**Notes on Using SDA**

Survey Documentation and Analysis (SDA) was developed by the Survey Methods Program at UC Berkeley and is currently maintained by the Institute for Scientific Analysis. It’s an online statistical package that is freely available wherever you have internet access. SDA is easy to use and it doesn’t take very long to get up and running on SDA. It has a very good context-sensitive help menu that is easily accessible.

One of the limitations of SDA is that it’s not easy to develop data sets in SDA format. While SDA is freely available for use, it requires a site license to create your own SDA data sets. Developing SDA data sets is labor intensive and has a fairly steep learning curve. That means that one normally relies on data sets that others have developed. Fortunately, there are many high quality SDA data sets available on the internet. Here are some places to go to find SDA data sets.

* [SDA Archive](http://sda.berkeley.edu/archive.htm) located at UC Berkeley including:
  + General Social Survey cumulative data file from 1972 through 2014
  + American National Election Studies from 1996 through 2012 and the ANES cumulative data file from 1952 through 2012
  + Census microdata
    - U.S. from 1990 through 2008
    - California for 1990 and 2000
* Inter-University Consortium for Political and Social Research (ICPSR) including:
  + many data set located in the ICPSR [general archive](http://www.icpsr.umich.edu)
  + [data-driven learning guides](http://www.icpsr.umich.edu/icpsrweb/instructors/biblio/resources) which are instructional exercises that can be used in the classroom
  + Investigating Community and Social Capital, an [instructional module](http://www.icpsr.umich.edu/icpsrweb/instructors/icsc/index.jsp) using SDA
  + Voting Behavior: the 2012 Election, another [instructional module](http://www.icpsr.umich.edu/icpsrweb/instructors/setups2012/index.jsp) using SDA
* Field Poll data located at [UC Berkeley](http://ucdata.berkeley.edu/data_record.php?recid=3#analyze) from the 1950’s through 2014 (for public users) and through 2016 (for UC and CSU affiliates)

In the series of exercises I developed we’re going to use the 2014 General Social Survey. The GSS cumulative data file is freely available by clicking [here](http://sda.berkeley.edu/sdaweb/analysis/?dataset=gss14). The 2014 file can be extracted by using the SELECTION FILTER(S) box. All you need to do is to enter *year(2014)* in the box and SDA will extract the 2014 file for your analysis.

It’s important to weight the data so the sample better represents the population from which the sample is selected. SDA automatically enters the weight variable in the WEIGHTS box which means that you don’t have to remind students to use it.

The exercises I developed provide students with all they need to know to run SDA. They probably won’t even need to access the help menu.

There are a couple of things that need to be mentioned about SDA. SDA doesn’t have a command to run a t test. The t test is a special case of one-way analysis of variance when the independent variable is a dichotomy. In that case, t is the square root of F. I explain that in the exercise but it will probably need reinforcement in class.

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. SDA will compute standard errors using either complex standard errors for samples like the GSS or simple random sample (SRS) errors when the sample is a simple random sample. However, complex standard errors are beyond the scope of any introductory statistics course. I consulted with a statistician at UC Berkeley and his advice for introductory statistics classes was to assume 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. I suggest you not try to try to explain this in class or that you go over it briefly.

The default when you open the GSS data file is to compute complex standard errors. Therefore, 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. However, they only need to make this change when they plan on using tests of significance. I incorporated this into the exercises but only when they were going to use tests of significance.

You can download a version of any SDA data set by clicking on DOWNLOAD CUSTOM SUBSET in the upper left of the SDA screen. Subsets can be downloaded as a text or csv file along with the data definitions in SPSS, SAS, or STATA format. These can then be combined to create the data file in these formats. Codebooks can also be downloaded.

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

If you would like to contact me, please email me at [ednelson@csufresno.edu](mailto: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.