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. After you have taken this course you should know when to use each technique, how to set up SPSS runs using the technique, how to interpret what you see in computer printouts using the technique, and how to summarize what you have found in a written report. I also want you all to be able to critically evaluate books and 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 you to do it by the specified due date. In some cases, I will ask you to read and comment on an article, a book chapter, or a table in one or the other 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 the same day as the mid-term and the final papers respectively). If after doing an assignment you have doubts about what you did or how you wrote it up, come to see either me or the TA, and we will be happy to review your assignments with you. 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 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 as it will save you a lot of time and be easier for me to understand your analysis. 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. If you have a lot of extra time on your hands to devote to this course, that might be a good idea. In a folder called DATA (see MyFiles) 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.doc). Once you are registered for the course, all related files can be found by going to https://wfs.bc.edu:443/jbw/SC703.

GRADING

Grading will be based my assessment of how much of the material that we have covered you seem to understand, how well you understand it, and how well you communicate that you understand it based on your mid-term and your final papers. This will include how much of what I cover in class you manage to cover in your statistical runs, how thoroughly, accurately, and clearly you write up the results of your statistical computer runs. It will also depend on how impressed I am by your comments on the articles that I assign for you to read and to critically assess. I also urge those you who have the time and the ability to go beyond the basic material that we cover in class to do so. I am not suggesting that you get ahead of the class and cover material in one assignment that the rest of the class will be covering in connection with a subsequent assignment. If you think that might be the case, ask me or the TA. You will be encouraged to find ways to communicate that you have learned interesting things from the reading you have done for the course that we did not get to in class. It will also include exploring some options in connection with this or that statistical technique that we do not have time to cover in class. However, when you do that extra credit work, it must be work that you can do yourself. I do not want to add to the burden on the TA the job of teaching you material that moves beyond the basic material that we cover in class. It is important that you learn to figure things out for yourself because before long you will find yourself in that situation often. I do not grade you on how well informed you seem to be in class. I want you to feel free to ask for clarification at any time or to ask a “dumb” question at any time without feeling that there will be a penalty for showing your ignorance. If you do not understand something that I seem to think is important, please ask, and the sooner you ask the better.

You should allocate the space in your reports so that you at least touch on all of the important concepts and issues that come up in class. As I need to be able to check your work, grading will 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. It is entirely unacceptable to “borrow” another student’s assignments from this year or a prior year as a “guide” to your analysis. You can ask a person to help you get your SPSS run to work, to help you understand material in the runs that I distribute in class, and how to write up results found in the outputs discussed in class. I encourage you to work with friends in the class on those things and to help one another on those tasks. But you should not be reading preliminary or final drafts of the answers from other students who have taken or are taking the. Similarly, you should not be asking other students to read and make suggestions on your early drafts to answers to questions in the various assignments that you will be dealing with this semester. 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 or to explain something in one of the textbooks or any of the printouts discussed in class. If you suspect that what you are doing might be stepping over the line, come to see me and we can discuss whether or not that is the case. If you are not fluent in English, you have permission (and encouragement) to get help with your English 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. I take this issue very seriously.

MID-TERM PAPER (Due Thursday February 28 @3:00 pm). This paper will consist of Assignments 1, 2, and 3. It will make up 40% of your grade for the course. I will need both a hard copy and an electronic copy. What you write must be in 12 point font, single spaced, and the body of your paper must not run more than 60 pagers including the imbedded tables. Before each question based on runs that you will submit, please insert the syntax used to generate the tables used to address that question. Put any and all extra credit work in bold so I will slow down to make take note of it and to assess it importance and accuracy with more than the usual amount of care. You do not need to factor in your extra credit work when it comes to the 60-page limit on your mid-term paper. If your goal is to get an A in the course, it is probably a good idea to do some extra credit work. This will help compensate for errors that even the best students sometimes make. If your goal is an A- or a B+, skip the extra credit effort and focus on the required material, that is, material at least briefly mentioned in class. You are encouraged to eliminate all tables from your report that you do not at least briefly mention. You can shrink tables a bit, but not to the point that your readers (the TA and me) are going to find it a bit difficult to read. The 60-page upper limit on the midterm paper is an upper limit, not a lower. Some strong students can make it very clear that they understand all of the important material that we cover in less than 60 pages. The penalty for a late paper will be minimal until midnight of February 28th (so long as both the electronic and hard copies are in by then). After that you will get two grades, one will be for what it would have been if it had been in on time and another grade that will reflect the penalty. That second grade is what will be average with your grade on the final paper to determine your overall grade for the course. The penalty will be only one grade level so long as your late paper is in by March 6th (midnight).

FINAL PAPER (Due Thursday May 9 @ 3:00 pm). This final paper will consist of Assignments 4, 5, and 6. It will make up 60% of your grade for the course. I will need both a hard copy and an electronic copy. What you write must be in 12 point font, single spaced, and the body of your paper must not run more than 60 pages including the imbedded tables. Before each question insert the syntax used to generate the tables used to address that question. Any (optional) extra credit work that you do must be typed in bold. You can shrink tables, but not to the point that your reader is going to find it hard to read. The 60-page limit is an upper limit, not a lower limit and it does not include spaced allocated to extra credit work. The penalty for a late paper will be minimal until midnight of May 9th (so long as both the electronic and hard copies are in by then) and after that it is just one grade level, so long as the late paper is in by midnight May 14th.

COURSE TEXTS

REQUIRED

1. Polit, Denise F. 2010. Statistics and Data Analysis for Nursing Research (2nd Edition). Upper Saddler River, NJ: Pearson Education. (More Basic) The first edition was published in 1996. The basic statistics have not changed that much and I am sure it will be a lot less expensive, but I would go with the current edition based on SPSS 16.0 (rather than a much earlier version of SPSS) if you can afford it.

2. Norusis, Marija J. 2012. IBM SPSS Statistics 19.0 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. However, the 2009 edition refers to the statistical package as PASW. This does not mean it differs in a major way from the earlier editions. For the 19th edition of this book the term PASW has been replaced by IBM SPSS. I do not think you will run into problems if you are working with and edition of this manual based on the 16th or later edition.

SUGGESTED FOR AN EXCELLENT, BUT A MORE ADVANCED TREATMENT THAN THE POLIT BOOK:
SUGGESTED FOR AN EXCELLENT, BUT A BIT MORE ADVANCED TREATMENT THAN THE WARNER BOOK:
Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*. Fifth Edition. Boston: Pearson/Allyn 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 ALTERNATIVE FOR A RELATIVELY BASIC TREATMENT

SUGGESTED FOR SPECIFIC TOPICS
Keith, Timothy K. 2006. *Multiple Regression and Beyond*. Upper Saddler River, NJ: Pearson Education. 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)

Namboodiri, Krishnan. 1984. *Matrix Algebra: An Introduction*. Sage. (GSSW). This is a good place to go if you want to learn the basics of matrix algebra.

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, “how to” documents, and datasets can be downloaded from MyFiles by going to [https://wfs.bc.edu/jbw/SC703](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 articles 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. **I recommend that you start with SPSS 19.0 Statistical Procedures Companion.**

Bottom Line: 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. 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 in one of the two required books, see me or the TA for the course for help in figuring out where to find relevant material to read. On a few highly specialized topics you will need to get it from the class notes alone or from an item on reserve at O’Neill. Past experience suggests that 80% of the students in the class will get 90% of the basic stats they need from the two required books for the course. If you want more depth try Tabachnick, Barbara G., & Linda S. Fidell. 2007. *Using Multivariate Statistics*. Fifth Edition. Boston: Pearson/Allyn and Bacon. If you think you might want to go on to courses like Sc704, Sc705, Sc706 I would suggest that you spend at least some time with one of the more advanced sources such as Tabachnick & Fidell.
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:

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

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 in 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 prior to each set of tables paste in the syntax commands used to generate those tables.

READING:

Textbooks

*Polit, 2010:
  Chapter 4: Preparing Your Data
  Chapter 5: Transforming Your Data
  Chapter 9: Correlation and Simple Regression.
  Chapter 10: Multiple Regression (a good basic introduction.)
  Chapter 14: Missing Values (Good discussion of alternatives for dealing with missing values)

*Norusis, 2012.
  Chapter 4: Preparing Your Data (This is very basic. Review this when we discuss data cleaning)
  Chapter 11: Correlation
  Chapter 12: Bivariate Linear 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.

Warner, 2008 (O’Neill)
  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

  This gives the details of SPSS syntax for this procedure. (O’Neill, stacks; GSSW, stacks)
*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. Do not be scared by your lack of ability to fully understand the sections that make use of matrix algebra. You will not need to understand the matrix algebra to understand most of what is discussed in this chapter. I do recommend that at some point you take the time to learn the basics of matrix algebra. I suggest you use the Namboodiri (1984) book below, but there are many other books that cover the topic just as well. For a very brief (I think too brief) introduction to matrix algebra see Appendix A of this book. (O'Neill)

See this book for a more detailed overview of matrix algebra. (GSSW)

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)

**SPSS 11.0 Syntax Reference Guide. 2001.**
This gives the details of SPSS syntax for this procedure. (O'Neill, stacks; GSSW, stacks)

**Articles**


Pampel, Fred C., John B. Williamson, and Robin Stryker. 1999. "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)

**Book**
"Wilkinson, Richard and Kate Pickett. 2010. *The Spirit Level: Why Greater Equality Makes Societies Stronger.* New York: Bloomsbury Press (O'Neill). You will be asked to make use of this book some assignments. You can probably get a copy at a good price on Amazon.com. I have a few copies that I can loan, but I probably will not have enough for everyone. So my plan is to loan one copy to two students to share. It is up to you to pick the person you will share with."
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>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logit Coefficients</td>
<td>Assessing Goodness of Fit of the Model</td>
</tr>
<tr>
<td>Standard Error of the Logit Coefficient</td>
<td>Wald Statistic</td>
</tr>
</tbody>
</table>

Reading:

Chapter 21, Binary Logistic Regression (O’Neill)

*Polit, 2010.
Chapter 13, pp. 389-412.
This is a simple introduction to Logistic Regression. (O’Neill)

*Norusis, 2012.
Chapter 16: 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

Mertler and Vannatta, 2005.
Chapter 11 (O’Neill)

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

This gives the details of SPSS syntax for this procedure. (O’Neill, stacks; GSSW, stacks)

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:

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

READING

Textbooks

Chapter 16: Discriminant Analysis (O’Neill)

*Polit, 2010.  
Chapter 11, pp. 293-298: This is a simple introduction to discriminant analysis.

*Norusis, 2012.  
Chapter 15: 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.

Chapter 10  
(O’Neill)

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

Tabachnick & Fidell. 2007.  
Chapter 11: This is a more advanced treatment of discriminant analysis 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>
</tbody>
</table>

READING

Textbooks

Chapter 18: Principal Components and Factor Analysis (O’Neill)

*Polit, 2010.  
Chapter 13: Factor Analysis. This source provides a simple overview.

*Norusis, 2012. 
Chapter 18: 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.

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)

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

Articles


SUGGESTIONS FOR THE MID-TERM AND FINAL PAPERS (AS WELL AS ASSIGNMENTS 1-6)

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. You are not obligated to show me the TA comments on your assignments.

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 page limit for each assignment. So that I can read your work, I am going to require that in all assignments your writing be in a 12-point font, and the words and numbers in the tables must be very easy to read by me in hard copy (and I find fine print hard to read!).

4. The most important time during the course to avoid getting behind is toward the end, specifically when doing assignments 5 and 6. Those assignments often take longer than 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 material in bold.

6. Why do 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, and (4) as an insurance policy (particularly for those who received an A on the mid-term paper). That said, there is typically at least one person in the class who manages to get an A on either the mid-term or the final paper without having done any extra credit work. This can happen if your presentation of the basic material covered in class is thorough and nearly flawless.

7. Is there a penalty for turning in a late mid-term for final paper? Except for very unusual circumstances (hospitalization, etc.) THERE IS A PENALTY when a paper comes in late. Think of it this way: (1) I do not want to create an incentive for turning in a paper late in a effort to get a better grade and (2) I want to protect students from the serious completion problems that in the past have come up when I have allowed people to turn in their assignments late. If you are considering turning in your assignment late, I urge you to first review the discussion of grading earlier in this syllabus. Due to the technical nature of what we cover in this course many students start to forget material that was once very clear. Within a few weeks after the end of the semester it can become very hard to finish assignments and after a few months, it becomes very hard. The good news is that no student has ever turned in a final paper more than 10 years after taking the course.

8. If you find yourself in a situation in which it becomes obvious that you are not going to be able to finish your mid-term or for final paper on time, my advice is to focus on the most important material that we have covered and leave out some of the more difficult material. If you do not leave out too much of that material, you are very likely to pass albeit with a lower grade.