Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[14.1 19.4 21.9 22.5 25.1 25.6 26.1 26.5 27.1 27.9]

[30.8 31.1 31.1 32.5 32.5 33.4 34.1 34.1 34.1 34.2]

[35.3 36.1 37.6 38.6 39.7 41.2 41.7 42.9 43.4 43.8]

[45.1 45.1 46.1 46.1 46.4 46.5 46.6 47.1 47.9 47.9]

[48.1 48.1 49.1 50.1 50.2 50.4 50.6 51.5 51.7 51.9]

[51.9 52.8 52.9 52.9 53.1 53.5 53.6 53.8 54. 54.1]

[54.1 54.3 54.9 55.4 55.8 56.1 57.1 58.1 58.1 58.5]

[59.1 59.1 60.6 60.7 61.6 61.6 62.3 62.4 62.7 63.1]

[63.4 66.8 66.9 67.1 67.1 68.9 68.9 71.1 72.7 73.1]

[73.2 73.3 74.5 75.3 79.1 80.8 82.4 82.6 88.9 94.1]

Шаг 1. Найти размах вариации

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

Определим максимальное и минимальное значение имеющихся значений: $\mathrm{x}_{\min} = 14.1;\ \mathrm{x}_{max} = 94.1$

$$R = \mathrm{x}_{max} - \mathrm{x}_{min} = 94.1 - 14.1 = 80.0$$

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 ⌈ ⌉ означает округление вверх, в данном случае не обязательно до целого числа

$$h = \left\lceil \frac{R}{k} \right\rceil = \left\lceil \frac{80.0}{7} \right\rceil = \lceil 11.4285 \rceil = 12$$

Шаг 4. Найти узлы ряда:

$$a_0 = x_{min} = 14.1$$

$$a_i = a_0 + i * h = 14.1 + i * 12, i = 1, \dots, 7$$

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

$(a_{i-1};\ a_i)$: [14.1; 26.1); [26.1; 38.1); [38.1; 50.1); [50.1; 62.1); [62.1; 74.1); [74.1; 86.1); [86.1; 98.1)

- построим гистограмму относительных частот;

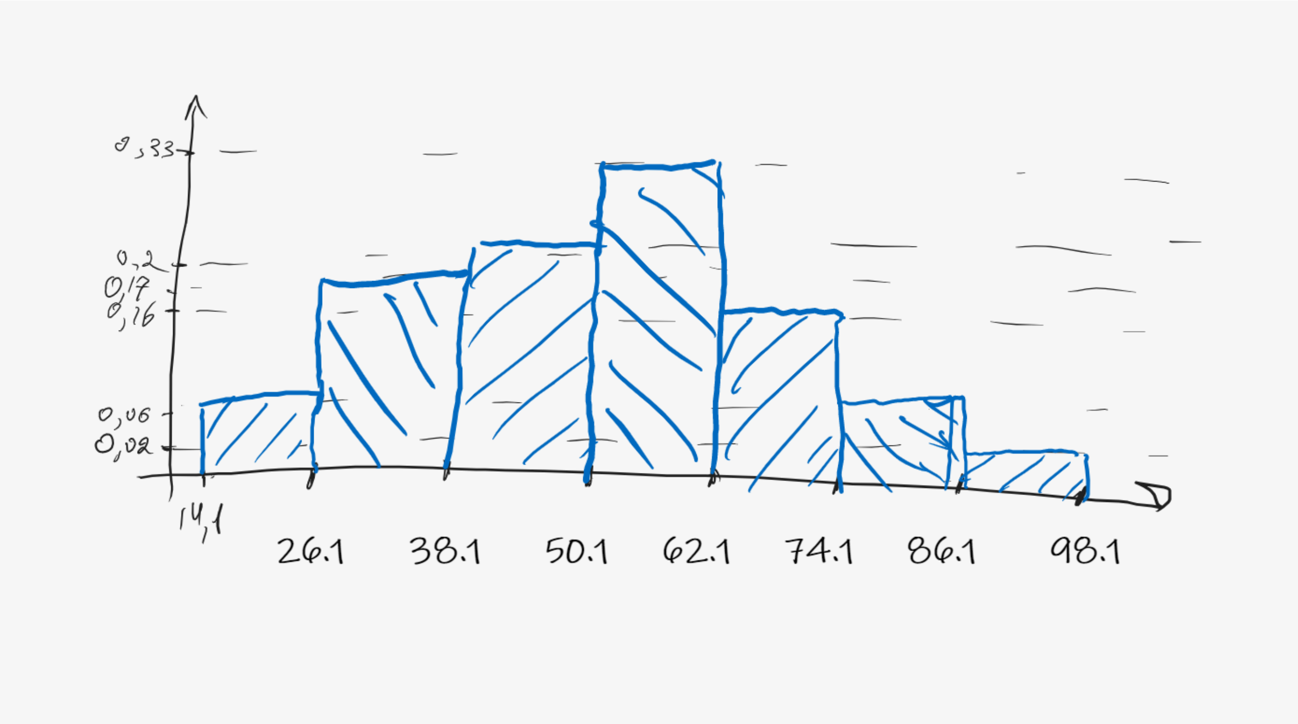
Найти частоты f_i – число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

$f_i = n_i, n_i$ - 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:

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

$[a_{i-1};\ a_i)$	[14.1, 26.1)	[26.1, 38.1)	[38.1, 50.1)	[50.1, 62.1)	[62.1, 74.1)	[74.1, 86.1)	[86.1, 98.1)
n_i	6	17	20	33	16	6	2
n	100	100	100	100	100	100	100
w_i	0.06	0.17	0.2	0.33	0.16	0.06	0.02



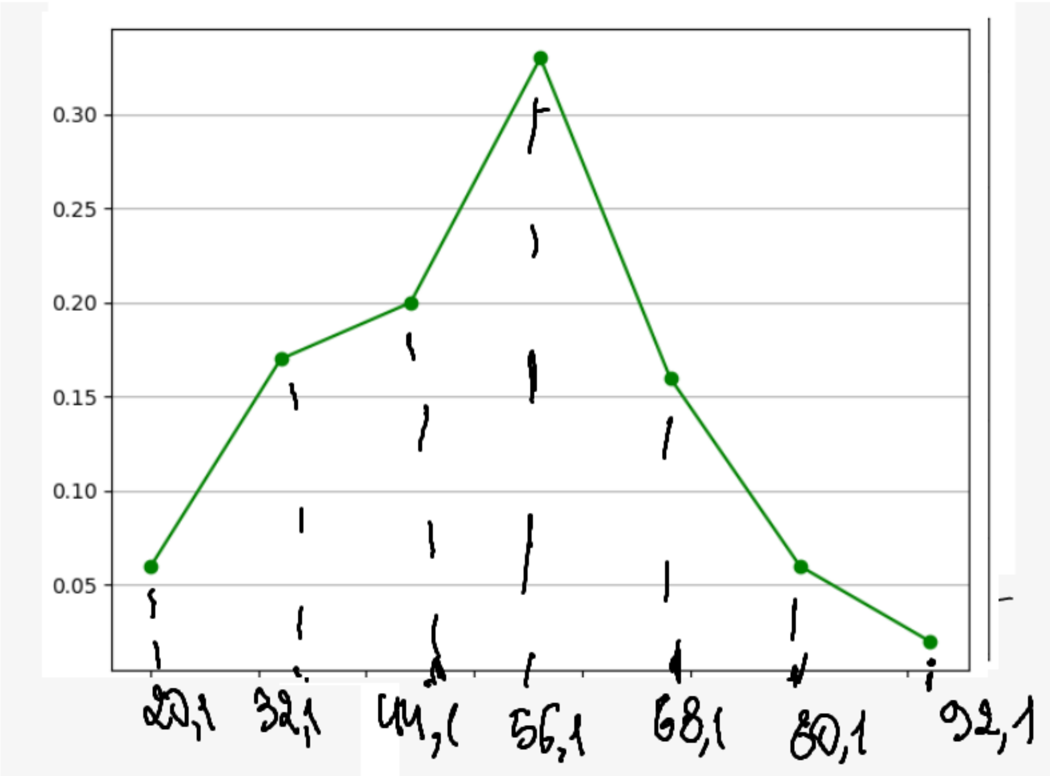
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	20.10	32.10	44.10	56.10	68.10	80.10	92.10
n_i	6.00	17.00	20.00	33.00	16.00	6.00	2.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.06	0.17	0.20	0.33	0.16	0.06	0.02

- 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:

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



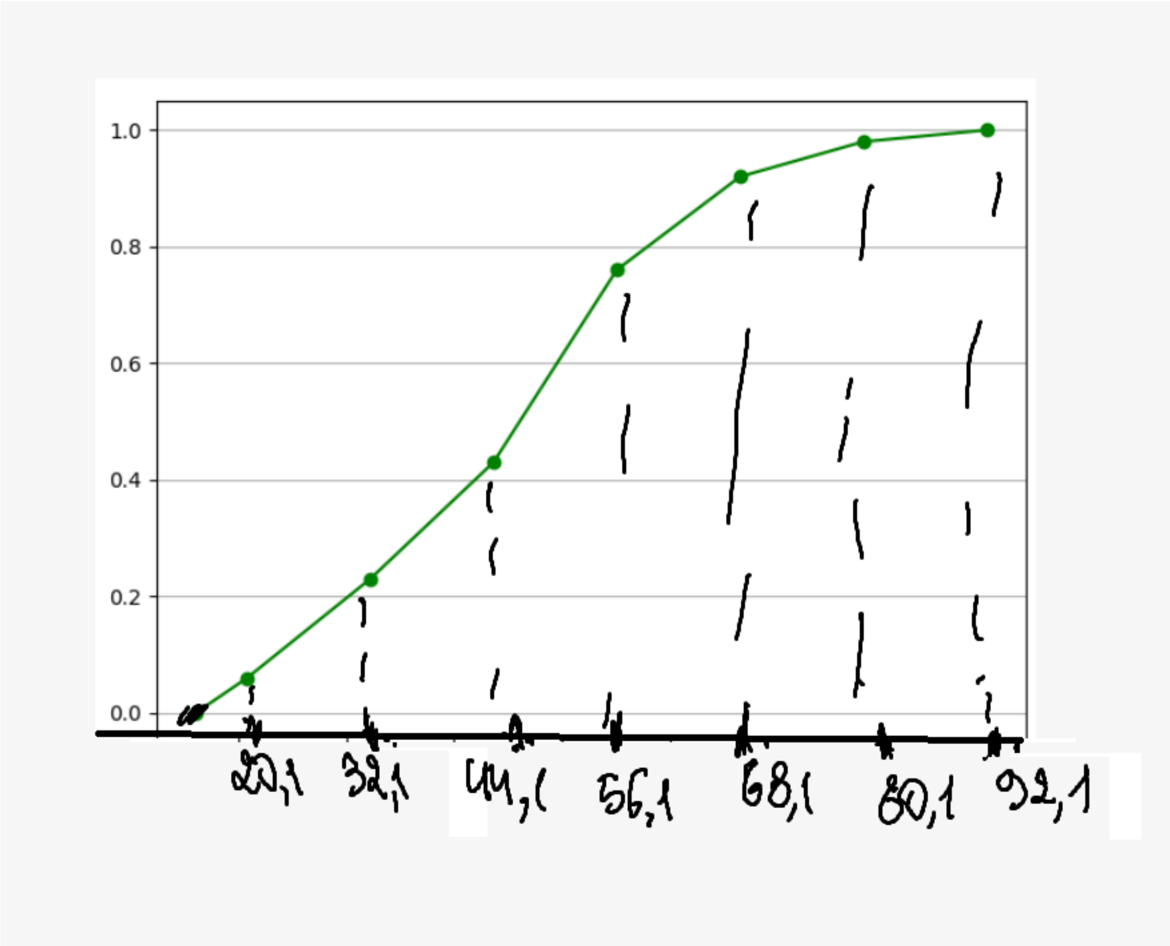
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [6, 17, 20, 33, 16, 6, 2]$$

$$x_i = [20.1, 32.1, 44.1, 56.1, 68.1, 80.1, 92.1]$$

$$F(x) = \begin{cases} 0.0, x <= 20.1, \\ 0.06, 20.1 <= x <= 32.1, \\ 0.23, 32.1 <= x <= 44.1, \\ 0.43, 44.1 <= x <= 56.1, \\ 0.76, 56.1 <= x <= 68.1, \\ 0.92, 68.1 <= x <= 80.1, \\ 0.98, 80.1 <= x <= 92.1, \\ 1.0, x > 92.1; \end{cases}$$



- 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;

$$\bar{X} = \sum_{i=1}^{7}(w_i * x_i) = 0.06 * 20.1 + 0.17 * 32.1 + 0.2 * 44.1 + 0.33 * 56.1 + 0.16 * 68.1 + 0.06 * 80.1 + 0.02 * 92.1 = 1.206 + 5.457 + 8.82 + 18.513 + 10.896 + 4.806 + 1.842 = 51.54$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7} (x_i * w_i) = 51.54$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7} \left(x_i - X_{cp}\right)^2 * w_i$$
$$= (20.1 - 51.54)^2 * 0.06 + (32.1 - 51.54)^2 * 0.17 + (44.1 - 51.54)^2 * 0.2 + (56.1 - 51.54)^2 * 0.33 + (68.1 - 51.54)^2 * 0.16 + (80.1 - 51.54)^2 * 0.06 + (92.1 - 51.54)^2 * 0.02 = 267.2064$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 267.2064 = 269.9056$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{267.2064} \approx 16.3465$$

исправленное выборочное с.к.о s

$$s = \sqrt{S^2} = \sqrt{269.9055} \approx 16.4287$$

- 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 = [54.3 58.1 45.1 46.1 62.3 63.4 88.9 46.1 60.6 62.4]

Y = [14.1 25.1 49.1 25.6 50.1 48.1 46.6 59.1 53.1 52.8]

1) Оценить тесноту линейной корреляционной зависимости между признаками

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
	54.30	14.10	765.63	2948.49	198.81
	58.10	25.10	1458.31	3375.61	630.01
	45.10	49.10	2214.41	2034.01	2410.81
	46.10	25.60	1180.16	2125.21	655.36
	62.30	50.10	3121.23	3881.29	2510.01
	63.40	48.10	3049.54	4019.56	2313.61
	88.90	46.60	4142.74	7903.21	2171.56
	46.10	59.10	2724.51	2125.21	3492.81
	60.60	53.10	3217.86	3672.36	2819.61
	62.40	52.80	3294.72	3893.76	2787.84
Сумма	587.30	423.70	25169.11	35978.71	19990.43

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.

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

$$= \frac{\frac{25169.11}{10} - \frac{587.3}{10} * \frac{423.7}{10}}{sqrt\left(\frac{35978.71}{10} - \left(\frac{587.3}{10}\right)^2\right) * sqrt\left(\frac{19990.43}{10} - \left(\frac{423.7}{10}\right)^2\right)} = 0.1638$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 58{,}73$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 42{,}37$$

$$\sigma_{вx}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 148{,}6581 \quad => \quad \sigma_{вx} \approx 12{,}1925$$

$$\sigma_{вy}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 203{,}8261 \quad => \quad \sigma_{вy} \approx 14{,}2768$$

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -246322{,}099$$

$$y_x = 0.1918 * x + 31.1023$$

$$r_{xy} = 0.1638$$