

Homework from Chapter 1

Ex. 1.1 Find the correlation matrix and covariance matrix of the data in Table 1.1.

Ex. 1.2 Fill in the missing values in Table 1.1 with appropriate mean values, and recalculate the correlation matrix of the data.

Ex. 1.3 Examine both the normal probability plots of each variable in the archaeological data in Table 1.3 and the chi-square plot of the data. Do the plots suggest anything unusual about the data?

Ex. 1.4 Convert the covariance matrix given below into the corresponding correlation matrix.

$$\mathbf{R} = \begin{bmatrix} 3.8778 & 2.8110 & 3.1480 & 3.5062 \\ 2.8110 & 2.1210 & 2.669 & 2.5690 \\ 3.1480 & 2.669 & 2.6550 & 2.8341 \\ 3.5062 & 2.5690 & 2.8341 & 3.2352 \end{bmatrix}$$

Ex. 1.5 For the small set of (10×5) multivariate data given below, find the (10×10) Euclidean distance matrix for the rows of the matrix. An alternative to Euclidean distance that might be used in some cases is what is known as city block distance (think New York). Write some R code to calculate the city block distance matrix for the data.

$$\mathbf{R} = \begin{bmatrix} 3 & 6 & 4 & 0 & 7 \\ 4 & 2 & 7 & 4 & 6 \\ 4 & 0 & 3 & 1 & 5 \\ 6 & 2 & 6 & 1 & 1 \\ 1 & 6 & 2 & 1 & 4 \\ 5 & 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 6 & 1 \\ 1 & 1 & 5 & 4 & 4 \\ 7 & 0 & 1 & 3 & 3 \\ 3 & 3 & 0 & 5 & 1 \end{bmatrix}$$

Ex. 1.6 A selection of four receipts from a university bookstore was obtained in order to investigate the nature of book sales. Each receipt provided, among other things, the number of books sold and the total amount of each sale.

The receipt ID	The number of books	The dollars of Sales	Students
101	4	142	yes
521	5	252	yes
746	4	148	no
857	3	158	yes

a. Identify and interpret the number of units, the number of variables, the types of variables, and the levels of measurements of variables, respectively.

b. Describe the table in a data matrix by using statistical softwares.

c. Read the observations X_2 and X_4 with variables of the number of books and the dollars of sales using statistical software.

Ex. 1.7 Energy consumption in 2015, by state, from the major sources

x_1 =petroleum, x_2 =natural gas, x_3 =hydroelectric power, x_4 =nuclear electric power is recorded in quadrillions (10^5) of BTU. The resulting mean and covariance matrix are

$$\bar{\mathbf{x}}^T = [.766 \quad .508 \quad .438 \quad .161]$$

$$\mathbf{S} = \begin{bmatrix} .856 & .635 & .173 & .096, \\ .635 & .568 & .128 & .067 \\ .173 & .127 & .171 & .039 \\ .096 & .067 & .039 & .043 \end{bmatrix}$$

a. Using the summary statistics, determine the sample mean and variance of a state's total energy consumption for these major sources.

b. Determine the sample mean and variance of the excess of petroleum consumption over natural gas consumption. Also find the sample covariance of this variable with the total variable in part a.

c. Find the variance-covariance matrix between $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.