

**Research Methods II**  
SYA 6306  
Spring 2005  
Dr. Richard Tardanico  
<http://www.fiu.edu/~tardanico>

---

### Getting in touch

[Richard.Tardanico@fiu.edu](mailto:Richard.Tardanico@fiu.edu)

Tel: 305/348-2247

Office hours: Wednesdays, 12:30am-1:30pm; or by appointment

Course web site: <http://www.fiu.edu/~tardanico>

Web resources for learning *Stata-9*: <http://www.ats.ucla.edu/stat/seminars>  
<http://www.ats.ucla.edu/stat/>

---

### What it's about

This course explores how to analyze social and policy data by means of multiple regression—a statistical method for examining how an outcome variable depends on several explanatory variables—which is a basic tool for quantitative social and policy research. The course's prerequisite is SYA 6305 (Research Methods I) or an equivalent course.

While this course focuses on quantitative data analysis, throughout the semester we will discuss the pro's and con's of quantitative versus qualitative research methods (see the various questions posed at the end of this syllabus). As Einstein said, "Not everything that can be counted counts and not everything that counts can be counted."

### Required books

The required books are: Mendenhall & Sincich, *A Second Course in Statistics: Regression Analysis*, 6<sup>th</sup> ed. (Prentice Hall, 2003); Allison, *Multiple Regression: A Primer* (Pine Forge, 1999); Pampel, *Logistic Regression: A Primer* (Pine Forge, 2000); Long & Freese, *Regression Models for Categorical Dependent Variables Using Stata*; and King et al., *Designing Social Inquiry*.

There will also be readings from articles and from the following books: Ragin, *Constructing Social Research*; Moore & McCabe, *Introduction to the Practice of Statistics*; and Agresti & Finlay, *Statistical Methods for the Social Sciences*.

Recommended is Hamilton, *Statistics with Stata – Updated for Version 9*.

### Required software

The previous semester's materials presented an introduction to Stata, the statistical software program that will be integral to this course. For a comparison of Stata, SPSS, and SAS, see the following report:

[www.ats.ucla.edu/stat/stata/faq/compare\\_packages.htm](http://www.ats.ucla.edu/stat/stata/faq/compare_packages.htm)

Students will be expected to have routine access to a professional version of *Stata-9*. We will continue to cover the basics of using *Stata* via material located on the course web site: <http://www.fiu.edu/~tardanic>

We will also learn how to use the program via the superb free, downloadable resources for learning *Stata* (as well as SPSS, SAS, and other statistical programs) that are available at the following web sites of UCLA-Academic Technology Services: <http://www.ats.ucla.edu/stat/>  
<http://www.ats.ucla.edu/stat/seminars/>  
<http://www.ats.ucla.edu/stat/stata/>  
<http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter1/statareg1.htm>  
See the modules *Learning Stata I & II* and *Training Videos for Stata*.

Other excellent resources are Hamilton, *Statistics with Stata* and Long & Freese, *Regression Models for Categorical Dependent Variables Using Stata*.

The course's objective, however, is learning regression analysis in the context of debates concerning the various approaches—quantitative and qualitative—to social and policy research. Learning statistical software is subordinate to this pursuit.

### **Homework, mini-projects & grades**

Final grades will be based on the following:

- *More or less weekly homework problems. **These will be graded pass/fail and collectively worth 15% of the final grade.*** Passing all such assignments earns an "A," all but one of them a "B," and so on. Homework problems assigned at a given class session are due at the start of the next class session.
  - *Five mini-projects. **Each of these will be worth 15% of the final grade (for a total of 75% of the final grade).*** Each mini-project will combine assigned problems and questions with regression analysis of a data set. Every student will select in consultation with the instructor a data set to analyze for mini-projects #1-4 (on regression with a quantitative dependent variable). Students will either continue with the same data set or select another one for mini-project #5 (on regression with a categorical dependent variable).
  - *Mandatory attendance of Sociology & Anthropology graduate students at departmental colloquia, **worth 10% of the final grade.***
- 

### **Tentative Schedule**

#### **Jan. 12: What is regression analysis?**

- Mendenhall & Sincich, chapter 2, "Introduction to Regression Analysis"
- Mendenhall & Sincich, chapter 3, "Simple Linear Regression"
- Ragin, *Constructing Social Research*, chapter 6, "Using Quantitative Methods to Study Covariation" & "Afterword"
- King et al., *Designing Social Research*, chapter 1, "The Science in Social Science"; and Chapter 2, "Descriptive Inference"

- *Recommended:* Mendenhall & Sincich, chapter 1, "A Review of Basic Concepts"; Allison, chapter 5, "Bivariate Regression"; and "Review Concepts" on course web site
- *Assignment:* (1) Go to the Stata website (by clicking the link on the *Stata* program) and subscribe to the *Stata listserv digest*. (2) Select an article using OLS regression from either *American Sociological Review*, *American Journal of Sociology*, or *Social Forces* since the year 2000. Briefly summarize how the research problem is conceptualized and how the data are described and analyzed. Then copy (by pen or pencil) the format of the main tables that describe the data and report the regression results. (3) Download the "Preparation checklist for ASR Manuscripts" and re-type the following: the guidelines for each type of in-text citation and one of each type of reference example. (4) Re-type "Organizing Your Computer Work for Serious Data Analysis" at the end of this syllabus. (5) Use StatTransfer to transfer the data from one of the Mendenhall & Sincich problems into Stata. (6) Do Stata assignments #1 ("Estimate tables") and #2 ("Log files"). (7) Mendenhall & Sincich problems 3.9, 3.11, 3.17b, 3.25, 3.32, 3.34, 3.44, 3.45, 3.48, 3.62, 3.68, 3.74, 3.75.

### **Jan. 19 & 26: Multiple regression**

- Allison, chapter 1, "What Is Multiple Regression?"
- Agresti & Finlay, *Statistical Methods for the Social Sciences*, chapter 10, "Introduction to Multivariate Relationships"
- King et al., *Designing Social Research*, chapter 3, "Causality and Causal Inference"
- Allison, chapter 2, "How Do I Interpret Multiple Regression Results?"
- Mendenhall & Sincich, chapter 4, "Multiple Regression"
- *Assignment:* (1) Do Stata assignment #3 ("Create variables"). (2) Mendenhall & Sincich problems, week #1 - 4.1, 4.4, 4.13, 4.14, 4.15, 4.21, 4.22, 4.25, 4.26; week #2 - 4.31, 4.35, 4.41, 4.42, 4.43, 4.55, 4.57, 4.59, 4.61, 4.62, 4.68, 4.79, 4.85.
- *Mini-project #1 assigned on Jan. 19: due in class on Feb. 9.*

### **Feb. 2 & Feb. 9: Model building**

- Allison, chapter 3, "What Can Go Wrong with Multiple Regression?"
- Allison, chapter 8, "How Can Multiple Regression Handle Nonlinear Relationships?"
- Mendenhall & Sincich, chapter 5, "Model Building"
- *Assignment:* (1) week #1 - Do Stata assignments #4 ("Scatterplots"), #5 ("Outreg"), & #6 ("Log transforms"). (2) Mendenhall & Sincich problems, week #1 - 5.1, 5.4, 5.6, 5.9, 5.14; week #2 - 5.17 (using Stata to do the calculations), 5.19, 5.21, 5.25, 5.31, 5.32, 5.37, 5.38, 5.43.
- *Mini-project #1 due in class on Feb. 9.*

### **Feb. 16: Selecting variables**

- Mendenhall & Sincich, chapter 6, "Variable Screening Methods"
- King et al., *Designing Social Research*, chapter 4, "Determining What to Observe"
- Scott Long & Jeremy Freese, "Replication Standards for Quantitative Social Science: Why Not Sociology?"  
<http://www.jeremyfreese.com/docs/freese-reproducibility-webdraft.pdf>

- *Assignment:* (1) Do Stata assignment #7 (“Do-files”). (2) Mendenhall & Sincich problems 6.1, 6.5.
- *Mini-project #2 assigned: due in class on Mar. 2.*

### **Feb. 23: Regression diagnostics, part 1**

- Allison, chapter 6, “What Are the Assumptions of Multiple Regression?”; and chapter 7, “What Can Be Done about Multicollinearity?”
- Mendenhall & Sincich, chapter 7, “Some Regression Pitfalls”
- King et al., *Designing Social Inquiry*, chapter 5, “Understanding What to Avoid”
- *Assignment:* (1) Do Stata assignment #8 (“Diagnostics A”). (2) Mendenhall & Sincich problems 7.1, 7.2, 7.3, 7.4, 7.5, 7.7, 7.13, 7.14, 7.20.

### **Mar. 2: Regression diagnostics, part 2**

- Mendenhall & Sincich, chapter 8, “Residual Analysis”
- *Assignment:* (1) Do Stata assignment #9 (“Diagnostics B”). (2) Mendenhall & Sincich problems 8.1, 8.2, 8.3, 8.7, 8.8, 8.13, 8.18, 8.19, 8.20, 8.23, 8.25.
- *Mini-project #2 due in class.*
- *Mini-project #3 assigned: due in class on Mar. 16.*

### **Mar. 9 & 16: Regression with categorical dependent variables—binary dependent variables**

- Moore & McCabe, chapter 15, “The Logistic Regression Model”
- Pampel, chapter 1, “The Logic of Logistic Regression”
- Pampel, chapter 2, “Interpreting Logistic Regression Coefficients”
- Pampel, chapter 3, “Estimation and Model Fit”
- Long & Freese, chapter 4, “Models for Binary Outcomes”
- Hoetker, “The Use of Logit and Probit Models in Strategic Management Research: Critical Issues.” *Strategic Management Journal* (forthcoming) [http://www.business.uiuc.edu/ghoetker/documents/Hoetker\\_logit\\_in\\_strategy.pdf](http://www.business.uiuc.edu/ghoetker/documents/Hoetker_logit_in_strategy.pdf)
- *Assignment:* (1) Select an article using logistic, ordinal, or multinomial regression from either *American Sociological Review*, *American Journal of Sociology*, or *Social Forces* since the year 2000. Briefly summarize how the research problem is conceptualized and how the data are described and analyzed. Then copy (by pen or pencil) the format of the main tables that describe the data and report the regression results. (2) Do Stata assignment #10 (“Categorical dependent variables, part 1”). (3) Moore & McCabe problems 15.1, 15.2, 15.21, 15.22, 15.23, 15.24; Mendenhall & Sincich problems 9.13, 9.16, 9.17.
- *Mini-project #3 due in class on Mar. 16.*
- *Mini-project #4 assigned on Mar. 16: due in class on March 30.*

### **March 30 & April 6: Regression with categorical dependent variables—ordinal & nominal dependent variables**

- Agresti & Finlay, chapter 15, “Logistic Regression Models: Modeling Categorical Responses, section 15.6, “Logit Models for Ordinal Variables”
- Long & Freese, chapter 5, “Models for Ordinal Outcomes”
- Long & Freese, chapter 6, “Models for Nominal Outcomes”
- *Recommended:* King et al., *Designing Social Inquiry*, chapter 6, “Increasing the Number of Observations”

- *Assignment:* (1) Do Stata assignment #11 (“Categorical dependent variables, part 2”) and #12 (“Check & Clean Data”).
  - *Mini-project #4 due in class on Mar. 30.*
  - *Mini-project #5 assigned on Apr. 6: due at scheduled time of final exam.*
- 

**Organizing Your Computer Work for Serious Data Analysis**  
(from Long & Freese, *Regression Models for Categorical Dependent Variables Using Stata*)

1. Ensure replicability by using do-files and log files for everything.
  2. Document your do-files.
  3. Keep a research log.
  4. Develop a system for naming files.
  5. Use new names for new variables and files.
  6. Use labels and notes.
  7. Double-check every new variable.
  8. Practice good archiving.
- See also Scott Long & Jeremy Freese, “Replication Standards for Quantitative Social Science: Why Not Sociology?”  
<http://www.jeremyfreese.com/docs/freese-reproducibility-webdraft.pdf>

**Social Research Study Questions**

“Blindly plugging into statistical formulas has caused a lot of confusion.”  
-- Freeman, Pisani & Purves, *Statistics*, 3e

1. What is *social research*? What are the principal differences between social research and other ways of representing social life?
2. What is the *scientific method*? What steps does the scientific method—or scientific methods—apply in conducting social research?
3. What is a *research strategy*? What are the differences between research strategies that *particularize* and those that *generalize*? What are the potential similarities between such strategies?
4. What is the *social construction of reality*? How does it pertain to the scientific method, social research/research strategies in general, and to other ways of representing social life—including the promises and risks of the various approaches?
5. What are *data*? What are interplays between data and the social construction of reality? Is everything worthwhile *measurable*?

6. What is *statistics*? What is the difference between *descriptive statistics* and *inferential statistics*? How do descriptive statistics and inferential statistics pertain to the principal kinds of *research strategy*?
7. What are advantages and disadvantages of using statistics in social research?
8. What are the intersections between the uses of statistics in social research and the social construction of reality? Conversely, what are the intersections between the “non-uses” of statistics in social research and the social construction of reality?

### **Statistical Methods: Some Pro's & Con's**

#### ***Some advantages of using statistics***

- Summarizes complex data
- Makes assumptions explicit
- Imposes explicit standards of evidence and comparison
- Raises the possibility of chance associations
- Emphasizes skepticism about hypotheses and findings
- Facilitates the testing of competing hypotheses and the building of theories
- Permits the examination of certain questions that couldn't otherwise be examined

#### ***Some pitfalls of using statistics***

- The use of statistics represents a strategic tool in the social construction of reality. Thus its use in general must be situated in the historical/geographic context of bureaucratization, state formation and geopolitical competition, industrial/ technological revolution, commodification, and urbanization; and the biases of statistical premises and the tendency of the statistically inclined research establishment to claim intellectual/policy hegemony on the basis of a “scientific approach” must be critically examined.
- The use of statistics impedes the examination of certain questions that otherwise would be examined, and obfuscates crucial kinds of social, cultural, and political analysis.
- Theory and substantive importance must guide the use of statistics (although the data must inform the theory as well [e.g., making sense of unanticipated nonlinearities or outliers]).
- Statistical research needs to emphasize theoretical/substantive significance and the magnitude of relationships between variables, rather than mere “statistical significance” as narrowly defined by mainstream statistical methodology. The research needs to recognize the arbitrariness of institutionalized significance-test standards and to consider alternative criteria for statistical significance.
- Statistical research needs to test a study's findings, not just against its null hypothesis but also against competing theories/hypotheses with the objective of long-term theory building.
- We need to use statistics wisely as one of many tools in social research.

### **Reminders**

- How were the numbers produced—in the sense of culture and power, and according to the (cultural) rules of scientific method?
- Is the sample random and representative of the population? Insofar as this is not true, then the use of inferential statistics is invalid.
- What are the overall pattern and striking deviations (i.e. -the shape, center, and spread) of the distribution? Are there outliers?
- Do the numbers make sense? (adapted from Moore, *The Basic Practice of Statistics*):
  - What's the explicit or implicit agenda behind them?
  - Is any essential information left out?
  - Are the numbers consistent?
  - Are the numbers plausible, including are they too good to be true?
  - Is the math correct?
  - What do the numbers signify about the social relations being studied?
- Always take “outlying” observations, “non-significant” findings, and otherwise “contrary” findings seriously: What insights do they potentially convey about the social relations that you’re studying, and possibly about social relations more generally?
- You’ve estimated something’s magnitude or likelihood. Don’t lose sight of *uncertainty*: What’s the estimated *range* of magnitude or likelihood? What does this range imply about the social relations being examined?
- Are all worthwhile things measurable? What do your conclusions imply about the social relations and public policies that you’re studying, and perhaps about social relations and public policies in other spheres?

### ***Benchmarks for assessing the usefulness of social statistics***

- Does the use of statistical methods in any given instance notably improve our understanding of social relations and public policies?
- In any given instance, what insights does the use of statistical methods provide (or not) in comparison with insights provided by other methods of social research, and in comparison with insights provided by other ways of interpreting the world?

### ***A strategic point in interpreting & summarizing your study's results***

- What are the wider, comparative ramifications of your study for understanding social relations and social policy/political practice?