DEPARTMENT OF POLITICAL SCIENCE
POSC/UAPP 816
STATISTICAL METHODS FOR SOCIAL AND POLICY SCIENCES

Fall 1998
H. T. Reynolds

PLEASE READ THIS MATERIAL CAREFULLY
IT EXPLAINS THE COURSE PHILOSOPHY
AND
REQUIREMENTS

PURPOSE:

Here's problem Americans have been struggling with for quite awhile: how do we help people escape poverty? One school of thought, led by Charles Murray, claims that many of the programs the government has created to help the poor actually encourage them to go and stay on welfare. State and national policies, for example, provide such generous benefits that many people prefer to remain on the public dole rather than look for a job or accept job training. Other programs encourage teenagers to raise children out-of-wedlock, shirk the responsiblity of parenthood, or leave school early. Proponents of this view claim that only by getting tough with welfare recipients, by making them stay in school or take jobs (however menial) will we begin to solve the problem.

Their critics reply that such policies are both too harsh and wrongheaded. Too harsh because they don't address the real causes of poverty and misguided because they will cost more than we already spend without reducing the number of poor.

Who is right? How can we decide whether welfare programs help or hurt the poor? What sort of methods should we use? Indeed, one wonders if we can answer these questions at all.

What is needed is an objective way to assess this assertion. Since many of the claims and counter-claims rest on supposedly hard numerical data, it's only natural for policy analysts to turn to statistics. Statistical methods cannot provide a definitive test, but if carefully applied they might at least shed light on it.

Using statistics to solve scholarly and policy problems is the main purpose of POSC/UAPP 816. As a follow up to a first course in applied statistics, the class provides a framework for using data to address substantive debates. Briefly stated, the class has two goals: first, to give every one an understanding of the statistical tools commonly used in the behavioral and policy sciences; and second, instill in each participant the motivation, confidence, and skills necessary to confront the technical controversies that dominate the political and social agenda. Although the course emphasizes "applications" rather than theory, methods rather than theorems, we will consider in detail the assumptions that underlie various procedures so that everyone will understand when statistics are informative rather than misleading.

With these considerations in mind, the specific topics include among others:

- methods for translating verbal theories into testable hypotheses;
- a review of the principles of statistical inference;
- graphical techniques for both presenting and analyzing data;
- a thorough discussion of the model that underlies the analysis of variance, regression, and similar methods;
- a fairly through study of the assumptions underlying statistical procedures and the substantive and policy consequences of analyses that violate them;
- an exploration of the difference between correlation and causation;
- tools for building statistical models of various social, economic, and political

processes;
- an understanding of time series and cross-classified data;
- the use of MINITAB and SPSS, general-purpose statistical computer programs;
- an evaluation of published empirical research;
- ways to "simulate" data on a computer in order to better comprehend the behavior of various statistical methods.

PRIOR EXPERIENCE:

The course has been “designed” for students who do not have a strong aptitude or interest in statistics but who need the experience to meet graduation requirements or who come across quantitative research in their employment or both. I do assume that everyone has had the equivalent of a one semester course in applied statistics. Experience working with, say, a spreadsheet program like "Lotus" is not sufficient; you need at least a rudimentary understanding of descriptive statistics and statistical inference. The key to success, however, is a good attitude and willingness to try and try again to learn a few basic concepts.

In this class we use a computer program, MINITAB, to carry out most of the calculations. I will provide most of instruction, but you can pick up a lot on your own, even if you’ve never used a desktop or personal computer before. Doing so won’t be difficult because you have access to several resources:

- I have prepared several handouts that explain the personal computers, Windows (the so-called operating system) and MINITAB.
- POSC/UAPP 815 classes were video taped last semester. These tapes, especially those of the first two or three meetings, explain Windows and the desktop version of MINITAB. The tapes may still be available in the reserve room of the library.
- Assignment 1 and perhaps 2 provide hands-on explanation and practice.
- The Computing Center offers several brief (usually two hour) introductory classes on Windows and other aspects of personal computing. Although these classes do not explain MINITAB, they still show you other useful techniques such as how to print results.

REQUIREMENTS:

Like a foreign language, statistics understanding comes only from active, not passive, learning; from practicing and using methods on "real" data. Consequently, everyone will be doing statistics constantly.

In order to encourage this activity, the course requirements consist of frequent assignments. These problems, which will be handed out virtually every week, encourage you to practice and apply the techniques discussed in class. They will usually require you to use MINITAB.

You will be told more precisely when these assignments are due. In every case you will

have sufficient time to complete them.

Here is the grading system I usually use:

* ✓ = satisfactory
* ☹ = acceptable but you should look at my comments to make sure that you fully understand an idea.
* ✗ = something has been done or expressed incorrectly. Correct the indicated errors and resubmit. Again the point is to make certain that you understand the material.

For some assignments I may substitute numerical grades for the check system. This will be true especially for the last assignment that asks you to apply the methods and concepts discussed during the semester to a relatively large body of data. Your grade will be calculated on the basis of the sum of the points earned on the assignments.

If you receive an X on an assignment that is just checked, you should make every effort to understand what went wrong and to hand in a correction. In other words, I'll ask you to do the problems until I'm sure you understand the concepts and can arrive at the "correct" results.

CALCULATOR:

If possible you should purchase a "statistical" calculator, one that accumulates sums of squares and cross products and that converts automatically to "scientific notation" (so you won't have to keep track of large numbers). If you read ***MINITAB for Windows***, you will discover that MINITAB can be used as a calculator.

OFFICE HOURS AND GETTING HELP:

In order to expedite the answering of questions, I rely as heavily as possible on electronic mail. In many cases I will be able to give you a prompt response so you may not have to wait until my next scheduled office hour to get an answer.

I will also hold regular office hours in Smith 434 on Monday and Thursday in the afternoon usually from about 3:00 to 4:00 and Wednesday after class for a short time. But of course I am more than willing to meet at other times of mutual convenience.

If you are having trouble understanding a topic or doing an assignment, please let me know immediately. (But don't ask your friends, even to check answers.) When seeking assistance, though, try to do these things:

- Bring or send (via E-mail if possible) a computer print out (if applicable). Attempting to reconstruct what you have done from memory is usually futile; you

almost always have to have a "hard copy" of your work for me to figure out your mistake. (I will explain how you can send copies of your work session via e-mail.)

- READ DIRECTIONS CAREFULLY. More than half of the questions I answer deal with someone's failure to read an assignment carefully.
- Make an initial effort to do the problem. Write down what you do and don't know. BE NEAT!
- Don't use the assignment sheets for notes or scratch paper. Keep them as neat as possible.

WHAT YOU NEED:

Since you are take at least a second semester of applied statistics it is worth starting your own reference library and acquiring some statistical software. For that reason I urge you to obtain the following:

* Class notes and data. (This is easy: they will be handed out in class or are available on the course web page.)
* McKenzie, Schaefer, and Farber, ***The Student Edition of MINITAB for Windows*** for use at home or office computer.
* Agresti and Finlay, ***Statistical Methods for the Social Sciences*** 3rd Edition.
* (As many as you can afford.) Sage Publications ***Quantitative Applications in the Social Sciences***: Jacoby, "Statistical Graphics for Univariate and Bivariate Data" (No. 117); Henkel, "Tests of Significance" (No. 4); Schroeder, Sjoquist, and Stephan, "Understanding Regression Analysis" (No. 57); Lewis-Beck, ***Applied Regression***; Berry and Feldman, ***Multiple Regression in Practice***; Achen, ***Interpreting and Using Regression***.
* A statistical calculator
* A subscription to the ***New York Times***, available at the Newark Newsstand. Even if you are a "distance" learner you should subscribe.

It seems to me that investment in a first-rate text such as ***Statistical Methods for the Social Sciences*** makes sense even for those people who do not intended to "do" statistics during their careers. After all, for the rest or your lives most of you will be discussing, if not formally evaluating, arguments that rest on at least some statistical methods. Having a good reference book ought to help greatly.

TOPICS

Note: It may take two or more classes to cover a topic, and you will always be told which readings are applicable.

REVIEW OF STATISTICAL INFERENCE

Topic I — Introduction: MINITAB, SPSS, Desktop Computing.
We'll explore some tricks for using programs, moving data from the internet to the desktop, and similar topics.

Topic II — Review of Statistical Inference.
We'll cover basic concepts such as populations, samples, parameters, statistics, sampling distributions, standard errors, critical regions, observed and critical test statistics, and types of errors. Small Sample Tests of Means

Topic III — Difference of Means Tests and Confidence Intervals.
Now we'll continue reviewing statistical inference methods and estimation by looking at more hypothesis tests, confidence intervals, power, (very briefly) sample sizes, and the like.

Topic IV — Continency Tables
As a transition to linear models we'll consider some simple types of association and (perhaps) underlying models appropriate for categorical data. In particular, we'll discuss the chi square test of significance and a couple of measures of association for two variable contingency tables. Then we'll take a quick look at multi-dimensional tables and talk about partial associations and interaction.

TWO VARIABLE REGRESSION

Topic V — Simple Regression
Two variable regression provides a very simple form of the general linear model. In this case we have a quantitative dependent variable and a single explanatory factor or predictor or independent variable. We describe a model that shows a linear relation between the two and explain its various parts. We'll also define the correlation coefficient and describe its uses and misuses.

Topic VI — Statistical Inference for Regression
This section covers tests of significance and confidence intervals for regression parameters.

Topic VII — Regression Diagnostics
Frequently a set of data do not "fit" a particular model very well or the data do not meet important assumptions. We'll pause briefly to discuss the assumptions that underlie regression and look at a) methods for detecting "violations" of these assumptions and b) procedures for adjusting or transforming the data or model or both so that the assumptions are met.

.

MULTIPLE REGRESSION

Topic VIII An Introduction to Multivariate Relationships
Building on the material covered in the section on contingency tables we'll discuss association versus causation; statistical controls as an analogue to the classical randomized experiment; and types of structural or causal models

Topic IX The Multiple Regression Model
We add more predictors or independent variables to the linear model and consider the interpretation of partial regression coefficients. Other topics include the multiple correlation coefficient, multicolinearity, and (simultaneous) inference and estimation.

Topic X Multiple Regression in More Detail
There are a number of straightforward extensions to the multiple regression model including the use of so-called dummy variables (to represent qualitative factors), standardization and standardized coefficients, interaction, and variable selection strategies.

Topic XI Analysis of Variance
If time permits we will briefly discuss the relationship between regression and analysis of variance, an extremely important tool used in experimental sciences.

Topic XII Analysis of Covariance
We can build, estimate, test, and interpret models having mixtures of quantitative and qualitative independent variables. Most of these models are simple applications of what has been covered above.

ADDITIONAL REGRESSION TOPICS

Topic XIII Resistant Regression.
The form of regression analysis used most commonly in the social and policy sciences may not be appropriate for all types of data. Hence, we look at alternatives that will always be appropriate: robust regression models.

Topic XIV Time Series and Intervention Analysis
We can study data that have been collected over time such as presidential approval scores or stock market prices or monthly crime statistics. It is also possible to assess in mathematical terms the effects of policy interventions, such as the passage of a law, on trends.

Topic XV Regression on Qualitative Dependent Variables

The topics considered so far assume that the dependent variable--the factor being "explained"--is a quantitative variable. Some people believe that if a theory does not include quantitative dependent variables, it is probably not sufficiently developed to be amenable to statistical analysis. But most social and policy scientists would not agree with this position so we look (very briefly) at methods for building regression type models for qualitative (sometimes called ordinal or nominal scales) dependent variables. Some of the ideas developed in the contingency table section lend themselves to logistic regression.

Go to Applied Statistics Page