

PO7001 Quantitative Methods I for Political Science

(a.k.a. UCD POL40030 Introduction to Statistics)

Trinity College Dublin, Autumn 2009

Professor Kenneth Benoit

<http://kenbenoit.net/quant1/>

Version: September 17, 2009

1 Objectives and Learning Outcomes

In this course you learn the basics of statistical analysis. Using the freely available [R statistical package](#), we go step by step from describing simple and more complex data, to issues of random samples, to the types and requirements of statistical inference, and finally to linear statistical models. In the end you should be comfortable with R and be able to perform basic regression analyses. Instead of focusing on the many methods and issues related to regression analysis, however, the focus is on the fundamental logic of statistical description and inference. The course thus forms a good foundation for further training in statistical modelling.

You will learn how to use the free, multi-platform statistical software package [R](#), a powerful, invaluable tool that will serve you well for quantitative analysis and graphical presentation of data at any level.

This course is primarily about data analysis and on introducing the linear model. The focus is on practice, and this focus is reflected in the choice of texts and in the emphasis on applied coursework. The emphasis is on developing the ability to use quantitative analysis techniques when faced with the need in practical research. Consequently the learning method combines lectures and reading with hands-on statistical programming exercises using real datasets.

The learning outcomes associated with this nine-week course are aimed at students being able to:

- Understand data concepts and basic descriptive quantitative analysis tools;
- Work with real datasets to perform basic quantitative analyses;
- Graph data effectively for presentation and analysis;
- Recognize and understand the basics of the linear regression model;
- Use the R statistical software package for analyzing and graphing data;
- Understand sufficient theoretical and practical material to build on in a second, more advanced quantitative methods course.

2 Prerequisites

There is formally no prerequisite for this course except an open mind. It will help a great deal, however, if you have a basic knowledge of mathematics, in particular algebra. Also since the practice component is done in the [R statistical package](#), the more comfortable you are with computers, the better, since R involves minor programming, manipulation of data files, and editing text files.

3 Logistics

Meetings. Classes will meet nine weeks for one session per week, on Wednesdays from 9-12:00 in College Green. The class will be mostly lectures and presentations by me, with the rest devoted to practical data analysis relating to weekly problem sets. For this reason I encourage students to bring their laptop computers to class, although this is not an integral requirement. (Since electrical outlets are limited in the classroom, please have your batteries charged ahead of time!)

Computer Software. For all applied work, we will use R, which may be downloaded from <http://www.r-project.org> (see also Verzani, 2005, Appendix A). You should download and install this right away, so that you can get as much hands-on practice as possible. A nice short video tutorial of R can be found at <http://www.decisionsciencenews.com/?p=261>.

Grading. Grading will be based on three components.

1. **Problem sets: 50%.** The only way to properly learn statistics is by hands-on training. You will need to work with actual data and produce your own statistical analysis - just the theory will never be sufficient. For that reason, a substantial part of the grading will be based on regular homework assignments. In total, there will be five problem sets each counting for 10% of your final grade. These will be made available on-line on Wednesdays (when assigned) and must be submitted to the class page at <http://turnitin.com> before class the following Wednesday. Each problem set will consist of a number of problems combining computer analysis with interpretation and analytical problems. Computer output, when supplied, should include both the commands used as well as results. Computer results should be indicated clearly. We strongly encourage you to submit your homework as a single pdf file.
2. **Course paper: 50%.** The second part of the grade will be based on a course paper, which is due two weeks after the last class. For this paper you will be required to produce your own regression analysis, in a subject of your choice, and present the results and interpretation. Before the 5th class, you will have to submit a short (maximum one page) description of a plan for a paper, describing the key research question and where or how you think you will find the necessary data. This will allow me to give feedback before you dive into the details of your paper. The final paper will count as fifty percent of your final grade for the course. The course paper will be due **16 December 2009**, and should be submitted as per the homework assignments to <http://turnitin.com>.
3. **Examination: 0%.** There no exam for this course.

Reading materials are described separately below.

The **home page** for this course will be http://www.kenbenoit.net/?page_id=509 and you will be able to find this course handout there, links to exercises, and additional material. In fact you are encouraged to shred this document since the on-line version will always be more current, and have more current links as well (to notes, slides, readings, and exercises).

I have set up a **calendar** for the course that you can subscribe to from the web page that has all of the dates and topics on it. This calendar can be subscribed to using a calendar client from webcal://ical.me.com/kbenoit/P07001%20Quant%201.ics or alternatively, viewed directly using a browser via <http://ical.me.com/kbenoit/P07001%20Quant%201>. One important date to note is that there is a one week break between weeks 8 and 9, because of Trinity's traditional autumn "reading week", so that there will be no class on November 11.

4 Texts

This course will assign a variety of reading materials, some essential and some supplementary. Readings are absolutely *central* to this course and you will not learn anything if you attempt to rely on the lectures alone.

The key textbook of this course is [Verzani \(2005\)](#), which I recommend you purchase. This textbook provides an excellent introduction to statistics and to R, but is not specifically written for social science students.¹ For more targeted textbooks, or simply to have some additional sources to help you understand the material, you might want to check out [Agresti and Finlay \(1997\)](#), [Agresti and Franklin \(2007\)](#) or [Healey \(2005\)](#). These books generally spend somewhat more time describing basic statistical tests, but you will still need [Verzani \(2005\)](#) for the chapters on probability distributions and simulations and for the introduction to R. The homework assignments will be based on social science examples. For the sections of regression analysis, a further useful book, with more explanation although at times slightly too technical, is [Kutner et al. \(2005\)](#). The relevant part of this book is equivalent to [Kutner, Nachtsheim and Neter \(2004\)](#), so you can check either in the library. A final text that I cannot recommend more highly is [Gelman and Hill \(2007\)](#), a relatively advanced but meticulously thorough treatment of regression analysis, starting from very basic and ending with very advanced topics.

You may find some of the readings difficult or uncomfortable. This is completely normal. Your response should not be avoidance but rather a renewed effort to understand the material by (1) reading it with even greater care, (2) rereading it several times, (3) seeking other readings that might make the primary texts more comprehensible, and (4) working with other students in study groups. It is also perfectly normal in methods classes that you do not absorb all a text has to offer upon the first reading, but rather return to it several times over the years and learn new things as your knowledge accumulates.

Other sources will be available on-line through the web page for this course, including the HTML version of this document. In the HTML syllabus, for instance, many of the articles may be viewed or downloaded from the web site. (Class exercises will also be on-line.)

5 Schedule

1. Introduction and Dealing with Data

Required Reading: [Verzani \(2005, Ch 1\)](#)

Recommended Reading: [Crawley \(2005, Ch 2\)](#)

Homework: TBA

2. Univariate Data

Required Reading: [Verzani \(2005, Ch 2\)](#)

Recommended Reading: [Crawley \(2005, Ch 3–5\)](#)

Homework: TBA

3. Bi- and Multivariate Distributions

¹Although it is recommended that you acquire the book, if you have trouble finding the book quick enough, you can download an earlier version online from <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>. Please note that this version is not as extensive and not as clear as the book, so you might not want to rely on it entirely. The book chapters also do not match exactly.

Required Reading: Verzani (2005, Chs 3–4)
Recommended Reading: Crawley (2005, Ch 6)
Homework: TBA

4. **Probability Distributions**

Required Reading: Verzani (2005, Ch 5)
Homework: TBA

5. **Simulations**

Required Reading: Verzani (2005, Ch 6)
Homework: TBA

6. **Confidence Intervals and Probability Testing**

Required Reading: Verzani (2005, Chs 7–8)
Homework: TBA

7. **Goodness of Fit (χ^2)**

Required Reading: Verzani (2005, Ch 10)
Homework: TBA

8. **Linear Regression I: Basic Model**

Required Reading: Verzani (2005, Ch 10)
Recommended Reading: Crawley (2005, Ch 8); Kutner et al. (2005, ch 1)
Homework: TBA

9. **Linear Regression II: Diagnostics, Interaction Terms**

Required Reading: Verzani (2005, Ch 10); Brambor, Clark and Golder (2006)
Recommended Reading: Kutner et al. (2005, Chs 6, 8.2, 8.5)
Homework: TBA

10. **Analysis of Variance (ANOVA)**

Required Reading: Verzani (2005, Ch 11)
Recommended Reading: Crawley (2005, Ch 9–10)
Homework: TBA

References

- Agresti, Alan and Barbara Finlay. 1997. *Statistical methods for the social sciences*. Prentice Hall.
- Agresti, Alan and Christine Franklin. 2007. *Statistics: the art and science of learning from data*. New Jersey: Pearson/Prentice Hall.
- Brambor, Thomas, William Roberts Clark and Matt Golder. 2006. "Understanding interaction models: improving empirical analyses." *Political Analysis* 14(1):63–82.
- Crawley, Michael J. 2005. *Statistics: An Introduction Using R*. Colchester: John Wiley & Sons, Ltd.
- Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Analytical Methods for Social Research Cambridge: Cambridge University Press.
- Healey, Joseph F. 2005. *Statistics: a tool for social research*. 7th ed. Wadsworth.

Kutner, Michael H., Christopher J. Nachtsheim and John Neter. 2004. *Applied linear regression models*. McGraw-Hill.

Kutner, Michael H., Christopher J. Nachtsheim, John Neter and William Li. 2005. *Applied linear statistical models*. 5th ed. McGraw-Hill.

Verzani, John. 2005. *Using R for introductory statistics*. Boca Raton, FL: Chapman & Hall/CRC.