

Теория Вероятностей. Домашнее задание №2

Бободжонов Комронджон, Р3213, ИДЗ-19.2, Вариант 1

Дано:

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

- Найти уравнение прямой регрессии y на x ;
- Построить уравнение эмпирической линии регрессии и случайные точки выборки (X, Y)

$X \backslash Y$	2,3	3,8	5,3	6,8	7,3	8,8	10,3	11,8	m_x
210	—	4	3	5	—	—	—	—	12
340	—	6	7	8	—	—	—	—	21
470	—	—	10	12	11	—	—	—	33
600	—	—	—	—	5	4	3	—	12
730	—	—	—	—	—	6	8	—	14
860	—	—	—	—	—	—	3	5	8
m_y	—	10	20	25	16	10	14	5	100

Решение:

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:

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

$$\sum_{i=1}^6 \sum_{j=1}^8 m_{ij} x_i = \sum_{i=1}^6 m_{x_i} x_i = 49470$$

$$\sum_{i=1}^6 \sum_{j=1}^8 m_{ij} y_j = \sum_{j=1}^8 m_{y_j} y_j = 722$$

$$\sum_{i=1}^6 \left(x_i \sum_{j=1}^8 m_{ij} y_j \right) = \sum_{j=1}^8 \left(y_j \sum_{i=1}^6 m_{ij} x_i \right) = 391106$$

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

$$\bar{x} = \frac{\sum \sum m_{ij} x_i}{n} = \frac{\sum m_{x_i} x_i}{n} = \frac{49470}{100} = 494,7$$

$$\bar{y} = \frac{\sum m_{y_j} y_j}{n} = \frac{722}{100} = 7,22$$

Выборочные дисперсии находим по формулам:

$$s_x^2 = \frac{1}{n-1} \left(\sum m_{x_i} x_i^2 - \frac{1}{n} \left(\sum m_{x_i} x_i \right)^2 \right) = \frac{1}{99} \left(27943900 - \frac{1}{100} (49470)^2 \right) = 35061,5252$$

$$s_y^2 = \frac{1}{n-1} \left(\sum m_{y_j} y_j^2 - \frac{1}{n} \left(\sum m_{y_j} y_j \right)^2 \right) = \frac{1}{99} \left(5670,7 - \frac{1}{100} (722)^2 \right) = 4,624848$$

	j	1	2	3	4	5	6	7	8	9	10	11	12	13
i	$X \backslash Y$	2,3	3,8	5,3	6,8	7,3	8,8	10,3	11,8	m_x	$m_x x_i$	$\sum_{j=1}^k m_{ij} y_j$	$m_{x_i} x_i^2$	$x_i \sum_{j=1}^k m_{ij} y_j$
1	210	–	4	3	5	–	–	–	–	12	2520	65,1	529200	13671
2	340	–	6	7	8	–	–	–	–	21	7140	114,3	2427600	38862
3	470	–	–	10	12	11	–	–	–	33	15510	214,9	7289700	101003
4	600	–	–	–	–	5	4	3	–	12	7200	102,6	4320000	61560
5	730	–	–	–	–	–	6	8	–	14	10220	135,2	7460600	98696
6	860	–	–	–	–	–	–	3	5	8	6880	89,9	5916800	77314
7	m_y	0	10	20	25	16	10	14	5	100	49470	722	27943900	391106
8	$m_y y_j$	0	38	106	170	116,8	88	144,2	59	722	–	–	–	–
9	$\sum_{i=1}^m m_{x_i} x_i$	0	2880	7710	9410	8170	6780	10220	4300	49470	–	–	–	–
10	$m_{ij} y_j^2$	0	144,4	561,8	1156	852,64	774,4	1485,26	696,2	5670,7	–	–	–	–
11	$y_j \sum_{i=1}^m m_{x_i} x_i$	0	10944	40863	63988	59641	59664	105266	50740	391106	–	–	–	–

Корреляционный момент вычисляем по формуле:

$$S_{xy} = \frac{1}{n-1} \left(\sum \sum m_{ij} x_i y_j - \frac{1}{n} (\sum m_{xi} x_i) (\sum m_{yj} y_j) \right) = \frac{1}{99} \left(391106 - \frac{1}{100} (49470 * 722) \right) \approx 342,753$$

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

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

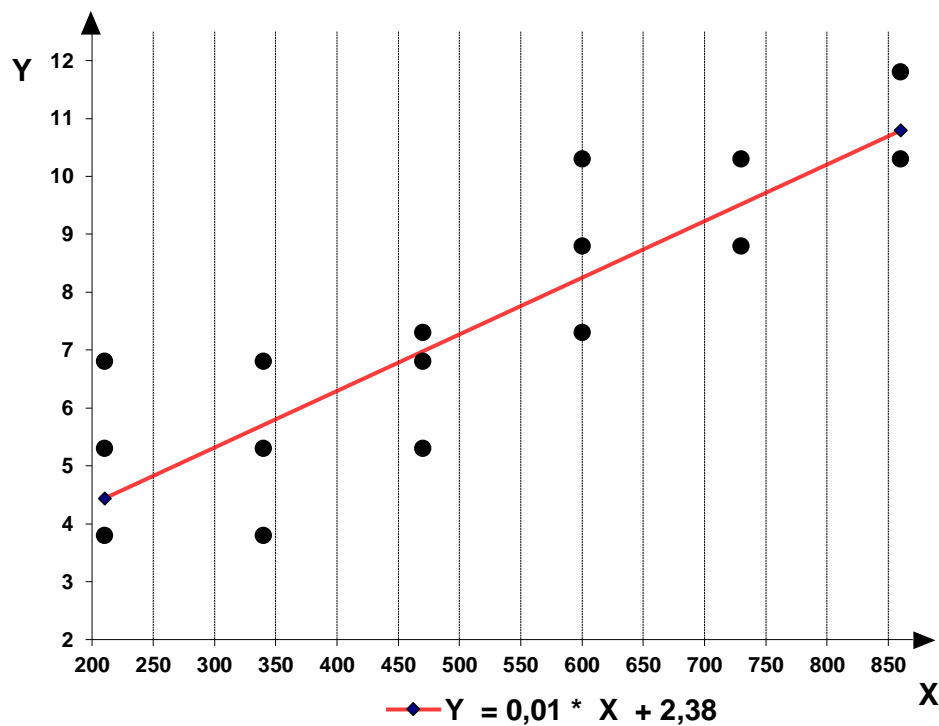
где $S_x = \sqrt{35061,5252} \approx 187,247$; $S_y = \sqrt{4,624848} \approx 2,151$;

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{342,753}{187,247 * 2,151} = \frac{342,753}{402,768297} \approx 0,850993$$

Составляем уравнение эмпирической линии регрессии у на х:

$$y = 7,22 + 0,850993 * \frac{2,151}{187,247} (x - 494,7)$$

$$y = 0,00987x + 2,38392$$



Теория Вероятностей. Домашнее задание №2

Бободжонов Комронджон, Р3213, ИДЗ-19.1, Вариант 1

Дано:

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

16.8	17.9	21.4	14.1	19.1	18.1	15.1	18.2	20.3	16.7
19.5	18.5	22.5	18.4	16.2	18.3	19.1	21.4	14.5	16.1
21.5	14.9	18.6	20.4	15.2	18.5	17.1	22.4	20.8	19.8
17.2	19.7	16.3	18.7	14.4	18.8	19.5	21.6	15.3	17.3
22.8	17.4	22.2	16.5	21.7	15.4	21.3	14.3	20.5	16.4
20.6	15.5	19.4	17.5	20.9	23.0	18.9	15.9	18.2	20.7
17.9	21.8	14.2	21.2	16.1	18.4	17.5	19.3	22.7	19.6
22.1	17.6	16.7	20.4	15.7	18.1	16.6	18.3	15.5	17.7
19.2	14.8	19.7	17.7	16.5	17.8	18.5	14.0	21.9	16.9
15.8	20.8	17.1	20.1	22.6	18.9	15.6	21.1	20.2	15.1

Решение:

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

14.0	14.1	14.2	14.3	14.4	14.5	14.8	14.9	15.1	15.1
15.2	15.3	15.4	15.5	15.5	15.6	15.7	15.8	15.9	16.1
16.1	16.2	16.3	16.4	16.5	16.5	16.6	16.7	16.7	16.8
16.9	17.1	17.1	17.2	17.3	17.4	17.5	17.5	17.6	17.7
17.7	17.8	17.9	17.9	18.1	18.1	18.2	18.2	18.3	18.3
18.4	18.4	18.5	18.5	18.5	18.6	18.7	18.8	18.9	18.9
19.1	19.1	19.2	19.3	19.4	19.5	19.5	19.6	19.7	19.7
19.8	20.1	20.2	20.3	20.4	20.4	20.5	20.6	20.7	20.8
20.8	20.9	21.1	21.2	21.3	21.4	21.4	21.5	21.6	21.7
21.8	21.9	22.1	22.2	22.4	22.5	22.6	22.7	22.8	23.0

б) Находим размах варьирования: $\omega = x_{max} - x_{min} = 23 - 14 = 9$

Выборку разобьём на 9 интервалов. Длина частичного интервала: $h = \frac{\omega}{l} = \frac{9}{9} = 1$

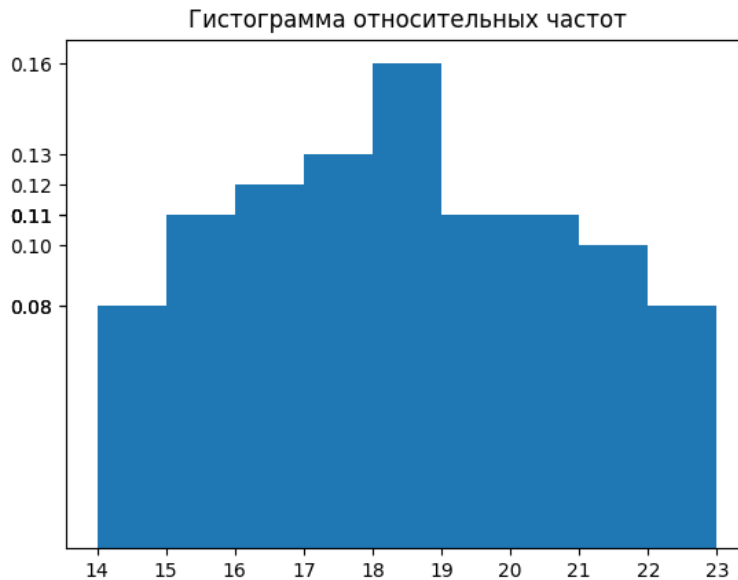
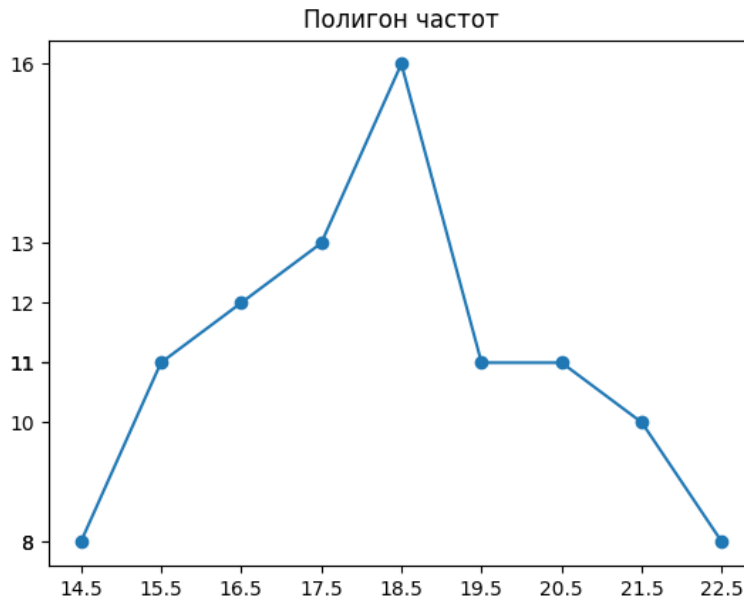
Номер частичного интервала l_i	Границы интервала $x_i - x_{i+1}$	Середина интервала $x'_i = \frac{x_i + x_{i+1}}{2}$	Частота интервала n_i	Относительная частота $W_i = \frac{n_i}{n}$	Плотность относительной частоты $\frac{W_i}{h}$
1	14 – 15	14.5	8	0.08	0.08
2	15 – 16	15.5	11	0.11	0.11
3	16 – 17	16.5	12	0.12	0.12
4	17 – 18	17.5	13	0.13	0.13
5	18 – 19	18.5	16	0.16	0.16
6	19 – 20	19.5	11	0.11	0.11
7	20 – 21	20.5	11	0.11	0.11
8	21 – 22	21.5	10	0.10	0.10
9	22 – 23	22.5	8	0.08	0.08
\sum_i	–	–	100	1	–

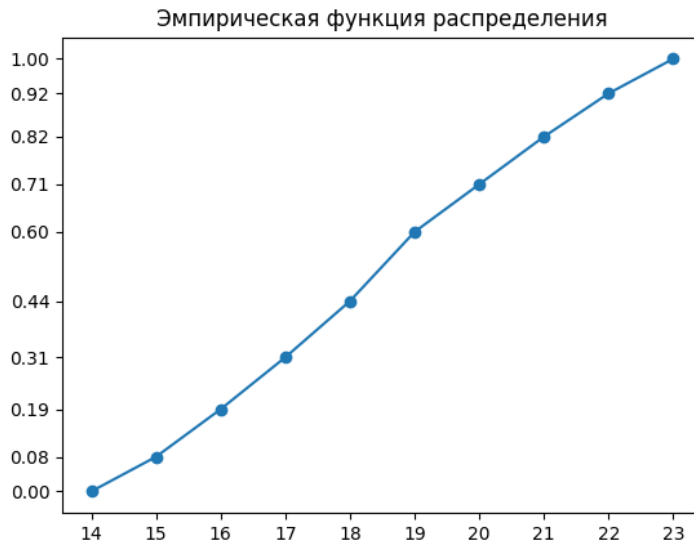
в) Строим полигон частот, гистограмму относительных частот и график эмпирической функции распределения.

Находим значения эмпирической функции распределения $F^*(x) = \frac{n_x}{n}$: $F^*(14) = 0$;

$F^*(15) = 0,08$; $F^*(16) = 0,19$; $F^*(17) = 0,31$; $F^*(18) = 0,44$; $F^*(19) = 0,60$;

$F^*(20) = 0,71$; $F^*(21) = 0,82$; $F^*(22) = 0,92$; $F^*(23) = 1$.





г) Находим выборочное среднее и выборочную дисперсию:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x'_i n_i = 18,43$$

$$D_B = \frac{1}{n} \sum_{i=1}^k (x'_i - \bar{x})^2 n_i = \frac{1}{n} \sum_{i=1}^k (x'_i)^2 n_i - \bar{x}^2 = 5,605$$

$$\sigma_B = \sqrt{D_B} = 2,367$$

Расчетная таблица:

m_i	Границы интервала $x_i; x_{i+1}$	Середина интервала x'_i	Частота интервала n_i	$n_i x'_i$	$(x'_i)^2$	$n_i (x'_i)^2$
1	14 – 15	14,5	8	116	210,25	1682,00
2	15 – 16	15,5	11	170,5	240,25	2642,75
3	16 – 17	16,5	12	198	272,25	3267,00
4	17 – 18	17,5	13	227,5	306,25	3981,25
5	18 – 19	18,5	16	296	342,25	5476,00
6	19 – 20	19,5	11	214,5	380,25	4182,75
7	20 – 21	20,5	11	225,5	420,25	4622,75
8	21 – 22	21,5	10	215	462,25	4622,50
9	22 – 23	22,5	8	180	506,25	4050,00
\sum_i	–	–	100	1843	–	34 527

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

$$\tilde{D}_B = \frac{n}{(n-1)} D_B = \frac{100}{99} * 5,605 = 5,6617$$

$$\tilde{\sigma}_B = \sqrt{\tilde{D}_B} = 2,379$$

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 и z_{i+1} , 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. Все $n_i > 5$, значит объединять интервалы не требуется.

i	Границы интервала $x_i; x_{i+1}$		$x_i - \bar{x}$	$x_{i+1} - \bar{x}$	Границы интервала $z_i; z_{i+1}$	
	x_i	x_{i+1}			$z_i = \frac{x_i - \bar{x}}{\sigma_B}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}}{\sigma_B}$
1	14	15	—	—3,43	—	—1,441782261
2	15	16	—3,43	—2,43	—1,441782261	—1,021437579
3	16	17	—2,43	—1,43	—1,021437579	—0,601092896
4	17	18	—1,43	—0,43	—0,601092896	—0,180748214
5	18	19	—0,43	0,57	—0,180748214	0,239596469
6	19	20	0,57	1,57	0,239596469	0,659941152
7	20	21	1,57	2,57	0,659941152	1,080285834
8	21	22	2,57	3,57	1,080285834	1,500630517
9	22	23	3,57	—	1,500630517	—

Находим теоретические вероятности P_i и теоретические частоты $n'_i = nP_i = 100P_i$. Составляем расчетную таблицу.

i	Границы интервала $z_i; z_{i+1}$		$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = 100P_i$
	z_i	z_{i+1}				
1	—	—1,441782261	—0,5	—0,425318	0,074682	7,4682
2	—1,441782261	—1,021437579	—0,425318	—0,346476	0,078842	7,8842
3	—1,021437579	—0,601092896	—0,346476	—0,226111	0,120365	12,0365
4	—0,601092896	—0,180748214	—0,226111	—0,0717174	0,1543936	15,43936
5	—0,180748214	0,239596469	—0,0717174	0,0946784	0,1663958	16,63958
6	0,239596469	0,659941152	0,0946784	0,245354	0,1506756	15,06756
7	0,659941152	1,080285834	0,245354	0,359993	0,114639	11,4639
8	1,080285834	1,500630517	0,359993	0,433274	0,073281	7,3281
9	1,500630517	—	0,433274	0,5	0,066726	6,6726
\sum_i	—	—	—	—	1	100

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

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

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$(n_i - n'_i)^2/n'_i$	n_i^2	n_i^2/n'_i
1	8	7,4682	0,5318	0,28281124	0,037868729	64	8,569668729
2	11	7,8842	3,1158	9,70820964	1,231349996	121	15,34715
3	12	12,0365	—0,0365	0,00133225	0,000110684	144	11,96361068
4	13	15,43936	—2,43936	5,95047721	0,385409577	169	10,94604958
5	16	16,63958	—0,63958	0,409062576	0,024583708	256	15,38500371
6	11	15,06756	—4,06756	16,54504435	1,098057307	121	8,030497307
7	11	11,4639	—0,4639	0,21520321	0,018772251	121	10,55487225

8	10	7,3281	2,6719	7,13904961	0,974201991	100	13,64610199
9	8	6,6726	1,3274	1,76199076	0,264063597	64	9,591463597
\sum_i	100	100	-	-	$\chi^2_{\text{набл}} = 4,034417841$	-	104,0344178

Контроль: $\frac{\sum n_i^2}{n} - n = \frac{\sum (n_i - n'_i)^2}{n} = 104,0344178 - 100 = 4,0344178$.

По таблице критических точек распределения χ^2 , уровню значимости $\alpha = 0,025$ и числу степеней свободы $k = l - 3 = 9 - 3 = 6$ находим: $\chi^2_{\text{кр}} = 14,4$

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

е) Если СВ X генеральной совокупности распределена нормально, то с надежностью $\gamma = 0,9$ можно утверждать, что математическое ожидание α СВ X покрывается доверительным интервалом

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

В нашем случае

$$\bar{x} = 18,43, \tilde{\sigma}_B = 2,379, n = 100, t_\gamma = 1,6604, \delta = 0,395$$

Доверительным интервалом для α будет (18,035; 18,825).

Доверительный интервал, покрывающий среднее квадратичное отклонение σ с заданной надежностью γ , $(\tilde{\sigma}_B(1 - q); \tilde{\sigma}_B(1 + q))$.

При $\gamma = 0,9$ и $n = 100$ имеем: $q = 0,102$. Доверительным интервалом для σ будет (2,136; 2,622).