R Codes for Chapter-1

Importing data from internet

When downloading data from internet, use "**read.table()".** In the arguments of the function:

- header: if TRUE, tells R to include variables names when importing,
- sep: tells R how the entries in the data set are separated.
    - sep=",": when entries are separated by COMMAS
    - sep="\t": when entries are separated by TAB
    - sep=" ": when entries are separated by SPACE

E.g:- The following command is used to import the data for Plastic Hardness example (exercise 1.22)

```
> data<-read.table("http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/
data/textdatasets/KutnerData/Chapter%20%201%20Data%20Sets/CH01PR22.txt ",
header= FALSE , sep="")

> data
    V1 V2
1  199 16
2  205 16
3  196 16
4  200 16
5  218 24
6  220 24
7  215 24
8  223 24
9  237 32
10 234 32
11 235 32
12 230 32
13 250 40
14 248 40
15 253 40
16 246 40
```

Importing data from the computer:

First, you need to save data in a folder in your computer. Then use **read.table()** as follows.

```
> datart=read.table("R:\\Teaching\\2017\\MA 542 F\\R Codes\\
Plastic Hardnes.csv",header = FALSE)

> datart
    V1 V2
1  199 16
2  205 16
.

.
```

This is only a part of the output.

Fitting the Simple Linear Regression (SLR) Model

The command **"lm"** can be used to fit the SLR model in R. To perform use the command:

lm (response ~ Predictor)

Here the terms response and Predictor in thecommand should be replaced by the names of the response and predictor variables, respectively, used in the analysis.

Ex. Plastic Hardness (Problem 1.22), Y=Hardness in Brinell units, X=Elapsed time in hours.

```
> Hardness=data[,1]
> Time=data[,2]
```

The following command crates a data frame, which is needed for most of the commands.

```
> dataf=data.frame(Hardness, Time)
> dataf
   Hardness Time
1       199   16
2       205   16
3       196   16
```

To fit a simple linear regression model, use the command:

```
> SLR=lm(Hardness~Time,data=dataf)
> SLR

Call:
lm(formula = Hardness ~ Time, data = dataf)

Coefficients:
(Intercept)         Time
    168.600        2.034
```

This output indicates that the fitted model is given by $\widehat{Y} = 168.600 + 2.034\,X$.

We can access more details about the fitted model by typing:

```
> summary(SLR)

Call:
lm(formula = Hardness ~ Time, data = dataf)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1500 -2.2188  0.1625  2.6875  5.5750

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 168.60000    2.65702   63.45  < 2e-16 ***
Time          2.03438    0.09039   22.51 2.16e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.234 on 14 degrees of freedom
Multiple R-squared:  0.9731,   Adjusted R-squared:  0.9712
F-statistic: 506.5 on 1 and 14 DF,  p-value: 2.159e-12
```

Extracting Estimators:

```
> b0=summary(SLR)$coefficients[1,1]
> b0
[1] 168.6

> b1=summary(SLR)$coefficients[2,1]
> b1
[1] 2.034375
```

The following command extracts the **least square estimator of the error standard deviation** $\hat{\sigma}$.

```
> sigmahat=summary(SLR)$sigma #Least square estimator.
> sigmahat
[1] 3.234027
```

We need to calculate the MLE of the error standard deviation manually.

```
> DoFR=df.residual(SLR) #Extracting error degrees of freedom:
> DoFR
[1] 14
>
> mle_sigmahat=sqrt(summary(SLR)$sigma^2*DoFR/(length(Hardness)))
> mle_sigmahat
[1] 3.025155
```

Fitted Values:

To calculate the fitted values, use the following command.

```
> Fitvals=fitted.values(SLR)
> Fitvals
      1       2       3       4       5       6       7       8       9 
201.150 201.150 201.150 201.150 217.425 217.425 217.425 217.425 233.700 
     10      11      12      13      14      15      16 
233.700 233.700 233.700 249.975 249.975 249.975 249.975 
```

Residuals:

Residuals for the fitted regression model are calculated as follows.

```
> Res=residuals(SLR)
> Res
     1      2      3      4      5      6      7      8      9     10 
-2.150  3.850 -5.150 -1.150  0.575  2.575 -2.425  5.575  3.300  0.300 
    11     12     13     14     15     16 
 1.300 -3.700  0.025 -1.975  3.025 -3.975 
```

MLE of σ in a different way:

```
> mles=sqrt(sum(Res*Res)/(length(Hardness)))
> mles
[1] 3.025155
```

Checking the Properties of residuals:

1. $\sum_{i=1}^{n} e_i = 0$

```
> sumei=sum(Res)
> sumei
[1] -1.998401e-15
```

2. $\sum_{i=1}^{n} X_i e_i = 0$

```
> sumXiei=sum(Time*Res)
> sumXiei
[1] -6.306067e-14
```

3. $\sum_{i=1}^{n} \widehat{Y}_i e_i = 0$

```
> sumyihatei=sum(Fitvals*Res)
> sumyihatei
[1] -5.782042e-13
```

4. $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{Y}_i$

```
> sumyi_sumyihat=sum(Hardness)-sum(Fitvals)
> sumyi_sumyihat
[1] 0
```

5. Fitted Regression line passes through the point $(\bar{X}, \bar{Y})$.

```
> XbarYbar=mean(Hardness)-(b0+b1*mean(Time))
> XbarYbar
[1] 0
```