

Из 300 однотипных предприятий региона сделали 10% механическую выборку. Данные о выпуске продукции и валовой прибыли этих предприятий собраны в следующую таблицу:

№ предприятия	Выпуск продукции (млн. руб.)	Валовая прибыль (млн. руб.)	№ предприятия	Выпуск продукции (млн. руб.)	Валовая Прибыль (млн. руб.)
1	65	15.7	16	52	14,6
2	78	18	17	62	14,8
3	41	12.1	18	69	16,1
4	54	13.8	19	85	16,7
5	66	15.5	20	70	15,8
6	80	17.9	21	71	16,4
7	45	12.8	22	64	15
8	57	14.2	23	72	16,5
9	67	15.9	24	88	18,5
10	81	17.6	25	73	16,4
11	92	18.2	26	74	16
12	48	13	27	96	19,1
13	59	16.5	28	75	16,3
14	68	16.2	29	101	19,6
15	83	16.7	30	76	17,2

По данным таблицы выполните следующие задания.

Задание 1. Сгруппируйте предприятия по выпуску продукции, образовав 6 групп с равными интервалами.

- 1) Постройте гистограмму распределения частот. Что можно сказать о симметричности (асимметричности) полученного распределения?
- 2) Найдите средний выпуск продукции, моду, медиану выпуска продукции на предприятиях региона. Подтвердился ли вывод о симметричности (асимметричности) по вычисленным средним показателям?
- 3) Рассчитайте дисперсию выпуска продукции и сделайте вывод об однородности совокупности предприятий по выпуску продукции.

Решение:

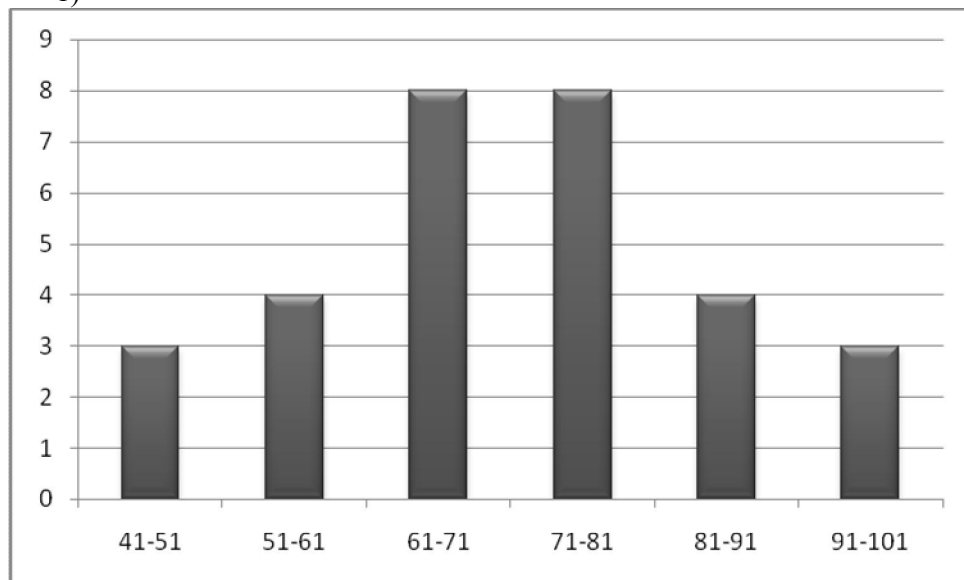
$$k = 6$$

Определим величину интервала :

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{101 - 41}{6} = 10$$

$x_i \div x_{i+1}$	41-51	51-61	61-71	71-81	81-91	91-101
середина интервала x_i	46	56	66	76	86	96
n_i (частота)	3	4	8	8	4	3

1)



Данное распределение симметрично.

2)

$x_i \div x_{i+1}$	41-51	51-61	61-71	71-81	81-91	91-101	сумма
середина интервала x_i	46	56	66	76	86	96	
n_i (частота)	3	4	8	8	4	3	30
$x_i n_i$	138	224	528	608	344	288	2130

Средний выпуск продукции:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{2130}{30} = 71$$

Мода:

$$Mo = x_0 + \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})} \cdot h = 61 + \frac{16 - 4}{(16 - 4) + (16 - 4)} \cdot 20 = 71$$

Медиана:

$$Me = x_0 + \frac{\sum n_i - n'_{Mo-1}}{n_{Mo}} \cdot h = 61 + \frac{15 - 7}{16} \cdot 20 = 71$$

Все три характеристики равны между собой, следовательно, вывод о симметричности данного распределения подтверждается.

3)

$x_i \div x_{i+1}$	41-51	51-61	61-71	71-81	81-91	91-101	сумма	среднее
середина интервала x_i	46	56	66	76	86	96		
n_i (частота)	3	4	8	8	4	3	30	
$x_i n_i$	138	224	528	608	344	288	2130	71
$x_i^2 n_i$	6348	12544	34848	46208	29584	27648	157180	5239,333

Дисперсия:

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 = 5239,333 - 5041 = 198,3333$$

Среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{198,3333} = 14,08309$$

Коэффициент вариации:

$$v = \frac{\sigma_x}{\bar{x}} \cdot 100\% = \frac{14,08309}{71} \cdot 100\% = 19,84\%$$

Таким образом, по полученным расчетам коэффициент вариации менее 33%, а следовательно, данную выборку можно считать однородной.

Задание 2. Определите долю предприятий региона с большим объемом выпуска продукции (не менее 80 млн. руб.). Каково примерное количество таких предприятий в рассматриваемом регионе? С вероятностью 0,997 определите, в каких пределах будет находиться доля крупных предприятий и число таких предприятий?

Решение:

Таких предприятий в 10%-ной выборке получилось 8, следовательно, если считать выборку репрезентативной, то в генеральной совокупности примерно должно быть 80 таких предприятий.

Доля предприятий региона с большим выпуском:

$$w = \frac{8}{30} = 0,2667$$

$$p = 0,997 \Rightarrow t = 3$$

$$\Delta w = t \mu_w = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} = 3 \cdot \sqrt{\frac{0,2667 \cdot 0,7333}{30} (1 - 0,1)} = 0,2298$$

$$w_{gen} = w \pm \Delta w$$

$$w - \Delta w < w_{\text{ген}} < w + \Delta w$$

$$0.2667 - 0.2298 < w_{\text{ген}} < 0.2667 + 0.2298$$

$$0.0369 < w_{\text{ген}} < 0.4965$$

Задание 3. Сгруппируйте предприятия по валовой прибыли, образовав 5 групп с равными интервалами.

- 1) Постройте кривую распределения и определите ее вид.
- 2) Найдите среднюю величину, моду и медиану валовой прибыли. Проверьте правильность определения вида кривой: при правосторонней асимметрии должно выполняться равенство $I_i < M_e < \bar{X}$, а при левосторонней $\bar{X} < M_e < M_o$.
- 3) Рассчитайте дисперсию валовой прибыли продукции и сделайте вывод об однородности совокупности предприятий по валовой прибыли.

Решение:

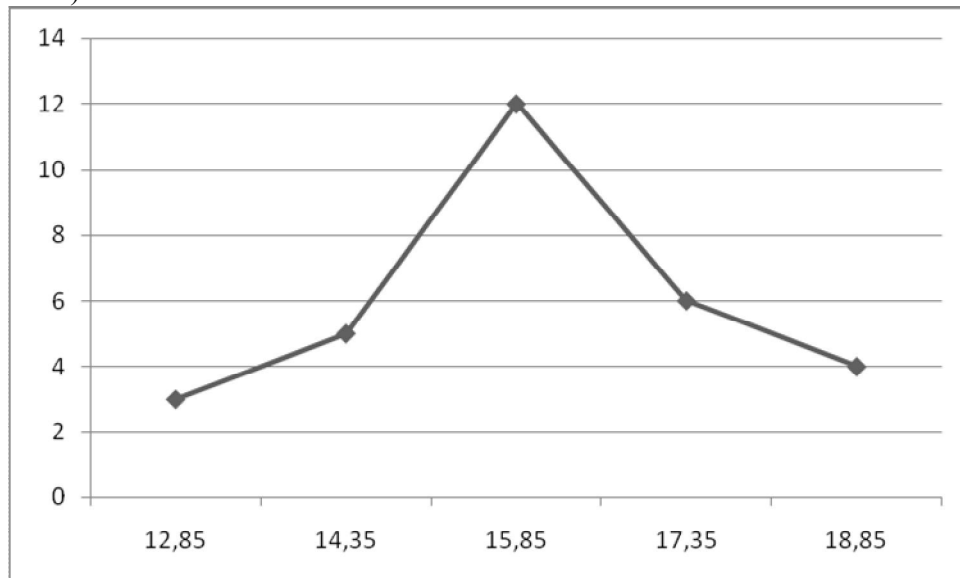
$$k = 5$$

Определим величину интервала:

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{19.6 - 12.1}{5} = 1.5$$

$x_i \div x_{i+1}$	12,1-13,6	13,6-15,1	15,1-16,6	16,6-18,1	18,1-19,6	сумма
x_i	12,85	14,35	15,85	17,35	18,85	
n_i	3	5	12	6	4	30
$x_i n_i$	38,55	71,75	190,2	104,1	75,4	480
$x_i^2 n_i$	495,3675	1029,613	3014,67	1806,135	1421,29	7767,075

1)



Кривая является асимметричной.

- 2) Средняя величина валовой прибыли:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{480}{30} = 16$$

Мода:

$$M_o = x_0 + \frac{n_{M_o} - n_{M_o-1}}{(n_{M_o} - n_{M_o-1}) + (n_{M_o} - n_{M_o+1})} \cdot h = 15,1 + \frac{12 - 5}{(12 - 5) + (12 - 6)} \cdot 1,5 = 15,91$$

Медиана:

$$M_e = x_0 + \frac{\frac{\sum n_i}{2} - n'_{M_o-1}}{n_{M_o}} \cdot h = 15,1 + \frac{15 - 8}{12} \cdot 1,5 = 15,98$$

Так как $M_o < M_e < \bar{X}$, следовательно, имеет место быть правосторонняя асимметрия.

- 3) Дисперсия:

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 = 258,9025 - 256 = 2,9025$$

Среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{2,9025} = 1,703673$$

Коэффициент вариации:

$$v = \frac{\sigma_x}{\bar{x}} \cdot 100\% = \frac{1,703673}{16} \cdot 100\% = 10,65\%$$

Таким образом, по полученным расчетам коэффициент вариации менее 33%, а следовательно, данную выборку можно считать однородной.

Задание 4. С вероятностью 0,866 определите ошибку выборки для суммы валовой прибыли на одно предприятие и границы, в которых будет находиться средняя сумма прибыли одного предприятия в генеральной совокупности.

Решение:

\bar{x} -генеральная средняя

$$\bar{x} = \tilde{x} \pm \Delta_x$$

$$\Delta_x = t\mu_x$$

$$p = 0.866 \Rightarrow t = 1.5$$

$$\mu_x = \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{2.9025}{30} (1 - 0.1)} = 0.295$$

$$\Delta_x = t\mu_x = 1.5 \cdot 0.295 = 0.443$$

$$16 - 0.443 < \bar{x} < 16 + 0.443$$

$$15.557 < \bar{x} < 16.443$$

Задание 5. Предположим, что исследование несет описательный характер, и нам не важна высокая точность результатов. Какое значение вероятности принимается для того, чтобы выборка была репрезентативной? Рассчитайте, сколько предприятий нужно было бы отобрать из генеральной совокупности в этом случае.

Решение:

$$p = 0.9 \Rightarrow t = 1,6499$$

$$\Delta_x = t\mu_x = \sqrt{\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right)}$$

$$\frac{\sigma_x^2}{n} \left(1 - \frac{n}{N}\right) = \mu_x^2 t^2$$

$$\frac{1}{n} - \frac{1}{N} = \frac{\mu_x^2 t^2}{\sigma_x^2}$$

$$\frac{1}{n} = \frac{1}{N} + \frac{\mu_x^2 t^2}{\sigma_x^2}$$

$$n = \frac{1}{\frac{1}{N} + \frac{\mu_x^2 t^2}{\sigma_x^2}} = \frac{1}{\frac{1}{300} + \frac{0.295^2 \cdot 1.6499^2}{2.9025}} = \frac{1}{0.0033 + 0.0816} = \frac{1}{0.0849} \approx 12$$

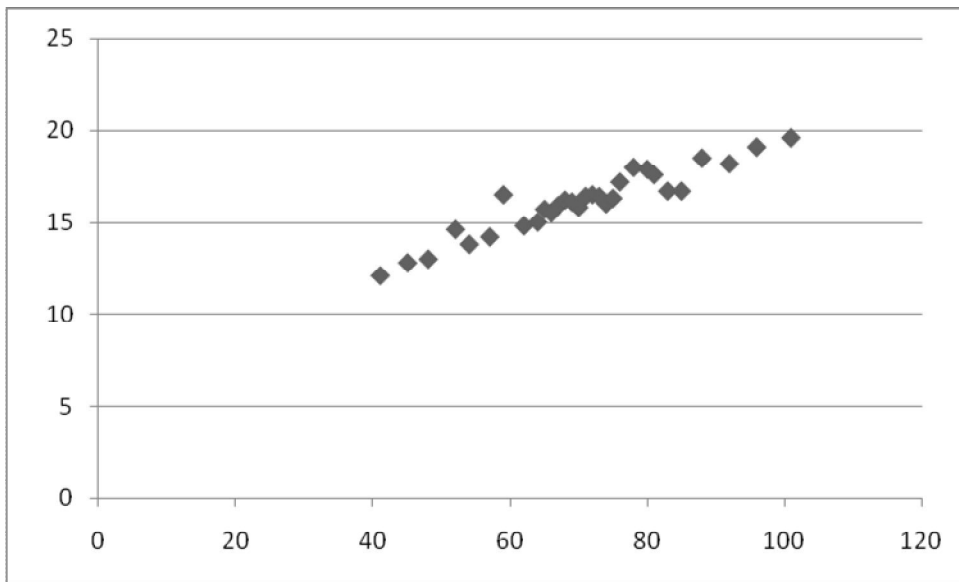
Задание 6. Определите факторную и результативную переменные в данном исследовании.

- 1) Постройте точечную диаграмму по совокупности предприятий.
- 2) По построенной диаграмме сделайте вывод о характере зависимости прибыли от выпуска продукции.
- 3) Какую модель регрессии целесообразно выбрать для исследования данной зависимости?

Решение:

Факторная переменная – выпуск продукции, результативная – валовая прибыль.

1)



- 2) Данная зависимость близка к линейной.
- 3) Целесообразно выбрать модель линейной регрессии.

Задание 7. Используя группировку по валовой прибыли, сделайте типологическую выборку 10 предприятий (пропорционально количеству предприятий в группе). Составьте таблицу зависимости прибыли от выпуска продукции на выбранных 10 предприятиях региона.

Решение:

$x_i \div x_{i+1}$	12,1-13,6	13,6-15,1	15,1-16,6	16,6-18,1	18,1-19,6
x_i'	12,85	14,35	15,85	17,35	18,85
n_i	3	5	12	6	4
n_i'	$3 \cdot 10 / 30 = 1$	$5 \cdot 10 / 30 = 1,67$	$12 \cdot 10 / 30 = 4$	$6 \cdot 10 / 30 = 2$	$4 \cdot 10 / 30 = 1,33$

То есть для сохранения пропорциональности из первого интервала надо взять 1 предприятие, из второго – 2 предприятия, из третьего – 4 предприятия, из четвертого – 2 предприятия, из пятого – 1 предприятий.

Получаем следующую выборку:

№ пред.	Выпуск продукции (млн.руб)	Валовая прибыль (млн.руб)
7	45	12,8
8	57	14,2
16	52	14,6
20	70	15,8
9	67	15,9
26	74	16
18	69	16,1
30	76	17,2
10	81	17,6
24	88	18,5

Задание 8. На основе выборки, сделанной в задании 7:

- 1) Рассчитайте параметры парной линейной регрессии,
- 2) Постройте график регрессии на том же поле, что и точечная диаграмма (задание б), и сделайте визуальный вывод об адекватности построенной модели,
- 3) Сделайте прогноз возможной валовой прибыли на предприятии с выпуском продукции 33 млн. руб.

Решение:

- 1) Для расчета параметров используем метод наименьших квадратов:

$$\begin{cases} na + b \sum x = \sum y \\ a \sum x + b \sum x^2 = \sum yx \end{cases}$$

Таблица расчетов:

	x	y	x ²	yx
	45	12,8	2025	576
	57	14,2	3249	809,4
	52	14,6	2704	759,2
	70	15,8	4900	1106
	67	15,9	4489	1065,3
	74	16	5476	1184
	69	16,1	4761	1110,9
	76	17,2	5776	1307,2
	81	17,6	6561	1425,6
	88	18,5	7744	1628
сумма	679	158,7	47685	10971,6

Получаем систему:

$$\begin{cases} 10a + 679b = 158,7 \\ 679a + 47685b = 10971,6 \end{cases}$$

Решаем методом Крамера:

$$\Delta = \begin{vmatrix} 10 & 679 \\ 679 & 47685 \end{vmatrix} = 15809 \quad \Delta_a = \begin{vmatrix} 158,7 & 679 \\ 10971,6 & 47685 \end{vmatrix} = 117893,1 \quad \Delta_b = \begin{vmatrix} 10 & 158,7 \\ 679 & 10971,6 \end{vmatrix} = 1958,7$$

$$a = \frac{\Delta_a}{\Delta} = \frac{117893,1}{15809} = 7,457341 \quad b = \frac{\Delta_b}{\Delta} = \frac{1958,7}{15809} = 0,123898$$

Уравнение линейной регрессии будет иметь вид:

$$y = 7,457 + 0.124x$$

2)

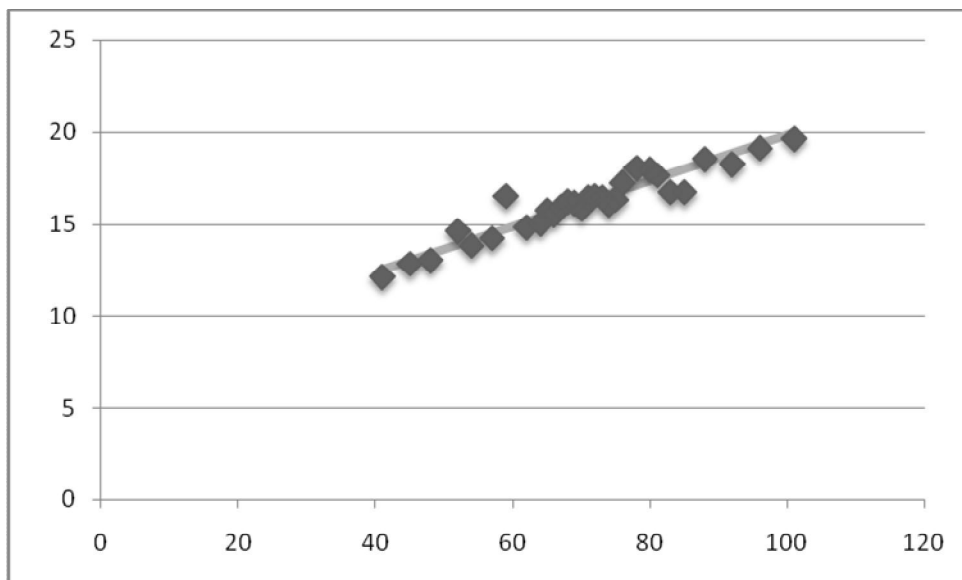


График уравнения линейной регрессии достаточно верно отражает зависимость выпуска продукции от валовой прибыли, значит модель подобрана правильно, а значит, ее можно считать адекватной.

3) $y(33) = 7,457 + 0.124 \cdot 33 = 11,549$

То есть если выпуск продукции будет составлять 33 млн. руб., то прибыль в среднем составит 11,549 млн. руб.