

商業智慧與人力資源 (Data Analysis and HR)

Spring 2019

政治大學統計系余清祥

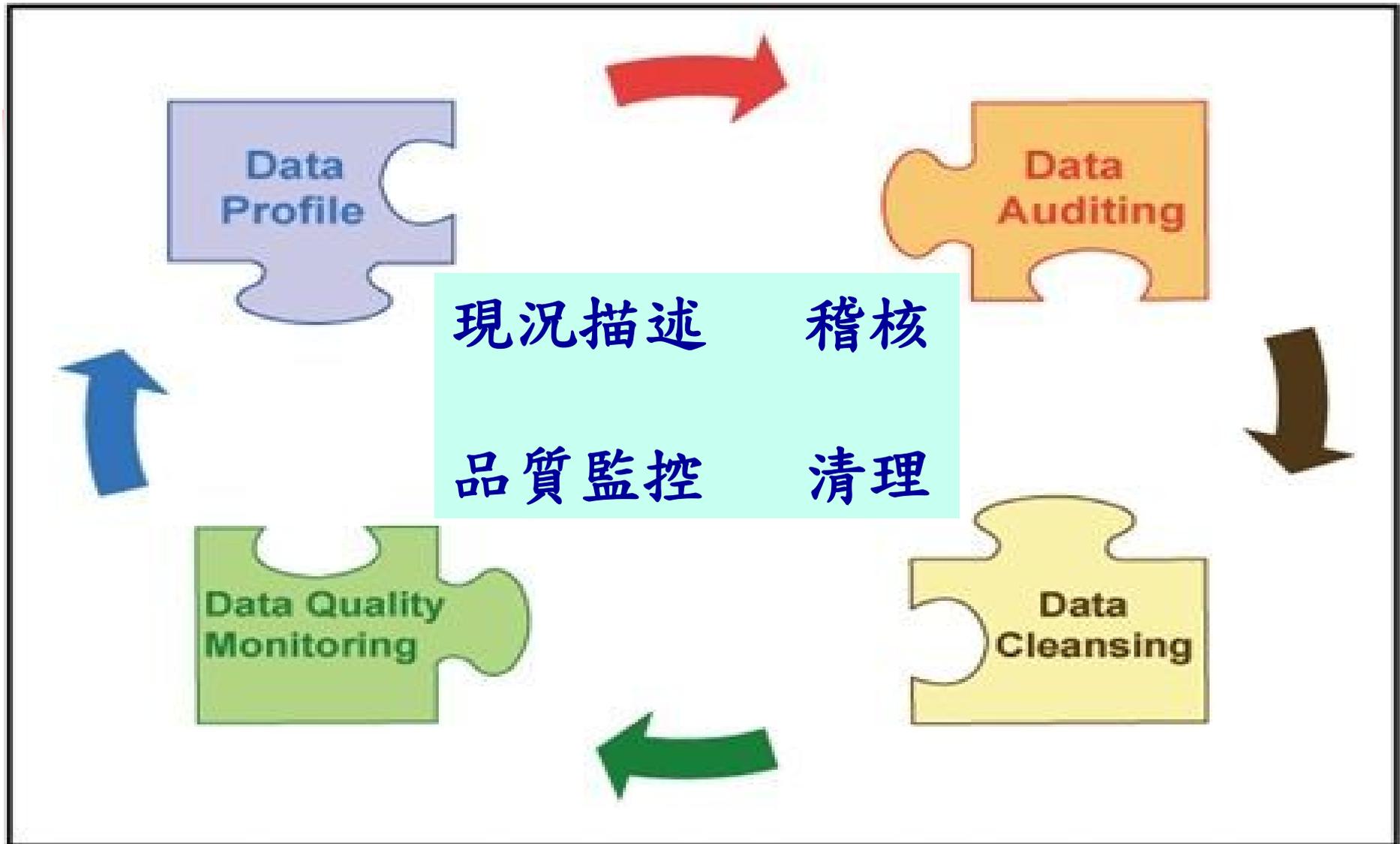
2019年6月20日

Email: csyue@nccu.edu.tw

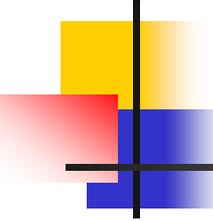
Website: csyue.nccu.edu.tw



商業智慧績效的關鍵－資料品質管理



資料來源：耐特普羅資訊 NetPro的Trinity資料品質管理循環



數量分析

- 透過數理模型描述觀察結果：

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

或

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

- 數量化模型的關鍵：

→ 量化目標值 y ：定義問題！

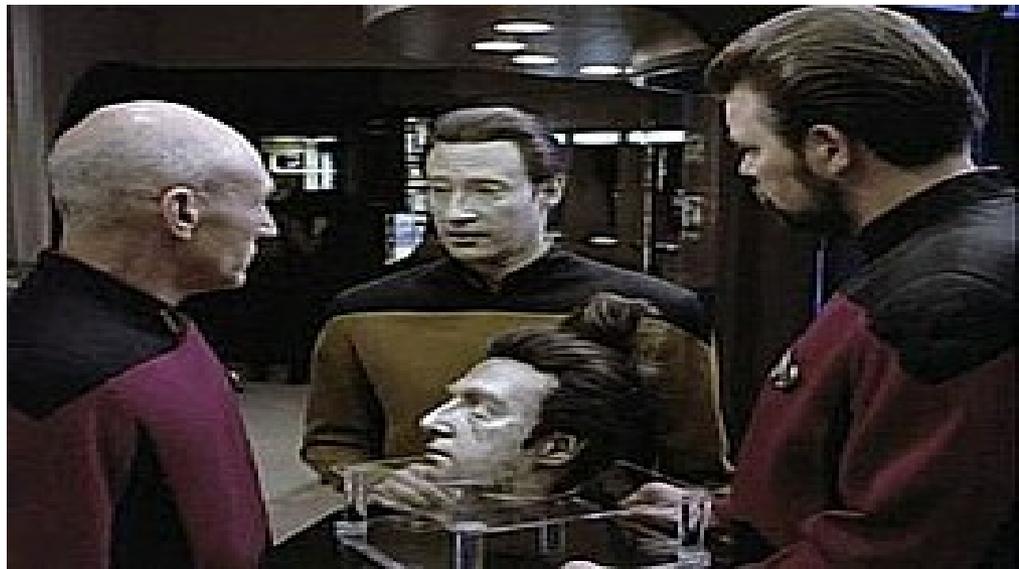
→ 選取關鍵變數： x_1, x_2, \dots, x_p

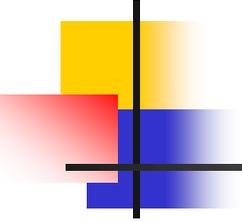
→ 建立量化模型：統計學習、機器學習。

關鍵因素：Y 與 X

- 量化目標為變數 Y ，與 Y 有關的資訊 X 。
 - 目標未必可量化、量化過程間的折損；
 - 無法蒐集資料、 X 與 Y 的關聯不足。
- 研究需要來來回回、嘗試錯誤。

Star Trek: The Next
Generation Data



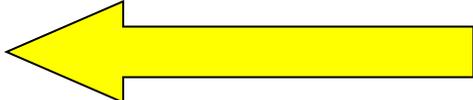


定義問題

蒐集資料

分析資料

詮釋結果



絕大多數的
資料分析重心



資料分析策略

■ 「觀察」、「推論」、「驗證」三步驟

→ 首先檢查資料品質，避免Garbage in, garbage out 的窘境，通常會佔用一半以上的時間。

→ 同時進行其他探索性資料分析(EDA)，一步一步找出資料的重要特性，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

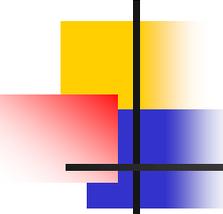
資料輸入錯誤、尋找可能的離群值。

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).

統計與知識

- 統計整理資訊的方法屬於歸納法 (Induction)，從龐雜的資料找出共同趨勢，並區分資料具有以下哪一種特性：



探索性資料分析(EDA)

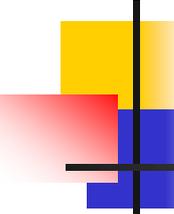
- EDA首要目的在於資料偵錯、獲得資料的大略資訊、驗證已知結果。
- 圖形、表格在EDA中扮演重要的角色；並由分析結果中尋找合適的下一步分析方法。
- 使用任何的統計方法前，先確定該方法需要的假設條件是否滿足。





EDA的目標

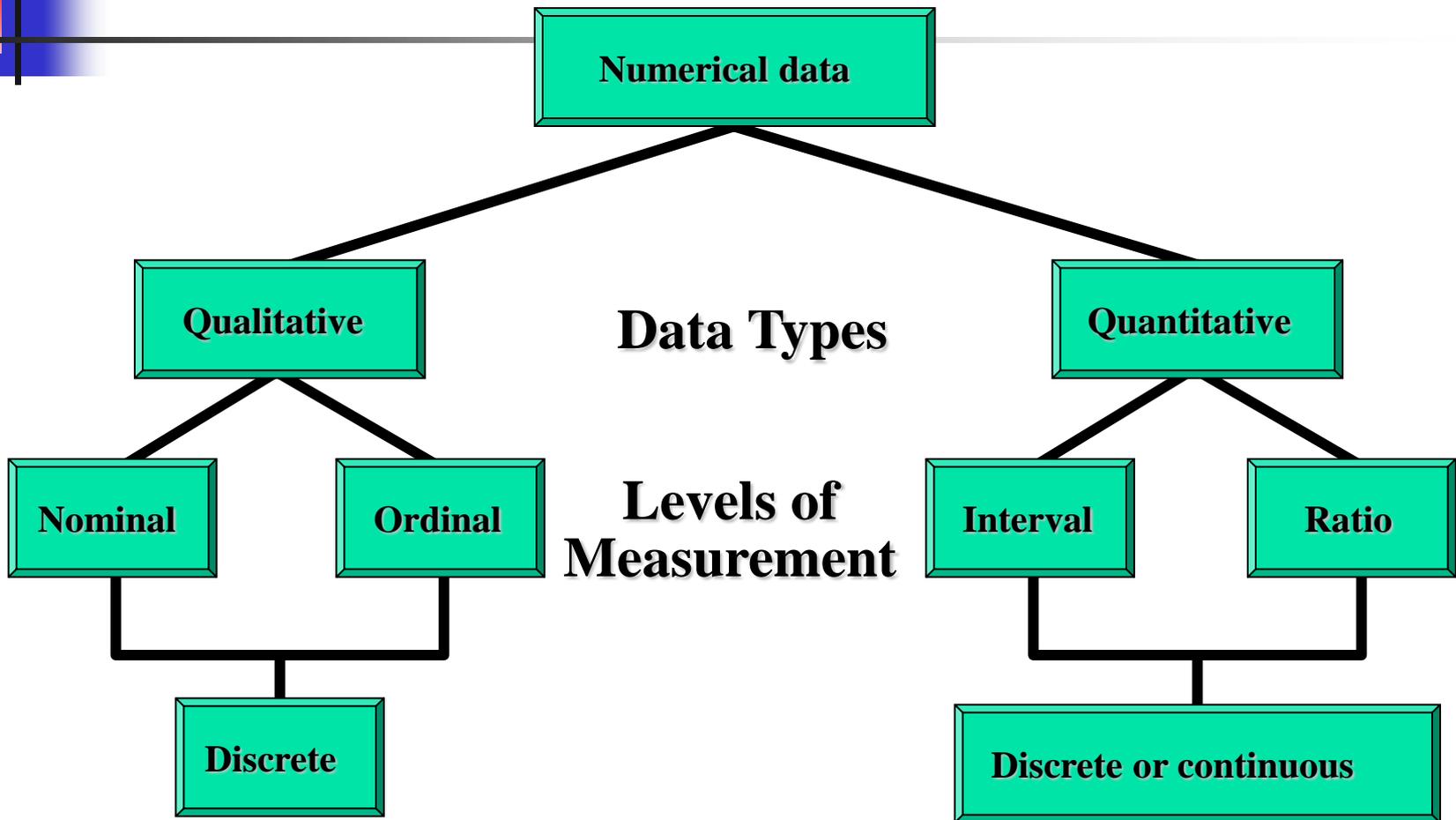
- 資料偵錯
 - 資料輸入錯誤、尋找可能的離群值。
- 初步探索資料的特性
 - 資料的集中、散佈趨勢。
- 驗證已知的結果
 - 是否與已知的結果相同？



EDA的工作項目

- 分析方法與資料本質有關，EDA首先確認資料類型、資料量等特性。
 - 比對編碼簿(Codebook)、異常數值；
 - 樣本與母體的差異（樣本代表性）；
 - 資料基本特性（如期望值、變異數）。
- 資料品質往往是數量分析的成敗關鍵，進行EDA時也應交叉比對分析結果，以不同面向獲得類似結論。

Types of Data(資料類型)



- 並非所有資料都適合常見的統計方法，任意使用分析方法會得出啼笑皆非的結果。

→ 已知 $A > B$, $B > C$ 是否代表 $A > C$?

	甲城市	乙城市	丙城市
A候選人	1	2	3
B候選人	2	3	1
C候選人	3	1	2

註：1代表最喜歡，3代表最不喜歡。



■ 數字定義可能引起的問題：

→ 評審給候選人A、B、C、D的評分。

評審	1	2	3	4	5	6	7	總分
A	4	1	2	4	1	2	4	18
B	3	4	1	3	4	1	3	19
C	2	3	4	2	3	4	2	20
D	1	2	3	1	2	3	1	13



- 候選人D分數明顯最低，刪除後評審重新對候選人A、B、C評分。
→ A的分數反而最高。

評審	1	2	3	4	5	6	7	總分
A	3	1	2	3	1	2	3	15
B	2	3	1	2	3	1	2	14
C	1	2	3	1	2	3	1	13

樣本代表性



- 樣本代表性意指抽出的樣本，其特性與母體非常類似，足以由樣本來代表整個母體。
- 檢查樣本代表性是分析抽樣調查、問卷資料的首要步驟，若樣本與母體差異過大，則以樣本的資訊推測母體的特性，將顯得不合適。

樣本代表性(續)

■ 通常用於檢查樣本代表性的問項：

1. 性別比例
2. 年齡結構
3. 居住地區
4. 教育程度
5. 職業別
6. 婚姻狀態
7. 其他因素



樣本代表性(續)

- 台灣地區的人口統計資料可到內政部統計處查詢，網址<http://www.moi.gov.tw/stat/>

→ 內政統計年報

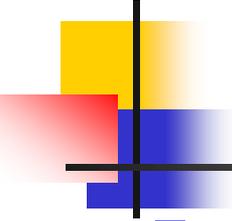
→ 性別統計資料

→ 重要參考指標

註：也可參考主計總處資料，

網址<http://www.dgbas.gov.tw/>



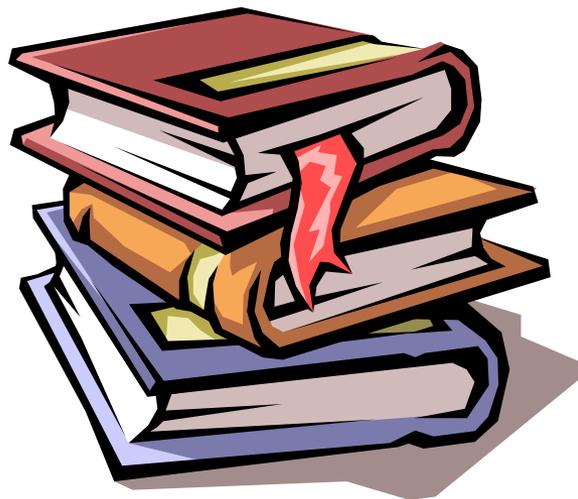


樣本代表性的實例

- 美國1930年代調查業巨人《文摘》(Literary Digest)，在1936年發出一千萬問卷，預言共和黨總統候選人Alfred Landon將大獲全勝，擊敗民主黨羅斯福。
 - 調查對象為擁有汽車與電話的家庭。
 - 但蓋洛普等三大業者1948年也錯誤預測共和黨杜依會獲勝。
 - 使用配額抽樣法，訪問五百位家庭主婦、二十位農夫與三百位老人。
- 註：資料來源【真實的謊言】

分析案例

- 政治大學統計系大學部三年級學生在民國90年，調查政治大學學生對流行性雜誌的偏好，包括雜誌價格、內容、出版期數等問項。



1. 研究對象：國立政治大學全體學生—因為我們身處在政大校園裡面，對於同校同學們，他們對於雜誌的傾向為和是我們所關心的話題。而且目前坊間得雜誌更是千奇百樣，尤其以流行性雜誌一本接一本的發行，更為獨特。所以我們選擇以政大全體學生來做為我們研究的對象，來了解他們對於各類型雜誌及流行性雜誌的偏好情形。

2. 抽樣方法：先分層抽樣在行群集抽樣

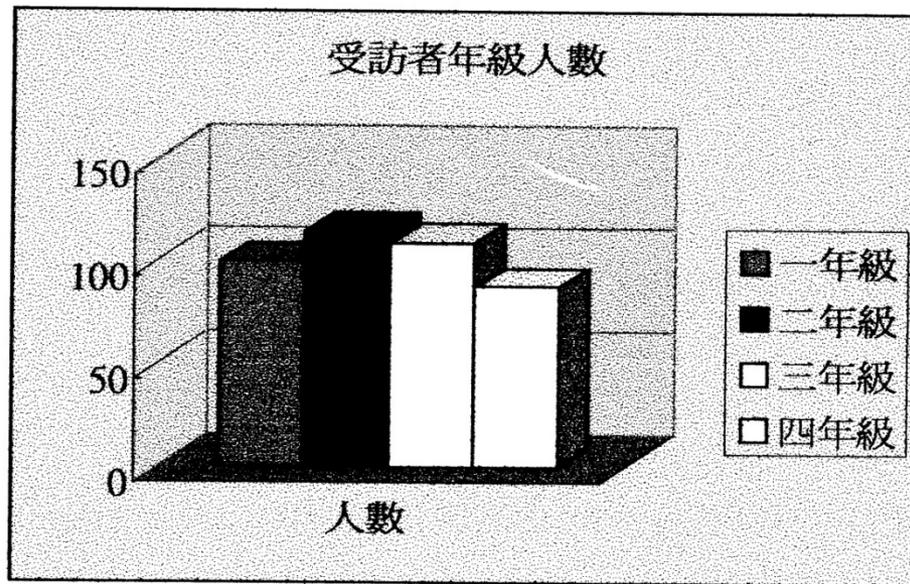
本次抽樣調查因我們認為年級以及學院在雜誌的閱讀習慣上會有差別，而同一學院內的系別不會有差異，所以把學院及年級同時做為分層標準，再針對每一學院的比例抽出等比例的樣本數，進行調查。母體即為全政大之學生。其中，抽出各學院的人數比為：

學院	人數	有效份數
文學院	60	52
社會科學院	120	95
商學院	150	126
傳播學院	40	34
法學院	30	21
理學院	30	24
國際事務學院	15	15
外國語文學院	55	45
總計	500	412

$$\text{有效問卷率} = 500 - (\text{無效問卷} + \text{未回收問卷}) / 500 = 0.824$$

* 無效問卷：答題不完整、單題複選皆列入無效問卷（71份）

* 未回收問卷：17份



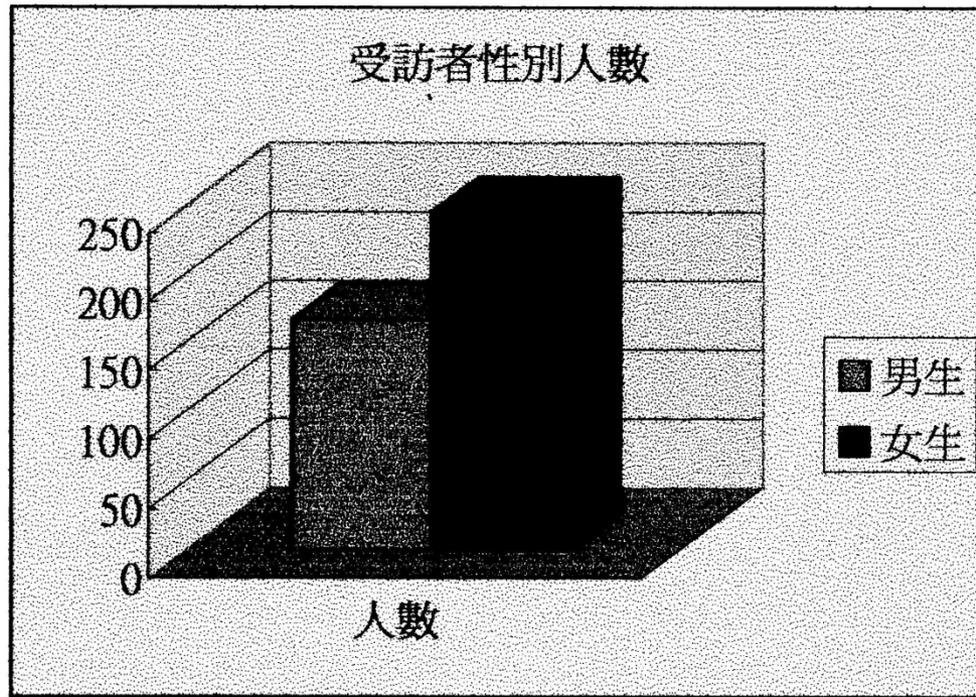
各年級中，以二年級受訪人數所佔比例最高，約有 27.9%，其次為三年級 26.5%。最少則為四年級，只有 21.4%。

	一年級	二年級	三年級	四年級
人數	100	115	109	88

※檢查年級的分布比例

	男	女	總計	百分比	樣本百分比
一年級	904	1470	2374	24.0%	24.3%
二年級	922	1587	2509	25.4%	27.9%
三年級	963	1376	2339	23.7%	26.5%
四年級	1108	1553	2661	26.9%	21.4%

利用 Goodness-of-fit test，得到 $X^2=7.059 > X^2_{3,0.9}=6.25$ ，我們發現在此一部分的檢查，樣本跟母體的比例是有差異的，樣本代表性可能不太足夠。但是由於在母體中的四年級包含了延畢生，所以在人數方面自然會比真正四年級人數高，因此如果在抽樣時能加入這項影響因素的話，整份問卷會更有其代表性。



在受訪者中，男生佔了41%，而女生的比例則為59%。性別比，男：女約為2：3。

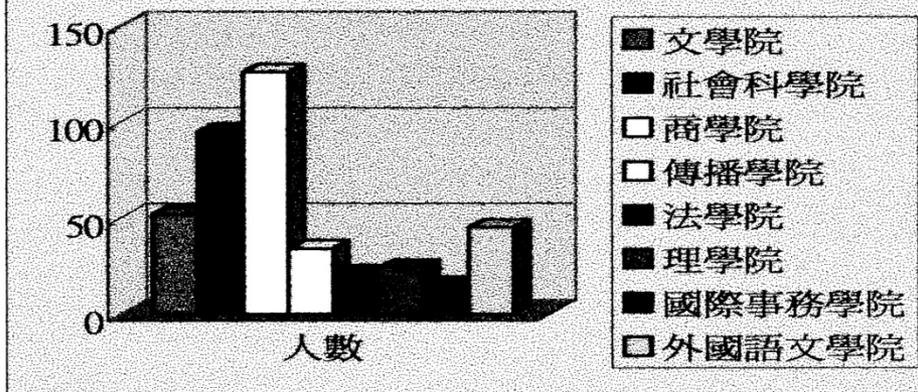
	男生	女生
人數	167	245

※檢查男女比例

全政大為母體下，男女比例為40.5：59.5 (3897：5986)；而樣本比例為41：59 (167：245)

由Goodness-of-fit test 檢定的結果， $X^2=0.21024 < X^2_{1,0.9}=2.71$ ，可知男女的採樣方面的小誤差是可接受的，即在男女方面，樣本代表性為足夠的。

受訪者學院人數



受訪者學院以商學院居多，佔30.8%，其次為社會科學院佔23.1%，最少的為國際事務學院，佔3.6%

	文學院	社會科學院	商學院	傳播學院
人數	52	95	126	34
	法學院	理學院	國際事務學院	外國語文學院
人數	21	24	15	45

※檢查學院的分布比例

	男	女	總計	百分比	樣本百分比
文學院	330	856	1186	12.00%	12.4%
社會科學院	975	1369	2344	23.72%	23.1%
商學院	1250	1704	2954	29.89%	30.8%
傳播學院	269	597	866	8.76%	8.3%
法學院	304	305	609	6.16%	5.1%
理學院	422	126	548	5.55%	5.8%
國際事務學院	94	174	268	2.71%	3.6%
外國語文學院	253	855	1108	11.21%	10.9%

利用Goodness-of-fit test，得到 $X^2=2.5264 < X^2_{7,0.9}=12.02$ ，所以能推之我們的假設「母體比例與樣本比例」是相同的。在學院這一方面可以說樣本對母體的代表性是足夠的。

2. 獨立性檢定

根據我們的目的及興趣，利用基本資料和問卷內容中的項目做獨立性檢定，以及利用問卷內容中題目對題目做獨立性檢定，以探討兩個分類間的相關性程度如何。其中基本資料有：

* 一個月可使用的金錢：\$0-\$4000、\$4001-\$8000、\$8000-\$12000、\$12001以上

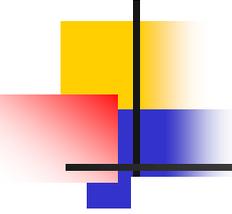
* 學院：文學院、社會科學院、商學院、傳播學院、法學院、理學院、國際事務學院、外國語文學院、

* 年級：一年級、二年級、三年級、四年級

* 性別：男生、女生

分類一	分類二	chi-square 值	P-value
一個月可使用金錢	多久看一次雜誌	15.396	0.081
一個月可使用金錢	每月平均花多少錢買雜誌	23.982	0.020
一個月可使用金錢	流行性雜誌的合理價格	26.252	0.010
一個月可使用金錢	流行性雜誌的合理頁數	28.921	0.004
學院	多久看一次雜誌	39.381	0.009
學院	最常閱讀的雜誌類型	63.440	0.080
學院	流行性雜誌內容重點	62.538	0.021
學院	流行性雜誌的出刊頻率	39.397	0.075
年級	平常有無看雜誌	9.285	0.026
年級	最常閱讀的雜誌類型	1.367	0.071
年級	不閱讀流行性雜誌的原因	8.950	0.030
年級	閱讀流行性雜誌的原因	18.707	0.096
年級	流行性雜誌的合理頁數	22.187	0.035
性別	最常閱讀的雜誌類型	61.759	0.000
性別	有無看過流行性雜誌	27.503	0.000
性別	閱讀流行性雜誌的原因	15.001	0.005
性別	流行性雜誌內容重點	14.103	0.029
性別	流行性雜誌的合理價格	11.122	0.025
性別	流行性雜誌的出刊頻率	10.890	0.028
多久看一次雜誌	流行性雜誌出刊的頻率	9.243	0.682
每月平均花多少錢買雜誌	流行性雜誌的合理價格	114.091	0.000
有無看過流行性雜誌	是否會因增加優惠而購買	14.858	0.000
不閱讀流行性雜誌的原因	是否會因增加優惠而購買	2.159	0.142

上述的獨立性檢定過程可見於後面的【分項說明】或【附錄】由於我們所使用的抽樣方法為先分層在群集抽樣，所以原先設定若以400份有效問卷做為樣本數時，樣本錯誤率 $\alpha = 0.05$ 應比實際為小，所以我們把 α 擴大為0.1，若其p-value $< \alpha$ ，則表示兩種分類間有相關性存在。



抽取1,000份樣本的原因

- 民意、市場調查的多為封閉問卷，有興趣的多為某個問項佔的比例，例如：某位候選人的支持程度→二項分配。

→信心水準95%及最大誤差不大於3%：

$$1.96 \times \sqrt{\frac{p(1-p)}{n}} \leq 0.03$$

$$\Leftrightarrow \sqrt{n} \geq \frac{1.96 \sqrt{p(1-p)}}{0.03} \cong \frac{1.96 \times 1/2}{0.03}$$

$$\Leftrightarrow n \geq 1,067$$



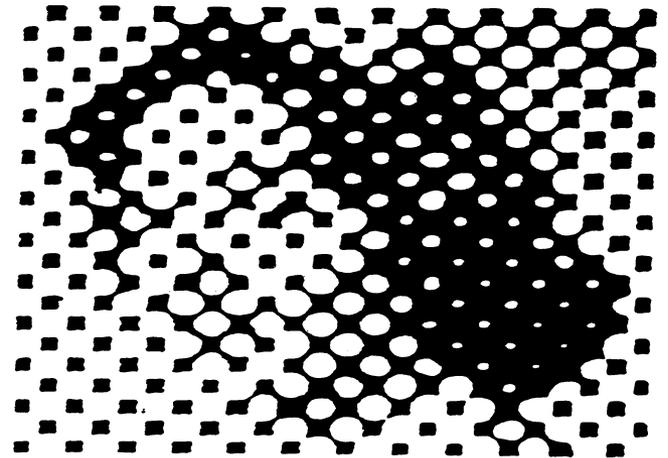
500,000



2,000



1,000



250

樣本對母體之代表性

敘述統計量(範例)

例題一、試以文字詮釋以下隨機抽出某公司業務部門20位員工的年齡：

41	25	25	33	27	31	42
35	36	32	36	41	34	29
34	31	34	35	32	35	

→ 平均數 = 33.4，中位數 = 34.0，
標準差 = 4.75，全距 = 17。



敘述統計量(續)

例題二、試以文字詮釋以下隨機抽出某公司20位員工去年請假的天數：

0 0 0 0 0 0 0 0 1 1
1 2 2 3 4 5 5 6 7 42



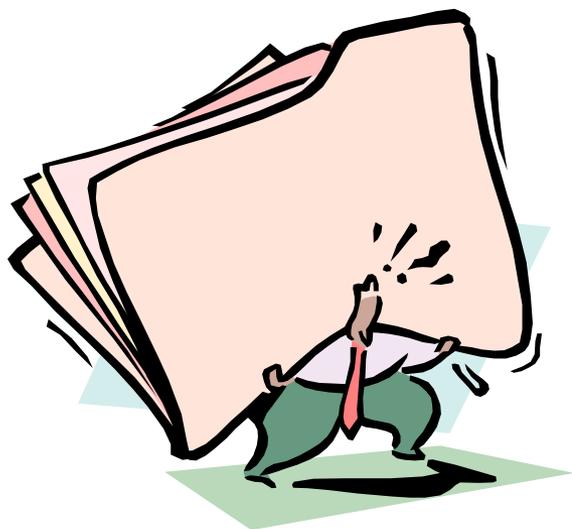
→ 你/妳看到了甚麼現象？

敘述統計量(續)

例題三、街頭隨機訪問20位成年受訪者去年閱讀某月刊的期數：

0 1 11 0 0 0 2 12 0 0
12 1 0 0 0 0 12 0 11 0

→ 請問這是甚麼樣的月刊？



敘述統計量(續)

例題四、以下為隨機抽出某地區16位孕婦的身高(單位：公尺)：

1.57 1.55 1.60 1.52 1.68 1.57 1.62 1.55
1.65 1.52 2.55 1.60 1.55 1.60 1.62 1.57

→ 請問你/妳看到資料有何特性？



圖形與表格

- 除了基本的敘述統計量外，圖形與表格可以輔助判斷資料的特性。

→ 常見的圖形：Boxplot、Histogram

- 這些圖表看似簡單，但仔細判讀仍可發現重要訊息，甚至不需進階統計分析，即能約略猜出分析的結論。

註：A picture is worth a thousand words!



圖形有時描述地更傳神！

A Picture



is worth
=
=

Creamy, delicious, yummy,
fudge ice cream, smooth,
chocolate-chip mint ice
cream, strawberry ice
cream with real chunks of
strawberry, colored sugar
sprinkles, waffle sugar
cone, sweet, wonderful,
tastes great, cold, nice to
eat, dessert, good yummy
toppings, chocolate
sprinkles, comforting,
good fun, dipping, terrific,

A thousand words.

©2003 E. Aoyama

Source: http://www.rapides.k12.la.us/nitro/visual_literacy.htm



Source: <http://rbonini.files.wordpress.com/2007/06/200706290853.jpg>

REMAIN

48.11%

LEAVE

51.89%

382 / 382 districts reporting

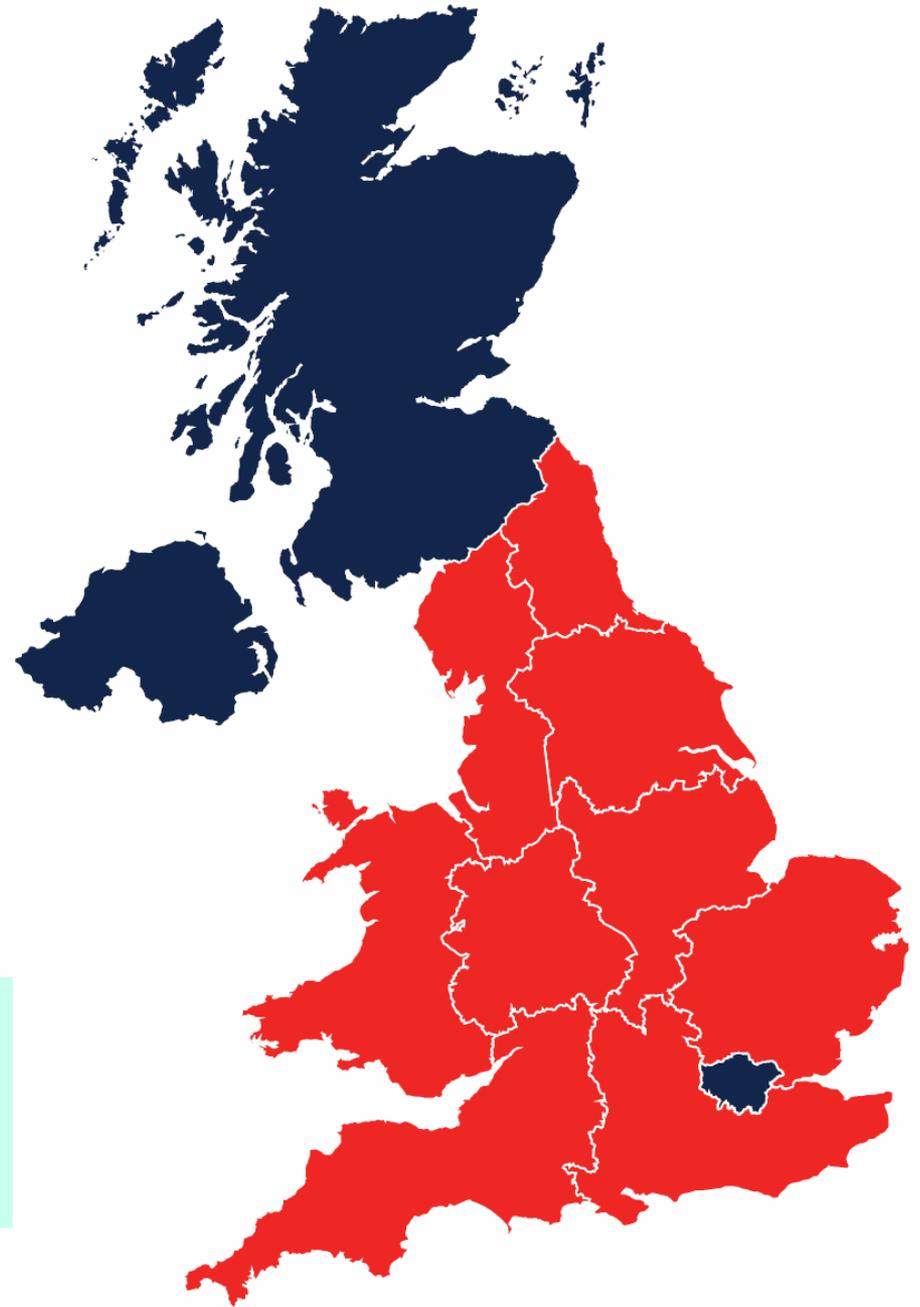
Breakdown by region

South East

Remain 2,391,718

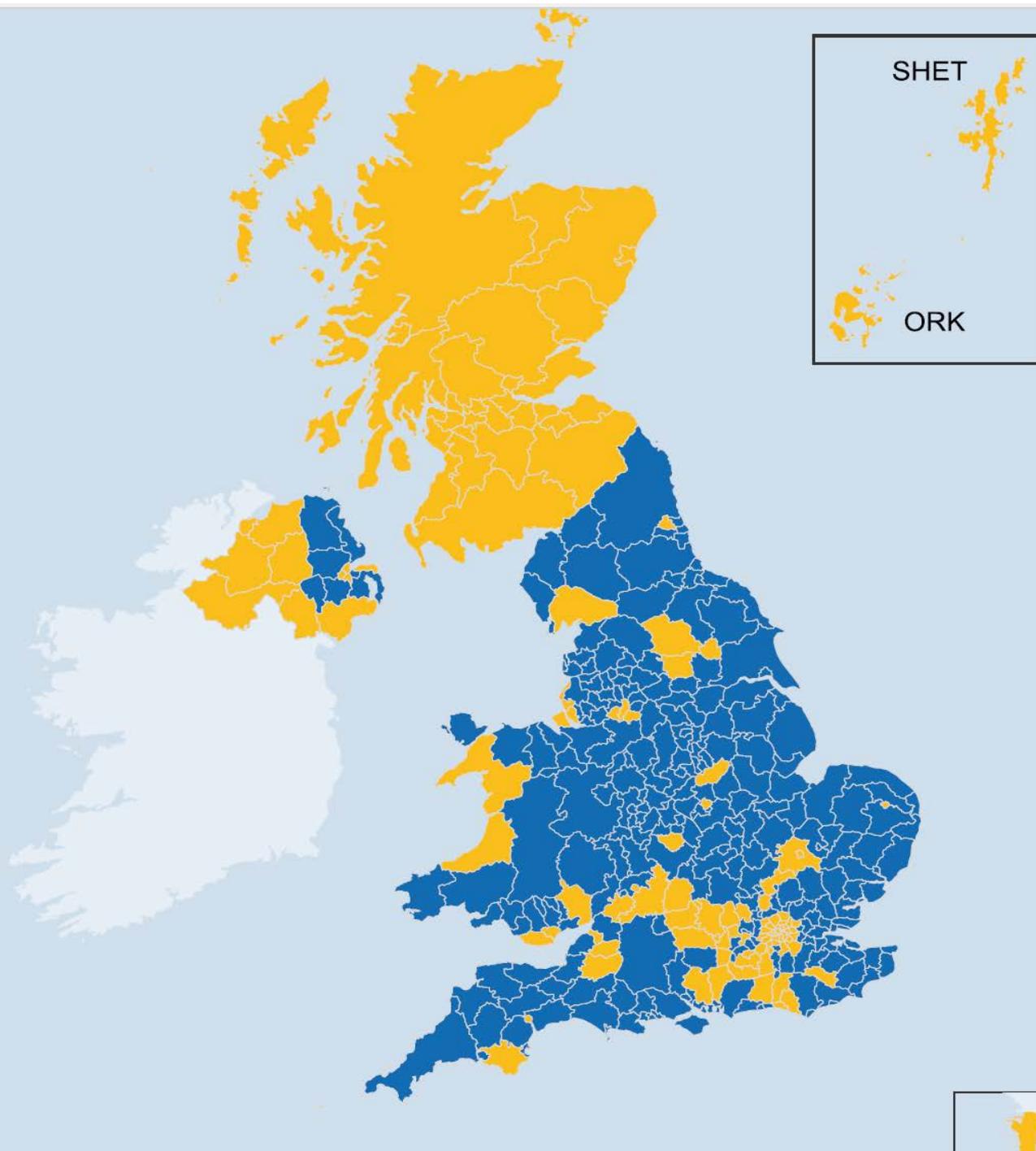
Leave 2,567,965

67 / 67 districts reporting



英國脫歐公投結果
(CNN)

英國脫歐公投結果(BBC)



England

Leave **53.4%**
15,188,406 VOTES

Northern Ireland

Leave **44.2%**
349,442 VOTES

Scotland

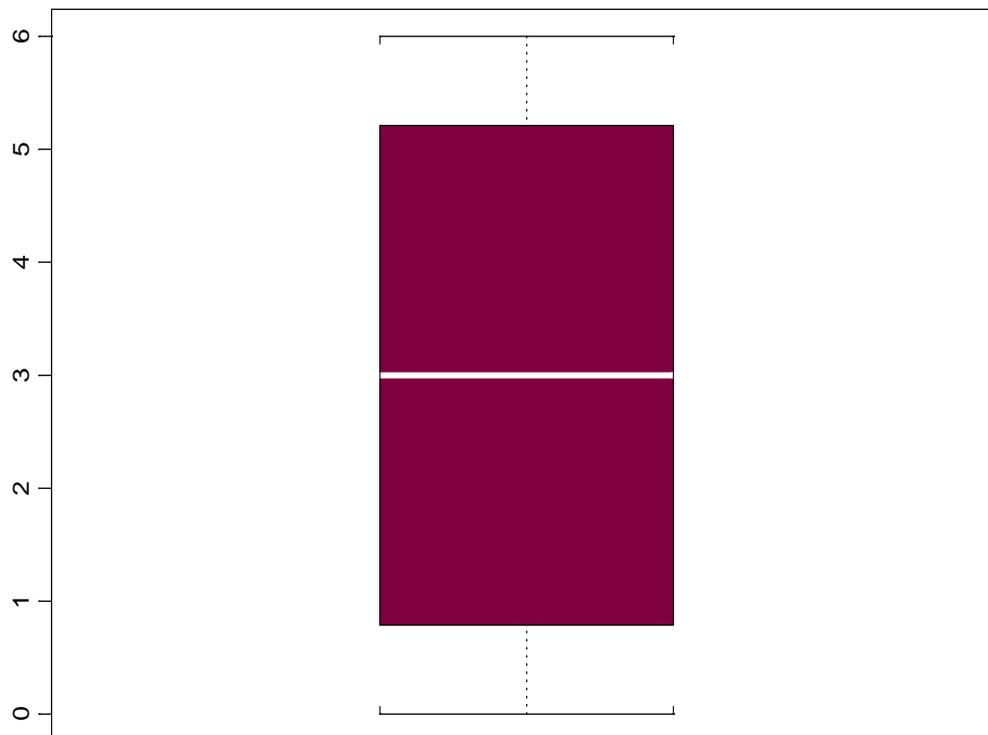
Leave **38.0%**
1,018,322 VOTES

Wales

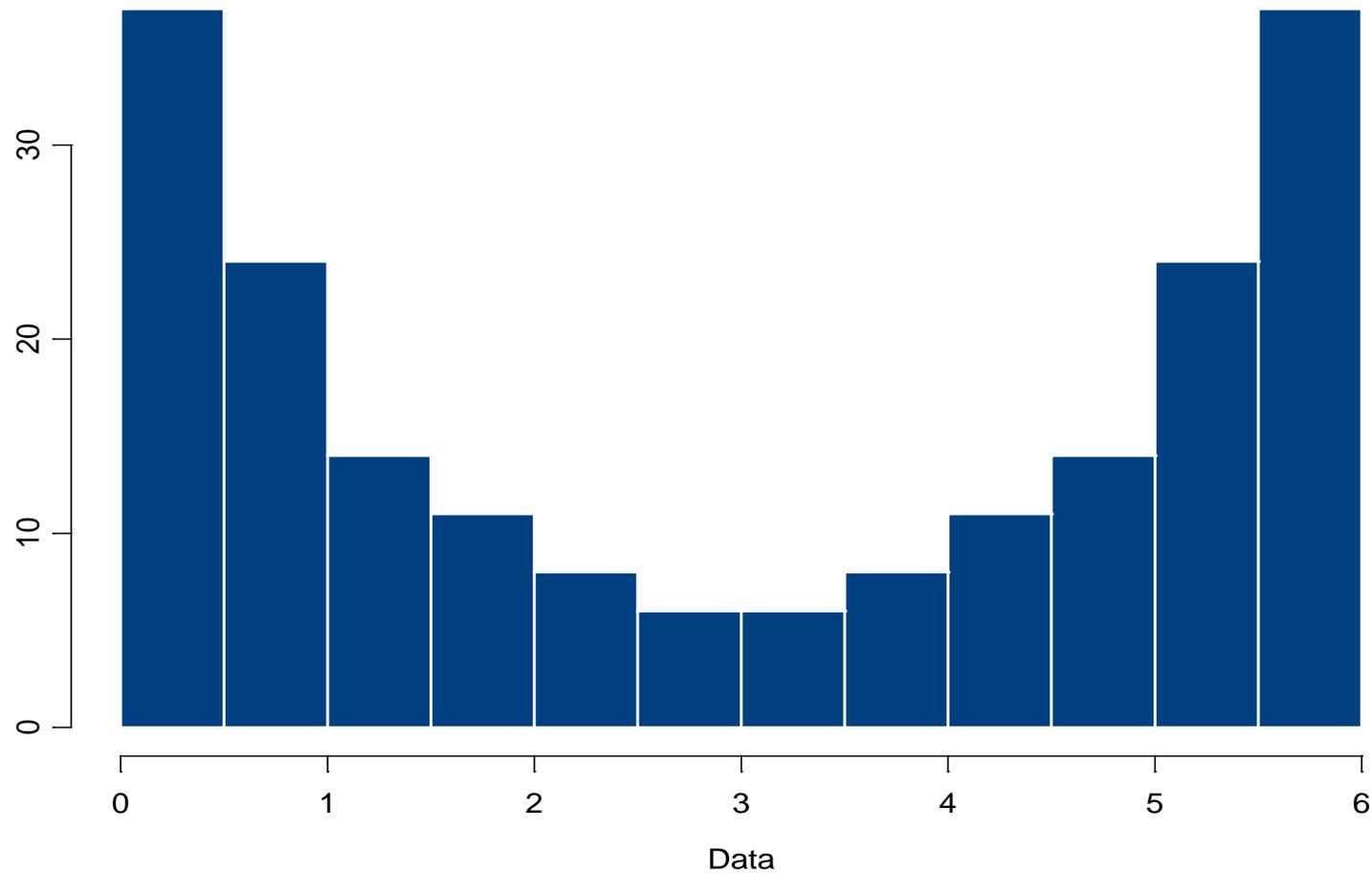
Leave **52.5%**
854,572 VOTES

表格與圖形(範例)

例題五、大略敘述下圖資料的特性。



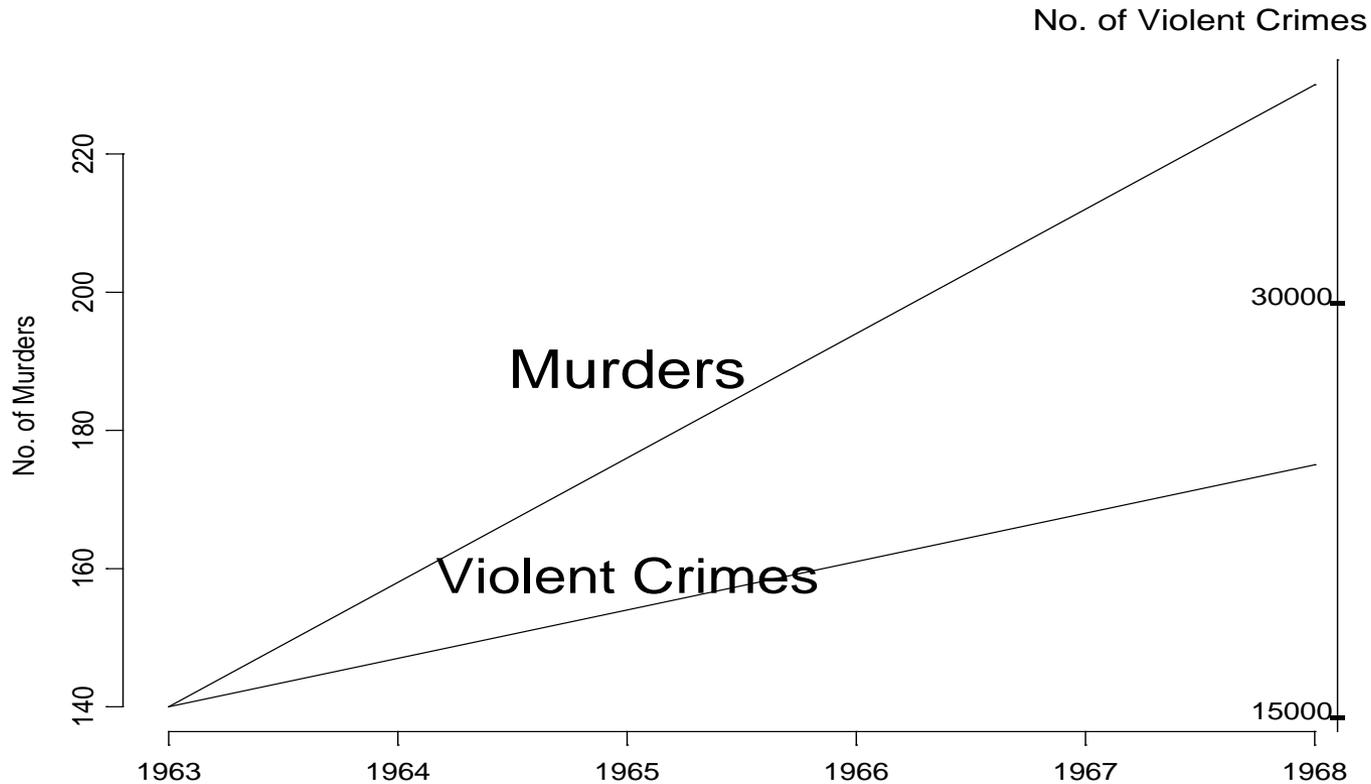
原始資料(你/妳猜對了嗎?)



問題：常態分配的箱型圖具有哪些特性？

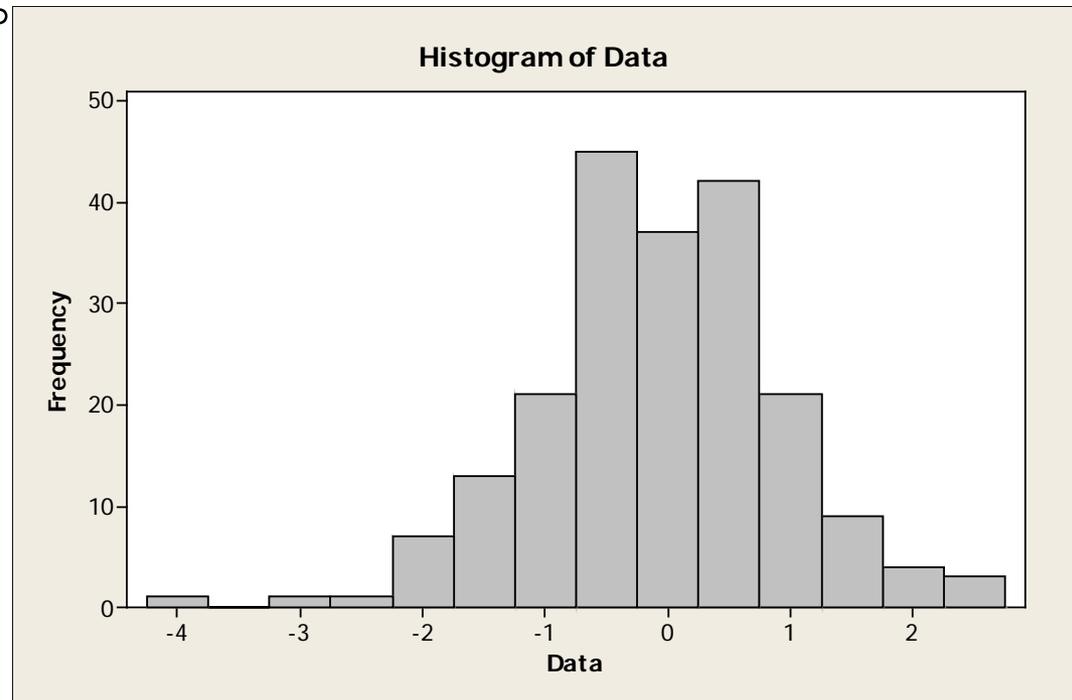
表格與圖形(續)

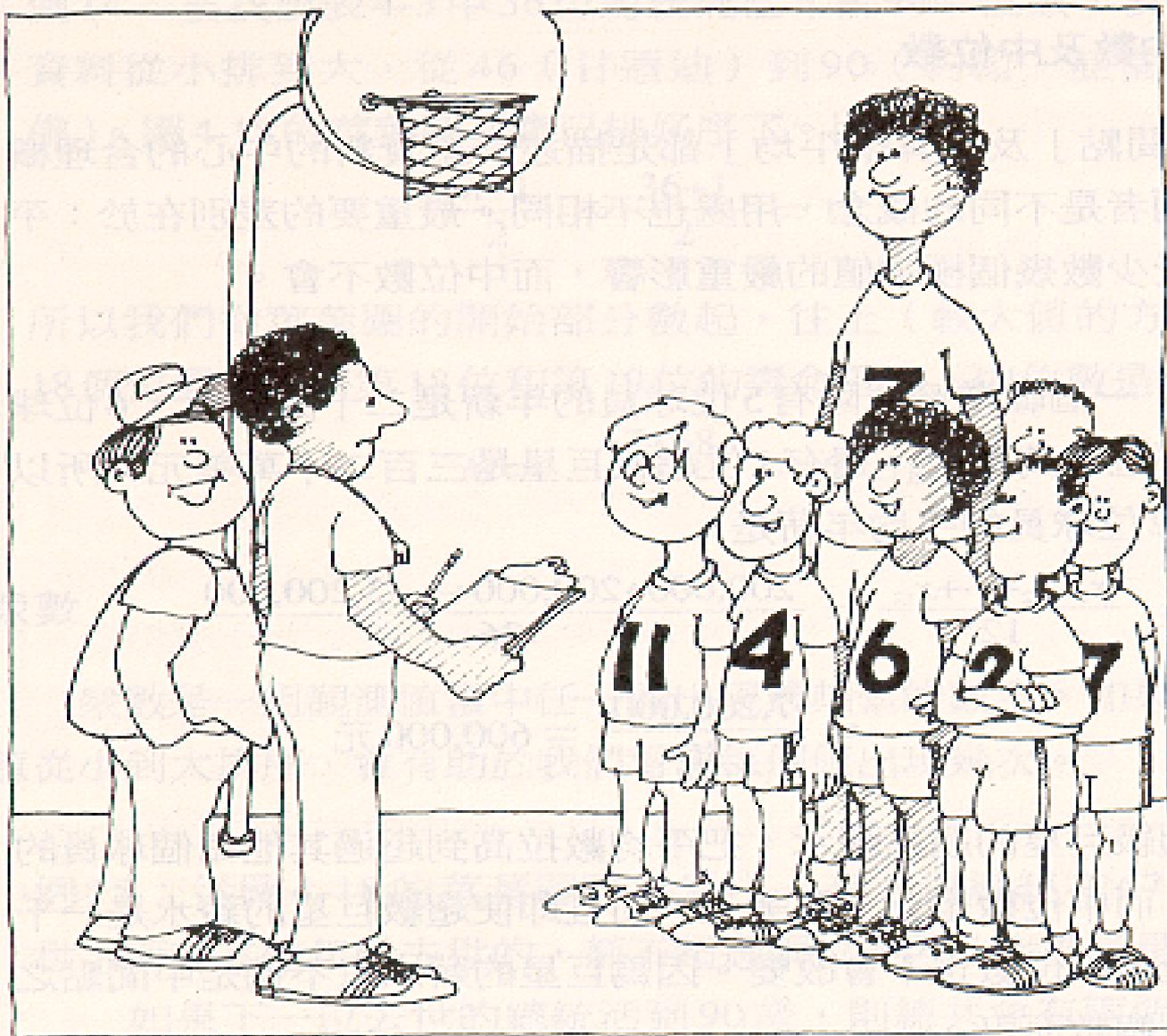
例題六、謀殺率上升嚴重嗎？



平均數與中位數

- 左偏(Left-skewed)：當少數觀察明顯值較一般觀察值小時(如下圖)，平均數將被這些觀察值拉下，但中位數較不受影響，此時中位數大於平均數。
- 同理，少數觀察值較大時則稱為右偏(Right-skewed)。

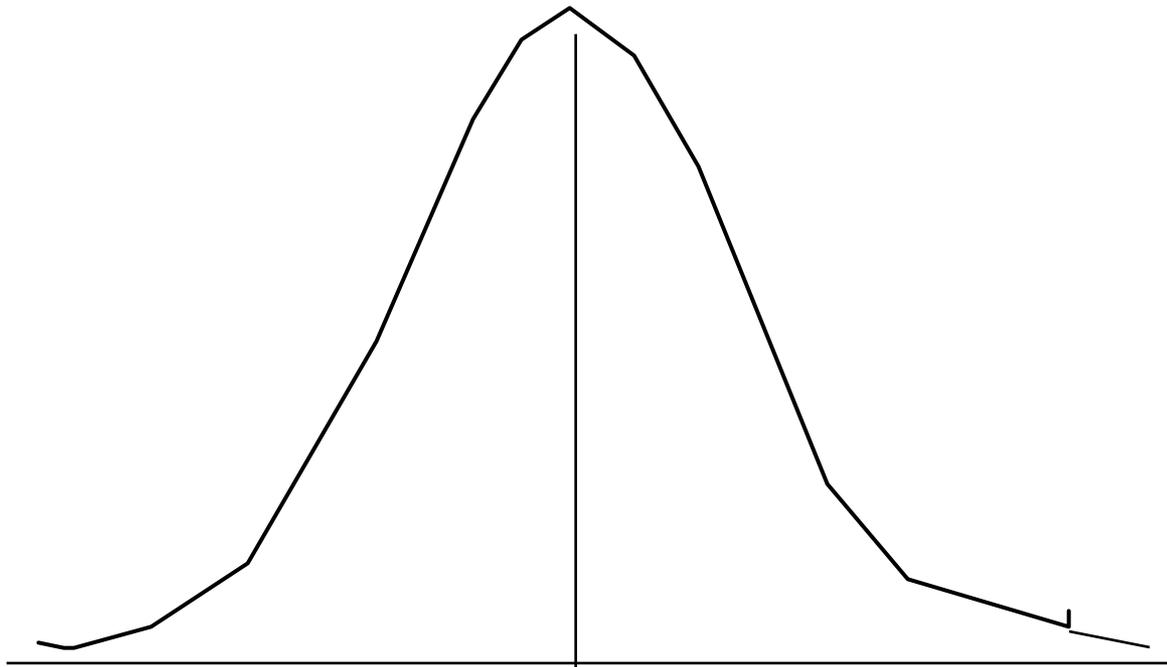




「我們是應該宣布我們的平均高度來嚇死對手，還是宣布我們的中位數高度來消除他們的戒心呢？」

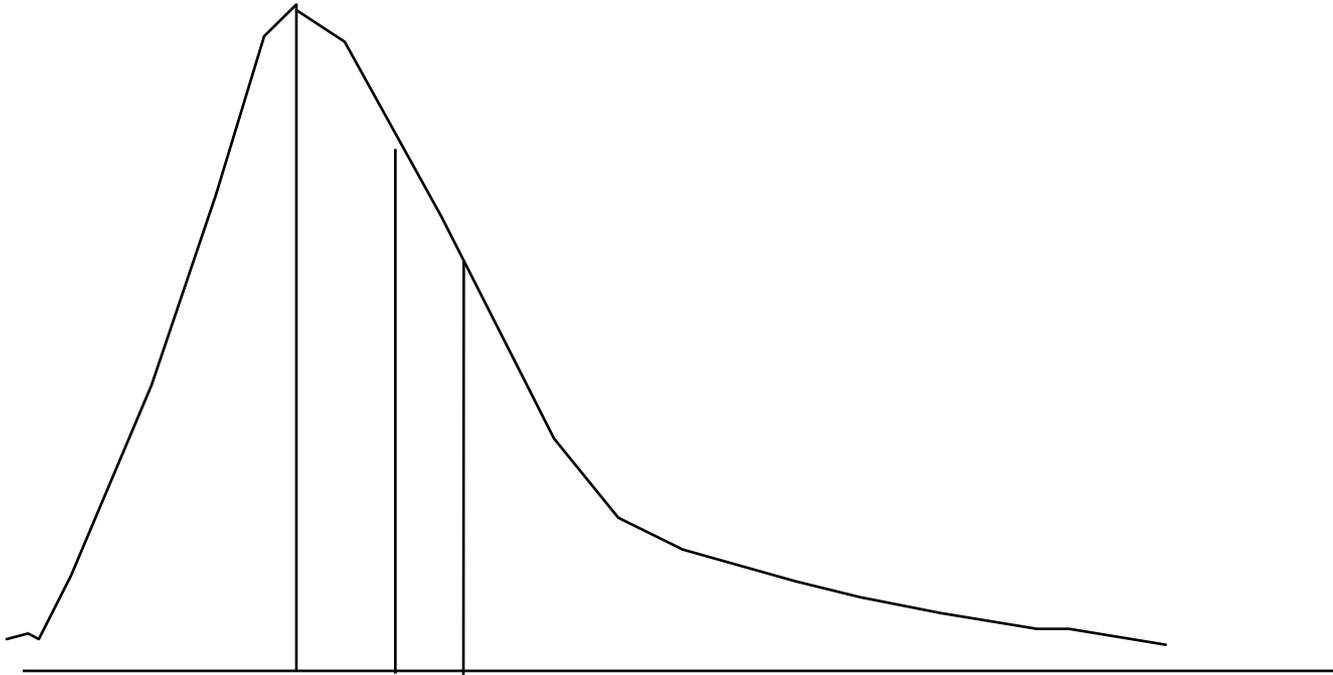
對稱分配(Symmetric Distribution)

■ 無偏態 眾數 = 中位數 = 平均數



右偏分配(Skewed to the Right)

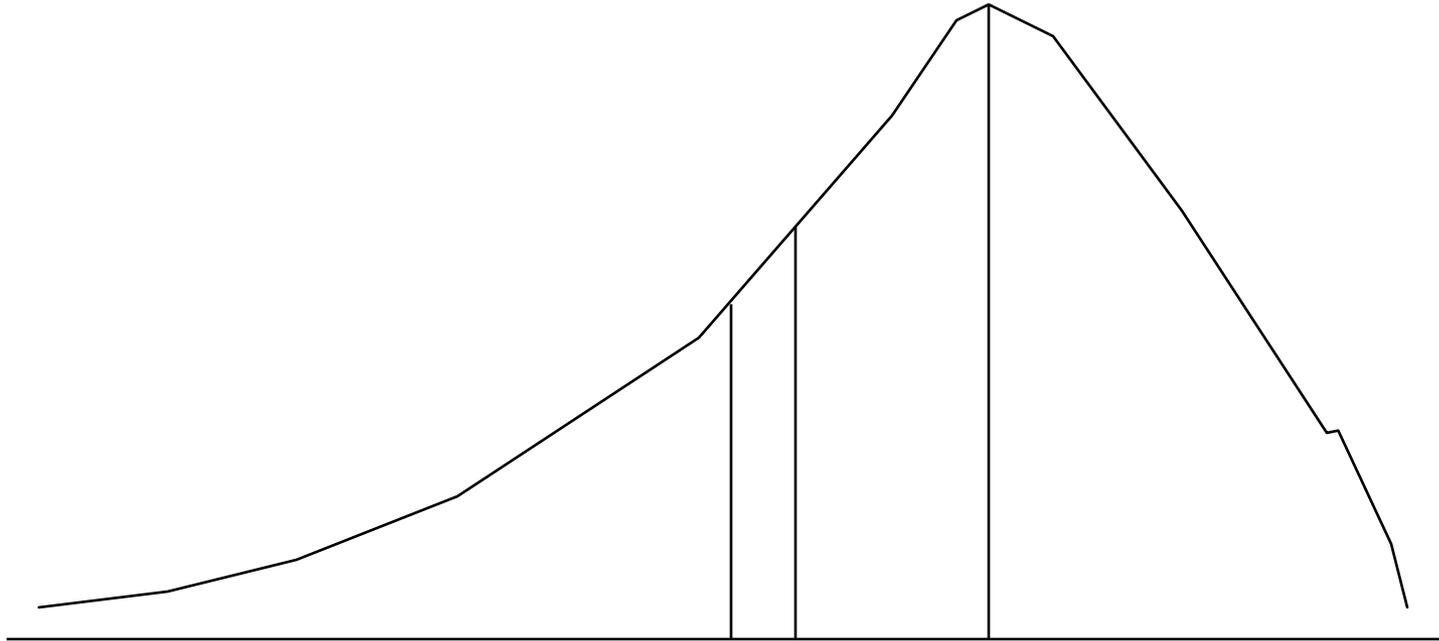
- 右偏分配：平均數和中位數在眾數的右邊。



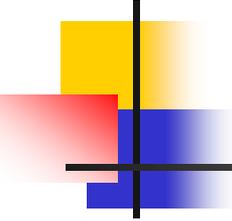
- 眾數 < 中位數 < 平均數

左偏分配(Skewed to the Left)

- 左偏分配：平均數和中位數在眾數的左邊。



- 平均數 < 中位數 < 眾數

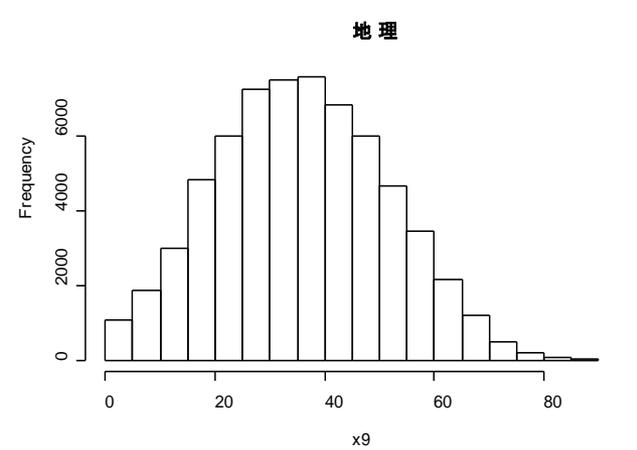
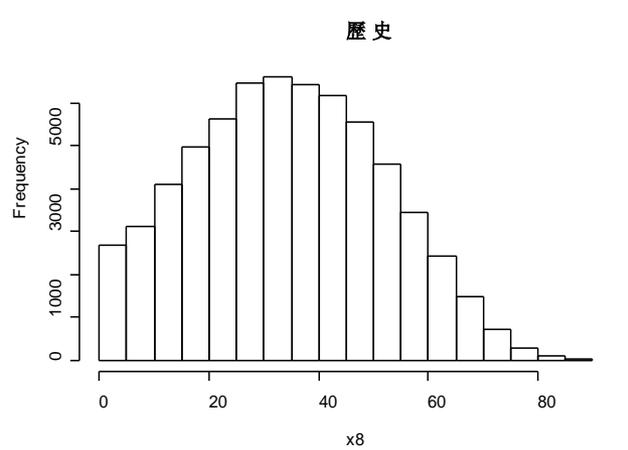
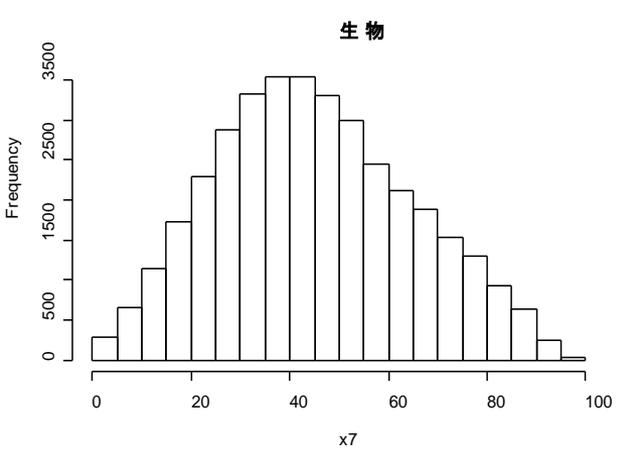
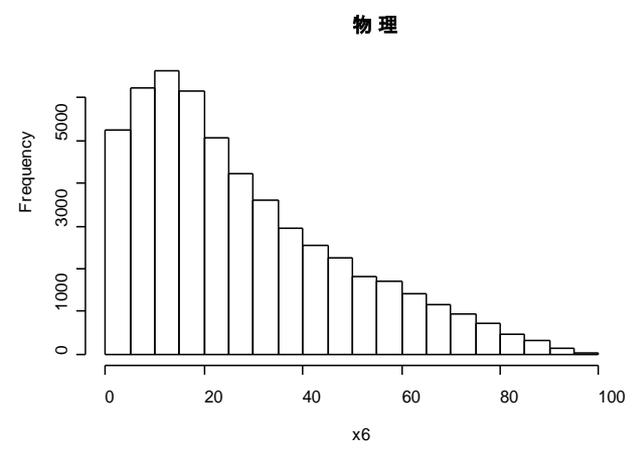
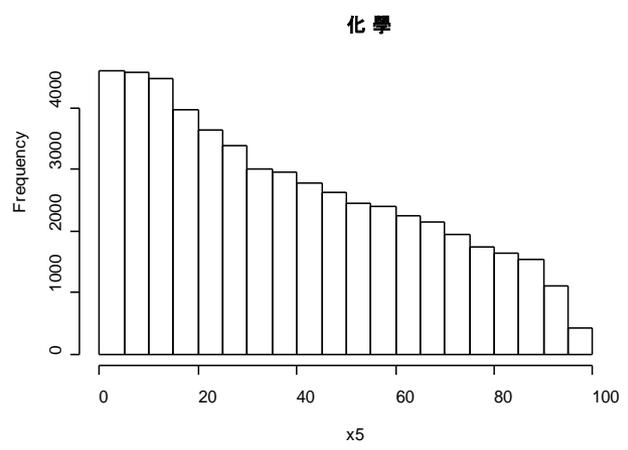
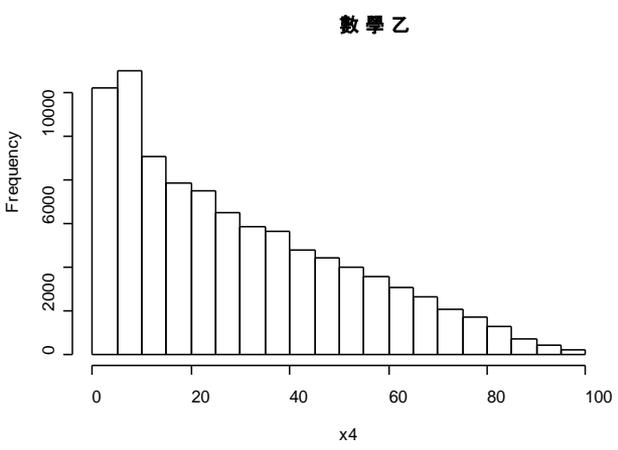
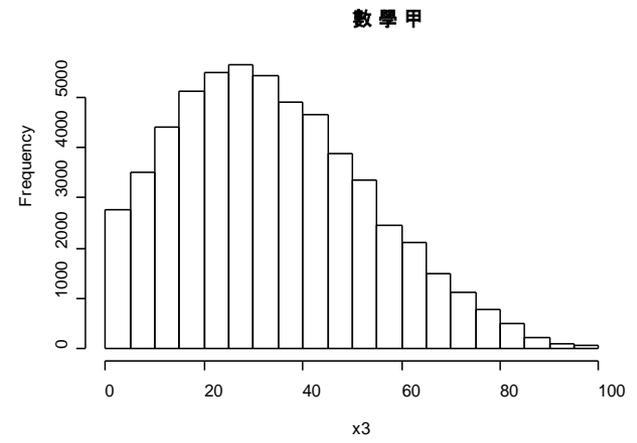
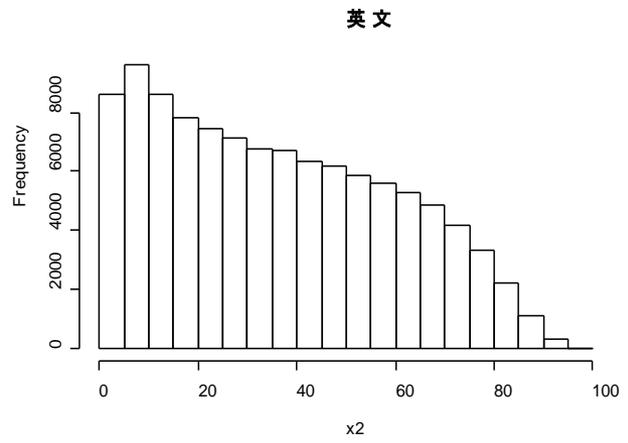
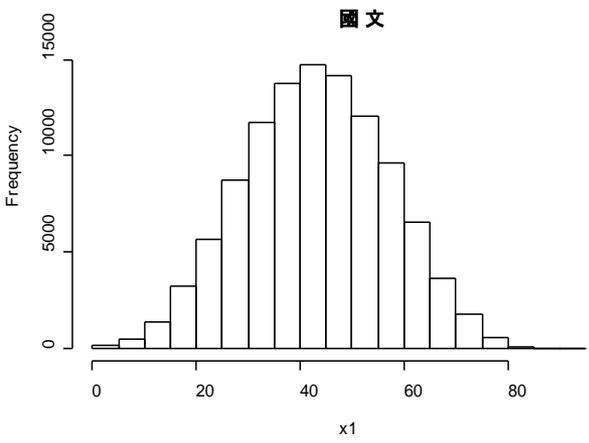


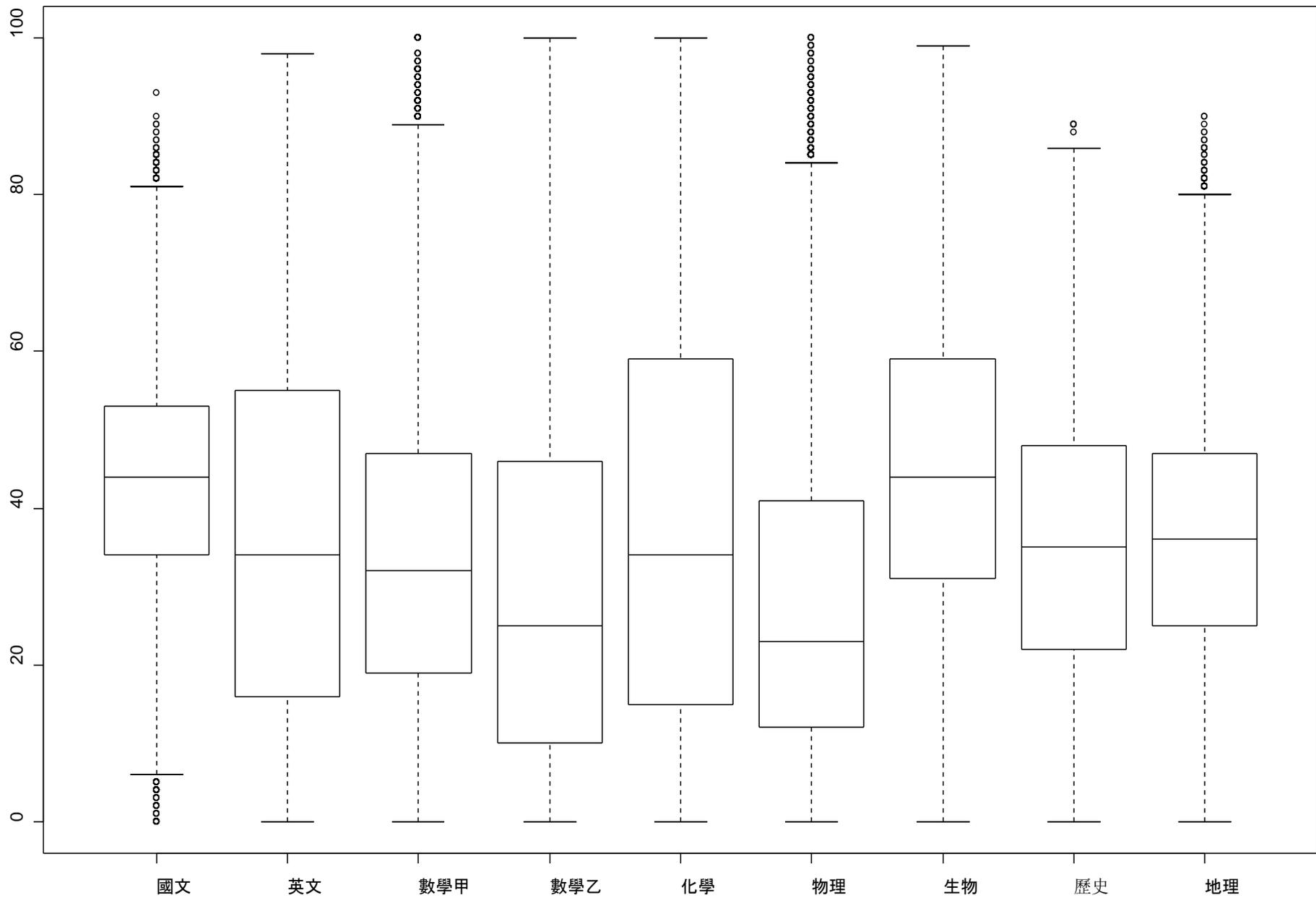
如何藉由統計獲取資訊？

- 如果想瞭解民國94年指定考試各科的特性，可以藉助哪些工具？
 - 例如：那一科的分數最不平均，像是哪一科大多數人都考得不好，只有少數人分數分高。
 - 平均數明顯大於中位數，稱為右偏(skewed to the right)；反之，若平均數明顯小於中位數，稱為左偏(skewed to the left)。平均數等於中位數，則為兩側對稱。

民國 94 年大學指定考試各科成績

	國文	英文	數學甲	數學乙	化學	物理	生物	歷史	地理
Min.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.00
12%	27.00	8.00	11.00	4.00	8.00	6.00	22.00	13.0	18.00
1st Qu.	34.00	16.00	22.00	12.00	15.00	12.00	32.00	28.0	30.00
Median	44.00	34.00	34.00	29.00	34.00	23.00	45.00	39.0	39.00
Mean	43.56	36.68	36.36	34.36	38.88	28.75	46.16	38.7	39.51
3rd Qu.	53.00	56.00	49.00	56.00	60.00	41.00	60.00	50.0	49.00
88%	60.00	69.00	59.00	61.00	76.00	57.00	71.00	56.0	55.00
Max.	93.00	98.00	100.00	100.00	100.00	100.00	99.00	89.0	90.00
st.d.	13.88	23.88	18.72	25.97	27.00	21.50	19.39	16.20	14.46



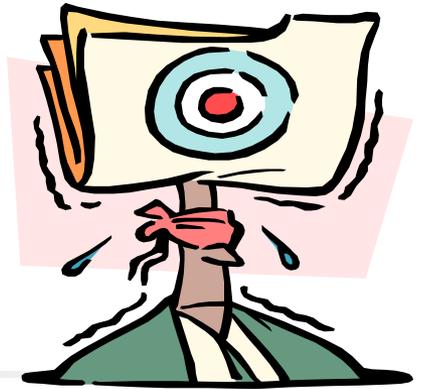


結果詮釋與推論限制

- 分析結果若能合理闡述，可達到「畫龍點睛」之效，但最忌諱忽略關鍵點，純粹就數字面來詮釋，反而變成「畫虎不成反類犬」。
- 研究結果的推論也需注意，不能僅從統計結果來看，也需瞭解推論結果代表的意義。



驟下結論(範例)

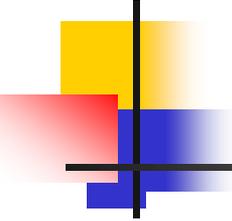


- 多數車禍發生在車速40~60公里/時，僅有少數在車速超過100公里。
→ 開快車比較安全？
- 美國亞歷桑那州死於肺結核的比例最高。
→ 亞歷桑那州的天氣易於感染肺結核？
- 調查小學生的拼字能力，發現腳愈大的拼字能力也較強。
→ 腳的大小影響拼字能力？

驟下結論(續)

- 2001年大陸調查發現長壽者中，排行老大者最多。
 - 排行老大較長壽？
 - 抑或是排行老大者佔了多數？
- 英國公務統計顯示在家裡生產者，發生意外的比例較在醫院生產者高，因此孕婦都應該在醫院生產。
 - 為什麼有些孕婦會在醫院以外的地方生產？





報紙上常看到的敘述！

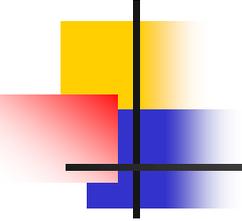
- 2004年九月底某學校公布使用手機的肇事率，發現在車內用免持聽筒打電話，比用手持聽筒還要危險。
 - 由於民眾以為使用免持聽筒就能夠一面開車、一面通話，但往往更容易分心，肇事率比用手持聽筒高出將近1倍。
 - 你/妳覺得這裡有甚麼疑問？

經過北醫2年多來的研究顯示，打手持聽筒的民眾，出車禍的風險比不打電話的人多出2.24倍，而用免持聽筒的人，出事風險又比沒打手機的人高出6.5倍，完全顛覆「用免持聽筒」比較安全的印象。北醫急診部主任謝屈平：「免持聽筒的人，你可能第一個觀念會認為說比較安全，反而疏忽掉路況，另外外也有研究過說，真正的危險性是在於對話，並不是在於免持或者是手持的這一部分。」

醫師說，比較國內外研究可以發現，行車途中只要撥打手機，危險性就會提高，而且結束通話之後5到10分鐘，出事比例仍比沒打手機的駕駛多出3倍。北醫急診部主任謝屈平：「主要在於你還在思索剛才那通電話的內容，還是會影響到你開車的表現。」

北醫附設醫院急診部主任蔡卓城說，他們除了分析各醫院急診的車禍傷患之外，也親自前往馬路旁、實地拍攝駕駛在車內使用行動電話的狀況。

(TVBS 新聞「開車打手機 免持聽筒更危險！」、
中時電子報「接、撥、掛 最易肇事時」2004年9月22日報導)

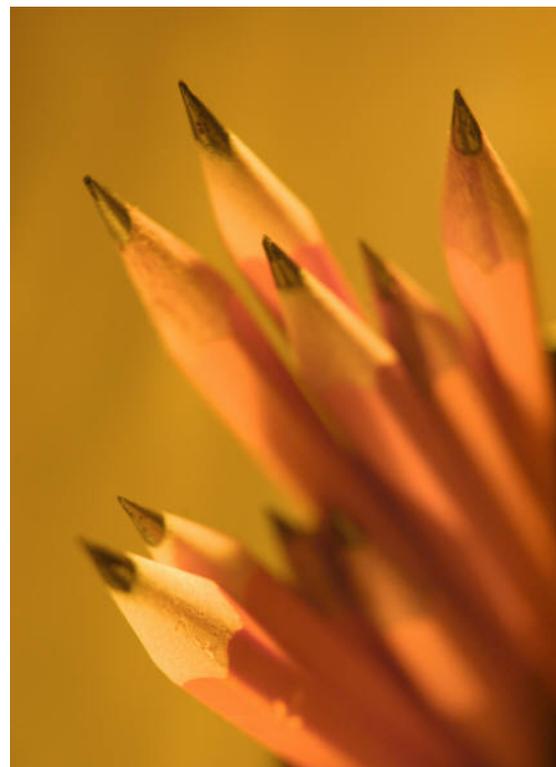


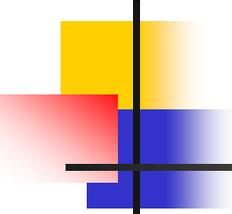
祝大家一切順利！



問卷設計的原則

- (1) 確定整份問卷的目標
- (2) 每個問題定義明確、
用字簡明易懂
- (3) 問卷內容合乎邏輯
- (4) 避免誘導、假設、敏感
的內容
- (5) 與事後編碼及分析配合





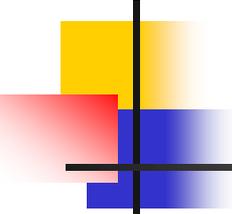
問卷設計的原則(一)

(1) 確定整份問卷的目標

- 每個問題是否與整體目標相關

→ 避免東問一題、西問一題。不妨以到一個陌生的地方問路為例，你/妳的目標是找到某個地標或建築物。

- 設計問題也可以拼圖為例，每一塊小拼圖的個別角色、幾塊拼圖合成之後的角色又如何。



問卷設計的原則(二)

(2) 每個問題定義明確、用字簡明易懂

- 使用的文字、敘述方式以出現在日常生活為原則，不使用過於艱澀、但也避免過於口語化的用詞，也避免一詞多義。

例如：請問你/妳常接觸的傳播媒體？

→或可改為「請問你/妳最常接受新聞的來源」。

問卷設計的用字



- 詞意統一(Uniformity of meaning)
- 詞意明確(Preciseness of meaning)
- 避免偏見及誤導(Freedom from undue influence of prejudice or bias)
- 避免非理性及情緒化反應(Freedom from tendency to arouse irrational or extremely emotional response)
- 避免雙重否定的用詞

問卷設計與解決問題

- 如能以問卷調查蒐集有代表性的資訊，就效率、成本大多較為划算。
- 關鍵：抽樣設計及問卷內容！
- 範例：文山指南郵局（政大）調查該局的服務滿意度。
- 動機：為什麼想要了解滿意度？
- 目的：年度評鑑、提升業績？
- 註：詢問「是否滿意」通常沒有實質助益。

問卷範例

- 請問您對本郵局的滿意程度為何？

1.滿意 2.普通 3.不滿意

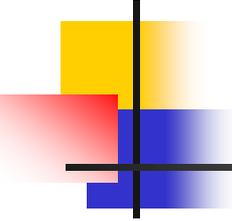
→ 銀行郵局項目不只一項，例如可分為郵務、儲蓄、劃撥等項目，以便受訪者就各項目填寫意見。

- 請問您是否為素食者？

1.是 2.否

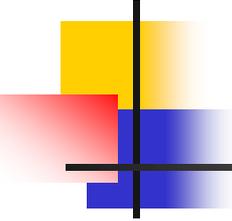
→ 素食的定義不明！





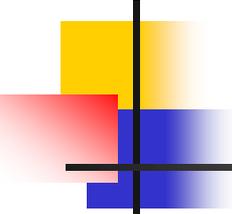
問卷範例(續)

- 您家中共有幾人工作？
 - 未註明調查所指的時間，也未說明是否也包含受訪者。
 - 「工作」指的是全職、兼職？
- 可能的修改方式：
 - 今年9月30日您家中共有幾人（包括您自己）擁有經常性（每週平均20小時以上）的工作？



問卷範例(續)

- 您經常做什麼運動？
→ 每一個人對運動的定義不同。
- 您最近是否頭痛或生病？
- 您是否會想藉由網路來獲得醫藥服務？
例如：保健資訊，線上購藥，藥物諮詢。
→ 項目太多，不知指的是哪一項。



關於敏感等相關議題

- 假設性議題

- 例：如果中了樂透、如果罹患絕症

- 敏感性（或攻擊性、私密性）議題

- 例：收入、性取向、職場暴力

- 自覺狀況或主觀判斷

- 自覺健康狀況（體檢前後差異大）

- 理想 vs. 實際（飲食、生活習慣）

1.(A)請問您喝酒的經驗是?(可複選) 紹興酒/陳紹 其他酒

(B)最近30天內您共喝了_____瓶酒,其中紹興酒/陳紹佔_____%。

(C)如果您喝紹興/陳紹,今年喝的量與去年相比,預計會

增加 減少 一樣 不喝

2.您曾經買或喝紹興/陳紹,下列因素的重要性為何?(從未喝過或未買者可免填)

	極強	強	中	弱	無
別人請客	<input type="checkbox"/>				
請客人	<input type="checkbox"/>				
送禮	<input type="checkbox"/>				
口感好	<input type="checkbox"/>				
習慣	<input type="checkbox"/>				
價格適中	<input type="checkbox"/>				
購買方便	<input type="checkbox"/>				
解悶	<input type="checkbox"/>				
強身	<input type="checkbox"/>				
流行	<input type="checkbox"/>				

3.如果您減少喝或不喜歡紹興/陳紹,下列因素的重要性為何?(多喝者可免填)

	極強	強	中	弱	無
別人請客時,減少提供紹興/陳紹	<input type="checkbox"/>				
自己請客,減少提供紹興/陳紹	<input type="checkbox"/>				
收到紹興/陳紹禮品的機會減少	<input type="checkbox"/>				
酒的選擇增多	<input type="checkbox"/>				
知名度不足(媒體少報導)	<input type="checkbox"/>				
身體狀況不允許	<input type="checkbox"/>				
飲用後,會頭痛,打嗝難聞	<input type="checkbox"/>				
品質不穩定	<input type="checkbox"/>				
口感不好,氣味不好	<input type="checkbox"/>				

4.下列的酒會取代您喝紹興/陳紹的意願為何?(每一項單獨與紹興/陳紹相比)

	極強	強	中	弱	無
紅葡萄酒	<input type="checkbox"/>				
白葡萄酒	<input type="checkbox"/>				
啤酒	<input type="checkbox"/>				
白蘭地(含XO)/威士忌	<input type="checkbox"/>				
高粱酒/白酒	<input type="checkbox"/>				
水果酒/清酒	<input type="checkbox"/>				

5.要讓紹興類酒銷路更好,您認為下列方法有效程度為何?

	極強	強	中	弱	無
創造流行,塑造形象	<input type="checkbox"/>				
大力廣告	<input type="checkbox"/>				
贈品	<input type="checkbox"/>				
折價	<input type="checkbox"/>				
增加銷售點,方便購買	<input type="checkbox"/>				
改變口感	<input type="checkbox"/>				
改變包裝	<input type="checkbox"/>				
其他_____	<input type="checkbox"/>				

6.(A)請問您喝紹興酒或陳紹，那一種為主？紹興酒為主 陳紹為主 紹興/陳紹一樣 不喝

(B)您對下列酒的整體印象是：

		紹興/陳紹				紅葡萄酒				白葡萄酒				啤酒				白蘭地/威士忌				高粱酒/白酒			
		非常贊成	沒意見	非常反對																					
健康	康	<input type="checkbox"/>																							
浪漫	漫	<input type="checkbox"/>																							
休閒	閒	<input type="checkbox"/>																							
高雅	貴	<input type="checkbox"/>																							
實惠	惠	<input type="checkbox"/>																							
活力	力	<input type="checkbox"/>																							
青春	春	<input type="checkbox"/>																							
豪爽	爽	<input type="checkbox"/>																							
新潮	潮	<input type="checkbox"/>																							
傳統/古板		<input type="checkbox"/>																							
帥氣	氣	<input type="checkbox"/>																							
典雅	雅	<input type="checkbox"/>																							
和諧	諧	<input type="checkbox"/>																							
補身體	體	<input type="checkbox"/>																							
料理酒	酒	<input type="checkbox"/>																							

7.(A)您想購買紹興/陳紹，下列那些地方曾讓您買不到？(可複選)

- 沒買過 餐廳 路邊攤 便利商店 雜貨店 KTV 酒廊
PUB 超市 量販店 洋酒專賣店 去買的地方，都買到 其他_____

(B)上個月中您本人喝或用紹興酒/陳紹的場合為何？

	次數	百分比(飲用量/紹興/陳紹總用量)
婚喪大宴	_____	_____
KTV、酒廊、PUB	_____	_____
平常宴客	_____	_____
聚餐吃飯	_____	_____
自己小飲	_____	_____
※料理食物(作業者才填)	_____	_____
其他	_____	_____

8.(A)您目前的行業是：(單選)

- 軍公教 農林漁牧業 礦產及土石採取業 製造業 水電燃氣
營造、建築業 交通、運輸及倉儲業 金融保險不動產業 資訊、通信業(製造除外)
貿易 顧問、公關公司 文化傳播娛樂、出版業 餐飲業 家庭主婦
學生 自由業 服務業(商店,百貨....) 其他_____

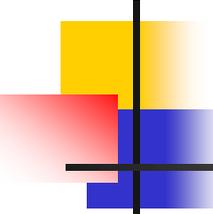
(B)您的工作型態為：(單選)

- 業務人員 行政事務人員 勞務人員 服務職 知識性工作
家庭主婦 學生 享清福

若您對本問卷有任何建議，請寫在背面空白處。

問卷資料輸入與偵錯

- 以問卷資料輸入為例：
 - 根據編碼簿(Codebook) 篩選可能問題；
 - 檢查輸入錯誤（1% ~ 5% Error）；
 - 翻閱原始問卷。
- 其他注意事項
 - 檢查離群值及其合理性；
 - 其他選項（文字、項目歸類）的整理；
 - 複選題及排序題。



樣本代表性與加權調整

- 如果無法通過樣本代表性的檢定，通常採取加權調整加以補救，常見方法有二：
 - 事後分層加權(Post-stratification)：將樣本某些重要特質，經過加權轉換，使得該特質與母體一致。(Joint Probability Distribution!)
 - 反覆多重加權(Raking)：在臺灣較為常用，一次只調整一個特質（變數），通過檢定後再考慮另一變數。

參考資料：「樣本代表性檢定與最小差異加權：以2001年台灣選舉與民主化調查為例」選舉研究(2003)

表1 訪問成功樣本之代表性檢定：性別（加權前）

	樣 本		母群	檢 定 結 果
	人 數	百分比	百分比	
男	1012	50.0495%	50.8202%	卡方值=0.456 $p > 0.05$ 樣本與母群一致
女	1010	49.9505%	49.1798%	
合 計	2022	100.0%	100.0%	

*「母群」依據2000年戶口普查資料。

表2 訪問成功樣本之代表性檢定：年齡（加權前）

	樣 本		母群	檢 定 結 果
	人 數	百分比	百分比	
20—29歲	444	21.9585%	24.6171%	卡方值=16.643 $p < 0.05$ 樣本與母群不一致
30—39歲	470	23.2443%	24.5794%	
40—49歲	450	22.2552%	21.9066%	
50—59歲	273	13.5015%	12.0141%	
60歲以上	385	19.0406%	16.8827%	
合 計	2022	100.0001%	99.9999%	

* 「母群」依據2000年戶口普查資料。