

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

**Attempt ALL Questions.**

**Group A [5 × 3 = 15]**

- 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?
- Obtain the mean and variance of multinomial distribution.
- Differentiate between parametric and non parametric test:
- What is hypothesis? Differentiate between simple and composite hypothesis.
- What do you understand by Most Powerful Test (MPT)

**Group B [5 × 6 = 30]**

- What do you know about multinomial distribution? Obtain the moment generating function of multinomial distribution.
- 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.

- 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]**

- 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?
- What do you understand by Extreme Value Distributions?
- Discuss conjugate prior with its families.
- What do you understand by unbiased test?
- Show that exponential family distribution has MLR.

**Group B [5 × 6 = 30]**

- Show that the negative Binomial Distribution is a special case of Generalized Power Series Distribution (GPSD).
- 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  $x_1, x_2, \dots, x_n$ . Also, compute its mean and variance.

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

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2 + 1}{4} & \text{if } 0 \leq x < 1 \\ \frac{x + 2}{4} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

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

- Let a random variable has normal distribution with unknown ' $\mu$ ' and known variance '2' i.e.  $X \sim N(\mu, 2)$ . 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  $x_1, x_2, \dots, x_n$  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

- $H_0: \sigma > \sigma_0$
- $H_0: \sigma < \sigma_0$
- $H_0: \sigma \neq \sigma_0$

\*\*\*

Tribhuvan University  
 Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 First Assessment 2080

**Subject: Statistical Method for Data Science**

**Full Marks: 45**

**Course No: MDS 553**

**Pass Mark: 22.5**

**Level: MDS / I year/ II Semester**

**Time: 2hrs.**

**Attempt all the questions**

**Group A [5×3=15]**

- Define generalized power series distribution and show that the negative binomial distribution is a special case of GPSD.
- The number of virus infected computers of five different capacity of hard disk is given below:

Capacity of hard disk(GB)	500	320	1000	2000	400
No. of virus infected	11	15	20	3	1

Test whether the computers of five hard disk are uniformly infected using Kolmogorov Smirnov test.

- An important part of the customer service responsibilities of a telephone company relates to the speed with which troubles in residential service can be repaired. Suppose past data indicate that the likelihood is 0.70 that troubles in residential service can be repaired on the same day. For the first five troubles reported on a given day, what is the probability that
  - All five will be repaired on the same day?
  - At least three will be repaired on the same day?
  - Fewer than two will be repaired on the same day?
- What do you mean by multinomial distribution and find the moment generating function of the distribution.
- The average number of claims per hour made to the Gnecco and Trust Insurance Company for damages or losses incurred in moving is 3.1. What is the probability that in any given hour
  - Fewer than three claims will be made?
  - Exactly three claims will be made?
  - Three or more claims will be made?

**Group B [5×6=30]**

- Define negative binomial distribution and find the mean and variance of the distribution.

**OR**

Define Binomial distribution and find the mean and variance of the distribution.

- Define errors in hypothesis testing. Ten accountants were given intensive coaching and four tests were conducted in a month. The scores of tests 1 and 4 are given below.

Accountants	1	2	3	4	5	6	7	8	9	10
Marks in 1 <sup>st</sup> test	50	42	51	42	60	41	70	55	62	38
Marks in 4 <sup>th</sup> test	62	40	61	52	68	51	64	63	72	50

Does the score from 1<sup>st</sup> to 4<sup>th</sup> test shows an improvement using Wilcoxon Matched pairs signed rank test? Test at the 5% level of significance.

8. An IQ test was given to a random sample of 15 male and 20 female students of a university. Their scores were recorded as follows:

Male: 56, 66, 62, 81, 75, 73, 83, 68, 48, 70, 60, 77, 86, 44, 72

Female: 63, 77, 65, 71, 74, 60, 76, 61, 67, 72, 64, 65, 55, 89, 45, 53, 68, 73, 50, 81

Use median test at 0.01 level of significance to determine whether IQ of male and female students are same in the university.

9. State and prove Neymann-Pearson's Lemma.

OR

For testing  $H_0: \theta=1.5$  against  $H_1: \theta=2.5$  on the basis of single observation drawn from uniform distribution with probability density function,

$$f(x) = 1/0; 0 < X < 0$$

Calculate the probabilities of type first error, type second error and power of the test, if critical regions are (i)  $W=\{x: 0.8 \leq x\}$  and (ii)  $W=\{x: 1.5 \leq x \leq 2\}$ .

10. The following data represents the operating times in hours for three types of scientific pocket calculators before a recharge is required:

Calculator A	4.9	6.1	4.3	4.6	5.3
Calculator B	5.5	5.4	6.2	5.8	5.5
Calculator C	6.4	6.8	5.6	6.5	6.3

Use Kruskal-Wallis H test, at the 0.05 level of significance, to test the hypothesis that the operating times for all three calculators are equal.

\*\*\*

Tribhuvan University  
 Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 first Reassessment 2080

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

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

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

**Attempt all the questions**

**Group A [5×3=15]**

- Define generalized power series distribution and show that the binomial distribution is a special case of GPSD.
- Differentiate between parametric and non-parametric test.
- A multiple choice test has 5 questions. There are 4 choices for each question. A student, who has not studied for the test, decides all the answer of all questions randomly. What is the probability that he will get
  - At most two questions correct?
  - At least one questions correct?
- Obtain the correlation coefficient of multinomial distribution.
- Nepal Rastra bank is responsible for printing the country's paper money. It has an impressively small printing error only 0.5 percent of all bills are too flawed for circulation. What is the probability that out of a batch of 1000 bills,
  - None are too flawed for circulation.
  - Ten are too flawed for circulation.
  - Fifteen are too flawed for circulation.

**Group B [5×6=30]**

- Define binomial distribution and find the mean and variance of the distribution.

**OR**

- Define Poisson distribution and find the mean and variance of the distribution.
- Define null and alternative hypothesis.

The same C programming papers were marked by two teachers A and B. The final score were recorded as follows:

Teacher A	73	89	82	43	80	73	66	45	50	55
Teacher B	88	78	91	48	85	74	77	31	40	28

- Using median test at 5% level of significance to determine if the marks distribution of two teachers differ significantly.
- A random sample of size 25 is drawn from a normal population with mean  $\mu$  and standard deviation 3. In testing  $H_0: \mu=20$  against  $H_1: \mu>20$  it is decided that  $H_0$  will be rejected if the sample mean is greater than 21.4. Calculate (i) probability of type I error (ii) probability of type II error when (a)  $\mu=21$  and (b) 22.
  - 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.
- 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.

\*\*\*

Tribhuvan University  
 Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 Second Assessment 2080

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

**Full Marks: 45**  
**Pass Mark: 22.5**  
**Time: 2hrs.**

*Candidates are required to give their answers in their own word as far as practicable*

**Attempt ALL questions.**

**Group A [5 x 3 = 15]**

1. A random sample of size 14 from a Poisson distribution with mean ' $\mu$ ' yields the following data values:

10	5	7	7	5	5	9
8	3	7	8	6	8	4

If the prior distribution of ' $\mu$ ' is Exponential (0.5), find the posterior distribution of  $\mu$ .

2. A manufacturer took a sample of 9. The life of the batteries in hours until failure was  
 30      40      30      50      42      30      48      46      44  
 At the 0.05 level of significance, is there evidence that the mean life of batteries is more than 40 hours?

3. Define compound distribution.

If  $X$  has an exponential distribution with parameter  $\alpha=12$  and  $N$  has a Poisson distribution with  $\lambda=55$ .

Calculate the expected value and variance of  $S$ .

4. 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.  $\alpha \leq 1-\beta$ .

5. Shoes are produced by two machines A and B. 50% of the shoes were produced by machine A with an estimate of 10% of them being defective. On machine B, 20% of the shoes produced are defective. If a shoe taken at random is found to be defective, what is the probability that the shoe was produced by machine (i) A.? (ii) B.?

**Group B [5x6=30]**

6. A continuous random variable  $X$  has pdf.  $f(x)$  where,

$$\begin{aligned} f(x) &= K x; 0 \leq x < 1 \\ &= K(2-x); 1 \leq x < 2 \\ &= 0; \text{ otherwise.} \end{aligned}$$

- a) Find the value of  $K$   
 b) Mean and variance of  $X$   
 c)  $P(0.75 \leq X \leq 1.5)$

**OR**

Let  $X$  has a mixed distribution with distribution function;  $F(x)$  defined as follows:

$$\begin{aligned} F(x) &= 0 \text{ if } x \leq 0 \\ &= (2x/3) \text{ if } 0 \leq x < 1 \\ &= (x+1)/3 \text{ if } 1 \leq x < 2 \\ &= 1 \text{ if } x \geq 2 \end{aligned}$$

Find the mean and variance of variable  $X$

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.

8. Define extreme value distribution in brief and also write the applications of the distribution.
9. Let a random variable has normal distribution with unknown ' $\mu$ ' and known variance '3' i.e.  $X \sim N(\mu, 3)$ . Derive the Likelihood Ratio Test (LRT) for testing the null hypothesis  $H_0: \mu=11$  against  $H_1: \mu \neq 11$  at a 5% level of significance.

$$n = 20 \\ s = 12 \\ \bar{x} \\ \alpha = 0.01$$

10. The scores of three matched groups under the six conditions are given below

Group	Conditions					
	I	II	III	IV	V	VI
A	9	5	2	5	6	7
B	6	4	3	4	6	5
C	5	1	3	3	6	5

Analyze the data using Friedman's test to identify if there is significant difference in variation between matched groups. Use 5% level of significance.

OR

A researcher company has designed three different systems to clean up oil spills. The following table contains the results, measured by how much surface area (in square meters) is cleared in 1 hour. The data were found by testing each method in several trials. Are the three systems equally effective? Use Kruskal Wallis H test at  $\alpha=5\%$ .

System A	55	60	63	56	59	55
System B	57	53	64	49	62	
System B	66	52	61	57		

\*\*\*

Tribhuvan University  
Institute of Science and Technology  
2079  
★

Master Level / I Year / II Ind 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 far as practicable.*

**Attempt All Questions**

**Group A**

[ $5 \times 3 = 15$ ]

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**

[ $5 \times 6 = 30$ ]

6. Let  $X$  follows binomial distribution with parameters  $n$  and  $\theta$ , and  $\theta$  follows Beta distribution with parameters  $\alpha$  and  $\beta$ . Find the posterior distribution of  $\theta$  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  $X$ .

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{4} & \text{if } 0 \leq x < 1 \\ \frac{x+1}{4} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2. \end{cases}$$

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.  $\alpha \leq 1 - \beta$ .

**Tribhuvan University**  
**Institute of Science and Technology**  
**SCHOOL OF MATHEMATICAL SCIENCE**  
First Assessment 2081

**Subject: Statistical Method for Data Science**

**Course No: MDS 553**

**Level: MDS / I year/ II Semester**

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

**Attempt all the questions**

**Full Marks: 45**

**Pass Mark: 22.5**

**Time: 2hrs.**

**Group A [5x3=15]**

1. Define generalized power series distribution and show that the binomial distribution is a special case of GPSD.
2. The chances of X, Y, Z becoming manager will be of a certain company are 4:2:3. The probabilities that bonus scheme will be introduced if X, Y, Z becomes manager are 0.3, 0.5 and 0.8 respectively. If the bonus scheme has been introduced, what is the probability that Z is appointed as the manager?
3. The Orange county Dispute settlement center handles various kinds of disputes but most are marital disputes. In fact 96 percent of the disputes handled by the DSC are of a marital nature.
  - a) What is the probability that, out of 80 disputes handled by the DSC, exactly five are non-marital?
  - b) What is the probability that, out of 70 disputes handled by the DSC, at least two are non-marital?
4. What do you mean by multinomial distribution and find the moment generating function of the distribution.
5. A random sample of size 12 from a Poisson distribution with mean ' $\lambda$ ' yields the following data values:

11	15	9	0	8	3	13
9	5	8	9	10		

If the prior distribution of ' $\lambda$ ' is Exponential (0.8) then find the posterior distribution of  $\lambda$ .

**Group B [5x6=30]**

6. Define negative binomial distribution and find the mean and variance of the distribution.

**OR**

- Define Poisson distribution and find the mean and variance of the distribution.

7. Define null and alternative hypothesis.
- MBA students were selected by a big organization and they were given an IQ test immediately after the selection and their scores out of 50 were noted. A two months training was arranged for them and on completion IQ test was given and scores recorded.

Students	1	2	3	4	5	6	7	8	9
Scores (Before)	31	35	38	25	40	45	40	28	18
Scores (After)	32	30	32	37	38	37	42	40	27

On the basis of the results can it be concluded that the training has benefited the students? Use the 0.01 level of significance.

8. An IQ test was given to a random sample of 15 male and 20 female students of a university. Their scores were recorded as follows;

Male: 56, 66, 62, 81, 75, 73, 83, 68, 48, 70, 60, 77, 86, 44, 72

Female: 63, 77, 65, 71, 74, 60, 76, 61, 67, 72, 64, 65, 55, 89, 45, 53, 68, 73, 50, 81

Use median test at 0.1 level of significance to determine whether IQ of male and female students are same in the university.

9. Following table gives the monthly sales(in thousand rupees) of a certain firm in four regions by its four salesmen:

Region	Salesmen			
	A	B	C	D
R <sub>1</sub>	5	4	4	7
R <sub>2</sub>	7	8	5	4
R <sub>3</sub>	9	6	6	7
R <sub>4</sub>	8	8	5	7

Analyze the data using Friedman's test to identify if there is significant difference between four regions. Use 5% level of significance.

**OR**

Four training methods were compared to see whether they lead to the same level of productivity after training. A total of 24 trainees are used in this study where the training methods are A, B, C and D; and six individuals chosen at random, are assigned to each method. The productivity scores corresponding to each method are shown below.

Method A	44	53	39	50	40	45
Method B	59	43	47	51	39	49
Method C	41	37	43	40	52	37
Method D	45	30	49	39	55	39

From the above data, can we conclude that all the four methods are identical in terms of productivity score? Use Kruskal Wallish H test at 5% level of significance.

10. The number of claims in a week arising from a certain group of insurance policies has a Poisson distribution with mean  $\mu$ . Seven claims were incurred in the last week. The prior distribution of  $\mu$  is uniform on integers 8, 10, 12.

- a) Determine the posterior distribution of  $\mu$ . b) Find the mean and variance of the posterior distribution.

\*\*\*

Tribhuvan University  
Institute of Science and Technology  
2081



MDS / I Year / Second 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 far as practicable.*

The figures in the margin indicate full marks. The symbols have their usual meanings.

**Attempt all questions.**

**Group A**

**[ $5 \times 3 = 15$ ]**

1. The director of admissions at a large university advises parents of incoming students about the cost of text books during a typical semester. A sample of 100 students enrolled in the university indicates a sample mean cost of \$ 315.40 with a sample standard deviation of \$ 43.20. Using the 0.10 level of significance, is there evidence that the population mean is above \$ 300.
2. Write the applications of the extreme value distributions.
3. Define compound negative exponential distribution.  
Suppose number of claims has a Binomial (98, 0.02) distribution and claim sizes have a Gamma (10, 0.2) distribution then find the mean and variance of aggregate claim amount 'S'.
4. The probability that player A will win any game is 35%, the probability that player B will win is 25% and the probability that C will win is 40%. If they play 7 games, what is the probability that A will win 2 games, player B will win 2 games and player C will win 3 games?
5. Two independent random samples of unemployed men and women are drawn and the ages of the 4 unemployed women and 5 unemployed men are recorded as follows:

Women	60	63	36	44	
Men	53	39	22	33	24

Test whether the average age of unemployed women is significantly high. Use Mann-Whitney U test at 5% level of significance.

**Group B**

**[ $5 \times 6 = 30$ ]**

6. Show that the negative binomial distribution is a special case of GPSD and also find the mean and variance of the distribution.
7. The number of claims in a week arising from a certain group of insurance policies has a Poisson distribution with mean  $\lambda$ . Nine claims were incurred in the last week. The prior distribution of  $\lambda$  is uniform on integers 4, 7, 9.
  - a) Determine the posterior distribution of  $\lambda$ .
  - b) Find the mean and variance of the posterior distribution.

8. Suppose it is desired to test the null hypothesis  $H_0: \mu=31$  against the alternative hypothesis  $H_1: \mu \neq 31$  on the basis of random sample of size 17 from a normal population  $N(\mu, 2)$ . The decision rule is; reject null hypothesis if the sample mean is less than 31.65 or more than 31.31.
- Find the probability of type first error.
  - Find the probability of type second error (i) when  $\mu=32$  and (ii)  $\mu=32.3$

**OR**

Let  $x_1, x_2, \dots, x_n$  be a random sample of size  $n$  drawn from a normal population with mean  $\mu$  and variance  $\sigma^2$  where variance is known. For testing  $H_0: \mu = \mu_0$  against  $H_1: \mu = \mu_1$ , find the best critical region if  $\mu_1 < \mu_0$  and also find the power of the test.

9. The time taken to perform a particular task,  $X$  hours, has the probability density function

$$\begin{aligned} f(x) &= 10kx^2; 0 \leq x < 0.6 \\ &= 9k(1-x); 0.6 \leq x < 1 \\ &= 0; \text{ otherwise.} \end{aligned}$$

- Find the value of constant  $k$ .
- Mean and variance of  $X$ .
- Determine the probability that the time will be between 0.3 and 0.9.

10. For the following scores of 3 matched groups, apply Kruskal-Wallis test to test the hypothesis that the three groups are not significantly different

Group	Scores				
	A	B	C	D	E
A	96	128	83	61	101
B	82	124	132	135	109
C	115	149	166	147	-

**OR**

A survey was conducted in four hospitals in a Lalitpur 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
I	91	73	93	77
II	16	15	10	17
III	58	71	51	62
IV	19	26	20	19

Analyze the data using Friedman two way ANOVA test.

**Tribhuvan University**  
**Institute of Science and Technology**  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 First Assessment 2082

**Subject:** Statistical Method for Data Science

**Full Marks:** 45

**Course No:** MDS 553

**Pass Mark:** 22.5

**Level:** MDS / I year/ II Semester

**Time:** 2hrs.

**Attempt all the questions**

**Group A** [5×3=15]

- Define generalized power series distribution and show that the Poisson distribution is a special case of GPSD.

$\Phi(\lambda)$

- A random sample of size 14 from a Poisson distribution with mean ' $\lambda$ ' yields the following data values:

8	12	5	9	4	8	10
7	5	9	10	7	12	18

If the prior distribution of ' $\lambda$ ' is Exponential (0.6) then find the posterior distribution of  $\lambda$ .

- The latest nationwide political poll indicate that for Americans who are randomly selected, the probability that they are conservative is 0.55, the prob. that they are liberal is 0.30 and the probability that they are middle of the road is 0.15. Assuming that these probabilities are accurate, answer the following questions. Pertaining to a randomly chosen group of 10 Americans.

- What is the probability that at most two are liberal?
- What is probability that none are conservative?
- What is the probability that at least eight are middle of the road?

posterior ( $\lambda$ )

- Obtain the moment generating function of multinomial distribution.

- Airline passengers arrive randomly and independently at the passenger screening facility at a major international airport. The average arrival rate is 6 passengers per minute.
  - What is probability that more than four passengers will arrive in a minute?
  - What is the probability that three or four passengers arrive in a minute?
  - What is the probability that at least one passengers arrive in two minutes?

**Group B** [5×6=30]

- Define negative binomial distribution and find the mean and variance of the distribution.

**OR**

Define Poisson distribution and find the mean and variance of the distribution.

- The number of claims in a week arising from a certain group of insurance policies has a Poisson distribution with mean  $\alpha$ . Five claims were incurred in the last week. The prior distribution of  $\alpha$  is uniform on integers 4, 6, 7.
  - Determine the posterior distribution of  $\alpha$ .
  - Find the mean and variance of the posterior distribution.

8. An IQ test was given to a random sample of 15 male and 20 female students of a university. Their scores were recorded as follows,  
 Male: 56, 66, 62, 81, 75, 73, 83, 68, 48, 70, 60, 77, 86, 44, 72  
 Female: 63, 77, 65, 71, 74, 60, 76, 61, 67, 72, 64, 63, 65, 89, 49, 63, 68, 77, 50, 81  
 Use median test at 0.01 level of significance to determine whether IQ of male and female students are same in the university.

9. A random sample of 20 volume based internet connected have following speed of internet connection in mps:  
 2.7, 2.9, 3.0, 3.1, 2.8, 3.0, 2.9, 3.0, 2.6, 3.1, 3.0, 2.9, 3.3, 3.0, 2.8, 2.9, 3.0, 2.9  
 Apply the Kolmogorov Smirnov test for testing that the speed are equally distributed.

OR

Test the hypothesis of no difference between the ages of male and female employees of certain IT company, using the Mann-Whitney U test for the sample data below. Use  $\alpha=0.1$

Male	35	12	43	18	26	44	40	15	42	12	33	10	38	14	25	21	26	15	
Female	30	7	41	16	34	11	31	8	36	13	32	9	25	21	47	10	28	6	24

10. The following data represents the operating times in hours for three types of scientific pocket calculators before a recharge is required

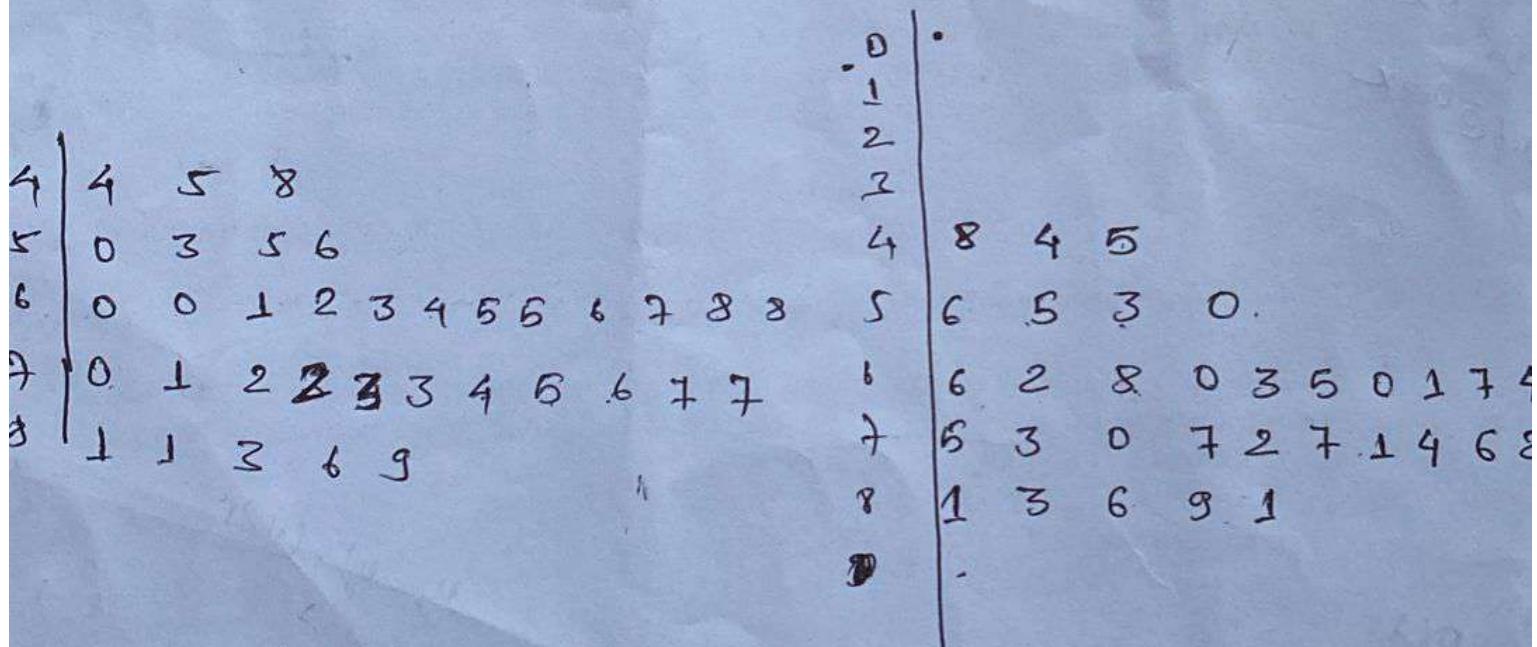
Calculator A 4.94 6.11 2.43 4.34 4.62 5.36

Calculator B 5.53 5.47 6.21 3.58 11 5.58 5 3.25 4.83

Calculator C 6.41 5.68 18 5.61 9 6.51 6 6.31 4 6.61 7

Use Kruskal-Wallis H test, at the 0.05 level of significance, to test the hypothesis that the operating times for all three calculators are equal.

\*\*\*



Tribhuvan University  
 Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 Second Assessment 2082

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

**Full Marks:** 45  
**Pass Mark:** 22.5  
**Time:** 2hrs.

Attempt all the questions

**Group A** [5×3=15]

- ✓ 1. Write the applications of the extreme value distributions
2. Ministry of Tourism and Civil Aviation has claimed that the average length of stay of tourist in Nepal is less than 13 days. To test this claim a researcher asked 9 tourists about their length of stay in Nepal and their length of stay in days were 10, 15, 11, 5, 7, 4, 8, 14 and 11. On the basis of this sample result can we conclude that the average length of stay is less than 13 days?
3. Define compound distribution
- ✓ If  $X$  has an exponential distribution with parameter  $\theta=10$  and  $N$  has a Poisson distribution with  $\lambda=50$ . Calculate the expected value and variance of  $S$ .
- ✓ 4. 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.  $\alpha \leq 1-\beta$ .
- ✓ 5. Hinton Press hypothesizes that the average 14500 hours. From a sample of 25 presses, the company finds a sample mean of 13000 hours with standard deviation of 2100 hours. At 1% significance level, should the company conclude that the average life of the presses is less than the hypothesized 14500 hours?

**Group B** [5×6=30]

$$N(2) \quad N(3.16^2) \quad \sigma = \sqrt{1.1}$$

6. Define error in hypothesis testing

Ten accountants were given intensive coaching and four tests were conducted in a month. The scores of tests 1 and 4 are given below:

Accountants	1	2	3	4	5	6	7	8	9	10
Marks in 1 <sup>st</sup> test	50	42	51	42	60	41	70	55	62	38
Marks in 4 <sup>th</sup> test	62	40	61	52	68	51	64	63	72	50

$$\frac{x-4}{\sqrt{5}}$$

Does the score from 1<sup>st</sup> to 4<sup>th</sup> test shows an improvement using Wilcoxon Matched pairs signed rank test? Test at the 5% level of significance

7. Suppose it is desired to test the null hypothesis  $H_0: \mu=35$  against the alternative hypothesis  $H_1: \mu \neq 35$  on the basis of random sample of size 16 from a normal population  $N(\mu, 1.1)$ . The decision rule is; reject null hypothesis if the sample mean is less than 34.65 or more than 35.35.

- Find the probability of type first error
- Find the probability of type second error (i) when  $\mu=36$  and (ii)  $\mu=36.1$ .

12    2    10    10    8    10    6    8    10    12  
 9.8    1    6.8    6.8    7.5    6.5    2    3.5    6.5    9.5  
 ↑                      ↑

$$x^\alpha < 1 - \beta$$

$$t(1-\alpha) > \beta$$

$$1 - \beta - \alpha \geq 1 - \beta$$

$$1 - \beta \geq \alpha$$

$$t(1-\alpha)(1-\beta) \geq x^\alpha \beta$$

$$1 - \beta - \alpha + \alpha \beta > x^\alpha \beta$$

$$2 \leq n \leq 9, k=3$$

$$2 \leq n \leq 5, k=4$$

8. A continuous random variable  $X$  has pdf.  $f(x)$  where,

$$\begin{aligned} f(x) &= Kx; 0 \leq x < 1 \\ &= K(2-x); 1 \leq x < 2 \\ &= 0, \text{ otherwise.} \end{aligned}$$

- a) Find the value of  $K$
- b) Mean and variance of  $X$
- c)  $P(0.75 \leq X \leq 1.5)$

9. Let a random sample of size 40 is selected from a  $N(\mu, 16)$ . In testing  $H_0: \mu=50$  against  $H_1: \mu=52$  at  $\alpha=5\%$ .

- a) Find best critical region of size  $\alpha$ .
- b) Find probability of type Second error.
- c) Find power of the test.

OR

If  $x>2$  is a critical region for testing  $H_0: \lambda=1.5$  against  $H_1: \lambda=2.5$ , based on a sample of size 1 drawn from a Poisson population  $P(\lambda)$

- a) Find the probability of type first and type second errors and power of the test.
- b) Find the probability of type second error and
- c) Find the power of the test.

10. The scores of three matched groups under the six conditions are given below

Group	Conditions					
	I	II	III	IV	V	VI
A	9	5	2	5	6	7
B	6	4	3	4	6	5
C	5	1	3	3	6	5

Analyze the data using Friedman's test to identify if there is significant difference in variation between matched groups. Use 5% level of significance.

OR

A researcher company has designed three different systems to clean up oil spills. The following table contains the results, measured by how much surface area (in square meters) is cleared in 1 hour. The data were found by testing each method in several trials. Are the three systems equally effective? Use Kruskal Wallis H test at  $\alpha=5\%$ .

System A	55	60	63	56	59	55
System B	57	53	64	49	62	
System C	66	52	61	57		

\*\*\*

1/2 1-0.75

0.8079

Tribhuvan University  
 Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 Second Re-Assessment 2082

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

**Full Marks:** 45  
**Pass Mark:** 22.5  
**Time:** 2hrs.

*Candidates are required to give their answer in their own words as far as practicable. The figures in the margin indicate full marks.*

**Attempt all the questions**

**Group A [5×3=15]**

1. A television documentary on over eating claimed that American are about 10 pounds overweight on average. To test this claim, 18 randomly selected individuals were examined, their average excess weights were found to be 12.4 pounds and the sample s.d. was 2.7 pounds. At a 10% level of significance, is there any reason to doubt the validity of the claimed 10 pounds value?

2. Write the applications of the extreme value distributions.

3. Weights, in kg, of 10 students are given below

38      40      45      53      47      43      55      48      52      49

Can we say that variance of distribution of weights of all students from which the above sample of 10 students was drawn is equal to 20 square kg?

4. Define compound distribution.

If X has an exponential distribution with parameter  $\theta=7$  and N has a Poisson distribution with  $\lambda=55$ .

Calculate the expected value and variance of S.

5. 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.  $\alpha \leq 1-\beta$ .

**Group B [5×6=30]**

6. Define null and alternative hypothesis.

MBA students were selected by a big organization and they were given an IQ test immediately after the selection and their scores out of 50 were noted. A two months training was arranged for them and on completion IQ test was given and scores recorded.

Students	1	2	3	4	5	6	7	8	9
Scores (Before)	31	35	38	25	40	45	40	28	18
Scores (After)	32	30	32	37	38	37	42	40	27

On the basis of the results can it be concluded that the training has benefited the students using Wilcoxon Matched pairs signed rank test at 0.01 level of significance?

7. Let  $X$  has a mixed distribution with DF,  $F(x)$  defined as follows:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2 + 1}{4} & \text{if } 0 \leq x < 1 \\ \frac{x + 2}{4} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

Obtain mean and variance of  $X$

OR

A continuous random variable  $X$  has pdf.  $f(x)$  where,

$$\begin{aligned} f(x) &= K x; \quad 0 \leq x < 1 \\ &= K (2-x); \quad 1 \leq x < 2 \\ &= 0; \text{ otherwise.} \end{aligned}$$

- a) Find the value of  $K$
- b) Mean and variance of  $X$
- c)  $P(0.25 \leq X \leq 1.72)$

8. Let  $x_1, x_2, \dots, x_n$  be a random sample of size  $n$  drawn from a normal population with mean  $\mu$  and variance  $\sigma^2$  where variance is known. For testing  $H_0: \mu = \mu_0$  against  $H_1: \mu = \mu_1$ , find the best critical region if  $\mu_1 > \mu_0$  and also find the power of the test.

9. Let a random sample of size 38 is selected from a  $N(\mu, 15)$ . In testing  $H_0: \mu = 49$  against  $H_1: \mu = 51$  at  $\alpha = 10\%$ .
- a) Find best critical region of size  $\alpha$ .
  - b) Find probability of type Second error.
  - c) Find power of the test.

10. Following table gives the monthly sales(in thousand rupces) of a certain firm in three regions by its four salesmen:

Region	Salesmen			
	A	B	C	D
X	5	4	4	7
Y	7	8	5	4
Z	9	6	6	7

Analyze the data using Friedman's test to identify if there is significant difference between three regions. Use 5% level of significance.

OR

Four training methods were compared to see whether they lead to the same level of productivity after training. A total of 24 trainees are used in this study where the training methods are A, B, C and D; and six individuals chosen at random, are assigned to each method. The productivity scores corresponding to each method are shown below.

Method A	44	53	39	50	40	45
Method B	59	43	47	51	39	49
Method C	41	37	43	40	52	37
Method D	45	30	49	39	55	39

From the above data, can we conclude that all the four methods are identical in terms of productivity score? Use Kruskal Wallish H test at 10% level of significance.

\*\*\*

Tribhuvan University  
 Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCE**  
 First Re-assessment 2082

**Subject:** Statistical Method for Data Science  
**Course No:** MDS553  
**Level:** MDS / I year/ II Semester

**Full Marks:** 45  
**Pass Mark:** 22.5  
**Time:** 2hrs.

Attempt all the questions

**Group A** [5×3=15]

1. Define generalized power series distribution and show that the Binomial distribution is a special case of GPSD.  
 2. A random sample of size 12 from a Poisson distribution with mean ' $\mu$ ' yields the following data values:

9	11	15	8	5	7	11	
6	7	5	11	9			

- If the prior distribution of ' $\mu$ ' is Exponential (0.7) then find the posterior distribution of  $\mu$ .  
 3. Nepal Rastra bank is responsible for printing the country's paper money. It has an impressively small printing error only 0.5 percent of all bills are too flawed for circulation. What is the probability that out of a batch of 1000 bills,  
 a) None are too flawed for circulation. b) Ten are too flawed for circulation. C) Fifteen are too flawed for circulation.  
 4. Obtain the correlation coefficient of multinomial distribution.  
 5. Differentiate between parametric and non-parametric test.

**Group B** [5×6=30]

6. Define binomial distribution and find the mean and variance of the distribution.

**OR**

- Define negative binomial distribution and find the mean and variance of the distribution.  
 7. Define errors in hypothesis testing.  
 8. Ten accountants were given intensive coaching and four tests were conducted in a month. The scores of tests 1 and 4 are given below.

Accountants	1	2	3	4	5	6	7	8	9	10
Marks in 1 <sup>st</sup> test	50	42	51	42	60	41	70	55	62	38
Marks in 4 <sup>th</sup> test	62	40	61	52	68	51	64	63	72	50

8. A random sample of size 25 is drawn from a normal population with mean  $\mu$  and standard deviation 3. In testing  $H_0: \mu=20$  against  $H_1: \mu>20$  it is decided that  $H_0$  will be rejected if the sample mean is greater than 21.4. Calculate (i) probability of type I error (ii) probability of type II error when (a)  $\mu=21$  and (b) 22.

**OR**

For testing  $H_0: \theta=1.5$  against  $H_1: \theta=2.5$  on the basis of single observation drawn from uniform distribution with probability density function,  $f(x) = 1/\theta$  ;  $0 < X < \theta$

Calculate the probabilities of type first error, type second error and power of the test, if critical regions are (i)  $w=\{x: 0.8 \leq x\}$  and (ii)  $w=\{x: 1.5 \leq x \leq 2\}$ .

9. An IQ test was given to a random sample of 15 male and 20 female students of a university. Their scores were recorded as follows:  
 Male: 56, 66, 62, 81, 75, 73, 83, 68, 48, 70, 60, 77, 86, 44, 72  
 Female: 63, 77, 65, 71, 74, 60, 76, 61, 67, 72, 64, 65, 55, 89, 45, 53, 68, 73, 50, 81  
 Use median test at 0.01 level of significance to determine whether IQ of male and female students are same in the university.

10. 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.