153

STA 6167 – Exam 1 – Spring 2016 – **PRINT** Name ANSWER KEY

Unless stated otherwise, for all significance tests, use $\alpha = 0.05$ significance level.

Q.1. A regression model was fit, relating estimated cost of de-commissioning oil platforms (Y, in millions of \$) to 2 predictors: Total number of piles/legs (X_1) and water depth (X_2, in 100s of feet). The model was fit, based on n = 17 oil platforms. Consider the models:

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2$ Model 2: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

ANOVA	Model 1			ANOVA	Model2	
	df	SS			df	SS
Regression	5	7397		Regression	2	7163
Residual	11	551		Residual	14	785
Total	16	7948		Total	16	7948
	Coefficients	andard Error			Coefficients	andard Errc
Intercept	-20.399	21.040		Intercept	-26.064	5.335
totpiles	-0.357	0.821		totpiles	0.957	0.274
wtrdph	6.155	7.048		wtrdph	4.401	1.095
tp2	0.047	0.046				
wtrdph2	-0.077	0.598				
tp*wd	-0.070	0.233				

p.1.a. Test whether the linear (main effects) model is appropriate. $H_0: \beta_{11} = \beta_{22} = \beta_{12} = 0$

8

$$F_{obs} = \frac{\frac{785-551}{14-11}}{\frac{551}{11}} = \frac{234(11)}{3(551)} = \frac{2574}{1653} \quad 1.557$$

$RR: F_{obs} \geq F_{.05,3,11} = 3.587$ 3

Do not reject H_0. 2

p.1.b. For each model, obtain the predicted value and residual for rig 17 (Y = 78.5, X_1 = 44, ~~X_2 = 13~~)

Model 1 $\hat{Y} = -20.399 - .357(44) + 6.155(13) + .047(44^2) - .077(13^2) - .070(44)(13)$ 4
$= 81.85 \Rightarrow e_1 = 78.5 - 81.85 = -3.35$ 2

Model 2: $\hat{Y} = -26.064 + .957(44) + 4.401(13) = 73.26$ 4
$\Rightarrow e_2 = 78.5 - 73.26 = 5.24$ 2

Model 1: Predicted = 81.85 Residual = -3.35 Model 2: Predicted = 73.26 Residual = 5.24

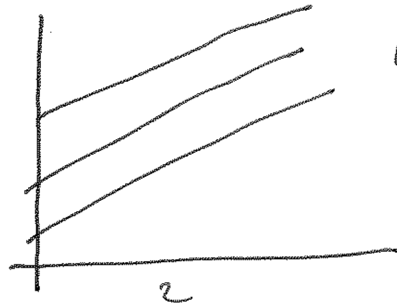
23

Q.2. A study related subsidence rate (Y) to water table depth (X_1) for 3 crops: pasture ($X_2 = 0$, $X_3 = 0$), truck crop ($X_2 = 1$, $X_3 = 0$), and sugarcane ($X_2 = 0$, $X_3 = 1$). Note the total sum of squares is TSS = 35.686, and n = 24.

p.2.a. Give a model that allows **separate intercepts** for each crop type, with a **common slope for water table depth** among crop types. Sketch the graph for this model. For this model, SSE = 1.853. Give the error degrees of freedom.

4 $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

3 $df_E = 24 - 4 = 20$



Parallel lines

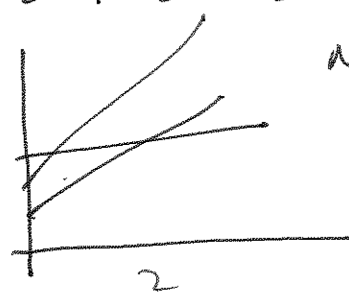
2

p.2.b. Give a model that allows **separate intercepts** for among crop types, with **separate slopes for water table depth** among crop types. Sketch the graph for this model. For this model, SSE = 1.261. Give the error degrees of freedom.

4 $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3$

$df_E = 24 - 6 = 18$

3



Non-Parallel lines

2

p.2.c. Test the null hypothesis that the (simpler) model in p.2.a. is appropriate. That is, the extra parameters in the second model are not significantly different from 0.

$$F_{obs} = \frac{\frac{1.853 - 1.261}{20 - 18}}{\frac{1.261}{18}} = \frac{18(0.592)}{2(1.261)} = 4.225$$

8

$RR: F_{obs} \geq F_{.05,2,18} = 3.555$

3

Reject Ho

2

p.2.d. For the model in p.2.a., compute R^2 $\quad 1 - \frac{1.853}{35.686} = 1 - .052 = .948$ 4

31

Q.3. A study investigated meteorological effects on condition of wheat yield in Ohio (Y), based on a series of n = 24 years. The predictors were: Average October/November temperature (X_1), September precipitation (X_2), October/November precipitation (X_3), and percent September sunshine (X_4). The best 1-,2-,3-, and 4-variable models (minimum SSE) are given below.

Model	p'	SSE	Cp	AIC	BIC
X3	2	1738	4.27	106.78	109.13
X2,X3	3	1531	3.37	105.73	109.27
X1,X2,X3	4	1395	3.48	105.50	110.21
X1,X2,X3,X4	5	1361	5.00	106.91	112.80

p.3.a. Complete the table. Use MSE of the full (X_1,X_2,X_3,X_4) models as the estimate of σ^2 when computing C_p

(5) $C_p = \frac{SSE}{MSE_c} + 2p' - n = \frac{1531}{71.63} + 2(3) - 24 = 3.37$

$\frac{1361}{19} = 71.63$

(5) $AIC = n \ln\left(\frac{SSE}{n}\right) + 2p' = 24 \ln\left(\frac{1361}{24}\right) + 2(5) =$

(5) $BIC = 24 \ln\left(\frac{1395}{24}\right) + \ln(24) * (4) = 110.21$

(2) (2) (2)

p.3.b. Give the best model based on each criteria. C_p X_1, X_2, X_3 AIC X_1, X_2, X_3 BIC X_3

p.3.c. The following output gives the regression coefficients for the X_1,X_2,X_3 model. Give the fitted value and residual for the first year (Y = 92, X_1 = 46, X_2 = 1.6, X_3 = 6.3).

	Coefficients
Intercept	16.59
tempon.x1	1.06
rains.x2	2.38
rainon.x3	2.96

$\hat{Y} = 16.59 + 1.06(46) + 2.38(1.6) + 2.96(6.3) = 87.8$

(4)

(2)

Fitted Value 87.8 Residual $92 - 87.8 = 4.2$

p.3.d. For the model in p.3.c., we obtain:

$\sum_{t=2}^{24}(e_t - e_{t-1})^2 = 3206 \qquad d_L(n=24, p=3) = 1.10 \quad d_U(n=24, p=3) = 1.66$

$DW = \frac{3206}{1395} = 2.30$

(4)

Test H_0: Errors are not autocorrelated versus H_A: Errors are autocorrelated

(2)

Circle the Best Answer Reject H_0 Accept H_0 Test is inconclusive

33

Q.4. A simple linear regression model was fit relating Weight (Y, in pounds) to Height (X, in inches) for a random sample of n=52 National Hockey League players. The total sum of squares, TSS = 9237, and $R^2 = 0.262$.

p.4.a. Complete the following ANOVA table.

ANOVA					
Source	*df*	*SS*	*MS*	*F*	*F(0.05)*
Regression	1 (2)	2420.0 (4)	2420.0 (3)	17.75 (3)	4.034 (3)
Residual	50 (2)	6816.7 (4)	136.3 (3)	#N/A	#N/A
Total	~~[illegible]~~ 51	9236.7	#N/A	#N/A	#N/A

$\frac{SSR}{TSS} = .262 \Rightarrow SSR = .262(9236.7) =$

p.4.b. Complete the following table and use it to conduct the F-test for Lack-of-Fit (there are c = 9 distinct heights).

$H_0: \mu_j = \beta_0 + \beta_1 X_j \quad j = 1,\ldots,c \qquad H_A: \mu_j \neq \beta_0 + \beta_1 X_j$

$SSLF = \sum_{j=1}^{c} n_j \left(\overline{Y}_j - \hat{Y}_j\right)^2 \qquad df_{LF} = c-2 \qquad SSPE = \sum_{j=1}^{c}(n_j - 1)S_j^2 \qquad df_{PE} = n-c$

Height	n	Y-bar	Y-hat	n*(YB-YH)^	(n-1)S^2
70	2	193.00	190.31	14.42	128.00
71	6	198.50	194.12	114.95	785.50
72	7	197.86	197.93	0.04	330.86
73	13	199.15	201.74	86.87	1357.69
74	11	202.27	205.55	117.91	1782.18
75	6	211.50	209.35	27.61	433.50
76	4	219.50	213.16	160.65	1451.00
77	2	216.50	216.97	0.44	24.50
78	1	222.00	220.78	~~[illegible]~~ 1.4	0.00
Sum	52	#N/A	#N/A	~~534.82~~ 525.29	6293.23

$F = \frac{SSLF/(c-2)}{SSPE/(n-c)} = \frac{\cancel{534.82}\ 525.29/7}{6293.23/43} = \frac{43(\cancel{534.82}\ 525.29)}{7(6293.23)} = \cancel{0.522}\ 0.513$ (8)

(3) 40 2.249

Test Statistic: ~~0.522~~ .513 Reject H_0 if Test Stat falls in Range: $> F_{.05,7,43} \approx 2.232$

50 2.199

35

Q.5. A simple linear regression model is fit, relating Orlando June Total Precipitation (Y) to Mean Temperature (X) over an n = 45 year period. The following table gives the results.

ANOVA

	df	*SS*
Regression	1	32.59982
Residual	43	489.2068
Total	44	521.8067

	Coefficients	*andard Errc*
Intercept	61.3280	31.8335
meanTemp	-0.6626	0.3915

p.5.a. Use the t-test to test H_0: $\beta_1 = 0$ vs H_A: $\beta_1 \neq 0$ at $\alpha = 0.10$ significance level

4

Test Statistic: $t = \frac{-.6626}{.3915} = -1.692$

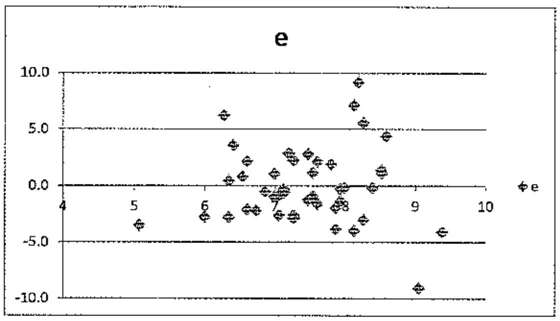
3

1.684

Reject H_0 if Test Stat falls in range: $|t| \geq t_{.05,43} \approx 1.681$

1.676

A plot of the residuals displays a possible case of unequal variances. The regression of the squared residuals on X is given below:



	df	*SS*
Regression	**1**	**1326.60**
Residual	**43**	**14436.63**
Total	**44**	**15763.24**

p.5.b. Conduct the Breusch-Pagan test to test H_0: Equal Variances H_A: Variance is related to X. (Use $\alpha = 0.05$):

7

$$X^2_{BP} = \frac{SSReg^*/2}{(SSE/n)^2} = \frac{1326.60/2}{\left(\frac{489.2068}{45}\right)^2} = \frac{663.30}{118.18} = 5.61$$

3

Test Statistic: $X^2_{BP} = 5.61$ Reject H_0 if Test Statistic falls in the Range: $\geq X^2_{.05,1} = 3.841$

17

Q.6. A multiple regression model is fit with 3 predictors and an intercept, based on a sample of n=25 observations. How large must $R^2/(1-R^2)$ be to reject H_0:$\beta_1 = \beta_2 = \beta_3 = 0$?

8

$$F = \frac{R^2/p}{1-R^2/(n-p')} = \frac{R^2}{1-R^2}\left(\frac{n-p'}{p}\right) \geq F_{.05,3,21} = 3.072$$

$$\Rightarrow \frac{R^2}{1-R^2}\left(\frac{21}{3}\right) \geq 3.072 \Rightarrow \frac{R^2}{1-R^2} \geq \frac{3.072}{7} = .4389$$

Q.7. It is possible to fail to reject H_{01}: $\beta_1 = 0$ and H_{02}: $\beta_2 = 0$ based on t-tests in multiple linear regression model with p > 2 predictors, but still reject H_{012}: $\beta_1 = \beta_2 = 0$, controlling for the remaining p-2 predictors. **True** or **False**

4

12