Homework 10 Due date: June 1

1. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{60}$ be a random sample of size 60 from a four-variate normal distribution having mean μ and covariance Σ. Specify each of the following completely.

   (a) The distribution of $\bar{\boldsymbol{X}}$.

   (b) The distribution of $(\boldsymbol{X}_1 - \mu)'\Sigma^{-1}(\boldsymbol{X}_1 - \mu)$.

   (c) The distribution of $60(\bar{\boldsymbol{X}} - \mu)'\Sigma^{-1}(\bar{\boldsymbol{X}} - \mu)$.

   (d) The distribution of $59S$ for sample covariance matrix S.

   (e) Let $\boldsymbol{Y}_i = [e_1'\boldsymbol{X}_i, e_2'\boldsymbol{X}_i]'$ where e_1 and e_2 are the two first eigenvectors of Σ. What is the distribution of $\boldsymbol{Y}_i$ and what is the distribution of the sample mean $\bar{\boldsymbol{Y}}$?

   (f) Suppose the sample variance-covariance matrix of $\boldsymbol{Y}_i$ is S_y. What is the relationship between S and S_y? What is the distribution of S_y (or its scaled form)?

2. Given the penguin data from 楊明澐 in Homework1, please first remove missing values.
   (a) Make a scatter plot between Culmen Length and Culmen Depth. Color the points according to species.
   (b) Evaluate if the Culmen Length and Culmen Depth jointly follow a bivariate normal distribution for each of the three species. You should make a chi-square plot for each one and comment on it.