Multivariate Statistics Questions and Solutions

Jules

September 13, 2025

Contents

1 Basics of Multivariate Data

Question 1

Question: Define a multivariate data sample and the sample mean vector.

Solution: A multivariate data sample consists of n observations on p variables. We can represent this data as a $n \times p$ matrix X, where x_{ij} is the i-th observation of the j-th variable. The sample mean vector is a $p \times 1$ vector $\bar{\mathbf{x}}$ where each element $\bar{x}_j$ is the average of the n observations for the j-th variable.

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$$

Question 2

Question: Given the following dataset with 2 variables:

$$X = \begin{pmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{pmatrix}$$

Calculate the sample mean vector.

Solution: The sample mean vector $\bar{\mathbf{x}}$ is calculated as:

$$\bar{x}_1 = \frac{2+4+6}{3} = 4$$

$$\bar{x}_2 = \frac{3+5+7}{3} = 5$$

So, the sample mean vector is:

$$\bar{\mathbf{x}} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

Question 3

Question: Define the sample variance-covariance matrix and the sample correlation matrix. Explain the relationship between them.

Solution: The sample variance-covariance matrix, denoted by S, is a $p \times p$ symmetric matrix where the diagonal elements s_{jj} are the variances of each variable and the off-diagonal elements s_{jk} are the covariances between variables j and k.

$$S = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

The sample correlation matrix, R, is a $p \times p$ matrix where the elements r_{jk} are the sample correlation coefficients between variables j and k.

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}}$$

The relationship is $R = D^{-1/2} S D^{-1/2}$, where D is a diagonal matrix of the variances from S.

Question 4

Question: For the dataset in Question 2, calculate the sample variance-covariance matrix.

Solution: First, we calculate the deviations from the mean:

$$X - \bar{\mathbf{x}} = \begin{pmatrix} 2-4 & 3-5 \\ 4-4 & 5-5 \\ 6-4 & 7-5 \end{pmatrix} = \begin{pmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{pmatrix}$$

The sum of squared products matrix is:

$$(X - \bar{\mathbf{x}})^T (X - \bar{\mathbf{x}}) = \begin{pmatrix} (-2)^2 + 0^2 + 2^2 & (-2)(-2) + 0(0) + 2(2) \\ (-2)(-2) + 0(0) + 2(2) & (-2)^2 + 0^2 + 2^2 \end{pmatrix} = \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix}$$

The variance-covariance matrix S is this matrix divided by $n-1=2$:

$$S = \frac{1}{2}\begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$$

Question 5

Question: From the result of Question 4, calculate the sample correlation matrix.

Solution: We have $s_{11}=4$, $s_{22}=4$, and $s_{12}=4$.

$$r_{11} = r_{22} = 1$$

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}} = \frac{4}{\sqrt{4\cdot 4}} = \frac{4}{4} = 1$$

The correlation matrix is:

$$R = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Question 6

Question: What is a feature space? How do you visualize a multivariate data sample in it?

Solution: A feature space is a p-dimensional space where each dimension corresponds to one of the p variables (features) of the dataset. Each observation $\mathbf{x}_i$ is represented as a point in this space. For $p=2$ or $p=3$, we can create a scatter plot of the n points. For $p>3$, we can use techniques like scatter plot matrices to visualize pairs of variables.

Question 7

Question: Define the Mahalanobis distance. How does it differ from the Euclidean distance?

Solution: The Mahalanobis distance between two points $\mathbf{x}$ and $\mathbf{y}$ in a p-dimensional space with covariance matrix S is:

$$D_M(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T S^{-1}(\mathbf{x}-\mathbf{y})}$$

It differs from Euclidean distance by accounting for the covariance among variables. It is scale-invariant and corrects for correlation.

Question 8

Question: Given a mean vector $\bar{\mathbf{x}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, a covariance matrix $S = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and a point $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, calculate the Mahalanobis distance from $\mathbf{x}_0$ to the mean.

Solution: First, find the inverse of S:

$$S^{-1} = \frac{1}{1-0.25}\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} = \frac{4}{3}\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$$

The Mahalanobis distance is:

$$D_M^2 = (\mathbf{x}_0-\bar{\mathbf{x}})^T S^{-1}(\mathbf{x}_0-\bar{\mathbf{x}}) = \begin{pmatrix} 1 & 1 \end{pmatrix} \frac{4}{3}\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{4}{3}\begin{pmatrix} 1 & 1 \end{pmatrix}\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = \frac{4}{3}(0.5+0.5) = \frac{4}{3}$$

$$D_M = \sqrt{4/3} \approx 1.1547$$

Question 9

Question: Describe the geometric shape of points that are at a constant statistical distance from the mean.

Solution: The set of points $\mathbf{x}$ that are at a constant Mahalanobis distance c from the mean vector $\bar{\mathbf{x}}$ forms an ellipsoid in the p-dimensional space. The equation for this ellipsoid is:

$$(\mathbf{x}-\bar{\mathbf{x}})^T S^{-1}(\mathbf{x}-\bar{\mathbf{x}}) = c^2$$

The center of the ellipsoid is $\bar{\mathbf{x}}$, and its axes are determined by the eigenvectors and eigenvalues of the covariance matrix S.

Question 10

Question: Write down the equation for an ellipsoid of constant statistical distance for a 2-dimensional case with a diagonal covariance matrix $S = \begin{pmatrix} s_{11} & 0 \\ 0 & s_{22} \end{pmatrix}$.

Solution: For a diagonal covariance matrix, $S^{-1} = \begin{pmatrix} 1/s_{11} & 0 \\ 0 & 1/s_{22} \end{pmatrix}$. The equation for the ellipsoid is:

$$(\mathbf{x} - \bar{\mathbf{x}})^T S^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = \frac{(x_1 - \bar{x}_1)^2}{s_{11}} + \frac{(x_2 - \bar{x}_2)^2}{s_{22}} = c^2$$

This is the standard equation of an ellipse centered at $(\bar{x}_1, \bar{x}_2)$ with axes parallel to the coordinate axes.

Question 11

Question: Define the total variance and the generalized variance. What does a generalized variance of zero imply?

Solution: The total variance is the sum of the variances of all variables, which is the trace of the covariance matrix S.

$$\text{Total Variance} = \text{tr}(S) = \sum_{j=1}^{p} s_{jj}$$

The generalized variance is the determinant of the covariance matrix S.

$$\text{Generalized Variance} = |S|$$

A generalized variance of zero implies that the covariance matrix is singular, which means there is at least one linear dependency among the variables.

Question 12

Question: Calculate the total variance and generalized variance for the covariance matrix from Question 4.

Solution: The covariance matrix is $S = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$. Total variance = tr(S) = 4 + 4 = 8. Generalized variance = —S— = (4)(4) - (4)(4) = 0. The zero generalized variance indicates that the variables are perfectly correlated.

Question 13

Question: Show that the sample correlation matrix is symmetric and has 1s on the diagonal.

Solution: The correlation coefficient $r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj} s_{kk}}}$. Symmetry: $r_{kj} = \frac{s_{kj}}{\sqrt{s_{kk} s_{jj}}} = \frac{s_{jk}}{\sqrt{s_{jj} s_{kk}}} = r_{jk}$ since $s_{kj} = s_{jk}$. Diagonal elements: $r_{jj} = \frac{s_{jj}}{\sqrt{s_{jj} s_{jj}}} = \frac{s_{jj}}{s_{jj}} = 1$.

Question 14

Question: If a variable is added to a dataset, what happens to the dimensions of the sample mean vector and sample variance-covariance matrix?

Solution: If we add a variable, the number of variables p becomes $p + 1$. The sample mean vector $\bar{\mathbf{x}}$ will have its dimension increase from $p \times 1$ to $(p + 1) \times 1$. The sample variance-covariance matrix S will have its dimensions increase from $p \times p$ to $(p + 1) \times (p + 1)$.

Question 15

Question: Consider a dataset with 3 variables. The covariance matrix is given by:

$$S = \begin{pmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{pmatrix}$$

Find the correlation between variable 1 and 3.

Solution: We need to find r_{13}. We have $s_{13} = 4$, $s_{11} = 25$, and $s_{33} = 9$.

$$r_{13} = \frac{s_{13}}{\sqrt{s_{11}s_{33}}} = \frac{4}{\sqrt{25 \cdot 9}} = \frac{4}{\sqrt{225}} = \frac{4}{15} \approx 0.2667$$

The correlation between variable 1 and 3 is approximately 0.2667.

2 Geometric Interpretation

Question 1

Question: Explain the concept of a sample space in the context of multivariate analysis. How does it differ from the feature space?

Solution: In multivariate analysis, the sample space is an n-dimensional space where n is the number of observations. Each of the p variables can be represented as a vector in this space. So, we have p vectors in an n-dimensional space. This is in contrast to the feature space, which is a p-dimensional space where each of the n observations is represented as a point.

Question 2

Question: What is a vector projection? Provide the formula for projecting a vector $\mathbf{y}$ onto a vector $\mathbf{x}$.

Solution: A vector projection of a vector $\mathbf{y}$ onto a vector $\mathbf{x}$ is the component of $\mathbf{y}$ that lies in the direction of $\mathbf{x}$. The formula is:

$$\text{proj}_{\mathbf{x}}\mathbf{y} = \frac{\mathbf{y}^T\mathbf{x}}{\mathbf{x}^T\mathbf{x}}\mathbf{x}$$

This gives a vector in the direction of $\mathbf{x}$. The scalar value $\frac{\mathbf{y}^T\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is the coordinate of the projection.

Question 3

Question: Given vectors $\mathbf{y} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, find the projection of $\mathbf{y}$ onto $\mathbf{x}$.

Solution: We use the formula for projection:

$$\mathbf{y}^T\mathbf{x} = (3)(1) + (4)(1) = 7$$

$$\mathbf{x}^T\mathbf{x} = (1)^2 + (1)^2 = 2$$

$$\text{proj}_{\mathbf{x}}\mathbf{y} = \frac{7}{2}\mathbf{x} = \frac{7}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3.5 \\ 3.5 \end{pmatrix}$$

Question 4

Question: How can we interpret the length of a vector in Euclidean space? What does the squared length of a mean-centered vector represent?

Solution: The length of a vector $\mathbf{x} = (x_1, ..., x_n)^T$ is given by $L_{\mathbf{x}} = \sqrt{\sum_{i=1}^n x_i^2}$. A mean-centered vector $\mathbf{d} = \mathbf{x} - \bar{x}\mathbf{1}$ has elements $d_i = x_i - \bar{x}$. The squared length of this vector is:

$$L_{\mathbf{d}}^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

This is $(n-1)$ times the sample variance of the variable x.

Question 5

Question: Calculate the length of the vector $\mathbf{d} = \mathbf{y} - \bar{y}\mathbf{1}$, where $\mathbf{y} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Solution: First, calculate the mean $\bar{y} = (1 + 2 + 3)/3 = 2$. The mean-centered vector is:

$$\mathbf{d} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - 2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

The length of $\mathbf{d}$ is:

$$L_{\mathbf{d}} = \sqrt{(-1)^2 + 0^2 + 1^2} = \sqrt{2}$$

Question 6

Question: Define the cosine of the angle between two vectors. What do values of 1, 0, and -1 signify?

Solution: The cosine of the angle θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is:

$$\cos(\theta) = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$$

- $\cos(\theta) = 1$ means the vectors point in the same direction ($\theta = 0°$). - $\cos(\theta) = 0$ means the vectors are orthogonal ($\theta = 90°$). - $\cos(\theta) = -1$ means the vectors point in opposite directions ($\theta = 180°$).

Question 7

Question: Find the cosine of the angle between the two vectors from Question 3, $\mathbf{y} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Solution:

$$\mathbf{x}^T\mathbf{y} = 7$$
$$\|\mathbf{x}\| = \sqrt{1^2 + 1^2} = \sqrt{2}$$
$$\|\mathbf{y}\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$$
$$\cos(\theta) = \frac{7}{5\sqrt{2}} \approx 0.9899$$

Question 8

Question: How does the concept of cosine angle relate to the sample correlation coefficient?

Solution: The sample correlation coefficient r between two variables x and y is the cosine of the angle between their mean-centered vectors in the n-dimensional sample space. If $\mathbf{d}_x = \mathbf{x} - \bar{x}\mathbf{1}$ and $\mathbf{d}_y = \mathbf{y} - \bar{y}\mathbf{1}$, then:

$$r_{xy} = \frac{\mathbf{d}_x^T\mathbf{d}_y}{\|\mathbf{d}_x\|\|\mathbf{d}_y\|} = \cos(\theta)$$

Question 9

Question: Explain how a linear combination of variables can be viewed as a projection.

Solution: Consider a linear combination of p variables, $c_1x_1 + ... + c_px_p$. In the sample space, we have p vectors $\mathbf{x}_1, ..., \mathbf{x}_p$. The linear combination forms a new vector $\mathbf{z} = c_1\mathbf{x}_1 + ... + c_p\mathbf{x}_p$. This vector $\mathbf{z}$ lies in the subspace spanned by the original variable vectors. Each observation's value for this new variable, z_i, is its value on the new axis defined by the linear combination. This can be seen as a projection of the observation points onto this new axis.

Question 10

Question: Project the first observation vector $\mathbf{x}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ from the feature space onto the vector $\mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Solution: This is a projection in the feature space.

$$\mathbf{x}_1^T\mathbf{v} = (2)(1) + (3)(-1) = -1$$
$$\mathbf{v}^T\mathbf{v} = (1)^2 + (-1)^2 = 2$$
$$\text{proj}_{\mathbf{v}}\mathbf{x}_1 = \frac{-1}{2}\mathbf{v} = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$$

Question 11

Question: What is the geometric interpretation of the sample variance?

Solution: Geometrically, the sample variance of a variable is proportional to the squared length of its mean-centered vector in the sample space. A larger variance means the vector is longer, indicating more spread in the data.

$$s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{n-1} \|\mathbf{x} - \bar{x}\mathbf{1}\|^2$$

Question 12

Question: What is the geometric interpretation of the sample covariance?

Solution: The sample covariance between two variables x and y is proportional to the dot product of their mean-centered vectors.

$$s_{xy} = \frac{1}{n-1} (\mathbf{x} - \bar{x}\mathbf{1})^T (\mathbf{y} - \bar{y}\mathbf{1})$$

The sign of the covariance is determined by the angle between these vectors. If the angle is less than 90 degrees, the covariance is positive. If it's greater than 90 degrees, it's negative.

Question 13

Question: Given two mean-centered vectors $\mathbf{d}_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ and $\mathbf{d}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, calculate their dot product. What does this imply about their sample covariance?

Solution:

$$\mathbf{d}_1^T \mathbf{d}_2 = (-1)(-1) + (0)(1) + (1)(0) = 1$$

Since the dot product is positive, the sample covariance between the two corresponding variables is positive. The sample covariance would be $1/(n-1) = 1/2$.

Question 14

Question: Describe how you would find a projection of a data set that maximizes the variance of the projected points.

Solution: This is the core idea of Principal Component Analysis (PCA). We want to find a direction (a unit vector $\mathbf{a}$) such that when we project the data points onto this direction, the variance of the projected points is maximized. The projected values are given by $X\mathbf{a}$. The variance of these projected values is proportional to $\mathbf{a}^T S \mathbf{a}$, where S is the covariance matrix. To maximize this quantity subject to $\|\mathbf{a}\| = 1$, we find the eigenvector of S corresponding to the largest eigenvalue. This eigenvector is the direction of maximum variance.

Question 15

Question: If two vectors representing two variables are orthogonal in the sample space after being mean-centered, what does this imply about their correlation?

Solution: If the mean-centered vectors $\mathbf{d}_x$ and $\mathbf{d}_y$ are orthogonal, their dot product is zero: $\mathbf{d}_x^T \mathbf{d}_y = 0$. Since the sample correlation is the cosine of the angle between these vectors, and the angle is 90 degrees, the correlation is $\cos(90°) = 0$. This means the two variables are uncorrelated.

3 Properties of Random Vectors

Question 1

Question: Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors and A, B be matrices of constants. State the property for the expectation of a linear combination of random vectors, $E(A\mathbf{X} + B\mathbf{Y})$.

Solution: The expectation of a linear combination of random vectors is the linear combination of their expectations.

$$E(A\mathbf{X} + B\mathbf{Y}) = AE(\mathbf{X}) + BE(\mathbf{Y})$$

This property holds assuming the dimensions of the matrices and vectors are compatible for addition and multiplication.

Question 2

Question: Let $\mathbf{X}$ be a $p \times 1$ random vector with mean $E(\mathbf{X}) = \boldsymbol{\mu}$. Let A be a $q \times p$ matrix of constants and $\mathbf{b}$ be a $q \times 1$ vector of constants. Show that $E(A\mathbf{X} + \mathbf{b}) = A\boldsymbol{\mu} + \mathbf{b}$.

Solution: Using the definition of expectation for a vector:

$$E(A\mathbf{X} + \mathbf{b}) = \int \cdots \int (A\mathbf{x} + \mathbf{b}) f(\mathbf{x}) d\mathbf{x}$$

where $f(\mathbf{x})$ is the joint pdf of $\mathbf{X}$.

$$= \int \cdots \int A\mathbf{x} f(\mathbf{x}) d\mathbf{x} + \int \cdots \int \mathbf{b} f(\mathbf{x}) d\mathbf{x}$$

$$= A\left(\int \cdots \int \mathbf{x} f(\mathbf{x}) d\mathbf{x}\right) + \mathbf{b}\left(\int \cdots \int f(\mathbf{x}) d\mathbf{x}\right)$$

Since $\int \cdots \int \mathbf{x} f(\mathbf{x}) d\mathbf{x} = E(\mathbf{X}) = \boldsymbol{\mu}$ and $\int \cdots \int f(\mathbf{x}) d\mathbf{x} = 1$, we have:

$$E(A\mathbf{X} + \mathbf{b}) = A\boldsymbol{\mu} + \mathbf{b}$$

Question 3

Question: Define the covariance matrix of a random vector $\mathbf{X}$ with mean $\boldsymbol{\mu}$.

Solution: The covariance matrix of a random vector $\mathbf{X}$, denoted by Σ or $\mathrm{Cov}(\mathbf{X})$, is a $p \times p$ matrix defined as:

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

The (i, j)-th element of Σ is the covariance between X_i and X_j, and the (i, i)-th element is the variance of X_i.

Question 4

Question: Show that $\mathrm{Cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$.

Solution: Starting from the definition:

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

$$= E[\mathbf{X}\mathbf{X}^T - \mathbf{X}\boldsymbol{\mu}^T - \boldsymbol{\mu}\mathbf{X}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T]$$

Using the linearity of expectation:

$$= E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X}\boldsymbol{\mu}^T) - E(\boldsymbol{\mu}\mathbf{X}^T) + E(\boldsymbol{\mu}\boldsymbol{\mu}^T)$$

Since $\boldsymbol{\mu}$ is a constant vector:

$$= E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X})\boldsymbol{\mu}^T - \boldsymbol{\mu}E(\mathbf{X}^T) + \boldsymbol{\mu}\boldsymbol{\mu}^T$$

$$= E(\mathbf{X}\mathbf{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T = E(\mathbf{X}\mathbf{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

Question 5

Question: Let a random vector $\mathbf{X}$ be partitioned into two sub-vectors $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$. Describe the structure of the mean vector $\boldsymbol{\mu}$ and the covariance matrix Σ in terms of the sub-vectors.

Solution: The mean vector $\boldsymbol{\mu}$ is partitioned similarly:

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{pmatrix} E(\mathbf{X}_1) \\ E(\mathbf{X}_2) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$$

The covariance matrix Σ is partitioned into blocks:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11} = \text{Cov}(\mathbf{X}_1)$, $\Sigma_{22} = \text{Cov}(\mathbf{X}_2)$, and $\Sigma_{12} = \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \Sigma_{21}^T$.

Question 6

Question: Let a random vector $\mathbf{X} = (X_1, X_2, X_3)^T$ have a mean vector $\boldsymbol{\mu} = (2, 3, 5)^T$. Partition the vector into $\mathbf{X}_1 = (X_1, X_2)^T$ and $\mathbf{X}_2 = (X_3)$. What are the corresponding partitioned mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$?

Solution: The partitioned mean vectors are simply the corresponding parts of the original mean vector:

$$\boldsymbol{\mu}_1 = E(\mathbf{X}_1) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$\boldsymbol{\mu}_2 = E(\mathbf{X}_2) = (5)$$

Question 7

Question: Define statistical independence for two random vectors $\mathbf{X}$ and $\mathbf{Y}$. What does this imply about their joint probability density function?

Solution: Two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are statistically independent if their joint probability density function (pdf) can be factored into the product of their marginal pdfs.

$$f(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})$$

for all values of $\mathbf{x}$ and $\mathbf{y}$.

Question 8

Question: If two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are independent, what can be said about their cross-covariance matrix, $\text{Cov}(\mathbf{X}, \mathbf{Y})$?

Solution: If $\mathbf{X}$ and $\mathbf{Y}$ are independent, their cross-covariance matrix is a zero matrix.

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T] = \mathbf{0}$$

Question 9

Question: If $\mathbf{X}$ and $\mathbf{Y}$ are independent random vectors, show that $\text{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$. Does the converse hold? Explain.

Solution: If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $E(\mathbf{X}\mathbf{Y}^T) = E(\mathbf{X})E(\mathbf{Y}^T) = \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T$.

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X}\mathbf{Y}^T) - \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T = \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T - \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T = \mathbf{0}$$

The converse is not true in general. Zero covariance implies no linear relationship, but there could still be a non-linear relationship, meaning the vectors are not independent. The exception is for multivariate normal distributions, where zero covariance does imply independence.

Question 10

Question: Consider a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from a population with mean $\boldsymbol{\mu}$ and covariance Σ. What is the expected value of the sample mean vector $\bar{\mathbf{X}}$?

Solution: The expected value of the sample mean vector $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$ is the population mean vector $\boldsymbol{\mu}$.

$$E(\bar{\mathbf{X}}) = \boldsymbol{\mu}$$

Question 11

Question: Prove that $E(\bar{\mathbf{X}}) = \boldsymbol{\mu}$.

Solution: Using the linearity of expectation:

$$E(\bar{\mathbf{X}}) = E\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\right) = \frac{1}{n}E\left(\sum_{i=1}^n \mathbf{X}_i\right) = \frac{1}{n}\sum_{i=1}^n E(\mathbf{X}_i)$$

Since each $\mathbf{X}_i$ is from the same population, $E(\mathbf{X}_i) = \boldsymbol{\mu}$ for all i.

$$E(\bar{\mathbf{X}}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\mu} = \frac{1}{n}(n\boldsymbol{\mu}) = \boldsymbol{\mu}$$

Question 12

Question: What is the covariance matrix of the sample mean vector, $\mathrm{Cov}(\bar{\mathbf{X}})$?

Solution: The covariance matrix of the sample mean vector $\bar{\mathbf{X}}$ is the population covariance matrix Σ divided by the sample size n.

$$\mathrm{Cov}(\bar{\mathbf{X}}) = \frac{1}{n}\Sigma$$

Question 13

Question: Prove that $\mathrm{Cov}(\bar{\mathbf{X}}) = \frac{1}{n}\Sigma$.

Solution:

$$\mathrm{Cov}(\bar{\mathbf{X}}) = \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\right) = \frac{1}{n^2}\mathrm{Cov}\left(\sum_{i=1}^n \mathbf{X}_i\right)$$

Since the observations are independent, the covariance of the sum is the sum of the covariances:

$$= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Cov}(\mathbf{X}_i)$$

Since $\mathrm{Cov}(\mathbf{X}_i) = \Sigma$ for all i:

$$= \frac{1}{n^2}(n\Sigma) = \frac{1}{n}\Sigma$$

Question 14

Question: What is the expected value of the sample covariance matrix S?

Solution: The sample covariance matrix $S = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ is an unbiased estimator of the population covariance matrix Σ. Therefore, its expected value is Σ.

$$E(S) = \Sigma$$

Question 15

Question: Show that the sample covariance matrix S is an unbiased estimator of the population covariance matrix Σ, i.e., $E(S) = \Sigma$.

Solution: This proof is more involved. First, we write:

$$(n-1)S = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T = \sum \mathbf{X}_i\mathbf{X}_i^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T$$

Taking the expectation:

$$(n-1)E(S) = \sum E(\mathbf{X}_i\mathbf{X}_i^T) - nE(\bar{\mathbf{X}}\bar{\mathbf{X}}^T)$$

We know $E(\mathbf{X}_i\mathbf{X}_i^T) = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T$ and $E(\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = \text{Cov}(\bar{\mathbf{X}}) + E(\bar{\mathbf{X}})E(\bar{\mathbf{X}})^T = \frac{1}{n}\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T$.

$$(n-1)E(S) = \sum_{i=1}^{n}(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T) - n(\frac{1}{n}\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T)$$

$$= (n\Sigma + n\boldsymbol{\mu}\boldsymbol{\mu}^T) - (\Sigma + n\boldsymbol{\mu}\boldsymbol{\mu}^T)$$
$$= n\Sigma - \Sigma = (n-1)\Sigma$$

Therefore, $E(S) = \Sigma$.

4 Multivariate Normal Distribution

Question 1

Question: Write down the probability density function (pdf) of a p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Solution: The pdf for a random vector $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

This is valid for $\mathbf{x} \in \mathbb{R}^p$, and it requires that the covariance matrix Σ be positive definite (and thus invertible).

Question 2

Question: What are the main properties of the multivariate normal distribution? List at least three.

Solution: 1. **Linear combinations are normal:** If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then any linear combination $A\mathbf{X} + \mathbf{b}$ is also normally distributed. 2. **Marginal distributions are normal:** All subsets of the components of $\mathbf{X}$ have multivariate normal distributions. 3. **Zero covariance implies independence:** If two subsets of components of $\mathbf{X}$ have a zero covariance matrix, then they are statistically independent. 4. **Conditional distributions are normal:** The conditional distribution of one subset of components, given the values of another subset, is also multivariate normal.

Question 3

Question: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and let A be a $q \times p$ matrix of constants. Show that the linear combination $A\mathbf{X}$ is also multivariate normal. What are its mean and covariance matrix?

Solution: The resulting distribution of $\mathbf{Y} = A\mathbf{X}$ is multivariate normal. We can find its mean and covariance as follows: Mean:

$$E(\mathbf{Y}) = E(A\mathbf{X}) = AE(\mathbf{X}) = A\boldsymbol{\mu}$$

Covariance:

$$\text{Cov}(\mathbf{Y}) = \text{Cov}(A\mathbf{X}) = A\text{Cov}(\mathbf{X})A^T = A\Sigma A^T$$

So, $\mathbf{Y} = A\mathbf{X} \sim N_q(A\boldsymbol{\mu}, A\Sigma A^T)$. A formal proof involves using moment generating functions or characteristic functions.

Question 4

Question: Let $\mathbf{X} \sim N_2\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}4 & 1\\1 & 9\end{pmatrix}\right)$. Let $A = \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}$. Find the distribution of $A\mathbf{X}$.

Solution: Let $\mathbf{Y} = A\mathbf{X}$. The distribution of $\mathbf{Y}$ is normal with mean $A\boldsymbol{\mu}$ and covariance $A\Sigma A^T$.

$$A\boldsymbol{\mu} = \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}3\\-1\end{pmatrix}$$

$$A\Sigma A^T = \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}4 & 1\\1 & 9\end{pmatrix}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}^T = \begin{pmatrix}5 & 10\\3 & -8\end{pmatrix}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} = \begin{pmatrix}15 & -5\\-5 & 11\end{pmatrix}$$

So, $A\mathbf{X} \sim N_2\left(\begin{pmatrix}3\\-1\end{pmatrix}, \begin{pmatrix}15 & -5\\-5 & 11\end{pmatrix}\right)$.

Question 5

Question: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$. If a subset of components of $\mathbf{X}$ has zero covariance with another subset, what does this imply about the independence of these subsets?

Solution: For the multivariate normal distribution, zero covariance is a necessary and sufficient condition for independence. If we partition $\mathbf{X}$ into $\mathbf{X}_1$ and $\mathbf{X}_2$ and their cross-covariance Σ_{12} is a zero matrix, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are statistically independent.

Question 6

Question: Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ be a partitioned multivariate normal random vector with corresponding partitioned mean $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$ and covariance $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. What is the marginal distribution of $\mathbf{X}_1$?

Solution: One of the key properties of the MVN is that its marginal distributions are also normal. The marginal distribution of $\mathbf{X}_1$ is obtained by simply taking the corresponding blocks of the mean vector and covariance matrix.

$$\mathbf{X}_1 \sim N_{p_1}(\boldsymbol{\mu}_1, \Sigma_{11})$$

where p_1 is the dimension of $\mathbf{X}_1$.

Question 7

Question: Given $\mathbf{X} \sim N_3 \left(\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 5 & 2 & 1 \\ 2 & 4 & -1 \\ 1 & -1 & 3 \end{pmatrix} \right)$, find the marginal distribution of $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$.

Solution: We select the components corresponding to X_1 and X_3 from the mean vector and covariance matrix. Mean: $\boldsymbol{\mu}_{1,3} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Covariance matrix: $\Sigma_{1,3} = \begin{pmatrix} 5 & 1 \\ 1 & 3 \end{pmatrix}$. So, $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 5 & 1 \\ 1 & 3 \end{pmatrix} \right)$.

Question 8

Question: State the formula for the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ for a partitioned multivariate normal vector.

Solution: The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is also multivariate normal, with: Mean: $E(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ Covariance: $\text{Cov}(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ Note that the conditional covariance does not depend on the value of $\mathbf{x}_2$.

Question 9

Question: Let $\mathbf{X} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \right)$. Find the conditional distribution of X_1 given $X_2 = 1$.

Solution: Here, $\mathbf{X}_1 = X_1$, $\mathbf{X}_2 = X_2$, $\boldsymbol{\mu}_1 = 0, \boldsymbol{\mu}_2 = 0$, $\Sigma_{11} = 4, \Sigma_{12} = 2, \Sigma_{22} = 2$. Conditional Mean: $E(X_1 | X_2 = 1) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = 0 + 2 \cdot (1/2) \cdot (1 - 0) = 1$. Conditional Variance: $\text{Var}(X_1 | X_2 = 1) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 4 - 2 \cdot (1/2) \cdot 2 = 4 - 2 = 2$. So, $(X_1 | X_2 = 1) \sim N(1, 2)$.

Question 10

Question: What is the distribution of the quadratic form $(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})$ when $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$?

Solution: The quadratic form $(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})$ follows a chi-square distribution with p degrees of freedom.

$$(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_p^2$$

Question 11

Question: Explain how the result from Question 10 is used to construct a confidence ellipsoid for the population mean vector $\boldsymbol{\mu}$.

Solution: From the central limit theorem, the sample mean $\bar{\mathbf{X}}$ is approximately $N_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma)$. Thus, $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T S^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$ is approximately χ_p^2 for large n. A $100(1-\alpha)\%$ confidence ellipsoid for $\boldsymbol{\mu}$ is the set of all $\boldsymbol{\mu}$ that satisfy:

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T S^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq \chi_{p,\alpha}^2$$

where $\chi_{p,\alpha}^2$ is the upper (100α)th percentile of the χ_p^2 distribution.

Question 12

Question: For a bivariate normal distribution ($p = 2$), what is the equation for a 95

Solution: The equation for a 95

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq \chi^2_{2,0.05}$$

where n is the sample size, $\bar{\mathbf{x}}$ is the sample mean vector, S is the sample covariance matrix, and $\chi^2_{2,0.05} \approx 5.99$ is the critical value from a chi-square distribution with 2 degrees of freedom.

Question 13

Question: Show that any linear combination of the components of a multivariate normal vector $\mathbf{X}$, say $\mathbf{a}^T\mathbf{X}$, follows a univariate normal distribution.

Solution: This is a special case of the property in Question 3, where A is a $1 \times p$ matrix (a row vector $\mathbf{a}^T$). Let $Y = \mathbf{a}^T\mathbf{X}$. The mean of Y is $E(Y) = E(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T E(\mathbf{X}) = \mathbf{a}^T\boldsymbol{\mu}$. The variance of Y is $\text{Var}(Y) = \text{Cov}(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\text{Cov}(\mathbf{X})\mathbf{a} = \mathbf{a}^T\Sigma\mathbf{a}$. Since Y is a scalar, its distribution is univariate normal: $Y \sim N(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\Sigma\mathbf{a})$.

Question 14

Question: If all marginal distributions of a random vector $\mathbf{X}$ are normal, is $\mathbf{X}$ necessarily multivariate normal? Explain.

Solution: No. If $\mathbf{X}$ is multivariate normal, then all its marginals are normal. However, the converse is not true. It is possible to construct a joint distribution where the marginals are normal, but the joint distribution is not multivariate normal. A key feature of the MVN distribution is that the dependency structure is fully captured by the covariance matrix, which is not true for all distributions.

Question 15

Question: Describe the shape and orientation of the contours of constant density for a multivariate normal distribution. What determines them?

Solution: The contours of constant density for a multivariate normal distribution are ellipsoids. - The center of the ellipsoids is the mean vector $\boldsymbol{\mu}$. - The orientation of the ellipsoids is determined by the eigenvectors of the covariance matrix Σ. The eigenvectors are the principal axes of the ellipsoids. - The lengths of the axes are determined by the eigenvalues of Σ. Larger eigenvalues correspond to longer axes, indicating greater variance in the direction of the corresponding eigenvector.

5 Inference

Question 1

Question: For a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from a population with probability density function (pdf) $f(\mathbf{x}|\boldsymbol{\theta})$, write down the likelihood function $L(\boldsymbol{\theta})$.

Solution: The likelihood function is the joint pdf of the observed data, viewed as a function of the parameters $\boldsymbol{\theta}$. Assuming the observations are independent and identically distributed, the likelihood function is:

$$L(\boldsymbol{\theta}|\mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_{i=1}^{n} f(\mathbf{x}_i|\boldsymbol{\theta})$$

Question 2

Question: What is the principle of maximum likelihood estimation (MLE)?

Solution: The principle of maximum likelihood estimation is to find the value of the parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood function $L(\boldsymbol{\theta})$. This value, denoted $\hat{\boldsymbol{\theta}}$, is the one that makes the observed data most probable. In practice, it is often easier to maximize the log-likelihood function, $\ln L(\boldsymbol{\theta})$.

Question 3

Question: For a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$ with known Σ, derive the MLE for $\boldsymbol{\mu}$.

Solution: The log-likelihood function (ignoring constants) is:

$$\ln L(\boldsymbol{\mu}) = -\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})$$

To find the maximum, we take the derivative with respect to $\boldsymbol{\mu}$ and set it to zero.

$$\frac{\partial \ln L(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \sum_{i=1}^{n} \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) = \mathbf{0}$$

$$\Sigma^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}) = \mathbf{0}$$

$$\sum \mathbf{x}_i - n\boldsymbol{\mu} = \mathbf{0} \implies n\boldsymbol{\mu} = \sum \mathbf{x}_i$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum \mathbf{x}_i = \bar{\mathbf{X}}$$

So, the MLE for $\boldsymbol{\mu}$ is the sample mean vector $\bar{\mathbf{X}}$.

Question 4

Question: For a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$, the MLEs are $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ and $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$. How does $\hat{\Sigma}$ relate to the sample covariance matrix S? Is $\hat{\Sigma}$ an unbiased estimator of Σ?

Solution: The MLE for Σ is $\hat{\Sigma} = \frac{1}{n}\sum(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$. The sample covariance matrix is $S = \frac{1}{n-1}\sum(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$. The relationship is $\hat{\Sigma} = \frac{n-1}{n}S$. The MLE $\hat{\Sigma}$ is a biased estimator of Σ. Its expectation is $E(\hat{\Sigma}) = \frac{n-1}{n}E(S) = \frac{n-1}{n}\Sigma$. The sample covariance matrix S is the unbiased estimator.

Question 5

Question: Define the Wishart distribution. What is its relationship to the multivariate normal distribution?

Solution: The Wishart distribution is a multivariate generalization of the chi-square distribution. It is the distribution of the sample sum of squares and cross-products matrix for a sample from a multivariate normal population. If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is a random sample from $N_p(\mathbf{0}, \Sigma)$, then the matrix $W = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T$ follows a Wishart distribution with n degrees of freedom, denoted $W \sim W_p(n, \Sigma)$.

Question 6

Question: State two properties of the Wishart distribution.

Solution: 1. **Additivity:** If $W_1 \sim W_p(n_1, \Sigma)$ and $W_2 \sim W_p(n_2, \Sigma)$ are independent, then $W_1 + W_2 \sim W_p(n_1 + n_2, \Sigma)$. 2. **Expectation:** If $W \sim W_p(n, \Sigma)$, then its expected value is $E(W) = n\Sigma$. 3. **Transformation:** If $W \sim W_p(n, \Sigma)$ and A is a $q \times p$ matrix, then $AWA^T \sim W_q(n, A\Sigma A^T)$.

Question 7

Question: Let $W_1 \sim W_p(n_1, \Sigma)$ and $W_2 \sim W_p(n_2, \Sigma)$ be independent Wishart matrices. What is the distribution of $W_1 + W_2$?

Solution: By the additivity property of the Wishart distribution, the sum of two independent Wishart matrices with the same scale matrix Σ is also a Wishart matrix with degrees of freedom added.

$$W_1 + W_2 \sim W_p(n_1 + n_2, \Sigma)$$

Question 8

Question: State the Central Limit Theorem for p-dimensional random vectors.

Solution: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from a population with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. For large n, the sample mean vector $\bar{\mathbf{X}}$ is approximately normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\frac{1}{n}\Sigma$.

$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{d} N_p(\mathbf{0}, \Sigma) \quad \text{as } n \to \infty$$

Question 9

Question: What is Hotelling's T^2 statistic used for? State the one-sample hypothesis test it is used for.

Solution: Hotelling's T^2 statistic is a multivariate generalization of the Student's t-statistic. It is used for hypothesis testing on the mean vector(s) of one or more multivariate normal populations. The one-sample test is used to test the null hypothesis that the population mean vector $\boldsymbol{\mu}$ is equal to a specific vector $\boldsymbol{\mu}_0$.

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Question 10

Question: Define the one-sample Hotelling's T^2 statistic in terms of the sample mean, hypothesized mean, sample covariance matrix, and sample size.

Solution: The one-sample Hotelling's T^2 statistic is defined as:

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T S^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

where n is the sample size, $\bar{\mathbf{X}}$ is the sample mean vector, $\boldsymbol{\mu}_0$ is the hypothesized mean vector, and S is the sample covariance matrix.

Question 11

Question: How is Hotelling's T^2 statistic related to the F-distribution? This relationship is used to find critical values for the test.

Solution: Under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the T^2 statistic follows a scaled F-distribution:

$$\frac{n-p}{p(n-1)} T^2 \sim F_{p,n-p}$$

where p is the number of variables and n is the sample size. This allows us to find a critical value for the test using the F-distribution.

Question 12

Question: A sample of size $n = 20$ from a bivariate normal population ($p = 2$) yields a T^2 statistic of 10.5. At a significance level of $\alpha = 0.05$, would you reject the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$? Assume $F_{2,18,0.05} = 3.55$.

Solution: First, we find the critical value for the T^2 test. Critical Value $= \frac{p(n-1)}{n-p} F_{p,n-p,\alpha} = \frac{2(20-1)}{20-2} F_{2,18,0.05} = \frac{38}{18} \cdot 3.55 \approx 2.11 \cdot 3.55 \approx 7.49$. The rejection rule is to reject H_0 if the observed $T^2 > 7.49$. Since our observed statistic is $T^2 = 10.5$, which is greater than 7.49, we reject the null hypothesis.

Question 13

Question: What is the two-sample Hotelling's T^2 test used for? State the null hypothesis.

Solution: The two-sample Hotelling's T^2 test is used to determine if two population mean vectors are equal. It is a multivariate generalization of the two-sample t-test. The null hypothesis is:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Question 14

Question: What assumptions are required for the two-sample Hotelling's T^2 test to be valid?

Solution: The main assumptions are: 1. The two samples are independent random samples from their respective populations. 2. Both populations are multivariate normal. 3. The covariance matrices of the two populations are equal ($\Sigma_1 = \Sigma_2$).

Question 15

Question: How does Hotelling's T^2 statistic relate to the Mahalanobis distance?

Solution: Hotelling's T^2 statistic is proportional to the squared Mahalanobis distance. Specifically, the one-sample T^2 is n times the squared Mahalanobis distance between the sample mean vector $\bar{\mathbf{X}}$ and the hypothesized mean vector $\boldsymbol{\mu}_0$, using the sample covariance matrix S to estimate the population covariance.

$$T^2 = n \cdot D_M^2(\bar{\mathbf{X}}, \boldsymbol{\mu}_0)$$