STATISTICS 174: FALL 2000: FINAL EXAM

DECEMBER 16 2000, 12:00–3:00 p.m.

Open book. Calculators allowed.

The exam is expected to be your own unaided work and you are not allowed to consult with each other or to receive any other form of outside assistance. You should sign the "pledge" to this effect on the front of your answer book. You may ask me questions at any time, and I encourage you to do that if you feel there is any ambiguity about the questions or if in some other way you are not sure what is being asked of you. You may take this question paper away with you but please show ALL workings, including rough working, in your answer book and do not take any of that away with you. In questions of a more verbal and descriptive nature, you are not expected to provide very long and detailed answers but please make sure you indicate clearly how you reached whatever conclusions you did.

There is no limit on how many questions, or parts of questions, you may answer but you are expected to do about two thirds of the whole exam, so that gives you some choice about which questions to do. Some parts of some questions depend on earlier parts of the same question, but if you are able to answer the later parts based on the information given, you will still receive credit for them even if you did not answer the earlier parts. The three questions are of equal weight overall. Please budget your time carefully and do not spend excessive time on any single question or part-question. Showing you understand the methods correctly is more important than getting the numerical calculations completely correct, though you should still take as much care as you reasonably can with the numerical calculations.

1. Consider the hypothetical data set of Table 1.

i	1	2	3	4	5	6	7
x_i	–2	–2	–1	0	1	2	2
y_i	2	2	4	3	1	0	–1

Table 1: Data for question 1

(a) We wish to fit a model of form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i,$$

where ϵ_i are independent $N(0, \sigma^2)$ errors. Using the matrix formulation of the linear model, calculate the least squares estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$.

(b) Suppose s^2 is the usual unbiased estimator of σ^2. Calculate s, and hence the standard errors of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$. Would you recommend removing any of these three parameters from the model?

(c) Writing $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$, suppose we are interesting in finding a confidence set for the maximum of f, with coverage probability at least $1 - \alpha$ for some given α. Show that one possible confidence set takes consists of all x such that

$$|\widehat{\beta}_1 + 2\widehat{\beta}_2 x + 3\widehat{\beta}_3 x^2| < st^* \sqrt{a + bx^2 + cx^4} \tag{1}$$

where t^* is an appropriate percentage point from a t distribution and a, b and c are constants. Show that $b = -\frac{647}{552}$ and find similar expressions for a and c. Why is this problem more complicated than the corresponding one for a quadratic function?

(d) An alternative derivation of (1) is that, for any fixed x, and whatever the true values of β_1, β_2 and β_3, the inequality

$$|(\widehat{\beta}_1 - \beta_1) + 2(\widehat{\beta}_2 - \beta_2)x + 3(\widehat{\beta}_3 - \beta_3)x^2| < st^* \sqrt{a + bx^2 + cx^4} \tag{2}$$

holds with probability $1 - \alpha$. (You do not need to give a separate proof of this.) Suppose now, however, we want equation (2) to hold simultaneously for all values of x. This could be achieved by changing the definition of t^*. Explain how and why.

2. The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \tag{3}$$

is being fitted with the x_{i1}, x_{i2} values in Table 2. As usual, we assume the ϵ_i values are independent errors of zero mean and common variance σ^2.

i	1	2	3	4	5	6
x_{i1}	–4	–2	–1	0	1	6
x_{i2}	4	–3	–1	–1	2	–1

Table 2: Data for question 2

(a) Suppose the statistician ignores the x_{i2} coordinates and fits the model (3), in effect assuming $\beta_2 = 0$. What will be the effect on the bias and

variance of $\widehat{\beta}_1$? Under what condition is the mean squared error of $\widehat{\beta}_1$ smaller using this estimate than if the statistician fitted the correct model including both β_1 and β_2? (*Note:* The answer depends on the value of σ^2, in a way that you should determine.)

(b) Repeat the calculations of part (a) but in the case when the objective of the analysis is not to estimate β_1, but to predict a future set of observations. Specifically, assume that a future set of observations is taken at x_{i1}, x_{i2} values the same as above, but independently, and the criterion for prediction is the average mean squared error over the predictions.

(c) Suppose, instead of an ordinary least squares estimator of β_1, we were to use ridge regression (but still only using x_{i1} as a covariate). Show how to derive the bias, variance and mean squared error of the estimator of β_1 under this scenario.

Note: You may use the results of examples or problems discussed during the course, but if you do so, be explicit about what results you are using.

3. The data in Table 3 represent measurements over 16 successive years of two measures of rubber production (Y1 and Y2) in terms of four covariates (X1: Car production; X2: Gross national product; X3: Dispoasable personal income; X4: motor fuel consumption). The objective of this exercise is to predict Y1 as a linear function of X1, X2, X3 and X4.

   Here is a sample SAS program:

```
options ls=77 ps=58;
data rubber;
infile 'rubber.dat';
input num y1 y2 x1 x2 x3 x4;
run;
;
proc reg;
model y1=x1 x2 x3 x4 / selection=rsquare;
run;
;
proc reg;
model y1=x1 x2 x3 x4 / vif collin influence;
run;
;
```

   Following this is the (very slightly edited) output of the above program:

Number	Y1	Y2	X1	X2	X3	X4
1	0.909	0.871	1.287	0.984	0.987	1.046
2	1.252	1.220	1.281	1.078	1.064	1.081
3	0.947	0.975	0.787	1.061	1.007	1.051
4	1.022	1.021	0.796	1.013	1.012	1.046
5	1.044	1.002	1.392	1.028	1.029	1.036
6	0.905	0.890	0.893	0.969	0.993	1.020
7	1.219	1.213	1.400	1.057	1.047	1.057
8	0.923	0.918	0.721	1.001	1.024	1.034
9	1.001	1.014	1.032	0.996	1.003	1.014
10	0.916	0.914	0.685	0.972	0.993	1.013
11	1.173	1.170	1.291	1.046	1.027	1.037
12	0.938	0.952	1.170	1.004	1.001	1.007
13	0.965	0.946	0.817	1.002	1.014	1.008
14	1.106	1.096	1.231	1.049	1.032	1.024
15	1.011	0.999	1.086	1.023	1.020	1.030
16	1.080	1.093	1.001	1.035	1.053	1.029

Table 3: Data for question 3

```
                    The REG Procedure
                      Model: MODEL1
                  Dependent Variable: y1

                 R-Square Selection Method

Number in
  Model      R-Square    Variables in Model

      1        0.7351    x3
      1        0.6637    x2
      1        0.4035    x1
      1        0.3652    x4
-----------------------------------------
      2        0.8265    x1 x3
      2        0.7807    x2 x3
      2        0.7597    x3 x4
      2        0.7448    x1 x2
      2        0.6689    x2 x4
      2        0.5622    x1 x4
-----------------------------------------
```

	3	0.8470	x1 x2 x3
	3	0.8374	x1 x3 x4
	3	0.7848	x2 x3 x4
	3	0.7474	x1 x2 x4
	4	0.8490	x1 x2 x3 x4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	4	0.16063	0.04016	15.47
Error	11	0.02856	0.00260	
Corrected Total	15	0.18918		

Source	Pr > F
Model	0.0002
Error	
Corrected Total	

Root MSE	0.05095	R-Square	0.8490
Dependent Mean	1.02569	Adj R-Sq	0.7942
Coeff Var	4.96770		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.94132	0.85034	-3.46	0.0053
x1	1	0.13170	0.06087	2.16	0.0534
x2	1	0.71618	0.77620	0.92	0.3760
x3	1	2.68615	0.98688	2.72	0.0199
x4	1	0.34859	0.90342	0.39	0.7070

Variable	DF	Variance Inflation
Intercept	1	0
x1	1	1.30157
x2	1	3.71202
x3	1	2.81265

x4	1	1.85587	

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	-Proportion of Variation- Intercept	x1
1	4.96239	1.00000	0.00000909	0.00148
2	0.03695	11.58929	0.00039809	0.81398
3	0.00041080	109.90845	0.22365	0.17049
4	0.00016527	173.27996	0.06325	0.00480
5	0.00008737	238.32222	0.71269	0.00925

Collinearity Diagnostics

	---------Proportion of Variation---------		
Number	x2	x3	x4
1	0.00001049	0.00000650	0.00000754
2	0.00024746	0.00020002	0.00025429
3	0.33428	0.00824	0.01666
4	0.02602	0.39387	0.59779
5	0.63944	0.59769	0.38529

Output Statistics

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
1	-0.0398	-1.2912	0.6126	1.9229	-1.6238
2	0.0177	0.4776	0.5095	2.9335	0.4868
3	-0.0465	-2.5933	0.8115	0.6529	-5.3810
4	0.0500	1.1341	0.2325	1.1459	0.6242
5	-0.0594	-1.3606	0.2082	0.8700	-0.6977
6	0.0118	0.2551	0.2435	2.0609	0.1447
7	0.0381	0.8488	0.2448	1.5060	0.4833
8	-0.0586	-1.4606	0.3161	0.8956	-0.9930
9	0.0454	0.9559	0.1378	1.2064	0.3821
10	0.0505	1.1735	0.2621	1.1451	0.6994
11	0.0750	1.7668	0.1720	0.5001	0.8053
12	-0.0337	-0.7708	0.2915	1.7033	-0.4945
13	0.005967	0.1273	0.2294	2.0732	0.0695
14	0.004854	0.1067	0.2741	2.2062	0.0656
15	-0.0223	-0.4368	0.0709	1.5772	-0.1207

16	-0.0390	-0.9717	0.3833	1.6633	-0.7661

Output Statistics

	DFBETAS				
Obs	Intercept	x1	x2	x3	x4
1	-0.1442	-0.8585	0.5891	0.5894	-0.9513
2	-0.4257	-0.0192	-0.0831	0.1654	0.2836
3	-0.3711	2.6335	-4.1989	3.7749	-0.3697
4	-0.2498	-0.3814	-0.0716	-0.0608	0.3844
5	-0.0393	-0.5522	0.1536	-0.1151	0.0623
6	0.0286	0.0031	-0.0869	0.0089	0.0388
7	-0.2384	0.2268	-0.0576	0.1194	0.1250
8	0.4748	0.6235	0.4970	-0.5620	-0.2975
9	0.2654	0.0947	0.0023	-0.0637	-0.1781
10	0.1735	-0.3239	-0.2032	0.0302	0.0088
11	0.1954	0.3709	0.4160	-0.2657	-0.2634
12	-0.4118	-0.2257	-0.1598	0.1769	0.3325
13	0.0254	-0.0277	0.0068	0.0164	-0.0442
14	0.0246	0.0150	0.0390	-0.0076	-0.0479
15	-0.0276	-0.0144	-0.0251	0.0076	0.0376
16	0.2883	0.1910	0.1993	-0.6298	0.2296

Sum of Residuals	0
Sum of Squared Residuals	0.02856
Predicted Residual SS (PRESS)	0.11536

Answer the following questions:

(a) Show how the R-Square values can be translated into (i) Root MSE, (ii) AIC, (iii) BIC statistics, and hence give your recommendation for which of the 16 possible models you would select using forward or backward selection, AIC and BIC.

Note: Do not try to work out Root MSE, AIC and BIC for all 16 models. Do a small number of these to show you understand the method, and then concentrate the remaining calculations on what appear to be the most promising models.

(b) Comment on the issue of multicollinearity as it affects this regression analysis.

(c) Are there influential or outlying observations in this data set? Comment with reference to each of the standard diagnostics tabulated in the above analysis.

SOLUTIONS

[Numbers in square brackets are points per part-question; 33 points for each full question, 99 points total.]

1. (a) [10] We have

$$X = \begin{bmatrix} 1 & -2 & 4 & -8 \\ 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 2 & 4 & 8 \end{bmatrix}, \quad X^TX = \begin{bmatrix} 7 & 0 & 18 & 0 \\ 0 & 18 & 0 & 66 \\ 18 & 0 & 66 & 0 \\ 0 & 66 & 0 & 258 \end{bmatrix},$$

$$(X^TX)^{-1} = \begin{bmatrix} \frac{66}{138} & 0 & -\frac{18}{138} & 0 \\ 0 & \frac{258}{288} & 0 & -\frac{66}{288} \\ -\frac{18}{138} & 0 & \frac{7}{138} & 0 \\ 0 & -\frac{66}{288} & 0 & \frac{18}{288} \end{bmatrix}, \quad X^Ty = \begin{bmatrix} 11 \\ -13 \\ 17 \\ -43 \end{bmatrix},$$

$$\widehat{\beta} = (X^TX)^{-1}X^Ty = \begin{bmatrix} \frac{420}{138} \\ -\frac{516}{288} \\ -\frac{79}{138} \\ \frac{84}{288} \end{bmatrix} = \begin{bmatrix} 3.0435 \\ -1.7917 \\ -0.5275 \\ 0.2917 \end{bmatrix}.$$

Hence $\widehat{\beta}_0 = 3.0435$, $\widehat{\beta}_1 = -1.7917$, etc.

Comment. Virtually everyone missed the trick for inverting X^TX. Maybe if columns 2 and 3 of X had been interchanged, so that X^TX looked like

$$\begin{bmatrix} 7 & 18 & 0 & 0 \\ 18 & 66 & 0 & 0 \\ 0 & 0 & 18 & 66 \\ 0 & 0 & 66 & 258 \end{bmatrix},$$

it would have been easier? In block-diagonal form, all that's needed is to invert (separately) the two 2×2 matrices.

(b) [9] $s^2 = \sum e_i^2/(n-p) = \sum e_i^2/3$ (since $n = 7, p = 4$), and $\sum e_i^2 = \sum y_i^2 - \sum \widehat{y}_i^2$. Since

$$\begin{aligned} \widehat{y} &= X(X^TX)^{-1}X^Ty, \\ \widehat{y}^T\widehat{y} &= y^TX(X^TX)^{-1}X^Ty = \widehat{\beta}^T(X^Ty) \\ &= \frac{420 \times 11}{138} + \frac{516 \times 13}{288} - \frac{79 \times 17}{138} - \frac{84 \times 13}{288} \end{aligned}$$

$$= \frac{3277}{138} + \frac{3096}{288}$$
$$= 34.49637...$$

and $\sum y_i^2 = 35$, we have $s = \sqrt{(35 - 34.49637..)/3} = .40972..$ The three standard errors are

$$s\sqrt{\frac{258}{288}} = .3878, \quad s\sqrt{\frac{7}{138}} = .0923, \quad s\sqrt{\frac{18}{288}} = .1024,$$

with corresponding t ratios –4.62, –6.20, 2.85 and in fact the last of these is not statistically significant, at the 5% level, if judged against the correct t_3 distribution ($t_{3;.975} = 3.182$). So by this criterion, β_3 should be dropped from the model, but not the other parameters.

(c) [9] *On the assumption* that a maximum of f corresponds to setting $0 = f'(x) = \beta_1 + 2\beta_2 x + 3\beta_3 x^2$ (more below on this point) we treat it as a test of the null hypothesis $\beta_1 + 2\beta_2 x + 3\beta_3 x^2 = 0$, as in Fieller's method. The test statistic $\widehat{\beta}_1 + 2\widehat{\beta}_2 x + 3\widehat{\beta}_3 x^2$ has variance (referring to the $(X^T X)^{-1}$ matrix)

$$\sigma^2 \left[\frac{258}{288} + 4 \cdot \frac{7}{138} \cdot x^2 + 9 \cdot \frac{18}{288} \cdot x^4 - 6 \cdot \frac{66}{288} \cdot x^2 \right],$$

where the last contribution results from the covariance of $\widehat{\beta}_1$ and $\widehat{\beta}_3$ (the other covariances are 0). Simplifying, this reduces to

$$\sigma^2 \left[\frac{43}{48} - \frac{647}{552} x^2 + \frac{9}{16} x^4 \right]$$

so $a = 43/48$, $b = -647/552$, $c = 9/16$. The result (1) then follows from the standard formula for a t test, with $t^* = t_{3;1-\alpha/2}$.

The problem is more complicated than for a quadratic function in two ways: (i) since in equality (1) corresponds to solving a quartic (rather than quadratic) equation in x, the form of the confidence set is even less likely to be an interval, (ii) even if there is a local maximum of f it will not be a global maximum (unless $\beta_3 = 0$), and will be accompanied by a local minimum somewhere else. Therefore, all that can really be said is that, with probability $1 - \alpha$, the confidence set derived from (1) covers the stationary points of f — at this level of statistical testing, we cannot distinguish between local maxima and minima.

(d) [5] According to Scheffé's method, t^* must be replaced by $\sqrt{qF_{q;n-p;1-\alpha}}$ where q is the dimension of the subspace (of the 4-dimensional parameter space) spanned by all functions of form $\beta_1 + 2\beta_2 x + 3\beta_3 x^2$ as x ranges over the real line. In this case $q = 3$, so the answer is $\sqrt{3F_{3;3;1-\alpha}}$.

2. Here,

$$X = \begin{bmatrix} 1 & -4 & 4 \\ 1 & -2 & -3 \\ 1 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 2 \\ 1 & 6 & -1 \end{bmatrix}, \quad X^T X = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 58 & -13 \\ 0 & -13 & 32 \end{bmatrix},$$

$$(X^T X)^{-1} = \begin{bmatrix} \frac{1}{6} & 0 & 0 \\ 0 & \frac{32}{1687} & -\frac{13}{1687} \\ 0 & -\frac{13}{1687} & \frac{58}{1687} \end{bmatrix}.$$

In particular, the variance of $\widehat{\beta}_1$ is $32\sigma^2/1687$, compared with $\sigma^2/58$ in the case when x_{i2} is left out of the model (because the inverse of the matrix $\begin{bmatrix} 6 & 0 \\ 0 & 58 \end{bmatrix}$ is $\begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{58} \end{bmatrix}$), and the difference between the two variances is $\sigma^2(32/1687 - 1/58) \approx .00173\sigma^2$.

(a) [11] If we rewrite the model in the form

$$y = X_1\gamma_1 + X_2\gamma_2 + \epsilon,$$

where X_1 is the first two columns of X, X_2 is the last column of X, $\gamma_1 = (\ \beta_0 \quad \beta_1\ )^T$, $\gamma_2 = \beta_2$, then using the result of problem 5.5 (15.4) in the course notes, the bias of $\widehat{\gamma}_1$ is $(X_1^T X_1)^{-1} X_1^T X_2 \gamma_2$, where

$$X_1^T X_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -4 & -2 & -1 & 0 & 1 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ -3 \\ -1 \\ -1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -13 \end{bmatrix}.$$

Hence, the bias in $\widehat{\beta}_0$ is 0 and the bias in $\widehat{\beta}_1$ is $-\frac{13}{58}\beta_2$. The estimate of β_1 will have smaller mean squared error than the full model if

$$\left(\frac{13\beta_2}{58}\right)^2 < .00173\sigma^2.$$

It's also possible to calculate the bias term directly: noting that $\sum x_{i1} = 0$, the estimator under the reduced model is $\widehat{\beta}_1 = \sum x_{i1} y_i / \sum x_{i1}^2$, and its mean is $\sum x_{i1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) / \sum x_{i1}^2 = \beta_1 + k\beta_2$ where $k = \sum x_{i1} x_{i2} / \sum x_{i1}^2$.

(b) [11] According to the final result of problem 5.5 (15.4), the question is to determine whether

$$\gamma_2 C \gamma_2 < \sigma^2, \tag{4}$$

where

$$C = X_2^T X_2 - X_2^T X_1 (X_1^T X_1)^{-1} X_1 X_2^T.$$

However $X_2^T X_2 = 32$ (the sum of squares of the X_2 vector), and

$$X_2^T X_1 (X_1^T X_1)^{-1} X_1 X_2^T = \begin{bmatrix} 0 & -13 \end{bmatrix} \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{58} \end{bmatrix} \begin{bmatrix} 0 \\ -13 \end{bmatrix} = \frac{169}{58}.$$

Hence $C = 32 - \frac{169}{58} = 29.086$ so the criterion (4) reduces to: use the simpler model if $29.086\beta_2^2 < \sigma^2$.

We can also do this part without direct reference to earlier problems. For the ith prediction, $y_i^* = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i^*$. We also have $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1}$. Therefore, $\widehat{y}_i - y_i^* = (\widehat{\beta}_0 - \beta_0) + (\widehat{\beta}_1 - \beta_1) x_{i1} - \beta_2 x_{i2} - \epsilon_i^*$. The variance of this expression is

$$\sigma^2 \left(\frac{1}{n} + \frac{x_{i1}^2}{\sum x_{i1}^2} + 1 \right),$$

and the sum of variances over all i is $\sigma^2(n+2)$. The mean of $\widehat{y}_i - y_i^*$ is $\beta_2(k x_{i1} - x_{i2})$ so the sum of squared means is

$$\begin{aligned} \beta_2^2 \sum (k x_{i1} - x_{i2})^2 &= \beta_2^2 \left(k^2 \sum x_{i1}^2 - 2k \sum x_{i1} x_{i2} + \sum x_{i2}^2 \right) \\ &= \beta_2^2 \left(\sum x_{i2}^2 - \frac{(\sum x_{i1} x_{i2})^2}{\sum x_{i1}^2} \right), \end{aligned}$$

after substituting the value of k from (a). Adding the last two expressions gives the total mean squared prediction error under the reduced model; the same result under the full model (no bias) is $(n+3)\sigma^2$ as follows from general results that we discussed in Chapter 3.

(c) [11] I'm interpreting the ridge estimator of γ_1 here as $(X_1^T X_1 + cI)^{-1} X_1^T y$ without rescaling X_1. The bias in this estimator is

$$\begin{aligned} & (X_1^T X_1 + cI)^{-1} X_1^T E(y) - \gamma_1 \\ = \; & (X_1^T X_1 + cI)^{-1} X_1^T (X_1 \gamma_1 + X_2 \gamma_2) - \gamma_1 \\ = \; & (X_1^T X_1 + cI)^{-1} \{ (X_1^T X_1 + cI - cI)\gamma_1 + X_1^T X_2 \gamma_2 \} - \gamma_1 \\ = \; & (X_1^T X_1 + cI)^{-1} (-c\gamma_1 + X_1^T X_2 \gamma_2) \\ = \; & \begin{bmatrix} \frac{1}{6+c} & 0 \\ 0 & \frac{1}{58+c} \end{bmatrix} \left\{ -c \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 0 \\ -13 \end{bmatrix} \beta_2 \right\}, \end{aligned}$$

so, in particular, the bias in the estimator of β_1 is

$$-\frac{c\beta_1 + 13\beta_2}{58 + c}.$$

The covariance matrix in the ridge estimator of γ_1 is, from the standard formula for ridge regression estimation,

$$(X_1^T X_1 + cI)^{-1} X_1^T X_1 (X_1^T X_1 + cI)^{-1} \sigma^2.$$

Since $X_1^T X_1$ is diagonal, we see at once that the variance of the estimator of β_1 is

$$\frac{58}{(58 + c)^2}\sigma^2.$$

As usual, the mean squared error is computed from the formula

$$\begin{aligned} \text{MSE} &= \text{Bias}^2 + \text{Variance} \\ &= \left(\frac{c\beta_1 + 13\beta_2}{58 + c}\right)^2 + \frac{58}{(58 + c)^2}\sigma^2. \end{aligned}$$

3. (a) [16] In the first place, we can see quickly from the table of R^2 values that it suffices to confine attention to five models if we include (i) the model with no covariates, the others being (ii) x3 alone, (iii) x1 and x3, (iv) x1, x2 and x3, (v) x1, x2, x3 ad x4. Note that either forward or backward selection would consider only these models. Since the total sum of squares $\sum(y_i - \bar{y})^2$ is .18918, we can compute the SSE for each of the five models from $SSE = (1 - R^2).18918$. Since we also have

$$AIC = n \log \frac{SSE}{n} + 2p, \qquad BIC = n \log \frac{SSE}{n} + p \log n,$$

the complete table is as in Table 4.

Variables	R^2	SSE	DF	MSE	AIC	BIC
None	0	.18918	15	.01261	–69.00	–68.23
x3	.7351	.05011	14	.00358	–88.26	–86.71
x1 x3	.8265	.03282	13	.00252	–93.03	–90.71
x1 x2 x3	.8470	.02894	12	.00241	–93.04	–89.95
x1 x2 x3 x4	.8490	.02857	11	.00260	–91.25	–87.39

Table 4: Solution for problem 3(a)

For testing one row of the table against the next, the successive F statistics are (for row 1 against row 2) $F =$

$(.18918 - .05011)/(.05011/14) = 38.85$, and then successively 6.85, 1.61, 0.14. The first two are significant but the rest are not. (Although you were unable to calculate actual p-values during the exam, they are .00002, .02, .23, .72.) Therefore, the preferred model by either forward or backward selection is (x1, x3). BIC also selects this model but AIC *just* prefers the model with x1, x2 and x3.

(b) [7] Assuming we really do fit the full model: the VIFs are not large enough to indicate a problem (largest is 3.7, for x2). However the three largest condition indexes are all over 100, with a largest of 238, and this does indicate some problem with multicollinearity. It appears that the intercept, x2, x3 and x4 all contribute substantial variance to the component of $X^T X$ with smallest eigenvalue, so from this, it is not possible to be more specific about the nature of the problem.

(c) [10] Recalling $p = 5, n = 16$, the critical values for the various diagnostics are: for h_i, $2p/n = .625$; for studentized residuals, $t_{10,.975} = 2.228$; for $|DFFITS|$, $2\sqrt{p/n} = 1.118$, for $|DFBETAS|$, $2/\sqrt{n} = 0.5$, for $|1 - COVRATIO|$, $3p/n = .9375$. Therefore, the outlying or influential observations measured by various criteria are:

i. h_i: 3
ii. RStudent: 3
iii. $DFFITS$: 1, 3
iv. $DFBETAS$: 1,3,5,8 (x1); 1,3 (x2); 1,3,8,16 (x3); 1 (x4)
v. $COVRATIO$: 2,6,13,14

Taking all the criteria together, it looks as though observation 3 is the only one for which we need to be seriously concerned about the issue of outliers and influence.