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 $R = D^{-1/2}SD^{-1/2}$, 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.

Question 16

Question: Prove that the sample generalized variance, $|S|$, is zero if and only if the columns of the mean-centered data matrix X_c are linearly dependent.

Solution: The sample covariance matrix is given by $S = \frac{1}{n-1} X_c^T X_c$, where X_c is the $n \times p$ mean-centered data matrix. The generalized variance is $|S| = |\frac{1}{n-1} X_c^T X_c| = (\frac{1}{n-1})^p |X_c^T X_c|$. Thus, $|S| = 0$ if and only if $|X_c^T X_c| = 0$. The matrix $X_c^T X_c$ is a $p \times p$ matrix. Its determinant is zero if and only if the matrix is singular. A matrix A is singular if and only if there exists a non-zero vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}$. So, $|X_c^T X_c| = 0$ if and only if there exists a non-zero $p \times 1$ vector $\mathbf{a}$ such that $X_c^T X_c \mathbf{a} = \mathbf{0}$. Multiplying by $\mathbf{a}^T$ on the left, we get $\mathbf{a}^T X_c^T X_c \mathbf{a} = 0$, which is $\|X_c \mathbf{a}\|^2 = 0$. This implies $X_c \mathbf{a} = \mathbf{0}$. $X_c \mathbf{a}$ is a linear combination of the columns of X_c. If $X_c \mathbf{a} = \mathbf{0}$ for a non-zero $\mathbf{a}$, this is the definition of the columns of X_c being linearly dependent. Conversely, if the columns of X_c are linearly dependent, there exists a non-zero $\mathbf{a}$ such that $X_c \mathbf{a} = \mathbf{0}$, which implies $X_c^T X_c \mathbf{a} = \mathbf{0}$, which means $X_c^T X_c$ is singular and its determinant is zero. Therefore, $|S| = 0$.

Question 17

Question: Prove that the sample covariance matrix S is positive semi-definite. That is, show $\mathbf{a}^T S \mathbf{a} \geq 0$ for any constant vector $\mathbf{a}$. When does strict equality, $\mathbf{a}^T S \mathbf{a} = 0$, hold?

Solution: Let $\mathbf{a}$ be any $p \times 1$ constant vector. Consider the quadratic form $\mathbf{a}^T S \mathbf{a}$.

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

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

Let $y_i = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{a}$, which is a scalar. Then $\mathbf{a}^T (\mathbf{x}_i - \bar{\mathbf{x}})$ is also a scalar, y_i. So the expression becomes:

$$= \frac{1}{n-1} \sum_{i=1}^{n} y_i^2$$

Since $y_i^2 \geq 0$, the sum is non-negative. Thus, $\mathbf{a}^T S \mathbf{a} \geq 0$, and S is positive semi-definite.

Strict equality, $\mathbf{a}^T S \mathbf{a} = 0$, holds if and only if $\sum y_i^2 = 0$, which means $y_i = 0$ for all $i = 1, \ldots, n$. $y_i = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{a} = 0$ for all i. This means the linear combination of the mean-centered variables, represented by vector $\mathbf{a}$, is zero for all observations. This corresponds to the columns of the mean-centered data matrix X_c being linearly dependent (as shown in Q16), where $\mathbf{a}$ is the vector of coefficients of the linear combination.

Question 18

Question: Show that the total sample variance and the sample generalized variance are invariant under a rotation of the data. A rotation is a linear transformation $\mathbf{y}_i = R\mathbf{x}_i$ where R is an orthogonal matrix ($R^T R = I$) with determinant $|R| = 1$.

Solution: Let the original data be $\mathbf{x}_i$ with sample mean $\bar{\mathbf{x}}$ and sample covariance S_x. The transformed data is $\mathbf{y}_i = R\mathbf{x}_i$. The new sample mean is $\bar{\mathbf{y}} = \frac{1}{n} \sum R\mathbf{x}_i = R(\frac{1}{n} \sum \mathbf{x}_i) = R\bar{\mathbf{x}}$. The new sample covariance matrix S_y is:

$$S_y = \frac{1}{n-1} \sum (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T = \frac{1}{n-1} \sum (R\mathbf{x}_i - R\bar{\mathbf{x}})(R\mathbf{x}_i - R\bar{\mathbf{x}})^T$$

$$= \frac{1}{n-1} \sum R(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T R^T = R \left(\frac{1}{n-1} \sum (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \right) R^T = R S_x R^T$$

Total Variance: The new total variance is $\mathrm{tr}(S_y) = \mathrm{tr}(RS_xR^T)$. Using the cyclic property of the trace, $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$:

$$\mathrm{tr}(S_y) = \mathrm{tr}(R^T RS_x) = \mathrm{tr}(IS_x) = \mathrm{tr}(S_x)$$

So the total variance is invariant under rotation.

Generalized Variance: The new generalized variance is $|S_y| = |RS_xR^T|$. Using the property $|ABC| = |A||B||C|$:

$$|S_y| = |R||S_x||R^T| = |R||S_x||R| = |R|^2|S_x|$$

For a rotation matrix, $|R| = 1$, so $|R|^2 = 1$.

$$|S_y| = |S_x|$$

So the generalized variance is also invariant under rotation.

Question 19

Question: Let S be a positive definite sample covariance matrix. It has a unique Cholesky decomposition $S = LL^T$, where L is a lower triangular matrix with positive diagonal elements. Explain how this decomposition can be used to simulate multivariate data that exhibits a specific covariance structure.

Solution: The Cholesky decomposition is a powerful tool for simulating multivariate data with a desired correlation structure. The process is as follows: 1. Start with a vector $\mathbf{z} = (z_1, \ldots, z_p)^T$ of p independent standard normal random variables (i.e., $z_j \sim N(0, 1)$). The mean of this vector is $E[\mathbf{z}] = \mathbf{0}$ and its covariance matrix is $\mathrm{Cov}(\mathbf{z}) = I$, the identity matrix. 2. Let the desired covariance structure for the simulated data be given by the matrix S, and let the desired mean be $\bar{\mathbf{x}}$. 3. First, find the Cholesky decomposition of the target covariance matrix, $S = LL^T$. 4. Transform the vector of independent standard normal variables $\mathbf{z}$ using the lower triangular matrix L and the target mean $\bar{\mathbf{x}}$:

$$\mathbf{x} = \bar{\mathbf{x}} + L\mathbf{z}$$

The resulting vector $\mathbf{x}$ will have the desired properties. Let's check its mean and covariance:

$$E[\mathbf{x}] = E[\bar{\mathbf{x}} + L\mathbf{z}] = \bar{\mathbf{x}} + LE[\mathbf{z}] = \bar{\mathbf{x}} + L\mathbf{0} = \bar{\mathbf{x}}$$

$$\mathrm{Cov}(\mathbf{x}) = \mathrm{Cov}(\bar{\mathbf{x}} + L\mathbf{z}) = L\mathrm{Cov}(\mathbf{z})L^T = LIL^T = LL^T = S$$

This procedure generates a random vector $\mathbf{x}$ from a distribution with the specified sample mean and sample covariance matrix S. Repeating this process n times would generate a full dataset. This technique is fundamental in Monte Carlo simulations for risk management, financial modeling, and statistical power analysis.

Question 20

Question: The sample correlation matrix R is related to the covariance matrix S by $S = D^{1/2}RD^{1/2}$, where D is the diagonal matrix of sample variances s_{jj}. Prove that the generalized variance $|S|$ can be expressed as the product of the individual sample variances and the determinant of the correlation matrix, i.e., $|S| = (\prod_{j=1}^{p} s_{jj})|R|$. What does this imply about the relationship between inter-correlation among variables and the generalized variance?

Solution: We start with the relationship $S = D^{1/2}RD^{1/2}$. Taking the determinant of both sides:

$$|S| = |D^{1/2}RD^{1/2}|$$

Using the property that $|ABC| = |A||B||C|$:

$$|S| = |D^{1/2}||R||D^{1/2}| = |D^{1/2}|^2|R|$$

The matrix $D^{1/2}$ is a diagonal matrix with elements $\sqrt{s_{jj}}$ on the diagonal. The determinant of a diagonal matrix is the product of its diagonal elements.

$$|D^{1/2}| = \prod_{j=1}^{p} \sqrt{s_{jj}}$$

Therefore, $|D^{1/2}|^2 = \left(\prod_{j=1}^{p} \sqrt{s_{jj}}\right)^2 = \prod_{j=1}^{p} s_{jj}$. Substituting this back into the equation for $|S|$:

$$|S| = \left(\prod_{j=1}^{p} s_{jj}\right)|R|$$

This completes the proof.

Implication: This result shows that the generalized variance $|S|$ is a combination of the total individual variances (product of s_{jj}) and the inter-correlations among the variables (captured by $|R|$). The determinant of the correlation matrix, $|R|$, ranges from 0 to 1. - If the variables are uncorrelated, $R = I$ and $|R| = 1$. In this case, $|S| = \prod s_{jj}$, and the generalized variance is maximized for a given set of individual variances. - If the variables are perfectly correlated (linearly dependent), then $|R| = 0$, which implies $|S| = 0$. - As the magnitude of correlation among variables increases, $|R|$ decreases from 1 towards 0, which in turn reduces the generalized variance $|S|$. Therefore, higher inter-correlation leads to a smaller "effective" volume of the data cloud in the feature space, as measured by the generalized variance.

Question 21

Question: Prove that the sample covariance matrix $\mathbf{S}$ is positive definite if the columns of the mean-corrected data matrix $\mathbf{X}_c = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T$ are linearly independent.

Solution: The sample covariance matrix is given by $\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^T\mathbf{X}_c$. To show that $\mathbf{S}$ is positive definite, we must show that for any non-zero vector $\mathbf{a} \in \mathbb{R}^p$, the quadratic form $\mathbf{a}^T\mathbf{S}\mathbf{a}$ is strictly positive.

$$\mathbf{a}^T\mathbf{S}\mathbf{a} = \mathbf{a}^T\left(\frac{1}{n-1}\mathbf{X}_c^T\mathbf{X}_c\right)\mathbf{a} = \frac{1}{n-1}(\mathbf{X}_c\mathbf{a})^T(\mathbf{X}_c\mathbf{a}) = \frac{1}{n-1}\|\mathbf{X}_c\mathbf{a}\|^2$$

The term $\|\mathbf{X}_c\mathbf{a}\|^2$ is the squared Euclidean norm of the vector $\mathbf{y} = \mathbf{X}_c\mathbf{a}$. The norm is zero if and only if the vector is a zero vector. So, $\mathbf{a}^T\mathbf{S}\mathbf{a} = 0$ if and only if $\mathbf{X}_c\mathbf{a} = \mathbf{0}$.

The expression $\mathbf{X}_c\mathbf{a}$ is a linear combination of the columns of the mean-corrected matrix $\mathbf{X}_c$. The condition that the columns of $\mathbf{X}_c$ are linearly independent means that the only solution to $\mathbf{X}_c\mathbf{a} = \mathbf{0}$ is the trivial solution $\mathbf{a} = \mathbf{0}$. Therefore, for any non-zero vector $\mathbf{a}$, we must have $\mathbf{X}_c\mathbf{a} \neq \mathbf{0}$, which implies $\|\mathbf{X}_c\mathbf{a}\|^2 > 0$. This means that for any non-zero $\mathbf{a}$, $\mathbf{a}^T\mathbf{S}\mathbf{a} > 0$. Thus, $\mathbf{S}$ is positive definite. This condition typically holds as long as the number of observations n is greater than the number of variables p.

Question 22

Question: Let S_x be the sample covariance matrix for an $n \times p$ data matrix X. If we create a new data matrix Y by adding a constant vector $\mathbf{c}$ to every observation row, so $\mathbf{y}_i^T = \mathbf{x}_i^T + \mathbf{c}^T$, what is the sample covariance matrix S_y of the new data? Prove your result.

Solution: First, let's find the sample mean of the new data, $\bar{\mathbf{y}}$.

$$\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i + \mathbf{c}) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\right) + \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{c}\right) = \bar{\mathbf{x}} + \frac{1}{n}(n\mathbf{c}) = \bar{\mathbf{x}} + \mathbf{c}$$

Now, let's compute the new sample covariance matrix S_y.

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

Substitute the expressions for $\mathbf{y}_i$ and $\bar{\mathbf{y}}$:

$$\mathbf{y}_i - \bar{\mathbf{y}} = (\mathbf{x}_i + \mathbf{c}) - (\bar{\mathbf{x}} + \mathbf{c}) = \mathbf{x}_i - \bar{\mathbf{x}}$$

The deviation from the mean is unchanged by the shift. Therefore, the new covariance matrix is:

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

The sample covariance matrix is invariant to a constant shift in the data. This makes sense, as covariance is a measure of spread, and shifting the data does not change its spread.

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, $c_1 x_1 + ... + c_p x_p$. 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.

Question 16

Question: Let X_c be the $n \times p$ mean-centered data matrix. The projection of the n observation points onto a p-dimensional vector $\mathbf{a}$ results in a new vector of data points $X_c\mathbf{a}$. Show that the sample variance of these projected points is given by $\mathbf{a}^T S\mathbf{a}$.

Solution: The vector of projected points is $\mathbf{z} = X_c\mathbf{a}$. The mean of the projected points is $\bar{z} = 0$. The sample variance of the projected points is:

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 = \frac{1}{n-1}(X_c\mathbf{a})^T(X_c\mathbf{a}) = \mathbf{a}^T\left(\frac{1}{n-1}X_c^T X_c\right)\mathbf{a} = \mathbf{a}^T S\mathbf{a}$$

This result is fundamental to Principal Component Analysis (PCA).

Question 17

Question: Let $\mathbf{d}_j$ and $\mathbf{d}_k$ be two mean-centered data vectors. The simple linear regression coefficient of variable j on variable k is $b_{jk} = \frac{\mathbf{d}_j^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{d}_k}$. Interpret this geometrically.

Solution: In the n-dimensional sample space, the vectors $\mathbf{d}_j$ and $\mathbf{d}_k$ represent the two variables. The formula for the regression coefficient b_{jk} is identical to the scalar component of the projection of vector $\mathbf{d}_j$ onto vector $\mathbf{d}_k$. The projection of $\mathbf{d}_j$ onto the line defined by $\mathbf{d}_k$ is:

$$\text{proj}_{\mathbf{d}_k} \mathbf{d}_j = \left(\frac{\mathbf{d}_j^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{d}_k} \right) \mathbf{d}_k = b_{jk} \mathbf{d}_k$$

The vector of predicted values, $\hat{\mathbf{d}}_j = b_{jk} \mathbf{d}_k$, is this projection.

Question 18

Question: Consider the projection matrix $P = \frac{1}{n} \mathbf{1}\mathbf{1}^T$. Show that P is idempotent and symmetric, and interpret its effect and the effect of $(I - P)$.

Solution: P is symmetric since $P^T = (\frac{1}{n}\mathbf{1}\mathbf{1}^T)^T = \frac{1}{n}\mathbf{1}\mathbf{1}^T = P$. P is idempotent since $P^2 = (\frac{1}{n}\mathbf{1}\mathbf{1}^T)(\frac{1}{n}\mathbf{1}\mathbf{1}^T) = \frac{1}{n^2}\mathbf{1}(\mathbf{1}^T\mathbf{1})\mathbf{1}^T = \frac{n}{n^2}\mathbf{1}\mathbf{1}^T = P$. $P\mathbf{x} = \bar{x}\mathbf{1}$, so it projects a vector onto the mean vector. $(I - P)\mathbf{x} = \mathbf{x} - \bar{x}\mathbf{1}$, so it produces the mean-centered vector.

Question 19

Question: From a geometric perspective, what is the goal of Canonical Correlation Analysis (CCA)?

Solution: Geometrically, CCA seeks to find a pair of vectors, one in the subspace spanned by the first set of variables and one in the subspace spanned by the second set, such that the angle between these two new vectors is minimized. The cosine of this minimum angle is the first canonical correlation.

Question 20

Question: The CCA canonical vectors are found by solving $(S_{12}S_{22}^{-1}S_{21})\mathbf{a} = \lambda S_{11}\mathbf{a}$. Interpret the matrix $M = S_{12}S_{22}^{-1}S_{21}$.

Solution: The matrix M is the covariance matrix of the predicted values of the first set of variables, $\hat{X}_1$, when they are predicted from the second set of variables using multivariate multiple regression. It represents the component of the variance in the first set of variables that is explained by the second set.

Question 21

Question: Justify that $(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})$ is a valid statistical distance measure that is scale-invariant.

Solution: The squared Mahalanobis distance is a valid distance as it is non-negative, zero only if $\mathbf{x} = \bar{\mathbf{x}}$, and symmetric. It is scale-invariant because if we scale the data by $\mathbf{z}_i = D^{-1/2}\mathbf{x}_i$, the new covariance matrix is the correlation matrix R. The distance computed for $\mathbf{z}$ is $D_z^2 = (\mathbf{z} - \bar{\mathbf{z}})^T R^{-1} (\mathbf{z} - \bar{\mathbf{z}})$, which can be shown to be equal to the original distance D_x^2.

Question 22

Question: Use the geometric interpretation in n-space to justify that generalised sample variance is a joint measure of variation. What is its major weakness?

Solution: In n-space, $|S|$ is proportional to the squared volume of the parallelepiped formed by the p mean-centered data vectors. This volume is large when the vectors are long (high variance) and not collinear (low correlation), so it captures joint variation. Its weakness is extreme sensitivity to the scale of the variables.

Question 23

Question: (True/False) For any $\mathbf{X} \in \mathbb{R}^p$, the Mahalanobis distance $(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$ is non-negative. Justify.

Solution: True. Since $\boldsymbol{\Sigma}$ must be positive definite for its inverse to exist, $\boldsymbol{\Sigma}^{-1}$ is also positive definite. The quadratic form $\mathbf{v}^T \boldsymbol{\Sigma}^{-1} \mathbf{v}$ is therefore positive for any non-zero vector $\mathbf{v} = \mathbf{X} - \boldsymbol{\mu}$.

Question 24

Question: Draw a constellation graph for the provided student marks data and identify the best, worst, and most average students.

Solution: A constellation graph plots each observation as a star with rays proportional to the variable values. Calculations show: - Best student (furthest from origin): Student 9, with scores (4, 10, 8) and distance 13.42. - Worst student (closest to origin): Student 4, with scores (2, 2, 3) and distance 4.12. - Most average student (closest to the mean vector (2.55, 5.4, 5.3)): Student 3, with scores (2, 4, 5).

Question 25

Question: In multiple regression, show that the vector of residuals $\mathbf{e} = (I - H)\mathbf{y}$ is orthogonal to the vector of fitted values $\hat{\mathbf{y}} = H\mathbf{y}$.

Solution: We need to show $\hat{\mathbf{y}}^T\mathbf{e} = 0$.

$$\hat{\mathbf{y}}^T\mathbf{e} = (H\mathbf{y})^T((I - H)\mathbf{y}) = \mathbf{y}^T H^T (I - H)\mathbf{y}$$

The hat matrix H is symmetric ($H^T = H$) and idempotent ($H^2 = H$).

$$\mathbf{y}^T H(I - H)\mathbf{y} = \mathbf{y}^T (H - H^2)\mathbf{y} = \mathbf{y}^T(\mathbf{0})\mathbf{y} = 0$$

Thus, the vectors are orthogonal.

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 $\text{Cov}(\mathbf{X})$, 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 $\text{Cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$.

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:

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

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 $E(\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = \mathrm{Cov}(\bar{\mathbf{X}}) + E(\bar{\mathbf{X}})E(\bar{\mathbf{X}})^T = \frac{1}{n}\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T$.

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

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

Question 16

Question: Prove the multivariate Cauchy-Schwarz inequality: $(E[\mathbf{X}^T\mathbf{Y}])^2 \le E[\mathbf{X}^T\mathbf{X}]E[\mathbf{Y}^T\mathbf{Y}]$.

Solution: Consider the scalar random variable $Z(t) = (\mathbf{X} - t\mathbf{Y})^T(\mathbf{X} - t\mathbf{Y}) \ge 0$. Its expectation must also be non-negative: $E[Z(t)] = E[\mathbf{X}^T\mathbf{X}] - 2tE[\mathbf{X}^T\mathbf{Y}] + t^2E[\mathbf{Y}^T\mathbf{Y}] \ge 0$. This is a quadratic in t that is always non-negative, so its discriminant must be ≤ 0. $(-2E[\mathbf{X}^T\mathbf{Y}])^2 - 4(E[\mathbf{Y}^T\mathbf{Y}])(E[\mathbf{X}^T\mathbf{X}]) \le 0 \implies (E[\mathbf{X}^T\mathbf{Y}])^2 \le E[\mathbf{X}^T\mathbf{X}]E[\mathbf{Y}^T\mathbf{Y}]$.

Question 17

Question: Let $\mathbf{X}$ be a random vector with mean $\boldsymbol{\mu}$ and covariance Σ. Find the matrix A that minimizes the mean squared error $E[\|A\mathbf{X} - \mathbf{b}\|^2]$ for a constant vector $\mathbf{b}$.

Solution: We want to minimize $J(A) = E[\|A\mathbf{X} - \mathbf{b}\|^2]$. By differentiating with respect to A and setting to zero, we find the optimal matrix is $A = \mathbf{b}\boldsymbol{\mu}^T(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T)^{-1}$.

Question 18

Question: Show that the sample mean vector $\bar{\mathbf{X}}$ and the vector of deviations $(\mathbf{X}_i - \bar{\mathbf{X}})$ are uncorrelated.

Solution: We need to show $\mathrm{Cov}(\bar{\mathbf{X}}, \mathbf{X}_i - \bar{\mathbf{X}}) = \mathbf{0}$. This is equivalent to showing $E[\bar{\mathbf{X}}(\mathbf{X}_i - \bar{\mathbf{X}})^T] = \mathbf{0}$ since $E[\mathbf{X}_i - \bar{\mathbf{X}}] = \mathbf{0}$. $E[\bar{\mathbf{X}}\mathbf{X}_i^T - \bar{\mathbf{X}}\bar{\mathbf{X}}^T] = (\frac{1}{n}\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T) - (\frac{1}{n}\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T) = \mathbf{0}$.

Question 19

Question: Let $\mathbf{X}$ be a random vector with mean $\boldsymbol{\mu}$ and covariance Σ. Let A be a symmetric matrix. Show that $E[\mathbf{X}^TA\mathbf{X}] = \mathrm{tr}(A\Sigma) + \boldsymbol{\mu}^TA\boldsymbol{\mu}$.

Solution: $E[\mathbf{X}^TA\mathbf{X}] = E[\mathrm{tr}(A\mathbf{X}\mathbf{X}^T)] = \mathrm{tr}(AE[\mathbf{X}\mathbf{X}^T])$. Since $E[\mathbf{X}\mathbf{X}^T] = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T$, this becomes $\mathrm{tr}(A(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T)) = \mathrm{tr}(A\Sigma) + \mathrm{tr}(A\boldsymbol{\mu}\boldsymbol{\mu}^T) = \mathrm{tr}(A\Sigma) + \boldsymbol{\mu}^TA\boldsymbol{\mu}$.

Question 20

Question: For a partitioned covariance matrix $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, prove $|\Sigma| = |\Sigma_{22}||\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|$. Provide a statistical interpretation of $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Solution: The result is proven using a block matrix decomposition. Statistically, $\Sigma_{11.2}$ is the conditional covariance matrix of $\mathbf{X}_1$ given $\mathbf{X}_2$. It is the covariance of the residuals from regressing $\mathbf{X}_1$ on $\mathbf{X}_2$.

Question 21

Question: A sample of size 10 yielded $\mathbf{S} = \begin{pmatrix} 0.85 & 0.63 & 0.17 \\ 0.63 & 0.57 & 0.13 \\ 0.17 & 0.13 & 0.17 \end{pmatrix}$. Compute an unbiased estimate of the covariance matrix of $\begin{pmatrix} X_1 - X_2 \\ X_1 - X_3 \end{pmatrix}$.

Solution: Let $\mathbf{Y} = A\mathbf{X}$ where $A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$. An unbiased estimator for $\mathrm{Cov}(\mathbf{Y}) = A\Sigma A^T$ is $A\mathbf{S}A^T$.

$$A\mathbf{S}A^T = \begin{pmatrix} 0.16 & 0.18 \\ 0.18 & 0.68 \end{pmatrix}$$

Question 22

Question: Let $\mathbf{X} \sim N_3(\boldsymbol{\mu}, \mathbf{\Sigma})$, with $\mathbf{\Sigma} = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix}$ and $\mathbf{a} = (1, -1, 1)^T$. Find the correlations between each X_i and $Y = \mathbf{a}^T\mathbf{X}$.

Solution: We need $r_i = \frac{\text{Cov}(X_i, Y)}{\sqrt{\text{Var}(X_i)\text{Var}(Y)}}$. $\text{Var}(Y) = \mathbf{a}^T\mathbf{\Sigma}\mathbf{a} = 15$. The vector of covariances is $\text{Cov}(\mathbf{X}, Y) = \mathbf{\Sigma}\mathbf{a} = (5, -3, 7)^T$. The variances are $\text{Var}(X_1) = 4, \text{Var}(X_2) = 4, \text{Var}(X_3) = 9$. The vector of correlations is $\mathbf{r} \approx (0.6455, -0.3873, 0.6030)^T$.

Question 23

Question: Let $\mathbf{X}$ be a random vector with covariance $\mathbf{\Sigma}$. Let A and B be constant matrices. Derive the expression for $\text{Cov}(A\mathbf{X}, B\mathbf{X})$.

Solution:

$$\text{Cov}(A\mathbf{X}, B\mathbf{X}) = E[(A\mathbf{X}-A\boldsymbol{\mu})(B\mathbf{X}-B\boldsymbol{\mu})^T] = E[A(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T B^T] = AE[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T]B^T = A\mathbf{\Sigma}B^T$$

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: $\mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 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: $\mathrm{Var}(X_1|X_2 = 1) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 4 - 2 \cdot (1/2) \cdot 2 = 4 - 2 = 2$. 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.

Question 16

Question: Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ be partitioned as $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, where $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Derive the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$.

Solution: The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is 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: $\text{Cov}(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Question 17

Question: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$. Use the moment generating function (MGF) to prove that for a constant matrix A of size $q \times p$, the random vector $\mathbf{Y} = A\mathbf{X}$ is distributed as $N_q(A\boldsymbol{\mu}, A\Sigma A^T)$.

Solution: The MGF of $\mathbf{Y} = A\mathbf{X}$ is $M_{\mathbf{Y}}(\mathbf{s}) = E[e^{\mathbf{s}^T(A\mathbf{X})}] = M_{\mathbf{X}}(A^T\mathbf{s})$. Substituting $\mathbf{t} = A^T\mathbf{s}$ into the MGF of $\mathbf{X}$ gives the MGF of a $N_q(A\boldsymbol{\mu}, A\Sigma A^T)$ distribution.

Question 18

Question: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and let $\Sigma = P\Lambda P^T$ be the spectral decomposition of Σ. Find the distribution of the transformed vector $\mathbf{Y} = P^T(\mathbf{X} - \boldsymbol{\mu})$.

Solution: $\mathbf{Y}$ is a linear transformation of a normal vector, so it is normal. $E[\mathbf{Y}] = \mathbf{0}$. $\text{Cov}(\mathbf{Y}) = P^T\Sigma P = P^T(P\Lambda P^T)P = \Lambda$. So, $\mathbf{Y} \sim N_p(\mathbf{0}, \Lambda)$.

Question 19

Question: Let $W \sim W_p(n, I)$. Describe a Monte Carlo simulation to estimate the 95th percentile of the distribution of the largest eigenvalue of W/n.

Solution: 1. Set parameters p, n, N. 2. For $i = 1, \ldots, N$: Generate n vectors $\mathbf{z}_j \sim N_p(\mathbf{0}, I)$ and form $W_i = \sum \mathbf{z}_j \mathbf{z}_j^T$. 3. Calculate the largest eigenvalue $\lambda_{max,i}$ of W_i/n. 4. The 95th percentile is the $0.95 \times N$ value in the sorted list of $\lambda_{max,i}$.

Question 20

Question: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, and let $\mathbf{a}$ and $\mathbf{b}$ be two constant vectors. Find the covariance of $Y_1 = \mathbf{a}^T\mathbf{X}$ and $Y_2 = \mathbf{b}^T\mathbf{X}$. What is the condition for independence?

Solution: $\text{Cov}(Y_1, Y_2) = \mathbf{a}^T\Sigma\mathbf{b}$. They are independent if and only if this covariance is zero.

Question 21

Question: Let $\mathbf{X} \sim \mathcal{N}_2\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}5 & -4\\-4 & 5\end{pmatrix}\right)$. Determine the 90% constant-density ellipse and its axes.

Solution: The ellipse is given by $5(x_1-1)^2+8(x_1-1)(x_2-1)+5(x_2-1)^2 \leq 41.49$. The eigenvalues are $\lambda_1 = 1, \lambda_2 = 9$. The major axis is in the direction $(1, -1)^T$ with semi-axis length ≈ 6.44. The minor axis is in the direction $(1, 1)^T$ with semi-axis length ≈ 2.15.

Question 22

Question: (True/False) If $aX + bY$ is normal for all a, b, then (X, Y) is bivariate normal. Justify.

Solution: True. This is the Cramér-Wold device.

Question 23

Question: (True/False) Let $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and g be continuous. Then $g(\mathbf{X})$ is univariate normal. Justify.

Solution: False. Only true if g is linear. Counterexample: $X^2 \sim \chi^2_1$.

Question 24

Question: Given candy bar weights $(X_1, X_2, X_3)^T \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. (a) Find $P(X_2 > 8 | X_1 = 2, X_3 = 10)$. (b) Determine $P(X_1 - 2X_2 + X_3 < 5)$.

Solution: (a) The conditional distribution of X_2 given (X_1, X_3) is $N(5.917, 3.306)$. $P(X_2 > 8) \approx 0.126$. (b) The linear combination $W = X_1 - 2X_2 + X_3$ is distributed $N(0, 25)$. $P(W < 5) \approx 0.8413$.

Question 25

Question: Suppose $\mathbf{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with given parameters. Obtain the conditional distribution of $(X_1, X_3)^T$ given $X_2 = 2$.

Solution: The conditional distribution is $N_2\left(\begin{pmatrix}0.5\\3.5\end{pmatrix}, \begin{pmatrix}0.5 & 1.5\\1.5 & 2.5\end{pmatrix}\right)$.

Question 26

Question: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$. Show that the Mahalanobis distance is invariant under any full-rank linear transformation of the data, $\mathbf{Y} = A\mathbf{X}$.

Solution: The squared Mahalanobis distance for $\mathbf{Y}$ is $D_y^2 = (\mathbf{Y} - \boldsymbol{\mu}_y)^T\Sigma_y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_y)$. Substituting the definitions of the transformed parameters shows that $D_y^2 = D_x^2$.

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.

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

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 relationship is $\hat{\Sigma} = \frac{n-1}{n}S$. The MLE $\hat{\Sigma}$ is a biased estimator of Σ. The sample covariance matrix S is the unbiased estimator.

Question 5

Question: Define the Wishart distribution.

Solution: The Wishart distribution, $W_p(n, \Sigma)$, is the distribution of the matrix $A = \sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T$, where each $\mathbf{X}_i$ is an independent random vector from $N_p(\mathbf{0}, \Sigma)$. It is a multivariate generalization of the chi-square distribution.

Question 6

Question: Let $\mathbf{A} \sim W_p(n, \mathbf{\Sigma})$. Prove that $\mathbb{E}(\mathbf{A}) = n\mathbf{\Sigma}$.

Solution: $\mathbb{E}(\mathbf{A}) = \mathbb{E}\left[\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T\right] = \sum_{i=1}^{n}\mathbb{E}[\mathbf{X}_i\mathbf{X}_i^T]$. Since $\mathbb{E}[\mathbf{X}_i\mathbf{X}_i^T] = \mathrm{Cov}(\mathbf{X}_i) + \mathbb{E}[\mathbf{X}_i]\mathbb{E}[\mathbf{X}_i]^T = \mathbf{\Sigma} + \mathbf{0} = \mathbf{\Sigma}$, the result follows.

Question 7

Question: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Prove or disprove that the distribution of $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ converges to a chi-square distribution as $n \to \infty$.

Solution: True. By the Law of Large Numbers, $\mathbf{S} \xrightarrow{p} \mathbf{\Sigma}$. By the multivariate CLT, $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{d} N_p(\mathbf{0}, \mathbf{\Sigma})$. By Slutsky's theorem, the statistic converges in distribution to $\mathbf{Y}^T\mathbf{\Sigma}^{-1}\mathbf{Y}$ where $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, which is the definition of a χ^2_p distribution.

Question 8

Question: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Show that the asymptotic distribution of $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$ is χ^2_p.

Solution: By the multivariate CLT, $\mathbf{Y}_n = \sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{Y} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$. By the continuous mapping theorem, the distribution of $\mathbf{Y}_n^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}_n$ converges to that of $\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}$. Let $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2} \mathbf{Y} \sim N_p(\mathbf{0}, I)$. Then $\mathbf{Y}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} = \mathbf{Z}^T \mathbf{Z} \sim \chi^2_p$.

Question 9

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

Solution: It is used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The test statistic is $T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T S^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$.

Question 10

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

Solution: It is used to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. It assumes two independent random samples from multivariate normal populations with a common covariance matrix $\Sigma_1 = \Sigma_2 = \Sigma$.

Question 11

Question: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from $\mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For testing $H_0 : \mu_1 = \mu_2 - \mu_3$, construct a test based on Hotelling's T^2 statistic.

Solution: The hypothesis is $H_0 : \mathbf{C}\boldsymbol{\mu} = 0$, where $\mathbf{C} = (1, -1, 1)$. Let $\mathbf{Y}_i = \mathbf{C}\mathbf{X}_i$. The problem reduces to a one-sample test for the mean of the univariate random variable Y. The statistic is $T^2 = n(\mathbf{C}\bar{\mathbf{X}})^T (\mathbf{C} S \mathbf{C}^T)^{-1} (\mathbf{C}\bar{\mathbf{X}})$, which is equivalent to a standard t-test.

Question 12

Question: Let a population distribution be $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Based on a random sample of size n, develop a large sample LRT for $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$.

Solution: The LRT statistic is $\Lambda = \frac{\sup_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \Sigma_0)}{\sup_{\boldsymbol{\mu}, \Sigma} L(\boldsymbol{\mu}, \Sigma)}$. It can be shown that $-2 \ln \Lambda = n(\mathrm{tr}(\hat{\Sigma}\Sigma_0^{-1}) - \ln(|\hat{\Sigma}\Sigma_0^{-1}|) - p)$, which is approximately $\chi^2_{p(p+1)/2}$ for large n.

Question 13

Question: Explain Bonferroni simultaneous confidence intervals for $\mu_1, \ldots, \mu_p$ and compare them to T^2 intervals.

Solution: To get a $100(1-\alpha)\%$ simultaneous confidence level, construct each of the p individual intervals at the $100(1-\alpha/p)\%$ level using a t-distribution critical value: $\bar{x}_i \pm t_{n-1, \alpha/(2p)} \sqrt{s_{ii}/n}$. Bonferroni intervals are generally wider (more conservative) than T^2 intervals but are more flexible and easier to compute.

Question 14

Question: Derive the two-sample Hotelling's T^2 statistic using the likelihood ratio test framework.

Solution: The LRT for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ (assuming $\Sigma_1 = \Sigma_2 = \Sigma$) is $\Lambda = \frac{\sup_{\boldsymbol{\mu}, \Sigma} L(\boldsymbol{\mu}, \Sigma)}{\sup_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma)}$. The resulting statistic $\Lambda^{2/n}$ is a monotonic function of the two-sample Hotelling's $T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T S_{pooled}^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$.

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.

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

Question 16

Question: For a random sample from a $N_p(\boldsymbol{\mu}, \Sigma)$ population, derive the likelihood ratio test (LRT) for the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, with Σ unknown. Show that the LRT statistic is a monotonic function of Hotelling's T^2 statistic.

Solution: The LRT statistic is $\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{n/2}$. We can show that $\frac{|\hat{\Sigma}_0|}{|\hat{\Sigma}|} = 1 + \frac{T^2}{n-1}$, where $T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T S^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$. So, $\Lambda^{2/n} = \left(1 + \frac{T^2}{n-1}\right)^{-1}$. Since Λ is a decreasing function of T^2, the tests are equivalent.

Question 17

Question: Compare the Mean Squared Error (MSE) of the MLE for variance, $\hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$, and the sample variance, $s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$.

Solution: $\text{MSE}(s^2) = \frac{2\sigma^4}{n-1}$ and $\text{MSE}(\hat{\sigma}^2) = \frac{(2n-1)\sigma^4}{n^2}$. For $n > 1$, $\text{MSE}(s^2) > \text{MSE}(\hat{\sigma}^2)$, so the MLE is preferred in terms of MSE.

Question 18

Question: Propose a test statistic for the multivariate Behrens-Fisher problem ($H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ when $\Sigma_1 \neq \Sigma_2$) and explain the difficulty.

Solution: For large samples, a reasonable test statistic is $T^2_{BF} = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$, which is approximately χ^2_p. The difficulty for small samples is that the exact distribution of this statistic is unknown.

Question 19

Question: Derive the LRT for the sphericity hypothesis $H_0 : \Sigma = \sigma^2 I$.

Solution: The LRT statistic is $\Lambda = \left(\frac{|\hat{\Sigma}|}{(\frac{\text{tr}(\hat{\Sigma})}{p})^p}\right)^{n/2}$. This is a power of the ratio of the geometric mean to the arithmetic mean of the eigenvalues of $\hat{\Sigma}$.

Question 20

Question: Discuss the concept of power for the one-sample Hotelling's T^2 test.

Solution: Power is the probability of correctly rejecting H_0. For Hotelling's T^2, it depends on the non-centrality parameter $\delta^2 = n(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$. Power increases with δ^2.

Question 21

Question: To test for an "equal correlation" structure, $H_0 : \mathbf{P} = (1-\rho)I + \rho\mathbf{1}\mathbf{1}^T$, the test statistic for a large sample is given by $T = \frac{n-1}{1-\bar{r}^2}((\sum\sum_{i<k}(r_{ik} - \bar{r})^2) - \gamma\sum_{k=1}^{p}(\bar{r}_k - \bar{r})^2)$, where $\gamma = \frac{p(p-1)}{2(p-2)^2}$ and other terms are sample estimates. It is known that $T \sim \chi^2_{(p+1)(p-2)/2}$. Use this to test H_0 at 1% level of significance for a sample of size 150 with correlation matrix $R = \begin{pmatrix} 1 & .7501 & 0.6392 & 0.6363 \\ .7501 & 1 & 0.6925 & 0.7386 \\ 0.6392 & 0.6925 & 1 & 0.6625 \\ 0.6363 & 0.7386 & 0.6625 & 1 \end{pmatrix}$.

Solution: Here $p = 4$ and $n = 150$. We calculate the mean correlation $\bar{r} \approx 0.6865$. We then calculate the sum of squared differences of correlations from the mean, and the sum of squared differences of column-average correlations from the mean. Plugging these values into the formula for T gives $T \approx 2.41$. The degrees of freedom are $(4+1)(4-2)/2 = 5$. The critical value is $\chi^2_{5,0.01} = 15.086$. Since $2.41 < 15.086$, we do not reject H_0.

Question 22

Question: (True/False) For any fixed vector $\mathbf{d} \in \mathbb{R}^p$, the distribution of the squared Mahalanobis distance $(\mathbf{d} - \mathbf{X})^T \mathbf{\Sigma}^{-1} (\mathbf{d} - \mathbf{X})$, where $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, has a non-central chi-square distribution. Justify your answer.

Solution: True. Let $\mathbf{Y} = \mathbf{X} - \mathbf{d} \sim N_p(\boldsymbol{\mu} - \mathbf{d}, \mathbf{\Sigma})$. The quadratic form is $\mathbf{Y}^T \mathbf{\Sigma}^{-1} \mathbf{Y}$. Let $\mathbf{Z} = \mathbf{\Sigma}^{-1/2} \mathbf{Y} \sim N_p(\mathbf{\Sigma}^{-1/2}(\boldsymbol{\mu} - \mathbf{d}), I)$. The form is $\mathbf{Z}^T \mathbf{Z}$, which is a sum of squared normal variables with unit variance and non-zero mean. This follows a non-central chi-square distribution, $\chi^2_p(\delta)$, with non-centrality parameter $\delta^2 = (\boldsymbol{\mu} - \mathbf{d})^T \mathbf{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{d})$.

Question 23

Question: Let $\mathbf{A} \sim W_p(n, \mathbf{\Sigma})$. Then using the definition of Wishart distribution or otherwise, prove or disprove $\mathbb{E}(\mathbf{A}) = n\mathbf{\Sigma}$.

Solution: This is true. By the definition of the Wishart distribution, the matrix $\mathbf{A}$ can be represented as the sum of n independent outer products of random vectors, $\mathbf{A} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T$, where each $\mathbf{X}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. We want to compute the expectation of $\mathbf{A}$. By the linearity of expectation:

$$\mathbb{E}(\mathbf{A}) = \mathbb{E}\left[\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T\right] = \sum_{i=1}^n \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^T]$$

The expectation $\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^T]$ is the second moment matrix of the random vector $\mathbf{X}_i$. This is related to the covariance matrix by $\text{Cov}(\mathbf{X}_i) = \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^T] - \mathbb{E}[\mathbf{X}_i]\mathbb{E}[\mathbf{X}_i]^T$. Since $\mathbf{X}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, its mean is $\mathbb{E}[\mathbf{X}_i] = \mathbf{0}$ and its covariance is $\text{Cov}(\mathbf{X}_i) = \mathbf{\Sigma}$. Therefore, $\mathbb{E}[\mathbf{X}_i \mathbf{X}_i^T] = \text{Cov}(\mathbf{X}_i) + \mathbf{0}\mathbf{0}^T = \mathbf{\Sigma}$. Substituting this back into the sum:

$$\mathbb{E}(\mathbf{A}) = \sum_{i=1}^n \mathbf{\Sigma} = n\mathbf{\Sigma}$$

The proof is complete.

Question 24

Question: Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Let $\bar{\mathbf{X}}$ and $\mathbf{S}$ be the sample mean and covariance matrix. Prove or disprove that the distribution of $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$ converges to a chi-square distribution as $n \to \infty$.

Solution: True. 1. By the multivariate Central Limit Theorem, we know that $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ converges in distribution to a multivariate normal random vector $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. 2. By the Law of Large Numbers, the sample covariance matrix $\mathbf{S}$ is a consistent estimator of the true covariance matrix $\mathbf{\Sigma}$. This means $\mathbf{S}$ converges in probability to $\mathbf{\Sigma}$ as $n \to \infty$, and therefore $\mathbf{S}^{-1} \xrightarrow{p} \mathbf{\Sigma}^{-1}$. 3. Let $\mathbf{Y}_n = \sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$. The statistic is $\mathbf{Y}_n^T \mathbf{S}^{-1} \mathbf{Y}_n$. 4. By Slutsky's Theorem, if $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}$ and $\mathbf{S}^{-1} \xrightarrow{p} \mathbf{\Sigma}^{-1}$, then the joint distribution of $(\mathbf{Y}_n, \mathbf{S}^{-1})$ converges to that of $(\mathbf{Y}, \mathbf{\Sigma}^{-1})$. 5. By the continuous mapping theorem, the distribution of the statistic converges to the distribution of $\mathbf{Y}^T \mathbf{\Sigma}^{-1} \mathbf{Y}$. 6. This resulting quadratic form, where $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, is the definition of a chi-square random variable with p degrees of freedom, χ^2_p.

Question 25

Question: Show that an approximate distribution of $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$ is a χ^2-distribution with p-degrees of freedom for a large $n - p$, where $\bar{\mathbf{X}}$ is the sample mean vector of a random sample of size n from any population having covariance matrix $\mathbf{\Sigma}$.

Solution: This is a direct application of the multivariate Central Limit Theorem. 1. Let $\mathbf{Y}_n = \sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$. The CLT states that as $n \to \infty$, the distribution of $\mathbf{Y}_n$ approaches a multivariate normal distribution $N_p(\mathbf{0}, \mathbf{\Sigma})$, regardless of the underlying population distribution (as long as it has a finite covariance matrix). 2. The statistic in question is $T_n = (\sqrt{n}(\bar{\mathbf{X}}-\boldsymbol{\mu}))^T\boldsymbol{\Sigma}^{-1}(\sqrt{n}(\bar{\mathbf{X}}-\boldsymbol{\mu})) = \mathbf{Y}_n^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_n$. 3. By the continuous mapping theorem, as $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y} \sim N_p(\mathbf{0},\boldsymbol{\Sigma})$, the distribution of the continuous function $g(\mathbf{Y}_n) = \mathbf{Y}_n^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}_n$ converges to the distribution of $g(\mathbf{Y}) = \mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y}$. 4. Let $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}\mathbf{Y} \sim N_p(\mathbf{0}, I)$. Then $\mathbf{Y}^T\boldsymbol{\Sigma}^{-1}\mathbf{Y} = (\boldsymbol{\Sigma}^{1/2}\mathbf{Z})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{1/2}\mathbf{Z}) = \mathbf{Z}^T\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{Z} = \mathbf{Z}^TI\mathbf{Z} = \mathbf{Z}^T\mathbf{Z}$. 5. $\mathbf{Z}^T\mathbf{Z} = \sum_{i=1}^p Z_i^2$, where Z_i are independent $N(0,1)$ variables. This is the definition of a chi-square distribution with p degrees of freedom. Thus, for large n, the statistic is approximately distributed as χ_p^2.

Question 26

Question: Suppose $\binom{1}{2}, \binom{2}{-2}, \binom{-2}{1}, \binom{1}{-3}, \binom{-2}{3}$ are five observations from a 2-dimensional Normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Use the above data to explain what is Wishart distribution.

Solution: The Wishart distribution describes the distribution of the sample sum of squares and cross-products (SSCP) matrix. 1. **Calculate the sample mean:** $\bar{x}_1 = (1+2-2+1-2)/5 = 0$. $\bar{x}_2 = (2-2+1-3+3)/5 = 0.2$. $\bar{\mathbf{x}} = \binom{0}{0.2}$. 2. **Calculate the deviation vectors:** $\mathbf{d}_1 = \binom{1}{1.8}, \mathbf{d}_2 = \binom{2}{-2.2}, \mathbf{d}_3 = \binom{-2}{0.8}, \mathbf{d}_4 = \binom{1}{-3.2}, \mathbf{d}_5 = \binom{-2}{2.8}$. 3. **Calculate the SSCP matrix:** The SSCP matrix is $A = \sum_{i=1}^5 \mathbf{d}_i\mathbf{d}_i^T$. For example, $\mathbf{d}_1\mathbf{d}_1^T = \binom{1}{1.8}\binom{1}{1.8}^T = \begin{pmatrix} 1 & 1.8 \\ 1.8 & 3.24 \end{pmatrix}$. Summing these five outer product matrices gives the observed SSCP matrix A. 4. **Explain the Distribution:** If the original data came from a $N_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution, then the matrix A that we calculated is a single realization from a Wishart distribution with $n-1=4$ degrees of freedom and scale matrix $\boldsymbol{\Sigma}$. We write this as $A \sim W_2(4,\boldsymbol{\Sigma})$.

Question 27

Question: Three observations from $\mathcal{N}_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$ are (1,2), (1.2, 1.8), (2,3). Find a 95% minimum volume confidence region for $\boldsymbol{\mu}$.

Solution: First calculate $\bar{\mathbf{x}} = (1.4, 2.267)^T$ and $\mathbf{S} = \begin{pmatrix} 0.28 & 0.38 \\ 0.38 & 0.523 \end{pmatrix}$. The confidence region is given by $n(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{p(n-1)}{n-p}F_{p,n-p,\alpha}$. Here $n=3, p=2, \alpha=0.05$. We need $F_{2,1,0.05} = 199.5$. The critical value is $\frac{2(2)}{1}\cdot 199.5 = 798$. The confidence region is the ellipse defined by $3((\bar{\mathbf{x}}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})) \le 798$.

Question 28

Question: Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be a random sample from $\mathcal{N}_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$. For testing $H_0 : \boldsymbol{\mu} = (1,2,3)^T$, construct a test based on Hotelling's T^2 statistic and write its distribution under H_0.

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

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

where $\boldsymbol{\mu}_0 = (1,2,3)^T$. Under H_0, this statistic is distributed as a scaled F-distribution:

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

Here, $p=3$, so the distribution is $\frac{n-3}{3(n-1)}T^2 \sim F_{3,n-3}$. We reject H_0 for large values of T^2.

Question 29

Question: Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be a random sample from $\mathcal{N}_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$. For testing $H_0 : \mu_1 = \mu_2 - \mu_3$, construct a test based on Hotelling's T^2 statistic and write its distribution under H_0.

Solution: This hypothesis can be written in the form $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where $\mathbf{C}$ is a contrast matrix, $\mathbf{C} = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix}$. Let $\mathbf{Y}_i = \mathbf{C}\mathbf{X}_i$. Then $\bar{\mathbf{Y}} = \mathbf{C}\bar{\mathbf{X}}$ and the sample covariance is $S_Y = \mathbf{C}S\mathbf{C}^T$. Since the linear combination is a scalar, this reduces to a one-sample test for a univariate mean. The Hotelling's T^2 statistic for this hypothesis is:

$$T^2 = n(\mathbf{C}\bar{\mathbf{X}}-\mathbf{0})^T(S_Y)^{-1}(\mathbf{C}\bar{\mathbf{X}}-\mathbf{0}) = n(\mathbf{C}\bar{\mathbf{X}})^T(\mathbf{C}S\mathbf{C}^T)^{-1}(\mathbf{C}\bar{\mathbf{X}})$$

This is equivalent to a standard one-sample t-test, where $t = \frac{\mathbf{C}\bar{\mathbf{X}}}{\sqrt{S_Y/n}}$. The distribution of the T^2 statistic is $\frac{n-1}{1(n-1)}T^2 \sim F_{1,n-1}$, which is the square of a t-distribution with $n-1$ degrees of freedom.

Question 30

Question: For sample data with $\bar{\mathbf{x}} = (1,0,2)^T$ and $S = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}$ from a sample of size 10, can we assert that $\mu_1 = (\mu_2 + \mu_3)/2$? Test at a 5% level.

Solution: The hypothesis is $H_0 : \mu_1 - 0.5\mu_2 - 0.5\mu_3 = 0$. This is of the form $\mathbf{C}\boldsymbol{\mu} = 0$ with $\mathbf{C} = (1, -0.5, -0.5)$. We construct a t-test. $\mathbf{C}\bar{\mathbf{x}} = 1 - 0 - 1 = 0$. $\mathbf{C}S\mathbf{C}^T = (1, -0.5, -0.5)S(1, -0.5, -0.5)^T = 2.25$. The t-statistic is $t = \frac{\sqrt{n}\mathbf{C}\bar{\mathbf{x}}}{\sqrt{\mathbf{C}S\mathbf{C}^T}} = \frac{\sqrt{10}\cdot 0}{\sqrt{2.25}} = 0$. The critical value is $t_{9,0.025} = 2.262$. Since $|t| < 2.262$, we do not reject H_0. We can assert that the relationship holds.

Question 31

Question: Define the 2-sample Hotelling T^2-statistic for testing equality of mean vectors from two normal populations, and find its distribution.

Solution: Let there be two independent random samples of size n_1 and n_2 from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ respectively. The test statistic for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T S_{pooled}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$$

where $S_{pooled} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$ is the pooled covariance matrix. The distribution of the statistic under H_0 is:

$$\frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 \sim F_{p, n_1+n_2-p-1}$$

Question 32

Question: For 42 observations on two variables, $\bar{\mathbf{x}} = \binom{0.564}{0.603}$ and $\mathbf{S} = \begin{pmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{pmatrix}$. Test $H_0 : \boldsymbol{\mu} = (0.5, 0.6)^T$ at $\alpha = 0.05$.

Solution: $n = 42, p = 2, \boldsymbol{\mu}_0 = (0.5, 0.6)^T$. $\bar{\mathbf{x}} - \boldsymbol{\mu}_0 = \binom{0.064}{0.003}$. $S^{-1} = \frac{1}{0.00007323}\begin{pmatrix} 0.0146 & -0.0117 \\ -0.0117 & 0.0144 \end{pmatrix}$. $T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T S^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \approx 42 \cdot 13655 \cdot (0.0034) \approx 1948$. The critical value is $\frac{p(n-1)}{n-p} F_{p,n-p,\alpha} = \frac{2(41)}{40} F_{2,40,0.05} \approx 2.05 \cdot 3.23 = 6.62$. Since T^2 is much larger than the critical value, we reject H_0.

Question 33

Question: For an iid sample of size n=5 from an $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution with $\boldsymbol{\Sigma} = \begin{pmatrix} 3 & \eta \\ \eta & 1 \end{pmatrix}$, where η is known, and $\bar{\mathbf{x}} = \binom{1}{0}$. For what value of η would $H_0 : \boldsymbol{\mu} = (0,0)^T$ be rejected at the 5% level?

Solution: The test statistic is $T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi^2_p$. $T^2 = 5\binom{1}{0}^T \frac{1}{3-\eta^2}\begin{pmatrix} 1 & -\eta \\ -\eta & 3 \end{pmatrix}\binom{1}{0} = \frac{5}{3-\eta^2}$. The critical value is $\chi^2_{2,0.05} = 5.99$. We reject if $T^2 > 5.99 \implies \frac{5}{3-\eta^2} > 5.99 \implies 5 > 5.99(3 - \eta^2) \implies 0.8347 > 3 - \eta^2 \implies \eta^2 > 2.165$. So, $|\eta| > \sqrt{2.165} \approx 1.47$. Also note that for Σ to be positive definite, $|\Sigma| = 3 - \eta^2 > 0$, so $|\eta| < \sqrt{3} \approx 1.732$. Thus, the hypothesis is rejected if $1.47 < |\eta| < 1.732$.