Manipal Academy of Higher Education

First Semester ME (BIG DATA AND DATA ANALYTICS)

BDA –605: Probability and Statistical Inferences

Scheme of Evaluation

1.

A. Describe the Von Mises's Statistical definition of probability with its advantages and limitations

- **Definition: 2 Marks**
- **Any two advantages or limitations: 3 marks**

B. State the Bayes theorem with a detailed explanation on terms involved in the theorem. Provide an example of the theorem

- **Bayes theorem: 3 Marks**
- **Explanation of terms and example: 2 Marks**

(5+5=10 Marks)

2.

A. Define the following concepts with the help of an example

(i) Independent events

(ii) Conditional probability

(iii) Mutually exclusive events

B. Following table shows the relationship between gender and eye color for a group of 167 German men.

		Eye Colour					
		Black	Brown	Blue	Green	Gray	Total
Gender	Female	20	30	10	15	10	85
	Male	25	15	12	20	10	82
	Total	45	45	22	35	20	167

(i) What is the probability that a randomly selected person will have green eyes?

(ii) What is the probability that a randomly selected female will have brown eyes?

(iii) What is the probability that a randomly selected person will have brown eyes OR will be male?

((3x2)+1+1+2=10 Marks)

- **Definition of each term: 1.5 Marks**
- **Examples: 0.5 Marks**
- **Correct answers to the first two questions: 1 Mar each**
- **Correct answer to the last question: 2 Mark**

3.

A

(i) Define univariate random variable. What are the two major classifications of a univariate random variable? Give two examples for the each classification.

(2 marks)

- **Definition of univariate random variable : 1 Mark**
- **Classification+ Examples for each : 1 Mark**

(ii) For an univariate random variable X, define the following terms.

a. probability mass function
b. probability density function
c. distribution function **(3 marks)**

- **Definition for each term: 1 marks each**

B

a. Let X be a random variable which denotes "the number of heads observed" defined for the random experiment of throwing a coin three times and observing the sequence of heads and tail. Obtain the functional form of the probability mass function (p.m.f) and Cumulative Distribution Function (CDF) for the random variable X .

(3 marks)

- **Obtaining for functional form of the p.m.f : 1.5 marks**
- **Obtaining for the cumulative distribution function : 1.5 marks**

b. Let the distribution function of a random variable X is defined as follows,

$$F(x)=\begin{cases}0 & if \quad x<0\\ x & if \quad 0\le x\le 1\\ 1 & if \quad x>1\end{cases}$$

Find $P\left(\frac{2X+6}{3}>2.2\right)$? **(2 marks)**

- **Identification of CDF : 1 mark**
- **Final value : 1 Mark**

4.

A. Write a short note on Poisson distribution

B. Calculate the values of mean and variance of the binomial distribution specified by the parameters $n = 5$ and $1\text{-}p = 0.4$? **(5+5=10 marks)**

- **Describing all the characteristics of Poisson distribution : 5 Marks**
- **Identifying the formulas for mean and variance of binomial distribution : 5 marks**

5.

A. Define the following terms

a. Parameter

b. Estimate

c. Statistic

d. Sampling distribution

e. Standard error

B. What you mean by an unbiased estimator? Let X be a random variable following a normal distribution with mean μ and variance σ^2. Propose an estimator for the population mean μ. Check whether the proposed estimator is an unbiased estimator for the population mean μ.

(5+5=10 marks)

- **Defining of all the terms : 1 marks each**

- **Defining unbiased estimator: 2 marks**
- **Problem solving : 3 marks**

6. Let X be a random variable following a normal distribution with mean μ and variance σ^2.

   Then,

   A. Give the sampling distribution of the sample mean.

   B. State the central limit theorem

   C. When a batch of a chemical product is prepared, the amount of an impurity (in grams) in the batch is a random variable X with: $\mu = 4.0g$ and $\sigma^2 = (1.5g)^2$. Suppose that $n = 50$ batches are prepared (independently). What is the probability that the sample mean impurity amount will be greater than 4.2grams? (Hint: $\Phi(4.2) = 0.8272$)

(3+2+5=10 marks)

- **Deriving the sampling distribution of the sample mean: 3 marks**
- **Stating the central limit theorem: 2 marks**
- **Problem solving : 5 marks**

7.

A. Derive the expression for $100(1-\alpha)\%$ confidence interval for the mean of a normal distribution $N(\mu, \sigma^2)$ when σ is unknown and the sample size is small.

- **Derivation of the expression of the confidence interval : 5 marks**

B.

a. Give the expression for $100(1-\alpha)\%$ confidence interval for the Population proportion *p*.

b. A medical study showed 57 of 300 persons failed to recover from a particular disease. Find 95 % C.I for the mortality rate of the disease.

(5+(2.5+2.5)=10 marks)

- **Stating the expression : 2.5 marks**
- **Problem solving : 2.5 marks**

8.

A. Discuss about the different type of errors in testing of hypothesis and how to manage these errors.

Defining two types of errors : 4 Marks

Managing errors : 1 Mark

B. What do you mean by critical region and discuss the decision criteria based critical region

Definition of critical region: 2 Mark

Decion criteria : 1 Mark

C. Mention the steps involved in hypothesis testing

Defining the steps: 2 marks

(5+3+2=10 Marks)

9. Discuss the following tests in detail

A. ANOVA

- **Explanation of ANOVA with hypothesis, assumptions, criteria of decision: 5 Marks**

B. Mann-Whitney U test

- **Explanation of Mann-Whitney U test with hypothesis, assumptions, criteria of decision: 5 Marks**

(5+5=10 Marks)

10. Suppose we measure the thickness of plaque (mm) in the carotid artery of 10 randomly selected patients with mild atherosclerotic disease. Two measurements are taken, thickness of plaque before treatment with Vitamin E (baseline) and after two years of taking Vitamin E daily. Do the data provide enough evidence to indicate that there is a difference in plaque before and after treatment with vitamin E for two years? [α = 0.01, $t_{1-\frac{\alpha}{2}}(9)$= 3.25]

Before	0.66	0.72	0.85	0.62	0.59	0.63	0.64	0.70	0.73	0.68
After	0.61	0.65	0.79	0.63	0.51	0.55	0.62	0.67	0.65	0.64

- **Identifying the test: 2 Marks**
- **Test statistics: 2 Marks**
- **Method of Calculations: 3 Marks**
- **Correct answer: 1 Mark**
- **Style f interpretation: 2 Marks**

(10 Marks)