II- Year II- Semester	Name of the Course	L	T	P	C
BS2201	Probability and Statistics	2	1	0	3

Course objectives:

1. To **Classify** the concepts of data science and its importance (L4) or (L2)
2. To **Interpret** the association of characteristics and through correlation and regression tools (L4)
3. To **Understand** the concepts of probability and their applications, **apply** discrete and continuous probability distributions (L3)
4. To **Design** the components of a classical hypothesis test (L6)
5. To **Infer** the statistical inferential methods based on small and large sampling tests (L4)

UNIT-I

Descriptive statistics and methods for data science: 8 hrs

Data science-Statistics Introduction-Population vs Sample-Collection of data-primary and secondary data-Types of variable: dependent and independent Categorical and Continuous variables-Data visualization-Measures of Central tendency-Measures of Variability (spread or variance)-Skewness Kurtosis.

UNIT-II

Correlation and Curve fitting: 10 hrs

Correlation- correlation coefficient-Rank correlation-Regression coefficient and properties-regression lines-Multiple regression-Method of least squares-Straight line-parabola-Exponential-Power curves.

UNIT-III

Probability and Distributions: 10 hrs

Probability-Conditional probability and Baye's theorem-Random variables-Discrete and Continuous random variables-Distribution function-Mathematical Expectation and Variance-Binomial, Poisson, Uniform and Normal distributions.

UNIT-IV

Sampling Theory: 10 hrs

Introduction–Population and samples-Sampling distribution of Means and Variance (definition only)-Central limit theorem (without proof)-Point and Interval estimations, Good estimator, Unbiased estimator, Efficiency estimator-Maximum error of estimate.

UNIT-V

Test of Hypothesis: 10 hrs

Introduction–Hypothesis-Null and Alternative Hypothesis-Type I and Type II errors-Level of significance-One tail and two-tail tests-Tests concerning one mean, two means, and proportions using

Z test, Tests concerning one mean, two means using t test, also chi-square and F tests use for small samples.

Course Outcomes

Upon successful completion of the course, the student will be able to

CO1: **Classify** the concepts of data science and its importance (L4) or (L2) (**Understand, Analyze)**

CO2: **Interpret** the association of characteristics and through correlation and regression tools (L4) **Analyze**

CO3: **Understand** the concepts of probability and their applications, **apply** discrete and continuous probability distributions (L3) **Understand, Apply**

CO4: **Design** the components of a classical hypothesis test (L6) **Understand, Design, create**

CO5: **Infer** the statistical inferential methods based on small and large sampling tests (L4) **Understand, Analyze**

Text books:

1. **Miller and Freund's,** Probability and Statistics for Engineers, 7/e, Pearson, 2008.
2. **S. C. Gupta and V. K. Kapoor,** Fundamentals of Mathematical Statistics, 11/e, Sultan Chand & Sons Publications, 2012

Reference books

1. **Shron L. Myers, Keying Ye, Ronald E Walpole,** Probability and Statistics Engineers and the Scientists, 8th Edition, Pearson 2007.
2. **Jay I. Devore,** Probability and Statistics for Engineering and the Sciences, 8th Edition, Cengage.
3. **Sheldon M. Ross,** Introduction to probability and statistics Engineers and the Scientists, 4th Edition, Academic Foundation, 2011.
4. **Johannes Ledolter and Robert V. Hogg,** Applied statistics for Engineers and Physical Scientists, 3rd Edition, Pearson, 2010.
5. **T. K. V. Iyenger**, Probability and Statistics, S. Chand & Company Ltd, 2015.

e- Resources & other digital material

1. https://www.youtube.com/watch?v=COI0BUmNHT8&list=PLyqSpQzTE6M_JcleDbrVyPnE0PixKs2JE (For Probability and Statistics)
2. https://www.youtube.com/watch?v=VVYLpmKRfQ8&list=PL6C92B335BD4238AB (For Probability and Statistics)
3. https://www.mathsisfun.com/data/standard-normal-distribution-table.html (Information about Normal distribution)
4. https://www.statisticshowto.com/tables/t-distribution-table/(Information about T- distribution)

Statistical Tables to be allowed in examinations:

1. Normal distribution table
2. T- distribution table

Table CO-PO Mapping:

CO-PO mapping Matrix

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO-1	PSO-2
CO1	2	2												
CO2	2	3												
CO3	2	2												
CO4	2	2												
CO5	2	3												

Micro-Syllabus of Probability and Statistics

UNIT-I:Descriptive statistics and methods for data science: 8 hrs

Data science-Statistics Introduction-Population vs Sample-Collection of data-primary and secondary data-Types of variable: dependent and independent Categorical and Continuous variables-Data visualization-Measures of Central tendency-Measures of Variability (spread or variance)-Skewness Kurtosis.

Unit	Module	Micro content	No of hrs
1a.Descriptive Statistics	Introduction-Population vs Sample	Collection of data-primary and secondary data	2
		Population	
		Sample	
	Types of variable	dependent and	2
		independent	
		Categorical	
		Continuous variables	
	Data visualization	-Data visualization	1
1b.methods for data science	Measures of Central tendency and Measures of Variability	Measures of Central tendency	1
		Measures of Variability	2
		Skewness Kurtosis.	

UNIT-II: Correlation and Curve fitting: 10 hrs

Correlation-correlation coefficient-Rank correlation-Regression coefficient and properties-regression lines-Multiple regression-Method of least squares-Straight line-parabola-Exponential-Power curves.

Unit	Module	Micro content	No of hrs
2.Correlation and Curve fitting	Correlation	correlation coefficient	4
		Rank correlation	
	Regression	Regression coefficient	4
		properties	
		regression lines	
		Multiple regression	
	Method of least squares	Straight line	4

		Parabola.	
		Exponential curves	
		Power curves.	

UNIT-III: Probability and Distributions: 10 hrs

Probability-Conditional probability and Baye's theorem- Random variables -Discrete and Continuous random variables-Distribution function-Mathematical Expectation and Variance-Binomial, Poisson, Uniform and Normal distributions.

Unit	**Module**	**Micro content**	**No of hrs**
3. Probability and Distributions	Probability	Conditional probability	2
		Baye's theorem	
	Random variables	Discrete Random variables	1
		Continuous Random variables	1
		Distribution function	1
		Mathematical Expectation and variance	1
	Distributions	Binomial distribution.	2
		Poisson distribution	
		Uniform distribution	
		Normal distribution	

UNIT-IV: Sampling Theory: 10 hrs

Introduction–Population and samples-Sampling distribution of Means and Variance (definition only)-Central limit theorem (without proof)-Point and Interval estimations, Good estimator, Unbiased estimator, Efficiency estimator-Maximum error of estimate.

Unit	**Module**	**Micro content**	**No of hrs**
4.Sampling Theory	Introduction	Population samples	1
		Central limit theorem (without proof	
	Sampling distributions	Sampling distribution of Means	4
		Sampling distribution of Variance	
	Estimation	Point estimations	5
		Interval estimation	
		Good estimator	
		Unbiased estimator	
		Efficiency estimator	
		Maximum error of estimate.	

UNIT-V: Test of Hypothesis: **10 hrs**

Introduction–Hypothesis-Null and Alternative Hypothesis-Type I and Type II errors-Level of significance-One tail and two-tail tests-Tests concerning one mean, two means, and proportions using Z test, Tests concerning one mean, two means using t test, also chi-square and F tests use for small samples.

Unit	**Module**	**Micro content**	**No of hrs**
5. Test of Hypothesis	Hypothesis	Null Hypothesis	**2**
		Alternative Hypothesis	
		Type I and Type II errors	
		Level of significance	
		One tail and two-tail tests	
	Test for large samples	Tests concerning one mean using Z test	**4**
		Tests concerning one two means using Z test.	
		Tests concerning proportions using Z test	
	Tests for small samples	Tests concerning one mean, two means using t test	**4**
		chi-square test	
		F test	

****