**Ramakrishna Mission Vivekananda Educational and Research Institute**
PO Belur Math, Howrah, West Bengal 711 202
**School of Mathematical Sciences**
**Department of Computer Science**

MSc Big Data Analytics : Batch 2019-21, Semester II, Final
DA210: Multivariate Statistics
Dr. Sudipta Das

Student Name (in block letters): Date: 13 July 2020
Student Roll No: Max Marks: 100
Signature: Time: 3hrs

---

*Answer all questions*

1. (20 points)

   Suppose $\mathbf{X} = [X_1, X_2, X_3]' \sim N_3(\mu, \Sigma)$ where

   $$\mu = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{bmatrix}.$$

   Let $Z_1 = X_2 - X_3, Z_2 = X_2 + X_3$ and $(Z_3|Z_1, Z_2) \sim N_1(Z_1 + Z_2, 10)$. Compute the distribution of

   $$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}.$$

2. (20 points)

   There are two variables, verbal and performance scores for $n = 100$ elderly subjects aged $60-64$ on an Adult Intelligence test. Assume that the data are from a bivariate normal distribution. The maximum likelihood estimates of unknown mean vector $\mu$ and unknown covariance matrix $\Sigma$ are calculated as

   $$\begin{bmatrix} 55.24 \\ 34.97 \end{bmatrix} \text{ and } \begin{bmatrix} 208.46 & 125.73 \\ 125.73 & 178.50 \end{bmatrix},$$

   respectively. Test the hypothesis

   $$H_0 : \mu = \begin{bmatrix} 60 \\ 30 \end{bmatrix},$$

   at the 5% level of significance.

3. (20 points)

   Let $M$ be an exponential random variable with rate $\lambda$. Let $c = (c_1, c_2, c_3)^T$ be a vector of real-valued constants such that $c_i \neq 0$, for $i = 1, 2, 3$. Suppose that $X = (X_1, X_2, X_3)^T = cM$. Find the principal components of standardized $X$ and draw the scree plot.

4. (20 points)

Identify the classification region for a bivariate observation $\mathbf{x} = [x_1, x_2]^T$, where the first component, $(X_1)$ follows an exponential distribution with rates $\lambda_1$ and $\lambda_2$ for populations $\pi_1$ and $\pi_2$, respectively; and the second component, $(X_2)$ has a Bernoulli distribution with parameters $p_1$ and $p_2$ for populations $\pi_1$ and $\pi_2$, respectively. One can assume the following.

- $X_1$ and $X_2$ are mutually independent
- $\lambda_2 > \lambda_1$
- Prior probabilities are same
- Cost ratio is 1

5. (20 points)

The admission officer of a business school has used an "index" of undergraduate grade point average (GPA) and graduate management aptitude test (GMAT) scores to help decide which applicants should be admitted to the school's graduate programs. Depending on the GPA and GMAT scores applicants are categorized as $\pi_1$ : admit; $\pi_2$ : do not admit; and $\pi_3$ : borderline. The data on the scores provide the following information

$$n_1 = 31, n_2 = 28, n_3 = 26.$$

$$\bar{x}_1 = \begin{bmatrix} 4.42 \\ 561.23 \end{bmatrix}, \bar{x}_2 = \begin{bmatrix} 3.58 \\ 443.07 \end{bmatrix}; \bar{x}_3 = \begin{bmatrix} 3.99 \\ 441.23 \end{bmatrix} \text{ and } S_{pooled} = \begin{bmatrix} 0.05111 & -3.0188 \\ -3.0188 & 3665.9011 \end{bmatrix}.$$

In the Scatter plot of ($z_1$ = GPA score, $z_2$ =GMAT score) specify the region of admission.

==================================================================

You may need following values:

$$F_{2,98}(0.05) = 3.08920, F_{98,2}(0.05) = 19.48552, F_{2,99}(0.05) = 3.08824,$$
$$F_{99,2}(0.05) = 19.48563, F_{2,100}(0.05) = 3.087296, F_{100,2}(0.05) = 19.48573$$

==================================================================

This exam has total 5 questions, for a total of 100 points and 0 bonus points.