

數據科學方法

課程介紹

What is data science?

- Data science = data help science
- Answer scientific questions by (or at least with the aid of) data

Example: 台中市空氣品質

- 台中市空氣品質是否改善？
- Data available [here](#)

Example: 自駕車

- [How Data Science Is Driving The Driverless Car](#)
- [BDD100K](#) dataset by UC Berkeley

Example: 天氣預報

- Numerical weather model + observational data
- [The data science behind weather prediction](#)

Sub-fields

1. Statistics
2. Big data
3. Machine learning
4. Artificial intelligence
5. Data assimilation

這門課會學到什麼？

1. 資料科學主要子領域
2. 根據問題與資料特性選用合適分析工具
3. Python資料科學工具箱

評分方式

- 作業 (約8次) 40%
- 期中考 (11/13?) 30%
- 期末報告 (1/8) 30%

作業

- 程式作業為主
- 將ipynb與pdf檔上傳至課程iLearning
- 約8次
- 嚴禁抄襲!

期中考

- Open book
- 可帶任意參考資料
- 可使用電腦上網
- 禁止交流 (如交談、使用通訊軟體等)
- 作弊提交獎懲委員會

期中考

- 觀念是非題 60分
 - 答案 + 理由 (只有答案不給分)
- 實作(軟體+繪圖)題 50分
 - 繳交ipynb檔至iLearning

期末報告

- 題目與資料於第12週公告
- 實作 + 結果解說 (嚴禁嘴泡與抄襲!)
- Mini thesis report (約5頁)
 - 問題描述
 - 採用方法
 - 結果與解釋
- Oral presentation (10分鐘內)

課程大綱

1. Overview
2. Data manipulation by Python
3. Causality and experiments
4. Randomness and probability
5. Random sampling
6. Statistical modeling
7. Estimation

8. Statistical inferencing
9. Introduction to machine learning
10. Regression
11. Classification
12. Clustering
13. Dimension reduction
14. Data visualization and explorative data analysis
15. Kalman filtering

參考資料

- [Computational and Inferential Thinking](#)
- [Introductory Statistics with Randomization and Simulation](#)
- [Practical Statistics for Data Scientists: 50 Essential Concepts](#)
- [Python for Data Analysis, 2nd Edition](#)
- [All of statistics](#)