Stat ST465/665, Project 2

Be sure to label all plots and include your code in the indicated place.

Problem

1. **(16 points)** Read in data set **lumber.txt**, *courtesy of W. Galligan, presented in Table 4.3 of Applied Multivariate Statistical Analysis by R. Johnson and D. Wichern, Pearson, New Jersey, 2019..* The random variables are different measures of stiffness in a set of boards. The first is determined by response to a shock, the second is response to vibrations, and the third and fourth are static. Units are not given.

   (a) Display a matrix scatter plot.

   (b) Is the data matrix $\boldsymbol{X}$ likely to be full rank? Provide an explanation of your answer using the matrix scatter plot.

   (c) Compute and display $\bar{x}$, $\boldsymbol{S_X}$, and $\boldsymbol{R_X}$.

   (d) Define new random variables $Y_1 = 2X_1 - 3X_3$, $Y_2 = X_1 + X_2 + X_3$, $Y_3 = 4X_3$, and $Y_4 = 2X_3 - 2X_4$. Find the matrix $\boldsymbol{A}$ (by hand) such that $\underset{\sim}{Y} = \boldsymbol{A}\underset{\sim}{X}$ and display .

   (e) Compute and display the corresponding data matrix $\boldsymbol{Y}$, $\bar{\underset{\sim}{y}}$, $\boldsymbol{S_Y}$, and $\boldsymbol{R_Y}$ for the new variables.

   (f) Include your code.

2. (**6 points**) Read in the data set **data_for_pr2.txt**, *Data set generated by Don, 2023.*

   (a) Compute and display R.

   (b) Explain the values in R.

   (c) Include your code.

3. **(18 points)** Read in data set **w-nat-track-rec.xlsx**, *(National Track Records for Women, IAAF/ATFS Track and Field Handbook for Helsinki 2005)*, call the data matrix X.

(a) Construct Y by eliminating the first column of X and display.

(b) Set $R = \mathbf{corr}(Y)$ and display.

(c) Compute $\det(R)$ and display. Does $\det(R)$ indicate anything important about R? If so, what?

(d) Solve $R\underset{\sim}{x} = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 2 \\ -1 \\ 1 \\ -2 \end{pmatrix}$ and display $\underset{\sim}{x}$.

(e) Compute the norm $\left\| R\underset{\sim}{x} - \begin{pmatrix} 1 \\ 3 \\ 1 \\ 2 \\ -1 \\ 1 \\ -2 \end{pmatrix} \right\|$ and display. Is the value what you expect to see? Why?

(f) For $\underset{\sim}{z} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$, compute $\underset{\sim}{w} = R^5 \underset{\sim}{z}$ and display.

(g) Compute the projection of $\underset{\sim}{z}$ on $\underset{\sim}{w}$ and display.

(h) Include your code.

3

4. **(19 points)** Read in the data set **Forbes-ten-companies.xlsx**, *World's 10 Largest Companies, The Forbes Global, 2000, April 18, 2005.* and call the data matrix X.

   (a) Produce a matrix scatter plot for X.

   (b) Produce a boxplot for X.

   (c) Could scale be an issue in computing covariance? Justify your answer.

   (d) Compute and display $\bar{x}$, S, and R.

   (e) Compute and display the standardized data matrix X^*.

   (f) Produce a boxplot for X^*. Does this indicate a scale issue in X^*.

   (g) Compute and display $\bar{x}^*$, S^*, and R^*.

   (h) Do R and S^* have the expected relationship? Explain.

   (i) Do R^* and S^* have the expected relationship? Explain.

   (j) Include your code.

5. **(6 points)** Read in the data set **pr2_second_data_set.txt**, *Data set generated by Don, 2022.* Call the variables A, B, $\overline{C}$, D, and $\overline{E}$.

   (a) Display a matrix scatter plot.

   (b) Indicate which of the subplots provide evidence for dependency between the random variables.

   (c) Include your code.