Исходная таблица

[[42.8 46.6 33.6 34.6 50.8 51.9 77.4 34.6 69.1 50.9]

[2.6 13.6 37.6 14.1 38.6 36.6 35.1 47.6 41.6 41.3]

[67.6 55.6 7.9 47.6 39.1 45.6 55.4 71.1 59.6 49.8]

[42.5 41.4 42.3 61.6 22.6 24.6 15. 44.6 63. 49.2]

[16.4 42.6 63.8 15.6 40.4 40. 43.4 70.9 19.6 40.2]

[43.9 51.2 21. 35. 47. 44.3 41.4 42. 50.1 30.2]

[26.1 42.6 19.6 32.3 50.1 40.4 11. 28.2 21. 61.7]

[42.1 19.3 46.6 61.2 21.9 55.3 23.8 36.4 36.6 55.6]

[38.9 69.3 29.7 61.8 81.9 22.6 35.6 38.7 82.6 22.6]

[36.4 57.4 14.6 31.4 34.9 57.4 33.6 10.2 22.6 27.1]]

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[2.6 7.9 10.2 11. 13.6 14.1 14.6 15. 15.6 16.4]

[19.3 19.6 19.6 21. 21. 21.9 22.6 22.6 22.6 22.6]

[23.8 24.6 26.1 27.1 28.2 29.7 30.2 31.4 32.3 33.6]

[33.6 34.6 34.6 34.9 35. 35.1 35.6 36.4 36.4 36.6]

[36.6 37.6 38.6 38.7 38.9 39.1 40. 40.2 40.4 40.4]

[41.3 41.4 41.4 41.6 42. 42.1 42.3 42.5 42.6 42.6]

[42.8 43.4 43.9 44.3 44.6 45.6 46.6 46.6 47. 47.6]

[47.6 49.2 49.8 50.1 50.1 50.8 50.9 51.2 51.9 55.3]

[55.4 55.6 55.6 57.4 57.4 59.6 61.2 61.6 61.7 61.8]

[63. 63.8 67.6 69.1 69.3 70.9 71.1 77.4 81.9 82.6]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 2.6$; $x_{max} = 82.6$

$$R = x_{max} - x_{min} = 82.6 - 2.6 = 80.0$$

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

$$h = \left\lceil \frac{R}{k} \right\rceil = \left\lceil \frac{80.0}{7} \right\rceil = \lceil 11.42857 \rceil = 12$$

Шаг 4. Найти узлы ряда:

$$a_0 = x_{min} = 2.6$$

$$a_i = a_0 + i * h = 2.6 + i * 12, i = 1, \ldots, 7$$

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

$[a_{i-1};\ a_i)$: $[2.6;\ 14.6)$; $[14.6;\ 26.6)$; $[26.6;\ 38.6)$; $[38.6;\ 50.6)$; $[50.6;\ 62.6)$; $[62.6;\ 74.6)$;

$$[74.6;\ 86.6)$$

- построим гистограмму относительных частот;

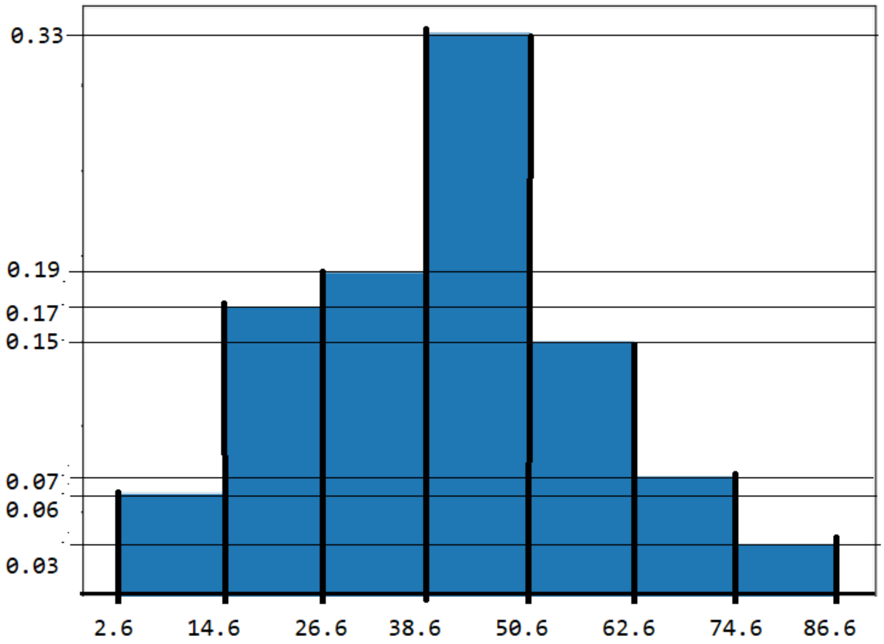
Найти частоты
f_i– число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

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

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

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

$[a_{i-1}; a_i)$	[2.6, 14.6)	[14.6, 26.6)	[26.6, 38.6)	[38.6, 50.6)	[50.6, 62.6)	[62.6, 74.6)	[74.6, 86.6)
n_i	6	17	19	33	15	7	3
n	100	100	100	100	100	100	100
w_i	0.06	0.17	0.19	0.33	0.15	0.07	0.03



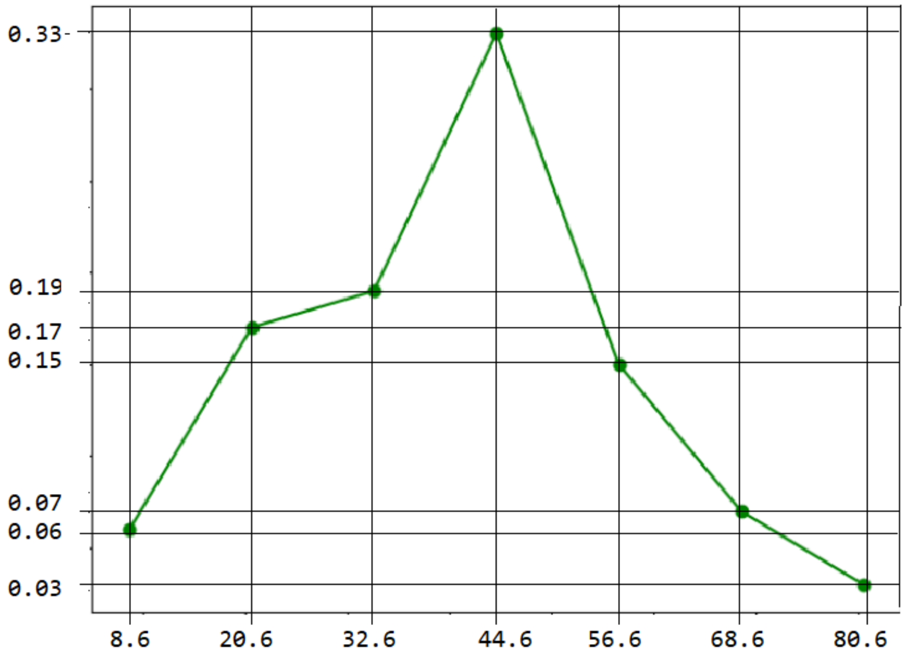
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	8.60	20.60	32.60	44.60	56.60	68.60	80.60
n_i	6.00	17.00	19.00	33.00	15.00	7.00	3.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.06	0.17	0.19	0.33	0.15	0.07	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:

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



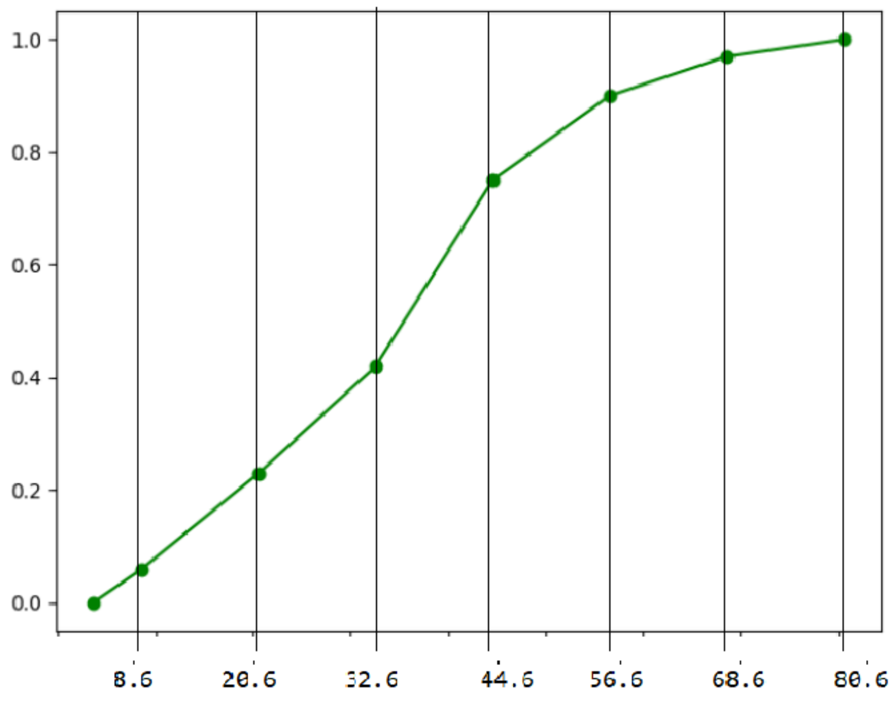
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [6, 17, 19, 33, 15, 7, 3]$$

$$x_i = [8.6, 20.6, 32.6, 44.6, 56.6, 68.6, 80.6]$$

$$F(x) = \begin{cases} 0.0, x <= 8.6, \\ 0.06, 8.6 <= x <= 20.6, \\ 0.23, 20.6 <= x <= 32.6, \\ 0.42, 32.6 <= x <= 44.6, \\ 0.75, 44.6 <= x <= 56.6, \\ 0.9, 56.6 <= x <= 68.6, \\ 0.97, 68.6 <= x <= 80.6, \\ 1.0, x > 80.6; \end{cases}$$



- 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;

$$\bar{X} = \sum_{i=1}^{7}(w_i * x_i) = 0.06 * 8.6 + 0.17 * 20.6 + 0.19 * 32.6 + 0.33 * 44.6 + 0.15 * 56.6 + 0.07 * 68.6 + 0.03 * 80.6 = 0.516 + 3.502 + 6.194 + 14.718 + 8.49 + 4.802 + 2.418 = 40.64$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7}(x_i * w_i) = 40.64$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7}\left(x_i - X_{cp}\right)^2 * w_i = (8.6 - 40.64)^2 * 0.06 + (20.6 - 40.64)^2 * 0.17 + (32.6 - 40.64)^2 * 0.19 + (44.6 - 40.64)^2 * 0.33 + (56.6 - 40.64)^2 * 0.15 + (68.6 - 40.64)^2 * 0.07 + (80.6 - 40.64)^2 * 0.03 = 288.1584$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 288.1584 \approx 291.0691$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{288.1584} \approx 16.9752$$

исправленное выборочное с. к. о s

$$s = \sqrt{S^2} \approx \sqrt{291.0691} \approx 17.06075$$

- 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

$$X = [42.8\ 46.6\ 33.6\ 34.6\ 50.8\ 51.9\ 77.4\ 34.6\ 69.1\ 50.9]$$

$$Y = [\,2.6\ 13.6\ 37.6\ 14.1\ 38.6\ 36.6\ 35.1\ 47.6\ 41.6\ 41.3\,]$$

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
	42.80	2.60	111.28	1831.84	6.76
	46.60	13.60	633.76	2171.56	184.96
	33.60	37.60	1263.36	1128.96	1413.76
	34.60	14.10	487.86	1197.16	198.81
	50.80	38.60	1960.88	2580.64	1489.96
	51.90	36.60	1899.54	2693.61	1339.56
	77.40	35.10	2716.74	5990.76	1232.01
	34.60	47.60	1646.96	1197.16	2265.76
	69.10	41.60	2874.56	4774.81	1730.56
	50.90	41.30	2102.17	2590.81	1705.69
Сумма	492.30	308.70	15697.11	26157.31	11567.83

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.

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

$$\frac{\frac{15697.11}{10} - \frac{492.3}{10} * \frac{308.7}{10}}{\sqrt{\frac{26157.31}{10} - \left(\frac{492.3}{10}\right)^2} * \sqrt{\frac{11567.83}{10} - \left(\frac{308.7}{10}\right)^2}} = 0.2526$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 49.23$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 30.87$$

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

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

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -150403.299$$

$$y_x = 0.2601 * x + 18.0638$$

$$r_{xy} = 0.2526$$