

Subject Code: 02MB0456
Subject Name: Biostatistics and Bioinformatics
M. Sc. Semester - II

Objective: Students are expected to have the advanced learning of Biostatistics and Bioinformatics which will enable them to apply these concepts in day to day life. As the name suggests, the course is divided into two parts: Biostatistics (the use of statistics to interpret biological data) and Bioinformatics (the use of informatics system to extract biological information and interpret results). The course also discusses application of several web servers that can be routinely used for various microbiological applications.

Credits Earned: 5 Credits

Course Outcomes: After completion of this course, student will be able to

- Recognize importance of Biostatistics in interpreting the biological data and design suitable experiments and understand the ways to report the results in a scientific way.
- Understand the errors obtained between different sets of experiments and calculate it precisely.
- Comprehend the ways to utilize informatics system to derive useful biological information.
- Use Bioinformatic tools to analyze different protein or nucleotide sequences to reach meaningful conclusions.

Pre-requisite of course: NA.

Teaching and Examination Scheme

Teaching Scheme (Hours)			Credits	Theory Marks			Tutorial/ Practical Marks		Total Marks
Theory	Tutorial	Practical		ESE (E)	Mid Sem (M)	Internal (I)	Viva (V)	Practicals (P)	
5	0	2	6	50	30	20	25	25	150

Contents:

Unit	Topics	Contact Hours
1	Quantitative methods in biology, sampling methods, Sampling errors Data collection, data organization and graphical representation.	10
2	Measures of Central Tendency and Dispersion, Probability distribution (Binomial and Normal), Hypothesis testing. Statistical Tests: Z-test, T-test, ANOVA, chi squared test.	15
3	History of Bioinformatics, Study, Scope and Applications. Branches emerged from Bioinformatics. Study of Databases of Nucleic Acids and Proteins: Primary, Secondary and composite. Submission and Retrieval of entry from database. Protein structure visualization and prediction: Tools and Methodology.	17
4	Study of Sequence Alignment Methods for searching homologous sequences: Local alignment (Smith and Watermann) and Global Alignment (Needleman and Wunch) and their use in Pairwise alignment (BLAST and FASTA) and Multiple sequence alignment(MSA). Application of MSA in phylogenetics: Building phylogeny (Principle and Procedure for taxonomic analysis of microorganisms) and Interpretation using different graphical forms (Rooted, Unrooted, Cladogram, Dendogram, Phylogram, Phenogram).	18
	Total Hours	60

References:

1. *Statistical Methods*, Gupta SP. Sultanchand & Sons.
2. *Fundamentals of Statistics*, Goon, Gupta and Dasgupta –World Press, Kolkata.
3. *Bioinformatics: A practical guide to the analysis of genes and proteins*. (2001) 2nd Edition, Baxevanis AD and Ouellette BFF. John Wiley & Sons, New York.
4. *Bioinformatics: Sequence and Genome Analysis*, 2nd Edition (2001), David W. Mount, Cold Spring Harbor Laboratory Press.

Suggested Theory distribution:

The suggested theory distribution as per Bloom's taxonomy is as per follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process.

Distribution of Theory for course delivery and evaluation					
Remember	Understand	Apply	Analyze	Evaluate	Create
10%	20%	20%	20%	25%	5%

Instructional Method:

- d. The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, etc.
- e. The internal evaluation will be done on the basis of continuous evaluation of students in the class-room in the form of attendance, assignments, verbal interactions etc.
- f. Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory.

List of Experiments**Sr. No. Experiments**

- 1 To study Major Nucleic Acid Databases NCBI, DDBJ and EMBL and their File formats
- 2 To study EST, UniSTS, SNP databases of Nucleic Acids.
- 3 To study Uniprot and PDB databases of Proteins.
- 4 To study submission protocol for Genbank and ESTs.
- 5 To perform analysis of pairwise alignment using NCBI-BLAST.
- 6 To perform analysis of alignment of multiple sequences using ClustalOmega/Phylip.
- 7 To Study Protein Structural Classification database: SCOP and CATH.
- 8 To study protein structure visualisation using RasMol.
- 9 To analyse protein structure using Expasy tools.
- 10 To study proteomics using various tools of Expasy.
- 11 To study Protein structure prediction using Expasy tools.
- 12 To study metabolic pathways using KEGG.