GR5205_HW4

Jennifer Lieu

10/29/2018

Problem 1.

```
setwd("/Users/jenniferlieu/Desktop/Linear Regression Models Data 1")
hw4 <- read.table("HW4Problem1.txt", header=T, sep='')
```

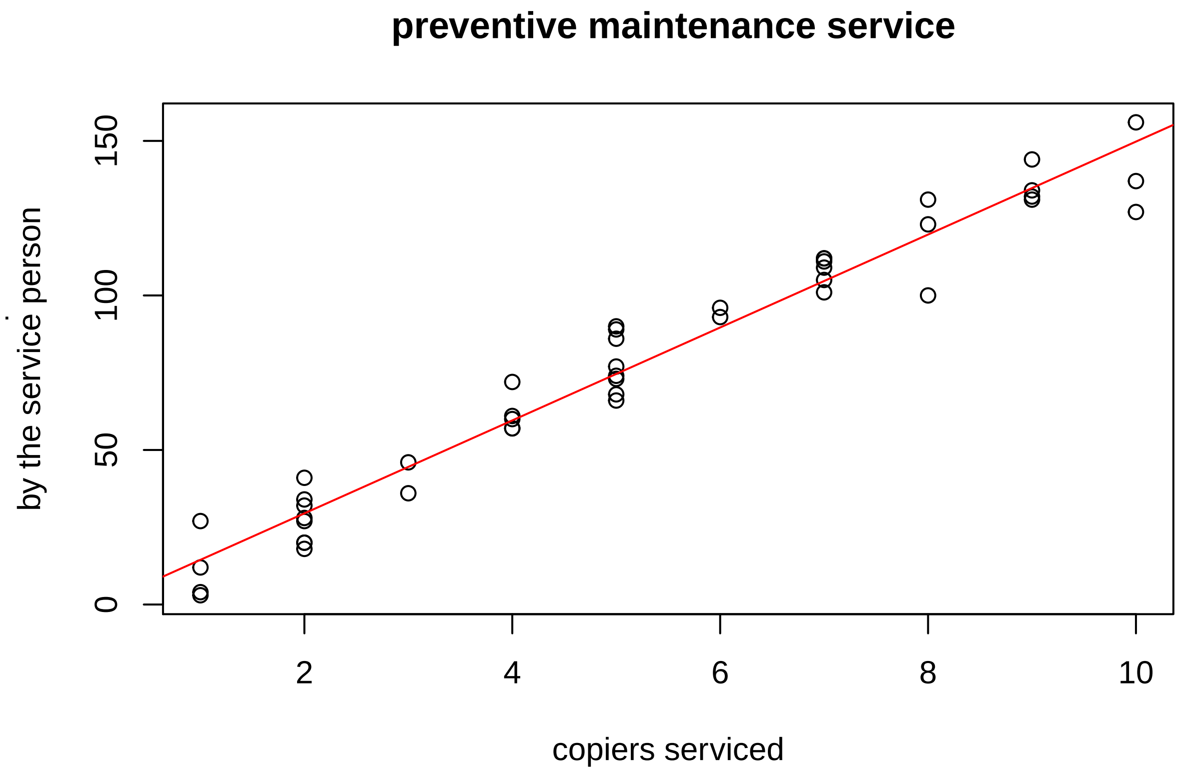
i.

```
copierslm <- lm(hw4$y~hw4$x)
summary(copierslm)
```

```
##
## Call:
## lm(formula = hw4$y ~ hw4$x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## hw4$x        15.0352     0.4831  31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

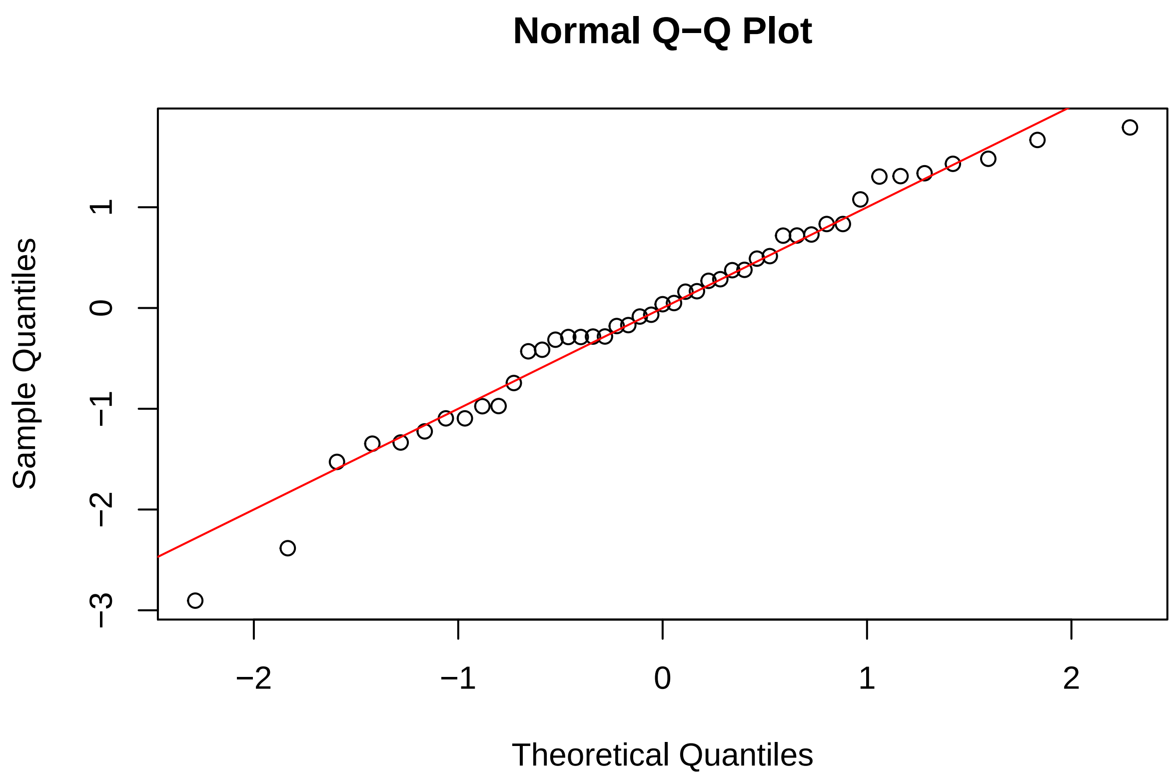
ii.

```
# scatter plot + linear model
plot(hw4$x, hw4$y, main = " preventive maintenance service", xlab= "copiers serviced", ylab= "minutes sp
by the service person")
abline(-0.5802 ,15.0352, col="red")
```

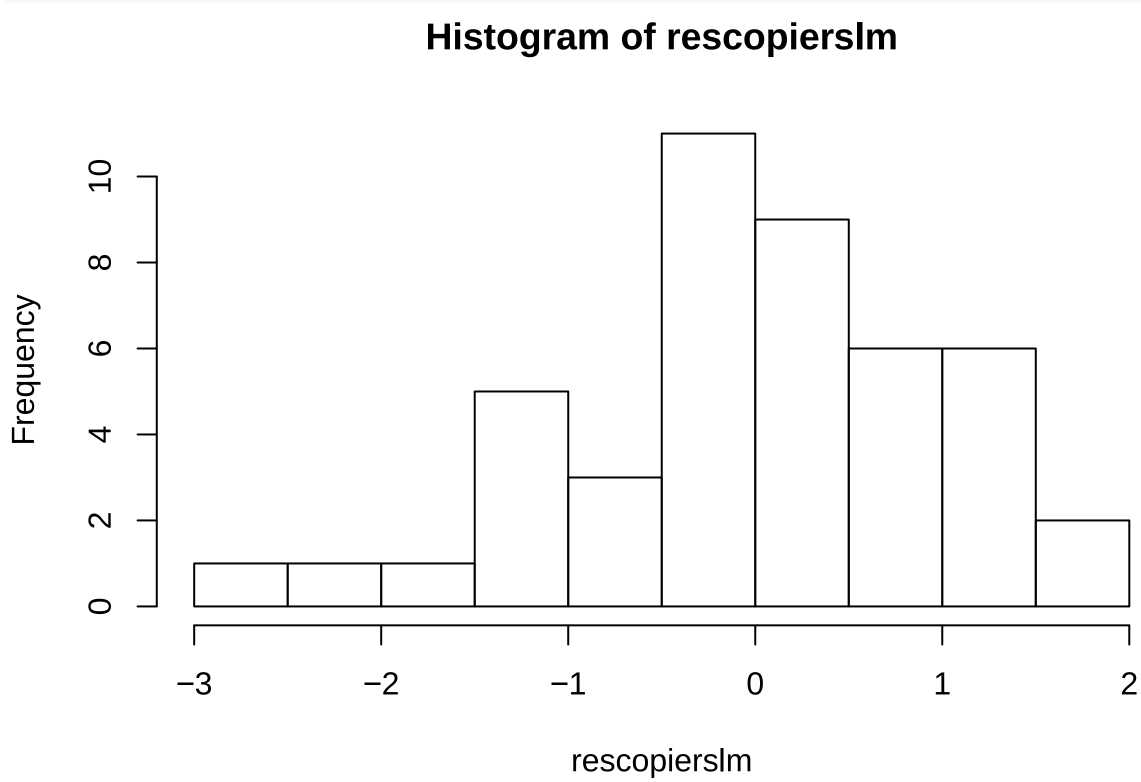



```r
# deleted residuals
rescopierslm <- rstudent(copierslm)

# QQ plot of deleted residuals
qqnorm(rescopierslm)
abline(0,1, col="red")
```

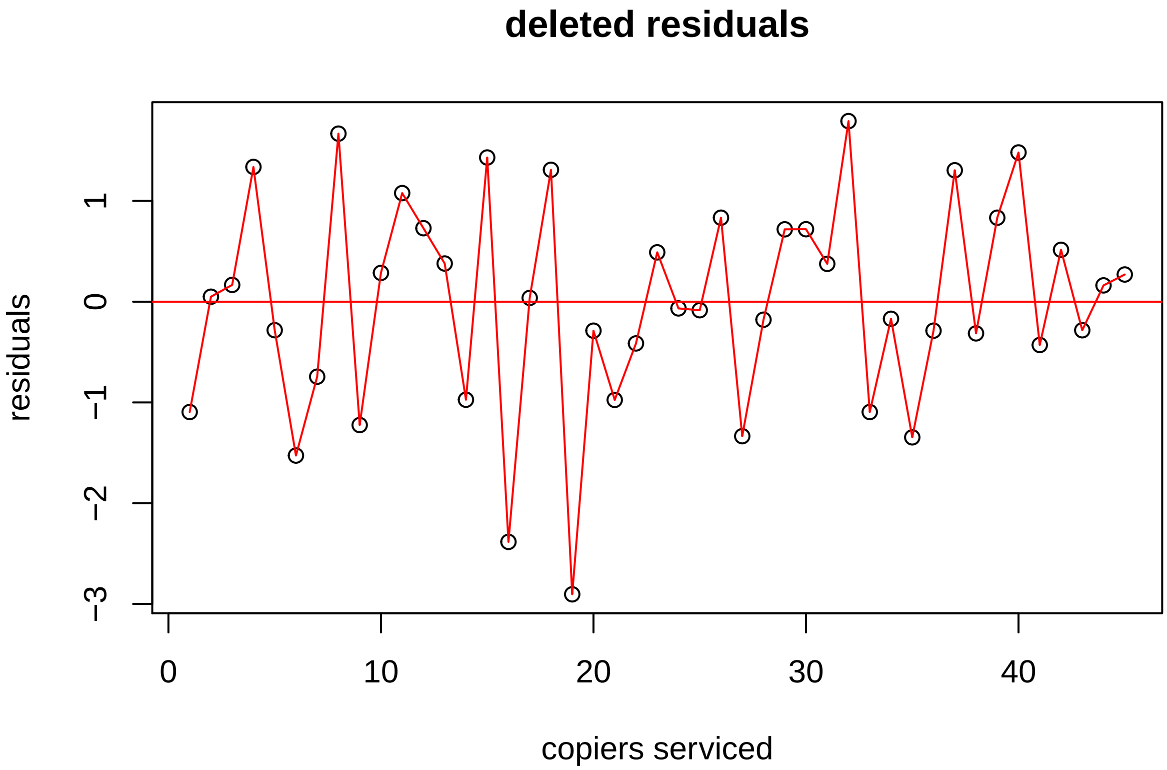



```
# histogram of deleted residuals
hist(rescopierslm)
```

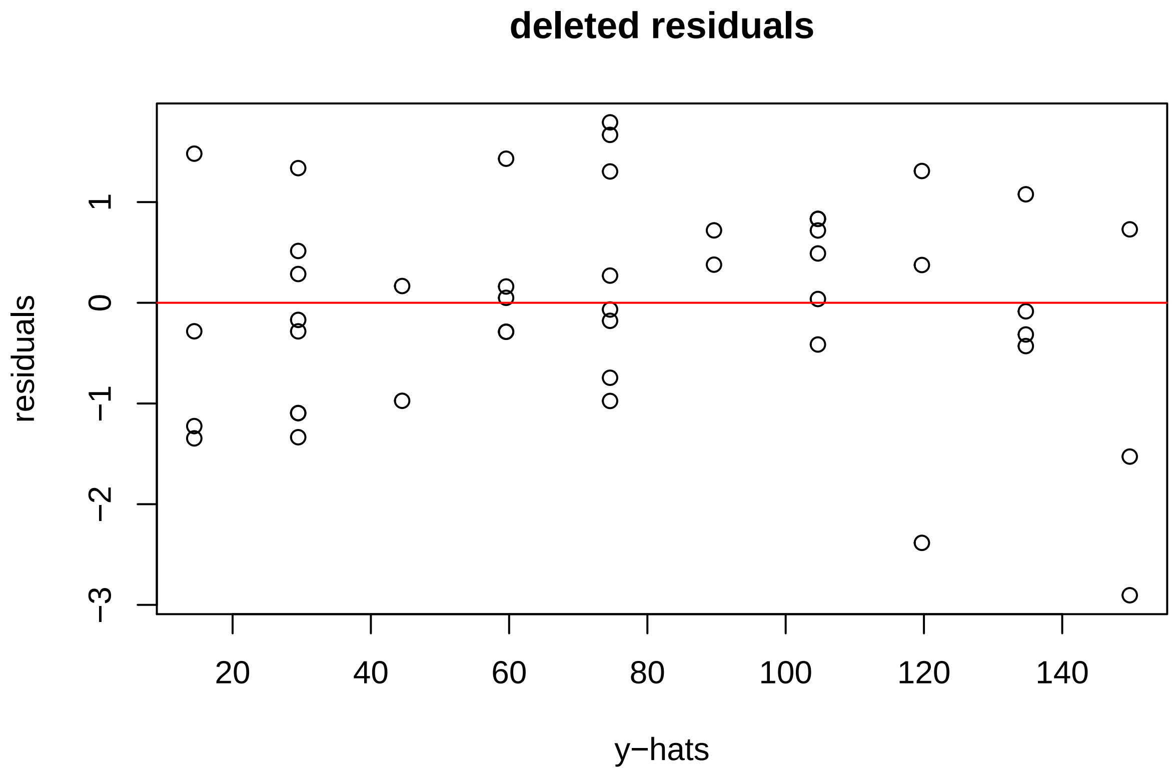



```
#line plots of deleted residuals
plot(1:45,rescopierslm, main="deleted residuals", xlab="copiers serviced", ylab="residuals")
```

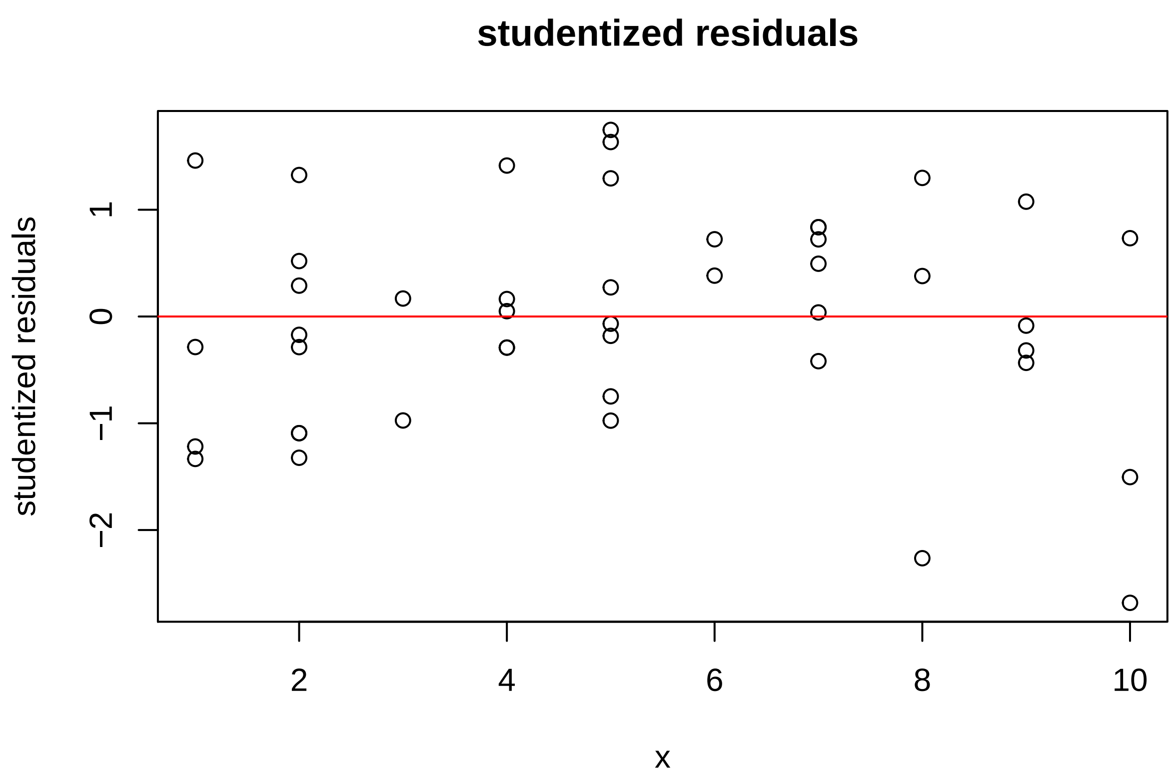
```r
lines(1:45, rescopierslm,  col="red")
abline(0,0, col="red")
```




```r
#the studentized deleted residuals verses predicted values ^y
plot(-0.5802+15.0352*hw4$x,rescopierslm, main="deleted residuals", xlab="y-hats", ylab="residuals")
abline(0,0, col="red")
```



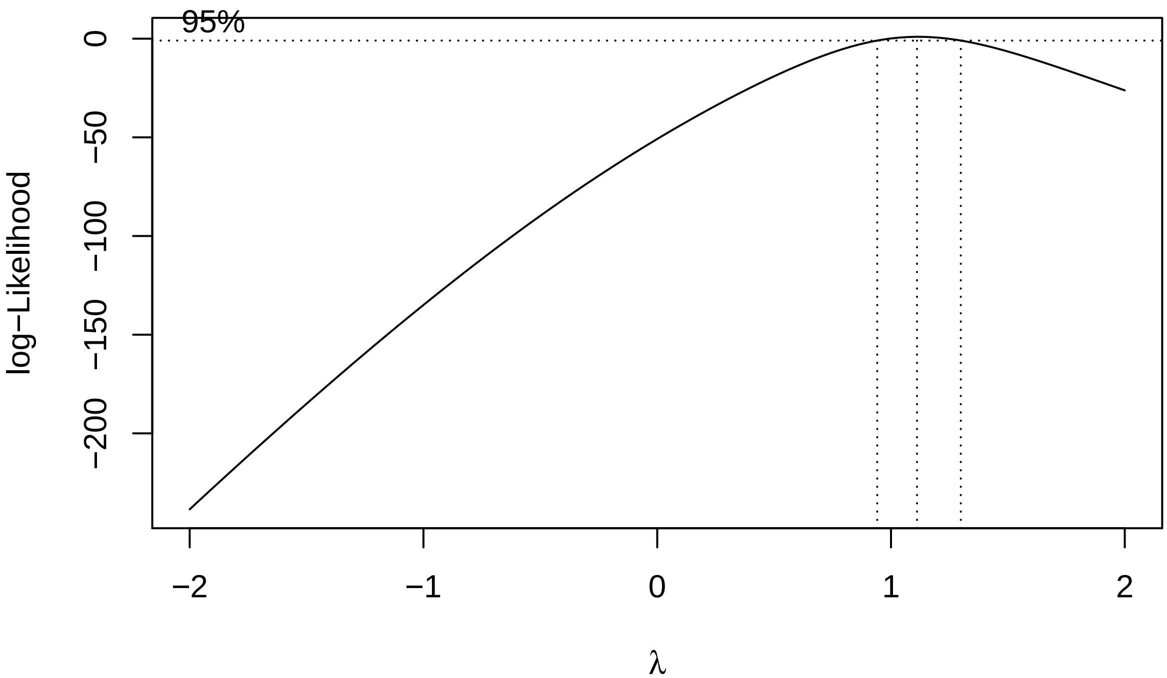

```
#studentized residuals vs. predictor variable x
studres <-rstandard(copierslm)
plot(hw4$x,studres, main=" studentized residuals", xlab="x", ylab="studentized residuals")
abline(0,0, col="red")
```



For the linear model, we notice that the linear model seems to fit the data proposed. According to the QQ plot, it seems that the model has a little bit of a left skew, which means that the variance is not constant. The histogram verifies that the distribution is almost-normal with a left skew. The line plot looks like there is a specific pattern to it, so it is not necessarily a linear model. Especially because we have discrete values, rather than continuous for the x. The last two residual plots look random, however, there are less negative values randomly placed near the center which suggests that the model might not be linear.

iii.

```
library(MASS)
copiers = boxcox(copierslm)
```




```
copiers.lambda = copiers$x[which(copiers$y==max(copiers$y))]
copiers.lambda
```

```
## [1] 1.111111
```

The lambda is close to 1, so we know that a linear model is appropriate.

Problem 2.

```
hw42 <- read.table("1_22.txt", header=F, sep = "")
time <- hw42$V2
hardness <- hw42$V1

n <- length(hardness)
c <- 4
sse.R <- anova(lm(hardness~time))[[2]][2]
fac.x <- factor(time)
sse.F <- anova(lm(hardness~fac.x))[[2]][2]
f.calc <- ((sse.R-sse.F)/(c-2))/(sse.F/(n-c))
1-pf(f.calc,c-2,n-c)
```

```
## [1] 0.4621597
```

We fail to reject the null hypothesis that linearity is not satisfied for the data.

4.

i.

```r
x<- c(10,20,30,40,50)
e <- rnorm(5,mean = 0,sd = (.8*x)^.5)
y<-20+10*x+e
y
```

```
## [1] 119.8998 216.1285 322.8241 418.3987 530.5185
```

ii. and iii.

```r
x<- c(10,20,30,40,50)
b <- rep(NA, 10000)
bweight <- rep(NA, 10000)

for(i in 1:10000){
e <- rnorm(5,mean = 0,sd = (.8*x)^.5)
y<-20+10*x+e
unweighted.model <- lm(y~x)
enew <- residuals(unweighted.model)
s.hat <- 1/(.8*x)
weighted.ls <- lm(y~x, weights = s.hat)

b[i] <- coef(unweighted.model)[2]
bweight[i] <- coef(weighted.ls)[2]
}

mean(b)
```

```
## [1] 10.00026
```

```r
var(b)
```

```
## [1] 0.0240546
```

```r
mean(bweight)
```

```
## [1] 10.00163
```

```r
var(bweight)
```

```
## [1] 0.01976351
```

iv. Both seem to be unbiased. However, the weighted beta seems to be more precise becuase it is closer to 10. Also, the var of my bweight is smaller.