SANDRA REBECCA BABTALE

PROBABILITY AND STATISTICS REVIEW

QUESTION ①

Given $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $E(\mathbf{X}) = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

(a) Show that $E(A\mathbf{X}) = AE(\mathbf{X})$

$$A\mathbf{X} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}X_1 + a_{12}X_2 \\ a_{21}X_1 + a_{22}X_2 \end{pmatrix}$$

$$E(A\mathbf{X}) = \begin{pmatrix} E(a_{11}X_1 + a_{12}X_2) \\ E(a_{21}X_1 + a_{22}X_2) \end{pmatrix}$$

$$= \begin{pmatrix} E(a_{11}X_1) + E(a_{12}X_2) \\ E(a_{21}X_1) + E(a_{22}X_2) \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}E(X_1) + a_{12}E(X_2) \\ a_{21}E(X_1) + a_{22}E(X_2) \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix}$$

$$= AE(\mathbf{X})$$

(b) If

$$Y = AX = \begin{pmatrix} a_{11}X_1 + a_{12}X_2 \\ a_{21}X_1 + a_{22}X_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

$$\Rightarrow Y_1 = a_{11}X_1 + a_{12}X_2 \quad , \quad Y_2 = a_{21}X_1 + a_{22}X_2$$

$$\Rightarrow Var(Y_1) = Var(a_{11}X_1 + a_{12}X_2)$$

$$= Var(a_{11}X_1) + Var(a_{12}X_2) + 2Cov(a_{11}X_1, a_{12}X_2)$$

$$= a_{11}^2 Var(X_1) + a_{12}^2 Var(X_2) + 2a_{11}a_{12}Cov(X_1, X_2)$$

$$\Rightarrow Var(Y_2) = Var(a_{21}X_1 + a_{22}X_2)$$

$$= Var(a_{21}X_1) + Var(a_{22}X_2) + 2Cov(a_{21}X_1, a_{22}X_2)$$

$$= a_{21}^2 Var(X_1) + a_{22}^2 Var(X_2) + 2a_{21}a_{22}Cov(X_1, X_2)$$

$$Cov(Y_1, Y_2) = Cov(a_{11}X_1 + a_{12}X_2, a_{21}X_1 + a_{22}X_2)$$

$$= E[(a_{11}X_1 + a_{12}X_2)(a_{21}X_1 + a_{22}X_2)] - E[a_{11}X_1 + a_{12}X_2] \cdot E[a_{21}X_1 + a_{22}X_2]$$

$$= E[a_{11}a_{21}X_1^2 + a_{11}a_{22}X_1X_2 + a_{12}a_{21}X_2X_1 + a_{12}a_{22}X_2^2] - \left[(a_{11}E[X_1] + a_{12}E[X_2]) \cdot (a_{21}E[X_1] + a_{22}E[X_2])\right]$$

$$= a_{11} a_{21} E[X_1^2] + a_{11} a_{22} E[X_1 X_2] + a_{12} a_{21} E[X_2 X_1]$$

$$+ a_{12} a_{22} E[X_2^2] - \left(a_{11} a_{21} E^2[X_1] + a_{11} a_2 E[X_1] E[X_2] \right)$$

$$- \left(a_{12} a_{21} E[X_2] E[X_1] + a_{12} a_{22} E^2[X_2] \right)$$

$$= a_{11} a_{21} \left(E[X_1^2] - E^2[X_1] \right) + a_{12} a_{22} \left(E[X_2^2] - E^2[X_2] \right)$$

$$+ 2 a_{11} a_{22} \left(E[X_1 X_2] - E[X_1] E[X_2] \right)$$

$$\Rightarrow \mathrm{Cov}(Y_1, Y_2) = a_{11} a_{21} \mathrm{Var}(X_1) + a_{12} a_{22} \mathrm{Var}(X_2) + 2 a_{11} a_{22} \mathrm{Cov}(X_1, X_2)$$

2.

```
data =[-0.4326;-1.6656;0.1253;0.2877;-1.1465;1.1909;1.1892;-0.0376;0.3273;0.1746];
n = length(data);
%Sample mean
Sample_mean = mean(data)
```

```
Sample_mean = 0.0013
```

```
%Standard deviation
standard_deviation = std(data)
```

```
standard_deviation = 0.9034
```

```

% Confidence interval
ci = 0.95;
top = 1-(1-ci)/2;
bottom = (1-ci)/2;
t_top = tinv(top,n-1);
t_bottom = tinv(bottom,n-1);

Confidence_interval = [Sample_mean+(t_bottom*standard_deviation)/sqrt(n) Sample_mean+(t
```

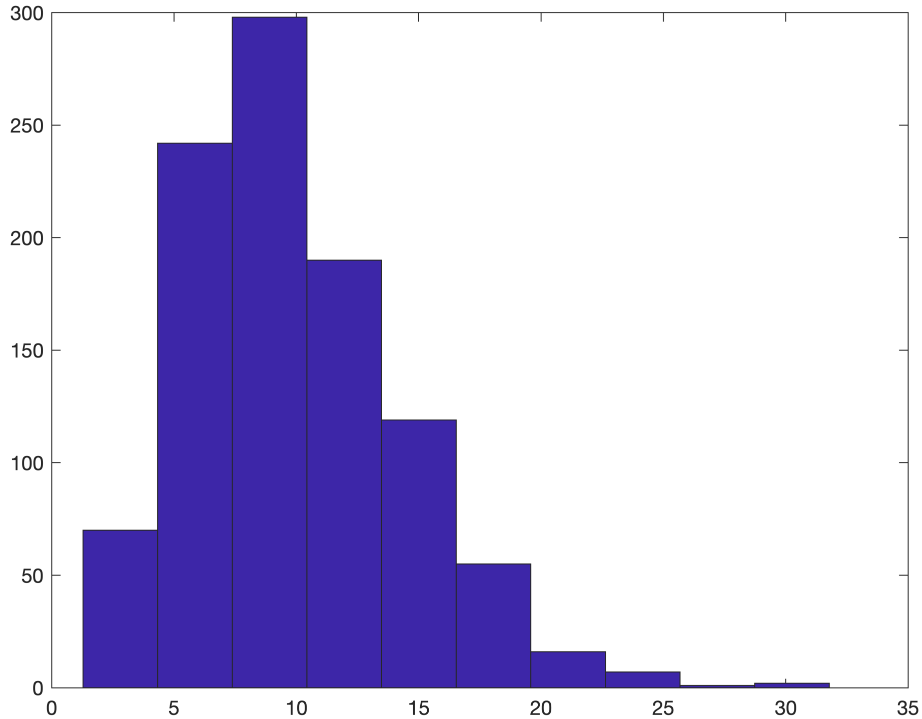
```
Confidence_interval = 1×2
   -0.6450    0.6475
```

3.

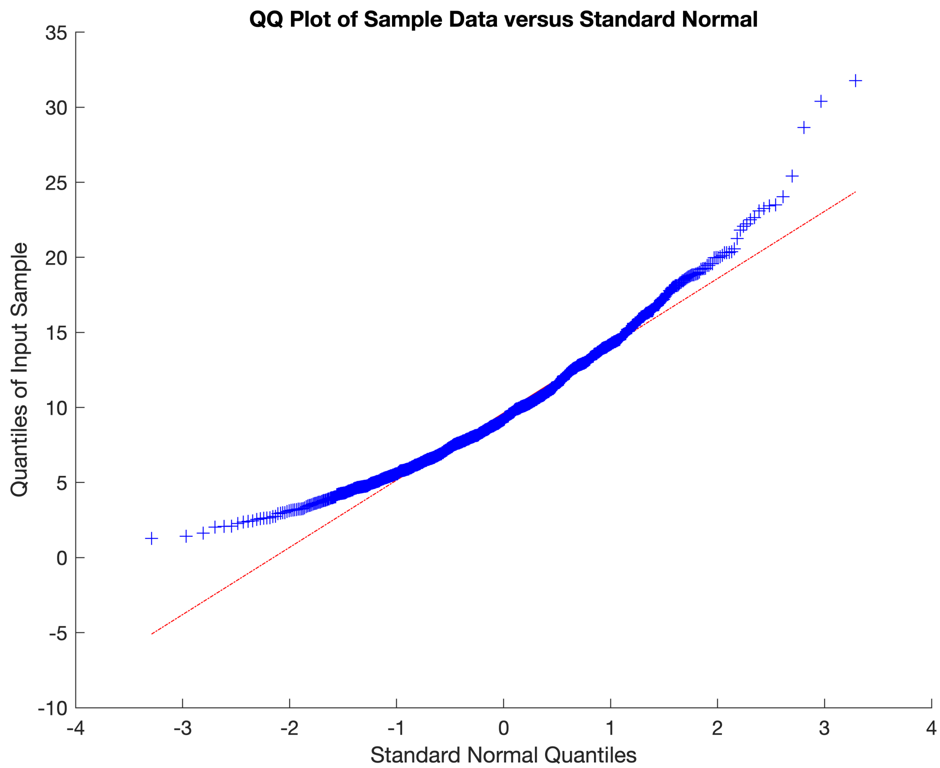
```



N = 1000;
n5 = 5; n50 =50;
[m5] = generate(N,n5);
[m50] = generate(N,n50);
figure(1)
hist(m5)
```




```
figure(2)
qqplot(m5)
```



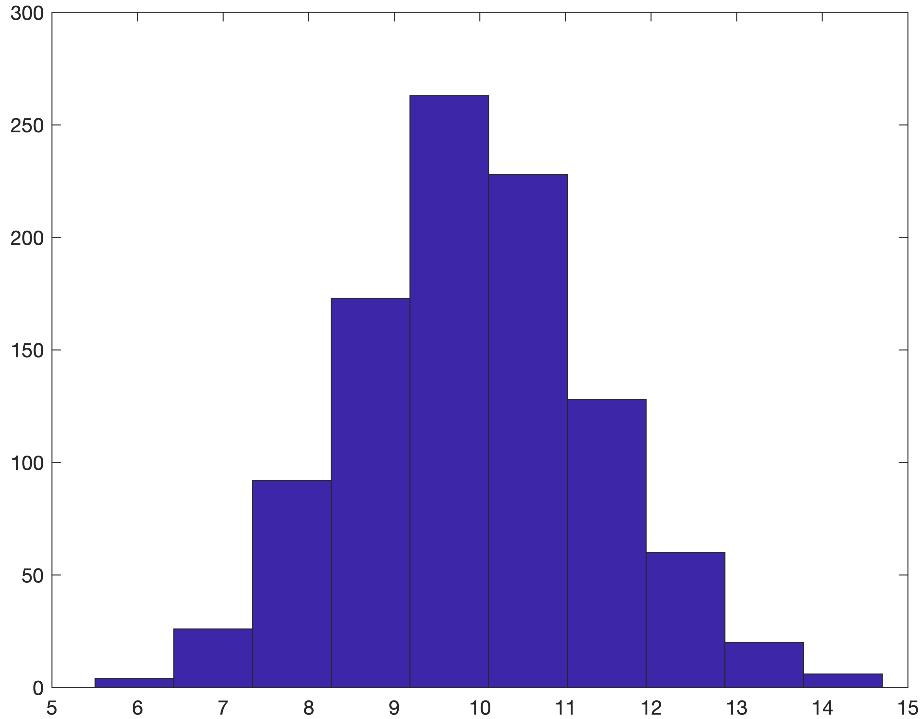
The histogram and Q-Q plot above show that the averages are not approxiately normally distributed.

This is so because from the histogram we can see that it is skewed to the right (there is a sharp drop on the left and a slow one on the right) and not symmetric about the mean as it would have if the averages were normarlly distributed.

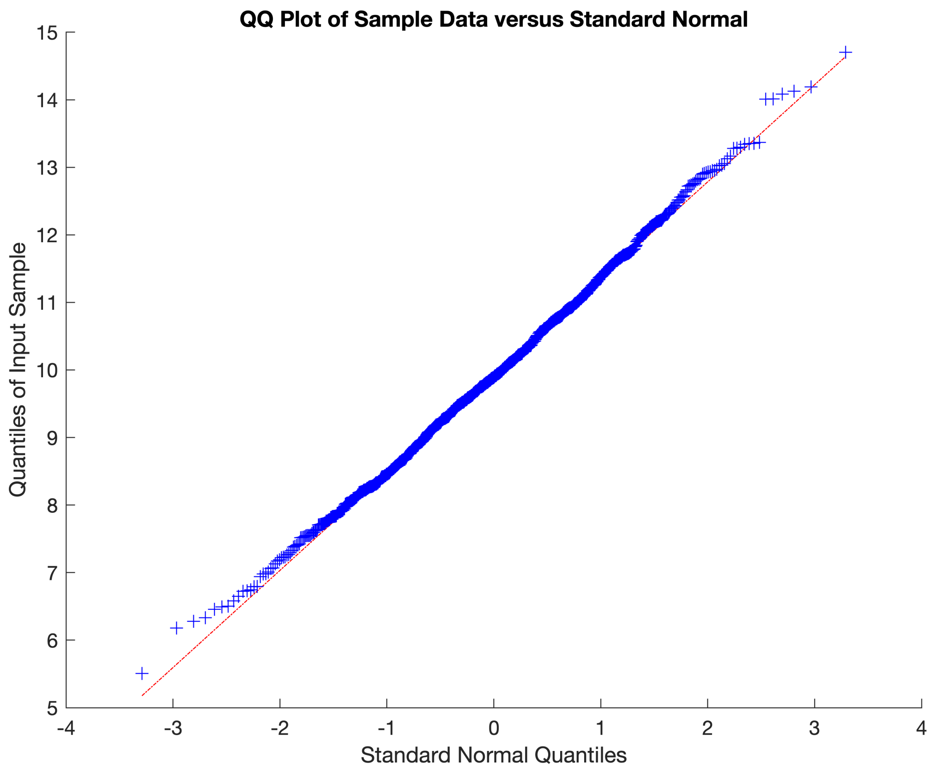
From the Q-Q plot, we see a curvature formed by plot of the averages and since they don't follow a straight line this clearly shows that the averages are not normally distributed since normally distributed data appears as roughly a straight line.

4.

```
figure(3)
hist(m50)
```




```
figure(4)
qqplot(m50)
```



For this case, the histogram and Q-Qplot show that the averages are approximately noramally distributed.

This is so because from the histogram we can see that although the result is not perfectly symmetric, the mean, median and mode are almost around the same point and we also see that there is no longer the combination of a sharp drop on one side and a slow decline on the other.

From the Q-Q plot we can see a that the plot roughly follows a straight diagonal line (although the ends of the plot often start to deviate from the straight line)and this show that the averages are approximately normally distributed.

```
function [ave] = generate(N,n)
    ave = [];
    for i=1:N
        m = mean(exprnd(10,n,1));
        ave = [ave m];
    end
end
```