STA 6167 – Exam 1 – Spring 2016 – **PRINT** Name ______________________

Unless stated otherwise, for all significance tests, use α = 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		7397		Regression		7163
Residual		551		Residual		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: $\beta_{11} = \beta_{22} = \beta_{12} = 0$

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: Predicted = _________ Residual = _________ Model 2: Predicted = _________ Residual = _________

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.

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.

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.

p.2.d. For the model in p.2.a., compute R^2

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		105.73	109.27
X1,X2,X3	4	1395	3.48	105.50	
X1,X2,X3,X4	5	1361	5.00		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

p.3.b. Give the best model based on each criteria. C_p _______________ AIC _____________ BIC _____________

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**

Fitted Value ____ Residual ______________________________________

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

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

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

Circle the Best Answer Reject H_0 Accept H_0 Test is inconclusive

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					
Residual				**#N/A**	**#N/A**
Total		**9236.7**	**#N/A**	**#N/A**	**#N/A**

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**		**0.00**
Sum	**52**	**#N/A**	**#N/A**		

Test Statistic: ________________________ Reject H_0 if Test Stat falls in Range: __________________________

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

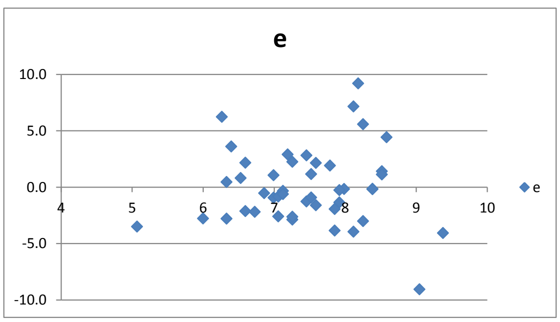
	Coefficients	*andard Err*
Intercept	61.3280	31.8335
meanTemp	-0.6626	0.3915

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

Test Statistic: ____

Reject H_0 if Test Stat falls in range: ______________________________

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 $\alpha = 0.05$):

Test Statistic: _____________________ Reject H_0 if Test Statistic falls in the Range: _____________________

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 $H_0: \beta_1 = \beta_2 = \beta_3 = 0$?

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**