[Course description for Scientific PhD courses]

[Please remove information in square brackets, before sending the course description to the following e-mail: [gradschool@science.au.dk](mailto:gradschool@science.au.dk). The course description will be announced at the GSST website and at the national PhD course website]

**Name of course: Basic data handling and statistics using R**

**ECTS credits: ????**

**Course parameters:**

*Language*: English

*Level of course*: PhD course

*Time of year*: March/April 2019

*No. of contact hours/hours in total incl. preparation, assignment(s) or the like [please indicate both]:Lectures and preparation 15 hours, exercises 12 hours, mentoring and reporting and mentoring 45 hours.*

*Capacity limits*: Minimum 10 and maximum 25 students

**Objectives of the course:**

The aim of the course is to introduce the PhD student to basic notions of statistical analysis and data handling using the statistical software R.

**Learning outcomes and competences:**

At the end of the course, the student **should be able to**:

* Import data into ‘R’ (own data, automatically collected data, external data bases) and organize the data efficiently in data base structure.
* Perform advanced data treatment, e.g. quality control and transformations.
* Perform advanced statistical analysis on the data with the use of appropriate statistical tools.
* Present the results graphically or in other appropriate ways.

**Compulsory programme:**

The course is divided into three parts. The first part consists of 3 days of lectures and hand-on exercises. The second part consists of 2 days of “mentoring”, where the PhD student works independently on own data, guided by the teacher. The last part consists of a short written report based on the students won data to be handed in no later than 15 April 2019. For ECTS to be awarded, PhD students must take active part in all parts of the course.

**Course contents:**

* Introduction to R:
  + Importing data
  + Data manipulation
  + Effective data visualization using ggplot2
* Linear modeling:
  + T-test and ANOVA
  + Simple linear regressions
* Generalized linear models (GLM) using R:
  + Main error distribution families (normal, Poisson, binomial, negative-binomial, etc.)
* Model validation and selection
* Variance partitioning
* Non-linear modeling
* Detecting and dealing with spatial auto-correlation
* Canonical analyzes:
  + Principal component analysis (PCA)
  + Redundancy Analysis (RDA)

**Prerequisites:**

Basic skills in R + basic notions of R programming + some stats experience??

**Name of lecturer[s]:**

Philippe Massicotte, University of Laval, Canada (lectures, exercises and mentoring)

Floris M. van Beest and Niels M. Schmidt (reporting)

**Type of course/teaching methods:**

Lectures and hands-on exercises, written report.

**Literature:**

Lecture notes written by Philippe Massicotte (distributed electronically during the course).

**Course homepage:**

None.

**Course assessment:**

PhD students will be evaluated based on their active participation in all course elements and on the written report.

**Provider:**

Department of Bioscience, Roskilde

**Special comments on this course:**

Students are expected to bring your own computer with newest version of R installed.

**Time:**

11-15 march 2019. Written report to be handed in no later than 1 May 2019.

**Place:**

Department of Bioscience, Frederiksborgvej 399, 4000 Roskilde

**Registration:**

[The following is an example of what you might write in connection with registration, please feel free to adjust the text to your needs]

Deadline for registration is ???. Information regarding admission will be sent out no later than ??? [e.g. two days later].

For registration: ???

If you have any questions, please contact Niels Martin Schmidt, e-mail: nms@bios.au.dk

Til annoncering på webside – til boks i øverste højre hjørne:

**PLEASE NOTE**

Deadline for registration is ???.

If you have any questions, please contact ???, e-mail: ???