Introduction to the course

BT307   **Biological Data Analysis**   **(2-0-2-6)**

Syllabus

Data, descriptive statistics, and visualization: Introduction to different types of data in biology; Descriptive statistics like mean, median, mode, quartiles, standard deviation, standard error; Different types of plots like scatter plot, bar graph, line graph, pie chart, box plot, frequency histogram; Understanding error bars. Probability and probability distributions: basic concepts of probability, conditional probability, Bayes theorem; binomial, multinomial, Poisson, exponential, and Gaussian distribution; Sampling distribution and central limit theorem. Hypothesis testing: Student's t-test, Z-test, Chi-squared test, ANOVA.

Correlation, regression and estimation: Pearson correlation; Regression: linear, non-linear, single and multivariate; concept of likelihood and method of maximum likelihood. Tools for data of high throughput experiments: principle component analysis; Clustering of data: K-means algorithm, hierarchical clustering; Visualization tools: heat map, volcano plot. Laboratory component: R and MS Excel based exercises on graphical visualization of data, different tests of hypothesis, estimation of correlation, regression, PCA, clustering.

Text Books/References

1. G. P. Quinn, and M. J. Keough, *Experimental Design and Data Analysis for Biologists*, 1st Edition, Cambridge University Press, 2002.

2. R. C. Elston and W. D. Johnson, *Basic Biostatistics for Geneticists and Epidemiologists: A Practical Approach*, 1st Edition, Wiley, 2008.

3. G. Hartvigsen, *A Primer in Biological Data Analysis and Visualization Using R*, 1st Edition, Columbia University Press, 2014.

4. M. C. Whitlock, and D. Schluter, *The Analysis of Biological Data*, 2nd Edition, W. H. Freeman & Company, 2014.

5. S. Ross, *A First Course in Probability*, 9th Edition, Pearson Education India, 2014.

6. M. D. Ugarte, A. F. Militino, and A. T. Arnholt, *Probability and Statistics with R*, 2nd Edition, CRC Press, 2016.

Course Instructors

Prof. Shankar Prasad Kanaujia

First half: 4th January – 23 February 2024

Prof. Biplab Bose

Second half: 4th March – 26th April 2024

Tentative distribution of classes

S. No.	Date/ Day	Time	Topic to be covered
1	04.01.2024 (Thursday)	10:00 – 12:00 PM	Introduction to the course
2	05.01.2024 (Friday)	4:00 – 5:00 PM	Types of data in biology
3	08.01.2024 (Monday)	3:00 – 4:00 PM	,,
4	11.01.2024 (Thursday)	10:00 – 12:00 PM	Descriptive statistics
5	12.01.2024 (Friday)	4:00 – 5:00 PM	,,
6	15.01.2024 (Monday)	3:00 – 4:00 PM	Different types of plots
7	18.01.2024 (Thursday)	3:00 – 4:00 PM	,, (Monday time table)
8	19.01.2024 (Friday)	4:00 – 5:00 PM	,,
9	22.01.2024 (Monday)	3:00 – 4:00 PM	Basic concepts of probability
10	25.01.2024 (Thursday)	10:00 – 12:00 PM	Lab 1 (types of data)
11	29.01.2024 (Monday)	3:00 – 4:00 PM	Conditional probability
12	01.02.2024 (Thursday)	10:00 – 12:00 PM	Lab 2 (descriptive statistics)

Tentative distribution of classes

S. No.	Date/ Day	Time	Topic to be covered
13	02.02.2024 (Friday)	4:00 – 5:00 PM	Bayes theorem
14	03.02.2024 (Saturday)	3:00 – 4:00 PM	Quiz
15	05.02.2024 (Monday)	3:00 – 4:00 PM	Binomial distribution
16	08.02.2024 (Thursday)	10:00 – 12:00 PM	Lab 3 (types of plots)
17	09.02.2024 (Friday)	4:00 – 5:00 PM	Multinomial distribution
18	12.02.2024 (Monday)	3.00 – 4:00 PM	Poisson distribution
19	15.02.2024 (Thursday)	10:00 – 12:00 PM	Lab 4 (probability distribution function)
20	16.02.2024 (Friday)	4:00 – 5:00 PM	Central limit theorem
21	19.02.2024 (Monday)	3:00 – 4:00 PM	Student's t-test, Z-test, Chi-square test
22	22.02.2024 (Thursday)	10:00 – 12:00 PM	Lab 5 (probability distribution function)
23	23.02.2024 (Friday)	4:00 – 5:00 PM	ANOVA
24	24.02.2024 – 01.03. 2024	XXX	Mid semester examination

Evaluation scheme

Weightage: 50% each (first and second halves)

Marks distribution (1st Half):
Quiz: 1x10 marks = 10 marks
Lab assignments: 5 x 1 mark = 5 marks
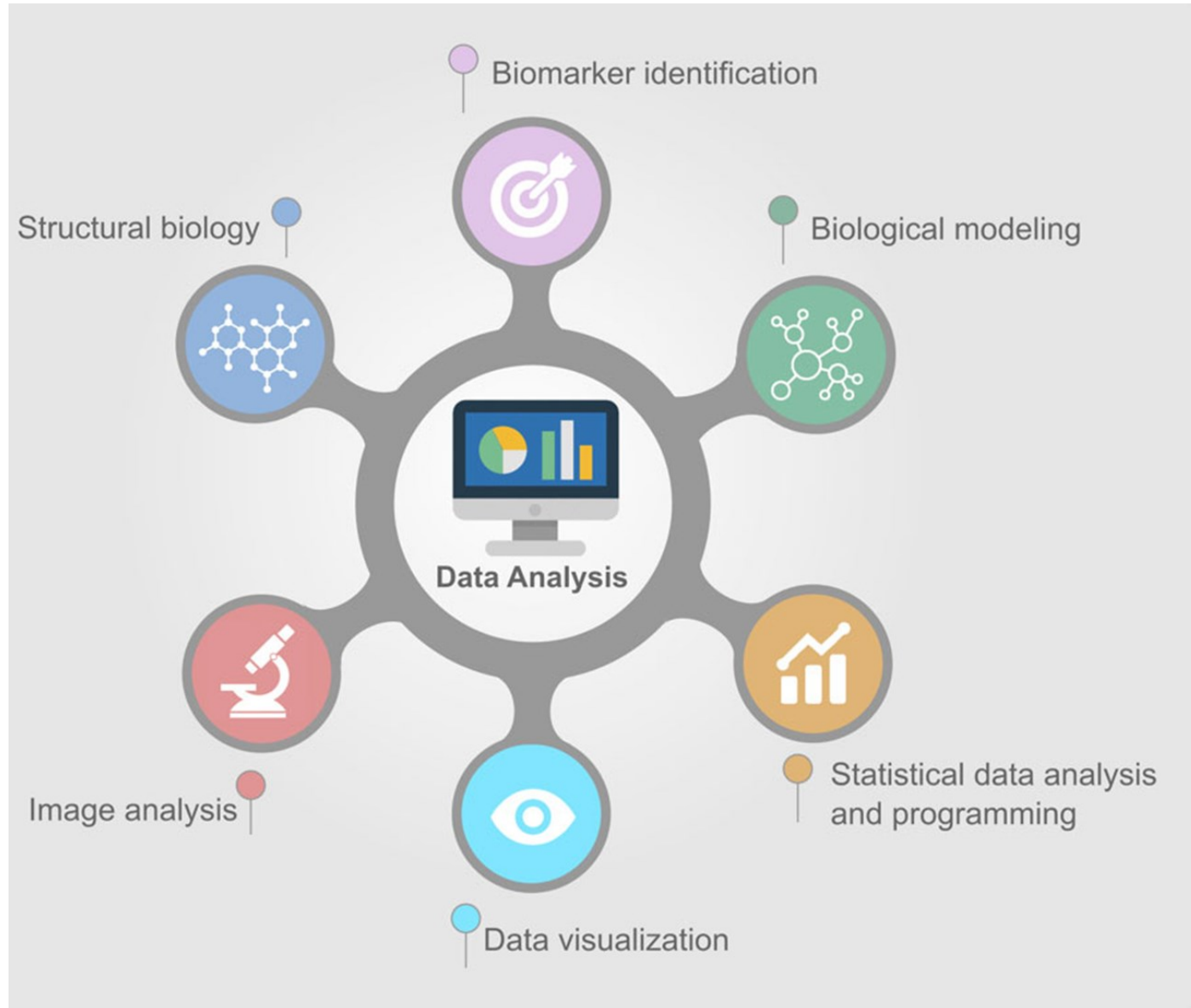Mid semester examination: 1 x 30 marks = 30 marks
Attendance: 5 marks
Total marks: 50 marks

Lecture notes availability

MOODLE

https://www.iitg.ac.in/moodle/login/index.php

(bt307_iitg2024)

Installing R and R Studio

For R, go to the following link

www.r-project.org

For Rstudio, go the following link

www.rstudio.com

Thank You