

Leiden University | Political Science

Quantitative Data Collection and Analysis

Research Master in Political Science & Public Administration
Semester 2, Block 4, 5 ECTS, Level 600, Course Code: 6446QNDCA
Tuesday, April 1 - May 20, 2014, 9:00-11:00, Room SA31

Instructor: Michael F. Meffert
Office: 5B11 | Office Hours: by appointment
071-527-3862 | m.f.meffert@fsw.leidenuniv.nl

Course Description

The aim of this course is to offer a short introduction to research designs for testing causal hypotheses, in particular experiments, and to review and apply basic statistical techniques such as principle components analysis (PCA), ANOVA, OLS and logistic regression that are commonly used in political science and public administration research. The course builds upon the knowledge gained in introductory statistics courses and further develops students' analytical skills in a series of practical assignments, with special attention to checking statistical assumptions and the interpretation and presentation of results.

Method of Instruction

Lectures, discussion, and assignments.

Required Readings

Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. London: Sage

Additional journal articles or book chapters (see *Course Schedule*).

Optional Readings/References

Kellstedt, Paul M., and Guy D. Whitten. 2009. *The Fundamentals of Political Science Research*. Cambridge: Cambridge University Press.

- A basic and general introduction to quantitative research methods / linear regression.

Box-Steffensmeier, Janet M., Henry E. Brady, and David Collier. 2008. *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.

- A very useful handbook that provides short and highly accessible introductions to a wide variety of research designs and statistical topics.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

- A more advanced but still accessible statistics book (using R)

Kennedy, Peter. 2008. *A Guide to Econometrics*. 6th ed. Malden, MA: Blackwell.

- A more advanced but accessible discussion of key issues in econometrics.

Personality Project, <https://personality-project.org/>, Sections “Psychometric Theory” and “R Guide”.

- Excellent textbook (draft) and introduction to R (and library “psych”).

Muenchen, Robert A. 2011. *R for SAS and SPSS Users*. New York: Springer

- Helpful for prolific SPSS users. Available at SpringerLink: <http://link.springer.com/>

Software

The course will utilize the software package R (current version: 3.0.3, 2014-03-06), which is available at: <http://cran.r-project.org/>. The functions of the basic software package can be extended by downloading (once) and loading (always when needed) additional 'libraries'. An overview of important packages can be found under <http://cran.r-project.org/web/views/>. In (the likely) case that the current version of R is not and cannot be installed on university computers through 'official' channels, the software can be downloaded and installed on a USB stick or directly on the desktop (the USB stick approach has the advantage that the additional libraries remain available at all times). Note that plenty of introductions to R are available, both books and websites, but one of the best remains the official *An Introduction to R* at: <http://www.cran.r-project.org/doc/manuals/R-intro.pdf>

Assignments, Research Paper, and Participation

The main focus of this course is on building practical (statistical) data analysis skills. For this reason, the main workload is a series of individual homework assignments with (mostly) statistical analyses. The final grade is based on five assignments, a research paper, and class participation.

5 Assignments:	60% (12% each)
Research Paper:	30%
Participation:	10%

Assignments. Students are expected to complete 5 short (!) individual homework assignments during the course. With the exception of the first assignment, the assignments involve the application of a statistical technique using data of their own choice (such as NKO data). The statistical assignments have two components: (1) a written justification, description, summary, and interpretation of an analysis using a specific statistical technique (with the R script/syntax that re-produces and replicates the analyses in the appendix) and (2) a short PowerPoint presentation that summarizes the key findings for potential presentation in class. The assignments need to be submitted electronically via Ephorus and as hard copy in class by the given date and time (usually at the beginning of class). The PowerPoint presentation should usually be submitted the day before class. Late submissions are *not* accepted.* The specific topics of the five assignments are:

- 1) Research Design: Randomized Experiment
- 2) Exploratory Data Analysis: Principle Components Analysis (PCA)
- 3) Hypothesis Testing of Group Differences: t Test *or* ANOVA
- 4) Hypothesis Testing of Causal Relationships (1): Multiple Regression (OLS)
- 5) Hypothesis Testing of Causal Relationships (2): Logistic Regression

During the course, handouts with the specific requirements for each assignment will be available on Blackboard.

Research Paper. At the end of the course, students will write an individual research paper (ca. 3000 words) which can take one of two forms, (1) an empirical research paper or (2) a more detailed (quantitative) research proposal.

Option 1: For the empirical research paper, the basic idea is to choose a theory-based research question, do a short literature review and present a testable hypothesis, then use the appropriate data and statistical technique to test this hypothesis, and report the results. It will usually take the following format:

- Introduction
- Theory and Review (incl. conceptualization of variables & hypothesis)
- Methods (incl. operationalization of variables, sampling etc.)
- Results (incl. descriptive statistics & statistical test)
- Discussion and Conclusion

Note that it is possible to use one of the homework assignments as a starting point and to develop it further into a research paper.

Option 2: For the research proposal, the idea is to choose a theory-based research question, do a short literature review and present a testable hypothesis, and then propose a

more detailed quantitative research design that would allow you to test this hypothesis, e.g. an experiment. It will usually take the following format:

- Introduction (incl. scientific and social relevance)
- Theory and Review (incl. conceptualization of variables & hypothesis)
- Methods (incl. operationalization of variables, specific research design, sampling, data collection)
- Planned analyses

In either case, a short, one-page proposal for the research paper is due **May 6, 2014**. The final research paper is due on Monday, **May 26, 2014**. In both cases, late submissions are not accepted.*

***Deadlines.** Assignment and paper deadlines are final and late submissions are not accepted. Properly documented emergencies and *in advance* requested and permitted extensions are exempt from this rule.

Participation. The seminar requires active and informed participation of the students in class discussions. Students are expected to read the assigned readings before each class meeting and be prepared to present their homework assignments. Class attendance is mandatory. Students who miss more than one class will automatically fail the course. Properly documented emergencies and in advance requested and permitted absences are exempt from this rule.

If you have a physical, psychological, medical, or learning disability that may impact on your ability to carry out the assigned course work, please contact the staff in the Department of Political Science. All information and documentation of disability is confidential.

Plagiarism

Plagiarism is understood as presenting, intentionally or otherwise, someone else's words, thoughts, analyses, argumentations, pictures, techniques, computer programs, etc., as your own work. Plagiarism is not allowed and has serious consequences. Students must be familiar with Leiden University's rules about plagiarism. They are available at:

<http://www.regulations.leiden.edu/education-students/plagiarism.html>

The departmental rules and procedures with regard to plagiarism can be found at:

<http://www.socialsciences.leiden.edu/politicalscience/students/postgraduate/regulations/plagiarism.html>

Important note: Plagiarism occurs in both of the following situations:

- Quoting work from other (and outside) sources without attribution;
- Copying the work of others when completing individual assignments.

Course Schedule

April 1, 2014	Course Introduction & Short Review of R	[SA35]
----------------------	--	---------------

Methods

Field, Ch. 1-3 (Introductory Chapters).

April 8, 2014	Research Design: Causality & Experiments	[SA31]
----------------------	---	---------------

Key Concepts: Causal Theories, Surveys, Experiments

Methods

McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5: 31-61.

Gaines, Brian J., James H. Kuklinski, and Paul J. Quirk. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15 (1):1-20.

Field, Ch. 4 (Exploring Data with Graphs) & 5 (Exploring Assumptions)

Applications

Tomz, Michael, and Robert P. Van Houweling. 2008. "Candidate Positioning and Voter Choice." *American Political Science Review* 102 (3): 303-318.

Further/Recommended Literature

Goodin, Robert E., Werner Güth, and Rupert Sausgruber. 2008. "When to Coalesce: Early Versus Late Coalition Announcements in an Experimental Democracy." *British Journal of Political Science* 38 (1): 181-191.

Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.

>> Assignment 1: Research Design - Randomized Experiment (due April 15, 2014)

April 15, 2014	Measurement & Describing Data (1)	[SA31]
-----------------------	--	---------------

Key Concepts: Operationalization, Reliability & Validity, Visual Data Analysis, Principle Components Analysis (PCA)

Methods

Field, Ch. 17 (Exploratory Factor Analysis).

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529-546.

Kritzer, Herbert M. 1996. "The Data Puzzle: The Nature of Interpretation in Quantitative Research." *American Journal of Political Science* 40 (1): 1-32.

Kastellec, Jonathan P., and Eduardo L. Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5 (4): 755-771.
[See also: <http://tables2graphs.com>]

Recommended

Epstein, Lee, Andrew D. Martin, and Matthew M. Schneider. 2006. "On the Effective Communication of the Results of Empirical Studies, Part I." *Vanderbilt Law Review* 59: 1811-1872.

Epstein, Lee, Andrew D. Martin, and Christina L. Boyd. 2007. "On the Effective Communication of the Results of Empirical Studies, Part II." *Vanderbilt Law Review* 60: 799-846.

>> Assignment 2: Exploratory Data Analysis - Factor Analysis (due April 22, 2014)

April 22, 2014	Statistical Inference & Significance Testing	[SA31]
-----------------------	---	---------------

Key Concepts: Statistical Tests, t Test, ANOVA

Methods

Field, Ch. 9 (Comparing Two Means) & 10 (Comparing Several Means: ANOVA).

Cumming, Geoff, and Sue Finch. 2005. "Inference by Eye: Confidence Intervals and How to Read Pictures of Data." *American Psychologist* 60 (2): 170-180.

Cohen, Jacob. 1994. "The Earth is Round ($p < .05$)." *American Psychologist* 49 (12): 997-1003.

Recommended

Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52 (3): 647-674.

>> Assignment 3: Hypothesis Testing of Group Differences - t Test or ANOVA (due April 29, 2014)

April 29, 2014 **Linear Regression (1) - Basics** **[SA31]**

Key Concepts: Correlation, Bivariate & Multivariate OLS Regression, Assumptions

Methods

Field, Ch. 6 (Correlation) & 7 (Regression).

Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology*: 291-313.

King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30 (3) 666-687.

Downs, George W., and David M. Rocke. 1979. "Interpreting Heteroscedasticity." *American Journal of Political Science* 23 (4): 816-828.

May 6, 2014 **Linear Regression (2) - Interpretation & Interactions** **[SA31]**

>> Deadline: Proposal for Research Paper (single page, as hardcopy in class)

Key Concepts: Interactions

Methods

Baron, Reuben M., and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6): 1173-1182.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63-82.

Recommended

Berry, William D., Matt Golder, and Daniel Milton. 2012. "Improving Tests of Theories Positing Interaction." *Journal of Politics* 74 (3): 653-671.

Friedrich, Robert J. 1982. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science* 26 (4): 797-833.

Aiken, Leona S., and Stephen G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage.

Hayes, Andrew F. 2013. *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach*. New York: Guilford Press.

>> Assignment 4: Multiple Regression (OLS) (due May 13, 2014)

May 13, 2014	Logistic Regression	[SA31]
---------------------	----------------------------	---------------

Key Concepts: Linear Probability Model, Maximum Likelihood, Logistic Regression

Methods

Review/How-To: Field, Ch. 8 (Logistic Regression).

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 347-361.

Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57 (1): 263–277.

Recommended

Imai, Kosuke, Gary King, and Olivia Lau. 2012. *Zelig: Everyone's Statistical Software*. Core Reference Manual. Available at: http://r.iq.harvard.edu/docs/zelig/4.1-3/zelig_core_reference.pdf

Zelig Homepage: <http://projects.iq.harvard.edu/zelig>

Menard, Scott. 1995. *Applied Logistic Regression Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA: Sage.

>> Assignment 5: Logistic Regression (due May 20, 2014)

May 20, 2014	Measurement & Describing Data (2)	[SA31]
---------------------	--	---------------

Key Concepts: Multidimensional Scaling (MDS)

Methods

Bartholomew, David J., Fiona Steele, Irimi Moustaki, and Jane I. Galbraith. 2008. *Analysis of Multivariate Social Science Data*. 2nd ed. Boca Raton, FL: CRC Press, Chapter 3: Multidimensional Scaling (pp. 55-81)

Weisberg, Herbert F. 1974. "Dimensionland: An Excursion into Spaces." *American Journal of Political Science* 18 (4): 743-776.

Recommended

Kruskal, Joseph B., and Myron Wish. 1978. *Multidimensional Scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills, CA: Sage.

Jacoby, William G., and David A. Armstrong. 2014. "Bootstrap Confidence Regions for Multidimensional Scaling Solutions." *American Journal of Political Science* 58 (1): 264–278.

May 26, 2014 **Research Paper due**

Appendix – Additional Topics

Linear Regression (3) – Nonrecursive Models (2SLS)

Key Concepts: Reciprocal Causation, Nonrecursive Models, Two Stage Least Squares

Methods

James, Lawrence R., and B. Krishna Singh. 1978. "An Introduction to the Logic, Assumptions, and Basic Analytic Procedures of Two-Stage Least Squares." *Psychological Bulletin* 85 (5): 1104-1122.

Application

Aarts, Kees, and Holli A. Semetko. 2003. "The Divided Electorate: Media Use and Political Involvement." *Journal of Politics* 65 (3): 759-784.

Recommended

Berry, William D. 1984. *Nonrecursive Causal Models*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-037. Newbury Park, CA: Sage.

Bartels, Larry M. 1991. "Instrumental and 'Quasi-Instrumental' Variables." *American Journal of Political Science* 35 (3): 777-800.

Sovey, Allison J., and Donald P. Green. 2011. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science* 55 (1): 188-200.

Postscript – Final Recommendation

Schrodt, Philip A. 2014. "Seven Deadly Sins of Contemporary Quantitative Political Analysis." *Journal of Peace Research* 51 (2): 287–300.