**Extended Notes for Instructors for Exercise STAT16S\_SDA**

This is the third in a series of exercises for an introductory course in statistics in the social sciences.[[1]](#footnote-1) This series uses SDA (Survey Documentation and Analysis) which is an online statistical package written by the Survey Methods Program at UC Berkeley. SDA can be used without cost wherever one has an internet connection. Students can be shown how to use SDA in approximately ten minutes making it unnecessary to spend valuable class time on how to use the statistical package. There is also an extensive help menu available to users of SDA. I have prepared notes on using SDA which can be accessed by clicking [here](http://ssric.org/files/notes_on_using_sda_0.docx).

The data set used in this series of exercises is the General Social Survey’s 2014 Cumulative Data File (1972 to 2014) which is available without cost by clicking [here](http://sda.berkeley.edu/sdaweb/analysis/?dataset=gss14). For this exercise we will only be using the 2014 General Social Survey. The exercises show students how to select the 2014 survey from the cumulative data set. A weight variable is automatically applied to the data set so it better represents the population from which is sample was selected.

The General Social Survey is a large, national probability sample of adults (18 years and older) living in the United States conducted by the National Opinion Research Center (NORC) at the University of Chicago. The GSS started in 1972 and was conducted annually through 1994 and biannually since then. Many of the questions in the GSS have been repeated from previous years providing important trend data. The most recent GSS was 2014. The sample size for the 2014 survey was approximately 2,500.

More information about the GSS can be found on the [NORC - General Social Survey website](http://gss.norc.org/).[[2]](#footnote-2) At the website you will find the documentation for the survey, survey questionnaires, a bibliography, useful Frequently Asked Questions, and more. You can also download the complete GSS in either SPSS or Stata format. You can create a free account on the GSS Data Explorer where you can search the GSS by variable and topic.

In the exercise and the extended notes, variable names appear in italics and SDA commands are in all capitals to make them easily recognizable. You could modify this if you wish.

The goal of this exercise is to explore linear regression with dummy variables. The exercise also gives students practice in using REGRESSION in SDA. This exercise isn’t a comprehensive discussion of dummy variable regression. For example, we don’t discuss how to compute the multiple regression coefficients. We don’t go into all the assumptions underlying multiple linear regression. We don’t go into all the details regarding the use of dummy variables. A good introductory treatment of these assumptions can be found in Colin Lewis-Beck and Michael Lewis-Beck*, Applied Regression – An Introduction* in the Sage Series on Quantitative Applications in the Social Sciences (number 22). In this exercise I assume that students have a basic understanding of tests of significance. You may want to add your own materials to this exercise or delete sections that go into more detail than you want. This exercise builds on the previous two exercises ([STAT14S](http://www.ssric.org/node/610) and [STAT15S](http://www.ssric.org/node/611)) so it is best used along with those exercises.

It’s important that students click on the arrow next to OUTPUT OPTIONS and then click on the circle next to SRS to select it. This will be on the line that says SAMPLE DESIGN. The General Social Survey is not a simple random sample. Rather it is an area probability cluster sample. This means that standard errors for the GSS are larger than what you would get assuming simple random sampling. This, in turn, affects tests for statistical significance. However, complex standard errors are beyond the scope of any introductory statistics course. Therefore, we’re going to proceed assuming simple random sampling even though we know it will produce an underestimate of the standard errors. In the exercise I simply tell students to make this change and do not try to explain it.

You have permission to use this exercise and to revise it to fit your needs. Feel free to revise the exercise in any way you want. Just recognize the source of the original exercise. Please send me a copy of the revised exercise so I can see how others are using it.

If you would like to contact me, please email me at ednelson@csufresno.edu. I’m Professor Emeritus at California State University, Fresno in the Sociology department. I taught research methods, statistics, and critical thinking before retiring and now teach a critical thinking course part time.

1. The first in the series used SPSS as the statistical package for the exercises. The second used PSPP as the statistical package. Both series of exercises are on the Social Science Research and Instructional Council’s [website](http://ssric.org/tr/onlinetextbooks). [↑](#footnote-ref-1)
2. It will ask you to log in when you click on the link. Wait several seconds and click on the X in the upper right and the site will open. [↑](#footnote-ref-2)