Исходная таблица

[[64.5 68.3 55.3 72.5 73.6 99.1 56.3 90.8 72.6 56.3]

[19.2 35.3 59.3 60.2 58.3 56.8 69.3 63.3 63. 36.8]

[89.3 77.3 29.6 69.3 60.8 67.3 77.1 92.8 81.3 48.8]

[24.3 63.1 64. 83.3 44.3 46.3 36.7 66.3 84.7 71.5]

[38.1 64.3 85.5 37.3 62.1 61.7 65.1 92.6 41.3 70.9]

[65.6 72.9 42.7 56.7 68.7 66. 63.1 63.7 71.8 61.7]

[47.8 64.3 41.3 54. 71.8 62.1 32.7 49.9 42.7 51.9]

[63.8 41. 68.3 82.9 43.6 77. 45.5 58.1 58.3 83.4]

[60.6 91. 51.4 83.5 53.6 44.3 57.3 60.4 99.2 77.3]

[44.3 58.1 79.1 36.3 53.1 56.6 79.1 55.3 32.1 44.3]]

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[19.2 24.3 29.6 32.1 32.7 35.3 36.3 36.7 36.8 37.3]

[38.1 41. 41.3 41.3 42.7 42.7 43.6 44.3 44.3 44.3]

[44.3 45.5 46.3 47.8 48.8 49.9 51.4 51.9 53.1 53.6]

[54. 55.3 55.3 56.3 56.3 56.6 56.7 56.8 57.3 58.1]

[58.1 58.3 58.3 59.3 60.2 60.4 60.6 60.8 61.7 61.7]

[62.1 62.1 63. 63.1 63.1 63.3 63.7 63.8 64. 64.3]

[64.3 64.5 65.1 65.6 66. 66.3 67.3 68.3 68.3 68.7]

[69.3 69.3 70.9 71.5 71.8 71.8 72.5 72.6 72.9 73.6]

[77. 77.1 77.3 77.3 79.1 79.1 81.3 82.9 83.3 83.4]

[83.5 84.7 85.5 89.3 90.8 91. 92.6 92.8 99.1 99.2]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 19.2$; $x_{max} = 99.2$

$R = x_{max} - x_{min} = 99.2 - 19.2 = 80.0$

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:

$$a_0 = x_{min} = 19.2$$

$$a_i = a_0 + i * h = 19.2 + i * 12, i = 1, \ldots, 7$$

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

$[a_{i-1}; a_i)$: $[19.2; 31.2)$; $[31.2; 43.2)$; $[43.2; 55.2)$; $[55.2; 67.2)$; $[67.2; 79.2)$; $[79.2; 91.2)$;

$$[91.2; 103.2)$$

- построим гистограмму относительных частот;

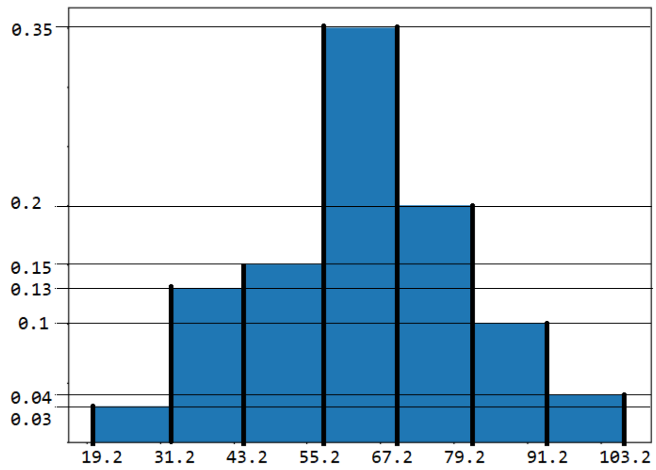
Найти частоты f_i– число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

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

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

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

$[a_{i-1}; a_i)$	[19.2, 31.2)	[31.2, 43.2)	[43.2, 55.2)	[55.2, 67.2)	[67.2, 79.2)	[79.2, 91.2)	[91.2, 103.2)
n_i	3	13	15	35	20	10	4
n	100	100	100	100	100	100	100
w_i	0.03	0.13	0.15	0.35	0.2	0.1	0.04



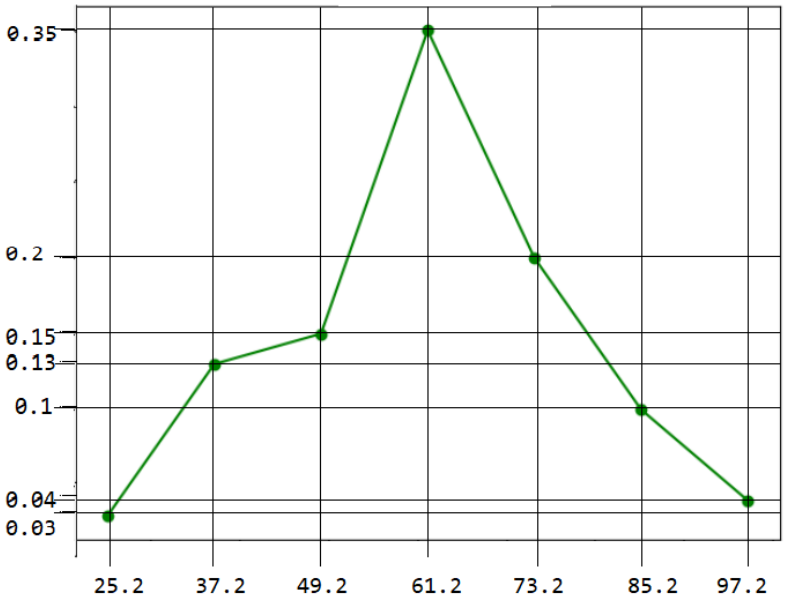
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

x_i	25.20	37.20	49.20	61.20	73.20	85.20	97.20
n_i	3.00	13.00	15.00	35.00	20.00	10.00	4.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.03	0.13	0.15	0.35	0.20	0.10	0.04

- 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:

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



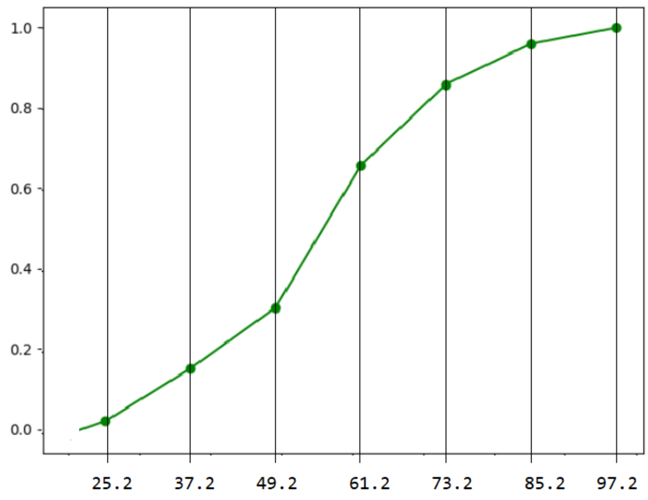
- найдем эмпирическую функцию распределения и построим ее график;

$$n = 100$$

$$n_x = [3, 13, 15, 35, 20, 10, 4]$$

$$x_i = [25.2, 37.2, 49.2, 61.2, 73.2, 85.2, 97.2]$$

$$F(x) = \begin{cases} 0.0, & x <= 25.2, \\ 0.03, & 25.2 <= x <= 37.2, \\ 0.16, & 37.2 <= x <= 49.2, \\ 0.31, & 49.2 <= x <= 61.2, \\ 0.66, & 61.2 <= x <= 73.2, \\ 0.86, & 73.2 <= x <= 85.2, \\ 0.96, & 85.2 <= x <= 97.2, \\ 1.0, & x > 97.2; \end{cases}$$



- вычислим все точечные статистические оценки числовых характеристик признака: среднее $\overline{X}$; выборочную дисперсию и исправленную выборочную дисперсию; выборочное с.к.о. и исправленное выборочное с.к.о. s;

$$\bar{X} = \sum_{i=1}^{7} (w_i * x_i)$$
$$= 0.03 * 25.2 + 0.13 * 37.2 + 0.15 * 49.2 + 0.35 * 61.2 + 0.2 * 73.2 + 0.1 * 85.2 + 0.04 * 97.2$$
$$= 0.756 + 4.836 + 7.38 + 21.42 + 14.64 + 8.52 + 3.888 = 61.44$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7} (x_i * w_i) = 61.44$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7} \left(x_i - X_{cp}\right)^2 * w_i$$
$$= (25.2 - 61.44)^2 * 0.03 + (37.2 - 61.44)^2 * 0.13 + (49.2 - 61.44)^2 * 0.15 + (61.2 - 61.44)^2 * 0.35 + (73.2 - 61.44)^2 * 0.2 + (85.2 - 61.44)^2 * 0.1 + (97.2 - 61.44)^2 * 0.04 =$$
$$= 273.5424$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 273.5424 \approx 276.3055$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{273.5424} \approx 16.5391$$

исправленное выборочное с. к. о s

$$s = \sqrt{S^2} \approx \sqrt{276.3055} \approx 16.6224$$

- 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

$$X = [64.5\ 68.3\ 55.3\ 72.5\ 73.6\ 99.1\ 56.3\ 90.8\ 72.6\ 56.3]$$

$$Y = [19.2\ 35.3\ 59.3\ 60.2\ 58.3\ 56.8\ 69.3\ 63.3\ 63.\ \ 36.8]$$

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
64.50	19.20	1238.40	4160.25	368.64
68.30	35.30	2410.99	4664.89	1246.09
55.30	59.30	3279.29	3058.09	3516.49
72.50	60.20	4364.50	5256.25	3624.04
73.60	58.30	4290.88	5416.96	3398.89
99.10	56.80	5628.88	9820.81	3226.24
56.30	69.30	3901.59	3169.69	4802.49
90.80	63.30	5747.64	8244.64	4006.89
72.60	63.00	4573.80	5270.76	3969.00
56.30	36.80	2071.84	3169.69	1354.24
Сумма 709.30	521.50	37507.81	52232.03	29513.01

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.

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

$$\frac{\frac{37507.81}{10} - \frac{709.3}{10} * \frac{521.5}{10}}{\sqrt{\frac{52232.03}{10} - \left(\frac{709.3}{10}\right)^2} * \sqrt{\frac{29513.01}{10} - \left(\frac{521.5}{10}\right)^2}} = 0.2454$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 70.93$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 52.15$$

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

$$\sigma_{вy}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 231.6785 \quad => \quad \sigma_{вy} \approx 15.221$$

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = -366149.169$$

$$y_x = 0.2695 * x + 33.0343$$

$$r_{xy} = 0.2454$$