

Quantitative Methods II

Assignment 12

December 4, 2011

Revised

One of the things I think I should do is introduce you to various sources of available data. This week, I am introducing you to the General Social Survey (GSS), which has some rather stunning facets. The surveys began in 1972, is biennial, and contains 5364 variables. These variables include a core set of questions that are asked each year, plus a variety of special-interest questions. As the unit of analysis is the individual, you can look at correlations between sets of demographic and attitudes. Furthermore, as the attitudes are usually ordinal variables, this dataset fits well with the objectives of this week.

As this dataset contains survey responses, there are quite a few missing values, which correspond to refusals to answer the question. The other drawback to this data is that the variable names are often confusing. For instance, what does the variable `astrosci` indicate? In order to handle this, GSS provides a codebook. I am providing the variable names and meanings in the question.

The entire problem set needs the subset of the GSS data available on the course's website. The data filename is `gss2`.

If you wish, you can download the entire GSS dataset from the GSS homepage.¹ However, it is only available in SPSS and STATA formats. If you have neither of these two programs, you will have to use GSS's NESSTAR program, which is available at the website above.

One important note: You will need to determine how to make sure your program recognizes the correct ordering for the ordinal variables. The only ordinal variable we will examine for this problem set is the `harmgood` variable. In R, you should refer to Forsberg, Chapter 10.

¹The URL is <http://www.norc.org/GSS+Website/>.

When you turn in this assignment, you will email to me the usual two files: your write-up and your R script.

As always, the email needs to include the subject line:

POLS6123: Assignment 11

Also as always, if you have problems, let me know. I am happy helping.

PROBLEM: BETTER DEAD THAN RED? [[5]]

I am curious about what causes a person to think it is acceptable to have a communist speak at a public meeting (`spkcom`). I hypothesize that the variables of interest are a belief in life after death (`afterlif`), frequency in attending religious services (`attend`), and the gender of the respondent (`male`).

Thus, the research model is

$$\text{spkcom} \sim \text{afterlif} + \text{attend} + \text{male}$$

Part A: What type of variable is the dependent variable?

Part B: What type of regression will you need to do?

Part C: Fit the research model and provide the results in a nicely formatted table.

Part D: Let Bob definitely believe in the afterlife (`yes, definitely`), attend church services every week (`every week`), and be a male. What is the probability that Bob would allow a communist to speak? (Ask yourself: Will you need to transform the result?)

PROBLEM: BETTER DEAD THAN MUSLIM? [[5]]

I am also curious if there is a relationship between the same independent variables from the previous problem and the willingness to allow an anti-American, Muslim clergyman to teach in a college (`colmslm`).

Part A: What is the research model?

Part B: What type of regression will you need to do?

Part C: Fit the research model and provide the results in a nicely formatted table.

Part D: Is Bob more likely to allow a communist to speak or allow an anti-American Muslim clergyman to teach at a college? (Make sure you provide the actual predictions.)

Part E: Do the results make sense to you; that is, do they have *face validity*?

PROBLEM: BETTER DEAD THAN A RED MUSLIM? [5]

I am curious if one can get a better prediction regarding a person's willingness to allow an anti-American, Muslim clergyman to teach in a college (`colmslm`) from his or her gender (`male`), income (`rincome`), and presidential vote in 2004 (`pres04`).

Part A: What is the research model?

Part B: Fit the research model and provide the results in a nicely formatted table.

Part C: Does this model seem to fit the data better than the model in the previous problem? How do you know? Make sure you cite the correct numbers from the output.

Part D: What is the probability that Bob is willing to allow an anti-American Muslim clergyman to teach at a college? Note that you will now need to know that Bob voted for Kerry, has an income of between \$10,000 and \$14,999, and is still male.

PROBLEM: BETTER DEAD THAN A RED MUSLIM SCIENTIST? [[5]]

I am curious if I can predict a person's feelings about science (`harmgood`)² from knowing how often he or she prays (`pray`), how much he or she believes in heaven (`heaven`), he or she has taken any college-level science courses (`colsci`), and whether or not he or she favors the death penalty for murder (`cappun`).

Part A: What is the research model?

Part B: What type of regression will you need to do?

Part C: Fit the research model and provide the results in a nicely formatted table.

Part D: What does your model predict about Bob's feelings towards science? What is the probability that Bob strongly agrees that science does more harm than good? For this, let us also add that Bob prays once a day, definitely believes in heaven (yes, definitely), did not take a college science course, and opposes capital punishment.

²The variable concerns the question: Do you agree that science does more harm than good?

PROBLEM: BETTER DEAD THAN A NOMINAL RED MUSLIM SCIENTIST? [5]

Fit the same research model in the previous problem, but use the nominal regression. Answer each of the four parts of of that previous problem.

Part E: How are these estimates different from those of the previous problem? What does this suggest to you?