

Problem Sheet 2

Mean Vectors & Covariance Matrices

Your Name (Reg No)

Consider the data set `crossSecData` (You may download the data set from <https://github.com/davidrajdg1/MultivariateAnalysis/blob/master/DataSets/crossSecData.xlsx>). The details of the variables are given below.

- Id : Patient Id
- Diabetes: Diabetes Present : 1; Absent : 0
- Fat_intake: The amount of fat intake
- Height : Height of the patient
- Weight : Weight of the patient
- SBP : Systolic Blood Pressure
- DBP : Diastolic blood pressure

Important: Don't forget to save the dataset in your working directory. Set the working directory before you proceed to answer any questions below.

Load the data set Here:

```
library(readxl)
data= read_excel('crossSecData.xlsx')
head(data, n = 5) #Show only first 5 rows of the data
```

Table 1: First 5 rows of the Data Set

Id	Diabetes	Fat_intake	Height	Weight	SBP	DBP
1	0	25	162	57	134	90
2	0	150	151	102	110	70
3	0	34	150	56	124	70
4	0	24	150	59	130	80
5	0	50	156	53	126	80

Question 1 What is a Random vector?

Question 2: What is the Random vector in this data set?

Answer:

```
meanVec = colnames(data)
```

The random vector of the data is (Id, Diabetes, Fat_intake, Height, Weight, SBP, DBP)^T.

Question 3: Generate a new data set from the `data`, consists of Fat_intake, Height, SBP and DBP only. Save the data set as 'newCrossSec' and print only first 10 rows of the new data.

Answer:

Question 4: What is the Random Vector in the new data set?

Answer:

Question 5: What is the sample mean vector in the new data set?

Answer:

```
#Load the new data set as X
X = 'Save the newCrossSec here'
#using the matrix approach find the mean vector. Note that you have generate a vector of all ones. Hint
one = 'Edit this place to generate vector of ones'
SampleMeanVec = 'Edit this place to place the sample mean vector'
```

Some useful hints:

Remember to find the sample covariance matrix, we need to find the matrix X^* where

$$X^* = \begin{pmatrix} X_{11} - \bar{X}_1 & X_{12} - \bar{X}_2 & \dots & X_{1p} - \bar{X}_p \\ X_{21} - \bar{X}_1 & X_{22} - \bar{X}_2 & \dots & X_{2p} - \bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} - \bar{X}_1 & X_{n2} - \bar{X}_2 & \dots & X_{np} - \bar{X}_p \end{pmatrix} \quad (1)$$

Note that, this matrix can be obtained by

$$X^* = \begin{pmatrix} X_{11} - \bar{X}_1 & X_{12} - \bar{X}_2 & \dots & X_{1p} - \bar{X}_p \\ X_{21} - \bar{X}_1 & X_{22} - \bar{X}_2 & \dots & X_{2p} - \bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} - \bar{X}_1 & X_{n2} - \bar{X}_2 & \dots & X_{np} - \bar{X}_p \end{pmatrix} \quad (2)$$

$$= \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} - \begin{pmatrix} \bar{X}_1 & \bar{X}_2 & \dots & \bar{X}_p \\ \bar{X}_1 & \bar{X}_2 & \dots & \bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{X}_1 & \bar{X}_2 & \dots & \bar{X}_p \end{pmatrix} \quad (3)$$

$$= \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1} (\bar{X}_1 \quad \bar{X}_2 \quad \dots \quad \bar{X}_p)_{1 \times p} \quad (4)$$

Question 6: Find the sample covariance matrix of the new data set step by step using the lecture.