Exercise 4

TKO_7093 Statistical Data Analysis
BIMA3015 Statistics in Biomedical Sciences

1. Apply a multivariable test and obtain P-value for the following 3 datasets.
a) Dice data: 2, 3, 5, 4, 4, 3 4, 2, 3, 5, 2, 3 3, 1, 4, 4, 3, 5
b) Nordic countries: Fi,Sw,Fi,No,Sw,Fi No,Sw,No,Fi,Fi,Fi Sw,Fi,No,Sw,Sw,No

2. In this exercise we consider articles

Article 4: X. Xing et al., Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing, PMID: 33571124

Article 5: L.L. Liang et al., Covid-19 mortality is negatively associated with test number and government effectiveness, PMID: 32709854

Identify type of variables and methods used to create
a) Article_4 and Figure 3 (J).
b) Article_5 and Figure 1.

3. In this exercise we consider file simulated_data_3_5.csv. It has four groups of 100 observation: F, G, H and I. Apply multivariable test and obtain P-value
a) to compare columns F, G and H.
b) to compare columns F, G and I.

4. Which multiple correction methods are used (if any) in Articles 1-5?

5. In this exercise we use `Horse Colic data set` from https://archive.ics.uci.edu/ml/datasets/Horse+Colic (the `horse-colic.data` file under `Data Folder`), which contained in the UCI Machine Learning Repository. Download the data, read it to your software and make sure that the missing values (question marks) are handled correctly.

Does the horse colic data provide statistical evidence that the mean `rectal temperature` or `age` or `pulse` are different between colic horses treated without surgery and those treated with surgery?