Nate Bitting
Assignment 4

Assignment #4: Statistical Inference in Linear Regression (50 points)

This assignment will be made available in both pdf and Microsoft docx format. Answers should be typed into the docx file, saved, and converted into pdf format for submission. **Color your answers in green so that they can be easily distinguished from the questions themselves.**

Throughout this assignment keep all decimals to four places, i.e. X.xxxx.

Any computations that involve "the log function", denoted by log(x), are always meant to mean the natural log function (which will show as ln() on a calculator). The only time that you should ever use a log function other than the natural logarithm is if you are given a specific base.

In this assignment we will review model output from SAS and perform the computations related to statistical inference for linear regression. By performing this computations we are ensuring that we understand how the numbers in this SAS output are computed.

Nate Bitting
Assignment 4

Model 1: Let's consider the following SAS output for a regression model which we will refer to as Model 1.

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	4	2126.00904	531.50226		<.0001
Error	67	630.35953	9.40835		
Corrected Total	71	2756.36857			

Root MSE	3.06730	**R-Square**	
Dependent Mean	37.26901	**Adj R-Sq**	
Coeff Var	8.23017		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	11.33027	1.99409	5.68	<.0001
X1	1	2.18604	0.41043		<.0001
X2	1	8.27430	2.33906	3.54	0.0007
X3	1	0.49182	0.26473	1.86	0.0676
X4	1	-0.49356	2.29431	-0.22	0.8303

Number in Model	C(p)	R-Square	AIC	BIC	Variables in Model
4	5.0000	0.7713	166.2129	168.9481	X1 X2 X3 X4

(1) (5 points) How many observations are in the sample data?
72 observations

(2) (5 points) Write out the null and alternate hypotheses for the t-test for Beta1.
$\mathbf{H_0: \beta_1 = 0}$
$\mathbf{H_a: \beta_1 \neq 0}$

(3) (5 points) Compute the t- statistic for Beta1.
$\boldsymbol{t = \beta_1 / SE = 2.18604 / 0.41043 = 5.3262}$

Nate Bitting
Assignment 4

(4) (5 points) Compute the R-Squared value for Model 1.

$$R_2 = \frac{SS_R}{SS_T} = \frac{2126.0090}{2756.36857} = 0.7713$$

(5) (5 points) Compute the Adjusted R-Squared value for Model 1.

$$Adjusted\ R_2 = \frac{SS_R(n-p)}{SS_T(n-1)} = \frac{2126.0090(72-5)}{2756.36857(72-1)} = 0.7271$$

(6) (5 points) Write out the null and alternate hypotheses for the Overall F-test.

$\mathrm{H}_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$

$\mathrm{H}_a: \beta_j \neq 0$, for at least one value of j

(7) (5 points) Compute the F-statistic for the Overall F-test.

$F = MSM \,/\, MSE = 531.50226 \,/\, 9.40835 = 56.4926$

Nate Bitting
Assignment 4

Model 2: Now let's consider the following SAS output for an alternate regression model which we will refer to as Model 2.

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	6	2183.75946	363.95991	41.32	<.0001
Error	65	572.60911	8.80937		
Corrected Total	71	2756.36857			

Root MSE	2.96806	**R-Square**	0.7923
Dependent Mean	37.26901	**Adj R-Sq**	0.7731
Coeff Var	7.96388		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	14.39017	2.89157	4.98	<.0001
X1	1	1.97132	0.43653	4.52	<.0001
X2	1	9.13895	2.30071	3.97	0.0002
X3	1	0.56485	0.26266	2.15	0.0352
X4	1	0.33371	2.42131	0.14	0.8908
X5	1	1.90698	0.76459	2.49	0.0152
X6	1	-1.04330	0.64759	-1.61	0.1120

Number in Model	C(p)	R-Square	AIC	BIC	Variables in Model
6	7.0000	0.7923	163.2947	166.7792	X1 X2 X3 X4 X5 X6

(8) (5 points) Now let's consider Model 1 and Model 2 as a pair of models. Does Model 1 nest Model 2 or does Model 2 nest Model 1? Explain.

Two models are nested if one model contains all of the terms of the other and at least one additional term. Therefore, Model 1 is nested within Model 2.

(9) (5 points) Write out the null and alternate hypotheses for a nested F-test using Model 1 and Model 2.

$\mathbf{H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0}$

$\mathbf{H_a}$: $\boldsymbol{at\ least\ one\ \beta \neq 0}$

Nate Bitting
Assignment 4

(10) (5 points) Compute the F-statistic for a nested F-test using Model 1 and Model 2.

$$\mathbf{F} = \frac{(\mathbf{SSE_1} - \mathbf{SSE_2})/\,(\mathbf{p_2} - \mathbf{p_1})}{\mathbf{SSE_2}/(\mathbf{n} - \mathbf{p_2})} = \frac{(\mathbf{630.35953} - \mathbf{572.60911})/\mathbf{2}}{\mathbf{572.60911}/(\mathbf{72} - \mathbf{7})} = \mathbf{3.2778}$$

Here are some additional questions to help you understand other parts of the SAS output.

(11) (0 points) Compute the AIC values for both Model 1 and Model 2.

$$\mathbf{AIC} = \boldsymbol{n}\mathbf{ln}\left(\frac{\mathbf{SS_{Res}}}{\boldsymbol{n}}\right) + \mathbf{2}\boldsymbol{p}$$

Model 1: ***166.2129***

Model 2: ***163.2947***

Python code:

```
import math

ssr1 = 630.35953
ssr2 = 572.60911

def aic(ssr, n, p):
    return n * math.log(ssr/n) + 2 * p

print 'AIC of Model 1: ' + str(aic(ssr1, 72, 5))
print 'AIC of Model 2: ' + str(aic(ssr2, 72, 7))
```

(12) (0 points) Compute the BIC values for both Model 1 and Model 2.

$$\mathbf{BIC} = \boldsymbol{n}\mathbf{ln}\left(\frac{\mathbf{SS_{Res}}}{\boldsymbol{n}}\right) + \boldsymbol{p}\mathbf{ln}(\boldsymbol{n})$$

Model 1: ***177.5963***

Model 2: ***179.2313***

Python code:

```
import math

ssr1 = 630.35953
ssr2 = 572.60911

def bic(ssr, n, p):
    return n * math.log(ssr/n) + p * math.log(n)

print 'BIC of Model 1: ' + str(bic(ssr1, 72, 5))
print 'BIC of Model 2: ' + str(bic(ssr2, 72, 7))
```

Nate Bitting
Assignment 4

(13) (0 points) Compute the Mallow's Cp values for both Model 1 and Model 2.

$$\textbf{Mallow's } \mathbf{C_p} = \left(\frac{\mathbf{SSE}}{\boldsymbol{MSE}}\right) - \boldsymbol{n} + \mathbf{2}\boldsymbol{p}$$

Model 1: *5.0000*

Model 2: *3.0000*

Python code:

```
import math

ssr1 = 630.35953
mse1 = 9.40835
ssr2 = 572.60911
mse2 = 8.80937

def mallows_cp(sse, mse, n, p):
    return (sse/mse) - n + 2 * p

print "Mallow's Cp of Model 1: " + str(mallows_cp(ssr1, mse1, 72, 5))
print "Mallow's Cp of Model 2: " + str(mallows_cp(ssr2, mse2, 72, 5))
```

(14) (0 points) Verify the t-statistics for the remaining coefficients in Model 1.
I manually calculated the t-statistics for each variable in Model 1 and they are all correct.

(15) (0 points) Verify the Mean Square values for Model 1 and Model 2.
I manually calculated the MSE for both Model 1 and Model 2 and they are both correct.

(16) (0 points) Verify the Root MSE values for Model 1 and Model 2.
I manually calculated the Root MSE for both Model 1 and Model 2 and they are both correct.