

# Resampling

Wolfgang Kremser, Nina Schmitzberger

Paris-Lodron Universität Salzburg

30. April 2018

# Overview

## Introduction

What is resampling?

Revision: Hypothesis Testing

Confidence Intervals

Easy Example

## bootstrap

Introduction

Easy Example

Implementation and R packages

## permutation

Introduction

Easy Example

Implementation and R packages



## What is resampling?

### Resampling

- Determining statistical properties of sampling functions
- Repeated sampling of a baseline sample
  - baseline sample becomes population
- Calculate sampling function repeatedly on these sub-samples
- Examine resulting distribution properties
- Useful for more precise approximations of required statistic
- Mostly used in context with hypothesis testing



## What is a hypothesis test?

- Method of statistical inference
- Compare data set obtained by sampling
  - against other sampled data set
  - against synthetic data set from an idealized model
- Propose hypothesis  $H_1$  for statistical relationship
- Compare against null hypothesis  $H_0$  (no relationship)
- Comparison statistically significant  $\rightarrow$  relationship would be unlikely under  $H_0$ , discard  $H_0$



## Steps followed for constructing a hypothesis test:

1. Stating relevant null- and the alternative hypothesis
2. Considering statistical assumptions, picking appropriate test, stating test statistic  $T$
3. Selecting significance level  $\alpha$
4. Computing observed value  $t$  for  $T$
5. Deciding whether to reject or accept null hypothesis



## What is a Confidence Interval (CI)?

- Consists of a range of numbers such that  $t$  realization of  $T$  will be included with fixed probability
- Often used for testing hypotheses
- $(1-\alpha)$  CI is calculated for required distribution,  $t$  is calculated from sample
- Accept  $H_0$  is included, reject if not
- Thresholds for two sided hypothesis test  $\approx$  Confidence interval



Example  $\longrightarrow$  R



## Resampling methods

- bootstrap
- permutation



## What is bootstrapping?

- Any test that relies on sampling with replacement
- Estimating properties of estimator
- Used for constructing hypothesis tests
- Alternative to inference based on parametric model assumption
  - when assumption of parametric model is in doubt or complicated
  - when parametric inference is impossible
  - when complicated formulas for calculation are required



Example  $\longrightarrow$  board



## Implementation and R packages

- Own implementation
- Package *boot*
- Package *resample*
- Package *bootstrap*

⇒ switch to R



## What is a permutation test?

- statistical significance test
- sample without replacement
- advantage: exists for any test statistic  
⇒ minimizes loss
- Thinking exercise: Review example from before without replacement



Example  $\longrightarrow$  board



## Implementation and R packages

- Own implementation
- Package *resample*
- Package *coin*
- Package *ImPerm*

⇒ switch to R



THANKS FOR LISTENING