

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Домашнее задание №2

по дисциплине

«Теория вероятностей»

Вариант 14

Выполнил:

Студент группы Р3213

Султанов А.Р.

Проверила:

Селина Е.Г.

г. Санкт-Петербург

2023г.

ИДЗ-19.1

В результате эксперимента получены данные, записанные в виде статистического ряда. Требуется:

- а) записать значения результатов эксперимента в виде вариационного ряда;
- б) найти размах варьирования и разбить его на 9 интервалов;
- в) построить полигон частот, гистограмму относительных частот и рафи эмпирической функции распределения;
- г) найти числовые характеристики выборки, \bar{x} , D_v ;
- д) приняв в качестве нулевой гипотезу H_0 : генеральная совокупность, из которой извлечена выборка, имеет нормальное распределение, проверить ее, пользуясь критерием Пирсона при уровне значимости $\alpha = 0,025$;
- е) найти доверительные интервалы для математического ожидания и среднего квадратичного отклонения при надежности $\gamma = 0,95$.

0,026	0,034	0,028	0,036	0,030	0,038	0,041	0,038	0,030	0,028
0,028	0,030	0,034	0,038	0,040	0,036	0,034	0,023	0,032	0,026
0,034	0,032	0,024	0,036	0,032	0,026	0,030	0,028	0,038	0,034
0,038	0,041	0,028	0,026	0,030	0,034	0,032	0,040	0,036	0,032
0,030	0,036	0,034	0,032	0,023	0,032	0,028	0,032	0,026	0,038
0,026	0,032	0,028	0,040	0,038	0,030	0,032	0,024	0,036	0,030
0,024	0,032	0,030	0,036	0,028	0,041	0,032	0,038	0,034	0,026
0,041	0,034	0,023	0,038	0,026	0,030	0,028	0,036	0,040	0,028
0,030	0,026	0,034	0,028	0,024	0,036	0,032	0,030	0,038	0,034
0,028	0,034	0,040	0,036	0,030	0,038	0,023	0,034	0,032	0,026

- а) Располагаем значения результатов эксперимента в порядке возрастания, т.е. записываем вариационный ряд:

0,023	0,023	0,023	0,023	0,024	0,024	0,024	0,024	0,026	0,026
0,026	0,026	0,026	0,026	0,026	0,026	0,026	0,026	0,028	0,028
0,028	0,028	0,028	0,028	0,028	0,028	0,028	0,028	0,028	0,028
0,030	0,030	0,030	0,030	0,030	0,030	0,030	0,030	0,030	0,030
0,030	0,030	0,030	0,032	0,032	0,032	0,032	0,032	0,032	0,032
0,032	0,032	0,032	0,032	0,032	0,032	0,032	0,034	0,034	0,034
0,034	0,034	0,034	0,034	0,034	0,034	0,034	0,034	0,034	0,034
0,036	0,036	0,036	0,036	0,036	0,036	0,036	0,036	0,036	0,036
0,038	0,038	0,038	0,038	0,038	0,038	0,038	0,038	0,038	0,038
0,038	0,040	0,040	0,040	0,040	0,040	0,041	0,041	0,041	0,041

б) Находим размах варьирования:

$$\omega = x_{\max} - x_{\min} = 0,041 - 0,023 = 0,018.$$

По формуле $h = \omega/l$, где l - число интервалов, вычисляем длину частичного интервала $h = 0,018/9 = 0,002$

В качестве границы первого интеграла можно выбрать значение $x_{\min} + dh$, $d = \overline{1, l}$.

Находим середины интервалов: $x'_i = (x_i + x_{i+1})/2$

Подсчитываем число значений результатов эксперимента, попавших в каждый интервал, т.е. находим частоты интервалов n_i .

Далее вычисляем относительные частоты $W_i = n_i/n$ ($n = 100$) и их плотности W_i/h .

Все полученные результаты помещаем в таблицу:

i	Границы интервала	Середина интервала x'_i	Частота интервала n_i	Относительная частота W_i	Плотность относительной частоты W_i/h
1	0,023 - 0,025	0,024	8	0,08	40

2	0,025 - 0,027	0,026	10	0,10	50
3	0,027 - 0,029	0,028	12	0,12	60
4	0,029 - 0,031	0,030	13	0,13	65
5	0,031 - 0,033	0,032	14	0,14	70
6	0,033 - 0,035	0,034	13	0,13	65
7	0,035 - 0,037	0,036	10	0,10	50
8	0,037 - 0,039	0,038	11	0,11	55
9	0,039 - 0,041	0,040	9	0,09	45

в) Строим полигон частот и гистограмму относительных частот. Находим значения эмпирической функции распределения:

$$F^*(x) = n_x / n$$

$$F^*(0.023) = 0$$

$$F^*(0.025) = 0,08$$

$$F^*(0.027) = 0,18$$

$$F^*(0.029) = 0,3$$

$$F^*(0.031) = 0,43$$

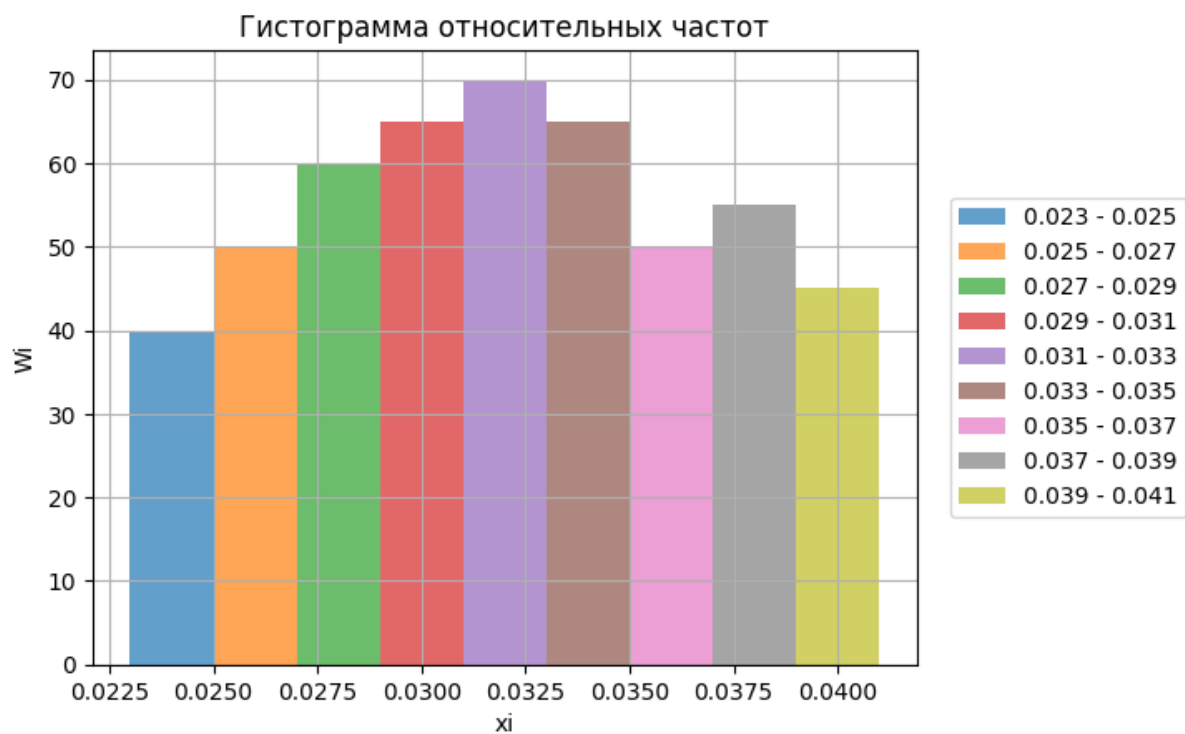
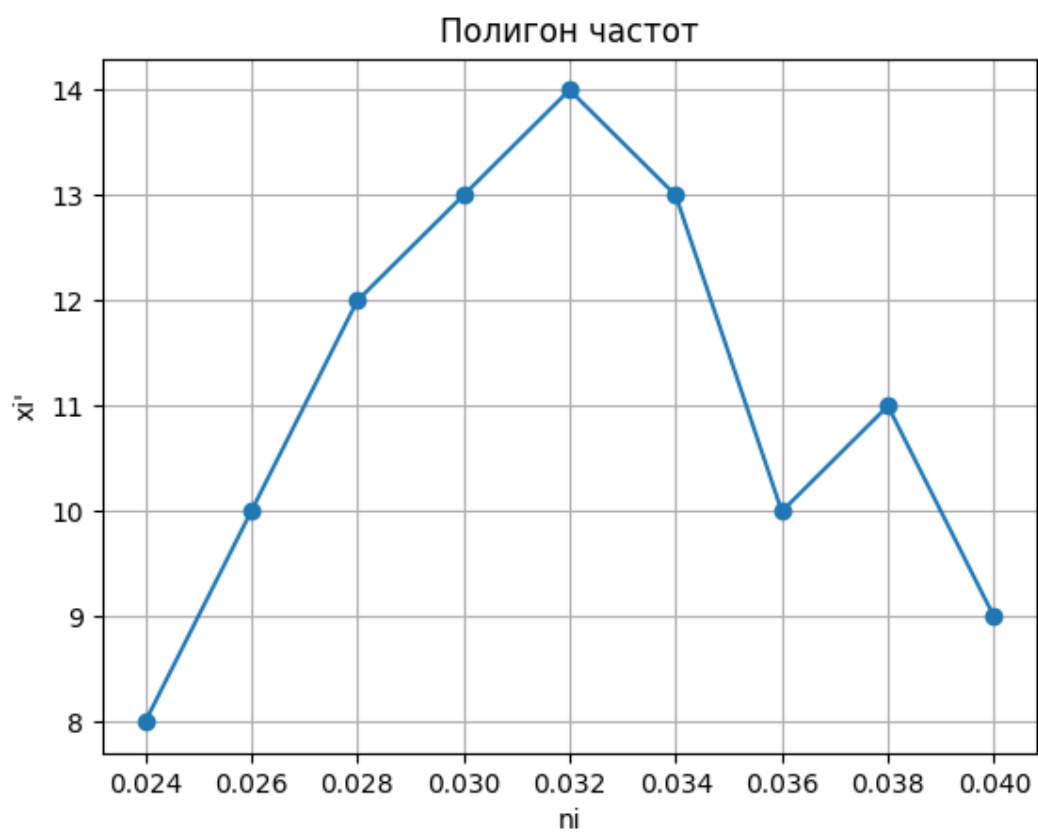
$$F^*(0.033) = 0,57$$

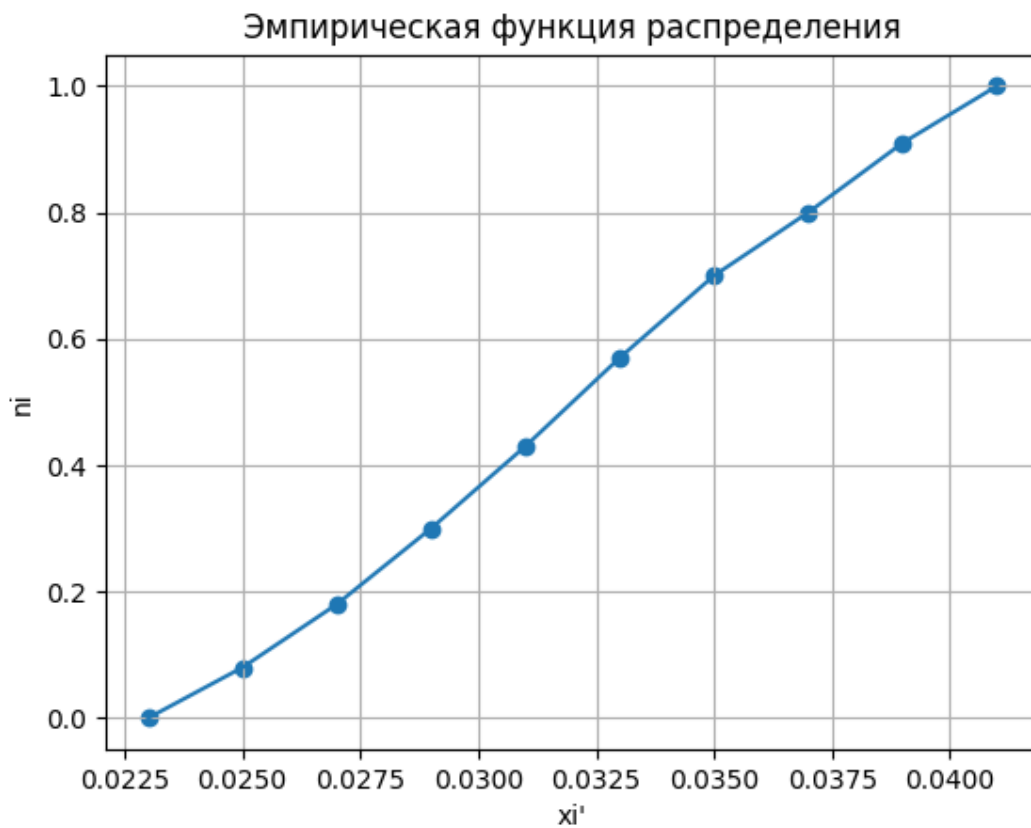
$$F^*(0.035) = 0,7$$

$$F^*(0.037) = 0,8$$

$$F^*(0.039) = 0,91$$

$$F^*(0.041) = 1,0$$





г) Находим выборочное среднее:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x'_i n_i$$

И выборочную дисперсию:

$$D_B = \frac{1}{n} \sum_{i=1}^k (x'_i - \bar{x})^2 n_i = \frac{1}{n} \sum_{i=1}^k (x'_i)^2 n_i - \bar{x}^2$$

Для этого составляем расчетную таблицу. Из нее получаем:

$$\bar{x} = 3,206/100 = 0,03206$$

$$D_B = 0,105084/100 - 0,03206^2 = 0,000023,$$

$$\sigma_B = \sqrt{D_B} = 0,004796$$

m_i	Границы интервала	x_i'	n_i	$n_i x_i'$	$(x_i')^2$	$n_i (x_i')^2$
1	0,023 - 0,025	0,024	8	0,192	0,000576	0,004608
2	0,025 - 0,027	0,026	10	0,260	0,000676	0,006760
3	0,027 - 0,029	0,028	12	0,336	0,000784	0,009408
4	0,029 - 0,031	0,03	13	0,390	0,000900	0,011700
5	0,031 - 0,033	0,032	14	0,448	0,001024	0,014336
6	0,033 - 0,035	0,034	13	0,442	0,001156	0,015028
7	0,035 - 0,037	0,036	10	0,360	0,001296	0,012960
8	0,037 - 0,039	0,038	11	0,418	0,001444	0,015884
9	0,039 - 0,041	0,04	9	0,360	0,001600	0,014400

Выборочная дисперсия является *смещенной оценкой генеральной дисперсии*, а исправленная дисперсия - *несмещенной оценкой*:

$$D_{\tilde{B}} = \frac{n}{n-1} D_B = \frac{100}{99} * 0,000023 = 0,00002323$$

$$\sigma_{\tilde{B}} = \sqrt{D_{\tilde{B}}} = 0,00481975$$

д) Согласно критерию Пирсона необходимо сравнить эмпирические и теоретические частоты. Эмпирические частоты даны. Найдем теоретические частоты. Для этого пронумеруем X , т.е. перейдем к СВ:

$$z = (x - \bar{x}) / \sigma_B$$

и вычислим концы интервалов:

$$z_i = (x_i - \bar{x}) / \sigma_B, z_{i+1} = (x_{i+1} - \bar{x}) / \sigma_B,$$

причем наименьшее значение z , т.е. z_1 , положим стремящимся к $-\infty$, а наибольшее, т.е. z_{m+1} , - к $+\infty$.

Результаты занесем в таблицу.

i	x_i	x_{i+1}	$x_i - \bar{x}$	$x_{i+1} - \bar{x}$	z_i	z_{i+1}
1	0,023	0,025		-0,00706		-1,47
2	0,025	0,027	-0,00706	-0,00506	-1,47	-1,06
3	0,027	0,029	-0,00506	-0,00306	-1,06	-0,64
4	0,029	0,031	-0,00306	-0,00106	-0,64	-0,22
5	0,031	0,033	-0,00106	0,00094	-0,22	0,2
6	0,033	0,035	0,00094	0,00294	0,2	0,61
7	0,035	0,037	0,00294	0,00494	0,61	1,03
8	0,037	0,039	0,00494	0,00694	1,03	1,45
9	0,039	0,041	0,00694		1,45	

Находим теоретические вероятности P_i и теоретические частоты:

$n'_i = nP_i = 100P_i$. Составляем расчетную таблицу:

i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_i) - \Phi(z_{i+1})$	$n'_i = 100P_i$
1		-1,47	-0,5	-0,4292	0,0708	7,08
2	-1,47	-1,06	-0,4292	-0,3554	0,0738	7,38
3	-1,06	-0,64	-0,3554	-0,2389	0,1165	11,65
4	-0,64	-0,22	-0,2389	-0,0871	0,1518	15,18
5	-0,22	0,2	-0,0871	0,0793	0,1664	16,64
6	0,2	0,61	0,0793	0,2291	0,1498	14,98
7	0,61	1,03	0,2291	0,3485	0,1194	11,94
8	1,03	1,45	0,3485	0,4265	0,078	7,8
9	1,45		0,4265	0,5	0,0735	7,35

Вычислим наблюдаемое значение критерия Пирсона. Для этого составим расчетную таблицу. Последние два столбца служат для контроля вычислений по формуле

$$\chi^2_{\text{набл}} = \frac{1}{n} \sum_{i=1}^k n_i^2 - n$$

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$	n_i^2	$\frac{n_i^2}{n'_i}$
1	8	7,08	0,92	0,8464	0,12	64	9,04
2	10	7,38	2,62	6,8644	0,93	100	13,55
3	12	11,65	0,35	0,1225	0,011	144	12,361
4	13	15,18	-2,18	4,7524	0,313	169	11,133
5	14	16,64	-2,64	6,9696	0,419	196	11,779
6	13	14,98	-1,98	3,9204	0,262	169	11,282
7	10	11,94	-1,94	3,7636	0,315	100	8,375
8	11	7,8	3,2	10,24	1,313	121	15,513
9	9	7,35	1,65	2,7225	0,37	81	11,02
\sum_i	100	100			4,053		104,053

Контроль: $\frac{\sum n_i^2}{n} - n = \frac{\sum (n_i - n'_i)^2}{n} = 104,053 - 100 = 4,053$

По таблице критических точек распределения χ^2 , уровню значимости $\alpha = 0,025$ и числу степеней свободы $k = l - 3 = 9 - 3 = 6$ (l - число интервалов) находим $\chi^2 = 14,4$

Так как $\chi^2_{\text{набл}} < \chi^2$, то гипотеза H_0 о нормальном распределении генеральной совокупности принимается.

е) Если СВ X генеральной совокупности распределена нормально, то с надежностью γ можно утверждать, что математическое ожидание a СВ X покрывается доверительным интервалом

$$\left(\bar{x} - \frac{\sigma_{\tilde{B}}}{\sqrt{n}} t_{\gamma}; \bar{x} + \frac{\sigma_{\tilde{B}}}{\sqrt{n}} t_{\gamma} \right), \text{ где } \delta = \frac{\sigma_{\tilde{B}}}{\sqrt{n}} t_{\gamma} - \text{точность оценки.}$$

В нашем случае $\bar{x} = 0,03206$, $\sigma_{\tilde{B}} = 0,00481975$, $n = 100$.

Из прил. 4 для $\gamma = 0,95$ находим: $\Phi(t_{\gamma}) = \frac{0,95}{2} = 0,475$, $t_{\gamma} = 1,96$, $\delta = 0,00094$. Доверительным интервалом для a будет $(0,03112; 0,033)$. Доверительный интервал, покрывающий среднее квадратичное отклонение σ с заданной надежностью γ , $(\sigma_{\tilde{B}}(1 - q); \sigma_{\tilde{B}}(1 + q))$, где q находится по данным γ и n из прил. 9. При $\gamma = 0,95$ и $n = 100$ имеем: $q = 0,143$. Доверительным интервалом для σ будет $(0,00413; 0,00551)$.

ИДЗ-19.2

Дана таблица распределения 100 заводов по производственным средствам X (тыс. ден. ед.) и по сточной выработке Y (т). Известно, что между X и Y существует линейная корреляционная зависимость. Требуется:

- найти уравнение прямой регрессии на x ;
- построить уравнение эмпирической линии регрессии и случайные точки выборки (X, Y) .

$X \backslash Y$	14	17	20	23	26	29	32	35	m_x
1,8	2	4	6						12
2,4		2	7	6					15
3			6	8	5				19
3,6				8	14	4			26
4,2					3	6	8		17
4,8							5	6	11
m_y	2	6	19	22	22	10	13	6	100

Для подсчета числовых характеристик (выборочных средних \bar{x} и \bar{y} , выборочных средних квадратичных отклонений s_x и s_y и выборочного корреляционного момента s_{xy}) составляем расчетную таблицу. При заполнении таблицы осуществляем контроль по строкам и столбцам:

X \ Y	14	17	20	23	26	29	32	35	mx	mxi xi	sum mij yj	xi^2 mxi	xi sum mij yj
1,8	2	4	6						12	21,6	216	38,88	388,8
2,4		2	7	6					15	36	312	86,4	748,8
3			6	8	5				19	57	434	171	1302
3,6				8	14	4			26	93,6	664	336,96	2390,4
4,2					3	6	8		17	71,4	508	299,88	2133,6
4,8							5	6	11	52,8	370	253,44	1776
my	2	6	19	22	22	10	13	6	100	332,4	2504	1186,56	8739,6
myj yj	28	102	380	506	572	290	416	210	2504				
sum mij xi	3,6	12	45,6	67,2	78	39,6	57,6	28,8	332,4				
yj^2 myj	392	1734	7600	11638	14872	8410	13312	7350	65308				
yj sum mij xi	50,4	204	912	1545,6	2028	1148,4	1843,2	1008	8739,6				

$$\sum_{i=1}^6 m_{xi} = \sum_{j=1}^8 m_{y_j} = n = 100$$

$$\sum_{i=1}^6 \sum_{j=1}^8 m_{ij} x_i = \sum_{i=1}^6 m_{x_i} x_i = 332,4$$

$$\sum_{i=1}^6 \sum_{j=1}^8 m_{ij} y_j = \sum_{j=1}^8 m_{y_j} y_j = 2504$$

$$\sum_{i=1}^6 (x_i \sum_{j=1}^8 m_{ij} y_j) = \sum_{j=1}^8 (y_j \sum_{i=1}^6 m_{ij} x_i) = 8739,6$$

Вычисляем выборочные средние \bar{x} и \bar{y} ; $i = \overline{1, 6}$; $j = \overline{1, 8}$:

$$\bar{x} = \frac{\sum \sum m_{ij} x_i}{n} = \frac{\sum m_{x_i} x_i}{n} = \frac{332,4}{100} = 3,324$$

$$\bar{y} = \frac{\sum m_{y_j} y_j}{n} = \frac{2504}{100} = 25,04$$

Выборочные дисперсии находим по формулам:

$$s_{x_i}^2 = \frac{n}{n-1} (\sum m_{x_i} x_i^2 - \frac{1}{n} (\sum m_{x_i} x_i)^2) = \frac{1}{99} (1186,56 - \frac{1}{100} (332,4)^2) =$$

$$\approx 0,825$$

$$s_{y_j}^2 = \frac{n}{n-1} (\sum m_{y_j} y_j^2 - \frac{1}{n} (\sum m_{y_j} y_j)^2) = \frac{1}{99} (65308 - \frac{1}{100} (2504)^2) =$$

$$\approx 26,342$$

Корреляционный момент вычисляем по формуле

$$s_{xy} = \frac{1}{n-1} (\sum \sum m_{ij} x_i y_j - \frac{1}{n} (\sum m_{x_i} x_i) (\sum m_{y_j} y_j)) =$$

$$= \frac{1}{99} (8739,6 - \frac{1}{100} (332,4 * 2504)) = 4,205$$

Оценкой теоретической линии регрессии является эмпирическая линия регрессии, уравнение которой имеет вид

$$y = \bar{y} + r_{xy} \frac{s_y}{s_x} (x - \bar{x}),$$

где $s_x = \sqrt{0,825} \approx 0,908$; $s_y = \sqrt{26,342} \approx 5,132$;

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{4,205}{0,908 * 5,132} \approx 0,902$$

Составляем уравнение эмпирической линии регрессии y на x :

$$y = 25,04 + 0,902 * \frac{5,132}{0,908} (x - 3,324)$$

$$y = 5,098x + 8,094$$

Линия регрессии и случайные точки

