Homework 4_1

Jyoti Chaudhary

February 24, 2018

Solution 1.a

```
# Fitting the MLR model using LM function

xr16054 <- read.csv(paste(getwd(),"/xr16054.csv",sep = ""),header=T)

xr16054_fit <- lm(formula = xr16054$Salary ~ xr16054$GPA + xr16054$Activities, data = xr16054)

summary(xr16054_fit)
```

```
##
## Call:
## lm(formula = xr16054$Salary ~ xr16054$GPA + xr16054$Activities,
##     data = xr16054)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1087 -0.6306 -0.1198  0.5621  2.2754
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         24.3092     3.1919   7.616 0.000125 ***
## xr16054$GPA          3.8416     1.2342   3.113 0.017016 *
## xr16054$Activities   1.6810     0.5291   3.177 0.015560 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.448 on 7 degrees of freedom
## Multiple R-squared:  0.8245, Adjusted R-squared:  0.7743
## F-statistic: 16.44 on 2 and 7 DF,  p-value: 0.002267
```

The MLR equation from above R output is:

Y-hat = 24.3092 + 3.8416 * X1 + 1.6810 * X2

where Beta0 = 24.3092 Beta1 = 3.8416 Beta2 = 1.6810

X1 = GPA X2 = Activities Y-hat = Predicted Salary

Interpretation of regression Coefficients: a. For a unit increase in GPA and a constant Activities, the starting salary will increase by 3.8416 units. b. For a unit increase in number of Activities with constant age, the starting salary will increase by 1.6810 units. c. As per the regression equation, if GPA is zero and number of activities is zero, the starting salary equals 24.3092. This interpretation is invalid as the parameter values in this case are outside the range of test data.

Solution (1.b)

For X1 = 3.6 and X2 = 3, the estimated salary will be:

Y-predicted = 24.3092 + 3.8416 * 3.6 + 1.6810 * 3 = 43.18196

Hence, Dave's estimated starting salary = 43.18196

Solution (1.c)

```r
# Below is the code to calculate SSE = e'e where e = (y - y-hat) = Observed y minus predicted y

# code to prepare a 10 by 1 matrix for regressor GPA and ACTIVITIES
matrix_X1 <- matrix(xr16054$GPA)
matrix_X2 <- matrix(xr16054$Activities)
# Code to prepare a 10 by 1 matrix
matrix_Beta0 <- matrix(rep(1,length(xr16054$Activities)))
# code to combine columns of all the above matrices
X <- cbind(matrix_Beta0, matrix_X1, matrix_X2)
# Matrix for the response variable
Y <- matrix(xr16054$Salary)

Beta_hat <- solve(t(X)%*%X)%*%(t(X)%*%Y)

# As e = y - (X * Beta-hat)
e <- Y -(X%*%Beta_hat)
# SSE = e'e
sse <- t(e) %*% (e)

# Standard Error = sqrt(SSE / n - (k + 1))
SE <- sqrt(sse/(10-(2+1)))
SE
```

```
##          [,1]
## [1,] 1.447934
```

From above output, it can be seen that Standard Error for the model = 1.448. This value can also be found from R output: "Residual standard error: 1.448"

Solution 3.a

```r
# Fitting the MLR model

StreetVN <- read.sas7bdat(paste(getwd(),"/STREETVN.sas7bdat",sep = ""))

StreetVN_fit <- lm(formula = StreetVN$earnings ~ StreetVN$age + StreetVN$hours, data = StreetVN)

summary(StreetVN_fit)
```

```
##
## Call:
## lm(formula = StreetVN$earnings ~ StreetVN$age + StreetVN$hours,
##     data = StreetVN)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1105.1  -322.1   -61.0   331.9   721.2
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -20.352    652.745  -0.031  0.97564
## StreetVN$age      13.350      7.672   1.740  0.10738
## StreetVN$hours   243.714     63.512   3.837  0.00236 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 547.7 on 12 degrees of freedom
## Multiple R-squared:  0.5823, Adjusted R-squared:  0.5126
## F-statistic: 8.363 on 2 and 12 DF,  p-value: 0.005314
```

Solution 3.c

From above model fit summary output, the regression equations =

Y-hat = -20.352 + 13.350 * X1 + 243.714 * X2 where Beta0 = -20.352 Beta1 = 13.350 Beta2 = 243.714

X1 = Age X2 = Hours Y-hat = Predicted Earnings

Interpretation of regression Coefficients: a. For a unit increase in age, keeping all other regressors constant, the Earnings of Street Vendors increase by a factor of 13.350

b. For a unit increase in hours worked per day, keeping all other regressors constant, the Earnings of Street Vendors increase by a factor of 243.714

c. for age = zero and hours worked per day equals zero, the estimated earnings will be -20.352. This interpretation has no practical significance since the age and hours worked per day are outside the range of test data used for model fitting and is not even practically feasible as age and hours worked cannot be zero.

Solution 3.9 →

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

where y = mean / predicted annual Earnings
x_1 = age
x_2 = hours worked

Solution 3.6 →

From the R-output for model fit,

$$y = -20.352 + 13.35 x_1 + 243.714 x_2$$

Solution 2.9 →

* The linear mode equation $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$ can be written in matrix form as below :—

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}_{4\times1} = \begin{pmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ 1 & X_{31} & X_{32} \\ 1 & X_{41} & X_{42} \end{pmatrix}_{4\times3} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}_{3\times1} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}_{4\times1}$$

After inputting the data →

$$\begin{pmatrix} 40 \\ 46 \\ 38 \\ 39 \end{pmatrix}_{4\times1} = \begin{pmatrix} 1 & 3.2 & 2 \\ 1 & 3.6 & 5 \\ 1 & 2.8 & 3 \\ 1 & 2.4 & 4 \end{pmatrix}_{4\times3} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}_{3\times1} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}_{4\times1}$$

Sol 2.b → x- Matrix

$$\begin{bmatrix} 1 & 3.2 & 2 \\ 1 & 3.6 & 5 \\ 1 & 2.8 & 3 \\ 1 & 2.4 & 4 \end{bmatrix}_{4\times3}$$

Sol 2.c → X' matrix (transpose)

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 3.2 & 3.6 & 2.8 & 2.4 \\ 2 & 5 & 3 & 4 \end{bmatrix}_{3\times4}$$

Sol 2.d → Y- vector

$$\begin{bmatrix} 40 \\ 46 \\ 38 \\ 39 \end{bmatrix}_{4\times1}$$

Solution 2.e → X'x matrix →

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 3.2 & 3.6 & 2.8 & 2.4 \\ 2 & 5 & 3 & 4 \end{bmatrix}_{3\times4} \begin{bmatrix} 1 & 3.2 & 2 \\ 1 & 3.6 & 5 \\ 1 & 2.8 & 3 \\ 1 & 2.4 & 4 \end{bmatrix}_{4\times3} = \begin{bmatrix} 4 & 12 & 14 \\ 12 & 36.8 & 42.4 \\ 14 & 42.4 & 54 \end{bmatrix}_{3\times3}$$

Scanned by CamScanner

Solution 2.f → $X'Y$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 3.2 & 3.6 & 2.8 & 2.4 \\ 2 & 5 & 3 & 4 \end{bmatrix}_{3\times4} \begin{bmatrix} 40 \\ 46 \\ 38 \\ 39 \end{bmatrix}_{4\times1} = \begin{bmatrix} 163 \\ 493.6 \\ 580 \end{bmatrix}_{3\times1}$$

Solution 2.g →

$$X'X\beta = X'Y$$

$$\beta = (X'X)^{-1} X'Y$$

From solution 2.e, $X'X = \begin{bmatrix} 4 & 12 & 14 \\ 12 & 36.8 & 42.4 \\ 14 & 42.4 & 54 \end{bmatrix}_{3\times3}$

From 2.f → $X'Y = \begin{bmatrix} 163 \\ 493.6 \\ 580 \end{bmatrix}_{3\times1}$

Now, for any matrix ~~A~~ A, $AA^{-1} = I$

Now, we need to find the inverse of $(X'X)$

$$(X'X)^{-1} = \frac{1}{|X'X|} \times \mathrm{Adj}(X'X)$$

Scanned by CamScanner

Determinant of $X'X$

$$\begin{bmatrix} 4 & 12 & 14 \\ 12 & 36.8 & 42.4 \\ 14 & 42.4 & 54 \end{bmatrix}$$

$$= 4(36.8\times 54 - 42.4\times 42.4) - 12(12\times 54 - 14\times 42.4) + 14(12\times 42.4 - 14\times 36.8)$$

$$= \frac{18944}{25} - \frac{3264}{5} - \frac{448}{5} = \frac{384}{25} = 15.36$$

det. $|X'X| = 15.36$

Below steps are to find ~~the~~ the Adj $(X'X)$

	C_1	C_2	C_3	C_1	C_2
R_1	~~4~~	~~12~~	~~14~~	~~4~~	~~12~~
R_2	12	36.8	42.4	12	36.8
R_3	14	42.4	54	14	42.4
R_1	4	12	14	4	12
R_2	12	36.8	42.4	12	36.8

$(36.8\times 54 - 42.4\times 42.4) = 189.44$

$(42.4\times 14 - 12\times 54) = -54.4$

$(12\times 42.4 - 36.8\times 14) = -6.4$

$54\times 4 - 14\times 14 = 20$

$14\times 12 - 42.4\times 4 = -1.6$

$12 \times 42.4 - 14 \times 36.8 = -6.4$

$14 \times 12 - 4 \times 42.4 = -1.6$

$4 \times 36.8 - 12 \times 12 = 3.2$

$$(X'X)^{-1} = \frac{1}{|X'X|} \operatorname{Adj}(X'X)$$

$$= \frac{1}{15.36} \begin{bmatrix} 189.44 & -54.4 & -6.4 \\ -54.4 & 20 & -1.6 \\ -6.4 & -1.6 & 3.2 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 12.33 & -3.5416 & -0.4166 \\ -3.5416 & 1.302 & -0.10416 \\ -0.4166 & -0.10416 & 0.2083 \end{bmatrix}$$

$$\beta = (X'X)^{-1} X'Y$$

$$\cancel{\beta = \begin{bmatrix} 12.33 & -3.5416 & -0.4166 \end{bmatrix} \begin{bmatrix} 40 \\ 46 \\ 38 \\ 39 \end{bmatrix}_{4\times 1}}$$

$$\beta = \begin{bmatrix} 12.33 & -3.5416 & -0.4166 \\ -3.5416 & 1.302 & -0.10416 \\ -0.4166 & -0.10416 & 0.2083 \end{bmatrix}_{3\times 3} \begin{bmatrix} 163 \\ 493.6 \\ 580 \end{bmatrix}_{3\times 1}$$

Solving the matrix multiplication

$$\begin{bmatrix} 12.33 \times 163 + (-3.5416 \times 493.6) + (-0.4166 \times 580) \\ (-3.5416 \times 163) + (1.302 \times 493.6) + (-0.10416 \times 580) \\ (-0.4166 \times 163) + (-0.10416 \times 493.6) + (0.2083 \times 580) \end{bmatrix}_{3 \times 1}$$

$$= \begin{bmatrix} 20.5 \\ 5 \\ 1.5 \end{bmatrix}_{3 \times 1}$$

Hence, the above is the β-vector.