ERSH 8310, Fall 2009

SPSS Lab Exercise: Analysis of Variance and Sampling Distributions

September 2, 2009

To further illustrate the concepts discussed this past week, the following lab exercises have been created to function as a live example of how we view statistics. In particular, this lab will emphasize sampling distributions, partitioning variance, and evaluating hypotheses. In working with statistics today, we will compare what we observe from actual data with what we understand from the statistical distributions that underlie our tests.

Exercise #1: Collecting and inputting data into SPSS

We will use our vigilance task data from Lecture #3. Here is the story:

- ☐ There are a = 4 conditions, namely, 4, 12, 20, and 28 hours without sleep.
- ☐ There are n = 4 subjects randomly assigned to each of the different treatment conditions.
- ☐ The vigilance task score represents the number of failures to spot objects on a radar screen during a 30-minute test period.

Here are the data:

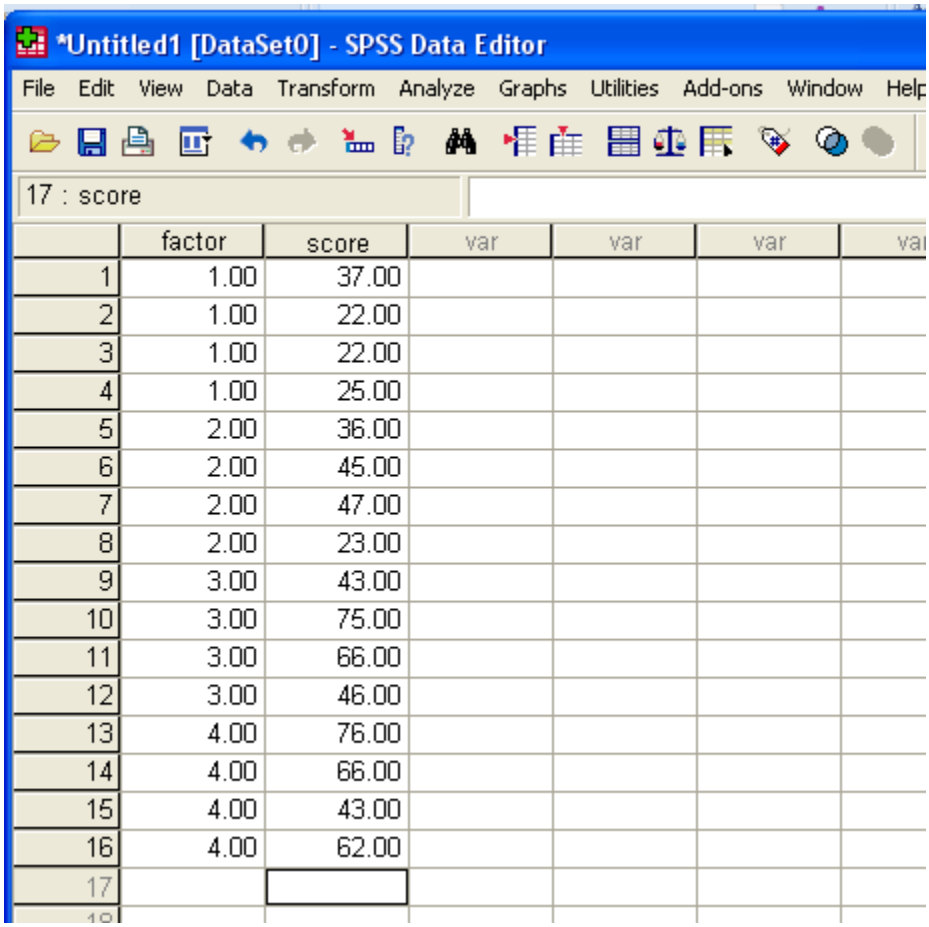
4 hr	12 hr	20 hr	28 hr
a_1	a_2	a_3	a_4
37	36	43	76
22	45	75	66
22	47	66	43
25	23	46	62

Exercise Goal: to type data into SPSS in the format needed for the analysis.

Steps:

1. Open SPSS, when asked select the option “type in data.”
2. Create two variables:
    a. One for the factor for each observation (call this FACTOR).
    b. One for the vigilance task score for each observation (call this SCORE).

3. Type in the data, using one row per observation (denote each observation's treatment condition with a 1, 2, 3, or 4 in the FACTOR variable).3
4. Save your file when finished (be sure to save your data often…crashes are very disheartening).
5. When done, your data should look like this:



	factor	score	var	var	var	var
1	1.00	37.00				
2	1.00	22.00				
3	1.00	22.00				
4	1.00	25.00				
5	2.00	36.00				
6	2.00	45.00				
7	2.00	47.00				
8	2.00	23.00				
9	3.00	43.00				
10	3.00	75.00				
11	3.00	66.00				
12	3.00	46.00				
13	4.00	76.00				
14	4.00	66.00				
15	4.00	43.00				
16	4.00	62.00				
17						

6. Close SPSS (for now).

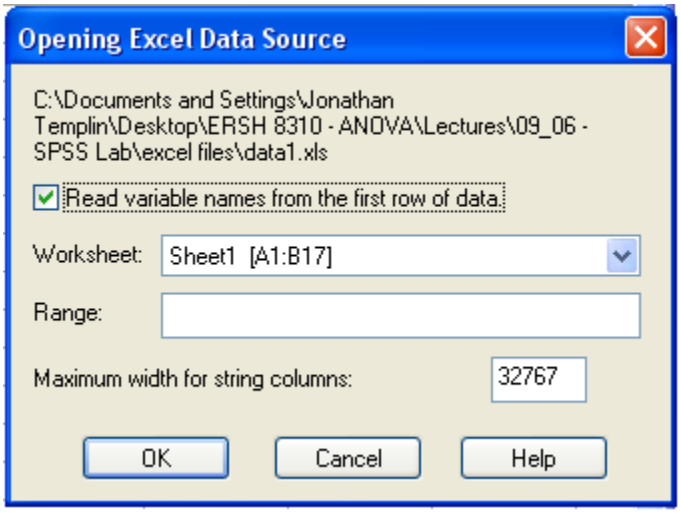
<u>Exercise #2: Understanding the sampling distribution under the null hypothesis (and creating data where we know the null hypothesis is true).</u>

Exercise Goal: To create data similar to what we have collected where we know the null hypothesis is true and determine how the distribution of F should look.

1. Determine how the data should look: how many groups and how many observations per group.
    a. Number of groups:
    b. Number of observations per group:
2. Create fake data. To do this, we will use Microsoft Excel and generate data randomly.
    a. Open Excel.
    b. In Cell A1, type "factor"
    c. In Cell B1, type "score"

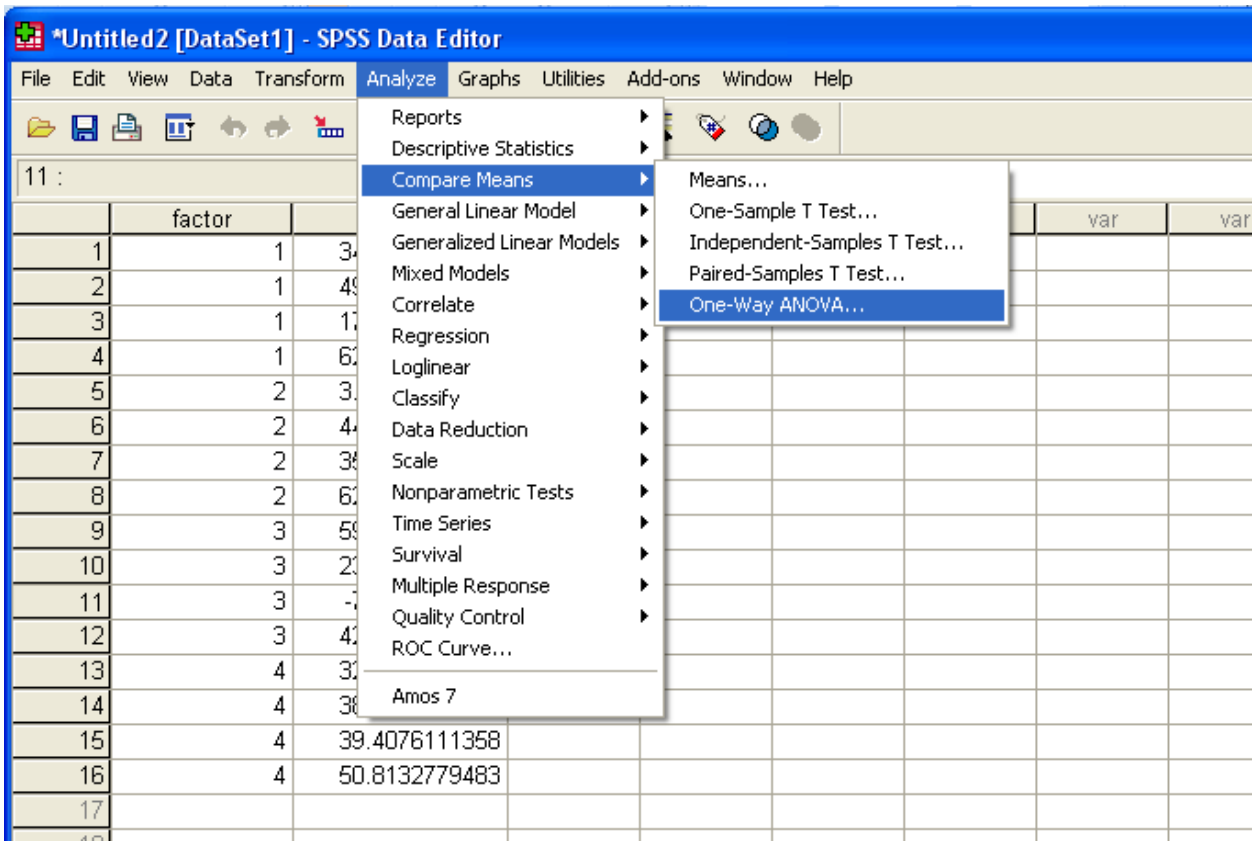
d. In cells A2-A17, repeat the condition numbers as you did in the SPSS file (1,1,1,1,2,2,…).
e. In cell B2, type =NORMINV(RAND(),45,18). This creates data that comes from a normal distribution with a mean of 45 and a standard deviation of 18.
f. Copy B2 and paste it into cells B3-B17. This does the same for all cells – so all factors will have the same mean
    i. This is our null hypothesis, only this time, we know it's true.
g. Save the file on the desktop as "data1.xls"
h. Close Excel.

3. Import the "fake" data into SPSS.
    a. Open SPSS, choose the "Open an existing data source" option.
    b. When the "Open Data" box appears:
        i. Click on the desktop button on the left.
        ii. Change the "Files of type" to be "Excel (*.xls)".
        iii. Find the "data1.xls" file you saved the "fake" data to.
        iv. Press the "Open" button.
    c. When the "Opening Excel Data Source" box appears, be sure it looks like this:

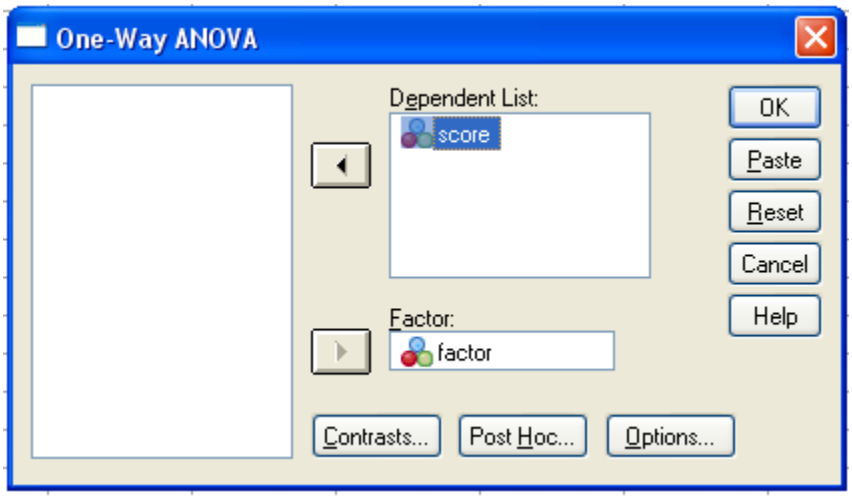


    d. Press "OK" and your data should have been imported into SPSS.

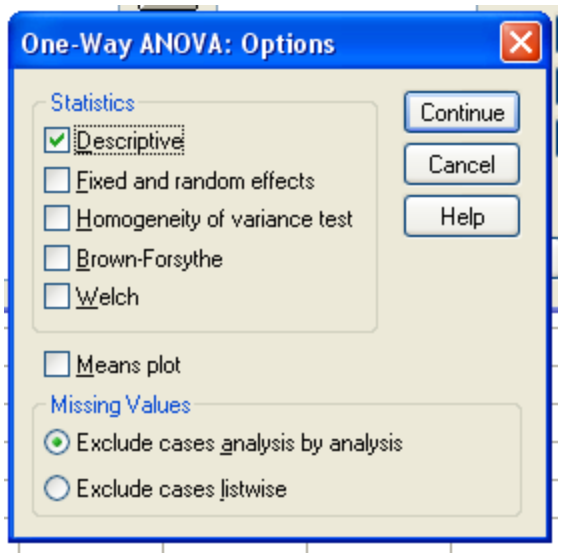
4. Find the F-ratio for your data (recall, under the null hypothesis, this should be roughly a value of 1.0 – anything different will reflect sampling error).
    a. Go to the Analyze…Compare Means…One-Way ANOVA… menu:



b. In the One-Way ANOVA box, put the “score” variable in the “dependent list” box and put the “factor” variable in the “factor” box.



c. Click on the “Options” button and check the “Descriptive” box – this gives the means and standard deviations for each factor level.
    i. Click OK



d. Click OK on the One-Way ANOVA box.
e. Check the output window for your statistic.

5. Your output should look like:

ANOVA

score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	332.790	3	110.930	.236	.869
Within Groups	5633.722	12	469.477		
Total	5966.512	15			

6. Write down the value you see for F and give that to me –I will form a histogram of all the F values in the class. NOTE: because each data set is the result of a random process, your F value will not be identical to the F values of those around you.

<u>Exercise #3: Evaluating the null hypothesis in comparison with our sampling distribution.</u>

Exercise Goal: To compute the F-value for the example data, compare it with the distribution the class generated in Exercise #2, find the theoretical p-value, and to make a decision regarding the null hypothesis.

1. Open the data you entered in Exercise 1 (you can do this by double clicking on the file).
2. Run step 4 in Exercise 2 – compute the F-value (and find the p-value under the column Sig.)
    a. F-value:
    b. p-value:
3. How does the F-value compare with the F-values generated in Example 2?
    a. Proportion of F-values larger:
4. If our Type-I error rate was set to be 0.01, would you reject the null hypothesis?
    a. Why?