

Foundations of Statistics

Trainer



Dr. András Aszódi
VBCF BioComp, Vienna

- Holds a PhD in chemistry
- He has extensive computational biology experience, both in academia and industry
- He develops short courses on biostatistics and scientific programming

Objective

The main goal of this course is to explain the ideas behind statistical hypothesis testing – to understand "why we do the tests the way we do them". To this end we first review the foundations (sampling theory, discrete and continuous distributions, Central Limit Theorem), then we focus on classical hypothesis testing itself. Basic concepts (significance, confidence intervals, error types, test power) will be introduced using Student's t-test as an example. Finally, an overview of the most popular statistical tests will be given.

Description

The participants will run simple hands-on exercises to try out the concepts discussed. On the second day they will also have the opportunity to run t-tests and power calculations on their own data. We use the R programming language for the exercises. However, the main focus of the workshop is to get a better understanding of statistical hypothesis testing in general, and less on how a particular test function is run in R.

Day 1:

- Sampling theory: obtaining information about a population via sampling
Sample characteristics (location, dispersion, skewness, quantiles), estimation of the mean, standard deviation etc.
- Probability theory: Discrete and continuous probability distributions
- The importance of the Normal Distribution: Central Limit Theorem

Day 2:

- Basic principles of statistical hypothesis testing (t-test as an example)
- Types of errors, power calculations, estimation of necessary sample size
- "Cookbook of tests": Distribution tests, parametric vs. non-parametric tests (t-test vs. Wilcoxon-Mann-Whitney) F-test, counting statistics, contingency tables (chi-square), correlation tests

Out of scope:

Due to the complexity of the topic, analysis of variance (ANOVA) techniques will not be addressed. Because of time constraints we will not have the opportunity to discuss research projects of the participants.

Conditions

Basic familiarity with R and simple plotting commands is required:

- Using the R interpreter, either the command-line program or in R Studio
<https://rstudio.com/>
- Know how to invoke R functions, pass optional / named parameters
- Some familiarity with simple plotting commands

Organizational Information

Language / Format	English / On campus
Target group	Doctoral Candidates at all stages and Postdocs (R2/R3) from all faculties
Date	Wednesday-Thursday, 19-20 June 2024, 9:00 – 13:00
Registration	For registration click here