

Описательная статистика

Семинар 1

Выборочное среднее значение

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$

Функция в Excel =срзнач()

Выборочная дисперсия

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

Функция в Excel =дисп()

Медиана

- Среднее значение выборки данных при их перечислении в порядке возрастания.
- Если множество содержит нечетное количество элементов, медианой станет значение, для которого количество элементов множества с меньшим значением такое же, как и с большим.
- Если множество содержит четное количество элементов, вы можете просто взять среднее двух расположенных в середине элементов.

**Для симметричного
множества данных:**

**Среднее, медиана
и мода
равны.**

Среднеквадратическое (стандартное) отклонение по выборке

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

Функция в Excel =СТАНДОТКЛОН()

Что можно узнать по среднему значению и стандартному отклонению

Для симметричной гистограммы эмпирическое правило гласит все данные расположены:

- около 68% в диапазоне $\bar{x} - s$ до $\bar{x} + s$
- около 95% в диапазоне
- около 99,7% в диапазоне

Ошибка среднего значения

$$\Delta x = \frac{S}{\sqrt{n}}$$

Например:

- Найдите : Среднее – $2s$ и среднее $+2s$
- Любые данные, лежащие за пределами двукратного стандартного отклонения, называются **резко отклоняющимися**.
- Много полезного можно почерпнуть, выяснив причины появления резко отклоняющихся значений. Любая компания должна добиваться. Чтобы причины «хороших» отклонений реализовывались чаще, а «плохих» - реже.

(100- γ)% доверительный интервал для среднего

Критическое
значение
статистики
Стьюдента

$$x_{теор} = \mu \in \bar{x} \pm \Delta x \cdot t_{cr}(\gamma, n - 1)$$

Теоретическое среднее значение рассматриваемой величины с вероятностью (100- γ)% находится в указанном интервале

Эксцесс (kurtosis)

- мера остроты пика распределения случайной величины
- коэффициент эксцесса нормального распределения равен нулю
- эксцесс положителен, если пик распределения около математического ожидания острый, и отрицателен, если пик гладкий.

Функция в Excel =ЭКСЦЕСС()

Интервал

- **Интервал множества данных – это разность между максимальным и минимальным значением из всех членов множества (измеряется в %).**

Асимметрия

- Асимметрия или коэффициент асимметрии (термин был впервые введен Пирсоном, 1895) является мерой несимметричности распределения.
- Если этот коэффициент отчетливо отличается от 0, распределение является асимметричным.
- Плотность нормального распределения симметрична относительно среднего.

Функция в Excel = СКОС()

Асимметрия

- Асимметричность более, чем $+1$, указывает на высокую степень положительного смещения.
- Асимметричность более, чем -1 , указывает на высокую степень отрицательного смещения
- Асимметричность более, от -1 до $+1$, характеризует относительно симметричное множество данных.