

### Исходная таблица

[[54.2 58. 45. 46. 62.2 63.3 88.8 46. 80.5 62.3]  
[14. 25. 49. 25.5 50. 48. 46.5 59. 53. 52.7]  
[79. 67. 19.3 59. 50.5 57. 66.8 82.5 71. 38.5]  
[53.9 52.8 53.7 73. 34. 36. 26.4 56. 74.4 61.2]  
[27.8 54. 75.2 27. 51.8 51.4 54.8 82.3 31. 60.6]  
[55.3 62.6 32.4 46.4 58.4 55.7 52.8 53.4 61.5 51.4]  
[37.5 54. 31. 43.7 61.5 51.8 22.4 39.6 32.4 41.6]  
[53.5 30.7 58. 72.6 33.3 66.7 35.2 47.8 48. 73.6]  
[50.3 80.7 41.1 73.2 43.3 34. 47. 50.1 94. 67. ]  
[34. 47.8 68.8 26. 42.8 46.3 68.8 45. 21.8 34.7]]

### Решение:

- Составим интервальное распределение выборки

Выстроим в порядке возрастания, имеющиеся у нас значения

[[14. 19.3 21.8 22.4 25. 25.5 26. 26.4 27. 27.8]  
[30.7 31. 31. 32.4 32.4 33.3 34. 34. 34. 34.7]  
[35.2 36. 37.5 38.5 39.6 41.1 41.6 42.8 43.3 43.7]  
[45. 45. 46. 46. 46.3 46.4 46.5 47. 47.8 47.8]  
[48. 48. 49. 50. 50.1 50.3 50.5 51.4 51.4 51.8]  
[51.8 52.7 52.8 52.8 53. 53.4 53.5 53.7 53.9 54. ]  
[54. 54.2 54.8 55.3 55.7 56. 57. 58. 58. 58.4]  
[59. 59. 60.6 61.2 61.5 61.5 62.2 62.3 62.6 63.3]  
[66.7 66.8 67. 67. 68.8 68.8 71. 72.6 73. 73.2]  
[73.6 74.4 75.2 79. 80.5 80.7 82.3 82.5 88.8 94. ]]

Шаг 1. Найти размах вариации

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

определим максимальное и минимальное значение имеющихся значений:  $x_{min} = 14.0$ ;  $x_{max} = 94.0$

$$R = x_{max} - x_{min} = 94.0 - 14.0 = 80.0$$

Шаг 2. Найти оптимальное количество интервалов

Скобка  $\lfloor \rfloor$  означает целую часть (округление вниз до целого числа).

$$k = 1 + \lfloor 3,222 * \lg(N) \rfloor$$

$$k = 1 + \lfloor 3,222 * \lg(100) \rfloor = 1 + \lfloor 6.444 \rfloor = 1 + 6 = 7$$

Шаг 3. Найти шаг интервального ряда

Скобка  $\lceil \rceil$  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:

$$a_0 = x_{min} = 14.0$$

$$a_i = a_0 + i * h = 14.0 + i * 12, i = 1, \dots, 7$$

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

$[a_{i-1}; a_i): [14.0; 26.0); [26.0; 38.0); [38.0; 50.0); [50.0; 62.0); [62.0; 74.0); [74.0; 86.0);$   
 $[86.0; 98.0)$

- построим гистограмму относительных частот;

Найти частоты

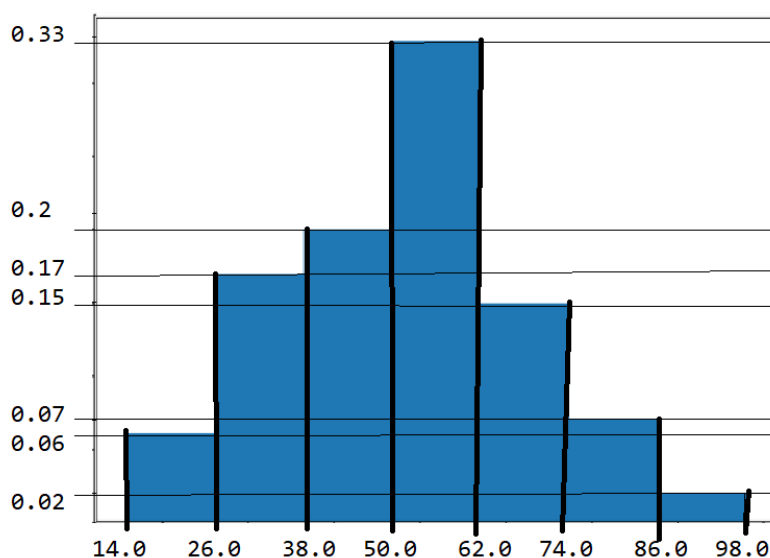
$f_i$  – число попаданий значений признака в каждый из интервалов  $[a_{i-1}, a_i)$

$f_i = n_i, n_i$  – 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:

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

$[a_{i-1}; a_i)$	[14.0, 26.0)	[26.0, 38.0)	[38.0, 50.0)	[50.0, 62.0)	[62.0, 74.0)	[74.0, 86.0)	[86.0, 98.0)
$n_i$	6	17	20	33	15	7	2
$n$	100	100	100	100	100	100	100
$w_i$	0.06	0.17	0.2	0.33	0.15	0.07	0.02



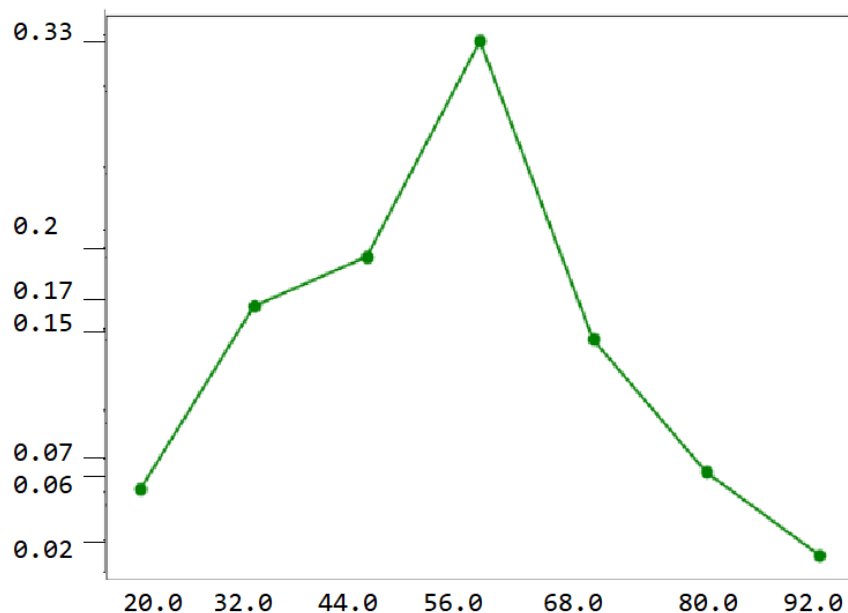
- Перейдем от составленного интервального распределения к точечному выборочному распределению, взяв за значение признака середины частичных интервалов.

$x_i$	20.0	32.0	44.0	56.0	68.0	80.0	92.0
$n_i$	6	17	20	33	15	7	2
$n$	100	100	100	100	100	100	100
$w_i$	0.06	0.17	0.2	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:

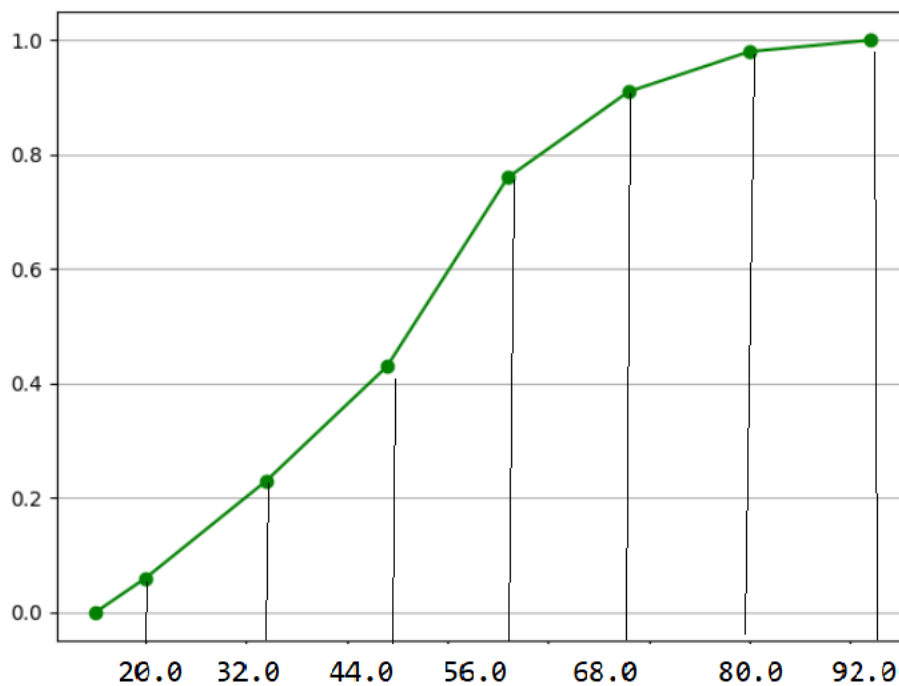
$$x_i = \frac{a_{i-1} + a_i}{2}, i = 1, \dots, 7$$



- найдем эмпирическую функцию распределения и построим ее график;

$F(x)$

$$= \begin{cases} 0.0, & x \leq 20.0, \\ 0.06, & 20.0 < x \leq 32.0, \\ 0.23, & 32.0 < x \leq 44.0, \\ 0.43, & 44.0 < x \leq 56.0, \\ 0.76, & 56.0 < x \leq 68.0, \\ 0.91, & 68.0 < x \leq 80.0, \\ 0.98, & 80.0 < x \leq 92.0, \\ 1.0, & x > 92.0; \end{cases}$$



- 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.  $s$ ;

$$\begin{aligned}\bar{X} &= \sum_{i=1}^7 (w_i * x_i) \\ &= 0.06 * 20.0 + 0.17 * 32.0 + 0.2 * 44.0 + 0.33 * 56.0 + 0.15 * 68.0 \\ &\quad + 0.07 * 80.0 + 0.02 * 92.0 = \\ &= 1.2 + 5.44 + 8.8 + 18.48 + 10.2 + 5.6 + 1.84 = \\ &= 51.56\end{aligned}$$

Выборочная средняя:

$$X_{cp} = \sum_{i=1}^7 (x_i * w_i) = 51.56$$

Выборочная дисперсия:

$$\begin{aligned}D &= \sum_{i=1}^7 (x_i - X_{cp})^2 * w_i = \\ &= (20.0 - 51.56)^2 * 0.06 + (32.0 - 51.56)^2 * 0.17 + (44.0 - 51.56)^2 * 0.2 \\ &\quad + (56.0 - 51.56)^2 * 0.33 + (68.0 - 51.56)^2 * 0.15 + (80.0 - 51.56)^2 \\ &\quad * 0.07 + (92.0 - 51.56)^2 * 0.02 = \\ &= 272.6064\end{aligned}$$

Исправленная выборочная дисперсия

$$S^2 = \frac{N}{N-1} * D = \frac{100}{99} * 272.6064 \approx 275.36$$

Выборочное среднее квадратичное отклонение:

$$\sigma = \sqrt{D} = \sqrt{272.6064} \approx 16.5108$$

исправленное выборочное с. к. о  $s$

$$s = \sqrt{S^2} \approx \sqrt{275.36} \approx 16.594$$

- 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  $X$

$x_i$	$y_i$	$x_i \cdot y_i$	$x_i^2$	$y_i^2$
54.20	14.00	758.80	2937.64	196.00
58.00	25.00	1450.00	3364.00	625.00
45.00	49.00	2205.00	2025.00	2401.00
46.00	25.50	1173.00	2116.00	650.25
62.20	50.00	3110.00	3868.84	2500.00
63.30	48.00	3038.40	4006.89	2304.00
88.80	46.50	4129.20	7885.44	2162.25
46.00	59.00	2714.00	2116.00	3481.00
80.50	53.00	4266.50	6480.25	2809.00
62.30	52.70	3283.21	3881.29	2777.29
<u>Сумма</u>	606.30	422.70	26128.11	38681.35

$$X = [54.2 \ 58. \ 45. \ 46. \ 62.2 \ 63.3 \ 88.8 \ 46. \ 80.5 \ 62.3]$$

$$Y = [14. \ 25. \ 49. \ 25.5 \ 50. \ 48. \ 46.5 \ 59. \ 53. \ 52.7]$$

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  $Y$ .

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

$$\frac{\frac{26128.11}{10} - \frac{606.3}{10} \cdot \frac{422.7}{10}}{\sqrt{\frac{38681.35}{10} - \left(\frac{606.3}{10}\right)^2} \cdot \sqrt{\frac{19905.79}{10} - \left(\frac{422.7}{10}\right)^2}} = 0.2526$$

2) Составим выборочное уравнение прямой регрессии  $Y$  на  $X$

2) [линейное уравнение регрессии](#)  $Y$  на  $X$ :

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

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 60.63$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 42.27$$

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

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

$$\bar{\mu}_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} = -253670.199$$



$$y_x = 0.2601 * x + 26.4983$$

$$r_{xy} = 0.2526$$