[[54.1 57.9 44.9 45.9 62.1 62.2 88.7 45.8 80.4 63.2]

[13.9 24.9 48.9 47.9 46.4 58.9 52.9 52.6 25.4 49.9]

[78.9 65.9 19.2 58.9 50.4 56.9 66.7 82.4 70.9 38.4]

[53.8 52.7 53.6 72.6 33.9 35.9 26.3 55.9 74.3 61.1]

[27.7 53.9 75.1 26.9 51.8 51.3 54.7 82.2 30.9 60.5]

[55.2 62.5 32.3 46.3 58.3 55.6 52.7 53.1 61.4 51.3]

[37.4 53.9 30.9 43.6 61.4 51.7 22.3 39.5 32.3 41.5]

[53.4 30.6 57.9 75.2 33.2 66.6 35.1 47.7 47.9 73.]

[50.2 80.6 41. 73.1 43.2 33.9 46.9 50. 93.9 66.9]

[33.9 47.7 68.7 25.9 42.7 46.2 68.7 44.9 21.7 33.9]]

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[13.9 19.2 21.7 22.3 24.9 25.4 25.9 26.3 26.9 27.7]

[30.6 30.9 30.9 32.3 32.3 33.2 33.9 33.9 33.9 33.9]

[35.1 35.9 37.4 38.4 39.5 41. 41.5 42.7 43.2 43.6]

[44.9 44.9 45.8 45.9 46.2 46.3 46.4 46.9 47.7 47.7]

[47.9 47.9 48.9 49.9 50. 50.2 50.4 51.3 51.3 51.7]

[51.8 52.6 52.7 52.7 52.9 53.1 53.4 53.6 53.8 53.9]

[53.9 54.1 54.7 55.2 55.6 55.9 56.9 57.9 57.9 58.3]

[58.9 58.9 60.5 61.1 61.4 61.4 62.1 62.2 62.5 63.2]

[65.9 66.6 66.7 66.9 68.7 68.7 70.9 72.6 73. 73.1]

[74.3 75.1 75.2 78.9 80.4 80.6 82.2 82.4 88.7 93.9]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 13.9$; $x_{max} = 93.9$

$$R = x_{max} - x_{min} = 93.9 - 13.9 = 80.0$$

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:

$$a_0 = x_{min} = 13.9$$

$$a_i = a_0 + i * h = 13.9 + i * 12, i = 1, \dots, 7$$

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

$$[a_{i-1}; a_i): [13.9; 25.9); [25.9; 37.9); [37.9; 49.9); [49.9; 61.9); [61.9; 73.9);$$

$$[73.9; 85.9); [85.9; 97.9)$$

- построим гистограмму относительных частот;

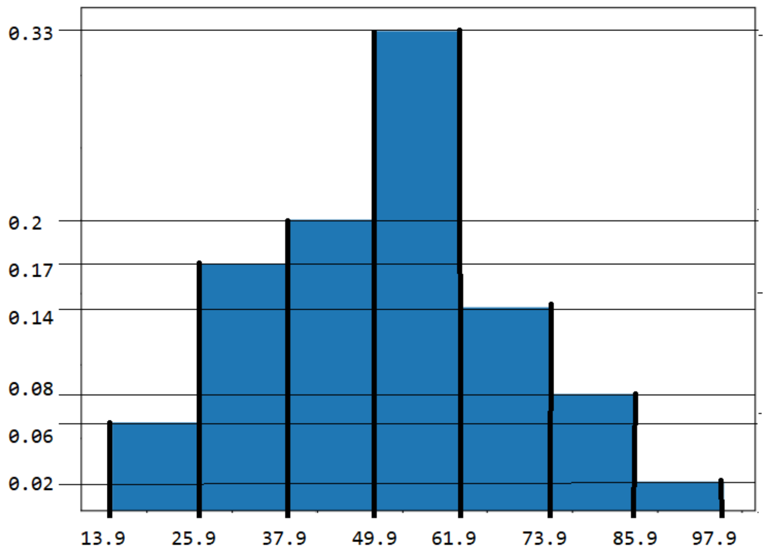
Найти частоты f_i– число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

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

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

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

$[a_{i-1}; a_i)$	[13.9, 25.9)	[25.9, 37.9)	[37.9, 49.9)	[49.9, 61.9)	[61.9, 73.9)	[73.9, 85.9)	[85.9, 97.9)
n_i	6	17	20	33	14	8	2
n	100	100	100	100	100	100	100
w_i	0.06	0.17	0.2	0.33	0.14	0.08	0.02



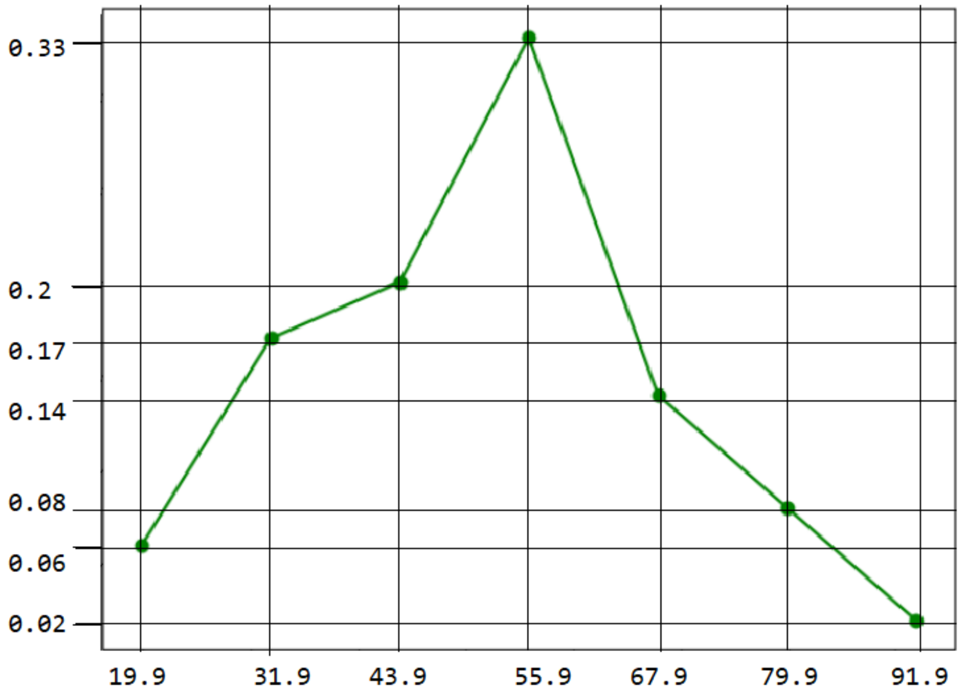
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	19.9	31.9	43.9	55.9	67.9	79.9	91.9
n_i	6	17	20	33	14	8	2
n	100	100	100	100	100	100	100
w_i	0.06	0.17	0.2	0.33	0.14	0.08	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:

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



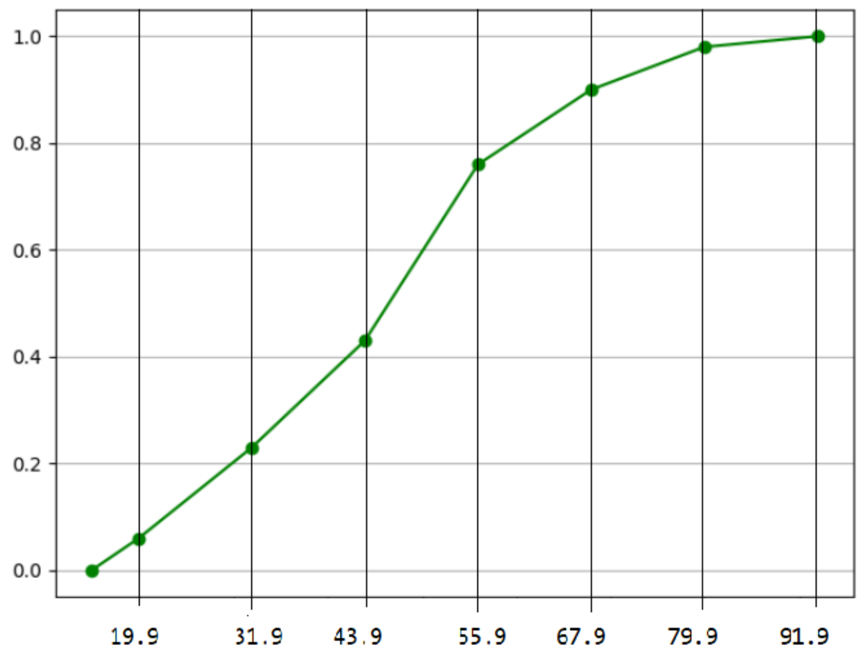
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [6, 17, 20, 33, 14, 8, 2]$$

$$x_i = [19.9, 31.9, 43.9, 55.9, 67.9, 79.9, 91.9]$$

$$F(x) = \begin{cases} 0.0, & x <= 19.9, \\ 0.06, & 19.9 <= x <= 31.9, \\ 0.23, & 31.9 <= x <= 43.9, \\ 0.43, & 43.9 <= x <= 55.9, \\ 0.76, & 55.9 <= x <= 67.9, \\ 0.9, & 67.9 <= x <= 79.9, \\ 0.98, & 79.9 <= x <= 91.9, \\ 1.0, & x > 91.9; \end{cases}$$



- 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;

$$\bar{X} = \sum_{i=1}^{7}(w_i * x_i)$$
$$= 0.06 * 19.9 + 0.17 * 31.9 + 0.2 * 43.9 + 0.33 * 55.9 + 0.14 * 67.9 + 0.08 * 79.9 + 0.02 * 91.9$$
$$= 1.194 + 5.423 + 8.78 + 18.447 + 9.506 + 6.392 + 1.838$$
$$= 51.58$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7}(x_i * w_i) = 51.58$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7}\left(x_i - X_{cp}\right)^2 * w_i$$
$$= (19.9 - 51.58)^2 * 0.06 + (31.9 - 51.58)^2 * 0.17 + (43.9 - 51.58)^2 * 0.2 + (55.9 - 51.58)^2 * 0.33 + (67.9 - 51.58)^2 * 0.14 + (79.9 - 51.58)^2 * 0.08 + (91.9 - 51.58)^2 * 0.02 = 277.9776$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 277.9776 \approx 280.7854$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{277.9776} \approx 16.6727$$

исправленное выборочное с. к. о s

$$s = \sqrt{S^2} \approx \sqrt{280.7854} \approx 16.7567$$

- 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

$$X = [54.1\ 57.9\ 44.9\ 45.9\ 62.1\ 62.2\ 88.7\ 45.8\ 80.4\ 63.2]$$

$Y = [13.9\ 24.9\ 48.9\ 47.9\ 46.4\ 58.9\ 52.9\ 52.6\ 25.4\ 49.9]$

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
	54.10	13.90	751.99	2926.81	193.21
	57.90	24.90	1441.71	3352.41	620.01
	44.90	48.90	2195.61	2016.01	2391.21
	45.90	47.90	2198.61	2106.81	2294.41
	62.10	46.40	2881.44	3856.41	2152.96
	62.20	58.90	3663.58	3868.84	3469.21
	88.70	52.90	4692.23	7867.69	2798.41
	45.80	52.60	2409.08	2097.64	2766.76
	80.40	25.40	2042.16	6464.16	645.16
	63.20	49.90	3153.68	3994.24	2490.01
Сумма	605.20	421.70	25430.09	38551.02	19821.35

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.

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

$$\frac{\frac{25430.09}{10} - \frac{605.2}{10} * \frac{421.7}{10}}{\sqrt{\frac{38551.02}{10} - \left(\frac{605.2}{10}\right)^2} * \sqrt{\frac{19821.35}{10} - \left(\frac{421.7}{10}\right)^2}} = -0.0460$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 60.52$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 42.17$$

$$\sigma_{вx}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 192.4316 \quad \Rightarrow \quad \sigma_{вx} \approx 13.872$$

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

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -252669.831$$

$$y_x = -0.0474 * x + 45.0381$$

$$r_{xy} = -0.0460$$