STAT 420 Spring 2014

Homework 4: Due March 6 by 7:00pm

Exercise 1

DO NOT use a computer for this problem. Consider the following salary data for 10 tenure-track professors in the Department of Psychology at Anytown State University in thousands of dollars per year (y), x_1 (years at Anytown State University), and x_2 (number of publications in the last 5 years).

salary (y)	44	46	46	61	62	63	64	69	79	86
years (x_1)	3	1	1	4	5	2	3	4	2	5
publications (x_2)	4	3	5	1	4	2	2	5	8	6

Consider the multiple linear regression model: $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$ with $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

In this case we have that

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 10 & 30 & 40 \\ 30 & 110 & 120 \\ 40 & 120 & 200 \end{pmatrix} \qquad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 620 \\ 1960 \\ 2600 \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.95 & -0.15 & -0.1 \\ -0.15 & 0.05 & 0 \\ -0.1 & 0 & 0.025 \end{pmatrix} \qquad \sum_{i=1}^{10} (y_i - \hat{y}_i)^2 = 896 \qquad \sum_{i=1}^{10} (y_i - \bar{y})^2 = 1756$$

(a) Obtain the least-squares estimates $\hat{b}_0$, $\hat{b}_1$, and $\hat{b}_2$.

(b) Perform the significance of the regression test at a 10% level of significance. Specify the null and the alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision.

(c) Construct a 90% confidence interval for b_1.

(d) Test $H_0 : b_1 = 0$ versus $H_1 : b_1 \neq 0$ at a 10% level of significance.

(e) Test $H_0 : b_2 = 5$ versus $H_1 : b_2 < 5$ at a 10% level of significance. Report the value of the test statistic, the critical value(s), and the decision.

(f) Construct a 90% prediction interval for the salary of a tenure-track professor who has been at Anytown State University for 2 years and published 4 times in the last 5 years.

Exercise 2

DO NOT use a computer for this problem.

The data below describe 8 houses sold recently in Anytown.

selling price: y	135	220	190	215	260	245	265	310
size (thousands sq. feet): x_1	1.6	1.7	1.8	2.1	1.9	2.2	2.3	2.4
backyard (1=yes): x_2	0	1	1	0	1	1	0	1
# bedrooms: x_3	2	2	2	3	3	3	4	5

Consider the multiple linear regression model: $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + e_i$ with $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

In this case we have that

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 8 & 16 & 5 & 24 \\ 16 & 32.6 & 10 & 50 \\ 5 & 10 & 5 & 15 \\ 24 & 50 & 15 & 80 \end{pmatrix} \qquad \mathbf{X}'\mathbf{y} = \begin{pmatrix} 1840 \\ 3770 \\ 1225 \\ 5860 \end{pmatrix} \qquad \hat{\mathbf{b}} = \begin{pmatrix} 15 \\ 50 \\ 40 \\ 30 \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 17.08333 & -12.5 & -0.33333 & 2.75 \\ -12.5 & 10 & 0 & -2.5 \\ -0.33333 & 0 & 0.53333 & 0 \\ 2.75 & -2.5 & 0 & 0.75 \end{pmatrix} \qquad \sum_{i=1}^{8}(y_i - \hat{y}_i)^2 = 2000 \qquad \sum_{i=1}^{8}(y_i - \bar{y})^2 = 19700$$

(a) Perform the significance of the regression test at a 5% level of significance. Specify the null and the alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision.

(b) Test $H_0 : b_2 = 0$ versus $H_1 : b_2 \neq 0$ at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision.

(c) Test $H_0 : b_1 = b_2 = 0$ vs. $H_1 : b_j \neq 0$ for some $j \in \{1, 2\}$ at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision.
Hint: the Null model is the simple linear regression model. You have

$$\sum_{i=1}^{8} x_{i3}, \qquad \sum_{i=1}^{8} x_{i3}^2, \qquad \sum_{i=1}^{8} y_i, \qquad \sum_{i=1}^{8} x_{i3} y_i$$

Find SSE for the null model to perform test.

(d) Construct a 95% prediction interval for the selling price of a house that has 2 thousand square feet, a backyard, and 3 bedrooms.

Exercise 3

Suppose the multiple linear regression model: $y_i = b_0 + \sum_{j=1}^{8} b_j x_{ij} + e_i$ with $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ was fit to $n = 36$ data points. Furthermore, suppose that we have

```
> sum( lm( y ~ 1 )$residuals^2 )
[1] 204
> sum( lm( y ~ x1+x3+x5+x6+x7)$residuals^2 )
[1] 138
> sum( lm( y ~ x1+x2+x3+x4+x5+x6+x7+x8)$residuals^2 )
[1] 108
```

(a) Perform the significance of the regression test at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision.

(b) Test $H_0 : b_2 = b_4 = b_8 = 0$ at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision.