HW4_code

Hengchang Hu

10/2/18

Chapter 2.2

(a)

```
indicators <- read.table("C:/Users/hugo1/Documents/MA575/HW4/indicators.txt", header = TRUE)
# import the dataset

OLS_priceloan <- lm(PriceChange~LoanPaymentsOverdue, indicators)
# get the least square model

esti_beta1_pl <- coef(summary(OLS_priceloan))[, "Estimate"]["LoanPaymentsOverdue"][[1]]

print(paste("The estimation of beta_1 is ", esti_beta1_pl))
```

```
## [1] "The estimation of beta_1 is  -2.24851978357815"
```

```
# show estimation value of beta_1

std_beta1_pl <- coef(summary(OLS_priceloan))[, "Std. Error"]["LoanPaymentsOverdue"][[1]]

print(paste("The Standard error of beta_1 is ", std_beta1_pl))
```

```
## [1] "The Standard error of beta_1 is  0.903311332218041"
```

```
# show standard error value of beta_1

print(paste("The 95% CI of beta_1 is [", esti_beta1_pl -
              qt(0.975,16) * std_beta1_pl, esti_beta1_pl
            + qt(0.975,16) * std_beta1_pl, "]"))
```

```
## [1] "The 95% CI of beta_1 is [ -4.16345426359378 -0.333585303562513 ]"
```

```
# 95% confidence interval of beta_1
```

From above we can tell that there is evidence of a significant negative linear association.

(b)

```
newdata <- list("LoanPaymentsOverdue" = 4) # type in newdata

esti_pl_4 <- predict(OLS_priceloan, newdata)[[1]] # estimation value

print(paste("When X = 4, E(Y|X) = ", esti_pl_4)) # show estimation value
```

```
## [1] "When X = 4, E(Y|X) =  -4.47958536542933"
```

```
S_pl <- summary(OLS_priceloan)$sigma # get the estimate of sigma

SXX_pl <- (S_pl / std_beta1_pl) ^ 2 # get SXX

lowerbound_pl <- esti_pl_4 - qt(0.975, 16) * S_pl * sqrt(1 / 18 +
    (4 - mean(indicators["LoanPaymentsOverdue"][,1])) ^ 2 / SXX_pl)
# get the lower bound of 95% CI for E(Y|X=4)

upperbound_pl <- esti_pl_4 + qt(0.975, 16) * S_pl * sqrt(1 / 18 +
    (4 - mean(indicators["LoanPaymentsOverdue"][,1])) ^ 2 / SXX_pl)
# upper bound

print(paste("95% CI for E(Y|X=4) is [", lowerbound_pl, upperbound_pl, "]"))
```

```
## [1] "95% CI for E(Y|X=4) is [ -6.64884919960484 -2.31032153125382 ]"
```

```
# 95% CI for E(Y|X=4)
```

From above we can tell that 0% is not in the 95% CI for E(Y|X=4), so it is not a feasible value.

Chapter 2.3

(a)

Through the output of R, $\hat{\beta}_0$ is 0.6417099, $se(\hat{\beta}_0)$ is 0.1222707, Degree of Freedom is 28. Therefore, the 95% CI for β_0 is [0.3912497, 0.8921701].

(b)

From the output of R, we can know that $\hat{\beta}_1 = 0.0112916$, and the $se(\hat{\beta}_1) = 0.0008184$. Since the statistics $T = \frac{(\hat{\beta}_1 - \beta_1)}{se(\hat{\beta}_1)}$ = 1.5782014, and the acceptance region is $\left[t(0.025, n-2), t(0.975, n-2)\right]$ = [-2.0484071, 2.0484071]. Therefore the T statistics fall into acceptance region, can not reject the null hypothesis.

(c)

When $X = 130$, $E(Y|X = 130) = 0.6417099 + 0.0112916 \times 130$ = 2.1096179. And the 95% PI for $E(Y|X = 130)$ is

$$\hat{y}_p \pm t_{(1-a/2, n-2)} \times S\sqrt{(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x_i - \bar{x})^2)}}$$

which is [1.4228861, 2.7963497]

Chapter 2.6

(a)

$$\begin{aligned}(y_i - \hat{y}_i) &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\ &= y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i \\ &= (y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})\end{aligned}$$

(b)

$$\begin{aligned}(\hat{y}_i - \bar{y}) &= \hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y} \\ &= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y} \\ &= \hat{\beta}_1 (x_i - \bar{x})\end{aligned}$$

(c)

$$\begin{aligned}\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^{n} \big[(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})\big] \hat{\beta}_1 (x_i - \bar{x}) \\ &= \hat{\beta}_1 \Big(\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big) \\ &= \hat{\beta}_1 \Big(SXY - \frac{SXY}{SXX} SXX \Big) \\ &= 0\end{aligned}$$