



45840



Library
Department of Allied Health Sciences
Brainware University
Barasat, Kolkata- 700125

BRAINWARE UNIVERSITY

Term End Examination 2025-2026

Programme – M.Sc.(BI)-2025

Course Name – R Programming

Course Code - MBI27502A (T)

(Semester II)

Full Marks : 40

Time : 2:0 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 10=10

1. Choose the correct alternative from the following :

- (i) Select the basic unit of gene expression data stored in R.
 - a) RGB pixel values
 - b) Expression values per gene per sample
 - c) Protein sequence characters only
 - d) Folder names
- (ii) Select the representation of missing values commonly used in R.
 - a) Null
 - b) Zero
 - c) NA
 - d) Blank space
- (iii) Select the risk of inappropriate imputation.
 - a) Faster file upload
 - b) Introduction of bias
 - c) Increased memory size
 - d) More colour options
- (iv) Choose the primary measure indicating the central location of a dataset.
 - a) Mean
 - b) Range
 - c) Standard deviation
 - d) Variance
- (v) Choose the measure sensitive to extreme values.
 - a) Median
 - b) Mode
 - c) Mean
 - d) Interquartile range
- (vi) Choose the base function used to visualize summary of quartiles and outliers.
 - a) barplot()
 - b) boxplot()
 - c) pairs()
 - d) contour()
- (vii) Choose the key feature of ggplot2 as a visualization system.
 - a) Command-line only
 - b) Grammar of graphics
 - c) Supports only bar plots
 - d) No layering possible
- (viii) Select the type of data produced in early stages of sequencing pipelines.
 - a) Images
 - b) Raw sequence reads
 - c) Weather data
 - d) Staff photos
- (ix) Select a standard downstream process after alignment.

- a) Pizza ordering
- b) Variant calling
- c) Manual scoring
- d) Random sorting
- (x) Select a practical benefit of automated reporting in pipelines.
- a) More guesswork
- b) Efficient result summarization
- c) Unexplained data loss
- d) Manual redrawing of graphs

Group-B

(Short Answer Type Questions)

3 x 5=15

2. Describe numeric, character, and logical data types in R and their typical uses in biological data analysis. (3)
3. Explain how scatter plots can be used to compare expression levels of two genes across a set of samples and identify correlations. (3)
4. Discuss the importance of identifying missing values in biological datasets before any analysis. (3)
5. Explain the concept of mode and its role in identifying the most common biological category or trait in a dataset. (3)
6. Summarize the steps one would take to preprocess raw sequencing reads in a practical pipeline. (3)

OR

Summarize how one would visualize gene expression results using R or Python in a bioinformatics pipeline. (3)

Group-C

(Long Answer Type Questions)

5 x 3=15

7. Discuss how filter() can be applied to remove samples with missing "Weight_g" values in a plant growth dataset and provide the R code. (5)
8. Explain how to use faceting with facet_wrap() or facet_grid() in ggplot2 to display subplots for different levels of a categorical variable. (5)
9. Summarize how a bioinformatics analysis pipeline ensures reproducibility and efficiency in data processing. (5)

OR

Summarize the role of quality control and preprocessing in a bioinformatics pipeline. (5)

Library
Department of Allied Health Sciences
Brainware University
Barasat, Kolkata- 700125