Q.1. 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					

$n = 135$

$p = 3 \quad p' = 4$

$\sum_i Y_i = 509.66$

$\Rightarrow \frac{(\sum Y_i)^2}{n} = \frac{509.66^2}{135} = 1924.10$

$SS_{TOTALC} = 3081.235 - 1924.10 = 1157.14$

p.1.a Complete the following Analysis of Variance table.

2 each 6 each 2 each 3 3

Source	df	SS	MS	F_obs	F(0.05)
Regression	3	305.975	101.992	15.698	≈2.674
Residual	131	851.16	6.497	#N/A	#N/A
Total (Corr)	134	1157.14	#N/A	#N/A	#N/A

$Y'PY = \hat{\beta}'X'Y = -11.064(509.666) + 1.027(418.083) + 0.183(38360.354) + .106(3959.497)$

$= -5638.945 + 429.371 + 7019.945 + 419.704 = 2230.075$

$Y'Y - Y'PY = 3081.235 - 2230.075 = 851.16$

$Y'(P - \frac{1}{n}J)Y = 2230.075 - 1924.10 = 305.975$

p.1.b. Obtain a 95% Confidence Interval for the effect of being a fisherman on expected total Mercury, controlling for Weight and Fish Meals/Week.

3 3 6

$1.027 \pm 1.978\sqrt{6.497(.063336)} = 1.027 \pm \underbrace{1.978(.6415)}_{1.2688}$

$-.242 \quad 2.296$

$(\quad , \quad)$

3

p.1.c. What proportion of the variance in Total Mercury is "explained" by this set of predictors?

$R^2 = \frac{305.975}{1157.14} = 0.2644$ 5

45

Q.2. For the multiple regression model, with an intercept term, complete the following parts, **showing all of your work.** NOTE: $[(X'X)^{-1}]' = (X'X)^{-1}$

p.2.a. $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e}$ $\quad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ prove that $\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \mathbf{e}'\mathbf{e}$

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = PY$$

15

$$P' = (X(X'X)^{-1}X')' = X(X'X)^{-1}X' = P \qquad PP = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = P$$

$$\hat{Y}'\hat{Y} = (PY)'(PY) = Y'P'PY = Y'PPY = Y'PY$$

$$e = Y - \hat{Y} = (I-P)Y \qquad (I-P)' = I' - P' = I - P$$

$$(I-P)(I-P) = I - P - P + PP = I - P - P + P = I - P$$

$$\Rightarrow e'e = [(I-P)Y]'(I-P)Y = Y'(I-P)'(I-P)Y = Y'(I-P)(I-P)Y$$
$$= Y'(I-P)Y$$

$$Y'Y = Y'IY = Y'PY + Y'(I-P)Y = \hat{Y}'\hat{Y} + e'e$$

p.2.b. Derive the sampling distributions of $\hat{\mathbf{Y}}$ and $\mathbf{e}$

15

$\hat{Y} = PY \sim$ Normal w/

$$E\{\hat{Y}\} = PE\{Y\} = PX\beta = X(X'X)^{-1}X'X\beta = X\beta$$
$$V\{\hat{Y}\} = P\sigma^2 IP' = \sigma^2 PP' = \sigma^2 PP = \sigma^2 P$$
$$\hat{Y} \sim N(X\beta, \sigma^2 P)$$

$e = (I-P)Y \sim$ Normal w/

$$E\{e\} = (I-P)E\{Y\} = (I-P)X\beta = X\beta - PX\beta = X\beta - X\beta = 0$$

$$V\{e\} = V\{(I-P)Y\} = (I-P)\sigma^2 I(I-P)' = \sigma^2 (I-P)(I-P)'$$
$$= \sigma^2 (I-P)(I-P) = \sigma^2 (I-P) \qquad e \sim N(\underset{\sim}{0}, \sigma^2 (I-P))$$

30

p.2.c. Derive the sampling distributions of $\frac{1}{\sigma^2}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}$ and $\frac{1}{\sigma^2}\mathbf{e}'\mathbf{e}$

15 $\hat{Y}'\hat{Y} = Y'PY \qquad P\,I\,PI = PP = P = PI$

$\Rightarrow \frac{1}{\sigma^2}\hat{Y}'\hat{Y} \sim \chi^2$ w/ $df = \text{rank}(P) = tr(P) = tr(X(X'X)^{-1}X')$

$= tr(X'X(X'X)^{-1}) = tr(I_{p'}) = p'$

$\Omega = \frac{1}{2\sigma^2}\beta'X'PX\beta = \frac{1}{2\sigma^2}\beta'X'X\beta \qquad (PX = X)$

$\Rightarrow \frac{1}{\sigma^2}\hat{Y}'\hat{Y} \sim \chi^2(p', \frac{1}{2\sigma^2}\beta'X'X\beta)$

$e'e = Y'(I-P)Y \qquad (I-P)I(I-P)I = (I-P)I$

$\Rightarrow \frac{1}{\sigma^2}e'e \sim \chi^2$ w/ $df = \text{rank}(I-P) = \text{trace}(I-P) = n-p'$

$\Omega = \frac{1}{2\sigma^2}\beta'X'(I-P)X\beta = 0 \quad (IX\beta = PX\beta = X\beta) \Rightarrow \frac{1}{\sigma^2}e'e \sim \chi^2_{n-p'}$

p.2.d. Show that $\frac{1}{\sigma^2}\hat{\mathbf{Y}}'\hat{\mathbf{Y}}$ and $\frac{1}{\sigma^2}\mathbf{e}'\mathbf{e}$ are independent.

8 $A = P \qquad V = I \qquad B = (I-P)$

$AVB = PI(I-P) = P(I-P) = P - PP = P - P = 0$

23

Q.3. A forensic study related Hand (X_1) and Foot (X_2) lengths to Height (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		75	176.062		1199.70
142.185	270.1992		2276.80		176.062	414.390402		2819.37
INV(X'X)			beta-hat		INV(X'X)			beta-hat
5.586829	-2.93992		8.9042		5.087522	-2.16153487		9.3261
-2.93992	1.550753		3.7408		-2.161535	0.920784244		2.8413
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.3.a. Compute $\mathbf{Y}'\left(\frac{1}{n}\mathbf{J}\right)\mathbf{Y}$ and the Total (Corrected) Sum of Squares.

6 $Y'\frac{1}{n}JY = \frac{1}{n}\left(\sum_i Y_i\right)^2 = \frac{1}{75}(1199.70)^2 = 19190.40$

6 $SS_{TOTAL\,C} = Y'(I - \frac{1}{n}J)Y = 19208.28 - 19190.40 = 17.88$

p.3.b. Compute the Residual (Error) Sum of Squares for ~~each~~ model. 3.

5 $Y'P_1Y = 8.9042(1199.70) + 3.7408(2276.80) = 19199.42$

5 $Y'P_2Y = 9.3261(1199.70) + 2.8413(2819.37) = 19199.20$

5 $Y'P_3Y = 7.4414(1199.70) + 2.3760(2276.80) + 1.7253(2819.37) = 19201.38$

$SSE_1 = 19208.28 - 19199.42 = 8.86$

$SSE_2 = 19208.28 - 19199.20 = 9.08$

6 $SSE_3 = 19208.28 - 19201.38 = 6.90$

23

p.3.c. Compute $R(\beta_1 \mid \beta_0), R(\beta_2 \mid \beta_0), R(\beta_1 \mid \beta_0, \beta_2), R(\beta_2 \mid \beta_0, \beta_1)$

$R(\beta_1 \mid \beta_0) = 19199.42 - 19190.40 = 9.02$ 5

$R(\beta_2 \mid \beta_0) = 19199.20 - 19190.40 = 8.80$ 5

$R(\beta_1 \mid \beta_0, \beta_2) = 19201.38 - 19199.20 = 2.18$ 5

$R(\beta_2 \mid \beta_0, \beta_1) = 19201.38 - 19199.42 = 1.96$ 5

p.3.d. Use the general linear test for Model 3 to test $H_0: \beta_1 - \beta_2 = 0$ vs $H_A: \beta_1 - \beta_2 \neq 0$

$K' = [0 \quad 1 \quad -1 \quad]$ $\qquad K'\hat{\beta} = 2.3760 - 1.7253 = 0.6507$

$K'(X'X)^{-1}K = 2.467531 + 1.465136 \; 2(-1.15897) = 6.2506$

$\Rightarrow Q = \dfrac{(0.6507)^2}{6.2506} = .0677$ 4 $\qquad k = 1$

$s^2 = MSE_3 = \dfrac{6.90}{75-3} = 0.0958$ 5

T.S. $F_{obs} = \dfrac{Q/k}{s^2} = \dfrac{.0677}{.0958} = 0.7071$ 3

3

RR: $F_{obs} \geq F_{.05,1,72} \approx 3.975$ Fail to reject H_0

40

Q.4. A simple linear regression model was fit, relating Weight (Y, in pounds) to Height (X, in inches) among a sample of n=9 Women's NBA basketball players. There were 3 distinct height levels, with 3 players per height. Use the tables below to conduct the F-test for Lack of Fit.

$H_0 : E\{Y_{ij}\} = \beta_0 + \beta_1 X_j$ vs $H_A : E\{Y_{ij}\} = \mu_j \neq \beta_0 + \beta_1 X_j$

$c=3$ $p'=2$ $n-c=6$ $c-p'=1$

Height	Weight	Ybar(Grp)	Yhat(Grp)	PE	LF
68	130	150	145.83	$(-20)^2=400$	$4.17^2=17.39$
68	165	150	145.83	225	17.39
68	155	150	145.83	25	17.39
72	180	165	173.33	225	69.39
72	165	165	173.33	0	69.39
72	150	165	173.33	225	69.39
76	200	205	200.83	25	17.39
76	185	205	200.83	400	17.39
76	230	205	200.83	625	17.39
				2150	312.51

	Coefficients
Intercept	-321.667
Height	6.875

$\hat{Y}_{68} = 145.83$

$\hat{Y}_{72} = 173.33$

$\hat{Y}_{76} = 200.83$

3 each 8 each 3 each 3 3

Source	df	SS	MS	F	F(.05)
Lack of Fit	1	312.51	312.51	0.87	5.987
Pure Error	6	2150	358.33	#N/A	#N/A
Error	7	2462.51	351.79	#N/A	#N/A

37

Q.5. 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.

$n = 63 \quad p' = 3 \quad p = 2$

ANOVA	2 each		3 each	3	3
	df	SS	MS	F	F(.05)
Regression	2	2472.2	1236.1	807.03	3.150
Residual	60	91.9	1.53	#N/A	#N/A
Total	62	2564.1	#N/A	#N/A	#N/A
	Coefficients	tandard Err	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.5.a. Complete the ANOVA table.

p.5.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.

$\hat{Y}_{DFW} = 115.906 - 0.117(5.6) - 2.183(32.9) = 43.43$ (5)

p.5.c. For DFW, we obtain the following values: $x_0 = \begin{bmatrix} 1 \\ 5.6 \\ 32.9 \end{bmatrix}$ $\qquad x_0'(\mathbf{X'X})^{-1}x_0 = 0.0454$.

Compute the 95% **Prediction Interval** for DFW's mean January temperature. $t_{.025,\,60} = 2.000$

(3) (5)

$43.43 \pm 2.000\sqrt{1.53(1+.0454)} = 43.43 \pm \underbrace{2.000(1.2647)}_{2.53}$

$= (40.90,\ 45.96)$ (3)

32