STA 6207 – Practice Problems – Multiple Regression

Part A: Estimating and Testing

QA.1. You obtain the following partial output from a regression program. Fill in all missing parts.

X'X				X'Y
9.0000	19.8602	26.0985		34.4348
19.8602	45.6772	57.5914		76.4129
26.0985	57.5914	77.5334		100.3257
INV(X'X)				
7.2817	-1.1916	-1.5660		
-1.1916	0.5400	0.0000		
-1.5660	0.0000	0.5400		

Regression Statistics					
R Square	(a)				
Standard Error	0.0621				
Observations	(b)				
ANOVA					
	df	*SS*	*MS*	*F*	*F(.05)*
Regression	(c)	0.2173	(d)	(e)	(f)
Residual	6	0.0232	0.0039		
Total	8	0.2405			
	Coefficients	*Standard Error*	*t Stat*	*t(.025)*	
Intercept	2.5823	0.1676	15.4048	(k)	
FL*	(g)	0.0456	(i)	(k)	
FC*	0.2540	(h)	(j)	(k)	

p.1.a. R^2 = ______________________ p.1.b. n = ____________ p.1.c. df_{Reg} = ________________

p.1.d. MS(Regression) = ______________________ p.1.e. F_{obs} = __________________________________

p.1.f. Critical F-value (α = 0.05) = ____________ p.1.g. $\hat{\beta}_1$ = ____

p.1.h. $s\left\{\hat{\beta}_2\right\}$ = ______________________ p.1.i. t-stat for H_0: $\beta_1 = 0$: vs H_A: $\beta_1 \neq 0$: _____________________

p.1.j. t-stat for H_0: $\beta_2 = 0$: vs H_A: $\beta_2 \neq 0$: ___________________ p.1.k. Critical t-value (α = 0.05) ____________

QA.2. A multiple linear regression model is fit, relating height (Y, mm) to hand length (X_1, mm) and foot length (X_2, mm), for a sample of n = 20 adult males. The following partial computer output is obtained, for model 1 with 2 predictors.

ANOVA					
	df	*SS*	*MS*	*F*	*F(0.05)*
Regression		37497			
Residual		19772		#N/A	#N/A
Total		57269	#N/A	#N/A	#N/A
	Coefficients	*Standard Err*	*t Stat*	*P-value*	
Intercept	1055.78	132.86	7.95	0.0000	
Hand	1.26	0.55	2.28	0.0357	
Foot	1.71	0.39	4.42	0.0004	

p.2.a Complete the table. Do you reject the null hypothesis H_0: $\beta_1 = \beta_2 = 0$? **Yes** or **No**

p.2.b. Give the predicted height of a man with a hand length of 210mm and a foot length of (260mm). Just give the point estimate, not confidence interval for the mean or a prediction interval.

p.2.c. Give an unbiased estimate of the error variance σ^2

p.2.d. The coefficient of determination represents the proportion of variation in heights “explained” by the model with hand and foot length as predictors. What is the proportion explained for this model?

QA.3. For the Analysis of Variance in model 2, with *n* observations and *p* predictors, complete the following parts.

p.3.a. Write the Regression and Residual sums of squares as quadratic forms.

p.3.b. Derive the distributions of SSRegression/σ^2 and SSResidual/σ^2

p.3.c. Show that SSRegression/σ^2 and SSResidual/σ^2 are independent

p.3.d. What is the sampling distribution of MSRegression/MSResidual when $\beta_1 = \ldots = \beta_p = 0$?

QA.4. A multiple regression model is fit, based on Model 2, with p predictors and an intercept. Define the projection matrix as: $\mathbf{P} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$, where $\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}$ Define $\frac{1}{n}\mathbf{J}_n = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$ where $\mathbf{J}_n$ is $n \times n$

p.4.a. Show that **P** and $\frac{1}{n}\mathbf{J}_n$ are symmetric and idempotent. (Hint: **(X'X)**$^{-1}$ is symmetric). **SHOW ALL WORK**.

p.4.b. Obtain $\mathbf{P}\frac{1}{n}\mathbf{J}_n$ and show that $\mathbf{P}-\frac{1}{n}\mathbf{J}_n$ is idempotent. **SHOW ALL WORK**

p.4.c. Obtain the rank of **P**, $\frac{1}{n}\mathbf{J}_n$, and $\mathbf{P}-\frac{1}{n}\mathbf{J}_n$ **SHOW ALL WORK**

p.4.d. What is the sampling distribution of $\frac{1}{\sigma^2}\mathbf{Y}'\left(\mathbf{P}-\frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}$? **SHOW ALL WORK**

p.4.e. Show that $\frac{1}{\sigma^2}\mathbf{Y}'\left(\mathbf{P}-\frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}$ and $\frac{1}{\sigma^2}\mathbf{Y}'(\mathbf{I}-\mathbf{P})\mathbf{Y}$ are independent. **SHOW ALL WORK**.

QA.5. Use the following output to obtain the quantities given below:

X				Y
1	0	2		4
1	5	2		6
1	10	2		9
1	0	8		7
1	5	8		10
1	10	8		12

(X'X)^-1		
0.8796	-0.0500	-0.0926
-0.0500	0.0100	0.0000
-0.0926	0.0000	0.0185

X'Y
48
290
270

Beta-hat
2.7222
0.5000
0.5556

P					
0.5833	0.3333	0.0833	0.2500	0.0000	-0.2500
0.3333	0.3333	0.3333	0.0000	0.0000	0.0000
0.0833	0.3333	0.5833	-0.2500	0.0000	0.2500
0.2500	0.0000	-0.2500	0.5833	0.3333	0.0833
0.0000	0.0000	0.0000	0.3333	0.3333	0.3333
-0.2500	0.0000	0.2500	0.0833	0.3333	0.5833

Y'Y	426.00
Y'PY	425.67
Y'(I-P)Y	0.33
Y'(J/n)Y	384.00
Y'(P-J/n)Y	41.67

Total Corrected: Sum Of Squares Degrees of Freedom

Regression: Sum Of Squares Degrees of Freedom

Residual: Sum Of Squares Degrees of Freedom

S^2 $\quad s\{\hat{\beta}_1\}$

Testing H_0: $\beta_1 = \beta_2 = 0$ F-stat Num df Den df

Predicted Value for Y_2 $\quad s\{\hat{Y}_2\}$ $\quad s\{e_2\}$

QA.6. Use the following output to obtain the quantities given below:

X		
1	0	10
1	2	10
1	4	10
1	0	20
1	2	20
1	4	20

Y
10
16
22
22
25
25

(X'X)^-1		
1.9167	-0.1250	-0.1000
-0.1250	0.0625	0.0000
-0.1000	0.0000	0.0067

X'Y		Beta-hat
120		4.250
270		1.875
1920		0.800

P					
0.5833	0.3333	0.0833	0.2500	0.0000	-0.2500
0.3333	0.3333	0.3333	0.0000	0.0000	0.0000
0.0833	0.3333	0.5833	-0.2500	0.0000	0.2500
0.2500	0.0000	-0.2500	0.5833	0.3333	0.0833
0.0000	0.0000	0.0000	0.3333	0.3333	0.3333
-0.2500	0.0000	0.2500	0.0833	0.3333	0.5833

Y'Y	2574.00
Y'PY	2552.25
Y'(J/n)Y	2400.00

Complete the following elements of the regression model:

X'X		
	40	
		90

ANOVA		
Source	df	SS
Total (Uncorr)		
Model		
Mu		
Regression		
Residual		
Total (Corr)		

S^2 = ________________ $\quad s\{\beta_2\} =$ ________________

Tests of (TS=Test statistic, RR=Rejection Region) each based on α = 0.05 significance level:

$H_0: \beta_2 = 0 \quad H_A: \beta_2 \neq 0$ TS: ____________ RR:________

$H_0: \beta_1 = \beta_2 = 0 \quad H_A: \beta_1 \neq 0$ and/or $\beta_2 \neq 0$ TS: ____________ RR:___________

Predicted Value for Y_4 based on each of these two forms and its residual (show work):

$\mathbf{X\beta}: Y_4 =$ ______________________________ $\mathbf{PY}: Y_4 =$ _________________

$e_4 =$ _______________________

$s\left\{Y_4\right\} =$ ____________________________ $s\left\{e_4\right\} =$ ______________________

QA.7. A large electronics retailer is interested in the relationship between net revenue of plasma TV sales (Y, \$1000s) , and the following 4 predictors: X_1= shipping costs (\$/unit), X_2= print advertising (\$1000s), X_3= electronic media ads (\$1000s), and X_4= rebate rate (% of retail price). A sample of n=50 stores is selected and the resulting (partial) regression output is obtained:

ANOVA

	df	*SS*	*MS*	*F*	*F(0.05)*
Regression		259411.8			
Residual		224539.0			
Total	49	483950.8			

	Coefficients	*andard Err*	*t Stat*	*P-value*
Intercept	4.31	70.82	0.0608	0.9518
ShipCost	-0.08	4.68	-0.0175	0.9861
PrintAds	2.27	1.05	2.1562	0.0364
WebAds	2.50	0.85	2.9535	0.0050
Rebate%	16.70	3.57	4.6766	0.0000

INV(X'X)

1.005224	-0.029489	-0.006808	-0.002514	-0.019146
-0.029489	0.004386	-0.000011	-0.000282	0.000021
-0.006808	-0.000011	0.000221	-0.000031	-0.000228
-0.002514	-0.000282	-0.000031	0.000143	-0.000002
-0.019146	0.000021	-0.000228	-0.000002	0.002555

p.7.a. Complete the ANOVA table.

p.7.b. Give the prediction for net revenue, when ShipCost=10, PrintAds=50, WebAds=40, Rebate%=15.

p.7.c. Controlling for all other factors, give a 95% confidence interval for the change in expected net revenue ($1000s) when Rebate% is increased by 1.

p.7.d. Test H_0: $\beta_{PrintAds} - \beta_{WebAds} = 0$ vs H_A: $\beta_{PrintAds} - \beta_{WebAds} \neq 0$ at $\alpha = 0.05$ significance level:

p.7.d.i. Test Statistic: p.7.d.ii. Rejection Region

p.7.e. What proportion of variation in revenues is "explained" by the regression model?

QA.8. You obtain the following partial output from a regression program. Fill in all missing parts.

X'X				X'Y
9.0000	19.8602	26.0985		34.4348
19.8602	45.6772	57.5914		76.4129
26.0985	57.5914	77.5334		100.3257
INV(X'X)				
7.2817	-1.1916	-1.5660		
-1.1916	0.5400	0.0000		
-1.5660	0.0000	0.5400		

Regression Statistics					
R Square	(a)				
Standard Error	0.0621				
Observations	(b)				
ANOVA					
	df	*SS*	*MS*	*F*	*F(.05)*
Regression	(c)	0.2173	(d)	(e)	(f)
Residual	6	0.0232	0.0039		
Total	8	0.2405			
	Coefficients	*Standard Error*	*t Stat*	*t(.025)*	
Intercept	2.5823	0.1676	15.4048	(k)	
FL*	(g)	0.0456	(i)	(k)	
FC*	0.2540	(h)	(j)	(k)	

p.8.a. R^2 = ______________________ p.8.b. n = ____________ p.8.c. df_{Reg} = ________________

p.8.d. MS(Regression) = ______________________ p.8.e. F_{obs} = ___________________________________

p.8.f. Critical F-value ($\alpha = 0.05$) = ____________ p.8.g. $\hat{\beta}_1 =$ ____

p.8.h. $s\left\{\hat{\beta}_2\right\} =$ _______________________ p.8.i. t-stat for H_0: $\beta_1 = 0$: vs H_A: $\beta_1 \neq 0$: _____________________

p.8.j. t-stat for H_0: $\beta_2 = 0$: vs H_A: $\beta_2 \neq 0$: ___________________ p.8.k. Critical t-value ($\alpha = 0.05$) ____________

QA.9. A linear regression model is fit, relating mean January temperatures (Y, in °F) to Elevation (X_1, in 100s of feet) and Latitude (X_2, in degrees north latitude) for a random sample of n = 63 weather stations in Texas. The (partial) computer results are given below.

ANOVA					
	df	*SS*	*MS*	*F*	*F(.05)*
Regression		**2472.2**			
Residual		**91.9**		**#N/A**	**#N/A**
Total		**2564.1**	**#N/A**	**#N/A**	**#N/A**
	Coefficients	*Standard Erro*	*t Stat*	*P-value*	
Intercept	**115.906**	**2.478**	**46.776**	**0.0000**	
ELEV.C	**-0.117**	**0.013**	**-8.877**	**0.0000**	
LAT	**-2.183**	**0.082**	**-26.580**	**0.0000**	

p.9.a. Complete the ANOVA table.

p.9.b. Dallas/Fort Worth International Airport (DFW) was not in the sampled locations, and is located at an elevation of $X_1 = 5.6$ and a latitude of $X_2 = 32.9$. Give the predicted value for DFW.

p.9.c. For DFW, we obtain the following values: $x_0 = \begin{bmatrix} 1 \\ 5.6 \\ 32.9 \end{bmatrix}$ $\qquad x_0'\left(\mathbf{X'X}\right)^{-1}x_0 = 0.0454$.

Compute the 95% **Prediction Interval** for DFW's mean January temperature.

QA.10. For the multiple regression model, with an intercept term, complete the following parts, **show all of your work.**

p.10.a. $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e} \quad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad \overline{\mathbf{Y}} = \frac{1}{n}\mathbf{J}_n\mathbf{Y}$ prove that $\left(\mathbf{Y}-\overline{\mathbf{Y}}\right)'\left(\mathbf{Y}-\overline{\mathbf{Y}}\right) = \left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right)'\left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right) + \mathbf{e}'\mathbf{e}$

p.10.b. Derive the sampling distributions of $\hat{\mathbf{Y}}-\overline{\mathbf{Y}}$ and $\mathbf{e}$

QA.11. A linear regression model was fit, relating weekly number of passengers (Y, in 10000s) to number of street cars in operation (X_1, in 100s) and number of miles street cars ran (X_2, in 10000 miles) over a period of n=20 consecutive weeks. The following EXCEL spreadsheet summarizes the model.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \varepsilon_t \qquad \varepsilon_t \sim NID\left(0, \sigma^2\right)$$

p.11.a. Complete the sheet.

Note: VIF = Variance Inflation Factor and DW = Durbin-Watson Statistic for Autocorrelation

X'X				X'Y			
20.00	25.89	17.81		76.27			
25.89	37.31	26.46		112.98			
17.81	26.46	19.21		81.06			
INV(X'X)				Beta-hat	SE{B-hat}	t	t(.025)
0.7614	-1.2257	0.9829		-0.73			
-1.2257	3.1660	-3.2261		2.69			
0.9829	-3.2261	3.5860		1.18			
Y'Y	Y'J/nY	Y'PY					
349.0592	290.8923	344.7835					
SSE	MSE	SSR	MSR	F	F(.05)		
r(X1,X2)	VIF	DW_Num	DW_Den	DW			
0.957434		4.906587					

p.11.b. The critical values for the Durbin-Watson test with n=20 and p=2 are: d_L = 1.10 and d_U = 1.54. Does the assumption of uncorrelated errors seem reasonable? **Yes** or **No**

p.11.c. Based on the VIF, is there evidence of serious multicollinearity? **Yes** or **No**

Part B: General Linear Hypothesis Tests

QB.1. A forensic study related Hand (X_1) and Foot (X_2) lengths to Stature (Y) for a sample of n = 75 adult females (each variable in 100s of mms). Consider the following three models.

$$M_1 : E\{Y\} = \beta_0 + \beta_1 X_1 \qquad M_2 : E\{Y\} = \beta_0 + \beta_2 X_2 \qquad M_3 : E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Model 1 (Hand)					Model 2 (Foot)			
X'X			X'Y		X'X			X'Y
75	142.185		1199.70		143450	6686.317		45568.10
142.185	270.1992		2276.80		6686.317	414.390402		2819.37
INV(X'X)			beta-hat		INV(X'X)			beta-hat
5.586829	-2.93992		8.9042		2.81E-05	-0.0004537		0.0022
-2.93992	1.550753		3.7408		-0.00045	0.009733694		6.7688
Y'Y		Model 3 (Hand,Foot)						
19208.28		X'X				X'Y		
		75	142.185	176.062		1199.70		
		142.185	270.1992	334.2884		2276.80		
		176.062	334.2884	414.3904		2819.37		
		INV(X'X)				beta-hat		
		6.640061	-1.95728	-1.24223		7.4414		
		-1.95728	2.467531	-1.15897		2.3760		
		-1.24223	-1.15897	1.465136		1.7253		

p.1.a. Compute $\mathbf{Y}'\left(\frac{1}{n}\mathbf{J}\right)\mathbf{Y}$ and the Total (Corrected) Sum of Squares.

p.1.b. Compute the Residual (Error) Sum of Squares for each model.

p.1.c. Compute $R(\beta_1 | \beta_0), R(\beta_2 | \beta_0), R(\beta_1 | \beta_0, \beta_2), R(\beta_2 | \beta_0, \beta_1)$

p.1.d. Use the general linear test for Model 3 to test $H_0 : \beta_1 = \beta_2 \quad \text{vs} \quad H_A : \beta_1 \neq \beta_2$

QB.2. A firm has 2 types of expenditures that can varied in their marketing plan: advertising and in-store promotion. A regression model is fit, relating Y=weekly sales to levels of these expense variables (X_1=advertising, X_2=in-store promotion). The model fit is: $E(Y) = \beta_0+\beta_1X_1+\beta_2X_2$. Set up the **K’** matrix and **m** vector for testing: (a) whether mean sales are 500 when no advertising or in-store promotion is conducted, and (b) the effects of increasing X_1 and X_2 by 1 unit have the same effect on mean sales. That is, H_0^A: β_0=500 $\quad$ H_0^B: $\beta_1=\beta_2$.

QB.3. A marketing department is interested in the effects of changing advertising levels for television and internet on sales. They vary X_1=TV ad \$, and X_2=internet ad \$ and obtain the following regression results:

X'X				**X'Y**
20	**416.5343**	**406.487**		**3676.373**
416.5343	**9546.826**	**8733.245**		**78940.14**
406.487	**8733.245**	**9111.308**		**77022.41**
(X'X)^(-1)				**betahat**
0.800494	**-0.01832**	**-0.01815**		**98.54071**
-0.01832	**0.00127**	**-0.0004**		**2.093241**
-0.01815	**-0.0004**	**0.001303**		**2.050869**
SS(Resid)				
608.6247				

Give the analysis of variance.

Set up and conduct the general linear test that the effects of changing each type of advertising are equal in terms of sales at the α=0.05 significance level.

QB.4. A researcher fits a **simple linear regression** model, relating yield of a chemical process to temperature when all inputs beside temperature are at a specific level. She wishes to test the following two hypotheses simultaneously (the temperature range the experiment was conducted was: 55ºF – 85ºF):

- The average yield increases by 2 units when temperature increases by 1ºF
- The average yield is 400 when the temperature is set to 70ºF

p.4.a. For model 2, fill in the following matrix and vectors that she is testing (this is her null hypothesis):

$$H_0 : K'\beta - m = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} \begin{bmatrix} \quad \\ \quad \end{bmatrix} - \begin{bmatrix} \quad \\ \quad \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}$$

p.4.b. She obtains the following results from fitting the regression based on *n* = 18 measurements while conducting the experiment:

$$\left(K'\beta-m\right)'\left(K'\left(X'X\right)^{-1}K\right)^{-1}\left(K'\beta-m\right)=640 \quad Y'(I-P)Y=3200$$

p.4.c. Conduct her test at the α = 0.05 significance level.

- Test Statistic:
- Reject H_0 if the Test Statistic falls in the range: ______________________________________

QB.5. A researcher fits a **multiple linear regression** model, relating yield (Y) of a chemical process to temperature (X_1), and the amounts of 2 additives (X_2 and X_3, respectively). She fits the following model:

$$E(Y)=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3$$

She wishes to test the following three hypotheses simultaneously:

- The mean response when X_1=70, X_2=10, X_3=10 is 80
- The average yield increases by 4 units when temperature increases by 1°F, controlling for X_2 and X_3
- The partial effect of increasing each additive is the same (controlling for all other factors)

p.5.a. Fill in the following matrix and vectors that she is testing (this is her null hypothesis):

$$H_0: K'\beta-m=\begin{bmatrix}0\\0\\0\end{bmatrix}\Rightarrow\begin{bmatrix} & & & \\ & & & \\ & & & \end{bmatrix}\begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}-\begin{bmatrix} \\ \\ \\ \end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

p.5.b. She obtains the following results from fitting the regression based on *n* = 24 measurements while conducting the experiment:

$$\left(K'\beta-m\right)'\left(K'\left(X'X\right)^{-1}K\right)^{-1}\left(K'\beta-m\right)=1800 \quad Y'(I-P)Y=7800$$

p.5.c. Conduct her test at the α = 0.05 significance level.

- Test Statistic:

Reject H_0 if the Test Statistic falls in the range: ____________________________________

QB.6. A research firm is interested in the effects of 4 types of advertising (**T**elevision, **R**adio, **N**ewspaper, and **I**nternet) on a firm's sales. They hold all other variables constant over the study period (such as price and store promotion). The sample is based on n=30 sales periods. They fit the following 2 regressions based on Model 1 (note that SS(Total Corrected)=5000):

$$\text{Model1: } E(Y) = \beta_0 + \beta_T T + \beta_R R + \beta_N N + \beta_I I \qquad SS(\text{Re}\, g_1) = 4000$$
$$\text{Model2: } E(Y) = \beta_0 + \beta_A A \qquad A = T + R + N + I \qquad SS(\text{Re}\, g_2) = 3700$$

p.6.a. Test H_0: $\beta_T = \beta_R = \beta_N = \beta_I = 0$ at α = 0.05 significance level.

Test Statistic ____ Rejection Region ___________________

p.6.b. Set up the test of H_0: $\beta_T = \beta_R = \beta_N = \beta_I$ in the form of a general linear test by giving K', β, and m, and the degrees of freedom. Note that there are several ways K' can be formed.

p.6.c. Test H_0: $\beta_T = \beta_R = \beta_N = \beta_I$ at α = 0.05 significance level.

Test Statistic ____ Rejection Region ___________________

QB.7. A study was conducted, relating female heights (Y, in 100s of mm) to hand length (X_1, in 100s of mm) and foot length (X_2 in 100s of mm), based on a sample of n = 15 adult females. The following model was fit, with matrix results given below.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad \varepsilon \sim NID\left(0, \sigma^2\right) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ We wish to test $H_0: \beta_1 = \beta_2 \quad H_A: \beta_1 \neq \beta_2$

X'X				**X'Y**
15	**28.494**	**35.212**		**239.227**
28.494	**54.16752**	**66.9457**		**454.7339**
35.212	**66.9457**	**82.74338**		**561.9917**
(X'X)^(-1)				**Beta-hat**
119.146	**-171.059**	**87.696**		**1.249**
-171.059	**544.165**	**-367.476**		**10.096**
87.696	**-367.476**	**260.008**		**-1.908**
Y'Y	**Y'PY**			
3817.66	**3817.525**			

p.7.a. Set this null hypothesis in the form

H_0: $\mathbf{K'\boldsymbol{\beta} - m = 0}$

p.7.b. Obtain the estimate of $\mathbf{K'\boldsymbol{\beta} - m}$:

p.7.c. Obtain $\mathbf{K'(X'X)^{-1}K}$

p.7.d. Obtain the estimate of σ^2

p.7.e. Compute the test statistic, give the rejection region, and conclusion for the test:

Test Statistic: ____________________ Rejection Region: ____________________ Reject H_0? **Yes** or **No**

QB.8. A regression model is fit, relating total team payroll (Y, in millions of £^{s}) to offensive goals scored (X_1) and defensive goals allowed (X_2) for the n=20 teams during the 2013 English Premier League season. For this problem, we will treat this as a sample from a population of all possible league teams.

Rank	Team	Y	X0	X1	X2
1	Man City	233	1	66	34
2	Chelsea	179	1	75	39
3	Manchester United	181	1	86	43
4	Arsenal	154	1	72	37
5	Liverpool	132	1	43	28
6	Tottenham	96	1	66	46
7	Aston Villa	72	1	47	69
8	Newcastle United	62	1	45	68
9	Sunderland	58	1	41	54
10	Everton	63	1	55	40
11	Fulham	67	1	50	60
12	Swansea City	49	1	47	51
13	West Brom	54	1	53	57
14	Stoke City	60	1	34	45
15	Norwich	75	1	41	58
16	West Ham	56	1	45	53
17	Southampton	47	1	49	60
18	QPR	78	1	30	60
19	Reading	46	1	43	73
20	Wigan	44	1	47	73

X'X				X'Y	
20	1035	1048		1806	
1035	57445	52509		104325	
1048	52509	58202		85017	
INV(X'X)				Beta-hat	
2.994125	-0.02661	-0.02991		88.84439	
-0.026608	0.000336	0.000176		1.953057	
-0.029907	0.000176	0.000397		-1.90105	
Ybar	Y'Y				
90	220320				

p.8.a. Complete the following ANOVA table.

ANOVA					
	df	*SS*	*MS*	*F*	*F(0.05)*
Regression					
Residual					
Total					

p.8.b. Test whether the offensive goals scored and defensive goals allowed effects are of equal magnitude, but opposite direction: $H_0: \beta_1 = -\beta_2$

Rank	Team	Y	X0	X1	X2
1	Man City	233	1	66	34
2	Chelsea	179	1	75	39
3	Manchester United	181	1	86	43
4	Arsenal	154	1	72	37
5	Liverpool	132	1	43	28
6	Tottenham	96	1	66	46
7	Aston Villa	72	1	47	69
8	Newcastle United	62	1	45	68
9	Sunderland	58	1	41	54
10	Everton	63	1	55	40
11	Fulham	67	1	50	60
12	Swansea City	49	1	47	51
13	West Brom	54	1	53	57
14	Stoke City	60	1	34	45
15	Norwich	75	1	41	58
16	West Ham	56	1	45	53
17	Southampton	47	1	49	60
18	QPR	78	1	30	60
19	Reading	46	1	43	73
20	Wigan	44	1	47	73

X'X				X'Y	
20	1035	1048		1806	
1035	57445	52509		104325	
1048	52509	58202		85017	
INV(X'X)				Beta-hat	
2.994125	-0.02661	-0.02991		88.84439	
-0.026608	0.000336	0.000176		1.953057	
-0.029907	0.000176	0.000397		-1.90105	
Ybar	Y'Y				
90	220320				

Test Statistic: ______________________ Rejection Region: __________________ Reject H_0? **Yes** / **No**

QB.9. A forensic study related Hand (X_1) and Foot (X_2) lengths to Stature (Y) for a sample of n = 75 adult females (each variable in 100s of mms). Consider the following three models.

$$M_1 : E\{Y\} = \beta_0 + \beta_1 X_1 \qquad M_2 : E\{Y\} = \beta_0 + \beta_2 X_2 \qquad M_3 : E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Model 1 (Hand)					Model 2 (Foot)			
X'X			X'Y		X'X			X'Y
75	142.185		1199.70		143450	6686.317		45568.10
142.185	270.1992		2276.80		6686.317	414.390402		2819.37
INV(X'X)			beta-hat		INV(X'X)			beta-hat
5.586829	-2.93992		8.9042		2.81E-05	-0.0004537		0.0022
-2.93992	1.550753		3.7408		-0.00045	0.009733694		6.7688
Y'Y		Model 3 (Hand,Foot)						
19208.28		X'X				X'Y		
		75	142.185	176.062		1199.70		
		142.185	270.1992	334.2884		2276.80		
		176.062	334.2884	414.3904		2819.37		
		INV(X'X)				beta-hat		
		6.640061	-1.95728	-1.24223		7.4414		
		-1.95728	2.467531	-1.15897		2.3760		
		-1.24223	-1.15897	1.465136		1.7253		

p.9.a. Compute $\mathbf{Y}'\left(\frac{1}{n}\mathbf{J}\right)\mathbf{Y}$ and the Total (Corrected) Sum of Squares.

p.9.b. Compute the Residual (Error) Sum of Squares for each model.

p.9.c. Compute $R(\beta_1 \mid \beta_0), R(\beta_2 \mid \beta_0), R(\beta_1 \mid \beta_0, \beta_2), R(\beta_2 \mid \beta_0, \beta_1)$

p.9.d. Use the general linear test for Model 3 to test $H_0 : \beta_1 = \beta_2 \quad \text{vs} \quad H_A : \beta_1 \neq \beta_2$

QB.10. Consider a sequence of regression models to be fit, each based on n observations:

Model 0: $Y_i = \beta_0 + \varepsilon_i$

Model 1: $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

Model 2: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

p.10.a. Set-up $\boldsymbol{P}_0$, the projection matrix for model 0.

p.10.b. Obtain $R(\beta_0)$ in terms of the data $Y_1,\ldots,Y_n$.

p.10.c. Suppose $Y'P_0Y = 500 \quad Y'P_{01}Y = 750 \quad Y'P_{02}Y = 600 \quad Y'P_{012}Y = 1000$. Complete the following table:

Variable	Sequential SS	Partial SS
X1		
X2		
Variable	**Sequential SS**	**Partial SS**
X2		
X1		

p.10.d. Suppose SS(Total Corrected) = 1000.

- P.10.d.i. Give the proportion of variation in Y that is explained by X_1 alone
- P.10.d.i.. Give the proportion of variation in Y that is not explained by X_1 that is explained by X_2

QB.11. A study considered noise level of the Teheran-Karaj express train (Y, in dB) in terms of distance to the center of the track (X_1, in meters) and speed of the train (X_2, in km/h), with n = 50. Consider the following models (in matrix form).

Model 0: $E\{Y\} = \beta_0$ Model 01: $E\{Y\} = \beta_0 + \beta_1 X_1$ Model 02: $E\{Y\} = \beta_0 + \beta_2 X_2$

Model 012: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ $\mathbf{Y'Y} = 348654$

p.11.a. The sum of the speeds of the 50 observations is $\sum_{i=1}^{50} Y_i = 4174.2$. For model 0, obtain:

$\mathbf{X_0'X_0}$, $\mathbf{X_0'Y}$, $\left(\mathbf{X_0'X_0}\right)^{-1}$, $\hat{\boldsymbol{\beta}}_0$, $\mathbf{Y'P_0Y}$

$\mathbf{X_0'X_0} =$ ______ $\mathbf{X_0'Y} =$ ________ $\left(\mathbf{X_0'X_0}\right)^{-1} =$ ________ $\hat{\boldsymbol{\beta}}_0 =$ ________ $\mathbf{Y'P_0Y} =$ ________

p.11.b. For Models 01, 02, and 012, you obtain the following

$\mathbf{X_*'Y}$, $\hat{\boldsymbol{\beta}}_*$ Compute $\mathbf{Y'P_{01}Y}$, $\mathbf{Y'P_{02}Y}$, $\mathbf{Y'P_{012}Y}$ and MSE_{012}

X01'Y	Beta-hat01		X02'Y	Beta-hat02		X012'Y	Beta-hat012
4174.2	88.5825		4174.2	75.7838		4174.2	80.1494
186706	-0.1133		332604.84	0.0967		186706	-0.1158
						332604.84	0.1073

$\mathbf{Y'P_{01}Y} =$ __________ $\mathbf{Y'P_{02}Y} =$ __________ $\mathbf{Y'P_{012}Y} =$ __________ $MSE_{012} =$ __________

p.11.c. Obtain the Sequential and Partial sums of squares for X_1 and X_2, and their corresponding F-statistics.

Variable	**Sequential SS**	**Sequential F**		**Partial SS**	**Partial F**
X1					
X2					

QB.12. Consider the general linear test $H_0 : \mathbf{K'\boldsymbol{\beta}} = \mathbf{0}$ where **K'** has q ≤ p' linearly independent rows.

p.12.a. Derive the mean vector and variance-covariance matrix of $\mathbf{K'}\hat{\boldsymbol{\beta}}$.

p.12.b. Show that $Q = \left(\mathbf{K'}\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{K'}\left(\mathbf{X'X}\right)^{-1}\mathbf{K}\right]^{-1}\mathbf{K'}\hat{\boldsymbol{\beta}}$ and SSE are independent. Hint: Write Q = **Y'AY**.

QB.13. A study was conducted, relating an abrasivity index measure (Y) to p = 4 predictors: UCS (X_1), BTS (X_2), and two brittleness indices: B_1 (X_3) and B_3 (X_4) in a sample of igneous rocks. The model fit is given below along with computations.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \quad \varepsilon \sim NID\left(0, \sigma^2\right)$$

X'X						X'Y
40	102.81	51.975	62.766	93.597		109.60
102.81	322.0451	121.9227	156.505	223.8523		263.04
51.975	121.9227	80.71724	93.99107	136.4427		152.32
62.766	156.505	93.99107	122.2703	153.0712		186.27
93.597	223.8523	136.4427	153.0712	241.7843		261.22
INV(X'X)						Beta-hat
1.509	-0.071	1.554	-0.634	-0.994		5.671
-0.071	0.023	0.045	-0.017	-0.008		-0.255
1.554	0.045	3.028	-1.150	-1.624		4.972
-0.634	-0.017	-1.150	0.482	0.605		-1.305
-0.994	-0.008	-1.624	0.605	0.930		-2.859
Y'Y	Y'(J/n)Y	Y'PY				
331.28	300.30	322.05				

p.13.a. Compute SSE, df_E and MSE

p.13.b. Compute SSR, df_R and MSR

p.13.c. Test H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (abrasivity index is not associated with any of the predictors)

Test Statistic ___________________________ Rejection Region ____________________________ P > < .05

p.13.c. Test H_0: $\beta_3 = \beta_4$ (The coefficients for the two brittleness indices are equal)

K' = **m =**

Test Statistic ___________________________ Rejection Region ____________________________ P > < .05

QB.14. Regression models were fit, relating various crime rates for U.S. states to a set of 25 predictors. The researchers fit the full model with all 25 predictors (say X_1,…,X_{25}) and then the best 4 predictor model (say X_1,…,X_4). For the outcome Total Crime, the authors report the following coefficients of multiple determination.

$$R^2\left(X_1,...,X_{25}\right) = .913 \quad R^2\left(X_1,...,X_4\right) = .774 \qquad \text{Compute} \quad R^2\left(X_5,...,X_{25} \mid X_1,...,X_4\right)$$

Part C: Models with Qualitative Variables and Interactions

QC.1. A linear regression model is fit, relating apartment rental prices (Y, in \$100) to square footage for for 5 apartments in each of 4 luxury neighborhoods (all apartments were built in the same decade). We consider the following 3 models, where X_1 is the square footage (100s of ft^2); X_2 =1 if neighborhood A, 0 otherwise; X_3=1 if neighborhood B, 0 otherwise; and X_4=1 if neighborhood C, 0 otherwise.

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1$ Model 2: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

Model 3: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_1 X_2 + \beta_6 X_1 X_3 + \beta_7 X_1 X_4$

The ANOVA Tables from each model are given below.

ANOVA	Model1			Model2			Model3	
	df	*SS*		*df*	*SS*		*df*	*SS*
Regression	1	201.2		4	228.7		7	249.6
Residual	18	90.3		15	62.7		12	41.8
Total	19	291.4		19	291.4		19	291.4

p.1.a. Test whether the "square footage effect" is the same for each neighborhood by completing the following parts (homogeneity of regressions):

p.1.a.i. H_0: H_A:

p.1.a.ii. Conduct the test

Test Statistic ________________________________ Rejection Region ____________________________

p.1.b. Assuming no interaction between neighborhood and square footage, test whether the neighborhoods have different means, controlling for square footage by completing the following parts (homogeneity of regressions):

p.1.b.i. H_0: H_A:

p.1.b.ii. Conduct the test

Test Statistic ________________________________ Rejection Region ____________________________

QC.2. Write the (full rank, additive) multiple regression equation for determining if the linear relationship of Y = response time as a function of X = strength of signal has the same slope for three groups. Define all variables.

QC.3. A regression model is fit, relating time to complete a task (Y, in minutes) to nationality of the team (X_1=1 if US, 0 if non-US) and complexity (X_2, on a TACOM scale) for nuclear power plant operators. The model fit is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

ANOVA										
	df	*SS*	*MS*	*F*	*gnificance F*		INV(X'X)			
Regressio	3	3234355	1078118	79	0.0000		0.7998	-0.7998	-0.1410	0.1410
Residual	66	903715	13693				-0.7998	1.5996	0.1410	-0.2820
Total	69	4138069					-0.1410	0.1410	0.0258	-0.0258
							0.1410	-0.2820	-0.0258	0.0515
	Coefficients	*andard Err*	*t Stat*	*P-value*						
Intercept	-406.1	104.7	-3.88	0.0002						
Nation	-386.0	148.0	-2.61	0.0112						
Complexi	117.2	18.8	6.24	0.0000						
N*C	108.0	26.6	4.06	0.0001						

p.3.a. Test whether the slopes (with respect to complexity scores) are equivalent for US and non-US power plants.

H_0: ________________ H_A: ________________ Test Stat: ______________________ P-Value: ________________

p.3.b. Give the estimated mean time to complete a task with complexity of $X_2 = 5$ for US and non-US plants.

US: ____ non-US: ____

p.3.c. Compute a 95% Confidence Interval for the difference of the means estimated in p.3.b.

QC.4. Regression models were fit, relating height (Y, in mm) to hand length (X_1, in mm), foot length (X_2, in mm) and gender (X_3=1 if male, 0 if female) based on a sample of 80 males and 75 females. Consider these 4 models:

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3$ Model 2: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Model 3 (Males Only): $E\{Y\} = \delta_0 + \delta_1 X_1 + \delta_2 X_2$ Model 4 (Females Only): $E\{Y\} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$

ANOVA	Model1	
	df	SS
Regression	5	1201091
Residual	149	157138
Total	154	1358229
	Coefficients	Standard Error
Intercept	744.14	83.68
Hand	2.38	0.51
Foot	1.73	0.39
Male	-304.72	125.47
MaleHand	0.91	0.68
MaleFoot	0.65	0.52

ANOVA	Model2	
	df	SS
Regressio	3	1193101
Residual	151	165128
Total	154	1358229
	Coefficients	Standard Error
Intercept	582.16	60.55
Hand	2.81	0.34
Foot	2.06	0.26
Male	39.61	8.50

ANOVA	Model3	
	df	SS
Regression	2	208298
Residual	77	88305
Total	79	296603
	Coefficients	Standard Error
Intercept	439.42	97.49
Hand	3.29	0.47
Foot	2.38	0.35

ANOVA	Model4	
	df	SS
Regression	2	110552
Residual	72	68833
Total	74	179385
	Coefficients	Standard Error
Intercept	744.14	79.67
Hand	2.38	0.49
Foot	1.73	0.37

p.4.a. Confirm the equivalence of the regression coefficients (but not standard errors) based on the appropriate models (Hint: set up the fitted equations based on the two models):

Females:

Males:

p.4.b. Test H_0: $\beta_{13} = \beta_{23} = 0$ (No interactions between Hand and Gender or Foot and Gender).

Test Statistic: __________________ Rejection Region: ______________________ p-value **>** or **<** 0.05?

p.4.c. Use Bartlett's Test to test whether the error variances among the individual regressions are equal:

$$B = \frac{1}{C}\left[\nu \ln\left(MSE\right) - \sum_{i=1}^{t} \nu_i \ln\left(s_i^2\right)\right] \qquad C = 1 + \frac{1}{3(t-1)}\left[\sum_{i=1}^{t} \nu_i^{-1} - \nu^{-1}\right]$$

Test Statistic B = _____________________ Rejection Region: ________________ p-value **>** or **<** 0.05?

p.4.d. What fraction of the total variation in height is explained by the set of predictors: hand length, foot length, and gender (but no interactions)?

p.4.e. Compute the standard deviations among the 80 Male heights and among the 75 Female heights (ignoring hand and foot length).

Males: SD = ____________________________ Females: SD = __________________________________

QC.5. A study was conducted to determine whether having been exposed to an advertisement claiming a natural ingredient is contained in a perfume had an effect on subjects' rating of the perfume's scent. There were 112 subjects of which, 56 were exposed to the ad, and 56 were not. We fit the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1,\dots,112 \quad X_i = \begin{cases} 1 \text{ if Subject } i \text{ was exposed to the ad} \\ 0 \text{ if Subject } i \text{ was not exposed to the ad} \end{cases}$$

X'X			X'Y		Y'Y
112	56		587		3683.05
56	56		337		

p.5.a. First, we fit a model with only an intercept term, what will $P_0 = X_0\left(X_0'X_0\right)^{-1}X_0'$ be (symbolically, do not write out a 112x112 matrix!)? Compute R(β_0).

P_0 = ______________________________ R(β_0) = ______________________________

p.5.b. Compute $\left(\mathbf{X'X}\right)^{-1}$ and $\hat{\boldsymbol{\beta}}$ NOTE: Write $\left(\mathbf{X'X}\right)^{-1}$ as $\frac{1}{|\mathbf{X'X}|}\mathbf{A}$ for the appropriate $\mathbf{A}$

p.5.c. Compute R(β_0 , β_1) , R(β_1 | β_0), and MSResidual

R(β_0 , β_1) = ________________ R(β_1 | β_0) = _________________ MSResidual = __________________

p.5.d. Use the t-test and the F-test to test H_0: $\beta_1 = 0$ vs H_A: $\beta_1 \neq 0$

t-Statistic: _____________________ Rejection Region: ____________________________

F-Statistic: _____________________ Rejection Region: ____________________________

QC.6. A regression model is fit, relating Weight (Y, in pounds) to Gender (X_1=1 if Male, 0 if Female) and Height (X_2, in inches) among professional NBA and WNBA players. The model fit is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

ANOVA						INV(X'X)			
	df	SS	MS	F		3.2360	-3.2360	-0.0446	0.0446
Regression	3	534045	178015	750		-3.2360	4.2777	0.0446	-0.0577
Residual	640	151857	237			-0.0446	0.0446	0.0006	-0.0006
Total	643	685902				0.0446	-0.0577	-0.0006	0.0008
	Coefficients	Standard Error	t Stat	P-value					
Intercept	-212.05	27.71	-7.65	0.0000					
Male	-67.82	31.86	-2.13	0.0337					
Height	5.32	0.38	13.91	0.0000					
M*H	1.01	0.43	2.35	0.0192					

p.6.a. Test whether the slopes (with respect to height) are equivalent for male and female pro basketball players.

H_0: ________________ H_A: _______________ Test Stat: ____________________ P-Value: ________________

p.6.b. Give the estimated mean weight for a player with height X_2 = 72 inches for male and female pro basketball players.

Male ____________________________________ Female: ____________________________________

p.6.c. Compute a 95% Confidence Interval for the difference of the means estimated in p.6.b.

QC.7. A study measured Total Mercury levels (Y, in mg/g) in a sample of n=135 Kuwaiti men. The independent variables were: X_1=1 if fisherman, 0 if not; X_2 = Weight (kg); and X_3 = # Fish Meals/Week. The matrix results are given below.

X'X					X'Y
135	100	9876	881		509.666
100	100	7280	845		418.083
9876	7280	728452	64639		38360.354
881	845	64639	9529		3959.497
INV(X'X)					beta-hat
0.967339	-0.060848	-0.012685	0.002005		-11.064
-0.060848	0.063336	0.000480	-0.003248		1.027
-0.012685	0.000480	0.000171	-0.000033		0.183
0.002005	-0.003248	-0.000033	0.000432		0.106
Y'Y					
3081.235					

p.7.a Complete the following Analysis of Variance table.

Source	df	SS	MS	F_obs	F(0.05)
Regression					
Residual				#N/A	#N/A
Total (Corr)			#N/A	#N/A	#N/A

p.7.b. Obtain a 95% Confidence Interval for the effect of being a fisherman on expected total Mercury, controlling for Weight and Fish Meals/Week.

p.7.c. What proportion of the variance in Total Mercury is “explained” by this set of predictors?

QC.8. A regression model is fit, relating weight (Y, in pounds) to height (X_1 in inches) and gender (X_2=1 if male, 0 if female) among a random sample of NBA/WNBA basketball players. The relationship between weight and height is fit first, separately for males and females, then combined in the model: $E\{Y_i\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2}$

p.8.a. Complete following table and test $H_0 : \sigma^2_{\varepsilon M} = \sigma^2_{\varepsilon F}$ using Bartlett’s Test.

$$B = \frac{1}{C}\left[\nu \ln(MSE) - \sum_{i=1}^{t} \nu_i \ln\left(s_i^2\right)\right] \qquad C = 1 + \frac{1}{3(t-1)}\left[\sum_{i=1}^{t} \nu_i^{-1} - \nu^{-1}\right] \quad \text{Under } H_0 \quad B \overset{\bullet}{\sim} \chi^2_{t-1}$$

Regression	n	SSE	df	MSE
Males	15	4729.8		
Females	15	3997.4		
All	30			

Test Statistic ____________________________ Rejection Region ________________________

p.8.b. The following (partial tables) include the estimated coefficients and standard errors for the males and females separately as well as the combined model. Complete the tables.

Males			
	Coefficients	Standard Err	t Stat
Intercept	-381.24	139.75	-2.73
Height	7.57	1.76	4.30

Females			
	Coefficients	Standard Err	t Stat
Intercept	-267.36	89.55	-2.99
Height	6.18	1.24	4.99

All			
	Coefficients	Standard Err	t Stat
Intercept	-267.36	93.56	-2.86
Height	6.18	1.29	4.77
Male		163.62	
Ht*M		2.13	

p.8.c. A final model is fit among all players, relating Weight to Height without including gender or the interaction.

$$E\{Y_i\} = \beta_0 + \beta_1 X_{i1} \qquad SSE = 9038.7 \quad \text{Test } H_0: \beta_2 = \beta_3 = 0$$

Test Statistic ____ Rejection Region ____________________________

QC.9. A regression model was fit based on a sample of n=117 Black Holes. The response was Bolometric Luminosity (Y), with predictors: Black Hole Mass (X_1) and Black Hole Type (X_2 = 1 if Radio Quiet Quasar (RQQ), 0 if Radio Loud Quasar (RLQ)), and a cross-product term to allow for a possible interaction between Mass and Type. Note that there were 20 RQQ and 97 RLQ Black Holes.

$$\text{Model 1: } \hat{Y}_i = 39.356 + 0.791X_{i1} + 2.047X_{i2} - 0.257X_{i1}X_{i2} \qquad SSE_1 = 30.964$$

$$\text{Model 2: } \hat{Y}_i = 39.453 + 0.780X_{i1} \qquad SSE_2 = 31.187$$

p.9.a. Based on Model 1, give the fitted equations relating Bolometric Luminosity to Mass, seperately by Quasar Type.

RQQ: _______________________________________ RLQ ______________________________

P.9.b. Test whether the true relationship between Bolometric Luminosity and Mass is the same for RQQs and RLQs.

H_0: H_A:

Test Statistic: ___________________ Rejection Region: ___________________ P-value **>** or **<** .05

p.9.c. When the regressions are fit seperately, the fitted equations are the same as you should have in part p.7.a. The residual variances (MSE's) for the models are: RQQ: $s_Q^2 = 0.088$ RLQ: $s_L^2 = 0.309$. Use Bartlett's test to test whether the true variances are equal.

$$B = \frac{1}{C}\left[\nu \ln(MSE) - \sum_{i=1}^{t} \nu_i \ln\left(s_i^2\right)\right] \qquad C = 1 + \frac{1}{3(t-1)}\left[\sum_{i=1}^{t} \nu_i^{-1} - \nu^{-1}\right] = 1.01$$

Test Statistic: ___________________ Rejection Region: ___________________ P-value **>** or **<** .05

Part D: Models with Curvature and Response Surfaces

QD.1. A second-order response surface is fit with 2 independent variables (including all main effects, cross-product, and squared terms) and n=20 observations. Give the degrees of freedom for regression and residual, as well as the rejection region for testing H_0: $E\{Y\} = \beta_0$

df(Regression) = ________ df(Residual) = __________ Rejection Region: ________________________

QD.2. A regression model is fit, relating the number of breeding pairs of penguins to the year, over a period of years. The researchers use Y= $\log_{10}$(# breeding pairs) and X = (Year – mean(Year)) . They fit 3 Models:

Model 1: $E\{Y\} = \beta_0$ Model 2: $E\{Y\} = \beta_0 + \beta_1 X$ Model 3: $E\{Y\} = \beta_0 + \beta_1 X + \beta_2 X^2$

Model1	Model2	Model3
beta-hat	beta-hat	beta-hat
4.4197	4.4197	4.5390
	0.0199	0.0095
		-0.0012
X'Y	X'Y	X'Y
48.6162	48.6162	48.6162
	21.0521	21.0521
		4300.8246

p.2.a. Compute $R(\beta_0)$, $R(\beta_0, \beta_1)$, $R(\beta_0, \beta_1, \beta_2)$, $R(\beta_1 \mid \beta_0)$, and $R(\beta_2 \mid \beta_0, \beta_1)$ (use 4 decimal places)

$R(\beta_0)$ = ____________________ $R(\beta_0, \beta_1)$ = __________________ $R(\beta_0, \beta_1, \beta_2)$__________________

$R(\beta_1 \mid \beta_0)$ = ______________________________ $R(\beta_2 \mid \beta_0, \beta_1)$ = _____________________________

p.2.b. Compute the fitted values and residuals for the following years, for each model:

Year	X	Y	Fit1	Residual1	Fit2	Residual2	Fit3	Residual3
1981	-9.09	4.17						
1989	-1.09	4.62						
1998	7.91	4.41						

QD.3. An experiment to study the effect of temperature (x) on the yield of a chemical reaction (Y), was conducted. There was a total of n = 30 experimental runs, each using one of 2 catalysts (z=0 if catalyst 1, z=1 if catalyst 2). There were 5 evenly-spaced temperatures, coded as x = -2, -1, 0, +1, +2. There were 3 replicates per temperature/catalyst. The model fit was:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 z + \varepsilon \qquad \varepsilon \sim NID\left(0, \sigma^2\right)$$

You are given the following results:

Parameter	Estimate	Std. Err.		(X'X)^(-1)			
β0	29.83	0.33		0.114	0.000	-0.024	-0.067
β1	0.95	0.13		0.000	0.017	0.000	0.000
β2	0.41	0.11		-0.024	0.000	0.012	0.000
β3	-0.32	0.36		-0.067	0.000	0.000	0.133
SSResidual	25						

p.3.a. Test whether there is evidence of difference in catalysts, controlling for temperature.

H_0: _______________ H_A: ________________ Test Stat: _______________ Rej. Region: ______________

p.3.b. Can we conclude that the relationship is not linear? Obtain a 95% Confidence Interval for the relevant parameter, and interpret.

Confidence Interval ________________________ Conclude that the relation is linear? **Yes** or **No**

p.3.c. Obtain the estimated mean yield when catalyst 2 is used and at the standard temperature (x = 0), and compute a 95% CI for the mean.

Point Estimate: ______________________ 95% CI: ______________________________________

p.3.d. At what (centered) temperature do you estimate the yield to be maximized?

QD.4. A response surface was fit, relating (coded) Nitrogen (X_N), Phosphorous (X_P) and Number of Days (X_D) on the percent crude oil removed from an experimental oil spill (Y). The following 3 models were fit, based on n = 20 experimental spills:

Model 1: $E\{Y\} = \beta_0 + \beta_N X_N + \beta_P X_P + \beta_D X_D \qquad SSRes_1 = 2945$

Model 2: $E\{Y\} = \beta_0 + \beta_N X_N + \beta_P X_P + \beta_D X_D + \beta_{NP} X_N X_P + \beta_{ND} X_N X_D + \beta_{PD} X_P X_D \qquad SSRes_2 = 2504$

Model 3: $E\{Y\} = \beta_0 + \beta_N X_N + \beta_P X_P + \beta_D X_D + \beta_{NP} X_N X_P + \beta_{ND} X_N X_D + \beta_{PD} X_P X_D + \beta_{NN} X_N^2 + \beta_{PP} X_P^2 + \beta_{DD} X_D^2 \qquad SSRes_3 = 368$

p.4.a. Use Models 1 and 2 to test whether any of the interaction terms are significant, after controlling for main effects: $H_0 : \beta_{NP} = \beta_{ND} = \beta_{PD} = 0$

Test Statistic _____________________ Rejection Region __________________ Reject H_0? **Yes** or **No** p.4.b. Use Models 2 and 3 to test whether any of the quadratic terms are significant, after controlling for main effects and interactions: $H_0 : \beta_{NN} = \beta_{PP} = \beta_{DD} = 0$

Test Statistic ______________________ Rejection Region ___________________ Reject H_0? **Yes** or **No**

p.4.c. The coded and actual levels are given below. The model was fit based on the coded values (-1, 0, 1) and several axial points.

Var\CodedVals	**-1**	**0**	**1**
Nitrogen	**0**	**10**	**20**
Phosphorous	**0**	**1**	**2**
Days	**7**	**17.5**	**28**

Give the actual levels, corresponding to the models' intercepts: Nit = _______, Phos = _______, Days = ______

QD.5. A study related Personal Best Shot Put distance (Y, in meters) to best preseason power clean lift (X, in kilograms). The following models were fit, based on a sample of n = 24 male collegiate shot putters:

Model 1: $E\{Y\} = \beta_0 + \beta_1 X$ $\quad SSE_1 = 43.41 \quad R_1^2 = .686 \quad \hat{Y}(X) = 4.4353 + 0.0898X$

Model 2: $E\{Y\} = \beta_0 + \beta_1 X + \beta_2 X^2$ $\quad SSE_2 = 37.41 \quad R_2^2 = .729 \quad \hat{Y}(X, X^2) = 12.08 + 0.3285X - 0.00084X^2$

p.5.a. Use Model 2 to test H_0: $\beta_1 = \beta_2 = 0$ (Y is not related to X)

Test Statistic_____________________ Rejection Region: _____________________ Reject H_0? **Yes** or **No**

p.5.b. Use Models 1 and 2 to test H_0: $\beta_2 = 0$ (Y is linearly related to X)

Test Statistic: ____________________ Rejection Region: _____________________ Reject H_0? **Yes** or **No**

p.5.c. Give an estimate of the level of *X* is that maximizes *E*{*Y*}.

X^* = _____________________________

QD.6. A study related Freight Volume (Y) in Shanghai to GDP (X_1) and Fixed Investment (X_2) over a period of n = 11 years. The authors fit the following 3 models:

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1$ Model 2: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ Model 3: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_{11} X_1^2$

X1'X1				X1'Y
11	101.9252			73.9559
101.9252	1165.029			735.7308
INV(X1'X1)				Beta1
0.4801	-0.0420			4.6037
-0.0420	0.0045			0.2287

Y'Y	Ybar
511.3755	6.7233

X2'X2				X2'Y
11	101.9252	268.8874		73.9559
101.9252	1165.029	3330.45		735.7308
268.8874	3330.45	14551.43		2053.5629
INV(X2'X2)				Beta2
0.5045	-0.0506	0.0023		4.7251
-0.0506	0.0076	-0.0008		0.1860
0.0023	-0.0008	0.0002		0.0112

X3'X3				X3'Y
11	101.9252	1165.029		73.9559
101.9252	1165.029	15072.8		735.7308
1165.029	15072.8	209541.9		8799.5043
INV(X3'X3)				Beta3
2.1808	-0.4891	0.0231		4.3315
-0.4891	0.1220	-0.0061		0.3003
0.0231	-0.0061	0.0003		-0.0037

p.6.a. Compute $SSTotal_{Corrected}$

p.6.b. Compute SSRegression and SSResidual for each model.

p.6.c. Compute $R(\beta_2 \mid \beta_0, \beta_1)$ and $R(\beta_{11} \mid \beta_0, \beta_1)$

p.6.d. Test H_0: $\beta_{11} = 0$ vs H_A: $\beta_{11} \neq 0$ (Note there are 2 ways of doing this).

p.6.e. What proportion of the variation in Y that is not explained by X_1 is explained by X_2?

QD.7. Show that for simple regression, when we have n_i observations at the i^{th} distinct level of X, the Pure error sum of squares can be written as

$$SSPE = \sum_{i=1}^{c} (n_i - 1) S_i^2 \quad \text{where } S_i^2 \text{ is the sample variance of } Y_{i1}, \ldots, Y_{in_i}$$

p.7.a. An experiment was conducted to study the relationship of between yield from a chemical reaction (y) and the reaction temperature (x). The following data were obtained from n=12 runs. The fitted equation based on n=12 runs was Y-hat = 92.68-0.15x. Complete the table by filling in values for X=100.

Level(i)	**1**	**2**	**3**	**4**	**5**	**6**
n(i)	**1**	**2**	**2**	**3**	**2**	**2**
x(i)	**60**	**70**	**80**	**90**	**100**	**110**
y(i)	**51**	**82,78**	**90,96**	**100,89,99**	**82,84**	**54,52**
y-bar(i)	**51**	**80**	**93**	**96**		**53**
S^2(i)	**0**	**8**	**18**	**37**		**2**
y-hat(i)	**83.81**	**82.34**	**80.86**	**79.38**		**76.42**
n*(ybar-yhat)^2	**1076.77**	**10.92**	**294.84**	**828.62**		**1097.44**
(n-1)S^2	**0**	**8**	**18**	**74**		**2**

p.7.b. Conduct the Lack-of-Fit F-test by completing the following table: (H_0: Linear Model is appropriate)

Source	**df**	**SS**	**MS**	**F**	**F(.05)**
Lack-of-Fit					
Pure Error					

p.7.c. Do you reject the hypothesis that a linear fit is appropriate at the 0.05 significance level? **Yes / No**

p.7.d. Based on the same dataset, a quadratic model is fit based on the original (non-centered) X values:

$E(Y) = \beta_0+\beta_1X+\beta_2X^2$

- Give the fitted value for X=100
- Compute simultaneous 95% CIs for β_1 and β_2 based on Bonferroni's adjustment.

	Estimate	Std. Error
Intercept	-427.018	32.774
X	12.260	0.772
X^2	-0.072	0.004

- Based on your simultaneous CIs give an approximate confidence interval for the X value where Y is maximized.

QD.8. A response surface model related yield of methyl ester from waste canola oil (Y) to 3 factors: Time (X_1, in minutes (15,30,45)), Temperature (X_2, in degrees Celsius (240,255,270)) and Methanol/Oil Ratio (X_3 (1,1.5,2)).

Two models are considered are given below, along with the partial ANOVA tables, based on n=19 cases.

$$\text{Model 1: } E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2$$
$$\text{Model 2: } E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2$$

p.8.a. Complete the ANOVA Tables.

ANOVA	Full Model				
	df	*SS*	*MS*	*F*	*F(.05)*
Regression		17395.68			
Residual		452.7547			
Total		17848.44			
ANOVA	Reduced Model				
	df	*SS*	*MS*	*F*	*F(.05)*
Regression		17143.29			
Residual		705.149			
Total		17848.44			

p.8.b. Test whether all terms that include X_3 can be excluded from the model.

Null Hypothesis:

Test Statistic: ______________________________ Rejection Region: ___________________________

p.8.c. The fitted equation for model 2 can be written as follows. Solve for the value $\mathbf{x}^*$ that maximizes the response (in terms of $\mathbf{B}_1$ and $\mathbf{B}_2$)?

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_{12} X_1 X_2 + \hat{\beta}_{11} X_1^2 + \hat{\beta}_{22} X_2^2 = \hat{\beta}_0 + \mathbf{B}_1 '\mathbf{x} + \mathbf{x}'\mathbf{B}_2\mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \mathbf{B}_1 = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \qquad \mathbf{B}_2 = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} \end{bmatrix}$$

p.8.d. The estimated regression coefficients, $\mathbf{B}_1$ and $\mathbf{B}_2$ and $\mathbf{B}_2^{-1}$ are given below. Obtain the optimum levels of Temperature (X_1) and Time (X_2).

	Coefficients		**B1**		B2	
Intercept	2950.9606		-21.61		0.0284	0.0424
Time	-21.6095		-22.25		0.0424	0.0417
Temp	-22.2503					
TimTem	0.0849				INV(B2)	
Time2	0.0284				-67.74	68.85
Temp2	0.0417				68.85	-46.02

QD.9. A second-order response surface is to be fit with 4 predictors. How many experimental runs will be needed so that the full model will have 20 Error degrees of freedom?

Part E: Model Building

QE.1. A regression model is to be fit, relating mean January High temperature (Y) to 3 potential predictors (ELEVation, LATitude, and LONGitude). The following results are obtained:

						df	*SS*	*MS*
	b	*SE(b)*	*t Stat*	*P-value*	Regression	1	2311.213	2311.213
Intercept	60.94146	0.363461	167.6699	0	Residual	367	8428.007	22.9646
ELEV	-0.00188	0.000187	-10.0321	4.32E-21	Total	368	10739.22	
Intercept	129.2194	1.1704	110.4105	0.0000	Regression	1	9768.30	9768.30
LAT	-2.2656	0.0373	-60.7646	0.0000	Residual	367	970.92	2.65
Intercept	87.8721	9.7737	8.9906	0.0000	Regression	1	261.75	261.75
LONG	-0.2990	0.0988	-3.0280	0.0026	Residual	367	10477.47	28.55

[redacted]								
Intercept	129.5893	1.2998	99.7002	0.0000	Regression	2	9769.44	4884.72
ELEV	0.0000	0.0001	0.6563	0.5120	Residual	366	969.78	2.65
LAT	-2.2796	0.0430	-53.0544	0.0000				
[redacted]								
Intercept	105.6653	6.5092	16.2332	0.0000	Regression	2	854.10	427.05
ELEV	0.0023	0.0001	16.2047	0.0000	Residual	366	1049.01	2.87
LONG	-0.7846	0.0676	-11.6090	0.0000				
[redacted]								
Intercept	117.5506	2.9467	39.8925	0.0000	Regression	2	9814.92	4907.46
LAT	-2.3027	0.0374	-61.5047	0.0000	Residual	366	924.30	2.53
LONG	0.1297	0.0302	4.2967	0.0000				
[redacted]								
Intercept	57.9594	7.2895	7.9510	0.0000	Regression	3	9976.42	3325.47
ELEV	-0.0014	0.0002	-8.7908	0.0000	Residual	365	762.80	2.09
LAT	-2.0491	0.0446	-45.9090	0.0000	Total	368	10739.22	
LONG	0.6718	0.0675	9.9520	0.0000				

p.1.a. Based on Stepwise Regression with SLS=SLE=0.05, what will be the sequence of models selected and the final model. Give BRIEFLY the reason for each step.

p.1.b. Compute SBC and C_p for the 3 2-variable models and the 3-variable model (C_p for the 3-variable model will be 4 by definition). Based on each criteria which model is selected?

$$C_p = \frac{SS(\mathrm{Res})_p}{MS(\mathrm{Res})_{Full}} + 2p' - n \qquad SBC(p') = n\ln\left(SS(\mathrm{Res})_p\right) + [\ln(n)]p' - n\ln(n)$$

ELEV, LAT C_p= ________________________ SBC = ________________________________

ELEV, LONG C_p= ________________________ SBC = ________________________________

ELEV, LAT C_p= ________________________ SBC = ________________________________

ELEV, LAT, LONG C_p= ____ SBC = ________________________________

p.1.c. Which model will have the highest adjusted-R^2?

p.1.d. Give the Sequential and Partial sums of squares for each variable (for the ordering: ELEV, LAT, LONG) by completing the following table:

Variable	Sequential SS	Partial SS
ELEV		
LAT		
LONG		

QE.2. A study looked at the relationship between stack loss (Y, a measure of ammonia escaping a process), and 3 protential predictors: airflow (Air), cooling temperature (temp), and acid concentration (acid).

$E(Y) = \beta_0 + \beta_{Air}Air + \beta_{Temp}Temp + \beta_{Acid}Acid$

p.2.a Complete the following table where:

$$C_p = \frac{SS(Res)_p}{s^2} + 2p' - n \qquad AIC = n\ln\left(SS(Res)_p\right) + 2p' - n\ln(n)$$

SS(Total Corr) **20.69**

Independent Vars	SS(Res)	R-Square	R^2-Adj	Cp	AIC
Air	**3.19**	**0.85**	**0.84**	**13.34**	**-35.57**
Temp	**4.83**	**0.77**	**0.75**	**28.93**	**-26.86**
Acid	**17.38**	**0.16**	**0.12**	**148.26**	**0.03**
Air,Temp	**1.89**	**0.91**	**0.90**	**2.95**	**-44.59**
Air,Acid	**3.09**	**0.85**	**0.83**	**14.39**	**-34.23**
Temp,Acid	**4.75**	**0.77**	**0.74**		**-25.21**
Air,Temp,Acid	**1.79**				

p.2.b. Which model is best by each of the following criteria? Why do you choose that model for that criteria?

p.2.b.i. Adjusted-R^2:

p.2.b.ii. C_p:

p.2.b.iii. AIC:

p.2.c. Give the following sums of squares:

p.2.c.i. R(Air | Intercept, Temp, Acid):

p.2.c.ii. R(Temp, Acid | Intercept, Air)

p.2.d. Test H_0: $\beta_{Temp} = \beta_{Acid} = 0$ versus H_A: β_{Temp} and/or $\beta_{Acid} \neq 0$ at the $\alpha = 0.05$ significance level:

p.2.d.i. Test Statistic:

p.2.d.ii. Rejection Region:

QE.3. A potentially cubic regression model is fit, relating Y to X. We get the following fits for all possible models:

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	13.0256	7.7852	1.67	0.1107
X	8.6304	0.6659	12.96	0.0000

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	39.4330	1.4266	27.64	0.0000
X-square	0.4383	0.0077	56.99	0.0000

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	52.0023	2.0269	25.66	0.0000
X-cube	0.0225	0.0006	35.70	0.0000

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	46.6514	1.8053	25.84	0.0000
X	-1.9882	0.4183	-4.75	0.0002
X-square	0.5309	0.0202	26.29	0.0000

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	39.9909	2.0364	19.64	0.0000
X	2.2993	0.3301	6.97	0.0000
X-cube	0.0173	0.0008	21.03	0.0000

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	43.4009	1.2529	34.64	0.0000
X-square	0.2893	0.0308	9.40	0.0000
X-cube	0.0078	0.0016	4.91	0.0001

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	45.0566	2.1990	20.49	0.0000
X	-0.8961	0.9760	-0.92	0.3714
X-square	0.3910	0.1151	3.40	0.0034
X-cube	0.0047	0.0038	1.23	0.2338

p.3.a. We fit a Stepwise Regression Model with SLE = SLS = 0.20

p.3.a.i. What variable is entered at Step 1? Why?

p.3.a.ii. What happens at Step 2? Why?

p.3.a.iii. What happens at Step 3? Why?

QE.4a. True or False: In all possible regressions, the model chosen based on R^2-Adj criterion will always give the model as the MS(Residual) criterion. **True** or **False**

QE.4b. True or False: In Stepwise Regression, it is possible for a predictor to enter a model at an early stage, then be removed at a later stage. **True** or **False**

QE.5. All possible regressions are fit among models containing 3 potential independent variables (X_1= quay cranes/berth, X_2 = terminal (yard) cranes/berth, and X_3 = berth length. The response is the Throughput/berth (Y in 1000s of TEU). The models are based on a sample n=15 Chinese ports.

Vars	SSResid	SSReg
X1	152359	283731
X2	124788	311302
X3	434546	1544
X1,X2	96491	339599
X1,X3	112392	323698
X2,X3	86155	349935
X1,X2,X3	47908	388182
SSTotal(C)	436090	

$$C_P = \frac{SS(\text{Re}\,s)_{\text{Model}}}{MS(\text{Res})_{\text{Complete}}} + 2p' - n$$

$$AIC = n\ln\left(SS(\text{Re}\,s)_{\text{Model}}\right) + 2p' - n\ln(n)$$

$$SBC = BIC = n\ln\left(SS(\text{Re}\,s)_{\text{Model}}\right) + \ln(n)\,p' - n\ln(n)$$

p.5.a. Compute SBC for the model with X_1 as the only predictor.

p.5.b. Compute Adjusted-R^2 for the model with X_1 and X_3.

p.5.c. Compute C_p for the model with X_2 and X_3.

p.5.d. Will C_p choose model (X_2,X_3) or model (X_1,X_2,X_3)? Why?

QE.6. A series of models were fit, relating Average January High Temperature (Y, in degrees F) to Elevation (X_1, in 100s ft above sea level), and Latitude (degrees North Lat) for n = 369 weather stations in Texas. Latitude and Elevation were centered in the regression models.

$$C_P = \frac{SS(\text{Re}\,s)_{\text{Model}}}{MS(\text{Res})_{\text{Complete}}} + 2p' - n$$

$$AIC = n\ln\left(SS(\text{Re}\,s)_{\text{Model}}\right) + 2p' - n\ln(n)$$

$$SBC = BIC = n\ln\left(SS(\text{Re}\,s)_{\text{Model}}\right) + \ln(n)\,p' - n\ln(n)$$

Variables in Model	SS(RES)	C_p	AIC	SBC
ELEV (E)	7986.3	4764.2	1138.6	1146.4
LAT (L)	1168.0		429.2	437.0
E,L	616.2	32.8	195.2	207.0
E,L,E*L	603.9	26.9		205.4
E,L,E^2,L^2	575.0	10.3	173.7	
E,L,E*L,E^2,L^2	565.2	6	169.3	192.8

p.6.a. Complete the table.

p.6.b. Based on each criteria, which model do you choose?

C_p: ______________________________ AIC: _____________________ SBC: _______________________

QE.7. Regression models are fit, relating price of Compact Hybrid Cars (Y, in $1000s) to Acceleration ($X_1$, in km/hour/sec) and Miles per Gallon (max of Gas and Electric mph) for n=25 models from years 2009-2013.

Consider the following models with Residual Sums of Squares for each model ($SSTotal_{Corr}$ = 2196)

Model 1: $E\{Y\} = \beta_0 + \beta_1 X_1 \quad SSE_1 = 1307$ Model 2: $E\{Y\} = \beta_0 + \beta_2 X_2 \quad SSE_2 = 1747$

Model 3: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad SSE_3 = 953$ Model 4: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad SSE_4 = 854$

Note:

$$C_P = \frac{SS(\mathrm{Res})_{\mathrm{Model}}}{MS(\mathrm{Res})_{\mathrm{Complete}}} + 2p' - n$$

$$AIC = n\ln\left(SS(\mathrm{Res})_{\mathrm{Model}}\right) + 2p' - n\ln(n)$$

$$SBC = BIC = n\ln\left(SS(\mathrm{Res})_{\mathrm{Model}}\right) + \ln(n)p' - n\ln(n)$$

p.7.a. Compute C_p for Model 1.

p.7.b. Compute AIC for model 2

p.7.c. Compute SBC for models 3 and 4. Which model is preferred based on that criteria?

QE.8. A model is fit relating January Mean Temperature (Y, in Fahrenheit) to Elevation (X_1, in 100s of feet above sea level) and Latitude (X_2 = Degrees North Latitude – 30) for a random sample of n = 29 weather stations in Texas. Two models are fit:

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

The matrix form for Model 1 is given below.

Model1				
X1'X1				X1'Y
29	424.37	39.91667		1329.4
424.37	12333.16	1024.833		17777.29
39.91667	1024.833	183.5892		1468.948
INV(X1'X1)				Beta-hat1
0.070532	-0.00215	-0.00333		50.6488
-0.00215	0.000217	-0.00074		-0.0954
-0.00333	-0.00074	0.010317		-2.47845
Y'Y				
62025.94				

p.8.a. Compute $\mathbf{Y}'\mathbf{P}_1\mathbf{Y}$, $\mathbf{Y}'\left(\frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}$, the Error and Regression Sums of Squares, and the estimate of σ for Model 1.

$\mathbf{Y}'\mathbf{P}_1\mathbf{Y} =$ ___________ $\mathbf{Y}'\left(\frac{1}{n}\mathbf{J}_n\right)\mathbf{Y} =$ ___________ $SSE =$ ___________ $SS\text{Reg} =$ ___________ $s =$ ___________

p.8.b. Compute a 95% Confidence Interval for β_2

p.8.c. For Model 2, we get the following results. Obtain the fitted value for a location with an Elevation of 500 feet above sea level and at 30 degrees North Latitude based on each model. Note the units with which X_1 and X_2 been "operationalized" with.

Beta-hat2	Y'P2Y
51.12363	62002.74
-0.15163	
-2.30609	
0.000163	
-0.14301	
0.026796	

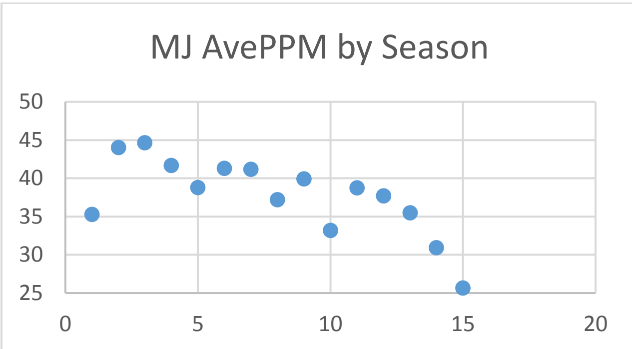
Model 1 ________________________________ Model 2 ____

p.8.d. Test $H_0: \beta_{11} = \beta_{22} = \beta_{12} = 0$

Test Statistic ____ Rejection Region ______________________________ p.8.e. Compute AIC for each model. Which model do you select based on the AIC criteria.

$$AIC = n\ln(SSE) + 2p' - n\ln(n)$$

QE.9. Over Michael Jordan's (Pro Basketball player, not UCBerkley Stat/CS Professor) career, he played 15 seasons. A plot of his average Points per 48 Minutes (regulation game) versus season is given below.



Consider the following orders of polynomials models (treating his seasons as a random sample of all seasons he could have played over the same ages / physical conditions): 2nd, 4th, 5th, 6th .

$$C_P = \frac{SS(\mathrm{Res})_{\mathrm{Model}}}{MS(\mathrm{Res})_{\mathrm{Complete}}} + 2p' - n$$

Note: $AIC = n\ln\left(SS(\mathrm{Res})_{\mathrm{Model}}\right) + 2p' - n\ln(n)$

$$SBC = BIC = n\ln\left(SS(\mathrm{Res})_{\mathrm{Model}}\right) + \ln(n)p' - n\ln(n)$$

Complete the following table. Which order model is selected by C_p? by BIC? ________________

Poly Order	# Parms	df_Err	SSE	R^2	Cp	AIC	BIC=SBC
2	3		110.16	0.6926		35.908	38.032
4	5		55.61	0.8448	4.737		33.195
5	6		52.72		6.231	30.854	35.102
6	7		45.69	0.8725	7.000	30.707	

QE.10. A study relating shipping fuel use (Y, tons/day) to speed (X, knots) for container ships travelling between Tokyo and Xiamen was fit by the following regression models based on n = 20 experimental runs.

Model 1: $E\{Y\} = \beta_0 + \beta_1\left(X - \overline{X}\right)$ $\qquad SSE_1 = 182.9746$

Model 2: $E\{Y\} = \beta_0 + \beta_1\left(X - \overline{X}\right) + \beta_2\left(X - \overline{X}\right)^2$ $\qquad SSE_2 = 27.0208$

Model 3: $E\{Y\} = \beta_0 + \beta_1\left(X - \overline{X}\right) + \beta_2\left(X - \overline{X}\right)^2 + \beta_3\left(X - \overline{X}\right)^3$ $\qquad SSE_3 = 26.1413$

p.10.a. Complete the following table.

Model	p'	Cp	AIC	BIC
Linear			48.272	50.264
Quadratic		2.569		15.042
Cubic		4.000	13.356	

Note:

$$C_p = \frac{SSE(\text{Model})}{MSE(\text{Complete})} - (n - 2p') \qquad AIC = n\ln\left(\frac{SSE(\text{Model})}{n}\right) + 2p' \qquad BIC = n\ln\left(\frac{SSE(\text{Model})}{n}\right) + \ln(n)p'$$

p.10.b. Which model would you choose based on the 3 criteria? C_p ____________ AIC __________ BIC____________

Part F: Multicollinearity

QF.1. True or False: When the independent variables have been set up in a controlled experiment to be uncorrelated among themselves, the Variance Inflation Factor for each predictor will be 0. **True** or **False**

QF.2. A regression model is fit, relating January mean temperature (Y) to ELEVation, LATitude, and LONGitude for n=369 weather stations in Texas (the data are aggregated over a period of years). The following table gives the Regression and Residual sums of squares for each model. All models contain an intercept.

Model	SS(REG)	SS(RES)
ELEV	5785.3	7986.3
LAT	12603.3	1168.3
LONG	2239.5	11532.1
ELEV,LAT	13155.1	616.5
ELEV,LONG	7669.7	6101.9
LAT,LONG	13087.0	684.6
ELEV,LAT,LONG	13156.8	614.8

p.2.a. Compute R(LONG), R(LONG|ELEV), R(LONG|LAT), R(LONG|ELEV,LAT)

p.2.b. Based on the simple linear regression, relating Y to LONG: $E\{Y\} = \beta_0 + \beta_{LONG}LONG$, test

H_0: $\beta_{LONG} = 0$ H_A: $\beta_{LONG} \neq 0$

Test Statistic: ______________________________ Rejection Region: __________________________

p.2.c. Based on the simple linear regression, relating Y to ELEV, LAT, LONG:

$E\{Y\} = \beta_0 + \beta_{ELEV}ELEV + \beta_{LAT}LAT + \beta_{LONG}LONG$, test H_0: $\beta_{LONG} = 0$ H_A: $\beta_{LONG} \neq 0$

Test Statistic: ______________________________ Rejection Region: __________________________

p.2.d. The Coefficient of Determination, when LONG is regressed on ELEV and LAT is 0.843. Compute the Variance Inflation Factor for LONG.

Part G: Weighted and Generalized Least Squares

QG.1. A regression model is fit, where X is the dose individuals receive and Y is a measure of therapeutic response. The following table gives the sample size, mean, and variance for each level of dose (the overall mean is 25). Consider the simple linear regression model $Y=X\beta+\varepsilon$

X	Sample Size	Mean	Variance
0	4	11	4
2	9	16	5
4	4	24	3
8	9	30	5
16	4	49	4

p.1.a. Obtain the weighted least squares estimate of β $\hat{\beta}_W = (X^{*'}X^{*})^{-1}X^{*'}Y^{*} \quad X^{*} = WX, Y^{*} = WY$

p.1.b. Obtain the fitted values in the original scale: $\hat{Y}_W = X\hat{\beta}_W$

p.1.c. Test whether the relationship between dose and response is linear (α=0.05).

QG.2. A regression model is fit, where X is the dose individuals receive and Y is a measure of therapeutic response. The following table gives the sample size, mean, and variance for each level of dose (the overall mean is 25). Consider the simple linear regression model $Y=X\beta+\varepsilon$

X	Sample Size	Mean	
0	9	10	
2	16	15	
4	4	27	
8	16	30	
16	9	48	

p.2.a. Obtain the weighted least squares estimate of β $\hat{\beta}_W = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* \quad X^* = WX, Y^* = WY$

p.2.a.1. $W =$

p.2.a.ii. $X^* =$

p.2.a.iii. $Y^* =$

p.2.a.iv. $\hat{\beta}_W =$

p.2.b. Obtain the fitted values in the original scale: $\hat{Y}_W = X\hat{\beta}_W$

QG.3. A study related the spread in shotgun pellets (Y, sqrt(AREA)) to distance shot (X, in feet) for a particular shotgun and bullet type (different from the in-class example).

p.3.a. Based on the residuals versus fitted plot (Plot 2 on page 1), which violation(s) if any is/are violated?

i) Linearity ii) Equal Variances iii) Independence

p.3.b. We wish to fit an Estimated Weighted Least Squares Regression of the following model:

$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad x_i = X_i - \overline{X}$ with weights $w_i = \frac{1}{s_j^2}$ where s_j is SD of Y for distance group of shot i

The data are given below, give the form of **W** used in estimating $\hat{\boldsymbol{\beta}}_{\mathbf{W}} = (\mathbf{X'WX})^{-1}\mathbf{X'WY}$

X	x	x^2	y1	y2	y3	y4	y5	y6	SD(grp)
10	-20	400	3.01	3.02	3.29	3.00	3.20	3.11	0.119
20	-10	100	5.57	5.00	5.42	5.73	5.29	5.10	0.278
30	0	0	8.09	6.80	7.95	8.62	8.41	8.62	0.685
40	10	100	10.81	10.19	13.01	11.17	11.33	9.35	1.232
50	20	400	16.07	14.90	17.47	14.21	13.13	11.93	1.996

$$
\mathbf{W} = \begin{bmatrix}
\rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 \\
\rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 \\
\rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 \\
\rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 \\
\rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6 & \rule{1.5cm}{0.4pt}\,\mathbf{I}_6
\end{bmatrix}
$$

p.3.c. From the results below, complete the following table. (SSE* based on transformed residuals)

X'WX				**X'WY**	
519.5807125	**-9180.684522**	**178240.8398**		**1899.813**	
-9180.684522	**178240.8398**	**-3451225.692**		**-29592.3**	
178240.8398	**-3451225.692**	**68848673.11**		**580927.4**	
INV(X'WX)				**Beta_W**	**SE(B_W)**
0.02146079	**0.00100725**	**-0.00000507**		**8.020371**	
0.00100725	**0.00023816**	**0.00000933**		**0.286385**	
-0.00000507	**0.00000933**	**0.00000050**		**0.00203**	
SSE*	**s^2**				
25.09147082					

QG.4. A regression is fit, relating Gross Profits (Y, in \$100M) to amount bet in Slot Machines (X_1, in \$1B) and amount bet on table games (X_2, in \$1B) for all Atlantic City casinos annually for 1978-2012 (n=35). The results for the regression coefficients and their standard errors are given below, for the model fit by Ordinary Least Squares.

	Coefficients	*andard Err*
Intercept	0.193	0.818
Slots	0.159	0.023
Tables	0.552	0.165

Model: $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \varepsilon_t \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t \qquad u_t \sim iid\left(0, \sigma^2\right)$

p.4.a. Compute the test statistic and give the rejection region for testing H_0: $\beta_1 = 0$ vs H_A: $\beta_1 \neq 0$

Test Statistic: ________________________ Rejection Region: _________________________________

p.4.b. Compute a 95% Confidence Interval for β_2.

p.4.c. The Durbin-Watson test results in strongly rejecting the null hypothesis H_0: $\rho = 0$ (See plot 2, page 1). A transformation, that when applied to the X-matrix and Y-vector produces uncorrelated errors:

$$\mathbf{Y}^* = \mathbf{T}^{-1}\mathbf{Y} = \mathbf{T}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{T}^{-1}\boldsymbol{\varepsilon}:$$

$$\mathbf{T}^{-1} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} \quad \text{For this data, } \hat{\rho} = \frac{\frac{1}{n}\sum_{t=1}^{n} e_t^2}{\frac{1}{n}\sum_{t=2}^{n} e_t e_{t-1}} = \frac{1.103}{1.562} \quad \mathbf{Y} = \begin{bmatrix} 0.90 \\ 1.51 \\ \vdots \\ 4.98 \\ 3.60 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 0.41 & 0.39 \\ 1 & 1.01 & 1.02 \\ \vdots & \vdots & \vdots \\ 1 & 26.28 & 6.42 \\ 1 & 24.42 & 5.70 \end{bmatrix}$$

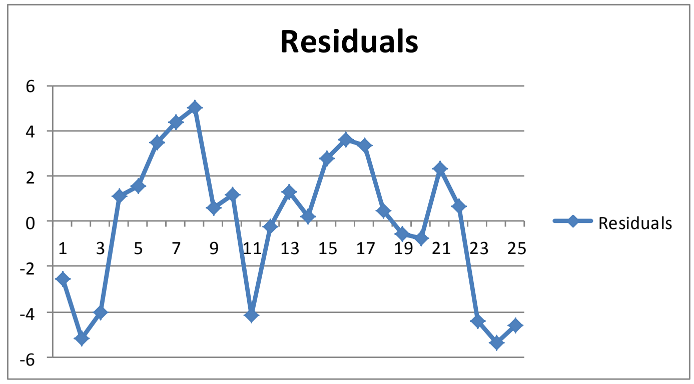
Compute the first 2 and last rows of the estimated transformed vector and matrix: $\mathbf{T}^{-1}\mathbf{Y}$ and $\mathbf{T}^{-1}\mathbf{X}$

p.4.d. The results from the estimated Generalized least squares fit are given below. Repeat parts p.3.a. and p.3.b. using them.

INV(X*'X*)				Beta_EGLS	SE(B_EGLS)
1.70995	0.00314	-0.24162		-0.0322	
0.00314	0.00228	-0.00878		0.1891	
-0.24162	-0.00878	0.07111		0.4478	
SSE*	s^2				
20.2328					

QG.5. A regression model was fit, relating the share of big 3 television network prime-time market share (Y, %) to household penetration of cable/satellite dish providers (X = MVPD) for the years 1980-2004 (n=25). The regression results and residual versus time plot are given below.

ANOVA				
	df	*SS*	*MS*	*F*
Regressio	1	7073.7	7073.7	685.7
Residual	23	237.3	10.3	
Total	24	7311.0		
	Coefficients	*tandard Err*	*t Stat*	*P-value*
Intercept	112.029	2.090	53.61	0.0000
mvpd	-0.863	0.033	-26.19	0.0000



p.5.a. Compute the correlation between big 3 market share and MVPD.

p.5.b. The residual plot appears to display serial autocorrelation over time. Conduct the Durbin-Watson test, with null hypothesis that residuals are autocorrelated.

$$\sum_{t=2}^{25}\left(e_t - e_{t-1}\right)^2 = 161.4 \qquad d_L\left(\alpha = 0.05, n = 25, p = 1\right) = 1.29 \quad d_U\left(\alpha = 0.05, n = 25, p = 1\right) = 1.45$$

Test Statistic: ____ Reject H_0? **Yes** or **No**

p.5.c. Data were transformed to conduct estimated generalized least squares (EGLS), to account for the auto-correlation. The parameter estimates and standard errors are given below. Obtain 95% confidence intervals for β_1, based on Ordinary Least Squares (OLS) and EGLS. Note that the error degrees' of freedom are 23 for OLS and 22 for EGLS (estimated the autocorrelation coefficient).

beta-egls	**SE(b-egls)**
110.577	**3.469**
-0.845	**0.055**

OLS 95% CI: __________________________________ EGLS 95% CI: ____________________________________

QG.6. A study compared surveyed velocities measured for n = 9 rock glaciers in Western Canada (Y, in meters/year) and their long term velocities, based on glacier length and age (X, in meters/year). The number of surveyed points on the glaciers varies, and **Y is the mean of those measurements**. Assume the variances of the individual points are all equal, and equal to σ^2. Note the "effective" sample size in 9 here, since we are working with the 9 means as the data.

Glacier	#points	Y	X
1	5	0.35	0.35
2	5	0.3	0.15
3	6	0.32	0.07
4	5	0.08	0.08
5	5	0.06	0.05
6	4	0.07	0.1
7	6	0.16	0.27
8	4	0.17	0.31
9	10	0.26	0.39

X'X			X'Y
9	1.77		1.77
1.77	0.4899		0.4036
INV(X'X)			b
0.3839	-1.3869		0.1197
-1.3869	7.0522		0.3914
Y'Y	Y'PY	SSE	
0.4519	0.3698	0.0821	

X*'X*			X*'Y*
50	10.73		10.39
10.73	3.1817		2.5309
INV(X*'X*)			b*
0.0724	-0.2441		0.1343
-0.2441	1.1376		0.3427
Y*'Y*	Y*'P*Y*	SSE*	
2.6917	2.2623	0.4294	

p.6.a. Give the variance of Y_5 (as a function of σ^2) :

p.6.b. We want to obtain the Weighted Least Squares Estimator:

$$\hat{\boldsymbol{\beta}}_{WLS} = \left(\mathbf{X'V^{-1}X}\right)^{-1}\mathbf{X'V^{-1}Y} = \left(\mathbf{X^{*}'X^{*}}\right)^{-1}\mathbf{X^{*}'Y^{*}} \qquad V\{\mathbf{Y}\} = \sigma^2\mathbf{V} \quad \mathbf{X}^* = \mathbf{WX} \quad \mathbf{Y}^* = \mathbf{WY}$$

Set up the **W** matrix (note that **X** also has a column of 1^s for the intercept).

p.6.c. Based on the Ordinary Least Squares (middle portion), and Weighted Least Squares (right-side portion), obtain estimated standard errors for $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{1,WLS}$ and 95% Confidence Intervals for β_1:

OLS: Standard Error: __________ CI:______________________

WLS: Standard Error: ________ CI:__________________

QG.7. A study is conducted, relating dose of a weight loss drug (X) to weight reduction (Y). There were n = 4 doses, with varying numbers of subjects per dose. The doses were X = 1, 2, 3, 4, and the sample sizes were r = 6, 8, 2, 4, respectively. A simple linear regression model, relating Y to X, with the error variance for individual subjects being σ^2. Observations are independent across subjects.

p.7.a. Set up the matrix computations to obtain the variance-covariance matrix for the weighted least squares estimator as a function of σ^2. All numbers, no symbols.

p.7.b. Set up the matrix computations to obtain the variance-covariance matrix for the ordinary least squares estimator as a function of σ^2. All numbers, no symbols.

QG.7. An economic history report studied the relationship between marginal product of labour (MPL) and estimated population of sampled communities (Pop) over an n=28 decade period from 1250-1529. Due to theoretical reasons, the model was fit on the log scale, with Y=ln(MPL) and X=ln(Pop). The author fit **weighted least squares**, where the **weight used was the number of communities (C)** which were used for population estimates by decade (the more communities used, the better the overall estimate, and C ranged from 2 to 14). The model fit was: E{**Y**} = **X**β.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \qquad \mathbf{C}^{1/2} = \mathrm{diag}\left\{\sqrt{C_i}\right\} \qquad \mathbf{Y}^* = \mathbf{C}^{1/2}\mathbf{Y} \qquad \mathbf{X}^* = \mathbf{C}^{1/2}\mathbf{X}$$

The following results are obtained for the transformed $\mathbf{Y}^*$ and $\mathbf{X}^*$. Complete the following table that would be obtained by standard software packages that fit WLS. Hint: in the unweighted case,

$$\mathbf{Y}'(1/n)\mathbf{J}\mathbf{Y} = \mathbf{Y}'\mathbf{P}_0\mathbf{Y} \quad \text{and} \quad \mathbf{Y}'\mathbf{P}\mathbf{Y} = \mathbf{Y}'\mathbf{P}_{01}\mathbf{Y}$$

X*'X*			X*'Y*
203	836.3475		917.10
836.3475	3470.013		3748.55
INV(X*'X*)			
0.7031	-0.1695		
-0.1695	0.0411		
Y*'Y*	Y*'P01*Y*	Y*'P0*Y*	
4182.513		4143.912	

Source	SS	df	MS	F	F(.05)
Regression					
Error			#N/A	#N/A	#N/A
Total			#N/A	#N/A	#N/A

	Estimate	Std Error	t	LB	UB
Intercept					
ln(Pop)					

QG.8. A regression model, relating mean temperature (Y, in F) to Year-1946 (X) for Years 1946-2014 has a Durbin-Watson statistic of DW = 0.628. Two regression models were fit, one with ordinary least squares (OLS) and one with estimated generalized least squares (GLS) with an AR1 process for errors.

Which of the following sets of estimates and standard errors do you believe are OLS and GLS?

	Estimate	Std.Error
(Intercept	75.0941	0.2158
year0	0.0385	0.0057

	Estimate	Std. Error
(Intercept	75.0789	0.1793
year0	0.0389	0.0048