**STAT 40001/STAT 50001 Statistical Computing Fall 2022**

**Lab-10**

1. ***faraway*** package in R contains a data set ***prostate*** which describes 97 men with prostate cancer who were due to receive a radical prostatectomy. Test whether the participants are younger than 65 years.
2. Napa Valley Marathon Times by Age and Gender for 2015 are provided with this assignment.
3. Import the data in R
4. How many runners are older than 50 years of age?
5. Display the age distributions of the runner by gender.
6. Are men older than women?
7. The average completion time for all runners is 4.361 hours. Test whether the completion time for men is lower than 4.361 hours.
8. The average age for all runners is 41.33 years. Test whether women are younger than 41.33 years.
9. The *birthwt* data in MASS package were collected at Baystate Medical Center, Springfield, Massachusetts. Import the data and identify the proportion of low birthweights based on the Race.

Variables provided in the data

Low: indicator of birth weight less than 2.5 kg.

Age:mother's age in years.

Lwt: mother's weight in pounds at last menstrual period.

Race:mother's race (1 = white, 2 = black, 3 = other).

Smoke:smoking status during pregnancy.

Pt1:number of previous premature labours.

Ht:history of hypertension.

Ui:presence of uterine irritability.

Ftv:number of physician visits during the first trimester.

Bwt:birth weight in grams.

1. The data set **Beijing Multi-Site Air-Quality Data** provided in the UCI Machine Learning Repository includes hourly air pollutants data from 12 nationally-controlled air-quality monitoring sites. The air-quality data are from the Beijing Municipal Environmental Monitoring Center. The meteorological data in each air-quality site are matched with the nearest weather station from the China Meteorological Administration. The time period is from March 1st, 2013 to February 28th, 2017. Missing data are denoted as NA.

*Source: Zhang, S., Guo, B., Dong, A., He, J., Xu, Z. and Chen, S.X. (2017) Cautionary Tales on Air-Quality Improvement in Beijing. Proceedings of the Royal Society A, Volume 473, No. 2205, Pages 04-57.*

1. Import the data from Wanliu and Tiantan stations in R and create a single data frame.
2. Display the PM10 values by creating a side-by-side box plot for both stations.
3. Test for significance difference in PM10 values in Wanliu and Tiantan.