

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ -
МСХА имени К.А. ТИМИРЯЗЕВА**

**О.Б. ТАРАСОВА, Б.Ш. ДАШИЕВА,
К.А. КОЗЛОВ, И.А. АНОХИН, А.Ю. ВЕТОШКИН**

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

МОСКВА 2025

УДК 519.2
ББК 22.17
М34

*Рекомендовано к изданию учебно-методической
комиссией института экономики и управления АПК
ФГБОУ ВО «РГАУ-МСХА имени К.А. Тимирязева»
(протокол № 8 от 17 марта 2025 г.).*

Рецензент:

Архипова М.Ю. – профессор департамента статистики и анализа данных
федерального государственного автономного образовательного учреждения
высшего образования «Национальный исследовательский университет
«Высшая школа экономики»,
доктор экономических наук, профессор

**Тарасова О.Б., Дашиева Б.Ш., Козлов К.А.,
Анохин И.А., Ветошкин А.Ю.**

М34 **Математическая статистика:** учебное пособие /
О.Б. Тарасова, Б.Ш. Дашиева, К.А. Козлов, И.А. Анохин,
А.Ю. Ветошкин. – М.: Издательство «Научный консультант», 2025.
– 130 с.

ISBN 978-5-907933-41-5

В учебном пособии изложены методы математической статистики, приведены примеры их применения для обработки и анализа данных в агрохимии и почвоведении.

Учебное пособие предназначено для аудиторной и самостоятельной работы студентов аграрных вузов, обучающихся по направлению подготовки 35.03.03 «Агрохимия и агропочвоведение». Оно может быть также полезно студентам других направлений подготовки, аспирантам, преподавателям и специалистам, применяющих методы математической статистики.

УДК 519.2
ББК 22.17

ISBN 978-5-907933-41-5

© Тарасова О.Б., 2025
© Дашиева Б.Ш., 2025
© Козлов К.А., 2025
© Анохин И.А., 2025
© Ветошкин А.Ю., 2025
© ФГБОУ ВО РГАУ-МСХА
имени К.А. Тимирязева, 2025
© Оформление: Издательство
«Научный консультант», 2025

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	6
РАЗДЕЛ I СТАТИСТИЧЕСКИЕ РЯДЫ РАСПРЕДЕЛЕНИЯ И ИХ ХАРАКТЕРИСТИКИ	7
Методические указания	7
Типовые задачи.....	9
Задача 1.1 Построение рядов распределения по одному признаку.....	9
Задача 1.2 Расчет средней арифметической и медианы в ранжированном ряду распределения.....	15
Задача 1.3 Расчет средней арифметической, моды и медианы в дискретном ряду распределения	16
Задача 1.4 Расчет средней арифметической, моды и медианы в интервальном ряду распределения	19
Задача 1.5 Анализ взаимосвязи средней арифметической, моды и медианы в умеренно асимметричных рядах	23
Задача 1.6 Расчет квартилей в интервальном ряду распределения.....	24
Задача 1.7 Расчет и анализ показателей вариации в ранжированном ряду распределения.....	25
Задача 1.8 Расчет и анализ показателей вариации в дискретном ряду распределения.....	28
Задача 1.9 Исчисление показателей вариации в интервальном ряду распределения	30
Задача 1.10 Правило сложения дисперсий (разложения объемов вариаций)	33
Задача 1.11 Расчет начальных и центральных моментов интервального ряда распределения	38
Задача 1.12 Расчет показателей формы распределения	41
Задача 1.13 Расчет средней арифметической и вариации альтернативного признака	42
Вопросы для проверки знаний по разделу I	43
РАЗДЕЛ II ВЫБОРОЧНЫЙ МЕТОД. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ.....	45
Методические указания	45
Типовые задачи.....	48
Задача 2.1 Точечная и интервальная оценка параметров генеральной совокупности по данным выборки.....	48
2.1.1 Точечная и интервальная оценка среднего значения признака в генеральной совокупности по данным большой выборки	48

2.1.2 Точечная и интервальная оценка средней величины генеральной совокупности по данным малой выборки	51
Задача 2.2 Критерий χ^2 -Пирсона как критерий согласия	52
2.2.1 Проверка статистической гипотезы о соответствии фактического выборочного распределения теоретическому соотношению	52
2.2.2 Проверка статистической гипотезы о соответствии фактического распределения нормальному распределению	54
2.2.3 Проверка статистической гипотезы о соответствии фактического выборочного распределения распределению Пуассона	57
Задача 2.3 Проверка статистической гипотезы по критерию χ^2 -Пирсона о независимости двух эмпирических распределений	59
Задача 2.4 Проверка статистической гипотезы по критерию χ^2 -Пирсона об однородности двух эмпирических распределений	62
Задача 2.5 Проверка статистических гипотез относительно двух средних величин по данным малых выборок	65
2.5.1 Выборки независимые с равными численностями ($n_1=n_2$), дисперсии генеральных совокупностей равны ($\sigma_1^2=\sigma_2^2$)	65
2.5.2 Выборки независимые с разной численностью ($n_1\neq n_2$), дисперсии генеральных совокупностей равны ($\sigma_1^2=\sigma_2^2$)	67
2.5.3 Выборки независимые, равной численности ($n_1=n_2$), дисперсии в генеральных совокупностях не равны ($\sigma_1^2\neq\sigma_2^2$)	69
2.5.4 Выборки независимые, разной численности ($n_1\neq n_2$), дисперсии в генеральных совокупностях не равны ($\sigma_1^2\neq\sigma_2^2$)	71
Задача 2.6 Проверка статистических гипотез относительно средней разности при парной зависимости между единицами совокупности	73
Вопросы для проверки знаний по разделу II	74
РАЗДЕЛ III СТАТИСТИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ СВЯЗЕЙ	76
Методические указания	76
Типовые задачи	79
Задача 3.1 Однофакторный дисперсионный анализ	79
3.1.1 Однофакторный дисперсионный анализ при случайном формировании выборок равной численности	79
3.1.2 Однофакторный дисперсионный анализ при неслучайном формировании выборок равной численности	83
Задача 3.2 Проверка статистических гипотез относительно средних по данным двух выборок на основе результатов однофакторного дисперсионного анализа по критерию Q-Тьюки	87

Задача 3.3 Однофакторный дисперсионный анализ при случайном формировании выборок неравной численности с попарным сравнением средних по критерию Шеффе	89
Задача 3.4 Двухфакторный дисперсионный анализ при неслучайном формировании равночисленных выборок.....	94
Задача 3.5 Определение показателей связи при парной линейной корреляции	98
Задача 3.6 Проверка гипотез и статистическая оценка показателей связи при парной линейной корреляции	102
Задача 3.7 Определение показателей связи при множественной линейной корреляции.....	107
Вопросы для проверки знаний по разделу III	113
ЛИТЕРАТУРА	115
ПРИЛОЖЕНИЯ	116

ПРЕДИСЛОВИЕ

Учебное пособие по математической статистике разработано в соответствии с требованиями ФГОС ВО по направлению подготовки 35.03.03 «Агрохимия и агропочвоведение».

В учебном пособии представлены методические указания и типовые задачи, учитывающие опыт преподавания дисциплины «Математическая статистика» на кафедре статистики и кибернетики РГАУ-МСХА имени К.А. Тимирязева. Задания адаптированы к учебнику «Математическая статистика» (под ред. А.П. Зинченко и А.В. Уколовой, 2018 г.). Это позволяет студентам изучать разделы дисциплины в сочетании теории с практикой, правильно уяснять поставленную цель заданий и ее реализацию при самостоятельном освоении и выполнении конкретной задачи. Ряд задач выполняется на одном исходном материале и носит сквозной характер, что способствует системному, последовательному освоению методов математической статистики и всестороннему подходу применительно к одному объекту изучения.

Предлагаемые типовые задачи по всем разделам курса «Математическая статистика» рассчитаны на выработку у студентов умений и навыков построения и графического изображения статистических рядов распределения, расчета описательных статистик, статистического оценивания параметров распределения, проведения проверки статистических гипотез, а также применения методов дисперсионного и корреляционно-регрессионного анализа. Задачи содержат краткие методические указания по выполнению задания, подробные алгоритмы расчета показателей, интерпретации полученных результатов и выводы. По каждому разделу имеется перечень контрольных вопросов. Типовые задачи рассчитаны как на аудиторную, так и на самостоятельную внеаудиторную работу студентов аграрных вузов, обучающихся по направлению «Агрохимия и агропочвоведение».

РАЗДЕЛ I

СТАТИСТИЧЕСКИЕ РЯДЫ РАСПРЕДЕЛЕНИЯ И ИХ ХАРАКТЕРИСТИКИ

Методические указания

Объектом изучения математической статистики являются *статистические совокупности* – множества единиц каких-либо массовых явлений в обществе и природе, объединенных общим содержанием. Например, группа студентов вуза представляет собой статистическую совокупность, так как ее единицы (студенты) имеют общее качество – заняты в учебном процессе с целью получения высшего образования.

Каждая единица совокупности обладает различными свойствами, которые могут быть выражены количественной характеристикой (например, оценки студента на экзаменах, число пропусков занятий и т.д.) или качественной характеристикой (студент вне учебы работает или не работает, проживает в общежитии или дома и т.д.)

Все единицы совокупности можно упорядочить по одному или нескольким признакам, то есть построить *ряды распределения*.

Различают следующие виды рядов распределений: а) по качественным признакам и б) по количественным признакам. По качественным признакам упорядочение единиц может быть проведено по градациям качества (например, распределение земельных участков по типам почв или распределение студентов по месту проживания). Упорядочение единиц совокупности по количественным признакам (условное обозначение признаков – $x, y, z, \dots$) проводится по возрастанию или убыванию значений количественного признака (например, распределение студентов по оценке на экзамене) и может быть реализовано в виде 1) ранжированного ряда, 2) дискретного ряда, 3) интервального ряда и 4) кумулятивного ряда распределения.

Ранжированным рядом называется распределение единиц совокупности в порядке возрастания ($x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$) или убывания значений количественного признака ($x_1 \geq x_2 \geq x_3 \geq \dots \geq x_n$). Ранжированный ряд имеет два элемента: номера единиц совокупности по ранжиру и значения признака в определенном порядке. В целях наглядной иллюстрации статистические ряды распределения могут быть изображены в виде графиков. График ранжированного ряда называется огивой Гальтона. Он показывает интенсивность изменения признака в совокупности (равномерное или скачкообразное изменение признака, слабая или высокая интенсивность изменения, наличие или отсутствие резко отличающихся значений).

Дискретным рядом называется ряд распределения единиц совокупности в порядке возрастания или убывания градаций дискретного признака с указанием частот встречаемости этих градаций в совокупности. Элементы дискретного ряда: градации значений признака в порядке возрастания или убывания – x_i ; число единиц совокупности с данными

градациями признака (так называемые частота встречаемости или статистический вес – f_i). График дискретного ряда распределения – *полигон*.

Интервальным рядом распределения называется ряд распределения единиц по признаку, заключенному в интервалы, с указанием частот встречаемости в каждом интервале. Элементы интервального ряда: градации значений признака в порядке возрастания или убывания, заключенные в интервалы – x_i ; число единиц совокупности в каждом интервале – f_i . График интервального ряда распределения – *гистограмма*.

По данным графиков полигона и гистограммы можно проанализировать равномерность распределения единиц совокупности по изучаемому признаку, наличие наиболее и наименее часто встречающихся значений признака, форму распределений (симметричная или асимметричная, близка к нормальному распределению или сильно отличается от него и т.п.).

Кумулятивный ряд распределения – это ряд накопленных частот значений признака в совокупности. Накопленная частота (S_f) – это частота конкретного значения признака, сложенная со всеми частотами предыдущих значений признака. По значениям признака и накопленным частотам строят и изображают график в виде *кумуляты*.

Обобщение значений признака по единицам совокупности позволяет получить *статистический показатель*. Показатель является характеристикой статистической совокупности в целом, в то время как значение признака – характеристикой отдельной единицы совокупности.

К статистическим показателям относятся 1) *средние величины*, характеризующие типические черты совокупности (их называют также показателями центральной тенденции); 2) *показатели вариации*, характеризующие изменчивость признака в совокупности, и 3) *показатели формы распределения*, характеризующие асимметричность (скошенность) и островершинность (плосковершинность) изучаемого распределения относительно нормального распределения. Многие из них являются частным случаем *статистических моментов*.

Средние величины могут быть получены на основе всех индивидуальных значений признака – параметрические (степенные) средние – $\bar{x}$: средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая, средняя кубическая), или на основе отдельных значений признака – непараметрические (структурные) средние: модальное значение признака (мода – x_{Mo}) – значение признака с наибольшей частотой встречаемости; медианное значение признака (медиана – x_{Me}) – значение признака, делящее ряд распределения единиц совокупности пополам; квартили Q_1, Q_2, Q_3 – значения признака, которые делят ряд распределения на четыре части, равные по численности единиц); квинтили ($K_1...K_4$) – значения признака, которые делят ряд распределения на пять частей, равных по численности единиц совокупности).

Показатели вариации могут характеризовать максимальную колеблемость признака (размах вариации) и среднюю колеблемость признака в совокупности (среднее линейное отклонение, среднее квадратичное

отклонение, дисперсия признака). Именованные показатели (кроме дисперсии), деленные на среднюю величину $\bar{x}$, позволяют получить относительные показатели вариации (коэффициент осцилляции, относительное линейное отклонение, коэффициент вариации).

Для обобщенной характеристики статистических рядов используют статистические моменты распределения. Статистический момент – это показатель статистической совокупности, характеризующий среднюю из k -степеней отклонений вариант от некоторой постоянной величины a . Исчисляют *начальные и центральные* моменты. В начальных моментах принимают величину $a = 0$. В центральных моментах за постоянную величину a принимают среднюю арифметическую. Центральные моменты 3 и 4 порядков положены в основу расчета показателей формы распределений: коэффициента скошенности (асимметрии) – As и коэффициента островершинности (эксцесса) – E .

Типовые задачи

Задача 1.1 Построение рядов распределения по одному признаку

Условие: в таблице 1.1 представлены данные о числе микроорганизмов в почве (тыс. кол/га) и плотности дерново-подзолистой почвы в пахотном слое (г/см³).

Построить и изобразить графически: 1) дискретный ряд распределения образцов по числу микроорганизмов; 2) ранжированный и интервальный ряды распределения образцов по плотности дерново-подзолистой почвы; 3) сделать выводы.

Таблица 1.1

Исходные данные для построения рядов распределения

Номер образца	Число микроорганизмов, тыс. кол/га	Номер образца	Плотность дерново- подзолистой почвы в пахотном слое, г/см ³
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
1	40	1	1,19
2	34	2	1,24
3	39	3	1,10
4	41	4	1,44
5	38	5	1,39
6	41	6	1,27
7	36	7	1,30
8	41	8	1,16
9	40	9	1,33
10	40	10	1,25
11	38	11	1,18
12	42	12	1,28
13	39	13	1,39
14	36	14	1,23

Номер образца	Число микроорганизмов, тыс. кол/га	Номер образца	Плотность дерново-подзолистой почвы в пахотном слое, г/см ³
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
15	44	15	1,18
16	38	16	1,26
17	36	17	1,31
18	43	18	1,28
19	35	19	1,36
20	40	20	1,33
21	35	21	1,22
22	38	22	1,40
23	36	23	1,46
24	42	24	1,14
25	38	25	1,21
26	39	26	1,32
27	39	27	1,37
28	42	28	1,22
29	39	29	1,25
30	38	30	1,34

Решение:

1. Построим *дискретный ряд распределения* единиц совокупности (образцов) по дискретному признаку (числу микроорганизмов). Для этого выпишем в таблицу 1.2 в порядке возрастания все встречающиеся значения признака – x_i (графа 1) и подсчитаем, сколько раз встречается каждое значение в совокупности (частоту встречаемости признака) – f_i (графа 2).

Таблица 1.2

Дискретный ряд распределения образцов по числу микроорганизмов

Число микроорганизмов (значение признака – x_i)	Число образцов (частота – f_i)
<i>1</i>	<i>2</i>
34	1
35	2
36	4
38	6
39	5
40	4
41	3
42	3
43	1
44	1
Итого	30

Изобразим дискретный ряд графически в виде *полигона* (рисунок 1.1). Для построения полигона в прямоугольной системе координат находим положение точек, абсциссами которых являются значения признака, ординатами – соответствующие частоты встречаемости признака, и соединим

эти точки отрезками прямой. Для получения замкнутой линии из точек крайних значений признака можно опустить перпендикуляры на ось абсцисс.

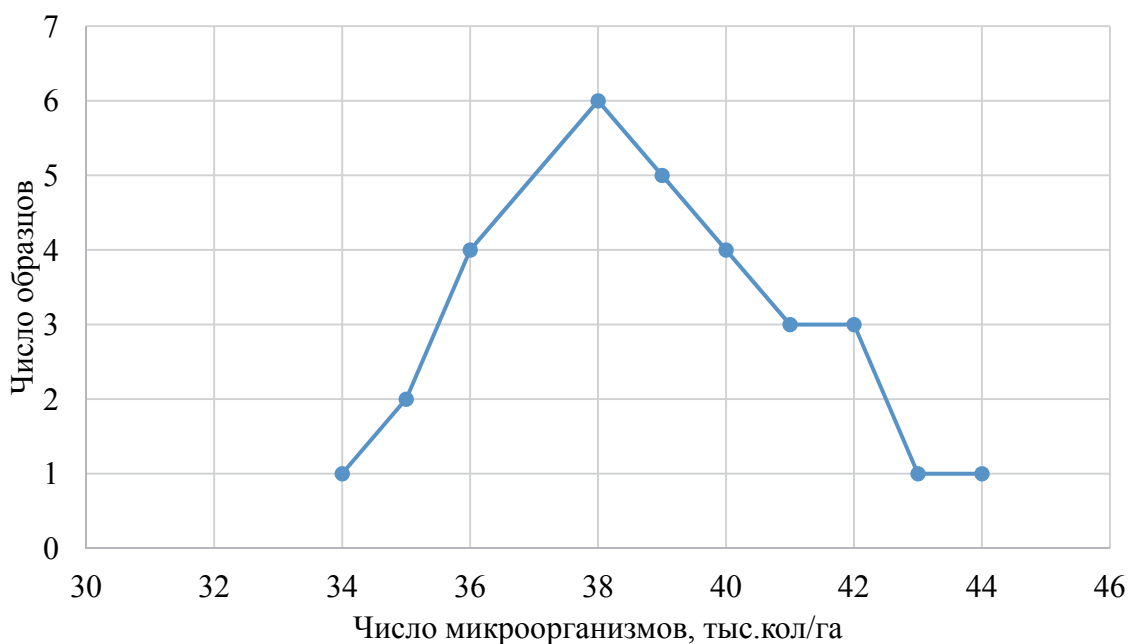


Рисунок 1.1 – Полигон распределения образцов почвы по числу микроорганизмов

Вывод: дискретный ряд распределения и его график показывают, что распределение единиц совокупности по значениям признака неравномерно. Наиболее часто в образцах встречается 38 тыс. кол/га микроорганизмов. Группы образцов с минимальным числом микроорганизмов (34-35) и максимальным (43-44) малочисленны. Распределение по форме близко к нормальному распределению.

2. Построим *ранжированный ряд распределения* образцов по плотности дерново-подзолистой почвы, то есть расположим образцы в порядке возрастания плотности, указав номер образца по ранжиру (таблица 1.3).

Таблица 1.3

Ранжированный ряд распределения образцов по плотности дерново-подзолистой почвы (пахотный слой – г/см³)

Номер по ранжиру	Номер образца	Плотность дерново-подзолистой почвы, г/см ³
<i>1</i>	<i>2</i>	<i>3</i>
1	3	1,10
2	24	1,14
3	8	1,16
4	15	1,18
5	11	1,18
6	1	1,19
7	25	1,21
8	28	1,22
9	21	1,22

Номер по ранжиру	Номер образца	Плотность дерново-подзолистой почвы, г/см ³
1	2	3
10	14	1,23
11	2	1,24
12	29	1,25
13	10	1,25
14	16	1,26
15	6	1,27
16	18	1,28
17	12	1,28
18	7	1,30
19	17	1,31
20	26	1,32
21	20	1,33
22	9	1,33
23	30	1,34
24	19	1,36
25	27	1,37
26	5	1,39
27	13	1,39
28	22	1,40
29	4	1,44
30	23	1,46

Изобразим ранжированный ряд графически посредством *огивы Гальтона*, для чего в прямоугольной системе координат построим точки, абсциссы которых – это номера единиц (образцов) в ранжированном ряду, ординаты – соответствующие им значения признака (плотности почвы), и соединим их отрезками прямой (рисунок 1.2).

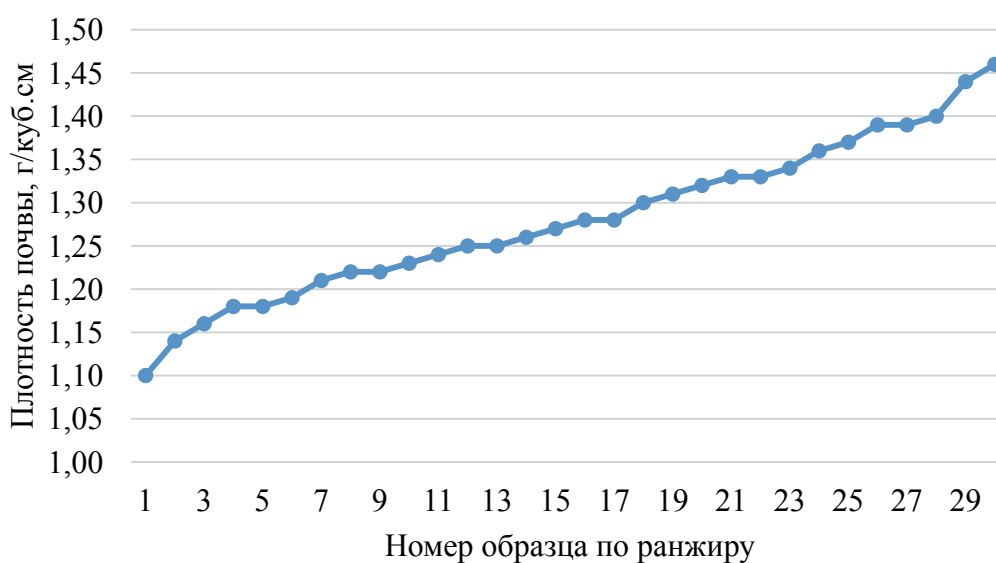


Рисунок 1.2 – Огива распределения образцов по плотности дерново-подзолистой почвы (огива Гальтона)

Вывод: ранжированный ряд распределения и его график показывают, что в изучаемой совокупности значение признака от единицы к единице изменяется в основном плавно, постепенно. Минимальное значение признака – $1,10 \text{ г/см}^3$, максимальное значение признака – $1,46 \text{ г/см}^3$.

3. *Интервальный ряд распределения* строится в следующей последовательности:

- а) определение группировочного признака, т. е. признака, по которому следует формировать группы;
- б) построение ранжированного ряда по группировочному признаку;
- в) анализ характера изменения признака в ранжированном ряду (по огиве Гальтона);
- г) определение числа групп;
- д) определение шага интервала (h);
- е) определение границ интервалов;
- ж) подсчет числа единиц (частот встречаемости) в каждом интервале.

При неравномерном (скачкообразном) изменении признака формируются группы с неравными интервалами. Число групп определяется по количеству «скачков», т.е. резких переходов от одних значений признака к другим.

При равномерном изменении признака формируются группы с равными интервалами, и их число определяется по формуле:

$$m = 1 + 3,322 \cdot \lg N, \quad (1.1)$$

где m – число групп,

N – число единиц в совокупности (в примере 30 образцов).

Шаг интервала рассчитывается по формуле:

$$h = \frac{x_{\max} - x_{\min}}{m} \quad (1.2)$$

где h – шаг интервала,

$x_{\max}$ – максимальное значение признака,

$x_{\min}$ – минимальное значение признака,

m – число групп.

В нашем примере группировочным признаком является плотность почвы. Ранжированный ряд по плотности дерново-подзолистой почвы изображен на рисунке 1.2. Изменение признака плавное, постепенное (на графике нет резких переходов от одного значения плотности почвы к другому), поэтому определяем число интервалов по формуле $m = 1 + 3,322 \cdot \lg N = 1 + 3,322 \cdot \lg 30 = 5,9 \approx 6$. Учитывая, что число групп должно быть более 5, определим 6 групп.

Определим величину шага интервала:

$$h = \frac{1,46 - 1,10}{6} = 0,06 \text{ г/см}^3.$$

В каждом интервале две границы: нижняя и верхняя. Разность между верхней и нижней границей интервала равна шагу интервала h . При плавном,

постепенном изменении значения признака для первой группы нижняя граница интервала равна первому значению признака в ранжированном ряду. В нашем примере значение $x_{min} = 1,10$. Прибавляя к этому значению шаг интервала h , получаем верхнюю границу первой группы $1,10+0,06=1,16$. Верхняя граница первого интервала является одновременно нижней границей второго интервала. Прибавляя к ней шаг интервала, определяем верхнюю границу второго интервала $1,16+0,06=1,22$. Второй интервал в нашей задаче имеет границы: 1,16-1,22. Аналогично определим границы остальных четырех интервалов, и запишем в графу 2 таблицы 1.4.

Подсчет числа единиц (частот встречаемости) в каждом интервале проводится по ранжированному ряду. Если значение признака попадает на границу групп (например, первой и второй групп), то, как правило, единицу учитывают по верхней границе (в первой группе) по принципу «включительно». Результаты подсчета в нашем примере запишем в графу 3 таблицы 1.4.

Таблица 1.4

Интервальный ряд распределения образцов по плотности дерново-подзолистой почвы (пахотный слой), г/см³

Номер группы	Интервалы по плотности почвы, г/см ³ x_i	Число образцов f_i
1	2	3
1	1,10-1,16	3
2	1,16-1,22	6
3	1,22-1,28	8
4	1,28-1,34	6
5	1,34-1,40	5
6	1,40-1,46	2
Итого	x	30

Изобразим интервальный ряд распределения графически в виде *гистограммы*. Для построения гистограммы на оси абсцисс прямоугольной системы координат откладываем интервалы значений признака и на них строим прямоугольники с высотами, пропорциональными частотам интервалов (рисунок 1.3).

Вывод: интервальный ряд распределения и его график показывают, что распределение единиц по группам неравномерное. Наиболее часто встречаются единицы со значением признака в интервале 1,22-1,28 г/см³. Наименьшее число единиц приходится на интервал 1,40-1,46 г/см³. Распределение единиц по форме близко к нормальному распределению.

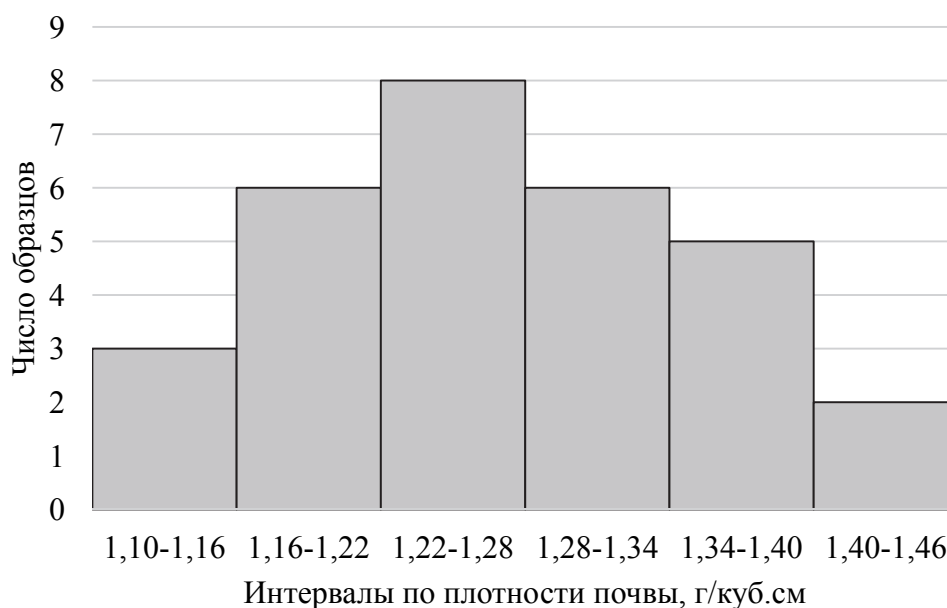


Рисунок 1.3 – Гистограмма распределения образцов по плотности дерново-подзолистой почвы

Задача 1.2 Расчет средней арифметической и медианы в ранжированном ряду распределения

Условие: имеются данные ранжированного ряда образцов по плотности дерново-подзолистой почвы (п. 2. задачи 1.1).

Определить среднюю арифметическую и медиану в ранжированном ряду.

Решение:

1. *Средняя арифметическая в ранжированном ряду* определяется по *простой форме*, так как в ранжире каждому значению признака соответствует частота встречаемости, равная 1 ($f_i = 1$):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}. \quad (1.3)$$

где x_i – индивидуальные значения признака (варианты),

n – численность единиц совокупности.

Рассчитаем среднюю арифметическую:

$$\bar{x} = \frac{1,10 + 1,14 + \dots + 1,46}{30} = 1,28 \text{ г/см}^3$$

Вывод: в среднем по всем образцам плотность почвы составила 1,28 г/см³.

2. *Медиана в ранжированном ряду*, как характеристика середины ряда распределения, при нечетном числе единиц определяется как значение признака единицы, делящей ранжир пополам; при четном числе единиц совокупности равна полусумме двух значений признака с серединными адресами.

Адрес медианы определяется по формуле:

$$n_{Me} = \frac{(N + 1)}{2}. \quad (1.4)$$

Определим адрес медианы:

$$n_{Me} = \frac{(30 + 1)}{2} = 15,5.$$

Адрес медианы равен 15,5, это означает, что медиана находится между 15 и 16 номерами образцов по ранжиру.

Образцы под номерами 15 и 16 имеют плотность почвы 1,27 и 1,28 г/см³, соответственно, медиана будет равна:

$$x_{Me} = \frac{1,27 + 1,28}{2} = 1,275 \text{ г/см}^3.$$

Вывод: Половина образцов дерново-подзолистой почвы имеет плотность 1,275 и менее г/см³, а другая половина образцов – плотность 1,275 и более г/см³.

Задача 1.3 Расчет средней арифметической, моды и медианы в дискретном ряду распределения

Условие: имеются данные дискретного ряда распределения образцов по числу микроорганизмов (результаты построения п. 1 задачи 1.1. представлены в таблице 1.2).

Определить среднюю арифметическую величину, моду, медиану.

Решение:

1. *Средняя арифметическая* в дискретном ряду определяется по *взвешенной форме*, так как каждая градация признака имеет разные частоты встречаемости:

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}, \quad (1.5)$$

где x_i – индивидуальные значения признака (варианты),

f_i – частоты встречаемости индивидуальных значений признака в совокупности (статистические веса значений признака).

Таблица 1.5

Исходные и расчетные данные для определения средних величин в дискретном ряду распределения

Число микроорганизмов (значение признака – x_i)	Число образцов (частота встречаемости значения признака – f_i)	Расчетные данные	
		$x_i f_i$	Накопленные частоты (S_f)
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
34	1	34	1
35	2	70	3
36	4	144	7
38	6	228	13

Число микроорганизмов (значение признака – x_i)	Число образцов (частота встречаемости значения признака – f_i)	Расчетные данные	
		$x_i f_i$	Накопленные частоты (S_f)
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
39	5	195	18
40	4	160	22
41	3	123	25
42	3	126	28
43	1	43	29
44	1	44	30
Итого	30	1167	x

Расчеты выполним в следующей последовательности:

а) Составим макет таблицы 1.5 и перенесем в него исходные данные из таблицы 1.2.

б) Рассчитаем произведения индивидуальных значений признака и их частот встречаемости ($x_i f_i$) и запишем в графу 3 таблицы 1.5, а затем найдем сумму произведений и запишем ее в итоговой строке данной графы.

в) Подставив в формулу (1.5) данные из таблицы 1.5 (итоги граф 3 и 2), получаем:

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} = \frac{1167}{30} = 38,9 \approx 39 \text{ тыс. кол./га.}$$

Вывод: образцы изучаемой совокупности в среднем содержат по 39 тыс. кол/га микроорганизмов в почве.

2. Определим *модальное значение* признака (x_{Mo}) в дискретном ряду.

Расчетный метод.

Проанализировав частоты встречаемости, отмечаем максимальное значение – 6. Эту частоту имеет значение признака $x_i=38$. Следовательно, $x_{Mo}=38$.

Графический метод.

На графике полигона (рисунок 1.1) отмечаем самую высокую точку по оси ординат. Из этой точки опустим перпендикуляр на ось абсцисс. Фиксируем значение признака – 38 (рисунок 1.4).

Вывод: в изучаемой совокупности наиболее часто встречаются образцы, содержащие по 38 тыс. кол/га микроорганизмов в дерново-подзолистой почве.

3. Определим *медианное значение* (x_{Me}) в дискретном ряду.

Расчетный (интерполяционный) метод.

а) Следует установить адрес единицы, находящейся в середине ряда. Ее адрес будет равен:

$$n_{Me} = \frac{(N + 1)}{2} = \frac{(30 + 1)}{2} = 15,5.$$

Так как в нашей совокупности четное число единиц, то адрес медианы состоит из адресов двух центральных единиц. Таким образом, медианное

значение признака равно половине суммы индивидуальных значений единиц с адресом n_{15} и n_{16} .

Примечание: В рядах распределений с нечетным числом единиц совокупности имеется конкретная единица, делящая ряд пополам.

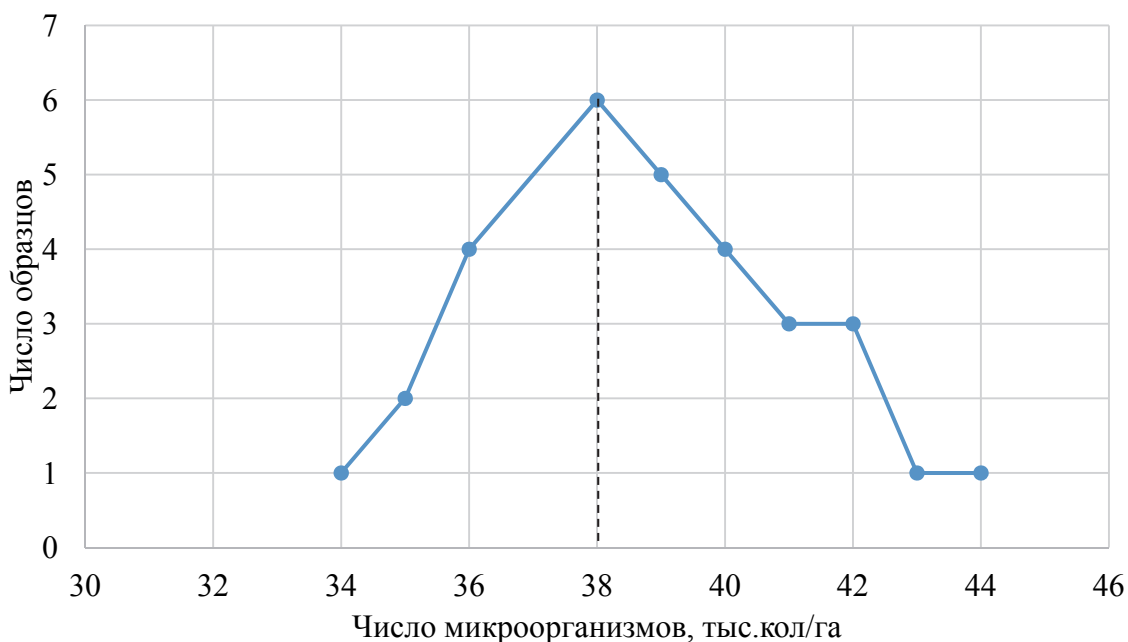


Рисунок 1.4 – Полигон распределения образцов почвы по числу микроорганизмов

б) Для нахождения значения признака единицы с адресом n_{15} рассчитаем накопленные частоты и запишем в таблицу 1.5 в графу 4. Для расчета накопленных частот (S_f) сложим последовательно частоту каждого значения признака с суммой всех предыдущих частот. Для первого значения признака $x_1=34$ частота встречаемости $f_1=1$. Поскольку предыдущих значений признака нет, то накопленная частота для значения 34 равна также 1. Для значения $x_2=35$ накопленная частота равна сумме частот 1 и 2. Следовательно, S_f для $x_2=35$ равна $1+2=3$. Для $x_3=36$ сумма накопленных частот S_f равна $3+4=7$ и т.д. Следует помнить, что сумма накопленных частот всей совокупности S_f равна общей численности совокупности $\sum_{i=1}^n f_i$.

в) Определим по сумме накопленных частот, что единицы с адресами от № 15 по № 16 совокупности имеют значение 39. Именно в этом диапазоне находятся интересующие нас единицы n_{15} и n_{16} . Следовательно, $x_{Me}=39$.

Графический метод.

а) Изобразим ряд распределения графически в виде кумулятивной кривой, для чего в прямоугольной системе координат строим точки, абсциссами которых являются значения признака, ординатами — накопленные частоты. Точки соединим отрезками прямой (рисунок 1.5).

б) Максимальную ординату, соответствующую общему числу единиц совокупности, делим пополам, проводим горизонтальную линию до пересечения с кумулятивной кривой. Из точки пересечения опускаем

перпендикуляр на ось абсцисс. Точка пересечения перпендикуляра и оси абсциссы соответствует медианному значению признака, равному 39.

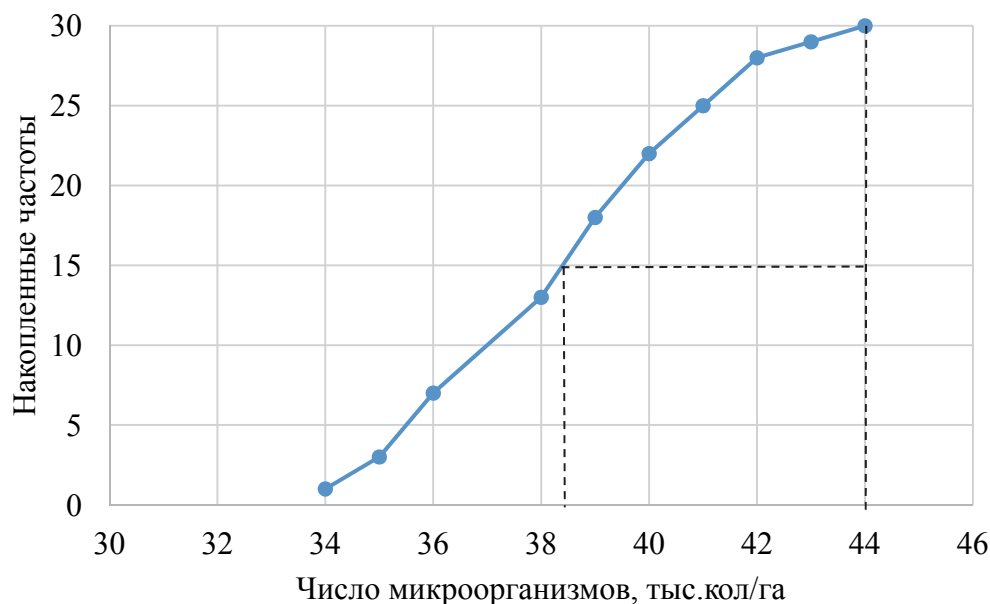


Рисунок 1.5 – Кумулятивная кривая дискретного ряда распределения

Вывод: половина образцов почвы содержат микроорганизмов 39 тыс. кол/га и менее, половина образцов – 39 и более тыс. кол/га.

Задача 1.4 Расчет средней арифметической, моды и медианы в интервальном ряду распределения

Условие: имеются данные интервального ряда распределения образцов по плотности дерново-подзолистой почвы (результаты решения задачи 1.1 в таблице 1.4.)

Определить среднюю арифметическую, моду и медиану плотности почвы и сделать выводы.

Решение:

Составим макет таблицы 1.6 и перенесем в него исходные данные из таблицы 1.4.

1. Расчет *средней арифметической величины* проводится по *взвешенной* форме, аналогично расчету в дискретном ряду, с тем различием, что за индивидуальные значения признака (x_i) в интервальном ряду условно принимаем срединные значения интервалов. Срединные значения интервалов определяем, как полусумму значений нижней и верхней границ интервалов (графа 4 таблицы 1.6).

Например,

$$x_1 = \frac{1,10+1,16}{2} = 1,13; x_2 = \frac{1,16+1,22}{2} = 1,19 \text{ и т. д.}$$

Таблица 1.6

Исходные и расчетные данные для определения средней арифметической величины, моды и медианы в интервальном ряду

Номер группы	Интервал по плотности почвы, г/см ³	Число образцов (f_i)	Срединное значение интервала (x_i)	Произведение вариант на частоту ($x_i f_i$)	Накопленные частоты (S_f)
1	2	3	4	5	6
1	1,10-1,16	3	1,13	3,39	3
2	1,16-1,22	6	1,19	7,14	9
3	1,22-1,28	8	1,25	10,00	17
4	1,28-1,34	6	1,31	7,86	23
5	1,34-1,40	5	1,37	6,85	28
6	1,40-1,46	2	1,43	2,86	30
Итого	х	30	х	38,10	х

а) Найдем произведение вариант (x_i) на частоту (f_i) (графа 5) и сумму произведений (итог графы 5).

в) Рассчитаем среднюю арифметическую взвешенную величину по формуле (1.5):

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} = \frac{38,1}{30} = 1,27 \text{ г/см}^3.$$

Вывод: средняя плотность дерново-подзолистой почвы в изучаемой совокупности образцов составляет 1,27 г/см³.

Примечание: средняя арифметическая по данным интервального ряда (1,27 г/см³) незначительно отличается от средней, определенной по данным ранжированного ряда (1,28 г/см³). Расхождение обусловлено тем, что в интервальном ряду расчеты проводились по усредненным значениям x_i в каждом интервале, а не по конкретным x_i .

2. Определим модальное значение признака (x_{Mo}) в интервальном ряду
Расчетный (интерполяционный) метод.

Модальное значение определяется поэтапно: сначала определяется модальный интервал, затем в нем находится x_{Mo} .

Определяем модальный интервал. Модальным будет интервал с наибольшей частотой встречаемости признака. В примере модальный интервал (от 1,22 до 1,28), так как он имеет максимальную частоту – 8.

Вычисляем модальное значение признака, для чего используем формулу:

$$x_{Mo} = x_0 + h \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}, \quad (1.6)$$

где x_0 – начальное значение модального интервала,

f_{Mo} – частота модального интервала,

f_{Mo-1} – частота интервала, предшествующего модальному интервалу,

f_{Mo+1} – частота интервала, следующего за модальным интервалом,

h – шаг интервала.

Мода составит:

$$x_{Mo} = 1,22 + 0,06 \cdot \frac{(8 - 6)}{(8 - 6) + (8 - 6)} = 1,25 \text{ г/см}^3.$$

Вывод: наиболее часто в образцах совокупности плотность дерново-подзолистой почвы составляет $1,25 \text{ г/см}^3$.

Графический метод.

С этой целью используется гистограмма распределения (рисунок 1.3). Перенесем график на рисунок 1.6.

Графическим методом модальное значение признака будет определяться в следующей последовательности:

а) определим модальный интервал, т.е. столбик гистограммы с наибольшей высотой;

б) точку, соответствующую верхней границе модального интервала, соединяем отрезком прямой с точкой, соответствующей верхней границе предыдущего интервала;

в) точку, соответствующую нижней границе модального интервала, соединяем с точкой, соответствующей нижней границе интервала, последующего за модальным;

г) из точки пересечения прямых опускаем перпендикуляр на ось абсцисс и фиксируем модальное значение признака. Искомая точка на графике соответствует модальному значению – $1,25 \text{ г/см}^3$.

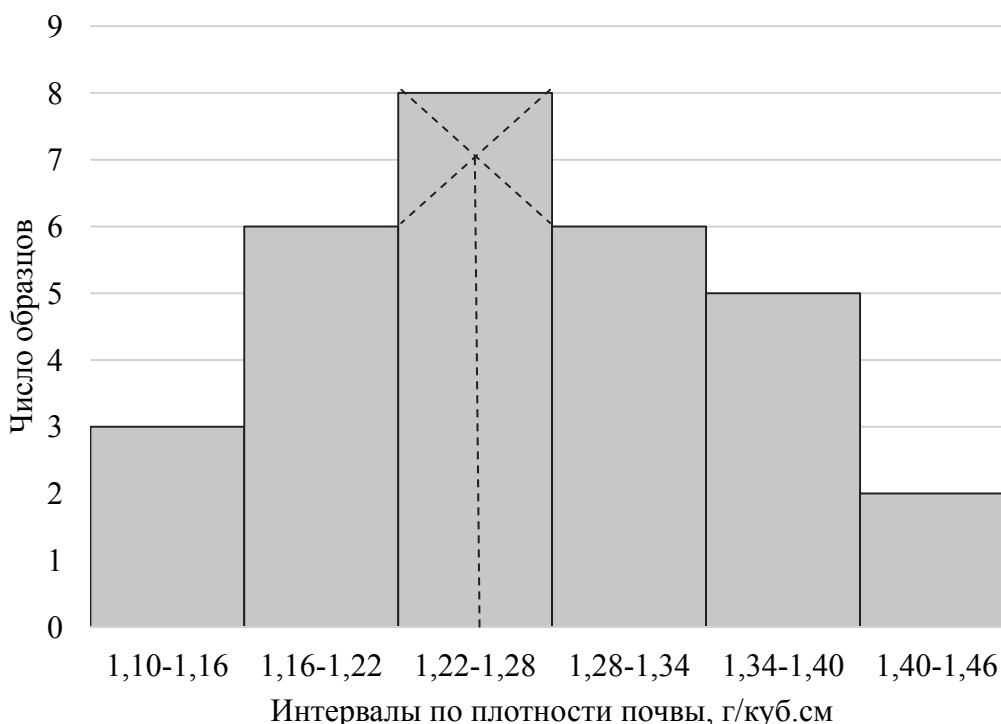


Рисунок 1.6 – Определение моды графическим методом

3. Рассчитаем *медианное значение* признака (x_{Me}) в интервальном ряду:
Расчетный (интерполяционный) метод

Определение медианы в интервальном ряду по алгоритму схоже с определением модального значения: сначала определяется медианный интервал, а затем в нем по формуле рассчитывается конкретное значение медианы.

Для определения медианного интервала для каждого интервала определим накопленную частоту (таблица 1.6, графа 6). Установим адрес медианы:

$$n_{Me} = \frac{(N + 1)}{2} = \frac{(30 + 1)}{2} = 15,5.$$

Таким образом, медианное значение признака равно половине суммы индивидуальных значений единиц с адресом n_{15} и n_{16} . Обе единицы находятся в интервале 1,22-1,28. Он является медианным.

Рассчитаем медианное значение признака по формуле:

$$x_{Me} = x_0 + h \cdot \frac{\frac{N(+1)}{2} - S_{Me-1}}{f_{Me}}, \quad (1.7)$$

где x_0 – начальное значение медианного интервала,

h – шаг интервала,

N – общее число единиц совокупности,

$(+1)$ – используется в рядах с нечетным числом единиц совокупности,

S_{Me-1} – накопленная частота домедианного интервала,

f_{Me} – частота медианного интервала.

Подставив необходимые величины в формулу (1.7), получаем:

$$x_{me} = 1,22 + 0,06 \cdot \frac{\frac{30}{2} - 9}{8} = 1,265 \text{ г/см}^3.$$

Графический метод.

Применение аналогично рассмотренному выше примеру дискретного ряда с тем различием, что за варианты принимаются верхние границы интервалов.

Графическим методом медианное значение признака будет определяться в следующей последовательности:

а) изобразим ряд распределения графически в виде кумулятивной кривой (рисунок 1.7). На оси абсцисс отмечаем точки, соответствующие верхним границам интервалов, на оси ординат – накопленные частоты. Точки соединяем отрезками прямой.

б) из максимальной точки кумуляты (30), соответствующей общему числу единиц, опускаем перпендикуляр на ось абсцисс, делим перпендикуляр пополам и из точки деления проводим горизонтальную линию до пересечения с кумулятой. Далее из точки пересечения с кумулятой опускаем перпендикуляр на ось ОХ. Точка пересечения на оси ОХ указывает значение медианы, равное 1,265.

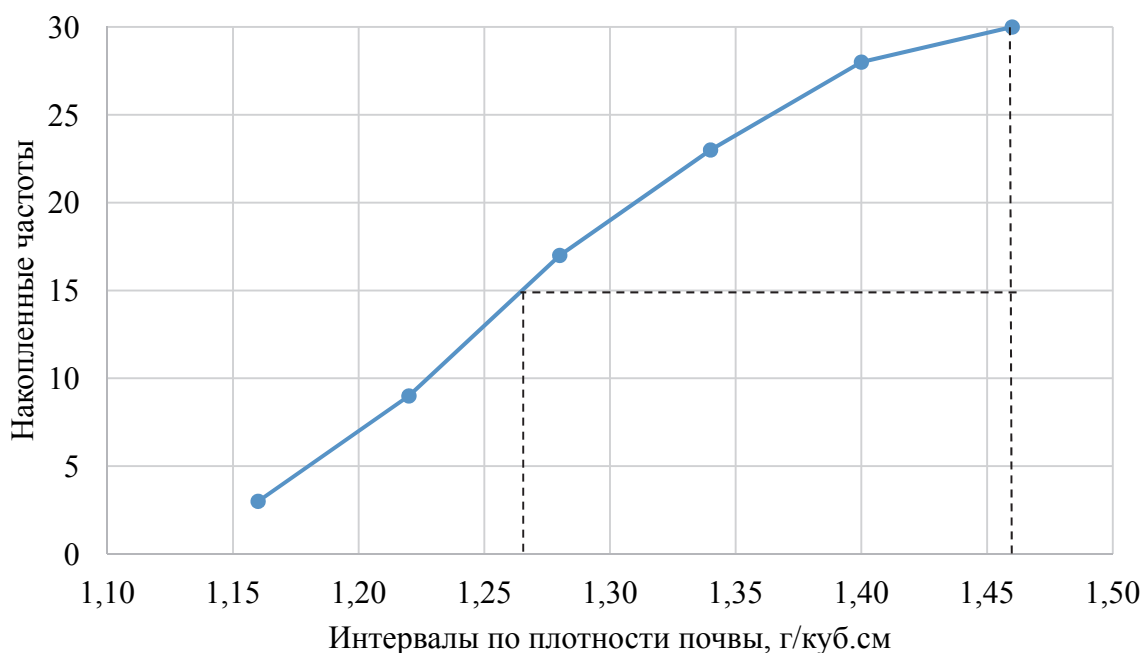


Рисунок 1.7 – Медиана на кумулятивной кривой интервального ряда распределения

Вывод: половина образцов изучаемой совокупности имеют плотность $1,265 \text{ г/см}^3$ и менее, половина – $1,265$ и более г/см^3 .

Задача 1.5 Анализ взаимосвязи средней арифметической, моды и медианы в умеренно асимметричных рядах

Условие: имеются показатели интервального ряда распределения – средняя арифметическая, мода и медиана (результаты решения задачи 1.4).

Проанализировать фактические соотношения между показателями и сделать выводы.

Решение:

В одномодальном симметричном ряду распределения средняя арифметическая, мода и медиана совпадают, т.е. $\bar{x}_{\text{арифм.}} = x_{\text{Ме}} = x_{\text{Мо}}$

Для умеренно асимметричных рядов справедливо приближенное соотношение между медианой, модой и средней арифметической.

$$x_{\text{Ме}} = \frac{1}{3} \cdot x_{\text{Мо}} + \frac{2}{3} \cdot \bar{x}_{\text{арифм.}} \quad (1.8)$$

После преобразования формулы получим соотношения:

$$x_{\text{Мо}} = \bar{x}_{\text{арифм.}} - 3 \cdot (\bar{x}_{\text{арифм.}} - x_{\text{Ме}}), \quad (1.9)$$

$$x_{\text{Мо}} = \bar{x}_{\text{арифм.}} + 3 \cdot (x_{\text{Ме}} - \bar{x}_{\text{арифм.}}). \quad (1.10)$$

По данным задачи 1.4 $\bar{x}_{\text{арифм.}} = 1,27$; $x_{\text{Ме}} = 1,265$; $x_{\text{Мо}} = 1,25$.

Вывод: характеристики близки, но не совпадают, следовательно, распределение асимметрично по форме. Так как $\bar{x}_{\text{арифм.}} > x_{\text{Ме}} > x_{\text{Мо}}$, в

распределении наблюдается правосторонняя скошенность (правая ветвь кривой распределения длиннее левой).

По соотношению (1.10) получаем: $1,27 + 3 \cdot (1,265 - 1,27) = 1,255$. Значения показателя, рассчитанные по соотношению (1.10) и полученные по исходным данным, отличаются незначительно, следовательно, можно заключить, что асимметрия умеренная.

Задача 1.6 Расчет квартилей в интервальном ряду распределения

Условие: имеются данные интервального ряда распределения (таблица 1.4).

Определить: нижний (1-й), средний (2-й) и верхний (3-й) квартили. Сделать выводы.

Решение.

1) Первый квартиль (Q_1) разделяет совокупность в соотношении 1:3, второй квартиль (Q_2) в отношении 2:2, третий квартиль (Q_3) в отношении 3:1. В соответствии с этим определим адреса квартилей по накопленным частотам. Так как второй квартиль равен медиане, рассмотренной выше (задача 1.4), определим адреса первого и третьего квартилей.

Адрес для первого квартиля $n_1 = \frac{1}{4} \cdot 30 = 7,5$. Первым квартилем является полусумма значений признака с номерами 7 и 8 в ранжированном ряду: $Q_1 = \frac{x_7 + x_8}{2}$. Адрес третьего квартиля $n_3 = \frac{3}{4} \cdot 30 = 22,5$. Третьим квартилем является полусумма значений признака с номерами 22 и 23 в ранжированном ряду: $Q_3 = \frac{x_{22} + x_{23}}{2}$.

2) В квартильных интервалах квартили определяются по следующим формулам:

– нижний квартиль:

$$Q_1 = x_0 + h \frac{\frac{N}{4} - S_{Q_1-1}}{f_{Q_1}}; \quad (1.11)$$

– средний квартиль:

$$Q_2 = x_{Me} = x_0 + h \frac{\frac{N}{2} - S_{Me-1}}{f_{Me}}; \quad (1.12)$$

– верхний квартиль:

$$Q_3 = x_0 + h \frac{\frac{3N}{4} - S_{Q_3-1}}{f_{Q_3}}. \quad (1.13)$$

Определим квартили:

– нижний квартиль:

$$Q_1 = 1,16 + 0,06 \frac{\frac{30}{4} - 3}{6} = 1,205;$$

– средний квартиль (медиану):

$$Q_2 = x_{Me} = 1,22 + 0,06 \frac{\frac{30}{2} - 9}{8} = 1,265;$$

– верхний квартиль:

$$Q_3 = 1,28 + 0,06 \frac{\frac{3 \cdot 30}{4} - 17}{6} = 1,335.$$

Вывод: четвертая часть образцов имеет плотность почвы менее 1,205 г/см³, половина образцов – не менее 1,265 г/см³, у три четверти образцов плотность почвы менее 1,335 г/см³, и одна четверть – свыше 1,335 г/см³.

Задача 1.7 Расчет и анализ показателей вариации в ранжированном ряду распределения

Условие: имеются данные ранжированного ряда распределения образцов по плотности дерново-подзолистой почвы (таблица 1.3).

Определить показатели вариации: размах вариации, среднее линейное отклонение, дисперсию (по основной формуле и рабочей формулам), среднее квадратическое отклонение, коэффициент осцилляции, относительное линейное отклонение и коэффициент вариации.

Решение:

Составим макет таблицы 1.7 и запишем исходные данные в графы 1 и 2.

Таблица 1.7

Исходные и расчетные данные для определения показателей вариации в ранжированном ряду распределения

Исходные данные		Расчетные данные			
Номер по ранжиру	Плотность дерново-подзолистой почвы, г/см ³ x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$	x_i^2
1	2	3	4	5	6
1	1,10	-0,18	0,18	0,0324	1,2100
2	1,14	-0,14	0,14	0,0196	1,2996
3	1,16	-0,12	0,12	0,0144	1,3456
4	1,18	-0,10	0,10	0,0100	1,3924
5	1,18	-0,10	0,10	0,0100	1,3924
6	1,19	-0,09	0,09	0,0081	1,4161
7	1,21	-0,07	0,07	0,0049	1,4641
8	1,22	-0,06	0,06	0,0036	1,4884
9	1,22	-0,06	0,06	0,0036	1,4884

Исходные данные		Расчетные данные			
Номер по ранжиру	Плотность дерново-подзолистой почвы, г/см ³ x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$	x_i^2
1	2	3	4	5	6
10	1,23	-0,05	0,05	0,0025	1,5129
11	1,24	-0,04	0,04	0,0016	1,5376
12	1,25	-0,03	0,03	0,0009	1,5625
13	1,25	-0,03	0,03	0,0009	1,5625
14	1,26	-0,02	0,02	0,0004	1,5876
15	1,27	-0,01	0,01	0,0001	1,6129
16	1,28	0,00	0,00	0,0000	1,6384
17	1,28	0,00	0,00	0,0000	1,6384
18	1,30	0,02	0,02	0,0004	1,6900
19	1,31	0,03	0,03	0,0009	1,7161
20	1,32	0,04	0,04	0,0016	1,7424
21	1,33	0,05	0,05	0,0025	1,7689
22	1,33	0,05	0,05	0,0025	1,7689
23	1,34	0,06	0,06	0,0036	1,7956
24	1,36	0,08	0,08	0,0064	1,8496
25	1,37	0,09	0,09	0,0081	1,8769
26	1,39	0,11	0,11	0,0121	1,9321
27	1,39	0,11	0,11	0,0121	1,9321
28	1,40	0,12	0,12	0,0144	1,9600
29	1,44	0,16	0,16	0,0256	2,0736
30	1,46	0,18	0,18	0,0324	2,1316
Итого	38,40	x	2,20	0,2356	49,3876

1. *Размах вариации* определяется как разность между максимальным и минимальным значением признака:

$$R = x_{\max} - x_{\min}. \quad (1.14)$$

Определим размах вариации:

$$R = 1,46 - 1,10 = 0,36 \text{ г/см}^3.$$

Вывод: По образцам максимальные различия составляют 0,36 г/см³.

2. Вычислим *среднее линейное отклонение* по формуле:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}, \quad (1.15)$$

где x_i – индивидуальное значение признака,

$\bar{x}$ – средняя арифметическая,

n – численность единиц совокупности.

Вычислим среднее линейное отклонение в следующей последовательности:

- а) определим отклонение каждой варианты от средней арифметической, определенной в задаче 1.2, $\bar{x}=1,28$ (графа 3);
 б) определим абсолютные отклонения от средней (графа 4);
 в) итог графы 4 делим на число единиц в совокупности:

$$\bar{d} = \frac{2,20}{30} = 0,073.$$

Вывод: среднее отклонение индивидуальных значений признака от средней величины составляет 0,073 г/см³.

3. Расчет дисперсии по *основной формуле* производится следующим образом:

$$\sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (1.16)$$

Определим дисперсию:

- а) для этого вычислим квадраты отклонений от средней (графа 5);
 б) итог графы 5 делим на число единиц в совокупности:

$$\sigma_x^2 = \frac{0,2356}{30} = 0,0079.$$

Вывод: средний квадрат отклонений вариант от средней составил 0,0079 (г/см³)².

4. Расчет дисперсии по *рабочей формуле* производится следующим образом:

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2 = \overline{x^2} - (\bar{x})^2. \quad (1.17)$$

Для этого:

- а) найдем квадраты значений признака (графа 6);
 б) вычислим дисперсию:

$$\sigma^2 = \frac{49,3876}{30} - \left(\frac{38,40}{30} \right)^2 = 0,0079.$$

5. *Среднее квадратическое отклонение*, определяется как квадратный корень из дисперсии:

$$\sigma = \sqrt{\sigma^2}. \quad (1.18)$$

Рассчитаем *среднее квадратическое отклонение*:

$$\sigma = \sqrt{0,0079} = 0,089.$$

Вывод: среднее квадратическое отклонение показывает, что все варианты отклоняются от средней величины в среднем на 0,089 г/см³.

6. Относительные показатели вариации определяются следующим образом:

– *коэффициент осцилляции*:

$$K_{oc} = \frac{R}{\bar{x}} \cdot 100; \quad (1.19)$$

– *относительное линейное отклонение* (линейный коэффициент вариации):

$$K_{oc} = \frac{\bar{d}}{\bar{x}} \cdot 100; \quad (1.20)$$

– коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} \cdot 100. \quad (1.21)$$

Рассчитаем относительные показатели вариации:

– коэффициент осцилляции:

$$K_{oc} = \frac{0,36}{1,28} \cdot 100 = 28,1\%;$$

Вывод: относительная максимальная колеблемость плотности почвы по образцам средняя.

– относительное линейное отклонение:

$$K_{\bar{d}} = \frac{0,073}{1,28} \cdot 100 = 5,7\%;$$

Вывод: средняя колеблемость плотности почвы по образцам сравнительно низкая и составляет 5,7%, что свидетельствует об однородности совокупности.

– коэффициент вариации:

$$V = \frac{0,089}{1,28} \cdot 100 = 6,9\%;$$

Вывод: коэффициент вариации указывает на то, что мера колеблемости в данном ряду распределения низкая (менее 33%), и совокупность является однородной по изучаемому признаку.

Задача 1.8 Расчет и анализ показателей вариации в дискретном ряду распределения

Условие: имеются данные дискретного ряда распределения по числу микроорганизмов в почве (таблица 1.2).

Определить показатели вариации: размах вариации, среднее линейное отклонение, дисперсию (по основной формуле), среднее квадратическое отклонение, коэффициент осцилляции, относительное линейное отклонение и коэффициент вариации.

Решение:

Составим макет таблицы 1.8 и запишем исходные данные в графы 1 и 2.

Таблица 1.8

Исходные и расчетные данные для определения показателей вариации в дискретном ряду распределения

Исходные данные		Расчетные данные				
Число микроорганизмов (x_i)	Число образцов (f_i)	$x_i - \bar{x}$	$ x_i - \bar{x} $	$ x_i - \bar{x} f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
1	2	3	4	5	6	7
34	1	-5	5	5	25	25

Исходные данные		Расчетные данные				
Число микроорганизмов (x_i)	Число образцов (f_i)	$x_i - \bar{x}$	$ x_i - \bar{x} $	$ x_i - \bar{x} f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
1	2	3	4	5	6	7
35	2	-4	4	8	16	32
36	4	-3	3	12	9	36
38	6	-1	1	6	1	6
39	5	0	0	0	0	0
40	4	1	1	4	1	4
41	3	2	2	6	4	12
42	3	3	3	9	9	27
43	1	4	4	4	16	16
44	1	5	5	5	25	25
Итого	30	x	x	59	x	183

1. Определим размах вариации:

$$R = 44 - 34 = 10 \text{ тыс. кол/га микроорганизмов.}$$

Вывод: По образцам максимальные различия составляют 10 тыс. кол/га.

2. Вычислим среднее линейное отклонение по формуле:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i}{\sum_{i=1}^n f_i}, \quad (1.22)$$

где x_i – значение признака;

$\bar{x}$ – средняя арифметическая;

f_i – частота встречаемости признака в совокупности.

Вычислим среднее линейное отклонение в следующей последовательности:

а) определим отклонение каждой варианты от средней $\bar{x}=39$ (графа 3);

б) определим абсолютные отклонения от средней (графа 4);

в) определим абсолютные отклонения, взвешенные соответствующими частотами, и их сумму (графа 5);

г) итог графы 5 делим на итог графы 2:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i}{\sum_{i=1}^n f_i} = \frac{59}{30} = 1,97.$$

Вывод: среднее отклонение индивидуальных значений признака от средней величины составляет 1,97 тыс. кол/га.

3. Расчет дисперсии по основной формуле:

$$\sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}. \quad (1.23)$$

Определим дисперсию:

а) для этого вычислим квадраты отклонений от средней (графа 6);

б) определим взвешенные частотами квадраты отклонений и их сумму (графа 7);

в) итог графы 7 делим на итог графы 2:

$$\sigma_x^2 = \frac{183}{30} = 6,10.$$

Вывод: средний квадрат отклонений вариант от средней составил 6,10 (тыс. кол/га)².

4. Рассчитаем среднее квадратическое отклонение:

$$\sigma = \sqrt{6,10} = 2,47.$$

Вывод: среднее квадратическое отклонение показывает, что все варианты отклоняются от средней величины в среднем на 2,47 тыс. кол/га.

Примечание: Если распределение близко к нормальному, или симметричному, то между среднеквадратичным и среднелинейным отклонением существует взаимосвязь $\sigma = 1,25L$ или $L = 0,8\sigma$.

5. Рассчитаем относительные показатели вариации:

– коэффициент осцилляции:

$$K_{oc} = \frac{10}{39} \cdot 100 = 25,6\%.$$

Вывод: относительная максимальная колеблемость числа микроорганизмов по образцам средняя.

– относительное линейное отклонение:

$$K_{\bar{a}} = \frac{1,97}{39} \cdot 100 = 5,0\%.$$

Вывод: средняя колеблемость количества микроорганизмов по образцам сравнительно низкая и составляет 5,0%, что свидетельствует об однородности совокупности.

– коэффициент вариации:

$$V = \frac{2,47}{39} \cdot 100 = 6,3\%.$$

Вывод: коэффициент вариации указывает на то, что мера колеблемости в данном ряду распределения низкая (менее 33%), и совокупность является однородной по изучаемому признаку.

Задача 1.9 Исчисление показателей вариации в интервальном ряду распределения

Условие: имеются данные интервального ряда распределения образцов по плотности дерново-подзолистой почвы в пахотном слое (г/см³) (таблица 1.4).

Определить показатели вариации. Дисперсию вычислить по основной и рабочей формулам.

Решение:

1. Определим размах вариации:

$$R = x_{max} - x_{min} = 1,46 - 1,10 = 0,36 \text{ г/см}^3.$$

Вывод: В изучаемой совокупности образцов максимальные различия в плотности составляют 0,36 г/см³.

2. Вычислим среднее линейное отклонение.

а) определим отклонение каждой варианты от средней $\bar{x}=1,27 \text{ г/см}^3$ (данное значение мы взяли из выполненной задачи 1.4) и запишем результаты в таблицу (графа 5, таблица 1.9);

б) определим абсолютное отклонение, взвешенное соответствующими частотами, и их сумму (графа 5);

в) рассчитаем среднее линейное отклонение (итог графы 6 делим на итог графы 3):

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i}{\sum_{i=1}^n f_i} = \frac{2,12}{30} = 0,07.$$

Вывод: среднее отклонение признака от средней величины составляет 0,07, т.е. в среднем по образцам плотность дерново-подзолистой почвы отклоняется от средней плотности на 0,07 г/см³.

3. Исчислим дисперсию по основной формуле:

а) вычислим квадраты отклонений от средней (графа 7);

б) определим взвешенные частотами квадраты отклонений и их сумму (графа 8).

Примечание: квадраты отклонений следует считать при достаточно большой степени точности (до 0,0000).

в) итог графы 8 делим на итог графы 3:

$$\sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i} = \frac{0,2112}{30} = 0,007 \text{ (г/см}^3\text{)}^2.$$

4. Определим дисперсию по рабочей формуле.

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 f_i}{\sum_{i=1}^n f_i} - \left(\frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \right)^2 = \overline{x^2} - (\bar{x})^2. \quad (1.24)$$

Для этого:

а) найдем квадраты значений признака (графа 9);

б) рассчитаем взвешенные частотами квадраты значений признака и их сумму (графа 10);

в) вычислим дисперсию:

$$\sigma^2 = \frac{48,5982}{30} - \left(\frac{38,10}{30} \right)^2 = 0,007 \text{ (г/см}^3\text{)}^2.$$

5. Рассчитаем среднее квадратическое отклонение как корень квадратный из дисперсии:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,007} = 0,084 \text{ г/см}^3.$$

Вывод: среднее квадратическое отклонение показывает, что все варианты отклоняются от средней величины в среднем на 0,084 г/см³.

6. Рассчитаем коэффициент осцилляции:

$$K_{oc} = \frac{R}{\bar{x}} \cdot 100\% = \frac{0,36}{1,27} \cdot 100 = 28,3 \ %.$$

Вывод: относительная колеблемость крайних значений плотности почвы по образцам умеренная.

Таблица 1.9

К расчету показателей вариации в интервальном ряду

Но- мер груп- пы	Интер- валы по Плот- ности почвы (г/см ³)	Число образц ов	Средин- ное значение интер- вала	Отклоне- ние варианты от средней	Взвешенное абсолютное отклонение	Квадрат отклоне- ний	Взвешен- ный квадрат отклоне- ний	Квадрат вариан- ты	Взвешен- ный квадрат варианты	Произ- ведение вариан- та на частоту
		f_i	x_i	$x_i - \bar{x}$	$ x_i - \bar{x} f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$	x_i^2	$x_i^2 f_i$	$x_i f_i$
1	2	3	4	5	6	7	8	9	10	11
1	1,10- 1,16	3	1,13	-0,14	0,14	0,42	0,0196	0,0588	1,2769	3,39
2	1,16- 1,22	6	1,19	-0,08	0,08	0,48	0,0064	0,0384	1,4161	7,14
3	1,22- 1,28	8	1,25	-0,02	0,02	0,16	0,0004	0,0032	1,5625	10,00
4	1,28- 1,34	6	1,31	0,04	0,04	0,24	0,0016	0,0096	1,7161	7,86
5	1,34- 1,40	5	1,37	0,10	0,10	0,50	0,0100	0,0500	1,8769	6,85
6	1,40- 1,46	2	1,43	0,16	0,16	0,32	0,0256	0,0512	2,0449	2,86
Итого	х	30	х	х	2,12	х	0,2112	х	48,5982	38,10

7. Рассчитаем относительное линейное отклонение:

$$K_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\% = \frac{0,07}{1,27} \cdot 100 = 5,5 \, \%.$$

Вывод: средняя колеблемость плотности по образцам низкая и составляет 5,5%, что свидетельствует об однородности совокупности.

8. Определим коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{0,084}{1,27} \cdot 100 = 6,6 \, \%.$$

Вывод: средняя мера колеблемости в данном ряду распределения невысокая (менее 33%), и совокупность является однородной по изучаемому признаку.

Задача 1.10 Правило сложения дисперсий (разложения объемов вариаций)

Условие: имеются результаты опытов по урожайности картофеля (ц/га) при различных видах удобрений (таблица 1.10).

Таблица 1.10

Урожайность картофеля при разных видах удобрений, ц/га

Удобрение	Урожайность картофеля по повторностям, ц/га				Итого	Средняя урожайность по виду удобрения
	1	2	3	4		
Без удобрений	240	262	273	263	1038	259,5
РК	325	378	346	327	1376	344,0
Навоз	362	384	360	330	1436	359,0
Навоз + РК	401	419	425	441	1686	421,5
Итого	1328	1443	1404	1361	5536	346,0

Рассчитать общий, межгрупповой и внутригрупповой объемы вариации, применив основные и рабочие формулы. По расчетным данным доказать справедливость теоремы разложения объемов вариации.

Определить общую, межгрупповую и внутригрупповую дисперсии, эмпирический коэффициент детерминации и корреляционное отношение. Сделать выводы о связи видов удобрений и урожайности картофеля.

Решение:

Правило сложения дисперсий: при сгруппированных единицах совокупности общая дисперсия ($\sigma_{\text{общ}}^2$) равна сумме межгрупповой ($\sigma_{\text{мгр}}^2$) и средней из внутригрупповых дисперсий ($\bar{\sigma}_{\text{внгр}}^2$):

$$\sigma_{\text{общ}}^2 = \sigma_{\text{мгр}}^2 + \bar{\sigma}_{\text{внгр}}^2. \quad (1.25)$$

При этом *общая дисперсия* представляет собой средний квадрат отклонений значений признака от средней по всей совокупности.

Межгрупповая дисперсия – это средний квадрат отклонений групповых средних от общей средней величины по всей совокупности. *Средняя из внутригрупповых дисперсий* представляет собой среднюю арифметическую взвешенную из внутригрупповых дисперсий. Взвешивание проводится по числу единиц совокупности в каждой группе. Каждая из внутригрупповых дисперсий представляет средний квадрат отклонений индивидуальных значений признака в пределах каждой группы от своей групповой средней величины.

Каждая из дисперсий рассчитывается как отношение объемов вариации к численности единиц совокупности:

$$\sigma_{\text{общ}}^2 = \frac{W_{\text{общ}}}{N}; \quad \sigma_{\text{мгр}}^2 = \frac{W_{\text{мгр}}}{N}; \quad \bar{\sigma}_{\text{внгр}}^2 = \frac{W_{\text{внгр}}}{N}. \quad (1.26)$$

Поскольку знаменатели дробей равны между собой, равенство справедливо также для числителей, т.е. общий объем вариации равен сумме межгруппового и внутригруппового объемов вариации. Таким образом, на первом этапе расчетов докажем равенство объемов вариаций, на втором – равенство дисперсий.

Вычисление объемов вариаций проведем двумя способами: по основным и рабочим формулам.

Для расчета объемов вариаций необходимо определить общую среднюю по совокупности и групповые средние по каждой группе.

Общая средняя величина по совокупности равна:

$$\bar{x}_{\text{общ}} = \frac{\sum_{i=1}^m \sum_{j=1}^n x_{ij}}{N} = \frac{240 + 262 + \dots + 425 + 441}{16} = \frac{5536}{16} = 346,0 \text{ ц/га.}$$

Групповые средние равны:

$$\bar{x}_1 = \frac{\sum_{j=1}^{n_1} x_{1j}}{n_1} = \frac{240 + 262 + 273 + 263}{4} = \frac{1038}{4} = 259,5 \text{ ц/га;}$$

$$\bar{x}_2 = \frac{\sum_{j=1}^{n_2} x_{2j}}{n_2} = \frac{325 + 378 + 346 + 327}{4} = \frac{1376}{4} = 344,0 \text{ ц/га;}$$

$$\bar{x}_3 = \frac{\sum_{j=1}^{n_3} x_{3j}}{n_3} = \frac{362 + 384 + 360 + 330}{4} = \frac{1436}{4} = 359,0 \text{ ц/га;}$$

$$\bar{x}_4 = \frac{\sum_{j=1}^{n_4} x_{4j}}{n_4} = \frac{401 + 419 + 425 + 441}{4} = \frac{1686}{4} = 421,5 \text{ ц/га.}$$

Полученные значения запишем в последнюю графу таблицы 1.10.

1. *Основные формулы* расчета общего и межгруппового объемов вариации:

– общий объем вариации:

$$W_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_0)^2; \quad (1.27)$$

– межгрупповой объем вариации:

$$W_{\text{мгр}} = \sum_{i=1}^m (\bar{x}_i - \bar{x}_0)^2 n_i. \quad (1.28)$$

По основным формулам объемы вариации составят:

$$W_{\text{общ}} = (240 - 346)^2 + (262 - 346)^2 + \dots + (441 - 346)^2 = 58108;$$

$$W_{\text{мгр}} = (259,5 - 346,0)^2 \cdot 4 + (344,0 - 346,0)^2 \cdot 4 + \\ + (359,0 - 346,0)^2 \cdot 4 + (421,5 - 346,0)^2 \cdot 4 = 53422,0.$$

Для расчета внутригруппового объема вариации по совокупности в целом предварительно определим внутригрупповой объем вариации в каждой группе по формуле:

$$W_{\text{внгр}i} = \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2. \quad (1.29)$$

Объем вариации в первой группе равен:

$$W_{\text{внгр}1} = \sum_{j=1}^n (x_{1j} - \bar{x}_1)^2 = (240,0 - 259,5)^2 + (262,0 - 259,5)^2 + \\ + (273,0 - 259,5)^2 + (263,0 - 259,5)^2 = 581,0.$$

Объем вариации в второй группе составляет:

$$W_{\text{внгр}2} = \sum_{j=1}^n (x_{2j} - \bar{x}_2)^2 = (325,0 - 344,0)^2 + (378,0 - 344,0)^2 + \\ + (346,0 - 344,0)^2 + (327,0 - 344,0)^2 = 1810,0.$$

Объем вариации в третьей группе равен:

$$W_{\text{внгр}3} = \sum_{j=1}^n (x_{3j} - \bar{x}_3)^2 = (362,0 - 359,0)^2 + (384,0 - 359,0)^2 + \\ + (360,0 - 359,0)^2 + (330,0 - 359,0)^2 = 1476,0.$$

Объем вариации в четвертой группе равен:

$$W_{\text{внгр}4} = \sum_{j=1}^n (x_{4j} - \bar{x}_4)^2 = (401,0 - 421,5)^2 + (419,0 - 421,5)^2 + \\ + (425,0 - 421,5)^2 + (441,0 - 421,5)^2 = 819,0.$$

Складывая $W_{\text{внгр}1}$, $W_{\text{внгр}2}$, $W_{\text{внгр}3}$, $W_{\text{внгр}4}$, получаем внутригрупповой объем вариации по совокупности:

$$W_{\text{внгр}} = W_{\text{внгр}1} + W_{\text{внгр}2} + W_{\text{внгр}3} + W_{\text{внгр}4}. \quad (1.30)$$

Общий внутригрупповой объем вариации по совокупности в целом равен:

$$W_{\text{внгр}} = 581,0 + 1810,0 + 1476,0 + 819,0 = 4686,0.$$

Рассчитанные показатели доказывают справедливость равенства объемов вариации:

$$W_{\text{общ}} = W_{\text{мгр}} + W_{\text{внгр}}; \quad (1.31)$$

$$58108 = 53422 + 4686.$$

2. Рассчитаем объемы вариации признака *по рабочим формулам*.

– общий объем вариации:

$$W_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N}; \quad (1.32)$$

– межгрупповой объем вариации

$$W_{\text{мгр}} = \sum_{i=1}^m \left(\frac{(\sum_{j=1}^n x_{ij})^2}{n_i} \right) - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N}; \quad (1.33)$$

– внутригрупповой объем вариации

$$W_{\text{внгр}} = \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - \sum_{i=1}^m \left(\frac{(\sum_{j=1}^n x_{ij})^2}{n_i} \right). \quad (1.34)$$

Для проведения расчетов объемов вариации по рабочим формулам необходимы дополнительные характеристики, которые получим в таблице 1.11.

Таблица 1.11

Таблица квадратов

Удобрение	Урожайность картофеля по повторностям, ц/га				Сумма квадратов	Квадрат суммы $(\sum_{j=1}^n x_{ij})^2$
	1	2	3	4		
Без удобрений	57600	68644	74529	69169	269942	1077444
РК	105625	142884	119716	106929	475154	1893376
Навоз	131044	147456	129600	108900	517000	2062096
Навоз + РК	160801	175561	180625	194481	711468	2842596
Сумма квадратов	455070	534545	504470	479479	1973564	x

Общий объем вариации равен:

$$W_{\text{общ}} = 1973564 - \frac{5536^2}{16} = 1973564,0 - 1915456,0 = 58108,0.$$

Определим межгрупповой объем вариации:

$$W_{\text{мгр}} = \left(\frac{1038^2}{4} + \frac{1376^2}{4} + \frac{1436^2}{4} + \frac{1686^2}{4} \right) - \frac{5536^2}{16} = 1968878,0 - 1915456,0 = 53422.$$

Внутригрупповой объем вариации по рабочей формуле равен:

$$W_{\text{внгр}} = 1973564 - \left(\frac{1038^2}{4} + \frac{1376^2}{4} + \frac{1436^2}{4} + \frac{1686^2}{4} \right) = 1973564 - 1968878 = 4686.$$

Таким образом, по рабочим формулам также доказали справедливость равенства:

$$W_{\text{общ}} = W_{\text{мгр}} + W_{\text{внгр}};$$

$$58108 = 53422 + 4686.$$

3. Для расчета дисперсий необходимо разделить объемы вариаций на численность единиц совокупности:

$$\begin{aligned}\sigma_{\text{общ}}^2 &= \frac{W_{\text{общ}}}{N} = \frac{58108}{16} = 3631,750; \\ \sigma_{\text{мгр}}^2 &= \frac{W_{\text{мгр}}}{N} = \frac{53422}{16} = 3338,875; \\ \bar{\sigma}_{\text{внгр}}^2 &= \frac{W_{\text{внгр}}}{N} = \frac{4686}{16} = 292,875.\end{aligned}$$

Таким образом, получено, что общая дисперсия ($\sigma_{\text{общ}}^2$) равна сумме межгрупповой ($\sigma_{\text{мгр}}^2$) и средней из внутригрупповых дисперсий ($\bar{\sigma}_{\text{внгр}}^2$):

$$\begin{aligned}\sigma_{\text{общ}}^2 &= \sigma_{\text{мгр}}^2 + \bar{\sigma}_{\text{внгр}}^2 \\ 3631,750 &= 3338,875 + 292,875.\end{aligned}$$

7. Для оценки связи вариации результативного признака с группировочным признаком определяют коэффициент детерминации, рассчитываемый как удельный вес межгрупповой дисперсии в общей дисперсии:

$$\eta^2 = \frac{\sigma_{\text{мгр}}^2}{\sigma_{\text{общ}}^2} \cdot 100. \quad (1.35)$$

Определим коэффициент детерминации:

$$\eta^2 = \frac{3338,875}{3631,750} \cdot 100 = 91,9\%.$$

Коэффициент детерминации η^2 показывает, что 91,9% вариации урожайности картофеля в совокупности обусловлено различиями в используемых удобрениях, остальные 8,1% – другими факторами.

Извлекая корень квадратный из коэффициента детерминации получим корреляционное отношение:

$$\eta = \sqrt{\eta^2}. \quad (1.36)$$

Корреляционное отношение составит:

$$\eta = \sqrt{0,919} = 0,959.$$

По предложенной Чеддоком шкале (таблица 1.12) значение 0,959 свидетельствует об очень сильном влиянии вида удобрения на урожайность картофеля.

Таблица 1.12

Шкала Чеддока

η	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
влияние	слабое	умеренное	заметное	сильное	очень сильное

Задача 1.11 Расчет начальных и центральных моментов интервального ряда распределения

Условие: имеются данные интервального ряда распределения образцов дерново-подзолистой почвы по плотности (пахотный слой – г/см³) (задача 1.1 таблица 1.4).

Определить начальные и центральные статистические моменты первых четырех порядков. Проверить наличие взаимосвязи между ними.

Решение.

Общий вид статистического момента распределения следующий:

$$M_k = \frac{\sum_{i=1}^n (x_i - a)^k f_i}{\sum_{i=1}^n f_i}, \quad (1.37)$$

где x_i – значение признака,

f_i – частота встречаемости признака в совокупности,

a – некоторая постоянная величина,

k – показатель степени.

В практических расчетах обычно используют моменты первых четырех порядков ($k=1,2,3,4$).

В начальных моментах принимают величину $a=0$. Формула начального момента:

$$m_k = \frac{\sum_{i=1}^n x_i^k f_i}{\sum_{i=1}^n f_i}. \quad (1.38)$$

Начальный момент первого порядка представляет собой среднюю арифметическую, и применяется для характеристики центральной тенденции:

$$m_1 = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}.$$

Начальные моменты второго, третьего и четвертого порядков применяют для расчета центральных моментов на основе формул взаимосвязи.

В центральных моментах за постоянную величину a принимают среднюю арифметическую. Центральные моменты имеют вид:

$$\mu_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^k f_i}{\sum_{i=1}^n f_i}. \quad (1.39)$$

Центральный момент второго порядка является дисперсией и используется для количественной характеристики вариации:

$$\mu_2 = \sigma^2 = \frac{\sum_{i=1}^l (x_i - \bar{x})^2 f_i}{\sum_{i=1}^l f_i}.$$

Центральные моменты 3 и 4 порядков положены в основу расчета показателей формы распределений: коэффициента скошенности (асимметрии) – As и коэффициента островершинности (экспесса) – E .

Составим макет таблицы и запишем исходные данные в таблицу 1.13.

1. Определим *начальные моменты* первых четырех порядков. Найдем взвешенные значения вариантов в степени k при $k=2,3,4$ и их суммы

(графы 4-7). Поделив итоги граф 4, 5, 6 и 7 на итог графы 2, получим соответственно начальные моменты:

– первого порядка:

$$m_1 = \bar{x} = \frac{\sum_{i=1}^l x_i f_i}{\sum_{i=1}^l f_i} = \frac{38,1}{30} = 1,27;$$

– второго порядка:

$$m_2 = \overline{x^2} = \frac{\sum_{i=1}^l x_i^2 f_i}{\sum_{i=1}^l f_i} = \frac{48,5982}{30} = 1,6199;$$

– третьего порядка:

$$m_3 = \overline{x^3} = \frac{\sum_{i=1}^l x_i^3 f_i}{\sum_{i=1}^l f_i} = \frac{65,2584}{30} = 2,0753;$$

– четвертого порядка:

$$m_4 = \overline{x^4} = \frac{\sum_{i=1}^l x_i^4 f_i}{\sum_{i=1}^l f_i} = \frac{80,1017}{30} = 2,6701.$$

2. Рассчитаем *центральные моменты* 2-4 порядков по основным формулам. Для этого в таблице 1.13. найдем значения отклонений от средней арифметической в степени k при $k=2,3,4$ (графы 8-10), взвешенные на число единиц в каждой группе. Необходимые величины подставим в формулу (1.39) и получим центральные моменты:

– второго порядка:

$$\mu_2 = \sigma^2 = \frac{\sum_{i=1}^l (x_i - \bar{x})^2 f_i}{\sum_{i=1}^l f_i} = \frac{0,2112}{30} = 0,00704;$$

– третьего порядка:

$$\mu_3 = \frac{\sum_{i=1}^l (x_i - \bar{x})^3 f_i}{\sum_{i=1}^l f_i} = \frac{0,0022}{30} = 0,00007;$$

– четвертого порядка:

$$\mu_4 = \frac{\sum_{i=1}^l (x_i - \bar{x})^4 f_i}{\sum_{i=1}^l f_i} = \frac{0,0032}{30} = 0,00011.$$

3. Центральные моменты 2-4 порядков на основе формул их взаимосвязи с начальными моментами определяются следующим образом:

$$\mu_2 = m_2 - m_1^2; \quad (1.40)$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3; \quad (1.41)$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4. \quad (1.42)$$

Рассчитаем центральные моменты 2-4 порядков:

$$\mu_2 = 1,6199 - 1,27^2 = 0,00704;$$

$$\mu_3 = 2,0753 - 3 \cdot 1,6199 \cdot 1,27 + 2 \cdot 1,27^3 = 0,00007;$$

$$\mu_4 = 2,6701 - 4 \cdot 2,0753 \cdot 1,27 + 6 \cdot 1,6199 \cdot 1,27^2 - 3 \cdot 1,27^4 = 0,00011.$$

Примечание: Небольшие различия в величине центральных моментов, вычисленных по основным формулам и формулам взаимосвязи с начальными моментами, могут быть объяснены округлениями в расчетах.

Таблица 1.13

К расчету начальных и центральных моментов интервального ряда распределения

Исходные данные			Расчетные данные						
Интервалы по плотности	f_i	x_i	Взвешенные значения вариант в степени k				Взвешенные значения отклонений вариант от средней в степени k		
			$x_i f_i$	$x_i^2 f_i$	$x_i^3 f_i$	$x_i^4 f_i$	$(x_i - \bar{x})^2 f_i$	$(x_i - \bar{x})^3 f_i$	$(x_i - \bar{x})^4 f_i$
1	2	3	4	5	6	7	8	9	10
1,10-1,16	3	1,13	3,39	3,8307	4,3287	4,8914	0,0588	-0,0082	0,0012
1,16-1,22	6	1,19	7,14	8,4966	10,1110	12,0320	0,0384	-0,0031	0,0002
1,22-1,28	8	1,25	10,00	12,5000	15,6250	19,5313	0,0032	-0,0001	0,0000
1,28-1,34	6	1,31	7,86	10,2966	13,4885	17,6700	0,0096	0,0004	0,0000
1,34-1,40	5	1,37	6,85	9,3845	12,8568	17,6138	0,0500	0,0050	0,0005
1,40-1,46	2	1,43	2,86	4,0898	5,8484	8,3632	0,0512	0,0082	0,0013
Итого	30	x	38,10	48,5982	62,2584	80,1017	0,2112	0,0022	0,0032

Задача 1.12 Расчет показателей формы распределения

Условие: имеются результаты расчетов начальных и центральных моментов по интервальному ряду распределения (задача 1.11).

Определить показатели формы распределения: коэффициенты скошенности и островершинности. Сделать выводы.

Решение:

Форму распределения характеризуют показатели асимметрии и островершинности, которые показывают степень отклонения формы конкретного распределения (дискретного или интервального ряда) от нормально распределенной совокупности.

Коэффициент скошенности (асимметрии) характеризует степень асимметричности распределения. Для оценки асимметрии применяются два показателя.

Первый коэффициент скошенности определяется как отношение центрального момента 3-го порядка к среднему квадратическому отклонению в 3-ей степени.

$$As = \frac{\mu_3}{\sigma^3}. \quad (1.43)$$

При правосторонней скошенности коэффициент асимметрии положительный, при левосторонней – отрицательный.

Второй коэффициент (предложил К. Пирсон) рассчитывается как отношение разности средней арифметической и моды распределения к среднему квадратическому отклонению:

$$As = \frac{\bar{x} - x_{Mo}}{\sigma}. \quad (1.44)$$

Определим коэффициенты скошенности (асимметрии):

$$As = \frac{0,00007}{0,084^3} = 0,125;$$

$$As \text{ Пирсона} = \frac{1,27 - 1,25}{0,084} = 0,238.$$

В нашей задаче оба коэффициента положительны, поэтому можно сделать вывод о положительной, правосторонней скошенности. Коэффициент асимметрии Пирсона характеризует, прежде всего, асимметрию середины распределения, а показатель, рассчитанный по центральному моменту – асимметрию крайних значений распределения. Поскольку коэффициент Пирсона в нашей задаче больше по значению, можно заключить, что в центре распределения асимметрия выражена сильнее. Оба коэффициента несильно отличаются от 0, поэтому следует сделать вывод о невысокой, умеренной асимметрии. Этот вывод наглядно подтверждает график на рисунке 1.3. (гистограмма распределения в задаче 1.1.)

При изучении формы вершины распределения за стандарт принимают вершину кривой нормального распределения. Отличающиеся от стандарта

распределения делят на островершинные и плосковершинные. Для измерения степени островершинности (эксцесса) используют нормированное отношение $\frac{\mu_4}{\sigma^4}$. Для нормального распределения нормированное отношение $\frac{\mu_4}{\sigma^4} = 3$. Следовательно, если нормированное отношение $\frac{\mu_4}{\sigma^4}$ будет больше трех – распределение островершинно, меньше трех – распределение плосковершинно. Поэтому, непосредственно для характеристики островершинности применяют коэффициент островершинности (эксцесса):

$$E = \frac{\mu_4}{\sigma^4} - 3. \quad (1.45)$$

Рассчитаем коэффициент эксцесса:

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{0,00011}{0,084^4} - 3 = -0,831.$$

Коэффициент эксцесса отрицательный, значительно отличается от 0, поэтому указывает на плосковершинность по сравнению с нормально распределенной совокупностью.

Вывод: интервальный ряд распределения имеет умеренную правостороннюю скошенность и плосковершинность по сравнению с нормальным распределением.

Задача 1.13 Расчет средней арифметической и вариации альтернативного признака

Условие: из 30 образцов 6 образцов имеют плотность дерново-подзолистой почвы ниже 1,20 г/см³.

Определить среднюю долю образцов, имеющих плотность почвы ниже 1,20 г/см³, и вариацию этой доли в совокупности.

Решение

Средняя арифметическая альтернативного признака приравнивается доле единиц, обладающих этим признаком:

$$\bar{x} = p. \quad (1.46)$$

Определим среднюю арифметическую альтернативного признака (доли образцов с плотностью ниже 1,20 г/см³):

$$\bar{x} = \frac{6}{30} = 0,20.$$

Дисперсия альтернативного признака равна произведению доли признака, обладающего альтернативным признаком (p) и доли единиц, не обладающих этим признаком (q):

$$\sigma^2 = p \cdot q. \quad (1.47)$$

Определим дисперсию альтернативного признака:

$$\sigma^2 = p \cdot q = 0,20 \cdot 0,80 = 0,16.$$

Рассчитаем среднее квадратическое отклонение признака:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,16} = 0,4.$$

Вывод: в среднем доля образцов с плотностью ниже $1,20 \text{ г/см}^3$ равна 0,20 при средней колеблемости признака $0,4 \text{ г/см}^3$.

Вопросы для проверки знаний по разделу I

1. Что такое статистическая совокупность?
2. Что есть определяющее свойство?
3. Какие признаки относятся к количественным?
4. В чем отличие признаков дискретных и непрерывных?
5. В чем отличие первичных и вторичных признаков?
6. Что такое ранжированный ряд распределения?
7. Как выглядит табличная форма ранжированного ряда?
8. Как называется графическая форма ранжированного ряда и как она выглядит?
9. Каково содержание колонок в дискретном вариационном ряду?
10. Что откладывается на координатных осях при построении полигона распределения?
11. Как определить число интервалов?
12. По каким признакам чаще всего строится интервальный вариационный ряд?
13. Как определить шаг интервала?
14. Как определить границы интервалов?
15. Что откладывается на координатных осях при построении гистограммы распределения?
16. Что произойдет с дисперсией признака, если все значения признака одновременно уменьшить на 10 и увеличить в 2 раза?
17. В партии продукции на некачественную приходится 10%. Определить коэффициент вариации.
18. Какой показателей вариации может быть использован для сравнения вариации нескольких признаков? Почему?
19. Если в распределении коэффициент эксцесса равен 0, чему равно отношение $\frac{\sum(x_i - \bar{x})^4}{\sigma^4}$?
20. Определить объем вариации, если среднее квадратическое отклонение равно 5, а совокупность состоит из 4-х единиц.
21. Какие показатели характеризуют среднюю колеблемость признака в совокупности?
22. Определить дисперсию альтернативного признака, если известно, что средняя в 3 раза превышает среднее квадратическое отклонение.
23. Что произойдет с величиной моды, если все значения признака уменьшить на 10?
24. Имеется распределение с параметрами: средняя арифметическая равна 10, среднее квадратическое отклонение равно 4. Используя

математические свойства средней арифметической и среднего квадратического отклонения, получить распределение с параметрами: средняя арифметическая равна 0, среднее квадратическое – 1.

25. Во сколько раз межгрупповая вариация меньше внутригрупповой, если корреляционное отношение равно 0,25?
26. Групповые средние по совокупности равны соответственно 25, 100, 250. Группы равной численности. Найти общую среднюю.
27. Найти межгрупповую вариацию по следующим данным:

№ группы	Средняя по группе	Число единиц в группе
1	5	5
2	7	5
3		
По совокупности в целом	10	15

28. Определите межгрупповую вариацию, если по 3-м группам имеются следующие значения признака:
- 1 группа: 2, 3, 4;
- 2 группа: 10, 12, 11;
- 3 группа: 25, 26, 24.
29. Постройте график ранжированного ряда, если известно, что совокупность имеет пять единиц и вариация признака равна нулю.

РАЗДЕЛ II

ВЫБОРОЧНЫЙ МЕТОД.

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Методические указания

Статистическая оценка означает, что выводы относительно статистических параметров (средней, дисперсии, коэффициента вариации) генеральной совокупности делаются на основе изучения выборочных данных. Исследуемый параметр генеральной совокупности при этом может быть охарактеризован одним числом (*точечная оценка*), либо указывается интервал, в котором он находится с определенной вероятностью (*интервальная оценка*). Как точечная, так и интервальная оценка предполагают, что сначала значения соответствующих параметров определяются по выборке. Такие значения носят название *оценок*.

При *точечной* оценке параметра генеральной совокупности (Q_0) его значение приравнивается к оценке (Q_B), но в силу того, что оценка определена с использованием случайно сформированной выборки, она (оценка) обязательно должна быть дополнена показателем ее средней ошибки (m_Q). Алгоритм расчета средней ошибки зависит от того, для какого параметра она рассчитывается. В итоге результаты точечной оценки параметра генеральной совокупности записываются следующим образом: $Q_0 = Q_B$ со средней ошибкой $m_Q = a$.

При *интервальной* оценке параметра генеральной совокупности так же, как и при точечной, сначала необходимо определить его значение по выборке, т.е. оценку. Значение оценки принимается за центр интервала. Далее устанавливается величина случайной предельной ошибки оценки (E_Q), на основе которой и устанавливаются границы интервала. В нем с некоторой гарантией (доверительным уровнем вероятности) находится значение исследуемого параметра в генеральной совокупности. Следовательно, интервальная оценка параметра генеральной совокупности предполагает следующую запись:

$$Q_0 = Q_B \pm E_Q \quad (2.1)$$

или

$$Q_B - E_Q \leq Q_0 \leq Q_B + E_Q \quad (2.2)$$

где $(Q - E_Q)$ – нижняя граница интервала,

$(Q + E_Q)$ – верхняя граница интервала.

При расчете предельных значений ошибки используется формула:

$$E_Q = m_Q \cdot t_P, \quad (2.3)$$

где m_Q – средняя ошибка оценки;

t_P – коэффициент (нормированное отклонение), зависящий от гарантии (доверительного уровня вероятности) того, что исследуемый

параметр генеральной совокупности не выйдет за установленные границы интервала.

Отсюда следует, что для расчета предельных границ ошибки вначале следует найти среднюю ошибку оценки, затем определиться с гарантией (доверительным уровнем вероятности) и в зависимости от ее величины использовать соответствующий коэффициент.

В практических исследованиях гарантия (доверительный уровень вероятности) берется не ниже 0,90 (0,90; 0,95; 0,99; 0,999). Порядок нахождения коэффициента (нормированного отклонения) зависит от численности выборки, которая используется для статистической оценки параметра генеральной совокупности. При $n \geq 30$ коэффициент определяется из таблицы «Значение интеграла вероятностей при разных значениях t », так как ошибки оценок при выборках численностью более 30 единиц распределяются в соответствии с законом нормального распределения. (Приложение 1) и его величина зависит только от доверительного уровня вероятности. При численности выборки менее 30 единиц следует воспользоваться таблицей «Значение двустороннего критерия -Стьюдента» (Приложение 2), где значение коэффициента зависит не только от доверительного уровня вероятности, но и от численности выборки (степеней свободы вариации).

Статистические гипотезы относительно тех или иных свойств статистических совокупностей (законов распределения, параметров распределения) могут быть проверены по данным выборочного наблюдения. С этой целью необходимо выдвинуть так называемую нулевую гипотезу (H_0). Решение о принятии или отказе от нулевой гипотезы должно быть основано на сопоставлении фактического значения того или иного математического критерия (критерия -нормального распределения, t -Стьюдента, F -распределения, χ^2 и др.) с его критическим значением. При выборе критерия необходимо учитывать характер решаемой задачи по проверке гипотезы, объем выборок и другие условия. В результате проверки по данным выборок нулевую гипотезу следует принять, если фактическое значение критерия попадает в область допустимых значений критерия. Если же оно попадает в критическую область, от нулевой гипотезы следует отказаться.

Проверка гипотез относительно распределений по критерию χ^2 -Пирсона имеет три области применения:

1) как *критерий согласия*, по которому проверяется согласие (соответствие) фактического выборочного распределения известному теоретическому распределению. При этом теоретическое распределение может быть представлено в виде функции (например, функции нормального распределения, распределения Пуассона и т.п.) или в виде теоретического соотношения частот n_i (например, соотношения 3:1 или 9:3:3:1).

2) как *критерий независимости*, по которому проверяется независимость распределения частот совокупности (n_i) по одному признаку от распределения частот этой совокупности по другому признаку.

3) как критерий однородности, по которому проверяется, однородны ли по структуре частот две совокупности, распределенные по одному признаку.

В соответствии с выбранной областью применения критерия выдвигается нулевая гипотеза (H_0) относительно свойств одного или двух распределений. Далее устанавливается область применения критерия и по данным выборочной совокупности определяется фактическое значение критерия ($\chi^2_{\text{факт}}$). При доверительном уровне вероятности суждения и численности выборки (числу степеней свободы) по таблицам определяется теоретическое значение критерия $\chi^2_{\text{таб}}$. При сравнении $\chi^2_{\text{факт}}$ и $\chi^2_{\text{таб}}$ определяем, в какой области (критической или согласия) находится выборочное значение критерия. Если $\chi^2_{\text{факт}} \leq \chi^2_{\text{таб}}$ выборочное значение находится в области согласия, нулевая гипотеза принимается. Напротив, если $\chi^2_{\text{факт}} \geq \chi^2_{\text{таб}}$, выборочное значение находится в критической области и нулевая гипотеза должна быть отвергнута, а следовательно – принята альтернативная гипотеза.

Проверка статистических гипотез относительно одной средней величины или двух средних величин по данным выборок может быть проведена по параметрическому критерию - *Стьюдента* (по малым выборкам $n < 30$) или *t-нормального распределения* (по выборкам достаточно большого объема $n \geq 30$).

В конкретных задачах при выборе алгоритма расчетов по критерию - *Стьюдента* необходимо учитывать:

1) схему формирования выборок (выборки независимые или зависимые);

2) равенство или неравенство объемов выборок;

3) равенство или неравенство дисперсий в генеральных совокупностях.

С целью проверки статистических гипотез относительно двух средних величин по данным двух независимых выборок необходимо сформулировать нулевую гипотезу (H_0) относительно средних величин. Она может быть основана на предположении, что выборки взяты из генеральных совокупностей с равными средними величинами: $H_0: \tilde{x}_1 = \tilde{x}_2$, или $H_0: \tilde{x}_1 - \tilde{x}_2 = 0$. Проверяемой (нулевой) гипотезе противостоит некоторая альтернативная гипотеза (H_1), которая может быть сформулирована в общем виде как $H_1: \tilde{x}_1 \neq \tilde{x}_2$ (ненаправленная гипотеза) или $H_1: \tilde{x}_1 > \tilde{x}_2$, $H_1: \tilde{x}_1 < \tilde{x}_2$ (направленная гипотеза).

Фактическое (выборочное) значение критерия рассчитывают следующим образом:

$$t_{\text{факт}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{m}_{1,2}}, \quad (2.4)$$

где $\bar{m}_{1,2}$ – обобщенная средняя ошибка двух выборочных средних.

Фактическое значение критерия необходимо сопоставить с критическим значением при заданном уровне значимости (t_α). Если выборочное значение критерия попадает в критическую область ($t_{\text{факт}} > t_\alpha$),

нулевая гипотеза о равенстве средних должна быть отвергнута; если же выборочное значение критерия попадает в область допустимых значений ($t_{\text{факт}} \leq t_{\alpha}$), нулевую гипотезу следует принять.

Нулевая гипотеза о равенстве средних в двух генеральных совокупностях может быть проверена путем сравнения разности между выборочными средними ($\bar{x}_1 - \bar{x}_2$) с предельной случайной ошибкой (*наименьшей существенной разностью* – НСР) при заданном уровне значимости (ε_{λ}). Если разность между выборочными средними находится в пределах НСР ($|\bar{x}_1 - \bar{x}_2| \leq \varepsilon_{\lambda}$), нулевую гипотезу следует принять. Если же разность между средними по выборкам выходит за пределы НСР ($|\bar{x}_1 - \bar{x}_2| > \varepsilon_{\lambda}$), нулевая гипотеза должна быть отвергнута.

Так как выборочные дисперсии, как правило, неравны, в ходе проверки статистических гипотез относительно средних величин в двух генеральных совокупностях необходимо проверить гипотезу относительно их дисперсий при помощи критерия *-распределения*. Допустим, что нулевая и альтернативная гипотезы сформулированы следующим образом:

$$H_0: \sigma_1^2 = \sigma_2^2;$$

$$H_1: \sigma_1^2 \neq \sigma_2^2.$$

Для проверки нулевой гипотезы о равенстве дисперсий в генеральных совокупностях следует определить фактическое (выборочное) значение критерия F как отношение выборочных дисперсий:

$$F_{\text{факт}} = \frac{S_1^2}{S_2^2}, \quad (2.5)$$

где $S_1^2 > S_2^2$ и сравнить его с критическим значением при заданном уровне значимости (F_{λ}). Если фактическое значение критерия попадает в область допустимых значений ($F_{\text{факт}} < F_{\lambda}$), нулевая гипотеза о равенстве дисперсий в генеральных совокупностях принимается. Если же фактическое значение критерия попадает в критическую область ($F_{\text{факт}} \geq F_{\lambda}$), от нулевой гипотезы следует отказаться.

Типовые задачи

Задача 2.1 Точечная и интервальная оценка параметров генеральной совокупности по данным выборки

2.1.1 Точечная и интервальная оценка среднего значения признака в генеральной совокупности по данным большой выборки

Условие: имеются данные выборочного наблюдения о плотности дерново-подзолистой почвы (таблица 2.1). Численность выборки составляет $n = 35$.

Провести точечную и интервальную оценку среднего значения признака генеральной совокупности (среднего значения плотности дерново-подзолистой почвы по всем 140 образцам).

Решение:

1. Определим квадраты значений признака (x_i^2) и запишем их в таблицу 2.1.

Таблица 2.1

Плотность дерново-подзолистой почвы по данным выборочного наблюдения

Номер образ-ца	Плотность почвы, г/см ³	Квадрат значения признака	Номер образ-ца	Плотность почвы, г/см ³	Квадрат значения признака
	x_i	x_i^2		x_i	x_i^2
1	1,44	2,0736	19	1,50	2,2500
2	1,02	1,0404	20	1,07	1,1449
3	1,06	1,1236	21	1,37	1,8769
4	1,37	1,8769	22	1,11	1,2321
5	1,24	1,5376	23	1,07	1,1449
6	1,23	1,5129	24	1,32	1,7424
7	1,32	1,7424	25	1,06	1,1236
8	1,18	1,3924	26	1,47	2,1609
9	1,10	1,2100	27	1,54	2,3716
10	1,44	2,0736	28	1,26	1,5876
11	1,50	2,2500	29	1,29	1,6641
12	1,37	1,8769	30	1,16	1,3456
13	1,42	2,0164	31	1,33	1,7689
14	1,12	1,2544	32	1,44	2,0736
15	1,18	1,3924	33	1,28	1,6384
16	1,20	1,4400	34	1,30	1,6900
17	1,36	1,8496	35	1,39	1,9321
18	1,29	1,6641	Итого	44,80	58,0748

2. Подсчитаем $\sum_{i=1}^n x_i$ и $\sum_{i=1}^n x_i^2$ и запишем в таблицу 2.1.

3. Вычислим среднюю величину по данным выборочной совокупности:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{44,80}{35} = 1,28 \text{ г/см}^3.$$

4. Найдем выборочную дисперсию:

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2 = \overline{x^2} - \bar{x}^2 = \frac{58,0748}{35} - 1,28^2 = 0,02088.$$

5. Для получения несмещенной оценки дисперсии выборочную дисперсию следует умножить на поправочный коэффициент $\frac{n}{n-1}$:

$$S^2 = \sigma^2 \cdot \frac{n}{n-1}. \quad (2.6)$$

Рассчитаем несмещенную оценку дисперсии:

$$S^2 = 0,02088 \cdot \frac{35}{35 - 1} = 0,021494.$$

Примечание: несмещенную оценку дисперсии необходимо рассчитывать прежде всего для совокупностей менее 30 единиц.

Несмещенная оценка дисперсии может быть исчислена также без определения выборочной дисперсии по следующей формуле:

$$S^2 = \frac{\sum_{i=1}^n x^2 - n\bar{x}^2}{n - 1}. \quad (2.7)$$

Рассчитаем несмещенную оценку дисперсии по формуле 2.7:

$$S^2 = \frac{58,0748 - 35 \cdot 1,28^2}{35 - 1} = 0,021494.$$

6. Формула расчета средней ошибки выборочной средней:

$$m_x = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}. \quad (2.7)$$

Рассчитаем среднюю ошибку выборочной средней по вышеприведенной формуле:

$$m_{\bar{x}} = \sqrt{\frac{0,021494}{35} \left(1 - \frac{35}{140}\right)} = \sqrt{0,000461} = 0,021461.$$

Примечание: если численность выборки по отношению к численности единиц генеральной совокупности чрезвычайно мала, то поправочным коэффициентом $\left(1 - \frac{n}{N}\right)$ можно пренебречь.

7. При *точечной* оценке генеральная средняя приравняется оценке (выборочной средней) с указанием размера средней ошибки. Проведем точечную оценку средней в генеральной совокупности:

$$\tilde{x} = \bar{x} = 1,28 \text{ г/см}^3 \text{ при } m_x = 0,02 \text{ г/см}^3.$$

8. По таблице «Значение интеграла вероятностей при разных значениях t » (Приложение 1) найдем теоретическое значение t при доверительном уровне вероятности $P = 0,95$: $t = 1,96$.

9. Предельная случайная ошибка выборочной средней определяется, исходя из выражения $t = \frac{\bar{x} - \tilde{x}}{m_{\bar{x}}} = \frac{\varepsilon_{\bar{x}}}{m_{\bar{x}}}$, следующим образом:

$$\varepsilon_{\bar{x}} = t \cdot m_{\bar{x}}. \quad (2.8)$$

Определим предельную случайную ошибку выборочной средней:

$$\varepsilon_{\bar{x}} = t \cdot m_{\bar{x}} = 1,96 \cdot (\pm 0,021461) = \pm 0,042064 \approx 0,04 \text{ (г/см}^3\text{)}.$$

10. При *интервальной* оценке генеральная средняя приравняется к интервалу:

$$\tilde{x} = \bar{x} \pm \varepsilon_x. \quad (2.9)$$

Проведем интервальную оценку, т.е. найдем интервал, в котором с заданным уровнем вероятности находится средняя генеральной совокупности:

$$\tilde{x} = \bar{x} \pm \varepsilon_x = 1,28 \pm 0,04,$$

или

$$\bar{x} - \varepsilon_x \leq \tilde{x} \leq \bar{x} + \varepsilon_x;$$

$$1,24 \leq \tilde{x} \leq 1,32 \text{ (г/см}^3\text{)}$$

11. Сделаем вывод. С доверительным уровнем вероятности 0,95 можно утверждать, что средняя плотность дерново-подзолистой почвы в генеральной совокупности находится в пределах от 1,24 до 1,32 г/см³.

2.1.2 Точечная и интервальная оценка средней величины генеральной совокупности по данным малой выборки

Условие: имеются данные выборочного наблюдения о плотности дерново-подзолистой почвы (таблица 2.2)

Провести точечную и интервальную оценку средней генеральной совокупности.

Таблица 2.2

Плотность дерново-подзолистой почвы, г/см³

Номер образца	Плотность почвы, г/см ³	Квадрат значения признака	Номер образца	Плотность почвы, г/см ³	Квадрат значения признака
	x_i	x_i^2		x_i	x_i^2
1	1,07	1,1449	6	1,30	1,6900
2	1,11	1,2321	7	1,33	1,7689
3	1,18	1,3924	8	1,37	1,8769
4	1,23	1,5129	9	1,44	2,0736
5	1,20	1,4400	10	1,47	2,1609
x	x	x	Итого	12,70	16,2926

Решение:

1. Определим квадраты значений признака (x_i^2) и запишем их в таблицу 2.2

2. Подсчитаем суммы $\sum_{i=1}^n x_i$ и $\sum_{i=1}^n x_i^2$ и запишем их в итоговую строку таблицы 2.2.

3. Вычислим среднюю величину плотности древесно-подзолистой почвы в образцах по данным выборочной совокупности:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{12,70}{10} = 1,27.$$

4. Найдем выборочную дисперсию:

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = \overline{x^2} - \bar{x}^2 = \frac{16,2926}{10} - 1,27^2 = 0,01636.$$

5. Исчислим несмещенную оценку дисперсии:

$$S^2 = \sigma^2 \cdot \frac{n}{n-1} = 0,01636 \cdot \frac{10}{10-1} = 0,018178.$$

6. Исчислим среднюю ошибку выборочной средней:

$$m_{\bar{x}} = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0,018178}{10} \left(1 - \frac{10}{140}\right)} = \sqrt{0,001688} = 0,041085.$$

7. Проведем *точечную оценку* средней в генеральной совокупности:

$$\tilde{x} = \bar{x} = 1,27 \text{ г/см}^3 \text{ при } m_x = 0,04 \text{ г/см}^3.$$

8. По таблице «Значения двустороннего критерия -Стьюдента (Приложение 2) найдем теоретическое значение t при доверительном уровне вероятности $P = 0,95$ и числе степеней свободы вариации $\lambda = n - 1 = 9$: $t_{0,05} = 2,2622$.

9. Определим предельную случайную ошибку выборочной средней:

$$\varepsilon_{\bar{x}} = t_{0,05} \cdot m_{\bar{x}} = 2,2622 \cdot (\pm 0,041085) = \pm 0,092941 \approx 0,09 \text{ (г/см}^3\text{)}$$

10. Проведем *интервальную оценку* средней в генеральной совокупности:

$$\tilde{x} = \bar{x} \pm \varepsilon_x = 1,27 \pm 0,09 \text{ г/см}^3,$$

или

$$\bar{x} - \varepsilon_x \leq \tilde{x} \leq \bar{x} + \varepsilon_x, 1,18 \leq \tilde{x} \leq 1,36 \text{ г/см}^3.$$

11. Сделаем вывод. С доверительным уровнем вероятности 0,95 можно утверждать, что средняя в генеральной совокупности находится в пределах от 1,18 до 1,36 г/см³.

Задача 2.2 Критерий χ^2 -Пирсона как критерий согласия

2.2.1 Проверка статистической гипотезы о соответствии фактического выборочного распределения теоретическому соотношению

Условие: имеются данные фактического распределения образцов по содержанию гумуса в почве.

Проверить соответствие фактического распределения ожидаемому соотношению 1:2:4:5:4:3:1 при уровне значимости 0,05.

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

H_0 : фактическое распределение образцов по содержанию гумуса в почве соответствует ожидаемому соотношению 1:2:4:5:4:3:1;

H_1 : фактическое распределение образцов по содержанию гумуса в почве не соответствует ожидаемому соотношению 1:2:4:5:4:3:1.

2. Составим макет таблицы 2.3 и запишем исходные данные задачи в графы 1 и 2. В графе 3 укажем ожидаемое соотношение. Получим суммы по графам 2 и 3.

3. Фактическое значение критерия χ^2 -Пирсона определяется по следующей формуле:

$$\chi_{\text{факт}}^2 = \sum \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}, \quad (2.10)$$

где n_i – фактическая частота в i -ой группе;

$\tilde{n}_i$ – гипотетическая частота в i -ой группе.

Таблица 2.3

Фактическое и теоретическое распределение образцов по содержанию гумуса в почве

Интервал по содержанию гумуса, %	Фактическое число образцов n_i	Ожидаемое соотношение	Гипотетическое число образцов $\tilde{n}_i$	$n_i - \tilde{n}_i$	$(n_i - \tilde{n}_i)^2$	$\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$
1	2	3	4	5	6	7
до 2,6	7	1	7	0	0	0,000
2,6-3,0	19	2	14	5	25	1,786
3,0-3,4	24	4	28	-4	16	0,571
3,4-3,8	32	5	35	-3	9	0,257
3,8-4,2	27	4	28	-1	1	0,036
4,2-4,6	21	3	21	0	0	0,000
свыше 4,6	10	1	7	3	9	1,286
Итого	140	20	140	0	x	3,936

С этой целью рассчитаем, сколько единиц совокупности (образцов) составляет одну часть в ожидаемом (теоретическом) соотношении, разделив итог графы 2 на итог графы 3: $140/20 = 7$ образцов. Рассчитаем

теоретическое (гипотетическое) число образцов по группам, умножая 7 образцов на число частей в каждой группе. Результаты запишем в графу 4, округляя до целых результаты умножения, так как число образцов – дискретная величина. Правильность расчетов покажет равенство итогов граф 2 и 4.

В графе 5 находим разности фактических и гипотетических частот, в графе 6 – их квадраты, в графе 7 – отношение квадратов к теоретическим частотам графы 4.

Итог графы 7 представляет собой фактическое значение критерия χ^2 -Пирсона: $\chi^2_{\text{факт}} = 3,936$.

4. Теоретическое значение критерия χ^2 -Пирсона находят с учетом уровня значимости и числа степеней свободы:

$$v = l - k, \quad (2.11)$$

где l – число градаций значений признака (число групп),

k – число независимых линейных ограничивающих связей.

Линейная ограничивающая связь заключается в том, что общее число единиц в эмпирическом и теоретическом распределениях должно быть равным. В данном примере имеется одна независимая линейная ограничивающая связь. Поэтому, $v = 7 - 1 = 6$.

В приложении 5 при уровне значимости $\lambda=0,05$ и числу степеней свободы $v = 6$, находим теоретическое значение критерия χ^2 -Пирсона 12,59.

5. Сопоставляем фактическое значение критерия χ^2 -Пирсона 3,936 с теоретическим значением χ^2 -Пирсона 12,59, и приходим к выводу, что теоретическое значение меньше фактического. Выборочная точка находится в области допустимых значений. Нулевая гипотеза принимается.

6. Практически значимый вывод: фактическое распределение образцов по содержанию гумуса в почве согласуется с теоретическим с вероятностью ошибки в 5 случаях из 100.

2.2.2 Проверка статистической гипотезы о соответствии фактического распределения нормальному распределению

Условие: имеется эмпирический ряд распределения совокупности образцов по содержанию гумуса в почве с параметрами $\bar{x} = 3,64\%$, $S = 0,64\%$ (таблица 2.4)

Проверить соответствие фактического распределения образцов по содержанию гумуса в почве нормальному распределению. Уровень вероятности суждения 0,95.

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

H_0 : эмпирическое распределение соответствует нормальному распределению;

H_1 : эмпирическое распределение не соответствует нормальному распределению.

2. Исходя из предположения о соответствии эмпирического распределения нормальному, определим для каждого интервала ожидаемые (гипотетические) частоты.

С этой целью:

а) определим срединные значения интервалов, как полусумму нижней и верхней границ в каждой группе, и запишем их в таблицу 2.4 (графа 3).

б) найдем отклонения срединных значений интервалов (x_i) от средней величины $\bar{x} = 3,64\%$ (графа 4).

в) вычислим для каждого интервала нормированное отклонение (t_i) как отношение данных графы 4 к среднему квадратичному отклонению $s = 0,64\%$. Результаты запишем в графу 5.

г) значение функции плотности нормального распределения (графа 6) для нормированного отклонения каждого интервала можно определить, используя данные таблицы «Значение функции $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ » (Приложение 4), или же используя мастер функций в MS Excel (функция НОРМ.СТ.РАСП).

д) рассчитаем произведение числа единиц в совокупности на длину интервала h , выраженную в долях среднего квадратического отклонения:

$$C = N \frac{h}{\sigma}. \quad (2.12)$$

Величина C составит:

$$C = 140 \cdot \frac{0,4}{0,64} = 87,5 \approx 88.$$

Таблица 2.4

Эмпирическое и теоретическое распределения образцов по содержанию гумуса в почве

Интервал по содержанию гумуса, %	Число образцов n_i	Срединное значение интервала x_i	Отклонение от средней $(x_i - \bar{x})$	Нормированное отклонение $t_i = \frac{x_i - \bar{x}}{S}$	Значение функции плотности нормального распределения $f(t)$	Гипотетическая частота нормального распределения		$n_i - \tilde{n}_i$	$(n_i - \tilde{n}_i)^2$	$\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$
						расчетная $\tilde{n}_i$	с округлением $\tilde{n}_i$			
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>
до 2,6	7	2,40	-1,24	-1,9375	0,0611	5,3768	6	1	1	0,1667
2,6-3,0	19	2,80	-0,84	-1,3125	0,1686	14,8368	15	4	16	1,0667
3,0-3,4	24	3,20	-0,44	-0,6875	0,3150	27,7200	28	-4	16	0,5714
3,4-3,8	34	3,60	-0,04	-0,0625	0,3982	35,0416	36	-2	4	0,1111
3,8-4,2	25	4,00	0,36	0,5625	0,3406	29,9728	30	-5	25	0,8333
4,2-4,6	21	4,40	0,76	1,1875	0,1971	17,3448	18	3	9	0,5000
свыше 4,6	10	4,80	1,16	1,8125	0,0772	6,7936	7	3	9	1,2857
Итого	140	x	x	x	x	x	140	x	x	4,5349

е) вычислим гипотетические частоты по формуле:

$$\tilde{n}_i = C \cdot f(t), \quad (2.13)$$

и запишем в графу 7.

Например, для первого интервала $88 \cdot 0,0611 = 5,3768$, для второго интервала $88 \cdot 0,1686 = 14,8368$ и т.д.

ж) округлим гипотетические частоты до целых чисел. Результаты запишем в графу 8. Подсчитаем сумму гипотетических частот и проверим ее равенство фактическому числу единиц совокупности: $\sum \tilde{n}_i = \sum n_i = 140$.

Примечание. Если число единиц в интервале менее 5, то интервалы объединяют с соседними так, чтобы частоты в эмпирическом и гипотетическом рядах распределений были не менее 5 единиц.

3. Определим разность фактических и гипотетических частот, проверим равенство их сумм нулю $\sum (n_i - \tilde{n}_i) = 0$ (графа 9).

4. Вычислим для каждого интервала квадраты разностей частот (графа 10).

5. Определим в каждой группе отношение квадрата разности частот к соответствующей гипотетической частоте $\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$, получим из них сумму, соответствующую фактическому значению критерия:

$$\chi_{\text{факт}}^2 = \sum \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} = 4,5349.$$

6. Определим число степеней свободы вариации признака по формуле:

$$v = l - k - m, \quad (2.14)$$

где l – число интервалов,

k – число независимых линейных ограничивающих связей,

m – число параметров, используемых при определении гипотетических частот.

В рассматриваемом примере число интервалов с учетом укрупнения равно 7, имеется одна линейная ограничивающая связь (равенство сумм фактических и гипотетических частот), при исчислении гипотетических частот использованы два параметра ($\bar{x}$, S): $v = 7 - 1 - 2 = 4$.

7. По математической таблице «Критическое значение критерия χ^2 -Пирсона» (Приложение 5) найдем критическое значение χ^2 при $\alpha = 0,05$, $v = 4$: $\chi_{\text{крит}}^2 = 9,49$

8. Сопоставим фактическое и критическое значение χ^2 -Пирсона. Фактическое значение критерия 4,53 меньше его критического значения 9,49, следовательно, находится в области допустимых значений критерия. Следует принять нулевую гипотезу о соответствии фактического распределения нормальному закону распределения с вероятностью ошибки в 5 случаях из 100.

9. Практически значимый вывод: распределение образцов по содержанию гумуса в почве согласуется с законом нормального распределения с вероятностью ошибки 5 %.

2.2.3 Проверка статистической гипотезы о соответствии фактического выборочного распределения распределению Пуассона

Условие: имеется эмпирический ряд распределения тракторов по числу поломок (таблица 2.5).

Проверить нулевую гипотезу о соответствии эмпирического распределения распределению Пуассона. Уровень вероятности суждения 0,95.

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

H_0 : эмпирическое распределение соответствует распределению Пуассона;

H_1 : эмпирическое распределение не соответствует распределению Пуассона.

2. Исходя из предположения о соответствии эмпирического распределения теоретическому определим *гипотетические частоты*.

Гипотетические частоты для возможных значений признака определяем, используя функцию:

$$P_{(x=m)} = \frac{\lambda^m e^{-\lambda}}{m!}, \quad (2.15)$$

где λ – параметр распределения Пуассона (за его приближенную оценку принимаются значения $\bar{x}$ и S^2 , в рассматриваемом примере $\lambda \approx \bar{x} \approx S^2 \approx 0,60$);

e – постоянная величина, основание натурального логарифма;

m – возможные значения признака, соответствующие x_i .

С этой целью:

а) определим произведение значений признака на частоту встречаемости ($x_i \cdot n_i$) и сумму произведений (итог графы 3);

б) рассчитаем среднее значение признака совокупности по средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{1600}{2650} = 0,604.$$

в) рассчитаем дисперсию признака, используя итог графы 5 и квадрат средней арифметической;

$$S^2 = \frac{\sum_{i=1}^n x_i^2 n_i}{\sum_{i=1}^n n_i} - \bar{x}^2 = \frac{2540}{2650} - 0,604^2 = 0,594.$$

г) определим λ^m и запишем значения в графу 6: для значения x_1 : $\lambda^m = 0,60^0 = 1,000$; для x_2 : $\lambda^m = 0,60^1 = 0,600$ и т.д.

д) значение $e^{-\lambda}$ найдем по таблице значений e^{-x} (приложение 6) или, используя мастер функций в MS Excel (функция EXP): $e^{-\lambda} = e^{-0,60} = 0,549$.

е) определим произведение $\lambda^m e^{-\lambda}$ (графа 7).

ж) определим $m!$ (графа 8).

Таблица 2.5

Эмпирическое и теоретическое распределения тракторов по числу поломок

Число поломок	Число тракторов	Объем явления	Квадрат значения признака	Произведение квадрата значения на частоту	λ^m	$\lambda^m e^{-\lambda}$	$m!$	$P_{(x=m)} = \frac{\lambda^m e^{-\lambda}}{m!}$	Гипотетическая частота	Разность частот	Взвешенный квадрат разности частот
x_i	n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$					$\tilde{n}_i = p_i N$	$n_i - \tilde{n}_i$	$\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>
0	1450	0	0	0	1,0000	0,549	1	0,549	1455	-5	0,0172
1	860	860	1	860	0,6000	0,329	1	0,329	873	-13	0,1936
2	290	580	4	1160	0,3600	0,198	2	0,099	262	28	2,9924
3	40	120	9	360	0,2160	0,119	6	0,020	52	-12	2,7692
4	10	40	16	160	0,1296	0,071	24	0,003	8	2	0,5000
Итого	2650	1600	x	2540	x	x	x	1,000	2650	0	6,4724

з) определим вероятность существования каждого значения признака (графа 9), и проверим равенство $\sum P_i = 1,000$.

и) рассчитаем гипотетические частоты распределения Пуассона (графа 10) по следующей формуле:

$$\tilde{n}_i = P_i \cdot N, \quad (2.16)$$

где N – общая сумма частот (общее количество тракторов).

Найдем сумму теоретических частот, и проверим ее равенство сумме фактических частот $\sum \tilde{n}_i = \sum n_i = N = 2650$.

3. Определим разности *фактических* и *гипотетических* частот (графа 11), и проверим равенство их сумм нулю $\sum (n_i - \tilde{n}_i) = 0$.

4. Определим в каждой группе отношение квадрата разности частот к соответствующей гипотетической частоте $\frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$ (графа 12).

5. Итог графы 12 – это фактическое значение критерия χ^2 -Пирсона.

$$\chi_{\text{факт}}^2 = \sum \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} = 6,4724.$$

6. Определим число степеней свободы по формуле (2.14). В рассматриваемом примере число интервалов (градаций признака) равно 5, ограничивающих линейных связей – 1 ($\sum n_i = \sum \tilde{n}_i = N$), при исчислении гипотетических частот использован один параметр – λ : $\nu = 5 - 1 - 1 = 3$.

7. По таблице «Критическое значение критерия χ^2 -Пирсона» (Приложение 5), найдем критическое значение χ^2 при $\alpha = 0,05$, $\nu = 3$: $\chi_{\text{крит}}^2 = 7,81$.

8. Сопоставим фактическое и табличное значение χ^2 . Фактическое значение критерия 6,47 меньше его критического значения 7,81, следовательно, находится в области допустимых значений критерия. Следует принять нулевую гипотезу о соответствии фактического распределения распределению Пуассона с вероятностью ошибки 5 случаев из 100.

9. Практически значимый вывод: фактическое распределение тракторов по числу поломок соответствует распределению Пуассона с вероятностью ошибки 5%.

Задача 2.3 Проверка статистической гипотезы по критерию χ^2 -Пирсона о независимости двух эмпирических распределений

Условие: имеются эмпирические распределения выборочной совокупности колосьев по двум признакам: числу обработок ядохимикатами и степени поражения колосьев (таблица 2.6).

Проверить статистическую гипотезу о независимости эмпирических распределений в генеральной совокупности.

Таблица 2.6

**Эмпирические распределения колосьев по числу обработок
ядохимикатами и степени поражения (n_{ij})**

Обработка ядохимикатами	Степень поражения				Итого ($\sum_{j=1}^m n_j$)	Про- цент к итогу
	непора- женные	пораженные				
		слабо	средне	сильно		
Не обработаны	170	250	340	480	1240	17,5
Обработаны 1 раз	760	680	590	410	2440	34,4
Обработаны 2 раза	1100	1020	930	360	3410	48,1
Итого ($\sum_{i=1}^l n_i$)	2030	1950	1860	1250	7090	100,0
Процент к итогу	28.6	27,5	26.2	17,6	100.0	X

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

H_0 : эмпирические распределения независимы;

H_1 : эмпирические распределения зависимы.

2. Подсчитаем для обоих распределений сумму частот в каждом интервале и общее число единиц в совокупности: $\sum_{i=1}^l n_i$, $\sum_{j=1}^m n_j$, и $\sum_{i=1}^l \sum_{j=1}^m n_{ij} = n$ (итоговые столбец и строка таблицы 2.6).

3. Определим процентное отношение частот каждого интервала к общему числу единиц в совокупности (последняя графа и строка таблицы 2.6).

4. Исчислим и запишем в таблицы 2.7 гипотетические частоты каждого интервала обоих распределений. При этом, исходя из нулевой гипотезы о независимости распределений, предполагаем, что распределение колосьев по степени поражения в пределах каждого интервала по числу обработок ядохимикатами соответствует итоговым процентам по строке, а распределение колосьев по числу обработок в пределах каждого интервала по степени поражения соответствует итоговым процентам по столбцу.

Например, гипотетическое распределение колосьев по степени поражения в пределах интервала необработанных колосьев составляет: непораженных – $\frac{1240}{100 \%} \cdot 28,6\% = 355$; пораженных слабо – $\frac{1240}{100 \%} \cdot 27,5 \% = 341$; пораженных средне – $\frac{1240}{100 \%} \cdot 26,2 \% = 325$; пораженных сильно – $\frac{1240}{100 \%} \cdot 17,7 \% = 219$. Аналогичный результат можно получить, используя процентное отношение последнего столбца таблицы 2.6 ($\frac{2030}{100 \%} \cdot 17,5 \% = 355$ и т.д.).

5. Подсчитаем для обоих распределений сумму гипотетических частот в каждом интервале $\sum_{i=1}^l \tilde{n}_i$, $\sum_{j=1}^m \tilde{n}_j$ (итоговая графа и строка таблицы 2.7). Суммы гипотетических частот по интервалам должны быть равны суммам фактических частот. Соответственно, общее число единиц гипотетических и

эмпирических распределений должно быть равным ($\sum_{i=1}^l \sum_{j=1}^m \tilde{n}_{ij} = \sum_{i=1}^l \sum_{j=1}^m n_{ij} = n$).

Таблица 2.7

Гипотетические распределения колосьев ($\tilde{n}_{ij}$)

Обработка ядохимикатами	Степень поражения				Итого ($\sum_{j=1}^m \tilde{n}_j$)	Про- цент к итогу
	непора- женные	пораженные				
		слабо	средне	сильно		
Не обработаны	355	341	325	219	1240	17,5
Обработаны 1 раз	699	671	640	430	2440	34,4
Обработаны 2 раза	976	938	895	601	3410	48,1
Итого ($\sum_{i=1}^l \tilde{n}_i$)	2030	1950	1860	1250	7090	100,0
Процент к итогу	28.6	27,5	26.2	17,7	100.0	x

6. Найдем разности между фактическими и гипотетическими частотами и запишем их в таблицу 2.8. Поскольку суммы фактических и гипотетических частот по интервалам равны, суммы разностей должны равняться нулю.

7. Фактическое значение критерия χ^2 -Пирсона, применяемый как критерий независимости, определяется по следующей формуле:

$$\chi_{\text{факт}}^2 = \sum \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}, \quad (2.17)$$

где n_{ij} – фактическая частота i -ого интервала по первому признаку и j -го интервала по второму признаку;

$\tilde{n}_{ij}$ – гипотетическая частота по i -ой и j -ой группам.

Таблица 2.8

**Разности между фактическими и гипотетическими частотами
($n_{ij} - \tilde{n}_{ij}$)**

Обработка ядохимикатами	Степень поражения				Итого
	непора- женные	пораженные			
		слабо	средне	сильно	
Не обработаны	-185	-91	+15	+261	0
Обработаны 1 раз	+61	+9	-50	-20	0
Обработаны 2 раза	+124	+82	+35	-241	0
Итого	0	0	0	0	0

Рассчитаем фактическое значение критерия χ^2 -Пирсона:

$$\chi_{\text{факт}}^2 = \frac{(-185)^2}{355} + \frac{(-91)^2}{341} + \frac{15^2}{325} + \frac{261^2}{219} + \frac{61^2}{699} + \frac{9^2}{671} + \frac{(-50)^2}{640} + \frac{(-20)^2}{430} + \frac{124^2}{976} + \frac{82^2}{938} + \frac{35^2}{895} + \frac{(-241)^2}{601} = 563,65.$$

8. Найдем число степеней свободы вариации по формуле:

$$\gamma = l - k, \quad (2.18)$$

где l – общее число интервалов в распределениях;

k – число независимых ограничивающих линейных связей.

Общее число интервалов в распределении (l) равно произведению числа интервалов по первому признаку на число интервалов по второму признаку:

$$l = ab, \quad (2.19)$$

где a и b – число интервалов по каждому признаку.

Линейные ограничивающие связи заключаются в том, что все итоги в эмпирическом и гипотетическом распределении должны совпадать. Следовательно, общее число ограничивающих линейных связей равно сумме числа интервалов по первому признаку (a) и числа интервалов по второму признаку (b). Однако свободно изменять численность могут только $(a + b - 1)$ интервалов, так как численность последнего интервала должна быть такой, чтобы было справедливо равенство: $(\sum n_{ij} = \sum \tilde{n}_{ij} = n)$. Следовательно,

$$k = a + b - 1. \quad (2.20)$$

Тогда число степеней свободы:

$$\gamma = ab - (a + b - 1) = (a - 1) \cdot (b - 1). \quad (2.21)$$

Для рассматриваемого примера: $df = (3 - 1) \cdot (4 - 1) = 6$.

9. По таблице «Критическое значение критерия χ^2 -Пирсона» (Приложение 5), найдем критическое значение χ^2 при уровне значимости $\lambda = 0,05$ и числе степеней свободы $df = 6$: $\chi^2_{0,05} = 12,59$.

10. Фактическое значение ($\chi^2_{\text{факт.}} = 563,65$) значительно превышает его критическое значение ($\chi^2_{0,05} = 12,59$). Нулевую гипотезу о независимости эмпирических распределений следует отвергнуть и принять альтернативную гипотезу с вероятностью ошибки в 5 случаях из 100.

11. Практически значимый вывод: обработка ядохимикатами является существенным фактором снижения заболеваемости растений при вероятной ошибке суждения 5 %.

Задача 2.4 Проверка статистической гипотезы по критерию χ^2 -Пирсона об однородности двух эмпирических распределений

Условие: имеются два эмпирических распределения образцов по плотности почвы (таблица 2.9).

Требуется: проверить статистическую гипотезу об однородности двух распределений.

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

H_0 : распределения однородны по плотности почвы;

H_1 : распределения неоднородны по плотности почвы.

Таблица 2.9

Распределение образцов по плотности почвы

Интервалы по плотности почвы, г/см ³	Число образцов			$n_{1i} \cdot n_2$	$n_{2i} \cdot n_1$	$n_{1i} \cdot n_2 - n_{2i} \cdot n_1$	$(n_{1i} \cdot n_2 - n_{2i} \cdot n_1)^2$	$\frac{(n_{1i} \cdot n_2 - n_{2i} \cdot n_1)^2}{n_{1i} + n_{2i}}$
	Массив почвы 1	Массив почвы 2	всего					
	n_{1i}	n_{2i}	$n_{1i} + n_{2i}$					
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
1,08-1,15	4	5	9	160	150	10	100	11,11
1,15-1,22	5	9	14	200	270	-70	4900	350,00
1,22-1,29	7	10	17	280	300	-20	400	23,53
1,29-1,36	6	8	14	240	240	0	0	0,00
1,36-1,43	5	6	11	200	180	20	400	36,36
1,43-1,50	3	2	5	120	60	60	3600	720,00
Итого	$n_1 = 30$	$n_2 = 40$	70	1200	1200	0	x	1141,00

2. Фактическое значение критерия χ^2 -Пирсона, применяемого как критерий однородности, определяется по следующей формуле:

$$\chi_{\text{факт.}}^2 = \frac{1}{n_1 n_2} \sum_{i=1}^l \frac{(n_{1i} n_2 - n_{2i} n_1)^2}{n_{1i} + n_{2i}}, \quad (2.22)$$

где n_1 – численность единиц в первом распределении;

n_2 – численность единиц в втором распределении;

n_{1i} – численность единиц в i -ом интервале первого распределения;

n_{2i} – численность единиц в i -ом интервале второго распределения.

С этой целью:

Определим число единиц в каждой совокупности $\sum_{i=1}^l n_{1i} = n_1 = 30$; $\sum_{i=1}^l n_{2i} = n_2 = 40$ (итоговая строка графы 2 и 3).

Определим в каждом интервале сумму частот двух распределений $n_{1i} + n_{2i}$ и общее число единиц в обоих распределениях $\sum_{i=1}^l (n_{1i} + n_{2i}) = n_1 + n_2 = 70$ (графа 4).

Рассчитаем произведения частот первого распределения на число единиц во втором распределении ($n_{1i} \cdot n_2$) и сумму произведений (графа 5).

Исчислим произведения частот второго распределения на число единиц в первом распределении ($n_{2i} \cdot n_1$) и сумму произведений (графа 6). Суммы граф 5 и 6 должны быть равны: $\sum_{i=1}^l n_{1i} \cdot n_2 = \sum_{i=1}^l n_{2i} \cdot n_1 = 1200$.

Найдем по каждому интервалу разность произведений частот на число единиц в совокупностях ($n_{1i} \cdot n_2 - n_{2i} \cdot n_1$) и сумму разностей, она должна равняться 0 (графа 7).

Определим квадраты разностей произведений частот ($(n_{1i} \cdot n_2 - n_{2i} \cdot n_1)^2$) (графа 8).

Рассчитаем взвешенные квадраты разностей $\frac{(n_{1i} \cdot n_2 - n_{2i} \cdot n_1)^2}{n_{1i} + n_{2i}}$ и их сумму (графа 9).

Исчислим фактическое значение критерия χ^2 :

$$\chi_{\text{факт.}}^2 = \frac{1}{n_1 n_2} \sum_{i=1}^l \frac{(n_{1i} n_2 - n_{2i} n_1)^2}{n_{1i} + n_{2i}} = \frac{1}{30 \cdot 40} \cdot 1141,00 = 0,95.$$

3. Число степеней свободы вариации определяется по формуле (2.11). В данном случае имеется одна линейная ограничивающая связь: сумма частот всех интервалов должна быть равна общему числу единиц в обеих совокупностях. Определим число степеней свободы вариации: $\nu = 6 - 1 = 5$.

4. По математической таблице «Критическое значение критерия χ^2 -Пирсона» (Приложение 5), найдем критическое значение χ^2 при уровне значимости $\lambda = 0,05$ и $\nu = 5$: $\chi_{0,05}^2 = 11,07$.

5. Так как фактическое значение критерия χ^2 -Пирсона ($\chi_{\text{факт.}}^2 = 0,95$) не превышает критическое ($\chi_{0,05}^2 = 11,07$), то выдвинутая гипотеза об однородности двух эмпирических распределений принимается с вероятностью ошибки 5%.

6. Практически значимый вывод: изучаемые распределения образцов по плотности почвы однородны при вероятности ошибки суждения 5%.

Задача 2.5 Проверка статистических гипотез относительно двух средних величин по данным малых выборок

2.5.1 Выборки независимые с равными численностями ($n_1=n_2$), дисперсии генеральных совокупностей равны ($\sigma_1^2=\sigma_2^2$)

Условие. имеются выборочные данные о содержании органического вещества в разных массивах почвы (таблица 2.10).

Проверить статистическую гипотезу относительно средних генеральных совокупностей. Уровень значимости 0,05.

Решение:

1. Сформулируем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2$ содержание органического вещества в разных массивах почвы в генеральной совокупности одинаково;

$H_1: \tilde{x}_1 \neq \tilde{x}_2$ содержание органического вещества в разных массивах почвы в генеральной совокупности различно.

2. Рассчитаем средние величины по выборкам:

$$\bar{x}_1 = \frac{\sum_{i=1}^n x_{1i}}{n} = \frac{23,10}{10} = 2,31\%; \quad \bar{x}_2 = \frac{\sum_{i=1}^n x_{2i}}{n} = \frac{22,5}{10} = 2,25\%.$$

Таблица 2.10

Содержание органического вещества в разных массивах почвы

Массив почвы № 1				Массив почвы № 2			
но- мер про- бы	содержание органи- ческого вещества, %	откло- нение от среднего	квадрат откло- нения	номер пробы	содержание органи- ческого вещества, %	откло- нение от среднего	квадрат откло- нения
	x_{1i}	$x_{1i} - \bar{x}_1$	$(x_{1i} - \bar{x}_1)^2$		x_{2i}	$x_{2i} - \bar{x}_2$	$(x_{2i} - \bar{x}_2)^2$
1	2,00	-0,31	0,0961	1	2,01	-0,24	0,0576
2	2,66	0,35	0,1225	2	2,24	-0,01	0,0001
3	2,18	-0,13	0,0169	3	2,47	0,22	0,0484
4	2,13	-0,18	0,0324	4	2,00	-0,25	0,0625
5	2,43	0,12	0,0144	5	2,63	0,38	0,1444
6	2,72	0,41	0,1681	6	2,53	0,28	0,0784
7	2,48	0,17	0,0289	7	2,10	-0,15	0,0225
8	2,12	-0,19	0,0361	8	1,77	-0,48	0,2304
9	2,06	-0,25	0,0625	9	2,30	0,05	0,0025
10	2,32	0,01	0,0001	10	2,45	0,20	0,0400
Итого	23,10	x	0,5780	Итого	22,50	x	0,6868

3. Вычислим несмещенные оценки выборочных дисперсий:

$$S_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n - 1} = \frac{0,5780}{9} = 0,064; \quad S_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n - 1} = \frac{0,6868}{9} = 0,076.$$

4. Обобщенная средняя ошибка двух выборочных средних определяется следующим образом:

$$\bar{m}_{1-2} = \sqrt{m_1^2 + m_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}. \quad (2.23)$$

Определим обобщенную среднюю ошибку двух выборочных средних:

$$\bar{m}_{1-2} = \sqrt{m_1^2 + m_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{0,064 + 0,076}{10}} = \sqrt{0,014} = 0,118.$$

5. Рассчитаем фактическое значение критерия t по формуле (2.4):

$$t_{\text{факт.}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{m}_{1-2}} = \frac{|2,31 - 2,25|}{0,118} = \frac{0,06}{0,118} = 0,508.$$

6. Табличное значение критерия определим с учетом формирования независимой выборки. Численности выборок равны. Выясним, равны ли дисперсии в генеральных совокупностях.

$H_0: \sigma_1^2 = \sigma_2^2$ (дисперсии в генеральных совокупностях равны);

$H_1: \sigma_1^2 \neq \sigma_2^2$ (дисперсии в генеральных совокупностях не равны).

С этой целью проведем проверку гипотез по критерию F -распределения, рассчитав фактическое значение по формуле (2.5):

$$F_{\text{факт.}} = \frac{S_2^2}{S_1^2}, \text{ где } S_2^2 > S_1^2.$$

Рассчитаем $F_{\text{факт.}}$:

$$F_{\text{факт.}} = \frac{0,076}{0,064} = 1,1875.$$

По математико-статистической таблице «Таблица 5%-ного распределения F » (Приложение 3), найдем критическое значение F_λ при $\lambda = 0,05$ и числе степеней свободы для S_2^2 , равным $\nu_1 = n_2 - 1 = 10 - 1 = 9$, и для S_1^2 , равным $\nu_2 = n_1 - 1 = 10 - 1 = 9$: $F_{0,05} = 3,18$.

Фактическое значение F -критерия ($F_{\text{факт.}} = 1,1875$) меньше табличного ($F_{0,05} = 3,18$), следовательно, принимается нулевая гипотеза о равенстве генеральных дисперсий ($\sigma_1^2 = \sigma_2^2$).

7. Найдем по таблице критических значений двустороннего критерия - Стьюдента (Приложение 2) значение t_λ при установленном уровне значимости $\lambda = 0,05$.

Число степеней свободы определяется по формуле:

$$df = 2n - 2. \quad (2.24)$$

Определим число степеней свободы: $df = 2 \cdot 10 - 2 = 18$.

Табличное значение критерия t -Стьюдента составит $t_{0,05} = 2,1009$.

8. Поскольку $t_{\text{факт.}} < t_{0,05}$, нулевую гипотезу о равенстве средних в генеральных совокупностях принимаем, следовательно средние в генеральных совокупностях равны с вероятностью ошибки 5%.

9. Практически значимый вывод: содержание органического вещества в двух массивах почвы в генеральной совокупности одинаково. Вероятность ошибки суждения – 5%.

2.5.2 Выборки независимые с разной численностью ($n_1 \neq n_2$), дисперсии генеральных совокупностей равны ($\sigma_1^2 = \sigma_2^2$)

Условие: имеются выборочные данные по двум массивам почвы о содержании цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы (таблица 2.11.).

Проверить статистическую гипотезу относительно средних величин генеральных совокупностей. Уровень значимости 0,05.

Решение:

1. Сформулируем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2$ среднее содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы в двух массивах почвы в генеральной совокупности одинаково;

$H_1: \tilde{x}_1 \neq \tilde{x}_2$ среднее содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы в двух массивах почвы в генеральной совокупности различно.

2. Исчислим средние величины по выборкам:

$$\bar{x}_1 = \frac{\sum_{i=1}^n x_{1i}}{n_1} = \frac{720}{6} = 120; \bar{x}_2 = \frac{\sum_{i=1}^n x_{2i}}{n_2} = \frac{1925}{7} = 275.$$

Таблица 2.11

Содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы

Массив почвы № 1				Массив почвы № 2			
номер пробы	содержание микроорганизмов, тыс. кол/г	отклонение от среднего	квадрат отклонения	номер пробы	содержание микроорганизмов, тыс. кол/г	отклонение от среднего	квадрат отклонения
	x_{1i}	$x_{1i} - \bar{x}_1$	$(x_{1i} - \bar{x}_1)^2$		x_{2i}	$x_{2i} - \bar{x}_2$	$(x_{2i} - \bar{x}_2)^2$
1	99	-21	441	1	218	-57	3249
2	124	4	16	2	309	34	1156
3	128	8	64	3	321	46	2116
4	97	-23	529	4	305	30	900
5	111	-9	81	5	292	17	289
6	161	41	1681	6	216	-59	3481
-	-	-	-	7	264	-11	121
Итого	720	-	2812	Итого	1925	-	11312

3. Вычислим выборочные дисперсии:

$$S_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1 - 1} = \frac{2812}{6 - 1} = 562,4; S_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n_2 - 1} = \frac{11312}{7 - 1} = 1885,3.$$

4. Определим обобщенную среднюю ошибку двух выборочных средних:

$$\bar{m}_{1-2} = \sqrt{m_1^2 + m_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{562,4}{6} + \frac{1885,3}{7}} = \sqrt{363,0} = 19,053.$$

5. Рассчитаем фактическое значение критерия t :

$$t_{\text{факт.}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{m}_{1-2}} = \frac{|120 - 275|}{19,053} = \frac{155}{19,053} = 8,135.$$

6. Табличное значение критерия определим с учетом формирования независимой выборки. Численности выборок не равны. Выясним, равны ли дисперсии в генеральных совокупностях.

$H_0: \sigma_1^2 = \sigma_2^2$ (дисперсии в генеральных совокупностях равны);

$H_1: \sigma_1^2 \neq \sigma_2^2$ (дисперсии в генеральных совокупностях не равны).

С этой целью проведем проверку гипотез по критерию F-распределения:

$$F_{\text{факт.}} = \frac{S_2^2}{S_1^2}, \text{ где } S_2^2 > S_1^2.$$

Рассчитаем $F_{\text{факт.}}$:

$$F_{\text{факт.}} = \frac{1885,3}{562,4} = 3,35.$$

По математико-статистической таблице «Таблица 5%-ного распределения F» (Приложение 3), найдем критическое значение F_λ при $\lambda = 0,05$ и числе степеней свободы для S_2^2 , равным $\nu_1 = n_2 - 1 = 7 - 1 = 6$, и для S_1^2 , равным $\nu_2 = n_1 - 1 = 6 - 1 = 5$: $F_{0,05} = 4,95$.

Фактическое значение F-критерия ($F_{\text{факт.}} = 3,35$) меньше табличного ($F_{0,05} = 4,95$), следовательно, принимается нулевая гипотеза о равенстве генеральных дисперсий ($\sigma_1^2 = \sigma_2^2$).

7. Найдем по таблице критических значений двустороннего критерия - Стьюдента (Приложение 2) значение t при уровне значимости 0,05.

При этом число степеней свободы определяется по формуле:

$$df = n_1 + n_2 - 2. \quad (2.25)$$

Определим число степеней свободы: $df = n_1 + n_2 - 2 = 6 + 7 - 2 = 11$.

Табличное значение критерия t-Стьюдента составит $t_{0,05} = 2,2010$.

8. Поскольку $t_{\text{факт.}} > t_{0,05}$, нулевую гипотезу о равенстве средних в генеральных совокупностях отвергаем. Следовательно, принимаем альтернативную гипотезу о том, что средние в генеральных совокупностях не равны с вероятностью ошибки суждения 5 %.

9. Практически значимый вывод: изучаемые массивы почвы достоверно различаются по среднему содержанию цитратиспользующих микроорганизмов. Ошибка суждения составляет 5%.

2.5.3 Выборки независимые, равной численности ($n_1=n_2$), дисперсии в генеральных совокупностях не равны ($\sigma_1^2 \neq \sigma_2^2$)

Условие. представлены выборочные данные о содержании органического вещества в разных массивах почвы (таблица 2.12).

Проверить статистическую гипотезу относительно средних величин генеральных совокупностей. Уровень значимости 0,05.

Таблица 2.12

Содержание органического вещества в разных массивах почвы

Массив почвы № 1				Массив почвы № 2			
но- мер про- бы	содержание органи- ческого вещества, %	откло- нение от среднего	квадрат откло- нения	номер пробы	содержание органи- ческого вещества, %	откло- нение от среднего	квадрат откло- нения
	x_{1i}	$x_{1i} - \bar{x}_1$	$(x_{1i} - \bar{x}_1)^2$		x_{2i}	$x_{2i} - \bar{x}_2$	$(x_{2i} - \bar{x}_2)^2$
1	2,91	0,46	0,2116	1	2,43	-0,21	0,0441
2	1,96	-0,49	0,2401	2	2,81	0,17	0,0289
3	2,04	-0,41	0,1681	3	2,70	0,06	0,0036
4	2,88	0,43	0,1849	4	2,52	-0,12	0,0144
5	2,46	0,01	0,0001	5	2,74	0,10	0,0100
Итого	12,25	x	0,8048	Итого	13,20	x	0,1010

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2$ содержание органического вещества в разных массивах почвы в генеральной совокупности одинаково;

$H_1: \tilde{x}_1 \neq \tilde{x}_2$ содержание органического вещества в разных массивах почвы в генеральной совокупности различно.

2. Рассчитаем средние величины по выборкам:

$$\bar{x}_1 = \frac{\sum_{i=1}^n x_{1i}}{n} = \frac{12,25}{5} = 2,45\%; \quad \bar{x}_2 = \frac{\sum_{i=1}^n x_{2i}}{n} = \frac{13,20}{5} = 2,64\%.$$

3. Вычислим выборочные дисперсии:

$$S_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n-1} = \frac{0,8048}{5-1} = 0,20120; \quad S_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n-1} = \frac{0,1010}{5-1} = 0,02525.$$

4. Определим обобщенную среднюю ошибку двух выборочных средних:

$$\bar{m}_{1-2} = \sqrt{m_1^2 + m_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{0,2012 + 0,02525}{5}} = \sqrt{0,04529} = 0,213.$$

5. Рассчитаем фактическое значение критерия t :

$$t_{\text{факт.}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{m}_{1-2}} = \frac{|2,45 - 2,64|}{0,213} = \frac{0,19}{0,213} = 0,892.$$

6. Для характеристики независимых выборок с точки зрения их формирования выясним равенство дисперсий в генеральных совокупностях.

$H_0: \sigma_1^2 = \sigma_2^2$ (дисперсии в генеральных совокупностях равны);

$H_1: \sigma_1^2 \neq \sigma_2^2$ (дисперсии в генеральных совокупностях не равны).

С этой целью проведем проверку гипотез по критерию F-распределения:

$$F_{\text{факт.}} = \frac{S_1^2}{S_2^2}, \text{ где } S_1^2 > S_2^2.$$

Рассчитаем $F_{\text{факт.}}$:

$$F_{\text{факт.}} = \frac{0,20120}{0,02525} = 7,97.$$

По математико-статистической таблице «Таблица 5%-ного распределения F» (Приложение 3), найдем критическое значение F_λ при $\lambda = 0,05$ и числе степеней свободы для S_1^2 , равным $\nu_1 = n_1 - 1 = 5 - 1 = 4$, и для S_2^2 , равным $\nu_2 = n_2 - 1 = 5 - 1 = 4$: $F_{0,05} = 6,39$.

Фактическое значение F-критерия ($F_{\text{факт.}} = 7,97$) больше табличного ($F_{0,05} = 6,39$). Следовательно, нулевая гипотеза о равенстве генеральных дисперсий ($\sigma_1^2 = \sigma_2^2$) отвергается, и принимается альтернативная гипотеза о том, что генеральные дисперсии не равны ($\sigma_1^2 \neq \sigma_2^2$).

7. Определим число степеней свободы вариации с учетом того, что дисперсии в генеральных совокупностях не равны:

В случае, когда дисперсии генеральных совокупностей не равны, число степеней свободы вариации определяется с учетом поправки:

$$df = (n_1 + n_2 - 2) \cdot \left(0,5 + \frac{S_1^2 \cdot S_2^2}{S_1^4 + S_2^4} \right), \quad (2.26)$$

где n_1 и n_2 – численности выборочных совокупностей,

S_1^2 и S_2^2 – дисперсии выборочных совокупностей.

Определим число степеней свободы:

$$df = (5 + 5 - 2) \cdot \left(0,5 + \frac{0,2012 \cdot 0,02525}{0,2012^2 + 0,02525^2} \right) = 8 \cdot 0,624 \approx 5.$$

8. По таблице критических значений двустороннего критерия t-Стьюдента (Приложение 2) найдем значение t при $\lambda = 0,05$ и $df = 5$: $t_{0,05} = 2,5706$.

9. Фактическое значение критерия Стьюдента ($t_{\text{факт.}} = 0,892$) меньше его критического значения ($t_{0,05} = 2,5706$). Следовательно, фактическое значение t находится в области допустимых значений, и нулевая гипотеза о равенстве средних в генеральных совокупностях принимается с вероятностью ошибки суждения 5%.

10. Практически значимый вывод: содержание органического вещества в двух массивах почвы одинаково. Уровень ошибки суждения 5%.

2.5.4 Выборки независимые, разной численности ($n_1 \neq n_2$), дисперсии в генеральных совокупностях не равны ($\sigma_1^2 \neq \sigma_2^2$)

Условие. представлены данные о содержании цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой двух массивов почвы (таблица 2.13).

Проверить статистическую гипотезу относительно средних величин генеральных совокупностей. Уровень значимости 0,05.

Таблица 2.13

Содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы

Массив почвы № 1				Массив почвы № 2			
но- мер про- бы	содержание микро- организмов, тыс. кол/г	откло- нение от среднего	квадрат откло- нения	номер пробы	содержание микро- организмов, тыс. кол/г	откло- нение от среднего	квадрат откло- нения
	x_{1i}	$x_{1i} - \bar{x}_1$	$(x_{1i} - \bar{x}_1)^2$		x_{2i}	$x_{2i} - \bar{x}_2$	$(x_{2i} - \bar{x}_2)^2$
1	418	115	13225	1	137	-52	2704
2	427	124	15376	2	124	-65	4225
3	136	-167	27889	3	162	-27	729
4	338	35	1225	4	204	15	225
5	142	-161	25921	5	312	123	15129
6	186	-117	13689	6	195	6	36
7	228	-75	5625	-	-	-	-
8	549	246	60516	-	-	-	-
Итого	2424	x	163466	Итого	1134	x	23048

Решение

1. Выдвинем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2$ среднее содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы в двух массивах почвы в генеральной совокупности одинаково.

$H_1: \tilde{x}_1 \neq \tilde{x}_2$ среднее содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы в двух массивах почвы в генеральной совокупности различно.

2. Рассчитаем средние величины по выборкам:

$$\bar{x}_1 = \frac{\sum_{i=1}^n x_{1i}}{n_1} = \frac{2424}{8} = 303; \bar{x}_2 = \frac{\sum_{i=1}^n x_{2i}}{n_2} = \frac{1134}{6} = 189.$$

3. Вычислим выборочные дисперсии:

$$S_1^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1 - 1} = \frac{163466}{8 - 1} = 23352,3; S_2^2 = \frac{\sum (x_{2i} - \bar{x}_2)^2}{n_2 - 1} = \frac{23048}{6 - 1} = 4609,6.$$

4. Определим обобщенную среднюю ошибку двух выборочных средних:

$$\bar{m}_{1-2} = \sqrt{m_1^2 + m_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{23352,3}{8} + \frac{4609,6}{6}} = \sqrt{3687,3} = 60,7.$$

5. Рассчитаем фактическое значение критерия t :

$$t_{\text{факт.}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{m}_{1-2}} = \frac{|303 - 189|}{60,7} = \frac{114}{60,7} = 1,878.$$

6. Проведем проверку гипотезы о равенстве дисперсий в генеральных совокупностях с целью определения условий формирования выборок с разной численностью.

$H_0: \sigma_1^2 = \sigma_2^2$ (дисперсии в генеральных совокупностях равны);

$H_1: \sigma_1^2 \neq \sigma_2^2$ (дисперсии в генеральных совокупностях не равны).

С этой целью проведем проверку гипотез по критерию F -распределения:

$$F_{\text{факт.}} = \frac{S_1^2}{S_2^2}, \text{ где } S_1^2 > S_2^2.$$

Рассчитаем $F_{\text{факт.}}$:

$$F_{\text{факт.}} = \frac{23352,3}{4609,6} = 5,07.$$

По математико-статистической таблице «Таблица 5%-ного распределения F » (Приложение 3), найдем критическое значение F_λ при $\lambda = 0,05$ и числе степеней свободы для S_1^2 , равным $\nu_1 = n_1 - 1 = 8 - 1 = 7$, и для S_2^2 , равным $\nu_2 = n_2 - 1 = 6 - 1 = 5$: $F_{0,05} = 4,88$.

Фактическое значение F -критерия ($F_{\text{факт.}} = 5,07$) больше табличного ($F_{0,05} = 4,88$). Следовательно, нулевая гипотеза о равенстве генеральных дисперсий ($\sigma_1^2 = \sigma_2^2$) отвергается, и принимается альтернативная гипотеза о том, что генеральные дисперсии не равны ($\sigma_1^2 \neq \sigma_2^2$).

7. Табличное значение критерия -Стьюдента определяется как средняя взвешенная из табличных значений t -Стьюдента для отдельных выборок. В качестве весов выступают квадраты средних ошибок выборок:

$$t_\lambda = \frac{t_1 \cdot m_1^2 + t_2 \cdot m_2^2}{m_1^2 + m_2^2}, \quad (2.27)$$

где t_1 и t_2 определяют по таблице критических значений двустороннего критерия -Стьюдента (Приложение 2) при $\lambda = 0,05$ и $df_j = n_j - 1$, где j – номер выборки: t_1 (при $\lambda = 0,05$; $df_1 = 8 - 1 = 7$) = 2,3646; t_2 (при $\lambda = 0,05$; $df_2 = 6 - 1 = 5$) = 2,5706.

Определим квадраты средних ошибок выборок:

$$m_1^2 = \frac{S_1^2}{n_1} = \frac{23352,3}{8} = 2919,0357; \quad m_2^2 = \frac{S_2^2}{n_2} = \frac{4609,6}{6} = 768,2667.$$

Тогда табличное значение критерия -Стьюдента составит:

$$t_{0,05} = \frac{2,3646 \cdot 2919,0357 + 2,5706 \cdot 768,2667}{2919,0357 + 768,2667} = 2,4075$$

8. Фактическое значение критерия Стьюдента ($t_{\text{факт}} = 1,878$) меньше его критического значения ($t_{0,05} = 2,4075$). Следовательно, фактическое значение t находится в области допустимых значений, и нулевая гипотеза о равенстве средних в генеральных совокупностях принимается с вероятностью ошибки суждения 5%.

9. Практически значимый вывод: среднее содержание цитратиспользующих микроорганизмов в пахотном горизонте дерново-подзолистой почвы в двух массивах почвы в генеральной совокупности одинаково. Уровень ошибки суждения 5%.

Задача 2.6 Проверка статистических гипотез относительно средней разности при парной зависимости между единицами совокупности

Условие: имеются данные выборочного наблюдения содержания йода в черноземе обыкновенном кинетическим и потенциометрическими методами определения (таблица 2.14).

Проверить статистическую гипотезу относительно средней разности между попарно связанными единицами генеральной совокупности. Уровень значимости 0,05.

Таблица 2.14

Содержание йода в почве

но- мер про- бы	Содержание йода в почве, мг/кг		Разность содержания йода, мг/кг	Квадрат разности содержания
	Метод определения			
	кинетический	потенциометрический		
	x_{1i}	x_{2i}		
1	9,25	9,31	-0,06	0,0036
2	8,96	9,04	-0,08	0,0064
3	9,32	9,23	0,09	0,0081
4	9,68	9,76	-0,08	0,0064
5	9,52	9,08	0,44	0,1936
6	8,72	8,62	0,10	0,0100
7	8,63	8,55	0,08	0,0064
Итого	х	х	0,49	0,2345

Решение.

1. Сформулируем нулевую и альтернативную гипотезы:

$H_0: \bar{d} = 0$ два метода определения содержания йода в почве показывают одинаковый результат;

$H_1: \bar{d} \neq 0$ два метода определения содержания йода в почве показывают разный результат.

2. Определим среднюю разность по данным выборки:

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} = \frac{0,49}{7} = 0,07 \text{ мг/кг.}$$

3. Исчислим дисперсию средней разности:

$$S_d^2 = \frac{\sum d_i^2 - n\bar{d}^2}{n-1} = \frac{0,2345 - 7 \cdot 0,07^2}{7-1} = 0,0334.$$

4. Рассчитаем ошибку средней разности:

$$m_d = \sqrt{\frac{S_d^2}{n}} = \sqrt{\frac{0,0334}{7}} = \sqrt{0,004771} = 0,069.$$

5. Вычислим фактическое значение критерия t:

$$t_{\text{факт.}} = \frac{|\bar{d}|}{m_d} = \frac{0,07}{0,069} = 1,014$$

6. Найдем по таблице критических значений двустороннего критерия - Стьюдента (Приложение 2) значение t при уровне значимости 0,05 и числе степеней свободы вариации $df = n - 1 = 7 - 1 = 6$: $t_{0,05} = 2,4469$.

7. Фактическое значение $t_{\text{факт.}} = 1,014$ меньше его критического значения ($t_{0,05} = 2,4469$), поэтому нулевую гипотезу о равенстве средней разности между попарно связанными наблюдениями в генеральной совокупности нулю принимаем. Вероятность ошибки суждения 5%.

8. Практически значимый вывод: два метода определения содержания йода в почве показывают одинаковый результат. Ошибка суждения 5%.

Вопросы для проверки знаний по разделу II

1. Какая совокупность называется выборочной? генеральной?
2. Каково содержание выборочного метода?
3. Какие способы отбора единиц из генеральной совокупности в выборочную Вам известны?
4. Какие ошибки выборки могут иметь место при проведении выборочного наблюдения?
5. Какие ошибки нельзя исключить при применении выборочного метода?
6. Какие ошибки называются ошибками репрезентативности?
7. Каково содержание оценки?
8. В чем состоит свойство несмещенности оценки?
9. В чем суть точечной оценки параметра генеральной совокупности?
10. В чем суть интервальной оценки параметра генеральной совокупности?
11. Что следует понимать под «статистической гипотезой»?
12. Какова последовательность проведения проверки статистической гипотезы?
13. Какие гипотезы проверяются с использованием критерия χ^2 в качестве критерия согласия?
14. Какая нулевая гипотеза выдвигается при проверке гипотезы о соответствии фактического распределения ожидаемому?
15. От чего зависит табличное значение критерия χ^2 при использовании его в качестве критерия согласия?

16. В каком случае критерий χ^2 используется как критерий однородности?
17. Как формулируется нулевая гипотеза при применении критерия χ^2 как критерия независимости?
18. При каком аспекте использования критерия χ^2 при расчете его фактического значения не используются ожидаемые (теоретические численности)?
19. От каких величин зависит, главным образом, фактическое значение критерия χ^2 при использовании его в области применения как критерия независимости?
20. Как установить число степеней свободы при нахождении табличного значения критерия χ^2 как критерия однородности?
21. В чем разница между выборками зависимыми и независимыми?
22. Как формулируется нулевая гипотеза о средних при независимых выборках?
23. Какие четыре ситуации возможны при проверке гипотезы относительно двух средних при независимых выборках?
24. В чем особенность расчета числа степеней свободы при неравных дисперсиях?
25. Что такое НСР?
26. Как формулируется нулевая гипотеза при зависимых выборках?
27. Каков алгоритм расчета фактического значения критерия при зависимых выборках?
28. С какой целью при проверке гипотезы относительно 2-х средних используется критерий F-Фишера? Каковы особенности проверки гипотезы в случае $F_{\text{факт.}} > F_{\text{табл.}}$?
29. Найти табличное значение критерия, если по двум вариантам обработки почвы получены следующие значения урожайности проса:

1 вариант	12,7	12,9	14,7	15,5
2 вариант	14,4	14,0	15,9	16,8

Выборки зависимые.

30. При неизменных дисперсиях $S_1 = 21$, $S_2 = 22$ в каком случае большая вероятность принять нулевую гипотезу ($\bar{x}_1 = \bar{x}_2$) при больших или при малых выборках и почему?
31. Вы доказали, что справедлива альтернативная гипотеза $H_a: \bar{x}_1 \neq \bar{x}_2$. Какова вероятность ошибочности такого вывода? $P = 0,99$.
32. По двум вариантам эксперимента получены средние $\bar{x}_1 = 10$, $\bar{x}_2 = 20$. Есть ли смысл в статистической оценке различий между средними. Если есть, то почему?
33. Найти разницу между выборочными средними, если $t_{\text{факт.}} = 5$, $m_{1-2} = 8$.

РАЗДЕЛ III

СТАТИСТИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ СВЯЗЕЙ

Методические указания

Дисперсионный анализ применяется для проверки статистических гипотез относительно средних величин в нескольких генеральных совокупностях.

При проверке гипотез методом дисперсионного анализа используется критерий F -распределения, который представляет собой отношение дисперсий двух выборок, сделанных из одной нормально распределенной генеральной совокупности:

$$F = \frac{S_1^2}{S_2^2}, \quad (3.1)$$

где $S_1^2 > S_2^2$.

Нулевая гипотеза основана на предположении, что все выборки взяты из нормально распределенных генеральных совокупностей с равными средними величинами ($H_0: \tilde{x}_1 = \tilde{x}_2 = \dots = \tilde{x}_m$). Критическая область критерия определяется как $F_{\text{факт}} \geq F_\alpha$. Это означает, что если $F_{\text{факт}} \geq F_\alpha$, от нулевой гипотезы о равенстве средних величин в генеральных совокупностях следует отказаться. Если же $F_{\text{факт}} \leq F_\alpha$, то нулевая гипотеза должна быть принята.

Дисперсионный анализ следует проводить по этапам:

- 1) сформулировать статистические гипотезы (нулевую и альтернативную),
- 2) рассчитать фактическое значение критерия $F_{\text{факт}}$, для чего:
 - а) определить общий объем вариации изучаемого признака ($W_{\text{общ}}$) и разложить его на составляющие части – вариацию между выборками (межгрупповую – $W_{\text{межгр}}$) и вариацию внутри выборок (остаточную – $W_{\text{ост}}$);
 - б) определить число степеней свободы для всех объемов вариации (df),
 - в) рассчитать дисперсии (S^2) по каждому источнику вариации, как отношение объемов вариации (W) к соответствующему числу степеней свободы (df);
 - г) рассчитать фактическое значение критерия Фишера по формуле:

$$F = \frac{S_{\text{межгр}}^2}{S_{\text{ост}}^2}, \quad (3.2)$$

где $S_{\text{межгр}}^2 > S_{\text{ост}}^2$.

- 3) определить по таблицам теоретическое значение критерия $F_{\text{теор}}$.
- 4) сопоставить фактическое значение критерия ($F_{\text{факт}}$) с теоретическим ($F_{\text{теор}}$) и сделать вывод о принятии или отвержении нулевой гипотезы.

В ходе дисперсионного анализа необходимо учитывать особенности формирования выборок. Выборки могут быть сформированы по одному, двум и большему числу признаков (однофакторный, двухфакторный дисперсионный анализ и т.д.). Различают выборки со случайным формированием единиц совокупности (независимые выборки) и с неслучайным формированием (зависимые выборки). Независимые выборки

могут быть сформированы как с равной, так и неравной численностью. Формирование зависимых выборок предопределяет их равную численность.

Следует различать также выборки, сформированные по качественному признаку (сорт, вид удобрений, способ обработки и т.д.) и выборки, сформированные с учетом градаций количественных признаков (дозы удобрений, затраты кормов на одну голову скота и т.д.)

Если выборки сформированы по качественному признаку, а также по количественному признаку, градации которого представляют все интересующие исследователя уровни, на основе результатов дисперсионного анализа могут быть проверены статистические гипотезы относительно достоверности различий средних величин всех попарных сочетаний выборок.

При этом, если выборки равночисленны ($n_1=n_2=\dots=n_m$), для попарного сравнения применяется критерий Q-Тьюки. В случае неравночисленных выборок необходимо применять метод контрастов с использованием критерия Ψ -Шеффе.

Корреляционно-регрессионный анализ применяется для определения количественных характеристик корреляционной связи, проявляющейся в среднем по достаточно большому числу наблюдений. С этой целью, прежде всего, связь между переменными должна быть выражена посредством математического уравнения соответствующего вида, называемого уравнением корреляционной связи или уравнением регрессии. Например, линейная связь между двумя переменными (парная связь) выражается уравнением:

$$\tilde{y} = a + bx, \quad (3.3)$$

между несколькими переменными (множественная связь):

$$\tilde{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m. \quad (3.4)$$

В уравнениях регрессии y – зависимая переменная, $x_1, x_2, \dots, x_m$ – независимые переменные, $a, b, b_1, b_2, \dots, b_m$ – параметры уравнения. Параметр a – условное начало. В уравнении парной линейной регрессии параметр b называется коэффициентом полной регрессии; он показывает, на сколько единиц в среднем изменится значение зависимой переменной при изменении независимой на единицу. В уравнениях множественной линейной регрессии параметры $b_1, b_2, \dots, b_m$ называются коэффициентами чистой регрессии; каждый из них показывает, на сколько единиц в среднем изменится значение зависимой переменной при изменении соответствующей независимой переменной на единицу при условии, что другие независимые переменные, включенные в уравнение, не изменяются.

Следует помнить, что вопрос об установлении вида уравнения регрессии является одним из наиболее сложных в корреляционно-регрессионном анализе. При парных связях он может быть решен посредством графиков путем нанесения фактических значений зависимой переменной при соответствующих значениях независимой переменной на корреляционное поле. При достаточно большом числе единиц наблюдения определенное представление о виде уравнения дает построение рядов

распределения по двум признакам, или так называемых корреляционных таблиц.

После установления вида уравнения необходимо исчислить показатели корреляционной связи. В частности, следует решить уравнение регрессии, то есть найти значения его параметров. При этом параметры уравнения могут быть определены методом наименьших квадратов, в соответствии с которым сумма квадратов отклонений фактических значений зависимой переменной от значений, определенных по уравнению, должна быть минимальной:

$$\sum (y - \tilde{y})^2 = \min. \quad (3.5)$$

Поэтому для определения параметров решают систему нормальных уравнений. Число нормальных уравнений в системе равно числу параметров уравнения регрессии. Так, для парной линейной связи система нормальных уравнений имеет вид:

$$\begin{cases} \sum y = na + b \sum x \\ \sum yx = a \sum x + b \sum x^2 \end{cases} \quad (3.6)$$

После определения параметров уравнения должна быть дана количественная характеристика тесноты связи между переменными, включенными в уравнение. При парной линейной зависимости тесноту связи характеризуют коэффициентом детерминации (r^2) и коэффициентом корреляции (r), при множественной – соответственно, коэффициентом множественной детерминации (R^2) и коэффициентом множественной корреляции (R).

Коэффициент детерминации может быть исчислен по следующей основной формуле:

$$r^2 = \frac{\sigma_{\text{воспр}}^2}{\sigma_{\text{общ}}^2}, \quad (3.7)$$

где $\sigma_{\text{воспр}}^2$ – дисперсия воспроизведенная уравнением регрессии, характеризующая колеблемость зависимой переменной под влиянием независимых переменных, включенных в уравнение регрессии;

$\sigma_{\text{общ}}^2$ – общая дисперсия зависимой переменной.

Следует помнить, что коэффициент детерминации показывает, какая доля общей вариации зависимой переменной обусловлена влиянием изучаемых независимых переменных; его величина заключена в пределах от 0 до 1.

Соответственно и коэффициент корреляции, представляющий собой корень квадратный из коэффициента детерминации:

$$r = \sqrt{\frac{\sigma_{\text{воспр}}^2}{\sigma_{\text{общ}}^2}}, \quad (3.8)$$

изменяется от 0 до 1. Знаки «+» и «-» при коэффициенте полной регрессии означает направление связи: «+» – связь прямая, «-» – связь обратная.

Равенство коэффициентов нулю означает, что связь между переменными отсутствует, равенство коэффициентов единице означает, что между переменными существует функциональная зависимость. Анализ градаций коэффициента корреляции по шкале Чеддока позволяет дать характеристику тесноты связи:

Таблица 3.1

Шкала Чеддока					
r	до 0,30	0,31-0,50	0,51-0,70	0,71-0,90	0,91-0,99
сила связи	практически отсутствует	слабая	средняя	сильная (тесная)	очень сильная

В практических расчетах для определения коэффициентов детерминации и корреляции используют ряд рабочих формул.

Когда корреляционный анализ проводится по данным выборочного наблюдения, возникает необходимость в проверке статистических гипотез относительно показателей корреляционной связи и в статистической оценке показателей связи в генеральных совокупностях.

Типовые задачи

Задача 3.1 Однофакторный дисперсионный анализ

3.1.1 Однофакторный дисперсионный анализ при случайном формировании выборок равной численности

Условие: имеются результаты полевых опытов зависимости урожайности картофеля (ц/га) от видов удобрений (таблица 3.2). Делянки формировались в случайном порядке.

Проверить статистическую гипотезу о равенстве средних уровней урожайности по удобрениям при уровне значимости 0,05.

Таблица 3.2

Урожайность картофеля при разном внесении удобрений (по результатам полевых опытов)

Вид удобрения	Повторности (делянки)					Итого $\sum_{j=1}^{n_i} x_j$	Средняя урожайность картофеля $\bar{x}_i$
	1	2	3	4	5		
Без удобрений	220	248	237	250	228	1183	236,6
Навоз	327	330	354	343	370	1724	344,8
Навоз+РК	430	427	440	436	418	2151	430,2
Итого $\sum_{i=1}^m x_i$	977	1005	1031	1029	1016	5058	337,2

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2 = \tilde{x}_3$ (средние урожайности картофеля по удобрениям в генеральной совокупности равны);

$H_1: \tilde{x}_1 \neq \tilde{x}_2 \neq \tilde{x}_3$ (средние урожайности картофеля по удобрениям в генеральной совокупности не равны).

2. Определим модель дисперсионного анализа в соответствии с источниками вариации урожайности и условиями формирования выборок. В задаче два источника вариации урожайности, так как делянки формировались в случайном порядке:

1) удобрение (признак, по которому проведена группировка делянок),

2) остальные (случайные) факторы: индивидуальные особенности растений, небольшие различия в температуре воздуха и почвы, влажности почвы и т.п.

Таким образом, имеем однофакторную модель со случайным распределением равночисленных единиц в группах.

Общий объем вариации раскладывается на объем вариации фактора (видов удобрений) и остаточный объем вариации:

$$W_{\text{общ}} = W_{\text{фактора}} + W_{\text{ост}}. \quad (3.9)$$

3. Введем условные обозначения: i – номер строки, j – номер столбца, ij – номер единицы совокупности i -ой строки и j -го столбца, N – общая численность совокупности, n – число повторностей, m – число вариантов опыта, x_{ij} – индивидуальные значения признака (урожайности), $\sum_{i=1}^m x_{ij}$ – сумма i -ых значений признака по j -му столбцу, $\sum_{j=1}^n x_{ij}$ – сумма j -ых значений признака по i -ой строке, $\sum_{i=1}^m \sum_{j=1}^n x_{ij}$ – сумма значений признака во всей совокупности. При этом верно равенство $\sum_{j=1}^n \sum_{i=1}^m x_{ij} = \sum_{i=1}^m \sum_{j=1}^n x_{ij}$.

4. Рассчитаем показатели таблицы 3.3. Для этого все индивидуальные значения признака из таблицы 3.2 (x_{ij}), а также суммы значений по строкам ($\sum_{j=1}^n x_{ij}$) и по столбцам ($\sum_{i=1}^m x_{ij}^2$) возведем в квадрат.

Таблица 3.3

Квадраты значений признаков и их сумм

Вид удобрения	Повторности					Сумма квадратов $\sum_{j=1}^n x_{ij}^2$	Квадрат суммы $(\sum_{j=1}^n x_{ij})^2$
	1	2	3	4	5		
Без удобрений	48400	61504	56169	62500	51984	280557	1399489
Навоз	106929	108900	125316	117649	136900	595694	2972176
Навоз+РК	184900	182329	193600	190096	174724	925649	4626801

Вид удобрения	Повторности					Сумма квадратов $\sum_{j=1}^{n_i} x_{ij}^2$	Квадрат суммы $(\sum_{j=1}^{n_i} x_{ij})^2$
	1	2	3	4	5		
Сумма квадратов $\sum_{i=1}^m x_{ij}^2$	340229	352733	375085	370245	363608	$\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}^2$ =1801900	$\sum_{i=1}^m (\sum_{j=1}^{n_i} x_{ij})^2$ =8998466

Проверка правильности расчетов: сумма квадратов значений признака по строкам должна быть равна сумме квадратов значений по столбцам ($\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}^2 = \sum_{j=1}^n \sum_{i=1}^m x_{ij}^2$). Эта сумма будет составлять сумму квадратов всех значений признака в совокупности.

5. Расчет объемов вариации по источникам варьирования проведем по рабочим формулам, рассмотренным в задаче 1.10.

Общий объем вариации определяется по формуле:

$$W_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}^2 - \frac{(\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij})^2}{N}. \quad (3.10)$$

Общий объем вариации урожайности картофеля составит:

$$W_{\text{общ}} = 1801900 - \frac{5058^2}{15} = 1801900 - 1705558 = 96342,4.$$

Объем вариации по вариантам опыта определим по формуле:

$$W_{\text{фактора}} = \frac{\sum_{i=1}^m (\sum_{j=1}^{n_i} x_{ij})^2}{n_i} - \frac{(\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij})^2}{N}. \quad (3.11)$$

Объем вариации по вариантам опыта составит:

$$W_{\text{фактора}} = \frac{8998466}{5} - \frac{5058^2}{15} = 1799693 - 1705558 = 94135,6$$

Согласно правилу разложения объемов вариации объем вариации, обусловленный остальными (случайными) факторами, равен разности общего объема вариации и вариации фактора:

$$W_{\text{ост}} = W_{\text{общ}} - W_{\text{фактора}}. \quad (3.12)$$

Остаточный объем вариации составит:

$$W_{\text{ост}} = 96342,4 - 94135,6 = 2206,8.$$

6. Рассчитаем число степеней свободы для каждого источника варьирования. В нашей задаче общее число наблюдений $N=15$, равное произведению числа групп $m=3$ и числа повторностей $n=5$.

Для общего объема вариации число степеней свободы определяется по формуле:

$$v_0 = N - 1. \quad (3.13)$$

Для общего объема вариации число степеней свободы равно:

$$v_0 = 15 - 1 = 14.$$

Число степеней свободы для факторной дисперсии определяется по формуле:

$$v_{\text{фактора}} = m - 1. \quad (3.14)$$

Число степеней свободы для факторной дисперсии составит:

$$v_{\text{фактора}} = 3 - 1 = 2.$$

Число степеней свободы для остаточной дисперсии:

$$v_{\text{ост}} = v_0 - v_{\text{фактора}}. \quad (3.15)$$

Число степеней свободы для остаточной дисперсии составит:

$$v_{\text{ост}} = v_0 - v_{\text{фактора}} = 14 - 2 = 12.$$

7. Рассчитаем необходимые для дальнейшего анализа дисперсии признака по следующим формулам:

– дисперсия по фактору:

$$S_{\text{фактора}}^2 = \frac{W_{\text{фактора}}}{v_{\text{фактора}}}; \quad (3.16)$$

– дисперсия остаточная:

$$S_{\text{ост}}^2 = \frac{W_{\text{ост}}}{v_{\text{ост}}}. \quad (3.17)$$

Факторная дисперсия составит:

$$S_{\text{фактора}}^2 = \frac{94135,6}{2} = 47067,8.$$

Остаточная дисперсия составит:

$$S_{\text{ост}}^2 = \frac{2206,8}{12} = 183,9.$$

Примечание: общую дисперсию не вычисляем, так как она не участвует в дальнейшем анализе.

8. Определим фактическое значение F-критерия по фактору по формуле (3.2):

$$F_{\text{факт}} = \frac{S_{\text{фактора}}^2}{S_{\text{ост}}^2} = \frac{47067,8}{183,9} = 255,9.$$

9. Определим теоретическое значение критерия F–распределения при заданном уровне значимости 0,05 (Приложение 3). Для этого в таблице приложения 3 находим по горизонтали в первой строке число степеней свободы для большего среднего квадрата (факторной дисперсии) – $v_1 = 2$, и опускаем перпендикуляр до строки, указывающей число степеней свободы для меньшего среднего квадрата (остаточной дисперсии) – $v_2 = 12$. На пересечении находим значение 3,88. Таким образом, $F_{\text{теор}}(0,05; 2; 12) = 3,88$.

10. Результаты решения запишем в таблицу 3.4.

Таблица 3.4

Расчет и анализ дисперсий

Источник вариации	Объем вариации W	Число степеней свободы ν	Дисперсия S^2	Значение F-распределения	
				фактическое	табличное
Вид удобрения	94135,6	2	47067,8	255,9	3,88
Случайные факторы	2206,8	12	183,9	x	x
Итого	96342,4	14	x	x	x

11. Сделаем вывод. Сопоставив фактическое и табличное значение критерия F. Фактическое значение критерия превышает табличное. Так как фактическое значение попадает в критическую область критерия, нулевая гипотеза о равенстве выборочных средних величин в генеральных совокупностях должна быть отвергнута. Принимаем альтернативную гипотезу. Практически значимый вывод: с вероятностью 0,95 изучаемые виды удобрений существенно различаются по урожайности картофеля.

3.1.2 Однофакторный дисперсионный анализ при неслучайном формировании выборок равной численности

Условие: проведен полевой опыт, в котором изучалось влияние способа внесения гербицида (аценита) на урожайность пшеницы. Способы: I – под культивацию, II – ленточный, III – под боронование (таблица 3.5). При подборе делянок (повторностей) учитывалось плодородие почв.

Установить, достоверны ли различия в урожайности культуры по вариантам опыта. Уровень вероятности суждения 0,95.

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

H_0 : $\tilde{x}_1 = \tilde{x}_2 = \tilde{x}_3$ (между средними величинами в генеральных совокупностях нет достоверных различий, следовательно, способ внесения гербицида не оказывает влияние на уровень урожайности пшеницы);

H_1 : $\tilde{x}_1 \neq \tilde{x}_2 \neq \tilde{x}_3$ (между средними величинами в генеральных совокупностях есть достоверные различия, обусловленные влиянием способа внесения аценита).

Таблица 3.5

Урожайность пшеницы при разных способах внесения аценита, ц/га

Способ внесения аценита	Повторности					Итого $\sum_{j=1}^n x_j$	Средняя $\bar{x}_i$
	1	2	3	4	5		
I	30	28	26	27	33	144	28,8
II	38	41	42	39	40	200	40,0
III	35	38	39	38	42	192	38,4

Способ внесения аценита	Повторности					Итого $\sum_{j=1}^n x_j$	Средняя $\bar{x}_i$
	1	2	3	4	5		
Итого $\sum_{i=1}^m x_i$	103	107	107	104	115	$\sum_{i=1}^m \sum_{j=1}^n x_{ij}$ =536	$\bar{x}_0$ =35,7

2. Определим модель дисперсионного анализа.

Группировочный признак один (способ внесения аценита), следовательно, модель однофакторная. Так как в качестве повторностей рассматривались делянки, имеющие общность условий – одинаковое плодородие почв, то повторности следует признать зависимыми выборками. Таким образом, в нашей задаче три источника вариации:

1) способ внесения аценита (признак, по которому проведена группировка делянок);

2) плодородие почв по повторностям;

3) остальные (случайные) факторы.

Общий объем вариации равен сумме объемов вариации по источникам варьирования:

$$W_{\text{общ}} = W_{\text{фактора}} + W_{\text{повт}} + W_{\text{ост.}} \quad (3.18)$$

3. Условные обозначения примем аналогичными для задачи 3.1.1. Составим таблицу квадратов (таблица 3.6), и сделаем в ней вычисления. Для этого все значения, а также суммы значений по строкам и столбцам из таблицы 3.5 возведем в квадрат.

Таблица 3.6

Квадраты значений признаков и их сумм

Способ внесения аценита	Повторности					Сумма квадратов $\sum_{j=1}^n x_{ij}^2$	Квадрат суммы $(\sum_{j=1}^n x_{ij})^2$
	1	2	3	4	5		
I	900	784	676	729	1089	4178	20736
II	1444	1681	1764	1521	1600	8010	40000
III	1225	1444	1521	1444	1764	7398	36864
Сумма квadrato в $\sum_{i=1}^m x_{ij}^2$	3569	3909	3961	3694	4453	$\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2$ =19586	$\sum_{i=1}^m (\sum_{j=1}^n x_{ij})^2$ =97600
Квадрат суммы $(\sum_{i=1}^m x_{ij})^2$	10609	11449	11449	10816	13225	$\sum_{j=1}^n (\sum_{i=1}^m x_{ij})^2$ =57548	$(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2$ =287296

Сумма квадратов значений признака по строкам $(\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2)$ должна быть равна сумме квадратов значений по столбцам $(\sum_{j=1}^n \sum_{i=1}^m x_{ij}^2)$. Эта сумма будет составлять сумму квадратов всех значений признака в совокупности.

4. Расчет объемов вариации по источникам варьирования проведем по рабочим формулам, рассмотренным в задаче 1.10.

Общий объем вариации урожайности пшеницы равен:

$$W_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} = 19586 - \frac{287296}{15} = 19586 - 19153,0667 = 432,9333.$$

Объем вариации по вариантам опыта определим по формуле:

$$W_{\text{фактора}} = \frac{\sum_{i=1}^m (\sum_{j=1}^n x_{ij})^2}{n_i} - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} = \frac{97600}{5} - \frac{287296}{15} = 19520 - 19153,0667 = 366,9333.$$

Объем вариации по повторностям определяется по формуле:

$$W_{\text{повт}} = \frac{\sum_{j=1}^n (\sum_{i=1}^m x_{ij}^2)}{m} - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N}. \quad (3.19)$$

Объем вариации по повторностям равен:

$$W_{\text{повт}} = \frac{57548}{3} - \frac{287296}{15} = 19182,6667 - 19153,0667 = 29,6$$

Объем вариации, обусловленный остальными (случайными) факторами равен согласно правилу разложения объемов вариации разности общего объема вариации и объемов вариации фактора и повторностей:

$$W_{\text{ост}} = W_{\text{общ}} - W_{\text{фактора}} - W_{\text{повт}}. \quad (3.20)$$

Остаточный объем вариации составит:

$$W_{\text{ост}} = 432,9333 - 366,9333 - 29,6 = 36,4.$$

5. Рассчитаем число степеней свободы для каждого источника варьирования.

В нашей задаче общее число наблюдений $N=15$, число групп $m=3$, число повторностей $n=5$.

Для общего объема вариации число степеней свободы равно:

$$v_0 = N - 1 = 15 - 1 = 14.$$

Число степеней свободы для факторной дисперсии:

$$v_{\text{фактора}} = m - 1 = 3 - 1 = 2.$$

Число степеней свободы для дисперсии повторностей определяется следующим образом:

$$v_{\text{повт}} = n - 1. \quad (3.21)$$

Число степеней свободы для дисперсии повторностей равно:

$$v_{\text{повт}} = n - 1 = 5 - 1 = 4.$$

Тогда число степеней свободы для остаточной дисперсии будет определено следующим образом:

$$v_{\text{ост}} = v_0 - v_{\text{фактора}} - v_{\text{повт}}. \quad (3.22)$$

Число степеней свободы для остаточной дисперсии составит:

$$v_{\text{ост}} = 14 - 2 - 4 = 8.$$

6. Рассчитаем необходимые для дальнейшего анализа дисперсии признака:

– дисперсия по фактору по формуле (3.16):

$$S_{\text{фактора}}^2 = \frac{W_{\text{фактора}}}{v_{\text{фактора}}} = \frac{366,9333}{2} = 183,4667 \approx 183,47;$$

– дисперсия по повторностям будет определена следующим образом:

$$S_{\text{повт}}^2 = \frac{W_{\text{повт}}}{v_{\text{повт}}}; \quad (3.23)$$

дисперсия по повторностям составит:

$$S_{\text{повт}}^2 = \frac{W_{\text{повт}}}{v_{\text{повт}}} = \frac{29,6}{4} = 7,4;$$

– дисперсия остаточная по формуле (3.17):

$$S_{\text{ост}}^2 = \frac{W_{\text{ост}}}{v_{\text{ост}}} = \frac{36,4}{8} = 4,55.$$

Примечание: Общую дисперсию не вычисляем, так как она не участвует в дальнейшем анализе.

7. Определим фактическое значение F-критерия для каждого источника вариации:

– по фактору по формуле (3.2):

$$F_{\text{факт}} = \frac{S_{\text{фактора}}^2}{S_{\text{ост}}^2} = \frac{183,47}{4,55} = 40,32;$$

– по повторностям определяется по формуле:

$$F_{\text{повт}} = \frac{S_{\text{повт}}^2}{S_{\text{ост}}^2}. \quad (3.24)$$

Фактическое значение F-критерия по повторностям составит:

$$F_{\text{повт}} = \frac{S_{\text{повт}}^2}{S_{\text{ост}}^2} = \frac{7,4}{4,55} = 1,63.$$

8. Определим теоретические значения критерия F–распределения для каждого источника вариации при заданном уровне значимости 0,05 (Приложение 3). Для первого источника вариации в приложении 3 находим по горизонтали в первой строке число степеней свободы для большего среднего квадрата – факторной дисперсии ($v_1 = 2$), и опускаем перпендикуляр до строки, указывающей число степеней свободы для меньшего среднего квадрата – остаточной дисперсии ($v_2 = 8$). На пересечении находим значение 4,46. Для дисперсии повторностей

теоретическое значение F находим на пересечении $\nu_1 = 4$ (по столбцу) и $\nu_2 = 8$ (по строке), равное 3,84.

9. Результаты решения запишем в таблицу 3.7.

Таблица 3.7

Расчет и анализ дисперсий

Источник вариации	Объем вариации W	Число степеней свободы ν	Дисперсия S^2	Значение F -распределения	
				фактическое	табличное
Способ внесения гербицида	366,9333	2	183,47	40,32	4,46
Плодородие почвы	29,6000	4	7,40	1,63	3,84
Случайные факторы	36,4000	8	4,55	x	x
Итого	432,9333	14	x	x	x

10. Сделаем вывод. Сопоставив фактическое и табличное значение критерия F , по первому источнику вариации (способ внесения гербицида) фактическое значение критерия превышает табличное. Так как фактическое значение попадает в критическую область критерия, нулевая гипотеза о равенстве выборочных средних величин в генеральных совокупностях должна быть отвергнута и принята альтернативная гипотеза. Применительно к содержанию задачи заключаем, что с вероятностью 0,95 способ внесения гербицида существенно влияет на уровень урожайности культуры. По второму источнику вариации (плодородие почв) фактическое значение критерия F ниже табличного, следовательно, попадает в область допустимых значений. Принимаем нулевую гипотезу как верную и делаем вывод о том, что различия в урожайности по повторностям нельзя признать существенными.

Задача 3.2 Проверка статистических гипотез относительно средних по данным двух выборок на основе результатов однофакторного дисперсионного анализа по критерию Q-Тьюки

В задачах 3.1.1 и 3.1.2. с помощью критерия F была доказана существенность различий между средними величинами по группам (вариантам опыта), поэтому возможна и целесообразна дальнейшая оценка достоверности различий при сопоставлении двух средних величин. Применительно к задаче 3.1.1 представляет практический интерес решение вопроса о достоверности различий в урожайности картофеля между 1 и 2, 2 и 3, 1 и 3 видами удобрения. Поскольку численность единиц в каждой группе

одинакова, для оценки этих различий возможно применение критерия Q-Тьюки. Проведем проверку гипотез в следующей последовательности:

1. Расположим выборочные средние величины в порядке возрастания, то есть построим ранжированный ряд и запишем в таблицу 3.8.

Таблица 3.8

Ранжированный ряд групповых средних величин

Средние величины в ранжированном ряду	$\bar{x}_1$ (без удобрений)	$\bar{x}_2$ (навоз)	$\bar{x}_3$ (навоз+РК)
Значения средних урожайностей картофеля, ц/га	236,6	344,8	430,2

2. Сформулируем статистические гипотезы относительно средних в ранжированном ряду (таблица 3.8) при парном их сравнении:

Нулевые гипотезы

$$H_0: \tilde{x}_1 = \tilde{x}_2$$

$$H_0: \tilde{x}_2 = \tilde{x}_3$$

$$H_0: \tilde{x}_1 = \tilde{x}_3$$

Альтернативные гипотезы

$$H_1: \tilde{x}_1 \neq \tilde{x}_2$$

$$H_1: \tilde{x}_2 \neq \tilde{x}_3$$

$$H_1: \tilde{x}_1 \neq \tilde{x}_3$$

3. Вычислим среднюю ошибку выборок, опираясь в расчетах на остаточную дисперсию признака, обусловленную случайными факторами:

$$m = \sqrt{\frac{S_{\text{ост}}^2}{n}} = \sqrt{\frac{183,9}{5}} = \sqrt{36,78} = 6,065.$$

4. Фактические значения критерия Q-Тьюки определяем как отношение разности средних (рассматриваемой как предельная ошибка) к средней ошибке выборок. При наличии трех разностей средних рассчитаем три фактических значения критерия и запишем в таблицу 3.9.

$$Q_{\text{факт1}} = \frac{|\bar{x}_1 - \bar{x}_2|}{m}; Q_{\text{факт2}} = \frac{|\bar{x}_2 - \bar{x}_3|}{m}; Q_{\text{факт3}} = \frac{|\bar{x}_1 - \bar{x}_3|}{m}. \quad (3.25)$$

Таблица 3.9

Фактическое и табличное значение критерия Q-Тьюки

Разности средних величин		Значение разности, ц/га	Значение критерия Q-Тьюки	
			фактическое	теоретическое
Первого порядка	$ \bar{x}_1 - \bar{x}_2 $	108,2	17,840	3,082
	$ \bar{x}_2 - \bar{x}_3 $	85,4	14,081	3,082
Второго порядка	$ \bar{x}_1 - \bar{x}_3 $	193,6	31,921	3,773

5. Теоретическое значение критерия определяем по таблице «Стьюдентизированный размах при уровне значимости 0,05» (Приложение 7) с учетом заданного уровня значимости 0,05, числа степеней свободы для

остаточной дисперсии ($\nu_{\text{ост}} = 12$) и порядка разностей (K). Теоретические значения Q-Тьюки также запишем в таблицу 3.9.

К разностям первого порядка относятся разности между соседними средними величинами в ранжированном ряду. В рассматриваемой задаче к ним относятся разности $|\bar{x}_1 - \bar{x}_2|$ и $|\bar{x}_2 - \bar{x}_3|$. Для них значение порядка разностей $K=2$, а критическое значение критерия Q-Тьюки при степени свободы 12 составляет 3,082.

К разностям второго порядка относятся разности между средними, расположенными в ранжированном ряду через одну величину ($|\bar{x}_1 - \bar{x}_3|$). Для этих разностей значение порядка $K=3$. Критическое значение Q-Тьюки при числе степеней свободы 12 равно 3,773.

6. Сравним фактические и критические значения критерия Q-Тьюки и сделаем вывод. Нулевые гипотезы о равенстве средних по всем парам должны быть отвергнуты, так как фактические значения критерия Q-Тьюки больше соответствующих критических значений. Принимаются альтернативные гипотезы о неравенстве средних в генеральных совокупностях.

7. Практически значимый вывод: с вероятностью ошибки 5 случаев из 100 можно утверждать, что урожайность картофеля существенно различается по видам внесенных удобрений. Самая высокая урожайность картофеля при внесении навоза+РК, на среднем уровне урожайность картофеля при внесении навоза и самая низкая урожайность картофеля без внесения удобрений.

Задача 3.3 Однофакторный дисперсионный анализ при случайном формировании выборок неравной численности с попарным сравнением средних по критерию Шеффе

Условие: имеются данные по урожайности ячменя разных доз фосфорных удобрений (таблица 3.10).

Таблица 3.10

Урожайность ячменя в сосудах, г

Доза фосфорных удобрений	Урожайность ячменя в сосудах, г					Итого $\sum_{j=1}^{n_i} x_j$	Средняя $\bar{x}_i$
	1	2	3	4	5		
1	32,6	33,1	31,9	34,6	33,8	166,0	33,2
2	39,6	40,1	38,2	-	-	117,9	39,3
3	49,6	50,1	52,1	50,6	-	202,4	50,6
Итого $\sum_{i=1}^m x_i$	121,8	123,3	122,2	85,2	33,8	486,3	41,0

Оценить достоверность различий в урожайности по дозам фосфорных удобрений при уровне вероятности суждения 0,95.

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2 = \tilde{x}_3$ (различий в средней урожайности по дозам фосфорных удобрений нет);

$H_1: \tilde{x}_1 \neq \tilde{x}_2 \neq \tilde{x}_3$ (средняя урожайность по дозам фосфорных удобрений достоверно различна).

2. Определим модель дисперсионного анализа. Группировочный признак – один (доза фосфорных удобрений), следовательно, модель однофакторная. Так как в условии задачи специально не оговаривается наличие каких-либо общих условий при отборе сосудов, следует признать выборки независимыми. Таким образом, в нашей задаче два источника вариации: 1) доза фосфорных удобрений (признак, по которому проведена группировка сосудов); 2) остальные (случайные) факторы: индивидуальные особенности растений и т.п.

Общий объем вариации равен сумме объемов вариации по источникам варьирования: $W_{\text{общ}} = W_{\text{фактора}} + W_{\text{ост.}}$

3. Все необходимые данные для расчета объемов вариации представлены в таблице 3.11.

Таблица 3.11

Квадраты значений признаков и их сумм

Доза фосфор- ных удобре- ний	Повторности					Сумма квад- ратов $\sum_{j=1}^{n_i} x_{ij}^2$	Квадрат суммы $\sum_{j=1}^{n_i} x_{ij}$ ²
	1	2	3	4	5		
1	1062,76	1095,61	1017,61	1197,16	1142,44	5515,58	27556,00
2	1568,16	1608,01	1459,24	-	-	4635,41	13900,41
3	2460,16	2510,01	2714,41	2560,36	-	10244,94	40965,76
Сумма квadra- тов $\sum_{i=1}^m x_{ij}^2$	14835,24	15202,89	14932,84	7259,04	1142,44	$\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}^2$ =20395,93	$\sum_{i=1}^m (\sum_{j=1}^{n_i} x_{ij})^2$ =82422,17

Расчет объемов вариации по источникам варьирования проведем по рабочим формулам.

Общий объем вариации урожайности равен:

$$W_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} = 20395,93 - \frac{486,3^2}{12} = 688,62.$$

Объем вариации урожайности, обусловленный дозами удобрений:

$$W_{\text{фактора}} = \sum_{i=1}^m \left(\frac{(\sum_{j=1}^n x_{ij})^2}{n_i} \right) - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} =$$

$$= \left(\frac{27556,00}{5} + \frac{13900,41}{3} + \frac{40965,76}{4} \right) - \frac{486,3^2}{12} = 678,80.$$

Объем вариации, обусловленный прочими, случайными факторами равен:

$$W_{\text{ост}} = W_{\text{общ}} - W_{\text{фактора}} = 688,62 - 678,80 = 9,82.$$

4. Рассчитаем число степеней свободы для каждого источника варьирования.

Общее число наблюдений $N=12$, число групп $m=3$.

Для общего объема вариации число степеней свободы равно:

$$v_0 = N - 1 = 12 - 1 = 11.$$

Число степеней свободы для факторной дисперсии:

$$v_{\text{фактора}} = m - 1 = 3 - 1 = 2.$$

Число степеней свободы для остаточной дисперсии:

$$v_{\text{ост}} = v_0 - v_{\text{фактора}} = 11 - 2 = 9.$$

5. Рассчитаем необходимые для дальнейшего анализа дисперсии признака.

– дисперсия по фактору:

$$S_{\text{фактора}}^2 = \frac{W_{\text{фактора}}}{v_{\text{фактора}}} = \frac{678,80}{2} = 339,40;$$

– дисперсия остаточная:

$$S_{\text{ост}}^2 = \frac{W_{\text{ост}}}{v_{\text{ост}}} = \frac{9,82}{9} = 1,09.$$

6. Определим фактическое значение F-критерия:

$$F_{\text{факт}} = \frac{S_{\text{фактора}}^2}{S_{\text{ост}}^2} = \frac{339,40}{1,09} = 311,38.$$

7. Определим теоретическое значение критерия F–распределения при заданном уровне значимости 0,05 (Приложение 3). Теоретическое значение F находим на пересечении $v_1 = 2$ (по столбцу) и $v_2 = 9$ (по строке), равное 4,26.

8. Результаты решения запишем в таблицу 3.12.

9. Сделаем вывод. Сопоставим фактическое и табличное значение критерия F. Так как фактическое значение критерия F превышает табличное и находится в критической области, нулевая гипотеза о равенстве выборочных средних величин в генеральных совокупностях должна быть

отвергнута. Принимаем альтернативную гипотезу. С вероятностью 0,95 можно утверждать, что между уровнями урожайности по дозам фосфорных удобрений имеются достоверные различия.

Таблица 3.12

Расчет и анализ дисперсий

Источник вариации	Объем вариации W	Число степеней свободы ν	Дисперсия S^2	Значение F-распределения	
				фактическое	табличное
Доза удобрения	678,80	2	339,40	311,38	4,26
Случайные факторы	9,82	9	1,09	x	x
Итого	688,62	11	x	x	x

10. С целью проверки статистических гипотез относительно достоверности попарных различий между сортами при разных по численности выборок применим *критерий F-Шеффе*.

Проверку гипотез проведем в следующей последовательности:

а) сформулируем статистические гипотезы относительно средних величин в генеральных совокупностях при парном их сравнении:

Нулевые гипотезы

Альтернативные гипотезы

$$H_0: \tilde{x}_1 = \tilde{x}_2$$

$$H_1: \tilde{x}_1 \neq \tilde{x}_2$$

$$H_0: \tilde{x}_2 = \tilde{x}_3$$

$$H_1: \tilde{x}_2 \neq \tilde{x}_3$$

$$H_0: \tilde{x}_1 = \tilde{x}_3$$

$$H_1: \tilde{x}_1 \neq \tilde{x}_3$$

б) рассчитаем фактические разности выборочных средних, сопоставив попарно средние урожайности по дозам удобрений по следующим формулам:

$$\varphi_{1,2} = |\bar{x}_2 - \bar{x}_1|; \varphi_{2,3} = |\bar{x}_3 - \bar{x}_2|; \varphi_{1,3} = |\bar{x}_3 - \bar{x}_1|. \quad (3.26)$$

Фактические разности выборочных средних равны:

$$\varphi_{1,2} = |\bar{x}_2 - \bar{x}_1| = |39,3 - 33,2| = 6,1;$$

$$\varphi_{2,3} = |\bar{x}_3 - \bar{x}_2| = |50,6 - 39,3| = 11,3;$$

$$\varphi_{1,3} = |\bar{x}_3 - \bar{x}_1| = |50,6 - 33,2| = 17,4.$$

в) для каждой разности найдем ее ошибку выборки по формуле:

$$S_\varphi = \sqrt{S_{\text{ост}}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad (3.27)$$

где $S_{\text{ост}}^2$ – остаточная дисперсия,

n_1 – численность 1-ой выборки,

n_2 – численность 2-ой выборки.

Для первой разности:

$$S_{\varphi_{1,2}} = \sqrt{S_{\text{ост}}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{1,09 \left(\frac{1}{5} + \frac{1}{3} \right)} = 0,762;$$

Для второй разности:

$$S_{\varphi_{2,3}} = \sqrt{S_{\text{ост}}^2 \left(\frac{1}{n_2} + \frac{1}{n_3} \right)} = \sqrt{1,09 \left(\frac{1}{3} + \frac{1}{4} \right)} = 0,797;$$

Для третьей разности:

$$S_{\varphi_{1,3}} = \sqrt{S_{\text{ост}}^2 \left(\frac{1}{n_1} + \frac{1}{n_3} \right)} = \sqrt{1,09 \left(\frac{1}{5} + \frac{1}{4} \right)} = 0,700.$$

г) рассчитаем фактические значения критерия Шеффе как отношение разности средних (рассматриваемой как предельная ошибка) к ее ошибке выборки:

$$F_{\varphi_i} = \frac{\varphi_i}{S_{\varphi_i}} \quad (3.28)$$

Фактические значения критерия Шеффе составят:

$$\begin{aligned} F_{\varphi_{1,2}} &= \frac{\varphi_{1,2}}{S_{\varphi_{1,2}}} = \frac{6,1}{0,762} = 8,01; \\ F_{\varphi_{2,3}} &= \frac{\varphi_{2,3}}{S_{\varphi_{2,3}}} = \frac{11,3}{0,797} = 14,18; \\ F_{\varphi_{1,3}} &= \frac{\varphi_{1,3}}{S_{\varphi_{1,3}}} = \frac{17,4}{0,700} = 24,86. \end{aligned}$$

д) теоретическое значение критерия определяется следующим образом:

$$F_{\varphi_{\text{табл}}} = \sqrt{F_{\alpha}(L - 1)},$$

где L – число групп,

F_{α} – критерий Фишера при заданном уровне значимости, числа степеней свободы для факторной дисперсии ($\nu_1 = 2$) и числа степеней свободы для остаточной дисперсии ($\nu_2 = 9$), равный 4,26.

Теоретическое значение критерия равно:

$$F_{\varphi_{\text{табл}}} = \sqrt{4,26(3 - 1)} = 2,92.$$

е) сравним фактические значения критерия Шеффе с критическим значением и сделаем вывод. Нулевая гипотеза о равенстве средних в первой ($\tilde{x}_1 - \tilde{x}_2$), второй ($\tilde{x}_2 - \tilde{x}_3$) и третьей ($\tilde{x}_1 - \tilde{x}_3$) парах должна быть отвергнута, так как фактические значения критерия Шеффе больше критического значения. Принимаем альтернативную гипотезу. Следовательно, с вероятностью 0,95 можно утверждать, что урожайность ячменя при дозах фосфорных удобрений 1 и 2, 2 и 3, 1 и 3 различается существенно.

Задача 3.4 Двухфакторный дисперсионный анализ при неслучайном формировании равночисленных выборок

Условие: имеются данные полевого опыта, в котором изучалось влияние дозы извести и дозы P_2O_5 на урожайность зерновых культур (таблица 3.13). Известно, что при выборе земельных участков для повторностей опыта учитывалось плодородие почв.

Статистически оценить результаты опыта при уровне значимости 0,05.

Таблица 3.13

Урожайность зерновых культур в зависимости от дозы извести и дозы P_2O_5 , ц с 1 га

Доза извести	Доза P_2O_5 , кг/га	Земельные участки (повторности)				Итого $\sum_{j=1}^{n_i} x_j$	Средняя по варианту опыта $\bar{x}_i$
		1	2	3	4		
0,5 рН	0	31,6	31,9	35,1	36,8	135,4	33,9
	360	39,6	39,9	42,6	44,7	166,8	41,7
	720	48,2	49,1	53,0	56,0	206,3	51,6
2,0 рН	0	33,4	33,7	37,2	38,6	142,9	35,7
	360	42,8	43,2	45,4	46,7	178,1	44,5
	720	50,4	53,0	59,7	60,3	223,4	55,9
Итого по повторности	$\sum_{i=1}^m x_i$	246,0	250,8	273,0	283,1	$\sum_{i=1}^m \sum_{j=1}^n x_{ij} = 1052,9$	x
Средняя по повторности	$\bar{x}_j$	41,0	41,8	45,5	47,2	x	$\bar{x}_0 = 43,9$

Решение:

1. Выдвинем нулевую и альтернативную гипотезы:

$H_0: \tilde{x}_1 = \tilde{x}_2 = \tilde{x}_3 = \tilde{x}_4 = \tilde{x}_5 = \tilde{x}_6$ (средние урожайности по вариантам опыта в генеральной совокупности равны);

$H_1: \tilde{x}_1 \neq \tilde{x}_2 \neq \tilde{x}_3 \neq \tilde{x}_4 \neq \tilde{x}_5 \neq \tilde{x}_6$ (средние урожайности по вариантам опыта не равны).

2. Определим модель дисперсионного анализа в соответствии с источниками вариации урожайности и условиями формирования выборок. При наличии двух группировочных признаков (доза извести – фактор А и доза P_2O_5 – фактор В), подбора делянок с учетом плодородия почв имеем модель двухфакторного дисперсионного анализа с неслучайным формированием групп равной численности.

При этом типе модели общий объем вариации может быть представлен как сумма:

$$W_{\text{общ}} = W_{\text{фактора А}} + W_{\text{фактора В}} + W_{\text{взаимод АВ}} + W_{\text{повт}} + W_{\text{ост}}. \quad (3.29)$$

3. Введем условные обозначения и рассчитаем необходимые показатели таблицы 3.14.

Таблица 3.14

Квадраты значений признаков и их сумм

Доза извести	Доза P_2O_5 , кг/га	Повторности				Сумма квадратов $\sum_{j=1}^n x_{ij}^2$	Квадрат суммы $(\sum_{j=1}^n x_{ij})^2$
		1	2	3	4		
0,5 рН	0	998,56	1017,61	1232,01	1354,24	4602,42	18333,16
	360	1568,16	1592,01	1814,76	1998,09	6973,02	27822,24
	720	2323,24	2410,81	2809,00	3136,00	10679,05	42559,69
2,0 рН	0	1115,56	1135,69	1383,84	1489,96	5125,05	20420,41
	360	1831,84	1866,24	2061,16	2180,89	7940,13	31719,61
	720	2540,16	2809,00	3564,09	3636,09	12549,34	49907,56
Сумма квадратов $\sum_{i=1}^m x_{ij}^2$		10377,52	10831,36	12864,86	13795,27	$\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2$ =47869,01	$\sum_{i=1}^m (\sum_{j=1}^n x_{ij})^2$ =190762,67
Квадрат суммы $(\sum_{i=1}^m x_{ij})^2$		60516,00	62900,64	74529,00	80145,61	$\sum_{j=1}^n (\sum_{i=1}^m x_{ij})^2$ =278091,25	$(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2$ =1108598,41

По первому признаку (доза извести) выделено 2 группы ($m=2$), по второму признаку (доза P_2O_5) выделено 3 группы ($k=3$). В каждой подгруппе выборка представлена 4 наблюдениями ($n=4$). Общая численность совокупности равна $N = m \cdot k \cdot n = 2 \cdot 3 \cdot 4 = 24$. $\sum_{j=1}^n x_j$ – суммы урожайностей по каждому варианту опыта (строке); $\sum_{i=1}^m x_i$ – суммы урожайностей по каждой повторности (столбцу).

4. Рассчитаем объемы вариации урожайности зерновых культур в соответствии с источниками вариации.

Расчет объемов вариации проводим в два этапа.

– на первом этапе обеспечим разложение: $W_{\text{общ}} = W_{\text{фактора}} + W_{\text{повт}} + W_{\text{ост}}$;

– на втором этапе: $W_{\text{фактора}} = W_A + W_B + W_{AB}$.

Первый этап реализуется по типу однофакторной модели. Используя для расчетов данные таблицы 3.14, определим:

$$W_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} = 47869,01 - \frac{1108598,41}{24} =$$

$$= 47869,01 - 46191,60 = 1677,41;$$

$$W_{\text{фактора}} = \frac{\sum_{i=1}^m (\sum_{j=1}^n x_{ij})^2}{n} - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} = \frac{190762,67}{4} - \frac{1108598,41}{24} =$$

$$= 47690,67 - 46191,60 = 1499,07;$$

$$W_{\text{повт}} = \frac{\sum_{j=1}^n (\sum_{i=1}^m x_{ij}^2)^2}{m \cdot k} - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N} = \frac{278091,25}{2 \cdot 3} - \frac{1108598,41}{24} =$$

$$= 46348,54 - 46191,60 = 156,94;$$

$$W_{\text{ост}} = W_{\text{общ}} - W_{\text{фактора}} - W_{\text{повт}} = 1677,41 - 1499,07 - 156,94 = 21,4.$$

5. Определим для каждого объема вариации число степеней свободы:

$$v_0 = N - 1 = 24 - 1 = 23;$$

$$v_{\text{фактора}} = mk - 1 = 2 \cdot 3 - 1 = 5;$$

$$v_{\text{повт}} = n - 1 = 4 - 1 = 3;$$

$$v_{\text{ост}} = v_0 - v_{\text{фактора}} - v_{\text{повт}} = 23 - 5 - 3 = 15.$$

6. Определим дисперсии для первого этапа анализа:

$$S_{\text{фактора}}^2 = \frac{W_{\text{фактора}}}{v_{\text{фактора}}} = \frac{1499,07}{5} = 299,81;$$

$$S_{\text{ост}}^2 = \frac{W_{\text{ост}}}{v_{\text{ост}}} = \frac{21,40}{15} = 1,43.$$

7. Фактическое значение критерия F-распределения по факторам в целом:

$$F_{\text{факт}} = \frac{S_{\text{фактора}}^2}{S_{\text{ост}}^2} = \frac{299,81}{1,43} = 210,14.$$

8. Табличное значение $F_{\text{табл}}$ при уровне значимости 0,05 и числе степеней свободы $v_1 = 5$; $v_2 = 15$ составит 2,90. Фактическое значение критерия больше табличного, следовательно, с вероятностью 0,95 можно утверждать, что различия между средними величинами по вариантам опыта достоверны и целесообразна дальнейшая работа по оценке достоверности влияния каждого фактора и их взаимодействия.

9. Проведем второй этап разложения объемов вариаций:

$$W_{\text{фактора}} = W_A + W_B + W_{AB}.$$

Для расчета объемов вариации необходимо составить новую таблицу 3.15, в которой заголовки строк содержат градации по первому фактору, а столбцов – по второму. В таблицу запишем суммы урожайности по повторностям из таблицы 3.13 (графа «Итого $\sum_{j=1}^{n_i} x_j$ »).

Таблица 3.15

Таблица сумм урожайности по вариантам опыта $\sum_{j=1}^{n_i} x_j$

Доза извести	Доза P_2O_5 , кг/га			Сумма А
	0	360	720	
0,5 рН	135,4	166,8	206,3	508,5
2,0 рН	142,9	178,1	223,4	544,4
Сумма В	278,3	344,9	429,7	$\sum x_{ij} = 1052,9$

10. Дальнейшие расчеты проводим по аналогии с расчетами первого этапа работы, представляя вместо $W_{\text{общ}}$ объем вариации по факторам в целом $W_{\text{фактора}}$, вместо $W_{\text{фактора}}$ объем вариации по первому фактору W_A , вместо $W_{\text{повт}}$ объем вариации по второму фактору W_B , вместо $W_{\text{ост}}$ объем вариации взаимодействия факторов W_{AB} :

$$W_A = \frac{\sum_{i=1}^m A^2}{n \cdot k} - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N}; \quad (3.30)$$

$$W_B = \frac{\sum_{j=1}^n B^2}{n \cdot m} - \frac{(\sum_{i=1}^m \sum_{j=1}^n x_{ij})^2}{N}; \quad (3.31)$$

$$W_{AB} = W_{\text{фактора}} - W_A - W_B. \quad (3.32)$$

Объем вариации по первому фактору W_A :

$$W_A = \frac{508,5^2 + 544,4^2}{4 \cdot 3} - \frac{1108598,41}{24} = 46245,30 - 46191,60 = 53,70.$$

Объем вариации по второму фактору W_B :

$$W_B = \frac{278,3^2 + 344,9^2 + 429,7^2}{4 \cdot 2} - \frac{1108598,41}{24} = 47631,12 - 46191,60 = 1439,52.$$

Объем вариации взаимодействия факторов W_{AB} :

$$W_{AB} = 1499,07 - 53,70 - 1439,52 = 5,85.$$

11. Определим для вычисленных в пункте 10 объемов вариаций соответствующее число степеней свободы по формулам:

$$v_A = m - 1; \quad (3.33)$$

$$v_B = k - 1; \quad (3.34)$$

$$v_{AB} = (mk - 1) - (m - 1) - (k - 1). \quad (3.35)$$

Число степеней свободы по первому фактору А:

$$v_A = 2 - 1 = 1;$$

Число степеней свободы по второму фактору В:

$$v_B = 3 - 1 = 2;$$

Число степеней свободы взаимодействия факторов АВ:

$$v_{AB} = 5 - 1 - 2 = 2.$$

12. Объединим результаты вычислений, проведенные на первом и втором этапах, в таблицу 3.16 и рассчитаем необходимые показатели.

Таблица 3.16

Анализ дисперсий

Источник вариации	Объем вариации W	Число степеней свободы ν	Дисперсия S^2	Значение F-распределения	
				фактическое	табличное
Факторы в целом	1499,07	5	299,81	210,14	2,90
в том числе					
- доза извести	53,7	1	53,70	37,64	4,54
- доза P_2O_5	1439,52	2	719,76	504,48	3,68
- взаимодей. факторов	5,84	2	2,92	2,05	3,68
Повторности	156,94	3	52,31	36,67	3,29
Остаточная вариация	21,40	15	1,43	x	x
Общая вариация	1677,41	23	x	x	x

13. Сопоставление $F_{\text{факт.}}$ и $F_{\text{табл.}}$ показывают, что существенны различия в средней урожайности зерновых культур по первому и второму фактору. Выдвинутая нулевая гипотеза подтверждается по другому источнику вариации: взаимодействию факторов; следовательно, по данным опыта не доказано влияние взаимодействия факторов. Вероятность ошибки суждения сравнительно низка – 5 %. Превышение фактического значения критерия F над табличным свидетельствует о существенных различиях делянок по плодородию почв. Сделанные выводы позволяют продолжить дисперсионный анализ для оценки достоверности различий урожайности зерновых культур при парном сравнении дозы извести или дозы P_2O_5 .

Задача 3.5 Определение показателей связи при парной линейной корреляции

Условие. В таблице 3.17 представлены данные выборочного наблюдения сельскохозяйственных организаций (СХО) о дозе внесенных минеральных удобрений в пересчете на 100% питательных веществ (кг на 1 га посева пшеницы) и урожайности пшеницы (ц/га убранной площади).

Таблица 3.17

Данные для определения показателей корреляционной связи

Номер СХО	Исходные данные		Расчетные величины		
	урожайность пшеницы, ц/га, y	доза внесенных минеральных удобрений, кг/га, x	yx	y^2	x^2
1	2	3	4	5	6
1	32	72	2304	1024	5184
2	30	145	4350	900	21025

Номер СХО	Исходные данные		Расчетные величины		
	урожайность пшеницы, ц/га, y	доза внесенных минеральных удобрений, кг/га, x	yx	y^2	x^2
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
3	45	172	7740	2025	29584
4	29	119	3451	841	14161
5	39	148	5772	1521	21904
6	28	103	2884	784	10609
7	21	91	1911	441	8281
8	17	16	272	289	256
9	35	120	4200	1225	14400
10	24	118	2832	576	13924
11	30	113	3390	900	12769
12	20	28	560	400	784
13	28	69	1932	784	4761
14	35	107	3745	1225	11449
15	31	156	4836	961	24336
16	24	50	1200	576	2500
17	20	68	1360	400	4624
18	24	73	1752	576	5329
19	25	83	2075	625	6889
20	31	102	3162	961	10404
21	27	131	3537	729	17161
22	37	88	3256	1369	7744
23	39	150	5850	1521	22500
24	36	104	3744	1296	10816
25	38	170	6460	1444	28900
26	18	47	846	324	2209
27	26	154	4004	676	23716
28	32	86	2752	1024	7396
29	24	117	2808	576	13689
30	28	83	2324	784	6889
Итого	873	3083	95309	26777	364193
Средняя	29,10	102,77	3176,97	892,57	12139,77

Требуется: построить уравнение регрессии и определить показатели тесноты связи урожайности пшеницы с дозой внесенных минеральных удобрений.

Решение:

1. Построим график, чтобы установить форму связи между признаками (рисунок 3.1). Для этого на корреляционное поле нанесем фактические значения зависимой переменной (y – урожайность пшеницы, ц с 1 га убранной площади) при соответствующих значениях независимой переменной (x – доза внесенных минеральных удобрений в пересчете на 100% питательных веществ, кг на 1 га посева пшеницы).

Расположение точек на корреляционном поле позволяет сделать вывод о том, что связь между переменными близка к линейной.

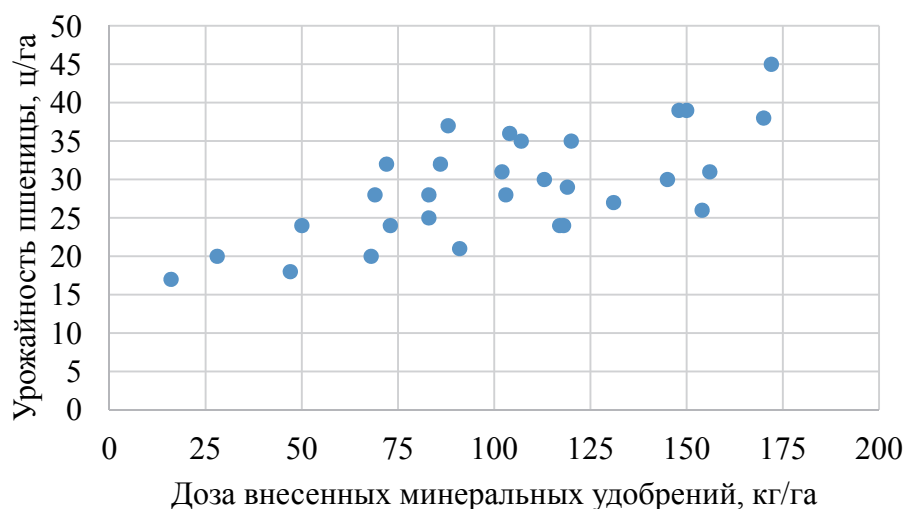


Рисунок 3.1 – Зависимость урожайности пшеницы от дозы внесенных минеральных удобрений

Запишем уравнение связи урожайности пшеницы с дозой внесенных минеральных удобрений в пересчете на 100% питательных веществ:

$$\tilde{y} = a + bx,$$

где a и b – неизвестные параметры уравнения.

2. Определим оценки параметров уравнения регрессии с использованием метода наименьших квадратов. Составим систему нормальных уравнений, необходимых для определения неизвестных параметров a и b :

$$\begin{cases} \sum y = na + b \sum x \\ \sum yx = a \sum x + b \sum x^2 \end{cases}$$

3. Рассчитаем величины, которые нужны для решения системы нормальных уравнений и определения показателей тесноты связи (таблица 3.17): $\sum y, \sum x, \sum yx, \sum y^2, \sum x^2$.

4. Из таблицы 3.17 подставим в систему нормальных уравнений соответствующие конкретные величины:

$$\begin{cases} 873 = 30a + 3083b \\ 95309 = 3083a + 364193b \end{cases}$$

5. Решим систему уравнений и определим неизвестные параметры a и b , для чего:

а) разделим оба уравнения на коэффициенты при a (первое уравнение – на 30, второе – на 3083):

$$\begin{cases} 29,1 = a + 102,76667b \\ 30,91437 = a + 118,12942b \end{cases}$$

б) вычтем из второго уравнения первое:

$$1,81437 = 15,36275b;$$

в) определим параметр b :

$$b = \frac{1,81437}{15,36275} = 0,118102 \approx 0,118.$$

г) подставим в одно из уравнений значение параметра b и исчислим параметр a :

$$a = 29,1 - 102,76667 \cdot 0,118102 = 29,1 - 12,13695 = 16,96305 \approx 16,96.$$

6. Подставим найденные значения параметров a и b в уравнение связи:

$$\tilde{y} = 16,96 + 0,118x.$$

7. Для расчета показателей тесноты связи определим величины:

– среднее квадратическое отклонение по урожайности:

$$\sigma_y = \sqrt{y^2 - \bar{y}^2} = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{26777}{30} - \left(\frac{873}{30}\right)^2} = \sqrt{45,757} = 6,76;$$

– среднее квадратическое отклонение по дозе внесенных минеральных удобрений:

$$\sigma_x = \sqrt{x^2 - \bar{x}^2} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{364193}{30} - \left(\frac{3083}{30}\right)^2} = \sqrt{1578,779} = 39,73;$$

– среднее произведение зависимой и независимой переменной:

$$\overline{xy} = \frac{\sum xy}{n} = \frac{95309}{30} = 3176,97;$$

– произведение средних величин зависимой и независимой переменной:

$$\bar{x} \cdot \bar{y} = \frac{\sum x}{n} \cdot \frac{\sum y}{n} = \frac{3083}{30} \cdot \frac{873}{30} = 102,77 \cdot 29,1 = 2990,61.$$

8. Коэффициент корреляции определяется по формуле:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \quad (3.36)$$

Рассчитаем коэффициент корреляции:

$$r_{xy} = \frac{3176,97 - 2990,61}{39,73 \cdot 6,76} = 0,694.$$

9. Рассчитаем коэффициент детерминации:

$$r_{xy}^2 = 0,694^2 = 0,481.$$

На рисунке 3.2 отображены линия регрессии, добавленная на диаграмму рассеяния, уравнение прямой и коэффициент детерминации.

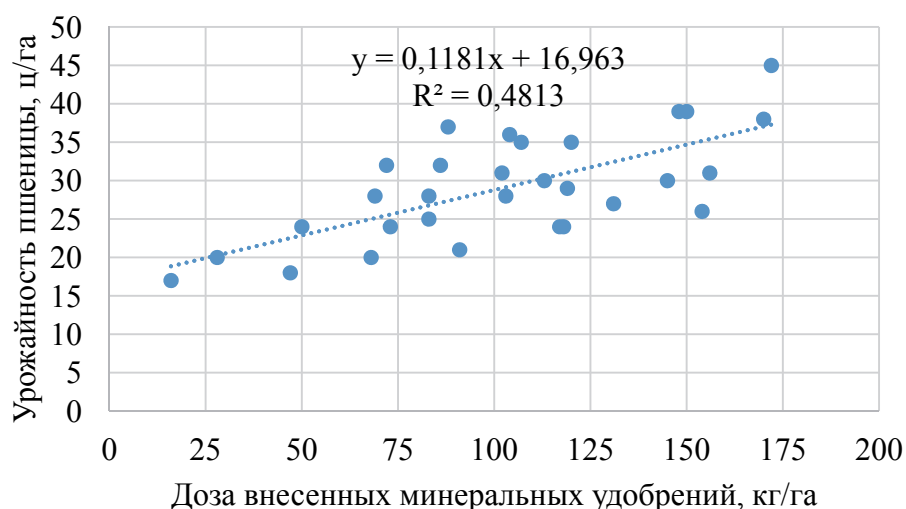


Рисунок 3.2 – Зависимость урожайности пшеницы от дозы внесенных минеральных удобрений

10. Сделаем вывод. Судя по уравнению регрессии: $\tilde{y} = 16,96 + 0,118x$, при повышении дозы внесенных минеральных удобрений на 1 кг в расчете на 1 га посева пшеницы урожайность пшеницы в среднем по совокупности повысится на 0,118 ц с 1 га. Коэффициент корреляции 0,694 показывает, что связь между изучаемыми признаками средняя. Коэффициент детерминации 0,481 показывает, что 48,1% вариации урожайности пшеницы объясняется вариацией дозы внесенных минеральных удобрений.

Задача 3.6 Проверка гипотез и статистическая оценка показателей связи при парной линейной корреляции

Условие. по данным выборочной совокупности решено уравнение и определены показатели связи урожайности пшеницы с дозой внесенных минеральных удобрений (см. задачу 3.5).

Требуется проверить гипотезы и провести интервальную оценку коэффициентов регрессии и коэффициента корреляции.

Решение:

1. Запишем решенное в задаче 3.5 уравнение связи урожайности с дозой внесенных минеральных удобрений: $\tilde{y} = 16,96 + 0,118x$. Так как уравнение регрессии построено по выборочным данным, то оно является лишь оценкой генерального уравнения $\tilde{y} = \alpha + \beta x$.

Уравнение в целом и выборочные оценки параметров могут оказаться незначимыми для генеральной совокупности в силу ошибок выборки, поэтому необходима проверка их достоверности (значимости).

Проверить достоверность уравнения регрессии – значит, установить: соответствует ли математическая модель, выражающая зависимость между переменными связям в генеральной совокупности и достаточно ли

включенных в уравнение объясняющих переменных для описания зависимой переменной.

Проверка достоверности уравнения в целом проводится на основе дисперсионного анализа по критерию F-Фишера.

2. Проведем F-тест. Сформулируем рабочую гипотезу о равенстве генеральных дисперсий: дисперсии, воспроизведенной уравнением регрессии ($\sigma_{\text{регр}}^2$), и остаточной дисперсии ($\sigma_{\text{ост}}^2$), а также альтернативную ей:

$$H_0: \sigma_{\text{регр}}^2 = \sigma_{\text{ост}}^2,$$

$$H_1: \sigma_{\text{регр}}^2 \neq \sigma_{\text{ост}}^2.$$

Выберем уровень значимости критерия $\alpha = 0,05$.

Для расчета F-критерия нам нужно рассчитать несмещенные оценки дисперсии, воспроизведенной регрессией, и остаточной дисперсии. Для этого необходимо определить соответствующие объемы вариации.

Общий объем вариации мы можем найти, умножив выборочную дисперсию зависимой переменной на численность выборки:

$$W_{\text{общ}} = W_y = \sigma_y^2 \cdot n = 45,757 \cdot 30 = 1372,7.$$

Далее определим выровненные значения для каждой единицы выборки ($\tilde{y}_i$), подставив индивидуальные значения независимой переменной в уравнение регрессии. Затем рассчитаем остатки ($e_i = y_i - \tilde{y}_i$) и их квадраты (e_i^2) для каждого наблюдения (таблица 3.18).

Таблица 3.18

Определение выровненных значений и остатков

Номер СХО	Расчетные величины		
	$\tilde{y}_i$	e_i	e_i^2
1	2	3	4
1	25,456	6,544	42,824
2	34,070	-4,070	16,565
3	37,256	7,744	59,970
4	31,002	-2,002	4,008
5	34,424	4,576	20,940
6	29,114	-1,114	1,241
7	27,698	-6,698	44,863
8	18,848	-1,848	3,415
9	31,120	3,880	15,054
10	30,884	-6,884	47,389
11	30,294	-0,294	0,086
12	20,264	-0,264	0,070
13	25,102	2,898	8,398
14	29,586	5,414	29,311
15	35,368	-4,368	19,079
16	22,860	1,140	1,300
17	24,984	-4,984	24,840
18	25,574	-1,574	2,477
19	26,754	-1,754	3,077

Номер СХО	Расчетные величины		
	$\tilde{y}_i$	e_i	e_i^2
1	2	3	4
20	28,996	2,004	4,016
21	32,418	-5,418	29,355
22	27,344	9,656	93,238
23	34,660	4,340	18,836
24	29,232	6,768	45,806
25	37,020	0,980	0,960
26	22,506	-4,506	20,304
27	35,132	-9,132	83,393
28	27,108	4,892	23,932
29	30,766	-6,766	45,779
30	26,754	1,246	1,553
Итого	872,594	x	712,080

Сумма всех квадратов будет являться остаточным объемом вариации:

$$W_{\text{ост}} = W_e = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 = \sum_{i=1}^n (e_i)^2 = 712,1.$$

Поскольку мы знаем общий объем вариации, то вычитая из него остаточный, получим объясненную регрессией вариацию:

$$W_{\text{регр}} = W_{\text{общ}} - W_{\text{ост}} = 1372,7 - 712,1 = 660,6.$$

Определим число степеней свободы:

$$v_{\text{общ}} = n - 1 = 30 - 1 = 29;$$

$$v_{\text{регр}} = m = 1;$$

$$v_{\text{ост}} = n - m - 1 = 30 - 1 - 1 = 28.$$

А затем, поделив объемы вариации $W_{\text{регр}}$ и $W_{\text{ост}}$ на соответствующие им степени свободы $v_{\text{регр}}$ и $v_{\text{ост}}$, получим несмещенные оценки дисперсии:

$$S_{\text{регр}}^2 = \frac{W_{\text{регр}}}{v_{\text{регр}}} = \frac{660,6}{1} = 660,6;$$

$$S_{\text{ост}}^2 = \frac{W_{\text{ост}}}{v_{\text{ост}}} = \frac{712,1}{28} = 25,4.$$

Результаты оформим в таблицу 3.19.

Таблица 3.19

Данные для определения показателей корреляционной связи

Источник вариации	Объем вариации	Число степеней свободы	Дисперсия
регрессия	660,6	1	660,6
остаток	712,1	28	25,4
общая	1372,7	29	x

Рассчитаем фактическое значение критерия Фишера:

$$F_{\text{факт}} = \frac{S_{\text{регр}}^2}{S_{\text{ост}}^2} = \frac{660,6}{25,4} = 26,0.$$

Теоретическое значение критерия F найдем при уровне значимости 0,05 по таблице 5%-ного уровня распределения F (Приложение 3). Для этого в таблице находим число степеней свободы для регрессионной дисперсии – $\nu_1 = \nu_{\text{рег}} = 1$, и опускаем перпендикуляр до строки, указывающей число степеней свободы для остаточной дисперсии – $\nu_2 = \nu_{\text{ост}} = 28$. На пересечении находим значение 4,20. Таким образом, $F_{\text{теор}}(0,05; 1; 28) = 4,20$

Фактическое значение критерия F (26,0) превысило его теоретическое значение (4,20): $F_{\text{факт}} > F_{\text{теор}}$. Следовательно, принимается альтернативная гипотеза $H_1: \sigma_{\text{рег}}^2 \neq \sigma_{\text{ост}}^2$, и уравнение в целом значимо для генеральной совокупности.

3. Проверим значимость параметров и коэффициента корреляции по t-критерию.

Сформулируем нулевую и альтернативную гипотезы:

$$H_0: \alpha = 0; \beta = 0; \rho = 0,$$

$$H_1: \alpha \neq 0; \beta \neq 0; \rho \neq 0.$$

Выбранный уровень значимости критерия равен 0,05 ($\alpha = 0,05$).

Определим средние ошибки для условного начала (a), коэффициента полной регрессии (b) и коэффициента корреляции (r) по формулам:

$$m_a = \sqrt{S_{\text{ост}}^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} = \sqrt{S_{\text{ост}}^2 \frac{\sum x_i^2}{n(\sigma_x^2 \cdot n)}} = \sqrt{25,4 \frac{364193}{30 \cdot (1578,779 \cdot 30)}} = 2,552;$$

$$m_b = \sqrt{\frac{S_{\text{ост}}^2}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{S_{\text{ост}}^2}{\sigma_x^2 \cdot n}} = \sqrt{\frac{25,4}{1578,779 \cdot 30}} = 0,023;$$

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0,694^2}{30 - 2}} = 0,136.$$

Далее рассчитаем фактические значения критериев:

$$t_a = \frac{a}{m_a} = \frac{16,96}{2,552} = 6,6;$$

$$t_b = \frac{b}{m_b} = \frac{0,118}{0,023} = 5,1;$$

$$t_r = \frac{r}{m_r} = \frac{0,694}{0,136} = 5,1.$$

Критическое значение t найдем по таблице значений критерия t-Стьюдента (Приложение 2). При уровне значимости $\alpha = 0,05$ и остаточном числе степеней свободы $df = \nu_{\text{ост}} = 28$: $t_{\text{крит}}(0,05; 28) = 2,0484$.

Полученные результаты оформим в таблицу 3.20. Как видно из таблицы каждый из параметров уравнения оказался значимым, как и коэффициент корреляции для генеральной совокупности. Результаты корреляционно-регрессионного анализа можно распространять на всю

генеральную совокупность, использовать уравнение в целях прогнозирования.

Таблица 3.20

Оценка значимости параметров уравнения парной линейной регрессии и коэффициента парной корреляции

Параметр	Средняя ошибка, m	$t_{\text{факт}}$	$t_{\text{крит}}$
a	2,552	6,6	2,048
b	0,023	5,1	2,048
r	0,136	5,1	2,048

4. Если параметры уравнения оказались значимыми, то возможна их интерпретация и распространение выводов на генеральную совокупность. Определим предельную ошибку параметров уравнения и коэффициента корреляции:

$$\varepsilon_a = t_{\alpha; v_{\text{ост}}} \cdot m_a = 2,0484 \cdot 2,552 = 5,23;$$

$$\varepsilon_b = t_{\alpha; v_{\text{ост}}} \cdot m_b = 2,0484 \cdot 0,023 = 0,047;$$

$$\varepsilon_r = t_{\alpha; v_{\text{ост}}} \cdot m_r = 2,0484 \cdot 0,136 = 0,279.$$

5. Проведем интервальную оценку параметров уравнения регрессии и коэффициента корреляции. Определим доверительные интервалы для генеральных параметров:

$$a - t_{\alpha; v_{\text{ост}}} \cdot m_a \leq \alpha \leq a + t_{\alpha; v_{\text{ост}}} \cdot m_a;$$

$$b - t_{\alpha; v_{\text{ост}}} \cdot m_b \leq \beta \leq b + t_{\alpha; v_{\text{ост}}} \cdot m_b;$$

$$r - t_{\alpha; v_{\text{ост}}} \cdot m_r \leq \rho \leq r + t_{\alpha; v_{\text{ост}}} \cdot m_r.$$

С 95% уровнем вероятности суждения можно утверждать, что условное начало попадет в интервал от 11,73 до 22,19:

$$16,96 - 5,23 \leq \alpha \leq 16,96 + 5,23;$$

$$11,73 \leq \alpha \leq 22,19.$$

Генеральный коэффициент полной регрессии будет находиться в пределах:

$$0,118 - 0,047 \leq \beta \leq 0,118 + 0,047;$$

$$0,071 \leq \beta \leq 0,165;$$

а коэффициент корреляции:

$$0,694 - 0,279 \leq \rho \leq 0,694 + 0,279;$$

$$0,415 \leq \rho \leq 0,973.$$

Вывод: с уровнем доверительной вероятности 0,95 можно утверждать, что величина коэффициента регрессии, характеризующего связь урожайности с дозой минеральных удобрений, в генеральной совокупности будет находиться в пределе от 0,071 до 0,165, а коэффициент корреляции урожайности с дозой минеральных удобрений в генеральной совокупности находится в пределе от 0,415 до 0,973.

Построим уравнение линейной регрессии с использованием встроенного инструмента «Регрессия» в «Пакете анализа» в MS Excel и проверим правильность проведенных расчетов. В результате получим следующие результаты (рисунок 3.3).

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	0,694					
R-квадрат	0,481					
Нормированный R-квадрат	0,463					
Стандартная ошибка	5,043					
Наблюдения	30					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	1	660,6	660,6	26,0	0,00002	
Остаток	28	712,1	25,4			
Итого	29	1372,7				
	Коэффициенты	Стандар тная ошибка	t- статис тика	P- Значение	Нижние 95%	Верхние 95%
Y-пересечение	16,963	2,553	6,644	0,00000	11,733	22,193
Переменная X 1	0,118	0,023	5,097	0,00002	0,071	0,166

Рисунок 3.3 – Вывод итогов по регрессии

Критические значения F и t можно определить, воспользовавшись встроенными функциями: «=F.ОБР.ПХ (α ; $\nu_{регp}$; $\nu_{ост}$)» и «=СТЮДЕНТ.ОБР.2Х (α ; $\nu_{ост}$)». В результате использования инструмента «Регрессия» получены те же параметры уравнения, ошибки, значения критериев F , t , получены аналогичные интервальные оценки для параметров в генеральной совокупности.

Задача 3.7 Определение показателей связи при множественной линейной корреляции

Условие. по совокупности сельскохозяйственных организаций (СХО) представлены данные наблюдения об урожайности пшеницы, ц/га (y); дозе внесенных минеральных удобрений в пересчете на 100% питательных веществ, кг/га (x_1); бонитете почвенного покрова, балл (x_2) (таблица 3.21).

Решить уравнение и определить показатели тесноты связи урожайности пшеницы с дозой внесенных минеральных удобрений и бонитетом почвенного покрова.

Решение:

1. Исходные данные выпишем в таблицу 3.21.

Таблица 3.21

Исходные и расчетные данные для уравнения связи

№ п/п	Исходные данные			Расчетные данные					
	y	x_1	x_2	y^2	x_1^2	x_2^2	$y x_1$	$y x_2$	$x_1 x_2$
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
1	32	72	60	1024	5184	3600	2304	1920	4320
2	30	145	59	900	21025	3481	4350	1770	8555
3	45	172	106	2025	29584	11236	7740	4770	18232
4	29	119	99	841	14161	9801	3451	2871	11781
5	39	148	90	1521	21904	8100	5772	3510	13320
6	28	103	79	784	10609	6241	2884	2212	8137
7	21	91	63	441	8281	3969	1911	1323	5733
8	17	16	40	289	256	1600	272	680	640
9	35	120	98	1225	14400	9604	4200	3430	11760
10	24	118	48	576	13924	2304	2832	1152	5664
11	30	113	58	900	12769	3364	3390	1740	6554
12	20	28	50	400	784	2500	560	1000	1400
13	28	69	92	784	4761	8464	1932	2576	6348
14	35	107	79	1225	11449	6241	3745	2765	8453
15	31	156	53	961	24336	2809	4836	1643	8268
16	24	50	55	576	2500	3025	1200	1320	2750
17	20	68	44	400	4624	1936	1360	880	2992
18	24	73	58	576	5329	3364	1752	1392	4234
19	25	83	71	625	6889	5041	2075	1775	5893
20	31	102	69	961	10404	4761	3162	2139	7038
21	27	131	40	729	17161	1600	3537	1080	5240
22	37	88	82	1369	7744	6724	3256	3034	7216
23	39	150	100	1521	22500	10000	5850	3900	15000
24	36	104	85	1296	10816	7225	3744	3060	8840
25	38	170	89	1444	28900	7921	6460	3382	15130
26	18	47	39	324	2209	1521	846	702	1833
27	26	154	69	676	23716	4761	4004	1794	10626
28	32	86	85	1024	7396	7225	2752	2720	7310
29	24	117	56	576	13689	3136	2808	1344	6552
30	28	83	73	784	6889	5329	2324	2044	6059
Сум -ма	873	3083	2089	26777	364193	156883	95309	63928	225878
Сред -нее	29	103	70	893	12140	5229	3177	2131	7529

2. Для изучения наличия и направления связи урожайности с факторами эффективно применение результативной группировки по урожайности с рассмотрением средних значений факторных признаков (таблица 3.22).

Таблица 3.22

Урожайность пшеницы и влияющие на нее факторы

Группы СХО по урожайности пшеницы	Число хозяйств	В среднем на 1 СХО		
		урожайность пшеницы, ц с 1 га	доза минеральных удобрений, кг на 1 га	бонитет почвенного покрова, балл
Низшая	11	22	77	54
Средняя	13	30	108	73
Высшая	6	39	139	92
В среднем	х	31	108	73

Данные таблицы 3.22 показывают, что при переходе от низшей группы к высшей возрастание урожайности пшеницы (на 77,3%) сопровождается параллельным повышением всех факторных показателей: доза минеральных удобрений возрастает на 80,5%, бонитет почвенного покрова – на 70,4%.

Для определения формы связи можно сопоставить по каждой группе средние значения результативного и факторных признаков. Постоянство этих соотношений по группам будет свидетельствовать о наличии линейной формы связи (таблица 3.23).

Таблица 3.23

Соотношение урожайности и ее факторов по группам

Группы СХО по урожайности пшеницы	Приходится урожайности пшеницы, кг	
	на 1 кг минеральных удобрений	на 1 балл почвы
Низшая	0,288	0,410
Средняя	0,282	0,419
Высшая	0,281	0,424
В среднем	0,283	0,418

Из таблицы 3.23 видно, что соотношения между урожайностью пшеницы и дозой минеральных удобрений, а также между урожайностью пшеницы и бонитетом почвенного покрова в разных группах в основном стабильны. Это указывает на линейный характер их связи.

3. Перед составлением уравнения связи необходимо исключение возможной коллинеарности факторов. С этой целью рассмотрим матрицу парных коэффициентов корреляции, которая получена в результате анализа данных в MS Excel (инструмент анализа «Корреляция»).

4. Матрица парных коэффициентов для нашего примера свидетельствует об отсутствии коллинеарности факторов – дозы минеральных удобрений и бонитета почвенного покрова, так как $r_{x_1x_2} < 0,6 - 0,7$. А также коэффициент парной корреляции между этими факторами не превышает по своему значению коэффициенты парной корреляции

урожайности с каждым из них ($r_{x_1x_2} < r_{yx_1}$, $r_{x_1x_2} < r_{yx_2}$), что позволяет оба фактора включить в одну модель.

Таблица 3.24

Матрица парных коэффициентов корреляции

Показатель	Урожайность пшеницы	Доза минеральных удобрений	Бонитет почвенного покрова
Урожайность пшеницы	1	0,694	0,793
Доза минеральных удобрений	0,694	1	0,482
Бонитет почвенного покрова	0,793	0,482	1

5. В итоге, уравнение множественной линейной регрессии будет иметь следующий вид:

$$\tilde{y} = a + b_1x_1 + b_2x_2,$$

где y – урожайность пшеницы, ц с 1 га;

x_1 – доза минеральных удобрений в пересчете на 100% питательных веществ, кг на 1 га;

x_2 – бонитет почвенного покрова, балл.

a , b_1 , b_2 – неизвестные параметры уравнения.

6. Для определения параметров уравнения регрессии с использованием МНК составим систему нормальных уравнений с тремя неизвестными:

$$\begin{cases} \sum y = na + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1x_2 \\ \sum yx_2 = a \sum x_2 + b_1 \sum x_1x_2 + b_2 \sum x_2^2 \end{cases}$$

Необходимые для решения системы уравнений суммы, квадраты и взаимные произведения рассчитаны в таблице 3.21.

Из таблицы 3.21 подставим в систему нормальных уравнений соответствующие конкретные величины:

$$\begin{cases} 873 = 30a + 3083b_1 + 2089b_2 \\ 95309 = 3083a + 364193b_1 + 225878b_2 \\ 63928 = 2089a + 225878b_1 + 156883b_2 \end{cases}$$

Оценки параметров уравнения регрессии можно получить с использованием инструмента «Регрессия» в «Пакете анализа» в MS Excel. В результате применения встроенного инструмента получены следующие коэффициенты уравнения связи:

$$a = 7,579149, b_1 = 0,069162, b_2 = 0,206988$$

Запишем уравнение регрессии в решенном виде и дадим интерпретацию полученных параметров:

$$\tilde{y} = 7,579 + 0,069x_1 + 0,207x_2,$$

$a = 7,579$ – условное начало отсчета результативного признака при нулевом значении факторов, для нашего примера – это возможное значение урожайности при нулевых значениях дозы минеральных удобрений и бонитета почвенного покрова, если такие уровни факторов могут быть в производстве;

$b_1 = 0,069$ – коэффициент чистой регрессии при факторе x_1 , показывает, что с увеличением дозы минеральных удобрений в пересчете на 100% питательных веществ на 1 кг/га урожайность пшеницы в среднем увеличивается на 0,069 ц/га при фиксированном (среднем) уровне другого признака, включенного в уравнение (бонитета почвенного покрова);

$b_2 = 0,207$ – коэффициент чистой регрессии при факторе x_2 ; показывает, что с повышением бонитета почвенного покрова на 1 балл урожайность пшеницы в среднем увеличится на 0,207 ц/га при условии, что доза минеральных удобрений зафиксирована на среднем уровне.

7. Определим коэффициенты множественной детерминации и корреляции между урожайностью и двумя факторами, используя матрицу парных коэффициентов корреляции, по следующим формулам:

– коэффициент детерминации:

$$R^2_{yx_1x_2} = \frac{r^2_{yx_1} + r^2_{yx_2} - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}; \quad (3.37)$$

$$R^2_{yx_1x_2} = \frac{0,482 + 0,629 - 2 \cdot 0,694 \cdot 0,793 \cdot 0,482}{1 - 0,232} = 0,755.$$

– коэффициент корреляции:

$$R_{yx_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}}; \quad (3.38)$$

$$R_{yx_1x_2} = \sqrt{\frac{0,482 + 0,629 - 2 \cdot 0,694 \cdot 0,793 \cdot 0,482}{1 - 0,232}} = 0,869.$$

Величина коэффициента корреляции $R_{yx_1x_2} = 0,869$ указывает на то, что связь между признаками тесная. Коэффициент множественной детерминации $R^2_{yx_1x_2} = 0,755$ показывает, что факторы, включенные в линейное уравнение связи, объясняют 75,5% вариации урожайности в исследуемой совокупности хозяйств.

7. Для сравнения оценок роли различных факторов в формировании результативного признака рассчитаем коэффициенты эластичности, β -

коэффициенты и коэффициенты отдельного определения. Необходимые данные возьмем из таблицы 3.21.

а) Коэффициенты эластичности определяются по следующим формулам:

$$\Theta_1 = b_1 \cdot \frac{\bar{x}_1}{\bar{y}}; \Theta_2 = b_2 \cdot \frac{\bar{x}_2}{\bar{y}}. \quad (3.39)$$

– для x_1 (дозы минеральных удобрений):

$$\Theta_1 = b_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 0,069162 \cdot \frac{103}{29} = 0,244.$$

– для x_2 (бонитета почвенного покрова)

$$\Theta_2 = b_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 0,206988 \cdot \frac{70}{29} = 0,495.$$

Коэффициенты эластичности показывают, что при увеличении дозы минеральных удобрений на 1% урожайность пшеницы в среднем увеличится на 0,244%, а при повышении бонитета почвенного покрова на 1% урожайность в среднем увеличится на 0,495%.

б) Определим β -коэффициенты по формулам:

$$\beta_1 = b_1 \cdot \frac{\sigma_{x_1}}{\sigma_y}; \beta_2 = b_2 \cdot \frac{\sigma_{x_2}}{\sigma_y} \quad (3.40)$$

Сначала рассчитаем средние квадратические отклонения для y , x_1 , и x_2 :

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2} = \sqrt{\frac{26777}{30} - \left(\frac{873}{30}\right)^2} = 6,764;$$

$$\sigma_{x_1} = \sqrt{\frac{\sum x_1^2}{n} - \left(\frac{\sum x_1}{n}\right)^2} = \sqrt{\frac{364193}{30} - \left(\frac{3083}{30}\right)^2} = 39,734;$$

$$\sigma_{x_2} = \sqrt{\frac{\sum x_2^2}{n} - \left(\frac{\sum x_2}{n}\right)^2} = \sqrt{\frac{156883}{30} - \left(\frac{2089}{30}\right)^2} = 19,510.$$

Итак,

$$\beta_1 = b_1 \cdot \frac{\sigma_{x_1}}{\sigma_y} = 0,069162 \cdot \frac{39,734}{6,764} = 0,406;$$

$$\beta_2 = b_2 \cdot \frac{\sigma_{x_2}}{\sigma_y} = 0,206988 \cdot \frac{19,510}{6,764} = 0,597.$$

β -коэффициенты показывают, что если доза минеральных удобрений увеличится на величину своего среднеквадратического отклонения (на σ_1), то урожайность увеличится пшеницы в среднем на $0,406\sigma_y$. Изменение бонитета почвенного покрова на σ_2 приведет к изменению урожайности на $0,597\sigma_y$.

в) Коэффициенты отдельного определения рассчитаем по формулам:

$$d_1^2 = r_{yx_1} \cdot \beta_1; d_2^2 = r_{yx_2} \cdot \beta_2. \quad (3.41)$$

Коэффициенты отдельного определения равны:

$$d_1^2 = r_{yx_1} \cdot \beta_1 = 0,694 \cdot 0,406 = 0,282;$$

$$d_2^2 = r_{yx_2} \cdot \beta_2 = 0,793 \cdot 0,597 = 0,473.$$

Проверим выполнение равенства:

$$d_1^2 + d_2^2 = R^2_{yx_1x_2};$$

$$0,282 + 0,473 = 0,755.$$

Сделаем вывод. Коэффициенты отдельного определения показывают долю фактора в воспроизведенной вариации. В нашем случае из 75,5% воспроизведенной уравнением вариации урожайности пшеницы 28,2% приходится на долю дозы минеральных удобрений, и 47,3% – на долю бонитета почвенного покрова.

Результаты сравнительного анализа влияния факторов оформим в виде таблицы (таблица 3.25).

Таблица 3.25

Влияние факторов на урожайность культуры

Факторы	Значения коэффициентов		
	эластичности ε_i	бета- коэффициент в β_i	коэффициентов отдельного определения d_i^2
доза минеральных удобрений – x_1	0,244	0,406	0,282
бонитет почвенного покрова – x_2	0,495	0,597	0,473

Подведем итог. По всем определенным коэффициентам подтверждается приоритетность фактора x_2 . Следовательно, наибольшие возможности в изменении урожайности пшеницы связаны с изменением бонитета почвенного покрова.

Вопросы для проверки знаний по разделу III

1. Какие задачи можно решать с помощью дисперсионного анализа?
2. Какой статистический критерий применяется в дисперсионном анализе и каково его содержание?
3. Каково содержание нулевой гипотезы при дисперсионном анализе?
4. Какие модели дисперсионного анализа можно формировать?
5. Какие дисперсии используются для расчета фактического значения критерия F-Фишера?
6. От чего зависит табличное значение критерия F-Фишера?

7. В Вашей задаче оказалось, что $F_{\text{факт}} = F_{\text{табл}}$ (при $\alpha = 0,05$). Какая гипотеза при этом принимается? Какая гипотеза будет принята при $\alpha = 0,01$?
8. Если в ходе расчетов оказалось, что $S_{\text{факт}}^2 < S_{\text{ост}}^2$, то каков порядок Ваших дальнейших действий?
9. Определите фактическое значение критерия F-Фишера, если $W_{\text{общ}} = 100$, $W_{\text{ост}} = 20$. Число групп (вариантов) $m=5$, число повторностей $n=4$.
10. Определите межгрупповую дисперсию и фактическое значение критерия F-Фишера при условии, что общий объем вариации $W_{\text{общ}} = 50$, объем остаточной (внутригрупповой) $W_{\text{ост}} = 10$. Число вариантов (групп) $m=5$, число повторностей $n=3$.
11. Определите число повторностей (наблюдений) в группах, если при $\alpha = 0,05$, $F_{\text{табл}} = 3,86$. Общее число наблюдений $N=12$. Число наблюдений по группам одинаково.
12. Во сколько раз $S_{\text{межгр}}^2$ должна быть больше $S_{\text{внгр}}^2$, чтобы принять альтернативную гипотезу, при условии, что число групп (вариантов) $m=3$, число повторностей $n=3$. Формирование групп независимое.
13. Определить общий объем вариации, если $S_{\text{межгр}}^2 = 10$, $S_{\text{внгр}}^2 = 5$. Число групп (вариантов) $m=5$, число повторностей $n=3$. Формирование групп независимое.
14. Определить $F_{\text{табл}}$ ($\alpha = 0,05$), если общее число наблюдений $N=30$. Число групп (вариантов) $m=5$. Группы имеют одинаковое число наблюдений. Формирование групп зависимое.
15. Может ли в дисперсионном анализе $F_{\text{факт}}$ быть меньше единицы и почему?
16. Каковы этапы построения корреляционного уравнения связи?
17. Как выбрать вид уравнения связи?
18. Каким методом определяются коэффициенты уравнения?
19. Каково содержание коэффициентов уравнения?
20. В чем отличие коэффициентов полной и чистой регрессии?
21. Какие существуют стандартизированные коэффициенты регрессии?
22. Каков алгоритм расчета стандартизированных коэффициентов регрессии?
23. Каким образом проводится оценка существенности показателей связи между признаками в корреляционно-регрессионном анализе?
24. Проанализируйте уравнение связи удоя коров в центнерах на 1 корову (y) и расхода кормов на 1 корову (x): $y = -20,07 + 0,815x$

ЛИТЕРАТУРА

1. Зинченко, А.П. Математическая статистика: учебник / А.П. Зинченко, М.В. Кагирова, Ю.Н. Романцева, О.Б. Тарасова, А.В. Уколова и др.; под ред. А.П. Зинченко и А.В. Уколовой. – М.: Издательство РГАУ-МСХА, 2018. – 199 с.
2. Дмитриев, Е.А. Математическая статистика в почвоведении: учебник / Е.А. Дмитриев. – М.: Книжный дом «ЛИБРОКОМ», 2023. – 334 с.
3. Самсонова, В.П. Математическая статистика в почвоведении и экологии. Практикум: учебное пособие / В.П. Самсонова, С.Е. Дядькина, Ю.Л. Мешалкина. – Москва: Издательско-торговая корпорация «Дашков и К°», 2021. – 154 с.
4. Кремер, Н. Ш. Математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. — Москва : Издательство Юрайт, 2024. — 259 с. — (Высшее образование). — ISBN 978-5-534-01654-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536959>.
5. Математическая статистика: практикум / О. Б. Тарасова [и др.]; Российский государственный аграрный университет - МСХА имени К. А. Тимирязева (Москва). — Электрон. текстовые дан. — Москва: РГАУ-МСХА им. К. А. Тимирязева, 2018 — 91 с. — Коллекция: Учебная и учебно-методическая литература.
6. Бурнаева, Э.Г. Обработка и представление данных в MS Excel : учебное пособие / Э.Г. Бурнаева, С.Н. Леора. — 2-е изд., стер. — Санкт-Петербург : Лань, 2018. — 156 с. — ISBN 978-5-8114-1923-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://reader.lanbook.com/book/108304#1>. — Режим доступа: для авториз. пользователей.
7. Трушков, А.С. Статистическая обработка информации. Основы теории и компьютерный практикум + CD : учебное пособие / А.С. Трушков. — Санкт-Петербург : Лань, 2020. — 152 с. — ISBN 978-5-8114-4322-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://reader.lanbook.com/book/126947#1> (дата обращения: 09.05.2024). — Режим доступа: для авториз. пользователей.
8. Малугин, В.А. Математическая статистика: учебное пособие для вузов / В.А. Малугин. – Москва: Издательство Юрайт, 2021. – 218 с. – (Высшее образование). – Текст: электронный. — URL: <https://urait.ru/viewer/matematiceskaya-statistika-473498#page/2>.
9. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel : учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2021. — 353 с. — (Высшее образование). — ISBN 978-5-534-01672-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/viewer/statistika-raschety-v-microsoft-excel-471895#page/1>

ПРИЛОЖЕНИЯ

Приложение 1

Значения интеграла вероятностей $F(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt$ при разных значениях t

t	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0718
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4909	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	6817	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8788	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9089
1,7	9108	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9425	9438	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9608	9615	9625	9634
2,1	9643	9652	9660	9669	9676	9684	9692	9700	9707	9715

Продолжение приложения 1

t	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
2,2	9722	9729	9736	9743	9749	9755	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9932
2,4	9836	9840	9845	9849	9853	9857	9861	9866	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9924	9926	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	99730	99739	99747	99755	99763	99771	99779	99786	99793	99800
3,1	99807	99813	99819	99825	99831	99837	99842	99847	99853	99858
3,2	99863	99867	99872	99876	99880	99884	99888	99892	99896	99900
3,3	99903	3,6	99911	3,9	999904	4,4	9999892	5,0		99999943
3,4	99933	3,7	99937	4,0	999937	4,6	9999957	5,0		99999996
3,5	99953	3,8	99957	4,2	999973	4,8	9999984	6,0		99999999

Некоторые специальные значения t для наиболее часто встречающихся в расчетах вероятностей p :

P	0,5000	0,7500	0,9000	0,9500	0,9900	0,9950	0,9990
t	0,6745	1,1504	1,6449	1,9600	2,5759	2,8771	3,2901

*

Приложение 2

Значение двустороннего критерия t -Стьюдента при уровне значимости 0,10; 0,05 и 0,01

$d.f.$	Уровень значимости λ			$d.f.$	Уровень значимости λ		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1718	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Приложение 3

Таблица 5%-ного уровня распределения F

	ν_1 – степени свободы для большего среднего квадрата								
ν_2	1	2	3	4	5	6	7	8	9
1	161	200	216	225	230	234	237	239	241
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38
3	10,13	9,55	9,28	9,19	9,01	8,94	8,88	8,84	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,26
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21
32	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19
34	4,13	3,28	2,88	2,65	2,49	2,38	2,30	2,23	2,17
36	4,11	3,26	2,86	2,63	2,48	2,36	2,28	2,21	2,15
38	4,10	3,25	2,85	2,62	2,46	2,35	2,26	2,19	2,14
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,14	2,09
48	4,04	3,19	2,80	2,56	2,41	2,30	2,21	2,14	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07
55	4,02	3,17	2,78	2,54	2,38	2,27	2,18	2,11	2,05
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,02
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01

Продолжение приложения 3

	ν_1 – степени свободы для большего среднего квадрата								
ν_2	1	2	3	4	5	6	7	8	9
80	3,86	3,11	2,72	2,48	2,33	2,21	2,12	2,05	1,99
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,95
150	3,91	3,06	2,67	2,43	2,27	2,16	2,07	2,00	1,94
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90
1000	3,85	3,00	2,61	2,38	2,22	2,10	2,02	1,95	1,89
	ν_1 – степени свободы для большего среднего квадрата								
ν_2	10	11	12	14	16	20	30	40	50
1	242	243	244	245	246	248	250	251	252
2	19,39	19,40	19,41	19,42	19,43	19,44	19,46	19,47	19,47
3	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,60	8,58
4	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,71	5,70
5	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,46	4,44
6	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,77	3,75
7	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,34	3,32
8	3,34	3,31	3,28	3,23	3,20	3,15	3,08	3,05	3,03
9	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,82	2,80
10	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,67	2,64
11	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,53	2,50
12	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,42	2,40
13	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,34	2,32
14	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,27	2,24
15	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,21	2,18
16	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,16	2,13
17	2,45	2,41	2,38	2,33	2,29	2,23	2,15	2,11	2,08
18	2,41	2,37	2,34	2,29	2,25	2,19	2,11	2,07	2,04
19	2,38	2,34	2,31	2,26	2,21	2,15	2,07	2,02	2,00
20	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,99	1,96
21	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,96	1,93
22	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,93	1,91
23	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,91	1,88
24	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,89	1,86
25	2,24	2,20	2,16	2,11	2,06	2,00	1,92	1,87	1,84
26	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,85	1,82
27	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,84	1,80
28	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,81	1,78
29	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,80	1,77
30	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,79	1,76
32	2,14	2,10	2,07	2,02	1,97	1,91	1,82	1,76	1,74
34	2,12	2,08	2,05	2,00	1,95	1,89	1,80	1,74	1,71
36	2,10	2,06	2,03	1,98	1,93	1,87	1,78	1,72	1,69
38	2,09	2,05	2,02	1,96	1,92	1,85	1,76	1,71	1,67

Продолжение приложения 3

	ν_1 – степени свободы для большего среднего квадрата								
ν_2	10	11	12	14	16	20	30	40	50
40	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,69	1,66
42	2,06	2,02	1,99	1,94	1,89	1,82	1,73	1,68	1,64
44	2,05	2,01	1,98	1,92	1,88	1,81	1,72	1,66	1,63
46	2,04	2,00	1,97	1,91	1,87	1,80	1,71	1,65	1,62
48	2,03	1,99	1,96	1,90	1,86	1,79	1,70	1,64	1,61
50	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,63	1,60
55	2,00	1,97	1,93	1,88	1,83	1,76	1,67	1,61	1,58
60	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,59	1,56
65	1,98	1,94	1,90	1,85	1,80	1,73	1,63	1,57	1,54
70	1,97	1,93	1,89	1,84	1,79	1,72	1,62	1,56	1,53
80	1,95	1,91	1,88	1,82	1,77	1,70	1,60	1,54	1,51
100	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,51	1,48
125	1,90	1,86	1,83	1,77	1,72	1,65	1,55	1,49	1,45
150	1,89	1,85	1,82	1,76	1,71	1,64	1,54	1,47	1,44
200	1,87	1,83	1,80	1,74	1,69	1,62	1,52	1,45	1,42
400	1,85	1,81	1,78	1,72	1,67	1,60	1,49	1,42	1,38
1000	1,84	1,80	1,76	1,70	1,65	1,58	1,47	1,41	1,36

Приложение 4

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

Целые и десятичные доли	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,940	0925	0909	0893	0878	0863	0848	0833	0818	0804

Продолжение приложения 4

Целые и десятичные доли	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
1,8	0790	0775	0761	0748	0734	0721	0707	0695	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0022	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0016	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	0,0001	0,0001	0,0001	0,0001						
4,1	0,0001338									

Приложение 5

Критическое значение критерия χ^2 -Пирсона

<i>d. f.</i>	Вероятности значения χ^2 , превышающего табличное			
	$\lambda=0,10$	$\lambda=0,05$	$\lambda=0,025$	$\lambda=0,01$
<i>l</i>	2	3	4	5
1	2,71	3,84	5,02	6,63
2	4,61	5,99	7,38	9,21
3	6,25	7,81	9,35	11,34
4	7,78	9,49	11,14	13,28
5	9,24	11,07	12,83	15,09
6	10,64	12,59	14,45	16,81
7	12,02	14,07	16,01	18,48
8	13,36	15,51	17,53	20,09
9	14,68	16,92	19,02	21,67
10	15,99	18,31	20,48	23,21
11	17,28	19,68	21,92	24,72
12	18,55	21,03	23,34	26,22
13	19,81	22,36	24,74	27,69
14	21,06	23,68	26,12	29,14
15	22,31	25,00	27,49	30,58
16	23,54	26,30	28,85	32,00
17	24,77	27,59	30,19	33,41
18	25,99	28,87	31,53	34,81
19	27,20	30,14	32,85	36,19
20	28,41	31,41	34,17	37,57
21	29,62	32,67	35,48	38,93
22	30,81	33,92	36,78	40,29
23	32,01	35,17	38,08	41,64
24	33,20	36,42	39,36	42,98
25	34,38	37,65	40,65	44,31
26	35,56	38,89	41,92	45,64
27	36,74	40,11	43,19	46,96
28	37,92	41,34	44,46	48,28
29	39,09	42,56	45,72	49,59
30	40,26	43,77	46,98	50,89
40	51,80	55,76	59,34	63,69
50	63,17	67,50	71,42	76,15
60	74,40	79,08	83,30	88,38
70	85,53	90,53	95,02	100,42
80	96,58	101,88	106,63	112,63
90	107,56	103,14	118,14	124,12
100	118,50	124,34	129,56	135,81

Таблица значений e^{-x}

x	e^{-x}	x	e^{-x}	x	e^{-x}	x	e^{-x}
1	2	1	2	1	2	1	2
0,00	1,0000	0,24	0,7866	0,48	0,6188	0,72	0,4868
01	0,9900	25	0,7788	49	0,6126	73	0,4819
02	0,9802	26	0,7711	0,50	0,6065	74	0,4771
03	0,9704	27	0,7634	51	0,6005	75	0,4724
04	0,9608	28	0,7558	52	0,5945	76	0,4677
05	0,9512	29	0,7483	53	0,5886	77	0,4630
06	0,9418	0,30	0,7408	54	0,5827	78	0,4584
07	0,9324	31	0,7334	55	0,5769	0,7854	0,4559
08	0,9231	32	0,7261	56	0,5712	79	0,4538
09	0,9139	33	0,7189	57	0,5655	0,80	0,4493
0,10	0,9048	34	0,7118	58	0,5599	81	0,4449
11	0,8958	35	0,7047	59	0,5543	82	0,4404
12	0,8869	36	0,6977	0,60	0,5488	83	0,4360
13	0,8781	37	0,6907	61	0,5434	84	0,4317
14	0,8694	38	0,6839	62	0,5379	85	0,4274
15	0,8607	39	0,6771	63	0,5326	86	0,4232
16	0,8521	0,40	0,6703	64	0,5273	87	0,4190
17	0,8437	41	0,6637	65	0,5220	88	0,4148
18	0,8353	42	0,6570	66	0,5169	89	0,4107
19	0,8270	43	0,6505	67	0,5117	0,90	0,4066
0,20	0,8187	44	0,6440	68	0,5066	91	0,4025
21	0,8106	45	0,6376	69	0,5016	92	0,3985
22	0,8025	46	0,6313	0,70	0,4966	93	0,3946
23	0,7945	47	0,6250	71	0,4916	94	0,3906
0,95	0,3867	1,21	0,2982	1,47	0,2299	0,72	0,1791
96	0,3829	22	0,2952	48	0,2276	73	0,1773
97	0,3791	23	0,2923	49	0,2254	74	0,1755
98	0,3753	24	0,2894	1,50	0,2231	75	0,1738
99	0,3716	25	0,2865	51	0,2209	76	0,1720
1,00	0,3679	26	0,2837	52	0,2187	77	0,1703
01	0,3642	27	0,2808	53	0,2165	78	0,1686
02	0,3606	28	0,2780	54	0,2144	79	0,1670
03	0,3570	29	0,2753	55	0,2122	1,80	0,1653
04	0,3535	1,30	0,2725	56	0,2101	81	0,1637
05	0,3499	31	0,2698	57	0,2080	82	0,1620
06	0,3465	32	0,2671	1,5708	0,2070	83	0,1604
07	0,3430	33	0,2645	58	0,2060	84	0,1588
08	0,3396	34	0,2618	59	0,2039	85	0,1572
09	0,3362	35	0,2592	1,60	0,2019	86	0,1557

Продолжение приложения 6

1	2	1	2	1	2	1	2
1,10	0,3329	36	0,2567	61	0,0199	87	0,1541
11	0,3296	37	0,2541	62	0,1979	88	0,1526
12	0,3263	38	0,2516	63	0,1959	89	0,1511
13	0,3230	39	0,2491	64	0,1940	1,90	0,1496
14	0,3198	1,40	0,2466	65	0,1920	91	0,1481
15	0,3166	41	0,2441	66	0,1901	92	0,1466
16	0,3135	42	0,2417	67	0,1882	93	0,1451
17	0,3104	43	0,2393	68	0,1864	94	0,1437
18	0,3073	44	0,2369	69	0,1845	95	0,1423
19	0,3042	45	0,2346	1,70	0,1827	96	0,1409
1,20	0,3012	46	0,2322	71	0,1809	97	0,1395

Стюдентизированный размах ($p=1-\lambda=0,95$).

v	K								
	2	3	4	5	6	7	8	9	10
1	17,97	26,98	32,82	37,08	40,41	43,12	45,40	47,36	49,07
2	6,085	8,331	9,798	10,88	11,74	12,44	13,03	13,54	13,99
3	4,501	5,910	6,825	7,502	8,037	8,478	8,853	9,177	9,462
4	3,927	5,040	5,757	6,287	6,707	7,053	7,347	7,602	7,826
5	3,635	4,602	5,218	5,673	6,033	6,330	6,582	6,802	6,995
6	3,461	4,339	4,896	5,305	5,628	5,895	6,122	6,319	6,493
7	3,344	4,165	4,681	5,060	5,359	5,606	5,815	5,998	6,158
8	3,261	4,041	4,529	4,886	5,167	5,399	5,597	5,767	5,918
9	3,199	3,949	4,415	4,756	5,024	5,244	5,432	5,595	5,739
10	3,151	3,877	4,327	4,654	4,912	5,124	5,305	5,461	5,599
11	3,113	3,820	4,256	4,574	4,823	5,028	5,202	5,353	5,487
12	3,082	3,773	4,199	4,508	4,751	4,950	5,119	5,265	5,395
13	3,055	3,735	4,151	4,453	4,690	4,885	5,049	5,192	5,318
14	3,033	3,702	4,111	4,407	4,639	4,829	4,990	5,131	5,254
15	3,014	3,674	4,076	4,367	4,595	4,782	4,940	5,077	5,198
16	2,998	3,649	4,046	4,333	4,557	4,741	4,897	5,031	5,150
17	2,984	3,628	4,020	4,303	4,524	4,705	4,858	4,991	5,108
18	2,971	3,609	3,997	4,277	4,495	4,673	4,824	4,956	5,071
19	2,960	3,593	3,977	4,253	4,469	4,645	4,794	4,924	5,038
20	2,950	3,578	3,958	4,232	4,445	4,620	4,768	4,896	5,008
21	2,919	3,532	3,901	4,166	4,373	4,541	4,684	4,807	4,915
30	2,888	3,486	3,845	4,102	4,302	4,464	4,602	4,720	4,824
40	2,858	3,442	3,791	4,039	4,232	4,389	4,521	4,635	4,735
60	2,829	3,399	3,737	3,977	4,163	4,314	4,441	4,550	4,646
120	2,800	3,356	3,685	3,917	4,096	4,241	4,363	4,468	4,560
∞	2,772	3,314	3,633	3,858	4,030	4,170	4,286	4,387	4,474

Продолжение приложения 7

v	K									
	11	12	13	14	15	16	17	18	19	20
1	50,59	51,96	53,20	54,33	55,36	56,32	57,22	58,04	58,83	59,96
2	14,39	14,75	15,08	15,38	15,65	15,91	16,14	16,37	16,57	16,77
3	9,717	9,946	10,15	10,35	10,53	10,69	10,84	10,98	11,11	11,24
4	8,027	8,208	8,373	8,525	8,664	8,794	8,914	9,028	9,134	9,233
5	7,168	7,324	7,466	7,596	7,717	7,828	7,932	8,030	8,122	8,208
6	6,649	6,789	6,917	7,034	7,143	7,244	7,338	7,426	7,508	7,587
7	6,302	6,431	6,550	6,658	6,759	6,852	6,939	7,020	7,097	7,170
8	6,054	6,175	6,287	6,389	6,483	6,571	6,653	6,729	6,802	6,870
9	5,867	5,983	6,089	6,186	6,276	6,359	6,437	6,510	6,579	6,644
10	5,722	5,833	5,935	6,028	6,114	6,194	6,269	6,339	6,405	6,467
11	5,605	5,713	5,811	5,901	5,984	6,062	6,134	6,202	6,265	6,326
12	5,511	5,615	5,710	5,798	5,878	5,953	6,023	6,089	6,151	6,209
13	5,431	5,533	5,625	5,711	5,789	5,862	5,931	5,995	6,055	6,112
14	5,364	5,463	5,554	5,637	5,714	5,786	5,852	5,915	5,974	6,029
15	5,306	5,404	5,493	5,574	5,649	5,720	5,785	5,846	5,904	5,958
16	5,256	5,352	5,439	5,520	5,593	5,662	5,727	5,786	5,843	5,897
17	5,212	5,307	5,392	5,471	5,544	5,612	5,675	5,734	5,790	5,842
18	5,174	5,267	5,352	5,429	5,501	5,568	5,630	5,688	5,743	5,794
19	5,140	5,231	5,315	5,391	5,462	5,528	5,589	5,647	5,701	5,752
20	5,108	6,199	5,282	5,357	5,427	5,493	5,553	5,610	5,663	5,714
30	4,917	5,001	5,077	5,147	5,211	5,271	5,327	5,379	5,429	5,475
40	4,824	4,904	4,977	5,044	5,106	5,163	5,216	5,266	5,313	5,358
60	4,732	4,808	4,878	4,942	5,001	5,056	5,107	5,154	5,199	5,241
120	4,641	4,714	4,781	4,842	4,898	4,950	4,998	5,044	5,086	5,126
∞	4,552	4,622	4,685	4,743	4,796	4,845	4,891	4,934	4,974	5,012

Учебное издание

*Тарасова Ольга Борисовна
Дашиева Баярма Шагдаровна
Козлов Кирилл Александрович
Анохин Игорь Александрович
Ветошкин Артем Юрьевич*

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Сдано в набор 01.06.2025
Формат 60×88/16.
Усл.печ.л. 8,2

Подп. в печ. 01.07.2025
Бумага офсетная.
Тираж 500 экз.

Издательство «Научный консультант» предлагает авторам:
издание рецензируемых сборников трудов научных
конференций; печать монографий, методической и иной литературы

ISBN 978-5-907933-41-5



9 785907 933415 >

*Издательство Научный консультант
123007, г. Москва, Хорошевское ш., 35к2, офис 508.
Тел.: +7 (926) 609-32-93, +7 (499) 195-60-77 www.n-ko.ru keyneslab@gmail.com*