

University of North Carolina  
Chapel Hill

# Soci708-001 Statistics for Sociologists

Fall 2009

Professor François Nielsen

Last modified 24 Aug 2009

## 1 Times, Places & Contacts

### Classes Meet

- Tue, Thu 11:00–12:15 PM in Hamilton 151

### Instructor

- Professor François Nielsen – Email: francois\_nielsen@unc.edu Office: Hamilton 163 Hours: by appointment Phone: 962-5064 Fax: 962-7568
- Web site <http://www.unc.edu/~nielsen>

### Teaching Assistant

- Andrew Ritchey – Email: ritcheya@email.unc.edu Office: Hamilton 254 (also check computer lab in Hamilton 228) Hours: Tue, Thu 12:15–1:15 PM or by appointment

## 2 Course Description & Goals

The course covers descriptive statistics and graphs for exploratory analysis of data distributions, including descriptive uses of normal distributions; exploratory approaches to association for both categorical and continuous variables; elements of probability theory including Bayes' Theorem and the algebra of expectations; probability distributions for continuous variables with special emphasis on the normal distribution; sampling distributions; principles of statistical inference, including confidence intervals and hypothesis testing, in large and small samples; inference for population means and proportions; the use of the chi-square statistic for dependence and homogeneity tests involving categorical variables; simple and multiple linear regression including statistical inference, some discussion of data transformations, and the analysis of residuals; elements of matrix algebra and the matrix representation of the general linear model. (Analysis of variance is not treated as a topic separate from multiple regression.) The course includes learning how to use the microcomputer for data entry and analysis using the statistical packages STATA.

The purposes of the course are

1. To enable the student to intelligently and critically read research articles based on quantitative research, as found in mainstream sociology journals such as *American Sociological Review*, *American Journal of Sociology*, or *Social Forces*, and equivalent journals in other fields.
2. To elucidate the steps in the process that leads from a research idea to actual data analysis. One of the major obstacle that once confronted some graduate students in the completion of master's thesis or dissertation has been the problem of "getting the data into the computer". The advent of microcomputers, spreadsheet software and interactive statistical packages has greatly simplified the task of entering data, creating permanent data files and analyzing them. The course will teach how to do this with minimal aggravation.
3. To lay the theoretical foundations for the next courses in the methods sequence of the graduate program in sociology (Sociology 709 and 711). For this reason, the course places greater emphasis on some theoretical topics, such as the algebra of expectations and matrix algebra, than is found in more elementary coverages of these materials.
4. To convey the flavor of the contemporary approach to data analysis and the use of statistical methods, which is an emerging synthesis based on the combination of the traditional rigorous approach of classical statistics with the influence of exploratory data analysis (as developed, for example, by John Tukey) and the liberating effect of inexpensive computing power associated with the advent of microcomputers. The contemporary approach is characterized by a greater attention paid to the ways data deviate from the assumptions of classical methods, the influence of outliers and influential cases, the robustness properties of various methods, and by the use of visual tools for exploration and diagnosis.

### 3 Readings

#### 3.1 Books

Books for the course are available at Student Stores. The main text is:

Moore, David S., George P. McCabe and Bruce A. Craig. 2009. *Introduction to the Practice of Statistics*. (6e.) New York: W. H. Freeman & Co.

You can also find an earlier edition of this text. This will most likely work well for learning about the materials, but it may be inconvenient in identifying problems for the homework assignments, as the selection and numbering of problems have changed between editions. You can minimize the inconvenience by collaborating with other students in doing the homework assignments.

In addition, the following book is *recommended* (but not required). You can get the hard copy edition at the Student Stores or an electronic version at a location TBA.

Hamilton, Lawrence C. 2009. *Statistics with STATA: Updated for Version 10.0*. Duxbury.

### 3.2 Additional Readings

Additional readings are available as PDF files from the course site. A password may be needed to access some of these documents.

### 3.3 Class Notes and Presentations

PDF files with PowerPoint-like class notes will be linked to this syllabus for each week. See course schedule. You can download or print these notes as you wish. I must reserve the right to revise/update the notes and presentations at any time during the semester.

### 3.4 Computing Resources

Access to statistical programs is available in two main computer labs.

**Odum Lab in Hamilton 228** This is the sociology graduate students computer lab. Andrew Ritchey will show you how to use the machines there if you are not a sociology graduate student.

**Odum Institute (IRSS) Computer Lab** This is the statistical computer lab at the Odum Institute in Manning Hall (basement). This is a completely up-to-date statistical laboratory providing access to many programs and expert statistical help.

## 4 Exams & Requirements

### 4.1 Honor Code

THE HONOR CODE WILL BE IN FORCE FOR ALL EXAMINATIONS AND ASSIGNMENTS

For more information on student duties under the Honor Code please see

<http://honor.unc.edu/students/rights.html>

Please note in particular the following section:

[It is the responsibility of students] To maintain the confidentiality of examinations by divulging no information concerning an examination, directly or indirectly, to another student yet to write that same examination.

However in the preparation of homework collaboration among students is allowed and encouraged.

### 4.2 Requirements

The midterm and the final are multiple-choice exams. You need to bring

- a “bubble sheet” for electronic scanning, available free of charge at Student Stores

- a pencil (#2 preferred) with an *effective* eraser (soft, pink,..., *not* hard and all dried up!) so you can erase your answer completely if you change your mind
- A calculator that can calculate a square root
- (Optional) A “cheat sheet” containing any information you wish; limit is 1 sheet (2 pages) for the midterm and 2 sheets (4 pages) for the final

For examples of multiple-choice questions of the type used in the exams click on

- Description of Midterm (to be updated)
- Description of Final (to be updated)

Your final grade in the class will be based on two exams and seven home work assignments.

#### **(1) Midterm (25% of course grade)**

The midterm will take place during regular class time on

- Tue 20 Oct 11:00–12:15 PM

The midterm will consist of multiple choice questions.

#### **(2) Final (35% of course grade)**

The final will take place on

- Sat 12 Dec 12:00 Noon–3:00 PM

The final will consist of multiple-choice questions. The final is cumulative but materials covered after the Midterm are emphasized.

#### **(3) Homework Assignments (40% of course grade)**

There will be seven homework assignments, due on dates listed in the outline. Your grade for the homework will be based on the best 5 out of the 7 assignments. An answer to a problem will receive partial credit if it shows a serious attempt at solving the problem, even if the answer is incorrect.

You are allowed to freely collaborate with others in preparing the assignment. However you should turn in your own copy of the assignment.

To turn in your assignment please hand in a hard copy of your work to Andrew Ritchey at class time. Please do not send email attachments.

### **4.3 Posting of Grades**

Exam scores and final grades will be made available in a manner TBA.

## 5 Outline & Readings Schedule

Key to Readings:

IPS6e refers to Moore & al. refers to Moore, David S., George P. McCabe and Bruce A. Craig. 2009. *Introduction to the Practice of Statistics*. (6e.)

SWS refers to Hamilton, Lawrence C. 2009. *Statistics with STATA*

Important Dates – PRESENCE REQUIRED:

1. Tue 20 Oct 11:00–12:15 PM – Midterm
2. Sat 12 Dec 12:00 Noon–3:00 PM – Final

Date	Class	Topic	Notes	Readings (IPS6e)	Due
Tue Aug 25	1	First contact	Mo		
Thu Aug 27	2	Data: Distributions	M1	Ch1, 1.1–1.3	
		✿			
Tue Sep 01	3	cont'd			
Thu Sep 03	4	Data: Relationships	M2	Ch2, 2.1–2.6	
		✿			
Tue Sep 08	5	cont'd			HW 1
Thu Sep 10	6	cont'd			
		✿			
Tue Sep 15	7	cont'd			
Thu Sep 17	8	Producing Data	M3	Ch3, 3.1–3.3	
		✿			
Tue Sep 22	9	cont'd			HW 2
Thu Sep 24	10	Probability	M4	Ch4, 4.1–4.5	
		✿			
Tue Sep 29	11	cont'd			
Thu Oct 01	12	cont'd			
		✿			
Tue Oct 06	13	cont'd			HW 3
Thu Oct 08	14	cont'd			
		✿			
Tue Oct 13	15	Sampling Distributions	M5	Ch5, 5.1–5.2	
Thu Oct 15	16	cont'd			HW 4
		✿			
Tue Oct 20	17	MIDTERM			
Thu Oct 22		Fall Break – NO CLASS			
		✿			
Tue Oct 27	18	Introduction to Inference	M6	Ch6, 6.1–6.4	
Thu Oct 29	19	cont'd			
		✿			
Tue Nov 03	20	Inference for Distributions	M7	Ch7, 7.1–7.3	
Thu Nov 05	21	cont'd			
		✿			
Tue Nov 10	22	cont'd			HW 5
Thu Nov 12	23	Inference for Proportions	M8	Ch8, 8.1–8.2	
		✿			
Tue Nov 17	24	cont'd			
Thu Nov 19	25	Inference for Categorical Data	M9	Ch9, 9.3, 9.1, 9.2	
		✿			
Tue Nov 24	26	Inference for Regression	M10	Ch10, 10.1–10.2	HW 6
Thu Nov 26		Thanksgiving – NO CLASS			
		✿			
Tue Dec 01	27	cont'd			
Thu Dec 03	28	Multiple Regression	M11	Ch11, 11.1–11.2	HW 7
		✿			
Tue Dec 08	29	cont'd			
Sat Dec 12		FINAL – 12:00–3:00 PM			