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

Course Code				**23STVAL201**			
Title of the Course				**Applied Statistical Methods for Cognitive Systems**			
Offered to: (Programme/s)				**B.Sc. (Hons) Computer Science with Cognitive Systems**			
L	**2**	**T**	**0**	**P**	**0**	**C**	**2**
Year of Introduction:		**2024-25**		**Semester:**		**3**	
Course Category:		**VAC**		**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 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 Objectives:

S. No	COURSE OBJECTIVES
1	Model the relationship between a dependent variable and one or more independent variables.
2	Analyze the relationship between a dependent variable and multiple independent variables, considering the effects of other variables.
3	Develop a comprehensive understanding of probability and univariate random variables, including their probability distributions, functions, and characteristics, to model and analyze real-world phenomena.
4	Master the concept of mathematical expectation and its properties, enabling the calculation of mean, variance, and other statistical measures for effective decision-making.
5	Acquire proficiency in generating functions, their applications in probability theory, and their role in deriving probability distributions and moments.

Course Outcomes

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

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	**Perform regression analysis:** fit regression models to data, interpret the results, and make predictions.	K2	1	1
CO2	**Apply regression techniques to various fields:** use regression analysis to solve problems in fields such as economics, finance, marketing, and engineering.	K3	1	1
CO3	**Define and understand key probability concepts:** define and explain concepts such as probability spaces, event, etc.,	K1	1	1
CO4	**Apply probability theory to real-world problems:** use probability theory to model and analyze real-world phenomena, such as coin tosses, card draws, and statistical experiments.	K3	1	1
CO5	**Understand key Random Variables:** use of random variables, probability distributions, expected value, variance, and covariance.	K1	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	**3**							**3**	
CO2	**2**							**2**	
CO3	**2**							**2**	
CO4	**3**							**3**	
CO5	**2**							**2**	

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

Course Structure:

Unit – 1: Regression (10 hours)

Concept of Regression, Linear and Non-Linear regression. Linear Regression – Regression lines, Regression coefficients and it properties. Correlation vs Regression. Explained and Unexplained variations. Coefficient of determination. Concept of Multiple and Partial correlation coefficient.

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. **Financial modeling:** Predicting stock prices, interest rates, and economic indicators.

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.

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.

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 underfitting.

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

Unit – 2: Probability **(10 Hours)**

Basic Concepts of Probability, random experiments, outcome, sample space, exhaustive events, favorable, mutually exclusive and equally likely outcomes. Mathematical, Statistical, axiomatic definitions of probability. Addition and multiplication laws of probability (statements only). Conditional probability and Baye's theorem (statement only) and problems.

Examples/Applications/Case Studies:

1: Quality Control in Manufacturing

a. **Random experiment:** Inspecting a randomly selected product from a production line.
b. **Sample space:** {Defective, Non-defective}
c. **Event:** Selecting a defective product.
d. **Probability:** The probability of selecting a defective product is calculated based on the number of defective products in a sample.
e. **Bayes' Theorem:** Can be used to update the probability of a product being defective based on additional information, such as the results of a quality control test.

2: Medical Diagnosis

a. **Random experiment:** Conducting a medical test on a patient.
b. **Sample space:** {Positive, Negative}
c. **Event:** The test result is positive.
d. **Conditional probability:** The probability of a disease given a positive test result is calculated using Bayes' Theorem, considering the sensitivity and specificity of the test.

e. **Applications:** Bayes' Theorem is widely used in medical diagnosis to assess the likelihood of a disease based on test results and prior probabilities.

Exercises/Project:

1: Coin Tossing

Problem:

a. Toss a coin 100 times and record the number of heads and tails.
b. Calculate the experimental probability of getting a head.
c. Compare the experimental probability with the theoretical probability of getting a head.

Analysis:

a. Discuss the concept of random experiments, outcomes, and sample space in the context of coin tossing.
b. Explain the difference between theoretical and experimental probability.
c. Analyze the relationship between the number of trials and the accuracy of the experimental probability.

2: Card Drawing

Problem:

a. Draw a card from a standard deck of 52 cards without replacement.
b. Calculate the probability of drawing an Ace.
c. Calculate the probability of drawing a heart.
d. Calculate the probability of drawing an Ace of Hearts.

Analysis:

a. Discuss the concept of mutually exclusive and exhaustive events in the context of card drawing.
b. Apply the addition and multiplication laws of probability to calculate the probabilities.
c. Explore the concept of conditional probability by calculating the probability of drawing an Ace given that the card drawn is a heart.

Specific Resources: (web): https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/resources/lecture-notes/

Unit – 3 Random Variables, Mathematical Expectation and Generating Function (10 Hours)

Random variables-Definition and types of random variable, probability mass function and probability density function. Distribution function and its properties. Simple problems.

Mathematical expectations-Introduction Mathematical Expectation expected value of function of a random variable, properties of expectation and variance, Covariance. Simple problems.

Generating functions-Definition and Properties of Moment Generating Function, Cumulative Generating Function, Characteristic Function, Probability Generating Function (p.g.f.). Chebyshev's Inequality (statement only) and Chebychev's inequality (statement only).

Examples/Applications/Case Studies:

1. Quality control - Number of defective items in a sample.
2. Recording the time taken to complete a task - Time in seconds or minutes.
3. Temperature in Celsius or Fahrenheit.

Exercises/Project:

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.

Specific Resources: (web):https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/resources/lecture-notes/

Textbook:

1. S. C. Gupta, Fundamentals of Statistics, 8th Edition, 2023, Himalaya Publishing House Pvt. Ltd Ramdoot, Dr. Bhalerao Marg, Girgaon, Mumbai – 400 004, Maharashtra, India

References Books:

1. **Business Statistics A First Course, 8e Paperback – 30 October 2022, David. Levine** (Author)
2. Business Statistics: Problems & Solutions by J.K. Sharma (Author), Vikas Publishing House Pvt Ltd. Noida, UP, India

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

23STVAL201: Applied Statistical Methods for Cognitive Systems

Offered to B.Sc.(Hons) Computer Science with Cognitive Systems

Max. Marks: 35 **Semester III** **Max. Time: 2Hrs**

Section - A

Answer any THREE from the following **3 X 5M = 15M**

1. Write the properties of regression coefficients.
2. What does R-squared represent in the context of regression analysis?
3. Define various definitions of probability.
4. Define distribution function and state its properties.
5. Define Moment Generating Function (MGF) and write its properties.

Section - B

Answer any TWO from the following **2 X 10M = 20M**

6. 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.
7. 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.

8. 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)

***