Statistical Sciences

Honours Multivariate Analysis
Continuous Assessment 2

Instructions:

- You will be divided into groups for this assessment.
- Your **.pdf** report may be compiled using any software you like (Rmarkdown, LaTeX, MSWord, etc.), as long as the presentation is neat.
- Do NOT paste R output verbatim, this will be penalised. If you want to include R output, typeset it properly or present it in a table.
- To help the reader easily assimilate the information, round values to a small number of decimal places (unless there is a reason for expressing a more exact value).

The file `CA2.csv` contains 100 observations on 12 unknown variables. Consider this as some data matrix $\boldsymbol{X}$. Using Singular Value Decomposition, find lower rank approximations of $\boldsymbol{X}$ for all ranks from 1 – 12.

For each approximation $\tilde{\boldsymbol{X}}_k$ of rank k, calculate the error, $\boldsymbol{\Delta}_k = \boldsymbol{X} - \tilde{\boldsymbol{X}}_k$.

1. Consider the rank 4 approximation. Report the mean vector of the approximation error, i.e. $\bar{\boldsymbol{\Delta}}_4$.
2. Compare the correlation matrix of $\boldsymbol{X}$ with that of $\tilde{\boldsymbol{X}}_2$ and briefly interpret.
3. Calculate the Frobenius norm, defined as

$$||\boldsymbol{A}||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$

   for $\boldsymbol{\Delta}_k,\ k = 1, \ldots, 12$. Plot the Frobenius norm as a function of k and briefly describe your findings.
4. Plot the percentage of the total variation in $\boldsymbol{X}$ retained in $\tilde{\boldsymbol{X}}_k$ for $k = 1, \ldots, 12$. Again, briefly interpret.
