STAT 2509B
Test#2

(March 7th, 2012)

All solutions must be shown clearly.

1. Agency revenues. An economic consultant was retained by a large employment agency in a metropolitan area to develop a regression model for predicting monthly agency revenues (y). She decided that three economic indicators for the area were potentially useful as independent variables, namely, average weekly overtime hours of production workers in manufacturing (x_1), number of job vacancies in manufacturing (x_2), and index of help wanted advertising in newspapers (x_3). Monthly observations on agency revenues and the three independent variables were obtained for the past 25 months. The ANOVA table for the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ is as follows:

Source	***d.f.***	***SS***	***MS***
Regression	3	5409.89	1803.30
Error	21	16.35	0.78
Total	24	5426.24	

The consultant decided to screen the independent variables to determine the best set for predicting agency revenues. The regression and the error sums of squares for all possible regression models were found to be as follows:

Independent variables in the model	SSR	SSE
x_1	2970.64	2455.60
x_2	3654.85	1771.39
x_3	3584.54	1841.70
x_1, x_2	5123.80	302.44
x_1, x_3	5409.59	16.65
x_2, x_3	3741.30	1684.94
x_1, x_2, x_3	5409.89	16.35

[6] (a) Determine the subset of variables that is selected as best using **max R^2 criterion**. Show your steps.

[5] (b) Determine the subset of variables that is selected as best using **min MSE criterion**. Show your steps.

[10] (c) Determine the subset of variables that is selected as best using **Mallows C_p criterion**. Show your steps.

2. A pharmaceutical company testing a new pain-killing drug tests it on 20 people suffering from arthritis. The time elapsed, in minutes, from taking the drug until a noticeable relief in pain is detected is to be predicted from the dosage (in grams) and the age of the patient (in years). The results are given in the following table:

Time (y)	**Dosage(x_1)**	**Age(x_2)**	**Time (y)**	**Dosage(x_1)**	**Age(x_2)**
11	2	59	35	7	33
3	2	57	28	7	49
20	2	22	23	7	29
25	2	12	21	7	32
27	2	18	33	7	20
15	5	40	27	10	43
10	5	64	8	10	61
34	5	27	3	10	69
14	5	54	12	10	62
34	5	22	14	10	61

[6] (a) State all assumptions which are necessary for the statistical inference.

[8] (b) Use matrices to compute the estimates of the population parameters β_0, β_1, β_2 and hence obtain the fitted least squares prediction line.

Hint: $\mathbf{X^TX} = \begin{bmatrix} 20 & 120 & 834 \\ 120 & 890 & 5472 \\ 834 & 5472 & 41158 \end{bmatrix}$, $\mathbf{X^TY} = \begin{bmatrix} 397 \\ 2327 \\ 13688 \end{bmatrix}$, $\mathbf{Y^TY} = \sum y_i^2 = 9927$

$$\mathbf{(X^TX)^{-1}} = \begin{bmatrix} 0.3863076905 & -0.021679047 & -0.004945645 \\ -0.021679047 & 0.0073707559 & -0.000540659 \\ -0.004945645 & -0.000540659 & 0.0001963933 \end{bmatrix}, \quad \sum y_i = 397$$

[22] (c) Set up the ANOVA table and hence test for the significance of the model. Use $\alpha = 0.05$.

[8] (d) Test whether x_2 term (i.e. whether the age) contributes to the given model. Use t-test with $\alpha = 0.05$.

[6] (e) Find the values of the coefficient of determination, r^2, and the adjusted r^2 and interpret their meanings in this problem.

3. A linear model relating y (number of vehicle sales per month) to independent variables x_1 (price per gallon), x_2 (interest rate) and x_3 is

[4] $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3 + \varepsilon,$$

where the independent variable (vehicle model) is defined by the dummy variable

$$x_3 = \begin{cases} 1, & if \quad standard \\ 0, & if \quad luxury \end{cases}$$

Interpret the meanings of the parameters β_3 and β_5 .

4. When studying the amount of heat evolved during curing in calories per gram of cement, the following variables were considered: (x_1) tricalcium aluminate, tricalcium silicate, (x_3) tetracalcium alumino ferrite and (x_4) dicalcium silicate in percent weight of the clinkers from which the cement was made.

[5] (a) Find the multiple linear regression equation relating the amount of heat evolved during curing of cement to the independent variables x_1, x_2, x_3 and x_4 using SAS output (next page).

[2] (b) Based on the SAS output, define your full model and your reduced model.

[10] (c) Test whether variables x_3 and x_4 contribute to the model. Use SAS output provided on the next page. Use α = 0.05. (Use partial F-test).

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	2657.65659	1328.82830	228.504	0.0001
Error	10	57.50448	5.75045		
C Total	12	2715.76308			

Root MSE	2.40004	R-square	0.9787
Dep Mean	95.43308	Adj R-sq	0.9744
C.V.	2.52170		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	52.577349	2.28617433	22.998	0.0001
X1	1	1.468506	0.12130082	12.105	0.0001
X2	1	0.662250	0.04585472	14.442	0.0001

Model: MODEL1
Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	2667.05844	666.97461	111.479	0.0001
Error	8	47.98464	5.98235		
C Total	12	2715.76308			

Root MSE	2.44501	R-square	0.9824
Dep Mean	95.43308	Adj R-sq	0.9736
C.V.	2.56233		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	62.405369	70.07095821	0.891	0.3991
X1	1	1.5511(0	0.74476987	2.083	0.0708
X2	1	0.510108	0.72572880	0.705	0.5009
X3	1	0.10'909	0.70470805	0.135	0.8959
X4	1	-0.144061	0.70905286	-0.203	0.8447