Multivariate Statistical Analysis Mid Term

November 10, 2008

2 pages, 17 problems, 100 points

1. A study of a patent medicine, *MagicPill*, for basketball players has been conducted. This medicine is claimed to enable Taiwanese basketball players to jump higher. A sample of 16 collegiate players had taken *MagicPill* for 2 weeks, and can jump an average height of 56 cm, with a sample standard deviation 15 cm. Find the 95% confidence interval of the population mean for all collegiate basketball players in Taiwan if they take *MagicPill*. (6%)
2. Explain the meaning of the 95% confidence interval obtained in Problem 1. (6%)
3. Following Problem 1, assume that population data gathered by university coaches across Taiwan show a normal jump height would be 50 cm, with a standard deviation 15 cm. How large a sample will we need to have a 90% power of detecting the Problem 1 distance of jump heights, using α= 0.05 as a critical value. In this problem, we may assume that the jump height is with a normal distribution. (6%)
4. Explain the geometrical meaning of sample correlation coefficient $r_{ik} = \dfrac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}}$ for an $n \times p$ data array $\mathbf{X} = \lfloor \mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_p \rfloor$. (6%)
5. Explain the geometrical meaning of the generalized sample variance $|\mathbf{S}|$ for an $n \times p$ data array $\mathbf{X} = \lfloor \mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_p \rfloor$. (6%)
6. Show that the ellipse equation $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2$ can be reduced to $\dfrac{y_1^2}{c^2\lambda_1} + \dfrac{y_2^2}{c^2\lambda_2} = 1$, $\mathbf{y} = [y_1 \quad y_2]' = \begin{bmatrix} \mathbf{e}_1' \\ \mathbf{e}_2' \end{bmatrix}(\mathbf{x}-\boldsymbol{\mu})$, where $\mathbf{x}$, $\boldsymbol{\mu}$ are two-dimensional vectors, λ_1. λ_2 and $\mathbf{e}_1$, $\mathbf{e}_2$ are the eigenvalues and corresponding eigenvectors of the 2×2 matrix $\boldsymbol{\Sigma}$, respectively. (6%)
7. Verify that $\lambda_1 = 16$, $\lambda_2 = 3$ and $\mathbf{e}_1' = \dfrac{1}{\sqrt{13}}[3 \quad 2]$, $\mathbf{e}_2' = \dfrac{1}{\sqrt{13}}[-2 \quad 3]$ are the eigenvalues and corresponding eigenvectors of $\begin{bmatrix} 12 & 6 \\ 6 & 7 \end{bmatrix}$, respectively. (4%)
8. 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 $\boldsymbol{\mu}_0' = [-1 \quad 1]$ for bivariate normal distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Show that the

Hotelling's T^2 test can not reject the null hypothesis at level of significance $\alpha = 0.10$. Suppose that the sample size is 30, and the sample mean $\overline{\mathbf{x}}' = [-0.5 \quad 0.5]$, the sample covariance matrix $\mathbf{S} = \begin{bmatrix} 12 & 6 \\ 6 & 7 \end{bmatrix}$. (6%)

9. Find the semi-major and semi-minor axes and directions of the Hotelling's 90% T^2 confidence region for Problem 8. (6%)
10. Explain the meaning of the 90% T^2 confidence region in Problem 9. (6%)
11. Determine the 90% simultaneous T^2 confidence intervals for Problem 8. (6%)
12. Explain the meaning of the 90% simultaneous T^2 confidence intervals in Problem 11. (6%)
13. Determine the 90% Bonferroni simultaneous confidence intervals for Problem 8. (6%)
14. Assume that $x_1, x_2, \cdots, x_n$ are independent samples drawn from a univariare normal population. For these samples, starting from the definition of likelihood, prove that the maximum likelihood estimates of the population mean $\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, variance $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$, and the corresponding maximum likelihood $\max_{\mu,\sigma^2} L(\mu,\sigma^2) = \frac{1}{(2\pi)^{n/2}\hat{\sigma}^n} e^{-n/2}$. Don't get the answer directly from the multivariate results given in the textbook. (6%)
15. Consider a null hypothesis $H_0 : \mu = \mu_0$ and its alternative hypothesis $H_1 : \mu \neq \mu_0$. Let samples $x_1, x_2, \cdots, x_n$ be those in Problem 14. Starting from the definition of likelihood, prove that the maximum likelihood estimate of the population variance under H_0 is $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu_0)^2$ and the corresponding maximum likelihood under H_0 is $\max_{\sigma^2} L(\sigma^2) = \frac{1}{(2\pi)^{n/2}\hat{\sigma}_0^n} e^{-n/2}$. Don't get the answer directly from the multivariate results given in the textbook. (6%)
16. Define the ratio of $\max_{\sigma^2} L(\sigma^2)$ in Problem 15 to $\max_{\mu,\sigma^2} L(\mu,\sigma^2)$ in Problem 14 as the likelihood ratio Λ. Using the results of Problems 14 and 15, prove that $\Lambda^{2/n} = \left(1 + \frac{t^2}{n-1}\right)^{-1}$, $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$., $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$. Don't get the answer directly from the multivariate results given in the textbook. (6%)
17. If the null hypothesis H_0 in Problem 15 is to be rejected at significance level α. What value must be set as the critical value for the Λ in Problem 16? Don't get the answer directly from the multivariate results given in the textbook. (6%)