R in statistical analysis for geneticists		
Institute of Plant Genetics Polish Academy of Sciences	**prof. dr hab. Idzi Siatkowski idzi@up.poznan.pl** Coordinator	**Lecturers: team of scientists from the Department of Mathematical and Statistical Methods, Poznan University of Life Sciences**

General information:

Number / form (s) / type (s) of classes	20 didactic hours (lectures and practical courses)
Didactic cycle	Spring-summer 2019
Language	English
ECTS credits	2

Objective of the course:
Understanding the basic computational and graphic R capabilities used in statistics for geneticists

Topics:

1. Introduction to R and RStudio
2. Calculations in R
3. Data preparation
4. Visualizations
5. Statistical testing
6. Assessing dependencies between variables
7. Linear and multiple linear regression
8. Exam

Effects of course (in terms of knowledge, skills)
PhD student:
- has knowledge about R and RStudio
- knows the advantages of R and RStudio
- can install R and RStudio and the appropriate packages
- knows the methods of searching for help
- can do math calculations
- understands the concept of variable, vector, matrix and data frame
- can prepare data for calculations
- can load data frames, text data and excel data
- able to visualize data
- can represent a mathematical function in form of a graph
- knows the rules for the use of statistical tests
- can do statistical tests for mean values
- can interpret the results obtained after testing
- can statistically compare two or more groups
- can perform a test for dependencies between variables
- understands the importance of linear and multiple regression
- can perform linear and multiple regression

– can interpret results obtained after applying linear and multiple regression

Course content:

1. **Introduction to R and RStudio (1 hour)**
   1.1. What is R?
   1.2. The advantages of R
   1.3. Installing R and RStudio
   1.4. Packages
   1.5. Help and documentation
2. **Calculations in R (3 hours**)
   2.1. Simple mathematical calculations
   2.2. Variables
   2.3. Numeric and character vectors, matrices and data frames
3. **Data preparation (2 hours)**
   3.1. Data frames
   3.2. Text files
   3.3. Excel files
   3.4. Data of various formats
4. **Visualizations (3 hours)**
   4.1. Graphic presentation of data
   4.2. Function graphs
5. **Testing (4 hours)**
   5.1. Introduction
   5.2. Two populations - normal distributions
   5.3. Two populations - any distributions
   5.4. Analysis of variance – ANOVA
   5.5. Multiple comparison tests
6. **Dependencies (1 hour)**
   6.1. Correlations
   6.2. Contingency tables
7. **Linear and multiple linear regression (4 hours)**
   7.1. Linear regression
   7.2. Multiple linear regression
   7.3. Selection of variables
8. **Exam (2 hours)**

Teaching methods / techniques:

- **lectures in English, using multimedia techniques**
- **exercises at computers**

Evaluation of learning outcomes:

- **practical exam at computers**