SPSS Tutorial Lab - ANOVA

==========

1. Go to http://calcnet.mth.cmich.edu/org/spss/toc.htm

2. I've attached a dataset (hsb.xls) that you will use throughout this lab. You can also download the dataset from **http://dl.dropbox.com/u/31778479/hsb.xls** . A brief explanation of the variables in the dataset is provided below.

3. Complete the **lab tasks** listed towards the end of this message. To help you complete each task, you may choose to follow along with the videos on the website. I've gone ahead and listed the movie clips that may help you with each lab task:

Movie clips that may be helpful:
A) Overview of SPSS, Getting Started
B) Open files in SPSS
C) Define/Modify Variables
D) Transformation of Variables
E) Transformation of Variables
F) Frequencies and Descriptive
G) Explore Procedures
H) Nonparametric Runs and 1-sample (from 2:15 - 3:55)
I) One-way ANOVA
J) Nonparametric Independent Samples
X) Output: pivot table, Saving, Exporting

4. Save and send me the output you generate from the lab. Make sure you answer any questions found in the lab tasks. You can either print the output or export it (as shown in lab task X) and email it to me.

==========

Dataset summary
This dataset (hsb.xls) contains a sample of 200 observations from a survey of high school seniors in 1980. The dependent variables of interest are a set of standardized test scores in reading, writing, math, science, and social studies. The independent variables (factors) of interest are gender, race, economic status, school type, and program type.

Variable: explanation
ID: Identification number
MALE: gender (0=female, 1=male)
RACE: Ethnicity (1=African-American, 2=Asian, 3=Hispanic, 4=Caucasian)
SES: Socioeconomic status (0=low, 1=middle, 2=high)
SCHTYP: Type of school attended (0=private, 1=public)
PROGRAM: Type of classes taken in high school (0=vocational, 1=general, 2=academic)
READ: Reading test score
WRITE: Writing test score
MATH: Math test score
SCIENCE: Science test score
SOCST: Social studies test score

==========

LAB TASKS

A) Open .xls in SPSS (be sure to keep the header row as variable names when you convert the file). If you have problems importing the file into SPSS, just download the SPSS formatted version at http://dl.dropbox.com/u/31778479/hsb.sav

B) Go into **variable view** and set all appropriate variable **types** (e.g., numeric, string), **decimals**, and **measures** (scale, ordinal, nominal). If you do not set these values appropriately, you may have problems completing this lab (you may be unable to use some variables in some analyses).

C) Right now, many of the variables are coded as numeric values. When we analyze the data, we may prefer the computer to print out labels rather than numbers. For example, we may prefer to see "male" and "female" rather than 0 and 1. In the **variable view**, add labels to the variables as follows:

male variable:
A value of 0 = female
A value of 1 = male

ses variable:
A value of 0 = low
A value of 1 = middle
A value of 2 = high

schtyp variable:
A value of 0 = private
A value of 1 = public

program variable:
A value of 0 = vocational
A value of 1 = general
A value of 2 = academic

race variable:
A value of 1 = AfAm
A value of 2 = Asian
A value of 3 = Hisp
A value of 4 = White

If you click the **VIEW menu**, I believe there's an option to display these variable labels in your data.

D) The **race** contains 4 values. Suppose we are only concerned with two groups of ethnicities: white and minority. Recode the **race** variable into a new variable called **minority** where the old values of 1-3 (African-American, Asian, and Hispanic) are now 1 (to represent minority groups) and the old value 4 is now 0 (to represent non-minority). Note: To do this, you'll need to go to the **TRANSFORM->Recode to different variables** menu.

E) Our data contains scores from 5 different standardized tests. Let's create a single total test

score. Create a variable named **test** that represents the sum of the reading, writing, math, science, and social studies test scores.

F) Find the frequencies for the **race** and **schtyp** variables. Find the mean and standard deviations of the **math** and **test** variables. To find the frequencies and descriptive statistics, you'll need to go to the **ANALYZE menu**.

G) Explore **math** by the factor **race** (using the ANALYZE-->EXPLORE menu). Get descriptive statistics, a normality test, and a histogram. Does it appear as though math test scores differ for the various ethnic groups? We're going to run an ANOVA on the **math** variable to compare scores by ethnicity. Based on the normality test and histograms you just created, are you concerned about the normality or equal variance assumptions for math test scores? (Go ahead and write this answer somewhere on your output. You can do this after you print it).

H) Run a nonparametric Kolmogorov-Smirnoff test (ANALYZE-->NONPARAMETRIC-->1-sample K-S test) to determine if we can safely assume **math** test scores follow a normal distribution. Based on the p-value of this test, what conclusion can you make? (Write this somewhere on your output)

I) Conduct an ANOVA to determine if **math** scores differ among the ethnic groups. Go into the POST HOC option and run all possible Bonferroni or Tukey tests. Also, go into the OPTIONS and choose to display descriptive statistics and a homogeneity of variances test. What conclusions can you make from all the output?

J) I hypothesize that students who took an academic program in high school outperformed students who took general or vocational programs. To test my hypothesis, conduct another ANOVA to compare overall **test** scores by **schtyp**. You should include all possible Bonferroni or Tukey tests. Based on the output, what conclusions can you make? Are you concerned about any assumptions needed to conduct an ANOVA? Conduct a Kruskal-Wallis test (ANALYZE-->NONPARAMETRIC-->K-INDEPENDENT SAMPLES). Does the p-value from this nonparametric test support the decision you made from your ANOVA?

K) Do you want to compare test scores by gender, race, school type, or economic status? What test score are you interested in? Choose another analysis to conduct and briefly state any conclusions you can make from the output.

X) Save your output and export it as a pdf file. Email that file to me (or print it out and bring to class)