homework three

Yi (Chris) Chen

October 8, 2017

homework three

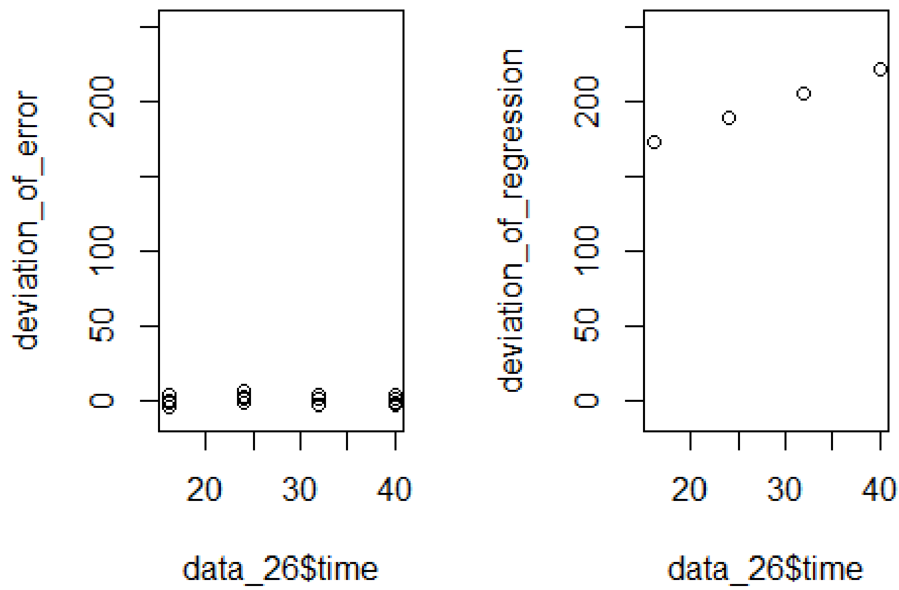
problem 2.23

```
# 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
```

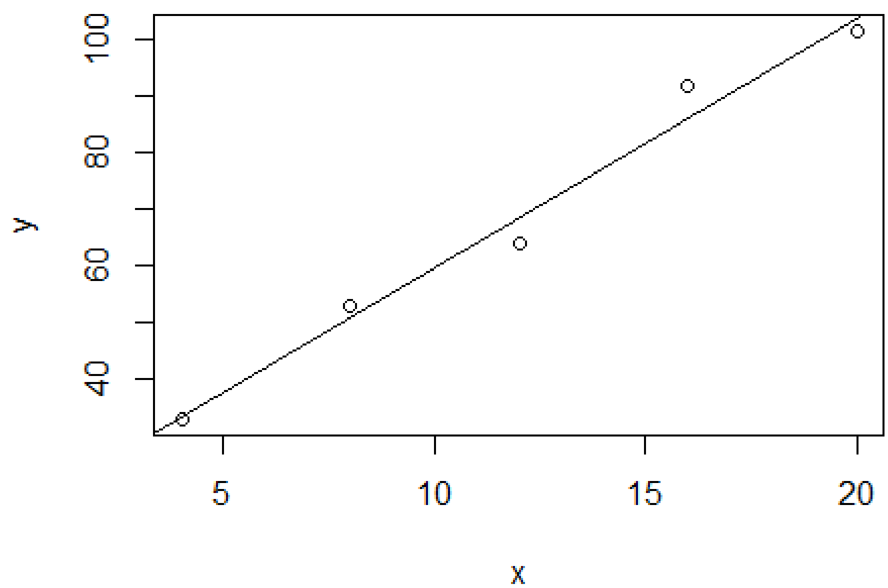
problem 2.26

```
# 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))
```

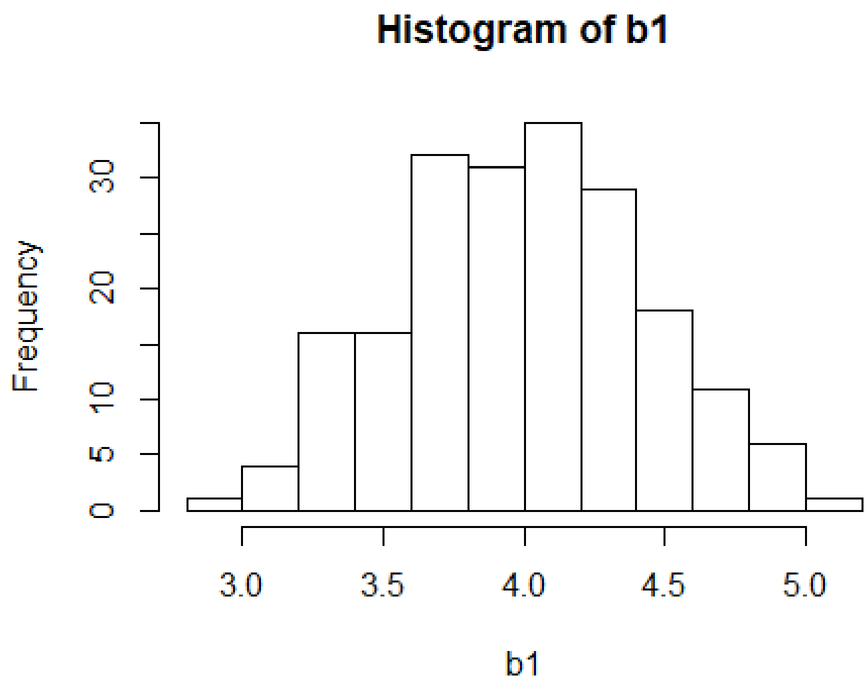


problem 2.66

```
# 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)
```

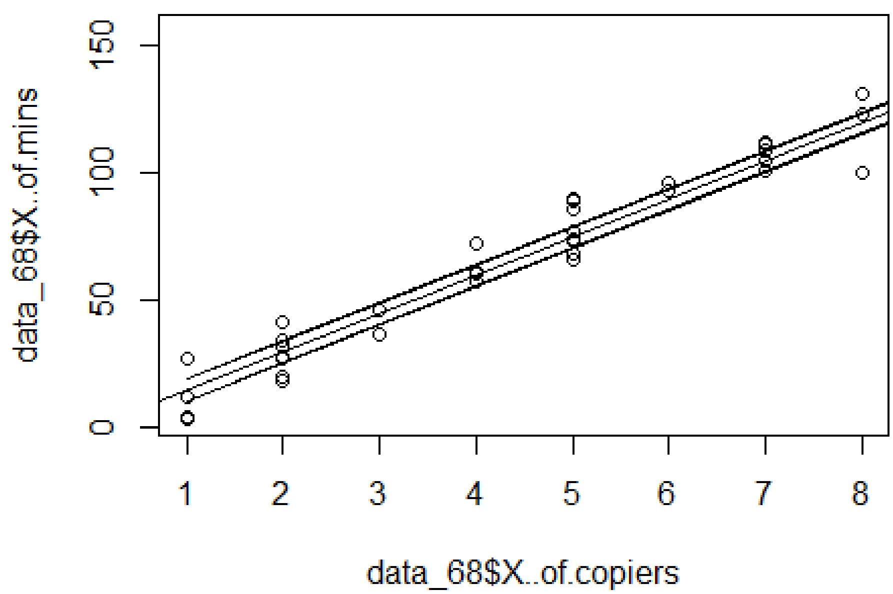



```
#problem d
y_true <- 4 * 10 + 20
times <- vector()
for(i in 1:200){
        residules <- rnorm(n = 5, mean = 0, sd = 5)
        x <- c(4,8,12,16,20)
        y <- x * 4 + 20 + residules
        reg_66 <- lm(y ~ x)
        y_h <- reg_66$coefficients[1] + 10 * reg_66$coefficients[2]
        confidenc_interval_up <- 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)
```

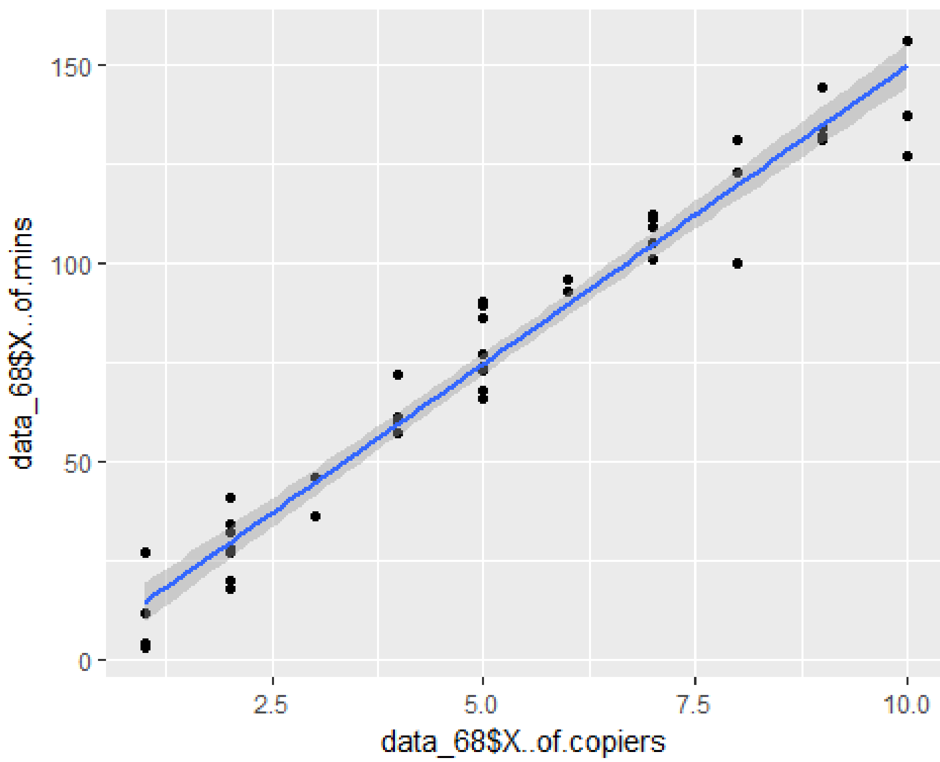
problem 2.68

```
data_68 <- read.table('1.20.txt',header = FALSE, col.names = c('# of mins','#
of copiers'))
```

```r
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)
```




```r
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)
```



data_68$X..of.mins
data_68$X..of.copiers
0
50
100
150
2.5
5.0
7.5
10.0