

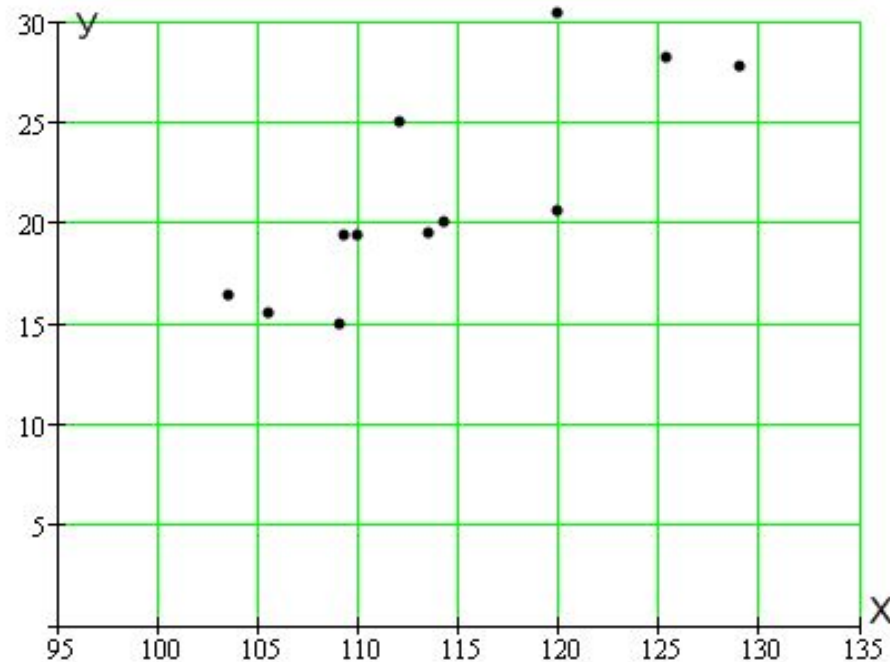
# ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННОГО АНАЛИЗА



## Пример 1.

Исследование связи между IQ и успеваемостью по математике 12 школьников

| №                | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| IQ (X)           | 120 | 112 | 110 | 120 | 103 | 126 | 113 | 114 | 106 | 108 | 128 | 109 |
| Оценка теста (Y) | 31  | 25  | 19  | 21  | 17  | 28  | 18  | 20  | 16  | 15  | 27  | 19  |



# § 1. Коэффициент корреляции Пирсона.

Соотношение между коэффициентами корреляции и типами шкал исследуемых переменных.

| Типы шкал                              |  | Мера связи                        |
|--|--|-----------------------------------|
| <i>Переменная X</i>                    | <i>Переменная Y</i>                    |                                   |
| Интервальная (или отношений)           | Интервальная (или отношений)           | Коэффициент Пирсона               |
| Ранговая, интервальная (или отношений) | Ранговая, интервальная (или отношений) | Коэффициент Спирмена              |
| Ранговая                               | Ранговая                               | Коэффициент Кендалла              |
| Дихотомическая                         | Дихотомическая                         | Коэффициент $\phi$                |
| Ранговая, интервальная                 | Дихотомическая                         | Точечный бисериальный коэффициент |

Ковариация  $X$  и  $Y$

$$S_{XY} = \frac{\sum (x_i - \bar{x}_B)(y_i - \bar{y}_B)}{n-1}$$

Коэффициент корреляции Пирсона

$$r_{XY} = \frac{S_{XY}}{S_x S_y}$$

Расчетная формула для вычисления коэффициента корреляции Пирсона

$$r_{XY} \in [-1; 1]$$

$$r_{XY} = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \cdot \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}$$

## Интерпретация значений коэффициента $r_{XY}$

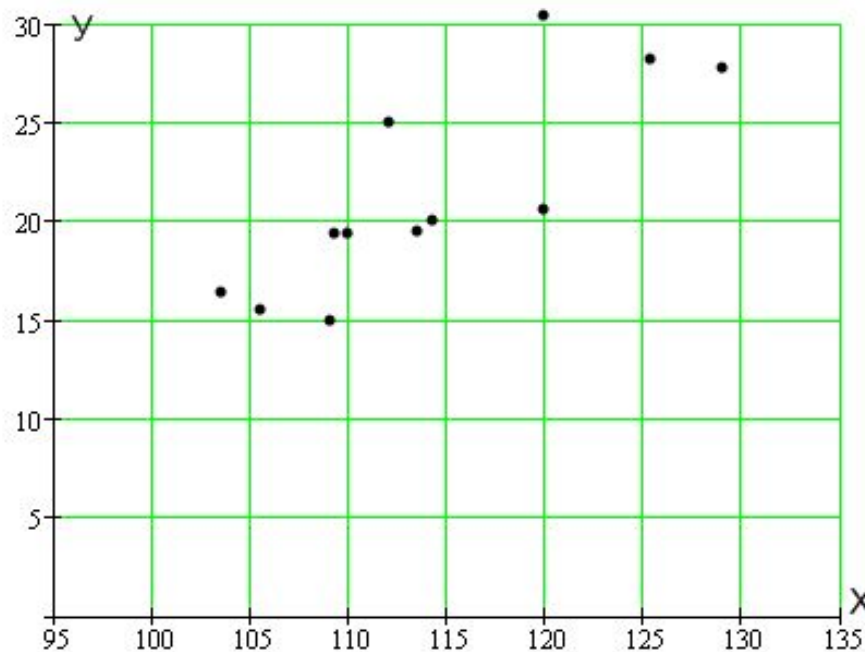
| $r_{XY}$       | Связь между переменными |
|----------------|-------------------------|
| $[-1; -0,9)$   | Очень сильная обратная  |
| $[-0,9; -0,7)$ | Сильная обратная        |
| $[-0,7; -0,5)$ | Средняя обратная        |
| $[-0,5; -0,2)$ | Слабая обратная         |
| $[-0,2; 0)$    | Очень слабая обратная   |
| 0              | Нет связи               |
| $(0; 0,2]$     | Очень слабая прямая     |
| $(0,2; 0,5]$   | Слабая прямая           |
| $(0,5; 0,7]$   | Средняя прямая          |
| $(0,7; 0,9]$   | Сильная прямая          |
| $(0,9; 1]$     | Очень сильная прямая    |

## Пример 1 (продолжение).

| $N$ | $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i \cdot y_i$ |
|-----|-------|-------|---------|---------|-----------------|
| 1   | 120   | 31    | 14400   | 961     | 3720            |
| 2   | 112   | 25    | 12544   | 625     | 2800            |
| 3   | 110   | 19    | 12100   | 361     | 2090            |
| 4   | 120   | 21    | 14400   | 441     | 2520            |
| 5   | 103   | 17    | 10609   | 289     | 1751            |
| 6   | 126   | 28    | 15876   | 784     | 3528            |
| 7   | 119   | 18    | 14161   | 324     | 2142            |
| 8   | 114   | 20    | 12996   | 400     | 2280            |
| 9   | 106   | 16    | 11236   | 256     | 1696            |
| 10  | 108   | 15    | 11664   | 225     | 1620            |
| 11  | 128   | 27    | 16384   | 729     | 3456            |
| 12  | 109   | 19    | 11881   | 361     | 2071            |
|     | 1375  | 256   | 158251  | 5756    | 29674           |

$$\begin{aligned}
 r_{XY} &= \\
 &= \frac{12 \cdot 29674 - 1375 \cdot 256}{\sqrt{[12 \cdot 158251 - (1375)^2] \cdot [12 \cdot 5756 - (256)^2]}} \approx \\
 &\approx 0,750674
 \end{aligned}$$

**Вывод:** Между уровнем IQ и успеваемостью по математике существует **сильная прямая** связь.



## §2. Коэффициент корреляции $\varphi$ .

Таблица сопряженности признаков 2\*2:

| Значения признаков |   | Признак X |       | Итого |
|--------------------|---|-----------|-------|-------|
|                    |   | 0         | 1     |       |
| Признак Y          | 1 | $a$       | $b$   | $a+b$ |
|                    | 0 | $c$       | $d$   | $c+d$ |
| Итого              |   | $a+c$     | $b+d$ | $n$   |

$$\varphi = \frac{bc - ad}{\sqrt{(a + c)(b + d)(a + b)(c + d)}}$$



**Пример 2.** Наблюдения за 10 студентами 3 курса университета по переменным: совмещение работы и учебы ( $X$ ) и успеваемость ( $Y$ ).

| № студента | $X$ (подрабатывает - 1, нет - 0) | $Y$ (сессия сдана - 1, нет - 0) |
|------------|----------------------------------|---------------------------------|
| 1          | 1                                | 0                               |
| 2          | 1                                | 1                               |
| 3          | 0                                | 1                               |
| 4          | 1                                | 1                               |
| 5          | 0                                | 0                               |
| 6          | 1                                | 0                               |
| 7          | 0                                | 1                               |
| 8          | 0                                | 1                               |
| 9          | 1                                | 1                               |
| 10         | 0                                | 0                               |

| Значения признаков |   | Признак X |   | Итого |
|--------------------|---|-----------|---|-------|
|                    |   | 0         | 1 |       |
| Признак Y          | 1 | 3         | 3 | 6     |
|                    | 0 | 2         | 2 | 4     |
| Итого              |   | 5         | 5 | 10    |

$$\varphi = \frac{3 \cdot 2 - 3 \cdot 2}{\sqrt{5 \cdot 5 \cdot 6 \cdot 4}} = 0$$

**Вывод:** Нет связи между признаками X и Y.

### §3. п.3.1. Коэффициент ранговой корреляции Спирмена $r_S$

$$r_S = 1 - \frac{6}{n^3 - n} \cdot \sum_{i=1}^n (x_i - y_i)^2$$

Пример 3.  $X$  - успеваемость,  $Y$  - поведение.

| №  | $X_i$ | $Y_i$ | $(x_i - y_i)^2$ |
|----|-------|-------|-----------------|
| 1  | 3     | 5     | 4               |
| 2  | 8     | 2     | 36              |
| 3  | 4     | 3     | 1               |
| 4  | 1     | 4     | 9               |
| 5  | 10    | 9     | 1               |
| 6  | 6     | 1     | 25              |
| 7  | 2     | 8     | 36              |
| 8  | 5     | 6     | 1               |
| 9  | 7     | 10    | 9               |
| 10 | 9     | 7     | 4               |
|    |       |       | 126             |

$$n = 10$$

$$r_S = 1 - \frac{6}{10^3 - 10} \cdot 126 \approx 0,24$$

**Вывод:** Слабая прямая связь.

## п.3.2. Тау ( $\tau$ ) Кендалла.

$P$  – общее число совпадений  $Q$  – общее число инверсий

$$\tau = \frac{4P}{n(n-1)} - 1$$

$$\tau = 1 - \frac{4Q}{n(n-1)}$$

**Пример 4.**  $X$  – количество пропущенных занятий,  $Y$  –

рейтинг.

| Студент | $X$ | $Y$ | Ранг по $X$ | Ранг по $Y$ | Совпадения | Инверсии |
|---------|-----|-----|-------------|-------------|------------|----------|
| A       | 5   | 69  | 1           | 2           | 1          | 1        |
| B       | 4   | 66  | 2           | 3           | 0          | 1        |
| C       | 2   | 78  | 3           | 1           | 0          | 0        |
|         |     |     |             |             | $P=1$      | $Q=2$    |

$$\tau = \frac{4}{3(3-1)} - 1 = -\frac{1}{3}$$