Regression HW 1

Jun Lu

Due: Thursday, Sept. 5

Table of Contents

Q.1. (5 pts)

I have read Lab 1. I have installed relevant software, and am ready to use software for homework, labs, and data analysis.

Q.2, Problem 1.20 (10 pts)

```
# Load data
cm <- read.table("CH01PR20.txt", header=FALSE)
colnames(cm) <- c("time", "copiers")
head(cm, 3)
```

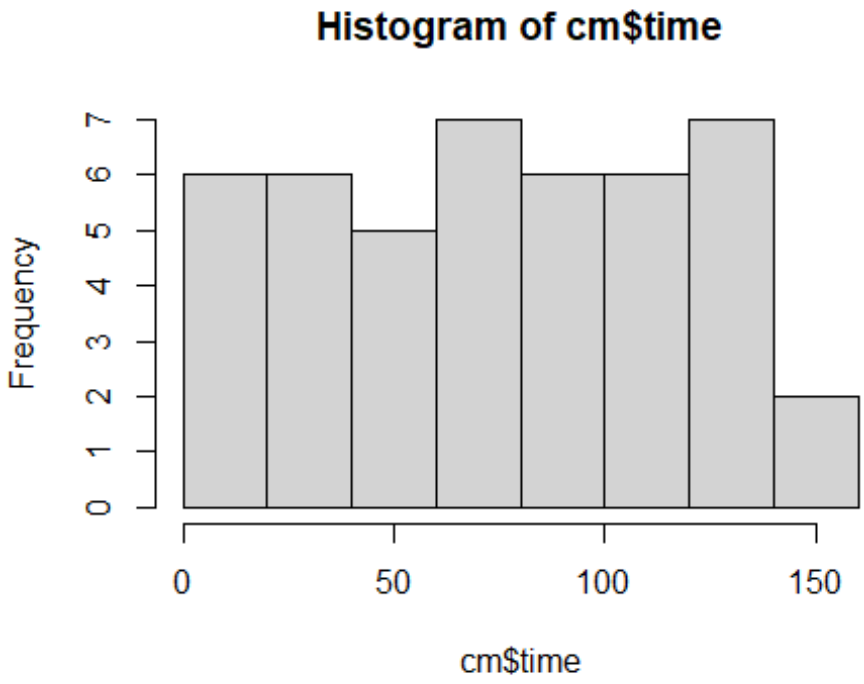
```
##   time copiers
## 1   20       2
## 2   60       4
## 3   46       3
```

```
tail(cm, 2)
```

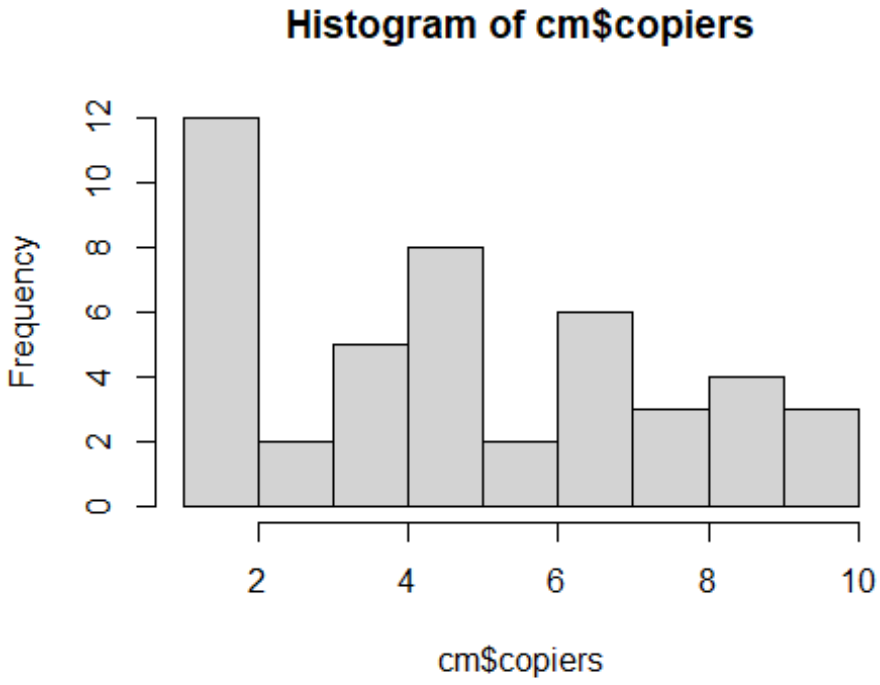
```
##    time copiers
## 44   61       4
## 45   77       5
```

Graphical summary (optional, not part of HW)

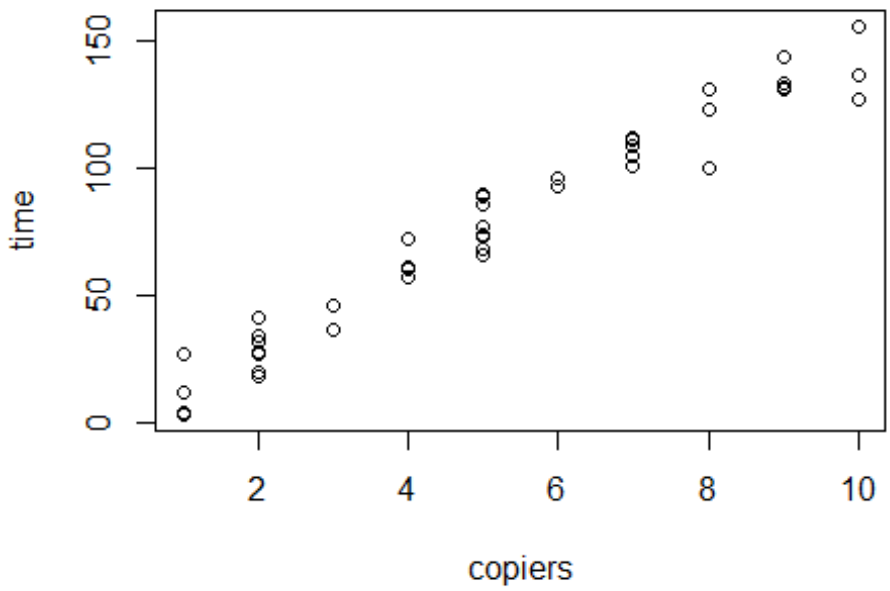
```
hist(cm$time)
```




```
hist(cm$copiers)
```




```
plot( time ~ copiers, data=cm)
```



- The response variable "time" is roughly uniformly distributed. There is no clear outliers in "time".
- The predictor "copiers" is skewed to the right.There is no clear outliers in "copiers".
- The scatter plot shows a relatively strong linear association between these two variable. the association is positive.

a. Estimated regression function (You do NOT have to copy the questions. I often include a brief description for my convenience.)

```
cm.SLR <- lm(time ~ copiers, data = cm)
cm.SLR

##
## Call:
## lm(formula = time ~ copiers, data = cm)
##
## Coefficients:
## (Intercept)      copiers
##     -0.5802      15.0352
```

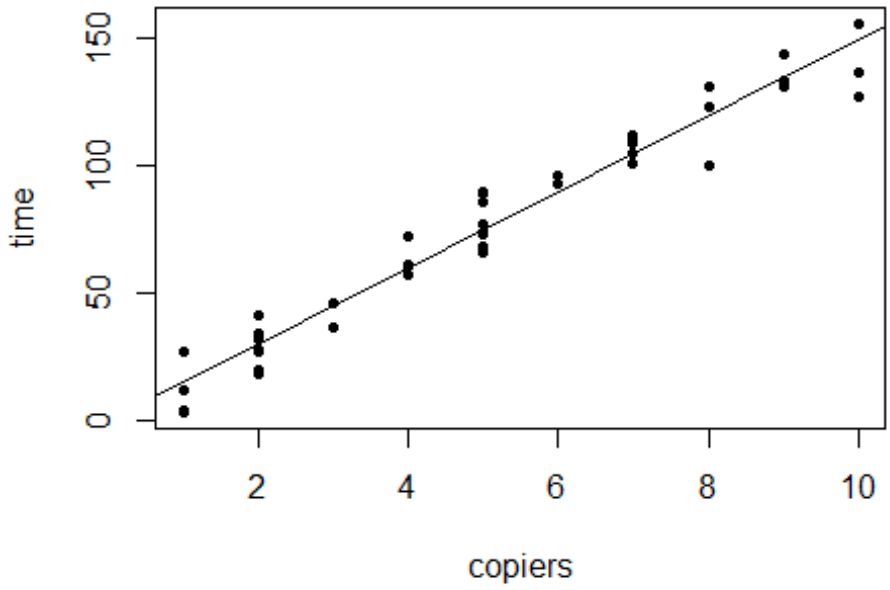
The estimated regression line (function) is:

time_hat = -0.58 + 15.04*(copiers).

$$\widehat{time} = -0.58 + 15.04 * (copiers)$$

b. Plot data with regression line

```
plot(time ~ copiers, data=cm, pch=20)
abline(cm.SLR$coefficients)
```



- The above plot shows the data with the esimated regression line. The SLR model seems to fit the data well.

c. Interpret the intercept

$$\widehat{time} = -0.58 + 15.04 * (copiers)$$

As shown in the output in part (a):

- The estimated intercept is -0.58. This means that the mean service time is estimated to be -0,58 when there is no copier. In this context, the negative intercept does NOT provide relevant information. We do not expect a negative service time when there is no copiers to repair.
- The estimated slope is 15.04. For each additional copier, the mean service time is estimated to increase 15.04 min.

d. Predict y at x = 5.

$$\widehat{time} = -0.58 + 15.04 * (copiers) = -0.58 + 15.04 * (5) = 74.62$$

- For 5 copiers, the mean service time is predicted (esimated) to be 74.62 mins.

Q.3. Problem 1.24 (5 pts)

a. Residuals and Sum of the squared residuals

```
head(cm.SLR$residuals)
```

```
##           1           2           3           4           5           6
##  -9.4903394   0.4391645   1.4744125  11.5096606  -2.4550914 -12.7723238
```

```
sum(cm.SLR$residuals^2)
```

```
## [1] 3416.377
```

- The sum of the squared residuals is 3416.377. This is the minimum value of Q in Expression 1.8 on p.15.

b. Estimate σ^2 and σ.

```
sum(cm.SLR$residuals^2)/(45-2)
```

```
## [1] 79.45063
```

```
sqrt(sum(cm.SLR$residuals^2)/(45-2))
```

```
## [1] 8.913508
```

- The estimate of σ^2 is 79.45
- The estimate of sigma is 8.91. The unit on sigma is "minute", same as the unit of the response variable.