620-371: Linear Models

Practice Class 10

12th May, 2009

In a manufacturing plant, filters are used to remove pollutants. We are interested in comparing the lifespan of 5 different types of filters. Six filters of each type are tested, and the time to failure in hours is given in the dataset (on the website) `filters` (in `csv` format).

1. Use the `read.csv` function to read the data. Then convert the `type` component into a factor.

   Solution:

```
> filters <- read.csv("../data/filters.csv")
> filters$type <- factor(filters$type)
```

2. Construct X and $\mathbf{y}$ matrices for this linear model.

   Solution:

```
> y <- filters$life
> X <- matrix(0, 30, 6)
> X[, 1] <- 1
> X[filters$type == 1, 2] <- 1
> X[filters$type == 2, 3] <- 1
> X[filters$type == 3, 4] <- 1
> X[filters$type == 4, 5] <- 1
> X[filters$type == 5, 6] <- 1
```

3. Using the algorithm given in the lecture slides, find a conditional inverse for $X^T X$.

   Solution:

```
> library(car)
> XtXc <- matrix(0, 6, 6)
> XtXc[2:6, 2:6] <- inv((t(X) %*% X)[2:6, 2:6])
> XtXc

     [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,]    0 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
[2,]    0 0.1666667 0.0000000 0.0000000 0.0000000 0.0000000
[3,]    0 0.0000000 0.1666667 0.0000000 0.0000000 0.0000000
[4,]    0 0.0000000 0.0000000 0.1666667 0.0000000 0.0000000
[5,]    0 0.0000000 0.0000000 0.0000000 0.1666667 0.0000000
[6,]    0 0.0000000 0.0000000 0.0000000 0.0000000 0.1666667
```

4. Use `ginv` to find another conditional inverse for $X^T X$.

   Solution:

```
> library(MASS)
> XtXc2 <- ginv(t(X) %*% X)
```

5. Verify that $X(X^T X)^c X^T$ is the same for your two conditional inverses.

   Solution:

```
> sum((X %*% XtXc %*% t(X) - X %*% XtXc2 %*% t(X))^2)

[1] 5.650678e-31
```

6. Find two solutions for the normal equations.

   Solution:

```
> b <- XtXc %*% t(X) %*% y
> b

         [,1]
[1,]   0.0000
[2,] 249.1667
[3,] 187.5000
[4,] 166.0000
[5,] 357.3333
[6,] 361.3333

> b2 <- XtXc2 %*% t(X) %*% y
> b2

          [,1]
[1,] 220.22222
[2,]  28.94444
[3,] -32.72222
[4,] -54.22222
[5,] 137.11111
[6,] 141.11111
```

7. Express one of your solutions in terms of the other.

   Solution: $\mathbf{b}_2 = \mathbf{b} + (I - (X^T X)^c X^T X)\mathbf{b}_2$.

```
> diag(rep(1, 6)) - XtXc %*% t(X) %*% X

     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]   -1    0    0    0    0    0
[3,]   -1    0    0    0    0    0
[4,]   -1    0    0    0    0    0
```

```
[5,]   -1    0    0    0    0    0
[6,]   -1    0    0    0    0    0
```

8. Is $\mu - \tau_1 + \tau_5$ estimable?

   Solution:

```
> tt <- as.vector(c(1, -1, 0, 0, 0, 1))
> t(tt) %*% XtXc %*% t(X) %*% X

     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0   -1    0    0    0    1
```

   No, it is not estimable.

9. Is $\tau_1 - \frac{1}{2}\tau_3 - \frac{1}{2}\tau_4$ estimable?

   Solution: Yes, as it is a treatment contrast.

10. Verify that your two solutions for the normal equations produce the same estimate of $\tau_4 - \tau_5$.

    Solution:

```
> b[5] - b[6]

[1] -4

> b2[5] - b2[6]

[1] -4
```

11. Calculate s^2.

    Solution:

```
> n <- dim(filters)[1]
> r <- 5
> e <- y - X %*% b
> s2 <- sum(e^2)/(n - r)
> s2

[1] 15304.2
```

12. Calculate a 95% confidence interval for the difference in lifespan between filter types 3 and 4.

    Solution:

```
> tt <- as.vector(c(0, 0, 0, 1, -1, 0))
> hw <- qt(0.975, df = n - r) * sqrt(s2 * t(tt) %*% XtXc %*% tt)
> c(tt %*% b - hw, tt %*% b + hw)
```

```
[1] -338.43399  -44.23268
```

13. Fit a `lm` model to the data and verify your answers to questions 11 and 12.

    Solution:

```
> options(contrasts = c("contr.treatment", "contr.poly"))
> model <- lm(life ~ type, data = filters)
> deviance(model)/model$df.residual

[1] 15304.2

> library(gmodels)
> estimable(model, c(0, 0, 1, -1, 0), conf.int = 0.95)

              Estimate Std. Error   t value DF   Pr(>|t|) Lower.CI  Upper.CI
(0 0 1 -1 0) -191.3333   71.42409 -2.678835 25 0.01287281 -338.434 -44.23268
```