APPLIED MULTIVARIATE ANALYSIS

FREQUENTLY ASKED QUESTIONS

AMIT MITRA & SHARMISHTHA MITRA

DEPARTMENT OF MATHEMATICS & STATISTICS,
INDIAN INSTITUTE OF TECHNOLOGY KANPUR.

[1] The variance covariance matrix of a 3-dimensional random vector $\underset{\sim}{X} = (X_1, X_2, X_3)'$ is given by

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

(a) Find the correlation matrix.

(b) Find the correlation between X_1 and $\frac{X_2}{2} + \frac{X_3}{2}$.

[2] Suppose $\underset{\sim}{X} = (X_1, ..., X_P)'$ is a p-dimensional random vector with $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ and $Cov(\underset{\sim}{X}) = \Sigma$. Find the covariance matrix of the random vector $\underset{\sim}{Z} = \left(\underset{\sim}{c}_1' \underset{\sim}{X}, ..., \underset{\sim}{c}_k' \underset{\sim}{X}\right)$; where $\underset{\sim}{c}_j' \in \Re^P$ are vectors of constants.

[3] Show that $|S| = s_{11}.....s_{pp}|R|$, where S is the sample variance covariance matrix and R is the sample correlation matrix.

[4] Suppose the random vector $\underset{\sim}{X}$ is such that $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ and $Cov(\underset{\sim}{X}) = \Sigma$. Find $E\left(\underset{\sim}{X}\, \underset{\sim}{X}'\right)$. Let $\underset{\sim}{Y}$ be another random vector with $E(\underset{\sim}{Y}) = \underset{\sim}{\delta}$ and $Cov(\underset{\sim}{X}, \underset{\sim}{Y}) = \Sigma_{XY}$. Derive $E\left(\underset{\sim}{Y}\, \underset{\sim}{X}'\right)$.

[5] Suppose the observed data matrix for a 3-dimensional random vector is given by

$$\mathcal{X} = \begin{pmatrix} -1 & 2 & 5 \\ 3 & 4 & 2 \\ -2 & 2 & 3 \end{pmatrix}.$$

(a) For the observations on variable X_1, find the projection on $\underset{\sim}{1}$.

(b) Find the deviation vectors and link them with the sample standard deviations.

(c) Calculate the angle between the deviation vectors $\underset{\sim}{d}_1$ and $\underset{\sim}{d}_2$.

(d) Using the deviation vectors $\underset{\sim}{d}_1$, $\underset{\sim}{d}_2$ and $\underset{\sim}{d}_3$, find $\mathcal{X} - \underset{\sim}{\bar{x}}\, \underset{\sim}{1}'$ and verify whether it is of full rank.

(e) Find the generalized sample variance and the total sample variance.

[6] Suppose the mean vector and covariance matrix of $\underset{\sim}{X} = \left(X_1, X_2, X_3, X_4\right)'$ is given by

$$\underset{\sim}{\mu} = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \ \Sigma = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix}.$$

Let $\underset{\sim}{X}_{(1)} = \left(X_1, X_3\right)'$ and $\underset{\sim}{X}_{(2)} = \left(X_2, X_4\right)'$ be 2 subvectors; $A = \begin{pmatrix} 1, & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}$.

(a) Find $Cov\left(A \underset{\sim}{X}_{(1)}\right)$, $Cov\left(B \underset{\sim}{X}_{(2)}\right)$ and $Cov\left(A \underset{\sim}{X}_{(1)}, B \underset{\sim}{X}_{(2)}\right)$.

(b) Find the joint distribution of $A \underset{\sim}{X}_{(1)}$ and $B \underset{\sim}{X}_{(2)}$ if $\underset{\sim}{X} \sim N_4\left(\underset{\sim}{\mu}, \Sigma\right)$.

(c) With $\underset{\sim}{X} \sim N_4\left(\underset{\sim}{\mu}, \Sigma\right)$, find the marginal distributions of $\underset{\sim}{X}_{(1)}$ and $\underset{\sim}{X}_{(2)}$ and the conditional distribution of $\underset{\sim}{X}_{(2)}$ given $\underset{\sim}{X}_{(1)}$.

[7] Suppose the covariance matrix of a 3-dimensional random vector $\underset{\sim}{X}$ is given by

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix}; |\rho| < \frac{1}{\sqrt{2}}$$

Suppose the underlying random vector is $N_3\left(\underset{\sim}{0}, \Sigma\right)$, find the joint distribution and the marginal distributions of the principal components.

[8] Determine the population principal components Y_1 and Y_2 for the covariance matrix $\Sigma = \begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}$. Further, find ρ_{Y_1, X_1} and ρ_{Y_1, X_2}.

[9] Let $\underset{\sim}{X}_1, \ldots., \underset{\sim}{X}_n$ be a random sample from $N_p\left(\underset{\sim}{0}, \Sigma\right)$, $\Sigma > 0$. Define the $p \times n$ data matrix X as $X = \left(\underset{\sim}{X}_1, \ldots., \underset{\sim}{X}_n\right)$

(a) Find the distribution of $\underset{\sim}{U}'\left(I_n - \frac{1}{n}\underset{\sim}{1}\underset{\sim}{1}'\right)\underset{\sim}{U}$, where $\underset{\sim}{U} = \left(U_1, \ldots, U_n\right)'$ with $U_i = \underset{\sim}{a}'\underset{\sim}{X}_i$, $i = 1(1)n$, $\underset{\sim}{a} \in \Re^p$, $\underset{\sim}{a} \neq \underset{\sim}{0}$.

(b) Find the distribution of $X\underset{\sim}{b}, \underset{\sim}{b} \in \Re^n$ such that $\underset{\sim}{b}'\underset{\sim}{b} = 1$.

(c) Find the distribution of $\underset{\sim}{b}' X' \Sigma^{-1} X \underset{\sim}{b}$.

[10] Let $Y_0, Y_1, \ldots, Y_p$ be independent and identically distributed random variables with mean 0 and variance σ^2. Define $X_i = Y_0 + Y_i;\, i = 1(1)p$.

(a) Show that there is a principal component of $\underset{\sim}{X} = \left(X_1, \ldots, X_p\right)'$ that is proportional to $\bar{X} = \frac{1}{p}\underset{\sim}{1}'\underset{\sim}{X}$.

(b) Show that the above principal component is in fact the first principal component.

[11] Let $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be a random sample from a p-dimensional multivariate population with mean vector $\underset{\sim}{\mu}$ and covariance matrix Σ. Let $X = \left(\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n\right)$ be the $p \times n$ data matrix. Prove or disprove

$$\text{“}\, n S_n = X\left(I_n - \frac{1}{n}\underset{\sim}{1}_n \underset{\sim}{1}_n'\right)X' \,\text{”}$$

where, S_n is the sample variance covariance matrix with divisor n.

[12] Let $\underset{\sim}{X} \sim N_p\left(\underset{\sim}{0}, \Sigma\right)$, where Σ is a singular matrix of rank $r < p$ and $\exists$ a non singular $p \times p$ matrix $H \ni$

$$H \Sigma H' = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

If B is a g-inverse of Σ, find the distribution of $\underset{\sim}{X}' B \underset{\sim}{X}$.

[13] Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ be i.i.d. $N_m\left(\underset{\sim}{\mu}, \sigma^2 I_m\right)$ and B is $k \times m$ matrix of constants with $B B^T = I_k$.

(a) Find the distribution of

(i) $\sum_{j=1}^{n} B \underset{\sim}{X}_j$, (ii) $\sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)^T\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)$ and (iii) $\sum_{j=1}^{n}\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)\left(\underset{\sim}{X}_j - \underset{\sim}{\mu}\right)^T$.

(b) Let $\underset{\sim}{Y} = B \underset{\sim}{X}_n$, find the distribution of $Z = \left(\underset{\sim}{X}_n^T \underset{\sim}{X}_n - \underset{\sim}{Y}^T \underset{\sim}{Y}\right)$. Are Z and $\underset{\sim}{X}_n$ independent? Are Z and $\underset{\sim}{Y}$ independent?

[14] Let $\underset{\sim}{X}_i,\ i = 1, \ldots, n$ be independently distributed as $N_P(\underset{\sim}{\mu}, \Sigma)$. Find the distribution of $\sum_{i=1}^{n} a_i \underset{\sim}{X}_i$; where $a_1, \ldots, a_n$ are real.

[15] Let $\underset{\sim}{X} = \left(X_1, X_2, X_3\right)'$ be distributed as $N_3(\underset{\sim}{\mu}, \Sigma)$,

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}; \ -1/2 < \rho < 1.$$

[16] Find the joint probability density function of $\left(X_1 + X_2, X_1 - X_2\right)'$.

[17] Let $\underset{\sim}{X} \sim N_2\left(\underset{\sim}{\mu}, \Sigma\right)$ with $\underset{\sim}{\mu}$ and $\Sigma = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$. Find the distribution of

$$Y = X_1^2 + \frac{3}{2} X_2^2 - 2 X_1 X_2 .$$

[18] Suppose $\underset{\sim}{Y} \sim N_n\left(X \underset{\sim}{\mu}, I_n\right)$, where X is a $n \times p$ matrix of constants and $\underset{\sim}{\mu}$ is a $p \times 1$ vector of constants. Find the distribution of $\underset{\sim}{Y}' \left(I_n - X\left(X'X\right)^{-1} X'\right)\underset{\sim}{Y}$.

[19] Suppose $\underset{\sim}{X} \sim N_2\left(\underset{\sim}{\mu}, \Sigma\right)$ with $\underset{\sim}{\mu} = (2,2)'$ and $\Sigma = I_2$. Consider $A = (1,1)$ and $B = (1,-1)$. Verify whether $A\underset{\sim}{X}$ and $B\underset{\sim}{X}$ are independent.

[20] Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ be a random sample from a population which is $N_P(\underset{\sim}{\mu}, \Sigma)$.

(a) Derive the sufficient statistic for $\underset{\sim}{\mu}$ when $\Sigma = \Sigma_0$ is known.

(b) Derive the sufficient statistic for Σ when $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$.

(c) Check whether the derived sufficient statistic are unbiased estimators for the corresponding unknown parameters.

[21] Suppose that the distribution of the $m \times m$ random matrix A is Wishart $W_m\left(n, \Sigma\right)$, $\Sigma > 0$. Let A & Σ be partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \& \, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

A_{11} & Σ_{11} are k x k, A_{12} & Σ_{12} are k x m-k, A_{21} & Σ_{21} are m-k x k and A_{22} & Σ_{22} are m-k x m- k. Find the distributions of A_{11} & A_{22}.

[22] Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ be a random sample from a population which is $N_m(\underset{\sim}{\mu}, \Sigma)$. Define the $n \times m$ data matrix as

$$X = \begin{pmatrix} \underset{\sim}{X}_1^{'} \\ \vdots \\ \underset{\sim}{X}_n^{'} \end{pmatrix}.$$

Prove that $n\,S_n = \left(X - \underset{\sim}{1}\,\overline{\underset{\sim}{X}}^{'}\right)^{'}\left(X - \underset{\sim}{1}\,\overline{\underset{\sim}{X}}^{'}\right).$

[23] Suppose that the distribution of the $m \times m$ random matrix A is Wishart $W_m(n, \Sigma)$, $\Sigma > 0$. Let Φ be an $m \times m$ symmetric matrix of full rank. Prove that

$$E\left(\exp\left(tr\frac{i}{2}A\Phi\right)\right) = \prod_{j=1}^{m}\left(1 - i\,\lambda_j\right)^{-n/2},$$

where, $\lambda_1, ..., \lambda_m$ are the eigen values of $\Sigma^{1/2}\,\Phi\,\Sigma^{1/2}$.

[24] Let A be a Wishart $W_m(n, \Sigma)$. For a $k \times m$ non random matrix of full row rank, M, find the characteristic function of $M\,AM^{'}$.

[25] Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ be a random sample from a population which is $N_m(\underset{\sim}{\mu}, \Sigma)$. Derive the distribution of

$$\overline{\underset{\sim}{X}}^{'}\,S\,\overline{\underset{\sim}{X}} \Big/ \overline{\underset{\sim}{X}}^{'}\,\Sigma\,\overline{\underset{\sim}{X}}.$$

[26] Let A be a Wishart $W_m(n, \Sigma)$. Find an unbiased estimator of Σ^{-1}.

.

[27] Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n$ be a random sample from $N_m(\underset{\sim}{\mu}, \Sigma)$, $\Sigma > 0$. Define a transformation $\underset{\sim}{X} \to \underset{\sim}{Y} = A\,\underset{\sim}{X} + \underset{\sim}{\beta}$, A is a $m \times m$ non singular matrix of constants and $\underset{\sim}{\beta}$ is a vector of constants. Show that Hotelling's T^2 statistic for testing $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ against $H_A : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ based on $X = \left(\underset{\sim}{X}_1, \underset{\sim}{X}_2, ..., \underset{\sim}{X}_n\right)$ and that based on $Y = \left(\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, ..., \underset{\sim}{Y}_n\right)$ are the same.

[28] Suppose $\underset{\sim}{X}_1^{(1)},....,\underset{\sim}{X}_{n_1}^{(1)}$ be a random sample from $N_P\left(\underset{\sim}{\mu}^{(1)},\Sigma\right)$ and $\underset{\sim}{X}_1^{(2)},....,\underset{\sim}{X}_{n_2}^{(2)}$ be a random sample from $N_P\left(\underset{\sim}{\mu}^{(2)},\Sigma\right)$, $\Sigma>0$.

(a) Under the condition that $\underset{\sim}{\mu}^{(1)}=\underset{\sim}{\mu}^{(2)}$, find the distribution of $\left(\overline{\underset{\sim}{X}}_1-\overline{\underset{\sim}{X}}_2\right)' S_p^{-1}\left(\overline{\underset{\sim}{X}}_1-\overline{\underset{\sim}{X}}_2\right)$. Where, $\overline{\underset{\sim}{X}}_1$ and $\overline{\underset{\sim}{X}}_2$ denote the sample mean vectors and S_p is the pooled sample covariance matrix.

(b) Derive the appropriate test statistic based on Hotelling's T^2 statistic for testing $H_0:\underset{\sim}{\mu}^{(1)}=\underset{\sim}{\mu}^{(2)}$ against $H_A:\underset{\sim}{\mu}^{(1)}\neq\underset{\sim}{\mu}^{(2)}$.

(c) Obtain $100(1-\alpha)\%$ confidence regions for $\underset{\sim}{\mu}^{(1)}$, $\underset{\sim}{\mu}^{(2)}$ and $\underset{\sim}{\mu}^{(1)}-\underset{\sim}{\mu}^{(2)}$.

[29] Let $\underset{\sim}{X}_1,\underset{\sim}{X}_2,...,\underset{\sim}{X}_n$ be a random sample from a population which is $N_{2m}\left(\underset{\sim}{\mu},\Sigma\right)$, $\Sigma>0$. Derive the testing procedure for testing $H_0:\mu_i=\mu_{i+m}; i=1(1)m$ against H_A: at least one such relation does not hold.

[30] Let $\underset{\sim}{X}_1,\underset{\sim}{X}_2,...,\underset{\sim}{X}_n$ be a random sample from $N_m\left(\underset{\sim}{\mu},\Sigma\right)$, $\Sigma=diag\left(\sigma_{11},...,\sigma_{mm}\right)$. Obtain a simultaneous confidence interval for $\mu_1-\mu_2$ and $\mu_1+\mu_2$, such that the joint confidence is exactly $100(1-\alpha)\%$.

[31] Let $\underset{\sim}{X}_1,\underset{\sim}{X}_2,...,\underset{\sim}{X}_n$ be a random sample from $N_m\left(\underset{\sim}{\mu},\Sigma\right)$, $\Sigma>0$. Using Bonferroni's approach, construct simultaneous confidence intervals of confidence level at least 90% for $\mu_1-\mu_m$ and $\mu_2-\mu_{m-1}$ under the following scenarios;

(a) the two contrasts are given equal importance and

(b) importance of the contrast $\mu_1-\mu_m$ is three times that of the contrast $\mu_2-\mu_{m-1}$.