

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Комин Андрей Эдуардович

Должность: ректор

Дата подписания: 15.07.2026 11:45:47

Уникальный программный ключ:

f6c6d686f0c899fdkca1ed01140b0c8a16f116f657d9a0c7e9

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Приморская государственная сельскохозяйственная академия»
Институт землеустройства и агротехнологий

Кафедра физики
и высшей
математики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания по дисциплине (модулю) к аудиторным занятиям и самостоятельной работе для обучающихся по направлениям подготовки: 35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение; 35.03.07 Технология производства и переработки сельскохозяйственной продукции

Электронное издание

Уссурийск, 2019

Составитель: Савельева Е.В., кандидат технических наук, доцент кафедры физики и высшей математики.

Математическая статистика: методические указания по дисциплине (модулю) к аудиторной и самостоятельной работе для обучающихся по направлениям подготовки 35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение; 35.03.07 Технология производства и переработки сельскохозяйственной продукции [Электронный ресурс]: / Е.В. Савельева; ФГБОУ ВО ПГСХА. - Электрон. текст. дан. – Уссурийск: ПГСХА, 2019.- 77 с. - Режим доступа: [www. de.primacad.ru](http://www.de.primacad.ru).

Рецензент: Лосев А.С., к.ф.-м. наук, доцент Федерального государственного бюджетного учреждения науки Института прикладной математики Дальневосточного отделения РАН

Методические указания составлены в соответствии с требованиями стандарта ФГОС 3++ по направлениям подготовки 35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение; 35.03.07 Технология производства и переработки сельскохозяйственной продукции, содержат задания для самостоятельной работы обучающихся и методические указания по их выполнению.

Издается по решению методического совета ФГБОУ ВО Приморская ГСХА

ВВЕДЕНИЕ

Основная задача математической статистики сбор и обработка экспериментальных данных, с целью изучения закономерностей случайных явлений, носящих массовый характер. Например, распределение пахотных земель по урожайности. Обследование всей совокупности заменяют обследованием выборки из генеральной совокупности. Желательно, чтобы результаты обследования выборки отражали характерные, основные черты изучаемого признака; для этого выборка должна быть репрезентативной.

В связи с возрастающей ролью вероятностно-статистических методов в сельском хозяйстве, задания составлены таким образом, чтобы обучающийся овладел практическими навыками по использованию статистических методов обработки и анализа сельскохозяйственных показателей.

В методических указаниях рассматривается четыре основных раздела дисциплины (модуля):

- 1) исследование вариационных рядов;
- 2) проверка статистических гипотез;
- 3) дисперсионный анализ;
- 4) корреляционный и регрессионный анализ.

В каждом разделе даны индивидуальные задания для самостоятельной работы и методические указания для их выполнения.

Методические указания могут быть использованы для обучающихся по направлениям аграрных технологий для любых форм обучения.

РАЗДЕЛ 1. Исследование вариационных рядов.

1.1. Вопросы для самопроверки.

1. Понятие генеральной и выборочной совокупности. Повторная и бесповторная выборки. Репрезентативная выборка.
2. Понятие варианты и частоты (веса) варианты. Относительная частота варианты. Построение таблицы статистического распределения.
3. Дискретный и интервальный вариационные ряды: понятие и алгоритмы построения.
4. Накопленная частота и накопленная относительная частота. Эмпирическая функция распределения.
5. Полигон и гистограмма (частот и относительных частот). Кумулятивная кривая.
6. Выборочное среднее (простое и взвешенное). Мода, ее нахождение для дискретного и интервального вариационных рядов. Медиана, её вычисление для дискретного и интервального вариационных рядов.
7. Среднее линейное отклонение (простое и взвешенное). Дисперсия (простая и взвешенная). Среднее квадратическое отклонение (простое и взвешенное).
8. В чём отличие точечной оценки числовой характеристики от интервальной?
9. Определение доверительного интервала. Определение доверительной вероятности и уровня значимости.
10. Как определяется предельная ошибка выборки при нахождении доверительного интервала для генеральной средней?
11. Нахождение доверительного интервала для генеральной средней по большой выборке ($n > 30$).
12. Построение доверительного интервала для $X_{ген.}$ по малой выборке ($n < 30$).
13. Как связаны величина доверительного интервала и доверительная вероятность?

1.2. Индивидуальное задание №1

Задание №1. Исследование дискретных вариационных рядов.

- 1) Составить вариационный дискретный ряд;
- 2) Построить полигон;
- 3) Вычислить числовые характеристики: а) характеристики положения - среднюю выборочную, моду, медиану; б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации;
- 4) Вычислить показатели точности выборочной средней;
- 5) Найти доверительные интервалы для генеральной средней с надежностью 0,99;0,95.
- 6) Составить содержательную интерпретацию результатов первичной обработки

ВАРИАНТ № 1 – 5. Представлены данные густоты всходов кукурузы на 10 м²

№1		№2		№3		№4		№5	
58	57	64	61	60	65	63	58	57	65
60	59	61	58	58	66	67	63	59	66
59	60	58	60	65	61	60	60	60	61
63	59	59	64	61	62	57	60	57	60
58	52	60	59	58	61	59	58	52	61
62	59	60	63	60	60	64	59	59	60
63	59	62	63	61	62	63	64	59	65
59	58	64	63	57	59	59	58	54	57
62	63	62	62	60	62	57	56	63	63
63	57	63	61	59	61	59	60	54	61

ВАРИАНТ № 6 – 10. Представлены данные высоты растений пшеницы

№6		№7		№8		№9		№10	
27	32	32	31	32	28	37	35	32	28
32	39	32	32	37	26	27	32	35	33
31	43	26	35	33	27	35	32	32	35
35	28	32	33	32	32	37	33	28	32
36	36	33	31	23	26	33	32	36	33
35	39	36	38	28	32	31	31	35	33
26	28	36	27	32	35	28	43	33	35
35	39	35	33	28	33	35	28	32	36
32	35	31	35	36	27	33	33	33	28

28	31	32	27	35	33	32	32	36	38
----	----	----	----	----	----	----	----	----	----

ВАРИАНТ №11 - 15. Дана выборочная совокупность урожайности картофеля в ц/га

№11		№12		№13		№14		№15	
143	126	148	153	142	142	164	132	138	140
143	126	140	130	139	124	127	137	139	131
128	120	120	138	130	114	125	151	120	131
134	145	140	150	124	142	132	137	140	131
143	143	140	140	135	139	125	137	131	120
126	103	120	124	135	111	132	132	100	132
143	134	138	130	135	130	137	151	107	107
107	107	148	148	136	139	172	137	138	120
134	145	124	138	139	139	119	120	138	142
145	134	148	148	150	143	137	138	131	138

ВАРИАНТ № 16 – 20. Данные по высоте растений ячменя

№16		№17		№18		№19		№20	
53	53	52	55	56	49	50	52	52	55
54	53	52	53	51	55	53	55	53	54
51	54	56	54	54	53	54	54	55	53
52	55	53	53	56	53	53	56	51	52
56	55	50	54	52	54	54	55	54	55
54	51	55	53	55	54	52	54	53	54
54	56	54	55	54	53	56	55	54	53
55	52	53	52	51	55	53	54	51	50
53	54	55	52	55	52	53	50	53	52
58	57	56	58	55	57	56	58	57	57

Задание 2. Исследование интервальных рядов.

Требуется:

- 1) Составить интервальный ряд распределения;
- 2) Построить полигон и гистограмму;
- 3) Вычислить числовые характеристики: а) характеристики положения - среднюю выборочную, моду, медиану; б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации.
- 4) Вычислить числовые характеристики упрощенным способом, с помощью метода условного нуля;
- 5) Найти показатели точности выборочной средней;
- 6) Найти доверительные интервалы для генеральной средней с надежностью

0,99;0,95.

7) Составить содержательную интерпретацию результатов первичной обработки

ВАРИАНТ 1. Дана выборочная совокупность урожайности гречихи в ц/га.

5,7	5,1	4,6	8,1	9,6	6,6	9,8	8,9	8,0	5,5
10,6	5,6	4,7	5,5	7,4	9,5	5,9	6,2	4,9	4,9
6,6	5,3	6,5	4,5	7,3	9,5	6,6	5,3	5,3	6,6
9,4	5,5	8,4	6,6	4,4	4,9	7,3	7,8	9,8	9,0
4,4	5,7	6,8	8,6	6,9	7,2	6,2	5,6	6,9	5,6
6,7	9,6	6,2	7,7	9,8	9,7	7,9	5,5	8,6	8,1
7,4	6,6	5,6	7,8	4,7	7,5	6,1	4,8	6,5	7,8
9,0	7,2	8,7	5,3	6,4	5,5	8,3	8,5	7,9	8,4

ВАРИАНТ 2. Дана выборочная совокупность массы 10 растений сои в граммах на опытном поле.

12	13	12,5	12	14	11	12	13	12
11	13,5	13	11	13	12,5	10	12	12,5
13	12	10	10	12	13	13	13	15
12	11	11	13	13	12	13,5	11	13
12	12	12	14	12	10	12,5	12	12
14	13	12	12	12,5	11	13	11,5	13,5
13	15	13	11	15	12	12	12,5	13

ВАРИАНТ 3. Высота растений пшеницы с использованием в качестве удобрений золу Лучегорской ГРЕС.

30	23	26	28	30	29	25	28
30	27	24	29,5	37	29	24	23
21	30	31	29	29	23	24	26
27	28	30	28	24	26	23	26
22	28	29	25	29	23	22	28
31	23	28	29	30	27	25	27
26	21	37	25	23	24	30	31

ВАРИАНТ 4. Дана выборочная совокупность густоты всходов кукурузы на 10 м^2 .

117	10	62	71	108	100	107	74
88	113	68	103	98	103	98	86
87	131	74	86	97	120	87	99
72	99	89	84	82	100	88	88
110	100	101	90	100	99	76	72
106	103	98	99	104	79	110	108
103	108	87	79	110	89	105	103
89	98	73	92	83	98	106	104

ВАРИАНТ 5. Дана выборочная урожайность сои по Приморскому краю ц/га.

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

ВАРИАНТ 6. Дана выборочная совокупность учета всходов ячменя с 1 м^2 на опытном поле.

72	63	45	33	49	22	42	48	52	50
50	59	30	65	50	45	52	53	71	52
56	50	59	41	57	51	41	62	68	81
27	64	49	57	50	46	35	80	77	60
74	65	48	89	64	50	49	64	84	69
59	51	55	37	46	39	56	55	75	68
77	59	72	35	35	68	54	71	77	68

ВАРИАНТ 7. Данные по высоте растений кукурузы, начиная с третьего листа на 10 м^2 на опытном поле.

64	78	86	81	77	78	80
70	79	75	80	85	75	77
73	80	81	76	81	78	83
80	80	86	78	84	77	81
86	81	87	73	81	86	85
78	86	78	81	80	90	80
76	85	79	85	77	91	70
80	75	73	80	83	78	81

ВАРИАНТ 8. Дана выборочная совокупность урожайности картофеля в ц/га по Приморскому краю.

92	86	71	89	80	84	75	82
91	82	60	91	81	70	82	77
96	93	65	80	80	95	94	77
100	79	81	89	78	75	98	83
88	88	90	95	77	90	100	90
82	81	86	82	74	82	94	75
81	85	96	70	75	80	75	92

ВАРИАНТ 9. Дана выборочная совокупность массы 10 растений сои в граммах на опытном поле.

12	10	12	12	10	15	14,5	9	15
13	13	13	13	14	10	10	9,5	14,5
12	11	12	10	12	12	15	10	10
13	13	12	16	17	15	12	8	10
13	11	12	13	14	12	13	11	15
13	9	11	13	12	12	12	13	12
12	13	14	15	12	10	11	16	13

ВАРИАНТ 10. Даны сведения об урожайности сена многолетних трав в ц/га.

55,9	26,0	39,3	69,3	59,8
32,0	32,9	57,3	28,2	54,1

29,1	58,1	17,4	45,9	71,8
28,4	43,4	59,3	39,0	49,4
40,1	50,5	26,8	34,3	66,0
17,1	47,5	25,1	38,8	30,6
35,5	60,2	21,5	34,1	59,5
49,9	45,0	27,3	53,8	70,5

ВАРИАНТ 11. Дана выборочная совокупность урожайности проса в ц/га.

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

ВАРИАНТ 12. Посевные участки равные по площади дали урожай хлопка - сырца в ц .

20	21	23	18	27	25	25	26	20	22
22	20	18	27	26	25	24	24	20	21
23	20	21	22	24	25	20	18	19	20
21	22	22	23	25	24	24	25	26	28
20	21	21	20	18	19	21	22	23	28

ВАРИАНТ 13. Данные по высоте растений ячменя на делянках с различной обработкой почвы

79	83	81	78	81	72	76
80	80	64	77	76	76	73
77	79	63	83	83	83	78
83	84	85	80	89	85	80

74	73	79	77	90	88	79
81	79	80	64	72	73	83
76	77	83	90	76	80	78

ВАРИАНТ 14. Высота растений пшеницы с использованием в качестве удобрений золу Лучегорской ГРЭС.

29	26	27	30	37	28	25	27	28
28	31	24	28	22	27	30	29	24
26	23	29	24	33	26	34	31	26
30	31	26	25	25	22	28	32	27
19	25	29	30	29	28	21	18	26
26	24	31	28	36	27	31	29	24
24	25	26	22	31	29	28	27	29

ВАРИАНТ 15. Дана выборочная совокупность урожайности овса в ц/га.

15,8	14,1	14,0	10,5	15,2	9,0	16,3	13,4
9,3	14	11,6	13,0	17,6	13,4	18,5	13,9
16,0	11,2	11,2	10,1	10,0	17,7	8,2	16,5
18,4	16,3	18,5	14,5	10,5	15,2	10,3	10,2
16,0	13,8	10,2	14,8	12,4	10,7	14,9	15,1
15,5	15,7	11,8	16,5	15,0	8,0	14,3	9,7
11,9	18,3	14,7	15,0	16,2	11,1	15,2	11,3
15,4	14,2	14,4	15,1	18,3	17,9	12,5	18,4

ВАРИАНТ 16. Данные по высоте растений ячменя на делянках с различной обработкой почвы.

90	65	86	82	86	83	83	78
83	71	87	84	87	82	77	79
78	78	80	83	87	83	80	83
79	73	88	77	79	75	81	82
86	81	90	85	80	80	83	86
81	83	85	82	75	90	93	88

83	82	87	79	86	91	84	86
90	93	82	80	89	77	80	95

ВАРИАНТ 17. Учёт всходов кукурузы в фазе 3-го листа на 10 погонных метрах на опытном поле.

59	84	98	63	50	77	54
75	100	83	54	52	66	59
82	82	59	68	66	104	80
80	64	70	63	60	87	49
69	45	75	69	62	70	50
60	100	83	66	104	54	59
63	70	54	62	87	80	70
64	75	63	60	70	59	75

ВАРИАНТ 18. Даны сведения об урожайности пшеницы (ц/га).

14,3	10,1	15,2	19,9	15,9	8,4	17,7	12,6
19,1	15,7	12,1	16,9	21,7	10,4	16,3	14,5
13,5	13,8	18,6	9,7	19,0	12,6	17,8	14,4
12,2	10,3	20,6	17,3	12,3	10,2	12,5	14,8
21,7	15,1	13,3	13,7	18,2	9,0	16,4	12,1
19,2	14,4	13,5	18,3	10,9	11,8	14,6	15,7
13,0	17,4	10,4	8,2	14,0	18,4	14,8	10,5
17,4	13,9	12,7	11,5	13,5	15,7	12,5	20,3
11,8	18,8	15,6	8,8	16,2	14,8	10,7	9,3

ВАРИАНТ 19. Дана выборочная совокупность урожайности риса в ц/га.

43,6	50,6	42,0	35,4	37,9	38,6	47,1	32,2
25,2	40,0	38,9	5,5	37,8	37,9	35,8	36,2
46,6	42,4	38,2	36,0	28,9	36,5	38,1	35,5
36,2	39,2	39,1	38,2	33,8	30,5	39,9	39,2
30,8	39,1	29,8	39,0	27,8	30,4	39,4	38,9
41,9	39,0	30,0	39,3	33,0	43,2	48,5	38,8

35,0	34,9	38,7	27,8	39,3	38,7	49,9	35,5
24,0	35,3	32,8	28,9	37,2	40,6	30,7	34,2

ВАРИАНТ 20. Дана выборочная совокупность урожайности сои по Приморскому краю в ц/га.

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

1.3. Решение типового примера.

Пример 1.1. Исследование дискретных рядов

При взвешивании новорождённых телят холмогорской породы были получены следующие данные ($n=20$; $\min=23$; $\max=45$):

42,26,30,33,28,35,40,33,35,28,37,30,40,30,33,33,37,37,33,30.

Требуется:

- 1) Составить вариационный дискретный ряд;
- 2) Построить полигон;
- 3) Вычислить числовые характеристики: а) характеристики положения - среднюю выборочную, моду, медиану; б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации;
- 4) Вычислить показатели точности выборочной средней;
- 5) Найти доверительные интервалы для генеральной средней с надёжностью 0,99;0,95.
- 6) Составить содержательную интерпретацию результатов первичной обработки

1) Составим дискретный ряд:

подсчитаем объем выборки $n=20$;

найдем: $x_{\min}=26$; $x_{\max}=42$,

ранжируем ряд: 26,28,28,30,30,30,33,33,33,33,33,35,35,35,35,37,37, 37,40,42

№	x_i —вес телят (кг)	частоты f_i	Относительные частоты ω_i
1	26	1	1/20
2	28	2	2/20
3	30	3	3/20
4	33	5	5/20
5	35	4	4/20
6	37	3	3/20
7	40	1	1/20
8	42	1	1/20
ИТОГО		20	1

2) Построим полигон распределения – по оси ОХ отложим значения веса телят, по ОУ значения частот (рис.1.1)

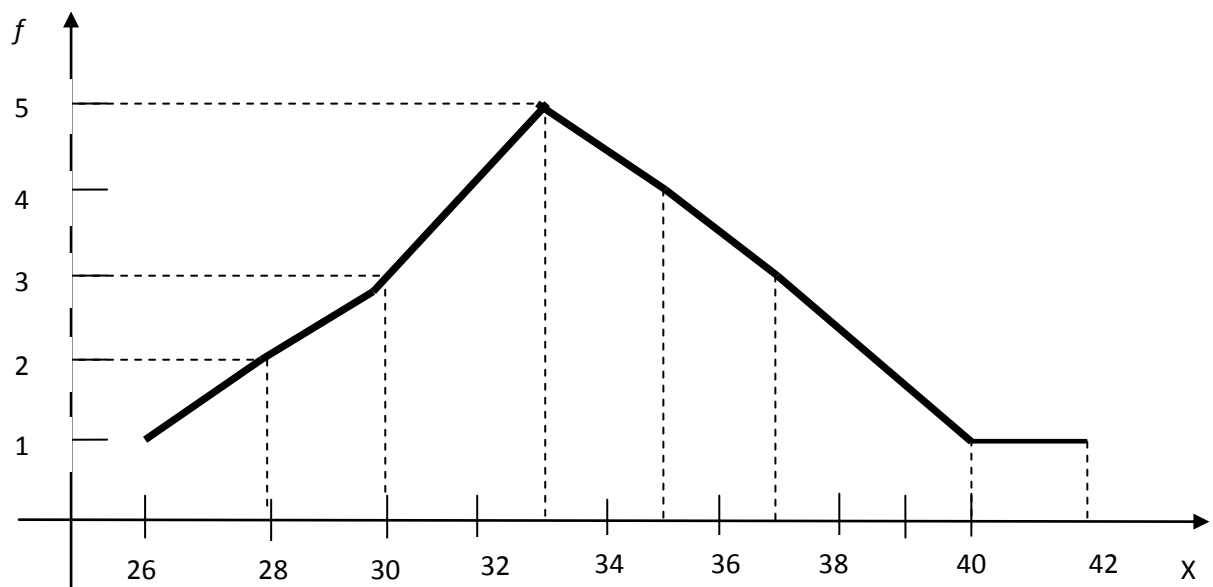


Рисунок 1.1

3) **Вычислить числовые характеристики:**

а) характеристики положения - среднюю выборочную, моду, медиану.

б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации;

Для подсчета числовых характеристик составим расчетную таблицу 1.1:

Таблица 1.1.

№	x_i	частота f_i	$x_i f_i$	$x_i - x_{\bar{e}}$	$(x_i - x_{\bar{e}})^2$	$(x_i - x_{\bar{e}})^2 f_i$
1	26	1	26	-5,4	29,2	29,2
2	28	2	56	-3,4	11,6	23,1
3	30	3	90	-1,4	2,0	5,9
4	33	5	165	1,6	2,6	12,8
5	35	4	140	3,6	13,0	51,8
6	37	3	111	5,6	31,4	94,1
7	40	1	40	8,6	74,0	74,0
8	42	1	42	10,6	112,4	112,4
Σ		20	628			403,36

а) характеристики положения - среднюю выборочную, моду, медиану.

1. Средняя взвешенная (итоговую сумму 4 –ого столбца разделим на объем выборки =20).

$$\bar{x}_{\bar{e}} = \frac{\sum_{i=1}^n x_i f_i}{n}; \quad \bar{x}_{\bar{e}} = \frac{628}{20} = 31,4 \approx 31 \text{ кг}$$

2. Вычислим моду – для дискретного ряда признак с наибольшей частотой

$$M_o = 33 \text{ кг}$$

3. Вычислим медиану - для дискретного ряда:

$$M_e = \begin{cases} \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right), & \text{если } n - \text{четное;} \\ x_{\frac{n+1}{2}}, & \text{если } n - \text{нечетное.} \end{cases}$$

В нашем случае $n=20$ – четное, тогда получим:

$$M_e = 1/2(x_{n/2} + x_{n/2+1}) = 1/2(x_{10} + x_{11}) = 1/2(33 + 33) = 33$$

$$M_e = 33 \text{ кг}$$

б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации;

1. Выборочную дисперсию (итоговую сумму 7-ого столбца разделим на объем выборки $=20$).

Т.к. выборка малая $n=20 < 30$, вычислим дисперсию по упрощенной формуле: $S^2 = \frac{\sum (x_i - \bar{x}_e)^2 f_i}{n-1}$ $S^2 = \frac{403,36}{19} \approx 21$

2. Среднее квадратическое отклонение

$$S_e = \sqrt{\frac{\sum (x_i - \bar{x}_e)^2 \cdot f_i}{n-1}}; \quad S_e = \sqrt{21} \approx 4 \text{ кг}$$

3. Коэффициент вариаций

$$V_e = \frac{S_e}{\bar{x}_e} \cdot 100\%; \quad V_e = \frac{4}{31} \cdot 100\% \approx 12\%$$

Коэффициент вариаций меньше 20%, следовательно, выборка слабо изменчива, однородна.

Вывод: Среднее квадратическое отклонение показывает, что вес по данной совокупности колеблется в пределах ± 4 кг., а коэффициент вариации равен $\pm 12\%$ по отношению к среднему уровню, что указывает на среднюю изменчивость.

4) показатели точности выборочной средней;

$$\text{Абсолютная ошибка средней: } S_{\bar{x}} = \frac{S_e}{\sqrt{n}} = \frac{4}{\sqrt{20}} \approx 0,9;$$

$$\text{Относительная ошибка выборочной средней: } S_{\bar{x}\%} = \frac{S_{\bar{x}}}{\bar{x}_e} \cdot 100\% = \frac{0,9}{31} \cdot 100\% \approx 3\%$$

Вывод: в нашем опыте $S_{\bar{x}}\% \approx 3\% < 6\%$, следовательно, точность исследования высокая, поэтому полученные выводы отображают сущность изучаемого явления.

5) Найти доверительные интервалы для генеральной средней с надежностью 0,99; 0,95.

Доверительный интервал для генеральной средней (малая выборка $n < 30$) имеет вид:

$$\bar{x}_в - \frac{t_{\gamma, n-1} \cdot S_в}{\sqrt{n}} \leq X_{ген.} \leq \bar{x}_в + \frac{t_{\gamma, n-1} \cdot S_в}{\sqrt{n}}.$$

$$\bar{x}_в - t_{\gamma, n-1} \cdot S_{\bar{x}} \leq X_{ген.} \leq \bar{x}_в + t_{\gamma, n-1} \cdot S_{\bar{x}}.$$

1. Найдем доверительный интервал с надежностью $\gamma = 0,95$, подставим ошибку средней $S_{\bar{x}} \approx 0,9$; величину $t_{\gamma, n-1}$ найдем по таблице Стьюдента при $\gamma = 0,95$, $n-1 = 20-1 = 19$, получим: $t_{0,95,19} = 2,09$

Получим:

$$31 - 2,09 \cdot 0,9 \leq X_{ген.} \leq 31 + 2,09 \cdot 0,9 \Rightarrow 29,92 \leq X_{ген.} \leq 33,82$$

2. Аналогично для $\gamma = 0,99$, найдем $t_{0,99,19} = 2,86$, получим доверительный интервал:

$$31 - 2,86 \cdot 0,9 \leq X_{ген.} \leq 31 + 2,86 \cdot 0,9 \Rightarrow 29,43 \leq X_{ген.} \leq 33,37$$

Вывод: доверительный интервал генеральной средней с надежностью 95% - (29,92; 33,82);

с надежностью 99% интервал (29,43; 33,37) покрывает генеральную среднюю веса телят.

6. Содержательная интерпретация результатов первичной обработки по условию задачи

1. Интерпретация $\bar{x}_g$.

Вес телят – величина случайная, ее выборочные значения изменяются в частности от 26 кг до 42 кг, однако среднее значение равно 31 кг

2. Интерпретация среднего квадратического отклонения S_g .

Так как среднее квадратическое отклонение $S_g=4$ кг., то можно сделать вывод, что отклонение веса отдельно взятого теленка в среднем составляет ± 4 кг.

3. Интерпретация значения коэффициента вариации V .

Значение $V=12\%$ - это больше 10%, что указывает на среднюю изменчивость.

4. Интерпретация доверительного интервала

$$\bar{x}_g - \frac{t_{\gamma} \cdot s}{\sqrt{n}} < a < \bar{x}_g + \frac{t_{\gamma} \cdot s}{\sqrt{n}}.$$

Так как $29\text{кг} \leq X_{ген.} \leq 33\text{кг}$ при заданной надежности $\gamma=0,95$, то можно утверждать с вероятностью 0,95, что средний вес телят генеральной совокупности будет колебаться в пределах от 29кг до 33кг, то есть из 100 телят примерно 95 будут иметь средний вес в указанных пределах;

С вероятностью 99 % можно утверждать, что средняя генеральная веса телят будет находится в интервале (24кг; 38кг)

Пример 1.2. Исследование интервальных вариационных рядов.

Даны сведения об урожайности пшеницы (ц/га).

14,3	10,1	15,2	19,9	15,9	8,4	17,7	12,6
19,1	15,7	12,1	16,9	21,7	10,4	16,3	14,5
13,5	13,8	18,6	9,7	19,0	12,6	17,8	14,4
12,2	10,3	20,6	17,3	12,3	10,2	12,5	14,8
21,7	15,1	13,3	13,7	18,2	9,0	16,4	12,1
19,2	14,4	13,5	18,3	10,9	11,8	14,6	15,7
13,0	17,4	10,4	8,2	14,0	18,4	14,8	10,5

17,4	13,9	12,7	11,5	13,5	15,7	12,5	20,3
11,8	18,8	15,6	8,8	16,2	14,8	10,7	9,3

Требуется:

- 1) Составить интервальный ряд распределения;
- 2) Построить полигон и гистограмму;
- 3) Вычислить числовые характеристики: а) характеристики положения - среднюю выборочную, моду, медиану; б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации.
- 4) Вычислить числовые характеристики упрощенным способом, с помощью метода условного нуля;
- 5) Найти показатели точности выборочной средней;
- 6) Найти доверительные интервалы для генеральной средней с надежностью 0,99; 0,95.
- 7) Составить содержательную интерпретацию результатов первичной обработки

Решение.

1) Составим вариационный ряд.

Так как объем выборки $n=72 > 30$ и значения признака в выборке не повторяются, то составляем интервальный ряд распределения.

Для составления интервального ряда (таблица 1.2) выполним следующие действия:

- 1)** Найдем длину (шаг) интервала по формуле $h = \frac{x_{\max} - x_{\min}}{1 + 3,2 \lg n}$, для этого

находим в выборке $x_{\max}$ и $x_{\min}$: $x_{\max} = 21,7$, $x_{\min} = 8,2$

Получим: $h = \frac{21,7 - 8,2}{1 + 3,2 \cdot \lg 72} \approx 2$

- 2)** В первый столбец записываем интервалы, начало первого интервала

$a_1 = x_{\min} = 8,2$ и конец $b_1 = a_1 + h = 8,2 + 2 = 10,2$; $a_2 = b_1$; $b_2 = b_1$ и т.д.

составляем столько интервалов, пока в последний интервал не попадет наибольшее значение, т.е. $x_{\max} = 21,7$

Таблица 1.2.

№ п/п	$a_i - b_i$	сводка	m_i	x_i	S_H
1	[8,2-10,2)	☐┐	7	9,2	7
2	[10,2-12,2)	☐☐┐	12	11,2	19
3	[12,2-14,2)	☐☐☐	16	13,2	35
4	[14,2-16,2)	☐☐☐	15	15,2	50
5	[16,2-18,2)	☐☐	9	17,2	59
6	[18,2-20,2)	☐☐	9	19,2	68
7	[20,2-22,2)	☐	4	21,2	72
Σ			72		

3) В третий столбец запишем значения частот, для этого нужно просмотреть по одному все элементы выборки в протоколе наблюдений, и подсчитать количество значений, соответствующих каждому интервалу (то есть m_i), так в первый интервал [8,2-10,2) попадает 7 значений выборочной совокупности.

Для облегчения подсчета частот составляют второй столбец (сводка). Границы интервалов выписывают в столбец. Затем просматривают данные, записанные в том порядке, в котором они были получены. Правее интервала, в который попало значение, черточку. Черточки удобно ставить так, чтобы пять попаданий образовывали «конверт», для каждого интервала

подсчитывается число полных «конвертов» и число черточек.

4) В четвертый столбец записываем середины полученных интервалов, которые рассчитываются по формуле: $x_i = \frac{a_i + b_i}{2}$, середина первого интервала

$$x_1 = \frac{a_i + b_i}{2} = \frac{8,2 + 10,2}{2} = 9,2 \text{ и т.д.}$$

5) В пятый столбец записывают значение накопленной частоты.

Накопленной частотой S_H в точке x называют суммарную частоту членов статистической совокупности со значениями признака, меньшими, чем x .

У первого интервала она равна частоте $S_{H_1} = m_1$; $S_{H_2} = m_1 + m_2$; $S_{H_3} = m_1 + m_2 + m_3 \dots\dots$

Накопленная частота последнего интервала равна объему выборки.

2) Построим полигон и гистограмму.

Представляем данный интервальный ряд графически; строим многоугольник распределения и гистограмму. По горизонтальной оси откладываем интервалы, по вертикальной – частоту.

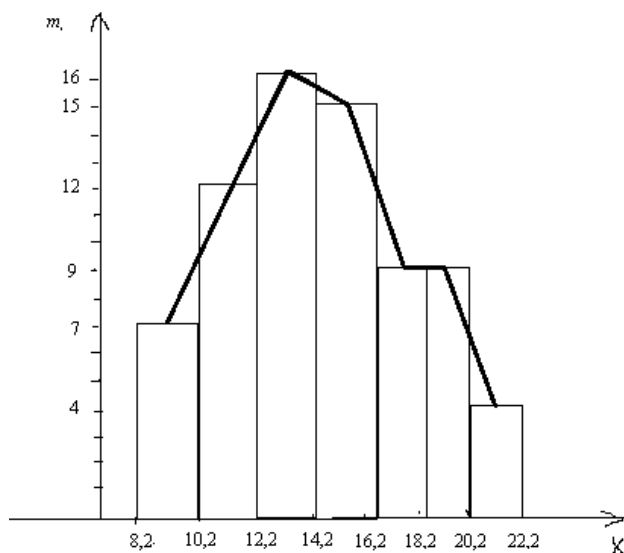


Рисунок 1.2

Многоугольник распределения строится по точкам: (9,2;7); (11,2;12); (13,2;16); (15,2;15); (17,2;9); (19,2;9); (21,2;4) – середины верхних оснований

прямоугольников. Точки соединяются прямыми линиями, полученную ломаную линию называют многоугольником распределения.

Гистограмма состоит из прямоугольников, основаниями которых являются интервалы, а высотами соответствующие частоты (рис. 1.2).

Данные гистограммы показывают характерную для многих признаков форму распределения: чаще встречаются значения средних интервалов признака, реже – крайние (малые и большие) значения признака. Форма этого распределения близка к нормальному закону распределения, которое образуется, если на варьирующую переменную влияет большое число факторов, ни один из которых не имеет преобладающего значения.

3) Вычислить числовые характеристики:

- а) характеристики положения - среднюю выборочную, моду, медиану.
- б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации;

Для подсчета числовых характеристик составим расчетную таблицу 1.2:

Таблица 1.2

№	x_i	частота f_i	$x_i f_i$	$x_i - x_{\bar{e}}$	$(x_i - x_{\bar{e}})^2$	$(x_i - x_{\bar{e}})^2 f_i$
1	9,2	7	64,4	-5,39	29,0521	203,3647
2	11,2	12	134,4	-3,39	11,4921	137,9052
3	13,2	16	211,2	-1,39	1,9321	30,9136
4	15,2	15	228	0,61	0,3721	5,5815
5	17,2	9	154,8	2,61	6,8121	61,3089
6	19,2	9	172,8	4,61	21,2521	191,2689
7	21,2	4	84,8	6,61	43,6921	174,7684
Σ	106,4	72	1050,4	4,27	114,6047	805,1112

а) характеристики положения - среднюю выборочную, моду, медиану.

1) Среднее выборочное:

$$\bar{x}_e = \frac{\sum_{i=1}^n x_i f_i}{n} \quad \bar{x}_e = \frac{1050,4}{72} \approx 14,6 \text{ ц/га}$$

2) Мода в интервальном ряду вычисляется по формуле:

$$Mo = x_{Mo} + h \frac{f_{Mo} - f_{Mo-1}}{2f_{Mo} - f_{Mo-1} - f_{Mo+1}},$$

x_{Mo} —нижняя граница модального интервала=12,2

h — длина интервала =2;

f_{Mo} — частота модального интервала =16;

f_{Mo-1} — частота интервала, предшествующего модальному = 12;

f_{Mo+1} — частота интервала, следующего за модальным=15.

Модальный интервал находится по наибольшей частоте. Модальный интервал 3-й, т.к. у этого интервала наибольшая частота $x_{H.Mo} = 12,2$, $h=2$,

$$M_0 = 12,2 + 2 \cdot \frac{4}{32 - 12 - 15} = 13,8 \text{ ц/га}$$

1) Медиана в интервальном ряду находится по формуле: $Me = x_{Me} + h_{Me} \frac{\frac{n}{2} - S_{Me-1}}{f_{Me}}$

где $x_{H.Me}$ - начало медиального интервала;

h - длина интервала;

S_{Me-1} - накопленная частота интервала предшествующая медианному;

m_{Me} - частота медианного интервала;

$n/2$ — половина объема выборки.

Медианный интервал находится по накопленной частоте, находим интервал, в который попадает значение признака с порядковым номером

$\frac{n}{2}$, т.е. $\frac{72}{2} = 36$. Это интервал 4-ий $x_{H.Me} = 14,2$; $h = 2$; $\frac{n}{2} = \frac{72}{2} = 36$; $S_{Me} = 35$;
 $m_{Me} = 50$;

$$M_e = 14,2 + 2 \cdot \frac{36 - 35}{50} = 14,24 \text{ ц/га}$$

б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициент вариации;

1) Выборочная дисперсия, т.к. объем выборки большой $72 > 30$, дисперсию вычисляем без исправлений

$$D_B = \frac{\sum_{i=1}^n (x_i - \bar{x}_e)^2 \cdot f_i}{n} \quad D_B = \frac{805,1112}{72} \approx 11,18$$

2) Выборочное среднее квадратическое отклонение:

$$\sigma_e = \sqrt{D_e} \quad \sigma_e = \sqrt{11,18} \approx 3,34 \text{ ц/га}$$

3) Коэффициент вариации – показатель изменчивости и однородности выборки.

$$V_e = \frac{\sigma_e}{\bar{x}_e} \cdot 100\% \quad V_e = \frac{3,34}{14,59} \cdot 100\% \approx 22,9\%$$

Коэффициент вариаций больше 20%, следовательно выборка считается сильно изменчивой, неоднородной.

4) Вычислить числовые характеристики упрощенным способом, с помощью метода условного нуля

За условное начало можно взять середину любого интервала, но лучше середину модального интервала. Условные варианты находятся по формулам:

$$u_1 = \frac{x_1 - A}{h}, \text{ а далее } u_2 = u_1 + 1 \text{ и т.д. В нашем случае } A = 13,2.$$

$$u_1 = \frac{9,2-13,2}{2} = -2; \quad u_2 = -1; \quad u_3 = 0; \quad u_4 = 1; \quad u_5 = 2; \quad u_6 = 3; \quad u_7 = 4,$$

заполним 5 –ый столбец, условные варианты простые числа.

Для вычисления числовых характеристик составляем таблицу

№ п/п	$a_i - b_i$	частота f_i	середина интервала x_i	u_i	$u_i f_i$	$u_i^2 f_i$
1	8,2-10,2	9,2	7	-2	-14	28
2	10,2-12,2	11,2	12	-1	-12	12
3	12,2-14,2	13,2	16	0	0	0
4	14,2-16,2	15,2	15	1	15	15
5	16,2-18,2	17,2	9	2	18	36
6	18,2-20,2	19,2	9	3	27	81
7	20,2-22,2	21,2	4	4	16	64
Σ			72		50	236

1. Вычисляем среднеарифметические условные варианты по следующим формулам:

$$\bar{u} = \frac{\sum u_i f_i}{n} = \frac{50}{72} = 0,694; \quad \bar{u}^2 = \frac{\sum u_i^2 f_i}{n} = \frac{236}{72} = 3,278$$

2. Среднее арифметическое значение признака вычисляем по формуле:

$$\bar{x}_g = A + h \cdot \bar{u}; \quad x_g = 13,2 + 2 \cdot 0,694 = 14,588$$

3. Выборочная дисперсия: $D = S^2 = h^2 (\bar{u}^2 - (\bar{u})^2) = 4(3,278 - (0,694)^2) \approx 11,18$

6) Найти показатели точности выборочной средней;

$$\text{Абсолютная ошибка средней: } S_{\bar{x}} = \frac{S_g}{\sqrt{n}} = \frac{3,34}{\sqrt{72}} \approx 0,4; \text{ц/га}$$

Относительная ошибка выборочной средней: $S_{\bar{x}\%} = \frac{S_{\bar{x}}}{\bar{x}_e} \cdot 100\% = \frac{0,4}{14,6} \cdot 100\% \approx 3\%$

Вывод: в нашем опыте $S_{\bar{x}\%} \approx 3\% < 6\%$, следовательно, точность исследования высокая, поэтому полученные выводы отображают сущность изучаемого явления.

7) Найти доверительные интервалы для генеральной средней с надежностью 0,99; 0,95.

Доверительные интервалы для генеральной средней находим по формулам:

$$X_{ген.} \in \left(\bar{x}_e - t \cdot S_{\bar{x}}; \bar{x}_e + t \cdot S_{\bar{x}} \right) \quad \text{или} \quad X_{ген.} \in \left(\bar{X}_e - \frac{t\sigma_e}{\sqrt{n}}; \bar{X}_e + \frac{t\sigma_e}{\sqrt{n}} \right).$$

Если выборка **большая** ($n > 30$), то выборочная средняя $\bar{x}_e$ приближённо имеет нормальный закон распределения и значение t по заданной надежности γ находят с помощью таблицы значений *интегральной функции Лапласа* (таблица 3 приложения)

1. Найдем доверительный интервал с надежностью $\gamma = 0,95$, подставим ошибку средней $S_{\bar{x}} \approx 0,4$; выборочную среднюю $\bar{x}_e = 14,6$ при $\gamma = 0,95$, значение $t = 1,96$

Получим:

$$14,6 - 1,96 \cdot 0,4 \leq X_{ген.} \leq 14,6 + 1,96 \cdot 0,4 \Rightarrow 13,8ц / за \leq X_{ген.} \leq 15,4ц / за$$

2. Аналогично для $\gamma = 0,99$, найдем $t = 2,58$, получим доверительный интервал:

$$14,6 - 2,58 \cdot 0,4 \leq X_{ген.} \leq 14,6 + 2,58 \cdot 0,4 \Rightarrow 13,6ц / за \leq X_{ген.} \leq 15,6ц / за$$

Вывод: доверительный интервал генеральной средней с надежностью 95% - (13,8 ц/за; 15,4ц/за); с надежностью 99% интервал (13,6ц/за ; 15,6ц/за) покрывает генеральную среднюю.

7. Содержательная интерпретация результатов первичной обработки по условию задачи

1. Интерпретация $\bar{x}_g$.

Урожайность пшеницы – величина случайная, ее выборочные значения изменяются в частности от 8,2 ц/кг до 21,7 ц/кг, однако среднее значение равно 14,53 ц/кг

2. Интерпретация среднего квадратического отклонения σ_g .

Так как среднее квадратическое отклонение $\sigma_g = \sqrt{11,18} \approx 3,34$ ц/га то можно сделать вывод, что отклонение урожайности в среднем составляет ± 3 ц/кг.

3. Интерпретация значения коэффициента вариации V.

Значение $V=22\%$ - это больше 20%, следовательно выборка считается сильно изменчивой, неоднородной.

4. Интерпретация доверительного интервала

Так как $13,8 \text{ ц/га} \leq X_{ген.} \leq 15,4 \text{ ц/га}$ при заданной надежности $\gamma=0,95$, то можно утверждать с вероятностью 0,95, что средний урожай генеральной совокупности будет колебаться в пределах от 13,8 ц/кг до 15,4 ц/кг.

С вероятностью 99 % можно утверждать, что средняя генеральная урожайности будет находится в интервале (13,6 ц/кг; 15,6 ц/кг)

Основные формулы для вычисления числовых характеристик, показателей точности средней приведены в таблице 1.6.

Таблица 1.6.

Основные числовые характеристики.		
Наименование характеристик	для простого дискретного (несгруппированного ряда)	для дискретного (сгруппированного ряда) или интервального
$\bar{x}_g$ Выборочная средняя	$\bar{x}_g = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{x}_g = \frac{\sum_{i=1}^n x_i \cdot f_i}{n};$
$D_g = S_b^2$ Дисперсия	$D_g = \frac{\sum (x_i - \bar{x}_g)^2}{n}; (n > 30)$ $S_b^2 = \frac{\sum (x_i - \bar{x}_g)^2}{n-1}; (n < 30)$	$D_g = \frac{\sum (x_i - \bar{x}_g)^2 \cdot f_i}{n}; (n > 30)$ $S_b^2 = \frac{\sum (x_i - \bar{x}_g)^2 \cdot f_i}{n-1}; (n < 30)$
$\sigma_g = S_b$ Среднее	$\sigma_g = \sqrt{\frac{\sum (x_i - \bar{x}_g)^2}{n}}; (n > 30)$	$\sigma_g = \sqrt{\frac{\sum (x_i - \bar{x}_g)^2 \cdot f_i}{n}};$

квадратическое отклонение	$S_{\sigma} = \sqrt{\frac{\sum (x_i - \bar{x}_{\sigma})^2}{n - 1}}; (n < 30)$	$S_{\sigma} = \sqrt{\frac{\sum (x_i - \bar{x}_{\sigma})^2 \cdot f_i}{n}};$
Мода	Мода – значение признака, имеющего максимальную частоту	$Mo = x_{Mo} + h \frac{f_{Mo} - f_{Mo-1}}{2 f_{Mo} - f_{Mo-1} - f_{Mo+1}}$
Медиана		$Me = x_{Me} + h_{Me} \frac{\frac{n}{2} - S_{Me-1}}{f_{Me}}$
Показатель изменчивости, однородности выборки.		
V_{σ} Коэффициент вариации	$V_{\sigma} = \frac{S_{\sigma}}{\bar{x}_{\sigma}} \cdot 100\%;$ 1. если $V < 10\%$, то выборка считается однородной, изменчивость незначительная; 2. если $10\% < V < 20\%$, то выборка считается средне однородной и средне изменчивость; 3. если $V > 20\%$, то выборка считается неоднородной, сильно изменчивой.	
Показатели точности выборочной средней.		
$S_{\bar{x}}$ Абсолютная ошибка средней выборочной	$S_{\bar{x}} = \frac{S_{\sigma}}{\sqrt{n}};$ Ошибка выборочной средней является мерой отклонения выборочной средней $\bar{x}_{\sigma}$ от средней генеральной $\bar{x}_z$. Ошибки выражают в тех же единицах измерения, что и признак и приписывают к соответствующим средним со знаком: $\pm$, т.е. $(\bar{x}_{\sigma} \pm S_{\bar{x}})$	
$S_{\bar{x}\%}$ Относительная ошибка выборочной (показатель точности)	$S_{\bar{x}\%} = \frac{S_{\bar{x}}}{\bar{x}_{\sigma}} \cdot 100\%;$ Этот показатель указывает на степень отображения свойств генеральной совокупности. Чем меньше ошибки в исследованиях, тем выше точность исследования. 1. если $S_{\bar{x}\%} \leq 3\%$, точность опыта высокая; 2. если $S_{\bar{x}\%} = 3\% - 6\%$, точность опыта средняя; 3. если $S_{\bar{x}\%} > 6\%$, точность опыта низкая, исследования считаются ненадежными.	

РАЗДЕЛ 2. Проверка статистических гипотез

2.1. Вопросы для самопроверки.

1. Понятие и виды статистических гипотез.
2. Виды ошибок, возникающие при проверке гипотез.
2. Уровень значимости; мощность критерия.
3. Понятие статистического критерия, наблюдаемое значение критерия.
4. Допустимая и критическая области, критическая точка.
6. Общий вид алгоритма проверки статистических гипотез.
7. Методика проверки гипотезы о нормальном законе распределения.
8. Методика проверки гипотез о равенстве параметров двух выборочных совокупностей (дисперсии, средних).

2.2. Индивидуальное задание №2

Задание 1. По данным исследований статического интервального ряда (индивидуальное задание №1 (задание №2)) проверить гипотезу о нормальном распределении генеральной совокупности по критерию χ^2 .

Задание 2. По приведенным ниже данным сравнить средние урожаи некоторой культуры, получавших различные виды минеральных удобрений №1 и №2. Для проверки нулевой гипотез о равенстве средних принять уровень значимости $\alpha = 0,05$.

Вариант №1

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,6
№2	7	17,1	4,0

Вариант №2

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	17,3	4,6
№2	10	16,1	3,0

Вариант №3

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	11	12,3	3,6
№2	8	14,1	5,0

Вариант №4

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	12	11,3	4,6
№2	10	15,1	6,2

Вариант №5

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	14	12,3	5,6
№2	12	10,1	4,2

Вариант №6

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	8	18,3	2,6
№2	12	20,1	3,2

Вариант №7

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,6
№2	11	18,1	4,2

Вариант №8

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,7
№2	7	18,1	5,0

Вариант №9

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	15	16,3	3,6
№2	12	17,1	4,0

Вариант №10

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,6
№2	7	19	5,2

Вариант №11

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,6
№2	7	17,1	4,0

Вариант №12

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	17,3	4,6
№2	10	16,1	3,0

Вариант №13

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	11	12,3	3,6
№2	8	14,1	5,0

Вариант №14

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	12	11,3	4,6
№2	10	15,1	6,2

Вариант №15

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	14	12,3	5,6
№2	12	10,1	4,2

Вариант №16

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	8	18,3	2,6
№2	12	20,1	3,2

Вариант №17

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,6
№2	11	18,1	4,2

Вариант №18

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,7
№2	7	18,1	5,0

Вариант №19

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	15	16,3	3,6
№2	12	17,1	4,0

Вариант №20

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	9	16,3	3,6
№2	7	19	5,2

2.3. Решение типового примера.

Пример 2.1. По данным исследований статического интервального ряда $x_{\text{с}} \approx 15 \text{ ц/га}$, $\sigma_{\text{с}} \approx 3 \text{ ц/га}$ (индивидуальное задание №1 (задание №2)) проверить гипотезу о нормальном распределении генеральной совокупности по выборочным данным урожайности пшеницы (ц/га) по критерию χ^2

Таблица 2.1

№ п/п	$a_i - b_i$	f	x_i	S_H
1	[8,2-10,2)	7	9,2	7
2	[10,2-12,2)	12	11,2	19
3	[12,2-14,2)	16	13,2	35
4	[14,2-16,2)	15	15,2	50
5	[16,2-18,2)	9	17,2	59
6	[18,2-20,2)	9	19,2	68
7	[20,2-22,2)	4	21,2	72
Σ		72		

Требуется проверить гипотезу соответствия распределения «урожайности пшеницы» нормальному закону распределения на уровнях значимости:

$$\alpha_1 = 0,05; \alpha_2 = 0,01$$

Решение.

1. Нулевая гипотеза H_0 : распределение признака «урожайности пшеницы» соответствует нормальному закону, уровни значимости $\alpha_1 = 0,05; \alpha_2 = 0,01$

2. Выберем критерий согласия Пирсона χ^2 (хи – квадрат).

По таблице распределения χ^2 (приложение 2) находим $\chi_{\text{теорет}}^2$ теоретический критерий на уровне значимости 0,05 с числом степеней свободы: $\nu = (k-1)(l-1)$, где k – число наложенных связей для проверки нормального закона распределения ($k=2$), l – число интервалов (групп).

При нахождении l необходимо учесть следующее условие применимости критерия согласия Пирсона: малочисленные частоты ($f < 5$) следует

присоединить к частотам соседних интервалов. В этом случае соответствующие теоретические частоты также складываются, а при определении l следует принять число групп (интервалов), получившихся после объединения, таким образом объединив 6 и 7 интервалы, т.к. частота 7-ого интервала $[20,2-22,2)$ - $f=4<5$, получим $l=6$ (см.таблицу 2.2)

Таблица 2.2

№ п/п	$a_i - b_i$	f
1	$[8,2-10,2)$	7
2	$[10,2-12,2)$	12
3	$[12,2-14,2)$	16
4	$[14,2-16,2)$	15
5	$[16,2-18,2)$	9
6	$[18,2-22,2)$	13
Σ		72

Тогда число степеней свободы $\nu = (k-1)(l-1) = (2-1) \cdot (6-1) = 5$

Отсюда по таблице (Приложения 5) найдем $\chi^2_{теор}$:

-на уровне значимости $\alpha_1 = 0,05 \Rightarrow \chi^2_{теор} = \chi^2_{0,05,5} = 11,07$,

- на уровне значимости $\alpha_2 = 0,01 \Rightarrow \chi^2_{0,01,5} = 12,83$

4.Вычислить фактическое значение $\chi^2_{набл}$ по данным выборки. Для упрощенного вычисления $\chi^2_{набл}$ составим таблицу 2.3.

Таблица 2.3.

№	Интервалы		частоты	нормированное отклонение		плотность нормального распределения		вероятность	теор. частоты	частоты	наблюдаемое $\chi^2_{набл}$ значение
	a_i	b_i		$\frac{a_i - x_B}{\sigma_B}$	$\frac{b_i - x_B}{\sigma_B}$	$\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$	$\phi\left(\frac{b_i - x_B}{\sigma_B}\right)$			теор. частоты округ.	
			f					P_i	f_i^0	f_i^0	$\sum_{i=1}^k \frac{(f_i - f_i^0)^2}{f_i^0}$

1	8,2	10,2	7	-2,27	-1,6	-0,4884	-0,4452	0,04	3,11	3	5,33
2	10,2	12,2	12	-1,6	-0,93	-0,4452	-0,3238	0,13	8,74	9	1
3	12,2	14,2	16	-0,93	-0,27	-0,3238	-0,1064	0,22	15,65	16	0
4	14,2	16,2	15	-0,27	0,4	-0,1064	0,1554	0,26	18,85	19	0,84
5	16,2	18,2	9	0,4	1,07	0,1554	0,3577	0,20	14,57	15	2,4
6	18,2	22,2	13	1,07	2,4	0,3577	0,4918	0,14	10,08	10	0,1
7											
Σ			72					1		72	9,68

Складывая значения последнего столбца, получим:

$$\chi^2_{\text{набл.}} = \sum_{i=1}^k \frac{(f_i - f_i^0)^2}{f_i^0} = 9,68$$

5. Сравним $\chi^2_{\text{набл}}$ и $\chi^2_{\text{теор}}$ на обоих уровнях значимости:

при $\alpha_1 = 0,05$, получим $\chi^2_{\text{набл}} = 1,5 < \chi^2_{\text{теор}1} = 11,07$

при $\alpha_2 = 0,01$, получим $\chi^2_{\text{набл}} = 1,5 > \chi^2_{\text{теор}2} = 12,83$, следовательно, на обоих уровнях значимости основная гипотеза принимается поэтому можно считать, что выборка сделана из генеральной совокупности признака «урожайность пшеницы» подчиняется нормальному закону распределения.

Пример 2.2. По приведенным ниже данным сравнить средние урожаи некоторой культуры, получавших различные виды минеральных удобрений №1 и №2. Для проверки нулевой гипотез о равенстве средних принять уровень значимости $\alpha = 0,05$.

Вид минеральных удобрений	Количество опытных образцов культуры	Средний урожай опытных культур ц/га	Среднеквадратическое отклонение урожаев, ц/га
№1	10	16,2	3,8
№2	8	17,7	4,2

Решение.

1. Проверим гипотезу о равенстве дисперсий:

а) основная гипотеза: H_0 - дисперсий двух совокупностей равны;
альтернативная гипотеза H_1 - дисперсий не равны. Проверить гипотезу, при уровне значимости $\alpha = 0,05$;

б) Выберем критерий Фишера и вычислим критическое значение критерия $k_{крит} = F_{крит}$ из таблицы F – распределения по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы ν_1 и ν_2 (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1 = 8 - 1 = 7$, $\nu_2 = 10 - 1 = 9$, $\nu_1 \geq \nu_2$), найдем:

$$k_{крит} = F_{0,05}(7;9) = 3,29$$

в) Наблюдаемое значение критерия вычисляется по формуле: $k_{набл} = F_{набл} = \frac{S_1^2}{S_2^2}$,

в числителе подставлена большая дисперсия.

$$S_1^2 = 4,8 \geq S_2^2 = 3,8; F_{набл} = 4,2^2 / 3,8^2 = 1,22,$$

г) $F_{набл} = 1,22 \leq F_{кр} = 4,88$, гипотеза о равенстве дисперсий не отвергается (выборки из одной генеральной совокупности). то есть основания предполагать, что выборки со средним $\tilde{X}_1 = 16,2$ и $\tilde{X}_2 = 17,7$ получены из одной генеральной совокупности, а расхождение между $\tilde{X}_1$ и $\tilde{X}_2$ случайно.

2. Проверим гипотезу о равенстве средних, оценим разность между средними:

а) основная гипотеза: $H_0: |\tilde{X}_1 - \tilde{X}_2| = 0$ - средние двух совокупностей равны;

альтернативная гипотеза H_1 - средние не равны. Проверить гипотезу, при уровне значимости $\alpha = 0,05$;

б) Выберем t критерий Стьюдента и вычислим критическое значение критерия $k_{крит} = t_{крит}$ из таблицы – t - распределения по заданному уровню значимости и числу степеней свободы (приложение 3).

В пункте (1) принята гипотеза о равенстве дисперсий, следовательно число степеней свободы вычислим по формуле из таблицы 1:

$$\nu = n_1 + n_2 - 2, \quad \nu = 10 + 8 - 2 = 16 \quad t_{кр.}(0,05;16) = 2,12$$

в) Наблюдаемое значение критерия $k_{набл} = t_{набл.}$ вычислим по формуле из таблицы 2 (объемы выборок не равны $n_1 \neq n_2$):

$$k_{набл} = t_{набл.} = \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{S^2}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

Вычислим оценку дисперсий:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \Rightarrow S^2 = \frac{(10 - 1) \cdot 3,8^2 + (8 - 1) \cdot 4,2^2}{10 + 8 - 2} = 15,84.$$

Подставим полученное значение S^2 и другие величины в формулу:

$$t_{набл.} = \frac{|16,2 - 17,7|}{\sqrt{15,84}} \cdot \sqrt{\frac{10 \cdot 8}{10 + 8}} = 0,79.$$

3. Сделаем вывод:

$t_{набл.}$ меньше, чем $t_{кр.}$, поэтому нулевая гипотеза не отвергается, разность между средними не существенна, изменение фактора не влияет на средние показатели. При равенстве дисперсий изменение вида удобрений не влияет на средний урожай, можно выбрать любой вид удобрения.

Пример 2.3. Оценить влияние обеспеченности рационов протеином на изменение живой массы свиноматок в начале лактации. С этой целью сравнить средние значения потерь массы по вариантам опыта, применяя t -критерий Стьюдента. Уровень значимости принять равным 0,05.

Содержание сырого протеина в рационе, %	Количество свиноматок, гол.	Средние потери живой массы в первые 20 дней лактации, кг/гол.	Среднеквадратическое отклонение потерь живой массы, кг/гол.
11,8	5	7,5	0,8
14,3	5	1,0	0,3

Решение.

1. Проверим гипотезу о равенстве дисперсий:

а) основная гипотеза: H_0 - дисперсий двух совокупностей равны;

альтернативная гипотеза H_1 - дисперсий не равны. Проверить гипотезу, при уровне значимости $\alpha = 0,05$;

б) Выберем критерий Фишера и вычислим критическое значение критерия $k_{крит} = F_{крит}$ из таблицы F – распределения по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы ν_1 и ν_2 (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1 = 5 - 1 = 4$, $\nu_2 = 5 - 1 = 4$, $\nu_1 \geq \nu_2$), найдем:

$$k_{крит} = F_{0,05}(4;4) = 6,39$$

в) Наблюдаемое значение критерия вычисляется по формуле: $k_{набл} = F_{набл} = \frac{S_1^2}{S_2^2}$,

в числителе подставлена большая дисперсия.

$$S_1^2 = 0,8 \geq S_2^2 = 0,3; F_{набл} = 0,8^2 / 0,3^2 = 7,11$$

г) $F_{набл} = 7,11 > F_{кр} = 6,39$, гипотеза о равенстве дисперсий отвергается (выборки сделаны из разных генеральных совокупностей).

Проверим гипотезу о равенстве средних, оценим разность между средними:

а) основная гипотеза: $H_0 : |\tilde{X}_1 - \tilde{X}_2| = 0$ - средние двух совокупностей равны;

альтернативная гипотеза H_1 - средние не равны. Проверить гипотезу, при уровне значимости $\alpha = 0,05$;

б) Выберем t критерий Стьюдента и вычислим критическое значение критерия $k_{крит} = t_{крит}$ из таблицы – t - распределения по заданному уровню значимости и числу степеней свободы (приложение 3).

В пункте (1) гипотеза о равенстве дисперсий отвергнута, следовательно по таблице 1, число степеней свободы вычислим по формуле:

$$\nu = (n_1 + n_2 - 2) \cdot \left(0,5 + \frac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4} \right),$$

$$\nu = (5 + 5 - 2) \left(0,5 + \frac{0,8^2 \cdot 0,3^2}{0,8^4 + 0,3^4} \right) = 5,10.$$

$$t_{кр.}(0,05;5,1) = 2,12$$

в) Наблюдаемое значение критерия по таблице 2 вычислим по формуле (объемы выборок равны, $n_1 = n_2 = 5$):

$$t_{\text{набл.}} = \frac{|\overline{X_1} - \overline{X_2}|}{\sqrt{S^2}} \cdot \sqrt{n}, \text{ где } S^2 = S_1^2 + S_2^2 - \text{оценка дисперсий.}$$

$$t_{\text{набл.}} = \frac{|7,5 - 1,0|}{\sqrt{0,8^2 + 0,3^2}} \cdot \sqrt{5} = 17,0.$$

2. *Сделаем вывод:*

$t_{\text{набл.}} = 17 > t_{\text{кр.}} = 2,12$, поэтому нулевая гипотеза отвергается, разность между средними существенна, изменение фактора влияет на средние показатели. Следовательно, изменение обеспеченности рационов протеином существенно влияет на массу свиноматок в начале лактации, следует выбрать рацион с высоким содержанием протеина 14,3%, для которого средние потери живой массы низкие 1кг/гол при малой изменчивости $\pm 0,3$ кг/гол

РАЗДЕЛ 3. Дисперсионный анализ

3.1. Вопросы для самопроверки.

1. Сущность дисперсионного анализа.
2. Модель однофакторного анализа.
3. Проверка нулевой гипотезы при дисперсионном анализе.
4. Как определить общую, факторную и остаточные суммы?
5. Как определить обобщенную ошибку среднего, ошибку разности?
6. Для чего производится статистическая обработка результатов исследования?
7. Как сделать вывод о существенности различий между средними по вариантам?

3.2. Индивидуальное задание №3

Вариант 1-10 (задание таблицы 1,1а)

Провести дисперсионный анализ данных однофакторного вегетационного опыта с культурами по изучению влияние азотных удобрений на урожайность лука, г/сосуд, установить, значимо ли различие видов азотных удобрении на урожай лука. Нулевая гипотеза равна нулю

$H_0: d = 0$, т.е. все различия между средними по вариантам статистически не существенны.

Влияние азотных удобрений на урожай лука.

Таблица 3.1

№ п/п	Варианты, свет л/к	Повторность						
		1	2	3	4	5	6	7
1	Без азота (st)	15,8	16,1	16,0	17,2	14,9	15,5	16,1
2	Сульфат аммония	29,4	30,1	27,3	28,1	25,4	27,6	26,1
3	Аммиачная селитра	46,1	13,9	44,2	44,1	46,0	44,4	48,6
4	Мочевина	38,4	39,9	40,2	38,5	39,6	40,1	39,3

Данные урожайности по столбцам:

Таблица 3.1а.

№ варианта	Повторность					№ варианта	Повторность			
	I	II	III	IV			I	II	III	IV
1	1	2	1	1		6	1	2	3	6
2	2	3	2	2		7	2	3	5	6
3	3	4	3	3		8	1	2	3	7
4	1	2	4	1		9	2	3	5	7
5	4	5	5	4		10	1	2	4	7

Вариант 11-20 (задание таблицы 2,2а)

Провести дисперсионный анализ данных однофакторного вегетационного опыта с культурами по изучению влияния удобрений на урожайность томатов, г/сосуд, установить, значимо ли различие видов удобрений на урожай томатов. Нулевая гипотеза равна нулю $H_0: d = 0$ т.е. все различия между средними по вариантам статистически не существенны.

Влияние удобрений на урожайность томатов, г/сосуд.

Таблица 3.2.

№ п/п	Варианты опыта	Повторность						
		1	2	3	4	5	6	7
1	Без удобрений	95	101	85	100	91	87	108

2	<i>NPK (фон)</i>	185	200	190	195	180	191	201
3	<i>NPK+Cu</i>	215	236	234	225	198	215	245
4	<i>NPK+B</i>	370	354	361	370	351	385	370

Данные урожайности по столбцам:

Таблица 3.2а

№ варианта	Повторность					№ варианта	Повторность			
	I	II	III	IV			I	II	III	IV
11	1	2	3	4		16	1	2	3	6
12	2	3	4	5		17	2	3	5	6
13	3	4	5	6		18	1	2	3	7
14	1	2	3	5		19	2	3	5	7
15	4	5	6	7		20	1	2	4	7

3.3. Решение типового примера.

Пример 3.1. Необходимо обработать данные вегетационного опыта с водными культурами по изучению действия соотношения натрия, фосфора и калия при питании рассады томатов на урожай плодов. Нулевая гипотеза равна нулю $H_0: d = 0$ - все разности между средними по вариантам статистически не существенны.

Решение

1. Составляем расчетную таблицу, располагая в ней исходные данные по рядам и столбцам, определяем суммы и средние по вариантам, общую сумму и среднее значение результативного признака по опыту (табл. 3.3).

Таблица 3.3.

Ранний урожай плодов								
Соотношения N:P ₂ O ₅ :K ₂ O		Урожай, X				Число наблюдений n	Суммы V	Средние
		I	II	III	IV			
1	1:1:1 (st)	454	470	430	500	4	1854	463,5
2	1:2:1	502	550	490	507	4	2049	512,3
3	1:2:2	601	670	550	607	4	2428	607,0
4	2:1:1	407	412	475	402	4	1696	424,0
5	2:2:1	418	470	460	412	4	1760	440,0

	Общая сумма $\sum X = 9787$	$N = \sum n = 20$	$\sum X = \sum V = 9787$	$\bar{x} = \sum_{j=1}^k X / N = 489,4$
--	--------------------------------	-------------------	--------------------------	--

2. Вычисляем суммы квадратов отклонений и определяем фактическое и теоретическое значение критерия Фишера F_{ϕ} и F_{05} .

Для вычисления суммы квадратов отклонений исходные даты целесообразно преобразовать по соотношению $X_1 = X - A$, приняв за условную среднюю A число 500, близкое к среднему урожаю по опыту $\bar{x} = 489,4$ (табл. 3.6).

Таблица 3.4.

Таблица преобразованных дат						
Варианты	Повторности, $X_i = X - A$				Суммы V	V^2
	I	II	III	IV		
1 1:1:1 (st)	-46	-30	-70	0	-146	21316
2 1:2:1	2	50	-10	7	49	2401
3 1:2:2	101	170	50	107	428	183184
4 2:1:1	-93	-88	-25	-98	-304	92416
5 2:2:1	-82	-30	-40	-88	-240	57600
Общая сумма	-	-	-	-	$\sum x_1 = -213$	45369

Вычисления суммы квадратов отклонений проведем в такой последовательности:

Общее число наблюдений

$$N = \sum n = 20$$

Корректирующий фактор

$$C = \left(\sum X_1 \right)^2 : N = (213)^2 : 20 = 2268$$

Общая сумма квадратов отклонений

$$C_y = \sum X_1^2 - C = (46^2 + 30^2 + \dots + 88^2) - 2268 = 104941$$

Сумма квадратов для вариантов

$$C_v = \sum V^2 : n - C$$

$$C_v = (146^2 + 49^2 + \dots + 240^2) : 4 - 2268 = 86961$$

Остаточную сумму квадратов, соответствующую случайному варьированию вычисляем по разности:

$$C_z = C_y - C_v = 104941 - 86961 = 1798$$

Составляем таблицу дисперсионного анализа (табл.3.5).

Таблица 3.5.

Результаты дисперсионного анализа					
Дисперсия	Сумма квадратов	Степени свободы	Средний квадрат	F_{ϕ}	F_{05}
Общая C_y	104941	19	-	-	-
Вариантов C_v	86961	4	21740	18,14	3,06
Остаток (ошибки) C_z	17980	15	1199	-	-

Теоретическое значение F_{05} находят в таблице, на 5%-ном уровне значимости по 4 степеням свободы для дисперсии вариантов и 15 степеням для остатка. При $F_{\phi} > F_{05}$ в опыте есть существенные различия по вариантам и нулевая гипотеза $H_0: d = 0$ отвергается.

Следовательно, действие соотношения натрия, фосфора и калия при питании рассады томатов имеет существенное влияние на урожай плодов.

3. Определяем ошибку опыта и существенность частных различий:

а) ошибка опыта

$$S_{\bar{x}} = \sqrt{\frac{S_z^2}{n}} = \sqrt{\frac{1199}{4}} = 17,3$$

б) ошибка разности средних

$$S_d = \sqrt{\frac{2S_z^2}{n}} = \sqrt{\frac{2 * 1199}{4}} = 24,5$$

в) наименьшую существенную разность для 5%-ного уровня значимости в абсолютных и относительных показателях

$$HCP_{05} = t_{05} * S_d = 2,13 * 24,5 = 52,2;$$

$$HCP_{05} = \frac{t_{05} * S_d}{\bar{x}} * 100 = \frac{52,2}{489,4} * 100 = 10,7\%.$$

Значение критерия t_{05} возьмем из *приложения 1* для 15 степеней свободы дисперсии остатка (ошибки).

г) Вычислим разности между средней урожайностью плодов по стандарту и частными средними:

- разность между средней урожайностью плодов по стандарту (1:1:1) и средней урожайностью при питании в соотношения (1:2:1)

$$d_1 = \bar{x}_{st} - \bar{x}_2 = 463,5 - 512,3 = 48,8$$

в процентах % - $d_{1\%} = \frac{d_1}{\bar{x}_{st}} 100\% = \frac{48,8}{463,5} 100\% \approx 48,8\%$

- разность между средней урожайностью плодов по стандарту (1:1:1) и средней урожайностью при питании в соотношения (1:2:2)

$$d_2 = \bar{x}_{st} - \bar{x}_3 = 463,5 - 607 = 143,5$$

в процентах % - $d_{2\%} = \frac{d_2}{\bar{x}_{st}} 100\% = \frac{143,5}{463,5} 100\% \approx 30,9\%$

- разность между средней урожайностью плодов по стандарту (1:1:1) и средней урожайностью при питании в соотношения (2:1:1)

$$d_3 = \bar{x}_{st} - \bar{x}_4 = 463,5 - 424 = -39,5$$

в процентах % - $d_{3\%} = \frac{d_3}{\bar{x}_{st}} 100\% = \frac{-39,5}{463,5} 100\% \approx -8,5\%$

- разность между средней урожайностью плодов по стандарту (1:1:1) и средней урожайностью при питании в соотношения (2:2:1)

$$d_4 = \bar{x}_{st} - \bar{x}_5 = 463,5 - 440 = -23,5$$

в процентах % - $d_{4\%} = \frac{d_4}{\bar{x}_{st}} 100\% = \frac{-23,5}{463,5} 100\% \approx -5,1\%$

Итоги результатов опыта и статистической обработки данных записываем в таблицу 3.6.

Таблица 3.6.

Итоговая таблица дисперсионного анализа

Варианты	Урожай, г/сосуд	Разность со стандартом	
		d	%
1 1:1:1 (st)	463,5	-	-
2 1:2:1	512,3	48,8	10,5
3 1:2:2	607,0	143,5	30,9
4 2:1:1	424,0	-39,5	-8,5
5 2:2:1	440,0	-23,5	-5,1
HCP _{0,5}	-	52,2	10,7

Вывод.

Сделаем оценку наименьшей существенной разности (HCP) и разностей средних:

- если фактическая разность $d > HCP$, то она существенна, значима;

- если $d < HCP$ – несущественна, незначима, получим:

1) $d_3 > HCP$ ($143,5 > 52,5$), следовательно увеличенное питание рассады фосфором и калием обеспечивает получение более высоких урожаев (соотношение 1:2:2);

2) $d_2 < HCP$, $d_3 < HCP$, $d_4 < HCP$, следовательно остальные соотношения не существенно влияют на урожай;

3) отрицательные значения разностей $d_3 = -39,5$ и $d_4 = -23,5$ указывают - при усилении азотного питания имеет место тенденция к снижению урожая, однако статистически оно несущественно.

РАЗДЕЛ 4. Элементы корреляционного и регрессионного анализа.

4.1. Вопросы для самопроверки.

1. Понятие статистической и корреляционных зависимостей.

3. В чём состоит различие между функциональной и корреляционной зависимостью?

4. Основные задачи регрессионного и корреляционного анализа.

5. Что такое корреляционное поле, как оно строится?

- 6.Опишите, как составляется корреляционная таблица.
- 7.Что называется эмпирической линией регрессии?
- 8.В каком виде ищут уравнение линии регрессии?
9. В чём состоит сущность метода наименьших квадратов для определения параметров линии регрессии?
10. Нахождение коэффициентов регрессии, что они показывают?
- 11.Определение коэффициент корреляции, его свойства.
12. Что показывает коэффициент детерминаций?
- 13.Каким образом исследовать значимость коэффициента корреляции, с помощью какого критерия?

4.2. Индивидуальное задание №4

Задание 1. С помощью корреляционного анализа определите влияние признака *X* на признак *У*. Для этого постройте линейное уравнение регрессии, рассчитайте коэффициент корреляции и детерминаций, оцените достоверность коэффициента корреляции с помощью *t*-критерия Стьюдента, проверив гипотезу $H_0: r_{ген}=0$.

Вариант 1.

Зависимость между толщиной стебля (мм) и высотой растения (см) раннеспелого сорта сои Искра 226:

Толщина стебля (мм)	3,6	3,3	3	3,1	3,3	3,4	3,7	3,4	3,2	3,6
Высота растения (см)	54	49,5	40,7	39,5	31,6	40,4	40	43,3	35,3	45,9

Вариант 2.

Зависимость между толщиной стебля (мм) и высотой растения (см) раннеспелого сорта сои Венера высокая:

Толщина стебля (мм)	3,9	3,1	3,6	3,1	2,8	3,4	4	3	3,5	3,5
Высота растения см)	40	47,4	30,5	36	44,1	31,6	42,6	40,2	41,4	39,3

Вариант 3.

Зависимость между числом бобов на растении (шт) и урожайностью (ц/га) раннеспелого сорта сои Тимирязевская 144:

Число бобов (шт)	32	18	20	18	19	22	16	23	19	24
------------------	----	----	----	----	----	----	----	----	----	----

Урожайность (ц/га)	21	17,6	16,5	12,8	17,9	19	15,4	17,9	18,2	24
--------------------	----	------	------	------	------	----	------	------	------	----

Вариант 4.

Зависимость между числом бобов на растении (шт) и урожайностью (ц/га) среднеспелого сорта сои Днепропетровская:

Число бобов (шт)	23	20	16	18	17	29	28	20	14	18
Урожайность (ц/га)	23,2	21,2	16,2	17,6	14,8	30,8	20,1	19,8	13,1	15,6

Вариант 5.

Зависимость между вегетационным периодом (дн) и урожайностью семян (ц/га) сорта сои Приморская 529:

Вегетационный период (дн)	97	98	99	95	94	94	98	93	94	96
Урожайность семян (ц/га)	21	17,6	16,5	12,8	17,9	19	17,9	18,2	15,4	24

Вариант 6.

Зависимость между содержанием масла (%) и белка (%) в семенах сои сорта Янтарный:

Содержание масла (%)	18,7	19	18,7	18	17,6	18,6	18,9	18,5	19,8	18,3
Содержание белка (%)	39,7	40,6	39,7	40,2	42,2	39,2	40,8	40,3	39,3	40,3

Вариант 7.

Зависимость между высотой стебля (см) и полеганием гречихи (балл) скороспелой северной группы:

Высота стебля (см)	5,4	5,4	5,2	3,9	5,5	4,8	5,8	5,9	4,2	4,1
Полегание (балл)	3	3,3	3,6	3,3	3,3	3,3	3,3	3,6	3,6	3

Вариант 8.

Зависимость между числом ветвей (шт) и количеством соцветий (шт) гречихи произрастающей в Приморском крае:

Число ветвей (шт)	3,2	3,5	3,6	4,1	3,6	3,5	3,6	3,3	4,2	3,6
Количество соцветий (шт)	28,4	19,1	23,2	19,3	26,4	18,6	25,3	26,7	22,4	18,5

Вариант 9.

Зависимость между урожайностью образцов гречихи с 1м² от вегетационного периода (дн):

Вегетационный период (дн)	65	66	69	67	71	73	69	67	66	68
Урожайность	40	43,3	97,1	98,6	139,8	62	76,2	80,7	64,4	103,9

Вариант 10.

Зависимость между количеством белка (%) от вегетационного периода (дн):

Вегетационный период (дн)	65	65	71	67	67	66	66	67	68	69
Белок (%)	15,9	17,5	17,9	17,4	18	17,1	17,5	16,8	17,2	15,3

Вариант 11.

Зависимость между количеством белка (%) от вегетационного периода (дн):

Вегетационный период (дн)	65	65	71	67	67	66	66	67	68	69
Белок (%)	15,9	17,5	17,9	17,4	18	17,1	17,5	16,8	17,2	15,3

Вариант 12.

Зависимость между количеством жира (%) в зерне пшеницы от вегетационного периода (дн):

Вегетационный период (дн)	69	70	68	67	68	67	69	61	68	71
Жирность (%)	3,2	2,9	3,9	2,2	2,6	2,9	3,8	3,5	2,6	2,5

Вариант 13.

Зависимость между массой 1000 зерен (г) пшеницы и энергии прорастания (%):

Масса 1000 зерен (г)	26,15	26,1	25,34	25,23	25,04	25,06	24,65	24,57
Энергия прорастания (%)	93	90	88	86	83	86	80	79

Вариант 14.

Зависимость между массой 1000 зерен (г) и полегания (балл) образцов гречихи, произрастающей в Приморском крае:

Масса 1000 зерен (г)	19	18,1	20,4	21,3	19,4	18,5	22,1	20	23,6	24,6
Энергия прорастания (%)	3,6	2,6	3	3,3	2,6	3	3	3	3,3	3,6

Вариант 15.

Зависимость между массой 1000 зерен (г) и плёчатости (%) образцов гречихи среднеспелой Южной группы:

Масса 1000 зерен (г)	21,6	24,8	22,4	22,5	21	21,8	18,3	24,6	24,4	23,2
Пленчатости (%)	22,6	24,1	21	18,1	21	23,2	21,4	25,3	24,3	21,9

Вариант 16.

Зависимость между жаростойкостью (балл) и массой 1000 зерен (г) овса скороспелой Северной группы:

Жаростойкость (балл)	2	2,3	4	2	2,3	2	3,3	2	1,6	3
Масса 1000 зерен (г)	25,4	26,7	23,2	24,9	25,4	24,4	24,6	22,3	23,1	23

Вариант 17.

Зависимость между толщиной (мм) и высотой (см) стебля гречихи:

Толщина стебля (мм)	0,32	0,37	0,4	0,38	0,4	0,43	0,43	0,4	0,55	0,43
Высота растения (см)	3,9	5,4	5,4	5,2	3,9	5,2	4,8	5,8	5,9	4,2

Вариант 18.

Зависимость между толщиной стебля (мм) и высотой растения (см) раннеспелого сорта сои Искра 226:

Толщина стебля (мм)	3	3,4	3,5	2,9	3,2	3,6	2,9	3,6	3,6	3,4
Высота растения (см)	34,5	40,6	31,4	30,7	37,9	37,8	29,9	38,8	29,7	32,8

Вариант 19.

Зависимость между толщиной стебля (мм) и высотой растения (см) среднераннего сорта сои Венера высокая:

Толщина стебля (мм)	3,4	3,6	3,8	3,8	3,6	3,6	4,1	4,1	3,5	3,4
Высота растения (см)	43,1	43,2	44,6	34,6	53,4	50,6	43,7	39,6	44,9	48,8

Вариант 20.

Зависимость между числом бобов на растении (шт) и урожайностью (ц/га) среднераннего сорта сои Днепропетровская:

Число бобов (шт)	22	26	20	16	15	17	22	17	18	27
Урожайность (ц/га)	15,9	27,4	17,3	15,6	15,1	14,8	21,5	17	17,9	24

Задание 2. Исследовать с помощью корреляционного анализа.

Вариант №1

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

X/Y	80	83	86	90	93	n _i
24	3			1		4
25	1	5	1		3	10
26		2	4	5	7	18
27		1	10	12	15	38

n_j	4	8	15	18	25	70
-------	---	---	----	----	----	----

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 28 г.
- 4) Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 88 %.

Вариант №2

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 м² от Y вегетационного периода (дн).

x/Y	16	17	18	19	20	n_i
65	2		1	1		4
67	3	5	6			14
69	1		2	3	9	15
73		1	5	7	14	27
n_j	6	6	14	11	23	60

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить величину урожая гречихи, если вегетационный период равен 15 дней.
- 4) Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 68 ц/га.

Вариант №3

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

x/Y	34	43	45	50	53	n_i
3	11	2		1		14
4	4		4	6	12	26
5		3	2	3	10	18

n_j	15	5	6	10	22	58
-------	----	---	---	----	----	----

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить высоту растения, если толщина стебля равна 6 мм.
- 4) Найти какой толщины должен быть стебель сои, чтобы его высота составляла 47 (см).

Вариант №4

В таблице приведены данные о зависимости между X числом бобов на растении (шт) от Y урожайностью сои (ц/га).

x/Y	10	12	16	18	20	n_i
12	11	2			1	14
16	4		4	6	12	26
18		4	2	3		9
22	2		5	7	17	31
n_j	17	6	11	16	30	80

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить урожайность сои, если число бобов на растении равна 20 (шт)
- 4) Найти число бобов на растении при урожайности сои 15 ц/га.

Вариант №5

В таблице приведены данные о зависимости между X (%) содержанием масла от Y (%) белка в семенах сои.

x/Y	38	39	40	42	44	n_i
16			3	7	10	20
17		1	4	6	12	23
18	7	8		3		18
20	6	10	1		2	19
n_j	13	19	8	16	24	80

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить содержание белка в семенах сои (%), если содержание масла составляет 19 (%).
- 4) Найти содержание масла (%) если содержание белка 43 (%).

Вариант №6

В таблице приведены данные о зависимости между X (балл) жаростойкостью от Y (г) массы 1000 зерен овса.

x/Y	22	23	24	25	26	n_i
1			3	7	10	20
2		1	4	6	12	23
3	7	8		3		18
4	11	10	1		2	24
n_j	18	19	8	16	24	85

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить массу 1000 зерен овса при балле жаростойкости 5.
- 4) Найти жаростойкость (балл), если масса 1000 зерен овса составляет 21 г.

Вариант №7

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

x/Y	80	83	86	90	92	n_i
24	3			1		4
25	1	5	1		3	10
26		2	4	10	7	23
27		1	10	12	15	38
n_j	4	8	15	23	25	75

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 28 г.
- 4) Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 88 %.

Вариант №8

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 м^2 от Y вегетационного периода (дн).

x/Y	16	17	18	19	20	n_i
65	2		1	1		4
67	3	5	6			14
69	1		2	3	9	15
71		6	5	7	14	32
n_j	6	11	14	11	23	65

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить величину урожая гречихи, если вегетационный период равен 21 день.
- 4) Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 70 ц/га.

Вариант №9

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

x/Y	34	43	45	50	53	n_i
3	11	2		1		14
5	4		4	6	12	26
6		3	2	5	10	20
n_j	15	5	6	12	22	60

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить высоту растения, если толщина стебля равна 4 мм.
- 4) Найти какой толщины должен быть стебель сои, чтобы его высота составляла 52 (см).

Вариант №10

В таблице приведены данные о зависимости между X числом бобов на растении (шт) от Y урожайностью сои (ц/га).

x/Y	10	12	16	18	20	n_i
12	11	2			1	14
16	4		4	6	12	26
18		5	2	3		10
22	2		9	7	17	35
n_j	17	7	15	16	30	85

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить урожайность сои, если число бобов на растении равна 14 (шт)
- 4) Найти число бобов на растении при урожайности сои 14 ц/га.

Вариант №11

В таблице приведены данные о зависимости между X (%) содержанием масла от Y (%) белка в семенах сои.

x/Y	38	39	40	42	44	n_i
16			3	7	10	20
17		1	4	6	12	23
18	7	8		3		18
20	6	10	1		7	24
n_j	13	19	8	16	29	85

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.

3) Определить содержание белка в семенах сои (%), если содержание масла составляет 15 (%).

4) Найти содержание масла (%) если содержание белка 41 (%).

Вариант № 12

В таблице приведены данные о зависимости между X (балл) жаростойкостью от Y (г) массы 1000 зерен овса.

x/Y	21	23	24	25	26	n_i
1			3	7	10	20
2		1	4	6	12	23
3	7	8		3		18
4	11	10	1		2	24
n_j	18	19	8	16	24	85

Требуется:

1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

2) Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3) Определить массу 1000 зерен овса при балле жаростойкости 5.

4) Найти жаростойкость (балл), если масса 1000 зерен овса составляет 21 г.

Вариант №13

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

x/Y	81	83	86	90	93	n_i
24	3			1		4
25	1	5	1		3	10
26		2	4	5	7	18
28		1	10	12	15	38
n_j	4	8	15	18	25	70

Требуется:

1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

2) Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3) Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 23 г.

4) Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 87 %.

Вариант № 14

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 м^2 от Y вегетационного периода (дн).

x/Y	15	17	18	19	20	n_i
65	2		1	1		4
67	3	5	6			14
69	1		2	3	9	15
74		1	5	7	14	27
n_j	6	6	14	11	23	60

Требуется:

1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.

3) Определить величину урожая гречихи, если вегетационный период равен 16 дней.

4) Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 68 ц/га.

Вариант № 15

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

x/Y	38	43	45	50	53	n_i
2	11	2		1		14
4	4		4	6	9	23
5		3	2	3	10	18
n_j	15	5	6	10	19	55

Требуется:

1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.

- 3) Определить высоту растения, если толщина стебля равна 3 мм.
- 4) Найти какой толщины должен быть стебель сои, чтобы его высота составляла 44 (см).

Вариант №16

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

x/Y	80	83	86	90	93	n_i
24	3			1		4
25	1	5	1		3	10
26		2	4	5	7	18
27		1	10	12	15	38
n_j	4	8	15	18	25	70

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 28 г.
- 4) Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 88 %.

Вариант №17

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 м² от Y вегетационного периода (дн).

x/Y	16	17	18	19	20	n_i
65	2		1	1		4
67	3	5	6			14
69	1		2	3	9	15
73		1	5	7	14	27
n_j	6	6	14	11	23	60

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить величину урожая гречихи, если вегетационный период равен 15 дней.
- 4) Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 68 ц/га.

Вариант №18

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

x/Y	34	43	45	50	53	n_i
3	11	2		1		14
4	4		4	6	12	26
5		3	2	3	10	18
n_j	15	5	6	10	22	58

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить высоту растения, если толщина стебля равна 6 мм.
- 4) Найти какой толщины должен быть стебель сои, чтобы его высота составляла 47 (см).

Вариант №19

В таблице приведены данные о зависимости между X числом бобов на растении (шт) от Y урожайностью сои (ц/га).

x/Y	10	12	16	18	20	n_i
12	11	2			1	14
16	4		4	6	12	26
18		4	2	3		9
22	2		5	7	17	31
n_j	17	6	11	16	30	80

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.

- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить урожайность сои, если число бобов на растении равна 20 (шт)
- 4) Найти число бобов на растении при урожайности сои 15 ц/га.

Вариант №20

В таблице приведены данные о зависимости между X (%) содержанием масла от Y (%) белка в семенах сои.

x/Y	38	39	40	42	44	n_i
16			3	7	10	20
17		1	4	6	12	23
18	7	8		3		18
20	6	10	1		2	19
n_j	13	19	8	16	24	80

Требуется:

- 1) Найти коэффициент корреляции, сделать вывод о тесноте связи.
- 2) Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
- 3) Определить содержание белка в семенах сои (%), если содержание масла составляет 19 (%).
- 4) Найти содержание масла (%) если содержание белка 43 (%).

4.3. Решение типового примера.

Пример 4.1. В результате исследования зависимости площади посева X (га) на валовой сбор Y (ц) получены следующие данные:

Первичные данные валовому сбору (ц) и площадей посева (га)

№	Посевные площади, га (X)	Валовой сбор, ц (Y)
1	11,9	4,2
2	20,4	5,1
3	17,1	4,8
4	18,3	4,7
5	22,6	5,4
6	19,1	4,7
7	16,8	4,5

8	15,5	4,4
9	20,9	5,2
10	24,4	5,5
11	18,7	4,8
12	22,0	4,5
13	18,0	4,7
14	15,0	4,2
15	18,7	4,6
16	20,3	4,8
17	20,6	4,9
18	19,0	4,6
19	21,3	5,1
20	13,2	3,9
21	22,3	5,1
22	15,0	4,5
23	22,3	4,9
24	16,0	4,3
25	13,9	4,2

Полагая, что между X и Y имеет место линейная зависимость, определить:

- 1) линейное уравнение регрессии, объясните значение коэффициента регрессий b_{yx} ;
- 2) выборочный коэффициент корреляции и детерминации, сделайте выводы
- 3) проверить значимость коэффициента корреляции при уровне значимости 0,05.

Решение.

Строим точки $(x_i; y_i)$ на координатной плоскости (рисунок 4.1)



Рисунок 4.1.

1) Найдем параметры линейной регрессий.

Предположим, что точки расположены приближённо вдоль некоторой прямой (рис.4.4), уравнение которой ищем в виде:

$$y = a + bx.$$

Уравнение является уравнением линейной регрессий

Для нахождения параметров линейной регрессий воспользуемся формулами:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} ; \quad a = \bar{y} - b\bar{x}.$$

где $\bar{x} = \frac{\sum x}{n}$; $\bar{y} = \frac{\sum y}{n}$.

Обозначим: $\Delta x_i = x_i - \bar{x}$; $\Delta y_i = y_i - \bar{y}$, тогда имеем:

$$b = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2}.$$

Угловым коэффициент b в уравнении регрессии называется выборочным коэффициентом регрессии Y по X . Обозначим его b_{yx} , тогда уравнение регрессии можно записать в виде:

$$y = b_{yx}(x - \bar{x}) + \bar{y} \quad \text{или} \quad y - \bar{y} = b_{yx}(x - \bar{x}).$$

Коэффициент регрессии Y по X показывает на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу.

Для вычислений удобно составить расчетную таблицу:

N	x_i	y_i	Δx_i	Δy_i	Δx_i^2	Δy_i^2	$\Delta x_i \Delta y_i$	x_i^2	$x_i y_i$	$\bar{y}_i$
1	2	3	4	5	6	7	8	9	10	11
1	11,9	4,2	-6,6	-0,5	44,0	0,3	3,3	141,6	50,0	4,0
2	20,4	5,1	1,9	0,4	3,5	0,2	0,7	416,2	104,0	4,9
-	-	-	-	-	-	-	-	-	-	-
24	16	4,3	-2,5	-0,4	6,4	0,2	1,0	256,0	68,8	4,4
25	13,9	4,2	-4,6	-0,5	21,5	0,3	2,3	193,2	58,4	4,2
Σ	463,3	117,6	0,0	0,0	256,0	3,7	27,2	8841,9	2206,6	117,6

Найдём среднюю выборочную по X и по Y . Для этого используем суммы, записанные в последней строке первого и второго столбцов:

$$\bar{x} = \frac{\sum x}{n} = \frac{463,3}{25} = 18,5; \quad \bar{y} = \frac{\sum y}{n} = \frac{117,6}{25} = 4,7.$$

Заполняем таблицу с четвёртого по восьмой столбец. В столбцах 6,7,8 подсчитываем суммы. Суммы, полученные по столбцам 6 и 8, подставляем в формулу для нахождения b :

$$b = \frac{27,2}{256} = 0,11; \quad a = 4,7 - 0,11 \cdot 18,5 \approx 2,7$$

Тогда искомое уравнение регрессии: $y=2,7+0,11x$, где

$b_{yx} = 0,11$ – коэффициент регрессии Y по X.

Вывод: При увеличении площади посева на 1 га., валовой сбор в среднем увеличивается на 0,11 ц.

Теперь по уравнению регрессии рассчитаем теоретические значения $\hat{y}_i$ и заполним столбец 9. Например: $y_1 = 0,1 + 0,2 \cdot 11,9 = 3$.

2) Вычислим коэффициенты корреляции и детерминации

Для оценки тесноты связи между ними по результатам выборки вычисляют статистику, называемую коэффициентом корреляции.

R (или r) – коэффициент корреляции Устанавливает, есть ли связь между признаками, и насколько она тесная $-1 \leq R \leq 1$

Если же модуль коэффициента корреляции ~ 1 , то связь близка к линейной (прямой).

Коэффициент корреляции вычисляют по формулам:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n \Delta x_i \cdot \Delta y_i}{\sqrt{\sum_{i=1}^n \Delta x_i^2 \cdot \sum_{i=1}^n \Delta y_i^2}}.$$

Итак, получаем: $r = \frac{340,2}{\sqrt{1377,3 \cdot 86,4}} = 0,878$.

Вывод:

1. Коэффициент корреляций близок к единице, следовательно связь между признаками весьма тесная;
2. Коэффициент корреляции положителен ($r > 0$), следовательно связь между признаками прямая, т.е. чем больше площадь посева, тем больше валовой сбор;

Найдем коэффициент *детерминации* $d=r^2 = 0,76$ который показывает, что вариация валового сбора обусловлена на 76% влиянием площади посева.

Остальные 24% вариации валового сбора обусловлены неучтенными факторами.

1) Оценим достоверность коэффициента корреляции с помощью *t*-критерия Стьюдента.

Фактическое значение *t*-критерия равно:

$$t_{\text{факт}} = \frac{r\sqrt{n-k}}{\sqrt{1-r^2}} = \frac{0,902 \cdot \sqrt{20-2}}{\sqrt{1-0,902^2}} = 8,86.$$

Критическое (табличное) значение *t*-критерия при уровне значимости 0,05 и при $\nu = n - k = 20 - 2 = 18$ степенях свободы вариации составляет $t_{\text{кр}} = 2,1009$ (таблица «Значения *t*-критерия Стьюдента при уровне значимости 0,10, 0,05 и 0,01»).

Фактическое значение критерия выше табличного $t_{\text{факт.}} > t_{\text{кр.}}$, следовательно, вывод о достоверности связи между признаками подтверждается, связь между признаками достоверна и уравнение регрессии в полной мере отражает эту связь.

Построим найденную прямую регрессии в программе Microsoft Excel (рис.4.2).

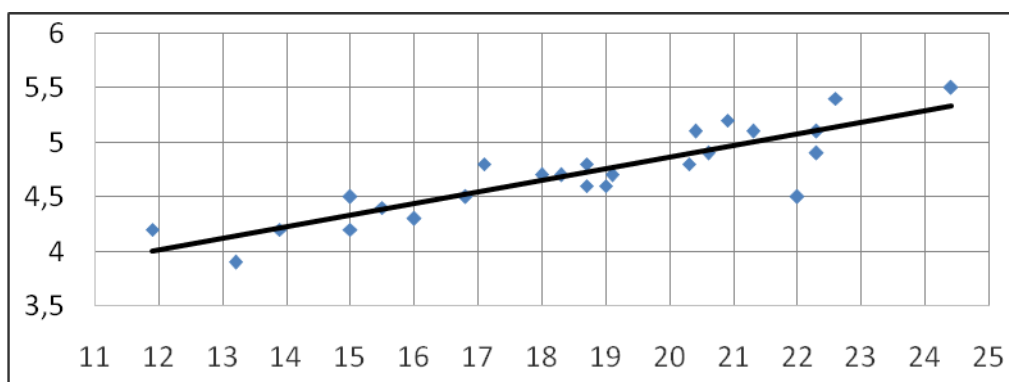


Рисунок 4.2.

Пример 4.2. В таблице приведены данные о количестве внесённых удобрений *X* (ц/га) и урожайность *Y* (ц/га) на 100га пахотной земли:

							n_i
<i>X / Y</i>	10	12	14	16	18	20	

10	9	4	1				14
30	1	10	9	3			23
50		2	6	14	6		28
70			1	10	18	6	35
n_j	10	16	17	27	24	6	100

Требуется:

1. Найти коэффициент корреляции, сделать вывод о тесноте связи.
2. Составить уравнения прямых регрессии Y на X и X на Y , построить графики.
3. Определить величину среднего урожая, если внести по 80ц удобрений на гектар.
4. Найти, сколько надо внести удобрений, чтобы получить урожай 15ц с гектара.

Решение.

1. Для нахождения коэффициента корреляции используем формулу

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x \cdot S_y} \quad \text{или} \quad r = \frac{\mu}{S_x \cdot S_y}.$$

Найдём средние выборочные значения и средние квадратические отклонения переменных X и Y :

$$\bar{x} = \frac{1}{100} (10 \cdot 14 + 30 \cdot 23 + 50 \cdot 28 + 70 \cdot 35) = 46,80;$$

$$\bar{y} = \frac{1}{100} (10 \cdot 10 + 12 \cdot 16 + 14 \cdot 17 + 16 \cdot 27 + 18 \cdot 24 + 20 \cdot 6) = 15,14;$$

$$S_x^2 = \overline{x^2} - (\bar{x})^2; \quad S_x = \sqrt{S_x^2}; \quad S_y^2 = \overline{y^2} - (\bar{y})^2; \quad S_y = \sqrt{S_y^2}.$$

$$S_x^2 = \frac{1}{100} (10^2 \cdot 14 + 30^2 \cdot 23 + 50^2 \cdot 28 + 70^2 \cdot 35) - (46,80)^2 = 445,76; \quad \text{Таким}$$

$$\text{образом: } S_x = \sqrt{445,76} \approx 21,113.$$

$$S_y^2 = \frac{1}{100} (10^2 \cdot 10 + 12^2 \cdot 16 + 14^2 \cdot 17 + 16^2 \cdot 27 + 18^2 \cdot 24 + 20^2 \cdot 6) - (15,14)^2 = 8,0204. \quad \text{Значит: } S_y = \sqrt{8,0204} \approx 2,832.$$

Чтобы найти ковариацию μ вычислим величину $\overline{xy}$.

$$\begin{aligned} \overline{xy} = \frac{1}{100} (10 \cdot 10 \cdot 9 + 10 \cdot 30 \cdot 1 + 10 \cdot 12 \cdot 4 + 30 \cdot 12 \cdot 10 + 50 \cdot 12 \cdot 2 + 14 \cdot 10 \cdot 1 + 30 \cdot 14 \cdot 9 + \\ + 14 \cdot 50 \cdot 6 + 70 \cdot 14 \cdot 1 + 16 \cdot 30 \cdot 3 + 50 \cdot 16 \cdot 14 + 70 \cdot 16 \cdot 10 + 18 \cdot 50 \cdot 6 + \\ + 18 \cdot 70 \cdot 18 + 6 \cdot 70 \cdot 20) = 759. \end{aligned}$$

$$\text{Тогда получаем, что } \mu = \overline{xy} - \bar{x} \cdot \bar{y} = 759 - 46,80 \cdot 15,14 = 50,448.$$

$$\text{Следовательно, коэффициент вариации } r = \frac{50,448}{21,113 \cdot 2,832} \approx 0,844.$$

Вывод: связь между признаками тесная, прямая.

2. Найдём коэффициенты регрессии, используя формулы:

$$b_{yx} = \frac{\mu}{S_x^2} = \frac{50,448}{445,76} \approx 0,113; \quad b_{xy} = \frac{\mu}{S_y^2} = \frac{50,448}{8,0204} \approx 6,290.$$

Составим теперь уравнения регрессий:

а) уравнение регрессии Y на X :

$$\begin{aligned} y - \bar{y} = b_{yx}(x - \bar{x}) &\Rightarrow y - 15,14 = 0,113(x - 46,8) \Rightarrow \\ &\Rightarrow y = 0,113x + 9,852. \end{aligned}$$

б) уравнение регрессии X на Y :

$$\begin{aligned} x - \bar{x} = b_{xy}(y - \bar{y}) &\Rightarrow x - 46,80 = 6,290(y - 15,14) \Rightarrow \\ &\Rightarrow x = 6,290y - 48,431. \end{aligned}$$

Графики найденных регрессий приведены на рис. 4.3.

Составим теперь уравнения регрессий:

а) уравнение регрессии Y на X :

$$y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow y - 15,14 = 0,113(x - 46,8) \Rightarrow \\ \Rightarrow y = 0,113x + 9,852.$$

б) уравнение регрессии X на Y :

$$x - \bar{x} = b_{xy}(y - \bar{y}) \Rightarrow x - 46,80 = 6,290(y - 15,14) \Rightarrow \\ \Rightarrow x = 6,290y - 48,431.$$

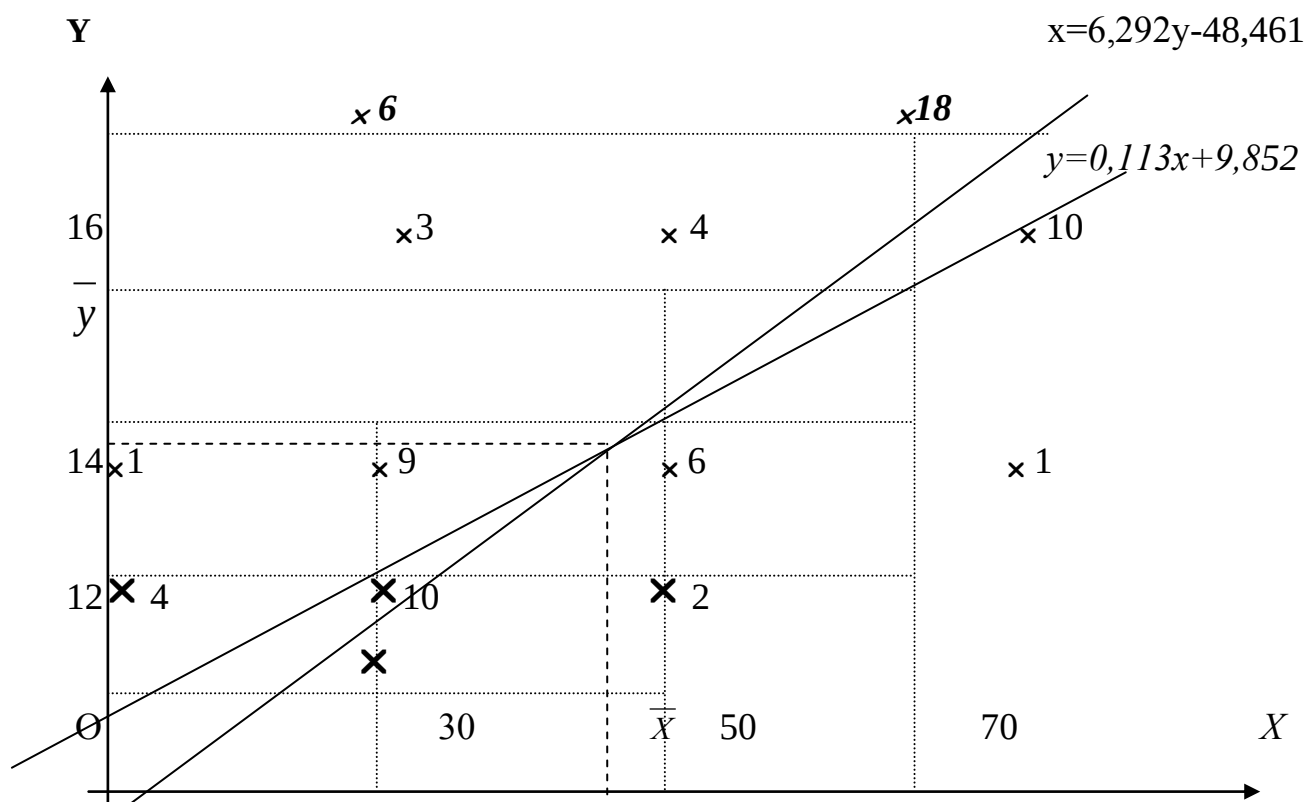


Рисунок 4.3.

3. Чтобы найти урожайность, используя уравнение регрессии Y на X , найдём значение y при $x=80$: $y = 0,113 \cdot 80 + 9,852 = 18,892 \approx 19$ (ц/га).

Значит, если будет внесено на 1 гектар 80 центнеров удобрений, то получим урожай примерно 19 центнеров с гектара.

4. Требуемое количество удобрений найдем, используя уравнение регрессии X на Y : $x = 6,290 \cdot 15 - 48,431 = 45,919 \approx 46$. Для достижения урожайности в 15 ц/га необходимо внести 46 ц/га удобрений.

Литература.

1. Основная литература.

1. Доспехов, Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований): учебник / Б.А. Доспехов. – М.: ИД Альянс, 2011. – 352 с.
2. Иода, Е.В. Статистика: учеб. пособие / Е.В. Иода.— М.: Вузовский учебник: ИНФРА-М, 2016.— 303 с.
3. Косач, О.И. Статистика [Электронный ресурс]: учеб. пособие / О.И. Косач, О.Ю. Косач; ФГБОУ ВПО «Примор. гос. с.-х. акад.». - Электрон. текст. дан. – Уссурийск, 2012. – 1 электрон. опт. диск.

2. Дополнительная литература:

1. Шириков, В.Ф. Математическая статистика: учеб. пособие / В.Ф. Шириков, С.М. Зарбалиев.— М.: КолосС, 2009.— 480 с.: ил.
2. Емельянов, Г.В. Задачник по теории вероятностей и математической статистике: учеб. пособие / Г.В. Емельянов, В.П. Скитович. – СПб.: Лань, 2007. – 336 с.

Введение.....	3
Раздел . 1. Исследование вариационных рядов	6
1.1. Вопросы для самопроверки.....	6
1.2. Индивидуальное задание №1 «Исследование вариационных рядов».....	6
1.3. Решение типового примера.....	11
Раздел . 2. Проверка статистических гипотез	24
2.1. Вопросы для самопроверки.....	24
2.2. Индивидуальное задание №2 «Проверка статистических гипотез»	25
2.3. Решение типового примера.....	30
Раздел . 3. Дисперсионный анализ.....	35
3.1. Вопросы для самопроверки.....	35
3.2. Индивидуальное задание №3 «Дисперсионный анализ»	36
3.3. Решение типового примера.....	46
Раздел . 4. Корреляционный и регрессионный анализ.....	56
4.1. Вопросы для самопроверки.....	56
4.2. Индивидуальное задание №4 «Корреляционный и регрессионный анализ»	57
4.3. Решение типового примера.....	60
Литература.....	70
Приложение.....	72

Приложение

Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}} dt$

Таблица 1

х	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2708	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3696	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3883	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4034	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4951
2,6	4953	4955	4956	4067	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

х		х		х		х	
3,0	0,49865	3,5	0,49977	4,0	0,499968	4,5	0,4999966
3,1	0,49903	3,6	0,49984	4,1	0,499979	4,6	0,4999979
3,2	0,49931	3,7	0,49989	4,2	0,499987	4,7	0,4999987
3,3	0,49952	3,8	0,49993	4,3	0,499991	4,8	0,4999992
3,4	0,49966	3,9	0,49995	4,4	0,499995	4,9	0,4999995

Значения критерия *Стьюдента* при различных уровнях значимости (α)

Таблица 2

Число степеней свободы, V	Уровни значимости (α)		
	0,05	0,01	0,001
1	12,7	63,66	-
2	4,30	9,93	31,60
3	3,18	5,84	12,94
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
ТО	1,96	2,58	3,29

Теоретические значения распределения χ^2 при разных степенях свободы (V) и критических уровнях значимости

Таблица 3

Число степеней свободы, V	Уровень значимости			Число степеней свободы, V	Уровень значимости		
	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,001$		$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,001$
1	3,84	6,63	10,8	18	28,87	34,81	42,3
2	5,99	9,21	13,8	19	30,14	36,19	43,8
3	7,81	11,34	16,3	20	31,41	37,57	45,3
4	9,49	13,28	18,5	21	32,67	38,93	
5	11,07	12,83	20,5	22	33,92	40,29	
6	12,59	16,81	22,5	23	35,17	41,64	
7	14,07	18,48	24,3	24	36,42	42,98	
8	15,51	20,09	26,1	25	37,65	44,31	
9	16,92	21,67	27,9	26	38,89	45,64	
10	18,31	23,21	29,6	27	40,11	46,96	
11	19,68	24,72	31,3	28	41,34	48,28	
12	21,03	26,22	32,9	29	42,56	49,59	
13	22,36	27,69	34,5	30	43,77	50,89	
14	23,68	29,14	36,1				
15	25,00	30,58	37,7	50	67,50	76,15	
16	26,30	32,00	39,3	80	101,88	112,3	
17	27,59	33,41	40,8	100	124,34	135,8	

Критические точки распределения Фишера-Снедекора
(k_1 - число степеней свободы большей дисперсии;
 k_2 - число степеней свободы меньшей дисперсии)

Таблица 4

Уровень значимости $\alpha = 0,01$												
k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5989	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,44	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости $\alpha = 0,05$												
k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Савельева Екатерина Владимировна

Математическая статистика: методические указания по дисциплине (модулю) к аудиторной и самостоятельной работе для обучающихся по направлениям подготовки 35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение; 35.03.07 Технология производства и переработки сельскохозяйственной продукции

Электронное издание

ФГОБУ ВО Приморская ГСХА

Адрес: 692510. г. Уссурийск, пр. Блюхера, 44.