Homework 3 Exercise 2

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

2020/4/16

Question 1

```
library(mgcv)
cars$speed2 <- (cars$speed)^2
model1 = lm(dist~speed + speed2, data = cars)
summary(model1)
```

```
##
## Call:
## lm(formula = dist ~ speed + speed2, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.720  -9.184  -3.188   4.628  45.152
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.47014   14.81716   0.167    0.868
## speed        0.91329    2.03422   0.449    0.656
## speed2       0.09996    0.06597   1.515    0.136
##
## Residual standard error: 15.18 on 47 degrees of freedom
## Multiple R-squared:  0.6673, Adjusted R-squared:  0.6532
## F-statistic: 47.14 on 2 and 47 DF,  p-value: 5.852e-12
```

```
AIC(model1)
```

```
## [1] 418.7721
```

```
model2 = lm(dist~speed, data = cars)
summary(model2)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

```
AIC(model2)
```

```
## [1] 419.1569
```

From the results, the model with speed2 is a better fit for the data.

Question 2

From the hypothesis test result in Question 1, the p values are not so good. We can take the reaction time instead. Due to the different reaction time, it could lead to a non-linear effection to the model. Also, we can test the combination of the speed and the reaction time to see their influence to the regression.

Question 3

```
myf <- function (X, y) {
  qrx <- qr(X) ## returns a QR decomposition object
  Q <- qr.Q(qrx,complete=TRUE) ## extract Q
  R <- qr.R(qrx) ## extract R
  f <- t(Q)%*%y
  beta <- backsolve(R,f)
  return(beta)
}
```

Question 4

```
test1 <- myf(model.matrix(dist ~ speed + speed2, cars), data.matrix(cars[,c("dist")]))
test1
```

```
##           [,1]
## [1,] 2.4701378
## [2,] 0.9132876
## [3,] 0.0999593
```

The coefficients are the same as Question 1, so it works.

Question 5

```
myf2 <- function (X, y) {
  qrx <- qr(X) ## returns a QR decomposition object
  Q <- qr.Q(qrx,complete=TRUE) ## extract Q
  R <- qr.R(qrx) ## extract R
  f <- t(Q)%*%y
  beta <- backsolve(R,f)

  res <- y-X%*%beta
  s <- as.vector(t(res)%*%res/(nrow(X)-ncol(X)))
  var <- solve(R)%*%t(solve(R))*s
  return(list(beta=beta, std=sqrt(as.matrix(diag(var),ncol=ncol(X)))))
}
```

```
test2 <- myf2(model.matrix(dist ~ speed + speed2, cars), data.matrix(cars[,c("dist")]))
test2$std
```

```
##                    [,1]
## (Intercept) 14.81716473
## speed        2.03422044
## speed2       0.06596821
```

Question 6

```
mat= model.matrix(dist ~ speed + speed2, cars)
pt(test2$beta / test2$std, nrow(mat)-ncol(mat), lower.tail=FALSE)
```

```
##                   [,1]
## (Intercept) 0.43415754
## speed       0.32776122
## speed2      0.06820122
```