

Exercise 2

TKO_7093 Statistical Data Analysis
BIMA3015 Statistics in Biomedical Sciences

1. Create vectors

A: 5, 8, 7, 6, 8, 4 B: 1.3, 2.1, 1.8, 1.2, 1.4, 2.3 C: y, y, n, y, n, n

Combine the vectors into a data matrix with 3 columns and 6 rows. From the data matrix

- print element (3,2).
- print the 4th row.
- create submatrix of 2 last columns and rows 2-5.
- transpose the data matrix so that it has 6 columns and 3 rows.

2. Read the following two articles

Article 1: L.E. Juarez-Orozco et al., Machine learning in the integration of simple variables for identifying patients with myocardial ischemia, PMID: 29790017

Article 2: W. He et al., High-salt diet inhibits tumour growth in mice via regulating myeloid-derived suppressor cell differentiation, PMID: 32265505

and identify used variables. Especially, find out how many samples were used, and list all variables and determine their type.

3. This exercise deals with human heights (cm). Find out if the following data points are normally distributed or not

170, 192, 184, 168, 176, 181, 163 ?

How about these

170, 170, 170, 170, 192, 192, 192, 192, 184, 184, 184, 184, 168, 168, 168, 168, 176, 176, 176, 176, 181, 181, 181, 181, 163, 163, 163, 163 ?

Can you explain the difference in your results?

4. Create histogram and density plot of the two datasets given in Exercise 3.

5. We consider world temperature (<https://climate.nasa.gov/vital-signs/global-temperature/>). Download the dataset and consider No_smoothing variable. Define mean and median for the data. Is the variable normally distributed? How about, is the variable normally distributed if measurements from year 2000 onwards are only considered?