

## BSH452 BIOSTATISTICS AND BIOINFORMATICS

### Course Outcomes:

*After successful completion of the course, students will be able to :*

- CO 1. Apply statistical methods to analyse and interpret the biological data.
- CO 2. Represent the data of experimental and field studies through graphs and diagrams.
- CO 3. Understand statistical concepts and learn to use a variety of statistical tests
- CO 4. Know how to use bioinformatics for DNA and protein sequence analysis through bioinformatics tools and databases
- CO 5. Understand microarray technique for gene expression analysis.
- CO 6. Understand the concept of protein folding and structure based targeted drug design

### Unit I (13 hours)

**Biological data**-frequency distribution, graphical and diagrammatic representations; Measures of Central tendency - Mean, Median and Mode; Measure of Dispersion - Range, Variance, Standard deviation, Coefficient of variation, Diversity Index.

Populations versus sample - sampling techniques; Standard error, Confidence limits. Random experiment-probability. Binomial, poisson and Normal distributions and their applications in genetics.

### Unit II (13 hours)

Simple linear Regression and Correlation analysis. Analysis of variance, principles of experimental design. Multiple regression.

**Tests of significance**- Normal,  $X^2$ , (Chi-square), 't' and F tests; Testing for goodness of fit. One-way analysis of variance (ANOVA) and Two-way analysis of variance. Statistical packages.

### Unit III (13 hours)

Introduction to bioinformatics, databases, search engines, internet tools and World Wide Web (WWW). Molecular modeling database at NCBI, major web resources for bioinformatics - Biological database types and their functioning, microbiological databases, primary sequence databases, carbohydrate databases, RNA databases, genome databases, organism databases, biodiversity. **Sequence database**: Introduction, nucleotide sequence database, protein sequence databases, EMBL nucleotide sequence databases, structure databases. Phylogeny - Tree definitions, distance matrix methods and parsimony and bootstrapping. **DNA and protein sequence Analysis**, FASTA, BLAST and GCG Wisconsin/Emboss packages. Genomics and proteomics.

### Unit IV (13 hours)

Microarray techniques, Gene Expression analysis, Protein Folding, Lattice models, Comparative modeling, threading, folds and function, Distributed Computing approach, genome@home, folding@home, proteomics, protein structure based targeted drug design – small molecular interactions and docking.