Computer Lab. 5 – Week 6

1 Problems

1. Consider the linear regression model

$$\mathbf{X} = Z\boldsymbol{\beta} + \boldsymbol{\mathcal{W}}\,,$$

as discussed on Lecture Slides Chapter 2, Page 11. Show that the Ordinary Least Squares estimators of the regression parameters are:

$$\widehat{\boldsymbol{\beta}} = (Z'Z)^{-1}Z'\mathbf{X}\,.$$

2. As mentioned on Lecture Slides Chapter 2, Page 15,

$$\frac{\widehat{\beta}_i - \beta_i}{\sqrt{c_{ii}MSE}} \sim \mathtt{t}_{(n-q)}\,,$$

where c_{ii} denotes the i^{th} diagonal element of matrix $C = (Z'Z)^{-1}$. Use the results given on Lecture Slides Chapter 2, Pages 14 and 15 to prove this.

3. *Hint for Assignment 3 - Bonus Question.* Consider the statement of Assignment 3 Bonus Question on Lecture Slides Chapter 2, Page 25. Based on information provided on Page 79 of the reference [1], we know that

$$\frac{n\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-k)} \quad \text{and} \quad \frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'Z'Z(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})}{\sigma^2} \sim \chi^2_{(k)}\,.$$

Utilise this information as well as the results given on Lecture Slides Chapter 2, Page 14 to show that

$$\frac{(n-k)\,(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'Z'Z(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})}{kn\,\widehat{\sigma}^2} \sim F_{(k,n-k)}\,.$$

2 Getting Started the Project

1. Open the page `https://www.google.com.au/trends/`.
2. Type a search keyword, say 'wine'.
3. Choose a country, say 'Australia'.
4. Choose a compare search keyword, say 'beer'.
5. Save the data file with extension '.csv' on your computer, say 'test.csv'.
6. Rearrange your data file such that each column contains the heading and data only.
7. Set your current working directory in R where the file has been saved.

```
setwd("C:/")
```

8. Store the data from the file to a variable in R:

   ```
   WBAust <- read.csv("test.csv")
   ```

9. Look at the structure of the variable:

   ```
   head(WBAust)
   ```

10. All students should form groups of size 1-3 to carry out the project. The list of each group members should be sent to 'Ali.Eshragh@newcastle.edu.au' by Wednesday, September 7, 11:55pm.

11. Those students that do not join a group, should attend the consultation session in V239 on Thursday, September 8, 11:30am-12:30pm to discuss the possibility of forming new groups together.

12. Each group should send their suggested data file with a short description to 'Ali.Eshragh@newcastle.edu.au' by Wednesday, October 5, 11:55pm. You are free to find your dataset from any permitted resource. 'Google-Trend' is only one suggestion.

3 Linear Regression in R

1. Find the time range of dataset:

   ```
   WBAust[1,1]
   dim(WBAust)
   WBAust[152,1]
   ```

2. Construct two time series for wine and beer consumption in Australia within the range of observation times:

   ```
   WineTS <- ts(WBAust[,2], start = c(2004,1), end = c(2016,8), frequency = 12)
   WineTS
   BeerTS <- ts(WBAust[,3], start = c(2004,1), end = c(2016,8), frequency = 12)
   BeerTS
   ```

3. Plot the time series of `WineTS` and `BeerTS`:

   ```
   plot.ts(WineTS, type="l")
   plot.ts(BeerTS, type="l")
   plot.ts(cbind(WineTS, BeerTS), type="l")
   ts.plot(cbind(WineTS, BeerTS), type="l")
   ```

4. Construct a time variable for observation times:

   ```
   TimeTS <- time(WineTS)
   TimeTS
   ```

5. Plot the scatterplot matrix:

   ```
   pairs(cbind(Wine=WineTS, Time=TimeTS, Beer=BeerTS))
   ```

6. Fit two regression models

$$\begin{cases} \text{Model 1: } \text{Wine}_t = \beta_1 + \beta_2 t + \beta_3 \text{beer}_t + \mathcal{W}_t \\ \text{Model 2: } \text{Wine}_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \text{beer}_t + \mathcal{W}_t \end{cases} :$$

```
fit1 <- lm(WineTS∼TimeTS + Beer)
TimeTS2 <- TimeTS^2
fit2 <- lm(WineTS∼TimeTS + TimeTS2 + Beer)
```

7. Summary of the regression parameters estimations:

```
summary(fit1)
summary(fit2)
```

8. Compare the values of AIC for two models:

```
AIC(fit1)/152 - log(2*pi))
AIC(fit2)/152 - log(2*pi))
```

4 Assignment 3: R Coding

Submission: *You should email your R-code to 'Ali.Eshragh@newcastle.edu.au' by Sunday, September 11, 11:55pm. Write your full name and student ID No. on the top of your code and your file name must be your complete surname.*

1. Choose a dataset from 'Google-Trends' and run statistical analysis similar to Sections 2 and 3 of this Computer Lab Notes.

References

[1] R.H. Shumway and D.S. Stoffer, *Time Series Analysis and Its Applications With R Examples*, Springer, New York, 2010.