Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[7.2 11.2 16.5 19. 19.6 22.2 22.7 23.2 23.6 24.2]

[25. 27.9 28.2 28.2 29.6 29.6 30.5 31.2 31.2 31.2]

[31.3 32.4 33.2 34.7 35.7 36.8 38.3 38.8 40. 40.5]

[40.9 42.2 42.2 43.2 43.2 43.5 43.6 43.7 44.2 45.]

[45. 45.2 45.2 46.2 47.2 47.3 47.5 47.7 48.6 49.]

[49. 49.9 50. 50. 50.2 50.6 50.7 50.9 51.1 51.2]

[51.2 51.4 52. 52.5 52.9 53.2 54.2 55.2 55.2 55.6]

[56.2 56.2 56.6 57.8 58.4 58.7 58.7 59.4 59.5 59.8]

[60.5 63.9 64. 64.2 64.2 66. 66. 68.2 69.8 70.3]

[70.4 71.6 72.4 76.2 77.7 77.9 79.5 79.7 86. 91.2]

Шаг 1. Найти размах вариации

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

Опрделим максимальное и минимальное значение имеющихся значений: $x_{min} = 7.2$; $x_{max} = 91.2$

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

$$h = \left\lceil\frac{R}{k}\right\rceil = \left\lceil\frac{84.0}{7}\right\rceil = \lceil 12.0 \rceil = 12$$

Шаг 4. Найти узлы ряда:

$$a_0 = x_{\min} = 7{,}2$$

$$a_i = a_0 + i * h = 7.2 + i * 12, i = 1, \dots, 7$$

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

$[a_{i-1};\ a_i)$: $[7.2;\ 19.2)$; $[19.2;\ 31.2)$; $[31.2;\ 43.2)$; $[43.2;\ 55.2)$; $[55.2;\ 67.2)$; $[67.2;\ 79.2)$;

$[79.2;\ 91.2)$

- построим гистограмму относительных частот;

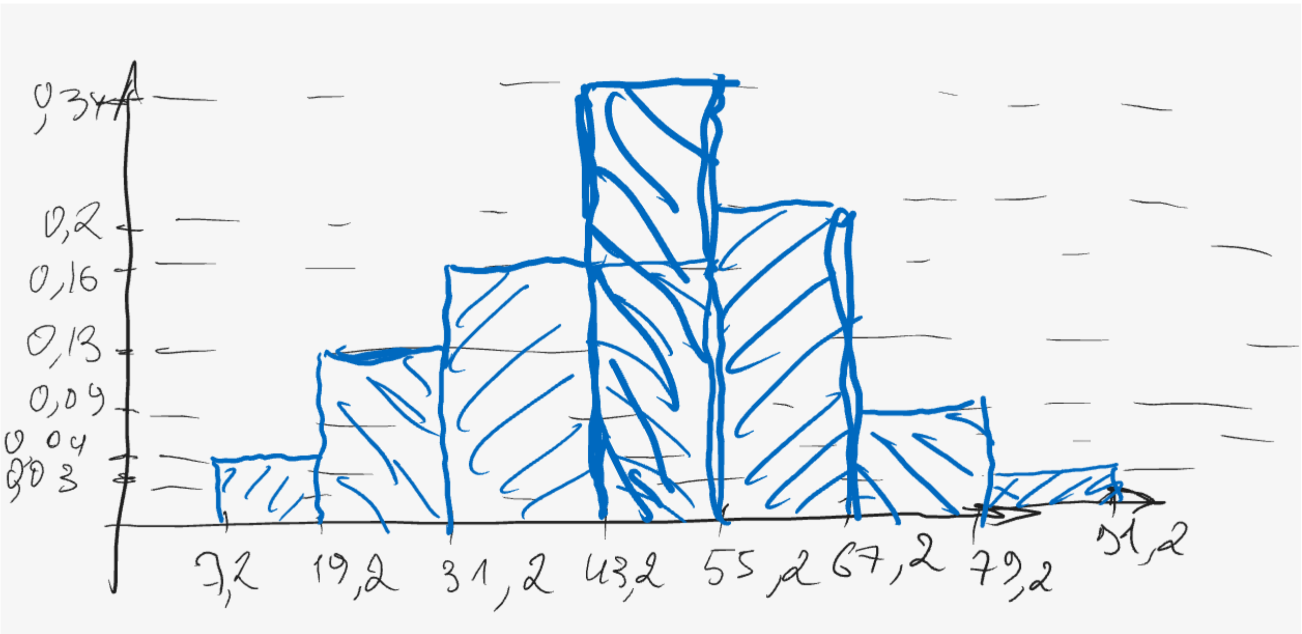
Найти частоты f_i – число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

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

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

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

$[a_{i-1}; a_i)$	[7.2, 19.2)	[19.2, 31.2)	[31.2, 43.2)	[43.2, 55.2)	[55.2, 67.2)	[67.2, 79.2)	[79.2, 91.2)
n_i	4	13	16	34	20	9	3
n	100	100	100	100	100	100	100
w_i	0.04	0.13	0.16	0.34	0.2	0.09	0.03



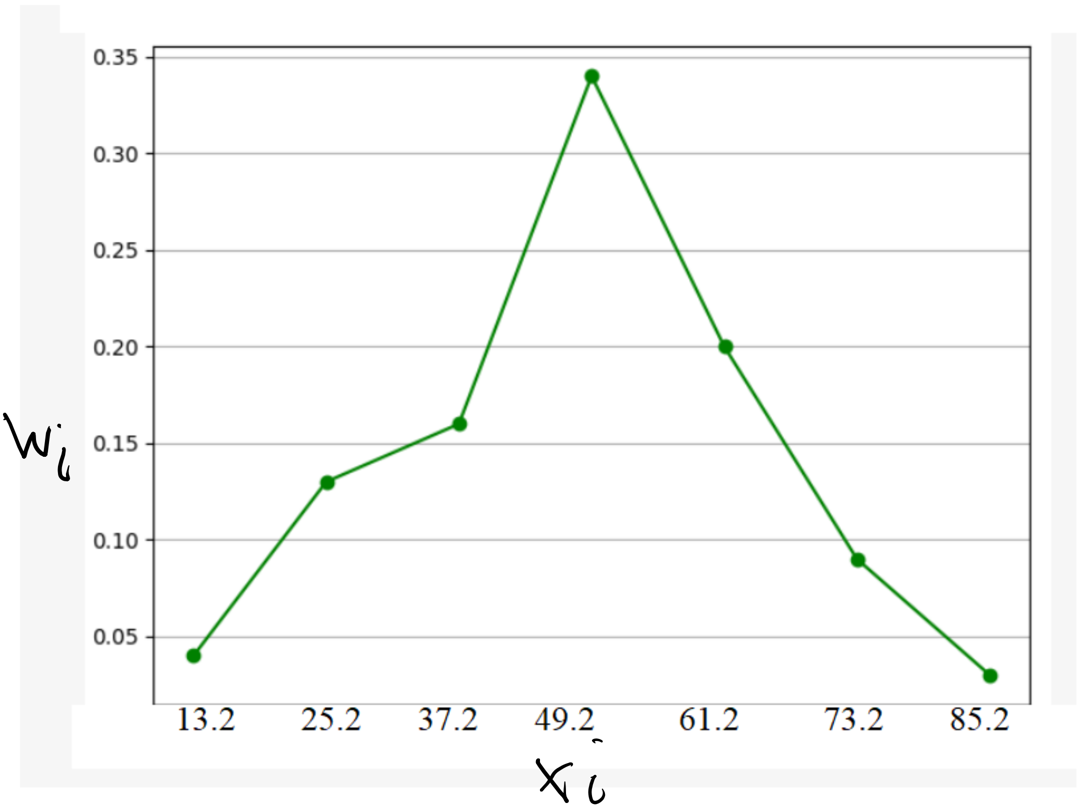
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	13.20	25.20	37.20	49.20	61.20	73.20	85.20
n_i	4.00	13.00	16.00	34.00	20.00	9.00	3.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.04	0.13	0.16	0.34	0.20	0.09	0.03

- 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:

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



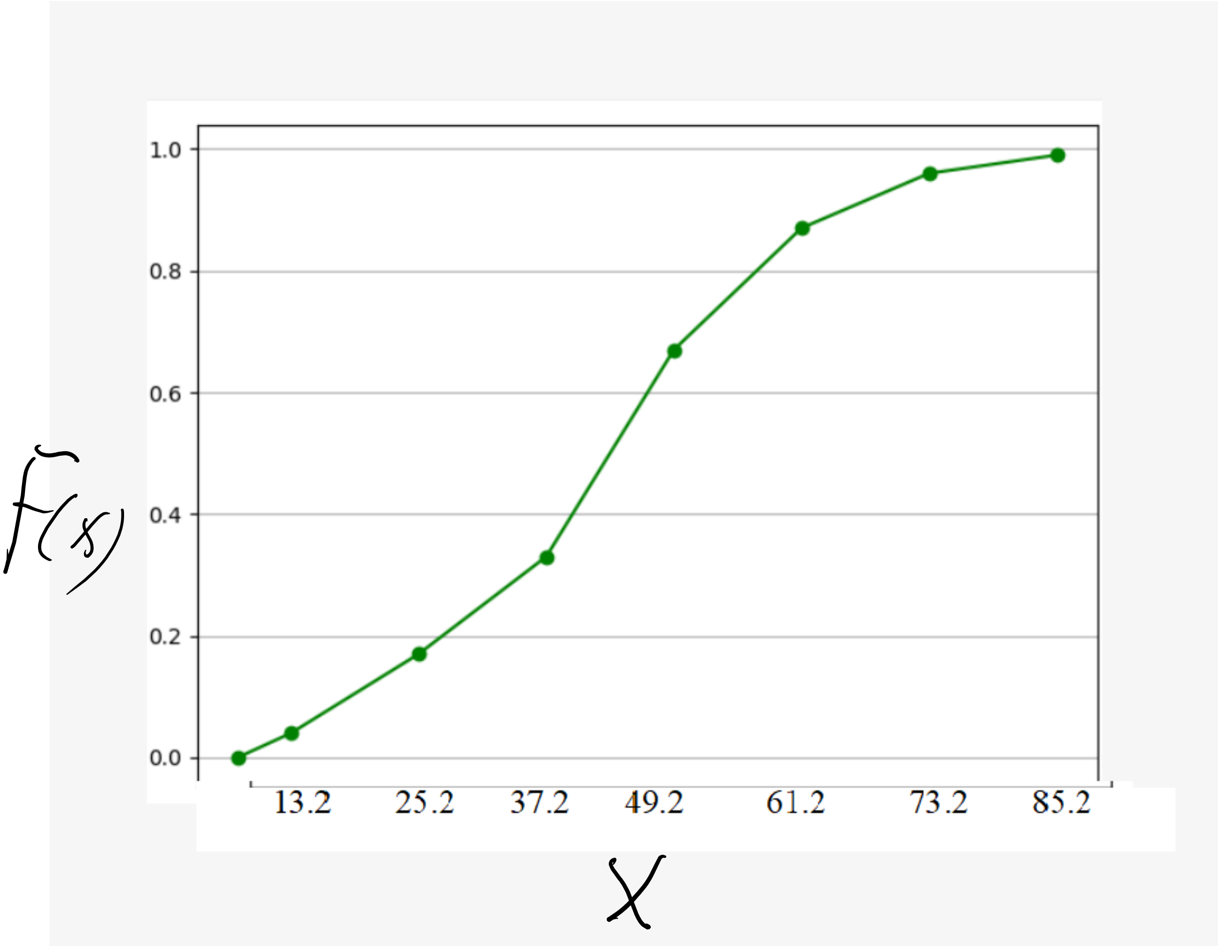
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [4, 13, 16, 34, 20, 9, 3]$$

$$x_i = [13.2, 25.2, 37.2, 49.2, 61.2, 73.2, 85.2]$$

$$F(x) = \begin{cases} 0.0, & x <= 13.2, \\ 0.04, & 13.2 <= x <= 25.2, \\ 0.17, & 25.2 <= x <= 37.2, \\ 0.33, & 37.2 <= x <= 49.2, \\ 0.67, & 49.2 <= x <= 61.2, \\ 0.87, & 61.2 <= x <= 73.2, \\ 0.96, & 73.2 <= x <= 85.2, \\ 0.99, & x > 85.2; \end{cases}$$



- 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;

$$\bar{X} = \sum_{i=1}^{7}(w_i * x_i)$$
$$= 0.04 * 13.2 + 0.13 * 25.2 + 0.16 * 37.2 + 0.34 * 49.2 + 0.2 * 61.2 + 0.09 * 73.2 + 0.03 * 85.2 =$$

$$= 0.528 + 3.276 + 5.952 + 16.728 + 12.24 + 6.588 + 2.556 =$$

$$= 47.868$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7}(x_i * w_i) = 47.868$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7} \left(x_i - X_{cp}\right)^2 * w_i =$$

$$= (13.2 - 47.868)^2 * 0.04 + (25.2 - 47.868)^2 * 0.13 + (37.2 - 47.868)^2 * 0.16 + (49.2 - 47.868)^2 * 0.34 + (61.2 - 47.868)^2 * 0.2 + (73.2 - 47.868)^2 * 0.09 + (85.2 - 47.868)^2 * 0.03 =$$

$$= 268.7987$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 268.7987 = 271.5138$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{268.7987} \approx 16.3950$$

исправленное выборочное с.к.о s

$$s = \sqrt{S^2} = \sqrt{271.5138} \approx 16.4776$$

- 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 = [51.4 55.2 42.2 43.2 59.4 60.5 86. 43.2 77.7 59.5]

Y = [11.2 22.2 46.2 47.2 45.2 43.7 56.2 50.2 49.9 22.7]

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
	51.40	11.20	575.68	2641.96	125.44
	55.20	22.20	1225.44	3047.04	492.84
	42.20	46.20	1949.64	1780.84	2134.44
	43.20	47.20	2039.04	1866.24	2227.84
	59.40	45.20	2684.88	3528.36	2043.04
	60.50	43.70	2643.85	3660.25	1909.69
	86.00	56.20	4833.20	7396.00	3158.44
	43.20	50.20	2168.64	1866.24	2520.04
	77.70	49.90	3877.23	6037.29	2490.01
	59.50	22.70	1350.65	3540.25	515.29
Сумма	578.30	394.70	23348.25	35364.47	17617.07

1) Оценить тесноту линейной корреляционной зависимости между признаками

Коэффициент корреляции Пирсона вычисляется по формуле:

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

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

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

$$\frac{\frac{23348.25}{10} - \frac{578.3}{10} * \frac{394.7}{10}}{\sqrt{\frac{35364.47}{10} - \left(\frac{578.3}{10}\right)^2} * \sqrt{\frac{17617.07}{10} - \left(\frac{394.7}{10}\right)^2}} = 0.26415$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 57{,}83$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = 39{,}47$$

$$\sigma_{вx}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 192{,}1381 \quad => \quad \sigma_{вx} \approx 13{,}864$$

$$\sigma_{вy}^2 = \frac{1}{n} \sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 203{,}8261 \quad => \quad \sigma_{вy} \approx 14{,}2767$$

$$\bar{\mu}_{xy} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -225920{,}185$$

$$y_x = 0.2721 * x + 23.7362$$

$$r_{xy} = 0.2642$$