MA 331 - Hw #1

"I pledge my honor I have abided by the Stevens's honor code"
-Yasmin Sahi
2/12/21

#2) $$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= (x_1^2 + x_2^2 + \dots + x_n^2) - 2(x_1 + x_2 + \dots + x_n)\bar{x} + \bar{x}^2 \cdot n$$

$$= (x_1^2 + x_2^2 + \dots + x_n^2) - 2(n\bar{x})\bar{x} + \bar{x}^2 n$$

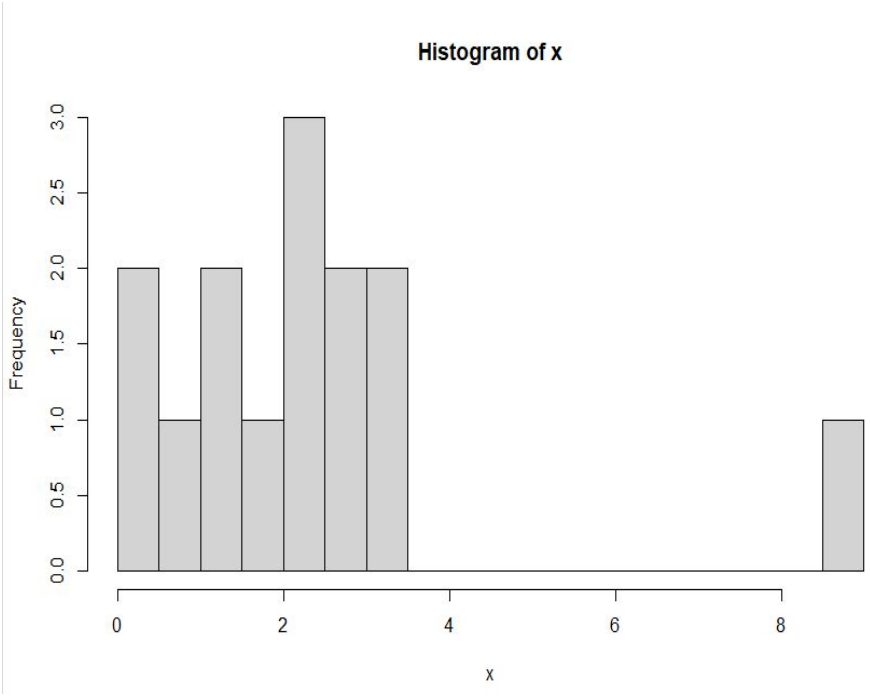
$$= \boxed{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \Rightarrow \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

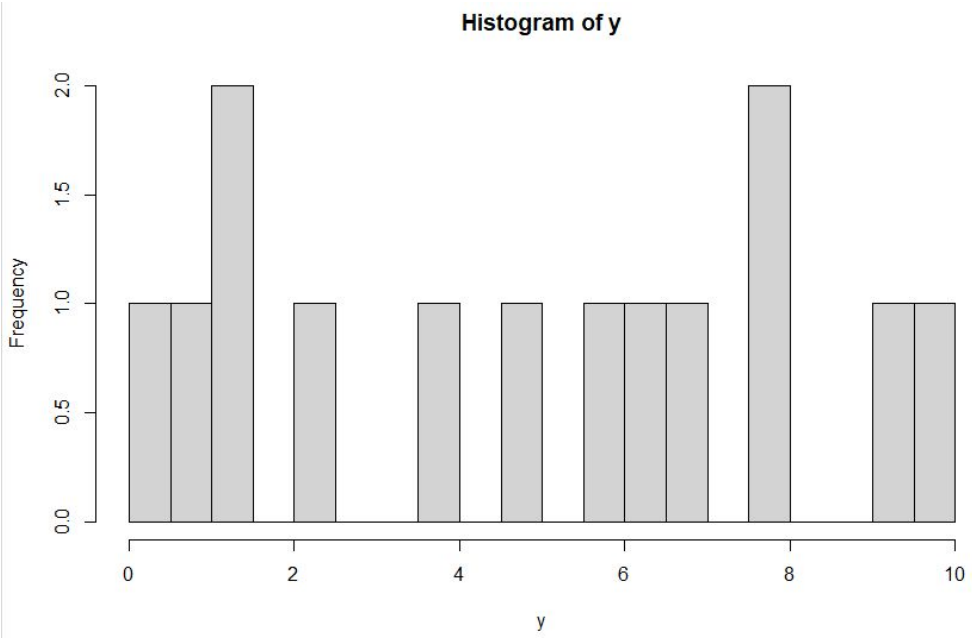
#1) **Data X**

- Apx. Uniform
- Outlier (9.0)
- Excluding outlier apx. left skewed
- range ~8.8

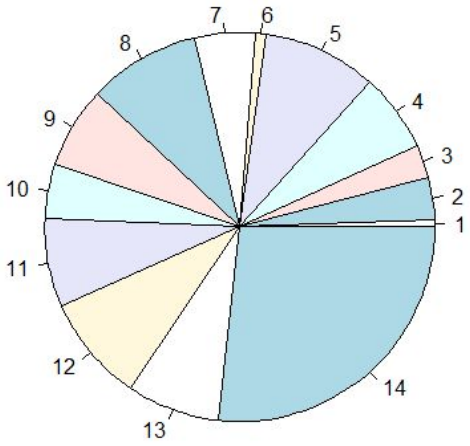
Data Y

- Bimodal (~10.5, ~8) ← modes
- No outliers
- Roughly symmetric
- ~ Range = 9.7

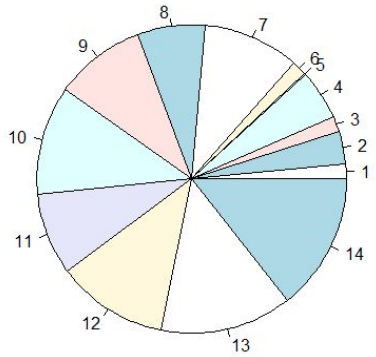




Piechart for data X



Piechart for data Y



1ii) Boxplot X

Summary: Min Q1 Median Q3 Max
0.20 1.275 2.250 2.90 9.00

Variance: 4.568407

Outliers: x = 9.0

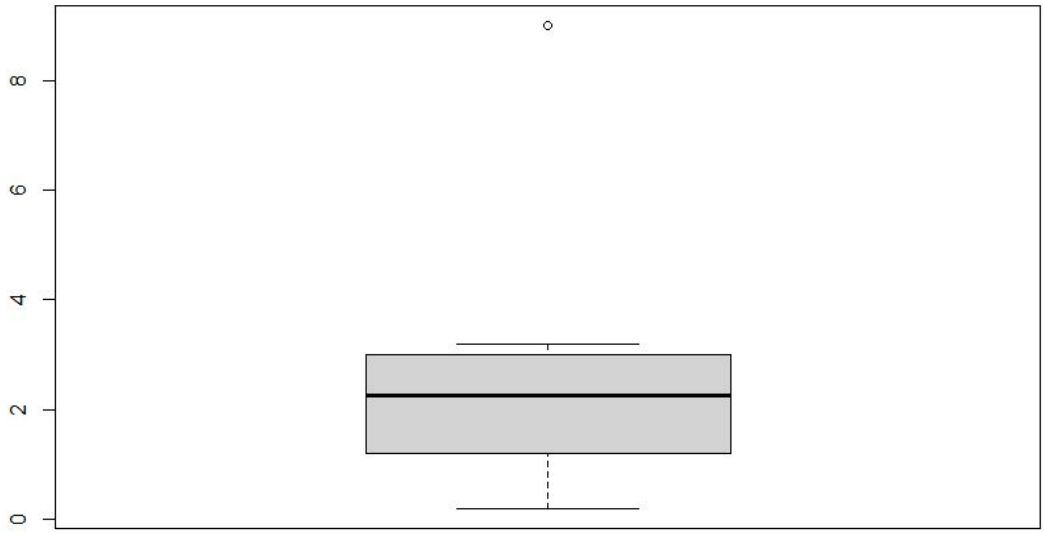
Boxplot Y

Outliers: N/A

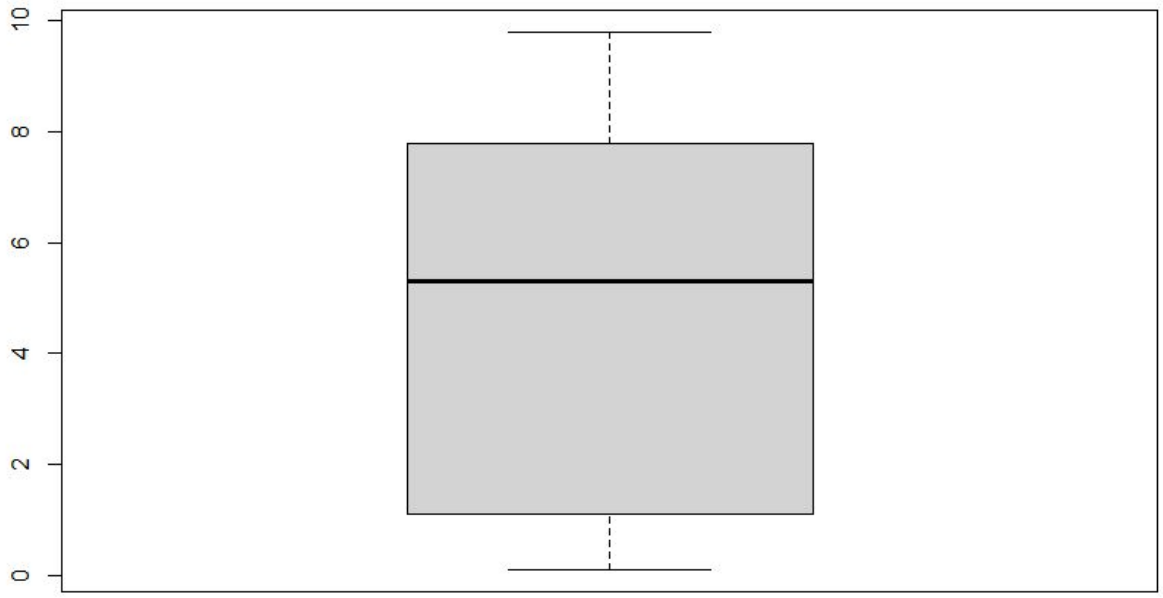
Variance: 11.17143

Summary: Min Q1 Median Q3 Max
0.10 1.400 5.300 7.575 9.800

Boxplot for Data X



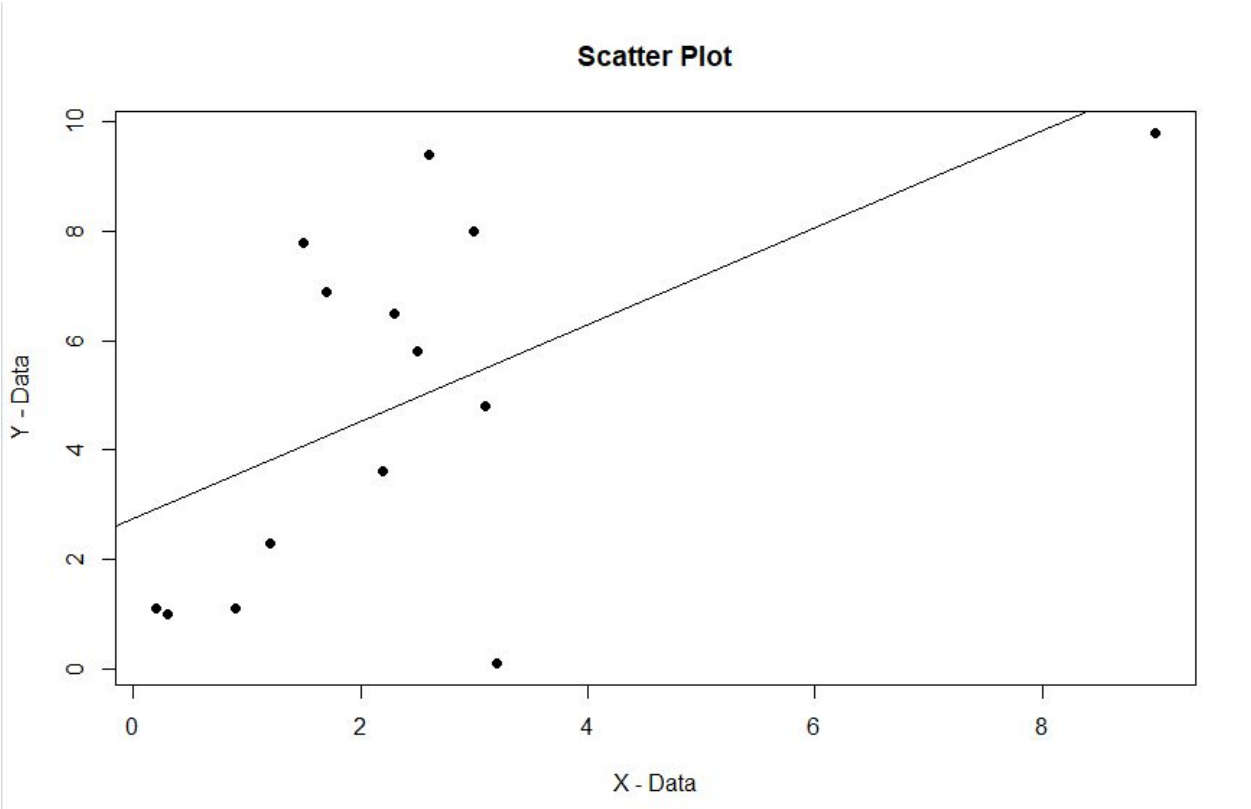
Boxplot for Data Y



1 iii) $r(x,y) = 0.5679153$

Correlation Coef.

Moderately strong positive linear correlation between (x, y). Note outlier (9, 9.8).

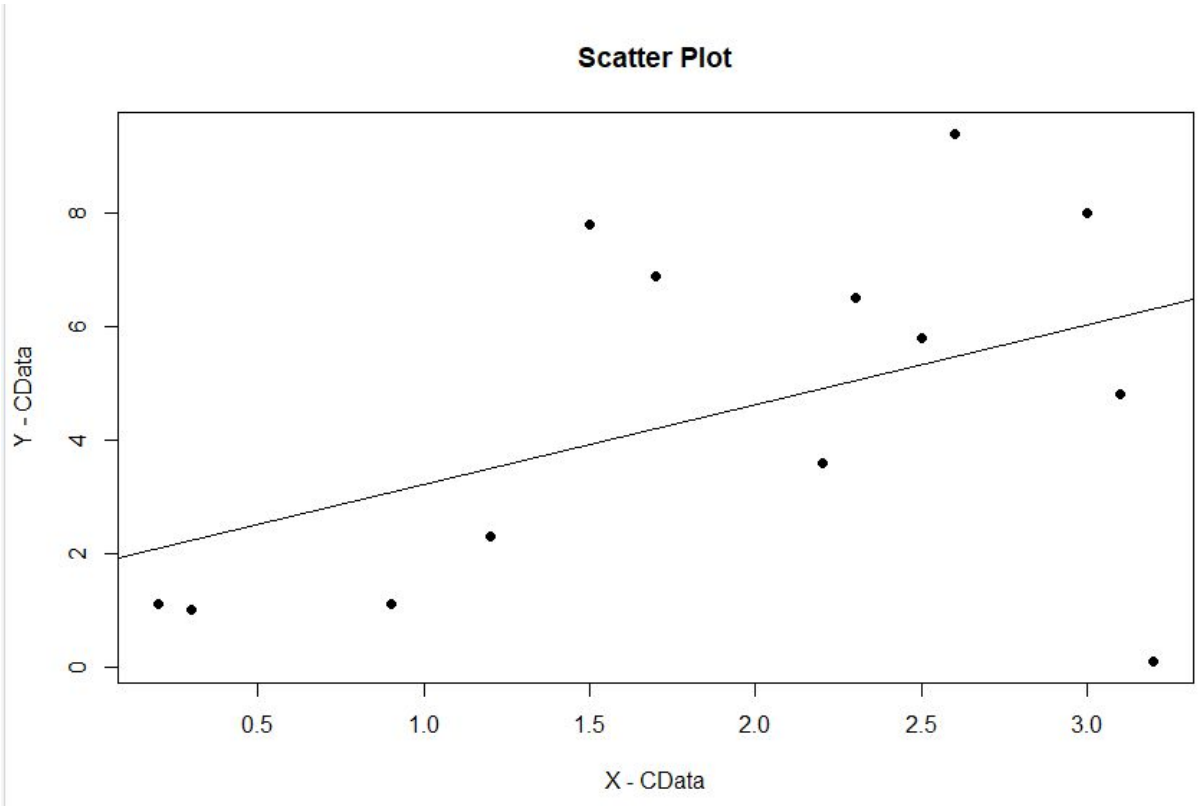


1 iv) Outlier (9, 9.2)

$r_c(x,y) = 0.4586256$

Correlation coeff w/o outlier

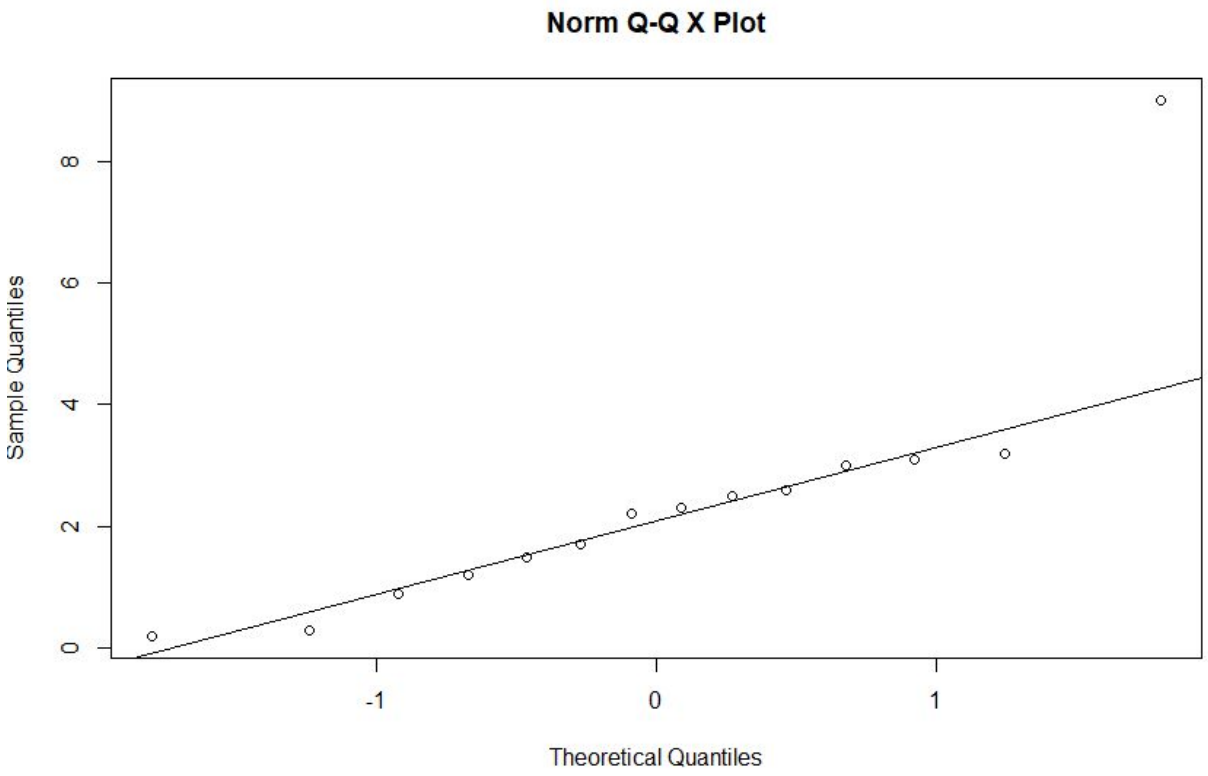
Linear regression model with outlier removed (corrected data, ie CData)

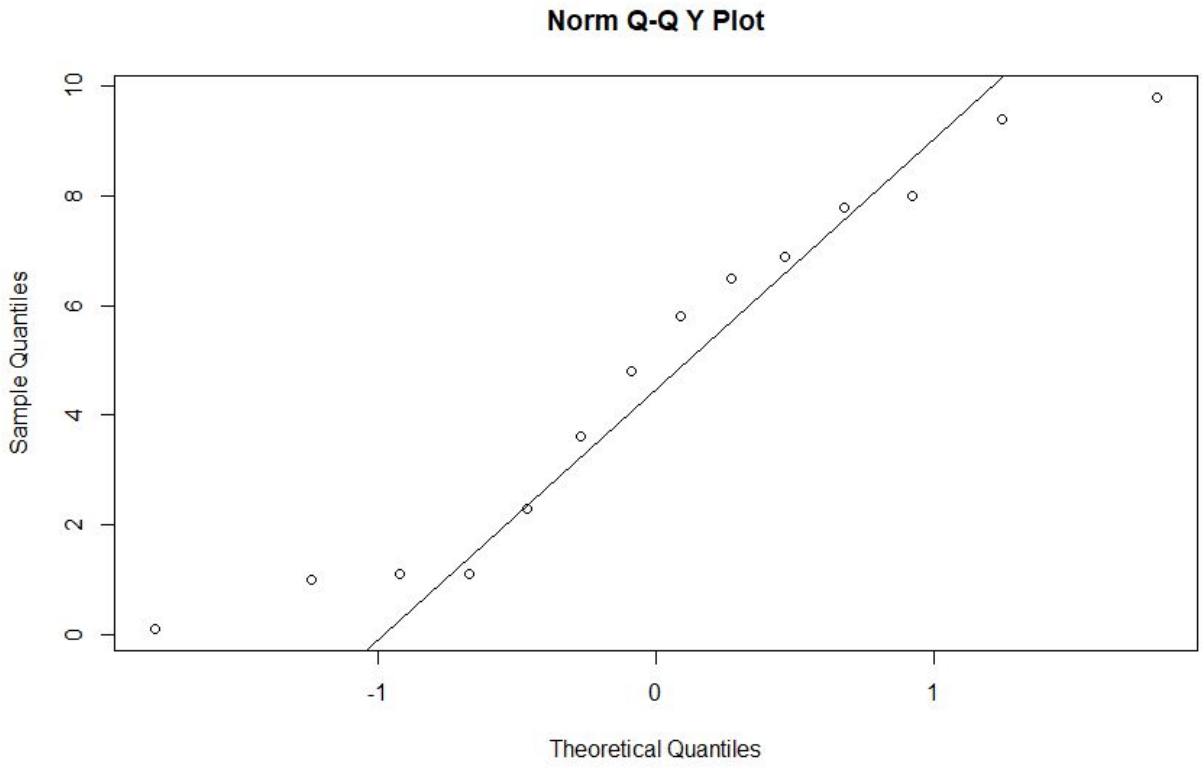


1v) It seems the correlation co-ef of (iii) *including outlier* is higher than that of (iv) *exluding outlier.

> iii better correlates with data than (iv)

1vi) The values of X are more likely to be normal then Y, as the plot of the theoretical normal line (qqline) of data x more closely fits the norm apx of data X than the (qqline) of data Y & normal apx. of data Y.


Norm Q-Q X Plot
Sample Quantiles
0
2
4
6
8
-1
0
1
Theoretical Quantiles


Norm Q-Q Y Plot
Sample Quantiles
0
2
4
6
8
10
-1
0
1
Theoretical Quantiles