

財務大數據分析 (BA)

BIG DATA ANALYSIS IN FINANCE

課程資訊

授課教師: 雲慕書老師
辦公室: 364
課程網頁:

上課時間: Thu. 13:10 - 16:00
Email: mushuyun@ccu.edu.tw
Office Hours: Thu. 16:00-17:00

課程目的

本課程主要目的在於透過專案實做來理解財金資料的結構、資料清理、與常用的計量分析方式。**本課程預設你會使用 AI 工具。**現在的 AI 已經可以寫出語法正確的程式，因此相較於過去偏重語法記憶的學習方法，本課程的重點放在三件 AI 替代不了的事：

- **資料架構的理解**：一筆資料代表什麼？合併之後應該剩幾筆？結果是對的還是只是「跑得動」？
- **程式的判讀**：讀懂一段別人（或 AI）寫的程式，說出它做了什麼、會輸出幾筆、錯在哪裡。
- **分析結果的詮釋**：這個係數可以解釋成因果嗎？有沒有遺漏變數？標準誤可不可信？

換句話說，這門課考的不是 AI 產出的品質，而是**你能不能檢查它**。

涵蓋主題

- 資料的形狀與資料整理：Data Step、Proc SQL 合併、時間對齊與 look-ahead bias
- 與 AI 協作寫程式的驗證方法
- 關聯分析與決策樹：有關聯不等於可預測
- 財務應用實作：投資組合分析、迴歸分析與虛擬變數、固定效果
- 因果推論與內生性：遺漏變數、測量誤差、因果互倒、工具變數、叢集標準誤
- 貝氏統計與常見的推論陷阱

修課要求

- 本課程著重資料整理與數據分析，要求修課同學具備基本統計基礎。
- 本課程要求學習 SAS 程式 (**必需每週練習**)。不定期會有隨堂小考，題型以**讀程式、判斷輸出、找出錯誤**為主，不是默寫語法。
- 課程採分組進行，**全班分為 8 組**，第一週完成分組。
- 每組在期中提出研究提案，期末報告研究結果。詳見下方「研究專案」。
- 學期後半每週上課前需填寫**進度 checkpoint**（本週做了什麼／目前 obs 數／卡在哪）。**遲交不補、不寄提醒信**。

研究專案：骨幹＋外接

為了讓每組的進度可以用同一套標準檢查，專案採「**骨幹＋外接**」的資料架構：

- **骨幹**：TEJ 資料庫。
- **外接**：你自己找的變數，接到骨幹上 (政府資料開放平台、中華民國統計資訊網、台北市資料大平台、商業大數據平台)。外接分三級：**A 級** (有統編或公司代碼，優先選)、**B 級** (地理型，可做但標準誤要處理)、**C 級** (只有全國、只有年度的總量，會被年度固定效果吸收，(不建議))。
- 想做完全自訂資料的題目，請於第五週前當面說服老師，個案核准。

期末口試

期末每組簡報 15 分鐘，簡報後從**公開題庫 10 題中隨機抽 3 題**，由老師指定組員回答。題庫第一週就發給大家，內容全部是「你的資料、你的變數、你的決策」——例如：你的樣本原始幾筆、最後剩幾筆、中間刪掉了什麼？你的骨幹和外接資料用什麼 key 合併？這段程式是誰寫的，如果是 AI，你改了哪裡？**分數壓在口試，不在簡報**。簡報只佔 10 分因為現在用 AI 生成 ppt 和講稿沒有成本。

評分標準

- 課堂參與 10%
包含出席、課堂互動。無法出席請於課前寫信告知。
問問題、回答問題一次 1%，每週記一次。
- 隨堂小考 (讀程式／抓錯) 10%
- 期中考 20%
- 每週進度 checkpoint 10%
- 期中提案 15%
- **期末口試 (隨機抽題) 25%**
- 期末書面／簡報 10%

checkpoint 缺交累計 3 次，期末報告與口試分數乘 0.8。不補交、不個別提醒、不受理事後補件。

期中考 (10/29) 為上機考，開放使用 AI。題型三種：讀一段 AI 產出的程式，說明它做了什麼、會輸出幾筆；找出一段「跑得動但結果錯」的程式錯在哪、後果的方向；給一個資料需求，用 AI 完成，但必須附上你怎麼驗證它是對的。

授課材料

- 「資料挖礦與大數據分析」，簡禎富，許嘉裕著。前程文化出版，ISBN：9789865774257。¹
- 「SAS 一財金計量之應用」，王萬成著。新陸書局出版，ISBN：9789866333835。
- 課堂程式檔分兩種版本：上課發**課堂版** (關鍵處留空，由你在課堂上判斷後填入)，**完整版**於課後上傳。

¹配合教育部針對保護智慧財產權觀念之宣導，請尊重智慧財產權，勿非法影印教師指定之教科書籍。

課程大綱 (暫訂)

每週三小時的固定結構：前 100 分鐘講授與實作，最後 50 分鐘為**上機練習 (前半段)** 或**專案 Lab (後半段)**。

09/10 【W1】 課程介紹：資料的形狀

- 檔案系統速成：路徑、副檔名、解壓縮、libname
- SAS 操作界面；一筆資料代表什麼
- **分組完成 (8 組)**

09/17 【W2】 Data Step 與資料匯入

- Data Step 是「一次一筆的迴圈」
- proc import、keep/drop/rename、format 與真值的差別

09/24 【W3】 排序、分組與轉置

- by 群組、proc rank、proc means、long 與 wide 互轉

10/01 【W4】 Proc SQL 合併 I

- 「一次一筆」與「一次一整張表」的對比；left join；一對多

10/08 【W5】 Proc SQL 合併 II

- 時間對齊、財報公布延遲、頻率不同的資料怎麼接
- look-ahead bias：用到當下還不可能知道的資訊

10/15 【W6】 迴歸分析與虛擬變數

- OLS 實作、N-1 個虛擬變數、交乘項、產業別合併

10/22 【W7】 固定效果+與 AI 協作的驗證方法

- class 與 absorb、雙向固定效果
- 怎麼檢查一段 AI 寫的程式：先預測筆數，再比對

10/29 【W8】 期中考 (上機，**開放 AI**)

11/05 【W9】 期中提案報告 (全組上台)

11/12 【W10】 關聯分析+專案 Lab

- support / confidence / lift；手算一條規則
- 詮釋陷阱：關聯規則說的到底是什麼
- **每週進度 checkpoint 開始**

11/19 【W11】 有關聯不等於可預測+專案 Lab

- 台股漲跌三步驟：同期關聯 → 加入時間差 → 跨年度驗證

11/26 【W12】 決策樹+專案 Lab

- proc hpsplit、剪枝、驗證樣本、規則輸出

12/03 【W13】投資組合分析＋內生性 I＋專案 Lab

- beta 估計、portfolio sorting、double sorting
- 遺漏變數偏誤：Job Training 的例子

12/10 【W14】內生性 II＋專案 Lab

- 測量誤差與衰減偏誤、因果互倒、工具變數
- 叢集標準誤：外接變數只有很少的不同值時，t 值為什麼會灌水

12/17 【W15】期末口試 I（4 組）

12/24 【W16】期末口試 II（4 組）

12/31 自主學習

01/07 自主學習