Class 2

What is empirical research?

Main idea in empirical research is that data is used to test a theory.

(Data can be many sources: public opinion data, census data, crime statistics, experimental data)

A scientific method is used to conduct empirical research.

A. Correct way to do empirical research:
1.) generate idea
2.) do background research on what has been done before on this topic and understand what you have to contribute
3.) formulate theory and testable hypotheses
4.) think about ways to empirically test the hypotheses
5.) collect data
6.) conduct statistical tests
7.) present analysis of the tests

B. The way we will do it:
1.) Look at the provided data
2.) generate idea
3.) formulate theory and testable hypotheses
4.) use the data provided to statistically test the hypotheses, etc.

(note that we have skipped 4 and 5 above and added 1)

Note that in empirical research we do not know beforehand if our theory and hypotheses will be proven correct. This is only determined after your empirical investigation is completed.

Unlike non-empirical research where a researcher can make a persuasive (subjective) argument to prove her argument without validating data.

However, although empirical research is “suppose” to be scientific, it is often a “subjective science.”

When conducting the statistical analysis the researcher has to make many decisions that are subjective.

C. Class Materials

u:\msu\course\pls\201

Here you will find:
   today’s class notes
   SPSS instructions
Data files
Codebooks
assignments
etc.

D. Research paper

1st thing we need to do for the paper is to examine the codebook.

The difference between a data file and codebook.
1. A data file is where the answers to a list of questions are located.

2. A codebook is where the questions to the answers are located. It is the survey questions.

3. A variable is just one particular answer to a question or one survey question

Data files are usually divided into two categories:

1.) Demographic information – what are the characteristics of the respondent (how old, gender, ethnic background, place of resident, how much income, education level, etc.)

2.) Personal opinion questions. What do you think about ______________?

Papers of the nature we are writing usually include one personal opinion question and several demographic questions.

Example: Abortion. What are the characteristics of people who are pro-life or pro-choice?

People who are pro-life tend to be male, elderly, and religious.

{Abortion will not be a topic you can choose since I will use this for in class discussion)

Your personal views should not be incorporated in the papers.

Hence, we are interested in what types of people have certain opinions.

Types is a form of generalization. So we will be making generalizations about people. i.e., people from small rural areas tend to like country music. It is not always true, but on average it tends to be true.

We will be using two data files:

1.) 2004 National Election Study (Major study funded by the National Science Foundation and conducted out of the University of Michigan). Panel study that is done every four years. Released after Presidential elections.
We will be using this one for our research papers.

2.) The GSS (2002) is another large survey conducted by the University of Chicago. The original data set covered the years from 1972-2002. It contained over 4000 variables. I have cut this to only include 2002, but it still has over 800 questions. This survey deals with a lot of social issues. There are a lot of missing values for the variables since years have been cut out.

We will be using the GSS for in class assignments only

Let’s look at the codebook for the NES

u:\msu\course\pls201\NES\Codebook

Sometimes we will not use a codebook but rather only the datafile. The GSS codebook is too big to be of any use.

For the paper we will have an:
1.) Introduction
You will need at least five variables but probably six, due to the creation of an index variable, plus an alternative rival variable. (one dependent variable and four to five independent variables, i.e., you will have to select at least five questions from the NES)

What is the difference between a dependent variable and an independent variable?

Dependent variable is the variable of interest, or the variable we are trying to explain.

Example: Attitudes on abortion is our dependent variable and age, gender and religiousness are our independent variables.

This is 2 to 3 pages. You will need to do some research on your topic before you write this.

2.) Theory and testable hypotheses
Theory is the hardest part. You are allowed to use a previously published theory. But just think about the process that is involved.

You will have to generate three hypotheses from your theory and one alternative rival hypothesis.

Alternative rival hypothesis is a different explanation that the one posited in your theory.

You will need at least five references to validate your theory.

This part should be three to four pages.

3.) Empirical analysis.
Crosstabs
Bivariate correlations
Regression analysis

4.) Conclusion
Sums up what you did in parts 1, 2, and 3
One page

E.) Enter SPSS under “Math Apps”

Pull up u:\msu\course\pls\GSS\gss2002.sav

Data view vs. Variable view
Data view is the actual data file
Variable view allows us to work with the variables.
Variable labels is a short description of the variable.

Also “utilities” on the top menu under “variables” gives us a description of the variables

Look under “view” and “value labels”

“Map out one respondent”

Lets look at the values of a single variable:

1.) Under “graphs” and “bar graph” and select “simple”

Let’s select “marital” variable and put it under “category axis.”

Double click on the graph to get a chart editor.

Under elements: Data label mode.
Under transform: Can change type of chart

Let’s look under frequencies

Go to “Analyze” then “Descriptive Statistics” to “Frequencies.”