# Extended Notes for Instructors for Exercise STAT15S\_pspp

The data set used in this exercise is gss14\_subset\_for\_classes\_STATISTICS\_pspp.sav which is a subset of the 2014 General Social Survey (GSS). Some of the variables in the GSS have been recoded to make them easier to use and some new variables have been created. The data may be downloaded at this site. You have permission to use this exercise and to revise it to fit your needs. Please send a copy of any revision to the author.

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.

This exercise is based on a subset of 218 variables from the 866 variables in the 2014 GSS.[[1]](#footnote-1) The data have been weighted according to the instructions from the National Opinion Research Center.

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 by topic and even analyze the data online.

For this exercise we have renamed the variables to make it easier for students to find related variables. For example, the first seven variables in the GSS deal with abortion (a for short). These seven variables start with a1, a2, a3, a4, a5, a6, and a7 which are followed by an underscore (\_). Following the underscore is the name of the variable from the GSS (abany, abdefect, abhlth, abnomore, abpoor, absingle, and abrape). Thus the first variable is named a1\_abany and the second variable is a2\_abdefect. The advantage of this system is that students can easily find categories of variables that have the same focus. At the same time, you also know the name of the variable in the GSS and can easily use the NORC codebook. Looking the variable up in the NORC codebook will give you the exact question wording.

There are 26 categories of variables which are listed below.

1. abortion (a) – 7 variables
2. aged (aged) – 1 variable which asks if the aged should live with their children
3. c (crime) – 3 variables which ask about fear of crime, capital punishment and the courts
4. ci (confidence in institutions) – 13 variables asking about confidence in American institutions such as organized labor, the press, and television
5. demographics (d) – 30 variables including, for example, age, sex, race
6. financial (f) – 11 variables which include income, satisfaction with financial situation and various opinion questions
7. guns (g) – 5 questions which include gun ownership and gun permits
8. helping others (help) – 12 questions focusing on the importance of helping others versus taking care of one’s self
9. happiness (hap) – 4 questions dealing with happiness and satisfaction focusing on marriage, life in general, and job satisfaction
10. health (hlth) – 2 questions including perceived state of respondent’s health and the importance of providing health care for everyone
11. immigration (i) – 8 opinion questions on immigration and immigrants
12. m (marijuana) – 1 questions asking about the legalization of marijuana
13. national spending (nat) – 29 questions asking whether we are spending too little, too much, or about the right amount on various issues such as defense, crime, and the environment. Embedded within these questions is a true experiment in which question wording is randomly assigned to respondents.
14. political (p) – 9 questions including party identification, political views, voting, and other opinion questions.
15. political efficacy (pe) – 4 questions asking whether the respondent thinks that they can influence what the government does and the degree to which they are informed about political issues
16. pornography (porn) – 1 question asking about how the distribution of pornography should be controlled
17. religion (r) – 21 questions including religious preference, religiosity, and opinion on various religious issues
18. sex (s) – 13 questions asking about gender, sexual activity and opinion about homosexuality and sexual behaviors
19. sociability (soc) – 4 questions about how much the respondent interacts with others
20. sex roles (sr) – 5 opinion questions about the roles of men and women
21. suicide (sui) – 5 questions asking about whether people should have the right to take their own life and whether patients with an incurable disease ought to be allowed to die
22. tolerance (t) – 18 questions about whether different groups of people (militarists, atheists, racists, communists, homosexuals, and anti-American Muslim clergy) ought to be allowed to make a public speech, have their books in libraries, and teach in colleges
23. trust, fairness (tf) – 5 questions asking whether people can be trusted and fairness issues
24. television (tv) – 1 questions asking how many hours per day respondents watch tv
25. weight (w) – 3 weight variables. We’re using W1\_WTSS to weight the data so the data more accurately represent the population
26. year (year) – the year (2014) of the survey

The goal of this exercise is to explore multiple linear regression. The exercise also gives you practice in using LINEAR REGRESSION, FREQUENCIES, BIVARIATE CORRELATION, and SELECT CASES. The exercise does not explain how to use these PSPP commands. Rather it gives students practice in using them.

The 15 exercises in this set (STAT1S\_pspp through STAT16S\_pspp with STAT4S omitted) were originally written for SPSS. However, some colleges including many community colleges do not have a site license for SPSS. The cost of SPSS is often prohibitive so I rewrote the exercises to run in PSPP which is free statistical software sponsored by the Free Software Foundation. For more information on PSPP, go to their [website](http://www.gnu.org/software/pspp/). Their website says that “GNU PSPP is a program for statistical analysis of sampled data. It is a Free replacement for the proprietary program SPSS, and appears very similar to it with a few exceptions.”

PSPP will open your SPSS data (.sav) file. You can also run SPSS syntax (.sps) files in PSPP. Sometimes PSPP will not run a particular SPSS command. I have rewritten some of the exercises to make sure that all the syntax files included with the exercises will run. That means I had to eliminate some parts of certain exercises. I also eliminated one exercise (STAT4S) which was heavily dependent on charts and graphs since PSPP does not have much capability in this area. I kept the original numbering of the exercises to make it easier to make comparisons to the SPSS version.

There are two documents that will help you in running PSPP. The first is “Notes on Using PSPP” which tell you how to download PSPP and gives you an introduction to running PSPP. The second is “Differences between PSPP and SSRIC” which alerts you to differences I encountered when rewriting these exercises for PSPP. I have used SPSS for many years and only started working with PSPP in the last several months so I don’t claim to be anything close to an expert in PSPP. If you discover things about PSPP that I haven’t noted in these documents please let me know by emailing me at ednelson@csufresno.edu.

There are some types of analysis that you cannot do from the PSPP graphical interface. In part 1 of the exercise we want to select out the cases for which the variable tv1\_tvhours is less than or equal to 13. To do this we need to run the commands from the syntax window. In the exercise I explain how to open a syntax file and tell students to cut and paste the commands into the syntax file and then I explain how to run the syntax file. This should be fairly easy for students to carry out.

This exercise isn’t a comprehensive discussion of multiple linear 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 discuss all the ways to test for multicolinearity. 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, 2015). In this exercise I assume that students have a basic understanding of a test of significance. You may want to add your own materials to this exercise or delete sections that go into more detail than you want.

To avoid students overwriting the data file, it is important that you make the data file a read only file.

Along with the exercise (.docx) file and these extended notes is the PSPP syntax needed for the exercise. This, of course, should be removed when preparing the exercise for the students. 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.

The data set includes a large number of variables that are not used in this particular exercise. I’m preparing other exercises that will use this same data set and will focus on using SPSS[[3]](#footnote-3) and exploring other areas of interest which include the following.

1. opinions about abortion
2. religious preference and religiosity and other aspects of religion
3. trust and fairness
4. tolerance towards those who hold unpopular views
5. confidence in various institutions
6. sexual behavior and opinions
7. opinions about immigration and immigrants
8. fear of crime and opinions about capital punishment and the courts
9. sex roles
10. opinions about the redistribution of wealth
11. political behavior and opinions
12. political efficacy
13. conditions under which suicide would be allowed
14. opinions about national spending
15. gun ownership and gun control
16. sociability
17. happiness and satisfaction with various aspects of life
18. willingness to help others
19. health and availability of health care
20. legalization of marijuana
21. social control of pornography
22. television viewing

Some of the exercises will also focus on various methodological and statistical issues including the following.

1. developing hypotheses
2. analysis of two-variable and three-variable relationships
3. spuriousness
4. measurement validity
5. measurement reliability
6. percentages
7. Chi Square
8. measures of association (Lambda, Cramer’s V, Gamma, Somers’ d, Goodman and Kruskal’s tau-b, Goodman and Kruskal’s tau-c)
9. interpreting data tables

Some of the exercises will focus on critical thinking skills such as the following.

1. developing arguments to support your hypotheses
2. developing tables (called dummy tables) that show what your tables should look like if your hypothesis is true
3. deciding if your data support your hypothesis

These exercises will be posted to this site as they are created.

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. There are actually more than 218 variables in the data set. There are 218 from the GSS and several variables that were created from these 218 variables. [↑](#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)
3. As of this writing, only the statistics exercises have been rewritten for PSPP. [↑](#footnote-ref-3)