

# Statistical Thinking and Data Treatment

Ľubica and Ján Krauskovi, Dominik Heger

Masaryk University

*hegerd@chemi.muni.cz*

STDT 01 Intro

**This course is an introduction to statistics.**

**This course is an introduction to statistics.**

## What is statistics?

The science of drawing conclusions from data.

**This course is an introduction to statistics.**

## **What is statistics?**

The science of drawing conclusions from data.

## **The goal.**

To be an informed user of numerical information.

## We will

- Think logically about quantitative evidence.
- Translate real-world situations into mathematical questions.
- Learn few important statistical and probabilistic concepts.
- Draw inference from the numbers.

## We will

- Think logically about quantitative evidence.
- Translate real-world situations into mathematical questions.
- Learn few important statistical and probabilistic concepts.
- Draw inference from the numbers.

## We will not learn (For that you need to call mathematician.)

- Statistical theory (statistics is mathematical discipline).
- Numerical analysis.
- Bestiary of techniques or named probability distributions.

## What the name means?

**Datum**, data (f.)

**Statistics** is the science of drawing conclusions from data.

**Inference** - drawing conclusions based on data from random samples.

**Probability** - understanding and quantifying randomness. What should happen just by chance?

**Descriptive Statistics** - describing and summarizing data.

Thinking how to draw conclusions and how to treat (collect, order, analyze, summarize) data.

**Why summarizing data? Is it not best to have all data?**



## Why summarizing data? Is it not best to have all data?

```
8 2 0 3 1 4 5 8 2 1 7 2 7 3 8 5 5 2 9 0 6 3 1 6 4
0 8 7 3 3 1 9 7 5 2 5 7 6 9 8 0 3 6 2 5 1 2 7 5 2
2 3 3 8 6 6 1 4 2 4 0 2 6 1 8 9 5 2 6 9 8 3 4 0 1 0
4 7 5 5 6 3 0 7 7 1 9 1 6 1 7 4 1 7 1 3 7 9 3 3 7
1 9 3 9 5 3 4 9 5 5 2 7 5 8 0 3 4 8 8 1 2 7 5 3 4
2 8 7 8 1 4 1 4 9 4 2 4 1 5 2 9 4 6 2 1 5 2 8 1 9
8 4 8 5 1 3 9 6 6 0 7 2 1 9 0 2 2 0 6 7 0 6 0 1 3 0
0 3 8 8 4 7 5 1 5 1 7 3 4 5 2 0 7 4 7 9 6 6 7 7 4
3 5 3 1 9 3 7 4 9 5 0 2 0 1 4 6 2 5 4 5 8 5 0 9 2
3 4 5 9 5 2 7 9 8 9 0 5 5 8 5 1 7 7 3 5 5 4 7 7 2
4 1 5 3 0 9 1 3 7 2 5 8 7 7 1 3 6 3 9 7 8 7 9 1 7
7 2 9 5 6 7 8 5 4 5 3 4 5 4 1 9 8 6 7 5 7 9 3 1 8
5 9 2 8 9 8 6 4 4 1 5 3 7 7 0 8 0 2 5 6 0 6 1 2 0
1 3 3 3 9 0 5 2 8 7 4 0 9 0 3 7 3 1 7 9 4 5 5 2 8
4 6 0 1 0 8 6 2 1 0 0 5 0 3 1 5 4 9 0 3 7 4 7 0 1
7 7 0 6 6 3 2 8 8 5 8 9 5 6 4 0 5 9 1 8 0 5 4 9 4
3 3 8 5 7 5 7 4 3 4 5 7 9 6 9 5 0 7 7 6 6 8 8 5 9
9 1 7 1 3 6 9 2 2 9 1 9 4 2 3 3 0 8 1 8 7 7 6 4 7 2
6 2 2 8 0 9 4 5 3 7 2 5 4 6 6 5 6 6 5 0 4 6 5 6 8
1 7 5 9 0 0 2 0 5 6 5 8 5 1 9 5 3 3 7 4 0 5 8 2 4
0 3 9 6 9 4 7 3 5 7 0 6 5 4 7 1 8 5 3 2 8 0 9 8
```

## Why summarizing data? Is it not best to have all data?

```
8 2 0 3 1 4 5 8 2 1 7 2 7 3 8 5 5 2 9 0 6 3 1 6 4
0 8 7 3 3 1 9 7 5 2 5 7 6 9 8 0 3 6 2 5 1 2 7 5 2
2 3 3 8 6 6 1 4 2 4 0 2 6 1 8 9 5 2 6 9 8 3 4 0 1 0
4 7 5 5 6 3 0 7 7 1 9 1 6 1 7 4 1 7 1 3 7 9 3 3 7
1 9 3 9 5 3 4 9 5 5 2 7 5 8 0 3 4 8 8 1 2 7 5 3 4
2 8 7 8 1 4 1 4 9 4 2 4 1 5 2 9 4 6 2 1 5 2 8 1 9
8 4 8 5 1 3 9 6 6 0 7 2 1 9 0 2 2 0 6 7 0 6 0 1 3 0
0 3 8 8 4 7 5 1 5 1 7 3 4 5 2 0 7 4 7 9 6 6 7 7 4
3 5 3 1 9 3 7 4 9 5 0 2 0 1 4 6 2 5 4 5 8 5 0 9 2
3 4 5 9 5 2 7 9 8 9 0 5 5 8 5 1 7 7 3 5 5 4 7 7 2
4 1 5 3 0 9 1 3 7 2 5 8 7 7 1 3 6 3 9 7 8 7 9 1 7
7 2 9 5 6 7 8 5 4 5 3 4 5 4 1 9 8 6 7 5 7 9 3 1 8
5 9 2 8 9 8 6 4 4 1 5 3 7 7 0 8 0 2 5 6 0 6 1 2 0
1 3 3 3 9 0 5 2 8 7 4 0 9 0 3 7 3 1 7 9 4 5 5 2 8
4 6 0 1 0 8 6 2 1 0 0 5 0 3 1 5 4 9 0 3 7 4 7 0 1
7 7 0 6 6 3 2 8 8 5 8 9 5 6 4 0 5 9 1 8 0 5 4 9 4
3 3 8 5 7 5 7 4 3 4 5 7 9 6 9 5 0 7 7 6 6 8 8 5 9
9 1 7 1 3 6 9 2 2 9 1 9 4 2 3 3 0 8 1 8 7 7 6 4 7 2
6 2 2 8 0 9 4 5 3 7 2 5 4 6 6 5 6 6 5 0 4 6 5 6 8
1 7 5 9 0 0 2 0 5 6 5 8 5 1 9 5 3 3 7 4 0 5 8 2 4
0 3 9 6 9 4 7 3 5 7 0 6 5 4 7 1 1 8 5 3 2 8 0 9
```

Graphical descriptions

Numerical summaries

## Why summarizing data? Is it not best to have all data?

```
8 2 0 3 1 4 5 8 2 1 7 2 7 3 8 5 5 2 9 0 6 3 1 6 4
0 8 7 3 3 1 9 7 5 2 5 7 6 9 8 0 3 6 2 5 1 2 7 5 2
2 3 3 8 6 6 1 4 2 4 0 2 6 1 8 9 5 2 6 9 8 3 4 0 1 0
4 7 5 5 6 3 0 7 7 1 9 1 6 1 7 4 1 7 1 3 7 9 3 3 7
1 9 3 9 5 3 4 9 5 5 2 7 5 8 0 3 4 8 8 1 2 7 5 3 4
2 8 7 8 1 4 1 4 9 4 2 4 1 5 2 9 4 6 2 1 5 2 8 1 9
8 4 8 5 1 3 9 6 6 0 7 2 1 9 0 2 0 6 7 0 6 0 1 3 0
0 3 8 8 4 7 5 1 5 1 7 3 4 5 2 0 7 4 7 9 6 6 7 7 4
3 5 3 1 9 3 7 4 9 5 0 2 0 1 4 6 2 5 4 5 8 5 0 9 2
3 4 5 9 5 2 7 9 8 9 0 5 5 8 5 1 7 7 3 5 5 4 7 7 2
4 1 5 3 0 9 1 3 7 2 5 8 7 7 1 3 6 3 9 7 8 7 9 1 7
7 2 9 5 6 7 8 5 4 5 3 4 5 4 1 9 8 6 7 5 7 9 3 1 8
5 9 2 8 9 8 6 4 4 1 5 3 7 7 0 8 0 2 5 6 0 6 1 2 0
1 3 3 3 9 0 5 2 8 7 4 0 9 0 3 7 3 1 7 9 4 5 5 2 8
4 6 0 1 0 8 6 2 1 0 0 5 0 3 1 5 4 9 0 3 7 4 7 0 1
7 7 0 6 6 3 2 8 8 5 8 9 5 6 4 0 5 9 1 8 0 5 4 9 4
3 3 8 5 7 5 7 4 3 4 5 7 9 6 9 5 0 7 7 6 6 8 8 5 9
9 1 7 1 3 6 9 2 2 9 1 9 4 2 3 3 0 8 1 8 7 7 6 4 7 2
6 2 2 8 0 9 4 5 3 7 2 5 4 6 6 5 6 6 5 0 4 6 5 6 8
1 7 5 9 0 0 2 0 5 6 5 8 5 1 9 5 3 3 7 4 0 5 8 2 4
0 3 9 6 9 4 7 3 5 7 0 6 5 4 7 1 1 8 5 3 2 8 0 9 8
```

Graphical descriptions

Numerical summaries

First - **single variable** (education level, income, favorite color)

## Why summarizing data? Is it not best to have all data?

```
8 2 0 3 1 4 5 8 2 1 7 2 7 3 8 5 5 2 9 0 6 3 1 6 4
0 8 7 3 3 1 9 7 5 2 5 7 6 9 8 0 3 6 2 5 1 2 7 5 2
2 3 3 8 6 6 1 4 2 4 0 2 6 1 8 9 5 2 6 9 8 3 4 0 1 0
4 7 5 5 6 3 0 7 7 1 9 1 6 1 7 4 1 7 1 3 7 9 3 3 4
1 9 3 9 5 3 4 9 5 5 2 7 5 8 0 3 4 8 8 1 2 7 5 3 4
2 8 7 8 1 4 1 4 9 4 2 4 1 5 2 9 4 6 2 1 5 2 8 1 9
8 4 8 5 1 3 9 6 6 0 7 2 1 9 0 2 0 6 7 0 6 0 1 3 0
0 3 8 8 4 7 5 1 5 1 7 3 4 5 2 0 7 4 7 9 6 6 7 7 4
3 5 3 1 9 3 7 4 9 5 0 2 0 1 4 6 2 5 4 5 8 5 0 9 2
3 4 5 9 5 2 7 9 8 9 0 5 5 8 5 1 7 7 3 5 5 4 7 7 2
4 1 5 3 0 9 1 3 7 2 5 8 7 7 1 3 6 3 9 7 8 7 9 1 7
7 2 9 5 6 7 8 5 4 5 3 4 5 4 1 9 8 6 7 5 7 9 3 1 8
5 9 2 8 9 8 6 4 4 1 5 3 7 7 0 8 0 2 5 6 0 6 1 2 0
1 3 3 3 9 0 5 2 8 7 4 0 9 0 3 7 3 1 7 9 4 5 5 2 8
4 6 0 1 0 8 6 2 1 0 0 5 0 3 1 5 4 9 0 3 7 4 7 0 1
7 7 0 6 6 3 2 8 8 5 8 9 5 6 4 0 5 9 1 8 0 5 4 9 4
3 3 8 5 7 5 7 4 3 4 5 7 9 6 9 5 0 7 7 6 6 8 8 5 9
9 1 7 1 3 6 9 2 2 9 1 9 4 2 3 3 0 8 1 8 7 7 6 4 7 2
6 2 2 8 0 9 4 5 3 7 2 5 4 6 6 5 6 6 5 0 4 6 5 6 8
1 7 5 9 0 0 2 0 5 6 5 8 5 1 9 5 3 3 7 4 0 5 8 2 4
0 3 9 6 9 4 7 3 5 7 0 6 5 4 7 1 1 8 5 3 2 8 0 9 8
```

Graphical descriptions

Numerical summaries

First - **single variable** (education level, income, favorite color)

Next - **relation between two variables** (education level and income)

**No statistical treatment can ever make up for well measured data!**

**No statistical treatment can ever make up for well measured data!**

Dobře naměřená data nelze nahradit žádnou statistikou!

# Information sources

<https://courses.edx.org/courses/BerkeleyX/Stat.2.1x/>

<http://www.stat.berkeley.edu/~stark/SticiGui/>