

Self-Assessment

Weeks 11: ANOVA and Regression

1. Linked below are the blood pressure data files that were presented in previous self-assessments.

SPSS

<http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week09/BloodPressureDrugs.sav>

Excel

<http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week09/BloodPressureDrugs.xlsx>

Perform the following analysis using ANOVA and regression comparing diastolic blood pressure among the four drug treatments (Ziac, Losartan, Lisinopril 40mg, Lisinopril 12.5mg), i.e.,

Systolic Blood Pressure = b_0 + Drug Treatment

(a) Compare the ANOVA summary table between regression and ANOVA. What similarities do you notice?

(b) Compare model fit between ANOVA and regression. What similarities do you notice?

Note: If ANOVA results do not present an R^2 value, it can be calculated by taking the ratio of the model sums of squares (SS) divided by the total SS. For example, if the model SS is 25 and the total SS is 100, the $R^2 = 25/100 = .25$. If using SPSS Unianova (General Linear Model) command, use the ratio of the corrected model SS to the corrected total SS to find the model R^2 .

(c) Compare both Bonferroni and Scheffé CIs for the pairwise comparisons. What similarities do you notice?

(d) After studying the above results, what conclusions do you draw about ANOVA and regression?

2. Below is a data file containing the following variables for cars taken between 1970 and 1982:

mpg: miles per gallon
engine: engine displacement in cubic inches
horse: horsepower
weight: vehicle weight in pounds
accel: time to accelerate from 0 to 60 mph in seconds
year: model year (70 = 1970, to 82 = 1982)
origin: country of origin (1=American, 2=Europe, 3=Japan)
cylinder: number of cylinders

SPSS Data: http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week04/cars_missing_deleted.sav

(Note: There are underscore marks between words in the SPSS data file name.)

Other Data Format: If you prefer a data file format other than SPSS, let me know.

For this problem we wish to know whether MPG differs among car origins and number of cylinders. The regression model for this study follows:

Predicted MPG = b_0 + origin of car + number of cylinders

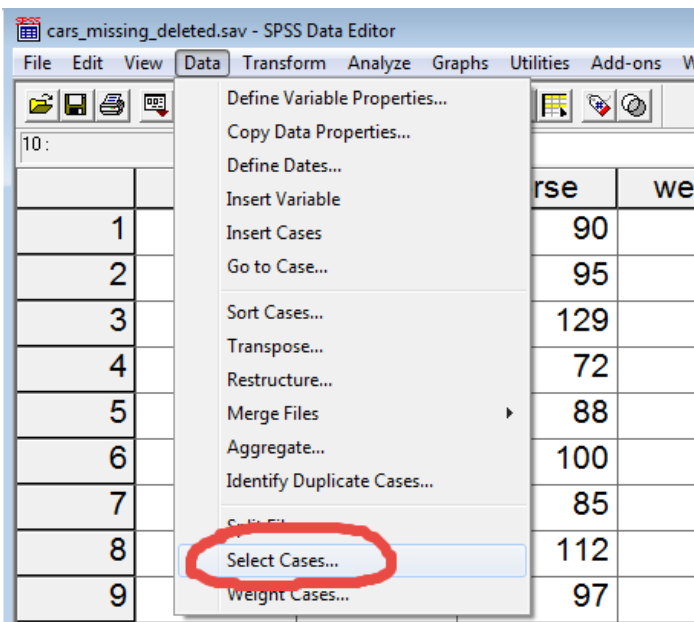
Origin of car is categorical. Number of cylinders may appear to be ratio, but since observed categories of this variable are limited, it is best to treat this variable as categorical. Note the following number of cylinders reported:

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
3 Cylinders	4	1.0	1.0	1.0
4 Cylinders	199	50.9	50.9	51.9
5 Cylinders	3	.8	.8	52.7
6 Cylinders	83	21.2	21.2	73.9
8 Cylinders	102	26.1	26.1	100.0
Total	391	100.0	100.0	

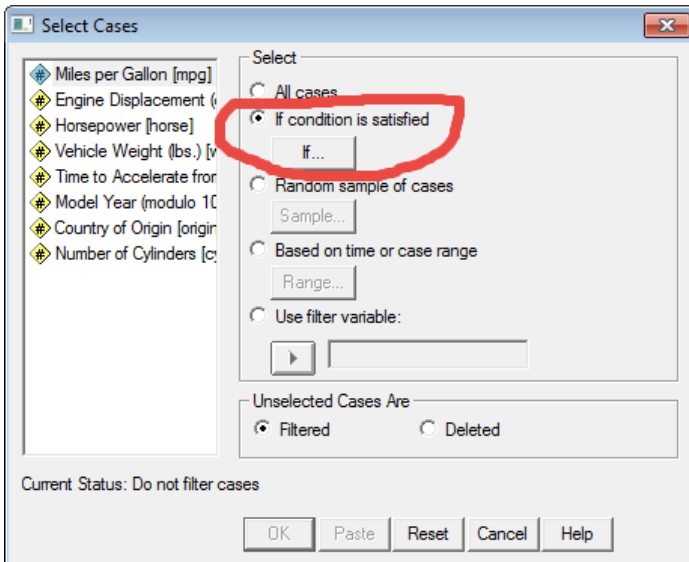
As the frequency display above shows, the number of cylinders include 3, 4, 5, 6, and 8. However, only 4 cars had 3 cylinders and only 3 cars had 5 cylinders. Given the small sample sizes for these categories, it is best to remove these cases from the regression analysis. There are several ways to accomplish this. Four approaches are (a) manually delete these cases after sorting all cases on number of cylinders, (b) telling SPSS to treat these 7 cases as missing values so they will not be included in any analysis (use Recode into Same Variable and set 3 Cylinders and 5 Cylinders as system missing), (c) defining 3 and 5 Cylinders as missing values in the variable missing values, or (d) using the Select Cases command to filter these cases from all analyses. Other possibilities also exist.

Of these four, option (d) works well and does not require deletion of any cases. This option is explained below.

Step 1: Open the Select Case window



Step 2: Choose the select If option



Step 3: Define the filter so SPSS can determine which cases NOT to select.

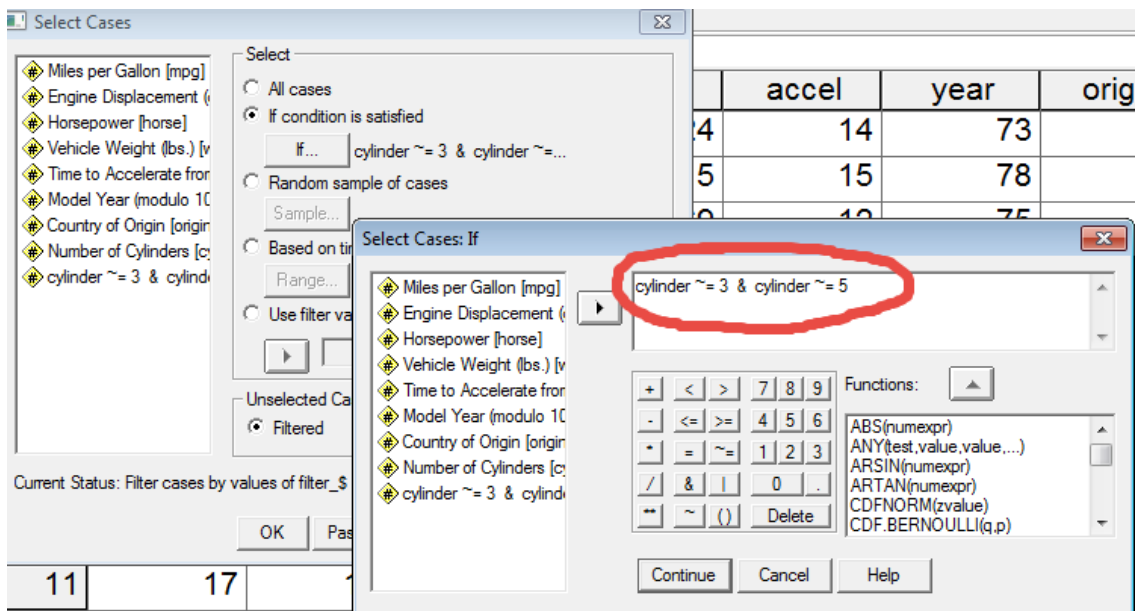
We do not want cylinders of 3 or 5, so in the Select Cases IF box, write

cylinder \neq 3

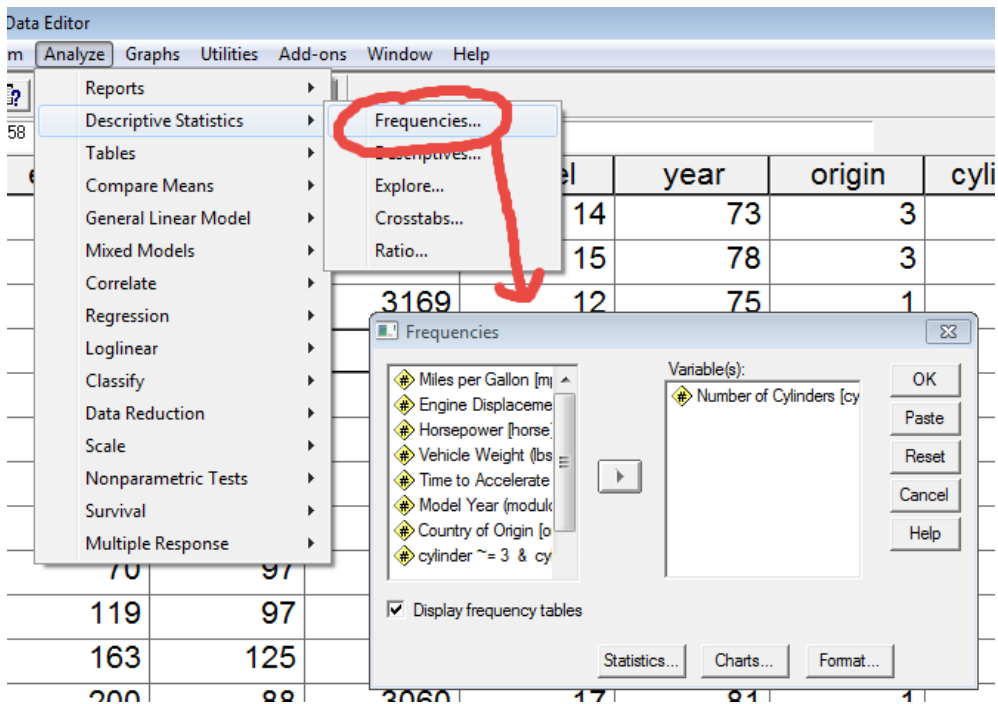
The symbol \neq means “not equal”; this tells SPSS not to select any cases in which cylinders are 3. Also, write

cylinder \neq 5

so SPSS knows not to select cases when cylinders are 5. To combine these two, we use the ampersand symbol, &, which means select all cases which are not 3 and 5 cylinders. See image below.



Once these cases are defined, click Continue then OK to process this command. Next, check that the appropriate cases were selected by running the Frequency command for cylinders as shown below.



And the results should look like this:

➔ **Frequencies**

Statistics

Number of Cylinders

N	Valid	384
	Missing	0

Number of Cylinders

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4 Cylinders	199	51.8	51.8	51.8
	6 Cylinders	83	21.6	21.6	73.4
	8 Cylinders	102	26.6	26.6	100.0
	Total	384	100.0	100.0	

Note that no cases of cylinders equal to 3 or 5 were selected.

(a) Compare the ANOVA summary table between regression and ANOVA. What similarities do you notice for the overall model test? Note that when running General Linear Model in SPSS, you must specify a custom model that does NOT include the interaction between Origins and Cylinders. Also, when comparing overall models in the ANOVA summary table, use the Corrected Model line from the General Linear Model output since this line tests the complete model.

(b) Compare model fit between ANOVA and regression. What similarities do you notice?

Note: If ANOVA results do not present an R^2 value, it can be calculated by taking the ratio of the model sums of squares (SS) divided by the total SS. For example, if the model SS is 25 and the total SS is 100, the $R^2 = 25/100 = .25$. If using SPSS Unianova (General Linear Model) command, use the ratio of the corrected model SS to the corrected total SS to find the model R^2 .

(c) Compare partial F tests (i.e., F ratio, degrees of freedom) for Origins and Number of Cylinders between ANOVA and Regression. Recall that to obtain the partial F test in Regression, you must test the variable contribute by testing the ΔR^2 value.

(d) After studying the above results, what conclusions do you draw about ANOVA and regression when used to analyze more than one predictor?

3. Using the same cars data provided in Question 2, perform a two-way ANOVA on MPG with Origins and Number of Cylinders as the predictors, i.e.,

$$\text{Predicted MPG} = b_0 + \text{origin of car} + \text{number of cylinders}$$

and ANOVA report results in APA style. Set alpha = .05 and use the Bonferroni adjustment for any multiple comparisons that are performed.