

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РФ
ФГБОУ ВО КОСТРОМСКАЯ ГСХА

Кафедра частной зоотехнии, разведения и генетики

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ
ПО ВЫПОЛНЕНИЮ КОНТРОЛЬНЫХ РАБОТ

2-е издание, исправленное и дополненное

*Для контактной и самостоятельной работы студентов,
обучающихся по направлению подготовки 36.04.02 Зоотехния*

КАРАБАЕВО
Костромская ГСХА
2021

УДК 51-76
ББК 28 + 22.172
М 34

Составители: сотрудники Костромской ГСХА канд. с.-х. наук, доцент, доцент кафедры частной зоотехнии, разведения и генетики *С.Г. Белокуров*, канд. ветеринар. наук, доцент, заведующий кафедрой эпизоотологии, паразитологии и микробиологии *М.С. Трескин*.

Рецензент: канд. ветеринар. наук, доцент, заведующий кафедрой внутренних незаразных болезней, хирургии и акушерства Костромской ГСХА *В.В. Решетняк*.

Рекомендовано методической комиссией факультета ветеринарной медицины и зоотехнии в качестве учебно-методического пособия по выполнению контрольных работ для контактной и самостоятельной работы студентов, обучающихся по направлению подготовки 36.04.02 Зоотехния

М 34 Математические методы в биологии : учебно-методическое пособие по выполнению контрольных работ / сост. С.Г. Белокуров, М.С. Трескин. — 2-е изд., исправл. и доп. — Караваево : Костромская ГСХА, 2021. — 60 с. ; 20 см. — 50 экз. — Текст непосредственный.

В общедоступной форме изложены материалы, необходимые для проведения практических занятий и самостоятельного изучения дисциплины, методики вычисления основных статистических параметров и коэффициентов, применяемых в зоотехнических и ветеринарных исследованиях, разведении и селекции сельскохозяйственных животных на примере больших и малых выборок, при альтернативной изменчивости. Представлены вспомогательные таблицы по определению необходимого объема совокупности, достоверности статистических параметров, и достоверности разности двух независимых рядов. Показаны приемы вычисления и пользования критерия хи-квадрат. Разобрана методика проведения дисперсионного анализа.

Учебно-методическое пособие по предназначено для контактной и самостоятельной работы магистров направления подготовки 36.04.02 Зоотехния.

УДК 51-76
ББК 28 + 22.172

ОГЛАВЛЕНИЕ

Введение	5
1. Содержание дисциплины	8
2. Понятие о генеральной и выборочной совокупности.....	13
3. Статистический анализ в многочисленных выборках.....	14
3.1. Вариационный ряд и вариационная кривая	14
3.2. Средняя арифметическая (\bar{X}).....	19
3.3. Среднее квадратическое (стандартное) отклонение (σ)	21
3.4. Коэффициент вариации (C_v)	22
3.5. Изучение взаимосвязи между признаками.....	23
3.6. Оценка достоверности статистических показателей	28
3.7. Критерий хи-квадрат (χ^2)	30
4. Статистический анализ в малочисленных выборках	32
4.1. Среднее арифметическое, стандартное отклонение, коэффициент вариации, ошибка среднего арифметического.....	32
4.2. Коэффициент корреляции	33
5. Статистический анализ альтернативных признаков.....	35
5.1. Определение ошибки выборочной доли.....	36
5.2. Вычисление достоверности разности между выборочными долями или процентами:.....	36
5.3. Коэффициент корреляции	37
6. Дисперсионный анализ	38
6.1. Техника расчетов при дисперсионном анализе однофакторного комплекса для малых групп	39
6.2. Использование Microsoft Excel для проведения дисперсионного анализа	42
7. Краткий словарь терминов, применяемых в математической биостатистике.....	46
8. Контрольные вопросы.....	48
9. Рекомендуемая литература	51
Приложение	52

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

x	значения признака у отдельного объекта (варианта)
i	индекс варианты ($1, 2, \dots, n$),
n	объем выборочной совокупности
N	объем генеральной совокупности
μ, \tilde{x}	средняя арифметическая генеральной совокупности
M, \bar{X}	средняя арифметическая выборочной совокупности
$M_{ВЗВ}, \bar{X}$	средняя взвешенная выборочной совокупности
G, \bar{x}_g	средняя геометрическая
H, \bar{x}_h	средняя гармоническая
S, \bar{x}_q	средняя квадратическая
Q, \bar{x}_Q	средняя квадратическая
Mo	мода
Me	медиана
p, q	доля отношение числа объектов с признаком (а) к общему объему выборки
s_x, σ	среднее квадратичное (стандартное) отклонение
s_x^2, σ^2	выборочная дисперсия (варианса)
C_v	коэффициент вариации (изменчивости)
C_s	показатель точности оценки средней величины
lim	лимиты указания минимального и максимального значения признака в совокупности
t	нормированное отклонение
r	коэффициент корреляции
$R_{x/y}$	коэффициент регрессии
r_w	коэффициент повторяемости
h^2	коэффициент наследуемости
γ_x^2	квадрат корреляционного отношения, основной показатель силы влияния
s, m	статистическая ошибка, с индексом соответствующим параметру, например: $s_{\bar{X}}(m_M)$ — ошибка средней арифметической, m_D — ошибка разницы
t	критерий достоверности
df, v, k	число степеней свободы
p	уровень вероятности безошибочного прогноза
α	уровень значимости
χ^2	критерий соответствия распределений
Σ	знак суммирования

ВВЕДЕНИЕ

Целью освоения дисциплины «Математические методы в биологии» является расширение и углубление базовых знаний и навыков по вопросам выбора и применения математических и статистических методов обработки экспериментальных данных в биологии, что позволит магистру обладать универсальными и профессиональными компетенциями, способствующими его успешной исследовательской деятельности

Задачи курса:

– изучить математическую основу алгоритмов, используемых в биологических исследованиях;

– научиться составлять репрезентативные выборки, адекватно выбирать методы обработки экспериментальных данных;

– овладеть методами обработки результатов эксперимента;

– научиться формулировать и проверять статистические гипотезы.

В результате освоения дисциплины аспирант должен

знать: классические и современные математические и статистические методы, основные математические модели, используемые в биологии;

уметь: производить статистическую обработку результатов эксперимента, устанавливать характер и тип распределения объектов с разными параметрами признака, выявлять изменчивость признака, оценивать значимость различия показателей в разных совокупностях, определять величину и направление связи между переменными величинами признаков объектов совокупности, изучать степень влияния того или иного фактора на изменчивость анализируемого признака и прогнозировать показатели-отклики при заданных значениях воздействующих факторов, формулировать и проверять выдвигаемые статистические гипотезы, организовать и провести научный эксперимент, обобщать результаты опыта и формулировать выводы;

владеть: современными математическими методами, используемыми в биологических исследованиях, навыками работы с современными пакетами прикладных программ, используемых для статистической обработки данных экспериментов в области биологии.

Развитие современной биологической науки, сопровождается накоплением большого количества информации по многим вопросам общей, прикладной генетики и селекции. При этом остро стоит проблема классификации данных, их упорядочение и систематизация, а так же научно-практический анализ, завершающийся формулировкой

реальных предложений для дальнейшего развития той или иной отрасли.

Специалистам (ветеринарным врачам, зоотехникам) в своей повседневной производственной деятельности приходится иметь дело с большими группами животных: вид, порода, стадо, линии, семейства и др., которые характеризуются большим разнообразием признаков, позволяющим отличать их друг от друга. В то же время, отдельные животные, входящие в состав этих групп, так же различаются между собой, но в целом им свойственны сходные признаки, типичные для вида, породы, стада, линии, семейства и др. Таким образом, при изучении и работе с большими группами животных выясняется значительная изменчивость признаков, что создает большие трудности для их характеристики и анализа. Наряду с этим установлено, что все биологические явления, свойства и признаки у множества особей одного вида имеют статистическую природу, т.е. подчиняются математическим закономерностям, Это вызвало необходимость их приложения к изучению биологических объектов.

Союз математики и биологии составляет содержание биометрии науки о статистическом анализе групповых свойств биологических объектов. Термин «биометрия» (от лат. *bios* жизнь и *metron* мера) введен в 1889 г. английским ученым Ф. Гальюном, впервые применившим математический анализ при изучении явлений изменчивости и наследственности.

Затем Г. Дункер предложил другое название — «вариационная статистика» (от лат. *variatio* — изменение, колебание и *status* — состояние, положение вещей), — которое также вошло в обиход. Так появились разные названия одного и того же предмета, хотя буквальный смысл их не одинаков. Оба термина недостаточно точны. Отсюда попытки заменить их термином «биологическая статистика» (А. В. Леонтович, 1909; П. Ф. Рокицкий, 1973) или математическая биостатистика который в настоящее время является наиболее часто используемым и введенным в курс обязательных дисциплин в образовательный стандарт специальности «Ветеринария». В дальнейшем описании, мы будем использовать все выше указанные термины, как синонимы.

Современная биометрия — это особый раздел биологии и генетики, позволяющий использовать не только присущий им комплекс специфических методов исследований, но и в значительной мере опирающийся на различные математические приемы анализа, оценки и характеристики признаков и свойств живых объектов. Прежде всего, это относится в оценке продуктивных и племенных качеств животных,

доли влияния на признак различных факторов в том числе наследственности и среды, при вычислении величин характеризующих связи между признаками, к оценке эффективности методов селекции и прогнозирования ее результатов, используя для этого сравнительно небольшие выборки а выводы, полученные при их изучении с достаточной точностью и надежностью распространить на большие по численности линии, семейства, стада, породы, виды.

Цель биометрического анализа: получение комплекса статистических параметров и коэффициентов, характеризующих изучаемую совокупность по одному или нескольким признакам.

Задачи биометрического анализа:

1. Выявить особенности и типы варьирования количественных и качественных признаков, характер распределения особей с разным уровнем признака.

2. Определить степень фенотипического уровня проявления признаку особей в совокупности путем вычисления средних величин (M , G , S , H , M_0 , M_e).

3. Установить степень фенотипической и генотипической изменчивости признаков с помощью σ , Cv , σ^2 , lim , t .

4. Изучить величину фенотипической и генотипической взаимосвязи между различными признаками и ее направление с использованием r , R и др.

5. Сравнить группы по величине средних с применением метода статистических ошибок, критериев достоверности (t_D , χ^2 и др.).

6. Определить долю влияния различных факторов на фенотипическую и генотипическую изменчивость признака с использованием дисперсионного и факторного анализа.

7. Дать характеристику изучаемой популяции по комплексу генетических и статистических параметров, степени гомо- и гетерозиготности, h^2 , проявлению гетерозиса и инбредной депрессии, типу наследования признаков.

Предмет биометрического анализа — изменчивость изучаемого признака в совокупности.

Источники информации. Наиболее распространенным источником, характеризующим практически все массивы сельскохозяйственных животных, являются первичные данные зоотехнического, племенного и ветеринарного учета, которые систематически заполняются специалистами хозяйств по установленной форме в различных журналах,

ведомостях, племенных карточках, а также в электронном виде с помощью прикладных компьютерных программ, автоматизирующих систему статистических данных.

Другим источником информации служат данные научно-производственных опытов, проводимые на ограниченном числе животных, а так же результаты специальных экспериментов на лабораторных животных при изучении различных биологических показателей, При этом информация записывается в протокольных книгах, дневниках наблюдений, ведомостях и специальных бланках.

Первичным материалом могут служить фотографии животных; микрофотографии клеток тканей, кариотипов; мазков крови, срезов тканей и др.

1. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Предмет, методы и задачи дисциплины

Практические занятия

Первичная обработка экспериментальных данных. Проверка статистических гипотез: о соответствии эмпирического распределения объектов в совокупности теоретически ожидаемому; о равенстве математических ожиданий двух нормальных распределений с известными дисперсиями.

Вопросы для самостоятельной подготовки

1. Основные понятия теории вероятностей, случайные события.
2. Что такое вероятность? По какой формуле она вычисляется?
3. Условная вероятность и независимость событий.
4. Случайные величины и функция распределения вероятностей, дискретные случайные величины.
5. Непрерывные случайные величины и функция плотности вероятностей.
6. Характеристики распределений случайной величины
7. Примеры распределения случайных величин.
8. Что такое математическое ожидание?
9. Характеристики распределений случайной величины
10. Распределение Бернулли.
11. Биномиальное распределение.
12. Распределение Пуассона.
13. Непрерывное и дискретное равномерное распределение.
14. Нормальное распределение.
15. Распределение χ^2 .

16. Распределение Стьюдента
17. Распределение Фишера.
18. Понятие случайной выборки. Примеры реальных биологических экспериментов.
19. Для чего может применяться математическая статистика в биологии?
20. Перечислите основные описательные статистики, используемые в предварительном анализе данных.
21. Что такое генеральная совокупность?
22. Что такое выборки, как они составляются?
23. Как составляется вариационный ряд?
24. Какие бывают типы распределения и вариационных кривых?
25. Перечислите средние величины и их использование.
26. Как вычисляется средняя арифметическая величина в малых больших выборках?
27. Какими свойствами обладают средние величины?
28. Какие показатели характеризуют разнообразие признаков?
29. Как вычисляется среднее квадратическое отклонение в малых и больших выборках?
30. Как вычисляется среднее квадратическое отклонение для альтернативных признаков?
31. Что такое ошибки репрезентативности? Чем отличаются они от ошибок измерения и вычисления?
32. Что такое доверительные вероятности?
33. Какие доверительные вероятности можно использовать в биологических, зоотехнических и ветеринарных исследованиях?
34. Как определяется достоверность выборочных показателей?
35. Как оценивается достоверность разности между средними величинами двух выборок?

Раздел 2. Корреляционно-регрессионный анализ.

Практические занятия.

1. Корреляционно-регрессионный анализ: функциональная, стохастическая, корреляционная зависимости.
2. Оценка достоверности коэффициента корреляции; доверительные интервалы для коэффициентов корреляции.
3. Коэффициенты и уравнения регрессии; построение прогноза по уравнению регрессии и оценка его точности и надежности.

Вопросы для самостоятельной подготовки

1. Как вычисляется коэффициент фенотипической корреляции а малых и больших выборках?
2. В чем заключается различие связи между признаками при положительных и отрицательных значениях коэффициента корреляции?
3. Как вычисляется коэффициент корреляции для альтернативных признаков?
4. В каких случаях используется коэффициент ранговой корреляции?
5. Приведите формулы вычисления коэффициентов генетической корреляции.
6. Что характеризуют коэффициенты регрессии? В чем различие между коэффициентами $R_{x/y}$ и $R_{y/x}$
7. В чем различие между коэффициентами r и R ?

Индивидуальные домашние задания

1. Рассчитать основные параметры выборки, согласно индивидуального задания.
2. Рассчитать достоверность разности согласно индивидуального задания.
3. Построить кривые распределения согласно индивидуального задания.
4. Провести расчет коэффициент корреляции, согласно индивидуального задания.
5. Провести оценку достоверности коэффициента корреляции, согласно индивидуального задания.
6. Определить доверительные интервалы коэффициента корреляции, согласно индивидуального задания.
7. Построить прогноз по уравнению регрессии и произвести оценка его точности и надежности, согласно индивидуального задания.
8. Раздел 3. Дисперсионный анализ.

Практические занятия.

1. Дисперсионный анализ: анализ компонентов общего разнообразия: факториальное и случайное разнообразие. Однофакторный дисперсионный комплекс (фиксированная и случайная модели); критерий достоверности.
2. Организация и анализ многофакторного дисперсионного комплекса (фиксированная и случайная модели); коэффициент внутриклассовой корреляции

Вопросы для самостоятельной подготовки

1. В чем заключается цель дисперсионного анализа? Что называется общей, факториальной и остаточной дисперсией?
2. Какие бывают дисперсионные комплексы? Чем они характеризуются?
3. Однофакторный дисперсионный анализ.
4. Многофакторный дисперсионный анализ.
5. Как составляют однофакторный дисперсионный комплекс и вычисляют вспомогательные величины?
6. Какие показатели используются для оценки достоверности влияния изучаемого фактора?

Индивидуальные домашние задания

1. Произвести дисперсионный анализ количественных признаков однофакторного статистического комплекса, согласно индивидуального задания.
2. Произвести дисперсионный анализ количественных признаков двухфакторного статистического комплекса, согласно индивидуального задания.
3. Вычислить коэффициент внутриклассовой корреляции, согласно индивидуального задания.
4. Определить достоверность разницы между средними арифметическими дисперсионным методом
5. Определить достоверность разности между долями по градациям статистического комплекса качественных признаков

Раздел 4. Анализ качественных признаков.

Практические занятия

Анализ качественных признаков: вероятность, частоты, частности; малые частоты, ϕ -преобразование Фишера; организация и анализ дисперсионных комплексов по признакам с альтернативной изменчивостью.

Индивидуальные домашние задания

Произвести дисперсионный анализ качественных признаков однофакторного статистического комплекса, согласно индивидуального задания.

1. Произвести дисперсионный анализ качественных признаков двухфакторного статистического комплекса, согласно индивидуального задания.

2. Определить достоверность между долями методом угловой трансформации с использованием критерия Фишера

Раздел 5. Методы непараметрической статистики

Практические занятия

Методы непараметрической статистики: χ^2 критерий, метод Смирнова-Колмогорова, Вилкоксона-Манна-Уитни; критерий Краскела-Уоллиса.

Индивидуальные домашние задания

1. Определить зависимость между категориальными признаками с использованием χ^2 -критерия, согласно индивидуального задания
2. Провести оценку вывода о статистической недостоверности между средними с помощью U-критерия Вилкоксона
3. Провести проверку соответствия эмпирическое распределение предполагаемой модели с помощью критерия согласия Колмогорова
4. Провести односторонний дисперсионный анализ Краскела-Уоллиса, согласно индивидуальному задания
5. Проанализировать результаты опыта с помощью X-критерия Ван-дер-Вардена
6. Оценить различия между попарно связанными членами сравниваемых выборок с помощью критерия знаков z

Раздел 6. Принципы построения исследования

Практические занятия

Принципы построения исследования: рандомизация; выбор адекватного метода, критерия.

Вопросы для самостоятельной подготовки

1. Основные принципы проведения научных исследований
2. Какие главные разделы научного знания существуют и какие науки они в себя включают?
3. Каковы основополагающие характеристики исследования?
4. Каковы особенности научного познания?
5. Какова последовательность характеристик исследований?
6. Какие вы знаете подходы к изучению объекта в развитии?
7. Что называют методикой научных исследований?
8. По каким факторам различаются исследований?
9. Каковы правила конструирования определения?
10. Какие группы вопросов существуют и сколько их?

11. Что предполагает морфологический анализ?
12. Что такое доказательство и его виды?
13. Перечислите виды классификации.
14. Какие научные знания называют прикладными?
15. Какие научные исследования называют фундаментальными?
16. Перечислите принципы корректной и эффективной классификации.
17. Что означает обобщение?
18. Какие бывают подходы?
19. Охарактеризуйте формы исследований.
20. В чем особенности эксперимента?
21. Что такое методология?
22. Что включает в себя методологический аппарат?
23. Какое исследование может быть подлинно научным?
24. Какие принципы необходимо выполнять для подбора метода исследования?
25. Какие существуют логические законы в научной работе?
26. Что означают умозаключения, как их можно квалифицировать?
27. Какие вы знаете правила проверки в каждом случае умозаключения о причинной зависимости?
28. Что такое аргументирование?
29. Какими способами можно опровергнуть или снизить ценность исследования? Опишите их.
30. В чем сущность метода мозговой атаки?

2. ПОНЯТИЕ О ГЕНЕРАЛЬНОЙ И ВЫБОРОЧНОЙ СОВОКУПНОСТИ

Совокупностью называется любое множество, отличающихся друг от друга и в то же время сходных в некоторых существенных отношениях биологических объектов.

Наблюдения, проводимые над биологическими объектами, могут охватывать всех членов изучаемой совокупности, но могут и ограничиваться лишь некоторой ее частью. В первом случае достигается исчерпывающая и достоверная информация об объекте. Однако, к этому методу прибегают не всегда из-за невозможности и нецелесообразности, сопряженных со значительными затратами труда, средств и времени. Поэтому в подавляющем большинстве изучению подвергают какую-то часть совокупности — выборку.

Общая совокупность называется *генеральной* и представляет собой большой массив биологических объектов. Практически ее объем ограничен и может быть различным в зависимости от объекта наблюдений и от задачи, поставленной перед исследователем. Так, весь крупный рогатый скот, обитающий на земном шаре, представляет собой генеральную совокупность. Если требуется изучить какую-то породу, то в этом качестве будут выступать все животные, входящие в нее, независимо от места нахождения. Структурные единицы породы (линии, семейства, типы) так же могут рассматриваться как генеральная совокупность.

Под *выборочной совокупностью* подразумевается часть генеральной совокупности, предназначенной для ее характеристики. Объем выборки может быть сравнительно большим ($n \geq 30$) и малым ($n < 30$). Это необходимо учитывать, т.к. приемы статистической обработки данных будут разными.

Для того, чтобы по выборочной совокупности можно было судить о генеральной совокупности, она должна обладать следующими свойствами:

- однородностью (включение животных только одного пола, возраста, породы и т.д.);
- рандомностью (случайность, т. е. любой член генеральной совокупности имеет одинаковую вероятность попасть в нее);
- репрезентативностью (отражать с заданной точностью и надежностью по параметрам выборочной совокупности аналогичных показателей в генеральной совокупности) 9

Все единицы совокупности нужно характеризовать с той степенью точности (в тех единицах измерения), которая подходит для данного конкретного признака.

3. СТАТИСТИЧЕСКИЙ АНАЛИЗ В МНОГОЧИСЛЕННЫХ ВЫБОРКАХ

3.1. Вариационный ряд и вариационная кривая

При характеристике количественных признаков и большом числе вариантов, первичные данные необходимо упорядочить, т.е. сгруппировать в определенном порядке. Допустим, что для характеристики свиноматок крупной белой породы по плодовитости на ферме зарегистрировано 100 опоросов, имеющих в каждом следующее количество поросят в помете:

9, 11, 15, 12, 11, 14, 11, 10, 11, 13, 8, 10, 8, 10, 11, 8, 11, 7, 9, 10, 9, 7, 8, 7, 9, 11, 12, 12, 12, 11, 12, 14, 13, 11, 15, 10, 8, 7, 8, 5, 9, 5, 10, 9, 6, 10, 10, 6, 13, 6, 10, 13, 12, 9, 6, 10, 11, 9, 9, 10, 12, 7, 8, 7, 11, 10, 8, 7, 9, 10, 11, 9, 9, 13, 15, 10, 7, 8, 10, 9, 9, 12, 14, 13, 9, 10, 11, 10, 7, 9, 10, 11, 13, 14, 8, 9, 8, 11, 8, 14.

Значения вариант представляют ряд цифр. При внимательном рассмотрении нетрудно обнаружить, что каждая цифра встречается не один раз. В свою очередь это означает, что в данной выборке имеются животные с одинаковой плодовитостью. Возникает необходимость упорядочения вариант, т.е. расположить их в порядке возрастания. Для чего следует найти минимальное (min) и максимальное (max) значения признака. Наименьшее число поросят в помете — 5, а наибольшее — 15 голов.

Ранжируем выборку в пределах от 5 до 15 вариант и находим их частоты. Вначале определяем величину классового промежутка, который обозначается латинской буквой (k). В данном примере признак варьирует дискретно и величина классового промежутка равна 1 поросятку. Остается выяснить, сколько свиноматок дали 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 и 15 поросят, т.е. разнести варианты по классам. Разноску вариант по классам следует делать, не выбирая из всей выборки свиноматок только с указанной плодовитостью, это трудоемко и нецелесообразно, а начиная с первой варианты, подряд, отмечая каждую из них с помощью условных знаков - точек и черточек.

Удобным оказывается шифр частот:

• : :: ::: |: |: |: □ □ □ □ □.

Закончив разноску вариант по классам, шифр частот переводят в цифры.

Таблица 1 — Техника разноски вариант по классам

Класс		Шифр частоты	Частоты f
№	граница		
1	5	•	2
2	6	::	4
3	7	:	9
4	8	□	12
5	9	□	17
6	10	□	18

Класс		Шифр частоты	Частоты f
№	граница		
7	11	☒.	15
8	12	☒:	8
9	13	☒:.	7
10	14	☒::	5
11	15	☒!:	3

Таким образом, сгруппировав значения изучаемого признака в классы, мы получили два ряда цифр (классы и частоты), представляющих собою вариационный ряд, который показывает, как изменяется признак от минимальной до максимальной величины, а также частоту вариант в каждом классе.

Дадим характеристику этой же группе свиноматок по живой массе (кг):

197, 247, 207, 223, 215, 229, 232, 226, 228, 220, 226, 238, 236, 226, 225, 247, 195, 220, 252, 236, 220, 259, 240, 245, 219, 252, 256, 223, 245, 235, 256, 234, 228, 222, 262, 268, 219, 217, 224, 215, 259, 253, 247, 209, 222, 226, 230, 211, 220, 223, 257, 227, 274, 230, 266, 247, 258, 228, 254, 261, 259, 228, 247, 225, 226, 215, 211, 226, 226, 229, 247, 228, 238, 225, 232, 223, 220, 236, 247, 226, 222, 230, 229, 238, 236, 233, 230, 239, 228, 245, 215, 238, 220, 233, 219, 229, 217, 227, 223.

Составим вариационный ряд распределения свиноматок по этому признаку. Для этого вначале определим размах изменчивости — лимит (lim), т.е. разницу между максимальным и минимальным значением признака. В нашем примере max = 274 кг, min = 195 кг, lim = 274 – 195 = 79 кг. При этом величину классового промежутка (k) рассчитывают по формуле:

$$k = \frac{\text{max} - \text{min}}{\text{желательно количество классов}}$$

Количество классов в вариационном ряду зависит от степени изменчивости изучаемого признака и объема выборки (n). Чем больше выборка, тем больше классов нужно брать при составлении вариационного ряда. Обычно придерживаются следующих правил:

Таблица 2

Количество вариант в выборке	31-40	41-60	61- 100	101-200	>201
Количество классов	5-6	6-8	7-10	8-12	10-15

При этом величина классового промежутка должна быть как можно ближе к единице измерения изучаемого признака. В данном примере, если мы возьмем:

7 классов, то величина $k = 79/7 = 11,3$ кг, 8 классов соответственно $79/8=9,9$ кг, 9 классов $79/9=8,8$ кг, 10 классов $79/10=7,9$ кг.

Последние две величины классового промежутка близки к единице измерения, т.е. 1 кг. Следует помнить, что чем больше будет взято количество классов, тем точнее расчеты.

Последние две величины классового промежутка близки к единице измерения, т.е. 1 кг. Следует помнить, что чем больше будет взято количество классов, тем точнее расчеты.

Отсюда берем величину $k = 79/10 = 7,9$ кг, и ее можно округлить до 8 кг.

Далее нужно определить границы классов. Они должны быть удобными для разности вариант и обработки вариационного ряда. Нижней границей первого класса является минимальное значение признака (195 кг), а верхней — величина, большая на k ($195+8=203$ кг). Затем определяем только верхние границы каждого последующего класса, постоянно прибавляя величину (k), в последний класс войдет тах значение признака. Разница между верхней границей предыдущего класса и нижней границей следующего класса равна единице измерения (1 кг). Вариационный ряд будет выглядеть следующим образом:

Таблица 3

№	Граница класса		Шифр частот (разноска)	Частота f
	нижняя	верхняя		
1	195	203	•	2
2	204	211	••	4
3	212	219	⊗⊗	18

№	Граница класса		Шифр частот (разноска)	Частота f
	нижняя	верхняя		
4	220	227	⊗⊗⊗: : :	34
5	228	235	⊗: : :	16
6	236	243	⊗	10
7	244	251	: : :	8
8	252	259	: :	5
9	260	267	:	2
10	268	275	.	1

Анализируя любой вариационный ряд можно сделать представление не только о границах изменчивости изучаемого признака, но и о характере его изменчивости. Минимальное количество вариантов оказалось в крайних классах, а максимальное в классах, находящихся ближе к середине вариационного ряда.

Распределение вариантов в любом вариационном ряду можно изобразить графически. Если классы вариационного ряда не содержат несколько значений, а составлены по отдельным значениям вариантов, то кривая их распределения называется полигоном распределения. Для его построения на оси абсцисс наносятся значения классов, а на оси ординат частоты (число вариантов в каждом классе). Соединение точек дает ломаную кривую, которую следует доводить до нулевых классов, в которых нет вариантов. Если же классы вариационного класса включают несколько значений признака, их частоту нужно выражать не точкой, а столбиком, основание которого должно включать все значения признака класса. Такой график называется гистограммой.

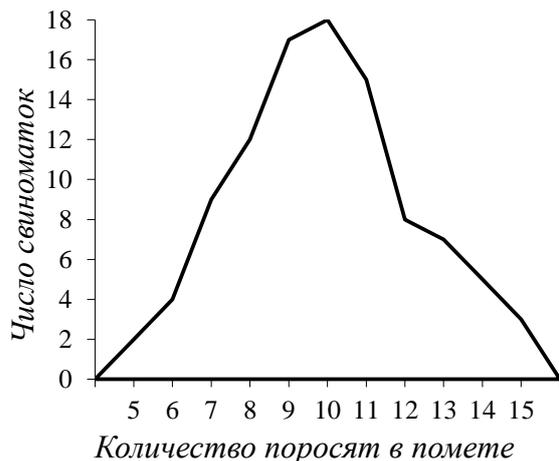


Рисунок 1 — Полигон распределения свиноматок по плодовитости

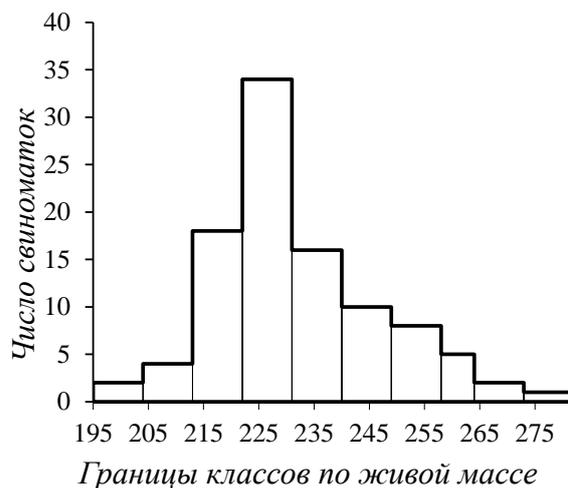


Рисунок 2 — Гистограмма распределения свиноматок по живой массе

Из гистограммы можно получить полигон, соединив середины верхних сторон столбиков линиями. В практике животноводства вариационную кривую можно использовать для выяснения результативности отбора.

3.2. Средняя арифметическая (\bar{X})

Это основной параметр, который показывает, какое значение признака наиболее характерно в целом для какой-либо совокупности. Она используется для сравнения пород, стад, линий, семейств и других разных групп, а так же характеризует средний фенотип и генотип популяции.

Среднюю арифметическую \bar{X} , \bar{X} определяют, как отношение суммы значений всех вариантов выборки к их числу объема выборки:

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

где x_i — значение варианты, n — объем выборки.

При отсутствии вычислительной техники расчеты удобнее вести с помощью вариационного ряда. Рассмотрим это на примере анализа живой массы свиноматок, который был составлен ранее.

Таблица 4

	Класс		Среднее значение W_i	Частота f	Отклонение	fa	fa^2
	граница						
	нижняя	верхняя					
1	195	203	199	2	-3	-6	18
2	204	211	207,5	4	-2	-8	16
3	212	219	215,5	19	-1	-9	9
4	220	227	A=223,5	28	0	0	0
5	228	235	231,5	21	+1	21	21
6	236	243	239,5	10	+2	20	40
7	244	251	247,5	10	+3	30	90
8	252	259	255,5	11	+4	44	176
9	260	267	263,5	3	+5	15	75
10	268	275	271,5	2	+6	12	72
				$\Sigma n=100$		$\Sigma fa=119$	$\Sigma fa^2=346$

Среднюю арифметическую в этом случае вычисляют по формуле:

$$\bar{X} = A \pm kb$$

где A — условная средняя, k — классовый промежуток, b — поправка к условной средней, которую определяем по формуле:

$$b = \pm \frac{\Sigma fa}{n}$$

где f — число вариантов в классе, a — отклонение от условной средней, n — число вариантов в выборке.

Если она получается со знаком минус, то ее отнимают, если плюс, то прибавляют к условной средней. За условную среднюю обычно принимают среднее значение класса с наибольшей частотой вариант. В нашем примере взято среднее значение 4-го класса (223,5). Для наглядности его нужно выделить двумя жирными линиями. Далее определяем условное отклонение для каждого класса. В четвертой строке вариационного ряда в шестой колонке ставят нуль

$$a = \frac{W_i - A}{k} = \frac{223,5 - 223,5}{9} = 0$$

Далее в сторону уменьшения средних значений классов ставим $-1, -2, -3$, а в сторону увеличения соответственно $+1, +2, +3 \dots +6$. Заполняем последующие колонки таблицы. Средняя арифметическая

величина именованная, т.е. выраженная в тех единицах измерения, что и признак, для которого она вычислена.

$$b = +\frac{119}{100} = +1,19$$

$$\bar{X} = 223,5 + 8 \cdot 1,19 = 234,2 = 233 \text{ кг,}$$

Наиболее целесообразным при проведении расчета средне арифметической, как и всех других статистических величин является использование электронных таблиц Microsoft Excel. Все исходные данные предварительно нужно разместить в столбцах таблицы и используя встроенные статистические, математические и логические функции выполнить вычисления используя ссылки на массивы данных.

Вычисления можно проводить с помощью меню вставка функции или вводя имена функций в конкретную ячейку после знака «=» (приложение Г).

3.3. Среднее квадратическое (стандартное) отклонение (σ)

Это основной показатель, характеризующий степень изменчивости изучаемого признака.

$$\sigma = \pm k \sqrt{\frac{\sum f a^2}{n} - b^2} = \pm 9 \sqrt{\frac{\sum 517}{100} - 1,19^2} = \pm 9 \cdot 1,94 = 17,4 \approx 17 \text{ кг}$$

Он применяется для построения кривых распределения признака в совокупности, определения интенсивности отбора, селекционного дифференциала и эффекта селекции. С его помощью можно выяснить границы вариационного ряда, если не известны лимиты. Установлено, что разнообразие признака располагается в следующих интервалах от средней арифметической:

$$\bar{X} \pm 1 \sigma — 68,3\% \text{ всех вариантов}$$

$$\bar{X} \pm 2 \sigma — 95,5\%$$

$$\bar{X} \pm 3 \sigma — 99,7\%$$

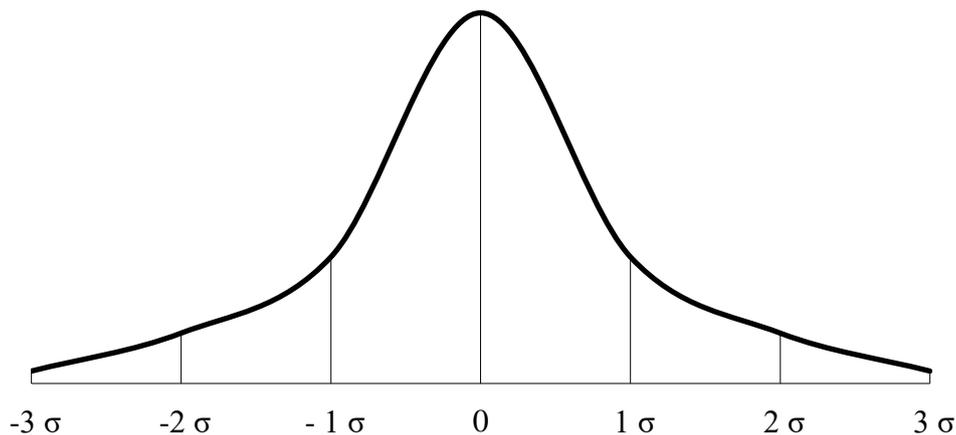


Рисунок 3 — Плотность нормального распределения

Следует учитывать, что в пределах $\bar{X} \pm 0,67 \sigma$ находится 50% особей всей совокупности. Это так называемые типичные особи, которые отличаются высокой приспособленностью к конкретным средовым факторам, что выражается в их более высоких воспроизводительных качествах, сохранности, долголетию и т.д. Если же отдельные варианты отклоняются от \bar{X} за $\pm 3\sigma$, то они не принадлежат данному вариационному ряду, а являются членами другой биологической совокупности и из анализа их необходимо исключить.

3.4. Коэффициент вариации (C_v)

Среднее квадратическое отклонение — величина именованная и выражается в определенных единицах измерения (л, кг, %, см, шт. и т.д.). Однако, чтобы имелась возможность сравнения степени изменчивости разных признаков, этот показатель нужно выразить в относительных единицах — в процентах к своей средней арифметической величине. Такой параметр называется коэффициентом вариации или коэффициентом изменчивости. Формула для его вычисления имеет следующий вид

$$C_v = \frac{\sigma}{\bar{X}} \cdot 100\% = \frac{17}{234} \cdot 100\% = 7,3\% \approx 7\%$$

Коэффициентом изменчивости необходимо пользоваться и тогда, когда сравнивается степень варьирования одинаковых признаков разных совокупностей при наличии значительной разницы между средними этих групп. Так, если в первом стаде средний удой 2000 кг, $\sigma_1 = \pm 500$ кг, во втором соответственно 1000 кг и 800 кг. Если судить по

величине среднего квадратического отклонения, то большая изменчивость должна быть во втором стаде, т.к. $\sigma_2 = 800$ кг против $\sigma_1 = 500$ кг. Однако, вычислив C_v мы увидим, что большая изменчивость будет в первом стаде

$$C_v = \frac{500}{2000} \cdot 100\% = 25\% \qquad C_v = \frac{800}{4000} \cdot 100\% = 20\%$$

Принято считать, что признаки со слабой изменчивостью характеризуются C_v не менее 10%, средней 11-20%, высокой — более 21%.

3.5. Изучение взаимосвязи между признаками

Признаки и свойства животных находятся в определенной взаимосвязи. Наблюдениями установлено, что увеличение удоя сопровождается уменьшением жира в молоке; чем больше масса яйца, тем больше масса цыпленка; существует связь между устойчивостью и восприимчивостью матерей и их дочерей по ряду различных заболеваний и т.д. Особенностью такой связи является то, что каждому значению признака соответствует несколько значений другого. Из приведенных примеров видно, что корреляция может быть различной. Она бывает положительной и отрицательной. При положительной увеличение (уменьшение) одного признака сопровождается увеличением (уменьшением) второго, а при отрицательной — увеличение одного ведет к уменьшению другого.

Различным может быть не только характер, но и степень связи. Для оценки силы и направления взаимосвязи между признаками вычисляют коэффициент корреляции (r). Он колеблется от 0 до ± 1 . При этом знак, стоящий перед цифровым значением r , свидетельствует о характере корреляции. Если r ниже 0,5, то связь считается слабой, при величине равной 0,51-0,7 — средней, если коэффициент более 0,71 — связь высокая. Следовательно, чем ближе значение r к единице, тем сильнее взаимосвязь; чем ближе к нулю — тем слабее связь между признаками.

3.5.1. Вычисление коэффициента корреляции

В качестве примера, вычислим величину корреляции между живой массой телок (кг) костромской породы в возрасте 12 мес. и величиной их удоя за 1 лактацию (кг).

Таблица 5

| Удой (кг) –
ж. масса (кг) |
|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| 4177 – 220 | 3903 – 255 | 3361 – 197 | 5074 – 270 | 3703 – 240 |
| 4188 – 227 | 4557 – 222 | 4345 – 262 | 4110 – 269 | 4548 – 244 |
| 4238 – 254 | 3865 – 200 | 4512 – 258 | 4433 – 276 | 4917 – 240 |
| 4179 – 260 | 4427 – 226 | 3971 – 259 | 4360 – 224 | 3746 – 230 |
| 4616 – 260 | 4778 – 250 | 4504 – 278 | 4651 – 253 | 4286 – 230 |
| 3819 – 280 | 5253 – 243 | 5548 – 263 | 4008 – 272 | 5038 – 240 |
| 4465 – 223 | 4427 – 226 | 3982 – 225 | 4215 – 264 | 5490 – 248 |
| 4374 – 225 | 4673 – 255 | 5451 – 278 | 4609 – 210 | 4628 – 290 |
| 4451 – 223 | 3865 – 200 | 4355 – 232 | 4231 – 235 | 3151 – 280 |
| 4394 – 224 | 5379 – 226 | 4765 – 255 | 4600 – 290 | 4493 – 265 |
| 4082 – 262 | 5180 – 260 | 4155 – 221 | 4101 – 233 | 3694 – 266 |
| 5145 – 260 | 3652 – 276 | 5253 – 243 | 3588 – 230 | 3572 – 240 |
| 4360 – 280 | 4682 – 261 | 3877 – 274 | 4623 – 269 | 3743 – 243 |
| 4511 – 245 | 3973 – 279 | 4131 – 250 | 5074 – 270 | 4896 – 264 |
| 5348 – 265 | 5213 – 224 | 3879 – 238 | 4816 – 250 | 4899 – 230 |
| 5124 – 253 | 4613 – 252 | 5419 – 271 | 3720 – 261 | 4731 – 231 |
| 4691 – 270 | 4955 – 225 | 5294 – 240 | 3946 – 238 | 5238 – 270 |
| 3199 – 265 | 5027 – 230 | 3601 – 212 | 4177 – 220 | 4460 – 270 |
| 3592 – 239 | 4815 – 299 | 4778 – 250 | 4900 – 268 | 4532 – 270 |
| 4815 – 248 | 4196 – 256 | 3168 – 242 | 4455 – 283 | 4262 – 257 |

Работа с выборкой идет в том же порядке, как это было при составлении вариационных рядов и вычисления его основных параметров. Но так как признака два, то удой обозначим буквой – x , а живую массу – y . Определяем границы классов для каждого вариационного ряда. Потом составляем корреляционную решетку (табл.). Записываем вверху по горизонтали границы классов по живой массе (y) от минимального до максимального значения, а по вертикали аналогично располагаем, границы по величине удоя (x), т.е. расположим их взаимоперпендикулярно. Затем разносим варианты по классам, при этом учитываем одновременно сразу значения двух признаков. Чтобы показатели первой первотелки, поместить в необходимую клеточку, нужно найти в какой класс она попадет по величине удоя и в какой по живой массе, и на их пересечении ставим точку и т.д. Подсчитываем частоту вариант f_x и f_y ,

выделяем жирными линиями классы с условной средней A_x (4352-4591) и A_y (220-230).

Таблица 6

$y \backslash x$	197	209	220	231	242	253	264	275	286	297	f_x	a_x	$f_x a_x$	$f_x a_x^2$
	208	219	230	241	252	263	274	285	296	307				
3151-3391	•	I			•		•	•	II		4	-5	-20	100
3392-3631		•	•	•							4	-4	-16	64
3632-3871	•		•	•	•	•	•	•			9	-3	-27	81
3872-4111			•	•		•	•	•			11	-2	-22	44
4112-4351			•	•	•	•	•				14	-1	-14	14
4352-4591			•	•	•	•	•	•			19	0	0	0
4592-4831		•		•	•	•	•		•	•	17	+1	17	17
4832-5071			•	•			•				6	+2	12	24
5072-5311		III	•	•	•	•	•	•	IV		10	+3	30	90
5312-5551			•		•	•	•	•			6	+4	24	96
f_y	3	2	23	12	13	19	16	9	2	1				$\sum f_x a_x^2 = 530$
a_y	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	100			
$f_y a_y$	-6	-2	0	12	26	57	64	45	12	7	$\sum f_y a_y = 215$			
$f_y a_y^2$	12	2	0	12	52	171	256	225	72	49	$\sum f_y a_y^2 = 851$			

$$b_x = b \frac{\sum f_x a_x}{n} = -\frac{16}{100} = -0,16$$

$$b_y = b \frac{\sum f_y a_y}{n} = +\frac{215}{100} = +2,15$$

В решетке образовались четыре квадранта. Далее определяем ряд условных отклонений a_x и a_y . Вычисляем произведение $f_x a_x$ и $f_y a_y$; $f_x a_x^2$ и $f_y a_y^2$ и их суммы.

Коэффициент корреляции рассчитывается по формуле:

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

где r — коэффициент корреляции, $\sum fa_x a_y$ — сумма произведений числа вариант на отклонение по ряду x и ряду y одновременно. Остальные символы в формуле известны.

Однако, среднее квадратическое отклонение берется не в именованных величинах, а в количестве классовых промежутков, т.е. цифра, полученная после извлечения квадратного корня на k (величина классового промежутка) не умножается и формула этого параметра имеет следующий вид

$$\sigma = \pm \sqrt{\frac{\sum fa^2}{n} - b^2},$$

Чтобы вычислить величину $\sum fa_x a_y$, нужно обратиться к корреляционной решетке, в которой как мы выяснили образовались четыре квадрата, их необходимо пронумеровать римскими цифрами. В каждом из них необходимо найти этот показатель. При этом частоту f в каждой клеточке умножаем на a_x и a_y одновременно. —

Таблица 7

I квадрант	II квадрант	III квадрант	IV квадрант
$1(-5)(-2) = 10$	$1(-5)2 = -10$	$\underline{1(-1)1 = -1}$	$2 \cdot 1 \cdot 1 = 2$
$1(-4)(-1) = 4$	$1(-5)4 = -20$	$\Sigma = -1$	$5 \cdot 1 \cdot 2 = 10$
$\underline{2(-3)(-2) = 12}$	$1(-5)5 = -25$		$2 \cdot 1 \cdot 3 = 15$
$\Sigma = +26$	$2(-4)1 = -8$		$1 \cdot 1 \cdot 4 = 4$
	$1(-3)1 = -3$		$2 \cdot 1 \cdot 6 = 12$
	$1(-3)2 = -6$		$1 \cdot 1 \cdot 7 = 7$
	$1(-3)3 = -9$		$1 \cdot 2 \cdot 1 = 2$
	$1(-3)4 = -12$		$2 \cdot 2 \cdot 4 = 16$
	$2(-3)5 = -30$		$1 \cdot 3 \cdot 1 = 3$
	$3(-2)1 = -6$		$2 \cdot 3 \cdot 2 = 12$
	$2(-2)3 = -12$		$3 \cdot 3 \cdot 3 = 9$
	$3(-2)4 = -24$		$3 \cdot 3 \cdot 4 = 36$
	$1(-2)5 = -10$		$1 \cdot 2 \cdot 3 = 24$
	$1(-1)1 = -1$		$1 \cdot 4 \cdot 4 = 16$
	$1(-1)2 = -2$		$\underline{1 \cdot 4 \cdot 5 = 20}$
	$5(-1)3 = -15$		$\Sigma = +213$
	$\underline{2(-1)4 = -8}$		
	$\Sigma = -201$		

Сумма произведений отклонений числа вариант по ряду x и ряду y в первом и четвертом квадратах положительная, во втором и третьем

— отрицательная. Таким образом, мы получили $\sum f_{x_1 y_1} = (+26) + (-201) + (-1) + (+213) = +37$. Все необходимое для вычисления r известно, находим его значение

$$r = \frac{+37 - 100 \cdot (-0,16) \cdot 2 \cdot 15}{100 \cdot 2,29 \cdot 1,97} = \frac{+71,4}{451,13} = +0,16$$

Следовательно, между удоем первотелок и их живой массой в возрасте 12 мес. существует слабая положительная связь. Так как коэффициент корреляции вычислен не по генеральной, а выборочной совокупности, он имеет свою статистическую ошибку

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}} = \pm \frac{1 - 0,16^2}{\sqrt{100}} = \pm 0,09$$

Далее определяем степень достоверности r

$$t_r = \frac{r}{m_r} = \frac{0,16}{0,09} = 1,78$$

При числе степеней свободы $\nu = n - 2 = 100 - 2 = 98$, находим t_{st} (приложение), они равны $1,98 \div 2,63 \div 3,39$. Так как найденное значение t_r меньше t_{st} , то не представляется возможным утверждать, что у исследованных коров костромской породы существует слабая положительная связь между их удоем за I лактацию и живой массой в возрасте одного года, хотя тенденция к этому имеется.

3.5.2. Вычисление коэффициента регрессии

При установлении связи между признаками часто возникает необходимость выяснить насколько увеличится один из признаков, если второй, связанный с ним, изменится на единицу измерения. К чему может привести односторонний отбор только по одному признаку? Это можно выяснить путем вычисления коэффициента регрессии R по следующим формулам:

$$R_{x/y} = r \frac{\sigma_x}{\sigma_y} = 0,16 \frac{2,26 \cdot 240}{1,97 \cdot 11} = 4,00 \text{ кг}$$

$$R_{y/x} = r \frac{\sigma_y}{\sigma_x} = 0,16 \frac{1,97 \cdot 11}{2,26 \cdot 240} = 0,01 \text{ кг}$$

При этом σ вычисляется с поправкой на величину классового промежутка. Следовательно, при увеличении живой массы первотелок в

возрасте 12 мес. на 1 кг, их удой за I лактацию увеличится на 4 кг, а при увеличении удоя на 1 кг, живая масса в анализируемом возрастном периоде увеличится на 0,01 кг (100 грамм).

Коэффициенты корреляции и регрессии вычисляются не только для установления связи между двумя варьирующими признаками одного и того же организма. Они используются для вычисления коэффициентов наследуемости h^2 (доля генетической изменчивости в общей фенотипической) и повторяемости r_w (степень соответствия оценок признака за разные промежутки времени). При этом r и R вычисляются между одноименными признаками в двух смежных поколениях (например, удой матерей — удой дочерей и т.д.). Техника их вычисления такая же как и при расчете коэффициента корреляции. При этом можно воспользоваться простейшими формулами:

$$h^2 = 2r_{M-D} \qquad h^2 = 2R_{M-D} \qquad r_w = r \qquad \text{и др.}$$

3.6. Оценка достоверности статистических показателей

3.6.3. Определение необходимого объема совокупности

В основу расчетов необходимого объема выборки положены такие параметры, как \bar{X} , C_v , σ , при этом можно пользоваться как научными литературными данными, так и личными наблюдениями. Наиболее простой способ — это использование коэффициента изменчивости, тогда достоверный объем выборки можно рассчитать по формуле:

$$n = \frac{C_v^2 \cdot t^2}{\varepsilon^2},$$

где C_v — коэффициент вариации; t — критерий достоверности (приложение); ε — допустимая погрешность (1, 3 или 5%)

3.6.4. Уровни вероятности и значимости

Всего выделяют три уровня вероятности безошибочного прогноза, обозначенные $P_1 \geq 0,95$; $P_2 \geq 0,99$ и $P \geq 0,999$, которым соответствуют три уровня значимости $\alpha \leq 0,05$; $\alpha \leq 0,01$ и $\alpha \leq 0,001$. Уровень значимости обозначает вероятность получения ошибки от установленных с определенной вероятностью результатов. С его помощью можно установить, в каком проценте случаев возможна ошибка в результатах.

Если уровень значимости $P \leq 0,05$ (5%-й), то это значит, что в силу случайности ошибка будет в 5% случаев. Вероятность $P \geq 0,95$ показывает, что из 100 повторений в 95 будут получены ожидаемые результаты.

Требования к уровню значимости:

$\alpha \leq 0,05$	Биологически вопросы, научно-хозяйственные опыты, поисковые эксперименты
$\alpha \leq 0,01$	Вопросы экономического и производственного характера, по которым будут даны рекомендации, детализация биологических явлений и закономерностей
$\alpha \leq 0,001$	Изучение действия опасных для жизни препаратов и заключение о дозах безвредности

3.6.5. Статистические ошибки (m)

В практической работе основные параметры совокупности вычисляют не по генеральной совокупности, а по части ее — выборке. Вследствие этого возникают ошибки, называемые статистическими, которые представляют собой величину, на которую параметр выборочной совокупности отличается от одноименного в генеральной совокупности. Для каждого параметра вычисляется своя статистическая ошибка. Для нашего примера они равны:

статистическая ошибка для средней арифметической

$$m_{\bar{x}} = \mp \frac{\sigma}{\sqrt{n}} = \frac{17}{\sqrt{100}} = 1,7 \approx 2 \text{ кг}$$

статистическая ошибка для среднего квадратического отклонения

$$m_{\sigma} = \mp \frac{\sigma}{\sqrt{2n}} = \frac{17}{\sqrt{200}} = 1,2 \approx 1 \text{ кг}$$

статистическая ошибка для коэффициента вариации

$$m_{c_v} = \mp \frac{c_v}{\sqrt{2n}} = \frac{7}{\sqrt{200}} = 0,49 \approx 0,5\%$$

Имея значения параметра, и, получив его статистическую ошибку, мы можем определить границы, в которых будет находиться данный параметр в генеральной совокупности. Статистики подсчитали, что он может находиться в пределах $\pm 3 m_x$ от выборочного. В данном примере это представлено следующим образом:

$$\tilde{x} = 234 - 3 \cdot 2 = 228,$$

$$234 + 3 \cdot 2 = 240$$

то есть генеральная средняя находится в пределах от 228 до 240 кг.

3.6.6. Ошибки разности выборочных средних

В практической работе и исследованиях нередко приходится встречаться с необходимостью сравнения двух групп (совокупностей) по ряду показателей. Для оценки достоверности разности между средними арифметическими двух выборок применяется критерий достоверности (t_d), который вычисляется по формуле:

$$t_d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m_1^2 + m_2^2}} \geq t_{st}$$

где \bar{X}_1 и \bar{X}_2 — средние арифметические, m_1 и m_2 — их статистические ошибки, t_{st} — стандартное значение критерия, определяемого по таблице Стьюдента (приложение) с учетом числа степеней свободы ($v = n_1 + n_2 - 2$) для трех уровней вероятности безошибочного прогноза, n_1 и n_2 — численность сравниваемых групп.

Например: Вычислить критерий достоверности разности по живой массе баранов породы прекос разных конституциональных типов ($n=200$ гол. в каждой группе), шерстно-мясной $77 \pm 2,2$ кг и мясной тип $85 \pm 2,6$ кг.

$$t_d = \frac{85 - 77}{\sqrt{2,2^2 + 2,6^2}} = 2,23$$

$$v = 200 + 200 - 2 = 398$$

$$P \geq 0,95$$

Следовательно, бараны мясного типа достоверно превосходят по живой массе своих сверстников шерстно-мясного типа.

3.7. Критерий хи-квадрат (χ^2)

Этот показатель показывает степень отличия наблюдаемых показателей от тех, которые должны были бы получаться в соответствии с существующими закономерностями. Он используется при сравнении достоверности разности между двумя и более группами, для изучения

влияния определенных факторов в изменчивости признаков. Критерий соответствия χ^2 вычисляется по формуле:

$$\chi^2 = \sum \frac{(p_{\text{э}} - p_{\text{т}})^2}{p_{\text{т}}}$$

где $p_{\text{э}}$ — эмпирическая частота (фактически наблюдаемое число); $p_{\text{т}}$ — теоретическая частота (теоретически ожидаемое число).

Допустим от 150 овцематок, при условии получения единцов, получено 150 ягнят, в т.ч. 80 баранчиков и 70 ярочек. Соответствует ли это теоретически ожидаемому 1:1 (75 баранчиков и 75 ярочек). Составим для удобства таблицу:

Таблица 8

Классы	Распределение	
	фактическое	теоретическое
Баранчики	80	75
Ярочки	70	75

Разница между фактическим и теоретически ожидаемым количеством баранчиков. $p_{\text{э}} - p_{\text{т}} = 80 - 75 = 5$, $(p_{\text{э}} - p_{\text{т}})^2 = 5^2 = 25$; ярочек соответственно $70 - 75 = -5$, $5^2 = 25$. Тогда:

$$\chi^2 = \frac{25}{150} + \frac{25}{150} = \frac{50}{150} = 0,33$$

Обратимся к таблице стандартных значений χ^2 (приложение), учтя при этом число степеней свободы $\nu = 2 - 1 = 1$. Полученное значение $\chi^2 = 0,33$ значительно ниже стандартного — 3,84, что свидетельствует о том, что наблюдаемое соотношение количества баранчиков и ярочек соответствует теоретическому ожиданию.

При четырёхпольных таблицах вычисления χ^2 пользуются непосредственно эмпирическими частотами по формуле:

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

где p — частоты, размещенные в четырехпольной таблице следующим образом:

p_1	p_2
p_3	p_4

Имеются две линии: высоколейкозная (АКР) и низколейкозная (СС57). В обычных условиях из 44 мышей АКР выжило 22 и пало 22, а в линии СС57

выжило 29, а пало 6. Необходимо оценить методом χ^2 различие между линиями. Составим таблицу:

Таблица 9

Линия	Выжило	Пало	Всего
АКР	22 (p_1)	22 (p_2)	44 (p_1+p_2)
СС57	29 (p_3)	6 (p_4)	35 (p_3+p_4)
Всего	51 (p_1+p_3)	28 (p_2+p_4)	$n = 79$

$$\chi^2 = \frac{(22 \cdot 6)^2 \cdot n}{(22 + 22) \cdot (29 + 6) \cdot (22 + 29) \cdot (22 + 6)}$$

Число степеней свободы $\nu = (2 - 1) (2 - 1) = 1$. Табличное значение χ^2 при $\nu = 1$ составляет $3,8 \div 6,6 \div 10,8$ (приложение), т.к. фактическая величина (9,2) больше табличного значения 6,6, но меньше, чем 10,8, нулевая гипотеза отвергается с высоким уровнем вероятности $p \geq 0,99$. Значит, между линиями существуют достоверные различия.

4. СТАТИСТИЧЕСКОЙ АНАЛИЗ В МАЛОЧИСЛЕННЫХ ВЫБОРКАХ

4.1. Среднее арифметическое, стандартное отклонение, коэффициент вариации, ошибка среднего арифметического

Малочисленными считают выборки, где число членов менее 30 ($n=3 \div 30$). В таких совокупностях значения вариант не группируется в классы, а вычисление статистических параметров удобнее вести с использованием метода условной средней. За ее величину может быть взято значение любой или наиболее часто встречающейся варианты.

Например: имеются показатели живой массы ягнят при рождении: 3,0; 4,5; 3,0; 3,5; 4,2; 4,6; 5,0; 3,0; 3,8; 3,2 кг. Значение вариант удобнее располагать не в ширину, а в столбик, во втором следует показать отклонение (Δ), в третьем квадрат отклонения значения каждой варианты от условной средней (Δ^2), которая выделяется рамкой.

Таблица 10

	Значение вариант (ν)	Отклонение $\Delta = \nu - A$	Δ^2
1	3,0	0	0
2	4,5	+1,5	2,25

	Значение вариант (v)	Отклонение $\Delta = v - A$	Δ^2
3	3,0	0	0
4	3,5	+0,5	1,25
5	4,2	+1,2	1,44
6	4,6	+1,6	2,56
7	5,0	+2,0	4,00
8	3,0	0	0
9	3,8	+0,8	0,64
10	3,2	+0,2	0,04
	$\Sigma v = 37,8$	$\Sigma \Delta = +7,8$	$\Sigma \Delta^2 = 11,18$

Средняя арифметическая $\bar{X} = A \pm \frac{\Sigma \Delta}{n} = 3,0 + \frac{7,8}{10} = 3,78 \approx 3,8$ кг

где A — условная средняя

Поправка к условной средней

$$\frac{\Sigma \Delta}{n} = +7,8$$

если со знаком «+», то прибавляется,
если со знаком «-», то вычитается

Среднее квадратическое отклонение

$$\sigma = \pm \sqrt{\frac{\Sigma \Delta^2}{n - 1}} = \pm \sqrt{\frac{11,18}{9}} = +1,1 \text{ кг}$$

Коэффициент вариации

$$C_V = \frac{\sigma}{\bar{X}} \cdot 100\% = \frac{1,1}{3,8} \cdot 100\% = 28,9\%$$

Статистическая ошибка средней арифметической

$$m_x = \pm \frac{\sigma}{\sqrt{n - 1}} = \pm \frac{1,1}{\sqrt{9}} = 0,37$$

4.2. Коэффициент корреляции

Коэффициент корреляции в малочисленной выборке вычисляется по формуле:

$$r = \frac{\alpha_1 + \alpha_2 - d}{2\sqrt{\alpha_1 \cdot \alpha_2}}$$

где α — сумма квадратов центральных отклонений: α_1 — для первого признака, α_2 — для второго

$$\alpha = \Sigma\Delta^2 - \frac{(\Sigma\Delta)^2}{n}$$

где Δ — отклонение вариант от условной средней: Δ_1 — для первого признака, Δ_2 — для второго)

$$d = \Delta_1 - \Delta_2$$

Пример: Определить направление и величину связи между содержанием жира (v_1) и белка (v_2) в молоке коров костромской породы ($n = 10$ гол)

Таблица 11

	Содержание		Δ_1	Δ_2	Δ_1^2	Δ_2^2	d	d^2
	жира (v_1)	белка (v_2)						
1	4,5	3,6	0	+0,1	0	0,01	+0,9	0,81
2	4,0	3,2	-0,5	-0,3	0,25	0,09	+0,8	0,64
3	4,6	3,5	+0,1	0	0,01	0	+1,1	1,21
4	5,0	3,6	+0,5	+0,1	0,25	0,01	+1,4	1,96
5	4,0	3,3	-0,5	-0,2	0,25	0,04	+0,7	0,49
6	4,1	3,3	-0,4	-0,2	0,16	0,04	+0,8	0,64
7	4,5	3,5	0	0	0	0	+1,0	1,00
8	4,8	3,6	+0,3	+0,1	0,09	0,01	+1,2	1,44
9	4,9	3,7	+0,4	+0,2	0,16	0,04	+1,2	1,44
10	5,2	3,8	+0,7	+0,3	0,49	0,04	+1,4	1,96
			$\Sigma\Delta_1=+0,6$	$\Sigma\Delta_2=+0,1$	$\Sigma\Delta_1^2=1,66$	$\Sigma\Delta_2^2=0,33$	$\Sigma d=10,5$	$\Sigma d^2=11,59$

Условные средние составляют $A_1 = 4,5\%$; $A_2 = 3,5\%$

Для проведения расчетов составим вспомогательную таблицу.

Таблица 12

Показатель	Признак		d
	1	2	
$\Sigma\Delta$	0,6	0,1	10,5
$(\Sigma\Delta)^2$	0,36	0,01	110,25
$(\Sigma\Delta)^2/n$	0,036	0,001	11,025
$\Sigma\Delta^2$	1,66	0,33	11,59
α	1,624	0,329	0,565

Коэффициент корреляции $r = \frac{1,624 + 0,329 - 0,565}{2\sqrt{1,624 \cdot 0,329}} = \frac{1,36}{1,45} = 0,937 \approx +0,94$

Статистическая ошибка коэффициента корреляции $m_r = \pm \sqrt{\frac{1 - r^2}{n - 2}} = \pm \sqrt{\frac{1 - 0,94^2}{8}} = 0,12$

Критерий достоверности коэффициента корреляции $t_r = \frac{r}{m_r} = \frac{0,94}{0,12} = 7,83$

По таблице Стьюдента находим, что $t_r = 7,83$ при $\nu = 8$ превышает стандартное значение при вероятности $p \geq 0,999$, что свидетельствует о статистической достоверности коэффициента корреляции. Значит между содержанием жира и белка в молоке у коров костромской породы существует высокая, устойчивая, положительная связь. Следовательно, с увеличением содержания жира в молоке увеличивается и содержание белка.

5. СТАТИСТИЧЕСКИЙ АНАЛИЗ АЛЬТЕРНАТИВНЫХ ПРИЗНАКОВ

Признаки, имеющие только два возможных взаимоисключающих состояния, называются альтернативными (пол мужской — женский, здоровые — больные животные и др.). Средняя арифметическая для

альтернативных признаков отражает долю или процент особей, имеющих (р) и не имеющих данный признак (q) в совокупности.

5.1. Определение ошибки выборочной доли

Пример: в стаде из 1000 коров заболело маститом 50 голов, а 950 остались здоровыми. Совокупность состоит из двух групп: первая — больные, вторая — здоровые животные. Обозначим численность первой группы как p_1 , второй — p_0 , а общее число — n .

Доля больных животных (т.е. имеющих данный признак) определяется по формуле:

$$p = \frac{p_1}{n} = \frac{50}{1000} = 0,05 \text{ или } 5\%$$

Доля здоровых животных (т.е. без данного признака) составляет:

$$q = \frac{p_0}{n} = \frac{950}{1000} = 0,95 \text{ или } 95\%$$

Среднее квадратическое отклонение вычисляют по формуле:

$$\sigma = \sqrt{p \cdot q} = \sqrt{0,95 \cdot 0,05} = 0,218 \text{ или } 21,8\%$$

Величина статистической ошибки доли величин одинакова как для доли больных, так и для доли здоровых животных:

$$m_p = m_q = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0,05 \cdot 0,95}{1000}} = 0,007 \text{ или } 0,7\%$$

5.2. Вычисление достоверности разности между выборочными долями или процентами:

Достоверность разницы между выборочными долями определяют по формуле:

$$t_{Dp} = \frac{|p_1 - p_2|}{\sqrt{m_{p_1}^2 + m_{p_2}^2}}$$

где p_1 и p_2 — доли особей, имеющих данный альтернативный признак; m_{p_1} и m_{p_2} ошибки выборочной доли в сравниваемых группах

Например: в семействе коровы Красы 998 учтено в трех поколениях 40 потомков, и у 20 из них выявлен лейкоз. А из 36 потомков семейства коровы Симпатии 200 — 9 голов. Необходимо установить различаются ли эти два семейства по восприимчивости к лейкозу.

Определим долю больных потомков: $P_x=20 : 40=0.5$ и $P_y=9 : 36=0.25$

Величины статистических ошибок соответственно будут равны:

$$m_{p_1} = \sqrt{\frac{0,5 \cdot (1 - 0,5)}{40}} = \sqrt{\frac{0,5 \cdot 0,5}{40}} = \sqrt{0,00625} = 0,08$$

$$m_{p_2} = \sqrt{\frac{0,25 \cdot (1 - 0,25)}{36}} = \sqrt{\frac{0,25 \cdot 0,75}{36}} = \sqrt{0,00521} = 0,07$$

Полученные данные подставляем в формулу:

$$t_{Dp} = \frac{0,5 - 0,25}{\sqrt{0,08^2 + 0,07^2}} = \frac{0,25}{\sqrt{0,0113}} = \frac{0,25}{0,106} = 2,36$$

По таблице Стьюдента с учетом числа степеней свободы $\nu = n_1 + n_2 - 2 = 40 + 36 - 2 = 74$ находим $t_{st} = 1,98 \div 2,63 \div 3,39$. Так как $t_d = 2,36$, это соответствует первому уровню вероятности безошибочного прогноза. Следовательно, разница заболеваемости лейкозом между потомками в анализируемых семействах достоверна. Это значит, что потомки в семействе Красы 200 отличаются большей восприимчивостью к лейкозу

5.3. Коэффициент корреляции

Для его определения первичные данные размещают в четырехпольной корреляционной решетке. Коэффициент корреляции (r_a) вычисляют по формуле:

$$r_a = \frac{(P_1 \cdot P_4 - P_2 \cdot P_3) - \frac{n}{2}}{\sqrt{(P_1 + P_2)(P_3 + P_4)(P_1 + P_3)(P_2 + P_4)}}$$

где P_1, P_2, P_3, P_4 — частоты, распределившиеся в четырех клетках корреляционной решетки.

Например: Определить величину связи между резистентностью цыплят к пуллорозу(белому поносу) и степенью их отселекционированности по данному признаку.

Таблица 13

Группа цыплят	После заражения живой культурой		Итого
	выжило	пало	
Исходная	115(P ₁)	105(P ₂)	P ₁ +P ₂ =220
Отселекционированная	560(P ₃)	58(P ₄)	P ₃ +P ₄ =618
Итого	P ₁ +P ₃ =675	P ₂ +P ₄ =163	n=838

Подставив табличные данные в формулу получим:

$$r_a = \frac{(115 \cdot 58 - 560 \cdot 105) - \frac{838}{2}}{\sqrt{220 \cdot 618 \cdot 675 \cdot 163}} = \frac{-52549}{122306} = -0,43$$

Статистическая ошибка коэффициента корреляции:

$$m_r = \frac{1 - r_a^2}{\sqrt{n}} = \frac{1 - 0,43^2}{\sqrt{838}} = 0,006$$

Достоверность коэффициента корреляции:

$$t_r = \frac{r}{m_r} = \frac{0,43}{0,006} = 71,67$$

Так как t_r значительно выше $t_{st}=3,29$, то коэффициент корреляции высоко достоверен ($p \geq 0,999$)

Следовательно, с увеличением интенсивности селекции снижается заболевание у цыплят пуллорозом.

6. ДИСПЕРСИОННЫЙ АНАЛИЗ

На формирование признаков у сельскохозяйственных животных влияют генетические факторы, факторы внешней среды, внутренние факторы организма, методы селекции. Они вызывают варьирование изучаемого признака, как в сторону увеличения, так и в сторону уменьшения. Степень и направленность воздействия различных факторов неодинакова, поэтому исключительно важно определить долю влияния каждого из них на изменчивость признака. Это можно выяснить с помощью дисперсионного анализа, в основу которого положен математический метод вычисления дисперсии или изменчивости. При этом в анализе используют общую дисперсию по всей выборке, факториальную и остаточную.

Дисперсионный анализ позволяет проводить изменение степени наследственности признака, сравнение наследственной и комбинативной способности лидерных животных, решает задачи по определению величины влияния генетических и паратипических факторов на признак, прогнозировать развитие признака при заданном или имеющемся комплексе условий или симптомов.

Дисперсионному анализу могут быть подвергнуты однофакторные и многофакторные статистические комплексы количественных и качественных признаков.

6.1. Техника расчетов при дисперсионном анализе однофакторного комплекса для малых групп

На *первом этапе* дисперсионного анализа определяют общие, факториальные, остаточные дисперсии (при многофакторном анализе и частные факториальные дисперсии, обусловленные взаимодействием факторов между собой)

Общая дисперсия (C_y) определяется суммой квадратов отклонений вариант результативного признака (V) от средней арифметической (M_0), вычисленной для выборочной совокупности:

$$C_y = \Sigma(V - M_0)^2$$

Факториальную дисперсию (C_x) определяют как сумму квадратов отклонений частной средней арифметической (M_i) от общей средней арифметической по каждой градации признака с учетом количества вариант результативного признака (n) по формуле:

$$C_x = \Sigma n_i (M_i - M_0)^2$$

Остаточная, или случайная дисперсия (C_z) вычисляется как сумма квадратов отклонений варианты каждой градации признака от частной средней арифметической по формуле:

$$C_z = (V - M_i)^2$$

Для комплексов всех типов действует равенство:

$$C_y = C_x + C_z$$

На *втором этапе* дисперсионного анализа определяют доли влияния факторов (учтенных и случайных) на вариабельность признака по формулам:

$$\eta_x^2 = C_x / C_y \quad \eta_z^2 = C_z / C_y$$

Эта величина может быть выражена в процентах и показывает, какая доля из общей изменчивости признака обусловлена влиянием анализируемого в опыте фактора, а какая случайного или неучтенного.

Для определения степени выраженности связи между частными средними (корреляционное отношение) рассчитывают показатель криволинейной связи по формуле:

$$\eta_x = \sqrt{\eta_x^2}$$

На *третьем этапе* определяют скорректированные (взвешенные) дисперсии или варианты (σ^2), как отношение дисперсий к соответствующему количеству степеней свободы. Полученные величины необходимы для определения критерия достоверности (F).

На *четвертом этапе* определяют достоверность полученной доли влияния воздействующего фактора на изменчивость признака с помощью критерия достоверности Р. Фишера как отношение факториальной дисперсии к остаточной, который сравнивают с табличными данными. Если $F_{\text{эмп}} \geq F_{\text{теор}}$, то дисперсия вызванная влиянием фактора достоверна, что позволяет говорить о достоверном влиянии фактора на результативный признак.

Вычислительные работы по проведению четырех этапов дисперсионного анализа оформляют в виде сводной таблицы в которую вводят по трем графам (x, z, y) значения дисперсии, доли влияния, число степеней свободы, скорректированные дисперсии, коэффициент достоверности Р. Фишера и табличные значения при трех уровнях значимости.

Пример: Определить долю влияния возраста овецематок (n=25) романовской породы на величину их многоплодия. Градации по возрасту следующие: 1, 2, 3, 4, 5 лет. В каждой градации 5 овцематок.

Для определения факториальной, остаточной и общей дисперсий при проведении дисперсионного анализа однофакторного комплекса целесообразно использовать вспомогательных величин, указанных в последнем столбце таблицы — взвешенный квадрат объема градаций (H), сумма квадратов всех вариантов ($\sum V^2$), сумма отношений квадратов объемов градаций и числа наблюдений по каждой градации ($\sum h_x$).

Таблица 14 — Однофакторный дисперсионный комплекс с одинаковым числом наблюдений по градациям фактора

Показатели	Градации фактора А (возраст, лет)					$l_A=5$
	1	2	3	4	5	
Даты (число ягнят) V_i	2	2	3	2	1	$H = (\sum V_i)^2/n = 92,16$
	1	3	3	2	2	
	1	2	2	2	2	
	2	1	3	1	1	
	1	3	2	2	2	
n_i	5	5	5	5	5	$\sum n_i = n = 25$
Объем градаций $\sum V_i$	7	11	13	9	8	$\sum V_i = \sum V = 48$
$(\sum V_i)^2$	49	121	169	81	64	$\sum (\sum V_i)^2 = 484$
$h_x = (\sum V_i)^2/n_i$	9,8	24,2	33,8	16,2	12,8	$\sum h_x = 96,8$
$\sum V_i^2$	11	27	35	17	14	$\sum \sum V_i^2 = \sum V^2 = 104$
Частое среднее $M_i = \sum V_i/n_i$	1,4	2,2	2,6	1,8	1,6	Общее среднее $M_o = \sum V/n = 1,92$

Таблица 15 — Сводная таблица дисперсионного анализа

Показатель		Варьирование					
		факториальное (межгрупповое под влиянием фактора)		остаточное (внутригрупповое под влиянием случайных факторов)		общее	
		x		z		y	
дисперсия	C	$C_x = \sum h_x - H$	4,64	$C_z = \sum V^2 - \sum h_x$	7,2	$C_y = \sum V^2 - H$	11,84
степень влияния	η^2	$\eta^2_x = C_x/C_y$	0,392	$\eta^2_z = C_z/C_y$	0,608	$\eta^2_y = C_y/C_y$	1
число степеней свободы	ν	$\nu_x = l_A - 1$	4	$\nu_z = n - l_A$	20	$\nu_y = n - 1$	24
варианса	σ^2	$\sigma_x^2 = C_x/\nu_x$	1,16	$\sigma_z^2 = C_z/\nu_z$	0,36	—	
критерий достоверности и критерий Фишера	F	$F_{эмп} = \sigma_x^2/\sigma_z^2$	3,22	—		—	
		$F_{теор} 5\%$	2,87	—		—	
		$F_{теор} 1\%$	4,43	—		—	
		$F_{теор} 0,1\%$	7,10	—		—	

Вычисленные значения η^2 дают ответ на первую задачу поставленную в дисперсионном анализе и свидетельствуют о том, что влияние случайных и неучтенных факторов на плодовитость овец значительно выше — 0,608 (60,8%), чем влияние возраста овцематок — 0,392 (39,2%) Однако возраст овцематок оказывает достоверное влияние на плодовитость, о чем свидетельствует превышение расчетного значения критерия достоверности табличных значений при $P < 0,05$ (5% уровень значимости)

Частные средние дисперсионного анализа (M_i) показывают, как изменяется плодовитость в зависимости от возраста и свидетельствуют о криволинейной связи между этими показателями.

Корреляционное отношение указывает на значительную величину связи между возрастом и плодовитостью:

$$\eta_x = \sqrt{\eta_x^2} = \sqrt{0,392} = 0,63$$

6.2. Использование Microsoft Excel для проведения дисперсионного анализа

Дисперсионный анализ можно провести и используя набор функций Microsoft Excel. Для этого в соответствующие ячейки таблицы помещают исходные данные и вводя формулы статистических и математических функций, используя ссылки на введенные и расчетные данные:

	A	B	C	D	E	F	G	H
1	Однофакторный дисперсионный комплекс с одинаковым числом наблюдений по градациям фактора							
2		Градации фактора А (возраст)					$l_A =$	=СЧЁТ(В3:F3)
3		1	2	3	4	5		
4		2	2	3	2	1	H	=H10^2/H9
5	V_i	1	3	3	2	2		
6		1	2	2	2	2		
7		2	1	3	1	1		
8		1	3	2	2	2		
9		n_i	=СЧЁТ(В4:В8)					n
10	$\sum V_i$	=СУММ(В4:В8)					$\sum V$	=СУММ(В10:F10)
11	$(\sum V_i)^2$	=В10^2					$\sum (\sum V_i)^2$	=СУММ(В11:F11)
12	h_x	=В11/В9					$\sum h_x$	=СУММ(В12:F12)
13	$\sum V_i^2$	=СУММ(В4^2;В5^2;В6^2;В7^2;В8^2)					$\sum V^2$	=СУММ(В13:F13)
14	M_i	=В10/В9					M_o	=H10/H9

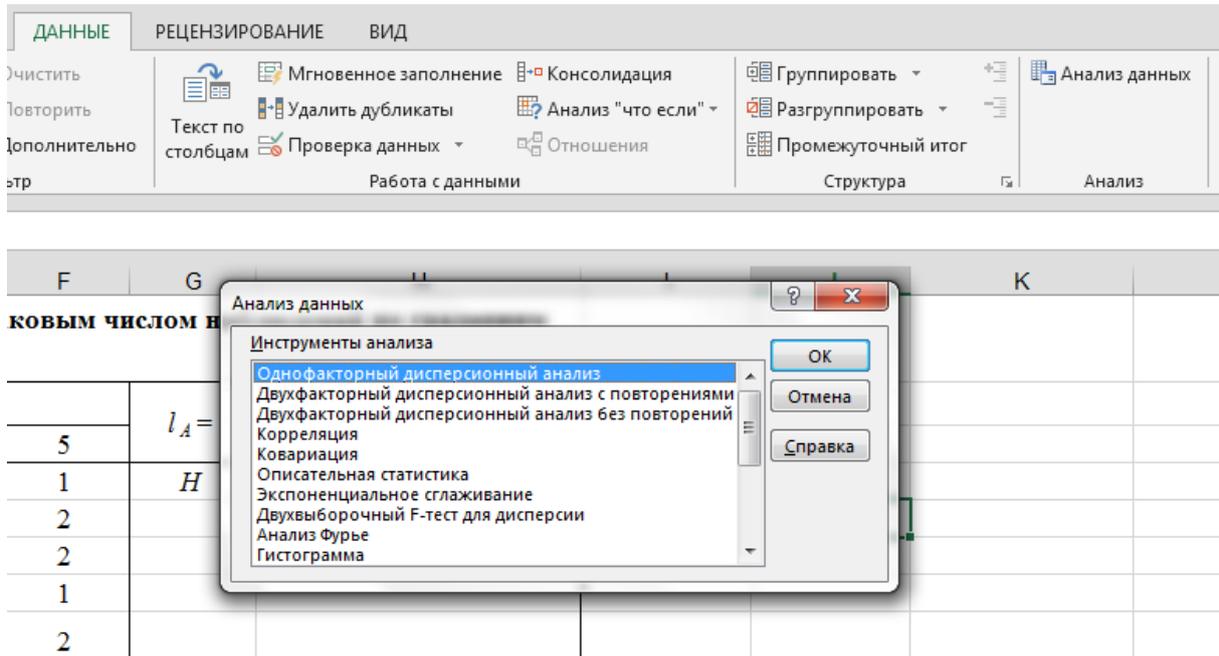
Формулы столбцах С-Е соответствуют указным для столбца В, и отличаются только буквой. Формулы необходимо ввести в ячейки В9:В14 и после этого скопировать в соседние.

При составлении сводной таблицы используют ссылки на ранее проведенные расчеты. Вычисление теоретических значений критерия достоверности Фишера в данном случае не является обязательным, так как, воспользовавшись формулой, представленной в ячейке В27 можно непосредственно вычислить вероятность F-распределения без использования таблиц.

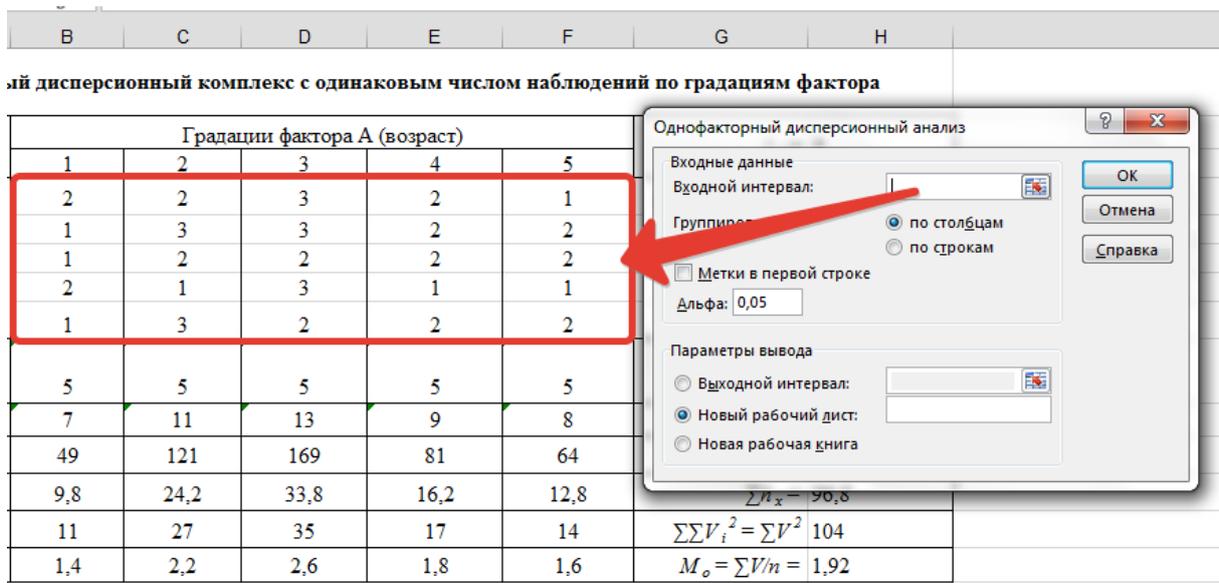
	А	В	С	Д
16	Сводная таблица дисперсионного анализа			
17	Показатель	Варьирование		
18		х	z	у
19	С	=Н12-Н4	=Н13-Н12	=Н13-Н4
20	η^2	=В19/Д19	=С19/Д19	=Д19/Д19
21	ν	=Н2-1	=Н9-Н2	=Н9-1
22	σ^2	=В19/В21	=С19/С21	–
23	$F_{\text{эмп}}$	=В22/С22	–	–
24	$F_{\text{теор}} 5\%$	=F.ОБР.ПХ(0,05;В21;С21)	–	–
25	$F_{\text{теор}} 1\%$	=F.ОБР.ПХ(0,01;В21;С21)		
26	$F_{\text{теор}} 0,1\%$	=F.ОБР.ПХ(0,001;В21;С21)		
27	Р	=F.РАСП.ПХ(В23;В21;С21)	–	–
28	η	=КОРЕНЬ(В22)	–	–

Проведение дисперсионного анализа, как и других сложных статистических вычислений можно существенно упростить если использовать входящий в состав Microsoft Excel набор средств анализа, который находится на вкладке «Данные».

Для проведения анализа, разобранный в нашем примере, необходимо выбрать однофакторный дисперсионный анализ.



Входным интервалом будет являться диапазон данных, в которых находятся сведения о плодовитости овцематок, сгруппированные по возрастам (показатели внесенные в ячейки В4:F8):



В качестве выходного интервала можно выбрать новый рабочий лист или указать пустую ячейку или диапазон на открытом. Дисперсионный анализ будет выполнен в автоматическом режиме а его результаты представляются в виде таблицы данных:

	A	B	C	D	E	F	G
20	Однофакторный дисперсионный анализ						
21							
22	ИТОГИ						
23	<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
24	Столбец 1	5	7	1,4	0,3		
25	Столбец 2	5	11	2,2	0,7		
26	Столбец 3	5	13	2,6	0,3		
27	Столбец 4	5	9	1,8	0,2		
28	Столбец 5	5	8	1,6	0,3		
29							
30							
31	Дисперсионный анализ						
32	<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
33	Между группами	4,64	4	1,16	3,22222222	0,0340156	2,866081402
34	Внутри групп	7,2	20	0,36			
35							
36	Итого	11,84	24				

В первой таблице отображаются значения объема градаций, частных средних и дисперсий в каждой градации фактора

Данные второй таблицы во многом аналогичны сводной таблице дисперсионного анализа. В строке «Между группами» отображается межгрупповое варьирование под влиянием фактора, изучаемого в опыте; в строке «Внутри групп» — внутригрупповое, под влиянием случайных и неучтенных факторов; в строке «Итого» — общее варьирование. Обозначение «SS» соответствуют дисперсии (факториальной и остаточной), «MS» — variance. Приводится эмпирическое (F) и табличное (F-критическое) значение критерия Фишера при выбранном ранее уровне значимости (в данном случае 5%), а также вероятность ошибочной оценки (P-Значение). При использовании пакета анализа Microsoft Excel также нет необходимости пользоваться таблицами распределения Фишера.

7. КРАТКИЙ СЛОВАРЬ ТЕРМИНОВ, ПРИМЕНЯЕМЫХ В МАТЕМАТИЧЕСКОЙ БИОСТАТИСТИКЕ

Альтернативные признаки — взаимоисключающие друг друга (больной–здоровый, высокий–низкий и т.п.).

Аномалия — отклонение от нормы.

Асимметрия — отсутствие или нарушение симметрии при распределении вариант по классам вариационного ряда.

Варианта — отдельное значение варьирующего признака.

Варьирование, вариабельность — особая форма биологической изменчивости, выражающаяся в виде внутригруппового индивидуального разнообразия.

Вариационный ряд — двойной ряд чисел, в котором произведена группировка их в классы в зависимости от величины изучаемого признака.

Вариация — разнообразие, видоизменение одного и того же признака, наблюдаемое в совокупности.

Вероятность — объективная возможность осуществления чего–либо, степень надежности или достоверности чего–либо.

Выборка — часть генеральной совокупности, предназначенная для ее характеристики.

Гистограмма — изображение вариационного ряда в виде столбиковой диаграммы, в которой высоты столбиков соответствуют частотам классов.

Группировка — объединение особей, вариант для совместного изучения, расположение вариант в таблицы или ряды.

Дисперсия — характеристика вариации признака, выражаемая средним квадратом отклонений вариант от средней арифметической данной совокупности.

Доверительные вероятности — степени надежности безошибочной оценки параметров генеральной совокупности по данным выборки.

Изменчивость — одно из общих свойств жизни, выражающееся в форме внутригруппового разнообразия.

Корреляция — связь между двумя варьирующими величинами, не имеющая функционального характера.

Кривая распределения — линейный график вариационного ряда.

Критерий — мерило надежности, достоверности суждений, выводов, их соответствие истине, реальной действительности.

Лимиты — границы вариации признака, максимальная и минимальная варианты совокупности или ряда.

Медиана — непараметрическая средняя, которая находится точно в середине вариационного ряда, делит его на две равные части.

Мода — наиболее часто встречающаяся варианта.

Параметр — величина, входящая в математическую формулу и сохраняющая свое постоянное значение в условиях данной совокупности.

Популяция — внутригрупповая группировка, или некоторая совокупность одного и того же вида.

Признак — любая особенность в строении, функциях, по которой можно отличить один объект наблюдения от другого.

Разнообразие — неоднородность особей по каким-либо признакам.

Ранжирование — размещение чего-либо в определенном порядке в направлении возрастания или убывания значений изучаемого признака.

Ранг — порядковый номер ранжированных значений признака.

Регрессия — математическое выражение линии связи между коррелируемыми признаками.

Репрезентативность — степень соответствия выборочных параметров, параметрам генеральной совокупности.

Рендамичность — случайность, неотобранность, при которой любой член генеральной совокупности имеет одинаковую вероятность войти в выборочную совокупность.

Совокупность (генеральная) — это теоретически бесконечно большая группа особей.

Совокупность (выборочная) — часть генеральной совокупности, предназначенная для ее характеристики.

Степени свободы — числа независимых величин, участвующих в образовании обобщенных статистических показателей.

Уровни значимости — показатели, используемые при проверке статистических гипотез, связанные с такими значениями вероятности, при которых появление ожидаемых событий в данных условиях считается практически невозможным. Чем меньше уровень значимости, тем меньше вероятность отвергнуть гипотезу.

Частота — абсолютная численность вариант, указывающая на то, как часто они встречаются в совокупности.

Явление — отдельный факт, единичное событие, их множество составляет массовое явление.

8. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Основные понятия теории вероятностей, случайные события.
2. Что такое вероятность? По какой формуле она вычисляется?
Условная вероятность и независимость событий.
3. Случайные величины и функция распределения вероятностей, дискретные случайные величины.
4. Непрерывные случайные величины и функция плотности вероятностей.
5. Характеристики распределений случайной величины
6. Примеры распределения случайных величин.
7. Что такое математическое ожидание?
8. Что такое закон распределения случайной величины?
9. Распределение Бернулли.
10. Биномиальное распределение.
11. Распределение Пуассона.
12. Непрерывное и дискретное равномерное распределение.
13. Нормальное распределение.
14. Распределение χ^2 .
15. Распределение Стюдента
16. Распределение Фишера.
17. Понятие случайной выборки. Примеры реальных биологических экспериментов.
18. Для чего может применяться математическая статистика в биологии?
19. Перечислите основные описательные статистики, используемые в предварительном анализе данных.
20. Что такое среднее арифметическое? Какими свойствами оно обладает?
21. Что такое мода? Ее использование в биологической статистике.
22. Что такое процентиля, квартили и медиана?
23. Для чего используется доверительный интервал?
24. Что такое дисперсия?
25. Что такое среднее квадратическое отклонение?
26. Что такое коэффициент вариации?
27. Что отражает стандартная ошибка среднего?
28. В чем разница между параметрическими и непараметрическими критериями?
29. Статистические данные: матрица экспериментальных данных, переменные и наблюдения, количественные и ранговые переменные.

30. Дескриптивные и графические методы анализа данных. Гистограмма: эмпирическая функция распределения.
31. Столбчатые диаграммы
32. Секторные диаграммы.
33. Понятие статистической оценки.
34. Свойства оценок: несмещенность, состоятельность, эффективность.
35. Метод максимального правдоподобия и точечное оценивание характеристик распределения (эмпирическая частота, выборочное среднее, выборочная дисперсия).
36. Интервальное оценивание.
37. Доверительный интервал. Доверительные интервалы для математического ожидания и дисперсия нормального распределения.
38. Логика проверки статистических гипотез. Ошибки первого и второго рода, уровень значимости и мощность критерия.
39. Одновыборочные и двухвыборочные критерии. Сравнение параметров биномиальных и пуассоновских распределений.
40. Критерий согласия χ^2
41. Критерий Колмогорова-Смирнова.
42. Непараметрические процедуры проверки гипотез.
43. Линейный регрессионный анализ, множественная линейная регрессия.
44. Метод наименьших квадратов.
45. Доверительные интервалы и проверка гипотез в линейном регрессионном анализе.
46. Корреляционный и дисперсионный анализ.
47. Однофакторный дисперсионный анализ.
48. Многофакторный дисперсионный анализ.
49. Методика планирования эксперимента.
50. Основные принципы проведения научных исследований
51. Какие главные разделы научного знания существуют и какие науки они в себя включают?
52. Каковы основополагающие характеристики исследования?
53. Каковы особенности научного познания?
54. Какова последовательность характеристик исследований?
55. Какие вы знаете подходы к изучению объекта в развитии?
56. Что называют методикой научных исследований?
57. По каким факторам различаются исследования?
58. Каковы правила конструирования определения?
59. Какие группы вопросов существуют и сколько их?

60. Что предполагает морфологический анализ?
61. Что такое доказательство и его виды?
62. Перечислите виды классификации.
63. Какие научные знания называют прикладными?
64. Какие научные исследования называют фундаментальными?
65. Перечислите принципы корректной и эффективной классификации.
66. Что означает обобщение?
67. Какие бывают подходы?
68. Охарактеризуйте формы исследований.
69. В чем особенности эксперимента?
70. Что такое методология?
71. Что включает в себя методологический аппарат?
72. Какое исследование может быть подлинно научным?
73. Какие принципы необходимо выполнять для подбора метода исследования?
74. Какие существуют логические законы в научной работе?
75. Что означают умозаключения, как их можно квалифицировать?
76. Какие вы знаете правила проверки в каждом случае умозаключения о причинной зависимости?
77. Что такое аргументирование?
78. Какими способами можно опровергнуть или снизить ценность исследования? Опишите их.
79. В чем сущность метода мозговой атаки?
80. Методика планирования эксперимента.

9. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Биометрия в MS Excel : учебное пособие / Е. Я. Лебедько, А. М. Хохлов, Д. И. Барановский, О. М. Гетманец. — 2-е изд., стер. — Санкт-Петербург : Лань, 2020. — 172 с.
2. Горлач, Б.А. Теория вероятностей и математическая статистика [Электронный ресурс] : учебно-методическое пособие. — Электрон. дан. — СПб. : Лань, 2013. — 320 с.
3. Дудников С.А. Количественная эпизоотология: основы прикладной эпидемиологии и биостатистики. — Владимир: Демиург, 2004. — 460 с.
4. Ивантер Э. В., Коросов А. В. Элементарная биометрия: Учеб. пособие / Э.В. Ивантер, А.В. Коросов; ПерГУ. — Петрозаводск, 2005. — 104 с.
5. Лакин Г.Ф. Биометрия: Учеб. пособие для биол. спец. вузов — 4-е изд., перераб. и доп. — М. :Высш. шк., 1990. — 352 с.
6. Ланг Т.А. Как описывать статистику в медицине. Аннотированное руководство для авторов редакторов и рецензентов / Т.А. Ланг, М. Сесик; пер. с англ. под. ред. В.П. Леонова. — М.: Практическая медицина, 2001. — 485 с.
7. Лях Ю.Е., Гурьянов В.Г., Хоменко В.Н., Панченко О.А. Основы компьютерной биостатистики: анализ информации в биологии, медицине и фармации статистическим пакетом MedStat. — Д. :Папакица Е.К., 2006. —214 с.
8. Меркурьева Е.К. Генетика с основами биометрии [Текст] : учебное пособие для вузов / Е. К. Меркурьева, Г. Н. Шангин-Березовский. - М : Колос, 1983. - 400 с.
9. Обмачевская, С. Н. Медицинская информатика. Курс лекций : учебное пособие для вузов / С. Н. Обмачевская. — 3-е изд., стер. — Санкт-Петербург : Лань, 2021. — 184 с.
10. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTIKA. — М. :МедиаСфера, 2002. — 312 с.
11. Степанов, В. Г. Применение методов непараметрической статистики в исследованиях сельскохозяйственной биологии и ветеринарной медицины : учебное пособие / В. Г. Степанов. — Санкт-Петербург : Лань, 2019. — 56 с.
12. Юнкеров В.И., Григорьев С.Г. Математико-статистическая обработка данных медицинских исследований. — СПб. :ВМедА, 2002. — 266 с.

Определение объема выборки для получения
достоверных результатов

Коэффициент изменчивости C_v , %	Значение уровня вероятности $P \geq$		
	0,95	0,99	0,999
	Объем выборки, гол		
Допустимая погрешность опыта — 5%			
8	10	17	28
10	16	27	44
12	23	39	63
14	31	53	85
16	41	69	112
18	52	88	141
20	64	108	174
Допустимая погрешность опыта — 3%			
8	28	48	77
10	44	75	121
12	64	108	174
14	87	147	237
16	114	192	310
18	144	243	392
20	178	300	480
Допустимая погрешность опыта — 1%			
8	2566	433	697
10	400	676	1089
12	576	973	1568
14	784	1325	2134
16	1024	1730	2708
18	1296	2190	3528
20	1600	2704	4356

Приложение Б

Значения критерия достоверности по Стьюденту-Фишеру

Число степеней свободы, ν	Доверительная вероятность (P) Уровень значимости (α)		
	$P = 0,095$ $\alpha = 0,05$	$P = 0,099$ $\alpha = 0,01$	$P = 0,0999$ $\alpha = 0,001$
	2	4,303	9,925
3	3,182	5,841	12,941
4	2,776	4,604	8,610
5	2,571	4,032	6,859
6	2,447	3,707	5,959
7	2,365	3,499	5,405
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
11	2,201	3,106	4,437
12	2,179	3,055	4,318
13	2,160	3,012	4,221
14	2,145	2,977	4,140
15	2,131	2,947	4,073
16	2,120	2,921	4,015
17	2,110	2,898	3,965
18	2,101	2,878	3,922
19	2,093	2,861	3,883
20	2,086	2,845	3,850
22	2,074	2,819	3,792
25	2,060	2,787	3,725
30	2,042	2,750	3,646
35	2,030	2,724	3,591
40	2,021	2,704	3,551
45	2,014	2,690	3,520
50	2,008	2,678	3,496
55	2,004	2,669	3,476
60	2,000	2,660	3,460
70	1,994	2,648	3,435
80	1,989	2,638	3,416
90	1,986	2,631	3,402
100	1,982	2,625	3,390
120	1,980	2,617	3,373
>120	1,960	2,5758	3,2905

Стандартные значения критерия хи-квадрат (χ^2)

Число степеней свободы, ν	Доверительная вероятность (P) Уровень значимости (α)		
	$P = 0,095$ $\alpha = 0,05$	$P = 0,099$ $\alpha = 0,01$	$P = 0,0999$ $\alpha = 0,001$
1	3,84	6,64	10,83
2	5,99	9,21	13,82
3	7,81	11,34	16,27
4	9,49	13,28	18,46
5	11,07	15,09	20,52
6	12,59	16,81	22,46
7	14,07	18,48	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88
10	18,31	23,21	29,59
11	19,68	24,72	31,26
12	21,03	26,22	32,91
11	22,36	27,69	34,53
14	23,68	29,14	36,12
15	25,00	30,58	37,70
16	26,30	32,00	39,25
17	27,59	33,41	40,79
18	28,87	34,81	42,31
19	30,14	36,19	43,82

Приложение Г

*Основные статистические, математические и логические функции
электронных таблиц Microsoft Excel*

Синтаксис	Выполняемая операция
ABS	Определяет абсолютное значение числа
ДИСП	Оценивает дисперсию по выборке,
ДОВЕРИТ	Вычисляет доверительный интервал для среднего значения по генеральной совокупности,
ЕСЛИ	Проверяет выполняется ли условие, Определяет одно значение если оно выполняется и другое если нет
КВАДРОТКЛ	Вычисляет сумму квадратов отклонений,
КОРЕНЬ	Вычисляет квадратный корень числа
КОРРЕЛ	Вычисляет коэффициент корреляции между двумя множествами данных,
МАКС	Вычисляет максимальное значение из списка аргументов
МИН	Вычисляет минимальное значение из списка аргументов
МЕДИАНА	Вычисляет медиану заданных чисел,
МОДА	Вычисляет значение моды множества данных,
ОКРУГЛ	Производит округление до указанного разряда
СРГЕОМ	Вычисляет среднее геометрическое,
СРГАРМ	Вычисляет среднее гармоническое,
СРЗНАЧ	Вычисляет среднее арифметическое аргументов,
СРОТКЛ	Вычисляет среднее абсолютных значений отклонений точек данных от среднего,
СТАНДОТ-КЛОН	Оценивает стандартное отклонение по выборке,
СТЕПЕНЬ	Вычисляет степень числа
СТЮДРАСП	Вычисляет t -распределение Стьюдента,
ТТЕСТ	Вычисляет вероятность, соответствующую критерию Стьюдента,

ДЛЯ ЗАМЕТОК

Учебно-методическое издание

Математические методы в биологии : учебно-методическое пособие по выполнению контрольных работ / сост. С.Г. Белокуров, М.С. Трескин. — 2-е изд., исправл. и доп. — Караваево : Костромская ГСХА, 2021. — 60 с. ; 20 см. — 50 экз. — Текст непосредственный.

*Учебно-методическое пособие по выполнению контрольных работ
издаётся в авторской редакции*

© Федеральное государственное бюджетное образовательное учреждение высшего образования "Костромская государственная сельскохозяйственная академия" 156530, Костромская обл., Костромской район, пос. Караваево, уч. городок, д. 34

Компьютерный набор. Подписано в печать 30/07/2021. Заказ № 800.
Формат 60x84/16. Тираж 50 экз. Усл. печ. л. 3,6. Бумага офсетная.
Отпечатано 30/07/2021. Цена 89,00 руб.

вид издания: 2-е изд., испр. и доп. (электронная версия)
(редакция от 15.06.2021 № 800)

Отпечатано с готовых оригинал-макетов в академической типографии на цифровом дубликаторе. Качество соответствует предоставленным оригиналам.

(Электронная версия издания - I:\подразделения \рио\издания\2021\800.pdf)



2021*800

Цена 89,00 руб.

ФГБОУ ВО КОСТРОМСКАЯ ГСХА



2021*800

(Электронная версия издания - I:\подразделения \рио\издания\2021\800.pdf)