MATH-650 Assignment 6

Saket Choudhary (USCID: 2170058637) (skchoudh@usc.edu)

09/15/2015

Problem 19

```
pH.data <- read.csv('case0702.csv', header=T)
logT <- log(pH.data$Time)
pH.data$logT <- logT
n <- nrow(pH.data)
```

Simple linear regression model for $log(pH)$ is given by:

$$pH = \beta_0 + \beta_1 log(T)$$

Part (a)

```
fit <- lm(pH~logT, data=pH.data)
s <- summary(fit)
b0 <- fit$coefficients[1]
b1 <- fit$coefficients[2]
se0 <- s$coefficients[3]
se1 <- s$coefficients[4]
t0 <- s$coefficients[5]
t1 <- s$coefficients[6]
p0 <- s$coefficients[7]
p1 <- s$coefficients[8]
```

Thus, $\beta_0 = 6.983626(S.E = 0.048532, p - value = 6.0839895 \times 10^{-15})$ and $\beta_1 = -0.7256578(S.E = 0.0344263, p - value = 2.6951582 \times 10^{-8})$

Part (b)

```
Xbar <- mean(log(pH.data$Time))#1.190
sx2 <- var(log(pH.data$Time))#0.6344
mu <- b0 +b1*log(5)
```

Thus, $\hat{\mu}\{Y|log(5)\} = 5.8157249$

Part (c)

```
sigmahat <- 0.08226
se = sigmahat *sqrt(1/n+(log(5)-Xbar)^2/(n-1*sx2))
```

$SE[\hat{\mu}\{Y|log(5)\}] = 0.0283496$

Problem 25

Part (a)

$$SS(\beta_0, \beta_1) = \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\frac{\partial SS}{\partial \beta_0} = -2 \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial SS}{\partial \beta_1} = -2 \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i) X_i$$

$$\frac{\partial SS}{\partial \beta_0} = -2 \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\sum_{i=1}^{N} Y_i = \sum_{i=1}^{N} (\beta_0 + \beta_1 X_i)$$

$$\sum_{i=1}^{N} Y_i = \beta_0 n + \beta_1 \sum_{i=1}^{N} X_i$$

$$\frac{\partial SS}{\partial \beta_1} = -2 \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i) X_i = 0$$

$$\sum_{i=1}^{N} Y_i X_i = \sum_{i=1}^{N} (\beta_0 + \beta_1 X_i) X_i$$

$$\sum_{i=1}^{N} Y_i X_i = \beta_0 \sum_{i=1}^{N} X_i + \sum_{i=1}^{N} \beta_1 X_i^2$$

Part (b)

$$\hat{\beta_0} = \overline{Y} - \hat{\beta_1} \overline{X}$$

And,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_i - \overline{X})^2}$$

$$= \frac{\sum_{i=1}^{N}(X_iY_i - \overline{X}Y_i + \overline{X}\overline{Y} - X_i\overline{Y})}{\sum_{i=1}^{N}(X_i^2 - 2\overline{X}X_i + \overline{X}^2)}$$

$$= \frac{\sum_{i=1}^{N} X_i\,Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{N} X_i^2 - n\overline{X}^2}$$

Thus,

$$\hat{\beta}_1 \sum_{i=1}^{N} X_i^2 - n\hat{\beta}_1\overline{X}^2 = \sum_{i=1}^{N} X_i\,Y_i - n\overline{X}\overline{Y}$$

$$\hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + n\overline{X}\overline{Y} - n\hat{\beta}_1\overline{X}^2 = \sum_{i=1}^{N} X_i\,Y_i$$

$$\hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + n\overline{Y}(\overline{X} - \hat{\beta}_1\overline{X}) = \sum_{i=1}^{N} X_i\,Y_i$$

Substituting (3) in (4), we get

$$\hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + n\overline{X}(\overline{Y} - \hat{\beta}_1\overline{X}) = \sum_{i=1}^{N} X_i\,Y_i$$

$$\hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + n\overline{X}(\hat{\beta}_0)\overline{X} = \sum_{i=1}^{N} X_i\,Y_i$$

$$\hat{\beta}_1 \sum_{i=1}^{N} X_i^2 + \hat{\beta}_0 \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} X_i\,Y_i$$

Thus, (5) is same as (2)

Now, From (3)

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1\overline{X}$$

$$\frac{\sum_{i=1}^{N} Y_i}{N} = \hat{\beta}_0 + \hat{\beta}_1 \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\sum_{i=1}^{N} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_i$$

Thus, (6) is same as (1)

Consider second order differentials:

$$\frac{\partial^2 SS}{\partial \beta_0^2} = -2 * -n = 2n$$

$$\frac{\partial^2 SS}{\partial \beta_0 \partial \beta_1} = 2\sum_{i=1}^{N} X_i$$

$$\frac{\partial^2 SS}{\partial \beta_1 \partial \beta_0} = 2\sum_{i=1}^{N} X_i$$

$$\frac{\partial^2 SS}{\partial \beta_1^2} = 2\sum_{i=1}^{N} X_i^2$$

$$Hessian(H) = \begin{bmatrix} 2 & 2\sum_{i=1}^{N} X_i \\ 2\sum_{i=1}^{N} X_i & 2\sum_{i=1}^{N} X_i^2 \end{bmatrix}$$

$$2n > 0 \text{and,}$$

$$det(H) = 4(n\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2) = 4n(n-1)Var(X) > 0 \text{ always}$$

Thus, the values of $\hat{\beta}_0, \hat{\beta}_1$ indeed guarantee a minima since the Hessian is positive definitee.

Part (c)

Problem 13

Intercept = 0.3991 Standard Error = 0.1185 df = 22 $t_{0.975,22} = 2.073$

Upper limit = 0.3991 + 2.073 * 0.1185 = 0.6447505 Lower limit = 0.3991 - 2.073 * 0.1185 = 0.1534495

Thus, the 95% CI for Intercept is [0.1534495, 0.6447505]

Also following is the R code:

```
intercept <- 0.3991
se <- 0.1185
df <- 22
t975 <- qt(0.975,df)
limit.upper <- intercept + t975*se
limit.lower <- intercept - t975*se
limit.upper

## [1] 0.644854
```

```
limit.lower

## [1] 0.153346
```