

「數量方法與金融應用」 ——分析與決策

Spring 2026

授課教師：統計系余清祥

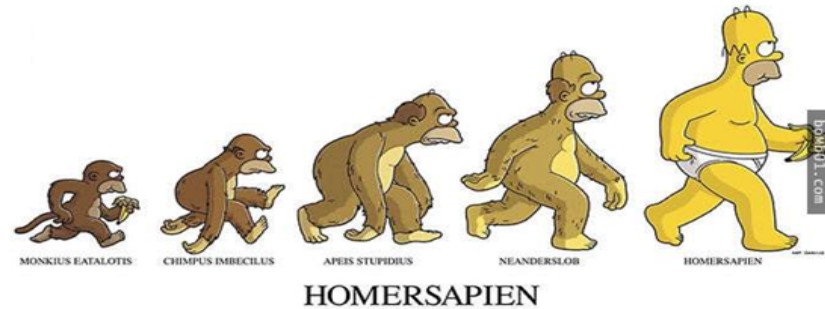
2026年3月7日

授課內容：統計分析

課程下載：csyue.nccu.edu.tw



資訊與決策



- 過去擁有土地、資金等的資本家，因為勞力及資本密集具有優勢；21世紀是知識經濟的時代，擁有及解讀資訊者掌握絕對優勢。

(Amazon、Google！)

- 問題：人類歷史（包括石器時代）各年代的生存關鍵是什麼？

註：天下文化《人類大命運》
《人類大歷史》。



天下文化 遠見

人類大命運

從智人到神人

Homo Deus

A Brief History of Tomorrow

by Yuval Noah Harari

哈拉瑞 — 著 林俊宏 — 譯

天下文化

人類大歷史

從野獸到扮演上帝

Sapiens

[From Animals Into Gods]

A Brief History of Humankind

by Yuval Noah Harari

哈拉瑞 — 著 林俊宏 — 譯

資料分析策略

■ 「觀察」、「推論」、「驗證」三步驟

- 先檢查資料品質，避免Garbage in, garbage out，通常會佔用一半以上的時間。
- 接著是探索性資料分析(EDA)，逐步找出資料重要特性，作為進一步推論(如迴歸分析)的依據。
- 驗證性資料分析(CDA)則是最後步驟，分析結果應與EDA接近，否則需重頭檢查。

資料偵錯

資料輸入錯誤、尋找可能的離群值。

初步探索資料特性

資料的集中、散佈趨勢

驗證已知的結果

是否與已知的結果相同？

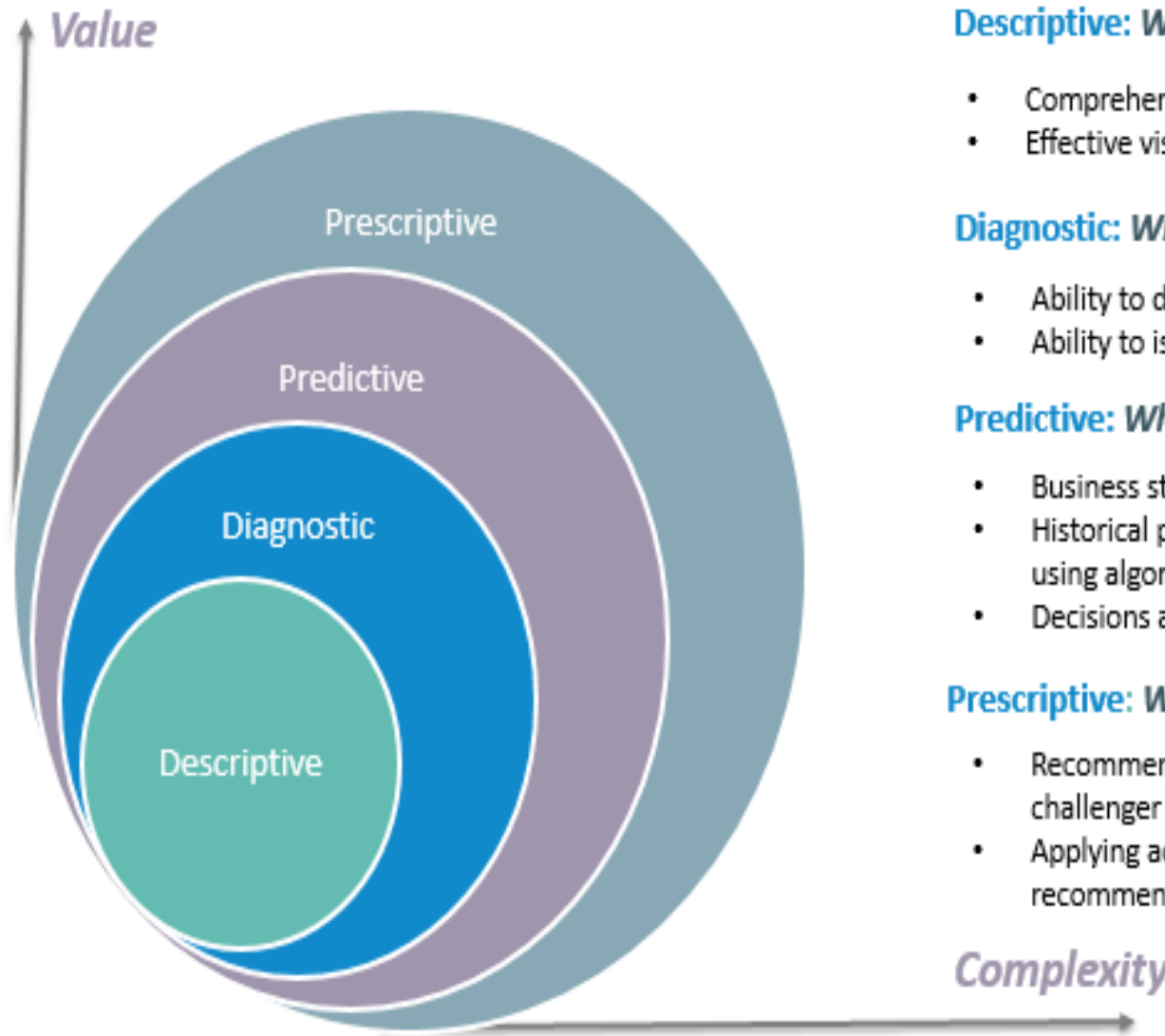


EDA and CDA

- EDA (or **Descriptive analytics**) tells us what happened up from the data.
- CDA can be separated into two parts:
 - **Predictive analytics** give us clues about the future, given data and domain knowledge.
 - **Prescriptive analytics** provide suggestions for optimizing the future.

1. **Descriptive.** Traditional HR metrics are largely efficiency metrics (turnover rate, time to fill, cost of hire, number hired and trained, etc.). The primary focus here is on cost reduction and process improvement. Descriptive HR analytics reveal and describe *relationships* and *current and historical data patterns*. This is the foundation of your analytics effort. It includes, for example, dashboards and scorecards; workforce segmentation; data mining for basic patterns; and periodic reports.
2. **Predictive.** Predictive analysis covers a variety of techniques (statistics, modeling, data mining) that use current and historical facts to make predictions about the future. It's about probabilities and potential impact. It involves, for example, models used for increasing the probability of selecting the right people to hire, train, and promote.
3. **Prescriptive.** Prescriptive analytics goes beyond predictions and outlines decision options and workforce optimization. It is used to analyze complex data to predict outcomes, provide decision options, and show alternative business impacts. It involves, for example, models used for understanding how alternative learning investments impact the bottom line (rare in HR).

4 types of Data Analytics



What is the data telling you?

Descriptive: *What's happening in my business?*

- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: *Why is it happening?*

- Ability to drill down to the root-cause
- Ability to isolate all confounding information

Predictive: *What's likely to happen?*

- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: *What do I need to do?*

- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations



<https://corticare.com/wp-content/uploads/revslider/home-new/clinical-trials-bg.jpg>

如何建立模型？



<https://www.bigskyassociates.com/hs-fs/hub/283820/file-2615489650-jpg/statistical-modeling-methods-for-security-data-analysis.jpg.jpg?width=217&name=statistical-modeling-methods-for-security-data-analysis.jpg.jpg>

驗證性資料分析

- 透過量化模型描述觀察結果：

$$\text{觀察現象} = \text{模型} + \text{誤差}$$

或是

$$y = f(x) + \text{error} ; \text{觀察值} = \text{訊號} + \text{雜訊}$$

- 數量化模型的關鍵：

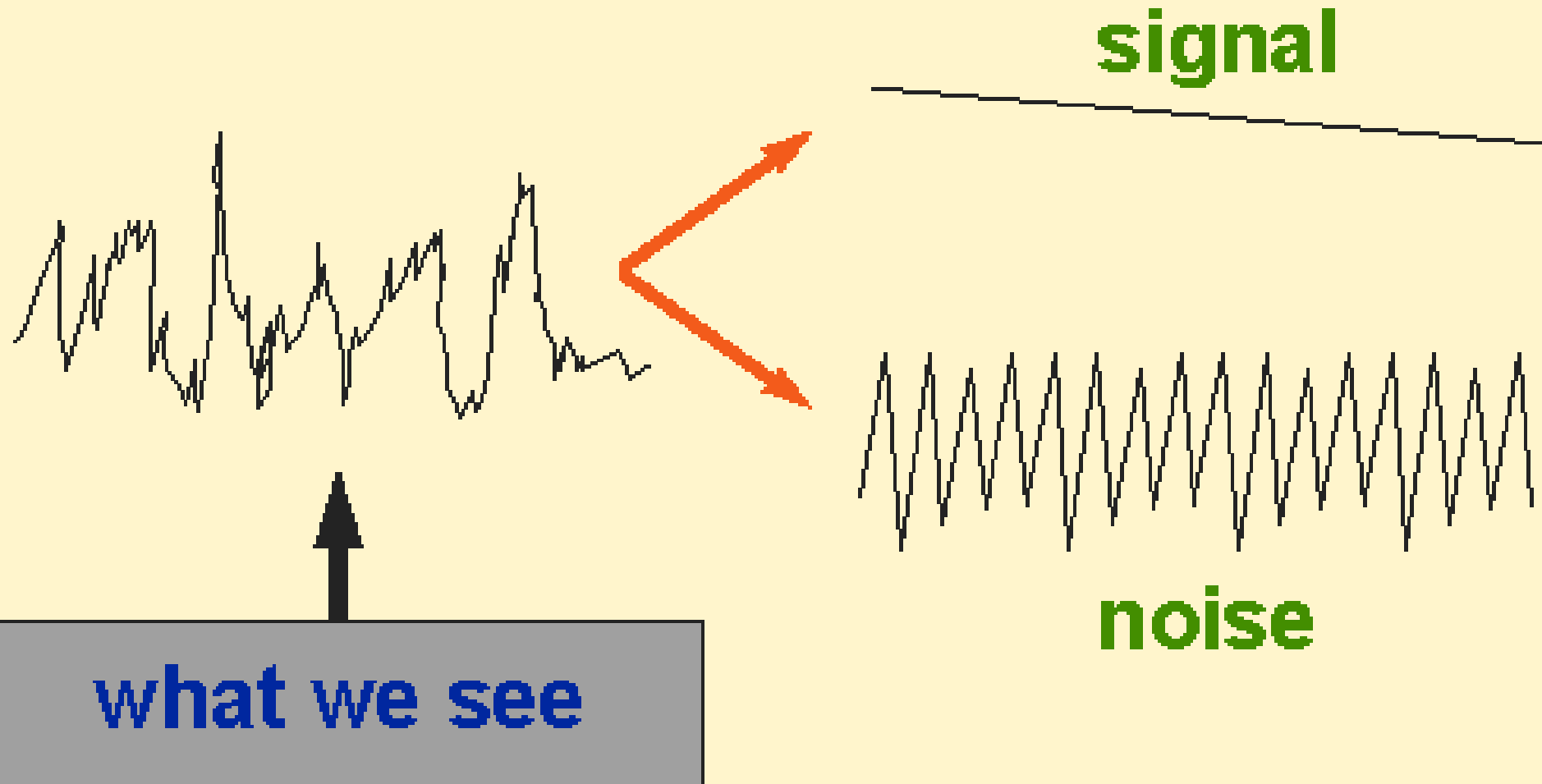
→ 量化目標值 y ：定義問題！

→ 選取關鍵變數： $x_1, x_2, \dots, x_p$

→ 建立量化模型：統計學習、機器學習。

觀察現象 = 模型 + 誤差

What we observe can be divided into:



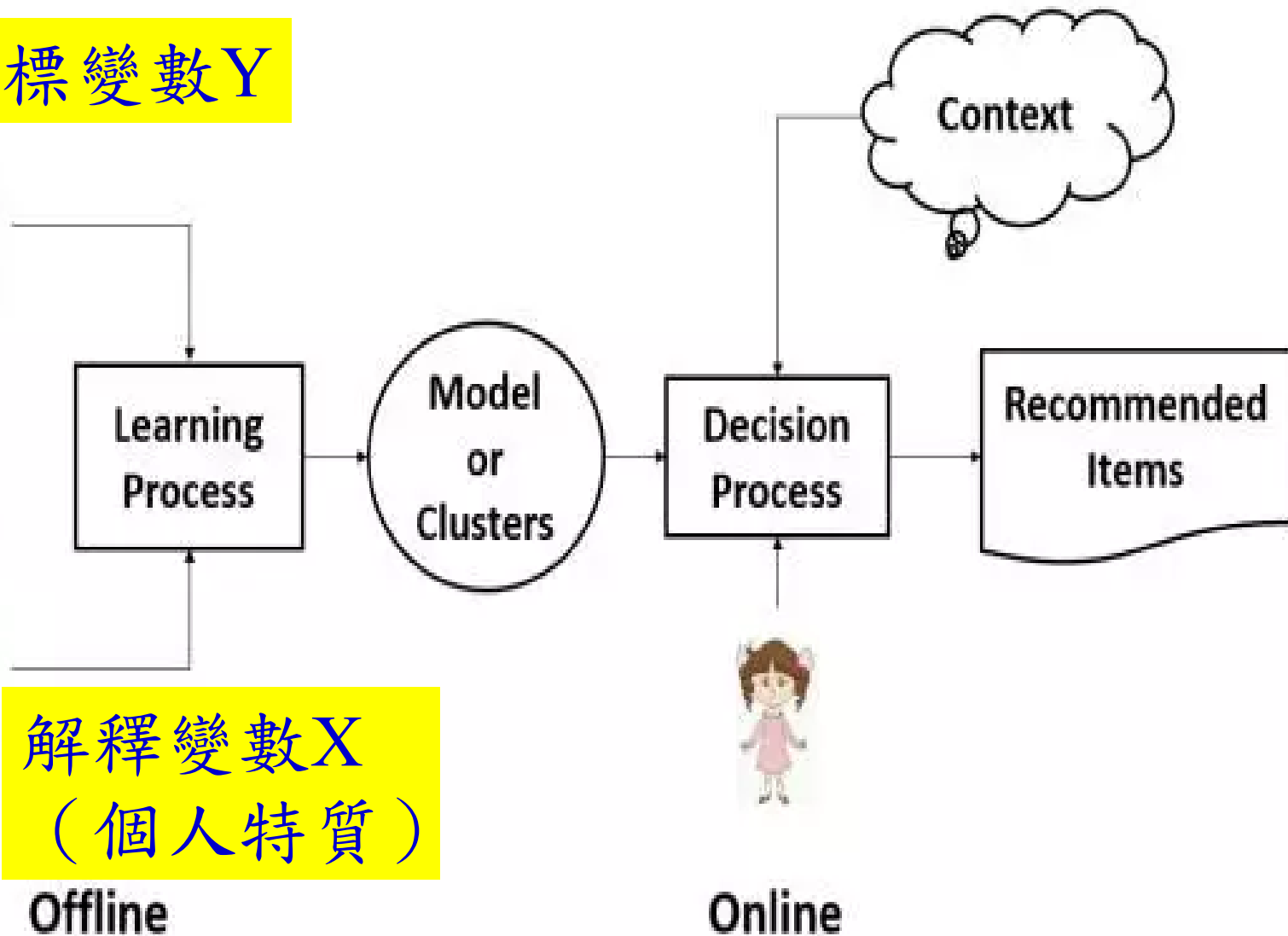
推薦系統(Recommendation System)分析流程

目標變數Y



解釋變數X
(個人特質)

Offline

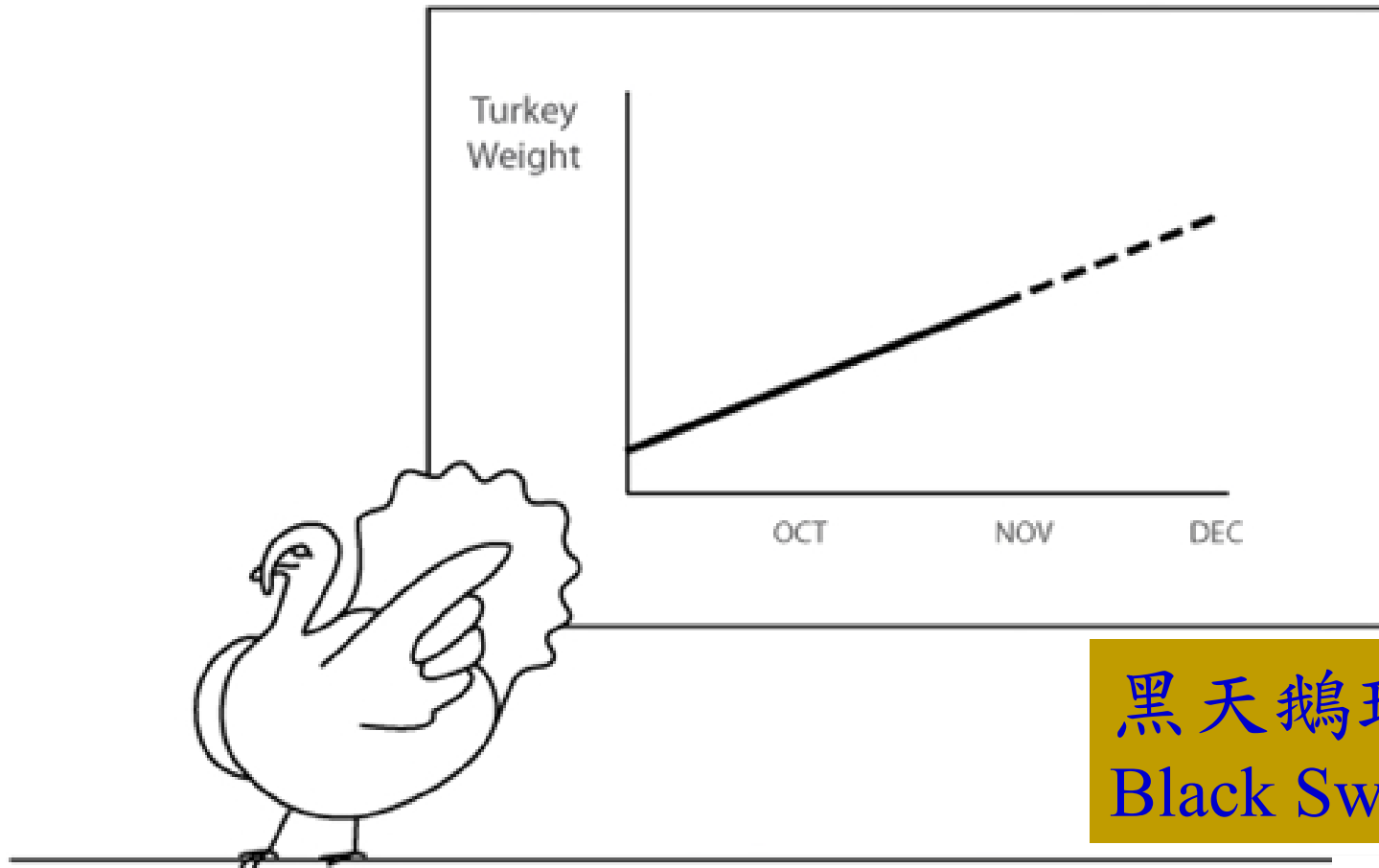


Online

解決問題的考量因素？

- 生活上遇到問題時，解決過程大致包括：
 - 確定問題目標及需求 Y ；
 - 蒐集相關資訊 X （資料來源、資料可信度、資料量及其品質）；
 - 資料整理、分析比對；
 - 決策評估標準 $f(X)$ （不同決策衍生出的優勢及劣勢、短期及中長期的影響）。

THANKSGIVING PREDICTIVE ANALYTICS



黑天鵝理論
Black Swan Theory

"I see no reason why excellent growth shouldn't continue.."

<https://i0.wp.com/timoelliott.com/blog/wp-content/uploads/2013/07/turkey-predictive-analytics.png?ssl=1>

Q：灰犀牛理論？



<https://cdn.technasia.com/wp-content/uploads/2016/10/bs1.jpg>



All models are approximations.
Essentially, all models are wrong, but
some are useful. However, the
approximate nature of the model
must always be borne in mind.

— George E. P. Box —

AZ QUOTES

<https://www.azquotes.com/quote/989358>

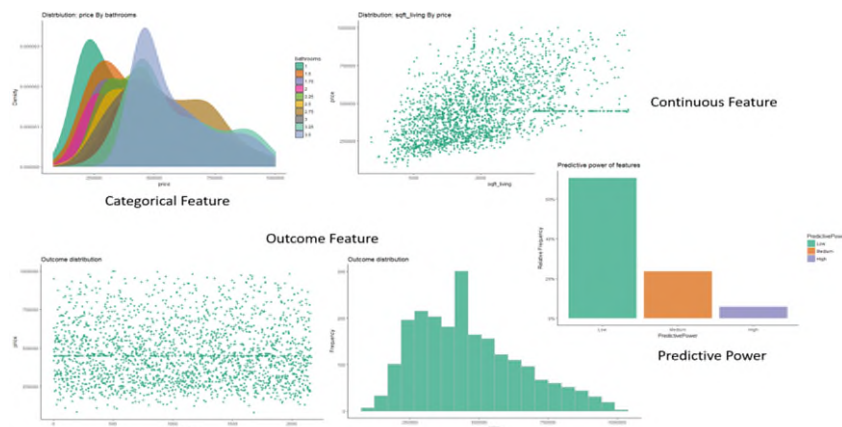
孟子·盡心下》：「盡信書，則不如無書」

探索性資料分析(EDA)

■ EDA首要目的在於資料偵錯、獲得資料的大略資訊、驗證已知結果。

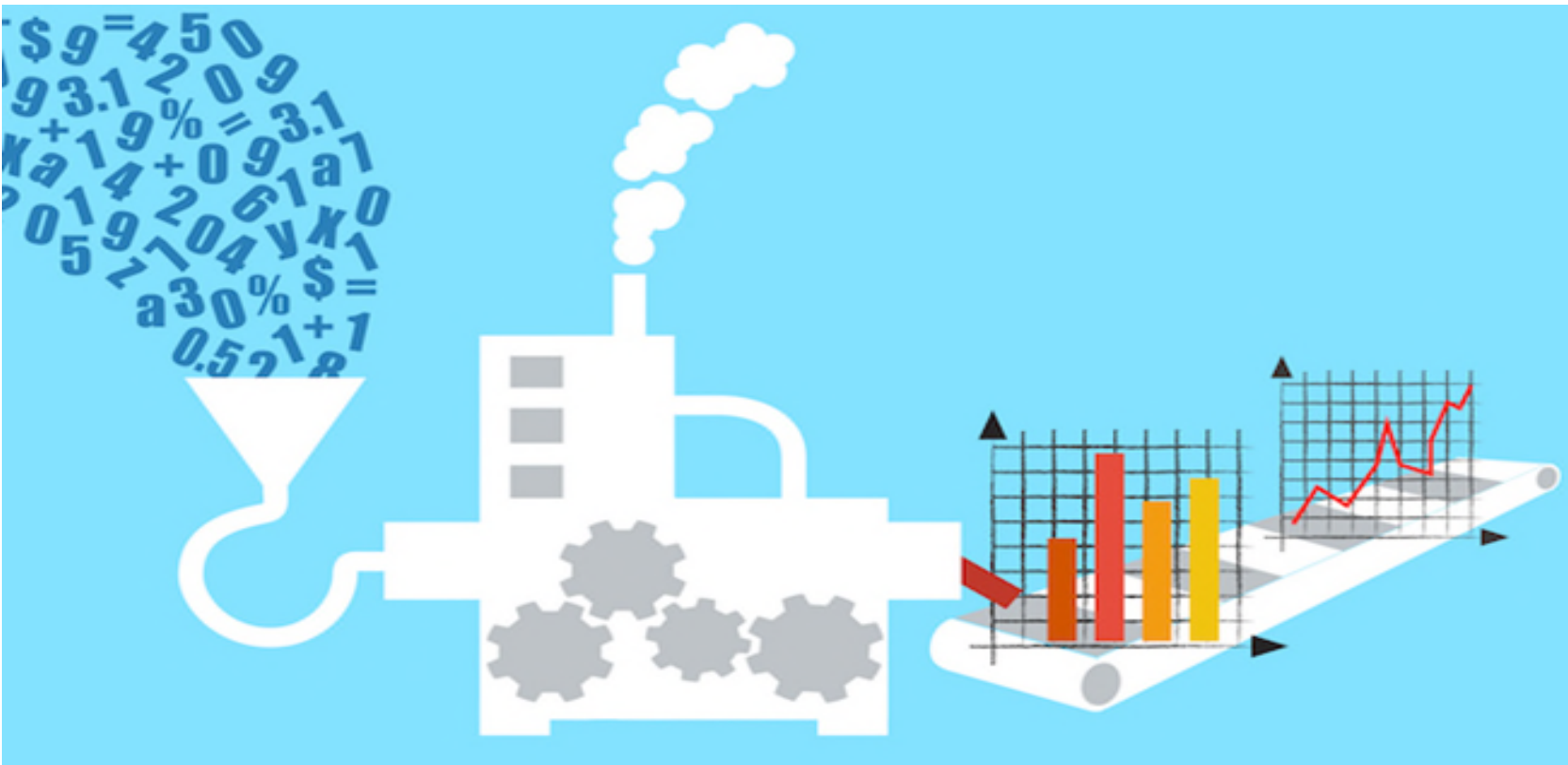
→ 圖形、表格在EDA中扮演重要的角色；並由分析結果中尋找合適的下一步分析方法。

→ 使用任何的統計方法前，先確定該方法需要的假設條件是否滿足。



探索性資料分析(資料驅動)

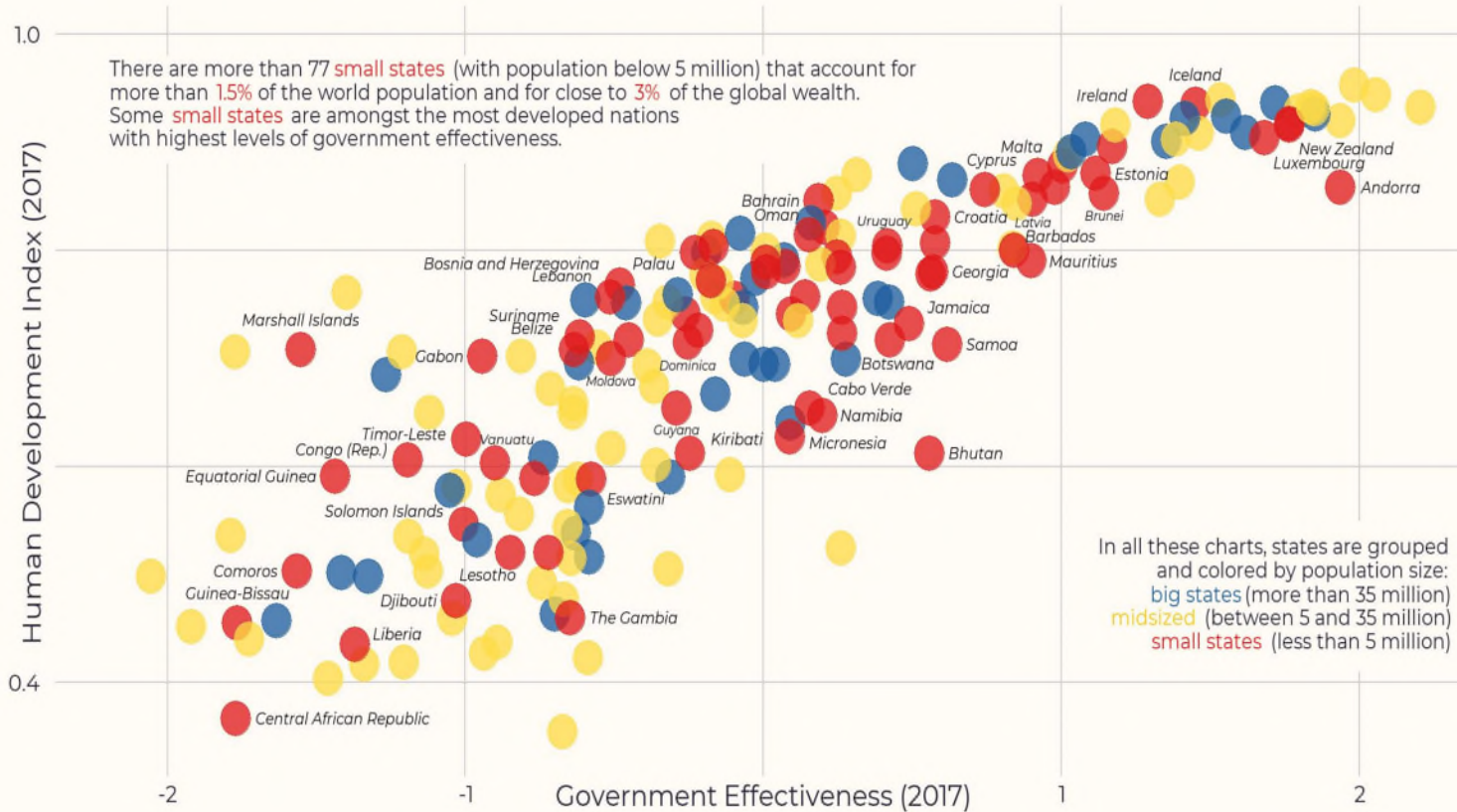
Exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics ... EDA is for seeing what the data can tell us beyond the formal modeling. ---Wikipedia



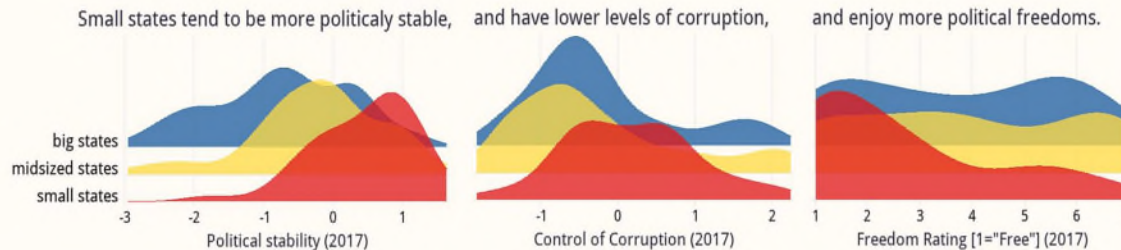
https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.aiche.org%2Facademy%2Fwebinars%2Fapplied-statistics-exploratory-data-analysis&psig=AOvVaw36ZuxAJqz27dLqU5IFzBMO&ust=1570108849384000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCJC1qLXV_eQCFQAAAAAaAAAAA BAJ

第五個V：資料視覺化(Visualization)

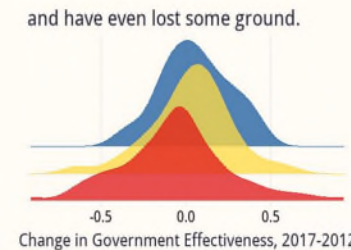
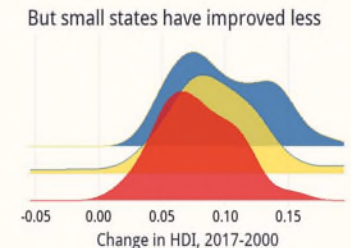
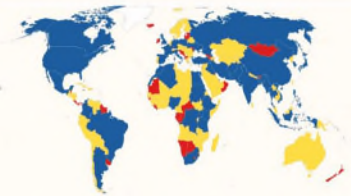
Small States Can Be Big Players in Development and Good Governance



These charts show the distribution densities of different variables by three groups based on state size:

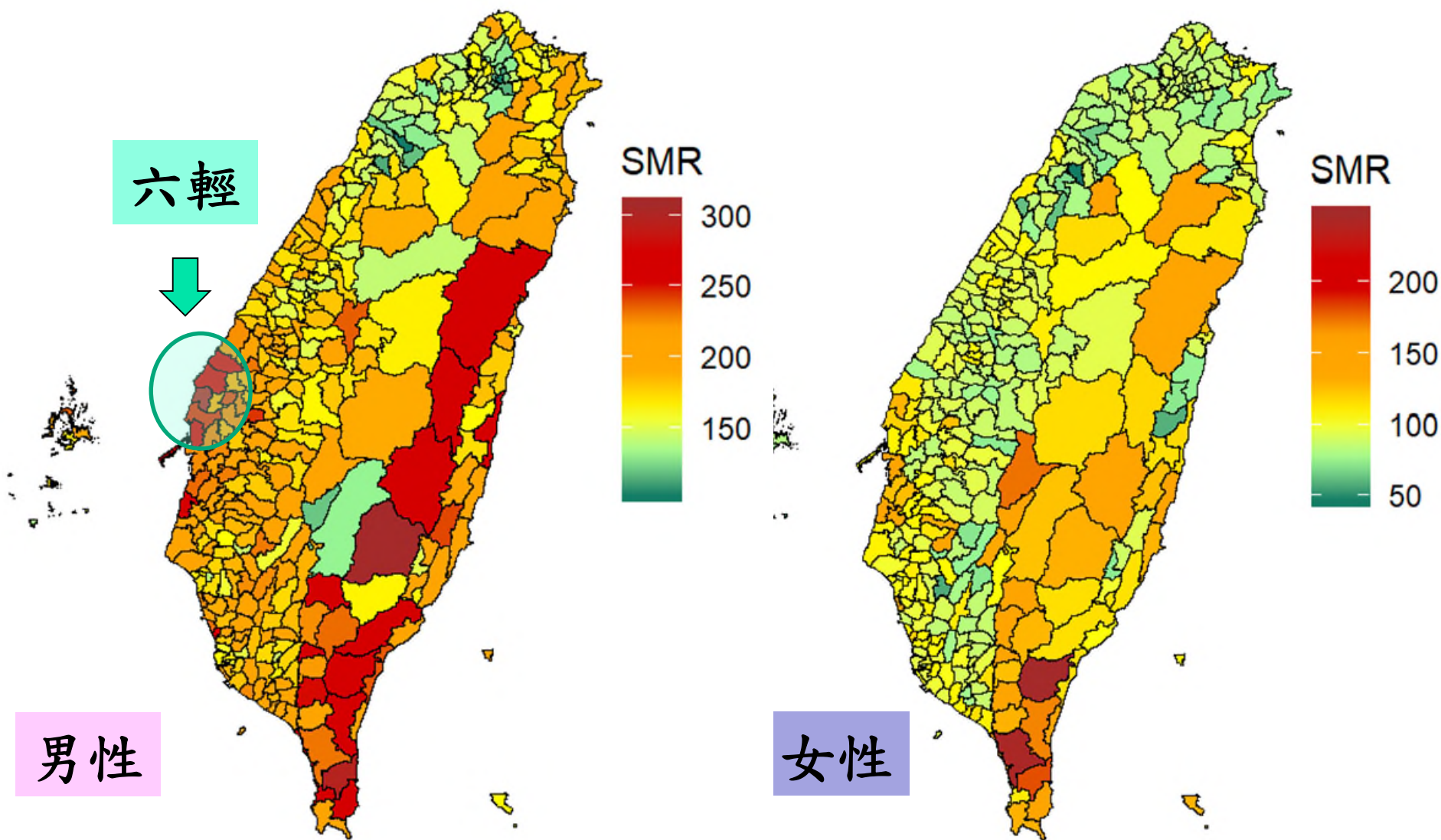


Data: Human Development Index from Human Development Reports. Government Effectiveness, Control of Corruption, and Political Stability from Worldwide Governance Indicators. Freedom Rating from Freedom House.



الجائزة العالمية
لفن عرض للبيانات
WORLD DATA
VISUALIZATION PRIZE

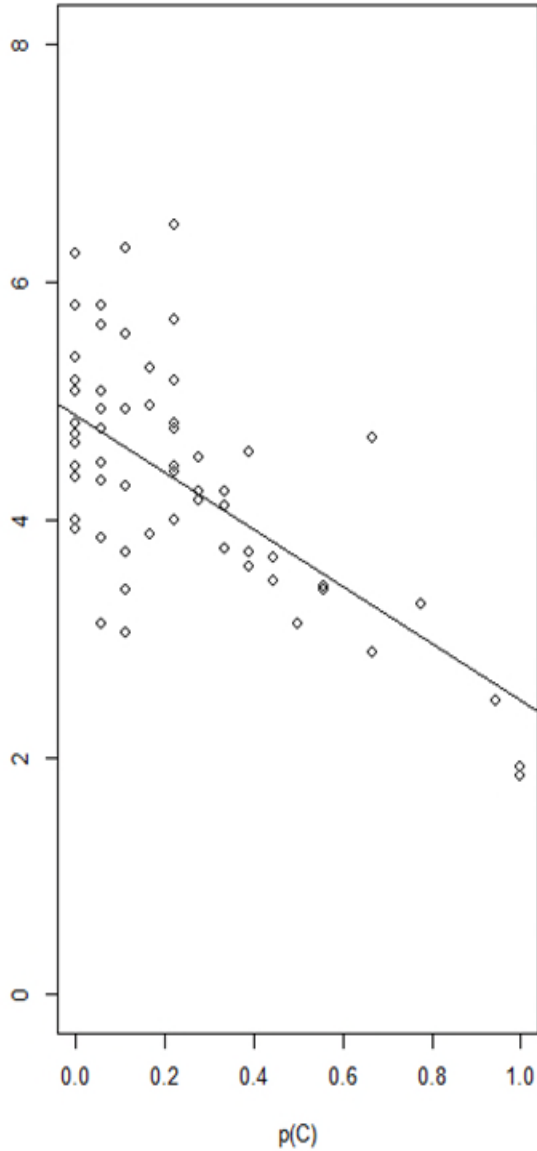
台灣2014~18年鄉鎮市區癌症死亡SMR



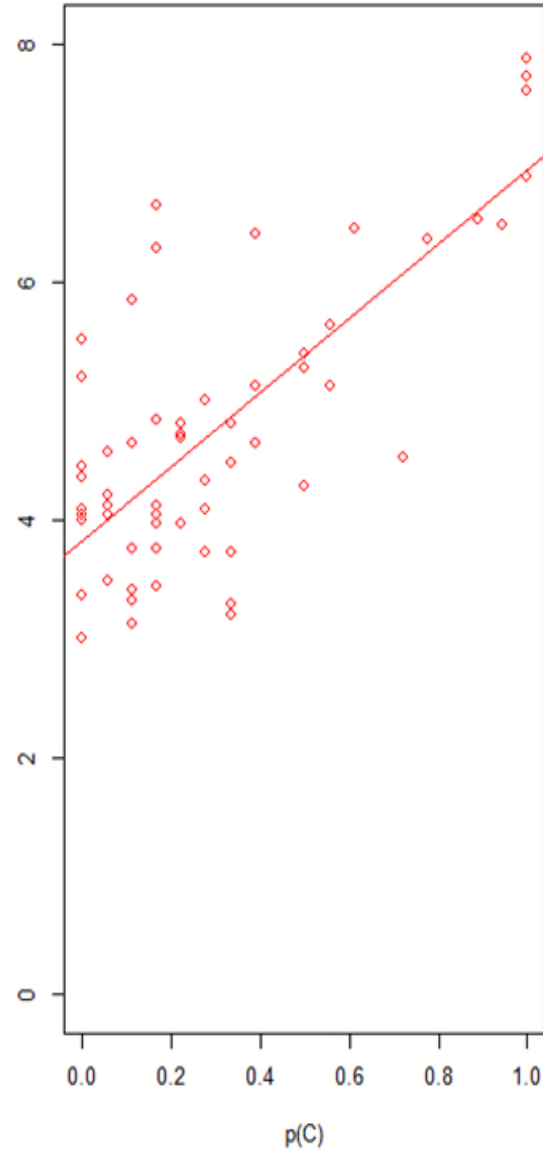
註：標準死亡比(SMR, Standard Mortality Ratio)可視為死亡率比值。

「囚犯困境」合作意願與報酬

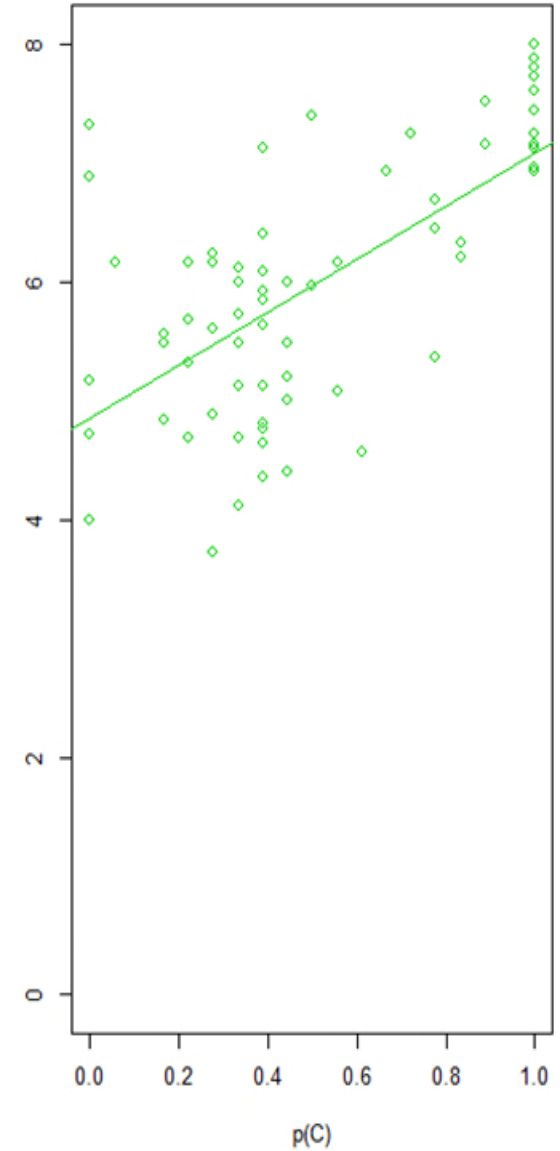
p(C) vs. Avg. Payoff (RM)



p(C) vs. Avg. Payoff (WH)

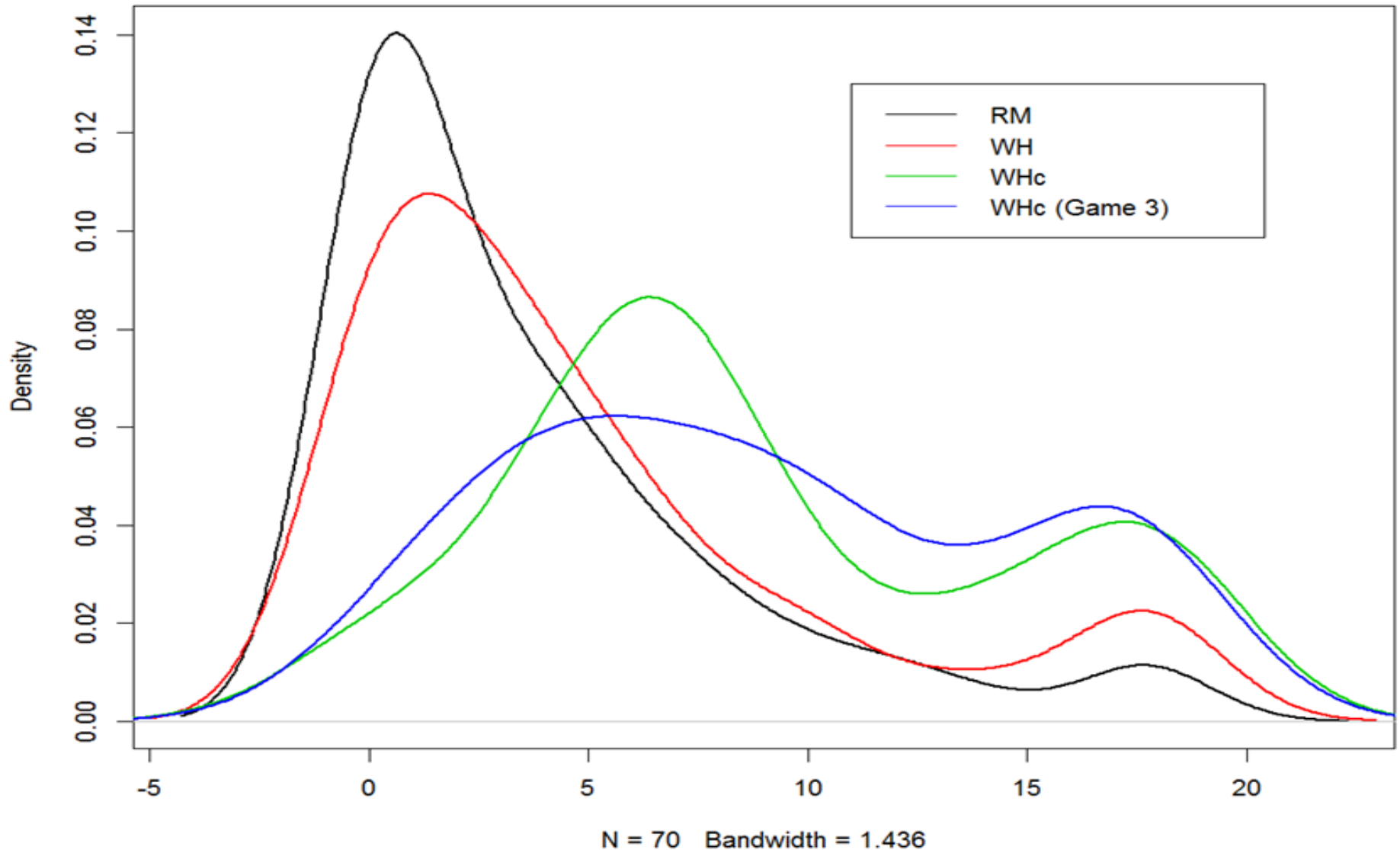


p(C) vs. Avg. Payoff (WHc)



「囚犯困境」合作意願分佈圖

Density of $p(C)$ in Individual Level (Game 2, rounds 6~23)



解決問題(Problem Solving)的流程

定義問題



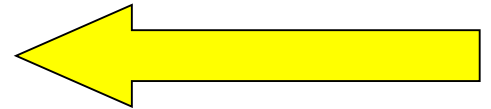
蒐集資料



分析資料



詮釋結果



絕大多數
(統計)教學重心



資料分析的陷阱

■ 如同定義問題，資料分析中也需注意類似陷阱，通常問題有三個主要來源：

(1) Source of Bias (資料偏誤)

→ 蒐集資料、抽樣、以及資料品質

(2) Errors in Methodology (方法瑕疵)

→ 使用錯誤的分析方法

(3) Problems with Interpretation (過度詮釋)

→ 不適當地套用分析結果至不同母體



殘差分析

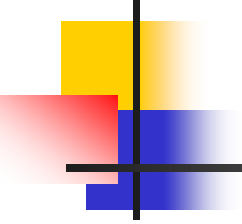
■ 可透過EDA挑選合適的分析方法，通常會再以AIC、BIC等準則比較模型優劣。

→ 交叉驗證、回測法也是常見的評估作法。

■ 估計和預測準確性之外，統計分析會檢查模型假設，確定方法的合理性。

→ 迴歸分析的最後一個步驟，一般會檢查殘差的幾個假設條件。

Q：哪些假設需要檢查、原因？



Identify questions of interest, review
design of study and scope of inference

Explore data graphically

Data Analysis Strategy

Fit model

Check model

Carry out inferences

Communicate results

迴歸分析(續)：何者 R^2 較大？

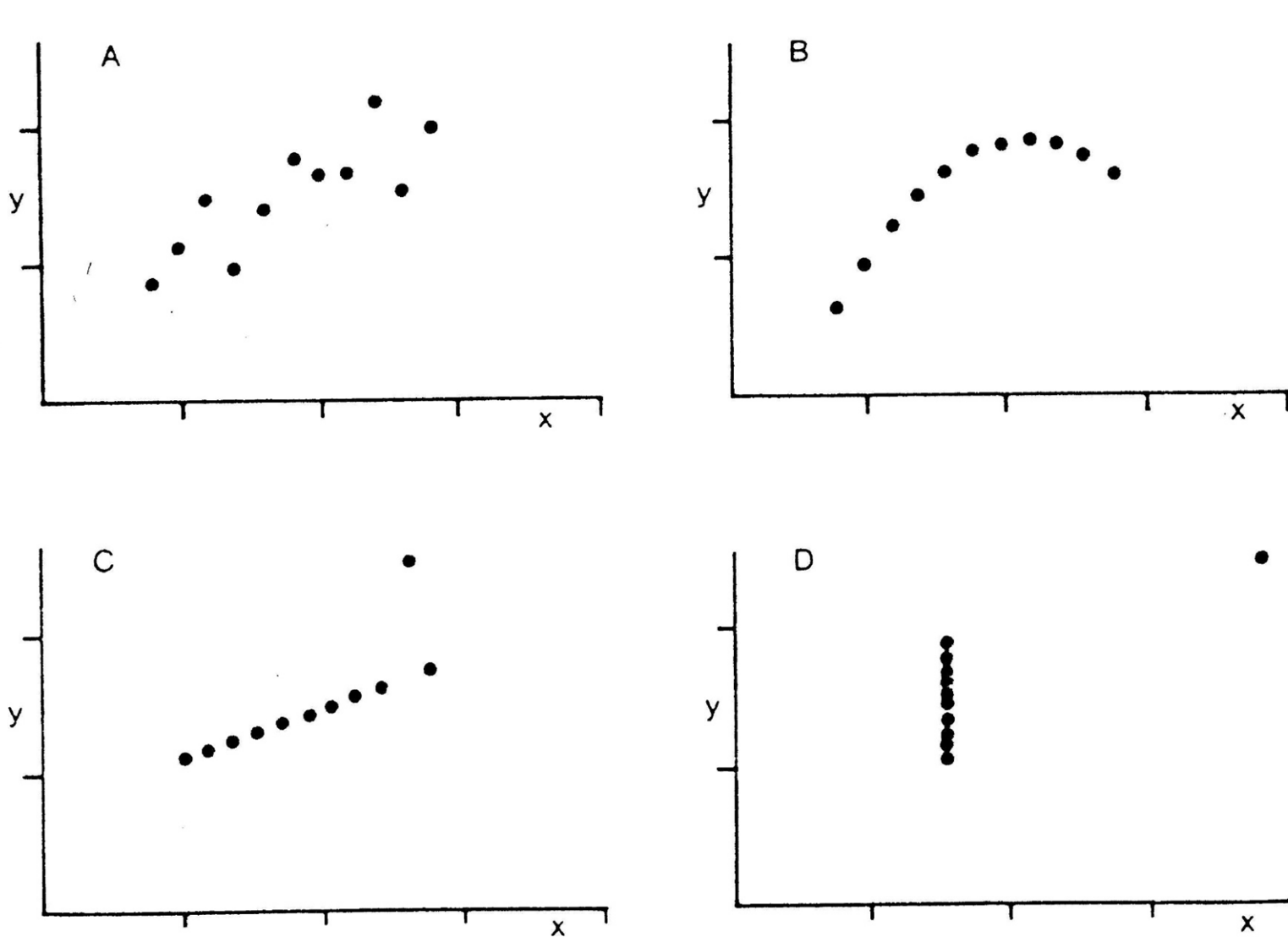
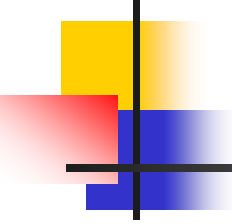


Figure C.2 Four bivariate sets of data.



迴歸分析(續)

- 你相信嗎？！ ($R^2 \cong 0.67$)
→ 由四個圖形中的資料，配適出截距及斜率完全相同的迴歸方程式。
(Anscombe, 1973, 文章中的資料)
- 但你/妳覺得哪一個圖形中
使用迴歸分析？



統計與深度學習模型的準確率小提琴圖

