

The Chinese University of Hong Kong  
Faculty of Science  
Science Academy for Young Talent

Summer Courses 2023  
Course Outline

*CUSA1026 Statistical Modeling and Big Data Analytics*  
*統計模型及大數據分析*

**Introduction:** Data from various fields, such as climatology, finance and sports, exhibit different properties. This course aims to use the R-package (a statistical software) to visualize the properties of the data, fit the data into various statistical models, evaluate model performance and perform model predictions. Topics include exploratory data analysis, time series models, hidden Markov models, Poisson process and analysis of big data problems. Students will gain hands-on experience in statistical programming at the computer lab.

各種領域的數據（如氣候學，金融及運動）會展示不同的特質。本課程目標是透過統計軟件 R 去透視數據多方面的特性，從而用適當的統計模型去解釋，評估模型的表現及作出數據預測。本課程涵蓋範圍包括：探索性數據分析，時間序列模型，隱馬爾可夫模型，泊松過程和大數據問題的分析。學生將親身體驗統計程式的編寫。

**Medium of Instruction:** Cantonese supplemented with English  
粵語輔以英語

**Organising Unit:** Department of Statistics, Faculty of Science, CUHK

**Teachers:**



**Dr. LIU, Kin Yat (廖健壹博士)**

Lecturer

Department of Statistics, CUHK

Room 116, Lady Shaw Building, CUHK

E-mail: kinyatliu@cuhk.edu.hk

## Course Content:

28 August 2023 (Monday)	<p><u>Lecture:</u></p> <ul style="list-style-type: none"> <li>Basics in Statistical Modeling: Random Variables, Probability Distributions</li> <li>Sports Data: Properties, Poisson Process, Implied Probability and Odds</li> <li>Exploratory Data Analysis (EDA): Scatter plot, Box plot, Histogram, Quartile-quartile Plot, Correlation and Autocorrelation</li> </ul> <p><u>Computer Lab Activities:</u></p> <ul style="list-style-type: none"> <li>R programming: The Basics, Sports Data Modeling, EDA</li> </ul> <p><u>Assessment:</u></p> <ul style="list-style-type: none"> <li>Data Modeling in R</li> </ul>
29 August 2023 (Tuesday)	<p><u>Lecture:</u></p> <ul style="list-style-type: none"> <li>Climate Data: Properties, Seasonal ARIMA Model, Model Prediction</li> <li>Financial Data: Properties, Hidden Markov Model, GARCH Model</li> <li>Monte Carlo Simulation</li> <li>Big Data Problems and Analysis</li> </ul> <p><u>Computer Lab Activities:</u></p> <ul style="list-style-type: none"> <li>R programming: Estimation of Time Series Models,</li> </ul> <p><u>Case Discussion and Assessment:</u></p> <ul style="list-style-type: none"> <li>Data Modeling in R</li> </ul>
30 August 2023* (Wednesday)	Make-up Class

<b>Duration</b>	2 whole day sessions (total 14 contact hours)
<b>Date</b>	28, 29 August 2023 30 August 2023* (make up class)
<b>Time</b>	9:30 am – 1:00 pm & 2:00 pm – 5:30 pm
<b>Teaching Mode</b>	Face to Face (The Chinese University of Hong Kong)
<b>Enrollment</b>	30
<b>Expected Applicants</b>	Students who are promoting to or studying S4 – S6
<b>Tuition Fee</b>	HKD 2,940.00
<b>Credit</b>	1 Academy Unit Certificate of completion will be awarded to students who pass the assessment (if applicable) and attain at least 75% attendance.
<b>Grading Methods</b>	Pass or Fail

\* This date is reserved for make-up classes in case there is any cancellation of classes due to unexpected circumstances.