

Course description for the ClimBEco course “Linear Regression using R” 21/3 - 12/4 2016, Lund University

A PhD level course describing how to analyse linear dependencies in data using linear regression. Basic tools for model selection and validation as well as hypothesis testing will be covered.

Lecturer: Umberto Picchini
Computer exercises: Anna Lindgren

Course outline

The course consists of 8 lectures during 4 weeks (see schedule below and [course webpage](#)) covering:

Week 1 - Simple linear regression

Week 2 - Multiple regression

Week 3 - Variable selection and regression diagnostics

Week 4 - Outliers and residual analysis

Course points: 2.5 ects points

Description

Regression analysis deals with modelling how one characteristic (height, weight, price, concentration, etc) varies with one or several other characteristics (sex, living area, expenditures, temperature, etc). Linear regression is an important and very common basic statistical tool for modelling dependences among variables. This course aims at providing an overview of basic statistics and linear regression methods. The focus is on the practical application of the methods, while providing a basic theoretical background. The computer exercises and project use the free R software/language and the commands needed to produce the desired output and answer the relevant statistical questions. Important questions studied during the course include: variables significance, model selection and validation, residual analysis, drawing conclusions from regression models and how to use the model for prediction.

The course is given as a subset of the 7 week advanced Masters level course *FMSN30/MASM22: Linear and Logistic Regression*.

Course Content

The course consists of lectures covering the theory behind linear regression combined with computer exercises and a project where R is used to solve practical regression problems.

The course covers: *Least squares and maximum-likelihood-methods; Multiple linear regression; hypothesis testing, confidence intervals and prediction; Methods for model validation: residuals, outliers, influential observations, multi-collinearity; Choice of regressors.*

Assessment

The grades will be determined by the quality of the final project. The grades are Fail, Pass and Pass with distinction.

Course Material

The course text book is:

Rawlings, J.O., Pantula, S.G., Dickey, D.A.: *Applied Regression Analysis - A Research Tool*, 2ed, Springer, available as [e-book](#),

For students at Lund, it is available as an e-book from the Lund University Library.

Additional course material will be available online through the course website.

Application

Some basic knowledge of statistics, e.g. *MASB11 Biostatistisk grundkurs* or similar, is recommended.

Places in the course are limited to 15, admission priority will be given to ClimBEco PhD students.

Application for the course will be on a first come basis. Apply by sending an email to Johan Lindström (johanl@maths.lth.se) before February 29th 2016.

Preliminary Schedule

21/3	13:15-15:00	Lecture 1
22/3	10:15-12:00	Lecture 2
22/3	13:15-15:00	Computer Exercise
29/3	13:15-15:00	Lecture 3
30/3	10:15-12:00	Lecture 4
31/3	10:15-12:00	Computer Exercise
4/4	13:15-15:00	Lecture 5
6/4	10:15-12:00	Lecture 6
7/4	10:15-12:00	Computer Exercise
11/4	13:15-15:00	Lecture 7
12/4	10:15-12:00	Lecture 8