

Instructions

Due Week 6 · Friday, October 2

Prof. Jack Reilly

Fall 2026

For this assignment, you must turn in two documents:

1. Your answers (as a PDF), and
2. A plain-text .do file containing the Stata code you used to arrive at your answers.

Use the provided subset of the General Social Survey (GSS) to answer these questions. The GSS cumulative file is a large social sciences study conducted since 1972 asking a variety of social questions over that time. To work with this data, you'll need to use some additional Stata commands.

For instance, the year a respondent was interviewed is found in the `YEAR` variable. If you wanted to look at individual years in the dataset, you could do this via dropping or keeping certain years in the data. For instance, if you want to only look at individuals interviewed in the year 1984, type:

```
keep if YEAR==1984
```

Which would keep ONLY individuals interviewed in 1984, and eliminate all other individuals from the dataset. Alternately, if you wanted to drop a certain year (say, 1996), you could drop that year by coding:

```
drop if YEAR==1996
```

The GSS also contains responses to questions that need to be recoded. For instance, if a variable for education contains values of 1 for high school, 2 for college, 3 for post-college, 4 for "did not respond" and 5 for "data missing", you may wish to turn categories 4 and 5 into formal missing data (stored as a "." in Stata). You can do this with the `recode` command. This code will generate a new variable in your dataset with all the 4 and 5 values changed to ".":

```
recode education 4=. 5=., gen(education_recode)
```

Finally, remember that if you want to save your current dataset, you may do so with the `save` command. The following code saves your current dataset to your working directory:

```
save working.dta, replace
```

And the following code saves the most recent graphic to your working directory as a .pdf labeled “graph1.pdf”:

```
graph export graph1.pdf, replace
```

Questions

1. Locate the seven-category party ID variable (PARTYID).
 - Clean it appropriately, creating an ordinal variable called `pid7clean`. (*To clean the variable, use the `tab` command to first view it, and then, recode the variable so that value 7 is coded as “missing”*)
 - With this new variable, calculate the average partisanship in the whole GSS sample and report this average in your word file.
 - Calculate the average partisanship in the GSS sample for the years 1980, 1990, 2000, 2010, and 2021. How did average partisanship change over time?
2. Locate and investigate the variables for income (REALINC) and education (EDUC). If either of these variables requires cleaning, clean the variable in the same fashion you did for party identification in #1. Calculate summary statistics for each variable. What is the mean of income and education?
3. For the whole sample, regress `pid7clean` on income and regress `pid7clean` on education. Copy your regression results into your word file. What do these regressions tell you?
4. Regress `pid7clean` on income and `pid7clean` on education for the following years: 1980, 1990, 2000, 2010, 2021. Copy your regression results into your word file. What do these regressions tell you?