Контрольная работа из учебно-методического пособия «Теория вероятностей и математическая статистика» под ред. проф. Н.Ш. Кремера. – М.:ВЗФЭИ, 2008.

ВАРИАНТ 1

(для студентов, номера личных дел которых оканчиваются цифрой 1)

Контрольная работа №3

Задача 1. В первом ящике 2 красных и 5 синих папок, во втором – 4 красных и 3 синих. Из первого ящика переложили 2 папки во второй, после чего из второго ящика наудачу достали одну папку. Какова вероятность того, что она красного цвета?

Решение. Введем полную группу гипотез:
$H1$ = (Из первого ящика во второй переложили 2 красные и 0 синих папок),
$H2$ = (Из первого ящика во второй переложили 1 красную и 1 синюю папку),
$H3$ = (Из первого ящика во второй переложили 0 красных и 2 синие папки).

Найдем вероятности гипотез по классическому определению вероятности:
$$P(H1)=\frac{C_2^2}{C_7^2}=\frac{1}{21},\quad P(H2)=\frac{2\cdot 5}{C_7^2}=\frac{10}{21},\quad P(H3)=\frac{C_5^2}{C_7^2}=\frac{10}{21}.$$

После перекладывания папок во втором ящике станет (при выполнении соответствующей гипотезы):
$H1$ - 6 красных и 3 синих папки.
$H2$ - 5 красных и 4 синих папки.
$H3$ - 4 красных и 5 синих папок.

Введем событие A = (Из второго ящика вынули красную папку). Найдем условные вероятности по классическому определению вероятности (отношение числа красных папок к общему числу папок):
$$P(A\,|\,H1)=\frac{6}{6+3}=\frac{6}{9}=\frac{2}{3},\quad P(A\,|\,H2)=\frac{5}{5+4}=\frac{5}{9},\quad P(A\,|\,H3)=\frac{4}{4+5}=\frac{4}{9}.$$

Найдем вероятность события A по формуле полной вероятности:
$$P(A)=\sum P(A\,|\,Hi)P(Hi)=\frac{1}{21}\cdot\frac{2}{3}+\frac{10}{21}\cdot\frac{5}{9}+\frac{10}{21}\cdot\frac{4}{9}=\frac{32}{63}\approx 0,508.$$

Ответ: 0,508.

Задача 2. Вероятность сдачи студентом контрольной работы в срок равна 0,7. Найти вероятность того, что из 5 студентов вовремя сдадут контрольную работу:

а) ровно 3 студента; б) хотя бы один студент.

Решение. Имеем схему Бернулли с параметрами $p=0,7$ (вероятность того, что контрольная будет сдана студентом в срок), $n=5$ (число студентов). Будем использовать формулу Бернулли (вероятность того, что из n студентов ровно k сдадут работы в срок): $P_n(k)=C_n^k p^k(1-p)^{n-k}=C_5^k\cdot 0,7^k\cdot 0,3^{5-k}$.

Найдем вероятность того, что из 5 студентов вовремя сдадут контрольную работу ровно 3 студента.

$$P_5(3)=C_5^3\cdot 0,7^3\cdot 0,3^2=\frac{5!}{3!2!}\cdot 0,7^3\cdot 0,3^2=\frac{4\cdot 5}{1\cdot 2}\cdot 0,7^3\cdot 0,3^2=0,3087.$$

Найдем вероятность того, что из 5 студентов вовремя сдаст контрольную работу хотя бы один студент.

$$P_5(k\ge 1)=1-P_5(k<1)=1-P_5(0)=1-C_5^0\cdot 0,7^0\cdot 0,3^5=1-0,3^5\approx 0,998.$$

Ответ: а) 0,3087; б) 0,998.

Задача 3. Всхожесть хранящегося на складе зерна равна 80%. Отбираются 400 зерен. Определить вероятность того, что из отобранных зерен взойдут:

а) ровно 303; б) от 250 до 330.

Решение. Имеем схему Бернулли с параметрами: $n=400$, $p=0,8$ (вероятность прорасти для каждого семени), $q=1-p=0,2$.

А) $k=303$.

Используем приближенную формулу: локальную теорему Лапласа:

$P_n(k)\approx\frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right)$, значения функции берутся из таблицы. Подставляем:

$$P_{400}(303)\approx\frac{1}{\sqrt{400\cdot 0,8\cdot 0,2}}\varphi\left(\frac{303-400\cdot 0,8}{\sqrt{400\cdot 0,8\cdot 0,2}}\right)=0,125\cdot\varphi(-2,13)=0,125\cdot 0,0413\approx 0,005..$$

Б) $k_1=250$, $k_2=330$.

Применим интегральную формулу Муавра-Лапласа. Вычисляем:

$$x_1=\frac{k_1-np}{\sqrt{npq}}=\frac{250-400\cdot 0,8}{\sqrt{400\cdot 0,8\cdot 0,2}}=-8,75,$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{330 - 400 \cdot 0,8}{\sqrt{400 \cdot 0,8 \cdot 0,2}} = 1,25.$$

Подставляем в формулу:

$$P_n(k_1, k_2) \approx \Phi(x_2) - \Phi(x_1) = \Phi(1,25) - \Phi(-8,75) = \Phi(1,25) + \Phi(8,75) =$$
$$= 0,3944 + 0,5 = 0,8944.$$

Ответ: 0,8944.

Задача 4. Котировки акций могут быть размещены в Интернете на трех сайтах. Материал есть на первом сайте с вероятностью 0,7, на втором – с вероятностью 0,6, на третьем – с вероятностью 0,8. Студент переходит к новому сайту только в том случае, если не найдет данных на предыдущем. Составить закон распределения числа сайтов, которые посетит студент.

Найти:

а) функцию распределения этой случайной величины и построить ее график;

б) математическое ожидание и дисперсию этой случайной величины.

Решение. Обозначим за X дискретную случайную величину, равную числу посещенных студентом сайтов. Она может принимать значения 1, 2 и 3. Найдем соответствующие вероятности.

$X = 1$, если студент нашел информацию на первом сайте, $P(X = 1) = 0,7$.

$X = 2$, если студент не нашел информацию на первом сайте (вероятность 0,3) и нашел на втором (0,6), $P(X = 2) = 0,3 \cdot 0,6 = 0,18$.

$X = 3$, если студент не нашел информацию ни на первом, ни на втором сайте, и стал смотреть третий, $P(X = 3) = 0,3 \cdot 0,4 = 0,12$.

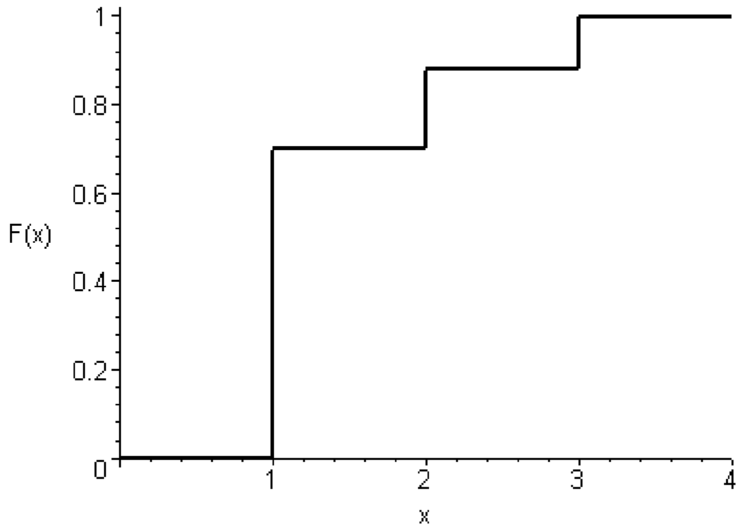
Получаем закон распределения:

x_i	1	2	3
p_i	0,7	0,18	0,12

Найдем функцию распределения $F(x) = P(X < x)$, то есть

при $x \le 1$, $F(x) = 0$,
при $1 < x \le 2$, $F(x) = 0+0,7 = 0,7$,
при $2 < x \le 3$, $F(x) = 0,7+0,18=0,88$,
при $x>3$, $F(x) =0,88+0,12 = 1$.

Построим график $F(x)$:



Математическое ожидание:

$$MX = \sum x_i p_i = 1\cdot 0,7 + 2\cdot 0,18 + 3\cdot 0,12 = 1,42$$

Дисперсия

$$DX = \sum x_i^2 p_i - (MX)^2 = 1\cdot 0,7 + 4\cdot 0,18 + 9\cdot 0,12 - 1,42^2 = 2,5 - 1,42^2 = 0,4836.$$

Задача 5. Случайная величина X имеет нормальный закон распределения с параметрами a и σ^2.

Найти:

а) параметр σ^2, если известно, что математическое ожидание $M(X)$=5 и вероятность $P(2 < X < 8) = 0,9973$;

б) вероятность $P(X < 0)$.

Решение. Используем формулу для нахождения вероятности попадания нормальной случайной величины в интервал:

$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$, где $\Phi(x)$ - функция Лапласа (значения берутся из таблицы).

Получаем

$$P(2 < X < 8) = \Phi\left(\frac{8-5}{\sigma}\right) - \Phi\left(\frac{2-5}{\sigma}\right) = \Phi\left(\frac{3}{\sigma}\right) - \Phi\left(\frac{-3}{\sigma}\right) = \Phi\left(\frac{3}{\sigma}\right) + \Phi\left(\frac{3}{\sigma}\right) = 2\Phi\left(\frac{3}{\sigma}\right) = 0,9973,$$

$$\Phi\left(\frac{3}{\sigma}\right) = 0,49865,$$

$$\frac{3}{\sigma} = 3,$$

$$\sigma = 1.$$

Стандартное отклонение данного распределения равно 1. Дисперсия $\sigma^2 = 1$.

Вычислим вероятность того, что случайная величина X примет значение меньшее 0:

$$P(-\infty < X < 0) = \Phi\left(\frac{0-5}{1}\right) - \Phi\left(\frac{-\infty-5}{1}\right) = \Phi(-5) - \Phi(-\infty) =$$

$$= -\Phi(5) + \Phi(\infty) = -0,5 + 0,5 = 0.$$

Вероятность практически равна нулю.

Контрольная работа №4

Задача 1. Для проверки качества поступившей партии зерна по схеме собственно-случайной бесповторной выборки произведено 10%-ное обследование. В результате анализа установлено следующее распределение данных о влажности зерна:

Процент влажности	Менее 8	8–10	10–12	12–14	14–16	16––18	18––20	Более 20	Итого
Число проб	7	15	30	35	25	18	7	3	140

Найти:
а) вероятность того, что средний процент влажности зерна в партии отличается от ее среднего процента в выборке не более чем на 0,5% (по абсолютной величине);
б) границы, в которых с вероятностью 0,95 заключена доля зерна, влажность которого менее 12%;
в) объем выборки, при которой те же границы для доли зерна, полученные в пункте б), можно гарантировать с вероятностью 0,9876; дать ответ на тот же вопрос, если никаких предварительных данных о рассматриваемой доле нет.

Решение. Вычислим сначала числовые характеристики выборки. Построим соответствующий простой вариационный ряд, выбрав в качестве вариант середины интервалов:

x_i	7	9	11	13	15	17	19	21	**Сумма**
n_i	7	15	30	35	25	18	7	3	**140**

Найдем среднее:

$$\bar{x} = \frac{1}{n}\sum x_i n_i = \frac{1}{140}1783 = 12,736$$

Найдем исправленную дисперсию:

$$S^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 n_i = \frac{1}{139}1262,626 \approx 9,084.$$

Найдем исправленное среднеквадратичное отклонение: $S \approx 3,014$.
Расчеты в таблице ниже:

x_i	7	9	11	13	15	17	19	21	**Сумма**
n_i	7	15	30	35	25	18	7	3	**140**
$x_i n_i$	49	135	330	455	375	306	133	63	**1783**
$(x_i - \bar{x})^2 n_i$	230,289	209,333	90,3811	2,44464	128,175	327,314	274,689	204,895	**1262,626**

а) Найдем вероятность того, что средний процент влажности зерна в партии отличается от ее среднего процента в выборке не более чем на 0,5% (по абсолютной величине).
То есть, надо найти доверительную вероятность, соответствующую предельной ошибке выборки 0,5%.

Предельная ошибка: $\Delta_x = t \cdot \mu_x = t \cdot \frac{S}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$. Подставляем данные:

$$0,5 = t \cdot \frac{3,014}{\sqrt{140}}\sqrt{1-0,1},$$

$$0,5 = t \cdot 0,242,$$

$$t \approx 2,069,$$

Тогда соответствующая вероятность $\gamma = 2\Phi(t) = 2\Phi(2,07) = 2 \cdot 0,4807 = 0,9614$.

б) Найдем границы, в которых с вероятностью 0,95 заключена доля зерна, влажность которого менее 12%.
Выборочная доля зерна, влажность которого менее 12%, равна

$$w = \frac{7+15+30}{140} = \frac{52}{140} = \frac{13}{35} = 0,371.$$

Предельная ошибка для доли $\Delta_w = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$. Коэффициент

$$t = \Phi^{-1}(\gamma/2) = \Phi^{-1}(0,95/2) = \Phi^{-1}(0,475) = 1,96.$$

Получаем:

$$\Delta_w = 1,96\sqrt{\frac{0,371(1-0,371)}{140}(1-0,1)} \approx 0,076.$$

Тогда границы, в которых с вероятностью 0,95 заключена доля зерна, влажность которого менее 12% имеют вид:

$$w - \Delta_w < p < w + \Delta_w,$$

$$0,371 - 0,076 < p < 0,371 + 0,076,$$

$$0,295 < p < 0,447.$$

в) Найдем объем выборки, при которой те же границы для доли зерна, полученные в пункте б), можно гарантировать с вероятностью 0,9876
То есть нужно найти объем выборки, при котором предельная ошибка будет также равна $\Delta_w = 0,076$. Формула для объема выборки:

$$n = \frac{t^2 N w(1-w)}{\Delta_w^2 N + t^2 w(1-w)}.$$

Коэффициент $t = \Phi^{-1}(\gamma/2) = \Phi^{-1}(0,9876/2) = \Phi^{-1}(0,4938) = 2,5$, $N = \frac{140}{0,1} = 1400$.

Подставляем все данные:

$$n = \frac{2,5^2 \cdot 1400 \cdot 0,371 \cdot 0,629}{0,076^2 \cdot 1400 + 2,5^2 \cdot 0,371 \cdot 0,629} \approx 214.$$

Дадим ответ на тот же вопрос, если никаких предварительных данных о рассматриваемой доле нет. Тогда рекомендуется брать $w=0,5$. Получаем:

$$n=\frac{2,5^2\cdot 1400\cdot 0,5\cdot 0,5}{0,076^2\cdot 1400+2,5^2\cdot 0,5\cdot 0,5}\approx 227.$$

Задача 2. По данным задачи 1, используя χ^2-критерий Пирсона, на уровне значимости α = 0,05 проверить гипотезу о том, что случайная величина X – процент влажности зерна – распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

Решение. Пронормируем случайную величину X, то есть перейдем к величине $Z=\frac{x-\bar{x}}{S}$, вычислим концы интервалов по формулам $z_i=\frac{x_i-\bar{x}}{S}$, $z_{i+1}=\frac{x_{i+1}-\bar{x}}{S}$.

Вычислим теоретические (выравнивающие частоты) $n_i'=nP_i$, где $n=140$, $P_i=\Phi(z_{i+1})-\Phi(z_i)$ - вероятность попадания в интервал (z_i, z_{i+1}), $\Phi(z)$ - функция Лапласа. Для нахождения значений составим расчетную таблицу (последние два интервала объединим как малочисленные):

x_i	x_{i+1}	n_i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	P_i	n_i'	$\frac{(n_i-n_i')^2}{n_i'}$
	8	7	$-\infty$	-1,571	-0,5	-0,442	0,0581	8,128	0,157
8	10	15	-1,571	-0,908	-0,442	-0,318	0,124	17,355	0,319
10	12	30	-0,908	-0,244	-0,318	-0,096	0,2216	31,018	0,033
12	14	35	-0,244	0,4195	-0,096	0,1626	0,259	36,259	0,044
14	16	25	0,4195	1,0831	0,1626	0,3606	0,198	27,726	0,268
16	18	18	1,0831	1,7467	0,3606	0,4597	0,099	13,866	1,233
18		10	1,7467	$+\infty$	0,4597	0,5	0,0403	5,649	3,352
								Сумма	**5,406**

Сравним эмпирические и теоретические частоты, используя критерий Пирсона:

$$\chi^2=\sum\frac{(n_i-n_i')^2}{n_i'}=5,406.$$

По таблице критических точек распределения χ^2 по уровню значимости $\alpha=0,05$ и числу степеней свободы k = 7 - 3 = 4, находим $\chi^2_{\text{кр.}}$ = 9,5. Так как $\chi^2_{\text{набл.}}$ = 5,406 < $\chi^2_{\text{кр.}}$ = 9,5, то нет оснований отвергнуть гипотезу о нормальном распределении данной величины.

Построим теоретическую нормальную кривую

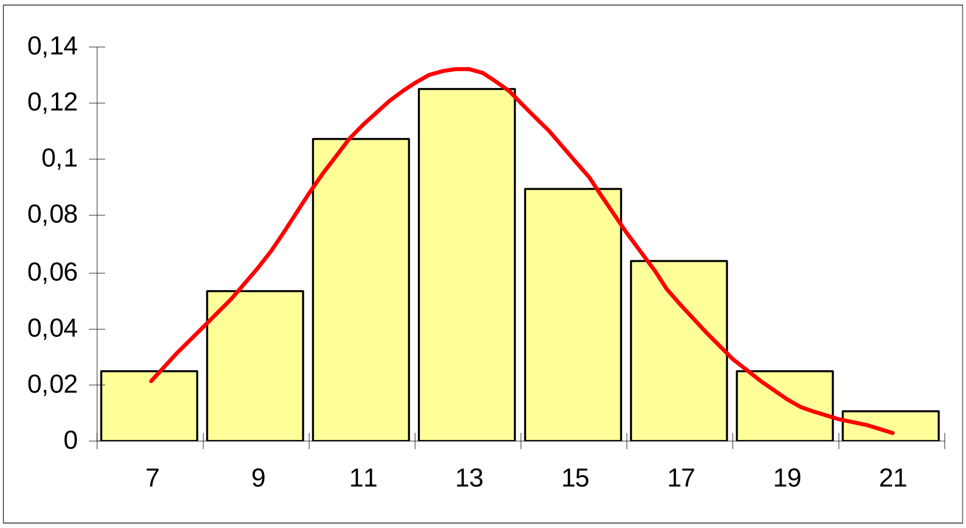
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)=\frac{1}{3,014\sqrt{2\pi}}\exp\left(-\frac{(x-12,736)^2}{2\cdot 9,084}\right)$$

и гистограмму на одном чертеже.

Составим расчетную таблицу:

x_i	7	9	11	13	15	17	19	21
частота n_i	7	15	30	35	25	18	7	3
плотность относит. частоты $\frac{n_i}{h\cdot n}=\frac{n_i}{2\cdot 140}$	0,025	0,0536	0,1071	0,125	0,0893	0,0643	0,025	0,0107
плотность распр. $f(x_i)$	0,0216	0,0614	0,1121	0,1319	0,0998	0,0486	0,0153	0,0031

По графику видно, что теоретическое и эмпирическое распределение согласуются хорошо.



Задача 3. Распределение 60 предприятий по затратам рабочего времени X (тыс. человеко-дней (чел. дн.)) и выпуску продукции Y (млн. руб.) представлены в таблице:

x \ y	30–40	40–50	50–60	60–70	70–80	Итого:
10–25	1	3	2			6
25–40	3	6	4	1		14
40–55		3	7	6	1	17
55–70		1	6	4	4	15
70–85			2	5	1	8
Итого:	4	13	21	16	6	60

Необходимо:

1) Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$, построить эмпирические линии регрессии;

2) Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции; на уровне значимости $\alpha = 0,05$ оценить его значимость и сделать вывод о тесноте и направлении связи между переменными X и Y;

в) используя соответствующее уравнение регрессии, оценить средний выпуск продукции предприятия с затратами рабочего времени 55 тыс. чел. дн.

Решение. Составим корреляционную таблицу, переходя к серединами интервалов.

$x \backslash y$	35	45	55	65	75	**Итого**
17,5	1	3	2	0	0	**6**
32,5	3	6	4	1	0	**14**
47,5	0	3	7	6	1	**17**
62,5	0	1	6	4	4	**15**
77,5	0	0	2	5	1	**8**
Итого	**4**	**13**	**21**	**16**	**6**	**60**

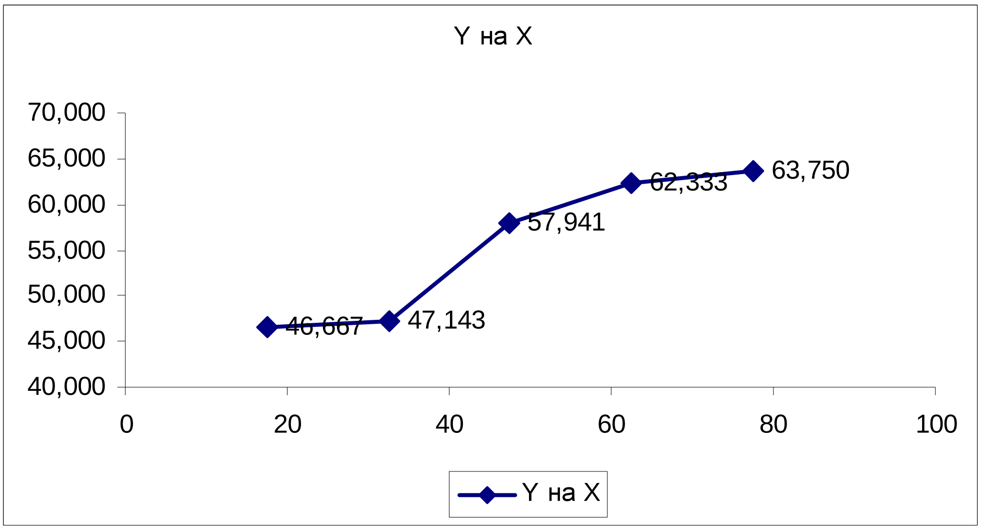
1) Найдем групповые средние по формулам: $\overline{x}_j = \dfrac{\sum_{i=1}^{5} x_i n_{ij}}{n_j}$; $\overline{y}_i = \dfrac{\sum_{j=1}^{5} y_j n_{ij}}{n_i}$.

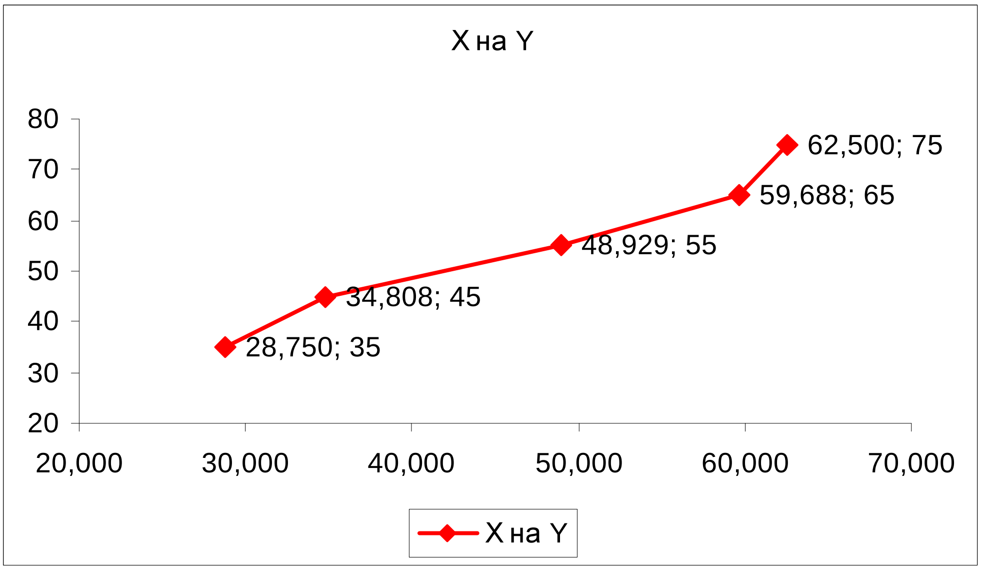
Вычисления проведем в Excel, получаем:

$\overline{x}_j$	28,750	34,808	48,929	59,688	62,500
y_j	35	45	55	65	75

x_i	$\overline{y}_i$
17,5	46,667
32,5	47,143
47,5	57,941
62,5	62,333
77,5	63,750

Построим эмпирические линии регрессии (Y на X, X на Y).





Из вида эмпирических линий регрессии можно заключить, что между переменными наблюдается линейная зависимость.

Найдем уравнения прямых линий регрессии. Вычислим необходимые величины (расчеты в таблицах ниже):

x_i	17,5	32,5	47,5	62,5	77,5	**Сумма**
n_i	6	14	17	15	8	**60**
$x_i \cdot n_i$	105	455	807,5	937,5	620	**2925**
$x_i^2 \cdot n_i$	1837,5	14787,5	38356,25	58593,75	48050	**161625**

y_j	35	45	55	65	75	**Сумма**
n_j	4	13	21	16	6	**60**

$y_j \cdot n_j$	140	585	1155	1040	450	**3370**
$y_j^2 \cdot n_j$	4900	26325	63525	67600	33750	**196100**

$$\sum_{i=1}^{5} x_i n_i = 2925, \ \sum_{i=1}^{5} x_i^2 n_i = 161625,$$

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i n_i}{n} = \frac{2925}{60} = 48,75,$$

$$s_x^2 = \frac{\sum_{i=1}^{5} x_i^2 n_i}{n} - \overline{x}^2 = \frac{161625}{60} - 48,75^2 = 317,1875$$

$$\sum_{i=1}^{5} y_j n_j = 3370, \ \sum_{j=1}^{5} y_j^2 n_j = 196100,$$

$$\overline{y} = \frac{\sum_{i=1}^{5} y_j n_j}{n} = \frac{3370}{60} = 56,167,$$

$$s_y^2 = \frac{\sum_{j=1}^{5} y_j^2 n_j}{n} - \overline{y}^2 = \frac{196100}{60} - 56,167^2 = 113,6389.$$

$$\sum_{i=1}^{5} \sum_{j=1}^{5} x_i y_j n_{ij} = 171100$$

$$\mu = \overline{xy} - \overline{x} \cdot \overline{y} = \frac{\sum_{i=1}^{5} \sum_{j=1}^{5} x_i y_j n_{ij}}{n} - \overline{x} \cdot \overline{y} = \frac{171100}{60} - 48,75 \cdot 56,167 = 113,542$$

$$b_{yx} = \frac{\mu}{s_x^2} = \frac{113,542}{317,1875} = 0,358$$

$$b_{xy} = \frac{\mu}{s_y^2} = \frac{113,542}{113,6389} = 0,999$$

Уравнения прямых регрессии:

$$y_x - \overline{y} = b_{yx}(x - \overline{x}),$$

$$y_x - 56,167 = 0,358(x - 48,75),$$

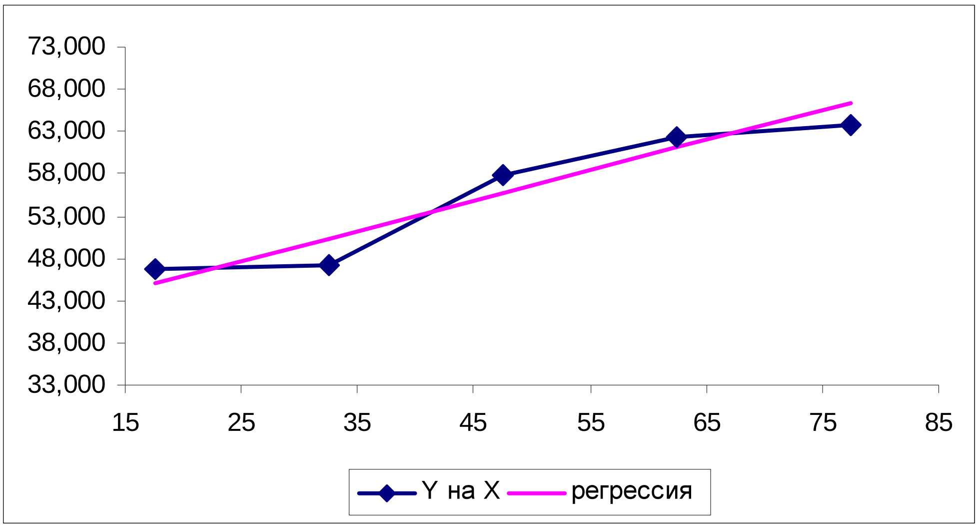
$$y_x = 0,358x + 38,715.$$

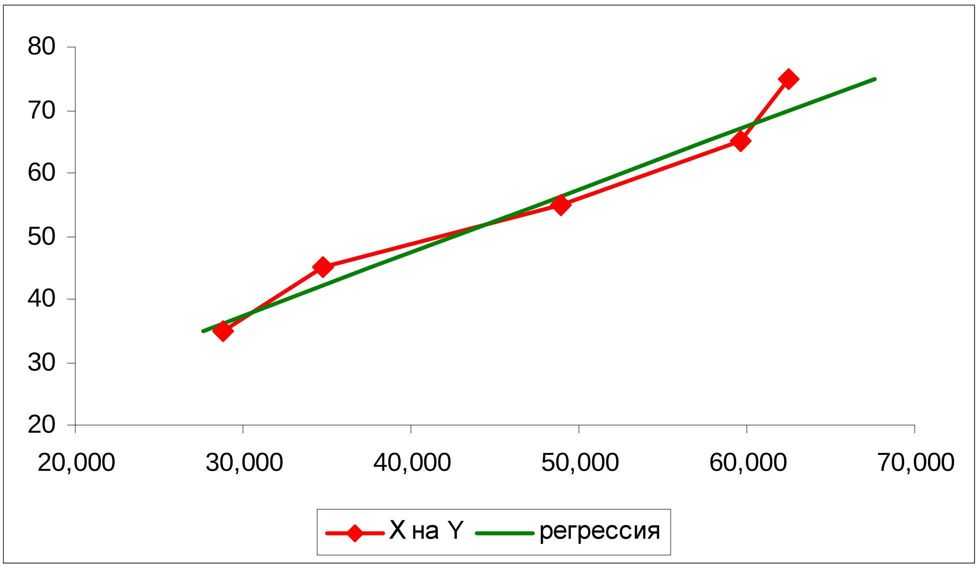
$$x_y - \overline{x} = b_{xy}(y - \overline{y}),$$

$$x_y - 48,75 = 0,999(y - 56,167),$$

$$x_y = 0,999y - 7,361.$$

Построим графики линий регрессии на одном чертеже с эмпирическими линиями регрессии.





Экономическая интерпретация полученных уравнений:
$y_x = 0,358x + 38,715$ - при увеличении числа затрат времени на 1 тыс. человеко-дней, средний выпуск продукции увеличивается на 0,358 млн. рублей.
$x_y = 0,999y - 7,361$ - при увеличении выпуска продукции на 1 млн. рублей, число затраты времени увеличиваются в среднем на 0,999 тыс. человеко-дней

Вычислим коэффициент корреляции $r = \sqrt{b_{yx}b_{xy}} = \sqrt{0,358 \cdot 0,999} \approx 0,598$

На уровне значимости $\alpha = 0,05$ оценим значимость коэффициента корреляции. Вычислим значение критерия $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,598\sqrt{60-2}}{\sqrt{1-0,598^2}} \approx 5,68$

По таблице критерия Стьюдента для уровня значимости 0,05 находим $t_{0,95;58} = 2,002$. Так как наблюдаемое значение 5,68 больше критического, коэффициент корреляции значим.

Связь между переменными X и Y относительно тесная, прямая по направлению.

Используя соответствующее уравнение регрессии, оценим средний выпуск продукции предприятия с затратами рабочего времени 55 тыс. чел. дней.
$y_x(55) = 0,358 \cdot 55 + 38,715 = 58,405$ млн. рублей.