

Introduction

C. Maugis-Rabuseau & P. Neuvial

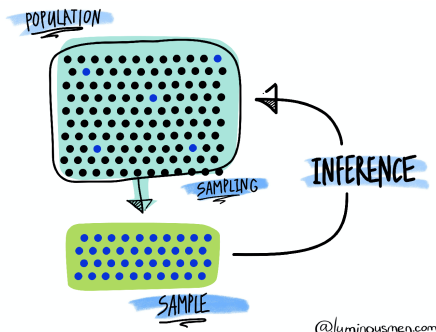
IMT

2021-06-30

Descriptive vs inferential statistics

Context: observations sampled from a population

- Descriptive statistics: summarize information on the observations (the *sample*)
- Inferential statistics: draw conclusions on the population from the observations



Descriptive vs inferential statistics

Descriptive statistics

- mean, variance
- correlation
- histograms
- boxplots

Inferential statistics

- confidence intervals
- hypothesis tests
- parameter estimation

Descriptive vs inferential statistics

Descriptive statistics

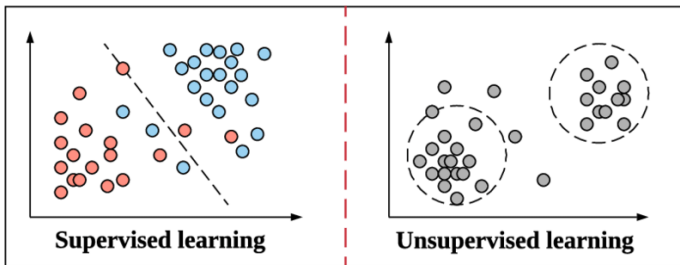
- mean, variance
- correlation
- histograms
- boxplots

Inferential statistics

- confidence intervals
- hypothesis tests
- parameter estimation

We will mainly focus on **inferential statistics**

Supervised vs unsupervised learning



Outline of the course

4 half-days with course and practicals:

- 1 Introduction to the linear model
- 2 Beyond the linear model: variable selection and generalized linear model
- 3 Supervised learning (classification)
- 4 Unsupervised learning (clustering)

Outline of the course

4 half-days with course and practicals:

- 1 Introduction to the linear model
- 2 Beyond the linear model: variable selection and generalized linear model
- 3 Supervised learning (classification)
- 4 Unsupervised learning (clustering)

Goals

Gain insight into

- how and when to use statistical methods
- how to estimate the risk/performance of a statistical method, and related issues in statistical learning:
 - underfitting and overfitting
 - model selection/penalization
 - ROC curves