

ПРИДНЕСТРОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. Т.Г.Шевченко

Аграрно-технологический факультет

Кафедра плодовоовощеводства и виноградарства

***ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ
В АГРОНОМИИ***

Методические указания
по выполнению лабораторных работ по курсу
для студентов заочного отделения специальностей
«Агрономия», «Плодовоовощеводство и виноградарство»,
«Защита растений», «Технология производства и
переработки сельскохозяйственной продукции»

Тирасполь, 2010 г

УДК 630 (478.9)
ББК 4с (2М,2)
М54

Составители:

Н.Н. Трескина – кан. с.-х. наук
Е.М. Коваль – канд. с.-х. наук

Рецензенты:

Т.Н.Звездина, канд. с.-х. наук, доцент кафедры **зоологии**
Н.И.Шульман, канд. биол. наук, доцент кафедры защиты растений и экологии

Основы научных исследований в агрономии: Методические
Г 49 указания по выполнению лабораторных работ по курсу для студентов
заочного отделения специальностей «Агрономия»,
«Плодоовощеводство и виноградарство», «Защита растений»,
«Технология производства и переработки сельскохозяйственной
продукции» /Сост.: Н.Н.Трескина, Е.М.Коваль, - Тирасполь, 2010. –
50 с.

*Методические указания составлены в соответствии с учебным
планом и программой по курсу «Основы научных исследований в агрономии».*

УДК 630 (478.9)
ББК 4с (2М,2)

Рекомендовано научно-методическим советом ПГУ им. Т.Г.Шевченко

©ПГУ им. Т.Г.Шевченко, 2010
© Составители:
Н.Н. Трескина, Е.М. Коваль, 2010

Введение

Агрономическая наука занимается разработкой агротехнических приемов повышения продуктивности сельскохозяйственных культур и улучшения качества урожая, созданием новых форм, сортов и гибридов культурных растений, устойчивых к действию неблагоприятных факторов внешней среды. При планировании научной работы и обработке результатов опытов используются различные статистические методы. В настоящее время без статистических методов невозможно дать научную оценку результатам исследований, сделать достоверные обоснованные выводы.

В процессе выполнения лабораторных работ студенты должны освоить основные методы математической обработки полученных данных, научиться делать научно-обоснованные выводы.

Лабораторная работа №1 **СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПРИ КОЛИЧЕСТВЕННОЙ ИЗМЕНЧИВОСТИ**

В научной агрономии используются общепринятые приемы научных исследований: наблюдение и эксперимент. **Наблюдение** – это количественная или качественная регистрация интересующих исследователя сторон развития явления, констатация наличия того или иного признака. При проведении наблюдения исследователь только наблюдает за развитием биологических объектов. В **эксперименте (опыте)** исследователь искусственно вызывает явления или изменяет условия таким образом, чтобы лучше выяснить сущность явления. Эксперимент является ведущим приемом исследования. Наблюдение и эксперимент в исследовательской работе наиболее часто используются в едином комплексе.

На результаты исследований всегда оказывают влияние внешние условия (микроклимат фитоценоза, микрорельеф земельного участка, неоднородность почвенного покрова), которые не поддаются учету и создают определенные сложности при проведении статистической обработки результатов наблюдений или экспериментов. Под влиянием внешних факторов даже генетически одинаковые особи отличаются друг от друга. Свойство особей (растений, насекомых, животных) отличаться друг от друга по какому либо признаку в однородной совокупности называют **изменчивостью, или варьированием**. Различают количественную и качественную изменчивость. **Количественная изменчивость** может быть измерена и имеет числовое выражение (количество плодов, стеблей, цветков, длина стебля, масса плода, химический состав плодов); **качественная изменчивость** не поддается измерению (окраска и форма плодов, листьев).

При проведении наблюдения над количественным варьирующим признаком, который обычно обозначается символом X , мы получаем ряд значений $X_1, X_2, X_3, \dots, X_n$. Каждое из этих значений называется *варианта*, или *дата*.

Для оценки количественной изменчивости используют следующие статистические показатели:

1. **Средняя арифметическая** представляет собой обобщенную абстрактную характеристику выборки и обозначается символом $\bar{x}$. Для выборки малого объема определяется **простая средняя арифметическая** по формуле:

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

где X – значение варьирующего признака,
 n – общее число значений признака.

Для выборок большого объема (более 30 значений признака) рассчитывается **взвешенная средняя арифметическая**:

$$\bar{x} = \frac{f_1 \cdot X_1 + f_2 \cdot X_2 + \dots + f_n \cdot X_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum f \cdot X}{n}$$

где f – частота встречаемости значения признака.

Для расчета взвешенной средней арифметической рассчитывается вариационный ряд. **Вариационным рядом** называется такой ряд данных, в котором указаны возможные значения варьирующего признака X в порядке возрастания и соответствующие им частоты f .

2. **Дисперсия** s^2 показывает квадрат среднего отклонения значения изучаемого признака X от средней арифметической $\bar{x}$.

3. **Стандартное отклонение** s показывает величину среднего отклонения значения признака X от средней арифметической $\bar{x}$.

Дисперсия и стандартное отклонение являются основными мерами вариации признака.

4. **Коэффициент вариации** V – относительный показатель изменчивости, показывающий степень варьирования признака. Если коэффициент вариации не превышает 10% изменчивость принято считать незначительной, если коэффициент вариации более 20% – значительной.

5. **Ошибка выборочной средней** s_x является мерой отклонения выборочной средней от средней генеральной совокупности. Она возникает вследствие неполной репрезентативности (представительности выборки) выборки, т.е. не все значения, которые признак принимает в генеральной совокупности, попадают в выборку. Ошибка выборочной средней прямо пропорциональна размаху варьирования признака и обратно

пропорциональна объему выборки. Чем больше величина стандартного отклонения, т.е. варьирование признака, тем больше величина ошибки. Чем больше объем выборки, т.е. чем больше возможных значений варьирующего признака попало в выборку, тем меньше ошибка выборочной средней.

Пример. Масса плодов перца сорта Золотой Юбилей, г.

Целью данного наблюдения является определение выравненности признака «масса плодов». Для достижения этой цели следует провести статистический анализ полученных данных.

92	96	86	102	93	77	98	95	83	90
97	82	104	94	99	95	86	105	91	110
93	89	103	75	91	87	100	83	88	97
99	105	93	96	84	98	92	105	107	94
85	87	104	99	91	103	79	93	88	100

При проведении наблюдения получен достаточно большой объем данных: $n=50$. При большом объеме данных (более 30) проведение математической обработки затруднено, поэтому их необходимо сгруппировать и построить вариационный ряд. Для построения вариационного ряда полученные значения признака X следует разбить на ряд интервалов. Для определения оптимального количества интервалов k , на которые будет разбита выборка, из общего числа наблюдений n извлекается квадратный корень. Если полученное значение не целое, его следует округлить до целого, т.к. количество интервалов не может быть выражено дробным числом. В нашем случае

$$k = \sqrt{n} = 7$$

Интервалы, кроме первого и последнего, должны быть одинаковых размеров. Величина первого и последнего интервалов может быть немного больше или меньше в связи с тем, что в них попадает меньше значений признака. Для успешной разбивки выборки на группы рассчитывают шаг интервала i . Для этого находят самое большое $X_{\max}$ и самое маленькое $X_{\min}$ значение признака, находят между ними разность, которую делят на количество интервалов k . Полученное частное округляем до того же уровня точности, которое мы использовали для записи значений признака. Т.е. если значения признака даны в целых числах, округляем до целых; если значения признака даны с точностью до десятых, то округляем до десятых и т.д. В нашем случае

$$i = \frac{X_{\max} - X_{\min}}{k} = \frac{110 - 75}{7} = 5$$

Разбиваем весь массив данных на интервалы. Интервал представляет собой ряд данных, для более краткой записи интервала используют только первое и последнее число (нижняя и верхняя границы) между которыми ставят знак интервала $\div$. Нижней границей первого

интервала является **Xmin**, для определения верхней границы первого интервала к **Xmin** прибавляют шаг интервала **i**. В нашем случае первый интервал имеет следующий вид **75 ÷ 80**. Второй интервал начинается с числа, идущего после верхней границы интервала, т.е. нижней границей второго интервала будет число **81**. Для определения верхней границы второго интервала к нижней границе мы прибавляем «**i - единица точности**». В нашем примере значения признака даны в целых, т.е. единицей точности является 1. Таким образом, верхнюю границу второго интервала мы определяем как **81+(5-1) = 81+4=85**. Этот же алгоритм мы используем для определения последующих интервалов.

Мы получили следующие интервалы:

- I. **75 ÷ 80**
- II. **81 ÷ 85**
- III. **86 ÷ 90**
- IV. **91 ÷ 95**
- V. **96 ÷ 100**
- VI. **101 ÷ 105**
- VII. **106 ÷ 110**

На следующем этапе необходимо определить сколько значений признака входит в каждый интервал. Для того чтобы не переписывать еще раз значения признака, используют метод штрихов или конвертиков. Более простым является метод штрихов. Сущность этого метода заключается в том, что вместо значения признака напротив соответствующего интервала мы рисуем штрих, затем подсчитываем количество штрихов и определяем сколько значений признака находятся в полученных интервалах, т.е. определяем частоту встречаемости признака **f**. Сумма частот $\sum f$ должна быть равна общему числу значений признака **n**, т.е.

$$\sum f = n = 50$$

Заполняем вспомогательную расчетную таблицу

Интервал	Частота, f	X	$f \cdot X$	X^2	$f \cdot X^2$
75 ÷ 80	2	78	156	6084	12168
81 ÷ 85	5	83	415	6889	34445
86 ÷ 90	9	88	792	7744	69696
91 ÷ 95	14	93	1302	8649	121086
96 ÷ 100	10	98	980	9604	96040
101 ÷ 105	8	103	824	1069	84872
106 ÷ 110	2	108	216	11664	23328
Суммы	50	-	4685	-	441635

X – середина интервала, для определения X складываем верхнюю и нижнюю границы интервала и полученное слагаемое делим на 2. Так, для первого интервала $X=(75+80):2=155:2=77,5$ округляем до целых 78; для второго $X=(81+85):2=166:2=83$ и т.д.

Таким образом, заполнив столбцы « X » и « f », мы получаем вариационный ряд, где каждому значению признака X соответствует частота f .

Проводим статистическую обработку данных наблюдения:

Средняя арифметическая взвешенная $\bar{x} = (\sum f \cdot x) \div n = 4685 \div 50 = 94$

Сумма квадратов отклонений $\sum (X - \bar{x})^2 = \sum f \cdot X^2 - (\sum f \cdot X)^2 : n =$
 $= 441635 - 4685^2 : 50 = 441625 - 21949225 : 50 = 441635 - 438984 = 2651$

Дисперсия $S^2 = \frac{\sum (X - \bar{x})^2}{n - 1} = \frac{2651}{50 - 1} = \frac{2651}{49} = 54$

Стандартное отклонение $s = \sqrt{S^2} = \sqrt{54} = 7,3$

Коэффициент вариации $V = \frac{s}{\bar{x}} \cdot 100 = \frac{7,3}{99} \cdot 100 = 7,3\%$

Таким образом, значение коэффициента вариации меньше 10%, изменчивость признака незначительна, следовательно, признак «масса плода» у перца сорта Золотой юбилей выровнен.

Ошибка выборочной средней $S_x = \frac{s}{\sqrt{n}} = \frac{7,3}{7} = 1,0$

Задание. Согласно N варианта провести статистический анализ результатов наблюдения.

Шифр задания к работе

№ варианта	Номера колонок	№ варианта	Номера колонок	№ варианта	Номера колонок
1	1 2 3 4 5	11	1 3 4 7 9	21	3 7 8 9 10
2	2 3 4 5 6	12	4 5 8 9 10	22	1 4 5 6 9
3	3 4 5 6 7	13	1 2 3 6 9	23	1 5 6 7 10
4	4 5 6 7 8	14	2 3 4 8 10	24	2 4 5 7 8
5	5 6 7 8 9	15	3 4 5 9 10	25	1 5 7 9 10
6	6 7 8 9 10	16	1 3 5 6 8	26	2 3 4 8 10
7	1 3 5 7 9	17	4 6 7 9 10	27	3 5 6 8 9
8	2 4 6 8 10	18	2 5 6 8 9	28	2 4 6 7 9
9	1 2 5 6 9	19	3 5 6 8 10	29	3 4 5 8 9
10	2 3 5 7 10	20	1 2 4 5 7	30	4 6 7 8 9

Пример. Масса зеленца огурца гибрида Струмок, г

Значение признака по колонкам

Номера колонок									
1	2	3	4	5	6	7	8	9	10
72	107	118	100	120	80	97	88	119	85
120	87	93	129	93	99	76	107	92	128
97	130	106	82	125	100	102	87	107	90
81	98	108	91	104	89	94	91	84	118
124	73	96	107	96	97	109	129	101	104
92	106	103	83	109	125	86	77	108	84
110	117	90	109	97	90	117	105	105	115
84	97	102	75	123	113	103	93	116	97
105	115	99	100	88	83	96	122	94	104
127	92	85	102	111	101	91	102	88	79

Выполнение работы.

1. Записать название примера, номер варианта.
2. Заполнить рабочую таблицу.

В таблице «шифр задания к работе» в колонке «№ варианта» находим № варианта, заданный преподавателем. Из соседней колонки «Номера колонок» в таблицу записываем попавшие в задание номера (в каждую графу по номеру). В колонку «Значение признака, х» выписываем значения признака, расположенные под соответствующим номером.

Номера колонок	Значение признака, х

3. Подсчитать общее количество значений признака

n =

4. Определить количество интервалов

$$k = \sqrt{n}$$

5. Рассчитать величину интервала

$$i = \frac{X_{\max} - X_{\min}}{k}$$

6. Заполнить расчетную таблицу

Интервал	Разноска значений признака	Частота, f	X	$f \cdot X$	X^2	$f \cdot X^2$
Суммы			-		-	

7. Провести статистическую обработку данных

Средняя арифметическая $\bar{x} = \sum (f \cdot X) : n =$

Сумма квадратов отклонений $\sum (X - \bar{x})^2 = \sum f \cdot X^2 - (\sum f \cdot X)^2 : n =$

Дисперсия $S^2 = \frac{\sum (X - \bar{x})^2}{n - 1} =$

Стандартное отклонение $s = \sqrt{S^2} =$

Коэффициент вариации $V = \frac{s}{\bar{x}} \cdot 100 =$

Ошибка выборочной средней $S_x = \frac{s}{\sqrt{n}} =$

Лабораторная работа №2
ОЦЕНКА СУЩЕСТВЕННОСТИ РАЗНОСТИ СРЕДНИХ
ДВУХ ВЫБОРОК ПРИ КОЛИЧЕСТВЕННОЙ
ИЗМЕНЧИВОСТИ

Понятия «совокупность» и «выборка» являются одними из основных в математической статистике. **Совокупностью**, или **генеральной совокупностью** называется вся группа объектов подлежащих изучению. Наиболее полные и точные данные получают при анализе всей генеральной совокупности. Однако не всегда технически возможно и целесообразно исследовать по тому или иному признаку всю совокупность. В этих случаях изучают часть совокупности. Часть объектов, которая попала в исследование, называют **выборочной совокупностью**, или **выборкой**. На основе статистических показателей малой выборки делается заключение обо всей совокупности.

Для сравнения данных двух или нескольких выборок используют различные статистические методы, которые сводятся к проверке статистических гипотез. **Статистической гипотезой** называют научное предположение о тех или иных статистических законах распределения значений признака X , которое может быть проверено на основе выборки. В агрономических исследованиях обычно проверяется *нулевая гипотеза*, которая предполагает, что между выборками нет существенных различий и они принадлежат к одной генеральной совокупности.

Для оценки существенности разности между средними используют следующие методы:

По критерию существенности. Критерий существенности определяется как отношение разности средних к ее ошибке. Критерий существенности сравнивается с критерием Стьюдента. Значение критерия Стьюдента берется из статистической таблицы с учетом принятого уровня значимости и числа степеней свободы, которое рассчитывается по формуле

$\nu = n_1 + n_2 - 2$, где n_1 - число наблюдений в первой выборке, n_2 - число наблюдений во второй выборке.

Если $t_{\text{факт}} \geq t_{\text{теор}}$, разница между выборками существенна; если $t_{\text{факт}} < t_{\text{теор}}$, разница несущественна.

По величине наименьшей существенной разницы (НСР). Наименьшая существенная разность сравнивается с разностью между средними $d = x_1 - x_2$.

Если $d \geq \text{НСР}$, разница между выборками существенна; если $d < \text{НСР}$, различие несущественно.

Пример. Содержание β -каротина (мг/100 г) в корнеплодах моркови сортов Артек (вариант1) и Стелла (вариант 2).

В выборках по 10 корнеплодов определили содержание β -каротина. Был получены следующие данные:

Артек	12,8	12,1	13,0	13,5	11,9	12,7	13,4	12,9	12,3	13,2
Стелла	16,1	15,8	16,3	17,6	14,9	16,5	17,2	16,9	15,4	16,7

Следует определить, различаются ли сорта моркови по содержанию β -каротина в корнеплодах.

1. *Определяем статистические характеристики данных варианта 1.*

Выборка малого объема, содержит лишь 10 значений ($n=10$), поэтому рассчитываем **простую среднюю арифметическую**:

$$\bar{x} = \frac{\sum X}{n} = \frac{12,8 + 12,1 + 13,0 + 13,5 + 11,9 + 12,7 + 13,4 + 12,9 + 12,3 + 13,2}{10} = \frac{127,8}{10} = 12,8$$

$$\text{Корректирующий фактор } C = \frac{(\sum X)^2}{n} = \frac{127,8^2}{10} = \frac{16332,84}{10} = 1633,3$$

Сумма квадратов отклонений

$$\sum (X - \bar{x})^2 = \sum X^2 - C = 12,8^2 + 12,1^2 + 13,0^2 + 13,5^2 + 11,9^2 + 12,7^2 + 13,4^2 + 12,9^2 + 12,3^2 + 13,2^2 - 1633,3 = 163,8 + 146,1 + 169,0 + 182,2 + 141,6 + 161,3 + 179,6 + 166,4 + 151,3 + 174,2 - 1633,3 = 2,2$$

$$\text{Дисперсия } S^2 = \frac{\sum (X - \bar{x})^2}{n - 1} = \frac{2,2}{10 - 1} = \frac{2,2}{9} = 0,2$$

$$\text{Стандартное отклонение } s = \sqrt{S^2} = \sqrt{0,2} = 0,4$$

$$\text{Коэффициент вариации } V = \frac{s}{\bar{x}} \cdot 100 = \frac{0,4}{12,8} \cdot 100 = 3$$

$$\text{Ошибка выборочной средней } S_x = \frac{s}{\sqrt{n}} = \frac{0,2}{\sqrt{10}} = \frac{0,2}{3,2} = 0,06$$

2. Определяем статистические характеристики данных варианта 2.

Средняя арифметическая

$$\bar{x} = \frac{\sum X}{n} = \frac{16,1 + 15,8 + 16,3 + 17,6 + 14,9 + 16,5 + 17,2 + 16,9 + 15,4 + 16,7}{10} =$$

$$= \frac{163,4}{10} = 16,3$$

$$\text{Корректирующий фактор } C = \frac{(\sum X)^2}{n} = \frac{163,4^2}{10} = \frac{26699,6}{10} = 2670,0$$

Сумма квадратов отклонений

$$\sum (X - \bar{x})^2 = \sum X^2 - C = 16,1^2 + 15,8^2 + 16,3^2 + 17,6^2 + 14,9^2 + 16,5^2 + 17,2^2 +$$

$$+ 16,9^2 + 15,4^2 + 16,7^2 - 2670,0 = 259,2 + 249,6 + 265,7 + 309,8 + 222,0 + 272,2 + 295,8 +$$

$$+ 285,6 + 237,2 + 278,9 - 2670,0 = 6,0$$

$$\text{Дисперсия } S^2 = \frac{\sum (X - \bar{x})^2}{n - 1} = \frac{6,0}{10 - 1} = \frac{6,0}{9} = 0,7$$

$$\text{Стандартное отклонение } s = \sqrt{S^2} = \sqrt{0,7} = 0,8$$

$$\text{Коэффициент вариации } V = \frac{s}{\bar{x}} \cdot 100 = \frac{0,8}{16,3} \cdot 100 = 5$$

$$\text{Ошибка выборочной средней } S_x = \frac{s}{\sqrt{n}} = \frac{0,8}{\sqrt{10}} = \frac{0,8}{3,2} = 0,25$$

3. Проводим оценку существенности разности между средними:

а) по критерию существенности

$$t_{\text{факт}} = \frac{d}{s_d} = \frac{x_2 - x_1}{\sqrt{s_{x1}^2 + s_{x2}^2}} = \frac{16,3 - 12,8}{\sqrt{0,06^2 + 0,25^2}} = \frac{3,5}{\sqrt{0,004 + 0,062}} = \frac{3,5}{\sqrt{0,066}} = \frac{3,5}{0,26} = 13,5$$

при $v = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$

$t_{05} = 2,10$

$t_{\text{факт}}$ больше t_{05} , следовательно между вариантами есть существенная разница.

б) по величине наименьшей существенной разности (НСР):

$$d = x_2 - x_1 = 16,3 - 12,8 = 3,5$$

$$НСР_{05} = t_{05} \cdot s_d = 2,10 \cdot 0,26 = 0,5$$

d больше $НСР_{05}$, следовательно между вариантами есть существенная разница.

Вывод: По содержанию β -каротина в корнеплодах сорт Стелла существенно превосходит сорт Артек.

Задание. Согласно N задания провести оценку существенности разности между средними 2 выборок.

Шифр задания к работе

№ задания	№№ колонок	№ задания	№№ колонок	№ задания	№№ колонок	№ задания	№№ колонок
1	1 6	7	3 9	13	5 8	19	2 6
2	2 7	8	4 10	14	1 9	20	3 7
3	3 8	9	5 7	15	2 10	21	4 8
4	4 9	10	1 8	16	3 5	22	5 10
5	5 6	11	2 9	17	4 7	23	2 8
6	1 7	12	3 10	18	5 9	24	1 10

Пример. Содержание крахмала (%) в клубнях картофеля сортов Астрал и Ягодка.

Астрал					Ягодка				
Номера колонок					Номера колонок				
1	2	3	4	5	6	7	8	9	10
15,3	16,7	17,1	14,8	15,6	17,9	18,2	17,8	19,4	19,3
14,9	14,3	17,3	15,4	15,2	18,4	19,4	18,3	18,9	18,0
16,8	15,9	14,5	16,3	14,9	19,1	19,8	19,8	18,2	18,9
15,7	16,8	17,0	17,1	17,4	17,8	18,3	19,1	18,0	19,2
16,2	17,1	16,4	16,8	17,0	18,2	17,9	18,6	19,4	19,0
17,4	14,8	15,2	15,7	16,8	18,4	18,8	18,2	17,9	18,4
14,3	15,7	16,1	16,2	14,5	19,2	19,4	19,4	19,7	19,5
15,9	16,8	17,2	16,7	15,1	18,8	18,7	19,0	18,5	18,3
16,2	14,9	15,9	15,6	15,9	17,9	19,2	18,7	19,8	18,7
17,1	15,2	16,3	14,9	16,2	19,4	18,0	18,0	19,6	19,8

Выполнение работы.

2. Записать название примера, номер задания.

№	Астрал	
№	Ягодка	

3. Провести статистический анализ каждой выборки.

Статистические показатели	Астрал	Ягодка
$n =$		
$\bar{x} = \frac{\sum X}{n} =$		
$C = \frac{(\sum X)^2}{n} =$		
$\sum (X - \bar{x})^2 = \sum X^2 - C =$		
$s^2 = \frac{\sum (X - \bar{x})^2}{n - 1} =$		
$s = \sqrt{s^2} =$		

$V = \frac{s}{\bar{x}} \cdot 100 =$		
$s_x = \frac{s}{\sqrt{n}} =$		

3. Провести оценку существенности разности:

а) по критерию существенности

$$t_{\text{факт}} = \frac{d}{s_d} = \frac{x_2 - x_1}{\sqrt{s_{x1}^2 + s_{x2}^2}} =$$

при $v = n_1 + n_2 - 2 =$

$t_{05} =$

б) по величине наименьшей существенной разности:

$d = x_2 - x_1 =$

$$HCP_{05} = t_{05} \cdot s_d =$$

Вывод:

Лабораторная работа №3

ДИСПЕРСИОННЫЙ АНАЛИЗ ДАННЫХ ОДНОФАКТОРНОГО ПОЛЕВОГО ОПЫТА, ПРОВЕДЕННОГО МЕТОДОМ РЕНДОМИЗИРОВАННЫХ ПОВТОРЕНИЙ

Дисперсионный анализ был разработан английским ученым Робертом Фишером, который открыл закон распределения средних квадратов или дисперсий. При дисперсионном анализе одновременно обрабатываются данные нескольких вариантов, составляющих единый статистический комплекс, структура которого определяется схемой и методикой эксперимента. Используется дисперсионный анализ для проверки нулевой гипотезы, которая предполагает, что между изучаемыми вариантами нет существенных различий, т.е. они относятся к одной генеральной совокупности.

Схема дисперсионного анализа зависит от количества изучаемых факторов, типа сравнительного эксперимента, а в полевых опытах и от метода размещения вариантов по делянкам опыта.

Для возможности применения дисперсионного анализа опыт должен быть проведен в нескольких повторностях. *Повторность* – это количество одноименных делянок каждого варианта. Количество повторностей определяется целями и задачами эксперимента, выравненностью почвенного плодородия, биологическими особенностями культурных растений. В полевых однофакторных опытах оптимальное количество повторностей - от 4 до 6. Одним из основных факторов внешней среды, оказывающих значительное влияние на величину случайной ошибки эксперимента, является варьирование почвенного плодородия. Снизить уровень ошибки можно правильным подбором метода размещения вариантов по делянкам полевого опыта. В полевых однофакторных опытах наиболее часто используется метод рендомизированных повторений (*Повторение* – это земельный участок с полным набором вариантов опыта.) Сущность этого метода заключается в том, что внутри каждого повторения варианты размещаются по делянкам случай по жребию или по таблице случайных чисел.

Дисперсионный анализ характеризуется строгой логичностью и последовательностью выполнения математических операций. В основе дисперсионного анализа лежит *расчленение общей суммы квадратов отклонений (общего варьирования) и общего числа степеней свободы на компоненты*, количество которых определяется структурой статистического комплекса.

Если, согласно нулевой гипотезе, все выборки принадлежат к одной генеральной совокупности, то обобщенной характеристикой данной совокупности будет являться средняя арифметическая $\bar{x}$. Большинство

значений изучаемого признака отклоняются от средней арифметической в ту или иную сторону. Эти отклонения могут быть обусловлены как влиянием случайных факторов, так и влиянием изучаемых. Сумма всех положительных и отрицательных отклонений равна 0, поэтому все отклонения фактических значений признака от средней арифметической необходимо возвести в квадрат и лишь затем сложить. Полученная сумма квадратов отклонений и называется *общее варьирование* и обозначается C_y .

В полевых однофакторных опытах, заложенных методом рендомизированных повторений, общее варьирование C_y делится на 3 компонента: варьирование вариантов C_v , варьирование повторений C_p и случайное варьирование, или варьирование ошибки C_z , т.е.

$$C_y = C_v + C_p + C_z$$

Варьирование вариантов обусловлено действием изучаемых факторов и определяет варьирование признака между выборками.

Варьирование повторений обусловлено влиянием почвенного плодородия и определяет варьирование признака между повторениями.

Варьирование случайное обусловлено действием случайных факторов и определяет варьирование признака внутри выборок.

Общее число степеней свободы ν_y также разлагается на 3 части: степени свободы для вариантов ν_v , степени свободы для повторений ν_p и для случайного варьирования ν_z , т.е.

$$\nu_y = \nu_v + \nu_p + \nu_z$$

Сумму квадратов вариантов и сумму квадратов ошибки делят на соответствующие им степени свободы и получают две дисперсии: дисперсию вариантов и дисперсию ошибки:

$$S_v^2 = \frac{C_v}{\nu_v} \qquad S_z^2 = \frac{C_z}{\nu_z}$$

Отношение дисперсии вариантов к дисперсии ошибки дает фактическое значение критерия Фишера F :

$$F = \frac{S_v^2}{S_z^2}$$

Для оценки действия изучаемых факторов фактическое значение критерия Фишера сравнивают с теоретическим, которое берут из статистической таблицы с учетом принятого уровня значимости (5%) и числа степеней свободы вариантов и ошибки.

Если $F_{\text{факт.}} < F_{\text{теор.}}$, нулевая гипотеза принимается, следовательно между вариантами нет существенных различий.

Если $F_{\text{факт.}} \geq F_{\text{теор.}}$, нулевая гипотеза отвергается, следовательно между вариантами есть существенные различия.

Критерий Фишера лишь показывает, что между изучаемыми вариантами есть существенные различия, но не указывает между какими средними имеются эти различия.

Для оценки существенности различий между средними рассчитывают **наименьшую существенную разность (НСР)**, которая указывает предельную величину ошибки для разности двух выборочных средних d . Если разность двух выборочных средних больше или равна НСР, то она существенна, если она меньше НСР – несущественна.

Пример. Влияние доз регулятора роста фумаран на урожайность редиса сорта Розово-красный с белым кончиком.

Вариант	Повторения		
	1	2	3
Контроль – без обработки	12,8	13,0	12,7
Фумаран – 0,05%	14,8	14,6	15,4
Фумаран – 0,15%	16,5	17,3	17,0
Фумаран – 0,25%	21,8	22,9	22,3

Нулевая гипотеза предполагает, что фумаран не оказывает существенного влияния на урожайность редиса сорта Розово-красный с белым кончиком. Для проверки нулевой гипотезы необходимо провести дисперсионный анализ.

Решение:

Заполняем вспомогательную расчетную таблицу:

Вариант	Повторения			Суммы, V	Средние, $\bar{x}$
	1	2	3		
Контроль – без обработки	12,8	13,0	12,7	38,5	12,8
Фумаран – 0,05%	14,8	14,6	15,4	44,8	14,9
Фумаран – 0,15%	16,5	17,3	17,0	50,8	16,9
Фумаран – 0,25%	21,8	22,9	22,3	67,0	22,3
Сумма, P	65,9	67,8	67,4	$\sum X = 201,1$	-

Проверяем правильность заполнения расчетной таблицы: сумма сумм значений признака по вертикали должна быть равна сумме сумм значений признака по горизонтали, т.е. $65,9 + 67,8 + 67,4 = 38,5 + 44,8 + 50,8 + 67,0 = 201,1$

Таким образом, сумма всех значений признака «урожайность редиса» равна 201,1, т.е. $\sum X = 201,1$.

Рассчитываем:

1. Общее число наблюдений $N = l \cdot n = 4 \cdot 3 = 12$,
где l - число вариантов,
 n - количество повторений.

2. Корректирующий фактор

$$C = (\sum X)^2 \div N = (201,1)^2 \div 12 = 40441,2 : 12 = 3370,1$$

3. Суммы квадратов отклонений:

общая: для определения общей суммы квадратов отклонений следует каждое значение признака «урожайность редиса» (X) следует возвести в квадрат и затем сложить. Таким образом, мы получим выражение

$\sum X^2$. От полученной суммы отнимаем корректирующий фактор. В нашем примере общая сумма квадратов отклонений рассчитывается следующим образом:

$$\begin{aligned} C_y = \sum X^2 - C &= 12,8^2 + 13,0^2 + 12,7^2 + 14,8^2 + 14,6^2 + 15,4^2 + \\ &+ 16,5^2 + 17,3^2 + 17,0^2 + 21,8^2 + 22,9^2 + 22,3^2 - 3370,1 = \\ &= 163,9 + 169,0 + 161,3 + 219,0 + 213,2 + 237,2 + 272,2 + 299,3 + 289,0 + \\ &+ 475,2 + 524,4 + 497,3 - 3370,1 = 3520,9 - 3370,1 = 150,8 \end{aligned}$$

вариантов: для определения выражения $\sum V^2$ необходимо сначала возвести в квадрат все суммы V , затем их сложить. Затем полученную сумму делим на число повторений и вычитаем корректирующий фактор. В нашем примере сумма квадратов отклонений вариантов рассчитывается следующим образом:

$$\begin{aligned} C_v = \sum V^2 \div n - C &= (38,5^2 + 44,8^2 + 50,8^2 + 67,0^2) : 3 - 3370,1 = \\ &= (1482,2 + 2007,0 + 2580,6 + 4489,0) : 3 - 3370,1 = 10558,8 : 3 - 3370,1 = \\ &= 3519,6 - 3370,1 = 149,5 \end{aligned}$$

повторений: для определения выражения $\sum P^2$ необходимо сначала возвести в квадрат все суммы P , затем их сложить. Полученную сумму делим на количество вариантов и вычитаем корректирующий фактор. В нашем примере сумма квадратов отклонений повторений рассчитывается следующим образом:

$$C_p = \sum P^2 \div l - C = (65,9^2 + 67,8^2 + 67,4^2) : 4 - 3370,1 =$$

$$= (4342,8 + 4596,8 + 4542,8) : 4 - 3370,1 = 13482,4 : 4 - 3370,1 =$$

$$= 3370,6 - 3370,1 = 0,5$$

остатка:

$$C_z = C_y - C_v - C_p = 150,8 - 149,5 - 0,5 = 0,8$$

4. Степени свободы:

общее число $\nu_y = N - 1 = 12 - 1 = 11$

для вариантов $\nu_v = l - 1 = 4 - 1 = 3$

для повторений $\nu_p = n - 1 = 3 - 1 = 2$

остатка $\nu_z = \nu_y - \nu_v - \nu_p = 11 - 3 - 2 = 6$

5. Заполняем таблицу дисперсионного анализа

Виды варьирования	Суммы квадратов	Степени свободы	Дисперсия	Критерий F	
				факт.	05
Общее	150,8	11	-	-	-
Вариантов	149,5	3	49,8	498	5,14
Повторений	0,5	2	-	-	-
Случайное (остаток)	0,8	6	0,1	-	-

Дисперсия, или средний квадрат вариантов рассчитывается как отношение суммы квадратов отклонений вариантов к степеням свободы вариантов, т.е. $S_v^2 = \frac{149,5}{3} = 49,8$

Дисперсия, или средний квадрат остатка рассчитывается как отношение суммы квадратов отклонений остатка к степеням свободы остатка, т.е. $S_z^2 = \frac{0,8}{6} = 0,1$

Фактическое значение критерия Фишера рассчитывается по формуле

$$F = \frac{S_v^2}{S_z^2} = \frac{49,8}{0,1} = 498$$

Значение критерия F берется из статистической таблицы с учетом степеней свободы вариантов и остатка. Фактическое значение критерия Фишера больше теоретического, следовательно, нулевая гипотеза отвергается и между вариантами есть существенные различия.

Т.к., между вариантами есть существенные различия, определяем наименьшую существенную разность

6. Определяем ошибку разности средних.

$$s_d = \sqrt{\frac{2s_z^2}{n}} = \sqrt{\frac{2 \cdot 0,1}{3}} = \sqrt{\frac{0,2}{3}} = \sqrt{0,07} = 0,26$$

7. Рассчитываем наименьшую существенную разность (НСР):

$$НСР_{05} = t_{05} \cdot s_d = 2,45 \cdot 0,26 = 0,6$$

Значение критерия t_{05} берется из статистической таблицы для числа степеней свободы остатка. При $v = 6$ $t_{05} = 2,45$

Таблица средних значений

Варианты опыта	Среднее значение	Отклонение от контроля, d
Контроль – без обработки	12,8	-
Фумаран – 0,05%	14,9	2,1
Фумаран – 0,15%	16,9	4,1
Фумаран – 0,25%	22,3	9,5

Для определения *отклонения от контроля*, от средних значений опытного варианта отнимаем среднее значение контрольного варианта, т.е. $14,9 - 12,8 = 2,1$; $16,9 - 12,8 = 4,1$; $22,3 - 12,8 = 9,5$.

Полученные отклонения сравниваем с $НСР_{05} = 0,6$. Все отклонения от контроля больше НСР, следовательно, опытные варианты существенно превосходят контрольный. Можно попарно сравнить 2 и 3, 2 и 4, 3 и 4 варианты.

Вывод: Обработка регулятором роста фумаран *существенно* повышает урожайность редиса сорта Розово-красный с белым кончиком. Увеличение концентрации рабочего раствора с 0,05 для 0,25% *существенно* повышает урожайность.

Задание:

Получив N задания от преподавателя, провести дисперсионный анализ результатов однофакторного полевого опыта.

Шифр задания к работе

№ задания	Повторения	№ задания	Повторения	№ задания	Повторения	№ задания	Повторения
1	1 2 3	11	2 4 6	21	5 6 9	31	4 7 8
2	2 3 4	12	4 6 8	22	1 3 4	32	5 8 9
3	3 4 5	13	1 2 4	23	2 4 5	33	1 3 6
4	4 5 6	14	2 3 5	24	3 5 6	34	2 5 8
5	5 6 7	15	3 4 7	25	4 6 7	35	3 7 9
6	6 7 8	16	5 6 8	26	5 7 8	36	4 5 8
7	7 8 9	17	6 7 8	27	6 8 9	37	1 5 8
8	1 3 5	18	1 2 5	28	1 4 5	38	1 6 9
9	3 5 7	19	2 3 7	29	2 5 6	39	2 6 8
10	5 7 9	20	3 4 8	30	3 6 7	40	3 6 8

Пример. Урожайность картофеля сорта Ягодка в зависимости от схем посадки микроклубней (т/га)

Вариант	Повторения, X								
	1	2	3	4	5	6	7	8	9
60x20 см - контроль	49,5	52,1	47,9	49,3	49,7	50,2	49,1	51,8	49,4
50x20 см	47,2	48,7	46,9	47,9	47,8	48,0	47,7	48,1	45,6
40x20 см	46,7	47,0	46,2	46,3	46,0	47,2	46,9	47,3	47,1
30x20 см	43,7	44,5	44,9	45,0	43,0	44,3	45,2	46,1	44,2

Выполнение работы:

1. Записать условие примера, N задания, заполнить таблицу.

Варианты опыта	Повторение			Суммы, V	Средние, x
60x20 см - контроль					
50x20 см					
40x20 см					
30x20 см					
Суммы				$\sum X$	-

2. Провести дисперсионный анализ

Общее число наблюдений $N = l \cdot n =$, где l - число вариантов,
 n - количество повторений

Корректирующий фактор $C = (\sum X)^2 \div N =$

Суммы квадратов:

общая $C_y = \sum X^2 - C =$

вариантов $C_v = \sum V^2 \div n - C =$

повторений $C_p = \sum P^2 \div l - C =$

остатка $C_z = C_y - C_v - C_p =$

Степени свободы:

общее число $\nu_y = N - 1 =$

для вариантов $\nu_v = l - 1 =$

для повторений $\nu_p = n - 1 =$

остатка $\nu_z = \nu_y - \nu_v - \nu_p =$

Таблица дисперсионного анализа

Виды варьирования	Суммы квадратов	Степени свободы	Дисперсия	Критерий F	
				факт.	05
Общее			-	-	-
Вариантов					
Повторений			-	-	-
Случайное (остаток)				-	-

Ошибка разности средних $S_d = \sqrt{\frac{2 S_z^2}{n}} =$

$HCP_{05} = t_{05} \cdot S_d =$

Значение критерия t_{05} берется из статистической таблицы для числа степеней свободы остатка.

Таблица средних значений

Варианты опыта	Среднее значение	Отклонение от стандарта, d
60x20 см - контроль		-
50x20 см		
40x20 см		
30x20 см		

Вывод:

ДИСПЕРСИОННЫЙ АНАЛИЗ ДВУХФАКТОРНОГО ПОЛЕВОГО ОПЫТА, ПРОВЕДЕННОГО МЕТОДОМ РЕНДОМИЗИРОВАННЫХ ПОВТОРЕНИЙ

В многофакторных опытах изучают действие и взаимодействие нескольких изучаемых факторов на изменчивость результативного признака. Эффект от совместного применения изучаемых факторов может быть больше (синергизм) или меньше (антагонизм) суммы эффектов от раздельного применения каждого из них. В первом случае существует положительное взаимодействие факторов, во втором – отрицательное. Когда факторы не взаимодействуют, прибавка от их совместного применения равна сумме прибавок от раздельного применения (аддитивизм).

Дисперсионный анализ многофакторного опыта проводят в два этапа. На первом этапе данные эксперимента обрабатывают как однофакторный опыт. Ко второму этапу переходят если фактическое значение критерия Фишера больше или равно теоретическому, т.е. если нулевая гипотеза отвергается. Если фактическое значение критерия Фишера меньше теоретического дисперсионный анализ можно завершить, т.к. между вариантами нет существенных различий.

Алгоритм второго этапа дисперсионного анализа полевого опыта зависит от метода размещения вариантов по делянкам опыта. В полевых многофакторных опытах наиболее часто используются метод рендомизированных повторений и метод расщепленных делянок.

Если опыт заложен методом рендомизированных повторений, на втором этапе дисперсионного анализа варьирование вариантов делится на компоненты, соответствующие источникам варьирования – главные эффекты изучаемых факторов и их взаимодействия, на соответствующие компоненты разлагается также и число степеней свободы вариантов.

При дисперсионном анализе многофакторных опытов сумма квадратов отклонений для вариантов S_v и число степеней свободы вариантов расчленяется на компоненты в зависимости от структуры опыта:

в двухфакторном опыте - на 3:

$$S_v = S_A + S_B + S_{AB}$$

$$V_v = v_a + v_b + v_{ab}$$

а в трехфакторном – на 7 компонентов:

$$S_v = S_A + S_B + S_C + S_{AB} + S_{AC} + S_{BC} + S_{ABC}$$

$$V_v = v_a + v_b + v_c + v_{ab} + v_{ac} + v_{bc} + v_{abc}$$

Для оценки существенности различий между средними в двухфакторном полевом опыте, заложенном методом рендомизированных повторений, рассчитывают 3 значения $НСП_{05}$:

$НСП_{05A}$ – для оценки действия фактора А,

$НСП_{05B}$ – для оценки действия фактора В,

$НСП_{05AB}$ – для оценки взаимодействия факторов.

Пример. Влияние способов возделывания на урожайность сортов капусты (т/га).

Способ возделывания (Фактор А)	Сорт капусты (Фактор В)	Повторения, X			
		1	2	3	4
Рассадный	Волна	65,2	67,4	62,3	66,1
	Арривист	74,7	77,2	71,9	76,4
	Вестри	87,5	89,1	85,3	86,9
Безрассадный	Волна	50,1	47,2	49,4	53,5
	Арривист	61,4	64,1	65,2	59,3
	Вестри	68,7	70,4	65,4	69,1

В опыте изучаются 2 фактора и их взаимодействие:

фактор А – способ выращивания капусты

фактор В – сорт капусты

фактор АВ – взаимодействие факторов А и В.

В двухфакторном опыте проверяется 3 нулевые гипотезы: по фактору А, по фактору В и по их взаимодействию. В нашем примере нулевые гипотезы формулируются следующим образом:

по фактору А – *способ возделывания не влияет на урожайность капусты;*

по фактору В – *изучаемые сорта капусты существенно не различаются по урожайности;*

по фактору АВ – *изучаемые факторы не взаимодействуют.*

Для проверки нулевых гипотез проводим дисперсионный анализ.

I. Обрабатываем данные как при проведении однофакторного опыта.

1. Заполняем вспомогательную таблицу:

Способ возделывания (Фактор А)	Сорт капусты (Фактор В)	Повторения, X				Суммы, V	Сред ние, $\bar{x}$
		1	2	3	4		
Рассадный	Волна	65,2	67,4	62,3	66,1	261,0	65,2
	Арривист	74,7	77,2	71,9	76,4	300,2	75,0
	Вестри	87,5	89,1	85,3	86,9	348,8	87,2
Безрассадный	Волна	50,1	47,2	49,4	53,5	200,2	50,0
	Арривист	61,4	64,1	65,2	59,3	250,0	62,5
	Вестри	68,7	70,4	65,4	69,1	273,6	68,4
Суммы, P		407,6	415,4	399,5	411,3	$\sum X = 1633,8$	-

Проверим правильность заполнения таблицы: сумма сумм значений признака по вертикали должна быть равна сумме сумм значений признака по горизонтали, т.е. $\sum X = 407,6 + 415,4 + 399,5 + 411,3 = 261,0 + 300,2 + 348,8 + 200,2 + 250,0 + 273,6 = 1633,8$

2. Рассчитываем суммы квадратов:

$$\sum X^2 = 65,2^2 + 67,4^2 + 62,3^2 + 66,1^2 + 74,7^2 + 77,2^2 + 71,9^2 + 76,4^2 + 87,5^2 + 89,1^2 + 85,3^2 + 86,9^2 + 50,1^2 + 47,2^2 + 49,4^2 + 53,5^2 + 61,4^2 + 64,1^2 + 65,2^2 + 59,3^2 + 68,7^2 + 70,4^2 + 65,4^2 + 69,1^2 = 4251,0 + 4542,8 + 3881,3 + 4369,2 + 5580,1 + 5959,8 + 5169,6 + 5837,0 + 7656,2 + 7938,8 + 7276,1 + 7551,6 + 2510,0 + 2227,8 + 2440,4 + 2862,2 + 3767,0 + 4108,8 + 4251,0 + 3516,5 + 4719,7 + 4956,2 + 4277,2 + 4774,8 = 114425,1$$

$$\sum V^2 = 261,0^2 + 300,2^2 + 348,8^2 + 200,2^2 + 250,0^2 + 273,6^2 = 68121,0 + 90120,0 + 121661,4 + 40080,0 + 62500,0 + 74857,0 = 457339,4$$

$$\sum P^2 = 407,6^2 + 415,4^2 + 399,5^2 + 411,3^2 = 166137,8 + 172557,2 + 159600,2 + 169167,7 = 667462,9$$

3. Проводим дисперсионный анализ

Определяем общее число наблюдений $N = la \cdot lb \cdot n = 2 \cdot 3 \cdot 4 = 24$

la – число градаций фактора А; lb – число градаций фактора В;

n – количество повторений

Корректирующий фактор:

$$C = (\sum X)^2 : N = 1633,8^2 : 24 = 2669302,4 : 24 = 111220,9$$

Суммы квадратов отклонений:

$$\text{общая } Cy = \sum X^2 - C = 114425,1 - 111220,9 = 3204,2$$

вариантов

$$Cv = \sum V^2 \div n - C = 457339,4 : 4 - 111220,9 = 114334,8 - 111220,9 = 3113,9$$

повторений

$$Cp = \sum P^2 \div (la \cdot lb) - C = 667462,9 : (2 \times 3) - 111220,9 = 667462,9 : 6 - 111220,9 = 111243,8 - 111220,9 = 22,9$$

$$\text{остатка } Cz = Cy - Cv - Cp = 3204,2 - 3113,9 - 22,9 = 67,4$$

Степени свободы:

$$\text{общее число } \nu_y = N - 1 = 24 - 1 = 23$$

$$\text{для вариантов } \nu_v = la \cdot lb - 1 = 2 \times 3 - 1 = 6 - 1 = 5$$

$$\text{для повторений } \nu_p = n - 1 = 4 - 1 = 3$$

$$\text{остатка } \nu_z = \nu_y - \nu_v - \nu_p = 23 - 5 - 3 = 15$$

4. Составляем таблицу дисперсионного анализа:

Виды варьирования	Сумма квадратов	Степени свободы	Дисперсия	Fфакт	F ₀₅
Общая	3204,2	23	-	-	-
Вариантов	3113,9	5	622,8	138,4	2.90
Повторений	22,9	3	-	-	-
Случайное (остаток)	67,4	15	4,5	-	-

Дисперсия, или средний квадрат вариантов рассчитывается как отношение суммы квадратов отклонений вариантов к степеням свободы вариантов, т.е. $S_v^2 = \frac{3113,9}{5} = 622,8$.

Дисперсия, или средний квадрат остатка рассчитывается как отношение суммы квадратов отклонений остатка к степеням свободы остатка, т.е. $S_z^2 = \frac{67,4}{15} = 4,5$.

Фактическое значение критерия Фишера рассчитывается по формуле

$$F = \frac{S_v^2}{S_z^2} = \frac{622,8}{4,5} = 138,4$$

Теоретическое значение критерия Фишера берется из статистической таблицы для 5% уровня значимости и числа степеней свободы вариантов и остатка. При $\nu_v=5$ и $\nu_z=15$ $F_{05}=2.90$

Фактическое значение критерия Фишера больше, чем теоретическое, следовательно, нулевая гипотеза отвергается и между вариантами есть существенные различия. Т.к. между вариантами есть существенные различия, переходим ко второму этапу дисперсионного анализа.

II. 5. Вычисляем суммы квадратов отклонений для факторов А, В и взаимодействия АВ.

Таблица для определения влияния главных эффектов и взаимодействий

А (способ возделывания)	В (сорта)			Суммы А
	Волна	Арривист	Вестри	
Рассадный	261,0	300,2	348,8	910,0
Безрассадный	200,2	250,0	273,6	723,8
Суммы В	461,2	550,2	622,4	$\sum X=1633,8$

Заносим в таблицу суммы значений признака по вариантам, которые берем из первой вспомогательной таблицы (графа «Суммы, V»). Складываем суммы по вертикали и по горизонтали. Проверяем правильность расчетов: сумма сумм значений признака по вертикали равна сумме сумм значений признака по горизонтали и равна $\sum X$, рассчитанной на первом этапе.
Т.е. $\sum X = 461,2 + 550,2 + 622,4 = 910,0 + 723,8 = 1633,8$

Суммы квадратов отклонений:

фактора А

$$C_A = \sum A^2 \div (lb \cdot n) - C = (910,0^2 + 723,8^2) \div (3 \cdot 4) - 111220,9 = (828100,0 + 523886,4) \div 12 - 111220,9 = 1351986,4 \div 12 - 111220,9 = 112665,5 - 111220,9 = 1444,6$$

фактора В

$$C_B = \sum B^2 \div (la \cdot n) - C = (461,2^2 + 550,2^2 + 622,4^2) \div (2 \cdot 4) - 111220,9 = (212705,4 + 302720,0 + 387381,8) \div 8 - 111220,9 = 902807,2 \div 8 - 111220,9 = 112850,9 - 111220,9 = 1630,0$$

взаимодействия АВ $C_{AB} = C_V - C_A - C_B = 3113,9 - 1444,6 - 1630,0 = 39,3$

Степени свободы:

для фактора А $\nu_A = la - 1 = 2 - 1 = 1$

для фактора В $\nu_B = lb - 1 = 3 - 1 = 2$

для взаимодействия АВ $\nu_{AB} = (la - 1) \cdot (lb - 1) = (2 - 1) \cdot (3 - 1) = 2$

6. Составляем таблицу дисперсионного анализа

Виды варьирования	Сумма квадратов	Степени свободы	Дисперсия	Fфакт	F ₀₅
Фактора А	1444,6	1	1444,6	321,0	4,54
Фактора В	1630,6	2	815,3	181,2	3,60
Взаимодействия АВ	39,3	2	19,6	4,3	3,60
Остаток	67,4	15	4,5	-	-

Значение F₀₅ берем из статистической таблицы, исходя из степеней свободы для дисперсии главных эффектов факторов А, В и взаимодействия АВ и остатка.

Определяем фактическое значение критерия Фишера:

для фактора А $F_A = \frac{S_A^2}{S_z^2} = \frac{1444,6}{4,5} = 321,0$

для фактора В $F_B = \frac{s_B^2}{s_z^2} = \frac{815,3}{4,5} = 181,2$

для взаимодействия АВ $F_{AB} = \frac{s_{AB}^2}{s_z^2} = \frac{19,6}{4,5} = 4,3$

Теоретическое значение критерия Фишера:

для фактора А при $v_A=1$ и $v_z=15$ $F_{05}=4,54$

для фактора В при $v_B=2$ и $v_z=15$ $F_{05}=3,60$

для взаимодействия АВ при $v_{AB}=2$ и $v_z=15$ $F_{05}=3,60$

Проводим попарно сравнение фактического и теоретического значения критериев Фишера:

Для фактора А фактическое значение критерия Фишера больше, чем теоретическое, следовательно, нулевая гипотеза отвергается и между вариантами есть существенные различия, т.е. способ возделывания культуры существенно влияет на урожайность капусты.

Для фактора В фактическое значение критерия Фишера больше, чем теоретическое, следовательно, нулевая гипотеза отвергается и между вариантами есть существенные различия, т.е. сорта капусты существенно различаются по урожайности.

Для взаимодействия АВ фактическое значение критерия Фишера больше, чем теоретическое, следовательно, нулевая гипотеза отвергается и между вариантами есть существенные различия, т.е. факторы взаимодействуют.

Т.к. между вариантами есть существенные различия определяем наименьшую существенную разность (НСР).

7. Для оценки существенности частных различий (взаимодействие АВ) определяем:

а) ошибку разности средних:

$$s_d = \sqrt{\frac{2s_z^2}{n}} \sqrt{\frac{2 \cdot 4,5}{4}} = \sqrt{\frac{9}{4}} = \sqrt{2,25} = 1,5$$

в) наименьшую существенную разность для взаимодействия АВ на 5% уровне значимости:

$$НСР_{05} = t_{05} \cdot s_d = 2,13 \cdot 1,5 = 3,2$$

Значение критерия t_{05} берут из статистической таблицы для числа степеней свободы ошибки. При $v_z=15$ $t_{05}=2,13$

7. Для оценки существенности главных эффектов вычисляем:
а) ошибку разности средних для фактора А:

$$s_d^A = \sqrt{\frac{2s_z^2}{n \cdot lb}} = \sqrt{\frac{2 \cdot 4,5}{4 \cdot 3}} = \sqrt{\frac{9}{12}} = \sqrt{0,75} = 0,9$$

- б) ошибку разности средних для фактора В:

$$s_d^B = \sqrt{\frac{2s_z^2}{n \cdot la}} = \sqrt{\frac{2 \cdot 4,5}{4 \cdot 2}} = \sqrt{\frac{9}{8}} = \sqrt{1,12} = 1,1$$

- в) наименьшую существенную разность для фактора А:

$$HCP_{05}^A = t_{05} \cdot s_d^A = 2,13 \cdot 0,9 = 1,9$$

- с) наименьшую существенную разность для фактора В:

$$HCP_{05}^B = t_{05} \cdot s_d^B = 2,13 \cdot 1,1 = 2,3$$

8. Составляем итоговую таблицу. В таблицу заносим средние значения признака по вариантам из первой вспомогательной таблицы (графа «Средние, $\bar{x}$ »). Затем находим средние значения из средних.

***Влияние способов возделывания на урожайность
сортов капусты (т/га).***

Способ возделывания (фактор А)	Сорт капусты (фактор В)			Средние по фактору А
	Волна	Арривист	Вестри	
Рассадный	65,2	75,0*	87,2*	75,8
Безрассадный	50,0	62,5	68,4	60,3
Средние по фактору В	57,6	68,8	77,8	-

$$HCP_{05}^A = 1.9$$

$$HCP_{05}^B = 2.3$$

$$HCP_{05} = 3,2$$

1. С помощью $HCP_{05}^A = 1.9$ сравниваем средние по фактору А, т.е. значения 75,8 и 60,3. Находим разность $d = 75,8 - 60,3 = 15,5$. Разность между средними больше НСР, следовательно, между вариантами есть существенные различия.

2. С помощью $HCP_{05}^B = 2.3$ сравниваем средние по фактору В, т.е. значения 57,6; 68,8 и 77,8. Разность по урожайности между сортами Волна и Арравист $68,8-57,6=11,2$, она больше НСР, следовательно, по урожайности сорт Арравист существенно превосходит сорт Волна; разность по урожайности между сортами Арравист и Вестри $77,8-68,8=11,0$, она больше НСР, следовательно, сорт Вестри по урожайности существенно превосходит сорта Волна и Арравист.

3. С помощью $HCP_{05} = 3,2$ сравниваем средние значения признака, взятые из вспомогательной таблицы. Нецелесообразно сравнивать попарно 6 значений признака, поэтому выбираем самые высокие значения и сравниваем их между собой. Отмечаем самые высокие значения звездочками, находим между ними разность: $87,2-75,0=12,2$. Разность больше НСР, следовательно, она существенна.

Вывод:

1. При возделывании капусты рассадным способом урожайность существенно выше, чем при возделывании безрассадным способом.

2. Сорт Вестри по урожайности существенно превосходит сорта Арравист и Волна; урожайность сорта Арравист существенно выше, чем у сорта Волна.

3. Максимальная доказуемая урожайность отмечена при выращивании сорта Вестри рассадным способом.

Задание. Согласно примера и N задания провести дисперсионный анализ данных двухфакторного полевого опыта, проведенного методом рендомизированных повторений

Шифр задания к работе

N задани я	Фактор		Повтор ение	N задани я	Фактор		Повтор ение
	A	B			A	B	
1	0 1 2	0 1	1 2 3 4	21	01	01 2	1 2 3 4
2	0 1 2	0 1	2 3 4 5	22	01	01 2	2 3 4 5
3	0 1 2	0 1	3 4 5 6	23	01	01 2	3 4 5 6
4	0 1 2	0 1	1 3 5 6	24	01	01 2	1 3 5 6
5	0 1 2	0 1	2 4 5 6	25	01	01 2	2 4 5 6
6	0 1 2	0 1	1 4 5 6	26	01	01 2	1 4 5 6
7	0 1 2	0 1	1 2 4 6	27	01	01 2	1 2 4 6
8	0 1 2	0 1	1 3 4 5	28	01	01 2	1 3 4 5
9	0 1 2	0 1	2 3 4 6	29	01	01 2	2 3 4 6
10	0 1 2	0 1	2 3 5 6	30	01	01 2	2 3 5 6
11	0 1 2	1 2	1 2 3 4	31	1 2	0 1 2	1 2 3 4
12	0 1 2	1 2	2 3 4 5	32	1 2	0 1 2	2 3 4 5

13	0 1 2	1 2	3 4 5 6	33	1 2	0 1 2	3 4 5 6
14	0 1 2	1 2	1 3 5 6	34	1 2	0 1 2	1 3 5 6
15	0 1 2	1 2	2 4 5 6	35	1 2	0 1 2	2 4 5 6
16	0 1 2	1 2	1 4 5 6	36	1 2	0 1 2	1 4 5 6
17	0 1 2	1 2	1 2 4 6	37	1 2	0 1 2	1 2 4 6
18	0 1 2	1 2	1 3 4 5	38	1 2	0 1 2	1 3 4 5
19	0 1 2	1 2	2 3 4 6	39	1 2	0 1 2	2 3 4 6
20	0 1 2	1 2	2 3 5 6	40	1 2	0 1 2	2 3 5 6

Пример. Влияние регуляторов роста на урожайность сортов томата (т/га)

Фактор А (сорт)	Фактор В (регулятор роста)	Повторения, Х					
		1	2	3	4	5	6
0	0	82,9	78,4	85,2	83,4	80,2	77,8
	1	89,6	92,4	87,9	86,7	91,7	93,5
	2	95,6	98,2	94,8	93,2	97,1	96,6
1	0	66,6	62,9	65,1	68,2	69,1	67,2
	1	73,3	73,9	70,1	76,8	75,2	77,1
	2	75,6	79,2	75,3	74,8	79,0	76,3
2	0	78,1	80,2	79,4	83,1	82,9	79,1
	1	89,7	94,5	85,1	87,2	88,4	93,7
	2	97,1	95,2	94,1	98,0	92,4	91,9

Фактор А: 0 - Факел;
1 – Баллада;
2 – Дар.

Фактор В: 0 – без обработки;
1 – Завязь;
2 – Циркон.

Выполнение работы

1. Составляем вспомогательную таблицу.

Фактор А	Фактор В	Повторения, Х				Суммы, V	Средние, $\bar{x}$
		1	2	3	4		
Суммы Р						$\sum X =$	-

2. Вычисляем суммы квадратов отклонений:

Общее число наблюдений: $N = la \cdot lb \cdot n =$

Корректирующий фактор: $C = (\sum X)^2 : N$

Суммы квадратов отклонений:

общая $C_y = \sum X^2 - C =$

вариантов $C_v = \sum V^2 \div n - C =$

повторений $C_p = \sum P^2 \div (la \cdot lb) - C =$

остатка $C_z = C_y - C_v - C_p =$

Степени свободы:

общее число $\nu_y = N - 1 =$

для вариантов $\nu_v = la \cdot lb - 1 =$

для повторений $\nu_p = n - 1 =$

остатка $\nu_z = \nu_y - \nu_v - \nu_p =$

3. Составляем таблицу дисперсионного анализа:

Виды варьирования	Сумма квадратов	Степени свободы	Дисперсия	Fфакт	F ₀₅
Общая			-	-	-
Вариантов					
Повторений			-	-	-
Остаток				-	-

4. Вычисляем суммы квадратов отклонений для факторов А, В и взаимодействия АВ.

Таблица для определения главных эффектов и взаимодействий

А	В			Суммы А
Суммы В				$\sum X =$

Суммы квадратов отклонений:

фактора А $C_A = \sum A^2 \div (lb \cdot n) - C =$

фактора В $C_B = \sum B^2 \div (la \cdot n) - C =$

взаимодействия АВ $C_{AB} = C_V - C_A - C_B$

Степени свободы:

для фактора А $\nu_A = la - 1$

для фактора В $\nu_B = lb - 1$

для взаимодействия АВ $\nu_{AB} = (la - 1) \cdot (lb - 1)$

5. Составляем таблицу дисперсионного анализа

Виды варьирования	Сумма квадратов	Степени свободы	Дисперсия	Fфакт	F ₀₅
Фактора А					
Фактора В					
Взаимодействия АВ					
Остаток				-	-

Значение F₀₅ берем из статистической таблицы, исходя из степеней свободы для дисперсии главных эффектов А, В и взаимодействия АВ и остатка.

6. Для оценки существенности частных различий определяем:

а) ошибку разности средних:

$$s_d = \sqrt{\frac{2s_z^2}{n}}$$

в) наименьшую существенную разность для 5% уровня значимости:

$$HCP_{05} = t_{05} \cdot s_d$$

7. Для оценки существенности главных эффектов и взаимодействия вычисляем:

а) ошибку разности средних для фактора А:

$$s_d^A = \sqrt{\frac{2s_z^2}{n \cdot lb}}$$

б) ошибку разности средних для фактора В:

$$s_d^B = \sqrt{\frac{2s_z^2}{n \cdot la}}$$

в) наименьшую существенную разность для фактора А:

$$HCP_{05}^A = t_{05} \cdot s_d^A$$

с) наименьшую существенную разность для фактора В:

$$HCP_{05}^B = t_{05} \cdot s_d^B$$

8. Составляем итоговую таблицу:

Влияние регуляторов роста на урожайность сортов томата (т/га)

Фактор А	Фактор В			Средние по фактору А
Средние по фактору В				-

$$HCP_{05} =$$

$$HCP_{05}^A =$$

$$HCP_{05}^B =$$

Вывод:

ПРЯМОЛИНЕЙНАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

В агрономических исследованиях практически никогда не встречаются функциональные связи, когда одному значению факториального признака X соответствует одно строго определенное значение результативного признака Y . Обычно наблюдаются такие зависимости, когда каждому значению факториального признака X соответствует множество значений результативного признака Y . Такие связи называются **Корреляционными**, или вероятностными, или стохастическими. Корреляционные связи обнаруживаются только при массовом изучении признака.

По **форме** корреляция может быть *линейной*, или прямолинейной, и *нелинейной*, или криволинейной. Под линейной корреляционной зависимостью между признаками X и Y понимают такую зависимость, когда при одинаковых приращениях признака X признак Y изменяется на одинаковую величину. При нелинейной корреляции при одинаковых приращениях признака X признак Y изменяется на неодинаковую величину.

По **направлению** корреляция может быть *прямой*, или положительной, и *обратной*, или отрицательной. Если при увеличении X величина Y в среднем увеличивается, то корреляция называется положительной или прямой, а если с увеличением X значение Y уменьшается – отрицательной или обратной.

Связь между 2 признаками называют *простой*, между тремя и более признаками – *множественной*.

Под **регрессией** понимают изменение результативного признака Y при определенном изменении одного или нескольких факториальных признаков.

Наиболее простой является линейная корреляционная зависимость между двумя признаками. Для описания линейной корреляционной зависимости используют следующие коэффициенты:

- коэффициент корреляции,
- коэффициент регрессии,
- коэффициент детерминации.

Коэффициент корреляции r показывает тесноту и направление связи, изменяется в пределах от -1 до $+1$, рассчитывается по формуле

$$r = \frac{\sum (X - \bar{x}) \cdot (Y - \bar{y})}{\sqrt{\sum (X - \bar{x})^2 \cdot \sum (Y - \bar{y})^2}}$$

Если коэффициент корреляции равен 1, связь из корреляционной переходит в функциональную; если он равен 0 - между изучаемыми признаками нет прямолинейной корреляционной зависимости, но может быть криволинейная.

Знак коэффициент корреляции показывает направление связи, числовое выражение – тесноту связи. Если перед коэффициентом корреляции стоит знак «+», то связь положительная или прямая; если знак «-» - отрицательная, или обратная. Если коэффициент корреляции меньше 0,3 корреляционная зависимость считается слабой, больше 0,7 – сильной.

Коэффициент детерминации d_{yx} показывает долю изменений признака Y, которая зависит от изучаемого признака X. Коэффициент детерминации рассчитывается по формуле:

$$d_{yx}=r^2$$

Коэффициент регрессии b_{yx} показывает на какую величину и в каком направлении изменяется признак Y при изменении признака X на единицу изменения, рассчитывается по формуле:

$$b_{yx} = \frac{\sum (X - \bar{x}) \cdot (Y - \bar{y})}{\sum (X - \bar{x})^2}$$

Коэффициент регрессии b_{xy} показывает на какую величину и в каком направлении изменяется признак X при изменении признака Y на единицу изменения, рассчитывается по формуле:

$$b_{xy} = \frac{\sum (X - \bar{x}) \cdot (Y - \bar{y})}{\sum (Y - \bar{y})^2}$$

Для оценки значимости корреляционной связи рассчитывают **ошибку коэффициента корреляции s_r** и **критерий существенности t_r** .

Ошибку коэффициента корреляции рассчитывают по формуле:

$$s_r = \sqrt{\frac{1-r^2}{n-2}}$$

Критерий существенности определяют по формуле:

$$t_r = \frac{r}{s_r}$$

Критерий существенности сравнивают с критерием Стьюдента. Значение критерия Стьюдента берется из статистической таблицы с учетом принятого уровня значимости и числа степеней свободы, которое рассчитывается по формуле $\nu = n - 2$, n – число парных наблюдений.

Если $t_{\text{факт}} \geq t_{\text{теор}}$, связь между признаками значима; если $t_{\text{факт}} < t_{\text{теор}}$, связь незначима.

Уравнение линейной регрессии Y по X имеет следующий вид

$$Y = \bar{y} + b_{yx}(X - \bar{x})$$

где $\bar{y}$ и $\bar{x}$ - средние арифметические для ряда Y и X ,
 b_{yx} – коэффициент регрессии Y по X .

Пример. Зависимость урожайности кукурузы (Y , т/га) от засоренности посевов (X , шт/м²).

В результате наблюдений были получены следующие данные:

Количество сорных растений на 1 м ² (X)	12,0	18,1	25,2	30,0	32,5	40,0	45,4	50,6	51,0	62,6
Урожайность, т/га (Y)	6,6	6,4	6,3	6,2	6,2	6,0	5,8	5,2	4,7	4,0

Следует определить, существует ли зависимость между засоренностью посевов и урожайностью кукурузы.

1. Рассчитываем средние арифметические по ряду X (количество сорных растений на 1 м²) и ряду Y (урожайность кукурузы).

$$\bar{x} = \frac{\sum X}{n} = \frac{12,0 + 18,1 + 25,2 + 30,0 + 32,5 + 40,0 + 45,4 + 50,6 + 51,0 + 62,6}{10} = \frac{367,4}{10} = 36,7$$

$$\bar{y} = \frac{\sum Y}{n} = \frac{6,6 + 6,4 + 6,3 + 6,2 + 6,2 + 6,0 + 5,8 + 5,2 + 4,7 + 4,0}{10} = \frac{57,4}{10} = 5,7$$

2. Заполняем вспомогательную таблицу

Признаки		Отклонения		Квадраты отклонений		Произведения ($Y - \bar{y}$) · ($X - \bar{x}$)
Y	X	($Y - \bar{y}$)	($X - \bar{x}$)	($Y - \bar{y}$) ²	($X - \bar{x}$) ²	
6,6	12,0	0,9	-24,7	0,81	610,1	-22,2
6,4	18,1	0,7	-18,6	0,49	346,0	-13,0
6,3	25,2	0,6	-11,5	0,36	132,2	-6,9
6,2	30,0	0,5	-6,7	0,25	44,9	-3,4
6,2	32,5	0,5	-4,2	0,25	17,6	-2,1
6,0	40,0	0,3	3,3	0,09	10,9	1,0
5,8	45,4	0,1	8,7	0,01	75,7	0,9

5,2	50,6	-0,5	13,9	0,25	193,2	-7,0
4,7	51,0	-1,0	14,3	1,00	204,5	-14,3
4,0	62,6	-1,7	25,9	2,89	670,8	-44,0
-	-	-	-	$\Sigma=6,40$	$\Sigma=2305,9$	$\Sigma=-111$

3. Рассчитываем коэффициенты:

корреляции

$$r = \frac{\sum (Y - \bar{y}) \cdot (X - \bar{x})}{\sqrt{\sum (Y - \bar{y})^2 \cdot \sum (X - \bar{x})^2}} = \frac{-111}{\sqrt{6,40 \cdot 2305,9}} = \frac{-111}{\sqrt{14757,8}} = \frac{-111}{121,5} = -0,91$$

Коэффициент корреляции имеет знак «-», следовательно, корреляционная связь *обратная*, или *отрицательная*; числовое значение коэффициента корреляции больше 0,7, следовательно, корреляционная связь *сильная*.

детерминации $d_{yx} = r^2 = -0,91^2 = 0,83$ или 83%.

Коэффициент детерминации равен 83%, следовательно, варьирование результативного признака Y на 83% определяется влиянием факториального признака X.

регрессии: для расчета уравнения регрессии необходимо рассчитать один или два коэффициента регрессии: b_{yx} и b_{xy} .

Коэффициент регрессии b_{yx} показывает в каком направлении и на какую величину изменяется признак Y при увеличении признака X на единицу измерения. В нашем примере коэффициент регрессии b_{yx} показывает на какую величину изменяется урожайность кукурузы при увеличении количества сорных растений на 1 штуку на 1 м².

Коэффициент регрессии b_{xy} показывает в каком направлении и на какую величину изменяется признак X при увеличении признака Y на единицу измерения. В нашем примере коэффициент регрессии b_{xy} показывает на какую величину изменяется количество сорных растений на 1 м² при увеличении урожайности кукурузы на 1 т/га.

В нашем примере нет смысла рассчитывать коэффициент регрессии b_{xy} .

$$b_{yx} = \frac{\sum (Y - \bar{y}) \cdot (X - \bar{x})}{\sum (X - \bar{x})^2} = \frac{-111}{2305,9} = -0,05$$

Таким образом, при увеличении количество сорных растений на 1 штуку на 1 м² урожайность кукурузы снижается на 0,05 ц/га.

4. Проводим оценку существенности коэффициента корреляции.

ошибка коэффициента корреляции

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0,91^2}{10-2}} = \sqrt{\frac{1-0,83}{8}} = \sqrt{\frac{0,17}{8}} = \sqrt{0,021} = 0,14$$

критерий существенности $t_{факт} = \frac{r}{s_r} = \frac{-0,91}{0,14} = -6,5$

Знак критерия существенности не несет смысловой нагрузки, поэтому мы его не учитываем.

При $v = n - 2 = 10 - 2 = 8$ $t_{05} = 2,31$

Критерий существенности больше чем критерий Стьюдента, следовательно между изучаемыми признаками существует *значимая* или *существенная* корреляционная зависимость.

5. Выводим уравнение линейной регрессии

$$Y = \bar{y} + b_{yx} \cdot (X - \bar{x}) = 5,7 - 0,05(X - 36,7) = 5,7 - 0,05X + 1,8 = 7,5 - 0,05X$$

Таким образом, уравнение регрессии имеет следующий вид:

$$Y = 7,5 - 0,05X$$

Подставляя в полученное уравнение экстремальные значения X (т.е. X_{min} и X_{max}), определяем соответствующие им значения Y.

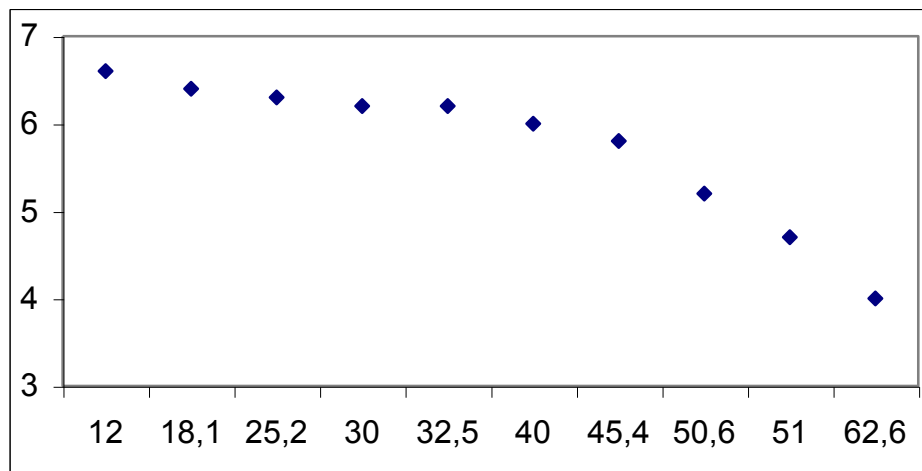
В нашем примере X_{min}=12,0, подставляем его в уравнение регрессии Y=7,5-0,05X=7,5-0,05×12,0=7,5-0,6=6,9, т.е. Y=6,9;

X_{max}=62,6, подставляем его в уравнение регрессии Y=7,5-0,05X=7,5-0,05×62,6=7,5-3,1=4,4, т.е. Y=4,4

6. Строим теоретическую линию регрессии Y по X.

На корреляционном поле размещаем точечные фактические значения данных эксперимента, но не соединяем их.

На основании полученных точек ($X_{\min}$, $Y_{\min}$ и $X_{\max}$, $Y_{\max}$) строим теоретическую линию регрессии. Теоретическая линия регрессии не должна выходить за пределы расчетных точек, т.к. полученное уравнение верно лишь для изучаемого диапазона значений факториального признака X .



Вывод. Установлена обратная значимая корреляционная зависимость между изучаемыми признаками. С увеличением засоренности урожайность кукурузы снижается.

Задание. Согласно N примера и задания провести корреляционно-регрессионный анализ.

Пример 1. Число развитых колосков в колосе (X , шт) и число зерен в колосе (Y , шт)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
38	18	28	19	30	13	32	17	18	10
29	12	39	15	22	11	21	9	28	13
32	13	25	13	28	9	34	15	32	15
42	17	29	13	38	15	27	12	41	17
30	16	27	14	32	13	41	17	32	14
33	15	26	12	35	15	31	14	28	13
36	16	34	15	26	13	28	13	31	14
41	16	40	16	31	14	31	14	29	13
30	13	34	14	22	10	36	16	31	15
28	13	29	16	31	14	31	14	18	9

Пример 2. Зависимость урожайности озимой пшеницы (Y, ц/га) от пораженности бурой ржавчиной (X, %)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
61,1	20	49,3	36	50,3	22	49,7	21	55,1	19
50,3	27	52,5	22	50,4	18	30,2	64	53,0	25
51,1	21	55,3	8	43,3	54	32,5	55	50,2	41
48,4	46	43,8	54	46,5	45	48,1	26	51,1	35
49,0	31	48,1	43	55,3	10	41,0	49	50,7	42
49,0	33	39,6	64	50,3	20	53,0	15	51,5	34
48,5	37	41,3	57	40,2	60	51,5	24	45,8	54
49,0	32	55,3	12	55,2	10	45,2	34	50,5	43
41,6	60	48,0	43	43,8	54	42,0	34	58,8	12
45,0	54	40,6	56	48,5	36	41,6	49	40,2	60

Пример 3. Зависимость прибавки урожая картофеля при внесении полного минерального удобрения(Y, ц/га) от количества осадков за май-июль (X, мм)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
150	235	46	113	45	121	203	245	30	60
136	204	125	212	84	142	65	153	33	74
42	120	77	145	83	138	74	138	47	94
216	238	201	247	143	221	34	84	42	99
37	96	112	176	136	198	90	150	88	150
95	145	112	188	100	168	80	137	60	143
82	140	37	88	95	140	33	74	144	220
48	119	42	95	103	178	46	112	45	108
42	115	39	83	31	78	103	165	27	64
96	156	54	113	36	85	136	198	42	100

Выполнение работы.

1. Записать название примера, номер задания.

2. Заполнить вспомогательную таблицу

Признаки		Отклонения		Квадраты отклонений		Произведения
Y	X	$(Y - \bar{y})$	$(X - \bar{x})$	$(Y - \bar{y})^2$	$(X - \bar{x})^2$	$(Y - \bar{y}) \cdot (X - \bar{x})$
Σ	Σ	-	-	Σ	Σ	Σ

3. Рассчитать средние арифметические по ряду X и Y.

n =

n – количество парных значений

$$\bar{x} = \sum X \div n =$$

$$\bar{y} = \sum Y \div n =$$

4. Рассчитать коэффициенты:

корреляции

$$r = \frac{\sum (Y - \bar{y}) \cdot (X - \bar{x})}{\sqrt{\sum (Y - \bar{y})^2 \cdot \sum (X - \bar{x})^2}} =$$

регрессии
$$b_{yx} = \frac{\sum (Y - \bar{y}) \cdot (X - \bar{x})}{\sum (X - \bar{x})^2} =$$

детерминации $d_{yx} = r^2 =$

5. Провести оценку существенности коэффициента корреляции.

ошибка коэффициента корреляции
$$S_r = \sqrt{\frac{1 - r^2}{n - 2}} =$$

критерий существенности
$$t_{факт} = \frac{r}{S_r} =$$

При $v = n - 2 =$ $t_{05} =$

6. Вывести уравнение линейной регрессии

$$Y = \bar{y} + b_{yx} \cdot (X - \bar{x}) =$$

Подставляя в полученное уравнение экстремальные значения X, определяем значения Y и строим теоретическую линию регрессии Y по X.



Вывод:

Значение критерия t на 5 и 1% уровне значимости

Число степеней свободы	Уровень значимости	
	05	01
1	12.71	63.33
2	4.30	9.93
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.37	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3.17
11	2.20	3.11
12	2.18	3.06
13	2.16	3.01
14	2.15	2.98
15	2.13	2.95
16	2.12	2.92
17	2.11	2.90
18	2.10	2.88
19	2.09	2.86
21	2.08	2.83
22	2.07	2.82
23	2.07	2.81
24	2.06	2.80
25	2.06	2.79
26	2.06	2.78
27	2.05	2.77
28	2.05	2.76
29	2.05	2.76
30	2.04	2.75
50	2.01	2.68
100	1.98	2.63
более 100	1.96	2.58

Значение критерия F на 5-ном уровне значимости

Степени свободы для меньшей дисперсии (знаменатель)	Степени свободы для большей дисперсии (числитель)							
	1	2	3	4	5	6	7	8
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
6	5.99	5.14	4.76	4.33	4.39	4.27	4.21	4.15
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85
13	4.64	3.80	3.41	3.18	3.02	2.92	2.84	2.77
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70
15	4.54	3.60	3.29	3.06	2.90	2.70	2.70	2.64
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45

Литература

1. Доспехов Б.А. Методика полевого опыта. М.: Агропромиздат, 1985.
 2. Методика полевого опыта в овощеводстве и бахчеводстве (Под ред. В.Ф.Белика). М.: Агропромиздат, 1974.
 3. Опытное дело в полеводстве. М.: Россельхозиздат, 1982.
 4. Пересыпкин В.Ф., Коваленко С.И. Практикум по методике опытного дела в защите растений. М.: Агропромиздат, 1989.
 5. Лакин Г.Ф. Биометрия. М.: Высшая школа, 1980.
- Потапов В.А., Качин В.И., Курсоков А.Г. Методы обработки экспериментальных данных в полеводстве. М.: Колос, 1997.

Содержание

Введение.....	3
Лабораторная работа № 1. Статистическая обработка результатов наблюдений при количественной изменчивости.....	3
Лабораторная работа № 2. Оценка существенности разности средних двух выборок при количественной изменчивости.....	10
Лабораторная работа № 3. Дисперсионный анализ данных однофакторного полевого опыта, заложенного методом рэндомизированных повторений.....	16
Лабораторная работа № 4. Дисперсионный анализ данных двухфакторного полевого опыта, заложенного методом рэндомизированных повторений.....	25
Лабораторная работа № 5. Прямолинейная корреляция и регрессия.....	37
Литература.....	45
Приложения.....	46

Учебное издание

*ТРЕСКИНА НАТАЛЬЯ НОВОМИРОВНА
КОВАЛЬ ЕФИМИЙ МИХАЙЛОВИЧ*

***ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ
В АГРОНОМИИ***

Методические указания
по выполнению лабораторных работ по курсу
для студентов заочного отделения специальностей
«Агрономия», «Плодоовощеводство и виноградарство»,
«Защита растений», «Технология производства и
переработки сельскохозяйственной продукции»

Компьютерная верстка и набор *Трескина Н.Н.*
Усл.-печатных . листов 2,0. Тираж 50 экз.