Nanyang Technological University, SPMS, MAS

MH4501 Multivariate Analysis, Semester 2 AY 2022-23

Assignment 2 (Due date: 4:30pm, February 23, 2023)

Remark: NTU places very high importance on honesty in academic work submitted by students, and adopts a policy of **ZERO** tolerance on cheating and plagiarism.

1. Let $x_1, \ldots, x_n$ be a random sample of size n from a p-variate normal distribution having mean μ and covariance Σ. Let $C \in \mathbb{R}^{q\times p}$ and $\beta \in \mathbb{R}^q$.

   (a) What is the distribution of $C\bar{x}$?

   (b) What is the distribution of $n(C\bar{x} - C\mu)'(C\Sigma C')^{-1}(C\bar{x} - C\mu)$?

   (c) What is the distribution of $n(C\bar{x} - C\mu)'(CSC')^{-1}(C\bar{x} - C\mu)$?

   (d) What is the approximate distribution of $n(C\bar{x} - C\mu)'(CSC')^{-1}(C\bar{x} - C\mu)$? (consider n is large)

   (e) We examine the hypothesis testing on linear combinations of μ. Consider the test statistic

   $$T^2 = n(C\bar{x} - \beta)'(CSC')^{-1}(C\bar{x} - \beta)$$

   for testing $H_0: \ C\mu = \beta$ *v.s.* $H_1: \ C\mu \neq \beta$. Describe the testing rules for

   i. n is small/moderate (general);

   ii. n is large.

2. Use the college test data (Table 5.2 in textbook), attached with this assignment ("T5-2.dat").

   (a) Find $n, \ p, \ \bar{x}, \ S$. You can answer with the aid of computer, but do not paste your code.

   (b) Test the null hypothesis $H_0: \ \mu = (525, 52, 25)'$ versus $H_1: \ \mu \neq (525, 52, 25)'$ at the $\alpha = 0.05$ significance level. Suppose $(525, 52, 25)'$ represent average scores for thousands of college students over the last 10 years. Is there reason to believe that the group of students represented by the scores in this dataset is scoring differently?

   (c) Give a 95% confidence region for μ (in terms of $\bar{x}$ and S).

   (d) Determine the lengths and directions for the axes of the 95% confidence region for μ.

   (e) Calculate Bonferroni simultaneous 95% confidence intervals for $\mu_i, \ i = 1, 2, 3$.

3. We still use the college test data, but consider the **large** n asymptotic statistics. Try to use the sample statistics in Q2(a) and not to rely on computer.

   (a) Test the null hypothesis $H_0: (\mu_1, \mu_3) = (525, 25)$ versus $H_1: (\mu_1, \mu_3) \neq (525, 25)$ at the $\alpha = 0.01$ significance level.

   (b) Give a 99% confidence ellipse for (μ_1, μ_3).

   (c) Calculate Bonferroni simultaneous 99% confidence intervals for μ_1 and μ_3.

4. In the first phase of a study of the cost of transporting milk from farms to dairy plants, a survey was taken of firms engaged in milk transportation. Cost data on X_1=fuel, X_2=repair, and X_3=capital, all measured on a per-mile basis, are stored in the attached data "T6-10.dat" for $n_1 = 36$ gasoline and $n_2 = 23$ diesel trucks.

   (a) Test for differences in the mean cost vectors. Set $\alpha = 0.01$.

   (b) Construct 99% simultaneous confidence intervals for the pairs of mean components. Which costs, if any appear to be quite different?

5. Researchers have suggested that a change in skull size over time is evidence of the interbreeding of a resident population with immigrant populations. Four measurements were made of male Egyptian skulls for three different time periods: period 1 is 4000 B.C., period 2 is 3300 B.C., and period is 1850 B.C. The data "T6-13.dat" are attached. The measured variables are (unit is mm)

$$X_1 = maixmum\ breadth\ of\ skull; \quad X_2 = basibregmatic\ height\ of\ skull;$$
$$X_3 = basialveolar\ length\ of\ skull; \quad X_4 = nasal\ height\ of\ skull.$$

   (a) Construct a MANOVA of the Egyptian skull data. Use $\alpha = 0.05$.

   (b) Construct 95% simultaneous confidence intervals to determine which mean components differ among the populations represented by the 3 periods.

[Optional Survey] How do you rate the difficulty of this assignment (1=easy, 5=hard)? Any feedback about lectures, assignments, tutorials, and labs so far?