**Approved in 38th BoA Meeting (22-01-2021)**

Course Number: BE303
Course Name: Applied Biostatistics
Credits: 3-0-2-4
Prerequisites: IC252, IC272
Intended for: B. Tech M.Tech Integrated Dual Degree Bioengineering students
Distribution: Core for Integrated Dual Degree Bioengineering students, elective for other B.Tech students
Semester: $4^{th}$ semester

---

**Preamble:** Probability theory and statistical analysis are fundamental concepts in medical and biological sciences due to the random processes involved. To form new knowledge and to test hypotheses in these sciences, this random nature renders experiments imperative, which are based on a set of trials which form a composed experiment. The desired function of novel medical and biological techniques and instruments can only be verified by experimental analysis in combination with statistical tools to evaluate the proper functioning. The foundations laid in this course provide the necessary vocabulary and facilitates the understanding of fundamental core concepts in medical and biological statistical analysis and experimentation.

**Objectives:** This course is intended to cover fundamental concepts in biostatistics, data science in biology, and how to apply these concepts using the R statistical programming language. In general, this course will emphasize applied statistical theory on biological data analysis. The student should be able to understand fundamental terms (e.g., incidence, prevalence), and to evaluate different types of medical studies and data collection, for example to be aware of importance of sample size determination. Important aspects of medical tests (e.g., sensitivity and specificity) and its impact on device development will be understood by the student. After the course the student can apply concepts of various statistical tests and understands concepts of multivariate analysis. Students are aware of problems and fallacies of statistical analysis.

**Module 1- Study design, data acquisition, and presentation**:

Recognize and give examples of different types of data arising in public health and clinical studies. Types of medical studies and introduction of different study designs (descriptive vs analytical, control groups etc.) and sampling methods (randomization). Risk studies (descriptive, case-control and cohort studies) Measure of important probabilities (incidence, prevalence, sensitivity-specificity-predictivity, morbidity, etc.). Example of study design at the example of a clinical trial for approval process. (10 hours)

**Module 2- Null hypothesis, Statistical testing:**

Why are there statistical tests in medial and biological sciences? Formulation and examples of null hypothesis for medical and bioengineering applications, subsequent decisions, and error types I + II.
Overview of tests: location (e.g., t-test), dispersion (e.g., ANOVA (f-test)), comparison of frequencies (chi-square), variance analysis (e.g., inter- and intra-class variance) and regression analysis (linear and logistic regression (ROC-analysis)). *Parametric tests* are used

only where a normal distribution is assumed. These are the t-test (paired or unpaired), ANOVA (one-way non-repeated, repeated; two-way, three-way), linear regression and Pearson correlation.
*Non-parametric tests* are used when continuous data are not normally distributed or when dealing with discrete variables. These are chi-squared, Fisher's exact tests, Wilcoxon's matched pairs, Mann–Whitney U-tests, Kruskal–Wallis tests and Spearman rank correlation. Comparison of different tests. Analyze required sample size (calculate the power of a test) and analysis of correct application of a test using specific examples. (15 hours)

**Module 3: Advanced tools – Introduction to multivariate analysis**

Multivariate analysis is concerned with the interrelationships among several variables. A specific example for the motivation of multivariate analysis will be given. The course includes the following methods: cluster analysis, principal components analysis, factor analysis, discriminant analysis, etc. Designs of Experiments which will include, one-way, two-way ANOVA, MANOVA study design, confounding and standardization. Difference between multiple linear regression and multivariate regression (e.g., using an example of the Framingham heart study) (7 hours)

**Module 4: Applications of biostatistics, quality, and potential fallacies**

Introduction to survival analysis, estimation of survival curves, and proportional hazards model (e.g., example of life expectancy under different conditions). Analysis of different parametric survival functions.

Quality considerations: Quality of statistical models and quality of data. Dealing with erroneous data, missing values, bias of observer etc.

Demonstration of statistical fallacies due to data (biased sample, inadequate sample size, incomparable objects), erroneous analysis (linear analysis of nonlinearities, biased data selection, misuse of p-values, etc.) and errors or misuse of presentations (misuse of percentages or wrong base for percentages, misuse of means etc. Misuse of graphical representation. (10 hours)

**Lab: Tentative list of experiments:**

1. Introduction to R using RStudio
2. Analysis of data used in public health with real life examples
3. T-test with biological data set
4. ANOVA test with biological data set
5. Application of Wilcoxon's Signed Rank test
6. Application of Mann Whitney U test
7. Kruskal-Wallis-test
8. Meta-analysis of the disease data
9. Equivalence study
10. Cross sectional study
11. Regression analysis
12. Multivariate Methods
13. Categorical data analysis

**Textbook:**

1. Abhaya Indrayan and Rajeev Kumar Malhotra, Medical biostatistics, Chapman and Hall, 2017.

**Reference book:**


1. Fundamentals of Biostatistics, 8th Edition. Bernard Rosner. Cengage Learning Inc., 2015
2. Biostatistics and Epidemiology: A Primer for Health and Biomedical Professionals, Wassertheil-Smoler, Springer-Verlag 2014.
3. Design and Analysis of Clinical Trials: Concepts and Methodologies, Shein-Chung Chow, Jen-Pei Liu, 3rd edition, Wiley, 2016.


1. **Similarity content declaration with existing courses:**

| Sl. No. | Course Code | Similarity Content | Approx. % of content |
|---|---|---|---|
| 1 | IC252 | Null hypothesis, type I+II error | 7 % |
| 2 | IC272 | Regression analysis | 5 % |
| 3 | HS550 | Regression analysis, ANOVA | 50% - 60% |

2. **Justification of new course proposal if cumulative similarity content is >30%:**
HS550 and BE303 are mutually exclusive.