

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения (табл.2)

[26.7 31.8 34.3 34.9 36.9 37.5 38. 38.5 39.5 40.3]

[43.2 43.5 43.5 44.9 44.9 45.8 46.5 46.5 46.5 46.5]

[47.7 48.5 50. 51. 52.1 53.6 54.1 55.3 55.8 56.2]

[57.5 57.5 58.5 58.5 58.8 58.9 59. 59.5 60.3 60.3]

[60.5 60.5 61.5 62.5 62.6 62.8 63. 63.9 63.9 64.3]

[64.3 65.2 65.3 65.3 65.5 65.9 66. 66.2 66.4 66.5]

[66.5 66.7 67.3 67.8 68.2 68.5 69.5 70.5 70.5 70.9]

[71.5 71.5 73.1 73.7 74. 74. 74.7 74.8 75.1 75.8]

[79.2 79.3 79.5 79.5 81.3 81.3 83.5 85.1 85.5 85.6]

[85.7 86.9 87.7 91.5 92.8 93. 93.2 94.8 95. 99.9]

Шаг 1. Найти размах вариации

$$R = x_{max} - x_{min}$$

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 26,7$; $x_{max} = 99,6$

$$R = x_{max} - x_{min} = 99,9 - 26,7 = 73,2$$

Шаг 2. Найти оптимальное количество интервалов

Скобка $[]$ означает целую часть (округление вниз до целого числа).

$$k = 1 + [3,222 * \lg(N)]$$

$$k = 1 + [3,222 * \lg(100)] = 1 + [6,444] = 1 + 6 = 7$$

Шаг 3. Найти шаг интервального ряда

Скобка $\lceil \rceil$ означает округление вверх, в данном случае не обязательно до целого числа

$$h = \lceil \frac{R}{k} \rceil = \lceil \frac{73.2}{7} \rceil = \lceil 10.4571 \rceil = 11$$

Шаг 4. Найти узлы ряда:

$$a_0 = x_{min} = 26.7$$

$$a_i = a_0 + i * h = 26.7 + i * 11, i = 1, \dots, 7$$

Заметим, что поскольку шаг h находится с округлением вверх, последний узел $a_k \geq x_{max}$

$(a_{i-1}; a_i)$: [26.7; 37.7); [37.7; 48.7); [48.7; 59.7); [59.7; 70.7); [70.7; 81.7); [81.7; 92.7); [92.7; 103.7)

- построим гистограмму относительных частот;

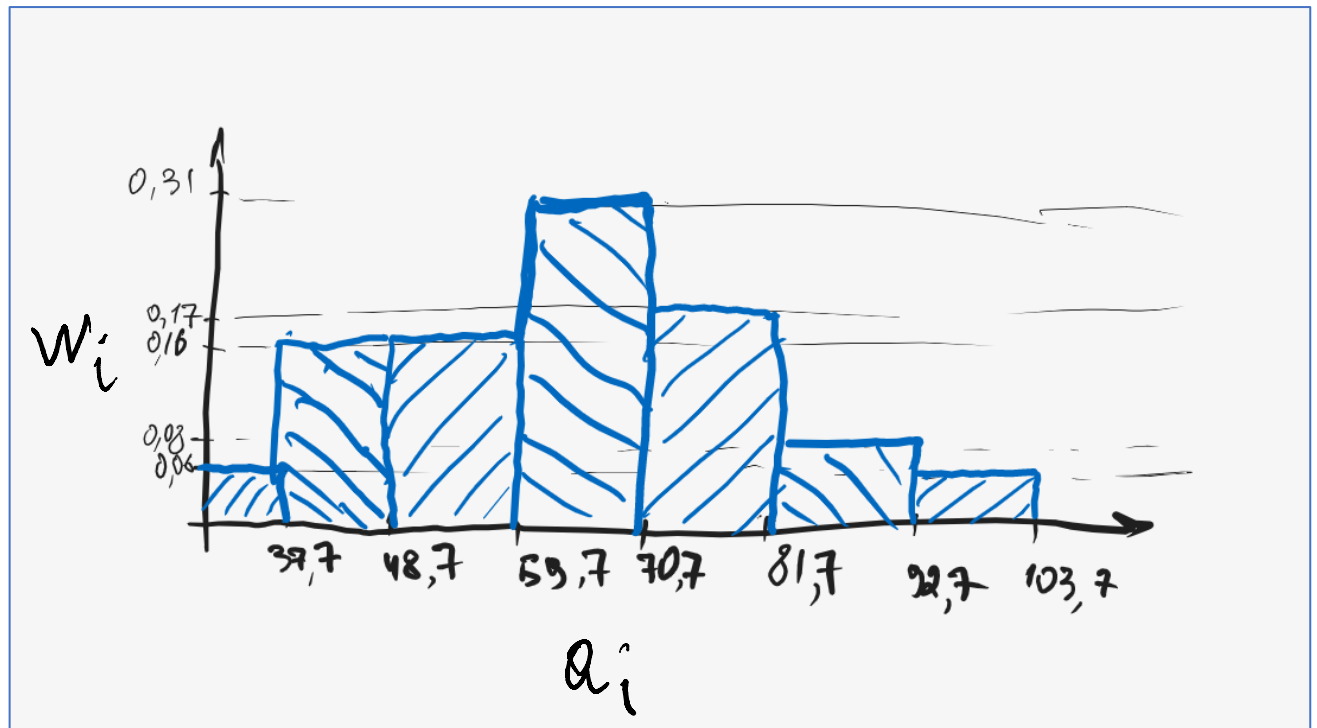
Найти частоты f_i – число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

$f_i = n_i$, n_i - количество точек на интервале $[a_{i-1}; a_i)$

Относительная частота интервала $[a_{i-1}; a_i)$ - это отношение частоты f_i к общему количеству исходов:

$$w_i = \frac{f_i}{100}, i = 1, \dots, 7$$

$[a_{i-1}; a_i)$	[26.7, 37.7)	[37.7, 48.7)	[48.7, 59.7)	[59.7, 70.7)	[70.7, 81.7)	[81.7, 92.7)	[92.7, 103.7)
n_i	6	16	16	31	17	8	6
n	100	100	100	100	100	100	100
w_i	0.06	0.16	0.16	0.31	0.17	0.08	0.06



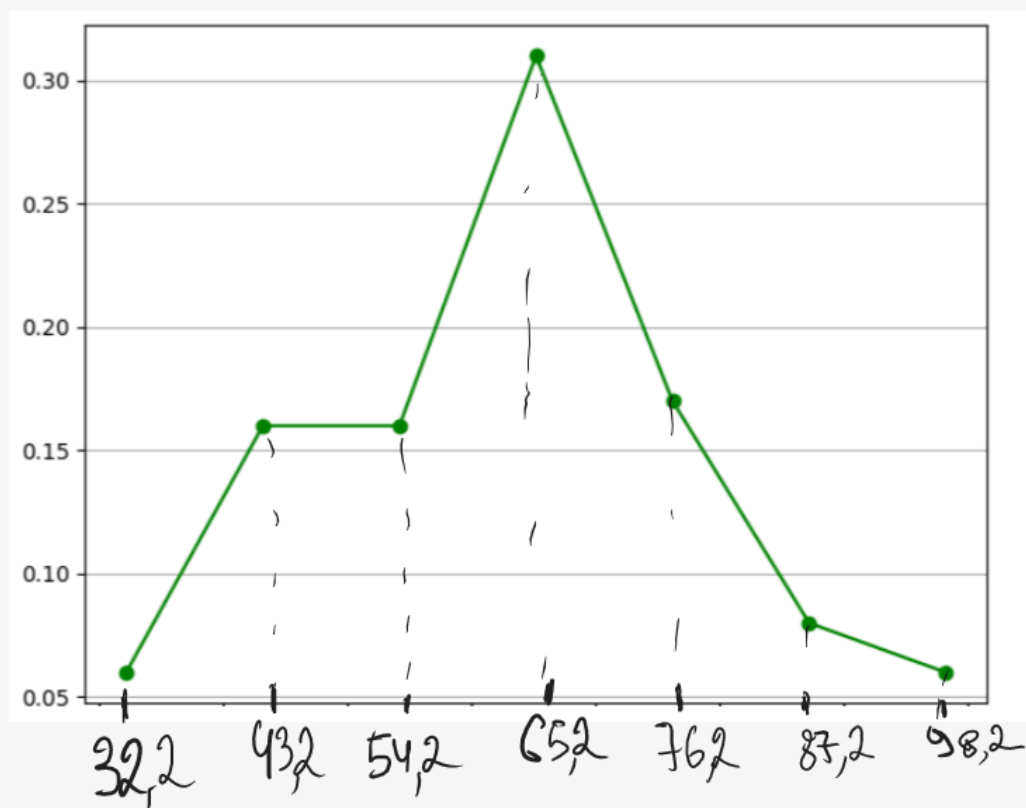
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	32.20	43.20	54.20	65.20	76.20	87.20	98.20
n_i	6.00	16.00	16.00	31.00	17.00	8.00	6.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.06	0.16	0.16	0.31	0.17	0.08	0.06

- Построим полигон относительных частот и найдем эмпирическую функцию распределения, построим ее график:

Полигон относительных частот интервального ряда – это ломаная, соединяющая точки (x_i, w_i) , где x_i - середины интервалов:

$$x_i = \frac{a_{i-1} + a_i}{2}, i = 1, \dots, 7$$



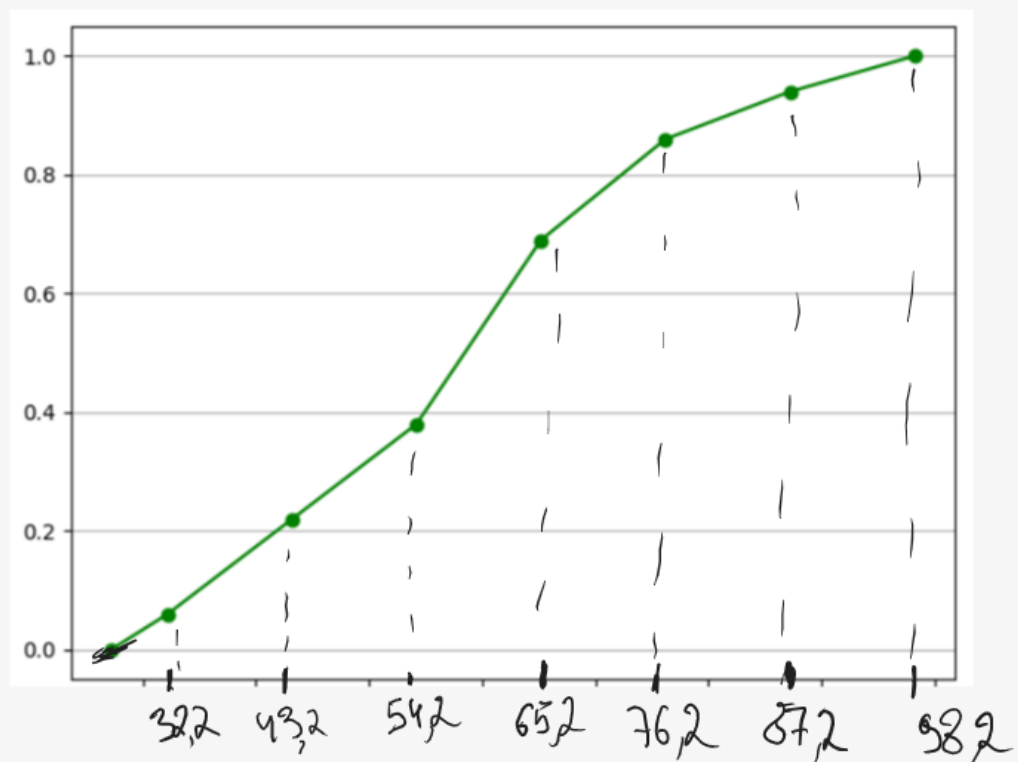
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [6, 16, 16, 31, 17, 8, 6]$$

$$x_i = [32.2, 43.2, 54.2, 65.2, 76.2, 87.2, 98.2]$$

$$F(x) = \begin{cases} 0.0, & x \leq 32.2, \\ 0.06, & 32.2 \leq x \leq 43.2, \\ 0.22, & 43.2 \leq x \leq 54.2, \\ 0.38, & 54.2 \leq x \leq 65.2, \\ 0.69, & 65.2 \leq x \leq 76.2, \\ 0.86, & 76.2 \leq x \leq 87.2, \\ 0.94, & 87.2 \leq x \leq 98.2, \\ 1.0, & x > 98.2; \end{cases}$$



- вычислим все точечные статистические оценки числовых характеристик

признака: среднее \bar{X} ; выборочную дисперсию и исправленную

выборочную дисперсию; выборочное с.к.о. и исправленное выборочное с.к.о. s ;

$$\bar{X} = \sum_{i=1}^7 (w_i * x_i) = 0.06 * 32.2 + 0.16 * 43.2 + 0.16 * 54.2 + 0.31 * 65.2 + 0.17 * 76.2 + 0.08 * 87.2 + 0.06 * 98.2 = 1.932 + 6.912 + 8.672 + 20.212 + 12.954 + 6.976 + 5.892 = 63.55$$

Выборочная средняя:

Выборочная средняя:

$$X_{cp} = \sum_{i=1}^7 (x_i * w_i) = 63.550000000000004$$

Выборочная дисперсия:

$$\begin{aligned} D &= \sum_{i=1}^7 (x_i - X_{cp})^2 * w_i = \\ &= (32.2 - 63.55)^2 * 0.06 + (43.2 - 63.55)^2 * 0.16 + (54.2 - 63.55)^2 * 0.16 + (65.2 - 63.55)^2 * 0.31 + (76.2 - 63.55)^2 * 0.17 \\ &\quad + (87.2 - 63.55)^2 * 0.08 + (98.2 - 63.55)^2 * 0.06 = 284.0475 \end{aligned}$$

Исправленная выборочная дисперсия:

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 284.0475 = 286.917$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{284.0475} \approx 16.8537$$

исправленное выборочное с.к.о s:

$$s = \sqrt{S^2} = \sqrt{286.9167} = 16.9386$$

- считая первый столбец таблицы выборкой значений признака X, а второй - выборкой значений Y, оценить тесноту линейной корреляционной зависимости между признаками и составить выборочное уравнение прямой регрессии Y на X

X = [66.7 70.5 57.5 58.5 74.7 75.8 99.9 58.5 93. 74.8]

Y = [26.7 37.5 61.5 38. 62.5 60.5 59. 71.5 65.5 65.2]

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
66.70000	26.70000	1780.89000	4448.89000	712.89000
70.50000	37.50000	2643.75000	4970.25000	1406.25000
57.50000	61.50000	3536.25000	3306.25000	3782.25000
58.50000	38.00000	2223.00000	3422.25000	1444.00000
74.70000	62.50000	4668.75000	5580.09000	3906.25000
75.80000	60.50000	4585.90000	5745.64000	3660.25000
99.90000	59.00000	5894.10000	9980.01000	3481.00000
58.50000	71.50000	4182.75000	3422.25000	5112.25000
93.00000	65.50000	6091.50000	8649.00000	4290.25000
74.80000	65.20000	4876.96000	5595.04000	4251.04000
Сумма	729.90000	547.90000	40483.85000	55119.67000
			32046.43000	

1) Оценить тесноту линейной корреляционной зависимости между признаками

Коэффициент корреляции Пирсона вычисляется по формуле:

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma(x) \cdot \sigma(y)},$$

где x_i – значения, принимаемые в выборке X , y_i – значения, принимаемые в выборке Y ;
 \bar{x} – среднее значение по X , \bar{y} – среднее значение по Y .

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma(x) \cdot \sigma(y)} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sqrt{x^2 - (\bar{x})^2} \cdot \sqrt{y^2 - (\bar{y})^2}} =$$

$$=$$

$$\frac{\frac{40483.85}{10} - \frac{729.9}{10} * \frac{547.9}{10}}{\sqrt{\left(\frac{55119.67}{10} - \left(\frac{729.9}{10}\right)^2\right)} * \sqrt{\left(\frac{32046.43}{10} - \left(\frac{547.9}{10}\right)^2\right)}} = 0.2547$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) [линейное уравнение регрессии](#) Y на X :

$$y_x - \bar{y} = r_{xy} \cdot \frac{\sigma_{by}}{\sigma_{bx}} (x - \bar{x}) \Rightarrow y_x = r_{xy} \cdot \frac{\sigma_{by}}{\sigma_{bx}} \cdot x + (\bar{y} - \bar{x} \cdot r_{xy} \cdot \frac{\sigma_{by}}{\sigma_{bx}})$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 72,99$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 54,79$$

$$\sigma_{\text{ex}}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = 184,42 \Rightarrow \sigma_{\text{ex}} \approx 13,5803$$

$$\sigma_{\text{ey}}^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 = 202,6989 \Rightarrow \sigma_{\text{ey}} \approx 14,2372$$

$$\bar{\mu}_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} = -395863.825$$

$$y_x = 0.26711 * x + 35.2933$$

$$r_{xy} = 0.25478$$