

# GFQR 1037 Hands on Little and Big Data

**Instructor:** Dr. PENG, Heng   **Office:** FSC1205   **Phone:** 3411-7021  
**Email:** hpeng@hkbu.edu.hk   **Home page:** www.math.hkbu.edu.hk/~hpeng

**Time and Place:** Tuesday 9:30-11:20 OEE702B, Fri. 10:30-11:20 OEE1017.  
**Office Hours:** Tuesday 11:30am-2:30pm and Friday 11:30am-2:30pm or by appointments.

**Teaching Assistant:** LAI, Pak Wing  
**Email:** Pakwing\_lai@hkbu.edu.hk

**Text Book:** Charles Wheelan *Naked Statistics*, W.W. Norton & Company, New York—London, 2013.

## Reference:

- Alan Agresti, Christine Franklin and Bernhard Klingenberg *Statistics: The art and Science of Learning from Data*, Pearson, 2018.
- Allen B. Downey, *Think Stats*, O'Reilly, 2011.

## Aims and Objective:

No matter what the major of a student is, in his studies he will meet situations in which he needs to handle data. For example, a student may identify the unknown author of an ancient book by the patterns of word usages, while another student may discover a new training programme for an athlete by analysing the historical data. For students of non- science majors, they tend to believe that handling data would be difficult since they would associate data with Mathematics and Statistics, which in turn would be associated with some painful memories. This belief, however, is not

true. Actually, handling data can be simple, easy, and playful when it is done by computer software.

This course introduces some simple methods of data analysis using computer software for small and large amounts of data. The students will use computer to carry out the procedures so that they can forget the formulae; the emphasis is on interpretation, explanation, argument, and reasoning. Only minimal theory behind the methods will be introduced for their interests.

## **Subject Content in Outline**

- Data in daily life
  - History data sets
  - Data in nowadays
  - Data in different scientific and social fields
  - Data in Hong Kong
- Data Types, Collection and Representation
  - Data Types: Qualitative data, Nominal and Ordinal data; Quantitative data, Discrete and Continuous data
  - Data collection methods
  - Data Structure in R
  - Data representation: image data, text data, voice data, video data
- Outliers, Bias and Randomness of Data
  - What are outliers, and how to detect
  - Human collection bias of data
  - Signal-Noise Ratio of data
- Visualizing data
- Data Simplification, Summary Data and Large Sample Laws
- Basic machine learning for big data
- Using computer to make decisions based on a small amount of data at hand

- Using computer to predict new outcomes based on historical data

**Software:** R, Python

### **Course Intended Learning Outcomes(CILOs)**

1. Identify the applications and limitations of various quantitative methods of data analysis.
2. Evaluate practical situations in different aspects and select appropriate quantitative reasoning for solving problems in students' own interests / majors.
3. Manipulate statistical computer software to analyse data.
4. Interpret the results from statistical computer software.

### **Teaching & Learning Activities (TLAs)**

- Lecture

Instructor will give simple real-life problems in lectures to motivate the concepts, followed by discussions of rigorous technical details. Students will then be required to consolidate the knowledge by further practices and through discussion within lectures. The instructor will serve as guide for the students' exploration of their own interests / majors within the boundaries of instructor-framed options.

- In-Class activities

A wide range of applications taken from daily life will be presented to students for further discussion in class. Students are required to contribute in the discussion by expressing their own opinions about what kind of statistical knowledge can be applied and what cannot be applied, and then they will be guided to figure out how statistical software is helpful in solving the problems. This hands-on experience helps students build up confidence in using statistical arguments for quantitative reasoning. Students are encouraged to find the solutions of the problems in their own interests / majors and discuss with the instructor and the other students.

- Assignments

In assignments, students will be given some scenarios in which statistics can be applied for quantitative reasoning in daily life problems. The assignments

require students to explore deeper into the topics. This not only allows students to consolidate their knowledge obtained during lectures but also improves students' communication skills via discussion and exploration.

**Assignments:** Problems will be assigned at class meetings and will be due in class on Fridays of following weeks. No late homework will be accepted. Missed homework will receive a grade zero. The homework will be graded, and each assignment carries equal weight. Verbatim copying of homework is absolutely *forbidden*.

### **Assessment Methods:**

- Continuous Assessment (30%)

Assignments are designed to help students understand the topics. Each assignment will not only allow the instructor to keep track of how well the students master the concepts but, more importantly, also allow students to comprehend the newly introduced statistical notions, as well as to practice the newly acquired quantitative reasoning skills and tools.

- Project Presentation and Report (50%)

Students present topics related to the applications of statistical and quantitative reasoning in daily life or big data. The project contents should be related to the student's major discipline. Students of different disciplines should be assigned within groups for the interdisciplinary works. The division of labour within the groups will be a good opportunity for the students to learn the effective cooperation with people of different backgrounds. Members have to show their individual contributions on the project. Each group submits a written report after the presentation. The report should include a section of self-reflection, which allows members to demonstrate their understanding of the whole course from the perspectives of their major disciplines. The weightings of presentation and report are 40

- Quizzes (20%)

Two quizzes are designed to assess how far students have achieved their intended learning outcomes. Some questions of the quizzes will be based on those seen in the assignments in order to examine whether students can apply the methodology they learned in class to solve problems appropriately. The rest will be used to assess students' abilities on tackling new problems.