Grown-ups love figures. When you tell them that that you’ve made a new friend, they never ask you any questions about essential matters. They never say to you, “What does his voice sound like? What games does he love best? Does he collect butterflies?” Instead, they demand: “How old is he? How many brothers has he? How much does he weigh? How much money does his father make?” Only from these figures do they think they have learned anything about him.

- *The Little Prince* (quoted in Freedman, Pisani & Purves, *Statistics*, 3e)
What it’s about

This course provides a graduate-level introduction to applied statistics within the framework of social research and analysis. The course complements the fall-semester graduate seminars in social theory and social research methods. Its objective is to present basic conceptual and practical tools in social statistics so that—whether or not you intend to pursue a career of doing quantitative studies—you’ll be better equipped, first, to critically assess social (and policy) research carried out from a wide array of methodological perspectives; and second, to make sound methodological decisions and wise interpretations in carrying out your own research projects. Throughout the course we’ll be emphasizing the question, “In what ways does using statistics promote or impede insight into meaningful aspects of social relations and public policies?”

Books, software & supplements

The required books are: Moore & McCabe, Introduction to the Practice of Social Statistics, 5th ed; Babbie, The Basics of Social Research; and Ragin, Constructing Social Research. Strongly recommended as supplementary reading are Utts, Seeing Through Statistics; Kohler & Kreuter, Data Analysis Using Stata; and Hamilton, Statistics with Stata Version 9.

The earlier editions of Moore/McCabe (along with those of Freedman et al., Statistics) are widely praised as having pioneered a new way of teaching and learning statistics: one that minimizes formulas and derivations while stressing conceptual and real-world issues. Among these are the notion that statistics is simply a means for reaching one kind of understanding of the world; the design of samples and experiments; the analysis of real data; the assessment of social, scientific, and policy debates; and the cultivation of sound judgement and humanitarian criteria in assessing and using statistics. This will be our approach. Babbie’s The Basics of Social Research provides a bridge between statistics and social research. Ragin’s Constructing Social Research complements our Department’s cross-disciplinary approach by comparing three basic research methodologies—“multivariate” statistical, case study, and “small-N” comparative—and proposes points of synthesis. Constructing Social Research, then, also provides a bridge from this course to the graduate seminars in social theory and social research methods. Utts’s Seeing Through Statistics is an engagingly written, non-technical guide to judging and using statistics. It keeps big, real-world matters front and center.

Several other resources will nurture us through the semester. Associated with Moore/McCabe is a web site (http://bcs.whfreeman.com/ips/5e/) and a CD-ROM that contain helpful materials, not least of which are conceptually-oriented quizzes on each chapter and “applets” that permit hands-on, interactive learning.

We are using the statistical software Stata, which is fast gaining ground in the social, health, and policy sciences. In contrast to the still more commonly used SPSS and SAS—which were born in an era of bulky mainframe computers and clunky top-down organizations—Stata is a product of the PC-age, decentralized work, and flexible specialization. It is much more intuitive as well as much faster and more agile than SPSS or SAS. Most important for us, Stata is designed for “exploratory data analysis” (EDA)—the interactive, “modern” approach (innovated by Bell Labs & Princeton statistician John Tukey) that is epitomized in Moore/McCabe. Stata makes
learning statistics easier and in the long run makes advanced statistical techniques (including survey methods) far more accessible than do SPSS or SAS. And if you do need to learn SPSS, by the end of this semester you’ll be capable of using an SPSS introductory book to teach yourself the point-and-click version in a day or so.

UCLA’s Academic Technology Services has created a superb website that makes available an impressive set of free, downloadable materials for learning statistics in tandem with Stata or SPSS, SAS, or specialized statistical software. The subsite http://www.ats.ucla.edu/stat/seminars/ contains introductions to these and other statistical software programs, including movies. The Stata subsite is located at: http://www.ats.ucla.edu/stat/stata/.

Software, though, is just a tool. The focus of this course is learning statistics as one way to describe and understand significant aspects of social relations.

**Classroom policy, projects, exams & grades**

- It is assumed that students will attend all class sessions & arrive on time.
- Cell phones must be turned off during class sessions, and using your lab computer for emailing during that time is prohibited.
- Questions, comments & discussion are enthusiastically encouraged.
- Graded assignments:
  - Students are responsible for all materials covered in the assigned readings & problems, as well as all materials covered in class sessions.
  - All graded assignments must be completed in order to earn a passing grade in the course. Late assignments—that is, anything turned in after the stipulated date and time—will be graded and returned after the final exam. The grades for late assignments will be severely reduced.
  - Moore/McCabe homework problems on more or less a weekly basis. These include practice not only in solving statistical problems but in using and interpreting statistics wisely. Hands-on practice, done virtually every day, is the only way to learn statistics (and software) as well and as fast as we need to do in this course. The Moore/McCabe problems are meant to foster active learning. In that spirit, most of them have answers, but not step-by-step solutions, in the back of the book. Some of them, moreover, are pegged to the web site’s and CD-ROM’s interactive applets, which in general should be used frequently for active learning of concepts. Grading: pass/fail and worth 20% of the final grade. Passing all of them earns an “A”, all but one of them a “B”, and so on, worth a total of 10% of your course grade. *Homework problems assigned at a given class session are due at the start of the next class session.*
  - One project based on a data set to be chosen in consultation with the instructor. Students will use Stata to conceptualize and apply the statistical methods we will have covered, and will interpret the results
as well as explore the pro’s and con’s of statistical social research. 

Grading: each the project is worth 40% of the course grade. The project is due at the start of the class session on November 10.

- Two take-home exams, which will combine statistical problems with essays focusing on the development of sober judgement in selecting, applying, interpreting, and critiquing statistics. Grading: each take-home exam is worth 25% of your course grade (for a total of 50%). Each take-home exam is due at the start of the class session on the due date.

Preparing for class sessions

Each class session will cover the minimum technical information that’s necessary to learn statistics and Statat, and the maximum possible to put social statistics within the frameworks of social theory and social research methodologies.

Regarding the basics of statistics, we’ll stick closely to Moore/McCabe’s textbook presentation, emphasizing the broadest conceptual issues. At the start of each session we’ll review some of the homework problems and/or the take-home exam you will have completed. As much as possible we’ll use the problems and exams to raise the big issues about doing social research.

Here’s how to prepare for each session:

- Complete the assigned Moore/McCabe problems and/or the assigned take-home exam or mini-project.

- Review the “Social Research Study Questions” and the “Statistical Methods: Some Pro’s & Con’s” (both of which are attached to the syllabus) before the first class session and throughout the semester. We’ll refer to them frequently.

- The assignments from Babbie and Ragin must be read before the corresponding class sessions. These provide social-science methodological context to the statistical work.

- The assigned readings from Moore/McCabe should be skimmed before each session, then should be thoroughly covered when doing the homework problems (including use of the instructor-supplied, Stata-formatted text data sets).

- Doing the web or CD-ROM quiz corresponding to each Moore/McCabe chapter assignment will be helpful to make sure you’ve mastered the course material.

- Kohler & Kreuter, Data Analysis Using Stata and Hamilton, Statistics with Stata Version 9 are excellent for learning Stata, as well as for learning statistics. The ATS-UCLA web site for “Stata Resources” is another excellent resource for learning Stata. See also “Stata survival commands” and “Stata practice” on the course web site.

- The Moore/McCabe problems assigned in any given class session are due at the start of the next class session.
• Everything else: Do whatever works best for you.

---

**Tentative Schedule**

“Statistics is the art of making numerical conjectures about puzzling questions.” – Freedman, Pisani & Purvis, *Statistics*, 3e

“... variation itself is nature’s only irreducible essence.” – Stephen Jay Gould, “The Median Isn’t the Message”

**Homework assignments:** format the data for Stata, if necessary; turn in only the Stata commands (by means of a “command log”) & your interpretation of the statistical output, not the output itself.

**Aug. 30**

**Univariate distributions: graphical & numerical summaries**


Babbie, chap. 1 (pages 6-25); Ragin, chaps. 1 & 2 (see “Social Research Study Questions” & “Statistical Methods: Some Pro's & Cons” attached to syllabus); Moore/McCabe, “To Students: What Is Statistics?” (pages xxi-xxiv); chap. 1 (pages 4-21, 23-24)

Recommended: Utts, chaps. 1 & 2

• Go to the website for “ASR Manuscript Submission Information for Authors” and navigate to the “Preparation Checklist for ASR Manuscripts.” Then type the guidelines for text/content, an example of each kind of text citation, and an example of each kind of bibliographic reference.

• See “Stata survival commands” and “Stata practice” on the course’s web site. See also Kohler/Kreuter or Hamilton.

• What is “social research”? What are its main goals and strategies?

• Moore/McCabe problems: 1.2, 1.4, 1.6, 1.7, 1.17, 1.23, 1.25, 1.26, 1.27, 1.29, 1.35, 1.39
Sept. 6  

*Univariate distributions: graphical & numerical summaries*  
(continued)

Ragin, chap. 3; Moore/McCabe, chap. 1 (pages 40-56, 64-84)

Recommended: Utts, chaps. 7 (pages 106-23), 8 & 9; Babbie, chap. 2 (pages 42-56)

- Utts, chap. 9: type Section 9.1’s guidelines for “Well-Designed Statistical Pictures” and Section 9.4’s “most common problems in plots, graphs, and pictures.”

- Learn how to document your statistical work by means of a “log” file and “command log file” (see “Stata survival commands” on the course web site).

- Beginning this week, always turn in statistical work by making the command log into a “do-file” (see Kohler/Kreuter, pages 37-39) and turning in the do-file to document the commands. A do-file should allow you (or anyone else) to replicate statistical output.

- Moore/McCabe problems: 1.43, 1.44, 1.45, 1.47, 1.49, 1.51, 1.52, 1.55, 1.56, 1.60, 1.75, 1.76, 1.77, 1.78, 1.83, 1.85, 1.86, 1.88, 1.89, 1.93, 1.96, 1.98, 1.112, 1.119, 1.120, 1.137, 1.147

- Briefly answer these questions: Why do univariate distributions matter—in terms of substantive issues in social science and in terms of statistics? Why should we graph a univariate distribution before we numerically summarize it? What are the advantages of the 5-number distribution?

Sept. 13  

*Bivariate distributions: scatterplots, correlation & least-squares regression*

Ragin, chap. 6; Moore/McCabe, chap. 2

Recommended: Utts, chaps. 10, 11 & 12

- See the course document “Profile.do” and Kohler/Kreuter, pages 25-27, on making a “profile do-file.” Make a profile do-file (including the precaution of automatically logging your statistical work) and turn it in. As begun last week, continue turning in a do-file for each homework assignment.

- How is large-scale quantitative research similar to and different from other kinds of social research?

- Moore/McCabe problems: 2.1, 2.3, 2.4, 2.5, 2.8, 2.9, 2.13, 2.14, 2.29, 2.30, 2.32, 2.33, 2.38, 2.39, 2.40
Sept. 20  \textit{Bivariate distributions: scatterplots, correlation & least-squares regression} (continued)

Moore/McCabe, chap. 2

Recommended: Utts, chaps. 10, 11 & 12

- Moore/McCabe problems: 2.42, 2.53, 2.54, 2.57, 2.66, 2.67, 2.68, 2.79, 2.80, 2.81, 2.85, 2.86, 2.88, 2.89, 2.93, 2.95, 2.113

- Briefly answer these questions: What can be misleading about measures of correlation and regression, and why? When might there be a strong bivariate relationship but a low correlation or regression coefficient? What are the advantages of regression versus correlation? What are the basic ways of establishing and explaining causation? What critical questions must we ask concerning the possible relationship between two variables?

Sept. 27  \textit{Producing data}

Babbie, chaps. 4, 5, & 7; Ragin, chap. 4; Moore/McCabe, chap. 3

Recommended: Utts, chaps. 3, 4, & 5

* Exam #1 (take-home) assigned: due at start of class on October 11.


- What is “statistical” significance? Give a social-science example of how it can be different from “practical” or “theoretical” significance? What is bias? What is variability? Why is randomization important? Why is comparative design important? What are the advantages of experimental research and the reasons for these advantages? What are the disadvantages of experimental research? What are the main sampling designs? What are the advantages and disadvantages of each sampling design? What are possible “sampling” and “non-sampling” sources of error in surveys? Regarding qualitative research, what is “theoretical sampling”?

Oct. 4  \textit{Probability & inference}

Moore/McCabe, chap. 4

Recommended: Utts, chaps. 15, 16 & 17

* Project assigned: due at the FIU-scheduled date/time of the final exam.
**Moore/McCabe problems:** 4.5, 4.7, 4.16, 4.17, 4.22, 4.24, 4.33, 4.43, 4.44, 4.45, 4.59, 4.61, 4.63, 4.72, 4.74

- Why is randomization important? What is a random variable? What is a probability distribution? Why is the reason for using the term “expected value of random variable” instead of “mean of a random variable”? What is the Law of Large Numbers? How many trials are needed to guarantee a mean outcome close to the population mean? What are “independent observations” (or events), and when can this requirement be relaxed? What is “conditional probability,” and what would be a social-science example?

**Oct. 11**

**Sampling distributions; exam #1 due**

Moore/McCabe, chap. 5

Recommended: Utts, chap. 18

- Moore/McCabe problems: 5.1, 5.3, 5.17, 5.19, 5.51, 5.55, 5.63

- What is a sampling distribution and a population distribution, and what would be a social-science example of both? What is a count and a sample proportion, and a social-science example of each? What is the “binomial setting,” and what is a social-science example? What is a binomial distribution and a sampling distribution of a count? When should we use a binomial sampling distribution? Why are sample means used in statistical inference? What is the sampling distribution of a sample mean, and what is a social-science example? How do we compute the mean and standard deviation of a sample mean? What is the Central Limit Theorem, and why is it important? Why is it that the sample mean of an SRS from a normal population has a normal distribution, and what principle does this illustrate?

**Oct. 18**

**Introduction to inference**

Babbie, chap. 7 (pages 197-206); Ragin, chap. 5; Moore/McCabe, chap. 6

Recommended: Utts, chaps. 20, 21, 22 & 23

- Moore/McCabe problems: 6.1, 6.2, 6.3, 6.9, 6.13, 6.14, 6.20, 6.36, 6.37, 6.40, 6.41, 6.45, 6.46, 6.57, 6.58, 6.74, 6.76, 6.77, 6.82, 6.86, 6.87

- What is “statistical inference”? For what kinds of data is it valid or invalid? What is the difference between statistical significance and “practical” or “theoretical” significance? What is a confidence interval, and how is it computed? What are the data requirements for a valid confidence interval, and how do we check these requirements? How can we reduce a confidence interval? How do we test a hypothesis? What is wrong with accepting a null hypothesis? What is
wrong with accepting an alternative hypothesis? What is a $P$-value? What is the Bonferroni procedure, and why and when should it be used? What is wrong with “searching for statistical significance”? When should statistically insignificant results be reported and explained? What are Type I and Type II errors, and what is an example of each? How does the logic of quantitative-inferential analysis differ from that of comparative case-study analysis?

Oct. 25  
**Inference for distributions**

Moore/McCabe, chap. 7

Recommended: Utts, chaps. 20, 21, 22 & 23

- Moore/McCabe problems: 7.2, 7.3, 7.5, 7.24, 7.40, 7.41, 7.42, 7.53, 7.57, 7.132, 7.133

- What is a $t$ distribution, when is it used, and how is it computed? How do $t$ distributions differ from $z$ distributions, and at one point do they become more or less identical? When do we use one-sample and two-sample $t$ tests, and what is a social-science example for each? What are the data requirements for such tests? What in general are test alternatives if the data requirements are not met?

Nov. 1  
**Inference for proportions**

Moore/McCabe, chap. 8

Recommended: Utts, chap. 19

- Moore/McCabe problems: 8.5, 8.7, 8.15, 8.45, 8.47, 8.53, 8.77

- What is a social-science example for single proportion and for two proportion inference? What are the data requirements?

Nov. 8  
**Inference for two-way tables**

Moore/McCabe, chap. 9 (pages 610-26, 628)

Recommended: Utts, chap. 12

- Do assignment on checking/cleaning data (check_data.do).


- What are the data requirements for two-way tables? What are social-science examples of two-way tables? What is Simpson’s Paradox, what is the reason for it, and how can we guard against it? What is the chi-square statistic? How do we test a hypothesis for a two-way table?
Nov. 15  **Inference for regression**

Moore/McCabe, chap. 10 (pages 656-75, 676-77)

Recommended: Utts, chap. 10

- Moore/McCabe problems: 10.1, 10.2, 10.21, 10.22, 10.28, 10.44, 10.45, 10.46
- What is simple linear regression? How does it differ from correlation? When might there be a strong bivariate relationship but a weak correlation or regression coefficient? What are the data requirements for regression? What do the following mean: DATA=FIT + RESIDUAL? What is the fundamental difference between how regression is treated in Moore/McCabe chaps. 2 and 10? How do we test a hypothesis for a regression coefficient? What does it mean if the plotted residuals of a regression model are not randomly distributed?

Nov. 22  **Inference for regression (continued)**

Ragin & Driscoll, "Afterword"; Moore/McCabe, chap. 11, chap. 1 (pages 51-55), chap. 2 (pages 187-203)

Recommended: Utts, chap. 10

* Exam #2 assigned: due at the start of class on Dec. 6.

- Moore/McCabe problems: 11.4, 11.5, 11.8, 11.24, 11.43-11.51
- What is the advantage of multiple regression over simple regression? Explain why or why not a strong/weak bivariate relationship may not result in a strong/weak multivariate relationship, including what this has to do with our previous reading on causal relations?

Nov. 29  **Analysis of Variance**

Moore/McCabe, chap. 12 (pages 746-765), chap. 13

Recommended: Utts, chap. 26

- Moore/McCabe problems: 12.2, 12.6, 12.10, 12.48, 12.49, 13.19, 13.20
- Why you should use ANOVA? When you should use one-way ANOVA? When you should use two-way ANOVA?
Dec. 6 Review; exam #2 due

Review Moore/McCabe, chaps. 1-11; Ragin, chaps. 1-6, Afterword; work on mini-project #2; Babbie (as relevant)

Dec. 13 Review; project due at the FIU-scheduled date/time of the final exam

Social Research Study Questions

“Blindly plugging into statistical formulas has caused a lot of confusion.”
- Freedman, 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 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 is 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 thing’s 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 any application of social statistics_

• Does the use of statistical methods in any given instance notably improve our intellectual 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 the understanding of social relations and social policy/political practice?