Data Analysis Project (DAP) – Planning Worksheet – Choosing GSS Variables LAVC / Raskoff / Stat

---

We are using the 2014 General Social Survey (GSS) data for this project. There is a file in Canvas that we can use with the SPSS software on the lab computers and the data are accessible online. Before we go there, you need to think about what you are interested in – what about people/society do you find puzzling? Interesting? Disturbing? The GSS data has a lot of opinion data on many different topics, thus it is possible that you can analyze what people think about topics you find interesting.

Your goal is to choose four variables – one nominal, one ordinal, and two interval variables – that you think are related to each other in some way.

To use the GSS site, go to https://gssdataexplorer.norc.org/ and click on the "Explore GSS data" button… Be sure to filter the years to only include 2014 – move the circle/slider all the way over to 2014 – press search – you should have 893 variables… Feel free to make a (free) account to 'keep' your variables in while you are choosing them.

Put a keyword (or subject) into the search box to find variables interesting to you… Search for only one at a time - education, work, income, sex, and any other words that you might be interested in about opinions (guns, abortion, political, economic, …). Make sure any variable you choose does have data from 2014. Feel free to choose more than 4 – you can always narrow them down as you work on the project. Look at all the variables that come up to see if you want to choose one (or more) to analyze – click on them to see a frequency distribution or summary statistics.

Once you narrow down your variables to four (one nominal, one ordinal, and two interval), write both the long name and the short name and the categories below. Look at the categories to identify the level of measurement for each variable.

Please note that the gender variable is called 'sex' and the best income variable is 'realinc', should you choose to use either.

Variable Name (Short, no spaces)	Long Variable Name (Description)	Categories (Answer responses, do not include non-answers)	Level of Measurement	Notes

Data Analysis Project (DAP) – Planning Worksheet – **LAVC / Raskoff / Stat**

In the SPSS software, we will use the GSS data set. Here is the process to do that, whether you are in class or in the library:

- In Canvas, click on the (home page) link to the GSS data discussion page
- In the GSS discussion page, click on the GSS2014.sav file – it will open SPSS and load the data.
- Once in the data, have SPSS do what you need it to do.. There's a separate handout listing each thing, step-by-step.

As you continue to work on the project, narrow your focus to those four variables. Once you have completed Part 1, this grid is to plan your hypothesis tests for Part 2:

Final List of Variables to use in the DAP

Variable Name (Short, no spaces)	Long Variable Name (Description)	Categories (Answer responses, do not include non-answers)	Level of Measurement	Independent or Dependent? (to which other variable?)	Notes (in SPSS, which # is this variable?)

Recap of Tests and Results:

Hypothesis Test	Independent Variable	Dependent Variable	Which Test?	Results – significant?	Patterns?
#1					
#2					

Data Analysis Project (DAP) – Planning Worksheet –Part 1 **LAVC / Raskoff / Stat**

DAP, Part 1 consists of your write-up for each variable, identifying the variables and their distribution of responses. We will use the SPSS software that is on the lab computers – and also on some computers in the LARC computer lab! (If you want to play with the program outside of class, ask the lab person which computers have SPSS.)

For DAP, Part 1:

For your nominal and ordinal variables, have SPSS generate a frequency distribution and graph (pie or bar).

For your interval variable, have SPSS generate summary statistics (mean, median, mode, sd, …) and graph (histogram).

Outline for DAP, Part 1

1. Introduction – What variables are you using and why did you choose them? (Why are they interesting to you?)
2. Variable 1 – Nominal – Describe the distribution of your data for this variable. Interpret the frequency distribution and appropriate measure(s) of central tendency (Mode).
3. Variable 2 – Ordinal – Describe the distribution of your data for this variable. Interpret the frequency distribution and appropriate measure(s) of central tendency (Mode, Median) and variability (Range).
4. Variable 3 – Interval – Describe the distribution of your data for this variable. Interpret the frequency distribution and appropriate measure(s) of central tendency (Mean, Median, Mode, skew, best measure) and variability (Range, sd).
5. Variable 4 – Interval – Describe the distribution of your data for this variable. Interpret the frequency distribution and appropriate measure(s) of central tendency (Mean, Median, Mode, skew, best measure) and variability (Range, sd).
6. Conclusions – Where will you take these variables in Part 2? Which do you think are related to which other(s)? Is there one (dependent) variable that might be affected by the other three (independent) variables? Or do you have two (or more) different sets of relationships to explore?
7. Appendix – SPSS printouts (This should be uploaded to Canvas as a separate file.)

To submit this in Canvas, go to the assignment and upload two files – one for your write-up per this outline (#1-#6) and another file with the output (#7). Your output can also be uploaded as more than one file – one for each variable, if you prefer.

Statistical Technique	*Nominal variables*	*Ordinal Variables*	*Interval Variables*
Distribution Patterns	Frequency & Percent Distribution; Pie Chart or Bar Graph		Histogram or Line Chart
Central Tendency	Mode	Mode, Median	Mode, Median, Mean
Variability	N/A	Range	Range, SD

Data Analysis Project (DAP) – Planning Worksheet – Part 2 LAVC / Raskoff / Stat

---

DAP, Part 2 consists of two hypothesis tests. You choose which variables will be tested with which other variables. To do this, you need to identify which variables are independent and dependent and then do the appropriate test.

For DAP, Part 2:

For your first hypothesis test, decide which variables you will test – which two do you think are related? Identify which is independent and which is dependent. Have SPSS generate the appropriate hypothesis test and interpret the results.

For your second hypothesis test, you may use the same dependent variable and use one of your other variables as a new independent variable, or you may choose the other two variables you have not yet tested – if you think they are related in some way. Have SPSS generate the appropriate hypothesis test and interpret the results.

Outline for DAP, Part 2

1. Introduction – What variables are you using and why might they be related?
2. Hypothesis #1 – describe your two variables, which is independent and dependent, and interpret the test results. Are they related? If so, describe the relationship. If not, why might they not be related?
3. Hypothesis #2 – describe your two variables, which is independent and dependent, and interpret the test results. Are they related? If so, describe the relationship. If not, why might they not be related?
4. Conclusions – What did you learn about these variables and their (potential) relationships to each other? If you did two tests with the same dependent variable, which independent variable was more influential? What can we learn about people and society based on your analysis?
5. Appendix – SPSS printouts (This should be uploaded to Canvas as a separate file.)

To submit this in Canvas, go to the assignment and upload two files – one for your write-up per this outline (#1-#4) and another file with the output (#5). Your output can also be uploaded as two files – one for each test.

<table>
<tr><td colspan="2" rowspan="2">Which Hypothesis Test to do?</td><td colspan="3">Independent Variable</td></tr>
<tr><td>Nominal variable</td><td>Ordinal Variable</td><td>Interval Variable</td></tr>
<tr><td rowspan="3">Dependent Variable</td><td>Nominal variable</td><td colspan="2" rowspan="2">Crosstabs – Chi-Square
Percent comparison, Phi, C, Cramer's V; Lambda</td><td rowspan="2">(Group into ordinal categories & do Chi-Square etc… or use Means test)</td></tr>
<tr><td>Ordinal Variable</td></tr>
<tr><td>Interval Variable</td><td colspan="2">Means test – t or ANOVA</td><td>Scatterplot – Correlation & Regression
r, r^2, regression, etc…</td></tr>
</table>