

STAT636 - Homework 2

Daniel Osorio - dcosorih@tamu.edu

Department of Veterinary Integrative Biosciences

Texas A&M University

1. Find the maximum likelihood estimates of the 2×1 mean vector μ and the 2×2 covariance matrix Σ based on the random sample

$$X = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

2. Let X_1, X_2, \dots, X_{60} be a random sample of size $n = 60$ from a $N_2(\mu, \Sigma)$ population. Specify each of the following.
 - (a) The distribution of $(X_1 - \mu)' \Sigma^{-1} (X_1 - \mu)$.
 - (b) The distributions of \bar{X} and $\sqrt{n}(\bar{X} - \mu)$.
 - (c) The distribution of $n(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)$
 - (d) The approximate distribution of $n(\bar{X} - \mu)' S^{-1} (\bar{X} - \mu)$
3. Consider the used car data. For each of 10 used cars, we have the numeric variables Age (age of the car) and Price (sale price of car, in \$1,000s)
 - (a) Determine the power transformation $\hat{\lambda}_1$ that makes the x_1 values approximately normal. Construct a Q-Q plot for the transformed data.
 - (b) Determine the power transformation $\hat{\lambda}_2$ that makes the x_2 values approximately normal. Construct a Q-Q plot for the transformed data.
 - (c) Determine the power transformations $\hat{\lambda}' = [\hat{\lambda}_1, \hat{\lambda}_2]$ that make the $[x_1, x_2]$ values approximately multivariate normal. Compare the results with those from above.
4. Consider the **Advertising** data. For each of 200 strategies, we have three numeric variables that influence the sales: **TV**, **radio**, and **Newspaper**.
 - (a) Construct univariate Q-Q plots for each of the three variables. Also make the three pairwise scatterplots. Does the multivariate normal assumption seem reasonable?
 - (b) Determine the 95% confidence ellipsoid for μ . Where is it centered? What are its axes and corresponding half-lengths?
 - (c) Compute 95% T2 simultaneous confidence intervals for the three mean components.
 - (d) Compute 95% Bonferroni simultaneous confidence intervals for the three mean components.
 - (e) Carry out a Hotelling's T^2 test of the null hypothesis $H_0 : \mu' = [150.0, 20.0, 30.0]$ at $\alpha = 0.05$. What is the test statistic, critical value, and the p-value? What is your conclusion regarding H_0 ?
 - (f) Is $\mu' = [150.0, 20.0, 30.0]$ inside the 95% confidence ellipse you computed in part (b)? Is this consistent with your findings in part (e)? Hint: It should be.

- (g) Use the bootstrap to test the same null hypothesis as in part (e), now using this as your test statistic

$$\Lambda = \left(\frac{|S|}{|S_0|} \right)^{n/2},$$

where

$$S = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})'$$

is the sample covariance matrix, and

$$S_0 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})'$$

is the sample covariance matrix under the assumption that H_0 is true. So that all our answers match, first do `set.seed(2)`, and use $B = 500$ bootstrap iterations. What is the p-value?