Multivariate Statistical Analysis Mid Term

November 14, 2011

2 pages, 16 problems, 100 points

1. Suppose that a sample covariance matrix $\mathbf{S}=\begin{bmatrix}1 & -2 & 0\\ -2 & 5 & 0\\ 0 & 0 & 2\end{bmatrix}$, calculate its generalized variance and total sample variance. (6%)
2. Assume that a ample covariance matrix $\mathbf{S}=\begin{bmatrix}13 & -4 & 2\\ -4 & 13 & -2\\ 2 & -2 & 10\end{bmatrix}$, calculate the cosine of the angle between the two deviation vectors $\mathbf{d}_1=\mathbf{y}_1-\bar{x}_1\mathbf{1}$ and $\mathbf{d}_2=\mathbf{y}_2-\bar{x}_2\mathbf{1}$. (6%)
3. Verify that $\lambda_1=9$, $\lambda_2=9$, $\lambda_3=18$ and $\mathbf{e}_1'=\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\end{bmatrix}$, $\mathbf{e}_2'=\begin{bmatrix}1/3\sqrt{2} & -1/3\sqrt{2} & -4/3\sqrt{2}\end{bmatrix}$, $\mathbf{e}_3'=\begin{bmatrix}2/3 & -2/3 & 1/3\end{bmatrix}$ are the eigenvalues and eigenvectors of the sample covariance matrix given in Problem 2. (6%)
4. Compute the major axes and directions formed by the ellipsoid $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})=c^2$, where $\mathbf{x}'=\begin{bmatrix}x_1 & x_2 & x_3\end{bmatrix}$, $\boldsymbol{\mu}'=\begin{bmatrix}0 & 1 & 1\end{bmatrix}$, $\boldsymbol{\Sigma}=\begin{bmatrix}13 & -4 & 2\\ -4 & 13 & -2\\ 2 & -2 & 10\end{bmatrix}$, and c is a positive real constant. Use the results of Problem 3. (6%)
5. Find the major axes and directions of the ellipsoid which contains 95% of the probability for multivariate normal distribution $N_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ have been defined in Problem 4. Use the results of Problem 3. (6%)
6. Set up a null hypothesis $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$ and an alternative hypothesis $H_1:\boldsymbol{\mu}\neq\boldsymbol{\mu}_0$, with $\boldsymbol{\mu}_0'=\begin{bmatrix}0 & 1 & 1\end{bmatrix}$ for multivariate normal distribution $N_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Determine if the Hotelling's T^2 test can or can not reject the null hypothesis at the level of significance level $\alpha=0.05$. Suppose that the sample size is 43, the sample mean $\bar{\mathbf{x}}'=\begin{bmatrix}0.5 & 0.5 & -0.5\end{bmatrix}$, and the sample covariance matrix $\mathbf{S}$ has been given in Problem 2. You may apply $\mathbf{S}^{-1}=\frac{1}{162}\begin{bmatrix}14 & 4 & -2\\ 4 & 14 & 2\\ -2 & 2 & 17\end{bmatrix}$ (6%)
7. Find the major axes and directions of the Hotelling's 95% T^2 confidence region for Problem 6. (6%)
8. Determine the 95% simultaneous T^2 confidence intervals for Problem 6. (6%)
9. Determine the 95% Bonferroni simultaneous confidence intervals for Problem 6. Note that if your t corresponding to the desired degrees of freedom can not be found directly in the table, use the following linear interpolation formula to get an approximate value: $t=t_1+m(\nu-\nu_1)$, $m=(t_2-t_1)/(\nu_2-\nu_1)$, where t_1 and t_2 are the two t values corresponding to degrees of freedom ν_1 and ν_2, respectively,

with ν_1 and ν_2 being the two ν values in the table and closest to the desired ν. (6%)

10. Determine the lower control limit (LCL) and the 95% upper control limit (LCL) of the T^2 control chart for future observation which is based on a sample of size n = 43 from a multivariate normal population. Assume that the sample mean is $\bar{\mathbf{x}}' = [0 \quad 1 \quad 1]$ and the sample covariance matrix is the **S** given in Problem 2. (6%)
11. Given that a random vector $\mathbf{X}$ follows $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ have been defined in Problem 4. Find the probability distribution of $\mathbf{a}'\mathbf{X}$ for $\mathbf{a}' = [1 \quad 1 \quad 2]$. (6%)
12. Let $\mathbf{X}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [0 \quad 1 \quad 1]$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Which of the following random samples are independent? Explain. (a) X_1 and X_2 (3%) (b) (X_1, X_2) and X_3 (3%)
13. Let x have an exponential density $p(x) = \begin{cases} \theta e^{-\theta x}, x \geq 0 \\ 0, \quad x < 0 \end{cases}$. Suppose that n random samples x_1, …, x_n are drawn. Find the maximum-likelihood estimate $\hat{\theta}$ for θ. (6%)
14. Assume a null hypothesis $H_0 : \theta = \theta_0$ for the random number in Problem 13. Find the likelihood ratio Λ in terms of $\theta_0 / \hat{\theta}$. (6%)
15. Given the data $\mathbf{X} = \begin{bmatrix} 2 & 6 \\ 4 & 4 \\ - & 8 \end{bmatrix}$ with missing components, use the prediction-estimation algorithm (the EM algorithm) to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Determine the initial estimate and iterate to find the *first* revised estimates. (8%)
16. Given the fact that if a random variable x is $N(\mu, \sigma^2)$, then $\frac{(n-1)s^2}{\sigma^2}$ follows χ^2_{n-1} distribution. Use the probability statement $\Pr\left\{ \chi^2_{n-1}(\frac{\alpha}{2}) < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{n-1}(1-\frac{\alpha}{2}) \right\} = 1-\alpha$ to find the confidence interval for the population variance σ^2. (8%)