Контрольная работа по математической статистике
Вариант: $m=2, n=6$

1 Численная обработка данных одномерной выборки.

Выборка X объемом $N=100$ измерений задана таблицей:

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7
m_{x_i}	5	13	$20+(m+n)$	$30-(m+n)$	19	10	3

где x_i - результаты измерений, m_{x_i} - частоты, с которыми встречаются значения x_i,
$\sum_{i=1}^{7} m_{x_i}=100$, $x_i=0,2\cdot m+(i-1)\cdot 0,3\cdot n$.

13.1.1 Построить полигон относительных частот $W_i=\frac{m_{x_i}}{N}$.

13.1.2 Вычислить среднее выборочное $\overline{X}$, выборочную дисперсию D_x и среднее квадратическое отклонение σ_x.

13.1.3 По критерию χ^2 проверить гипотезу о нормальном распределении генеральной совокупности при уровне значимости $\alpha=0,05$.

<u>Примечание.</u> Для расчетов $\overline{X}$ и D_x рекомендуется перейти к условным значениям $u_i=\frac{x_i-c_x}{0,3\cdot n}$ и, взяв за ложный нуль c_x значение с наибольшей частотой, использовать суммы $\sum_{i=1}^{7} m_{x_i}\cdot u_i$ и $\sum_{i=1}^{7} m_{x_i}\cdot u_i^2$.

Решение. Для значений $m=2, n=6$ получаем таблицу

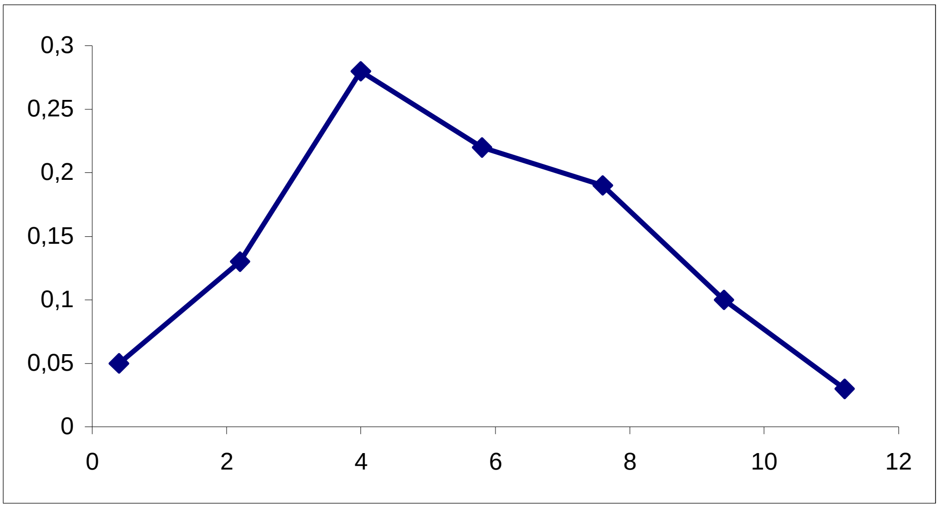
x_i	0,4	2,2	4,0	5,8	7,6	9,4	11,2
m_{x_i}	5	13	28	22	19	10	3

Построим полигон относительный частот $W_i=\frac{m_{x_i}}{N}$. Вычислим относительные частоты и занесем в таблицу:

x_i	0,4	2,2	4	5,8	7,6	9,4	11,2

m_{x_i}	5	13	28	22	19	10	3
W_i	0,05	0,13	0,28	0,22	0,19	0,1	0,03

Полигон имеет вид:



Вычислим среднее выборочное $\overline{X}$, выборочную дисперсию D_x и среднее квадратическое отклонение σ_x.

Перейдем к условным значениям $u_i = \frac{x_i - c_x}{1,8}$ и, взяв за ложный нуль c_x значение с наибольшей частотой $c_x = 4$. Составим расчетную таблицу:

u_i	-2	-1	0	1	2	3	4	**Сумма**
m_{x_i}	5	13	28	22	19	10	3	**100**
$m_{x_i} \cdot u_i$	-10	-13	0	22	38	30	12	**79**
$m_{x_i} \cdot u_i^2$	20	13	0	22	76	90	48	**269**

Получаем $\overline{X} = \left(\frac{1}{N} \sum_{i=1}^{7} m_{x_i} \cdot u_i \right) \cdot 1,8 + c_x = 0,79 \cdot 1,8 + 4 = 5,422$.

$$D_x = 1,8^2 \cdot \left(\frac{1}{N} \sum_{i=1}^{7} m_{x_i} \cdot u_i^2 - \left(\frac{1}{N} \sum_{i=1}^{7} m_{x_i} \cdot u_i \right)^2 \right) = 1,8^2 \cdot \left(2,69 - 0,79^2 \right) \approx 6,694.$$

$\sigma_x = \sqrt{D_x} \approx 2,587$.

По критерию χ^2 проверим гипотезу о нормальном распределении генеральной совокупности при уровне значимости $\alpha = 0,05$.

Выдвинем гипотезу H_0: генеральная совокупность имеет нормальное распределение с параметрами $a = 5,422$ и $\sigma = 2,587$.
Рассчитываем теоретические частоты n_i^0 по формуле
$n_i^0 = \frac{nh}{\sigma_x}\varphi(u_i)$, где $u_i = \frac{x_i - \overline{X}}{\sigma_x}$, $h = 1,8$ – шаг, $\varphi(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$.
Вычисления представим в виде таблицы:

x_i	u_i	$\varphi(u_i)$	n_i^0
0,4	-1,94	0,0608	4,23
2,2	-1,25	0,1826	12,71
4	-0,55	0,3429	23,86
5,8	0,15	0,3945	27,45
7,6	0,84	0,2803	19,50
9,4	1,54	0,1219	8,48
11,2	2,23	0,0332	2,31

Наблюдаемое значение критерия вычислим по формуле
$\chi^2_{набл} = \sum_{i=1}^{7}\frac{(n_i - n_i^0)^2}{n_i^0}$ и вычисления представим в виде таблицы.

x_i	n_i	n_i^0	$\lvert n_i - n_i^0 \rvert$	$(n_i - n_i^0)^2$	$\frac{(n_i - n_i^0)^2}{n_i^0}$
0,4	5	4,23	0,770	0,592	0,140
2,2	13	12,71	0,295	0,087	0,007
4	28	23,86	4,141	17,152	0,719
5,8	22	27,45	5,449	29,689	1,082
7,6	19	19,50	0,503	0,253	0,013
9,4	10	8,48	1,518	2,305	0,272
11,2	3	2,31	0,690	0,476	0,206
				Сумма	**2,438**

По таблице критических значений $\chi^2_{кр}$ при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = l - 3 = 7 - 3 = 4$ найдем $\chi^2_{кр} \approx 9,5$. Так как
$\chi^2_{набл} = 2,438 < \chi^2_{кр} = 9,5$, нулевую гипотезу можно принять при данном уровне значимости.

2 Построение уравнения прямой регрессии.

Двумерная выборка результатов совместных измерений признаков x и y объемом $N=100$ измерений задана корреляционной таблицей:

	y_1	y_2	y_3	y_4	y_5	m_{x_i}
x_1	2	3	–	–	–	5
x_2	3	8	2	–	–	13
x_3	–	$8+m$	$12+n$	–	–	$20+(m+n)$
x_4	–	–	$16-m$	$14-n$	–	$30-(m+n)$
x_5	–	–	9	10	–	19
x_6	–	–	3	6	1	10
x_7	–	–	–	1	2	3
m_{y_j}	5	$19+m$	$42+n-m$	$31-n$	3	$N=100$

где $x_i = 0,2\cdot m + (i-1)\cdot 0,3\cdot n$, $y_j = 0,5\cdot m + (j-1)\cdot 0,2\cdot n$.

13.2.1 Найти $\overline{Y}$ и σ_y для выборки

y_j	y_1	y_2	y_3	y_4	y_5
m_{y_j}	5	$19+m$	$42+n-m$	$31-n$	3

(Расчеты $\overline{Y}$ и σ_y можно провести аналогично расчетам $\overline{X}$ и σ_x в задаче 13.1.2)

13.2.2 Построить уравнение прямой регрессии Y на X в виде $\overline{y_x} = ax+b$. $\overline{X}$ и σ_x следует взять из задачи 13.1.2.

13.2.3 На графике изобразить корреляционное поле, то есть нанести точки (x_i, y_j) и построить прямую $\overline{y_x} = ax+b$.

Примечание. Уравнение регрессии сначала рекомендуется найти в виде $\frac{\overline{y_x}-\overline{Y}}{\sigma_y} = r\cdot\frac{x-\overline{X}}{\sigma_x}$, где r - выборочный коэффициент корреляции, для расчета которого можно воспользоваться методом четырех полей.

Решение. Заполним корреляционную таблицу данными согласно варианту $m=2, n=6$.

Получаем:

	1	**2,2**	**3,4**	**4,6**	**5,8**	m_{x_i}
0,4	2	3				5
2,2	3	8	2			13
4		10	18			28
5,8			14	8		22
7,6			9	10		19
9,4			3	6	1	10
11,2				1	2	3
m_{y_j}	5	21	46	25	3	

Найдем выборочное среднее $\overline{Y}$ и выборочное среднее квадратическое отклонение σ_y для выборки

y_j	1	2,2	3,4	4,6	5,8
m_{y_j}	5	21	46	25	3

Расчеты $\overline{Y}$ и σ_y проведем аналогично расчетам $\overline{X}$ и σ_x в задаче 13.1.2.

Перейдем к условным значениям $v_j = \dfrac{y_j - c_y}{1,2}$ и, взяв за ложный нуль c_y значение с наибольшей частотой $c_y = 3,4$. Составим расчетную таблицу:

v_j	-2	-1	0	1	2	**Сумма**
m_{v_j}	5	21	46	25	3	**100**
$m_{v_j} \cdot v_j$	-10	-21	0	25	6	**0**
$m_{v_j} \cdot v_j^2$	20	21	0	25	12	**78**

Получаем $\overline{Y} = \left(\dfrac{1}{N}\sum_{i=1}^{7} m_{y_j} \cdot v_j\right) \cdot 1,2 + c_y = 0 \cdot 1,2 + 3,4 = 3,4$.

$$D_y = 1,2^2 \cdot \left(\frac{1}{N}\sum_{i=1}^{7} m_{y_j} \cdot v_j^2 - \left(\frac{1}{N}\sum_{i=1}^{7} m_{y_j} \cdot v_j\right)^2\right) = 1,2^2 \cdot (0,78 - 0) \approx 1,123.$$

$\sigma_y = \sqrt{D_y} \approx 1,06$.

Построим уравнение прямой регрессии Y на X в виде $\overline{y_x} = ax + b$. Уравнение регрессии сначала найдем в виде $\frac{\overline{y_x} - \overline{Y}}{\sigma_y} = r \cdot \frac{x - \overline{X}}{\sigma_x}$, где r - выборочный коэффициент корреляции.

Вычислим r по формуле:

$$r_в = \frac{\sum m_{ij} x_i y_j - N \cdot \overline{X} \cdot \overline{Y}}{N \cdot \sigma_x \cdot \sigma_y} = \frac{2059,48 - 100 \cdot 5,422 \cdot 3,4}{100 \cdot 2,587 \cdot 1,06} \approx 0,788$$

Получаем:

$$\frac{\overline{y_x} - 3,4}{1,06} = 0,788 \cdot \frac{x - 5,422}{2,587},$$

$$\overline{y_x} = 0,323x + 1,649.$$

На графике изобразим корреляционное поле, то есть точки (x_i, y_j) и прямую $\overline{y_x} = 0,323x + 1,649$.