I. COURSE DOMAIN AND BOUNDARIES

The goals of this course are to provide students with the skills needed to be critical consumers of analyses reported in empirical research, to be able to use the SAS system to conduct select parametric and nonparametric bivariate and multivariate analyses, to correctly interpret SAS output and discuss results in writing, and to integrate understanding of various statistical techniques with research design. This course provides a broad foundation that prepares students for more advanced courses focused on specific multivariate approaches. The course complements “Introduction to Advanced Research” to help facilitate students’ integration of research methods with the appropriate statistical analyses. This course is NOT a traditional statistics course. The focus is on the application and interpretation of techniques using the SAS computer package rather than learning mathematical formulae (although these will be introduced). The course is also complemented by Foundations of Data Management that allow for further understanding of the use of SAS and how to program data.

Students should have a solid grounding in select basic statistical concepts (e.g., hypothesis testing, central limit theorem and normal distribution, and basic concept behind probability related to significance testing). Various concepts/tests are introduced with an emphasis on how they allow us to answer research questions with differing types of data (e.g., coefficient alpha, odds ratios, Two-way ANOVA, logistic regression, multivariate regression and an introduction to longitudinal analyses (e.g., survival analyses, multilevel statistics). Ethical issues related to social science research (including sensitivity to issues of gender, disability, culture, race, socioeconomic class, sexuality, religion and oppression), particularly as they relate to the appropriate use of statistics, will be addressed throughout the course. Students will also be
exposed to resources (books, further classes, and on-line opportunities) to develop a more advanced understanding of the statistical analyses as they progress in their career.

II. CORE COMPETENCIES AND PRACTICE BEHAVIORS

To develop an understanding of:
1. the relationship between statistical analysis and research design;
2. the use of scientific data in decision-making;
3. the correct application of basic descriptive and select inferential statistics
4. the use of SAS to perform basic statistical analyses
5. how to interpret statistical findings
6. how statistics can be misused and the ethical consequences of that misuse;
7. how to have fun exploring the use of statistics.

Students should demonstrate:
1. the ability to construct an analysis plan appropriate for a specific research question and design;
2. the ability to articulate how research can support decision-making;
3. the ability to select and perform the appropriate statistical test (among those covered);
4. a beginning knowledge of the use of SAS to perform basic statistical analyses;
5. the ability to interpret test results and identify potential implications, including consideration of statistical power;
6. how to identify and appropriately resolve ethical issues regarding statistics;
7. a reduction in initial fear of the use statistics.

III. BROWN SCHOOL ACADEMIC POLICIES

**Academic Integrity and Professional Conduct:** All sentences must be comprised of the student’s own words. Ideas, information, and concepts that originated with any other source must always be noted as such (based on APA format), including past work that you have authored. Please use plagiarism tools if you are unsure if your paper is in violation. Material that is not correctly cited (or content that is copied from another student’s work) is considered plagiarized and provides grounds for academic discipline. Please review the Ph.D. Program Policy and Procedures Manual (on InsideBrown) and the Graduate School of Arts and Sciences Academic Integrity Manual for further guidance on all matters of academic integrity and professional conduct for graduate students at Washington University.

**Special Needs:** If you have a learning disability, sensory, or physical disability or other impairment, and you may need special assistance in lectures, reading, written assignments, and/or exam taking, please contact the Brown School Director of Student Affairs who can provide coordination of accommodations at Washington University and the Brown School. The Disability Resource Center, a University-wide resource, provides diagnostic and academic accommodations support and referrals.

**English Language Proficiency:** If your English language proficiency is such that you may need special assistance in lectures, reading, written assignments, and/or exam taking, please communicate these needs to your instructor who may refer you to the English Language Program.
(ELP), a University-wide resource which provides classes and academic English language support designed to increase non-native English speaking students' English language proficiency and to facilitate their academic success at Washington University. You may also find the Academic Assistance resources available through the Office for International Students and Scholars to be helpful.

Professional Use of Electronic Devices in the Classroom: Computers or other electronic devices, including “smart pens” (devices with an embedded computer and digital audio recorder which records the classroom lecture/discussion and links that recording to the notes taken by the student), may be used by students at the discretion of the faculty member to support the learning activities in the classroom. These include such activities as taking notes and accessing course readings under discussion. If a student wishes to use a smart-pen or other electronic device to audio record lectures or class discussions, they must notify the instructor in advance of doing so. Permission to use recording devices will be at the discretion of the instructor, unless this is an accommodation approved by Disability Resources.

Nonacademic use of laptops and other devices is distracting and seriously disrupts the learning process for everyone. Neither computers nor other electronic devices are to be used in the classroom during class for non-academic reasons. This use includes emailing, texting, social networking, and use of the Internet. The use of cell phones during class time is prohibited, and they should be set on silent before class begins. In the case of an emergency, please step out of the room to take the call. The instructor has the right to hold students accountable for meeting these expectations, and failure to do so may result in a loss of participation points, a loss of the privilege of computer use in the classroom, or being asked to leave the classroom.

Religious Holidays: The Brown School recognizes the individual student’s choice in observing religious holidays that occur during periods when classes are scheduled. Students are encouraged to arrange with their instructors to make up work missed as a result of religious observance, and instructors are asked to make every reasonable effort to accommodate such requests.

IV. WASHINGTON UNIVERSITY ACADEMIC SUPPORT POLICIES

Accommodations based upon sexual assault: The University is committed to offering reasonable academic accommodations to students who are victims of sexual assault. Students are eligible for accommodation regardless of whether they seek criminal or disciplinary action. Depending on the specific nature of the allegation, such measures may include but are not limited to: implementation of a no-contact order, course/classroom assignment changes, and other academic support services and accommodations. If you need to request such accommodations, please direct your request to Kim Webb (kim_webb@wustl.edu), Director of the Relationship and Sexual Violence Prevention Center. Ms. Webb is a confidential resource; however, requests for accommodations will be shared with the appropriate University administration and faculty. The University will maintain as confidential any accommodations or protective measures provided to an individual student so long as it does not impair the ability to provide such measures.

If a student comes to me to discuss or disclose an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if I otherwise observe or become aware of such an allegation, I will keep the information as private as I can, but as a
faculty member of Washington University, I am required to immediately report it to my
Department Chair or Dean or directly to Ms. Jessica Kennedy, the University’s Title IX
Coordinator. If you would like to speak with the Title IX Coordinator directly, Ms. Kennedy can
be reached at (314) 935-3118, jwkenndedy@wustl.edu, or by visiting her office in the Women’s
Building. Additionally, you can report incidents or complaints to Tamara King, Associate Dean
for Students and Director of Student Conduct, or by contacting WUPD at (314) 935-5555 or your
local law enforcement agency. You can also speak confidentially and learn more about available
resources at the Relationship and Sexual Violence Prevention Center by calling (314) 935-
8761 or visiting the 4th floor of Seigle Hall.

Bias Reporting: The University has a process through which students, faculty, staff and
community members who have experienced or witnessed incidents of bias, prejudice or
discrimination against a student can report their experiences to the University’s Bias Report and
Support System (BRSS) team. See: brss.wustl.edu

Mental Health: Mental Health Services’ professional staff members work with students to
resolve personal and interpersonal difficulties, many of which can affect the academic
experience. These include conflicts with or worry about friends or family, concerns about eating
or drinking patterns, and feelings of anxiety and depression. See: shs.wustl.edu/MentalHealth.

V. READINGS

On-line
We will be using portions of several texts, articles and a few other readings on specific
techniques all on Blackboard. Some of the articles are studies that happen to use the statistical
method for that week and were chosen for this reason NOT as exemplars of research on a given
topic. These articles are offered as a guide for writing up results for publication. Students are
encouraged to review them and bring questions about them relevant to the material covered in
class or lab to further their understanding of the application of statistics to social science
research.

RECOMMENDED:
(1) APA Style Manual: Students are strongly encouraged to access the latest version of the APA
manual. It is available in the library and parts can be accessed on-line.

VI. ORGANIZATION OF COURSE

The class includes a lecture, discussion and microcomputer lab application. Although we will be
covering a rather wide range of techniques, the underlying concepts are largely cumulative in
nature (for example the use of the ANOVA builds upon the use of the t-test, etc.). There is a
strong emphasis on exposure to categorical data methods in the last half of the class rather than
progressing further into linear regression. This is due both to the fact that the focus of the
subsequent required PhD statistics class is on linear regression and because categorical outcomes
are common in data in social work, public health and other social sciences. There are two “flex
weeks” built into the course. If needed, this will allow us to spend more time on a subject
already covered. If not, we will proceed with an introduction to a special topic like Propensity
Score techniques. While data sets are provided to support the labs, students are highly
encouraged to locate a dataset they are interested in analyzing no later than Week 5 so that they
can use their own data to work on the labs and be better prepared for completion of the final paper.

VII. ROLE OF FACULTY AND STUDENT

Instructor: Class meetings will be primarily lecture/discussion format. The instructor will be responsible for the preparation and response to student inquiries regarding the readings, exercises, and the course project. Additionally, the instructor will be available after class and by appointment for questions regarding the course. Lecture powerpoints will be posted AFTER the class they are discussed. The instructor is also responsible for timely feedback on assignments.

Teaching Assistant: Autumn Asher, a doctoral student, will be the lab instructor for this course. Contact information is on the top of the syllabus. Lab sessions will be used to practice application of the techniques discussed. The TA also proctors course quizzes which are completed during lab time. The TA will also assist students with assignments during office hours.

Students: This is an applied course that requires significant attention to out-of-class activities as well as readings. Students will get lost without regular attention to readings and attendance at all scheduled meetings. Students are expected to seek assistance and clarification when needed, complete course work and assigned readings, and provide feedback regarding the effectiveness of the class. Students are encouraged to ask questions regarding articles/text passages in the online folder or other outside articles read. If the questions are from an article not provided to the class, students are asked to e-mail the instructor at least 2 days prior to class so that the instructor can read the article and respond and provide a copy of the article to the class. Lab assignments are due at the start of the class following the lab.

Students are also responsible for notifying the instructor in advance of the need to miss a class (e.g., religious holiday) and are responsible for obtaining notes from other students and completing assignments. In case of emergency or illness the student is responsible for notifying the instructor as soon as possible and for requesting extensions for assignments due to illness or emergencies as needed.

Although there are many options for statistical packages, we will use SAS on microcomputers. Examples of computer output will be discussed in class and reviewed again during labs. A complimentary data programming course is offered during the first semester. The expectation is that students will become independent and proficient in basic SAS use. SAS is also available in the PhD computer lab. Students are encouraged to branch out and explore the use of other statistical packages such as SPSS, STATA, R, etc. but these are not covered during class or lab.

Because we cover such a broad range of techniques, students will have access to data sets that can be used for the specific lab assignments. The electronic data sets will be placed on a g-drive or other student drive for easy access and contain no confidential material. Students will receive a handout by the second week of class that describes the content of the data sets that are not in the texts. Students must use an outside dataset for the final project. Once identified, students are strongly encouraged to use their data set for labs whenever appropriate to the statistical technique
covered in that week to increase their familiarity with the data. Students will receive assistance from the TA in locating secondary data. Lab time will be dedicated to the computer and technical writing skills necessary to complete assignments and the final project.

VIII. ASSIGNMENTS AND GRADING CRITERIA

Students will receive a letter grade based upon the percentage of points obtained in completion of homework assignments, quizzes, and a final paper.

Weighting for final course grade:

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labs (weeks 1-8, 10-11, 13-14)</td>
<td>35%</td>
</tr>
<tr>
<td>3 Quizzes</td>
<td>15%</td>
</tr>
<tr>
<td>2 Final Paper prep assignments</td>
<td>10%</td>
</tr>
<tr>
<td>Final Paper</td>
<td>40%</td>
</tr>
</tbody>
</table>

Grading Scale:

<table>
<thead>
<tr>
<th>%</th>
<th>Grade</th>
<th>%</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>98-100</td>
<td>A+</td>
<td>77-79.9</td>
<td>C+</td>
</tr>
<tr>
<td>94-97.9</td>
<td>A</td>
<td>74-76.9</td>
<td>C</td>
</tr>
<tr>
<td>90-93.9</td>
<td>A-</td>
<td>70-73.9</td>
<td>C-</td>
</tr>
<tr>
<td>87-89.9</td>
<td>B+</td>
<td>84-86.9</td>
<td>B</td>
</tr>
<tr>
<td>80-83.9</td>
<td>B-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lab Assignments:
Topics for the labs and outlines of requirements are listed under the computer lab section for that week on the syllabus. **The assignment may be altered slightly by the instructor and/or TA to better fit the appropriate level of challenge for the class.** Each lab write-up is due the following week in class (morning). These are short assignments to practice the use of SAS to conduct various analyses, the technical skills involved in interpretation and writing up discussion of results, and to help you progress toward the final paper. Analysis assignments can be completed with data provided by the instructor (outlined above) or other data identified by the student. If the student chooses to use their own data they must check to see that the dataset has the type of variables needed to complete that week’s assignment-otherwise they should use the data provided. There are fourteen pass/fail lab and final paper prep assignments that together comprise 45% of your grade as noted above.

With the exception of the first lab which is used to revisit some basic SAS commands and descriptive procedures and the two assignments designed to provide you assistance with your final paper, the other 11 labs must include: (1) a research question, (2) hypothesis (when appropriate), (3) brief description of the dataset itself (n size, what it was collected for…), (3) mention of how checking the assumptions required to run a given statistical test was done and what the checks told you about any alterations to data needed, (5) operationalization of variables – your reader should know about any recoding done and what the variable codes mean, (6) presentation of results (APA style), (7) discussion/interpretation of results (what do the results say or not say, what are the limitations of the data used, what might you do to further explore the question- need more data, different analyses….). Please use APA guidelines for any citations
and reporting of data in tables and text (this will help you prepare for article writing). The on-line articles are helpful resources as examples for writing up results as well.

Although ungraded, each lab and paper prep assignment completed on time receives 5 points. The goal is skill mastery and therefore students may be asked to redo assignments and turn in a second time if the first submission is unsatisfactory (these ‘redos’ are not graded as late). Choosing not to redo the assignment when asked will result in only half credit for that assignment. Labs turned in late without a valid excuse (e.g., illness…) will also receive half credit.

Additional Challenge: The difficulty of course assignments may vary for some students. Some of the labs may be particularly easy for some to execute-particularly early in the semester. However, the goal is not just to execute a given statistical test but to do so in combination with methodological considerations and critical thought about interpretation. Students who desire more challenge on a given lab are encouraged to develop a brief introduction or rationale for the study question (drawing on published literature) as you would in an article and discuss potential or hypothetical implications of results and further ideas for more research. Students are also encouraged to use their own data for as many of the lab assignments as possible and will find it useful to review the recommended as well as the required readings.

Quizzes:
Three brief quizzes will be given during lab. These quizzes will only cover the weeks prior to the quiz and are intended to check on students’ understanding of the terminology and appropriate reasons to use the tests covered. Study guides are provided one week before the quiz.

Final Paper:
The purpose of this paper is twofold: (1) to assess the student’s ability to apply at least one of the covered statistical techniques—beyond simple bivariate techniques covered—to a research question of their choice; (2) to help synthesize data analysis skills with research methods. It provides students with practice in writing in the form of a journal article rather than a class paper. The paper should be no more than 15 double-spaced pages (excluding abstract, tables, references, and any appendices). The focus is on practicing skills not on mastery of a substantive area so the literature review will be brief. In other words, include enough literature to provide some background for the research question(s), but don’t worry about making a brilliant and exhaustive summary of the topic. Students may use the same topic as the paper in Introduction to Advanced Research or a data source identified in your data programming course, however students may not use completed analyses from a prior project or a different course—the analysis must be your own original work done for this class. The bulk of the grade will be based on the correct selection, execution, and interpretation of the analyses (see sections 1.3c, 1.4, 1.5 below). The paper should include the appropriate APA documentation in the text and tables as well as references. Students are encouraged to limit their choice of analysis technique for the final paper to those covered by week 12 in order to have sufficient time to finish the paper.

STUDENTS ARE ENCOURAGED TO USE THE WRITING LAB FOR ASSISTANCE WITH GRAMMAR/SPELLING. Writing counts in the grading just like it will when you submit articles for publication! A detailed grading rubric is available on Blackboard but the following includes all the elements expected.
1.1 Abstract (150 to 200 words maximum)

1.2 Objectives and Background of Project

A. Introduction (1 page)
   What is the problem/issue that the analyses will address?

B. Brief background literature (1-2 pages)
   How does your effort build on or add to this literature?

C. Specific research questions &/or hypotheses to be addressed in this study.

1.3 Methodology (3-4 pages)

A. Sample (who, how many, how obtained)
   a. Human Subjects concerns addressed

B. Operationalization of variables used

C. Data analysis plan
   a. Power analysis (if applicable)

1.4 Results (APA format) with tables and/or graphs (2-4 pages)

1.5 Discussion: Relation to Background Literature, Implications & Limitations & A conclusion (3-4 pages)

1.6 References

IX. COURSE OUTLINE

Class 1: August 28, 2018 Statistics and social science research: Diving in!
This class will introduce the topics to be covered, expectations and organization of the course. Then we dive into the importance and means of conducting univariate exploration of your data and touch on the underlying issues of power and error. We will begin discussion of how to assess for violations of assumptions of statistical tests (level of measurement, normality, outliers), the concept of robustness, and “fixing” data problems. I have chosen to go over “power” issues early as they really impact interpretation of any of the techniques we will cover. I know we have not yet learned how to use the test statistics, but understanding these issues are basic to their interpretation. We will discuss how these concepts are linked to the study design and selection of the test statistic. You have been exposed to most of these concepts before, but we will discuss their ideal and practical application.


Recommended reading if you want further explanation of underlying power formulas:

Lab Assignment 1:
This lab DOES NOT require a research question.
A. Choose two continuous level variables from the data set and create a frequency distribution, use PROC MEANS and request VAR, KURTOSIS and SKEWNESS options in SAS to compute means, range, standard deviation, variance, and examine departures from normality. Use PROC UNIVARIATE with NORMAL and PLOT options to examine the results. Assignment: Describe the distribution, variance, presence of outliers, and any issues related to normality for these
variables. Will you be able to use a parametric approach with these variables? What, if any, “fixes” are required to provide the most accurate results?

B. Now, try running a power analysis exercise for a t-test using PROC POWER. Student will select a continuous level variable and a dichotomous variable from a data set and run descriptive statistics (Use PROC SORT to sort the data by the independent variable. Use PROC MEANS with a “by” statement). 

Assignment: Pretend you are using the variable you selected for a project. Write a research question and hypothesis appropriate for comparing the means of 2 groups. Why might you use a one-tailed instead of a two-tailed hypothesis? Describe how the results of your power analysis might impact your research design.

C. Next choose two categorical variables: like age and a dichotomous outcome. Use PROC UNIVARIATE to examine the mean and median values for age. RECODE AGE using an IF/THEN statement into a 2 or 3 category variable. Use PROC FREQ to print out a cross tabulation of your recoded age variable by report. 

Assignment: Was there a large difference between mean and median age? How did you make the decision to collapse age into the categories you chose (i.e. justify the categories you selected)? Do you see any patterns in the data presented in the crosstabulation?

Class 2: Tests of Association
Many texts do not cover categorical techniques until the end, but why wait? Sometimes the research question of interest involves understanding whether or not two variables are associated. For example, is being a graduate student associated with liking cheese? In this class we will discover the wonder of the chi-square family and discuss when to use the Pearson Chi-square, Exact tests, Trend or McNemar’s Chi-square.


Computer Lab 2: 
Assignment: From here on you will want that research question! Remember to include any preliminary assumption checks in all labs before discussing results!!
(1) Students will use a dataset (TBA) and select 2 dichotomous variables they think might be associated. Use PROC FREQ with the CHISQ OPTION.
Write a research question that is appropriate for the variables selected. Use APA style to report the results appropriate to the sample size. Interpret the results including significance and the direction of the association.
(2) Next students should choose an ordinal and dichotomous variable and use a trend test. Write a research question that is appropriate for the variables selected. Use APA style to report the results appropriate to the sample size. Interpret the results including significance and the direction of the association.

Class 3: Correlations
More associations! Unlike chi-square, correlations describe relationships between ordinal or continuous data –like income and IQ. In addition to bivariate analyses, correlations are used in other applications like variable reduction and scale reliability. We will briefly cover the bivariate correlation then discuss coefficient alpha. Later, after examining tests of difference, we will return to the concepts of correlation when we discuss Principal Components Analysis and simple regression.


On-line folder article examples that include correlation:


Computer Lab 3:
Assignment: Remember to include any preliminary assumption checks in all labs before discussing results!!
(1) Students will identify 3 variables that may vary together. Use PROC PLOT to verify a linear relationship. Use PROC CORR to run a 3 by 3 correlation matrix. Assignment: Write a research question that is appropriate for the variables selected. Enter the results in a table form as illustrated in a journal or use the APA manual. Discuss significant relationships, including the strength and the direction. Is the association practically important? Why or why not?
(2) Students will also be given a dataset based upon a survey and be asked to run a coefficient alpha. First, assume that you know nothing about the survey and just run the procedure based on all items. Assignment: Interpret the result; is the reliability coefficient acceptable? If not, examine the correlation coefficients and see if there may be subscales present. If so, re-run the procedure using the subscales. Summarize the results in a table using APA style.

Class 4: T-tests, ANOVAs and non-parametric alternatives
Something old and something new 😊. This week we will move back to the familiar world of parametric statistics and normality for a while. This takes us from associations to group differences when the outcome of interest is a continuous or ordinal variable. In this class we will review and extend the concepts learned about t-test and explore the One-way ANOVA procedure. We will discuss its use, a priori versus post hoc procedures for identifying where
differences are, and the Kruskal-Wallis test as an alternative when assumptions cannot be met. Well because the F table we see in the ANOVA reappears in regression so it’s nice to get used to all the basic pieces.

**Required readings:** Drake & Jonson-Reid, Chp 13, pp 275-288 [RESERVE](#), Chapter 13; Field & Miles (2010), pp 466-482.

**On-line folder Article examples that include t-test/ANOVA**


**Computer Lab 4:**
Assignment: Remember your research question(s) and to include any preliminary assumption checks in all labs before discussing results!!

1. Students will be asked to select a dichotomous (independent) and a continuous (dependent) variable from one of the class datasets. Use PROC Univariate to explore assumptions for the use of a t-test. If the dependent variable violates the assumptions too badly choose another variable. Use PROC TTEST to analyze. Then use the Wilcoxon-Mann-Whitney test. Assignment: Comment on the assumptions. Write a research question and the hypothesis for a 2-tailed test. Discuss the results of the TTEST compared to the Wilcoxon-Mann-Whitney test. Did you reject the null hypothesis? What is the practical difference?

2. Students will use data provided in lab to select variables to use for an ANOVA. Develop a question that deals with the difference between 3 or more groups. Make sure you check and report on assumptions. Use SAS to conduct an ANOVA and post hoc using the Tukey method. Students should also attempt to use the same variables to run a Kruskal-Wallis test using the NPAR1WAY procedure. Assignment: Write a research question that fits the variables you have selected. Report the findings of the ANOVA in a table according to the APA style. Write a paragraph explaining the results of the ANOVA and post hoc analyses.

**Class 5: Factorial ANOVA**
Life is not simple so often our stats cannot be either! ANOVA is great when you only have a 1xR design, but it’s limited when it comes to social science research. For example, there are often cases in which one wants to compare the outcomes according to two different populations (e.g., IQ scores for 3 different stimulation groups according to gender). We will discuss the application of the Factorial ANOVA in this class. We will also introduce the “interaction term”.


Computer Lab 5:
Students will take a brief quiz (15-20 minutes) during lab on materials through class 4.

Assignment: Remember to include any preliminary assumption checks in all labs before discussing results!! Students will use data provided in lab to run a Two-way (factorial) ANOVA. Try running the parametric and non-parametric versions even if your data meet assumptions Write a research question for your ANOVA. Look at the possible interaction term. Then write 1 or 2 paragraphs interpreting the results of each analysis. As these data are being used to evaluate the effect of a program, what might you say to the program operators based on your analyses?

Class 6: Introduction to Principal Components and Regression
Bringing ideas together….For regression we want our variables of interest to be correlated—but not toooo correlated. One of the advanced uses of correlation techniques are used in variable reduction (principal components) to handle issues like wanting to include variable in a later multivariate regression that may be too correlated to enter as separate independent variables. In this class we will review correlation and discuss the extension of correlation to simple regression. We will introduce multicollinearity and see how Principal Components might help us address this problem.

Readings:  Drake & Jonson-Reid, Chp 13, 295-298 & Chp 14, 309-315 RESERVE

Computer Lab 6:
(1) Students will use census data or other data provided and run a principal components procedure using at least 7 variables. *Assignment*: Discuss your results. How many factors did you keep? What type of variables seemed to hang together? Did the resulting factors make intuitive sense?
(2) Using data set provided in lab or your own dataset to calculate and interpret two separate bivariate regression models using the same dependent variables for both models. Then run a
multiple regression model in which both independent variables are entered simultaneously.

**Assignment:** Write a research question appropriate for regression. Do the regression coefficients the independent variables change when they are both in the equation? Identify model fit indicators and report those and coefficient magnitude. Include the regression equation in your write-up.

**Class 7: Multivariate Regression Continued.**

So last week we just took a peak at a regression output, but this week we will more thoroughly explore aspects of a multivariate linear regression model and take a peak at related diagnostics.

**Readings:** Drake & Jonson-Reid, Chp 14 pp 303-308; Hatcher, L (2013). *Advanced statistics in research: Reading, understanding, and writing up data analysis results.* Shadow Finch Media; pp 249-293. [RESERVE]


**Computer Lab 7:**

Students will use either their own data or a data set provided to construct a multivariate regression model with at least 3 variables and one interaction term.

**Assignment:** Write a research question appropriate for regression. Students should report on both assumptions and diagnostics but they are not expected to “fix” the model if it fails only to know what that means and next steps. Write the regression equation in equation format. Interpret and report the regression findings using APA format.

**Class 8: Variations on Regression & ANOVA: ANCOVA**

We will wrap up our understanding of the variations in the ANOVA Regression families that deal with controlling for factors outside group membership – one technique sometimes used when an RCT does not go as planned

**Required readings:** Drake & Jonson-Reid, pp316-320 [RESERVE].


**Lab 8**

Students will take use their own data or data provided to run an ANCOVA.
Assignment: Write an appropriate research question. Students should include discussion of assumptions and interpretation of results using APA style.

Class 9 Flex week
This week is for you to decide! We can go on and cover an additional form of regression OR we can review. I will be asking for a class vote 2 weeks prior. If you wish to forge ahead, we will introduce the Poisson variation on regression. Why? Well sometimes we are interested in understanding a count of things like number of reports or number of hospital visits. Sort of seems continuous but not really and they are certainly not going to be normally distributed. You will not be tested on this content. There will be no lab assignment so you can work on your project.


Computer Lab 9:
(1) Students will take a quiz during first 25 minutes of lab covering weeks 5-8.
(2) Bring your data set you plan to use for the final project (or have it saved to the network). Please turn in a brief description of your selected data set as you written lab assignment: this should include the n size, the unit of analyses (person, community…), and a brief description of the data collection methods (interview, survey, administrative data, census data…). You should be sure you have successfully downloaded the data and begin work on selecting variables you will use and running descriptive and bivariate statistics. Please use this time to assess and clean your data so you can move to constructing an appropriate research plan. A complete data analysis plan including a research question, operationalization of variables, and description of the statistical technique you will use is due by the end of Lab on Week 12 (see further note in Week 12) so you can receive feedback in time for the paper!

Class 10: Return to the categorical: Inference and Measures of Association, Cochran Mantel-Haenszel & Loglinear models
First, we discuss the use of the Mantel-Haenszel Chi-square and loglinear techniques that go beyond the simple chi-square for higher order cross-tabulation. We will discuss the odds ratio and some of its cousins like relative risk why they are used and how to interpret them this will lead us gently to logistic regression which we will cover in more detail next week.


Lab 10
(1) Students will be guided in the selection of variables from a dataset (TBA) and compute the CMH statistic for a set of 2 x 2 tables. Assignment: Write a research question. Students should write up the results in a paragraph. What did you learn about the relationships between the variables you examined?
(2) Students will use the same outcome and 2 independent variables from exercise #1 and construct a logistic regression model using PROC LOGISTIC. Enter the independent variables one at a time. Assignment: Write a research question. Students should compare the model chi-square with one as compared to the 2 variables and interpret the results in a paragraph. What did you learn about the relationships between the variables and the outcome you examined? Were the variable coefficients significant? Look at the Odds Ratios are they practically and statistically significant?

Class 11: Logistic Regression: Continuous variables, Interactions, and ROC
We will review the interpretation of interaction terms in a logistic regression model. We will also discuss means of interpreting continuous variables and calculating Odds Ratios to compare various variables in a model. The use of ROC analysis to assess predictive use of a model will be introduced.

Required readings: Drake & Jonson-Reid, Chp 14, pp331-340 RESERVE.
Other on-line reading: Allison, P. (2012). pp 40-80 from Chapter 2 from Logistic Regression Using SAS.


Lab 11
Students will use the same data as last week, but this time they will add age as a continuous variable in the model and then re-run the model with an interaction term in PROC LOGISTIC. Request the output for ROC for both models. Assignment: Write a research question. Interpret both models, including the overall fit and the results for the continuous variable and interaction term. Was the interaction term significant? Did it alter the model fit or predictive value in a significant way?

Class 12 Flex Week Review or Extensions of Logistic Regression: An Introduction to Propensity Scores and small sample issues
Another “its up to you” moment! There are a LOT of variations on a theme for logistic regression—much more than we can cover here. If you wish review, we can review. If you wish to move on we will! If we move on we will talk about options like how to handle clustering in
data. In this class we will also introduce the use of logistic regression in creation of a propensity score and an introduction to propensity score matching and/or exact logistic regression for small samples.

**Required readings if we move on (on-line):**


**On-line folder Article examples:**


**Recommended further reading:**

**Lab 12**
Students will use the lab to work on their project. **Assignment:** This time is to be used to complete the work begun in week nine. By the end of the lab you will: (1) Turn in the research question and operationalized list of dependent and independent variables to be used; (2) Identify the analysis technique and describe progress in data cleaning and preparation.

**Class 13: Controlling for Time**
In this class we will begin our two week discussion of techniques for looking at longitudinal data. This week we will look at the repeated measures ANOVA—useful for questions of maintenance of effect over time or measures at discrete time intervals and an extension of the ANOVA. We will also explore an alternative extension of logistic regression for repeated dichotomous outcomes using Generalized Estimating Equations.

**Required Reading:** Drake & Jonson-Reid, Chp 14 341-346; Allison (2012) pp226-229
RESERVE


Computer Lab 13: Using a data set provided or your own, students will either (A) run an One-way repeated measures ANOVA. Or (B) use the GEE method if you have repeated dichotomous events. Assignment: Write a research question appropriate to the technique. Then write 1 or 2 paragraphs interpreting the results of the analysis.

Class 14: Survival Analyses
No, this is not about choosing who remains on the “island” after a challenge… Imagine you want to know what the chances are of a person becoming poor. If the outcome is “yes” or “no” then you might use a chi-square method right? But, what if your chance (or risk) of being poor changes over time? Herein lies the beauty of survival analyses. We will only have time to look at a bivariate approach in-depth, but we will discuss how this can be extended into the multivariate Cox Regression models in our final week.

Readings provided on-line: Drake & Jonson-Reid, Chp 14, pp347-357


Computer Lab 14:
(1) Students will take a quiz during first 20 minutes of lab on weeks 10-11 & 13.
(2) Students will be given a small dataset with a start and end dates, a dichotomous independent variable and a dichotomous count of an event. Students will create a time variable using SAS. Students will use PROC LIFETEST to produce a survival curve and significance test. Assignment: Interpret the output. Were there significant differences in the occurrence of the event over time for the 2 groups? Look at the survival curve, does the occurrence of the event seem proportional over time for both groups?

The remainder of the lab should be used to work on final modeling for your paper.

Class 15: Introduction to Cox Regression, Final Questions…
NOTE: Final Paper and Final Lab is due by start of class!!
Congratulations! You have made it through the first statistics course in your PhD program! This week we will see how some of the things you learned in week 14 are then applied in a multivariate survival model. We will overview assumptions and briefly discuss the similarities and differences between a hazard ratio and our old friend the odds ratio. Then we will chat about future courses, answer any remaining questions and send you on!

A SAS User’s Guide to Stata:
http://www.cpc.unc.edu/research/tools/data_analysis/sas_to_stata/index.html

**On-line folder Article examples:** Jonson-Reid, M., Kohl, P. & Drake, B. (2012). *Child and Adult Outcomes of Chronic Child Maltreatment* *Pediatrics*


**No lab!**