Исходная таблица

[[56.7 60.5 47.5 48.5 64.7 65.8 91.3 83. 48.5 64.8]

 [16.5 27.5 51.5 28. 52.5 50.5 49. 55.5 61.5 55.2]

 [81.5 69.5 21.8 61.5 53. 59.5 69.3 73.5 85. 41.]

 [56.4 55.3 56.2 75.5 36.5 38.5 26.9 76.9 58.5 63.7]

 [30.3 56.5 77.7 29.5 54.3 53.9 57.3 33.5 84.8 63.1]

 [57.8 65.1 34.9 60.9 58.2 55.3 55.9 53.9 64. 48.9]

 [40. 56.5 33.5 46.2 64. 54.3 24.9 44.9 42.1 44.1]

 [56. 33.2 60.5 75.1 35.8 69.2 37.7 50.5 50.3 75.6]

 [52.8 83.2 43.6 75.7 45.8 36.5 49.5 96.5 52.6 69.5]

 [36.5 50.3 71.3 28.5 45.3 48.8 71.3 24.3 47.5 36.5]]

Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[16.5 21.8 24.3 24.9 26.9 27.5 28. 28.5 29.5 30.3]

 [33.2 33.5 33.5 34.9 35.8 36.5 36.5 36.5 36.5 37.7]

 [38.5 40. 41. 42.1 43.6 44.1 44.9 45.3 45.8 46.2]

 [47.5 47.5 48.5 48.5 48.8 48.9 49. 49.5 50.3 50.3]

 [50.5 50.5 51.5 52.5 52.6 52.8 53. 53.9 53.9 54.3]

 [54.3 55.2 55.3 55.3 55.5 55.9 56. 56.2 56.4 56.5]

 [56.5 56.7 57.3 57.8 58.2 58.5 59.5 60.5 60.5 60.9]

 [61.5 61.5 63.1 63.7 64. 64. 64.7 64.8 65.1 65.8]

 [69.2 69.3 69.5 69.5 71.3 71.3 73.5 75.1 75.5 75.6]

 [75.7 76.9 77.7 81.5 83. 83.2 84.8 85. 91.3 96.5]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений: $x_{min} = 16.5$; $x_{max} = 96.5$

$$x_{max} - x_{min} = 96.5 - 16.5 = 80.0.$$

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:

$$a_0 = x_{min} = 16.5$$

$$a_i = a_0 + i * h = 16.5 + i * 12, i = 1, \dots, 7$$

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

$[a_{i-1};\ a_i)$: [16.5; 28.5); [28.5; 40.5); [40.5; 52.5); [52.5; 64.5); [64.5; 76.5);

[76.5; 88.5); [88.5; 100.5)

- построим гистограмму относительных частот;

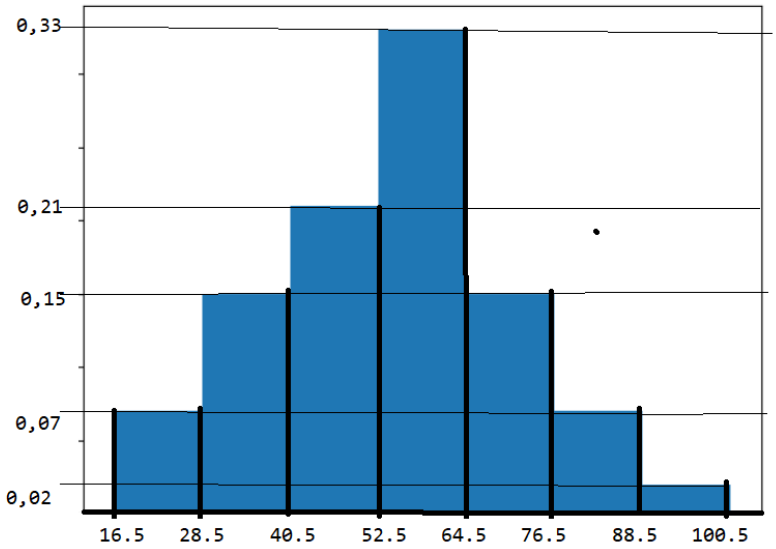
Найти частоты f_i – число попаданий значений признака в каждый из интервалов $[a_{i-1}, a_i)$

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

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

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

$[a_{i-1}; a_i)$	[16.5, 28.5)	[28.5, 40.5)	[40.5, 52.5)	[52.5, 64.5)	[64.5, 76.5)	[76.5, 88.5)	[88.5, 100.5)
n_i	7	15	21	33	15	7	2
n	100	100	100	100	100	100	100
w_i	0.07	0.15	0.21	0.33	0.15	0.07	0.02



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

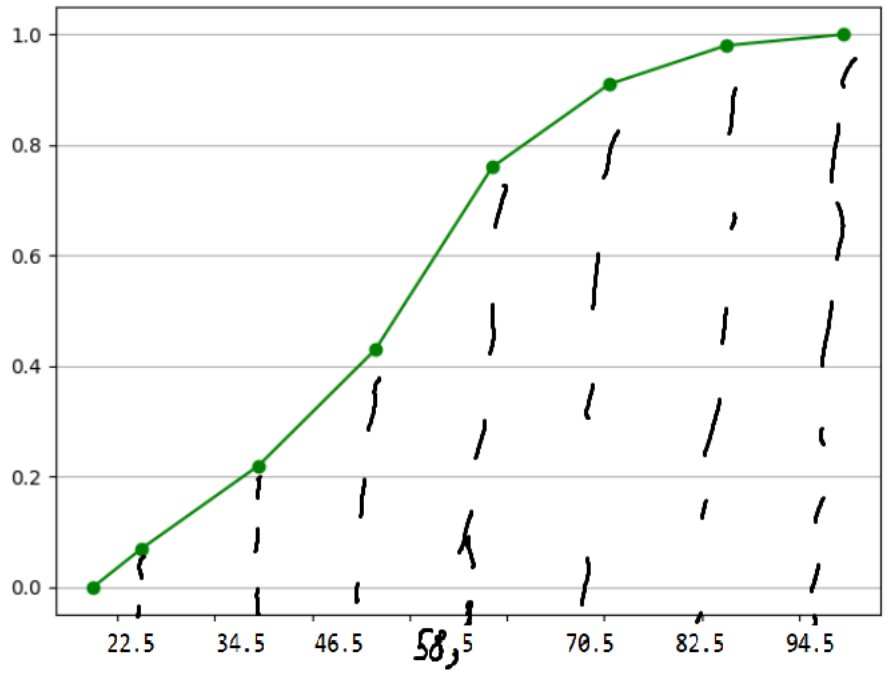
x_i	22.50	34.50	46.50	58.50	70.50	82.50	94.50
n_i	7.00	15.00	21.00	33.00	15.00	7.00	2.00
n	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w_i	0.07	0.15	0.21	0.33	0.15	0.07	0.02

- 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:

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

$$F(x) = \begin{cases} 0.0, & x <= 22.5, \\ 0.07, & 22.5 <= x <= 34.5, \\ 0.22, & 34.5 <= x <= 46.5, \\ 0.43, & 46.5 <= x <= 58.5, \\ 0.76, & 58.5 <= x <= 70.5, \\ 0.91, & 70.5 <= x <= 82.5, \\ 0.98, & 82.5 <= x <= 94.5, \\ 1.0, & x > 94.5; \end{cases}$$



- 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;

$\bar{X} = \sum_{i=1}^{7}(w_i * x_i) = 0.07 * 22.5 + 0.15 * 34.5 + 0.21 * 46.5 + 0.33 * 58.5 + 0.15 * 70.5 + 0.07 * 82.5 + 0.02 * 94.5 =$

$$= 1.575 + 5.175 + 9.765 + 19.305 + 10.575 + 5.775 + 1.89 =$$

$$= 54.06$$

Выборочная средняя:

$$X_{\text{ср}} = \sum_{i=1}^{7}(x_i * w_i) = 54.06$$

Выборочная дисперсия:

$$D = \sum_{i=1}^{7}(x_i - X_{cp})^2 * w_i =$$

$$= (22.5 - 54.06)^2 * 0.07 + (34.5 - 54.06)^2 * 0.15 + (46.5 - 54.06)^2 * 0.21 + (58.5 - 54.06)^2 * 0.33 + (70.5 - 54.06)^2 * 0.15 + (82.5 - 54.06)^2 * 0.07 + (94.5 - 54.06)^2 * 0.02 =$$

$$= 275.4864$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 275.4864 \approx 278.2690$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{275.4864} \approx 16.5977829844832$$

исправленное выборочное с.к.о s

$$s = \sqrt{S^2} = \sqrt{278.26909090909095} \approx 16.68139954887152$$

- 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 = [56.7 60.5 47.5 48.5 64.7 65.8 91.3 83. 48.5 64.8]

Y = [16.5 27.5 51.5 28. 52.5 50.5 49. 55.5 61.5 55.2]

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
56.70000	16.50000	935.55000	3214.89000	272.25000
60.50000	27.50000	1663.75000	3660.25000	756.25000
47.50000	51.50000	2446.25000	2256.25000	2652.25000
48.50000	28.00000	1358.00000	2352.25000	784.00000
64.70000	52.50000	3396.75000	4186.09000	2756.25000
65.80000	50.50000	3322.90000	4329.64000	2550.25000
91.30000	49.00000	4473.70000	8335.69000	2401.00000
83.00000	55.50000	4606.50000	6889.00000	3080.25000
48.50000	61.50000	2982.75000	2352.25000	3782.25000
64.80000	55.20000	3576.96000	4199.04000	3047.04000
631.30000	447.70000	28763.11000	41775.35000	22081.79000

Сумма

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.

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

$$\frac{\frac{28763.11}{10} - \frac{631.3}{10} * \frac{447.7}{10}}{\sqrt{\frac{41775.35}{10} - \left(\frac{631.3}{10}\right)^2} * \sqrt{\frac{22081.79}{10} - \left(\frac{447.7}{10}\right)^2}} = 0.2525$$

2) Составим выборочное уравнение прямой регрессии Y на X

2) линейное уравнение регрессии Y на X:

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 63,13$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 44,77$$

$$\sigma_{вх}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 192,1381 \quad => \quad \sigma_{вх} \approx 13,8619$$

$$\sigma_{ву}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 203,8261 \quad => \quad \sigma_{ву} \approx 14,2768$$

$$\bar{\mu}_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = 279756,699$$

$$y_x = 0.2601 * x + 28.3479$$

$$r_{xy} = 0.2525$$