

# ЛЕКЦИЯ № 7

КОРРЕЛЯЦИОННО-  
РЕГРЕССИОННЫЙ  
АНАЛИЗ.

МОДЕЛИРОВАНИЕ  
СВЯЗЕЙ.

# § 1. ОСНОВНЫЕ ПОНЯТИЯ

**Функциональные  
(жёстко детерминированные)**

С изменением значений одной переменной ( $x$ ) другая переменная ( $y$ ) изменяется определённым образом (по определённому закону)

$$y = f(x_1, x_2, \dots, x_n)$$

**Статистические  
(стохастически детерминированные)**

Связь между величинами, при которой с изменением значений одной переменной вторая с определённой вероятностью принимает значения в определённых границах, а другие статистические величины, например, средние, могут изменяться по определённому закону

$$\tilde{y} = f(x_1, x_2, \dots, x_n)$$

Корреляционная связь

# Классификация корреляционных связей



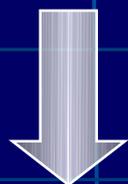


ПРЯМАЯ



$$x_1 > x_2 \Rightarrow \tilde{y}_1 > \tilde{y}_2$$

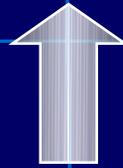
$$x_1 > x_2 \Rightarrow \tilde{y}_1 < \tilde{y}_2$$



ОБРАТНАЯ



ПРЯМОЛИНЕЙНАЯ

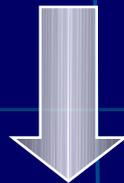


ЗАВИСИМОСТЬ ВЫРАЖЕНА  
УРАВНЕНИЕМ ЛИНЕЙНОЙ  
ЗАВИСИМОСТИ:

$$\tilde{y} = a + b \cdot x$$

---

ЗАВИСИМОСТЬ ВЫРАЖЕНА  
УРАВНЕНИЕМ ДРУГОЙ ФУНКЦИИ



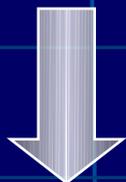
КРИВОЛИНЕЙНАЯ

ОДИН ФАКТОР



$$\tilde{y} = f(x)$$

$$\tilde{y} = f(x_1, x_2, \dots, x_n)$$



МНОГО  
ФАКТОРОВ

## **§ 2. Статистические методы моделирования связей**

- **метод сопоставления параллельных рядов;**

- **метод аналитических группировок;**

- **корреляционный анализ;**

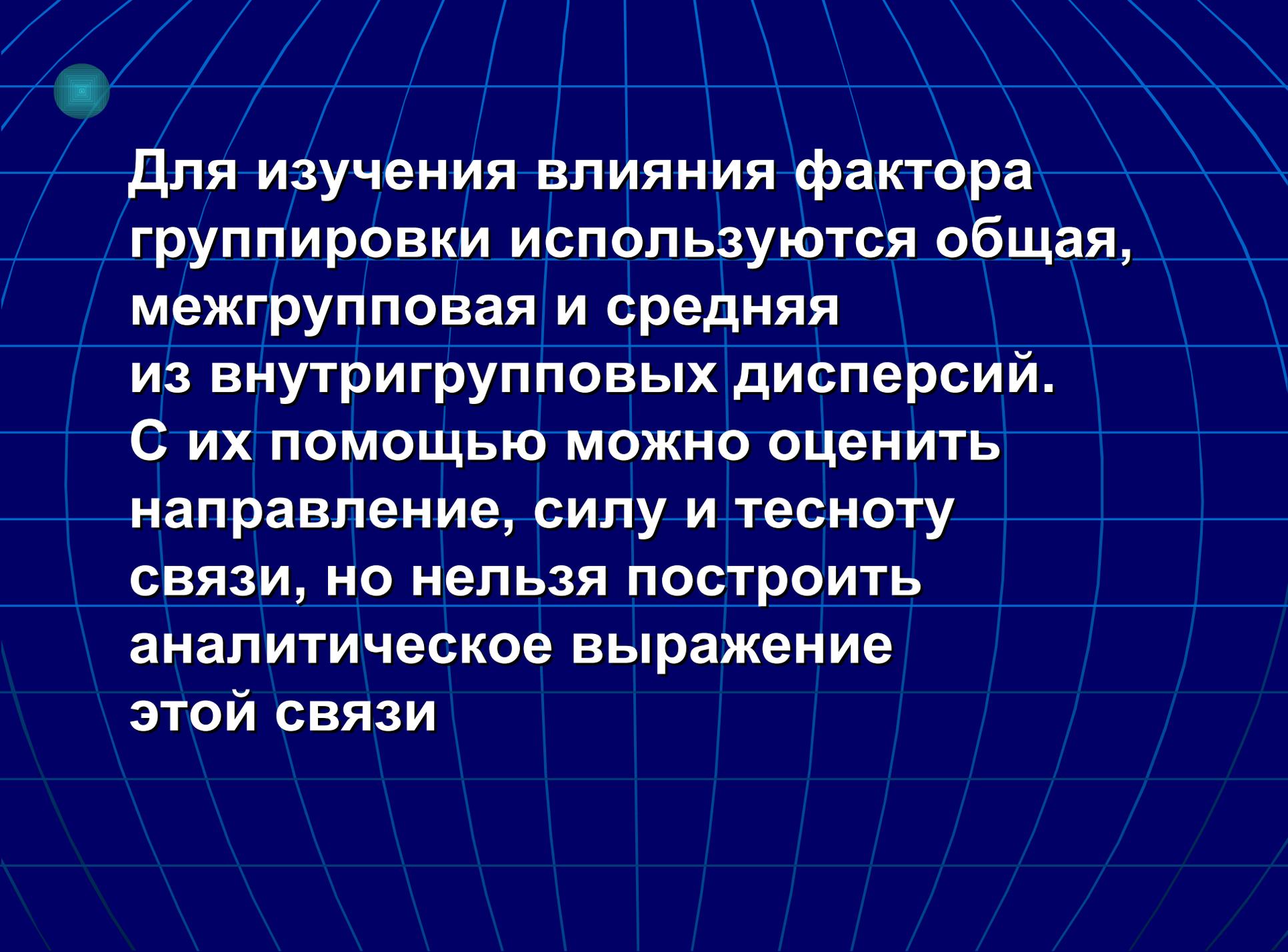
- **регрессионный анализ;**

- **некоторые непараметрические методы (для оценки связи атрибутивных признаков).**



**Позволяет установить наличие связи между признаками, направление связи.**

**Для этого факторы располагают по возрастанию или убыванию и прослеживают изменение величины результата.**



**Для изучения влияния фактора группировки используются общая, межгрупповая и средняя из внутригрупповых дисперсий. С их помощью можно оценить направление, силу и тесноту связи, но нельзя построить аналитическое выражение этой связи**

• **Корреляционный анализ** измеряет тесноту известной связи между факторами и результатом, оценивает факторы, оказывающие наибольшее влияние.

• **Регрессионный анализ** осуществляет выбор модели связи, определяет расчётные значения функции, устанавливает степень влияния признаков

# **Корреляционно -регрессионный анализ**

**осуществляет построение  
аналитического выражения  
зависимости признаков,  
оценивает это аналитическое  
выражение, оценивает  
существующие между факторами  
и результатом связи, рассчитывает  
теоретические значения функции**

# САМОСТОЯТЕЛЬНО ИЗУЧИТЬ:

- НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ  
ОЦЕНКИ СВЯЗИ МЕЖДУ  
АТТРИБУТИВНЫМИ  
ПРИЗНАКАМИ;
- МЕТОДЫ ОЦЕНКИ  
АДЕКВАТНОСТИ УРАВНЕНИЙ  
ПАРНОЙ КОРРЕЛЯЦИИ.

**УРАВНЕНИЕ КОРРЕЛЯЦИИ ЭТО-  
математическое выражение связи  
признаков, которое представляет  
собой приближение (аппроксимацию)  
изменений условной средней  
величины результативного признака  
с изменением факторов**

$$\tilde{y} = f(x_1; x_2, \dots, x_n)$$

# УРАВНЕНИЯ ПАРНОЙ КОРРЕЛЯЦИИ

$$1. \tilde{y} = a + b \cdot x$$

$$2. \tilde{y} = a + b \cdot x + c \cdot x^2$$

$$3. \tilde{y} = a + b \cdot \log x$$

$$4. \tilde{y} = a + \frac{b}{x} \quad \text{u m.đ.}$$

Для определения параметров  $a$  и  $b$  в уравнении линейной парной корреляции применяют метод наименьших квадратов решают систему линейных неоднородных уравнений:

$$\begin{cases} n * a + \sum x_i = \sum y_i \\ a * \sum x_i + b * \sum x_i^2 = \sum x_i * y_i \end{cases}$$

$$b = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$a = \bar{y} - b * \bar{x}$$

# Экономический смысл параметров уравнения линейной корреляции

*Вариация результата*

*на*

*единицу вариации*

*фактора*

***b***

*Минимальное значение фактора,  
при котором  
возможно изменение результата*

***a***  
***b***

**Последовательность  
выполнения корреляционного  
и регрессионного анализа**

- 1. Графическое изображение исходных данных в виде ломаной линии для выбора типа модели;**
- 2. Расчёт параметров в аналитическом выражении типа модели (уравнении);**
- 3. Проверка адекватности (соответствия фактическим данным) построенной модели;**
- 4. Оценка силы и тесноты связи с помощью коэффициентов.**

**Коэффициенты  
для определения  
силы, тесноты и  
направления  
связи**

Коэффициент детерминации:  $\eta = \sqrt{1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y_i - \bar{y})^2}}$ ,

подкоренное выражение показывает долю вариации результативного признака под влиянием вариации фактора.  
Линейный коэффициент корреляции:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y}, \text{ где}$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}; \sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}$$

# Шкала Чеддока

| $ r_{xy} $    | 0,1-0,3 | 0,3-0,5   | 0,5-0,7  | 0,7-0,9 | 0,9-0,99         |
|---------------|---------|-----------|----------|---------|------------------|
| сила<br>связи | слабая  | умеренная | заметная | сильная | очень<br>сильная |

**ПРОВЕРКА  
АДЕКВАТНОСТИ  
ЛИНЕЙНОЙ  
РЕГРЕССИОННОЙ  
МОДЕЛИ**

**Состоит в проверке  
равенства коэффициентов:**

- **детерминации;**

- **линейного коэффициента**

**корреляции**

**Состоит в проверке  
равенства коэффициентов:  
детерминации;  
линейного коэффициента  
корреляции**

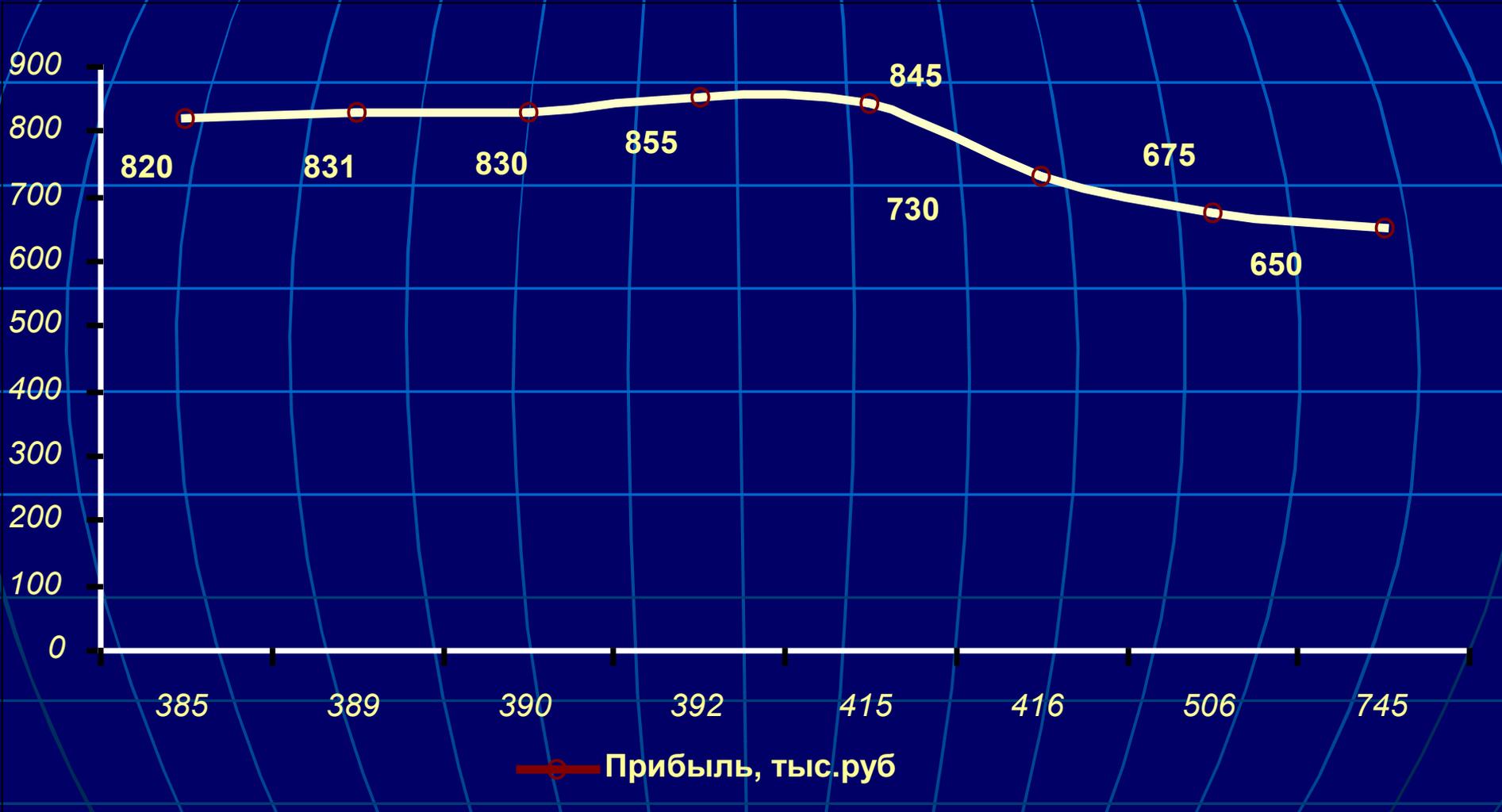
# Пример

| № пр-ия | Затраты на выпуск продукции, тыс.руб | Прибыль, тыс.руб |
|---------|--------------------------------------|------------------|
| 1       | 385                                  | 820              |
| 2       | 389                                  | 831              |
| 3       | 390                                  | 830              |
| 4       | 392                                  | 855              |
| 5       | 415                                  | 845              |
| 6       | 416                                  | 730              |
| 7       | 506                                  | 675              |
| 8       | 745                                  | 650              |

# Пример

Найти корреляционную зависимость признаков, оценить направление, силу и тесноту связи признаков.

# ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ ЗАВИСИМОСТИ ПРИБЫЛИ ОТ ЗАТРАТ



Рассчитаем параметры уравнения линейной парной корреляции, проверим адекватность построенной модели, оценим силу, тесноту и направление связи:

| № п/п | Затраты на выпуск продукции, тыс.руб | Прибыль, тыс.руб | $x_i - \bar{x}$ | $y_i - \bar{y}$ | $(x_i - \bar{x}) * (y_i - \bar{y})$ | $(x_i - \bar{x})^2$ |
|-------|--------------------------------------|------------------|-----------------|-----------------|-------------------------------------|---------------------|
| 1     | 385                                  | 820              | -69,75          | 40,5            | -2824,875                           | 4865,0625           |
| 2     | 389                                  | 831              | -65,75          | 51,5            | -3386,125                           | 4323,0625           |
| 3     | 390                                  | 830              | -64,75          | 50,5            | -3269,875                           | 4192,5625           |
| 4     | 392                                  | 855              | -62,75          | 75,5            | -4737,625                           | 3937,5625           |
| 5     | 415                                  | 845              | -39,75          | 65,5            | -2603,625                           | 1580,0625           |
| 6     | 416                                  | 730              | -38,75          | -49,5           | 1918,125                            | 1501,5625           |
| 7     | 506                                  | 675              | 51,25           | -104,5          | -5355,625                           | 2626,5625           |
| 8     | 745                                  | 650              | 290,25          | -129,5          | -37587,375                          | 84245,0625          |
|       | 3638                                 | 6236             |                 |                 | -57847                              | 107271,5            |
|       |                                      |                  |                 |                 |                                     |                     |

$$\bar{x} = \frac{3638}{8} = 454,75; \bar{y} = \frac{6236}{8} = 779,5$$

$$\tilde{y} = \frac{\sum(x_i - \bar{x}) * y_i}{\sum(x_i - \bar{x})^2} = 1024,728 - 0,539258 \cdot x$$

$$b = -0,53926$$

$$a = 1024,728$$

| № пр-ия | Затраты на выпуск продукции, тыс.руб | Прибыль, тыс.руб | $(x_i - \bar{x})^2$ | $(y_i - \bar{y})^2$ | $\tilde{y}_i$ | $(\tilde{y}_i - \bar{y})^2$ | $(\tilde{y}_i - y_i)^2$ |
|---------|--------------------------------------|------------------|---------------------|---------------------|---------------|-----------------------------|-------------------------|
| 1       | 385                                  | 820              | 4865,0625           | 1640,25             | 817,11325     | 1414,7562                   | 8,333351543             |
| 2       | 389                                  | 831              | 4323,0625           | 2652,25             | 814,95621     | 1257,1431                   | 257,4030853             |
| 3       | 390                                  | 830              | 4192,5625           | 2550,25             | 814,41696     | 1219,1938                   | 242,8312759             |
| 4       | 392                                  | 855              | 3937,5625           | 5700,25             | 813,33844     | 1145,0400                   | 1735,685623             |
| 5       | 415                                  | 845              | 1580,0625           | 4290,25             | 800,93551     | 459,4809                    | 1941,679676             |
| 6       | 416                                  | 730              | 1501,5625           | 2450,25             | 800,39625     | 436,6532                    | 4955,631662             |
| 7       | 506                                  | 675              | 2626,5625           | 10920,25            | 751,86303     | 763,8022                    | 5907,924996             |
| 8       | 745                                  | 650              | 84245,063           | 16770,25            | 622,98037     | 24498,3960                  | 730,0606485             |
|         | 3638                                 | 6236             | 107271,5            | 46974               | 6236          | 31194,46537                 | 15779,55032             |

$$b = -0,5392$$

$$a = 1024,728$$

$$\sigma_x = 115,797$$

$$\sigma_y = 76,627$$

$$\eta = 0,815$$

$$r_{xy} = 0,815$$

$$\eta^2 \cdot 100\% = 66,41\%$$







# ВЫВОДЫ

1. Для изучения влияния фактора на результат в статистике применяют парную линейную корреляцию и методы корреляционного и регрессионного анализа;
2. Для оценки тесноты связи используют коэффициент детерминации;
3. Для оценки силы связи используют коэффициент корреляции или теоретическое корреляционное отношение и шкалу Чеддока;
4. Для уравнения парной линейной корреляции если  $b > 0$ , то связь - прямая, если  $b < 0$ , то связь - обратная;
5. Применение корреляционно-регрессионного анализа начинается с построения графического изображения зависимости и выбора типа связи.