STAT 53533: R Assignments 03
Upload your answers**, as a single file,** in Blackboard by
11:59 PM on Wednesday, December 11, 2024

A. Separate your answer in two parts (within the same file) as follows:

1. Part 1 must contain complete answers for all questions **without** any R code. It must have all numbers/values/diagrams as asked in the questions and should be accompanied by necessary explanations.

2. Part 2 contains only the R code and nothing else. **Any output/numbers/explanations written inside the R Code will NOT be graded**.

3. **You must type all your answers and explanations for submission, as a professional report.** Scanned copy of any **handwritten answers/explanations will not be graded.**

Consider the factor application dataset. The dataset consists of measurements on a large number of variables collected from a relatively small number of observations. Each row represents data from one observation. Each column represents data for one specific variable across all observations. Now, answer the following questions. [1+2+3+3+1=10 points]

(i) For this dataset, report n and p.

(ii) Use Parallel Analysis to choose the number of factors k using 95% quantile as summary criterion. Use 2000 simulated datasets. Report the comparison table and justify what the number of factors should be.

(iii) With the number of factors as decided in Part (ii), estimate the factor loading matrix using PCA method and then improve it using varimax criterion.

(a) Report the overall improvement as the ratio of values of varimax criterion before and after applying the criterion.

(b) Report the improvement for each factor by showing the ratio of factor-specific variances before and after applying the criterion.

(c) Which factor showed maximum improvement due to varimax? Which factor showed minimum improvement due to varimax?

(iv) Now, using the improved loading matrix from Part (iii), complete the following table:

How many variables have at most 30% variability explained by factors	The variable which is worst explained by the factors	Communality of 236^{th} variable
Proportion of overall variability explained by 2^{nd} factor	Correlation between 23^{rd} and 317^{th} variables	Specific variance of 840^{th} variable

(v) Report the factor scores for 16^{th} observation.