Lab #5 Report

Nick Huo & Nahom Ayele

2022-10-19

Set-up

```
library(carData)
data("mtcars")
```

Question 1

Part 1.

```
model_ln <- lm(log(mpg) ~ cyl+disp+hp+wt, data = mtcars)
summary(model_ln)
```

```
##
## Call:
## lm(formula = log(mpg) ~ cyl + disp + hp + wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14897 -0.07696 -0.02464  0.07057  0.24981
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.9056288  0.1222253  31.954  < 2e-16 ***
## cyl         -0.0378448  0.0290719  -1.302 0.203993
## disp         0.0001174  0.0005198   0.226 0.823043
## hp          -0.0010687  0.0005384  -1.985 0.057401 .
## wt          -0.1816048  0.0450111  -4.035 0.000404 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1114 on 27 degrees of freedom
## Multiple R-squared:  0.8782, Adjusted R-squared:  0.8601
## F-statistic: 48.66 on 4 and 27 DF,  p-value: 5.844e-12
```

$F - statistic = 48.66$; $DF = 4, 27$. Since we have a extremely small $p - value$ (< 0.0001), we are able to reject H_0 as there is very strong evidence to suggest that one of the estimators has a slope different from 0.

Part 2.

```
lin_eft <- function(b) {
    return(sign(b) * abs(exp(b)-1))
}
for (p in 1:4) {
    this_name <- labels(model_ln)[p]
    coef <- model_ln$coefficients[1+p]
    print(paste("The linear effect of", this_name,"is",round(lin_eft(coef)*100,2),"%"))
}
```

```
## [1] "The linear effect of cyl is -3.71 %"
## [1] "The linear effect of disp is 0.01 %"
## [1] "The linear effect of hp is -0.11 %"
## [1] "The linear effect of wt is -16.61 %"
```

- *cyl*: As engine cylinder count increases by 1, mpg decreases by 3.7% on average while holding other variables constant.
- *disp*: As engine displacement increases by 1 cu.in., mpg increases by 0.01% on average while holding other variables constant.
- *hp*: As engine gross horsepower increases by 1 hp, mpg decreases by 0.11% while holding other variables constant.
- *wt*: As car weight increases by 1000 lbs, mpg decreases by 16.6% while holding other variables constant.

Part 3.

```
table_df <- data.frame("t-statistic"=round(summary(model_ln)$coefficients[2:5,3],2),
              "p-value"=round(summary(model_ln)$coefficients[2:5,4],4))
knitr::kable(
  table_df, booktabs = TRUE,
  caption = 't-statistic and p-value of the covariates'
)
```

Table 1: t-statistic and p-value of the covariates

	t.statistic	p.value
cyl	-1.30	0.2040
disp	0.23	0.8230
hp	-1.98	0.0574
wt	-4.03	0.0004

Part 4.

```
confint(model_ln,c("cyl","disp","hp","wt"),level=.95)
```

```
##              2.5 %       97.5 %
## cyl  -0.0974953761  0.0218057024
```

```
## disp -0.0009491483  0.0011839064
## hp   -0.0021734121  0.0000360317
## wt   -0.2739598493 -0.0892496889
```

Part 5.

```r
hatvalues(model_ln)[1:5]
```

```
##         Mazda RX4     Mazda RX4 Wag        Datsun 710    Hornet 4 Drive
##        0.08527051        0.08775434        0.08713971        0.08997459
## Hornet Sportabout
##        0.17577861
```

Part 6.

```r
this_car<-data.frame(cyl=6, disp=190, hp=180, wt=1.5)
predict(model_ln, newdata=this_car, interval="confidence", level=0.9)
```

```
##       fit      lwr      upr
## 1 3.23609 3.119155 3.353026
```

Part 7.

```r
predict(model_ln, newdata=this_car, interval="prediction", level=0.9)
```

```
##       fit      lwr      upr
## 1 3.23609 3.013252 3.458929
```

Question 2

Part 1.

```r
X <- cbind(rep(1,length(mtcars$mpg)), data.matrix(mtcars[,c(2:4,6)]))
Y <- log(mtcars$mpg)
B_hat <- solve(t(X) %*% X) %*% t(X) %*% Y
B_hat
```

```
##              [,1]
##       3.905628840
## cyl  -0.037844837
## disp  0.000117379
## hp   -0.001068690
## wt   -0.181604769
```

Part 2.

We know that $\hat{\beta} \sim MVN(\beta, \sigma^2(X^T X)^{-1})$. We can find the standard errors of the coefficients by looking at the square root of the variance of $\hat{\beta}$.

```
SE <- sqrt(sigma(model_ln)**2 * diag(solve(t(X) %*% X)))
test_statistics <- B_hat/SE
test_statistics
```

```
##            [,1]
##      31.9543398
## cyl  -1.3017685
## disp  0.2258188
## hp   -1.9849078
## wt   -4.0346693
```

We got the same results as compared to Q1 part 3.

Part 3.

```
ME <- qt(0.975,length(mtcars$mpg-5)) * sqrt(sigma(model_ln)**2 * diag(solve(t(X) %*% X)))
data.frame(low=round(B_hat-ME,5),high=round(B_hat+ME,5))[2:5,]
```

```
##           low     high
## cyl  -0.09706  0.02137
## disp -0.00094  0.00118
## hp   -0.00217  0.00003
## wt   -0.27329 -0.08992
```

The answers we got here are very similar to the answers to Q1 part 4, but there are some differences after certain decimal places.

Part 4.

```
H <- X %*% solve(t(X) %*% X) %*% t(X)
diag(H[1:5,1:5])
```

```
##         Mazda RX4     Mazda RX4 Wag        Datsun 710    Hornet 4 Drive
##        0.08527051        0.08775434        0.08713971        0.08997459
## Hornet Sportabout
##        0.17577861
```

We get the same results compared to Q1.

Part 5.

```
x_new <- matrix(c(1, 6, 190, 180, 1.5), nrow = 1)
ME_conf <- qt(0.95,length(mtcars$mpg-5)) * sqrt(sigma(model_ln)**2 * (x_new) %*% solve(t(X) %*% X) %*% t
data.frame(low=x_new%*%B_hat-ME_conf, high=x_new%*%B_hat+ME_conf)
```

```
##      low     high
## 1 3.1198 3.352381
```

We get the same results compared to Q1.

Part 6.

```
x_new <- matrix(c(1, 6, 190, 180, 1.5), nrow = 1)
ME_pred <- qt(0.95,length(mtcars$mpg-5)) * sqrt(sigma(model_ln)**2 * (1+ (x_new) %*% solve(t(X) %*% X) %
data.frame(low=x_new%*%B_hat-ME_pred, high=x_new%*%B_hat+ME_pred)
```

```
##        low     high
## 1 3.014482 3.457699
```

We get the same results compared to Q1.