П. С. Катмаков, В. П. Гавриленко, А. В. Бушов

БИОМЕТРИЯ

УЧЕБНОЕ ПОСОБИЕ ДЛЯ ВУЗОВ

Под общей редакцией **П. С. Катмакова**

2-е издание, переработанное и дополненное

Допущено УМО вузов РФ по агрономическому образованию в качестве учебного пособия для студентов, обучающихся по направлению подготовки «Технология производства и переработки сельскохозяйственной продукции»

Москва • Юрайт • 2019

УДК 57.087.1(075.8)
ББК 28.0я73
К29

Авторы:

Катмаков Петр Сергеевич — профессор, доктор сельскохозяйственных наук, профессор Ульяновского государственного аграрного университета имени П. А. Столыпина;

Гавриленко Владимир Петрович — профессор, доктор сельскохозяйственных наук, профессор Ульяновского государственного аграрного университета имени П. А. Столыпина;

Бушов Александр Владимирович — профессор, доктор биологических наук, профессор Ульяновского государственного аграрного университета имени П. А. Столыпина.

Рецензенты:

Карамаев С. В. — доктор сельскохозяйственных наук, профессор, заведующий кафедрой зоотехники факультета биотехнологии и ветеринарной медицины Самарской государственной сельскохозяйственной академии;

Каменек В. М. — доктор биологических наук, профессор, заведующий кафедрой генетики человека и экологии Ульяновского государственного университета.

Катмаков, П. С.

К29 Биометрия : учеб. пособие для вузов / П. С. Катмаков, В. П. Гавриленко, А. В. Бушов ; под общ. ред. П. С. Катмакова. — 2-е изд., перераб. и доп. — М. : Издательство Юрайт, 2019. — 177 с. — (Серия : Университеты России).

ISBN 978-5-534-10022-8

В общедоступной форме приведены основные и наиболее распространенные математико-статистические методы, применяемые в биологических исследованиях. Изложены закономерности распределения случайных величин, методика построения вариационных рядов, техника вычисления и оценка выборочных показателей, корреляционный, регрессионный, дисперсионный анализ и другие вопросы биометрии. Приведены примеры, пользуясь которыми даже слабо подготовленный по математике специалист сможет применить в своей работе рекомендуемые способы обработки материалов. Даны сведения, необходимые для вычисления биометрических показателей

Содержание учебного пособия соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования.

Для подготовки студентов, обучающихся по направлениям «Технология производства и переработки сельскохозяйственной продукции», «Зоотехния», «Водные биоресурсы и аквакультура».

УДК 57.087.1(075.8)
ББК 28.0я73

ISBN 978-5-534-10022-8

Оглавление

Список условных обозначений

$\bar{X}$ — средняя арифметическая
m_x, m_σ, m_{cv} — ошибка средних величин ($\bar{X}$, σ, Cv)
t_x, t_σ, t_{cv} — достоверность выборочных показателей
G — средняя геометрическая
S — средняя квадратическая
H — средняя гармоническая
Mo — мода
Me — медиана
C, s^2 — дисперсия
M_S — средний квадрат
σ — среднее квадратическое отклонение
Cv — коэффициент изменчивости
σ^2 — варианса
r — коэффициент корреляции
R, b — коэффициент регрессии
r_s — коэффициент корреляции рангов
η — корреляционное отношение
t — нормированное отклонение
r_a — тетрахорический показатель связи
ρ — полихорический коэффициент связи
r_G — генетический коэффициент корреляции
Ω — множество всех возможных исходов эксперимента (пространство элементарных исходов)
P_A — вероятность наступления события А
q — критерий Ньюмена — Кейлса
q' — критерий Даннета
Δ — определитель матрицы
A — обозначение матрицы
$A^т$ — транспонированная матрица А
A^{-1} — матрица обратная к А
b — вектор
E — единичная матрица
S_b — квадратическая ошибка коэффициента регрессии b,
$\hat{y}$ — обозначение уравнения регрессии
R — коэффициент множественной корреляции
R^2 — коэффициент множественной детерминации

Предисловие

Изучая закономерности раяличных процессов и явлений в живой природе, биологи все чаще обращаются к математике. Для этих целей и служит биометрия, основу которой составляют приемы вариационно-статистического анализа материала. Биометрия основывается на анализе массовых данных, на законе больших чисел и теории вероятностей, которые выявляют закономерности проявления случайных событий на фоне массового материала.

Объектом биометрии служит варьирующий признак, учтенный в группе особей, имеющей достаточную численность и являющейся однородной по ряду других признаков. Например, если исследуют такой основной признак, как продуктивность, то животные, включенные в группу для изучения, должны быть одного вида, возраста, одной породы и находиться в аналогичных условиях кормления и содержания.

Варьирование любого признака у особей однородной группы обусловлено, с одной стороны, различиями в их наследственности, а с другой — влиянием внешних и внутренних факторов. В результате многофакторного воздействия реакция организмов неодинакова, что приводит к индивидуальному варьированию величины признака в пределах даже достаточно однородной группы.

Биометрия дает возможность вычислить средние значения и определить степень их достоверности, вскрыть степень фенотипической и генотипической изменчивости отдельных признаков, выраженную математически, установить степень достоверности опыта, выявить наличие и характер корреляции между различными признаками, а также долю влияния различных факторов на фенотипическую и генотипическую изменчивости признака с использованием дисперсионного и факторного анализа.

При изучении животных по тому или иному признаку не всегда представляется возможным исследовать весь их массив, всю генеральную совокупность. В этомслучае применяется так называемый метод выборочного обследования, который позволяет оценить генеральную совокупность путем отбора меньшей численности животных. Однако при этом выборочная совокупность должна правильно отражать качества и особенности животных, составляющих генеральную совокупность. Такое условие обеспечивается отбором части животных из генеральной совокупности по принципу случайной выборки. Метод случайного отбора членов выборки называется рандомизацией. Она дает равную возможность любому члену генеральной совокупности войти в состав рандомизированной выборки.

Тем не менее следует знать, что биометрический метод, будучи чисто математическим, является только вспомогательным методом, дающим возможность охарактеризовать лишь фактическую сторону явления. Этот метод исследования не вскрывает причин получаемых различий, поэтому использовать его можно только в соединении с глубоким биологическим анализом изучаемых явлений.

В результате изучения дисциплины обучающийся должен:

знать

- правила построения вариационных рядов;
- технику нахождения основных параметров в малой и большой выборках;
- способы графического изображения структуры разнообразия вариационного ряда;
- закономерности соотносительной изменчивости и регрессии для использования их в практической работе;
- основные положения дисперсионного анализа;

уметь

- полученные цифровые материалы обрабатывать по определенной системе и в правильной последовательности;
- графически изображать вариационные ряды;
- применять типы кривых распределения с использованием свойств нормального распределения для решения ряда биологических задач;
- пользоваться техникой дисперсионного анализа и методом хи-квадрат для анализа результатов генетических экспериментов;
- обобщать и статистически обрабатывать результаты экспериментов, формулировать выводы и предложения;

владеть

- статистическими методами обработки экспериментального материала;
- навыками самостоятельной работы с научной литературой.

Глава 1
ПОНЯТИЕ О БИОМЕТРИИ

Биометрия (от греч. *bios* — жизнь, *metron* — мера) — наука о применении математических методов в биологических исследованиях при изучении групповых свойств биологических объектов. Содержанием биометрии является обработка данных наблюдений и экспериментов в биологических исследованиях, а ее методом — теория вероятностей и математическая статистика. Предмет биометрии составляют те или иные варьирующиеся по размеру (изменяющиеся, колеблющиеся в определенных пределах) признаки или свойства объектов совокупности. Биометрию можно использовать при планировании и обработке лишь тех биологических экспериментов и наблюдений, результаты которых могут быть подведены под теоретическое понятие статистической совокупности. Под это понятие можно подвести только те результаты, которые удовлетворяют одновременно двум требованиям:

1) произведено не одно, а несколько наблюдений; минимальное число наблюдений — два;

2) результаты наблюдений (или частоты событий) обладают свойством статистической устойчивости (однородностью).

При проведении эксперимента первое условие выполнить довольно легко. Для этого при изучении группы животных достаточно измерить интересующий признак у каждого представителя группы. Результат каждого измерения записывают и таким образом получают не одно, а несколько значений признака — совокупность.

Выполнение второго условия для полученной совокупности значений признака проверить и обеспечить труднее. Пока нет единственно верного для всех ситуаций научного метода, который позволял бы дать вполне определенный и однозначный ответ на вопрос о наличии свойства статистической устойчивости в результатах измерений. Однако есть несколько требований, выполнение которой при проведении экспериментов обязательно:

1) обеспечение для каждого объекта изучаемой группы одинаковой вероятности быть измеренным, т. е. случайный отбор;

2) исключение возможности попадания в опыт объектов, по существу не принадлежащих к изучаемой группе;

3) независимость отдельных наблюдений друг от друга, т. е. чтобы одно измерение не изменяло объекта и не отражалось на результатах последующих измерений;

4) проведение измерений в одинаковых условиях одинаковыми методами.

Невыполнение любого из этих требований неизбежно приводит к нарушению или полному исчезновению свойства статистической устойчивости результатов, что делает применение биометрии бессмысленным.

1.1. Основные направления применения биометрии в генетике и селекции животных

Внедрение в биологию математических методов, в том числе биометрии и статистики, как специфических разделов современной математики отражает необходимость интенсификации познания в эпоху научно-технического прогресса.

Развитие современного животноводства сопровождается накоплением большого количества информации по многим вопросам общей, прикладной генетики и селекции. В задачу науки входят классификация этих данных, их упорядочение и систематизация, научный анализ, завершающийся формулировкой практических предложений для дальнейшего развития и совершенствования той или иной отрасли животноводства. Необходимы своевременный анализ фенотипической и генетической изменчивости и наследуемости признаков, характеризующих большие массивы животных (породу, стадо), моделирование селекционного процесса в динамике по поколениям и прогноз дальнейшего развития селекционных признаков и увеличения генетического потенциала популяции в целом и отдельных групп животных, составляющих популяцию.

Основные направления применения биометрии в генетике и селекции животных[1]:

- определение степени фенотипического уровня признаков у особей совокупности путем вычисления таких параметров, как средние величины: средняя арифметическая ($\bar{X}$), геометрическая (G), квадратическая (S), гармоническая (H), мода (Mo), медиана (Me);
- определение степени фенотипической и генотипической изменчивости признаков с помощью среднего квадратического отклонения (σ), коэффициентов изменчивости (Cv), варианс (σ^2);
- выявление особенностей и типов варьирования количественных и качественных признаков и характера распределения особей с разным уровнем признаков (нормальное, ассиметричное, эксцессивное, биномиальное, пуассоново, трансгрессивное и др.). Для этого используют уравнения, функции и статистические параметры распределения;
- определение величины фенотипической и генетической корреляции между различными признаками и ее направления с использова-

[1] *Меркурьева Е. К.* [и др.]. Генетика. М. : ВО «Агропромиздат», 1991.

нием коэффициентов корреляции (R), регрессии(R, b), корреляционного отношения (η), ранговых коэффициентов связи (r_s);

- определение доли влияния различных факторов на фенотипическую и генетическую изменчивость признака с использованием дисперсионного и факторного анализа;
- сравнение групп по величине средних, степени изменчивости, по вариансам, частотам, коэффициентам связи, теоретическому и эмпирическому распределению с применением метода статистических ошибок, критерия достоверности Фишера, Стьюдента, метода χ^2 и путем проверки состояния генного равновесия в популяциях и определения генетического расстояния или сходства;
- определение характеристик популяции по комплексу генетических и статистических параметров: степени гомо- и гетерозиготности, генетическому равновесию, коэффициентам наследуемости и постоянства, проявлению гетерозиса и инбредной депрессии; проверка генетических гипотез о типе наследования (доминантности, рецессивности, кодоминантности).

1.2. Понятие о качественных и количественных признаках

Сельскохозяйственные животные обладают большим разнообразием морфологических, физиологических, хозяйственно-полезных признаков. Многие из них имеют значение для практики животноводства, и на улучшение и совершенствование их направлена племенная работа. В то же время большое число признаков не играет практической роли и не служит объектом селекционного воздействия. Все хозяйственно-полезные признаки животных подразделяются на качественные и количественные.

К *качественным* признакам относятся: пол (мужской и женский), окраска шерстного покрова (альбиносность, пигментированность, пятнистость и др.), тип шерстного покрова (грубая, тонкая шерсть овец, смушки), рогатость или комолость, тип нервной деятельности (сильный, неуравновешенный подвижный; сильный, уравновешенный подвижный; сильный уравновешенный инертный; слабый), телосложения (конституция грубая, нежная, рыхлая, плотная, крепкая и др.).

Многие качественные признаки имеют два возможных альтернативных (взаимоисключающих) состояния, например пол мужской или женский; альбиносность — пигментированность; здоровые — больные животные. Некоторые качественные признаки могут иметь три — четыре состояния (тип движения лошади, тип телосложения, тип нервной деятельности). Различия в состоянии качественного признака позволяют разделять животных на группы, несходные по его выраженности. Для характеристики качественных признаков, как правило, достаточно провести глазомерную оценку и дать словесное описание их у конкретного животного или у группы животных.

Количественные, или мерные признаки отличаются тем, что они могут быть измерены и выражены в килограммах, сантиметрах, процентах и т. п. К количественным признакам относятся, например, удой, содержание жира и белка в молоке, живая масса животного, возраст, плодовитость, скорость бега, тонина шерстного волокна и др. Переход от одного количественного уровня признака к другому составляет непрерывный ряд величин.

1.3. Генеральная и выборочная совокупность

Любая целостная группа биологических объектов обладает некоторыми свойствами, которые характеризуют ее в целом, в то время как любой объект, входящий в данную группу, не обладает и не может обладать ни одним из групповых свойств.

Величина признака у отдельной особи называется *вариантой* и обозначается буквами (v) или (x). Значение варианты зависит от многих факторов. Например, суточный удой коровы зависит от генетических факторов, физиологического состояния организма, условий кормления и содержания, климатических факторов и др. При этом возможны многообразные сочетания факторов, положительно и отрицательно влияющих на признак и обусловливающие его разнообразие.

На практике почти никогда не исследуют все объекты изучаемой категории и не измеряют признак у всех объектов, входящих в изучаемую группу. Это обусловлено тем, что группа может быть представлена очень большим числом объектов; измерения требуют больших затрат времени и средств или связаны с нанесением ущерба здоровью и продуктивности животных, а иногда — с уничтожением объекта (например, при изучении действия ионизирующей радиации на животных или изучении особенностей строения и функции внутренних органов животных). Однако самое главное заключается в том, что в настоящее время довольно хорошо разработаны методы, позволяющие по части характеризовать целое, по результатам измерения признака у части объектов, входящих в изучаемую группу, охарактеризовать всю группу в целом с достаточной (для практического использования полученных данных) степенью точности и надежности выводов.

Чтобы освоить эти методы, необходимо ознакомиться с важнейшими понятиями биометрии: генеральной совокупностью и выборочной совокупностью (выборкой).

Генеральная совокупность — это все объекты изучаемой категории, т. е. такая группа объектов, которую нужно описать по тому или иному признаку, по-другому, это — большой массив животных, интересующих исследователя (например, животных бестужевской или другой породы крупного рогатого скота). Конечной целью исследования генеральной совокупности является нахождение параметров (показателей), характеризующих ее свойства. При этом необходимо иметь ясное представление о составе генеральной совокупности. Этого можно достигнуть двумя способами:

1) перечислением всех объектов, входящих в генеральную совокупность (используется редко);

2) указанием некоторых общих для всех объектов свойств, входящих в генеральную совокупность. Второй способ применяется чаще, так как дает возможность, не перечисляя объекты, определить в отношении любого из них, принадлежит ли он к изучаемой совокупности. Например, если целью исследования является характеристика по ряду хозяйственно-полезных признаков (живой массе, настригу шерсти и др.) группы баранов-производителей, использовавшихся в данный случной сезон в хозяйствах данного района, то генеральную совокупность составляют все бараны, которые использовались в этот сезон в хозяйствах района.

Следует также учитывать, что генеральная совокупность может быть представлена объектами, которых еще нет фактически, но которые в принципе могут быть получены. Так, при оценке племенных качеств производителя генеральную совокупность образуют все его возможные потомки. В этом случае любое количество фактически полученных потомков производителя, использующихся для характеристики его племенной ценности, будет лишь частью генеральной совокупности. Это относится и к большинству опытов по изучению степени влияния различных факторов на животных. Например, при изучении влияния на продуктивность или рост животных определенного рациона генеральную совокупность образуют все животные, которые в принципе могли бы получать данный рацион. Группа опытных животных будет представлять лишь часть генеральной совокупности.

Из этих примеров очевидно, что генеральную совокупность, как правило, приходится характеризовать по ее части — так называемой выборке.

Выборочная совокупность (*выборка*) — это часть генеральной совокупности, выделенная по специальным правилам и предназначенная для ее характеристики. Таким образом, изучение выборочной совокупности является не самоцелью, а лишь средством изучения и характеристики генеральной совокупности.

Выборка должна быть *типичной*, т. е. правильно отражать генеральную совокупность. Например, при изучении молочной продуктивности коров нельзя включать в выборку больных животных и животных с атрофией сосков, так как они нетипичны для изучаемой совокупности. Выборка должна быть однородной (одна порода, один пол и т. д.).

Выборка должна быть представительной — *репрезентативной*. Точного метода отбора, гарантирующего ее представительность, не существует. Наиболее желательным является метод случайного отбора, обеспечивающий одинаковую вероятность для каждого объекта генеральной совокупности стать объектом выборки. Рекомендуется пронумеровать все объекты и отобрать некоторые из них по таблице случайных чисел.

Существуют и другие способы отбора, обеспечивающие случайность, например механический — в выборку включается или каждый 10-й, или каждый 20-й объект и т. д.

В некоторых научных экспериментах не всегда имеет смысл прибегать к случайной выборке. При исследовании влияния двух кормов на рост или продуктивность животных можно использовать пары однояйцевых близнецов. В этом случае разница между близнецами каждой пары будет обусловлена именно тем, что их кормили неодинаковыми кормами.

Один из главных параметров выборки — это ее объем (численность), т. е. количество объектов, входящих в изучаемую совокупность. Различают малую и большую выборки. *Большими* называют выборки с численностью 30 особей и более, малыми — численностью менее 30 особей. Объем выборки имеет значение при определении параметров генеральной совокупности по параметрам выборки. Кроме того, различия между малой и большой выборкой сказываются в технике (приемах) расчета основных параметров совокупности, хотя наличие современной счетно-вычислительной техники фактически ликвидировало эту разницу.

При отсутствии счетно-вычислительной техники и наличии многозначных показателей обработка методами больших и малых выборок имеет значение: расчет в больших группах ведется не прямым способом, а путем группировки.

Контрольные вопросы и задания

1. Что такое биометрия?
2. Расскажите об основных направлениях применения биометрии в генетике и селекции животных.
3. Какие признаки называются качественными?
4. Какие признаки называются количественными?
5. Какие качественные признаки могут иметь три — четыре состояния?
6. Что такое генеральная совокупность?
7. Что такое выборки?Как они составляются?
8. Какие выборки называются большими, и какие — малыми?
9. Что является объектом биометрии?
10. Какие совокупности вы знаете?

Глава 2
ОБЩИЕ ПРАВИЛА ПОСТРОЕНИЯ ВАРИАЦИОННЫХ РЯДОВ

Собранный экспериментальный материал обрабатывают в следующем порядке. Предположим, что мы собрали данные о промерах высоты в холке 100 коров бестуженской породы и получили следующие значения.

124	124	129	135	130	127	126	132	124	125
133	127	131	136	125	131	128	127	137	126
122	128	134	122	132	128	129	123	126	125
128	127	130	125	129	122	126	127	119	132
125	127	130	129	134	128	134	132	132	126
134	131	135	128	125	128	133	123	121	125
128	126	129	127	134	129	126	123	124	128
130	126	131	135	129	130	125	130	121	123
137	127	128	129	135	132	135	129	124	130
131	133	132	127	134	127	136	123	128	133

Каждая такая цифра называется вариантой (величина признака у отдельной особи) и обозначается буквой (v). Общее число вариант обозначается буквой (n).

Для построения вариационного ряда необходимо выполнить следующее.

1. Подсчитать количество вариант (n), входящих в данный вариационный ряд (в нашем примере $n = 100$);

2. Просмотрев весь материал, найти минимальное и максимальное значения признака (высоты в холке бестужевских коров) для определения размаха изменчивости или лимита ($\lim = \max V - \min V$). Для нашего примера наибольшая варианта — $\max v = 137$ см, минимальная — $\min v = 119$ см; размах изменчивости — $\lim = 137 - 119 = 18$ см;

3. Установить величину интервала (классного промежутка), который обозначается буквой k (иногда i).

Классный промежуток — это величина, на которую отличается значение одного класса от значения другого (соседнего). Количество классов определяется произвольно. Оно зависит от желаемой степени точности всех расчетов и от объема выборки. Чем больше число классов, на которые разбит вариационный ряд, тем выше будет точность вычисленных с его помощью параметров. Чем больше численность совокупности (выборки), тем на большее число классов следует ее разбить.

В зависимости от объема выборки (n) рекомендуется иметь следующее число классов (l) (табл. 2.1).

Таблица 2.1

Рекомендованное число классов (l)

Объем выборки n, до	Рекомендованное число классов l
40—60	6—8
60—100	7—10
100—200	9—12
200—500 и более	12—17

В приведенном примере n (100) находится между 60 и 100, чему соответствует 7—10 классов. Для нахождения значения классного промежутка в этом случае желательно взять 10 классов, тогда

$$k=\frac{\lim}{L}=\frac{\max V-\min V}{\text{число классов}}=\frac{137-119}{10}=1{,}8\approx 2\,\text{см}.$$

Если значение классного промежутка — дробное число (в данном примере 1,8 см), то ее следует округлить до ближайшего целого числа, лучше в бóльшую сторону (2,0 см).

4. Построить расчетную таблицу (вариационную решетку).

5. Установить границы классов. Классы обозначаются буквой W. За нижнюю (первую) границу первого класса обычно принимают минимальное значение признака в выборке (119). Первые границы каждого последующего класса определяют путем прибавления к значению первой границы предыдущего класса величины классного промежутка (до получения величины максимального значения признака). Верхние (вторые) границы каждого класса, в зависимости от того, в каких единицах выражается признак, берут или на 1, или на 0,1, или на 0,01 меньше первых границ последующих классов. Затем следует определить срединные значения классов ($W_{ср}$). Срединное (центральное) значение каждого класса определяется как полусумма нижних границ данного и следующего классов, можно также для этого к нижней границе данного класса прибавлять половину классного промежутка.

В нашем примере $W_{ср}$ первого класса составит $\frac{119+121}{2}=120$ или первая граница класса $+\frac{k}{2}=119+1=120$.

Достаточно рассчитать срединные значения только одного первого класса, срединные значения остальных определяются последовательным прибавлением величины классного промежутка K:$W_{ср}$ второго класса будет равным 120 + 2 = 122 см, третьего — 122 + 2 = 124 см и т. д.

6. Варианты изучаемой совокупности (v) разносят по классам. Разноску удобнее проводить с помощью точек и «конвертов», например:

1 · 2 ·· 3 ∴ 4 ∷ 5 ⁙ 6 ⊓ 7 ⊐ 8 ⊔ 9 ⊠ 10 ☒

После распределения всех вариант подсчитывают их количество в каждом классе, т. е. частоту классов, обозначаемую буквой P. В результате наш вариационный ряд принимает вид, показанный в табл. 2.2.

Таблица 2.2

Вариационный ряд по величине высоты в холке коров бестужевской породы

W	Границы классов									
	119—120	121—122	123—124	125—126	127—128	129—130	131—132	133—134	135—136	137—138
Центральное значение класса $W_{ср}$	120	122	124	126	128	130	132	134	136	138
Разноска	·	⁙	☒	☒⊓	☒☒·	☒⊓	☒··	☒	⊐	··
Частота вариант P	1	5	10	16	21	16	12	10	7	2

В таком виде вариационный ряд готов к математической обработке с целью нахождения нужных параметров совокупности.

Вариационные ряды можно изобразить графически в виде *гистограммы или линейной кривой (полигон распределения).* Для этого, используя систему координат, строят график: на горизонтальной оси (ось абсцисс) откладывают первые границы классов, на вертикальной (ось ординат) — частоты. Изобразив частоты каждого класса в виде столбиков, получают ступенчатую фигуру, называемую гистограммой. Во втором случае при пересечении перпендикуляров, восстановленных из значений середины классов с горизонтальными линиями, проведенными из соответствующих им частот, ставят точки, которые затем соединяют ломаной линией, называемой вариационной кривой (рис. 2.1).

Ступенчатость гистограммы и ломаный вид вариационной кривой объясняются небольшим количеством наблюдений в вариационном ряду. Если же число наблюдений будет стремиться к бесконечности, то вариационная кривая приобретает плавный характер и превращается в теоретическую, характеризующую распределение членов генеральной совокупности с теоретическим значением частот. При этом вариационная кривая по всей форме напоминает биномиальную кривую.

Следовательно, распределение частот у биологических объектов характеризуется нормальным распределением. Изменчивость биологических признаков не хаотична, носит не случайный характер, а вполне закономерна. Чем ближе варианты к средней арифметической, тем чаще они встречаются, и наоборот, чем больше уклоняются варианты от средней арифметической, тем реже они встречаются в генеральной совокупности.

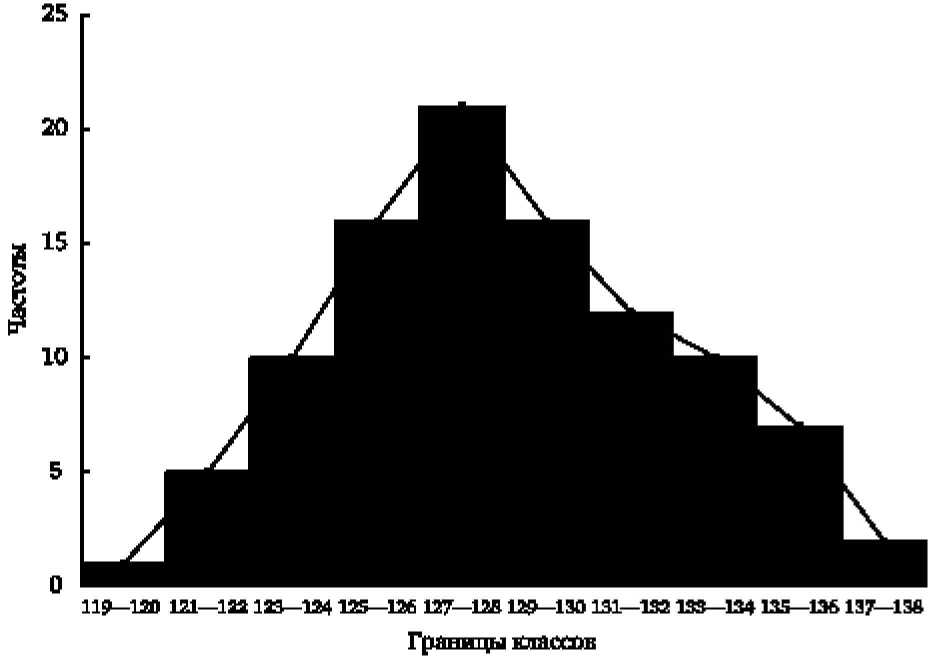

Рис. 2.1. **Графическое изображение вариационного ряда (гистограмма и линейная кривая распределения 100 коров бестужевской породы по промерам высоты в холке)**

Задания для самостоятельной работы

Задание 1. Составить вариационный ряд и изобразить его графически в виде гистограммы и полигона распределения (вариационной кривой), пользуясь следующими данными.

Масса парной туши, кг, при убое свиней с живой массой 100 составила:

62,9	63,1	64,7	64,6	63,9	62,3	64,4	67,5	66,3	64,8
56,0	64,0	65,9	65,7	65,0	66,9	65,0	66,2	68,0	65,1
65,3	65,7	62,4	66,2	64,2	64,5	67,5	70,3	62,8	66,3
65,3	65,1	65,0	68,8	62,0	65,3	61,8	63,1	67,4	63,2
66,2	66,5	60,0	66,7	61,8	59,3	66,5	63,5	65,7	66,2
66,4	65,6	62,2	67,0	66,9	62,6	64,6	67,6	66,2	58,7
56,0	64,0	65,9	65,7	65,0	66,9	65,0	66,2	68,0	65,1
65,3	65,7	62,4	66,2	64,2	64,5	67,5	70,3	62,8	66,3
65,3	65,1	65,0	68,8	62,0	65,3	61,8	63,1	67,4	63,2
65,3	65,1	65,0	68,8	62,0	65,3	61,8	63,1	67,4	63,2
66,2	66,5	60,0	66,7	61,8	59,3	66,5	63,5	65,7	66,2
66,4	65,6	62,2	67,0	66,9	62,6	64,6	67,6	66,2	58,7

Задание 2. По приведенным ниже данным о живой массе, кг, коров швицкой породы составить вариационный ряд и изобразить его графически в виде гистограммы и линейной кривой:

529	497	530	500	545	436	565	515	495	481
500	520	562	518	552	550	479	487	491	505
495	501	493	507	523	557	545	470	509	515
529	504	452	535	535	559	469	493	527	530
490	541	556	485	514	511	521	527	543	510
547	529	538	475	483	583	487	497	520	505
518	472	520	539	507	512	465	518	538	515
541	510	527	515	524	480	531	462	517	478
518	472	520	539	507	512	465	518	538	515
500	520	562	518	552	550	479	487	491	505
490	541	556	485	514	511	521	527	543	510

Задание 3. Составить вариационный ряд и изобразить его графически по данным, касающимся удоя коров симментальской породы за 305 дней третьей лактации, кг:

3684	3946	4028	2636	5194	4578	4190	4786	3652	3828
4183	3645	4292	4362	4202	4313	3862	4467	4166	5443
4269	4640	4624	4929	5017	5332	4285	4200	4179	4436
3298	4276	4532	4954	3972	3276	3822	4880	3507	3102
5292	4744	3869	3788	4682	3845	3623	4402	4123	4601
4175	4812	2915	4195	3992	4224	4616	3386	3592	4081
4764	3575	3413	4326	4281	4539	3674	5780	4379	4424

Задание 4. Составить вариационный ряд и изобразить его графически по данным живой массы бычков симментальской породы при рождении, кг:

45	47	44	36	56	45	40	33	45	46
46	48	39	40	40	45	49	45	50	40
44	49	39	45	45	37	47	52	60	34
44	45	50	53	38	44	40	38	43	41
42	37	47	31	51	48	50	48	50	46
37	44	41	48	40	41	41	36	36	38
44	52	46	61	40	38	38	45	46	40
45	46	32	55	45	45	40	37	53	50
48	45	32	36	38	39	47	38	37	40

Задание 5. Составить вариационный ряд и изобразить его графически со следующими данными промера обхвата груди у коров красной горбатовской породы, см:

194	186	185	182	186	185	194	179	187	180
172	180	181	171	171	167	190	196	183	191
186	178	189	177	172	191	174	185	180	176
189	177	173	187	180	191	203	190	192	186
184	172	179	189	174	167	172	183	186	174
193	195	188	179	197	185	196	183	180	180
173	182	190	168	182	191	184	186	184	178
187	178	180	173	180	173	178	172	170	179
176	182	185	180	178	180	191	188	176	184
187	178	180	173	180	173	178	172	170	179
173	182	190	168	182	191	184	186	184	178
194	186	185	182	186	185	194	179	187	180

Задание 6. Составить вариационный ряд и изобразить его графически по данным яйценоскости кур за месяц, шт.:

22	26	25	27	26	24	27	20	25	22	25
26	25	17	24	19	27	24	25	23	25	23
23	22	26	24	29	24	23	22	23	27	27
28	17	26	24	30	27	22	27	25	27	22
26	21	24	20	26	24	24	25	24	13	25
26	22	26	20	30	23	24	24	22	22	26
19	26	26	24	25	22	24	24	25	27	22

Задание 7. Составить вариационный ряд и изобразить его графически по данным среднего процента жира за лактацию у коров костромской породы, %:

4,02	4,31	3,61	4,01	4,40	3,75	4,01	4,51	4,05	3,71
4,01	4,05	4,28	3,91	4,21	4,02	3,30	3,80	3,92	4,25
4,01	4,27	3,95	4,28	3,95	4,26	3,81	4,01	4,29	3,82
3,83	4,11	3,85	4,21	3,90	4,10	4,05	4,15	3,86	4,16
3,96	4,18	3,83	4,20	4,12	4,15	4,13	4,15	4,05	3,99
4,05	4,01	4,11	3,92	4,12	3,95	4,05	4,01	4,05	4,01
4,12	4,03	4,12	4,05	4,03	4,12	4,05	4,12	4,03	4,02
4,01	4,13	4,05	4,03	4,02	4,03	4,02	3,46	3,59	4,02

Задание 8. Составить вариационный ряд и изобразить его графически по данным суточных удоев коров бестужевской породы, кг:

21,9	21,4	27,7	17,7	12,3	21,8	23,4	25,7	21,2	20,2
23,8	24,1	26,9	21,4	20,7	18,5	22,5	23,0	18,5	25,7
20,1	21,3	15,7	24,8	19,3	22,2	22,9	14,9	26,1	20,5
14,6	27,8	22,4	16,7	22,9	25,3	22,7	19,7	15,2	21,3
22,1	20,5	19,7	24,5	29,6	22,3	19,1	23,5	25,9	17,2
15,5	18,1	23,9	25,4	20,4	13,2	19,6	24,4	18,2	24,8
24,2	20,9	20,1	16,5	20,9	23,2	27,2	21,1	26,3	18,6
17,2	17,8	31,2	25,0	20,7	18,3	23,7	16,1	16,2	21,6
23,0	20,7	25,3	13,9	17,3	21,8	14,1	19,0	21,9	18,7
28,5	21,2	19,9	24,8	22,7	16,4	20,6	23,5	22,2	19,5

Задание 9. Составить вариационный ряд и изобразить его графически по данным настрига шерсти у овец асканийской породы, кг:

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

Задание 10. Составить вариационный ряд и изобразить его графически по данным плодовитости свиноматок крупной белой породы по второму опоросу, голов:

11	12	14	10	13	13	11	12	10	12
11	14	12	10	11	8	13	13	12	12
16	12	13	14	10	11	11	12	16	13
10	11	14	11	13	10	12	12	13	9
18	12	10	13	12	14	14	11	12	10
15	13	13	16	12	12	12	10	13	13
13	14	12	14	14	14	14	14	12	12
14	12	14	12	15	13	11	12	14	10
12	16	10	12	16	10	13	10	9	11
14	11	11	11	14	11	10	13	11	13

Задание 11. Составить вариационный ряд и изобразить его графически по данным времени наступления овуляции у коров бестужевской породы, ч от начала охоты:

12	15	20	13	19	20	17	25	22	24
25	21	30	26	26	27	28	30	26	26
28	30	27	26	28	30	35	31	33	34
35	32	33	31	40	45	50	37	27	15
38	39	28	40	21	18	26	20	34	16
22	31	27	33	28	33	24	16	28	18
18	28	33	30	20	19	18	20	19	34
19	22	27	26	19	24	36	30	26	16
29	31	19	23	31	29	40	19	30	19

Задание 12. Составить вариационный ряд и изобразить его графически по данным чисел пульсовых ударов у свиней:

65	70	68	73	80	68	70	75	69	66
70	75	73	72	68	74	69	78	67	69
73	69	73	68	73	70	68	74	76	79
68	68	60	68	69	70	74	73	75	60
65	66	70	71	73	74	75	67	69	68
71	72	71	69	70	66	75	66	69	68
67	69	75	74	66	65	70	71	71	77
69	68	69	66	69	72	68	69	66	69
71	74	70	77	71	70	66	67	69	73
65	72	73	69	74	69	71	70	68	75

Задание 13. Составить вариационный ряд и изобразить его графически по данным удоя коров красной степной породы за 305 дней первой лактации, кг:

4208	2438	4204	4110	5398	3557	3851	4540	3953	4552
4220	3205	4921	3540	3311	4155	3380	4160	3240	4510
4985	5421	4087	5260	3160	4525	3555	3210	3440	5200
3185	5635	2030	3270	3822	3138	3540	3850	2625	4625
3870	3250	2245	4760	4187	3290	2755	5000	5438	4270
3662	3141	3990	3360	3260	3214	2960	4591	5215	3377
3919	3660	4110	3843	4055	3550	3168	4600	4616	4135
4101	3266	3553	4105	3664	3943	4210	3977	4220	4368
3344	3195	3470	3691	2998	3619	4050	3278	3843	3930

Задание 14. Составить вариационный ряд и изобразить его графически в виде гистограммы и линейной кривой по данным живой массы коров черно-пестрой породы, кг:

380	427	372	382	390	380	384	397	398	400
403	438	432	480	489	470	450	400	398	420
370	378	480	473	450	478	455	519	450	540
451	444	586	500	466	462	515	501	537	529
582	554	551	544	500	513	549	511	534	584
391	395	460	531	509	540	390	500	510	560
405	430	553	510	480	515	410	490	543	549
463	463	484	473	399	471	451	462	515	574
500	510	395	490	548	490	486	473	524	590
530	494	460	504	503	408	508	430	390	433

Задание 15. Составить вариационный ряд и изобразить его графически по данным промера обхвата груди у кобыл буденновской породы, см:

164	180	159	163	165	165	160	160	172	170
190	160	155	172	190	150	165	150	188	170
170	188	180	180	173	189	182	180	158	158
180	173	189	191	161	158	172	177	187	169
185	170	180	183	166	156	180	166	170	163
173	176	159	178	170	163	169	171	162	168
168	168	165	163	185	168	178	182	166	181
181	184	176	169	190	177	172	158	179	159

Контрольные вопросы

1. Что такое вариационный ряд?
2. Как составляют вариационный ряд?
3. Какими способами можно графически изобразить вариационные ряды?
4. Как определяют величину классового промежутка?
5. Как устанавливают границы классов?
6. Как устанавливают число классов?
7. Как находят средние (центральные) значения классов?
8. Как проводят разноску вариант по классам?
9. Как находят размах изменчивости признака?
10. Что называют вариантой, варьированием?

Глава 3
ВЫЧИСЛЕНИЕ СРЕДНИХ ВЕЛИЧИН

Средние величины — важные биометрические показатели, используемые в науке и практике. Имеется несколько средних величин: средняя арифметическая, средняя квадратическая, средняя геометрическая, средняя гармоническая, мода, медиана и др.

3.1. Вычисление средней арифметической в малочисленных выборках

Средняя арифметическая — показатель средней величины признака данной группы особей, характеризующий среднюю вариацию этого признака. Средняя арифметическая величина именованная; ее значение выражается в тех же единицах измерений, что и признак вариационного ряда (в сантиметрах, граммах, килограммах и т. д.). Средняя арифметическая — абстрактное число. Если среднее поголовье овец на фермах сельскохозяйственных предприятий района оказалось равным 415,4, то такое число точно характеризует среднюю величину овцефермы, хотя в действительности существование 0,4 овцы невозможно. Средняя арифметическая величина в малочисленных выборках вычисляется прямым способом, который заключается в суммировании всех вариант $(x_1+x_2+x_3+...+x_n)$ с последующим делением суммы на число вариант в совокупности (n):

$$\bar{X}=\frac{x_1+x_2+x_3+...+x_n}{n}, \text{ или } \bar{X}=\frac{\sum X_n}{n},$$

где $\bar{X}$ — средняя арифметическая; x — величина вариант, n — число вариант, $\sum$ — знак суммирования (сумма).

Например, живая масса у отдельных коров составляет, кг: 620, 598, 606, 550, 583, 611,

$$\bar{X}=\frac{620+598+606+550+583+611}{6}=\frac{3568}{6}=594{,}6.$$

Следует обратить внимание на одно из важных свойств средней арифметической. Отдельные варианты отклоняются от средней. Одни из них больше средней (т. е. имеют положительные отклонения), другие — меньше (их отклонения отрицательные). Но сумма положительных и отрицательных отклонений всегда равна нулю.

Задания для самостоятельной работы

Задание 1. Живая масса поросят свиноматки № 1 при рождении (крупноплодность) составляла, кг: 1,2; 1,5; 1,0; 1,3; 1,4; 1,3; 0,9; 1,4; 1,3, а поросят свиноматки № 2, кг: 1,2; 1,3; 1,0; 0,8; 1,3; 0,9; 1,0; 1,1; 1,2; 1,0. Вычислить отдельно среднюю живую массу поросят свиноматки № 1 и № 2.

Задание 2. Суточные привесы в группе телят бестужевской породы составляли, г: 667, 521, 644, 443, 759, 576, 820, 691, 487, 722. Вычислить среднюю арифметическую.

Задание 3. В группе из 10 коров черно-пестрой породы суточные удои отдельных животных составляли, кг молока: 12, 16, 20, 24, 28, 19, 14, 16, 18, 17. Вычислить среднюю арифметическую для изученной группы.

Задание 4. Живая масса у 10 коров симментальской породы составляет, кг: 620, 598, 606, 550, 583, 611, 660, 585, 608, 623. Вычислить среднюю арифметическую.

Задание 5. Живая масса у 20 быков бестужевской породы в пятилетнем возрасте составляет, кг: 920, 806, 784, 761, 855, 820, 781, 768, 800, 950, 780, 1092, 930, 786, 905, 767, 770, 923, 820, 782. Определить среднюю живую массу быков.

Задание 6. Масса парной туши при убое 15 свиней крупной белой породы составила, кг: 56; 64; 65,9; 65,7; 65; 66,9; 66,2; 68; 65,1; 66,4; 65,6; 62,2; 67; 60; 61,8. Вычислить среднюю арифметическую.

Задание 7. Время наступления овуляции у 10 коров бестужевской породы составила,ч от наступления охоты: 12, 15, 20, 13, 19, 20, 17, 25, 22, 24. Определить среднюю арифметическую.

Задание 8. Число пульсовых ударов у 15 свиней беркширской породы составила: 65, 70, 68, 73, 80, 68, 70, 75, 69, 66, 70, 75, 73, 78, 67. Вычислить среднюю арифметическую.

Задание 9. Плодовитость свиноматок крупной белой породы по второму опоросу составила: 11, 12, 14, 10, 13, 13, 11, 12, 10, 8, 12, 10, 11, 14, 16. Определить среднюю плодовитость свиноматок.

Задание 10. Настриг шерсти у 15 овец асканийской породы составила, кг: 8,5; 8,2; 9,6; 9,2; 7,5; 8,2; 9; 8,4; 9,5; 9,4; 7; 8; 8,5; 9,4; 9,7. Вычислить среднюю арифметическую.

Задание 11. Величина обхвата груди коров бестужевской породы по третьей лактации составила, см: 194, 186, 185, 179, 190, 186, 179, 180, 187, 192, 185, 189, 176, 180, 186. Вычислить среднюю арифметическую.

Задание 12. Удой за 305 дней лактации у 15 коров бестужевской породы в ОПХ «Тимирязевское» составил, кг: 4203, 2438, 4204, 4110, 5398, 3557, 3851, 4540, 3958, 4552, 4220, 3205, 4921, 3540, 3311. Вычислить среднюю арифметическую.

Задание 13. Средняя жирность молока у коров черно-пестрой породы за вторую лактацию составила, %: 3,4; 3,8; 3,8; 3,7; 3,4; 3,3; 3,4; 3,5; 3,8; 3,1; 3,4; 4,1; 3,5; 3,6; 3,2. Вычислить среднюю арифметическую.

Задание 14. Живая масса коров симментальской породы при бонитировке в племзаводе «Родина» Вешкаймского района по первой лактации составила (кг): 580, 427, 582, 554, 551, 544, 500, 513, 549, 511, 534, 584, 444, 466, 515. Определить среднюю живую массу коров.

Задание 15. Удой за 305 дней лактации укоров черно-пестрой породы по данным бонитировки составил, кг: 3900, 4230, 4080, 2660, 3356, 3923, 4806, 3693, 3246, 2948, 3649, 4092, 4837, 3674, 3441. Вычислить среднюю арифметическую.

3.2. Вычисление взвешенной средней арифметической

В том случае, когда варианты имеют неодинаковый математический вес, для определения среднего значения показателя предварительно вносится поправка (именуемая весом варианты) и вычисляется взвешенное среднее арифметическое.

Средняя взвешенная представляет собой результат усреднения средних арифметических нескольких совокупностей. Она вычисляется по формуле

$$\overline{X}_{\text{взв}} = \frac{\overline{X}_1 \cdot n_1 + \overline{X}_2 \cdot n_2 + \ldots + \overline{X}_i \cdot n_i}{n_1 + n_2 + \ldots + n_i} = \frac{\sum \overline{X}_i \cdot n_i}{\sum n_i},$$

где $\overline{X}_{\text{взв}}$ — средняя взвешенная; $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_i$ — средние арифметические первой, второй совокупностей и т. д.; n_1, $n_2, \ldots$, n_i — вес (объем) этих совокупностей.

В качестве примера возьмем вычисление среднего процента жира в молоке коровы за лактацию. За средний процент жира в период лактации нельзя взять среднее арифметическое из ежемесячных показателей жирности молока. Здесь отдельные варианты имеют разный математический вес, так как определенному проценту жира за каждый месяц лактации соответствует разное количество надоенного за этот период молока, которое и служит мерой веса каждой варианты ряда.

В этом случае взвешенное среднее арифметическое вычисляется следующим образом: $\overline{X}_{\text{взв}} = \frac{\sum X_n}{n}$, где X — значение варьирующего признака (процент жира за каждый месяц лактации); n — математический вес этого признака (удой за каждый месяц лактации).

Таблица 3.1

Порядок вычисления среднего процента жира в молоке за лактацию

Показатели	Месяцы лактации									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Процент жира x, %	4,0	4,1	4,2	4,2	4,3	4,3	4,4	4,6	4,5	4,7
Удой за месяц, кг (n)	757	748	620	638	666	609	524	501	472	414
Однопроцентное молоко ($x \cdot n$)	3028	3067	2604	2680	2864	2620	2306	2305	2124	1946

$$\sum n = 5949; \sum x \cdot n = 25544; \bar{X}_{\text{вз}} = \frac{25544}{5949} = 4{,}29\%.$$

Для определения среднего процента жира в молоке за лактацию (табл. 3.1) необходимо общее количество однопроцентного молока (сумма произведений содержания жира по месяцам на месячные удои) разделить на общий удой молока за лактацию.

Приведем еще пример. Имеются средние арифметические величины живой массы поросят в двухмесячном возрасте от отдельных свиноматок одного семейства. Необходимо вычислить среднюю живую массу поросенка по этому семейству. Здесь взвешенное среднее арифметическое вычисляется следующим способом (табл. 3.2):

$$\bar{X}_{\text{вз}} = \frac{\sum \bar{X}_i \cdot n_i}{\sum n_i},$$

где $\bar{X}$ — среднее арифметическое для отдельных групп; n — численность особей в группах.

Таблица 3.2

Определение средней взвешенной живой массы поросенка

Индивидуальный номер свиноматок	Средняя живая масса поросят $\bar{X}$ в двухмесячном возрасте, кг	Количество поросят n	$\bar{X} \cdot n$
604	15,3	39	596,7
88	16,1	29	466,9
371	13,1	23	301,3
1004	13,6	19	258,4

$$\bar{X}_{\text{вз}} = \frac{1623{,}3}{110} = 14{,}8 \text{ кг}; \sum n = 110; \sum x \times n = 1623{,}3$$

Задания для самостоятельной работы

Задание 1. В хозяйстве от трех быков-производителей получено 58 дочерей. Показатели продуктивности их следующие: средний удой 20 дочерей Ветерка — 3250 кг за лактацию при жирности молока равной 4,3 %; средний удой 23 дочерей Метеора — 4115 кг, жирность молока — 3,8 %; средний удой 15 дочерей Грома — 2756 кг, содержание жира — 4,53 %. Определить среднее содержание жира в молоке дочерей всех трех производителей. Потомство какого из трех быков-производителей является наиболее продуктивным?

Задание 2. На трех птицефермах насчитывалось 3000, 1500 и 7900 несушек, причем за год ими было снесено соответственно 214 500, 14 835, 1 185 000 яиц. Вычислить среднюю взвешенную по трем птицефермам вместе.

Задание 3. Известны средняя живая масса и число коров в трех хозяйствах. Они составляли в первом хозяйстве $\overline{X}_1 = 420$ кг, $n_1 = 1000$ голов; во втором — $\overline{X}_2 = 460$ кг, $n_2 = 500$ голов; в третьем — $\overline{X}_3 = 520$ кг, $n_3 = 2000$ голов. Вычислить среднюю живую массу коров по данным всех трех хозяйств.

Задание 4. Средняя тонина шерсти, мк, в трех отарах овец хозяйства составляла 20, 25, 19, а количество животных — соответственно 2100, 2400, 1000. Вычислить взвешенную среднюю арифметическую тонины шерсти.

Задание 5. В 100 кг кормовой смеси содержится следующее количество отдельных кормов, кг: сено — 50, резаная солома — 10, подсолнечниковый жмых — 20, пшеничные отруби — 20 с содержанием переваримого протеина, %, соответственно 3, 1, 33, 11. Определить содержание протеина в смеси путем вычисления взвешенной средней величины.

Задание 6. Средние арифметические ($\overline{X}$) показатели тонины шерсти в отдельных стадах овец цигайской породы госплемзавода СХК «Волга» Цильнинского района составляют 20, 22, 25, 18, 24 мк, а количество овец в этих стадах — 210, 150, 240, 100, 300 голов. Вычислить взвешенную среднюю арифметическую тонины шерсти всех овец.

Задание 7. В четырех хозяйствах Цильнинского района средний удой коров черно-пестрой породы за третью лактацию составил 3621, 4025, 3391, 4112 кг, а численность поголовья — 450, 613, 722, 540 голов. Вычислить взвешенную среднюю арифметическую удоев коров данных хозяйств.

Задание 8. Вычислить взвешенную среднюю арифметическую для годовой яйценоскости кур четырех птицеферм Ульяновского района, если на них насчитывалось 5432, 6233, 4242, 7252 несушек, при их средней яйценоскости 151, 147, 200, 172 яиц.

Задание 9. Средняя живая масса коров, кг, черно-пестрой породы по первой лактации в пяти хозяйствах Ульяновского района составила 554, 515, 478, 500, 466 кг, а число коров в них — соответственно 350, 420,600, 390, 550 голов. Определить взвешенную среднюю арифметическую живой массы коров по данным хозяйствам.

Задание 10. Средний удой коров бестужевской породы в четырех хозяйствах Ульяновского района был равным 4203, 3557, 4110, 2438 кг, а численность коров в них — 380, 427, 480, 400 голов. Вычислить взвешенный средний удой коров по данным хозяйствам.

Задание 11. Величина обхвата груди коров бестужевской породы в трех племенных хозяйствах по данным бонитировки составила, см: 194, 186, 180, а поголовье коров в данных хозяйствах, голов: 580, 450, 500. Вычислить взвешенную среднюю арифметическую обхвата груди коров.

Задание 12. В опытно-производственном хозяйстве Ульяновского НИИСХ средний удой 15 дочерей быка-производителя Клена по первой лактации был равен 4215 кг при жирности молока 3,82 %, средний удой 22 дочерей Грача — 3880 кг с жирностью 4,12 %, средний удой дочерей Юга — 4550 кг с содержанием жира в молоке 3,76 %. Вычислить среднюю взвешенную по удою и содержанию жира в молоке дочерей всех трех быков-производителей.

Задание 13. Средняя плодовитость свиноматок крупной белой породы по второму опоросу в пяти хозяйствах Чердаклинского района составила (голов): 11, 12, 9, 8, 13, а поголовье свиноматок, голов: 60, 110, 80, 90, 50. Определить взвешенную среднюю арифметическую плодовитости свиноматок.

Задание 14. В четырех хозяйствах Николаевского района средний настриг шерсти у овец куйбышевской породы был равным, кг: 8,5; 6,4; 7,2; 6,8, а поголовье овец в них — соответственно, голов: 600, 450, 760, 800. Вычислить средний настриг шерсти овец по данным всех четырех хозяйств.

Задание 15. Средняя живая масса молодняка свиней крупной белой породы в возрасте четырех месяцев в пяти хозяйствах Ульяновского района составила, кг: 43, 32, 49, 41, 50, а их поголовье, голов — 200, 160, 320, 240, 300. Определить среднюю живую массу молодняка свиней по пяти хозяйствам вместе.

3.3. Вычисление средней величины для неизмеряемых признаков (непараметрическая средняя)

Многие признаки не имеют количественного измерения (интенсивность окраски шкурок цветного каракуля, норок и др.). По степени интенсивности развития признака животные могут быть ранжированы в порядке усиления или ослабления выраженности признака. Порядковый номер животного называется рангом.

Например,от двух баранов-производителей (№ 2 и № 3) каракульской породы и группы отобранных маток получено по восемь серых ягнят с различной интенсивностью окраски (от светлой до темно-серой). Следует выяснить, какой из производителей дает потомство с более темной мастью. Все потомки обоих баранов-производителей распределены в ранжированный ряд от светло-серой до темно-серой окраски шерсти с указанием номера отца.

Ранги	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Номер отца	5	6	5	6	5	5	6	5	5	5	6	5	6	6	6	6

На основании полученного ряда определяют средний ранг каждого производителя:

$$\bar{X}_2 = \frac{1+3+5+6+8+9+10+12}{8} = \frac{54}{8} = 6{,}5;$$

$$\bar{X}_3 = \frac{2+4+7+11+13+14+15+16}{8} = 10{,}2.$$

Второй производитель (№ 3) имел больше ягнят с темно-серой окраской, которая ценится дороже.

Задания для самостоятельной работы

Задание 1. Вычислить средний ранг (непараметрическую среднюю) для двух самцов норок № 201 и № 202 голубой окраски.

От каждого самца и группы самок получено 20 щенков с различной окраской меха, от почти белого до темно-голубого. Какой самец дает потомство с более темной окраской меха?

Составлен следующий ранжированный ряд потомков производителей в порядке усиления серого цвета:

Ранг	№ отца	Ранг	№ отца	Ранг	№ отца
1	201	8	201	15	202
2	202	9	201	16	202
3	201	10	201	17	201
4	202	11	202	18	202
5	201	12	201	19	202
6	201	13	202	20	202
7	202	14	202		

3.4. Вычисление средней геометрической

При планировании интенсивности роста, прироста популяции или продуктивности по определенным периодам, т. е. при определении темпа изменения признака используется средняя геометрическая *G*:

$$G=\sqrt[n]{x_1\cdot x_2\cdot x_3 \ldots x_n},$$

где n — число членов в выборке; $x_1 \ldots x_n$ — значения вариант.

Для использования указанной формулы проводят ее логарифмирование:

$$\lg G=\frac{\lg x_1+\lg x_2+\lg x_3+\ldots+\lg x_n}{n}.$$

Зная $\lg G$, по логарифмическим таблицам находят соответствующую ему величину G.

Средняя геометрическая наиболее правильно отражает фенотипический средний уровень признака, особенно если распределение членов выборки проявляет отклонение от нормального, например при ассиметричном распределении, т. е. когда частоты (P_i) смещены в крайние классы ряда.

Например, для пяти дат (1; 4; 5; 5; 5) среднюю геометрическую можно получить следующим образом:

x	1	4	5	5	5
$\lg x$	0,000	0,602	0,699	0,699	0,699

$$\sum \lg x=2{,}699;\ \lg G=\frac{\sum \lg x}{n}=\frac{2{,}699}{5}=0{,}5398;\ G=3{,}467.$$

3.5. Вычисление средней квадратической

Для определения средней квадратической применяют формулу $S=\sqrt{\frac{\sum x^2}{n}}$. Средняя квадратическая применима при вычислении площади, длины окружности или объема шара (величина эритроцита, жирового шарика, ядра клетки).

Например, необходимо определить среднюю окружность основания соска вымени пяти коров, если известен обхват соска каждой из них, см: 6; 5; 7; 8; 6. Используя приведенную выше формулу, вычислим:

$$S=\sqrt{\frac{6^2+5^2+7^2+8^2+6^2}{5}}=\sqrt{\frac{210}{5}}=\sqrt{42}=6{,}481\text{ см},$$

в то время как $\overline{X}$ дает несколько меньшую величину: (6 + 5 + 7 + 8 + + 6) : 5 = 32 : 5 = 6,40.

Приведем еще пример. Определили средний диаметр пяти зигот до начала их дробления и получили такие данные, мкм: 60; 70; 58; 65; 75. Находим, что

$$S=\sqrt{\frac{60^2+70^2+58^2+65^2+75^2}{5}}=\sqrt{\frac{21714}{5}}=\sqrt{4342{,}8}=65{,}9\text{ мкм}.$$

3.6. Вычисление средней гармонической

Среднюю гармоническую применяют при вычислении среднего уровня признака, характеризующего скорость какого-либо процесса (средняя скорость бега, интенсивность молокоотдачи при доении), а также в случае, если признак выражен индексом (число шерстинок на 1 мм^2 поверхности кожи). Для ее определения используют следующую формулу:

$$H=\frac{n}{\sum\frac{1}{x_i}}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}},$$

где $x_1, x_2 \ldots x_n$ — варианты признака; n — число периодов (отрезков времени).

Например, необходимо определить среднюю скорость молокоотдачи у коровы, если за 5 мин выдоено 12,5 кг молока, в том числе за первую минуту — 5 кг, вторую — 4 кг, третью — 2 кг, четвертую — 1 кг и за пятую минуту — 0,5 кг. Используя указанную формулу, получим:

$$H=\frac{5}{\frac{1}{5}+\frac{1}{4}+\frac{1}{2}+\frac{1}{1}+\frac{1}{0,5}}=\frac{5}{0,20+0,25+0,5+1,0+2}=\frac{5}{3,95}=1,26 \text{ кг/мин.}$$

Если же вычислить простую среднюю арифметическую, то скорость молокоотдачи составит $\overline{X}=(5+4+2+1+0,5):5=12,5:5=2,5$ кг/мин, т. е. полученная величина недостоверна.

3.7. Определение моды и медианы

Модой называют наиболее часто встречающуюся варианту в вариационном ряду. Класс, в котором находится мода, называют модальным. В вариационном ряду может быть несколько модальных классов. Мода может быть вычислена с помощью формулы

$$Mo=W_O+K\left(\frac{P_2-P_1}{2\cdot P_2-P_1-P_3}\right),$$

где W_O — нижняя граница модального класса; K — величина классного промежутка; P_1 — частота класса, предшествующего модальному; P_2 — частота модального класса; P_3 — частота класса, следующего за модальным. Подставляя данные из таблицы распределения величины высоты в холке коров бестужевской породы в формулу, получим

$$M_O=127+2\left(\frac{21-16}{2\cdot 21-16-16}\right)=127+2\cdot\frac{5}{10}=127+1=128 \text{ см.}$$

Медианой называют середину класса, который делит вариационный ряд на две части: одна имеет значение признака меньшее, чем медиана, другая — большее.

Вычисление медианы в больших выборках при неравномерном распределении вариантов по классам проводят по формуле

$$M_e = W_O + K\left(\frac{\frac{n}{2}-P_1}{P}\right),$$

где W_O — начало класса, в котором находится медиана; n — общее числовариант в группе; P_1 — сумма частот классов, предшествующих классу, где находится медиана; P — частота класса, в котором находится медиана. В данном примере (распределение величины высоты в холке 100 коров бестужевской породы) медиана равна:

$$Me = 127+2\left(\frac{\frac{100}{2}-32}{21}\right) = 127+2\cdot\frac{18}{21} = 127+2\cdot 0{,}85 = 128{,}7 \text{ см}.$$

3.8. Вычисление средней арифметической в больших выборках

Прямой метод вычисления $\bar{X}$ при большом числе вариант при отсутствии вычислительной техники требует много труда. Поэтому при биометрической обработке многочисленных выборок используются другие (непрямые) методы, в частности способ произведений, или «условной средней». В основу положено следующее свойство средних величин: алгебраическая сумма положительных и отрицательных отклонений отдельных вариантов от средней арифметической величины (сумма центральных отклонений) всегда равна нулю, т. е. $(v-\bar{X})=0$. При этом способе для вычисления средней арифметической величины используются вариационные ряды. Вычисление производится по формуле

$$\bar{X} = A + b\cdot k \text{ или } \bar{X} = A + \frac{\sum p\cdot a}{n}\cdot k,$$

где A — условная средняя (центральное значение модального класса); b — поправка к условной средней арифметической $(b=\frac{\sum p\cdot a}{n})$; k — величина классного промежутка.

Первый пример. Для вычисления средней арифметической величины по промерам высоты в холке коров бестужевской породы желательно использовать готовый вариационный ряд, составленный по значениям данных промеров (табл. 3.3) и добавить две графы под рубрикой — отклонения каждого класса от модального a и $p\times a$.

Затем нужно выделить условно средний (модальный) класс. За модальный класс, как правило, принимается класс с наибольшим количеством частот вариант. Модальный класс следует выделить путем заштриховки. Если в вариационном ряду имеется несколько классов с одинаковым максимальным значением частот, то в качестве модального берется один из них, расположенный к середине ряда.

Таблица 3.3

Вычисление средней арифметической величины промеров высоты в холке коров бестужевской породы

Границы классов	Частота варианта p	Отклонения a	$p \times a$	W
119—120	1	–4	–4	119,5
121—122	5	–3	–15	121,5
123—124	10	–2	–20	123,5
125—126	16	–1	–16	125,5
127—128	21	0	0	127,5
129—130	16	1	16	129,5
131—132	12	2	24	131,5
133—134	10	3	30	133,5
135—136	7	4	28	135,5
137—138	2	5	10	137,5
	$\sum p = 100$		$\sum p \times a = 53$	

Определяют условную среднюю арифметическую (A). За условную среднюю величину принимается середина модального класса, поскольку искомая $\overline{X}$ скорее всего будет близка к нему (в случае с прерывным варьированием — плодовитость свиноматки, яйценоскость и т. д.) или к начальной границе (A_1) модального класса прибавляют половину значения классного промежутка (в случае с непрерывным варьированием): $A = A_1 + \frac{K}{2}$. Значение классного промежутка (k) берется не округленное.

В нашем примере наибольшее количество частот вариант (p = 21) имеет пятый класс с границами 127—128 см, тогда

$$A = 127 + \frac{1{,}8}{2} = 127{,}9\,\text{см}.$$

Далее, приняв модальный класс за нулевой, находят отклонение срединного значения каждого класса от модального, выраженное в единицах классного промежутка. Эти отклонения обозначаются буквой a. Классы, расположенные в таблице вверх от модального класса, будут меньше него на один, два, три и т. д. классных промежутка. Аналогичные отклонения будут в классах, стоящих ниже от условного среднего, но только с положительным знаком.

Записав отклонения с их знаками в третью графу таблицы, умножают отклонения каждого класса (a) на соответствующую частоту (p) и произведения ($p \times a$) вписывают в четвертую графу таблицы. После этого суммируют все значения $p \times a$ с учетом их знака: сначала все положительные ($+p \times a$), затем все отрицательные ($-p \times a$) и вычитают из большей суммы меньшую, сохраняя знак большей величины.

Если бы условная средняя (A) совпадала с истинной средней ($\bar{X}$), то $\sum p \times a$ с минусом равнялась бы $\sum p \times a$ с плюсом, а их общая сумма была бы равна нулю. В нашем примере $\sum p \times a$ с минусом равняется –55, а $\sum p \times a$ с плюсом — 108, их алгебраическая сумма $\sum p \times a = 108 - 55 = 53$. Следовательно, условная средняя в какой-то мере отличается от истинной средней. Это несовпадение корректируется поправкой, обозначаемой буквой b.

Вычисляют поправку к условной средней арифметической по формуле $b = \frac{\sum p \times a}{n}$. Поправка ($b$) может быть как с положительным знаком, так и с отрицательным. Это зависит от знака $\sum p \times a$. Для нашего примера $b = \frac{53}{100} = 0{,}53$. Прибавив к условной средней поправку, получают истинную среднюю арифметическую:

$$\bar{X} = A + b \times k = 127{,}9 + 0{,}53 \times 1{,}8 = 128{,}8 \text{ см.}$$

Второй пример. Вычислить среднее арифметическое содержания жира в молоке у коров симментальской породы по следующим данным, %:

3,72	3,51	3,29	3,12	3,67	4,09	3,70	3,86	4,23	3,92	3,99	4,09	3,88	3,95	3,94
3,56	3,32	3,31	3,44	3,75	3,90	4,35	4,56	3,60	4,00	4,36	3,68	3,87	4,23	3,71
3,87	4,13	3,81	3,63	3,56	3,90	3,84	3,72	4,20	3,67	4,46	3,57	4,05	3,59	5,02
3,34	3,68	3,51	3,93	4,00	3,93	3,99	3,93	3,83	4,04	4,20	4,00	4,16	4,33	4,30
3,92	3,59	3,89	3,55	4,14	4,58	4,25	3,88	4,15	3,99	4,00	3,46	3,67	3,97	4,37

Методика определения $\bar{X}$ остается прежней с той лишь разницей, что в данном случае классный промежуток будет меньше единицы (табл. 3.4).

Таблица 3.4

Вычисление средней арифметической величины содержания жира в молоке коров

Классы W	Частоты P	Отклонения a	$p \times a$
3,10—3,29	2	–4	–8
3,30—3,49	5	–3	–15
3,50—3,69	15	–2	–30
3,70—3,89	14	–1	–14

Окончание табл. 3.4

Классы W	Частоты P	Отклонения a	p×a
3,90—4,09	21	0	0
4,10—4,29	9	1	9
4,30—4,49	6	2	12
4,50—4,69	2	3	6
4,70—4,89	0	4	0
4,90—5,09	1	5	5
	$\sum p=75$		$\sum p\times a=-35$

1. Подсчитывают количество животных в совокупности или число вариант: $n = 75$.

2. Находят максимальное (max) и минимальное (min) значения признака животных в исследуемой группе:

$$\max V=5{,}02;\ \min V=3{,}12.$$

3. По разности между максимальным и минимальным значением признака определяют лимит:

$$\lim=5{,}02-3{,}12=1{,}9.$$

4. Определяют величину классного промежутка:

$$K=\frac{\lim}{L}=\frac{1{,}9}{10}=0{,}19\approx 0{,}2.$$

5. Строят вариационную решттку.

6. Устанавливают границы классов.

7. Разносят варианты выборки по классам.

8. В вариационном ряду выделяют модальный (условно средний) класс.

9. Выделяют условную среднюю арифметическую:

$$A=3{,}90+\frac{0{,}2}{2}=3{,}90+0{,}1=4{,}0\,\%.$$

10. Определяют условные отклонения каждого класса от модального (a).

11. Подсчитывают сумму всех произведений:

$$\sum p\cdot a=(-67)+32=-35.$$

12. Вычисляют поправку к условной средней арифметической:

$$b=\frac{\sum p\cdot a}{n}=\frac{-35}{75}=-0{,}46.$$

13. Находят среднее арифметическое величины изучаемого признака: $\overline{X} = A + b \cdot k = 4{,}0 + (-0{,}46 \cdot 0{,}2) = 4{,}0 - 0{,}09 = 3{,}91$ % жира в молоке.

Задание для самостоятельной работы

Вычислить среднюю арифметическую в больших выборках, пользуясь данными заданий 1—15 главы 2.

Контрольные вопросы и задания

1. Какими свойствами обладают средние величины?
2. Перечислите средние величины. Как их используют?
3. Как вычисляется средняя арифметическая для малой выборки и для большой выборки?
4. Что такое средняя взвешенная? В каких случаях она применяется? Как ее вычислять?
5. Как вычисляют непараметрическую среднюю?
6. Как вычисляют среднюю геометрическую?
7. Как вычисляют среднюю квадратическую?
8. Как вычисляют среднюю гармоническую?
9. Что такое мода и медиана?
10. Как находят условную среднюю арифметическую?
11. Как находят поправочную величину к условной средней арифметической?
12. Что такое модальный класс? Как его находят?
13. Что показывает средняя арифметическая?
14. Как определяют отклонения класса от модального?
15. Может ли быть $\sum p \cdot a$ со знаком минус? Почему?

Глава 4
ПОКАЗАТЕЛИ РАЗНООБРАЗИЯ ПРИЗНАКОВ В СОВОКУПНОСТИ

Для правильного представления о развитии признака в данном вариационном ряду недостаточно определения одного среднего арифметического, так как особи в двух рядах с одинаковыми средними значениями признака могут значительно отличаться по степени изменчивости признака. Это свойство состоит в том, что совокупность образуют неодинаковые по величине признака объекты. Не следует путать разнообразие с изменчивостью. Изменчивость — биологическое явление, она всегда выступает как причина разнообразия животных в исследуемой группе. Изучается она в специальных опытах путем оценки степени разнообразия той или иной совокупности объектов. Однако не во всех опытах степень разнообразия является показателем изменчивости.

Оценка степени разнообразия животных в изучаемых группах — важнейший элемент селекционной работы. Большое значение придается ей в связи с переводом животноводства на промышленную технологию для обеспечения выпуска стандартной продукции.

Разнообразие объектов в совокупности характеризуется целым рядом показателей:

1) лимиты (lim), или пределы: $\lim = V_{max} - V_{min}$. Размах разнообразия определяются крайними вариантами, т. е. максимальным и минимальным значением признака в совокупности;

2) дисперсия C, или сумма квадратов центральных отклонений: $C = \sum(V - \overline{X})^2$, т. е.сумма квадратов разности между значением каждого отдельного варианта (V) и средней арифметической величиной ($\overline{X}$);

3) варианса (σ^2) — сигма в квадрате, или средний квадрат центральных отклонений. Вычисляется по формуле $\sigma^2 = \dfrac{C}{N-1}$, где $N-1$ — число степеней свободы ν (ν — греческая буква «ню»). Сумму квадратов центральных отклонений делят не на объем совокупности (N), а на число объектов, не соответствующих средней арифметической ($N-1$), так как считается, что один из объектов представлен средней арифметической величиной;

4) среднее квадратическое отклонение от средней арифметической величины, или сигма (σ). Сигма является основным показателем разнообразия значений признака в группе и используется для определения

целого ряда других параметров совокупности (коэффициента вариации, ошибки средней арифметической величины и др.), а также характеризует распределение объектов в совокупности: $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{C}{N-1}}$;

5) коэффициент вариации $C_v = \frac{\sigma}{\bar{X}} \cdot 100$, который показывает, какую часть сигма составляет от средней арифметической величины. Выражается в долях единицы или в процентах, т. е. является относительной величиной, что позволяет использовать его для сравнения степени разнообразия различных признаков и разных групп;

6) нормированное отклонение (t). Оно является показателем разнообразия признака отдельной варианты и представляет выраженное в долях сигмы взвешенное отклонение соответствующей варианты от средней арифметической: $t = \frac{V - \bar{X}}{\sigma}$.

Наиболее простой способ установления изменчивости признака в совокупности — определение минимальной (X_{min}) и максимальной (X_{max}) величин вариантов (лимит). Чем больше абсолютная разность между ($X_{max} - X_{min}$), тем более значительна изменчивость признака. Так, если содержание жира в молоке помесных коров колеблется от 3,3 до 4,5 % (lim = 1,2 %), а в молоке коров черно-пестрой породы — от 2,8 до 3,6 % (lim = 0,8 %), то это значит, что в группе помесных животных степень изменчивости жирномолочности выше. Однако лимит неточно отражает степень варьирования признака. Значения отдельных случайных вариант могут быть очень большими или очень маленькими, что сильно отражается на лимите, в то время как на закономерность изменчивости эти варианты влияют незначительно. И наоборот, ряды бывают с одинаковым или близким значением лимита, но с различной степенью варьирования, т. е. лимиты не отражают степень разнообразия внутри группы.

Например, изменчивость живой массы поросят при отъеме (кг):

1-я группа: 11, 12, 13, 15, 18, 22, 26, 28; ($\bar{X}$) = 18 кг, lim = 28 − 11 = 17;

2-я группа: 11,5; 18; 19; 19,5; 20; 20,5; 21; 26,5; ($\bar{X}$) = 19 кг, lim = 26,5 − 11,5 = 15.

В таком случае пользуются вычислением среднего квадратического отклонения.

4.1. Вычисление среднего квадратического отклонения в малочисленных выборках (n < 30)

Установление степени разнообразия признака в популяциях имеет большое значение в селекции. Наилучшим показателем разнообразия признака является среднее квадратическое отклонение σ, которое учитывает отклонение каждой варианты от средней арифметической. Чем больше σ, тем больше изменчивость признака в группе животных.

Сигма — величина именованная и выражается в тех же единицах, что и средняя арифметическая. Между средней арифметической и средним квадратическим отклонением существует закономерная связь, которая выражается в следующих правилах: в генеральной совокупности в пределах $\bar{X} \pm 1\sigma$ находится 68,3 % вариант совокупности, в пределах $\bar{X} \pm 2\sigma$ — 95,5 %, а в пределах $\bar{X} \pm 3\sigma$ — 99,7 %, или практически все варианты.

При небольшом числе вариант среднее квадратическое отклонение вычисляется по формуле

$$\sigma = \pm\sqrt{\frac{\sum(V-\bar{X})^2}{n-1}},$$

где V — варианта; $\bar{X}$ — средняя арифметическая; n —число вариант в совокупности.

Пример. Вычислить среднее квадратическое отклонение по данным о живой массе 20 быков бестужевской породы в возрасте 5 лет и старше (табл. 4.1).

$$\bar{X} = 840;\ \sum(V-\bar{X}) = 0;\ \sum(V-\bar{X})^2 = 143150.$$

Для получения необходимых величин для расчетаσ значения признака у животных совокупности вносят в вспомогательную таблицу. В первую графу вписывают варианты (живая масса быков). Суммировав их и разделив на число вариант, получают среднюю живую массу быка ($\bar{X}$): $\bar{X} = \frac{\sum V_i}{N} = 840$ кг. Затем нужно вычесть $\bar{X}$ из каждой варианты и разности ($V-\bar{X}$), т. е. отклонения вариант от средней, вписать во вторую графу.

Таблица 4.1

Вычисление среднего квадратического отклонения прямым способом (при малом числе вариант)

Живая масса быков V, кг	Отклонения $(V-\bar{X})$	Квадраты отклонений $(V-\bar{X})^2$	Живая масса быков V, кг	Отклонения $(V-\bar{X})$	Квадраты отклонений $(V-\bar{X})^2$
920	80	6400	780	–60	3600
806	–34	1156	1092	252	63 504
784	–56	3136	930	–90	8100
761	–79	6241	786	–54	2916
855	15	225	905	65	4225
820	–20	400	767	–73	5329
781	–59	3481	770	–70	4900
768	–72	5184	923	83	6889
800	–40	1600	820	–20	400
950	110	12 100	782	–58	3364

Для проверки правильности вычислений суммируют все разности $(V-\bar{X})$, их сумма должна быть равна нулю. Далее каждое отклонение возводят в квадрат и вписывают квадраты отклонений $(V-\bar{X})^2$ в третью графу. В отличие от отклонений, которые могут быть положительными и отрицательными, квадраты отклонений всегда положительны. Суммируя все показатели третьей графы, получают сумму квадратов отклонений $\sum(V-\bar{X})^2$, которую вписывают в итоге третьей графы.

В нашем примере $\sigma=\pm\sqrt{\frac{\sum(V-\bar{X})^2}{n-1}}=\sqrt{\frac{143\,150}{19}}=86{,}79\text{ кг}$.

Сигма является показателем разнообразия признака. Согласно правилу трех сигм, почти все варианты укладываются в интервал от -3σ до $+3\sigma$. В данном примере живая масса быков в генеральной совокупности должен находиться между $840-3\cdot86{,}79$ и $840+3\cdot86{,}79$, т. е. между 579,63 и 1100,37 кг, что соответствует действительности.

Задания для самостоятельной работы

Задание 1. Вычислить среднее квадратическое отклонение в малых выборках, пользуясь данными заданий 1—15 главы 3 (см. параграф 3.1).

4.2. Вычисление среднего квадратического отклонения в больших выборках

При вычислении среднего квадратического отклонения для многочисленных выборок составляется вариационный ряд и вычисление производится по формуле

$$\sigma=\pm k\sqrt{\frac{\sum p\cdot a^2}{n}-b^2}\ \text{или}\ \sigma=\pm k\sqrt{\frac{\sum p\cdot a^2}{n}-\left(\frac{\sum p\cdot a}{n}\right)^2},$$

где k — величина классного промежутка; p — частоты; a — отклонения от условного среднего класса, выраженные в числе классных промежутков; n — число вариант в выборке; Σ — знак суммирования; b — поправка к условной средней арифметической.

Вычисление среднего квадратического отклонения производится аналогично вычислению средней арифметической. Для этого требуются те же величины, что и для вычисления $\bar{X}$, дополнительно необходимо произвести определение b^2 и $\sum p\cdot a^2$. Поэтому к записанным уже при вычислении $\bar{X}$ столбцам частот (p), отклонений (a), их произведений ($p\cdot a$) следующим столбцом записываются произведения частот на квадраты отклонения ($p\cdot a^2$). Затем производится их суммирование, т. е. определяется $\sum p\cdot a^2$.

Произведем вычисление среднего квадратического отклонения (табл. 4.2) на примере, по которому уже вычисляли среднюю арифметическую (промер высоты в холке чистопородных коров бестуженской породы, см. главу 2).

Таблица 4.2

Вычисление среднего квадратического отклонения промера высоты в холке 100 коров бестуженской породы

Классы	p	a	$p\cdot a$	$p\cdot a^2$
119—120	1	–4	–4	16
121—122	5	–3	–15	45
123—124	10	–2	–20	40
125—126	16	–1	–16	16
127—128	21	0	0	0
129—130	16	1	16	16
131—132	12	2	24	48
133—134	10	3	30	90
135—136	7	4	28	112
137—138	2	5	10	50
	$\sum p=100$		$\sum p\cdot a=53$	$\sum p\cdot a^2=433$

Подставив вычисленные величины в формулу, получим:

$$\sigma=\pm k\sqrt{\frac{\sum p\cdot a^2}{n}-\left(\frac{\sum p\cdot a}{n}\right)^2}=1{,}8\cdot\sqrt{\frac{433}{100}-\left(\frac{53}{100}\right)^2}=$$

$$=1{,}8\cdot\sqrt{4{,}33}-0{,}2809=1{,}8\cdot\sqrt{4{,}0491}=1{,}8\cdot 2{,}012=\pm 3{,}62\text{ см.}$$

Как видно, среднее квадратическое отклонение нашего вариационного ряда равно $\pm 3{,}62$ см.

Крайние значения (лимиты) в генеральной совокупности будут находиться в пределах $\bar{X}\pm 3\sigma$, а в данном примере:

$$\bar{X}+3\sigma=128{,}8+3\cdot 3{,}62=128{,}8+10{,}86=139{,}66\text{ см;}$$

$$\bar{X}-3\sigma=128{,}8-3\cdot 3{,}62=128{,}8-10{,}86=117{,}94\text{ см.}$$

Расчет среднего квадратического отклонения на примере вариационного ряда по содержанию жира в молоке коров бестуженской породы (табл. 4,3).

$$\sigma=\pm k\sqrt{\frac{\sum p\cdot a^2}{n}-\left(\frac{\sum p\cdot a}{n}\right)^2}=0{,}2\cdot\sqrt{\frac{227}{75}-\left(\frac{-35}{75}\right)^2}=$$

$$=0{,}2\sqrt{3{,}03-(0{,}46)^2}=\pm 0{,}33\,\%;$$

$$\bar{X}+\sigma=3{,}91\pm 0{,}33\,\%.$$

Таблица 4.3

Вариационный ряд по содержанию жира в молоке коров

Классы	p	a	$p \cdot a$	$p \cdot a^2$
3,10—3,29	2	–4	–8	32
3,30—3,49	5	–3	–15	45
3,50—3,69	15	–2	–30	60
3,70—3,89	14	–1	–14	14
3,90—4,09	21	0	0	0
4,10—4,29	9	1	9	9
4,30—4,49	6	2	12	24
4,50—4,69	2	3	6	18
4,70—4,89	0	4	0	0
4,90—5,09	1	5	5	25
	$\sum p = 75$		$\sum p \cdot a = -35$	$\sum p \cdot a^2 = 227$

Итак, среднее квадратическое отклонение вариационного ряда по содержанию жира в молоке бестужевских коров равно ±0,33 %.

Задания для самостоятельной работы

Задание 1. Вычислить среднее квадратическое отклонение в больших выборках, пользуясь данными заданий 1—15 главы 2.

4.3. Вычисление среднего квадратического отклонения для альтернативных признаков

Показатель разнообразия для альтернативных признаков определяют с помощью среднего квадратического отклонения в абсолютных и относительных выражениях по формуле $\sigma = \sqrt{p \cdot g}$ или $\sigma = \sqrt{p \cdot (1-p)}$ и $\sigma = \sqrt{n \cdot p \cdot g}$, где p — доля особей, имеющих данный признак в совокупности; g — доля особей, лишенных данного признака, n — общее поголовье.

Например, требуется определить величину среднего квадратического отклонения по показателям наличия у животных желательного типа при разведении помесей, полученных при скрещивании грубошерстных овец с тонкорунными баранами. Из 1000 голов стада 650 животных было желательного, а 350 — нежелательного типа. Отсюда соотношения животных желательного и нежелательного типа составят:

$$p = \frac{650}{1000} = 0{,}65; \; g = \frac{350}{1000} = 0{,}35.$$

Правильность расчетов проверяют по формуле $p+g=1$. В данном примере $0,65+0,35=1$.

Среднее квадратическое отклонение будет равно:

$$\sigma=\sqrt{p\cdot g}=\sqrt{0,65\cdot 0,35}=\sqrt{0,2275}=0,476, \text{ или } 47,6\%;$$

$$\sigma=\sqrt{1000\cdot 0,65\cdot 0,35}=\sqrt{227,5}=15 \text{ голов.}$$

Задания для самостоятельной работы

Задание 1. Определить величину среднего квадратического отклонения по показателю коричневой окраски у норок, если у 60 зверей из 85 окраска была коричневой.

Задание 2. В отаре каракульских овец насчитывалось 745 черных и 245 серых животных. Определить величину среднего квадратического отклонения по показателю наличия животных серой окраски.

Задание 3. В хозяйстве было 1500 норок с жемчужной окраской и 2100 — с коричневой. Определить величину среднего квадратического отклонения по жемчужной окраске.

4.4. Вычисление коэффициента вариации

По средним квадратическим отклонениям сравнивают изменчивость двух или большего числа групп животных только по одноименным признакам и при условии, если средние $\overline{X}$ этих признаков у сравниваемых групп имеют не очень большую разность. Среднее квадратическое отклонение непригодно для сравнения степени варьирования разноименных признаков как внутри группы, так и при сопоставлении разных групп. Например, необходимо сравнить, по какому из признаков наиболее изменчиво стадо коров: по живой массе, удою за лактацию, содержанию жира в молоке или высоте в холке (табл. 4.4).

Средние квадратические отклонения (σ) этих признаков следующие: живой массы — 50 кг, удой за лактацию — 760 кг, содержание жира в молоке — 0,32 %, высоты в холке — 4,1 см. Чтобы сравнить степень изменчивости разноименных признаков, применяется коэффициент вариации C_v который вычисляется по формуле

$$C_v=\frac{\sigma}{\overline{X}}\cdot 100\,\%$$

и представляет величину σ, выраженную в процентах к $\overline{X}$.

Коэффициент вариации — величина отвлеченная и поэтому удобная для сравнения изменчивости различных признаков. Вычислим C_v для перечисленных выше признаков.

Таблица 4.4

Изменчивость основных хозяйственно-полезных признаков коров стада

Показатели	$\bar{X}$	σ	C_v, %
Живая масса, кг	509	50	9,8
Удой за лактацию, кг	4200	760	18,1
Содержание жира в молоке, %	4,03	0,32	7,9
Высота в холке, см	137	4,1	3,0

Коэффициент изменчивости будет равен:

- для живой массы $C_v = \frac{50 \cdot 100}{509} = 9{,}8\,\%$;
- для удоя коров за лактацию $C_v = \frac{760 \cdot 100}{4200} = 18{,}1\,\%$;
- для процента жира в молоке $C_v = \frac{0{,}32 \cdot 100}{4{,}03} = 7{,}9\,\%$;
- для высоты коров в холке $C_v = \frac{4{,}1 \cdot 100}{137} = 3{,}0\,\%$.

Коэффициенты вариации показывают, что данное стадо наиболее изменчиво по удою за лактацию и наименее — по высоте в холке.

Коэффициентом вариации удобно пользоваться также при сравнении степени варьирования одинаковых признаков у разных групп, если между средними ($\bar{X}_1 - \bar{X}_2$) этих признаков разница очень большая. Так, сравнивая стадо, средний удой которого равен 2000 кг (σ = 500 кг), со стадом при среднем удое 4000 кг (σ = 800 кг), находим, что для первого стада $C = \frac{500 \cdot 100}{2000} = 25\,\%$, а для второго — $C = \frac{800 \cdot 100}{4000} = 20\,\%$.

Больший коэффициент вариации удоев коров первого стада указывает на бóльшую их разнородность, хотя среднее квадратическое отклонение удоя коров первого стада, наоборот, было меньше, чем во втором стаде.

Для определения степени изменчивости признака количественное выражение коэффициента вариации распределяют на три группы: при C_v, равном приблизительно 10 %, степень изменчивости признака считается слабой; при C_v, составляющем приблизительно 15 %, — средней; если значение C_v будет выше 15 %, то степень вариации признака считается сильной.

Задания для самостоятельной работы

Вычислить коэффициенты вариации для больших и малых выборок, пользуясь данными заданий 1—15 (см. главы 2 и 3) и определить какой из сравниваемых признаков наиболее изменчив.

Задание 1. Вычислить C_v в больших выборках (см. главу 2) для:

а) массы парной туши свиней по данным задания № 1;

б) живой массы коров швицкой породы по данным задания № 2;

в) удоя коров симментальской породы за 305 дней третьей лактации по данным задания № 3;

г) живой массы бычков симментальской породы при рождении по данным задания № 4;

д) промера обхвата груди коров красной горбатовской породы по данным задания № 5;

е) яйценоскости кур за месяц по данным задания № 6;

ж) среднего процента жира за лактацию у коров костромской породы по данным задания № 7;

з) суточных удоев коров бестужевской породы по данным задания № 8;

и) настрига шерсти у овец асканийской породыпо данным задания № 9;

к) плодовитости свиноматок крупной белой породы по второму опоросу по данным задания № 10;

л) времени наступления овуляции у коров бестужевской породы по данным задания № 11;

м) чисел пульсовых ударов у свиней по данным задания № 12;

н) удоя коров красной степной породы за 305 дней первой лактации по данным задания № 13;

о) живой массы коров черно-пестрой породы по данным задания № 14;

п) промера обхвата груди у кобыл буденовской породы по данным задания № 15.

Задание 2. Вычислить C_v в малых выборках (см. главу 3) для:

а) живой массы поросят свиноматки № 1 и свиноматки № 2 по данным задания № 1;

б) суточных привесов в группе телят бестужевской породы по данным задания № 2;

в) суточных удоев коров черно-пестрой породы по данным задания № 3;

г) живой массы коров симментальской породы по данным задания № 4;

д) живой массы быков бестужевской породы в возрасте 5 лет и старше по данным задания № 5;

е) массы парной туши свиней крупной белой породы по данным задания № 6;

ж) времени наступления овуляции у коров бестужевской породы по данным задания № 7;

з) чисел пульсовых ударов свиней беркширской породы по данным задания № 8;

и) плодовитости свиноматок крупной белой породы по второму опоросу по данным задания № 9;

к) настрига шерсти у овец асканийской породы по данным задания № 10;

л) промера обхвата груди коров бестужевской породы по третьей лактации по данным задания № 11;

м) удоя коров бестужевской породы по третьей лактации по данным задания № 12;

н) содержания жира в молоке коров черно-пестрой породы по второй лактации по данным задания № 13;

о) живой массы коров симментальской породы по первой лактации по данным задания № 14;

п) удоя коров черно-пестрой породы за 305 дней лактации по данным задания № 15.

4.5. Вычисление нормированного отклонения

Среднее квадратическое отклонение и коэффициент вариации являются показателями разнообразия, характеризующими вариационный ряд в целом.Показателем, характеризующим отдельную варианту (или группу вариант), служит *нормированное отклонение* (t). Оно является также показателем разнообразия признака и представляет выраженное в долях сигмы взвешенное отклонение соответствующей варианты от средней арифметической:

$$t = \frac{v - \bar{X}}{\sigma}.$$

Нормированное отклонение является неименованной (безразмерной) величиной, что позволяет использовать его для сравнения отдельных объектов, принадлежащих к разным генеральным совокупностям, по степени развития у них того или иного признака, а также определять, какой признак у данного объекта развит лучше.

Пример. При отборе коров в племенную группу (или при покупке племенных животных) необходимо решить, какое из двух животных разного возраста лучше по молочной продуктивности. Известно, что одна корова за первую лактацию дала 3500 кг молока, а другая после третьего отела имела удой 4580 кг. Известно также, что средний удой первотелок в этом хозяйстве в данном году составил 2500 кг и σ = 500 кг, а удой полновозрастных животных (третий отел и старше) — 3500 кг и σ = 600 кг. Если для этого случая рассчитать абсолютные отклонения, то вторая корова будет выглядеть предпочтительнее первой:

$$d_1 = v_1 - \bar{X}_1 = 3500 - 2500 = 1000 \text{ кг};$$

$$d_2 = v_2 - \bar{X}_2 = 4580 - 3500 = 1080 \text{ кг}.$$

Однако этот вывод был бы ошибочным. Расчет нормированных отклонений t покажет их действительное место в соответствующих вариационных рядах по степени развития данного признака:

$$t_1 = \frac{v_1 - \bar{X}_1}{\sigma_1} = \frac{1000}{500} = 2{,}0; \; t_2 = \frac{v_2 - \bar{X}_2}{\sigma_2} = \frac{1080}{600} = 1{,}8.$$

Отсюда видно, что первотелка более продуктивна и ей следует оказать предпочтение.

Соответствующим образом решается вопрос о том, какой из двух или нескольких признаков развит лучше у данного животного. Если, например, по удою данная корова характеризуется значением $t = 2{,}0$; по жирномолочности — $t = 2{,}5$; по живой массе — $t = 1{,}7$, то совершенно ясно, что у нее сильнее развит признак жирномолочности.

Нормированное отклонение используется при решении селекционных задач: для оценки и сравнения интенсивности отбора и селекционной границы в разных стадах по разным признакам:

$$i = \frac{X_p - \bar{X}}{\sigma},$$

где i — нормированное отклонение, характеризующее интенсивность отбора; X_p — среднее значение признака в группе отобранных на племя животных; $\bar{X}$ — среднее значение признака в группе до отбора;

$$u = \frac{X_u - \bar{X}}{\sigma},$$

где u —величина отсекающей абсциссы, выраженная в долях сигмы. Селекционная граница определяется по формуле

$$X_u = \bar{X} + u \cdot \sigma.$$

Пример. Средний удой в стаде ($\bar{X}_1$) составил 3500 кг при среднем квадратическом отклонении σ_1 = 600 кг. Средняя жирномолочность коров в стаде ($\bar{X}_2$) — 3,70 %, при σ_2 = 0,20 %. Группа животных, отобранных из стада для племенного использования (племенное ядро), имела средний удой ($\bar{X}_{p1}$) 3800 кг и среднюю жирномолочность ($\bar{X}_{p2}$) 3,77 %. Рассчитав i, можно установить, по какому из признаков провели более интенсивный отбор:

$$i_1 = \frac{3800 - 3500}{600} = 0{,}5; \; i_2 = \frac{3{,}77 - 3{,}70}{0{,}20} = 0{,}35.$$

Отсюда видно, что в стаде большее внимание уделялось отбору по удою, чем по жирномолочности.

Значения i, u, P (доля отобранных на племя животных) связаны с функциональной зависимостью и сведены в специальную таблицу (табл. 4.5), используя которую, можно по известному параметру определить неизвестные. Например, селекционная граница по удою — 0,52, т. е. минимальное значение признака в отобранной группе составит:

$3500 - 0{,}52 \cdot 600 = 3188$ кг, а доля отобранных животных — 0,7, или 70 % поголовья.

Таблица 4.5

Таблица интенсивности отбора

P	u	i
0,9	–1,28	0,20
0,8	–0,84	0,35
0,7	–0,52	0,50
0,6	–0,25	0,64
0,5	0,00	0,80
0,4	0,25	0,97
0,3	0,52	1,16
0,2	0,84	1,40
0,1	1,28	1,76

Нормированное отклонение используется также при оценке производителей по качеству потомства, при сравнении показателей животных из разных совокупностей, при суждении о ходе выздоравливания животных и т. д. Показатель нормированного отклонения удобен как для оценки отдельных вариант, так и при характеристике сравниваемых групп.

Задания для самостоятельной работы

Задание 1. Какой вывод можно сделать на основе вычисления нормированного отклонения о ходе выздоравливания животного в результате его лечения по показателям табл. 4.6.

Таблица 4.6

Содержание общего белка в сыворотке крови и количество лейкоцитов в 1 мм³ крови

Показатели крови	Норма для здорового животного		Показатели в период лечения на		
	$\bar{X}$	σ	10-й день	20-й день	30-й день
Число лейкоцитов, тыс.	6,0	0,5	8,6	7,6	7,0
Общий белок, г %	6,5	1,0	4,6	5,7	6,5

Задание 2. От быка-производителя получены две дочери. Сравнив их со сверстницами по приведенным ниже данным (табл. 4.7) и используя нормированные отклонения, дайте заключение о качестве этих животных.

Таблица 4.7

Показатели продуктивности дочерей быка и их сверстниц

Показатели	Сверстницы		Дочери быка	
	$\bar{X}$	σ	первая	вторая
Удой за лактацию, кг	3500	400	3900	4100
Содержание жира, %	3,80	0,35	3,92	3,72
Живая масса, кг	450	50	455	500
Оценка экстерьера, балл	80	2,2	78	82

Задание 3. В хозяйстве сравнивают по суточным удоям две первотелки. Первая из них на втором месяце лактации продуцирует в сутки 10 кг молока, вторая на шестом месяце лактации — 5 кг. Суточный удой первотелок данного хозяйства выражается следующими показателями: за второй месяц лактации $\bar{X}_1 = 8$ кг, $\sigma_1 = 2$ кг, а за шестой месяц $\bar{X}_2 = 4$ кг, $\sigma_2 = 1$ кг.

Установить путем сравнения нормированного отклонения, какая из оцениваемых коров лучше по продуктивности.

Задание 4. Установить путем сравнения нормированных отклонений, какой из двух 12-месячных бычков общего происхождения лучше. Один из них, выращенный при менее благоприятных условиях, весит 200 кг; второй, находившийся в нормальных условиях, — 230 кг. Средние показатели живой массы бычков годовалого возраста составляют: в первом хозяйстве — $\bar{X}_1 = 210$ кг, $\sigma_1 = 20$ кг, а во втором — $\bar{X}_2 = 250$ кг, $\sigma_2 = 20$ кг.

Задание 5. Используя нормированные отклонения, сопоставить приведенные ниже данные (табл. 4.8) помесных бестужевско-голштинских коров по удою за лактацию и жирномолочности со средними показателями коров этой популяции, удой которых за пятую лактацию равен 3800 кг, σ = 845 кг; содержание жира в молоке 4,11 %, σ = 0,18.

Таблица 4.8

Молочная продуктивность помесных бестужевско-голштинских коров

Кличка и номер коровы	Удой за пятую лактацию, кг	Содержание жира в молоке, %
Утка 750	4333	4,30
Бурёна 7306	5631	4,04
Быстрая 1001	4110	3,95
Ветка 6964	5698	3,98
Ласточка 61	5250	4,01
Пальма 64	7513	4,53
Белка 66	4468	3,94
Кудесница 551	4346	4,10
Галка 061	4216	4,52

Контрольные вопросы и задания

1. Какие показатели характеризуют разнообразие признака?
2. Как вычисляется среднее квадратическое отклонение в малых выборках?
3. Как вычисляется среднее квадратическое отклонение в больших выборках?
4. Как вычисляется среднее квадратическое отклонение для альтернативных признаков?
5. В каких случаях вычисляют коэффициент вариации? Приведите его формулу.
6. Что такое нормированное отклонение?Для чего используется этот показатель?
7. Что показывает среднее квадратическое отклонение? В каких единицах измерения оно выражается?
8. Что показывает коэффициент изменчивости?
9. Какой биометрический показатель, характеризующий изменчивость, можно использовать для сравнения изменчивости разных признаков у животных?
10. Что понимается под изменчивостью признака?
11. Какая связь существует между средней арифметической и средним квадратическим отклонением?

Глава 5
РАСПРЕДЕЛЕНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

5.1. Случайные события. Классическое определение вероятности

Множество $\Omega=(\omega_1, \omega_2, \omega_n)$ всех возможных исходов эксперимента называется *пространством элементарных исходов*, а каждый его элемент называется элементарным исходом или *элементарным событием*. Событием называется любое подмножество A этого пространства: $A \subseteq \Omega$. Каждому элементарному исходу ω_i поставим в соответствие число $p_{i\geq 0}$, называемое его вероятностью, такое, что $\sum p_i = 1$.

Простейшим пространством элементарных исходов является так называемая классическая модель, в которой пространство конечно и все исходы эксперимента:

1) равновозможны (т. е. их вероятности предполагаются равными);

2) несовместны (т. е. никакие исходы не могут произойти одновременно);

3) в сумме образуют все пространство (т. е. никакие другие исходы, кроме перечисленных, не могут произойти).

В этом случае *вероятность события А* определяется по формуле $P_A = m/n$, где n — число элементов множества Ω (общее число исходов), а m — число элементов множества A (число исходов, благоприятствующих событию A).

Событие $\bar{A}$, состоящее из всех элементарных исходов, не входящих в A, называется противоположным событием к событию A. Оно происходит тогда и только тогда, когда событие A не произошло. Очевидно, что $PA + P\bar{A} = 1$.

5.2. Свойства вероятностей

1. Вероятность любого события есть число, заключенное между нулем и единицей, т. е. $0 \leq P(A) \leq 1$. Вероятность невозможного события равна нулю, а вероятность достоверного события равна единице.

2. Если события A и B несовместны, то

$$P(A+B)=P(A)+P(B).$$

3. Вероятность любого события A в сумме с вероятностью противоположного события $\bar{A}$ равна единице:

$$P(A)+P(\bar{A})=1.$$

Если вероятность интересующего нас события *А* по каким-либо причинам вычислить трудно, то можно попытаться рассчитать вероятность противоположного события, а затем с помощью свойства 3 вычислить искомую вероятность события *А*.

В генетике применяются правила сложения и умножения вероятностей.

Правило сложения вероятностей: *вероятность протекания двух взаимоисключающих событий равна сумме вероятностей каждого из них.*

Правило умножения вероятностей: *вероятность совместного осуществления двух или большего числа независимых событий равна произведению вероятностей каждого из отдельно взятых событий.*

Теория вероятностей успешно используется в генетике при вычислении вероятности проявления того или иного класса потомства.

Используя решетку Пеннета, можно определить вероятность появления того или иного типа потомства. Исходя из того, что гаметы, несущие аллель А и аллель а у гетерозиготного родителя, образуются равновероятно (с частотой 1/2), и согласно правилу умножения вероятностей можно установить вероятность появления того или иного генотипического класса:

Гаметы	1/2A	1/2a
1/2A	[redacted]	
1/2a		

Согласно правилу сложения вероятностей частоту появления каждого из генотипических классов в F_2 можно записать в виде формулы

$$1/4\,AA + (1/4\,Aa + 1/4Aa) + 1/4aa = 1/4\,AA + 1/2\,Aa + 1/4\,aa.$$

Соответственно, частота появления различных фенотипических классов будет равна:

$$(1/4\,AA + 1/2\,Aa) + 1/4\,aa = 3/4\,A + 1/4\,aa.$$

Если для дигибридного и тригибридного скрещивания частоту появления генотипических и фенотипических классов в F_2 можно рассчитать по решетке Пеннета, то при полигибридных скрещиваниях это выполнить сложно. При скрещивании особей, различающихся по пяти парам признаков, число возможных сочетаний гамет равно (4^5), т. е. 1024, а при шести парах различающихся признаков таких сочетаний будет (4^6), т. е. 4096.

Поэтому необходимо применить правило умножения вероятностей.

Пример. При полигибридном скрещивании растений различающихся по шести парам независимо наследуемых признаков (при условии полного доминирования): AaBbCcDdEeFf × AaBbCcDdEeFf, необходимо рассчитать частоту генотипа AaBbCcDdEeFf и AAbbCcDdeeFF и AabbCcDDeeFf. Поэтому частота гетерозиготных форм AaBвCcDdEeFf следующая. Вероятность генотипа Аа в потомстве от скрещивания Аа×Аа равна 1/2. Вероятность генотипа Bb от скрещивания Bb × Bb также будет равна 1/2, генотипа Cc от скрещивания Cc × Cc равна 1/2, генотипа Dd от скрещивания Dd×Dd — 1/2, генотипа Ee от скрещивания Ee × Ee — 1/2 и вероятность генотипа Ff от скрещивания Ff × Ff — 1/2. Следовательно, вероятность генотипа AaBbCcDdEeFf будет равна: 1/2 × 1/2 × 1/2 × 1/2 × 1/2 × 1/2 = 1/64.

Частоту гетерозиготных форм генотипа AAbbCcDdeeFF можно определить следующим образом. Вероятность генотипа AA в потомстве от скрещивания Aa × Aa равна 1/4. Вероятность генотипа bb от скрещивания Bb × Bb будет равна 1/4, а вероятность генотипа Cc от скрещивания Cc × Cc — 1/2, генотипа Dd от скрещивания Dd × Dd — 1/2, генотипа ee от скрещивания Ee × Ee — 1/4 и вероятность генотипа FF от скрещивания Ff × Ff — 1/4. Следовательно, вероятность генотипа AAbbCcDdeeFF будет равна: 1/4 × 1/4 × 1/2 × 1/2 × 1/4 × × 1/4 = 1/1024, а вероятность генотипа AabbCcDDeeFf — 1/2 × 1/4 × × 1/2 × 1/4 × 1/4 × 1/2 = 1/512.

5.3. Повторные независимые испытания. Формула Бернулли

Опыты называются независимыми, если вероятность каждого исхода любого опыта не изменяется от того, какие исходы имели другие опыты. Пусть производится n независимых опытов и известна $P(A) = p$ — вероятность появления события A в каждом из опытов ($P(\overline{A}) = 1 - p = q$). Тогда вероятность того, что в n независимых опытах событие A появится ровно k раз, равна по *формуле Бернулли*:

$$P_n(k) = C_n^k p^k q^{n-k}.$$

Число $k = k_0$, при котором вероятность $P_n(k)$ принимает наибольшее значение, называется *наивероятнейшим числом появления события*.

Если $(n + 1)p$ — число дробное, то k_0 равно целой части числа $(n + + 1)p$. Если же $(n + 1)p$ — число целое, то k_0 принимает два значения: $k_0 = (n + 1)p - 1$ и $k_0 = (n + 1)p$.

Пример[1]. Предположим, что 30 % студентов университета отличники. Какова вероятность того, что среди первых пяти встречных студентов окажется только один отличник? Какова вероятность того, что среди них есть хотя бы один отличник? Каково наиболее вероятное число отличников среди них?

Решение. Поскольку студентов в университете много (несколько тысяч), то по мере опроса нескольких из них пропорции в оставшейся части практически не изменяются. Поэтому можно считать опрос каждого студента независимым опытом. Всего опытов производится $n = 5$, а вероятность положительного ответа $p = 0{,}3$. Вычисление по формуле Бернулли дает следующий результат:

$$P_5(1) = C_5^1 0{,}3(0{,}7)^4 = 0{,}36015.$$

[1] *Крупин В. Г., Павлов А. Л., Попов Л. Г.* Высшая математика. Теория вероятностей, математическая статистика, случайные процессы. Сборник задач с решениями : учеб. пособие. М. : Издательский дом МЭИ, 2013.

Вероятность хотя бы одного правильного ответа проще вычислять, по противоположному событию: $P_5(k \geq 1)$ $1 - (0{,}7)^5 = 1 - 0{,}16807 = 0{,}83193$. Поскольку $(n + 1)p = (5 + 1)0{,}3 = 1{,}8$ (целая часть числа равна 1), наиболее вероятное число отличников среди пяти опрошенных $k_0 = 1$.

Ответ. 0,36015; 0,83193; 1.

5.4. Распределение Пуассона

Распределение Пуассона относится к числу важнейших теоретических распределений, имеющих практическое применение. Данное распределение принимает классическую форму в случае, если значения признака имеют дискретный характер $x = 0, 1, 2, = 0, 1, 2, 3, = 0 \ldots$ и являются результатом редко возникающего события, причем с увеличением значений признака вероятность наступления события уменьшается. Распределение Пуассона называют распределением редких событий. Это распределение наблюдается в тех совокупностях, число единиц которых достаточно велико, $N \geq 100$, а доля единиц, которые обладают большими значениями признака, мала. При этом средняя арифметическая величина и дисперсия, вычисленные на основании эмпирических данных, как правило, совпадают или незначительно различаются между собой. Распределение Пуассона характеризуется одним параметром — средней величиной $a = \bar{x}$ — средняя арифметическая ряда.

Распределение Пуассона выражается формулой

$$P(x) = \frac{a^x e^{-a}}{x!},$$

где $P(x)$ — веорятность того, что признак примет то или иное значение; $a = \bar{x}$ — средняя арифметическая ряда; $e = 2{,}7183$ — основание натуральных логарифмов; $x!$ — произведение натуральных чисел $1 \cdot 2 \cdot 3 \ldots x$ (факториал).

Последовательность расчета теоретических частот кривой распределения Пуассона:

- находят среднюю арифметическую ряда $a = \bar{x}$;
- по таблицам определяют e^{-a};
- для каждого значения x вычисляют теоретическую частоту по формуле

$$f' = N\frac{a^x e^{-a}}{x!} = NP(x),$$

где N — объем изучаемой совокупности.

По закону Пуассона распределяются редкие случайные события, современной биологии: частоты мутаций, численность перезимовавших клонов насекомых, рождение троен у одноплодных животных и т. п.

Пример. В стаде крупного рогатого скота бестужевской породы (N = = 400 коров) последовательно из года в год рождались тройни. Результаты представлены в табл. 5.1.

Таблица 5.1

Тройни в стаде скота бестужевской породы

Число родившихся троен x_i	Количество коров f_i	f_i	$(f_i - f_i)^2 f_i'$	f_i'
0	304	0	23,83	302
1	80	80	41,47	84
2	15	30	44,37	12
3	1	3	7,39	1
$\sum$	400	113	117,06	399

Вычислим среднюю арифметическую ряда и дисперсию:

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{113}{400} = 0{,}28; \ \sigma^2 = \frac{\sum (f_i - f_i')^2 f_i}{\sum f_i} = \frac{117{,}06}{400} = 0{,}29 .$$

Поскольку $\overline{x} = \sigma^2$, имеем основание полагать о подчинении данного распределения закону Пуассона. Поскольку $a = \overline{x} = 0{,}28$, по таблице приложения находим значение $e^{-0,28}$ = 0,7558.

Определим теоретические расчеты f_i':

$$f_0' = 400\frac{0{,}28^0 \cdot 0{,}7558}{0!} = 302;$$

$$f_1' = 400\frac{0{,}28 \cdot 0{,}1558}{1!} = 84;$$

$$f_2' = 400\frac{0{,}28^2 \cdot 0{,}7558}{2!} = 12;$$

$$f_3' = 400\frac{0{,}28^3 \cdot 0{,}7558}{3!} = 9.$$

5.5. Распределение Максвелла

Часто встречаются асимметричные распределения, которые не следуют закону Пуассона. Одним из таких распределений является *распределение, описываемое формулой Максвелла*

$$P(X) = \frac{2}{\sqrt{2\pi}} \cdot \frac{t^2}{a} \cdot e^{\frac{t^2}{2}} \cdot dx ,$$

где a = 0,6267$\overline{x}$ — параметр распределения, определяемый через среднюю арифметическую $\overline{x}$ варьирующего признака; $t = x_i/a$, здесь x_i — числовые значения случайной величины X; dx — разность между двумя смежными значениями переменной величины X.

То, что эмпирическое распределение подчинено закону Максвелла, является равенство между средним квадратическим отклонением и величиной $0{,}674a$, т. е. $S_x = 0{,}674a$, в отличие от распределения Пуассона, где $Sx^2 = \bar{x}$.

Чтобы рассчитать по формуле Максвелла теоретические (выравнивающие) частоты, нужно проделать следующее:

- определить среднюю арифметическую эмпирического вариационного ряда и параметр a;
- разделить каждую классовую варианту x_i на величину a, что даст значения t;
- найти для каждого значения $t = x_i/a$ по таблице приложений значение функции $f(t)$;
- определить значения t^2/a;
- умножить значения t^2/a на удвоенную величину t и на величину классового промежутка ($\lambda = dx$), т. е. определить $P = (t^2/a)2f(t)\lambda$;
- умножить значения P на общее число наблюдений n, получить теоретические частоты вариационного ряда, т. е. $f' = Pn$.

Пример[1]. Скрещивали мелкоплодную линию томатов с крупноплодной линией того же сорта. В первом поколении плоды оказались не среднего, а несколько меньшего размера. При скрещивании представителей первого поколения между собой, во втором поколении, масса отдельных плодов еще более приблизилась к массе плодов исходной мелкоплодной линии. Распределение массы плодов семенных гнезд, взятых с 928 растений расщепляющейся популяции томатов, показано в табл. 5.2.

Характеристики этого распределения следующие: $\bar{x} = 48{,}7$ г, $Sx = 23{,}8$, откуда $a = 0{,}6267$, $\bar{x} = 30{,}5$. Близость $S_x = 23{,}8$ к величине $0{,}674a = 20{,}6$ позволяет предположить, что данное распределение следует закону Максвелла. Расчет выравнивающих частот f' приведен в табл. 5.2. Значения $t = x_i/a$ получены так: $t_1 = 10/30{,}5 = 0{,}328 = 0{,}33$; $t_2 = 20/30{,}5 = 0{,}655 = 0{,}66$ и т. д. Значения $f(t)$ находят в таблице приложений, приведенных в специальной литературе по биометрии и математической статистике [7, 8, 12, 17, 18] или вычисляют по формуле

$$f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}},$$

где $f(t)$ — ординаты нормальной кривой.

$$f(t_1 = 0{,}33) = 0{,}3778; f(t_2 = 0{,}66) = 0{,}3209 \text{ и т. д.}$$

Значения в предпоследней графы этой таблицы рассчитаны следующим образом: $t_1^2 = (0{,}33)^2 = 0{,}109$; $t_1^2/a = 0{,}109/30{,}5 = 0{,}0035$; $2f(t) = 2 \cdot 0{,}3778 = 0{,}7556$; $f(t_1^2/a)2f(t)\lambda = 0{,}0035 \cdot 0{,}7556 \cdot 10 = 0{,}0264$. Умножая эту величину на $n = 928$, находим значение $f' = P_n = 0{,}0264 \cdot 928 = 25$ и так до конца ряда.

[1] *Лакин Г. Ф.* Биометрия : учеб. пособие для биол. спец. вузов. М. : Высшая школа, 1990.

Таблица 5.2

Распределение массы плодов семенных гнезд томатов

Классы x_i	Частоты f	$t = x_i/\sqrt{\bar{x}}$	$f(t)$	t^2	$t^2/\sqrt{\bar{x}}$	$(t^2/\sqrt{\bar{x}}) \times$ $\times 2f(t)x_i = p$	$p_n = f'$
10	28	0,33	0,3778	0,109	0,0035	0,0264	25
20	93	0,66	0,3209	0,430	0,0141	0,0905	84
30	186	0,98	0,2468	0,966	0,0316	0,1571	146
40	148	1,31	0,1691	1,716	0,0562	0,1900	176
50	176	1,61	0,1040	2,686	0,0880	0,1830	170
60	102	1,97	0,0573	3,869	0,1268	0,1450	134
70	74	2,30	0,0283	5,267	0,1727	0,0977	91
80	46	2,62	0,0129	6,880	0,2256	0,0582	54
90	28	2,95	0,0051	8,708	0,2855	0,0291	28
100	19	3,28	0,0018	10,745	0,3523	0,0127	12
110	14	3,61	0,0005	13,032	0,4273	0,0043	4
120	6	3,93	0,0002	15,476	0,5074	0,0020	2
130	5	4,25	0,0001	18,164	0,5955	0,0011	1
140	2	4,59	0,0000	21,068	0,6907	0,0001	1
150	1	4,92	0,0000	24,186	0,7930	0,0000	0
$\sum$	928	—	—	—	—	0,9972	928

Как результат, получен ряд теоретически вычисленных частот f'. Представленные графически в виде вариационных кривых эмпирические и вычисленные по закону Максвелла частоты этого ряда (рис. 5.1) свидетельствуют о том, что они согласуются между собой. Следовательно, формула для нахождения выравнивающих частот этого ряда выбрана правильно.

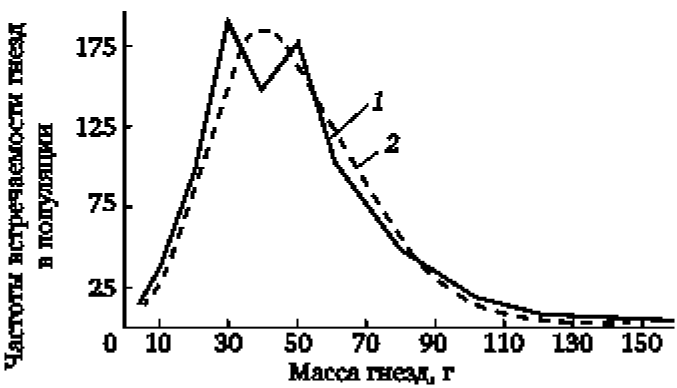

Рис. 5.1. **Эмпирическая (*1*) и вычисленная по формуле Максвелла (*2*) кривые распределения массы гнезд томатов**

5.6. Биномиальный закон распределения вероятностей

Биномиальным называется распределение, когда вероятности появления отдельных вариант соответствуют коэффициентам разложения бинома Ньютона. По биноминальному закону распределяются признаки, которые варьируют дискретно (число здоровых и больных особей в популяции, численность особей с доминантным и рецессивным признаком, и т. д.).

Частоты отдельных классов пропорциональны коэффициентам разложения бинома Ньютона $(p + q)^k$, где p и q — вероятность появления каждого признака, k — число классов, отличающихся по появлению признака. Если $p = 0{,}5$, $q = 0{,}5$, а k увеличивается, то биномиальная кривая приближается к нормальной кривой, которая является пределом биномиального распределения. При больших различиях значений p и q асимметрия биномиальной кривой более значительна. Характеризуют биномиальное распределение:

1) средняя арифметическая $\bar{x} = \dfrac{\sum xf}{n}$, где n — объем выборки; x — число классов альтернативного признака;

2) среднее квадратическое отклонение $\sigma = \sqrt{kpq}$, где k — число субвыборок; p и q — частоты появления каждого альтернативного признака.

Закон биноминального распределения выражается формулой $(p + q)^k$. Коэффициенты разложения бинома указывают на вероятность альтернативного признака. Их можно определить по треугольнику Паскаля:

Число наблюдений субвыборок, k	Биноминальные коэффициенты	Число возможных исходов (2^k)
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
и т. д.		

Биномиальный закон распределения вероятностей — это наиболее распространенный вид дискретного распределения.

В основе биномиального распределения лежит альтернативное проявление изучаемого признака: он может присутствовать у единичного объекта или отсутствовать, проявиться или нет. Если вероятности появления объектов разного качества приблизительно равны (например, когда общее число больных примерно равно числу здоровых особей), биномиальное распределение имеет симметричную, колоколообразную (но ступенчатую) форму, подобную нормальному распределению.

Пример. Важной характеристикой популяций животных является плодовитость поголовья маток, которая подчиняется биномиальному закону распределения вероятностей. Отдельное событие — это появление или (непоявление) детеныша. При этом вероятность реализации отдельного события «детеныш появится» составляет p, а вероятность события «детеныш не появится» равна $q = 1 - p$.

Рассмотрим результаты изучения плодовитости овцематок стада овец романовской породы (число ягнят на матку).

Исходные данные приведены в таблице.

Проведем расчет вероятности элементарных событий (появления — непоявления) ягнят. Для этого сначала вычислим общее число появившихся ягнят. Оно равно сумме всех произведений числа ягнят от отдельной овцематки (x) на частоту их встречаемости (f): $N_ф = \sum(x \cdot f) = 370$.

Плодовитость овцематок, x	Число овцематок (частота) f	Число эмбрионов $(x \cdot f)$
1	1	1
2	3	6
3	24	72
4	26	104
5	22	110
6	8	48
7	3	21
8	1	8
$k = 8$	$n = 88$	$N_ф = 370$

Затем рассчитаем общее число потенциальных зародышей (суммарное количество потенциальных ягнят от всех овцематок) $k = 8$; $N_п = n \cdot 8 = 88 \cdot 8 = 704$ экз. Наконец, определим долю реализованных исходов (рожденных ягнят) среди потенциальных:

$$p = N_ф / N_п = 370 / 704 = 0{,}53$$

и долю нерожденных:

$$q = 1 - p = 1 - 0{,}53 = 0{,}47.$$

Вычисленные характеристики сущности наблюдаемого процесса размножения овец романовской породы свидетельствуют о том, что вероятность рождения отдельного ягненка ($p = 0{,}53$) превышает вероятность его не рождения ($q = 0{,}47$), это означает, что в стаде имеются овцематки с более высокой плодовитостью и указывает на некоторую асимметрию распределения. Вычисляем показатели асимметрии и эксцесса и их ошибки по следующим формулам:

$$A=\frac{q-p}{\sqrt{n\cdot p\cdot q}}=\frac{0{,}47-0{,}53}{\sqrt{88\cdot 0{,}53\cdot 0{,}47}}=0{,}109;\ m_A=\sqrt{\frac{6}{n}}=0{,}26;$$

$$E=\frac{\frac{1}{p\cdot q}-6}{n}=\frac{\frac{1}{0{,}53\cdot 0{,}47}-6}{28}=-0{,}022.$$

Таким образом, вычисленные показатели асимметрии и эксцесса несущественные. По найденным вероятностям p и q можно рассчитать параметры биномиального распределения плодовитости овцематок по формулам: $\bar{x}=k\cdot p=8\cdot 0{,}53=4{,}2$ ягненка на овцематку; $\sigma=\sqrt{k\cdot p\cdot q}=$ $=\sqrt{8\cdot 0{,}53\cdot 0{,}47}=0{,}41$ ягненка на овцематку.

Ошибка средней величины признака: $m_x=\frac{\sigma}{\sqrt{n}}=\frac{0{,}41}{9{,}38}=0{,}150$; ошибка стандартного отклонения: $m_\sigma=\frac{\sigma}{\sqrt{2n}}=\frac{0{,}41}{\sqrt{2\cdot 88}}=0{,}031$.

Доверительный интервал для параметров биномиального распределения следующий: при уровне значимости t = 0,05 находим доверительные интервалы для $\bar{x}$ и σ:

$$\bar{x}=tm_x=4{,}2\pm 1{,}96\cdot 0{,}15=4{,}2\pm 0{,}295;$$

$$\sigma=\pm tm_\sigma=1{,}41\pm 1{,}96\cdot 0{,}106=1{,}41\pm 0{,}208.$$

Значение генерального параметра находится в диапазоне: для средней арифметической — от 3,91 до 4,49 ягненка на овцематку, для стандартного отклонения — от 1,20 до 1,62 ягненка на овцематку.

Пример. Найти вероятность рождения телят с заболеванием порфирия, которая вызывает повышенную светочувствительность и розовую окраску зубов и нормальных телят у крупного рогатого скота, где оба родителя гетерозиготные по гену порфирии (при получении, от высокопродуктивной коровы, пяти телят).

Для нахождения вероятности рождения здоровых и больных телят необходимо использовать формулу биноминального распределения: $(a+b)^n$, где a — вероятность рождения больного теленка, b — вероятность рождения здорового теленка; n — возможное число телят. Для данного случая формула биноминального распределения будет иметь вид

$$(a+b)^5=a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5.$$

В нашем примере величина a (вероятность рождения больного теленка) равна 1/4, b (вероятность рождения здорового) равна 3/4.

Первый член формулы (a^5) обозначает вероятность рождения всех пяти телят больных порфирией; второй член ($5a^4b$) — вероятность события, когда четверо телят будут больные и один здоровый; третий член ($10a^3b^2$) — вероятность события, когда трое телят больные и двое здоровые; четвертый член ($10a^2b^3$) — вероятность события, когда двое телят больные и трое здоровые; пятый член ($5ab^4$) — вероятность того, что один теленок будет больным и четыре здоровыми; шестой член (b^5) — вероятность того, что все пять телят будут здоровыми.

Подставляя числовые значения для каждого члена формулы можно определить вероятность того или иного события.

В нашем примере:

1) вероятность того, что все телята будут больными равна: $a^5 = (1/4)^5 = 1/1024 = 0{,}001$;

2) вероятность того, когда четверо телят будут больные и один здоровый равна: $5a^4b = 5(1/4)^4 \cdot (3/4) = 15/1024 = 0{,}015$;

3) вероятность того, что трое телят будут больными, а двое здоровыми равна: $10a^3 \cdot b^2 = 10(1/4)^3 \cdot (3/4)^2 = 90/1024 = 0{,}089$;

4) вероятность того, что двое телят будут больными, а трое — здоровыми равна: $10a^2 \cdot b^3 = 10(1/4)^2 \cdot (3/4)^3 = 270/1024 = 0{,}26$;

5) вероятность того, один будет больным, а четверо — здоровыми равна: $5a \cdot b^4 = 5(1/4) \cdot (3/4)^4 = 405/1024 = 0{,}395$;

6) вероятность того, что все пять телят будут здоровыми равна: $b^5 = (3/4)^5 = 243/1024 = 0{,}24$.

Проверка: $0{,}001 + 0{,}015 + 0{,}089 + 0{,}395 + 0{,}26 + 0{,}24 = 1$.

Нормальное распределение названо по имени авторов — распределение Гаусса — Лапласа. Нормальное распределение наиболее часто встречается на практике. Это наиболее распространенный тип распределения особей совокупности по классам вариационного ряда.

Для биологической практики наиболее важными являются следующие виды нормального распределения:

- эмпирическое нормальное распределение;
- общее (теоретическое) нормальное распределение;
- стандартное нормальное распределение.

Эмпирическое нормальное распределение — это распределение, полученное опытным путем на основе статистического исследования. Особенность данного распределения в том, что объем его совокупности всегда конечен.

Общее (теоретическое) нормальное распределение — это абстрактная математическая модель, которую используют в качестве стандарта для сравнения с эмпирическим (опытным) распределением по разным статистическим критериям. Особенность этого распределения: оно построено в предположении, что объем совокупности является бесконечно большим.

Нормальное распределение — это распределение непрерывной случайной величины X, характеризуемое плотностью вероятности:

$$f_N(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} \quad \text{при } -\infty < x < \infty,$$

где e и π — математические постоянные, равные 2,71 и 3,14 соответственно; a, σ^2 — параметры, которые равны соответственно математическому ожиданию и дисперсии распределения.

График общего (теоретического) нормального распределения представляет собой одновершинную колоколообразную симметричную вариационную кривую, которая асимптотически приближается к оси абсцисс (рис. 5.2).

Эту кривую строят в прямоугольной системе координат, где по горизонтали откладывается текущее числовое значение признака x_i в порядке возрастания, а по вертикали — значение функции $y = f(x)$, которое соответствует числу объектов (вариантов) с данным числовым значением признака.

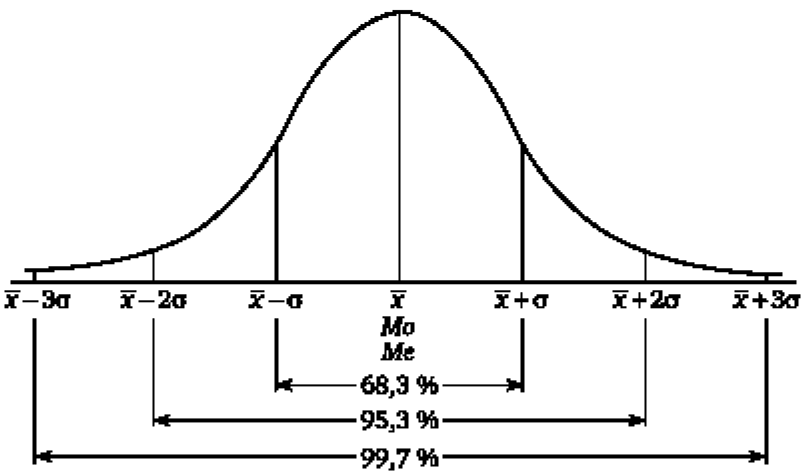

Рис. 5.2. **График нормального распределения**

Особенности нормального распределения:

- кривая распределения симметрична относительно средней арифметической (или $\bar{x}$);
- чем больше отклоняются числовые значения вариантов от средней арифметической, тем реже такие варианты встречаются в совокупности;
- площадь под кривой равна объему совокупности $N = 1$ (точнее, 0,9973).

В пределах площади, отсекаемой перпендикулярами к горизонтали в точках:

- $\pm\ 1\sigma$ находится 68,3 % всех наблюдений;
- $\pm\ 2\sigma$ находится 95,5 % всех наблюдений;
- $\pm\ 3\sigma$ находится 99,7 % всех наблюдений.

Для общего (теоретического) нормального распределения средняя арифметическая $\bar{x}$, мода *Mo* и медиана *Me* равны между собой.

Смещение кривой по горизонтальной оси определяется числовым значением средней арифметической.

Степень вытянутости кривой по вертикали зависит от значения стандартного отклонения. Чем меньше числовое значение имеет стандартное отклонение σ, тем более острой будет кривая.

Наибольшее значение функции y имеет место тогда, когда числовое значение конкретной варианты равно значению средней арифметической $x_i = \bar{x}$.

Стандартное нормальное распределение — это особая форма распределения, которую можно использовать в качестве стандарта при оценках любых данных, независимо от их размерности. По-другому такое распределение называют нормированным нормальным распределением.

Введение стандартного нормального распределения вызвано тем, что обычная формула общего нормального распределения для практического применения неудобна. По этой формуле положение кривой по горизонтали, ее размер и форма определяются и зависят от объема совокупности, значения средней арифметической и стандартного отклонения. Поэтому используют нормированное отклонение:

$$z = \frac{x - \bar{x}}{\sigma},$$

где x — числовое значение конкретной варианты; $\bar{x}$ — средняя арифметическая выборочной или генеральной совокупности; σ — стандартное отклонение.

Так же как и для общего теоретического нормального распределения, для стандартного распределения:

- 68,3 % площади под кривой соответствует $\pm 1\sigma$;
- 95,5 % площади под кривой соответствует $\pm 2\sigma$;
- 99,7 % площади под кривой соответствует $\pm 3\sigma$.

Первая функция нормального распределения (отсекающая ордината) выражается формулой

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} = 0{,}39894 \cdot 2{,}711828^{\frac{-z^2}{2}}.$$

Вторая функция нормального распределения $\varphi(z)$ показывает, какова площадь кривой, если ее отсечь ординатами y_0 и y_i.

Для стандартного нормального распределения существуют «зэт-таблицы», которые связывают значения «зэт» с площадью кривой, отсеченной ординатами y_0 и y_i. В биологических и зоотехнических исследованиях это соответствует числу особей изучаемой совокупности (в процентах) которые вошли в отсеченную часть кривой.

Третья функция нормального распределения $F(z)$ указывает, какая будет средняя величина признака в отсеченной части кривой. Она вычисляется по формуле

$$F(z) = \frac{f(z)}{0{,}5 \pm \varphi(z)},$$

где $f(z)$ — первая функция нормированного отклонения (отсекающая ордината); $\varphi(z)$ — вторая функция нормированного отклонения.

Если стандарт больше средней, то доля особей, превышающих значение стандарта равна

$$0{,}5-\varphi(z).$$

Если стандарт меньше средней, то доля особей, превышающих значение стандарта равна

$$0{,}5+\varphi(z).$$

Закономерности и три функции нормированного распределения варьирующего признака используются при планировании селекционного процесса. При этом определяют:

- теоретическую величину σ;
- теоретические частоты распределения животных в совокупности по данному признаку при неизвестном их фактическом распределении, но заданных исходных значениях x и n;
- количество животных (%), которое оставляется для племенного использования при заданном уровне отбора;
- средние величины признака у животных, вошедших в племенное ядро (стада, породы, линии, кросса) при заданном уровне отбора (при данном селекционном дифференциале);
- границу отбора (селекционный дифференциал).

Распределение Стьюдента — это нормальное распределение которое зависит от одного параметра v — числа степеней свободы.

Рассмотрено в 1908 г. математиком В. Госсетом, который взял себе псевдоним «студент» (англ. *student*).

Распределение Стьюдента иначе называется t_{st}-распределение.

Формула критерия:

$$t_{st}=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}},$$

где $\bar{x}$ — средняя арифметическая выборки; μ — средняя арифметическая генеральной совокупности (генеральная средняя); $\frac{\sigma}{\sqrt{n}}$ — ошибка репрезентативности средней арифметической (стандартная ошибка средней арифметической).

В формуле числитель представляет собой отклонение выборочной средней от генеральной средней μ, а знаменатель является стандартной ошибкой средней арифметической.

График распределения Стьюдента, как и график нормального распределения, представляет собой куполообразную симметричную кривую, но он более полог и имеет бóльшую площадь под кривой.

При увеличении объема выборки распределение Стьюдента стремится к нормальному распределению и переходит в него при объеме совокупности, равном бесконечности. На практике такой переход фактически наблюдается при объемах выборки более 100 вариантов.

Практическое значение распределения Стьюдента состоит в том, что по малым выборкам становится возможным проверять статистические гипотезы относительно параметров генеральной совокупности.

На практике значения критерия Стьюдента t_{st} берут из таблиц t_{st}. В этих таблицах в одном столбце даются значения числа степеней свободы v, а в других — значения критерия для стандартных уровней надежности (0,95; 0,99; 0,999) или уровней значимости (0,05; 0,01; 0,001).

Асимметричное и эксцессивное распределение (проверка нормальности распределения). В отличие от нормального распределения встречаются еще распределения асимметричные и эксцессивные. При асимметричном распределении на графике (рис. 5.3) получается кривая, скошенная вправо или влево. В первом случае распределение будет положительным, или правосторонним, а во втором случае — отрицательным, или левосторонним. Такой тип распределения может отражать влияние каких-то факторов (уровень кормления, интенсивность отбора), изменяющих нормальное распределение и вызывающих асимметрию, т. е. накопление частот в левой или правой части кривой.

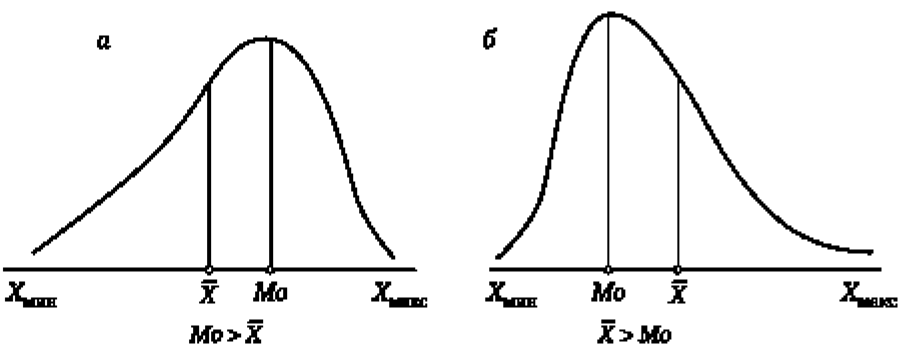

Рис. 5.3. **Асимметричное распределение:**
а — отрицательная асимметрия; *б* — положительная асимметрия

Асимметрия может быть и следствием неправильно сделанной выборки, что требует проведения нового отбора особей из генеральной совокупности.

Показатель асимметрии (As) вычисляется по формуле

$$As = \frac{\sum (X_i - \bar{X}_i)^3}{n \cdot \sigma^3}.$$

Ошибка показателя асимметрии вычисляется по формуле

$$m_A = \pm\sqrt{\frac{6}{n}}.$$

Эксцессивное распределение характеризуется значительным накоплением частот в классах, близких по величине к среднему значению признака (положительный эксцесс). На графике (рис. 5.4) это выражается крутовершинностью ветвей кривой. Эксцесс наблюдается также в виде плосковершинности и даже двухвершинности. 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.

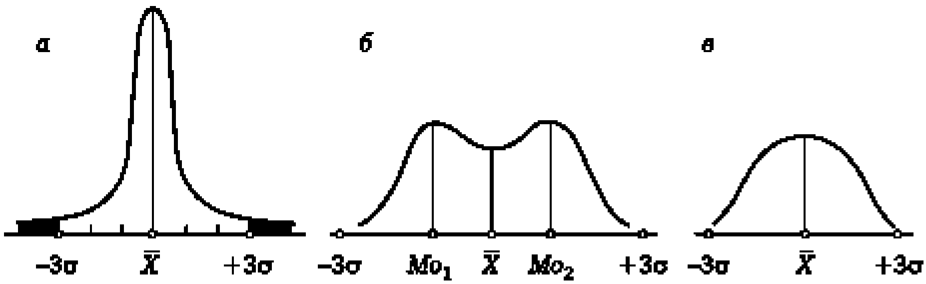

Рис. 5.4. **Эксцессивное распределение:**
а — положительное; *б* — отрицательное; *в* — плосковершинное

Двухвершинность в распределении молочного скота по содержанию жира в молоке может явиться следствием того, что в стаде имеются помесные животные. Оба типа эксцесса могут возникнуть и в результате неправильно проведенной выборки, что недопустимо.

Показатель эксцесса (E) определяется по формуле

$$E = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} \cdot -3.$$

Ошибка показателя эксцесса (m_E) определяется по формуле

$$m_E = \pm 2 \cdot \sqrt{\frac{6}{n}}.$$

Распределение считается достоверно нормальным, если абсолютная величина показателей асимметрии и эксцесса меньше их ошибок репрезентативности в 3 раза и более.

Пример. Из стада крупного рогатого скота голштинской породы по таблице случайных чисел проведена малая выборка в размере пяти коров с массовой долей жира в молоке (табл. 5.3). Требуется проверить эту выборку на нормальность распределения. Необходимые рассчеты приведены в табл. 5.3.

Средняя арифметическая выборки $\bar{x} = 4{,}102$.

Стандартное отклонение $\sigma = 0{,}234$. По формуле вычисляем показатель асимметрии:

$$As = \frac{\sum (x_i - \bar{x})^3}{n\sigma^3} = \frac{0{,}0093}{0{,}064} = 0{,}145.$$

Таблица 5.3

Значение массовой доли жира в молоке коров голштинской породы

Корова	МДЖ, %	$x_i - \bar{x}$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	3,98	–0,12	–0,00182	0,002097274
2	3,84	–0,26	–0,01798	0,000429982
3	4,35	0,248	0,015253	0,002176782
4	3,99	–0,11	–0,0014	0,002176782
5	4,35	0,248	0,015253	0,0037827420
$N = 5$	$\bar{x} = 4{,}10$	—	$\sum = 0{,}0093$	$\sum = 0{,}00691$

Ошибка показателя асимметрии:

$$m_A = \sqrt{\frac{6}{n}} = 1{,}095.$$

По формуле E вычисляем показатель эксцесса:

$$E = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} - 3 = \frac{0{,}0069}{0{,}1495} - 3 = -2{,}95.$$

Ошибка показателя эксцесса:

$$m_E = 2\sqrt{\frac{6}{n}} = 2{,}19.$$

Как показывают расчеты, показатель асимметрии в 7,5 раза меньше своей ошибки, что свидетельствует о достоверно нормальном распределении этой выборки. Показатель эксцесса свидетельствует о существенном отклонении от нормального распределения.

Трансгрессивное распределение. При трансгрессивном распределении классы одного варьирующего признака, например классы в минимальных его величинах, являются в то же время классами максимального значения другого вариационного ряда. Если изобразить его графически, то одна кривая как бы частично накладывается на другую, образуя трансгрессирующую зону (заштрихована) с одинаковыми классами для части обеих кривых. Это является первой особенностью трансгрессии вариационных рядов. Вторая особенность трансгрессии заключается в том, что средние арифметические $\bar{x}_1$ и $\bar{x}_2$ вариационных рядов достоверно различаются.

В зависимости от целей селекции можно увеличить или уменьшить степень трансгрессии. Для определения степени трансгрессии используют ее коэффициент, вычисляемый по формуле

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

где n_1 и n_2 — число членов каждой совокупности; p_1 и p_2 — доли частот, которые входят в трансгрессию.

Эти доли определяют с помощью функции нормального распределения $\varphi(z)$ по формулам

$$p_1 = \pm\varphi(z_1) \text{ и } p_2 = \pm\varphi(z_2),$$

где z_1 и z_2 — нормированные отклонения варьирующего признака, ордината которого отсекает трансгрессирующую часть кривой (рис. 5.5).

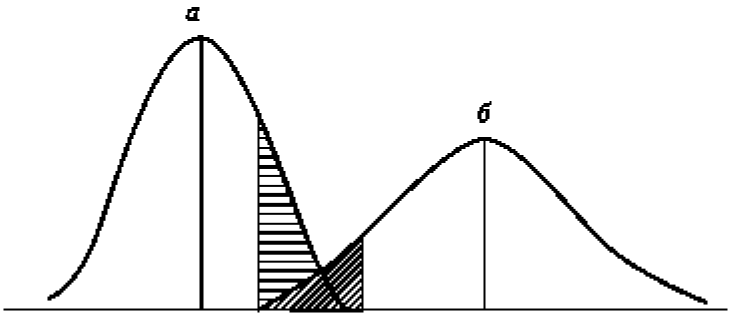

Рис. 5.5. **Трансгрессивное распределение**

5.7. Повторные измерения. Критерий Фридмана

Если одна и та же группа больных последовательно подвергается нескольким методам лечения или просто наблюдается в разные моменты времени, то применяют дисперсионный анализ повторных измерений. Однако для того чтобы использование дисперсионного анализа было правомерно, данные должны подчиняться нормальному распределению. Если в этом нет уверенности, то лучше воспользоваться критерием Фридмана — непараметрическим аналогом дисперсионного анализа повторных измерений.

Каждый больной ровно один раз подвергается каждому методу лечения (или наблюдается в фиксированные моменты времени). Результаты наблюдений у каждого больного упорядочиваются независимо от всех остальных. В результате, получается столько упорядоченных рядов, сколько больных участвует в исследовании. Далее, для каждого метода лечения (или момента наблюдения) вычисляется сумма рангов. Если разброс сумм велик, то различия статистически значимы. Формула критерия Фридмана имеет вид

$$x^2 = \frac{12}{nk(k+1)}S = \frac{12}{nk(k+1)}\sum\left(R_M - \frac{n(k+1)}{2}\right)^2.$$

Расчет критерия Фридмана:

1) значения для каждого больного располагаются по возрастанию, каждому значению присваивается ранг;

2) для каждого из методов лечения подсчитывается сумма присвоенных ему рангов;

3) вычисляется значение χ_r^2;

4) если рассчитанное значение χ_r^2 превышает критическое, то различия статистически значимы.

5.8. Множественное сравнение после применения критерия Фридмана

Как обычно, за выявлением различий между несколькими методами лечения должно последовать выяснение, в чем состоят эти различия, т. е. попарное сравнение методов лечения. Поскольку число больных, подвергшихся каждому методу лечения, одинаково, для этой цели легко приспособить критерий Ньюмена — Кейлса. Если считать один из методов лечения «контролем», то остальные можно сравнить с ним при помощи критерия Даннета. Для попарного сравнения методов лечения (или моментов наблюдения) применяется критерий Ньюмена — Кейлса:

$$q=\frac{R_A-R_B}{\sqrt{\frac{nl(l+1)}{12}}},\quad q=\frac{R_A-R_B}{\sqrt{\frac{s_{ВНУ}^2}{2}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}},$$

где R_A и R_B — суммы рангов для двух сравниваемых методов лечения; $s_{ВНУ}^2/2$ — внутригрупповая дисперсия; а n_A и n_B — численность групп; l — интервал сравнения; n — число больных.

Найденное значение q сравнивается с критическим из соответствующей таблицы для бесконечного числа степеней свободы. Если найденное значение больше критического, различие методов лечения (моментов наблюдения) статистически значимо.

Чувствительность критерия Ньюмена — Кейлса выше, чем критерия Стьюдента с поправкой Бонферрони.

При использовании этого критерия сначала необходимо с помощью дисперсионного анализа проверить нулевую гипотезу о равенстве всех средних. Если она отвергается, то все средние упорядочивают по возрастанию и сравнивают попарно, каждый раз вычисляя значение критерия Ньюмена — Кейлса.

5.9. Множественные сравнения с контрольной группой

Иногда задача заключается в том, чтобы сравнить несколько групп с единственной — контрольной. Для этого существуют специальные методы сравнения — это модификация критерия Стьюдента с поправкой Бонферрони и критерий Даннета. Как и другие методы множественного сравнения, их применяют только после того, как с помощью дисперсионного анализа отвергнута нулевая гипотеза о равенстве всех средних.

Критерий Даннета:

$$q' = \frac{R_{кон} - R_A}{\sqrt{\frac{nl(l+1)}{6}}},\quad q' = \frac{R_A - R_B}{\sqrt{\frac{s^2_{ВНУ}}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}},$$

где l — число всех групп, включая контрольную; $R_{кон}$ — сумма рангов в контрольной группе.

Остальные величины определяются, как в формуле для q. Значение q' сравнивается с критическим (табличным) для бесконечного числа степеней свободы.

Критерий Даннета — это вариант критерия Ньюмена — Кейлса для сравнения нескольких групп с одной контрольной. Он вычисляется как

$$q' = \frac{R_A - R_B}{\sqrt{\frac{s^2_{ВНУ}}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}.$$

Число сравнений равно числу групп, не считая контрольной, и существенно меньше числа сравнений в исходном критерии Ньюмена — Кейлса. Соответственно меньше и критические значения, как и в критерии Ньюмена — Кейлса, сначала средние значения для всех групп упорядочиваются, но только теперь — по абсолютной величине их отличия от контрольной группы. Затем контрольную группу сравнивают с остальными, начиная с наиболее отличной от контрольной. Если различия с очередной группой не найдены, то вычисления прекращают. Параметр l постоянен и равен числу групп, включая контрольную. Число степеней свободы вычисляют как в критерии Ньюмена — Кейлса: $\nu = N - m$.

5.10. Анализ выживаемости

Под *анализом выживаемости* подразумевается изучение закономерностей появления ожидаемого события в наблюдаемой выборке во времени. При этом ожидаемым событием может быть рецидив заболевания или, наоборот, выздоровление. Точка отсчета — это дата (час) выполнения определенной процедуры и т. п.

Время ожидания — это период времени от начального события до итогового. Это может быть выздоровление или летальный исход.

Отличие анализа выживаемости от других статистических методов оценки ожидаемого события — в способе построения выборки, т. е. в анализе времени до события они могут меняться. Время до события становится определенным только среди тех объектов, у которых событие произошло. До наступления события этот показатель остается неизвестным для всех остальных объектов. При этом по различным причинам изучаемые объекты могут выбыть из анализа. В таких случаях применяются специальные методы, позволяющие учитывать и использовать неполную информацию, т. е. цензурированные данные. Они используются в тех случаях, когда наблюдаемый параметр является временем до наступления события, а период наблюдения ограничен.

При анализе выживаемости вся информация о выборке содержится в соответствующей ей функции распределения вероятности (в данном случае — времен ожидания). Она используется в виде функции выживания.

Кумулятивная функция распределения $F(t)$ времен ожидания отражает вероятность того, что время ожидания события меньше t. Соответственно, функция выживания $S(t) = 1 - F(t)$ равна вероятности того, что событие не состоится ранее, чем по истечении времени t. Наиболее распространенными описательными методами исследования цензурированных данных являются построение таблиц дожития и метод Каплана — Мейера.

Таблицы дожития — один из наиболее традиционных методов исследования данных о выживаемости (происхождение интересующего нас события).

В таблицах дожития время наступления события разбивается на интервалы, для каждого из которых определяется число и доля объектов:

1) объекты, у которых событие не произошло на момент начала данного интервала времени;

2) объекты, у которых событие произошло в течение данного интервала;

3) объекты, которые были изъяты или цензурированы на данном интервале.

Таблица дожития является расширенной таблицей частот. Для получения надежных оценок основных показателей (функции выживания, плотности вероятности и интенсивности), размер группы должен быть не менее 30 объектов. На основании таблицы рассчитывается ряд индикаторов. Число изучаемых объектов — число объектов, у которых событие не произошло на момент начала данного интервала времени минус половина числа объектов, которые были изъяты или цензурированы. Доля «умерших» — отношение числа объектов, у которых событие произошло в течение данного интервала, к числу изучаемых объектов на данном временном интервале. Доля выживших — единица минус доля «умерших».

Функция выживания (*выживаемость*) — кумулятивная доля объектов, событие у которых не произошло на момент начала определенного интервала времени; ее рассчитывают как произведение долей выживших на всех предыдущих интервалах.

Плотность вероятности — оценка вероятности наступления события в каком-либо интервале; рассчитывается как отношение разности между значениями функции выживания на любом данном и последующем интервале к продолжительности данного интервала времени.

Функция интенсивности представляет собой вероятность того, что на данном интервале произойдет событие у того объекта, у которого оно еще не произошло на момент начала этого интервала. Эта функция вычисляется как отношение числа событий, происшедших в течение данного интервала, к числу объектов, у которых событие не произошло до момента времени, находящегося в середине этого интервала.

Медиана ожидаемого времени жизни — точка на оси времени, в которой значение функции выживания равна 0,5; медиана ожидаемого времени жизни совпадает с точкой выживания 50 % наблюдений только в том случае, если до этого момента времени цензурированных наблюдений не было (рис. 5.6).

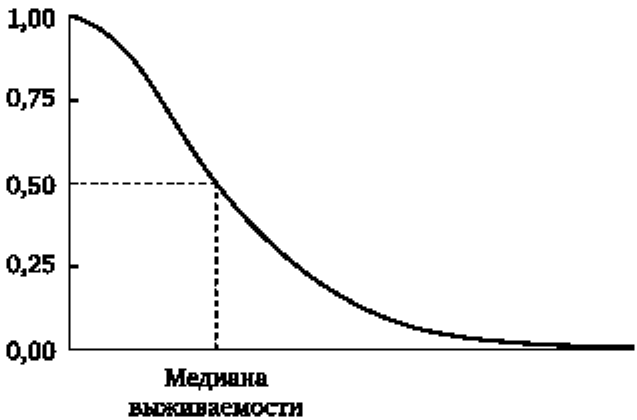

Рис. 5.6. **Кривая выживаемости**

Метод Каплана — Мейера используется для оценки доли объектов наблюдения (пациентов), у которых событие не произошло (функция выживания, выживаемость) для любого момента времени в течение всего периода наблюдения. Поскольку разбиение данных по временным интервалам (группировка) не производится, суть метода Каплана — Мейера несколько отличается от таблиц дожития. В то же самое время результаты, получаемые с помощью этих двух методов, принципиально близки по смыслу.

Оценка функции выживания в методе Каплана — Мейера представляет собой произведение выживаемости в данный момент времени на выживаемость в следующий момент времени, когда событие произошло. Как и таблицы дожития, метод Каплана — Мейера полностью применим к цензурированным данным. Для расчетов используется истинное количество объектов, у которых событие еще не произошло в любой момент времени, для которого производится оценка.

Формула Каплана — Мейера:

$$\hat{S}(t)=\prod j:\tau_j<t\frac{r_j-d_j}{r_j}=\prod j:\tau_j<t\left(1-\frac{d_j}{r_j}\right),$$

где τ_j — моменты времени «умерших», наблюдаемые в выборке; d_j — число умерших в момент τ_j; r_j — число объектов, умерших или цензурированных в момент τ_j или позже.

При этом справедливы следующие соотношения:

$$r_j=r_j-1-d_j-1-c_j-1;\ r_j=\sum_{1\geq j}(c_1-d_1),$$

где c_j — число цензурированных объектов в промежутке между j-м и $(j+1)$ интервалами. Объекты, цензурированные в момент τ_j, включаются в c_j.

В случае отсутствия цензурирования $\hat{S}(t) = S(t)$. $\hat{S}(t)$ имеет асимптотическое нормальное распределение. $\hat{S}(t)$ является асимптотически несмещенной оценкой $S(t)$. Дисперсия $\hat{S}(t)$ вычисляется по формуле Гринвуда:

$$\operatorname{var}\hat{S}(t)=[\hat{S}(t)]^2\sum_{j:\tau j<t}\frac{d_j}{(d_j-r_j)r_j}.$$

Метод Каплана — Мейера широко используется в клинических испытаниях, например с целью оценки эффективности нового лекарственного препарата в изучаемой группе по сравнению с контрольной группой.

Задания для самостоятельной работы

Задание 1. В стаде коров черно-пестрой породы численностью 6000 коров лимит жирномолочности составляет $\bar{X}_{min}$ = 3,4 % и $\bar{X}_{max}$ = = 4,2 % при средней жирномолочности 3,8 %. Следует определить теоретическую величину σ, распределение животных по данному признаку в стаде и установить, какое количество коров можно выделить в племенную группу, если границей отбора считать показатель селекционного дифференциала, равный $+0{,}25\sigma$.

Задание 2. Найти среднюю арифметическую по показателю содержания белка в молоке коров неплеменной части стада ($\bar{X}_p$) и племенной части ($\bar{X}_q$), если $\bar{X}_{общ}$ = 3,30 %, σ = 0,30 %, стандарт племенного класса — 3,60 % и t для этого стандарта равно 1,0.

Задание 3. По данным бонитировок в племенных хозяйствах удои коров черно-пестрой породы по высшей лактации варьируют в пределах от 3500 до 6500 кг, а средний удой составляет 5000 кг. Общее поголовье коров, по которым учтена высшая лактация, равно 6000 голов.

Имея эти данные, следует провести прогнозирование численности животных племенных стад с определенным уровнем их продуктивности. Для этого необходимо составить вариационный ряд с определением теоретических частот по классам удоя коров.

Задание 4. Определить теоретические частоты вариационного ряда по показателю настрига шерсти с тонкорунных овец, если эмпирические частоты вариационного ряда следующие:

W — середина классов	3,5	4,0	4,5	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5	n
P — эмпирические частоты	1	2	3	10	12	8	6	5	6	3	2	58

Задание 5. В популяции коров айрширской породы численностью 5000 коров известен лимит жирномолочности ($\bar{X}_{min}$ = 3,7 % и $\bar{X}_{max}$ = = 4,5 % при средней жирномолочности 4,05 %).Следует определить теоретическую величину σ, теоретическое распределение животных по этому признаку в популяции и установить, какое количество коров можно выделить в селекционную группу, если границей отбора взять показатель селекционного дифференциала, равный +0,2σ.

Задание 6. У пчел определенной породы длина хоботка $\bar{X}$ = 6,0 мм и σ = 0,3 мм, а для того, чтобы легко опылять клевер, требуется хоботок длиной 6,6 мм. Какое количество пчел будет опылять этот сорт клевера?

Задание 7. Из большой партии яиц необходимо отобрать лучшую часть с весом 70 г и более и предусмотреть, какие средние веса будет в отобранной и оставшейся, худшей части. Известно, что эта партия получена от породы кур, которая в условиях данной птицефабрики обычно дает яйца со средней массой $\bar{X}$ = 70 г и σ = 10 г.

Задание 8. Пчел, у которых длина хоботка $\bar{X}$ = 6,4 мм и σ = 0,3 мм, перевели на такой сорт клевера, для опыления которого им достаточно иметь хоботок 6,1 мм. Какое количество пчел будет опылять этот сорт клевера?

Задание 9. Из большого стада овец со средним настригом шерсти $\bar{X}$ = 4,0 кг и σ = 1,0 кг решено отобрать элитную часть овец, дающих 5,0 кг и более шерсти за год. Требуется предусмотреть, какой при этом можно ожидать средний настриг у отобранной и у оставшейся части овец.

Задание 10. При изучении соотношения полов жеребят у 231 матки орловской рысистой породы получено следующее распределение (по Н. А. Плохинскому):

Число жеребчиков p	0	1	2	3	4	5	6	7	8	9	10
Число кобылок q	10	9	8	7	6	5	4	3	2	1	0
Число матерей n_i	0	4	3	33	45	56	44	28	14	3	0

Всего маток Σn_i = 231

Используя формулу бинома Ньютона $(p + q)^n$, определить теоретическое распределение жеребчиков и кобылок в потомстве (приплоде) 231 матки.

Задание 11. В одном хозяйстве изучено распределение семейств по количеству больных маститом коров. Распределение семейств по количеству больных маститом коров следующее:

Количество больных (x)	5	4	3	2	1	0	$k = 5$
Число семейств (p)	1	1	3	12	5	3	$\Sigma p = 25$

Вычислить теоретическую частоту распределения семейств по комплексу больных маститом коров.

Задание 12. Изучить распределение 25 потомков кроликов по доминантной окраске шерсти. Каждый помет состоит из четырех крольчат. Возможные варианты числа доминантно окрашенных крольчат будут следующими: нет доминантных крольчат (О); один крольчонок; два, три, четыре доминантных крольчонка. Ниже приведен вариационный ряд фактического распределения пометов по числу крольчат с доминантной окраской:

Число доминантных крольчат в помете (x)	0	1	2	3	4	$k = 4$
Число пометов (p)	2	4	5	10	4	$\Sigma p = 25$

Установить по биному Ньютона теоретические частоты распределения окролов по окраске крольчат.

Задание 13. Обследовано 50 пометов норок с равным числом щенков — по пять голов. Среди пометов могли быть следующие варианты: 0 — отсутствие щенков с рецессивной окраской — 11 пометов; с одним рецессивным щенком — 10; с двумя — 25 пометов; с тремя — два помета; с четырьмя — один помет и один помет, в котором все пять щенков были рецессивного типа. Определить теоретическую вероятность рождения рецессивных щенков в количестве пяти, четырех, трех, двух, одного на помет или ноль и абсолютные числа теоретических частот.

Задание 14. В группе из 100 коров каждая корова имела по четыре отела. Среди них двойневые отелы имели всего 10 коров. Вариационный ряд распределения коров по количеству двойневых отелов имеет следующий вид:

Число двойневых отелов (x)	0	1	2	3
Число коров (n)	90	6	3	1

Определить теоретическое распределение коров по количеству двойневых отелов.

Задание 15. В популяции бестужевской породы вероятность появления наследственного уродства $p = 0{,}002$. Определить вероятность появления среди 200 телят 3; 2; 1; 0 уродств.

Задание 16. Исследовано 200 групп лошадей. В 109 группах не было зарегистрировано ни одной смерти от удара лошадью, в 65 группах было по одному такому случаю, в 22 группах — по два случая, в трех группах — по три и в одной группе — четыре случая смерти от удара лошади. Составить эмпирическое и теоретическое распределение числа смертей от удара лошади.

Задание 17. Пробонитировано 100 больших групп коров симментальской породы. В 55 группах не было обнаружено ни одного случая наличия трехсосковых коров, в 30 группах было по одному такому случаю, в 11 группах — по два случая, в трех группах — по три и в одной группе — четыре коровы с тремя сосками.

Составить теоретическое распределение числа коров с тремя сосками.

Контрольные вопросы и задания

1. Какие закономерности характеризуют нормальную кривую?
2. Какие три функции нормального распределения вы знаете? Что они показывают?
3. Какой вид имеет теоретическая кривая нормального распределения?
4. Какие параметры позволяют определить закономерности и три функции нормального распределения варьирующего признака?
5. Для каких целей используется распределение Пуассона?
6. Какую форму имеет распределение Пуассона при графическом изображении?
7. Дайте характеристику асимметричного и эксцессивного распределения вариационного ряда.
8. Что понимается под биноминальном распределением?
9. Дайте характеристику биноминального распределения.
10. Что такое треугольник Паскаля? Как его можно применить в биологических исследованиях?
11. Какие параметры служат характеристикой биноминального распределения?
12. Что такое трансгрессивное распределение? Как оно применяется в селекции растений и животных?
13. Какие особенности трансгрессии вариационных рядов вы знаете?
14. Как выглядит графически трансгрессивное распределение?
15. Какую формулу используют для определения степени трансгрессии?
16. Назовите основные свойства вероятностей. Какова их характеристика?
17. Перечислите типы статистических распределений вариационных рядов. Какова их характеристика?
18. Что понимается под анализом выживаемости?
19. Что характеризует время ожидания при анализе выживаемости?
20. Что такое таблицы дожития при анализе выживаемости?
21. Как выглядит кривая выживаемости?
22. В каких случаях используется метод Каплана — Мейера?

Глава 6
КОРРЕЛЯЦИЯ МЕЖДУ ПРИЗНАКАМИ

Наследование признаков часто рассматривается отдельно, независимо друг от друга. Между тем в природе многие явления, а в организме многие свойства и признаки находятся во взаимосвязи и взаимодействии. Некоторые же признаки совершенно независимы от продуктивности животных. Например, удой и содержание жира в молоке не связаны с мастью животных, степенью пигментации шерстного покрова, наличием или отсутствием рогов. При огромном разнообразии окраски оперения кур уровень яичной продуктивности у них более или менее одинаков.

Однако, как показывают исследования, случаи дискретного проявления признаков встречаются сравнительно редко. Гораздо чаще проявляются весьма сложные взаимосвязи различных признаков.

Вопрос о характере связей различных признаков интересовал ученых давно. Так, в 1836 г. французский биолог Ж. Кювье открыл закон корреляции развития органов и тканей, согласно которому изменение одной части организма влечет за собой изменение других его частей.

Огромный вклад в учение о корреляции признаков внес Ч. Дарвин, сформулировавший закон соотносительной (коррелятивной) изменчивости. Однако, как показал Дарвин, существующие в организме взаимосвязи не абсолютны, не вечны, а находятся под контролем естественного и искусственного отбора. Е. А. Богданов со всей определенностью указывал, что можно путем племенного подбора и отбора изменять прочно сложившиеся связи и создавать новые.

Взаимосвязь между отдельными признаками называется корреляцией. Установлено, что эта связь проявляется не у всех животных и не в одинаковой степени. Например, не всегда с увеличением живой массы коров повышаются надои молока. Иногда у отдельных особей можно наблюдать и противоположную взаимосвязь, т. е. с увеличением живой массы коров удой их снижается. Среди животных, у которых с повышением живой массы увеличивается обхват груди, встречаются такие, у которых грудь развита слабо. В массе, чем выше удой матерей, тем выше удой дочерей, но в отдельных случаях бывает, что от высокомолочной матери рождается низкопродуктивная дочь. С повышением плодовитости свиноматок средний вес поросенка в гнезде снижается, что делает их дальнейшее выращивание и откорм неперспективными. В то же время встречаются свиноматки, у которых при высокой плодовитости (15—18 поросят) в гнезде имеет место и высокая крупноплодность. Отбирая таких свиноматок, можно длительной селекцией закрепить в потомстве желаемую взаимосвязь полезных признаков (сочетание высокой плодовитости с достаточной крупноплодностью), т. е. создавать группы животных с желательными соотношениями признаков: высокий процент жира и высокий удой, большая длина и густота шерсти, высокая яйценоскость и крупное яйцо и т. д.

Обнаружить наличие или отсутствие корреляционной связи между какими-либо признаками возможно лишь на основании обследования достаточно большой группы животных и в результате обработки полученных цифр определенным биометрическим методом.

По форме корреляция может быть прямолинейной и криволинейной, по направлению — прямой (положительной) и обратной (отрицательной).

При прямолинейной связи равномерным изменениям одного признака соответствуют равномерные изменения второго признака при незначительных отклонениях. Например, при увеличении длины тела на 1 см ширина его также увеличивается на определенную величину.

При криволинейной связи с увеличением одного признака другой увеличивается до определенной величины, а затем уменьшается (или наоборот). Криволинейной является связь удоев с возрастом животных, удоя с месяцем лактации и др.

Положительной корреляцией называется такая, при которой с увеличением (или уменьшением) одного признака связанный с ним другой признак также увеличивается (или уменьшается). Например, с увеличением живой массы коров-матерей в среднем увеличивается и масса получаемых от них телят, с повышением живой массы увеличивается обхват груди, с повышением суточных удоев коров растут и их годовые удои и т. д.

Отрицательной называется такая корреляция, при которой с увеличением одного признака связанный с ним другой признак уменьшается или, наоборот, с уменьшением одного другой увеличивается. Так, с увеличением длины шерсти овец снижается ее густота, с увеличением удоя коров жирномолочность уменьшается, с увеличением количества скормленного корма переваримость снижается и т. д.

Для оценки связи между биологическими признаками и свойствами используется коэффициент корреляции, который обозначается буквой r (от англ. *relation* — отношение, зависимость). Он является величиной неименованной и определяет направление и силу только прямолинейной корреляции. Силу связи показывает абсолютная величина r, а ее направление — знак перед ней. В случае криволинейной корреляционной связи вычисляется другой показатель, который называется корреляционным отношением — η.

Размер положительного и отрицательного коэффициентов корреляции колеблется от нуля до единицы. При полной положительной корреляции $r = +1$ (связь становится функциональной). При отсутствии взаимосвязи $r = 0$. При полной отрицательной корреляции $r = -1$. Однако в природе не наблюдается как полной корреляции между признаками, так и абсолютного отсутствия ее. Эти понятия сугубо теоретические.

Различают низкую (слабую), среднюю и высокую (сильную, тесную) корреляционную зависимость:

- если $r > 0{,}7$, то связь считается сильной;
- если $r < 0{,}3$ — слабой;
- если $r = 0{,}3 — 0{,}7$ — средней.

Изучение связи между признаками имеет большое значение при решении генетико-селекционных вопросов. Установление фенотипической и генотипической связи между признаками позволяет вести косвенную селекцию по коррелирующим признакам и используется для прогноза селекции.

6.1. Вычисление коэффициента фенотипической корреляции для малочисленных выборок ($n \leq 30$)

В малых выборках коэффициент фенотипической корреляции вычисляют по формуле

$$r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{C_x \cdot C_y}} \text{ или } r = \frac{C_x + C_y - C_d}{2\sqrt{C_x \cdot C_y}},$$

где x и y — значения вариант первого и второго признака; n — число животных, изучаемых по двум признакам; C — сумма квадратов центральных отклонений по каждому признаку (например, C_x — сумма квадратов центральных отклонений по живой массе, C_y — по величине удоя и их разности — C_d); $d = x - y$.

Сумма квадратов центральных отклонений вычисляется по формуле

$$C_x = \sum x^2 - \frac{(\sum x)^2}{n}; \; C_y = \sum y^2 - \frac{(\sum y)^2}{n}; \; C_d = \sum d^2 - \frac{(\sum d)^2}{n}.$$

Пример. Нужно установить направление и силу связи между живой массой и обхватом груди свиноматок украинской степной белой породы (в малых выборках) $n = 15$ голов; x — варианты 1-го признака — живая масса, кг; y — варианты 2-го признака — обхват груди, см.

Данные о живой массе (x) и обхвате груди (y) свиноматок украинской степной белой породы:

x	y	x	y	x	y	x	y	x	y
200	140	223	144	220	145	225	138	253	146
248	145	227	142	226	140	243	144	226	142
251	158	220	145	291	160	228	143	265	151

Расчеты по определению коэффициента корреляции приведены в табл. 6.1.

Таблица 6.1

Расчет коэффициента корреляции в малых выборках

x	y	xy	x^2	y^2	$d(x-y)$	$d^2(x-y)^2$
200	140	28 000	40 000	19 600	60	3600
248	145	35 960	61 504	21 026	103	10 609
251	158	39 658	63 001	24 964	93	8649
233	144	33 552	54 289	20 736	89	7921
227	142	32 234	51 529	20 164	85	7226
220	145	31 900	48 400	21 025	75	5625
220	145	31 900	48 400	21 025	75	5625
226	140	31 640	51 076	19 600	86	7396
291	160	46 560	84 681	25 600	131	17 161
225	138	31 050	50 625	19 044	87	7569
243	144	34 992	59 049	20 736	99	9801
228	143	32 604	51 984	20 449	85	7225
253	146	36 938	64 009	21 316	107	11 449
226	142	32 092	51 076	20 164	84	7056
265	151	40 015	70 225	22 801	114	12 996
$\sum_x=3556$	$\sum_y=2183$	$\sum_{xy}=519\,095$	$\sum_{x^2}=849\,848$	$\sum_{y^2}=318\,250$	$\sum_d=1373$	$\sum_{d^2}=129\,908$

$$C_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 849\,848 - \frac{12\,645\,136}{15} = 6839;$$

$$C_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 318\,250 - \frac{4\,765\,489}{15} = 551;$$

$$C_d = \sum d^2 - \frac{\left(\sum d\right)^2}{n} = 129\,908 - \frac{1\,885\,129}{15} = 4233;$$

$$r = \frac{C_x + C_y - C_d}{2\sqrt{C_x \cdot C_y}} = \frac{6839 + 551 - 4233}{2\sqrt{6839 \cdot 551}} = \frac{3157}{3882} = 0{,}813.$$

Вывод: между живой массой и обхватом груди свиноматок украинской степной белой породы имеется высокая положительная связь.

Задания для самостоятельной работы

Задание 1. Вычислить коэффициент фенотипической корреляции между возрастом (x — количество опоросов) и числом поросят в помете (y) у свиноматок крупной белой породы:

x	y	x	y	x	y	x	y	x	y
2	9	7	10	6	11	3	14	4	13
1	7	3	11	1	6	2	9	5	10
5	11	2	8	4	12	1	8	3	11

Задание 2. Вычислить коэффициент фенотипической корреляции между живой массой (x) и удоем (y) у 15 коров черно-пестрой породы, кг:

x	y	x	y	x	y	x	y	x	y
520	3120	470	3100	550	3160	500	3180	480	2880
490	3130	560	3190	510	3170	530	3305	500	3200
540	3150	480	3110	530	3140	495	2950	550	3350

Задание 3. Вычислить коэффициент фенотипической корреляции между живой массой (x) и настригом шерсти (y) у пятнадцати овец по следующим данным, кг:

x	y	x	y	x	y	x	y	x	y
70	4,0	74	5,5	80	8,5	80	7,5	83	7,8
71	4,5	75	6,0	85	8,0	68	3,8	80	8,2
72	5,0	70	6,0	75	7,0	72	4,5	74	6,5

Задание 4. Вычислить коэффициент фенотипической корреляции между числом эритроцитов (x, млн) и содержанием гемоглобина (y, г %) в крови овец по приведенным ниже данным:

x	y	x	y	x	y	x	y	x	y
5,8	10,0	9,8	13,0	7,2	10,1	8,0	11,0	26,8	10,5
8,3	11,6	6,2	9,6	8,6	12,2	6,4	10,0	7,1	10,9
6,0	9,5	7,4	11,0	7,7	10,5	7,0	9,6	8,2	11,4

Задание 5. Определить по данным следующей выборки, имеется ли корреляция между плодовитостью самок серебристо-черных лисиц (x) и плодовитостью их дочерей (y), голов:

x	y	x	y	x	y	x	y	x	y
6	7	5	3	6	7	4	2	7	4
7	5	6	6	6	8	6	5	6	7
5	6	5	2	3	5	4	6	5	5

Задание 6. Вычислить коэффициент фенотипической корреляции между живой массой матерей в возрасте первого отела (x) и живой массой телят при рождении (y), кг:

x	y	x	y	x	y	x	y	x	y
380	20	412	28	480	49	413	31	423	23
479	45	483	45	475	44	415	29	485	48
500	48	446	44	390	23	438	33	426	32
405	26	395	28	453	35	450	42	487	46
463	48	493	50	487	38	395	30	473	33

Задание 7. Определите, какая корреляция имеется между незаменимыми аминокислотами лизином (x) и аргинином (y), входящим в состав белка молока коров, по данным следующей таблицы, мг %:

x	y	x	y	x	y	x	y	x	y
2,45	1,02	2,39	0,80	2,55	1,02	2,60	0,92	2,50	0,92
2,89	1,25	2,60	0,98	2,40	0,88	2,56	0,98	2,70	1,13
2,85	1,10	2,96	1,16	2,90	1,16	2,28	0,90	2,90	1,71
2,50	0,85	2,45	1,00	2,72	1,16	2,72	1,15	1,90	0,62
2,60	0,90	1,90	0,65	2,50	0,80	2,40	0,80	2,42	1,09
2,39	1,81	2,54	0,93	2,50	0,80	2,54	1,03	2,85	1,10

Задание 8. Вычислить коэффициент корреляции между содержанием в сперме быков липоидного фосфора (x, мг %) и концентрацией сперматозоидов (y, млрд в 1 мл):

x	y	x	y	x	y	x	y	x	y
1,20	21,0	1,02	24,3	1,76	28,8	0,92	18,2	1,75	29,0
1,29	25,3	1,35	31,7	1,34	25,3	1,24	23,3	1,82	19,8
1,90	19,0	2,06	35,8	1,18	22,3	1,68	26,3	1,46	32,5

Задание 9. Вычислить коэффициент корреляции между содержанием лейкоцитов в крови коров (x, тыс.) и количеством лимфоцитов (y, %):

x	y	x	y	x	y	x	y	x	y
11,2	71,6	16,6	75,5	11,0	90,0	11,5	73,0	11,6	82,5
8,8	72,0	10,6	82,0	12,6	68,8	9,3	74,8	12,4	70,6
8,2	67,5	11,2	76,3	13,5	75,4	10,5	76,5	13,0	73,8
10,0	78,1	13,6	74,1	14,0	74,3	13,4	73,4	11,2	75,3

Задание 10. Вычислить коэффициент корреляции между многоплодием матерей (x) и дочерей (y) свиней крупной белой породы:

x	y	x	y	x	y	x	y	x	y
10,7	11,9	9,5	10,6	9,5	10,6	9,7	10,8	9,2	10,2
11,9	13,2	10,4	11,6	9,8	10,9	9,4	10,4	10,8	12,0
9,2	10,2	9,2	10,2	9,0	10,0	11,9	13,2	10,4	11,6
10,4	11,6	10,8	12,0	8,8	9,8	10,4	11,6	11,9	13,2
10,8	12,0	9,7	10,8	9,0	10,0	9,5	10,6	10,4	11,6
10,8	12,0	9,7	10,8	9,8	10,9	8,8	9,8	9,4	10,4

Задание 11. Вычислить коэффициент корреляции между жирностью молока коров матерей (x) и дочерей (y) бестужевской породы,%:

x	y	x	y	x	y	x	y	x	y
4,40	3,94	3,93	3,85	4,08	3,92	4,11	4,04	4,35	4,20
3,58	4,36	3,70	3,81	4,91	4,07	4,36	3,98	3,90	4,05
4,87	4,35	4,20	3,94	4,08	4,36	3,86	4,12	3,76	3,92

Задание 12. Вычислить коэффициент корреляции между живой массой самок нутрий (x, кг) и средней массой щенят (y, г):

x	y	x	y	x	y	x	y	x	y
3,5	201	4,3	213	5,4	270	4,0	221	3,5	206
3,6	205	4,6	211	5,7	267	4,3	230	3,9	225
3,8	209	4,9	230	5,5	260	5,0	238	4,3	218
4,1	215	5,1	251	3,9	215	4,6	242	4,8	254

Задание 13. Вычислить коэффициент корреляции между средним удоем за три лактации матерей (x) и дочерей (y) коров, кг:

x	y	x	y	x	y	x	y	x	y
4376	4451	3688	4032	4035	4105	3514	4652	2580	4091
5078	4834	3949	4440	3212	3620	4144	3809	4145	4867
4299	4183	3957	3581	4145	4867	4738	4050	3820	4919
3733	3530	4000	3331	3868	4322	4552	5412	4144	3910
3089	3547	3661	3549	3820	4919	6013	4550	4552	5412

Задание 14. Вычислить коэффициент корреляции между длиной шерсти (x, см) и содержанием в организме овец связанного белкового йода (y, мг %):

x	y	x	y	x	y	x	y	x	y
15,5	2,1	13,0	5,1	10,2	4,5	9,5	5,8	12,0	4,4
15,0	3,6	13,2	5,6	11,5	4,3	10,0	5,1	11,4	4,7
15,2	4,0	9,0	7,1	9,5	4,7	10,5	5,0	13,2	4,3
14,0	4,2	12,5	3,9	10,0	4,8	9,5	5,5	12,2	4,7
13,0	5,5	9,0	6,0	10,5	4,5	10,5	5,1	10,8	5,2

Задание 15. Вычислить коэффициент корреляции между суточным удоем (x) и живой массой коров (y), кг:

x	y	x	y	x	y	x	y	x	y
28,8	512	21,0	479	12,3	380	21,9	428	31,2	560
20,2	472	25,5	515	21,4	465	17,8	447	23,9	459
21,4	489	21,7	451	18,9	485	20,0	412	27,0	548
20,6	482	20,8	450	21,8	458	21,1	560	20,9	457
23,7	468	25,3	500	20,9	413	27,5	542	25,9	517

6.2. Вычисление коэффициента фенотипической корреляции в больших выборках ($n \geq 30$)

Для определения коэффициента корреляции в большой выборке необходимо построить корреляционную решетку. По форме корреляционного поля делают предварительное заключение о характере связи, ее направлении и силе. Затем обрабатывают корреляционную решетку и находят коэффициент корреляции по формуле

$$r = \frac{\sum p \cdot a_x \cdot a_y - n \cdot b_x \cdot b_y}{n \cdot \sigma_x \cdot \sigma_y},$$

где r — коэффициент корреляции; a_x и a_y — отклонения классов от условного среднего по первому и второму признакам; p — частоты в корреляционной решетке; n — объем выборки; σ_x и σ_y — средние квадратические отклонения для каждого коррелирующего признака; b_x и b_y — поправки к условным средним по признакам x и y.

Например, необходимо вычислить коэффициент фенотипической корреляции между живой массой и обхватом груди коров бестужевской породы по следующим данным (табл. 6.2).

Таблица 6.2

Показатели живой массы и обхвата груди за лопатками коров бестужевской породы

Живая масса, кг	Обхват груди, см	Живая масса, кг	Обхват груди, см	Живая масса, кг	Обхват груди, см	Живая масса, кг	Обхват груди, см	Живая масса, кг	Обхват груди, см
602	196	590	202	650	205	659	195	653	203
500	180	725	205	663	210	566	188	670	204
495	188	635	195	670	199	821	216	695	210
494	196	555	194	735	210	643	197	667	205
597	208	570	197	680	207	650	198	597	197
443	191	695	200	600	205	637	204	578	192
770	206	640	200	645	210	828	210	532	192
860	215	670	205	625	204	582	193	645	203
470	180	560	197	625	205	634	192	560	197
560	194	654	201	590	197	760	202	605	200
735	216	655	202	680	210	849	218	625	200
650	212	600	200	590	192	736	208	660	210
575	200	610	200	635	203	787	210	527	192
615	202	610	200	600	203	745	206	670	208
697	211	612	202	630	204	864	214	666	205
591	205	690	216	685	211	732	206	645	205
515	187	555	198	678	205	654	206	573	197
615	200	785	215	553	190	687	197	680	195
675	210	625	197	600	193	715	205	590	204
598	205	563	195	680	202	640	200	713	200
680	206	670	197	615	194	750	208	705	202
601	200	680	214	740	210	686	203	540	196
570	194	670	212	600	194	730	207	750	201
685	208	615	198	670	201	600	202	674	198
533	195	675	201	660	202	642	198	520	190
593	192	580	197	715	206	721	206	590	200

Для построения корреляционной решетки необходимо предварительно выполнить следующие задачи:

1) условно обозначить один из признаков через x (живая масса, кг), а другой — через y (обхват груди, см);

2) подсчитать количество вариантов (животных) в выборке (n = 130).

Для каждого признака следует найти максимальное и минимальное значение и установить размах изменчивости.

Живая масса коров, кг: $\lim_x = \max_x - \min_x = 864 - 443 = 421$ кг.

Обхват груди, см: $\lim_y = \max_y - \min_y = 218 - 180 = 38$ см.

Вычислить величину классного промежутка по исходным признакам:

$$K_x = \frac{\lim_x}{l} = \frac{421}{10} = 42{,}1 \approx 40 \text{ кг};$$

$$K_y = \frac{\lim_y}{l} = \frac{38}{10} = 3{,}8 \approx 4 \text{ см}.$$

Для вычисления коэффициента корреляции между живой массой и обхватом их груди за лопатками необходимо построить корреляционную решетку (табл. 6.3) и заполнить ее в следующей последовательности (при этом желательно для обоих признаков брать одинаковое число классов, но допускается разница в два — три класса).

1. Построить классы по признаку x (живая масса) и по признаку y (обхват груди), использовав величину классного промежутка $k_x = 40$ кг, $k_y = 4$ см. В верхнюю строку решетки вписываются классы по обхвату груди,а с левой стороны — классы по живой массе, располагая их в порядке возрастания снизу вверх или сверху вниз;

2. Разнести варианты по клеткам корреляционной решетки с учетом значений у каждого объекта признаков одновременно.

Например, первая корова (варианта совокупности) имела живую массу 602 кг и обхват груди 196 см (признака два, а объект один и тот же). Находим соответствующие этим цифрам классы в решетке: по ряду x она относится к пятому классу (живая масса 600—639 кг), а по ряду y — также к пятому (обхват груди 196—199 см.). В квадрате на месте пересечения этих классов ставится точка. Таким образом, по решетке разносятся все остальные варианты.

Закончив разноску, нужно начертить корреляционную решетку заново, вписав в ее клетки соответствующие частоты и добавив для последующих расчетов четыре графы справа и четыре строки снизу.

3. Подсчитать количество вариантов по классам признака x, заполнить колонку P_x и по классам признака y, заполнив строку P_y (они должны быть равны между собой и соответствовать объему выборки $\sum n_1 = \sum n_2 = N = 130$ голов).

4. Выделить модальный класс с наибольшим количеством вариант и расположенный ближе к середине, как по признаку x, так и по признаку y, приняв эти классы за нулевое отклонение по колонкам a_x и a_y. Границы этих классов следует выделить полужирными линиями, в результате чего корреляционная решетка распадется на четыре квадранта (I—II—III—IV).

5. Проставить отклонения каждого класса от модального. Вверх по колонке a_x — отрицательное отклонение, вниз — положительное, влево, по строке a_y — отрицательное, вправо — положительное отклонения.

6. Заполнить колонку $P_x \cdot a_x$ и строку $P_y \cdot a_y$, умножив частоту на отклонение в каждом классе. Суммируя произведения со знаком минус и плюс, вычисляют их общую сумму $\sum p_x \cdot a_x = -46$ и $\sum p_y \cdot a_y = 9$.

7. Заполнить колонку $p_x \cdot a_x^2$, умножив a_x на $p_x \cdot a_x$, затем суммируя полученные произведения, записать как $\sum p_x \cdot a_x^2 = 504$ и соответственно произвести такие же действия и по строке $p_y \cdot a_y^2$. Получим $\sum p_y \cdot a_y^2 = 415$.

8. Для вычисления $\sum p \cdot a_x \cdot a_y$ (сумму произведений частот на отклонения по признаку x и y) необходимо:

а) в каждой ячейке, имеющей частоту, поставить множитель, полученный в результате умножения значений a_x и a_y. Множители к частотам записываются в виде степени. Например, в первом квадрате частота 6 расположена на пересечении отклонений по $a_y = -2$, по $a_x = -1$, перемножив $(-2)\cdot(-1)$, получим +2. Это и будет множитель к частоте шесть, который записывается в виде степени (6^2);

б) перемножив в каждом квадрате частоты на их множители, получают сумму этих произведений по каждому квадранту решетки (если в клетках частоты отсутствуют, то вычисления не производят). Например, первый квадрант (I) равняется 25 + 20 + 16 + 15 + 12 + 18 + + 6 + 24 + 28 + 12 + 4 + 6 + 14 + 3 =203; второй квадрант (II) = (–6) + (–4) + (–5) = –15; третий квадрант (III) = –1; четвертый квадрант (IV) = 4 + 10 + 3 + 10 + 16 + 6 + 8 + 3 + 9 + 30 + 4 + 8 + + 16 + 20 = 147;

в) суммировав результаты с положительными знаками (первый и четвертый квадранты), с отрицательными знаками (второй и третий квадранты), получаем общую сумму всех произведений $\sum p \cdot a_x \cdot a_y$ четырех квадрантов: первый квадрант + четвертый = = 203 +147 = +350; второй квадрант + третий = (–15) + (–1) = –16; $\sum p \cdot a_x \cdot a_y = 350 + (-16) = 334$. Полученный результат нужно подставить в формулу вычисления коэффициента корреляции.

Затем следует вычислить следующие показатели для формулы коэффициента корреляции:

1) поправку к условному среднему по признаку x:

$$b_x = \frac{\sum p_x \cdot a_x}{n} = \frac{-46}{130} = -0{,}35 \text{ кг};$$

2) поправку к условному среднему по признаку y:

$$b_y = \frac{\sum p_y \cdot a_y}{n} = \frac{9}{130} = 0{,}07 \text{ см};$$

3) среднее квадратическое отклонение (σ) по признаку x:

$$\sigma_x = \pm\sqrt{\frac{\sum p_x \cdot a_x^2}{n} - b_x^2} = \pm\sqrt{\frac{504}{130} - (-0{,}35)^2} = \pm\sqrt{3{,}88 - 0{,}12} = \pm 1{,}9;$$

4) среднее квадратическое отклонение (σ) по признаку y:

$$\sigma_y = \pm\sqrt{\frac{\sum p_y \cdot a_y^2}{n} - b_y^2} = \pm\sqrt{\frac{415}{130} - (0{,}07)^2} = \pm\sqrt{3{,}19 - 0{,}01} = \pm 1{,}7.$$

Подставляя полученные значения в формулу коэффициента корреляции, вычисляют его значение:

$$r = \frac{\sum p \cdot a_x \cdot a_y - n \cdot b_x \cdot b_y}{n \cdot \sigma_x \cdot \sigma_y} = \frac{334 - 130 \cdot (-0{,}35) \cdot 0{,}07}{130 \cdot 1{,}9 \cdot 1{,}7} = \frac{337{,}2}{419{,}9} = +0{,}80.$$

Это означает, что между живой массой коров и их обхватом груди существует положительная, сильная коррелятивная связь, так как коэффициент корреляции +0,80 близок к единице, т. е. с увеличением обхвата груди коров живая масса их повышается в большинстве случаев.

Ошибка коэффициента корреляции выборочного обследования вычисляется по формуле

$$m_r = \pm\frac{1 - r^2}{\sqrt{n}} = \frac{1 - 0{,}80^2}{\sqrt{130}} = \pm 0{,}03.$$

Критерий достоверности корреляции определяется по формуле $t_r = \frac{r}{m_r}$ как отношение величины коэффициента корреляции к своей ошибке:

$$t_r = \frac{0{,}80}{0{,}03} = 27.$$

Корреляция считается достоверной, если коэффициент корреляции равен своей удвоенной ошибке или превышает удвоенную ошибку, т. е. при $t_r = 2$.

Сопоставляем вычисленное значение t_r с величиной t_{st} из таблицы Стьюдента (см. приложение 1) при $\nu = 128$. Вычисленное t_r значительно больше табличной t_{st}(2,0–2,6–3,4).

Следовательно, корреляционная связь между живой массой тела и обхватом груди у коров является достоверной при $P < 0{,}001$.

Таблица 6.3

Расчет коэффициента корреляции между обхватом груди и живой массой коров бестужевской породы

Живая масса *x*, кг	Обхват груди *y*, см													
	180—183	184—187	188—191	192—195	196—199	200—203	204—207	208—211	212—215	216—219	p_x	a_x	$p_x \cdot a_x$	$p_x \cdot a_x^2$
440—479	1^{25}		1^{15}								2	–5	–10	50
480—519	1^{20}	1^{16}	1^{12}		1^{4}						4	–4	–16	64
520—559			2^{9}	4^{6}	2^{3}						8	–3	–24	72
560—599	I		1^{6}	7^{4}	7^{2}	3	3^{-2}	1^{-4}		II	22	–2	–44	88
600—639				6^{2}	3^{1}	12	5^{-1}				26	–1	–26	26
640—679				1	6	9	8	5	2		31	0	0	0
680—719					1^{-1}	5	4^{1}	5^{2}	1^{3}	1^{4}	17	1	17	17
720—759						1	5^{2}	4^{4}		1^{6}	11	2	22	44
760—799	III					1	1^{3}	1^{6}	1^{9}	IV	4	3	12	36
800—839								1^{8}		1^{16}	2	4	8	32
840—879									2^{15}	1^{20}	3	5	15	75
p_y	2	1	5	18	20	31	26	17	6	4	130	$\sum p_x \cdot a_x = -46$		
a_y	–5	–4	–3	–2	–1	0	1	2	3	4			$\sum p_x \cdot a_x^2 = 504$	
$p_y \cdot a_y$	–10	–4	–15	–36	–20	0	26	34	18	16	$\sum p_y \cdot a_y = 9$			
$p_y \cdot a_y^2$	50	16	45	72	20	0	26	68	54	64	$\sum p_y \cdot a_y^2 = 415$			

Задания для самостоятельной работы

Задание 1. Вычислить коэффициент корреляции между живой массой (x) и удоем за лактацию (y) у коров бестужевской породы по следующим данным, кг:

x	y	x	y	x	y	x	y	x	y
407	2049	545	2911	575	2988	478	3686	520	4773
542	4058	494	3236	575	3915	450	3606	607	3565
547	3077	525	2706	555	2712	450	3091	460	2790
667	5040	600	2813	547	3809	450	2855	485	2040
628	2242	486	3234	580	3647	557	4068	550	3319
530	3211	538	3192	555	3764	525	5025	572	3748
494	3206	588	3942	575	4484	475	2580	541	2788
486	3145	548	4560	608	4101	600	5618	525	3330
610	3212	585	5006	474	3412	579	4078	450	2838
456	2568	560	5605	491	3856	502	3180	485	2855

Задание 2. Вычислить коэффициент корреляции между живой массой кур-несушек (x, кг) и средней массой яйца (y, г) в потомстве петуха 695 русской белой породы:

x	y	x	y	x	y	x	y	x	y
1,5	58,1	2,3	66,2	2,2	59,5	1,5	58,1	1,9	53,1
2,2	59,1	2,0	60,6	2,2	57,5	2,2	59,5	1,9	64,1
2,2	57,5	2,2	61,4	2,2	52,0	2,2	57,5	1,8	60,6
2,0	58,8	1,7	58,5	2,1	62,0	2,3	66,2	2,4	64,6
2,2	52,0	2,4	64,6	1,7	58,8	2,0	60,6	1,7	58,0
2,1	62,0	1,8	60,6	2,2	63,6	1,7	58,5	2,2	60,0
1,7	58,8	1,9	64,1	2,1	66,1	2,1	64,6	2,0	60,6
2,2	63,6	1,8	62,2	1,7	68,3	1,9	64,1	2,2	63,6
2,1	66,1	1,9	53,1	1,7	58,8	1,9	53,1	2,1	62,0
1,7	68,3	1,8	57,8	2,1	62,0	1,8	57,8	1,5	58,1

Задание 3. Вычислить коэффициент корреляции между многоплодием матерей (x) и дочерей (y) у свиней крупной белой породы:

x	y	x	y	x	y	x	y	x	y
10,7	11,9	9,5	10,6	9,2	10,2	11,4	12,7	9,6	10,7
11,9	13,2	9,8	10,9	10,8	12,0	10,1	11,2	9,9	11,0
9,2	10,2	9,0	10,0	10,4	11,6	8,5	9,4	9,8	10,9

Окончание таблицы

x	y	x	y	x	y	x	y	x	y
10,4	11,6	8,8	9,8	11,9	13,2	10,0	11,1	9,0	10,0
10,8	12,0	9,0	10,0	10,4	11,6	10,0	11,2	9,1	10,1
9,5	10,6	9,7	10,8	10,8	12,0	10,6	11,8	10,3	11,4
10,4	11,6	9,4	10,4	9,7	10,8	10,3	11,4	9,8	10,9
9,2	10,2	11,9	13,2	9,8	10,9	9,2	10,2	9,0	10,0
10,8	12,0	10,4	11,6	8,8	9,8	10,1	11,2	9,4	10,4
9,7	10,8	9,5	10,6	9,4	10,4	9,8	10,9	9,7	10,8

Задание 4. Вычислить коэффициент корреляции между удоями за лактацию бестужевских коров (x) и их дочерей (y) (удои определены в одном и том же возрасте, кг):

x	y	x	y	x	y	x	y	x	y
3899	3725	3216	3919	3989	5678	5503	4500	5000	6011
3035	5422	4469	3208	3152	4227	5286	5046	5020	4984
3692	4701	3203	4918	4711	5000	4578	5109	3295	4110
3295	4622	4227	5211	4164	5286	5486	4050	3472	4111
3822	5123	4288	4346	3372	4475	4346	3447	3758	3670
5678	5143	3281	3486	5352	6037	3208	5486	2641	2760
3472	3475	3158	3483	3158	4064	2980	3483	3213	2641
3484	3401	4064	3764	3149	5363	3483	3463	2614	3940
4228	3864	2991	3158	2793	4183	3692	5206	3481	3419
5685	5315	2991	4394	3050	5288	5255	4851	3213	3725

Задание 5. Вычислить коэффициент корреляции между живой массой (x, кг) и яйценоскостью (y, г) кур-несушек русской белой породы:

x	y	x	y	x	y	x	y	x	y
2,1	225	1,9	191	2,1	170	2,2	215	2,3	207
2,0	193	2,0	201	2,3	232	2,2	180	2,3	205
2,4	271	2,0	200	2,0	208	1,9	193	2,0	213
2,2	208	2,1	210	1,9	189	2,3	241	1,9	190
2,2	201	2,1	220	1,8	179	2,0	207	2,0	200
1,7	212	2,3	246	1,7	163	2,4	241	2,2	203
2,0	189	2,2	219	2,0	201	2,1	199	2,1	221
1,8	200	1,8	175	1,8	185	2,5	220	2,2	230
2,5	256	2,2	217	1,9	194	2,0	198	2,3	234
1,9	183	2,1	213	1,7	165	2,1	200	2,2	212

Задание 6. Вычислить коэффициент корреляции между суточным удоем (x) и живой массой (y) коров черно-пестрой породы:

x	y	x	y	x	y	x	y	x	y
21,8	479	20,9	413	25,2	525	27,6	495	22,8	465
20,2	472	21,9	428	21,4	481	23,8	453	21,1	465
21,4	489	17,8	447	20,7	464	25,7	527	23,1	501
20,6	482	20,2	412	31,2	560	26,4	500	20,2	459
23,7	468	21,1	560	23,9	459	15,6	531	15,2	381
21,0	479	27,5	542	27,0	548	20,1	410	20,5	466
12,3	380	21,8	468	20,9	457	24,9	379	23,4	461
21,4	465	14,8	502	25,9	517	21,8	469	14,2	543
18,9	485	21,1	487	27,8	531	26,3	545	20,5	462
21,8	458	18,1	476	14,5	426	22,6	450	24,5	512

Задание 7. Вычислить коэффициент корреляции между живой массой (x, кг) и высотой в холке (y, см) у коров симментальской породы по следующим данным:

x	y	x	y	x	y	x	y	x	y
502	124	617	128	681	130	465	124	457	123
510	125	614	127	585	123	422	116	415	126
543	128	595	131	507	125	385	115	510	122
568	125	576	128	515	126	398	124	508	118
567	126	543	129	511	126	490	120	478	125
598	128	547	130	525	128	475	122	530	126
603	129	556	132	665	129	528	123	525	122
606	126	504	124	637	137	400	119	641	129
608	127	522	126	514	120	462	115	698	131
611	128	616	127	495	128	485	127	540	118

Задание 8. Вычислить коэффициент корреляции между высотой в холке (x) и обхватом груди (y) у кобыл русской рысистой породы по следующим данным, см:

x	y	x	y	x	y	x	y	x	y
161	176	155	173	166	182	154	175	155	170
160	175	156	171	152	178	152	167	156	170
150	167	165	180	155	172	158	182	160	180
156	175	155	170	156	179	149	160	164	184

Окончание таблицы

x	y	x	y	x	y	x	y	x	y
152	167	164	187	154	175	157	180	156	182
155	178	157	180	152	163	149	164	148	166
155	179	157	172	159	175	154	173	160	175
152	174	156	178	152	165	155	181	155	172
154	171	159	179	157	180	150	168	159	181
158	175	155	164	160	186	158	182	150	171

Задание 9. Вычислить коэффициент корреляции между длиной шерсти (x, см) и содержанием в организме овец связанного белкового йода (y, мг %):

x	y	x	y	x	y	x	y	x	y
11,0	4,7	10,0	5,5	15,5	2,1	9,0	6,0	9,5	5,5
12,0	4,6	9,0	5,8	15,0	3,6	10,2	4,5	10,5	5,1
11,5	5,5	9,2	6,0	15,2	4,0	11,5	4,3	12,0	4,4
11,2	4,7	9,0	5,6	14,0	4,2	9,5	4,7	11,5	4,7
9,5	5,6	12,0	5,0	13,0	5,5	10,0	4,8	13,0	4,3
9,0	6,3	12,0	5,4	13,0	5,1	10,5	4,5	12,5	4,7
12,5	4,5	13,0	3,6	13,0	5,6	9,5	5,8	10,0	5,2
9,0	5,6	12,5	4,5	9,0	7,1	10,0	5,1	9,0	6,5
13,0	4,1	15,0	2,0	12,5	3,9	10,5	5,0	11,2	5,0
9,8	5,3	9,0	5,6	15,0	3,6	9,0	6,0	13,0	4,3

Задание 10. Вычислить коэффициент корреляции между удоем за лактацию (x) и живой массой (y) у коров черно-пестрой породы:

x	y	x	y	x	y	x	y	x	y
7537	676	5651	543	7254	640	5374	770	4689	700
5279	530	5518	634	5389	535	7103	653	6278	600
6323	657	3351	676	4922	604	5280	668	4517	678
6121	620	5191	620	5726	720	5885	550	6044	649
6749	650	6039	540	5649	683	6388	615	5707	633
5068	614	5589	620	5556	578	5594	584	5206	570
7440	640	6392	630	3425	620	5373	519	5753	593
7208	603	5262	600	4634	632	4869	603	6537	660
5840	700	6146	703	5548	774	5087	729	6906	641
5095	636	6743	735	7638	677	4914	640	6671	640

Задание 11. Вычислить коэффициент корреляции между живой массой (x, кг) и обхватом груди (y, см) у коров бестужевской породы:

x	y	x	y	x	y	x	y	x	y
489	184	440	182	429	176	440	173	545	191
467	186	475	186	429	180	488	187	454	184
462	185	411	178	442	183	439	180	467	186
441	182	488	189	490	189	445	180	454	184
472	186	426	177	426	177	504	191	441	178
491	190	390	172	440	185	550	203	519	192
545	196	482	191	524	194	495	190	488	187
433	183	391	174	447	179	536	192	456	180
488	191	470	185	430	179	426	186	545	191
539	196	421	180	485	187	388	169	493	185

Задание 12. Вычислить коэффициент корреляции между удоем (x, кг) и содержанием жира в молоке (y, %) у полновозрастных коров бестужевской породы:

x	y	x	y	x	y	x	y	x	y
2049	3,80	3600	3,80	5605	4,36	3856	3,60	3180	3,59
4058	3,43	2911	4,01	2988	3,87	3686	3,70	4773	3,95
3077	3,78	3236	4,00	3915	4,14	3606	3,60	3565	3,47
5040	3,86	2706	3,80	2712	4,11	3091	3,80	2790	4,50
2433	3,80	2813	3,80	3809	3,34	2855	3,96	2940	3,87
3211	3,60	3234	4,30	3647	4,10	4068	4,20	3319	4,05
3206	4,00	3192	3,97	3764	3,76	5025	4,10	3748	3,90
3145	3,90	2942	3,55	4484	4,19	2580	4060	2788	3,80
3212	3,50	4560	3,89	4101	4,16	5618	4,50	3330	3,70
2568	3,90	5006	4,24	3412	3,74	4078	4,22	2838	3,80

Задание 13. Вычислить коэффициент корреляции между живой массой кобыл (x) и живой массой жеребят при рождении (y) орловской породы лошадей, кг:

x	y	x	y	x	y	x	y	x	y
483	51	510	50	534	56	450	47	480	53
487	48	486	54	550	57	544	57	510	49
481	58	526	53	500	46	520	51	550	56
462	42	450	44	545	57	597	53	470	44
538	55	470	45	491	50	592	52	492	51

Окончание таблицы

x	y	x	y	x	y	x	y	x	y
480	48	460	50	444	48	555	59	448	47
478	48	468	51	532	51	547	55	515	49
509	54	598	57	520	58	529	57	469	52
533	52	469	48	496	48	526	48	534	54
577	54	420	43	552	53	585	59	460	58

Задание 14. Вычислить коэффициент корреляции между живой массой (x) и удоем за лактацию (y) коров симментальской породы, кг:

x	y	x	y	x	y	x	y	x	y
523	4016	485	2975	495	2874	476	3483	572	5811
523	4181	500	3783	460	3644	575	3636	554	4800
485	4034	425	3183	470	3055	440	3399	515	3809
485	3905	478	2673	530	2793	503	3252	523	3572
500	2941	493	3735	508	2574	459	3702	600	4033
410	3022	450	3400	485	3355	548	4296	550	3984
520	4004	511	3340	460	3761	485	3149	575	3725
460	3767	570	3442	522	3109	450	3764	550	2931
500	2862	555	3425	580	3131	568	3465	485	3835
550	2562	410	3425	536	3509	602	3910	563	3221

Задание 15. Вычислить коэффициент корреляции между суточным удоем (x) и живой массой (y) коров бестужевской породы, кг:

x	y	x	y	x	y	x	y	x	y
28,8	512	12,3	380	31,2	560	26,4	500	23,4	451
20,2	472	21,4	465	23,9	459	15,6	531	16,0	445
21,4	489	18,9	485	27,0	548	20,1	410	23,8	458
20,6	482	21,8	458	20,9	457	24,9	379	24,3	524
23,7	468	20,9	413	25,9	517	21,8	469	19,6	487
21,0	479	21,9	428	27,8	513	26,3	545	15,5	416
25,5	515	17,8	447	14,5	426	22,6	450	21,6	418
21,7	451	20,0	412	27,6	495	15,2	396	14,2	393
20,8	450	21,1	560	23,8	453	23,4	469	20,1	455
25,3	500	27,5	542	25,7	527	24,8	521	21,4	462

6.3. Вычисление коэффициента корреляции для альтернативных признаков (r_a)

Рассмотренные выше способы вычисления корреляции пригодны для вычисления количественных сильно варьирующих признаков, которые можно выразить в килограммах, сантиметрах, процентах и т. д.

Очень часто встречаются признаки, имеющие качественное выражение. В этих случаях требуется отмечать наличие или отсутствие признака, т. е. проявление его в двух состояниях, например наличие сережек у свиней или их отсутствие, наличие или отсутствие рогов у животных. Такая изменчивость называется альтернативной и имеет качественное выражение. Между качественным или количественным выражением существует взаимосвязь. Усовершенствование анализа дает возможность выражать качественные признаки в количественных показателях, например, по количеству пигмента можно судить о степени окрашенности шерсти.

Корреляция между альтернативными признаками измеряется тетрахорическим показателем связи.

При изучении у каждой особи двух альтернативных признаков группа разбивается на четыре части: P_1 — особи, имеющие оба признака; P_2 — особи, имеющие первый признак, но не имеющие второго признака; P_3 — особи, не имеющие первого признака, но имеющие второй признак; P_4 — особи, не имеющие обоих признаков.

Тетрахорический показатель связи вычисляется по формуле

$$r_a = \frac{P_1 \cdot P_4 - P_2 \cdot P_3}{\sqrt{(P_1 + P_2)\cdot(P_3 + P_4)\cdot(P_1 + P_3)\cdot(P_2 + P_4)}}.$$

Величина r_a может находиться в границах от –1 до +1.

Пример. При изучении влияния отселекционированности на резистентность цыплят к пуллорозу оказалось, что после заражения живой культурой из 220 цыплят выжило 115, пало 105, а в отселекционированной группе выжило 560, пало 58. Определить тетрахорический показатель связи между резистентностью цыплят к пуллорозу и степенью отселекционированности стада по этому показателю.

Для вычисления тетрахорического показателя связи необходимо показатели разнести в четырехпольную корреляционную решетку (табл. 6.4).

Таблица 6.4

Корреляционная решетка для альтернативных признаков

Группа птицы	После заражения живой культурой		Итого
	выжило	пало	
Исходная	115P_1	105P_2	$P_1 + P_2 = 220$
Отселекционированная	560P_3	58P_4	$P_3 + P_4 = 618$
Итого	$P_1 + P_3 = 675$	$P_2 + P_4 = 163$	

Подставляя значения P в формулу, получим:

$$r_a = \frac{P_1 \cdot P_4 - P_2 \cdot P_3}{\sqrt{(P_1+P_2)\cdot(P_3+P_4)\cdot(P_1+P_3)\cdot(P_2+P_4)}} = \frac{115\cdot 58 - 560\cdot 105}{\sqrt{220\cdot 618\cdot 675\cdot 163}} =$$

$$= \frac{-52130}{122306} = -0{,}426.$$

Полученный показатель указывает на эффект селекции на резистентность цыплят к пуллорозу: с увеличением отселекционированности снижается заболевание цыплят пуллорозом.

Задания для самостоятельной работы

Задание 1. Вычислить коэффициент корреляции между безрогостью производителя и безрогостью потомства по следующим данным: от безрогого производителя получено 80 безрогих и 20 рогатых потомка, от рогатых — 40 безрогих и 35 рогатых.

Задание 2. Вычислить коэффициент корреляции между мастью матерей и мастью их потомства по следующим данным: от 300 красных матерей родилось 260 красных потомков и 40 палевых, а от 200 палевых матерей — 50 красных и 150 палевых.

Задание 3. Вычислить коэффициент корреляции между окраской шерсти и цветом глаз у кролика по следующим данным: от кролика с белой шерстью получено 29 потомка с красными и 11 с некрасными глазами, а от кролика с окрашенной шерстью — один с красными и 59 с некрасными глазами.

Задание 4. Установить связь между формой коконов матерей и потомства, если у матерей бабочек из остроконечного кокона было 18 потомков с этим признаком и 82 с нормальным коконом, а бабочка из нормального кокона дала два потомка с остроконечным и 98 с нормальным коконом.

6.4. Вычисления полихорического коэффициента связи

В практике селекционной работы и при анализе генетических данных иногда необходимо установить связь между качественными признаками, которые оценивают «на глаз», грубо, но зоотехнически вполне допустимо, например, связь между конституцией животных и степенью упитанности, конституцией и формой завитка смушка овец или связь признаков потомства с признаками родителей и т. п. В таких случаях связь устанавливают с помощью полихорического коэффициента связи ρ, который вычисляют по формуле

$$\rho = \frac{\alpha - 1}{\sqrt{(l_1 - 1)\cdot(l_2 - 1)}},$$

где

$$\alpha_1=\sum\left[\frac{\sum(P_{1.2}^2:P_2)}{P_1}\right]-\frac{(l_1-1)\cdot(l_2-1)}{n}$$

или

$$\alpha_2=\sum\left[\frac{\sum(P_{1.2}^2:P_1)}{P_2}\right]-\frac{(l_1-1)\cdot(l_2-1)}{n}.$$

Здесь 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.

При использовании данной формулы нет необходимости вычислять σ_1и σ_2, что значительно упрощает расчеты. Величина ρ колеблется от нуля до единицы, но этот коэффициент не выявляет направления корреляции, а только указывает на ее величину, поэтому он всегда имеет положительный знак. Полихорический коэффициент применяется и в тех случаях, когда один признак имеет несколько градаций.

Пример. Требуется определить величину связи между типом конституции коров-матерей и их дочерей. Обследовано 50 пар (мать — дочь). Пары распределены по клеткам корреляционной решетки (табл. 6.5).

Каждую частоту в клетке решетки обозначают $P_{1.2}$, возводят это число в квадрат $(P_{1.2})^2$ и затем делят на значение частот, размещенных в строчке P_2. Проведенная обработка дает для каждой клетки выражение $(P_{1.2})^2:P_2$. Полученную величину суммируют, получая $\sum\frac{(P_{1.2})^2}{P_2}$, и записывают сумму в графу, идущую после графы с частотами вариационного ряда P_1. Проведя вычисления по формуле $\left[\sum\frac{(P_{1.2})^2}{P_2}\right]:P_1$, получают итоговое число 2,23.

Подставляем данные в формулу α_1:

$$\alpha_1=\sum\left[\frac{\sum(P_{1.2}^2:P_2)}{P_1}\right]-\frac{(l_1-1)\cdot(l_2-1)}{n}=2{,}23-\frac{(4-1)\cdot(4-1)}{50}=2{,}05.$$

Используя величину α_1, находят полихорический коэффициент

$$\rho=\frac{\alpha-1}{\sqrt{(l_1-1)\cdot(l_2-1)}}=\frac{2{,}05-1}{\sqrt{(4-1)\cdot(4-1)}}=\frac{1{,}05}{\sqrt{9}}=\frac{1{,}05}{3}=0{,}35.$$

Для данного примера на основании величины ρ можно сделать вывод, что тип конституции дочерей коррелирует с типом конституции матерей, а полихорический коэффициент равен 0,35, т. е. наследственно обусловлен и может быть выражен коэффициентом наследуемости $h^2=2r$. В нашем примере $h^2=2\cdot0{,}35=0{,}70$ выявляет наследственную обусловленность конституции потомства.

Таблица 6.5

Полихорический коэффициент связи между типом конституции коров и их матерей

Дочери (1)		Матери (2)						
		Тип конституции				P_1	$\sum \frac{(R_{1,2})^2}{R_2}$	$\left[\sum \frac{(R_{1,2})^2}{R_2}\right] : R_1$
		грубая	нежная	плотная	рыхлая			
Тип конституции	грубая	$R_{1,2}=5$ $\frac{5^2}{5}=5$	—	—	$R_{1,2}=2$ $\frac{2^2}{15}=0,27$	7	5,27	$\frac{5,27}{7}=0,75$
	нежная	—	$R_{1,2}=15$ $\frac{15^2}{25}=9$	$R_{1,2}=5$ $\frac{5^2}{5}=5$	—	20	14,0	$\frac{14}{20}=0,70$
	плотная	—	$R_{1,2}=3$ $\frac{3^2}{25}=0,36$	—	—	3	0,36	$\frac{0,36}{3}=0,12$
	рыхлая	—	$R_{1,2}=7$ $\frac{7^2}{25}=1,96$	—	$R_{1,2}=13$ $\frac{13^2}{15}=11,27$	20	13,23	$\frac{13,23}{20}=0,66$
P_2		5	25	5	15	$n=50$	—	$\sum=2,23$

6.5. Вычисление рангового коэффициента корреляции Спирмена (r_s)

При обработке первичных материалов могут встретиться такие признаки, которые нельзя измерить ни точно, ни грубо, поэтому их выражают порядковым местом (рангом).

При вычислении рангового коэффициента корреляции составляется ранжированный ряд по возрастающей (или убывающей) степени выраженности признака. Порядковые номера животных в ранжированном ряду (их ранги) включаются в обработку.

Ранговый коэффициент корреляции выражают следующей формулой: $r_s = 1 - \frac{6 \cdot \sum D^2}{n \cdot (n^2 - 1)}$, где n — число сопоставляемых пар рангов; D — разность между парами рангов признака x с рангами признака y. Величина r_s изменяется от –1 до +1.

Пример. Определить, имеется ли связь между ростом (x) рысистых лошадей и скоростью их бега (y) на дистанцию 1600 м. Сравниваются ранги пяти лошадей по этим двум признакам (табл. 6.6).

Таблица 6.6

Вычисление рангового коэффициента корреляции между ростом и скоростью бега рысистых лошадей

Номер лошади	Ранги по x	Ранги по y	Разность рангов $D = (x - y)$	$D^2 = (x - y)^2$
1	1	4	$1 - 4 = -3$	$-3^2 = 9$
2	2	5	$2 - 5 = -3$	$-3^2 = 9$
3	3	2	$3 - 2 = +1$	$1^2 = 1$
4	4	3	$4 - 3 = +1$	$1^2 = 1$
5	5	1	$5 - 1 = +4$	$4^2 = 16$

$$r_s = 1 - \frac{6 \cdot 36}{5 \cdot (5^2 - 1)} = 1 - \frac{216}{120} = 1 - 1{,}8 = -0{,}80, \sum D^2 = 36.$$

Как видно из приведенного примера, связь между ростом и скоростью бега лошади на дистанцию большая и обратная: у лошадей более высокого роста в среднем скорость бега выше, чем у лошадей более низкого роста.

Пример. Выяснить, имеется ли связь между агрессивностью (x) норок и окраской их волосяного покрова (y). Было отобрано шесть норок, которые оценены по обоим признакам и распределены по рангам: агрессивность — от неагрессивного до очень агрессивного, а степень окраски волосяного покрова — от худшей до лучшей (табл. 6.7).

Таблица 6.7

Вычисление рангового коэффициента корреляции между агрессивностью и степенью окраски меха норок

Норки	Ранги по x	Ранги по y	Разность рангов $D=(x-y)$	$D^2=(x-y)^2$
А	Слабая 1	3	$1-3=-2$	$-2^2=4$
Б	2	1	$2-1=+1$	$1^2=1$
В	3	2	$3-2=+1$	$1^2=1$
Г	4	6	$4-6=-2$	$-2^2=4$
Д	5	4	$5-4=+1$	$1^2=1$
Е	Сильная 6	5	$6-5=+1$	$1^2=1$
				$\sum D^2=12$

$$r_s=1-\frac{6\sum(x-y)^2}{n\cdot(n^2-1)}=1-\frac{6\cdot 12}{6\cdot(6^2-1)}=\frac{72}{6\cdot 35}=1-\frac{72}{210}=1-0{,}34=+0{,}66.$$

Расчеты показывают, что между агрессивностью норок и степенью окраски их волосяного покрова имеется высокая положительная связь.

Задания для самостоятельной работы

Задание 1. При изучении зависимости густоты спермы от степени упитанности отобрано семь быков-производителей. Они ранжированы: по упитанности — от самой низкой до ожиревшего, по густоте спермы — от очень плохого до самого лучшего. По данным ранжированных рядов вычислить коэффициент ранговой корреляции:

Производители	А	Б	В	Г	Д	Е	Ж
Упитанность (x)....	1	2	3	4	5	6	7
Густота спермы (y)...	3	5	7	6	4	1	2

Задание 2. Приведены ранжированные ряды густоты шерсти овец от меньшей и упитанности от низшей до высшей. Что можно сказать о характере взаимосвязи между густотой шерсти и упитанностью овец?

Номера овец...	20	21	22	23	24	25	26
Густота шерсти.....	1	2	3	4	5	6	7
Упитанность...	2	3	4	1	7	4	5

Задание 3. При изучении связи качества меха с агрессивностью поведения у лисиц были отобраны девять самцов, которые занимают ранги от самого спокойного, т. е. слабой агрессивности, до сильного.

По качеству меха животные расположены также по рангу от худшего до лучшего. Можно ли вести селекцию по поведению норок с целью повышения экономической эффективности?

Лисы	А	Б	В	Г	Д	Е	Ж	З	И
Агрессивность	1	2	3	4	5	6	7	8	9
Качества меха	3	1	2	5	4	8	9	6	7

6.6. Вычисление коэффициента генетической корреляции

Немалое значение для селекционной практики имеет определение связей между признаками, обусловленными наследственностью. В основе выявления генетической детерминации наследования признаков лежит генетический анализ, при котором признаки потомков сравнивают с признаками их родителей. Такой анализ можно успешно 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 провести изучение в пределах одного семейства, а потребуется его осуществление на массовом материале с применением популяцонного и математического методов. Для этого применяется метод определения генетического коэффициента корреляции (r_G) между признаками, разработанный Хейзелем. Суть этого метода заключается в том, что на группах родственных животных (матери — дочери, отцы — сыновья, полусестры) вычисляют четыре коэффициента корреляции между двумя разными фенотипическими признаками (x и y) в пределах каждой сопоставляемой родственной группы и между группами.

Генетическая корреляция указывает на изменение вторичных признаков при селекции первичных признаков:

$$r_G = \sqrt{\frac{rx_D y_M \cdot ry_D x_M}{rx_D x_M \cdot ry_D y_M}};\quad r_G = \sqrt{\frac{Rx_D y_M \cdot Ry_D x_M}{Rx_D x_M \cdot Ry_D y_M}}.$$

Эта формула применяется для тех случаев, когда оба r в числителе имеют знак «плюс» или знак «минус», но в знаменателе оба r должны быть положительными, иначе пользоваться формулой Хейзеля нельзя. Если же под корнем в числителе один из r имеет знак «минус», а другой — знак «плюс», то формула видоизменяется:

$$r_G = \frac{(rx_D y_M + ry_D x_M):2}{\sqrt{rx_D x_M \cdot ry_D y_M}}.$$

Степень и характер генетических корреляций между признаками тесно связаны с эффектом селекции и должны учитываться при отборе родительских пар для получения потомства с наилучшей сочетаемостью желательных признаков.

Коэффициенты генетической корреляции имеют большое значение также при ранней индивидуальной оценке животных, при вычислении селекционных индексов.

Наличие отрицательной связи между x_D и x_M или между y_D и y_M указывает на сильное взаимодействие генотипа со средой или на сложный тип наследования (эпистаз, межаллельное взаимодействие) и, следовательно, по формуле Хейзеля нельзя выявить связь, так как эта формула основана на предположении об аддитивном действии генов коррелирующих признаков.

Пример. 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. 6.8.

Таблица 6.8

Данные для вычисления генетического коэффициента корреляции

Пара	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
5	3,4	3,3	4,5	4,2

Методом малой выборки определяют rx_Dx_M; ry_Dy_M; rx_Dy_M; ry_Dx_M. Тем же способом определяют фенотипические коэффициенты корреляции:

- содержание белка (дочери-матери): $rx_Dx_M = 0{,}65$;
- содержание жира (дочери-матери): $ry_Dy_M = 0{,}80$;
- содержание белка (дочери) — содержание жира (матери): $rx_Dy_M = 0{,}733$;
- содержание жира (дочери) — содержание белка (матери): $ry_Dx_M = 0{,}053$.

Подставляя эти данные в формулу, получают:

$$r_G = \sqrt{\frac{rx_Dy_M \cdot ry_Dx_M}{rx_Dx_M \cdot ry_Dy_M}} = \sqrt{\frac{0{,}733 \cdot 0{,}053}{0{,}65 \cdot 0{,}80}} = \sqrt{\frac{0{,}03906}{0{,}52}} = \sqrt{0{,}0751} = 0{,}274.$$

Следовательно, генетическая связь между белковомолочностью и жирномолочностью коров выражается $r_G = 0{,}274$, поэтому отбор животных по содержанию жира в молоке будет сопровождаться генетически обусловленным повышением белковомолочности.

Задания для самостоятельной работы

Задание 1. Вычислить коэффициенты фенотипических (rx/x'; ry/y'; rx/y'; ry/x') и генетической корреляции между живой массой кур и их яйценоскостью (табл. 6.9).

Таблица 6.9

Живая масса (кг) и яйценоскость (шт.) кур (x'; y') и их дочерей (x, y)

x'	y'	x'	y'	x	y	x	y
2,1	225	1,9	191	2,1	170	2,2	215
2,0	193	2,0	201	2,3	232	2,2	180
2,4	271	2,0	200	2,0	208	1,9	193
2,2	208	2,1	210	1,9	189	2,3	241
2,2	201	2,1	220	1,8	179	2,0	207
1,7	212	2,3	246	1,7	163	2,4	241
2,0	189	2,2	219	2,0	201	2,1	199
1,8	200	1,8	175	1,8	181	2,5	220
2,5	256	2,2	217	1,9	194	2,0	198
1,9	183	2,1	213	1,7	165	2,1	200

Задание 2. Вычислить коэффициент генетической корреляции между относительной массой белка и яйценоскостью кур русской белой породы, если коэффициенты фенотипических корреляций равны:

$$r_{x'/x} = \pm 0,42;\ r_{x'/y} = -0,23;$$
$$r_{y'/y} = \pm 0,65;\ r_{y'/x} = -0,28,$$

где x'/x — относительная масса белка у матерей и дочерей; y'/y — яйценоскость матерей и дочерей; x'/y — относительная масса белка у матерей и яйценоскость дочерей; y'/x — яйценоскость матерей и относительная масса белка у дочерей.

Какие выводы можно сделать о характере наследования указанных признаков? В каком случае наиболее эффективно их сочетание?

Задание 3. Определить генетическую корреляцию между содержанием белка и жира в молоке 16 коров и их 16 дочерей по следующим данным (табл. 6.10).

Таблица 6.10

Содержание белка и жира в молоке коров-матерей и их дочерей

Содержание белка, %				Содержание жира, %			
дочери (x)	матери (x')	дочери (x)	матери (x')	дочери (y)	матери (y')	дочери (y)	матери (y')
3,1	3,0	3,3	3,1	4,0	3,9	4,1	4,0
3,3	3,1	3,3	3,4	4,2	4,0	4,0	4,0

Содержание белка, %				Содержание жира, %			
дочери (x)	матери (x')	дочери (x)	матери (x')	дочери (y)	матери (y')	дочери (y)	матери (y')
3,2	3,2	3,4	3,3	4,1	4,0	4,1	4,0
3,0	3,1	3,3	3,2	4,0	3,8	3,8	3,9
3,4	3,3	3,2	3,3	4,5	4,2	4,2	4,0
3,1	3,1	3,2	3,1	4,3	4,3	4,2	4,1
3,2	3,2	3,0	3,1	4,1	4,2	4,0	4,0
3,5	3,4	3,3	3,2	4,2	4,0	3,8	3,9

6.7. Вычисление коэффициента прямолинейной регрессии

Регрессией называется изменение функций (зависимого признака) в зависимости от изменения аргумента. Регрессионный анализ имеет большое значение в изучении корреляционных связей.

Коэффициент прямолинейной регрессии R[1] указывает, насколько в среднем изменяется один из признаков при изменении другого на единицу измерения. В больших выборках этот показатель вычисляется по формулам

$$R_{x/y} = r \cdot \frac{\sigma_x}{\sigma_y}; \ R_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x}.$$

В малочисленных выборках определяется по формулам

$$R_{x/y} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}};$$

$$R_{y/x} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}.$$

Коэффициент регрессии отражает связь между признаками, в отличие от коэффициента корреляции, в именованных величинах (килограммы, сантиметры, проценты и т. д.) и применяется при планировании и прогнозировании уровня того или иного признака по заданному уровню данного признака. Регрессия между признаками может быть

[1] В литературе по биометрии, изданной в 1960—1970 гг., коэффициент регрессии обозначен как R. В современной литературе этот показатель обозначен буквой b. В данном параграфе используются эти обозначения.

выражена в виде эмпирического и теоретического рядов регрессии, в виде графика, а также через уравнения регрессии.

Пример. Требуется определить для нашего примера (по данным таблицы):

1) как изменяется живая масса свиноматок украинской степной белой породы при изменении их обхвата груди на 1 см;

2) на сколько сантиметров увеличится (уменьшится) обхват груди свиноматок при увеличении (уменьшении) живой массы их на 1 кг.

Среднее квадратическое отклонение равно: $\sigma_y = 6{,}266$; $\sigma_x = 25{,}798$, а корреляция между этими признаками: $r_{xy} = 0{,}905$. Подставим эти значения в формулы коэффициента регрессии. Регрессия живой массы по обхвату груди:

$$R_{x/y} = r \cdot \frac{\sigma_x}{\sigma_y} = 0{,}905 \frac{25{,}798}{6{,}266} = 3{,}72 \text{ кг};$$

$$R_{x/y} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}} = \frac{523\,343 - \frac{3582 \cdot 2183}{15}}{318\,249 - \frac{4\,765\,489}{15}} = 3{,}72 \text{ кг},$$

т. е. с увеличением обхвата груди на 1 см живая масса свиноматок увеличится на 3,72 кг.

Регрессия обхвата груди по живой массе равняется:

$$R_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x} = 0{,}905 \frac{6{,}266}{25{,}738} = 0{,}22 \text{ см};$$

$$R_{y/x} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{523\,343 - \frac{3582 \cdot 2183}{15}}{864\,656 - \frac{12\,830\,724}{15}} = \frac{2042{,}6}{9274{,}4} = 0{,}22 \text{ см},$$

т. е. с увеличением живой массы на 1 кг обхват груди увеличится в среднем на 0,22 см.

Между коэффициентами регрессии и корреляции имеется связь которая выражается формулой: $r_{xy} = \sqrt{R_{yx} R_{xy}}$. Подставим в эту формулу значения R_{yx} и R_{xy}, получим: $r_{xy} = \sqrt{0{,}22 \cdot 3{,}72} = 0{,}905$. Коэффициент корреляции совпадает с тем, который был вычислен ранее.

Пример. При изучении связи между содержанием жира (x, %) и белка (y, %) (табл. 6.11) в молоке джерсейских коров были рассчитаны следующие показатели: $\sigma_x = 0{,}513$; $\sigma_y = 0{,}274$; $r = 0{,}414$. Вычислить коэффициенты регрессии $R_{x/y}$ и $R_{y/x}$:

$$R_{x/y} = r \frac{\sigma_x}{\sigma_y} = 0{,}414 \cdot \frac{0{,}513}{0{,}274} = +0{,}76\,\%; \quad R_{y/x} = r \frac{\sigma_y}{\sigma_x} = 0{,}414 \cdot \frac{0{,}274}{0{,}513} = +0{,}22\,\%.$$

Это означает, что с увеличением содержания белка в молоке на 1 % жирномолочность повышается в среднем на 0,76 %, а с увеличением жирности молока на 1 % содержание белка возрастет в среднем на 0,22 %.

Таблица 6.11

Исходные данные для расчета коэффициента регрессии по живой массе и обхвату груди у свиноматок, $n = 15$

y	x	y^2	x^2	xy
140	200	19 600	40 000	28 000
145	248	21 025	61 504	35 960
158	287	24 964	63 001	39 658
144	223	20 736	49 729	32 112
142	227	20 164	51 529	32 234
145	220	21 025	48 400	31 900
145	220	21 025	48 400	31 900
140	226	19 600	51 076	31 640
160	291	25 600	84 681	46 560
138	225	19 044	50 625	31 050
144	243	20 736	59 049	34 992
143	228	20 449	51 984	32 604
146	253	21 316	64 009	36 938
142	226	20 164	51 076	32 092
151	265	22 801	70 225	40 015
$\sum y = 2183$	$\sum x = 3582$	$\sum y^2 = 318\,249$	$\sum x^2 = 864\,656$	$\sum xy = 523\,343$

6.8. Метод наименьших квадратов

Для определения параметров линейной регрессии применяется метод наименьших квадратов (МНК), который разработали К. Гаусс и П. Лаплас. Согласно МНК параметры уравнения регрессии подбираются таким образом, чтобы сумма квадратов отклонений наблюдений от линии регрессии была минимальной, т. е. $Q = (y_i - \overline{y})^2 \to \min$.

Необходимым условием минимума функции является равенство нулю частных производных по неизвестным параметрам a и b[1]. Для функции $Q = \sum(y_i - \overline{y})^2 = \sum(y_i - a - bx_i)^2$ находят частные производные по a и b и приравнивают их к нулю:

[1] *Четыркин Е. М., Калихман И. Л.* Вероятность и статистика. М. : Финансы и статистика, 1982. С. 226.

$$\begin{cases} \dfrac{dQ}{da} = -2\sum(y_i - a - bxi) = 0; \\ \dfrac{dQ}{db} = -2\sum(y_i - a - bxi)xi = 0; \end{cases}$$

Сокращают каждое уравнение на –2, раскрывают скобки и переносят члены с x в одну сторону, а с y — в другую, получают стандартную форму линейных уравнений:

$$\begin{cases} na + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i \end{cases}.$$

Разделив первое уравнение системы на n, получают $\bar{y} = a + b\bar{x}$, т. е. найденная прямая проходит через точку с координатами $\bar{x}$, $\bar{y}$. Решают эти уравнения относительно параметров a и b, например методом Крамера, используя следующие формулы:

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix} = n\sum x^2 - \left(\sum x\right)^2;$$

$$\Delta_a = \begin{vmatrix} \sum y & \sum x \\ \sum xy & \sum x^2 \end{vmatrix} = \sum y \sum x^2 - \sum x \sum xy; \; a = \Delta_a / \Delta;$$

$$\Delta_b = \begin{vmatrix} n & \sum y \\ \sum x & \sum xy \end{vmatrix} - n\sum xy \sum y \sum x; \; b = \Delta_b / \Delta.$$

В нашем примере по регрессии обхвата груди и живой массе у свиноматок украинской степной белой породы расчетные суммы следующие: $\sum x = 3582$; $\sum y = 2183$; $\sum x^2 = 864\,656$; $\sum xy = 523\,343$; $n = 15$.

В связи с этим вычисляем:

$$\Delta = \begin{vmatrix} 15 & 3582 \\ 3582 & 864\,656 \end{vmatrix} = 15 \cdot 864\,656 - 3582^2 = 139\,116.$$

$\Delta = 139\,116 \neq 0$, поэтому исходная матрица является неособенной и имеет единственное решение.

Далее находим определители Δ_a и Δ_b, а также a и b:

$$\Delta_a = \begin{vmatrix} 2183 & 3582 \\ 523\,343 & 864\,656 \end{vmatrix} = 2183 \cdot 864\,656 - 3582 \cdot 523\,343 =$$

$$= 12\,929\,422;$$

$$a = 12\,929\,422 / 139\,116 = 92{,}93986314;$$

$$\Delta b = \begin{vmatrix} 15 & 2183 \\ 3582 & 523\,343 \end{vmatrix} = 15 \cdot 523\,343 - 2183 \cdot 3582 = 30\,639;$$

$$b = 30\,639 / 139\,116 = 0{,}220240662.$$

Чтобы проверить правильность вычислений, нужно в систему уравнений

$$\begin{cases} na+b\sum x_i=\sum y_i \\ a\sum x_i+b\sum x_i^2=\sum x_i y_i \end{cases}$$

подставить значения a и b:

$$\begin{cases} 15\cdot 92{,}93986314+0{,}220240662\cdot 3582=2183 \\ 92{,}93986314\cdot 3582+0{,}220240662\cdot 864\,656=523\,343 \end{cases}$$

Из этого следует, что система решена правильно.

Уравнение регрессии имеет вид

$$\bar{y}_x=92{,}93986314+0{,}220240662x.$$

Для упрощения системы нормальных уравнений значения переменных могут быть выражены в отклонениях от средней. Обозначим эти отклонения $x'i$ и $y'i$. При этом система нормальных уравнений будет иметь вид

$$\sum x_i' y_i' = b\sum (x_i')^2,$$

так как $\sum y_i'$ и $\sum x_i'$ равны нулю. Из этого следует:

$$b=\frac{\sum x_i' y_i'}{\sum (x'_i)^2}.$$

Из уравнения $\bar{y}=a+b\bar{x}$ находим:

$$a=\bar{y}-b\bar{x}.$$

Для оценки уравнения регрессии этим способом определяем суммы квадратов и средних:

$$\sum y=2183,\ \sum x=3582;$$

$$\bar{y}=2183/15=145{,}533;\ \sum x^2=864\,656;$$

$$\bar{x}=3582/15=238{,}8;\ \sum xy=523\,343;$$

$$\sum (x_i')^2=\sum x^2-\frac{\left(\sum x\right)^2}{n}=846\,465-\frac{3582^2}{15}=9274{,}4;$$

$$\sum x_i' y_i' \sum xy-\frac{\sum x\sum y}{n}=523\,343-\frac{3582\cdot 2183}{15}=2042{,}6.$$

Находим a и b:

$$b=\frac{\sum x_i' y_i'}{\sum (x_i')^2}=\frac{2042{,}6}{9274{,}4}=0{,}220240662;$$

$$a=\bar{y}-b\bar{x}=145{,}5333333-0{,}220240662\cdot 238{,}8= =92{,}93986314.$$

Уравнение регрессии имеет вид

$$\bar{y}_x=92{,}93986314+0{,}220240662x.$$

Коэффициенты a и b, вычисленные способом отклонения переменных от средней и с помощью определителей Крамера, полностью совпадают. Поэтому получены идентичные уравнения регрессии.

6.9. МНК в матричной записи

Уравнение парной регрессии можно представить в матричной записи:

$$Y = X\beta + e.$$

Из нашего примера по регрессии обхвата груди и живой массе у свиноматок можно записать:

$$140=\beta_0 + 200\beta_1 + e_1;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$151=\beta_0 + 265\beta_1 + e_{15};$$

или

$$Y_i = \beta_0 + \beta_1 X_i + e_i \ (e_i = 1, 2, \ldots, 15)$$

для каждого из 15 наблюдений. При нахождении коэффициентов регрессии в уравнениях в матричной записи применяются операции транспонирования, произведения и обращения матриц.

Транспонированной называется матрица ($A^т$), в which которй столбцы исходной матрицы (A) заменяются строками с соответствующими номерами, $A = (a_{ij})$, то $A^т = (a_{ji})$. Из определения транспонированной матрицы следует, что если исходная матрица A имеет размер $m \times n$, то транспонированная матрица $A^т$ имеет размер $n \times m$.

Произведение матриц определено, если число столбцов первой матрицы равно числу строк второй.

Пусть $A = (a_{ij})\ m \times n$, $B = (b_{ij})\ n \times p$, тогда размерность произведения $A \times B$ равна $m \times p$. При этом матрица C (размера $m \times p$) называется произведением матриц A и B, если каждый ее элемент c_{ij} равен сумме произведений элементов i-й строки матрицы A на соответствующие элементы j-го столбца матрицы B:

$$c_{ij} = a_{i1} + b_{1j} + a_{i2}b_{2j} + \ldots + a_{ip} + b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj},$$

$$i=1,2,\ldots,m,\ j=1,2,\ldots,n.$$

Обращение матриц. Матрица A^{-1} называется обратной по отношению к квадратной матрице A, если при умножении этой матрицы на данную как слева, так и справа получается *единичная матрица*: $A \times A^{-1} = A^{-1} \times A = E$. Из определения следует, что обратная матрица является квадратной того же порядка, что и исходная матрица. Необходимым и достаточным условием существования обратной матрицы является невырожденность исходной матрицы. Матрица называется *невырожденной* или *неособенной*, если ее определитель ($\Delta \neq 0$).

Обратные матрицы используются для решения систем уравнений с несколькими неизвестными.

Доказано[1], что

$$X^{T}X = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix} \text{ и } X^{T}Y = \begin{vmatrix} \sum y \\ \sum xy \end{vmatrix},$$

т. е. получаем нормальные уравнения, которые можно записать так: $X^{T}Xb = X^{T}Y$.

Умножая обе части этого уравнения на $(X^{T}X)^{-1}$ получим $(X^{T}X)^{-1}X^{T}Xb = (X^{T}X)^{-1}X^{T}Y$, отсюда $b = (X^{T}X)^{-1}X^{T}Y$, поскольку $(X^{T}X)^{-1}(X^{T}X) = E$, где E — единичная матрица. В нашем примере:

$$(X^{T}X)^{-1} = \begin{vmatrix} 6{,}215359844 & -0{,}025748296 \\ -0{,}025748296 & 0{,}000107824 \end{vmatrix}.$$

Таким образом, выражая одномерную линейную модель, подлежащую оцениванию на основе данных нашего примера по регрессии обхвата груди и живой массы у свиноматок в форме матричной записи:

$$Y = X\beta + e,$$

МНК оценки (β_0, β_1), МНК оценки вектора $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ можно выразить формулой

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (X^{T}X)^{-1}X^{T}Y.$$

Умножая обратную матрицу $(X^{T}X)^{-1}$ на $X^{T}Y$ получим вектор столбец коэффициентов регрессии:

$$b = \begin{vmatrix} 6{,}215359844 & -0{,}025748296 \\ -0{,}025748296 & 0{,}000107824 \end{vmatrix} \times \begin{vmatrix} 2183 \\ 523343 \end{vmatrix} = \begin{vmatrix} 92{,}93986314 \\ 0{,}220240662 \end{vmatrix}.$$

При этом уравнение регрессии имеет вид

$$\bar{y}_x = 92{,}93986314 + 0{,}220240662x.$$

Таким образом, если вычисления провести с максимальным количеством знаков после запятой, то коэффициенты регрессии, вычисленные разными способами, полностью совпадают.

Коэффициент корреляции r_{xy} служит основой для вычисления *коэффициента детерминации* r_{xy}^2.

Коэффициент детерминации выступает как мера качества подбора линии регрессии и показывает, какая доля вариации одного признака зависит от варьирования другого признака.

[1] *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. М. : Статистика, 1973.

В нашем примере коэффициент корреляции r_{xy} между обхватом груди и живой массой у свиноматок равен 0,905. Следовательно, $r_{xy}^2 = 0{,}82$. Это означает, что 82 % общей дисперсии обхвата груди свиноматок (относительно средней) связано с изменением их живой массы.

6.10. Ошибки коэффициентов регрессии

Надежность получаемых по уравнению регрессии расчетных значений во многом определяется рассеянием наблюдений вокруг линии регрессии, т. е. остатков, которые определены как разность между наблюдаемыми значениями и прогнозируемыми, полученными с помощью расчетного уравнения регрессии:

$$e_i = y_i - \overline{y}'_i,\ i = 1, 2, \dots, n,$$

где y_i — наблюдаемая величина; $\overline{y}'_i$ — соответствующая прогнозируемая величина.

В качестве меры рассеяния служит дисперсия s^2 относительно регрессии:

$$s^2 = \frac{\sum e_i^2}{n-2},$$

где s^2 — выборочная оценка дисперсии; n – 2 — число степеней свободы. Сумму квадратов отклонений фактических наблюдений от линии регрессии $\sum e_i^2$ можно определить по формуле

$$\sum e_i^2 = \sum (y'_i)^2 - b \sum x'_i y'_i.$$

Дисперсия коэффициента регрессии b равна

$$s_b^2 = \frac{s^2}{\sum (x'_i)^2},$$

следовательно,

$$s_b = \frac{s^2}{\sqrt{\sum (x'_i)^2}}.$$

Значимость коэффициентов регрессии можно проверить с помощью обычной схемы проверки статистических гипотез. При этом предполагается, что отклонение от регрессии следует нормальному распределению. В парной линейной регрессии проверяется значимость коэффициента b. Нулевая гипотеза имеет вид $H_0 : \beta = 0$. Проверка осуществляется при помощи нормированного отклонения $t^* = \dfrac{b}{s_b}$. Если $t^* > t_\alpha$ (t — распределения Стьюдента) с числом степеней свободы $\nu = n - 2$, то нулевая гипотеза отклоняется. Это свидетельствует о том, что между соответствующими переменными существует линейная зависимость.

В нашем примере получена регрессия

$$\hat{y}_i = 92{,}93986314 + 0{,}220240662 x_i.$$

По приведенным выше формулам находим ошибку коэффициента регрессии и его значимость. По данным этого примера находим:

$$\sum (x_i')^2 = 9274{,}4; \ \sum (y_i')^2 = 549{,}733;$$
$$\sum x'y' = 2042{,}6; \ \sum e_i^2 = \sum (y_i')^2 b \sum x_i' y_i' =$$
$$= 549{,}733 - 0{,}220240662 \cdot 2042{,}6 = 99{,}86942284.$$

Далее находим дисперсию относительно регрессии:

$$s^2 = \frac{\sum e_i^2}{n-2} = \frac{2042{,}6}{15-2} = 7{,}68230769.$$

Квадратическая ошибка коэффициента регрессии:

$$s_b = \frac{7{,}68230769}{\sqrt{9274{,}4}} = \frac{7{,}68230769}{96{,}30368633} = 0{,}079771689,$$

следовательно,

$$t^* = \frac{b}{s_b} = \frac{0{,}220240662}{0{,}079771689}.$$

Критическое значение t_α при числе степеней свободы равном 14 ($t_{\alpha\ 0{,}05} = 2.1$), т. е. меньше расчетного. Поэтому гипотеза H_0 отклоняется. Из этого следует, что между обхватом груди и живой массой свиноматок, применительно к нашему примеру, существует линейная зависимость.

6.11. Множественная линейная регрессия

Зависимость между несколькими переменными величинами можно выразить уравнением множественной регрессии. Для двух независимых переменных (x и z) уравнение регрессии имеет вид

$$y = a + bx + cz,$$

где a — свободный член уравнения, b и c — параметры уравнения. Параметры этого уравнения можно найти, применяя способ наименьших квадратов:

$$an + b\sum x + c\sum z = \sum y;$$
$$n\sum x + b\sum x^2 + c\sum xZ = \sum xy;$$
$$a\sum z + b\sum xZ + c\sum z^2 = \sum yz.$$

Для составления системы уравнений необходимо по эмпирическим данным предварительно рассчитать следующие величины: $\sum x$, $\sum y$, $\sum z$, $\sum x^2$, $\sum z^2$, $\sum xy$, $\sum xz$, $\sum yz$.

Пример. Необходимо найти эмпирическое уравнение регрессии между живой массой y, высотой в холке x и обхватом груди z у 25 коров симментальской породы. Эмпирические исходные данные приведены в табл. 6.12. Необходимые суммы из этой таблицы надо подставить в уравнения системы:

$$25a + 3404b + 5005c = 16\,266;$$

$$3404a + 463\,800b + 682\,058c = 2\,217\,306;$$

$$5005a + 652\,058b + 100\,4583c = 3\,264\,917.$$

Таблица 6.12

Живая масса, высота в холке и обхват груди у коров симментальской породы

y	x	z	yx	yz	xz	$\hat{y}$	$y-\hat{y}$
600	135	193	81 000	115 800	26 055	628,6068858	–28,60688579
680	141	200	95 880	136 000	28 200	667,0086309	12,99136907
650	134	194	87 100	126 100	25 996	627,6210796	22,37892035
630	137	189	86 310	119 070	25 893	625,5805122	4,419487824
600	132	188	79 200	112 800	24 816	605,6575238	–5,6575238
700	142	216	99 400	151 200	30 672	710,4773171	–10,47731715
680	140	202	95 200	137 360	28 280	668,5218177	11,47818227
590	135	195	79 650	115 050	26 325	633,6048717	–43,60487168
661	136	197	89 896	130 217	26 792	642,0876567	18,91234334
610	132	185	80 520	112 850	24 420	598,160545	11,83945504
630	134	199	84 420	125 370	26 666	640,1160444	–10,11604438
580	131	189	75 980	109 620	24 759	604,6717177	–24,67171766
680	139	208	94 520	141 440	28 912	680,0309763	–0,030976322
615	130	190	79 950	116 850	24 700	603,6859115	11,31408848
700	137	212	95 900	148 400	29 044	683,0573499	16,94265007
640	132	214	84 480	136 960	28 248	670,6313404	–30,6313404
680	135	207	91 800	140 760	27 945	663,592787	16,40721297
610	131	184	79 910	112 240	24 104	592,1767529	17,82324707
700	142	218	99 400	152 600	30 956	715,475303	–15,47530304
670	140	207	93 800	138 690	28 980	681,0167825	–11,01678246
630	138	192	86 940	120 960	26 496	636,5622901	–6,5622901
680	139	211	94 520	143 480	29 329	687,5279552	–7,52795516
670	135	200	90 450	134 000	27 000	646,0998364	23,90016359
740	138	215	102 120	159 100	29 670	694,0391279	45,96087214
640	139	200	88 960	128 000	27 800	660,0390328	–20,03903275

3) $n=25$; $\sum y = 16\,266$; $\sum x = 3404$; $\sum z = 5005$; $\sum x^2 = 463\,800$;

4) $\sum z^2 = 1\,004\,583$; $\sum xy = 2\,217\,306$; $\sum xz = 682\,058$; $\sum yz = 3\,264\,917$.

Решаем эту систему относительно параметров a, b и c путем деления каждого уравнения на коэффициент при a. Получим:

$$a+136{,}16b+200{,}2c=650{,}24;$$

$$a+136{,}2514689b+200{,}3695652c=651{,}3824912;$$

$$a+136{,}2753247b+200{,}7158841c=652{,}3310689.$$

Затем, вычитая первое уравнение из второго, а второе из третьего, получим:

$$a+0{,}091468866b+0{,}169565217c=0{,}742491187;$$

$$a+0{,}023855815b+0{,}346318898c=0{,}948577744.$$

Далее разделим каждое из этих уравнений на коэффициент при b и определим разность между полученными уравнениями:

$$\begin{array}{r} _{-}\,b+1{,}853802672c=8{,}11742034 \\ b+14{,}51716892c=39{,}76295666 \\ \hline -12{,}66336625c=-31{,}64553631. \end{array}$$

Из разности уравнений:

$$c=\frac{-31{,}64553631}{-12{,}66336625}=2{,}498982946.$$

Подставив в одно из уравнений значение c, находим:

$$b+1{,}853802672\cdot 2{,}498982946=8{,}117420349,$$

т. е.

$$b+4{,}632621264=8{,}117420349,$$

откуда

$$b=8{,}117420349-4{,}632621264=3{,}484799086.$$

Для нахождения a в первое исходное уравнение вместо b и c подставим их значения:

$$25a+3404\cdot(3{,}484799086)+5005\cdot(2{,}498982946)=16\,266.$$

Отсюда

$$a=\frac{16\,266-24\,369{,}66573}{25}=\frac{-8103}{25}=-324{,}1466294.$$

Уравнение регрессии имеет вид

$$\hat{Y}=-324{,}1466294+3{,}484799086x+2{,}498982946z.$$

Поскольку переменные x и z измеряются в одних и тех же единицах (в данном случае в сантиметрах), можно отметить, что x сильнее воздействует на y, чем z, примерно в 1,4 раза, т. е. живая масса коров в большей степени зависит от высоты в холке, чем от обхвата груди.

Коэффициент множественной корреляции можно вычислить по следующей формуле:

$$R = \sqrt{1 - \frac{\sum e_i^2}{\sum (y_i - \overline{y})^2}},$$

где $\sum e_i^2 = \sum (y_i - \hat{y})^2$; $\sum (y_i - \overline{y})^2 = \sum y_i^2 - n\overline{y}^2$.

В нашем примере коэффициент множественной корреляции равен:

$$R = \sqrt{1 - \frac{10377{,}2476}{40335{,}76}} = 0{,}861816887,$$

а коэффициент множественной детерминации $R^2 = 0{,}742728347$.

Коэффициент множественной детерминации можно вычислить и другим способом по следующей формуле:

$$R^2 = \frac{\sum (\hat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2} = \frac{29958{,}682}{40335{,}76} = 0{,}742728347.$$

Из этого следует, что регрессия y на x и z на 74,3 % объясняет вариации значений y.

Для построения уравнения регрессии можно использовать матричный подход, рассмотренный выше, который применим и к более сложным линейным моделям. При этом линейная модель для двух независимых переменных в нашем примере имеет вид

$$\hat{y} = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + e,$$

где $\hat{y}$ — отклик, или живая масса коров; x_0 — фиктивная переменная, которая всегда равна единице; x_1 — высота коров в холке в см; x_2 — обхват груди, см; β — оценка метода наименьших квадратов.

Используя данные таблицы, составим необходимые векторы и матрицы для нашего примера:

$$\mathbf{Y} = \begin{vmatrix} 600 \\ 680 \\ 650 \\ \vdots \\ 740 \\ 640 \end{vmatrix}; \ \mathbf{X} = \begin{pmatrix} 1\,135\,193 \\ 1\,141\,200 \\ 1\,134\,194 \\ \vdots \\ 1\,138\,215 \\ 1\,139\,200 \end{pmatrix}; \ \boldsymbol{\beta} = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{vmatrix}; \ \mathbf{e} = \begin{vmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{24} \\ e_{25} \end{vmatrix}.$$

где $\mathbf{Y}$ — (25×1) — вектор; $\mathbf{X}$ — (25×3) — матрица; β — (3×1) — вектор; $\mathbf{e}$ — (25×1) — вектор.

Согласно изложенному выше, получим МНК оценки для β_0, β_1 и β_2:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

где **b** — вектор оценок элементов β, при условии, что матрица $X^T X$ не вырождена, т. е. ее определитель $\Delta \neq 0$.

Выше было указано, что:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix}; \ \mathbf{X}^T\mathbf{Y} = \begin{bmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \sum y_i x_{i2} \end{bmatrix}.$$

Из этого следует:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 135 & 141 & 134 & \cdots & 139 \\ 193 & 200 & 194 & \cdots & 200 \end{bmatrix} \begin{bmatrix} 1 & 135 & 193 \\ 1 & 141 & 200 \\ 1 & 134 & 194 \\ \vdots & \vdots & \vdots \\ 1 & 139 & 200 \end{bmatrix} =$$

$$= \begin{bmatrix} 25 & 3404 & 5005 \\ 3404 & 463800 & 682058 \\ 5005 & 682058 & 1004583 \end{bmatrix}.$$

Матрица, обратная к $X^T X$ следующая:

$$(X^T X)^{-1} = \begin{bmatrix} 61{,}3869 & -0{,}50133 & 0{,}03453 \\ -0{,}50133 & 0{,}00548 & -0{,}00123 \\ 0{,}03453 & -0{,}00123 & 0{,}00066 \end{bmatrix};$$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 135 & 141 & 134 & \cdots & 139 \\ 193 & 200 & 194 & \cdots & 200 \end{bmatrix} \begin{bmatrix} 600 \\ 680 \\ 650 \\ \vdots \\ 640 \end{bmatrix} = \begin{bmatrix} 16266 \\ 2217306 \\ 3264917 \end{bmatrix}.$$

Далее находим **b**:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} =$$

$$= \begin{bmatrix} 61{,}3869 & -0{,}50133 & 0{,}03453 \\ -0{,}50133 & 0{,}00548 & -0{,}00123 \\ 0{,}03453 & -0{,}00123 & 0{,}00066 \end{bmatrix} \begin{bmatrix} 16266 \\ 2217306 \\ 3264917 \end{bmatrix} =$$

$$= \begin{bmatrix} -324{,}1466294 \\ 3{,}484799086 \\ 2{,}498982946 \end{bmatrix}.$$

Найденное уравнение множественной регрессии имеет вид

$$\hat{Y} = -324{,}1466294 + 3{,}484799086x_1 + 2{,}498982946x_2.$$

Таким образом, уравнение множественной регрессии подобранное с помощью матричного анализа методом наименьших квадратов, идентично уравнению, подобранному этим методом, но другим способом.

6.12. Нелинейная регрессия

В биологии кроме линейных часто встречаются и нелинейные зависимости между переменными величинами. Во многих случаях зависимость между y и x может быть выражена уравнением параболы второго порядка:

$$y = a + bx + cx^2.$$

Параметры данного уравнения находят методом наименьших квадратов по следующей системе нормальных уравнений:

$$an + b\sum x + c\sum x^2 = \sum y;$$
$$a\sum x + b\sum x^2 + c\sum x^3 = \sum xy;$$
$$a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum yx^2.$$

Для решения этой системы относительно параметров a, b и c необходимо предварительно рассчитать $\sum x$, $\sum y$, $\sum xy$, $\sum x^2$, $\sum yx^2$, $\sum x^3$ и $\sum x^4$.

Пример. Исследованиями установлено, что удой коров симментальской породы сначала возрастает до пятой-шестой лактации, а затем уменьшается (табл. 6.13). В этой таблице приведены все суммы, необходимые для вычисления параметров a, b и c.

Составляем систему нормальных уравнений:

$$8a + 36b + 204c = 41{,}822,$$
$$36a + 204b + 1296c = 191{,}113,$$
$$204a + 1296b + 8772c = 1068{,}83.$$

Эту систему необходимо решить относительно коэффициентов a, b и c способом, который описан при вычислении параметров множественной регрессии.

Таблица 6.13

Динамика удоя коров симментальской породы по лактациям

x	y	xy	x^2	yx^2	x^3	x^4	$\hat{y}$
1	3,756	3,756	1	3,756	1	1	3,991
2	4,914	9,827	4	19,655	8	16	4,912
3	5,375	16,124	9	48,373	27	81	5,50
4	5,959	23,836	16	95,343	64	256	5,903
5	6,142	30,710	25	153,552	125	625	5,972

x	y	xy	x^2	yx^2	x^3	x^4	$\hat{y}$
6	5,958	35,749	36	214,494	216	1296	5,758
7	5,032	35,225	49	246,576	343	2401	5,259
8	4,485	35,885	64	287,077	512	4096	4,476

5) $n = 8$; $\sum x = 36$; $\sum y = 41{,}822$; $\sum x^2 = 204$; $\sum x^3 = 1296$; $\sum x^4 = 8772$; $\sum xy = 191{,}113$; $\sum yx^2 = 1068{,}83$

Решив эту систему находим: $a = 2{,}785104167$, $b = 1{,}3476399881$, $c = -0{,}14202877$.

Следовательно, уравнение параболы второго порядка имеет вид

$$\hat{y} = 2{,}785104167 + 1{,}3476399881x - 0{,}14202877x^2.$$

В табл. 6.13 в последнем столбце приведен ожидаемый удой коров (в тоннах) за любую лактацию, вычисленный по уравнению параболы второго порядка, который хорошо согласуется с эмпирическими данными.

Корреляционное отношение $\eta = 0{,}978$. Показатель криволинейной связи $\eta^2 = 0{,}956$. Эти показатели можно вычислить по следующей формуле:

$$\eta = \sqrt{1 - \frac{\sum e_i^2}{\sum (y - \hat{y})}} = \sqrt{1 - \frac{0{,}209}{4{,}774}} = \sqrt{0{,}956} = 0{,}978.$$

Задание для самостоятельной работы

Вычислить коэффициенты прямолинейной регрессии для большой и малой выборок, пользуясь данными заданий 1—15 главы 5.

Контрольные вопросы

1. Какие показатели применяются для измерения связи между признаками?
2. Как вычисляют коэффициент фенотипической корреляции в малых выборках?
3. Как вычисляют коэффициент фенотипической корреляции в больших выборках?
4. В чем заключается различие связи между признаками при положительных и отрицательных значениях коэффициента корреляции?
5. Как вычисляется коэффициент корреляции для альтернативных признаков?
6. В каких случаях используется коэффициент ранговой корреляции?

7. Как вычисляют коэффициент генетической корреляции?

8. Что характеризуют коэффициенты регрессии? В чем различие между коэффициентами $R_{x/y}$ и $R_{y/x}$?

9. В чем различие между коэффициентами r и R?

10. Какой может быть корреляция по направлению?

11. Какой может быть корреляция по величине?

12. В каких единицах измеряется коэффициент регрессии и корреляции?

13. Как характеризует связь между признаками коэффициент корреляции?

Глава 7
ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ. ОЦЕНКА ДОСТОВЕРНОСТИ ВЫБОРОЧНЫХ ПОКАЗАТЕЛЕЙ

Выше было отмечено, что конечной целью исследования является нахождение параметров генеральной совокупности, но в большинстве случаев генеральную совокупность изучают не непосредственно, а пользуясь выборкой. Ограниченность объема выборки и случайность отбора в нее объектов приводят к тому, что выборочные параметры отличаются от генеральных. Эти различия имеют объективную природу, т. е. возникают независимо от исследователя всегда, когда по части (выборке) пытаются охарактеризовать целое (генеральную) совокупность. Ошибки, возникающие при характеристике генеральной совокупности показателями, полученными при изучении выборки, называются ошибками репрезентативности. Не следует путать статистические ошибки с ошибками другого рода: ошибками типичности (когда выборка составлена неправильно и вследствие этого не является репрезентативной), ошибками прибора или инструмента, ошибками при измерении, ошибками в расчетах и т. д. Такого рода ошибки не вскрываются биометрическими методами, они должны быть устранены заранее. Более того, отсутствие таких ошибок в результате измерений дает основание для дальнейшей биометрической обработки материала с целью выявления статистических ошибок (ошибок репрезентативности), которых нельзя избежать при использовании выборочного метода и которые совершенно необходимо учитывать, чтобы сделать научно обоснованные выводы.

Ошибки репрезентативности показывают степень соответствия выборочных параметров параметрам генеральной совокупности. Чем меньше цифровые значения ошибки, тем точнее вычисленный параметр, тем ближе его значение к значению соответствующего параметра генеральной совокупности.

Согласно закону больших чисел практически маловероятно существенное отклонение выборочного параметра ($\bar{X}$, σ и др.) от соответствующего параметра генеральной совокупности, если число наблюдений достаточно велико. При изучении всех членов генеральной совокупности статистических ошибок быть не может, так как генеральный параметр находится (рассчитывается) непосредственно.

Пример. Если необходимо охарактеризовать по удою группу коров, закрепленных за каким-то оператором машинного доения, то, принимая во внимание ограниченный объем данной генеральной совокупности можно провести сплошное обследование, т. е. учесть удой у всех коров, входящих в данную группу. В этом случае рассчитанные по всем животным средняя арифметическая ($\bar{X}$) и среднее квадратическое отклонение (σ) будут генеральными параметрами.

Ошибки вычисляются для всех выборочных параметров и обычно обозначаются буквой m с подстрочным указанием знака того параметра, для которого они определяются: m_x; m_σ; m_{cv} и т. д. (в зарубежной литературе ошибку часто обозначают буквой S со знаком своего параметра: S_x; S_σ; S_{cv} и т. д.).

Если генеральная совокупность велика, то ее приравнивают к бесконечности (∞). В этом случае ошибку выборочной средней арифметической ($\bar{X}$) вычисляют по формуле

$$m_{\bar{X}} = \pm\frac{\sigma}{\sqrt{n}},$$

где $m_{\bar{X}}$ — ошибка средней арифметической ($\bar{X}$); σ — среднее квадратическое отклонение; n — численность выборки.

Из формулы следует, что величина ошибки средней арифметической зависит от значения σ и n, причем чем меньше разнообразие признака, тем меньше ошибка. При полной однородности совокупности по изучаемому признаку ($\sigma = 0$) средняя ошибка равна нулю, т. е. $\bar{X}$ выборки становится равной $\bar{X}$ генеральной совокупности. Величина средней ошибки обратно пропорциональна корню квадратному из объема выборки ($\sqrt{n}$). Поскольку в практической работе уровень варьирования признака (σ) изменить (уменьшить) обычно не представляется возможным, для повышения точности определения $\bar{X}_{ген}$ необходимо увеличить n.

В малых выборках $m_{\bar{X}}$ вычисляется по следующей формуле:

$$m_{\bar{X}} = \pm\frac{\sigma}{\sqrt{n-1}},$$

где $n-1=\nu$ — число степеней свободы.

Ошибка — величина именованная (см, кг, % и т. д.), выражается в тех же единицах измерения, что и средняя арифметическая. Средняя арифметическая величина, как и другие параметры совокупности, обычно записывается вместе со своей ошибкой: $\bar{X} \pm m_x$.

Допустим, в выборке из 100 коров среднесуточный удой $\bar{X} = 21{,}26$ кг, а $\sigma = \pm 3{,}68$. Ошибка средней арифметической в данном случае составит $m_{\bar{X}} = \pm\frac{\sigma}{\sqrt{n}} = \pm\frac{3{,}68}{\sqrt{100}} = \pm 0{,}368$ кг. Это означает, что средняя ошибка на 100 голов — 0,368 кг. Следовательно, среднесуточные удои изученной выборки характеризуются $\bar{X} \pm m_{\bar{X}} = 21{,}26 \pm 0{,}368$ кг.

Для следующего примера возьмем высоту в холке жеребцов чистокровной верховой породы, где $\bar{X} = 160{,}9$ см; $n = 93$; $\sigma = \pm 3{,}4$ см:

$$m_{\bar{X}} = \pm \frac{\sigma}{\sqrt{n}} = \pm \frac{3{,}4}{9{,}6} = \pm 0{,}35 \text{ см}; \ \bar{X} + m_{\bar{X}} = 160{,}9 \pm 0{,}35 \text{ см}.$$

Ошибки других выборочных показателей вычисляют по следующим формулам:

- среднего квадратического отклонения — $m_{\sigma} = \pm \dfrac{\sigma}{\sqrt{2n}}$;
- коэффициента вариации — $m_{cv} = \pm \dfrac{c_v}{\sqrt{2n}}$;
- коэффициента корреляции: для малой выборки — $m_r = \pm \sqrt{\dfrac{1 - r^2}{n - 2}}$;
- для большой выборки — $m_r = \pm \dfrac{1 - r^2}{\sqrt{n}}$;
- коэффициента регрессии — $m_{R_{y/x}} = \pm m_r \cdot \dfrac{\sigma_y}{\sigma_x}$; $m_{R_{x/y}} = \pm m_r \dfrac{\sigma_x}{\sigma_y}$.

Достоверность выборочных показателей (t) определяется отношением выборочного показателя к его средней ошибке по формулам

$$t_{\bar{X}} = \frac{\bar{X}}{m_{\bar{X}}}; \ t_{\sigma} = \frac{\sigma}{m_{\sigma}}; \ t_{cv} = \frac{cv}{m_{cv}}; \ t_r = \frac{r}{m_r}; \ t_R = \frac{R}{m_R}.$$

Задания для самостоятельной работы

Задание 1. Вычислить ошибки репрезентативности:

- средней арифметической для большой и малой выборок, пользуясь данными заданий 1—15 главы 3 (см. параграфы 3.1, 3.8);
- среднего квадратического отклонения для большой и малой выборок, пользуясь данными заданий 1—15 главы 4 (см. параграфы 4.1, 4.2);
- коэффициента вариации для большой и малой выборок, пользуясь данными заданий 1—15 главы 4 (см. параграф 4.4);
- коэффициента корреляции для больших и малых выборок, пользуясь данными заданий 1—15 главы 5 (см. параграфы 5.1, 5.2);
- коэффициента регрессии для большой и малой выборок, пользуясь данными заданий 1—15 главы 5 (см. параграф 5.7).

Задание 2. Вычислить достоверность выборочных показателей t для $\bar{X}$, σ, cv, r, R в больших и малых выборках, пользуясь данными задания 1 темы 6.

7.1. Вычисление доверительных границ для средней арифметической генеральной совокупности

Зная среднюю арифметическую ($\bar{X}$) и ошибку ($m_{\bar{X}}$) выборочной совокупности, можно с определенной степенью достоверности и точности определить те границы, в которых лежит средняя генеральной совокупности ($\tilde{X}$). Эти границы называют доверительными.

Предположим, что для изучения среднего настрига шерсти у мериносовых овец было сделано 100 выборок (при $n > 30$) и для каждой из них вычислено $\bar{X} \pm m_{\bar{X}}$. Доказано, что средние величины отдельных выборок группируются вокруг средней для генеральной совокупности ($\tilde{X}$), подчиняясь закону, согласно которому выборочная $\bar{X}$ отклоняется от $\tilde{X}$ генеральной совокупности:

- в 95 % случаев — не более чем на $1{,}96m$,
- в 99 % случаев — не более чем на $2{,}58m$,
- в 99,9 % случаев — не более чем на $3{,}29m$.

Приведенные выше показатели — 95, 99 и 99,9 % — называются доверительными вероятностями (P). Их обозначают обычно не в процентах, а в долях единицы ($P = 0{,}95$; $P = 0{,}99$; $P = 0{,}999$). Они указывают на вероятность безошибочного прогноза. Коэффициенты, стоящие при средней ошибке (1,96; 2,58 и 3,29), представляют собой нормированные отклонения (t), соответствующие приведенным доверительным вероятностям.

Используя эти показатели, можно по данным одной выборки определить доверительные границы, в пределах которых лежит средняя генеральной совокупности ($\tilde{X}$). Она находится между $\bar{X} - t \cdot m$ (нижняя граница) и $\bar{X} + t \cdot m$ (верхняя граница). Нормированное отклонение (t) зависит от доверительной вероятности, которую выбирают исходя из требований, предъявляемых к достоверности выводов (табл. 7.1).

Таблица 7.1

Три порога надежности или вероятности безошибочных прогнозов для больших выборок

Порог	Применение	Вероятность безошибочных прогнозов (P)	Число ошибочных случаев	Нормированное отклонение или критерий надежности (t)	Минимальный объем выборки
1	В производственных и научно-производственных исследованиях	0,95	5 из 100	1,96	30
2	В большинстве биологических, зоотехнических и ветеринарных исследований	0,99	1 из 100	2,58	100

Порог	Применение	Вероятность безошибочных прогнозов (P)	Число ошибочных случаев	Нормированное отклонение или критерий надежности (t)	Минимальный объем выборки
3	В работах с очень высокими требованиями к достоверности выводов	0,999	1 из 1000	3,29	200

Для малых выборок стандартные значения (показатель надежности) определяются по таблице Стьюдента (см. приложение 1).

Пример 1. Средний настриг шерсти в выборке мериносовых овец ($\bar{X} \pm m_{\bar{X}}$) составил 4,0 ± 0,2 кг. Установить доверительные границы для среднего настрига шерсти в генеральной совокупности мериносов.

Возьмем в качестве доверительной вероятности $P = 0,95$. Согласно данным табл. 7.1, при этой вероятности нормированное отклонение $t = 1,96$. Определяя доверительные границы для $\bar{X}$ генеральной совокупности мериносов, получим: нижняя граница составит $\bar{X} - 1,96 \cdot m = 4,0 - 1,96 \cdot 0,2 = 3,61$ кг; верхняя граница — $\bar{X} + 1,96 \cdot m = 4,0 + 1,96 \cdot 0,2 = 4,39$ кг. Это показывает, что генеральная средняя (т. е. средний настриг шерсти среди всего поголовья мериносов) находится в интервалах между 3,61 и 4,39 кг. Вероятность того, что данное утверждение правильно, составляет 95 %, а риск ошибки — 5 %.

Если считать, что 5 % риск ошибки слишком высок, то в качестве допустимого можно взять 1 % риск ошибки, а следовательно, 99 % вероятность достоверности утверждения. При P равном 0,99, $t = 2,58$. Рассчитывая в таком случае доверительные границы для $\bar{X}$ генеральной совокупности по формуле $\bar{X} \pm 2,58 \cdot m$, получим: нижняя граница равна 4,0 – 0,52 = 3,48, верхняя — 4,0 + 0,52 = 4,52. В этом случае можно утверждать, что генеральная средняя находится в границах 3,48—4,52. Точность второго утверждения по сравнению с предыдущим уменьшилось, так как границы расширились, но уменьшился и риск ошибки с 5 %-го до 1 %-го уровня. В случаях, требующих особо высокой достоверности выводов, в качестве доверительной вероятности значения используют P, равное 0,999. При этом риск ошибки снижается до одного случая из 1000, но доверительный интервал становится еще более широким.

Пример. Предположим, что поставлена задача — построить новый животноводческий комплекс, в котором будет внедрена вновь разработанная технология откорма до 15-месячного возраста около 5 тыс. голов бычков. Какова будет живая масса одного бычка (в среднем), выращенного в этих условиях?

Прежде чем начать строительство комплекса, необходимо произвести эксперимент на бычках из тех же хозяйств, которые будут поставлять их для откорма на комплексе в условиях новой технологии. Обработанные материалы этого эксперимента дали следующие выборочные показатели: n = 100 голов; $\bar{X} = 400$ кг и σ = 40 кг.

При планировании исследования был установлен обычный для научно-хозяйственных опытов (поисковых) критерий надежности $t_{0,95}$ = 1,96 (первый порог вероятности безошибочных прогнозов). Имея эти данные, мы можем определить ошибку средней арифметической и установить доверительные группы для генеральной средней: $m_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{100}} = \frac{40}{10} = 4$ кг. При установленном критерии надежности t = 1,96 доверительные границы для генеральной средней составят: $\bar{X}_{ген} = \bar{X}_{выб} \pm t \cdot m_{\bar{X}} = 400 \pm 1{,}96 \cdot 4 \approx 392 \div 408$ кг, т. е. к 15-месячному возрасту средняя живая масса бычков будет не меньше гарантированного минимума (392,2 кг), но и не больше вероятного максимума (407,8 кг), хотя отдельные животные достигнут к этому возрасту довольно разной живой массы: от $\bar{X} - 2\sigma$ $(400 - 2 \cdot 40 = 320)$ до $\bar{X} + 2\sigma$ $(400 + 2 \cdot 40 = 480$ кг). Полученные данные позволяют сделать вывод, что при откорме 5 тыс. телят за указанный период их общая живая масса составит не менее: 392 кг · 5000 = 1960 т. Эта цифра может служить плановым показателем производства мяса (в живой массе). В то же время необходимо готовить материальную производственную базу для переработки не 1960 т, а вероятного максимума: 408 кг · 5000 = 2040 т.

Задание для самостоятельной работы

Установить доверительные границы по всем показателям, пользуясь данными заданий 1—15 главы 2 для средней арифметической генеральной совокупности с учетом трех степеней вероятности безошибочных прогнозов (P).

7.2. Вычисление достоверности разности между средними величинами двух выборок

В биометрических исследованиях возникает необходимость сравнить средние арифметические двух групп животных (например, среднюю живую массу животных опытной и контрольной групп, среднюю продуктивность дочерей двух производителей и т. д.). Средние двух сравниваемых групп всегда в некоторой мере отличаются друг от друга. Поэтому важно установить, достоверна ли разность между средними. «Достоверность» — основное понятие в биометрии. Оно означает возможность обобщить данные опыта, перенести результаты исследования выборок на соответствующие генеральные совокупности.

Предположим, что в опыте выборочная средняя 1-й группы ($\bar{X}_1$) оказалось больше выборочной средней 2-й группы ($\bar{X}_2$), т. е. $\bar{X}_1 > \bar{X}_2$. Если разность выборочных средних достоверна, то это означает, что такую же по знаку разность, какая получена в эксперименте, следует ожидать (с принятым уровнем надежности выводов) и между соответствующими генеральными средними, т. е. большей по величине выборочной средней соответствует и бóльшая генеральная средняя, а меньшей по величине выборочной средней соответствует меньшая генеральная средняя. Это утверждение можно выразить формулой: если $\bar{X}_1 > \bar{X}_2$ и разность достоверна, то $\tilde{X}_1 > \tilde{X}_2$. В том случае, если разность недостоверна, остается неизвестным, находятся ли генеральные средние в таком же соотношении между собой, как и выборочные средние, или их значения противоположны, или они равны. Другими словами, если в эксперименте $\bar{X}_1 > \bar{X}_2$, но разность недостоверна, то не ясно $\tilde{X}_1 > \tilde{X}_2$, $\tilde{X}_1 = \tilde{X}_2$ или $\tilde{X}_1 < \tilde{X}_2$. Для выборок оптимального объема недостоверность разности означает принятие «нулевой гипотезы» о равенстве генеральных средних.

Для определения достоверности разности выборочных средних рассчитывают критерий достоверности разности t_d по формуле

$$t_d = \frac{d}{m_d} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m^2_{\bar{X}_1} + m^2_{\bar{X}_2}}},$$

где d — разность средних арифметических выборок, $d = \bar{X}_1 - \bar{X}_2$ (обычно из большей средней вычитают меньшую или берут d по модулю, т. е. положительной); m_d — ошибка разности: $\sqrt{m^2_{\bar{X}_1} + m^2_{\bar{X}_2}}$; $m^2_{\bar{X}_1}$, $m^2_{\bar{X}_2}$ — квадраты ошибок соответствующих средних арифметических.

Полученное фактическое значение t_d нужно сравнить с t табличным, взятым из таблицы стандартных значений критерия Стьюдента (см. приложение 1): для числа степеней свободы $\nu_d = n_1 + n_2 - 2$ в малых выборках и $\nu_d = n_1 + n_2$ в больших выборках, где n_1 — число вариант в одной группе, n_2 — число вариант во второй группе.

Существует три уровня достоверности разности: $P \geq 0{,}95$; $P \geq 0{,}99$; $P \geq 0{,}999$. При $t_d = 1{,}96$ разность достоверна в 95 случаях из 100 (95 %), при $t_d = 2{,}58$ она достоверна в 99 случаях из 100 (99 %), а при $t_d = 3{,}29$ разность достоверна в 999 случаях из 1000 (99,9 %). Если же величина t_d меньше 1,96, то разность между средними сравниваемых групп не может быть признана достоверной.

Пример. Предположим, что в предгорном районе проводится опыт, цель которого — определить оптимальный способ нагула молодняка крупного рогатого скота в летний период на низинах или на горных пастбищах. Для этого было сформировано три группы молодняка, аналогичных по полу, возрасту и породности. 1-я группа выпасалась на горных пастбищах, 2-я и 3-я — на низинах, однако животных 3-й группы во время сильной жары загоняли в помещения, оборудованные автопоилками, и подкармливали свежескошенной травой.

Материалы опыта обработали биометрически и получили следующие результаты:

$$1)\ n = 100 \text{ голов};\ \bar{X}_1 = 330 \text{ кг};\ \sigma_1 = 40 \text{ кг};\ m_{\bar{X}_1} = 4 \text{ кг};$$

2) $n = 64$ голов; $\bar{X}_2 = 340$ кг; $\sigma_2 = 56$ кг; $m_{\bar{X}_2} = 7$ кг;

3) $n = 100$ голов; $\bar{X}_3 = 380$ кг; $\sigma_3 = 60$ кг; $m_{\bar{X}_3} = 6$ кг.

Определим достоверность разности между выборочными средними $\bar{X}_1$ и $\bar{X}_2$, а также между $\bar{X}_1$ и $\bar{X}_3$. В первом случае:

$$t_d = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{m_{\bar{X}_1}^2 + m_{\bar{X}_2}^2}} = \frac{340-330}{\sqrt{4^2+7^2}} = \frac{10}{8} = 1,25.$$

Полученное значение $t_d = 1,25$ необходимо сравнить с $t_{\text{табл}}$, взятым из таблицы стандартных значений критерия Стьюдента. Для входа в таблицу необходимо рассчитать число степеней свободы по формуле $\nu_d = n_1 + n_2$. В нашем случае $\nu_d = n_1 + n_2 = 100 + 64 = 164$. В таблице (см. приложение 1) для $\nu = 120$ и больше находим значение $t_{\text{табл}}$ для трех уровней надежности выводов — 0,95; 0,99 и 0,999. Они будут следующими 1,96; 2,58 и 3,29. Сравнение t фактического с t табличными показывает, что t фактическое меньше любого из трех табличных значений t. Разность оказалась недостоверной. Это означает, что хотя у животных, выпасавшихся на низинных пастбищах, и были более высокие показатели живой массы, у нас нет оснований обобщить полученные результаты и перенести их на генеральные совокупности. Генеральными совокупностями в данном случае являются все животные, которых в принципе можно было бы откормить на горных (1-я совокупность) и на низинных (2-я совокупность) пастбищах. Осталось невыясненным, в каком соотношении находятся генеральные средние этих совокупностей. Учитывая, что в опыте использовались достаточно большие выборки ($n_1 = 100$ и $n_2 = 64$), можно принять предположение о равенстве генеральных средних.

Если всех животных (а не только взятых для опыта) нагуливать на низинных пастбищах, то их средняя масса будет равна средней живой массе всех животных, нагуливающихся на горных пастбищах.

Во втором случае:

$$t_d = \frac{\bar{X}_3 - \bar{X}_1}{\sqrt{m_{\bar{X}_1}^2 + m_{\bar{X}_3}^2}} = \frac{380-330}{\sqrt{4^2+6^2}} = \frac{50}{7,2} = 6,94.$$

Полученное значение $t_d = 6,94$ нужно сравнить с $t_{\text{табл}}$, которые для $\nu_d = 100 + 100 = 200$ составляют соответственно 1,96; 2,58 и 3,29. Фактическое значение t_d превышает любое из трех табличных. Обычно это отмечается либо подчеркиванием фактического значения t_d, либо соответствующим числом звездочек: если $t_{\text{факт}}$ выше t_1 — одной чертой или звездочкой, если выше t_2 — двумя, а выше t_3 — тремя черточками или звездочками. В нашем случае $t_{\text{факт}} > t_3 = 3,29$, что можно выразить как $t_d = \underline{\underline{\underline{6,94}}}$ или $t_d = 6,94^{***}$. Это означает, что разность достоверна с высшей степенью надежности выводов ($P > 0,999$) и, следовательно, полученные результаты ($\bar{X}_3 > \bar{X}_1$) можно обобщить и перенести на генеральные совокупности. Вывод о том, что большей выборочной средней будет соответствовать и бóльшая генеральная средняя научно обоснованная. При нагуле всех (а не только взятых для опыта) животных по третьей схеме (на низинных пастбищах с подкормкой) их средняя живая масса будет выше, чем средняя живая масса всего молодняка при нагуле на горных пастбищах. После экономической оценки изученных способов нагула третий способ можно рекомендовать для широкого использования.

Пример. Средняя высота в холке кобыл чистокровной верховой породы $\bar{X}_1 = 157,4$ см, ошибка средней арифметической $m_{\bar{X}_1} = 0,35$ см. $n_1 = 10$. Средняя высота в холке жеребцов чистокровной верховой породы $\bar{X}_2 = 160,9$ см, а ошибка средней арифметической $m_{\bar{X}_2} = 0,35$ см. $n_2 = 10$. Установить достоверность разности между этими группами животных по высоте в холке.

Критерий достоверности разницы по промеру высоты в холке двух групп сравниваемых животных равняется:

$$t_d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m_{\bar{X}_1}^2 + m_{\bar{X}_2}^2}} = \frac{160,9 - 157,4}{\sqrt{0,35^2 + 0,35^2}} = \frac{3,5}{0,49} = 7,1;$$

$$\nu = 10 + 10 - 2 = 18.$$

Разность оказалась достоверной, критерий достоверности разницы выше третьей степени вероятности $P_3 \geq 0,999$, т. е. из 1000 в 999 случаях наблюдалась разница в данном признаке. Биологически это объяснимо: проявляется половой диморфизм (самцы в среднем крупнее самок).

Пример. Сравнивается молочная продуктивность первотелок бестужевской и холмогорской пород (большие выборки). В одинаковых условиях кормления и содержания получены следующие показатели удоя за первую лактацию: по бестужевскому скоту — $\bar{X}_1 \pm m_{\bar{X}_1} = 2600 \pm 30$ кг; по холмогорскому скоту — $\bar{X}_2 \pm m_{\bar{X}_2} = 3200 \pm 40$ кг. Установить достоверность разности между удоями в этих группах. Вычисляя t_d, получаем:

$$t_d = \frac{d}{\sqrt{m_{\bar{X}_1}^2 + m_{\bar{X}_2}^2}} = \frac{3200 - 2600}{\sqrt{40^2 + 30^2}} = \frac{600}{\sqrt{2500}} = \frac{600}{50} = 12.$$

Критерий достоверности разницы ($t_d = 12$) значительно превышает величины, приведенные в приложении 1 (таблица Стьюдента). Поэтому можно с вероятностью, превышающей 99,9 %, утверждать, что холмогорские коровы более продуктивны по сравнению с бестужевскими.

7.3. Определение достоверности средней разности при изучении совокупностей с попарно связанными вариантами

Выше мы рассматривали примеры определения достоверности разности выборочных средних, когда выборки брались из двух разных генеральных совокупностей, независимых друг от друга.

Однако в практике зоотехнических исследований приходится иметь дело и с зависимыми друг от друга генеральными совокупностями, оценивать генеральные параметры по выборкам, в которых объекты одной так или иначе связаны с объектами другой. Такие ситуации являются типичными в случае проведения эксперимента методом периодов, когда, например, группу животных сначала наблюдают в определенный период — контрольный, а затем подвергают воздействию изучаемого фактора. Таким образом, получают две сонокупности данных наблюдений, в которых варианты попарно связаны: по каждому животному есть два значения признака: до воздействия (контроль) и после воздействия. То же происходит при проведении опытов методом аналогов: совокупности представлены разными животными, но каждому животному одной группы соответствует аналог (по полу, возрасту, сроку отела, живой массе и т. д.) в другой группе (полными аналогами являются монозиготные близнецы). В селекционно-генетических исследованиях при оценке производителей по качеству потомства попарно связанными будут показатели матерей и дочерей. Для таких условий существуют два способа определения достоверности средней разности.

1. Рассчитывают критерий достоверности разности выборочных средних, но ошибку разности определяют по более сложной формуле

$$t_d = \sqrt{m_1^2 + m_2^2 - 2r m_1 \cdot m_2},$$

где r — коэффициент корреляции между значениями признака в двух сравниваемых совокупностях. Из формулы видно, что если выборки независимы, то $r = 0$ и формула превращается в уже известную формулу для расчета ошибок разности $t_d = \sqrt{m_1^2 + m_2^2}$. Таким образом, критерий достоверности разности выборочных средних для совокупностей с попарно связанными вариантами рассчитывают по формуле

$$t_d = \frac{d}{m_d} = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{m_{\bar{X}_1}^2 + m_{\bar{X}_2}^2 - 2r m_{\bar{X}_1}^2 \cdot m_{\bar{X}_2}^2}}.$$

Сравнение с $t_{табл}$ для $\nu = n_1 + n_2 - 2$ далее проводится обычным способом.

2. Необходимость расчета коэффициента корреляции несколько усложняет применение данной формулы на практике. Существует второй, более простой способ сравнения генеральных средних: не путем оценки достоверности разности средних, а путем оценки достоверности средней разности, рассчитываемой по всем разностям между попарно связанными вариантами.

Пример. Оценить быка-производителя Жасмина 171 голштинской породы, использованного в учебном хозяйстве Ульяновской ГСХА, по потомству путем сравнения показателей его дочерей с показателями их матерей по удою за первую лактацию (табл. 7.2).

Среднюю разность рассчитывают по формуле

$$\bar{d} = \frac{\sum d}{n} = \frac{251{,}1}{19} = 13{,}22.$$

Она равна разности средних арифметических $(\bar{X}_1 - \bar{X}_2)$. Далее находят ошибку средней разности:

$$C_d = \sum d^2 - \frac{\left(\sum d\right)^2}{n}; \; C_d = 7454{,}55 - \frac{(251{,}1)^2}{19} = 4136{,}07;$$

$$m_d = \sqrt{\frac{C_d}{n \cdot (n-1)}} = \sqrt{\frac{4136{,}07}{19 \cdot (19-1)}} = 3{,}48.$$

Рассчитаем критерий достоверности средней разности:

$$t_d = \frac{\bar{d}}{m_{\bar{d}}} = \frac{13{,}22}{3{,}48} = 3{,}80.$$

Таблица 7.2

Оценка быка-производителя Жасмина 171 по качеству потомства

Пары дочь-мать	Удой за лактацию, ц		$d = v_1 - v_2$	d^2
	дочери v_1	матери v_2		
1	61,7	32,0	29,7	882,09
2	76,3	46,3	30,0	900,00
3	36,5	32,5	4,0	16,00
4	86,4	41,2	44,2	1953,64
5	62,0	35,7	26,3	691,69
6	52,9	61,9	—9,0	81,00
7	59,5	41,6	17,9	320,41
8	59,3	33,0	26,3	691,69
9	52,2	43,3	8,9	79,21
10	67,9	61,9	6,0	36,00
11	47,9	45,6	2,3	5,29
12	57,3	36,4	20,9	436,81
13	66,3	34,3	32,0	1024,00
14	42,3	37,4	4,9	24,01
15	31,4	34,3	–2,9	8,41
16	45,0	56,6	–11,6	134,56

Окончание табл. 7.2

Пары дочь-мать	Удой за лактацию, ц		$d=v_1-v_2$	d^2
	дочери v_1	матери v_2		
17	51,2	40,5	10,7	114,49
18	51,5	57,0	5,5	30,25
19	55,9	50,9	5,0	25,00
$n=19$			$\sum d=251,1$	$\sum d^2=7454,55$

Фактическое значение сравним с табличным. Для нахождения $t_{табл}$ число степеней свободы в данном случае следует брать $v = n - 1$ (а не $v = n_1 + n_2 - 2$), $t_{табл}$ для $v=n-1=19-1=18\cdot\{2,10 - 2,88 - 3,92\}$. $t_{факт} > t_{табл}$ для вероятности 0,99. Следовательно, разность достоверна с высоким уровнем надежности выводов ($P>0,99$). Данного производителя можно оценить как улучшателя, т. е. если получить от него не 19, а большее количество потомков на матерях такого же качества, что и в опыте, и в таких же условиях выращивания и кормления, то средняя продуктивность всех его дочерей будет выше средней продуктивности матерей.

Пример. Живая масса дочерей быка-производителя Юга 553 равна:

$$\bar{X}_1 \pm m_{\bar{X}_1} = 530 \pm 10,\ n = 20;$$

живая масса матерей равна:

$$\bar{X}_2 \pm m_{\bar{X}_2} = 500 \pm 12 \text{ кг},\ n = 20;\ r = 0,6;\ d = \bar{X}_1 - \bar{X}_2 = 530 - 500 = 30 \text{ кг};$$

$$m_d = \pm\sqrt{m^2_{\bar{X}_1} + m^2_{\bar{X}_2} - 2 \cdot r m^2_{\bar{X}_1} \cdot m^2_{\bar{X}_2}} = \pm\sqrt{10^2 + 12^2 - 2 \cdot 0,6 \cdot 10 \cdot 12} = \pm 10;$$

$$t_d = \frac{30}{10} = 3;\ v = 20 + 20 - 2 = 38.$$

По таблице Стьюдента находим, что $P>0,99$, следовательно, разность достоверна.

Задания для самостоятельной работы

Задание 1. Вычислить достоверность разницы между обхватом груди кобыл брабансонской и орлонской пород.

Данные о промерах обхвата груди кобыл брабансонской породы, см ($\bar{X}_1$):

190	205	193	188	193	185	190	200	186	188
188	190	190	183	188	184	188	183	196	191
191	200	190	190	191	206	195	190	200	198
195	188	189	186	200	216	195	192	200	207
194	190	185	190	192	182	199	200	200	190

Данные о промерах обхвата груди кобыл орловской породы, см ($\bar{X}_2$):

169	178	191	179	173	168	168	170	172	185
165	180	176	179	170	179	180	170	170	168
185	170	183	178	182	172	180	171	172	175
180	181	183	169	181	169	164	182	178	184
169	179	173	170	172	177	169	174	176	181

Задание 2. В результате экспериментов по изучению эффективности межпородного скрещивания получены данные о живой массе чистопородных и помесных поросят в двухмесячном возрасте. Определить достоверность различий по живой массе между чистопородными и помесными животными.

Живая масса чистопородных поросят, кг:

16	16	11	12	14	10	15	16	17	18	11	17	12	13	12	12
11	11	11	13	14	11	12	12	12	12	12	13	14	14	15	15
15	15	15	20	20	12	15	18	16	16	16	17	17	17	16	12

Живая масса помесных поросят, (кг):

18	18	19	17	18	20	21	22	20	16	18	17	19	20	22	20
16	15	16	20	20	21	20	19	16	17	17	16	19	20	17	18
19	20	18	17	15	16	17	18	20	21	20	21	20	19	18	14

Задание 3. Вычислить достоверность разницы между живой массой коров красной горбатовской и черно-пестрой пород.

Данные о живой массе коров красной горбатовской породы, кг:

489	475	470	432	495	467	411	421	488	536
462	488	429	456	426	441	506	438	430	388
473	534	442	485	407	491	475	490	440	425
545	450	426	485	433	426	440	439	488	390
524	445	539	482	447	504	440	391	500	550

Данные о живой массе коров черно-пестрой породы, кг:

444	536	500	466	462	515	501	537	529	582
500	544	513	549	511	534	584	601	571	612
515	522	517	572	559	501	505	452	417	493
554	551	510	411	491	500	429	372	500	487
483	493	407	371	449	562	417	423	437	448

Задание 4. Определить достоверность разницы по удою за первую лактацию между коровами бестужевской и черно-пестрой пород, лактировавших в условиях ОПХ «Тимирязевское» Ульяновского НИИСХ.

Данные об удое коров бестужевской породы, кг:

3900	4230	4080	2660	3356	3923	4806	3693	3246	2948
3648	4092	4837	3674	3441	3825	3259	4231	4742	3556
3256	3857	3879	3440	4862	3037	2990	4248	3637	2985
3218	3285	2837	3050	3625	2984	3028	4027	3089	4034
3235	4454	3135	5118	3295	2864	3972	3645	4358	5442

Данные об удое коров черно-пестрой породы, кг:

4342	3687	4284	3639	2323	4247	5154	5416	5005	5109
5670	5134	3305	2474	3032	3379	4353	3150	4217	5744
4396	3401	2728	5165	4842	2888	4640	2100	3630	2177
4130	5883	2312	4558	5142	4100	3415	4353	4188	5066
4247	5047	3403	6005	3508	4947	4089	3188	4779	3925

Задание 5. Вычислить достоверность разницы по содержанию жира в молоке за третью лактацию между коровами красной степной и черно-пестрой пород.

Данные о средней жирности молока коров красной степной породы, %:

3,31	3,62	3,70	3,55	3,77	3,48	3,56	3,38	4,51	3,68
3,52	3,45	3,94	3,29	3,24	3,29	4,12	3,43	3,38	4,82
3,33	3,76	3,77	3,73	3,46	3,70	3,31	3,76	3,54	3,83
3,80	3,71	3,62	3,42	3,88	3,83	3,79	3,45	3,33	3,55
3,84	3,53	3,78	3,98	3,51	3,45	3,53	3,27	3,50	3,53

Данные о средней жирности молока коров черно-пестрой породы, %:

3,42	3,82	3,88	3,77	3,43	3,32	3,48	3,49	3,53	3,81
3,41	4,12	3,54	3,63	3,24	3,74	3,20	3,26	3,71	3,45
3,30	3,43	3,35	3,59	3,08	3,29	3,41	3,48	4,09	3,62
3,26	3,86	4,28	3,62	3,51	3,61	3,33	3,30	4,03	3,80
3,64	3,29	3,80	3,38	3,06	3,55	3,52	3,52	3,66	3,06

Задание 6. Замечено, что если вымя коровы подвергнуть тепловому воздействию, т. е. теплым компрессам, то это мероприятие будет способствовать повышению жирномолочности. Для проверки данного наблюдения был поставлен опыт. В хозяйстве для этой цели выделили две группы коров. Первая из них была опытной, у которой вымя подвергалось обогреванию, и вторая контрольная — без обогревания. У опытной группы процент жира в молоке оказался равным

$\bar{X}_1 \pm m_{\bar{x}} = 3{,}87 \pm 0{,}07$, у контрольной — $\bar{X}_2 \pm m_{\bar{x}_2} = 3{,}37 \pm 0{,}09$. Вычислить ошибку и критерий достоверности выборочной разницы.

Задание 7. Определить достоверность разности между настригом шерсти овец в связи с различным типом гемоглобина и породой (табл. 7.3).

Таблица 7.3

Тип гемоглобина и настриг шерсти у овец

Порода	Тип гемоглобина					
	A		AB		B	
	n	настриг шерсти, кг	n	настриг шерсти, кг	n	настриг шерсти, кг
Советский меринос	14	5,39 ± 0,19	125	5,69 ± 0,06	268	5,45 ± 0,04
Эдельбаевская	21	3,10 ± 0,09	40	2,97 ± 0,06	75	3,05 ± 0,03

Задание 8. Определить при $P = 0{,}95$ достоверность разницы и доверительные границы показателей симментальского скота Ульяновской области разного типа конституции (табл. 7.4).

Таблица 7.4

Показатели продуктивности симментальских коров разного типа конституции

Показатели	Широкотелый тип ($\bar{X} \pm m_{\bar{x}}$)	Узкотелый тип ($\bar{X} \pm m_{\bar{x}}$)
Живая масса, кг	630 ± 10,2	563 ± 8,4
Наивысший удой за лактацию, кг	4536 ± 92,6	4180 ± 105,0
Содержание жира в молоке, %	3,91 ± 0,03	3,93 ± 0,04

Задание 9. При изучении влияния инбридинга (родственного спаривания) на мясные качества маток различных семейств получены следующие показатели среднесуточного прироста.

Показатели среднесуточного прироста у инбредного и аутбредного потомства свиней:

Семейство	n	$\bar{X} \pm m_{\bar{x}}$	Семейство	n	$\bar{X} \pm m_{\bar{x}}$
Волшебницы:			Черной птички:		
аутбредные	108	716 ± 6,5	аутбредные	44	667 ± 8,6
инбредные	76	718 ± 16,2	инбредные	8	829 ± 8,5

Одинаково ли влияние инбридинга в обоих семействах? Какова достоверность полученных показателей среднесуточного прироста между аутбредными и инбредными животными в обоих семействах?

Задание 10. Определить, существуют ли достоверные породные различия в активности ферментов крови коров (табл. 7.5).

Таблица 7.5

Активность ферментов крови коров в среднем за лактацию ($\bar{X} \pm m_{\bar{X}}$)

Порода	n	Амилаза, %	Щелочная фосфатаза, мг %	Альдолаза, ед.
Айрширская	17	12,86 ± 0,35	2,32 ± 0,19	23,62 ± 0,77
Голландская	20	13,75 ± 0,35	2,17 ± 0,15	24,82 ± 0,88
Черно-пестрая	18	10,37 ± 0,47	1,85 ± 0,14	25,98 ± 0,52

Задание 11. Достоверны ли различия в живой массе у овец гиссарской породы трех типов конституции при доверительных вероятностях 0,95 и 0,99 по следующим данным:

Тип конституции	Грубая	Крепкая	Нежная
Живая масса	82,4 ± 1,10	78,9 ± 0,84	69,9 ± 0,88

Задание 12. Определить достоверность разности содержания лейкоцитов (тыс.) в 1 мм^3 крови здоровых и больных пневмонией овец:

Здоровые	5,8	6,2	8,3	6,0	9,8	7,4	7,2	8,6	7,7	8,0
Больные	12,0	12,1	11,8	11,4	11,5	11,1	13,6	10,5	12,7	13,0

Задание 13. Определить достоверность разницы между годовыми удоями у коров красной степной породы различного возраста. Годовой удой по первой лактации $\bar{X} \pm m_{\bar{X}} = 1630 \pm 52$ кг, по третьей — $\bar{X} \pm m_{\bar{X}} = 3180 \pm 57$кг, по шестой — $\bar{X} \pm m_{\bar{X}} = 3320 \pm 71$ кг.

Задание 14. За 233 дня ($\bar{X}_1$) живая масса мелких (0,92 кг) при рождении поросят составил 100 кг при средней ошибке $m_{\bar{X}_1} = \pm 5$ дней. У крупных при рождении поросят (1,57 кг) такая же масса была достигнута за 212 дней ($\bar{X}_2$) при средней ошибке $m_{\bar{X}_2} = \pm 4$ дня. Вычислить критерий достоверности разности скороспелости мелких и крупных поросят при рождении.

Задание 15. При исследовании нуклеиновых кислот в клетках вымени лактирующих коров получены следующие результаты:

Порода	Концентрация нуклеиновых кислот, усл. ед.	
	ядро	цитоплазма
Голландская	164,0 ± 1,40	137,2 ± 2,10
Холмогорская	128,8 ± 2,08	73,5 ± 3,03

Определите достоверность полученных данных.

7.4. Определение необходимого объема выборки

При проведении научного эксперимента, прежде всего, возникает вопрос о том, какие объекты и в каком количестве необходимо включить в выборку, чтобы получить достаточно надежные, достоверные и объективные выводы, т. е. вопрос о минимальном объеме выборки. При этом нужно учитывать степень изменчивости изучаемого признака: чем она выше, тем больше должен быть объем выборки. Для определения объема выборки пользуются формулой: $n=\dfrac{N}{N\left(\dfrac{d}{t}\right)^2+1}$, где

n — искомый объем выборки; N — численность генеральной совокупности; $d=\dfrac{\Delta}{\sigma}$ — допускаемое расхождение между $\bar{X}_{\text{выб}}$ и $\bar{X}_{\text{генер}}$, выраженное в долях σ; Δ — абсолютная допускаемая разность, $\Delta=\bar{X}_{\text{генер}}-\bar{X}_{\text{выб}}$; σ — среднее квадратическое отклонение, которое можно ориентировочно определить по лимиту изменчивости признака $\sigma=\dfrac{X_{max}-X_{min}}{6}$ исходя из закономерности, что лимит равен $\bar{X}\pm3\sigma$; t — критерий достоверности, который соответствует взятому уровню вероятности ($t_{0,95}\approx2{,}0$; $t_{0,99}\approx2{,}6$; $t_{0,999}\approx3{,}3$).

Если генеральная совокупность имеет большую численность ($N\to\infty$), то формула упрощается: $n=t^2:d^2$.

Пример. Требуется определить объем выборки, который необходим для получения достоверности в показателе средней плодовитости свиноматок при $t=2$. Предположим, что плодовитость в популяции колеблется от 6 до 18 поросят. Тогда $\sigma=(18-6):6=2$ головам. Ставят условие, что $\bar{X}_{\text{выб}}$ отличалось от $\bar{X}_{\text{генер}}$ на $\Delta=0{,}5$ головы. Тогда $d=\dfrac{\Delta}{\sigma}=\dfrac{0{,}5}{2}=0{,}25$. Отсюда $n=\dfrac{t^2}{d^2}=\dfrac{2^2}{0{,}25^2}=\dfrac{4}{0{,}0625}=64$ головы.

Следовательно, чтобы получить достоверный показатель средней плодовитости животных при планируемых условиях, нужно включить в выборку 64 свиноматки.

С целью получения достоверной разности (t_d) между двумя средними арифметическими ($\bar{X}_1-\bar{X}_2=d$) используют формулу для определения объема двух выборок (n_1 и n_2):

$$n_1=\frac{t^2}{\Delta^2}\cdot\left(\sigma_1^2+\frac{\sigma_2^2}{a}\right),$$

где t — критерий достоверности; $a=n_2:n_1$ — отношения объема двух выборок, если планируется, что $n_1\neq n_2$, σ_1^2 и σ_2^2 для каждой группы; Δ — допускаемое расхождение между выборочными средними и генеральной средней сравниваемых групп.

Если изменчивость признака в обеих группах близкая или равная ($\sigma_1^2\approx\sigma_2^2$), то формула преобразуется и принимает вид $n_1=\dfrac{t^2}{\Delta^2}\cdot\sigma^2\left(1+\dfrac{1}{a}\right)$, а поскольку $d^2=\dfrac{\Delta^2}{\sigma^2}$, следовательно $\Delta^2=d^2\cdot\sigma^2$.

Подставляя это выражение в первый член формулы, получают:

$$n_1=\frac{t^2}{d^2}\cdot\left(1+\frac{1}{a}\right).$$

где $a = n_2 : n_1$ или $n_1 : n_2$. Если объемы сравниваемых выборок равны ($n_1 = n_2$), то формула упрощается и имеет вид

$$n_1 = n_2 = \frac{t^2}{d^2}\left(1+\frac{1}{1}\right) = \frac{2t^2}{d^2}.$$

Пример. При оценке производителя по продуктивности дочерей требуется определить число дочерей и сверстниц, которое необходимо включить в обработку для получения достоверной разности между дочерями производителя и их сверстницами.

Предположим, что изменчивость удоя сверстниц и дочерей близка и составляет 60 кг ($\sigma^2_{св} \approx \sigma^2_{доч}$). Уровень t берем при $P = 0{,}95$, т. е. $t = 2{,}0$; $\Delta = \overline{X}_{св} - \overline{X}_{доч} = 30$ кг. Сверстниц в стаде в 2 раза больше, чем дочерей быка, поэтому $n_{св} : n_{доч} = 2 = a$, откуда $n_{св} = 2n_{доч}$. Величину d^2 вычисляют так:

$$d^2 = \frac{\Delta^2}{\sigma^2} = \frac{30^2}{60^2} = 0{,}25.$$

Тогда

$$n_{св} = \frac{t^2}{d^2}\cdot\left(1+\frac{1}{2}\right) = \frac{2^2}{0{,}25^2}\cdot\left(1+\frac{1}{2}\right) = \frac{4\cdot 1{,}5}{0{,}0625} = 96 \text{ голов}.$$

Следовательно, сверстниц надо взять в количестве около 100 голов, а дочерей быка можно взять в 2 раза меньше, что обеспечит достоверность разности между удоем коров сравниваемых групп ($d = \overline{X}_{св} - \overline{X}_{доч}$).

7.5. Оценка достоверности различий между фактическими и теоретическими ожидаемыми данными методом хи-квадрат (χ^2)

Критерий хи-квадрат («критерий соответствия», «критерий Пирсона») используется для проверки гипотез путем сравнения фактического распределения с теоретическим. Вычисление критерия соответствия также основано на принципах нулевой гипотезы, которая предполагает, что между сравниваемыми частотами сопоставляемых рядов нет достоверных различий. С помощью этого метода можно оценить, являются ли отклонения, наблюдаемые в опыте, случайными. Если отклонения оказываются неслучайными, эксперимент следует повторить, а также применить другие методы генетического исследования.

Метод χ^2 не может быть применен, если значения величин в опыте (количество объектов в классах) выражены в процентах или в относительных числах (долях), а также если в выборке число особей в каком-либо из теоретических классов меньше пяти (оптимальным считается число особей не менее 50).

Критерий χ^2 вычисляется по формулам

$$\chi^2 = \sum \frac{(\Phi - T)^2}{T}; \quad \chi^2 = \frac{[(\Phi - T) - 0{,}5]^2}{T},$$

где Ф — фактическое число особей, Т — теоретически ожидаемое число особей; член 0,5 — поправка Йетса (на нее уменьшается абсолютная величина значения Ф – Т). Если n и ожидаемые величины велики, то пользуются первой формулой (без поправки).

Вычисление критерия соответствия основано на принципах нулевой гипотезы, которая предполагает, что между частотами сопоставляемых вариационных рядов нет достоверной разности, 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, нулевая гипотеза отвергается и принимается альтернативная ей гипотеза.

Для доказательства достоверности разности между частотами двух рядов сопоставляют величину χ^2, вычисленную по конкретным данным ($\chi^2_{факт}$), со стандартными значениями хи-квадрат ($\chi^2_{теор}$), с учетом числа степеней свободы ν (см. приложение 2). Если вычисленное нами значение $\chi^2_{факт}$ больше стандартного ($\chi^2_{теор}$) при значимости $P = 0{,}01$ и $P = 0{,}001$, то считают, что гипотеза не согласуется с полученными в опыте данными. В таких случаях нулевая гипотеза должна быть отброшена. Если вычисленная величина $\chi^2_{факт} < \chi^2_{теор}$ при $P = 0{,}01$, но больше нее при значимости $P = 0{,}05$, то согласие наблюдаемых данных с ожидаемыми являются сомнительными. Однако это не дает права отбросить нулевую гипотезу. Если же $\chi^2_{факт} < \chi^2_{теор}$ при $P = 0{,}05$, то соответствие наблюдаемых данных с ожидаемыми считается установленным.

Величина χ^2 зависит от числа степеней свободы (ν). Поэтому для каждого значения вероятности (P) дано несколько значений χ^2, расположенных в приложении под определенным уровнем значимости.

Пример. Использование χ^2 при изучении наследования качественных признаков. В опыте по моногибридному скрещиванию изучали наследование нормальных и зачаточных крыльев у плодовой мушки *Drosophila melanogaster*. Анализировали F_2 от скрещивания родительских форм, различающихся между собой по форме крыльев:

Р: нормальные крылья × зачаточные крылья.

В F_1 от реципрокных вариантов скрещивания наблюдалось единообразие: все мухи были с нормальными крыльями. При анализе F_2 получены следующие результаты: 285 мух с нормальными крыльями и 91 — с зачаточными.

Найдем теоретически ожидаемые результаты. Предполагалось расщепление 3 : 1. Рассчитаем количество мух в каждом классе, которое бы точно соответствовало этому соотношению. Сложим общее количество объектов и разделим на 4, узнав, сколько объектов приходится соответственно на одну и на три части: 285 + 91 = 376. Ожидаемые результаты — 282 : 94. Все расчеты оформляются в таблицу (табл. 7.6).

Таблица 7.6

Вычисление χ^2 при изучении наследования качественных признаков

Класс	Данные опыта (Ф)	Теоретически ожидаемые (Т)	(Ф – Т)	$(Ф - Т)^2$	$\frac{(Ф-Т)^2}{Т}$	χ^2
Нормальные крылья	285	282	+3	9	0,03	0,12
Зачаточные крылья	91	94	–3	9	0,09	

В рассматриваемом нами примере число степеней свободы (ν) на единицу меньше числа классов $\nu = l - 1 = 2 - 1 = 1$, где l — число классов. Следовательно, для решения задачи нужно использовать из приложения уровни «вероятности» и строку «$\nu = 1$». В этой строке стоят три значения $\chi^2_{табл}$ {3,8—6,6—10,8}. Вычисленное значение χ^2 значительно меньше табличных. Следовательно, наблюдаемое в опыте расщепление соответствует ожидаемому, поэтому нулевая гипотеза, т. е. расщепление в соотношении 3:1, остается в силе (0,12<3,8; разность недостоверна).

Пример. Использование критерия χ^2 при определении достоверности различий между двумя группами животных. Предположим, требуется оценить результаты испытания нового препарата для предупреждения инфекционного заболевания кроликов. Из 50 кроликов 20 получали профилактический препарат (опытная группа), а 30 — не получали (контрольная группа). В опытной группе заболело семь особей, а 13 остались здоровыми. В контрольной группе заболело 14 кроликов, остались здоровыми 16. Доказывают ли результаты опыта профилактическое действие препарата, или различие в числе заболевших кроликов зависит не от введения препарата, а от случайных причин? Обработка материала приведена в табл. 7.7.

Таблица 7.7

Расчет критерия χ^2 при определении достоверности различия между двумя группами

Группа животных	Число заболевших		Число здоровых		Всего
	наблюдаемых (Ф)	теоретически ожидаемых (Т)	наблюдаемых (Ф)	теоретически ожидаемых (Т)	
Опытная	7	8,4 ($Т_1$)	13	11,6 ($Т_2$)	20
Контрольная	14	12,6 ($Т_3$)	16	17,4 ($Т_4$)	30
Итого	21	21	29	29	50

Теоретически ожидаемые частоты Т для заболевших и здоровых животных в опытной и контрольной группах составляют:

$$Т_1 = \frac{20 \cdot 21}{50} = 8{,}4;\ Т_2 = \frac{20 \cdot 29}{50} = 11{,}6 \text{ или } Т_2 = 20 - 8{,}4 = 11{,}6;$$

$$Т_3 = \frac{30 \cdot 14}{50} = 12{,}6;\ Т_4 = \frac{30 \cdot 29}{50} = 17{,}4 \text{ или } Т_4 = 30 - 12{,}6 = 17{,}4.$$

Подставив полученные данные в формулу $\chi^2 = \frac{[(Ф-Т)-0,5]^2}{Т}$, получим:

$$\chi^2 = \frac{[(7-8,4)-0,5]^2}{8,4} + \frac{[(13-11,6)-0,5]^2}{11,6} + \frac{[(14-12,6)-0,5]^2}{12,6} +$$

$$+\frac{[(16-17,4)-0,5]^2}{17,4} = \frac{0,9^2}{8,4} + \frac{0,9^2}{11,6} + \frac{0,9^2}{12,6} + \frac{0,9^2}{17,4} =$$

$$= 0,09 + 0,07 + 0,06 + 0,04 = 0,26.$$

При расчетах по четырехпольным таблицам число степеней свободы равно единице. Сравнивая полученное значение χ^2 со стандартным, находим, что вычисленная величина (0,26) меньше стандартныхзначений в строке таблицы, соответствующей одной степени свободы. Следовательно, оснований для того, чтобы отбросить нулевую гипотезу, нет, т. е. профилактическое действие препарата не может считаться доказанным.

Пример. Применение критерия хи-квадрат при сравнении частот двух эмпирических рядов между собой. Требуется сопоставить методом χ^2 частоты распределения коров-матерей черно-пестрой породы и коров-дочерей помесного происхождения от отцов голштинской породы по форме вымени. Вариационные ряды по матерям и дочерям подразделяют на четыре класса в зависимости от формы вымени (чашевидное, ваннообразное, округлое и козье). Обследовано 200 коров-матерей и 210 их дочерей — помесей первого поколения от голштинских быков. Результаты обследования коров обеих групп приведены в табл. 7.8. Нулевой гипотезой для этого примера служит предположение, что скрещивание не приводит к улучшению формы вымени.

В этом случае следует использовать формулу

$$\chi^2_n = \frac{1}{n_1 \cdot n_2} \cdot \sum \left[\frac{(P_1 \cdot n_2 - P_2 \cdot n_1)^2}{P_1 + P_2} \right].$$

Здесь по каждому классу частоты первого ряда умножают на общее число наблюдений второго ряда ($P_1 \cdot n_2$), а частоты второго ряда — на общее число наблюдений первого ряда ($P_2 \cdot n_1$). Полученную разность возводят в квадрат и делят на сумму частот этого класса по обоим рядам $(P_1 \cdot n_2 - P_2 \cdot n_1)^2 : (P_1 + P_2)$. Сумма дробей по всем классам ряда умножают на коэффициент $1/n_1 + n_2$, получая в результате искомую величину χ^2.

Подставляя числовые значения в формулу получаем эмпирическую величину χ^2:

$$\chi^2_n = \frac{1}{n_1 \cdot n_2} \cdot \sum \left[\frac{(P_1 \cdot n_2 - P_2 \cdot n_1)^2}{P_1 + P_2} \right] =$$

$$= 136\,048 : 42\,000 = 3,239 \approx 3,24.$$

Число степеней свободы $\nu = l - 1 = 4 - 1 = 3$. Табличное $\chi^2_{0,05} = 7,81$. Вычисленное χ^2 по конкретным данным примера равно 3,2, т. е. меньше теоретического ($\chi^2 = 7,81$), следовательно, скрещивание с голштинскими быками несущественно повлияло на форму вымени коров-дочерей, т. е. сохраняется нулевая гипотеза (H_0).

Таблица 7.8

Сравнение распределения коров по форме вымени методом χ^2

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			
Чашевидное	30	50	$30 \cdot 210 = 6300$	$50 \cdot 200 = 10\,000$	$\frac{(6300 - 10\,000)^2}{410} = 33\,390$
Ваннообразное	50	70	$50 \cdot 210 = 10\,500$	$70 \cdot 200 = 14\,000$	$\frac{(10\,500 - 14\,000)^2}{410} = 29\,878$
Округлое	100	80	$100 \cdot 210 = 21\,000$	$80 \cdot 200 = 16\,000$	$\frac{(21\,000 - 16\,000)^2}{410} = 60\,975$
Козье	20	10	$20 \cdot 210 = 4200$	$10 \cdot 200 = 2000$	$\frac{(4200 - 2000)^2}{410} = 11\,805$
Всего	$\sum P_1 = 200$,	$\sum P_2 = 210$;	число классов $l = 4$.		

Задания для самостоятельной работы

Задание 1. В результате спаривания кур, имеющих листовидный гребень, с гетерозиготным петухом, гребень у которого розовидный, получено 106 цыплят с розовидным и 120 — с листовидным гребнем. При нулевой гипотезе, согласно которой данная пара признаков зависит от одной пары генов, в потомстве ожидается расщепление в соотношении 1:1. Вычислить критерий хи-квадрат и оценить соответствие между наблюдаемым и ожидаемым расщеплением.

Задание 2. Между собой спарены помеси первого поколения от черных гемпширских свиней и красного дюрок-джерсейского хряка. Среди помесного потомства был 81 поросенок черной масти и 26 — красной. При нулевой гипотезе (масть обусловлена одной парой генов) ожидается расщепление по масти в соотношении три особи черных и одна красная. Вычислить критерий хи-квадрат и оценить соответствие между наблюдаемыми и ожидаемыми данными.

Задание 3. При изучении защитного действия индол-3-пропиогидросамовой кислоты и экспериментальном заражении кроликов болезнью Ауески из 20 особей выжило восемь, пало 12, а при изучении терапевтического эффекта из 17 кроликов выжило шесть, пало 11. Проверьте гипотезу об эффективности терапевтического и защитного действия этого препарата.

Задание 4. Имеются мыши двух линий: высоколейкозной (AKR) и низколейкозной (CC57Br). В обычных условиях из 44 мышей в линии AKR выжило 22, пало 22; из 35 мышей линии CC57Br выжило 29, пало шесть. Оцените методом хи-квадрат, достоверно ли различие между линиями.

Задание 5. При испытании нового антибиотика на кроликах, больных пневмонией, получены следующие результаты: из больных, принимавших антибиотик, выжило 65, пало 25 животных; из не получавших антибиотик выжило 35 пало 25. Оцените, достоверно ли лечебное действие препарата.

Задание 6. Определить с помощью критерия хи-квадрат соответствие между наблюдаемым и ожидаемым распределением коров красной степной породы по типам лактоглобулинов (Lg) молока. Наблюдаемое распределение: LgA был у 26 особей, LgB — у 271 особи и LgAB — у 202 особей. Ожидаемое распределение: LgA — 0,065; LgB — 0,555; LgAB — 0,380.

Задание 7. Определить соответствие между наблюдаемым и ожидаемым распределением буйволов по типам трансферринов (Tf) сыворотки крови. Наблюдаемое распределение: TfA найдено в сыворотке крови 21 животного, TfB — в сыворотке крови 135, а TfAB — в сывортке 105 животных. Ожидаемое распределение: TfA — 0,079; TfB — 0,517; TfAB — 0,404.

Задание 8. Изучено наличие антигенов системы группы крови C у зебувидного скота и у его гибридов со швицами:

1) антиген C_1 обнаружен у 11 из 23 зебувидных коров, а у гибридов — 43 из 68. Оценить с помощью критерия хи-квадрат, достоверно ли различие по сравниваемому антигену между этими популяциями;

2) оценить на основании следующих данных различие по антигену C_2: у зебувидного скота антиген C_2 обнаружен у 12 из 33 коров, а у гибридов — у 43 из 48.

Задание 9. Вычислить критерий соответствия (хи-квадрат), если во втором поколении моногибридного скрещивания, состоящем из 8024 особей, получено 6023 особи с доминантным признаком и 2001 особь с рецессивным. Согласно теории, ожидается расщепление 3 : 1.

Задание 10. Для повышения оплодотворяемости коров, в сперму быка рекомендуют добавлять биостимулятор. Из 441 коровы, осемененной обработанной биостимулятором спермой, оплодотворилось 301. В контроле (сперма без биостимулятора) из 352 оплодотворилось 216 животных. Оцените методом хи-квадрат, достоверно ли действие биостимулятора.

Задание 11. У кур розовидная форма гребня доминирует над простой. При скрещивании кур с розовидной формой гребня между собой в потомстве получилось 705 — с розовидной формой гребня и 224 — с простой. Соответствует ли расщепление по фенотипу, полученное у потомков, теоретически ожидаемому?

Контрольные вопросы

1. Что такое ошибки репрезентативности? Чем они отличаются от ошибок вычисления и измерения?
2. Как вычисляют ошибку средней арифметической величины? Что она показывает?
3. Приведите формулы вычисления ошибок σ, cv, r, R, d.
4. Как изменяется величина m_x при изменении объема выборки и величины сигмы?
5. Что такое доверительные вероятности? Какие доверительные вероятности используют в биологических, зоотехнических и ветеринарных исследованиях?
6. Как определяется достоверность выборочных показателей?
7. Как оценивается достоверность разности между средними величинами двух выборок?
8. Как определяют доверительные границы при работе с большими и малыми выборками?
9. Как определяют необходимый объем выборки?
10. Что такое критерий соответствия (хи-квадрат)?Как он используется в генетических исследованиях?
11. От чего зависит величина ошибки средней арифметической?
12. Какое существует правило записи статистической величины и ее ошибки?
13. Что показывает t_d?
14. Как определяют уровень достоверности?

Глава 8
ДИСПЕРСИОННЫЙ АНАЛИЗ

Биологические особенности животных зависят от многих внешних и внутренних факторов. Так, масса телят при рождении и в последующие периоды определяется наследственными особенностями их матерей и отцов, кормлением, содержанием, а также условиями эмбрионального развития. Этим обусловливается большое разнообразие животных по их физиологическим и морфологическим свойствам. Поэтому необходимо было разработать соответствующие математические методы, с помощью которых можно было бы установить влияние отдельных факторов и оценить их относительную роль в общей изменчивости этих свойств. Это, прежде всего, метод дисперсионного анализа, разработанный английским математиком и биологом Р. Фишером.

8.1. Сущность дисперсионного анализа

Дисперсионный анализ является самостоятельным и важным разделом биологической статистики. С его помощью можно установить роль отдельных факторов в изменчивости того или иного признака, разложить общую изменчивость признака на составные части, обусловленные изучаемыми конкретными факторами (возраст родителей, крупность родителей, порода, уровень кормления и т. д.), а также вызываемый случайными, неконтролируемыми факторами (температура, свет и т. д.).

Таким образом, сущность дисперсионного анализа состоит в изучении статистического влияния одного или несколько факторов на результативный признак.

Результативный признак (например, величина удоя, плодовитость, крупноплодность поросят и т. д.), который изучается как результат влияния двух различных факторов:

1) организованных в исследованиях — разные породы, уровень кормления животных и т. д. ($\overline{X}$);

2) неорганизованных, т. е. случайных, не учитываемых в исследованиях ($\overline{Z}$).

В зоотехнии понятие «дисперсия» означает варьирование или изменчивость признака, возникающие под влиянием различных факторов, и обозначается через C. Дисперсия представляет сумму квадратов отклонений вариант от средней арифметической — $C = \sum(X - \bar{X})^2$. При проведении дисперсионного анализа в той или иной группе животных используется общая дисперсия, под которой понимается изменчивость, вызываемая всеми одновременно действующими факторами.

Общая дисперсия признака (C_y) в изучаемой группе животных может быть разложена: на дисперсию, возникающую под влиянием различных учтенных, организованных факторов, называемую факториальной (C_x), и дисперсию, возникающую под влиянием различных случайных (неучтенных), неорганизованных факторов, называемую остаточной (C_z). При этом сумма факториальной и сумма остаточной дисперсий всегда равна величине общей дисперсии: $C_y = C_x + C_z$.

При дисперсионном анализе вычисляют величину общей (C_y), факториальной (C_x) и остаточной (C_z) дисперсии. Если изменчивость возникает под влиянием нескольких факторов (возраста, живой массы, продолжительность плодоношения, кормления и т. д.) и требуется определить долю влияния каждого из этих учтенных факторов, то в таком случае факториальная дисперсия (C_x) может быть разделена на дисперсии каждого фактора раздельно (C_A, C_B, C_C) и совместную дисперсию.

При дисперсионном анализе обработке подвергаются выборочные данные, оформленные в статистические комплексы. Выборка может быть мало- и многочисленной. Статистические комплексы, в зависимости от того, сколько факторов включено в каждый, могут быть одно-, двух-, трехфакторными и с бóльшим числом факторов.

Следует отметить, что трехфакторный и более дисперсионный анализ используют в практике довольно редко вследствие ненадежности получаемых оценок силы влияния. Кроме того, существуют некоторые обязательные условия организации многофакторных комплексов. Одно из них — полная независимость изучаемых факторов друг от друга — обычно трудновыполнимо.

Для получения достоверных данных дисперсионного анализа при построении дисперсионных комплексов (особенно двух- и многофакторных) необходимо правильно выполнить следующие операции.

1. *Подбор факторов*. Фактор — это любое проявление признака, влияние которого требуется изучить на результативный признак (живая масса, плодовитость, крупноплодность и т. д.). При организации двух- и многофакторных комплексов свободный выбор факторов для исследования ограничен требованием полной независимости их между собой, т. е. они должны быть не коррелирующими. Нельзя брать, например, массу и размер животного (обхват груди и масса животного), крупноплодность и размер помета. Независимыми между собой могут быть следующие факторы: пол животных и уровень кормления, порода животных и переваримость питательных веществ, температура и влажность помещения и т. п.

2. *Разделение факторов на градации*. Фактором может быть любое воздействие на изучаемые объекты, в частности на животных. Обычно

приходится изучать разные степени действия фактора — разные его дозы, называемые градациями фактора. Например, при изучении наследственных влияний градациями фактора будут отдельные производители (или отдельные матки).

Для однофакторных, а также двух- и многофакторных комплексов они могут иметь количественные (количество стимуляторов в граммах, уровень кормления +20 % или –20 % к норме и т. д.) и качественные градации (пол — мужской и женский; упитанность у свиней — жирная, мясная, беконная; масть — красно-пестрая, светло-палевая, темно-палевая и т. д.).

Группа животных, которая подвергается воздействию определенной градации фактора, образует градацию дисперсионного комплекса. Таким образом, число градаций комплекса соответствует числу градаций фактора.

3. *Подбор особей*. Результаты дисперсионного анализа во многом зависят от правильного подбора особей как по количеству, так и по качеству. Подобранные по качеству особи должны отражать генеральную совокупность, для изучения которой проводится исследование.

Предположим, что исследуется влияние использовавшихся в стаде быков-производителей на удой их дочерей. Может случиться так, что дочери одного быка оценены по удою только за превую лактацию, дочери другого быка — по удою за вторую и третью лактации, а дочери третьего — за третью, четвертую и пятую лактации. Если не обратить внимание на тот факт, что первотелки в среднем дают за лактацию молока меньше, чем коровы более старших возрастов (удои повышаются до пяти — восьми отелов), то в приведенном примере даже при абсолютной одинаковости быков-производителей по их наследственным качествам, обусловливающим молочную продуктивность, можно прийти к неправильным выводам, отдав предпочтение третьему быку-производителю (дочери которого будут иметь более высокий удой как старшие по возрасту).

Очевидно, что для выяснения влияния производителей на удои дочерей необходимо учитывать возраст телок. В данном примере сильное влияние на результат могут оказать и коровы-матери (дочери одного быка могут быть получены только от лучших матерей, а другого — от плохих или средних), а также условия кормления и содержания животных (если дочери сравниваемых производителей выращивались и продуцировали в разных условиях). Чтобы избежать наложения действия других факторов на действие изучаемых градаций фактора (производителей), следует дочерей разных быков от матерей примерно одинакового качества вырастить в одинаковых условиях.

Градации комплекса составляют таким образом, чтобы обеспечить случайность действия всех остальных факторов (кроме изучаемого), их неорганизованность, равновероятную направленность (рандомность) действия, что создаст фон, на котором можно выявить закономерность действия, организованного в градации фактора.

Градации комплекса — это выборки, сделанные из заведомо разных генеральных совокупностей, поэтому наиболее приемлемым принципом формирования градаций комплекса является принцип случайного отбора в них отдельных объектов из совокупностей. Нарушение этого принципа всегда приводит к неправильным выводам.

В тех случаях, когда не удается избежать систематического влияния на объект неизучаемого фактора (не удается рендомизировать его влияние), прибегают к двухфакторному дисперсионному анализу. Хотя второй фактор сам по себе может не интересовать исследователя, его влияние учитывают, чтобы выявить степень влияния изучаемого фактора. Влияние фактора оценивают по изменениям у животных отдельного признака, который называют результативным признаком.

4. *Преобразование значений результативного признака.* Для облегчения счетной работы можно неудобные значения результативного признака (многозначные, дробные) преобразовать в удобные (малозначные и выражаемые целыми числами).

а) все значения признака можно умножить на одно и то же число: $0{,}30 \cdot 100 = 30$ или $0{,}45 \cdot 100 = 45$; в конечный результат необходимо внести поправки (разделить на 100).

б) все значения признака по градациям можно разделить на одно и то же число (на 2, 3 или 4 и т. д.) и внести в конце соответствующие поправки, т. е. умножить на то число, на которое было произведено деление;

в) от всех значений результативного признака можно отнять одно и тоже число (лучше вычитать наименьшее значение признака). Поправку в конечном результате при этом нужно вносить лишь для средней арифметической $\bar{X}$, прибавляя к ней ранее вычтенное число.

5. *Техника расчетов при дисперсионном анализе.* Расчет дисперсионных комплексов проводится по официальным рабочим формулам. Однако техника расчетов для малых и больших групп различна. Для малых групп (выборок) дисперсии рассчитываются проще, для больших — несколько сложнее. Группа в 40—50 вариантов (многозначных показателей) уже считается большой.

8.2. Вычисление дисперсии однофакторного комплекса при малочисленной выборке

Пример 1. Определить влияние возраста матерей на живую массу телят при рождении. Порядок вычисления и необходимые данные приведены в табл. 8.1.

$$C_y = \sum v^2 - H,$$

где $H = \dfrac{(\sum v)^2}{n}$.

Остаточную дисперсию (C_z) вычисляют по формуле

$$C_z = \sum v^2 - \sum h_x,$$

где $\sum h_x = \frac{(\sum v)^2}{n}$.

Факториальную дисперсию (C_x) вычисляют по формуле

$$C_x = \sum h_x - H;$$

$$H = \frac{(\sum v)^2}{n} = \frac{609^2}{17} = 21\,817.$$

$$C_y = \sum v^2 - H = 22\,067 - 21\,817 = 250;$$

$$C_x = \sum h_x - H = 21\,921 - 21\,817 = 104;$$

$$C_z = \sum v^2 - \sum h_x = 22\,067 - 21\,921 = 146.$$

Проверка правильности подсчетов производится суммированием: $C_y = C_x + C_z$, т. е. 104 + 146 = 250. В данном случае подсчеты сделаны правильно.

Степень (доля) влияния разных факторов на варьирующий признак определяется отношением между дисперсиями C_x и C_y, C_z и C_y; обозначают эти отношения через η^2. Так, доля влияния учтенных факторов равняется $\eta_x^2 = \frac{C_x}{C_y}$, а доля неучтенных факторов $\eta_z^2 = \frac{C_z}{C_y}$.

Число степеней свободы $v = l - 1 = 4 - 1 = 3$. Табличное $\chi^2_{0,05} = 7,81$. Вычисленное χ^2 по конкретным данным примера равно 3,2, т. е. меньше теоретического ($\chi^2 = 7,81$), следовательно, скрещивание с голштинскими быками несущественно повлияло на форму вымени коров-дочерей, т. е. сохраняется нулевая гипотеза (H_0).

Таблица 8.1

Техника расчета при дисперсионном анализе однофакторного комплекса для малых выборок

Показатели (v)	Градации фактора (R)			r = 3 $\sum$
	полновозрастные матери	возраст матерей 31—36 мес.	возраст матерей 25—30 мес.	
Живая масса при рождении	35, 36, 40, 38, 43, 42	38, 32, 40, 34, 35, 31	35, 37, 30, 31, 32	$\sum v = 609$
v^2	1225, 1296, 1600, 1444, 1849, 1764	1444, 1024, 1600, 1156, 1225, 961	1225, 1369, 900, 961, 1024	$\sum v = 22\,067$
n	6	6	5	$\sum n = 17$
$\sum v$	234	210	165	$\sum v = 609$
$(\sum v)^2$	$234^2 = 54\,756$	$210^2 = 44\,100$	$165^2 = 27\,225$	—

Показатели (v)	Градации фактора (R)			r = 3 Σ
	полновозрастные матери	**возраст матерей 31—36 мес.**	**возраст матерей 25—30 мес.**	
$h_x = \frac{(\sum v)^2}{n}$	$\frac{54756}{6} = 9120$	$\frac{44100}{6} = 7350$	$\frac{27225}{5} = 5445$	$\sum h_x = 21921$
$M = \frac{\sum v}{n}$	$\frac{234}{6} = 39$	$\frac{210}{6} = 35$	$\frac{165}{5} = 33$	$M_{общ} = \frac{609}{17} = 35{,}8$

Достоверность факториальной дисперсии, т. е. достоверно ли влияние и доля влияния фактора на изменчивость признака, определяется коэффициентом достоверности Фишера (F). Для вычисления коэффициента Фишера необходимо определить число степеней свободы (ν) и корректированную дисперсию — варианcу (σ^2).

Число степеней свободы для факториальной дисперсии (C_x) равно числу классов (l) или градаций (r) по фактору минус единица: $\nu_x = l_x - 1$; $\nu_x = 3 - 1 = 2$.

Для остаточной дисперсии (C_z) число степеней свободы равно численности выборки (n) минус число классов (l) или градаций (R): $\nu_z = n - l_x$; $\nu_z = 17 - 3 = 14$.

Число степеней свободы для общей дисперсии (C_y) равно численности выборки (n) без единицы: $\nu_y = n - 1$; $\nu_y = 17 - 1 = 16$.

Корректированную дисперсию, или варианcу (σ^2) (факториальную и остаточную) вычисляют делением дисперсии на соответстнующее число степеней свободы.

Факториальная варианса:

$$\sigma_x^2 = \frac{C_x}{\nu_x};\ \sigma_x^2 = \frac{104}{2} = 57.$$

Остаточная варианса:

$$\sigma_z^2 = \frac{C_z}{\nu_z};\ \sigma_z^2 = \frac{146}{14} = 10{,}4.$$

Коэффициент достоверности Фишера вычисляется делением факториальной корректированной дисперсии (вариансы) на остаточную корректированную дисперсию:

$$F = \frac{\sigma_x^2}{\sigma_z^2};\ F = \frac{57}{10{,}4} = 5{,}5.$$

Вычисленное (эмпирическое) значение F сравнивают с табличным (стандартным) значением F. Табличное значение F для данного примера при трех уровнях вероятности равно: $F_{0,95} = 3{,}7$; $F_{0,99} = 6{,}5$; $F_{0,999} = 11{,}8$.

В данном примере вычисленное F равно 5,5, следовательно, влияние возраста матерей на живую массу телят при рождении достоверно при уровне вероятности $P = 0{,}95$.

Пример 2. Установить влияние фактора породности на переваримость питательных веществ (результативный признак), если по принципу случайной выборки было отобрано для каждой градации фактора по три особи (подсвинки), имеющие следующие показатели коэффициента переваримости протеина: крупная белая порода — 78,33; 76,70; 76,37 %; крупная белая × ландрас — 81,20; 82,16; 83,50 % (табл. 8.3, 8.4).

Для вычисления общей, факториальной и случайной дисперсий необходимо определить:

$$H = \frac{(\sum v)^2}{n} = \frac{(478{,}26)^2}{6} = \frac{228\,732{,}62}{6} = 38\,122{,}10;$$

$$C_y = \sum v^2 - H = 38\,166{,}80 - 38\,122{,}10 = 44{,}70;$$

$$C_z = \sum v^2 - \sum h_x = 38\,166{,}80 - 38\,161{,}93 = 4{,}87;$$

$$C_x = \sum h_x - H = 38\,161{,}93 - 38\,122{,}1 = 39{,}83;$$

$$C_x + C_z = C_y;\ C_y = 39{,}83 + 4{,}87 = 44{,}70.$$

Таблица 8.3

Однофакторный дисперсионный комплекс при малой выборке

Показатели	Порода		$r = 2$ $\sum$
	A_1 (крупная белая)	A_2 (крупная белая × ландрас)	
v — варианты (значение результативного признака)	78,33 76,70 76,37	81,20 82,16 83,50	—
v^2 — квадраты вариант градаций	6135,59 5882,89 5832,37	6593,44 6750,26 6972,25	$(\sum v)^2 = 38166{,}80$
n — объемы градаций	3	3	$\sum n = 6$
$\sum v$ — сумма вариант каждой градации	231,40	246,86	$\sum v = 478{,}26$
$(\sum v)^2$ — квадрат суммы вариант градации	53 545,96	60 939,85	—
$h = \frac{(\sum v)^2}{n}$ — частная подсобная величина	17 848,65	20 313,28	38 161,93
$M = \frac{\sum v}{n}$ — средняя арифметическая вариант градации	77,13	82,29	$M_{общ} = 79{,}71$

Таблица 8.4

Сводка результатов дисперсионного анализа однофакторного комплекса для малых выборок

$h=\frac{(\sum v)^2}{n}=\frac{(478{,}26)^2}{6}=38122{,}10$	Показатель	Дисперсия		
		x	z	y
$C_y=\sum v^2-H=38166{,}80-38122{,}10=44{,}70$	C	39,83	4,87	44,70
$C_z=\sum v^2-\sum h_x=38166{,}80-38161{,}93=4{,}87$	η^2 η^2	$\frac{C_x}{C_y}$ 0,8911	$\frac{C_z}{C_y}$ 0,1089	$\frac{C_y}{C_y}$ 1,00
$C_x=\sum h-H=38161{,}93-38122{,}1=39{,}83$	%	89,11	10,89	100,0
$C_x+C_z=C_y;\ C_y=39{,}83+4{,}87=44{,}70$				
Число степеней свободы	ν	$\nu_x=l-1=$ $=2-1=1$	$\nu_z=n-l=$ $=6-2=4$	$\nu_y=n-1=$ $=6-1=5$
Корректированная дисперсия	σ^2	$\sigma_x^2=\frac{39{,}83}{1}=39{,}83;\ \sigma_z^2=\frac{4{,}87}{4}=1{,}22$		
Коэффициент достоверности	$F=\frac{39{,}83}{1{,}22}=32{,}64$			
Табличное значение	$F=\begin{cases}74{,}1-P_3=0{,}999\\21{,}2-P_2=0{,}99\\7{,}7-P_1=0{,}95\end{cases}$			

Следовательно, в данном примере установлено достоверное влияние фактора породности на результативный признак при ($P \geq 0{,}99$).

8.3. Вычисление дисперсии однофакторного комплекса при многочисленной выборке

Для вычисления общей дисперсии при многочисленной выборке пользуются следующей формулой:

$$C_y = \sum p_v a_v^2 - H,$$

где

$$H = \frac{(\sum p_v a_v)^2}{n}.$$

Формула факториальной дисперсии:

$$C_x = \sum h_x - H.$$

Формула остаточной дисперсии:

$$C_z = \sum p_v a_v^2 - \sum h_x.$$

Пример. Установить влияние живой массы матерей на живую массу телят при рождении. Порядок вычисления и необходимые данные приведены в табл. 8.5. Эта таблица составляется так же, как и корреляционная решетка. В крайней левой графе записывают классы варьирующего признака, а в верхней заглавной строке — классы изучаемого фактора. В клетки решетки обычным методом разносят варианты.

Чтобы вычислить дисперсии (факториальную и остаточную), необходимо к решетке добавить четыре графы по вертикали и четыре — по горизонтали, которые строятся внизу решетки для изучаемого фактора, а с правой стороны — для классов варьирующего признака. В заголовке вертикальных граф пишут последовательно: p_v, a_v, $p_v \cdot a_v$, $p_v \cdot a_v^2$, а в заголовке горизонтальных граф: p_A, $p_A \cdot a_v$, $(p_A \cdot a_v)^2$, $h_x = \frac{\sum (p_A \cdot a_v)^2}{p_A}$.

Далее производятся вычисление и заполнение добавочных к решетке граф. Ход вычисления будет понятен при рассмотрении табл. 8.5.

Вычисленные выражения $\sum p_A \cdot a_v$ приводятся для каждого столбца по классам A. Для получения указанных выражений требуется умножить частоты по клеткам каждого столбца на отклонение a_v; сумму этих произведений записывают по каждому классу в строку под строчкой p_A.

Таблица 8.5

Обработка однофакторного комплекса при многочисленной выборке

v (живая масса телят, кг)	Фактор A, живая масса матерей, кг				p_v	a_v	$p_v a_v$	$p_v \cdot a_v^2$
	400—449	450—499	500—549	550—599				
30—32,9	3	1	—	—	4	–3	–12	36
33—35,9	14	7	2	1	24	–2	–48	96
36—38,9	15	15	8	2	40	–1	–40	40
39—41,9	8	10	14	14	46	0	0	0
42—44,9	2	5	10	16	33	1	33	33
45—47,9	—	2	2	7	11	2	22	44
48—50,9	—	—	—	2	2	3	6	18
p_A	42	40	36	42	160		$\sum p_v \cdot a_v = -39$	$\sum p_v \cdot a_v^2 = 267$
$p_A \cdot a_v$	–50	–23	2	32				
$\sum (p_A \cdot a_v)^2$	$-50^2 = 2500$	$-23^2 = 529$	$2^2 = 4$	$32^2 = 1024$				
$h_x = \frac{\sum (p_A \cdot a_v)^2}{p_A}$	$\frac{2500}{42} = 59{,}5$	$\frac{529}{40} = 13{,}2$	$\frac{4}{36} = 0{,}1$	$\frac{1024}{42} = 24{,}4$	$\sum h_x = 97{,}2$			

1-й столбец:

$$\sum p_A \cdot a_v = 3\cdot(-3)+14\cdot(-2)+15\cdot(-1)+8\cdot 0+2\cdot 1=-50;$$

2-й столбец:

$$\sum p_A \cdot a_v = 1\cdot(-3)+7\cdot(-2)+15\cdot(-1)+10\cdot 0+5\cdot 1+2\cdot 2=-23;$$

3-й столбец:

$$\sum p_A \cdot a_v = 2\cdot(-2)+8\cdot(-1)+14\cdot 0+10\cdot 1+2\cdot 2=2;$$

4-й столбец:

$$\sum p_A \cdot a_v = 1\cdot(-2)+2\cdot(-1)+14\cdot 0+16\cdot 1+7\cdot 2+2\cdot 3=32.$$

Промежуточная величина H:

$$H=\frac{\sum(p_v \cdot a_v)^2}{n}=\frac{(-39)^2}{160}=9{,}5.$$

Дисперсии C_y, C_x, C_z вычисляют по приведенным выше формулам, подставляя в них данные из таблицы:

$$C_y=\sum p_v a_v^2 - H = 267-9{,}5=257{,}5;$$

$$C_x=\sum h_x - H = 97{,}2-9{,}5=87{,}7;$$

$$C_z=\sum p_v a_v^2 - \sum h_x = 267-97{,}2=169{,}8.$$

Доля учтенных факторов равняется: $\eta_x^2=\frac{C_x}{C_y}=\frac{87{,}7}{257{,}7}=0{,}341$, или 34,1 %.

Доля неучтенных факторов: $\eta_z^2=\frac{C_z}{C_y}=\frac{169{,}8}{257{,}5}=0{,}659$, или 65,9 %.

Число степеней свободы для факториальной дисперсии равно: $\nu_x=l_x-1=4-1=3$.

Для остаточной дисперсии число степеней свободы: $\nu_z = n - l_x = 160 - 4 = 156$.

Факториальная варианса: $\sigma_x^2=\frac{C_x}{\nu_x}=\frac{87{,}7}{3}=29{,}2.$

Остаточная варианса: $\sigma_z^2=\frac{C_z}{\nu_z}=\frac{169{,}8}{156}=1{,}1.$

Коэффициент достоверности Фишера:

$$F=\frac{\sigma_x^2}{\sigma_z^2}=\frac{29{,}2}{1{,}1}=26{,}5.$$

Табличное значение F для данного примера при трех уровнях вероятности равно (см. приложение 3): $F_{0,95}=2{,}7$; $F_{0,99}=3{,}9$; $F_{0,999}=5{,}7$.

Влияние живой массы матерей на живую массу телят при рождении вполне достоверно при всех градациях вероятности, так как вычисленное $F=26{,}5$.

8.4. Вычисление дисперсии двухфакторного комплекса при малочисленной выборке

Структура двухфакторного комплекса более сложная, чем у однофакторных. Для облегчения счетной работы в двухфакторных комплексах необходимо, чтобы изучаемые факторы были независимыми друг от друга, а частоты были пропорциональны (градациям). Изучая действие более одного фактора необходимо учитывать влияние на результативный признак не только фактора A и фактора B, но и их совместное влияние AB на варьирующий признак. Поэтому дисперсионный анализ двухфакторного комплекса должен определить не только C_y, C_x, C_z, C_A, C_B, но и дисперсию C_{AB}. Показатели C_y, C_x, C_z вычисляются так же, как и в однофакторном комплексе.

Для вычисления факториальной дисперсии C_A и C_B обработка данных ведется несколько иначе. Для получения $\sum h_A$ и $\sum h_B$ обрабатываются решетки, разделенные для фактора A и B, составляя подсобные таблицы. В равномерных и пропорциональных комплексах дисперсии вычисляются следующими формулами.

Общая дисперсия:

$$C_y = \sum v^2 - H,$$

где $H = \frac{\left(\sum v\right)^2}{n}$.

Общефакториальная дисперсия:

$$C_x = \sum h_x - H.$$

Дисперсия от фактора A:

$$C_A = \sum h_A - H.$$

Дисперсия от фактора B:

$$C_B = \sum h_B - H.$$

Дисперсия от фактора AB:

$$C_{AB} = C_x - C_A - C_B.$$

Остаточная дисперсия:

$$C_z = \sum v^2 - \sum h_x.$$

Пример. Выяснить влияние на яйценоскость кур кормов животного происхождения и продолжительность светового дня. Необходимые данные для вычисления приведены в табл. 8.6, а для вычисления факториальных дисперсий — в табл. 8.7.

Таблица 8.6

Обработка двухфакторного равномерного комплекса при малочисленной выборке

Показатель	Рацион без кормов животного происхождения		Рацион с кормами животного происхождения		$\sum$
	световой день 8—10 ч (B_1)	световой день 12—14 ч (B_2)	световой день 8—10 ч (B_1)	световой день 12—14 ч (B_2)	
v (яйценоскость)	157, 162, 166 158, 167, 164	170, 178, 185 180, 173, 175	188, 193, 192 184, 199, 190	190, 211, 198 204, 220, 207	$\sum v = 4411$
n_x	6	6	6	6	$\sum n_x = n = 24$
v^2	24 649, 26 244 27 556, 24 964 27 889, 26 896	28 900, 31 684 34 225, 32 400 29 929, 30 625	35 344, 37 249 36 884, 33 856 39 601, 36 100	36 100, 44 521 39 204, 41 616 48 400, 42 849	$\sum v^2 = 817\,685$
$\sum v_x$	974	1061	1146	1230	$\sum v_x = 4411$
$(\sum v_x)^2$	948 676	1 125 721	1 312 316	1 512 900	—
$h = \frac{(\sum v_x)^2}{n_x}$	$\frac{948\,676}{6} = 158\,113$	$\frac{1\,125\,721}{6} = 187\,620$	$\frac{1\,312\,316}{6} = 218\,719$	$\frac{1\,512\,900}{6} = 252\,150$	$\sum h_x = 816\,602$

Дисперсии выполняют по приведенным выше формулам, подставляя в них данные из табл. 8.6 и 8.7.

Таблица 8.7

Обработка комплекса по фактору А и В

Классы по факторам	n фактора	$\sum v_x$	$(\sum v_x)^2$	$h = \frac{(\sum hx)^2}{n}$
A_1	12	2035	4 141 225	$\frac{4\,141\,225}{12} = 345\,102$
A_2	12	2376	5 645 376	$\frac{5\,645\,376}{12} = 470\,448$
По фактору *А*	24	4411	—	$\sum h_A = 815\,550$
B_1	12	2120	4 494 400	$\frac{4\,494\,400}{12} = 374\,533$
B_2	12	2291	5 248 681	$\frac{5\,248\,681}{12} = 437\,390$
По фактору *В*	24	4411	—	$\sum h_B = 811\,923$

Величина

$$H = \frac{(\sum v)^2}{n} = \frac{(4411)^2}{24} = \frac{19\,456\,921}{24} = 810\,705;$$

$$C_y = \sum v^2 - H = 817\,685 - 810\,705 = 6980;$$

$$C_x = \sum h_x - H = 816\,602 - 810\,705 = 5897;$$

$$C_A = \sum h_A - H = 815\,550 - 810\,705 = 4845;$$

$$C_B = \sum h_B - H = 811\,923 - 810\,705 = 1218;$$

$$C_{AB} = C_x - C_A - C_B = 5897 - 4845 - 1218 = 166;$$

$$C_z = \sum v^2 - \sum h_x = 817\,685 - 816\,602 = 1083.$$

Доля влияния факторов *А* и *В* и совместного действия *АВ* на изменчивость признака в данном примере равнялось:

$$\eta_A^2 = \frac{C_A}{C_y} = \frac{4845}{6980} = 0{,}69, \text{ или } 69{,}5\ \%;$$

$$\eta_B^2 = \frac{C_B}{C_y} = \frac{1218}{6980} = 0{,}175, \text{ или } 17{,}5\ \%;$$

$$\eta_{AB}^2 = \frac{C_{AB}}{C_y} = \frac{166}{6980} = 0{,}018, \text{ или } 1{,}8\ \%.$$

Число степеней свободы: $v_x = l_A \cdot l_B - 1 = 2 \cdot 2 - 1 = 3$; $v_A = l_A - 1 = 2 - 1 = 1$; $v_B = l_B - 1 = 2 - 1 = 1$; $v_{AB} = v_A \cdot v_B = 1 \cdot 1 = 1$; $v_z = n - l_A \cdot l_B = 24 - 2 \cdot 2 = 20$; $v_y = n - 1 = 24 - 1 = 23$.

Корректированные дисперсии: $\sigma_x^2=\frac{5897}{3}=1966$; $\sigma_A^2=\frac{4845}{1}=4845$; $\sigma_B^2=\frac{1218}{1}=1218$; $\sigma_{AB}^2=\frac{166}{1}=166$; $\sigma_z^2=\frac{1083}{20}=54$.

Коэффициент Фишера, показывающий достоверность каждой дисперсии в данном примере: $F_x=\frac{1966}{54}=26,4$; $F_A=\frac{4875}{54}=90,4$; $F_B=\frac{1218}{54}=22,5$; $F_{AB}=\frac{166}{54}=3,08$.

Табличное значение F для нашего примера при трех уровнях вероятности равно:

- степеней свободы 20—3. $F_{0,95}=3,1$; $F_{0,99}=4,9$; $F_{0,999}=8,1$.
- степеней свободы 20—1. $F_{0,95}=4,3$; $F_{0,99}=8,1$; $F_{0,999}=14,8$.

Следовательно, дисперсия, вызванная кормлением и продолжительностью светового дня, достоверна при $P\geq 0,999$. Дисперсия от совместного действия факторов A и B недостоверна, так как вычисленное $F(3,08)$ ниже табличного значения F при наличии степеней свободы 20—1.

8.5. Вычисление дисперсии двухфакторного равномерного комплекса при многочисленной выборке

Для вычисления дисперсий двухфакторного комплекса при многочисленной выборке используются следующие формулы:

- общей дисперсии: $C_y=\sum p_v\cdot a_v^2-H$, где $H=\frac{(\sum p_v\cdot a_v)^2}{n}$;
- общефакториальной дисперсии: $C_x=\sum h_x-H$;
- дисперсии фактора A: $C_A=\sum h_A-H$;
- дисперсии фактора B: $C_B=\sum h_B-H$;
- дисперсии от совместного действия факторов A и B: $C_{AB}=C_x-$ $-C_A-C_B$;
- остаточной дисперсии: $C_z=\sum p_v\cdot a_v^2-\sum h_x$.

Дисперсии C_y, C_x и C_z вычисляют так же, как в однофакторном комплексе: составляют таблицу в форме корреляционной решетки, в которой получают все данные для вычисления этих дисперсий.

Пример. Определить влияние кормов животного происхождения и продолжительности светового дня на яйценоскость кур. Необходимые данные для вычисления дисперсий приведены в табл. 8.8.

Выражение $p_A\cdot a_x$ получается от умножения частоты каждой клетки решетки на отклонения a_v и суммирования этих произведений по столбцам факторов. Записываются эти суммы в столбике каждого фактора под строкой P_A:

- 1-й столбец: $\sum p_A\cdot a_y=8\cdot(-3)+14\cdot(-2)+4\cdot(-1)=-56$;
- 2-й столбец: $\sum p_A\cdot a_y=6\cdot(-2)+16\cdot(-1)+7\cdot 0+2\cdot 1=-26$;
- 3-й столбец: $\sum p_A\cdot a_y=2\cdot(-1)+7\cdot 0+14\cdot 1+5\cdot 2=22$;
- 4-й столбец: $\sum p_A\cdot a_y=5\cdot 1+15\cdot 2+10\cdot 3=65$.

Таблица 8.8

Техника расчета двухфакторного равномерного комплекса при многочисленной выборке

v (яйценоскость, шт.)	Рацион без кормов животного происхождения		Рацион с кормами животного происхождения		p_v	a_v	$p_v a_v$	$p_v \cdot a_v^2$
	световой день 8—10 ч (B_1)	световой день 12—14 ч (B_2)	световой день 8—10 ч (B_1)	световой день 12—14 ч (B_2)				
135—149	8	—	—	—	8	–3	–24	72
150—164	14	6	—	—	20	–2	–40	80
165—179	4	16	2	—	22	–1	–22	22
180—194	—	7	7		14	0	0	0
195—209	—	2	14	5	21	1	1	21
210—224	—	—	5	15	20	2	2	80
225—239	—	—	—	10	10	3	3	90
$p_A = n_x$	26	31	28	30	115	—	$\sum p_v \cdot a_v = 5$	$\sum p_v \cdot a_v^2 = 365$
$\sum p_A \cdot a_x$	–56	–26	+22	+65	—	—	—	—
$\sum (p_A \cdot a_x)^2$	3136	676	484	4225	—	—	—	—
$h_x = \frac{\sum (p_A \cdot a_x)^2}{p_A}$	$\frac{3136}{26} = 120{,}5$	$\frac{676}{31} = 21{,}8$	$\frac{484}{28} = 17{,}5$	$\frac{4225}{30} = 141$	$\sum h_x = 300{,}8$	—	—	—

Необходимые данные ($\sum h_A$ и $\sum h_B$) для вычисления факториальных дисперсий C_A и C_B приводятся в табл. 8.9.

Таблица 8.9

Обработка комплекса по фактору A и B

Классы по факторам	Факторы	$p_x \cdot a_x$	$(p_x \cdot a_x)^2$	$h = \frac{(p_x \cdot a_x)^2}{n}$
A_1	57	–82	$(-82)^2 = 6724$	$\frac{6724}{57} = 118$
A_2	58	87	$87^2 = 7569$	$\frac{7569}{58} = 131$
По фактору A	115	—	—	$\sum h_A = 249$
B_1	54	–34	$(-34)^2 = 1156$	$\frac{1156}{54} = 21{,}4$
B_2	61	39	$39^2 = 1521$	$\frac{1521}{61} = 25{,}0$
По фактору B	115	—	—	$\sum h_B = 46{,}4$

Дисперсии вычисляют по приведенным выше формулам, подставляя в них данные из табл. 8.8 и 8.9.

Величина $H = \frac{\left(\sum p_v \cdot a_v\right)^2}{n} = \frac{5^2}{115} = \frac{25}{115} = 0{,}22;$

$$C_y = p_v \cdot a_v^2 - H = 365 - 0{,}22 = 364{,}78;$$

$$C_x = \sum h_x - H = 300{,}8 - 0{,}22 = 300{,}58;$$

$$C_B = \sum h_B - H = 46{,}4 - 0{,}22 = 46{,}18;$$

$$C_{AB} = C_x - C_A - C_B = 300{,}58 - 248{,}78 - 46{,}18 = 5{,}84;$$

$$C_z = p_v \cdot a_v - \sum h_x = 365 - 300{,}8 = 64{,}2.$$

Доля влияния учтенных факторов:

$$\eta_x^2 = \frac{300{,}58}{364{,}78} = 0{,}824, \text{ или } 82{,}4\ \%.$$

Доля влияния фактора A:

$$\eta_A^2 = \frac{248{,}5}{364{,}78} = 0{,}682, \text{ или } 68{,}2\ \%.$$

Доля влияния фактора B:

$$\eta_B^2 = \frac{46{,}18}{364{,}78} = 0{,}126, \text{ или } 12{,}6\ \%.$$

Доля влияния фактора AB:

$$\eta^2_{AB} = \frac{5,64}{364,78} = 0,016, \text{ или } 1,6\ \%.$$

Доля влияния остаточного фактора:

$$\eta^2_z = \frac{64,2}{364,78} = 0,176, \text{ или } 17,6\ \%.$$

Число степеней свободы: $\nu_x = l_A \cdot l_B - 1 = 2 \cdot 2 - 1 = 3$; $\nu_A = l_A - 1 = 2 - 1 = 1$; $\nu_B = l_B - 1 = 2 - 1 = 1$; $\nu_{AB} = \nu_A \cdot \nu_B = 1 \cdot 1 = 1$; $\nu_z = n - l_A \cdot l_B = 115 - 2 \cdot 2 = 111$; $\nu_y = n - 1 = 115 - 1 = 114$.

Корректированные дисперсии:

$$\sigma^2_x = \frac{364,8}{3} = 121,6;\ \sigma^2_A = \frac{248,78}{1} = 248,78;\ \sigma^2_B = \frac{46,18}{1} = 46,18;$$

$$\sigma^2_{AB} = \frac{5,84}{1} = 5,84;\ \sigma^2_z = \frac{64,2}{111} = 0,58.$$

Коэффициент Фишера в данном примере:

$$F_x = \frac{121,6}{0,58} = 210;\ F_A = \frac{248,78}{0,58} = 429;\ F_B = \frac{46,18}{0,58} = 80;\ F_{AB} = \frac{5,84}{0,58} = 10.$$

Табличное значение F для этого примера при трех уровнях вероятности равно:

- степеней свободы 111—3 — $F_{0,95} = 2,7$; $F_{0,99} = 4,0$; $F_{0,999} = 5,9$.
- степеней свободы 111—1 — $F_{0,95} = 3,9$; $F_{0,99} = 6,8$; $F_{0,999} = 11,5$.

Коэффициент Фишера общефакториальной дисперсии и дисперсий от факторов A и B равен: 80—249 и превышает $F_{\text{табл}}$ при всех уровнях вероятности; дисперсия, вызванная совместным действием факторов A и B, превышает $F_{\text{табл}}$ при уровне вероятности $P \geq 0,99$, следовательно все дисперсии достоверны.

Задания для самостоятельной работы

Задание 1. Оценить силу влияния фактора X по приведенным ниже данным:

1) $C_y = 90$; $C_x = 20$;
2) $C_y = 90$; $C_x = 50$;
3) $C_y = 90$; $C_x = 70$.

Задание 2. По следующим данным установить, достоверно ли влияние фактора X:

1) $C_y = 90$; $C_x = 10$; $N = 100$; $r = 5$;
2) $C_y = 90$; $C_x = 10$; $N = 50$; $r = 5$;
3) $C_y = 90$; $C_x = 10$; $N = 10$; $r = 5$.

Задание 3. Оценить силу и достоверность влияния фактора *X* на изучаемый признак по материалам приведенного ниже дисперсионного комплекса:

	X	Варианты			
Градации	1	0,3	0,5	0,2	0,2
	2	0,3	0,4	0,2	0,5
	3	0,6	0,4	0,5	0,2
	4	0,3	0,6	0,4	0,4
	5	0,3	0,5	0,6	0,5

Задание 4. Определить силу влияния метода разведения на плодовитость свиноматок. При чистопородном разведении свиней крупной белой породы плодовитость составляла 10, 9, 11, 10, 11, 10, 10, 11 поросят. Плодовитость помесных свиноматок соответственно 12, 9, 11, 10, 13, 11, 15, 10 поросят в помете.

Задание 5. На основании данных, приведенных в табл. 8.10, оценить наследственные качества быков-производителей по удою за лактацию их дочерей методом дисперсионного анализа.

Таблица 8.10

Индивидуальные различия быков по удою их дочерей

Производители	Удой за лактацию (кг) дочерей				
	№ 1	№ 2	№ 3	№ 4	№ 5
Алмаз	4500	3800	3680	5200	3900
Луч	3800	3750	4200	3500	4100
Ветер	4100	4000	3900	4150	3850

Задание 6. Изучить действие добавки травяной муки (*B*) на плодовитость свиноматок, где установлено две градации этого фактора: (B_1) — первая группа 0 %, (B_2) — вторая группа +25 % к суточному рациону по питательности. Фактор (*A*) — порода свиноматок, его градации: (A_1) — крупная белая, (A_2) — миргородская.

Для каждой градации фактора было выбрано по две особи (супоросные свиноматки), которые за опытный период показали следующую плодовитость:

B_1 1-я группа 8 и 11 поросят

A_1 — крупная белая

B_2 2-я группа 9 и 13 поросят

B_1 1-я группа 10 и 15 поросят

A_2 — миргородская

B_2 2-я группа 11 и 17 поросят

Задание 7. Изучить действие белковой добавки животного происхождения на плодовитость свиноматок, если установлено три градации этого фактора: первая группа — 0 %, вторая группа — (+10 %)

и третья группа — (+20 %) добавки к суточной норме переваримого протеина. В качестве результативного признака была принята плодовитость свиноматок за опытный период. Для каждой градации фактора (0 %; +10 %; +20 %) было выбрано с сохранением принципа случайности по две особи (свиноматки), которые за опытный период показали разную плодовитость, выражающуюся в следующих числах: первая группа — 14 и 6; вторая — 20 и 16; третья — 10 и 6 поросят.

Контрольные вопросы

1. В чем заключается цель дисперсионного анализа?
2. Что называется общей факториальной и остаточной дисперсией?
3. Какие бывают дисперсионные комплексы? Чем они характеризуются?
4. Как составляют однофакторный дисперсионный комплекс при малочисленной выборке и вычисляют вспомогательные величины?
5. Какие показатели используются для оценки силы и достоверности влияния изучаемого фактора?
6. Как составляется однофакторный дисперсионный комплекс при многочисленной выборке?
7. Как составляется двухфакторный дисперсионный комплекс при малочисленной выборке?
8. Как составляется двухфакторный равномерный дисперсионный комплекс при многочисленной выборке?
9. Как можно выявить долю влияния одного из факторов на изменчивость признаков?
10. Что называют градациями фактора, объемом градации?
11. Как можно проверить правильность вычисленных дисперсий?
12. Как можно определить достоверность влияния изучаемого фактора на изменчивость признака при использовании дисперсионного анализа?

Контрольные вопросы и задания для зачета

1. Что такое генеральная совокупность?
2. Что такое выборка? Как ее составляют?
3. Какие выборки называют большими и какие — малыми?
4. Как составляют вариационный ряд?
5. Какими способами можно графически изобразить вариационные ряды?
6. Какие бывают типы распределения вариационных кривых?
7. Перечислите средние величины. Как их используют?
8. Как вычисляют среднюю арифметическую величину в малых и больших выборках?
9. Что такое средняя взвешенная? В каких случаях она применяется? Как ее вычисляют?
10. Как вычисляют непараметрическую среднюю?
11. Как вычисляют среднюю геометрическую?
12. Как вычисляют среднюю квадратическую?
13. Как вычисляют среднюю гармоническую?
14. Что такое мода и медиана?
15. Какими свойствами обладают средние величины?
16. Какие показатели характеризуют разнообразие признаков?
17. Как вычисляют среднее квадратическое отклонение в малых и больших выборках?
18. Как вычисляют среднее квадратическое отклонение для альтернативных признаков?
19. В каких случаях вычисляют коэффициент вариации? Приведите его формулу.
20. Что такое нормированное отклонение? Для чего используют этот показатель?
21. Какие показатели применяются для измерения связи между признаками?
22. Как вычисляют коэффициент фенотипической корреляции в малых выборках?
23. Как вычисляют коэффициент фенотипической корреляции в больших выборках?
24. В чем заключается различие связи между признаками при положительных и отрицательных значениях коэффициента корреляции?
25. Как вычисляется коэффициент корреляции для альтернативных признаков?
26. В каких случаях используется коэффициент ранговой корреляции?
27. Приведите формулы вычисления коэффициентов генетической корреляции.
28. Что характеризуют коэффициенты регрессии? В чем различие между коэффициентами $R_{x/y}$ и $R_{y/x}$?
29. В чем различие между коэффициентами r и R?
30. Что такое ошибки репрезентативности? Чем отличаются они от ошибок измерения и вычисления?

31. Как вычисляют ошибку средней арифметической величины?

32. Приведите формулы вычисления ошибок σ, Cv, r, R.

33. Как изменяется величина $m_{\bar{x}}$ при изменении объема выборки и величины сигмы?

34. Что такое доверительные вероятности?

35. Какие доверительные вероятности можно использовать в биологических, зоотехнических и ветеринарных исследованиях?

36. Как определяют доверительные границы при работе с большими и малыми выборками?

37. Как определяется достоверность выборочных показателей?

38. Как оценивается достоверность между средними величинами двух выборок?

39. Что такое критерий соответствия (хи-квадрат)? Как он используется в генетических исследованиях?

40. В чем заключается цель дисперсионного анализа? Что называется общей, факториальной, и остаточной дисперсией?

41. Какие бывают дисперсионные комплексы? Чем они характеризуются?

42. Как составляют однофакторный дисперсионный комплекс и вычисляют вспомогательные величины?

43. Какие показатели используются для оценки силы и достоверности влияния изучаемого фактора?

Библиографический список

1. *Абрамова, З. В.* Учебное пособие по генетике : задания по программированному обучению и контролю / З. В. Абрамова. — Л. ; Пушкин, 1997.

2. *Антипов, Г. П.* Методические указания к практическим занятиям по курсу «Генетика с основами биометрии» / Г. П. Антипов, А. П. Лисицын, В. В. Лавровский. — М. , 1986.

3. *Бакай, А. В.* Генетика / А. В. Бакай, И. И. Кочиш, Г. Г. Скрипниченко. — М. : Колос, 2006.

4. *Габитов, Н. А.* Методические разработки к курсу генетики для студентов зооинженерного и ветеринарного факультетов / Н. А. Габитов, М. К. Шайдуллов. — Казань, 1982.

5. *Гофман-Кадошников, П. Б.* Руководство к практическим занятиям по генетике / П. Б. Гофман-Кадошников, С. Х. Ларцева. — М. : Колос, 1976.

6. *Зубрич, А. С.* Практическое пособие по биометрии / А. С. Зубрич [и др.]. — Харьков, 1974.

7. *Крупин, В. Г.* Высшая математика. Теория вероятностей, математическая статистика, случайные процессы. Сборник задач с решениями: учебное пособие / В. Г. Крупин, А. Л. Павлов, Л. Г. Попов. — М. : Издательский дом МЭИ, 2013.

8. *Лакин, Г. Ф.* Биометрия / Г. Ф. Лакин. — М. : Высшая школа, 1980.

9. *Ларцева, С. Х.* Практикум по генетике / С. Х. Ларцева, М. К. Муксинов. — М. : Агропромиздат, 1985.

10. *Лебедько, Е. Я.* Биометрия в MS Excel: учеб. пособие / Е. Я. Лебедько [и др.]. — СПб. : Лань, 2018.

11. *Марченко, Г. Г.* Методические рекомендации к лабораторным занятиям по генетике сельскохозяйственных животных / Г. Г. Марченко, Л. П. Ефименко. — Саратов, 1981.

12. *Меркурьева, Е. Н.* Биометрия в селекции и генетике сельскохозяйственных животных / Е. Н. Меркурьева. — М. : Колос, 1970.

13. *Меркурьева, Е. К.* Генетика / Е. Н. Меркурьева [и др.]. — М. : Агропромиздат, 1991.

14. *Меркурьева, Е. К.* Генетика с основами биометрии / Е. К. Меркурьева, Г. Н. Шангин-Березовский. — М. : Колос, 1983.

15. *Мицель, А. А.* Прикладная математическая статистика: лабор. практикум / А. А. Мицель. — Томск : ТУСУР, 2015.

16. *Насыров, Ю. Я.* Методические разработки задач по вариационной статистике для студентов-зоотехников / Ю. Я. Насыров. — Кинель, 1968.

17. *Плохинский, Н. А.* Руководство по биометрии для зоотехников / Н. А. Плохинский. — М. : Колос, 1969.

18. *Плохинский, Н. А.* Биометрия / Н. А. Плохинский. — М. : изд-во Московского университета, 1970.

19. *Соломенко, Л. К.* Биометрия : пособие к практикуму для студентов-заочников зоотехнического факультета / Л. К. Соломенко. — Москва, 1971.

Приложения

Приложение 1

Стандартные значения критерия достоверности Стьюдента (t_{st}) при трех уровнях вероятности (P)

Число степеней свободы (ν)	Уровень вероятности (P)			Число степеней свободы (ν)	Уровень вероятности (P)		
	0,95	0,99	0,999		0,95	0,99	0,999
	значения				значения		
1	12,71	63,66	637,00	19	2,09	2,86	3,88
2	4,30	9,92	31,60	20	2,09	2,85	3,85
3	3,18	5,84	12,94	21	2,08	2,83	3,82
4	2,78	4,60	8,61	22	2,07	2,82	3,79
5	2,57	4,03	6,86	23	2,07	2,81	3,77
6	2,45	3,71	5,96	24	2,06	2,80	3,75
7	2,37	3,50	5,41	25	2,06	2,79	3,73
8	2,31	3,36	5,04	26	2,06	2,78	3,71
9	2,26	3,25	4,78	27	2,05	2,77	3,69
10	2,23	3,17	4,59	28	2,05	2,76	3,67
11	2,20	3,11	4,44	29	2,05	2,76	3,66
12	2,18	3,06	4,32	30	2,04	2,75	3,65
13	2,16	3,01	4,22	35—39	2,03	2,72	3,59
14	2,15	2,98	4,14	40—44	2,02	2,70	3,55
15	2,13	2,95	4,07	45—60	2,01	2,66	3,50
16	2,12	2,92	4,02	70—100	1,98	2,63	3,39
17	2,11	2,90	3,97	120 — ∞	1,96	2,58	3,29
18	2,10	2,88	3,92				

Приложение 2

Стандартные значения критерия χ^2 при трех уровнях вероятности (P)

Число степеней свободы (ν)	Уровень вероятности (P)			Число степеней свободы (ν)	Уровень вероятности (P)		
	0,95	0,99	0,999		0,95	0,99	0,999
	значения				значения		
1	3,8	6,6	10,8	16	46,3	32,0	39,3
2	6,0	9,2	13,8	17	27,6	33,4	40,8
3	7,8	11,3	16,3	18	28,9	34,8	42,3
4	9,5	13,3	18,5	19	30,1	36,2	43,8
5	11,1	15,1	20,5	20	31,4	37,6	45,3
6	12,6	16,8	22,5	25	37,7	44,3	52,6
7	14,1	18,5	24,3	30	43,8	50,9	59,7
8	15,5	20,1	26,1	40	55,8	63,7	73,4
9	16,9	21,7	27,9	50	67,5	76,2	86,7
10	18,3	23,2	29,6	60	79,1	88,4	99,6
11	19,7	24,7	31,3	70	90,5	100,4	112,3
12	21,0	26,2	32,9	75	96,2	106,4	118,5
13	22,4	27,7	34,5	80	101,9	112,3	124,8
14	23,7	29,1	36,1	90	113,1	124,1	137,1
15	25,0	30,6	37,7	100	124,3	135,8	149,4

Приложение 3

Стандартные значения критерия Фишера $F = \frac{\sigma_x^2}{\sigma_y^2}$ при трех уровнях вероятности
($P_1 = 0{,}999$; $P_2 = 0{,}99$; $P_3 = 0{,}95$)

Число степеней свободы (ν_y)	Число степеней свободы (ν_x)											
	1	2	3	4	5	6	7	8	9	10	11	12
3	167,5 34,1 10,1	148,5 30,8 9,6	141,1 29,5 9,3	137,1 28,7 9,1	134,6 28,2 9,0	132,9 27,9 8,9	131,8 27,7 8,9	130,6 27,5 8,8	130,0 27,4 8,8	129,5 27,2 8,8	128,9 27,1 8,8	128,3 27,1 8,7
4	74,1 21,2 7,7	61,2 18,8 6,9	56,1 16,7 6,6	53,4 16,0 6,4	51,7 15,5 6,3	50,5 15,2 6,2	49,8 15,0 6,1	49,0 14,8 6,0	48,6 14,7 6,0	48,2 14,5 6,0	47,8 14,5 5,9	47,4 14,4 5,9

Окончание таблицы

Число степеней свободы (ν_2)	Число степеней свободы (ν_1)											
	1	2	3	4	5	6	7	8	9	10	11	12
5	47,0 16,3 6,6	36,6 13,3 5,8	33,2 12,1 5,4	31,1 11,4 5,2	29,8 11,0 5,1	28,8 10,7 5,0	28,2 10,5 4,9	27,6 10,3 4,8	27,3 10,2 4,8	27,0 10,1 4,7	26,7 10,0 4,7	26,4 9,9 4,7
6	35,5 13,4 6,0	27,0 10,9 5,1	23,7 9,8 4,8	21,9 9,2 4,5	20,8 8,8 4,4	20,0 8,5 4,3	19,5 8,3 4,2	19,0 8,1 4,1	18,8 8,0 4,1	18,5 7,9 4,1	18,3 7,8 4,0	18,0 7,7 4,0
7	29,2 12,3 5,6	21,7 9,6 4,7	18,8 8,5 4,4	17,2 7,9 4,1	16,2 7,5 4,0	15,5 7,2 3,9	15,1 7,0 3,8	14,6 6,8 3,7	14,4 6,7 3,7	14,2 6,6 3,6	13,9 6,5 3,6	13,7 6,5 3,6
8	25,4 11,3 5,3	18,5 8,7 4,5	15,8 7,6 4,1	14,4 7,0 3,8	13,5 6,6 3,7	12,9 6,4 3,6	12,5 6,2 3,5	12,0 6,0 3,4	11,8 5,9 3,4	11,6 5,8 3,3	11,4 5,7 3,1	11,2 5,7 3,3
9	22,9 10,6 5,1	16,4 8,0 4,3	13,9 7,0 3,9	12,6 6,4 3,6	11,7 6,1 3,5	11,1 5,8 3,4	10,8 5,6 3,3	10,4 5,5 3,2	10,2 5,4 3,2	10,0 5,3 3,1	9,8 5,2 3,1	9,6 5,1 3,1
10	21,0 10,0 5,0	14,9 7,9 4,1	12,3 6,6 3,7	11,3 6,0 3,5	10,5 5,6 3,3	9,9 5,4 3,2	9,6 5,2 3,1	9,2 5,1 3,1	9,0 5,0 3,0	8,9 4,9 3,0	8,7 4,8 2,9	8,5 4,7 2,9
20	14,8 8,1 4,3	10,0 5,8 3,5	8,1 4,9 3,1	7,1 4,4 2,9	6,5 4,1 2,7	6,0 3,9 2,6	5,7 3,7 2,5	5,4 3,6 2,4	5,3 3,4 2,4	5,1 3,4 2,3	5,0 3,3 2,3	4,8 3,2 2,3
30	13,3 7,6 4,2	8,8 5,4 3,3	7,1 4,5 2,9	6,1 4,0 2,7	5,5 3,7 2,5	5,1 3,5 2,4	4,9 3,3 2,3	4,6 3,2 2,2	4,5 3,1 2,2	4,3 3,0 2,2	4,2 2,9 2,1	4,0 2,8 2,1
50	12,2 7,2 4,0	8,0 5,1 3,2	6,4 4,2 2,8	5,4 3,7 2,6	4,9 3,4 2,4	4,5 3,2 2,3	4,3 3,0 2,2	4,0 2,9 2,1	3,9 2,8 2,1	3,7 2,7 2,0	3,6 2,6 2,0	3,1 2,6 1,9
100	11,5 6,9 3,9	7,4 4,8 3,1	5,9 4,0 2,7	5,0 3,5 2,5	4,5 3,2 2,3	4,1 3,0 2,2	3,9 2,8 2,1	3,7 2,7 2,0	3,6 2,6 2,0	3,4 2,5 1,9	3,3 2,4 1,9	3,1 2,4 1,8
150	11,3 6,8 3,9	7,3 4,7 3,1	5,7 3,9 2,7	4,9 3,4 2,4	4,4 3,1 2,3	4,0 2,9 2,2	3,8 2,8 2,1	3,5 2,6 2,0	3,4 2,5 1,9	3,2 2,4 1,9	3,1 2,4 1,8	2,9 2,3 1,8

Число степеней свободы (ν_2)	Число степеней свободы (ν_1)												
	14	16	20	24	30	40	50	75	100	200	500	∞	t
3	127,7 26,9 8,7	127,1 26,8 8,7	126,5 26,7 8,7	125,9 26,6 8,6	125,6 26,5 8,6	125,3 26,4 8,6	125,0 26,4 8,6	124,7 26,3 8,6	124,4 26,2 8,6	124,1 26,2 8,5	123,8 26,1 8,5	123,5 26,1 8,5	12,9 5,8 3,2
4	47,0 14,2 5,9	46,6 14,1 5,8	46,2 14,0 5,8	45,8 13,9 5,8	45,6 13,8 5,7	45,4 13,7 5,7	45,2 13,7 5,7	45,0 13,6 5,7	44,7 13,6 5,7	44,5 13,5 5,7	44,3 13,5 5,6	44,1 13,5 5,6	8,6 4,6 2,8
5	26,1 9,8 4,6	25,8 9,7 4,6	25,4 9,6 4,6	25,1 9,5 4,5	24,9 9,4 4,5	24,8 9,3 4,5	24,6 9,2 4,4	24,5 9,1 4,4	24,3 9,1 4,4	24,1 9,1 4,4	24,0 9,0 4,4	23,8 9,0 4,4	6,9 4,0 2,6
6	17,7 7,6 4,0	17,5 7,5 3,9	17,2 7,4 3,9	16,9 7,3 3,8	16,8 7,2 3,8	16,6 7,1 3,8	16,5 7,1 3,8	16,4 7,0 3,7	16,2 7,0 3,7	16,1 6,9 3,7	15,9 6,9 3,7	15,8 6,9 3,7	6,0 3,7 2,4
7	13,5 6,4 3,5	13,2 6,3 3,5	13,0 6,2 3,4	12,7 6,1 3,4	12,6 6,0 3,4	12,5 5,9 3,3	12,3 5,9 3,3	12,2 5,8 3,3	12,1 5,8 3,3	12,0 5,7 3,3	11,8 5,7 3,2	11,7 5,7 3,2	5,3 3,5 2,4
8	11,0 5,6 3,2	10,8 5,5 3,2	10,5 5,4 3,2	10,3 5,3 3,1	10,2 5,2 3,1	10,1 5,1 3,1	10,0 5,1 3,0	9,9 5,0 3,0	9,7 5,0 3,0	9,6 4,9 3,0	9,5 4,9 2,9	9,4 4,9 2,9	5,0 3,4 2,3
9	9,4 5,0 3,0	9,2 4,9 3,0	8,9 4,8 2,9	8,7 4,7 2,9	8,6 4,6 2,9	8,5 4,6 2,8	8,4 4,5 2,8	8,3 4,5 2,8	8,1 4,4 2,8	8,0 4,4 2,7	7,9 4,3 2,7	7,8 4,3 2,7	4,8 3,3 2,3
10	8,3 4,6 2,9	8,1 4,5 2,8	7,8 4,4 2,8	7,6 4,3 2,7	7,5 4,3 2,7	7,4 4,2 2,7	7,3 4,1 2,6	7,2 4,1 2,6	7,1 4,0 2,6	7,0 4,0 2,6	6,9 3,9 2,6	6,8 3,9 2,5	4,6 3,2 2,2
20	4,7 3,1 2,2	4,5 3,0 2,2	4,4 2,9 2,1	4,2 2,9 2,1	4,1 2,8 2,0	4,0 2,7 2,0	3,9 2,6 2,0	3,8 2,6 1,9	3,7 2,5 1,9	3,6 2,5 1,9	3,5 2,4 1,8	3,4 2,4 1,8	3,9 2,8 2,1
30	3,9 2,7 2,0	3,7 2,7 2,0	3,6 2,5 1,9	3,4 2,5 1,9	3,3 2,4 1,8	3,2 2,3 1,8	3,1 2,2 1,8	3,0 2,2 1,7	2,9 2,1 1,7	2,8 2,1 1,7	2,7 2,0 1,6	2,6 2,0 1,6	3,6 2,8 2,0
50	3,3 2,5 1,9	3,2 2,4 1,8	3,0 2,3 1,8	2,9 2,2 1,7	2,7 2,1 1,7	2,7 2,0 1,6	2,6 1,9 1,6	2,5 1,9 1,5	2,3 1,8 1,5	2,2 1,8 1,5	2,1 1,7 1,5	2,0 1,7 1,4	3,5 2,7 2,0
100	3,0 2,3 1,8	2,8 2,2 1,7	2,7 2,1 1,7	2,5 2,0 1,6	2,4 1,9 1,6	2,3 1,8 1,5	2,2 1,7 1,5	2,1 1,6 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,7 1,5 1,3	1,6 1,4 1,3	3,4 2,6 2,0
150	2,8 2,2 1,8	2,7 2,1 1,7	2,5 2,0 1,6	2,4 1,9 1,6	2,2 1,8 1,5	2,2 1,7 1,5	2,0 1,7 1,4	1,9 1,6 1,4	1,8 1,5 1,3	1,7 1,4 1,3	1,5 1,4 1,2	1,4 1,3 1,2	3,4 2,6 2,0

Глоссарий

Биометрия — наука о применении математических методов в биологических исследованиях при изучении групповых свойств биологических объектов.

Большая выборка — выборка с численностью 30 особей и более.

Варианта — величина признака у отдельной особи.

Вариационный ряд — группировка вариант сырого ряда по классам.

Взвешенная средняя арифметическая — результат усреднения средних арифметических нескольких совокупностей.

Выборочная совокупность (выборка) — часть генеральной совокупности, выделенная по специальным правилам и предназначенная для ее характеристики.

Генеральная совокупность — большой массив животных, интересующих исследователя.

Градации комплекса — выборки, сделанные из заведомо разных генеральных совокупностей, поэтому наиболее применяемым принципом формирования градаций комплекса является принцип случайного отбора в них отдельных объектов из совокупностей.

Дисперсия — варьирование или изменчивость признака, возникающая под влиянием различных факторов. Дисперсия представляет сумму квадратов отклонений вариант от средней арифметической.

Достоверность — возможность обобщить данные опыта, перенести результаты исследования выборок на соответствующие генеральные совокупности.

Изменчивость — способность организмов изменяться под действием наследственных и ненаследственных факторов.

Качественные признаки — морфологические или биологические признаки, проявление которых легко может быть словесно охарактеризовано (масть, форма рогов, и т. д.).

Количественные признаки — признаки, которые не имеют четкого выражения, их изучают путем измерения, подсчета (живая масса, длина шерсти, жирность молока и т. д.).

Корреляция — взаимосвязь между отдельными признаками. По форме корреляция может быть прямолинейной и криволинейной, по направлению — прямой (положительной) и обратной (отрицательной).

Коэффициент изменчивости (вариации) — величина сигмы (σ), выраженная в процентах к средней арифметической; применяется для сравнения изменчивости различных признаков.

Коэффициент корреляции — величина, определяющая направление и силу прямолинейной корреляции.

Коэффициент прямолинейной регрессии — величина, указывающая, насколько в среднем изменяется один из признаков при изменении другого на единицу измерения.

Криволинейная связь — связь, при которой с увеличением одного признака другой увеличивается до определенной величины, а затем уменьшается (или наоборот).

Критерий хи-квадрат (критерий соответствия) — показатель, используемый для проверки гипотез путем сравнения фактического распределения с теоретическим.

Малая выборка — выборка с численностью менее 30 особей.

Медиана — середина класса, которая делит вариационный ряд на две части: одна имеет значение признака меньшее, чем медиана, другая — большее.

Мода — наиболее часто встречающаяся варианта в вариационном ряду.

Модальный класс — класс с наибольшим количеством частот вариант, т. е. животных.

Нормированное отклонение — показатель, характеризующий отдельную варианту. Оно является также показателем разнообразия признака и представляет выраженное в долях сигмы взвешенное отклонение соответствующей варианты от средней арифметической.

Общая дисперсия — сумма факториальной и остаточной дисперсий, или изменчивость, вызываемая всеми одновременно действующими факторами.

Остаточная дисперсия — дисперсия, возникающая под влиянием различных случайных (неучтенных), неорганизованных факторов.

Отрицательная корреляция — корреляция, при которой с увеличением одного признака связанный с ним признак уменьшается, или, наоборот, с уменьшением одного признака другой увеличивается.

Ошибки репрезентативности — ошибки, возникающие при характеристике генеральной совокупности показателями, полученными при изучении выборки.

Положительная корреляция — корреляция, при которой с увеличением (или уменьшением) одного признака связанный с ним другой признак также увеличивается (или уменьшается).

Признак (свойство) — единица морфологической, физиологической или биологической дискретности организма.

Прямолинейная связь — связь, при которой равномерным изменениям одного признака соответствуют равномерные изменения второго признака при незначительных отклонениях.

Ранжированный ряд — варианты, расположенные по мере их возрастания или убывания.

Регрессия — изменение функции в зависимости от изменения аргумента.

Совокупность —группа, генетически сходных между собой животных, взятая для исследования.

Среднее квадратическое отклонение (σ — сигма) — отклонение каждой варианты от средней арифметической. Сигма — величина именованная и выражается в тех же единицах, что и средняя арифметическая.

Средняя арифметическая — среднее значение признака, приходящееся на одно животное в совокупности (средняя масса тела, удой, настриг шерсти).

Сырой ряд — варианты выборочной совокупности, расположенные произвольно по мере их получения.

Условная средняя арифметическая — середина модального класса.

Фактор — любое проявление признака, влияние которого требуется изучить на результативный признак.

Факториальная дисперсия — дисперсия, возникающая под влиянием различных учтенных, организованных факторов.

Число степеней свободы для общей дисперсии — численность выборки минус единица.

Число степеней свободы для остаточной дисперсии — численность выборки минус число классов или градаций.

Число степеней свободы для факториальной дисперсии — число классов или градаций по фактору минус единица.

Учебное издание

Катмаков Петр Сергеевич,
Гавриленко Владимир Петрович,
Бушов Александр Владимирович

БИОМЕТРИЯ

Учебное пособие для вузов

Под общей редакцией *П. С. Катмакова*

Формат 70×100 1/16.
Гарнитура «Charter». Печать цифровая.
Усл. печ. л. 13,73.

ООО «Издательство Юрайт»
111123, г. Москва, ул. Плеханова, д. 4а.
Тел.: (495) 744-00-12. E-mail: izdat@urait.ru, www.urait.ru