Applied Statistical Regression – AS 2015

People:

Lecturer: **Dr. Marcel Dettling** (marcel.dettling@math.ethz.ch)

Assistants: **Christina Heinze** (heinze@stat.math.ethz.ch)
Preetam Nandy (nandy@stat.math.ethz.ch)
Alan Muro (muro@stat.math.ethz.ch)

Course Schedule:

Lectures will be held weekly at HG E1.2, on Mondays from 8.15-10.00. The contents will be explained by presenting slides and making notes on the blackboard.

Week	Date	L/E	Topics
01	14.09.2015	---	---
02	21.09.2015	L/L	**Linear Modeling & Smoothing**
03	28.09.2015	L/L	**Simple Linear Regression: Fitting & Inference**
04	05.10.2015	L/L	**Curvilinear Models, Variable Transformations**
05	12.10.2015	L/L	**Multiple Linear Regression: Model & Fitting**
06	19.10.2015	L/L	**Multiple Linear Regression: Inference & Prediction**
07	26.10.2015	L/L	**Extensions: Categorical Variables, Interactions**
08	02.11.2015	L/L	**Model Diagnostics: Standard Residual Plots**
09	09.11.2015	L/L	**Model Diagnostics: Advanced Techniques**
10	16.11.2015	L/L	**Multicollinearity & Variable Selection**
11	23.11.2015	L/L	**Modeling Strategies, Cross Validation**
12	30.11.2015	L/L	**Generalized Additive Modeling (GAM)**
13	07.12.2015	L/L	**Generalized Linear Modeling (GLM)**
14	14.12.2015	L/L	**Grouped Data, Poisson Regression**

Exercise Schedule:

Exercises will be held roughly bi-weekly, but on an irregular schedule, see below. They start on September 21, 2015 with an introduction to the statistical software package R, which is facultative for students already familiar with it. Please bring your own laptop for this introduction. Also further on, R will be used during the exercises so that you are expected to bring your laptop to the classes. The idea is that you work on the problems using the computer; the assistants will be there to give instructions and support. The main exercise slot is on Monday 10.15-11.55. Please check the course website for the location. If you have a valid reason for not being able to attend the exercise classes during this time slot, other slots (tentatively Monday 15.15-17.00 or Friday 10.15-11.55) might be offered. Please check back with Christina Heinze (heinze@stat.math.ethz.ch).

Series	Date	Topic	Hand-In	Discussion
01	21.09.2015	Data Analysis with R	---	21.09.2015
02	21.09.2015	Smoothing	28.09.2015	05.10.2015
03	05.10.2015	Simple Regression	19.10.2015	26.10.2015
04	26.10.2015	Multiple Regression	02.11.2015	02.11.2015
05	02.11.2015	Model Diagnostics	09.11.2015	16.11.2015
06	16.11.2015	Variable Selection	23.11.2015	30.11.2015
07	30.11.2015	Modeling Strategies	07.12.2015	14.12.2015
08	14.12.2015	GLM	---	14.12.2015

The solved exercises should be handed in at the end of the lecture of the due date or placed in the corresponding tray in HG J68 until 12.00am. Please note that only solutions to the exercises with your most important findings and answers shall be handed in, but no R script files and lengthy compilations of output or figures!

Software:

Theory and exercises will be based on the statistical software R. This is a freely available open source program that works on all platforms and has become worldwide standard for data analysis. It can be downloaded from CRAN (http://cran.r-project.org/). A good primer is the R tutorial that will be discussed in first exercise session of September 21, 2015. Further documentation on the use of R is available on CRAN, and in many textbooks.

Exercises and Exam:

There will be a written exam during the regular session that lasts for 120 minutes. It will be "Open Book", thus you are allowed to bring any written materials you wish. We also recommend bringing a pocket calculator. However, notebooks/computers are not allowed. In this regard, there will be some discrepancy between the exam and the exercises. The latter are mostly with R and try to best prepare you for applying regression analyses in scientific context. However, the paper and pencil exam will at best address your competency in applied regression and not be based on theory only.

The exercises will usually contain several problems and be rather long. They serve as a very important illustration to the theory that is presented in the lectures and thus, they are at the heart of learning how to work with regression analysis in practice. Furthermore, all exercises will be complemented by a detailed and commented sample solution. Because there are no conditions for obtaining the attendance certificate, it is completely up to you how many of the problems you want to solve. However, for becoming proficient in applying statistical regression techniques, thoroughly studying the exercises is key.

Written Material

There is a scriptum for this course. The scriptum, as well as the slides that are presented during the lectures, the exercise sheets, the sample solutions and some instructional datasets are also available for download from the course website at http://stat.ethz.ch/education/semesters/as2015/asr.

Credit Points:

There are no conditions for obtaining the attendance certificate. For obtaining the 5 ETCS credit points for this course, one needs to attend and pass the exam. PhD students who are after ETH credit points need to sign up with the lecturer at the beginning of the semester and hand in 5 well-solved exercises.

Literature:

1) **Linear Models with R**, Julian J. Faraway, Chapman & Hall/CRC (2005). ISBN-10: 1584884258. 229 pages, ca. 70$. There is a freely available version on CRAN, that is almost identical to the book, entitled **Practical Regression and Anova using R**: see http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf. For covering the topics from week 11 on, the second volume of Faraway's regression literature is required:

   Extending the Linear Model with R, Julian J. Faraway, Chapman & Hall/CRC (2006). ISBN-10: 158488424X. 312 pages, ca. 75$. There is no free online version.

2) **Applied Regression Analysis**, N. Draper and H. Smith, Wiley Interscience, 3rd Edition (1998). ISBN-10: 0471170828. 736 pages, ca. 100$.

3) **Introduction to Linear Regression Analysis**, D. Montgomery, E. Peck, G. Vining, Wiley-Interscience, 4th Edition (2006). ISBN-10: 0471754951. 640 pages, ca. 85$.

4) **Applied Regression Analysis and Generalized Linear Models**, J. Fox, Sage Publications, 2nd Edition (2008). ISBN-10: 0761930426. 688 pages, ca. 82$.