MATH 668-01 Homework 5

Due: Thursday, April 3, 2018

Instructions: *Each of the following problems must be submitted to the instructor on or before the due date. Partial credit may be given for incorrect answers which make some positive progress. Late homework will not be accepted.*

1. (10 points) For the data set "HeightData.txt" consider the multiple linear regression model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i$$

for $i = 1, \ldots, n$ where the response variable y_i is the height of the ith child, the explanatory variables are

$$\begin{aligned} x_{i1} &= g_i \\ x_{i2} &= g_i m_i \\ x_{i3} &= g_i f_i \\ x_{i4} &= 1 - g_i \\ x_{i5} &= (1 - g_i) m_i \\ x_{i6} &= (1 - g_i) f_i \end{aligned}$$

where

$$\begin{aligned} g_i &= \text{indicator for female gender (1 if the } i^{th} \text{ gender is female, 0 if male)} \\ m_i &= \text{height of the mother of the } i^{th} \text{ child} \\ f_i &= \text{height of the father of the } i^{th} \text{ child,} \end{aligned}$$

and $\varepsilon_1, \ldots, \varepsilon_n$ are independent $N(0, \sigma^2)$ random variables.

(a - 3 pts) Find the maximum likelihood estimate of $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{pmatrix}$ under the constraints that $\beta_2 = \beta_5$ and $\beta_3 = \beta_6$.

(b - 2 pts) Denote the constrained maximum likelihood estimate in part (a) as $\hat{\boldsymbol{\beta}}_c$. Show that

$$F = \frac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_c)\|^2/2}{\|\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2/(n-6)} = 0.407,$$

as computed in Example 8.4.1.

(c - 5 pts) Find values A, B, C, D, and E such that

$$A(\beta_2 - \beta_5)^2 + B(\beta_3 - \beta_6)^2 + C(\beta_2 - \beta_5)(\beta_3 - \beta_6) + D(\beta_2 - \beta_5) + E(\beta_3 - \beta_6) \leq 1$$

is a 95% confidence ellipse for $\begin{pmatrix} \beta_2 - \beta_5 \\ \beta_3 - \beta_6 \end{pmatrix}$.

(Hint: Use the fact that $\dfrac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\boldsymbol{\beta})^\top \left(\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top\right)^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\boldsymbol{\beta})/2}{\|\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2/(n-6)} \sim F_{2,n-6}$ if $y \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.)

2. (10 points) Suppose $\mathbf{X}_{(n)}$ is a $(n-1)\times p$ matrix with rank $p \le n-1$ and $\boldsymbol{y}_{(n)}$ is an $(n-1)$-dimensional vector. Let $\hat{\boldsymbol{\beta}}_{(n)} = \left(\mathbf{X}_{(n)}^\top \mathbf{X}_{(n)}\right)^{-1} \mathbf{X}_{(n)}^\top \boldsymbol{y}_{(n)}$.

Now, suppose that we add one observation and let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_{(n)} \\ \mathcal{X}_n^\top \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_{(n)} \\ y_n \end{pmatrix}$ where $\mathcal{X}_n$ is a p-dimensional vector. Letting $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}$, show that

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{(n)} + \frac{\mathbf{M}\mathcal{X}_n(y_n - \mathcal{X}_n^\top \hat{\boldsymbol{\beta}}_{(n)})}{1 + \mathcal{X}_n^\top \mathbf{M}\mathcal{X}_n}$$

where $\mathbf{M} = (\mathbf{X}^\top\mathbf{X})^{-1}$.

3. (10 points) Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

for the data set “HW5-3.txt”. (You can read the data from the file or create the data in R using the following commands.)

```
set.seed(123)
x0=1:100/100
x=c(x0,10)
y0=sqrt(1+.1*x0+rnorm(100,sd=.0001))
y=c(y0,0)
```

(a - 3 pts) Create a scatterplot with x on the horizontal axis and y on the vertical axis, and superimpose the regression line obtained by the method of least squares.
(b - 3 pts) Compute the residual, the studentized residual, and the externally studentized residual for the last (101st) observation.
(c - 1 pt) Using the data set with the last observation removed, re-create the scatterplot of the data set.
(d - 1 pt) Using the data set with the last observation removed, create a residual plot with the fitted values on the horizontal axis.
(e - 2 pts) Using the data set with the last observation removed, let $u_i = \dfrac{y_i^\lambda - 1}{\lambda}$ be a transformation of the response variable. What is the value of λ obtained using the Box-Cox method?