Semester: 2	
Programme: Data Science	
Course: Statistical inference	
Paper code: M2C4DS26021CT / M2C4DS26021CP	**Credits: 6**
Hours/week: Theory: 4 classes per week. Practical: 2 Classes per week	
Category: Core/MDC/SEC/VAC: Core	
Theory / Practical / Composite: Composite	
No of Module: 1	

Course Outcome:

Remember the core concepts, definitions, and paradigms of classical and Bayesian inference, parametric and non-parametric frameworks, and estimation criteria
Understand the theoretical foundations of point estimation, interval estimation, hypothesis testing, Type I/II errors, power functions, and linear model assumptions.
Apply parametric tests (Z, t, F, Chi sq.), non-parametric tests (Sign, Kolmogorov-Smirnov, Kruskal-Wallis, Siegel-Tukey), and resampling methods (Jackknife, Bootstrap, Monte Carlo) to solve inference problems.
Analyze structured data using Gauss-Markov linear models, ANOVA, ANCOVA, MANOVA, and regression analysis to test parameters and assess estimability.
Evaluate estimator performance with respect to unbiasedness, consistency, efficiency, and sufficiency, as well as model adequacy and test validity
Create statistical workflows in R to execute complete data analysis pipelines, from hypothesis formulation and estimation to complex modelling.

SYLLABUS

UNIT	CONTENT	HOURS or NUMBER OF CLASSES	CO Mapping	COGNITIVE LEVEL
1	**Unit 1: Elements of Statistical Inference** Problems and paradigms of Inference - Classical and Bayesian inference. Parametric and non-parametric inference. Estimation and Testing of hypothesis.	**4**	CO1	K1

2	***Unit 2: Estimation and Testing*** Criteria of a good estimator – sufficiency, unbiasedness, consistency, efficiency (concepts only). Minimum variance unbiased estimator (concept only). Methods of estimation – method of moments, method of maximum likelihood. Null and alternative hypothesis. Simple and composite hypothesis. Type I and Type II errors. Level and power of a test. Power function. P-value of a test. Parametric tests using Z, t (one-sample and two sample), F and $\chi 2$ tests. Large sample tests. Confidence interval and coverage. Confidence interval for mean, variance, proportion and their differences using pivots. Application of Monte Carlo methods in finding bias, mean squared error, level, power, coverage. Non-Parametric Tests: Pearsonian $\chi 2$ tests. Sign test (univariate and bivariate). Kolmogorov Smrinoff test, Kruskal–Wallis test, Siegel–Tukey test.	**24**	C01, C02, C05, C06	K1, K2, K5, K6
3	***Unit 3: Resampling Techniques*** Concept of jackknife and bootstrap. Bootstrap variance and confidence intervals. Cross validation studies. Application of Monte Carlo methods.	**10**	C04, C06	K4, K6
4	***Unit 4: Linear Models*** Gauss Markov linear model and assumptions. Normal equations and solvability. Linear parametric functions and estimability. BLUE. Fundamental theorem of least squares (statement only). ANOVA, ANCOVA and regression models with related testing problems. MANOVA.	**14**	C04, CO6	CO4, CO6
Practical will be based on the Theory content				
Total Practical Hours: 26 Hours (13 Labs × 4 Hours = 52 Hours).				
Text Books				
1. Casella, G., Berger, R.(2002): Statistical Inference. (2nd Ed.), Duxbury.				
2. Dalgard,P.(2008): Introductory Statistics with R,(2nd Ed.), Springer.				
3. Goon,A.M., Gupta,M.K., Dasgupta, B.(2002): Fundamentals of Statistics Vol 1 & 2(8th Ed.), Kolkata: The World Press.				

<table>
<tr><td colspan="3">4. Lehman,E.L., Casella,G.(1998): Theory of Point Estimation,(2nd Ed.), Springer.</td></tr>
<tr><td colspan="3">5. James,G., Witten,D., Hastie,T.,Tibshirani,R.(2021): An Introduction to Statistical Learning with Applications in R (2nd ed.),Springer.</td></tr>
<tr><td colspan="3">6. Efron, B., & Tibshirani, R. (1993): An Introduction to the Bootstrap, (1st Ed.), Champan and Hall.</td></tr>
<tr><td colspan="3">7. MIT-OpenCourseWare-Recorded lectures
https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/video_galleries/lecture-videos/</td></tr>
<tr><td colspan="3">8. MOOC-Coursera
https://www.coursera.org/learn/inferential-statistics</td></tr>
<tr><td>Evaluation</td><td colspan="2"></td></tr>
<tr><td colspan="3">Paper structure for theory semester exam</td></tr>
<tr><td>Marks</td><td>CIA:
Class Performance: 2
Attendance: 3
Mid-term Exam: 10
Total: 15
End Semester Exam: 45</td><td>Continuous Assessment: 40</td></tr>
<tr><td rowspan="2">Paper Structure for Theory Semester Exam</td><td>Short Questions: 5 marks each</td><td>Long Questions: 10 marks each</td></tr>
<tr><td>3 out of 5</td><td>3 out of 5</td></tr>
</table>

Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	**Remember** the core concepts, definitions, and paradigms of classical and Bayesian inference, parametric and non-parametric frameworks, and estimation criteria	K1
CO2	**Understand** the theoretical foundations of point estimation, interval estimation, hypothesis testing, Type I/II errors, power functions, and linear model assumptions.	K2
CO3	**Apply** parametric tests (Z, t, F, Chi sq.), non-parametric tests (Sign, Kolmogorov-Smirnov, Kruskal-Wallis, Siegel-Tukey), and resampling methods (Jackknife, Bootstrap, Monte Carlo) to solve inference problems.	K3
CO4	**Analyze** structured data using Gauss-Markov linear models, ANOVA, ANCOVA, MANOVA, and regression analysis to test parameters and assess estimability.	K4
CO5	**Evaluate** estimator performance with respect to unbiasedness, consistency, efficiency, and sufficiency, as well as model adequacy and test validity	K5
CO6	**Create** statistical workflows in R to execute complete data analysis pipelines, from hypothesis formulation and estimation to complex modelling.	K6