INTRODUCTION

From the major topic headings you will note that we will focus on four statistical procedures: regression analysis, logistic regression, discriminant analysis, and factor analysis. But the course is primarily a course on multiple regression and related procedures. In class I will generally focus on the basic concepts and the most essential material. As this course is designed for students who vary greatly with respect to the strength of their mathematical background and extent to which they expect to be using these skills in the future, I have attempted to build in some flexibility with respect to what you will cover in your assignments. Everyone will be expected to demonstrate competence with respect to the basic material covered in class. This will provide an exposure to what multivariate statistics is all about. Those of you have reasons to go beyond this basic material will be given an incentive (extra credit) for doing so. After you have taken this course you should know when to use each technique, how to set up SPSS runs using the technique, and how to interpret what you see in computer printouts using the technique. I also want you all to be able to critically evaluate journal articles using these techniques. To that end you will be asked to read and critically evaluate the statistical analysis in published articles using each of these techniques.

COURSE REQUIREMENTS

Approximately every other week I will give an assignment. I strongly urge that you do it by the specified due date. In some cases I will ask you to read and comment on an article that uses a particular statistical procedure. If you turn these assignments in ON TIME, the course TA will provide feedback and suggestions for revision for assignments 1, 2, 4, and 5, but not for assignments 3 or 6 (as they will be due at the same time as the mid-term and the final paper respectively). If after doing an assignment you have doubts about what you did or how you wrote it up, come to see me or the TA, and go over what you did in person. If you have any questions about the comments the TA has made, go see the TA for elaboration or come to see me. If you have doubts about a particular part of an assignment, it might make sense to say so or to ask the TA to take a particularly close look at that part of your write-up. I will suggest that you work on one of the datasets that I have prepared for the course. However, if you come to see me, it will be possible to obtain permission to use your own data at least for some of the assignments. In a folder called DATA you will find several datasets (GSS2004, PCNORC94, GSS2000, LES101, Wk228, and WSP362); for each there is an SPSS system file (file name ending in .sav). You will also find a corresponding codebook for each as a text file (file name ending in CB.rtf).

GRADING

Grading will be based on how much of the material in the lectures, printouts, and reading (including material covered in the reading that we do not have time to cover in class) that you can convince me that you have learned. The grade will be based on how much you cover and how well you seem to understand what you cover.

You should allocate the space so that you touch on all of the important concepts, but spend the most time (words) showing that you understand the more difficult concepts and issues. It is bad strategy to spend most of the space discussing very basic material, if you fully understand the more advanced material. As I need to be able to check your work, grading will also be influenced by the clarity not only of the writing, but also the clarity of efforts to present relevant documentation from your SPSS runs (including the relevant material from syntax files) to support assertions in the text of your paper. I need to be able to find that documentation very quickly. All of your written work must be your own work, but you can ask a friend for help in getting your SPSS runs to work. If you are not fluent in English you do have permission to get help on that issue from the Connors Family Learning Center (617 552 0611), Room 200 in O’Neill Library. Please go to the following website for information about the issue of academic integrity: http://www.bc.edu/integrity.

MID-TERM PAPER (Due Thursday February 26 @3:00 pm). This paper will consist of Assignments 1, 2, and 3. It will make up 40% of your grade for the course.

FINAL PAPER (Due Thursday May 7 @ 3:00 pm). This paper will consist of Assignments 4, 5, and 6. The Final Paper will make up 60% of your grade.

COURSE TEXTS
REQUIRED

Norusis, Marija J. 2008. *SPSS 16.0 Statistical Procedures Companion*. Upper Saddle River, NJ: Pearson/Prentice Hall. This will be a very useful book as it will discuss various statistics that are presented in your SPSS output. If you buy on line, you may want to buy a less expensive (and thus less recent edition). The changes from edition to edition tend to be very modest. I have selected the 16.0 edition as it correspondence to the current version of SPSS at BC, but you never know when they will update BC to a more current version.

SUGGESTED FOR A MORE ADVANCED TREATMENT
Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*. Fifth Edition. Boston: Pearson/Alyn and Bacon. This text is suggested for a more advanced treatment than you will find in Warner (2008) for some issues. I have compared this 5th edition with the prior (2001) 4th edition and the differences are quite minor. If you have access to a 4th edition do not hesitate to use it, but keep in mind that in the prior edition logistic regression is covered in Chapter 10 (not 12) and discriminant analysis is covered in Chapter 9 (not 11). (O'Neill) (GSSW)

SUGGESTED FOR A MORE BASIC TREATMENT
Polit, Denise F. 1996. *Data analysis & Statistics for Nursing Research*. Stanford, CT: Appleton & Lange. This will often be a useful source for a basic treatment of the topic. While the book is somewhat dated, it provides a good basic overview of all the topics we cover. (O'Neill)


SUGGESTED FOR SPECIFIC TOPICS

Keith, Timothy K. 2006. *Multiple Regression and Beyond*. Pearson Education: Boston. This book provides a thorough treatment of many of the issues we cover in this course. It is not written at an overly sophisticated level. You might find this book very useful for multiple regression, path analysis, and related procedures. (O'Neill)


TOPICS AND READINGS

Items on reserve at O'Neill are indicated by (O'Neill) at the end of the citation. Items marked as (GSSW) are in the library for the Graduate School of Social Work. Most assigned articles, SPSS runs, and datasets can be downloaded from MyFiles by going to https://wfs.bc.edu/jbw/SC703 . It is best to download files after I tell you to, so that you will have this year’s (and not last year’s version) of the file. Many can also be found using the BC online course reserves at <www.bc.edu/libraries>. The staff at O'Neill reserve room can help you download articles.

**Items that everyone will be required to read are marked with an asterisk.** Each of you will read items without an asterisk selectively depending on whether you are looking for a more basic or a more advanced treatment and depending on which version of SPSS you are working with. You will always want to read at least one of the SPSS sources in connection with each procedure. I recommend you start with *SPSS 16.0 Statistical Procedures Companion*.

I will be leaving it up to you to decide which of the various sources to use in connection with each procedure and topic. It will depend on how sophisticated a treatment you want. Most of you will eventually figure that out and do most of your basic reading from just two or possibly three sources. When I cover a technique, concept, statistic, option, etc. after class look it up and read about it in at least one of the sources that I have provided. If you are having trouble finding a relevant discussion see me or the TA for the course.
MULTIPLE REGRESSION & RELATED TECHNIQUES

OBJECTIVES: To be able to do multiple regression and path analysis using SPSS. After we finish this section of the course you should understand such concepts as:

<table>
<thead>
<tr>
<th>slope</th>
<th>tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>residual analysis</td>
</tr>
<tr>
<td>least squares line</td>
<td>standardized residuals</td>
</tr>
<tr>
<td>simple vs. multiple regression</td>
<td>assumptions about residuals</td>
</tr>
<tr>
<td>regression coefficient vs. partial regression coefficient</td>
<td>normality</td>
</tr>
<tr>
<td>the multiple correlation coefficient</td>
<td>linearity</td>
</tr>
<tr>
<td>the coefficient of determination</td>
<td>homoscedasticity</td>
</tr>
<tr>
<td>R-square</td>
<td>partial plots</td>
</tr>
<tr>
<td>adjusted R-square</td>
<td>the construction and use of dummy variables</td>
</tr>
<tr>
<td>unstandardized partial regression coefficients</td>
<td>interaction terms</td>
</tr>
<tr>
<td>standardized partial regression coefficients</td>
<td>analysis of interaction</td>
</tr>
<tr>
<td>beta weights</td>
<td>data transformation</td>
</tr>
<tr>
<td>t values</td>
<td>quadratic regression</td>
</tr>
<tr>
<td>hierarchical multiple regression</td>
<td>panel regression</td>
</tr>
<tr>
<td>stepwise multiple regression</td>
<td>path analysis</td>
</tr>
<tr>
<td>forward inclusion</td>
<td>path coefficients</td>
</tr>
<tr>
<td>backward elimination</td>
<td>direct effects</td>
</tr>
<tr>
<td>part correlation</td>
<td>indirect effects</td>
</tr>
<tr>
<td>partial correlation</td>
<td>spurious effects</td>
</tr>
<tr>
<td>collinearity</td>
<td>causal analysis</td>
</tr>
<tr>
<td>multicollinearity</td>
<td>path diagrams</td>
</tr>
</tbody>
</table>

You should be able to interpret an SPSS syntax file for a multiple regression run and you should be able to write a SPSS program (syntax file) to do multiple regression. When doing this here and all other assignments, please turn in the appropriate syntax file corresponding to any and all SPSS output tables presented. Extract the tables that you will be discussing from the SPSS output using “copy objects” and paste them into a word document at the point you discuss them. Just before each set of tables paste in the syntax commands used to generate those tables.

READING:

Textbooks

Polit, 1996.
Chapter 9: Correlation and Simple Regression. Review of material covered in Sc702.
Chapter 10: Multiple Regression (a good basic introduction.)
Chapter 14: Causal Modeling: Path Analysis and Linear Structural Relations Analysis. A basic introduction to some advanced topics; treatment toward the end of the chapter may be too brief to be of much use. (O’Neill)

*Norusis, 2006.
Chapter 12: Bivariate Liner Regression
Chapter 13: Multiple Linear Regression
This text provides a good statistical overview of regression analysis and what you get when you run it in SPSS, but it does not say anything about the syntax needed to set up your run. If you do not have a copy of version 16.0, an earlier version will also work (O’Neill & GSSW [Graduate School of Social Work Library]).

Chapter 9: Bivariate Regression
Chapter 10: Adding a Third Variable: Preliminary Exploratory Analysis
Chapter 11: Multiple Regression with Two Predictor Variables
Chapter 12: Dummy Predictor Variables and Interaction Terms in Multiple Regression
Chapter 14: Multiple Regression with More than Two Predictors

Chapter 3: Data Screening,
Chapter 7: Multiple Regression, and
**Chapter 8: Path Analysis.** (O'Neill)

**SPSS 11.0 Syntax Reference Guide. 2001.**
Chapter 12: This (or a more recent edition of this) manual can be used for information about the syntax you will need to do multivariate regression in SPSS. (O'Neill)

Chapter 3: Multiple Regression (Good basic treatment of topic)
Chapter 5: Nominal Independent Variables (an advanced version of dummy variable analysis)
Chapter 7: Nonadditive Relationships (good treatment of interaction)
Chapter 8: Causal Analysis I (a good discussion of path analysis). (O'Neill)

**Tabachnick & Fidell. 2007.**
Chapter 4: Cleaning Up Your Act: Screening Data Prior to Analysis (good discussion of various issues related to getting the data ready for statistical analysis)
Chapter 5: Multiple Regression (a place to go for a more advanced and more detailed treatment of the topic than you will find in the Polit book)
If the library does not have the 5th edition ask for the 4th edition (2001) keeping in mind that the chapter numbers may change. This book provides a more advanced treatment of multiple regression than you will find in Polit or Warner. You will need to know (or learn) some matrix algebra to fully comprehend some parts of these chapters. For a brief introduction to matrix algebra see Appendix A of this book. (O'Neill)

**Namboodiri, 1984.**
See this book for a more detailed overview of matrix algebra. (GSSW)

**Spicer, 2005.**
Chapter 4: (O'Neill)

**Keith, 2006.**
You might find this book very useful for this course and for future work using multiple regression, path analysis, and related procedures. (O'Neill)

**Articles**


Pampel, Fred C., John B. Williamson, and Robin Stryker. 1990. "Class Context and Pension Response to Demographic Structure in Advanced Industrial Democracies." *Social Problems* 37:535-550. (This is an example of multiple regression analysis checking for and reporting interaction effects.) (O'Neill) (MyFiles)


Shen, Ce. and John B. Williamson. 1999. "Maternal Mortality, Women's Status, and Economic Dependency in Less Developed Countries: A Cross-National Analysis." *Social Science & Medicine* 49:197-214. (This is an article illustrating the use of multiple regression analysis and path analysis.) (O'Neill) (MyFiles)
LOGISTIC REGRESSION

OBJECTIVES: To be able to do logistic regression using SPSS. After we finish this section of the course you should understand such concepts as:

<table>
<thead>
<tr>
<th>Logistic Regression Curve</th>
<th>Partial Correlation (the special version used in logistic regression)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit Coefficients</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Standard Error of the Logit Coefficient</td>
<td>Assessing Goodness of Fit of the Model.</td>
</tr>
<tr>
<td>Wald Statistic</td>
<td></td>
</tr>
</tbody>
</table>

READING:

Chapter 21, Binary Logistic Regression

**Polit, 1996.
Chapter 13, pp. 389-412.
This is a simple introduction to Logistic Regression. (O'Neill)

*Norusis, 2006.
Chapter 15: Logistic Regression Analysis
This is a good statistical overview of logistic regression and what you get when you run it in SPSS, but it does not say anything about syntax describing how to set up your run. If they do not have version 15.0, ask for version 14.0 (2005). (O'Neill) (GSSW).

Mertler and Vannatta, 2005.
Chapter 1 (O'Neill)

Tabachnick & Fidell. 2007.
Chapter 10 Logistic Regression.
This is a more advanced treatment of logistic regression than you will find in Polit or Warner.

Spicer, John. 2005
Chapter 5. An overview of logistic regression at a relatively elementary level. (O'Neill)

This gives the details of SPSS syntax for this. (O'Neill)

Articles:


Gay, David, and John Lynxwiler. 1999. "The Impact of Religiosity on Race Variation in Abortion Attitudes." Sociological Spectrum 19:359-377. (This is a simple example of logistic regression.) (O'Neill) (MyFiles)

DISCRIMINANT ANALYSIS

OBJECTIVES: To be able to do discriminant analysis (also called discriminant function analysis) using SPSS. You should understand such terms as:

- discriminant function
- discriminant score
- group centroids
- eigenvalue
- Wilks’ lambda
- canonical correlation

<table>
<thead>
<tr>
<th>discriminant function</th>
<th>unstandardized canonical discriminant function coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>discriminant score</td>
<td>standardized canonical discriminant function coefficients</td>
</tr>
<tr>
<td>group centroids</td>
<td>structure matrix</td>
</tr>
</tbody>
</table>

READING

Textbooks

  Chapter 16: Discriminant Analysis

Polit, 1996.  
  Chapter 14: This is a simple introduction to discriminant analysis. (O’Neill)

*Norusis, 2006.  
  Chapter 14: Discriminant Analysis  
  This is a good statistical overview of discriminant analysis and what you get when you run it in SPSS, but it does not say anything about how to set up your run. (O’Neill) (GSSW).

  Chapter 10  
  (O’Neill)

Spicer, 2005.  
  Chapter 5: This provides a good overview of discriminant analysis, but at a relatively elementary level (O’Neill).

SPSS 11.0 Syntax Reference Guide. 2001  
  This gives the details of SPSS syntax for this procedure. (O’Neill)

Tabachnick & Fidell. 2007.  
  Chapter 11: This is a more advanced treatment of logistic regression than you will find in Polit or Warner. (O’Neill)

Articles


FACTOR ANALYSIS
OBJECTIVES: To be able to do factor analysis using SPSS. By the time we finish this topic you should understand such terms as:

<table>
<thead>
<tr>
<th>common factors</th>
<th>oblique rotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>communality</td>
<td>factor scores</td>
</tr>
<tr>
<td>eigenvalue</td>
<td>factor score coefficient matrix</td>
</tr>
<tr>
<td>scree plot</td>
<td>principal components analysis</td>
</tr>
<tr>
<td>factor loadings</td>
<td>principal axis factoring</td>
</tr>
<tr>
<td>factor pattern matrix</td>
<td>plots of factor loadings</td>
</tr>
<tr>
<td>factor structure matrix</td>
<td>reproduced correlation matrix</td>
</tr>
<tr>
<td>factor rotation</td>
<td>anti-image correlation matrix</td>
</tr>
<tr>
<td>orthogonal rotation</td>
<td></td>
</tr>
</tbody>
</table>

READING

Textbooks

  Chapter 18: Principal Components and Factor Analysis

Polit, 1996.
Chapter 12: Factor Analysis. This source provides a simple overview. (O'Neill)

*Norusis, 2006.
Chapter 17: Factor Analysis. This is a good statistical overview of factor analysis and what you get when you run it in SPSS, but it does not say anything about how to set up your run. If they do not have version 15.0, ask for version 14.0 (2001). (O'Neill), (GSSW).

Chapter 9: This book provides an alternative to the Polit text written at about the same level. (O'Neill)

Tabachnick & Fidell. 2007.
Chapter 13: This is a more advanced treatment of factor analysis than you will find in Polit or Warner. (O'Neill)

Spicer, 2005.
Chapter 7: This provides a good overview and even discusses confirmatory factor analysis (O'Neill).

Articles


1. You are free to submit your original assignments without making changes. Alternatively, you can make some revisions if you so elect. It is not a problem if there are some comments from the TA on the version that you submit to be graded by me.

2. It is very important that you include the SPSS tables (that the TA and I will need to check what you have done). Do not include tables that you do not mention in your report. It works best to use the “copy objects” function to insert tables into the body of your write-up the segment at the appropriate place. It is important to make it very easy for us to find the relevant material in the tables presented. Highlighting specific numbers in a table that you discuss in the narrative can help a lot.

3. I will give you a suggested word limit for each assignment. This is my suggested number of words for what you will write for the specified assignment. In the past I have made it a mandatory upper word limit, but I am not going to do that this year. The reason for the change is that many students were spending a lot of time editing down their much longer first draft to get to the word limit that I had specified. I am now providing a suggest word limit to give you a ball park to aim for. If you go over or under that word range, that is your decision, but there will be no added credit for long winded assignments. I will be looking for evidence as to how well you understand the material and how much of the material you seem to understand well. A person who does that well in 2,500 words does not get more credit than a person who communicates the same thing in 7,500 words.

4. The most important time during the course to avoid getting behind is toward the end, specifically when doing assignment 6. That assignment often takes longer the others. By that time in the semester the course will seem to be moving very quickly and most students will experience increased time demands from other courses. The solution: be sure to keep up, particularly toward the end of the course.

5. Does it make sense to try to show that you have learned some tricks that we did not cover in class? If you have the time to do some of the optional supplementary reading and if you understand what you have read and if you can find a way to do some computer runs using this new material, then be sure to find a way to show this in your assignments. If you elect to learn optional advanced material, it is up to you to learn this new material on your own (do not expect the TA to teach it to you). If you do so and understand what you have read, then there will be some extra credit. I suggest that you find a way to work in some additional runs using some tricks (options, etc.) that we did not cover in class, options that are not covered in the runs we have (or will have) gone over in class. Be sure to include a discussion of the runs to illustrate that you understand how to interpret what you have found. There is no extra credit for just doing the run that does something new. Also there is no extra credit for doing something early, that is, doing it before we get to it in class. If you are not sure whether or not we are going to cover the topic in class later in the course, please ask. When you are doing something in the hope of getting some extra credit for your efforts, be sure to put that part of your discussion in bold so that the TA and I give special attention to that part of your assignment, mid-term paper, or final paper. What about the issue of space? I would assume that any extra credit material would be over an above the suggested word limit that I mention in connection with a specific assignment.

6. Why add extra credit material? (1) intellectual curiosity, (2) you will learn more from the course, (3) it will help compensate for credit lost due to errors that you do make or for the omission or inadequate coverage of topics that I feel are important. That said, there will be some who get an A on an assignment without doing any extra credit work so long as they do excellent work including thorough coverage with only a few very minor mistakes on the material that we cover in detail in class.

7. Is there a penalty for turning in a late paper? Except for very unusual circumstances (hospitalization, etc.) THERE IS A PENALTY when a paper comes in late. Think of it this way, if you have more time to work on the paper, to be fair to others, my expectations for the quality of work that I expect for a specified grade will move up. If the paper is a couple of days late, the penalty is modest; if is a few weeks late, it becomes substantial. In this course it is a very bad decision to turn in your paper late, not so much due to the grade reduction as the huge increase in the time it takes you to do it due to the forgetting curve.