\cdot(2 – 1) = 1.$$

При таком значении v табличные величины χ^2 (табл. 34) равны 3,8 при P=0,05, 6,6 при P=0,01 и 10,8 при P=0,001. Так как $\chi^2_{эмп}$ (11) больше даже максимального табличного значения (10,8), то нулевая гипотеза *отвергается* с высоким уровнем вероятности ($p > 0{,}999$). Заключение: животные, имеющие антиген H^a, более чувствительны к синдрому злокачественной гипертермии.

С помощью критерия χ^2 можно сравнивать частоты двух эмпирических рядов между собой. В этом случае используют следующую рабочую формулу:

$$\chi^2 = \frac{1}{n_1 + n_2} \cdot \sum \left[\frac{(p_1 \cdot n_2 - p_2 \cdot n_1)^2}{p_1 + p_2} \right] \tag{39}$$

Рассмотрим использование этой формулы на примере [2, *с. 217*]. Форма вымени у коров – признак очень важный для современной технологии автоматического доения. Поэтому селекционеры стремятся вести селекцию на предпочтительные формы вымени. Этот признак был изучен у коров-матерей (n_1=200) чёрно-пёстрой породы и их помесных дочерей (n_2=210), полученных от скрещивания матерей с быками-голштинами. Полученные данные приведены в первых трёх графах табл. 29 (см. следующую стр.).

Таблица 29

Сравнение методом χ^2 распределения коров по форме вымени в двух выборках.

Форма вымени	Вариационные ряды (p)		$p_1 \cdot n_2$	$p_2 \cdot n_1$	$\frac{(p_1 \cdot n_2 - p_2 \cdot n_1)^2}{p_1 + p_2}$
	матери	дочери			
Чаше-образ-ное	30	50	30·210= =6300	50·200= =10000	$\frac{(6300-10000)^2}{410} = 33390$
Ванно-образ-ное	50	70	50·210= =10500	70·200= =14000	$\frac{(10500-14000)^2}{410} = 29878$
Округ-лое	100	80	100·210 = =21000	80·200= =16000	$\frac{(21000-16000)^2}{410} = 60975$
Козье	10	10	20·210= =4200	10·200= =2000	$\frac{(4200-2000)^2}{410} = 11805$
Всего	$n_1=\Sigma p_1=$ $=200$	$n_2=\Sigma p_2=$ $=210$			Число классов l=4

В соответствии с нулевой гипотезой в этом исследовании предполагается, что *скрещивание не приводит к улучшению формы вымени.*

Алгоритм вычислений эмпирической величины χ^2 сводится к следующему:

1) по каждому классу («чашеобразное», «ваннообразное» и т.д.) частоты классов первого ряда умножают на *общее число* наблюдений второго ряда ($p_1 \cdot n_2$). Результат вносят в четвёртую графу;

2) Частоты классов второго ряда умножают на общее число наблюдений первого ряда ($p_2 \cdot n_1$) и данные заносят в пятую графу;

3) разность между этими величинами возводят в квадрат, делят на сумму всех частот в обоих рядах и результат деления заносят в шестую графу;

4) величины, внесённые в шестую графу суммируют и умножают на коэфициент $1/n_1 \cdot n_2$ (см. первую дробь формулы 39).

$$\chi^2 = \frac{1}{200 \cdot 210} \cdot (33390 + 29878 + 60975 + 11805) = \frac{136048}{42000} \approx 3{,}24$$

В результате получаем искомую величину χ^2=3,24;

5) число степеней свободы ν= *l*–1 = 4–l =3.

При *P*=0,05 табличное значение χ^2=7,81 Вычисленная величина χ^2 по конкретным данным примера равна 3,24, т.е. меньше теоретического. Следовательно, скрещивание чёрно-пёстрых коров с быками-голштинами не оказывает статистически достоверного влияния на форму вымени коров-дочерей. Нулевая гипотеза сохраняется.

Таким образом, метод χ^2 позволяет сравнивать распределение частот эмпирических и теоретических рядов между собой для качественных и количественных признаков, а также частоты между двумя эмпирическими рядами. Эмпирическая величина χ^2 бóльшая, чем табличное, указывает, что различие в частотах статистически достоверны. В это случае нулевая гипотеза отвергается. Если эмпирическая величина χ^2 меньше табличного значения, то различия между сравниваемыми выборками недостоверны, то нулевая гипотеза сохраняется. Она свидетельствует о том, что две сравниваемые по частотам выборки принадлежат одной генеральной совокупности.

СЛОВАРЬ НОВЫХ ТЕРМИНОВ

Ошибка выборочности (ошибка репрезентативности) – средняя величина расхождения между средним значением изучаемого признака в выборке и генеральной совокупности.

Закон больших чисел – чем больше число *n* некоторых случайных величин, тем их средняя арифметическая ближе к средней арифметической генеральной совокупности, тем меньше разница между $\bar{x}$ и μ.

Доверительные границы – границы *доверительного интервала,* т.е. значения параметра, равноотстоящие от средней арифметической выборочной совокупности, между которыми с определённой вероятностью находится средняя арифметическая генеральной совокупности

Доверительный интервал – интервал значений статистического параметра, в границах которого с определённой вероятностью находится средняя арифметическая генеральной совокупности.

Нулевая гипотеза – исходное предположение, согласно которому между двумя группами показателей (и двумя выборками, в которых эти показатели были получены) статистически достоверного различия нет, т.е. что обе группы вместе составляют один и тот же однородный материал, одну совокупность.

Критерий Фишера или **F-критерий** – статистический показатель, используемый для проверки гипотезы о равенстве дисперсий (вариации) двух генеральных совокупностей, имеющих нормальный закон распределения, и равный отношению варианс выборок из этих генеральных совокупностей.

Критерий хи-квадрат (критерий соответствия) – статистический критерий, показывающий соответствие экспериментально полученных частот признаков теоретически ожидаемым частотам.

Критерий Стьюдента или **критерий достоверности** t_d – статистический показатель, используемый для оценки достоверности различий между средними арифметическими двух выборочных совокупностей при заданном уровне доверительной вероятности и вычисляемый как частное от деления разности средних арифметических двух выборок $|\bar{x}_1 - \bar{x}_2|$ на корень квадратный из суммы квадратов ошибок средних арифметических $\sqrt{m_1^2 + m_2^2}$.

ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

1. Чем отличаются, и как связаны между собой выборочная и генеральная совокупности?

1.1. Отличаются ли друг от друга по закономерностям случайной вариации выборочная и генеральная совокупности?
1.2. В какой степени средняя арифметическая выборочной совокупности характеризует среднюю арифметическую генеральной совокупности?
1.3. Как колеблются $\bar{x}$ отдельных выборок вокруг средней арифметической генеральной совокупности?
1.4. Как вычисляется ошибка средней арифметической?
1.5. Что называется ошибкой выборочности? Каков её статистический смысл?
1.6. Как вычисляется средняя ошибка среднего квадратического отклонения?
1.7. Как вычисляется средняя ошибка коэффициента вариации?
1.8. Сформулируйте закон больших чисел.
1.9. Что называется t-распределением по Стьюденту?
1.10. Что называется t-критерием Стьюдента-Фишера?

2. Как определяют величину доверительно интервала для *μ* – средней арифметикой генеральной совокупности?

2.1. Изменяются ли доверительные границы и доверительный интервал для *μ* при разных величинах *n*?

3. Каков смысл нулевой гипотезы?

4. Как определяется достоверность различий между средними арифметическими двух выборочных совокупностей?

4.1.Какова рабочая формула для вычисления критерия достоверности различий между средними арифметическими двух выборочных совокупностей?
4.2.Как вычисляется число степеней свободы при вычислении критерия достоверности?
4.3.Как определяют достоверность различий между выборочными долями и процентами животных, в двух разных выборках?

5. Как определяют достоверность различий между средними квадратическими отклонениями и вариансами?

6. Для каких целей может быть использован метод хи-квадрат?

6.1. Какова общая формула критерия χ^2?
6.2. Какие значения может принимать критерий χ^2?
6.3. Каковы ограничения для использования критерия χ^2?
6.4. Как определяют число степеней свободы при использовании критерия χ^2?
6.5. Опишите алгоритм расчёта критерия χ^2 при определении достоверности различий между эмпирическими и теоретически ожидаемыми частотами вариант.
6.6. Опишите алгоритм расчёта критерия χ^2 при определении достоверности различий между двумя эмпирическими распределениями.

ЗАДАЧИ

1. При обследовании состояния гланд у студентов первого курса ветеринарного факультета из 84 студентов 39 человек оказались больными. При аналогичном обследовании второкурсников зоофака из 44 студентов больными оказались 28 человек.

Определите среднюю арифметическую и ошибку средней арифметической для обеих выборок. Определите, есть ли статистически достоверные различия в состоянии здоровья между группами.

2. Двенадцать черно-пестрых коров были покрыты джерсейским быком. Получены следующие данные о количестве молока за лактацию (в кг) и о жирности молока у матерей и дочерей:

Матери		Дочери	
количество молока	% жира	количество молока	% жира
1983	3,25	3509	5,29
3674	3,81	3110	6,04
3976	2,96	3181	5,24
3391	3,24	2997	5,25
4344	2,82	2991	5,14
3784	2,83	3720	4,72
3628	2,79	3268	4,54
3957	3,08	3595	4,97
2185	3,01	2939	5,13
4980	3,23	3213	4,98
2709	3,68	3240	5,58
2807	2,96	3388	4,81

Сравните удои и жирность молока дочерей и матерей.

3. Средний процент жира в молоке за лактацию коров холмогорских помесей был следующим: 3,4; 3,6; 3,2; 3,1; 2,9; 3,7; 3,2; 3,6; 4,0; 3,4; 4,1; 3,8; 3,4; 4,0; 3,3; 3,7; 3,5; 3,6; 3,4; 3,8. Определите $\bar{x}$, σ и *m*. Установите доверительные интервалы для $\bar{x}$ при вероятности *p*=0,99; при вероятности *p*=0,95.

4. При обследовании 150 взрослых мужчин средний рост был равен 167 см, а σ = 6 см. Определите, в каких пределах находится средняя арифметическая генеральной совокупности а) с вероятностью 0,99 и б) с вероятностью 0,95?

5. Были получены следующие данные о весе тушканчиков (*Dipus aegyptius*):

Самцы	186	190	165	182	182	182	180
	173	157	179	164	146	173	144
	156	156	165	160	160	161	144
	153	152	151	173			
Самки	162	163	190	188	147	146	145
	157	162	186	175	147	145	145
	155	174	180	148	175	145	144
	153	165	141	164			

Отличаются ли по весу самцы тушканчика от самок?

6. Температура тела тушканчиков оказалась следующей:

У самцов	37,5	37,9	37,4	37,8	36,8	37,8	37,5
У самок	37,8	38,1	37,0	37,5	37,7	37,8	37,6

Отличаются ли самцы и самки по температуре тела?

7. У серебристо-черных лисиц подмосковных звероводческих хозяйств было подсчитано количество желтых тел в яичниках:

В 1933–1935 гг.	5	3	8	4	6	4	5	9	5	4
	5	4	3	5	5	5	5	7	4	6
	5	7	5	3	3	5	4	5	4	5
В 1959–1962 гг.	6	7	4	5	5	6	7	5	6	8
	7	7	5	6	4	5	7	5	6	4
	5	7	6	5	5	5	6	6	6	5
	6	8	7	7	6	6	6	7	7	6
	5	9	8	8	5	8	5	6	7	8

Достоверно ли различие по числу желтых тел на самку за 2 периода времени?

8. У баранов мериносовой породы были произведены промеры рогов (в см):

Годовалый возраст	47	53	50	56	49	52	51	58	55
	50	48	51	51	48	60	51	57	57
	51	54	52	58	50	51	51	58	53
	52	49	59	61	50	52	51	63	62
	54	53	54	68	54	63	64	57	57
	60	57	60	69	57	56	54	54	55
	61	59	57	70	58	57	55	53	57
Трехлетний возраст	83	87	89	77	80	78	82	75	73
	71	76	86	90	84	88	72	73	68
	72	75	74	81	91	85	79	78	69
	68	73	68	71	75	71	76	72	66
	63	65	66	69	73	69	71	72	67
	70	64	51	67	70	66	62	65	74

Различаются ли по длине рогов бараны годичного и трехлетнего возраста?

9. Для 7 коров известны следующие данные об их убойном весе (в кг) в теплом состоянии ***x*** и после охлаждения ***y***:

x	322,6	250,6	287,3	408,1	338,0	213,5	323,3
y	318,9	247,0	279,7	403,0	334,7	209,3	319,2

Определите достоверность разницы между средним убойным весом в теплом состоянии и средним убойным весом после охлаждения.

10. Для определения pH применили 2 типа электродов. При первом показания pH: 5,78; 5,74; 5,84; 5,80; при втором – 5,82; 5,87; 5,96; 5,89. Следует ли отбросить нулевую гипотезу?

11. В результате испытания племенных качеств быка были получены следующие данные:

Удой матерей, кг	2500	2600	2400	2000	3000	2600	2400
Удой дочерей, кг	2900	3000	2600	2100	2800	2500	2400

Какой вывод о племенных качествах быка можно сделать на основании этих данных?

12. Пробы по 15 зерен кукурузы разных стадий зрелости проверяли на устойчивость к раздавливанию. Пробы дали следующие цифры (в единицах давления):
Первая проба: 42, 50, 36, 34, 45, 56, 42, 53, 25, 65, 33, 40, 39, 43, 42;
Вторая проба: 43, 44, 51, 40, 29, 49, 39, 59, 43, 48, 67, 44, 46, 54, 64.

Проверьте, достоверно ли различие между двумя $\overline{x}$?

13. Было изучено общее содержание азота в плазме крови крыс-альбиносов в возрасте 37 и 180 дней. Результаты выражены в граммах на 100 куб. см плазмы. В возрасте 37 дней 9 крыс, имели: 0,98; 0,83; 0,99; 0,86; 0,90; 0,81; 0,94; 0,92 и 0,87. В возрасте 180 дней 8 крыс имели: 1,20; 1,18; 1,33; 1,21; 1,20; 1,07; 1,13 и 1,12.

Установите доверительные интервалы для разницы с вероятностью p=0,95.

14. На 10 парах крыс определяли биологическую ценность белков земляного ореха – сырого *P* и жареного *R*. Пары данных (в условных единицах) были следующими:
61–55, 60–54, 56–47, 63–59, 56–51, 63–61, 59–57, 56–54, 44–63, 61–58. Достоверна ли разница? Какой метод можно применить для установления ошибки разницы? На сколько изменятся результаты, если исключить резко отличающуюся от остальных пару данных 44–63? Достаточны ли полученные данные для того, чтобы можно было сделать какой-либо вывод?

15. Получены следующие данные о длине крыльев у представителей 3 видов скворцов *Sturnus* из Индии (в мм). Определите, достоверна ли разница длин крыльев этих трех видов скворцов?

St. contra	120	120	121	122	125	126	126	125	122	123	122
St. gingi-niamus	129	123	128	125	126	127	129	125	124	129	128
St. fuscus	122	127	125	128	127	122	129	127			

16. Для изучения влияния рационов с добавкой 10 мкг витамина B_{12} на рост свиней было составлено 16 групп, в каждой из которых было по 6 голов. Средние суточные привесы в фунтах (на 100 фунтов живого веса) для каждой группы представлены в таблице:

Рационы	Пары групп							
	1	2	3	4	5	6	7	8
С B_{12}	1,60	1,68	1,75	1,64	1,75	1,79	1,78	1,77
Без B_{12}	1,56	1,52	1,52	1,49	1,59	1,56	1,60	1,56
Привес	0,04	0,16	0,23	0,15	0,16	0,23	0,18	0,21

Какова достоверность различий?

17. Имеются следующие данные об удоях 12 коров-матерей и их дочерей по полновозрастным лактациям (в кг):

Удой матерей	3770	3817	2450	3463	3500	5544	3112	3150	3118	3018	4291	3463
Удой дочерей	2991	4593	3529	4274	3103	3947	3491	3559	2916	4580	4510	4144

Достоверна ли разница между удоями матерей и дочерей? Какой метод сравнения можно применить?

18. В опыте по откорму 15 баранов получали ежедневно в качестве подкормки по 5 г фосфорной муки, 15 других баранов примерно того же возраста, веса и происхождения были контрольными. Суточный привес был следующим (в г):

Опытная группа	234	277	214	201	174	167	184	157	196	173	190	191	141	150	191
Контроль	183	154	175	159	157	189	198	165	176	124	173	182	204	151	147

Каким методом можно установить, достоверна ли разница между опытной и контрольной группами по суточному привесу? Определите эту разницу и выясните, достоверна ли она. Выясните также, отличались ли опытная и контрольная группы по вариансе и достоверно ли это различие.

19. При изучении суточных привесов 30 баранчиков выяснилось, что они происходят от 4 разных производителей. Данные о привесах потомков этих производителей были следующие:

Производитель	Привесы							
A	124	151	196	141	174	201	147	157
B	183	150	198	191	154	173	157	159
C	234	167	189	165	175	190	176	
D	173	184	277	214	182	191	204	

Определите $\bar{x}$ и σ для привесов каждой группы баранчиков, общую варианcу и варианcу между группами. Установите с помощью критерия *F* достоверность разницы между этими варианcами.

20. Были получены следующие (в мм) распределения по длине хвоста в 2 выборках оленьих мышей *Peromyscus*, взятых из географически удаленных мест. Достоверна ли разница между ними?

Классы (в мм)	1-я выборка (из Айовы)	2-я выборка (из Сев. Дакоты)
52–53	1	0
54–55	3	1
56–57	11	2
58–59	18	2
60–61	21	3
62–63	20	8
64–65	9	25
66–67	2	11
68–69	1	10
70–71	0	5
72–73	0	7
74–75	0	2
76–77	0	2

21. Для проверки последовательного действия противоглистных препаратов проверяли количество яиц гельминтов на грамм кала до и после дачи препарата. Знаком «+» обозначались случаи снижения количества гельминтов после дачи препарата, знаком «–» – увеличения или сохранения прежнего количества. В 9 пробах было 7 случаев «+» и 2 случая «–». Можно ли считать влияние препарата доказанным?

22. В таблице (см. следующую страницу) приведены годовые удои коров и удои коров- дочерей, рождённых от разных отцов.

Удои коров-матерей, кг	Удои коров дочерей, кг	
	бык-отец Фагот	бык-отец Баян
3039	4107	2750
3224	3478	3890
3444	3763	3568
3441	3836	3891
3629	3507	3695

Дайте статистически обоснованную оценку племенных качеств этих быков.

23. Был проведен опыт по подкормке 32 свиноматок препаратом афаромом, содержавшим железо и медь, в целях уменьшения процента мертворожденных поросят. От каждой матки получали 1 опорос, когда добавляли в корм афаром, и 1 опорос контрольный, когда добавки препарата не было. Маток покрывали всегда одними и теми же хряками. Получены следующие результаты:

Номера маток	Мертворожденные,%		Номера маток	Мертворожденные, %	
	при афароме	без афарома		при афароме	без афарома
1	0	8,3	17	0	35,7
2	0	12,5	18	0	63,6
3	0	9,1	19	0	9,1
4	18,2	22,2	20	0	10,0
5	0	10,0	21	22,2	40,0
6	25,0	33,3	22	0	0
7	10,0	0	23	0	0
8	11,1	0	24	0	0
9	0	16,7	25	0	0
10	0	28,6	26	0	0
11	0	25,0	27	0	0
12	11,1	0	28	0	0
13	11,1	0	29	0	0
14	0	25,0	30	0	0
15	0	9,1	31	0	0
16	0	14,3	32	0	0

Установите, достоверно ли различие между процентами мертворожденных поросят в опытной и контрольной группах?

Таблица 30

Таблица вероятностей при нормальном распределении. Доли площади под нормальной кривой в пределах от $-t$ до $+t$ [6, с. 294].

t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1113	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	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6680	6729	6778
1,0	6827	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	8029
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	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1.8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9960	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9981	9986	9990	9993	9995	9997	9998	9999	9999

Таблица 31

Значения *t* при различных уровнях значимости (*P*) [6, с. 297]

Число степеней свободы, *v*	Уровни значимости (*P*)		
	0,05	0,01	0,001
1	12,7	63,66	–
2	4,30	9,93	31,60
3	3,18	5,84	12,94
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
∞	1,96	2,58	3,29

Таблица 32

Значение F при уровне значимости **0,05** (v_1 – число степеней свободы для большей variансы, которая берётся числителем) [6, *с. 299*]

v_2	v_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	250	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,46	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,94	5,91	5,86	5,80	5,75	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	4,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	3,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,57	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,47	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,19	2,01

Таблица 32 (окончание)

v_2	v_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,04	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,01	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,94	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,92	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,74	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,65	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	202	1,96	1,91	1,83	1,75	1,66	1,55	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,46	1,00

Таблица 33

Значение F при уровне значимости **0,01** (v_1 – число степеней свободы для большей вариансы, которая берётся числителем) [6, *с. 301*]

v_2	v_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
1	4052	4999	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6261	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,47	99,50
3	34,12	30,82	29,46	28,71	28,42	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,50	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,84	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,38	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,23	6,88
7	12,25	9,55	8,47	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	5,99	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,20	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,65	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,25	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	3,94	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,70	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,51	3,16
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,35	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,21	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,10	2,75

Таблица 33 (окончание)

v_2	v_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,00	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	2,92	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,61	3,52	3,43	3,30	3,15	3,00	2,84	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,78	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,72	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,67	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,62	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,58	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,54	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,50	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,47	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,44	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,41	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	7,39	2,01
40	7,31	5,18	4,31	3,85	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,20	1,80
60	7,08	4,98	4,13	3,63	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,03	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,86	1,39
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,70	1,00

Таблица 34.

Теоретические значения $\chi^2_{теор}$ при разных степенях свободы (ν) и критических уровнях значимости [2]

Число степеней свободы, ν	Уровень значимости			Число степеней свободы, ν	Уровень значимости		
	P=0,05	*P*=0,01	*P*=0,001		*P*=0,05	*P*=0,01	*P*=0,001
1	3,84	6,63	10,8	18	28,87	34,81	42,3
2	5,99	9,21	13,8	19	30,14	36,19	43,8
3	7,81	11,34	16,3	20	31,41	37,57	45,3
4	9,49	13,28	18,5	21	32,67	38,93	
5	11,07	12,83	20,5	22	33,92	40,29	
6	12,59	16,81	22,5	23	35,17	41,64	
7	14,07	18,48	24,3	24	36,42	42,98	
8	15,51	20,09	26,1	25	37,65	44,31	
9	16,92	21,67	27,9	26	38,89	45,64	
10	18,31	23,21	29,6	27	40,11	46,96	
11	19,68	24,72	31,3	28	41,34	48,28	
12	21,03	26,22	32,9	29	42,56	49,59	
13	22,36	27,69	34,5	30	43,77	50,89	
14	23,68	29,14	36,1	…			
15	25,00	30,58	37,7	50	67,50	76,15	
16	26,30	32,00	39,3	80	101,88	112,3	
17	27,59	33,41	40,8	100	124,34	135,8	

9.5. СТАТИСТИЧЕСКАЯ СВЯЗЬ МЕЖДУ ПРИЗНАКАМИ И МЕТОДЫ ЕЕ ИЗМЕРЕНИЯ

9.5.1. Основные положения. Типы связей

В большинстве исследований животных, требуется выяснить связь между различными признаками, Например, для животновода может представлять интерес связь между жирностью молока и его белковостью у коров определённой породы, связь уровня удоя с уровнем кормления, связь между кровностью помесей и их продуктивностью и т.п.

В природе могут быть обнаружены связи двух типов: функциональные и корреляционные. Функциональные связи чаще обнаруживаются в физических и химических процессах. В биологии и, следовательно, в животноводстве, специалисты чаще встречаются с корреляционными связями.

Функциональная связь – это такая связь между признаками (показателями), при которой с изменением одного признака (показателя) на *определенную* величину, другой признак (показатель) изменяется *тоже на определенную величину*.

Примеры функциональной связи: 1) при изменении температуры газа на 5° его объем увеличится на определенную величину; 2) при изменении радиуса круга на 10 см площадь круга изменится также на определенную величину и т.п.

Корреляционная связь – это такая связь, при которой изменение одного признака *у ряда особей* на *определенную величину* сопровождается изменениями другого признака на *различные* (варьирующие) значения.

Пример корреляционной связи: если для группы из 30 коров питательность рациона увеличить на 1 кормовую единицу, то прибавка удоя в среднем по стаду составит какую-то определённую величину. Вместе с тем, каждого конкретного животного увеличение надоя молока будет разным. Следовательно, при изменении воздействующего фактора на строго определённую величину удой у отдельных животных будет варьировать в разной степени. Такая вариабельность удоя объясняется различной наследственностью животных, определяющей их физиологические особенности, а также воздействием многих других случайных факторов, которые влияют на животных.

В организме животного все процессы и все признаки взаимно связаны. Кроме того, каждый из них, в свою очередь, испытывает влияние внешней среды. Поэтому корреляционные связи в животноводстве являются широко распространенными и требуют углубленного изучения.

Корреляционные связи между признаками организмов изучают различными биологическими методами, которые позволяют вскрыть биологическую природу взаимосвязи. Но, кроме биологических методов, для выяснения величины, типа и направления связи используют методы математического анализа групп особей. Для этого используют несколько статистических коэффициентов, каждый из которых позволяет выявить различные стороны корреляционной связи.

По своим математическим особенностям корреляционная связь различных признаков может быть:

– прямой (положительной) и обратной (отрицательной);
– прямолинейной и криволинейной;
– простой и множественной;
– между количественными признаками и между качественными признаками.

Прямой (положительной) **связью** называется такая связь, при которой с увеличением (уменьшением) одного признака другой также увеличивается (уменьшается), т.е. изменение значений признаков происходит в *одном направлении*. Пример прямой (положительной) связи: по мере увеличения роста животных увеличивается их живая масса.

Обратной (отрицательной) **связью** называется такая связь, при которой с увеличением одного показателя другой уменьшается, т.е. изменение значений признаков происходит *в противоположных направлениях*. Пример обратной (отрицательной) связи: увеличение дозы γ-облучения животного вызывает уменьшение его плодовитости.

Прямолинейная связь – это такая связь, при которой изменение значений одного признака вызывает *пропорциональное* изменение значений другого признака.

Криволинейная связь – это такая связь, при которой изменение значений одного признака вызывает сначала увеличение, а затем уменьшение значения другого признака. Например, с увеличением возраста коров удой за лактацию вначале повышается до 5-8 отела, а затем по мере дальнейшего увеличения возраста он начинает снижаться. Графически криволинейные связи изображаются кривыми различного типа.

Простая корреляционная связь – это связь между *двумя* признаками (например, связь удоя и живой массы), *без учета других существующих связей*.

Множественная корреляционная связь выясняет связь сразу между несколькими показателями (удой, порода, возраст, живой вес).

Корреляционные связи могут выявляться не только **между количественными признаками** (например, связь величины удоя и питательностью рациона), но и **между качественными признаками** (например, связь между типом рациона родителей и полом потомства).

Коэффициенты связи могут быть вычислены как между признаками, измеряемыми одноименными мерами (живой вес и удой – в кг), так и измеряемые разноименными мерами (живой вес в кг и пол животного).

Следует подчеркнуть, что с помощью коэффициентов связи можно определить *величину* связи, *направление* и её *тип*. Однако, коэффициенты связи не вскрывают *причинности связи* между изучаемыми показателями. Причинность взаимосвязи между различными признаками и факторами может быть определена только физиологическими, биохимическими, биофизическими и другими исключительно биологическими методами.

Наиболее часто при обработке биологического материала применяют следующие коэффициенты связи:

1) r – коэффициент корреляции для малых и больших выборок,
2) b – коэффициент регрессии (обозначают также символом R),
3) r_a – коэффициент корреляции для альтернативных признаков,
4) r_s; – ранговый коэффициент корреляции Спирмена [1, *с.144-150*; 2, *с.178-192*; 5, *с.139*].

Эти частные коэффициенты корреляции будут рассмотрены ниже.

9.5.2. Общая характеристика коэффициентов корреляции r

Коэффициенты корреляции вычисляют для выявления связи между признаками и её количественной характеристики.

Все коэффициенты корреляции имеют ряд общих свойств:

1) Коэффициент корреляции r выражается десятичной дробью и может принимать значения от 0 до ±1.

2) Чем ближе значение r к 1, тем больше связь между данными признаками. Если r=0 или имеет небольшое дробное значение, то это указывает на отсутствие связи или на существование незначительной связи между признаками. Связь считают *малой*, если коэффициенты корреляции имеют уровень 0,2-0,3. Величина связи считается *средней* при коэффициенте корреляции, близком к 0,5. Если же коэффициент корреляции больше 0,7, то связь считается *значительной*, высокой.

3) Знаки (+) или (–) указывают на направление связи. Знак (+) означает, что связь между признаками прямая (положительная). При прямой связи с изменением одного признака другой – сопряжённый признак изменяется в том же направлении. Если коэффициент корреляции имеет знак (–), то это свидетельствует об обратной связи между признаками. Это значит, что при увеличении значений одного признака значения другого признака уменьшаются (и, наоборот, при уменьшении значений первого признака значения второго признака растут).

4) Коэффициент корреляции, r – это показатель, позволяющий определить величину и направление связи *прямолинейного* типа (или типа, близкого к прямолинейному). Для оценки *криволинейных* связей коэффициентом корреляции r пользоваться нельзя, так как он сильно недооценивает связь, а в ряде случаев даже не может ее уловить.

5) Коэффициент корреляции, взятый в квадрате (r^2), называется **детерминантой** и показывает *долю изменчивости признака* обусловленную влиянием изменчивости воздействующего фактора [5, *с.141*].

В общем виде формула коэффициента корреляции может быть представлена суммой произведений нормированных отклонении вариант каждого признака от своей средней арифметической, делённой на число наблюдений:

$$r = \frac{\sum t_x \cdot t_y}{n} \qquad (40)$$

где t_x и t_y – это нормированные отклонения каждой варианты по признаку x и признаку y. Подставляя формулу нормированных отклонений ($t = \frac{x_i - \bar{x}}{\sigma}$) t_x и t_y в формулу (40), получим следующую формулу:

$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \qquad (41)$$

Использование в этой формуле значений варьирующих признаков, выраженных через нормированное отклонение (т.е. в долях σ), позволяет вычислять связь между признаками, измеряемыми *разными* мерами (литры и проценты, килограммы и сантиметры и т.п.).

Помимо общей формулы (41) коэффициент корреляции может быть выражен и другими («рабочими») формулами. Рабочие формулы коэффициента корреляции применяют с учетом того, какую изучают выборку – большую или малую и какими значениями вариант (однозначными, многозначными или дробными) представлены анализируемые величины.

9.5.3. Коэффициент корреляции r, для малых выборок

Для малых выборок при малозначных показателях вариант удобно пользоваться следующими формулами [5, с.149]:

$$r = \frac{\sum x_i \cdot y_i - \dfrac{\sum x_i \cdot \sum y_i}{n}}{\sqrt{C_x \cdot C_y}}, \qquad (42)$$

где n – это число наблюдений, а C_x и C_y – дисперсии каждого признака, вычисляемые по формулам:

$$C_x = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \qquad (43)$$

$$C_y = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \qquad (43а)$$

Подставив значения C_x и C_y в формулу вычисления r, получим рабочую формулу:

$$r = \frac{\sum x_i \cdot y_i - \dfrac{\sum x_i \cdot \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}\right) \cdot \left(\sum y_i^2 - \dfrac{(\sum y_i)^2}{n}\right)}} \qquad (44)$$

Рассмотрим на примере как нужно пользоваться этой формулой при анализе корреляции признаков в малых выборках. Предположим нужно определить величину и направление связи между плодовитостью 10 свиноматок материнской группы и плодовитостью их дочерей.

Вычисление r для малой выборки сводится к несложным рассчётам. Представим в табличной форме показатели плодовитости каждой свиноматки, записывая эти данные *попарно* (мать-дочери), и вычислим те величины, которые представлены в формуле (44): $x \cdot y$, x^2, y^2 (табл. 35, столбцы *3*, *4* и *5*, соответственно) и их сумм (табл. 35, нижняя строка).

Таблица 35

Расчёт промежуточных величин для вычисления *r* между показателями в малой выборке

Плодовитость свиноматок-матерей, x_i	Плодовитость свиноматок - дочерей, y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	*2*	*3*	*4*	*5*
10	11	110	100	121
9	10	90	81	100
12	12	144	144	144
10	12	120	100	144
8	9	72	64	81
10	8	80	100	64
9	9	81	81	81
7	8	56	49	64
15	10	150	225	100
10	10	100	100	100
$\Sigma x_i = 100$	$\Sigma y_i = 99$	$\Sigma x_i \cdot y_i = 1003$	$\Sigma x_i^2 = 1044$	$\Sigma y_i^2 = 999$

Подставив полученные суммы в формулу (44), получим:

$$r = \frac{1003 - \dfrac{100 \cdot 99}{10}}{\sqrt{\left(1044 - \dfrac{100^2}{10}\right) \cdot \left(999 - \dfrac{99^2}{10}\right)}} = \frac{1003 - 990}{\sqrt{44 \cdot 18{,}9}} = \frac{13}{28{,}8} = +0{,}451$$

Таким образом, связь между плодовитостью дочерей и плодовитостью их матерей умеренная (средняя) и положительная (прямая). Последнее означает, что чем выше плодовитость матерей, тем выше плодовитость их дочерей.

Существуют и другие варианты рабочих формул для вычисления *r*, предназначенные для регрессионного анализа признаков, описываемых величинами малозначными (одно- и двухзначными) и многозначными (3 и более значащих цифры). Для знакомства с этими формулами следует обратиться к специальной литературе [5, *с.150-157*].

9.5.4. Коэффициент корреляции r для больших выборок

Если необходимо определить коэффициент корреляции двух признаков в большой выборке, то необходимо использовать рабочую формулу (45) и строить специальную **корреляционную решётку,** которая объединяет частоты *p* по обоим коррелирующим признакам (табл. 36).

$$r = \frac{\sum p \cdot a_x \cdot a_y - n \cdot b_x \cdot b_y}{n \cdot \sigma_x \cdot \sigma_y}, \tag{45}$$

где

p – частоты вариант каждого признака по клеткам решётки;

a_x и a_y – условные отклонения от нулевого класса *i*-тых вариант каждого признака;

n – объём выборки;

b_x и b_y – поправки для каждого ряда, вычисляемые по формулам (46):

$$b_x = \frac{\sum p_x \cdot a_x}{n} \tag{46}$$

$$b_y = \frac{\sum p_y \cdot a_y}{n} \tag{46a}$$

σ_x и σ_y – средние квадратические отклонения, выраженные в относительных величинах (т.е. без умножения квадратного корня на классовый интервал *k*):

$$\sigma_x = \sqrt{\frac{\sum p_x \cdot a_x^2}{n} - b_x^2} \tag{47}$$

$$\sigma_y = \sqrt{\frac{\sum p_y \cdot a_y^2}{n} - b_y^2} \tag{47a}$$

Рассмотрим пример вычисления *r* для большой выборки [4, с.117-122]. Пусть требуется определить – имеется ли связь между живым весом полновозрастных коров и их удоем за одну (третью) лактацию. Обрабатываются данные, полученные по 100 животным. Для каждой из коров имеются сведения о её живом весе и удое за лактацию. Для удобства обработки данные следует ранжировать по одному из признаков (т.е. распределить от минимальной величины до максимальной по мере возрастания его значений). В нашем примере данные были ранжированы по живой массе (табл. 36).

Алгоритм вычисления r для большой выборки следующий:

1) Составляем классы вариационного ряда отдельно для каждого признака.

Из имеющихся данных минимальный вес коровы равен 403 кг, а максимальный вес – 617 кг. Для удобства построения вариационного ряда расширим его лимиты до 400 кг и 620 кг. Размах варьирования признака составит 620–400 = 220 кг. Выборки объёмом n=100 следует делить на 8-12 классов. Разделим нашу выборку на 11 классов. В этом случае размер классового интервала первого признака будет равен

$$k_x = (x_{max} – x_{min}): 11=(620–400): 11= 20 \text{ кг.}$$

Максимальный удой за лактацию (5170 кг) установлен у коровы № 99, минимальный удой (3095 кг) – у коровы № 2. Для удобства построения вариационного ряда расширим его лимиты до 5200 кг и 3000 кг. Размах варьирования второго признака составит 5200–3000= 2200 кг. Размер классового интервала второго признака примем

$$k_y = (y_{max}–y_{min}): 11 = (5200–3000): 11= 200 \text{ кг.}$$

2) Строим корреляционную решетку (табл. 37), используя полученные величины классовых интервалов k_x и k_y. Классы по признаку x, «живая масса» помещаем в столбце 1 табл. 3. Классы по признаку у (удой) размещаем в строке A. Сведения о каждом животном разносим по клеткам решётки, учитывая величины каждого из двух признаков.

3) Для обоих признаков определяем частоты встречаемости вариант в каждом из классов (p_x и p_y), суммируя частоты по каждому из признаков. В результате образуются два вариационных ряда – по живому весу (x) и по удою (y).

Каждый из двух вариационных рядов обрабатываем методом «условной средней». Для этого по каждому признаку выделяем классы с условными средними A_x и A_y, которые удобнее брать исходя из центрального расположения класса и наибольшего числа частот. В данном примере классы с условной средней A следующие: по живому весу (A_x) класс с границами 500-519 кг, и по удою (A_y) класс 4000-4199 кг. Эти классы отграничим сверху и снизу чертами, отчего в решетке образуется фигура «креста». Эти два класса служат нулевыми классами.

Таблица 36

Данные по живой массе коров и их удою за третью лактацию

№ коровы	живая масса	удой	№ коровы	живая масса	удой
1	403	3210	*26*	492	3922
2	425	3095	*27*	492	4010
3	431	3215	*28*	494	4059
4	434	3450	*29*	495	4094
5	437	3560	*30*	495	4120
6	442	3286	*31*	496	4195
7	445	3465	*32*	497	4250
8	448	3620	*33*	497	4386
9	451	3687	*34*	499	4520
10	452	3750	*35*	502	3805
11	454	3846	*36*	502	3880
12	457	3963	*37*	505	3920
13	458	4351	*38*	506	3953
14	463	3492	*39*	508	3990
15	469	3667	*40*	509	4010
16	471	3759	*41*	509	4083
17	474	4250	*42*	510	4130
18	478	4382	*43*	511	4159
19	482	3602	*44*	511	4133
20	484	3644	*45*	513	4198
21	485	3797	*46*	515	4420
22	489	3684	*47*	515	4494
23	490	3795	*48*	518	4508
24	490	3856	*49*	518	4573
25	491	3990	*50*	519	4591

Окончание таблицы см. на следующей странице

4) Определяем условные отклонения от нулевых классов a_x и a_y (a = $(x_{mid} – A):k$ где x_{mid} – среднее значение каждого класса, определяемое по формуле $x_{mid} = x_{min} + 0{,}5{\cdot}k$). В сторону уменьшения признака от

Таблица 36 (окончание)
(данные ранжированы по первому признаку – живой массе коров)

№ коровы	живая масса	удой	№ коровы	живая масса	удой
51	520	4020	*76*	540	4000
52	520	4038	*77*	546	4150
53	521	4047	*78*	549	4111
54	521	4056	*79*	553	1463
55	523	4076	*80*	555	4196
56	524	4069	*81*	555	4265
57	524	4090	*82*	558	4608
58	524	4105	*83*	558	4710
59	525	4120	*84*	559	4786
60	525	4140	*85*	560	4208
61	526	4160	*86*	563	4300
62	528	4187	*87*	567	4366
63	528	4150	*88*	570	4400
64	530	4190	*89*	570	4490
65	530	4197	*90*	575	4550
66	533	4210	*91*	578	4593
67	533	4285	*92*	580	4611
68	535	4330	*93*	584	4300
69	535	4367	*94*	585	4700
70	535	4395	*95*	590	4810
71	535	4421	*96*	593	4860
72	536	4477	*97*	596	4946
73	537	4502	*98*	596	5061
74	537	4563	*99*	599	5170
75	539	4580	*100*	617	4905

нулевого класса отклонения получаются отрицательными, а в сторону увеличения признака – положительными.

5) Выполняем необходимые вычисления для определения величин $p_x{\cdot}a_x$, $p_x{\cdot}a^2_x$ (*графы 15* и *16*), $p_y{\cdot}a_y$, $p_y{\cdot}a^2_y$ (*строки R* и *S*).

Таблица 37

Корреляционная решётка для расчёта связи между живой массой (x) и

	1	2	3	4	5	6	7	8
A	Классы по живой массе, кг (признак x)	Классы по удою, кг (признак y)						
		3000-3199	**3200-3399**	**3400-3599**	**3600-3799**	**3800-3999**	**4000-4199** A_y	**4200-4399**
B		*I квадрант*						
C	**400-419**	–	1	–	–	–	–	–
D	**420-439**	1	1	2	–	–	–	–
E	**440-459**	–	1	1	3	2	–	1
F	**460-479**	–	–	1	2	–	–	2
G	**480-499**	–	–	–	5	3	5	2
H	**500-519 = =A_x**	–	–	–	–	5	6	5
I	**520-539**	–	–	–	–	–	15	5
J	**540-559**	–	–	–	–	–	5	1
K	**560-579**	–	–	–	–	–	–	3
L	**580-599**	–	–	–	–	–	–	1
M	**600-619**	–	–	–	–	–	–	–
N		*III квадрант*						
P	p_y	**1**	**3**	**4**	**10**	**10**	**31**	**20**
Q	a_y	**–5**	**–4**	**–3**	**–2**	**–1**	**0**	**1**
R	$p_y \cdot a_y$	**–5**	**–12**	**–12**	**–20**	**–10**	**0**	**20**
S	$p_y \cdot a^2_y$	**25**	**48**	**36**	**40**	**10**	**0**	**20**
	1	2	3	4	5	6	7	8

Окончание таблицы см. на следующей странице

Таблица 37 (окончание)

удоем за третью лактацию (y) для 100 коров

9	10	11	12	13	14	15	16
4400--4599	**4600--4799**	**4800-4999**	**5000-5199**	p_x	a_x	$p_x \cdot a_x$	$p_x \cdot a^2_x$
II квадрант							
–	–	–	–	**1**	**–5**	**–5**	**25**
–	–	–	–	**4**	**–4**	**–16**	**64**
–	–	–	–	**8**	**–3**	**–24**	**72**
–	–	–	–	**5**	**–2**	**–10**	**20**
1	–	–	–	**16**	**–1**	**–16**	**16**
–	–	–	–	**16**	**0**	**0**	**0**
5	–	–	–	**25**	**1**	**25**	**25**
–	3	–	–	**9**	**2**	**18**	**36**
4	–	–	–	**7**	**3**	**21**	**63**
–	2	3	2	**8**	**4**	**32**	**128**
–	–	1	–	**1**	**5**	**5**	**25**
				n*= *100	**–**	$\Sigma p_x \cdot a_x$= = +30	$\Sigma p_x \cdot a^2_x$= =474
10	**5**	**4**	**2**				
2	**3**	**4**	**5**	**–**			
20	**15**	**16**	**10**	$\Sigma\ p_y \cdot a_y = +22$			
40	**45**	**64**	**50**	$\Sigma\ p_y \cdot a^2_y = 378$			
9	*10*	*11*	*12*	*13*	*14*	*15*	*16*

6) Определяем суммы этих величин $\Sigma p_x \cdot a_x$ и $\Sigma p_x \cdot a^2_x$ (*ячейки N15* и *N16*), а также $p_y \cdot a_y$, $p_y \cdot a^2_y$ (*ячейки R13* и *S13*).

7) В формуле (45) для вычисления коэффициента корреляции требуется проставить b_x, и b_y (формулы (46) и (46а)). Эти значения определяем обычным приемом обработки вариационного ряда. В результате этого получаем

$$b_x = 30 : 100 = 0{,}30 \text{ и } b_x^2 = 0{,}0900$$

$$b_y = 22 : 100 = 0{,}22 \text{ и } b_y^2 = 0{,}0484.$$

8) Вычисляем σ_x и σ_y. При этом следует помнить, что в формулах (47) и (47а) средние квадратические отклонения σ_x и σ_y, выражаются в относительных величинах, т.е. без умножения квадратного корня на классовый промежуток k.

$$\sigma_x = \sqrt{\frac{\sum p_x \cdot a_x^2}{n} - b_x^2} = \sqrt{\frac{474}{100} - 0{,}09} = \sqrt{4{,}65} = 2{,}156 \approx 2$$

$$\sigma_y = \sqrt{\frac{\sum p_y \cdot a_y^2}{n} - b_y^2} = \sqrt{\frac{378}{100} - 0{,}0484} = \sqrt{3{,}731} = 1{,}931 \approx 2$$

9) Вычисляем выражение $\Sigma p \cdot a_x \cdot a_y$, необходимое для вычисления коэффициента корреляции. Эта сумма получается путем умножения каждого значения частот p по клеткам решетки на условные отклонения a_x и a_y. При этом действие умножения осуществляется только для тех частот, которые расположены в клетках *за пределами* нулевых классов (за пределами «креста»), т.е. в клетках I, II, III, IV квадрантов решетки (см. табл. 37). Следует отметить, что I и IV квадранты имеют всегда положительное значение $\Sigma p \cdot a_x \cdot a_y$, а II и III квадранты – всегда отрицательное значение $\Sigma p \cdot a_x \cdot a_y$. Для упрощения расчёта этих сумм удобнее построчно выписать все необходимые произведения и затем их сложить. Выполним построчное умножение $p \cdot a_x \cdot a_y$ по каждому квадранту и найдём суммы этих произведений (табл. 38):

Таблица 38

Вспомогательная таблица для расчёта $\Sigma p{\cdot}a_x{\cdot}a_y$

I квадрант

	$p{\cdot}a_x{\cdot}a_y$
1 (*C*) строка	1· –5· –4= +20
2 (*D*) строка	1· –4· –5= +20
–"–	1· –4· –4= +16
–"–	2· –4· –3= +24
3 (*E*) строка	1· –3· –4= +12
–"–	1· –3· –3= +9
–"–	3· –3· –2= +18
–"–	2· –3· –1= +6
4 (*F*) строка	1· –2· –3= +6
–"–	2· –2· –2= +8
5 (*G*) строка	5· –1· –2= +10
–"–	3· –1· –1= +3
	$\Sigma p{\cdot}a_x{\cdot}a_y$ = **+152**

II квадрант

	$p{\cdot}a_x{\cdot}a_y$
3 (*E*) строка	1· –3· 1= –3
4 (*F*) строка	2· –2· 1= –4
5 (*G*) строка	2· –1· 1= –2
–"–	1· –1· 2= –2
	$\Sigma p{\cdot}a_x{\cdot}a_y$ = **–11**

III квадрант

	$p{\cdot}a_x{\cdot}a_y$
	0
	$\Sigma p{\cdot}a_x{\cdot}a_y$ = **0**

IV квадрант

	$p{\cdot}a_x{\cdot}a_y$
7 (*I*) строка	5· 1· 1=5
–"–	5· 1· 2=10
8 (*J*) строка	1· 2· 1=2
–"–	3· 2· 3=18
9 (*K*) строка	3· 3· 1=9
–"–	4· 3· 2=24
10 (*L*) строка	1· 4· 1=4
–"–	2· 4· 3=24
–"–	3· 4· 4=48
–"–	2· 4· 5=40
11 (*M*) строка	1· 5· 4=20
	$\Sigma p{\cdot}a_x{\cdot}a_y$ = **+204**

Суммарное значение по всем 4-м квадрантам:

$$\Sigma p{\cdot}a_x{\cdot}a_y = +152 + (-11) + 0 + 204 = +345.$$

10) Подставляем полученные значения $\Sigma p{\cdot}a_x{\cdot}a_y$ =+345; n =100; b_x=0,3; b_y=0,22; σ_x=2,156; σ_y=1,931 в формулу (45) для вычисления r:

$$r = \frac{\sum p\cdot a_x\cdot a_y - n\cdot b_x\cdot b_y}{n\cdot\sigma_x\cdot\sigma_y} = \frac{345-100\cdot 0{,}3\cdot 0{,}22}{100\cdot 2{,}156\cdot 1{,}931} = \frac{338{,}4}{416{,}3} = +0{,}812$$

Таким образом, связь между величиной живого веса и удоем за лактацию значительная и положительная (т.е. чем больше живой вес, тем в среднем выше удой коров).

9.5.5. Ошибка коэффициента корреляции

Так как коэффициент корреляции вычислен не по генеральной, а по выборочной совокупности, он имеет ошибку выборочности [2, *с.194*]:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}, \text{ если } n<100 \qquad (48)$$

$$m_r = \frac{1-r^2}{\sqrt{n-1}}, \text{ если } n\geq 100 \qquad (48а)$$

В приведённом выше примере (живая масса коров – удой) ошибка коэффициента корреляции составит:

$$m_r = \frac{1-r^2}{\sqrt{n-1}} = \frac{1-0{,}812^2}{\sqrt{100-1}} = \frac{0{,}34}{9{,}95} = 0{,}03$$

9.5.6. Достоверность коэффициента корреляции

Когда известна ошибка коэффициента корреляции, можно определить степень достоверности *r*:

$$t_r = \frac{r}{m_r} \qquad (49)$$

При этом исходит из нулевой гипотезы, т.е. предполагают, что в генеральной совокупности связь между изучаемыми признаками отсутствует. Нулевая гипотеза может быть отвергнута, и значение *r* можно считать достоверным только в том случае, когда вычисленное значение t_r**,** будет равно табличному значению или превысит его при выбранной уровне доверительной вероятности (0,95, 0,99 или 0,999).

В приведённом выше примере

$$t_r = \frac{r}{m_r} = \frac{0{,}812}{0{,}03} = 27{,}1$$

С учетом числа степеней свободы ν = n–1 = 100–1 = 99 находим табличное значение t_r [2, с.194]. (В учебном пособии [1] указывают, что в этом случае *ν* = *n*–*2*). Сравнение вычисленного значения с табличным (табл. 15) доказывают высокую статистическую достоверность установленного коэффициента корреляции.

9.5.7. Определение величины и направления связи между признаками

По расположению частот в клетках корреляционной решетки можно судить о направлении связи, её типе и уровне. Если числа частот в основном группируются близко к диагонали, идущей из левого верхнего угла в правый нижний угол, то коэффициент корреляции будет иметь знак плюс (рис. 13, *А*). Если же частоты группируются вдоль другой диагонали, идущей из левого нижнего угла в правый верхний угол решетки, то связь отрицательная и r будет иметь знак (–) (рис. 13, *Б*).

А		Признак ***x***				
		2	3	4	5	6
Признак ***y***	3	×	×			
	4	×	×	×		
	5		×	×		
	6			×	×	
	7				×	
	8					×

Б		Признак ***x***				
		2	3	4	5	6
Признак ***y***	3					×
	4			×	×	
	5		×	×	×	
	6		×	×		
	7	×	×	×		
	8	×				

В		Признак ***x***				
		2	3	4	5	6
Признак ***y***	3		×	×	×	×
	4	×	×	×	×	
	5		×	×	×	×
	6	×	×	×	×	
	7	×	×	×	×	
	8	×	×			

Г		Признак ***x***				
		2	3	4	5	6
Признак ***y***	3	×	×			
	4		×	×	×	
	5		×	×	×	
	6			×	×	×
	7	×	×	×		
	8	×	×			

Рис.13. Распределение частот по двум коррелирующим признакам (x и y) при разных направлениях и величинах связи: А – связь положительная и большая; Б – связь отрицательная и большая; В – связь отрицательная и небольшая; Г – связь криволинейная.

Чем ближе группируются частоты к диагоналям (рис. 13, *А*, *Б*), тем больше значение *r*. При беспорядочном распределении частот в клетках решетки связь признаков будет незначительная (рис. 13, *В*).

Если частоты располагаются в решетке дугообразно или образуют фигуру подобную полумесяцу, то связь имеет криволинейный тип (рис. 13, *Г*) и вычислять r нецелесообразно.

9.5.8. Коэффициент регрессии, *b*

При анализе корреляционных связей большое значение имеет коэффициент регрессии b и регрессионный анализ. Если коэффициент корреляции показывает величину связи в относительных величинах (долях единицы), то с помощью регрессионного анализа, при котором вычисляется коэффициент регрессии, получают величину связи в именованных величинах.

Итак, **коэффициент регрессии** – это именованный статистический показатель, который свидетельствует насколько изменяется в среднем признак x если коррелирующий с ним признак y изменится на определённую величину.

Регрессионный анализ дает больше информации, чем коэффициент корреляции. Он позволяет установить связь между двумя признаками, а именно x по y и y по x, что достигается вычислением двух коэффициентов регрессии: b_{xy} и b_{yx}. По наличию знака минус или плюс у коэффициента регрессии определяют направление связи.

Регрессионную связь можно выражать с помощью уравнения, в котором при известной величине коэффициента регрессии и определенном значении одного из признаков (аргумента) устанавливают величину другого признака (функции).

Регрессионная связь может быть выражена графически в виде эмпирической и теоретической линии регрессии.

По характеру графического изображения линии регрессии можно судить о том, имеет ли место связь прямолинейного или близкого к ней типа, или связь носит криволинейный характер и в этом случае коэффициентом корреляции пользоваться нельзя.

Коэффициенты регрессии имеют большое значение в генетических и селекционных исследованиях. Наиболее широко их используют при вычислении так называемого коэффициента наследуемости (h^2) для количественных признаков. В этом случае определяют коэффициент регрессии по фенотипическим показателям признака у родственных групп животных. Например, определяют регрессию удоя дочерей по показателям удоя их матерей. Удвоенный коэффициент регрессии будет выражать **коэффициент наследуемости** данного признака: $h^2 = 2 \cdot b_{xy}$.

9.5.8.1. Коэффициент прямолинейной регрессии, *b*

Коэффициент прямолинейной регрессии *b*, представляет собой произведение коэффициента корреляции *r* на отношение стандартных отклонений σ_x и σ_y, вычисленных для каждого признака.

В общем виде коэффициенты регрессии для выборки выражаются следующим образом:

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} \qquad (50)$$

и

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} \qquad (51)$$

По формуле (11) определяют величину значения *x* при изменении, *y* на единицу. По формуле (12), наоборот, находят величину *y* при изменении *x* на единицу. Величины σ_x и σ_y в формуле вычисления *b* берут в именованном выражении. Между *r* и *b* имеется определенная зависимость. Она следует из формул регрессии, а именно – произведение регрессии *x* по *y* на регрессию *y* по *x* равно квадрату коэффициента корреляции: $b_{xy} \cdot b_{yx} = r^2$, откуда $r = \sqrt{b_{xy} \cdot b_{yx}}$. Эта формула указывает на то, что коэффициент корреляции представляет собой среднюю геометрическую из обоих значений коэффициентов регрессии данной выборки.

Существуют и другие рабочие формулы для вычисления коэффициента регрессии.

9.5.8.2. Вычисление коэффициента регрессии b для больших выборок

Формулы (50) и (51) удобно использовать для вычисления регрессии при большом числе наблюдений. Для примера вычислений используем фактические данные, которые мы анализировали, вычисляя коэффициент корреляции между живой массой и удоем в большой выборке коров (табл. 36 и 37, с. 133-136). Статистическая обработка корреляционной решетки для этого примера позволила вычислить коэффициент корреляции r, который оказался равен +0,812. При вычислении этого коэффициента были определены и средние квадратические отклонения:

$$\sigma_x = \sqrt{\frac{\sum p_x \cdot a_x^2}{n} - b_x^2} = \sqrt{\frac{474}{100} - 0{,}09} = \sqrt{4{,}65} = 2{,}156$$

$$\sigma_y = \sqrt{\frac{\sum p_y \cdot a_y^2}{n} - b_y^2} = \sqrt{\frac{378}{100} - 0{,}0484} = \sqrt{3{,}731} = 1{,}931$$

Однако, они были вычислены в относительных величинах, т.е. без умножения квадратного корня на классовый промежуток k. Чтобы выразить σ_x и σ_y в абсолютных величинах их нужно умножить на соответствующие величины классных интервалов (k_x=20 кг живой массы и k_y=200 кг удоя):

$$\sigma_x = 2{,}156 \cdot 20 \text{ кг} = 43{,}12 \text{ кг},$$

$$\sigma_y = 1{,}931 \cdot 200 \text{ кг} = 386{,}2 \text{ кг}.$$

Находим по данным значениям σ_x, σ_y и r величину регрессии b.

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = 0{,}812 \cdot \frac{43{,}12}{386{,}2} = 0{,}09 \text{ кг.}$$

Таким образом, в изученной выборке увеличение удоя (y) на 1 кг происходит при увеличении живой массы коров в среднем на 0,09 кг.

Вычисляем второе значение регрессии:

$$b_{yx} = r \cdot \frac{\sigma_x}{\sigma_y} = 0{,}812 \cdot \frac{386{,}2}{43{,}12} = 7{,}27 \text{ кг.}$$

Следовательно, с увеличением живой массы на 1 кг удой коров в среднем увеличивается на 7,27 кг.

9.5.8.3. Вычисление коэффициента регрессии *b* для малых выборок

Для анализа коэффициента регрессии в *малых* выборках, особенно если значения *x* и *y* выражаются малозначными числами, удобнее пользоваться следующими рабочими формулами:

$$b_{xy} = \frac{\sum x \cdot y - \frac{\sum x \cdot \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}} \qquad (52)$$

и

$$b_{yx} = \frac{\sum x \cdot y - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} \qquad (53)$$

Рассмотрим пример вычисления коэффициента регрессии для *малой* выборки с использованием формул (52) и (53).

Пусть требуется определить коэффициент регрессии между жирномолочностью дочерей и жирномолочностью их матерей (табл. 39).

Таблица 39

Вычисление коэффициента регрессии жирномолочности дочерей (*x*) по жирномолочности их матерей (*y*)

Номер пар	% жира в молоке дочерей, (*x*)	% жира в молоке матерей, (*y*)	$x{\cdot}y$	x^2	y^2
1	3,5	3,4	11,90	12,25	11,56
2	3,8	3,8	14,44	14,44	14,44
3	3,9	4,0	15,60	15,21	16,00
4	3,3	3,0	9,90	10,89	9,00
5	3,7	3,7	13,69	13,69	13,69
6	3,4	3,0	10,20	11,56	9,00
7	3,6	3,5	12,60	12,96	12,25
8	4,0	3,8	15,20	16,00	14,44
9	3,4	3,1	10,54	11,56	9,61
10	3,1	3,0	9,30	9,61	9,00
	Σx=35,7; $\bar{x}$ =3,57%	Σy=34,3; $\bar{y}$ =3,43%	$\Sigma x{\cdot}y$= =123,37	Σx^2 = =128,17	Σy^2= =118,99

Для получения исходных значений в таблице 39 оформляют графы со следующими показателями: x, y, $x{\cdot}y$, x^2 и y^2. Суммирование по этим столбцам дает необходимые для формулы регрессии величины.

Вычисляем по этим данным величину коэффициента регрессии x по y:

$$b_{xy} = \frac{\sum x \cdot y - \frac{\sum x \cdot \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}} = \frac{123{,}37 - \frac{35{,}7 \cdot 34{,}3}{10}}{118{,}99 - \frac{34{,}3^2}{10}} = \frac{0{,}919}{1{,}341} = 0{,}685\%$$

Полученный коэффициент регрессии указывает на то, что с увеличением жирномолочности коров-*матерей* на 1 % содержание жира в молоке их *дочерей* в среднем увеличится на 0,685%.

Если вычислять регрессию *матерей* (y) по показателям *дочерей* (x), то получаем следующую величину:

$$b_{yx} = \frac{\sum x \cdot y - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{123{,}37 - \frac{35{,}7 \cdot 34{,}3}{10}}{128{,}17 - \frac{35{,}7^2}{10}} = \frac{0{,}919}{0{,}721} = 1{,}274\%$$

т.е. для увеличения в среднем содержания жира в молоке коров-*дочерей* на 1% у их *матерей* это увеличение в среднем должно составлять 1,274%.

9.5.9. Коэффициент корреляции между альтернативными признаками, r_a

Корреляционная связь может быть определена не только между количественными признаками, но и между признаками качественными. Предположим, что перед вами стоит задача установить, имеется ли связь между полом помесных кур и пигментацией их ног. В опыте получено 50 петушков и 50 курочек. Из них 5 петушков имели чёрную окраску ног и 45 – белую, 49 курочек имели пигментированные ноги и 1 курочка – оказалась с белыми ногами.

При определении корреляции между альтернативными признаками строят четырёхпольную корреляционную решётку, в которой два класса будут по одному признаку (x), и два – по другому признаку (y). В каждое из полей решетки (a, b, c и d) вносят частоты вариант каждого признака (табл. 40):

Таблица 40

Четырёхпольная корреляционная решётка для определения коэффициента корреляции r_a между двумя качественными признаками

Первый признак (x)	Второй признак (y)		Сумма
	Вариант y_1	Вариант y_2	
Вариант x_1	$p_1=a$	$p_2=b$	$p_1+p_2=a+b$
Вариант x_2	$p_3=c$	$p_4=d$	$p_3+p_4=c+d$
Сумма	$p_1+p_3=a+c$	$p_2+p_4=b+d$	$p_1+p_2+p_3+p_4=$ $=a+b+c+d$

Величину связи между двумя качественными признаками определяют с помощью коэффициента корреляции r_a (коэффициента ассоциации). Для более точной оценки r_a в формулу вводят поправку Иейтса на непрерывность, равную половине объёма выборки, вычитаемой из числителя [3, *с.177*]:

$$r_a = \frac{|a \cdot d - b \cdot c| - 0{,}5 \cdot n}{\sqrt{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)}} \tag{54}$$

или

$$r_a = \frac{|p_1 \cdot p_4 - p_2 \cdot p_3| - 0{,}5 \cdot n}{\sqrt{(p_1+p_2)\cdot(p_3+p_4)\cdot(p_1+p_3)\cdot(p_2+p_4)}} \tag{54а}$$

Для вычисления коэффициента корреляции построим корреляционную решётку для имеющихся у нас данных (табл. 41):

Таблица 41

Четырёхпольная корреляционная решётка для расчёта коэффициента корреляции r_a между полом особи и типом пигментации ног у кур.

Пол птицы (x)	Тип пигментации ног (y)		Сумма
	белые	чёрные	
Самцы	45	5	50
Самки	1	49	50
Сумма	46	54	n=100

Подставим данные, из корреляционной решетки в формулу (54) и вычислим r_a.

$$r_a = \frac{|45 \cdot 49 - 5 \cdot 1| - 0{,}5 \cdot 100}{\sqrt{50 \cdot 50 \cdot 46 \cdot 54}} = \frac{2205 - 5 - 50}{\sqrt{6210000}} = \frac{2150}{2492} = 0{,}863.$$

Полученный коэффициент корреляции r_a, указывает на большую связь между типом пигментации и полом птиц.

9.5.10. Ранговый коэффициент связи по Спирмену, r_s

При ветеринарном или зоотехническом анализе животных могут встретиться признаки, которые нельзя измерить ни точно, ни грубо. Поэтому их вариации обозначают порядковым местом (рангом). Например, можно определить ранги быков-производителей по удою (x) их дочерей и эти данные сопоставить с рангами их дочерей по показателям живой массы (y).

Ранговый коэффициент корреляции по Спирмену выражают следующей формулой:

$$r_S = 1 - \frac{6 \cdot \sum (x - y)^2}{n \cdot (n^2 - 1)}, \qquad (55)$$

где x и y – ранги по каждому признаку; n – число членов в совокупности.

Из формулы (55) видно, что показателем корреляции служит сумма квадратов отклонений между парными рангами обоих признаков. Для её вычисления члены выборки записывают по порядку, по величине ранга одного из признаков, от максимального к минимальному уровню ранга. При этом ранги будут такие: 1, 2, 3, ... n.

Ранг, занимаемый членом совокупности по второму признаку, выписывают для каждого члена ряда с учетом фактического уровня второго признака. Если фактическая абсолютная величина признака у нескольких членов выборки будет одинаковая, то их нумеруют подряд и берут среднее из этих рангов. Например, для членов совокупности с одинаковым абсолютным значением признака, получивших порядковый ранг 3 и 4, будет записана для каждого члена величина среднего ранга, т.е. 3,5.

Рассмотрим на конкретном примере вычисление рангового коэффициента корреляции r_s [5, *с.142*]. Допустим, нужно определить ранговый коэффициент корреляции r_s между уровнем жирности молока и содержанием белка в молоке дочерей двух проверяемых быков-производителей. Затем на основании величин $\bar{x}$ и r_S, вычисленных для каждого производителя, сделать выводы о селекционной ценности каждого быка (табл. 42).

Таблица 42

Ранговое распределение коров-дочерей двух быков по показателям жирности молока и содержанию белка в молоке для определения величины связи r_s

№ коровы	Жир, (x)		Белок, (y)		Разность рангов (x–y)	Квадрат разности рангов $(x–y)^2$
	%	Ранг, (x)	%	Ранг, (y)		
Дочери быка Ворчуна (1)						
1	3,5	1	3,2	2,5	–1,5	2,25
2	3,4	2	3,0	6,5	–4,5	20,25
3	3,3	3,5	3,3	1	+2,5	6,25
4	3,3	3,5	3,0	6,5	–3,0	9,00
5	3,2	5,5	2,9	8,5	–3,0	9,00
6	3,2	5,5	3,1	4,5	+1,0	1,00
7	3,1	7	3,2	2,5	+4,5	20,25
8	3,0	8,5	3,1	4,5	+4,0	16,00
9	3,0	8,5	2,9	8,5	0	0,00
10	2,9	10	2,8	10	0	0,00
Сумма	**31,9**	–	**30,5**	–	–	**$\Sigma(x–y)^2$ =84,0**

Таблица 42 (окончание)

№ коровы	Жир, (x)		Белок, (y)		Разность рангов (x–y)	Квадрат разности рангов $(x–y)^2$
	%	Ранг, (x)	%	Ранг, (y)		
Дочери быка Валдая (2)						
1	3,8	1	3,5	3	–2	4
2	3,7	2,5	3,6	1,5	+1	1
3	3,7	2,5	3,6	1,5	+1	1
4	3,6	4,0	3,4	4,5	–0,5	0,25
5	3,5	5,5	3,4	4,5	+1	1
6	3,5	5,5	3,3	6,5	–1	1
7	3,3	7,5	3,3	6,5	+1	1
8	3,3	7,5	3,2	8	–0,5	0,25
9	3,2	9	3,1	9	0	0
10	3,1	10	3,0	10	0	0
Сумма	**34,8**	–	**33,4**	–	–	**$\Sigma (x–y)^2$=9,5**

При обработке таких данных следует помнить следующие правила:

1) если фактическая абсолютная величина признака у нескольких членов выборки будет одинаковая, то их нумеруют подряд и берут среднее из этих рангов;

2) ранг, занимаемый членом совокупности по второму признаку, выписывают для каждого члена ряда с учетом фактического уровня второго признака.

Например, члены совокупности с одинаковым абсолютным значением признака, записанные с порядковым рангом 3 и 4, получат среднее значение этих двух рангов, т.е. 3,5.

Используя расчётные данные таблицы 8, вычисляем величины $\bar{x}$ и r_s:

Расчёт показателей для дочерей быка Ворчуна:

$$\Sigma(x-y)^2 = 84{,}0; \quad r_s = 1 - \frac{6 \cdot \sum (x-y)^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 84}{10 \cdot (10^2 - 1)} = +0{,}491$$

$$\bar{x}_x = \frac{\sum x_1}{n_1} = \frac{31{,}9}{10} = 3{,}19\%$$

$$\bar{x}_y = \frac{\sum y_1}{n_1} = \frac{30{,}5}{10} = 3{,}05\%$$

Расчёт показателей для дочерей быка Валдая:

$$\Sigma(x-y)^2 = 9{,}5; \quad r_s = 1 - \frac{6 \cdot \sum (x-y)^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 9{,}5}{10 \cdot (10^2 - 1)} = +0{,}942$$

$$\bar{x}_x = \frac{\sum x_1}{n_1} = \frac{34{,}8}{10} = 3{,}48\%$$

$$\bar{x}_y = \frac{\sum y_1}{n_1} = \frac{33{,}4}{10} = 3{,}34\%$$

Из результатов обработки данных табл. 42 видно, что связь жирности молока и содержания белка в нём у дочерей двух быков различная. Она оказалась очень высокой у дочерей быка Валдая (r_s=+0,942) и умеренной у потомства Ворчуна (r_s=+0,491). Сопоставив средние арифметические белковости молока, средние арифметические жирности молока и коэффициенты корреляции двух анализируемых групп коров, можно сделать следующие выводы:

1) селекционный эффект на одновременное повышение белка и жира в молоке дочерей наиболее высок у потомков быка Валдая;

2) высокий коэффициент корреляция (r_s=+0,942) свидетельствует, что в потомстве быка Валдая косвенная селекция на белок в молоке по показателям жирности молока проявляется сильнее;

3) дочери быка Ворчуна по содержанию жира и содержанию белка в молоке, а также по степени связи между этими признаками представляют меньшую селекционную ценность.

Для анализа причин высокой или низкой связи между жирностью и белковостью молока у потомства разных отцов применяют специальные методы генетического анализа. Генетический анализ учитывает а) качество матерей, б) долю влияния внешних условий на изменчивость уровня белка и жира в молоке, в) различия в условиях кормления дочерей каждого производителя, а также другие факторы, влияние которые могло сказаться на показателях.

Из разобранного выше примера видно, что техника вычисления рангового коэффициента корреляции проста и для малых выборок очень удобна, так как позволяет рассчитывать коэффициент связи при неизвестном типе распределения членов выборки.

9.5.11. Полихорический показатель связи, ρ

В зоотехнической практике довольно часто приходится проводить ориентировочную оценку связи признаков, не прибегая к их точному измерению. Например, на глаз оценивают густоту сперматозоидов в эякуляте (густой, средний, разбавленный), степень упитанности, степень линьки, степень поражения кожи заболеванием и т.п. В тех случаях, когда оба коррелирующих признака определяются грубо, на глаз, пользуются полихорическим показателем связи ρ:

$$\rho = \frac{\alpha - 1}{\sqrt{(l_1 - 1)\cdot(l_2 - 1)}}, \qquad (56)$$

Значение $\boldsymbol{\alpha}$ вычисляют по формулам (57, 58):

$$\alpha = \sum \left[\frac{\sum (p_{1,2}^2 : p_2)}{p_1} \right] - \frac{(l_1 - 1) \cdot (l_2 - 1)}{n} \quad (57)$$

или

$$\alpha = \sum \left[\frac{\sum (p_{1,2}^2 : p_1)}{p_2} \right] - \frac{(l_1 - 1) \cdot (l_2 - 1)}{n} \quad (58)$$

где:

– l_1 и l_2 – число классов по каждому признаку;

– $p_{1,2}$ – частоты в клетках корреляционной решетки;

– p_1 и p_2, – частоты вариационного ряда каждого из признаков;

– n – число членов в выборке.

Рассмотрим на примере использование полихрического показателя связи ρ [4, *с.124-127*]. Определим, имеется ли связь между формой завитка у каракульских ягнят и типом их конституции. Данные получены для выборки, включающей 100 голов молодняка.

Составим корреляционную решетку, в которой классы по признаку «тип конституции» (1) расположим в первом столбце вертикально, а классы по признаку формы завитка (2) – горизонтально вверху, Распределим данные по обоим признакам каждого ягненка в полученной корреляционной решётке (табл. 43).

Для обработки решетки производят следующие расчеты:

1) в каждой клетке корреляционной решетки каждую частоту $p_{1,2}$ возводят в квадрат, получают $p_{1,2}^2$;

2) делят $p_{1,2}^2$ на значение p_2;

3) полученное в каждой клетке значение $p_{1,2}^2 : p_2$ суммируют по каждой строке решетки, получая $\sum (p_{1,2}^2 : p_2)$;

4) полученное значение $\sum (p_{1,2}^2 : p_2)$ по каждой строчке делят на соответствующие значения p_1, получая значения $\left[\frac{\sum (p_{1,2}^2 : p_2)}{p_1} \right]$, равные 0,37, 0,59 и 0,50;

Таблица 43

Корреляционная решетка для определения полихорического показателя связи между типом конституции (1) и формой завитка (2) у ягнят

Тип конституции (1)	Форма завитка (2)				p_1	$\sum(p_{1,2}^2 : p_2)$	$\left[\dfrac{\sum(p_{1,2}^2 : p_2)}{p_1}\right]$
	валек	боб	гривка	штопор			
Нежный	**5** $5^2=25$ $\frac{25}{22}=1{,}14$	**20** $20^2=400$ $\frac{400}{45}=8{,}89$	**5** $5^2=25$ $\frac{25}{25}=1$	–	**30**	1,14+8,89+ +1=11,03	11,03:30=0,37
Рыхлый	**2** $2^2=4$ $\frac{4}{22}=0{,}18$	**5** $5^2=25$ $\frac{25}{45}=0{,}56$	**15** $15^2=225$ $\frac{225}{25}=9$	**8** $8^2=64$ $\frac{64}{8}=8$	**30**	0,18+0,56+ +9+8=17,74	17,74:30=0,59
Крепкий	**15** $15^2=225$ $\frac{225}{22}=10{,}23$	**20** $20^2=400$ $\frac{400}{45}=8{,}89$	**5** $5^2=25$ $\frac{25}{25}=1$	–	**40**	10,23+8,89+ +1=20,12	20,12:40=0,50
p_2	**22**	**45**	**25**	**8**	n=**100**	–	Σ= =0,37+0,59+0,50= =1,46

5) полученные значения 0,37, 0,59 и 0,50 суммируют:

$$\sum\left[\frac{\sum(p_{1,2}^2 : p_2)}{p_1}\right]=1{,}46$$

6) рассчитанную величину 1,46 используют для вычисления ***α***, которая входит в формулу полихорического коэффициента. Поскольку в нашем примере l_1=3 и l_2=4, то

$$\alpha=\sum\left[\frac{\sum(p_{1,2}^2 : p_2)}{p_1}\right]-\frac{(l_1-1)\cdot(l_2-1)}{n}$$

$$\boldsymbol{\alpha}=1{,}46-\frac{(3-1)\cdot(4-1)}{100}=1{,}46-0{,}06=1{,}40.$$

Подставим полученное значение ***α*** в формулу (56) и вычислим ***ρ***:

$$\rho=\frac{\alpha-1}{\sqrt{(l_1-1)\cdot(l_2-1)}}=\frac{1{,}40-1}{\sqrt{(3-1)\cdot(4-1)}}=\frac{0{,}40}{\sqrt{6}}=0{,}16$$

Полученный коэффициент показывает, что связь формы завитка с типом конституции ягнят не велика.

Обработку решетки можно было бы сделать, используя формулу (58). В этом случае обработку следует проводить не по строчкам, а по столбцам.

Полихорический показатель связи – число всегда положительное. Следовательно, он *не выявляет направления* связи, а только указывает на ее величину. Чтобы определить направление связи следует проанализировать расположение частот в клетках решетки по отношению к диагоналям этой решетки.

Полихорический показатель связи может быть использован и для таких данных, когда один признак имеет *качественное* выражение, а другой – *количественное*, например, тип кормления и плодовитость облученных животных. Вместе с тем величина связи между качественными признаками, а также между качественными и количественными признаками может быть определена и по обычной формуле коэффициента корреляции, но тогда технически будет больше расчетов, так как требуется вычислить σ_1 и σ_2. Поэтому для сокращения объема вычислений целесообразно пользоваться полихорическим показателем связи.

9.5.12. Генетический коэффициент корреляции между признаками

В практике селекции большое значение имеет определение связей между наследственно обусловленными признаками. Генетическую обусловленность наследования признаков выявляют, сравнивая признаки потомков с признаками их родителей. Такой анализ можно осуществлять в отношении качественных признаков, которые легко прослеживаются у родителей и потомков данной семьи.

Если же необходимо определить наследование количественных признаков и их генетическую обусловленность, то для анализа недостаточно изучения данных в пределах одного семейства. В этом случае требуется анализ массового материала с применением популяционного и биометрического методов.

Для анализа наследования количественных признаков применяют разработанный Хейзелем (1943) метод определения генетического коэффициента корреляции (r_G) между признаками. Суть метода заключается в том, что на группах родственных животных (матери–дочери, отцы–сыновья, полусёстры) вычисляют четыре коэффициента корреляции между двумя разными фенотипическими признаками (x и y) в пределах каждой сопоставляемой родственной группы и между группами.

Формула для вычисления генетического коэффициента связи между признаками x и y основана на определении фенотипических коэффициентов корреляции между x и y, учтенных у животных родственных групп, например у матерей и дочерей. Это означает, что вычисляют фенотипические коэффициенты корреляции между:

1) признаками x дочерей и матерей, $r_{x_d x_m}$;

2) признаками y дочерей и матерей, $r_{y_d y_m}$;

3) признаками x дочерей и y матерей, $r_{x_d y_m}$;

4) признаками y дочерей и x матерей, $r_{y_d x_m}$.

Данные берут из первичного зоотехнического учета. В результате получения четырех величин r определяют генетический коэффициент связи между признаками x и y, используя следующую формулу:

$$r_G = \sqrt{\frac{r_{x_d y_m} \cdot r_{y_d x_m}}{r_{x_d x_m} \cdot r_{y_d y_m}}}\,, \qquad (59)$$

Формулу применяют для тех случаев, когда оба r в числителе имеют знак «+» или знак «–», но в знаменателе оба r должны быть положительными. Если же под корнем в числителе один из r имеет отрицательное значение (знак «–»), а другой – положителен (знак «+»), то формула видоизменяется:

$$r_G = \frac{(r_{x_d y_m} + r_{y_d x_m}) : 2}{\sqrt{r_{x_d x_m} \cdot r_{y_d y_m}}}\,, \qquad (60)$$

Отрицательная связь между x_d, и x_m или между y_d и y_m свидетельствует о сильном взаимодействии генотипа со средой или на сложный тип наследования (эпистаз, межаллельное взаимодействие). Поэтому в таких случаях по формуле Хейзеля связь между признаками установить нельзя, т.к. она основана на предположении о наличии аддитивного действия генов коррелирующих признаков.

Таблица 14

Данные для вычисления генетического коэффициента корреляции r_G [2, *с.190*]

Порядковый номер пары	Содержание белка (x)		Содержание жира (y)	
	дочери (x_d)	матери (x_m)	дочери (y_d)	матери (y_m)
1	3,1	3,0	4,0	3,9
2	3,3	3,1	4,2	4,0
3	3,2	3,2	4,1	4,0
4	3,0	3,1	4,0	3,8
8	3,4	3,3	4,5	4,2

Рассмотрим использование приведённых формул на примере. Пусть требуется вычислить генетический коэффициент корреляции между содержанием жира и белка в молоке коров-дочерей и их матерей. Исходные данные для пяти пар «мать–дочь» приведены в табл. 44.

Методом вычисления коэффициента корреляции для малой выборки (с. 129) определяют значения $r_{x_d x_m}$, $r_{y_d y_m}$, $r_{x_d y_m}$ и $r_{y_d x_m}$

Расчёт позволяет установить следующее:

– содержание белка (матери-дочери) – $r_{x_d x_m}$ =0,65;

– содержание жира (дочери-матери) – $r_{y_d y_m}$ =0,80;

– содержание белка (дочери) - содержание жира (матери) – $r_{x_d y_m}$ = 0,733;

– содержание жира (дочери) - содержание белка (матери) – $r_{y_d x_m}$ = 0,053.

Подставляя эти данные в формулу (21), получают

$$r_G = \sqrt{\frac{r_{x_d y_m} \cdot r_{y_d x_m}}{r_{x_d x_m} \cdot r_{y_d y_m}}} = \sqrt{\frac{0{,}73 \cdot 0{,}05}{0{,}65 \cdot 0{,}80}} = \sqrt{0{,}0702} \approx 0{,}27 .$$

Следовательно, генетическая обусловленность связи между белковомолочностью и жирномолочностью коров выражается r_G= 0,27. Поэтому отбор животных по содержанию жира в молоке будет сопровождаться некоторым повышением белковомолочности.

Рассмотренная группа коэффициентов, позволяющих определять величину и направление связи между признаками, показала, что для решения ряда вопросов, в зависимости от материала, могут быть использованы разные коэффициенты. Корреляционный и регрессионный анализ имеет большое значение для планирования и прогнозирования уровня того или иного признака.

9.5.13. Криволинейные связи, корреляционное соотношение

Ранее были рассмотрены корреляционные коэффициенты, характеризующие прямолинейные или близкие к ним связи. Вместе с тем, в природе существует множество связей криволинейного типа. Для их анализа используют другие коэффициенты.

Многие признаки сельскохозяйственных животных проявляют криволинейную связь с возрастом. Изменение живой массы животного от рождения до наступления половой зрелости выражается так называемой логистической кривой. У насекомых наблюдаются периодические вспышки усиления и затухания в размножении, что можно выразить периодической синусоидной кривой.

Криволинейную связь между признаками, можно охарактеризовать с помощью показателя, называемого **корреляционным соотношением** $\boldsymbol{\eta}$. Корреляционное соотношение позволяет определить величину связи, но не выявляет её направления. По этой причине η, изменяясь от 0 до +1, всегда имеет положительные значения.

Корреляционное соотношение применяют для измерения криволинейных связей. Вместе с тем его можно применять для определения величины и прямолинейных связей. При криволинейном типе связи величина η всегда больше величины r, вычисленной по данным той же самой выборки. Если же связь прямолинейна, то $\eta = r$.

Корреляционное соотношение может иметь два значения: y по x и x по y. Следовательно, оно выявляет неравенство связей, т.к. $\eta_{xy} \neq \eta_{yx}$. В общем виде формула корреляционного соотношения является отношением частных (групповых) средних квадратических отклонений к общим средним квадратическим отклонениям:

$$\eta_{yx} = \frac{\sigma_{yx}}{\sigma_y} \tag{61}$$

$$\eta_{xy} = \frac{\sigma_{xy}}{\sigma_x} \tag{62}$$

где

$$\sigma_{yx} = \sqrt{\frac{\sum p_x \cdot (\bar{y}_x - \bar{y}_{общ})^2}{n}} \tag{63}$$

$$\sigma_{xy} = \sqrt{\frac{\sum p_y \cdot (\bar{x}_y - \bar{x}_{общ})^2}{n}} \tag{64}$$

Для вычисления корреляционного соотношения по вариационным рядам, оформленным с распределением признаков на варьирующие классы (то есть при взвешенных рядах), пользуются следующей формулой:

$$\eta_{yx} = \sqrt{\frac{\sum (\bar{y}_{част} - \bar{y}_{общ})^2 \cdot p_x}{\sum (y - \bar{y}_{общ})^2 \cdot p_y}}\ , \tag{65}$$

где

$\bar{y}_{общ}$ – среднее арифметическое признака y;

$\bar{y}_{част}$ – средние величины признака y по классам признака x. Здесь зависимой функцией является признак y, а независимым аргументом – признак x.

Для определения η_{yx} составляют корреляционную решетку и определяют эмпирические величины регрессии признака у в зависимости от признака *x*. Рассмотрим такой расчёт на примере [2, *с.188*]. Пусть требуется определить влияние возраста (*x*) свиноматок на их плодовитость (*y*). Возраст выражен через число опоросов. В выборку вошло 75 опоросов (*n*), полученных от 50 свиноматок. В каждом опоросе учтено число родившихся поросят.

Алгоритм вычислений:

1) Составляем корреляционную решётку зависимости *y* от *x* (табл. 45).

Таблица 45

Корреляционная решётка зависимости плодовитости свиноматок (*y*) от их возраста, выраженного числом опоросов (*x*)

Плодовитость, голов		Возраст (по числу опоросов), *x*								p_y
границы класса	середина класса, *y*	1	2	3	4	5	6	7	8	
7-9	8	6	3	2	–	–	–	1	1	13
10-12	11	4	6	6	5	6	2	2	3	34
13-15	14	–	1	5	4	4	4	1	1	20
16-18	17	–	–	2	1	–	–	–	–	3
p_x		10	10	15	10	10	6	4	5	*n*=70

2) Вычисляют эмпирические величины регрессии изменения признака плодовитости (*y*) от изменения возраста (*x*). Для этого умножают частоту вариант каждой клетки решетки (p_{yx}) на соответствующую величину признака у каждого класса, то есть по столбцам ($p_{yx}\cdot y$). Получаемые произведения ($p_{yx}\cdot y$) суммируют, а полученную сумму делят на число p_x:

1-й столбец: (6·8+4·11):(6+4) = (48+44):10 = 92:10 = 9,2;
2-й столбец: (3·8+6·11+1·14):(3+6+1) = 10,4;
3-й столбец: (2·8+6·11+5·14+2·17):(2+6+5+2)=12,4
4-й столбец: (5·11+4·14+1·17):(5+4+1)=12,8
5-й столбец: (6·11+4·14):(6+4)=12,2
6-й столбец: (2·11+4·14):(2+4)=13
7-й столбец: (1·8+2·11+1·14):(1+2+1)=11,0
8-й столбец: (1·8+3·11+1·14):(1+3+1)=11,0

Результаты этих расчётов вписывают в строку «y/x – ряд регрессии, голов».

Таблица 46

Результаты расчётов в корреляционной решётке (табл.11)

Плодовитость, голов		Возраст (по числу опоросов)								p_y
границы класса	середина класса, у	1	2	3	4	5	6	7	8	
7-9	8	6	3	2	–	–	–	1	1	13
10-12	11	4	6	6	5	6	2	2	3	34
13-15	14	–	1	5	4	4	4	1	1	20
16-18	17	–	–	2	1	–	–	–	–	3
p_x		10	10	15	10	10	6	4	5	n=70
y/x – ряд регрессии, голов		9,2	10,4	12,4	12,8	12,2	13,0	11,0	11,0	
$\bar{y}_{общ} = \frac{809}{70}$ =10,8										

Полученный ряд изменений y по x показывает эмпирический ряд регрессии. Из него видно, что с 3-го опороса по 6-й наблюдается повышение плодовитости свиноматок, а с 7-го – снижение.

3) Определяют среднюю плодовитость свиноматок по всем возрастным группам, то есть $\bar{y}_{общ}$. Для этого перемножают частоты ряда у т.е. p_y по каждому классу на середину класса. В результате получают несколько частных произведений $y{\cdot}p_y$; их суммируют, получая $\Sigma y{\cdot}p_y$. Разделив вычисленную сумму (809) на общее число обследованных опоросов (n=70), получают среднюю плодовитость по всему корреляционному комплексу: $y_{общ} = \frac{\sum y \cdot p_y}{\sum p_y} = \frac{809}{70} = 10{,}8$ гол. Эти данные показаны в табл. 47.

Таблица 47

Вспомогательные данные к вычислению корреляционного соотношения

Середина классов по плодовитости	p_y	$y{\cdot}p_y$
8	13	8·13=104
11	34	11·34=374
14	20	14·20=280
17	3	17·3=51
	Σp_y=70	$\Sigma y{\cdot}p_y$ =809 $\bar{y}_{общ} = \dfrac{809}{70}$ =10,8

4) Рассчитывают квадраты отклонений каждого показателя у от общей средней $\bar{y}_{общ}$ и полученные величины умножают на частоты ряда *у*, т.е. вычисляют $(y - \bar{y}_{общ})^2 \cdot p_y$. Затем находят сумму $\sum (y - \bar{y}_{общ})^2 \cdot p_y$ (табл. 48).

Таблица 48

Вычисление $\sum (y - \bar{y}_{общ})^2 \cdot p_y$

$\bar{y}_{част}$	$(y - \bar{y}_{общ})^2 \cdot p_y$
8	$(8–10,8)^2·13 = 2,8^2·13 = 101,2$
11	$(11–10,8)^2·34 = 0,2^2·34 = 1,5$
14	$(14–10,8)^2·20 = 3,2^2·20 = 206,1$
17	$(17–10,8)^2·3 = 6,21^2·3 = 115,7$
	$\sum (y - \bar{y}_{общ})^2 \cdot p_y$ = **424,5**

5) Определяют $\sum (\bar{y}_{част} - \bar{y}_{общ})^2 \cdot p_x$, где $\bar{y}_{част}$ – средняя плодовитость по столбцам, полученная в показателях эмпирической регрессии (см. табл. 46, где $\bar{y}_{част}$ равны 9,2; 10,4; 12,4; 12,8; 12,2; 13,0; 11,0; 11,0). Вычисляют отклонения частных средних ($\bar{y}_{част}$) от общей средней ($\bar{y}_{общ}$), возводят их в квадрат и умножают на частоты для эмпирического ряда регрессии, т.е. на p_x, в соответствии с выражением $(\bar{y}_{част} - \bar{y}_{общ})^2 \cdot p_x$. Затем вычисляют сумму этих выражений:

$\sum(\bar{y}_{част}-\bar{y}_{общ})^2\cdot p_x$. Эти промежуточные расчеты приведены в таблице 49.

Таблица 49

Вычисление $\sum(\bar{y}_{част}-\bar{y}_{общ})^2\cdot p_x$

$\bar{y}_{част}$	p_x	$(\bar{y}_{част}-\bar{y}_{общ})^2\cdot p_x$
9,2	10	$(9{,}2–10{,}8)^2$•10=25,3
10,4	10	$(10{,}4–10{,}8)^2$•10=15,2
12,4	15	$(12{,}4– 10{,}8)^2$•15=38,9
12,8	10	$(12{,}8–10{,}8)^2$•10=40,4
12,2	10	$(12{,}8–10{,}8)^2$•10=19,9
13,0	6	$(13{,}0–10{,}8)^2$•6=29,3
11,0	4	$(11{,}0–10{,}8)^2$•4=0,18,
11,0	5	$(11{,}0–10{,}8)^2$•5=0,22
n=70		$\sum(\bar{y}_{част}-\bar{y}_{общ})^2\cdot p_x$ = **169,4**

Полученные величины вводят в формулу (26) корреляционного соотношения для обработки взвешенных рядов:

$$\eta_{yx}=\sqrt{\frac{\sum(\bar{y}_{част}-\bar{y}_{общ})^2\cdot p_x}{\sum(y-\bar{y}_{общ})^2\cdot p_y}}=\sqrt{\frac{169{,}4}{424{,}5}}=\sqrt{0{,}3990}=0{,}632$$

Полученная величина η_{yx}=0,632 указывает на большую зависимость плодовитости от возрастного показателя свиноматок, что следует из ряда регрессии y по x.

СЛОВАРЬ НОВЫХ ТЕРМИНОВ

Генетический коэффициента корреляции (r_G) – показатель степени связи двух наследуемых признаков в двух группах родственных животных (например, корреляцию содержания жира и белка в молоке матерей и дочерей).

Детерминанта – квадрат коэффициента корреляции (r^2), которая показывает долю изменчивости признака (функции) обусловленную влиянием изменчивости воздействующего фактора.

Корреляционная связь криволинейная – такая связь, при которой изменение значений одного признака вызывает сначала увеличение, а затем уменьшение значения другого признака.

Корреляционная связь множественная – связь сразу между несколькими изучаемыми показателями.

Корреляционная связь обратная (отрицательная) **–** такая связь, при которой с увеличением одного показателя другой уменьшается, т.е. изменение значений признаков происходит в противоположных направлениях.

Корреляционная связь простая – связь между двумя изучаемыми признаками, без учета других существующих связей.

Корреляционная связь прямая (положительная) **–** такая связь, при которой с увеличением (уменьшением) одного признака другой также увеличивается (уменьшается), т.е. изменение значений признаков происходит в одном направлении.

Корреляционная связь прямолинейная – такая связь, при которой изменение значений одного признака вызывает пропорциональное изменение значений другого признака.

Корреляционная связь – такая связь, при которой изменение одного признака у ряда особей на определенную величину сопровождается изменениями другого признака на различные (варьирующие) значения.

Корреляционное соотношением η – показатель тесноты связи между взаимосвязанными явлениями (их признаками).

Коэффициент корреляции – статистический показатель, характеризующий совместное распределение двух случайных величин, выражающий их взаимосвязь и равный сумме произведений нормированных отклонении вариант каждого признака от своей средней арифметической, делённой на число наблюдений:

$$r = \frac{\sum t_x \cdot t_y}{n}$$

Коэффициент прямолинейной регрессии, b – произведение коэффициента корреляции r на отношение стандартных отклонений σ_x и σ_y, вычисленных для каждого признака.

Коэффициент регрессии – именованный статистический показатель, который свидетельствует насколько изменяется в среднем признак x если коррелирующий с ним признак y изменится на определённую величину.

Полихорический показатель связи ρ – статистический показатель, используемый для приблизительной оценки величины (но не направления) связи между двумя признаками;

Ранговый коэффициент корреляции по Спирмену – показатель, характеризующий статистическую связь двух признаков, величины которых выражены рангами и позволяющий прогнозировать значения одного признака по значениям другого.

Функциональная связь – такая связь между признаками (показателями), при которой с изменением одного признака (показателя) на определенную величину, другой признак (показатель) изменяется тоже на определенную величину.

ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

1. Какие типы связей встречаются в природе?
2. Какую связь называют функциональной?
3. Какую связь называют корреляционной?
   - 3.1. Какой может быть корреляционная связь по своим математическим особенностям?
   - 3.2. Какую связь называют прямой? Как ещё называют такую связь? Приведите примеры.
   - 3.3. Какую связь называют обратной? Как ещё называют такую связь? Приведите примеры.
   - 3.4. Какую связь называют прямолинейной? Приведите примеры.
   - 3.5. Какую связь называют криволинейной? Приведите примеры.
   - 3.6. Какую связь называют простой? Приведите примеры.
   - 3.7. Какую связь называют множественной? Приведите примеры.
   - 3.8. Между какими признаками может быть выявлена корреляционная связь? Приведите примеры.
4. Что называется коэффициентом корреляции?
   - 4.1. Что показывает коэффициент корреляции?
   - 4.2. Какие значения может принимать коэффициент корреляции?

4.3. Какая корреляция называется малой, средней и большой?

4.4 Какую корреляционную связь **не** может характеризовать коэффициент корреляции?

5. Как вычисляют коэффициент корреляции для малой выборки?

5.1. Какова рабочая формула, позволяющая вычислить коэффициент корреляции для малой выборки?

5.2. Каков алгоритм вычисления коэффициента корреляции для малой выборки?

6. Как вычисляют коэффициент корреляции для большой выборки?

6.1. Какова рабочая формула, позволяющая вычислить коэффициент корреляции для малой выборки?

6.2. Каков алгоритм вычисления коэффициента корреляции для малой выборки?

7. Как вычисляют ошибку коэффициента корреляции?

7.1. Что показывает ошибка коэффициента корреляции?

8. Как вычисляют достоверность коэффициента корреляции?

8.1. Что показывает достоверность коэффициента корреляции?

9. Что называют коэффициентом регрессии?

9.1. Что показывает коэффициент регрессии?

9.2. По какой формуле вычисляют коэффициент регрессии?

10. Как вычисляют коэффициент корреляции между альтернативными признаками?

11. Как вычисляют полихорический коэффициент связи?

12. Как вычисляют ранговый коэффициент связи по Спирмену?

13. Как вычисляют корреляционное соотношение при криволинейных корреляционных зависимостях?

14. Как вычисляется генетический коэффициент корреляции между признаками?

ЗАДАЧИ

1. Длины первого молярного *x* и второго молярного *y* зубов у ископаемого млекопитающего *Phenacodus sp.* оказались следующими (в мм):

x	10,7	10,8	10,6	10,7	10,1	11,2	11,4	12,1	12,3	12,0	12,3	12,7	12,9
y	11,2	10,9	10,5	10,5	9,6	11,2	11,3	12,2	12,1	11,7	11,0	13,2	13,0
x	12,8	13,1	13,3	13,3	13,4	12,7	12,5	12,7	13,6	13,5	13,7	13,6	13,8
y	12,2	13,4	12,6	12,2	12,0	11,2	11,4	11,3	13,6	13,2	12,7	12,9	12,3

Определите коэффициент корреляции, оцените его достоверность и установите доверительные границы при $P = 0,05$.

2. У окуня озера Баторино измерены длина головы *x* и длина грудного плавника *y:*

x	66	61	67	73	51	59	48	47	58	44	41	54	52	47	51	45
y	38	31	36	43	29	33	28	25	36	26	21	30	28	27	28	26

Определите корреляцию между *x* и *y*.

3. Установить, есть ли корреляция между высотой головы *x* и длиной 3-го членика усика *y* у дрозофилы. Для этого с помощью окуляр-микрометра получены следующие данные по *x* и *y* (в делениях окуляр-микрометра):

x	15	16	15	15	16	16	17	18	18	17	17	17	15	16	15	15	15	17
y	29	31	32	33	32	33	33	36	36	35	35	35	35	33	31	31	31	35
x	15	13	15	14	17	15	16	15	15	16	15	16	15	16	18	17	14	15
y	33	30	32	31	35	33	33	32	30	33	33	33	30	31	34	34	31	33
x	14	15	15	13	15	16	14	15	15	15	14	15	15	15	16	18	15	14
y	31	31	33	30	30	33	30	33	31	32	30	31	31	32	33	35	32	32
x	15	15	14	16	17	15	15	15	14	15	14	15	17	15	17	15	14	15
y	32	31	31	33	35	32	31	34	30	33	32	32	35	31	36	33	33	33
x	18	17	17	18	17	17	16	17	18	18	16	16	17	17	16	16	17	16
y	35	36	34	35	33	32	34	34	34	35	35	33	34	33	35	33	33	33

Вычислите коэффициент корреляции и определите его достоверность.

4. Между живым и убойным весом свиней на материале 533 голов был получен r=0,986. Каковы доверительные границы этого коэффициента корреляции при вероятности 0,95?

5. Получены следующие данные о продолжительности беременности у кроликов породы шиншилла при различных размерах помета (число крольчат в помете *x* и длительность беременности в днях *y*):

x	1	8	3	5	7	8	4	8	3	4	4	8	8	5	7	6	6	5	6
y	33	30	31	31	31	32	31	31	32	33	32	31	31	31	31	30	31	32	32
x	6	5	7	8	10	6	7	6	7	6	5	10	7	8	8	6	5	6	5
y	32	32	31	32	31	31	30	31	31	32	31	30	32	32	31	31	31	32	30
x	6	8	6	5	8	7	6	5	9	5	3	4	7	8	9	5	6	2	2
y	31	31	32	32	31	30	32	31	31	31	32	32	31	31	31	31	31	32	33

Есть ли корреляция между длительностью плодоношения и размерами помета?

6. Учитывали плодовитость самок серебристо-черных лисиц *x* в совхозе «Белорусский» и плодовитость их дочерей *y:*

x	6	7	5	6	5	5	4	5	5	4	6	7	6	5	6	6	7	5	6	6	7	3	6	7	4
y	4	5	4	4	6	2	3	3	2	6	6	9	7	2	4	7	5	6	8	9	4	5	5	4	2
x	5	5	6	5	6	4	5	7	7	6	6	6	5	7	6	5	4	6	5	5	6	6	5	6	8
y	5	6	5	9	3	6	4	3	4	2	5	7	3	5	4	5	3	7	2	3	2	5	3	7	4
x	5	6	6	5	6	6	5	5	7	6	6	5	6	6	5	7	5	5	5	4	4	5	5	5	6
y	2	4	5	3	6	3	2	5	3	4	3	3	5	3	6	5	6	5	4	2	3	4	6	2	7
x	6	5	6	5	6	5	6	5	6	6	6	8	7	8	4	5	6	7	7	7	7	5	5	6	4
y	6	4	2	6	4	1	2	4	5	3	2	7	4	6	5	6	3	4	5	3	7	3	8	8	3

Есть ли корреляция между плодовитостью матерей и плодовитостью их дочерей?

7. Были получены следующие данные о весе ягнят-баранчиков (одинцов) *y* и весе баранов – их отцов – *x*: (в кг).

x	76,6	72,2	67,0	66,5	63,3	65,4	63,9	63,1	63,0	62,5	62,2
y	4,56	4,79	4,49	4,32	4,59	4,32	4,67	4,29	4,57	4,20	4,12
x	61,0	60,2	60,0	59,6	59,5	58,9	58,0	57,8	57,6	57,0	56,9
y	4,13	4,70	3,80	4,23	3,76	4,08	4,61	4,37	4,30	4,0	3,92
x	56,8	55,4	55,0	53,8	53,7	52,0	51,4	51,0	50,9	48,5	48,4
y	3,82	4,12	4,19	4,16	4,09	4,12	4,02	4,31	4,06	4,03	4,01

Есть ли корреляция между весом баранчиков и весом их отцов?

8. При объединении ряда данных о корреляции между длиной крыла и длиной хоботка у пчел r=0,721 (n=126 пчел). Каковы его доверительные границы при доверительной вероятности 99%?

9. У 100 серебристо-черных лисиц были измерены (в см) длина туловища *x* и длина хвоста *y*:

x	70	65	66	65	71	68	64	57	66	65	67	62	67	62	63	57	64	66	69	58
y	40	40	40	40	40	42	39	38	41	43	39	45	43	38	40	40	41	45	43	37
x	63	67	67	67	65	65	67	70	65	71	69	64	64	66	69	72	66	66	67	66
y	45	38	39	37	42	38	38	38	38	40	39	43	43	42	40	41	47	47	40	40
x	76	68	71	71	67	66	69	64	69	71	64	71	66	68	68	66	65	66	67	66
y	41	40	41	34	38	44	47	37	42	40	40	42	39	45	36	40	40	40	40	37
x	68	65	63	66	66	65	65	65	65	64	66	67	63	64	69	69	65	65	69	67
y	40	40	40	40	40	41	39	41	39	40	40	42	38	43	41	41	40	40	41	45
x	68	61	69	61	64	62	66	59	65	62	68	61	67	69	69	66	66	71	67	65
y	40	38	37	38	37	39	37	39	38	40	41	39	41	40	50	40	40	39	44	41

Есть ли корреляция между длиной туловища и длиной хвоста у лисиц? Если коэффициент корреляции окажется достоверным, определите доверительные границы для него при $P = 0,05$.

10. Были получены следующие данные о весе x (в кг) и длине туловища y (в см) 100 серебристо-черных лисиц:

x	4,7	4,6	5,2	5,1	5,3	5,3	4,6	4,8	5,8	5,7	4,5	5,7	5,0	4,8	4,7	5,2	4,6
y	70	65	69	70	66	68	65	71	69	68	57	73	65	67	71	62	69
x	5,5	5,5	4,8	4,7	6,0	5,1	5,2	4,5	5,0	5,0	4,9	5,5	5,9	5,6	5,2	5,7	5,3
y	62	63	67	64	64	66	68	69	58	63	67	74	67	67	70	65	71
x	5,4	5,3	4,6	5,6	5,1	4,9	5,2	5,3	50	5,3	5,6	5,0	5,1	5,5	5,6	5,2	5,0
y	63	64	64	66	63	69	62	72	66	66	67	67	66	63	67	62	71
x	5,5	5,6	5,0	6,7	4,7	5,3	5,1	5,0	5,0	5,1	4,8	5,0	6,0	5,5	4,6	4,5	4,5
y	67	66	66	69	64	69	70	62	68	68	72	68	67	66	69	65	65
x	5,4	5,0	4,9	5,0	5,7	5,9	5,6	5,1	5,1	4,6	4,4	6,9	5,6	5,9	5,1	4,5	
y	65	65	64	66	66	67	62	63	64	69	69	68	65	69	67	68	
x	4,8	5,5	6,0	5,3	4,8	5,3	5,1	5,4	4,7	5,0	5,9	5,0	5,9	5,6	5,9	5,1	
y	61	64	62	66	59	65	62	68	61	67	69	69	66	66	67	70	

Есть ли корреляция между весом и длиной туловища у лисиц?

11. Были получены следующие данные о весе *x* (в г) левой камеры сердца и длине ядер *y* (в мк) в мышцах сердца;

x	207	221	256	262	273	289	291	292	304	328	372	397	460	632
y	16,6	18,0	15,9	20,7	19,4	19,8	11,7	21,0	23,0	13,6	19,6	22,9	19,4	28,4

Ввиду резко асимметричного распределения вариант по ряду *x* примените для установления связи коэффициент ранговой корреляции.

12. На 12 экземплярах солонгая (*Mustela altaica*) были получены следующие данные о длине тела *x* (в мм), длине хвоста *y* (мм) и весе *z* (в г):

x	172	175	163	165	161	174	159	154	163	172	164	172
y	81	90	74	78	70	86	80	70	79	81	77	81
z	83	110	72	84	86	130	116	80	122	83	94	83

Вычислите коэффициенты ранговой корреляции. Есть ли связь между длиной тела и длиной хвоста и длиной тела и весом солонгаев? Оцените степень достоверности полученных коэффициентов ранговой корреляции.

13. Были получены следующие данные о весе *x* (в кг) и размерах шкурки – длина *y* и ширина *z* (в см) у бобров:

Самки	*x*	8,90	13,30	14,35	19,00	29,6	
	y	76	92	85	104	105	
	z	37	49	48	54	44	
Самцы	*x*	5,40	6.45	700	11,10	12,0	12,05
	y	69	71	83	85	84	83
	z	35	35	42	42	42	44
	x	14,65	17,15	17,25	19,05	19,55	29,9
	y	87	93	97	92	92	102
	z	44	49	46	48	50	50

Есть ли корреляция между этими признаками?

14. У 25 экземпляров днепровского ерша были изучены: длина тела *x* (в см), вес *y* (в г) и вес гонад *z* (в г):

x	10,0	10,0	10,4	10,4	10,5	10,5	10,6	10,7	10,7	10,7	10,8	10,8	10,9
y	19,0	20,0	28,0	35	27	26	28	28	30	27	29	27	31
z	2,2	2,1	3,1	4,4	3,5	2,9	3,8	3,2	2,1	2,6	3,4	2,8	3,2
x	11,0	11,0	11,0	11,0	11,1	11,2	11,3	11,3	11.5	12,5	12,6	13,6	
y	27	31	30	28	32	45	31	35	37	49	54	56	
z	3,2	3,3	4,0	3.4	2,6	3,2	2,8	4,6	3,8	6,7	7,0	8,7	

Определите коэффициенты корреляции между этими признаками и значения средних арифметических с их ошибками.

15. У 16 экземпляров щук были измерены длина тела *x* (в см), вес *y* (в г) и вес икры *z* (в г):

x *y* *z*	33,4 456 32	31,8 375 34	38,0 484 24	33,4 456 19	42,5 788 126	90,0 7900 744	38,0 9581 42	67,0 3550 579
x *y* *z*	35,4 478 49	42,8 783 138	36,0 365 22	50,5 1300 110	53,4 1998 287	62,0 2320 149	64,0 3650 461	71,0 3450 202

Определите коэффициенты корреляции между всеми тремя признаками (общие и частные).

16. Коровы холмогорские помеси 2-го поколения по высоте в холке ***x***, глубине груди *и* и ширине в моклоках *г* были следующими:

x *y* *z*	125 69 56	126 69 52	133 70 49	130 71 53	126 68 42	132 73 56	130 72 53	130 72 53	122 66 51	133 76 57	131 70 50
x *y* *z*	131 57 55	138 73 50	132 71 54	127 71 53	125 68 50	122 67 50	123 69 49	128 70 52	126 70 52	126 65 51	124 68 52
x *y* *z*	131 69 54	123 70 51	131 70 54	132 72 52	129 67 54	133 70 53	124 60 46	124 68 55	126 64 50	123 65 47	

Вычислите r_{xy}, r_{xz}, r_{yz}.

17. В опытах по кормлению 30 крыс в течение 28 дней были получены следующие данные (в г) (начальный вес *x*, количество скормленной пищи *y* (г), конечный вес крыс *z* (г):

x *y* *z*	25,8 98 14,8	15,8 116 9,7	18,1 104 11,3	13,3 99 26,0	20,1 153 44,7	10,1 98 21,0	17,1 103 25,2	21,0 112 13,7	23,7 133 38,5	11,2 80 5,8
x *y* *z*	10,2 87 17,7	16,4 138 40,0	15,9 96 17,1	8,0 102 3,0	26,0 155 37,3	2,4 107 9,7	7,5 142 36,3	15.9 110 21,2	10,7 80 4,5	6,4 83 4,0
x *y* *z*	16,9 105 20,2	12,2 96 20,5	13,4 90 18,9	15,0 24 26,4	13,8 153 25,4	17,8 82 9,4	20,4 88 21,2	7,9 66 9,2	16,0 118 41,1	12,8 135 31,3

Вычислите коэффициенты корреляции: 1) между x и y, 2) между x и z; 3) между y и z.

18. В 36 анализах крови определяли: x – число эритроцитов (в миллионах) и y – содержание гемоглобина (в %):

x	0,80	0,71	2,63	3,19	2,80	3,14	3,21	3,28	3,63	3,30	4,10	3,29
y	22	45	61	66	72	83	73	82	78	82	81	82
x	3,46	3,32	3,11	3,28	3,66	3,90	4,33	3,80	3,82	3,81	4,20	4,47
y	77	80	82	79	84	75	82	79	87	87	87	90
x	3,71	4,22	3,90	4,36	1,30	2,50	2,80	3,10	2,87	3,68	3,59	3,40
y	97	96	92	94	27	50	63	71	70	72	76	71

Определите коэффициент корреляции r_{xy}, r_{xz} и r_{yz}.

19. В 36 анализах крови определяли: x – число эритроцитов (в миллионах), y – содержание гемоглобина (в %) и z – оседание крови за 24 часа (в мм):

x	0,80	0,71	2,63	3,19	2,80	3,14	3,21	3,28	3,63	3,30	4,10	3,29
y	22	45	61	66	72	83	73	82	78	82	81	82
z	8	18	24	26	28	29	30	30	30	30	32	32
x	3,46	3,32	3,11	3,28	3,66	3,90	4,33	3,80	3,82	3,81	4,20	4,47
y	77	80	82	79	84	75	82	79	87	87	87	90
z	32	33	33	34	34	34	34	35	36	37	37	38
x	3,71	4,22	3,90	4,36	1,30	2,50	2,80	3,10	2,87	3,68	3,59	3,40
y	97	96	92	94	27	50	63	71	70	72	76	71
z	40	40	40	44	12	20	26	28	29	30	30	30

Определите коэффициент корреляции r_{xy}, r_{xz} и r_{yz}.

9.6. ДИСПЕРСИОННЫЙ АНАЛИЗ

9.6.1. Сложная обусловленность изменчивости признаков

Изменчивость признаков живых организмов зависит от многих внешних и внутренних факторов. Например, у разных поросят рост и развитие будут варьировать в зависимости от наследственных особенностей родителей, от числа детёнышей в помёте, в котором они родились, от их размеров, от времени рождения, от условий содержания, от количества получаемой пищи и других факторов.

Такая сложная обусловленность биологических свойств и признаков организмов привела к необходимости разработки соответствующих математических методов, с помощью которых можно было бы определять влияние отдельных факторов и оценивать их относительную роль в общей изменчивости признаков. Одним из таких методов является дисперсионный анализ.

9.6.2. Сущность дисперсионного анализа

Дисперсией называют само разнообразие в группе организмов и первичную меру, которая определяет степень этого разнообразия. Этой мерой является квадрат отклонений вариант от их средней величины в данной совокупности. При проведении дисперсионного анализа оперируют дисперсиями трёх типов: общей дисперсией C_y, факториальной, или частной дисперсией C_x, и случайной, или остаточной дисперсией C_z. Смысл этих величин рассматривается ниже.

Дисперсионный анализ (или **вариансный анализ** – *от англ.:* analysis of variance) – это статистический метод измерения связи изучаемого признака с факторами, оказывающими влияние на этот признак [7].

Сущность дисперсионного анализа заключается в изучении статистического влияния одного или нескольких факторов на результативный признак. Для этого общий размах вариации распределяют по источникам и оценивают достоверности влияния факторов. Степень влияния факторов на признак оценивают по удельному весу соответствующей факторной дисперсии в общей дисперсии признака.

9.6.2.1. Цели использования дисперсионного анализа

Дисперсионный анализ используют для установления достоверности и силы влияния факторов на признак, а также для определения относительной доли одного или нескольких факторов в общей изменчивости признака.

9.6.2.2. Основные понятия и термины

Результативный признак (***y***) – элементарное качество или свойство объектов, изучаемое как результат влияния факторов: организованных в исследовании (***x***), и всех неорганизованных в данном исследовании (***z***). Результативными признаками могут быть:

а) точно измеряемые особенности объектов (длина, рост, сила, резвость, содержание вещества в биологической жидкости);

б) неточно измеряемые признаки (густота спермы в баллах, умственные способности, конституциональная крепость);

в) комбинированные признаки (индексы продуктивности, отношение размеров тела);

г) качественные признаки (пол, масть, цвет глаз, болезнь).

Фактор – любое влияние, воздействие или состояние, разнообразие которых может так, или иначе отражаться на разнообразии результативного признака.

Градация фактора – степень действия фактора или состояние объектов изучения (пол, возраст, живая масса).

Градации дисперсионного комплекса – это опытные группы исследования. Каждая градация комплекса соответствует одной градации фактора и включает те объекты (с их вариантами), которые подверглись одной степени действия фактора или находились в одном из изучаемых состояний.

9.6.2.3. Градации факторов и их характер. Статистические комплексы и их типы

Обычно каждый изучаемый в эксперименте фактор имеет не одно, а несколько значений, которые называют **градациями фактора** или **уровнями фактора**. В пределах каждого уровня отдельные переменные (варианты) могут принимать разные значения. В тех более сложных случаях, когда в формировании общей изменчивости участвует несколько факторов, и каждый из них имеет свои уровни, отдельные переменные (варианты) внутри каждого такого уровня также могут принимать разные значения. Факторы, уровни которых не являются точно фиксированными или которые имеют вообще все возможные случайные градации, называют **случайными.**

Существуют разнообразные схемы (модели, статистические комплексы) дисперсионного анализа, различаемые 1) по числу анализируемых факторов и 2) характеру градаций внутри фактора.

По числу анализируемых факторов различают **одно-, двух-** и **трёхфакторный дисперсионный анализ.** Если факторов больше трёх, то дисперсионный анализ называют многофакторным. Каждый фактор может иметь несколько градаций (классов).

По характеру градаций внутри факторов различают 3 типа схем (моделей, статистических комплексов): а) с **фиксированными градациями,** б) со **случайными градациями** и в) **иерархическими градациями**. В пределах каждой градации записывают соответствующие величины варьирующего признака. Варьирование признака внутри градации исследуемого фактора носит случайный характер и обусловлено влиянием **остаточных** (неучтённых) факторов.

В зависимости от числа вариант и их численного соотношения в каждой отдельной градации дисперсионные комплексы разделяют на равномерные, пропорциональные и неравномерные.

Равномерным называется такой дисперсионный комплекс, в котором число наблюдений по градациям факторов одинаковое и их соотношения равны 1:1:1 и т.д. [4, *с.237*]. Пример равномерного двухфакторного дисперсионного комплекса показан в таблице 50.

Таблица 50

Равномерный двухфакторный дисперсионный комплекс (фактор А – доза облучения, фактор В – пол животного)

Фактор ***A*** (доза облучения, рентген)	A_1=0		A_2=100		A_3=200	
Фактор ***B*** (пол)	B_1=♂	B_2=♀	B_1=♂	B_2=♀	B_1=♂	B_2=♀
Варьирующий признак, ***x*** (плодовитость)	8 9 10	10 10 8	7 6 8	6 6 5	5 3 5	4 2 3
Частоты, n_i или ***p***	3	3	3	3	3	3

Обратите внимание на тот факт, что о равномерности дисперсионного комплекса говорят только в том случае, если в нём анализируют два фактора или более. Однофакторные комплексы не бывают неравномерными, т.к. в их структуре представлены классы только по одному фактору.

В таблице 50 приведен равномерный статистический комплекс, в котором рассматривается влияние двух факторов: доз облучения (***A***) и пола облученного животного (***B***) на показатель плодовитости (***x***).

Выборка имеет 18 наблюдений (n=18). Структура статистического комплекса включает три класса (градации) по фактору A: A_1 – облучения не проводилось, A_2 – облучение в дозе 100 рентген и A_3 – облучение в дозе 200 рентген и два класса (градации) по фактору B: B_1 – группа самцов, B_2 – группа самок. В каждую градацию фактора A входят факторы B_1 и B_2. Таким образом, в комплексе имеется 6 классов, или градаций ($l_a \cdot l_b$ =3·2=6). Все 18 подопытных мышей распределены равномерно, в каждой градации по 3 особи; следовательно, комплекс имеет равномерный, двухфакторный тип. В строчках x каждой градации (в каждой клетке) таблицы проставлены сведения о плодовитости мышей. Для подопытных самок это данные об их плодовитости, а для самцов указана плодовитость слученных с ними самок.

Равномерный комплекс является частным случаем пропорционального комплекса, который рассмотрен ниже.

Пропорциональным называется такой дисперсионный комплекс, в котором число наблюдений по градациям всех факторов различно, но соотношения частот вариант по градациям одного из факторов постоянно. Пример пропорционального дисперсионного комплекса показан в таблице 51.

Таблица 51

Пропорциональный двухфакторный комплекс

Фактор ***A*** (доза облучения, рентген)	A_1=0		A_2=100		A_3=200	
Фактор ***B*** (пол)	B_1=♂	B_2=♀	B_1=♂	B_2=♀	B_1=♂	B_2=♀
Варьирующий признак, ***x*** (плодовитость)	10 12	10 11 11 12	8 9 6	7 9 7 8 6 10	6 8	6 5 7 6
Частоты, n_i или p	2	4	3	6	2	4
Отношение частот ***B*** по классам ***A***	1:2		1:2		1:2	

В табл.51 дано распределение 21 подопытного животного по плодовитости в зависимости от фактора ***A*** (облучение) и с фактором ***B*** (пол животных). Соотношение частот вариант в градациях фактора ***B*** для каждого класса ***A*** одно и то же и составляет 1:2, хотя число наблюдений по классам различное.

В равномерных и в пропорциональных комплексах сумма частных факториальных дисперсий равна общефакториальной дисперсии, т.е. $C_A + C_B + C_{AB} = C_x$.

Неравномерным называется такой дисперсионный комплекс, в котором число наблюдений по градациям факторов неодинаковое. Пример такого комплекса показан в таблице 52.

Таблица 52

Неравномерный двухфакторный комплекс

	A_1 Чистопородные собаки		A_2 Помесные собаки	
	B_1 Инбредные	B_2 Кроссбредные	B_1 Инбредные	B_2 Кроссбредные
x_{ij} (живая масса)	35, 33, 30	36, 35, 37, 35, 36, 33	45, 40, 42, 43	45, 46, 48, 44
Частоты, n_i или p	3	6	4	4
Отношения частот ***B*** по ***A***	1:2		1:1	

В таблице 52 приведено распределение 17 собак по живой массе в зависимости от породности и генетического происхождения. Распределение частот фактора ***B*** по классам фактора ***A*** неравномерное: в классе A_1 оно равно 1:2, а в классе A_2 – 1:1.

Рабочие формулы и техника обработки статистического комплекса меняются в зависимости от его типа. Неравномерные комплексы анализировать сложнее.

9.6.2.4. Нулевая гипотеза в дисперсионном анализе

Как и в других случаях статистического анализа, при дисперсионном анализе следует исходить из первоначально принимаемой нулевой гипотезы. Согласно нулевой гипотезе, принимаемой до проведения дисперсионного анализа, исследуемый фактор **A** (или **B**, или **C** и т.д.) **не влияет** на изменчивость признака. Если нулевая гипотеза правильна, то σ^2_A должна быть равна нулю, т.е. вся вариация сводится только к случайной. То же относится к σ^2_B и σ^2_C. Для того чтобы отклонить нулевую гипотезу, нужно доказать что σ^2_A достоверно (т.е. с вероятностью не меньшей, чем 0,95 или с уровнем значимости 0,05) отличается от нуля. Достоверность значения $\boldsymbol{\sigma^2_A}$ можно установить путём его деления на величину **случайной ошибки** (в дисперсионном анализе роль случайной ошибки выполняет $\boldsymbol{\sigma^2_z}$, т.е. варианса, определяемая неконтролируемыми факторами (см. с. 182).

9.6.3. Общая схема статистического однофакторного дисперсионного комплекса и его анализ

9.6.3.1. Составление однофакторного комплекса

Чтобы понять смысл расчётов при дисперсионном анализе, важно ясно представить возможную вариацию признака в тех группах, на которые разбивается фактический материал.

Рассмотрим составление дисперсионного комплекса, в котором анализируется влияние одного фактора $\boldsymbol{A}$, принимающего разные градации (уровни) от 1 до $\boldsymbol{a}$:

$$1, 2, 3, \ldots \boldsymbol{i}, \ldots, \boldsymbol{a}$$

Отдельные наблюдения (варианты) разбиваются на группы согласно этим градациям фактора $\boldsymbol{A}$. Важно, что фактор только один (например, породная принадлежность скота: голштинфризы, симменталы и т.д., или – время проведения биохимического анализа: 6, 9, 12, 15, 18, 21 и 24 часа). Распределение вариант в таблице однофакторного дисперсионного комплекса представлено в таблице 53.

Таблица 53

Схема варьирования при различии по одному фактору [6, *с. 191*]

Градации фактора $\boldsymbol{A}$	Отдельные варианты (наблюдения), x_{ij}							Суммы по группам, $\boldsymbol{T_i}$	Среднее по группам, $\bar{x}_i$
	1	2	3		$\boldsymbol{j}$		$\boldsymbol{n}$		
1	x_{11}	x_{12}	x_{13}		x_{1j}		x_{1n}	$\sum x_1 = T_1$	$\bar{x}_1$
2	x_{21}	x_{22}	x_{23}		x_{2j}		x_{2n}	$\sum x_2 = T_2$	$\bar{x}_2$
…									
$\boldsymbol{i}$	x_{i1}	x_{i2}	x_{i3}		x_{ij}		x_{in}	$\sum x_i = T_i$	$\bar{x}_i$
…									
$\boldsymbol{a}$	x_{a1}	x_{a2}	x_{a3}		x_{aj}		x_{an}	$\sum x_a = T_a$	$\bar{x}_a$
								$\sum x_{ij} = T$	$\bar{x}$

Разные уровни (градации) фактора принято обозначать буквой ***i*,** а отдельные (варианты) наблюдения – буквой ***j***. Поэтому каждую варианту независимо от того где она находятся можно обозначить в общем виде как $\boldsymbol{x_{ij}}$.

Число групп (градаций фактора, уровней фактора) равно ***a***.

Число вариант (наблюдении) в каждой группе равно ***n***. Равное число вариант в группах не обязательно. При неравном числе можно исходить из среднего числа *n*.

Сумма вариант по каждой группе (графа «Суммы по группам») обозначены буквами ***T*** с соответствующими индексами. Общую сумму обозначают буквой ***T*** без индекса.

В последней графе таблицы указывают средние арифметические по группам ($\bar{x}_i$), а в самой нижней правой клетке – средняя арифметическая для всех вариант всех групп, $\bar{x}$.

9.6.3.2. Характеристика различного варьирования вариант в дисперсионном комплексе

Заполнив таблицу, приступают к разбору варьирования данных. Можно выделить три типа варьирования данных:

1) общее варьирование всех вариант $\boldsymbol{x_{ij}}$ вокруг общей средней $\bar{x}$ независимо от того, в какой группе эти варианты находятся;

2) варьирование групповых средних $\bar{x}_i$ (средних арифметических каждого уровня изучаемого фактора) вокруг общей средней $\bar{x}$;

3) варьирование вариант $\boldsymbol{x_{ij}}$ внутри каждой группы вокруг каждой групповой средней $\bar{x}_i$.

Для характеристики этих трёх варьирований при проведении дисперсионного анализа используют уже известные статистические величины:

– суммы квадратов отклонений от средней арифметической;

– средние квадраты отклонений, т.е. суммы квадратов отклонений, делённые на количество степеней свободы, а это – варианcы $\boldsymbol{\sigma^2}$.

9.6.3.3. Суммы квадратов отклонений

В однофакторных дисперсионных комплексах для всех трёх типов варьирования вычисляют три дисперсии (суммы квадратов отклонений), соответствующие трём видам разнообразия. Они будут следующими:

9.6.3.3.1. Общая сумма квадратов отклонений

Общую сумму квадратов отклонений (**общую дисперсию**), $\boldsymbol{C_y}$ вычисляют по формуле:

$$C_y = \sum_{ij} (x_{ij} - \bar{x})^2 \tag{66}$$

Значок $\boldsymbol{ij}$ под знаком суммы означает, что суммирование производится по всем вариантам всех групп.

9.6.3.3.2. Сумма квадратов отклонений для групповых средних

Сумму квадратов отклонений для групповых средних (**общефакториальную дисперсию**), $\boldsymbol{C_x}$ вычисляют по формуле

$$C_x = \sum_i n_i \cdot (\bar{x}_i - \bar{x})^2 \tag{67}$$

Чтобы эта величина была того же порядка, что и первая, введён множитель $\boldsymbol{n_i}$ т.е среднее число вариант в каждой группе (если число вариант во всех группах одинаков, то просто – $\boldsymbol{n}$).

9.6.3.3.3. Сумма квадратов отклонений от групповых средних внутри каждой группы

Сумму квадратов отклонений от групповых средних внутри каждой группы (т.е. для случайной вариации внутри групп), называемую также **остаточной** или **случайной дисперсией**, $\boldsymbol{C_z}$ вычисляют по формуле:

$$C_z = \sum_i \left[\sum_j (x_{ij} - \bar{x}_i)^2 \right] \tag{68}$$

Два знака сумм указывают, что суммация производится дважды: внутри каждой группы, т.е. по отдельным $\boldsymbol{j}$ (от 1 до $\boldsymbol{n}$) а затем по всем уровням $\boldsymbol{i}$ (от 1 до $\boldsymbol{a}$). Чем меньше величина случайной дисперсии $\boldsymbol{C_z}$, тем больше результативный признак обусловлен влиянием воздействующих факторов ($\boldsymbol{C_x}$) и тем больше выявляется и познаётся закономерность изменчивости признака, возникающая под действием этих факторов.

9.6.3.3.4. Разложение дисперсий

При дисперсионном анализе всегда проводят разложение общего влияния всех факторов (организованных и неорганизованных) на компоненты, соответствующие частным влияниям, действовавшим в изученном комплексе. Такое разложение составляет основу дисперсионного анализа и оказывается возможным, потому что дисперсии комплекса, отражающие силу влияний, обладают одним основным и постоянным групповым свойством аддитивности.

В равномерных и пропорциональных дисперсионных комплексах все дисперсии находятся в определенной взаимосвязи. Всегда сумма частных дисперсий, рассчитанных независимо, равна общей дисперсии:

$$C_y = C_x + C_z \tag{69}$$

В однофакторных комплексах общая дисперсия в точности равна сумме факториальной дисперсии и случайной дисперсии: $C_y = C_x + C_z$. Поэтому можно одну из трех дисперсий вычислять по двум другим:

$$C_x = C_y - C_z \tag{70}$$

$$C_z = C_y - C_x \tag{71}$$

$$C_y = C_x + C_z \tag{69a}$$

В двухфакторных и многофакторных комплексах суммарная общая дисперсия складывается из большего числа составляющих её частных дисперсий.

Аддитивность частных дисперсий положена в основу определения показателей силы влияния. Методы их вычисления рассмотрены ниже.

9.6.3.3.5. Универсальное использование дисперсий

Все конечные результаты дисперсионного анализа можно получить на основе только дисперсий (т.е. без вычисления варианс). Это обстоятельство и определяет название метода: дисперсионный анализ. По одним дисперсиям (без расчета варианс) можно определить: 1) показатели силы влияний; 2) показатель достоверности влияния; 3) доверительные границы генерального параметра силы влияния (приближенно); 4) ошибку репрезентативности показателя силы влияний; 5) ошибку репрезентативности частных средних по градациям дисперсионного комплекса.

Вычисление первых двух показателей – силы влияний и показателя достоверности влияния буден показано ниже. Для знакомства с методами вычисления трёх последних показателей: ошибки репрезентативности показателя силы влияний, доверительных границ генерального параметра силы влияния и ошибки репрезентативности частных средних по градациям дисперсионного комплекса следует обратиться к специальным учебникам по биометрии [3, 5, 8].

9.6.3.4. Степени свободы в дисперсионном анализе

Вариансы (средние квадраты отклонений) в дисперсионном анализе вычисляют для того, чтобы использовать традиционные методы определения достоверности влияний. Чтобы вычислить вариансы, надо разделить каждую сумму квадратов отклонений на соответствующие им числа степеней свободы v (в некоторых пособиях по биометрии используют другое обозначение – *df*), которые будут следующими:

– для общей дисперсии:

$$v_y = n – 1; \qquad (72)$$

– для дисперсии групповых средних:

$$v_x = a – 1, \qquad (73)$$

где a – число вариант в группе (градации фактора);

– для случайной вариации внутри групп:

$$v_z = n – a\ . \qquad (74)$$

Нужно подчеркнуть, что сумма чисел степеней свободы для групповых средних и для вариации внутри групп должна равняться числу степеней свободы для общей дисперсии:

$$(n – a) + (a – 1) = n – 1. \qquad (75)$$

9.6.3.5. Анализ дисперсионного комплекса

Общая схема дисперсионного анализа приведена в таблице 54.

Таблица 54

Схема однофакторного дисперсионного анализа (анализа варианcы)

Источник варьирования	Сумма квадратов отклонений,	Число степеней свободы, v	Средний квадрат отклонений, варианcы	№ формулы для вычисления
Общее (все варианты)	$\sum\limits_{ij}(x_{ij}-\bar{x})^2$	***n*–1**	$\dfrac{\sum\limits_{ij}(x_{ij}-\bar{x})^2}{n-1}$	(76)
Групповые средние (фактор А)	$\sum\limits_{i} n_i \cdot (\bar{x}_i-\bar{x})^2$	***a*–1**	$\dfrac{\sum\limits_{i} n_i \cdot (\bar{x}_i-\bar{x})^2}{a-1}$	(77)
Варианты внутри групп (случайные отклонения)	$\sum\limits_{i}\left[\sum\limits_{j}(x_{ij}-\bar{x}_i)^2\right]$	***n*–*a***	$\dfrac{\sum\limits_{i}\left[\sum\limits_{j}(x_{ij}-\bar{x}_i)^2\right]}{n-a}$	(78)

Вычислить суммы квадратов отклонений по исходным формулам, указанным во 2-м столбце таблицы возможно, но это потребует больших вычислений. Поэтому используют рабочую формулу, по которой сумма квадратов отклонений вычисляется по значениям вариант:

$$\sum x_i^2 - \frac{(\sum x_i)^2}{n} \qquad (79)$$

Если использовать обозначения, приведённые в таблице 53 (см. с. 176), т.е. $\sum \boldsymbol{x_i}$ для каждой группы (градации) фактора ***A*** обозначить как $\boldsymbol{T_i}$, сумму всех вариант – ***T***, число наблюдения в каждой отдельной группе (градации) обозначить как $\boldsymbol{n_i}$, а общее число вариант – ***n***, то **рабочие формулы** для вычисления дисперсий будут выглядеть следующим образом (формулы 80-82).

Общая дисперсия (общая сумма квадратов отклонений):

$$C_y = \sum_{ij} x_{ij}^2 - \frac{T^2}{n}; \tag{80}$$

Факториальная дисперсия (сумма квадратов отклонений для групповых средних):

$$C_x = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n}; \tag{81}$$

Случайная (внутригрупповая) дисперсия (сумма квадратов отклонений для вариант внутри групп:

$$C_z = \sum_{ij} x_{ij}^2 - \sum_i \frac{T_i^2}{n_i} \tag{82}$$

Варианса в дисперсионном анализе – это дисперсия, приходящаяся на один элемент свободного разнообразия, т.е. дисперсия, делённая на число степеней свободы. Поэтому, разделив указанные в формулах (80-82) три суммы квадратов отклонений (дисперсий) на соответствующие числа степеней свободы, получим средние квадраты отклонений (вариансы). Тогда **рабочие формулы** для вычисления варианс будут следующими (83-85).

Вариансы для общего варьирования:

$$\sigma_y^2 = \frac{\sum_{ij} x_{ij}^2 - \frac{T^2}{n}}{n-1} \tag{83}$$

Вариансы для групповых средних:

$$\sigma_x^2 = \frac{\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n}}{a-1} \tag{84}$$

Вариансы для случайных отклонений:

$$\sigma_z^2 = \frac{\sum_{ij} x_{ij}^2 - \sum_i \frac{T_i^2}{n_i}}{n-a} \tag{85}$$

9.6.3.6. Показатель силы влияний

Определение силы влияний по их результатам требуется в биологии, сельском хозяйстве, ветеринарии для выбора наиболее эффективных средств воздействия, для дозировки физических и химических агентов, стимуляторов, возбудителей, лекарственных препаратов, пищевых средств.

В биометрии применяют два метода определения силы влияния Первый метод (Н.А. Плохинского) оперирует величинами дисперсий [2, *с.221*; 3, *с.229*]. Во втором методе (Д.У. Снедекора) используют величины варианс [1, *с.153*; 2, *с. 221*; 3, *с. 229*].

9.6.3.6.1. Вычисление показателя силы влияний с помощью показателей дисперсии (метод Н.А.Плохинского)

Для измерения силы влияния в дисперсионном анализе используют аддитивные свойства частных дисперсий (сумм квадратов центральных отклонений): $C_x + C_z = C_y$.

Отношения частных дисперсий к общей:

$$\frac{C_x}{C_y} + \frac{C_z}{C_y} = \frac{C_y}{C_y} = 1 \qquad (86)$$

показывает долю участия отдельной частной дисперсии в образовании общей дисперсии. А так как каждая частная дисперсия соответствует одному из частных влияний, то отношение частной дисперсии к общей измеряет долю данного влияния в общем суммарном статистическом влиянии всех факторов, определяющих результативный признак. Поэтому доля (или процент) каждой частной дисперсии в общей их сумме принята за показатель силы влияния.

В однофакторном комплексе, чем большую долю в общей дисперсии занимает ее факториальная часть (C_x/C_y), тем большая часть общего разнообразия обусловлена разнообразием градаций фактора. Это означает, что фактор действует с большей силой, оставляя на долю случайных влияний меньшую часть общего разнообразия признака.

Таким образом, сила влияния фактора (или факторов), в дисперсионном анализе измеряется отношением частных дисперсий к дисперсии общей (формула 87). Эту величину принято обозначать греческой буквой «эта», возведённой в квадрат – η^2:

$$\eta_i^2 = \frac{C_i}{C_y} \quad (87)$$

Следует обратить внимание на то, что корень квадратный из величины η^2 даёт величину корреляционного соотношения η, выражающую степень связи между фактором и признаком

В однофакторном комплексе определяются два показателя силы влияния:

1) организованного фактора

$$\eta_x^2 = \frac{C_x}{C_y} \quad (88)$$

2) неорганизованных (случайных) факторов

$$\eta_z^2 = \frac{C_z}{C_y} \quad (89)$$

Сумма этих показателей равна единице:

$$\eta^2{}_x + \eta^2{}_z = 1 = \eta^2{}_y \quad (90)$$

9.6.3.6.2. Вычисление показателя силы влияний с помощью варианс (метод Д.У. Снедекора)

Способ Д.У. Снедекора в качестве показателя силы влияния использует отношение межгрупповой варианс $\sigma_x{}^2$ к варианс общего варьирования $\sigma_y{}^2$. В этом случае формула для определения η^2 будет иметь следующий вид:

$$\eta_x^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} \quad (91)$$

Если в градациях дисперсионного комплекса число вариант разное, то в величину $\sigma_x{}^2$ приходится вносить корректировки (см. [3, *с. 229*]).

Силу влияния регулируемого фактора можно определить непосредственно по разности $\sigma_x{}^2 - \sigma_z{}^2$, используя следующую формулу:

$$\eta_x^2 = \frac{\sigma_x^2 - \sigma_z^2}{\sigma_x^2 + (n-1)\cdot\sigma_z^2} \quad (92)$$

В учебнике [1] величина $\eta^2{}_x$ обозначена символом r_w.

9.6.3.7. Достоверность влияний

Показатель силы влияния, полученный в выборочном исследовании, характеризует ту степень влияния, которая проявилась не во всей генеральной совокупности, а лишь в группе исследованных объектов. По этой причине вычисленный показатель силы влияния может несколько отличаться от реальной силы влияния фактора действовавшей на всю генеральную совокупность.

Учет статистических ошибок репрезентативности в дисперсионном анализе производится в форме критерия достоверности выборочного показателя и доверительных границ генерального параметра силы влияния.

Если разнообразие частных средних в выборочном комплексе меньше критерия Фишера, то это выборочное разнообразие недостоверно. (Оно могло получиться в результате совершенно случайных отклонений от нулевого разнообразия соответствующих генеральных средних). В таких случаях вычисленный выборочный показатель силы влияний недостоверен, и прогноз генерального параметра остаётся неопределённым потому, что влияние исследуемого фактора на биологические объекты в генеральном комплексе осталось не доказанным.

При недостоверности показателя силы влияния эмпирический показатель полностью применим для характеристики влияния лишь только в пределах изученного комплекса, и не может доказывать наличие или отсутствие влияния фактора в генеральном комплексе.

Если разнообразие частных средних в выборочном комплексе равно или превышает критерий Фишера, то исследуемое выборочное разнообразие уже не случайно. По этой причине влияния изучаемого фактора статистически достоверно.

При достоверном влиянии эмпирический показатель силы влияния применим уже не только к выборочному комплексу, но и к генеральному. В таких случаях по вычисленному выборочному показателю вполне определенно можно заключить о наличии влияния в генеральном комплексе (при массовом применении фактора). Кроме того, можно определить возможную генеральную силу этого фактора в форме доверительных границ (при этом нижняя граница не будет отрицательной).

Достоверность влияния может иметь разную степень. Чем больше разнообразие действия градаций фактора, тем больше факториальное разнообразие отличается от случайного. Поэтому за меру достоверности влияния принят результат сопоставления степени двух разнообразий – факториального и случайного.

В качестве критерия достоверности различия факториального и случайного разнообразия используется отношение варианс (критерий Фишера *F*; см.: формулу 34, *с. 94-96*):

$$F = \frac{\sigma_x^2}{\sigma_z^2} \tag{34а}$$

При вычислении критерия Фишера F следует помнить, что делимым всегда выбирают большую из двух сравниваемых варианс, а делителем – меньшую. Поэтому, если факториальная варианса оказалась меньше случайной, показателем достоверности будет

$$F = \frac{\sigma_z^2}{\sigma_x^2} \tag{93}$$

Стандартные значения критерия F для двух показателей степеней свободы $v_1 = v_x$, $v_2 = v_z$ и для двух порогов доверительной вероятности p=0,95и p=0,99 приведены в табл. 32 и 33 (*с. 120* и *122*).

Вариансу σ^2_z иногда обозначают символом σ^2_e. Эта величина в дисперсионном анализе является мерилом случайной ошибки. Поэтому в некоторых учебниках её называют просто «ошибкой».

Сопоставление вычисленного (эмпирического) критерия Фишера F с его стандартными (табличными) значениями может дать два принципиально различных результата: 1) влияние факторов недостоверно и 2) влияние факторов достоверно. Рассмотрим эти возможные результаты подробнее.

9.6.3.7.1. Недостоверное влияние факторов

Вычисленный критерий F не достигает своего табличного значения, взятого в соответствии с выбранной доверительной вероятностью.

В таких случаях при требуемой доверительной вероятности невозможно сделать заключения о равенстве или различии соответствующих генеральных средних. Однозначный вывод невозможно сделать потому, что малое разнообразие выборочных частных средних может получиться при любом (большом или малом, или нулевом) разнообразии генеральных средних по градациям комплекса. В таких случаях нельзя дать определенного прогноза о генеральном влиянии фактора потому, что остается невыясненным, можно или нельзя ожидать с установленной доверительной вероятностью, что при массовом применении фактора получатся результаты, сходные с теми, которые получены в выборочном комплексе.

Следует остерегаться двух ошибочных мнений о недостоверном показателе силы влияния.

1) Нельзя считать, что получение недостоверного показателя силы влияния указывает на то, что «влияния вообще нет», что влияние фактора отсутствует в генеральных совокупностях. Получение при расчётах недостоверного показателя ни подтверждает, ни отрицает генеральное влияние.

2) Нельзя также считать, что при получении недостоверного показателя силы влияния в проведенном исследовании вообще ничего не получено и это исследование проведено без всякой пользы. Такой вывод ошибочен. Та мера влияния, которая получена при анализе, целиком относится к группе изученных объектов. Как экспериментальный факт она должна быть учтена и в данном анализе, и в других, дальнейших работах.

9.6.3.7.2. Достоверное влияние факторов

Эмпирический критерий равен или превышает свое стандартное значение при требуемой доверительной вероятности.

В таких случаях возможен определенный прогноз: генеральные средние по градациям комплекса неодинаковы и их разнообразие подобно тому, которое наблюдалось в выборочном комплексе. Разнообразие частных средних в выборочном комплексе теперь уже не может быть объяснено только случайностями выборочного исследования.

Достоверное влияние означает, что изученный фактор при его массовом применении в определенных градациях и в данных условиях будет оказывать влияние на результативный признак с доверительной вероятностью, найденной при оценке достоверности его силы влияния.

9.6.4. Примеры однофакторного дисперсионного анализа

Ниже показано как метод дисперсионного анализа практически используется для анализа данных полученных в экспериментах.

9.6.4.1. Пример однофакторного дисперсионного анализа при одинаковой численности вариант в группах

В лаборатории определяли содержание вещества x в крови четырёх поросят в различные часы суток. Полученные данные представлены в таблице 55. Такое представление данных в таблице является **стандартным** при дисперсионном анализе и удобным для дальнейших вычислений.

Таблица 55

Результаты лабораторного анализа крови у поросят

Часы суток	Концентрация вещества, x_{ij}				$T_i=\sum x_i$	n_i	$\bar{x}_i$	T_i^2
	1	2	3	4				
15	1,41	0,95	1,00	0,93	4,29	4	1,07	18,4041
18	1,17	1,10	0,84	1,01	4,21	4	1,03	16,9744
21	1,38	1,38	0,91	1,36	5,03	4	1,26	25,3009
24	0,62	0,48	0,43	0,62	2,15	4	0,54	4,6225
6	0,74	0,41	0,41	0,43	1,96	4	0,50	3,9601
9	0,76	0,59	0,74	0,46	2,55	4	0,64	6,5025
12	0,64	1,02	1,04	0,98	3,68	4	0,92	13,5422
					T=23,81 T^2=566,92	n=28		$\sum T_i^2$=89,3069

Таблица 56

Вспомогательная таблица для вычисления $\sum x_{ij}^2$

Концентрация вещества, x_{ij}					Квадрат концентрации вещества, $(x_{ij})^2$			
1	2	3	4		1	2	3	4
1,41	0,95	1,00	0,93		1,9881	0,9025	1,0000	0,8649
1,17	1,10	0,84	1,01		1,3689	1,2190	0,7056	1,0201
1,38	1,38	0,91	1,36		1,9044	1,9044	0,8281	1,8496
0,62	0,48	0,43	0,62		0,3844	0,2304	0,1849	0,3844
0,74	0,41	0,41	0,43		0,5476	0,1681	0,1681	0,1849
0,76	0,59	0,74	0,46		0,5778	0,3481	0,5476	0,2116
0,64	1,02	1,04	0,98		0,4096	1,0404	1,0816	0,9604
								$\sum x_{ij}^2$ = 22,9855

Восьмая (предпоследняя) графа таблицы 55 содержит значения $\bar{x}_i$ концентрации вещества в различные часы суток. Эти концентрации не одинаковы. Однако каждая средняя вычислена только по 4 вариантам. Поэтому вместо обычного приёма сравнения средних арифметических лучше применить дисперсионный анализ, который поможет вскрыть различия между группами в целом.

Так как в рабочих формулах (80, 82, 83, 85) имеется величина $\sum \boldsymbol{x^2_{ij}}$, то полезно составить вспомогательную таблицу, в которой каждое из всех значений $\boldsymbol{x_{ij}}$ возводятся в квадрат и после этого суммируются (табл. 56).

Общая сумма квадратов отклонений (формула 80) равна

$$C_y = \sum_{ij} x_{ij}^2 - \frac{T^2}{n} = 22{,}9855 - 566{,}92 : 28 = 22{,}9855 - 20{,}2471 = 2{,}7384.$$

Для вычисления суммы квадратов отклонений от групповых средних по формуле (81) надо было каждое значение $\boldsymbol{T_i^2}$ разделить на соответствующее $\boldsymbol{n_i}$. Но, так как во все часы дня было по 4 определения, то вычисления упрощаются:

$$C_x = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n} = \frac{1}{n_i} \sum T_i^2 - \frac{T^2}{n} = \frac{1}{4} \cdot 89{,}3069 - \frac{566{,}92}{28} = 2{,}0797$$

Сумму квадратов отклонений для случайных варьирований находим по формуле (82):

$$C_z = \sum_{ij} x_{ij}^2 - \sum_i \frac{T_i^2}{n_i} = 22{,}9855 - \frac{89{,}3069}{4} = 22{,}9855 - 22{,}3267 = 0{,}6588.$$

Рассчитываем показатель силы влияния организованного фактора (формула 88):

$$\eta_x^2 = \frac{C_x}{C_y} = 2{,}0797 : 2{,}7384 = 0{,}76$$

Рассчитываем показатель силы влияния неорганизованных факторов (формула 89):

$$\eta_z^2 = \frac{C_z}{C_y} = 0{,}6588 : 2{,}7384 = 0{,}24$$

Проверяем правильность вычислений по сумме показателей силы влияния организованных и неорганизованных факторов. Она должна быть равна единице:

$$\eta^2_x + \eta^2_z = 0{,}76 + 0{,}24 = 1.$$

Числа степеней свободы в разбираемом примере будет следующими:

для общей вариации: $v = n–1 = 28–1 = 27$;

для групповых средних: $v = a–1 = 7–1= 6$;

для случайных отклонений: $v = n–a = 28 – 7 = 21$.

Для вычисления средних квадратических отклонений (варианс) нужно суммы квадратов отклонений разделить на числа степеней свободы в соответствии с формулами (76-78 и 83-85). После этого может быть составлена сводная таблица дисперсионного анализа (табл. 57).

Таблица 57

Дисперсионный анализ данных о содержании вещества $\boldsymbol{x}$ в крови поросят в разные часы суток.

Источник варьирования	Сумма квадратов отклонений	Число степеней свободы	Средний квадрат отклонения (варианса)	Эмпирическое значение критерия Фишера, $\boldsymbol{F}$	$\boldsymbol{F}$ табличное	
					при Р=0,05	при Р=0,01
Общее варьирование	2,7384 $\boldsymbol{C_y}$	27 $\boldsymbol{n–1}$	–	$\frac{0{,}3466}{0{,}0314}$ = =11,0	2,57	3,81
Фактор А (время суток)	2,0797 $\boldsymbol{C_x}$	6 $\boldsymbol{a–1}$	0,3466			
Случайные отклонения	0,6588 $\boldsymbol{C_z}$	21 $\boldsymbol{n–a}$	0,0314			

Заключительным этапом дисперсионного анализа является сравнение двух средних квадратов отклонений – $\boldsymbol{\sigma^2_x}$ и $\boldsymbol{\sigma^2_z}$. Первая из этих величин (0,3466) характеризует варьирование групп по часам суток, т.е. влияние изучаемого фактора (времени суток), а другая (0,0193) – это случайное варьирование вариант (в данном примере это отдельные определения, внутри каждой группы).

Сравнение $\boldsymbol{\sigma^2}$ выполняют с помощью критерия достоверности $\boldsymbol{F}$. При нулевой гипотезе полагают, что $\boldsymbol{\sigma_1^2 = \sigma_2^2}$ и $\boldsymbol{F}$= 1. Чтобы отвергнуть нулевую гипотезу, надо доказать что неравенство $\boldsymbol{\sigma_1^2}$ и $\boldsymbol{\sigma_2^2}$ не случайно, т.е. выходит за пределы случайной ошибки. Для этого нужно, чтобы вычисленное эмпирическое значение $\boldsymbol{F}$ было равно или превышало табличную величину при избранном уровне доверительной вероятности. Граничные (контрольные) значения $\boldsymbol{F}$ приведены в специальных таблицах (32 и 33, *с.115-118*) для двух уровней значимости:

P=0,05 и P=0,01. Достоверным принимается такое фактически полученное значение ***F***, которое превышает табличное. В приведённом выше примере ***F***=11, что значительно больше табличного (2,57 и 3,81).

Следовательно, можно считать доказанным влияние времени суток на концентрации исследуемого вещества в крови поросят с уровнем значимости $P<0{,}01$, т.е. с вероятностью $p>0{,}99$. Первоначально принятая нулевая гипотеза об отсутствии влияния времени суток отвергается.

Выше были изложены этапы дисперсионного анализа. Для его проведения исходные данные, промежуточные расчёты и окончательные вычисления были организованы в три таблицы (табл. 55-57). Обычно же все исходные данные и промежуточные расчёты дисперсионного анализа представляют в виде одной таблицы (дисперсионного комплекса). Для получения величин, входящих в формулы, статистический комплекс дополняют строчками со следующими вспомогательными величинами:

– n_i – число вариант в градации (не путайте с величиной $\boldsymbol{n}$ – числом вариант во всём комплексе);

– $\sum x_i$ – сумма значений вариант в группе (градации);

– $(\sum x_i)^2$ – квадрат суммы значений вариант

– $h_x = \dfrac{(\sum x_i)^2}{n_i}$ – сумма значений вариант в группе (градации), возведённая в квадрат и разделённая на число вариант в этой группе (градации);

– $\sum \boldsymbol{x_i}^2$ – сумма квадратов значений вариант в каждой градации;

– $\bar{x}_i = \dfrac{\sum x_i}{n_i}$ – частные средние арифметические каждой градации;

Используя эти вспомогательные величины, рассчитывают следующие виличины:

– $H = \dfrac{\left(\sum x_i\right)^2}{n}$ – сумму значений всех вариант в дисперсионном комплексе, возведённую в квадрат и разделённую на число вариант во всём комплексе;

– $\sum(\sum \boldsymbol{x_i}^2)$ – сумму квадратов значений вариант во всём комплексе;

– $\bar{x}_{общ} = \dfrac{\sum\left(\sum x_i\right)}{n}$ – общую среднюю арифметическую для всего статистического комплекса.

На странице 195-196 показан дисперсионный комплекс данных вещества ***x*** в крови поросят (табл. 58; см стр. 193-194), который мы разобрали выше. (Таблицу 58, расположенную на страницах 193-194, нужно рассматривать сразу на двух страницах так, как это показано на вспомогательном рисунке, приведённом ниже).

Таблица 58

Таблица 58 (окончание)

Пример однофакторного дисперсионного анализа влияния времени суток (фактора А) на концентрацию вещества x в крови поросят

		Градация фактора А				(время суток, часы)			число градаций фактора $l_A = j = 7$
		1 6 часов	2 9 часов	3 12 часов	4 15 часов	5 18 часов	6 21 час	7 24 часа	
x_i – концентрация вещества x в крови поросят	первого	1,41	1,17	1,38	0,62	0,74	0,76	0,64	$H = \dfrac{\left(\sum x_i\right)^2}{n} = \dfrac{566{,}92}{28} = 20{,}25$
	второго	0,95	1,10	1,38	0,48	0,41	0,59	1,02	
	третьего	1,00	0,84	0,91	0,43	0,41	0,74	1,04	
	четвёртого	0,93	1,01	1,36	0,62	0,43	0,46	0,98	
n_i число вариант в градации		4	4	4	4	4	4	4	$n_i=4$ $\sum n_i = n = 28$
$\sum x_i$, сумма значений вариант в группе (градации)		4,29	4,12	5,03	2,15	1,99	2,55	3,68	$\sum(\sum x_i) = T = 23{,}81$ сумма всех i-тых вариант в j-том количестве градаций $T^2 = 2381^2 = 566{,}92$
$(\sum x_i)^2$, квадрат суммы значений вариант		18,40	16,97	25,30	4,62	3,96	6,50	13,54	$(\sum x_i)^2 = \sum T_i^2 =$ $=89{,}3069$
$h_x = \dfrac{\left(\sum x_i\right)^2}{n_i}$		18,40:4= =4,60	16,97:4= =4,24	25,30:4= =6,33	4,62:4= =1,16	3,96:4= =0,99	6,50:4= =1,63	13,54:4= =3,39	$\sum h_x = \sum\left(\dfrac{\left(\sum x_i\right)^2}{n_i}\right) = 22{,}34$
$\sum x_i^2$, сумма квадратов значений вариант в каждой градации		4,76	4,30	6,49	1,18	1,07	1,68	3,49	$\sum(\sum x_i^2) = 22{,}9855$ сумма квадратов значений вариант во всём комплексе
$\bar{x}_i = \dfrac{\sum x_i}{n_i}$, частные средние арифметические каждой градации		1,07	1,03	1,26	0,54	0,50	0,64	0,92	$\bar{x}_{общ} = \dfrac{\sum\left(\sum x_i\right)}{n} =$ $=23{,}9855:28=0{,}8566$ обшая средняя арифметическая для всего статистического комплекса

Таблица 58

Пример однофакторного дисперсионного анализа влияния времени

		Градация фактора А			
		1 6 часов	2 9 часов	3 12 часов	4 15 часов
$\boldsymbol{x_i}$ – концентрация вещества $\boldsymbol{x}$ в крови поросят	первого	1,41	1,17	1,38	0,62
	второго	0,95	1,10	1,38	0,48
	третьего	1,00	0,84	0,91	0,43
	четвёртого	0,93	1,01	1,36	0,62
$\boldsymbol{n_i}$ число вариант в градации		4	4	4	4
$\sum \boldsymbol{x_i}$, сумма значений вариант в группе (градации)		4,29	4,12	5,03	2,15
$(\sum \boldsymbol{x_i})^2$, квадрат суммы значений вариант		18,40	16,97	25,30	4,62
$h_x = \frac{(\sum x_i)^2}{n_i}$		18,40:4= =4,60	16,97:4= =4,24	25,30:4= =6,33	4,62:4= =1,16
$\sum \boldsymbol{x_i}^2$, сумма квадратов значений вариант в каждой градации		4,76	4,30	6,49	1,18
$\bar{x}_i = \frac{\sum x_i}{n_i}$, частные средние арифметические каждой градации		1,07	1,03	1,26	0,54

Таблица 58 (окончание)

суток (фактора А) на концентрацию вещества ***x*** в крови поросят

(время суток, часы)			число градаций фактора $l_A = j = 7$
5 18 часов	6 21 час	7 24 часа	
0,74 0,41 0,41 0,43	0,76 0,59 0,74 0,46	0,64 1,02 1,04 0,98	$H = \frac{(\sum x_i)^2}{n} = \frac{566{,}92}{28} = 20{,}25$
4	4	4	$n_i = 4$ $\sum n_i = n = 28$
1,99	2,55	3,68	$\sum(\sum x_i) = T = 23{,}81$ сумма всех *i*-тых вариант в *j*-том количестве градаций $T^2 = 2381^2 = 566{,}92$
3,96	6,50	13,54	$(\sum x_i)^2 = \sum T_i^2 =$ $= 89{,}3069$
3,96:4= =0,99	6,50:4= =1,63	13,54:4= =3,39	$\sum h_x = \sum\left(\frac{(\sum x_i)^2}{n_i}\right) = 22{,}34$
1,07	1,68	3,49	$\sum(\sum x_i^2) = 22{,}9855$ сумма квадратов значений вариант во всём комплексе
0,50	0,64	0,92	$\bar{x}_{общ} = \frac{\sum(\sum x_i)}{n} =$ $= 23{,}9855 : 28 = 0{,}8566$ общая средняя арифметическая для всего статистического комплекса

Другие примеры анализа однофакторного дисперсионного комплекса при одинаковой численности вариант в группах приведены в учебниках [1, *с.151-153* и 2, *с.222-223*].

9.6.4.2. Пример однофакторного дисперсионного анализа при различной численности вариант в группах

Рассмотрим такой пример из животноводческой практики [6, *с. 197*]. Изучали живую массу телят одного и того же возраста, но трёх разных пород. Число телят в группах было неодинаковым – 11, 12 и 8. Стояла задача выяснить влияет ли породная принадлежность на анализируемый признак. Полученные данные были сведены в таблицу 59.

Таблица 59

Живая масса одновозрастных телят трёх разных пород

Порода	Живая масса отдельных телят	$T_i=\sum x_i$	n_i	$\bar{x}_i$	T_i^2
А	120 120 121 122 122 122 123 125 125 126 126	1352	11	122,9	1827904
Б	123 124 125 125 126 127 127 127 128 128 129 129	1518	12	126,5	2304324
В	122 122 125 127 127 127 128 129	1007	8	125,9	1014049
		T=3877	n=31		–
		T^2=15031129			

Сумма квадратов всех вариант – $\sum x_{ij}^2$ равна:
$120^2 + 120^2 + 121^2 + 122^2 + \ldots + 127^2 + 128^2 + 129^2 = 485097$.

Вычисляем общую дисперсию (общую сумму квадратов отклонений, формула 80):

$$C_y = \sum_{ij} x_{ij}^2 - \frac{T^2}{n} = 485097 - \frac{15031129}{31} = 485097 - 484875 = 222$$

Вычисляем факториальную дисперсию (сумму квадратов отклонений для групповых средних, формула 81). Здесь нужно обратить внимание на то, что анализируемом дисперсионном комплексе число вариант в группах не одинаковое. Поэтому использовать величину

$\sum T_i^2$ нельзя и надо вычислять частные $\dfrac{T_i^2}{n_i}$ для каждой группы отдельно:

$$C_x = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n} = \frac{1352^2}{11} + \frac{1518^2}{12} + \frac{1007}{8} - 484875 =$$

$$= 484956–484875 = 81$$

Вычисляем случайную дисперсию (остаточную дисперсию, сумму квадратов отклонений от групповых средних внутри каждой группы, формула 82):

$$C_z = \sum_{ij} x_{ij}^2 - \sum_i \frac{T_i^2}{n_i} = 485956–484956=141$$

Вычисляем числа степеней свободы. Для общей дисперсии $v_y = n–1 = 31–1 = 30$. Для дисперсии групповых средних $v_x = a–1 = 3–1 = 2$. Для случайной вариации внутри групп $v_z = n–a = 30–2 = 28$.

Результаты всех расчётов вносим в сводную таблицу 60:

Таблица 60

Дисперсионный анализ живой массы одновозрастных телят трёх различных пород

Источник варьирования (порода животных)	Сумма квадратов отклонений	Число степеней свободы	Средний квадрат отклонения (варианса)	Эмпирическое значение критерия Фишера, ***F***	***F*** табличное	
					при P=0,05	при P=0,01
Общее варьирование	222	30	–	$\frac{40,5}{5}$ = =8,1	3,34	5,45
Фактор А (порода)	81	2	40,5			
Случайные отклонения	141	28	5,0			

Поскольку эмпирическое значение критерия ***F*** (8,1) превышает табличные значения, то следует сделать следующее заключение: влияние межпородных различий на живую массу телят доказано с уровнем значимости ***P***<0,01, т.е. с доверительной вероятностью ***p***>0,99.

9.6.4.3. Пример однофакторного дисперсионного анализа качественных признаков

При характеристике группы животных по качественному признаку указывают количество особей, имеющих альтернативные варианты этого признака. Например, стадо животных в 500 голов может быть охарактеризовано по таким качественным признакам как пол (470 коров и 30 быков), масть (320 чёрных, 20 белых и 160 чалых), состояние здоровья (460 здоровых и 40 больных лейкозом).

Если в группе из *n* особей у *m* особей имеется данный вариант качественного признака, то доля этого признака $p = m{:}n$ в группе животных имеет то же значение, что и средняя арифметическая для количественных признаков. Поэтому при дисперсионном анализе качественных признаков за общие и частные средние принимаются доли признака в общих и частных группах. Исходя из этого, дисперсия качественного признака может быть выражена уравнением $C=n{\cdot}p{\cdot}q$, где n – общее число животных; p – доля животных, имеющих первый вариант признака (например, здоровые); q – доля животных, имеющих второй вариант признака (больные); $q = 1 - p$.

Подставив в формулу $C=n{\cdot}p{\cdot}q$ величины p и q, выраженные через m и n, можно получить формулы дисперсии:

$$C_y = m - \frac{m^2}{n} = \sum m_x - H \qquad (94)$$

$$C_x = \sum h_x - H \qquad (95)$$

$$C_z = \sum m_x - \sum h_x \qquad (96)$$

где n – общее число животных; m – число животных, имеющих первый вариант признака (например, больные); m_x – число животных, имеющих второй вариант признака (здоровые). Вспомогательные величины H и $\sum h_x$ вычисляют по следующим формулам.

$$H = \frac{(\sum m_x)^2}{n} \qquad (97)$$

Для вычисления H нужно возвести в квадрат число всех животных обладающих анализируемым признаком, и разделить полученную величину на число особей во всей выборке.

$$\sum h_x = \frac{m_x^2}{n} \qquad (98)$$

Для вычисления $\sum h_x$ в каждой отдельно взятой градации фактора число животных, обладающих анализируемым признаком, возводят в квадрат, и делят на число особей во всей выборке.

Рассмотрим пример практического применения дисперсионного анализа качественных признаков. Предположим, что для борьбы с аскаридозом свиней был использован препарат сантонин в пяти увеличивающихся дозах. Результаты наблюдений можно представить в форме однофакторного дисперсионного комплекса. Пятью градациями фактора будут различные дозы сантонина, а результативным признаком – доля свиней, освобожденных от этой глистной инвазии.

Таблица 61

Однофакторный дисперсионный комплекс для качественных признаков

x	Градации фактора						$a = 6$
	0	1	2	3	4	5	
n_i (всего особей в группе)	20	30	50	21	20	20	$\sum n_i = n = 161$
m_i (освободились от инвазии)	0	9	28	11	11	10	$\sum m_i = m_x = 69$
$h_x = m_i^2 : n$	0	2,7	15,7	5,8	6,1	5,0	$\sum h_x = 35,3$
$p = m_i : n$	0	0,30	0,56	0,52	0,55	0,50	–
							$H = 69^2:161= =29,6$

Дальнейшие расчёты удобно представить следующей вспомогательной таблицей

Таблица 62

Вспомогательная таблица для расчёта дисперсионного комплекса качественных признаков

Статистические показатели	Характеристики дисперсии		
	общей, y	факториальной, x	случайной, z
C	$\sum m_x – H =$ $=69–29,6 = 39,4$	$\sum h_x – H =$ $=35,3–29,6 =5,7$	$\sum m_x – \sum h_x =$ $=69–35,3 = 33,7$
$\eta^2 = C_i : C_y$	1,00	0,15	0,85
v	$n–\mathbf{1} = 161–1 = 160$	$a–1 = 6–1 = 5$	$n–a = 161–6 = 155$
$\sigma^2 = C_i : v_i$,	–	1,14	0,22
$F = \sigma_x^2 : \sigma_z^2$	–	5,2	–

Эмпирическое значение критерия Фишера полученное путём вычислений оказалось равным 5,2. Табличные значения этого критерия при степенях свободы v_1= 5 и v_2= 155 равны 2,3 (при P>0,05), 3,1 (при P>0,01) и 4,4 (при P>0,001). Все они меньше эмпирического значения. Следовательно, результаты анализа показывают вполне достоверное действие сантонина. Вместе с тем анализ таблицы 61 показывает, что один сантонин не может дать стопроцентную дегельминтизацию: часть свиней (примерно 50%) остается зараженной. Кроме того, анализ показывает, что третья, четвертая и пятая дозы (все увеличивающиеся) уже ничего не прибавляют к результатам второй дозы: доля выздоровевших животных остается практически той же. Поэтому можно попытаться одновременно с сантонином использовать дополнительные лекарственные препараты или другие лечебные процедуры. Однако их результаты нужно будет анализировать уже с использованием двух- и трёхфакторного анализа. Для знакомства с техникой двух- и трёхфакторного дисперсионного анализа качественных признаков следует обратиться к специальным учебникам по биометрии, например, [3-6, 8].

СЛОВАРЬ НОВЫХ ТЕРМИНОВ

Градации дисперсионного комплекса – опытные группы исследования; каждая градация комплекса соответствует одной градации фактора и включает те объекты (с их вариантами), которые подверглись одной степени действия фактора или находились в одном из изучаемых состояний.

Градация фактора – степень действия фактора или состояние объектов изучения (пол, возраст, живая масса).

Дисперсионный анализ (вариансный анализ – *от англ.:* analysis of variance) – статистический метод качественного измерения связи изучаемого признака с факторами, оказывающими влияние на этот признак.

Дисперсионный комплекс неравномерный – дисперсионный комплекс, в котором число наблюдений по градациям факторов неодинаковое.

Дисперсионный комплекс пропорциональный – дисперсионный комплекс, в котором число наблюдений по градациям всех факторов различно, но соотношения частот вариант по градациям одного из факторов постоянно.

Дисперсионный комплекс равномерный – такой дисперсионный комплекс, в котором число наблюдений по градациям факторов одинаковое и их соотношения равны 1:1:1 и т.д.

Дисперсия общая – величина дисперсии признака, равная сумме дисперсий факториальной (межгрупповой) и случайной (внутригрупповой).

Дисперсия признака – 1) само разнообразие признака в группе организмов; 2) статистический параметр, который определяет степень этого разнообразия. Этот параметр представляет собой отношение суммы квадратов отклонений значений x_i от их средней арифметической к общей численности выборочной совокупности.

Дисперсия случайная (внутригрупповая, остаточная) – средняя арифметическая из частных дисперсий, т.е. дисперсий признака, выявляемых внутри групп объектов, у которых исследуется изменчивость признака.

Дисперсия факториальная (межгрупповая) – величина, равная среднему квадрату отклонений частных средних от общей средней.

Результативный признак (***y***) – элементарное качество или свойство объектов, изучаемое как результат влияния факторов: организованных в исследовании (***x***), и всех неорганизованных в данном исследовании (***z***).

ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

1. Что можно выяснить, используя дисперсионный анализ?
   1.2. Какие три типа дисперсий анализируют в дисперсионном анализе?

2. Что в дисперсионном анализе называют статистическим комплексом?

3. Какие комплексы выделяют исходя из числа учтённых признаков?
   3.1. Какой комплекс называют однофакторным, двухфакторным, многофакторным?
   3.2. Какие комплексы выделяют исходя из особенностей градаций признаков?
   3.3. Что означает понятие «фиксированная градация признаков»?
   3.4. Что означает понятие «случайная градация признаков»?
   3.4. Что означает понятие «иерархическая градация признаков»?

4. Какие комплексы выделяют на основании числа вариант в градациях?
   4.1. Какой комплекс называют равномерным?
   4.2 Какой комплекс называют неравномерным?

5. Каковы основные элементы и ход (алгоритм) дисперсионного анализа?

6. Как выполняют дисперсионный анализ при одинаковой численности вариант в группах?

7. Как выполняют дисперсионный анализ при различной численности вариант в группах?

8. Каким образом выполняют дисперсионный анализ качественных признаков?

ЗАДАЧИ

1. Изучали продолжительность развития эмбрионов у кроликов разных пород. Она оказалась следующей (в днях) [6]:

Породы	Продолжительность развития отдельных крольчат									
Альбиносы	30	36	31	30	34	32	34	32	33	32
	35	32	31	33	33	35	31	33	32	33
Шиншилла	31	32	30	34	32	31	30	31	30	31
	30	32	31	32	30	31	33	32	32	33
Голландские	30	29	30	31	30	30	30	31	31	31
	30	31	29	32	31	31	30	31	31	31
Польские	30	31	29	30	29	30	29	31	29	30
	30	30	31	30	30	30	31	30	31	30

Влияет ли породность на продолжительность развития эмбрионов кроликов?

2. У овец породы **А** изучали живой вес ягнят-одинцов при рождении (в кг), ношенных разное число дней [6]:

Длительность беременности	Живой вес ягнят									
145	3,8	2,9	3,3	3,6	3,8	3,7	4,8	5,1	3,4	3,3
146	3,7	2,9	3,3	3,6	3,9	3,7	4,7	5,0	3,4	3,2
147	3,9	4,1	4,4	5,0	3,0	2,9	4,0	3,2	4,2	4,3
148	4,0	5,2	4,3	2,9	4,1	3,9	3,2	3,9	4,1	4,0
149	4,0	5,3	4,2	3,0	4,0	3,9	4,2	3,3	4,0	4,1
150	4,1	4,3	5,4	3,1	4,0	4,0	4,3	3,9	4,0	4,1
151	4,3	4,2	5,5	4,2	4,1	4,1	4,4	3,5	4,1	3,6
152	4,3	3,6	4,4	5,5	4,0	4,1	4,5	4,1	4,2	4,3
153	4,4	4,7	3,9	4,6	5,7	4,3	4,8	4,9	4,7	4,7

Примените метод дисперсионного анализа для выяснения влияния длительности плодоношения на живой вес ягнят.

3. Получены следующие данные о плодовитости мышей при облучении рентгеновыми лучами [6]:

Группы	Число мышат от отдельных самок			
Контроль	10	12	11	10
Доза 100 р	8	10	7	9
Доза 200 р	7	9	6	4

Примените метод дисперсионного анализа, чтобы выяснить влияет ли облучение на плодовитость мышей.

4. У овец породы **В** изучали живой вес ягнят-одинцов при рождении (в кг), ношенных разное число дней [6]:

Длительность беременности	Живой вес отдельных ягнят									
145	4,1	5,1	3,5	2,8	4,2	4,1	4,0	3,9	4,6	3,5
146	4,2	4,4	4,0	2,9	4,1	4,2	4,4	4,1	4,0	5,1
147	4,1	5,0	2,8	3,9	4,2	4,3	4,4	4,1	4,1	5,1
148	4,4	5,7	3,9	4,5	4,4	4,3	3,8	4,1	4,5	4,4
149	4,3	5,6	3,0	3,9	4,1	4,2	4,3	4,7	4,5	4,4
150	4,5	5,0	5,2	4,6	4,3	3,0	4,7	4,6	4,0	5,1
151	4,6	5,3	5,5	4,4	4,3	3,2	4,0	4,5	5,0	5,2
152	4,6	5,4	6,1	4,8	4,4	3,2	4,8	4,7	4,0	4,2
153	4,8	5,5	5,2	4,9	4,5	3,4	4,9	4,4	5,1	5,3

Определите методами дисперсионного анализа влияние длительности плодоношения на живой вес ягнят.

5. В опытах по изучению влияния синэстрола в дозе 0,5 мг (инъекции масляного раствора и. кристаллов) на вес яйцеводов пятисуточных цыплят получены следующие результаты (в мг) [6]:

Тип влияния фактора	Вес отдельных яйцеводов					
Масляный раствор	125	160	200	141	254	113
Кристаллы	23	40	130	122	44	120
Контроль	5	6	7	9		

Примените метод дисперсионного анализа для установления влияния синэстрола. Сравните средние арифметические отдельных групп, пользуясь величиной статистической ошибки по данным дисперсионного анализа.

6. Годовые удои (в л) отдельных коров распределялись в зависимости от количества отелов следующим образом [6]:

Количество отелов	Годовые удои отдельных коров				
1	2115	2290	2230		
2	2238	2364	2310		
3	2462	2381	2236	2327	
4	2381	2472	2415		
5	2430	2375	2402	2405	
6	2504	2471	2371	2400	2628
7	2439	2508	2439	2784	

Определите методами дисперсионного анализа, влияет ли количество отелов на годовые удои коров?

7. Получены следующие данные о продолжительности эмбрионального развития коз при рождении их одинцами, двойнями, тройнями [6]:

Группы по количеству козлят от одной самки	Длительность эмбрионального развития отдельных козлят											
Одинцы	148	151	153	150	151	150	154	152	151	151	149	150
	152	151	150	152	149	148	151	152	152	152	151	151
Двойни	154	151	152	151	151	152	152	150	151	152	152	151
	152	148	150	151	151	153	152	149	151	148	149	150
Тройни	150	152	149	153	151	148	150	148	149	149		
	147	148	150	149	152	149	148	149	149	148		

Есть ли разница в продолжительности эмбрионального развития между козлятами-одинцами, двойнями и тройнями? Примените метод дисперсионного анализа и обычное сравнение средних арифметических.

8. Определяли концентрацию кальция в 3 листах на 4 растениях турнепса по 2 анализам на каждый лист. Получены следующие данные (в % к сухому веществу) [6]:

Растение	Лист	Анализы	
1	1	3,28	3,09
	2	3,52	3,48
	3	2,88	2,80
2	1	2,46	2,44
	2	1,87	1,92
	3	2,19	2,19
3	1	2,77	2,66
	2	3,74	3,44
	3	2,55	2,55
4	1	3,78	3,87
	2	4,07	4,12
	3	3,31	3,31

Примените дисперсионный анализ для установления роли индивидуальности растений и различий между листьями в изменчивости содержания кальция.

9. При кормлении тушканчиков сухой и влажной пищей получены следующие данные о температурах тела у тушканчиков [6]:

Отдельные наблюдения							
сухая пища				влажная пища			
36,9	36,8	37,0	36,6	37,3	36,8	37,3	37,1
36,7	36,7	36,8	36,6	36,7	37,0	37,0	36,9

Используя метод дисперсионного анализа, определите, зависит ли температура тела от типа кормления животных.

10. Изучали содержание гемоглобина (%) в крови кур разных пород. Получены следующие данные [6]:

Породы	Отдельные наблюдения	
	самцы	самки
Итальянские	87 92 86 91 90 93 90	53 59 50 52 62 60
Куропатчатые	91 90 88 89	64 68 70 60
Минорки	85 82 85 86 89 84	59 62 65 70 65 63
Бентамы	82 82 85	65 68 72

Определите методами дисперсионного анализа, влияет ли породность на содержание гемоглобина у самок и отдельно – у самцов.

11. Гибридные крысы вскармливались самками разных генотипов (четырех разных линий). В таблице приведены средние веса крыс по каждому помету на 28-й день вскармливания (в г) [6]:

Какое влияние на вес вскармливаемых крыс оказывали генотипические различия между самками-кормилицами и пометами, из которых происходили изучаемые крысята?

Крысы - кормилицы из линий			
1	2	3	4
62	55	53	42
68	42	62	54
64	60	50	61
65	51	53	48
60	65	57	40
60	62	59	51
52	64	47	41
49	62	53	50
48	56	40	44
37	70	46	55
36	67	61	45
68	60	55	52
59	53	56	53
58	56	45	42
54	61	57	54

12. У трёх петухов леггорнов определяли количество сперматозоидов. Пробы брали через каждые 2 дня. Получены следующие данные (в млн./мм^3) [6].

Номера петухов	Отдельные определения								
	1	2	3	4	5	6	7	8	9
1	4,1	3,5	4,1	3,6	47	4,4			
2	1,9	1,3	1,6	1,6	1,1	2,0	2,2	1,1	1,1
3	3,7	3,2	2,2	3,4	3,8	2,5	35	4,2	4,1

Определите методами дисперсионного анализа, есть ли статистически достоверные различия в количестве сперматозоидов в сперме разных петухов. Примените метод дисперсионного анализа и обычное сравнение средних арифметических.

13. В течение ряда лет подсчитывали количество жёлтых тел в яичниках серебристо-чёрных лисиц. Получены следующие данные [6]:

Годы	Количество желтых тел у отдельных самок									
1971-1974	5	6	5	5	3	4	6	5	4	6
1975-1978	4	5	6	6	3	4	5	5	6	5
1979-1982	5	4	3	6	6	5	6	5	5	5
1983-1986	4	7	4	5	6	5	5	6	5	5
1987-1990	8	5	6	4	5	5	7	6	6	5
1991-1994	5	8	7	6	5	5	6	6	5	6
1995-1998	5	5	7	5	7	6	6	6	8	6
1999-2002	6	6	5	7	5	7	6	8	6	6

Используя метод дисперсионного анализа, определите, изменилось ли количество жёлтых тел за 30 лет?

14. В четырёх разновозрастных группах мужчин измеряли скорость кровотока в сосудах (см/с). Были получены следующие результаты:

Возрастные группы мужчин	Варианты опыта (пробы)			Средние
	1	2	3	
Первая	7	10	12	9,67
Вторая	9	7	14	10,00
Третья	11	16	20	15,67
Четвертая	15	18	17	16,67

Установите, достоверны ли различия между средними показателями этих групп и если они достоверны, то какова сила влияния возрастных периодов на скорость кровотока.

15. При изучении влияний светового режима на развод гусениц дубового шелкопряда получены следующие результаты наблюдений [3]:

Варианты опыта	Количество гусениц к началу выкармливания	Число погибших гузениц на протяжении 5 дней				
		1	2	3	4	5
Контроль	150	9	8	7	8	17
Полная темнота	150	10	10	9	8	17
Свет 4 ч	150	9	8	8	9	16
Свет 8 ч	150	8	7	8	7	15
Свет 12 ч	150	9	8	8	7	17

Используя метод дисперсионного анализа, определите, влияет ли продолжительность светлого периода суток на выживаемость гусениц.

16. Сравнивали три группы свиней по плодовитости свиноматок:

Группы свиней	Получено живых поросят в шести опоросах						Среднее
Беркширы	6	8	10	7	6	7	7,33
Короткоухие	8	9	7	11	9	8	8,66
Немецкие	10	9	8	11	10	11	9,83

Установите, достоверны ли различия между средними показателями этих групп и если они достоверны, то какова сила влияния породных свойств на плодовитость свиноматок.

17. У кубышки (*Nymphea*) 4 раза в сутки определяли содержание каротиноидов [6]:

Часы суток	Определения									
	1	2	3	4	5	6	7	8	9	10
18	1,42	1,30	1,68	1,59	1,49	1,62	1,36	1,26	1,58	1,66
24	1,45	1,38	1,49	1,71	1,54	1,57	1,34	1,32	1,66	1,39
6	1,48	1,42	1,58	1,67	1,50	1,80	1,35	1,36	1,67	1,49
12	1,43	1,38	1,47	1,33	1,22	1.35	1,10	1,08	1,34	1,11

Влияет ли время суток на содержание каротиноидов в листьях кубышки?

18. У сирени (*Syringa emodi*) в разные часы суток изучали содержание каротиноидов. Результаты были следующие [6]:

Часы суток	Определения									
	1	2	3	4	5	6	7	8	9	10
18	0,60	0,64	0,69	0,52	0,65	0,52	0,58	0,63	0,69	0,48
24	0,61	0,72	0,72	0,58	0,46	0,52	0,72	0,59	0,71	0,79
6	0,58	0,59	0,66	0,46	0,47	0,56	0,70	0,60	0,60	0,64
12	0,70	0,57	0,67	0,87	0,52	0,66	0,68	0,59	0,65	0,66

Изменяется ли содержание каротиноидов в листьях сирени в течение суток?

19. Были получены следующие данные о содержании хлорофилла в листьях томата (в услов. ед.) в различные часы суток [6]:

Часы суток	Определения									
	1	2	3	4	5	6	7	8	9	10
18	0 22	0,27	0,22	0,23	0,20	0,28	0,32	0,28	0,29	0,27
24	0,23	0,24	0,24	0,26	0,23	0,25	0,26	0,28	0,30	0,23
6	0,25	0,23	0,19	0,24	0,20	0,27	0,30	0,27	0,20	0.26
12	0,24	0,28	0,25	0,27	0,22	0,27	0,32	0,30	0,29	0,27

Определите методами дисперсионного анализа, оказывает ли влияние время суток на содержание хлорофилла в листьях томата.

20. Получены следующие данные о содержании хлорофилла (в мг/дм^2) в листьях канатника (*Abutilon*) в разное время суток [6]:

Часы суток	Определения			
	1	2	3	4
15	3,06	2,88	2,83	2,41
18	3,20	2,97	2,50	3,03
21	1,82	1,73	1,33	2,25
24	1,67	1,26	1,52	1,36
6	2,76	1,26	1,46	1,32
9	2,78	2,70	2,49	1,66
12	2,41	3,22	1,90	2,00

Влияет ли время суток на содержание хлорофилла?

21. Были получены следующие данные о содержании *хлорофилла b* в листьях томата (в услов. ед.) в различные часы суток [6]:

Часы суток	Определения									
	1	2	3	4	5	6	7	8	9	10
18	0,087	0,106	0,091	0,097	0,077	0,103	0,118	0,104	0,099	0,108
24	0,086	0,092	0,095	0,100	0,086	0,096	0,101	0,113	0,113	0,092
6	0,093	0,090	0,091	0,089	0,081	0,090	0,118	0,106	0,088	0,096
12	0,088	0,091	0,104	0,113	0,079	0,097	0,114	0,107	0,103	0,093

Определите методами дисперсионного анализа, оказывают ли влияние различные часы суток на содержание *хлорофилла b* в листьях томата.

22. На опытной рисовой станции проводили опыты по изучению влияния удобрений на урожай риса. В 1-й опытной группе применяли удобрение P_2O_5 + N; во 2-й – то же удобрение с предпосевной обработкой почвы. Получены следующие данные (в ц/га) [6]:

Группы	Урожаи на делянках						
Контрольная	35	33	31	37	42	35	40
Опытная 1-я	43	48	54				
Опытная 2-я	36	31	42	36			

Определите методами дисперсионного анализа, оказывают ли влияние на урожайность риса различные методы обработки почвы.

Приложение

ГРЕЧЕСКИЙ АЛФАВИТ

Печатные буквы		Название букв	Рукописные буквы		Стандартный шрифт	
прописные	строчные		прописные	строчные	прописные	строчные
Α	α	альфа	Α	α	Α	α
Β	β	бэта	Β	β	Β	β
Γ	γ	гамма	Γ	γ	Γ	γ
Δ	δ	дэльта	Δ	δ	Δ	δ
Ε	ε	эпсилон	Ε	ε	Ε	ε
Ζ	ζ	дзэта	Ζ	ζ	Ζ	ζ
Η	η	эта	Η	η	Η	η
Θ	θ	тэта	Θ	θ	Θ	θ
Ι	ι	йота	Ι	ι	Ι	ι
Κ	κ	каппа	Κ	κ	Κ	κ
Λ	λ	ламбда	Λ	λ	Λ	λ
Μ	μ	мю	Μ	μ	Μ	μ
Ν	ν	ню	Ν	ν	Ν	ν
Ξ	ξ	кси	Ξ	ξ	Ξ	ξ
Ο	ο	омикрон	Ο	ο	Ο	ο
Π	π	пи	Π	π	Π	π
Ρ	ρ	ро	Ρ	ρ	Ρ	ρ
Σ	σ	сигма	Σ	σ	Σ	σ
Τ	τ	тау	Τ	τ	Τ	τ
Υ	υ	ипсилон	Υ	υ	Υ	υ
Φ	φ	фи	Φ	φ	Φ	φ
Χ	χ	хи	Χ	χ	Χ	χ
Ψ	ψ	пси	Ψ	ψ	Ψ	ψ
Ω	ω	омега	Ω	ω	Ω	ω

Рекомендуемая литература

Основная

Бакай А.В., Кочиш И.И., Скрипниченко Г.Г. **Генетика**. –М.: КолосС, 2006. –408 с.

Ветеринарная генетика. Учебник для высших с.-х., учебн. завед. Специальность – 310800, ветеринария /Петухов В.Л., Жигачев А.И., Назарова Г.А. – 2-е изд. –М.: Колос, 1996. –384 с.

Генетика. Учебник для высш. с.-х. учебн. завед. Специальность – 310700, зоотехния /Меркурьева Е.К., Абрамова З.В., Бакай А.В.,– М.: Агропромиздат, 1991, – 446 с.

Лакин П.Ф. **Биометрия**. –М.: Высшая школа. 1980. –293 с.

Терентьев П.В., Ростова Н.С. **Практикум по биометрии**. Учебное пособие. –Л.: Изд-во ЛГУ, 1977. – 152 с.

Руденко В.И. **Статистика: пособие студ. для подготовки к экзаменам** /В.И. Руденко. –М.: «Дашков и К», 2004. –188 с.

Дополнительная

Гланц С. **Медико-биологическая статистика**. –М.: «Практика», 1999. –460 с.

Жижин К.С. Медицинская статистика: Учебное пособие для ВУЗов – Ростов н/Д: Феникс, 2007. – 160 с.

Мазер К., Джинкс Дж. **Биометрическая генетика**. / Пер. с англ. –М.: Мир, 1985. –463 с.

Меркурьева Е.К. **Биометрия в селекции и генетике сельскохозяйственных животных**. – М.: Колос. 1970. – 424 с.

Рокицкий П.Ф. **Биологическая статистика**. Учебник для вузов. – Минск: «Вышейш. шк.», 1973. – 320 с.

Плохинский Н.А. **Биометрия**. Уч. для вузов.–М.: Изд-во МГУ, 1970. – 368 с.

Учебники по теории статистики

Андронов А.М. Теория вероятностей и математическая статистика. – М.: «Питер», 2004. –460 с.

Гусаров В.М. **Статистика**: **Учеб. пособие для вузов** –М.: ЮНИТИ-ДАНА, 2002. – 463 с.

Едронова В.Н.,. Едронова М.В. **Общая теория статистики**: Учебник. –М.: «Юрист», 2001. – 511 с.

Ефимова М.Р. **Общая теория статистики**: Учебник для вузов /М.Р.

Ефимова, Е.В. Петрова, В.Н. Румянцев. -2-е изд.,испр. и доп. –М.: ИНФРА-М, 2002. – 416 с.

Практикум по теории статистики: /Р.А. Шмойлова, А.Б. Гусынин, В.Г. Минашкин. –М.: Финансы и статистика, 2003. –416 с.

Статистика: **Курс лекций** / Л.П. Харченко, В.Г. Долженкова, В.Г. Ионин, др. –Новосибирск: Изд-во НГАЭиУ; М.: ИНФРА-М, 2000. – 310 с.

Теория статистики с основами теории вероятностей /И. И. Елисеева, В. С. Князевский, Л. И. Ниворожкина, Под ред.И.И.Елисеевой. –М.: ЮНИТИ, 2001. –446 с.

Теория статистики: Учебник / Р.А. Шмойлова, Е.Б. Шувалова, Н.Ю. Глубокова, др.; Под ред.Р.А.Шмойловой. -–М.: Финансы и статистика, 2001.–560 с.

Учебники по статистической обработке данных на ПК

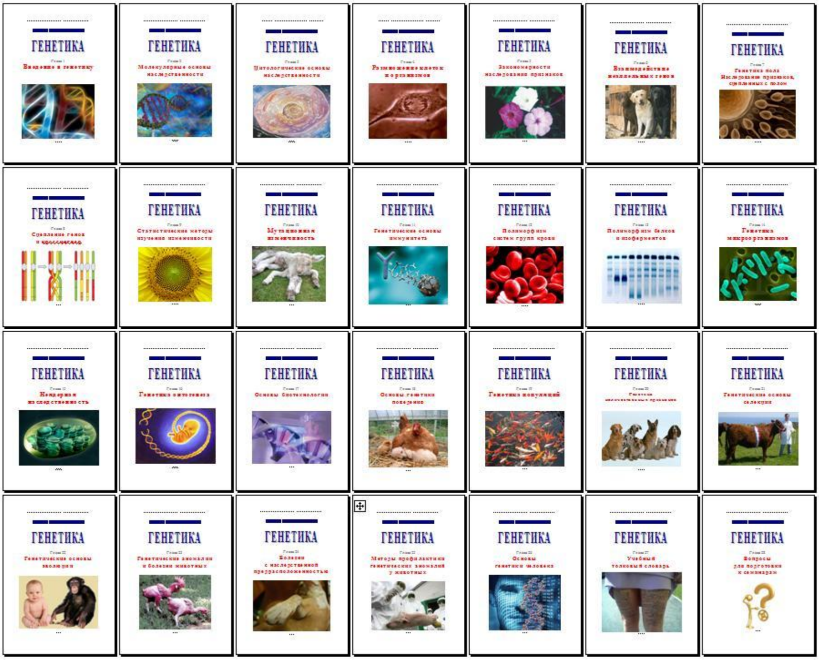
Боровиков В. **Statistica: Искусство анализа данных на компьютере**. Изд. 2-е, дополн. –СПб: «Питер», 2004. –678 с.

Макарова Н.В., Трофимец В.Я. **Exel: Учебное пособие**. –М.: Финансы и статистика, 2002. – 368 с.

Цитированная литература

1. **Ветеринарная генетика**. /Петухов В.Л., Жигачев А.И., Назарова Г.А. – 2-е изд. –М.: Колос, 1996. – 384 с.
2. **Генетика**. / Меркурьева Е.К., Абрамова З.В., Бакай А.В., Кочиш И.И. –М.: Агропромиздат, 1991, – 446 с.
3. Лакин П.Ф. **Биометрия**. –М.: Высшая школа. 1980. – 293 с.
4. Меркурьева Е.К. **Биометрия в животноводстве**. – М.: Колос. 1964. – 311 с.
5. Поляничкин А.А. **Популяционная генетика в птицеводстве**. –М.: Колос. 1980. –271 с.
6. Рокицкий П.Ф. **Биологическая статистика**. Учебник для вузов. – Минск: «Вышейшая школа», 1973. – 320 с.
7. **Статистический словарь** /Гл.ред. М.А.Королёв. –2-е изд. –М.: Финансы и статистика, 1989. –623 с.
8. Урбах В.Ю. **Статистический анализ в биологических и медицинских исследованиях**. –М.: Медицина, 1975. – 295 с.

Методические пособия профессора В.И.Крюкова
можно скачать с сайта www.labogen.ru



Об авторе

Крюков Владимир Иванович, русский, д.б.н., ст.н.с.; 170 опубликованных научных работ (2012), профессор Орловского государственного аграрного

университета (с 2002 г). Читает курсы лекций «Ветеринарная генетика», «Генетика и биометрия». В 2007-2010 гг – директор Инновационного научно-исследовательского центра ОрёлГАУ, с 2011 – заведующий лабораторией экологической генетики этого центра.

Родился в 1956 году в городе Душанбе. В 1977 году с отличием окончил биофак Таджикского государственного университета. После окончания университета направлен на работу в ООРИПР АН ТаджССР.

С 1978 по 1983 – стажёр-исследователь и очный аспирант Института эволюционной морфологии и экологии животных АН СССР им. А.Н. Северцова (г. Москва). В 1983 г. защитил кандидатскую диссертацию «Анализ полиморфизма Т-локуса в популяциях Mus musculus» по двум специальностям – «Генетика» и «Зоология». (автореферат диссертации см. на сайте www.labogen.ru)

С 1983 по 1990 г. научный сотрудник ООРИПР АН ТаджССР. Создал лабораторию экологической генетики. Исследовал генетические последствия антропогенного загрязнения экосистем. В октябре 1990 г. ВАКом присвоено учёное звание «старший научный сотрудник» по специальности «Генетика».

С 1990 по 1999 г. – ст.н.с., зав. лаб. радиобиологии Тульского НИИ сельского хозяйства. В 1995-1998 по совместительству руководил работой лаборатории медицинской генетики Тульского НИИ новых медицинских технологий МЗ РФ. Исследовал влияние электромагнитных полей различных частот на наследственность организмов. В 1999-2002 г. – доцент и профессор медицинского факультета Тульского государственного университета. В 2000 году защитил докторскую диссертацию «Генетический мониторинг антропогенного загрязнения окружающей среды» (диссертацию см. на сайте www.labogen.ru)