SRI DURGA MALLESWARA SIDDHARTHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous College in the jurisdiction of Krishna University, Machilipatnam)

DEPARTMENT OF STATISTICS

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL231**			
Title of the Course				**Discrete Probability Distributions**			
Offered to: (Programme/s)				**B.Sc. Hons Statistics**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**III**
Course Category:	**Major**			**Course Relates to:**	Local, Regional, National, Global		
Year of Revision:	**NA**			**Percentage:**	**NA**		
Type of the Course:				Skill Development			
Crosscutting Issues of the Course :				**NA**			
Pre-requisites, if any				**Probability & Random Variable**			

NA: Not Applicable

Course Description:

Discrete probability distributions explore the fundamentals of probability theory with a focus on discrete random variables. Students will learn about key concepts such as probability mass functions, expected values, variance, and common distributions including binomial, Poisson, and geometric. The course emphasizes practical applications and problem-solving techniques, providing a solid foundation for analyzing and interpreting data in various fields. By the end of the course, students will be equipped to model real-world scenarios using discrete probability distributions and apply statistical methods to draw meaningful conclusions from data.

Course Aims and Objectives:

S.NO	COURSE OBJECTIVES
1	**Define and Describe:** Clearly define and describe key concepts related to discrete probability distributions, including probability mass functions and cumulative distribution functions.
2	**Identify and Use Distributions:** Identify and apply common discrete probability distributions, such as binomial, Poisson, and geometric, in various scenarios.
3	**Perform Calculations:** Accurately perform calculations involving probabilities, expected values, and variances for discrete random variables.
4	**Interpret Results:** Interpret the results of probability computations and understand their implications in real-world contexts.
5	**Apply Problem-Solving Techniques:** Utilize discrete probability distributions to solve practical problems, employing appropriate statistical methods and techniques.

Course Outcomes

At the end of the course, the student will be able to…

CO NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Understand and explain key concepts in discrete probability distributions, such as probability mass functions, cumulative distribution functions, expected value, variance, and statistical moments.	K2	2	2
CO2	Apply common discrete distributions, including the binomial, Poisson, geometric, hypergeometric, and negative binomial distributions.	K3	7	1
CO3	Calculate probabilities, expected values, variances, and other statistical measures for discrete random variables.	K3	7	2

CO4	**a**pply discrete probability distributions to solve real-world problems in various fields, such as quality control, finance, and engineering.	**K3**	**7**	**1**
CO5	Evaluate statistical analyses involving discrete probability distributions.	**K5**	**7**	**1**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO 2**
CO1		**3**							**2**
CO2							**3**	**3**	
CO3							**3**		**2**
CO4							**3**	**3**	
CO5							**3**	**2**	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively
Course Structure:
Unit – 1: Uniform, Bernoulli and Binomial distributions (12Hrs)
Discrete Uniform distribution – definitions, mean, variance. Bernoulli distribution – definitions, mean, variance and its moment generating function. Binomial distribution – Definition, Moments of Binomial Distribution, Mode of Binomial Distribution, Recurrence Relation for the moments of Binomial distribution, Mean deviation about Mean of Binomial distribution, Moment generating function, Characteristic function, Cumulates and Probability generating function of Binomial distribution. Additive property of Binomial distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of Binomial distribution.
Examples/Applications/Case Studies:
1. Rolling a fair die. Each number (1-6) has an equal probability of appearing.
2. Random number generation, simulations, load balancing in computer networks.
3. Quality control, medical trials, decision making.
4. Number of heads in 10-coin flips, number of defective items in a sample.

Exercises/Projects:
1. Simulation**:** Simulate the tossing of a fair coin 1000 times and compare the observed frequency of heads to the theoretical probability.
2. **Gambling:** Analyze the probability of winning different types of lottery games using binomial distribution.
3. **Sports Analytics:** Analyze the performance of a basketball player or baseball hitter using binomial distribution and calculate relevant statistics.
4. **Data Analysis:** Collect data on a binary outcome (e.g., success/failure, yes/no) and fit a binomial distribution to the data. Analyze the goodness of fit.

Specific Resources: (web)
Stat Trek: Offers online tutorials and calculators for statistical concepts.
Link: https://stattrek.com/
Unit – 2: Poisson distribution (12Hrs)

Poisson distribution – Definition, Moments of Poisson distribution, Mode of Poisson distribution, Recurrence Relation for the moments of Poisson distribution, Moment generating function, Characteristic function, Cumulates and Probability generating function of Poisson distribution. Additive or reproductive property of Poisson distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of Poisson distribution. Poisson distribution as a limiting case of Binomial distribution.

Examples/Applications/Case Studies:

1. **Number of calls received by a call center in an hour.** The arrival of calls can be considered independent events occurring at a constant rate.
2. **Number of cars arriving at a toll booth in a minute.** Similar to call center example, car arrivals can be modelled as independent events with a constant rate.
3. **Case Study: Call Center Staffing**
   A call center wants to determine the optimal number of agents to handle incoming calls. By modelling the arrival of calls as a Poisson process, the call center can calculate the average number of calls per hour and the probability of different call volumes. This information can be used to optimize staffing levels and minimize customer wait times.

Exercises/Projects:

Project 1: Real-Life Application

Objective: Apply the Poisson distribution to model a real-life scenario.

1. **Choose a Scenario**: For instance, modeling the number of phone calls received by a customer service center per hour.
2. **Data Collection**: Collect data on the number of calls received over several hours.
3. **Model Fitting**: Fit a Poisson model to the data. Estimate the parameter λ\lambda and check if the model is a good fit using goodness-of-fit tests.
4. **Analysis**: Analyze the results. Compare the predicted probabilities (e.g., probability of receiving exactly 10 calls in an hour) with observed frequencies.

Project 2: Simulation and Visualization

Objective: Simulate Poisson random variables and visualize their properties.

1. **Simulation**: Write a program (using Python, R, etc.) to simulate Poisson random variables with different parameters.
2. **Visualization**:
   - Plot histograms of the simulated data for various λ values.
   - Show how the distribution changes with different λ.
   - Plot the theoretical PMF and compare it with the histogram from the simulation.
3. **Analysis**: Calculate and compare the empirical moments and cumulants with theoretical values. Verify the additive and reproductive properties through simulations.

Specific Resources: (web)

https://stattrek.com/probability-distributions/poisson.aspx

Unit – 3: Negative Binomial Distribution (12Hrs)

Negative Binomial distribution – Definition, Moments of Negative Binomial Distribution,Recurrence Relation for the moments of Negative Binomial distribution, Moment generating function, Characteristic function, Cumulants and Probability generating function of Negative Binomial distribution. Additive property of Negative Binomial distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of Negative Binomial distribution.Poisson distribution as a limiting case of Negative Binomial distribution.

Examples/Applications/Case Studies:

1. **Modelling the Number of Trials Until a Fixed Number of Successes**
   Scenario: Consider a factory where each machine has a probability p of producing a defective part. You are interested in the number of machines needed to find a fixed number r of defective parts.

Application: This scenario can be modelled using the Negative Binomial distribution, where r represents the number of defective parts (successes), and the trials represent machines inspected until the rth defective part is found.

2. **Modelling the Number of Failures before a Set Number of Successes**
   Scenario: In a clinical trial, suppose you are examining the number of patients who must be tested before achieving a certain number of successful treatments. If each treatment has a success probability p, the number of trials needed to achieve r successes follows a Negative Binomial distribution.
   Application: This distribution helps in understanding how many patients need to be treated before reaching the desired number of successful outcomes.
3. **Case Study: Insurance Claims**
   Scenario: Insurance companies often use the Negative Binomial distribution to model the number of claims made by policyholders before a certain number of claims is reached. For instance, an insurer might want to model the number of policies that must be written before observing a certain number of large claims. Application: This helps in risk assessment and setting premium levels based on the number of claims and the probability distribution of these claims.

Exercises/Projects:

1. Coin Tossing Simulation

Objective: Simulate a negative binomial experiment by tossing a coin repeatedly until a specified number of heads is obtained.

- Steps:

a) Define the desired number of heads (r).
b) Simulate coin tosses using a random number generator.
c) Count the number of tails (failures) before obtaining r heads.
d) Repeat the experiment multiple times to create a dataset.
e) Calculate the mean, variance, and other statistics of the simulated data.
f) Compare the results with the theoretical values for a negative binomial distribution with parameters r and p (probability of heads).

2. Simulation and Application of the Negative Binomial Distribution

Objective: To simulate the Negative Binomial distribution and understand its properties through empirical data and compare with theoretical values.

Instructions:

Simulation:

Write a program (in Python, R, etc.) to simulate random variables from a Negative Binomial distribution with given parameters r and p

Generate a large sample (e.g., 10,000 samples) for various sets of r and p values.

Empirical Analysis:

Compute the sample mean and variance from the simulated data.

Compare the empirical mean and variance with the theoretical values derived from the distribution's parameters.

Visualization:

Plot histograms of the simulated data for different values of r and p.

Create plots to compare the empirical distribution with the theoretical PMF.

Application:

Use the simulated data to solve a practical problem, such as modelling the number of trials needed to achieve a certain number of successes in a sequence of Bernoulli trials.

Analyse the skewness and kurtosis of the simulated data and compare these with theoretical values.

Specific Resources: (web)

https://www.khanacademy.org/math/statistics-probability

Unit – 4: Geometric Distribution (12Hrs)

Geometric distribution – Definition, Moments of Geometric distribution, Moment generating function, Characteristic function, Cumulants and Probability generating function of geometric distribution. Additive property of geometric distribution. Measure of skewness, kurtosis and problems. The first two moments are obtained through moment generating function, Recurrence relation for the probabilities of geometric distribution. Lack of memory property.

Examples/Applications/Case Studies:

1.Modelling the Number of Trials Until the First Success

Scenario: Consider a quality control test where each unit produced has a probability p of passing the test. You want to know the number of units that need to be tested until the first unit passes the test.

Application: This scenario is modelled using the Geometric distribution, where X represents the number of trials until the first success. The probability mass function (PMF) of the Geometric distribution is given by: $P(X=k)=q^{k-1}p,\ k=1,2,3,...$

where k is the trial number of the first success, and p is the probability of success in each trial.

2. Example: Modelling the Number of Calls Until a Sale

Scenario: In a telemarketing campaign, the probability of making a sale on any given call is p. You want to determine the number of calls needed to achieve the first sale.

Application: This scenario can be modelled using the Geometric distribution to determine the expected number of calls required. For instance, if p=0.1, you can calculate the expected number of calls needed using the mean of the Geometric distribution, which is $\frac{1}{p}$.

Exercises/Projects:

Simulation and Application of the Geometric Distribution

Objective:

To simulate the Geometric distribution and analyse empirical data to understand its properties and applications.

Instructions:

1. **Simulation**:
   - Write a program (using Python, R, etc.) to simulate random variables from a Geometric distribution with parameter p.
   - Generate a large sample (e.g., 10,000 samples) for various values of p.
2. **Empirical Analysis**:
   - Compute the sample mean and variance from the simulated data.
   - Compare the empirical mean and variance with the theoretical values E(X) and V(x)
3. **Visualization**:
   - Plot histograms of the simulated data for different values of p.
   - Create plots to compare the empirical PMF with the theoretical PMF.
4. **Application**:
   - Use the simulated data to solve a practical problem. For example, model the number of trials required to achieve the first success in a sequence of Bernoulli trials and use the results to make predictions or decisions.
   - Analyse and interpret the results in the context of the application. For instance, if you model the number of calls needed before achieving a sale, discuss how the results can inform sales strategies.
5. **Lack of Memory Property**:
   - Demonstrate the lack of memory property of the Geometric distribution through simulation. For example, simulate the number of trials needed after a certain number of trials have already occurred and verify that the distribution of the remaining trials follows the same Geometric distribution.

Unit – 5: Hyper Geometric Distribution (12Hrs)

Hyper Geometric Distribution – Definition, mean and variance, problems. Recurrence relation for probabilities. Limiting case of Hyper Geometric distribution to Binomial distribution.

Examples/Applications/Case Studies:

1. Example: Quality Control in Manufacturing

Scenario: Suppose a factory produces a batch of 100 items, 15 of which are defective. A quality control inspector randomly selects 10 items from the batch for inspection. You want to find the probability that exactly 3 out of the 10 inspected items are defective.

2. Case Study: Election Polling

Scenario: During an election, a poll is conducted to estimate the proportion of voters who support a particular candidate. Suppose there are 500 registered voters, of whom 200 are known to support the candidate. If a poll of 50 voters is randomly selected, what is the probability that exactly 20 of them support the candidate?

Exercises/Projects:

Simulation and Limiting Case to Binomial Distribution

Objective: To simulate the Hypergeometric distribution and examine its approximation to the Binomial distribution as the population size becomes large.

Instructions:

1. Simulation: Write a program (in Python, R, etc.) to simulate random variables from the Hypergeometric distribution. Use parameters N, K, and n to generate a large sample (e.g., 10,000 samples).
2. Empirical Analysis:
   Compute the sample mean and variance from the simulated data. Compare these with the theoretical values derived from the Hypergeometric distribution. Plot histograms of the simulated data for different parameter values.
3. Limiting Case Analysis:
   Simulate the Hypergeometric distribution with large N and fixed n (e.g., N=1,000 and n=50). Compare the Hypergeometric distribution with the corresponding Binomial distribution where $p = \frac{k}{N}$
4. Comparison:
   Compare the simulated data of the Hypergeometric distribution with the Binomial distribution using statistical tests or visual plots to evaluate the approximation.Show how the Binomial distribution approximates the Hypergeometric distribution as N increases while keeping n constant.
5. **Application**:
   Apply the results to a real-world scenario, such as quality control or polling, and discuss how the approximation affects decision-making.

TEXT BOOKS:

1. Gupta. S.C. &Kapoor,V.K. (2023) . Fundamentals of Mathematical Statistics, Sultan Chand & Sons Pvt. Ltd. New Delhi.

REFERENCES:

1. Bansilal and Arora (1989). New Mathematical Statistics, Satya Prakashan, New Delhi.
2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.
3. Mukhopadhyay, P. (2015). Mathematical Statistics.Publisher: BOOKS AND ALLIED (1 January 2016)

***

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code		**23STMAL231**	
Title of the Course		**Discrete Probability Distributions**	
Year of Introduction:	**2024-25**	**Semester:**	**III**

NA: Not Applicable

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a. Define Bernoulli distribution and write its applications (Co-1, K-1)

(OR)

b. State binomial distribution and write its applications (Co-1, K-1)

2. a. Define Poisson distribution and derive its mean and variance. (Co-2, K-1)

(OR)

b. Find mean deviation about mean of a Poisson distribution. (Co-2, K-1)

3. a. Obtain first four central moments of negative binomial distribution (Co-3, K-1)

(OR)

b. Find cumulates of negative binomial distribution (Co-3, K-1)

4. a. Define geometric distribution and obtain its mean and variance (Co-4, K-1)

(OR)

b. Define geometric Distribution and its applications (Co-4, K-1)

5. a. Explain hyper geometric Distribution and its applications (Co-5, K-2)

(OR)

b. Define geometric distribution and obtain its mean and variance (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. Derive recurrence relation for the moments of Binomial distribution. (Co-1, K-3)

(OR)

b. Fit a Binomial Distribution, by direct method for the following data: (Co-1, K-1)

X	0	1	2	3	4	5	6
F	7	64	140	210	132	75	12

7. a. Derive Poisson distribution as limiting form of Binomial distribution. (Co-2, K-3)

(OR)

b. Fit a Poisson distribution, by direct method for the following data: (Co-2, K-1)

No of deaths	0	1	2	3	4
F	122	60	15	2	1

8. a. Fit a Negative binomial distribution for the given hypothetical distribution (Co-3, K-1)

X	0	1	2	3	4	5
f	112	71	35	10	5	2

(OR)

b. Obtain MGF of Negative binomial distribution hence obtain mean and variance of it. (Co-3, K-1)

9. a.Explain memory less property of geometric distribution. (Co-4, K-2)

(OR)

b. Obtain mean, variance and MGF of geometric distribution. (Co-4, K-1)

10. a. Show that Binomial distribution as a limiting case of hyper geometric distribution. (Co-5, K-2)

(OR)

b. Explain recurrence relation for hyper geometric distribution. (Co-5, K-2)

****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10
(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAP231**			
Title of the Course				**Discrete Probability Distributions Using R**			
Offered to: (Programme/s)				**B.Sc. Hons Statistics**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**Major**			**Course Relates to:**	Local, Regional, National, Global		
Year of Revision:	**NA**			**Percentage:**	**NA**		
Type of the Course:				Skill development			
Crosscutting Issues of the Course :				**NA**			
Pre-requisites, if any				**Probability and Basic R programming**			

Course Description:
The course "Discrete Probability Distributions Using R" offers a comprehensive introduction to the theory and application of discrete probability distributions, with a focus on practical implementation using R. Students will explore key distributions such as Bernoulli, Binomial, Poisson, Geometric, and Hypergeometric, gaining a solid understanding of their properties, applications, and interrelationships. Through hands-on exercises in R, learners will simulate distributions, compute probabilities, and analyze real-world data, equipping them with the tools to apply statistical concepts in various fields. The course emphasizes both theoretical foundations and practical skills, making it ideal for students aiming to deepen their statistical knowledge and computational proficiency.

Course Aims and Objectives:

S.NO	COURSE OBJECTIVES
1	Understand the fundamental concepts of discrete probability distributions.
2	Calculate and interpret key statistical measures for discrete probability distributions.
3	Apply discrete probability distributions to real-world problems.
4	Perform hypothesis testing using discrete probability distributions.
5	Compare and contrast different discrete probability distributions.

Course Outcomes
At the end of the course, the student will be able to…

CO NO	COURSE OUTCOME	BTL	PO	PSO
CO1	apply R effectively to simulate, analyze, and visualize various discrete probability distributions (e.g., binomial, Poisson, geometric).	**K3**	**7**	**1**
CO2	calculate and interpret key statistical measures such as mean, and Variance of these distributions.	**K3**	**7**	**1**
CO3	apply appropriate distributions to solve problems and make informed decisions.	**K3**	**7**	**1**
CO4	critically analyze and solve problems involving discrete probability distributions.	**K4**	**7**	**2**
CO5	evaluate the appropriateness of different distributions for given scenarios.	**K5**	**7**	**1**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1							**3**	**3**	
CO2							**2**	**2**	
CO3							**3**	**3**	
CO4							**3**		**3**
CO5							**3**	**3**	

Use the codes 3,2,1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

This lab list covers the key areas of a ….(title of the course) course, providing hands-on practice with ….(technology/software)

Unit 1: Binomial Distribution (6Hrs)

Lab 1: Modelling Coin Tosses with a Binomial Distribution

Dataset: Simulated or real-world data on coin tosses.

Experiment:

1. Simulate a series of coin tosses using a random number generator.
2. Count the number of heads and tails.
3. Calculate the probability of success (e.g., heads) based on the simulation results.

Tasks:

1. Calculate the mean and variance of the binomial distribution.
2. Compare the simulated results with the theoretical values for a binomial distribution.
3. Explore the effect of changing the number of trials and the probability of success on the distribution.

Lab 2: Fitting of Binomial distribution

Dataset: Real world data set

Tasks:

1. Calculate the mean and variance of the binomial distribution and also calculate Probability of success.
2. Compare the expected results using recurrence relation for probabilities with the theoretical values for a binomial distribution.

Unit 2: Poisson distribution 6Hrs

Lab 1: Simulating and analysing a Poisson distribution

Dataset: Simulated data generated in R.

Experiment:

1. Generate a random sample from a Poisson distribution using the **rpois** function in R.
2. Calculate the mean, variance, and standard deviation of the sample.

Tasks:

1. Compare the calculated statistics with the theoretical values for a Poisson distribution.
2. Create a histogram to visualize the distribution of the simulated data.
3. Explore the effect of changing the rate parameter (lambda) on the distribution.

Lab 2: Fitting of Poisson distribution

Dataset: Real world data set

Tasks:

1. Calculate the mean and variance of the Poisson distribution and also calculate lambda value.
2. Compare the expected results using recurrence relation for probabilities with the theoretical values for a Poisson distribution.

Unit 3: Geometric Distribution 6Hrs

Lab 1: Simulating and Analysing a Geometric Distribution

Dataset: Simulated data generated in R.

Experiment:

1. Generate a random sample from a geometric distribution using the**rgeom** function in R.
2. Calculate the mean, variance, and standard deviation of the sample.

Tasks:
1. Compare the calculated statistics with the theoretical values for a geometric distribution.
2. Create a histogram to visualize the distribution of the simulated data.
3. Explore the effect of changing the probability of success (p) on the distribution.

Lab 2: Fitting of Geometric distribution
Dataset: Real world data set
Tasks:
1. Calculate the mean and variance of the geometric distribution and also calculate Probability of success.
2. Compare the expected results using recurrence relation for probabilities with the theoretical values for a geometric distribution.

Unit 4: Negative Binomial Distribution **6Hrs**
Lab 1: Simulating and Analysing a Negative Binomial Distribution
Dataset: Simulated data generated in R.
Experiment:
1. Generate a random sample from a negative binomial distribution using the rnbinom function in R.
2. Calculate the mean, variance, and standard deviation of the sample.

Tasks:
1. Compare the calculated statistics with the theoretical values for a negative binomial distribution.
2. Create a histogram to visualize the distribution of the simulated data.
3. Explore the effect of changing the size parameter (r) and the probability of success (p) on the distribution.

Lab 2: Fitting of Negative Binomial Distribution
Dataset: Real world data set
Tasks:
1. Calculate the mean and variance of the negative binomial distribution and also calculate r and p.
2. Compare the expected results using recurrence relation for probabilities with the theoretical values for a negative binomial distribution.

Unit 5: Hypergeometric Distribution

Lab 1: Simulating and Analysing a Hypergeometric Distribution
Dataset: Simulated data generated in R.
Experiment:
1. Generate a random sample from a hypergeometric distribution using the rhyper function in R.
2. Calculate the mean, variance, and standard deviation of the sample.

Tasks:
1. Compare the calculated statistics with the theoretical values for a hypergeometric distribution.
2. Create a histogram to visualize the distribution of the simulated data.
3. Explore the effect of changing the population size (N), the number of successes in the population (K), and the sample size (n) on the distribution.

Lab 2: Modelling Real-World Data with a Hypergeometric Distribution
Dataset: Real-world data on the number of successes in a sample drawn without replacement from a finite population.
Experiment:
1. Import the dataset into R.

2. Fit a hypergeometric distribution to the data using the dhyper function to calculate probabilities.

Tasks:

1. Calculate the estimated population size (N) and the number of successes in the population (K) based on the sample data.
2. Conduct a goodness-of-fit test (e.g., chi-square test) to assess how well the hypergeometric distribution fits the data.
3. Use the fitted hypergeometric model to make predictions or inferences about the population parameters.

Lab Manual:

1. Discrete Probability distributions using R programming –Sri K. Siva Naga Raju, 2024 First Edition, Department of Statistics, PBSCAS

References:

1. Zaven A. Karian, Edward J. Dudewicz, 2010, **Handbook of Fitting Statistical Distributions with R, 1st edition, Chapman and Hall/CRC.**
2. Vito Ricci, 2005, FITTING DISTRIBUTIONS WITH R, R-Cran Projects. https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

***

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipatnam)

Course Code				**23STMAL232**			
Title of the Course				**Continuous Probability Distributions**			
Offered to: (Programme/s)				**B.Sc. Hons Statistics**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**Major**			**Course Relates to:**	Local, Regional, National, Global		
Year of Revision:	**NA**			**Percentage:**	**NA**		
Type of the Course:				Skill Development			
Crosscutting Issues of the Course:				**NA**			
Pre-requisites, if any				**Probability & Random Variable**			

NA : Not Applicable

Course Description:

Continuous probability distributions explore the fundamentals of probability theory with a focus on continuous random variables. Students will learn about key concepts such as probability density functions, expected values, variance, and common distributions including rectangular, exponential, Laplace, beta, gamma, Cauchy, normal and log - normal . The course emphasizes practical applications and problem-solving techniques, providing a solid foundation for analyzing and interpreting data in various fields. By the end of the course, students will be equipped to model real-world scenarios using continuous probability distributions and apply statistical methods to draw meaningful conclusions from data.

Course Aims and Objectives:

S.NO	COURSE OBJECTIVES
1	**Define and Describe:** Clearly define and describe key concepts related to continuous probability distributions, including probability density functions and cumulative distribution functions.
2	**Identify and Use Distributions:** Identify and apply common continuous probability distributions, such as exponential, uniform, beta, gamma, and normal in various scenarios.
3	**Perform Calculations:** Accurately perform calculations involving probabilities, expected values, and variances for continuous random variables.
4	**Interpret Results:** Interpret the results of probability computations and understand their implications in real-world contexts.
5	**Apply Problem-Solving Techniques:** Utilize continuous probability distributions to solve practical problems, employing appropriate statistical methods and techniques.

Course Outcomes

At the end of the course, the student will be able to…

CO NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Understand and explain key concepts in continuous probability distributions, such as probability mass functions, cumulative distribution functions, expected value, variance, and statistical moments.	**K2**	**2**	**1**
CO2	Apply common continuous distributions, including the normal, exponential, uniform, gamma, beta, and t distributions.	**K3**	**7**	**2**
CO3	Calculate probabilities, expected values, variances, and other statistical measures for continuous random variables.	**K3**	**7**	**1**
CO4	Apply continuous probability distributions to solve real-world problems in various fields, such as quality control, finance, and engineering.	**K3**	**7**	**2**
CO5	Evaluate statistical analyses involving continuous probability distributions.	**K5**	**7**	**1**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO MATRIX									
CO NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PSO1	PSO2
CO1		3						3	
CO2							2		1
CO3							3	2	
CO4							3		3
CO5							3	3	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure:

Unit – 1: Continuous Uniform distribution (12Hrs)

Continuous Uniform distribution – Definition, distribution function, Moments of Continuous Uniform distribution,Median of Continuous Uniform distribution, Mean deviation about Mean of Continuous Uniform distribution, Moment generating function and Characteristic function,of Continuous Uniform distribution. Measure of skewness, kurtosis and problems.

Examples/Applications/Case Studies:

1. **Random Number Generation:** Many programming languages and statistical software use the continuous uniform distribution to generate random numbers within a specified range.
2. **Simulation Studies:** In various fields, such as engineering, finance, and social sciences, the continuous uniform distribution is often used to simulate random events or variables. For example, it can be used to simulate the arrival time of customers in a queueing system or the random errors in a measurement process.
3. **Sampling:** The continuous uniform distribution can be used to select random samples from a population, ensuring that each member of the population has an equal chance of being selected.

Exercises/Projects:

Project 1: Simulating Dice Rolls

Goal: Simulate the rolling of a fair six-sided die using the continuous uniform distribution.

Steps:

1. **Generate random numbers:** Use the runif() function in R to generate a large number of random numbers between 1 and 6.
2. **Round to integers:** Round the generated numbers to the nearest integer to simulate a die roll.
3. **Analyze results:** Calculate the frequency of each outcome, compare it to the expected frequency (1/6 for each outcome), and visualize the results using a histogram.

Project 2: Monte Carlo Simulation for Pi Estimation

Goal: Estimate the value of pi using a Monte Carlo simulation based on the continuous uniform distribution.

Steps:

1. **Generate random points:** Generate a large number of random points within a square with side length 2, centered at the origin.
2. **Calculate distances:** Calculate the distance of each point from the origin.
3. **Count points within the circle:** Determine how many of the generated points fall within a circle with radius 1, centered at the origin.
4. **Estimate pi:** Calculate the estimated value of pi using the formula: pi ≈ 4 * (number of points within the circle) / (total number of points)

Specific Resources (Web):

1. https://univ.jeanpaulcalvi.com/Posters/ConfAuchWeb/abramovitz2.pdf
2. **Stat Trek:** https://www.youtube.com/watch?v=bPEctT-xb1o

Unit – 2: Exponential Distribution (12Hrs)

Exponential distribution – Definition, Moments of Exponential distribution,Median of Exponential distribution, Moment generating function and Characteristic function of Exponential distribution.The first four moments are obtained through moment generating function, Memoryless property of Exponentialdistribution.Measure of skewness, kurtosis and problems.

Double Exponentialor Laplace Distribution- Definition, Moments of Double Exponential distribution,Moment generating function and Characteristic function of double Exponential distribution. The first four moments are obtained through characteristic function,Measure of skewness, kurtosis and problems.

Examples/Applications/Case Studies:

1. **Customer Service:**

Scenario: A call center receives an average of 20 calls per hour.

Analysis: The time between calls follows an exponential distribution. This can be used to model wait times, staffing requirements, and system performance.

2. **Equipment Failure:**

Scenario: A machine has a mean time between failures (MTBF) of 1000 hours.

Analysis: The time until the next failure follows an exponential distribution. This can be used to predict equipment reliability, maintenance schedules, and replacement costs.

Exercises/Projects:

1. **Queuing System Simulation**

Project Goal: Simulate a queuing system (e.g., a call center, a grocery store checkout) using the exponential distribution to model inter-arrival and service times.

Steps:

1. **Define parameters:** Determine the average arrival rate and service rate for the system.
2. **Generate inter-arrival and service times:** Use the exponential distribution to generate random inter-arrival and service times for customers.
3. **Simulate the system:** Create a simulation model to track customer arrivals, waiting times, and service times.
4. **Analyze performance:** Calculate metrics like average waiting time, system utilization, and queue length.

2. **Financial Modeling**

Project Goal: Use the exponential distribution to model the time until a financial event (e.g., a market crash, a default) occurs.

Steps:

1. **Determine mean time:** Estimate the average time until the event based on historical data or expert opinion.
2. **Model the time:** Represent the time until the event using an exponential distribution with the estimated mean as the rate parameter.
3. **Calculate probabilities:** Calculate the probability of the event occurring within a given time period.
4. **Risk assessment:** Use the model to assess the risk of the event and its potential impact.

Specific Resources: (web)

OpenIntro Statistics: https://www.openintro.org/book/os/

Unit – 3: Gamma and Beta Distributions (12Hrs)

Gamma distribution – Definition, Moments of Gamma Distribution, Moment generating function, Characteristic function, Cumulate generating function of Gamma distribution. Additive property of Gamma distribution. Measure of skewness, kurtosis and problems. The first four moments are obtained through cumulate generating function, limiting form of Gamma distribution.

Beta Distribution of First Kind and Second kind – Definition, Mean, Variance and Harmonic mean. Simple problems.

Examples/Applications/Case Studies:

Example1: Waiting Time for a Bus for Gamma Distribution

Scenario: A bus arrives at a stop every 15 minutes on average. The time between bus arrivals follows an exponential distribution. What is the probability that you will wait more than 30 minutes for the next bus?

Analysis:

1. The time between bus arrivals is exponentially distributed with a rate parameter $\lambda = 1/15$ (since the average waiting time is 15 minutes).
2. The gamma distribution with shape parameter $\alpha = 1$ and rate parameter $\lambda = 1/15$ is equivalent to the exponential distribution.
3. Therefore, the probability of waiting more than 30 minutes is $P(X > 30) = 1 - F(30)$, where $F(x)$ is the cumulative distribution function of the gamma distribution.

Example 2: Proportion of Defective Items for Beta Distribution

Scenario: A quality control inspector examines a sample of 10 items from a production line. If 3 of the items are defective, what is the probability that the proportion of defective items in the entire population is between 0.2 and 0.4?

Analysis:

1. The proportion of defective items in the population can be modeled using a beta distribution with shape parameters $\alpha = 4$ (number of successes) and $\beta = 7$ (number of failures).
2. The probability of the proportion being between 0.2 and 0.4 is $P(0.2 < X < 0.4) = F(0.4) - F(0.2)$, where $F(x)$ is the cumulative distribution function of the beta distribution.

Exercises/Projects:

Project 1: Modeling Customer Lifetime Value (CLTV) Using the Gamma Distribution

Goal: Use the gamma distribution to model the lifetime value of customers in a business.

Steps:

1. **Collect data:** Gather data on customer purchase history, demographics, and other relevant factors.
2. **Calculate customer lifetime:** Determine the lifetime of each customer based on their purchase history.
3. **Fit a gamma distribution:** Fit a gamma distribution to the customer lifetime data.
4. **Calculate CLTV:** Use the fitted gamma distribution to estimate the expected lifetime value of a customer.
5. **Analyze customer segments:** Identify different customer segments based on their CLTV and analyze their characteristics.

Project 2: Bayesian A/B Testing with Beta Priors

Goal: Use the beta distribution as a prior distribution in Bayesian A/B testing to compare the effectiveness of two different treatments or conditions.

Steps:

1. **Define prior distributions:** Choose appropriate beta priors for the success probabilities of the two treatments based on prior knowledge or assumptions.
2. **Collect data:** Gather data on the number of successes and failures for each treatment.
3. **Update posterior distributions:** Use Bayes' theorem to update the prior distributions based on the observed data.
4. **Calculate Bayes factor:** Calculate the Bayes factor to compare the evidence in favor of one treatment over the other.
5. **Make a decision:** Based on the Bayes factor and other factors, make a decision about whether one treatment is significantly better than the other.

Specific Resources: (web)

https://www.khanacademy.org/math/statistics-probability

Unit – 4: Normal Distribution (12Hrs)

Normal distribution – Definition, Properties of normal distribution, importance of normal distribution, Moment generating function, Characteristic function, Cumulant generating function of normal distribution. Additive property of Normal distribution. Mean, Median and Mode, Even and Odd order moments about mean of normal distribution. Linear combination of normal variates, points of inflexion of normal probability curve. Measure of skewness, and kurtosis.

Standard Normal Distribution – Definition, Moment generating function and Characteristic function, Mean and Variance, Area property and problems.

Examples/Applications/Case Studies:

Case Study 1: Analyzing Student Test Scores

Scenario: A teacher wants to analyze the distribution of test scores for a class of 30 students.

Steps:

1. **Collect data:** Gather the test scores for all 30 students.
2. **Calculate statistics:** Calculate the mean, median, mode, standard deviation, and variance of the test scores.
3. **Visualize the distribution:** Create a histogram and a normal probability plot to assess the normality of the data.
4. **Test for normality:** Use a statistical test, such as the Shapiro-Wilk test or the Kolmogorov-Smirnov test, to formally test for normality.
5. **Analyze results:** If the data is normally distributed, you can use statistical methods based on the normal distribution, such as t-tests or confidence intervals. If the data is not normally distributed, you may need to consider alternative methods, such as nonparametric tests.

Case Study 2: Quality Control in Manufacturing

Scenario: A manufacturing company wants to monitor the quality of a product by measuring its weight. The target weight is 100 grams, and the standard deviation is known to be 5 grams.

Steps:

1. **Collect data:** Measure the weight of a random sample of products.
2. **Calculate statistics:** Calculate the sample mean and standard deviation.
3. **Test for normality:** Use a statistical test to assess the normality of the data.
4. **Construct a confidence interval:** Construct a confidence interval for the population mean weight.
5. **Make a decision:** If the confidence interval includes the target weight of 100 grams, the manufacturing process is considered to be in control. If not, corrective actions may be necessary.

Exercises/Projects:

Project 1: Analyzing Stock Returns

Goal: Analyze the distribution of stock returns and test for normality.

Steps:

1. **Collect data:** Gather historical stock price data for a specific stock or index.
2. **Calculate returns:** Calculate the daily or weekly returns of the stock or index.
3. **Visualize the distribution:** Create a histogram and a normal probability plot to assess the normality of the returns.
4. **Test for normality:** Use a statistical test, such as the Shapiro-Wilk test or the Kolmogorov-Smirnov test, to formally test for normality.
5. **Analyze results:** If the returns are normally distributed, you can use statistical methods based on the normal distribution, such as hypothesis testing or risk analysis. If the returns are not normally distributed, you may need to consider alternative models or techniques.

Project 2: Modeling Customer Satisfaction

Goal: Model customer satisfaction scores using a normal distribution.

Steps:

1. **Collect data:** Gather customer satisfaction scores for a specific product or service.

2. **Analyze the distribution:** Create a histogram and a normal probability plot to assess the normality of the satisfaction scores.
3. **Fit a normal distribution:** Fit a normal distribution to the data to estimate the mean and standard deviation of the satisfaction scores.
4. **Calculate probabilities:** Use the fitted normal distribution to calculate the probability of a customer having a satisfaction score above a certain threshold.
5. **Make inferences:** Use the model to make inferences about customer satisfaction, such as identifying areas for improvement or assessing the impact of changes to the product or service.

Specific Resources: (web)

https://stattrek.com/probability-distributions/poisson.aspx

Unit – 5: Exact Sampling Distributions- (12Hrs)

χ2 - Distribution– Definition, Properties and applications.

Student 's t- distribution– Definition, Properties and applications.

F – Distribution – Definition, Properties and applications.

Examples/Applications/Case Studies:

Case Study 1: Sampling Distribution of Sample Mean

Scenario: A researcher wants to estimate the average height of adult males in a city. They randomly sample 100 adult males and measure their heights.

Steps:

1. **Collect data:** Collect the heights of the 100 adult males.
2. **Calculate sample mean:** Calculate the mean height of the sample.
3. **Repeat sampling:** Repeat steps 1 and 2 many times (e.g., 10,000 times) to create a distribution of sample means.
4. **Analyze sampling distribution:** Analyze the distribution of sample means. It should be approximately normal with a mean equal to the population mean and a standard deviation equal to the population standard deviation divided by the square root of the sample size (standard error of the mean).
5. **Make inferences:** Use the sampling distribution to make inferences about the population mean height. For example, construct a confidence interval or conduct a hypothesis test.

Case Study 2: Sampling Distribution of Sample Proportion

Scenario: A political pollster wants to estimate the proportion of voters who support a particular candidate. They randomly sample 500 voters and ask them their preference.

Steps:

1. **Collect data:** Record the number of voters who support the candidate and the total number of voters sampled.
2. **Calculate sample proportion:** Calculate the proportion of voters who support the candidate.
3. **Repeat sampling:** Repeat steps 1 and 2 many times (e.g., 10,000 times) to create a distribution of sample proportions.
4. **Analyze sampling distribution:** Analyze the distribution of sample proportions. It should be approximately normal with a mean equal to the population proportion and a standard deviation equal to the square root of (p*(1-p))/n, where p is the population proportion and n is the sample size.
5. **Make inferences:** Use the sampling distribution to make inferences about the population proportion. For example, construct a confidence interval or conduct a hypothesis test.

Exercises/Projects:

Project 1: Simulating the Sampling Distribution of the Sample Mean

Goal: Simulate the sampling distribution of the sample mean from a normally distributed population.

Steps:

1. **Define population parameters:** Specify the population mean (μ) and population standard deviation (σ) of the normal distribution.
2. **Generate random samples:** Draw random samples of a specified size (n) from the normal distribution using the rnorm() function in R.
3. **Calculate sample means:** Calculate the mean of each sample.
4. **Repeat sampling:** Repeat steps 2 and 3 a large number of times (e.g., 10,000) to create a distribution of sample means.
5. **Visualize sampling distribution:** Plot a histogram or density plot of the distribution of sample means.
6. **Compare with theory:** Compare the observed sampling distribution to the theoretical sampling distribution, which is a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Project 2: Estimating Population Mean with Confidence Intervals

Goal: Estimate the population mean from a sample and construct confidence intervals using the sampling distribution of the sample mean.

Steps:

1. **Collect data:** Collect a random sample of size n from the population.
2. **Calculate sample mean and standard deviation:** Calculate the sample mean ($\bar{x}$) and sample standard deviation (s) of the data.
3. **Calculate standard error:** Calculate the standard error of the mean using the formula $SE = s/\sqrt{n}$.
4. **Construct confidence intervals:** Construct confidence intervals for the population mean using the t-distribution with n-1 degrees of freedom. For example, a 95% confidence interval can be calculated as $\bar{x} \pm t(\alpha/2, n-1) * SE$, where $t(\alpha/2, n-1)$ is the t-value corresponding to the desired confidence level and degrees of freedom.
5. **Interpret results:** Interpret the confidence intervals and make inferences about the population mean.

Text Books:

2. Gupta. S.C. &Kapoor,V.K. (2023) . Fundamentals of Mathematical Statistics, Sultan Chand & Sons Pvt. Ltd. New Delhi.

References:

1. Bansilal and Arora (1989). New Mathematical Statistics, Satya Prakashan, New Delhi.
2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.
3. Mukhopadhyay, P. (2015). Mathematical Statistics.Publisher: BOOKS AND ALLIED (1 January 2016)

***

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipatnam)

Course Code		**23STMAL232**	
Title of the Course		**Continuous Probability Distributions**	
Year of Introduction:	**2024-25**	**Semester:**	**3**

NA : Not Applicable

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a.Define rectangular distribution write its mean and variance. (CO-1,K-1)

(OR)

b. Find the uniform distribution whose mean is 5.5 and variance is 8.25. (CO-1,K-1)

2. a. Define exponential distribution write its applications. (CO-2,K-1)

(OR)

b.Derive moment generating function of exponential distribution. (CO-2,K-1)

3. a. Derive additive property of gamma distribution. (Co-3, K-3)

(OR)

b. Define beta second kind and write its properties. (Co-3, K-1)

4. a. Define normal distribution and write its importance. (Co-4, K-1)

(OR)

b. Show that for a normal distribution Q.D:M.D:S.D=10:12:15. (Co-4, K-2)

5. a. Define chi-square distributions and its applications. (Co-5, K-1)

(OR)

b. Explain F-Distribution and its applications. (Co-5, K-2)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. Derive the moments of rectangular distribution. (CO-1,K-3)

(OR)

b. Derive mean deviation about mean of rectangular distribution. (CO-1,K-3)

7. a. State and prove the memoryless property of the exponential distribution. (CO-2,K-1)

(OR)

b. Derive moments and characteristic function of exponential distribution. (CO-2,K-3)

8. a. Derive cumulate generating function of gamma distribution. (Co-3, K-3)

(OR)

b. .Derive mean and variance of beta distribution first kind. (Co-3, K-3)

9. a. Explain normal distribution as a limiting case of binomial distribution. (Co-4,K-2)

(OR)

b.Explain properties of normal distribution. . (Co-4, K-2)

10. a. Derive Student's t Distribution. (Co-5, K-3)

(OR)

b. Derive the relation between F and χ^2 distribution. (Co-5, K-3)

****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10
(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAP232**			
Title of the Course				**Continuous Probability Distributions Using R**			
Offered to: (Programme/s)				**B.Sc. Hons Statistics**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**Major**			**Course Relates to:**	Local, Regional, National, Global		
Year of Revision:	**NA**			**Percentage:**	**NA**		
Type of the Course:				Skill development			
Crosscutting Issues of the Course :				**NA**			
Pre-requisites, if any				**Probability and Basic R programming**			

Course Description:

The course "Continuous Probability Distributions" with practical and hands-on experience using R. This course covers key concepts such as rectangular, exponential, Laplace, beta, gamma, Cauchy, normal and log - normal distributions, alongside statistical analysis and data visualization. Gain proficiency in applying these distributions to real-world problems, from hypothesis testing to regression analysis, using R's powerful tools. Ideal for data scientists, analysts, and statisticians, the course emphasizes both theoretical understanding and practical skills. By the end, you'll be adept at performing calculations, simulations, and visualizations to tackle complex data challenges with confidence. Enhance your statistical toolkit and analytical capabilities in this comprehensive course.

Course Aims and Objectives:

S.NO	COURSE OBJECTIVES
1	Understand the fundamental concepts of continuous probability distributions.
2	Calculate and interpret key statistical measures for continuous probability distributions.
3	Apply continuous probability distributions to real-world problems.
4	Perform hypothesis testing using continuous probability distributions.
5	Compare and contrast different continuous probability distributions.

Course Outcomes

At the end of the course, the student will be able to…

CO	COURSE OUTCOME	BTL	PO	PSO
CO1	apply R effectively to simulate, analyze, and visualize various continuous probability distributions (e.g., Uniform, Normal, Exponential, etc.,).	**K3**	**7**	**1**
CO2	calculate and interpret key statistical measures such as mean, and Variance of these distributions.	**K3**	**7**	**1**
CO3	apply appropriate distributions to solve problems and make informed decisions.	**K3**	**7**	**1**
CO4	critically analyze and solve problems involving continuous probability distributions.	**K4**	**7**	**2**
CO5	evaluate the appropriateness of different distributions for given scenarios.	**K5**	**7**	**2**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1							**3**	**3**	
CO2							**3**	**3**	
CO3							**3**	**3**	
CO4							**3**		**3**
CO5							**3**		**1**

Use the codes 3,2,1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

This lab list covers the key areas of a Continuous Probability Distributions, providing hands-on practice with R - Programming

Unit 1: Continuous Uniform distribution (6Hrs)

Lab 1: Understanding and Generating Continuous Uniform Distributions in R

Title: Exploring the Continuous Uniform Distribution in R

Dataset: Simulated data generated using R's runif() function.

Experiment: Investigate the properties of the continuous uniform distribution and its applications in R.

Tasks:

1. **Generate uniform random numbers:** Create a vector of 1000 random numbers from a uniform distribution with specified minimum and maximum values.
2. **Visualize the distribution:** Plot a histogram and density plot of the generated random numbers to visualize the uniform distribution.
3. **Calculate statistical properties:** Compute the mean, variance, and standard deviation of the generated random numbers.
4. **Compare with theoretical values:** Verify that the calculated statistical properties match the theoretical values for a uniform distribution.
5. **Applications:** Explore real-world applications of the uniform distribution, such as random sampling, Monte Carlo simulations, or hypothesis testing.

Lab 2: Simulating and Analyzing Uniformly Distributed Data

Title: Simulating and Analyzing Uniformly Distributed Data in R

Dataset: Simulated data generated using R's runif() function.

Experiment: Simulate and analyze data from a continuous uniform distribution to understand its characteristics and applications.

Tasks:

1. **Generate random samples:** Create multiple samples of varying sizes from a uniform distribution with specified parameters.
2. **Calculate statistics:** Compute the mean, variance, and standard deviation for each sample.
3. **Visualize the distribution:** Plot histograms and density plots for each sample to observe the distribution's shape and variability.

Unit 2: Exponential Distribution 6Hrs

Lab 1: Understanding and Generating Exponential Distributions in R

Title: Exploring the Exponential Distribution in R

Dataset: Simulated data generated using R's rexp() function.

Experiment: Investigate the properties of the exponential distribution and its applications in R.

Tasks:

1. **Generate exponential random numbers:** Create a vector of 1000 random numbers from an exponential distribution with a specified rate parameter (λ).
2. **Visualize the distribution:** Plot a histogram and density plot of the generated random numbers to visualize the exponential distribution.

3. **Calculate statistical properties:** Compute the mean, variance, and standard deviation of the generated random numbers.
4. **Compare with theoretical values:** Verify that the calculated statistical properties match the theoretical values for an exponential distribution.
5. **Applications:** Explore real-world applications of the exponential distribution, such as modeling waiting times, reliability analysis, or queuing systems.

Lab 2: Simulating and Analyzing Exponentially Distributed Data
Title: Simulating and Analyzing Exponentially Distributed Data in R
Dataset: Simulated data generated using R's rexp() function.
Experiment: Simulate and analyze data from an exponential distribution to understand its characteristics and applications.
Tasks:
1. **Generate random samples:** Create multiple samples of varying sizes from an exponential distribution with specified rate parameters.
2. **Calculate statistics:** Compute the mean, variance, and standard deviation for each sample.
3. **Visualize the distribution:** Plot histograms and density plots for each sample to observe the distribution's shape and variability.
4. **Hypothesis testing:** Conduct hypothesis tests to determine if a given sample comes from an exponential distribution.

Unit 3: Gamma and Beta Distributions 6Hrs
Lab 1: Understanding and Generating Gamma Distributions in R
Title: Exploring Gamma Distributions in R
Dataset: Simulated data generated using R's rgamma() functions.
Experiment: Investigate the properties of the gamma distributions and their applications in R.
Tasks:
3. **Generate gamma random numbers:** Create vectors of 1000 random numbers from gamma and beta distributions with specified parameters (shape and rate for gamma).
4. **Visualize the distributions:** Plot histograms and density plots of the generated random numbers to visualize the gamma distributions.
5. **Calculate statistical properties:** Compute the mean, variance, and standard deviation of the generated random numbers.
6. **Compare with theoretical values:** Verify that the calculated statistical properties match the theoretical values for gamma and beta distributions.
7. **Applications:** Explore real-world applications of the gamma distributions, such as modeling waiting times, survival analysis.
Lab 2: Understanding and Generating Beta Distributions in R
Title: Exploring Beta Distributions in R
Dataset: Simulated data generated using R's rbeta() functions.
Experiment: Investigate the properties of the gamma and beta distributions and their applications in R.
Tasks:
1. **Generate gamma and beta random numbers:** Create vectors of 1000 random numbers from gamma and beta distributions with specified parameters (shape1 and shape2 for beta).
2. **Visualize the distributions:** Plot histograms and density plots of the generated random numbers to visualize the beta distributions.
3. **Calculate statistical properties:** Compute the mean, variance, and standard deviation of the generated random numbers.
4. **Compare with theoretical values:** Verify that the calculated statistical properties match the theoretical values for gamma and beta distributions.

5. **Applications:** Explore real-world applications of the beta distributions, such as modeling beta distributions in Bayesian statistics.

Unit 4: Normal Distribution **6Hrs**

Lab 1: Understanding and Generating Normal Distributions in R

Title: Exploring the Normal Distribution in R

Dataset: Simulated data generated using R's rnorm() function.

Experiment: Investigate the properties of the normal distribution and its applications in R.

Tasks:

1. **Generate normal random numbers:** Create a vector of 1000 random numbers from a normal distribution with specified mean and standard deviation.
2. **Visualize the distribution:** Plot a histogram and density plot of the generated random numbers to visualize the normal distribution.
3. **Calculate statistical properties:** Compute the mean, variance, and standard deviation of the generated random numbers.
4. **Compare with theoretical values:** Verify that the calculated statistical properties match the theoretical values for a normal distribution.
5. **Applications:** Explore real-world applications of the normal distribution, such as modeling measurement errors, financial returns, or population characteristics.

Lab 2: Simulating and Analyzing Normally Distributed Data

Title: Simulating and Analyzing Normally Distributed Data in R

Dataset: Simulated data generated using R's rnorm() function.

Experiment: Simulate and analyze data from a normal distribution to understand its characteristics and applications.

Tasks:

1. **Generate random samples:** Create multiple samples of varying sizes from a normal distribution with specified mean and standard deviation.
2. **Calculate statistics:** Compute the mean, variance, and standard deviation for each sample.
3. **Visualize the distribution:** Plot histograms and density plots for each sample to observe the distribution's shape and variability.
4. **Hypothesis testing:** Conduct hypothesis tests to determine if a given sample comes from a normal distribution (e.g., Shapiro-Wilk test).

Unit 5: Exact Sampling Distributions

Lab 1: Understanding and Generating Chi-Squared Distributions in R

Title: Exploring the Chi-Squared Distribution in R

Dataset: Simulated data generated using R's rchisq() function.

Experiment: Investigate the properties of the chi-squared distribution and its applications in R.

Tasks:

1. **Generate chi-squared random numbers:** Create a vector of 1000 random numbers from a chi-squared distribution with a specified degrees of freedom (df).
2. **Visualize the distribution:** Plot a histogram and density plot of the generated random numbers to visualize the chi-squared distribution.
3. **Calculate statistical properties:** Compute the mean, variance, and standard deviation of the generated random numbers.
4. **Compare with theoretical values:** Verify that the calculated statistical properties match the theoretical values for a chi-squared distribution.
5. **Applications:** Explore real-world applications of the chi-squared distribution, such as goodness-of-fit tests, tests of independence).

Lab 2: Simulating and Analyzing t-Distributed Data
Title: Simulating and Analyzing t-Distributed Data in R
Dataset: Simulated data generated using R's rt() function.
Experiment: Simulate and analyze data from a t-distribution to understand its characteristics and applications.
Tasks:

1. **Generate random samples:** Create multiple samples of varying sizes from a t-distribution with specified degrees of freedom.
2. **Calculate statistics:** Compute the mean, variance, and standard deviation for each sample.
3. **Visualize the distribution:** Plot histograms and density plots for each sample to observe the distribution's shape and variability.

Lab Manual:
Continuous Probability distributions using R programming –Sri K. Siva Naga Raju, 2024 First Edition, Department of Statistics, PBSCAS

References:

3. Zaven A. Karian, Edward J. Dudewicz, 2010, **Handbook of Fitting Statistical Distributions with R, 1st edition, Chapman and Hall/CRC.**
4. Vito Ricci, 2005, FITTING DISTRIBUTIONS WITH R, R-Cran Projects. https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

***

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL233**			
Title of the Course				**Statistical Methods**			
Offered to: (Programme/s)				**B.Sc.(Honors) - Statistics**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basic Mathematics**			

Course Description:

This course introduces the concept of bivariate random variables, their distributions, and independence. It covers curve fitting techniques using least squares for various functions. Correlation analysis explores relationships between variables through scatter diagrams, Pearson's and Spearman's coefficients. Regression analysis focuses on modeling relationships and prediction, including linear and nonlinear regression. Finally, the course delves into multiple and partial correlation, examining complex relationships among variables.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Understand the joint behavior of two random variables and their probabilistic relationships.
2	Approximate complex relationships between variables using simpler mathematical functions.
3	Measure the strength and direction of the linear relationship between two variables.
4	Model the relationship between a dependent variable and one or more independent variables.
5	Analyze the relationship between a dependent variable and multiple independent variables, considering the effects of other variables.

Course Outcomes

At the end of the course, the student will be able to…

CONO	COURSE OUTCOME	BTL	PO	PSO
CO1	Apply probability concepts to solve problems involving bivariate random variables	K3	7	1
CO2	Apply the method of least squares to fit various types of curves to data.	K3	7	1
CO3	Calculate and interpret correlation coefficients to measure the strength and direction of linear relationships between variables.	K4	7	1
CO4	Build and interpret linear regression models to predict a dependent variable based on independent variables.	K4	1	2
CO5	Analyze the relationship between a dependent variable and multiple independent variables using multiple and partial correlation coefficients.	K4	7	2

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1							**3**	**3**	
CO2							**3**	**3**	
CO3							**2**	**2**	
CO4	**1**								**2**
CO5							**3**		**3**

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Unit – 1 Bi-variate Random Variables (12 Hours)

Bivariate random variables- Definition, Discrete and Continuous - joint, marginal and conditional distributions- its properties. Distribution functions of bivariate random variables and its properties. Independence of random variables, and simple problems.

Applications:

1. If a manufacturer produces electronic components, the joint distribution of voltage and current measurements can help in identifying defects and improving product quality. (Quality Control)
2. Understanding the relationship between the failure times of different components helps in assessing the overall system reliability and designing more robust systems. (Reliability Analysis)
3. In risk management, understanding the joint distribution of financial variables like asset returns and interest rates helps in assessing the risk of financial instruments. (Risk Management)
4. During clinical trials, bivariate analysis is used to assess the relationship between treatment effects and patient outcomes.
5. In education research, bivariate random variables can analyze the relationship between variables such as student performance and socioeconomic status.

Activity 1: Joint Probability Distribution Table

a. Create a hypothetical scenario involving two discrete random variables, such as the number of heads and tails obtained when flipping two coins.
b. Construct a joint probability distribution table showing the probabilities of all possible outcomes.
c. Calculate marginal probability distributions for each random variable.
d. Calculate conditional probabilities for different events.

Activity 2: Bivariate Normal Distribution Simulation

a. Use statistical software (e.g., R, Python, MATLAB) to generate random samples from a bivariate normal distribution with specified means, variances, and correlation coefficient.
b. Create a scatter plot of the generated data points to visualize the distribution.
c. Experiment with different parameter values to observe the changes in the shape of the distribution.
d. Calculate marginal distributions and conditional distributions to understand their relationships with the joint distribution.

Unit – 2 Curve fitting (12 hours)

Introduction, Principle of least squares, fitting of straight line, Second degree polynomial or parabola, k^{th} degree polynomial, family of exponential curves and fitting of power curve. Simple problems.

Applications:

1. Fitting a Straight Line - Forecasting population growth based on past data
2. Fitting a Second-Degree Polynomial (Parabola) - Modeling projectile motion and Analyzing the relationship between dosage and drug response
3. Fitting a kth Degree Polynomial - Signal processing for noise reduction
4. Fitting Exponential Curves - Studying compound interest
5. Fitting a Power Curve - Modeling economic growth

Activity 1: Fitting a Straight Line to Predict Sales Based on Advertising Spend

a. A dataset containing advertising spend and sales figures.

b. Statistical software or a spreadsheet application (such as Excel, R, Python with libraries like pandas and scikit-learn).

Activity 2: Fitting an Exponential Growth Curve to Bacterial Growth Data

a. A dataset containing time and bacterial population counts..

b. Statistical software or a spreadsheet application (such as Excel, R, Python with libraries like pandas and scipy).

Unit – 3 Correlation (12 hours)

Meaning, Types of Correlation, Measures of Correlation – Scatter diagram, Karl Pearson's Coefficient of Correlation,Bivariate frequency distribution, Correlation coefficient for Bivariate frequency distribution, Spearman's Rank Correlation Coefficient, Properties andProblems.

Applications:

1. By analyzing the correlation between asset returns, investors can build diversified portfolios to reduce risk.
2. Correlation between asset returns is crucial for accurately assessing portfolio risk.
3. Understanding these correlations helps in identifying risk factors and developing preventive measures.
4. the correlation between stress levels and job performance can provide insights into how stress affects work outcomes.
5. Correlation analysis helps in understanding the relationship between mental health variables (e.g., depression) and various life factors (e.g., social support).

Activity 1: Correlation Scavenger Hunt

a. Divide students into groups.

b. Provide each group with a set of images or descriptions of pairs of variables (e.g., height and weight, ice cream sales and temperature, shoe size and intelligence).

c. Ask each group to categorize the pairs as positively correlated, negatively correlated, or having no correlation.

d. Discuss the findings as a class and clarify any misconceptions.

Activity 2: Correlation Coefficient Estimation

a. Show students a series of scatter plots with varying degrees of correlation (strong positive, weak positive, no correlation, weak negative, strong negative).

b. Ask students to estimate the correlation coefficient for each scatter plot.

c. Discuss the relationship between the shape of the scatter plot and the value of the correlation coefficient.

d. Introduce the formula for calculating the correlation coefficient and explain its components.

Unit – 4 Regression (12 hours)

Concept of Regression, Linear and Non Linear regression. Linear Regression – Regression lines, Regression coefficients and it properties. Regressions lines for bivariate data and simple problems. Correlation vs Regression. Explained and Unexplained variations. Coefficient of determination.

Applications:

1. **Sales forecasting:** Predicting future sales based on factors like advertising expenditure, price, and competition.
2. **Market research:** Understanding customer behavior and preferences.
3. **Financial modeling:** Predicting stock prices, interest rates, and economic indicators.
4. **Risk assessment:** Evaluating the impact of different factors on financial outcomes.
5. **Medicine:** Developing predictive models for disease progression and treatment outcomes.
6. **Agriculture:** Optimizing crop yields based on various factors.
7. **Insurance:** Assessing risk and setting premiums.

Activity 1: Real-world Data Regression Project

a. Divide students into groups.
b. Assign each group a real-world dataset (e.g., housing prices, student grades, climate data).
c. Guide students in selecting appropriate dependent and independent variables.
d. Have students create scatter plots to visualize the relationship between variables.
e. Use statistical software to fit a regression model and interpret the coefficients.
f. Discuss the findings and potential applications of the model.

Activity 2: Regression Simulation

a. Generate simulated data with a known relationship between variables (e.g., linear relationship with added noise).
b. Fit a regression model to the simulated data.
c. Compare the estimated coefficients with the true values.
d. Experiment with different sample sizes and noise levels to observe the impact on the model's accuracy.
e. Discuss the concept of overfitting and under fitting.

Unit – 5 Multiple & Partial Correlation Coefficient (12 hours)

Introduction, Concept of Coefficient of Multiple correlation, properties, determination. Concept of Coefficient of Partial correlation and determination. Relation between Total, Multiple and Partial Correlation Coefficients.

Applications:

Multiple Correlation Coefficients

1. **Finance:** Assessing the relationship between a stock's return and multiple factors like market index, industry performance, and company-specific factors.
2. **Economics:** Analyzing the impact of various economic indicators on GDP growth.
3. **Marketing:** Understanding the relationship between sales and multiple marketing variables (e.g., advertising, price, promotions).
4. **Social Sciences:** Studying the influence of multiple factors on educational attainment, income, or crime rates.

Partial Correlation Coefficient

5. **Medicine:** Investigating the relationship between two variables while controlling for the influence of a third variable (e.g., the relationship between smoking and lung cancer, controlling for age).
6. **Psychology:** Analyzing the correlation between intelligence and income, while accounting for the effect of education.
7. **Economics:** Examining the relationship between consumption and income, controlling for other factors like wealth and population.
8. **Environmental Science:** Studying the impact of air pollution on health, while considering other factors like socioeconomic status and lifestyle

Activity 1: Real-world Data Analysis

a. Select a dataset with multiple variables (e.g., housing prices, student performance, economic indicators).
b. Calculate the multiple correlation coefficient between the dependent variable and multiple independent variables.
c. Calculate partial correlation coefficients between pairs of variables while controlling for other variables.
d. Interpret the results to understand the relationships between variables.
e. Discuss the implications of the findings.

Activity 2: Simulation and Visualization

a. Generate simulated data with correlated variables.

b. Calculate multiple and partial correlation coefficients for the simulated data.
c. Visualize the relationships between variables using scatter plots and correlation matrices.
d. Experiment with different correlation structures to observe the impact on the coefficients.
e. Discuss the limitations of correlation analysis and the importance of considering other statistical methods.

Text Book

1. Fundamentals of Mathematical Statistics, 12th Edition, 10th September 2020, S. C. Gupta and V. K. Kapoor, Sultan Chand & Sons, New Delhi.

Recommended References books:

1. Probability and Statistics,Volume I, D.Biswas, New central book Agency (P) Ltd, New Delhi.
2. An outline of Statistical theory, Volume Two,3rd Edition,2010(with corrections) A.M.Goon,M.K. Gupta, B.Dasgupta ,The World Press Pvt.Ltd., Kolakota.
3. Sanjay Arora and BansiLal:. New Mathematical Statistics, SatyaPrakashan , New Delhi.

Websites of Interest: http://onlinestatbook.com/rvls/index.html.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machlipatnam)

Course Code		**23STMAL233**	
Title of the Course		**Statistical methods**	
Offered to: (Programme/s)		**B.Sc. Hons Statistics**	
Year of Introduction:	**2024-25**	**Semester:**	**III**

NA: Not Applicable **Time: 3 hours** **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a. Define bivariate random variable and its types. (CO-1,K-1)

(OR)

b.Define distribution function and its properties. (CO-1,K-1)

2. a. Derive the fitting of straight line by using least square principle. (CO-2,K-3)

(OR)

b.Define curve fitting and explain principle of least square. (CO-2,K -1)

3. a.Define correlation and write the types of correlation. (Co-3, K-1)

(OR)

b. Derive the limits of Karl's person's coefficient of correlation. (Co-3, K-3)

4. a.Derive the regression equation Yon X. (Co-4, K-3)

(OR)

b. Define regression and regression coefficient. (Co-4, K-1)

5. a.Explain partial correlation. (Co-5, K-2)

(OR)

b. Explain total, multiple and partial correlation coefficients. (Co-5, K-2)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. Explain bivariate discrete random variable. (CO-1,K-2)

(OR)

b. Find i) Marginal Distribution of X and Y ii) conditional distribution of X given y=1

x \ Y	-1	0	1
0	1/15	2/15	1/15
1	3/15	2/15	1/15
2	2/15	1/15	2/15

for the bivariate probability distribution of X on Y . (CO-1,K-1)

7. a. Derive the fitting of second degree parabola. . (CO-2,K-3)

(OR)

b.Fit a power curve to the following data. (CO-2,K-1)

X	1	2	3	4	5	6
Y	1.6	3.5	4.6	5.2	5.7	6.1

8. a. Calculate Karl's Pearson coefficient of correlation for the following data. (CO-3,K-3)

X	9	8	7	6	5	4	3	2
Y	2	3	4	5	6	7	8	9

(OR)

b.Calculate Pearson rank correlation coefficient to the following data. (Co-3,K-3)

X	50	40	32	66	47	58	48	65	46	30
Y	32	50	42	47	58	30	44	55	56	59

9. a. Explain regression and Write its properties (Co-4, K-2)

(OR)

b.Find the regression equation Y on X and X on Y for the following data. (Co-4, K-1)

X	0	1	2	3	4	5	6
Y	6	32	15	18	14	9	8

a. Explain the multiple and partial correlation coefficient. (Co-5, K-2)

(OR)

b. In a trivariate distribution it is found that r_{12}=0.7, r_{13}=0.61, r_{23}=0.4. Find the value of $r_{23.1}$, $r_{13.2}$ and $r_{12.3}$. (Co-5, K-1)

****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipatnam)

Course Code				**23STMAP233**			
Title of the Course				**Statistical Methods LAB**			
Offered to: (Programme/s)				**B.Sc.(Honors) - Statistics**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basic Computers**			

Course Description:

This course focuses on the application of statistical techniques to analyze the relationship between two variables. Through hands-on exercises and data analysis, students will develop a strong foundation in bivariate analysis and regression techniques, preparing them for applications in various fields such as business, economics, and social sciences. It combines theoretical concepts with practical implementation using Microsoft Excel and SPSS.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Develop proficiency in using MS Excel to generate bivariate random variable data.
2	Apply the least squares method to fit various curves (linear, quadratic, exponential, power) to data.
3	Calculate and interpret correlation coefficients (Pearson's and Spearman's) to measure the strength and direction of relationships between variables.
4	Compute regression lines for both ungrouped and grouped data to model the relationship between variables.
5	Determine multiple and partial correlation coefficients to assess the complex relationships among multiple variables.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Create bivariate random variable tables using MS Excel and explore the relationship between the variables.	K6	1	1
CO2	Apply curve fitting techniques to fit different types of curves (linear, quadratic, exponential, power) to data using the least squares method.	K3	7	2
CO3	Compute Pearson's and Spearman's correlation coefficients to measure the strength and direction of relationships between variables.	K5	7	1
CO4	Calculate regression lines for both grouped and ungrouped data and interpret the results.	K4	7	2
CO5	Calculate and interpret multiple and partial correlation coefficients to understand complex relationships among variables.	K4	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1	**3**							**2**	
CO2							**3**		**3**
CO3							**2**	**3**	
CO4							**3**		**2**
CO5							**3**	**2**	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

List of Practical's

1. Create the Bivariate random variable table with random number generation method in MS Excel.
2. Fitting of the Straight line using principle of least squares method.
3. Fitting of the Second degree parabola using principle of least squares method.
4. Fitting of the Exponential curve using principle of least squares method.
5. Fitting of the Power curve using principle of least squares method.
6. Calculate the Karl Persons Correlation Coefficient between the two random variables
7. Calculate the Spearman's Rank Correlation Coefficient (With tie and Without tie ranks)
8. Calculate the Correlation Coefficient between the Bivariate frequency data of two random variables.
9. Computation of Regression lines for ungrouped data.
10. Computation of Regression lines for grouped data.
11. Computation of Multiple and Partial Correlation Coefficients.

Text Book

1. Fundamentals of Mathematical Statistics, 12th Edition, 10th September 2020, S. C. Gupta and V. K. Kapoor, Sultan Chand & Sons, New Delhi.
2. Bernd Held., 2016, Microsoft Excel Functions & Formulas, Third Edition, Mercury Learning & Information.

Recommended References books:

1. Probability and Statistics,Volume I, D.Biswas, New central book Agency (P) Ltd, New Delhi.
2. An outline of Statistical theory, Volume Two,3rd Edition,2010(with corrections) A.M.Goon,M.K. Gupta, B.Dasgupta ,The World Press Pvt.Ltd., Kolakota.
3. Sanjay Arora and BansiLal:. New Mathematical Statistics, SatyaPrakashan , New Delhi.
4. https://www.tutorialspoint.com/excel_data_analysis/excel_data_analysis_tutorial.pdf

Suggested Co-curricular Activities:

1. Training of students by related industrial experts
2. Assignments including technical assignments if any.
3. Seminars, Group Discussions, Quiz, Debate set on related topics.
4. Preparation of audio and videos on tools of diagrammatic and graphical representations.
5. Collection of material / figures / photos / author photos of related topics.
6. Invited lectures and presentations of stalwarts to those topics.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipantam)

Course Code				**23STMAL234**			
Title of the Course				**Inferential Statistics**			
Offered to: (Programme/s)				**B.Sc.(Honors) - Statistics**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121, 23STMAL122**			

Course Description:

Inferential statistics is a branch of statistics concerned with drawing conclusions about a population based on information obtained from a sample. This course will equip students with the tools and knowledge to analyze data, make informed decisions, and interpret results in various fields

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Master techniques for estimating population parameters using confidence intervals.
2	Conduct hypothesis tests to assess claims about population characteristics.
3	Apply inferential statistical methods to solve practical problems.
4	Develop proficiency in using statistical software for data analysis.
5	Foster critical thinking and problem-solving skills related to data analysis

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	acquaint with estimator, estimates, estimation techniques and its properties	K1	5	1
CO2	acquire knowledge of testing the hypothesis of different distributions	K2	5	1
CO3	apply the large sample techniques for data analysis	K3	6	2
CO4	Apply the small sample techniques for data analysis.	K3	6	2
CO5	apply the non-parametric techniques for data analysis	K3	6	2

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PSO1	PSO2
CO1					3			3	
CO2					3			3	
CO3						3			3
CO4						3			3
CO5						3			3

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure:

Unit – 1: Theory of Estimation **(12 hours)**

Definitions – Parameter, Sample, Estimation concept and its types, criteria of a good estimator – unbiasedness, consistency, efficiency, & sufficiency. Problems on unbiasedness and consistency, Statement of Neyman's factorization theorem and problems on sufficiency.

Estimation of parameters by the method of moments and maximum likelihood (M.L), properties of MLE's. Rao – Cramer Inequality, properties. Binomial, Poisson & Normal Population parameters estimate by MLE method. Confidence Intervals- parameters of normal distribution

Applications:

1. Estimating economic indicators like GDP, inflation, and unemployment rates.
2. Estimating consumer preferences and demand for products
3. Estimating public opinion on various issues
4. Estimating student performance and program effectiveness.

Activity 1:

a. Collect datasets on heights, weights, test scores.
b. Calculate the sample mean as an estimate of the population mean.
c. Discuss the concept of margin of error and confidence intervals.

Activity 2 :

a. Collect real-world dataset (e.g., census data, economic data).
b. Estimate population parameters (e.g., mean income, population size) using appropriate statistical methods.
c. Discuss the implications of the results in the real world

Unit–2: Testing of Hypothesis **(12 hours)**

Concepts of statistical hypotheses, null and alternative hypothesis, critical region, two types of errors, level of significance and power of a test. One and two tailed tests. P value, Neyman-Pearson's lemma. Examples in case of Binomial, Poisson, Exponential and Normal distributions.

Applications

1. Hypothesis testing is a powerful statistical tool used across various fields to make informed decisions based on data. Some Areas applications are

- Medicine and Healthcare
- Business and Economics
- Social Sciences
- Engineering and Science

2. Is a new drug more effective than an existing treatment for a particular disease?
3. Does a new marketing campaign increase sales compared to the previous one?
4. Is there a relationship between smoking and lung cancer?
5. Does a new manufacturing process produce products with higher quality?

Activity

Formulating Hypotheses

- Stating the null hypothesis (H_0).
- Stating the alternative hypothesis (H_1)
- Setting the Significance Level (α)
- Determining the Critical Value or P-value
- **Finding the critical value**.
- **Calculating the p-value**.

Unit–3: Large sample Tests **(12 hours)**

Large sample tests - Single mean and difference of two means, confidence Intervals for mean(s)

Large sample tests - Single proportion and difference of proportions
Large sample test - difference of standard deviations
Large sample test - Correlation coefficients**.**
Problems**.**

Applications

Large sample tests, typically employing the Z-test, are widely used in various fields due to their simplicity and robustness. Some application areas are

1. Business and Economics
2. Social Sciences
3. Healthcare
4. Engineering
5. Agriculture
6. Environmental science

Activity 1:

Student satisfaction: Conduct a survey among a large sample of students to assess satisfaction with various campus services, facilities, or academic programs.

Student behavior : Investigate habits like study patterns, social media usage, or **dietary** preferences among a representative student sample

Activity 2:

Generate random data: Create simulated datasets with different
Characteristics to experiment with different test scenarios.

Apply large sample tests: Practice conducting hypothesis tests on the simulated data.

Unit – 4: Small Sample tests **(12 hours)**

Student's t-test for single mean, difference of means, paired t-test and correlation coefficient. χ^2 test for goodness of fit, independence of attributes and significance of population variance. F-test for equality of variances. Problems

Applications

Small sample tests are essential when dealing with limited data. They find applications in various fields where large datasets are difficult or impossible to obtain.

1. Research and Development
2. Quality Control
3. Medical and Healthcare
4. Social Sciences

Activity 1 :

Psychology experiments: Conduct small-scale experiments on perception, memory, or cognitive abilities

Activity 2 :

Study habits: Investigate if there's an association between study habits (library, dorm, group study) and exam scores

Activity 3

Generating random data: Create simulated datasets with different variances to practice calculating F-statistics and making decisions

Understanding F-distribution: Explore the shape of the F-distribution for different degrees of freedom

Unit–5:Non-parametric tests **(12 hours)**

Advantages and disadvantages NP tests, comparison with parametric tests. One sample - Run test, Sign test and Wilcoxon – signed rank test.
Two independent sample tests - Median test, Wilcoxon – Mann –
Whitney U test, Wald Wolfowitz's runs test**.** Problems

Applications

These are statistical methods that don't rely on specific assumptions about the population distribution. They are used when parametric tests (like t-tests, ANOVA) cannot be applied due to violations of assumptions or data characteristics. Widely used fields are

1. Social sciences
2. Medicine
3. Biology

4. Economics
5. Education

Activity 1:

Compare satisfaction levels between different academic programs using the Mann-Whitney U test.

Activity 2:

Analyze if there's a difference in social media usage (measured by time spent) between genders using the Wilcoxon rank-sum test.

Activity 3:

Compare median satisfaction levels between different academic programs (e.g., science, humanities, arts)

TEXTBOOKS:

1. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi
2. A.M.Goon,M.K. Gupta, B.Dasgupta, An outline of Statistical theory, The World Press Pvt.Ltd., Kolakota

REFERENCES

1. K.M.Ramachandran and Chris P.Tsokos, Mathematical Statistics with Applications, Academic Press(Elsevier), Haryana
2. D. Biswas, Probability and Statistics, New central book Agency (P) Ltd, NewDelhi

*****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipatnam)

Course Code		**23STMAL234**	
Title of the Course		**Inferential Statistics**	
Offered to: (Programme/s)		**B.Sc.(Honors) - Statistics**	
Year of Introduction:	**2024-25**	**Semester:**	**3**

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a.Show that sample mean is an unbiased and consistent estimator of population mean. (CO-1,L-2)

(OR)

b. Define the terms i) Parameter ii) Sample iii) Estimation concept and its types (CO-1,K-1)

2. a. Define the following terms: (CO-2,K-1)

(i) Null hypothesis (ii) Alternative hypothesis (iii) critical region. (CO-2,K-1)

(OR)

b. Explain Type I and Type II errors. (CO-2,K-2)

3. a. Write the procedure for single mean in large sample tests. (CO-3,K-6)

(OR)

b. Write the procedure for difference of proportions in large sample tests (CO-3,K-6)

4. a. Write the procedure of F-test for equality of population variances. (CO-4,K-6)

(OR)

b. Write the procedure of χ^2-test for goodness of fit (CO-4,K-6)

5. a. Explain the procedure of Sign test for one sample. (CO-5,K-2)

(OR)

b. Explain the procedure of Median test for two independent samples. (CO-5,K-2)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. Explain the characteristics of a good estimator (CO-1,K-2)

(OR)

b. Find Maximum likelihood estimator for μ and σ^2 in normal population. (CO-1,K-1)

7. a. State and prove Neyman-Pearson's lemma (CO-2,K-1)

(OR)

b. If $x \geq 1$ is the critical region for testing $H_0: \theta=2$ vs $H_1: \theta=1$ on the basis of the single Observation from an exponential distribution with probability density function $f(x, \theta)= \theta e^{-\theta x}$. Obtain the value of Type I and Type II errors. (CO-2,K-1)

8. a. In a Survey of buying habits, 400 women shoppers are chosen at random on supermarket A' located in a certain section of the city. Their average weekly food expenditure is Rs.250 with a S.D. of Rs. 40. For 400 women shoppers are chosen at random on Supermarket 'B' in another section of the city, the average weekly food expenditure is Rs.220 with a S.D. of Rs 55.Test at 1% level of significance whether the average week food expenditure of the populations of shoppers are equal. (CO-3,K-3)

(OR)

b. Explain the procedure for difference of standard deviations. (CO-3,K-2)

9. a. Out of 8,000 graduates in a town 800 are females, out of 1,600 graduate employees 120 are females. Use χ^2 to determine if any distinction is made in appointment the basis of sex. (CO-4,K-3)

(OR)

b.Explain the procedure for paired t-test (CO-4,K-2)

10. a. Explain the procedure of Wald-Wolfowitz run test for two samples. (CO-5,K-2)

(OR)

b. Explain the procedure of Wilcoxon signed rank test for one samples (CO-5,K-2)

***

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipatnam)

Course Code				**23STMAP234**			
Title of the Course				**Inferential Statistics**			
Offered to: (Programme/s)				**B.Sc.(Honors) - Statistics**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121, 23STMAL122**			

Course Description:

An inferential statistics lab course is designed to provide hands-on experience in applying statistical methods to real-world data. While the theoretical concepts are typically covered in a lecture-based course, the lab focuses on the practical implementation of these concepts using statistical software.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	**Develop proficiency in statistical software:** Students will become proficient in using statistical software packages (like SPSS, R, Python, or Excel) to perform complex statistical analyses
2	**Apply statistical concepts to real data:** Students will analyze real-world datasets to draw meaningful conclusions and make informed decisions
3	**Understand the assumptions and limitations of statistical tests:** Students will learn to critically evaluate the appropriateness of different statistical tests for given data and research questions
4	**Develop critical thinking and problem-solving skills:** Students will learn to interpret statistical results, identify potential issues, and communicate findings effectively.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	apply statistical analysis that can test hypotheses under parametric approaches	K3	6	1
CO2	apply statistical analysis that can test hypotheses under non-parametric approaches.	K3	6	1
CO3	draw the inferences for various non – parametric methods for Two samples using excel	K3	6	2
CO4	draw the inferences of various large samples using Excel.	K3	6	2
CO5	draw the inferences of various small samples	K3	6	2

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1						**3**		**3**	
CO2						**3**		**3**	
CO3						**3**			3
CO4						**3**			3
CO5						**3**			3

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

Practical 1

Large sample test for single mean & difference of means

Applications

- **Quality control:** Testing if the average weight of a product meets the specified standard
- **Manufacturing:** Checking if the mean production rate of a machine is as expected
- **Healthcare:** Evaluating if a new treatment improves the average patient outcome compared to a known standard
- **Market research:** Comparing the average satisfaction levels of two product versions.
- **Education:** Assessing if there's a significant difference in test scores between two teaching methods

Activity - 1

Real-world data collection

- Collect data on a variable (e.g., height, weight, income) from a large Sample (n > 30).
- Calculate the sample mean and standard deviation.
- Formulate a hypothesis about the population mean.
- Conduct a Z-test to test the hypothesis.
- Interpret the results in context

Activity – 2

Comparative data collection

- Collect data on two groups (e.g., males vs. females, treatment group Vs. control group) for a variable of interest.
- Calculate sample means and standard deviations for both groups.
- Formulate a hypothesis about the difference in population means.
- Conduct a Z-test to test the hypothesis.
- Interpret the results in context

Practical 2

Large sample test for single proportion & difference of proportions

Applications

- **Market research:** Testing if the proportion of customers who prefer a new product exceeds a certain threshold
- **Medical research:** Evaluating the effectiveness of a vaccine by comparing the proportion of vaccinated individuals who contract a disease to a historical control group
- **Advertising:** Assessing the effectiveness of two different advertising campaigns by comparing the proportion of customers who make a purchase
- **Social sciences:** Comparing the proportion of individuals with a specific characteristic in two different groups (e.g., gender, age, education)

Activity 1

Real-world data collection:

- Collect data on a categorical variable (e.g., success/failure, yes/no) from a large sample (n > 30).
- Calculate the sample proportion.
- Formulate a hypothesis about the population proportion.
- Conduct a Z-test to test the hypothesis.

• Interpret the results in context

Activity 2

Comparative data collection:

- Collect data on a categorical variable for two groups (e.g., males vs. females, treatment group vs. control group).
- Calculate sample proportions for both groups.
- Formulate a hypothesis about the difference in population proportions.
- Conduct a Z-test to test the hypothesis.
- Interpret the results in context

Practical 3

Large sample test for difference of standard deviations

Applications

➢ **Process Variability:** Comparing the variability of two production processes can help identify which process is more consistent.

➢ **Data Reliability:** In experimental studies, comparing the variability of measurements can assess the reliability of the data collection method

Activity

Data Collection and Visualization:

- Collect data on two different groups (e.g., heights of male and female students).
- Calculate the mean and standard deviation for each group.
- Visualize the data using box plots or histograms to compare the spread.
- Discuss the implications of a larger or smaller standard deviation in this context.

Practical 4

Large sample test for correlation coefficient

Applications

➢ **Education:** Investigating the relationship between student test scores and class size.

➢ **Psychology:** Studying the correlation between IQ and income level.

➢ **Science:** Investigating the correlation between temperature and atmospheric pressure.

Activity

Real-world Data Analysis

- **Data collection:** Collect data on two variables from a large sample (e.g., height and weight of students, hours studied and exam scores).
- **Data cleaning:** Ensure data is clean and free from outliers.
- **Calculate correlation coefficient:** Use statistical software to calculate the correlation coefficient (r).
- **Hypothesis testing:**

1. Set up null and alternative hypotheses
2. Calculate the test statistic (t-value).
3. Determine the p-value.
4. Make a decision based on the p-value (reject or fail to reject H0).

- **Interpretation:** Explain the meaning of the results in context

Practical 5

Small sample test for single mean & difference of means

Applications

➢ **Quality control:** Testing the mean weight of a small batch of products

➢ **Psychology:** Studying the cognitive abilities of a small sample of individuals

➢ **Agriculture:** Comparing the yield of two new crop varieties on a small experimental plot.

➢ **Education:** Comparing the test scores of two small groups of students using different teaching methods.

Activity 1

Simulation-based activities

- Use statistical software to generate random samples from a normal distribution with a known mean.
- Calculate sample means and conduct t-tests for different sample sizes.
- Investigate the impact of sample size on the power of the test.
- Explore Type I and Type II errors through simulations

Activity 2

Comparative data collection

- Collect data on a variable for two small groups (e.g., treatment group vs. control group, male vs. female).
- Calculate sample means and standard deviations for both groups.
- Formulate a hypothesis about the difference in population means.
- Conduct a t-test for independent samples to test the hypothesis.
- Interpret the results in context.

Practical 6

Small sample test for correlation coefficient

Applications

➢ **Medical Research:** Investigating the correlation between a new drug's efficacy and patient age in a small clinical trial.

➢ **Psychology:** Studying the relationship between intelligence quotient (IQ) and creativity in a small group of individuals

Activity

Simulation Activity

- **Generate random data:** Use statistical software to generate pairs of random data with different correlation coefficients (e.g., strong positive, weak negative, no correlation).
- **Calculate correlation coefficients:** Calculate the correlation coefficient for each dataset.
- **Hypothesis testing:** Conduct hypothesis tests for each correlation coefficient.
- **Discussion:** Discuss the impact of sample size and the true population correlation on the results.

Practical 7

Paired t-test (paired samples).

Applications

Before-and-after designs:

➢ Evaluating the effectiveness of a treatment or intervention.

Examples: Weight loss programs, smoking cessation,
Educational programs.

Repeated measures designs:

➢ Analysing data collected over time from the same individuals.

Examples: Tracking blood pressure changes over time, monitoring patient symptoms before, during, and after treatment

Activity 1

Simulation-based Learning

- **Generate paired data:** Use statistical software to generate paired data with a specified mean difference.
- **Conduct paired t-tests:** Perform paired t-tests on the generated data for different sample sizes.
- **Explore power:** Investigate the impact of sample size and effect size on the power of the paired t-test.
- **Type I and Type II errors:** Discuss the concepts of Type I and Type II errors in the context of paired t-tests

Activity 2

Real-world Data Collection and Analysis

- **Collect paired data:** Students can collect paired data on themselves (e.g., heart rate before and after exercise, weight before and after a diet).
- **Data analysis:** Calculate the difference scores, mean difference, and standard deviation of the differences.
- **Hypothesis testing:** Formulate null and alternative hypotheses, calculate the t-statistic, and determine the p-value.
- **Interpretation:** Draw conclusions about the effectiveness of the intervention or change.

Practical 8

Small sample test for single variance (χ^2 test)

Applications

- **Quality control:** Testing if the variance of a manufacturing process meets specified quality standards.
- **Finance:** Assessing the volatility (variance) of a financial asset compared to a benchmark.
- **Engineering:** Evaluating the consistency of a measurement process.
- **Scientific research:** Testing if the variability of experimental data aligns with theoretical expectations

Activity

Real-world Data Collection and Analysis

- **Collect data:** Students can collect data on a variable from a small sample (e.g., reaction times, measurement errors).
- **Calculate variance:** Calculate the sample variance.
- **Hypothesis testing:** Formulate null and alternative hypotheses about the population variance.
- **Chi-square test:** Conduct a chi-square test for a single variance.
- **Interpretation:** Draw conclusions about the population variance based on the test results.

Practical 9

Small sample test for difference of variances (F test)

Applications

- Assessing the consistency of two different measurement methods.
- Comparing the variability of two manufacturing processes.
- Evaluating the reliability of two different test instruments

Activity 1

Real-world Data Collection and Analysis

- **Collect data:** Students can collect two sets of data from different groups (e.g., reaction times for two groups of people, measurement errors from two instruments).
- **Calculate variances:** Calculate the sample variances for both groups.
- **Hypothesis testing:** Formulate null and alternative hypotheses about the population variances.
- **F-test:** Conduct an F-test to compare the variances.
- **Interpretation:** Draw conclusions about the equality of population variances based on the test results.

Activity 2

Case Studies

- **Provide real-world case studies:** Present case studies involving F-tests for comparing variances (e.g., manufacturing processes, measurement instruments).
- **Data analysis:** Provide students with sample data and ask them to conduct an F-test.
- **Interpretation:** Discuss the implications of the results in the context of the case study.

Practical 10

χ^2 test for goodness of fit and independence of attributes

Applications

- **Market Research:** To determine if product preferences match historical data.
- **Quality Control:** To check if production output conforms to expected distribution.
- **Genetics:** To test if observed genetic ratios match expected Mendelian ratios.
- **Dice Rolling:** To verify if a dice is fair.

Activity 1

Product Preference Survey

- Conduct a survey asking people about their preferred brand of a product.
- Compare the observed frequencies with the expected frequencies based on market share data.
- Conduct a chi-square goodness of fit test to determine if the preferences match the market share.

Activity 2

Gender and Major

- Collect data on the gender and major of a group of students.
- Create a contingency table to display the observed frequencies.
- Calculate the expected frequencies under the assumption of independence.
- Conduct a chi-square test of independence to determine if there is a relationship between gender and major.

Activity 3

Smoking and Lung Cancer

- Collect data on smoking habits and lung cancer incidence.
- Create a contingency table to display the observed frequencies.
- Calculate the expected frequencies under the assumption of independence.
- Conduct a chi-square test of independence to determine if there is a relationship between smoking and lung cancer.

Practical 11

Nonparametric tests for single sample (run test, sign test and Wilcoxon Signed rank test)

Applications

Run Test

 Detecting **randomness:**

 Quality **control:** To check if a production process Is in control.

 Financial **analysis:** To analyse the randomness of Stock price movements.

Sign Test

 Comparing **median to a value:** Used to test if the Median of a sample is different from a specified Value.

 Before**-after comparisons:** To analyse changes in paired data when the magnitude of change is not important.

 Quality **control:** To assess if a process has shifted From a target value

Wilcoxon Signed-Rank Test

 Comparing **paired data:** Used to compare paired Observations when the data is not normally distributed.

 Before**-after designs:** To analyse changes in paired Data when the magnitude of change is important.

 Clinical **trials:** To compare treatment effects Between paired samples

Activity 1 Run Test

Coin Flipping Experiment: Toss a coin multiple times and record

The outcomes as heads (H) or tails (T). Count the number of runs (Consecutive sequences of the same outcome). Test if the sequence is Random

Activity 2 **Sign Test**

Medication Effectiveness: Collect data on patients' pain levels Before and after taking a new medication. Convert the differences to Plus or minus signs based on improvement or worsening. Test if the Median difference is different from zero.

Activity 3 **Wilcoxon Signed-Rank Test**

Weight Loss Program: Measure participants' weights before and After a weight loss program. Calculate the differences and rank their Absolute values. Test if the median weight loss is different from zero

Practical 12

Nonparametric tests for two independent samples (Median test, Wilcoxon –Mann- Whitney - U test, Wald - Wolfowitz's runs test)

Applications

Median Test

Comparing groups based on median: Determines if Two or more groups differ significantly in terms of their Medians.

Example: Comparing the median income of Different educational levels

Wilcoxon-Mann-Whitney U Test

Comparing two independent groups: Determines if There's a difference in distribution between two Independent groups.

Example: Comparing the pain levels of two Treatment groups

Wald-Wolfowitz Runs Test

Testing for randomness: Determines if a sequence of Data is random.

Example: Analyzing the pattern of stock price increases and decreases

Activity 1 **Median Test**

Plant Growth Experiment: Measure the heights of plants Under two different growing conditions. Calculate the median height for each group. Conduct a median test to compare plant growth

Activity 2 **Wilcoxon-Mann-Whitney U Test**

Drug Effectiveness: Collect data on pain levels for two Groups of patients, one receiving a placebo and the other a New drug. Conduct a Wilcoxon-Mann-Whitney U test to Compare pain levels between the groups.

Activity 3 **Wald-Wolfowitz Runs Test**

Quality Control: Inspect a sequence of products for defects. Convert the data to a sequence of defective (D) and non-defective (N) items. Conduct a Wald-Wolfowitz runs test to check for randomness in the defect pattern.

*****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMIL231**			
Title of the Course				**Random Variables & Mathematical Expectations**			
Offered to: (Programme/s)				**B.Sc.(Honors) – Data Science**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**Minor**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basic Mathematics and Probability**			

Course Description:

This course helps the students to familiarize with the ways in which we talk about uncertainty and estimate their situations in which probability arises. Also this course aims at providing basic knowledge about mathematical expectations & generating functions.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Develop a comprehensive understanding of univariate random variables, including their probability distributions, functions, and characteristics, to model and analyze real-world phenomena.
2	Master the concept of mathematical expectation and its properties, enabling the calculation of mean, variance, and other statistical measures for effective decision-making.
3	Acquire proficiency in generating functions, their applications in probability theory, and their role in deriving probability distributions and moments.
4	Explore the Law of Large Numbers and its implications, understanding the convergence of sample means to population means.
5	Grasp the Central Limit Theorem and its applications, recognizing the normal approximation of sample means and its importance.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Understand the concept of discrete and continuous random variables with the application of random variables in real time problems.	K2	2	1
CO2	Learn the variance and covariance of random variables in terms of expectation	K1	7	1
CO3	Understand the definitions of various generating functions and learns their applications.	K2	2	1
CO4	Learn how to apply the concepts of Weak Law of Large Numbers (WLLN) and Strong Law of Large Numbers (SLLN)	K1	7	1
CO5	Applying the concept of Central limit theorem in real life examples and various inequalities in expectation	K3	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1		**2**						**1**	
CO2							**3**	**2**	
CO3		**2**						**1**	
CO4							**3**	**3**	
CO5							**3**	**3**	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure:

Unit – 1 **Univarate Random Variables (12 Hours)**

Definition, Discrete and Continuous random variables -Probability mass function and Probability density function with illustrations. Distribution function and its properties. Calculation of moments, coefficient of skewness and kurtosis for a given probability mass function (PMF) and Probability Density function (PDF)

Applications:

6. A random experiment of tossing a coin can be modeled using a random variable that represents the outcome of getting a head.
7. Rolling a die - Number of dots on the upper face
8. Drawing a card from a standard deck - Face value of the card (Ace, 2, 3, ..., King)
9. Quality control - Number of defective items in a sample.
10. Recording the time taken to complete a task - Time in seconds or minutes.
11. Temperature in Celsius or Fahrenheit.

Activity 1: Simulation

e. Introduce random number generators and simulation tools.
f. Provide a random variable scenario (e.g., flipping a coin, rolling a die, generating random numbers).
g. Have students use technology to simulate the experiment and collect data.
h. Analyze the results and compare them to theoretical probabilities.

Activity 2: Probability Distributions

e. Introduce different types of probability distributions (discrete and continuous).
f. Provide examples of random variables.
g. Divide students into groups.
h. Assign each group a random variable.
i. Ask them to determine the possible values and their corresponding probabilities.
j. Create a probability distribution table or graph.
k. Discuss the characteristics of the distribution.

Unit – 2 Mathematical Expectations (12 Hours)

Definition, Mathematical expectation of a random variable, Properties of expectations. Moments and covariance using mathematical expectation and their properties. Addition and Multiplication theorems on expectation of two and n variables. Variance, variance of a Linear combination of Random variables

Applications:

1. **Fair Games:** Used in gambling and game theory to determine the expected payoff or loss.
2. **Decision Making:** Helps in evaluating options by calculating expected values of different choices.
3. **Insurance:** Calculating premiums based on expected claims.
4. **Finance:** Estimating expected returns on investments.

Activity 1: Lottery Game Simulation

a. Create a simple lottery game with different prize values and corresponding probabilities.
b. Divide students into groups.
c. Each group should simulate playing the lottery multiple times, recording their winnings.
d. Calculate the average winnings per game (experimental expected value).

e. Compare the experimental expected value with the theoretical expected value calculated using the probability distribution.
f. Discuss the implications of the results in terms of fair games and decision-making.

Activity 2: Insurance Premium Calculation

a. Present a hypothetical insurance scenario (e.g., car insurance).
b. Provide data on the probability of different types of accidents and corresponding claim amounts.
c. Calculate the expected claim amount per policyholder.
d. Determine the insurance premium that would cover the expected claims and administrative costs.
e. Discuss factors that might affect the actual premium charged (e.g., profit margin, risk assessment).
f. Explore the concept of risk aversion and how it influences insurance purchasing decisions.

Unit – 3 Generating Functions (12 Hours)

Definitions of Moment Generating Function (M.G.F.), Cumulant Generating Function (C.G.F), Probability Generating Function (P.G.F), Characteristic Function (c.f.) and their properties with applications.

Applications:

1. Generating functions are used to solve counting problems, like finding the number of ways to partition a number or the number of solutions to a linear equation.
2. In the study of branching processes, PGFs are used to analyze the size of populations over time.
3. Generating functions can be seen as a type of integral transform, similar to Laplace or Fourier transforms, with applications in various areas of mathematics and engineering.

Activity 1: Generating Function Applications in Probability

a. Provide students with probability problems involving sums of independent random variables or finding probabilities of specific outcomes.
b. Guide students in using generating functions to solve these problems.
c. Emphasize the efficiency of using generating functions compared to traditional methods.
d. Discuss the limitations of generating functions and when they might not be applicable.

Activity 2: Generating Functions and Combinatorics

a. Introduce the concept of generating functions as a tool for solving counting problems.
b. Provide examples of counting problems (e.g., number of ways to distribute identical objects to distinct recipients, number of non-negative integer solutions to a linear equation).
c. Show how to represent the problem as a generating function.
d. Demonstrate how to extract the desired information from the generating function.
e. Explore different types of generating functions (ordinary, exponential) and their applications.

Unit – 4 Law of Large Numbers & Inequality (12 Hours)

Weak Law of Large Numbers (WLLN) and Strong Law of Large Numbers (SLLN- Statement only) for identically and independently distributed (i.i.d.) random variables with finite variance.
Markov's inequality (Statements only), Khinchin's Theorem for WLLN (Statements only).

Applications:

1. In manufacturing processes, it can be used to assess the long-run average quality of products based on sample data.
2. In finance, it helps in estimating the long-term average return of an investment based on historical data.
3. To estimate the average claim amount based on a large number of claims.

Activity 1: Coin Tossing Simulation

a. Divide students into groups.
b. Each group tosses a coin a specified number of times (e.g., 10, 20, 50, and 100).
c. Calculate the proportion of heads for each trial.
d. Plot the proportion of heads against the number of tosses.

e. Discuss how the proportion of heads converges to 0.5 (the theoretical probability of heads) as the number of tosses increases.
f. Relate the experiment to the WLLN.

Activity 2: Markov's Inequality and Gambling

a. Introduce Markov's inequality and its statement.
b. Present a gambling scenario (e.g., a game with a random payoff).
c. Calculate the expected value of the payoff.
d. Use Markov's inequality to estimate the probability of winning a large amount.
e. Discuss the limitations of Markov's inequality and the need for tighter bounds.

Unit – 5 Central Limit Theorems (CLT) (12 Hours)

Chebychev's and Cauchy - Schwartz inequalities and their applications. Central limit theorems, Statement of De-Movire's Laplace theorem, Lindberg – Levy CLT and its applications, Statement of Liapounoff's CLT, relationship between CLT and WLLN.

Applications:

1. The CLT can be used to monitor process means and detect shifts in the process.
2. The distribution of portfolio returns is often assumed to be normal, based on the CLT applied to individual asset returns.
3. Analyzing biological data, such as population sizes or genetic variations, often involves applying the CLT.
4. In Market Research understanding consumer behavior often involves analyzing large datasets, where the CLT is applicable.

Activity 1: Dice Rolling Simulation

a. Divide students into groups.
b. Each group rolls a set of dice multiple times (e.g., 5, 10, 20, 30 dice).
c. Calculate the average of the dice rolls for each trial.
d. Create a histogram of the sample means.
e. Observe the shape of the distribution as the number of dice increases.
f. Discuss how the distribution of sample means approaches a normal distribution, illustrating the CLT.

Activity 2: Simulation of Sampling Distributions

a. Simulate a population with a known distribution (e.g., normal, uniform, exponential).
b. Draw random samples of different sizes from the population.
c. Calculate the sample mean for each sample.
d. Create a histogram of the sample means.
e. Observe the shape of the distribution and compare it to a normal distribution.
f. Explore the impact of sample size on the shape of the sampling distribution.

Text Book

1. Fundamentals of Mathematical Statistics, 12th Edition, 10th September 2020, S. C. Gupta and V. K. Kapoor, Sultan Chand & Sons, New Delhi.

Recommended References books:

4. Probability and Statistics,Volume I, D.Biswas, New central book Agency (P) Ltd, New Delhi.
5. An outline of Statistical theory, Volume Two,3rd Edition,2010(with corrections) A.M.Goon,M.K. Gupta, B.Dasgupta ,The World Press Pvt.Ltd., Kolakota.
6. Sanjay Arora and BansiLal:. New Mathematical Statistics, SatyaPrakashan , New Delhi.

Websites of Interest: http://onlinestatbook.com/rvls/index.html

Learning Outcomes: At the end of the course student able to learn

1. Differentiate between discrete and continuous random variables, calculate probabilities, and identify appropriate probability distributions for given scenarios.
2. Calculate expected values, variances, and covariance's of random variables and understand their significance in statistical analysis.
3. Apply probability generating functions and moment generating functions to derive probabilities, moments, and distributions of random variables.
4. Understand the concept of convergence in probability and apply the Law of Large Numbers to estimate population parameters from sample data.
5. Apply the Central Limit Theorem to approximate probabilities and construct confidence intervals for sample means.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

23STMIL231: Random Variables & Mathematical Expectations

Max. Marks: 70 **Semester III** **Max. Time: 3Hrs**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a. Define the Random variables and state its types. (Co-1, K-1)

(OR)

b. Define distribution function and state its properties (Co-1, K-1)

2. a. Show that the mathematical expectation of the sum of two random variables is the sum of their individual expectation. (Co-2, K-1)

(OR)

b. State and prove multiplication theorem on Mathematical expectationof two events. (Co-2, K-1)

3. a. Define Probability Generating Function(PGF) and write its properties. (Co-3, K-1)

(OR)

b. Define Characteristic Function (CF) and write its properties. (Co-3, K-1)

4. a. Explain the concept of Weak law of large numbers (WLLN). (Co-4, K-2)

(OR)

b. Explain the concept of Strong Law of Large Numbers (SLLN). (Co-4, K-2)

5. a. State the Liapounoff'scentral limit theorem. (Co-5, K-1)

(OR)

b. State the Lindberg – Levy's theorem and its assumptions. (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. A random variable has the following probability distribution

x	0	1	2	3	4	5	6	7	8
P(X= x)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine 'a'

(ii) Find P(X<3), P(X ≥ 3) and P(0<X<5)

(iii) Find the distribution function of X. (Co-1, K-3)

(OR)

b. The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f. $f(x) = 6x(1-x),\ 0 \le x \le 1$.

i) Check that f(x) is p.d.f.,

ii) Determine a number **b** such that P(X <**b**) = P(X >**b**). (Co-1, K-3)

7. a. State and prove the Linear combination of Random variables (Co-2, K-5)

(OR)

b. Given the following table:

X = x	-3	-2	-1	0	1	2	3
P(X = x)	0.05	0.1	0.3	0	0.3	0.15	0.1

Compute (i) E(X), (ii) E(2X+3), (iii) V(X) and (iv) V(2X+3) (Co-2, K-3)

8. a. Derive the relation between cumulants in terms central moments. (Co-3, K-3)

(OR)

b. Prove that $\mu_r^1 = \left[\frac{d^r M_X(t)}{dt^r}\right]_{t=0}$ (Co-3, K-3)

9. a. Examine whether the weak law of large numbers holds for the sequence $\{X_k\}$ of independent random variable defined as follows: $P\left(X_k = \pm 2^k\right) = 2^{-(2k+1)}$, $P\left(X_k = 0\right) = 1 - 2^{-2k}$ (Co-4, K-3)

(OR)

c. Write the statements of W.L.L.N and S.L.L.N. for the sequence ofi.i.d. Random variables (Co-4, K-3)

10. a. Use chebychev's inequality to determine how many times a fair coin must be tossed in order that the probability will be atleast 0.90 that the ratio of the observed number of heads to the number of tosses will lie between 0.4 and 0.6. (Co-5, K-3)

(OR)

b. State and prove Cauchy - Schwartz inequalities (Co-5, K-3)

***

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University, Machilipanam)

Course Code				**23STMIL231**			
Title of the Course				**Statistical Data Analysis Using SPSS - I**			
Offered to: (Programme/s)				**B.Sc.(Honors) – Data Science**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**III**
Course Category:	**Minor**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basics Computers**			

Course Description:

This course is designed to introduce students to the fundamentals of descriptive statistics and their application using SPSS statistical software. Students will learn how to summarize and describe data using various statistical measures and graphical representations. The course emphasizes hands-on experience in data analysis and interpretation.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Understand the basic concepts of descriptive statistics, including measures of central tendency, dispersion, and shape.
2	Learn how to import and manage data in SPSS
3	Master the use of SPSS to calculate and interpret descriptive statistics.
4	Create effective graphical representations of data using SPSS.
5	Apply descriptive statistical techniques to analyze real-world datasets.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Understand the SPSS packages and describing a variety of statistical variables in SPSS	K2	2	1
CO2	Integrate and access the data base from different source of file format.	K1	7	1
CO3	Construct various charts and diagrams in SPSS	K3	7	1
CO4	Construct the frequency tables in SPSS	K3	7	1
CO5	Recommend the best statistical tool for basic statistical analysis.	K3	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1		**2**						**3**	
CO2							**2**	**3**	
CO3							**1**	**3**	
CO4							**1**	**3**	
CO5							**2**	**3**	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

List of practical's

1) Data entry, Import and Export the data sets from various formats to SPSS
2) Construction of frequency table for univariate categorical data using SPSS
3) a) Construction of pie chart for nominal data using SPSS
b) Construction of pie chart for ordinal data using SPSS
4) a) Construction of bar chart for nominal data using SPSS
b) Construction of bar chart for ordinal data using SPSS
5) Create grouped frequency table from given raw data (non categorical) using SPSS
6) Computation of descriptive statistics for continuous data using SPSS
7) Construction of histogram for continuous data using SPSS
8) Construction of Box plots using SPSS
9) Construction of Crosstabs using SPSS

Question Paper Pattern for Core Lab Courses

(A) Semester End Lab Examination

23STMIP231: Statistical Data Analysis Using SPSS - I

Max.Marks: 35 Max.Time: 3Hours Pass. Min : 14

I. **Answer the following.** **Max. Marks: 30**

Q1
Q2
Q3
Q4
Q5

II Viva **3 Marks**

III Record **2 Marks**

(B) CONTINUOUS ASSESMENT (Internal) **15 Marks**

TOTAL: (A)+(B) = **50 Marks**

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMIL234**			
Title of the Course				**Random variables and Probability Distributions**			
Offered to: (Programme/s)				**B.B.A(Honors) – B.A**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**4**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:				**Course Relates to:**		**GLOBAL**	
Year of Introduction:	**2024**			**Percentage:**		**NA**	
Type of the Course: Minor							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basics of statistics**			

Course Description:

This course explores random variables as numerical representations of outcomes from random experiments. It delves into various probability distributions to model different typesof random phenomena, enabling the calculation of probabilities and statistical measures.

Students will learn to apply these concepts to real-world problems involving decision-making under uncertainty.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	To understand the definitions of discrete, continuous, and joint random variables, compute the mean, variance and covariance of random variables, know the definition of mass (density) function and distribution function of a random variable and be able to find one from the other, and be able to find the marginal mass (density) function and distribution functions from the joint mass (density) function and distribution function.
2	To differentiate between large and small samples and apply apt testing procedures
3	Knowledge to conceptualize the probabilities of events including axiomatic approach. Simultaneously, they will learn the notion of conditional probability including the concept of Bayes' theorem and applications of normal, t, F and chi- square distributions.

Course Outcomes

At the end of the course, the student will / will be…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Acumen to apply Random Variable and expectation to data standard discrete probability distribution to different situations.	K1	1	1
CO2	Apply the concept of mathematical expectation to find the mean, variance, and standard deviation of a probability distribution	K2	7	1
CO3	knowledge of important discrete distributions such as Binomial, Poisson, Geometric distributions and relations with some other distributions	K3	1	1
CO4	knowledge of important continuous distributions such as Uniform, Normal, Exponential and Gamma and relations with some other distributions	K3	1	1
CO5	Testing the Qualitative and Quantitative factors in case of one and two samples using standard normal variate, student's t ,F-statistic and chisquare test statistic and Quantitative factors in case of more than twosamples using ANOVA	K3	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate;K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1	3							3	
CO2							3	3	
CO3	3							1	
CO4	3							1	
CO5							3	3	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

I. Course Structure:

Unit– 1: Random variable **(12 hours)**

Introduction, Distribution Function, Discrete Random Variable, Continuous Random Variable, Two-Dimensional Random Variables, Transformation of One-Dimensional Random Variable, Transformation of Two-Dimensional Random Variable, Transformation of n-Dimensional Random Variable.

Applications:

- Hypothesis testing for population mean
- Constructing confidence intervals for the population mean when the population
- Comparing the means of two related samples (e.g., before and after)
- Assessing the overall significance of the regression model.
- Testing if observed data fits a particular theoretical distribution
- To determining if there is a relationship between two categorical variables.
- Hypothesis testing for population variance

Activity 1

- Discuss examples of random variables in real life (e.g., number of cars passing a traffic lightin an hour, height of students in a class, temperature)
- Roll a fair die multiple times and record the outcomes. Calculate probabilities of differentevents (e.g., getting a number greater than 4).
- Toss a coin multiple times and record the results. Calculate probabilities of different outcomes Analyze real-world datasets using statistical concepts related to random variables.
- Number of customers arriving at a store in an hour)

Activity 2

- Analyze real-world datasets using random variables.
- Identify appropriate probability distributions for the data.
- Draw conclusions and make inferences.

Unit– 2: Mathematical expectations:

Mathematical Expectation or Expected Value of a Random Variable, Expected Value of Function of A Random Variable, Properties of Expectation, Properties of Variance, Covariance, Some Inequalities Involving Expectation, Moments of Bi-variate Probability Distributions, Conditional Expectation and Conditional Variance

Applications:

- Gambling and Games of Chance in determining the expected payoff of a game,helping to assess its fairness.
- Insurance companies calculate expected losses to determine appropriate premiumamounts.
- Investment and Finance expected returns are calculated for different investmentportfolios to aid in decision-making
- Quality Control Expected defect rates help determines the economic impact of quality issues.
- Decision Making: Expected net present value (NPV) is a key metric in projectappraisal.

Activity

- Calculating the expected winnings in a card game.
- Determining the fair price for a game of chance.
- Analyzing the expected return on an investment.

Unit–3: Discrete Distributions

Introduction, Discrete Uniform Distribution, Bernoulli Distribution, Binomial Distribution, Poisson distribution, Definition, p.m.f, properties, applications and simple Problems.

Applications:

- Binomial: Used for binary outcomes (success/failure) with fixed number of trials.
- Poisson: Models the number of events occurring in a fixed interval of time or space.

Activity:

- Probability calculations: Students can calculate probabilities of various events using different discrete distributions and interpret the results.
- Distribution fitting estimate parameters of different discrete distributions and assessgoodness-of-fit.

Unit–4: Continuous Distribution

Introduction, Normal Distribution, Rectangular (or Uniform) Distribution, Definition, d. f, properties,applications and simple problems.

Applications:

- Quality control: Normal distribution is used to set control limits for process monitoring.
- Reliability analysis: Exponential distribution models the time between failures of components.
- Physics: Many physical phenomena, such as particle velocities and measurement errors, follow normal distribution

Activity:

- Probability calculations: Students can calculate probabilities of various events using different discrete distributions and interpret the results.
- Distribution fitting estimate parameters of different discrete distributions and assessgoodness-of-fit.

Unit–5: **Exact Sampling Distributions**

Introduction, M.G.F. of Chi-Square Distribution, Applications of Chi-Square Distribution, Student's 't' Distribution, Applications of t-Distribution, F-Distribution, Applications of F-Distribution, Relation between t and F Distributions, Relation Between F and x2 Distributions

Applications:

- Hypothesis testing for population mean
- Construct confidence intervals for population mean.
- Regression analysis: Test the significance of the overall regression model and individualpredictors.
- Test if two populations have equal variances.
- Compare the distribution of a categorical variable across different populations
- Analyze if two categorical variables are independent (e.g., gender and smoking habits).

Activity 1:

- **Test of Independence:** Analyze data on customer preferences (e.g., age, gender, productchoice) to determine if product preference is related to demographic factors.
- Test of Homogeneity: Compare the distribution of a categorical variable (e.g., car color)across different regions to see if there are differences.
- **Goodness of Fit:** Students can collect data on coin tosses, dice rolls, or card draws to test ifthe results follow a uniform distribution.

Activity 2:

- One-Sample T-Test: Compare the mean height of students in a class to the national average.
- Independent Samples T-Test: Compare the mean test scores of two different teaching

methods.

- Paired T-Test: Analyze the effectiveness of a weight loss program by comparing participants'weights before and after.

Activity 3:

- A**NOVA:** Compare the mean sales of different product brands to determine if there is a significant difference.
- **Equality of Variances:** Test if the variability in test scores between two different schools is significantly different.
- **Regression Analysis:** Analyze the relationship between variables (e.g., advertising expenditure and sales) and test the significance of the overall model

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMIL234**			
Title of the Course				**Random variables and Probability Distributions**			
Offered to: (Programme/s)				**B.B.A(Honors) – B.A**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**4**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:				**Course Relates to:**		**GLOBAL**	
Year of Introduction:	**2024**			**Percentage:**		**NA**	
Type of the Course: Minor							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basics of statistics**			

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. a. Define the Random variables and state its types. (Co-1, K-1)

(OR)

b. Define distribution function and write its properties (Co-1,K-1)

2. a. State and prove multiplication theorem of expectation for two variables (Co-2, K-1)

(OR)

b. State and prove Cauchy-Schwartz Inequality (Co-2, K-1)

3. a. . Define Bernoulli distributions and its applications (Co-3, K-1)

(OR)

b. Define binomial distributions and its applications (Co-3,K-1)

4. a. Define rectangular distributions and derive its mean, variance. (Co-4, K-1)

(OR)

b. Define normal distribution and write its properties. (Co-4, K-1)

5. a. Define chi-square distributions and its applications. (Co-5, K-1)

(OR)

b. Define the F-Distribution and its applications. (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. A random variable has the following probability distribution

x	0	1	2	3	4	5	6	7	8
P(X= x)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine 'a' (ii) Find P(X<3), P(X ≥ 3) and P(0<X<5)

(iii) Find the distribution function of X. (Co-1, K-3)

(OR)

b. The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f. $f(x) = 6x(1-x),\ 0 \le x \le 1$.

i) Check that f(x) is p.d.f.,

ii) Determine a number **b** such that P(X <**b**) = P(X >**b**). (Co-1,K-3)

7. a. state and prove Chebychev's inequalities.

(ii) A fair die is tossed 720 times. Use Chebychev's inequalities to find a lower bound for the probability of getting 100 to 140 possibility of six points on the face of die. (Co-2, K-1)

(OR)

b. Explain moment generating function and cumulate generating functions for random variables. (Co-2, K-2)

8. a. Explain Poisson distributions and writs its properties. (Co-3, K-2)

(OR)

b. Fit a Binomial Distribution, by direct method for the following data: (Co-3, K-1)

X	0	1	2	3	4	5	6
f	7	64	140	210	132	75	12

9. a. Fit a normal distribution and obtain the expected frequencies using ordinary method.

C.I	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
f	3	21	150	335	326	135	26	4

(Co-4, K-1)

(OR)

b. A normal population is having the mean $\mu = 30$ and S.D $\sigma = 5$. If X is a variable of this population then find the probabilities of (i)$26 \leq X \leq 40$ (ii) $X \geq 45$ (Co-4, K-3)

10. a. Derive Student's t Distribution. (Co-5, K-3)

(OR)

b. Derive the relation between F and χ^2 distribution. (Co-5, K-3)

****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMIP234**			
Title of the Course				**Random variables and Probability Distributions lab**			
Offered to: (Programme/s)				**B.B.A(Honors) – B.A**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:				**Course Relates to:**		**GLOBAL**	
Year of Introduction:	**2024**			**Percentage:**		**NA**	
Type of the Course: Minor							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**Basics of statistics**			

Practical Credits:1 2hrs/week

Practical Syllabus

1. (a)Fitting of Binomial Distribution (Direct Method)
   (b)Fitting of Binomial Distribution (Recurrence Method)

Applications

- Determining the probability of a certain number of defective items in a productionbatch
- Predicting the number of customers who will purchase a new product.
- Assessing the probability of a stock price going up or down a certain number of times
- Calculating the likelihood of a specific number of insurance claims.

Activity:

- what is the probability of getting exactly 5 heads when flipping a coin 10 times?
- A manufacturing process produces 5% defective items. What is the probability of finding 2defective items in a sample of 20?
- A basketball player has a free-throw shooting percentage of 80%. What is the probability ofmaking exactly 6 out of 8 free throws?

2. (a) Fitting of Poisson Distribution (Direct Method)
(b) Fitting of Poisson distribution (Recurrence Method)

Applications

* Telecommunications: Call arrivals, Network traffic
* Insurance: Accident claims Catastrophic events:

*Healthcare: Patient arrivals, Disease outbreaks

Activity:

- Predicting the number of calls per minute.
- Modeling the number of cars arriving at an intersection
- Analyzing customer arrival rates.
- Determining the probability of finding a certain number of defects in a unit.
- Estimating the number of bacteria in a food sample.

3. (a) Fitting of Normal Distribution (Areas Method)
(b) Fitting of Normal Distribution (Ordinates Method)

Applications:

- Monitoring product dimensions, weight, or other characteristics to ensure they fallwithin specified tolerances

- Pricing financial derivatives, valuing assets, and assessing investment risks.
- Modeling physical quantities like height, weight, or IQ scores, as well as naturalprocesses like temperature and rainfall.
- Making inferences about population parameters based on sample data, such asconducting hypothesis tests about means and proportions.

Activity:

- Quality Control: Identifying Defects, Process Optimization
- Finance and Economics: Stock Prices, Economic Indicators
- Natural Sciences: Physical Measurements, error Analysis
- Social Sciences: Survey Data, Intelligence Quotient (IQ):
- F-test for equality of population variances.

Applications:

- Determine if two manufacturing processes produce products with the samelevel of consistency.
- Verify the assumption of equal variances before conducting a t-test or ANOVA.
- Evaluate the homogeneity of experimental groups in terms of variability.
- Compare the volatility of two investment portfolios

Activity

- Collect samples from both processes, calculate the sample variances, andconduct an F-test to compare the population variances.
- Verify the assumption of equal variances before conducting a t-test orANOVA.
- Calculate the variances of the groups and conduct an F-test to determine if thegroups have similar variability, which is often a desirable condition for experimental designs.
- Calculate the sample variances of the returns for both portfolios and conductan F-test to determine if there is a significant difference in volatility.

4. Chi-Square test for goodness of fit

Applications:

- Determine if a dataset follows a specific probability distribution (e.g., normal,Poisson, binomial).
- Compare observed frequencies of categories to expected frequencies based ona theoretical or hypothesized distribution.
- Check if a sequence of data is truly random.
- Analyze if product preferences or customer demographics align with expectedmarket proportions.

Activity

- Collect data, calculate observed frequencies, compute expected frequencies based on the hypothesized distribution, and perform the chi-square goodness offit test.
- Collect categorical data, determine expected frequencies based on a specificproportion or distribution, and conduct the chi-square test to assess if the observed data fits the expected pattern.
- Convert the data into categories (e.g., even/odd, above/below median),calculate observed and expected frequencies under the assumption of randomness, and perform the chi-square test
- Collect data on product choices or demographic information, compare observed frequencies to expected values based on market research or industrystandards, and perform the chi-square test.

5. Chi-Square test for independence of attributes

Applications:

- Determine if there is an association between two categorical variables in a population.

- Analyze whether customer preferences for different product features are related todemographic factors like age, gender, or income.
- Investigate if there is a connection between a certain disease and specific risk factorslike smoking, diet, or exercise
- Examine the relationship between educational attainment and employment status, orbetween voting behavior and political affiliation.

Activity

- Collect data on the two categorical variables, create a contingency table, andconduct a chi-square test to assess the independence of the variables.
- Collect data on product preferences and demographic information, create acontingency table, and conduct a chi-square test to determine if there is a relationship between the variables.
- Collect data on disease status and risk factors, create a contingency table, andconduct a chi-square test to assess the independence of the variables
- Collect data on educational attainment and employment status (or voting behavior and political affiliation), create a contingency table, and conduct achi-square test to determine if there is a relationship between the variables.

6. Small sample test (t-test): One Sample.

Applications:

- Determine if a sample of products from a manufacturing process meets thespecified target value.
- Assess the effectiveness of a new drug or therapy by comparing the meanoutcome of a sample of patients to a known baseline value.
- Test a claim about the average income level of a specific population group.
- Monitor the quality of a production process by comparing the mean of asample of products to a specified standard.

Activity:

- Collect a sample of product measurements, calculate the sample mean, and conduct aone-sample t-test to compare it to the target value.
- Collect data on a sample of patients before and after treatment, calculate the mean difference, and conduct a one-sample t-test to determine if the treatment had a significanteffect.
- Collect a random sample of incomes from the population, calculate the sample mean,and conduct a one-sample t-test to determine if the sample mean supports or rejects the claim about the population mean.
- Collect a sample of products, calculate the sample mean, and conduct a one-sample t-test to determine if the process is in control or if there is evidence of a quality issue.

7. t – test Independent Sample and Paired sample

Independent Sample T-Test

Applications:

- Marketing: Comparing sales of two different product variants.
- Education: Comparing test scores of students from two different teaching methods.
- Healthcare: Comparing the effectiveness of two different treatments for a disease.
- Psychology: Comparing the reaction times of two different groups of participants.

Activity

- Data Collection: Collect data on two independent groups (e.g., male vs. femalestudents, experimental group vs. control group).
- Data Analysis: Calculate means and standard deviations for both groups.
- Hypothesis Testing: Formulate a null and alternative hypothesis.
- T-Test Calculation: Perform the t-test using statistical software or by hand.
- Interpretation: Interpret the results and draw conclusions about the difference betweenthe groups.

8. Paired Sample T-Test

Applications:

- Healthcare: Comparing blood pressure before and after medication, weight loss beforeand after a diet program.
- Education: Comparing student performance on a pre-test and post-test.
- Psychology: Comparing mood scores before and after a therapy session.
- Sports: Comparing athlete performance before and after a training program.

Activities:

- Data Collection: Collect paired data from the same individuals (e.g., pre-test and post-test scores, measurements before and after treatment).
- Data Analysis: Calculate the difference scores for each pair.
- Hypothesis Testing: Formulate a null and alternative hypothesis about the meandifference.
- T-Test Calculation: Perform the t-test using statistical software or by hand.
- Interpretation: Interpret the results and draw conclusions about the change in thepaired measurements.

Note: Training shall be on establishing formulae in Excel cells and derive the results. The excel outputshall be exported to MS word for writing inference.

II. References

1. S.C. Gupta, (2019), Seventh Edition, Fundamentals of Statistics, Mumbai: Himalaya Publishing House.
2. Sharma, J. K. (2013), *Business statistics*, New Delhi: Pearson Education.
3. Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2013), *Statisticsfor Business and Economics*, New Delhi: Cenga.ge Learning.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL236**			
Title of the Course				**Probability Foundation for AI**			
Offered to: (Programme/s)				**B.Sc.(Honors) -Artificial Intelligence**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any							

Course Description:

Inferential statistics is a branch of statistics concerned with drawing conclusions about a population based on information obtained from a sample. This course will equip students with the tools and knowledge to analyze data, make informed decisions, and interpret results in various fields

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Grasp the basic principles of probability, including sample spaces, events, probability axioms, and conditional probability.
2	Develop the ability to calculate probabilities of different events using various methods and techniques.
3	Learn about the concept of random variables, their types (discrete and continuous), and their probability distributions.
4	Understand the concepts of expected value, variance, and standard deviation and be able to calculate them for different distributions.
5	Study different types of probability distributions (binomial, Poisson, normal, etc.) and their applications.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Apply probability concepts to real-world problems	**K1**	**7**	**1**
CO2	Apply appropriate probability distributions to model real-world situations.	**K2**	**7**	**1**
CO3	**Apply expected values** to calculate various statistical measures like mean, variance, and standard deviation.	**K3**	**7**	**1**
CO4	**Understand the relationships** between different discrete distributions and their applications.	**K3**	**1**	**1**
CO5	**Apply continuous distributions** to model real-world phenomena and solve practical problems.	**K3**	**7**	**1**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1							**3**	**3**	
CO2							**3**	**3**	
CO3							**2**	**3**	
CO4	**2**							**2**	
CO5							**3**	**3**	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

II. Course Structure:

Unit– 1: Probability (12 hours)

Introduction, random experiments, sample space, events and algebra of events. Definitions of probability-classical, statistical, And axiomatic. Conditional Probability, Addition and Multiplication theorem of probability, independent events, Bayes' Theorem, and its applications.

Applications:

1. Predicting weather patterns based on historical data and statistical models.
2. Ensuring product reliability and consistency through statistical process control.
3. Clinical trials, drug development, and disease modeling involve probabilistic methods.
4. Opinion polls, election predictions, and demographic studies utilize probability.
5. Understanding odds, probabilities, and expected values in games of chance.

Activity:

a. Coin Flipping**:** A classic for exploring probability. Students can Predict outcomes, record results, and calculate Experimental probability.

b. Dice Rolling: Similar to coin flipping, but with more Outcomes. Students can explore different combinations and Probabilities.

c. **Card Games:** Explore probability concepts through card games Like drawing specific cards or combinations.

d. Data Analysis: Provide datasets and ask students to Calculate probabilities and draw conclusions.

Unit–2: Random variables (12 hours)

Introduction, discrete and continuous, illustrations and properties of Random variables, probability mass function, probability density function and cumulative distribution functions and its properties. Two Dimensional random variables: Joint, marginal and conditional probability mass function and probability density function, independence of random variables. Transformation of one dimensional random variable.

Applications

1. Risk assessment: Estimating the probability of financial losses.
2. Portfolio management: Evaluating investment returns and risks.
3. Option pricing: Determining fair values of financial derivatives.
4. Machine learning: Building probabilistic models for

Classification and regression.
5. Computer science: Network analysis, data mining, and Algorithm design.
6. Image and signal processing: Noise reduction and feature extraction.

Activity

1. Simulate rolling a die 100 times and count the frequency of each outcome.
2. Simulate bus arrival times for 10 buses and calculate average wait time.
3. Simulate customer purchases for 50 customers and calculate the Percentage buying each amount

Unit3: Mathematical Expectation (12 hours)

Introduction, variance and covariance of random variables and their properties,

Conditional expectations. Moment generating function and its properties, Cumulant generating function, characteristic function, and its properties. Chebyshev's inequality and its applications.

Applications

1. Expected value in portfolio management.
2. Expected value in insurance pricing.
3. Expected value in gambling. (e.g., roulette, blackjack)
4. Expected value in quality control.
5. Determining the expected number of defective items in a Production process.

Activity :

1. Students can calculate the expected value for each ticket based On its probability and prize.
2. They can then determine the expected total winnings based on The number of tickets purchased.
3. Introduce the concept of variance to measure the risk involved in buying Multiple tickets.
4. Discuss strategies for maximizing winnings based on expected value and probability.
5. Explore real-world lottery scenarios and compare them to the classroom game.
6. Discuss the concept of "house edge" and how it relates to lottery profits.

Unit–4: Discrete Probability Distributions (12 hours)

Uniform, Bernoulli, Binomial, Poisson, Geometric, Negative Binomial and Hyper-geometric distributions along with their characteristic properties, Applications and limiting/approximation cases.

Applications:

1. Bernoulli distribution**:**Modelling single trials with two outcomes (e.g., coin flips, success/failure).
2. Binomial distribution**:**Modelling the number of successes in a fixed number Of independent trials (e.g., number of heads in 10 coin flips).
3. Poisson distribution**:**Modelling the number of events occurring in a Fixedintervaloftime or space (e.g., number of customers arriving at a Store per hour).
4. Geometric distribution:Modelling the number of trials until the first Success (e.g., number of coin flips until the first head).

5.Negative binomial distribution:Modelling the number of failures before Aspecified number of successes (e.g., number of defective items before finding10 non- defective ones).

6.Hyper geometric distribution:Modeling sampling without replacement From a finite population (e.g., drawing cards from a deck Without replacement).

Activity :

Binomial Distribution Simulation:

- Simulate a binomial experiment (e.g., flipping a coin 10 times).
- Calculate the probability of different numbers of successes.
- Compare the simulated results with theoretical probabilities.

Poisson distribution Application:

- Collect data on the number of customers arriving at a store per hour.
- Fit a Poisson distribution to the data.
- Use the Poisson distribution to estimate the probability of different numbers of customers.

Geometric Distribution Problem:

- Solve word problems involving geometric distributions (e.g., finding the probability of getting the first head on the fourth coin toss).
- Calculate the expected number of trials until the first success.

Unit5: Continuous Probability Distributions (12 hours)

Normal, Exponential, Uniform, Beta, Gamma, distributions along with Their characterstic properties, applications and limiting/approximation cases.

Applications:

1.Normal Distribution**:** Widely used for modelling errors, measurement Uncertainties and natural phenomena.

2. Exponential Distribution: Useful for modelling waiting times, lifetimes, and Arrival processes.

3.Uniform Distribution: Applicable for random number generation and modeling Processes with equal probabilities within a range.

4 .Gamma Distribution: Versatile distribution for modelling waiting times, Survival data, and skewed distributions.

5. Beta Distribution: Used for modelling proportions and probabilities**.**

Activity :

1. **Area Under the Curve:**

- Provide students with different shapes (rectangles, triangles, trapezoids) and ask them to find the area.
- Explain the concept of probability as an area under a curve.
- Introduce the idea of a probability density function (PDF) as a curve whose total area is 1.

2. **Normal Distribution Exploration:**

- Provide students with data sets that follow a normal distribution (e.g., heights, weights, IQ scores).
- Calculate the mean and standard deviation.
- Use graphing calculators or statistical software to plot the data and overlay a normal curve.
- Discuss the characteristics of the normal curve (bell shape, symmetry).

3. **Uniform Distribution Simulation:**

- Simulate a random number generator to produce numbers between 0 and 1.
- Plot a histogram of the generated numbers.
- Explain the concept of a uniform distribution and compare the histogram to the theoretical uniform distribution.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL236**			
Title of the Course				**Probability Foundation for AI**			
Offered to: (Programme/s)				**B.Sc.(Honors) -Artificial Intelligence**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any							

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. (a) State and prove addition theorem of probability for two events. (CO-1, K-1)

(OR)

(b) Define axiomatic definition of probability. (CO-1, K-1)

2. a. Define the Random variables and state its types. (CO-2, K-1)

(OR)

b. Define distribution function and state its properties (Co-2, K-1)

3. a. State and prove multiplication theorem of expectation(Co-3, K-1)

(OR)

b. State and prove Cauchy-Schwartz Inequality (Co-3, K-1)

4. a. . Define Bernoulli distributions and write its applications (Co-1, K-1)

(OR)

b. Define binomial distributions and its applications (Co-4, K-1)

5. a. Define rectangular distributions and its mean, variance. (Co-5, K-1)

(OR)

b. Define normal distribution and write its properties. (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a) State and prove the addition theorem of probability for n events. (CO-1, K-1)

(OR)

(b) For two events A and B, prove that (CO-1, K-3)

(i) $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ (ii) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$ (iii) If $B \subset A$ then $P(A \cap \bar{B}) = P(A) - P(B)$ (iv) If $A \subset B$ then $P(\bar{A} \cap B) = P(B) - P(A)$

7. a. A random variable has the following probability distribution

x	0	1	2	3	4	5	6	7	8
P(X= x)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine 'a' (ii) Find P(X<3), P(X ≥ 3) and P(0<X<5)

(iii) Find the distribution function of X. (Co-2, K-3)

(OR)

b. The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f. $f(x) = 6x(1-x),\ 0 \le x \le 1$.

i) Check that f(x) is p.d.f.,

ii) Determine a number **b** such that P(X <**b**) = P(X >**b**). (Co-2, K-3)

8. a. state and prove Chebychev's inequalities.

(ii) A fair die is tossed 720 times. Use Chebychev's inequalities to find a lower bound for the probability of getting 100 to 140 possibility of six points on the face of die. (Co-3, K-3)

(OR)

b. Define moment generating function and cumulate generating functions. (Co-2, K-3)

9. a. Define Poisson distributions and its properties. (Co-4, K-1)

(OR)

b. Fit a Binomial Distribution, by direct method for the following data: (Co-4, K-1)

X	0	1	2	3	4	5	6
f	7	64	140	210	132	75	12

10. (a) Fit a normal distribution and obtain the expected frequencies using ordinary method.

C.I	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
f	3	21	150	335	326	135	26	4

(Co-5, K-1)

(OR)

b. A normal population is having the mean μ =30 and S.D σ = 5. If X is a variable of this population then find the probabilities of (i)$26 \le X \le 40$ (ii) $X \ge 45$ (Co-5, K-1)

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10
(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAP236**			
Title of the Course				**Statistical Data Analysis with R- Programming**			
Offered to: (Programme/s)				**B.Sc.(Honors) – Artificial Intelligence**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :							
Pre-requisites, if any							

Course Description:

An inferential statistics lab course is designed to provide hands-on experience in applying statistical methods to real-world data. While the theoretical concepts are typically covered in a lecture-based course, the lab focuses on the practical implementation of these concepts using statistical software.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	**Develop proficiency in statistical software:** Students will become proficient in using statistical software packages (like SPSS, R, Python, or Excel) to perform complex statistical analyses
2	**Apply statistical concepts to real data:** Students will analyze real-world datasets to draw meaningful conclusions and make informed decisions
3	**Understand the applications of probability distributions:** Students will learn to critically evaluate the appropriateness of different probability distributions for given data and research questions
4	**Develop critical thinking and problem-solving skills:** Students will learn to interpret statistical results, identify potential issues, and communicate findings effectively.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	apply statistical analysis that can test hypotheses under parametric approaches	K3	7	1
CO2	apply statistical analysis that can test hypotheses under non-parametric approaches.	K3	7	1
CO3	draw the inferences for various non – parametric methods for Two samples using excel	K4	7	2
CO4	draw the inferences of various large samples using Excel.	K4	7	2
CO5	draw the inferences of various small samples	K4	7	2

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1								**3**	3
CO2								**3**	3
CO3								**2**	3
CO4								**2**	3
CO5								**2**	3

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

1. i) Bayesian probabilities ii) boole's in equality
2. Generate the random numbers.
3. Fitting of binomial distribution by using random numbers.
4. Fitting of binomial distribution by using recurrence relation method.
5. Fitting of poison distribution by using random numbers.
6. Fitting of geometric distribution.
7. Fitting of negative binomial distribution.
8. Fitting of normal distribution by using random numbers.
9. Fitting of normal distribution by ordinates method.
10. Fitting of uniform distributions using random numbers.

III. References

1. S. C. Gupta & V. K. Kapoor: Fundamentals of Mathematical Statistics, SultanChand&Sons, New Delhi.
2. O.P.Gupta: Mathematical Statistics, KedarnathRamnath&Co.
3. P.N.Arora &S.Arora: QuantitativeAptitude Statistics–VolII,S.Chand&CompanyLtd.
4. K.Rohatgi&EhsanesSaleh: An Introductionto Probabilityand Statistics, JohnWiley&Sons.

IV. SuggestedCo-curricularActivities:

1. Training of students byrelated industrial experts
2. Assignmentsincludingtechnicalassignmentsifany.
3. Seminars,GroupDiscussions,Quiz,Debatesetconrelatedtopics.
4. Preparationofaudioandvideosontoolsofdiagrammaticandgraphicalrepresentations.
5. Collectionofmaterial/figures/photos/authorphotoesofrelatedtopics.
6. Invitedlecturesandpresentationsofstalwartstothose topics.
7. Visits/fieldtripsoffirms,researchorganizationsetc.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL237**			
Title of the Course				**ESSENTIAL STATISTICS FOR BUSINESS MANAGEMENT**			
Offered to: (Programme/s)				**BBA**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:		**2024-25**		**Semester:**			**3**
Course Category:		**MAJOR**		**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:		**2024 - 25**		**Percentage:**		**NA**	
Type of the Course:				**Theory**			
Crosscutting Issues of the Course :							
Pre-requisites, if any							

Course Description:

Descriptive statistics is a foundational course that introduces students to the fundamental methods for summarizing, organizing, and presenting data. This course provides a comprehensive overview of statistical concepts and techniques used to describe and understand data sets.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Understand the concepts of data organization, classification, tabulation and presentation of data.
2	Understand the impact of outliers on each measure of central tendency.
3	Understand the properties and applications of each measure of dispersion.
4	Understand the relationship between moments and also identify the shape of the distribution using skewness and kurtosis.
5	

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	To impart knowledge on Statistical concepts like Data Collection and Classification.	K1	1	1
CO2	Describe the central tendency value measurement	K2	1	1
CO3	Knowledge of various types of data, their organization and evaluation of summary measures	K3	1	1
CO4	Knowledge of various types of data such as non- central and central moments, measures of skewness and kurtosis.	K3	1	1
CO5	Statistically analyze the strengths of relationship between variables and to outline the vital area of regression models applicable in a wide variety of real time situations	K3	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX										
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	
CO1	2							**1**		
CO2	2							**1**		
CO3	2							**1**		
CO4	2							**1**		
CO5							**3**	**1**		

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

I. Course Structure:

Unit– 1: Introduction to Statistics (12 hours)

Definition of statistics. Importance, Scope and limitations of Statistics, Collection of data – Primary and Secondary, Classification, Tabulation. Construction of frequency distribution. Diagrammatic representation of data – Bar and Pie. Graphical representation of data – Histogram, frequency polygon and Ogive curves.

Applications :

1. **Market research:** Understanding consumer behavior, preferences, and market trends. Estimating consumer preferences and demand for products
2. **Financial analysis:** Making investment decisions, risk assessment, and portfolio management.Estimating student performance and program effectiveness.
3. **Sales forecasting:** Predicting future sales based on historical data.
4. **Quality control:** Ensuring product quality and consistency.

Activity 1 :

d. Conduct a class survey on favorite subjects, hobbies, or sports.
e. Collect data on the number of students present each day for a week.
f. Gather information about the height or weight of classmates.

Activity 2:

a. Introduce the concept of data cleaning by providing a dataset with errors or inconsistencies. Students can practice identifying and correcting these issues.
b. Find datasets online (weather data, sports statistics, population data) and represent them graphically
c. Provide students with datasets and ask them to choose the most appropriate graph to represent the data (bar graph, histogram, pie chart, line graph).

Unit–2: Measures of Central tendency (12 hours)

Objectives of average, Characteristics of a good average, Arithmetic mean, Geometric mean, Harmonic mean, Median and Mode – Merits, demerits, properties and applications.

Applications

1. Calculates the average sales for each product, helping identify top-performing items.
2. Calculates the average profit margin to assess overall business profitability.
3. Calculates the average inventory turnover rate to optimize stock levels and reduce carrying costs.
4. Calculates the mean loss amount for insurance claims, aiding in premium calculation and risk assessment.
5. Determines the middle revenue value, useful for understanding market penetration and pricing strategies.

Activity

1. Discuss everyday scenarios where central tendency is used (e.g., average temperature, median income).
2. Assign different groups to calculate mean, median, and mode for different data sets.
3. Provide students with business-related datasets (e.g., sales figures, customer demographics, financial performance). Have them calculate and interpret mean, median, and mode to draw conclusions.
4. Present a business problem (e.g., pricing strategy, product launch, investment decision) and provide relevant data. Have students calculate measures of central tendency.

Unit–3: Measures of Dispersion (12 hours)

Introduction, Characteristics of an ideal measures of dispersion. Absolute and Relative measures of dispersion – Rage, Quartile Deviation, Mean Deviation, Standard Deviation and Variance – Merits, demerits, properties and applications.

Applications

1. Standard deviation of returns is a key measure of investment risk. Higher volatility indicates higher risk.
2. Analyzing the dispersion of price changes and corresponding demand changes helps in pricing strategies.
3. Standard deviation helps in setting quality control limits and identifying process variations.
4. Understanding the variability of lead times helps in managing supply chain risks.
5. Understanding the dispersion of historical sales data helps in improving forecast accuracy.

Activity 1:

1. Provide sales data for a product over a period. Calculate range, variance, and standard deviation. Analyze the results to understand sales consistency.
2. Analyze financial data (e.g., profit margins, revenue growth) of different companies. Compare the dispersion of these metrics to assess business stability.
3. Discuss real-world business scenarios where dispersion is crucial (e.g., stock price volatility, customer churn rate).

Unit–4: (12 hours)

Moments – Concept, Central and Non – central moments. Interrelationship between central and non – central moments and vice versa.

Skewness – Karl Pearson's coefficient of skewness, Bowley's coefficient of skewness based on moments.

Kurtosis – Concept, Meaures of kurtosis based on moments and simple problems.

Applications

1. Understanding the skewness of sales data can improve forecasting accuracy and inventory management.
2. Kurtosis helps identify market anomalies and potential trading opportunities.
3. Kurtosis can be used to assess the consistency of a production process.
4. Understanding the skewness and kurtosis of customer behavior can help create targeted marketing campaigns and improve customer retention.

Activity 1 :

1. Provide students with real-world business datasets (e.g., stock prices, sales figures, customer demographics) and guide students through manual calculations of moments, skewness, and kurtosis to understand the underlying formulas.
2. Analyze financial transaction data to identify unusual patterns using moments, skewness, and kurtosis.

3. Analyze sales data to identify trends, outliers, and the shape of the distribution using these statistical measures

Unit–5: (12 hours)

Correlation – Introduction, types of correlation, Methods of studying correlation – Scatter diagram, Karl Pearson's coefficient of correlation and Spearman's rank correlation – Properties and Applications.

Regression – Introduction, lines of regression, coefficient of regression - Properties and Applications.

Applications

1. Correlation measures the strength and direction of the relationship between two variables. It helps businesses understand how variables are connected. Example: Correlation between advertising expenditure and sales.
2. Regression analysis helps predict the value of a dependent variable based on independent variables. Example, predicting sales based on advertising expenditure, price, and competition.
3. Regression models can be used to forecast market trends and demand.
4. Correlation between marketing expenditure and sales. Regression analysis to predict sales based on marketing campaigns.
5. Correlation between production factors and output. Regression analysis to optimize production processes.

Activity 1:

1. Collect data on sales, advertising expenditure, and competitor activity. Calculate correlation between these variables. Build a regression model to predict future sales.
2. Analyze the relationship between price and demand using regression analysis to determine optimal pricing strategy.
3. Analyze customer data (demographics, purchase history, etc.) to identify customer segments using correlation analysis.
4. Provide students with real-world business datasets (e.g., sales figures, marketing expenditure, customer demographics). Guide students through manual calculations of correlation coefficients and simple linear regression to understand the underlying formulas.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL237**			
Title of the Course				**ESSENTIAL STATISTICS FOR BUSINESS MANAGEMENT**			
Offered to: (Programme/s)				**BBA GENERAL**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:				**Theory**			
Crosscutting Issues of the Course :							
Pre-requisites, if any							

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. (a) Define Primary data and secondary data. (CO-1, K-1)

(OR)

(b) The daily wages of workers in a locality are given below. Construct Histogram to represent the data (CO-1, K-3)

Daily Wages	No. Of workers
200-225	07
225-250	12
250-275	21
275-300	26
300-325	18
325-350	10
350-375	04

2. a. Explain A.M, G.M and write its merits and demerits. (Co-2, K-2)

(OR)

b. Obtain Median to the following data (Co-2, K-1)

X	1	2	3	4	5	6	7	8	9
F	8	10	11	16	20	25	15	09	6

3. a. Explain Standard deviation and write its merits and demerits (Co-3, K-2)

(OR)

b. Find Q.D for the values 8,2,11,10,9,15,4 (Co-3, K-1)

4. a. . Define Skewness and write one example (Co-4, K-1)

(OR)

b. Define Kurtosis and write one example (Co-4, K-1)

5. a. Define correlation and write the properties of correlation (Co-5, K-1)

(OR)

b. Difference between correlation and regression. (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. Explain Classification and tabulation of data (Co-1, K-2)

(OR)

b. Construct ogive curve to the following data (Co-1, K-3)

Age	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of persons	50	70	100	180	150	120	70	59

7. a. Find median to the Following data (Co-2, K-1)

C.I	**40-50**	**50-60**	**60-70**	**70-80**	**80-90**
F	**5**	**12**	**23**	**8**	**2**

(OR)

b. Find G.M to the following data (Co-2, K-1)

C.I	4-8	8-12	12-16	16-20	20-24
F	2	4	8	5	3

8. a. Find S.D to the following data (Co-3, K-1)

C.I	0-20	20-40	40-60	60-80	80-100
F	10	25	40	15	10

(OR)

b. Explain Q.D and write its merits and demerits (Co-3, K-2)

9. a. Explain Relation between central moments in terms of Non-central moments(Co-4, K-2)

(OR)

b. Compute first four central moments and also find Sheppard's correction to the following data

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70
F	2	8	12	40	20	15	3

(Co-4, K-6)

10. (a) Derive the regression line of Y on X. (Co-5,K-3)

(OR)

(b)Calculate correlation coefficient to the following data: (Co-5, K-3)

X	10	15	12	17	13	16	24	14	22	20
Y	30	42	45	46	33	34	40	35	39	38

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAP237**			
Title of the Course				**EXCEL BASED STATISTICS LAB FOR BUSINESS MANAGEMNT**			
Offered to: (Programme/s)				**BBA**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:							
Crosscutting Issues of the Course :							
Pre-requisites, if any							

Course Description:

To equip students with practical skills in applying descriptive statistical methods to real-world datasets. This course focuses on hands-on experience in data collection, organization, analysis, and interpretation using various statistical techniques and software.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	**Statistical Software Proficiency:** Utilize statistical software (Excel, SPSS, Python, R, etc.) to perform calculations, create visualizations, and interpret results efficiently.
2	**Data Summarization:** Calculate and interpret measures of central tendency (mean, median, mode) and dispersion (range, variance, standard deviation).
3	**Data Distribution:** Analyze data distribution patterns, including skewness and kurtosis.
4	**Relationship Analysis:** Explore relationships between variables using correlation and simple linear regression.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Draw the suitable diagram and graphs of the given sample data	K3	7	1
CO2	**Apply** measures of central tendency to solve real-world problems.	K3	7	2
CO3	**Apply** measures of dispersion to assess data variability and make informed decisions.	K3	7	2
CO4	**Apply** the concepts of moments, skewness, and kurtosis to analyze real-world data sets.	K3	7	2
CO5	**Apply** correlation and regression analysis to solve real-world problems.	K3	7	2

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1							**2**	**1**	
CO2							**3**		**3**
CO3							**3**		**3**
CO4							**3**		**3**
CO5							**3**		**3**

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

1. Diagrammatic and Graphical representation of data – Bar, Pie, Histogram, Frequency polygon and Ogive curves.
2. Diagrammatic and Graphical representation of data – Bar, Pie, Histogram, Frequency polygon and Ogive curves using MS – Excel.
3. Computations of Measures of central tendency – Arithmetic mean, Geometric Mean, Harmonic Mean, Median and Mode
4. Computations of Measures of Dispersion – Quartile Deviation, Mean Deviation, Standard Deviation, Variance and Coefficient of variation.
5. Computations of Non – central, Central, $\beta_1 and\ \beta_2$ and Sheppard's correction for grouped data.
6. Computations of Karl Pearson's Coefficient of skewness and Bowley's Coefficient of skewness for grouped data.
7. Computations of Correlation coefficient for ungrouped data.
8. Forming regression lines for ungrouped data.
9. Univariate Data Analysis of a raw – data set using MS – Excel.

II. References

1. S. C. Gupta & V. K. Kapoor: Fundamentals of Mathematical Statistics, SultanChand & Sons, New Delhi.
2. O. P. Gupta: Mathematical Statistics, Kedar nath Ram nath & Co.
3. P. N. Arora & S. Arora: Quantitative Aptitude Statistics – Vol II, S. Chand & Company Ltd.
4. K. Rohatgi & Ehsanes Saleh: An Introduction to Probability and Statistics, John Wiley & Sons.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10
(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL235**			
Title of the Course				**Descriptive Statistics and Probability**			
Offered to: (Programme/s)				B.Sc. (H)-CSC (A & B)			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121**			

Course Description:
Probability distributions are a fundamental concept in statistics, providing a mathematical framework for understanding and analyzing random phenomena. This course delves into the theory and applications of probability distributions, equipping students with the tools to model and predict uncertain events.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Grasp the fundamental concepts of probability and its applications.
2	Understand the distinction between discrete and continuous random variables.
3	Learn about various probability distributions, including their properties, applications, and limitations.
4	Develop skills in calculating probabilities, expected values, and variances for different distributions.
5	Apply probability distributions to real-world problems in fields like finance, engineering, and science.

Course Outcomes
At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Calculate probabilities and analyze random events.	K1	7	2
CO2	Model real-world phenomena using appropriate probability distributions.	K2	7	2
CO3	Make informed decisions based on probabilistic information.	K3	7	2
CO4	Apply probability concepts to various fields and industries.	K3	7	1
CO5	Evaluate the assumptions underlying the use of exact sampling distributions.	K3	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PSO1	PSO2
CO1							2		3
CO2							3		3
CO3							3		2
CO4	✓						3	3	
CO5							2	1	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

III. Course Structure:

Unit – 1: Measures of Central Tendency

Mathematical averages- arithmetic mean, geometric mean and harmonic mean. Properties and applications. Positional Averages: Mode, Median – properties and problems.

Measures of Dispersion

Absolute and Relative measures of Dispersion-Range, standard deviation, and co-efficient of variation - Properties and applications.

Moments: Importance of moments, central and non-central moments and their inter-relationships (Statements only), Sheppard's corrections for central moments for grouped data. Measures of Skewness and Kurtosis with simple problems

Applications :

1. Calculating average income, average test scores, average product ratings.
2. Identifying the most popular product, determining the most common shoe size, analyzing categorical data
3. Quick estimate of data spread, used in quality control to monitor process variation.
4. Risk assessment in finance, quality control, understanding data variability.

Activities:

1. Discuss real-world applications of these measures in various fields (economics, finance, sports, etc.)
2. Collect data on a topic of interest (e.g., heights, exam scores, salaries) to create a tangible dataset for analysis..
3. Calculate mean, median, mode, range, variance, and standard deviation using both manual calculations and statistical software (Excel, SPSS).
4. Analyze different datasets and compare their findings.
5. Studies involving data analysis and ask students to apply their knowledge to solve problems.

Unit–2: Probability

Terminology - Random experiments, sample space, exhaustive, favorable, mutually exclusive, equally likely, conditional and independent events. Mathematical, Statistical and Axiomatic definitions of probabilities. Addition law of probabilities. Boole's inequalities and problems. Conditional Probability-multiplication law of probability. Pairwise independent events and conditions for mutual independence of n events and Baye's theorem and its applications

Applications

Probability, the mathematical study of chance and uncertainty, finds applications in a vast array of fields. Here are some key areas

- Business and Economics
- Social Sciences
- Engineering and Science
- Medicine
- Insurance
- Investment

6. Probability is fundamental to understanding information transmission and storage
2. Behavior analysis, decision-making, and social networks utilize probabilistic models.
3. Election forecasting, public opinion polling, and political risk assessment use probability.

Activities:

1. Coin Flipping and Dice Rolling Experiments
2. Students can analyze lottery systems to understand probability and expected value.
3. Students can collect real-world data (e.g., sports statistics, weather data) and analyze it Using probability concepts.
4. Provide real-world scenarios (e.g., insurance, finance, medicine) and ask students to apply probability concepts to solve problems
5. Utilize online resources for interactive probability games and simulations.

Unit–3: Univariate Random Variables:

Definition, Discrete and Continuous random variables -Probability mass function and Probability density function with illustrations. Distribution function and its properties.

Mathematical Expectations: Definition, Properties of Expectations - Addition and Multiplication theorems of expectation. Properties of Variance and Covariance. Cauchy-Schwartz Inequality, Chebyshev's Inequality (Statements only). Generating Functions. Definition of moment generating function (m.g.f), Cumulant generating function (c.g.f), Probability generating function (p.g.f) and Characteristic function (c.f) and their properties

Applications

1. **Demographics:** Age, income, education level, and gender distribution within a population.
2. **Finance:** Stock prices, interest rates, and return on investments
3. **Quality Control:** Product dimensions, weight, or defect rates.
4. **Engineering:** Reliability analysis.
5. **Telecommunications:** Signal processing, error correction

Activities:

1. Students can flip a coin multiple times and record the number of heads. They can then calculate the probability of getting different numbers of heads and compare it to the theoretical binomial distribution.
2. Students can roll a die multiple times and record the outcomes. They can then calculate the probability of different outcomes and compare it to the theoretical uniform distribution
3. Games like poker or blackjack can be used to explore probability concepts, such as conditional probability and expected value.
4. Students can calculate and graph CDFs for different random variables and understand their

significance
5. Assign group activities where students collect data, analyze it, and model it using appropriate random variables

Unit – IV: Probability Distributions:

Discrete Probability Distributions:

Definition Binomial distribution, properties and its applications, Poisson distributions –properties and its applications. Geometric distributions- properties and its Applications.-simple problems.

Continuous Probability Distributions:

Definition of Uniform distribution, properties, Normal distribution, properties and its applications, exponential distributions –properties and its applications

Applications:

1. Normal Distribution: Widely used for modelling continuous data, such as heights, weights, and measurement errors.
2. Binomial Distribution: Used for modelling the number of successes in a fixed number of Bernoulli trials (e.g., coin flips, quality control).
3. Poisson Distribution: Used for modeling the number of events occurring in a fixed interval of time or space (e.g., arrivals at a service counter, radioactive decay).
4. Exponential Distribution: Used for modeling the time between events in a Poisson process (e.g., time between customer arrivals).

Activities:

1. Dice Experiments:
   a. Roll two dice and find the sum of the numbers and Create a table to record the Possible outcomes and their frequencies
   b. Graph the probability distribution.
2. Coin Tossing:
   a. Toss a coin multiple times and record the number of heads
   b. Compare the theoretical and experimental probabilities.
3. Binomial Distribution:
   a. Simulate flipping a coin multiple times and count the number of heads.
   b. Use a graphing calculator or software to plot the binomial distribution.
4. Poisson Distribution:
   a. Simulate the number of cars arriving at an intersection in a given time Interval.
   b. Analyse the distribution and calculate probabilities.
5. Normal Distribution:
   a. Create a histogram of the data.
   b. Overlay a normal curve on the histogram.

Unit – V: Exact Sampling distributions:

Basic Concepts- Population, Sample, Parameter, Statistic, Sampling Distribution of a statistic and Standard Error and its uses. Chi-square distribution- definition, properties, Graph and applications. t-Distribution – definition, properties, Graph and applications. F-distribution – definition, properties, Graph and applications

Applications:

1. Hypothesis Testing with Small Sample Sizes
2. Non-Normal Population Distributions
3. Exact Confidence Intervals
4. Rare Event Probability Estimation
5. Power Analysis

Activities:

1. Sampling Distribution Simulation
2. Empirical Sampling Distributions with Dice
3. Classroom Polling
4. Case Study Analysis
5. Interactive Discussion and Reflection

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAL235**			
Title of the Course				**Descriptive Statistics and Probability**			
Offered to: (Programme/s)				B.Sc. (H)-CSC (A & B)			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121**			

Time: 3 hours **Maximum Marks: 70**

Section - A

Answer the following questions **5 X 4M = 20M**

1. (a) Explain the concept of censored data. (CO-1, K-2)

(OR)

b. The first four moments of a distribution about the value 5 are -4, 22,-117 and 560. Find the corresponding moments about the mean and also comment on the nature of the data. (CO-1, K-3)

2. (a) State and prove addition theorem of probability for two events. (CO-2, K-1)

(OR)

(b) Define axiomatic definition of probability. (CO-4, K-1)

3. a) Define the Random variables and state its types. (Co-3, K-1)

(OR)

b. Define distribution function and state its properties (Co-3, K-1)

4. a) Define binomial distributions and its applications (Co-4, K-1)

(OR)

b. Define rectangular distributions and its mean, variance. (Co-4, K-1)

5. a. Define chi-square distributions and its applications. (Co-5, K-1)

(OR)

b. Define the F-Distribution and its applications. (Co-5, K-1)

Section - B

Answer the following questions **5 X 10M = 50M**

6. a. Explain A.M, G.M, H.M and writs its merits and demerits (CO-1, K-2)

(OR)

b. Define moments. Establish the relationship between the moments about mean (Central moments in terms of moments about any arbitrary point and vice versa. (CO-1, K-2)

7. a) State and prove the addition theorem of probability for n events. (CO-2, K-1)

(OR)

(b) For two events A and B, prove that (CO-2, K-3)

(i) $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ (ii) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

(iii) If $B \subset A$ then $P(A \cap \bar{B}) = P(A) - P(B)$(iv) If $A \subset B$ then $P(\bar{A} \cap B) = P(B) - P(A)$

8. a. A random variable has the following probability distribution

x	0	1	2	3	4	5	6	7	8
P(X= x)	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine 'a' (ii) Find P(X<3), P(X ≥ 3) and P(0<X<5)

(iii) Find the distribution function of X. (Co-3, K-3)

(OR)

b. The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f. $f(x) = 6x(1-x),\ 0 \le x \le 1$.

i) Check that f(x) is p.d.f.,

ii) Determine a number **b** such that P(X <**b**) = P(X >**b**). (Co-3, K-3)

9. a. Fit a Binomial Distribution, by direct method for the following data: (Co-4, K-1)

X	0	1	2	3	4	5	6
f	7	64	140	210	132	75	12

(OR)

b. Obtain MGF of Normal distribution hence obtain mean and variance of it (Co-4, K-1)

10. a. Derive Student's t Distribution. (Co-5, K-3)

(OR)

b. Derive the relation between F and χ^2 distribution. (Co-5, K-3)

****

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMAP235**			
Title of the Course				**Descriptive Statistics and Probability Lab**			
Offered to: (Programme/s)				B.Sc. (H)-CSC (A& B)			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121**			

Course Description:

Probability distributions are a fundamental concept in statistics, providing a mathematical framework for understanding and analyzing random phenomena. This course delves into the theory and applications of probability distributions, equipping students with the tools to model and predict uncertain events.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	Grasp the fundamental concepts of probability and its applications.
2	Understand the distinction between discrete and continuous random variables.
3	Learn about various probability distributions, including their properties, applications, and limitations.
4	Develop skills in calculating probabilities, expected values, and variances for different distributions.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	Calculate probabilities and analyze random events.	K1	7	2
CO2	Model real-world phenomena using appropriate probability distributions.	K2	7	2
CO3	Make informed decisions based on probabilistic information.	K3	7	2
CO4	Apply probability concepts to various fields and industries.	K3	7	1
CO5	Evaluate the assumptions underlying the use of exact sampling distributions.	K3	7	1

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PSO1	PSO2
CO1							2		3
CO2							3		3
CO3							3		2
CO4							3	3	
CO5							2	1	

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

Unit– 1:

Measures of Central Tendency:

1. Collect a dataset of at least 30 observations. This could be from a real-world scenario (e.g., heights of students, exam scores, salaries)
2. Calculate the mean, median, and mode for the dataset
3. Discuss the suitability of each measure based on the data distribution

Applications:

1. **Market Research:** Determining average customer satisfaction, age, income, or spending Habits.
2. **Sales Analysis:** Identifying the most popular products or services.
3. **Teacher Effectiveness:** Evaluating average student ratings or achievement levels.

Activities/Examples/Case Study:

1. Case Study 1: Student Performance Analysis

Problem: A school administrator wants to assess the overall performance of students in a math class.

Case Study 2: Customer Satisfaction Survey

Problem: A company wants to measure customer satisfaction with its products.

Case Study 3: Income Distribution Analysis

Problem: A government agency wants to study income inequality in a city.

Measures of Dispersion:

1. Calculate the range, variance, and standard deviation for the dataset
2. Calculate the coefficient of variation to compare variability between datasets

Applications:

1. **Teacher Effectiveness:** Evaluating the variability of student outcomes within a class.
2. **Curriculum Evaluation:** Assessing the consistency of student learning across different Groups.
3. **Health Outcomes:** Assessing the dispersion of health indicators (e.g., blood pressure, Cholesterol levels).

Activities/Examples/Case Study:

Case Study 1: Quality Control in Manufacturing

Problem: A manufacturing company wants to assess the consistency of the weight of its products.

Case Study 2: Investment Portfolio Analysis

Problem: An investor wants to compare the risk of two different investment portfolios.

Unit – IV: Probability Distributions:

Discrete Probability Distributions:

1. Fitting of a Binomial distribution and calculate their expected frequencies
2. Fitting of a Poisson distribution and calculate their expected frequencies
3. Fitting of a Geometric distribution and calculate their expected frequencies

Applications:

1. **Risk assessment:** Evaluating the probability of financial losses or gains.
2. **Portfolio management:** Allocating investments to optimize returns and minimize risk.
3. **Algorithm analysis:** Analysing the expected running time of algorithms.
4. **Cryptography:** Designing secure encryption methods based on probability theory.

Activities/Examples/Case Study:

1.Case Study 1: Quality Control in Manufacturing

Problem: A manufacturing company produces light bulbs. The company wants to assess the probability of a light bulb being defective.

2.Case Study 2: Customer Satisfaction Survey

Problem: A company conducts a customer satisfaction survey with a Likert scale of 1-5 (1 being very dissatisfied, 5 being very satisfied).

3.Case Study 3: Network Traffic Analysis

Problem: A network administrator wants to analyze the number of packets arriving at a Network interface per second.

Continuous Probability Distributions:

1. Fitting of a Norma distribution by using areas method
2. Fitting of a Norma distribution by using ordinate method
3. Fitting of a Exponential distribution

Applications:

1 .**Confidence intervals:** Estimating the range of values within which a population Parameter lies.

2. **Statistical inference:** Making inferences about a population based on a sample.
3. **Simulation:** Modelling complex systems to understand their behaviour.

Activities/Examples/Case Study:

Case Study 1: Product Lifetime Analysis

Problem: A manufacturing company wants to analyze the expected lifespan of its products.

Case Study 2: Investment Portfolio Analysis

Problem: An investor wants to assess the risk and return of an investment portfolio.

Case Study 3: Weather Forecasting

Problem: A meteorologist wants to predict the probability of rainfall.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMDL101**			
Title of the Course				**BASIC STATISTICS**			
Offered to: (Programme/s)				**B.Com.(Honors) (Gen, TPP, Fin, BPM, Banking & CA) , ,BCA(Honors),cscs,B.sc (Honors)(BOT,AQUA,BIO,FST,MICRO)**			
L	**2**	**T**	**0**	**P**	**0**	**C**	**2**
Year of Introduction:	**2024-25**			**Semester:**		**3**	
Course Category:				**Course Relates to:**	**Local, Regional, National, Global**		
Year of Introduction:	**2024 - 25**			**Percentage:**	**NA**		
Type of the Course:							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121**			

Course Prerequisites (if any):

Course Description:

This course helps the students to familiarize with the ways in which we talk about descriptive statistics.

Course Objectives:

1) To compute various measures of central tendency, dispersion, skewness and kurtosis.
2) To get the knowledge regarding qualitative factors

Learning Outcomes: At the end of the course, the student will able to:

1) To familiar with the measures of central tendency and dispersion throw light on reliability of average and control of variability.
2) To find the association of the categorical data by using attributes.

Course Outcomes:		
Course Outcome	Upon successful completion of this course, students should have the knowledge and skills to:	**Program Outcomes Mapping**
CO 1	Understand the concepts of data organization, classification, tabulation and presentation of data.	2
CO 2	Understand the impact of outliers on each measure of central tendency.	2
CO3	Understand the properties and applications of each measure of dispersion.	2

Syllabus

Course Details

Unit	**Learning Units**	**Lecture Hours**
I	Meaning, scope and limitations of Statistics **Collection of data:** Primary and Secondary, Classification and Tabulation, Construction of frequency distribution.	**10**

	Graphical Representation: Histogram, Bar, Pie and Frequency polygon.	
II	**Measures of Central Tendency:** Features of good average, Arithmetic Mean, Median, Mode. Empirical relationship between Mean Median and Mode and skewness based on central values.	**10**
III	**Measures of Dispersion:** Range, Quartile Deviation (QD), Mean Deviation (MD), Variance, Standard Deviation (SD), relationship between QD, MD and SD. Familiarization of the concepts relating to Correlation and Linear Regression line.	**10**

Text Book:

1. Statistics (Theory, Methods, Application) D C Sancheti, V K Kapoor, Sultan Chand and Sons, New Delhi

2. Statistical Methods, S.P. Gupta, Sultan Chand and Sons, New Delhi

3. Statistics (Theory and Practice) B.N Gupta, Sahitya Bhavan, Agra

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10
(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23STMDL101**			
Title of the Course				**BASIC STATISTICS**			
Offered to: (Programme/s)				**B.Com.(Honors) (Gen, TPP, Fin, BPM, Banking & CA) , ,BCA(Honors),cscs,B.sc (Honors)(BOT,AQUA,BIO,FST,MICRO)**			
L	**2**	**T**	**0**	**P**	**0**	**C**	**2**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:				**Course Relates to:**		**Local, Regional, National, Global**	
Year of Introduction:	**2024 - 25**			**Percentage:**		**NA**	
Type of the Course:							
Crosscutting Issues of the Course :							
Pre-requisites, if any				**23STMAL121**			

Basic Statistics
23STMDL101

Max. Marks: 35
Min. Marks: 14 Pass

Section – A

Answer the following **3 x 5M = 15Marks**

1. a. Define Primary data and secondary data (CO-1, K-1))

(OR)

b. The daily wages of workers in a locality are given below. Construct Histogram to represent the data (CO-1, K-3)

Daily Wages	No. Of workers
200-225	07
225-250	12
250-275	21
275-300	26
300-325	18
325-350	10
350-375	04

2. a. Explain A.M, G.M and write its merits and demerits. (Co-2, K-2)

(OR)

b. Obtain Median to the following data (Co-2, K-1)

X	1	2	3	4	5	6	7	8	9
F	8	10	11	16	20	25	15	09	6

3. a. Explain Standard deviation and write its merits and demerits (Co-3, K-2)

(OR)

b. Find Q.D for the values 8,2,11,10,9,15,4 (Co-3, K-1)

Section – B

Answer the following **2x 10M = 20Marks**

4. a. Explain Classification and tabulation of data (Co-1, K-2)

(OR)

b. Construct ogive curve to the following data (Co-1, K-3)

Age	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of persons	50	70	100	180	150	120	70	59

5. **a.** Find median to the Following data (Co-2, K-1)

C.I	**40-50**	**50-60**	**60-70**	**70-80**	**80-90**
F	**5**	**12**	**23**	**8**	**2**

(OR)

b. Find G.M to the following data (Co-2, K-1)

C.I	4-8	8-12	12-16	16-20	20-24
F	2	4	8	5	3

6. a. Find S.D to the following data (Co-3, K-1)

C.I	0-20	20-40	40-60	60-80	80-100
F	10	25	40	15	10

(OR)

b. Explain Q.D and write its merits and demerits (Co-3, K-2)

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10
(An Autonomous college in the jurisdiction of Krishna University)

Course Code				**23ANMDL102**			
Title of the Course							
Offered to: (Programme/s)				**B.B.A(Honors) & BBA(B.A)**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:				**Course Relates to:**		**GLOBAL**	
Year of Introduction:	**2024**			**Percentage:**		**NA**	
Type of the Course: Practical – MDC							
Crosscutting Issues of the Course :							
Pre-requisites, if any							

30Hours

POWER BI

Prerequisite:

Course Objectives:

To import, transform and cleanse data using Power Query Editor, build a data model for self-service reporting, manipulate the model with DAX, publish and share visualizations.

Course Outcomes: At the end of this course, students should be able:

CO1: To explain the concept Power Pivot and interface with excel analytic way **(PO5, PO6)**

CO2: To choose the algorithms for combine data quickly from a variety of sources into your model **(PO5, PO6)**

CO3: To organize the data various sources, clean, merge, filter data and calculated methods **(PO4, PO5, PO6)**

CO4: To construct the model, relationships between in the models, user friendly models **(PO5, PO6)**

CO5: To decide BI environment, data clean, shaping, table relationships and analysis techniques **(PO4, PO5, PO6)**

Mapping of Course Outcomes (COs) with Programme Outcomes (POs) & PSOs

CO	BTL	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PSO1	PSO2
CO1	K2					3	3			
CO2	K3					3	2			
CO3	K3				1	3	3			
CO4	K3					3	2			
CO5	K5				1	3	3			

Syllabus		
Unit	**Learning Units**	**Lecture Hours**
I	**Introduction Power Pivot** Introduction of Pivot - Use Power Pivot – x Velocity in-memory analytics engine - Exploring the Data Model Management interface - Analyzing data using a pivot table	6

II	**Power BI Data Import and Data Cleaning** Working with Data - Import data from relational databases - Import data from text files - Import data from a data feed - Import data from other sources, Discover and import data from various sources -	6
III	**Data Cleaning Techniques** Data Munging - Getting, cleaning, and shaping data, Cleanse data - Merge, shape, and filter data - Group and aggregate data - Insert calculated columns.	6
IV	**Power BI Data Model** Creating data Model - Explain what a data model is - Create relationships between tables in the model - Create and use a star schema - Understand when and how to deformalize the data - Create and use linked tables	6
V	**Power BI Visuals and DAX** Adding calculations and measures - Incorporating time-basedanalysis	6

List of Experiments

1. Write the Procedure for preparing a Pivot in Excel and prepare a Dashboard using samplemarketing data.

a) Offline data and online data b) Online to Online using Google forms

2. Installation of Power BI and its procedure

3. Explain the procedure in importing various format files in Power BI, write its observations

4. Power BI Data Models (Schemas in Power BI)

5. How to edit data in power BI when data is Exported use few data cleaning techniques (Munging)

6. Advance Data Cleaning techniques, Data Munging and Data collection and collaborationtechniques.

7. Write the procedure in building an association (Power Query) identify various schemas inPower BI

8. Data Visualization (charts for a sample data) constructions and analysis

9. Step in preparing a dashboard for the organization

10. Constructing Quick Measures and Dax formulas

Text Books

1. Roger F Silva, **"Power BI Create and learn"**, Version – January 2024, ISBN: 9781726793216.

2. Brett Powell, **"Mastering Microsoft Power BI"**, Packt publishing, Birmingham, UK, ISBN: 978-1-78829-723-3.

Reference Book

1. Dan Clark, **"Beginning Power BI: A practical Guide to Self – Service Data Analytics with Excel 2016 and Power BI Desktop Second Edition."** ISBN: 978-1-4842-2576-9, A Press Publications.

2. Jeff Hutchinson, **"Microsoft Power BI Desktop – Creating Visual Reports"** – July - 2019, ISBN: 9781081588908, Independently Published.

SRI DURGA MALLESWARA SIDDHARATHA MAHILA KALASALA: VIJAYAWADA-10

(An Autonomous college in the jurisdiction of Krishna University)

Question Paper Pattern for SEE-Practical Courses

POWER BI

Course Code: 23ANMDL102 **Offered to:** BBA Business Analytics

Max. Marks: 35 **Max. Time: 2Hours**

Pass. Min: 14

(A) SEE (LAB) Model Question Paper

I	answer the following Lab Assessment Q1/Q2/Q3….	**30 Marks**
II	Viva	**2 Marks**
III	Record	**3 Marks**

(B) CONTINUOUS ASSESMENT: **15 MARKS**

TOTAL: (A) + (B) = **50MARKS**

----------END----------