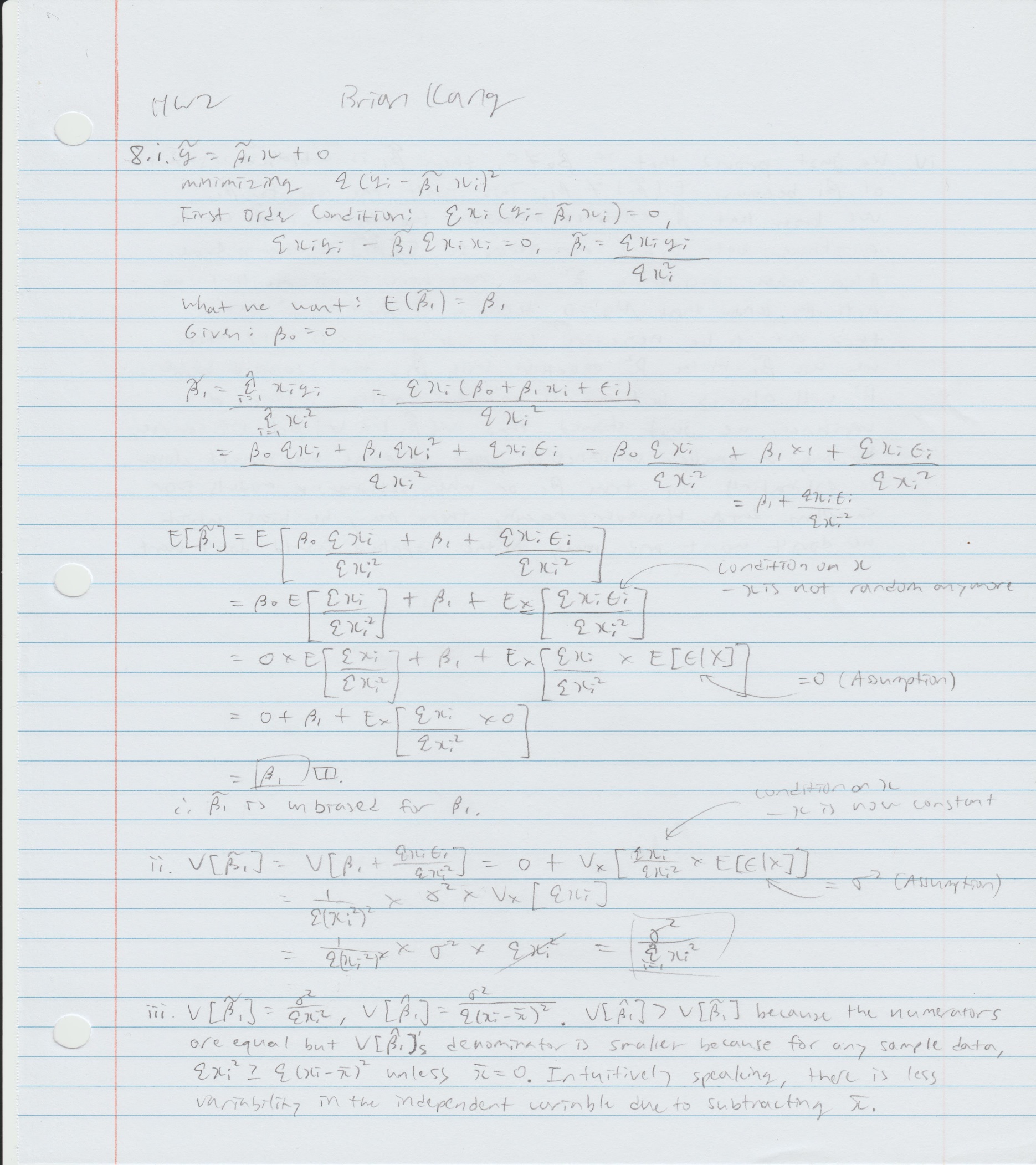
hw02

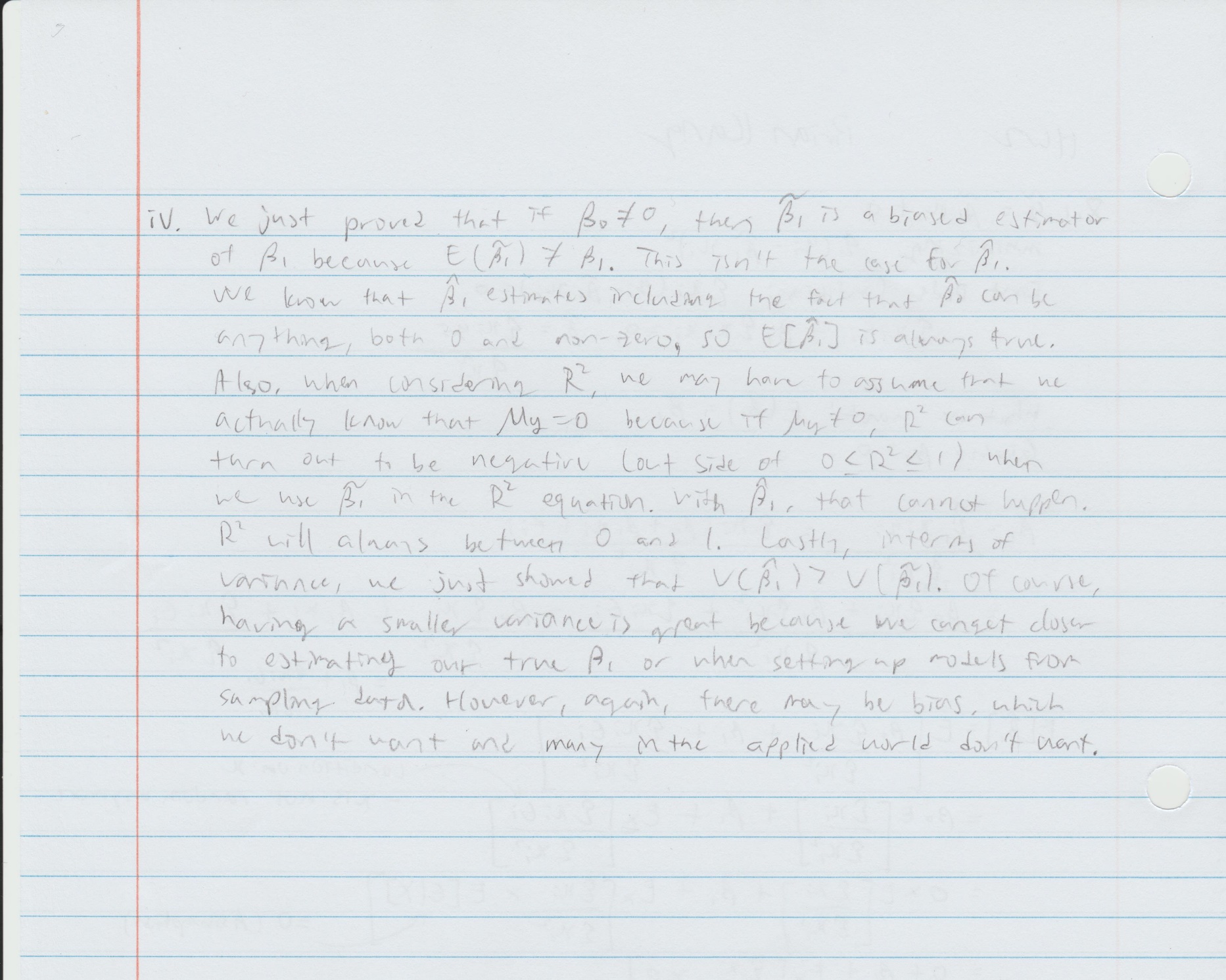
brian

January 27, 2019

THIS IS THE NUMERICAL PROBLEMS

(GO TO NEXT PAGE)





THIS IS THE COMPUTATIONAL EXERCISES

## R Markdown

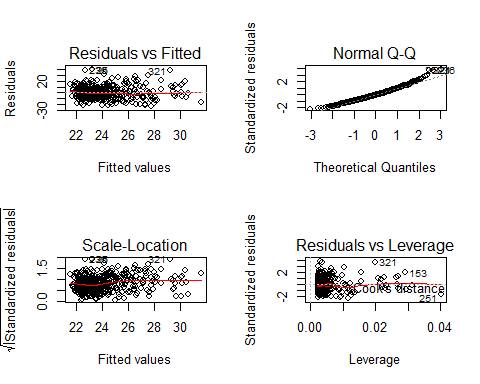
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

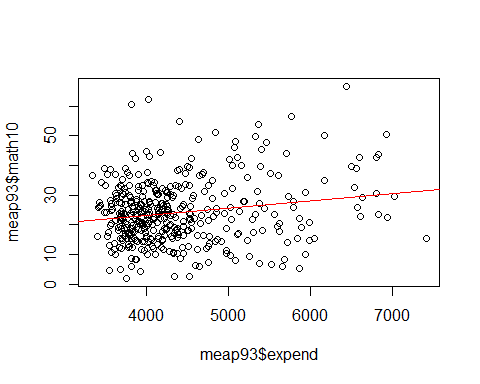
#install.packages("wooldridge")  
library(wooldridge)

## Warning: package 'wooldridge' was built under R version 3.5.2

data("meap93")  
  
# CE6)  
# i)  
lm.1 <- lm(math10 ~ expend, data = meap93)  
# print general data  
par(mfrow = c(2,2)) # disp all graphs at once  
plot(lm.1)



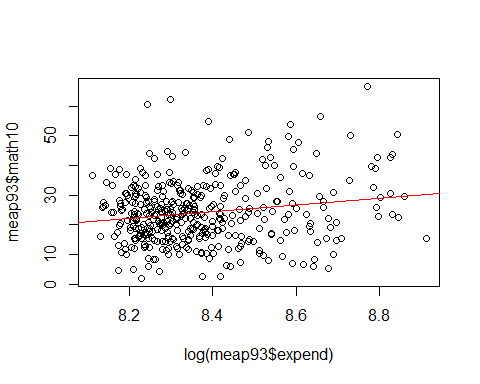
par(mfrow = c(1,1)) # back to normal  
plot(meap93$expend, meap93$math10)  
abline(lm.1, col = "red")



summary(lm.1)

##   
## Call:  
## lm(formula = math10 ~ expend, data = meap93)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.579 -7.175 -0.874 6.299 39.174   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.336e+01 2.934e+00 4.553 7e-06 \*\*\*  
## expend 2.456e-03 6.601e-04 3.720 0.000227 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.33 on 406 degrees of freedom  
## Multiple R-squared: 0.03296, Adjusted R-squared: 0.03058   
## F-statistic: 13.84 on 1 and 406 DF, p-value: 0.0002273

# If I had to choose one, I would say that there is some linear  
# relation between each additional dollar spent and the pass  
# rate, but honestly it is hard to say. From the scatterplot we  
# can see a wide spread in the dots (almost random association)  
# and a cluster on the lower left hand corner. Due to this  
# characteristic our R-squared value and our estimated slope are  
# near zero. But, we can still say that there is a reasonably  
# linear and positive relationship.  
# Just to supplement, looking at the qq-plot of our linear model,  
# we can see that our quantiles and residuals generally follow  
# our slope-intercept life, supporting the fact of linearity.  
# But at the higher end, we see a gradual increasing and parting  
# from the line, implying a change in behavior at the high end.  
  
# ii)  
lm.2 <- lm(math10 ~ log(expend), data = meap93)  
plot(log(meap93$expend), meap93$math10)  
abline(lm.2, col = "red")



summary(lm.2)

##   
## Call:  
## lm(formula = math10 ~ log(expend), data = meap93)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.343 -7.100 -0.914 6.148 39.093   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -69.341 26.530 -2.614 0.009290 \*\*   
## log(expend) 11.164 3.169 3.523 0.000475 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.35 on 406 degrees of freedom  
## Multiple R-squared: 0.02966, Adjusted R-squared: 0.02727   
## F-statistic: 12.41 on 1 and 406 DF, p-value: 0.0004752

# We have the regression slope value 11.164, meaning that with  
# each 1 percentage point increase in log(expend) we estimate  
# that math10 will increase by 11.164 percentage points.  
# In a linear model in this form we know (Appendix A) that  
# 100\*change in log(x) is approximately the %change in x.  
# And the UNIT change in y is regression slope/100\*100\*change in log(x)  
# which is about regression slope/100\* %change in x.  
# Thus, a 10%increase in expend is: regression slope/100\*10  
# = regression slope/10 increase in units of y,   
# which is percentage point change.  
  
# iii)  
# We estimate that with each 1 percentage point increase in   
# log(expend) we estimate that math10 will increase by 11.164   
# percentage points. Our estimated intercept is -69.341.  
# The sample size is 408, and our R-squared value is 0.02966,  
# almost zero.  
  
# iv)  
# Answer is in (ii) as well. If spending increases by 10%, the  
# estimated percentage point increase in math10 is 1.1164.  
  
# v)  
# Our math10 data corresponds to the math pass rate of 10th graders.  
# It is unlikely to ger an estimated pass rate that is over  
# 100%, which means everyone in the highschool passes math.  
# I don't even know what "over 100% of kids passed math" would mean.  
  
# CE9)  
data("countymurders")  
# all data in rows with year=1996  
dat96 <- countymurders[(countymurders$year==1996),]  
# i)  
# number of counties with zero murders in 1996  
nrow(dat96[dat96$murders>=0,])

## [1] 2197

# number of counties with at least one execution  
nrow(dat96[dat96$execs>=1,])

## [1] 31

# largest number of executions  
max(dat96$execs)

## [1] 3

# ii)  
lm.3 <- lm(murders ~ execs, data = dat96)  
summary(lm.3)

##   
## Call:  
## lm(formula = murders ~ execs, data = dat96)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -149.12 -5.46 -4.46 -2.46 1338.99   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.4572 0.8348 6.537 7.79e-11 \*\*\*  
## execs 58.5555 5.8333 10.038 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 38.89 on 2195 degrees of freedom  
## Multiple R-squared: 0.04389, Adjusted R-squared: 0.04346   
## F-statistic: 100.8 on 1 and 2195 DF, p-value: < 2.2e-16

# We estimate that with each increase in executions by 1, the   
# number of murders in a county will increase by 58.5555 and  
# where there is no executions extimate 5.4572 murders. The   
# sample size is 2197 counties and the R-squared value is  
# 0.04389, very close to zero.  
  
# iii)  
# Interpretation of the slope is in (ii) as well.   
# We estimate that with each increase in executions by 1, the   
# number of murders in a county will increase by 58.5555.  
# I think this estimation does not suggest a deterrent effect  
# of executions. We showed that almost half (1051/2197) counties  
# had no murders at all. In fact, there is only 30 counties  
# had both murder and executions together. This should not  
# should not determine the trend of all counties. Also, there  
# is one outlier with almost 1400 murders but 1 execution.  
  
# iv)  
# Interpretation of the intercept is in (ii) as well.  
# In counties where there is no executions, we extimate   
# 5.4572 murders.  
# The residual for a county with 0 murder and executions is  
# the intercept of the estimated model, 5.4572.  
  
# v)  
# Does execution cause murder? Not likely. The question itself  
# is weird and we know there are many underlying components  
# that increase both murders and potentially the number of  
# executions, meaning there is probable bias. Also, from our  
# data we can see that the R-squared value is near zero, which  
# indicates that there is no strong correlation. (Note, if we  
# get rid of the outliers, data will be more clear and value   
# of R-squared will probably increase a lot.)