STA 6167 – Exam 1 – Spring 2017 – **PRINT** Name ________________________

For all significance tests, use α = 0.05 significance level.

Q.1. A multiple linear regression model is fit relating a dependent variable to a set of 3 numeric predictor variables, based on a sample on n=20 experimental units. How large does R^2 need to be so that the null hypothesis $H_0:\beta_1=\beta_2=\beta_3=0$ will be rejected?

Q.2. An experiment is conducted with 3 numeric predictors and 2 categorical predictors, one with 3 levels, the other with 2 levels. There are no interaction or polynomial terms in the model, and the sample size is n = 30. Give the degrees of freedom for Regression and Error.

Df_{Reg} = ________________________ Df_{Error} = ______________________________

Q.3. It is possible for a dataset to reject H_0: $\beta_1 = \beta_2 = ... = \beta_p = 0$ based on the F-test, but fail to reject H_0: $\beta_i = 0$ for i=1,…,p based on the individual t-tests.

True or **False**

Q.4. A linear regression model is fit, relating salary (Y) to experience (X_1), gender (X_2=1 if female, 0 if male) and an experience/gender interaction term to employees in a large law firm. The fitted equation is

$\hat{Y} = 50000 + 2000X_1 + 1000X_2 - 100X_1X_2$. Give the predicted salaries for the following groups of individuals.

Males with 0 experience ________________________ Females with 0 experience ________________________

Males with 10 experience ________________________ Females with 10 experience ________________________

Q.5. A regression model was fit, relating blood alcohol elimination rate measurements (Y, in grams/litre/hour) to Gender (X_1=1 if female, 0 if male), breath alcohol elimination measurements (X_2 in mg/l/h) and a gender/breath interaction term. The sample was 59 adult Austrians. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$

p.5.a. Complete the following Analysis of Variance Table and test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

ANOVA					
	df	*SS*	*MS*	*F*	*F(.05)*
Regression		0.0478			
Residual				#N/A	#N/A
Total		0.0624	#N/A	#N/A	#N/A

p.5.b. Is the P-value for the test **Larger** or **Smaller** than 0.05?

p.5.c. What proportion of the variation in Blood alcohol elimination rate measurements is "explained" by the model?

p.5.d. The regression coefficient estimates are given below. Test $H_0 : \beta_i = 0 \quad H_A : \beta_i \neq 0$ for each coefficient.

	Coefficients	*Standard Error*	t_obs	*t(.025)*	*Reject H0?*
Intercept	0.0427	0.0154	#N/A	#N/A	#N/A
female	-0.0335	0.0229			
breath	1.5349	0.1951			
f*breath	0.4213	0.2744			

Q.6. Monthly mean temperatures for Boston (Y, in Fahrenheit) for the years 1920-2014 are fit using a linear regression model to Year (X_1=Year-1920) and 11 monthly dummy variables ($X_2 = 1$ if January, 0 otherwise,…, $X_{12} = 1$ if November, 0 otherwise, Note that December is the reference month). The ANOVA table and regression coefficient estimates are given below.

ANOVA					
	df	*SS*	*MS*	*F*	*gnificance*
Regression	12	270975.961	22581.330	2842.345	0.000
Residual	1127	8953.578	7.945		
Total	1139	279929.539			

	Coefficients	*tandard Errc*	*t Stat*	*P-value*
Intercept	33.343	0.323	103.344	0.000
year	0.013	0.003	4.406	0.000
month1	-4.563	0.409	-11.158	0.000
month2	-3.285	0.409	-8.033	0.000
month3	4.312	0.409	10.543	0.000
month4	14.171	0.409	34.649	0.000
month5	24.313	0.409	59.449	0.000
month6	33.787	0.409	82.616	0.000
month7	39.412	0.409	96.368	0.000
month8	37.906	0.409	92.688	0.000
month9	30.806	0.409	75.327	0.000
month10	20.758	0.409	50.757	0.000
month11	10.793	0.409	26.390	0.000

p.6.a. Give the predicted temperatures for December 1920, June (Month 6) 1920, December 2010, and June 2010.

	1920	2010
December		
June		

p.6.b. Compute a 95% Confidence Interval for the change in annual mean temperature, controlling for month.

Lower Bound: ___________________________ Upper Bound: ___________________________

p.6.c. Compute the Durbin-Watson statistic. $\sum_{t=2}^{1140}\left(e_t - e_{t-1}\right)^2 = 14094.8$

DW = ______________________

p.6.d. What proportion of the variation in temperature is explained by the model?

Q.7. A response surface model is fit, relating potato chip moistness (Y) to 3 factors: drying time (X_1), frying temperature (X_2), and frying time (X_3). There were n = 20 experimental runs (observations). The following 3 models were fit:

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad SSR_1 = 475.2 \quad SSE_1 = 145.2$

Model 2: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 \quad SSR_2 = 558.3 \quad SSE_2 = 62.1$

Model 3: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 \quad SSR_3 = 599.0 \quad SSE_3 = 21.4$

p.7.a. Test whether any of the 2-way interaction effects are significantly different from 0 , controlling for main effects.

H_0:

Test Statistic: ____________________ Rejection Region: _____________________ P-value: > or < 0.05

p.7.b. Test whether any of the quadratic effects are significantly different from 0, controlling for main effects and 2-factor interactions.

H_0:

Test Statistic: ____________________ Rejection Region: _____________________ P-value: > or < 0.05

Q.8. A regression model was fit, relating Price (Y, in \$1000s) to acceleration rate (X_1) and Miles per gallon (X_2) for a sample of n = 25 models of hybrid compact cars. The fitted equation and summary model statistics are given below.

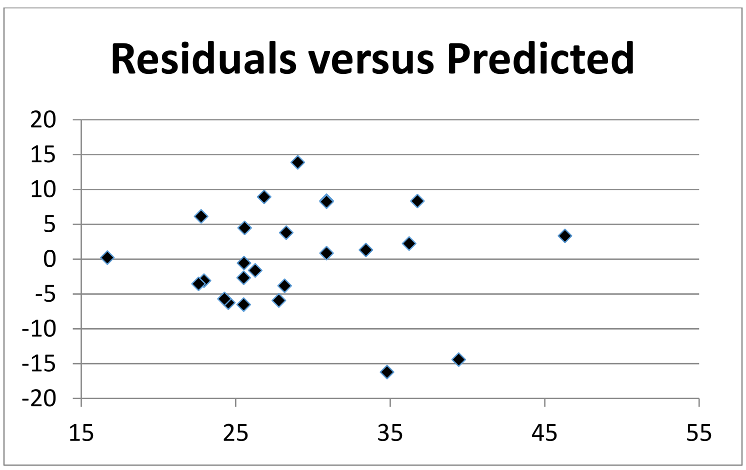
$$\hat{Y} = -26.35 + 4.50X_1 + 0.20X_2 \quad SSR = 957 \quad SSE = 1239$$

p.8.a. Test whether price is related to either acceleration rate and/or Miles per gallon. H_0: $\beta_1 = \beta_2 = 0$.

Test Statistic: ____________________ Rejection Region: _____________________ P-value: > or < 0.05

p.8.b. A plot of the residuals versus predicted values suggests a possible non-constant error variance. A regression of the squared residuals on X_1 and X_2 yields SS(Reg*) = 3786261. Test:

H_0: Equal Variance Among Errors $\sigma^2\{\varepsilon_i\} = \sigma^2 \ \forall i$ $\quad H_A$: Unequal Variance Among Errors $\sigma_i^2 = \sigma^2 h(\gamma_1 X_{i1} + \gamma_2 X_{i2})$



Test Statistic: ____________________ Rejection Region: _____________________ P-value: > or < 0.05

Q.9. A regression model is fit, relating energy consumption (Y) to 3 predictors: area (X_1), age (X_2), and effective number of guest rooms (X_3 = rooms*occupancy rate) for a sample of n = 15 hotel rooms.

p.9.a. Complete the following table of C_p, AIC, and BIC for all possible regressions involving X_1, X_2, and X_3.

Model	p*	SSE	C_p	AIC	BIC
X1		75.13		30.12	32.01
X2		327.08	57.31	58.07	59.96
X3		187.85	26.53	47.53	49.42
X1,X2		70.84	2.66		33.84
X1,X3		71.04	2.71	31.06	33.89
X2,X3		186.24	28.18	49.37	52.20
X1,X2,X3		67.85	4.00	32.18	

p.9.b. Which model is selected based on each criteria?

C_p: ____________________________ AIC: _______________________ BIC: ___________________________

p.9.c. To check for issues of multicollinearity, a regression relating each predictor on the other 2 predictors is fit. The largest R^2 of the 3 regressions is when X_1 is regressed on X_2 and X_3. That R_1^2 value is 0.468. Compute the Variance Inflation Factor VIF for X_1, where $VIF_1 = 1/\left(1-R_i^2\right)$. Does it exceed 10?