

# **СТАТИСТИКА.**

---

---

## **Описательная статистика.**

---

---

### **Лекция 2. Показатели вариации и способы их вычисления.**

---

---

**Автор: Равичев Л.В.**

**РХТУ им. Д.И.Менделеева**

**Кафедра управления технологическими инновациями**

**Москва - 2013**

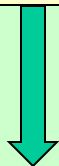
# Размах (амплитуда) колебаний

*Размах (амплитуда) колебаний (размах вариации)* - это разность между наименьшей и наибольшей вариантой.

$$R = X_{\max} - X_{\min}$$

Пример. Даны два ряда набора чисел:

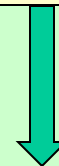
6, 10, 14, 26, 34



$$x_1 = \frac{90}{5} = 18$$

$$R_1 = 34 - 6 = 28$$

14, 16, 18, 20, 22.



$$x_2 = \frac{90}{5} = 18$$

$$R_2 = 22 - 14 = 8$$

# Квартильное отклонение

*Квартильное отклонение* применяется вместо размаха вариации, чтобы избежать недостатков, связанных с использованием крайних значений.

$$d_k = \frac{Q_3 - Q_1}{2}$$

где  $Q_1$  и  $Q_3$  – соответственно третья и первая квартили распределения.

# Среднее линейное отклонение

Для несгруппированных данных:

$$\bar{d} = \pm \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

где

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Взвешенное линейное отклонение:

$$\bar{d} = \pm \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i}{\sum_{i=1}^n f_i}$$

где

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n}$$

# Среднее квадратическое отклонение

**Простое квадратическое отклонение:**

$$\sigma = \pm \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

где

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

**Взвешенное квадратическое отклонение:**

$$\sigma = \pm \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}}$$

где

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n}$$

# Среднее квадратическое отклонение

**Пример.** Имеются следующие данные о распределении кип шерсти по весу при отгрузке:

| Вес одной кипы, кг. | Количество отгруженных кип, шт. |
|---------------------|---------------------------------|
| 86                  | 10                              |
| 90                  | 20                              |
| 94                  | 10                              |
| 96                  | 30                              |
| 100                 | 15                              |
| 110                 | 15                              |
| <b>ИТОГО</b>        | <b>100</b>                      |

Требуется определить среднюю арифметическую простую и взвешенную, среднее квадратическое отклонение простое и взвешенное.



# Среднее квадратическое отклонение (простое)

1. Средний вес одной кипы:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{86 + 90 + 94 + 96 + 100 + 110}{6} = 96 \text{ кг}$$

2. Среднее квадратическое простое отклонение:

Данные для расчета квадратичного отклонения

| Вес одной кипы,<br>кг. | Отклонение от<br>среднего значения | Квадраты<br>отклонений |
|------------------------|------------------------------------|------------------------|
| 86                     | -10                                | 100                    |
| 90                     | -6                                 | 36                     |
| 94                     | -2                                 | 4                      |
| 96                     | 0                                  | 0                      |
| 100                    | 4                                  | 16                     |
| 110                    | 14                                 | 196                    |
|                        | -                                  | Сумма = 352            |

$$\sigma = \pm \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$\sigma = \pm \sqrt{\frac{352}{6}} = \pm 7,66 \text{ кг}$$

$$\text{Вес кипы} = 96 \pm 7,66 \text{ кг}$$

# Среднее квадратическое отклонение (взвешенное)

Данные для расчета взвешенного квадратичного отклонения

$$\sigma = \pm \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}}$$

| Вес одной кипы, кг. | Количество | Общий вес, кг. | Отклонение от средней взвешенной | Квадраты отклонений | Произведение квадратов отклонений на количество |
|---------------------|------------|----------------|----------------------------------|---------------------|---|
| 86                  | 10         | 860            | -10,3                            | 106,09              | 1060,9  |
| 90                  | 20         | 1800           | -6,3                             | 39,69               | 793,8   |
| 94                  | 10         | 940            | -2,3                             | 5,29                | 52,9  |
| 96                  | 30         | 2880           | -0,3                             | 0,09                | 2,7   |
| 100                 | 15         | 1500           | 3,7                              | 13,69               | 205,4   |
| 110                 | 15         | 1650           | 13,7                             | 187,69              | 2815,4  |
| Сумма               | 100        | 9630           | -                                | -                   | 4931,1  |



# Среднее квадратическое отклонение (взвешенное)

1. Средний вес одной кипы (взвешенный):

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} = \frac{9630}{100} = 96,3 \text{ кг}$$

2. Среднее квадратическое отклонение (взвешенное):

$$\sigma = \pm \sqrt{\frac{4931,1}{100}} = \pm 7,02 \text{ кг}$$

$$\text{Вес кипы} = 96,3 \pm 7,02 \text{ кг}$$

# Относительные показатели вариации

Коэффициент осцилляции:

$$K_R = \frac{R}{X} 100\%$$

Относительное линейное отклонение:

$$K_d = \frac{d}{X} 100\%$$

Относительный показатель квартильной вариации:

$$K_{d_k} = \frac{d_k}{M_e} 100\%$$



$$K_Q = \frac{Q_3 - Q_1}{2Q_2} 100\%$$

# Коэффициент вариации

*Коэффициент вариации* представляет собой отношение среднего квадратического отклонения к средней арифметической и показывает величину отклонения (в процентах) от средней величины.

**Простое квадратическое отклонение:**

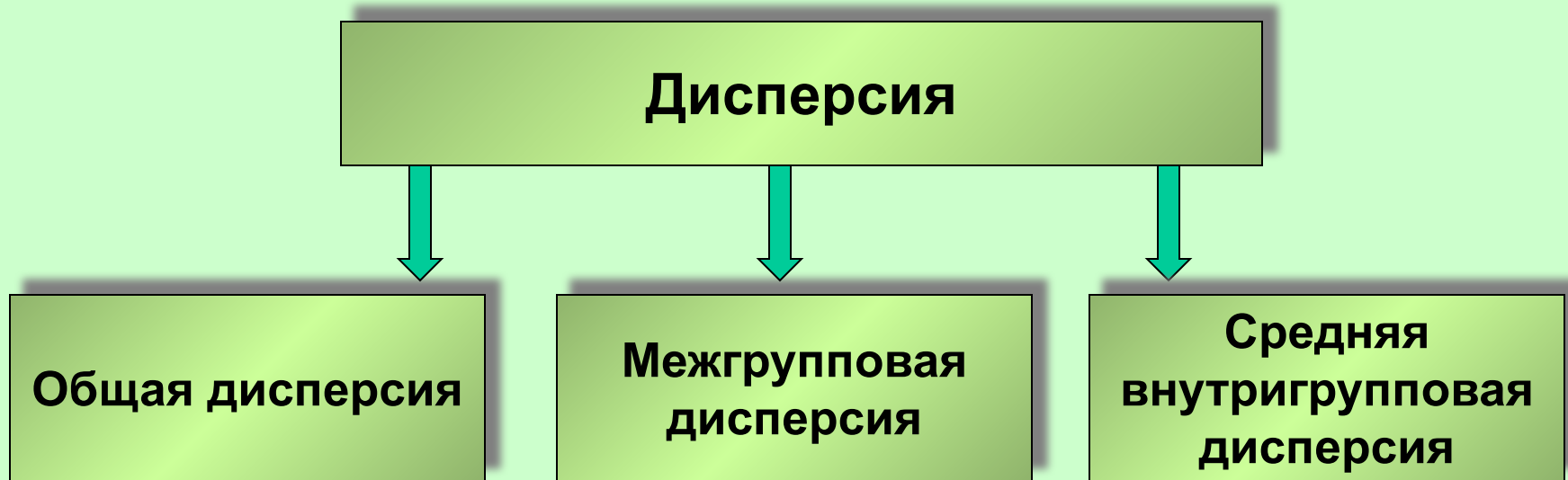
$$V_1 = \frac{\sigma}{x} 100\% \longrightarrow V_1 = \frac{7,66}{96} 100\% = 8,02\%$$

**Взвешенное квадратическое отклонение:**

$$V_2 = \frac{\sigma}{x} 100\% \longrightarrow V_2 = \frac{7,02}{96,3} 100\% = 7,29\%$$

# Дисперсия

*Дисперсия* – это средний квадрат отклонения всех значений признака ряда распределения от средней арифметической



# Общая дисперсия

*Общая дисперсия* характеризует вариацию признака под влиянием всех факторов, формирующих уровень признака у единиц данной совокупности.

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}$$

где  $\bar{x}$  - общая средняя для всей изучаемой совокупности.

# Межгрупповая дисперсия

*Межгрупповая дисперсия* отражает те различия в величине изучаемого признака, которые возникают под влиянием фактора, положенного в основу группировки.

$$D_{mg} = \frac{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2 m_i}{\sum_{i=1}^n m_i}$$

где  $\bar{x}_i$  - средняя по отдельной группе;  $m_i$  - число единиц в определенной группе.



# Средняя внутригрупповая дисперсия

*Средняя внутригрупповая дисперсия* характеризует случайную вариацию, возникающую под влиянием других, неучтенных факторов, и не зависит от условия, положенного в основу группировки.

$$\overline{D}_{вг} = \frac{\sum_{i=1}^n \sigma_i^2 m_i}{\sum_{i=1}^n m_i}$$

где  $\sigma_i^2$  - дисперсия по отдельной группе.

# Правило сложения дисперсий

*Величина общей дисперсии равна сумме межгрупповой дисперсии и средней внутригрупповой дисперсии.*

$$D_x = D_{MG} + \bar{D}_{BG}$$

**Пример №1.** Имеются следующие данные о времени простоя автомобиля под разгрузкой:

|                     |    |    |   |    |    |    |   |    |    |    |
|---------------------|----|----|---|----|----|----|---|----|----|----|
| № пункта разгрузки  | 1  | 2  | 3 | 4  | 5  | 6  | 7 | 8  | 9  | 10 |
| Число грузчиков     | 3  | 4  | 4 | 3  | 3  | 4  | 4 | 4  | 3  | 4  |
| Время простоя, мин. | 12 | 10 | 8 | 15 | 19 | 12 | 8 | 10 | 18 | 8  |

**Проверить закон сложения дисперсий.**

# Правило сложения дисперсий

Решение:

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i}$$

| Время простоя под погрузкой, мин., $x$ | Число выполненных разгрузок, $f$ | $x \cdot f$ | $x - \bar{x}$ | $(x - \bar{x})^2$ | $(x - \bar{x})^2 \cdot f$ |
|--|----------------------------------|-------------|---------------|-------------------|---------------------------|
| 8                                      | 3                                | 24          | -4            | 16                | 48                        |
| 10                                     | 2                                | 20          | -2            | 4                 | 8                         |
| 12                                     | 2                                | 24          | 0             | 0                 | 0                         |
| 15                                     | 1                                | 15          | 3             | 9                 | 9                         |
| 18                                     | 1                                | 18          | 6             | 36                | 36                        |
| 19                                     | 1                                | 19          | 7             | 49                | 49                        |
| Сумма                                  | 10                               | 120         | -             | -                 | 150                       |

$$\bar{x} = \frac{120}{10} = 12 \text{ мин.}$$



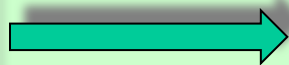
$$\sigma^2 = \frac{150}{10} = 15$$

# Правило сложения дисперсий

Расчет внутригрупповой дисперсии по первой группе  
(число грузчиков, участвующих в разгрузке – 3)

| Время простоя под погрузкой, мин., $x$ | Число выполненных разгрузок, $f$ | $x \cdot f$ | $x - \bar{x}_1$ | $(x - \bar{x}_1)^2$ | $(x - \bar{x}_1)^2 \cdot f$ |
|--|----------------------------------|-------------|-----------------|---------------------|-----------------------------|
| 12                                     | 1                                | 12          | -4              | 16                  | 16                          |
| 15                                     | 1                                | 15          | -1              | 1                   | 1                           |
| 18                                     | 1                                | 18          | 2               | 4                   | 4                           |
| 19                                     | 1                                | 19          | 3               | 9                   | 9                           |
| Сумма                                  | 4                                | 64          | -               | -                   | 30                          |

$$\bar{x} = \frac{64}{4} = 16$$



$$\sigma_1^2 = \frac{30}{4} = 7,5$$

# Правило сложения дисперсий

Расчет внутригрупповой дисперсии по второй группе  
(число грузчиков, участвующих в разгрузке – 4)

| Время простоя под погрузкой, мин., $x$ | Число выполненных разгрузок, $f$ | $x \cdot f$ | $x - \bar{x}_2$ | $(x - \bar{x}_2)^2$ | $(x - \bar{x}_2)^2 \cdot f$ |
|--|----------------------------------|-------------|-----------------|---------------------|-----------------------------|
| 8                                      | 3                                | 24          | -1,33           | 1,77                | 5,31                        |
| 10                                     | 2                                | 20          | 0,67            | 0,45                | 0,90                        |
| 12                                     | 1                                | 12          | 2,67            | 7,13                | 7,13                        |
| Сумма                                  | 6                                | 56          | -               | -                   | 13,34                       |

$$\bar{x}_2 = \frac{56}{6} = 9,33 \text{ мин.}$$



$$\sigma_2^2 = \frac{13,34}{6} = 2,22$$

# Правило сложения дисперсий

**Средняя внутригрупповая дисперсия:**

$$\overline{D}_{вг} = \frac{\sum_{i=1}^n \sigma_i^2 m_i}{\sum_{i=1}^n m_i} = \frac{7,5 \cdot 4 + 2,22 \cdot 6}{10} = 4,33$$

**Межгрупповая дисперсия:**

$$D_{мг} = \frac{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2 m_i}{\sum_{i=1}^n m_i} = \frac{(16 - 12)^2 \cdot 4 + (9,33 - 12)^2 \cdot 6}{10} = 10,67$$

**Общая дисперсия:**

$$D_x = D_{мг} + D_{вг} = 10,67 + 4,33 = 15,00$$



# Правило сложения дисперсий

**Пример №2.** Имеются следующие данные о результатах обследования рабочих предприятия по размеру месячной заработной платы:

| Группы рабочих по возрасту, лет | Число рабочих | Дисперсия заработной платы |
|---------------------------------|---------------|----------------------------|
| До 20                           | 100           | 300                        |
| 20-30                           | 120           | 400                        |
| 30 и старше                     | 150           | 500                        |

**Общая дисперсия заработной платы в обследованной совокупности рабочих составила 450. Определить, в какой степени вариация заработной платы рабочих предприятия зависит от возраста.**

# Правило сложения дисперсий

## Решение.

Средняя внутригрупповая дисперсия характеризует случайную вариацию под влиянием неучтенных факторов:

$$\bar{D}_{вг} = \frac{\sum_{i=1}^n \sigma_i^2 m_i}{\sum_{i=1}^n m_i} = \frac{300 \cdot 100 + 400 \cdot 120 + 500 \cdot 150}{100 + 120 + 150} = 413,5$$

Межгрупповая дисперсия отражает систематическую вариацию под влиянием фактора, положенного в основу группировки (возраста рабочих):

$$D_{мг} = D_x - \bar{D}_{вг} = 450 - 413,5 = 36,5$$

Соотношение дисперсий:

$$\frac{D_{мг}}{D_x} = \frac{36,5}{450} = 0,08 \text{ (8\%)}$$

# Преобразование формулы для расчёта общей дисперсии

Формула для расчёта общей дисперсии может быть преобразована:

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{\sum_{i=1}^n f_i} \xrightarrow{f_i = 1} D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n [x_i^2 - 2x_i\bar{x} + (\bar{x})^2]}{n} = \frac{\sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n(\bar{x})^2}{n}$$

Учитывая, что  $\sum_{i=1}^n x_i = n\bar{x}$ , и разделив полученное выражение на  $n$ , получаем:

$$D_x = \frac{\sum_{i=1}^n x_i^2}{n} - 2(\bar{x})^2 + (\bar{x})^2 = \bar{x}^2 - (\bar{x})^2$$

# Преобразование формулы для расчёта общей дисперсии

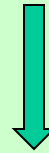
**Пример.** Дисперсия признака равна 600. Объем совокупности равен 10. Сумма квадратов индивидуальных значений признака равна 6250. Найти среднюю величину.

$$D_x = \overline{x^2} - (\overline{x})^2$$



$$\overline{x} = \sqrt{\overline{x^2} - D_x}$$

$$\overline{x^2} = \frac{\sum_{i=1}^n x_i^2}{n} = \frac{6250}{10} = 625$$



$$\overline{x} = \sqrt{625 - 600} = 5$$

# Вариации альтернативного признака

**Альтернативный признак** – качественный признак, имеющий две взаимоисключающие разновидности. Альтернативный признак принимает всего два значения: 1 – наличие признака; 0 – отсутствие признака.

$$p + q = 1$$

где  $p$  - доли единиц, обладающих признаком;  $q$  - доли единиц, не обладающих признаком.

**Среднее значение альтернативного признака:**

$$\bar{x} = \frac{(1 \cdot p) + (0 \cdot q)}{p + q} = p$$

**Дисперсия альтернативного признака:**

$$D_{an} = \frac{(1 - p)^2 \cdot p + (0 - p)^2 \cdot q}{p + q} = p \cdot q$$

# Вариации альтернативного признака

**Пример.** Удельный вес основных рабочих в трех цехах предприятия составил: 80, 75 и 90% общей численности рабочих. Определить дисперсию и среднее квадратическое отклонение доли основных рабочих по предприятию в целом, если численность всех рабочих трех цехов составила соответственно 100, 200 и 150 человек.

**Решение.**

1. Общая численность основных рабочих по предприятию:

$$\sum_{i=1}^n m_{oi} = 100 \cdot 0,8 + 200 \cdot 0,75 + 150 \cdot 0,9 = 365 \text{ чел.}$$

2. Доля основных рабочих по предприятию:

$$p = \frac{\sum_{i=1}^n m_{oi}}{\sum_{i=1}^n m_i} = \frac{365}{100 + 200 + 150} = \frac{365}{450} = 0,811$$

3. Дисперсия альтернативного признака:

$$D_{an} = p \cdot q = p \cdot (1 - p) = 0,811 \cdot (1 - 0,811) = 0,1533$$

4. Среднее квадратическое отклонение:

$$\sigma = \sqrt{D_{an}} = \sqrt{0,1533} = 0,3915$$