IE 522 HW10

Q1

7 Points

The dataset USMacroG in the AER package contains quarterly macroeconomic variables from 1950 to 2000. Click **HERE** for details. We want to study the relationship between consumption (dependent variable) and other variables (independent variables). In all of the regression fitting below, it's assumed that the constant regressor is included, significance level is 1% for tests, and error terms in the regression models are normally distributed.

```
library(AER)
data("USMacroG")
USMacroG=na.omit(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
```

Choose Files No file chosen

`USMacroG` contains the original variables before differencing, `Macro` contains the variables after differencing (that is, quarterly changes).

Q1.1

1 Point

Regress consumption on the other 11 variables. What regressors are significant at 0.05 significance level?

```
Call:
lm(formula = Macro$consumption ~ ., data = Macro)

Residuals:
    Min      1Q  Median      3Q     Max
-34.287  -6.922  -1.108   6.922  38.723

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.14859    3.07949  -1.022  0.30787
gdp           0.77061    0.04636  16.624  < 2e-16 ***
invest       -0.74133    0.05781 -12.822  < 2e-16 ***
government   -0.58702    0.08414  -6.976  4.9e-11 ***
dpi           0.10500    0.03550   2.958  0.00349 **
cpi          -0.39117    0.46705  -0.838  0.40334
m1            0.14953    0.11079   1.350  0.17873
tbill        15.04615  245.71289   0.061  0.95124
                        3.82193   3.05721  -1.250  0.21279
population   10.22284    3.99745   2.557  0.01133 *
inflation   -16.05374  245.71204  -0.065  0.94798
interest    -16.77342  245.72689  -0.068  0.94565
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.71 on 190 degrees of freedom
Multiple R-squared:  0.7533, Adjusted R-squared:  0.739
F-statistic: 52.74 on 11 and 190 DF,  p-value: < 2.2e-16
```

Choose Files No file chosen

The regressors that have the p value smaller than alpha will be significant. That is, gdp, invest, government, dpi, population are significant.

CURRENTLY UPLOADED FILES

Q1-1.R Download Remove

```r
#Q1-1
#Data Preparation
library(AER)
data("USMacroG")
USMacroG=na.omit(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)

#Regression on all models
fit1_1 = lm(Macro$consumption ~ ., Macro)
summary(fit1_1)

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:20 AM**

Choose Files No file chosen

Q1.2

1 Point

Using the best subset selection and C_p, what regressors are selected? Write down the equation of the best linear model selected (keep 2 digits after the decimal point for the coefficients). In this model, consumption is positively related to what variables and negatively related to what variables? Are the regressors in 1.1 selected?

Best subsets:
Subset selection object

```
          gdp invest government dpi cpi m1  tbill unemp population inflation interest
1  ( 1 ) "*" " "    " "      " " " " " " " " "   " "   " "      " "      " "
2  ( 1 ) "*" "*"    " "      " " " " " " " " "   " "   " "      " "      " "
3  ( 1 ) "*" "*"    "*"      " " " " " " " " "   " "   " "      " "      " "
4  ( 1 ) "*" "*"    "*"      "*" " " " " " " "   " "   " "      " "      " "
5  ( 1 ) "*" "*"    "*"      "*" " " " " " " "   " "   "*"      " "      " "
6  ( 1 ) "*" "*"    "*"      "*" " " " " " " "   " "   "*"      " "      "*"
7  ( 1 ) "*" "*"    "*"      "*" " " "*" " " "   " "   "*"      " "      "*"
8  ( 1 ) "*" "*"    "*"      "*" " " "*" " " "   "*"   "*"      " "      "*"
9  ( 1 ) "*" "*"    "*"      "*" "*" "*" " " "   "*"   "*"      " "      "*"
10  ( 1 ) "*" "*"    "*"      "*" "*" "*" " " "   "*"   "*"      "*"      "*"
11  ( 1 ) "*" "*"    "*"      "*" "*" "*" "*"   "*"   "*"      "*"      "*"
```

Based on the best subsets selection and Cp plot, gdp, invest, government, dpi, population, interest are selected.

```
> coef(best_subsets_1_2, j)
(Intercept)         gdp      invest  government         dpi  population    interest
 -3.0510494   0.7965592  -0.7532173  -0.5838572   0.1026401   8.5459340  -0.6407097
```

---

consumption = 0.8*gdp - 0.75*invest - 0.58*government + 0.1*dpi + 8.55*population - 0.64*interest -3.05

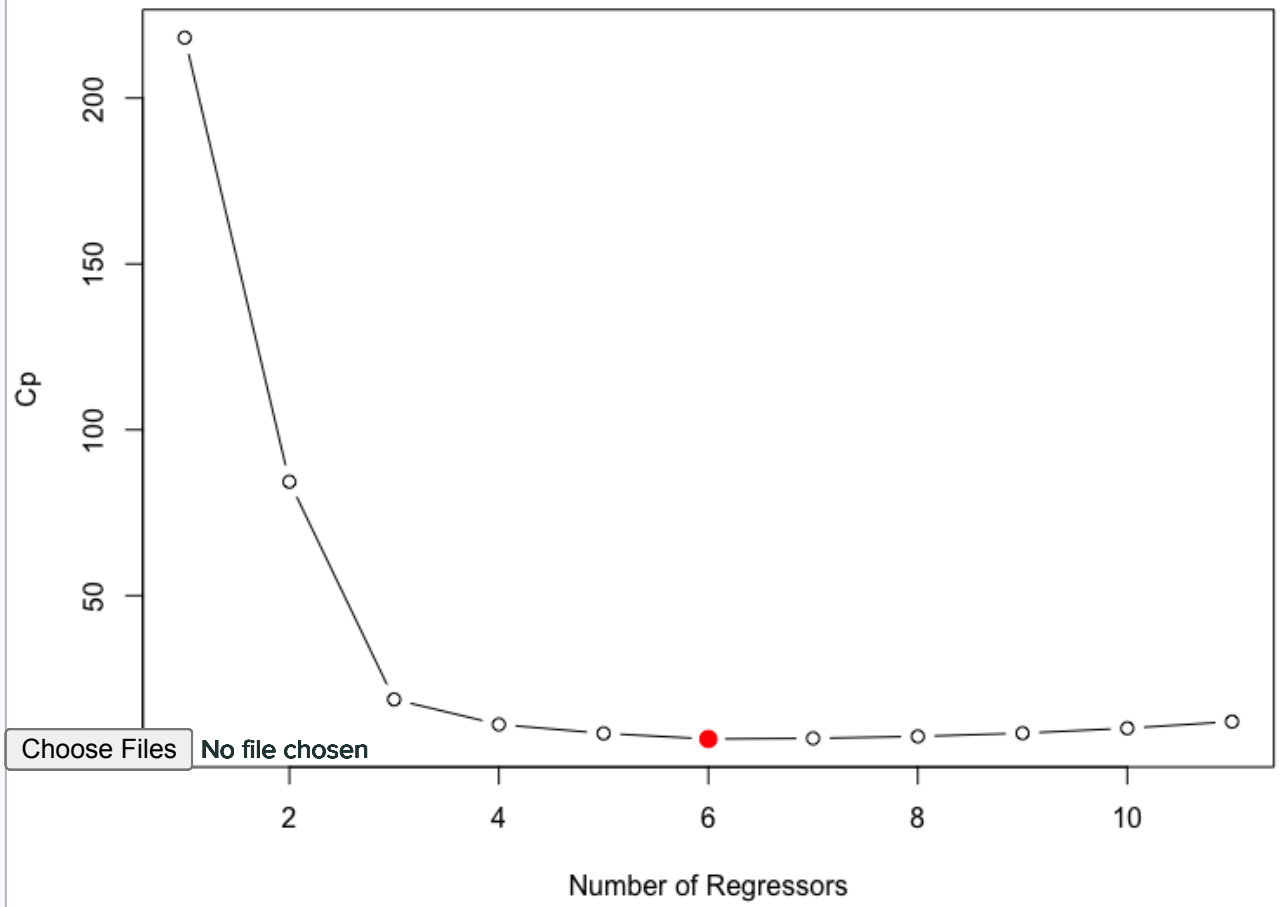
Positively related to: gdp, dpi, population.
Negatively related to: invest, government, interest.

Besides interest, all the regressors selected via Cp is selected in 1.1.

CURRENTLY UPLOADED FILES

Cp_Plot_Q1-2.png

Download Remove



Choose Files No file chosen

Q1-2.R

Download Remove

```
1  #Q1-2
2  #Data Preparation
3  library(AER)
4  data("USMacroG")
```

```r
USMacroG=na.omit(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)

#Best subset selection
library(leaps)
best_subsets_1_2 = regsubsets(Macro$consumption ~ ., Macro, nvmax = 11)
result1_2 = summary(best_subsets_1_2)
result1_2

#Best model via Cp
plot(result1_2$cp, xlab = "Number of Regressors", ylab = "Cp", type = "b")
j = which.min(result1_2$cp)
points(j, result1_2$cp[j], col = "red", cex = 2, pch = 20)

#Get coefficient
coef(best_subsets_1_2, j)


```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:20 AM**

Choose Files No file chosen

Q1.3

1 Point

Without running the forward and backward selection, from the result of the best subset selection, can you tell whether the forward and backward selection will give the same result as the best subset selection? Explain.

Selection Algorithm: exhaustive

gdp invest government dpi cpi m1 tbill unemp population inflation interest

```
1 ( 1 ) "*" " "   " "     " " " " " " " " "  " "  " "     " "     " "
2 ( 1 ) "*" "*"   " "     " " " " " " " " "  " "  " "     " "     " "
3 ( 1 ) "*" "*"   "*"     " " " " " " " " "  " "  " "     " "     " "
4 ( 1 ) "*" "*"   "*"     "*" " " " " " " "  " "  " "     " "     " "
5 ( 1 ) "*" "*"   "*"     "*" " " " " " " "  " "  "*"     " "     " "
6 ( 1 ) "*" "*"   "*"     "*" " " " " " " "  " "  "*"     " "     "*"
7 ( 1 ) "*" "*"   "*"     "*" " " "*" " " "  " "  "*"     " "     "*"
8 ( 1 ) "*" "*"   "*"     "*" " " "*" " " "  "*"  "*"     " "     "*"
9 ( 1 ) "*" "*"   "*"     "*" "*" "*" " " "  "*"  "*"     " "     "*"
10 ( 1 ) "*" "*"   "*"     "*" "*" "*" " " "  "*"  "*"     "*"     "*"
11 ( 1 ) "*" "*"   "*"     "*" "*" "*" "*"  "*"  "*"     "*"     "*"
```

The forward selection will be exactly identical with the outcome of the best subsets selection, since in best subsets selection, the variables that have been selected are continuously existed in the selection. That is, none of the variables are dropped after being selcted.

The backward selection will also be identical with the (reverse version of the) outcome of best subsets selection, since the vairables that have been dropped in the best subsets selection is not being selected again.

Choose Files No file chosen

CURRENTLY UPLOADED FILES

Q1-3.R Download Remove

```r
#Q1-3
#Data Preparation
library(AER)
data("USMacroG")
USMacroG=na.omit(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)

```

```
#Best subset selection
library(leaps)
best_subsets_1_3 = regsubsets(Macro$consumption ~ ., Macro, nvmax = 11)
result1_3 = summary(best_subsets_1_3)
result1_3

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:20 AM**

Q1.4

1 Point

Using the best subset selection and R^2_{adj}, what regressors are selected? Write down the equation of the best linear model selected (keep 2 digits after the decimal point for the coefficients). Is the best model selected here the same as the best model selected in 1.2 using C_p?

Selection Algorithm: exhaustive

Choose Files No file chosen

```
          income employment dpi cpi m1  tbill unemp population inflation interest
1  ( 1 ) "*" " "   " "      " " " " " " " " " " " "  " "      " "     " "
2  ( 1 ) "*" "*"   " "      " " " " " " " " " " " "  " "      " "     " "
3  ( 1 ) "*" "*"   "*"      " " " " " " " " " " " "  " "      " "     " "
4  ( 1 ) "*" "*"   "*"      "*" " " " " " " " " " "  " "      " "     " "
5  ( 1 ) "*" "*"   "*"      "*" " " " " " " " " " "  "*"      " "     " "
6  ( 1 ) "*" "*"   "*"      "*" " " " " " " " " " "  "*"      " "     "*"
7  ( 1 ) "*" "*"   "*"      "*" " " "*" " " " " "  "*"      " "     "*"
8  ( 1 ) "*" "*"   "*"      "*" " " "*" " " "*"  "*"      " "     "*"
9  ( 1 ) "*" "*"   "*"      "*" "*" "*" " " "*"  "*"      " "     "*"
```

10 (1) "*" "*" "*" "*" "*" " " "*" "*" "*" "*" "*"
11 (1) "*" "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"

Via best subsets sselection and their corresponding adjR2 value, gdp, invest, government, dpi, cpi, m1, unemp, population, interest.

> coef(best_subsets_1_4, i)
(Intercept) gdp invest government dpi cpi m1 unemp population interest
-3.1257250 0.7707953 -0.7435715 -0.5816921 0.1028569 -0.4621522 0.1662844 -3.0792789 10.3345440 -0.7364632

---

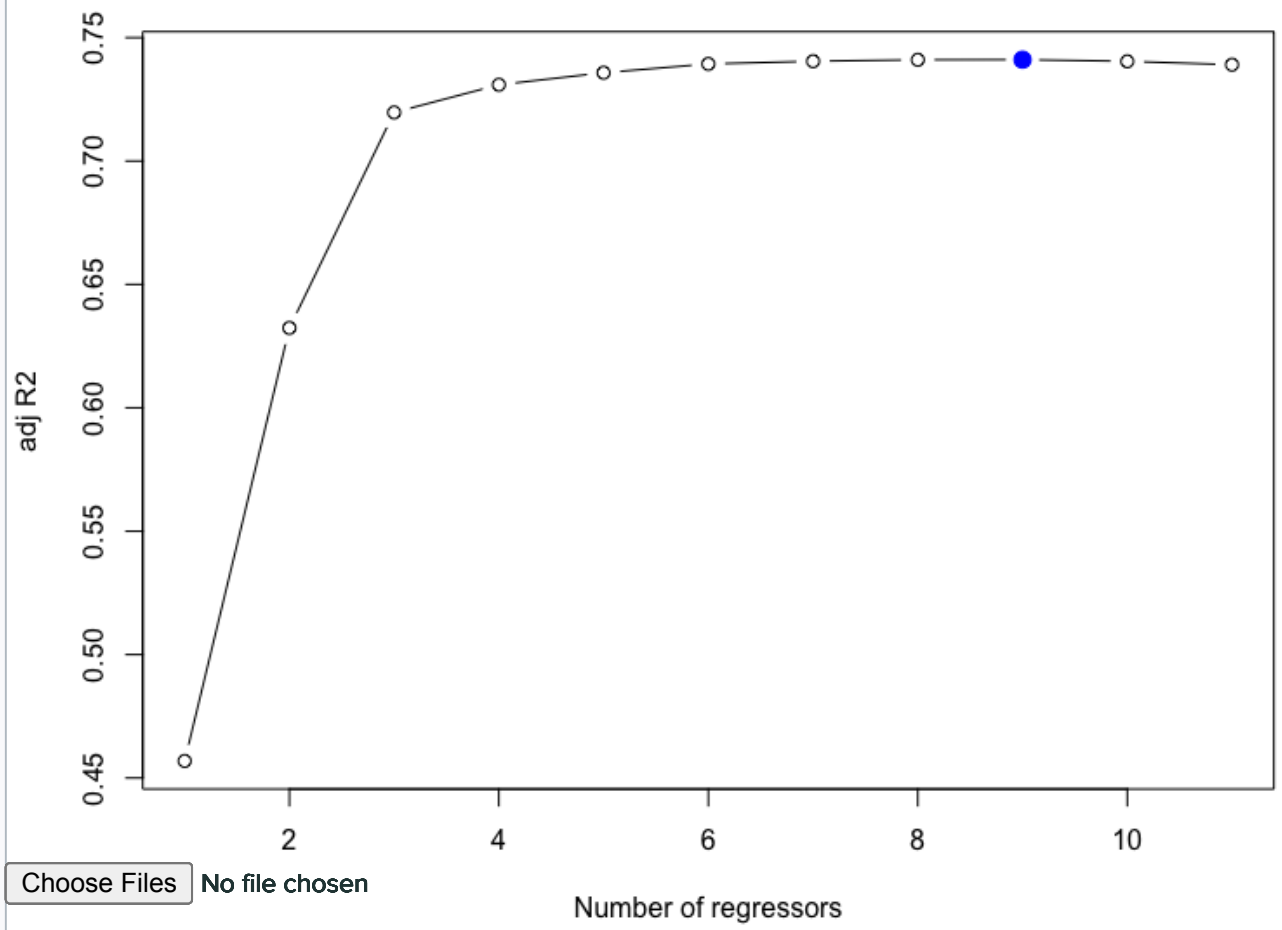
consumption = 0.77*gdp - 0.74*invest -0.58*government +0.1*dpi - 0.46*cpi + 0.17*m1 + -3.08*unemp + 10.34*population -0.74*interest -3.13

It is different from the model that we derived in Q1-2.

CURRENTLY UPLOADED FILES

adjR2_Plot_Q1_-4.png Download Remove

Choose Files No file chosen



Choose Files No file chosen

▾ Q1-4.R

Download ✖ Remove

```r
#Q1-4
#Data Preparation
library(AER)
data("USMacroG")
USMacroG=na.omit(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)
```

```r

#Best subset selection
library(leaps)
best_subsets_1_4 = regsubsets(Macro$consumption ~ ., Macro, nvmax = 11)
result1_4 = summary(best_subsets_1_4)
result1_4

#Best model via R2adj
plot(result1_4$adjr2, xlab = "Number of regressors", ylab = "adj R2", type =
'b')
i = which.max(result1_4$adjr2)
points(i, result1_4$adjr2[i], col = 'blue', cex = 2, pch = 20)

#Get coefficient
coef(best_subsets_1_4, i)


```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:21 AM**

Choose Files No file chosen

Q1.5

1 Point

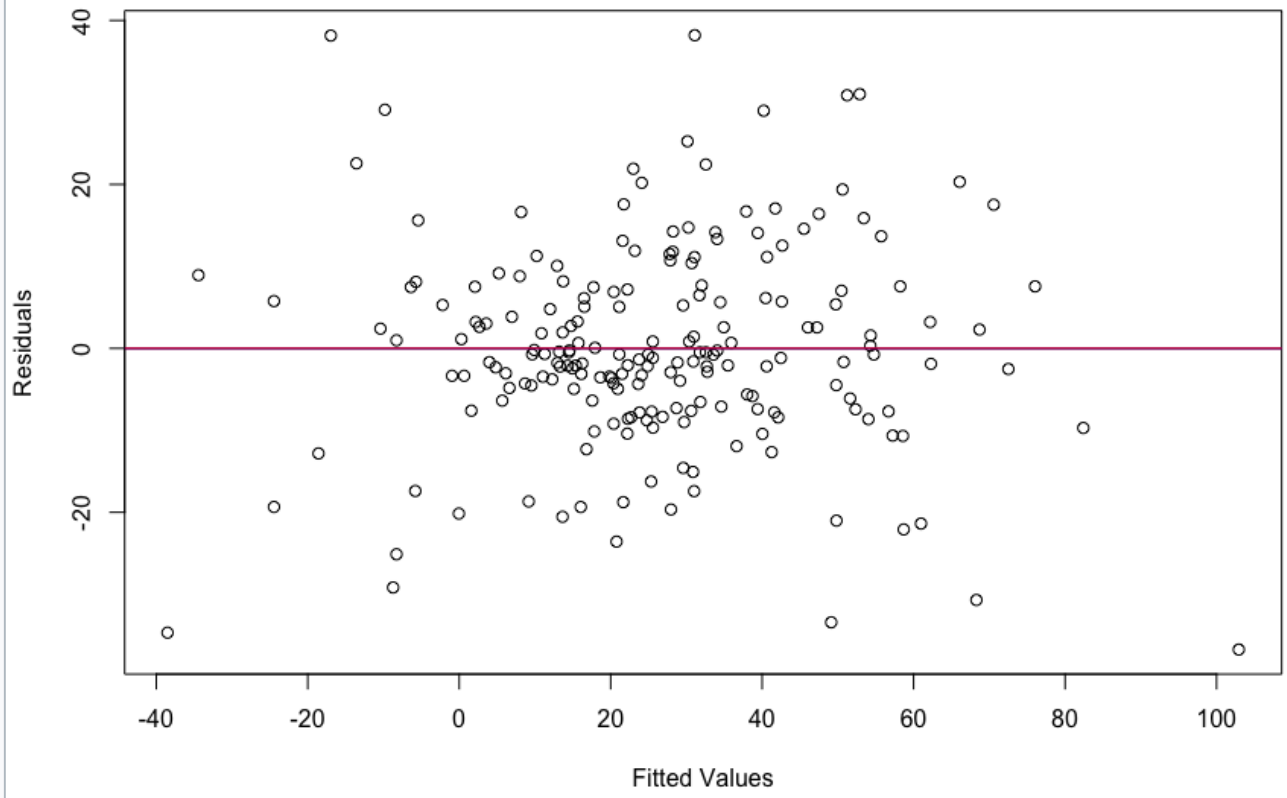
For the best model selected in 1.2, does the residual plot show obvious non-linearity and non-constant variances for the error terms?

From the residual plot, it shows an obvious nonlinearity, which shows its randomness. And it as well does not show the constant variance.

CURRENTLY UPLOADED FILES

Residual_Plot_Q1-5.png

Download Remove



Choose Files No file chosen

Q1-5.R

Download Remove

```
1  #Q1-5
2  #Data Preparation
3  library(AER)
4  data("USMacroG")
5  USMacroG=na.omit(USMacroG)
6  Macro=as.data.frame(apply(USMacroG,2,diff))
7  dim(Macro)
8
9  #Regress on gdp, invest, government, dpi, population, interest
```

```
fit1_5 = lm(Macro$consumption ~ Macro$gdp + Macro$invest + Macro$government
            + Macro$dpi + Macro$population + Macro$interest)
summary(fit1_5)
plot(fit1_5$fitted.values, fit1_5$residuals, xlab = "Fitted Values", ylab
     = "Residuals")
abline(0,0, col = "red")

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:21 AM**

Q1.6

1 Point

For the best model selected in 1.2, is there evidence of the violation of zero correlation assumption?

Series: fit1_6$residuals
ARIMA(4,1,0)

Choose Files No file chosen

Coefficients:

	ar1	ar2	ar3	ar4
	-0.8986	-0.7410	-0.4153	-0.1259
s.e.	0.0705	0.0904	0.0901	0.0709

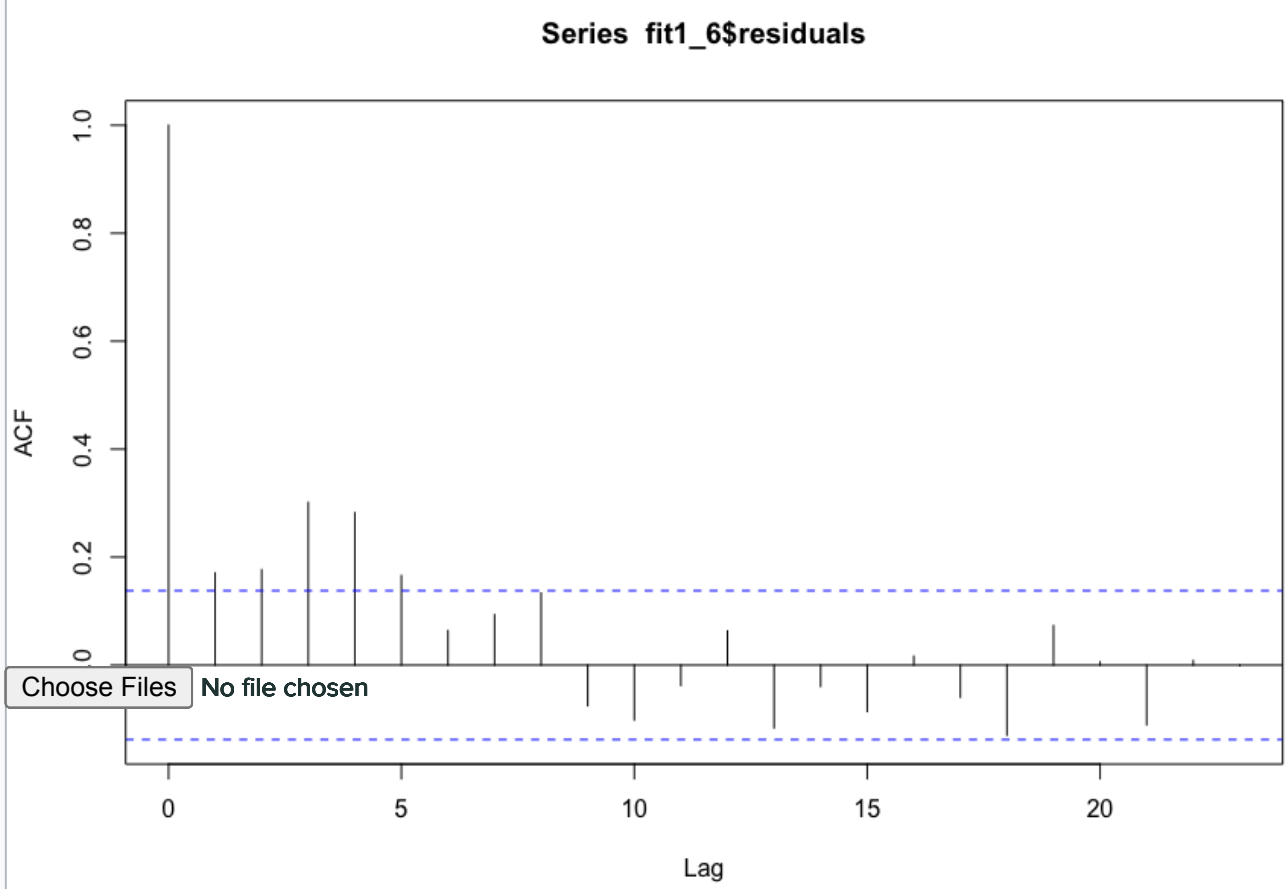
sigma^2 = 142.2: log likelihood = -781.91
AIC=1573.82 AICc=1574.13 BIC=1590.34

Yes, the autocorrelation value we get with first five lag are quite high that makes it violates

the zero correlation assumption. Ans the model selected via auto.arima shows that none of ma models is selected, means that ACF is not converging to 0.

CURRENTLY UPLOADED FILES

ACF_Plot_Q1_6.png Download Remove



Choose Files No file chosen

Q1-6.R Download Remove

```
1  #Q1-6
2  #Data Preparation
3  library(AER)
```

```r
data("USMacroG")
USMacroG=na.omit(USMacroG)
Macro=as.data.frame(apply(USMacroG,2,diff))
dim(Macro)

#Regress on gdp, invest, government, dpi, population, interest
fit1_6 = lm(Macro$consumption ~ Macro$gdp + Macro$invest + Macro$government
            + Macro$dpi + Macro$population + Macro$interest)
summary(fit1_6)

#Plot ACF
acf(fit1_6$residuals)

#Test via arima
library(forecast)
auto.arima(fit1_6$residuals)

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:21 AM**

Choose Files No file chosen

Q1.7

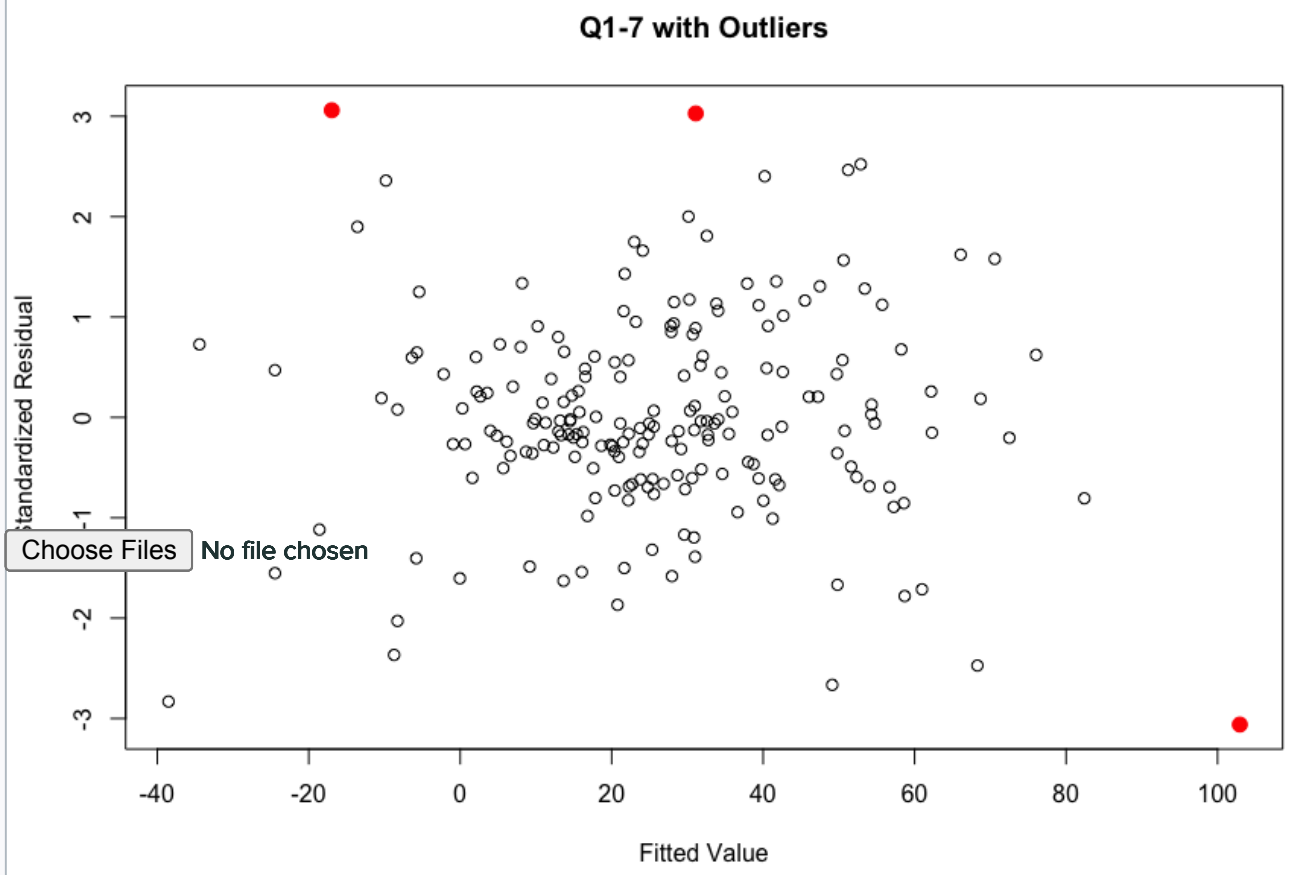
1 Point

For the best model selected in 1.2, plot the standardized residuals. What's the average of leverages in this model? Using "which()" function in R, identify observations that are outliers (with absolute standardized residual greater than 3) and observations that are high leverage points (with leverage greater than 2 times the average leverage). Is there any observation that's both an outlier and a high leverage point?

avg_of_leverage = K/n = 0.03465347
Outliers = 112 129 195
HLP = 4 5 99 100 112 121 141 148 151 171 175 191 198 199 200 201 202
The observation 112 is both Outlier and HLP.

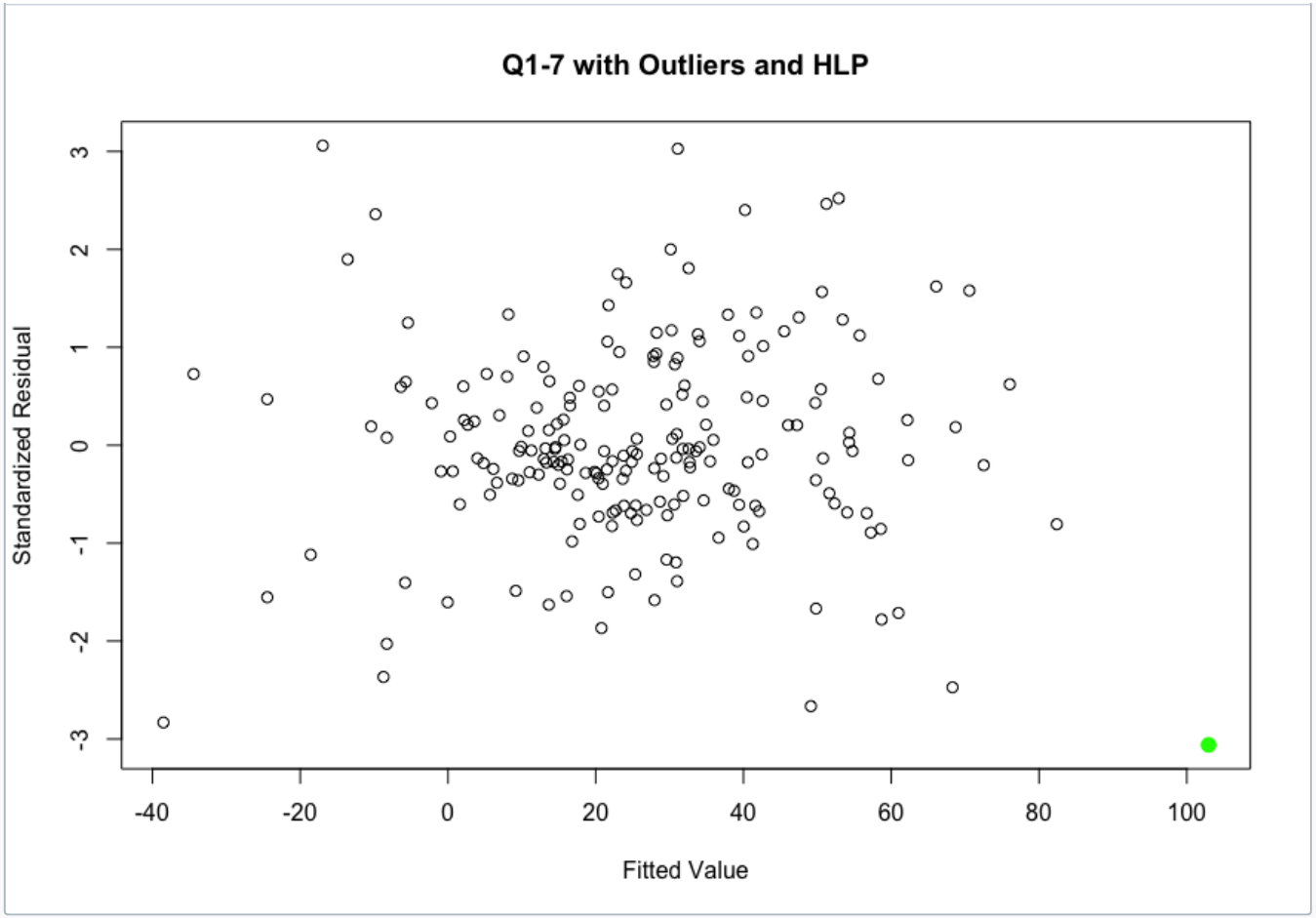
CURRENTLY UPLOADED FILES

Outliers finding Q1-7.png Download Remove



Choose Files No file chosen

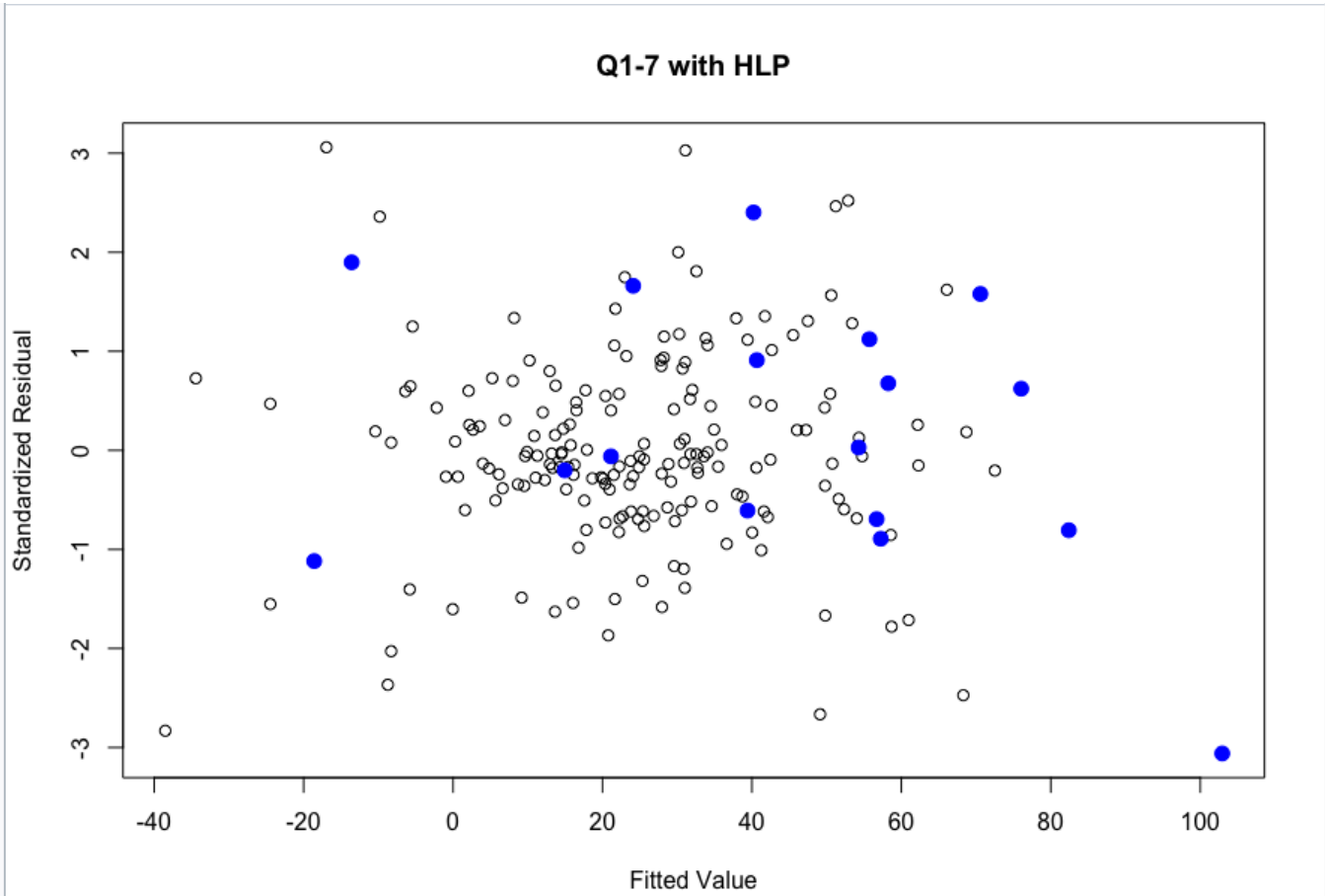
Outliers and HLP at once.png Download Remove



Choose Files No file chosen

Q1-7.png

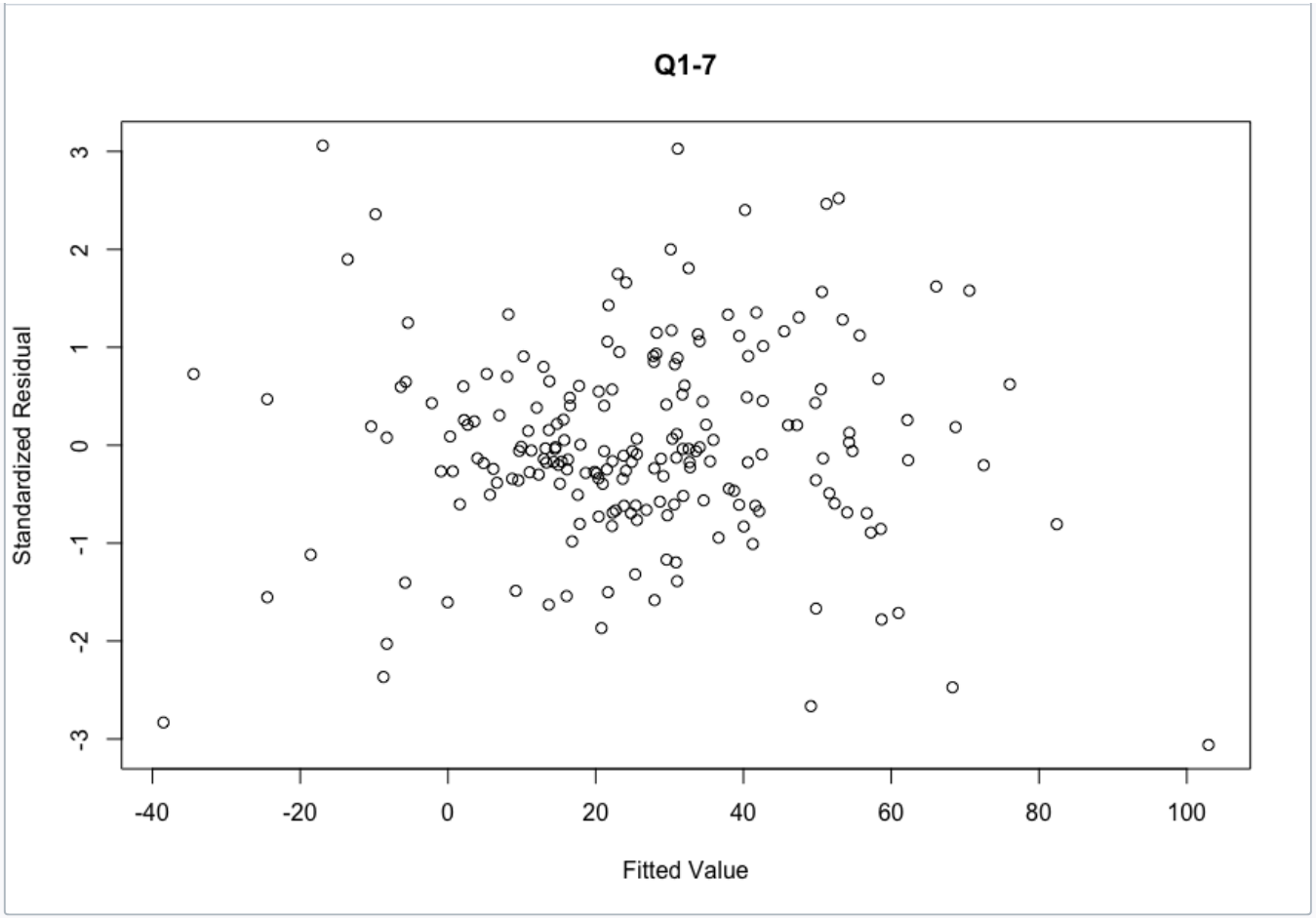
Download ✖ Remove



Choose Files No file chosen

Standardize Residual Plot Q1-7.png

Download ✖ Remove



Choose Files No file chosen

Download Remove

```
1  #Q1-7
2  #Data Preparation
3  library(AER)
4  data("USMacroG")
5  USMacroG=na.omit(USMacroG)
6  Macro=as.data.frame(apply(USMacroG,2,diff))
7  dim(Macro)
8
9  #Regress on gdp, invest, government, dpi, population, interest
10 fit1_7 = lm(Macro$consumption ~ Macro$gdp + Macro$invest + Macro$government
```

```
            + Macro$dpi + Macro$population + Macro$interest)
summary(fit1_7)

#Plot Standardize model
plot(fit1_7$fitted.values, rstandard(fit1_7), xlab = "Fitted Value", ylab =
'Standardized Residual'
     , main = "Q1-7")
abline(fit1_7)

#Calculate the average of leverages
K = length(fit1_7$coefficients)
n = length(Macro$consumption)
avg_of_leverage = K/n
avg_of_leverage

#Find Outlier
plot(fit1_7$fitted.values, rstandard(fit1_7), xlab = "Fitted Value", ylab =
'Standardized Residual'
     , main = "Q1-7 with Outliers")
abline(fit1_7)
i = which(abs(rstandard(fit1_7)) > 3)
points(fit1_7$fitted.values[i], rstandard(fit1_7)[i], col = 'red', cex = 2, pch
= 20)

#Find HLP
plot(fit1_7$fitted.values, rstandard(fit1_7), xlab = "Fitted Value", ylab =
'Standardized Residual'
     , main = "Q1-7 with HLP")
j = which(hatvalues(fit1_7) > 2*avg_of_leverage)
points(fit1_7$fitted.values[j], rstandard(fit1_7)[j], col = 'blue', cex = 2,
pch = 20)

#Find Outliers & HLP
plot(fit1_7$fitted.values, rstandard(fit1_7), xlab = "Fitted Value", ylab =
'Standardized Residual'
     , main = "Q1-7 with Outliers and HLP")
k = which((hatvalues(fit1_7) > 2*avg_of_leverage) & (abs(rstandard(fit1_7)) >
```

Choose Files No file chosen

```
   3))
42 points(fit1_7$fitted.values[k], rstandard(fit1_7)[k], col = 'green', cex = 2,
   pch = 20)
43
44 for (i in 1:10){
45   print(Macro$gdp[i])
46 }
47
48
```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:22 AM**

Q2

3 Points

The dataset Credit in R library ISLR can be used to study how (credit card) Balance depends on Income, (credit) Limit, (credit) Rating, (number of) Cards, Age and (whether the cardholder is a) Student. Convert Student into a quantitative variable so that the value is 0 for Yes and 1 for No.

Choose Files No file chosen

```
library(ISLR)
x=Credit
```

Q2.1

1 Point

Regress Balance on all the above 6 regressors. Write down the equation of the resulted linear regression model (keep one digit after the decimal point for the coefficients). Is Student

significant at 5% significance level? Is the coefficient for Student positive or negative? What does such a coefficient mean?

```
Call:
lm(formula = x$Balance ~ ., data = x)

Residuals:
    Min      1Q  Median      3Q     Max
-170.00  -77.85  -11.84   56.87  313.52

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -68.12425   28.84442  -2.362   0.0187 *
Income        -7.79508    0.23342 -33.395  < 2e-16 ***
Limit          0.19369    0.03238   5.981 4.98e-09 ***
Rating         1.09119    0.48480   2.251   0.0250 *
Cards         18.21190    4.31865   4.217 3.08e-05 ***
Age           -0.62406    0.29182  -2.139   0.0331 *
Student     -425.60994   16.50956 -25.780  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Choose Files No file chosen

```
Residual standard error: 98.61 on 393 degrees of freedom
Multiple R-squared:  0.9547, Adjusted R-squared:  0.954
F-statistic:  1380 on 6 and 393 DF,  p-value: < 2.2e-16
```

-68.1 -7.8 *Income + 0.2*Limit +1.1 *Rating + 18.2 *Cards -0.6 *Age - 425.61 *Student = Balance

---

Yes, since its p value is smaller than 0.05

Its -425 < 0.

Since the student variable is binary, and it is negatively correlated with balance, so as the balance is high, it is highly possible that the customeer is not a student.

CURRENTLY UPLOADED FILES

Q2-1.R

Download Remove

```r
#Q2-1
#Data Preparation
library(ISLR)
x = Credit
x = data.frame(x)
x$Student <- as.character(x$Student)
for (i in 1: length(x$Student)){
  if (isTRUE(x$Student[i] == "No")==TRUE)
    x$Student[i] = 1
  else if (isTRUE(x$Student[i] == "Yes")==TRUE){
    x$Student[i] = 0
  }
}
x$Student <- as.numeric(x$Student)
x = subset(x, select = c(Balance, Income, Limit, Rating, Cards, Age, Student) )

#Regression
fit2_1 = lm(x$Balance ~ ., x)
summary(fit2_1)

#Get coefficient
coef(fit2_1)

```

Choose Files No file chosen

Please select file(s) Select file(s)

Save Answer

Save Answer Last saved on **Nov 11 at 2:22 AM**

Q2.2

1 Point

Add the interaction term Limit:Student into the model in 2.1. Is the interaction significant at 0.05 significance level? Write down the equation of the resulted linear regression model. What does this interaction imply when (credit) Limit goes up by $1000?

```
Call:
lm(formula = x$Balance ~ x$Limit:x$Student + ., data = x)

Residuals:
    Min      1Q  Median      3Q     Max
-168.445  -73.669   -9.437   57.792  304.044

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.980e+02  4.187e+01  -4.730 3.14e-06 ***
Income          -7.836e+00  2.288e-01 -34.246  < 2e-16 ***
Limit            2.278e-01  3.274e-02   6.959 1.45e-11 ***
Rating           1.010e+00  4.752e-01   2.126   0.0341 *
Cards            1.836e+01  4.230e+00   4.342 1.80e-05 ***
Age             -6.293e-01  2.858e-01  -2.202   0.0283 *
Student         -2.802e+02  3.818e+01  -7.340 1.24e-12 ***
x$Limit:x$Student -3.101e-02  7.376e-03  -4.204 3.25e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 96.58 on 392 degrees of freedom
```

Choose Files No file chosen

Multiple R-squared: 0.9566,	Adjusted R-squared: 0.9559
F-statistic: 1236 on 7 and 392 DF, p-value: < 2.2e-16

Yes, since the p value is smaller than 0.05

-198 + Income * -7.8 + Limit * 0.2 + Rating * 1 + Cards * 18.4 + Age * -0.6 + Student *-280.2 + x$Limit:x$Student *-0.03 = Balance

---

Balance will increase by 1000*(0.2-0.03)*Student.

CURRENTLY UPLOADED FILES

Q2-2.R

Download Remove

```r
#Q2-2
#Data Preparation
library(ISLR)
x = Credit
x = data.frame(x)
x$Student <- as.character(x$Student)
for (i in 1: length(x$Student)){
  if (isTRUE(x$Student[i] == "No")==TRUE)
    x$Student[i] = 1
  else if (isTRUE(x$Student[i] == "Yes")==TRUE){
    x$Student[i] = 0
  }
}
x$Student <- as.numeric(x$Student)
x = subset(x, select = c(Balance, Income, Limit, Rating, Cards, Age, Student) )


#Regression
fit2_2 = lm(x$Balance ~ x$Limit:x$Student + ., x)
summary(fit2_2)

```

Choose Files No file chosen

```
#Get coefficient
coef(fit2_2)

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:23 AM**

Q2.3

1 Point

Add the second order term for Income and Limit into the model in 2.2. Are these two second order terms significant at 0.05 significance level? How does the R^2_{adj} change from the model in 2.1 to the model in 2.3?

```
Call:
lm(formula = x$Balance ~ x$Limit:x$Student + I(x$Income^2) +
    I(x$Limit^2) + ., data = x)

Residuals:
```

Choose Files No file chosen

```
     Min       1Q   Median       3Q      Max
-241.764  -48.320    7.922   46.089  194.406

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -6.541e+01  3.333e+01  -1.963 0.050374 .
I(x$Income^2)     -2.471e-02  2.870e-03  -8.608  < 2e-16 ***
I(x$Limit^2)       1.176e-05  6.971e-07  16.871  < 2e-16 ***
Income            -6.144e+00  4.043e-01 -15.197  < 2e-16 ***
Limit              1.631e-01  2.523e-02   6.463 3.07e-10 ***
```

```
Rating           6.551e-01  3.612e-01   1.814 0.070505 .
Cards            1.869e+01  3.199e+00   5.844 1.08e-08 ***
Age             -7.667e-01  2.163e-01  -3.544 0.000441 ***
Student          -1.898e+02  2.935e+01  -6.469 2.95e-10 ***
x$Limit:x$Student -5.353e-02  5.732e-03  -9.340  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 73.02 on 390 degrees of freedom
Multiple R-squared:  0.9753, Adjusted R-squared:  0.9748
F-statistic:  1714 on 9 and 390 DF,  p-value: < 2.2e-16
```

Yes, their p value is smaller than 0.05.

Adjusted R-squared of Q2-1 is 0.954, and the Adjusted R-squared of Q2-3 is 0.9748.

CURRENTLY UPLOADED FILES

Q2-3.R — Download — Remove

```r
#Q2-3
#Data Preparation
library(ISLR)
x = Credit
x = data.frame(x)
x$Student <- as.character(x$Student)
for (i in 1: length(x$Student)){
  if (isTRUE(x$Student[i] == "No")==TRUE)
    x$Student[i] = 1
  else if (isTRUE(x$Student[i] == "Yes")==TRUE){
    x$Student[i] = 0
  }
}
x$Student <- as.numeric(x$Student)
```

```r
x = subset(x, select = c(Balance, Income, Limit, Rating, Cards, Age, Student) )

#Regression
fit2_3 = lm(x$Balance ~ x$Limit:x$Student + I(x$Income^2) + I(x$Limit^2) + ., x)
summary(fit2_3)

```

Please select file(s) Select file(s)

Save Answer Last saved on **Nov 11 at 2:23 AM**

Save All Answers Submit & View Submission