

# МАТЕМАТИЧЕСКА ОБРАБОТКА НА ДАННИ

[dimitrova@tu-sofia.bg](mailto:dimitrova@tu-sofia.bg)  
[pct.tu-sofia.bg/dd/pik3](http://pct.tu-sofia.bg/dd/pik3)



# Обработка на данни

## ● Терминология

- преобразуване на **данни** в **информация** или **знание**
- подпомагано от компютър

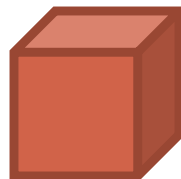
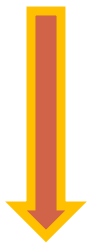
## ● Задачи

- разглеждане и обобщаване на **данните** с намерение за
- намиране на полезна **информация** и
- формиране на заключения и **знания**



Обект

Информация, Знание



Атрибути

размер

цвят

други свойства ...



Стойности на атрибутите

скали

мерни единици

представяне



ДАННИ



## Данни

**Данни** – числа или символи, които представят измерените стойности на наблюдаваните свойства на изследвания обект

## Информация

**Информация** – алгоритмично, статистически или логически извлечено значение на данните

## Знание

**Знание** – теоретично или практическо разбиране на определена тематика, основано на познаване на факти или ситуации.



## Процеси

Събиране на данни

Измерване

Наблюдение

Експериментиране

Подготовка на  
представителна  
извадка

Представяне

подходящ формат

кодова система

табулиране

Контрол и валидиране

Обобщаване

Агрегиране

според избрани

признаци

Статистичен анализ

Графична  
визуализация

Формулиране на  
хипотези

Проверка на хипотези

Разкриване на  
информация и знания



# Приложение

## ● Области на приложение

- експериментални изследвания
- моделиране и симулация на реални системи
- проектиране и създаване на реални системи
- управление на данни
- управление на производство
- създаване на интелигентни системи

## ● Задачи

- организиране на изследванията
- създаване на адекватни компютърни модели на реални системи
- добавяне на интелект към компютърните програми
- подпомагане на интелекта на хората



# Видове приложения

## ● Технически изчисления

- изчисляване на параметри на технически системи

## ● Статистическа обработка на данни

- събиране, анализ, интерпретация и представяне на данни

## ● Математическо моделиране

- моделиране на структурата и симулиране на поведението на реални системи посредством математически методи

## ● Модели и архитектури на данни

- структуриране и организация на данни

## ● Информационни системи и управление на бази данни

- търсене в големи масиви от данни и извличане на заявена информация

## ● Откриване на данни и знания

- преглеждане на големи масиви от данни и отделчнe на значима информация

## ● Визуализация на данни и информация

- предаване на абстрактни и конкретни съобщения посредством образи



# Технически изчисления

## ● Основни обекти

- константи
- променливи

## ● Типове данни

- количествени
  - цели числа
  - реални числа
- качествени
- структури от данни
  - дати
  - валути и др.

## ● Операции с данните

- аритметични операции
- логически операции
- функции
- изрази

## ● Офис системи за технически изчисления MS Excel





# Статистическа обработка на данни

## ● Описателна статистика

- описание на основни свойства на множества от експериментални данни

## ● Статистически анализ на данни

- откриване на нови свойства

## ● Проверка на хипотези

- потвърждаване или отхвърляне на хипотези за данните



# Описателна статистика

## Цели – установяване на отношения между елементите на извадка данни

- колко са подобни (**central tendency**)
- колко са различни (**statistical variability**)

## Процедура

- събиране на данни
- класифициране / категоризиране
- обобщаване
- представяне
- последващ статистически анализ

## Техники

- количествени мерки
- табулиране
- визуална интерпретация – графи, диаграми



## Типове данни

## Измервателни скали

### Качествени

- етикети, посочващи **категория** на принадлежност
- не се изчисляват с аритметични операции

### Количествени

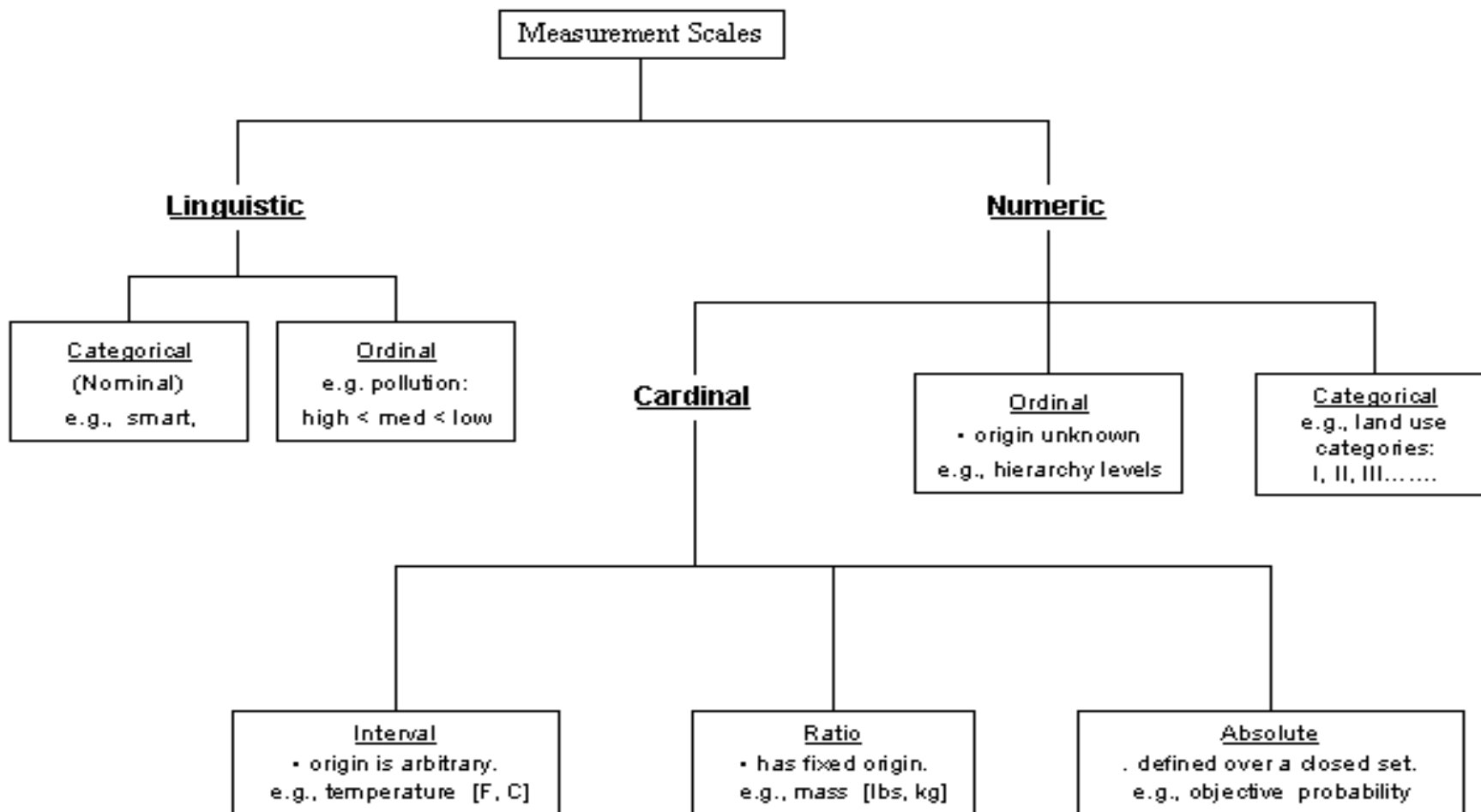
- числени стойности
- непрекъснатата или дискретизирана извадка

### Лингвистични

### Числови



# Измервателни скали



## Measurement Scales



## Измервателни скали

### ● Номинални скали

- разпределяне на данните в еднородни групи (*категории*)
- присвояването на число на всяка от групите представлява кодиране, а не количествена оценка за съответната група
- дихотомни скали – на групата се присвоява едно от две възможни състояния

### ● Рангови скали

- данните са подредени в зависимост от интензитета на разлеждан признак, без да има количествена оценка за него



# Измервателни скали

## ● Интервални скали

- количествена оценка на признак по скала с деления на равни интервали (метри, секунди и т.н.)
- нулева стойност не означава липса на значение, а е условно избрана

## ● Пропорционални скали

- количествена оценка на признак по скала, при която нулата е абсолютна

