Question 2

January 28, 2021

(a)

[1]:
```
x1 <- c(3,2,1,10,5,8,9)
x2 <- c(6,4,2,20,10,16,17)
#data matrix
X <- matrix (c(x1,x2),nrow=7,ncol=2,byrow=F)
X
```

A matrix: 7×2 of type dbl

3	6
2	4
1	2
10	20
5	10
8	16
9	17

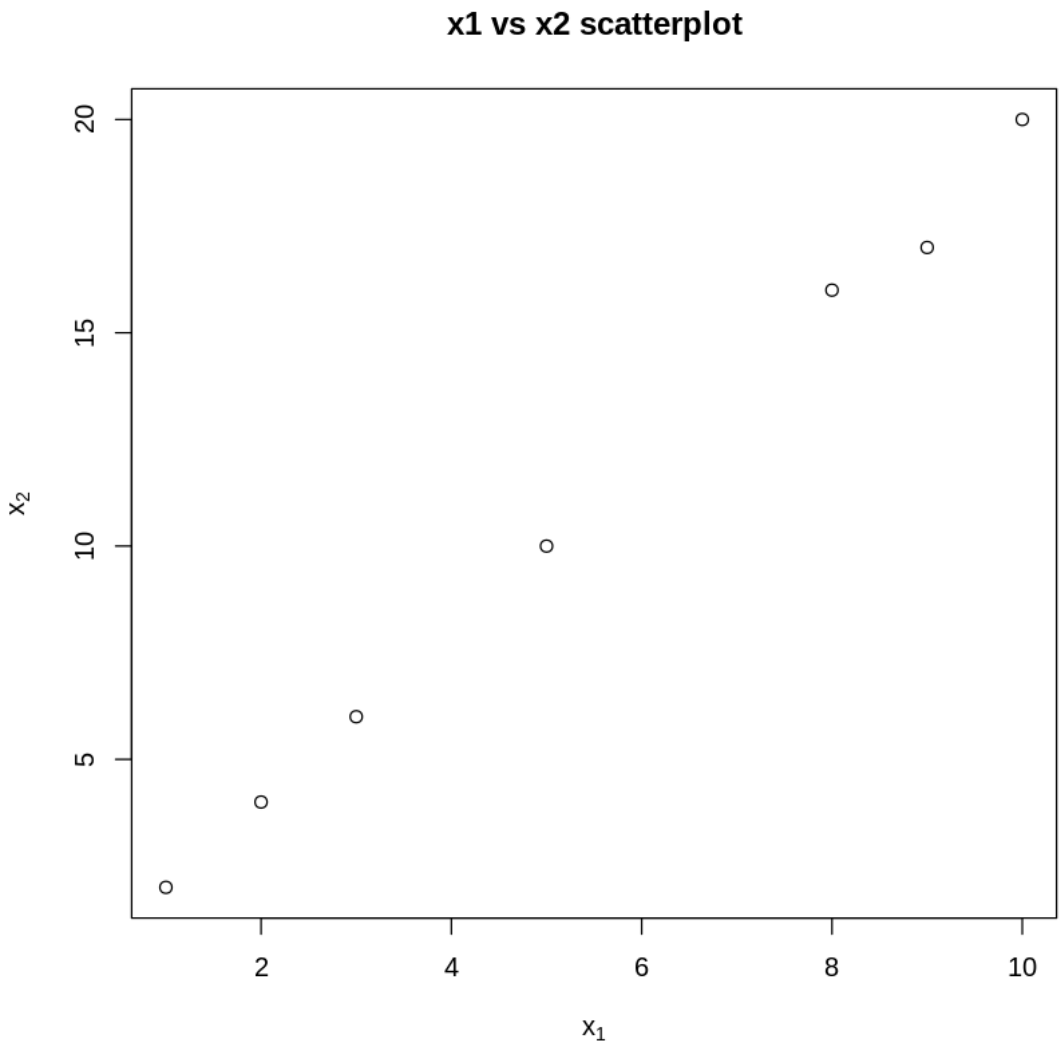
Data Matrix:

$$X = \begin{bmatrix} 3 & 6 \\ 2 & 4 \\ 1 & 2 \\ 10 & 20 \\ 5 & 10 \\ 8 & 16 \\ 9 & 17 \end{bmatrix} \tag{1}$$

(b)

Scatter plot:

[2]:
```
plot(X, main="x1 vs x2 scatterplot",
     xlab=expression(x[1]), ylab=expression(x[2]))
```



(c)

[3]:
```
#sample mean vector
x_bar <- colMeans(X)
x_bar
```

1. 5.42857142857143 2. 10.7142857142857

The sample mean vector is:

$$\vec{\bar{x}} = \begin{pmatrix} 5.42857142857143 \\ 10.7142857142857 \end{pmatrix} \tag{2}$$

[4]:
```
R <- cor(X)
R
```

A matrix: 2 × 2 of type dbl

1.0000000	0.9988343
0.9988343	1.0000000

The sample correlation matrix is:

$$R = \begin{bmatrix} 1.0000000 & 0.9988343 \\ 0.9988343 & 1.0000000 \end{bmatrix} \tag{3}$$

(d)

Each entry of the sample correlation matrix r_{ik} represents the sample correlation coefficient for the i^{th} and k^{th} random variables in the data matrix. In this data set, x_1 and x_2 have a positive correlation(one increase as another increases), and strength of the linear relationship is very high (r_{12} and r_{21} close to 1).

The values on the diagonal always equal to one because those are the correlation of variables with themselves.