Assignment #3: 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.

Students are expected to show all work in their computations. A good practice is to write down the generic formula for any computation and then fill in the values need for the computation from the problem statement.

Model 1: Let's consider the following statistical 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	56.49	<.0001
Error	67	630.35953	9.40835		
Corrected Total	71	2756.36857			

Root MSE	3.06730	**R-Square**	0.7713
Dependent Mean	37.26901	**Adj R-Sq**	0.7577
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	5.33	<.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	177.5963	X1 X2 X3 X4

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

Corrected Total = N – 1 = 71 + 1 = **72 observations**

(2) (5 points) Write out the null and alternate hypotheses for the t-test for Beta1.

Null Hypothesis (H0): Beta1 = 0
Alternate Hypothesis (Ha): Beta1 ≠ 0

(3) (5 points) Compute the t- statistic for Beta1.

t = Beta1/SE(Beta1) = 2.1860/0.41043 = **5.3261**

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

R2 = SSR/SST = 2126.0090/2756.3686 = **0.7713**

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

R2 Adjusted = $1 - \frac{SSE/(n-p)}{SST/(n-1)}$ = $1 - \frac{630.35953/(72-5)}{2756.36857/(72-1)}$ = 1-(9.408351/38.82209) = **0.7577**

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

Null Hypothesis (H0): Beta1 = Beta2 = Beta3 = Beta4

Alternate Hypothesis (Ha): Any Bi in i = 1,2,3,4 ≠ 0

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

F Overall = $\frac{SSR/k}{SSE/(n-p)}$ = $\frac{2126.00904/4}{630.35953/(72-5)}$ = 531.50226/9.408351 = **56.4926**

Model 2: Now let's consider the following statistical 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	179.2313	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.

Model 2 nests Model 1. This is because the predictor variables in Model 1 are a subset of the predictor variables in Model 2.

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

Null Hypothesis (H0): Beta5 = Beta6 = 0

Alternative Hypothesis (Ha): Any Bi in i = 5,6 ≠ 0

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

SS(Beta5, Beta6) = SS(FM)-SS(RM) = 2183.75946-2126.00904 = **57.7504**

SS(Beta5, Beta6)/df = 57.75047/2 = **28.8752**

MS(Error FM) = 572.60911/65 = **8.8094**

F Nested = 28.87521/8.80937 = **3.2778**

Double check: F Nested = $\frac{[SSE(RM)-SSE(FM)]/[\dim(FM)-\dim(RM)]}{SSE(FM)/[n-\dim(FM)]} = \frac{[630.35953-572.60911]/[7-5]}{572.60911/(72-7)} =$ 3.2778

Here are some additional questions to help you understand other parts of the statistical output. Completing these questions first will help you build confidence in your complications. The answers to these problems are available in the statistical output tables. You are simply verifying that you can reproduce the values.

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

Model 1: AIC = nln(SSE/n)+2p = 72ln(630.35953/72)+2(5) = 166.2129

Model 2: AIC = nln(SSE/n)+2p = 72ln(572.60911/72)+2(7) = 163.2947

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

Model 1: BIC = nln(SSE/n)+pln(n) = 72ln(630.35953/72)+5ln(72) = 177.5963

Model 2: BIC = nln(SSE/n)+pln(n) = 72ln(572.60911/72)+7ln(72) = 179.2313

(13) (0 points) Verify the t-statistics for the remaining coefficients in Model 1.

Intercept: 11.33027/1.99409 = 5.6819
Beta2: 8.27430/2.33906 = 3.5374
Beta3: 0.49182/0.26473 = 1.8578
Beta4: -0.49356/2.29431 = -0.2151

(14) (0 points) Verify the Mean Square values for Model 1 and Model 2.

Model 1 (Model): MS = SS/DF = 2126.00904/4 = 531.50226
Model 1 (Error): MS = SS/DF = 630.35953/67 = 9.408351
Model 2 (Model): 2183.75946/6 = 363.95991
Model 2 (Error): 572.60911/65 = 8.809371

(15) (0 points) Verify the Root MSE values for Model 1 and Model 2.

Model 1 Root MSE = $\sqrt{MSE}$ = $\sqrt{9.40835}$ = 3.06730
Model 2 Root MSE = $\sqrt{MSE}$ = $\sqrt{8.80937}$ = 2.96806