

# Conditional Expectation MVN

Ilke Kas

Due: Monday, February 24, 2025, by 23:59

1

Let  $X$  be a  $p$ -dimensional random vector with mean and variance given by:

$$E(X) = \mu, \quad \Sigma = E[(X - \mu)(X - \mu)^\top]$$

Show that:

$$\text{Var}(AX) = A\Sigma A^\top$$

where  $A \in \mathbb{R}^{d \times p}$  is a constant matrix.

**Solution:** Since  $A$  is a constant matrix, we know that the expected value of it can be expressed like this:

$$E(AX) = A \cdot \mu \tag{1}$$

We know the variance is calculated as in the following for the random vectors:

$$\text{Var}(AX) = E[(AX - E(AX)) \cdot (AX - E(AX))^\top] \tag{2}$$

If we substitute the  $E(AX)$  in equation (2) with equation (1), we will get:

$$\text{Var}(AX) = E[(AX - A\mu) \cdot (AX - A\mu)] \tag{3}$$

Realize that we can factor out the  $A$  constants in equation (4) as in the following:

$$\text{Var}(AX) = E[A \cdot (X - \mu) \cdot (X - \mu)^\top A^\top] \tag{4}$$

Again, by using the fact that in equation (1), we can rewrite the equation (4) as in the following:

$$\text{Var}(AX) = A \cdot E[(X - \mu)(X - \mu)^\top] A^\top \tag{5}$$

In the question definition we already know that  $\Sigma$  is defined as:

$$\Sigma = E[(X - \mu)(X - \mu)^\top] \tag{6}$$

Therefore, if we substitute the expression  $E[(X - \mu)(X - \mu)^\top]$  in the equation (5) with the  $\Sigma$  by using the fact in equation (6), we show what is asked for:

$$\text{Var}(AX) = A\Sigma A^\top \tag{7}$$

## 2

Suppose  $X \in \mathbb{R}^p$  and  $Y \in \mathbb{R}^q$  and:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q} \left( \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)$$

where  $\Sigma_{XX}$  and  $\Sigma_{YY}$  are positive definite.

(i) Find  $E(Y|X)$

**Solution:**

$$E(Y|X) = \mu_Y + \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot (X - \mu_X) \quad (8)$$

(ii) Find  $E(X|Y)$

**Solution:**

$$E(X|Y) = \mu_X + \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot (Y - \mu_Y) \quad (9)$$

(iii) What is  $\text{Var}(Y|X)$ ?

**Solution:**

$$\text{Var}(Y|X) = \Sigma_{YY} - \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot \Sigma_{XY} \quad (10)$$

(iv) What is  $\text{Var}(X|Y)$ ?

**Solution:**

$$\text{Var}(X|Y) = \Sigma_{XX} - \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot \Sigma_{YX} \quad (11)$$

(v) Suppose  $E(Y|X) = \beta_0 + \beta_1^\top X$ . What will be the expressions for  $\beta_0$  and  $\beta_1$  based on your answer in (i)?

**Solution:** We already know the  $E(Y|X)$  from part (i) in equation (8). Let's rearrange this equation by expanding term  $\Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot (X - \mu_X)$ :

$$E(Y|X) = \mu_Y + \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot X - \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot \mu_X = (\mu_Y - \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot \mu_X) + (\Sigma_{YX} \cdot \Sigma_{XX}^{-1}) \cdot X \quad (12)$$

When we write the  $E(Y|X)$  as in the equation (12), realize that the  $\beta_0$  and  $\beta_1$  values can be expressed as in the following:

$$\beta_0 = \mu_Y - \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \cdot \mu_X \quad (13)$$

and

$$\beta_1 = \Sigma_{YX} \cdot \Sigma_{XX}^{-1} \quad (14)$$

(vi) Suppose  $E(X|Y) = \alpha_0 + \alpha_1^\top Y$ . What will be the expressions for  $\alpha_0$  and  $\alpha_1$  based on your answer in (ii)?

We already know the  $E(X|Y)$  from part (ii) in equation (9). Let's rearrange this equation by expanding term  $\Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot (Y - \mu_Y)$ :

$$E(X|Y) = \mu_X + \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot Y - \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot \mu_Y = (\mu_X - \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot \mu_Y) + (\Sigma_{XY} \cdot \Sigma_{YY}^{-1}) \cdot Y \quad (15)$$

When we write the  $E(X|Y)$  as in the equation (15), realize that the  $\alpha_0$  and  $\alpha_1$  values can be expressed as in the following:

$$\alpha_0 = \mu_X - \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \cdot \mu_Y \quad (16)$$

and

$$\alpha_1 = \Sigma_{XY} \cdot \Sigma_{YY}^{-1} \quad (17)$$

(vii) From your answers in (v) and (vi), does regressing  $Y$  on  $X$  yield the same results as regressing  $X$  on  $Y$ ?

**Solution:** Regressing  $Y$  on  $X$  does not give us the same results as regressing  $X$  on  $Y$ . This is because the relationships and how one depends on the other are different in each case, which we can see in the different coefficients and intercepts for each regression. When we regress  $Y$  on  $X$ , we are trying to predict  $Y$  based on  $X$ . The coefficients show how much  $Y$  changes with a change in  $X$ . However, regressing  $X$  on  $Y$  is about predicting  $X$  using  $Y$ .

### 3

Let  $X \sim N_4(\mu, \Sigma)$ , where:

$$\mu = \begin{pmatrix} 4 \\ 1 \\ -5 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{bmatrix} 3 & -2 & 0 & 1 \\ -2 & 5 & 0 & 2 \\ 0 & 0 & 2 & -1 \\ 1 & 2 & -1 & 9 \end{bmatrix}$$

(i) Find the distribution of the following variables and determine if the variables are independent.

- (a)  $X_1$  and  $X_3$

**Solution:** We know that all subsets of the components of  $X$  have a multivariate normal distribution. For marginal ones, they also have normal distribution. Therefore, by using the corresponding means and the variances given in the question, we can find the distribution of  $X_1$  and  $X_3$  :

$$X_1 \sim N(\mu = 4, \sigma^2 = 3) \quad (18)$$

and

$$X_3 \sim N(\mu = -5, \sigma^2 = 2) \quad (19)$$

In order to see whether they are independent we can look at the covariance value  $\Sigma_{13}$  since we know that zero covariance implies that the corresponding components are independent. As seen from  $\Sigma_{13} = 0$ , we can say that  $X_1$  and  $X_3$  are independent.

- (b)  $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  and  $X_3$

**Solution:** We know that all subsets of the components of  $X$  have a multivariate normal distribution. Since  $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  is subset of  $X$ , we need to get the corresponding mean and the covariance matrix from  $\mu$  and  $\Sigma$  matrices provided in the question.

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\mu = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 3 & -2 \\ -2 & 5 \end{pmatrix}) \quad (20)$$

From the previous question, we know that:

$$X_3 \sim N(\mu = -5, \sigma^2 = 2)$$

Let's look at the dependency of this. As we mentioned, since zero covariance implies that the corresponding components are independent, we need to look at the covariance of these components:

$$\Sigma_{(X_1, X_2), X_3} = (X^{(1)} - \mu^{(1)}) \cdot (X^{(2)} - \mu^{(2)}) \quad (21)$$

where  $X^{(1)} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  and  $X^{(2)} = X_3$

$$\Sigma_{(X_1, X_2), X_3} = \begin{pmatrix} X_1 - 4 \\ X_2 - 1 \end{pmatrix} \cdot (X_3 + 5) = \begin{pmatrix} (X_1 - 4) \cdot (X_3 + 5) \\ (X_2 - 1) \cdot (X_3 + 5) \end{pmatrix} \quad (22)$$

Realize that

$$\Sigma_{(X_1, X_2), X_3} = \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} \quad (23)$$

Again by using the corresponding means and the variances given in the question, we can find that:

$$\Sigma_{(X_1, X_2), X_3} = \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (24)$$

Therefore, we can say that  $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  and  $X_3$  are independent.

- (c)  $Y = X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3$  and  $Z = X_2 - \frac{1}{2}X_4$

**Solution:**

Let's start by finding the distributions of Y and Z.

$$\mu_Y = E(Y) = E(X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3) \quad (25)$$

We can use the linearity property of the expected value:

$$\mu_Y = E(Y) = E(X_1) - \frac{1}{3} \cdot E(X_2) - \frac{1}{3} \cdot E(X_3) = 4 - \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 5 = \frac{16}{3} \quad (26)$$

We know that the  $Var(aX + bY) = a^2Var(X) + b^2Var(Y) - 2abCov(X, Y)$ . By using this fact, we can calculate the variance for our linear combination as in the following:

$$Var(Y) = Var(X_1) + \frac{1}{9} \cdot Var(X_2) + \frac{1}{9} \cdot Var(X_3) - \frac{2}{3} \cdot Cov(X_1, X_2) - \frac{2}{3} \cdot Cov(X_1, X_3) + \frac{2}{9} \cdot Cov(X_2, X_3) \quad (27)$$

$$Var(Y) = 3 + \frac{1}{9} \cdot 5 + \frac{1}{9} \cdot 2 - \frac{2}{3} \cdot (-2) - \frac{2}{3} \cdot 0 + \frac{2}{9} \cdot 0 = \frac{46}{9} \quad (28)$$

Therefore, we can say that the distribution of Y is:

$$Y \sim N\left(\frac{16}{3}, \frac{46}{9}\right) \quad (29)$$

Now, Let's find the distribution of the  $Z = X_2 - \frac{1}{2}X_4$ :

$$\mu_Z = E(Z) = E(X_2 - \frac{1}{2}X_4) \quad (30)$$

We can use the linearity property of the expected value:

$$\mu_Z = E(Z) = E(X_2) - \frac{1}{2} \cdot E(X_4) = 1 - \frac{1}{2} \cdot 2 = 0 \quad (31)$$

We know that the  $Var(aX + bY) = a^2Var(X) + b^2Var(Y) - 2abCov(X, Y)$ . By using this fact, we can calculate the variance for our linear combination as in the following:

$$Var(Z) = Var(X_2) + \frac{1}{4} \cdot Var(X_4) - Cov(X_2, X_4) = 5 + \frac{1}{4} \cdot 9 - 2 = \frac{21}{4} \quad (32)$$

Therefore, we can say that the distribution of Z is:

$$z \sim N(0, \frac{21}{4}) \quad (33)$$

Now, it is time to check the dependency of Y and Z: In order to check whether they are independent or not, we need to check whether the covariance of these two equal to 0 or not.

$$Cov(Y, Z) = Cov(X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3, X_2 - \frac{1}{2}X_4) \quad (34)$$

By using the linearity property of the covariance, we can calculate this in more simplified way:

$$Cov(Y, Z) = Cov(X_1, X_2 - \frac{1}{2}X_4) - \frac{1}{3}Cov(X_2, X_2 - \frac{1}{2}X_4) - \frac{1}{3}Cov(X_3, X_2 - \frac{1}{2}X_4) \quad (35)$$

Let's use the linearity property again to simplify each term again:

$$Cov(Y, Z) = Cov(X_1, X_2) - \frac{1}{2}Cov(X_1, X_4) - \frac{1}{3}Cov(X_2, X_2) + \frac{1}{6}Cov(X_2, X_4) - \frac{1}{3}Cov(X_3, X_2) + \frac{1}{6}Cov(X_3, X_4) \quad (36)$$

If we substitute the provided values in the question text, we will get:

$$Cov(Y, Z) = -2 - \frac{1}{2} \cdot 1 - \frac{1}{3} \cdot 5 + \frac{1}{6} \cdot 2 - \frac{1}{3} \cdot 0 + \frac{1}{6} \cdot -1 = -4 \quad (37)$$

Since the covariance value is not zero, the Y and Z values are not independent!

**(ii) Find the conditional distribution of  $X_2$  given  $X_1 = 2$  and  $X_4 = 10$ .**

**Solution:**

In order to answer this question, we can use the parts (i) and (iii) from question 2. Let's define some intermediate random variables and call them A and B so that  $A = X_2$  and  $B = \begin{pmatrix} X_1 \\ X_4 \end{pmatrix}$ . The conditional distribution of  $A = X_2$  given  $B = \begin{pmatrix} X_1 \\ X_4 \end{pmatrix}$  is given by:

$$A|B \sim N(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}) \quad (38)$$

As first step, let's find the mean values used in the equation by using  $\mu$  provided in the question:

$$\mu_A = \mu_2 = 1, \quad \mu_B = \begin{pmatrix} \mu_1 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \quad (39)$$

Now, let's find the covariance data from  $\Sigma$  given in the question:

$$\Sigma_{BB} = \begin{bmatrix} \Sigma_{11} & \Sigma_{14} \\ \Sigma_{41} & \Sigma_{44} \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 9 \end{bmatrix} \quad (40)$$

$$\Sigma_{AB} = [\Sigma_{21} \quad \Sigma_{24}] = [-2 \quad 2] \quad (41)$$

$$\Sigma_{AA} = \Sigma_{22} = 5 \quad (42)$$

The remaining step is to computing  $\Sigma_{BB}^{-1}$  by using the formula for the inverse of a 2x2 matrix:

$$\Sigma_{BB}^{-1} = \frac{1}{(3)(9) - (1)(1)} \begin{bmatrix} 9 & -1 \\ -1 & 3 \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 9 & -1 \\ -1 & 3 \end{bmatrix} \quad (43)$$

Now, we can calculate the expected value of conditional distribution as in (44).

$$E[X_2|X_1 = 2, X_4 = 10] = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(B - \mu_B) \quad (44)$$

Let's compute  $B - \mu_B$ :

$$B - \mu_B = \begin{pmatrix} 2 \\ 10 \end{pmatrix} - \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 8 \end{pmatrix} \quad (45)$$

Let's multiply  $\Sigma_{AB}$  we found in equation (41) with  $\Sigma_{BB}^{-1}$  we found in equation (43):

$$\Sigma_{AB}\Sigma_{BB}^{-1} = \begin{bmatrix} -2 & 2 \end{bmatrix} \cdot \frac{1}{26} \begin{bmatrix} 9 & -1 \\ -1 & 3 \end{bmatrix} \quad (46)$$

$$\Sigma_{AB}\Sigma_{BB}^{-1} = \frac{1}{26} [(-2)(9) + (2)(-1), (-2)(-1) + (2)(3)] \quad (47)$$

$$\Sigma_{AB}\Sigma_{BB}^{-1} = \frac{1}{26} [-18 - 2, 2 + 6] = \frac{1}{26} [-20 \quad 8] \quad (48)$$

Now, I am multiplying what I found in equation (48) with  $B - \mu_B$  we found in equation (45):

$$= \frac{1}{26} [-20 \quad 8] \begin{bmatrix} -2 \\ 8 \end{bmatrix} = \frac{1}{26} (-20 \times -2 + 8 \times 8) \quad (49)$$

$$= \frac{1}{26} (40 + 64) = \frac{104}{26} = 4 \quad (50)$$

So, since we know  $\mu_A$  value, let's sum it with what we find in here:

$$E[X_2|X_1 = 2, X_4 = 10] = 1 + 4 = 5 \quad (51)$$

The next step is calculating the conditional variance. We know that we can calculate it by using the following formula:

$$\text{Var}(X_2|X_1, X_4) = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \quad (52)$$

Since  $\Sigma_{BA} = \Sigma_{AB}^T$ :

$$\Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} = \frac{1}{26} [-20 \quad 8] \begin{bmatrix} -2 \\ 2 \end{bmatrix} \quad (53)$$

$$= \frac{1}{26} (-20 \times -2 + 8 \times 2) \quad (53)$$

$$= \frac{1}{26} (40 + 16) = \frac{56}{26} = \frac{28}{13} = \frac{65}{13} - \frac{28}{13} = \frac{37}{13} \quad (54)$$

Now, since we have both expected value and the variance of the conditional distribution, we can provide the distribution by saying:

$$X_2|(X_1 = 2, X_4 = 10) \sim N\left(5, \frac{37}{13}\right) \quad (55)$$

#### 4

The following data consists of the heights of 93 children born in Berkeley from Age 1 through 18. Use R to assess whether or not the data is distributed as a multivariate normal. Provide at least one plot to support your claim in each case.

```
install.packages("MASS")

## Installing package into '/home/ilkekas/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

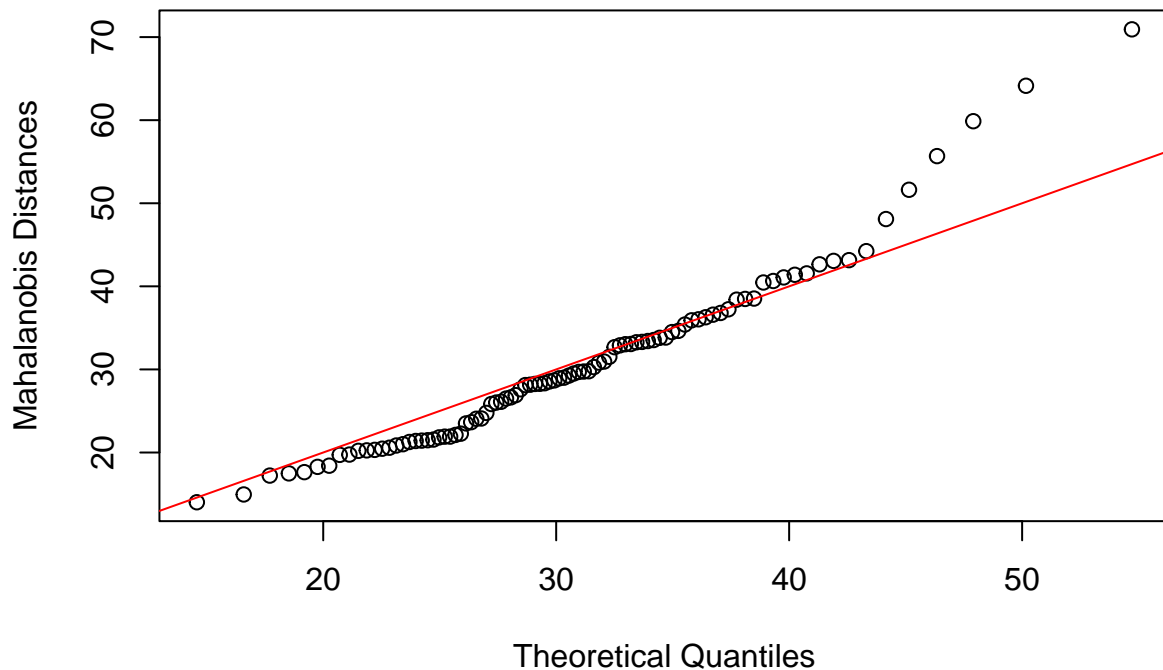
library(MASS)

# Read the data
data <- read.csv("Berkeley.csv")

# Compute Mahalanobis distances for a multivariate normality check
mahalanobis_distances <- mahalanobis(data, colMeans(data), cov(data))

# Q-Q Plot
qqplot(qchisq(ppoints(nrow(data))), df = ncol(data), mahalanobis_distances,
       main = "Chi-Square Q-Q Plot for Multivariate Normality",
       xlab = "Theoretical Quantiles", ylab = "Mahalanobis Distances")
abline(0, 1, col = "red")
```

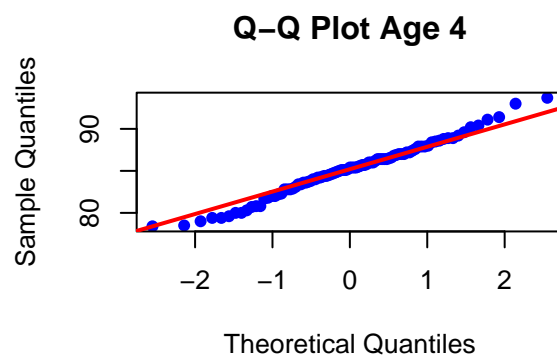
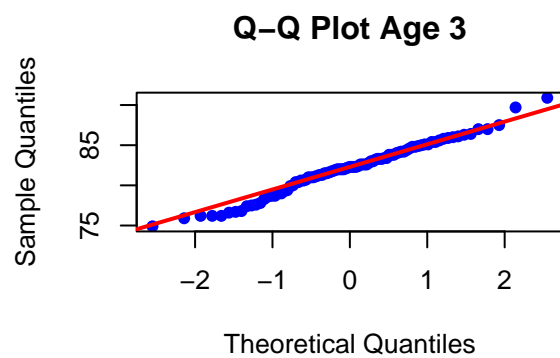
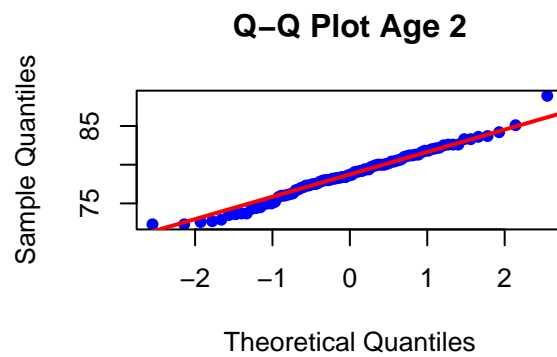
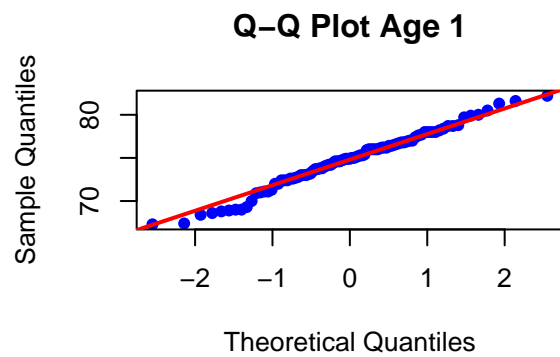
## Chi-Square Q-Q Plot for Multivariate Normality



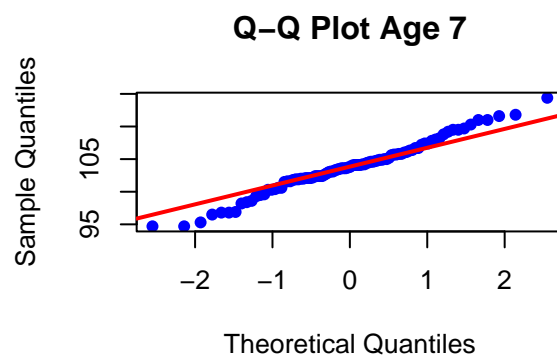
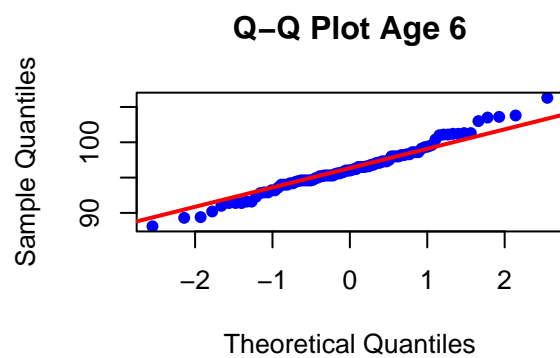
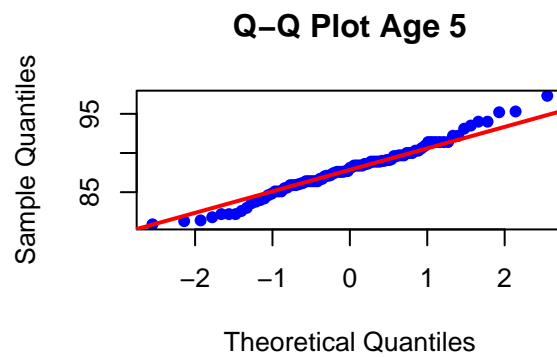
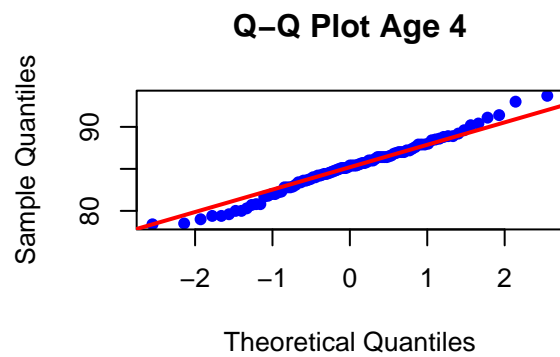
- In this plot, I used Mahalanobis distance to check if multivariate data fits a normal distribution. We know that this distance measures how far each data point is from the mean, considering the correlation between variables. Then, when I compared these distances to a Chi-Square distribution to see how well they match up. As seen in the plot, the data points line up nicely with the reference line in the middle of the plot. This shows us that the data behaves normally in that range. However, there are some data points stray from the line at the ends. It may show some outliers. Let's look at the individual random variables now:

```
par(mfrow = c(2, 2))
for (i in 1:4) {
  qqnorm(data[, i], main = paste("Q-Q Plot Age", i), col = "blue", pch = 16, cex = 1)
  qqline(data[, i], col = "red", lwd = 2)
}
```

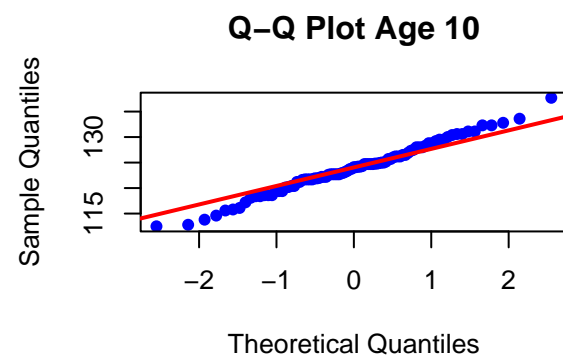
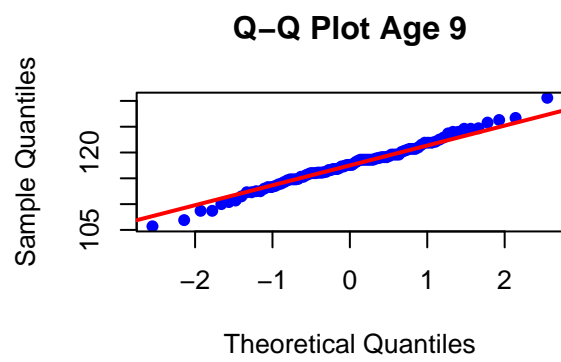
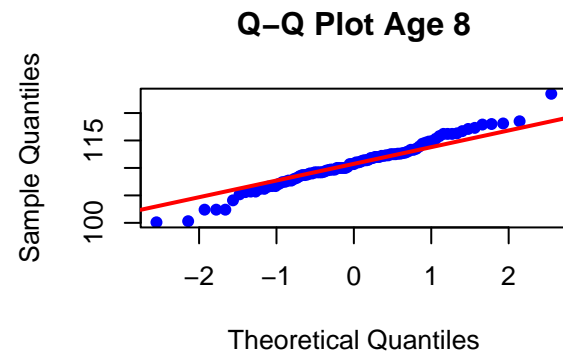
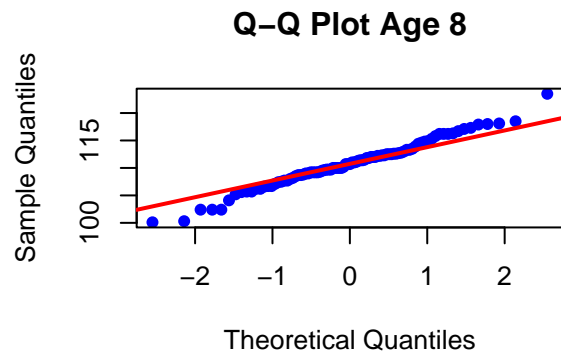


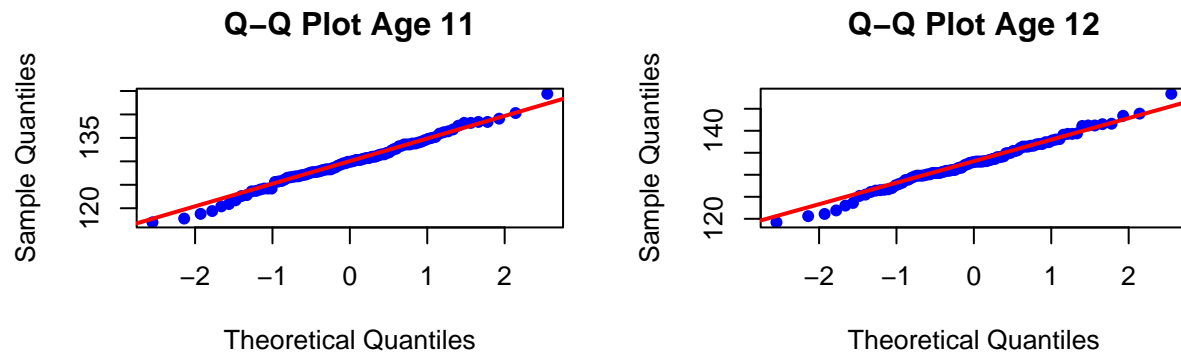


```
for (i in 4:8) {
  qqnorm(data[, i], main = paste("Q-Q Plot Age", i), col = "blue", pch = 16, cex = 1)
  qqline(data[, i], col = "red", lwd = 2)
}
```



```
for (i in 8:12) {
  qqnorm(data[, i], main = paste("Q-Q Plot Age", i), col = "blue", pch = 16, cex = 1)
  qqline(data[, i], col = "red", lwd = 2)
}
```





- Since we know that any subset of the multivariate random vector also have multivariate distribution, I wanted to look at the marginals ( since they also follow this rule). In here, I only printed 12 of them. These Q-Q plots show that for most ages, the data's distribution closely follows a normal distribution as most points align well with the red line. Especially near the center. However, deviations are visible in the tails for several age groups. This shows thatt the data might not be perfectly normal at the extreme values. Maybe some outliers or skewness. Overall, the data mostly looks normal, but the small differences at the ends could affect some kinds of statistical tests, especially those that don't handle outliers or uneven data well.