Tribhuvan University
Institute of Sciences and Technology
**SCHOOL OF MATHEMATICAL SCIENCES**
**First Assessment 2079**

*Subject:* **Statistical Methods for Data Science**
*Course No:* **MDS 553**
*Level:* **MDS /I Year/II Semester**

*Full Marks* 45
*Pass Marks* 22.5
*Time* 2 hrs

*Candidates are required to give their answer in their own words as far as practicable.*

**Attempt ALL Questions.**

**Group A** **[5 × 3 = 15]**

1. There card players play a series of matches, the probability that A will any game is 20%, the probability that player B will is 30%, and the probability that the player C will win is 50% If they play 6 games, what is the probability that player A will win 1 game, player B will win 2 games and player C will win 3 games?
2. Obtain the mean and variance of multinomial distribution.
3. Differentiate between parametric and non parametric test.
4. What is hypothesis? Differentiate between simple and composite hypothesis.
5. What do you understand by Most Powerful Test (MPT)

**Group B** **[5 × 6 = 30]**

6. What do you know about multinomial distribution? Obtain the moment generating function of multinomial distribution.
7. In a certain computer hardware manufacturing industry six different types of machines are working to cut pieces of wires. The number of wires of unequal length recorded in a day is as follows:

| **Machine** | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **No. of wire** | 2 | 0 | 4 | 8 | 5 | 11 |

Do these data provide sufficient evidence that the six machines equally cut the wires of unequal length? Apply Kolmogorov Smirnov test at 5% level of significance.

**OR**

The heart beating rate of 5 vegetarians and 5 non vegetarians are recorded below:

| **Vegetarians** | 56 | 67 | 82 | 60 | 75 |
|---|---|---|---|---|---|
| **Non vegetarians** | 53 | 42 | 75 | 58 | 65 |

Is the mean heart beating rate of non vegetarians significantly high? Use Mann Whitney U test.

8. Poverty Alleviation fund has provided grants on income generation program to different districts. A sample of 5 districts from Terai region, 9 from Hilly region and 6 from Himali region are selected and grants on income generation activities (in million Rs.) on different districts were recorded as follows:-

| **Himali** | 7 | 33 | 111 | 39 | 72 | 128 | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Hilly** | 176 | 266 | 213 | 135 | 95 | 54 | 86 | 75 | 45 |
| **Terai** | 138 | 290 | 66 | 98 | 208 | | | | |

Use Kruskal Wallis H test to test whether there is any significant difference in the mean grants in three geographical regions.

9. A survey was conducted in four hospitals in a Kathmandu to obtain the number of babies born over a 12 months period. This time period was divided into four seasons to test the hypothesis that the birth rate is constant over all the four seasons. The results of the survey were as follows:

| Hospital | No. of births | | | |
|---|---|---|---|---|
| | Winter | Spring | Summer | Fall |
| A | 92 | 72 | 94 | 77 |
| B | 15 | 16 | 10 | 17 |
| C | 58 | 71 | 51 | 62 |
| D | 19 | 26 | 20 | 18 |

Analyze the data using Friedman two way ANOVA test.

10. State and Prove Neymann- Pearson's Lemma (N-P Lemma).

**OR**

Find BCR in a normal distribution to test $H_0: \mu = \mu_0$ vs $H_1: \mu = \mu_1$. Given, probability of type I error = $\alpha$ and $\sigma = 1$.

***

Tribhuvan University
Institute of Sciences and Technology
**SCHOOL OF MATHEMATICAL SCIENCES**
**Second Assessment 2079**

*Subject*: **Statistical Methods for Data Science** *Full Marks*: **45**

*Course No*: **MDS 553** *Pass Marks*: **22.5**

*Level*: **MDS /I Year /II Semester** *Time*: **2 hrs**

*Candidates are required to give their answer in their own words as far as practicable.*

**Group A** **[5 ×3=15]**

1. If a probability is 0.40 that a child exposed to a certain disease will contain it, what is the probability that the tenth child exposed to the disease will be the third to catch it?
2. What do you understand by Extreme Value Distributions?
3. Discuss conjugate prior with its families.
4. What do you understand by unbiased test?
5. Show that exponential family distribution has MLR.

**Group B** **[5 × 6 = 30]**

6. Show that the negative Binomial Distribution is a special case of Generalized Power Series Distribution (GPSD).
7. Let x is distributed as a Poisson distribution with parameter $\lambda$ and $\lambda$ itself is distributed as a Gamma distribution with parameter $\alpha$ and $\beta$. Find the posterior distribution of $\lambda$ given x. Also, calculate its mean and variance.

**OR**

The time failure of transition is known to be exponential distribution with parameter $\lambda$. If the prior distribution of $\lambda$ is exponentially distributed with parameter $\theta$. Find the posterior distribution of $\lambda$ given $x_1, x_2, \ldots x_n$. Also, compute its mean and variance.

8. Discuss Compound Negative Exponential Distribution. Derive its moments.
9. Let X has a mixed distribution with DF, F(x) defined as follows:

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ \dfrac{x^2+1}{4} & if\ 0 \le x < 1 \\ \dfrac{x+2}{4} & if\ 1 \le x < 2 \\ 1 & if\ x \ge 2 \end{cases}$$

Obtain mean and variance of X and Sketch F(x) with respect to x.

10. Let a random variable has normal distribution with unknown '$\mu$' and known variance '2' i.e. X~N ($\mu$, 2). Derive the Likelihood Ratio Test (LRT) for testing the null hypothesis $H_0$: $\mu$=10 against $H_1$: $\mu \neq 10$ at a 5% level of significance.

**OR**

Let $x_1, x_2, \ldots, x_n$ be a sample from $N(0, \sigma^2)$. Test the UMP test exist of not for testing a hypothesis $H_0$: $\sigma = \sigma_0$ against

a) $H_0$: $\sigma > \sigma_0$
b) $H_0$: $\sigma < \sigma_0$
c) $H_0$: $\sigma \neq \sigma_0$

***

Tribhuvan University
**Institute of Science and Technology**
2079
✡

Master Level / I Year /IInd Semester/ Science
**Data Science (MDS 553)**
(Statistical Methods for Data Science)

Full Marks: 45
Pass Marks: 22.5
Time: 2 hours

*Candidates are required to give their answers in their own words as for as practicable.*
**Attempt All Questions**

## Group A

**[5 × 3 = 15]**

1. What do you mean by multinomial distribution? Obtain the expression for its mean.
2. Write the applications of Extreme Value Distribution.
3. Distinguish between prior and posterior distributions with an illustrative example.
4. Show that the Binomial distribution is a special case of Generalized Power Series Distribution.
5. Differentiate between Null and Alternative hypothesis.

## Group B

**[5 × 6 = 30]**

6. Let X follows binomial distribution with parameters n and θ, and θ follows Beta distribution with parameters α and β. Find the posterior distribution of θ given x. Also, find the mean and variance of posterior distribution.
7. Let X has a mixed distribution with DF; F(x) defined as follows. Obtain the mean and variance of X.

$$F(x)=\begin{cases} 0 & if \quad x<0 \\ \frac{x^2}{4} & if \quad 0\le x<1 \\ \frac{x+1}{4} & if \quad 1\le x<2 \\ 1 & if \quad x\ge 2. \end{cases}$$

8. An experiment designed to compare three preventive methods against corrosion yielded the following maximum depths of pits (in thousands of an inch) in pieces of wire subjected to the respective treatments:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Method I** | 77 | 54 | 67 | 74 | 71 | 66 | |
| **Method II** | 60 | 41 | 59 | 65 | 62 | 64 | 52 |
| **Method III** | 49 | 52 | 69 | 47 | 56 | | |

Use the 0.05 level of significance to test the hypothesis that the three samples come from identical population. Use Kruskal Wallis H test.

9. An agricultural experiment was conducted to compare the yields of paddy at 4 plots in Godawari by using the three different chemical fertilizers Nitrogen (N), Phosphorus (P), and Potash (K). The yields of paddy in (Qtl) were given in the following table:

| **Chemical Fertilizer** | **Plot** | | | |
|---|---|---|---|---|
| | **I** | **II** | **III** | **IV** |
| **N** | 122 | 83 | 138 | 121 |
| **P** | 81 | 89 | 79 | 65 |
| **K** | 80 | 82 | 65 | 58 |

Do the data provide sufficient evidence to support the null hypothesis that the population of yields of paddy corresponding to the three types of fertilizers do not differ in location? Use Friedman two-way ANOVA test.

**OR**

Two groups of data managers, one group consisting of trained ones, another groups are not trained have the following number of correction required.

| | | | | | |
|---|---|---|---|---|---|
| **Trained** | 78 | 64 | 75 | 45 | 82 |
| **Untrained** | 110 | 70 | 53 | 51 | |

Use Mann Whitney U test to test if there is a significantly difference between the two average number of correction of trained and untrained data manager.

10. State and Prove Neymann- Pearson's Lemma (N-P Lemma).

**OR**

Prove that for testing of hypothesis $H_0$: $\theta = \theta_0$ vs $H_1$: $\theta = \theta_1$, its power is never less than its size i. e. $\alpha \leq 1 - \beta$.