

Sciences Po  
Master of Public Affairs 2009-2010

Syllabus, Required Course, 2<sup>nd</sup> semester (5 credits)

## Statistical and data analysis for policymakers - Level 2

**Bruno Cautrès**

**Instructor :**

Dr. Bruno Cautrès  
CEVIPOF- Centre de recherches politiques de Sciences Po  
98 rue de l'Université  
75007 - PARIS  
[bruno.cautres@sciences-po.fr](mailto:bruno.cautres@sciences-po.fr)

**TA:**

Hyungsoo Woo  
[hyungsoo.woo@sciences-po.fr](mailto:hyungsoo.woo@sciences-po.fr)

**Course Description:**

This Level 2 course is an advanced course on multivariate statistical methods that may be of use for reading advanced journals and using statistical methods in the field of quantitative social sciences (mainly sociology, political science, and economics). The course also aims at making students more effective users of statistical tools in analyzing public policy problems and to give diagnostics about policy solutions. The interaction between statistics, policy analysis and decision-making will be highlighted. An essential part of the course is that students can work on data coming from major data bases of interest from the MPA perspective.

**Assessment:**

The course is evaluated on the following basis :

*Take-home assessment tasks* : you will have 2 assignments, with 2 weeks to complete each of the assignments: 50 % of your grade. One assignment will cover the first sessions, another one will cover the last ones. The assignments will be individual work and will be short. Each assignment will account for 10% of the final grade.

*A mid-term exam*, organized after session 5, which accounts for 30%.

*A final research paper*, done in groups of three students. The research paper will challenge the students to look for, work with, and explore real public policy/political questions. This group work consists of a paper of about 15 pages, with a high level of presentation and data analysis. It must closely resemble a research or policy paper: hypotheses, presentation of data and methods, statistical data analysis and implications of results for policy/political choice and decision-making. Students are responsible for researching the appropriate data (relating to a policy question) and

conducting the necessary statistical analysis. The instructor and TA will also present and introduce existing relevant data bases.. It must be clear that students must organize themselves immediately into groups and ask for regular meetings with both the instructor and TA about their data and work. The instructor will help them in accessing relevant data bases, in connection with the Sciences Po facilities (there is a Centre for Socio-Political Data, which is a member of CESSDA and ICPSR networks in data sharing and archives). The paper is assessed by groups, which means that there is only one grade per group. In case of difficulties in sharing the work, students should talk to their instructor. A solution will be found.

This paper should be delivered at the end of the semester (a week before the end) and should focus on a public policy/political or political economy issue or problem. It accounts for 50% of your final grade.

### **Monitoring of the course and of the students' work:**

A short meeting between student delegate, the instructor and/or TA and the MPA administration will be organized every two courses, to ensure the quality of the interaction between the course and the students.

### **Textbook and software:**

There are several textbooks that may be of use for advanced methods but two books are excellent and instructive:

Alan Agresti, Barbara Finlay. *Statistical methods for social sciences*. Prentice Hall, 1997 (3<sup>rd</sup> edition, which is the last one).

Damodar N. Gujarati. 2003. *Basic Econometrics*, 4<sup>th</sup> edition. NY: McGraw-Hill/Irwin.

Gujarati covers a more complete picture of multivariate techniques used in econometrics while Agresti/Finlay is a little bit more easy to read. In 2009/2010, Gujarati will be the reference textbook; a reading pack will be available, made of large extracts from the Gujarati textbook. In any case, the instructor will provide an electronic version of the book that may complete the reading pack. This book meets international standards, and is used quite extensively in most US universities.

Every session will be based on a PowerPoint presentation, made available to students through an ENTG workgroup.

Statistical software used for this course will be Stata.

**Course outline:**

**Session 1 : Introduction to the course**

**Session 2 : The bivariate statistical association : chi-square for tables and correlation for interval distributions.**

**Session 3 – The linear regression model : theory and basics.**

**Session 4- OLS estimates and their properties : how the linear model functions ?**

**Session 5 : Goodness of fit and statistical inference : how to evaluate the quality of a model ?**

**Session 6 : Multiple regression analysis : the logic of controlling for and ‘ceteris paribus’**

**Session 7 : Goodness of fit, statistical inference, models comparisons : how to choose a good model ?**

**Session 8 : Dummies variables, interactions, nonlinearities**

**Session 9 : Panel data and the differences-in-differences regression analysis : a statistical method for public policy analysis**

**Session 10 : Instrumental variables**

**Session 11 : Qualitative dependent variable I : the logistic regression model for binomial distribution.**

**Session 12 : Qualitative dependent variable II : the logistic regression model for multinomial distribution.**