

統計實務

Spring 2024

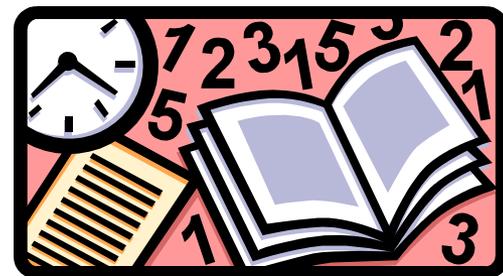
授課教師：統計系余清祥

日期：2024年4月17日

第九週：問卷資料的分析



問卷設計的統計分析

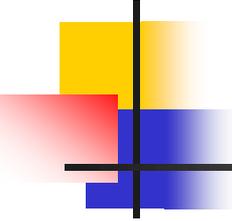


- 敘述性統計量(Descriptive Statistics)
- 相關性分析
- 卡方檢定(Chi-Square Test)
- 因素分析(Factor Analysis)
- 其他方法(例如：迴歸、時間數列、存活分析、類別資料分析)

敘述性統計量



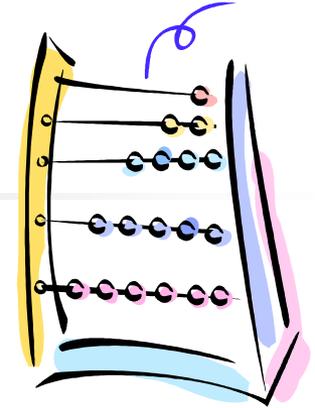
- 基本資料分析(Explanatory Data Analysis 或Initial Data Analysis)是資料分析中最基本、也是非常重要的一個步驟，資料分析的成敗往往在這個步驟中決定。
 - 敘述性統計量包括資料的基本特性，例如：平均數、標準差、所佔比例(圖表)等。
 - 一般的分類方式為：
集中趨勢量數、差異量數



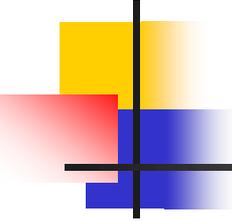
基本資料分析

- 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？

集中趨勢量數



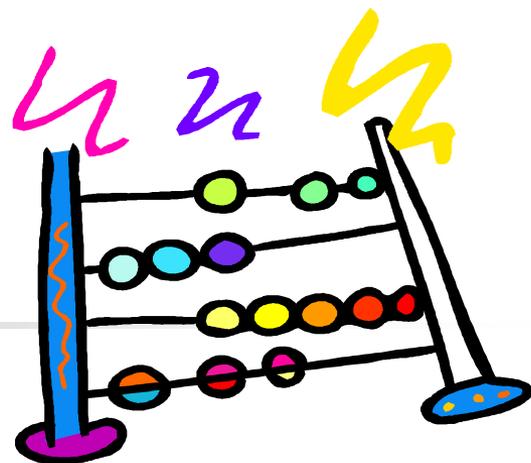
- 平均數(Mean;期望值)
 - 算術平均數(Average) ;
 - 加權平均數(Weighted Average) ;
 - 其他(幾何平均數、調和平均數)。
- 中位數(Median): 一半的數值比中位數大，一半的數值比中位數小。
- 眾數(Mode): 出現次數最多的數值



集中趨勢量數(範例)

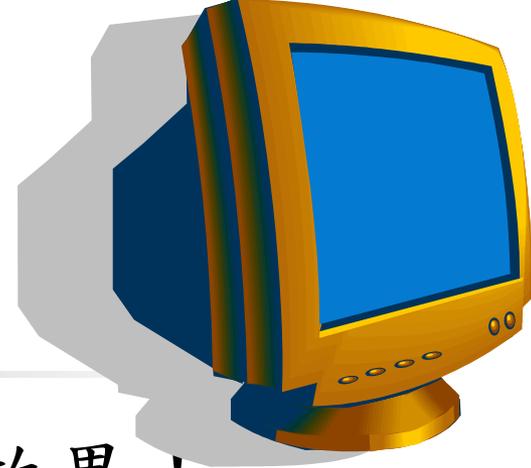
- 員工薪資為25, 30, 30, 30, 35, 43, 70, 80, 85
→ 中位數是35
- 25, 30, 30, 30, 35, 43, 70, 80, 85, 90
→ 中位數是 $\frac{35+43}{2} = 39$
- 員工薪資為25, 30, 30, 30, 35, 43, 70, 80, 85
→ 眾數是30
- 25, 30, 30, 35, 38, 43, 43, 80, 85
→ 眾數是30及43(眾數不唯一)

差異量數



- 全距(Range):
 - 最大與最小數值之差(Range = Max - Min)
- 四分位差(Quartile Deviation):
 - 四分位數(Quartile; Q_1): 3/4的數值比大 Q_1 , 1/4的數值比 Q_1 小。
 - 四分位差 = $Q_3 - Q_1$
- 變異數(Variance; σ^2)與標準差(Standard Deviation; σ)

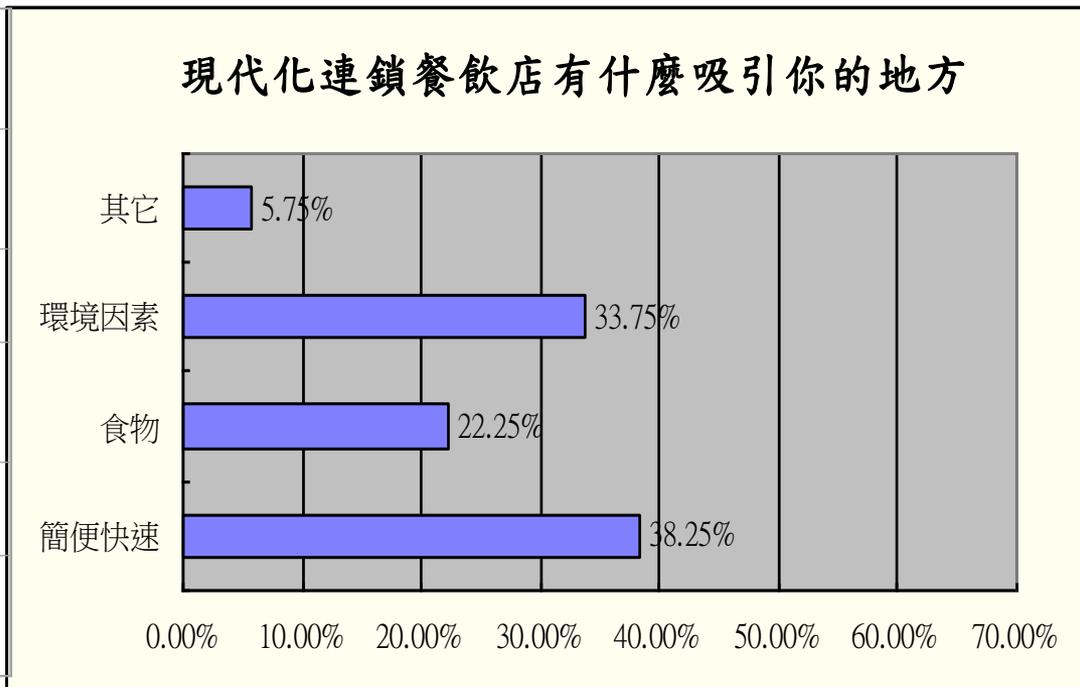
問卷分析範例



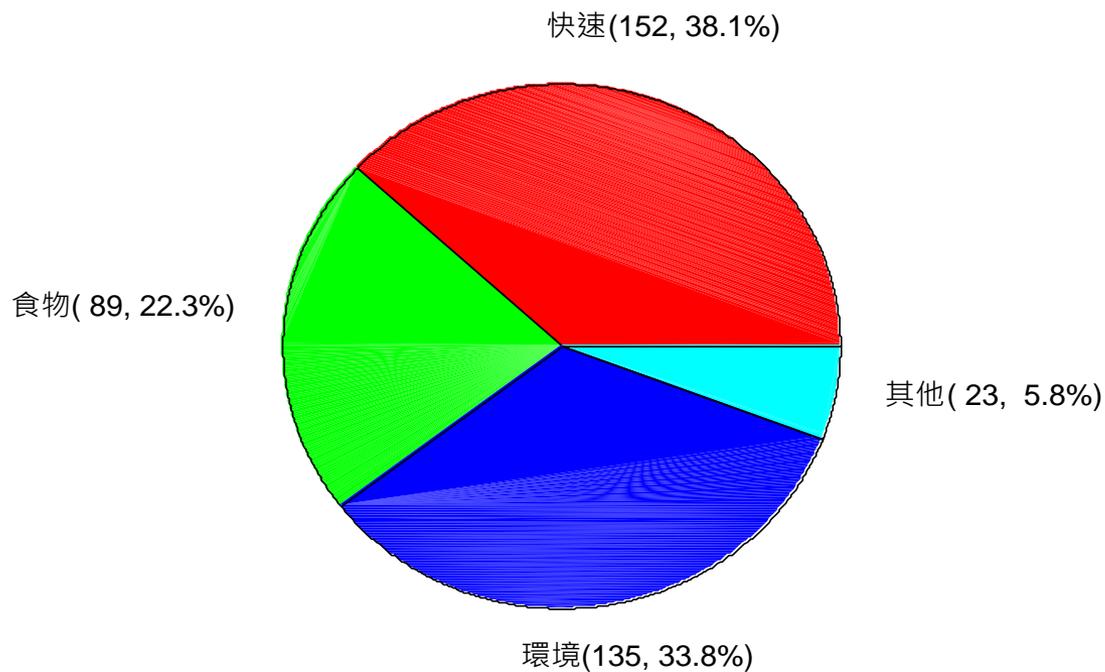
- 以表格、圖形展示資料更有效果！

長條圖(Bar Chart)

	選項	人數	百分比
1	簡便快速	153	38.25
2	食物	89	22.25
3	環境因素	135	33.75
4	其它	23	5.75
	N =	400	

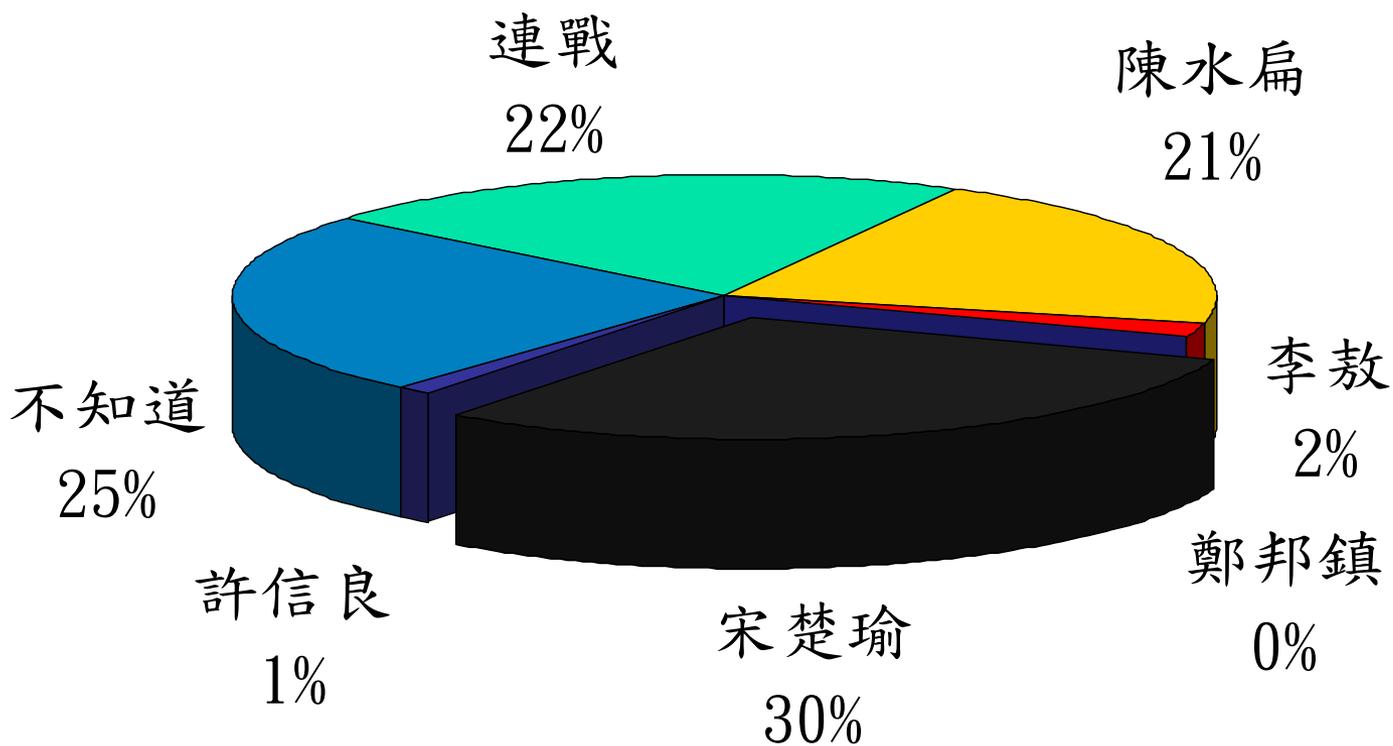


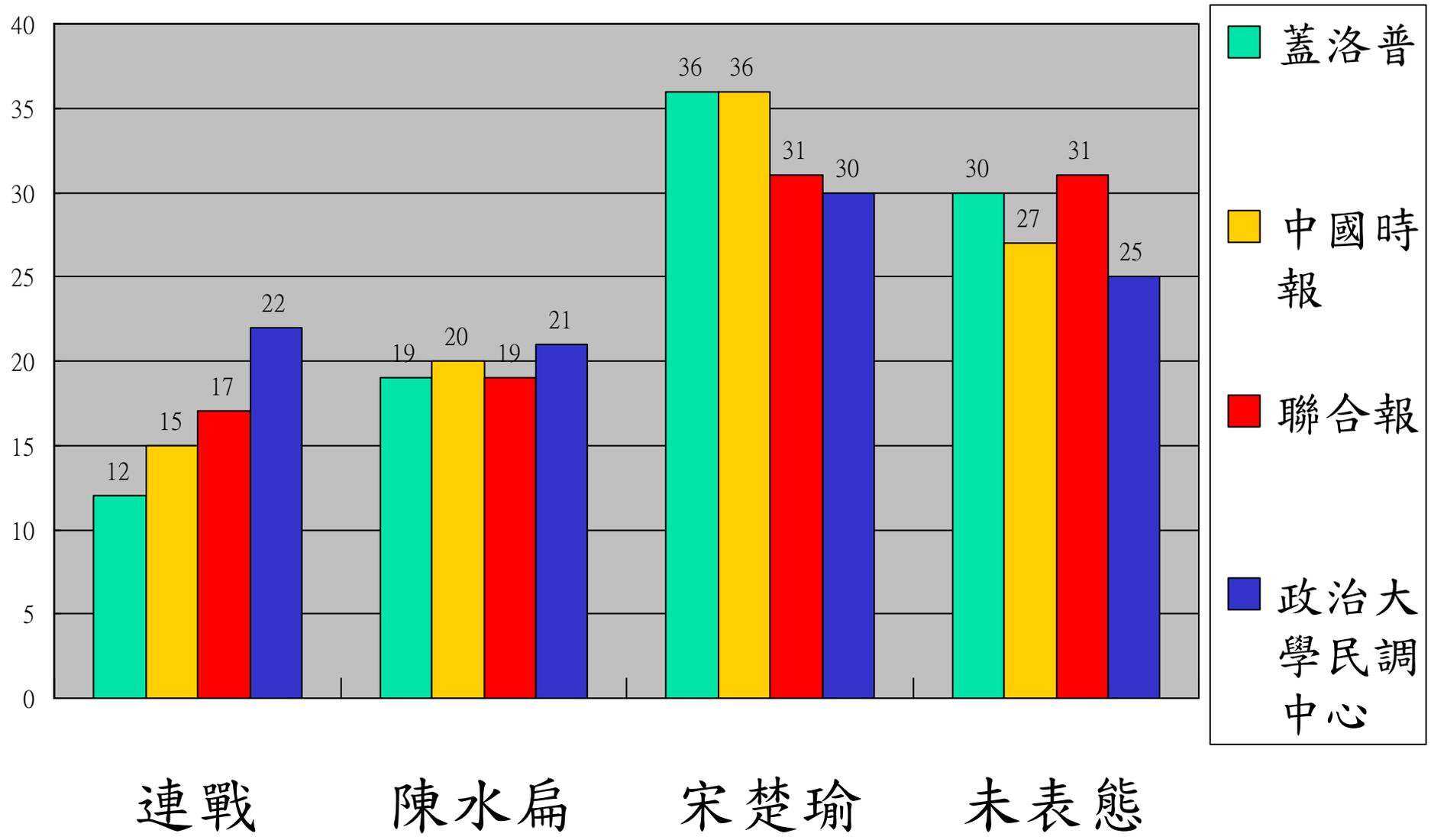
- 圓餅圖(Pie Chart)也是另一種圖形表示法。



圓餅圖的範例(立體)

總統候選人的支持比例(88年11月)





連戰

陳水扁

宋楚瑜

未表態

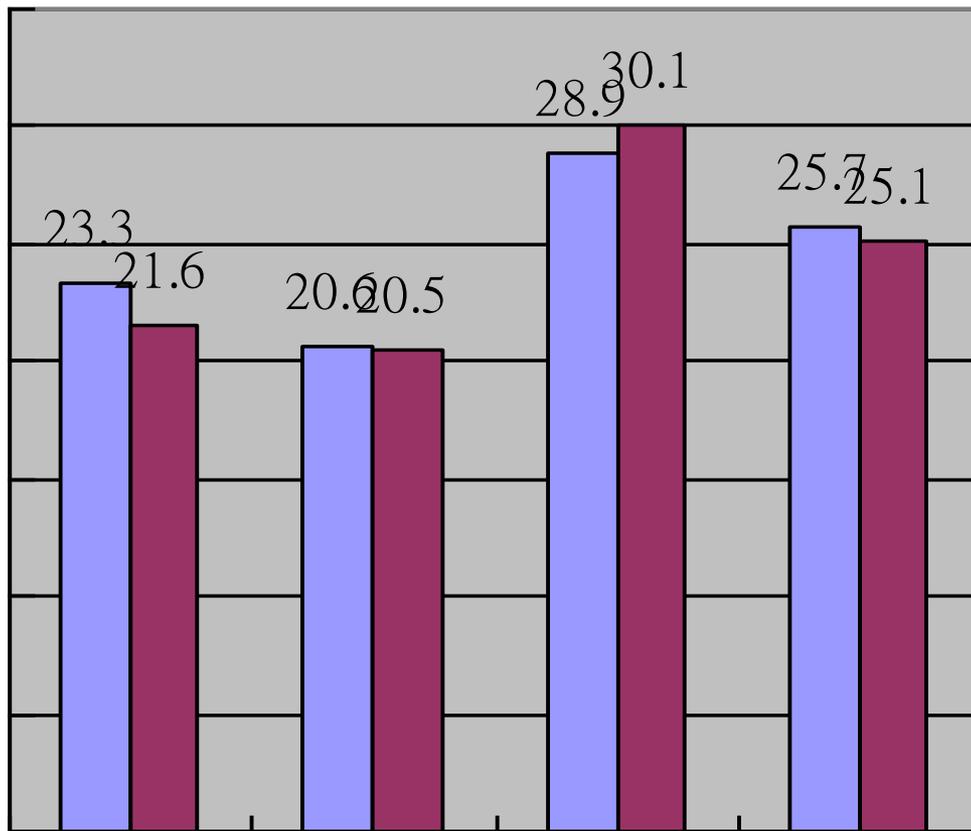
蓋洛普

中國時報

聯合報

政治大學民調中心

35
30
25
20
15
10
5
0



連戰

陳水扁

宋楚瑜

未回答

921地震災後
重建(105~106)

宋楚瑜宣布副
總統人選
(11/11~11/12)



香港基督教女青年會



香港大學行為健康教研中心

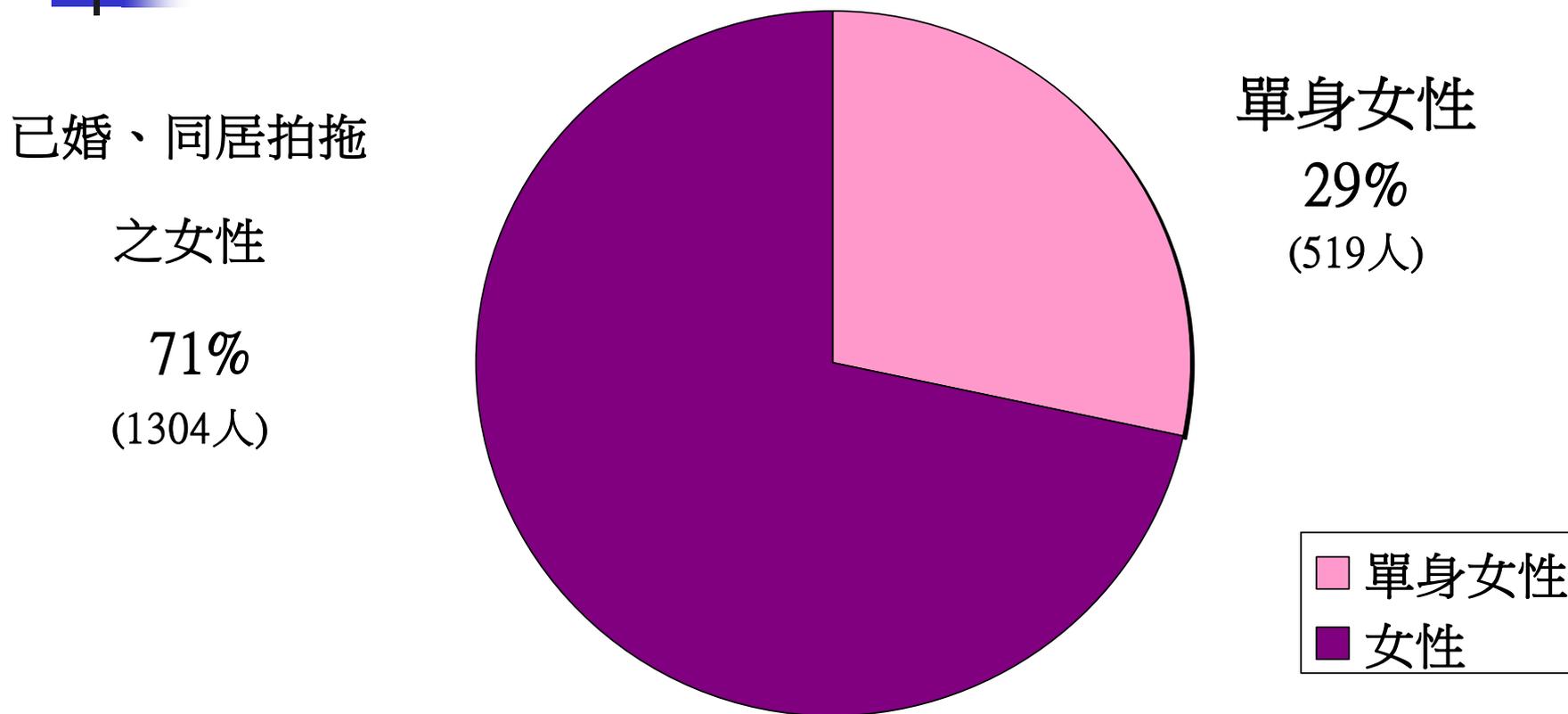
合辦

單身女性狀況及需要調查研究

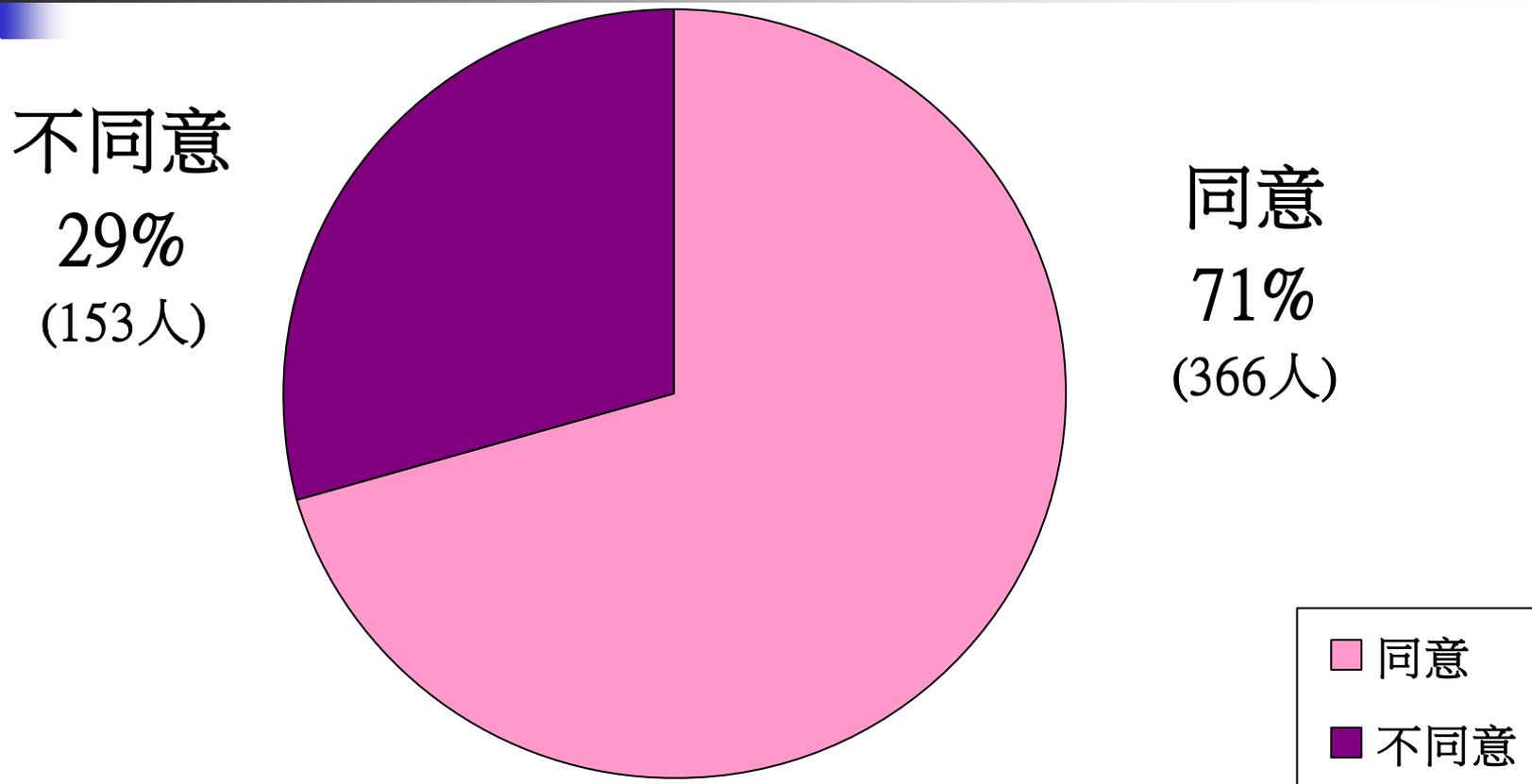
新聞發佈會

二零零四年二月八日

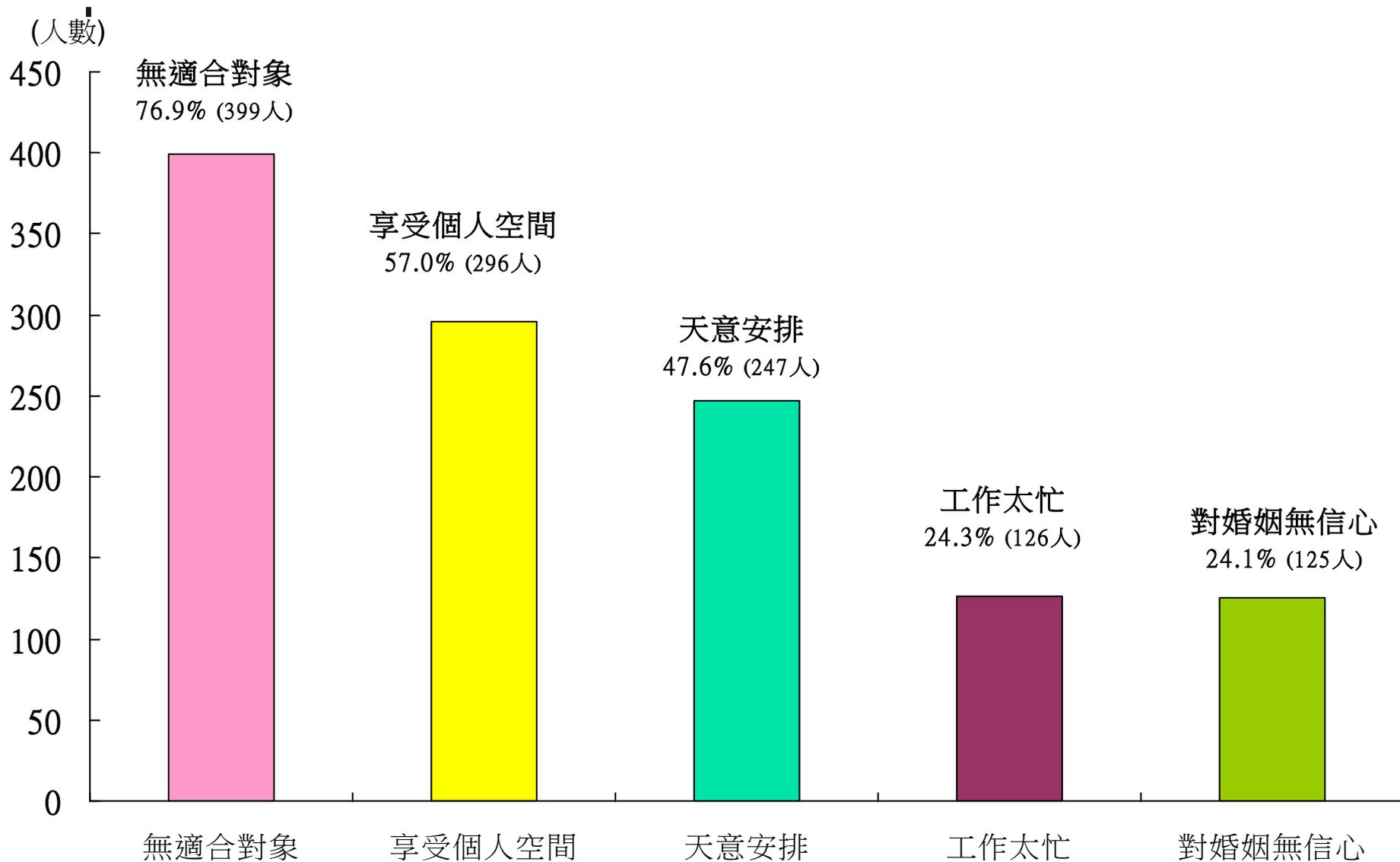
圖二：被訪單身女性人數



圖四：被訪單身女性同意單身是她們的選擇



圖六：被訪單身女性認為個人單身的原因



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？



驟下結論？

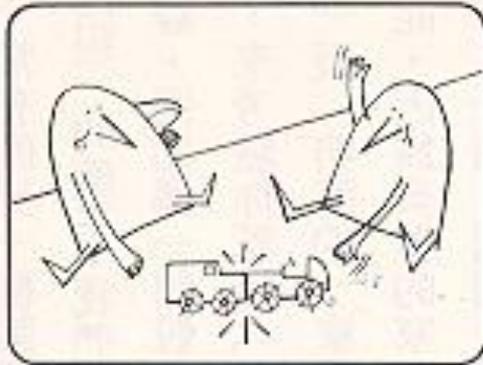


圖 1：根據統計顯示，多數車禍發生在車子行駛於一般車速的時候；只有少數車禍發生在車超速過每小時一百五十公里的時候。這是否表示開快車比較安全呢？

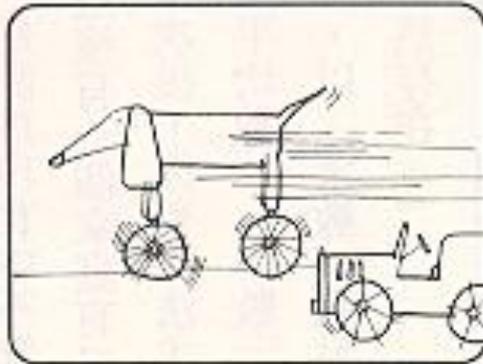


圖 2：絕對不是。統計上的關係通常和因果是無關的，多數人都以一般車速開車，自然多數車禍發生於一般車速。

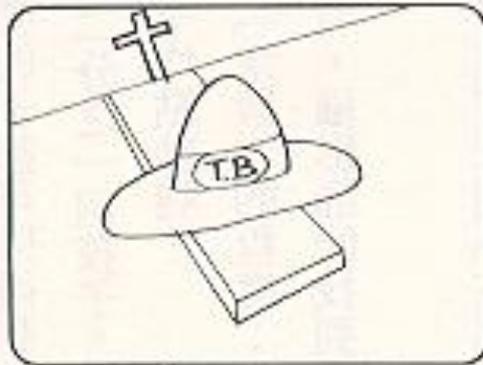


圖 3：如果統計顯示在亞利桑那州有較多的人死於肺結核，這是否表示和別州比較起來，亞利桑那州的天氣較容易感染肺結核呢？

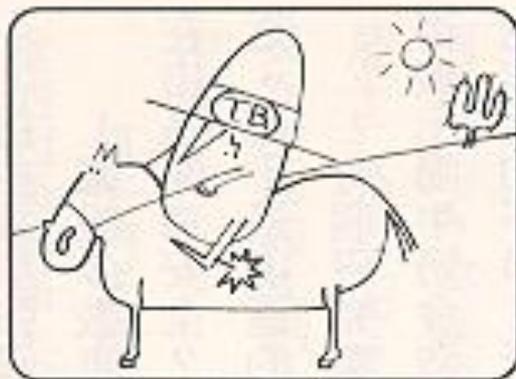


圖 4：剛好相反，正因為亞利桑那的天氣有助於肺炎患者，所以有上千患者去那兒休養，自然就提高了死於肺炎的平均人數。

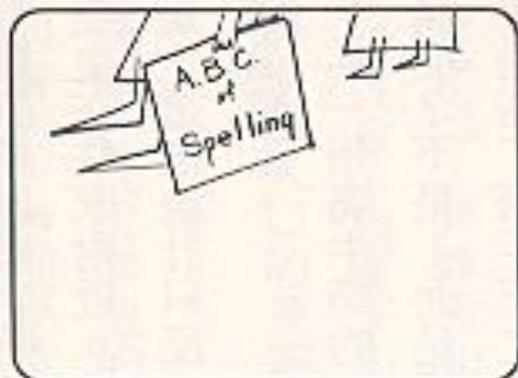


圖 5：有個調查研究顯示，腳比較大的兒童拼字能力也比較好。這是不是表示從一個人腳的大小，可以測量出他的拼字能力？

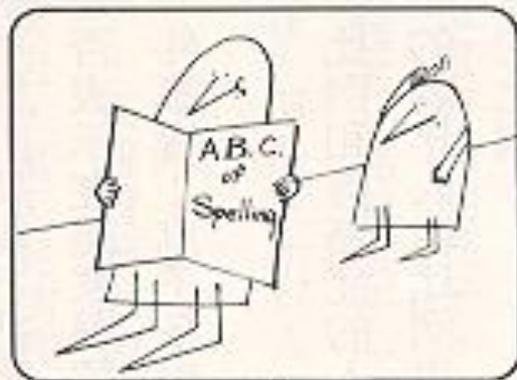
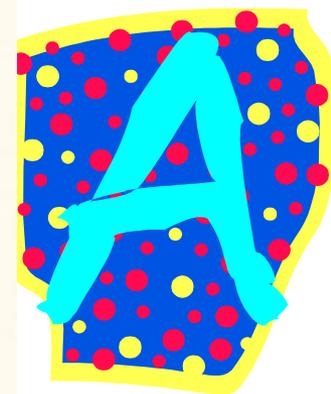
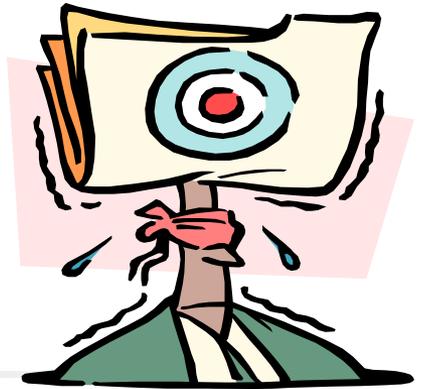


圖 6：當然不是。這個研究包括成長中的孩童，所以統計數字只是說明年齡較大的兒童（他們的腳當然也比較大）比年齡較小的兒童，拼字拼得比較好。



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。」

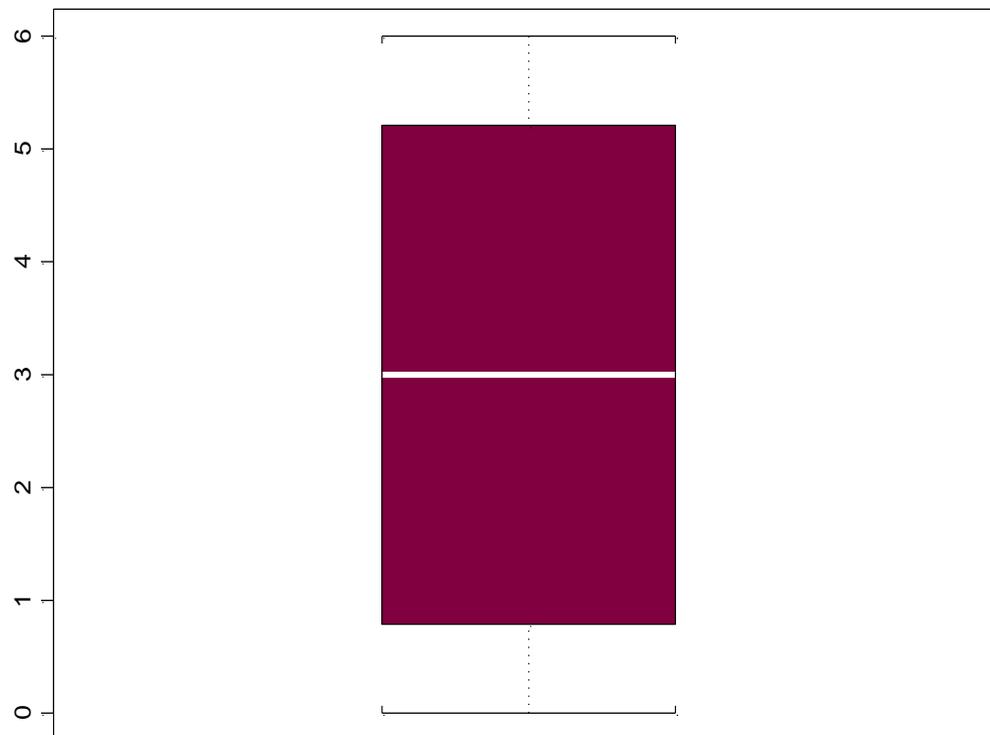
醫師說，比較國內外研究可以發現，行車途中只要撥打手機，危險性就會提高，而且結束通話之後5到10分鐘，出事比例仍比沒打手機的駕駛多出3倍。北醫急診部主任謝屈平：「主要在於你還在思索剛才那通電話的內容，還是會影響到你開車的表現。」

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。

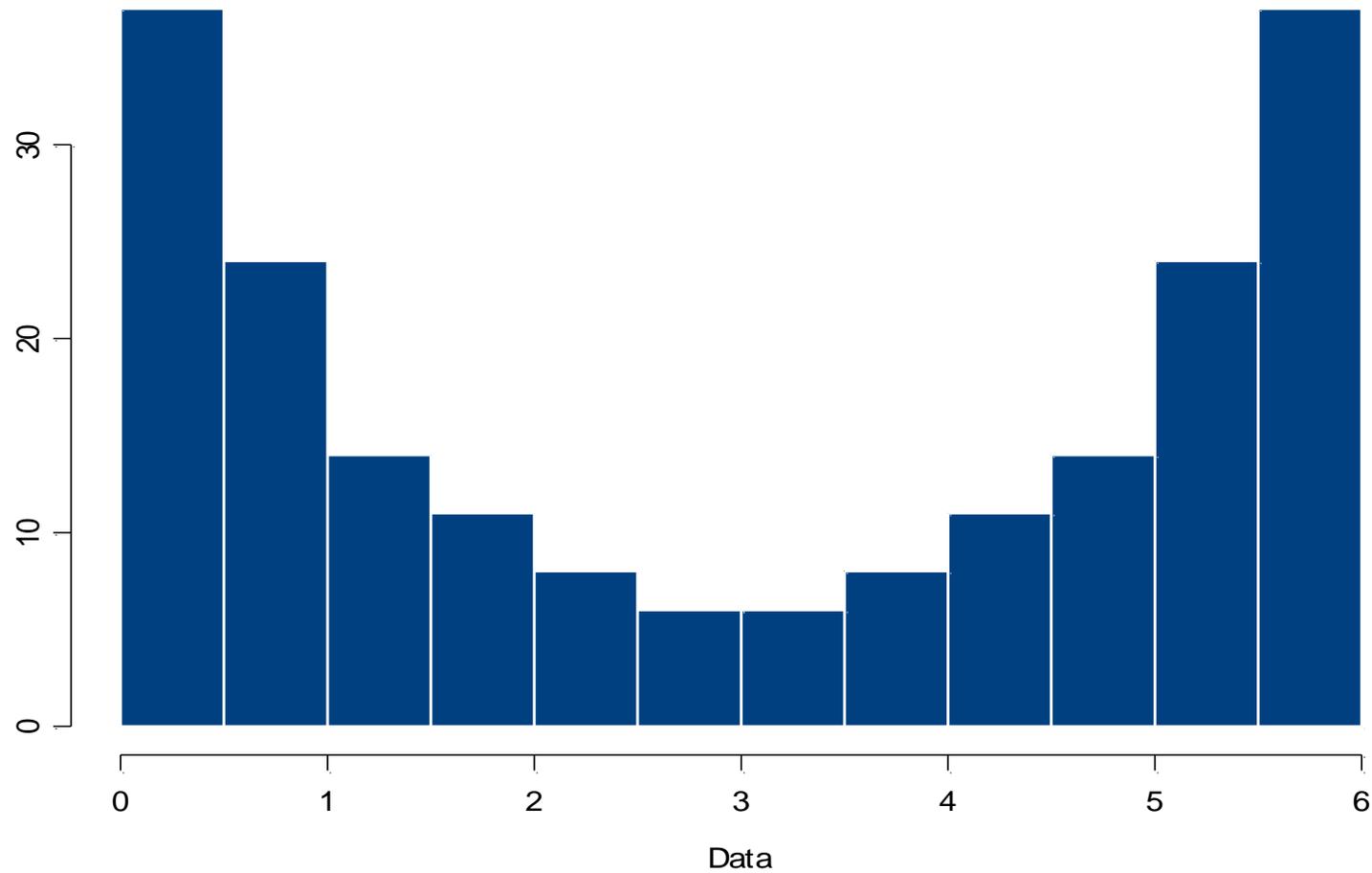
(TVBS 新聞「開車打手機 免持聽筒更危險！」、
中時電子報「接、撥、掛 最易肇事時」2004年9月22日報導)

表格與圖形(範例)

例題五、大略敘述下圖資料的特性。



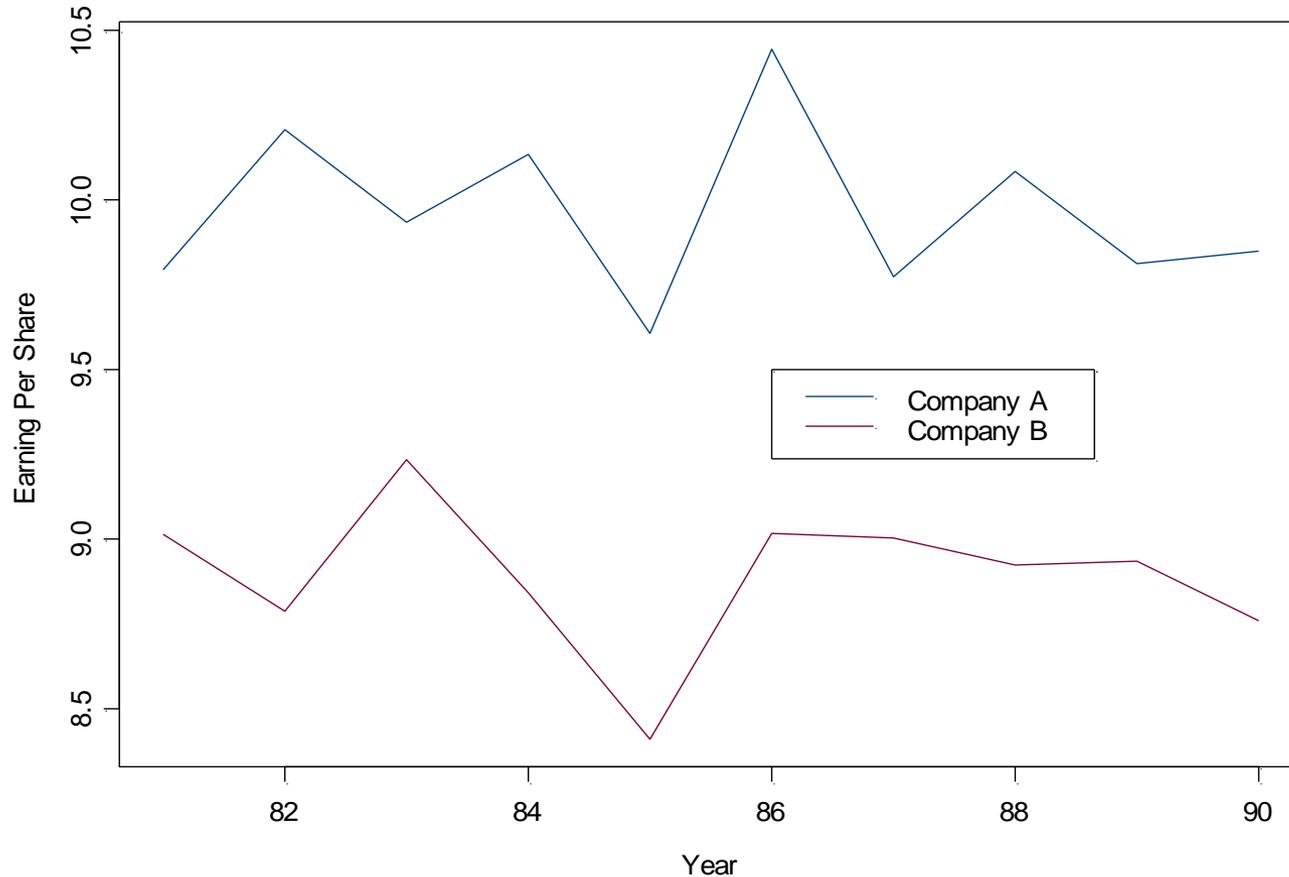
原始資料(你/妳猜對了嗎?)



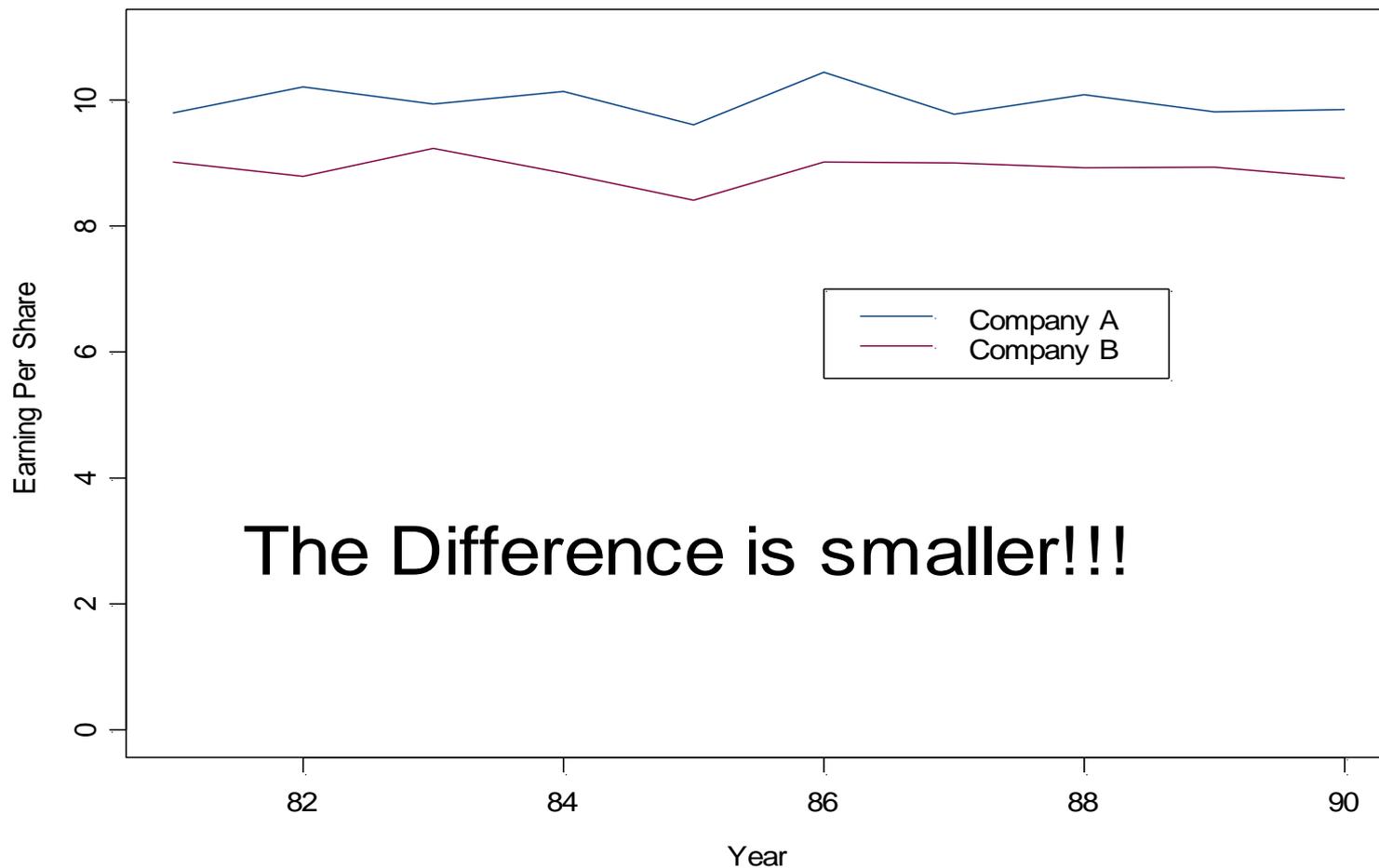
問題：常態分配的箱型圖具有哪些特性？

表格與圖形(續)

例題六、A、B兩公司的差距大嗎？

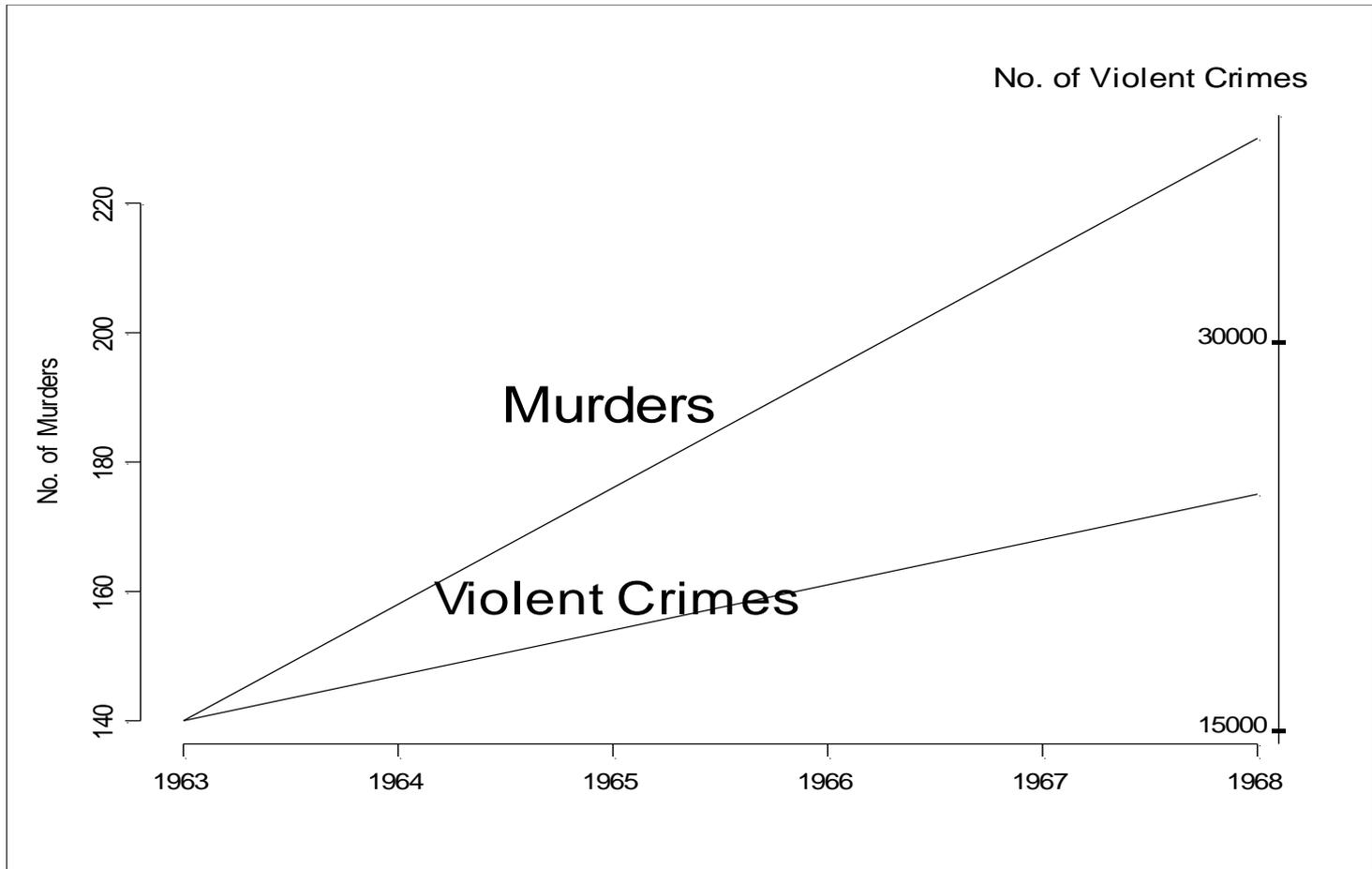


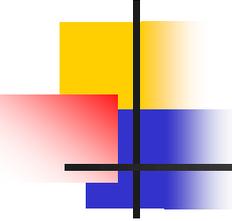
更動 Y 軸標示單位的效果！



表格與圖形(續)

例題七、謀殺率上升嚴重嗎？



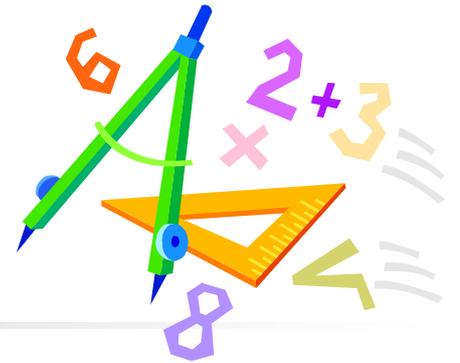


相關性分析

- 相關性分析的主旨在於找出數字大小的差異或關係，例如：
 - 比較平均數(Z或t 檢定)、變異數
 - 線性相關係數(Correlation Coefficient)
- 但相關性分析在一般的問卷分析中無法進行，因為資料為選項式，不見得具有大小關係。

	完全 不同意	不同意	無所謂同 不同意	同意	完全 同意
(R01)我喜歡在有挑戰的路段上開車	1	2	3	4	5
(R02)我要與眾不同的車子	1	2	3	4	5
(R03)加速感和操控感對我來說很重要	1	2	3	4	5
(R04)我喜歡買台灣製的車子	1	2	3	4	5
(R05)在買車前我會閱讀一些雜誌和指南	1	2	3	4	5
(R06)進口車的品質比國產車好	1	2	3	4	5
(R07)對我而言，車子只是一種交通工具	1	2	3	4	5
(R08)從一個人開車的風格，可以了解一個人	1	2	3	4	5
(R09)如果能得到我所要的性能，我可以忍受一些問題	1	2	3	4	5
(R10)我通常大概每二年左右就會買一輛新車	1	2	3	4	5
(R11)我喜歡讓別人來開車	1	2	3	4	5
(R12)銷售量高的車子一定是好車	1	2	3	4	5
(R13)我的朋友認為我很懂車	1	2	3	4	5
(R14)我喜歡為自己的車子選購配件	1	2	3	4	5
(R15)我絕對不會買第一年剛上市的车子	1	2	3	4	5
(R16)我偏好大且重的車	1	2	3	4	5
(R17)在買車時，我會買同級車中最好的車	1	2	3	4	5
(R18)對我而言，安全遠比造型重要	1	2	3	4	5
(R19)我偏好有名氣的车子	1	2	3	4	5

相關性分析(續)



- 有些問卷會要求受訪者填寫0到9(或1到5)的數字，代表喜好(或贊成)的程度。
- 因此資料具有連續的大小關係，相關分析、甚至多變量分析都可使用。
- 這些分析假設填寫的數字具有大小關係，也就是說填寫9分者是填寫1分者的9倍。(實際上，很難驗證這是否合理！)

相關性分析(續)

■ 例如：以下為某項調查的兩個問項

1. 請問您平均一週到現代連鎖餐飲店用餐的次數：

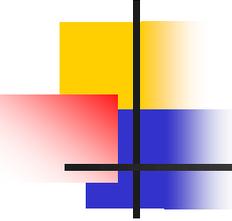
- _1. 0次
- _2. 1~3次
- _3. 3~6次
- _4. 6~10次
- _5. 10次以上

2. 請問您覺得現代連鎖餐飲店的價格如何？

- _1. 高很多
- _2. 高一些
- _3. 差不多
- _4. 低一些
- _5. 低很多

→ 計算出兩者的相關係數為0.08，
兩者間似乎不相關。





相關性分析(續)

- 建議：以圖形或表格找出較明確的資訊。

第 2 題：價格

第
1
題
：
次
數

	1	2	3	4	加總
1	34	47	4	2	87
2	67	199	14	0	280
3	3	22	2	0	27
4	1	3	0	0	4
5	1	1	0	0	2
加總	106	272	20	2	400

相關性分析(續)



■ 數字定義可能引起的問題：

→ 評審給候選人A、B、C、D的評分。

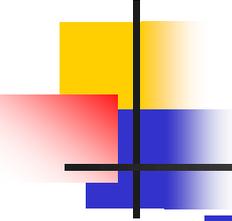
評審	1	2	3	4	5	6	7	總分
A	4	1	2	4	1	2	4	18
B	3	4	1	3	4	1	3	19
C	2	3	4	2	3	4	2	20
D	1	2	3	1	2	3	1	13

相關性分析(續)



- 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	2	1	2	2	1	2	14
C	1	3	3	1	3	3	1	13



卡方檢定

- 卡方檢定(Chi-Square Test) 在問卷分析中多為交叉分析(獨立性檢定)，測量各問項間是否相關。

→ 例如：隨機抽樣調查200位大學生，獲得其就讀年級與吸煙習慣的資料如下：

年級 習慣	一年級	二年級	三年級	四年級
吸煙	21	33	25	20
不吸煙	47	26	19	9

請問您對中央政府處理921大地震的救災工作
 這段期間的整體表現滿不滿意？vs. 請問您的
 戶籍是否在災區？

	很 滿意	有點 滿意	不太 滿意	很 不滿意	不知道 /拒答	總計
不是在災區	127 14.60	405 46.55	136 15.63	99 11.38	103 11.84	870
是在災區	64 10.88	274 46.60	102 17.35	85 14.46	63 10.71	588
總計	191 13.10	679 46.57	238 16.32	184 12.62	166 11.39	1458 100.0

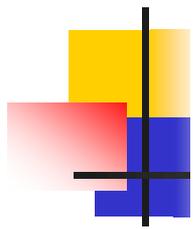
X 軸題目 : 1. 請問明年的總統大選，您比較支持誰？

Y 軸題目 : 2. 請問明年的總統大選，您認為哪一位候選人最有可能當選？

次數	連戰	陳水扁	宋楚瑜	許信良	李敖	鄭邦鎮	不知道/ 拒答	總計
百分比								
列百分比								
行百分比								
連戰	166	91	120	6	7	0	65	455
	15.90	8.72	11.49	0.57	0.67	0.00	6.23	43.58
	36.48	20.00	26.37	1.32	1.54	0.00	14.29	
	73.45	42.52	38.22	85.71	35.00	0.00	24.81	
陳水扁	8	47	12	0	3	1	27	98
	0.77	4.50	1.15	0.00	0.29	0.10	2.59	9.39
	8.16	47.96	12.24	0.00	3.06	1.02	27.55	
	3.54	21.96	3.82	0.00	15.00	100.00	10.31	
宋楚瑜	26	32	138	1	2	0	27	226
	2.49	3.07	13.22	0.10	0.19	0.00	2.59	21.65
	11.50	14.16	61.06	0.44	0.88	0.00	11.95	
	11.50	14.95	43.95	14.29	10.00	0.00	10.31	
許信良	0	0	0	0	0	0	0	0
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
李敖	0	0	0	0	0	0	0	0
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
鄭邦鎮	0	0	0	0	0	0	0	0
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
不知道/ 拒答	26	44	44	0	8	0	143	265
	2.49	4.21	4.21	0.00	0.77	0.00	13.70	25.38
	9.81	16.60	16.60	0.00	3.02	0.00	53.96	
	11.50	20.56	14.01	0.00	40.00	0.00	54.58	
總計	226	214	314	7	20	1	262	1044
	21.65	20.50	30.08	0.67	1.92	0.10	25.10	100

Chi Square : 395.490

P Value : 0.000



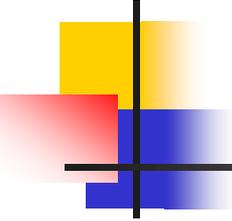
X 軸題目 : 2. 請問明年的總統大選，您認為哪一位候選人最有可能當選？

Y 軸題目 : 8. 請問您的教育程度？

次數	連戰	陳水扁	宋楚瑜	許信良	李敖	鄭邦鎮	不知道/拒答	總計
百分比								
列百分比								
行百分比								
國中及以下	103	43	75	0	0	0	181	402
	9.87	4.12	7.18	0.00	0.00	0.00	17.34	38.51
	25.62	10.70	18.66	0.00	0.00	0.00	45.02	
	22.64	43.88	33.19	0.00	0.00	0.00	68.30	
高中高職	192	34	70	0	0	0	57	353
	18.39	3.26	6.70	0.00	0.00	0.00	5.46	33.81
	54.39	9.63	19.83	0.00	0.00	0.00	16.15	
	42.20	34.69	30.97	0.00	0.00	0.00	21.51	
大專及以上	160	21	81	0	0	0	27	289
	15.33	2.01	7.76	0.00	0.00	0.00	2.59	27.68
	55.36	7.27	28.03	0.00	0.00	0.00	9.34	
	35.16	21.43	35.84	0.00	0.00	0.00	10.19	
拒答	0	0	0	0	0	0	0	0
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
總計	455	98	226	0	0	0	265	1044
	43.58	9.39	21.65	0.00	0.00	0.00	25.38	100

Chi Square : 101.760

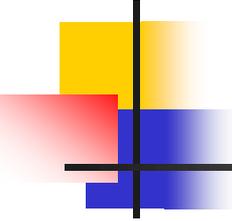
P Value : 0.000



卡方檢定(續)

- 卡方檢定要求每一格的期望個數不得小於5，因此個數太少者必須合併。

	1	2	3	4	加總
1	34	47	4	2	87
2	67	199	14	0	280
3	3	22	2	0	27
4	1	3	0	0	4
5	1	1	0	0	2
加總	106	272	20	2	400



卡方檢定(續)

- 可能的合併方式：

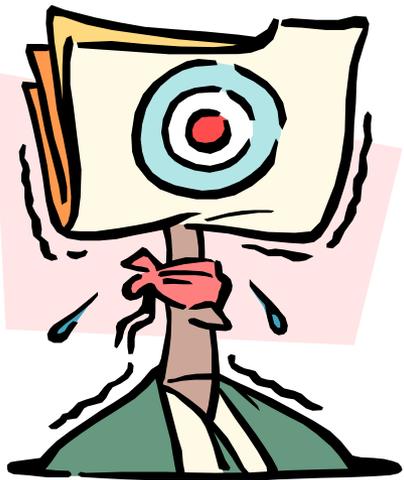
	1	2	3以上	加總
1	34	47	6	87
2	67	199	14	280
3以上	5	26	2	33
加總	106	272	22	400

→ 分析結果：行與列有顯著的關係！

價格 vs. 次數

列聯表的資訊

- 美國某州的婦運團體研究判刑的輕重是否存有性別歧視，隨機抽取男女各一百名判刑確定的嫌犯，結果如下：



	輕刑	重刑	總數
男嫌疑犯	40	60	100
女嫌疑犯	60	40	100