Self-Assessment

Weeks 6 and 7: Multiple Regression with a Qualitative Predictor; Multiple Comparisons

1. Suppose we wish to assess the impact of five treatments on an outcome Y. How would these five treatments be coded as dummy variables? Present actual data to illustrate the coding.

> **Each case for a given treatment would be coded as 1, and if a case does not occur for that treatment, it would be coded as 0.**

Y	Treatment 1	Treatment 2	Treatment 3	Treatment 4	Treatment 5	Treatment Received
Score	1	0	0	0	0	1
Score	0	1	0	0	0	2
Score	0	0	1	0	0	3
Score	0	0	0	1	0	4
Score	0	0	0	0	1	5

2. Below is a regression analysis with two variables:

Heart rate = beats per minute

Blood Pressure Medication = four drugs prescriptions (Losartan, Ziac, Lisinopril [12.5mg], and Lisinopril [40mg])

Dummy variables were created for Ziac, Lisinopril 12.5, and Lisinopril 40.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.741(a)	.550	.531	4.75098

a Predictors: (Constant), lisinopril_40, lisinopril_12_5, ziac

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1982.987	3	660.996	29.284	.000(a)
	Residual	1625.171	72	22.572		
	Total	3608.158	75			

a Predictors: (Constant), lisinopril_40, lisinopril_12_5, ziac
b Dependent Variable: heart_rate

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	61.714	1.037		59.527	.000
	ziac	-12.597	1.550	-.762	-8.127	.000
	lisinopril_12_5	-1.067	1.550	-.065	-.689	.493
	lisinopril_40	-.190	1.466	-.012	-.130	.897

a Dependent Variable: heart_rate

Coefficients(a)

Model		95% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	59.648	63.781
	ziac	-15.687	-9.507
	lisinopril_12_5	-4.157	2.023
	lisinopril_40	-3.113	2.732

a Dependent Variable: heart_rate

(a) Overall does it appear heart rate differs by blood pressure medication at the .05 level? Explain how you arrived at an answer for this question.

> **Yes, because the model F ratio, 29.28, which is significant at the .05 level (p=.000), tests whether there are mean differences among each of the four drugs if drug is the only variable in the question.**

(b) Provide literal interpretations for each of the four unstandardized regression coefficients presented in the coefficients table.

> **B0 = 61.71: This is the mean heart rate when taking Losartan**
> **B1 = -12.597: The mean heart rate is 12.59 lower under Ziac compared to the heart rate when taking Losartan.**
> **B2 = -1.067: The mean heart rate is 1.067 lower under Lisinopril 12.5 compared to the heart rate when taking Losartan.**
> **B3 = -0.19: The mean heart rate is 0.19 lower under Lisinopril 40 compared to the heart rate when taking Losartan.**

(c) What is the predicted mean heart rate for someone taking Lisinopril 40? For someone taking Ziac?

> **Lisinopril 40:**
> **Predicted Heart Rate = 61.714 -12.597(Ziac) -1.067(Lisin 12.5) -0.190(Lisin 40)**
> **Predicted Heart Rate = 61.714 -12.597(0) -1.067(0) -0.190(1)**
> **Predicted Heart Rate = 61.714 -0.190(1)**
> **Predicted Heart Rate = 61.524**
>
> **Ziac:**
> **Predicted Heart Rate = 61.714 -12.597(Ziac) -1.067(Lisin 12.5) -0.190(Lisin 40)**
> **Predicted Heart Rate = 61.714 -12.597(1) -1.067(0) -0.190(0)**
> **Predicted Heart Rate = 61.714 -12.597(1))**
> **Predicted Heart Rate = 49.117**

(d) Which regression coefficients are statistically significant at the .05 level for the dummy variables?

> **Only the Ziac coefficient is statistically significant (t = -8.12, p<.001).**

(e) Write a very brief, one sentence interpretation for each dummy variable coefficient in the SPSS coefficients table above. Take into account significance testing with an alpha = .05.

> **Ziac, b = -12.597: Significant so this means heart rate is lower when takin Ziac than when taking Losartan.**
> **Lisin 12.5, b = -1.067: Not significant so this means heart rate is similar between Lisinopril 12.5 and Losartan.**
> **Lisin 40, b = -0.190: Not significant so this means heart rate is similar between Lisinopril 40 and Losartan.**

(f) Model R^2 = .55 – what does this tell us?

> **About 55% of variance in heart rate can be predicted by drug prescription.**

(g) What is the predicted mean heart rate difference between Ziac and Lisinopril 12.5?

> **Predicted mean for Ziac = 49.117**
> **Predicted mean for Lisin 12.5 = 60.647**
> **Mean difference = 49.117-60.647 = -11.53**

(h) What is the interpretation for the 95% confidence interval for b3 (Lisin 40 dummy)?

> **One may be 95% confident that the true mean difference in heart rate between Lisinopril 40 and Losartan lies between -3.113 and 2.732.**

3. If one wishes to compare Y across five treatments, and the per comparison alpha = .10,

(a) how many pairwise comparisons are possible, and

> **If there are five treatments, then there are 10 total pairwise comparisons. The following formula can be used to compute the number of pairwise comparisons where N is the number of units/groups to compare:**
>
> **((N – 1) N) / 2 = number of pairwise comparison possible**
> **((5 – 1) 5) / 2**
> **(4*5)/2**
> **20/2 = 10**

(b) what would be the familywise error rate?

> **The formula to determine familywise error rate is**
>
> **Familywise error rate = 1 - (1 - alpha per comparison)^{C}**
>
> **where c is the number of comparisons to perform.**
>
> **Familywise error rate = 1 - (1 - alpha per comparison)^{C}**
> **Familywise error rate = 1 - (1 - .10)^{10}**
> **Familywise error rate = 1 - (.90)^{10}**
> **Familywise error rate = .6514**

(c) Interpret the familywise error rate calculated in (b) above.

The probability of committing a Type 1 error across the 10 comparisons is .6514 when each comparison as a per comparison alpha of .10.

4. Question 2 above presented SPSS results for heart rate compared across four drug prescriptions. The per comparison alpha was set at .05. The number of observations for the study is 76.

(a) There are six possible pairwise comparisons among the four drug prescriptions. This means there are six different confidence intervals that could be computed. Use a familywise error rate of .05 when considering these comparisons. Based upon this information, calculate and present the Bonferroni confidence interval for the Ziac vs. Losartan comparison.

For simplicity, I used the Bonferroni adjusted t critical values found in this table

http://www.bwgriffin.com/gsu/courses/edur8132/notes/bonferroni_critical_t_values.pdf

Bonferroni adjusted t = 2.721 (for 6 companions, alpha = .05, and df2 = 60; note cannot use next level of 120)

.95CI = b ± t (Bonferroni t) (SE b)
.95CI = -12.597 ± (2.721) (1.55)
Upper = -8.379
Lower = -16.814

A more precise critical t value can be obtained with Excel using the critical t value function:

=T.INV.2T(adjusted alpha, df)

The adjusted alpha would be .05/6 = .008333, and the df for this model is 76-3-1 = 72 (also reported in the ANOVA table from SPSS above), so

=T.INV.2T(.008333, 72)
= -2.7129

So with this critical t the CI would be

.95CI = b ± t (Bonferroni t) (SE b)
.95CI = -12.597 ± (2.7129) (1.55)
Upper = -8.392
Lower = -16.802

This value matches slightly better with the CI provided by SPSS.

(b) Using the same information presented in (a), construct and present a Scheffé confidence interval for the Ziac vs. Losartan comparison.

To calculate the CI, first one must find the critical F ratio:

df1 = J − 1 = 4 − 1 = 3
df2 = n − k − 1 = 76 − 3 − 1 = 72

where J is the number of groups, and k is the number of dummy variables in the regression equation.

We want an overall familywise error rate of .05.

Critical F can be found in Excel using the following function

=F.INV.RT(alpha level, df1 , df2)
=F.INV.RT(0.05,3,72)
= 2.7318

Next convert this critical F ratio to a Scheffé adjusted F ratio

Scheffé F = (J − 1) (original critical F)
Scheffé F = (3) (2.7318)
Scheffé F = 8.1954

Next convert this Scheffé F ratio to a critical Scheffé t value by taking the square root of the Scheffé F:

Scheffé t = $\sqrt{\text{Scheffé F}}$

Scheffé t = $\sqrt{8.1954}$

Scheffé t = 2.8627

Now use the CI formula to find the upper and lower limits:

.95CI = b ± t (Scheffé t) (SE b)
.95CI = -12.597 ± (2.8627) (1.55)
Upper = -8.1598
Lower = -17.034

5. 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:

Predicted MPG = b0 + origin of car with appropriate dummy variables

Present an APA styled regression analysis with DV = MPR and IV = origin. Set alpha = .01. You will have to create the dummy variables for origins. Also present both the Bonferroni and Scheffé confidence intervals comparing each of the origins.

APA styled response.

Table 1
Descriptive Statistics and Correlations Between MPG and Origin of Vehicle Manufacturer

Variable	Correlations		
	MPG	**USA**	**Europe**
MPG	---		
USA	-.56*	---	
Europe	.24*	-.59*	---
Mean	23.48	0.62	.17
SD	7.78	.48	.38

Note. USA (1 = made in USA; 0 = others) and Europe (1 = made in Europe, 0 = others) are dummy variables.
n = 391.
*p < .01.

Table 2
Regression Analysis of MPG by Origin of Vehicle

Variable	b	se b	99% CI	t
USA	-10.37	0.83	-12.51, -8.23	-12.56*
Europe	-2.85	1.06	-5.58, -0.12	-2.70*
Intercept	30.45	0.72	28.59, 32.31	42.42*

Note. R^2 = .33, adj. R^2 = .33, $F_{2,388}$ = 96.03*, MSE = 40.703, n = 391. USA (1 = made in USA; 0 = others) and Europe (1 = made in Europe, 0 = others) are dummy variables.
*p < .01.

Table 3
Multiple Comparison in MPG by Origin of Vehicle

Comparison	Estimated Mean Difference	Standard Error of Difference	99% Bonferroni Adjusted CI of Mean Difference	99% Scheffé Adjusted CI of Mean Difference
USA vs. Japan	-10.37*	0.83	-12.81, -7.93	-12.89, -7.85
Europe vs. Japan	-2.85	1.06	-5.97, 0.27	-6.07, 0.37
USA vs Europe	-7.52*	0.88	-10.11, -4.94	-10.20, -4.86

*$p < .01$, where p-values are adjusted using the Bonferroni or Scheffé method depending upon which column is viewed.

(Note to EDUR 8132 students: Normally one would not present both the Bonferroni and Scheffé corrected confidence intervals, but for convenience I have included both in one table.)

Regression results show that there are statistically significant mean differences in MPG across vehicles based on vehicle origin. More specifically, the table of multiple comparisons shows that vehicles made in the USA tend to have lower MPG than those made in Japan or Europe, and these differences are significant at the .01 level using the Bonferroni (or Scheffé) adjustment for comparisons. The MPG difference between European and Japanese cars is not statistically significant suggesting that vehicles from these two areas produce similar mean MPGs.

SPSS Output Used to Create the Above Tables

Correlations obtained using SPSS correlation command.

Descriptive Statistics

	Mean	Std. Deviation	N
Miles per Gallon	23.48	7.781	391
USA	.6240	.48499	391
Europe	.1739	.37952	391

Correlations

		Miles per Gallon	USA	Europe
Miles per Gallon	Pearson Correlation	1	-.564(**)	.243(**)
	Sig. (2-tailed)		.000	.000
	N	391	391	391
USA	Pearson Correlation	-.564(**)	1	-.591(**)
	Sig. (2-tailed)	.000		.000
	N	391	391	391
Europe	Pearson Correlation	.243(**)	-.591(**)	1
	Sig. (2-tailed)	.000	.000	
	N	391	391	391

** Correlation is significant at the 0.01 level (2-tailed).

Below are SPSS results using General Linear Model – this will be converted to regression. See Parameter Estimates table for regression results. General Linear Model is convenient to use because it can provide Bonferroni and Scheffe comparisons.

Descriptive Statistics

Dependent Variable: Miles per Gallon

Country of Origin	Mean	Std. Deviation	N
American	20.08	6.415	244
European	27.60	6.580	68
Japanese	30.45	6.090	79
Total	23.48	7.781	391

Tests of Between-Subjects Effects

Dependent Variable: Miles per Gallon

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7817.309(a)	2	3908.655	96.030	.000
Intercept	194029.594	1	194029.594	4767.050	.000
origin	7817.309	2	3908.655	96.030	.000
Error	15792.466	388	40.702		
Total	239224.740	391			
Corrected Total	23609.775	390			

a R Squared = .331 (Adjusted R Squared = .328)

Estimates

Dependent Variable: Miles per Gallon

			99% Confidence Interval	
Country of Origin	Mean	Std. Error	Lower Bound	Upper Bound
American	20.079	.408	19.021	21.136
European	27.603	.774	25.600	29.606
Japanese	30.451	.718	28.593	32.309

Parameter Estimates

Dependent Variable: Miles per Gallon

					99% Confidence Interval	
Parameter	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
Intercept	30.451	.718	42.423	.000	28.593	32.309
[origin=1]	-10.372	.826	-12.559	.000	-12.510	-8.234
[origin=2]	-2.848	1.055	-2.698	.007	-5.580	-.116
[origin=3]	0(a)	.	.	.	.	.

a This parameter is set to zero because it is redundant.

Origin 1 = American
Origin 2 = European
3 = Japanese

Pairwise Comparisons

Dependent Variable: Miles per Gallon

(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.(a)	99% Confidence Interval for Difference(a)	
					Lower Bound	Upper Bound
American	European	-7.524(*)	.875	.000	-10.108	-4.940
	Japanese	-10.372(*)	.826	.000	-12.811	-7.933
European	American	7.524(*)	.875	.000	4.940	10.108
	Japanese	-2.848	1.055	.022	-5.965	.269
Japanese	American	10.372(*)	.826	.000	7.933	12.811
	European	2.848	1.055	.022	-.269	5.965

Based on estimated marginal means
* The mean difference is significant at the .01 level.

a Adjustment for multiple comparisons: Bonferroni.

Multiple Comparisons

Dependent Variable: Miles per Gallon

	(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.	99% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	American	European	-7.52(*)	.875	.000	-10.20	-4.85
		Japanese	-10.37(*)	.826	.000	-12.89	-7.85
	European	American	7.52(*)	.875	.000	4.85	10.20
		Japanese	-2.85	1.055	.027	-6.07	.37
	Japanese	American	10.37(*)	.826	.000	7.85	12.89
		European	2.85	1.055	.027	-.37	6.07
Bonferroni	American	European	-7.52(*)	.875	.000	-10.11	-4.94
		Japanese	-10.37(*)	.826	.000	-12.81	-7.93
	European	American	7.52(*)	.875	.000	4.94	10.11
		Japanese	-2.85	1.055	.022	-5.96	.27
	Japanese	American	10.37(*)	.826	.000	7.93	12.81
		European	2.85	1.055	.022	-.27	5.96

Based on observed means.

* The mean difference is significant at the .01 level.