PARVATHANENI BRAHMAYYA
SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada–520010
Re-accredited at 'A+' by the NAAC

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 Revision:		**NA**		**Percentage:**		**NA**	
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course :				**NA**			
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…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	define and differentiate between discrete and continuous bivariate random variables and describe the joint, marginal, and conditional distributions and their properties.	K2	1	1
CO2	analyze the independence of two random variables through simple problem-solving using their joint distribution functions and assess their relationships.	K4	1	1
CO3	apply the principle of least squares to fit different types of curves such as straight lines, second-degree polynomials, and exponential curves to given data, and solve related problems.	K3	1	1
CO4	interpret and compute measures of correlation, including Karl Pearson's and Spearman's methods, and analyze the correlation between variables using scatter diagrams and bivariate frequency distributions.	K4	1	1
CO5	evaluate the properties of linear regression models by calculating regression lines, coefficients, and coefficients of determination, and distinguish between correlation and regression through practical examples.	K5	1	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**							**2**	
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 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 their properties. Independence of random variables, and simple problems.

Examples/Applications/Case Studies:

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 education research, bivariate random variables can analyze the relationship between variables such as student performance and socioeconomic status.

Exercises/Project:

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.

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.

Specific Resources: (web): https://open.umn.edu/opentextbooks

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 and Simple problems.

Examples/Applications/Case Studies:

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 Exponential Curves - Studying compound interest
4. Fitting a Power Curve - Modeling economic growth

Exercises/Project:

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).

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).

Specific Resources: (web):

https://www.originlab.com/index.aspx?go=Products/Origin/DataAnalysis/CurveFitting

Unit – 3 Correlation **(12 hours)**

Introduction, Meaning of Correlation, Scatter Diagram, Karl Pearson's Coefficient of correlation, Calculation of the Correlation Coefficient for a Bivariate Frequency Distribution Probable Error of Correlation Coefficient, Rank Correlation(with tie and without tie).Lag and Lead in correlation.

Examples/Applications/Case Studies:

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. Correlation analysis helps in understanding the relationship between mental health variables (e.g., depression) and various life factors (e.g., social support).

Exercises/Project:

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).

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.

Specific Resources: (web): https://www.khanacademy.org/math/senior-high-school-statistics-probability/xc50280f9c512251f:2nd-quarter

Unit – 4 Regression **(12 hours)**

Concept of Regression, Linear and Non-Linear regression. Linear Regression – Regression lines, Regression coefficients and it properties and simple problems. Correlation versus Regression. Explained and Unexplained variations. Coefficient of determination.

Examples/Applications/Case Studies:

1. **Sales forecasting:** Predicting future sales based on factors like advertising expenditure, price, and competition.
2. **Market research:** Understanding customer behavior and preferences.
3. **Insurance:** Assessing risk and setting premiums.

Exercises/Project:

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.

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.

Specific Resources: (web): https://www.khanacademy.org/math/senior-high-school-statistics-probability/xc50280f9c512251f:2nd-quarter

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

Introduction-Multiple and Partial Correlation, Coefficient of multiple correlation, properties of multiple correlation coefficient, coefficient of multiple determination. Coefficient of Partial correlation, Coefficient of Partial determination. Relationship between total , multiple and partial correlation coefficients and simple problems.

Examples/Applications/Case Studies:

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.

Partial Correlation Coefficient

3. **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).
4. **Psychology:** Analyzing the correlation between intelligence and income, while accounting for the effect of education.

Exercises/Project:

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.

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.

Specific Resources: (web): https://www.khanacademy.org/math/senior-high-school-statistics- probability/xc50280f9c512251f:2nd-quarter

Text Book

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

Recommended References books:

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

***

PARVATHANENI BRAHMAYYA
SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada–520010
Re-accredited at 'A+' by the NAAC

23STMAL233: Statistical Methods

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

Section - A

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

1. a. Define the joint probability function (X,Y) and marginal probability function of X &Y (CO-1, K-1)

   (OR)

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

2. a. Explain the Principle of least squares. (CO-2, K-2)

   (OR)

   b. Explain the power curve and convert the power curve into linear equation. (CO-2, K-2)

3. a. Explain the scatter diagram method. (CO-3, K-2)

   (OR)

   b. Explain the types of correlation. (CO-3, K-2)

4. a. Write the properties of regression coefficients. (CO-4, K-1)

   (OR)

   b. What does R-squared represent in the context of regression analysis? (CO-4, K-1)

5. a. Explain multiple correlation with examples. (CO-5, K-2)

   (OR)

   b. Explain partial correlation with examples. (CO-5, K-2)

Section - B

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

6. a. Two discrete random variables X and Y have joint probability density functions $P_{XY}(x,y)=\frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y!(x-y)!}$, $y=0,1,2,...,x$; $x=0,1,2,...$, where λ and p are constants with $\lambda>0$ and 0 < p < 1. Find the marginal probability density functions of X and Y. Also find the conditional probability density functions of X given Y = y. (CO-1, K-3)

   (OR)

   b. Suppose that two-dimensional continuous random variable (X, Y) has joint p.d.f. given by $f_{XY}(x,y)=\begin{cases}6x^2y, & 0<x<1,\ 0<y<1\\ 0, & otherwise\end{cases}$.

   Find $P\left(0<X<\frac{3}{4}, \frac{1}{3}<Y<2\right)$, $P(X+Y<1)$, $P(X<1/Y<2)$. (CO-1, K-3)

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

(OR)

b. Fit an exponential curve of the form $Y = ab^X$ to the following data: (CO-2, K-3)

X	1	2	3	4	5	6	7	8
Y	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

8. a. Calculate Karl Pearson's coefficient of correlation in the following series relating to cost of living and wages. (CO-3, K-3)

Wages (in Rs.)	100	101	103	102	100	99	97	98	96	95
Cost of living	98	99	99	97	95	92	95	94	90	91

(OR)

b. Derive the rank correlation coefficient formula. (CO3, K-3)

9. a. The lines of regression of a bivariate population are $8x - 10y + 66 = 0$ and $40x - 18y = 214$ then find (i) The mean values of X and Y (ii) which is Y on X and which is X on Y (iii) Correlation coefficient between X and Y (iv) Standard deviation of Y if Variance of X=9. (CO-4, K-3)

(OR)

b. Derive the regression line Y on X. (CO-4, K-3)

10. a. The data relating to the yield of dry bark (X_1), height (X_2) and girth (X_3) for 18 cinchona plants the following correlation coefficients were obtained: $r_{12} = 0.77, r_{13} = 0.72, r_{23} = 0.52$. Find the partial correlation coefficient $r_{12.3}$ and multiple correlation coefficient $R_{1.23}$. (CO-5, K-3)

(OR)

b. Show that $1 - R_{1.23}^2 = \left(1 - r_{12}^2\right)\left(1 - r_{13.2}^2\right)$. Derive that

(i) $R_{1.23} \geq r_{12}$ and (ii) $R_{1.23}^2 = r_{12}^2 + r_{13}^2$ *if* $r_{23} = 0$ (CO-5, K-3)

***