

Biological Data Analysis

Course Program

Teaching staff: Märt Möls, Krista Fischer, Lars Snipen, Egils Stalidzans, Anu Roos, Anne Selart.

Pre-course activities (distance learning part, November 20 – December 20, 2006)

Topics covered: Introduction to R (principles; data import; basic data manipulations; basic descriptive statistics; frequency tables; basic graphics); overview of basic descriptive statistics and their uses.

How? Student has to download and install R to his/her computer. One has to go through/repeat a provided sample session and complete a few exercises. As homework one has to import a dataset and give a description of the data using suitable descriptive statistics, tables and graphics.

What will be provided to students: Introductory written materials about R; an overview about basic descriptive statistics and their uses; a sample session and exercises; dataset and its description. E-mail support will be provided; a discussion list for course participants will be set up.

Face-to-face course (29th of January – 3rd of February)

Sunday, 28th of January

Arrival

1. day, Monday, 29th of January

9.00 – 12.15 Introduction to the course.

Summary and discussion of distance learning part

Presentation of a random sample of students' homeworks; questions; general discussion. (*Krista Fischer & assistants*)

Making conclusions from data. Population and sample, sampling variability, normal distribution. (*Krista Fischer*)

10.30-11.00 *Coffee break*

Lunch

13.15 – 14.45 Prediction interval & confidence interval. Calculating confidence intervals for expectations and percents. Presenting confidence intervals in graphs. (*Märt Möls*)

15.00 – 16.30 Computer practical: Looking at the distribution of the data. Calculating confidence intervals. Adding the description of uncertainty to graphics. (*Märt Möls & assistants*)

16.45 *Walking tour in Tartu*

2. day, Tuesday, 30th of January

- 9.00 – 10.30 Hypothesis testing. Basic principles: Null hypothesis and alternative hypothesis; Type I and Type II Errors; confidence level; p-value. Tests covered in detail: t-test (hypothesis about expectation, paired samples, unpaired samples), chi-square test. (*Märt Möls*)
- 10.45 – 12.15 Computer practical: Testing hypothesis, t-test. Choosing the right test, understanding the meaning of null hypothesis, interpreting results. (*Märt Möls & assistants*)
- Lunch*
- 13.15 – 14.45 Statistical dependence (association), testing statistical dependence. Introduction to simple linear regression, linear regression model, principles of estimation, determination and correlation coefficients. (*Krista Fischer*)
- 15.00 – 16.30 Computer practical: introduction to linear regression in R. (*Krista Fischer & assistants*)

3. day, Wednesday, 31st of January

- 9.00 – 10.30 Multiple regression, model building. Testing goodness-of-fit. Modeling non-linear relationships with linear regression. (*Krista Fischer*)
- 10.45 – 12.15 Computer practical: linear models. Estimation, model selection and interpretation of the results. (*Krista Fischer & assistants*)
- Lunch*
- 13.15 – 14.45 Indicator variables. 1-way ANOVA. Different parameterisations. Issues of multiple testing. Bonferroni method/Tukey HSD. (*Märt Möls*)
- 15.00 – 16.30 Computer practical: introduction to ANOVA models, interpretation of model parameters, variable selection in ANOVA models. (*Märt Möls & assistants*)

4. day, Thursday, 1st of February

- 9.00 – 10.30 2-way ANOVA, Analysis of Covariance. Interactions, model assumptions, checking model assumptions. Transformations to achieve normality. (*Märt Möls*)
- 10.45 – 12.15 Computer practical. Checking model assumptions. Interactions in ANOVA and regression models. (*Märt Möls & assistants*)
- Lunch*
- 13.15 – 14.45 Causality. Randomized experiments vs observational data. Basic experimental designs. (*Krista Fischer*)
- 15.00 – 16.30 Computer practical. Confounding, simple simulations. (*Krista Fischer & assistants*)

5. day, Friday, 2nd of February

9.00 – 10.30 Hierarchical clustering, k-means. Estimating the number of clusters. (*Lars Snipen*)

10.45 – 12.15 Computer practical: Hierarchical clustering, k-means. Estimating the number of clusters. (*Lars Snipen & assistants*)

Lunch

13.15 – 14.45 Principal Component Analysis. (*Lars Snipen*)

15.00 – 16.30 Computer practical. Principal Component Analysis. (*Lars Snipen & assistants*)

6. day, Saturday, 3rd of February

9.00 – 11.00 Test

11.00 – 12.00 Evaluation of the course, closure.

12.00 – 13.00 *Lunch*

13.00 – 19.00 *Excursion to Southern Estonia*

19.00 – *Dinner*

Sunday, 4th of February

Departure