Statistics 19c_SPSS.pdf	Michael Hallstone, Ph.D. hallston@hawaii.edu

Lecture 19c: Using SPSS for One Way ANOVA

The purpose of this lecture is to illustrate the how to create SPSS output for ANOVA.

Recall that ANOVA allows us to compare means of 3 or more groups (that correspond to different populations). Thus you will need a variable for the mean and a variable to chop up the mean into at least three groups. We will use the data from lecture 19 ANOVA. This means you will need two variables from your SPSS data set:

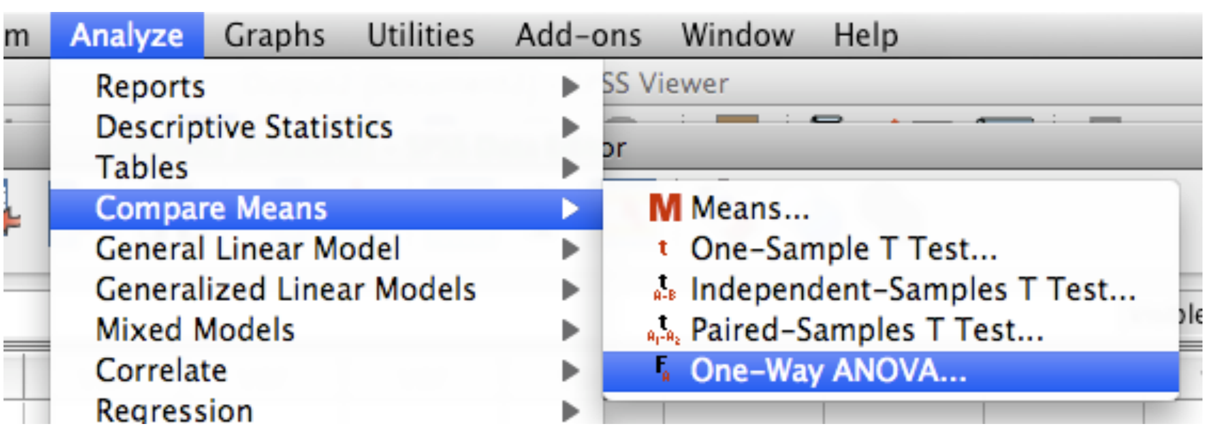
- One interval/ratio level variable to compute the mean. We have ***sales*** = *sales in thousands of $* (continuous ratio variable All students should have at least two ratio variables from the top part of the student survey. If you want to be boring you could also use age, income, or years of education – being boring or exciting will NOT affect your grade!

- One discrete variable with at least three groups in it. ***<u>A nominal or ordinal variable makes the most sense. Do not use another ratio variable, even if it is discrete.</u>*** All students should have a variable with at least 3 groups, and I also try to have everyone have a discrete variable with at least 3 groups on the top part of the student survey. ***In this example we have location= location of retail store in the City Hall (ordinal discrete variable where 1= front, 2= middle, and 3= back).***

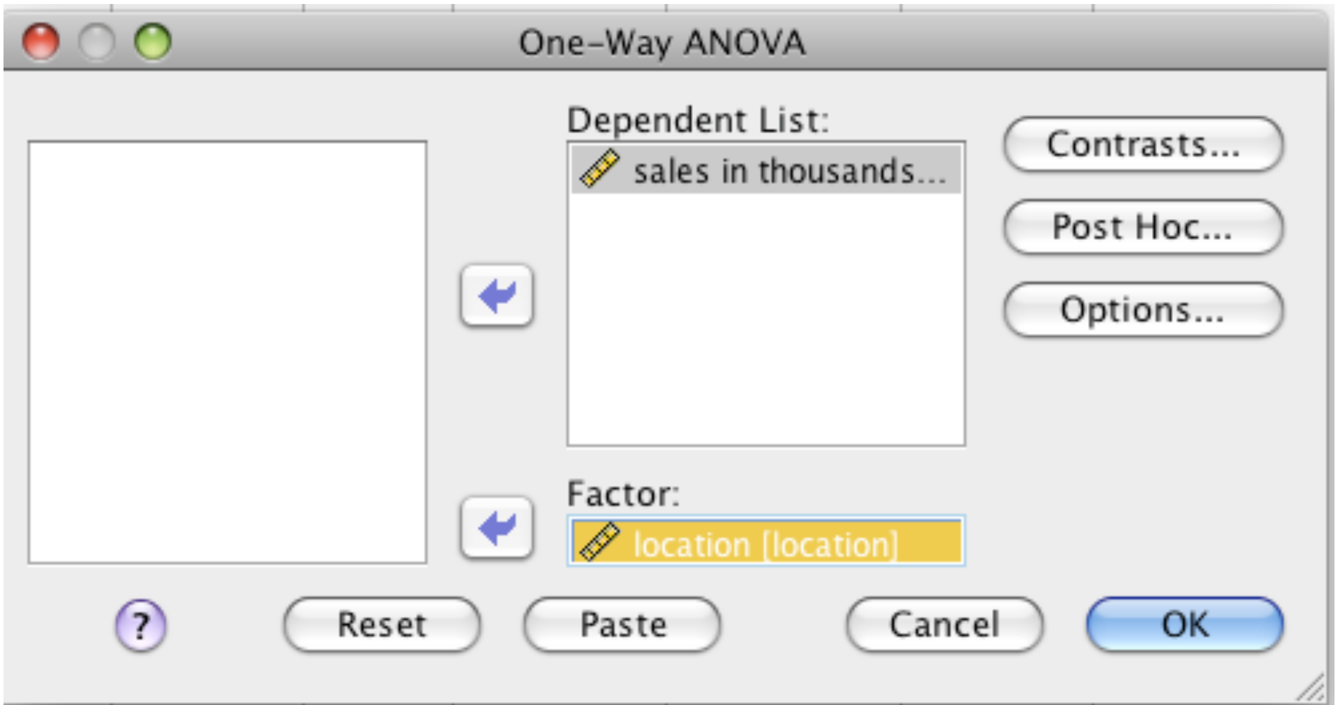
Here are the data:

	sales	location
1	45.00	1.00
2	56.00	1.00
3	47.00	1.00
4	51.00	1.00
5	50.00	1.00
6	45.00	1.00
7	55.00	2.00
8	50.00	2.00
9	53.00	2.00
10	59.00	2.00
11	58.00	2.00
12	49.00	2.00
13	54.00	3.00
14	61.00	3.00
15	54.00	3.00
16	58.00	3.00
17	52.00	3.00
18	51.00	3.00

How to have SPSS do it:

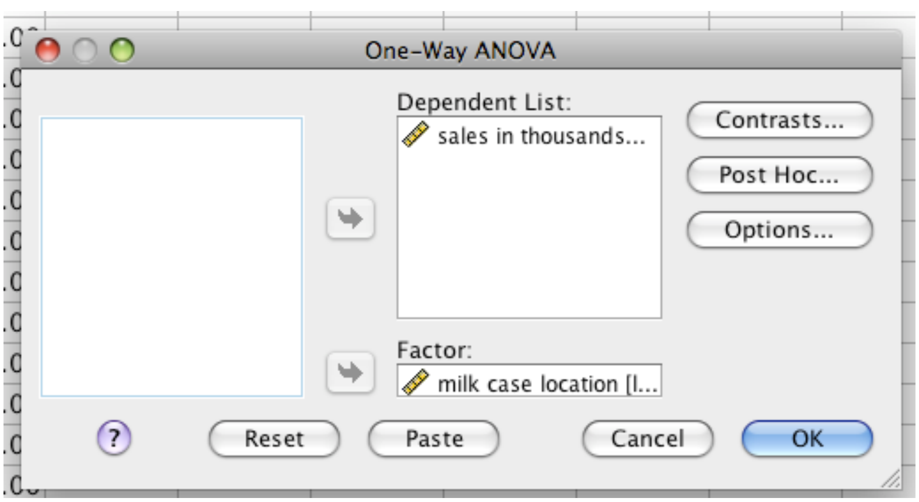


Move your interval/ratio level variable into the **Dependent List** box (our measures sales in thousands of $). Move your discrete variable with at least three groups into the **Factor** box.

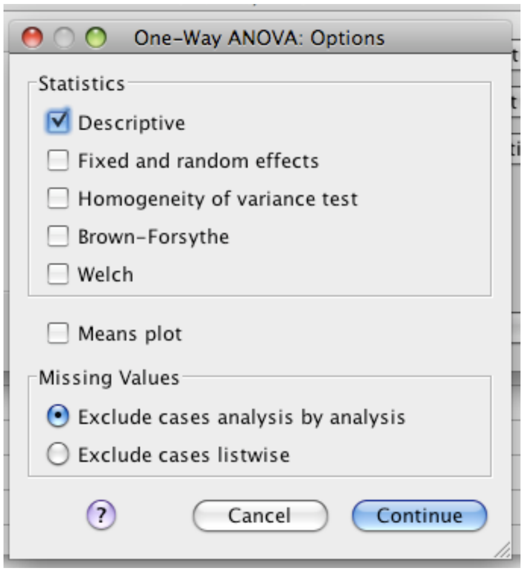


How to get the additional output [this is required for the test]

Note that you will not really need this output, but I will need it to grade your test. To get it push the OPTIONS button in this screen

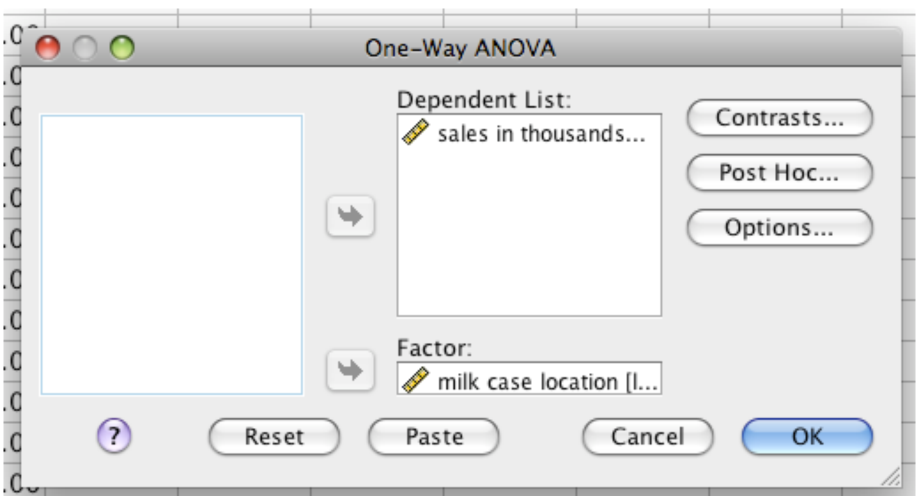


and select Descriptive Statistics.



Push Continue

You will be back at the original window that popped up. Push OK



then you will see two boxes of output.

Descriptives

sales in thousands$

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
front	6	49.0000	4.24264	1.73205	44.5476	53.4524	45.00	56.00
middle	6	54.0000	4.09878	1.67332	49.6986	58.3014	49.00	59.00
back	6	55.0000	3.79473	1.54919	51.0177	58.9823	51.00	61.00
Total	18	52.6667	4.66527	1.09961	50.3467	54.9866	45.00	61.00

ANOVA

sales in thousands$

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	124.000	2	62.000	3.780	.047
Within Groups	246.000	15	16.400		
Total	370.000	17			

This example comes from lecture 18. See it for the seven steps, but here is the language for step 7

Comparing our TR value to the critical value stated in step #5, we see that TR> 3.68. Therefore, reject the null and conclude with at least 95% confidence that at least one of the populations that corresspond to the three different City Hall retail store locations is likely to have greater or fewer mean sales than the other. Additional research is needed to define the nature of this relationship between total sales and location of retail store placement.

An example of an ANOVA where you fail to reject the null hypothesis.

Here I analyze real data from my drunk driving study in Honolulu. Recall we need two variables:

- One interval/ratio level variable to compute the mean. Here I will use total convictions separate from DUI convictions. Essentially it is a variable that takes all convictions and subtracts the DUI convictions. ***So the variable tells us how many convictions (other that DUI) each person has.***

- One discrete variable with at least three groups in it. We have ***occupation***= *type of job* (a nominal discrete variable with six categories: 1= white collar, 2= blue collar, 3= unemployed, 4= military, 5=retired, and 6= student).

  Test the theory that the mean number of convictions for the six populations that corresoond to the six occupation categories are equal. Or try to prove that at least one of the occupation populations has a different mean number of convictions than the others.

1. State the null and alternative hypothesis (H_0 and H_1).

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$
H_1: not all population means are equal

2. State level of significance or α "alpha."
For this example we'll use alpha =.05

3. Determine the test distribution to use – ANOVA tests use F distribution.

4. Define the rejection regions. And draw a picture!
$df_{numerator}$=k-1 $df_{denominator}$=T-k
k= # of samples T=total # of items in all samples
for "samples" think number of groups or categories the I/R variable is "chopped up" into. Or how many categories are there in your discrete variable. Occupation has 6 categories.
in this case $df_{numerator}$=6-1 =5 & $df_{denominator}$=435-5=430
$F_{(5,430,\ \alpha=.05)}$=2.21 (draw it out)

5. State the decision rule.
Reject the null if the TR >2.21, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.
In this case
TR= 1. 981 (by the way spss does the math below using the numbers in the Mean Square column)

$$TR= \frac{\hat{\sigma}^2 between}{\hat{\sigma}^2 within} = 31.103 \text{ divided by } 15.7 = 1.981$$

Descriptives

sales in thousands$

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	Upper Bound	Minimum
front	6	49.0000	4.24264	1.73205	44.5476	53.4524	45.00
middle	6	54.0000	4.09878	1.67332	49.6986	58.3014	49.00
back	6	55.0000	3.79473	1.54919	51.0177	58.9823	51.00
Total	18	52.6667	4.66527	1.09961	50.3467	54.9866	45.00

ANOVA

total convictions - dui convictions

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	155.514	5	31.103	1.981	.080
Within Groups	6750.841	430	15.700		
Total	6906.356	435			

7. **Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)**

Comparing our TR value to the critical value stated in step #5, we see that TR< 2.37. Therefore, FAIL TO REJECT the null. There is insufficient evidence to reject the theory that the mean number of convictions for the six populations that correspond to the six occupation categories are equal.