Исходная таблица

[[51.5 55.3 42.3 43.3 59.5 60.6 86.1 43.3 77.8 59.6]

[11.3 22.3 46.3 22.8 47.3 45.3 43.8 56.3 50.3 50.]

[76.3 64.3 16.6 56.3 47.8 54.3 64.1 79.8 68.3 35.8]

[51.2 50.1 51. 70.8 31.3 33.3 23.7 53.3 71.7 58.5]

[25.1 51.3 72.5 24.3 49.1 48.7 52.1 79.6 28.3 57.9]

[52.6 59.9 29.7 43.7 55.7 53. 50.1 50.7 58.8 46.7]

[34.8 51.3 28.3 41. 58.8 49.1 19.7 36.9 29.7 38.9]

[50.8 28. 35.3 69.9 30.6 64. 32.5 45.1 45.3 70.4]

[47.6 78. 38.4 70.5 40.6 31.3 44.3 47.4 91.3 64.3]

[31.3 45.1 66.1 23.3 40.1 43.6 66.1 42.3 19.1 31.3]]

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[11.3 16.6 19.1 19.7 22.3 22.8 23.3 23.7 24.3 25.1]

[28. 28.3 28.3 29.7 29.7 30.6 31.3 31.3 31.3 31.3]

[32.5 33.3 34.8 35.3 35.8 36.9 38.4 38.9 40.1 40.6]

[41. 42.3 42.3 43.3 43.3 43.6 43.7 43.8 44.3 45.1]

[45.1 45.3 45.3 46.3 46.7 47.3 47.4 47.6 47.8 48.7]

[49.1 49.1 50. 50.1 50.1 50.3 50.7 50.8 51. 51.2]

[51.3 51.3 51.5 52.1 52.6 53. 53.3 54.3 55.3 55.7]

[56.3 56.3 57.9 58.5 58.8 58.8 59.5 59.6 59.9 60.6]

[64. 64.1 64.3 64.3 66.1 66.1 68.3 69.9 70.4 70.5]

[70.8 71.7 72.5 76.3 77.8 78. 79.6 79.8 86.1 91.3]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 11.3;\ x_{max} = 91.3$

$$R = x_{max} - x_{min} = 91.3 - 11.3 = 80.0$$

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

$$h = \left\lceil \frac{R}{k} \right\rceil = \left\lceil \frac{80.0}{7} \right\rceil = \lceil 11.428571428571429 \rceil = 12$$

Шаг 4. Найти узлы ряда:

$$a_0 = x_{min} = 11.3$$

$$a_i = a_0 + i * h = 11.3 + i * 12, i = 1, \ldots, 7$$

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

$[a_{i-1};\ a_i)$: $[11.3;\ 23.3)$; $[23.3;\ 35.3)$; $[35.3;\ 47.3)$; $[47.3;\ 59.3)$; $[59.3;\ 71.3)$; $[71.3;\ 83.3)$;

$[83.3;\ 95.3)$

- построим гистограмму относительных частот;

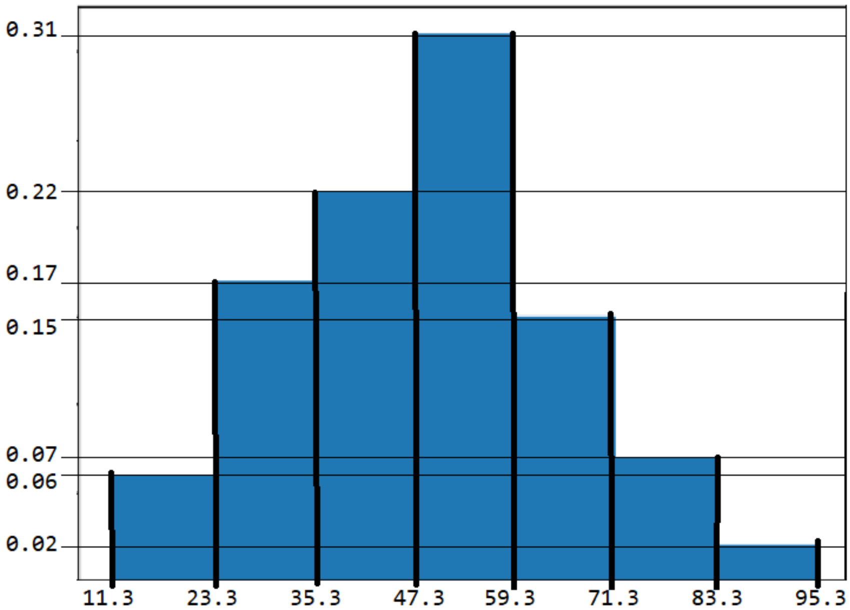
Найти частоты
f_i– число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

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

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

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

$[a_{i-1}; a_i)$	[11.3, 23.3)	[23.3, 35.3)	[35.3, 47.3)	[47.3, 59.3)	[59.3, 71.3)	[71.3, 83.3)	[83.3, 95.3)
n_i	6	17	22	31	15	7	2
n	100	100	100	100	100	100	100
w_i	0.06	0.17	0.22	0.31	0.15	0.07	0.02



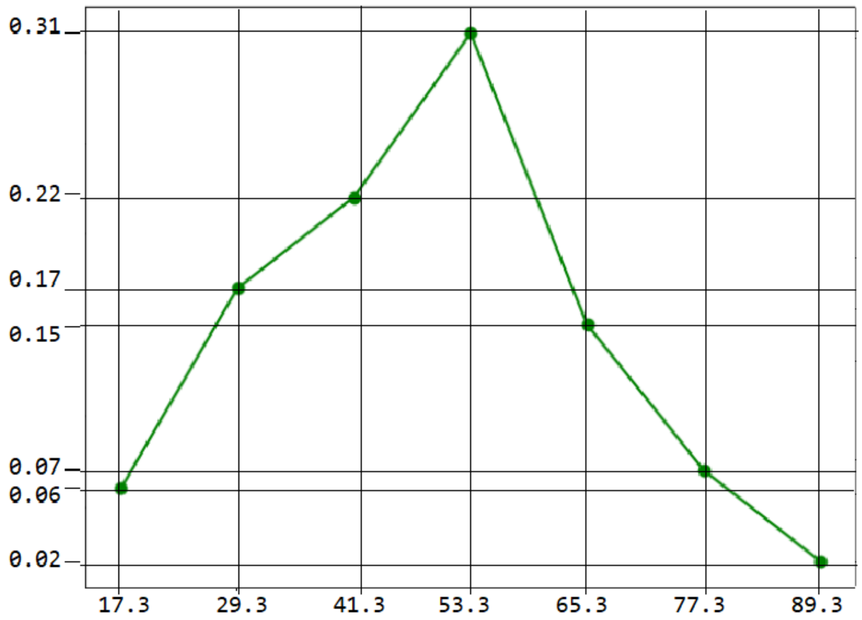
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	17.30	29.30	41.30	53.30	65.30	77.30	89.30
n_i	6.00	17.00	22.00	31.00	15.00	7.00	2.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.06	0.17	0.22	0.31	0.15	0.07	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:

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



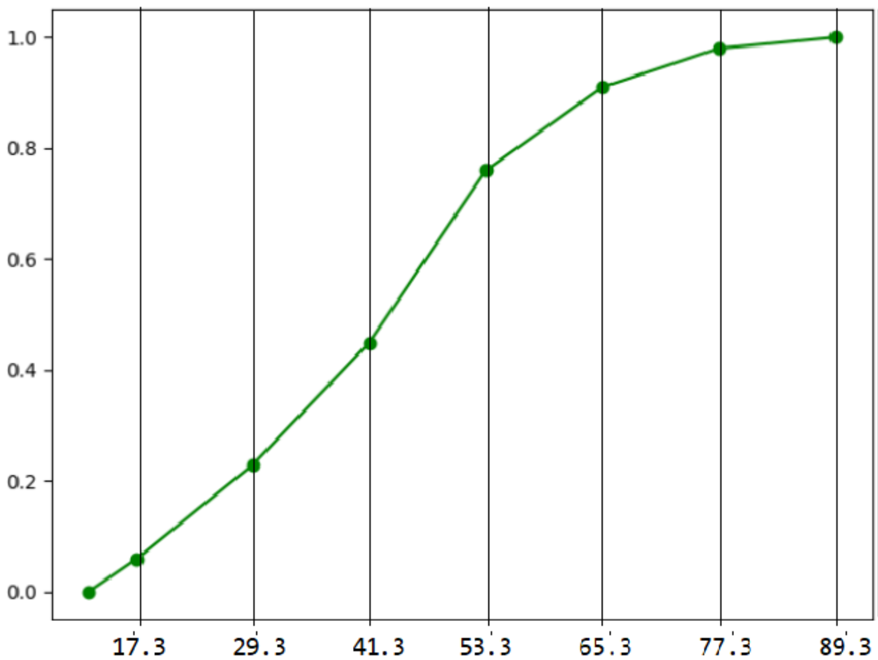
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [6, 17, 22, 31, 15, 7, 2]$$

$$x_i = [17.3, 29.3, 41.3, 53.3, 65.3, 77.3, 89.3]$$

$$
F(x) = \begin{cases}
0.0, & x <= 17.3, \\
0.06, & 17.3 <= x <= 29.3, \\
0.23, & 29.3 <= x <= 41.3, \\
0.45, & 41.3 <= x <= 53.3, \\
0.76, & 53.3 <= x <= 65.3, \\
0.91, & 65.3 <= x <= 77.3, \\
0.98, & 77.3 <= x <= 89.3, \\
1.0, & x > 89.3;
\end{cases}
$$



- вычислим все точечные статистические оценки числовых характеристик признака: среднее X̄; выборочную дисперсию и исправленную

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

$$\bar{X} = \sum_{i=1}^{7}(w_i * x_i)$$
$$= 0.06 * 17.3 + 0.17 * 29.3 + 0.22 * 41.3 + 0.31 * 53.3 + 0.15 * 65.3 + 0.07 * 77.3 + 0.02 * 89.3$$
$$= 1.038 + 4.981 + 9.086 + 16.523 + 9.795 + 5.411 + 1.786$$
$$= 48.62$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7}(x_i * w_i) = 48.62$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7}\left(x_i - X_{cp}\right)^2 * w_i$$
$$= (17.3 - 48.62)^2 * 0.06 + (29.3 - 48.62)^2 * 0.17 + (41.3 - 48.62)^2 * 0.22 + (53.3 - 48.62)^2 * 0.31 + (65.3 - 48.62)^2 * 0.15 + (77.3 - 48.62)^2 * 0.07 + (89.3 - 48.62)^2 * 0.02 = 273.2976$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 273.2976 \approx 276.0582$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{273.29760000000005} \approx 16.5317$$

$$\text{исправленное выборочное с. к. о } s$$

$$s = \sqrt{S^2} \approx \sqrt{276.0582} \approx 16.615$$

- 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

$$X = [51.5\ 55.3\ 42.3\ 43.3\ 59.5\ 60.6\ 86.1\ 43.3\ 77.8\ 59.6]$$

$Y = [11.3\ 22.3\ 46.3\ 22.8\ 47.3\ 45.3\ 43.8\ 56.3\ 50.3\ 50.]$

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
	51.50	11.30	581.95	2652.25	127.69
	55.30	22.30	1233.19	3058.09	497.29
	42.30	46.30	1958.49	1789.29	2143.69
	43.30	22.80	987.24	1874.89	519.84
	59.50	47.30	2814.35	3540.25	2237.29
	60.60	45.30	2745.18	3672.36	2052.09
	86.10	43.80	3771.18	7413.21	1918.44
	43.30	56.30	2437.79	1874.89	3169.69
	77.80	50.30	3913.34	6052.84	2530.09
	59.60	50.00	2980.00	3552.16	2500.00
Сумма	579.30	395.70	23422.71	35480.23	17696.11

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.

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

$$\frac{\frac{23422.71}{10} - \frac{579.3}{10} * \frac{395.7}{10}}{\sqrt{\frac{35480.23}{10} - \left(\frac{579.3}{10}\right)^2} * \sqrt{\frac{17696.11}{10} - \left(\frac{395.7}{10}\right)^2}} = 0.2526$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 57.93$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 39.57$$

$$\sigma_{вx}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 192.1381 \quad \Rightarrow \quad \sigma_{вx} \approx 13.8614$$

$$\sigma_{вy}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 203.8261 \quad \Rightarrow \quad \sigma_{вy} \approx 14.2768$$

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -226886.739$$

$$y_x = 0.26013008351805156 * x + 24.500664261799272$$

$$r_{xy} = 0.25256166320516293$$