Homework: Multivariate Inference

This assignment consists of three problems covering: 1. Multivariate mean testing, 2. Familywise (simultaneous) confidence intervals, and 3. Multivariate analysis of variance (MANOVA).

Unless otherwise stated, assume all random vectors are normally distributed and that covariance matrices are unknown.

Problem 1 — Multivariate Mean Testing (Hotelling's T^2)

Let $X_1, \dots, X_n$ be i.i.d. $\mathcal{N}_p(\mu, \Sigma)$ with unknown Σ, where $p = 3$ and $n = 40$.

You observe the following summary statistics:

$$\bar{X} = \begin{pmatrix} 1.20 \\ 0.85 \\ -0.40 \end{pmatrix}, \qquad S = \begin{pmatrix} 2.10 & 0.60 & -0.30 \\ 0.60 & 1.50 & 0.20 \\ -0.30 & 0.20 & 1.00 \end{pmatrix}.$$

Consider testing

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0, \qquad \mu_0 = \begin{pmatrix} 1.00 \\ 1.00 \\ 0.00 \end{pmatrix}.$$

(a) Write down the Hotelling's T^2 statistic explicitly in terms of $\bar{X}$, S, and μ_0.

(b) Compute the observed value of T^2.

(c) Derive the equivalent F-statistic and state its null distribution.

(d) At significance level $\alpha = 0.05$, state whether H_0 is rejected.

(e) Explain how this hypothesis test relates geometrically to the confidence ellipsoid for μ.

Problem 2 — Familywise Confidence Intervals

Let $X_1, \dots, X_n \sim \mathcal{N}_p(\mu, \Sigma)$ with unknown Σ, where $p = 3$ and $n = 60$. Let $\bar{X}$ and S denote the sample mean and covariance matrix.

You wish to construct **simultaneous** 95% **confidence intervals** for the components of

$$\mu = (\mu_1, \mu_2, \mu_3)^\top.$$

(a) Write down the $(1 - \alpha)$ confidence region for μ based on Hotelling's T^2.

(b) Derive the simultaneous confidence interval for a generic component μ_j, $j = 1, 2, 3$, induced by this confidence region.

(c) Construct Bonferroni-adjusted 95% confidence intervals for each μ_j.

(d) Compare the Hotelling-based intervals and Bonferroni intervals in terms of: - interval length, - use of correlation information, - geometric interpretation.

(e) Under what conditions would the two methods yield identical confidence intervals?

Problem 3 — MANOVA and Linear Hypotheses

A study measures $p = 3$ response variables for subjects assigned to $g = 3$ groups. Each group contains $n_k = 25$ observations.

Let

$$Y_{ik} \sim \mathcal{N}_p(\mu_k, \Sigma), \quad i = 1, \dots, n_k,\ k = 1, 2, 3,$$

with a common covariance matrix Σ.

(a) State the null hypothesis for a one-way MANOVA testing equality of the group mean vectors.

(b) Define the within-group sum of squares and cross-products matrix W and the between-group matrix B.

(c) Write down Wilks' Lambda statistic Λ in terms of W and B.

(d) Explain how Wilks' Lambda is related to the eigenvalues of $W^{-1}B$.

(e) Interpret a rejection of the MANOVA null hypothesis and contrast it with rejecting at least one univariate ANOVA conducted separately for each response variable.

Problem 4

Anderson 6.6

Problem 5

Anderson 6.24

Data for Anderson problems is in data folder on github