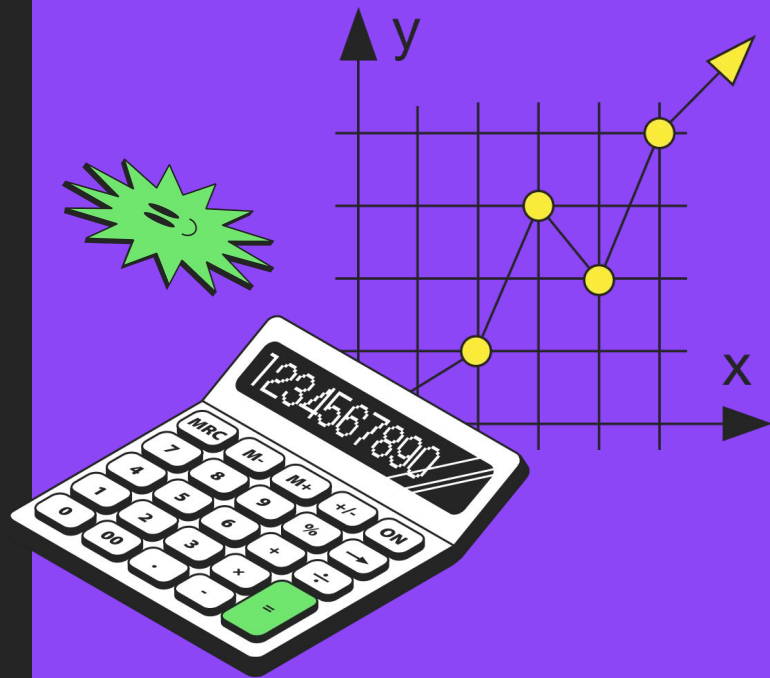


Теория вероятностей и математическая статистика

Урок 3

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



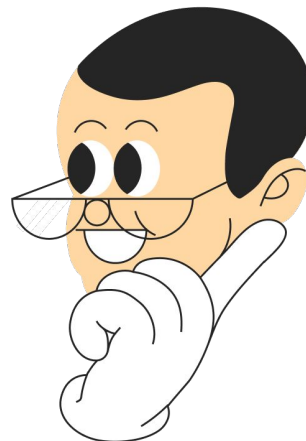


Генеральная совокупность — это множество, которое содержит данные обо всех объектах, соответствующих определённым характеристикам.



Что будет на уроке сегодня

- Понятие генеральной совокупности и выборки
- Математическое ожидание
- Дисперсия, среднее квадратичное отклонение. Смещённая и несмещённая оценка дисперсии
- Понятие моды, медианы, квартиля, перцентиля, дециля, квантиля
- Построение гистограммы, boxplot





Выборка — это случайным образом
выбранная часть генеральной совокупности.



Математическое ожидание

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

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



Оценка математического ожидания

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

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



Квадратичное отклонение

Другим важным показателем выборки считается 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

Дисперсия равна среднему квадратичному отклонению, возведённому в квадрат. Формула дисперсии (смещённая оценка):

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



Несмещённая оценка дисперсии

Вычисляется по формуле:

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



Отличие **несмещённой** оценки дисперсии от **смещённой**
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, делящее 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.



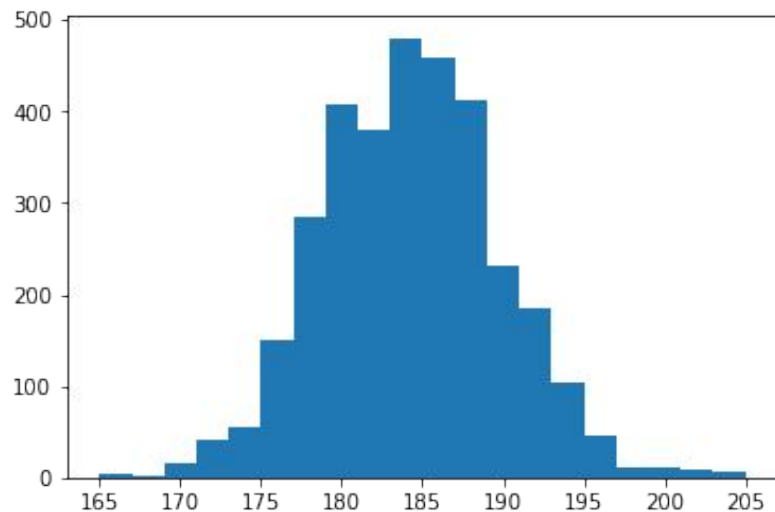
Интерквартильное расстояние — отрезок, равный разности третьей и первой квартили.



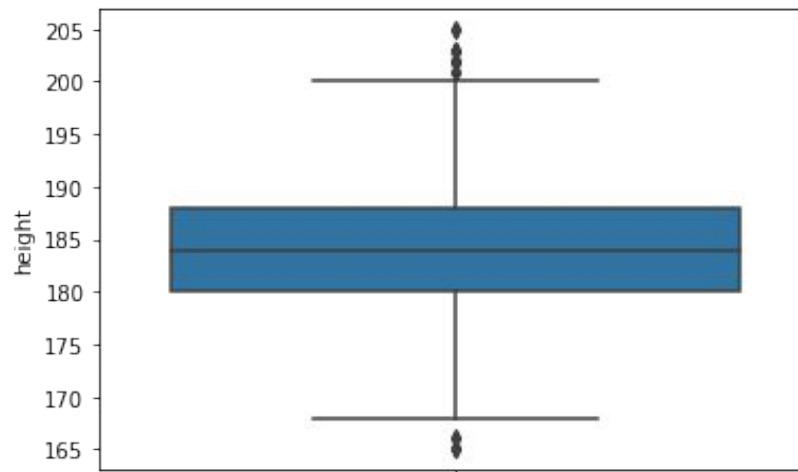
Помимо **квартилей**, в статистике используются **перцентили**, **децили** и **квантили**.

Несмещённая оценка дисперсии

Гистограмма



Boxplot



Заключение

- Генеральная совокупность и выборка
- Математическое ожидание
- Дисперсия, среднее квадратичное отклонение. Смещённая и несмещённая оценка дисперсии
- Мода, медиана, квартиль, перцентиль, дециль, квантиль
- Графическое представление данных: гистограмма, boxplot

