620-371: Linear Models

Assignment 3

Due: Monday, 25th May, 2009

This assignment is worth 5% of your total mark.

You may use R (and only R) for this assignment, but only for questions 1 and 2e. You may not use the \texttt{lm} function for question 1, but you may use it for question 2e. If you do use R, include a computer printout of your commands and R output.

1. We are interested in examining the yield of tomato plants that have been grown with certain types of fertiliser. A study is conducted and the following data obtained:

<table>
<thead>
<tr>
<th>Fertiliser</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>33</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>37</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>38</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We fit the model

\[ y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \]

where \( \mu \) is the overall mean and \( \tau_i \) is the effect of using the \( i \)th fertiliser.

(a) [2 marks] Find a conditional inverse for \( X^T X \), using the algorithm given in the lecture slides.

**Solution:**

\begin{verbatim}
> X <- matrix(c(rep(1, 17), rep(0, 12), rep(1, 4), rep(0, 12),
+                 rep(1, 3)), 12, 4)
> y <- as.vector(c(43, 45, 47, 46, 48, 33, 37, 38, 35, 54, 54,
+                 57))
> t(X) %*% X
[1,] 12  5  4  3
[2,]  5  5  0  0
[3,]  4  0  4  0
[4,]  3  0  0  3
\end{verbatim}
A conditional inverse is

\[
(X^T X)^c = \begin{bmatrix}
0 & 0 & 0 \\
0 & \frac{1}{5} & 0 \\
0 & 0 & \frac{1}{4} \\
0 & 0 & 0
\end{bmatrix},
\]

(b) [3 marks] Find two solutions to the normal equations, using only the conditional inverse you found in question 1a.

Solution:

\[
\begin{align*}
&> \text{XtXc <- diag(c(0, 1/5, 1/4, 1/3))} \\
&> \text{I <- diag(rep(1, 4))} \\
&> \text{b <- XtXc %*% t(X) %*% y} \\
&> \text{b} \\
&\quad [,1] \\
&\quad [1,] 0.00 \\
&\quad [2,] 45.80 \\
&\quad [3,] 35.75 \\
&\quad [4,] 55.00 \\
&> \text{b2 <- b + (I - XtXc %*% t(X) %*% X) %*% as.vector(c(1, 0, 0, + 0))} \\
&> \text{b2} \\
&\quad [,1] \\
&\quad [1,] 1.00 \\
&\quad [2,] 44.80 \\
&\quad [3,] 34.75 \\
&\quad [4,] 54.00
\end{align*}
\]

(c) [2 marks] Is \( \mu + \tau_1 - \tau_2 + \tau_3 \) estimable?

Solution: Yes; \( \mu + \tau_1 \) is estimable as it is an element of \( \mathbf{X}\beta \), and \( \tau_2 - \tau_3 \) is a treatment contrast, so their sum is estimable.

(d) [3 marks] Find a 95\% confidence interval for the estimable quantity \( \tau_2 - \tau_3 \).

Solution:

\[
\begin{align*}
&> \text{n <- 12} \\
&> \text{r <- 3} \\
&> \text{tt <- as.vector(c(0, 0, 1, -1))} \\
&> \text{s2 <- sum((y - X %*% b)^2)/(n - r)} \\
&> \text{hw <- qt(0.975, df = n - r) * sqrt(s2 * t(tt) %*% XtXc %*% tt)} \\
&> \text{c(tt %*% b - hw, tt %*% b + hw)} \\
\end{align*}
\]

(e) [2 marks] Test the hypothesis that fertiliser has no effect on yield.

Solution: This hypothesis can be written as \( H_0 : C\beta = 0 \), where
\[
C = \begin{bmatrix}
0 & 1 & -1 & 0 \\
0 & 0 & 1 & -1
\end{bmatrix}.
\]

```r
> library(car)
> C <- matrix(c(0, 0, 1, 0, -1, 1, 0, -1), 2, 4)
> Fstat <- (t(C %*% b) %*% inv(C %*% XtXc %*% t(C)) %*% C %*% b/2)/s2
> Fstat
[,1]
[1,] 81.6076
> pf(Fstat, 2, n - r, lower.tail = FALSE)
[,1]
[1,] 1.705178e-06
```

We reject the null hypothesis firmly.

2. We study the amount of rotting of a potato exposed to a variety of levels of oxygen, and a variety of temperatures. A small experiment is conducted and the following data obtained:

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Oxygen level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>16</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>18</td>
</tr>
</tbody>
</table>

We fit the model

\[
y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk},
\]

where \( \mu \) is the overall mean, \( \tau_i \) is the effect of the \( i \)th oxygen level, and \( \beta_j \) is the effect of the \( j \)th temperature level. A model was fitted in R using the `lm` function, from which the following output was derived:

```r
> options(contrasts = c("contr.treatment", "contr.poly"))
> model <- lm(rot ~ oxygen.f + temp.f, data = potato)
> summary(model)

Call:
lm(formula = rot ~ oxygen.f + temp.f, data = potato)

Residuals:

Min     1Q Median     3Q    Max
-10.4444 -2.8611  0.4444  3.0278  8.1111

Coefficients:
```
(Intercept) 10.222 2.410 4.242 0.000820 ***
oxygen.f2 -3.333 2.951 -1.130 0.277660
oxygen.f3 -3.333 2.951 -1.130 0.277660
temp.f16 11.556 2.410 4.796 0.000285 ***

---
Signif. codes: 0 ^ a˘A¨Y***^ a˘A´Z 0.001 ^ a˘A¨Y**^ a˘A´Z 0.01 ^ a˘A¨Y*^ a˘A´Z 0.05 ^ a˘A¨Y.^ a˘A´Z 0.1 ^ a˘A¨Y ^ a˘A´Z 1

Residual standard error: 5.111 on 14 degrees of freedom
Multiple R-squared: 0.6382, Adjusted R-squared: 0.5607
F-statistic: 8.233 on 3 and 14 DF, p-value: 0.002105

> anova(model)

Analysis of Variance Table
Response: rot

Df Sum Sq Mean Sq F value Pr(>F)
oxygen.f 2 44.44 22.22 0.8505 0.4481124
temp.f 1 600.89 600.89 22.9988 0.0002849 ***
Residuals 14 365.78 26.13

---
Signif. codes: 0 ^ a˘A¨Y***^ a˘A´Z 0.001 ^ a˘A¨Y**^ a˘A´Z 0.01 ^ a˘A¨Y*^ a˘A´Z 0.05 ^ a˘A¨Y.^ a˘A´Z 0.1 ^ a˘A¨Y ^ a˘A´Z 1

> linear.hypothesis(model, c(0, 1, 0, -1), 0)

Linear hypothesis test

Hypothesis:
oxygen.f2 - temp.f16 = 0

Model 1: rot ~ oxygen.f + temp.f
Model 2: restricted model

Res.Df RSS Df Sum of Sq F Pr(>F)
1 14 365.78
2 15 764.80 -1 -399.02 15.272 0.001577 **

---
Signif. codes: 0 ^ a˘A¨Y***^ a˘A´Z 0.001 ^ a˘A¨Y**^ a˘A´Z 0.01 ^ a˘A¨Y*^ a˘A´Z 0.05 ^ a˘A¨Y.^ a˘A´Z 0.1 ^ a˘A¨Y ^ a˘A´Z 1

(a) [2 marks] What information do you gather about the parameters of the model in equation 1, from the coefficients of the fitted R model?
**Solution:** From the coefficients, we deduce that

\[
\begin{align*}
\mu + \tau_1 + \beta_1 & \approx 10.22 \\
\tau_2 - \tau_1 & \approx -3.33 \\
\tau_3 - \tau_1 & \approx -3.33 \\
\beta_2 - \beta_1 & \approx 11.56
\end{align*}
\]

(b) [1 mark] Should we accept or reject the hypothesis that oxygen level has no effect on rot?

**Solution:** We should accept it; the \( p \)-value is 0.448.

(c) [1 mark] Should we accept or reject the hypothesis that temperature has no effect on rot?

**Solution:** We should reject it; the \( p \)-value is 0.00028.

(d) [2 marks] What hypothesis is the `linear.hypothesis` function testing, given in terms of the parameters of the model in equation 1?

**Solution:** The hypothesis is testing \( \tau_2 - \tau_1 = \beta_2 - \beta_1 \); in other words, whether the difference in effect between the temperature levels is equal to the difference in effect between the first two oxygen levels.

(e) [2 marks] Using R, test for the presence of interaction between the factors.

**Solution:**

```r
> model2 <- lm(rot ~ oxygen.f * temp.f, data = potato)
> anova(model2)
```

Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oxygen.f</td>
<td>2</td>
<td>44.44</td>
<td>22.22</td>
<td>0.7634</td>
<td>0.487453</td>
</tr>
<tr>
<td>temp.f</td>
<td>1</td>
<td>600.89</td>
<td>600.89</td>
<td>20.6412</td>
<td>0.000674 ***</td>
</tr>
<tr>
<td>oxygen.f:temp.f</td>
<td>2</td>
<td>16.44</td>
<td>8.22</td>
<td>0.2824</td>
<td>0.758816</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>349.33</td>
<td>29.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signif. codes: 0 \text{**} \text{***} 0.001 \text{***} \text{***} 0.01 \text{**} 0.05 \text{*} 0.1 \text{.}

There clearly is no interaction — the \( p \)-value is 0.76.

3. [4 marks] Show that for any matrix \( A \), the matrix \( A(A^TA)^{-}A^T \) is invariant to the choice of a conditional inverse (i.e. unique). You may use the property of a general matrix \( M \) that if \( M^TM = 0 \), then \( M = 0 \). (Hint: Let \( (A^TA)^{-}_1 \) and \( (A^TA)^{-}_2 \) be two different conditional inverses for \( A^TA \) and show that \( A(A^TA)^{-}_1 A^T - A(A^TA)^{-}_2 A^T = 0 \).)

**Solution:** Let \( M = A(A^TA)^{-}_1 A^T - A(A^TA)^{-}_2 A^T \). Then
\[ M^T M = (A^T A)_{c1}^T (A^T A)_{c2} - A^T A)_{c1} A^T A^T - A^T A)_{c2} A^T \] 
\[ = A^T A)_{c1}^T A^T A)_{c1} A^T - A^T A)_{c2} A^T A)_{c1} A^T \]
\[ - A^T A)_{c1}^T A^T A)_{c2} A^T + A^T A)_{c2} A^T A)_{c2} A^T \]
\[ = A^T A)_{c1} A^T - A^T A)_{c2} A^T - A^T A)_{c2} A^T + A^T A)_{c2} A^T \]
\[ = 0. \]

Hence \( M = 0 \) and the result follows.