# PRESIDENCY UNIVERSITY

## M.Sc. (STATISTICS): Semester 3 (2024)

### Course: APPLIED MULTIVARIATE ANALYSIS AND RESAMPLING (STAT0992)

#### PRACTICAL PROBLEM SET - II

*Date: Friday, August 16, 2024*

- *Use R to solve the problems.*
- *All relevant R programming code should be submitted for evaluation, together with a properly formulated report* **(as a pdf document),** *containing intermediate and final outputs (including plots, if any),* **with explanation wherever required, and comments if asked for.**
- *Deadline for submission:* **Friday, August 23, 2024 (11.59 p.m.)**
- *The filename must contain your name AND roll number.*
- *Submission should be made only to the* **Google Classroom** *for the course*
- *Extra credit may be given for individual effort.*
- *Submissions received after the deadline will be summarily rejected.*

1. In a study pertaining to women's nutrition, daily nutrient intake was measured for a random sample of 737 women aged 25-50 years. Five nutritional components were measured: *calcium, iron, protein, vitamin A*, and *vitamin C.* From analysis of this data, the sample mean vector for daily intake was calculated. The table below shows the recommended daily intake, and the sample means for all the variables:

| Variable | Recommended Daily Intake ($\mu_0$) | Mean Intake |
|---|---|---|
| Calcium | 1000 mg | 624.0 mg |
| Iron | 15mg | 11.1 mg |
| Protein | 60g | 65.8 g |
| Vitamin A | 800 µg | 839.6 µg |
| Vitamin C | 75 mg | 78.9 mg |

   a) In the light of this data, can it be inferred that women meet the recommended nutritional intake guidelines? Conduct a statistical test of inference of size 0.01 for this purpose, and comment.

   b) If women fail to meet the guidelines, identify nutrients for which the guidelines are not met, by constructing 99% simultaneous and Bonferroni confidence intervals for the mean intake.

   c) Generate profile plots for both categories of confidence intervals.

2. A shoe company evaluates new shoe models based on five criteria: *style, comfort, stability, cushioning*, and *durability*. Each of the first four criteria is evaluated on a scale of 1 to 20 and the durability criterion is evaluated on a scale of 1 to 10.  The company is considering phasing out an existing shoe model (Model 2), replacing it with a new prototype (Model 1). The company had the same subjects evaluate both Model 1 and Model

2 and looked to see if there was a significant difference between the two models which would help them decide whether to replace Model 2 with Model 1. The data is available in sheet no. 1 of the file Practical_Set_II_Data.xlsx. Answer the following questions:
  a) Is there a significant difference between the two shoe models at the 5% level of significance?
  b) If significant difference is seen in part (a), determine for which of the individual criteria there is a significant difference, using both 95% simultaneous confidence intervals and 95% Bonferroni confidence intervals.

3. A certain disease is characterized by *fever*, low *blood pressure*, and *body aches*. A pharmaceutical company is working on a new drug to treat this type of disease and wants to determine whether the drug is effective. They take a random sample of 20 people with this type of disease and 18 with a placebo. Based on the data (given in sheet no. 2 of the file Practical_Set_II_Data.xlsx), they want to determine whether the drug is effective at reducing these three symptoms.

   Conduct an appropriate multivariate test of significance for this purpose at the 5% level, assuming that the population covariance matrices corresponding to the two populations are
   a) equal.
   b) not equal.

4. A new type of corn seed has been developed by a team of agronomists who want to determine whether there was a significant difference between the types of soils that they are planted in (*loam*, *sandy*, *salty*, *clay*) based on the *yield* of the crop, amount of *water* required and amount of *herbicide* needed. Eight fields of each type were chosen for the analysis. Based on the data in sheet no. 3 of the file Practical_Set_II_Data.xlsx,
   a) determine whether there is a significant difference at the 5% level in the effects of the types of soil on the three variables;
   b) if there are significant differences between the four levels of the factor, perform appropriate tests of significance at the 5% level to answer the following questions:

     i.   Is there a significant difference between the clay and salty groups?
     ii.  Is there a significant difference between the loam and sandy groups?
     iii. Do the loam and sandy groups differ significantly in means from the clay and salty groups? If they do, obtain 95% simultaneous as well as Bonferroni confidence intervals for the individual means.

5. A study was conducted to see the impact of social-economic class (rich, middle, poor) and gender (male, female) on kindness and optimism using a sample of 24 people based on the data in sheet no. 4 of the file Practical_Set_II_Data.xlsx. Conduct appropriate MANOVA tests for the significance of the differences among the main effects of the two factors as well as their interactions on the two variables. Comment on the outcomes of the tests.

_____