Assignment 2 Homework

Jeffrey Chang

2/3/2020

EXERCISE 2.6

(a) Interpret the estimated slope and intercept. Do the estimates make sense? Why, or why not?

According to the function from EXERCISE 2.6, we can find out that if the temperature is below 30 degrees Fahrenheit, the sales of soda is negative which is impossible. So, the x shouldn't be 0 and below 30. We need to make estimate reasonable.

(b) On a day when the temperature at game time is forecast to be 80 F, predict how many sodas the vendor will sell.

```
## [1] "Y = -240 + 8*80 = 400"
```

(c) Below what temperature are the predicted sales zero?

```
## [1] "If y = 0 = -240+8x, x = 30. so when temperature below 30 and the sales will be zero."
```

(d) Sketch a graph of the estimated regression line.

```r
plot(30, 0, xlim = c(0,100), ylim = c (-200,600),
     main="Estimated Regression Line",xlab="X", ylab="Y")
abline(h = 0, v = 30, col = "gray", lwd = 2, lty = 2)
abline(-240,8, col = "red")
```

EXERCISE 2.7

(a) The estimated error variance q^2 = 2.04672. What is the sum of the squared least squares residuals?

According to the formula from PPT:

```
library(jpeg)
library(ggplot2)
library(reshape)
readImage<-readJPEG('Q2.7a.jpg')
longImage<-melt(readImage)
rgbImage<-reshape(longImage,timevar='X3',idvar=c('X1','X2'),direction='wide')
rgbImage$X1<- -rgbImage$X1
colorColumns<- rgbImage[, substr(colnames(rgbImage), 1, 5)== "value"]
with(rgbImage,plot(X2, X1, col = rgb(colorColumns), asp = 1, pch =".",axes=F,xlab='',ylab=''))
```

There is a simple modification that produces a consistent and unbiased estimator, and that is:

so that:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N-2}$$

$$E\left(\hat{\sigma}^2\right) = \sigma^2$$

```
q2 <- 2.04672
N <- 51
slsr <-q2*(N-2)
print(paste0("The sum of the squared least squares residuals = 2.04672(51-2) = ",
             round(slsr, 4)))
```

```
## [1] "The sum of the squared least squares residuals = 2.04672(51-2) = 100.2893"
```

(b) The estimated variance of b2 is 0.00098. What is the standard error of q^2? What is the sum of value of (Xi-X_bar)^2?

```
varb2 <- 0.00098
sd <- varb2^(1/2)
print(paste0("The standard error of q^2 = var(b2)^(1/2) =  ",
             round(sd, 4)))
```

```
## [1] "The standard error of q^2 = var(b2)^(1/2) =  0.0313"
```

```
# Use the equation Var(b2) = q^2/ (Xi-X_bar)^2
a272 <- q2/varb2
print(paste0("The sum of (Xi-X_bar)^2 = q^2/Var(b2) =  ",
             round(a272, 4)))
```

```
## [1] "The sum of (Xi-X_bar)^2 = q^2/Var(b2) =  2088.4898"
```

(c) Suppose the dependent variable Yi = the state's mean income (in thousands of dollars) of males who are 18 years of age or older and Xi the percentage of males 18 years or older who are high school graduates. If b2 = 0.18 interpret this result.

If b2 = 0.18, the result means that 1% increase of male 18 years or older who are high school graduates would increase $180 of mean income of males who are 18 years of age or older.

(d)Suppose x = 69.139 and y = 15.187, what is the estimate of the intercept parameter?

```
# b1 = (y_bar) - b2*(x_bar)
b2 <- 0.18
b1 <-round(c(15.187-b2*69.139), 4)
print( paste0("The estimate of the intercept parameter = ", b1))
```

```
## [1] "The estimate of the intercept parameter = 2.742"
```

(e) Given the resultin (b) and (d), what is Sum of (Xi)^2

```
# the function sum of (Xi-X_bar)^2 = Sum of (Xi)^2 - N*(X_bar)^2
print( paste0("Sum of (Xi)^2 = ", round(c(a272+N*69.139^2),4)))
```

```
## [1] "Sum of (Xi)^2 = 245878.7572"
```

(f)For the state of Arkansas, the value of Yi = 12.274 and the value of Xi = 58.3. Compute the least squares residual for Arkansas. (Hint: Use the information in parts (c) and (d).).

```
# e_i = Yi-Yi_hat = Yi-b1-b2*Xi
yi <- 12.274
xi <- 58.3
ei <- yi-b1-b2*xi
print( paste0("The least squares residual for Arkansas = ", round(ei,4)))
```

```
## [1] "The least squares residual for Arkansas = -0.962"
```

EXERCISE 2.10

(a)Explain why the econometric model above is a simple regression model like those discussed in this chapter.

Set Y=(Rj-Rf) and X=(Rm-Rf), so the mode will become Y=Aj+Bj*X+e, and this is a simple regression model.

(b) In the data file capm4.dat are data on the monthly returns of six firms (Microsoft, GE, GM, IBM, Disney, and Mobil-Exxon), the rate of return on the market portfolio (MKT), and the rate of return on the risk free asset (RISKFREE). The 132 observations cover January 1998 to December 2008. Estimate the CAPM model for each firm, and comment on their estimated beta values. Which firm appears most aggressive? Which firm appears most defensive?

```
library(foreign)
capm <- read.dta("capm.dta") # input Stata file

x <- capm[,c("mkt")]-capm[,c("riskfree")]
x <- data.frame(x)
colnames(x) <- c('x')
capm <- cbind(capm,x)

ydisrf <- data.frame(capm[,c("dis")]-capm[,c("riskfree")])
colnames(ydisrf) <- c('ydisrf')
capm <- cbind(capm,ydisrf)

ygerf <- data.frame(capm[,c("ge")]-capm[,c("riskfree")])
colnames(ygerf) <- c('ygerf')
capm <- cbind(capm,ygerf)

ygmrf <- data.frame(capm[,c("gm")]-capm[,c("riskfree")])
colnames(ygmrf) <- c('ygmrf')
capm <- cbind(capm,ygmrf)

yibmrf <- data.frame(capm[,c("ibm")]-capm[,c("riskfree")])
colnames(yibmrf) <- c('yibmrf')
capm <- cbind(capm,yibmrf)

ymsftrf <- data.frame(capm[,c("msft")]-capm[,c("riskfree")])
colnames(ymsftrf) <- c('ymsftrf')
capm <- cbind(capm,ymsftrf)

yxomrf <- data.frame(capm[,c("xom")]-capm[,c("riskfree")])
colnames(yxomrf) <- c('yxomrf')
capm <- cbind(capm,yxomrf)

capm <- capm[,c("date", "dis","ydisrf","ge","ygerf","gm","ygmrf","ibm","yibmrf","msft","ymsftrf","xom",

library(PoEdata)
moddis <- lm( ydisrf ~ x, data = capm)
b1dis <- round(coef(moddis)[[1]],4)
b2dis <- round(coef(moddis)[[2]],4)

modge <- lm( ygerf ~ x, data = capm)
b1ge <- round(coef(modge)[[1]],4)
```

```
b2ge <- round(coef(modge)[[2]],4)

modgm <- lm( ygmrf ~ x, data = capm)
b1gm <- round(coef(modgm)[[1]],4)
b2gm <- round(coef(modgm)[[2]],4)

modibm <- lm( yibmrf ~ x, data = capm)
b1ibm <- round(coef(modibm)[[1]],4)
b2ibm <- round( coef(modibm)[[2]],4)

modmsft <- lm( ymsftrf ~ x, data = capm)
b1msft <- round(coef(modmsft)[[1]],4)
b2msft <- round(coef(modmsft)[[2]],4)

modxom <- lm( yxomrf ~ x, data = capm)
b1xom <- round(coef(modxom)[[1]],4)
b2xom <- round(coef(modxom)[[2]],4)

betavalues<-data.frame(Disney=c(b2dis),GE=c(b2ge), GM=c(b2gm), IBM=c(b2ibm),Mircrosoft=c(b2msft),ExxonM
betavalues
```

```
##    Disney     GE     GM    IBM Mircrosoft ExxonMobil
## b2 0.8978 0.8993 1.2614 1.1882     1.3189      0.414
```

1. Microsoft, GM and IBM are aggressive, because beta values > 1.
2. Disney, GE, Exxon. are defensive, because their beta values < 1.
3. The most aggressive beta value is Microsoft's stock $= 1.3189$
4. the most defensive beta value is Exxon-Mobil's stock $= 0.414$

(c)Finance theory says that the intercept parameter Aj should be zero. Does this seem correct given your estimates? For the Microsoft stock, plot the fitted regression line along with the data scatter.

```
Aj<-data.frame(Disney=c(b1dis),GE=c(b1ge), GM=c(b1gm), IBM=c(b1ibm),Mircrosoft=c(b1msft),ExxonMobil=c(b
Aj
```

```
##     Disney      GE      GM    IBM Mircrosoft ExxonMobil
## Aj -0.0011 -0.0012 -0.0116 0.0059     0.0061     0.0079
```

```
plot(capm$x, capm$ymsftrf,
xlab="MKT-RISKFREE", ylab="MSFT-RISKFREE", main = "Scatter plot of Microsoft and market rate")
abline(b1msft,b2msft, col = "red")
```

(d)Estimate the model for each firm under the assumption that Aj = 0. Do the estimates of the beta values change much?

```
b10b2dis<- glm((capm$dis-capm$riskfree)~0 + capm$x)
b10b2ge <-glm((capm$ge-capm$riskfree)~0 + capm$x)
b10b2gm <-glm((capm$gm-capm$riskfree)~0 + capm$x)
b10b2ibm <-glm((capm$ibm-capm$riskfree)~0 + capm$x)
b10b2msft <-glm((capm$msft-capm$riskfree)~0 + capm$x)
b10b2xom <-glm((capm$xom-capm$riskfree)~0 + capm$x)

betavalues_2<-data.frame(Disney=c(b10b2dis$coefficients),GE=c(b10b2ge$coefficients), GM=c(b10b2gm$coeff
betavalues_2
```

```
##              Disney        GE       GM      IBM Mircrosoft ExxonMobil
## b2(b1=0) 0.8979193 0.8993424 1.262227 1.187795   1.318516  0.4134124
```

According to the result with aj=0, the estimate of bj changed a little, but the nature of aggressive or defensive is same.

EXERCISE 2.14

(a) Using the data for 1916–2008, plot a scatter diagram of VOTE against GROWTH. Does there appear to be positive association?

```
library(foreign)
fair4 <- read.dta("fair4.dta") # input Stata file
fair4_1<-fair4[-1:-9,]
plot(fair4_1$growth, fair4_1$vote, xlab = "Growth", ylab = "Vote", main = "Scatter Diagram of VOTE agai
```

```
library(PoEdata)
mod1 <- lm(fair4_1$vote~ fair4_1$growth, data = fair4_1)
mod1b1 <- round(coef(mod1)[[1]],4)
mod1b2 <- round(coef(mod1)[[2]],4)
```

Yes, It is positive association.

(b) Estimate the regression VOTE = b_1 + b_2*(GROWTH) + e by least squares using the data from 1916 to 2008. Report and discuss the estimation result. Sketch, by hand, the fitted line on the data scatter from (a).

```
readImage<-readJPEG('Q2.14b.jpg')
longImage<-melt(readImage)
rgbImage<-reshape(longImage,timevar='X3',idvar=c('X1','X2'),direction='wide')
rgbImage$X1<- -rgbImage$X1
colorColumns<- rgbImage[, substr(colnames(rgbImage), 1, 5)== "value"]
with(rgbImage,plot(X2, X1, col = rgb(colorColumns), asp = 1, pch =".",axes=F,xlab='',ylab=''))
```

(c)Fit the regression in (b) using the data from 1916 to 2004. Predict the VOTE share for the incumbent party based on the actual 2008 value for GROWTH. How does the predicted vote for 2008 compare to the actual result?

```
fair4_2<-fair4_1[-24,]
mod2 <- lm(fair4_2$vote~ fair4_2$growth, data = fair4_2)
mod2b1 <- round(coef(mod2)[[1]],4)
mod2b2 <- round(coef(mod2)[[2]],4)
print(paste0("Vote= ", mod2b1,"+",mod2b2, "*Growth"))
```

```
## [1] "Vote= 51.0533+0.878*Growth"
```

According the fair4.dta, the actual 2008 value for growth is 0.22. Putting 0.22 into the estimated equation.

```
vote2008 <-round((mod2b1+mod2b2*0.22),4)
vote2008
```

```
## [1] 51.2465
```

The actual vote in 2008 is 46.6, less than 51.2465. so the incumbent party did not obtain the majority of the vote in 2008.

(d) Economywide inflation may spell doom for the incumbent party in an election. The variable INFLATION is the growth in prices over the first 15 quarters of an administration. Using the data from 1916 to 2008, plot VOTE against INFLATION. Using the same sample, report and discuss the estimation results for the model VOTE = a_1 + a_2*INFLATION + e.

```
plot(fair4_1$inflation, fair4_1$vote, xlab = "Inflation", ylab = "Vote", main = "Scatter Diagram of VOTE
mod3 <- lm(fair4_1$vote~ fair4_1$inflation, data = fair4_1)
mod3b1 <- round(coef(mod3)[[1]],4)
mod3b2 <- round(coef(mod3)[[2]],4)
abline(mod3b1,mod3b2, col = "red")
```

A negative association between variables Vote and Inflation.

```
## [1] "The estimated equation is: Vote= 53.4077-0.4443*Inflation"
```

According to the dataset, 1% point increase in "INFLATION" during the incumbent party's first 15 quarters reduces the share of the incumbent party's vote by 0.444% points.

The estimated intercept shows that when inflation is 0% for that party's first 15 quarters, the expected share of votes won by the incumbent party is 53.4%; the incumbent party is predicted to maintain the majority vote when "INFLATION" is 0%, during its first 15 quarters.