Project 5.2

Jared Barber

3/31/2022

Exercise 12.01

An estimate for the linear regression is y = 0.5800 + (2.7122)x1 + (2.0497)x2

```
reg = lm(y~x1+x2, data =data.frame(Ex12.01))
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = data.frame(Ex12.01))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.85350 -0.35739 -0.00333  0.24429  1.15088
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.57999    0.60685   0.956    0.371
## x1           2.71224    0.20209  13.421 2.99e-06 ***
## x2           2.04971    0.04808  42.630 1.02e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.657 on 7 degrees of freedom
## Multiple R-squared:  0.9997, Adjusted R-squared:  0.9996
## F-statistic: 1.269e+04 on 2 and 7 DF,  p-value: 3.482e-13
```

Exercise 12.24

$$H_0 : B_1 = 2$$
$$H_A : B_1! = 2$$

The p-value obtained is 0.0084. Since this is small, reject the null hypothesis and conclude that B1 is not equal to 2.

```
TS = (2.71224-2)/0.20209
1-pt(TS,5)
```

```
## [1] 0.008420234
```

Exercise 12.31

From 12.01, R-squared is 0.9997. 99.97% of the variation in y can be explained by the linear model.

Exercise 12.05

a - An estimate for the linear regression is y = -102.71324 + (0.60537)x1 + (8.92364)x2 + (1.43746)x3 + (0.01361)x4

b- An estimate for power consumption with the given parameters is 287.562.

```
reg = lm(y~x1+x2+x3+x4, data =data.frame(Ex12.05))
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4, data = data.frame(Ex12.05))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.758  -9.952   3.350   6.627  23.311
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -102.71324  207.85885  -0.494    0.636
## x1             0.60537    0.36890   1.641    0.145
## x2             8.92364    5.30052   1.684    0.136
## x3             1.43746    2.39162   0.601    0.567
## x4             0.01361    0.73382   0.019    0.986
##
## Residual standard error: 15.58 on 7 degrees of freedom
## Multiple R-squared:  0.7447, Adjusted R-squared:  0.5989
## F-statistic: 5.106 on 4 and 7 DF,  p-value: 0.0303
```

```
y = -102.71324 + (0.60537)*75 + (8.92364)*24 + (1.43746)*90 + (0.01361)*98
y
```

```
## [1] 287.562
```

Exercise 12.19

s^2 = 242.7364

```
15.58^2
```

```
## [1] 242.7364
```

Exercise 12.27

A 95% CI for the mean response is (263.7879, 311.3357). It can be said with 95% confidence that the true mean is between 263.78 and 311.34 with the given parameters. A 95% CI for the predicted response is (243.7175, 331.4062) It can be said with 95% confidence that the next observation of y will fall within this range with the given parameters.

```
predict(reg, newdata = data.frame(x1=75,x2=24,x3=90,x4=98), interval = "confidence")
```

```
##        fit      lwr      upr
## 1 287.5618 263.7879 311.3357
```

```
predict(reg, newdata = data.frame(x1=75,x2=24,x3=90,x4=98), interval = "prediction")
```

```
##        fit      lwr      upr
## 1 287.5618 243.7175 331.4062
```

Exercise 12.33

From 12.05, the F-statistic is 5.106 with a p-value of 0.0303. Since the p-value is greater than the significance level, fail to reject the null. We cannot claim the regression is significant.

Exercise 12.37

H_0:B_1 = B_2 = 0

H_A: At least one of B_1 and B_2 is not zero.

Comparing models with and without B1 and B2 returns an F-statistic of 10.177 and a p-value of 0.0085. Since the p-value is small, reject the null hypothesis and claima that at B2 is not zero.

```
reg = lm(y~x1+x2+x3+x4, data =data.frame(Ex12.05))
reg2 = lm(y~x3+x4, data =data.frame(Ex12.05))

anova(reg2,reg)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x3 + x4
## Model 2: y ~ x1 + x2 + x3 + x4
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1      9 6639.2
## 2      7 1699.0  2    4940.2 10.177 0.008478 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exercise 12.13

a - An estimate of the regression is y = 350.9943 + (-1.2720)x1 + (-0.1539)x2

b - An estimate of wear with the given parameters is 140.8743.

```
reg = lm(y~x1+x2, data=data.frame(Ex12.13))
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = data.frame(Ex12.13))
##
## Residuals:
##       1       2       3       4       5       6
## -24.987  11.820  12.830  24.307 -20.460  -3.511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 350.99427   74.75307   4.695   0.0183 *
## x1           -1.27199    1.16914  -1.088   0.3562
## x2           -0.15390    0.08953  -1.719   0.1841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.5 on 3 degrees of freedom
## Multiple R-squared:  0.8618, Adjusted R-squared:  0.7696
## F-statistic: 9.353 on 2 and 3 DF,  p-value: 0.05138
```

```
y = 350.9943 + (-1.2720)*20 + (-0.1539)*1200
y
```

```
## [1] 140.8743
```

Exercise 12.43

Comparing the Adjusted R-squared values from the three models shows that both models with x2 have higher values than the model with only x1. Of the two models that include x2, the model with x1 and x2 has a slightly higher adjusted R-squared value, but the difference does not jsutify the complexity of the model. The third model, using only x2, is prefered.

```
reg = lm(y~x1, data=data.frame(Ex12.13))
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x1, data = data.frame(Ex12.13))
##
## Residuals:
##       1       2       3       4       5       6
## -30.257   7.005 -20.580  46.441 -14.020  11.412
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  227.827     26.009   8.759 0.000936 ***
## x1            -2.856      0.878  -3.253 0.031296 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.11 on 4 degrees of freedom
## Multiple R-squared:  0.7256, Adjusted R-squared:  0.6571
## F-statistic: 10.58 on 1 and 4 DF,  p-value: 0.0313
```

```
reg = lm(y~x2, data=data.frame(Ex12.13))
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x2, data = data.frame(Ex12.13))
##
## Residuals:
##      1      2      3      4      5      6
## -10.67  16.08  26.05  18.25 -31.94 -17.77
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 399.97425   61.03126   6.554   0.0028 **
## x2           -0.23067    0.05636  -4.093   0.0149 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.08 on 4 degrees of freedom
## Multiple R-squared:  0.8073, Adjusted R-squared:  0.7591
## F-statistic: 16.75 on 1 and 4 DF,  p-value: 0.01494
```

Exercise 12.45

a - an estimate for the linear regression is y = 5.959 + (-0.00004)Odometer + (0.3373)Octane + (-12.63)Van + (-12.98)Suv

b - Since the coefficients of Van and Suv are both negative, a sedan should have the best gas mileage.

c - The parameter estimate for van is -12.63(1.072). The parameter estimate for suv is -12.98(1.115). Since the difference between the two parameters is less than the standard error of either van or suv, there should not be a significant difference between a van and suv in terms of gas mileage.

```
#str(Ex12.45)

df = data.frame(mpg=c(Ex12.45$MPG),
                odo=c(Ex12.45$Odometer),
                oct=c(Ex12.45$Octane),
                van=c(rep(0,8),rep(1,9),rep(0,8)),
                suv=c(rep(0,17),rep(1,8)))

reg = lm(mpg~odo+oct+van+suv,data=df)
summary(reg)
```

```
## 
## Call:
## lm(formula = mpg ~ odo + oct + van + suv, data = df)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -3.938 -1.354 -0.372  1.448  4.046 
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  5.959e+00  7.058e+00   0.844 0.408437    
## odo         -3.773e-05  1.593e-05  -2.368 0.028070 *  
## oct          3.373e-01  8.380e-02   4.025 0.000663 ***
## van         -1.263e+01  1.072e+00 -11.776 1.89e-10 ***
## suv         -1.298e+01  1.115e+00 -11.647 2.30e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.178 on 20 degrees of freedom
## Multiple R-squared:  0.9046, Adjusted R-squared:  0.8856 
## F-statistic: 47.43 on 4 and 20 DF,  p-value: 6.256e-10
```

Exercise 12.46

a - An estimate for the linear regression is y = -206.6 + (0.0054)Income + (236.7)Female + (-49.24)Family. The since the coefficient for female is positive, the company would prefer female customers.

b - The p-value associated with income, 0.0649 is small, so income is an important factor.

```
#str(Ex12.46)
df = data.frame(profit=c(Ex12.46$profit),
                income=c(Ex12.46$income),
                female=c(rep(0,10),rep(1,10)),
                fam=c(Ex12.46$family.member))

reg = lm(profit~income+female+fam, data = df)
summary(reg)
```

```
##
## Call:
## lm(formula = profit ~ income + female + fam, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -347.24 -150.85    7.16  132.66  341.49
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.066e+02  1.637e+02  -1.262   0.2249
## income       5.433e-03  2.741e-03   1.982   0.0649 .
## female       2.367e+02  1.106e+02   2.141   0.0480 *
## fam         -4.924e+01  5.196e+01  -0.948   0.3574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 227.5 on 16 degrees of freedom
## Multiple R-squared:  0.3075, Adjusted R-squared:  0.1777
## F-statistic: 2.368 on 3 and 16 DF,  p-value: 0.1091
```

In addition to what is specified in the exercise, add an income*Gender interaction and test to see if it is significant.

The pvalue associated with the new parameter is 0.2872. Since this is a moderately large value, income*Gender interaction is not a significant factor.

```
reg = lm(profit~income*female+fam, data = df)
summary(reg)
```

```
##
## Call:
## lm(formula = profit ~ income * female + fam, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -302.43 -182.23  -14.96   99.02  370.12
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.207e+02  1.927e+02  -1.664   0.1168
## income         8.864e-03  4.133e-03   2.145   0.0488 *
## female         4.937e+02  2.575e+02   1.917   0.0744 .
## fam           -7.609e+01  5.706e+01  -1.334   0.2022
## income:female -6.185e-03  5.606e-03  -1.103   0.2872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 226 on 15 degrees of freedom
## Multiple R-squared:  0.3595, Adjusted R-squared:  0.1887
## F-statistic: 2.105 on 4 and 15 DF,  p-value: 0.1308
```