

Tutorial 1

1. Consider dataset `final_marks` which contains the scores of final examination of students from a class. Create R code for the question below.
 - (a) Import the dataset into R then name the first column as “ID” and the second column as “mark”.
 - (b) Create an R dataframe called “final” which includes only “mark” column. What is the main difference between “final” and “mark”?
 - (c) Define a function, which has output including summaries of variable “marks” (mean, three quartiles, variance, standard deviation, range).
 - (d) How many students scored ≥ 40 ?
 - (e) How many students scored as high as the maximum mark of the class?

2. Suppose a matrix $\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \\ 1 & 7 \\ 1 & 12 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 4 \\ 6 \\ 13 \\ 15 \\ 20 \end{pmatrix}$. Define $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Using the matrix operations in R to find $\hat{\beta}$.

3. A sequence is generated using the following recursive relation

$$x_n = 2x_{n-1} - x_{n-2}, \quad \text{for } n \geq 3,$$

with $x_1 = 0$ and $x_2 = 1$.

- (a) Use the loop function in R to find the 30th term of the series.
 - (b) Find the sum of the first 20 terms in this sequence.
4. Write a function that will calculate the mean, the second, the third and the fourth central moments if a given data vector on variable X . The r th central moment for $r \geq 2$ is defined by

$$M_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r, \quad \text{where } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Hence obtain the mean, the second, the third and the fourth central moments for the variable “mark” in the data frame “final”.

Answer

1d: 21; 1e: 1; 2: $\hat{\beta} = \begin{pmatrix} 3.8 \\ 1.45 \end{pmatrix}$; 3a: 29; 3b: 190; 4: 31.33163, 97.72931, -470.34752, 27442.12306.