# Foundation of Inference: Hypothesis testing

a) Introduction

Prof. Dr. Jan Kirenz

Statistical inference is primarily concerned with understanding and quantifying the **uncertainty of parameter estimates**.

Population vs sample

- Sometimes the dataset at hand represents the **population** for the entire research question.

- Our data is usually a (hopefully) **representative subset (sample)** of a larger population
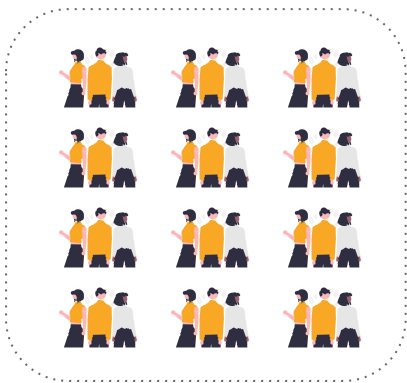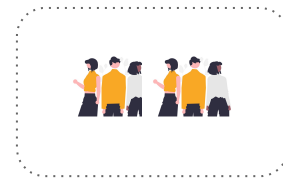
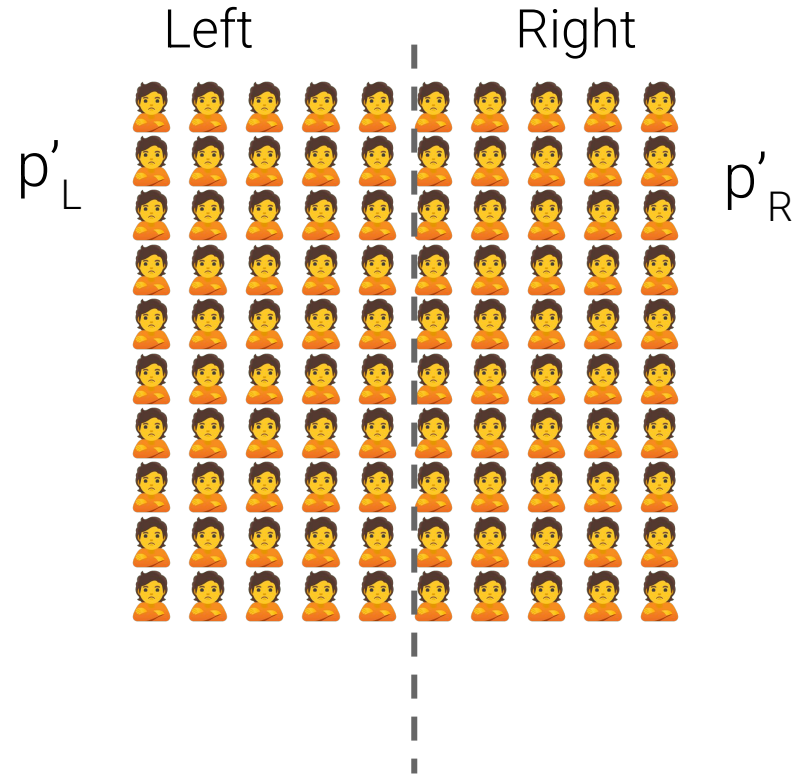How different is one dataset from another?
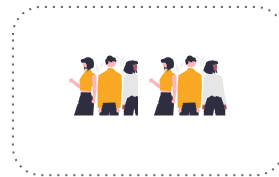
Notation for samples and population

- **p**   population proportion

- μ population mean

- $\hat{p}$ and **p'**   sample proportion
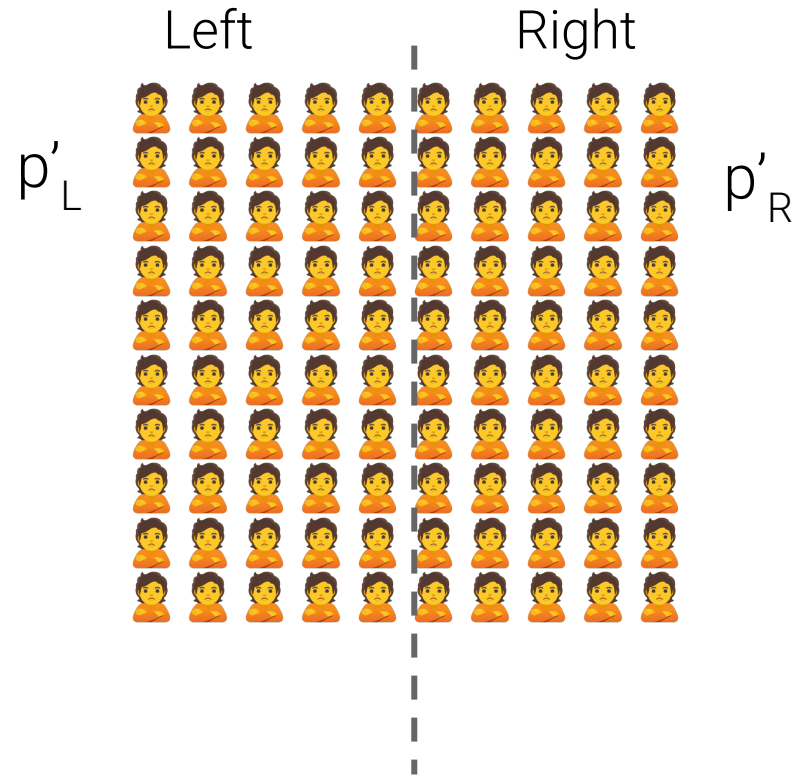
- $\bar{x}$ and **x'**   sample mean

Proportion of students who prefer reading books on screen

- What are the variables?

- What data type?

- Levels of the variables?

- What is the explanatory variable?

Left   Right

p'_L p'_R

- Variables and levels:
  - **side of the room**: left, right
  - **prefer to read books on screen**: yes; no
- Assumption about the relationship??
  - A: independent?
  - B: dependent?

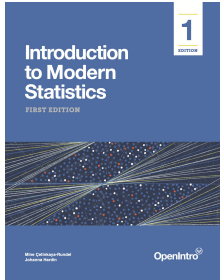Left   Right

p'_L p'_R

Studying **randomness** is a key focus of statistics.

Three different approaches for quantifying the **variability inherent in data**:

1. (Randomization) 🎲

2. (bootstrapping)

3. mathematical models *1+1=2*

Resources

The content of this lecture is mainly based on the excellent book (can be accessed for free)

- "Introduction to Modern Statistics" by Mine Çetinkaya-Rundel and Johanna Hardin (2021)

    https://openintro-ims.netlify.app/index.html