# Regression Analysis

Course Syllabus
Higher School of Economics, Department of Sociology

January – March 2012

## 1 Contact information

Instructor: Alexey Bessudnov
Email: bessudnov@gmail.com
Course webpage: http://sites.google.com/site/bessudnov/regressan
Office: 3 Kochnovsky proezd, room 419
Office hours: by appointment

## 2 Course overview and structure

The aim of this course is to provide students with an introduction to modern regression analysis. In the last decades regression analysis in its various modifications has become the major tool for statistical analysis in the social sciences, including economics, political science, and sociology. The knowledge of regression modelling is crucial for understanding modern social science literature and for conducting your own research. The course focuses on the discussion of linear and logit models that form the basis for more complicated analysis. In the end of the course you will be able to (1) understand and interpret linear and logit regression models presented in the literature, (2) estimate regression models using the statistical software Stata.

The course consists of 10 lectures and 10 seminars. The lectures focus on the explanation of statistical models. At the seminars you will learn how to conduct regression analysis in Stata.

## 3 Prerequisites

The course in general does not require mathematical knowledge beyond the high school level. However, students are expected to be comfortable with basic statistical concepts (such as mean, variance, standard error, confidence

interval, hypothesis testing) that are usually introduced in the introductory course on mathematical statistics and/or statistical data analysis.

The language of instruction is English. Students are expected to communicate in English in the class and prepare all the assignments in English.

No previous knowledge of Stata is required.

## 4 Course requirements

Course requirements consist of three parts.

1. Exercises. Students will be given home exercises after each class, except of the first (i.e., there will be nine exercises). Usually the exercises will require the statistical analysis in Stata of real data sets that will be available on the course webpage. The completion of an exercise should ideally take five or six hours every week (although this should be taken as an approximate estimate). Exercises must be handed in at the class next to the one when they were assigned. No late submissions are allowed.

2. First examination. An open book exam covers linear models introduced in the first part of the course.

3. Second examination. An open book exam covers logit models discussed in the second part of the course.

Although I do not check attendance, students are expected to attend all the lectures and seminars in order to successfully complete exercises and exams.

All assignments and exams are graded on the ten-point scale. The overall grade is a weighted average of the grades for exercises (50%) and two exams (25% each).

The HSE implements a strong anti-plagiarism policy. If I notice substantial overlaps in individual exercises or exam answers, both works will be graded zero. Disciplinary action will be taken.

## 5 Readings

1. A.Agresti and B.Finley. "Statistical Methods for the Social Sciences". Pearson, 2009. 4th ed. Ch.9-15.

2. A.Gelman and J.Hill. "Data Analysis Using Regression and Multilevel/ Hierarchical Models". Cambridge University Press, 2007. Ch.3-5, 9.

3. J.S.Long. "Regression Models for Categorical and Limited Dependent Variables". Sage, 1997. Ch.1-6.

4. J.S.Long and J.Freese. "Regression Models for Categorical Dependent Variables Using Stata". Stata Press, 2006. 2nd ed. Ch.1-6.

5. J.Fox. "Applied Regression Analysis and Generalised Linear Models". Sage, 2008. 2nd ed.

The textbooks are available in the university library. Additionally, all the slides from the lectures and seminars will be available on the course webpage.

## 6 Course outline

**Week 1.** Why regression? A gentle introduction to regression analysis (with examples and demonstrations).
Seminar: Introduction to Stata.
Fox, ch.1; Gelman/Hill, ch.2.

**Week 2.** Linear regression: the basics.
Agresti/Finley, ch.9-10; Gelman/Hill, ch.3; Fox, ch.2-6.

**Week 3.** Linear regression: nominal predictors, interactions, transformations, predicted values.
Agresti/Finley, ch.11,13; Gelman/Hill, ch.3-4; Fox, ch.5-7.

**Week 4.** Linear regression: assumptions and diagnostics.
Agresti/Finley, ch.14; Gelman/Hill, ch.3; Fox, ch.11-13.

**Mid-course exam.**

**Week 5.** Models for binary outcome variables. Linear probability model. Logit and probit.
Agresti/Finley, ch.15; Gelman/Hill, ch.5; Long, ch.3; Fox, ch.14.

**Week 6.** Logit models: odds ratios and predicted probabilities.
Agresti/Finley, ch.15; Gelman/Hill, ch.5; Long, ch.3; Long/Freese, ch.4; Fox, ch.14.

**Week 7.** Effective presentation of linear and logit models.
There are no readings for this week, but have a look at how models are presented in Gelman/Hill and Fox.

**Week 8.** Ordered logit.
Long, ch.5; Long/Freese, ch.5.

**Week 9.** Multinomial logit.
Long, ch.6; Long/Freese, ch.6.

**Week 10.** Causal inference and regression analysis.

Gelman/Hill, ch.9-10.

M.Gangl. 2010. "Causal inference in sociological research". *Annual Review of Sociology* 36: 21-47.

P.W.Holland. 1986. "Statistics and causal inference". *Journal of the American Statistical Association* 81(396): 945-960.

A.Abbott. 1998. "The causal devolution". *Sociological Methods and Research* 27(2): 148-181.

**Final exam.**