homework three

Yi (Chris) Chen

October 8, 2017

## homework three

### problem 2.23

# problem a  
setwd("C:/Users/cheny/Desktop/study/linear regression model/homework/homework three")  
data\_23 <- read.table('19.txt',header = FALSE,col.names=c('GPA','ATC'))  
reg\_23 <- lm(data\_23$GPA ~ data\_23$ATC)  
anova\_23 <- anova(reg\_23)  
anova\_23

## Analysis of Variance Table  
##   
## Response: data\_23$GPA  
## Df Sum Sq Mean Sq F value Pr(>F)   
## data\_23$ATC 1 3.588 3.5878 9.2402 0.002917 \*\*  
## Residuals 118 45.818 0.3883   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# problem c  
f\_value <- qf(0.99,1,118)  
anova\_23$`F value` > f\_value

## [1] TRUE NA

t\_value <- qt(1-0.01/2,118)  
t\_value\_star <- 3.040  
t\_value\_star > t\_value

## [1] TRUE

cat('conclude Ho')

## conclude Ho

# problem d   
sse <- 45.818  
ssr <- 3.588  
ssto <- 3.588 + 45.818  
r2 <- ssr/ssto  
cat('the relative reduction is', r2)

## the relative reduction is 0.07262276

# problem e  
r <- sqrt(r2)  
cat('the r is',r)

## the r is 0.2694861

### problem 2.26

# problem a  
data\_26 <- read.table('1.22.txt',header = FALSE, col.names = c('hardness','time'))  
reg\_26 <- lm(data\_26$hardness ~ data\_26$time)  
anova\_26 <- anova(reg\_26)  
anova\_26

## Analysis of Variance Table  
##   
## Response: data\_26$hardness  
## Df Sum Sq Mean Sq F value Pr(>F)   
## data\_26$time 1 5297.5 5297.5 506.51 2.159e-12 \*\*\*  
## Residuals 14 146.4 10.5   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# problem b  
f\_value <- qf(0.99,1,13)  
anova\_26$`F value` > f\_value

## [1] TRUE NA

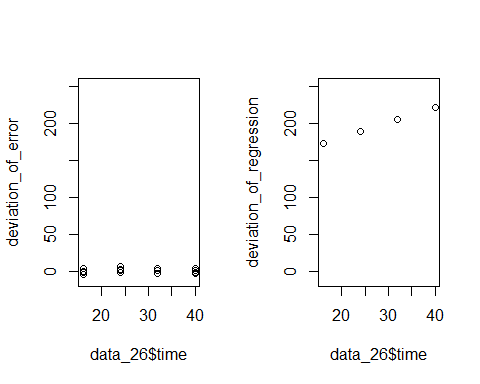
t\_value <- qt(1-0.01/2,13)  
t\_value\_star <- 22.51  
t\_value\_star > t\_value

## [1] TRUE

cat('conclude Ho')

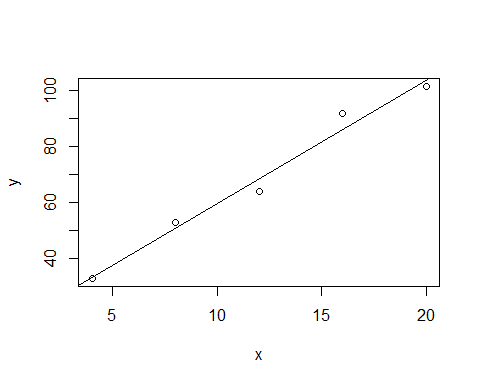
## conclude Ho

# problem c  
y\_hat <- data\_26$time \* reg\_26$coefficients[2] + reg\_26$coefficients[1]  
deviation\_of\_error <- data\_26$hardness - y\_hat  
y\_mean <- mean(data\_26$time)  
deviation\_of\_regression <- y\_hat - y\_mean  
  
par(mfrow = c(1,2))  
plot(deviation\_of\_error~data\_26$time,ylim=c(-10,250))  
plot(deviation\_of\_regression~data\_26$time,ylim=c(-10,250))



### problem 2.66

# problem a  
set.seed(1)  
residules <- rnorm(n = 5, mean = 0, sd = 5)  
x <- c(4,8,12,16,20)  
y <- x \* 4 + 20 + residules  
reg\_66 <- lm(y~x)  
plot(y~x)  
abline(reg\_66)



y\_h <- reg\_66$coefficients[1] + 10 \* reg\_66$coefficients[2]  
cat('y\_h is', y\_h)

## y\_h is 59.81546

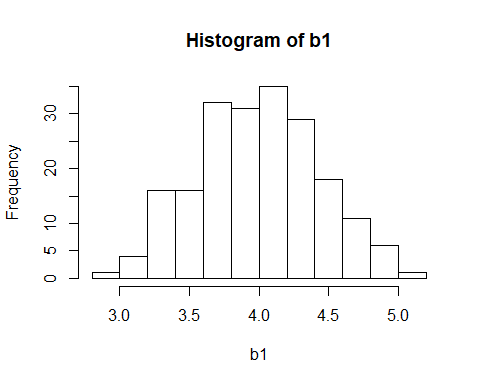
confidenc\_interval\_up <- y\_h + qt(1-0.05/2,5-2)\*sqrt(sum(reg\_66$residuals^2)\*(1/5+(10-mean(x))/sum((x-mean(x))^2)))  
confidenc\_interval\_down <- y\_h - qt(1-0.05/2,5-2)\*sqrt(sum(reg\_66$residuals^2)\*(1/5+(10-mean(x))/sum((x-mean(x))^2)))  
  
# problem b and c  
b1 <- vector()  
for(i in 1:200){  
 residules <- rnorm(n = 5, mean = 0, sd = 5)  
 x <- c(4,8,12,16,20)  
 y <- x \* 4 + 20 + residules  
 reg\_66 <- lm(y~x)  
 b1 <- c(b1,reg\_66$coefficients[2])  
}  
mean(b1)

## [1] 3.999999

sd(b1)

## [1] 0.4397091

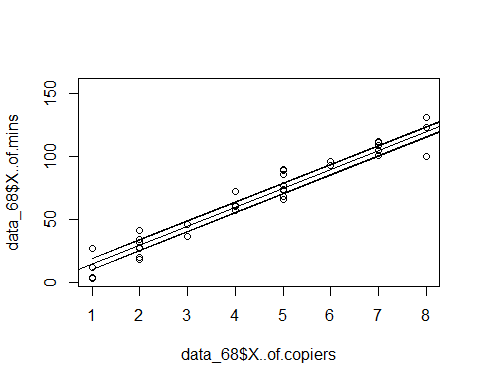
hist(b1)



#problem d  
y\_true <- 4 \* 10 + 20  
times <- vector()  
for(i in 1:200){  
 residules <- rnorm(n = 5, mean = 0, sd = 5)  
 x <- c(4,8,12,16,20)  
 y <- x \* 4 + 20 + residules  
 reg\_66 <- lm(y ~ x)  
 y\_h <- reg\_66$coefficients[1] + 10 \* reg\_66$coefficients[2]  
 confidenc\_interval\_up <- y\_h + qt(1-0.05/2,5-2)\*sqrt(sum(reg\_66$residuals^2)\*(1/5+(10-mean(x))/sum((x-mean(x))^2)))  
 confidenc\_interval\_down <- y\_h - qt(1-0.05/2,5-2)\*sqrt(sum(reg\_66$residuals^2)\*(1/5+(10-mean(x))/sum((x-mean(x))^2)))  
  
 if( y\_true <= confidenc\_interval\_up & y\_true >= confidenc\_interval\_down){  
 times <- c(times,1)  
 }else{  
 times <- c(times,0)  
 }  
}  
propotion <- sum(times)/length(times)

### problem 2.68

data\_68 <- read.table('1.20.txt',header = FALSE, col.names = c('# of mins','# of copiers'))  
reg\_68 <- lm(data\_68$X..of.mins ~ data\_68$X..of.copiers)  
y\_h <- data\_68$X..of.copiers \* reg\_68$coefficients[2] + reg\_68$coefficients[1]  
up\_band <- y\_h + sqrt(2\*qf(0.9,2,45-2))\*sqrt(sum(reg\_68$residuals^2)/43)\*sqrt(1/45 + (data\_68$X..of.copiers-mean(data\_68$X..of.copiers)^2)/(sum(data\_68$X..of.copiers-mean(data\_68$X..of.copiers)^2)))  
down\_band <- y\_h - sqrt(2\*qf(0.9,2,45-2))\*sqrt(sum(reg\_68$residuals^2)/43)\*sqrt(1/45 + (data\_68$X..of.copiers-mean(data\_68$X..of.copiers)^2)/(sum(data\_68$X..of.copiers-mean(data\_68$X..of.copiers)^2)))  
  
plot(data\_68$X..of.mins~data\_68$X..of.copiers,xlim=c(1,8))  
abline(reg\_68)  
points(up\_band~data\_68$X..of.copiers,type='l')  
points(down\_band~data\_68$X..of.copiers,type='l')  
  
# or it can be solved in this way  
library(ggplot2)



result <- ggplot(data\_68,aes(x=data\_68$X..of.copiers,y=data\_68$X..of.mins))+geom\_point() + stat\_smooth(method = lm)  
plot(result)

