

YESHIVA UNIVERSITY
WURZWEILER SCHOOL OF SOCIAL WORK

Fundamentals of Applied Statistics

SWK 8406

Fall 2018

Wednesdays: 6:00 – 7:50pm

Belfer Hall 821

Professor: Matthew J. Cuellar, Ph.D.

Office: Belfer Hall 909B

Phone: (212) 960 – 0827

Office Hours: By Appointment

e-mail: matthew.cuellar@yu.edu

Instructor's Web Page: <http://faculty.yu.edu/faculty/pages/cuellar-Matthew>

Ariella Tomback, Assistant Dean: (212) 960 – 5418; ariella.tomback@yu.edu

Course Description

The course provides the foundations necessary to understand elementary statistics and applications of statistics to many areas in natural and social sciences. It will concentrate on the interpretation and comprehension of graphical and statistical techniques that are essential components to all areas of research. The ability to understand basic arithmetic and high school algebra is required.

Objectives

At the conclusion of this course the student will be able to:

- 1) Read the social science research literature and understand some fundamental methodologies of research as regards: research design, descriptive statistics [univariate and bivariate] and measurement, statistical inferences drawn when comparing samples to populations, and when comparing samples to other samples.
- 2) Understand the importance and use of probability theory and distributions as a basis for statistical decision-making.
- 3) Review and interpret findings by other researchers in various areas of social science research

Course Requirements

1. **Quizzes (25% of final grade):** Quizzes will cover content from each week of class. The purpose of the quizzes is to account for student engagement in material. Quizzes focus on the conceptual understanding of the material. Quizzes can be completed in Canvas. There are 12 quizzes. Quizzes will be made available each week and are due Monday evenings at 11:59pm ET (unless otherwise specified for holidays; due dates are clearly outlined in Canvas). *Late quizzes will not be accepted unless there is an unavoidable emergency and I am notified in advance.*

2. **Homework (25% of final grade):** It is essential that students practice the skills covered in class sessions. Homework assignments will focus on the application of the material (e.g., running statistical tests) and understanding and writing up statistics in APA format. There are 12 homework assignments. Homework must be submitted via Canvas. Grades and feedback on homework assignments will be provided through Canvas and discussed at the beginning of each class, so you must keep up with the work! Homework is due Monday evenings at 11:59pm ET (unless otherwise specified for holidays; due dates are clearly outlined in Canvas). ***Late homework will not be accepted unless there is an unavoidable emergency and I am notified in advance.*** While you may work together on homework assignments, you **MUST** write up and submit the homework ***individually***.
3. **Final Exam (50% of your final grade):** There will be one final exam that will take place on the last scheduled meeting date. The exam will be held on Canvas during the designated meeting time and will consist of questions related to the theory used, assumptions made, and interpretation of the statistical data presented. Material from both the lectures and from the text may appear on the exam and it is cumulative. The final exam will be administered in Canvas on the last day of class.

Grading:	A \geq 94	A- = 90-93	B+ = 87-89
	B = 84-86	B- = 80-83	F < 80

Office Hours (By Appointment): I encourage you to consult with me about any questions or problems you have with the course material. Please Email me to ask questions or make an appointment to discuss the materials: matthew.cuellar@yu.edu. Office Hours are filled as needed.

Evaluation of the Course: At the end of the semester, a formal evaluative questionnaire regarding both the curriculum and my instruction will be administered online. The information gathered from this process will be used to improve future courses and instruction. Additionally, I strongly encourage you to provide feedback to me during the semester either in person or anonymously—I am here to teach you and want to do it well!

Required Texts:

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.

Kabacoff, R. (2015). *R in Action: Data analysis and graphics with R* (2nd ed.). ISBN: 9781617291388. Greenwich: Manning. (See here for a discount: <http://www.statmethods.net/about/author.html>)

Reid, H. M. (2014). *Introduction to statistics: Fundamental concepts and procedures of data analysis*. Los Angeles: Sage.

Class Schedule

Session 1 9/12/2018	Course overview Review of basic research concepts Basic R concepts Readings: Reid (2014), Chapters 1 & 5
Session 2 9/26/2018	Making friends with your data: Exploring, prescreening, and describing data Readings: Reid (2014), Chapters 2, 3, 4
Session 3 10/3/2018	Principles of inferential statistics Readings: Reid (2014), Chapter 9
Session 4 10/10/2018	Principles of inferential statistics--continued Readings: Reid (2014), Chapter 6
Session 5 10/17/2018	Comparing two group means Readings: Reid (2014), Chapter 10 Wright et al. (2011)
Session 6 10/24/2018	Comparing three or more group means Readings: Reid (2014), Chapters 11 & 12 & Appendix A
Session 7 10/31/2018	Factorial analysis of variance & interactions Analysis of covariance Readings: Reid (2014), Chapter 13
Session 8 11/7/2018	***NO CLASS MEETING: ONLINE MODULE WILL BE POSTED*** Categorical variables Readings: Reid (2014), Chapters 7 & 8
Session 9 11/14/2018	Effect size Readings: Reid (2014), Chapter 14, Appendix B Ellis (2010), Introduction, Chapters 1 & 2
Session 10 11/19/2018	***NO CLASS MEETING: ONLINE MODULE WILL BE POSTED*** Type I errors, statistical power & sample size selection Readings: Reid (2014), Appendix D Ellis (2010), Chapters 3 & 4
Session 11 11/28/2018	Bivariate linear regression Readings: Reid (2014), Chapter 15
Session 12 12/5/2018	Missing data Cox et al. (2014)

Session 13 Final Exam Review
12/12/2018

Session 14 FINAL EXAM
12/19/2018

Other Resources

Downloading R/RStudio: To download RStudio, you will first need to download base R. You can download R onto your computer from the following website:

<http://cran.cnr.berkeley.edu/>. After you have successfully installed R, you may install RStudio from the following website: <https://www.rstudio.com/>. If you run into an issue downloading or installing R or RStudio, you can troubleshoot on the provided websites or run a quick Google search on your issue. There are great instructions for getting RStudio up and running here: <https://ssdanalysis.com/>. Take a look at the bottom of this page for instructions on how to download R/RStudio. Throughout the course, if you have problems with RStudio on your personal computer, you will need to use the computers provided to you during class. It is your responsibility to review Kabacoff and reach out to me for help with the R language. You are expected to be able to run basic statistical analyses using the R language by the end of the class.

Learning Style: You might find it useful to assess and think about your learning style:

<http://www.educationplanner.org/students/self-assessments/learning-styles.shtml>

G*Power 3: G*Power 3 covers statistical power analyses for many different statistical tests and it works with the MAC and Windows OS. G*Power 3 and the manual can be obtained free from:

<http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3>

Online Statistics Videos: You can go to the following web sites and watch videos on various topics in statistics. This is free, but you do have to sign up to view the videos.

Against All Odds: Inside Statistics

<http://www.learner.org/resources/series65.html?pop=yes&vodid=252432&pid=139>

Learning Math: Data Analysis, Statistics, and Probability

<http://www.learner.org/resources/series158.html>

Statz Rappers (I kid you not)

<http://video.google.com/videoplay?docid=489221653835413043&pr=goog-sl>

Tables

The following web site and book provide valuable information about creating tables in APA style:

Nicol, A. A. M., & Pexman, P. M. (2010). Presenting your findings: A practical guide for creating tables (6th ed.). Washington, DC: American Psychological Association.

National Council on Family Relations: <http://oregonstate.edu/~acock/tables/>

Graphs and other Figures

The following book provides valuable information about creating graphs and other types of figures in APA style:

Nicol, A. A. M., & Pexman, P. M. (2010). *Displaying your findings: A practical guide for creating figures, posters, and presentations* (6th ed.). Washington, DC: American Psychological Association.

Web Sites and Related Resources

The following are online statistics books with links to other statistics resources on the web, and more:

www.gapminder.org
<http://onlinestatbook.com/>
<http://davidmlane.com/hyperstat/>
<http://stattrek.com/default.aspx>
<http://cast.massey.ac.nz/core/index.html?collection=Public>
<http://www.jerrydallal.com/LHSP/LHSP.HTM>
<http://www.sjsu.edu/faculty/gerstman/StatPrimer/> (also contains numerous data sets)
<http://www.sportsci.org/resource/stats/>
<http://www.visualstatistics.net/>
<http://faculty.vassar.edu/lowry/webtext.html>
<http://www.stat.berkeley.edu/users/stark/SticiGui/>
<http://www.bmj.com/collections/statsbk/>
<http://www.usi.edu/libarts/socio/stats.htm>
<http://www.seeingstatistics.com/>
<http://www2.chass.ncsu.edu/garson/pa765/statnote.htm>
<http://www.psychstat.missouristate.edu/introbook/sbk00.htm>
<http://www.itl.nist.gov/div898/handbook/eda/eda.htm>

Glossary of Statistical Terms

The following provide glossaries of statistical terms that you might find useful.

<http://www.statsoft.com/textbook/glosfra.html>
<http://www.geodata.soton.ac.uk/biology/lexstats.html>
<http://stattrek.com/Help/Glossary.aspx>
http://www.animatedsoftware.com/elearning/Statistics%20Explained/glossary/se_glossary.html
<http://dorakmt.tripod.com/mtd/glosstat.html>
<http://davidmlane.com/hyperstat/glossary.html>
<http://www.stats.gla.ac.uk/steps/glossary/alphabet.html>
<http://www.ablongman.com/html/abrami/glossary/glossary.html>
<http://dorakmt.tripod.com/mtd/glosstat.html>
Biographical Information About Statisticians
<http://www-gap.dcs.st-and.ac.uk/~history/BiogIndex.html>

Teaching Statistics

Here are good web sites for those of you who end up teaching statistics (Don't laugh. I didn't set out to do this.):

Journal of Statistics Education
<http://www.amstat.org/PUBLICATIONS/JSE/>

Statistics Education Research Journal

<http://www.stat.auckland.ac.nz/~iase/publications.php?show=serj>

Hulsizer, M., & Woolf, L. M. (2009). *A guide to teaching statistics: Innovations and best practices*. Hoboken, NJ: John Wiley.

<http://www.webster.edu/teachstats/>

Consortium (C) for the Advancement (A) of Undergraduate (U) Statistics (S) Education (E)

<http://www.causeweb.org/>

Available Data

The following provides a description of a few of the data sets available in Canvas. If you would like a dataset that is not available on Canvas, please email me directly.

General Social Survey

The following web page provides a detailed description of the General Social Survey.

<http://www.norc.org/GSS+Website/> This data set contains data for the cross-sectional portion of the 2008 General Social Survey. The data set is named gss2008.dta. The codebook and questionnaires for this data set are also contained in the GSS2008 folder.

Grades

This data set contains quiz grades, final exam grades, and final course grades for students who completed either Applied Stats at UT (1993 – 2015). The data set is named Stat Grades.dta.

Foster Parent Applicant Study

This is a study of 161 foster family applicants. The study design and variables are described in Orme et al. (2004). The data set is named Orme.dta and the article by Orme et al. (2004) is available upon request.

Bank Study

This data set and its description were obtained from SPSS (Norusis, 1988, pp. 126-127). This is a study of salaries at a Midwestern bank involved in EEO litigation. Data are available for 474 individuals hired between 1969 and 1971. This file is excerpted from H. V. Roberts' *Statistical Bases in the Measurement of Employment Discrimination in Comparable Worth: Issues and Alternatives* (ed. E. Robert Livernash). Washington, D.C.: Equal Employment Advisory Council, 1980). The data set is named bank.dta.

Employment Study

This data set and its description were obtained from Enders (2010). These are hypothetical data for 480 employees on eight work-related variables: age (years), job tenure (years), female (0 = male, 1 = female), psychological well-being, job satisfaction, job performance, turnover (intention to turnover—0 = no, 1 = yes), and IQ. The data set is named employment.dta.

Depression Study

This data set and its description were obtained from Afifi and Clark (1984, p. 3, 30-39). The data for the depression study have been obtained from a complex, random, multiethnic sample of 1000 adult residents of Los Angeles County. The study was a panel or longitudinal design where the

same respondents were interviewed four times between May 1979 and July 1980. About three fourths of the respondents were reinterviewed for all four interviews. The fieldwork for the survey was conducted by professional interviewers from the Institute for Social Science Research at UCLA. This ongoing research is an epidemiological study of depression and help seeking behavior among free living (noninstitutionalized) adults. The major objectives are to provide estimates of the prevalence and incidence of depression and to identify causal factors and outcomes associated with this condition. The factors examined include demographic variables, life events stressors, physical health status, health care utilization, medication use, life style and social support networks. The major instrument used for classifying depression is the Depression Index (CESD) of the National Institute of Mental Health, Center for Epidemiologic Studies. A discussion of this index and the resulting prevalence of depression in this sample are given in Frerichs, Aneshensel, and Clark (1981). The longitudinal design of the study offers advantages for assessing causal priorities since the time sequence allows us to rule out certain potential causal links. Nonexperimental data of this type cannot directly be used to establish causal relationships, but models based on an explicit theoretical framework can be tested to determine if they are consistent with the data. An example of such model testing is given in Aneshensel and Frerichs (1982). Only a subset of the factors measured on a sample of the respondents is included in order to keep the data set easily comprehensible. The data are from a subset of 294 respondents randomly chosen from the original 1000. Only data from the first time period are included. The data set is named depress.dta.

Magnets and Pain Relief

Magnetic fields have been shown to have an effect on living tissue as early as the 1930's. Plants have been shown to have an improved growth rate when raised in a magnetic field (Mericle et al., 1964). More recently, doctors and physical therapists have used either static or fluctuating magnetic fields to aid in pain management, most commonly for broken bones. In the case study presented here, Carlos Vallbona and his colleagues sought to answer the question "Can the chronic pain experienced by postpolio patients be relieved by magnetic fields applied directly over an identified pain trigger point?" This data set and its description were obtained from http://onlinestatbook.com/case_studies_rvls/magnets/index.html. (Go to this web site for additional information about this study.) Patients experiencing post-polio pain syndrome were recruited. Half of the patients were treated with an active magnetic device and half were treated with an inactive device. The patient was asked to subjectively grade pain at the trigger point under palpitation on a scale from 0 to 10 (0 is the least pain, increasing to 10). All patients rated their pain before ("PretestPain") and after ("PosttestPain") application of the device. Following the initial pain assessment, an envelope containing a device was randomly selected from the box containing active and inactive devices. This device was applied to the pain area for 45 minutes and then removed. The patient then evaluated his or her pain again at the region or trigger point. The treatment condition is indicated by the variable "Group;" subjects receiving treatment with the active magnet have a "1" on this variable; subjects treated with the inactive placebo have a "0." The data set is named Magnets&Pain.dta and the Vallabona et al. (1997) article is available upon request.

Instructor Reputation

How powerful are rumors? Frequently, students ask friends and/or look at instructor evaluations to decide if a class is worth taking. Kelley (1950) found that instructor reputation has a profound impact on actual teaching ratings. Towler and Dipboye (1998) replicated and extended this study by asking (a) Does an instructor's prior reputation affect student ratings? and (b) Does the size of this effect depend on student characteristics. This case study presents only data relevant to the former question. This data set and its description were obtained from http://onlinestatbook.com/case_studies_rvls/magnets/index.html. (Go to this web site for additional information about this study.)

Kelley, H. H. (1950). The warm-cold variable in first impression of persons. *Journal of Personality*, 18, 431-439.

Towler, A., & Dipboye, R. L. (1998). The effect of instructor reputation and need for cognition on student behavior (poster presented at American Psychological Society conference, May 1998).

Subjects were randomly assigned to one of two conditions. Before viewing the lecture, students were given a summary of the instructor's prior teaching evaluations. There were two conditions: Charismatic instructor ("1") and Punitive instructor ("2"): Summary given in the "Charismatic instructor" condition: Frequently at or near the top of the academic department in all teaching categories. Professor S was always lively and stimulating in class, and commanded respect from everyone. In class, she always encouraged students to express their ideas and opinions, however foolish or half-baked. Professor S was always innovative. She used differing teaching methods and frequently allowed students to experiment and be creative. Outside the classroom, Professor S was always approachable and treated students as individuals; Summary given in the "Punitive instructor" condition: Frequently near the bottom of the academic department in all important teaching categories. Professor S did not show an interest in students' progress or make any attempt to sustain student interest in the subject. When students asked questions in class, they were frequently told to find the answers for themselves. When students felt they had produced a good piece of work, very rarely were they given positive feedback. In fact, Professor S consistently seemed to grade students harder than other lecturers in the department. Then all subjects watched the same twenty-minute lecture given by the exact same lecturer. Following the lecture, subjects rated the lecturer. Subjects answered three questions about the leadership qualities of the lecturer. A summary rating score was computed and used as the variable "rating" here. The data set is named Instructor Reputation.dta.

Stress Reduction and Smoking Relapse

This data set and its description were obtained from Stevens (2002, pp. 661-652). This study by Christine Wynd (Journal of Advanced Nursing, 1992, 204-302), investigated the effect of stress reduction in preventing smoking relapse. Subjects who had completed a local smoking cessation program were randomly assigned to either an experimental group or an attention placebo control group. The subjects in the treatment group were exposed to a three session treatment, and were measured on 3 variables at each session: creative imagery, stress and smoking rate. In the data listing I have abbreviated these variables as CRE, STR and SMR. The subjects in the experimental group are coded as 1 and those in the control group as 0. The data set is named Wynd.dta and the article by Wynd is available upon request.

Changes in Attitudes of Grade School Children Toward School Subjects

This data set and its description were obtained from Stevens (1990, pp. 246-248) and a data disk supplied by Stevens. These are data collected on 189 third through sixth graders from a suburban Midwestern public school. Sex is the first variable on the file, grade is the second variable and teacher is the third variable (there were five teachers). The children were measured on preference toward the following subjects: mathematics, language arts, science, reading and social studies prior to an intervention. These are variables 4 through 8 on the file. An intervention was employed with the teachers in an attempt to change the children's preferences (attitudes). The children were posttested on these subject preferences again 4 months later. These are the remaining 5 variables on the file. The data set is named attitude.dta.

Additional Sources for Data

Inter-University Consortium for Political and Social Research (ICPSR) has a very rich array of data sets that you can use: <http://www.icpsr.umich.edu/icpsrweb/>

The following websites also have numerous public use data sets that you can use:

World Health Organization
<http://www.who.int/research/en/>

CDC Wide-ranging OnLine Data for Epidemiological Research
<http://wonder.cdc.gov/>

FedStats
<http://www.fedstats.gov/>

Centers for Disease Control and Prevention Public-Use Data Files
http://www.cdc.gov/nchs/data_access/ftpdata.htm

United States Department of Health and Human Services
<http://aspe.os.dhhs.gov/datacncl/datadir/>

Substance Abuse & Mental Health Data Archive
<http://www.icpsr.umich.edu/SAMHDA/>

Data.gov
<http://www.data.gov/>

National Center for Education Statistics
<http://nces.ed.gov/surveys/>

Roper Center for Public Opinion Research
http://www.ropercenter.uconn.edu/data_access.html

The Data and Story Library
<http://lib.stat.cmu.edu/DASL/DataArchive.html>

Carnegie Mellon University StatLib
<http://lib.stat.cmu.edu/datasets/>

Henry A. Murray Research Archive
<http://www.murray.harvard.edu/frontpage>

Sociometrics Data Archives
<http://www.socio.com/dataarchives.htm>

National Data Archive on Child Abuse and Neglect
<http://www.ndacan.cornell.edu/NDACAN/AboutNDACAN.html>

Center for AIDS Prevention Studies
<http://www.caps.ucsf.edu/tools/data/VCT/>

TheDataWeb
<http://www.thedataweb.org/datasets.html>

Rice Virtual Lab in Statistics
http://onlinestatbook.com/case_studies_rvls/index.html

Dr. Karl L. Wuench's SPSS-Data Page
<http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Data.htm>

Project on Human Development in Chicago Neighborhoods

http://dvn.iq.harvard.edu/dvn/dv/mra/faces/study/StudyPage.xhtml?studyId=307&studyListingIndex=1_7f5088797bc253a292ee18fd7b70&rvn=2

Health Information and Trends Survey (HINTS): How Americans Find and Use Cancer Information

<http://hints.cancer.gov/dataset.jsp>

National Survey of Children's Health

<http://nschdata.org/Content/Default.aspx>

National Survey of Children with Special Health Care Needs

<http://cshcndata.org/Content/Default.aspx>

The National Longitudinal Study of Adolescent Health (Add Health)

<http://www.cpc.unc.edu/projects/addhealth>

National Ambulatory Medical Care Survey

<http://www.cdc.gov/nchs/ahcd.htm>

National Hospital Discharge Survey

<http://www.cdc.gov/nchs/ahcd.htm>

National Nursing Home Survey

<http://www.cdc.gov/nchs/nnhs.htm>

National Immunization Survey

<http://www.cdc.gov/nis/datafiles.htm>

Surveillance Epidemiology and End Results Program

<http://seer.cancer.gov/data/options.html>

Behavioral Risk Factor Surveillance System

http://www.cdc.gov/brfss/technical_infodata/surveydata.htm

Youth Risk Behavior Surveillance System

<http://www.cdc.gov/HealthyYouth/yrbs/data/index.htm>

National Health and Nutrition Examination Survey

<http://www.cdc.gov/nchs/nhanes.htm>

National Health Interview Survey

<http://www.cdc.gov/nchs/nhis.htm>

Joint Canada/United States Survey of Health

<http://www.cdc.gov/nchs/nhis/jcush.htm>

Longitudinal Studies of Aging

<http://www.cdc.gov/nchs/lsao.htm>

State and Local Area Integrated Telephone Survey

<http://www.cdc.gov/nchs/slait.htm>

National Vital Statistics System

<http://www.cdc.gov/nchs/nvss.htm>

National Maternal and Infant Health Survey
<http://www.cdc.gov/nchs/nvss/nmihs.htm>

Panel Study of Family Dynamics
<http://psidonline.isr.umich.edu/default.aspx>

Multiple sub-studies/ data sets:
<http://psidonline.isr.umich.edu/Studies.aspx>

Secondary Data Analysis

The following books are good sources for the secondary data analysis. As a doctoral student, it is a very wise idea to stay up-to-date with secondary data sources relevant to your area of focus.

Boslaugh, S. (2007). *Secondary data sources for public health*. New York: Cambridge University Press.

Bulmer, M., Sturgis, P. J., & Allum, N. (2009). *Secondary analysis of survey data (Volumes 1 – 4)*. Los Angeles: Sage.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied survey data analysis*. Boca Raton, FL: CRC Press.

Smith, D. (2008). *Using secondary data in educational and social research*. Berkshire, England: Open University Press.

Vartanian, T. P. (2011). *Secondary data analysis*. New York: Oxford University Press