

**Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Тульский государственный университет»**

КАФЕДРА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

**Методические указания по выполнению
типового расчета**

Тула 2015

ВВЕДЕНИЕ

В методических указаниях представлены типовые задания по математической статистике на следующие темы: первичная обработка выборок, точечное оценивание, интервальное оценивание, статистическая проверка гипотез, корреляционный анализ, дисперсионный анализ, регрессионный анализ.

Задания взяты из практикума [1], сборника типовых расчетов [2], а также разработаны автором.

Все вычисления при решении эталонного варианта произведены только с помощью калькулятора. Однако такие вычисления могут быть эффективно автоматизированы. Для обработки выборок можно использовать пакеты статистических программ

ТИПОВОЕ ЗАДАНИЕ

Задание 1. Первичная обработка выборок

- 1.1. По выборке А вычислить относительные и накопленные частоты, построить полигон и гистограмму, составить эмпирическую функцию распределения и построить ее график, вычислить числовые характеристики вариационного ряда: выборочное среднее, дисперсию, стандартное отклонение. Найти моду и медиану.
- 1.2. По выборке В: вычислить относительные и накопленные частоты, построить полигон и гистограмму, вычислить числовые характеристики: выборочное среднее, дисперсию, стандартное отклонение.
- 1.3. Для случайных величин X, Y, Z выборки С найти числовые характеристики: выборочное среднее, дисперсию, стандартное отклонение.

Задание 2. Точечное оценивание

- 2.1. Пусть случайная величина имеет распределение Пуассона. Используя метод моментов и метод максимального правдоподобия получения точечных оценок, найти по выборке А значение оценки $\hat{\lambda}$ неизвестного параметра λ .
- 2.2. Пусть случайная величина имеет нормальное распределение. Используя метод моментов и метод максимального правдоподобия получения точечных оценок, найти по выборке В значения оценок $\hat{\alpha}$ и $\hat{\sigma}$ неизвестных параметров α, σ .

Задание 3. Интервальное оценивание

Пусть генеральная совокупность имеет нормальное распределение. Из генеральной совокупности сделана выборка В. Найти доверительные интервалы для оценок математического ожидания, дисперсии σ^2 и среднеквадратичного отклонения σ генеральной совокупности при доверительной вероятности p :

$$p = \begin{cases} 0.95, & \text{если } V \leq 8 \\ 0.99, & \text{если } 9 \leq V \leq 16, \\ 0.999, & \text{если } V \geq 16 \end{cases}$$

где V – номер варианта.

Задание 4. Статистическая проверка гипотез

4.1. По выборке С случайной величины Y при уровне значимости α проверить гипотезу о среднем значении $H_0: a = a_0$ при альтернативной гипотезе $H_1: a \neq a_0$ (σ неизвестно)

$$\alpha = \begin{cases} 0.01, & \text{если } V \leq 6 \\ 0.05, & \text{если } 7 \leq V \leq 12 \\ 0.1, & \text{если } 13 \leq V \leq 18 \\ 0.2, & \text{если } V \geq 19 \end{cases}, \quad a_0 = \begin{cases} [\bar{y}] - 4, & \text{если } V \leq 6 \\ [\bar{y}] - 2, & \text{если } 7 \leq V \leq 12 \\ [\bar{y}] + 2, & \text{если } 13 \leq V \leq 18 \\ [\bar{y}] + 4, & \text{если } V \geq 19 \end{cases}$$

$$* \begin{cases} >, & \text{если } V \leq 8 \\ \neq, & \text{если } 9 \leq V \leq 16 \\ <, & \text{если } V \geq 17 \end{cases}.$$

Примечание: $[]$ – функция антье (целая часть числа, это число не превосходящее).

4.2. По выборке С случайной величины Z при уровне значимости α проверить гипотезу о дисперсии $H_0: \sigma^2 = \sigma_0^2$ при альтернативной гипотезе $H_1: \sigma^2 \neq \sigma_0^2$

$$\alpha = \begin{cases} 0.2, & \text{если } V \leq 6 \\ 0.1, & \text{если } 7 \leq V \leq 12 \\ 0.05, & \text{если } 13 \leq V \leq 18 \\ 0.01, & \text{если } V \geq 19 \end{cases}, \quad \sigma_0^2 = \begin{cases} [s^2] - 8, & \text{если } V \leq 6 \\ [s^2] - 3, & \text{если } 7 \leq V \leq 12 \\ [s^2] + 3, & \text{если } 13 \leq V \leq 18 \\ [s^2] + 8, & \text{если } V \geq 19 \end{cases}$$

$$* \begin{cases} >, & \text{если } V \leq 8 \\ \neq, & \text{если } 9 \leq V \leq 16 \\ <, & \text{если } V \geq 17 \end{cases}$$

4.3. По выборке А при уровне значимости α проверить гипотезу о распределении Пуассона генеральной совокупности

$$\alpha = \begin{cases} 0.01, & \text{если } V \leq 6 \\ 0.05, & \text{если } 7 \leq V \leq 12 \\ 0.1, & \text{если } 13 \leq V \leq 18 \\ 0.2, & \text{если } V \geq 19 \end{cases}.$$

4.4. По выборке В при уровне значимости α проверить гипотезу о нормальном распределении генеральной совокупности

$$\alpha = \begin{cases} 0.2, & \text{если } V \leq 5 \\ 0.1, & \text{если } 7 \leq V \leq 12 \\ 0.05, & \text{если } 13 \leq V \leq 18 \\ 0.01, & \text{если } V \geq 19 \end{cases}.$$

Задание 5. Корреляционный анализ

Для двум столбцам X, Y и Z выборки С определить матрицу выборочных коэффициентов парной корреляции с учётом их значимости (принять уровень значимости $\alpha = 0.05$). Если коэффициент значим, сделать вывод о направлении и тесноте связи между случайными величинами. Построить корреляционные поля и линии регрессии.

Задание 6. Дисперсионный анализ

По выборке D проверить с уровнем значимости α гипотезу о равенстве средних значений для всех уровней факторов F (о значимости влияния фактора F), где

$$\alpha = \begin{cases} 0.01, & \text{если } V \leq 12 \\ 0.05, & \text{если } V \geq 13 \end{cases}.$$

Задание 7. Регрессионный анализ

По выборке E провести регрессионный анализ: определить параметры модели

$$y = \begin{cases} \theta_1 + \theta_2 x_1 + \theta_3 x_2, & \text{если } V \leq 8 \\ \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2, & \text{если } 9 \leq V \leq 16 \\ \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2, & \text{если } V \geq 17 \end{cases}$$

по методу наименьших квадратов, проверить гипотезы об однородности дисперсии, значимости параметров модели, адекватности модели (для всех вариантов принять $\alpha = 0.05$). Найти наибольшее и наименьшее значения критерия процесса в области D, где

$$D = \begin{cases} \text{квадрат } |x_1| \leq 1, |x_2| \leq 1 \text{ при } V \leq 8 \\ \text{квадрат } |x_1| + |x_2| \leq 1 \text{ при } 9 \leq V \leq 16 \\ \text{круг } x_1^2 + x_2^2 \leq 1 \text{ при } V \geq 17 \end{cases}$$

Задание 8. Оптимальное планирование эксперимента

Синтезировать точный D-оптимальный план эксперимента для моделей, представленных в задании 7. Шаг по каждому фактору принять 0,1. Число точек плана принять на единицу больше числа параметров модели.

ВЫБОРКИ

Выборка A25 (n = 138)

X	0	1	2	3	4	5	6
n_i	2	15	41	35	23	15	7

Выборка B25 (n = 160)

X	(59;61]	(61;63]	(63;65]	(65;67]	(67;69]	(69;71]	(71;73]
n_i	4	16	34	47	37	19	3

Выборка C25 (n = 24)

X	Y	Z		X	Y	Z
73	291	577		60	234	478
69	270	548		63	243	495
72	279	575		80	313	635
72	282	573		71	278	564
65	254	519		74	292	583
67	264	530		68	271	534
56	216	443		65	256	518
70	276	555		73	291	574
63	248	502		57	219	454
64	253	506		71	281	566
70	276	554		66	262	522
67	262	535		76	302	599

Выборка D25

N	F_1	F_2	F_3	F_4	F_5	F_6
1	75	104	96	92	76	92
2	86	89	88	89	89	87
3		92	105		90	88
4		90	90		77	82
5		81	91		75	90
6						86

Выборка E25

N	x_1	x_2	y_1	y_2
1	-1	-1	59	57
2	1	-1	68	72
3	-1	1	87	81
4	1	1	123	137
5	0	0	70	74

РЕШЕНИЕ ТИПОВОГО ВАРИАНТА

Задание 1. Первичная обработка выборок

1.1. По выборке A случайной величины X найдем относительные частоты и накопленные частоты (табл. 1):

Таблица 1

Вариационный ряд выборки A

x_i	n_i	$\frac{n_i}{n}$	w_i
0	2		
1	15		
2	41		
3	35		
4	23		
5	15		
6	7		
Σ	138	1,0000	-

Полигон и гистограмма показаны на рис. 1 и 2.

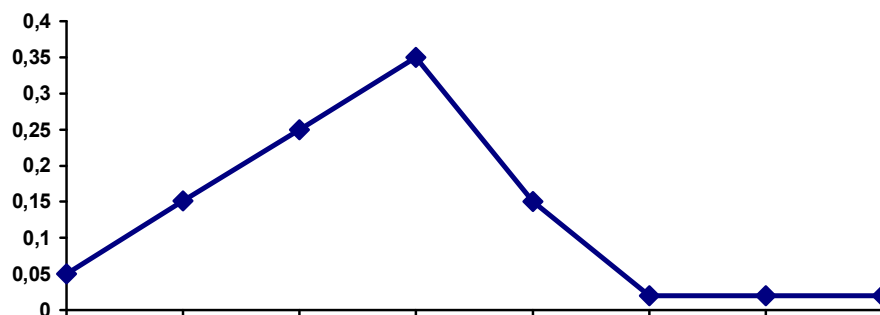


Рис. 1. Полигон выборки A

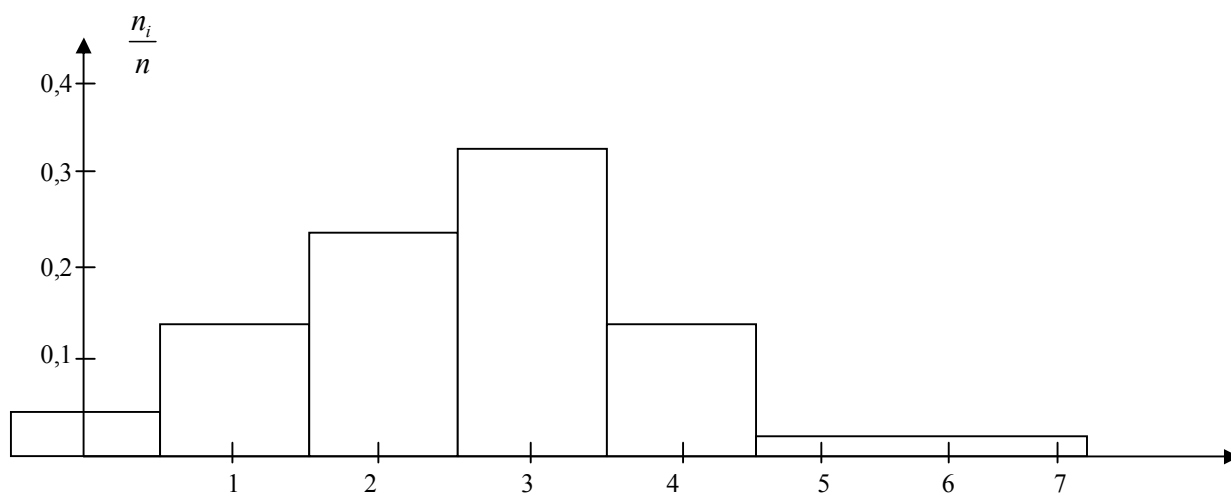


Рис. 2. Гистограмма выборки A

Эмпирическая (выборочная) функция распределения случайной величины X определяется по накопленным частотам (табл. 1):

$$F_B(x) = \begin{cases} 0 & x \leq 0 \\ 0.0417 & 0 < x \leq 1 \\ 0.1875 & 1 < x \leq 2 \\ 0.4375 & 2 < x \leq 3 \\ 0.7918 & 3 < x \leq 4 \\ 0.9376 & 4 < x \leq 5 \\ 0.9584 & 5 < x \leq 6 \\ 0.9792 & 6 < x \leq 7 \\ 1.0000 & x > 7 \end{cases}$$

График эмпирической функции распределения показан на рис. 3.

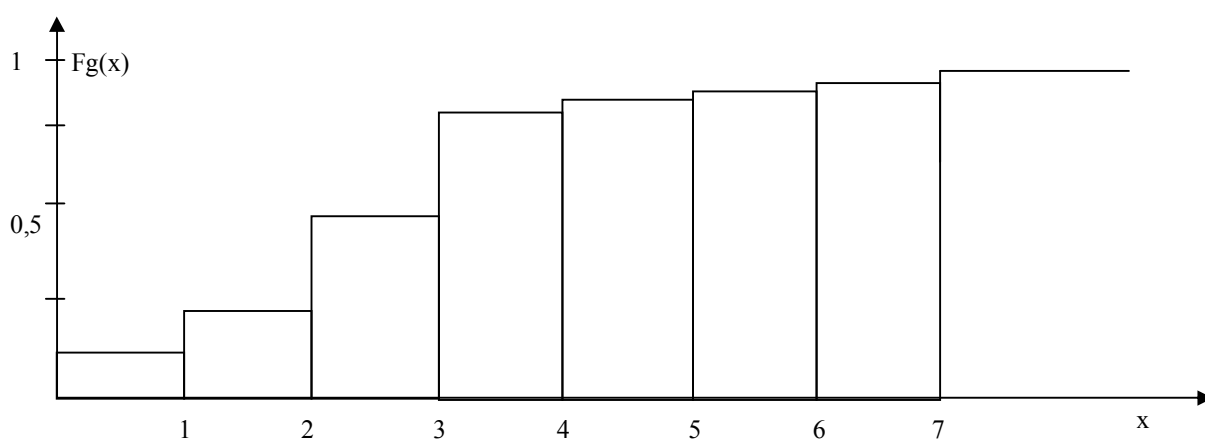


Рис. 3. Эмпирическая функция распределения выборки A

Выборочное среднее находится по формуле

$$\bar{x} = \frac{1}{n} \sum_{i=1}^w x_i n_i = \frac{1}{48} 128 = 2,67,$$

выборочная дисперсия

$$s^2 = \frac{1}{n} \sum_{i=1}^w x_i^2 n_i - \bar{x}^2 = \frac{1}{48} 430 - 2,67^2 = 1,83,$$

стандартное отклонение

$$s = \sqrt{s^2} = \sqrt{1,83} = 1,35.$$

Мода, т.е. значение, имеющее максимальную частоту, равна $M_0 = 3$ (табл. 1).

Медиана, т.е. значение срединного элемента вариационного ряда равна среднему арифметическому двух срединных элементов

$$M_e = \frac{3+4}{2} = 3,5.$$

1.2. По выборке B случайной величины X составим вариационный ряд (табл. 3) по интервалам значений, так как размах достаточно велик.

$$X_{\max} - X_{\min} + 1 = 80 - 60 + 1 = 21$$

Выберем длину интервала, равную двум.

Таблица 3

Вариационный ряд выборки B

Номер интервала	Границы интервала	Середина интервала x_i	n_i	n_i/n	w_i
1	59,5-61,5	60,5	1	0,0141	0,0141
2	61,5-63,5	62,5	1	0,0141	0,0282
3	63,5-65,5	64,5	5	0,0704	0,0986
4	65,5-67,5	66,5	6	0,0845	0,1831
5	67,5-69,5	68,5	17	0,2394	0,4225
6	69,5-71,5	70,5	14	0,1972	0,6197
7	71,5-73,5	72,5	9	0,1268	0,7465
8	73,5-75,5	74,5	10	0,1407	0,8872
9	75,5-77,5	76,5	3	0,0423	0,9295
10	77,5-79,5	78,5	3	0,0423	0,9718
11	79,5-81,5	80,5	2	0,0282	1,0000
Σ	-	-	71	1,0000	-

Полигон и гистограмма показаны на рис. 4, 5.

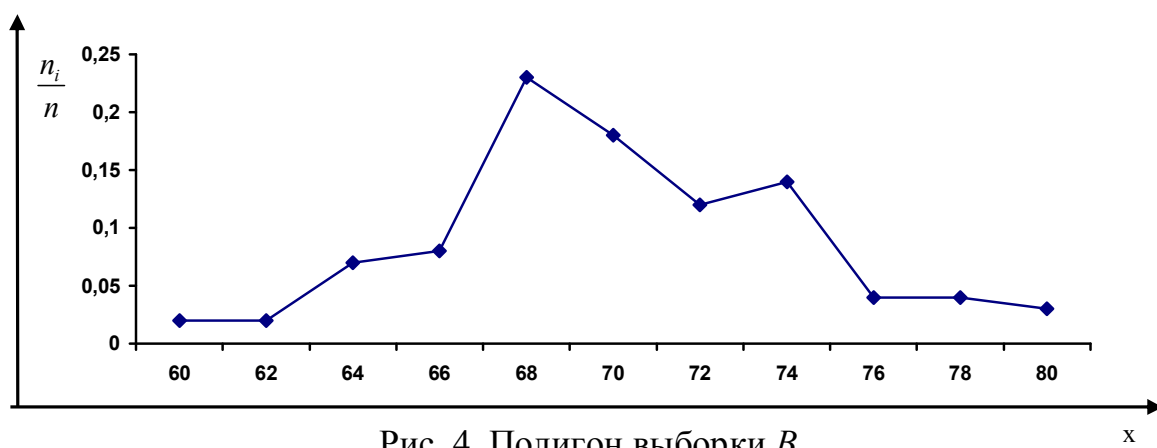


Рис. 4. Полигон выборки B

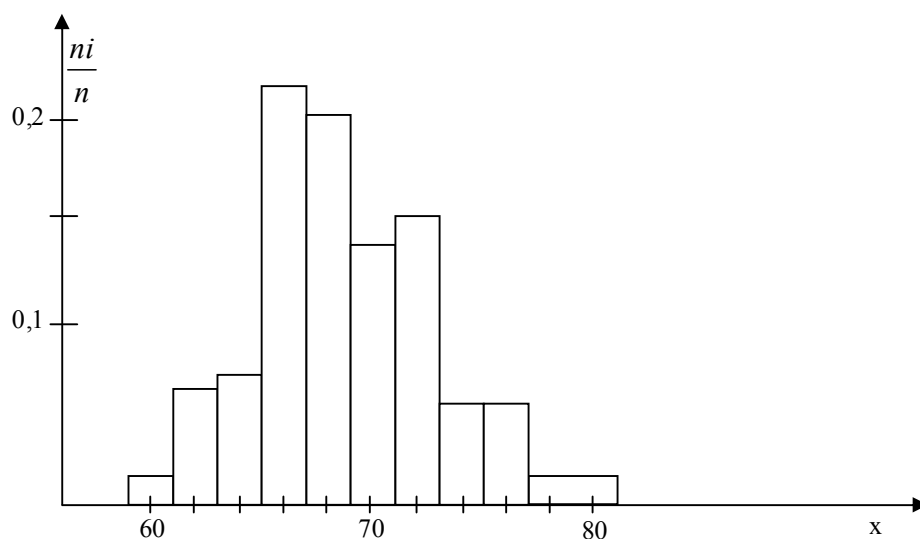


Рис. 5. Гистограмма выборки B

Эмпирическая (выборочная) функция распределения случайной величины X определяется по накопленным частотам (табл. 3):

$$F_{\xi}(x) = \begin{cases} 0 & x < 59.5 \\ 0.0141 & 59.5 \leq x < 61.5 \\ 0.0282 & 61.5 \leq x < 63.5 \\ 0.0986 & 63.5 \leq x < 65.5 \\ 0.1831 & 65.5 \leq x < 67.5 \\ 0.4225 & 67.5 \leq x < 69.5 \\ 0.6197 & 69.5 \leq x < 71.5 \\ 0.7465 & 71.5 \leq x < 73.5 \\ 0.8872 & 73.5 \leq x < 75.5 \\ 0.9295 & 75.5 \leq x < 77.5 \\ 0.9718 & 77.5 \leq x < 79.5 \\ 1.0000 & x \geq 79.5 \end{cases}$$

График эмпирической функции распределения изображен на рис. 6.

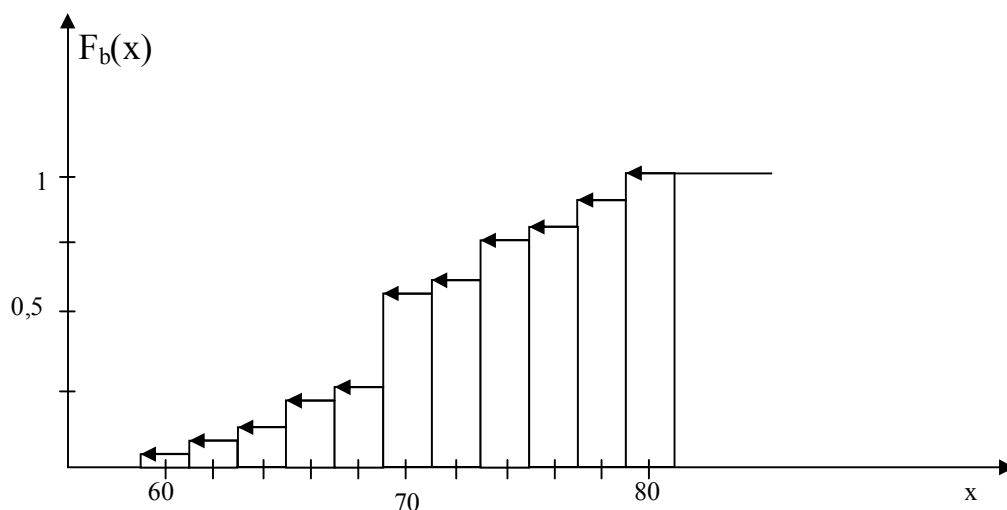


Рис. 6. Эмпирическая функция распределения выборки B

Для вычисления числовых характеристик вариационного ряда положим $c = 68,5$, соответствующей середине интервала, обладающей максимальной частотой, и шаг таблицы $k = 2$ и составим таблицу 4.

Таблица 4

Таблица для вычисления числовых характеристик вариационного ряда выборки B

Середина интервала x_i	n_i	$\frac{x_i - c}{k}$	$\frac{x_i - c}{k} n_i$	$\left(\frac{x_i - c}{k}\right)^2$	$\left(\frac{x_i - c}{k}\right)^2 n_i$
60,5	1	-4	-4	16	16
62,5	1	-3	-3	9	9
64,5	5	-2	-10	4	20
66,5	6	-1	-6	1	6
68,5	17	0	0	0	0
70,5	14	1	14	1	14
72,5	9	2	18	4	36
74,5	10	3	30	9	90
76,5	3	4	12	16	48
78,5	3	5	15	25	75
80,5	2	6	12	36	72
Σ	71	-	78	-	386

Выборочное среднее найдем по формуле

$$\bar{x} = \frac{k}{n} \sum_{i=1}^w \frac{x_i - c}{k} n_i + c = \frac{2}{71} 78 + 68,5 = 70,70,$$

выборочная дисперсия

$$s^2 = \frac{k^2}{n} \sum_{i=1}^w \left(\frac{x_i - c}{k} \right)^2 n_i - (\bar{x} - c)^2 = \frac{4}{71} 386 - (70,7 - 68,5)^2 = 16,91.$$

Стандартное отклонение

$$s = \sqrt{s^2} = \sqrt{16,91} = 4,11.$$

Мода находится по формуле

$$M_o = x_0 + k \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})},$$

где $x_0 = 67,5$ – начало модального интервала, т.е. интервала, имеющего максимальную частоту (табл. 3);

$k = 2$ – длина модального интервала;

$n_i = 17$ – частота модального интервала;

$n_{i-1} = 6$ – частота интервала, предшествующего модальному;

$n_{i+1} = 14$ – частота интервала, следующего за модальным.

Тогда

$$M_o = 67,5 + 2 \frac{17 - 6}{(17 - 6) + (17 - 14)} = 69,07.$$

Медиана вычисляется по формуле

$$M_e = x_o + k \frac{\frac{n}{2} - \sum_{j=1}^{i-1} n_j}{n_i},$$

где $x_o = 69,5$ – начало медианного интервала, т.е. интервала, в котором содержатся срединный элемент (табл. 3);

$k = 2$ – длина медианного интервала;

$n = 71$ – объем выборки;

$\sum_{j=1}^{i-1} n_j = 30$ – сумма частот интервалов, предшествующих медианному;

$n_i = 14$ – частота медианного интервала.

Тогда

$$M_e = 69,5 + 2 \frac{\frac{71}{2} - 30}{14} = 70,29.$$

1.3. Из выборки C видно, что составление вариационного ряда по значениям не имеет смысла, поскольку в выборке мало повторяющихся элементов.

Составим таблицу для вычисления необходимых сумм случайных величин X , Y , и Z выборки C (табл. 5).

Для этого определим $x_{\min} = 56$, $x_{\max} = 80$, $c_x = 68$, $y_{\min} = -313$, $y_{\max} = -216$, $c_y = -265$, $z_{\min} = 443$, $z_{\max} = 635$, $c_z = 540$, $n = 24$.

Таблица 5

Таблица для вычисления числовых характеристик вариационного ряда выборки C

x_i	$x_i - c_x$	$(x_i - c_x)^2$	y_i	$y_i - c_y$	$(y_i - c_y)^2$	z_i	$z_i - c_z$	$(z_i - c_z)^2$
73	5	25	-291	-26	676	577	37	1369
69	1	1	-270	-5	25	548	8	64
72	4	16	-279	-14	196	575	35	1225
72	4	16	-282	-17	289	573	33	1089
65	-3	9	-254	11	121	519	-21	441
67	-1	1	-264	1	1	530	-10	100
56	-12	144	-216	49	2401	443	-97	9409
70	2	4	-276	-11	121	555	15	225
63	-5	25	-248	17	289	502	-38	1444
64	-4	16	-253	12	144	506	-34	1156
70	2	4	-276	-11	121	554	14	196
67	-1	1	-262	3	9	535	-5	25
60	-8	64	-234	31	961	478	-62	3844
63	-5	25	-243	22	484	495	-45	2025
80	12	144	-313	-48	2304	635	95	9025
71	3	9	-278	-13	169	564	24	576
74	6	36	-292	-27	729	583	43	1849
68	0	0	-271	-6	36	534	-6	36
65	-3	9	-256	9	81	518	-22	484
73	5	25	-291	-26	676	574	34	1156
57	-11	121	-219	46	2116	454	-86	7396
71	3	9	-281	-16	256	566	26	676
66	-2	4	-262	3	9	522	-18	324
76	8	64	-302	-37	1369	599	59	3481
Σ	0	772	-	-53	13583	-	-21	47615

Выборочные средние

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n (x_i - c_x) + c_x = \frac{1}{24} 0 + 68 = 68,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n (y_i - c_y) + c_y = \frac{1}{24} (-53) - 256 = -267,21,$$

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n (z_i - c_z) + c_z = \frac{1}{24} (-21) + 540 = 539,13.$$

Выборочные дисперсии

$$\bar{S}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - c_x)^2 - (\bar{x} - c_x)^2 = \frac{1}{24} 772 - (68 - 68)^2 = 32,17,$$

$$\bar{S}_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - c_y)^2 - (\bar{y} - c_y)^2 = \frac{1}{24} 13583 - (-267,21 + 265)^2 = 561,08,$$

$$\bar{S}_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - c_z)^2 - (\bar{z} - c_z)^2 = \frac{1}{24} 47615 - (539,13 - 540)^2 = 1983,20.$$

Стандартные отклонения:

$$\bar{S}_x = \sqrt{\bar{S}_x^2} = \sqrt{32,17} = 5,67,$$

$$\bar{S}_y = \sqrt{\bar{S}_y^2} = \sqrt{561,08} = 23,69,$$

$$\bar{S}_z = \sqrt{\bar{S}_z^2} = \sqrt{1983,20} = 44,53.$$

Задание 2. Точечное оценивание

2.1. Метод моментов

Математическое ожидание распределения Пуассона $P(x = k) = \frac{a^k}{k!} e^{-a}$ равно $M_x = a$, где a – параметр распределения. Но выборочное среднее $\bar{x} = 2,67$ (см. задание 1). Согласно методу моментов, приравниваются теоретические моменты и их выборочные значения. В нашем случае $\hat{a} = 2,67$.

2.2. Метод максимального правдоподобия

Для получения оценки максимального правдоподобия составляется функция правдоподобия

$$L(a) = \prod_{i=1}^n \frac{a^{x_i}}{x_i!} e^{-a} = \frac{a^{\sum_{i=1}^n x_i}}{x_1! x_2! \dots x_n!} e^{-an}.$$

где n – объем выборки. Логарифмируя, получим

$$\ln L(a) = -\ln(x_1! x_2! \dots x_n!) + \sum_{i=1}^n x_i \ln a - an.$$

Найдем максимум функции $\ln L(a)$.

Используя необходимое условие экстремума, получим уравнение для нахождения оценки

$$\frac{d \ln L(a)}{d\hat{a}} = \sum_{i=1}^n x_i \frac{1}{\hat{a}} - n = 0,$$

откуда

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} = 2,67.$$

Достаточное условие экстремума выполняется

$$\frac{d^2 \ln L(a)}{da^2} = \frac{\sum_{i=1}^n x_i}{a^2} < 0.$$

Задание 3. Интервальное оценивание

Для выборки B в задании 1 были получены несмещенная оценка математического ожидания $\bar{x} = 70,70$ и смещенные оценки дисперсии $\bar{S} = 16,91$ и среднеквадратического отклонения $s = 4,11$. Найдем несмещенные оценки дисперсии и среднеквадратического отклонения

$$\begin{aligned}\bar{S}^2 &= \frac{n}{n-1} \bar{S}^2 = \frac{71}{70} 16,91 = 17,15, \\ \bar{S} &= \sqrt{\bar{S}^2} = \sqrt{17,15} = 4,14.\end{aligned}$$

Доверительный интервал для математического ожидания (среднего значения) a нормально распределенной случайной величины при неизвестной дисперсии σ^2 определяется из условия

$$P\left(\bar{x} - t_\gamma \frac{S}{\sqrt{n}} \leq a \leq \bar{x} + t_\gamma \frac{S}{\sqrt{n}}\right) = \gamma,$$

где γ - заданная доверительная вероятность;

t_γ - квантили распределения Стьюдента.

Для выборки B объем $n = 71$. Из таблицы распределения Стьюдента (3) при доверительной вероятности $\gamma = 0,98$ при числе степеней свободы $\nu = n - 1 = 71 - 1 = 70$ находим $t_\gamma = 2,38$ (интерполяция).

Точность оценки

$$\Delta = t_\gamma \frac{S}{\sqrt{n}} = 2,38 \frac{4,14}{\sqrt{71}} = 1,17.$$

Тогда доверительный интервал для математического ожидания

$$\begin{aligned}\bar{x} - \Delta &\leq a \leq \bar{x} + \Delta, \\ 70,70 - 1,17 &\leq a \leq 70,70 + 1,17, \\ 69,53 &\leq a \leq 71,87.\end{aligned}$$

Доверительный интервал для дисперсии σ^2 определяется из условия

$$P\left\{\frac{(n-1)S^2}{u^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{u_1}\right\} = \gamma.$$

Из таблицы χ^2 - распределения необходимо найти два числа u_1 и u_2 , удовлетворяющих условию

$$P(u_1 \leq \chi^2 \leq u_2) = \gamma.$$

Так как таких пар чисел существует бесконечное множество, то используем условие симметрии по вероятности

$$P(\chi^2 < u_1) = P(\chi^2 > u_2),$$

но

$$P(\chi^2 > u_2) = a_2 = \frac{1}{2}(1 - \gamma), \quad P(\chi^2 > u_1) = a_1 = 1 - a_2 = 1 - \frac{a}{2}(1 + \gamma).$$

Для выборки B $v = n - 1 = 71 - 1 = 70$. При $\gamma = 0,98$, $\alpha_1 = 0,99$, $\alpha_2 = 0,01$. По таблице χ - распределения $u_1 = 45,42$, $u_2 = 100,43$ (интерполяция). Тогда

$$\frac{(n-1)S^2}{u_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{u_1},$$

$$\frac{70 * 17,15}{100,20} \leq \sigma^2 \leq \frac{70 * 17,15}{45,54},$$

$$11,98 \leq \sigma^2 \leq 26,36.$$

Доверительный интервал среднего квадратического отклонения

$$\frac{\sqrt{n} - 1S}{\sqrt{u_2}} \leq \sigma \leq \frac{\sqrt{n} - 1S}{\sqrt{u_1}},$$

$$3,46 \leq \sigma \leq 5,13.$$

Задание 4. Статистическая проверка гипотез

4.1. При выполнении задания I получено $\bar{x} = 68$, $\bar{S}_x^2 = 32,17$, объем выборки $n = 24$.

Найдем несмещенные оценки дисперсии и среднеквадратического отклонения

$$S_x^2 = \frac{n}{n-1} \bar{S}_x^2 = \frac{24}{23} 32,17 = 33,57,$$
$$S_x = \sqrt{33,57} = 5,79.$$

По номеру варианта $V = 0$ определим $k = 0$, $\alpha = 0,01$, $a_0 = 65$ гипотеза $H_0: a = 65$, альтернативная гипотеза $H_1: a > 65$, при $\sigma = 5,79$.

Статистика $Z = \frac{\bar{x} - a_0}{\sigma} \sqrt{n}$ имеет нормальное распределение. По альтернативной гипотезе $H_1: a > 65$ найдем правостороннюю критическую область из равенства

$$P(Z < z_\alpha) = \Phi(z_\alpha) = 1 - \alpha = 1 - 0,01 = 0,99.$$

По таблице нормального распределения $z_\alpha = 2,33$.

Поэтому критическая область $Z > 2,33$. Значение статистики

$$Z = \frac{68 - 65}{5,79} \sqrt{24} = 2,54$$

Значение статистики принадлежит критической области. Следовательно, гипотезу H_0 необходимо отвергнуть и принять гипотезу $H_1: a > 65$.

4.2. При выполнении задания 1 получено $\bar{y} = -26,7,21$, $\bar{S}_y^2 = 561,08$.

Несмещенная оценка дисперсии

$$S_y^2 = \frac{n}{n-1} \bar{S}_y^2 = \frac{24}{23} 561,08 = 585,47$$

и среднеквадратического отклонения

$$S_y = \sqrt{585,47} = 24,20.$$

По номеру варианта $V = 0$ определим $k = 0$, $\alpha = 0,1$, $a_0 = -264$, гипотеза $H_0: a = -264$, альтернативная гипотеза $H_1: a < -264$.

Статистика $T = \frac{\bar{Y} - a_0}{S_y} \sqrt{n}$ имеет t -распределение Стьюдента с числом степеней свободы $\nu = n - 1 = 24 - 1 = 23$. По альтернативной гипотезе H_1 найдем левостороннюю критическую область по условию $P(T < t_\alpha) = 0,1$. Из таблицы t -распределения Стьюдента получаем $t_\alpha = -1,32$.

Отсюда критическая область $T < -1,32$.

Значение статистики

$$t = \frac{-267,21 + 264}{24,20} \sqrt{24} = -0,65.$$

Значение статистики не принадлежит критической области и нет оснований отвергать основную гипотезу $H_0: a = -264$.

Замечание. В случае $H_1: a > a_0$ используется правосторонняя критическая область $t > t_\alpha$, если же $H_1: a \neq a_0$, то критическая область примет вид $|t| > t_\alpha$.

4.3. При выполнении задания 1 получено $\bar{S}_Z^2 = 1983,20$.

Несмещенная оценка дисперсии

$$S_Z^2 = \frac{n}{n-1} \bar{S}_Z^2 = \frac{24}{23} 1983,20 = 2069,43,$$

среднеквадратического отклонения

$$S_Z = \sqrt{S_Z^2} = \sqrt{2069,49} = 45,49.$$

По номеру варианта $V = 0$ определим $k = 0$, $\alpha = 0,01$, $\sigma_o^2 = 2065$, гипотеза $H_0: \sigma^2 = 2065$ при альтернативной гипотезе $H_1: \sigma^2 > 2065$.

Статистика $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ имеет χ^2 -распределение с числом степеней свободы $\nu = n - 1 = 24 - 1 = 23$. По альтернативной гипотезе H_1 найдем правостороннюю критическую область, используя условие $P(\chi^2 > \chi_\alpha) = \alpha$ и таблицу χ^2 -распределения Пирсона

$$P(\chi^2 > \chi_\alpha) = 0.01, \quad \chi_\alpha = 41,64.$$

Отсюда критическая область $\chi^2 > 41,64$. Значение статистики

$$\chi^2 = \frac{23 \cdot 2069,43}{2065} = 23,05.$$

Это значение не принадлежит критической области, поэтому гипотеза $\sigma^2 = 2065$ не отвергается.

4.4. Произведем первичную обработку выборок, взятых из первого и второго столбцов случайной величины X выборки C . Здесь $x_{\min}^{(1)} = 56$, $x_{\max}^{(1)} = 73$, $x_{\min}^{(2)} = 57$, $x_{\max}^{(2)} = 80$, $c_1 = 65$, $c_2 = 68$, $n_1 = 12$, $n_2 = 12$.

Для вычисления числовых характеристик вариационного ряда составим таблицу 6 и таблицу 7.

Таблица 6

Таблица для вычисления числовых характеристик вариационного ряда, составленного из первого столбца случайной величины X выборки C

$x_i^{(1)}$	$x_i^{(1)} - C_1$	$(x_i^{(1)} - C_1)^2$
73	8	64
69	4	16
72	7	49
72	7	49
65	0	0
67	2	4
56	-9	81
70	5	25
63	-2	4
64	-1	1
70	5	25
67	2	4
Σ	28	322

Таблица 7

Таблица для вычисления числовых характеристик вариационного ряда, составленного из II столбца случайной величины X выборки C

$x_i^{(2)}$	$x_i^{(2)} - C_2$	$(x_i^{(2)} - C_2)^2$
60	-8	64
63	-5	25
80	12	144
71	3	9
74	6	36
68	0	0
65	-3	9
73	5	25
57	-11	121
71	3	9
66	-2	4
76	8	64
Σ	8	510

Выборочное среднее

$$\bar{x}^{(1)} = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i^{(1)} - c_1) + c_1 = \frac{1}{12} 28 + 65 = 67,33,$$

$$\bar{x}^{(2)} = \frac{1}{n_2} \sum_{i=1}^{n_2} (x_i^{(2)} - c_2) + c_2 = \frac{1}{12} 8 + 68 = 68,67.$$

Выборочная дисперсия

$$\bar{S}_{(1)}^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i^{(1)} - c_1)^2 - (\bar{x}^{(1)} - c_1)^2 = \frac{1}{12} 322 - (67,33 - 65)^2 = 21,40,$$

$$\bar{S}_{(2)}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (x_i^{(2)} - c_2)^2 - (\bar{x}^{(2)} - c_2)^2 = \frac{1}{12} 510 - (68,67 - 68)^2 = 42,05.$$

Найдем несмещенные оценки дисперсий

$$S_{(1)}^2 = \frac{n}{n-1} \bar{S}_{(1)}^2 = \frac{12}{11} 21,40 = 23,35,$$

$$S_{(2)}^2 = \frac{n}{n-1} \bar{S}_{(2)}^2 = \frac{12}{11} 42,05 = 45,87.$$

Проверим гипотезу о равенстве дисперсий $H_0: \sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе $H_1: \sigma_2^2 > \sigma_1^2$.

Уровень значимости $\alpha = 0.01$.

Статистика $F = \frac{S_{(2)}^2}{S_{(1)}^2}$ имеет F-распределение Фишера с числом степеней свободы

$\nu_1 = n_1 - 1 = 12 - 1 = 11, \nu_2 = n_2 - 1 = 12 - 1 = 11$. Найдем правостороннюю критическую область согласно условию $P(F > f_\alpha) = \alpha = 0.01$ и по таблице F-распределения: $f_\alpha = 4.49$.

Критическая область $F > 4.49$. Значение статистики $F = \frac{45,87}{23,35} = 1,96$ не принадлежит

критической области, поэтому принимаем гипотезу о равенстве дисперсий.

Проверим гипотезу о равенстве средних значений генеральных совокупностей $H_0: a_1 = a_2$ или $b = b_0 = a_1 - a_2 = 0$ при альтернативной гипотезе $H_1: b < 0$.

$$T = \frac{\bar{x}^{(1)} - \bar{x}^{(2)} - \delta_0}{\sqrt{\frac{(n_1 - 1)S_{(1)}^2 + (n_2 - 1)S_{(2)}^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Статистика имеет t -распределение Стьюдента с числом степеней свободы

$$\nu = n_1 + n_2 - 2 = 12 + 12 - 2 = 22.$$

По альтернативной гипотезе H_1 , используя условие $P(T < t_\alpha) = 0.01$ и таблицу t -распределения Стьюдента, найдем левостороннюю критическую область $T < -2.51$.

Значение статистики не принадлежит критической области, поэтому принимаем гипотезу H_0

$$t = \frac{67,33 - 68,67}{\sqrt{\frac{11 * 21,40 + 11 - 42,05 \left(\frac{1}{12} + \frac{1}{12} \right)}}} = -0,58$$

4.5. При выполнении задания 1 для выборки A получено $\bar{x} = 2,67$, $\bar{S}^2 = 1,83$. Найдем несмещенную оценку дисперсии

$$S^2 = \frac{n}{n-1} \bar{S}^2 = \frac{48}{47} 1,83 = 1,87,$$

С уровнем значимости $\alpha = 0.01$ проверим гипотезу о распределении Пуассона генеральной совокупности. Параметром распределения Пуассона

$$P(x = k) = \frac{a^k}{k!} e^{-a}$$

является a , при этом математическое ожидание и дисперсия $MX = DX = a$. Оценки среднего значения и дисперсии близки по значению, но не равны. В таблице распределения Пуассона ближайшими к ним значениями a являются 1, 2, 3.

Проверим гипотезу при $a = 2$. Составим таблицу 8.

Таблица 8

Таблица для нахождения статистики Q^2 для выборки A при $a = 2$

i	x_i	n_i	p_i	$m_i = np_i$	$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
1	0	2	0,1353	6,49	-4,49	20,16	3,11
2	1	7	0,2707	12,99	-5,99	35,88	2,76
3	2	12	0,2707	12,99	-0,99	0,98	0,08
4	3	17	0,1804	8,66	8,34	69,56	8,03
5	4	7	0,0902	4,33	2,67	7,13	1,65
6	≥ 5	3	0,0527	2,53	0,47	0,22	0,09
Σ	-	48	1,0000	48	-	-	15,72

В таблице 8 из-за того, что частоты последних значений вариационного ряда малы, они объединены в один интервал ($x_i \geq 5$).

По таблице распределения Пуассона вычислим вероятности

$$p_i = P(x = i) = \frac{a^i}{i!} e^{-a}, \quad i = 1, 2, \dots, 6$$

при $a = 2$

Статистика

$$Q^2 = \sum_{i=1}^6 \frac{(n_i - m_i)^2}{m_i} = 15,72$$

имеет χ^2 -распределение с числом степеней свободы $\nu = k - r - 1 = 6 - 1 - 1 = 4$, где k – количество интервалов, r – число параметров распределения.

Критическая область

$$P(Q^2 > \chi_\alpha^2) = \alpha = 0.01.$$

По таблице χ^2 -распределения $\chi_\alpha^2 = 13,277$. Значение статистики $Q^2 = 15,72$ принадлежит критической области $Q^2 > 13,277$, поэтому гипотезу о распределении Пуассона при $a = 2$ следует отвергнуть.

Проверим гипотезу при $a = 3$. Снова составим таблицу (табл. 9).

Таблица 9

Таблица для нахождения статистики Q^2 выборки A при $a = 3$

i	x_i	n_i	p_i	$m_i = np_i$	$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
1	0	2	0,0498	2,39	-0,39	0,15	0,06
2	1	7	0,1494	7,17	-0,17	0,03	0,00
3	2	12	0,2240	10,75	1,25	1,56	0,15
4	3	17	0,2240	10,75	6,25	39,06	3,63
5	4	7	0,1680	8,06	-1,06	1,12	0,14
6	≥ 5	3	0,1848	8,87	-5,87	34,46	3,89
Σ	-	48	1,0000	-	-	-	7,87

Так как $Q^2 = 7,87$ не принадлежит критической области $Q^2 > 13,277$, то нет оснований отвергать гипотезу о распределении Пуассона генеральной совокупности

4.6. При решении задания 1 для выборки B получены следующие числовые характеристики:

$$\bar{x} = 70,70, \quad S^2 = 17,15, \quad S = 4,14.$$

Проверим гипотезу о нормальном законе распределения генеральной совокупности при уровне значимости $\alpha = 0,1$. Примем следующие оценки параметров нормального распределения $\hat{a} = 70,70$, $\sigma = 4,14$. В вариационном ряде (табл. 3) частоты первых и последних интервалов слишком малы, поэтому объединим их с соседними интервалами так, чтобы осталось $k = 8$ интервалов. Для вычисления теоретических вероятностей составим табл. 10.

Таблица 10

Таблица для вычисления теоретических вероятностей нормально распределенной случайной величины

x_i	$\frac{x_i - a}{\sigma}$	$\Phi\left(\frac{x_i - a}{\sigma}\right)$	p_i
$-\infty$	$-\infty$	-0,5	-
62,5	-1,98	-0,47615	0,02385
64,5	-1,50	-0,43319	0,04296
66,5	-1,01	-0,34375	0,08944
68,5	-0,53	-0,20194	0,14181
70,5	-0,05	-0,01994	0,18200
72,5	0,43	0,16640	0,18634
74,5	0,92	0,32121	0,15481
∞	∞	0,5	0,17879
Σ	-	-	1,00000

Значения функции

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

взяты из таблицы 2 [2]

$$P_i = \Phi\left(\frac{x_{i+1} - a}{\sigma}\right) - \Phi\left(\frac{x_i - a}{\sigma}\right).$$

Статистика

$$Q^2 = \sum_{i=1}^k \frac{(n_i - m_i)^2}{m_i}$$

имеет χ^2 -распределение с $\nu = k - r - 1 = 8 - 2 - 1 = 5$ степенями свободы. Здесь $r = 2$, так как нормальное распределение имеет два параметра. Критическую область найдем из условия $P(Q^2 > \chi^2) = 0.1$. По таблице χ^2 -распределения $\chi_{\alpha}^2 = 9,236$ и критическая область будет $Q^2 > 9,236$.

Таблица 11

Таблица для нахождения статистики Q^2 выборки B

Интервалы	n_i	p_i	$m_i = np_i$	$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
$(-\infty; 63,5)$	2	0,02385	1,693	0,307	0,094	0,056
$(63,5; 65,5)$	5	0,04296	3,050	1,950	3,803	1,247
$(65,5; 67,5)$	6	0,08944	6,350	-0,350	0,123	0,019
$(67,5; 69,5)$	17	0,14181	10,069	6,931	48,039	4,771
$(69,5; 71,5)$	14	0,18200	12,922	1,078	1,162	0,090
$(71,5; 73,5)$	9	0,18634	13,230	-4,23	17,893	1,352
$(73,5; 75,5)$	10	0,15481	10,992	-0,992	0,984	0,090
$(75,5; +\infty)$	8	0,17879	12,694	-4,694	22,034	1,796
Σ	71	1,0000	71,000	-	-	9,361

Значение статистики $Q^2 = 9,361$ (табл. 11) принадлежит критической области $Q^2 > 9,236$, поэтому не принимаем гипотезу о нормальном распределении генеральной совокупности.

Задание 5. Корреляционный анализ

Ранее были определены случайные характеристики случайных величин X : $\bar{x} = 68$; $\overline{S_x} = 5,67$; Y : $\bar{y} = -267, 21$; $\overline{S_y} = 23,69$; Z : $\bar{Z} = 539,13$; $\overline{S_Z} = 44,53$.

Выборочная ковариация определяется по формулам

$$\text{cov}(X, Y)_B = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}, \quad \text{cov}(X, Z)_B = \frac{1}{n} \sum_{i=1}^n x_i z_i - \bar{x} \bar{z},$$

$$\text{cov}(Y, Z)_B = \frac{1}{n} \sum_{i=1}^n y_i z_i - \bar{y} \bar{z}.$$

Суммы найдем из таблицы 12.

Таблица 12

Таблица для нахождения сумм

i	x_i	y_i	z_i	$x_i y_i$	$x_i z_i$	$y_i z_i$
1	73	-291	577	-21243	42121	-167907
2	69	-270	548	-18630	37812	-147960
3	72	-279	575	-20088	41400	-160425
4	72	-282	573	-20304	41256	-161586
5	65	-254	519	-16510	33735	-131826
6	67	-264	530	-17688	35510	-139920
7	56	-216	443	-12096	24808	-95688
8	70	-276	555	-19320	38850	-153180
9	63	-248	502	-15624	31626	-124496
10	64	-253	506	-16192	32384	-128018
11	70	-276	554	-19320	38780	-152904
12	67	-262	535	-1754	35845	-140170
13	60	-234	478	-14040	28680	-111852
14	63	-243	495	-15309	31185	-120285
15	80	-313	635	-25040	50800	-198755
16	71	-278	564	-19738	40044	-156792
17	74	-292	583	-21608	43142	-170236
18	68	-271	534	-18428	36312	-144714
19	65	-256	518	-16640	33670	-132608
20	73	-291	574	-21243	41902	-167034
21	57	-219	454	-12483	25878	-99426
22	71	-281	566	-19951	40186	-159046
23	66	-262	522	-17292	34452	-136764
24	76	-302	599	-22952	45524	-180898
Σ	1632	-6413	12939	-439293	885902	-3482490

Тогда

$$\text{cov}(X, Y)_B = \frac{1}{24}(-439293) - 68(-267,21) = -133,60,$$

$$\text{cov}(X, Y)_B = \frac{1}{24}(885902) - 68 \cdot 539,12 = 251,74,$$

$$\text{cov}(Y, Z)_B = \frac{1}{24}(-3482490) - (-267,21)539,13 = -1042,82.$$

Выборочные коэффициенты парной корреляции найдем по формулам

$$(r_{XY})_B = \frac{\text{cov}(X, Y)_B}{\bar{S}_x \bar{S}_y} = -\frac{133,60}{5,67 \cdot 23,69} = -0,995,$$

$$(r_{XZ})_B = \frac{\text{cov}(X, Z)_B}{\bar{S}_x \bar{S}_z} = \frac{251,74}{5,67 \cdot 44,53} = 0,997,$$

$$(r_{YZ})_B = \frac{\text{cov}(Y, Z)_B}{\bar{S}_y \bar{S}_z} = -\frac{1042,82}{23,69 \cdot 44,53} = -0,989.$$

Для определения значимости найденных выборочных коэффициентов парной корреляции сравним их с табличным значением [4, с. 30] при числе степеней свободы $\nu = n - 2 = 24 - 2 = 22$ при уровне значимости $\alpha = 0,05 : r_{\text{табл}} = 0,408$. так как все коэффициенты парной корреляции $r_e > r_{\text{табл}}$, то все они значимы. Матрица выборочных коэффициентов парной корреляции представлена в таблице 13.

Таблица 13

Матрица выборочных коэффициентов парной корреляции

	X	Y	Z
X	1	-0,995	0,997
Y	-0,995	1	-0,989
Z	0,997	-0,989	1

Итак, между случайными величинами X и Z существует тесная положительная корреляционная связь, а между X и Y , Y и Z – тесная отрицательная корреляционная связь.

Так как корреляция между случайными величинами значима, то найдем уравнение регрессии

$$Y - \bar{Y} = k_1(X - \bar{X}),$$

где

$$k_1 = (r_{XY})_B \frac{\bar{S}_y}{\bar{S}_x} = -0,995 \frac{23,69}{5,67} = -4,16$$

и уравнение регрессии запишется в виде

$$y + 267,21 = -4,16(x - 68)$$

или

$$y = -4,16x + 15,47.$$

То же для случайных величин X и Z :

$$z - \bar{z} = k_2(x - \bar{x}),$$

где

$$k_2 = (r_{XZ})_B \frac{\bar{S}_Z}{\bar{S}_X} = 0,997 \frac{44,53}{5,67} = 7,83.$$

Тогда

$$z - 539,13 = 7,83 (x - 68)$$

и уравнение регрессии

$$z = 7,83x + 6,69.$$

Для случайных величин Y и Z

$$z - \bar{z} = K_2(y - \bar{y}),$$

где

$$K_3 = (r_{YZ})_B \frac{\bar{S}_Z}{\bar{S}_y} = -0,989 \frac{44,53}{23,69} = -1,86.$$

Тогда

$$z - 539,13 = -1,86 (y + 267,21)$$

или

$$z = -1,86y + 42,12$$

Корреляционные поля и графики уравнений регрессии показаны на рис. 7

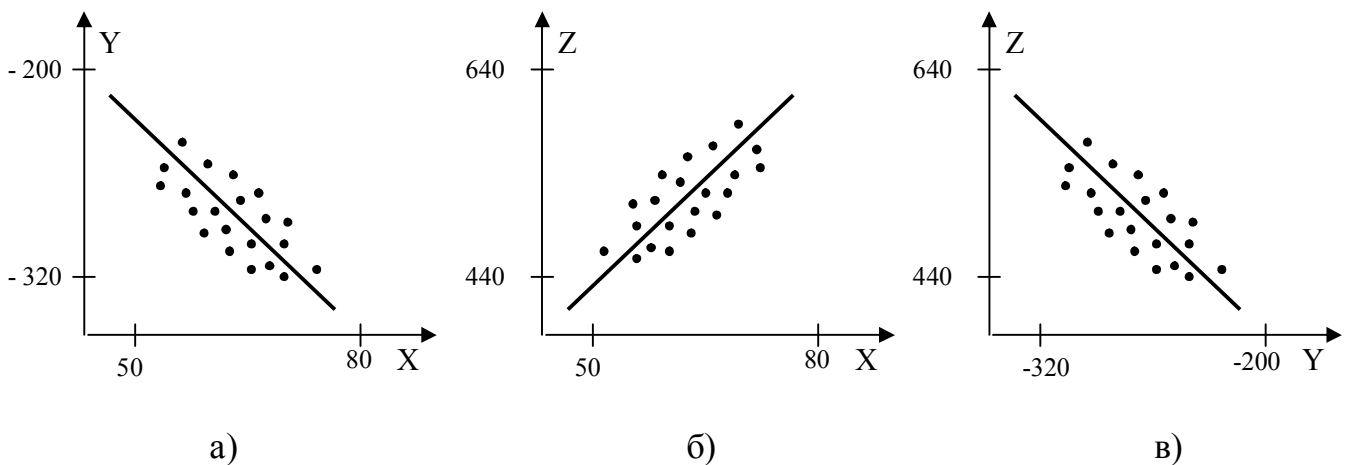


Рис. 7. Корреляционные поля и графики функций линейной регрессии случайных величин X и Y (а), X и Z (б), Y и Z (в)

Оценим ожидаемые значения случайных величин X и Z по значению случайной величины Y . Пусть, например, $x=75$. Тогда

$$y|_{x=75} = -4,16 \cdot 75 + 15,47 = -296,53,$$

$$z|_{x=75} = 7,83 \cdot 75 + 6,69 = 593,94.$$

Или

$$y|_{x=75} = -296,53 \pm 23,69,$$

$$z|_{x=75} = 593,94 \pm 44,53.$$

Задание 6. Однофакторный дисперсионный анализ

По выборке D проверить с уровнем значимости $\alpha = 0,01$ гипотезу о равенстве средних значений

$$H_0 : a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = a_7.$$

Найдем суммы по группам и общую сумму (табл. 14).

Таблица 14

Таблица для вычисления сумм выборки D

№ испытания	Уровни фактора							Σ
	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$	
1	75	104	96	92	76	92	89	
2	86	89	88	89	89	87	85	
3		92	105		90	88	93	
4		90	90		77	82		
5		81	91		75	90		
6						86		
Σ	161	456	470	181	407	525	267	2467

Здесь

$$m = 7, n_1 = 2, n_2 = 5, n_3 = 5, n_4 = 2, n_5 = 5, n_6 = 6, n_7 = 3.$$

Объем выборки:

$$N = \sum_{i=1}^7 n_i = 28.$$

Общее среднее

$$\bar{X} = \frac{2467}{28} = 88,11.$$

Сделаем замену переменных $y_{ji} = x_{ji} - c$, где c выберем близким к общему среднему, $c = [\bar{x}] = 88$.

Для определения суммы составим табл. 15.

Таблица 15

Таблица для вычисления сумм выборки D

$i \backslash j$	$F1$		$F2$		$F3$		$F4$		$F5$		$F6$		$F7$		Σ
	Y_{1i}	Y_{1i}^2	Y_{2i}	Y_{2i}^2	Y_{3i}	Y_{3i}^2	Y_{4i}	Y_{4i}^2	Y_{5i}	Y_{5i}^2	Y_{6i}	Y_{6i}^2	Y_{7i}	Y_{7i}^2	
1	-13	169	16	256	8	64	4	16	-12	144	4	16	1	1	
2	-2	4	1	1	0	0	1	1	1	1	-1	1	-3	9	
3			4	16	17	289			2	4	0	0	5	25	
4			2	4	2	4			-11	121	-6	36			
5			-7	49	3	9			-13	169	2	4			
6											-2	4			
T_i	-15		16		30		5		-33		-3		3		3
T_j^2/n_j	112,5		51,2		180		12,5		217,8		1,5		3		578,5
P_j		173		326		366		17		439		61		35	1417

Сумма квадратов отклонений элементов выборки относительно общего среднего равна

$$Q = \sum_{j=1}^m \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^m P_j - \frac{1}{N} \left(\sum_{j=1}^m T_j \right)^2 = 1417 - \frac{9}{28} = 1416,68,$$

где

$$P_j = \sum_{i=1}^{n_j} Y_{ji}^2, \quad T_j = \sum_{i=1}^{n_j} Y_{ji}.$$

Сумма квадратов отклонений групповых средних относительно общей средней (сумма квадратов межгрупповых отклонений) характеризует влияние исследуемых факторов

$$Q_1 = \sum_{j=1}^m n_j (\bar{x}_j - \bar{x})^2 = \sum_{j=1}^m \frac{1}{n_j} T_j^2 - \frac{1}{N} \left(\sum_{j=1}^m T_j \right)^2 = 578,5 - \frac{9}{28} = 578,18.$$

Сумма квадратов отклонений значений выборки относительно групповых средних (сумма квадратов внутригрупповых отклонений) характеризует влияние неучтенных факторов

$$Q_2 = \sum_{j=1}^m \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = Q - Q_1 = 1416,68 - 578,18 = 838,50.$$

Дисперсии

$$S_1^2 = \frac{Q_1}{m-1}, \quad S_2^2 = \frac{Q_2}{N-m}.$$

где S_1^2 - дисперсия, характеризующая влияние исследуемого фактора (факторная дисперсия);

S_2^2 - дисперсия, характеризующая влияния неучтенных факторов (остаточная дисперсия).

$$S_1^2 = \frac{578,18}{7-1} = 96,36 \quad , \quad S_2^2 = \frac{838,50}{28-7} = 39,93 \quad .$$

В качестве статистики используем величину

$$F = \frac{S_1^2}{S_2^2} = \frac{96,36}{39,93} = 2,41 \quad .$$

Случайная величина F имеет F -распределение со степенями свободы $\nu_1 = m - 1 = 6$, $\nu_2 = N - m = 28 - 7 = 21$.

По таблице F - распределения при уровне значимости $\alpha = 0,01$ определяем критическую область $F > 3,81$. Значение статистики $F = 2,41$ не входит в критическую область, поэтому нет основания отвергать гипотезу H_0 о равенстве средних значений по уровням. Итак, влияние фактора незначимо.

Если исходные данные интерпретировать следующим образом: m_j – число однотипных заводов разных городов $n_j = j = 1, 2, \dots$, m_i – число рабочих на j -ом заводе, изготавливающих однотипную деталь, x_{ij} , $I = 1, 2, \dots$, n_j , $j = 1, 2, \dots$, m_i – производительность труда при изготовлении однотипной детали на j -ом заводе i -ым рабочим, то задачу можно сформулировать так: установить, существенно ли различаются производительности труда рабочих определенной профессии на m заводах. В качестве фактора здесь выступает организация производства. Решение задачи позволяет сделать вывод, что производительность труда рабочих изучаемой профессии на m заводах одинакова, т.е. организация производства не оказывает существенного влияния.

Задание 7. Регрессионный анализ

Найдем средние значения критерия процесса (табл. 16)

Таблица 16

План эксперимента и средние значения критерия процесса

i	x_{1i}	x_{2i}	y_{1i}	y_{2i}	y_i
1	-1	-1	59	57	58
2	1	-1	68	72	70
3	-1	1	87	81	84
4	1	1	123	137	130
5	0	0	74	74	72

Параметры модели

$$y = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2$$

найдем по методу наименьших квадратов, который в матричном виде запишется

$$\bar{\theta} = (F^T P F)^{-1} F^T P Y,$$

где $\bar{\theta} = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ - вектор неизвестных параметров.

Распишем подробно матрицы

$$F = \begin{bmatrix} 1 & x_1 & x_2 & x_1 x_2 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \text{- матрица заданных функций;}$$

$$F^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 0 \\ -1 & -1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 \end{bmatrix} \quad \text{- транспонированная матрица заданных функций;}$$

$$P = \begin{bmatrix} 2 & & & & 0 \\ & 2 & & & \\ & & 2 & & \\ & & & 2 & \\ 0 & & & & 2 \end{bmatrix} \quad \text{- диагональная матрица весов;}$$

$$F^T P = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 0 \\ -1 & -1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 \end{vmatrix} \begin{vmatrix} 2 & & & & 0 \\ & 2 & & & \\ & & 2 & & \\ & & & 2 & \\ 0 & & & & 2 \end{vmatrix} = \begin{vmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & 2 & -2 & 2 & 0 \\ -2 & -2 & 2 & 2 & 0 \\ 2 & -2 & -2 & 2 & 0 \end{vmatrix}$$

$$F^T P F = \begin{vmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & 2 & -2 & 2 & 0 \\ -2 & -2 & 2 & 2 & 0 \\ 2 & -2 & -2 & 2 & 0 \end{vmatrix} \begin{vmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & 2 & -2 & 2 & 0 \\ -2 & -2 & 2 & 2 & 0 \\ 2 & -2 & -2 & 2 & 0 \end{vmatrix} \text{ - информационная матрица Фишера}$$

$$(F^T P F)^{-1} = \begin{vmatrix} \frac{1}{10} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & \frac{1}{8} \end{vmatrix} \text{ - дисперсионная матрица;}$$

$$(F^T P F)^{-1} F^T P = \begin{vmatrix} \frac{1}{10} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & \frac{1}{8} \end{vmatrix} \begin{vmatrix} 2 & 2 & 2 & 2 & 2 \\ -2 & 2 & -2 & 2 & 0 \\ -2 & -2 & 2 & 2 & 0 \\ 2 & -2 & -2 & 2 & 0 \end{vmatrix} = \begin{vmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & 0 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & 0 \end{vmatrix}$$

$$Y = \begin{vmatrix} 58 \\ 70 \\ 84 \\ 130 \\ 72 \end{vmatrix} \text{ - вектор критерия процесса.}$$

Тогда

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{pmatrix} = \bar{\theta} = (F^T P F)^{-1} F^T P Y = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & 0 \\ \frac{4}{1} & \frac{4}{1} & -\frac{4}{1} & \frac{4}{1} & 0 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix} \begin{pmatrix} 58 \\ 70 \\ 84 \\ 130 \\ 72 \end{pmatrix} = \begin{pmatrix} \frac{58+70+84+130+72}{5} \\ \frac{-58+70-84+130}{4} \\ \frac{-58-70+84+130}{4} \\ \frac{58-70-84+130}{4} \end{pmatrix} = \begin{pmatrix} 82.8 \\ 14.5 \\ 21.5 \\ 8.5 \end{pmatrix}$$

Итак, модель имеет вид:

$$y = 82,8 + 14,5x_1 + 21,5x_2 + 8,5x_1x_2.$$

Для проверки гипотез (однородности дисперсий, значимости параметров и адекватности модели) найдем дисперсии в точках плана (табл. 17) по формуле

$$S_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (y_{ki} - y_i),$$

где n_i – число опытов в точке плана (в нашем случае $n_i = 2 \quad \forall i$).

Таблица 17

Дисперсии в точках плана

i	y_{1i}	y_{2i}	y_i	$y_{1i} - y_i$	$y_{2i} - y_i$	$\sum_{n=1}^2 (y_{ki} - y_i)^2$	S_i^2
1	59	57	58	1	-1	2	2
2	68	72	70	-2	2	8	8
3	87	81	84	3	-3	18	18
4	123	137	130	-7	7	98	98
5	70	74	72	2	2	8	8
Σ							134

Дисперсия всего эксперимента

$$S_y^2 = \frac{\sum_{i=1}^w f_i S_i^2}{\sum_{i=1}^w f_i},$$

где $f_i = n_i - 1$ – число степеней свободы в i -ой точке плана;

$N = 5$ – число точек плана.

В нашем случае $f_i = 1 \quad \forall i$

$$S_y^2 = \frac{2+8+18+98+8}{5} = 26,80.$$

Проверим гипотезу об однородности дисперсий в точках плана по критерию Кохрена (в случае неодинакового числа опытов в точках плана применяется критерий Бартлетта)

$$G_{\text{расч}} = \frac{\max S_i^2}{\sum_{i=1}^w S_i^2} = \frac{98}{134} = 0,73 .$$

Табличное значение $G_{\text{табл}} = 0,928$ [3].

Так как $G_{\text{расч}} < G_{\text{табл}}$, то гипотеза об однородности дисперсий не отвергается.

Значимость параметров модели определяется по критерию Стьюдента. Для этого найдем дисперсионную матрицу

$$D = (F^T P F)^{-1} S_y^2 = \begin{vmatrix} \frac{1}{10} & & & \\ & \frac{1}{8} & & \\ & & \frac{1}{8} & \\ 0 & & & \frac{1}{8} \end{vmatrix} \cdot 26,80 = \begin{vmatrix} 2,68 & & & 0 \\ & 3,35 & & \\ & & 3,35 & \\ 0 & & & 3,35 \end{vmatrix}$$

Расчетные значения критерия Стьюдента

$$t_{j\text{расч}} = \frac{|\theta_j|}{\sqrt{d_{jj}}}, \quad j = 1, 2, 3, 4,$$

где d_{ji} – j -ый диагональный элемент дисперсионной матрицы.

Тогда

$$t_{1\text{расч}} = \frac{82,8}{\sqrt{2,68}} = 50,58,$$

$$t_{2\text{расч}} = \frac{14,5}{\sqrt{3,35}} = 7,92,$$

$$t_{3\text{расч}} = \frac{21,5}{\sqrt{3,35}} = 11,75,$$

$$t_{4\text{расч}} = \frac{8,5}{\sqrt{3,35}} = 4,64.$$

Табличное значение критерия Стьюдента $t_{\text{табл}} = 2,45$ при числе степеней свободы $\nu = \sum_{i=1}^w n_i - w = 10 - 4 = 6$ (3). Так как все $t_{j\text{расч}} > t_{\text{табл}}$, то гипотеза о значимости параметров модели не отвергается.

Проверку гипотезы об адекватности регрессионной модели осуществим по критерию Фишера

$$F = \frac{S_{ag}^2}{S_y^2},$$

где

$$S_{ag}^2 = \frac{\sum_{i=1}^w n_i (y_{i\text{мод}} - y_i)^2}{W} - \text{дисперсия адекватности.}$$

Для определения дисперсии адекватности составим табл. 18.

Таблица 18

Таблица для нахождения дисперсии адекватности

i	x_1	x_2	y_i	$y_{i\text{мод}}$	$y_{i\text{мод}} - y_i$	$(y_{i\text{мод}} - y_i)^2$
1	-1	-1	58	55,3	-2,7	7,29
2	1	-1	70	67,3	-2,7	7,29
3	-1	1	84	81,3	-2,7	7,29
4	1	1	130	127,3	-2,7	7,29
5	0	0	72	82,8	10,8	110,64
Σ						145,80

Тогда

$$S_{ag}^2 = \frac{2 \cdot 145,80}{5} = 58,32.$$

Расчетное значение критерия Фишера равно

$$F_{\text{расч}} = \frac{58,32}{26,80} = 2,18.$$

Табличное значение критерия Фишера берем из таблицы F -распределения [3] при степенях свободы $\nu = \sum_{i=1}^w f_i = 5$; $\nu_2 = w - m = 5 - 4 = 1$, где m – число значимых параметров модели: при уровне значимости $\alpha = 0,05$. Оно равно $F_{\text{табл}} = 230,16$. Так как $F_{\text{расч}} < F_{\text{табл}}$, то гипотеза об адекватности модели не отвергается.

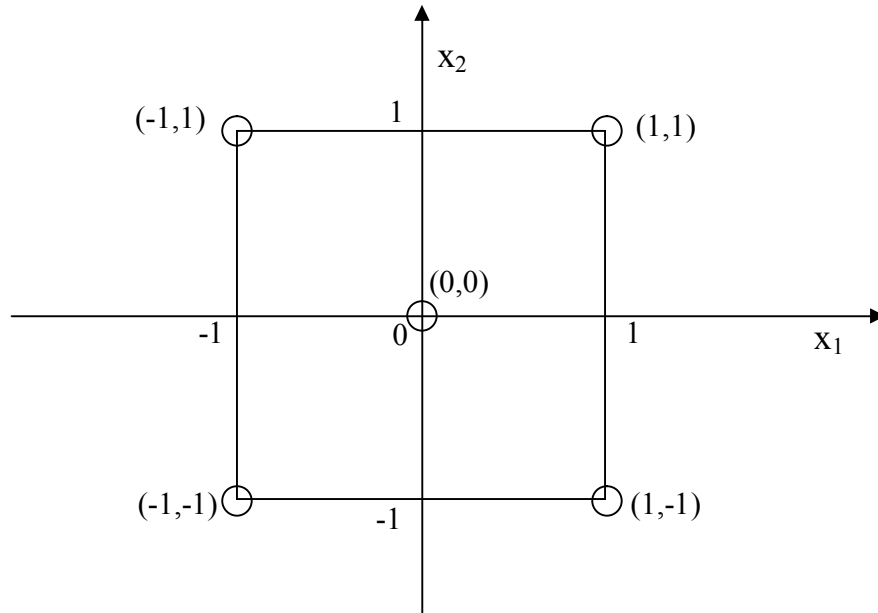
Найдем наибольшее значение критерия у в квадрате G : $|x_1| \leq 1$, $|x_2| \leq 1$.

Внутри квадрата

$$\begin{cases} \frac{dy}{dx_1} = 14,5 + 8,5x_2 = 0 \\ \frac{dy}{dx_2} = 21,5 + 8,5x_1 = 0 \end{cases}$$

Решая систему, получим $x_1 = -2,53$, $x_2 = -1,74$.

Точка $(-2,53; -1,71)$ не принадлежит G . На границе $x_1 = 1$ $y = 97,3 + 30,0x_2$ экстремума нет. Аналогично для остальных границ. В вершинах квадрата



(табл. 18) критерий процесса имеет наибольшее значение в точке $(1,1)$ и равен $127,3 \pm 5,18$.

Библиографический список

1. Колде Я.И. Практикум по теории вероятностей и математической статистики. – М.: Высшая математика, 1991 – 157 с.
2. Чудесенко В.Ф. Сборник заданий по специальным курсам высшей математики. – М.: Высшая школа, 1983 – 112с.
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: ВЦ АН СССР, 1968 – 475 С.
4. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий – М.: Наука, 1976 – 279 с.

Рассмотрено на заседании методической комиссии
кафедры математического моделирования
Протокол № _____ от ” ____ ” _____ 20 ____ г.
Председатель _____ Аверин В.В.

Нормоконтролер,
ответственный за стандартизацию на кафедре
_____ Пустовгар А.С.

Утверждено на заседании
кафедры математического моделирования.
Протокол № _____ от “ ____ ” _____ 20 ____ г.
Заведующий кафедрой _____ Маркин А.А.