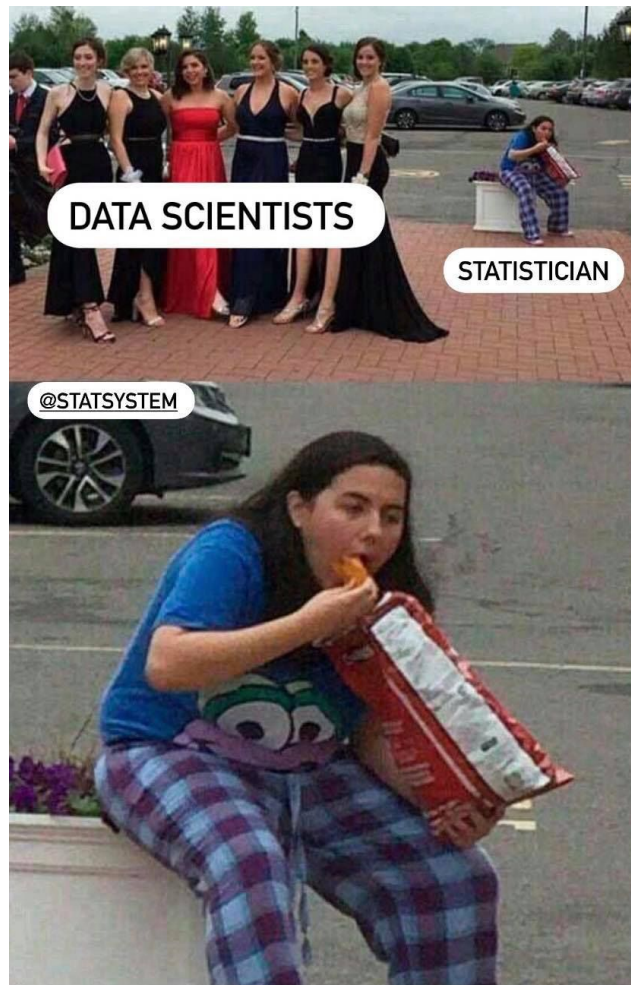


# Статистика



# Статистика

Наука за:

- Събиране
- Организиране
- Обобщаване
- Анализирне
- Интерпретатиране

на данни.

# Ползи от статистиката

- Вземане на решения, базирани на научни данни
- Стигане до изводи подкрепени от числени доказателства
- Разбиране на процеси чрез изучаване на данни
- Статистиката ни предпазва от влизане в заблуди.
- Извличане на знания от данни (data mining)

# Основни елементи на статистиката

- Събиране на данни
- Обобщаване на данни
- Интерпретатиране на данни
- Вземане на решения от данни

# Понятия в статистиката

- генерална съвкупност (ГС); популация - множество от обекти, които са подбрани по определени критерии и имат поне един общ признак
  - Пример: ученици в 11 клас, жители на Бургас, мъже между 18 и 45 години, жени между 18 и 45 години и др.
- Представителна (репрезентативна) извадка - част от генералната съвкупност, която достатъчно точно да възпроизвежда характеристиките на цялата генерална съвкупност

# Понятия в статистиката

- Обем на извадка - брой на елементите в извадката
- Променливи - изследваните характеристика/и на елементите от генералната съвкупност
- Данни - списък от стойности от наблюденията на изследваната характеристика

# Пример: Изследване на средния успех на учениците от ПГЕЕ

Променлива: среден успех за учебната година

Данни: 3.82; 3.94; 4.05; 5.28; 4.34; 6.00; 5.82; 5.01; 3.23;



# Видове променливи

- Качествени променливи - изследваната характеристика не се измерва числено; При наблюдението се определят категории - пример: Цвят на очи, пол, религиозна принадлежност и др.
- Количествени променливи - изследваната характеристика се измерва числено; пример: ръст, възраст, баланс в банкова сметка, успех по даден предмет и др.

# Описателна статистика

**Описателната статистика** се занимава с обобщаването на систематизирани данни. То е от особена важност при емпиричните изследвания и описването на резултати от експерименти.

Понятия:

- Средна стойност
- Медиана
- Мода

# Средна стойност

- Средна стойност на извадка

- означение:  $\bar{x}$

- Използва се само за количествени данни
- Съществено се влияе от всички данни
- Формула:

$$\bar{x} = \frac{\sum X}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

- Данни:  $X = x_1, x_2, \dots, x_n$

## Пример: Средна стойност

- 5 ученика получават оценки от контролно по математика: 3.85; 4.48; 5.42; 2.00; 5.26.
- Какъв е техният среден успех?
- $X = x_1; x_2; x_3; x_4; x_5 = 3.85; 4.48; 5.42; 2.00; 5.26$

$$\bar{x} = \frac{\sum X}{n} = \frac{x_1+x_2+x_3+...+x_n}{n} = \frac{3.85+4.48+5.42+2.00+5.26}{5} = \frac{21.01}{5} = 4.202$$

# Медиана

- Медиана - точка, за която 50% от данните са по-малки от нея
  - **Данните винаги се подреждат в нарастващ ред!**
  - Използва се само за количествени данни
- 
- При нечетен брой елементи, медианата съвпада със средния елемент на данните
  - При четен брой елементи, медианата е средно аритметичното на двата средни елемента

# Примери: Медиана

- Заплати във фирма:

650; 675; 705; 790; 833; 910; **950**; 950; 950; 1010; 1500; 2000; 2000;

Медианата е 950

- Заплати в друга фирма:

650; 650; 650; 710; 750; **750; 780**; 800; 890; 1250; 1500; 2000;

Медиана =  $(750 + 780) / 2 = 1530 / 2 = 765$

# Мода

- Мода - най-често срещаната стойност
- Данните могат да имат няколко моди
- Подходяща за всеки вид данни

Примери:

- 5 5 3 2 3 5 5 5 1 => Модата е 5 (среща се 5 пъти)
- 1 3 3 3 5 6 6 6 5 5 => Моды: 3 (среща се 3 пъти), 5 (среща се 3 пъти) и 6 (среща се 3 пъти)
- 1 2 3 4 5 => Няма мода

# Графично представяне на данни

- Зависи от вида на данните (количествени, качествени...)
- Зависи от това какво желаем да покажем
- Зависи от възможностите на софтуера
- Принципът е „Правилният инструмент за правилната задача”



# Графично представяне на данни

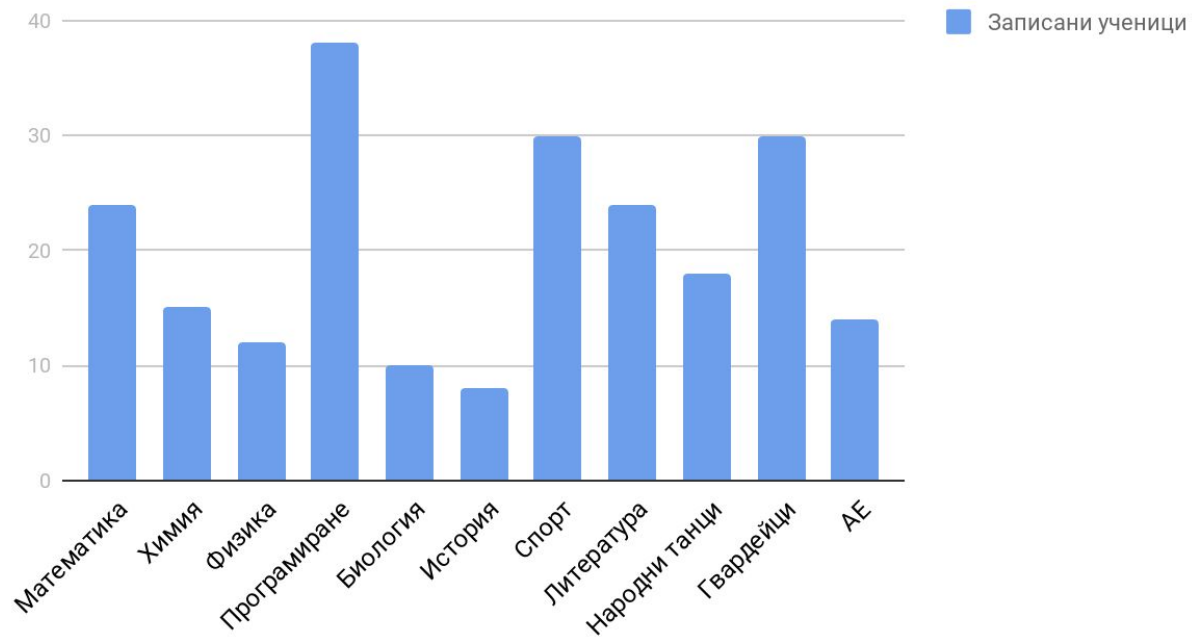
- Зависи от вида на данните (количествени, качествени...)
- Зависи от това какво желаем да покажем
- Зависи от възможностите на софтуера
- Принципът е „Правилният инструмент за правилната задача”

# Хистограма

- Всяка категория се нанася по хоризонтал и се чертаят правоъгълници, чиято височина е равна на честотата\*

# Пример: Хистограма

Извънкласни дейности

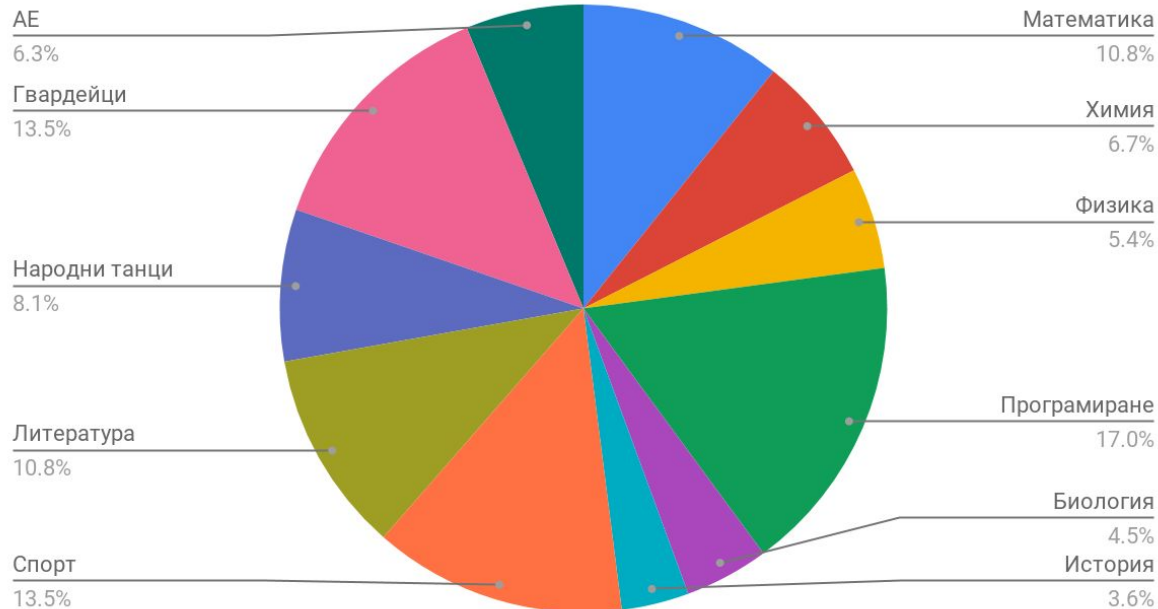


# Кръгова диаграма

- Кръг се разделя на сектори, като всеки сектор отговаря за различна категория. Лицето на сектора е пропорционално на честотата.

# Пример: Кръгова диаграма

## Записани ученици

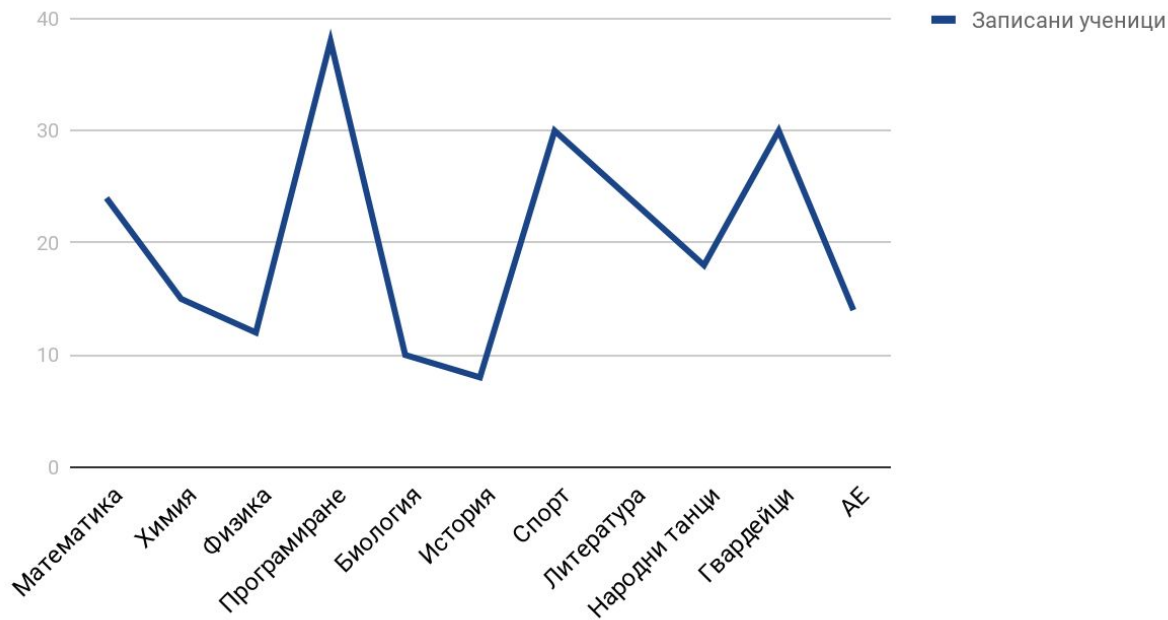


# Честотен полигон

- Графика, при която по хоризонтал се отбелязват точки, представляващи честотите на всяка категория. Посочените точки се съединяват с прави линии. Прилича на хистограма, но не ползва правоъгълници.

# Пример: Честотен полигон

Записани ученици



# Благодаря за вниманието!

Автор: Петър Р. Петров, учител по програмиране, ПГЕЕ  
„Константин Фотинов“, гр. Бургас