

11120KHCT100700 巨量資料軟體實務
Software Practice for Data Analytics (Spring 2023)
College of Education, National Tsing Hua University

教師: [郭孟倫 博士 Tonny Menglun Kuo, PhD](mailto:tonny@gapp.nthu.edu.tw) (tonny@gapp.nthu.edu.tw)
助教: 待公布 教室: GEN II綜二420
課程時間: Thursday 13:20-15:10 課程網站: eeClass
Office hour: Friday 10-12:00 a.m. (by Email)

課程簡介

本門課程旨在讓學生運用統計軟體解決複雜多變的研究問題，正確解讀統計軟體中的數據及報表，讓學生依照情境及對象撰寫適切的分析報告。本課程強調解構問題、資料間的關係以及動手實作，學生會學到兩套統計軟體基本功能(R語言與Tableau)，並可以利用這些軟體的功能來分析巨量資料軟體在教育上的應用。

領會過統計的魅力，卻不知如何應用在職場、研究與生活中？統計如何以視覺化方法呈現？統計軟體中複雜的數字和報表，該如何精確並公正的解讀？只追求統計顯著性p值(p-hacking)會造成哪些推論上的偏誤？撰寫分析報告該考慮到哪些因素調整呈現內容？如果上述內容是你的「痛」，你很適合選修這門課程！

本課程是心理統計以及教育大數據分析間的橋樑課程。本門課特別強調學生能夠解構研究問題，找出關鍵構念，分析與資料間的潛在關係，以量化統計的觀點回答研究問題。透過小組報告、微型講授、個人作業，學生會學習兩款熱門且常用的統計軟體，分別是：資料視覺化軟體Tableau、統計與資料科學家愛用的R程式語言。學生並不需要具備程式語言的基礎，但是需要有願意學習寫程式的心。我們會在課堂上逐步教授初階與中階的程式撰寫技巧。

學完這門課程後，將能夠分析研究問題所需要的統計模型、運用統計軟體呈現量化資料結果，有助於未來在畢業專題(Senior Project)、產業實習、或實際工作場中以有效率的方式分析統計資料。對學習行為與教育大數據的有興趣的學生，本門課也能幫助你養成初階教育數據分析的能力。想從統計小白變成統計小達人嗎？快加入這堂課來升級你的量化分析能力吧！

課程目標

1. 回顧初階與中階統計學概念。
2. 選用正確的統計模型來描述、解釋或預測資料並回答研究問題。
3. 以JASP或R軟體分析資料，掌握統計資料的分佈與特性。
4. 運用統計軟體與資料搜尋能力，養成量化問題解決能力。
5. 了解教育現場的不同利害關係人對於教育議題的不同角度，試圖從資料觀點提供新穎且有用的解決方法。
6. 依不同教育場域、情境或對象，撰寫適切的統計分析報告。

課程進度表 (課表為預排參考, 授課教師有權依進度及學生程度滾動調整內容)

日期	周次	課程主題	作業 / 備註
Feb 16	1	課程概述: 基本概念、術語、R語言的歷史、JASP與R操作環境介紹與相關分析	
Feb 23	2	R語言基礎(I): 操作介面、變量、資料型態 (levels, vector, matrix, list, array, and data frame)	
Mar 2	3	R語言基礎(II): factor, table, string, date, and time	
模組一: 資料轉換 (Data Transformation)			
Mar 9	4	資料轉換(I): dplyr (e.g., selecting, appending, sorting, sampling, filtering, pivoting)	
Mar 16	5	資料轉換(II): Merging, binning, reshaping, date operations, and string parsing	
模組二: 資料探索 (Data Exploration)			
Mar 23	6	相關: cor, cor.test, types of plots	
Mar 30	7	資料視覺化 (I): 資料類型與描述性分析與Tableau軟體簡介	
Apr 6	8	資料視覺化 (II): 資料易讀性與互動性	
模組三: 因果建模 (Causal Modeling)			
Apr 13	9	線性迴歸	
Apr 20	10	進階線性迴歸 (適配度、多元共線性、非常態資料)	
Apr 27	11	資料分析報告撰寫要點 專題演講: 教育大數據與校務研究	
模組四: 高維度資料建模 (Modeling Higher Dimensional Data)			
May 4	12	探索式分析: 主成分分析 (PCA)	
May 11	13	探索式分析: Clustering and K-means	
May 18	14	測量模型: 轉軸、解釋、平行分析	
May 25	15	小組期末報告 (課堂報告)	
June 1	16	小組期末報告 (繳交作業)	

課程評量

作業(繳交八次個人小作業, 須附上程式碼)	50%
課堂討論參與	20%
分組報告回顧統計概念	10%
期末小組專題	20%

參考書目

本課程沒有指定用書, 但可參考以下書籍。

Field, A. (2022). *Discovering Statistics Using R and RStudio*. Sage Publications.
(Field有很好的分析概念的理解, 初學R的學生很適合讀這本書)

Lander, J. P. (2017). *R for everyone: Advanced analytics and graphics (Second edition)*. Addison-Wesley.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl, K. C. (2017). *Data mining for business analytics: Concepts, techniques, and applications in R*. Wiley.

(Shmueli老師是商業分析的專家, 這本入門書雖是英文但是字句淺白, 容易閱讀。裡面的程式碼都有附註可以參考, 是實用的商業分析入門書籍)

Verhoef, P. C., Kooge, E., Walk, N., & Wieringa, J. E. (2022). *Creating value with data analytics in marketing: Mastering data science (Second edition)*. Routledge.

(Verhoef列出了許多消費者相關的指標, 並說明這些指標如何在數據中創造價值)

Yau, N. (2011). *Visualize this: The Flowing Data guide to design, visualization, and statistics*. Wiley Publications.

(雖然有點舊, 但是Yau一步一步教導資料視覺化的原則跟流程, 做圖時可以參考)

鄭中平、許清芳。(2015)。R在行為科學之應用。台北: 雙葉書廊。

(一本言簡意賅的中文入門書, 是以前我在承德的量表發展課參考的書。)

可供參考資料集

教育雲eduLRS學習行為資料庫	https://pads.moe.edu.tw/
臺灣學生學習成就評量資料庫 (Taiwan Assessment of Student Achievement, TASA)	https://tasal.naer.edu.tw/tasa
台灣教育長期追蹤資料庫(Taiwan Education Panel Survey, TEPS)	http://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=2&csid=7
「台灣教育長期追蹤資料庫」後續調查(TEPS-B)	http://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=2&csid=20

特殊教育長期追蹤資料庫(Special Needs Education Longitudinal Study, SNELS)	http://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=2&csid=18
國家教育研究資料庫(National Education Research Data Archive, NERDA)	https://www.naer.edu.tw/PageDoc/Detail?fid=440&id=3491
臺灣後期中等教育長期追蹤資料庫 (Taiwan Upper Secondary Education Database)	https://use-database.cher.ntnu.edu.tw/used/
各級學校學生運動參與情形調查	https://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=4&typeb=013&csid=126
Datashop	https://datashop.memphis.edu/
Newfoundland Labrador (Cnanadat) Download a Database of K-12 Schools	https://www.gov.nl.ca/education/faq/schooldatabase/
NCES national center for education statistics	https://nces.ed.gov/
Kaggle	https://www.kaggle.com/junyiacademy/datasets
THE WORLD BANK Education Statistics (EdStats)	https://datatopics.worldbank.org/education/home
California Department of Education	https://www.cde.ca.gov/ds/ad/downloadabledata.asp