



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ОСНОВЫ АНАЛИЗА ДАННЫХ В МЕЖДУНАРОДНЫХ ОТНОШЕНИЯХ

Лекция 2

Маргарита Бурова

Москва, 2018



ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

Генеральная совокупность — совокупность всех объектов, относительно которых предполагается делать выводы при изучении конкретной задачи



Выборка или выборочная совокупность — множество случаев (испытываемых, объектов, событий, образцов), с помощью определённой процедуры выбранных из генеральной совокупности для анализа

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА





РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРКИ

Репрезентативность — соответствие характеристик выборки характеристикам популяции или генеральной совокупности в целом



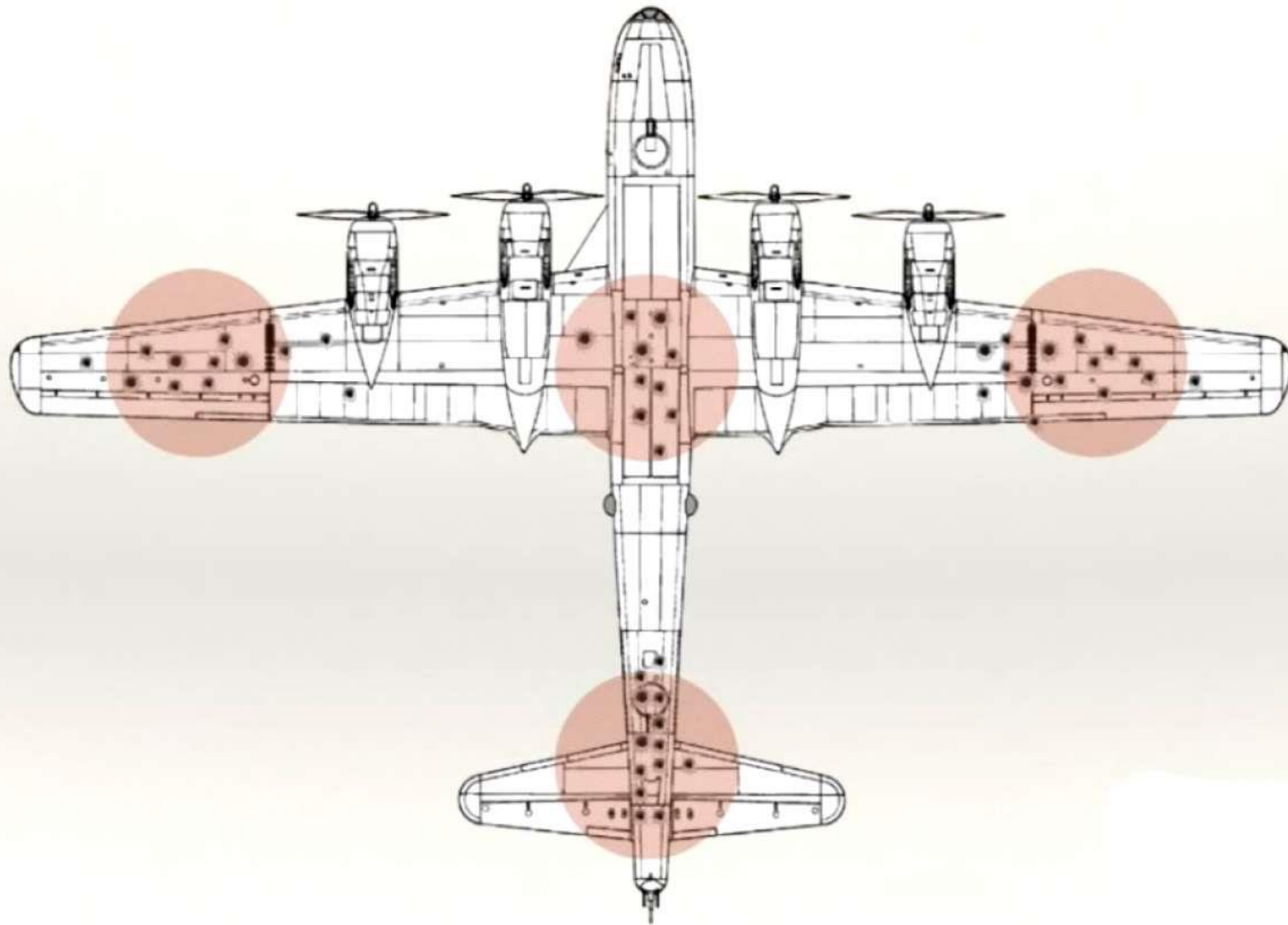
РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРКИ

Рузвельт и Лэндон на выборах 1936 года





ОШИБКА ВЫЖИВШЕГО



ОШИБКА ВЫЖИВШЕГО





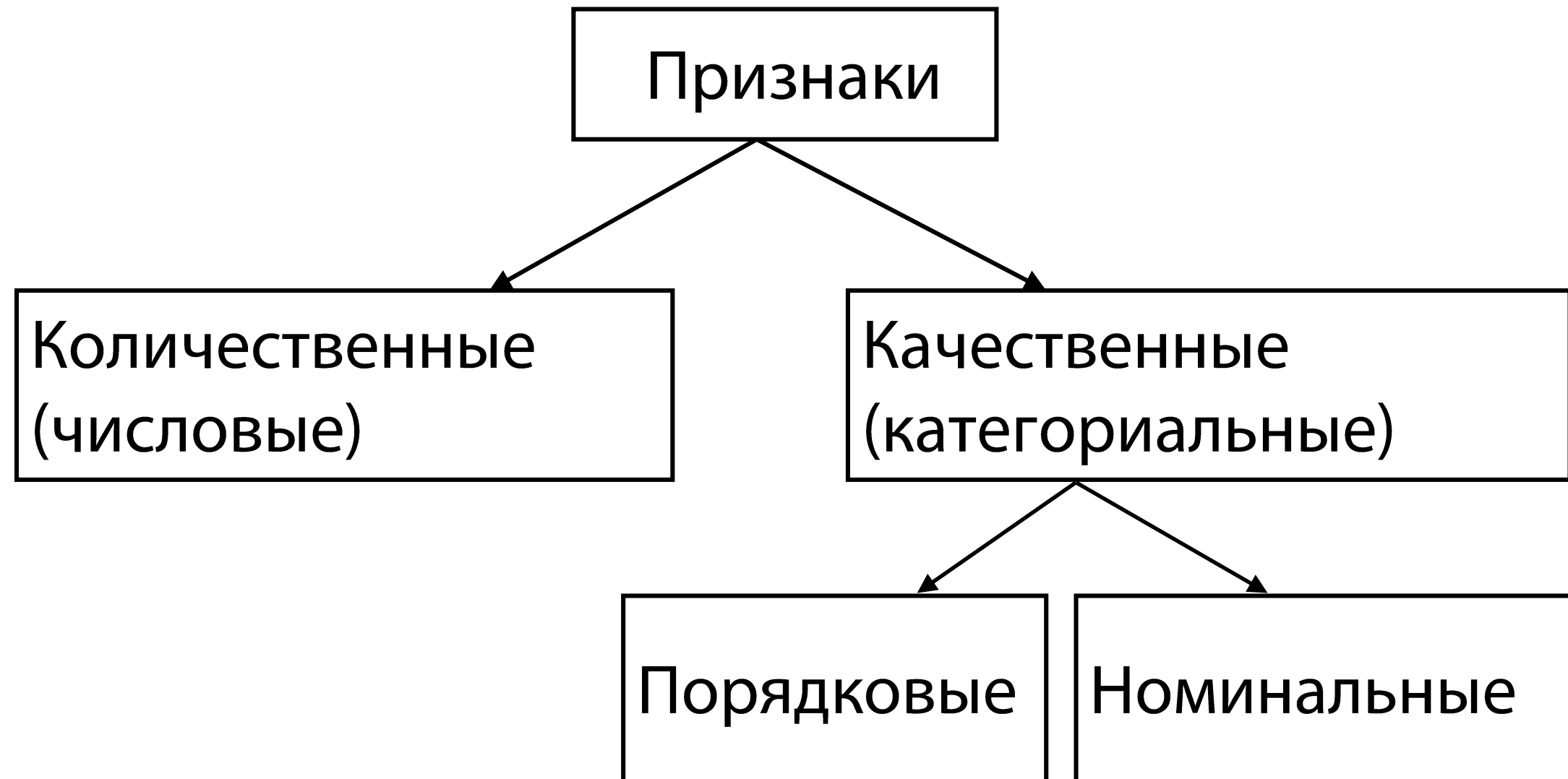
ОБЪЕКТЫ И ПРИЗНАКИ

Выборка состоит из объектов, объекты характеризуются признаками

	Возраст	Город	Уровень образования
Иванов П.А.	24	Санкт-Петербург	Высшее
Петрова К.В.	35	Москва	Кандидат наук
Семенова Н.К.	31	Иваново	Среднее специальное
Сидоров С.О.	28	Сургут	Доктор наук



ТИПЫ ПРИЗНАКОВ





НОМИНАЛЬНЫЕ ПРИЗНАКИ

Качественные признаки, не подлежащие упорядочиванию

Примеры:

- Город
- Темперамент человека
- Группа крови
- Цвет предмета



ПОРЯДКОВЫЕ ПРИЗНАКИ

Качественные признаки, которые могут быть ранжированы в убывающем или восходящем порядке

Примеры:

- Уровень образования
- Степень ожога
- Социально-экономический статус
- Спортивный разряд



КОЛИЧЕСТВЕННЫЕ ПРИЗНАКИ

Признаки, измеряемые с помощью чисел, имеющих содержательный смысл

Примеры:

- Рост
- Вес
- Зарплата



ТИПЫ ПРИЗНАКОВ

	Возраст	Город	Уровень образования
Иванов П.А.	24	Санкт-Петербург	Высшее
Петрова К.В.	35	Москва	Кандидат наук
Семенова Н.К.	31	Иваново	Среднее специальное
Сидоров С.О.	28	Сургут	Доктор наук

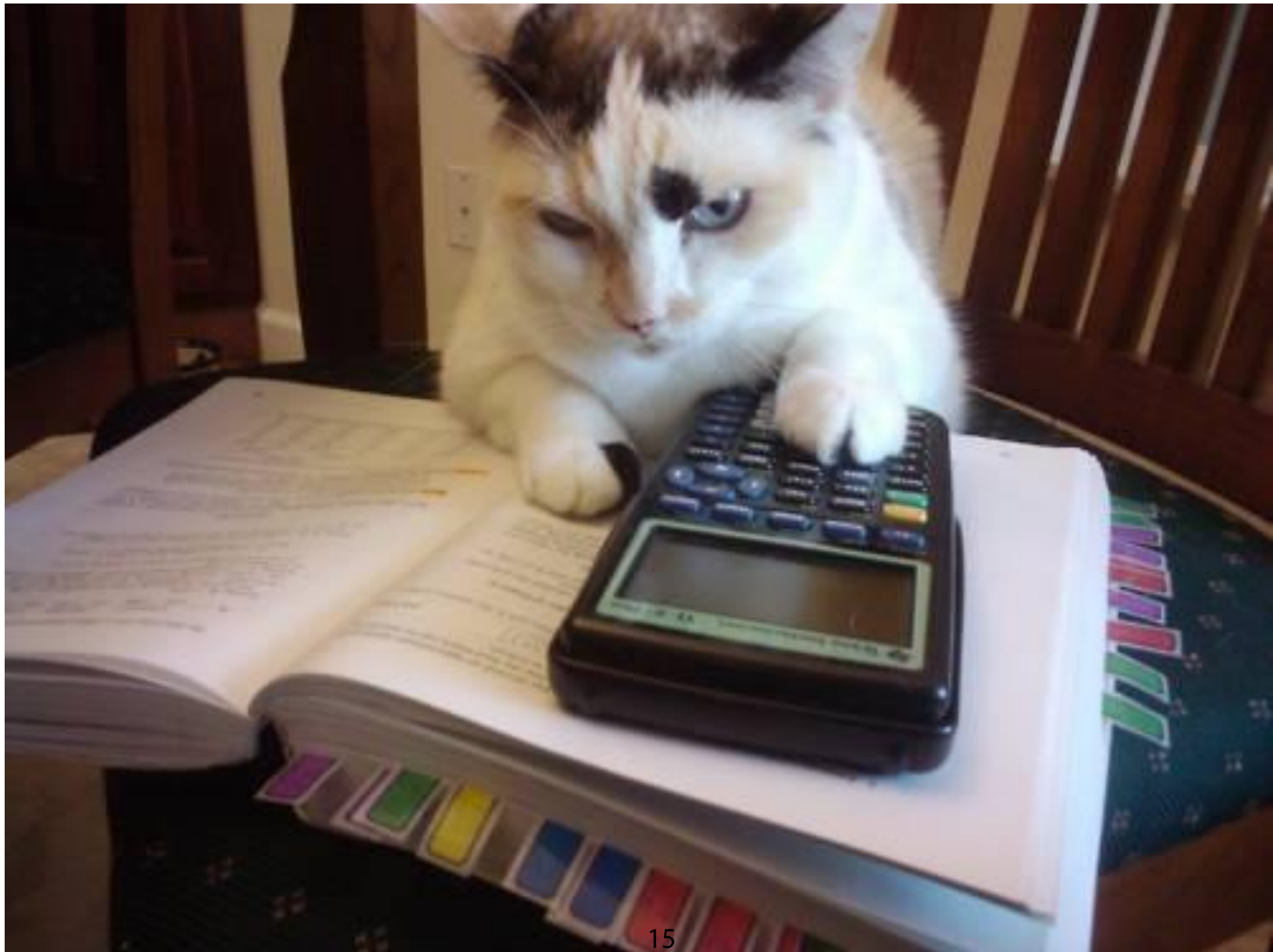
Количественный

Номинальный

Порядковый



ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ: КОТИКИ





МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ

- Среднее арифметическое
- Медиана
- Мода



СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

$$\text{Среднее} = \frac{\text{СУММА ЭЛЕМЕНТОВ}}{\text{КОЛИЧЕСТВО ЭЛЕМЕНТОВ}}$$



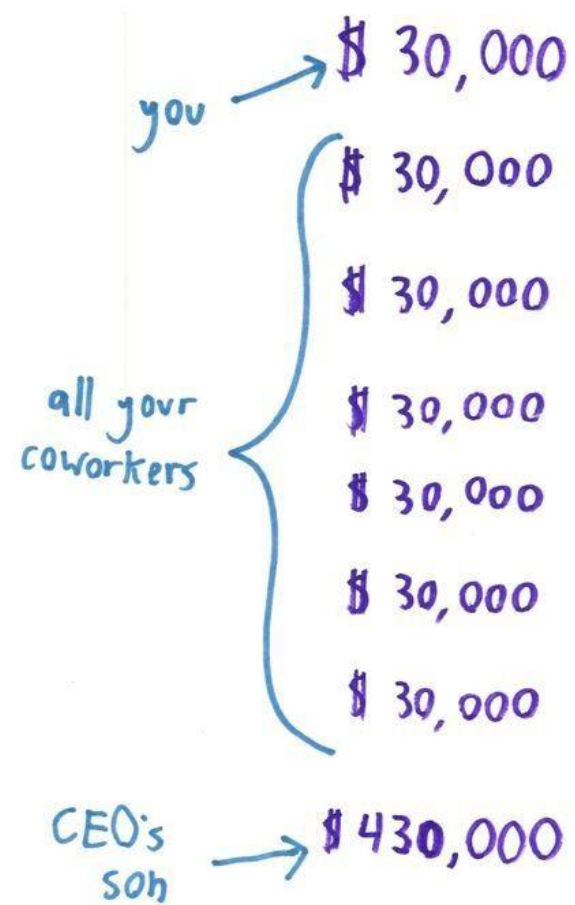
Среднее значение \bar{x}

Пример: 1,2,6,6,7

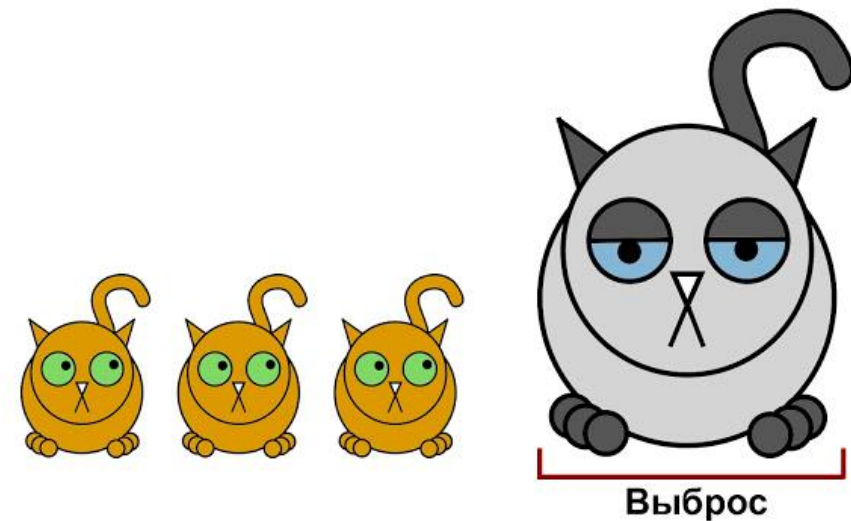
$$\text{Среднее} = \frac{1+2+6+6+7}{5} = \frac{22}{5} = 4,4$$

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ

Минус данной МЦТ: чувствительность к выбросам

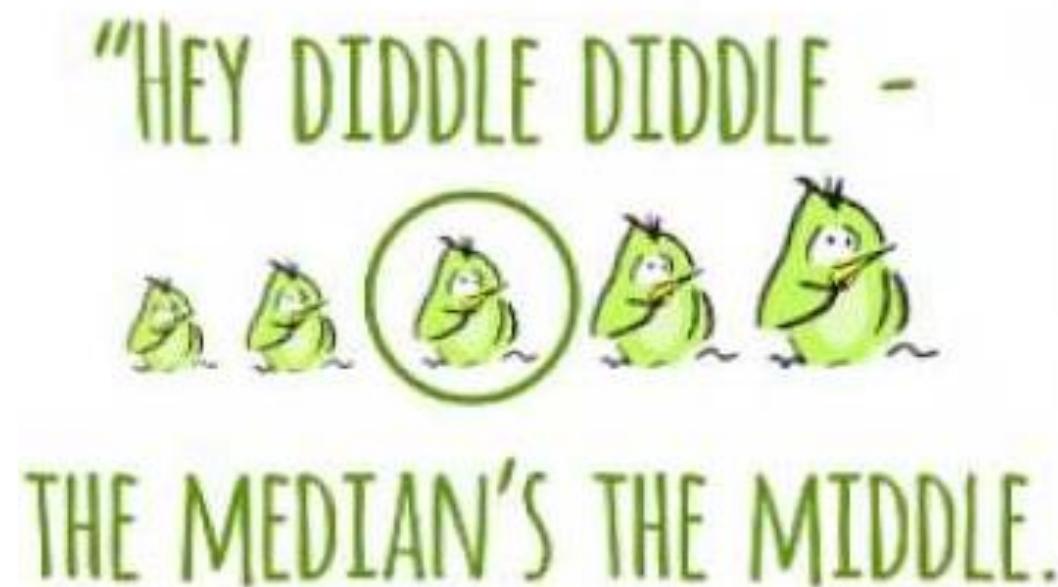


Average: \$80,000.



Алгоритм нахождения медианы:

1. Расположить значения по возрастанию
2. Если количество значений нечетное, то медианой будет центральное значение в ряду
3. Если количество значений четное, то для вычисления медианы необходимо найти среднее арифметическое двух центральных значений

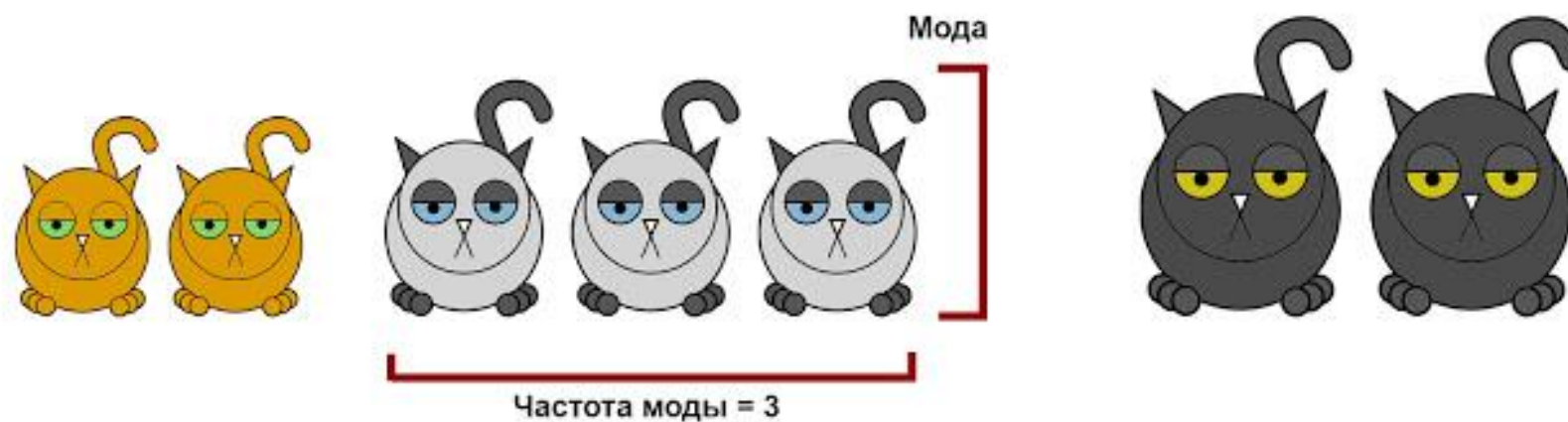




МЕДИАНА: ПРИМЕР

1. Дан числовой ряд: 1,5,3,9,11, 2, 14, 6
2. Расположим числа в порядке возрастания:
$$1,2,3,5,6,9,11,14$$
3. Найдем центральные числа: 5 и 6
4. Найдем их среднее арифметическое: $(5+6):2$
5. Получаем, что значение медианы равно 5,5

Мода-наиболее часто встречающееся значение





Пример вычисления моды:

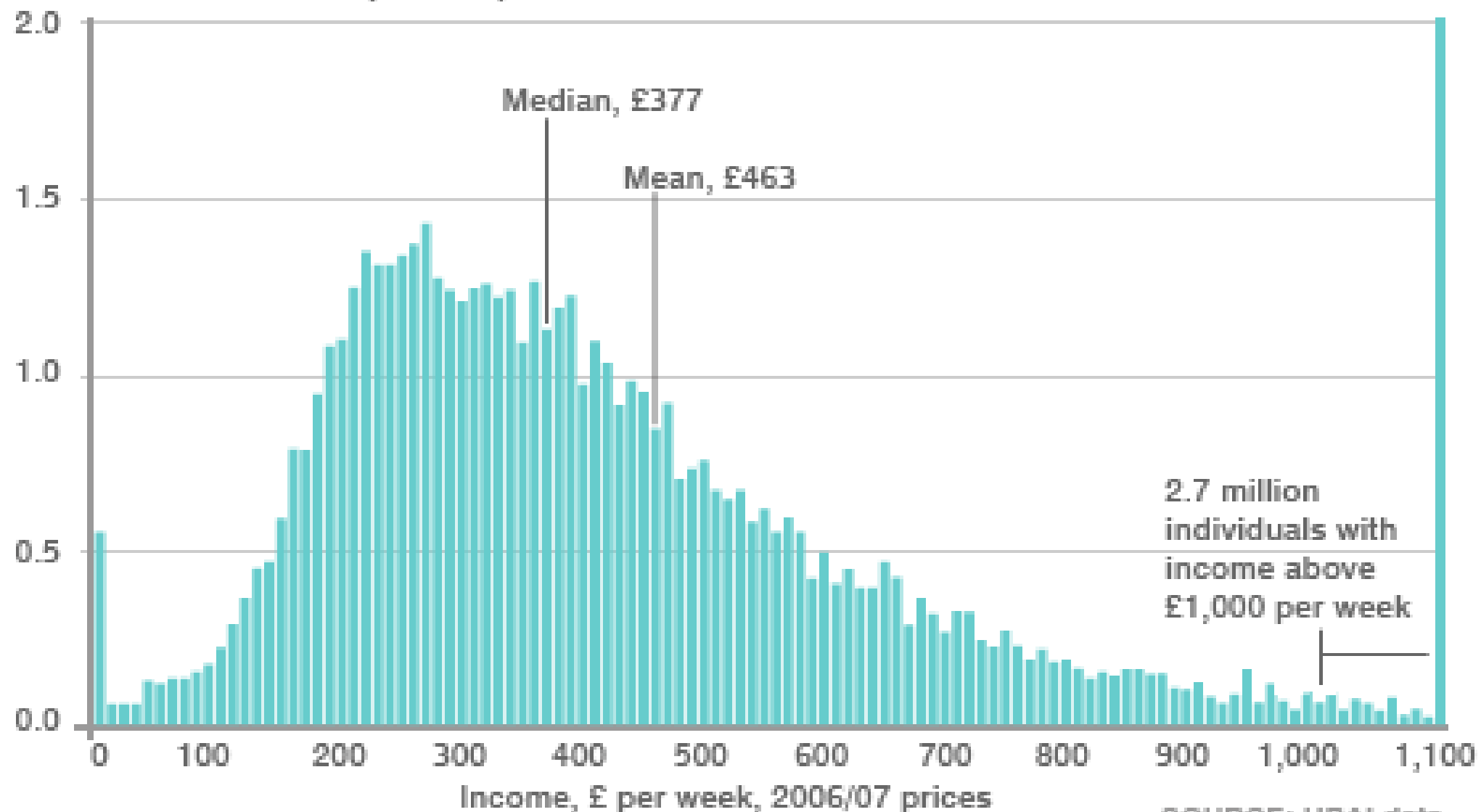
1. Пусть дан числовой ряд 1,6,1,7,1,4,5,5
2. Чаще всего в нем встречается единица
3. Получается, что мода данного ряда равна одному



ДОЛЖНЫ ЛИ СОВПАДАТЬ МЦТ?

THE UK INCOME DISTRIBUTION IN 2006 / 7

Number of individuals (millions)





МЕРЫ РАЗБРОСА

- Размах
- Межквартильный размах
- Стандартное отклонение
- Дисперсия



КВАНТИЛИ

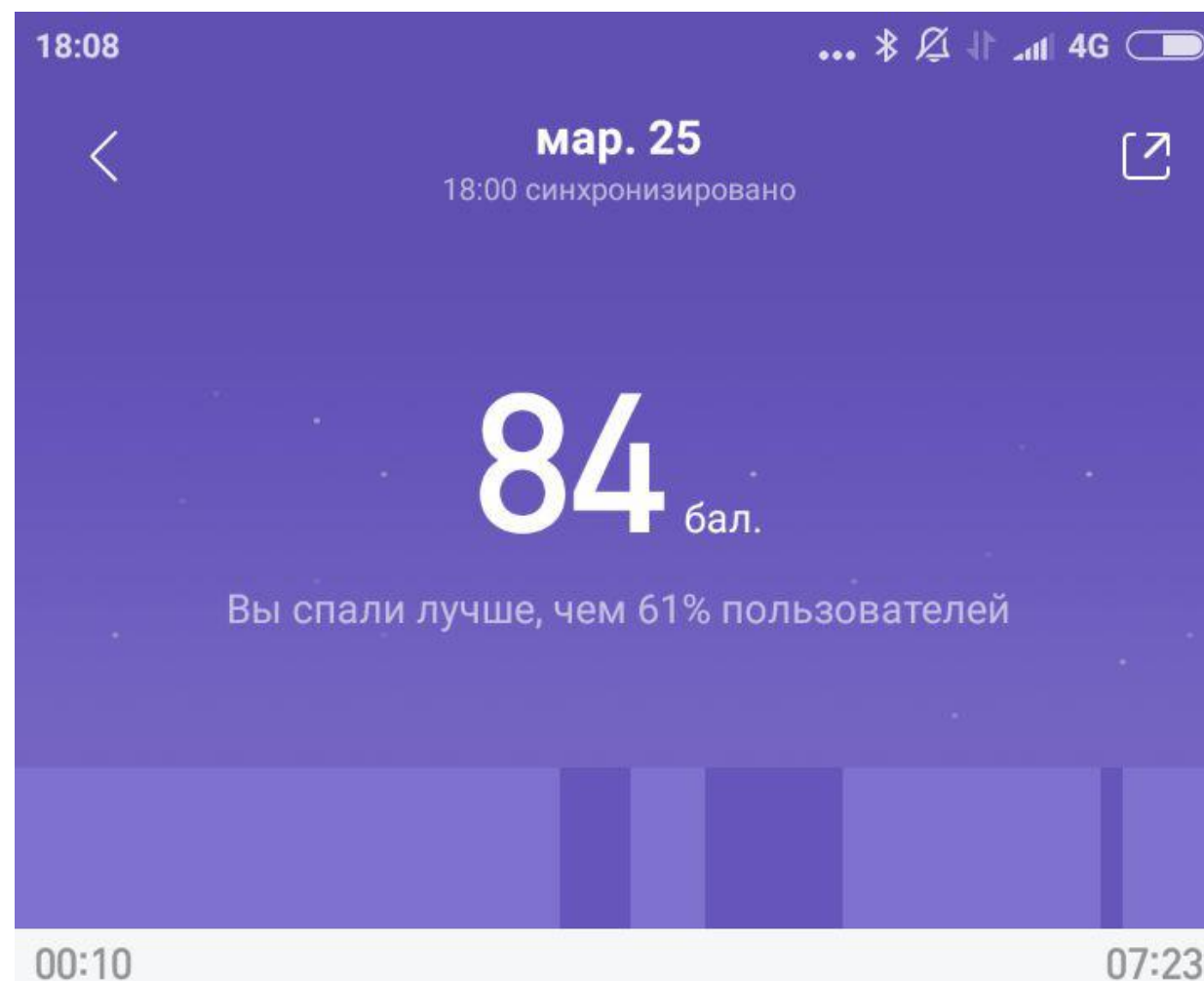
Кванти́ль в математической статистике — значение, которое заданная случайная величина не превышает с фиксированной вероятностью. Если вероятность задана в процентах, то квантиль называется проценти́лем или перценти́лем

- 0,25-квантиль называется первым (или нижним) квартилем (от лат. *quarta* — четверть);
- 0,5-квантиль называется медианой (от лат. *mediana* — середина) или вторым квартилем;
- 0,75-квантиль называется третьим (или верхним) квартилем.



ПЕРЦЕНТИЛИ

Я спала лучше, чем 61% пользователей.
Значит, 25 марта я находилась в 61-ом процентиле





МЕРЫ РАЗБРОСА: СТАНДАРТНОЕ ОТКЛОНЕНИЕ

Стандартное отклонение- показатель рассеивания значений случайной величины относительно её математического ожидания.



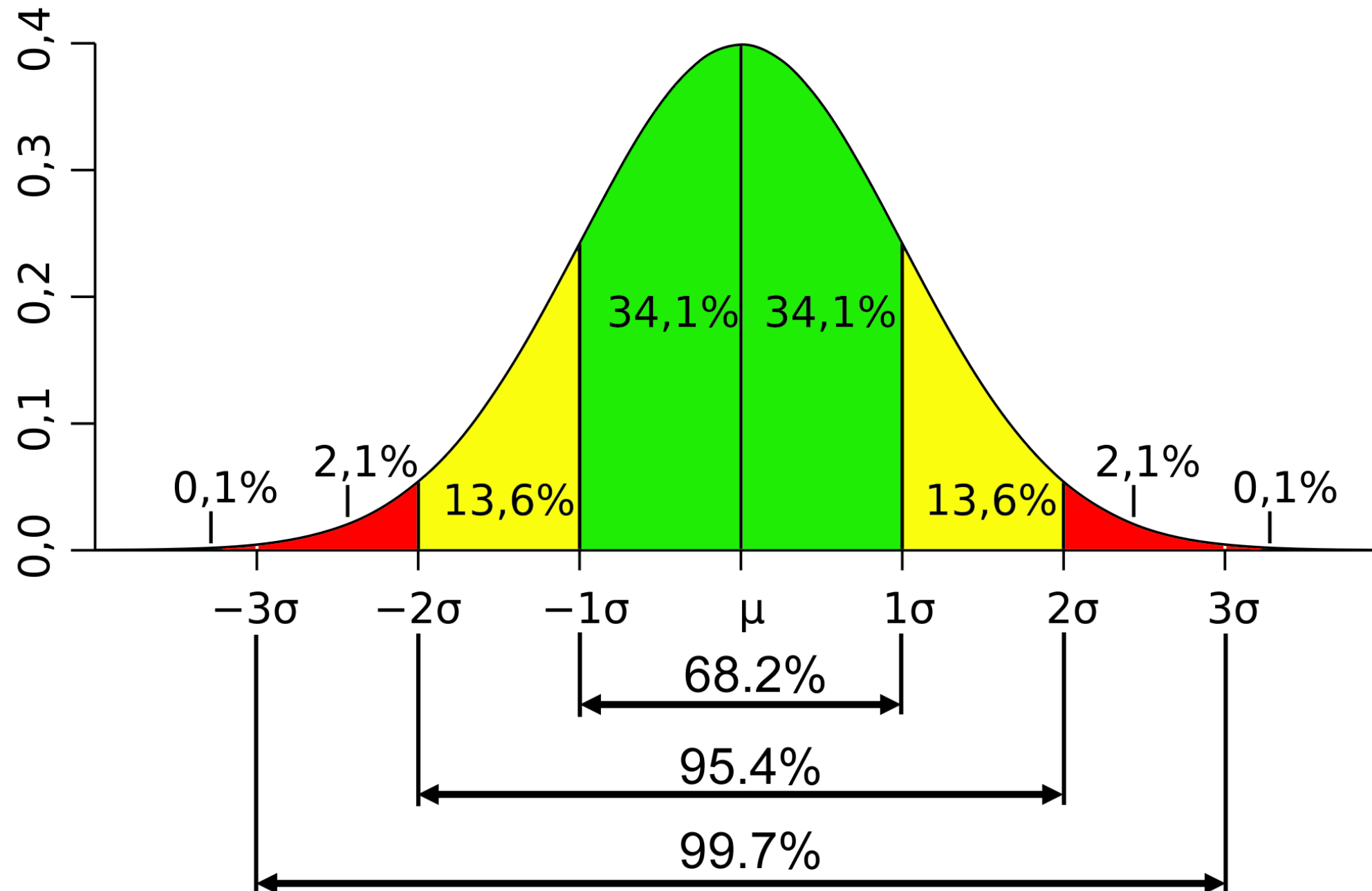


МЕРЫ РАЗБРОСА: ДИСПЕРСИЯ





НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ





МЕРЫ И ТИПЫ ПРИЗНАКОВ

Типы данных	Меры центра			Меры разброса		
	Мода	Медиана	Среднее	Размах	Q-Q	Ст.Откл.
Номинальные	✓	✗	✗	✗	✗	✗
Порядковые	✓	✓	✗	✓	✓	✗
Количественные	✓	✓	✓	✓	✓	✓



МЕРЫ И ТИПЫ ПРИЗНАКОВ

Типы данных	Меры центра			Меры разброса		
	Мода	Медиана	Среднее	Размах	Q-Q	Ст.Откл.
Номинальные	✓	✗	✗	✗	✗	✗
Порядковые	✓	✓	✗	✓	✓	✗
Количественные	✓	✓	✓	✓	✓	✓



ЧЕМ МЫ БУДЕМ ПОЛЬЗОВАТЬСЯ?





КАК УСТАНОВИТЬ?

<https://www.anaconda.com/download/>

Windows macOS Linux

Anaconda 5.2 For Windows Installer

Python 3.6 version *

[Download](#)

[64-Bit Graphical Installer \(631 MB\)](#) ⓘ
[32-Bit Graphical Installer \(506 MB\)](#)

Python 2.7 version *

[Download](#)

[64-Bit Graphical Installer \(564 MB\)](#) ⓘ
[32-Bit Graphical Installer \(443 MB\)](#)

[Behind a firewall?](#)
[*How to get Python 3.5 or other Python versions](#)
[How to Install ANACONDA](#)



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ