STAT435_HW2

Liyuan Tang

4/19/2020

Question 1

1.1

```r
library(ISLR)
#View(Auto)
lm.fit = lm(mpg~ cylinders+displacement+horsepower+weight+acceleration+year+factor(origin), data = Auto
summary(lm.fit)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + factor(origin), data = Auto[, -9])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0095 -2.0785 -0.0982  1.9856 13.3608
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.795e+01  4.677e+00  -3.839 0.000145 ***
## cylinders       -4.897e-01  3.212e-01  -1.524 0.128215
## displacement     2.398e-02  7.653e-03   3.133 0.001863 **
## horsepower      -1.818e-02  1.371e-02  -1.326 0.185488
## weight          -6.710e-03  6.551e-04 -10.243  < 2e-16 ***
## acceleration     7.910e-02  9.822e-02   0.805 0.421101
## year             7.770e-01  5.178e-02  15.005  < 2e-16 ***
## factor(origin)2  2.630e+00  5.664e-01   4.643 4.72e-06 ***
## factor(origin)3  2.853e+00  5.527e-01   5.162 3.93e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.307 on 383 degrees of freedom
## Multiple R-squared:  0.8242, Adjusted R-squared:  0.8205
## F-statistic: 224.5 on 8 and 383 DF,  p-value: < 2.2e-16
```

For the first column 'Estimate', it means the change in mpg with increaing 1 unit of the corresponding variable and holding all other variables fixed. The 'Intercept' means that the expected mean mpg for a US car is -17.95 when all other variables are 0. On average, the mpg of a car made in Europe is 2.63 higher than a US car and the mpg of a car made in Japan is 2.853 higher than a US car when holding all other variables fixed.

For ‘displacement’, ‘weight’, ‘year’, ‘orgin’ and ‘Intercept’, we can reject the null hypothesis that there is no linear association between that predictor and gas mileage on the 0.05 significance level.

1.2

```
mpg_pred = predict(lm.fit, newdata = Auto[, c(-1, -9)])
res.mse = sum((mpg_pred - Auto$mpg)^2) / length((mpg_pred))
cat('The resubstitution mean square error of this model is' , res.mse)
```

```
## The resubstitution mean square error of this model is 10.68212
```

1.3

```
newdata = data.frame(cylinders = 3, displacement = 100, horsepower = 85, weight = 3000, acceleration = 2
mpg.predict = predict(object = lm.fit, newdata)
cat('The predicted gas mileage is', mpg.predict)
```

```
## The predicted gas mileage is 27.89483
```

1.4

```
cat('The difference between the mpg of a Japanese car and the mpg of an American car is', coef(lm.fit)[
```

```
## The difference between the mpg of a Japanese car and the mpg of an American car is 2.853228
```

1.5

```
cat('The change in mpg associated with a 10-unit change in horsepower is', 10*lm.fit$coefficients[4])
```

```
## The change in mpg associated with a 10-unit change in horsepower is -0.1818346
```

Question 2

2.1

```
ame = ifelse(Auto$origin == 1, 1, 0)
eur = ifelse(Auto$origin == 2, 1, 0)
lm2.fit = lm(mpg~ ame + eur, data = Auto)
pred.jap = lm2.fit$coefficients[1]
pred.ame = lm2.fit$coefficients[1] + lm2.fit$coefficients[2]
pred.eur = lm2.fit$coefficients[1] + lm2.fit$coefficients[3]
print(lm2.fit$coefficients)
```

```
## (Intercept)         ame         eur
##   30.450633  -10.417164   -2.847692
```

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$.

$\hat{\beta}_0 =$ 30.4506329, $\hat{\beta}_1 =$ -10.4171635, $\hat{\beta}_2 =$ -2.8476917.

If the car is made in US $x_{i1} = 1$, $x_{i2} = 0$. If the car is made in Europe $x_{i1} = 0$, $x_{i2} = 1$. If the car is made in Japan $x_{i1} = 0$, $x_{i2} = 0$.

The predicted mpg for a Japanese car is 30.4506329. The predicted mpg for a American car is 20.0334694. The predicted mpg for a European car is 27.6029412.

2.2

```
jap = ifelse(Auto$origin == 3, 1, 0)
lm3.fit = lm(mpg~ jap + eur, data = Auto)
pred.ame = lm3.fit$coefficients[1]
pred.jap = lm3.fit$coefficients[1] + lm3.fit$coefficients[2]
pred.eur = lm3.fit$coefficients[1] + lm3.fit$coefficients[3]
print(lm3.fit$coefficients)
```

```
## (Intercept)         jap         eur
##   20.033469   10.417164    7.569472
```

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$.

$\hat{\beta}_0 =$ 20.0334694, $\hat{\beta}_1 =$ 10.4171635, $\hat{\beta}_2 =$ 7.5694718.

If the car is made in US $x_{i1} = 0$, $x_{i2} = 0$. If the car is made in Europe $x_{i1} = 0$, $x_{i2} = 1$. If the car is made in Japan $x_{i1} = 1$, $x_{i2} = 0$.

The predicted mpg for a Japanese car is 30.4506329. The predicted mpg for a American car is 20.0334694. The predicted mpg for a European car is 27.6029412.

2.3

```
ame = ifelse(Auto$origin == 1, 1, -1)
eur = ifelse(Auto$origin == 2, 1, -1)
lm4.fit = lm(mpg~ ame + eur, data = Auto)
pred.jap = lm4.fit$coefficients[1] - lm4.fit$coefficients[2] - lm4.fit$coefficients[3]
pred.ame = lm4.fit$coefficients[1] + lm4.fit$coefficients[2] - lm4.fit$coefficients[3]
pred.eur = lm4.fit$coefficients[1] - lm4.fit$coefficients[2] + lm4.fit$coefficients[3]
print(lm4.fit$coefficients)
```

```
## (Intercept)         ame         eur
##   23.818205   -5.208582   -1.423846
```

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$.

$\hat{\beta}_0 =$ 23.8182053, $\hat{\beta}_1 =$ -5.2085818, $\hat{\beta}_2 =$ -1.4238459.

If the car is made in US $x_{i1} = 1$, $x_{i2} = -1$. If the car is made in Europe $x_{i1} = -1$, $x_{i2} = 1$. If the car is made in Japan $x_{i1} = -1$, $x_{i2} = -1$.

The predicted mpg for a Japanese car is 30.4506329. The predicted mpg for a American car is 20.0334694. The predicted mpg for a European car is 27.6029412.

2.4

```
new.origin = Auto$origin
new.origin[new.origin == 3] = 0
lm5.fit = lm(mpg ~ new.origin, data = Auto)
pred.jap = lm5.fit$coefficients[1]
```

```r
pred.ame = lm5.fit$coefficients[1] + lm5.fit$coefficients[2]
pred.eur = lm5.fit$coefficients[1] + 2 * lm5.fit$coefficients[2]
print(lm5.fit$coefficients)
```

```
## (Intercept)  new.origin
##   25.239473   -1.845337
```

$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$.

$\hat{\beta}_0 = 23.8182053$, $\hat{\beta}_1 =$ -5.2085818.

If the car is made in US $x_{i1} = 1$. If the car is made in Europe $x_{i1} = 2$. If the car is made in Japan $x_{i1} = 0$.

The predicted mpg for a Japanese car is 25.2394727. The predicted mpg for a American car is 23.394136. The predicted mpg for a European car is 21.5487992.

2.5

The first three models give the same predicted mpg even the functions are different. But the last one is different. So when we use two dummy variables to build the model, we will always get the same predicted value no matter what default value we chose.

Question 3

3.1

```r
h1 = -165.1 + 4.8 * 64
cat('The weight is', h1)
```

```
## The weight is 142.1
```

3.2

$\beta_0^* = \beta_0 = -165.1$, $\beta_1^* = \hat{\beta}_1 * 12 = 4.8 * 12 = 57.6$. The weight is still 142.1, as the previous model.

3.3

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

$= \beta_0 + \beta_1 X_1 + \beta_2 \frac{1}{12} X_1 + \epsilon$

$= \beta_0 + (\beta_1 + \frac{1}{12}\beta_2) X_1 + \epsilon$

Thus, the general expression should be $\beta_1 + \frac{1}{12}\beta_2 = 4.8, \beta_0 = -165.1$

3.4

The mean squared errors for three models should be the same.