

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА И ПРОДОВОЛЬСТВИЯ
РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«ВИТЕБСКАЯ ОРДЕНА «ЗНАК ПОЧЕТА» ГОСУДАРСТВЕННАЯ
АКАДЕМИЯ ВЕТЕРИНАРНОЙ МЕДИЦИНЫ»

Т. В. Павлова, В. Ф. Соболева, Т. В. Видасова

БИОМЕТРИЯ

Учебно-методическое пособие по дисциплине «Биометрия»

для магистрантов по специальности
1-74 80 04 «Ветеринария»

Витебск
ВГАВМ
2022

УДК 636.082
ББК 45.3
П12

Рекомендовано к изданию методической комиссией
факультета ветеринарной медицины
УО «Витебская ордена «Знак Почета» государственная
академия ветеринарной медицины»
от 15 июня 2022 г. (протокол №5)

Авторы:

кандидат биологических наук, доцент *Т. В. Павлова*;
кандидат сельскохозяйственных наук, доцент *В. Ф. Соболева*;
кандидат сельскохозяйственных наук, доцент *Т. В. Видасова*

Рецензенты:

кандидат ветеринарных наук, доцент *В. А. Патафеев*;
кандидат ветеринарных наук, доцент *А. В. Богомольцев*

Павлова, Т. В.

Биометрия : учебно-методическое пособие по дисциплине
П12 «Биометрия» для магистрантов по специальности 1-74 80 04
«Ветеринария» / Т. В. Павлова, В. Ф. Соболева, Т. В. Видасова. –
Витебск : ВГАВМ, 2022. – 80 с.

Учебно-методическое пособие подготовлено в соответствии с
учебной программой по дисциплине «Биометрия» для магистрантов по
специальности 1-74 80 04 «Ветеринария», содержит методические указа-
ния по выполнению практических занятий.

УДК 636.082
ББК 45.3

© УО «Витебская ордена «Знак Почета»
государственная академия ветеринарной
медицины», 2022

Электронное учебное издание

Павлова Татьяна Владимировна,
Соболева Валентина Федоровна,
Видасова Татьяна Викторовна

БИОМЕТРИЯ

Для создания электронного издания
использовалось следующее программное обеспечение:
Microsoft Office Word 2007,
doPDF v 7.

Минимальные системные требования:
Internet Explorer 6 или более поздняя версия;
Firefox 30 или более поздняя версия;
Chrome 35 или более поздняя версия.
Скорость подключения не менее 1024 Кбит/с.

Ответственный за выпуск Т. В. Павлова
Технический редактор О. В. Луговая
Компьютерный набор В. Ф. Соболева
Компьютерная верстка Е. В. Морозова
Корректор Т. А. Никитенко

Дата размещения на сайте 15.09.2022 г.

Объем издания 2 604 Кб.

Режим доступа: <http://www.vsavm.by>

Технические требования: сетевое электронное издание.
Издатель: учреждение образования «Витебская ордена «Знак Почета»
государственная академия ветеринарной медицины».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/ 362 от 13.06.2014.
ЛП №: 02330/470 от 01.10.2014 г.
Ул. 1-я Доватора, 7/11, 210026, г. Витебск.
Тел.: (0212) 48-17-82.
E-mail: rio_vsavm@tut.by
<http://www.vsavm.by>

СОДЕРЖАНИЕ

	Введение	5
Тема 1.	Предмет и основные понятия биометрии. Понятие изменчивости	6
Тема 2.	Средние величины и способы их вычисления	8
Тема 3.	Вариационный ряд. Основные статистические параметры для больших выборок. Построение гистограммы и вариационной кривой	18
Тема 4.	Основные статистические величины для малых выборок	26
Тема 5.	Изучение связи между признаками. Вычисление коэффициента корреляции для больших и малых выборок по количественным признакам	28
Тема 6.	Вычисление коэффициента корреляции для альтернативных признаков	35
Тема 7.	Ранговый коэффициент связи по Спирмену	37
Тема 8.	Регрессивный анализ. Вычисление коэффициента прямой линейной регрессии	39
Тема 9.	Изучение криволинейного типа связи и вычисление корреляционного отношения	41
Тема 10.	Вычисление коэффициента генетической корреляции	45
Тема 11.	Полихорический и бисериальный показатели связи	47
Тема 12.	Вычисление критерия достоверности разности между варьирующими признаками	51
Тема 13.	Границы доверительного интервала. Определение необходимого объема выборки и правило отбрасывания резко отклоняющихся вариантов из совокупности	56
Тема 14.	Оценка достоверности методом хи-квадрат (χ^2)	60
Тема 15.	Определение частот аллелей и генотипов в популяции	64
Тема 16.	Дисперсионный анализ	66
Тема 17.	Обработка вариационных рядов на ПЭВМ. Анализ полученных данных	70
Приложение 1.	Стандартные значения критерия t_d по Стьюденту	75
Приложение 2.	Таблица значений χ^2 при разных уровнях вероятности	76
Приложение 3.	Стандартные значения критерия Фишера (F)	77
	Литература	79

ВВЕДЕНИЕ

Биометрия (вариационная статистика) – представляет собой один из разделов высшей математики, который занимается разработкой методов изучения изменчивости признаков. Биометрия основывается на законе больших чисел и теории вероятности.

Биометрия – вариационно-статистический метод – применяется при анализе массовых данных в области биологии. Развитие отрасли животноводства сопровождается накоплением большого количества информации, поэтому для успешного ведения научно-исследовательской и практической работы необходимо проводить классификацию данных, упорядочение и систематизацию, научный анализ, с формулировкой практических предложений для той или иной отрасли сельского хозяйства. Основная задача биометрии заключается в получении комплекса параметров и коэффициентов, характеризующих членов изучаемой группы по одному или нескольким признакам.

Биометрия позволяет систематизировать и обрабатывать числовые данные, получаемые при изучении биологических объектов в условиях экспериментов, при обобщении производственных первичных записей, проводимых в животноводстве или других отраслях сельского хозяйства, а также при анализе племенной документации. Вариационно-статистический метод применяется при генетическом анализе различных популяций (пород, стад, линий, семейств) в отношении количественных признаков. Позволяет установить степень наследуемости признаков, определить эффект селекции и интенсивность отбора. Математический анализ массовых данных, исходящий из биометрического метода, находит широкое применение при решении ряда теоретических и практических вопросов генетики, селекции и племенного дела [8, 9].

На основании анализа полученных статистических параметров даются рекомендации о применении определенных методов разведения, кормления, выращивания, лечения и продуктивного использования сельскохозяйственных животных.

Тема 1. ПРЕДМЕТ И ОСНОВНЫЕ ПОНЯТИЯ БИОМЕТРИИ. ПОНЯТИЕ ИЗМЕНЧИВОСТИ

Цель занятия: изучить основные термины и понятия биометрии, ознакомиться с видами изменчивости.

Основные понятия

Биометрия (вариационная статистика) – это наука, основанная на использовании математических методов в биологии.

Предмет биометрии – это группа биологических объектов.

Объект биометрии – это варьирующий признак, учтенный в группе особей, имеющей достаточную численность и однородной по ряду других основных признаков. При проведении исследований животные должны быть одного вида, возраста, одной породы и находиться в аналогичных условиях кормления и содержания.

Вариация (изменчивость) - это различие между единицами совокупности, между отдельными вариантами.

Варьирующий признак – это признак, принимающий различные значения у разных членов совокупности.

Варианта (x) – величина изучаемого признака у отдельного объекта совокупности.

Совокупность – группа особей, в пределах которой изучается варьирующий признак. В состав совокупности входят различные члены или единицы. К совокупности относится также сумма наблюдений или измерений. Каждое отдельное наблюдение, при котором устанавливается значение случайной переменной, является единицей этой совокупности.

Генеральная совокупность (N) – это теоретически бесконечно большая или приближающаяся к бесконечности совокупность всех единиц или членов, это группа животных, составляющих вид, породу в целом.

Выборочная совокупность, выборка (n) – это часть генеральной совокупности, она должна быть типичной, т.е. правильно отражать генеральную совокупность. Должна быть однородной, составляется по принципу случайного отбора. Различают многочисленные (большие) и малочисленные (малые) выборки, для которых используются различные методы обработки показателей признаков. Большими называют выборки с численностью свыше 30 особей, малыми – до 30 особей.

Объем совокупности – число единиц совокупности. Каждая единица совокупности характеризуется определенными признаками.

Репрезентативность – это достоверное отражение структуры генеральной совокупности, которое достигается принципом случайной выборки.

Рендоминизация – метод случайного отбора, когда имеется равная возможность любому члену совокупности попасть в выборку. Каждый член выборки должен быть отобран случайно.

Признак – это характерное свойство особи. Признаки подразделяются на количественные и качественные (альтернативные).

Количественные – это признаки, которые можно измерить и выразить в конкретных единицах измерения (кг, %, см, м/с, м³ и т. д.). К ним относятся большинство хозяйственно полезных признаков (удой, содержание жира и белка в молоке, скорость бега, затраты корма и др.), количество эритроцитов, лейкоцитов, гемоглобина и т.д. Количественные признаки могут изменяться и создавать непрерывный ряд показателей (масса, молочная продуктивность, пульс) или прерывистый, который также называют дискретным (количество сосков, количество детенышей и т. д.).

Качественные (альтернативные) – это признаки, которые не поддаются непосредственному измерению. У животных это пол (мужской или женский), окраска шерстного покрова (альбинос, пигментированность и др.), тип шерстного покрова (волнистая шерсть, прямая, длинная, короткая и др.), рогатость или комолость, тип нервной деятельности и т. д. Все качественные признаки наследуются по законам Менделя.

Изменчивость – это свойство живых организмов, которое заключается в способности дочерних организмов отличаться от родителей морфологическими, физиологическими, биохимическими особенностями и отклонениями в индивидуальном развитии. Она является одним из основных факторов, обуславливающих эволюцию. Изменчивость подразделяется на *определенную*, отличающую одни группы индивидуумов от других и возникающую вследствие влияния определенных факторов на всю группу особей, и *неопределенную*, свойственную отдельным особям и отличающую их друг от друга.

Выделяют изменчивость *наследственную* и *ненаследственную*. Наследственная изменчивость возникает в результате нового сочетания в потомстве наследственных особенностей родителей или появления мутаций. Она передается по наследству. Ненаследственная изменчивость не изменяет генотип и не передается по наследству.

К наследственной изменчивости относят следующие ее виды: *онтогенетическая*, *комбинативная*, *мутационная*, *коррелятивная*. Ненаследственная изменчивость представлена *модификационной*.

Онтогенетическая изменчивость – это совокупность последовательных изменений признаков и свойств особи в процессе ее индивидуального развития (онтогенеза).

Комбинативная изменчивость. Наблюдается обычно в потомстве, полученном в результате скрещивания животных различных пород и видов, и у растений разных сортов.

Мутационная изменчивость. Эта изменчивость характеризуется внезапным появлением у особи каких-либо новых особенностей, которых не было у его предков. Эти изменения называются мутациями, процесс их появления называется мутагенез, а особь, у которой мутировал тот или иной признак, – мутантом.

Коррелятивная изменчивость. Связана с взаимодействием признаков в организме. Связи между признаками бывают положительные (когда признаки изменяются в одну сторону, или оба уменьшаются, или оба увеличиваются) и отрицательные (когда признаки изменяются в противоположные стороны, при

этом усиление развития одного признака тормозит развитие другого).

Модификационная изменчивость. Возникает у растений и животных под непосредственным влиянием среды. Она не изменяет наследственность и не передается по наследству. Модификационная изменчивость не наследуется, но возможны *длительные модификации*, которые могут сохраняться в ряде поколений при сохранении условий, которые определили их появление [6, 11].

Контрольные вопросы:

1. Предмет и задачи биометрии.
2. Классификация признаков.
3. Дать определения основным понятиям и терминам биометрии.
4. Назовите виды изменчивости.
5. Что такое модификации? Какие они бывают?

Тема 2. СРЕДНИЕ ВЕЛИЧИНЫ И СПОСОБЫ ИХ ВЫЧИСЛЕНИЯ

Цель занятия: изучение различных видов и свойств средних величин. Получение навыков использования различных средних величин в зависимости от изучаемого признака и вида распределения.

Основные понятия

Наиболее распространенной характеристикой совокупности служит величина среднего значения признака. В отличие от индивидуальных числовых характеристик средние величины обладают большей устойчивостью, способностью характеризовать целую группу однородных единиц одним (средним) числом.

В зависимости от поставленной задачи выбирают следующие средние величины:

\bar{X}	– средняя арифметическая,
$\bar{X}_{взв}$	– средняя взвешенная,
\bar{X}_n	– средняя непараметрическая,
G	– средняя геометрическая,
H	– средняя гармоническая,
$H_{взв}$	– взвешенная средняя гармоническая,
S	– средняя квадратическая,
$S_{взв}$	– взвешенная средняя квадратическая,
Q	– средняя кубическая,
$Q_{взв}$	– взвешенная средняя кубическая,
M_o	– мода,
M_e	– медиана.

Для вычисления статистических параметров можно пользоваться традиционными способами, по нижеуказанным методикам, и современными – с использованием специальных пакетов программ для ПЭВМ. Наиболее часто используют пакет статистики MS Excel.

Свойства средних величин

Все средние величины имеют абстрактное значение и могут выражаться дробной величиной. Все средние позволяют получать среднее значение массовых варьирующих признаков, что позволяет характеризовать большой объем значений одним числом.

Средняя величина должна выбираться индивидуально, в зависимости от изучаемого объекта, так как неправильно выбранная средняя искажает величину среднего значения варьирующего признака и приводит к неверным выводам о генеральной совокупности (N).

Средние величины могут характеризовать только однородную совокупность вариантов. Если средняя получена на качественно неоднородном материале или выбрана неправильно, без учета специфики характеризуемого явления или процесса, она окажется искаженной. При наличии разнородных по составу данных их необходимо группировать в отдельные качественно однородные группы и вычислять групповые или частные средние.

Характеристика средней арифметической величины варьирующего признака. Средние арифметические величины используются наиболее часто для характеристики биологических объектов. Сопоставление средних арифметических позволяет вычислить разность между уровнями показателя в группах.

Средняя арифметическая может быть простой и взвешенной.

В *малой выборке* ($n < 30$) среднюю арифметическую вычисляют простым способом, как единство суммарного действия:

$$\bar{X} = \frac{\sum x_i}{n}, \quad (1)$$

где x_i – величина варьирующего признака;
 n – объем выборки.

$$\sum x_i = x_1 + x_2 + x_3 + \dots + x_n, \quad (2)$$

Вычисление среднего арифметического значения признаков для большой совокупности:

$$\bar{X} = A + k \frac{\sum fa}{n}, \quad (3)$$

где A – условная средняя величина;
 k – классный промежуток;
 f – частота;
 a – отклонение от нулевого класса;
 n – объем выборки.

Точность вычисления \bar{X} зависит от величины классного промежутка (k). Чем больше классов оформлено в вариационном ряду, тем точнее будет вели-

чина \bar{X} . Классный промежуток при построении вариационного ряда следует округлять до целой величины и желательно в сторону уменьшения.

Характеристика средней арифметической для альтернативных (качественных) признаков. Средней арифметической для *непараметрических или альтернативных признаков* служит показатель доли, которую составляют члены совокупности, имеющие данный альтернативный признак. Это можно выразить следующей формулой:

$$\bar{X}_n = \frac{p}{n}, \quad (4)$$

где p – число членов совокупности с наличием альтернативного признака;
 n – общее число членов выборки.

Средняя арифметическая величина для большой и малой выборок выражается в тех же единицах, что и сам варьирующий признак, который подвергается изучению, для альтернативных признаков – в процентах или долях единицы.

Характеристика средней взвешенной величины. Кроме простой средней арифметической величины (\bar{X}) иногда применяется средняя арифметическая величина, называемая взвешенной. Средняя взвешенная величина – это результат усреднения средних арифметических нескольких совокупностей. Вычисляется по формуле:

$$\bar{X}_{взв} = \frac{\bar{X}_1 \cdot n_1 + \bar{X}_2 \cdot n_2 + \dots + \bar{X}_i \cdot n_i}{n_1 + n_2 + \dots + n_i} = \frac{\sum \bar{X} \cdot n}{\sum n}, \quad (5)$$

где \bar{X}_1, \bar{X}_2 , и \bar{X}_i – средние арифметические 1-й, 2-й и i -й совокупности;
 $n_1; n_2$; и n_i – объемы этих совокупностей.

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 признаков животные могут быть ранжированы в порядке усиления или ослабления выраженности признака. Порядковый номер животного называется рангом. Средний ранг рассчитывается по формуле:

$$\bar{X} = \frac{\sum X_i}{n} \quad (6)$$

Затем сравнивают ранги оцениваемых животных и делают выводы.

Характеристика средней гармонической величины. Средняя гармони-

ческая (Н) применяется при вычислении среднего уровня признака, характеризующего скорость какого-либо процесса (средняя скорость бега, скорость передвижения рыб на нерест и т.д.), а также в случае, если признак выражен индексом (число чешуек на 1 см² поверхности кожи). Эту характеристику в отличие от средней арифметической, представляющей сумму вариантов, отнесенных к их числу, определяют как сумму обратных значений вариантов, деленных на их число. Для вычисления средней гармонической пользуются следующей формулой:

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_i}}, \quad (7)$$

где x_i – величина варьирующего признака;

n – число периодов времени.

Эту формулу используют в том случае, если периоды времени имеют одинаковую величину.

Если среднюю гармоническую вычисляют для совокупности, сгруппированной по классам, или для совокупности, у которой частные отрезки времени имеют разную величину, то пользуются формулой взвешенной средней гармонической, которая выражается следующим образом:

$$H_{\text{взв}} = \frac{p_1 + p_2 + \dots + p_n}{\frac{1}{x_1} \times p_1 + \frac{1}{x_2} \times p_2 + \dots + \frac{1}{x_n} \times p_n}, \quad (8)$$

где p – длительность периодов времени.

Величина H всегда меньше величины \bar{X} .

Характеристика средней геометрической величины. Средняя геометрическая необходима для определения среднего значения признака, если он характеризует темп роста, темп увеличения численности популяции, применяется, если признак выражен в долях единицы или в процентах и изменяется во времени и по периодам. С помощью ее можно определять средний прирост или среднее уменьшение какого-либо показателя за определенный отрезок времени. В этих случаях средняя арифметическая совершенно непригодна, так как она исказит показатель средней. Среднюю геометрическую используют для определения относительного прироста косяка рыб, привесов тела за большой период времени или за отдельный искомый.

Формула средней геометрической имеет следующий вид:

$$G = \sqrt[n]{X_1 \times X_2 \times X_3 \times \dots \times X_n}, \quad (9)$$

где x – варьирующий признак;

n – число наблюдений в выборке.

В формуле средней геометрической под корнем стоит произведение вариантов, число которых равно числу наблюдений. Корень имеет степень, соответ-

ствующую числу наблюдений. Для вычисления G необходимо произвести логарифмирование этой формулы.

Из-за сложности вычисления среднюю геометрическую на практике можно заменить средней арифметической величиной.

Характеристика средней квадратической величины. Среднее квадратическое значение варьирующего признака (S) используют в тех случаях, когда этот признак выражают площадью круга и для ее получения измеряют величину диаметра (диаметр икринок, диаметр клеточного ядра, альвеол или клеток крови):

$$S = \sqrt{\frac{\sum x_i^2}{n}}, \quad (10)$$

где x_i – величина варьирующего признака;
 n – число наблюдений в выборке.

При повторяемости отдельных вариантов используется взвешенная средняя квадратическая ($S_{взв}$):

$$S_{взв} = \sqrt{\frac{\sum n_i \times x_i^2}{\sum n_i}}, \quad (11)$$

где n_i – частота (число случаев).

Характеристика средней кубической величины. Для характеристики объемных признаков (объем икринок, эритроцитов и т.д.) более точной является средняя кубическая величина (Q), которая определяется по формуле:

$$Q = \sqrt[3]{\frac{\sum x_i^3}{n}}, \quad (12)$$

где x_i – величина варьирующего признака;
 n – число наблюдений в выборке.

При повторяемости отдельных вариантов используется взвешенная средняя кубическая ($Q_{взв}$):

$$Q_{взв} = \sqrt[3]{\frac{\sum (n_i \times x_i^3)}{\sum n_i}}, \quad (13)$$

где n_i – частота (число случаев).

Характеристика моды. Модой называется чаще всего встречающаяся величина варьирующего признака как количественного, так и качественного. В вариационных рядах, разбитых на классы по величине варьирующего признака, *модальным* будет класс, имеющий наибольшее число частот. В вариационном

ряду может быть несколько модальных классов. Мода вычисляется по формуле:

$$Mo = W_0 + k \times \left(\frac{f_1 - f_2}{2 \times f_2 - f_1 - f_3} \right), \quad (14)$$

где W_0 – нижняя граница модального класса;
 k – величина классного промежутка;
 f_1 – частота класса, предшествующего модальному;
 f_2 – частота модального класса;
 f_3 – частота класса, следующего за модальным.

Эту формулу используют для совокупностей, оформленных в виде классов варьирующего признака. В вариационных рядах, имеющих симметричное или близкое к нему размещение частот, величины моды и средней арифметической совпадают:

$$\bar{X} = Mo. \quad (15)$$

Характеристика медианы. Медианой называют середину класса, который делит вариационный ряд на две части. Одна часть имеет значение признака меньше Me , другая больше:

$$Me = W_0 + K \times \frac{\frac{n}{2} - f_1}{f_{Me}}, \quad (16)$$

где W_0 – нижняя граница класса, где находится медиана;
 K – величина классного промежутка;
 f_1 – сумма накопленных частот, предшествующих классу, где находится медиана;
 f_{Me} – частота класса, где находится медиана;
 n – объем выборки.

Используют показатель медианы чаще для характеристики качественных признаков. При определении медианы для количественных признаков при малом числе наблюдений члены выборки записываются подряд в возрастающем порядке. Срединный член такого ранжированного ряда будет служить показателем медианы.

Средние значения используют как главную характеристику любой группы особей по какому-либо фенотипическому признаку. Без среднего значения невозможно вести отбор, определять наследуемость признака, осуществлять подбор и т.д., т.е. невозможно вести племенную и селекционную работу [4,8, 9].

Примеры решения задач по определению средних величин

Вычисление среднего арифметического значения варьирующего признака в малой выборке ($n < 30$)

Пример 1. Необходимо определить средний вес тела у самок лосося, если вес отдельных особей составил (кг): 6,4; 6,8; 9,4; 9,6; 7,5; 8,6.

Небольшое поголовье ($n < 30$) условно считается малой выборкой. Вычисление простой средней арифметической проводим по формуле 1:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{6,4 + 6,8 + 9,4 + 9,6 + 7,5 + 8,6}{6} = 8,05 \text{ кг}$$

Средний вес тела у самок лосося в данной секции находится на уровне 8,05 кг.

Пример 2. Определить среднее число самцов в помете свиней, если из 2000 поросят 1200 оказались самцами. Для определения средней арифметической по альтернативным признакам используем формулу 4.

Средняя арифметическая количества самцов составит:

$$\bar{X}_{\text{альт}} = \frac{p}{n} = \frac{1200}{2000} = 0,6, \text{ или } 60 \, \%.$$

Вычисление взвешенной средней

Пример 3. Определить среднюю оплодотворяемость самок карпа в трех водоемах, если известно, что в первом водоеме в выборку попало 19 самок, во втором – 25, в третьем 27 самок, а средняя оплодотворяемость по водоемам составила соответственно 46,7; 70,0; 76,0 %.

Для решения задачи воспользуемся формулой 5:

$$\bar{X}_{\text{взв}} = \frac{\bar{X}_1 \times n_1 + \bar{X}_2 \times n_2 + \dots + \bar{X}_i \times n_i}{n_1 + n_2 + \dots + n_i} = \frac{\sum \bar{X} \times n}{\sum n}.$$

Оплодотворяемость самок в разных водоемах (%) – это усредненные данные (\bar{X}_i), а количество обследованных самок (гол) – это объемы совокупностей (n_i). Подставив данные в формулу, получим:

$$\bar{X}_{\text{взв}} = \frac{46,7 \times 19 + 70 \times 25 + 76 \times 27}{19 + 25 + 27} = 66 \, \%.$$

Для сравнения рассчитаем по формуле 1 простую среднюю арифметическую. $\bar{X} < \bar{X}_{\text{взв}}$; $\bar{X} = 64,2\%$; $\bar{X}_{\text{взв}} = 66\%$. Таким образом, средняя оплодотворяемость самок в изученных водоемах составляет 66%.

Вычисление средней величины для неизмеряемых признаков (непараметрическая средняя)

Многие признаки не имеют количественного измерения (интенсивность окраски шкурок пушных зверей). По степени интенсивности развития признака животные могут быть ранжированы в порядке усиления или ослабления окраски, при этом порядковый номер животного называется рангом.

Пример 4. От двух баранов-производителей (№5 и №6) каракульской породы и группы отобранных маток получено по 8 ягнят с различной интенсивностью окраски (от светло-серой до темной). Определить, какой из производителей даст потомство с более темной мастью.

Строим ранжированный ряд от светло-серой до темной окраски шерсти с указанием номера отца.

Ранги	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
№ отца	5	6	5	6	5	5	6	5	5	5	6	5	6	6	6

На основании полученного ряда определяют средний ранг каждого производителя, формулы 4,6:

$$\bar{X}_5 = \frac{1+3+5+6+8+9+10+12}{8} = 6,5;$$

$$\bar{X}_6 = \frac{2+4+7+11+13+14+15+16}{8} = 10,2.$$

Второй производитель имел больше ягнят с темной окраской, которая ценится дороже.

Вычисление средней гармонической величины

Пример 5. Определить среднюю скорость молокоотдачи у коровы, если за четыре минуты выдоено 9 кг молока, в том числе за первую минуту – 2 кг, за вторую, третью и четвертую минуты соответственно 3,3 и 1 кг.

При изучении скорости какого-либо процесса необходимо пользоваться формулой 7. Из условия задачи видно, что скорость молокоотдачи меняется на протяжении дойки. Количество затраченного времени берется как объем выборки ($n = 4$), а за x_i принимаем количество молока (кг) за каждую минуту дойки. Подставим значения в формулу:

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{4}{\frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{1}} = \frac{4}{0,5 + 0,33 + 0,33 + 1} = \frac{4}{2,166} = 1,85 \text{ кг/мин.}$$

Для сравнения вычислим по предложенным данным простую среднюю арифметическую ($\bar{X} = 2,25$ кг/мин). Следует помнить, что средняя арифметическая всегда больше средней гармонической ($\bar{X} > H$): $2,25 \text{ кг/мин} > 1,85 \text{ кг/мин}$.
 Ответ: Средняя скорость молокоотдачи составляет 1,85 кг/мин.

Пример 6. Определить среднюю скорость хода семги при движении из моря к нерестовой реке, если за первые сутки пути она прошла 20 км, за следующие 3 суток – по 15 км и за последние 2 суток – по 14 км.

В отличие от предыдущей, в настоящей задаче идет речь о разных промежутках времени, поэтому необходимо воспользоваться формулой 8 для определения гармонической взвешенной величины:

$$H_{\text{взв}} = \frac{p_1 + p_2 + \dots + p_n}{\frac{1}{x_1}p_1 + \frac{1}{x_2}p_2 + \dots + \frac{1}{x_n}p_n} = \frac{1+3+2}{\frac{1}{20}1 + \frac{1}{15}3 + \frac{1}{14}2} = 15,38 \text{ км/сут.}$$

Вычисление средней квадратической величины

Пример 7. Рассчитать средний диаметр семи зигот до начала их дробления, если известен диаметр каждой из них (мкм): 75; 65; 55; 70; 60; 59; 68.

Для определения диаметра шара пользуемся формулой 10:

$$S = \sqrt{\frac{\sum x_i^2}{n}}.$$

Объем выборки равен семи ($n = 7$). Подставим значения в формулу:

$$S = \sqrt{\frac{75^2 + 65^2 + 55^2 + 70^2 + 60^2 + 59^2 + 68^2}{7}} = 64,90 \text{ мкм.}$$

Для сравнения вычислим простую среднюю арифметическую величину. Если подтвердится неравенство $S > \bar{X}$, то задача решена верно. $\bar{X} = 64,57$ мкм. $64,90 \text{ мкм} > 64,57 \text{ мкм.}$

Пример 8. Определить средний диаметр ядра в клетках, если известно, что 5 клеток имели диаметр ядра 120 мкм, 3 клетки – 225 мкм и 8 клеток – 400 мкм.

Так как отдельные варианты в задании повторяются, воспользуемся формулой 11 для вычисления взвешенной средней квадратической:

$$S_{\text{взв}} = \sqrt{\frac{\sum n_i x_i^2}{\sum n_i}} = \sqrt{\frac{5 \times 120^2 + 3 \times 225^2 + 8 \times 400^2}{16}} = 306,58 \text{ мкм.}$$

Вычисление средней кубической величины

Пример 9. Определить средний объем 5 икринок щуки, если известно, что они имели следующий объем (мм³): 4,3; 5,1; 4,8; 5,2; 4,5 мм³.

Для решения задачи воспользуемся формулой 12.

$$Q = \sqrt[3]{\frac{\sum x_i^3}{n}} = \sqrt[3]{\frac{4,3^3 + 5,1^3 + 4,8^3 + 5,2^3 + 4,5^3}{5}} = 4,8 \text{ мм}^3.$$

Пример 10. Вычислить средний объем 18 наугад отобранных икринок сельди-черноспинки, если результаты измерений оказались следующими:

Объем икринок (x_i), мм ³	4,7	4,8	5,0	5,4	5,6	6,0
Количество икринок (n_i)	2	4	6	3	2	1

Определим средний объем икринок по формуле 13:

$$Q_{\text{взв}} = \sqrt[3]{\frac{\sum (n_i x_i^3)}{\sum n_i}}.$$

Предварительно вычислим:

$$\sum n_i x_i^3 = 2(4,7)^3 + 4(4,8)^3 + 6(5,0)^3 + 3(5,4)^3 + 2(5,6)^3 + 1(6,0)^3 = 2419,638.$$

$$\text{Отсюда } Q_{\text{взв}} = \sqrt[3]{\frac{2419,638}{18}} = \sqrt[3]{134,42} = 5,12 \text{ мм}^3.$$

Вычисление моды

Пример 11. Вычислить моду по показателю плодовитости свиноматок.

Класс по плодовитости	Количество поросят					K=2 L= 5
	6-7	8-9	10-11	12-13	14-15	
f	4	14	21	8	3	$\Sigma f=50$

В данном примере модальным классом является класс, имеющий частоту $f_2 = 21$. Начало этого класса, соответствующее $W_0 = 10$ голов, $f_1 = 14$, $f_2 = 21$, $f_3 = 8$, т.е. это частоты трех классов – модального и смежных с ним. Величина класса $K=2$.

Мода вычисляется по формуле 14:

$$M_o = W_0 + k \times \left(\frac{f_1 - f_2}{2 \times f_2 - f_1 - f_3} \right) = 10 + 2 \times \left(\frac{21 - 14}{2 \times 21 - 14 - 8} \right) = 10 + 0,7 = 10,7 \text{ голов}.$$

Вычисление медианы

Пример 12. Найти значение медианы для ряда распределения кустов помидор по числу плодов на каждом кусте:

№ куста	1	2	3	4	5	6	7	8	9	10	11
Число плодов	14	15	14	17	13	20	16	22	13	18	14

Для определения медианы при малом числе наблюдения располагают все варианты в возрастающем порядке. При этом данные будут расположены следующим образом:

13 13 14 14 14 **15** 16 17 18 20 22

Вариантом, выражающим медиану, будет шестой вариант.

Пример 13. Найти значение медианы для вариационного ряда по живой массе коров:

Класс по живой массе	Живая масса, кг					
	250-299	300-345	350-399	400-449	450-499	500-549
Частоты f	4	15	30	10	5	1
Накопленные частоты	4	19	49	59	64	65

Величина класса $k=50$ кг.

Ряд накопленных частот составляется путем последовательного сложения частоты каждого последующего класса с частотой предыдущего, при этом полученная сумма проставляется в каждый класс.

Медиана вычисляется по формуле 16:

$$Me = W_o + K \times \frac{\frac{n}{2} - f_1}{f_{Me}} = 350 + 50 \times \frac{\frac{65}{2} - 19}{30} = 350 + 50 \times \frac{32,5 - 19}{30} = 372,5 \text{ кг.}$$

Контрольные вопросы:

1. Перечислите средние величины.
2. Какими свойствами обладают средние величины?
3. Где используются средние величины?

Тема 3. ВАРИАЦИОННЫЙ РЯД. ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ПАРАМЕТРЫ ДЛЯ БОЛЬШИХ ВЫБОРОК. ПОСТРОЕНИЕ ГИСТОГРАММЫ И ВАРИАЦИОННОЙ КРИВОЙ

Цель занятия: изучить методику биометрической обработки данных для больших выборок, освоить способ построения вариационного ряда и определения основных показателей варьирующего признака.

Основные понятия

Распределение – это частота проявления определенных значений признака в совокупности.

Вариационный ряд – распределение объектов совокупности по классам в соответствии с величиной признака. Это изображение вариационного ряда в виде кривой, ординаты которой пропорциональны частотам вариационного ряда. По вариационному ряду можно судить не только о границах изучаемого количественного признака, но и о характере вариации.

Пример 1. **Вычисление основных показателей варьирующего признака для большой выборки ($n > 30$).**

Методика выполнения

1. Из всей выборки найти максимальную (*lim max*) и минимальную (*lim min*) величину, определить разность между ними, рассчитать число классов, которое зависит от объема выборки:

Число вариантов 25-40, 40-60, 60-100, 100-200;

Классы	Число вариант, f	a	fa	fa ²
575-599,9	26	1	26	26
600-624,9	16	2	32	64
625-649,9	9	3	27	81
650-674,9	7	4	28	112
675-699,9	5	5	25	125
	$\sum n=150$		$\sum fa=30$	$\sum fa^2=730$

7. В колонке, обозначенной буквой *f*, ставится число животных в каждом классе.

8. Определить класс условной средней (*A*). Это класс в середине вариационного ряда или близкий к нему, но с максимальным числом животных. Его обозначают 0 (ноль). Ставят это значение в колонке *a*, (*a* – отклонение от условной средней). Верхние от нуля классы нумеруют по порядку от 1 и выше с отрицательным знаком, вниз от 1 и ниже с положительным знаком. *A* – среднее значение нулевого класса (начало класса плюс половина классного промежутка), *A*=562,5.

9. Найти значение произведений *fa* в каждом классе.

10. Найти сумму произведений $\sum fa$ с учетом знака (+) или (–):

$$\sum fa = -112 + 142 = 30;$$

11. В колонке, обозначенной *fa*², найти по каждому классу произведение *fa* на *a*.

12. Найти сумму произведений – $\sum fa^2 = 730$.

13. Рассчитать основные показатели вариационного ряда:

- *средняя арифметическая* (\bar{X}) – показывает среднюю величину признака в группе; показывает, какое значение признака наиболее характерно в целом для данной совокупности. Она используется для сравнения пород, стад, линий, семейств и т. д. по какому-либо признаку.

Найти значение средней арифметической, пользуясь формулой 18:

$$\bar{X} = A + \frac{\sum fa}{n} \times k = 562,5 + 0,2 \times 25 = 567,5 \text{ кг}; \quad (18)$$

где \bar{X} – средняя арифметическая величина;

A – условный средний класс;

$\sum fa$ – сумма произведений числа животных в каждом классе и отклонений от условного среднего класса;

n – число животных в выборке;

k – классный промежуток.

- *среднее квадратическое отклонение* (σ) – показывает, в каких пределах каж-

дый член совокупности отклоняется от среднего арифметического. Чем больше сигма, тем больше изменчивость данного признака.

Определить среднее квадратическое отклонение по формуле 19:

$$\sigma = \pm \sqrt{\frac{\sum fa^2}{n} - \left(\frac{\sum fa}{n} \right)^2} \times k = \pm \sqrt{\frac{730}{150} - \left(\frac{30}{150} \right)^2} \times 25_{к2} = 55_{к2}, \quad (19)$$

где σ – среднее квадратическое отклонение;

$\sum fa^2$ – сумма произведений числа животных в каждом классе и квадрат отклонений от условного среднего класса;

$\sum fa$ – сумма произведений числа животных в каждом классе и отклонений от условного среднего класса;

n – число животных в выборке;

k – классный промежуток.

- *коэффициент вариации* (Cv) – среднее квадратическое отклонение, выраженное в процентах от средней арифметической, используется при сравнении разных признаков, чтобы показать, изменчивость какого признака выше. Определить коэффициент вариации (Cv) по формуле 20:

$$Cv = \frac{\sigma}{\bar{X}} \cdot 100\% = \frac{55}{567,7} \cdot 100\% = 9,69\%, \quad (20)$$

где σ – среднее квадратическое отклонение;

\bar{X} – средняя арифметическая величина.

- *ошибка средней арифметической* (m) – показывает, в каких пределах средняя арифметическая в данной выборке отклоняется от средней арифметической генеральной совокупности;

Определить ошибку средней арифметической по формуле 21:

$$m = \pm \frac{\sigma}{\sqrt{n}} = \pm \frac{55}{\sqrt{150n}} = 4,5_{к2}; \quad (21)$$

где m – ошибка средней арифметической величины;

σ – среднее квадратическое отклонение;

n – число животных в выборке.

Построение гистограммы

Пример 2. Осуществить графическое изображение вариационного ряда и проанализировать характер распределения вариантов.

Вариационный ряд по удою :

Классы (W), кг	13	15	17	19	21	23	25	27	29	31
Частоты (f) гол.	3	6	10	15	24	19	14	6	2	1

1. На горизонтальной оси откладываются классы, а на вертикальной оси – число животных.
2. Количество животных в каждом классе изображается в виде столбиков с основанием, равным величине классного промежутка и высотой, соответствующей числу животных в каждом классе. Получается ступенчатая кривая, или **гистограмма**.
3. Из середины каждого класса восстанавливается перпендикуляр высотой, равной числу животных. Вершины перпендикуляров соединяются прямыми линиями, получается линейная кривая.
4. Анализируется характер распределения вариантов (биномиальное, эксцесс, асимметрия, многовершинность, трансгрессия).

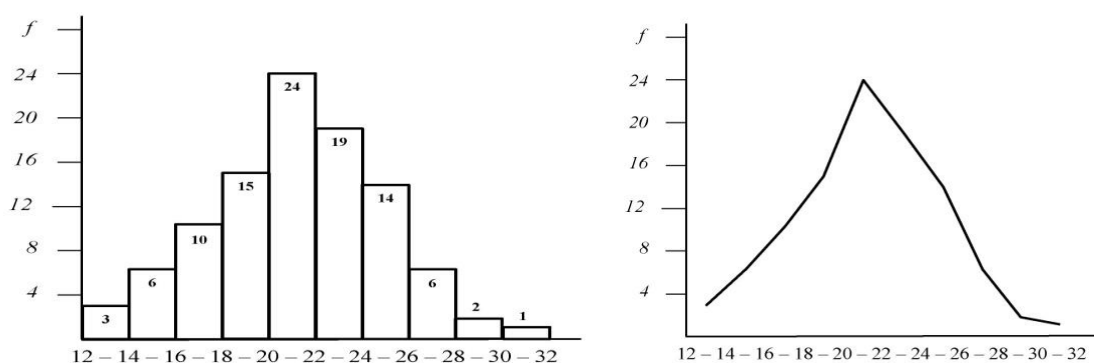


Рисунок 1 – Графическое изображение вариационного ряда по удою (гистограмма и линейная кривая)

Типы вариационных кривых

1. *Нормальное (биномиальное) распределение.* Наиболее распространенный тип распределения особей совокупности. Основное варьирование признака ограничивается лимитом, составляющим $\pm 3\sigma$ от среднего значения признака x . В эти границы входит 99,7% всех членов совокупности. Особей с величиной признака выше $+3\sigma$ или меньше -3σ встречается 0,3%.

Биномиальная кривая – это теоретическая кривая, построенная по коэффициентам бинома Ньютона при бесконечно большом числе вариантов. При этом в средних классах наблюдается повышенное число вариантов. Для данного вида распределения характерно симметричное распределение вариантов по обе стороны от средних величин.

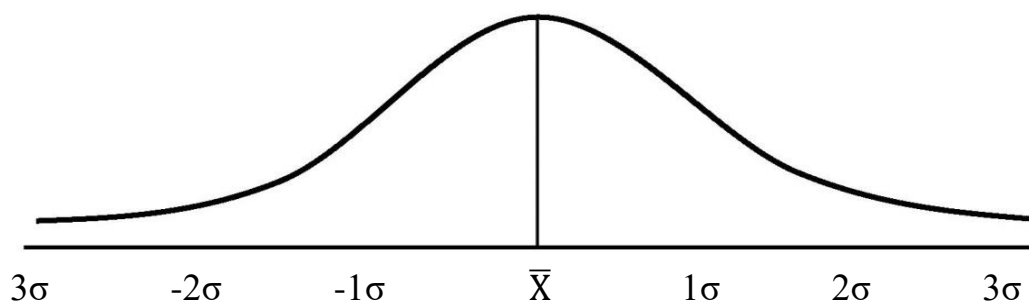


Рисунок 2 – Биномиальное распределение

2. *Асимметрия*. Это смещение вершины вправо или влево. Причинами асимметричного распределения являются резкая неоднородность условий развития отобранных для изучения особей, преобладание в отобранной группе особей с лучшими или худшими наследственными задатками, различная степень реакции особей на разные условия, отбор. Асимметрия может быть положительной и отрицательной. Положительная асимметрия или правосторонняя наблюдается, когда средняя арифметическая \bar{X} лежит правее моды M_o , отрицательная или левосторонняя – когда средняя арифметическая \bar{X} лежит левее моды M_o .

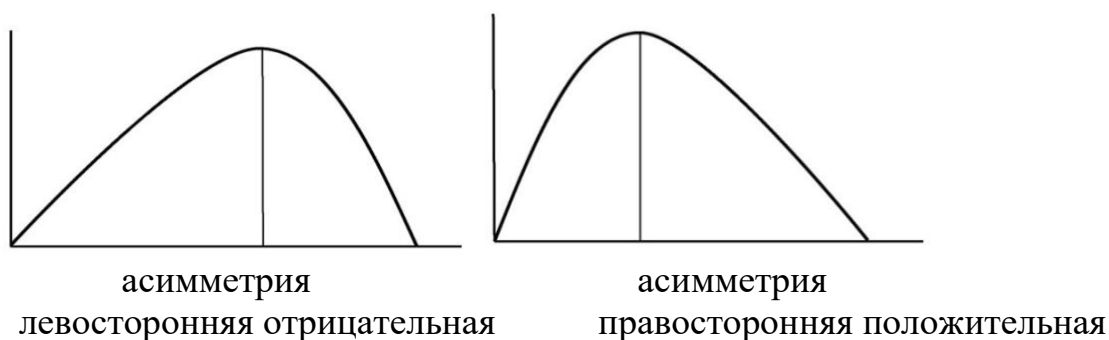


Рисунок 3 – Биномиальное распределение

3. *Эксцесс*. Это островершинная кривая с резким преобладанием вариантов в средних классах. Причинами возникновения эксцесса являются реакция организма на условия среды, влияние наследственности. Эксцесс характеризуется скоплением большого числа частот около тех значений варьирующего признака, которые близки к значениям средней арифметической, моды и медианы. Эксцессивная кривая имеет сильно вытянутую узкую вершинную часть и удлиненные пологие боковые ветви.

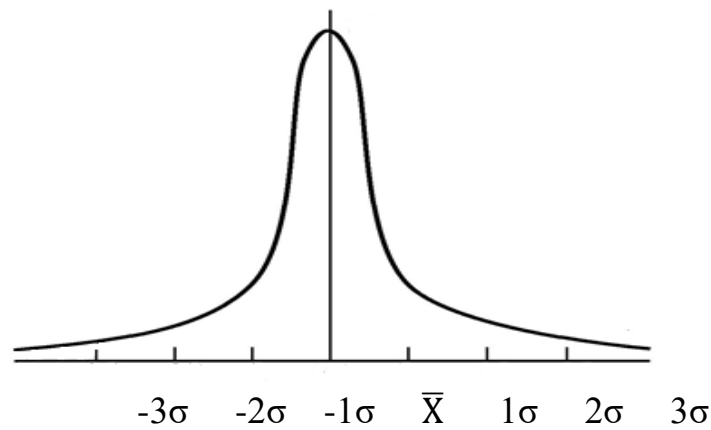


Рисунок 4 – Эксцесс

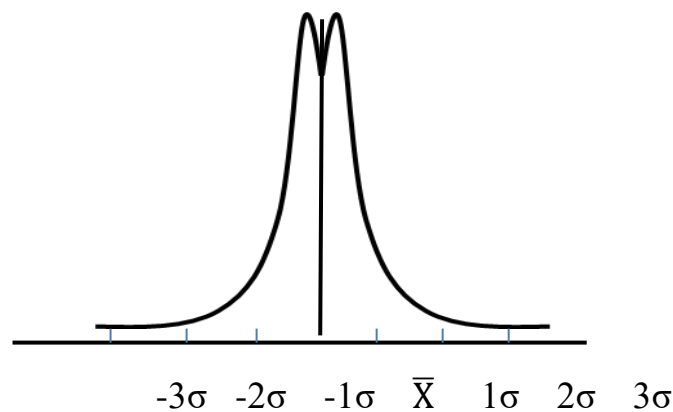


Рисунок 5 – Эксцесс отрицательный

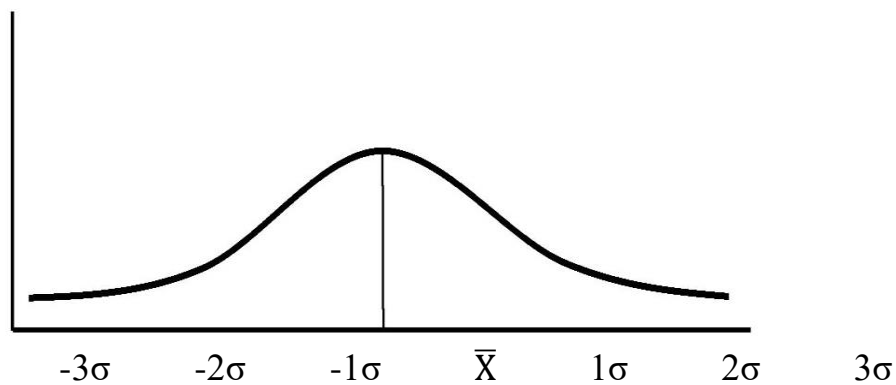


Рисунок 6 – Эксцесс плосковершинный

4. *Трансгрессия.* Это захождение вариационных рядов друг за друга. При трансгрессии часть животных имеет одинаковые классы. При большой трансгрессии, т.е. сильном захождении одного ряда за другой, общая вариационная кривая окажется близкой к биномиальной, отличаясь от нее несколько более тупой вершиной. При малой трансгрессии при смешении двух групп она получится двухвершинной, при смешении трех и более – многовершинной. Часть классов при трансгрессии общая.

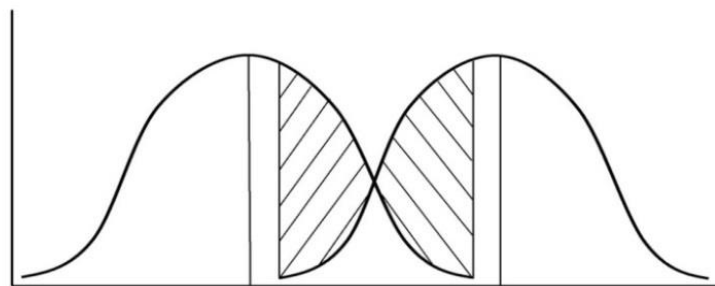


Рисунок 7 – Трансгрессия

5. *Двухвершинная и многовершинная кривая.* Такие виды кривых встречаются, когда совокупность состоит из разнородного материала. Причиной этого может быть появление объектов с новыми свойствами.



Рисунок 8 – Двухвершинная кривая

6. *Распределение Пуассона.* Относится к числу появления редких (маловероятных) событий – появление уродов, рождение монозиготных близнецов у одноплодных животных и т.д. Для него характерно, что варьирующий признак x может принимать только целые значения. В распределении Пуассона среднее значение признака x совпадает с величиной дисперсии σ^2 , таким образом, для него характерно наличие одного параметра \bar{X} , тогда как нормальное распределение имеет два параметра, полностью его характеризующих – \bar{X} и σ . При графическом изображении распределение Пуассона имеет резко асимметричную форму [6, 8].

Контрольные вопросы:

1. Методика построения вариационного ряда.
2. Типы вариационных кривых.
3. Основные показатели вариационного ряда, их свойства.

Тема 4. ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ ДЛЯ МАЛЫХ ВЫБОРОК

Цель занятия: научиться определять статистические параметры варьирующего признака в малых выборках.

Методика выполнения

Пример 1. Вычисление основных показателей варьирующего признака для малой выборки ($n < 30$).

1. Переписать индивидуальное задание в таблицу в виде одного столбика цифр:

Живая масса поросят, кг	Отклонения	Квадраты отклонений
x	$x - \bar{X}$	$(x - \bar{X})^2$
1,2	-0,15	0,0225
1,5	+0,15	0,0225
1,1	-0,25	0,0625
1,3	-0,05	0,0225
1,4	+0,05	0,0225
1,3	-0,05	0,0225
1,4	+0,05	0,0225
1,4	+0,05	0,0225
1,3	-0,05	0,0225
1,6	+0,25	0,0625
$\sum x = 13,5$	$(x - \bar{X}) = 0$	$\sum (x - \bar{X})^2 = 0,1850$

2. Найти среднее арифметическое значение признака в группе по формуле 22:

$$\bar{X} = \frac{\sum x}{n}, \quad (22)$$

где \sum – знак суммы;
 x – значение вариант;
 n – число животных.

$$\bar{X} = \frac{\sum x}{n} = \frac{13,5}{10} = 1,35 \text{ кг.}$$

3. Найти сумму квадратов отклонений каждого варианта от средней арифметической группы – $(x - \bar{X})^2$.

4. Найти среднее квадратическое отклонение в данной группе по формуле:

$$\sigma = \pm \sqrt{\frac{\sum (x - \bar{X})^2}{n - 1}}, \quad (23)$$

где σ – среднее квадратическое отклонение;
 x – значение вариант;
 \bar{X} – средняя арифметическая величина;
 n – число животных в выборке.

$$\sigma = \pm \sqrt{\frac{\sum (x - \bar{X})^2}{n - 1}} = \sqrt{\frac{0,1850}{9}} = 0,14_{\text{кг}}.$$

5. Найти коэффициент вариации по формуле 24:

$$C_v = \frac{\sigma}{\bar{X}} \cdot 100\% = \frac{0,14}{1,35} \cdot 100\% = 10,3\%.$$

6. Найти ошибку средней арифметической для малых выборок по формуле 25:

$$m = \pm \frac{\sigma}{\sqrt{n - 1}}, \quad (25)$$

где m – ошибка средней арифметической величины;
 σ – среднее квадратическое отклонение;
 n – число животных в выборке.

$$m = \pm \frac{\sigma}{\sqrt{n - 1}} = \frac{0,14}{\sqrt{9}} = 0,05_{\text{кг}}.$$

Пример 2. Определить крайние значения признака, пользуясь правилом «плюс-минус три сигмы».

Лимиты (\lim_{\max} , \lim_{\min}) – это максимальные и минимальные значения варьирующего признака, они вычисляются в абсолютных величинах. Вся изменчивость признака лежит от средней арифметической в пределах $\pm 3\sigma$ ($\bar{X} \pm 3\sigma$). В эти границы входит 99,7 % всех особей. Средняя арифметическая, увеличенная или уменьшенная на три сигмы, дает крайние значения признака (*lim min и lim max*) [1, 2, 7, 11].

Методика выполнения

Используя данные, рассчитанные при выполнении примера 1 из темы 3 по живой массе коров (средняя арифметическая – $\bar{X}=567,5$ кг и среднее квадратическое отклонение – $\sigma=55$ кг), вычислить крайние значения признака при нормальном распределении вариант (формулы 26, 27):

$$\lim \max = \bar{X} + 3\sigma, \quad (26)$$

$$\lim \min = \bar{X} - 3\sigma, \quad (27)$$

$$\lim \max = \bar{X} + 3\sigma = 567,5 + 3 \times 55 = 732,5 \text{ кг.}$$

$$\lim \min = \bar{X} - 3\sigma = 567,5 - 3 \times 55 = 402,5 \text{ кг.}$$

Контрольные вопросы:

1. Какие существуют выборки при статистическом анализе данных?
2. Какие параметры характеризуют малые выборки, их значение?
3. Привести формулы лимитов, средней арифметической, среднего квадратического отклонения, ошибки средней арифметической и коэффициента вариации для малых выборок.

Тема 5. ИЗУЧЕНИЕ СВЯЗИ МЕЖДУ ПРИЗНАКАМИ. ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ДЛЯ БОЛЬШИХ И МАЛЫХ ВЫБОРОК ПО КОЛИЧЕСТВЕННЫМ ПРИЗНАКАМ

Цель занятия: дать характеристику различных типов корреляционной связи, изучить свойства коэффициентов корреляции. Овладеть методом построения корреляционной решетки при вычислении коэффициента корреляции для больших выборок и методикой определения коэффициента корреляции для малых выборок.

Основные понятия

Изучение совокупности животных одновременно по нескольким признакам показывает, что между признаками существует взаимная связь. Взаимная связь признаков в их изменении называется *корреляция*.

Корреляция по форме бывает:

- *положительная* – увеличение или уменьшение одного признака сопровождается увеличением или уменьшением сопряженного признака;
- *отрицательная* – увеличение одного признака сопровождается уменьшением сопряженного признака и наоборот;
- *криволинейная* – с увеличением одного признака сопряженный признак сначала увеличивается, а затем уменьшается, или наоборот, или когда с увеличением одного признака другой увеличивается до определенного момента, а затем уменьшается. При криволинейной корреляции связь сначала положительная, а затем отрицательная;
- *прямолинейная* – когда равномерному изменению одного признака соответствует равномерное изменение другого.

Коэффициент корреляции (r) – количественная характеристика величины и направления корреляционной связи. Коэффициент корреляции является относительной величиной. Он показывает величину связи между двумя или несколькими признаками. Его величина изменяется в пределах от 0 до ± 1 . Чем ближе показатель к единице, тем больше связь между коррелирующими признаками.

Степень корреляционной связи может быть:

- 1) *слабая* – от 0 до 0,3;
- 2) *средняя* – от 0,3 до 0,7;
- 3) *сильная* – от 0,7 до 1,0.

При вычислении коэффициента фенотипической корреляции для многочисленных выборок наиболее часто используют следующую формулу:

$$r = \pm \frac{\frac{\sum f a_x a_y}{n} - \frac{\sum f a_x}{n} \times \frac{\sum f a_y}{n}}{\sigma'_x \times \sigma'_y}; \quad (28)$$

где \sum – знак суммы;
 $f a_x a_y$ – произведение числа животных на соответствующее значение a_x и a_y по каждому из квадратов;
 $f a_x; f a_y$ – произведение числа животных в каждом классе и отклонений от условного среднего класса;
 $\sigma'_x \sigma'_y$ – средние квадратические отклонения для каждого изучаемого признака, взятые без классного промежутка;
 n – число животных в группе.

Пример 1. Вычислить коэффициент корреляции между суточными удоями и живой массой коров. В хозяйстве изучено 100 коров ($n = 100$).

Методика выполнения

Данные по суточным удоям и живой массе заносим в таблицу:

Суточный удой (x) и живая масса (y) коров, кг

x	y	x	y	x	y	x	y	x	y	x	y
28,8	512	12,3	380	31,2	560	15,2	396	29,0	521	22,8	465
20,2	472	21,4	465	23,9	459	23,4	469	20,7	456	21,1	456
21,4	489	18,9	485	27,0	548	24,8	521	17,5	438	23,1	501
20,6	482	21,8	458	20,9	457	23,4	451	22,3	462	20,2	459
23,7	468	20,9	413	25,9	517	16,0	445	27,0	507	15,2	381
21,0	479	21,9	428	27,8	531	23,0	458	20,9	450	20,5	466
25,5	515	17,8	447	14,5	426	24,3	524	21,6	474	23,4	461
21,7	451	20,0	412	27,6	495	19,6	487	25,1	420	14,2	543
20,9	475	21,1	560	23,8	453	15,5	416	22,1	456	20,5	462

x	y	x	y	x	y	x	y	x	y	x	y
14,8	402	27,5	542	25,7	527	21,6	418	20,4	478	20,9	453
20,7	473	21,8	468	26,4	500	14,2	393	16,4	437	24,6	512
21,0	467	14,8	502	15,6	531	20,1	455	22,3	454	19,4	472
23,5	458	21,1	487	20,1	410	21,4	462	23,2	464	21,2	473
26,2	534	18,1	476	24,9	379	15,7	407	21,7	485	21,4	428
16,3	433	25,2	525	21,8	469	21,1	455	22,5	459	21,8	480
24,4	528	21,4	481	26,3	545	20,4	482	20,8	483	20,2	419
20,3	452	20,7	464	22,6	450	22,8	455				

1. В выборке случайно отобранных животных, характеризующихся по двум признакам, найти максимальное и минимальное значение каждого признака в отдельности и определить число классов по одному признаку (условно обозначенному X) и второму (обозначенному Y). Количество классов не должно различаться больше, чем на 1-2 класса.

Величина классового промежутка по первому признаку будет:

$$k_x = \frac{31,2-12,3}{10} = 1,89 \approx 2.$$

Величина классового промежутка по второму признаку

$$k_y = \frac{560-380}{10} = 18 \approx 20.$$

2. Построить классы (вариационные ряды) по признакам X и Y .

Строится корреляционная решетка. В верхнюю строку решетки вписываются классы одного из признаков X (удой), а с левой стороны – классы второго признака Y (живая масса).

3. Затем проводится разноска животных по ячейкам корреляционной решетки с учетом обоих признаков. Например, первая корова имеет удой 28,8 кг и массу 512 кг. По удою она должна быть отнесена в класс 28-29, а по живой массе – в класс 510-529. Животное с этими показателями помещают в ячейку, находящуюся на пересечении указанных классов по удою и живой массы.

Результат разnosки 100 коров показан в таблице. Расположение частот в корреляционной решетке указывает на положительную связь между суточным удоем и живой массой.

**Расчет коэффициента корреляции между суточными удоями
и живой массой коров**

$x \backslash y$	12-13,9	14-15,9	16-17,9	18-19,9	20-21,9	22-23,9	24-25,9	26-27,9	28-29,9	30-31,9	f	a_y	fa_y	fa_y^2
370-389	1	1					1				3	-4	-12	48
390-409		4		I					II		4	-3	-12	36
410-429		2	1		7		1				11	-2	-22	44
430-449			5								5	-1	-5	5
450-469					19	16					0	0	0	0
470-489				4	15						19	+1	+19	1
490-509		1		II I		1		3			5	+2	+10	20
510-529							6		2	IV	8	+3	+24	72
530-559		2						5			7	+4	+28	112
550-569					1					1	1	+5	+5	25
f	1	10	6	4	43	17	8	8	2	1	100			
a_x	-4	-3	-2	-1	0	+1	+2	+3	+4	+5				
fa_x	-4	-30	-12	-4	0	+17	+16	+24	+8	+5				
fa_x^2	16	90	24	4	0	17	32	72	32	25				

4. Закончив разnosку, нужно подсчитать в ячейках корреляционной решетки частоты. Затем выбирается условный средний класс по первому и второму признакам. В качестве условного среднего берут тот класс, в который входит наибольшее число вариантов. В данном примере условные средние классы – 20-21,9 по удою и 450-469 по живой массе. Перечеркнув эти классы, корреляционную решетку делят на четыре квадрата, обозначенных римскими цифрами (I, II, III, IV).

5. Затем выполняют обычные вычисления по каждому вариационному ряду по-рознь тем же способом, как при вычислении сигмы, f_x – частоты вариационного ряда по удою; f_y – частоты вариационного ряда по живой массе; a_x и a_y – отклонения от условного среднего класса.

Суммируем с учетом знака значения $f_x a_x$ по удою и $f_y a_y$ по живой массе: $\sum f_x a_x = +70 - 50 = +20$; $\sum f_y a_y = +87 - 51 = 36$.

Суммируем значения fa^2 для каждого из рядов $\sum f_x a_x^2 = 312$, $\sum f_y a_y^2 = 382$.

6. Рассчитываем σ'_x , σ'_y

$$\sigma'_x = \pm \sqrt{\frac{\sum f_x a_x^2}{n} - \left(\frac{\sum f_x a_x}{n} \right)^2}; \quad (29)$$

$$\sigma'_y = \pm \sqrt{\frac{\sum f_y a_y^2}{n} - \left(\frac{\sum f_y a_y}{n} \right)^2}; \quad (30)$$

где σ'_x , σ'_y – средние квадратические отклонения для каждого изучаемого признака (берутся без классных промежутков);
 \sum – знак суммы;
 $f_x a_x$, $f_y a_y$ – произведение числа животных в каждом классе и отклонений от условного среднего класса, для каждого признака;
 $f_x a_x^2$, $f_y a_y^2$ – сумма произведений числа животных в каждом классе и квадрат отклонений от условного среднего класса для каждого признака;
 n – число животных в группе.

$$\sigma'_x = \pm \sqrt{\frac{312}{100} - \left(\frac{20}{100} \right)^2} = \sqrt{3,12 - 0,04} = 1,75;$$

$$\sigma'_y = \pm \sqrt{\frac{382}{100} - \left(\frac{36}{100} \right)^2} = \sqrt{3,82 - 0,13} = 1,92;$$

7. Чтобы вычислить коэффициент корреляции (по формуле 28), необходимо знать $\sum f a_x a_y$, где f – число животных в одной клетке решетки; a_x – отклонение от условного среднего класса по молочности; a_y – отклонение от условного среднего класса по живой массе.

Вычисления производят отдельно по каждому из четырех квадратов. Частоту f в ячейке корреляционной решетки умножают на отклонение a_y и на отклонение a_x . Если в клетках частоты отсутствуют, вычислений не производят.

После умножения результаты суммируют и получают $\sum f a_x a_y$ по каждому квадрату.

I квадрат	II квадрат	III квадрат	IV квадрат
$1 \cdot (-4) \cdot (-4) = +16$	$1 \cdot (+2) \cdot (-4) = -8$	$4 \cdot (-1) \cdot (+1) = -4$	$1 \cdot (+1) \cdot (+2) = +2$
$1 \cdot (-3) \cdot (-4) = +12$	$1 \cdot (+2) \cdot (-2) = -4$	$2 \cdot (-3) \cdot (+2) = -6$	$6 \cdot (+2) \cdot (+3) = +36$
$4 \cdot (-3) \cdot (-3) = +36$		$2 \cdot (-3) \cdot (+4) = -24$	$2 \cdot (+4) \cdot (+3) = +24$
$2 \cdot (-3) \cdot (-2) = +12$	$\sum f a_x a_y = -12$	$\sum f a_x a_y = -34$	$5 \cdot (+3) \cdot (+4) = +60$
$1 \cdot (-2) \cdot (-2) = +4$			$1 \cdot (+5) \cdot (+5) = +25$
$5 \cdot (-2) \cdot (-1) = +10$			

$$\overline{\sum f a_x a_y = +90}$$

$$\sum f a_x a_y = +147$$

$$\sum f a_x a_y = 90 + 147 + -12 - 34 = +191$$

8. Подставляем значения $\sum f a_x a_y$, $f_x a_x$, $f_y a_y$, σ_x' , σ_y' в формулу 28:

$$r = \pm \frac{\frac{191}{100} - \frac{20}{100} \cdot \frac{36}{100}}{1,75 \cdot 1,92} = \frac{1,91 - 0,2 \cdot 0,36}{3,36} = \frac{1,91 - 0,072}{3,36} = 0,54;$$

Связь при $r = 0,5$ и выше считается значительной, при $r = 0,3 - 0,49$ – средней, $r < 0,3$ – малой.

Вычисленное значение коэффициента корреляции показывает, что между суточным удоем и живой массой коров существует значительная положительная связь. При отборе более крупных коров удои в стаде будут повышаться.

9. Рассчитать ошибку коэффициента корреляции (m_r) по формуле 31:

$$m_r = \frac{1 - r^2}{\sqrt{n}}, \quad (31)$$

где r – коэффициент корреляции;
 n – число животных в группе.

$$m_r = \frac{1 - r^2}{\sqrt{n}} = \frac{1 - 0,54^2}{\sqrt{100}} = \frac{0,708}{10} = 0,07.$$

Пример 2. Рассчитать коэффициент корреляции для малых выборок и определить величину, направление и степень связи между возрастом (x) и плодовитостью (y) 10 свиноматок белорусской мясной породы. В данном примере возраст выражен порядковым номером опороса.

Методика выполнения

1. Данные систематизируем и заполняем статистическую таблицу.

**Расчет коэффициента корреляции между возрастом
и плодовитостью свиноматок**

Признаки		xy	x^2	y^2	$d=x-y$	d^2
возраст x	число поросят y					
2	9	18	4	81	-7	49
1	7	7	1	49	-6	36
5	11	55	25	121	-6	36
7	10	70	49	100	-3	9
3	11	33	9	121	-8	64
2	8	16	4	64	-6	36
6	11	66	36	121	-5	25
1	6	6	1	36	-5	25
4	12	48	16	144	-8	64
3	14	42	9	196	-11	121
$\sum x=34$	$\sum y=99$	$\sum_{y=361} x$	$\sum_{^2=154} x$	$\sum_{y^2=1033}$	$d=-65$	$\sum_{^2=465} d$

2. Рассчитать сумму $\sum x$ и $\sum y$ по каждому признаку:
3. Найти произведение xy и вычислить сумму $\sum xy$.
4. Найти x^2 и y^2 по каждому признаку, определить $\sum x^2$ и $\sum y^2$.
5. Рассчитать $d=x-y$ и суммировать.
6. Отклонения d возвести в квадрат и суммировать.
7. Рассчитать дисперсию по первому признаку (C_x) и по второму (C_y) по формулам 32, 33:

$$C_x = \sum x^2 - \frac{(\sum x)^2}{n} = 154 - \frac{34^2}{10} = 154 - \frac{1156}{10} = 154 - 115,6 = 38,4; \quad (32)$$

$$C_y = \sum y^2 - \frac{(\sum y)^2}{n} = 1033 - \frac{99^2}{10} = 1033 - \frac{9801}{10} = 1033 - 980,1 = 52,9; \quad (33)$$

где C_x – сумма квадратов отклонений (дисперсия) по первому признаку;
 C_y – сумма квадратов отклонений (дисперсия) по второму признаку;
 \sum – знак суммы;
 x – значение первого признака (x);
 y – значение второго признака (y);
 n – число животных в группе.

8. Определить коэффициент корреляции для малых выборок по формуле 34:

$$r = \frac{\sum_{xy} - \frac{\sum x \times \sum y}{n}}{\sqrt{C_x C_y}} = \frac{361 - \frac{34 \times 99}{10}}{\sqrt{38,4 \cdot 52,9}} = \frac{361 - \frac{3366}{10}}{\sqrt{2031,36}} = \frac{361 - 336,6}{45,07} = +0,541. \quad (34)$$

9. Определить ошибку коэффициента корреляции по формуле 35:

$$m_r = \frac{1 - r^2}{\sqrt{n - 2}} = \frac{1 - 0,29}{\sqrt{8}} = \frac{0,71}{2,82} = 0,25, \quad (35)$$

где r – коэффициент корреляции;
 n – число животных в группе.

Коэффициент корреляции ($r = +0,541$) указывает на среднюю по силе и положительную по направлению связь между двумя изучаемыми признаками. Это свидетельствует о том, что при селекции свиноматок по возрасту будет наблюдаться повышение многоплодия [5, 6, 8, 11, 12].

Контрольные вопросы:

1. Понятие корреляции и примеры коррелятивной изменчивости.
2. Направление и степень корреляционной связи. Значение знака абсолютной величины коэффициента корреляции.
3. Методика вычисления коэффициента корреляции для больших и малых выборок.

Тема 6. ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ДЛЯ АЛЬТЕРНАТИВНЫХ ПРИЗНАКОВ

Цель занятия: научиться вычислять коэффициент корреляции для альтернативных признаков.

Основные понятия

Альтернативные признаки – это два взаимоисключающих варианта, например, пол животных – мужской или женский, скот – комолый или рогатый, состояние животных – здоровые или больные.

Корреляция между альтернативными признаками измеряется **тетрахорическим показателем связи** (r_a).

При изучении у каждой особи двух альтернативных признаков группа разбивается на 4 части:

- a** – особи, имеющие оба признака (+ +),
- b** – особи, имеющие первый признак, но не имеющие второго (+ –),
- c** – особи, не имеющие первого признака, но имеющие второй (– +),
- d** – особи, не имеющие обоих признаков (– –).

Пример. Изучить частоту заболеваемости лейкозом матерей-коров и их дочерей и рассчитать коэффициент корреляции для альтернативных признаков.

Методика выполнения

1. Для установления связи между альтернативными признаками первичные данные по каждому члену выборки помещают в 4-х клеточную решетку.

Дочери (y)	Матери (x)		Всего
	больные	здоровые	
Больные	a = 46	b = 70	a + b = 116
Здоровые	c = 114	d = 391	c + d = 505
Всего	a + c = 160	b + d = 461	n = a + b + c + d = 621

2. Вычисление проводят по формуле 36:

$$r_a = \frac{(ad - bc) - \frac{n}{2}}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{(46 \times 391 - 70 \times 114) - \frac{621}{2}}{\sqrt{(46+70)(114+391)(46+114)(70+391)}} = 0,147, \quad (36)$$

где a – особи, имеющие оба признака;
 b – особи, имеющие первый признак, но не имеющие второго;
 c – особи, не имеющие первого признака, но имеющие второй;
 d – особи, не имеющие обоих признаков;
 n – число животных в группе.

3. Определить ошибку коэффициента корреляции для альтернативных признаков (m_r) по формуле 37:

$$m_r = \frac{1 - r_a^2}{\sqrt{n}} = \frac{1 - 0,147^2}{\sqrt{621}} = 0,039, \quad (37)$$

где r – коэффициент корреляции;
 n – число животных в группе.

Таким образом, можно сделать вывод, что у больных матерей дочери тоже чаще болевают лейкозом, существует невысокая положительная связь [4, 7, 8].

Контрольные вопросы:

1. Какие признаки считаются альтернативными?
2. Как вычисляется коэффициент корреляции для альтернативных признаков?
3. Имеет ли смысловое значение знак при коэффициенте корреляции между альтернативными признаками?

Тема 7. РАНГОВЫЙ КОЭФФИЦИЕНТ СВЯЗИ ПО СПИРМЕНУ

Цель занятия: научиться вычислять ранговый коэффициент корреляции Спирмена.

Основные понятия

Ранговый коэффициент связи по Спирмену (r_s) используется при биометрической обработке признаков, которые нельзя измерить ни точно, ни грубо, поэтому их выражают порядковым местом (рангом). Ранг – это порядковый номер по величине признака. Величина рангового коэффициента связи изменяется от -1 до +1.

Пример 1. Установить имеется ли связь между ростом (x) рысистых лошадей и скоростью их бега (y) на дистанцию 1600 м. Сравниваются ранги 5 лошадей по этим двум признакам.

Методика выполнения

1. Первичные данные по каждому члену выборки разнести в таблицу, соблюдая принцип ранжирования по возрастающей (или убывающей) степени выраженности признака – x , определить ранг по признаку – y .

Номер лошади	Ранги по x	Ранги по y	Разность рангов ($x - y$)	$(x - y)^2$
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
				$\sum(x - y)^2 = 36$

2. Если в изучаемой группе два и более животных имеют одинаковую выраженность признака, то им дается средний ранг, определяемый как средняя арифметическая величина.

3. Найти разницу между рангами по признакам x и y , сохраняя знаки + и -.

4. Возвести в квадрат разницу между рангами по каждому члену выборки $(x - y)^2$ и суммировать.

5. Вычислить ранговый коэффициент корреляции Спирмена (r_s) по формуле 38:

$$r_s = 1 - \frac{6 \times \sum (x - y)^2}{n \times (n^2 - 1)} = 1 - \frac{6 \times 36}{5(5^2 - 1)} = 1 - \frac{216}{120} = 1 - 1,8 = -0,80. \quad (38)$$

где $(x - y)$ – разность между рангами по изучаемым признакам;

n – число сопоставляемых пар признаков.

Таким образом, между ростом лошадей и скоростью их бега имеется

связь, у лошадей более высокого роста в среднем скорость бега выше, чем у лошадей низкого роста.

Пример 2. Выяснить, имеется ли связь между агрессивностью норок и степенью опушенности меха. Отобрано 6 норок, которые оценены по обоим признакам и распределены по рангам: агрессивность от неагрессивного до очень агрессивного, а степень опушенности от худшей до лучшей.

Методика выполнения

1. Первичные данные по каждому члену выборки разнести в таблицу, соблюдая принцип ранжирования по возрастающей (или убывающей) степени выраженности признака – x , определить ранг по признаку – y .

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

2. Заполнить таблицу, найти разницу между рангами по каждому признаку, возвести ее в квадрат, суммировать. Полученные данные подставить в формулу 43 и вычислить ранговый коэффициент корреляции Спирмена (r_s):

$$r_s = 1 - \frac{6 \times \sum (x - y)^2}{n \times (n^2 - 1)} = 1 - \frac{6 \times 12}{6(6^2 - 1)} = 1 - \frac{72}{6 \times 35} = 1 - 0,3428 = +0,658.$$

Как видно из расчетов, установлена положительная связь между агрессивностью норок и густотой меха [2, 4, 7, 8].

Контрольные вопросы:

1. Особенности коэффициента корреляции Спирмена.
2. Принципы ранжировки объектов совокупности.
3. Форма записи объектов совокупности.

Тема 8. РЕГРЕССИВНЫЙ АНАЛИЗ. ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТА ПРЯМОЛИНЕЙНОЙ РЕГРЕССИИ

Цель занятия: изучить методику регрессивного анализа.

Основные понятия

Регрессивный анализ позволяет установить связь между двумя варьирующими признаками. Определить, как количественно меняется одна величина при изменении другой. Регрессия является двухсторонней, т.е. можно определить изменение x по изменению y и изменение y по изменению x .

Регрессия между признаками может быть выражена через уравнение регрессии и в виде графика.

Коэффициент регрессии ($R_{x/y}$) показывает, насколько в среднем изменяется один признак при изменении другого на единицу измерения.

Коэффициент регрессии применяется при планировании и прогнозировании уровня того или иного признака по заданному уровню другого признака. Регрессия между признаками может быть выражена в виде эмпирического и теоретического рядов регрессии, в виде графика, а также через уравнения регрессии.

С помощью уравнения прямолинейной регрессии можно определить величину одного признака при заданной величине другого признака (формулы 39-42):

$$X = \bar{X}_x + R_{x/y} \times (Y - \bar{Y}_y), \quad (39)$$

$$X = \bar{X}_x + r \times \frac{\sigma_x}{\sigma_y} (Y - \bar{Y}_y), \quad (40)$$

Если зависимым признаком будет Y , а независимым – X , то в уравнении произойдет замена мест между значениями признаков, а именно:

$$Y = \bar{Y}_y + R_{y/x} \times (X - \bar{X}_x), \quad (41)$$

$$Y = \bar{Y}_y + r \frac{\sigma_y}{\sigma_x} \times (X - \bar{X}_x), \quad (42)$$

где X – искомая величина признака x ;
 Y – заданная величина признака y ;
 \bar{X}_x и \bar{X}_y – средние арифметические признаков;
 R – коэффициент регрессии;
 r – коэффициент корреляции;
 σ – среднее квадратическое отклонение.

Пример 1. При изучении взаимосвязи между обхватом груди (x) и живой массой (y) у лошадей установлено: $r = +0,89$; $\sigma_x = 7,9$ см; $\sigma_y = 56,8$ кг. Рассчитать коэффициенты регрессии.

Методика выполнения

Рассчитайте коэффициенты регрессии по формулам 43, 44:

$$R_{x/y} = r \times \frac{\sigma_x}{\sigma_y}; \quad (43)$$

$$R_{y/x} = r \times \frac{\sigma_y}{\sigma_x}; \quad (44)$$

где r – коэффициент корреляции между изучаемыми признаками;
 σ_x – среднее квадратическое отклонение по первому признаку;
 σ_y – среднее квадратическое отклонение по второму признаку.

$$R_{y/x} = r \times \frac{\sigma_y}{\sigma_x} = +0,89 \times \frac{56,8}{7,9} = 6,4 \text{ кг}.$$

При увеличении или уменьшении обхвата груди на 1 см живая масса увеличится или уменьшится на 6,4 кг.

$$R_{x/y} = r \times \frac{\sigma_x}{\sigma_y} = +0,89 \times \frac{7,9}{56,8} = 0,12 \text{ см}$$

При увеличении или уменьшении живой массы на 1 кг обхват груди увеличится или уменьшится на 0,12 см.

Пример 2. Определить процент жира в молоке коров с удоем 4000 кг, если коэффициент корреляции между удоем и процентом жира в молоке равен (-0,2), σ по удою = 600 кг, σ по жиру = 0,4, средний удой по стаду – 3000 кг, средний процент жира по стаду – 3,8.

Методика выполнения

$$X = \bar{X}_x + r \times \frac{\sigma_x}{\sigma_y} (Y - \bar{Y}_y) = 3,8 + (-2) \times \frac{0,4}{600} \times (4000 - 3000) = 3,93\% \quad (40)$$

Пример 3. Построить график изменения одного признака (x) при заданном изменении другого признака (y).

При обработке данных о жирномолочности (x) и белкомолочности (y) стада получены следующие статистические параметры для этих признаков: жир (\bar{X}) = 3,6%; $\sigma_x = 0,2\%$; белок (\bar{Y}) = 3,0%; $\sigma_y = 0,1\%$; $r_{xy} = 0,30$.

Предположим, что содержание жира в молоке коров стада необходимо увеличить с 3,7 (x_1) до 3,8 (x_2) %. Требуется определить, каким может быть уровень белковомолочности в стаде при намеченном плане жирномолочности. Подставляя в уравнение $Y = \bar{Y}_y + r \frac{\sigma_y}{\sigma_x} \times (X - \bar{X}_x)$ известные данные, получаем предполагаемый уровень белковомолочности:

$$Y_1 = 3,0 + 0,3 \times \frac{0,1}{0,2} \times (3,7 - 3,6) = 3,015\%; Y_2 = 3,015 + 0,3 \times \frac{0,1}{0,2} \times (3,8 - 3,7) = 3,03\%;$$

Эти данные могут быть изображены графически в виде линии регрессии [2, 4, 7, 8].

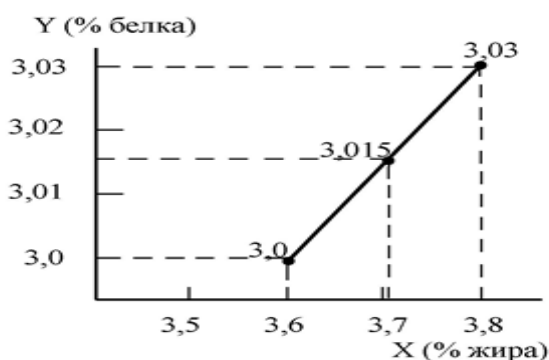


Рисунок 9 – Графическое изображение линии регрессии

Контрольные вопросы:

1. Что такое регрессия? Что характеризуют коэффициенты регрессии?
2. Чем различаются коэффициенты $R_{x/y}$ и $R_{y/x}$?
3. В чем заключается различие между коэффициентами корреляции и регрессии?
4. Напишите уравнение прямолинейной регрессии.
5. Как используются данные регрессивного анализа в животноводстве?

Тема 9. ИЗУЧЕНИЕ КРИВОЛИНЕЙНОГО ТИПА СВЯЗИ И ВЫЧИСЛЕНИЕ КОРРЕЛЯЦИОННОГО ОТНОШЕНИЯ

Цель занятия: ознакомиться с криволинейными типами связи, научиться вычислять силу криволинейной связи при помощи корреляционного отношения.

Основные понятия

Кроме прямолинейных и близких к ним связей, для определения которых можно применять коэффициент корреляции (r), существует многообразие связей криволинейного типа. С целью их установления используют другие формулы и коэффициенты. Многие признаки животных проявляют криволинейную связь с возрастом. Например, удои за лактацию коров имеют криволинейность, выражаемую параболой. Изменение живой массы животного от рождения до

наступления половой зрелости выражается так называемой логистической кривой. У насекомых наблюдаются периодические всплески усиления и затухания в размножении, что можно выразить периодической синусоидной кривой.

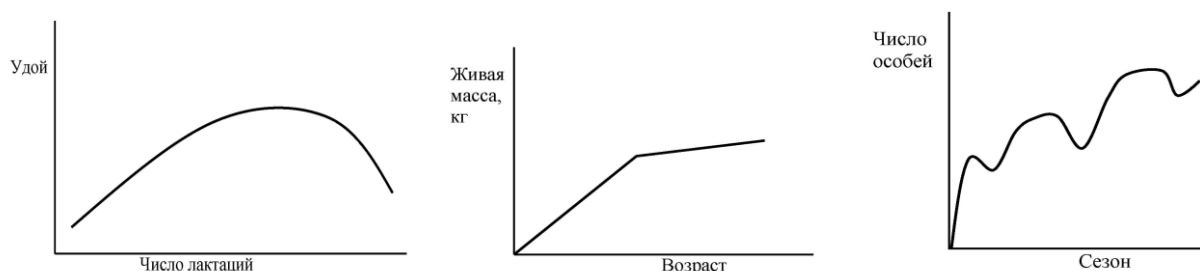


Рисунок 10 – Криволинейные типы связи

При наличии криволинейного типа связи между признаками, как правило, используют показатель корреляционного отношения η (эта), который позволяет определить величину связи, но не выявляет ее направления, т. е. этот коэффициент всегда положительный. Его величина изменяется от 0 до +1. Корреляционное отношение можно применять для измерения как криволинейных, так и прямолинейных связей.

При криволинейном типе связи величина η всегда будет больше величины r , вычисленной на том же материале. Если же связь прямолинейна, то $\eta = r$. Корреляционное отношение имеет два значения: y по x и x по y , следовательно, оно выявляет неравенство связей, так как $\eta_{xy} \neq \eta_{yx}$.

В общем виде формула корреляционного отношения включает показатели групповых (частных) средних квадратических и общих средних квадратических отклонений:

$$\eta_{y/x} = \frac{\sigma_{y/x}}{\sigma_y} ; \quad (45)$$

$$\eta_{x/y} = \frac{\sigma_{x/y}}{\sigma_x} ; \quad (46)$$

$$\sigma_{y/x} = \sqrt{\frac{1}{n} \sum f_x (\bar{y}_x - \bar{y}_{\text{общ}})^2} ; \quad (47)$$

$$\sigma_{x/y} = \sqrt{\frac{1}{n} \sum f_y (\bar{x}_y - \bar{x}_{\text{общ}})^2} . \quad (48)$$

Для вычисления корреляционного отношения по вариационным рядам, оформленным с распределением признаков на варьирующие классы (т. е. при взвешенных рядах), воспользуемся формулой

$$\eta_{y/x} = \sqrt{\frac{\sum (\bar{y}_{\text{част}} - \bar{y}_{\text{общ}})^2 \cdot f_x}{\sum (y - \bar{y}_{\text{общ}})^2 \cdot f_y}}; \quad (49)$$

где $\bar{y}_{\text{общ}}$ – среднее арифметическое признака y ;

$\bar{y}_{\text{част}}$ – средние величины признака y по классам признака x . Здесь зависимой функцией является признак y , а независимым аргументом – признак x .

Для определения η_{yx} составляют корреляционную решетку и определяют эмпирические величины регрессии признака y в зависимости от признака x .

Пример. Определить влияние возраста (x) самок лахвинского карпа на их рабочую плодовитость (y) (тыс. шт.). Возраст выражен продолжительностью использования самок (лет) с момента наступления половой зрелости. В выборку вошло 70 нерестов ($n > 30$), полученных от 50 самок карпа, и учтена рабочая плодовитость для каждого нереста.

Составляем корреляционную таблицу зависимости y -признака от x -признака.

Корреляционная решетка по определению влияния возраста самок лахвинского карпа на их рабочую плодовитость

у – рабочая плодовитость, тыс. шт.		х – возраст (по продолжительности использования), лет								f_y	$y \times f_y$	$\left(y - \bar{y}_{\text{общ}}\right)^2 \times f_y$
границы	серединна класса, у	1	2	3	4	5	6	7	8			
70 – 90	80	6	3	2				1	1	13	1040	
100 – 120	110	4	6	6	5	6	2	2	3	34	3740	1066,2
130 – 150	140		1	5	4	4	4	1	1	20	2800	11907,2
160 – 180	170			2	1					3	510	8878,1
f_x		10	10	15	10	10	6	4	5	$\Sigma f_y = 70$	$\Sigma y \times f_y = 8090$	$\Sigma \left(y - \bar{y}_{\text{общ}}\right)^2 \times f_y = 38327,2$
$\bar{y}_{\text{част.}}$ тыс. шт.		92	104	124	128	122	130	110	110			
$(\bar{y}_{\text{част.}} - \bar{y}_{\text{общ}})^2 \cdot f_x$		5569,6	1345,6	1058,4	1537,6	409,6	1244,2	125,44	156,8	$\Sigma (\bar{y}_{\text{част.}} - \bar{y}_{\text{общ}})^2 \cdot f_x = 11447,2$		

2. Вычисляем эмпирические величины регрессии изменения признака рабочей плодовитости (y) от изменения возраста (x) – $\bar{y}_{\text{част}}$. Для этого умножаем частоты клеток решетки (f_{yx}) на соответствующую величину признака y каждого класса, т. е. по столбцам, и для каждого столбца записываем получаемую сумму от умножения $f_{yx} \times y$ и делим на число f_{xi} :

1-й столбец: $(6 \cdot 80 + 4 \cdot 110) = (480 + 440) : 10 = 920 : 10 = 92$;

2-й столбец: $(3 \cdot 80 + 6 \cdot 110 + 1 \cdot 140) : 10 = 104$ и т.д.

Полученный ряд изменений y по x демонстрирует эмпирический ряд регрессии, из которого видно, что с 3-го нереста по 6-й наблюдается повышение рабочей плодовитости самок, а с 7-го – снижение.

3. Определяем среднюю рабочую продуктивность самок по всем возрастным группам, т.е. $\bar{y}_{\text{общ}}$. Для этого перемножаем частоты ряда y (f_y) по каждому классу на середину класса. Суммируем данные по всем четырем классам признака плодовитости и, разделив сумму на общее число обследованных нерестов ($n=70$), получаем среднюю плодовитость по всему корреляционному комплексу. Эти данные введены в корреляционную решетку (колонка « $y \cdot f_y$ »). Отсюда средняя рабочая плодовитость по всему корреляционному комплексу составит:

$$\bar{y}_{\text{общ.}} = \frac{\sum y \cdot f_y}{\sum f_y} = \frac{8090}{70} = 115,6. \quad (50)$$

4. Далее обработка данных сопровождается получением суммы квадратов отклонения для показателей y от общей средней и умножения на частоты ряда y , т.е. $\sum (y - \bar{y}_{\text{общ}})^2 \times f_y$ (последний столбец корреляционной решетки).

5. Затем производим вычисление суммы квадратов отклонений частных средних ($\bar{y}_{\text{част.}}$) от общей средней, умноженных на частоты для эмпирического ряда регрессии, т.е. на f_x , в соответствии с выражением

$\sum (\bar{y}_{\text{част.}} - \bar{y}_{\text{общ.}})^2 \times f_x$. Этот расчет приведен в последней строчке таблицы.

6. Полученные величины вводим в формулу 49 корреляционного отношения при обработке взвешенных рядов:

$$\eta_{y/x} = \sqrt{\frac{\sum (\bar{y}_{\text{част.}} - \bar{y}_{\text{общ.}})^2 \cdot f_x}{\sum (y - \bar{y}_{\text{общ.}})^2 \cdot f_y}} = \sqrt{\frac{11447,2}{38327,2}} = 0,55.$$

Полученная величина $\eta_{y/x}$ указывает на большую зависимость рабочей плодовитости самок лахвинского карпа от их возрастного показателя, что следует из ряда регрессии y по x [2, 3, 7, 8, 9].

Контрольные вопросы:

1. Что такое криволинейная связь?
2. Какие существуют типы криволинейной связи?
3. С какой целью рассчитывают корреляционное отношение?
4. Свойства корреляционного отношения.

Тема 10. ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТА ГЕНЕТИЧЕСКОЙ КОРРЕЛЯЦИИ

Цель занятия: научиться вычислять коэффициент генетической корреляции.

Основные понятия:

Генетическая корреляция – это показатель связи, указывающий на изменение вторичных признаков при селекции по первичным признакам. Она вычисляется при наличии родственных групп матерей и дочерей, отцов и сыновей, полусибсов, полных сибсов и близнецов. Показатели генетической корреляции имеют важное значение в селекционно-племенной работе. Эффект селекции тесно связан с генетической корреляцией между признаками. Показателем генетической корреляции между признаками является *коэффициент генетической корреляции* (r_G).

Коэффициенты генетической корреляции имеют большое значение при ранней индивидуальной оценке животных, при вычислении селекционных индексов и должны учитываться при отборе родительских пар для получения потомства с наилучшей сочетаемостью желательных признаков.

Коэффициент генетической корреляции (r_G) определяют по фенотипическим показателям коррелируемых признаков у родственных особей. Вычисляют фенотипические коэффициенты корреляции между признаками x дочерей и x матерей, y дочерей и y матерей, x дочерей и y матерей, y дочерей и x матерей. После получения четырех величин r определяют генетический коэффициент связи между признаками x и y , используя формулы Хейзеля 51 и 52:

$$r_{G_{xy}} = \sqrt{\frac{r_{xDyM} \cdot r_{yDxM}}{r_{xDxM} \cdot r_{yDyM}}}; \quad (51)$$

$$r_{G_{xy}} = \sqrt{\frac{R_{xDyM} \cdot R_{yDxM}}{R_{xDxM} \cdot R_{yDyM}}}, \quad (52)$$

- где xD, yD – фенотипическое выражение двух признаков у дочерей;
 xM, yM – фенотипическое выражение двух признаков у матерей;
 $r_{G_{xy}}$ – коэффициент генетической корреляции;
 $R_{G_{xy}}$ – коэффициент регрессии;
 r_{xDxM} – коэффициенты генетических корреляций между одним и
 r_{yDyM} тем же признаком у дочерей и матерей;
 r_{xDyM} – коэффициенты генетических корреляций между разноимен-
 r_{yDxM} ными признаками дочерей и матерей.

Если в числителе подкоренного выражения один из показателей отрицательный, то используется формула 53:

$$r_{G_{xy}} = \frac{(r_{xDyM} + r_{yDxM}) \times 0,5}{\sqrt{r_{xDxM} \times r_{yDyM}}}. \quad (53)$$

Если оба коэффициента корреляции в числителе отрицательные, то производят вычисления, не обращая внимания на знаки. В знаменателе оба коэффициента корреляции должны быть положительные. В противном случае пользоваться формулой нельзя.

Пример 1. Рассчитать коэффициент генетической корреляции между содержанием жира и белка в молоке коров-дочерей и их матерей.

Методика выполнения

Данные заносим в таблицу:

Пара	Содержание белка (x)		Содержание жира (y)	
	дочери (xD)	матери (xM)	дочери (yD)	матери (yM)
1-я	3,1	3,0	4,0	3,9
2-я	3,3	3,1	4,2	4,0
3-я	3,2	3,2	4,1	4,0
4-я	3,0	3,1	4,0	3,8
5-я	3,4	3,3	4,5	4,2

По методике расчета коэффициента корреляции для малых выборок вычисляем r_{xDxM} ; r_{yDyM} ; r_{xDyM} ; r_{yDxM} по формулам 51 и 52.

Содержание белка в молоке дочерей – в молоке матерей: $r_{xDxM} = 0,65$;

-содержание жира в молоке дочерей – в молоке матерей: $r_{yDyM} = 0,80$;

-содержание белка в молоке дочерей – содержание жира в молоке матерей: $r_{xDyM} = 0,733$;

-содержание жира в молоке дочерей – содержание белка в молоке матерей: $r_{yDxM} = 0,053$.

$$r_{G_{xy}} = \sqrt{\frac{r_{xDyM} \times r_{yDxM}}{r_{xDxM} \times r_{yDyM}}} = \sqrt{\frac{0,737 \times 0,053}{0,65 \times 0,80}} = \sqrt{0,0751} = 0,274.$$

Таким образом, между белкомолочностью и жирномолочностью коров-матерей и дочерей существует генетическая связь. Отбор животных по содержанию жира в молоке будет сопровождаться генетически обусловленным повышением содержания белка.

Пример 2. Вычислить коэффициент генетической корреляции между живой массой (x) и яйценоскостью (y) кур – матерей и дочерей.

При вычислении генетической корреляции используются данные фенотипических корреляций у кур:

- между живой массой дочерей в 32-недельном возрасте и годовой яйценоскостью матерей: $r_{xDyM} = + 0,092$;

- между живой массой матерей в 32-недельном возрасте и годовой яйценоскостью дочерей: $r_{xM_yD} = +0,164$;
- между живой массой дочерей и матерей в 32-недельном возрасте: $r_{xD_xM} = +0,410$;
- между годовой яйценоскостью дочерей и матерей: $r_{yD_yM} = +0,340$.

$$r_{G_{xy}} = \sqrt{\frac{0,092 \times 0,164}{0,41 \times 0,34}} = \sqrt{\frac{0,015088}{0,1394}} = \sqrt{0,1082} = +0,328.$$

Следовательно, при селекции по живой массе происходит увеличение яйценоскости [2,3,8,9].

Контрольные вопросы:

1. Сущность метода, разработанного Хейзелем.
2. Между какими признаками рассчитывают коэффициент генетической корреляции?
3. Использование коэффициента генетической корреляции в селекционной работе.

Тема 11. ПОЛИХОРИЧЕСКИЙ И БИСЕРИАЛЬНЫЙ ПОКАЗАТЕЛИ СВЯЗИ

Цель занятия: научиться вычислять полихорический коэффициент связи между признаками.

Основные понятия

При анализе генетических данных в практической селекции животных может возникать необходимость установить связь между качественными признаками, которые оценивают приблизительно. В этом случае применяют полихорический показатель связи.

Полихорический показатель связи (ρ) (ρ_o) – позволяет определить величину связи между двумя качественными признаками, имеющими несколько градаций. Например, связь между типом конституции каракульских овец и типами завитка у ягнят; между породой коров и формой вымени, между типом конституции и степенью упитанности и т.д.

Полихорический показатель связи (ρ) рассчитывают по формуле:

$$\rho = \frac{\alpha - 1}{\sqrt{(l_1 - 1) \times (l_2 - 1)}}, \quad (54)$$

где
$$\alpha = \sum \left[\frac{\sum (p_{1.2}^2 \div p_2)}{p_1} \right] - \frac{(l_1 - 1)(l_2 - 1)}{n}, \quad (55)$$

или
$$\alpha = \sum \left[\frac{\sum (p_{1.2}^2 \div p_1)}{p_2} \right] - \frac{(l_1 - 1)(l_2 - 1)}{n}, \quad (56)$$

где l_1 и l_2 – число классов по каждому признаку;
 $p_{1.2}$ – частоты вариационного ряда каждого из признаков;
 p_1 и p_2 – частоты ряда каждого из признаков;
 n – число наблюдений в выборке.

Величина полихорического коэффициента связи (ρ) колеблется от 0 до 1, этот коэффициент не выявляет направления корреляции, а только указывает на ее величину, поэтому он всегда имеет положительный знак.

Пример 1. Определить величину связи между типом конституции коров-матерей и их дочерей. Обследовано 50 пар (мать – дочь).

Методика выполнения

Распределяют пары по клеткам корреляционной решетки.

1. Каждое число 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 . Такая обработка дает для каждой клетки выражение $(p_{1.2})^2 : p_2$.

Полихорический коэффициент связи между типом конституции коров-дочерей и их матерей

Тип конституции		Тип конституции				p_1	$\sum \frac{(p_{1.2})^2}{p_2}$	$\left[\sum \frac{(p_{1.2})^2}{p_2} \right] / p_1$
		гру-бая	нежная	плот-ная	рыхлая			
	Грубая	$p_{1.2} = 5$ $5^2 = 5$ $\frac{5}{5} = 5$	–	–	$p_{1.2} = 2$ $2^2 = 0,27$ $\frac{2}{15} = 0,27$	7	5,27	$\frac{5,27}{7} = 0,75$
	Нежная	–	$p_{1.2} = 15$ $15^2 = 9$ $\frac{15}{25} = 9$	$p_{1.2} = 5$ $5^2 = 5$ $\frac{5}{5} = 5$	–	20	14,0	$\frac{14}{20} = 0,70$
	Плотная	–	$p_{1.2} = 3$ $3^2 = 0,36$ $\frac{3}{25} = 0,36$	–	–	3	0,36	$\frac{0,36}{3} = 0,12$
	Рыхлая	–	$p_{1.2} = 7$ $7^2 = 1,96$ $\frac{7}{25} = 1,96$	–	$p_{1.2} = 13$ $13^2 = 11,27$ $\frac{13}{15} = 11,27$	20	13,23	$\frac{13,23}{20} = 0,66$
	p_2	5	25	5	15	$n = 50$	–	$\sum = 2,23$

2. Полученную величину суммируют, получая $\sum \frac{(p_{1.2})^2}{p_2}$, и записывают суммы в графу, идущую после графы с частотами вариационного ряда p_1 .

3. Проведя вычисления по формуле $\left[\sum \frac{(p_{1.2})^2}{p_2} \right] \div p_1$, получают итоговое число 2,23.

4. Подставляя эти данные в формулу α_1 , получают:

$$\alpha_1 = \sum \left[\frac{\sum (p_{1.2}^2 \div p_2)}{p_1} \right] - \frac{(l_1 - 1)(l_2 - 1)}{n} = 2,23 - \frac{(4 - 1) \cdot (4 - 1)}{50} = 2,23 - \frac{9}{50} = 2,23 - 0,18 = 2,05.$$

5. Подставляя величину α_1 в формулу ρ , находят полихорический коэффициент:

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

На основании вычисленного полихорического показателя связи ρ можно сделать вывод, что тип конституции дочерей коррелирует с типом конституции матерей.

В данном примере полихорический коэффициент связи ρ равен 0,35, то есть наследственно обусловлен и может быть выражен коэффициентом наследуемости $h^2 = 2r = 2 \cdot 0,35 = 0,70$.

Следовательно, существует высокая связь между типами конституции дочерей и матерей.

Бисериальный показатель связи (r_b) – позволяет установить величину связи между количественным и качественным альтернативным признаком. В этом случае один признак измеряется точно и может принять любое качественное значение, а другой может принять только два крайних (альтернативных) выражения. Для определения бисериального показателя связи составляется корреляционная решетка, в которой будут два класса для альтернативных признаков (+ и –) и классы для количественного признака.

$$r_b = \frac{\frac{\sum p_+ \times ax}{n_+} - \frac{\sum p_x a}{n}}{\sqrt{\frac{C}{n_+} - \frac{C}{n}}}, \quad (57)$$

$$r_b = \frac{\frac{\sum p_+ \times ax}{n_-} - \frac{\sum p_x a}{n}}{\sqrt{\frac{C}{n_-} - \frac{C}{n}}}, \quad (58)$$

$$Cx = \sum p_x \times a^2 x - \frac{(\sum p_x \times ax)^2}{n}, \quad (59)$$

где p – частоты ряда в выборке, распределенные по количественному ряду;

p_+ – частоты первого альтернативного признака;

p_- – частоты второго альтернативного признака;

n – число наблюдений в выборке;

n_+ – число наблюдений по первому альтернативному признаку;

n_- – число наблюдений по второму альтернативному признаку;

a – условное отклонение класса для ряда по всей выборке по количественному признаку от нулевого класса;

Cx – величина, входящая в значение σ для вариационного ряда всей выборки.

Пример 2. Определить, имеется ли корреляционная связь между типами гемоглобина у телят джерсейской породы и величиной живой массы при рождении.

Для животных джерсейской породы характерно в основном наличие типов гемоглобина Hb AA и Hb AB, гомозигот с типом Hb BB в породе очень мало. Следует установить, имеется ли корреляция между величиной живой массы телочек при рождении и типами гемоглобина – гомозиготным Hb AA и гетерозиготным Hb AB.

Корреляционная решетка для определения связи типа гемоглобина у телок джерсейской породы с величиной живой массы

Типы гемоглобина	Класс по живой массе					
	18-20	21-23	24-26	27-29	30-32	
Hb AA (p+)	10	20	30	40	20	$\sum p+ = n+ = 120$
Hb AB (p-)	10	20	30	10	10	$\sum p- = n- = 80$
p_x	20	40	60	50	30	$n+ + n- = 200$
a_x	-2	-1	0	1	2	-
$p+ \cdot a_x$	-20	-20	0	40	40	$\sum p+ \cdot a_x = -40 + 80 = +40$
$p_x \cdot a_x$	-40	-40	0	50	60	$\sum p_x \cdot a_x = -80 + 110 = +30$
$p_x \cdot a^2_x$	80	40	0 условная средняя	50	120	$\sum p_x \cdot a^2_x = 290$

1. Выделяем класс с условной средней (A). В примере это класс с живой массой телок 24-26 ($p_x=60$).

2. Записываем строчку условных средних отклонений (a) от нулевого класса.

3. Записываем строчку со значениями произведений частот ($p+$) для гомозиготного типа гемоглобина на соответствующее условное отклонение (a_x). Итог по строчке суммируем $\sum p+ \cdot a_x = -40 + 80 = +40$.

4. Образует строчку произведений частот (p) по классам количественного признака на соответствующую величину условного отклонения (a), т.е. значения $p_x \cdot a_x$. Сумма этих произведений равна $\sum p_x \cdot a_x = -80 + 110 = +30$.

5. Умножаем значения $p_x \cdot a_x$ в каждом классе на a_x , образуется $p_x \cdot a^2_x$, суммирование по строчке дает $\sum p_x \cdot a^2_x = 290$.

6. Вычисляем величину Cx .

$$Cx = \sum p_x \cdot a^2_x - \frac{(\sum p_x \cdot a_x)^2}{n} = 290 - \frac{30^2}{200} = 290 - \frac{900}{200} = 285,5$$

7. Вычисляем для числителя

$$\frac{\sum p_{+} \cdot ax}{n_{+}} \text{ и } \frac{\sum px \cdot ax}{n}; \quad (60, 61)$$

$$\frac{\sum p_{+} \cdot ax}{n_{+}} = \frac{+40}{120} = 0,33; \quad \frac{\sum px \cdot ax}{n} = \frac{+30}{200} = 0,15.$$

8. Определяем бисериальный коэффициент связи

$$r_b = \frac{\frac{\sum p_{+} \cdot ax}{n_{+}} - \frac{\sum px \cdot ax}{n}}{\sqrt{\frac{C}{n_{+}} - \frac{C}{n}}} = \frac{0,33 - 0,15}{\sqrt{\frac{285,5}{120} - \frac{285,5}{200}}} = \frac{0,18}{\sqrt{2,3791 - 1,4275}} = \frac{0,18}{\sqrt{0,9516}} = 0,184.$$

Вычисленный бисериальный коэффициент указывает на незначительную связь между живой массой телок при рождении и типом гемоглобина в их крови [2, 3, 9].

Контрольные вопросы:

1. При анализе каких признаков применяют полихорический коэффициент связи?
2. В каких пределах варьирует величина полихорического коэффициента связи (ρ)?
3. Что позволяет установить бисериальный показатель связи?
4. Использование полихорического и бисериального показателей связи в селекционной работе.

Тема 12. ВЫЧИСЛЕНИЕ КРИТЕРИЯ ДОСТОВЕРНОСТИ РАЗНОСТИ МЕЖДУ ВАРЬИРУЮЩИМИ ПРИЗНАКАМИ

Цель занятия: сравнить средние арифметические двух групп животных и установить, достоверна ли разность между ними и при каком уровне вероятности.

Основные понятия

При изучении изменчивости какого-либо признака и расчете основных показателей обычно используют данные по некоторой части совокупности особей, т.е. выборку. Выводы, сделанные по относительно ограниченной группе животных, рассматривают как определенную закономерность. Поскольку выборочная совокупность составляет только часть генеральной совокупности, то в этом случае необходимо учитывать величину статистической ошибки.

Важное значение при сравнении двух средних арифметических имеет ошибка средней арифметической. Она показывает, в каких пределах средняя арифметическая в данной выборке отклоняется от средней арифметической генеральной совокупности, она зависит от объема выборки, чем больше объем выборки, тем меньше ошибка и наоборот. Две средние арифметические, полученные в результате случайной выборки из одной и той же статистической со-

вокупности, не могут отличаться друг от друга более чем на три ошибки. Если разница между средними арифметическими превышает ошибку в три раза и более, то такие средние относятся к разным статистическим совокупностям. Если разница между средними арифметическими не превышает трех ошибок, то это может свидетельствовать об отсутствии существенной разницы между средними, или о ее недостоверности.

Основные типы ошибок:

- 1) *технические* – это следствие просчетов, описок, арифметически неверных вычислений и т.п., ошибки по невнимательности;
- 2) *систематические* – это ошибки, появляющиеся в результате неточного измерения показателя за счет неисправного прибора, неопытности и т.д.
- 3) *статистические (ошибки репрезентативности)* обусловлены самим статистическим методом. Этот метод предусматривает характеристику генеральной совокупности на основе выборки, составленной по принципу случайности и типичности. Характеристика таким методом всегда будет неточной. Эта неточность возникает, когда генеральную совокупность (N) пытаются охарактеризовать по ее части (n). Максимально приблизиться к точной характеристике генеральной совокупности можно лишь при условии, что n – объем выборки будет приближен к N – объему генеральной совокупности. Чем меньше ошибки репрезентативности, тем более точно выборочные параметры будут характеризовать генеральную совокупность.

Величина статистической ошибки зависит от степени изменчивости признака: чем выше изменчивость, тем больше статистическая ошибка. Репрезентативность напрямую связана с численностью выборки. Чем больше объем выборки, тем ближе полученные статистические параметры к параметрам генеральной совокупности и меньше ошибка. При условии, что $n = N$ – ошибок репрезентативности не будет.

В биометрии принято статистические ошибки обозначать буквой m с подстрочным индексом того параметра, для которого она вычисляется ($m_{\bar{x}}$; m_{σ} ; m_{C_v} ; m_r ; m_R).

Величину ошибок репрезентативности определяют только для выборочных показателей и записывают величину выборочного показателя с величиной его ошибки со знаком « \pm »: $\bar{X} \pm m_{\bar{x}}$; $\sigma \pm m_{\sigma}$; $C_v \pm m_{C_v}$ и т.д. Такая форма записи показывает, что параметр генеральной совокупности может отклоняться от выборочного в ту или другую сторону на величину ошибки. Если ошибка репрезентативности (m) очень велика, то выборка (n) неправильно составлена и при помощи ее данных нельзя характеризовать генеральную совокупность (N).

Формулы статистических ошибок:

- ошибка средней арифметической: $m_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$; (60)

- ошибка среднего квадратического отклонения: $m_{\sigma} = \frac{\sigma}{\sqrt{2n}}$; (61)

$$\text{ - ошибка коэффициента вариации: } m_{Cv} = \frac{Cv}{\sqrt{2n}}; \quad (62)$$

$$\text{ - ошибки коэффициента корреляции: } m_r = \sqrt{\frac{1-r^2}{n-2}} \quad (n < 100); \quad (63)$$

$$m_r = \frac{1-r^2}{\sqrt{n-1}} \quad (n > 100); \quad (64)$$

$$\text{ - ошибка коэффициента регрессии: } m_{R1,2} = \frac{\sigma_1}{\sigma_2} \sqrt{\frac{1-r^2}{n-2}}; \quad (65)$$

$$m_{R2,1} = \frac{\sigma_2}{\sigma_1} \sqrt{\frac{1-r^2}{n-2}}; \quad (66)$$

$$\text{ - ошибка долей } p \text{ и } q: \quad m_p = m_q = \sqrt{\frac{p \cdot q}{n}}. \quad (67)$$

Для решения вопроса о существенности или несущественности разницы между двумя средними арифметическими определяют *критерий достоверности* (**td**), т.е. отношение разницы к ошибке. Его величина зависит от числа наблюдений (особей) (**n**) в выборке и связана с понятием *вероятности* (**P**), которая дает возможность считать данные, полученные в выборке достоверными. Критерий достоверности рассчитывается по формуле 68 :

$$t_d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m_1^2 + m_2^2}}, \quad (68)$$

где t_d – достоверность разности;

\bar{X}_1, \bar{X}_2 – средние арифметические величины изучаемого признака в разных группах;

m – ошибка средней арифметической.

Вероятность (**P**) – это объективная возможность наступления какого-либо события. Кроме понятия вероятности часто используется понятие *значимости*. Значимость отражает уровень риска и ошибки.

При $td = 1,96$	$P = 0,95$	первый порог достоверности;
$td = 2,58$	$P = 0,99$	второй порог достоверности;
$td = 3,29$	$P = 0,999$	третий порог достоверности.

Стандартные значения критерия достоверности (**td**) определяют по таблице Стьюдента, при этом учитывается число степеней свободы.

Число степеней свободы – это число наблюдений, уменьшенное на число ограничений: **n - 1**; **n - l** и т. д.

Критерий достоверности также позволяет определить границы доверительного интервала.

Пример 1. Сравнить молочную продуктивность коров голштинской (\bar{X}_1) и белорусской черно-пестрой (\bar{X}_2) пород. В одинаковых условиях кормления и содержания получены следующие показатели по удою:

$\bar{X}_1 \pm m_1 = 7218 \pm 103$ кг ($n=76$); $\bar{X}_2 \pm m_2 = 5634 \pm 95$ кг ($n=112$). Рассчитать критерий достоверности (td) для количественных признаков и определить, достоверна ли разность между средними арифметическими двух выборок по количественным признакам.

Методика выполнения

1. Найти разность между двумя средними арифметическими по формуле 69:

$$\bar{X}_1 - \bar{X}_2 = 7218 - 5634 = 1584 \text{ кг.} \quad (69)$$

2. Определить среднюю ошибку разности двух выборок (m_d) по формуле 67:

$$m_d = \sqrt{m_1^2 + m_2^2} = \sqrt{103^2 + 95^2} = 140. \quad (70)$$

3. Рассчитать критерий достоверности (td) по формуле 68:

$$td = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m_1^2 + m_2^2}} = \frac{1584}{140} = 11,3. \quad (68)$$

4. Эмпирический критерий Стьюдента (td) сравнить со стандартными значениями критерия t_{st} (приложение 1), с учетом степеней свободы ($v = n_1 + n_2 - 2$).

При сравнении больших выборок, если $td < 1,96$ – разница недостоверна, если $td > 1,96$, но $< 2,58$ – разница достоверна при $P > 0,95$, если $td > 2,58$ – то разница достоверна при $P > 0,99$, и при $td > 3,29$ – разность очень высокодостоверна при $P > 0,999$ (99,9 %). В данном примере разница очень высокодостоверна при $P > 0,999$, т.к. $td = 11,3$.

Пример 2. Оценить эффективность использования антиоксиданта (витамина Е) для повышения жизнеспособности телят, если препарат применялся на группе коров:

Группа коров	Родилось телят	Пало
Опытная (60 мг витамина Е на голову)	29	4
Контрольная	29	9

Методика выполнения

1. В данном случае каждая выборка состоит из двух групп животных: 1- больные (павшие) (p), 2- здоровые (q). Определить по каждой выборке долю павших животных (p):

$$p_{1,2} = \frac{p}{n} \quad (71)$$

$$p_1 = \frac{4}{29} = 0,14, \quad p_2 = \frac{9}{29} = 0,31,$$

где p_1 и p_2 – численность животных с определенной степенью развития признака;

n – общая численность выборки.

2. Определить по каждой выборке долю здоровых животных (q) по формуле 72:

$$q = 1 - p \quad (72)$$

$$q_1 = 1 - p_1 = 1 - 0,14 = 0,86; \quad q_2 = 1 - p_2 = 1 - 0,31 = 0,69.$$

3. Определить среднее квадратическое отклонение по формуле 73:

$$\sigma = \sqrt{p \times q}. \quad (73)$$

$$\sigma_1 = \sqrt{p_1 \times q_1} = \sqrt{0,14 \times 0,86} = 0,34; \quad \sigma_2 = \sqrt{p_2 \times q_2} = \sqrt{0,31 \times 0,69} = 0,46;$$

4. Рассчитать по каждой выборке среднюю ошибку m_1 и m_2 по формуле 67:

$$m = \sqrt{\frac{p \times q}{n}}. \quad (67)$$

$$m_1 = \sqrt{\frac{p_1 \times q_1}{n_1}} = \sqrt{\frac{0,14 \times 0,86}{29}} = 0,11; \quad m_2 = \sqrt{\frac{p_2 \times q_2}{n_2}} = \sqrt{\frac{0,31 \times 0,69}{29}} = 0,09.$$

5. Рассчитать критерий достоверности разности между выборочными долями двух выборочных совокупностей (td) по формуле 74:

$$td = \frac{p_1 - p_2}{\sqrt{m_1^2 + m_2^2}} = \frac{0,31 - 0,14}{\sqrt{0,11^2 + 0,09^2}} = \frac{0,17}{0,14} = 1,21. \quad (74)$$

По таблице Стьюдента (приложение 1) с учетом степеней свободы находим значения t_{st} . Если $td < 1,96$ – разница недостоверна.

В данном примере вычисленный критерий достоверности $td=1,21$, следовательно, использование антиоксиданта (витамина Е) не влияет на жизнеспособность телят [2, 3, 6, 8, 9].

Контрольные вопросы:

1. Основные типы ошибок, почему они возникают при биометрической обработке данных?
2. Использование статистических ошибок для установления достоверности разности двух выборок.
3. Что влияет на величину статистической ошибки?
4. Три порога вероятности безошибочного прогноза.

Тема 13. ГРАНИЦЫ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА. ОПРЕДЕЛЕНИЕ НЕОБХОДИМОГО ОБЪЕМА ВЫБОРКИ И ПРАВИЛО ОТБРАСЫВАНИЯ РЕЗКО ОТКЛОНЯЮЩИХСЯ ВАРИАНТ ИЗ СОВОКУПНОСТИ

Цель занятия: изучение методик установления границ доверительного интервала, определения объема выборочной совокупности для проведения исследований, правило отбрасывания резко отклоняющихся вариантов.

Основные понятия

Доверительными границами называют крайние значения, в пределах которых может находиться параметр генеральной совокупности при данной величине статистической ошибки (m) и уровнях (td). Доверительные границы устанавливаются с использованием критерия достоверности (td).

Доверительные границы определяют по следующим формулам:

$$\bar{X}_{\text{ген}} = \bar{X}_{\text{выб}} \pm td \cdot m_x; \quad \sigma_{\text{ген}} = \sigma_{\text{выб}} \pm td \cdot m_{\sigma}; \quad C_v = C_v \pm td \cdot m_{cv}. \quad (75, 76, 77)$$

Значения доверительных границ будут находиться в пределах от нижней до верхней границ варьирующего признака и зависеть от конкретного значения td и ошибки репрезентативности.

Пример 1. Установить границы доверительного интервала по формуле 75:

$$X_{\text{генер.}} = \bar{X}_{\text{выб.}} \pm td \cdot m_x$$

где $\bar{X}_{\text{выб.}}$ – средняя арифметическая изучаемого признака;
 td – значение критерия достоверности;
 m_x – значение ошибки средней арифметической.

Методика выполнения

Средняя плодовитость $\bar{X}_{\text{выб.}}$ свиноматок составила 10 поросят, ошибка m_x равна 0,5 головы, с вероятностью $P=0,95$ при $td=1,96$ можно утверждать, что $X_{\text{генер}}$ по плодовитости будет находиться в интервале:

от $x_{\text{выб.}} + 1,96 \cdot m = 10 + 1,96 \cdot 0,5 = 10 + 0,98 = 10,98 \approx 11$ голов.

до $x_{\text{выб.}} - 1,96 \cdot m = 10 - 0,98 = 9,02 \approx 9$ голов.

Таким образом, можно утверждать, что генеральная средняя плодовитости свиней находится в границах от 9 до 11 голов.

Определение необходимого объема выборки.

Различают две основные статистические совокупности: *генеральную* и *выборочную*.

Генеральная совокупность (N) – это массовый материал, т.е. это большое

количество изучаемых биологических объектов (косяк, стадо, линия, популяция, порода и т.д.). Объем генеральной совокупности стремится к бесконечности ($N \rightarrow \infty$). *Выборочная совокупность, или выборка (n), – это часть генеральной совокупности, позволяющая изучить и оценить ее.* Принято считать выборку малой при $n < 30$ и большой при $n > 30$, но n не может быть меньше двух особей. Следует помнить, что выборка – это лишь часть генеральной совокупности и для того, чтобы при помощи ее можно было правильно охарактеризовать всю генеральную совокупность, следует правильно отобрать животных в выборку. Отбор биологических объектов в выборку осуществляется по принципам рендоминизации (случайности) и репрезентативности (типичности).

Примером выборки могут служить: особи, отобранные из косяка или стада для характеристики его по каким-либо хозяйственно полезным признакам, пробы крови – для гематологического анализа, зерна – для оценки его всхожести и др.

При проведении различного рода исследований, научных экспериментов, обработки массового материала необходимо правильно определить количество изучаемых объектов. Правильно выбранный объем совокупности может обеспечить получение достоверных статистических параметров, что позволит более точно охарактеризовать всю генеральную совокупность.

Ошибки репрезентативности и величина критерия достоверности находят широкое применение для определения величины объема выборочной совокупности. При этом следует учитывать степень изменчивости изучаемого признака, чем она выше, тем больше должен быть объем выборки. Если известен объем генеральной совокупности (N), то для вычисления объема выборки используют формулу 77:

$$n = \frac{N}{N\left(\frac{d}{td}\right)^2 + 1}, \quad (77)$$

где n – объем выборки;

N – объем генеральной совокупности;

td – критерий достоверности по трем порогам вероятности

(1- $td = 1,96$; 2- $td = 2,58$; 3- $td = 3,3$);

d – допустимое расхождение между $\bar{X}_{выб.}$ и $\bar{X}_{ген.}$, выраженное в долях σ , или допустимая погрешность.

$$d = \frac{\Delta}{\sigma}, \quad (78)$$

где Δ – абсолютная допустимая разность, т.е. $\Delta = \bar{X}_{генер.} - \bar{X}_{выб.}$

σ – среднее квадратическое отклонение, которое можно определить по лимиту изменчивости признака, исходя из закономерности, что лимит равен $\bar{X} \pm 3\sigma$:

$$\sigma = \frac{\bar{X}_{max} - \bar{X}_{min}}{6}. \quad (79)$$

Если генеральная совокупность имеет большую численность ($N \rightarrow \infty$), то для определения объема выборки используется формула 80:

$$n = t^2 : d^2. \quad (80)$$

Пример 2. Определить объем выборки, необходимый для достоверности показателя средней плодовитости свиноматок при $td = 2$.

Методика выполнения

Предположительно, плодовитость в популяции колеблется от 6 до 18 поросят, тогда $\sigma = (18-6) : 6 = 2$ головы.

При этом $\bar{X}_{выб.}$ отличается от $\bar{X}_{генер.}$ на $\Delta = 0,5$ головы. Тогда $d = \frac{\Delta}{\sigma} = \frac{0,5}{2} = 0,25$. Отсюда $n = \frac{td^2}{d^2} = \frac{2^2}{0,25^2} = \frac{4}{0,0625} = 64$ головы. Следовательно, чтобы получить достоверный показатель средней плодовитости свиней, необходимо включить в выборку 64 свиноматки.

Для того чтобы получить достоверную разность между двумя средними арифметическими ($\bar{X}_1 - \bar{X}_2 = D$), используют формулу для определения объема двух выборок (n_1 и n_2):

$$n_1 = \frac{td^2}{\Delta^2} \left(\sigma_1^2 + \frac{\sigma_2^2}{a} \right); \quad (81)$$

где td – критерий достоверности;

$a = n_2 : n_1$ – отношение объемов двух выборок, если планируется, что $n_1 \neq n_2$;

σ_1^2 и σ_2^2 – среднее квадратическое отклонение для каждой группы;

Δ – допускаемое расхождение между выборочными средними и генеральной средней сравниваемых групп.

Если изменчивость признака в обеих группах близкая или равная ($\sigma_1^2 \cong \sigma_2^2$), то формула преобразуется и принимает вид:

$$n_1 = \frac{td^2}{\Delta^2} \cdot \sigma^2 \left(1 + \frac{1}{a} \right); \quad (82)$$

где
$$d^2 = \frac{\Delta^2}{\sigma^2}, \quad (83)$$

$$\Delta^2 = d^2 \sigma^2. \quad (84)$$

Подставляя это выражение в формулу, получают

$$n_1 = \frac{td^2}{d^2} \left(1 + \frac{1}{a} \right), \quad (85)$$

где $a = n_2 : n_1$ или $n_1 : n_2$. (86)

Если объемы сравниваемых выборок равны ($n_1=n_2$), то формула упрощается и имеет следующий вид:

$$n_1 = n_2 = \frac{td^2}{d^2} \left(1 + \frac{1}{1}\right) = \frac{2td^2}{d^2}. \quad (87)$$

Если численность N велика или неизвестна, то для определения объема выборки пользуются формулой 88:

$$n = \frac{t^2}{d^2}. \quad (88)$$

Пример 3. При оценке производителя по продуктивности дочерей требуется определить число дочерей и сверстниц, которое необходимо включить в обработку для получения достоверной разности между дочерьми производителя и их сверстницами.

Методика выполнения

Предположим, что изменчивость удоя сверстниц и дочерей близка и составляет 60 кг ($\sigma_{\text{св.}}^2 \approx \sigma_{\text{доч.}}^2$). Уровень td берем при $P=0,95$, тогда $td=2$; $\Delta = \bar{X}_{\text{св.}} - \bar{X}_{\text{доч.}} = 30$ кг. Сверстниц в стаде в два раза больше, чем дочерей быка, поэтому $n_{\text{св.}} : n_{\text{доч.}} = 2 = a$, откуда $n_{\text{св.}} = 2n_{\text{доч.}}$. Величину d^2 вычисляют по формуле 85:

$$d^2 = \frac{\Delta^2}{\sigma^2} = \frac{30^2}{60^2} = 0,25. \text{ Тогда количество сверстниц равно :}$$

$$n_{\text{св.}} = \frac{td^2}{d^2} \left(1 + \frac{1}{a}\right) = \frac{2^2}{0,25^2} \left(1 + \frac{1}{2}\right) = \frac{4 \cdot 1,5}{0,0625} = 96 \text{ голов.}$$

Следовательно, группа сверстниц должна быть численностью около 100 голов, а дочерей в 2 раза меньше. Это обеспечит достоверность разности между удоем коров сравниваемых групп ($D = \bar{X}_{\text{св.}} - \bar{X}_{\text{доч.}}$).

Правило отбрасывания резко отклоняющихся вариантов

При обработке материала часто возникает необходимость решения вопроса: принадлежат ли резко отклоняющиеся варианты к данной генеральной совокупности или они достоверно отклоняются и должны быть устранены, как представители другой генеральной совокупности. Для решения вопроса о правомочности исключения резко отклоняющихся вариантов (x_{\min} или x_{\max}) используется правило « $\pm 3\sigma$ » от среднего значения признака в данной совокупности.

Пример 4. Для исследований отобрано 10 здоровых коров. Минимальное (x_{\min}) число лейкоцитов в их крови составило 5900, а максимальное (x_{\max}) - 10000. Средняя арифметическая (\bar{X}) по 10 коровам - 6000 лейкоцитов, $\sigma=100$ лейкоцитов. Требуется определить, принадлежат ли крайние варианты к изучаемой генеральной совокупности.

Методика выполнения

Устанавливают теоретические границы варьирования по методу « $\pm 3\sigma$ ». Находят отклонения крайних вариант от средней:

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

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

Вычисление показало, что эмпирический минимум меньше, чем теоретический ($x_{эмп} = 5900$; $x_{теор} = 5700$), следовательно, минимальный вариант принадлежит генеральной совокупности, из которой сделана выборка (10 коров). Эмпирический максимальный вариант резко превосходит теоретический максимум ($x_{эмп} = 10000$; $x_{теор} = 6300$), поэтому животное с таким x_{max} из выборки исключается, как принадлежащее другой генеральной совокупности, которая характеризует животных, возможно, больных лейкозом, так как в крови у них содержится большее количество лейкоцитов, чем у здоровых [2, 3, 7, 8, 9].

Контрольные вопросы:

1. Что называется доверительными границами?
2. С какой целью определяют необходимый объем выборки?
3. Какие варианты называются резко отклоняющимися?

Тема 14. ОЦЕНКА ДОСТОВЕРНОСТИ МЕТОДОМ ХИ-КВАДРАТ (χ^2)

Цель занятия: освоение метода хи-квадрат, использование его при решении генетико-селекционных, ветеринарных и других задач.

Основные понятия

Критерий соответствия или критерий χ^2 применяется для сопоставления эмпирических и теоретических частот количественных и качественных признаков, при гибридологическом анализе, при оценке закономерности распределения частот в популяциях. Он применим для выборок численностью 20 особей и более. Его нельзя использовать, когда частоты выражаются в относительных величинах.

Формула критерия хи-квадрат (89) включает сумму дробей, полученную от деления квадрата разности эмпирических ($p_{эмп.}$) и теоретических частот ($p_{теор.}$) на частоты теоретические ($p_{теор.}$):

$$\chi^2 = \sum \frac{(p_{эмп.} - p_{теор.})^2}{p_{теор.}}, \quad (89)$$

где \sum – знак суммы;
 $p_{эмп.}$ – частоты данных, полученных в конкретном опыте;
 $p_{теор.}$ – предполагаемые частоты данных.

Величина χ^2 выражается любым положительным числом от 0 до ∞ .

С помощью этого метода можно определить достоверность расхождения между частотами обоих рядов или его случайности.

Чтобы установить достоверность разности между частотами 2-х рядов методом хи-квадрат, необходимо выбрать *нулевую гипотезу*, которая указывает на то, что между частотами сопоставимых вариационных рядов нет достоверной разницы, т. е. оба ряда являются выборками из одной генеральной совокупности.

Для доказательства достоверности разности между частотами 2-х рядов вначале рассчитывают величину хи-квадрат по конкретным данным ($\chi^2_{\text{эмп.}}$), а затем сравнивают эту величину с величиной $\chi^2_{\text{теор.}}$, взятой из специальной таблицы, с учетом числа степеней свободы.

Если же, $\chi^2_{\text{эмп.}}$ окажется больше $\chi^2_{\text{теор.}}$ при $p=0,05$ или $p=0,01$, ($\chi^2_{\text{эмп.}} > \chi^2_{\text{теор.}}$), то нулевая гипотеза отвергается, это означает, что частоты сравниваемых рядов достоверно отличаются друг от друга. Если же, $\chi^2_{\text{эмп.}} \leq \chi^2_{\text{теор.}}$, при $p=0,05$, нулевая гипотеза остается в силе.

В том случае, если нулевая гипотеза отбрасывается при $p=0,05$, то это значит, что все-таки нулевая гипотеза правильна и расхождение между частотами рядов случайно, то есть из 100 случаев в 5 случаях ряды принадлежат к одной и той же генеральной совокупности.

Таким образом, хи-квадрат позволяет сравнивать распределение опытных (фактических) и теоретических рядов между собой для качественных и количественных признаков. Если разница недостоверна, то сохраняется нулевая гипотеза.

1. Применение критерия χ^2 (хи-квадрат) при изучении наследования качественных признаков.

Пример 1. При спаривании белых хряков с белыми свиноматками во втором поколении получено 8024 поросенка, из них 6023 белой масти (доминантный признак) и 2001 черной (рецессивный признак). Определить достоверность разности в частотах фактического и теоретического распределения с помощью критерия χ^2 . В качестве нулевой гипотезы берется соотношение фенотипов 3 : 1.

Методика выполнения

При моногибридном скрещивании во втором поколении происходит расщепление 3:1, зная общее число особей, мы можем рассчитать теоретически ожидаемое их число с доминантным и рецессивным признаком. Затем подсчитываем, сколько имеется указанных особей фактически. Вносим эти данные в таблицу для вычисления χ^2 .

Полученную величину сравниваем с табличной (с учетом числа степеней свободы). Если полученное значение χ^2 выше табличного, то полученные данные не соответствуют теоретически ожидаемым при моногибридном скрещивании.

Фенотип	Наблюдаемое число особей ($p_{\text{эмп.}}$)	Ожидаемое число особей ($p_{\text{теор.}}$)	$p_{\text{эмп.}} - p_{\text{теор.}}$	$(p_{\text{эмп.}} - p_{\text{теор.}})^2$	$\frac{(p_{\text{эмп.}} - p_{\text{теор.}})^2}{p_{\text{теор.}}}$
доминантный	6023	6018	+5	25	0,0041
рецессивный	2001	2006	-5	25	0,0124
	$\Sigma = 8024$	$\Sigma = 8024$	-	-	$\Sigma = 0,0165$

$\chi^2 = 0,0165$. Полученную величину χ^2 сравниваем с табличной. В примере число степеней свободы (ν) равно 1, так как имеются 2 класса: $\nu = n - 1$.

Число степеней свободы $2 - 1 = 1$. Для определения соответствия в таблице используется строка $\nu = 1$, в которой имеются 3 значения χ^2 : 3,8; 6,6; 10,8.

Вычисленное значение $\chi^2 = 0,0165$, что значительно меньше табличных значений, следовательно, наблюдаемое в опыте расщепление соответствует ожидаемому, поэтому нулевая гипотеза остается в силе.

Пример 2. При скрещивании черных короткошерстных самцов (***AABB***) с белыми пуховыми самками (***aabb***) в F_1 получено гетерозиготное потомство ***AaBb*** – черное короткошерстное. Доминантными признаками являются черная шерсть и нормальная ее длина, рецессивные признаки – белая шерсть и пуховый тип шерсти. При скрещивании таких гетерозигот между собой в F_2 получено крольчат:

45 черных короткошерстных;
30 черных пуховых;
25 белых короткошерстных;
20 белых пуховых.

Определить достоверность разности в частотах фактического и теоретического распределения кроликов по цвету и типу шерсти. В качестве нулевой гипотезы берется соотношение фенотипов 9 : 3 : 3 : 1.

Методика выполнения

При скрещивании получено 120 потомков, теоретически ожидаемое расщепление в потомстве должно соответствовать соотношению четырех фенотипов (9 : 3 : 3 : 1), которое типично при дигибридном скрещивании. В соответствии с теоретически ожидаемым расщеплением можно ожидать следующее:

черных короткошерстных - $9/16 \cdot 120 = 67,5$;
черных пуховых – $3/16 \cdot 120 = 22,5$;
белых короткошерстных - $3/16 \cdot 120 = 22,5$;
белых пуховых – $1/16 \cdot 120 = 7,5$.

Заполнить таблицу:

Фенотип Число голов	черные коротко- шерстные	черные пуховые	белые ко- ротко- шерстные	белые пуховые	n
$p_{\text{эмп.}}$	67,7	22,5	22,5	7,5	$\sum = 120$
$p_{\text{теор.}}$	45	30	25	20	$\sum = 120$
$p_{\text{эмп.}} - p_{\text{теор.}}$	-22,5	+7,5	+2,5	-12,5	-
$(p_{\text{эмп.}} - p_{\text{теор.}})^2$	506,25	56,25	6,25	156,25	-
$\frac{(p_{\text{эмп.}} - p_{\text{теор.}})^2}{p_{\text{теор.}}}$	$\frac{506,25}{67,5} = 7,5$	$\frac{56,25}{22,5} = 2,5$	$\frac{6,25}{22,5} = 0,28$	$\frac{156,25}{7,5} = 20,8$	$\chi^2 = 31,08$

Число степеней свободы: $\nu = n - 1 = 4 - 1 = 3$. Табличное значение χ^2 при $\nu = 3$ и $p = 0,01$ составляет 11,3; эмпирическое значение $\chi^2 = 31,08$, т.е. больше $\chi^2_{\text{теор.}}(11,3)$, поэтому обнаружено достоверное расхождение между частотами, эмпирическими и теоретически возможными по соотношению фенотипов.

Вывод: полученная группа кроликов отклоняется по распределению фенотипов от законов Менделя при дигибридном скрещивании и отражает влияние каких-то факторов, изменяющих тип расщепления по фенотипу во втором поколении помесей.

2. Применение критерия χ^2 (хи-квадрат) при сравнении двух вариационных рядов - эмпирического и теоретического.

Пример 3. Требуется определить, достоверны ли различия по жирномолочности коров разных групп.

Методика выполнения

В качестве нулевой гипотезы принимается предположение, что между селекционной и пользовательной частями стада нет различий. Если число наблюдений в классах меньше 5, то такие частоты смежных рядов объединяют. Таким образом получено 6 классов (l).

Класс по жирно- молочности	Наблюдаемое число особей ($p_{\text{эмп.}}$) (селекционное ядро)	Теоретическое число особей ($p_{\text{теор.}}$) (пользовательное стадо)	$p_{\text{эмп.}}$ - $p_{\text{теор.}}$	$(p_{\text{эмп.}} - p_{\text{теор.}})^2$ $p_{\text{теор.}}$	$\frac{(p_{\text{эмп.}} - p_{\text{теор.}})^2}{p_{\text{теор.}}}$
3,31	1 } 5	3 } 8	-3	9	9 : 8 = 1,12
3,42					
3,68	5	10	-5	25	25 : 10 = 2,50
3,74	7	11	-4	16	16 : 11 = 1,45
3,78	10	9	+1	1	1 : 9 = 0,11
3,92	17	7	+10	100	100 : 7 = 14,3
4,20	2 } 6	3 } 5	+1	1	1 : 5 = 0,2
4,26					
	$\sum = 50$	$\sum = 50$			$\sum = 19,63$

Определяем достоверность (χ^2). Находим число степеней свободы:

$$v = l - 2 = 6 - 2 = 4$$

При v , равном 4 и P , равном 0,01, стандартное значение хи – квадрат равно 13,3 (таблица стандартных значений χ^2 при разных уровнях вероятности). Следовательно, нулевая гипотеза, т.е. предположение о том, что различий между племенным ядром и пользовательным стадом нет, отвергается [2, 3, 7, 8, 9, 11].

Контрольные вопросы:

1. Какие существуют методы оценки достоверности разности между средними величинами двух выборок?
2. Использование критерия соответствия хи-квадрат в генетических исследованиях.
3. Понятия вероятности и значимости в применении к χ^2 .

Тема 15. ОПРЕДЕЛЕНИЕ ЧАСТОТ АЛЛЕЛЕЙ И ГЕНОТИПОВ В ПОПУЛЯЦИИ

Цель занятия: изучить методику определения частот аллелей и генотипов.

Основные понятия

Популяцию можно определить как совокупность особей одного вида, заселяющих определенную территорию, свободно скрещивающихся друг с другом и в той или иной степени изолированных от других совокупностей особей данного вида. Панмиктической, или свободно размножающейся, называют такую популяцию, в которой происходит спаривание любых животных независимо от их генотипа [2, 10].

Для определения структуры свободно размножающейся популяции используют закон **Харди-Вайнберга**: при отсутствии отбора и свободе скрещивания популяция находится в равновесии, т.е. из поколения в поколение не изменяется и в ней сохраняется определенное соотношение генотипов и определяется формулами 90, 91:

$$p + q = 1 (100 \%), \quad (90)$$

$$p^2 AA + 2pqAa + q^2 aa = 1(100\%), \quad (91)$$

где p – частота встречаемости доминантного гена;
 q – частота встречаемости его рецессивного аллеля;
 p^2 – частота встречаемости доминантных гомозигот;
 $2pq$ – частота встречаемости гетерозигот;
 q^2 – частота встречаемости рецессивных гомозигот.

Пример 1. В популяции беспородных собак было найдено 245 животных корот-

коногих и 24 с нормальными ногами. Коротконогость у собак – доминантный признак (А), нормальная длина ног – рецессивный (а). Определить частоту аллелей А и а и генотипов АА, Аа и аа в данной популяции.

Методика выполнения

1. Находим общее количество собак: $245 + 24 = 269$.
2. Вычисляем процент рецессивных особей q^2aa :
 $q^2 = 24 : 269 \cdot 100\% = 9\% = 0,09$.
1. Определяем частоту аллеля *a*: $qa = \sqrt{0,09} = 0,3$.
2. Определяем частоту аллеля *A*: $pA = 1 - 0,3 = 0,7$.
3. Определяем частоту генотипа *AA*: $p^2 = 0,7^2 = 0,49$.
4. Определяем частоту генотипа *Aa*: $2pq = 2 \cdot 0,7 \cdot 0,3 = 0,42$.

Структура популяции будет выглядеть следующим образом:

$$0,49AA + 0,42Aa + 0,09aa = 1.$$

Пример 2. Рассчитать частоты аллелей и генотипов по локусу гемоглобина в стаде крупного рогатого скота при условии:

n = 1000 голов;

с генотипом HbA/HbA – 300 голов;

с генотипом HbB/HbB – 650 голов;

с генотипом HbA/HbB – 50 голов.

Методика выполнения

1. Определяем частоты аллелей *A* и *B*:

$$pA = \frac{2nAA + nAB}{2N} = \frac{2 \cdot 300 + 50}{2 \cdot 1000} = 0,3250 ;$$

$$qB = \frac{2nBB + nAB}{2N} = \frac{2 \cdot 650 + 50}{2 \cdot 1000} = 0,6750 ;$$

$$pA + qB = 0,3250 + 0,6750 = 1.$$

2. Рассчитываем статистическую ошибку:

$$mp = mq = \sqrt{\frac{q(1-q)}{2N}} = \sqrt{\frac{p \cdot q}{2N}} = \sqrt{\frac{0,3250 \cdot 0,6750}{2 \cdot 1000}} = \sqrt{0,00019} = 0,014 ,$$

$$pA \pm mq = 0,3250 \pm 0,014,$$

$$qB \pm mq = 0,6750 \pm 0,014.$$

3. Определяем критерий достоверности:

$$tdp = tdq = \frac{0,6750 - 0,3250}{\sqrt{0,014^2 + 0,014^2}} = \frac{0,35}{0,019} = 18,42 .$$

По таблице стандартных значений критерия Стьюдента 18,42 находим с учетом степеней свободы $td > 3,29$. Если ошибка более, чем в 3 раза меньше по-

казателя частоты, то приведенные выше частоты аллелей статистически достоверны.

4. Находим частоты генотипов в популяции:

$$p_{AA} = \frac{n_{AA}}{N} = \frac{300}{1000} = 0,300; \quad (92)$$

$$q_{BB} = \frac{n_{BB}}{N} = \frac{650}{1000} = 0,650; \quad (93)$$

$$z_{AB} = \frac{n_{AB}}{N} = \frac{50}{1000} = 0,050. \quad (94)$$

5. Проверяем состояние данной популяции на генное равновесие:

$$p^2 q^2 = \left(\frac{2pq}{2} \right)^2 = 300 \cdot 650 = \left(\frac{50}{2} \right)^2. \quad (95)$$

$195000 \neq 625$. В этом случае генное равновесие нарушено.

Пример 3. В обследованной популяции частота людей с группой крови **A** составила 0,15, с группой крови **B** – 0,48, с группой крови **0** – 0,01, с группой крови **AB** – 0,36. Определить частоты аллелей **A**, **B**, **0**.

Методика выполнения

A - 0,15

B - 0,48

0 – 0,01

AB – 0,36.

В **A** входят и **AA**, и **A0**,

поэтому частота **A** = $\sqrt{0,15 + 0,01} - \sqrt{0,01} = 0,4 - 0,1 = 0,3$.

В **B** входят и **BB**, и **B0**, поэтому частота **B** =

$= \sqrt{0,48 + 0,01} - \sqrt{0,01} = 0,6$.

Частота **0** = $\sqrt{0,01} = 0,1$.

Частоты аллелей : **A** - 0,3; **B** - 0,6; **0** – 0,01.

Контрольные вопросы:

1. Основные свойства панмиктической популяции.
2. Какими параметрами характеризуется генетическая структура популяции?
3. Какие существуют причины нарушения генного равновесия?

Тема 16. ДИСПЕРСИОННЫЙ АНАЛИЗ

Цель занятия: освоение метода дисперсионного анализа и приобретение навыка по использованию этого метода в селекции и ветеринарии.

Основные понятия

Дисперсия – это рассеяние или отклонение от среднего.

Сущность дисперсионного анализа – установление роли отдельных факторов в изменчивости признака.

Задача дисперсионного анализа – определение доли влияния на варьирующий признак каждого учтенного и неучтенного (или случайного) фактора, а также установление достоверности этого влияния.

Типы дисперсионного анализа:

- 1) однофакторный – изучение влияния одного фактора на признак;
- 2) двухфакторный – изучение влияния двух факторов;
- 3) многофакторный – изучение влияния нескольких факторов.

В дисперсионном анализе используют *три типа дисперсий*:

1. *Дисперсия общая (C_y)* – обусловлена влиянием всего многообразия факторов.

Общая дисперсия по всей выборке σ_y^2 может разделяться на:

- дисперсия, обусловленная действием одного фактора.
- дисперсия, обусловленная действием ряда факторов ($\sigma_x^2, \sigma_A^2, \sigma_B^2$).
- дисперсия, обусловленная совместным действием нескольких факторов (σ_{AB}^2).

2. *Дисперсия факториальная (C_x)* или межгрупповая – $\sigma_x^2, \sigma_A^2, \sigma_B^2, \sigma_{AB}^2$ и др. обусловлена влиянием изучаемого фактора.

3. *Дисперсия случайная (C_z)* внутригрупповая (остаточная) дисперсия σ_z^2 – обусловлена влиянием неучтенных факторов.

Пример 1. Определить долю влияния на плодовитость маток хряков-производителей разных линий.

Методика выполнения

Расчет однофакторного дисперсионного анализа

При определении доли влияния изучаемого фактора на изменчивость признака варианты распределяют по градациям, на которые подразделяется этот фактор, например, по линиям, по сезонам, по возрасту и т. д.

Матки № п/п	Градации фактора							
	линия Дельфина		линия Смарагда		линия Драчуна		линия Самсона	
	число поро- сят в помете x_1	$(x_1 - \bar{X}_1)^2$	число поросят в поме- те x_2	$(x_2 - \bar{X}_2)^2$	число поро- сят в помете x_3	$(x_3 - \bar{X}_3)^2$	число поро- сят в помете x_4	$(x_4 - \bar{X}_4)^2$
1	12	1	11	0,36	10	0,25	14	2,25
2	12	1	10	2,56	10	0,25	11	2,25
3	13	4	12	0,16	10	0,25	14	2,25
4	11	0	12	0,16	11	0,25	13	0,25
5	10	1	13	1,96	10	0,25	13	0,25
6	10	1	13	1,96	11	0,25	11	2,25

Матки № п/п	Градации фактора							
	линия Дельфина		линия Смарагда		линия Драчуна		линия Самсона	
	число поро- сят в помете x_1	$(x_1 - \bar{X}_1)^2$	число поросят в поме- те x_2	$(x_2 - \bar{X}_2)^2$	число поро- сят в помете x_3	$(x_3 - \bar{X}_3)^2$	число поро- сят в помете x_4	$(x_4 - \bar{X}_4)^2$
7	11	0	11	0,36	11	0,25	13	0,25
8	10	1	11	0,36	11	0,25	12	0,25
9	10	1	12	0,16	11	0,25	12	0,25
10	11	0	11	0,36	10	0,25	12	0,25
\bar{X}_i	11,0		11,6		10,5		12,5	
$\sum(x_i - \bar{X}_i)^2$		10		8,40		2,50		10,5
$\bar{X}_i - \bar{X}_0$	-0,4		0,2		-0,9		1,1	
$(\bar{X}_i - \bar{X}_0)^2$	0,16		0,04		0,81		1,21	

1. Находят среднюю арифметическую для всех вариантов – \bar{X}_0 . При этом сумму величин всех вариантов делят на их число.
2. Находят среднюю арифметическую для каждой группы $\bar{X}_1, \bar{X}_2, \bar{X}_3$ (и т. д.).
3. Определяют отклонения средних арифметических каждой группы от общей средней ($\bar{X}_1 - \bar{X}_0; \bar{X}_2 - \bar{X}_0; \bar{X}_3 - \bar{X}_0$ и т. д.). Возводят их в квадрат и суммируют, получается сумма квадратов для групповых средних арифметических $\sum(\bar{X}_i - \bar{X}_0)^2$. Находят факториальную дисперсию (Cx) по формуле 96:

$$Cx = \sum (\bar{X}_i - \bar{X}_0)^2 \cdot n_0, \quad (96)$$

где \sum – знак суммы;
 \bar{X}_i – средняя арифметическая по каждой группе;
 \bar{X}_0 – средняя арифметическая для всех вариантов.

4. Чтобы уравнивать численность вариантов при нахождении суммы квадратов отклонений средних арифметических и суммы квадратов отклонений от средних внутри групп, сумму квадратов отклонений средних арифметических умножают на число вариантов в группе, если оно во всех группах одинаково. Если число вариантов в каждой группе разное, то сумму квадратов отклонений средних арифметических умножают на n_0 (формула 97):

$$n_0 = \frac{1}{a-1} \times \left(N - \frac{\sum n_i^2}{N} \right), \quad (97)$$

где \sum – знак суммы;

a – число градаций;
 n_i – число вариантов в каждой градации;
 N – число вариантов во всех группах.

5. Находят отклонение каждой варианты группы от ее средней арифметической, после этого величины возводят в квадрат и суммируют вначале по каждой группе, а затем по всем группам. Получается сумма квадратов для вариантов $\sum(x_i - \bar{X}_i)^2$. Находят внутригрупповую дисперсию (C_z) по формуле 98:

$$C_z = \sum \sum (x_i - \bar{X}_i)^2, \quad (98)$$

где \sum – знак суммы;
 x_i – значение варианты;
 \bar{X}_i – средняя арифметическая по каждой группе.

6. Находят общую дисперсию (C_y) по формуле 99:

$$C_y = C_x + C_z, \quad (99)$$

где C_x – факториальная (межгрупповая) дисперсия;
 C_z – внутригрупповая дисперсия.

Для этого складывают сумму квадратов для групповой изменчивости (C_x) и сумму квадратов для внутригрупповой изменчивости (C_z).

Источник варьирования	Число степеней свободы	Сумма квадратов	Средний квадрат
Группы	4 – 1 = 3	22,2	7,4
Объекты	40 – 4 = 36	31,4	0,87
Общее	39	53,6	

Находят долю влияния изучаемого фактора на результирующий признак по формуле 100:

$$\eta_x^2 = \frac{C_x}{C_y}, \quad (100)$$

где C_x – факториальная (межгрупповая) дисперсия;
 C_y – общая дисперсия.

Для этого делят сумму квадратов межгрупповой изменчивости (C_x) на (C_y), получают долю влияния изучаемого фактора на изменчивость признака.

Влияние отцов на изменчивость плодовитости маток составляет 41,4% ($22,2 : 53,6 = 0,414$) от общей изменчивости. На долю прочих факторов и случайности выборки приходится 58,6% ($31,4 : 53,6 = 0,586$), что указывает на значительное влияние линий хряков на изменчивость плодовитости маток.

7. Находят внутригрупповую (случайную) (σ_z^2) и межгрупповую (факториальную) (σ_x^2) дисперсии.

Для нахождения *вариансы внутригрупповой дисперсии* сумму квадратов отклонений отдельных вариантов группы от их средних арифметических делят на сумму чисел степеней свободы каждой группы, которая меньше общего количества вариантов на число групп.

Чтобы найти *вариансу межгрупповой дисперсии*, сумму квадратов отклонений средних арифметических от общей средней корректируют на число вариантов в группе и делят на число степеней свободы, которое на единицу меньше числа групп.

8. Рассчитывают достоверность доли влияния изучаемого фактора на изменчивость признака с помощью критерия Фишера по формуле 101:

$$F_{\text{эмп.}} = \frac{\sigma_x^2}{\sigma_z^2}. \quad (101)$$

В таблице (приложение 3) с учетом степеней свободы находим значение *F теорет.* и сравниваем с полученным нами *F эмпирическим*.

Если $F_{\text{эмп.}} > F_{\text{теор.}}$, то влияние изученного фактора доказано и различия между группами существенны.

9. Для контроля за правильностью расчетов вычисляют квадрат отклонений каждой варианты от общей средней X_0 и полученные величины суммируют. Результат должен быть равен сумме квадратов отклонений от средних арифметических групп и сумме квадратов отклонений вариантов, а число степеней должно быть равно сумме соответствующих степеней свободы групп и вариант [2, 3, 8, 9, 11].

Контрольные вопросы:

1. Использование дисперсионного анализа в селекционной работе.
2. Основные понятия, условия и типы дисперсионного анализа.
3. Этапы дисперсионного анализа.
4. Как определяется доля влияния отдельных факторов и уровень их достоверности?

ТЕМА 17. ОБРАБОТКА ВАРИАЦИОННЫХ РЯДОВ НА ПЭВМ. АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ

Цель занятия: освоение метода биометрической обработки цифровых данных и приобретение навыка по использованию этого метода для анализа в селекции и ветеринарии.

Основные понятия

При проведении биометрической обработки данных при помощи ПЭВМ рекомендуется использовать в MS EXCEL опцию «Мастер функций» или пакет «Анализ данных».

Расчет популяционно-генетических параметров следует проводить по нижеприведенным формулам.

Среднее арифметическое значение количественных признаков (\bar{x}) определяют по формуле:

$$\bar{X} = \frac{\sum X_i}{n}, \quad (102)$$

где X_i – величина варьирующего признака;
 n – объем выборки, гол.

В MS EXCEL следует использовать в «Мастере функций» (f_x) функцию СРЗНАЧ (рис. 11).

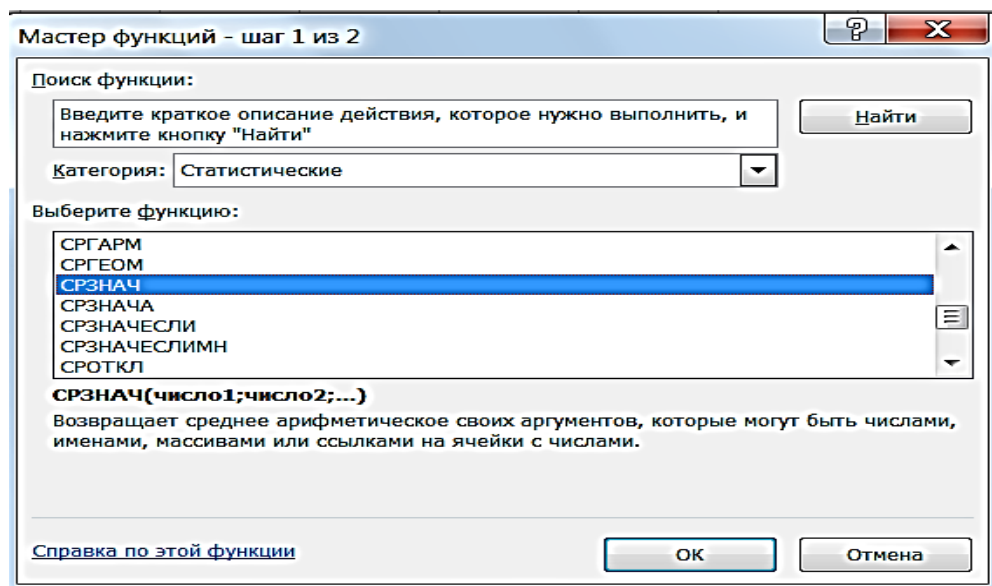


Рисунок 11 – Определение \bar{X} с помощью MS EXCEL

Количество значений в выборке (n) можно определить при помощи MS EXCEL, для этого в «Мастере функций» (f_x) используют функцию СЧЕТ (рис. 12).

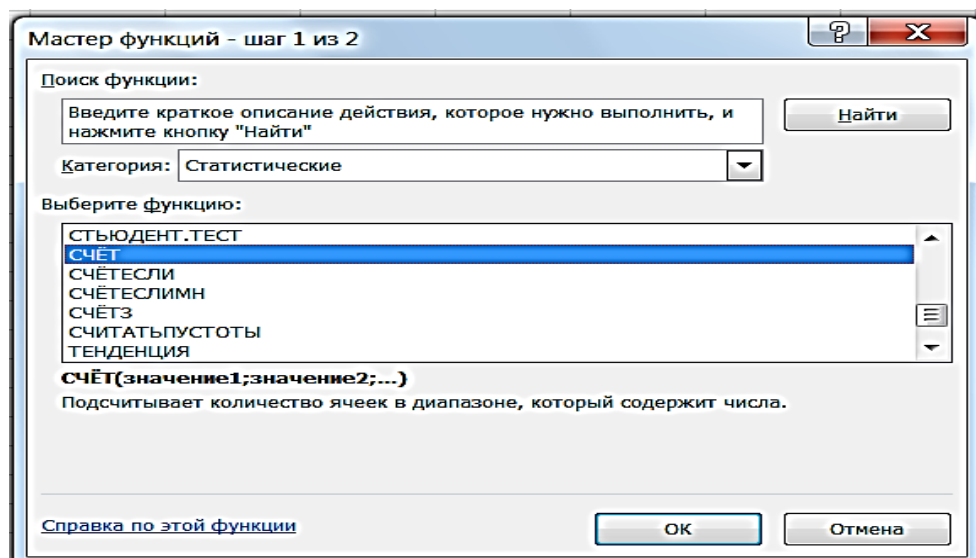


Рисунок 12 – Определение n с помощью MS EXCEL

Среднее квадратическое отклонение (σ) определяют по формуле 103:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}. \quad (103)$$

В MS EXCEL следует использовать в «Мастере функций» (f_x) функцию СТАНДОТКЛОН. В (рис. 13).

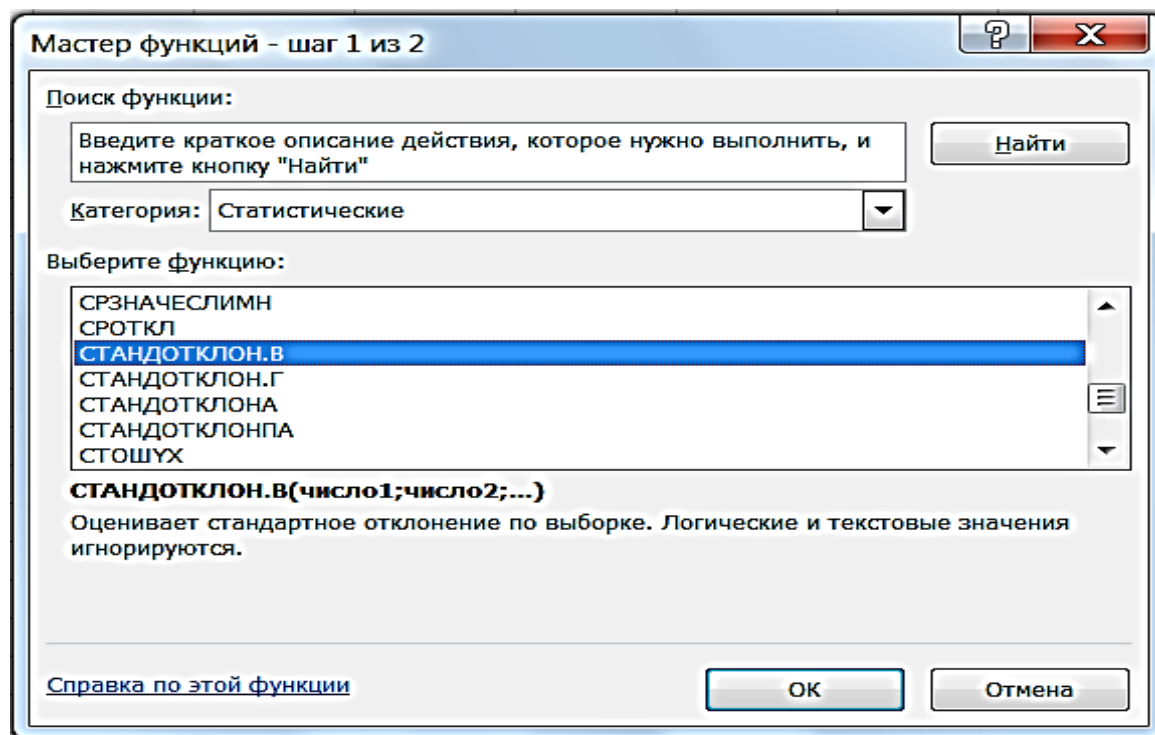


Рисунок 13 – Определение σ с помощью MS EXCEL

Ошибку средней арифметической ($m_{\bar{x}}$) вычисляют по формуле 104:

$$m_{\bar{x}} = \frac{\sigma}{\sqrt{n-1}}. \quad (104)$$

В MS EXCEL расчет ведут через строку формул, запись выглядит как $=\sigma/\text{корень}(n-1)$ (рис. 14).

Файл		Главная		Вставка		Разметка страницы		Формулы		Данные		Рецензирование		В	
Вставить		Буфер обмена		Шрифт		Выравнивание		Число		Общий					
СТАНДОТКЛОН.В \sqrt{x} \checkmark f_x =B183/КОРЕНЬ(B181)															
	A	B	C	D	E	F	G	H							
172		8195	4,02	3,27	8195	4,02	3,27								
173		7694	3,77	3,25	7694	3,77	3,25								
174		6501	3,65	3,51	7888	3,67	3,21								
175		8590	3,5	3,15	8590	3,5	3,15								
176		7590	4,22	3,2	7590	4,22	3,2								
177		7241	4,31	3,77	7907	4,25	3,55								
178		8302	4	3,26	8302	4	3,26								
179		6806	4,75	3,94	8294	3,8	3,64								
180															
181	n	174													
182	X	6948,42													
183	σ	1233,012													
184		=B183/КОРЕНЬ(B181)													
185															

Рисунок 14 – Определение m_x с помощью MS EXCEL
Коэффициент изменчивости (C_v) определяют по формуле 105:

$$C_v = \frac{\sigma}{X} \cdot 100 \% . \quad (105)$$

В MS EXCEL расчет ведут через строку формул, запись выглядит как = $\sigma/\bar{X} \cdot 100$ (рис. 15).

Файл		Главная		Вставка		Разметка страницы		Формулы		Данные		Рецензирование	
Вставить		Буфер обмена		Шрифт		Выравнивание		Общий		Число			
СТАНДОТКЛОН.В \sqrt{x} \checkmark f_x =B183/B182*100													
	A	B	C	D	E	F	G						
172		8195	4,02	3,27	8195	4,02	3,27						
173		7694	3,77	3,25	7694	3,77	3,25						
174		6501	3,65	3,51	7888	3,67	3,21						
175		8590	3,5	3,15	8590	3,5	3,15						
176		7590	4,22	3,2	7590	4,22	3,2						
177		7241	4,31	3,77	7907	4,25	3,55						
178		8302	4	3,26	8302	4	3,26						
179		6806	4,75	3,94	8294	3,8	3,64						
180													
181	n	174											
182	X	6948,42											
183	σ	1233,012											
184	m	93,47443											
185		=B183/B182*100											
186													

Рисунок 15 – Определение C_v с помощью MS EXCEL

Коэффициент корреляции (r) определяют через «Мастер функций» (f_x) MS EXCEL, функция КОРРЕЛ (рис. 16, 17).

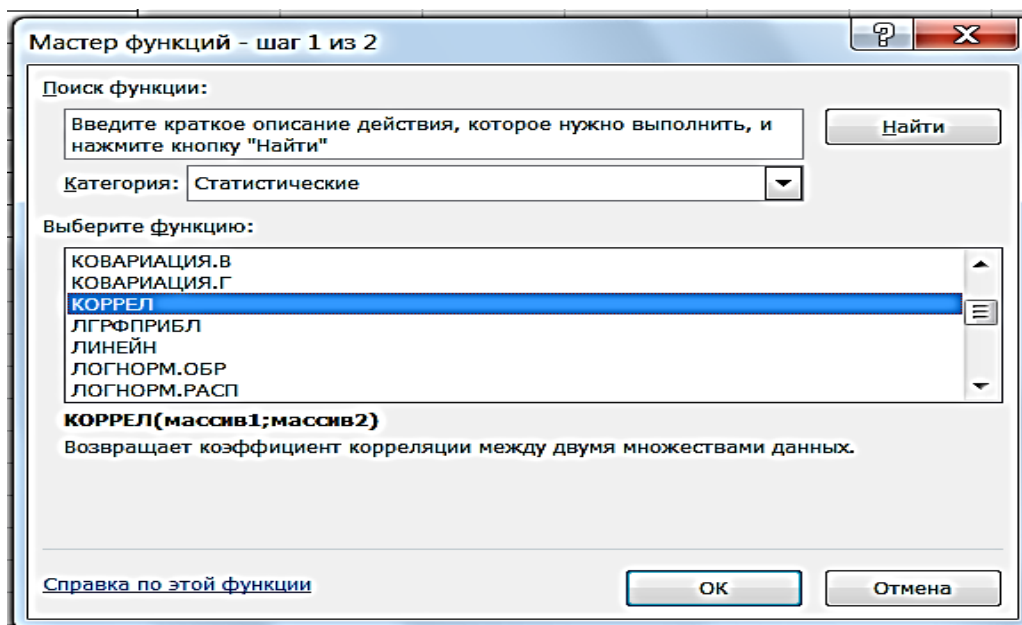


Рисунок 16 – Определение r с помощью MS EXCEL (шаг 1)

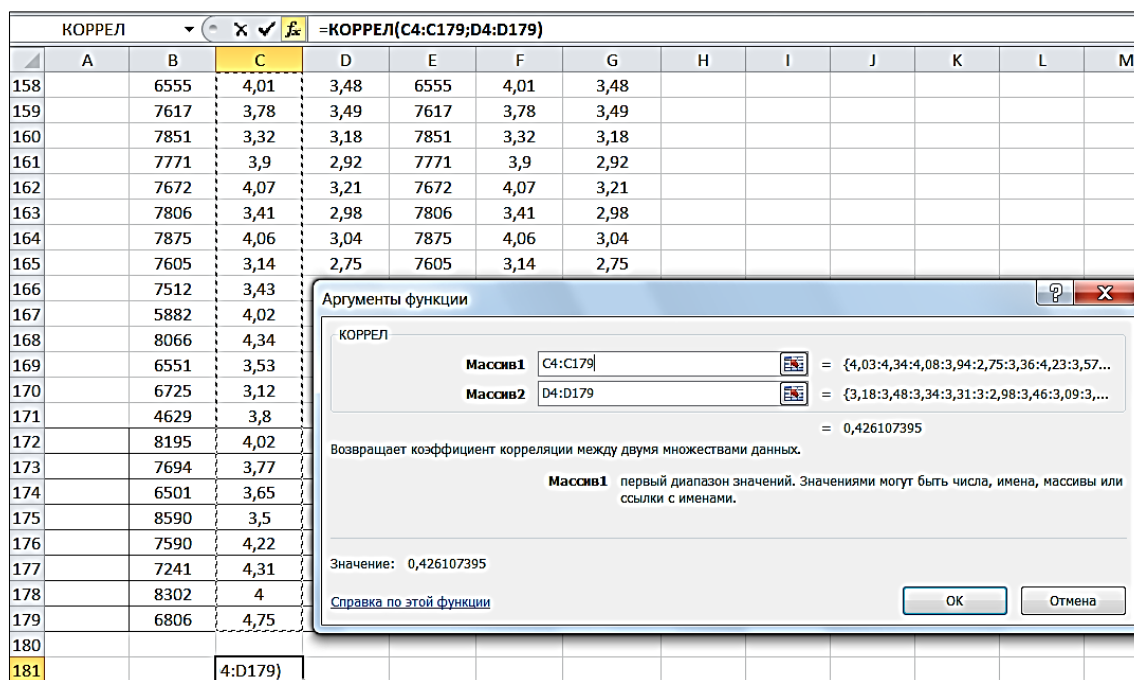


Рисунок 17 – Определение r с помощью MS EXCEL (шаг 2)

Контрольные вопросы:

1. Как проводится биометрическая обработка первичных цифровых данных?
2. Какие популяционно-генетические параметры можно определить с помощью MS EXCEL?

СТАНДАРТНЫЕ ЗНАЧЕНИЯ КРИТЕРИЯ t_d (по СТЬЮДЕНТУ)

Число степеней свободы	Вероятность (P)		
	0,95	0,99	0,999
1	12,7	63,66	-
2	4,30	9,93	31,60
3	3,18	5,84	12,94
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,75	3,66
30	2,04	2,75	3,65
31	1,96	2,58	3,29

СТАНДАРТНЫЕ ЗНАЧЕНИЯ χ^2 ПРИ РАЗНЫХ УРОВНЯХ ЗНАЧИМОСТИ

Степени свободы ($\nu = n - 1$)	Уровень значимости			Степени свободы ($\nu = n - 1$)	Уровень значимости		
	0,05	0,01	0,001		0,05	0,01	0,001
1	3,8	6,6	10,8	26	38,9	45,6	54,1
2	6,0	9,2	13,8	27	40,1	47,0	55,5
3	7,8	11,3	16,3	28	41,3	48,3	56,9
4	9,5	13,3	18,5	29	42,6	49,6	58,3
5	11,1	15,1	20,5	30	43,8	50,9	59,7
6	12,6	16,8	22,5	32	46,2	53,5	62,4
7	14,1	18,5	24,3	34	48,6	56,0	65,2
8	15,5	20,1	26,1	36	51,0	58,6	67,9
9	16,9	21,7	27,9	38	53,4	61,1	70,7
10	18,3	23,2	29,6	40	55,8	63,7	73,4
11	19,7	24,7	31,3	42	58,1	66,2	76,1
12	21,0	26,2	32,9	44	60,5	68,7	78,7
13	22,4	27,7	34,5	46	62,8	71,2	81,4
14	23,7	29,1	36,1	48	65,2	73,7	84,0
15	25,0	30,6	37,7	50	67,5	76,2	86,7
16	26,3	32,0	39,3	55	73,3	82,3	93,2
17	27,6	33,4	40,8	60	79,1	88,4	99,6
18	28,9	34,8	42,3	65	84,8	94,4	106,0
19	30,1	36,2	43,8	70	90,5	100,4	112,3
20	31,4	37,6	45,3	75	96,2	106,4	118,5
21	32,7	38,9	46,8	80	101,9	112,3	124,8
22	33,9	40,3	48,3	85	107,5	118,2	131,0
23	35,2	41,6	49,7	90	113,1	124,1	137,1
24	36,4	43,0	51,2	95	118,7	130,0	143,3
25	37,7	44,3	52,6	100	124,3	135,8	149,4

СТАНДАРТНЫЕ ЗНАЧЕНИЯ КРИТЕРИЯ ФИШЕРА $F = \sigma_x^2 / \sigma_z^2$

$V_2 \backslash V_1$	1	2	3	4
3	167,5	148,5	141,1	137,1
	34,1	30,8	29,5	28,7
	10,1	9,6	9,3	9,1
4	74,1	61,2	56,1	53,4
	21,2	18,8	16,7	16,0
	7,7	6,9	6,6	6,4
5	47,0	36,6	33,2	31,1
	16,3	13,3	12,1	11,4
	6,6	5,8	5,4	5,8
6	35,5	27,0	23,7	21,9
	13,4	10,9	9,8	9,2
	6,0	5,1	4,8	4,5
7	29,2	21,7	18,8	17,2
	12,3	9,6	8,5	7,9
	5,6	4,7	4,4	4,1
8	25,4	18,5	15,8	14,4
	11,3	8,7	7,6	7,0
	5,3	4,6	4,1	3,8
9	22,9	16,4	13,9	12,6
	10,6	8,0	7,0	6,4
	5,1	4,8	3,6	3,6
10	21,0	14,9	12,3	11,3
	10,0	7,9	6,6	6,0
	5,0	4,1	3,7	3,5
11	19,7	13,8	11,6	10,4
	9,7	7,2	6,2	5,7
	4,8	4,0	3,6	3,4
12	18,6	12,3	10,8	9,6
	9,3	6,9	6,0	5,4
	4,8	3,9	3,5	3,3
13	17,8	12,3	10,2	9,1
	9,1	6,7	5,7	5,2
	4,7	3,8	3,4	3,2
14	17,1	11,8	9,7	8,6
	8,9	6,5	5,6	5,0
	4,6	3,7	3,3	3,1
15	16,6	11,3	9,3	8,3
	8,7	6,4	5,4	4,9
	4,5	3,7	3,3	3,1

16	16,1	11,0	9,0	7,9
	8,5	6,2	5,3	4,8
	4,5	3,6	3,2	3,0
17	15,7	10,7	8,7	7,7
	8,4	6,1	5,1	4,7
	4,5	3,6	3,2	3,0
18	15,4	10,4	8,5	7,5
	8,3	6,0	5,1	4,6
	4,4	3,5	3,2	2,9
19	15,1	10,2	8,3	7,3
	8,2	5,9	5,0	4,5
	4,4	3,5	3,1	2,9
20	14,8	10,0	8,1	7,1
	8,1	5,8	4,9	4,4
	4,3	3,5	3,1	2,9
21	14,6	9,8	7,9	7,0
	8,0	5,8	4,9	4,4
	4,3	3,5	3,1	2,8
22	14,4	9,6	7,8	6,8
	7,9	5,7	4,8	4,3
	4,3	3,4	3,0	2,8
23	14,2	9,5	7,7	6,7
	7,9	5,7	4,8	4,0
	4,3	3,4	3,0	2,8

ЛИТЕРАТУРА

1. Бакай, А. В. Биометрические методы анализа качественных и количественных признаков / А. В. Бакай, И. И. Кочиш, Г. Г. Скрипниченко // Генетика : учебник для студентов вузов по специальности «Зоотехния» / А. В. Бакай, И. И. Кочиш, Г. Г. Скрипниченко. – Москва : КолосС, 2006. – С. 176–227.
2. Генетика : учебник для студентов высших учебных заведений по специальности «Зоотехния» / Е. К. Меркурьева [и др.]. – Москва : Агропромиздат, 1991. – 446 с. : ил., табл.
3. Шацкий, А. Д. Генетика с основами биометрии : учебное пособие для студентов учреждений высшего образования по специальности «Зоотехния» / А. Д. Шацкий, М. А. Шацкий. – Минск : ИВЦ Минфина, 2015. – 304 с.
4. Давыдович, Е. В. Генетика с основами биометрии : методические указания к лабораторным занятиям для студентов зооинженерного факультета очного и заочного отделений специальностей 1-74 03 01 – Зоотехния, 1-74 03 03 – Промышленное рыбоводство : в 2 ч. / Е. В. Давыдович, Т. В. Павлова. – Горки, 2014. – Ч. 1 : Генетико-статистические методы биометрии. – 87 с.
5. Давыдович, Е. В. Генетика с основами биометрии : методические указания к лабораторным для студентов зооинженерного факультета очного и заочного отделений специальностей 1-74 03 01 – Зоотехния, 1-74 03 03 – Промышленное рыбоводство : в 2 ч. / Е. В. Давыдович, Т. В. Павлова. – Горки, 2014. – Ч. 2 : Измерение связи между количественными признаками. – 55 с.
6. Иванова, О. А. Генетика : учебник для зоотехнических и ветеринарных факультетов сельскохозяйственных вузов / О. А. Иванова. – 2-е изд., перераб. и доп. – Москва : Колос, 1974. – 431 с.
7. Лакин, Г. Ф. Биометрия : учебное пособие для студентов биологических специальностей вузов / Г. Ф. Лакин. – 4-е изд., перераб. и доп. – Москва : Высшая школа, 1990. – 352 с.
8. Ларцева, С. Х. Практикум по генетике : учебное пособие для студентов высших сельскохозяйственных учебных заведений по специальности «Зоотехния» / С. Х. Ларцева, М. К. Муксинов. – Москва : Агропромиздат, 1985. – 288 с.
9. Меркурьева, Е. К. Генетика с основами биометрии : учебное пособие для студентов высших сельскохозяйственных учебных заведений по специальности «Зоотехния» / Е. К. Меркурьева, Г. Н. Шангин-Березовский. – Москва : Колос, 1983. – 400 с.
10. Меркурьева, Е. К. Генетические основы селекции в скотоводстве : учебное пособие для слушателей факультетов повышения квалификации сельскохозяйственных кадров / Е. К. Меркурьева. – Москва : Колос, 1977. – 239 с.

11. Петухов, В. Л. Генетика = Genetics : учебник / В. Л. Петухов, О. С. Короткевич, С. Ж. Стамбеков ; Семипалатинский государственный педагогический институт. – 2-е изд., испр. и доп. – Новосибирск : СемГПИ, 2007. – 628 с. : табл., ил.
12. Плохинский, Н. А. Биометрия : [учебное пособие для студентов биологических специальностей университетов] / Н. А. Плохинский. – 2-е изд. – Москва : Издательство Московского университета, 1970. – 368 с.