STAT401 Homework #3
Spring 2023

**Q1**. Assume that data $X = (X_1, X_2)'$ follows a multivariate normal distribution with mean $\mu$ and the variance-covariance matrix $\Sigma$ given by

$$\mu = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}.$$

(a) Determine the population principal components, $Y_1$ and $Y_2$.
(b) Calculate the proportion of the total population variance explained by the first principal component.
(c) Calculate the correlation between $X_1$ and $Y_2$.
(d) For an observation $x = (1, -1)'$, calculate the principal component score for the second principal component.
(e) Using a graph, explain the geometric meaning of the principal components in (a).

**Q2**. Sales and profits were measured for the world's 10 largest companies. (available as an attached file, "COMPANY.DAT")

(a) Determine the sample principal components and their variances using the covariance matrix *S*.
(b) Determine the sample principal components and their variances using the correlation matrix *R*.
(c) Which analysis results would you recommend between (a) and (b)?

**Q3**. Data (RADIOTHERAPY.DAT attached with this file) measures average ratings over the course of treatment for cancer patients undergoing radiotherapy. Among six variables, consider the following five variables.

| | |
|---|---|
| $X_1$ | Number of symptoms |
| $X_2$ | Amount of activity (1–5 scale) |
| $X_3$ | Amount of sleep (1–5 scale) |
| $X_4$ | Amount of food consumed (1–3 scale) |
| $X_5$ | Appetite (1–5 scale) |

(a) Choose whether you would conduct principal component analysis using the sample variance-covariance matrix $S$ or the sample correlation matrix $R$. Justify your choice.
(b) Find the sample principal components based on your choice of $S$ or $R$ in (a).
(c) Suggest an appropriate number of the principal components for this data set.
(d) Interpret the principal components chosen in (c).
(e) Given the results in (a) - (d), do you feel that the radiotherapy data can be summarized in fewer than five dimensions?