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 = b0 + 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, of 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: <u>http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week04/cars_missing_deleted.sav</u> (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 = b0 + 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:

Number of Cylinders									
		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.

🛅 cars_missing_deleted.sav - SPSS Data Editor									
File Edit View	Data Transform Analyze Graphs U	tilities Ad	d-ons W						
<mark>⊯∎⊜</mark> <u>■</u> 10:	Define Variable Properties Copy Data Properties Define Dates	F S	0						
	Insert Variable	rse	we						
1	Insert Cases	90							
2	Go to Case	95							
3	Sort Cases	129							
4	Restructure	72							
5	Merge Files	88							
6	Aggregate Identify Duplicate Cases	100							
7	C. In Fil	85							
8	Select Cases	112							
9	Weight Cases	97							

Step 1: Open the Select Case window

Step 2: Choose the select If option

Select Cases	
 Miles per Gallon [mpg] Engine Displacement (Horsepower [horse] Vehicle Weight (bs.) [v Time to Accelerate fror Model Year (modulo 10 Country of Origin [origin Number of Cylinders [c) 	Select C All cases If condition is satisfied If C Random sample of cases Sample C Based on time or case range Range C Use filter variable:
Current Status: Do not filter cas	Unselected Cases Are Filtered C Deleted ses OK Paste Reset Cancel Help

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

cylinders ~= 3

The symbol ~= means "not equal"; this tells SPSS not to select any cases in which cylinders are 3. Also, write

cylinders ~= 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.

Select Cases			23				
Miles per Gallon [mpg] Engine Displacement (C All cases				accel	vear	oria
Horsepower [horse] Vehicle Weight (bs.) [v	If condition is If	satisfied :vlinder ~= 3 & cvlinder ~=		4	14	73	
 Time to Accelerate fror Model Year (modulo 10) 	C Random sam	ple of cases		5	15	78	
Country of Origin [origin When the second	C Based on tir	Select Cases: If			40	76	×
	Range Image: Miles per Gallon [mpg] C Use filter va Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] Image: Miles per Gallon [mpg] <td></td> <td colspan="4">cylinder ~= 3 & cylinder ~= 5</td>			cylinder ~= 3 & cylinder ~= 5			
	Unselected Ca • Filtered	 Vehicle Weight (bs.) [v Time to Accelerate fron Model Year (modulo 10 Country of Origin [origin 	+ -		7 8 9 Func 4 5 6 ABS	tions:	<u> </u>
Current Status: Filter cases by	values of filter_\$ OK Pas	 Wumber of Cylinders [c] cylinder ~= 3 & cylinder 		8	ARS O Delete CDF	SIN(numexpr) AN(numexpr) NORM(zvalue) BERNOULLI(q,p)	Ŧ
11 17	7 1		Conti	inue	Cancel H	elp	

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.

Dat	a Edit	or										
m	Ana	lyze) Gra	phs Util	ities Ad	d-ons	Window I	Help					
?		Reports			۱.		-	-				
58		Descript	ive Statisti	cs	• 🕻	Frequencie	is	-				
		Tables				Essenptive	5		Vear	origin		a v/
		Compar	e Means		•	Explore	· •		year	Ongin		Cyl
		General	Linear Mo	del	•	Crosstabs	· •	14	73		3	
		Mixed M	lodels		•	Ratio		15	78		3	
	-	Correlat	e		• F	3160	J	12	75		1	
		Regressi	on		۰ <u>L</u>		ncies	12				82
		Loglinea	ir		۲.		incres .	_	Mariable (a)		l	
	1	Classify			<u>۲</u>	Miles	per Gallon [m	<u> </u>	Wanable(s).	of Cylinders Icy	0	K
	-	Data Rec	duction		• -	Horse	e Displaceme nower horse				Pas	te
		Scale			۲ <u> </u>	Vehic	le Weight (bs				Res	set
		Nonpara	ametric Te	sts	۲.	Time	to Accelerate	1 L	<u>}</u>		Can	
		Survival			• -	- Mode	l Year (moduk					
	-	Multiple	Response		• -	- Couri	er∼=3&cv/				He	
	_	70		97					1			
		119		97		Display	rfrequency tal	oles				
		163		125				S	tatistics Charts	Format		
		200		00		2060		17	Q1	I	1	

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, of 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.,

Predicted MPG = b0 + origin of car + 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.