Lab 20 R Script

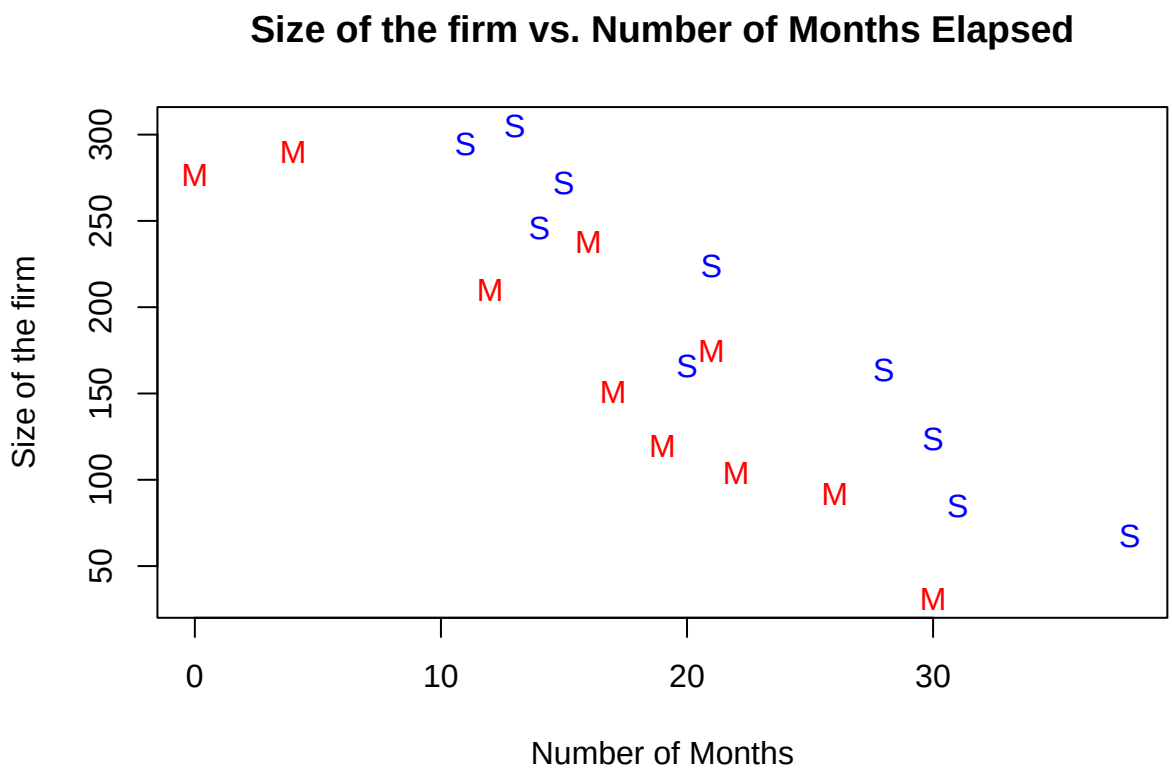
Alexander Hernandez

11/10/2022

1) Mutual and Stock Firms

a) Draw a Scatterplot of "Size of Firm vs Number of month elapsed"

```
stock = read.table("C:\\repos\\STAT 50001\\Lab 20\\stock.txt",
                   header=TRUE)
attach(stock)
plot(y, X1,
     col=ifelse(X2=="Mutual", "red", "blue"),
     pch=ifelse(X2=="Mutual", "M", "S"),
     main="Size of the firm vs. Number of Months Elapsed",
     ylab="Size of the firm", xlab="Number of Months")
```

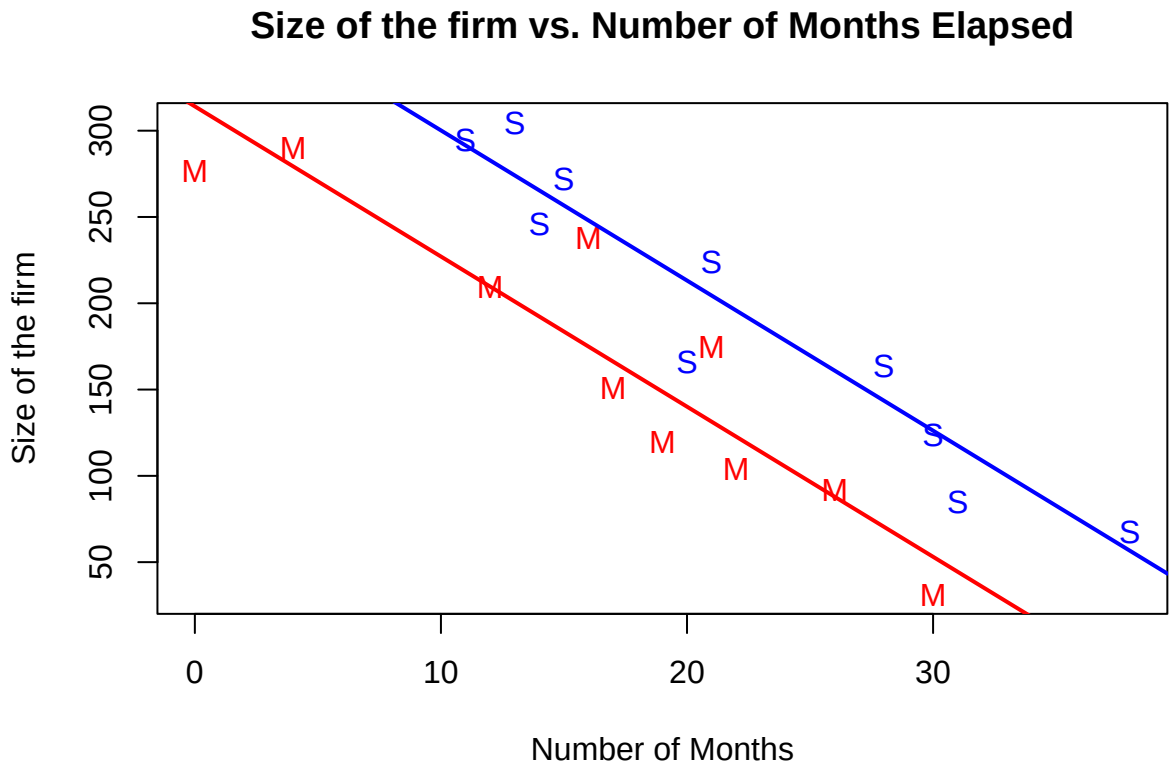


b) Fit a reression model with indicator variable and write out model

```
model1 = lm(X1~y+X2)
model1
```

```
##
## Call:
## lm(formula = X1 ~ y + X2)
##
## Coefficients:
## (Intercept)            y      X2Stock
##     314.079       -8.699        73.076
```

```
# y = 314.079 - 8.699(y)  if "Firm" is Mutual
# y = 387.155 - 8.699(y)  if "Firm" is Stock
plot(y, X1,
     col=ifelse(X2=="Mutual", "red", "blue"),
     pch=ifelse(X2=="Mutual", "M", "S"),
     main="Size of the firm vs. Number of Months Elapsed",
     ylab="Size of the firm", xlab="Number of Months")
abline(314.079, -8.699, col="red", lwd=2)
abline(387.155, -8.699, col="blue", lwd=2)
```



2) Weight, Health, and Fitness in "BodyFat"

a) Import "BodyFat" and access Variable Names

```
library(Lock5withR)
```

```
## Warning: package 'Lock5withR' was built under R version 4.2.2
```

```
names(BodyFat)
```

```
##  [1] "Bodyfat" "Age"     "Weight"  "Height"  "Neck"    "Chest"   "Abdomen"
##  [8] "Ankle"   "Biceps"  "Wrist"
```

```
attach(BodyFat)
```

b) Fit a model to predict Bodyfat using Height and Weight. Find predictors

```
model2b = lm(Bodyfat ~ Height + Weight)
summary(model2b)
```

```
##
## Call:
## lm(formula = Bodyfat ~ Height + Weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7697  -3.9527  -0.5364   4.0473  13.2829
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 71.48247   16.20086   4.412 2.65e-05 ***
## Height      -1.33568    0.25891  -5.159 1.32e-06 ***
## Weight       0.23156    0.02382   9.721 5.36e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.754 on 97 degrees of freedom
## Multiple R-squared:  0.494,  Adjusted R-squared:  0.4836
## F-statistic: 47.35 on 2 and 97 DF,  p-value: 4.48e-15
```

```
# All predictors are valid
```

c) Add Abdomen as a third predictor

```
model2b = lm(Bodyfat ~ Height + Weight + Abdomen)
summary(model2b)
```

```
##
## Call:
## lm(formula = Bodyfat ~ Height + Weight + Abdomen)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5219 -2.9969  0.0378  2.8933  9.2859
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -56.1329    18.1372  -3.095 0.002580 **
## Height        0.1018     0.2444   0.417 0.677750
## Weight       -0.1756     0.0472  -3.720 0.000335 ***
## Abdomen       1.0747     0.1158   9.279 5.27e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.199 on 96 degrees of freedom
## Multiple R-squared:  0.7332, Adjusted R-squared:  0.7249
## F-statistic: 87.96 on 3 and 96 DF,  p-value: < 2.2e-16
```

```
# With abdomen included, Height has a lower p-value and thus less relevant
```

d) Interpret the coefficient of Abdomen

```
# Abdomen's coefficient value shows that it affects bodyfat
# by 1.0747 per unit when predicting a value
```

3) Children Measurements

```
children = read.table("C:\\repos\\STAT 50001\\Lab 20\\pediatrician.txt",
                      header=TRUE)
attach(children)
```

```
## The following objects are masked from BodyFat:
##
##     Height, Weight
```

a) Construct a correlation matrix. Multicollinearity?

```
cor(children)
```

```
##                        Height    Weight Head_Circumference
## Height             1.0000000 0.7847652          0.8708869
## Weight             0.7847652 1.0000000          0.7796990
## Head_Circumference 0.8708869 0.7796990          1.0000000
```

```
# Yes, we do not want a relationship between Height and Weight
```

b) Find the LEast-squares regression equation with response variable = head circumference

```
model3 = lm(Head_Circumference ~ Height + Weight)
model3
```

```
##
## Call:
## lm(formula = Head_Circumference ~ Height + Weight)
##
## Coefficients:
## (Intercept)       Height       Weight
```

```
##    18.82425      0.78634      0.01281
```

```
# Head Circumference = 18.82425 + 0.78634(Height) + 0.01281(Weight)
```

c) Construct 95% confidence and prediction intervals for 27.5 height and 285 weight

```
predict(model3, data.frame(Height=27.5, Weight=285), interval = "conf")
```

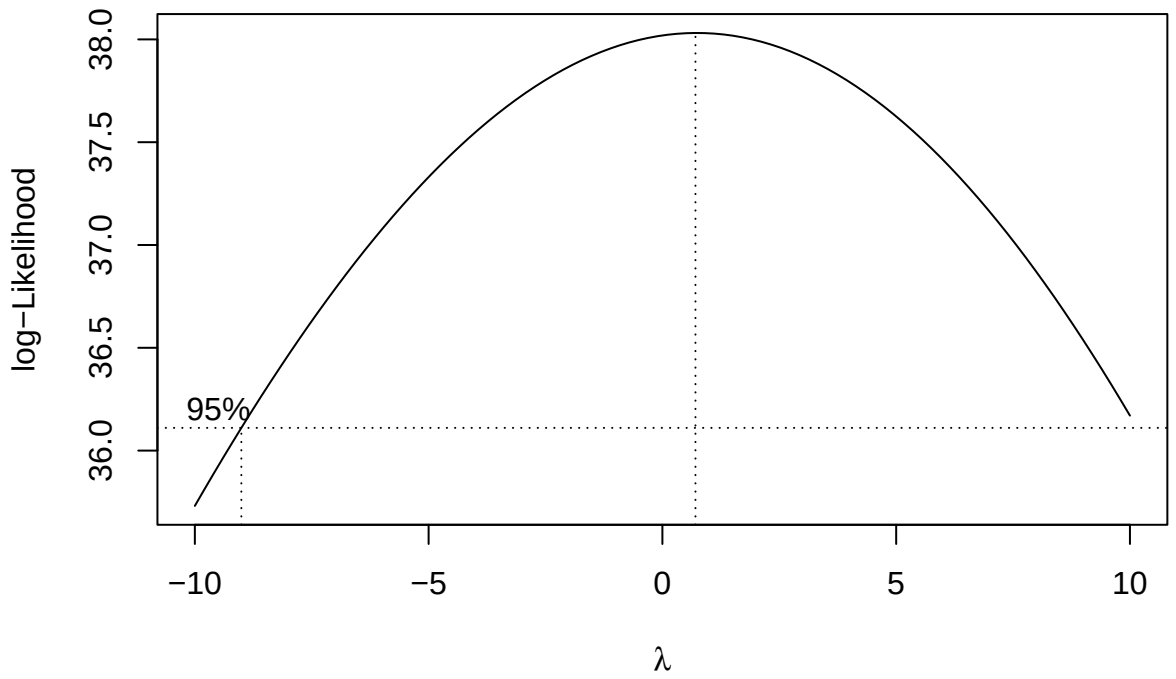
```
##        fit      lwr      upr
## 1 44.09898 43.39962 44.79834
```

```
predict(model3, data.frame(Height=27.5, Weight=285), interval = "pred")
```

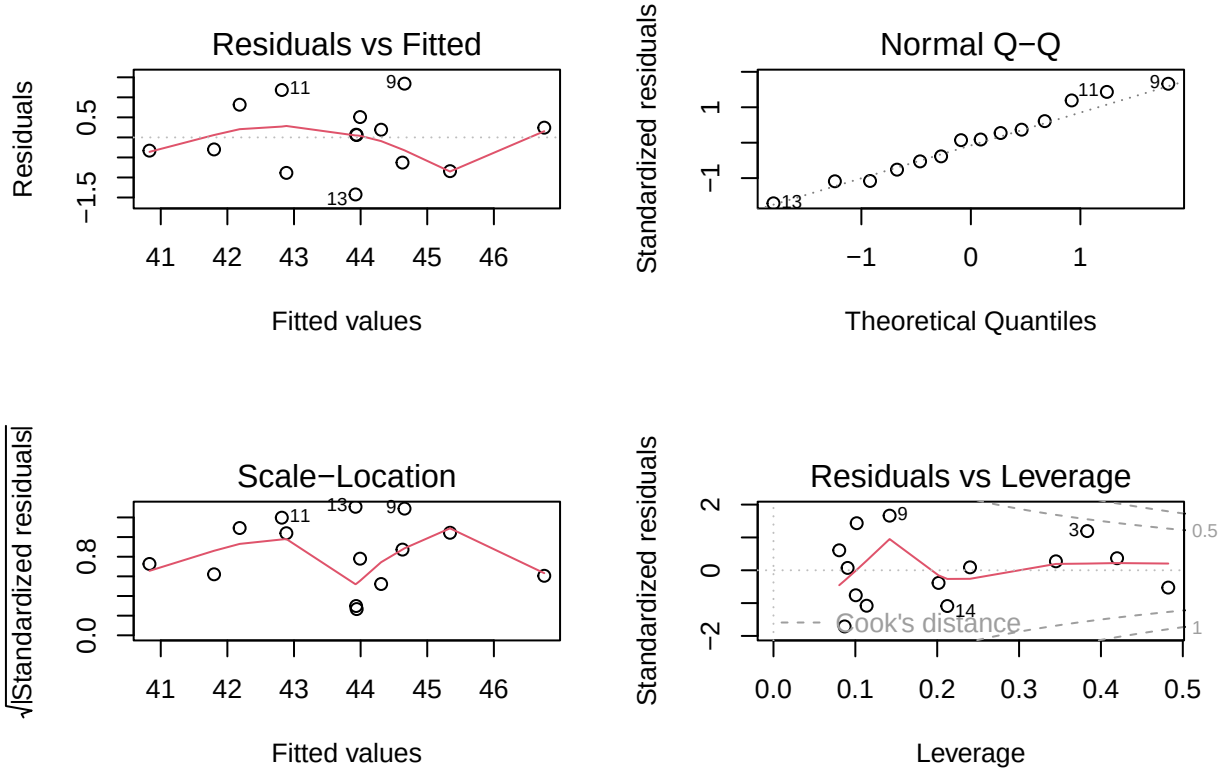
```
##        fit      lwr     upr
## 1 44.09898 42.05886 46.1391
```

d) Perform the residual analysis of the model

```
library(MASS)
b=boxcox(model3, lambda=seq(-10,10))
```




```
par(mfrow=c(2,2))
plot(model3)
```




```
# No transformation is needed
# There are further contributing factors that affect head circumference
```