GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering
Level: UG
Branch: Biotechnology
Subject Code: BE03004051
Subject Name: Biostatistics

w. e. f. Academic Year:	**2024-25**
Semester:	**3**
Category of the Course:	**Basic Science Course**

Prerequisite:	Basic mathematics (algebra and probability) Understanding of biological concepts and experimental design Familiarity with data types and measurements Basic computer skills (e.g., Excel, statistical software like R or SPSS)
Rationale:	Biostatistics enables objective analysis and interpretation of biological data. It is essential for validating experimental results, designing robust studies, analyzing trends, and making evidence-based conclusions in research. In biochemical processes, it aids in process optimization, quality control, clinical trials, and data-driven decision-making in biotechnology and public health..

Course Outcomes:

Sr. No.	CO statement	Marks% weightage
CO-1	Understand the mathematical basis and foundations of probability and statistics	15
CO-2	Apply statistical methods to solve biological problems	25
CO-3	Apply the statistical principles to design the research experiments	25
CO-4	Apply basic and modern statistical methods to analyze the big data in biology and clinical data	15
CO-5	Interpret statistical results and communicate findings effectively in the context of life science research.	20

Teaching and Examination Scheme:

Teaching / Learning Scheme (in Hours per semester)					Total Credits	Assessment Pattern and Marks					Total Marks
						Theory		Tutorial / Practical			
L	T	P	PBL*	Total no of hours per semester		ESE (E)	PA / CA (M)	PA/CA (I)	PBL (I)	ESE (V)	
45	0	30	45	120	4	70	30	20	30	50	200

*** Problem Based Learning (PBL) aims to accommodate learning beyond syllabus as per clause 9.4 of NBA manual.**

Content:

Sr. No.	Content	Total Hrs
1	**Introduction to Biostatistics and Organization of Data:** Fundamentals of biostatistics. Types of biological data (qualitative and quantitative). Graphical and pictorial data representation: bar charts, pie charts, histograms, scatter plots. Measures of central tendency (mean, median, mode) and dispersion (range, variance, standard deviation). Sampling methods and sample size determination. Calculation of coefficient of variation, mean error, relative error, concepts of precision, accuracy, and bias in measurements.	10
2	**Probability Distribution and Statistical Inference:** Introduction to probability theory and Bayes' theorem. Probability distributions: binomial, Poisson, and normal distributions, their properties, and applications in biological data analysis. Introduction to parametric and non-parametric hypothesis testing. Testing of hypotheses: types I and II errors, p-values, and statistical power. Tests of significance: z-test, t-test (single, paired, two-sample), F-test. Nonparametric procedures: chi-square test, sign test, Wilcoxon signed-rank test, Wilcoxon rank-sum (Mann-Whitney U) test. Analysis of Variance (ANOVA): one-way and two-way ANOVA. Correlation and regression analysis: simple and multiple regression models, interpretation of regression coefficients. Tests of significance for correlation coefficients. Introduction to multivariate statistical methods (PCA, cluster analysis).	10
3	**Experimental Design :** Principles of experimental design: randomization, replication, and blocking. Advanced techniques: Double-blind and double-dummy procedures for clinical and biological experiments. Design structures: Completely randomized design (CRD), randomized block design (RBD), Latin square design, factorial experiments (full and fractional designs), crossover designs, and parallel group designs. Optimization techniques and response surface methodology (RSM) for engineering and biological process design.	10
4	**Statistics in Data Analytics:** Application of statistical methods in biological and biomedical data analysis. Introduction to Big Data analytics in biological sciences and healthcare. Data analytics lifecycle: Discovery, data preparation, model planning, model building, evaluation and validation, result communication, and operationalization. Introduction to supervised and unsupervised learning for biological data: decision trees, logistic regression, clustering. Fundamentals of machine learning model validation: cross-validation, train/test split, confusion matrix, ROC curves. Overview of bioinformatics data analytics: handling omics data (genomics, proteomics, metabolomics) using statistical methods. Ethical considerations in biological data analytics and reporting standards (e.g., CONSORT, STROBE guidelines).	15
	TOTAL	**45**

Suggested Specification table with Marks (Theory): (For B.E. only)

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
20	**20**	**15**	**25**	**10**	**10**

R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Reference Books:

1. Daniel Wayne W., Biostatistics: A Foundation for Analysis in the Health Sciences, John Wiley & Sons, 2008, 9th Edition
2. Rosner Bernard, Fundamentals of Biostatistics, Brooks/Cole, 2011, 7th Edition
3. Motulsky H, Intuitive Biostatistics, Oxford University Press, 2009, 2nd Edition
4. Data Science and BIg Data Analytics. Discovery, Analyzing, Visualizing and Presenting Data, EMC Education Services, John Wiley & Sons, 2017

List of Experiments:

1. Collect and organize biological data (e.g., plant heights, cell counts) and represent it using bar charts, histograms, and scatter plots.
2. Calculate measures of central tendency (mean, median, mode) and dispersion (range, variance, standard deviation) for a given biological dataset.
3. Determine appropriate sampling methods and calculate required sample sizes for a hypothetical biological study.
4. Apply probability distributions (binomial, Poisson, normal) to analyze and interpret simple biological scenarios.
5. Conduct z-tests and t-tests (single, paired, two-sample) to compare means in various biological datasets.
6. Perform Chi-square tests for association and goodness-of-fit on categorical biological data.
7. Apply Wilcoxon signed-rank and Mann-Whitney U tests for non-parametric comparisons of biological data.
8. Conduct one-way and two-way ANOVA to analyze the effects of factors on biological outcomes.
9. Perform simple and multiple linear regression analysis on biological data, interpreting correlation coefficients and model parameters.
10. Design a completely randomized experiment (CRD) and a randomized block design (RBD) for a biological study.
11. Apply principles of randomization, replication, and blocking to a given experimental scenario.
12. Utilize statistical software to visualize and analyze omics data (e.g., genomics, proteomics).

Major Equipment: Nil

List of Open Source Software/learning website:

- R: Statistical programming language for all biostatistics tasks.
- RStudio: IDE for user-friendly R programming.
- Python: General-purpose language with libraries for data analysis (Pandas, NumPy, SciPy), visualization (Matplotlib, Seaborn), machine learning (Scikit-learn), and bioinformatics (Biopython).
- JASP: User-friendly GUI statistical software.
- Jamovi: Another GUI statistical software, built on R.
- GNU Octave: MATLAB-compatible numerical computation language.
- Cytoscape: For visualizing biological networks.
- ImageJ / Fiji: For quantitative biological image analysis.

Free Learning Websites for Microbiology

- Coursera: Online courses on Biostatistics, Bioinformatics, and Data Science.
- edX: Similar to Coursera, offering courses from top universities.
- NPTEL / Swayam: Free online courses from IITs on Biostatistics and Data Science.
- YouTube (University Channels): Lectures and tutorials on biostatistics and bioinformatics.
- Codecademy / freeCodeCamp: Platforms for learning programming languages like Python and R.
- Kaggle: Data science platform with datasets, notebooks, and tutorials.
- GitHub: Repository for open-source code and bioinformatics projects.
- Stack Overflow / Biostars / Bioinformatics Stack Exchange: Q&A platforms for programming and bioinformatics.
- Towards Data Science / Medium: Blogs with articles and code examples on data science and biostatistics.
- The Carpentries: Workshops on foundational computational and data science skills.

*** List of suggested activities for Problem Based Learning:**

- Collect and Organize Data: Gather a real biological dataset (e.g., plant growth measurements, cell counts, patient characteristics) and structure it effectively in a spreadsheet or database.
- Visualize Data: Create various plots like bar charts, histograms, scatter plots, and box plots to represent your collected biological data using statistical software.
- Calculate Descriptive Statistics: Compute and interpret measures of central tendency (mean, median, mode) and dispersion (variance, standard deviation, coefficient of variation) for different biological datasets.
- Apply Sampling Methods: Design a practical sampling strategy for a hypothetical biological experiment, justifying your choice of method and sample size.
- Solve Probability Problems: Tackle problems involving binomial, Poisson, and normal distributions as they apply to biological scenarios (e.g., genetic crosses, bacterial colony counts, physiological measurements).
- Perform Parametric Hypothesis Tests: Conduct z-tests, t-tests (single, paired, two-sample), and F-tests on biological data to assess statistical significance.

- Conduct Non-parametric Hypothesis Tests: Utilize chi-square tests, Wilcoxon signed-rank tests, and Mann-Whitney U tests for analyzing categorical or non-normally distributed biological data.
- Apply ANOVA: Perform one-way and two-way ANOVA to analyze data from experiments with multiple groups and interpret interaction effects.
- Analyze Correlation & Regression: Conduct simple and multiple linear regression on biological data, interpreting correlation coefficients, R-squared values, and regression parameters.
- Practice Experimental Design: Design and critically evaluate different experimental layouts like Completely Randomized Designs (CRD), Randomized Block Designs (RBD), and Factorial Designs for biological studies, emphasizing randomization and replication.
- Explore Response Surface Methodology (RSM): Analyze a case study involving biological process optimization using RSM concepts, perhaps with simulated or simplified data.
- Master Statistical Software: Become proficient in an open-source statistical software (R, Python, JASP, or Jamovi) by performing a range of statistical analyses from your syllabus.
- Undertake a Data Analytics Project: Work on a small project using a biological dataset (e.g., gene expression, public health data) to apply supervised or unsupervised machine learning techniques.
- Explore Omics Data: Use bioinformatics tools (e.g., Bioconductor packages, Galaxy) to perform basic exploration and visualization of publicly available omics datasets.
- Review Ethical Reporting: Critically assess a published biological research paper to identify how it adheres to (or deviates from) ethical reporting standards like CONSORT or STROBE guidelines.
- Simulate Statistical Consulting: Formulate a response to a hypothetical biological research question, outlining the most appropriate statistical test(s) and experimental design.

****