Honours Multivariate Analysis
Continuous Assessment 2

Instructions:

- You will be divided into groups for this assessment. Only 1 submission per group is required.

- 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.txt` 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 $\boldsymbol{X}_k$ of rank k, calculate the error, $\boldsymbol{\Delta}_k = \boldsymbol{X} - \boldsymbol{X}_k$.

1. Consider the rank 3 approximation. Report the mean vector of the approximation error, i.e. $\bar{\boldsymbol{\Delta}}_3$.

2. Standardise each variable in $\boldsymbol{\Delta}_3$ and plot a histogram of these 1200 standardised errors. Briefly comment on this distribution.

3. Compare the correlation matrix of $\boldsymbol{X}$ with that of $\boldsymbol{X}_3$ and briefly interpret.

4. 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.