Исходная таблица

[[54.5 58.3 45.3 46.3 62.5 63.6 46.4 89.1 80.8 62.6]

[14.3 25.3 49.3 25.8 61.8 48.3 59.3 46.8 53.3 53.]

[79.3 67.3 19.6 59.3 50.3 57.3 82.9 61.7 71.3 38.8]

[54.2 53.1 54. 73.8 50.8 36.3 56.3 25.7 74.7 61.5]

[28.1 54.3 75.5 27.3 34.3 51.7 82.6 55.1 31.3 60.9]

[55.6 62.9 32.7 46.7 52.1 56. 53.7 53.1 61.8 51.7]

[37.8 54.3 31.3 44. 58.7 52.1 39.9 22.7 32.7 41.9]

[53.8 31. 58.3 72.9 33.6 67. 48.1 35.5 48.3 73.4]

[50.6 81. 41.4 73.5 43.6 34.3 50.4 47.3 94.3 67.3]

[34.3 48.1 69.1 26.3 43.1 46.6 45.3 69.1 22.1 34.3]]

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[14.3 19.6 22.1 22.7 25.3 25.7 25.8 26.3 27.3 28.1]

[31. 31.3 31.3 32.7 32.7 33.6 34.3 34.3 34.3 34.3]

[35.5 36.3 37.8 38.8 39.9 41.4 41.9 43.1 43.6 44.]

[45.3 45.3 46.3 46.4 46.6 46.7 46.8 47.3 48.1 48.1]

[48.3 48.3 49.3 50.3 50.4 50.6 50.8 51.7 51.7 52.1]

[52.1 53. 53.1 53.1 53.3 53.7 53.8 54. 54.2 54.3]

[54.3 54.5 55.1 55.6 56. 56.3 57.3 58.3 58.3 58.7]

[59.3 59.3 60.9 61.5 61.7 61.8 61.8 62.5 62.6 62.9]

[63.6 67. 67.3 67.3 69.1 69.1 71.3 72.9 73.4 73.5]

[73.8 74.7 75.5 79.3 80.8 81. 82.6 82.9 89.1 94.3]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 14.3;\ x_{max} = 94.3$

$$R = x_{max} - x_{min} = 94.3 - 14.3 = 80.0$$

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

$$h = \left\lceil \frac{R}{k} \right\rceil = \left\lceil \frac{80.0}{7} \right\rceil = \lceil 11.4286 \rceil = 12$$

Шаг 4. Найти узлы ряда:

$$a_0 = x_{min} = 14.3$$

$$a_i = a_0 + i * h = 14.3 + i * 12, i = 1, \ldots, 7$$

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

$[a_{i-1};\ a_i)$: $[14.3;\ 26.3)$; $[26.3;\ 38.3)$; $[38.3;\ 50.3)$; $[50.3;\ 62.3)$; $[62.3;\ 74.3)$; $[74.3;\ 86.3)$;

$[86.3;\ 98.3)$

- построим гистограмму относительных частот;

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

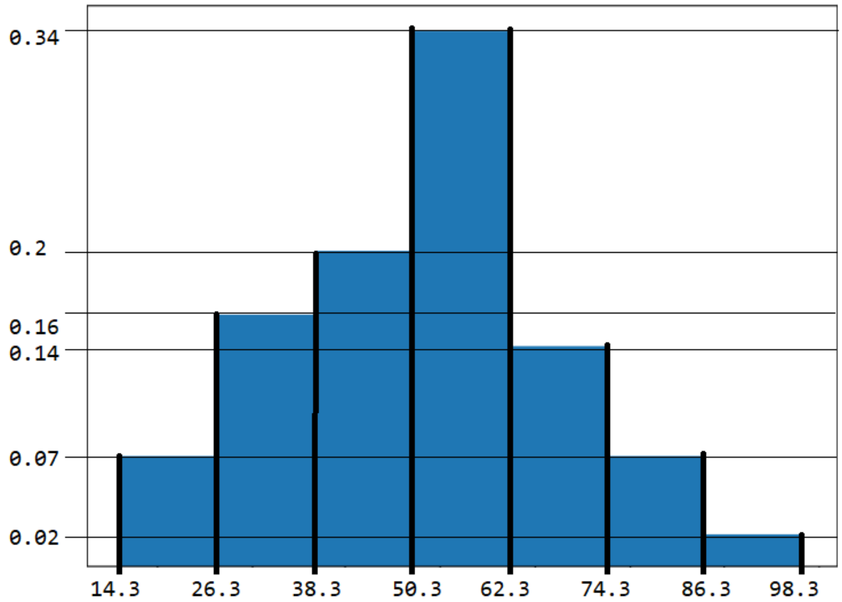
$$f_i = n_i, n_i - \text{ количество точек на интервале } [a_{i-1};\ a_i)$$

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

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

Названия строк по номерам (вписать в таблице)

$[a_{i-1}; a_i)$	[14.3, 26.3)	[26.3, 38.3)	[38.3, 50.3)	[50.3, 62.3)	[62.3, 74.3)	[74.3, 86.3)	[86.3, 98.3)
n_i	7	16	20	34	14	7	2
n	100	100	100	100	100	100	100
w_i	0.07	0.16	0.2	0.34	0.14	0.07	0.02



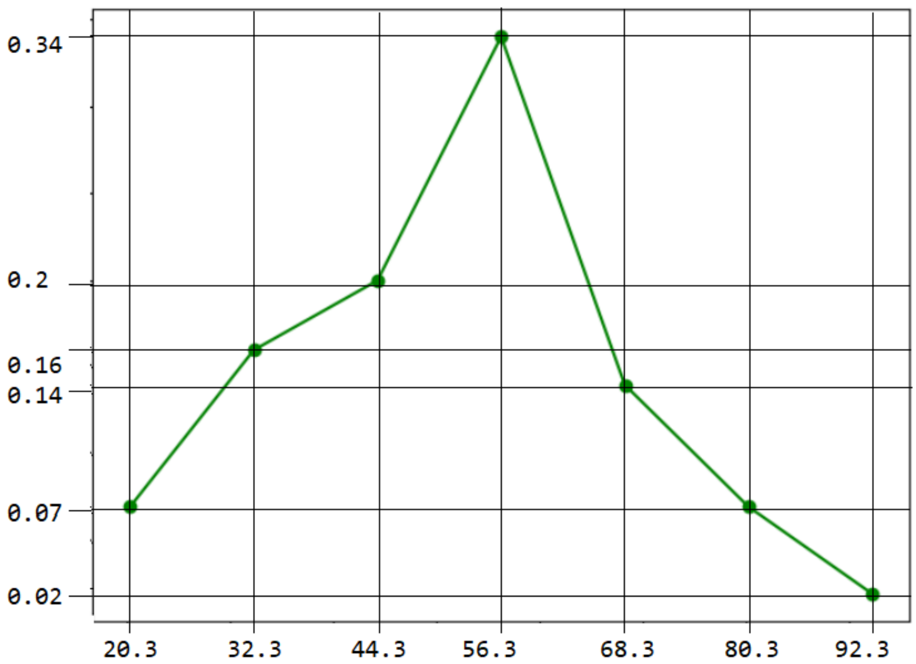
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	20.30	32.30	44.30	56.30	68.30	80.30	92.30
n_i	7.00	16.00	20.00	34.00	14.00	7.00	2.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.07	0.16	0.20	0.34	0.14	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:

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



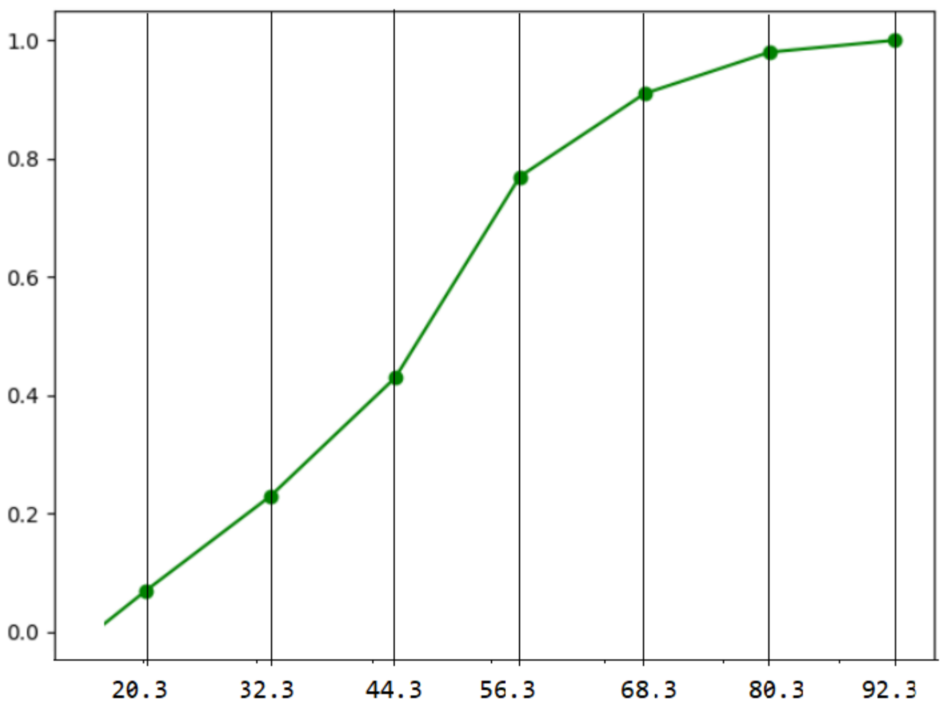
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [7, 16, 20, 34, 14, 7, 2]$$

$$x_i = [20.3, 32.3, 44.3, 56.3, 68.3, 80.3, 92.3]$$

$$F(x) = \begin{cases} 0.0, & x <= 20.3, \\ 0.07, & 20.3 <= x <= 32.3, \\ 0.23, & 32.3 <= x <= 44.3, \\ 0.43, & 44.3 <= x <= 56.3, \\ 0.77, & 56.3 <= x <= 68.3, \\ 0.91, & 68.3 <= x <= 80.3, \\ 0.98, & 80.3 <= x <= 92.3, \\ 1.0, & x > 92.3; \end{cases}$$



- 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;

$$\bar{X} = \sum_{i=1}^{7}(w_i * x_i)$$
$$= 0.07 * 20.3 + 0.16 * 32.3 + 0.2 * 44.3 + 0.34 * 56.3 + 0.14 * 68.3 + 0.07 * 80.3 + 0.02 * 92.3$$
$$= 1.421 + 5.168 + 8.86 + 19.142 + 9.562 + 5.621 + 1.846 = 51.62$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7}(x_i * w_i) = 51.62$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7}\left(x_i - X_{cp}\right)^2 * w_i =$$
$$= (20.3 - 51.62)^2 * 0.07 + (32.3 - 51.62)^2 * 0.16 + (44.3 - 51.62)^2 * 0.2 + (56.3 - 51.62)^2 * 0.34 + (68.3 - 51.62)^2 * 0.14 + (80.3 - 51.62)^2 * 0.07 + (92.3 - 51.62)^2 * 0.02 = 276.1776$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 276.1776 \approx 278.9678$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{276.1776} \approx 16.6186$$

исправленное выборочное с. к. о s

$$s = \sqrt{S^2} \approx \sqrt{278.9673} \approx 16.7023$$

- 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

$$X = [54.5\ 58.3\ 45.3\ 46.3\ 62.5\ 63.6\ 46.4\ 89.1\ 80.8\ 62.6]$$

$Y = [14.3\ 25.3\ 49.3\ 25.8\ 61.8\ 48.3\ 59.3\ 46.8\ 53.3\ 53.]$

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
	54.50	14.30	779.35	2970.25	204.49
	58.30	25.30	1474.99	3398.89	640.09
	45.30	49.30	2233.29	2052.09	2430.49
	46.30	25.80	1194.54	2143.69	665.64
	62.50	61.80	3862.50	3906.25	3819.24
	63.60	48.30	3071.88	4044.96	2332.89
	46.40	59.30	2751.52	2152.96	3516.49
	89.10	46.80	4169.88	7938.81	2190.24
	80.80	53.30	4306.64	6528.64	2840.89
	62.60	53.00	3317.80	3918.76	2809.00
Сумма	609.40	437.20	27162.39	39055.30	21449.46

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.

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

$$\frac{\frac{27162.39}{10} - \frac{609.4}{10} * \frac{437.2}{10}}{\sqrt{\frac{39055.3}{10} - \left(\frac{609.4}{10}\right)^2} * \sqrt{\frac{21449.46}{10} - \left(\frac{437.2}{10}\right)^2}} = 0.2454$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 60.94$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 43.72$$

$$\sigma_{вx}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 191.8464 \quad => \quad \sigma_{вx} \approx 13.8509$$

$$\sigma_{вy}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 233.5076 \quad => \quad \sigma_{вy} \approx 15.281$$

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -263713.441$$

$$y_x = 0.2707 * x + 27.2205$$

$$r_{xy} = 0.2454$$