STP533: Applied Multivariate Analysis

Homework 2

Name: Your name; NetID: Your ID

Due 11:59pm Friday February 14 2025

Question 1

Consider the data matrix

$$\mathbf{X} = \begin{bmatrix} -1 & 3 & -2 \\ 2 & 4 & 2 \\ 5 & 2 & 3 \end{bmatrix}$$

(a) Calculate the matrix of deviations (residuals), $\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T$. Is this matrix of full rank? Explain.

(b) Determine $\mathbf{S}$ and calculate the generalized sample variance $|\mathbf{S}|$. Interpret the latter geometrically.

(c) Using the results in (b), calculate the total sample variance.

(d) Using the results in (b), calculate the generalized sample variance for standardized variables $|\mathbf{R}|$.

Question 2

Use `R` to sketch the solid ellipsoid $(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})$ for the following three matrices

(a) $\mathbf{S} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

(b) $\mathbf{S} = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}$

(c) $\mathbf{S} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

Question 3

Consider the data matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 4 & 3 \\ 6 & 2 & 6 \\ 8 & 3 & 3 \end{bmatrix}$$

and linear combinations $\mathbf{Xb}$ and $\mathbf{Xc}$ with $\mathbf{b} = [1, 1, 1]^T$ and $\mathbf{c} = [1, 2, -3]^T$.

(a) Compute $\mathbf{Xb}$ and $\mathbf{Xc}$ first. Then evaluate their sample means, sample covariance.

(b) Now compute $\bar{\mathbf{X}}$ and $\mathbf{S}_{\mathbf{X}}$ first and use formula (page 10 of lecture 3) to derive the sample means and sample covariance of the linear combinations. Compare them with results from (a).

Question 4

Consider a bivariate normal distribution with $\mu_1 = 1$, $\mu_2 = 3$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ and $\rho_{12} = -0.8$.

(a) Write out the bivariate normal density.

(b) Write out the squared statistical distance expression $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ as a quadratic function of x_1 and x_2.

(c) Write out the probability density function of $X_1 - X_2$.

(d) Write out the conditional density of $X_1 | X_2 = x_2$.

Question 5

If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then what is the distribution of $(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$? Can you prove it?

Question 6

(a) Consider the annual rates of return (including dividens) on the Dow-Jones industrial average for the years 1996-2005. These data, multiplied by 100, are

$$-0.6, 3.1, 25.3, -16.8, -7.1, -6.2, 25.2, 22.6, 26.0.$$

Use these 10 observations to construct a Q-Q plot. Do these data seem to be normally distributed? Explain.

(b) Consider the following data with age x_1, measured in years, and the selling price x_2, measured in thousands of dollars, for $n = 10$ used cars:

x_1	1	2	3	3	4	5	6	8	9	11
x_2	18.95	19.00	17.95	15.54	14.00	12.95	8.94	7.49	6.00	3.99

Calculate the squared statistical distances $(\mathbf{x_i} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x_i} - \bar{\mathbf{x}})$ for $i = 1, \cdots, 10$. Then construct a chi-square plot. Are these data approximately bivariate normal? Explain.