Assignment 2

Shraddha Ganesh

18/01/2024

```r
library(readxl)
growth = read_excel("C:/Users/shrad/Downloads/Growth.xlsx")
knitr::opts_chunk$set(echo = TRUE)
```

Problem 1

i.

```r
#Mean Values.
mean1 = mean(growth$growth)
mean2 = mean(growth$tradeshare)
mean3 = mean(growth$yearsschool)
mean4 = mean(growth$oil)
mean5 = mean(growth$rev_coups)
mean6 = mean(growth$assasinations)
mean7 = mean(growth$rgdp60)
#Standard Deviations.
sd1 = sd(growth$growth)
sd2 = sd(growth$tradeshare)
sd3 = sd(growth$yearsschool)
sd4 = sd(growth$oil)
sd5 = sd(growth$rev_coups)
sd6 = sd(growth$assasinations)
sd7 = sd(growth$rgdp60)
#Minimum Values.
min1 = min(growth$growth)
min2 = min(growth$tradeshare)
min3 = min(growth$yearsschool)
min4 = min(growth$oil)
min5 = min(growth$rev_coups)
min6 = min(growth$assasinations)
min7 = min(growth$rgdp60)
#Maximum Values.
max1 = max(growth$growth)
max2 = max(growth$tradeshare)
max3 = max(growth$yearsschool)
max4 = max(growth$oil)
max5 = max(growth$rev_coups)
max6 = max(growth$assasinations)
max7 = max(growth$rgdp60)

table = data.frame(matrix(nrow=7,ncol=4))
colnames(table) = c("Mean","Standard Deviation","Minimum",
                   "Maximum")
rownames(table) = c("Growth (Percentage %)","Trade-Share (Average Value)"
                    ,"YearsShare (Avg.Years)","Oil (Dummy Variable)",
                    "Rev_Coups (Average Number)",
                    "Assasinations (Avg. Per Million)","RGDP60 (US Dollars)")
table[,1] = c(mean1,mean2,mean3,mean4,mean5,mean6,mean7)
table[,2] = c(sd1,sd2,sd3,sd4,sd5,sd6,sd7)
table[,3] = c(min1,min2,min3,min4,min5,min6,min7)
table[,4] = c(max1,max2,max3,max4,max5,max6,max7)
knitr::kable(table,"pipe",align=c("l","c","c","c","C"))
```

	Mean	Standard Deviation	Minimum	Maximum
Growth (Percentage %)	1.8691197	1.8161889	-2.811945	7.1568546
Trade-Share (Average Value)	0.5423919	0.2283326	0.140502	1.1279370
YearsShare (Avg.Years)	3.9592187	2.5534647	0.200000	10.0699997
Oil (Dummy Variable)	0.0000000	0.0000000	0.000000	0.0000000
Rev_Coups (Average Number)	0.1700666	0.2254557	0.000000	0.9703704
Assasinations (Avg. Per Million)	0.2819010	0.4941590	0.000000	2.4666667
RGDP60 (US Dollars)	3130.8125339	2522.9786371	366.999939	9895.0039062

```r
knitr::opts_chunk$set(echo = TRUE)
```

ii.

```r
p1_reg = lm(growth~tradeshare+yearsschool+rev_coups+assasinations+rgdp60,
            data=growth)
summary(p1_reg)
```

```
##
## Call:
## lm(formula = growth ~ tradeshare + yearsschool + rev_coups +
##     assasinations + rgdp60, data = growth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6897 -0.9459 -0.0565  0.8286  5.1534
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6268915  0.7830280   0.801  0.42663
## tradeshare     1.3408193  0.9600631   1.397  0.16786
## yearsschool    0.5642445  0.1431131   3.943  0.00022 ***
## rev_coups     -2.1504256  1.1185900  -1.922  0.05947 .
## assasinations  0.3225844  0.4880043   0.661  0.51121
## rgdp60        -0.0004613  0.0001508  -3.059  0.00336 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.594 on 58 degrees of freedom
## Multiple R-squared:  0.2911, Adjusted R-squared:   0.23
## F-statistic: 4.764 on 5 and 58 DF,  p-value: 0.001028
```

```
knitr::opts_chunk$set(echo = TRUE)
```

$$Growth = 0.6268915 + 1.3408193 \times TradeShare + 0.5642445 \times YearsSchool + -2.1504256 \times Rev_Coup + 0.3225844 \times Assassinations + -0.0004613 \times rgdp60$$

Interpretation:

The coefficient of Rev_Coup is -2.1504256. Therefore, while keeping the variables TradeShare, Yearsschool, Assassinations, and RGDP60 constant, when Rev_Coup increases by 1 unit, the dependent variable growth decreases by 2.1504256 units on an average.

iii.

Calculating the Average Predicted Growth Rate.

$$Growth = 0.6268915 + 1.3408193 \times TradeShare + 0.5642445 \times YearsSchool + -2.1504256 \times Rev_Coup + 0.3225844 \times Assassinations + -0.0004613 \times rgdp60$$

$$Growth = 0.6268915 + 1.3408193 \times 0.5423919 + 0.5642445 \times 3.9592187 + -2.1504256 \times 0.1700666 + 0.3225844 \times 0.2819010 + -0.0004613 \times 3130.8125339$$

$$Growth = 1.869086$$

iv.

$$New\ TradeShare\ Value = Mean\ TradeShare\ Value + 1$$

$$New\ TradeShare\ Value = 0.5423919 + 1$$

$$New\ TradeShare\ Value = 1.5423919$$

Calculating the Average Predicted Growth Rate.

$$Growth = 0.6268915 + 1.3408193 \times NewTradeShare + 0.5642445 \times YearsSchool + -2.1504256 \times Rev_Coup + 0.3225844 \times Assassinations + -0.0004613 \times rgdp60$$

$$Growth = 0.6268915 + 1.3408193 \times 1.5423919 + 0.5642445 \times 3.9592187 + -2.1504256 \times 0.1700666 + 0.3225844 \times 0.2819010 + -0.0004613 \times 3130.8125339$$

$$Growth = 3.209905$$

Problem 2

```
tuna = read_excel("C:/Users/shrad/Downloads/tuna.xlsx")
knitr::opts_chunk$set(echo = TRUE)
```

i.

```
p2_reg = lm(log(sal1)~apr1+apr2+apr3+disp+dispad,data=tuna)
summary(p2_reg)
```

```
##
## Call:
## lm(formula = log(sal1) ~ apr1 + apr2 + apr3 + disp + dispad,
##     data = tuna)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70001 -0.21573 -0.03785  0.26241  0.74457
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.9848     0.6464  13.900  < 2e-16 ***
## apr1         -3.7463     0.5765  -6.498 5.17e-08 ***
## apr2          1.1495     0.4486   2.562 0.013742 *
## apr3          1.2880     0.6053   2.128 0.038739 *
## disp          0.4237     0.1052   4.028 0.000209 ***
## dispad        1.4313     0.1562   9.165 6.04e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3397 on 46 degrees of freedom
## Multiple R-squared:  0.8428, Adjusted R-squared:  0.8257
## F-statistic: 49.33 on 5 and 46 DF,  p-value: < 2.2e-16
```

```
knitr::opts_chunk$set(echo = TRUE)
```

$$ln(SAL1) = 8.9848 + -3.7463 \times APR1 + 1.1495 \times APR2 + 1.2880 \times APR3 + 0.4237 \times DISP + 1.4313 \times DISPAD$$

ii.

Interpretation of $\beta 1$ - While keeping the variables APR2, and APR3, DISP and DISPAD constant, when APR1 increases by 1 unit, SAL1 decreases/falls by 100*3.7463 % or by 374.63%. That is, keeping everything else constant, when the unit price of the brand no.1 canned tuna increases by 1 unit, the unit sales of brand no.1 canned tuna decreases by 374.63% on an average. Therefore, the demand is elastic. Because, the decrease in demand is greater than the increase in price.

Interpretation of β_2 - While keeping the variables APR1, and APR3, DISP and DISPAD constant, when APR2 increases by 1 unit, SAL1 increases by 100*1.1495 % or by 114.95% on an average. That is, keeping everything else constant, when the unit price of the brand no.2 canned tuna increases by 1 unit, the unit sales of brand no.1 canned tuna increases by 114.95% (average). Therefore, one can say that brand nos.1 and 2 are supplementary goods. Hence, when the unit price of brand 2 increases,the unit price of brand 1 increases.

Interpretation of β_3 - While keeping the variables APR1, APR2, DISP, and DISPAD constant, when APR3 increases by 1 uit, SAL1 increases by 100*1.2880 % or by 128.80% (on average). That is, keeping everything else constant, when the unit price of the brand no.3 canned tuna increases by 1 unit, the unit sales of brand no.1 canned tuna increases by 128.80% (on average). Therefore, one can say that brand nos.1 and 3 are supplementary goods. Hence, when the unit price of brand 3 increases, the unit price of brand 1 increases.

iii.

Interpretation of β_4 and β_5 - The estimate of β_4 is 0.4237 and the estimate of β_5 is 1.4313. While APR1, APR2, APR3, and DISPAD is constant, if DISP takes the value 1, SAL1 increases by 100*(8.9848+0.4237)% of by 940.85% (on average). However, while APR1, APR2, APR3, and DISP is constant, if DISPAD takes the value 1, SAL1 increases by 100*(8.9848+1.4313)%, or by 1041.61% (on average). This is consistent with economic logic because when DISP takes value 1 there is store display but not a newspaper ad for the tuna. However, when DISPAD takes value 1, there is both a store display and a newspaper ad for the canned tuna. While just a store display does increase demand, which is consistent with the positive sign of β_4, posting a newspaper ad in addition to having a store display can increase the demand even more, which is consistent with the increase in demand from when DISPAD takes value 1. Hence, when there is just a store display, although demand increases by 940.85% (on average), with an additional newspaper ad, the demand increases by 1041.61%, which is a 100.76% more increase in demand.

iv.

a).

$H_0 : \beta_4 = 0, H_1 : \beta_4 \neq 0$

```
summary(p2_reg)
```

```
##
## Call:
## lm(formula = log(sal1) ~ apr1 + apr2 + apr3 + disp + dispad,
##     data = tuna)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70001 -0.21573 -0.03785  0.26241  0.74457
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.9848     0.6464  13.900  < 2e-16 ***
## apr1         -3.7463     0.5765  -6.498 5.17e-08 ***
## apr2          1.1495     0.4486   2.562 0.013742 *
## apr3          1.2880     0.6053   2.128 0.038739 *
## disp          0.4237     0.1052   4.028 0.000209 ***
## dispad        1.4313     0.1562   9.165 6.04e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3397 on 46 degrees of freedom
## Multiple R-squared:  0.8428, Adjusted R-squared:  0.8257
## F-statistic: 49.33 on 5 and 46 DF,  p-value: < 2.2e-16
```

```
knitr::opts_chunk$set(echo = TRUE)
```

t-statistic = 4.028, a = 0.05, degrees of freedom = 46, critical points = 1.67 and -1.67. Since our t-statistic (4.028) is greater than our critical value (1.67), we reject the null hypothesis at 5% level of significance. The p-value approach also provides the same result. Out p-value (0.000209) is less than our alpha value (0.05). Therefore, we reject the null hypothesis at 5% level of significance. Hence, the estimated slope is statistically significant and is not equal to zero.

b).

$H_0 : \beta_5 = 0, H_1 : \beta_5 \neq 0$

t-statistic = 9.165, a = 0.05, degrees of freedom = 46, critical points = 1.67 and -1.67. Since our t-statistic (9.165) is greater than our critical value (1.67), we reject the null hypothesis at 5% level of significance. The p-value approach also provides the same result. Our p-value (6.04e-12) is less than our alpha value (0.05). Therefore, we reject the null hypothesis at 5% level of significance. Hence, the estimated slope is statistically significant and is not equal to zero.

c).

$H_0 : \beta_4 = 0, \beta_5 = 0, H_1 : \beta_4 or \beta_5 \neq 0$

```
install.packages("car",repos = "http://cran.us.r-project.org")
```

```
## package 'car' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\shrad\AppData\Local\Temp\Rtmp6vZrPp\downloaded_packages
```

```
library(car)
nullhyp = c("disp","dispad")
linearHypothesis(p2_reg,nullhyp)
```

```
## Linear hypothesis test
##
## Hypothesis:
## disp = 0
## dispad = 0
##
## Model 1: restricted model
## Model 2: log(sal1) ~ apr1 + apr2 + apr3 + disp + dispad
##
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     48 15.0023
## 2     46  5.3073  2     9.695 42.015 4.172e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
knitr::opts_chunk$set(echo = TRUE)
```

a = 0.05, F-statistic = 42.015, Critical Value = 3.19. Since our F-statistic value (42.015) exceeds the critical value (3.19), we can reject the null hypothesis with 5% significance level. The p-value approach also provides the same result. Our p-value (4.172e-11) is less than our alpha value (0.05). Therefore, we can reject the null hypothesis at 5% significance level. Hence, $\beta_4 \ and \ \beta_5 \ are \neq 0$.

d).

$H_0 : \beta_5 \leq \beta_4, H_1 : \beta_5 > \beta_4$

Rearranging the Null and Alternative Hypothesis. $H_0 : \beta_4 - \beta_5 \geq 0, H_1 = \beta_4 - \beta_5 < 0$

$\theta = \beta_5 - \beta_4$

$$ln(SAL1) = \beta_0 + \beta_1 APR1 + \beta_2 APR2 + \beta_3 APR3 + \beta_4 DISP + (\theta + \beta_4) DISPAD + u$$
$$ln(SAL1) = \beta_0 + \beta_1 APR1 + \beta_2 APR2 + \beta_4 DISP + \theta DISPAD + \beta_4 DISPAD + u$$
$$ln(SAL1) = \beta_0 + \beta_1 APR1 + \beta_2 APR2 + \beta_4 (DISP + DISPAD) + \theta DISPAD + u$$

```
tuna$new = tuna$dispad + tuna$disp
new_reg = lm(log(sal1) ~ apr1 + apr2 + apr3 + new + dispad, data=tuna)
summary(new_reg)
```

```
##
## Call:
## lm(formula = log(sal1) ~ apr1 + apr2 + apr3 + new + dispad, data = tuna)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70001 -0.21573 -0.03785  0.26241  0.74457
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.9848     0.6464  13.900  < 2e-16 ***
## apr1         -3.7463     0.5765  -6.498 5.17e-08 ***
## apr2          1.1495     0.4486   2.562 0.013742 *
## apr3          1.2880     0.6053   2.128 0.038739 *
## new           0.4237     0.1052   4.028 0.000209 ***
## dispad        1.0075     0.1469   6.858 1.49e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3397 on 46 degrees of freedom
## Multiple R-squared:  0.8428, Adjusted R-squared:  0.8257
## F-statistic: 49.33 on 5 and 46 DF,  p-value: < 2.2e-16
```

```
knitr::opts_chunk$set(echo = TRUE)
```

$t - statistic = 1.0075/0.1469 = 6.858$

a = 0.05, t-statistic = 6.858, Critical Value = -1.67. Since our t-statistic (6.858) is greater than out critical value (-1.67), we reject the null hypothesis at 5% significance level. Our p-value approach also provides the same result. Our p-value (1.49e-08) is less than our alpha value (0.05). Therefore, we reject the null hypothesis at 5% level of significance. Hence, β_4 is neither equal to nor greater than β_5. The result is statistically significant.

Our p-value (6.04e-12) is less than our alpha value (0.05). Therefore, we reject the null hypothesis at 5% level of significance. Hence, the estimated slope is statistically significant and is not equal to zero.

e).

The hypothesis tests in part (iv) looks at the relationship between the variables "disp" and "dispad". The first two hypothesis tests looks at whether or not the variables impact sales. The subsequent hypothesis tests look at which of the two variables have a greater impact,if not equal, on sales. These tests are crucial for a supermarket executive because the primary objective of an organization is to make profit through sales. And it is crucial to identify and rank factors that positively and negative effect sales to study its impact on revenue. For example, from the first two hypothesis tests, we have identified that both display of product without a newspaper ad, and display of product with a newspaper ad affect sales positively, however, from the fourth test we know that having a newspaper advertisement in addition to having a product display effects the sales more positively. For example, in the fourth hypothesis test, we proved that just having a display is less effective than having both a display and a newspaper ad. This is a marketing insight that can help the executive increase profit through sales.

Problem 3

```
cocaine = read_excel("C:/Users/shrad/Downloads/cocaine.xlsx")
knitr::opts_chunk$set(echo = TRUE)
```

i.

Expectation for β_1 - β_1 is expected to be positive. An increasing number of grams sold per sale indicates an increase in demand which can increase the average price per gram.

Expectation for β_2 - β_2 is expected to be positive. As quality increases, or purity percentage increases, the value of each gram of cocaine increases. Therefore, as quality increases, the average price per gram is also likely to increase.

Expectation for β_3 - β_3 is expected to be positive. As time increases, the value for money is likely to decrease due to inflation. Hence, the average price of cocaine per gram is also likely to increase.

ii.

```
p3_reg = lm(price~quant+qual+trend,data = cocaine)
summary(p3_reg)
```

```
##
## Call:
## lm(formula = price ~ quant + qual + trend, data = cocaine)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.479 -12.014  -3.743  13.969  43.753
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 90.84669    8.58025  10.588 1.39e-14 ***
## quant       -0.05997    0.01018  -5.892 2.85e-07 ***
## qual         0.11621    0.20326   0.572   0.5700
## trend       -2.35458    1.38612  -1.699   0.0954 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.06 on 52 degrees of freedom
## Multiple R-squared:  0.5097, Adjusted R-squared:  0.4814
## F-statistic: 18.02 on 3 and 52 DF,  p-value: 3.806e-08
```

```
knitr::opts_chunk$set(echo = TRUE)
```

$$Price = 90.84669 + -0.05997 \times Quant + 0.11621 \times Qual + -2.35458 \times Trend$$

Interpreting the Results.

Interpretation of β_1 - The value of β_1 is -0.05997. Therefore, while the variables "Qual" and "Trend" are held constant, when "Quant" increases by 1 gram, the variable "Price" decreases by 0.05997 dollars on an average. This is, however, unlike my expectation. Unlike my expectations, when quantity sold increases by a gram, the average price per gram sold decreases.

Interpretation of β_2 - The value of β_1 is 0.11621. Therefore, while the variables "Quant" and "Trend" are held constant, when "Qual" increases by 1%, the variable "Price" increases by 0.11621 dollars on an average. This sign of β_2 matches my expectations. Hence, as quality of cocaine increases, the average price per gram of cocaine sold increases.

Interpretation of β_3 - The value of β_3 is -2.35458. Therefore, while the variables "Quant" and "Qual" are held constant, when "Trend" increases by 1 year, the variable "Price" decreases by2.35458 dollars on an average. This is, however, not consistent with my expectation. Hence, with every additional year passing by, the average price of every gram of cocaine sold decreases instead of increasing.

iii.

```
summary(p3_reg)
```

```
##
## Call:
## lm(formula = price ~ quant + qual + trend, data = cocaine)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -43.479 -12.014  -3.743  13.969  43.753
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 90.84669    8.58025  10.588 1.39e-14 ***
## quant       -0.05997    0.01018  -5.892 2.85e-07 ***
## qual         0.11621    0.20326   0.572   0.5700
## trend       -2.35458    1.38612  -1.699   0.0954 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.06 on 52 degrees of freedom
## Multiple R-squared:  0.5097, Adjusted R-squared:  0.4814
## F-statistic: 18.02 on 3 and 52 DF,  p-value: 3.806e-08
```

```
knitr::opts_chunk$set(echo = TRUE)
```

The r-squared value is 0.5097, therefore, 50.97% of the variation in the dependent variable “price” is explained jointly by variation in quantity, quality, and time.

iv.

$HO : \beta_1 \geq 0, H1 : \beta_1 < 0$

t-statistic = -5.892, a = 0.05, degrees of freedom = 52, critical point = -1.67.Since out t-statistic (-5.892) is less than the critical value (-1.67), we reject the null hypothesis. The p-value approach also provides the same result. Our p-value (2.85e-07) is less than out alpha value (0.05). Therefore, we reject the null hypothesis at 5% level of significance. Hence, The relationship between quantity and price is neither 0 nor positive. It is negative as believed by the sellers.

v.

$H0 : \beta_2 \leq 0, H1 : \beta_2 > 0$

t-statistic = 0.572, a = 0.05, degrees of freedom = 52, critical point = 1.67. Since our t-statistic (0.572) is less than out critical value (1.67), we fail to reject the null hypothesis. The p-value approach also provides the same result. Our p-value (0.5700) is greater than out alpha value (0.05). Therefore, we fail to reject the null hypothesis. Hence, it is possible that there is a either no relation or a negative relation between cocaine price and quality. And one cannot conclusively say that a premium is paid for better quality cocaine.

vi.

The value of β_3 = -2.35458. Therefore, with every successive year, the price of cocaine, on an average, decreases by 2.35458 dollars. The increase in supply of cocaine may increase competition and therefore reduce prices of cocaine. Furthermore, with ever successive year, technological improvements may reduce the costs of producing cocaine, which can further reduce prices of cocaine.

Problem 4

```
library(wooldridge)
#data("attend")
#?attend
```

i.

```
#Minimum Values.
mina = min(attend$atndrte)
mina
```

```
## [1] 6.25
```

```
minb = min(attend$priGPA)
minb
```

```
## [1] 0.857
```

```
minc = min(attend$ACT)
minc
```

```
## [1] 13
```

```
#Maximum Values.
maxa = max(attend$atndrte)
maxa
```

```
## [1] 100
```

```
maxb = max(attend$priGPA)
maxb
```

```
## [1] 3.93
```

```
maxc = max(attend$ACT)
maxc
```

```
## [1] 32
```

```
#Mean/Average Values.
meana = mean(attend$atndrte)
meana
```

```
## [1] 81.70956
```

```
meanb = mean(attend$priGPA)
meanb
```

```
## [1] 2.586775
```

```
meanc = mean(attend$ACT)
meanc
```

```
## [1] 22.51029
```

ii.

```
p4_reg = lm(atndrte~priGPA+ACT,data=attend)
summary(p4_reg)
```

```
##
## Call:
## lm(formula = atndrte ~ priGPA + ACT, data = attend)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.373  -6.765   2.125   9.635  29.615
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   75.700      3.884   19.49   <2e-16 ***
## priGPA        17.261      1.083   15.94   <2e-16 ***
## ACT           -1.717      0.169  -10.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.38 on 677 degrees of freedom
## Multiple R-squared:  0.2906, Adjusted R-squared:  0.2885
## F-statistic: 138.7 on 2 and 677 DF,  p-value: < 2.2e-16
```

$$atndrte = 75.700 + 17.261 \times priGPA + -1.717 \times ACT$$

The intercept value is 75.700. Therefore, for a student with 0 cumulative GPA prior to term and 0 ACT score, the attendance percentage is equal to 75.700. However, this value does not have a meaningful interpretation as the data set does not include a student with both 0 cumulative GPA prior to term and 0 ACT score.

iii.

The value of β_1 = 17.261. Therefore, while holding "ACT" constant, with every unit increase in the cumulative GPA earned prior the term, the attendance increases by 17.261% (on average). Perhaps, an increase in GPA motivates the student increase their academic engagement to further increase their GPA. The value of β_2 is -1.717. Therefore, while holding "priGPA" constant, when ACT scores increase by 1 unit, the attendance rate decreases by 1.717%. Although the percentage decrease is small, It is surprising to see a negative relationship between ACT scores and attendance rates. Perhaps, with an increase in their ACT scores, students are devoting greater time to self-study.

iv.

Calculating Predicted "atndrte".

$$atndrte = 75.700 + 17.261 \times priGPA + -1.717 \times ACT$$

$$atndrte = 75.700 + 17.261 \times 3.65 + -1.717 \times 20$$

$$atndrte = 104.3627\%$$

This value is misleading as the attendance percentage cannot be greater than zero. In the given data set, there is no student with an attendance percentage greater than 100. Hence, there is no student with 104.3627% attendance rate.

```
sum(attend$atndrte>100)
```

```
## [1] 0
```

v.

Calculating the attendance rate for Student A.

$$atndrte = 75.700 + 17.261 \times priGPA + -1.717 \times ACT$$

$$atndrte = 75.700 + 17.261 \times 3.1 + -1.717 \times 21$$

$$atndrte = 93.1521\%$$

Calculating the attendance rate for Student B.

$$atndrte = 75.700 + 17.261 \times priGPA + -1.717 \times ACT$$

$$atndrte = 75.700 + 17.261 \times 2.1 + -1.717 \times 26$$

$$atndrte = 67.3061\%$$

Calculating the difference in their attendance rates.

$$Difference\ in\ Attendance\ Rates = 93.1521 - 67.3061$$

$$Difference\ in\ Attendance\ Rates = 25.846\%$$

Problem 5

i.

$$H0 : \beta_1 = 0, H1 : \beta_1 \neq 0$$

$$t - statistic = \hat{\beta}_1 - \beta_1 / \hat{SE}(\hat{\beta}_1)$$

$$t - statistic = 0.567 - 0/1.23 = 0.4609756$$

Assuming the degrees of freedom is 20 or more, and with a alpha value of 0.5, since the critical value is lesser than 2 for a two sides test, we cannot reject the null hypothesis. Therefore, the coefficient of BDR is not statistically significant from zero.

ii.

Ideally, four-bathroom houses should cost more than three-bedroom houses. However, this is inconsistent with our results in part i). According to i), an addition bedroom does not effect the selling price of the house. Hence, our coefficient was not statistically significant. However, the coefficient of BDR is 0.567, that is, while holding the variables "Bath", "Hsize", "Lsize", "Age", and "Poor" constant, for every additional bedroom, the selling price of the house increases by (1000*0.567 = $567 dollars) on an average. Although the coefficient is positive, the value is relatively small. For example, from the regression equation, an additional bathroom has a greater positive effect on the selling price than an additional bedroom, while holding other independent variables constant.

iii.

Although square feet is a convenient unit of measurement of interpretation, the choice of scale depends on how lot size and selling price are related to each other. If the relationship is linear, then using square feet as the unit of measurement is appropriate. However, if the relationship is non- linear, using square feet can still be effective if it is accompanied with transformation of the independent variable (lot size). For example, if the relationship between the variables are multiplicative or exponential, a logarithm transformation can be helpful to capture the relationship while retaining the units of measurement. Based on just the regression equation, it is hard to say whether lot size and selling price share a linear relationship or not. If they do, square feet is an appropriate unit of measurement, if not, we could either use an alternative unit of measurement that captures the non-linear relationship between the variables, or transform the lot size variable.

iv.

$$H0 : \beta_1, \beta + 5 = 0, H1 : \beta_1, \beta_5 \neq 0$$

```
d1 = 2
d2 = Inf
critical_value = qf(0.10,d1,d2)
critical_value
```

```
## [1] 0.1053605
```

The F-stat given (2.38) is greater than the critical value (0.1053605). Therefore, we reject the null hypothesis at 10% significance level. Therefore, the coefficients of BDR and Age are statistically significant.