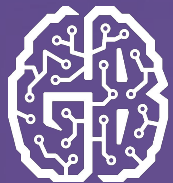


GeekBrains

Теория вероятностей и математическая статистика

Вебинары



GeekBrains

Урок 3

Теория вероятностей и математическая статистика

Описательная статистика. Качественные и количественные
характеристики популяции. Графическое представление данных

На этом уроке мы изучим:

1. Понятия генеральной совокупности и выборки.
2. Математическое ожидание.
3. Дисперсию, среднее квадратичное отклонение. Смещенную и несмещенную оценку дисперсии.
4. Понятия моды, медианы, квартиля, перцентиля, дециля, квантиля.
5. Построение гистограммы, boxplot.

Генеральная совокупность — это множество, которое содержит данные обо всех объектах, соответствующих определенным характеристикам.

**Выборка — это случайным образом
выбранная часть генеральной
совокупности.**

Одно из основных понятий в теории вероятностей — **математическое ожидание**.
Обозначается как **$M(X)$** (в статистике принято обозначение **μ**).

Математическое ожидание — среднее значение случайной величины (распределение вероятностей стационарной случайной величины) при стремлении количества выборок или измерений (иногда говорят — количества испытаний) к бесконечности.

Среднее арифметическое одномерной случайной величины конечного числа испытаний обычно называют оценкой математического ожидания.

$$M(X) = \frac{1}{n} \sum_{i=1}^n x_i$$

Другим важным показателем выборки является **среднее квадратичное отклонение**.

Оно показывает, насколько далеко наблюдения могут быть «разбросаны» относительно среднего значения.

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

Дисперсия равна среднему квадратичному отклонению, возведенному в квадрат. Формула дисперсии (смещенная оценка):

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

Несмещенная оценка дисперсии вычисляется по формуле:

$$\sigma_{\text{несмещ.}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Отличие **несмещенной** оценки дисперсии от **смещенной**
в том, что мы делим сумму квадратов разностей
всех величин со средним значением не на **n** , а на **$n - 1$** .

Мода — наиболее часто встречающееся в выборке значение.

Медиана — значение, которое делит выборку на две части так, что значения, которые меньше медианы, составляют половину (50 %) выборки.

Первый квартиль — значение, которое не превышают 25 % наблюдений в выборке.

Второй квартиль — синоним медианы.

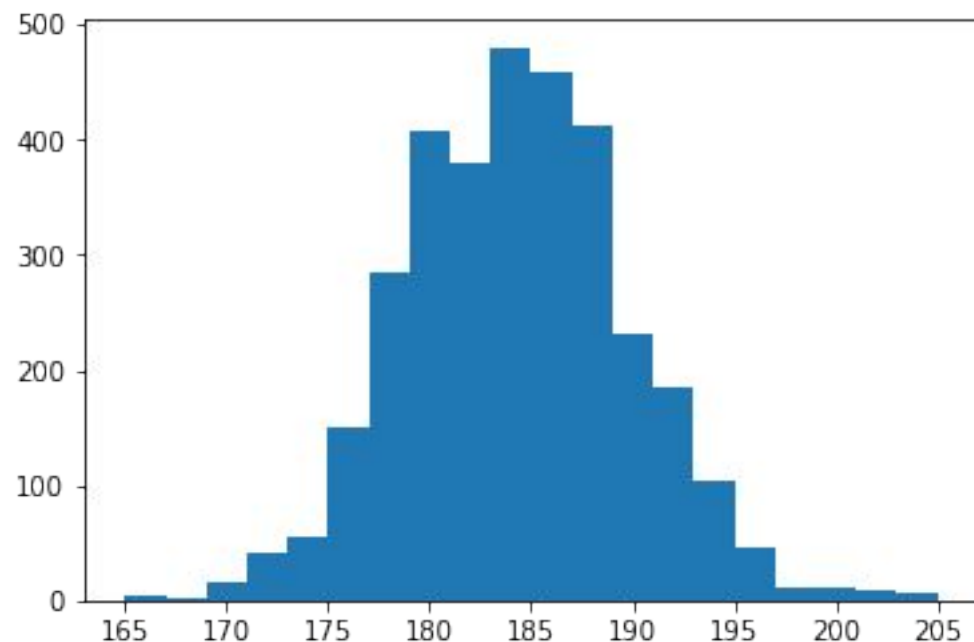
Третий квартиль — значение, которое не превышают 75 % наблюдений в выборке.

Интерквартильное расстояние — отрезок, равный разности третьей и первой квартили.

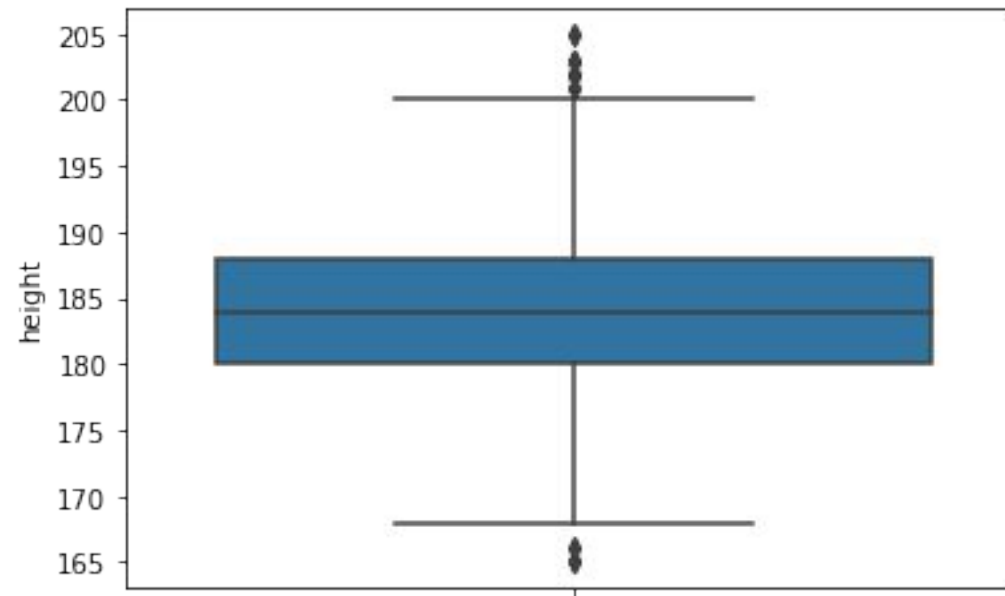
Помимо **квартилей**, в статистике используются **перцентили, децили и квантили**.

Графическое представление данных

Гистограмма



Boxplot



ИТОГИ

1. Генеральная совокупность и выборка.
2. Математическое ожидание.
3. Дисперсия, среднее квадратичное отклонение. Смещенная и несмещенная оценка дисперсии.
4. Мода, медиана, квартиль, перцентиль, дециль, квантиль.
5. Графическое представление данных: гистограмма, boxplot.