MATH 4630 / 6632 3.0 - Fall 2022
Assignment 3
(Due Date: December 2, 2022)

Question 1: Consider the data given in the EXCEL file tab "q1".

a. State the multivariable linear regression model with all the necessary assumptions.

b. Find the predicted model.

c. Test the significance of the model.

d. Regarless of your result in part (c), test if X_1 is significant? How about X_2?

e. Find a 95% woking Hotelling confidence region for the mean response when $X_1 = 192$ and $X_2 = 152$.

f. Find a 95% woking Hotelling prediction region for a new response when $X_1 = 192$ and $X_2 = 152$.

Question 2: Timm (1975) reported the results of an experiment in which subjects' respond time to "probe words" at five position (Y_1 is at the beginning of the sentence, Y_2 is in the first quartile of the sentence, Y_3 is in the middle of the sentence, Y_4 is in the third quartile of the sentence, and Y_5 is at end of the sentence). The data are recorded in the EXCEL file tab "q2".

a. Use the sample variance and obtain all the principle components.

b. Timm specifically required the reduction in dimension should cover at least 90% of the total variance. How many principle components are needed? Why?

c. Repeat parts (a) and (b) using the sample correlation matrix.

Question 3: Use the data in the EXCEL file tab "q2".

a. Find the canonical correlation between (x_1, x_2, x_3) and (x_4, x_5).

b. Test the significance canonical correlations.

c. Regardless of your answer in part (b), is each canonical correlation individually significant?

Question 4:

a. Show that

$$-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}_1)'\Sigma^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}_1) + \frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu}_x)'\Sigma^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}_2)$$
$$(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)'Sigma^{-1}\underset{\sim}{x} - \frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)'Sigma^{-1}(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2)$$

b. Let

$$\begin{aligned} f_1(x) &= (1 - |x|) \quad \text{for} \quad |x| \leq 1 \\ f_2(x) &= (1 - |x - 0.5|) \quad \text{for} \quad -0.5 \leq x \leq 1.5 \end{aligned}$$

1. Sketch the two densities in the same graph.
2. Identify the classification rule for the case $p_1 = p_2$ and $c(1|2) = c(2|1)$.
3. Identify the classification rule for the case $p_1 = 0.2$ and $c(1|2) = c(2|1)$.

Question 5: Use the data in the EXCEL file tab "q5". Assume the data is a sample from two multivariate normal distributions.

a. Identify the classification rule for the case $p_1 = p_2$ and $c(1|2) = c(2|1)$.

b. Identify the classification rule for the case $p_1 = 0.25$ and $c(1|2)$ is half of $c(2|1)$.

c. Identify the Bayesian rule using $p_1 = 0.6$.

d. Based on the rules given in part (a), which population will a new data point $(50, 48, 47, 49)$ be classified into? How about using the rules in part (b)? Using the rule in part (c)?