BA Assignment-3

Abhinay Chiranjeeth Marneni

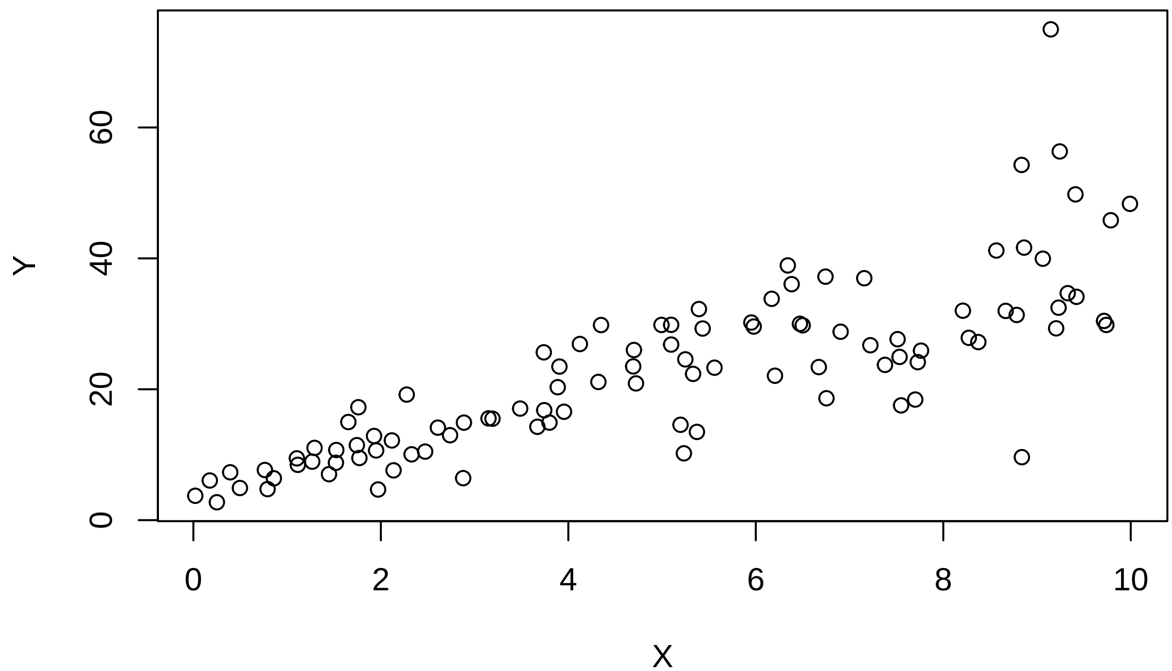
2022-11-13

1)To create the variables X and Y, run the following code.

```
set.seed(2017)
X=runif(100)*10
Y=X*4+3.45
Y=rnorm(100)*0.29*Y+Y
```

a) Creating Plot for X and Y variables, the plot will allow us to fit a linear model to explain Y based on X.

```
plot(X,Y)
```



```
# Yes, We are able to fit a linear model to Y given X.
```

b) Create a simple linear model of Y based on X.And find the accuracy of linear model.

```
Model<-lm(Y~X)
Model$coefficients
```

```
## (Intercept)           X
##    4.465490    3.610759
```

```
Accuracy<- 3.610759/4.465490
Accuracy  # The accuracy of above linear model is 0.8085919
```

```
## [1] 0.8085919
```

c) The correlation between X and Y is related to the Coefficient of Determination, R2, of the above linear model.

```
(cor(Y,X))^2  # To find Coefficient of Determination of model we using the                        code
```

```
## [1] 0.6517187
```

2) Utilize the mtcars dataset.

```
data<-data("mtcars")
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

a) James wants to buy a car. He and his friend, Chris, have different opinions about the Horse Power (hp) of cars. James think the weight of a car (wt) can be used to estimate the Horse Power of the car while Chris thinks the fuel consumption expressed in Mile Per Gallon (mpg), is a better estimator of the (hp). Who do you think is right? Construct simple linear models using mtcars data to answer the question.

```r
Weight <- lm(mtcars$wt ~ mtcars$hp) # James opinion
Mile <- lm(mtcars$mpg ~ mtcars$hp)  # Chris opinion
summary(Weight)
```

```
##
## Call:
## lm(formula = mtcars$wt ~ mtcars$hp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41757 -0.53122 -0.02038  0.42536  1.56455
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.838247   0.316520   5.808 2.39e-06 ***
## mtcars$hp   0.009401   0.001960   4.796 4.15e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7483 on 30 degrees of freedom
## Multiple R-squared:  0.4339, Adjusted R-squared:  0.4151
## F-statistic:    23 on 1 and 30 DF,  p-value: 4.146e-05
```

```r
summary(Mile)
```

```
##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$hp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7121 -2.1122 -0.8854  1.5819  8.2360
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
## mtcars$hp   -0.06823    0.01012  -6.742 1.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.863 on 30 degrees of freedom
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.5892
## F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

```
# By looking at the multiple r-squared values, Chris is right; Mile Per Gallon had a high r square valu
```

b) Build a model that uses the number of cylinders (cyl) and the mile per gallon (mpg) values of a car to predict the car Horse Power (hp). Using this model, what is the estimated Horse Power of a car with 4 calendar and mpg of 22.

```
Model1<-lm(hp ~cyl + mpg, data = mtcars)
Model1
```

```
##
## Call:
## lm(formula = hp ~ cyl + mpg, data = mtcars)
##
## Coefficients:
## (Intercept)          cyl          mpg
##      54.067       23.979       -2.775
```

To perdict the car horse power.

```
HP <- 54.067 + (23.979 *4) + (-2.775 * 22)
HP
```

```
## [1] 88.933
```

3) Install "mlbench" packages and Loading the "BostonHousing" dataset.

```
library(mlbench)
```

```
## Warning: package 'mlbench' was built under R version 4.2.2
```

```
data(BostonHousing)
```

a)Build a model to estimate the median value of owner-occupied homes (medv)based on the following variables: crime crate (crim), proportion of residential land zoned for lots over 25,000 sq.ft (zn), the local pupil-teacher ratio (ptratio) and weather the whether the tract bounds Chas River(chas). Is this an accurate model.

```
Accuracy1<- lm(medv~crim+zn+ptratio+chas,data=BostonHousing)
summary(Accuracy1)
```

```
##
## Call:
## lm(formula = medv ~ crim + zn + ptratio + chas, data = BostonHousing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.282  -4.505  -0.986   2.650  32.656
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.91868    3.23497  15.431  < 2e-16 ***
## crim        -0.26018    0.04015  -6.480 2.20e-10 ***
## zn           0.07073    0.01548   4.570 6.14e-06 ***
## ptratio     -1.49367    0.17144  -8.712  < 2e-16 ***
## chas1        4.58393    1.31108   3.496 0.000514 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.388 on 501 degrees of freedom
## Multiple R-squared:  0.3599, Adjusted R-squared:  0.3547
## F-statistic: 70.41 on 4 and 501 DF,  p-value: < 2.2e-16
```

```
# Due to the low R square value, the model is not accurate.
```

b)

##i) Imagine two houses that are identical in all aspects but one bounds the Chas River and the other does not. Which one is more expensive and by how much.

```
CR1 <- lm(medv ~ chas == 1, data = BostonHousing) # Price not next to river
summary(CR1)
```

```
##
## Call:
## lm(formula = medv ~ chas == 1, data = BostonHousing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.094  -5.894  -1.417   2.856  27.906
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    22.0938     0.4176  52.902  < 2e-16 ***
## chas == 1TRUE   6.3462     1.5880   3.996 7.39e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.064 on 504 degrees of freedom
## Multiple R-squared:  0.03072,    Adjusted R-squared:  0.02879
## F-statistic: 15.97 on 1 and 504 DF,  p-value: 7.391e-05
```

```r
CR2 <- lm(medv ~ chas == 0, data = BostonHousing) # Price next to river
summary(CR2)
```

```
##
## Call:
## lm(formula = medv ~ chas == 0, data = BostonHousing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.094  -5.894  -1.417   2.856  27.906
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     28.440      1.532  18.563  < 2e-16 ***
## chas == 0TRUE   -6.346      1.588  -3.996 7.39e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.064 on 504 degrees of freedom
## Multiple R-squared:  0.03072,    Adjusted R-squared:  0.02879
## F-statistic: 15.97 on 1 and 504 DF,  p-value: 7.391e-05
```

```r
# The cost of a house with a chas of 1 is higher than a house without a chas of 0
```

ii) Imagine two houses that are identical in all aspects but in the neighborhood of one of them the pupil-teacher ratio is 15 and in the other one is 18. Which one is more expensive and by how much.

```r
summary(HP2<-lm(medv~ptratio,data = BostonHousing))
```

```
##
## Call:
## lm(formula = medv ~ ptratio, data = BostonHousing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.8342  -4.8262  -0.6426   3.1571  31.2303
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   62.345      3.029   20.58   <2e-16 ***
## ptratio       -2.157      0.163  -13.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.931 on 504 degrees of freedom
## Multiple R-squared:  0.2578, Adjusted R-squared:  0.2564
## F-statistic: 175.1 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
# If ptratio increases the housing price decreases,The price of house which has ptratio of 15 is high
```

c) Which of the variables are statistically important.

```
summary(Accuracy1)
```

```
##
## Call:
## lm(formula = medv ~ crim + zn + ptratio + chas, data = BostonHousing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.282  -4.505  -0.986   2.650  32.656
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.91868    3.23497  15.431  < 2e-16 ***
## crim        -0.26018    0.04015  -6.480 2.20e-10 ***
## zn           0.07073    0.01548   4.570 6.14e-06 ***
## ptratio     -1.49367    0.17144  -8.712  < 2e-16 ***
## chas1        4.58393    1.31108   3.496 0.000514 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.388 on 501 degrees of freedom
## Multiple R-squared:  0.3599, Adjusted R-squared:  0.3547
## F-statistic: 70.41 on 4 and 501 DF,  p-value: < 2.2e-16
```

```
#  So the all varaiables in the model are Statistically important.
```

d) Use the anova analysis and determine the order of importance of these four variables.

```
anova(Accuracy1)
```

```
## Analysis of Variance Table
##
## Response: medv
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## crim        1  6440.8  6440.8 118.007 < 2.2e-16 ***
## zn          1  3554.3  3554.3  65.122 5.253e-15 ***
## ptratio     1  4709.5  4709.5  86.287 < 2.2e-16 ***
## chas        1   667.2   667.2  12.224 0.0005137 ***
## Residuals 501 27344.5    54.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## order of Importance of the varaiables
## 1) crim
## 2) zn
## 3) ptratio
## 4)chas
```