

卡方檢定

檢定原理

之前，我們所處理的資料大多是數量資料(quantitative data)，再透過計算其平均數或標準差，然後做估計或檢定等統計推論工作。

有時我們所遇到的資料為質的資料或類別資料(categorical data)，此時我們所擁有的訊息大多是各不同類別資料的所佔的個數或比例，所以此種資料亦稱為計數資料。

本單元即介紹一些分析類別資料的方法，通稱為「卡方檢定法」(chi-square tests)。

檢定原理

- 卡方適合度檢定通常只適用於大樣本，一般說來，當所有的 $E_{ij} \geq 5$ 時，卡方適合度檢定才適用。若有 $E_{ij} < 5$ 時，該組必須與相鄰組合併至 $E_{ij} \geq 5$ 為止，即組數亦必須減少。
(E_{ij} 指的是期望次數)

檢定原理

- 卡方檢定的基本概念是測試某類別資料出現的頻率 (frequency)，是否隨著另一個類別資料的不同層級而有不同的分布。舉例來說，購買者的性別與居住地之分布如下表：

	北部	中部	南部	總計
男	24	55	41	120
女	42	77	59	178
總計	66	132	100	298

檢定原理

- 該業者希望了解什麼呢？

顯然男女生的比例是否在北、中、南部有所不同？

- 依照兩變數間次數的分布情形，從而判斷兩變數是否有關聯，就是卡方檢定的應用。列聯表中自變數 (independent variable) 應列在上方(column)，而應變數(dependent variable)列在左方(row)。

檢定原理

- 卡方檢定的例子：

- 從下表中的次數資料，檢定男女生在購買房車時，對大小型車的選擇是否有所不同？

	大型車	中型車	小型車	總計
男	30 (實際次數) (24) (期望次數)	25 (24)	25 (32)	80
女	15 (21)	20 (21)	35 (28)	70
總計	45	45	60	150

檢定原理

- 檢定步驟如下：

- 建立假說：

- H_0 ：購車大小與性別無關

- H_1 ：購車大小與性別有關

- 依據虛無假設，計算出期望次數(E)。

- 卡方值計算如下：

$$x^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- 卡方值愈大，愈表示 H_0 之假設有誤，傾向拒絕 H_0 。

卡方檢定的關聯性檢測

- 依據質化資料的性質（名目或順序），而有不同的相關係數。
 - **Phi相關**：適用於兩個變項均為二分法之名目變項（例如：性別）。
 - **Cramer's V**：是名目變數相關性最常用的指標。其值介於0與1之間($0 \leq V \leq 1$)；數值愈接近 1 表示相關愈強。當用於2*2的列聯表時，V值與Phi值相同，而一般**Cramer's V**多用於比2*2大的列聯表。

卡方檢定的關聯性檢測

- **列聯相關係數(contingency coefficient)**：建議此數值應用在大於 5×5 以上的列聯表。
- **Eta相關**：如果兩個變數中一個是名目尺度，另一個為連續變數由不同數值區間所訂出的區間尺度（例如：依年齡分為四組），則選用Eta值。
- **McNemar相關**：用於重複試驗時，比較前後的差異性是否明顯。

卡方檢定的類型與範例

- 獨立性檢定(test of independence)

- 用以檢測同一樣本的兩個類別變項 (X 與 Y) 之間的關聯性。

- 配適度檢定(test of goodness of fit)

- 用以討論某一個變項是否與某個理論分配或母群分配相符合，或是檢定抽樣的次數分配是否為某種特殊形式。

- 同質性檢定 (test of homogeneity)

- 檢定二個以上樣本，在某一個變項的反應是否具有顯著差異。

- 改變顯著性檢定(test of significance of change)

- 用以檢定同一樣本在事件前後的反應，是否有顯著的差異。

卡方檢定--傳統作法

- 對於單選題之交叉表，通常得進行卡方獨立性檢定。以政黨傾向交叉地區別資料為例，其虛無假設（ H_0 ）與對立假設（ H_1 ）為：

H_0 ：政黨傾向與地區別無關

H_1 ：兩者有關

$\alpha=0.05$

傳統之作法為，先計算卡方值，其運算公式為：

$$\chi^2 = \sum_{allcell} \frac{(O - E)^2}{E}$$

		北區	中區	南區	東區	合計	列百分比
民進黨	觀察值	85	43	95	5	228	22.8%
	期望值	102.6	45.6	68.4	11.4		
國民黨	觀察值	101	37	51	15	204	20.4%
	期望值	91.8	40.8	61.2	10.2		
親民黨	觀察值	91	45	50	10	196	19.6%
	期望值	88.2	39.2	58.8	9.8		
台聯	觀察值	18	13	21	3	55	5.5%
	期望值	24.75	11	16.5	2.75		
新黨	觀察值	17	9	7	3	36	3.6%
	期望值	16.2	7.2	10.8	1.8		
無	觀察值	138	53	76	14	281	28.1%
	期望值	126.45	56.2	84.3	14.05		
合計		450	200	300	50	1000	100.0%
欄百分比		45.0%	20.0%	30.0%	5.0%	100.0%	

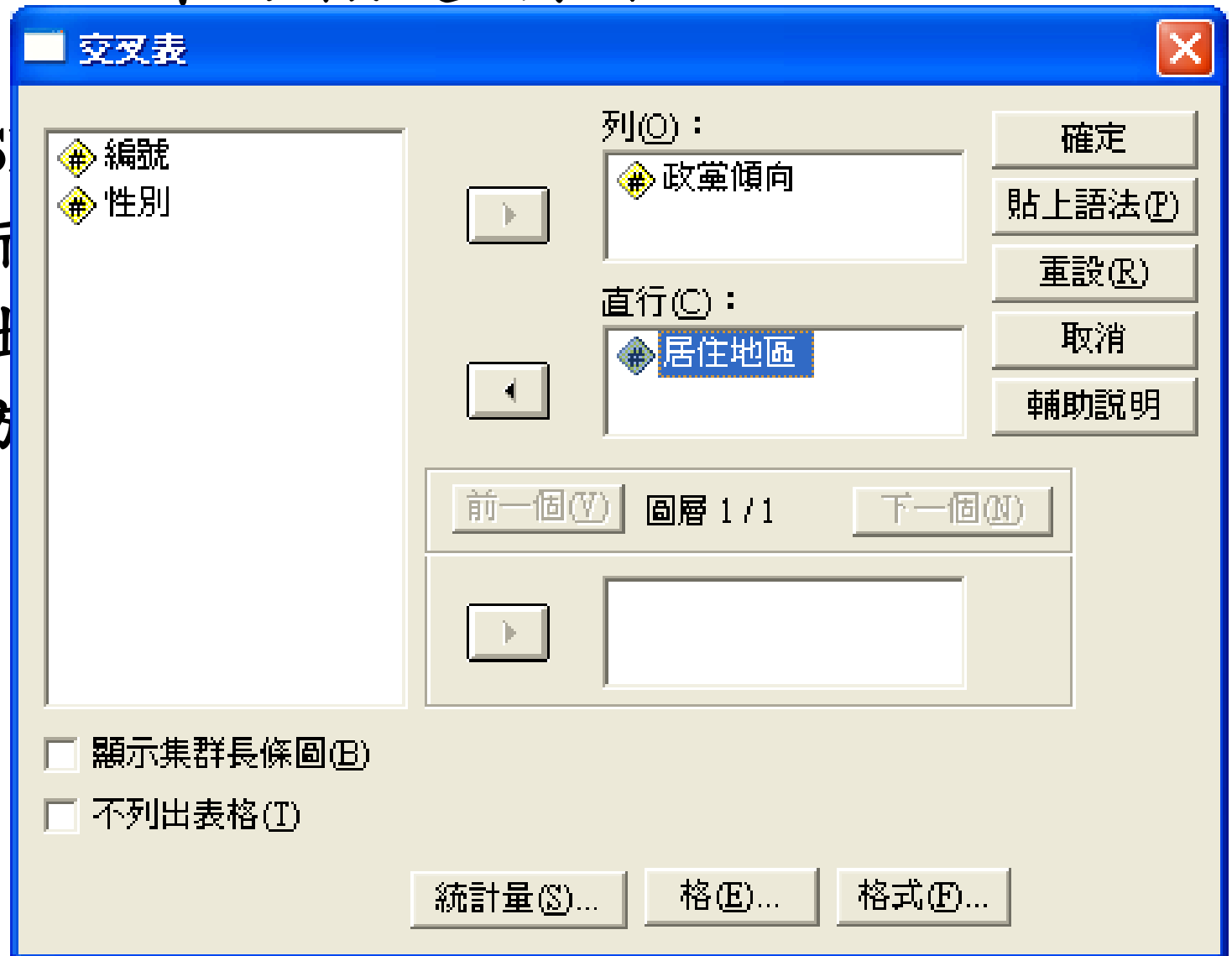
χ^2

2

- 計算出卡方值為32.743後，再計算自由度：
 $(r-1)*(c-1)$
r為列數、c為欄數。本例之自由度為 $3 \times 5 = 15$ 。
- 最後，依自由度查『卡方分配的臨界值』，比較所計算之卡方值，是否超過所指定顯著水準（ $\alpha=0.05$ ）的臨界值？若超過，則應棄卻欄變數與列變數並無關聯之虛無假設。反之，則否。
- 查『卡方分配的臨界值』，於自由度15、 $\alpha=0.05$ ，其臨界值為25.00。而我們所求算出之卡方值 $32.743 > 25.00$ ，故應棄卻政黨傾向與地區別無關之虛無假設。也就是說，政黨支持率會隨地區別不同而有顯著差異。

卡方檢定-利用SPSS

- 利用SPSS
還顯示
判斷出
對話方



- 按 **統計量(S)...** 鈕，選擇要求得「卡方統計量(H)」

交叉表：統計量 ✕

卡方統計量(H) 相關(R)

名義的 **次序的**

列聯係數(O) Gamma 參數(G)

Phi 與 Cramer's V 係數(F) Somers' d 值(S)

Lambda 值(L) Kendall's tau-b 統計量數(B)

不確定係數(U) Kendall's tau-c 統計量數(C)

名義變數對等距變數 Kappa 統計量數(K)

Eta 值(E) 風險(I)

McNemar 檢定(M)

Cochran's 與 Mantel-Haenszel 統計量(A)

檢定共同 odds 比率等於：

繼續
取消
輔助說明

- 按〔繼續〕鈕，回上一層對話方塊。再按〔確定〕鈕，則除了顯示層亦回表外
卡方檢定

居住地區		數值	自由度	漸近顯著性 (雙尾)		
	Pearson 卡方	32.743 ^a	15	.005		
	概似比	32.699	15	.005		
	線性對線性的關連	3.131	1	.077		
	有效觀察值的個數	1000				
a. 2格 (8.3%) 的預期個數少於 5。最小的預期個數為 1.80。						
	無	30.7%	26.5%	25.3%	28.0%	28.1%
總和		100.0%	100.0%	100.0%	100.0%	100.0%

- 還可獲致卡方檢定之結果：其卡方值為32.743，顯著性（雙尾）為0.005。

卡方檢定

	數值	自由度	漸近顯著性 (雙尾)
Pearson 卡方	32.743 ^a	15	.005
概似比	32.699	15	.005
線性對線性的關連	3.131	1	.077
有效觀察值的個數	1000		

a. 2格 (8.3%) 的預期個數少於 5。最小的預期個數為 1.80。

- 判斷檢定結果時很簡單，只須看此顯著性是否小於所指定顯著水準之 α 值。若是，即表示交叉表兩個變項間存有顯著關聯，可省去查表之麻煩。
- 本例，由於其顯著性 $0.005 < \alpha = 0.05$ ，所以應棄卻政黨傾向與地區別無關之虛無假設。故而，我們判定政黨支持率與居住地區存有顯著關聯。

判讀並解釋

- 判讀表內那一個百分比值得注意並解釋，於求縱向百分比之表內，應以橫向進行觀察。可拿一把尺來一橫，找出該列幾個較高之百分比即可。若該列之總百分比太低，也可以不予分析：

政黨傾向 * 居住地區 交叉表

居住地區內的 %

		居住地區				總和
		北區	中區	南區	東區	
政黨 傾向	民進黨	18.9%	21.5%	31.7%	10.0%	22.8%
	國民黨	22.4%	18.5%	17.0%	30.0%	20.4%
	親民黨	20.2%	22.5%	16.7%	20.0%	19.6%
	台聯	4.0%	6.5%	7.0%	6.0%	5.5%
	新黨	3.8%	4.5%	2.3%	6.0%	3.6%
	無	30.7%	26.5%	25.3%	28.0%	28.1%
總和		100.0%	100.0%	100.0%	100.0%	100.0%

- 分析時，先就最右側之總百分比，做一概述：

整體言，受訪者中，以無政黨傾向者居最高（28.1%）。有明顯政黨傾向之受訪者中，以民進黨的支持率最高（22.8%），其次依序為國民黨（20.4%）、親民黨（19.6%）、台聯（5.5%）、新黨（3.6%）。

- 然後，對各欄之數字高低，依序說明一下：

經由卡方檢定，其顯著性 $0.005 < \alpha = 0.05$ ，應棄卻政黨傾向與地區別無關之虛無假設。亦即，政黨支持率與居住地區存有顯著關聯。以地區別來看，北區主要是支持國民黨與親民黨（22.4%與20.2%），中區是親民黨（22.5%）與民進黨（21.5%）略勝於國民黨（18.5%），南區是民進黨獨大（31.7%），東區則還是國民黨（30.0%）與親民黨的天下（20.0%）。

- 最後，才對以橫向進行觀察，所標出之各該列的幾個較高之百分比進行說明：

以所支持之政黨別來看，民進黨之支持者主要集中於南區；國民黨與親民黨等泛藍政黨之支持者主要集中於北區與東區。而值得注意的是：居多數之無政黨傾向者，也是以北部多於中區與南區，這些人才是將來各黨要爭取之主要對象。

應注意下列事項

- 卡方檢定僅適用於類別資料（名目變數，如：性別、地區、政黨傾向、宗教信仰、是否有手機、……）。
- 各儲存格之期望次數不應少於5。通常要有80%以上的儲存格期望次數 ≥ 5 ，否則會影響其卡方檢定的效果。若有期望次數小於5時，可將其合併。

- 由於，各儲存格之期望次數不應少於5。通常要有80%以上的儲存格期望次數 ≥ 5 ，否則會影響其卡方檢定的效果。故而，SPSS之卡方檢定結果，會於最底下計算期望值 < 5 之儲存格比例：

卡方檢定

	數值	自由度	漸近顯著性 (雙尾)
Pearson 卡方	32.743 ^a	15	.005
概似比	32.699	15	.005
線性對線性的關連	3.131	1	.077
有效觀察值的個數	1000		

a. 2格 (8.3%) 的預期個數少於 5。最小的預期個數為 1.80。

- 本例，期望值（個數） < 5 之儲存格比例，僅8.3%而已，故並無需進行任何調整。
- 萬一，本例發生有期望值 < 5 之儲存格比例達20%以上之情形，要進行合併時，也應將政黨性質相近者進行合併。如，將台聯併入民進黨組成一新的集合：泛綠；而將國民黨、親民黨與新黨組成一新的集合：泛藍。
- 另外一種，避免期望值 < 5 之儲存格比例達20%以上的處理方法為：將樣本數比較少之『台聯』與『新黨』排除於分析之外。但這樣的作法較受爭議，因為無法表示出這兩黨之支持者的現象！

縮減組數--組距分組

- 無論是文字或數字，於交叉表中，均是將不重複出現之內容視為一個類別，去求算交叉表之相關統計數字。當碰上重複性較低之數字，很可能每一個數值均是唯一，而產生幾乎無法縮減其類別之情況。

	編號	性別	運動時間
1	1	1	120
2	2	1	10
3	3	2	0
4	4	2	120
5	5	1	120
6	6	1	15
7	7	1	150

- 以「分析(A)/描述性統計(E)/交叉表(C)...」求性別對運動時間之交叉表：



- 其結果，有很多種運動時間係獨立存在，產生一系列內容。由於組數太多，於資料分析時並無多大作用：

運動時間 * 性別 交叉表

			性別		總和
			男	女	
運動時間	0	個數	3	7	10
		性別內的%	5.1%	12.5%	8.7%
	10	個數	1	1	2
		性別內的%	1.7%	1.8%	1.7%
	15	個數	1	2	3
		性別內的%	1.7%	3.6%	2.6%
	30	個數	2	9	11
		性別內的%	3.4%	16.1%	9.6%
	40	個數		1	1
		性別內的%		1.8%	.9%
	45	個數	1		1
		性別內的%	1.7%		.9%
	60	個數	17	8	25
		性別內的%	28.8%	14.3%	21.7%
	70	個數		1	1
		性別內的%		1.8%	.9%
	90	個數	7	9	16
		性別內的%	11.9%	16.1%	13.9%

- 較理想之方式為：將每次運動時間分組，以縮減其組數。若以「轉換(T)/重新編碼(R)/成不同變數(D)...」，將每次運動時間分為0~30、31~60、61~90、91~120與121~五組：

重新編碼成不同變數：舊值與新值

舊值

- 數值(V)：
- 系統界定的遺漏值(S)
- 系統或使用者界定的遺漏值(U)
- 範圍(N)：
 到(T)
- 範圍(G)：
從最低值
- 範圍(E)：
 到最高值
- 全部其他值(O)

新值為

- 數值(L)：
- 複製舊值(P) 系統界定的遺漏值(Y)

舊值 --> 新值(D)：

新增(A) 變更(C) 刪除(R)

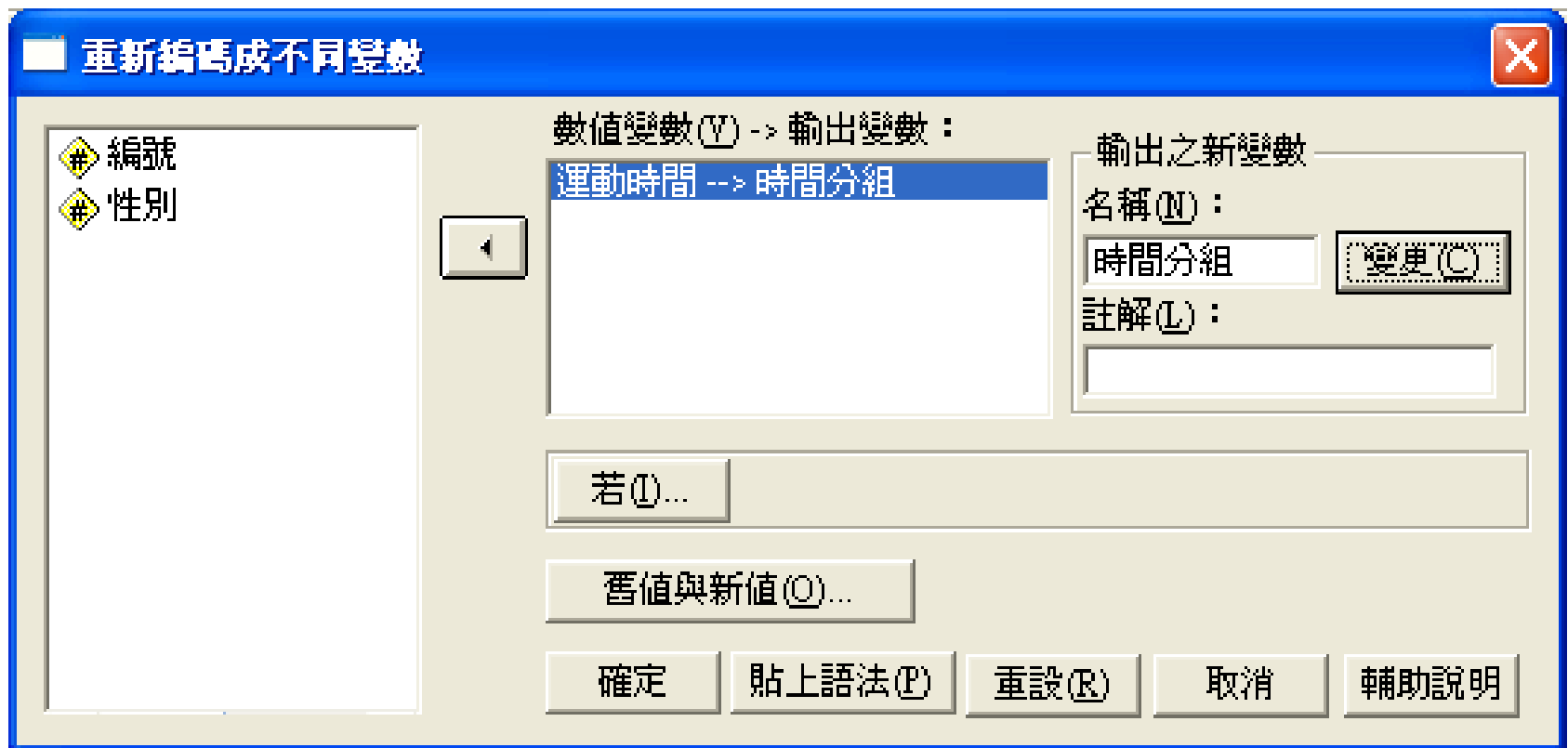
```
Lowest thru 30 --> 1
31 thru 60 --> 2
61 thru 90 --> 3
91 thru 120 --> 4
121 thru Highest --> 5
```

輸出變數為字串(B) 寬度(W)：

將數值字串轉換為數字 (5'->5)(M)

繼續 取消 輔助說明

- 將『重新編碼』結果安排到『時間分組』：



- 續以『時間分組』與『性別』重建一次交叉表：



- 即可得到經縮減組數後之交叉表：

時間分組 * 性別 交叉表

			性別		總和
			男	女	
時間分組	~30	個數	7	19	26
		性別內的 %	11.9%	33.9%	22.6%
	31~60	個數	18	9	27
		性別內的 %	30.5%	16.1%	23.5%
	61~90	個數	7	10	17
		性別內的 %	11.9%	17.9%	14.8%
	91~120	個數	20	15	35
		性別內的 %	33.9%	26.8%	30.4%
	121~	個數	7	3	10
		性別內的 %	11.9%	5.4%	8.7%
總和		個數	59	56	115
		性別內的 %	100.0%	100.0%	100.0%

- 期望值個數 < 5 之儲存格比例為 10%，並未超過 20%。表格無須再行合併以縮減組別。
- 卡方值為 11.312，自由度為 4，其顯著水準 $0.023 < \alpha = 0.05$ 。所以，應棄卻運動時間長短與性別無關之虛無假設。

卡方檢定

	數值	自由度	漸近顯著性 (雙尾)
Pearson 卡方	11.312 ^a	4	.023
概似比	11.629	4	.020
線性對線性的關連	4.044	1	.044
有效觀察值的個數	115		

a. 1格 (10.0%) 的預期個數少於 5。最小的預期個數為 4.87。

- 由表可知，全體受訪者運動時間主要以91~120分鐘居多30.4%。交叉分析後，可發現男性之運動時間主要以91~120分鐘居多（33.9%）；而女性之運動時間主要以0~30分鐘居多（33.9%）。
- 就相對比例言，在較長之運動時間部份（31~60、91~120分鐘與120分鐘~）的比例，男性明顯高過女性；而女性則在較短之運動時間部份（0~31與61~90分鐘）的比例，明顯高過男性。可見男性之運動時間普遍較女性來得長一點！

縮減類別

- 進行交叉分析表時，通常要有80%以上的儲存格期望次數 ≥ 5 ，否則會影響卡方檢定的效果。若有期望次數小於5時，可將其合併。

	編號	有手機	手機廠牌	性別
6	106	2	0	2
7	107	1	3	2
8	108	2	0	1
9	109	1	1	1
10	110	1	2	1
11	111	1	4	1

- 以『手機廠牌』交叉『性別』後，其結果為將無手機者（其『手機廠牌』=0）均納入分析，故資料並不正確；且一些廠牌的出現次數並不高。

手機廠牌 * 性別 交叉表

			性別		總和
			男	女	
手機廠牌	0	個數	26	46	72
		性別內的 %	33.8%	40.4%	37.7%
1. Nokia		個數	26	25	51
		性別內的 %	33.8%	21.9%	26.7%
2. Motorola		個數	7	17	24
		性別內的 %	9.1%	14.9%	12.6%
3. ERICSSON		個數	8	6	14
		性別內的 %	10.4%	5.3%	7.3%
4. SAGEM		個數	4	11	15
		性別內的 %	5.2%	9.6%	7.9%
5. Acer		個數	1	2	3
		性別內的 %	1.3%	1.8%	1.6%
6. SIEMENS		個數	2	3	5
		性別內的 %	2.6%	2.6%	2.6%
7. PHILIPS		個數	1	1	2
		性別內的 %	1.3%	.9%	1.0%
8. Panasonic		個數	1	2	3
		性別內的 %	1.3%	1.8%	1.6%
9. 其他		個數	1	1	2
		性別內的 %	1.3%	.9%	1.0%
總和		個數	77	114	191
		性別內的 %	100.0%	100.0%	100.0%

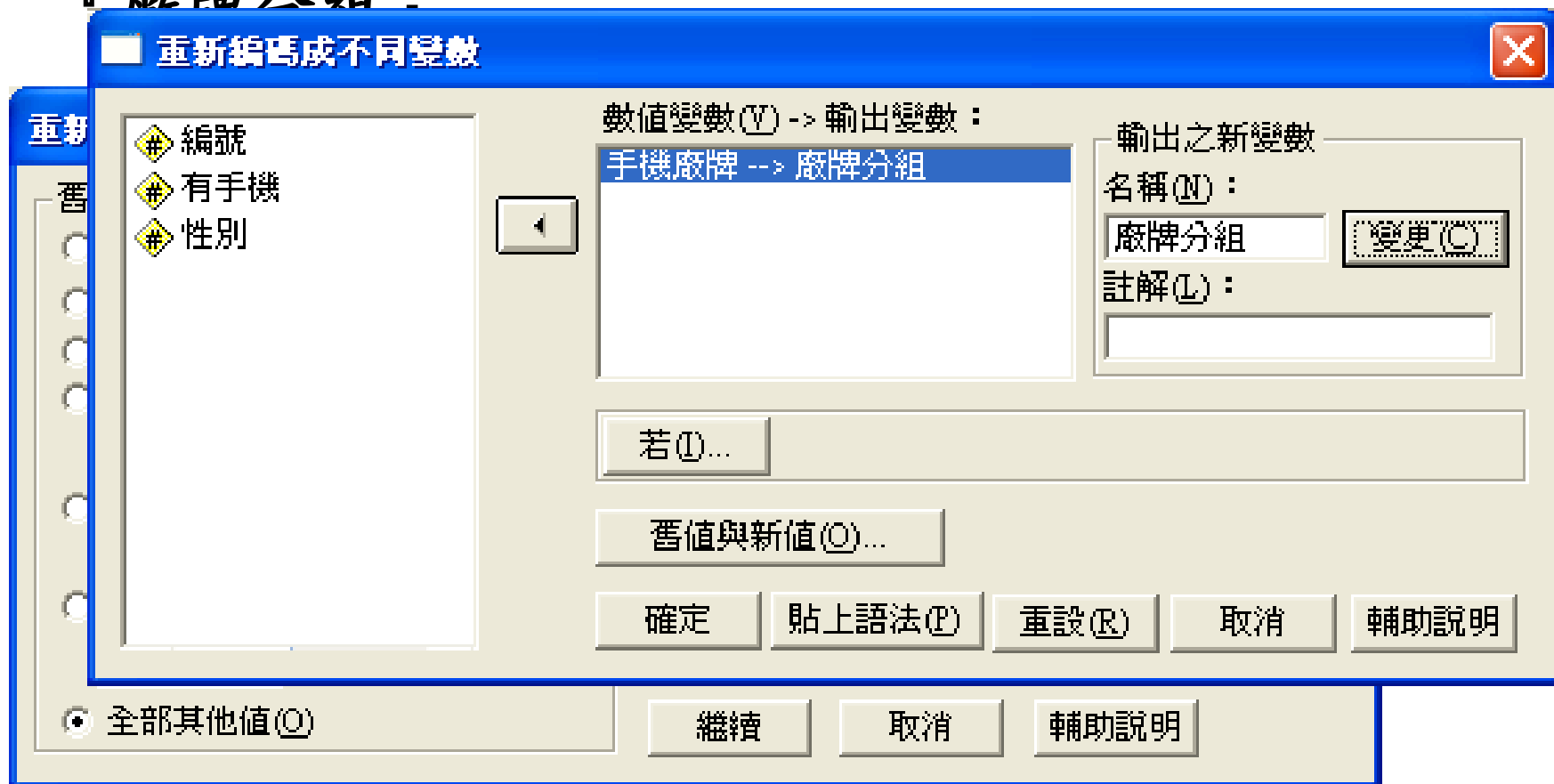
- 此外，其卡方檢定之結果顯示期望次數 < 5 者有 50%，超過 20%，故得將組數進行縮減。

卡方檢定

	數值	自由度	漸近顯著性 (雙尾)
Pearson 卡方	7.266 ^a	9	.609
概似比	7.297	9	.606
線性對線性的關連	.003	1	.955
有效觀察值的個數	191		

a. 10 格 (50.0%) 的預期個數少於 5。最小的預期個數為 81。

- 首先，執行「轉換(T)/重新編碼(R)/成不同變數(D)...」，先將答案為5~9者合併為『9.其他』（因其等之次數最高僅為5而已），存入另一新變數『廠牌分組』。



- 以縮減其組數，並將『廠牌分組』答案為0設定為遺漏值，將其排除掉：

遺漏值

無遺漏值 (N)

離散遺漏值 (D)

範圍加上一個選擇性的遺漏值 (R)

低 (L) : 高 (H) :

離散值 (S) :

確定

取消

輔助說明

- 續以『廠牌分組』與『性別』重建一次交叉表：



- 即可得到經縮減組數後之交叉表：

廠牌分組 * 性別 交叉表

			性別		總和
			男	女	
廠牌 分組	1. Nokia	個數	26	25	51
		性別內的 %	51.0%	36.8%	42.9%
	2. Motorola	個數	7	17	24
		性別內的 %	13.7%	25.0%	20.2%
	3. ERICSSON	個數	8	6	14
		性別內的 %	15.7%	8.8%	11.8%
	4. SAGEM	個數	4	11	15
		性別內的 %	7.8%	16.2%	12.6%
	9. 其他	個數	6	9	15
		性別內的 %	11.8%	13.2%	12.6%
總和		個數	51	68	119
		性別內的 %	100.0%	100.0%	100.0%

- 期望值個數 < 5 之儲存格比例為 0% ，表格無須再行合併以縮減組別。本例之卡方值為 6.033 ，自由度為 4 ，其顯著水準 $0.197 > \alpha = 0.05$ 。所以，無法棄卻使用之手機廠牌與性別無關之虛無假設。

卡方檢定

	數值	自由度	漸近顯著性 (雙尾)
Pearson 卡方	6.033 ^a	4	.197
概似比	6.167	4	.187
線性對線性的關連	.520	1	.471
有效觀察值的個數	119		

a. 0 格 (0%) 的預期個數少於 5。最小的預期個數為 6.00。

- 撰寫報告時，對於卡方檢定結果顯示兩變數間無關之交叉分析表，僅須就其最右側之欄百分比進行解釋即可：
- 整體言，受訪者所使用之手機廠牌，主要以Nokia為最高（42.9%），其次為Motorola（20.2%）；再其次為SAGEN（12.6%）與ERICSSON（11.8%）。
- 由於兩者無關，就不必再對交叉結果進行說明了！（通常，於報告中也會將此交叉表省略，以縮減篇幅。僅敘述一下其檢定結果並不顯著即可）

卡方檢定的類型與範例

- 獨立性檢定(test of independence)

用以檢測同一樣本的兩個類別變項 (X 與 Y) 之間的關聯性。

例如：電影製作公司可能有興趣知道不同型態電影的喜好度是否與觀眾年齡有關、股票分析師也許希望分析熱門股與產業別是否有關、政治分析家可能希望研究候選人的得票數是否會受到選民省籍的影響。這些種種探討兩變數間是否獨立的問題，均可利用卡方統計量來進行分析。

獨立性檢定(test of independence)

- 卡方獨立性檢定可以用來分析各種型態的資料，尤其是最常用於名目尺度的資料。在資料分析中的兩個變數代表兩種不同的分類方法，資料經過兩種不同的分類方法所觀察到的結果，通常可交叉整理成列聯表。
- 列聯表 (contingency table)：將資料用兩種不同分類方法，計次分類結果用行與列共同表列，如下頁表格所示。

$r \times c$ 列聯表

		B				
A	B ₁	B ₂	...	B _c	總計	
A ₁	O ₁₁	O ₁₂	...	O _{1c}	O _{1.}	
A ₂	O ₂₁	O ₂₂	...	O _{2c}	O _{2.}	
⋮	⋮	⋮	⋮	⋮	⋮	
A _r	O _{r1}	O _{r2}	...	O _{rc}	O _{r.}	
總計	O _{.1}	O _{.2}	...	O _{.c}	O _{..} = n	

獨立性檢定(test of independence)

● 卡方獨立性檢定過程

1. H_0 ：列分類與行分類獨立

2. H_1 ：列分類與行分類不獨立

3. 檢定統計量：
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

χ^2 ：自由度為 $(r-1)(c-1)$ 之卡方統計量。

4. 拒絕域： $\chi^2 > \chi_{(r-1)(c-1), \alpha}^2$

5. p值： $P(\chi_{(r-1)(c-1)}^2 > \chi^{*2})$

獨立性檢定(test of independence)

- 若虛無假設正確，則樣本資料在列聯表中之期望次數為行總和與列總和的乘積再除以樣本大小。

$$E_{ij} = n \cdot \frac{O_{i.}}{n} \cdot \frac{O_{.j}}{n} = \frac{O_{i.} \cdot O_{.j}}{n} \quad ; \quad i=1,2,\dots,r \quad j=1,2,\dots,r$$

- 與卡方適合度檢定相同，卡方獨立性檢定通常只適用於大樣本，一般說來，當所有的 $E_{ij} \geq 5$ 時，卡方適合度檢定才適用。若有 $E_{ij} < 5$ 時，很容易產生異常大的卡方檢定統計量，因此該行或該列分類就必須與相鄰行或列合併至 $E_{ij} \geq 5$ 為止。

獨立性檢定-理財方式與年齡

- 隨機選取的 500 名成人年齡與其最喜愛的投資方式列聯表，欲檢定人們喜好的投資理財方式是否會因年齡有所不同，可設定虛無與對立假設為
 - H_0 ：年齡大小與喜愛的投資方式無關
 - H_1 ：年齡大小與喜愛的投資方式有關
- 卡方檢定統計量之自由度為 $(r-1)(c-1) = (3-1)(4-1) = 6$ 。透過卡方分配表得知臨界值 $\chi_{6,0.05}^2 = 12.5916$ 。

獨立性檢定-理財方式與年齡

- 500 名成年人喜好的理財方式與年齡之觀察次數（期望次數）如下表所示：

年齡	理財方式				總計
	基金	股票	債券	定存	
21-35	36(38.5)	60(53.284)	30(36.652)	28(25.564)	154
35-55	46(39)	49(53.976)	41(37.128)	20(25.896)	156
> 55	43(47.5)	64(65.74)	48(45.22)	35(31.54)	190
總計	125	173	119	83	500

- 卡方檢定統計量為 6.9324 小於臨界值 12.5916，在 0.05 的顯著水準下，我們無法拒絕虛無假設；換言之，我們沒有足夠的證據顯示年齡大小與喜愛的投資方式有關。

獨立性檢定-理財方式與年齡

- 若虛無假設正確，則不同年齡與理財方式組合下之期望次數可計算為

$$E_{ij} = n \cdot \frac{O_{i\cdot}}{n} \cdot \frac{O_{\cdot j}}{n} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n}$$

- (21-35) 與「基金」：
$$E_{11} = \frac{154 \times 125}{500} = 38.5$$
- (21-35) 與「股票」：
$$E_{12} = \frac{154 \times 173}{500} = 53.284$$
- 其餘計算以此類推。

獨立性檢定

例題

◎ 假設某研究者想檢定民眾之性別與搭乘高鐵意願是否有顯著地相關，於是隨機抽樣得到500位民眾之性別與搭乘高鐵意願之情形如下表。請在 H_0 : 民眾之性別與搭乘高鐵意願無關之假設成立條件下，計算每個項目之期望次數。

性別	搭乘高鐵意願		總和
	無意願(N)	有意願(Y)	
男(M)	75	225	300
女(F)	25	175	200
總和	100	400	500

獨立性檢定

解

以 M 、 F 、 Y 、 N 分別表男性、女性、有意願、無意願。在 H_0 成立之條件下，兩兩事件均為獨立，即

$$P(M \cap Y) = P(M) \cdot P(Y) \quad P(M \cap N) = P(M) \cdot P(N)$$

$$P(F \cap Y) = P(F) \cdot P(Y) \quad P(F \cap N) = P(F) \cdot P(N)$$

因此在上述總樣本數500中，其期望個數分別為

$$E(M \cap Y) = 500 \times P(M \cap Y) = 500 \times P(M) \cdot P(Y) = 500 \times \frac{300}{500} \times \frac{400}{500} = 240$$

$$E(M \cap N) = 500 \times P(M \cap N) = 500 \times P(M) \cdot P(N) = 500 \times \frac{300}{500} \times \frac{100}{500} = 60$$

$$E(F \cap Y) = 500 \times P(F \cap Y) = 500 \times P(F) \cdot P(Y) = 500 \times \frac{200}{500} \times \frac{400}{500} = 160$$

$$E(F \cap N) = 500 \times P(F \cap N) = 500 \times P(F) \cdot P(N) = 500 \times \frac{200}{500} \times \frac{100}{500} = 40$$

獨立性檢定

解

因此可將觀察次數與期望次數如下表所示：

性別	搭乘高鐵意願		總和
	無意願(N)	有意願(Y)	
男(M)	75(60)	225(240)	300
女(F)	25(40)	175(160)	200
總和	100	400	500

獨立性檢定

定理1

在屬性一與屬性二之獨立性檢定中，令 O_{ij} 、 E_{ij} 分別表屬性一第 i 等級、屬性二第 j 等級之觀察次數與期望次數，則當樣本數 $n \rightarrow \infty$ 時，在 H_0 ：兩屬性相互獨立成立之條件下，

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2((r-1)(c-1))$$

其中， r 、 c 分別表屬性一及屬性二之等級數。

獨立性檢定

(三) 決策法則

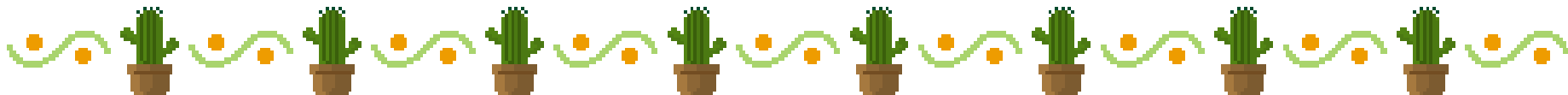
當 H_0 : 兩屬性獨立 & H_1 : 兩屬性相關，則以

$$\chi_0^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{為檢定值，其中}$$

$$E_{ij} = \frac{\sum_{s=1}^c O_{is} \times \sum_{s=1}^r O_{sj}}{n}$$

， n 為總樣本數，其檢定之拒絕

域為 $\{\chi_0^2 \geq \chi_\alpha^2((r-1)(c-1))\}$ ， P 值 = $P(\chi^2 \geq \chi_0^2)$ 。



獨立性檢定

例題

◎承上例，請以 $\alpha = 0.05$ 檢定此樣本資料是否顯示民眾之性別與搭乘高鐵意願與否無關。

解

其檢定之拒絕域為 $\{\chi^2 \geq \chi_{0.05}^2((2-1)(2-1)) = \chi_{0.05}^2(1) = 3.841\}$

檢定值

$$\begin{aligned}\chi^2 &= \sum_{j=1}^2 \sum_{i=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(75 - 60)^2}{60} + \frac{(25 - 40)^2}{40} + \frac{(225 - 240)^2}{240} + \frac{(175 - 160)^2}{160} \\ &= 11.71875 > 3.841\end{aligned}$$

落在拒絕域中，因此拒絕 H_0 ，即民眾之性別與搭乘高鐵意願有顯著地相關。

獨立性檢定

例題

- ◎ 假設某人想瞭解性別與其每月信用卡消費金額是否有顯著地相關，於是隨機抽查50位該地區信用卡消費者，得其結果如下表所示(單位：萬元)。

性別	每月消費金額				總和
	0~2	2~4	4~6	6以上	
男(M)	45	76	38	61	220
女(F)	70	112	72	26	280
總和	115	188	110	87	500

請問此資料是否顯示性別與其每月信用卡消費金額有顯著地相關？($\alpha = 0.05$)

獨立性檢定

解

依題意，可建立假設如下：

H_0 : 信用卡消費金額與性別無關 & H_1 : 信用卡消費金額與性別有關

利用卡方檢定之獨立性檢定，其拒絕域為

$$\{\chi_0^2 \geq \chi_{0.05}^2((4-1)(2-1)) = \chi_{0.05}^2(3) = 7.815\}$$

在 H_0 為真的條件下，資料與期望次數

$$E_{11} = \frac{220 \times 115}{500} = 50.6,$$

$$E_{12} = \frac{220 \times 188}{500} = 82.72,$$

$$E_{13} = \frac{220 \times 110}{500} = 48.4,$$

獨立性檢定

解

$$E_{14} = \frac{220 \times 87}{500} = 38.28, \quad E_{21} = \frac{280 \times 115}{500} = 64.4,$$

$$E_{22} = \frac{280 \times 188}{500} = 105.28, \quad E_{23} = \frac{280 \times 110}{500} = 61.6,$$

$$E_{24} = \frac{280 \times 87}{500} = 48.72,$$

由此可得其觀察次數與期望次數如下表所示。

性別	每月消費金額				總和
	0~2	2~4	4~6	6以上	
男(M)	45(50.6)	76(82.72)	38(48.4)	61(38.28)	220
女(F)	70(64.4)	112(105.28)	72(61.6)	26(48.72)	280
總和	115	188	110	87	500

獨立性檢定

解

因此得其檢定值

$$\chi_0^2 = \frac{(45-50.6)^2}{50.6} + \frac{(76-82.72)^2}{82.72} + \frac{(38-48.4)^2}{48.4} + \frac{(61-38.28)^2}{38.28} + \frac{(70-64.4)^2}{64.4} \\ + \frac{(112-105.28)^2}{105.28} + \frac{(72-61.6)^2}{61.6} + \frac{(26-48.72)^2}{48.72} = 30.152 > 7.815$$

落在拒絕域中，拒絕 H_0 ，即此地區民眾使用信用卡之消費金額與性別有顯著地相關。

範例一

- 以「銀行客戶資料」為例，探討客戶信用等級與居住區域是否有關係。建立假說如下：
 - H_0 ：客戶信用等級與居住區域無關
 - H_1 ：客戶信用等級與居住區域有關
- 操作：
 - 1. Analyze/Descriptive Statistics/Crosstabs
 - 2. 程式操作
 - 3. 分析結果

範例一——程式操作

銀行客戶 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1: 客戶編號 1

	客戶編號	性別	年齡	居住區域	信用等級	職業	所得	存款	貸
1	1	0	20	2	1	3	118.33	286.88	
2	2	0	24	2	2	3	189.68	408.26	
3	3	1	21	2	2	3	208.58	353.11	
4	4	0	20	1	2	3	185.66	370.38	
5	5	1	56	1	1	5	319.57	608.65	
6	6	1	35	1	3	5	319.78	588.73	
7	7	1	50	1	2	3	204.95	412.27	
8	8	1	48	3	1	3	98.95	206.92	
9	9	1	64	2	2	3	200.36	396.01	
10	10	0	55	2	2	3	185.98	419.72	
11	11	0	20	2	2	3	187.28	394.54	
12	12	1	29	2	3	5	318.43	562.52	
13	13	0	20	3	3	4	253.05	530.57	
14	14	1	21	3	2	3	201.90	334.51	
15	15	0	50	2	2	3	184.36	422.58	
16	16	1	52	3	3	4	278.47	485.03	
17	17	1	35	2	2	3	202.03	370.05	
18	18	0	28	3	2	4	229.28	504.73	
19	19	1	21	1	3	5	313.97	576.40	
20	20	0	57	2	2	3	180.37	417.69	
21	21	0	47	1	2	3	189.79	400.73	
22	22	0	21	1	2	4	224.51	447.60	
23	23	0	53	3	2	4	223.72	512.86	
24	24	0	20	2	3	4	253.78	530.41	

Data View / Variable View / SPSS Processor is ready



範例一—分析結果

居住區域 * 信用等級 Crosstabulation

		信用等級			Total
		1	2	3	
居住區域	1	11	37	16	64
	2	15	49	19	83
	3	15	27	10	52
Total		41	113	45	199

Chi-Square Tests^a

	Value.	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.065(a)	4	.547
Likelihood Ratio	2.921	4	.571
Linear-by-Linear Association	1.904	1	.168
N of Valid Cases	199		

Chi-sq = 3.065 , **Sig. = 0.547 > 0.05** , 無法拒絕 H_0 。

結論：客戶信用等級與居住區域無顯著關係。



範例二

- 根據全校成績資料檔，探討居住區域與入學方式之關係。建立假說如下：
 - H_0 ：全校學生入學方式與居住區域無關
 - H_1 ：全校學生入學方式與居住區域有關
- 操作：
 - 1. Analyze/Descriptive Statistics/Crosstabs
 - 2. 程式操作（如範例一）
 - 3. 改善前分析結果
 - 4. 改善方法
 - 5. 改善後分析結果

範例二—改善前分析結果

居住區域 * 入學方式 Crosstabulation

			入學方式			Total
			聯考	推薦甄試	保送	
居住區域	台灣北部	Count	54	8	0	62
		Expected Count	17.8	43.7	.5	62.0
	台灣中部 ^a	Count	12	39	0	51
		Expected Count	14.6	35.9	.4	51.0
	台灣南部	Count	0	51	0	51
		Expected Count	14.6	35.9	.4	51.0
	台灣東部	Count	0	49	0	49
		Expected Count	14.1	34.5	.4	49.0
	外島	Count	0	15	2	17
		Expected Count	4.9	12.0	.1	17.0
Total		Count	66	162	2	230
		Expected Count	66.0	162.0	2.0	230.0

Chi-Square Tests^a

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	175.372(a)	8	.000
Likelihood Ratio	181.679	8	.000
Linear-by-Linear Association	116.501	1	.000
N of Valid Cases	230		

a. 6 cells (40.0%) have expected count less than 5. The minimum expected count is .15.

注意：卡方分配要求每個格子(cell)之期望值儘量大於 5 個，一般卡方檢定的應用，期望次數低於 5 的方格數不能多於 20 %。



範例二—改善方法

● 先觀察次數分配

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	台灣北部	62	27.0	27.0	27.0
	台灣中部	51	22.2	22.2	49.1
	台灣南部	51	22.2	22.2	71.3
	台灣東部	49	21.3	21.3	92.6
	外島	17	7.4	7.4	100.0
	Total	230	100.0	100.0	

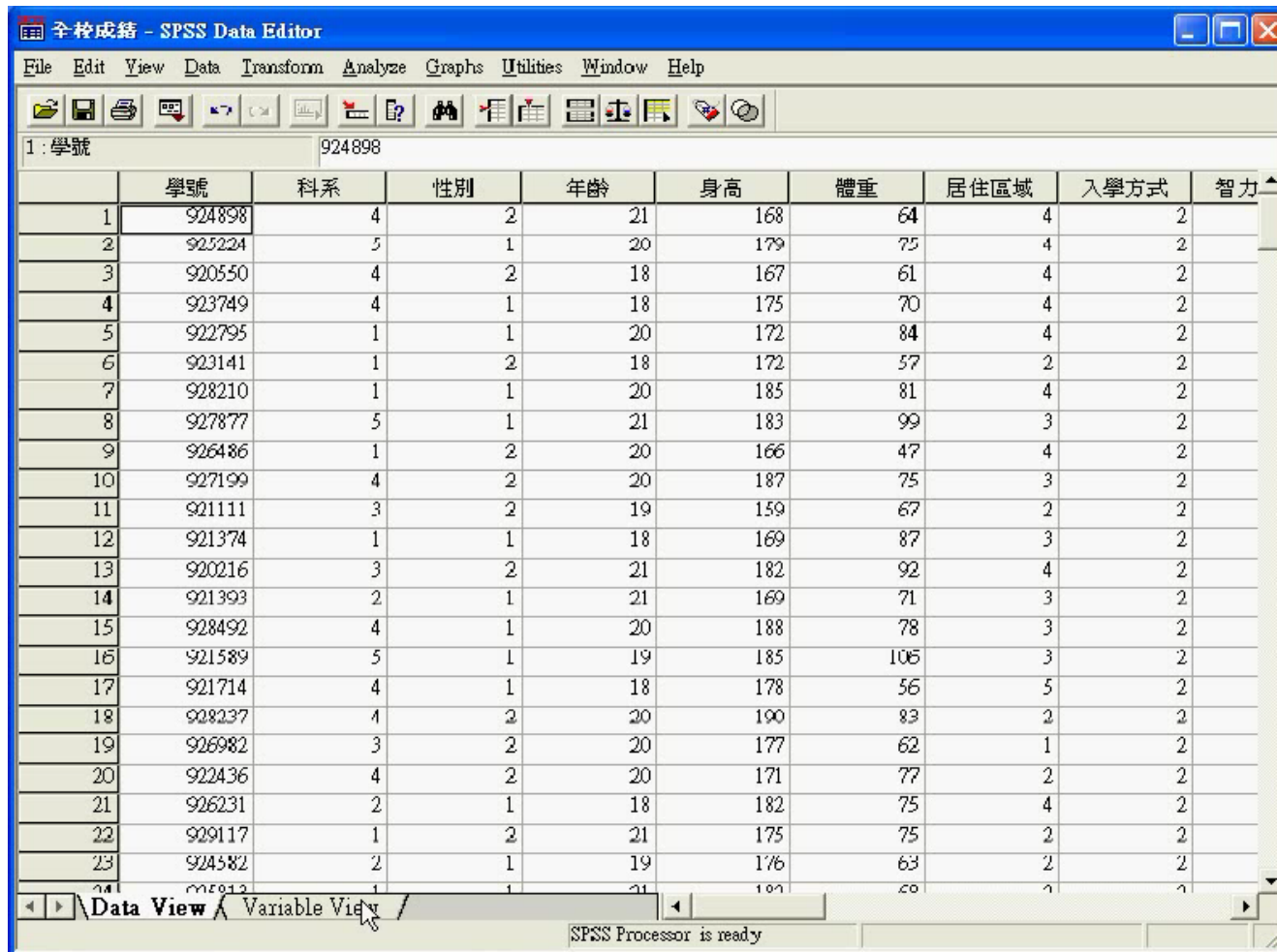
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	聯考	66	28.7	28.7	28.7
	推薦甄試	162	70.4	70.4	99.1
	保送	2	.9	.9	100.0
	Total	230	100.0	100.0	

● 針對樣本較少之類別

- 將目前分類歸併（例如：保送納入甄試）。
- 加以忽略（設定遺漏值）。

範例二—改善方法

操作程序如下：



The screenshot shows the SPSS Data Editor window titled "全校成績 - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main data grid has the following columns: 學號 (Student ID), 科系 (Department), 性別 (Gender), 年齡 (Age), 身高 (Height), 體重 (Weight), 居住區域 (Residence Area), 入學方式 (Admission Method), and 智力 (Intelligence). The data is organized into rows, with the first row highlighted in blue. The status bar at the bottom indicates "SPSS Processor is ready".

	學號	科系	性別	年齡	身高	體重	居住區域	入學方式	智力
1	924898	4	2	21	168	64	4	2	
2	925224	5	1	20	179	75	4	2	
3	920550	4	2	18	167	61	4	2	
4	923749	4	1	18	175	70	4	2	
5	922795	1	1	20	172	84	4	2	
6	923141	1	2	18	172	57	2	2	
7	928210	1	1	20	185	81	4	2	
8	927877	5	1	21	183	99	3	2	
9	926486	1	2	20	166	47	4	2	
10	927199	4	2	20	187	75	3	2	
11	921111	3	2	19	159	67	2	2	
12	921374	1	1	18	169	87	3	2	
13	920216	3	2	21	182	92	4	2	
14	921393	2	1	21	169	71	3	2	
15	928492	4	1	20	188	78	3	2	
16	921589	5	1	19	185	105	3	2	
17	921714	4	1	18	178	56	5	2	
18	928237	4	2	20	190	83	2	2	
19	926982	3	2	20	177	62	1	2	
20	922436	4	2	20	171	77	2	2	
21	926231	2	1	18	182	75	4	2	
22	929117	1	2	21	175	75	2	2	
23	924582	2	1	19	176	63	2	2	
24	925912	1	1	21	182	68	2	2	



範例二—改善後分析結果

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	137.506(a)	3	.000
Likelihood Ratio	160.355	3	.000
Linear-by-Linear Association	111.550	1	.000
N of Valid Cases	213		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 15.18...

居住區域 * 入學方式 Crosstabulation

			入學方式		Total
			聯考	推薦甄試	
居住區域	台灣北部	Count	54	8	62
		Expected Count	19.2	42.8	62.0
	台灣中部	Count	12	39	51
		Expected Count	15.8	35.2	51.0
	台灣南部	Count	0	51	51
		Expected Count	15.8	35.2	51.0
	台灣東部	Count	0	49	49
		Expected Count	15.2	33.8	49.0
Total		Count	66	147	213
		Expected Count	66.0	147.0	213.0

結論：全校學生入學方式與居住區域有顯著關係。



範例三

- 大發量販店委託某研究單位，想了解消費者對該店的滿意度與忠誠度的關係。該研究單位以非隨機之便利抽樣法發放1,000份問卷，調查消費者對該店的滿意度與忠誠度，結果整理如下表。試問消費者對該店的滿意度與忠誠度是否有顯著的關係？

		忠誠度		
		高	中	低
滿意度	很滿意	312	143	43
	無意見	118	97	85
	不滿意	25	82	95

範例三

- 建立假說如下：

- H_0 ：滿意度與忠誠度無關

- H_1 ：滿意度與忠誠度有關

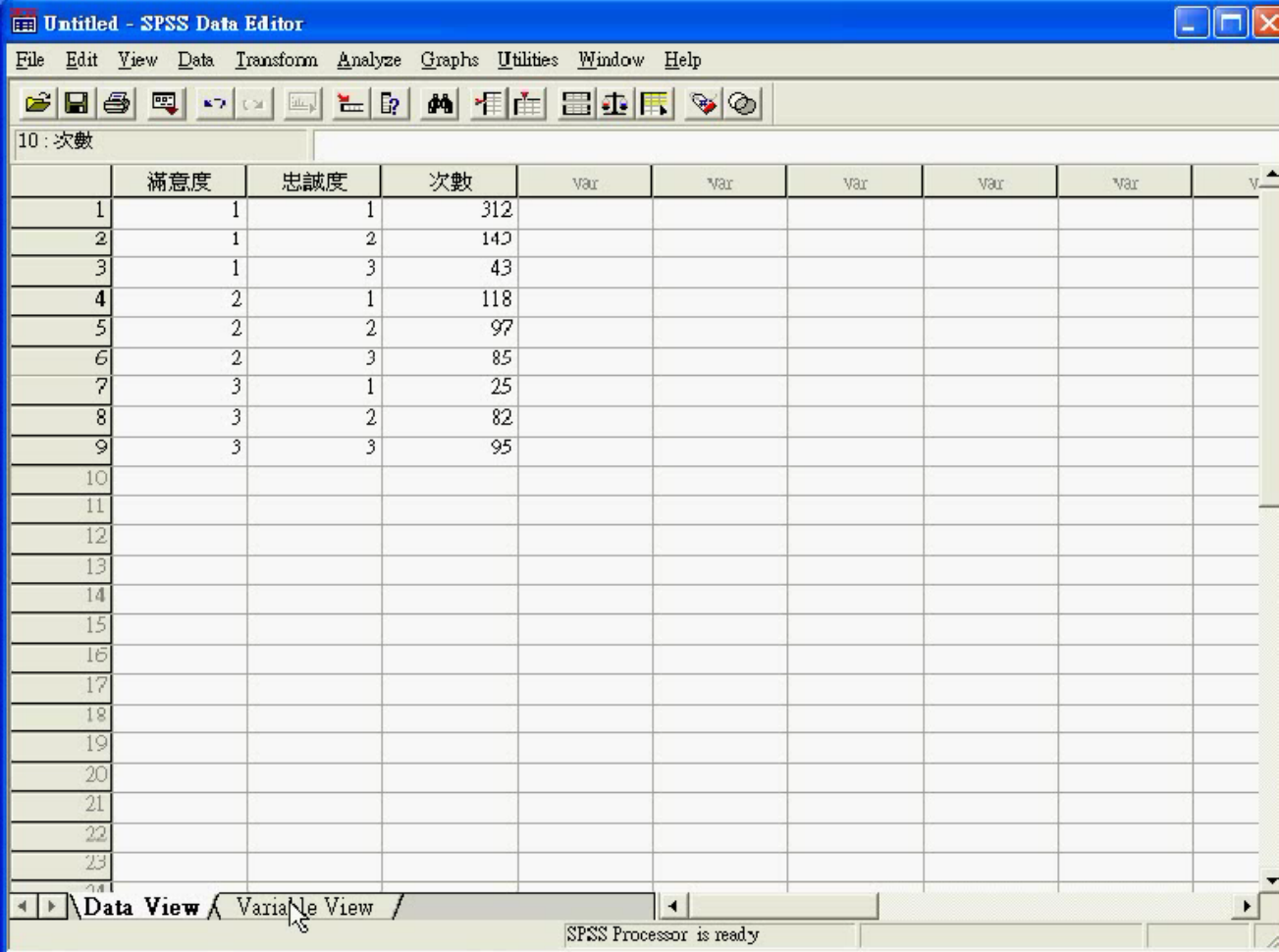
- 操作：

- 1. **Analyze/Descriptive Statistics/Crosstabs**

- 2. 程式操作

- 3. 分析結果

範例三一程式操作



The screenshot shows the SPSS Data Editor window titled "Untitled - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The data grid shows a table with 23 rows and 10 columns. The first three columns are labeled "滿意度", "忠誠度", and "次數". The first three rows of data are highlighted in blue. The status bar at the bottom indicates "SPSS Processor is ready".

	滿意度	忠誠度	次數	Var	Var	Var	Var	Var	Var
1	1	1	312						
2	1	2	140						
3	1	3	43						
4	2	1	118						
5	2	2	97						
6	2	3	85						
7	3	1	25						
8	3	2	82						
9	3	3	95						
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									



範例三—分析結果

滿意度 * 忠誠度 Crosstabulation

			忠誠度			Total
			1	2	3	
滿意度	1	Count	312	143	43	498
		Expected Count	226.6	160.4	111.1	498.0
	2	Count	118	97	85	300
		Expected Count	136.5	96.6	66.9	300.0
	3	Count	25	82	95	202
		Expected Count	91.9	65.0	45.0	202.0
Total		Count	455	322	223	1000
		Expected Count	455.0	322.0	223.0	1000.0

範例三—分析結果

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	191.709(a)	4	.000
Likelihood Ratio	207.042	4	.000
Linear-by-Linear Association	190.993	1	.000
N of Valid Cases	1000		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 4

Symmetric Measures^a

		Value	Approx. Sig.
Nominal by Nominal	Phi	.438	.000
	Cramer's V	.310	.000
	Contingency Coefficient	.401	.000
N of Valid Cases		1000	

Chi-sq = 191.709 , Sig. = 0.000 < 0.05 , 拒絕 H_0 。

結論：大發量販店消費者的滿意度與忠誠度有顯著的關係。其關聯性可由Cramer's V = 0.31(Sig. = 0.000 < 0.05)看出。



適合度檢定

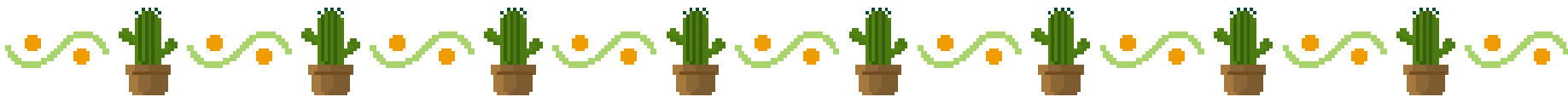
(一)適合度之概念

檢定某一資料之次數分配或相對次數分配是否合乎一特定之機率分配。

定理2

若 O_i 、 E_i 分別表適合度檢定之觀察次數與期望次數
且 k 表資料之類別個數、 m 表需估計母體之參數個數，
則當 $n \rightarrow \infty$ 時，在 H_0 成立條件下，

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k - m - 1) \quad , \quad \text{其中 } E_i = np_{i0} \quad \circ$$



適合度檢定

- 當決定顯著水準 α 為時，則其拒絕域為 $\{\chi_0^2 \geq \chi_\alpha^2(k-m-1)\}$ 如下圖1所示，卡方檢定為右尾檢定。

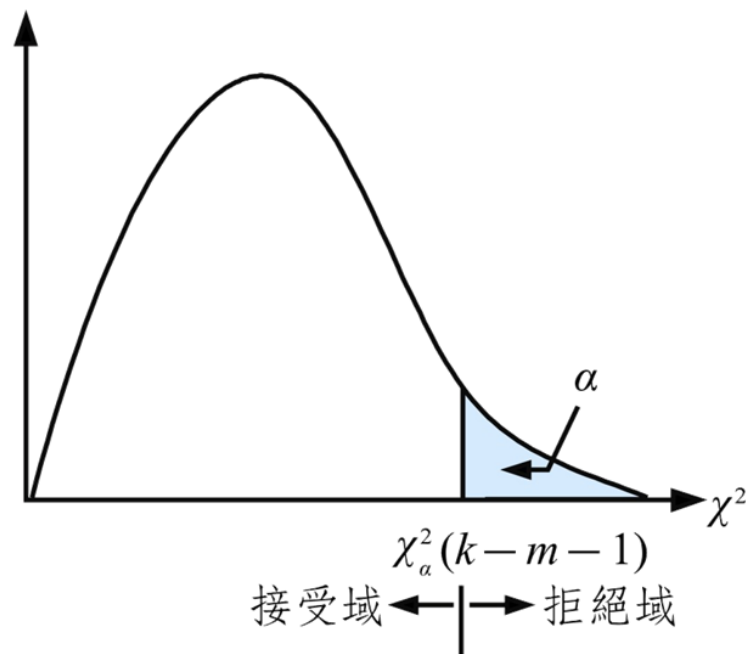


圖 適合度檢定之拒絕域

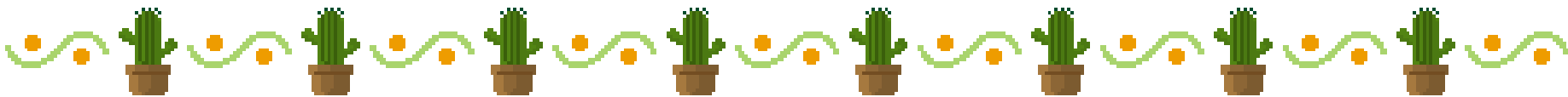
適合度檢定

(二) 決策法則

當 $H_0: p_1 = p_{10}, p_2 = p_{20}, \dots, p_k = p_{k0}$ & $H_1: \text{至少還有一個}$

$p_i \neq p_{i0}$ ，以 $\chi_0^2 = \sum_{i=1}^k \frac{(O_i - np_{i0})^2}{np_{i0}}$ 為檢定統計量，則

其檢定之拒絕域為 $\{\chi_0^2 \geq \chi_\alpha^2(k - m - 1)\}$ ， P 值 = $P(\chi^2 \geq \chi_0^2)$



適合度檢定

例題

- ◎ 某公司想瞭解消費者對其所生產三種不同口味的產品之喜好程度是否一致，於是隨機訪問此產品消費者120人，得其結果如下：

產品類別	1	2	3
喜好人數	35	42	43

請以 $\alpha=0.05$ 檢定該產品消費者對三種產品之喜好程度是否有顯著地差異？

適合度檢定

解

令 p_i 表第 i 種產品受喜好的比例

假設 $H_0: p_1 = p_2 = p_3 = \frac{1}{3}$ & $H_1: p_1, p_2, p_3$ 不完全相等

其拒絕域 $\{\chi_0^2 \geq \chi_{0.05}^2(3-1) = \chi_{0.05}^2(2) = 5.991\}$

在 H_0 成立條件下， $E_1 = E_2 = E_3 = 120 \times \frac{1}{3} = 40$

$$\chi_0^2 = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} = \frac{(35-40)^2 + (42-40)^2 + (43-40)^2}{40} = 0.95 < 5.991$$

不落在拒絕域中，接受 H_0 ，即三種不同口味之產品受消費者喜好程度無顯著地差異。

適合度檢定

例題

◎ 令 X 表同時投擲3個硬幣出現正面之個數，今重複投擲此3個硬幣100次，得其出現正面數的結果如下：

X	0	1	2	3
出現次數	15	38	35	12

請以 $\alpha=0.05$ 來檢定隨機變數 X 是否具二項分配 $b(3,0.5)$ 。

適合度檢定

解

依題意，可建立假設如下：

$H_0 : R.V.X \sim b(3, 0.5) \& X$ 不具有二項分配 $b(3, 0.5)$

若令 p_i 表出現 $i-1$ 個正面之機率，即 $p_i = P(X = i-1) = C_{i-1}^3 (0.5)^3$ ，則上述所建立之假設可改寫為：

$H_0 : p_1 = 0.125, p_2 = 0.375, p_3 = 0.375, p_4 = 0.125 \& H_1 : H_0$ 不成立

由於無估計參數，因此 $m=0$ ，且其卡方檢定之拒絕域為

$$\{\chi_0^2 \geq \chi_{0.05}^2(4-1) = \chi_{0.05}^2(3) = 7.815\}$$

適合度檢定

解

承上頁，在 H_0 成立條件下之期望次數

$$E_0 = 100 \cdot P(X = 0) = 100 \cdot C_0^3 (0.5)^3 = 12.5$$

$$E_1 = 100 \cdot P(X = 1) = 100 \cdot C_1^3 (0.5)^3 = 37.5$$

$$E_2 = 100 \cdot P(X = 2) = 100 \cdot C_2^3 (0.5)^3 = 37.5$$

$$E_3 = 100 \cdot P(X = 3) = 100 \cdot C_3^3 (0.5)^3 = 12.5$$

因此檢定值為

$$\chi^2 = \frac{(15-12.5)^2}{12.5} + \frac{(38-37.5)^2}{37.5} + \frac{(35-37.5)^2}{37.5} + \frac{(12-12.5)^2}{12.5} = 0.693 < 7.815$$

不落在拒絕域中，即勉強接受 $R.V.X \sim b(3, 0.5)$

適合度檢定

例題

◎ 隨機抽取某大學學生100人，得其身高之資料如下：

身高	140-150	150-160	160-170	170-180	180-190
人數	5	25	35	29	6

試以此資料檢定此大學學生身高是否呈現常態分配？ ($\alpha=0.05$)

適合度檢定

解

依題意，可建立假設如下：

H_0 : 學生身高呈現常態分配 & H_1 : 學生身高不具有常態分配
由於其母體平均數與變異數未知，需估計平均數與變異數，因此 $m=2$ ，且其拒絕域為

$$\{\chi_0^2 \geq \chi_{0.05}^2 (5-2-1) = 5.991\}$$

另外，所估計常態分配之參數，

$$\text{平均數 } \bar{x} = \frac{145 \times 5 + 155 \times 25 + 165 \times 35 + 175 \times 29 + 185 \times 6}{100} = 165.6$$

(其中145為140-150之平均數，155為150-160之平均數，依此類推。)

適合度檢定

解

承上頁，

$$\begin{aligned} \text{變異數 } s^2 &= \frac{1}{100-1} [(145-165.6)^2 \times 5 + (155-165.6)^2 \times 25 + (165-165.6)^2 \times 35 + (175-165.6)^2 \times 29 \\ &\quad + (185-165.6)^2 \times 6] = 98.6 \end{aligned}$$

由此可知

$$P(140 < X < 150) = P\left(\frac{140-165.6}{\sqrt{98.6}} < Z < \frac{150-165.6}{\sqrt{98.6}}\right) = P(-2.58 < Z < -1.57) = 0.0533,$$

$$\text{同理 } P(150 < Z < 160) = P(-1.57 < Z < -0.56) = 0.2295,$$

$$P(160 < Z < 170) = P(-0.56 < Z < 0.44) = 0.3523,$$

$$P(170 < Z < 180) = P(0.44 < Z < 1.45) = 0.2565,$$

$$P(180 < Z < 190) = P(1.45 < Z < 2.46) = 0.0666,$$

適合度檢定

解

$$E_1 = 100 \times 0.0533 = 5.33,$$

$$E_2 = 100 \times 0.2295 = 22.95,$$

$$E_3 = 100 \times 0.3823 = 38.23,$$

$$E_4 = 100 \times 0.2565 = 25.65,$$

$$E_5 = 100 \times 0.0666 = 6.66,$$

類別	140-150	150-160	160-170	170-180	180-190
觀察次數	5	25	35	29	6
期望次數	5.33	22.95	38.23	25.65	6.66

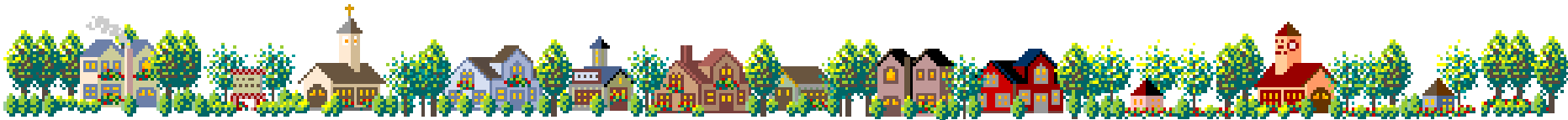
適合度檢定

解

得其檢定值

$$\chi^2 = \frac{(5-5.33)^2}{5.33} + \frac{(25-22.95)^2}{22.95} + \frac{(35-38.23)^2}{38.23} + \frac{(29-25.65)^2}{25.65} + \frac{(6-6.66)^2}{6.66}$$
$$= 0.979 < 5.991$$

不落在拒絕域中，因此可得結論：此大學學生身高呈現常態分配。



適合度檢定-常態分配

- 欲檢定旅館之旅客年齡是否為常態分配，該經理隨機蒐集了 100 個旅客之年齡資料，得到平均旅客年齡為 40 歲，標準差為 15 歲，可設立虛無假設與對立假設為

H_0 ：旅館旅客年齡之分配為常態分配

H_1 ：旅館旅客年齡之分配不為常態分配

- 在計算期望次數的過程中，須先用樣本平均估計常態母體之平均數與用樣本標準差估計常態母體標準差，所以 $m=2$ 。因此在此例中的卡方檢定統計量之自由度為 $k-1-m=3$ 。透過卡方分配表得知臨界值 $\chi_{3,0.05}^2 = 7.81$ 。

適合度檢定-常態分配

- 卡方檢定統計量值可計算如下表所示：

旅客年齡	觀察人數 (f_i)	機率 (p_{i0})	期望人數 ($E_i=100 \times p_{i0}$)	$\frac{(f_i - E_i)^2}{E_i}$
不足 20	5	0.0918	9.18	1.9033
20 - 不足30	21	0.1596	15.96	1.5916
30 - 不足40	30	0.2486	24.86	1.0627
40 - 不足50	25	0.2486	24.86	0.0008
50 - 不足60	15	0.1596	15.96	0.0577
60 以上	4	0.0918	9.18	2.9229
總計	100	1	100	$\chi^2=7.539$

- 卡方檢定統計量值為 7.539 略小於臨界值 7.81，在 0.05 的顯著水準下，我們無法拒絕虛無假設；換言之，我們沒有足夠的證據顯示該旅館旅客之年齡分配不為常態分配。

範例四

- 某研究單位想了解消費者對四種促銷工具的偏好是否有顯著不同？該研究單位回收240份有效問卷，問卷中要求每位受訪者回答最常使用的促銷工具為何？結果整理如下表。

促銷工具	分期付款	直接折扣	折價券	贈品
次數	55	115	24	46

- 建立假說如下：
 - H_0 ：消費者對四種促銷工具的偏好無差異
 - H_1 ：消費者對四種促銷工具的偏好有差異

範例四

- 操作：

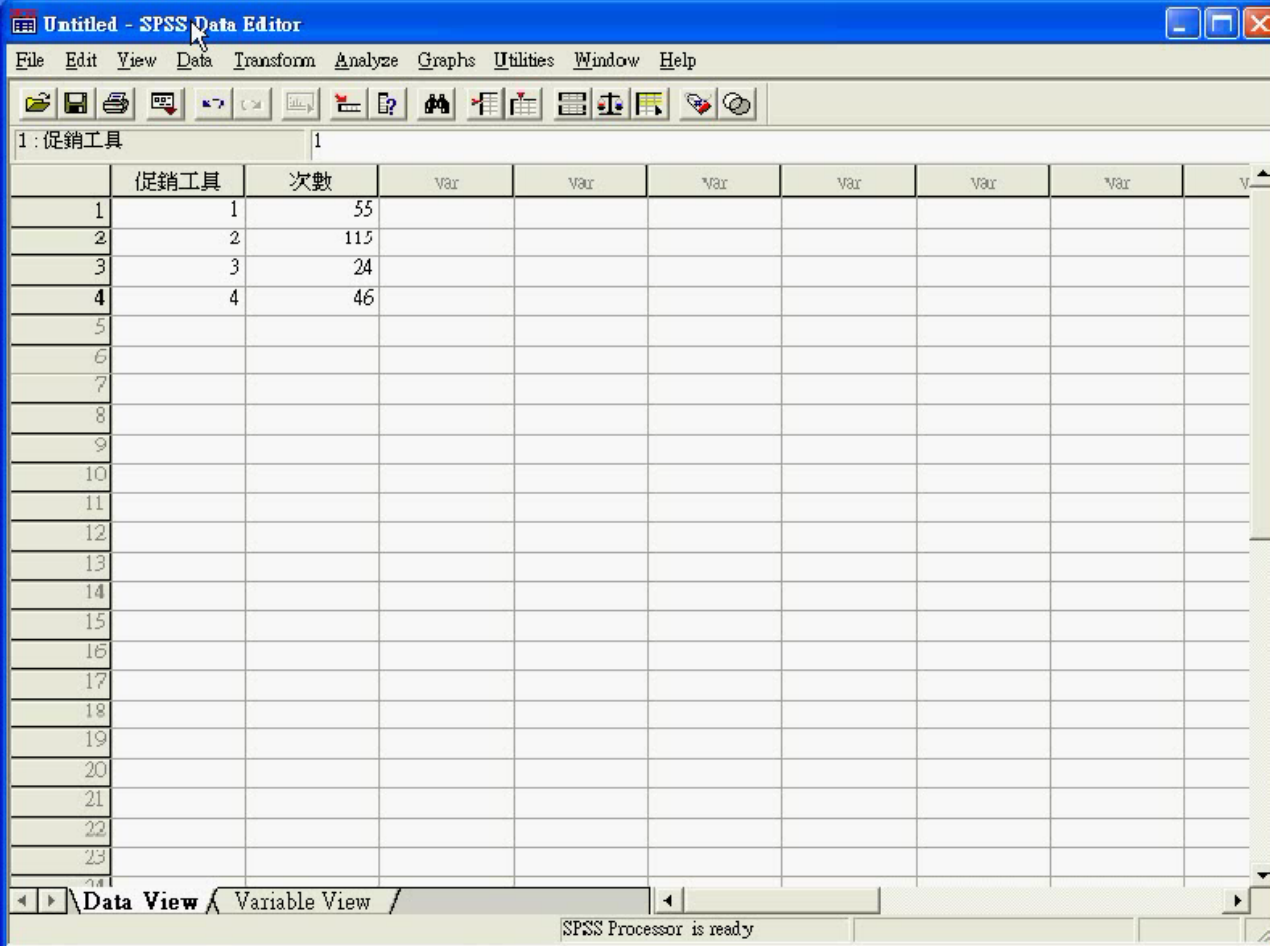
- 1. **Analyze/Nonparametric Tests/Chi-square**

- 2. 程式操作

- 3. 分析結果

範例四—程式操作

2. 程式操作



The screenshot shows the SPSS Data Editor window titled "Untitled - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The data grid is titled "1: 促銷工具" and contains the following data:

	促銷工具	次數	Var	Var	Var	Var	Var	Var	Var
1	1	55							
2	2	115							
3	3	24							
4	4	46							
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									

The status bar at the bottom indicates "SPSS Processor is ready".



範例四—分析結果

促銷工具

	Observed N	Expected N	Residual
分期付款	55	60.0	-5.0
直接折扣	115	60.0	55.0
折價卷	24	60.0	-36.0
贈品	46	60.0	-14.0
Total	240		

Test Statistics

	促銷工具
Chi-Square(a)	75.700
df	3
Asymp. Sig.	.000

Chi-sq = 75.7，**Sig. = 0.000 < 0.05**，拒絕 H_0 。

結論：消費者對四種促銷工具之偏好有顯著差異。



齊一性檢定

(一)齊一性檢定之概念

檢定兩個或兩個以上母體是否具有相同的機率分配或相同的比例。

定理3

在 r 個不同母體， c 個等級之齊一性檢定中，令 O_{ij} 、 E_{ij} 分別表第 i 個母體第 j 個等級之觀察次數與期望次數，則當樣本數 $n \rightarrow \infty$ 時，在 H_0 ：

不同母體具有相同機率分配或相同比例成立條件下，

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \chi^2((r-1)(c-1))$$

齊一性檢定

(二) 決策法則

當 H_0 ：不同母體具有相同機率分配或相同比例

& H_1 ： H_0 不成立，則以 $\chi_0^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

為檢定統計量，其中 $E_{ij} = \frac{\sum_{s=1}^c O_{is} \times \sum_{s=1}^r O_{sj}}{n}$ ， n 為總樣本

數，其檢定之拒絕域為 $\{\chi_0^2 \geq \chi_\alpha^2((r-1)(c-1))\}$ ，

P 值 = $P(\chi^2 \geq \chi_0^2)$ 。

齊一性檢定

例題

- ◎若一研究者想檢定新竹市民與新竹縣民對其地方政府之滿意程度是否相同，於是調查者分別從新竹市抽取600位市民，由新竹縣抽取500位縣民，調查他們對所屬地方政府之滿意程度，得其結果如下表所示(此表為虛擬數據)。

地區	滿意程度					總和
	非常滿意	滿意	普通	不滿意	非常不滿意	
新竹市	152	198	150	58	42	600
新竹縣	98	202	100	42	58	500
總和	250	400	250	100	100	1100

齊一性檢定

解

依題意，可建立假設如下：

H_0 ：新竹市民與新竹縣民對其所屬地方政府之滿意程度相同 &

H_1 ： H_0 不成立

將 H_0 以母體參數表示如下： $(P_{ij}$ 表第 i 個母體第 j 等級之機率值) $H_0 : P_{11} = P_{21}$ 、 $P_{12} = P_{22}$ 、 $P_{13} = P_{23}$ 、 $P_{14} = P_{24}$ 、 $P_{15} = P_{25}$ ，亦即兩母體具有相同之機率分配 ($P_j = P_{1j} = P_{2j}$)，而在此條件下，所估計的比例值 P_j 如下：

$$\hat{P}_1 = \frac{\text{第1行總數}}{\text{樣本總數}} = \frac{250}{1100}$$

齊一性檢定

解

$$\hat{P}_2 = \frac{\text{第2行總數}}{\text{樣本總數}} = \frac{400}{1100}$$

$$\hat{P}_3 = \frac{\text{第3行總數}}{\text{樣本總數}} = \frac{250}{1100}$$

$$\hat{P}_4 = \frac{\text{第4行總數}}{\text{樣本總數}} = \frac{100}{1100}$$

$$\hat{P}_5 = \frac{\text{第5行總數}}{\text{樣本總數}} = \frac{100}{1100}$$

因此各類別資料之期望次數如下：

$$E_{11} = \text{新竹市樣本數} \times \hat{P}_1 = 600 \times \frac{250}{1100} = 136.4$$

齊一性檢定

解

$$E_{12} = \text{新竹市樣本數} \times \hat{P}_2 = 600 \times \frac{400}{1100} = 218.2$$

$$E_{13} = \text{新竹市樣本數} \times \hat{P}_3 = 600 \times \frac{250}{1100} = 136.4$$

$$E_{14} = \text{新竹市樣本數} \times \hat{P}_4 = 600 \times \frac{100}{1100} = 54.4$$

$$E_{15} = \text{新竹市樣本數} \times \hat{P}_5 = 600 \times \frac{100}{1100} = 54.4$$

$$E_{21} = \text{新竹縣樣本數} \times \hat{P}_1 = 500 \times \frac{250}{1100} = 113.6$$

$$E_{22} = \text{新竹縣樣本數} \times \hat{P}_2 = 500 \times \frac{400}{1100} = 181.8$$

齊一性檢定

解

$$E_{23} = \text{新竹縣樣本數} \times \hat{P}_3 = 500 \times \frac{250}{1100} = 113.6$$

$$E_{24} = \text{新竹縣樣本數} \times \hat{P}_4 = 500 \times \frac{100}{1100} = 45.5$$

$$E_{25} = \text{新竹縣樣本數} \times \hat{P}_5 = 500 \times \frac{100}{1100} = 45.5$$

由此可知其觀察次數與期望次數如下表所示。

地區	滿意程度					總和
	非常滿意	滿意	普通	不滿意	非常不滿意	
新竹市	152(136.4)	198(218.2)	150(136.4)	58(54.5)	42(54.5)	600
新竹縣	98(113.6)	202(181.8)	100(113.6)	42(45.5)	58(45.5)	500
總和	250	400	250	100	100	1100

齊一性檢定

解

經計算後，可得其檢定值為

$$\chi_0^2 = \sum_{i=1}^2 \sum_{j=1}^5 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 17.88$$

在顯著水準 H_0 之條件下，其拒絕域為

$$\{\chi_0^2 \geq \chi_{0.05}^2((2-1)(5-1)) = \chi_{0.05}^2(4) = 9.488\}$$

因此檢定值落在拒絕域中，結論為拒絕 H_0 ，即上述資料顯示新竹市民與新竹縣民對其所屬地方政府之滿意程度有顯著地差異。

範例五

- 某高職學校想了解不同的年級其打工經驗是否有差異，於是調查不同年級（高一、高二、高三）共125位同學，並詢問其打工經驗（有、無），調查結果如下表。

	高一	高二	高三
有經驗	27	34	28
無經驗	15	12	9

- 此範例之目的在了解三個不同年級的學生，在打工經驗上的表現是否有差異。
- 建立假說如下：
 - H_0 ：不同年級學生打工經驗無差異
 - H_1 ：不同年級學生打工經驗有差異

範例五—操作

● 操作：

- 1.建資料檔：變數為「經驗」、「年級」、「次數」。
- 2.次數加權：Data/Weight Case/勾選Weight case by「次數」。
- 3.執行交叉列聯表分析：
Analyze/Descriptive/Crosstabs/Row:「經驗」；
Column:「年級」/Stastics勾選Chi-square/Cells 的
Counts勾選Observed及Expected。

範例五—分析結果

經驗 * 年級 Crosstabulation

			年級			Total
			高一	高二	高三	
經驗	有經驗	Count	27	34	28	89
		Expected Count	29.9	32.8	26.3	89.0
	無經驗	Count	15	12	9	36
		Expected Count	12.1	13.2	10.7	36.0
Total		Count	42	46	37	125
		Expected Count	42.0	46.0	37.0	125.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.506(a)	2	.471
Likelihood Ratio	1.481	2	.477
Linear-by-Linear Association	1.277	1	.259
N of Valid Cases	125		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.66.

Chi-sq = 1.506 , Sig. = 0.471 > 0.05 , 無法拒絕 H_0 。

結論：不同年級之打工經驗無顯著差異。



範例六

- 探討不同居住區域的居民，其目前使用手機的品牌是否有差異。建立假說如下：
 - H_0 ：不同居住區域的居民，其目前使用手機品牌之偏好無差異
 - H_1 ：不同居住區域的居民，其目前使用手機品牌之偏好有差異
- 操作：
 - 1.Analyze/Descriptive Statistics/Crosstabs
 - 2.程式操作：改善前、改善後。
 - 3.分析結果：改善前、改善後。

範例六一改善前程式操作

2004手機資料 - SPSS Data Editor

File Edit View Data Transform Analy Graphs Utilities Window Help

1: 編號 1

	編號	性別	所得	年齡	手機品牌	主要用途	居住區域	換購品牌	購買
1	1	1	130	42	1	2	3	2	
2	2	2	150	27	2	1	2	3	
3	3	1	140	45	1	2	2	2	
4	4	2	40	41	2	1	4	3	
5	5	1	80	63	2	1	2	3	
6	6	1	40	26	1	2	4	2	
7	7	1	100	22	1	2	2	3	
8	8	1	110	23	1	2	2	1	
9	9	2	20	32	2	3	3	3	
10	10	2	110	28	2	3	4	3	
11	11	2	100	24	2	1	4	3	
12	12	2	100	21	2	3	3	3	
13	13	2	50	60	3	1	4	4	
14	14	2	110	21	1	1	3	1	
15	15	2	30	20	1	1	2	2	
16	16	2	40	51	3	3	3	4	
17	17	2	60	35	2	3	2	3	
18	18	2	10	16	1	1	2	1	
19	19	2	10	27	2	3	4	3	
20	20	1	30	64	2	2	3	3	
21	21	2	90	44	3	3	4	4	
22	22	1	30	47	2	3	2	3	
23	23	2	110	49	1	1	2	4	
24	24	1	80	18	1	2	2	1	

Data View / Variable View /

SPSS Processor is ready



範例六一改善前分析結果

居住區域 * 目前使用之手機品牌 Crosstabulation

			目前使用之手機品牌			Total
			nokia	易利信	摩托羅拉	
居住區域	台北市	Count	10	0	0	10
		Expected Count	3.4	4.2	2.5	10.0
	北部區域	Count	27	26	0	53
		Expected Count	17.8	22.3	13.0	53.0
	中部區域	Count	26	28	14	68
		Expected Count	22.8	28.6	16.7	68.0
	南部區域	Count	4	30	24	58
		Expected Count	19.4	24.4	14.2	58.0
	東部區域	Count	0	0	11	11
		Expected Count	3.7	4.6	2.7	11.0
Total		Count	67	84	49	200
		Expected Count	67.0	84.0	49.0	200.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	93.371 ^a	8	.000
Likelihood Ratio	109.430	8	.000
Linear-by-Linear Association	71.766	1	.000
N of Valid Cases	200		

a. 6 cells (40.0%) have expected count less than 5. The minimum expected count is 2.45.

因期望次數小於5的方格高達40%，因而必須將某些方格捨棄或合併。在此捨棄台北市與東部區域（因期望次數皆小於5），電腦操作上採用遺漏值。



範例六一改善後程式操作

2004手機資料 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1: 編號 1

	編號	性別	所得	年齡	手機品牌	主要用途	居住區域	換購品牌	購買
1	1	1	130	42	1	2	3	2	
2	2	2	150	27	2	1	2	3	
3	3	1	140	45	1	2	2	2	
4	4	2	40	41	2	1	4	3	
5	5	1	80	63	2	1	2	3	
6	6	1	40	26	1	2	4	2	
7	7	1	100	22	1	2	2	3	
8	8	1	110	23	1	2	2	1	
9	9	2	20	32	2	3	3	3	
10	10	2	110	28	2	3	4	3	
11	11	2	100	24	2	1	4	3	
12	12	2	100	21	2	3	3	3	
13	13	2	50	60	3	1	4	4	
14	14	2	110	21	1	1	3	1	
15	15	2	30	20	1	1	2	2	
16	16	2	40	51	3	3	3	4	
17	17	2	60	35	2	3	2	3	
18	18	2	10	16	1	1	2	1	
19	19	2	10	27	2	3	4	3	
20	20	1	30	64	2	2	3	3	
21	21	2	90	44	3	3	4	4	
22	22	1	30	47	2	3	2	3	
23	23	2	110	49	1	1	2	4	
24	24	1	90	16	1	2	2	1	

Data View / Variable View

SPSS Processor is ready



範例六一改善後分析結果

居住區域 * 目前使用之手機品牌 Crosstabulation

			目前使用之手機品牌			Total
			nokia	易利信	摩托羅拉	
居住區域	北部區域	Count	27	26	0	53
		Expected Count	16.9	24.9	11.3	53.0
	中部區域	Count	26	28	14	68
		Expected Count	21.7	31.9	14.4	68.0
	南部區域	Count	4	30	24	58
		Expected Count	18.5	27.2	12.3	58.0
Total		Count	57	84	38	179
		Expected Count	57.0	84.0	38.0	179.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	41.452 ^a	4	.000
Likelihood Ratio	54.650	4	.000
Linear-by-Linear Association	39.011	1	.000
N of Valid Cases	179		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.25.

Chi-sq = 41.452 , Sig. = 0.000 < 0.05 , 拒絕 H_0 。

結論：不同居住區域的居民，其目前使用手機品牌之偏好有顯著差異。



卡方檢定的類型與範例

- 改變顯著性檢定(test of significance of change)

用以檢定同一樣本在事件前後的反應，是否有顯著的差異。

- 範例七

- 範例八



範例七

- 某大學社團想了解社團的活動方式，是否使得團員對社團生活的喜好態度有所改變，於是在學期初與學期末分別調查，詢問「是否喜歡社團生活」？調查結果如下表：

		期末	
		是	否
期初	是	30	25
	否	110	15

範例七

- 建立假說

- H_0 ：學期初及學期末，團員對社團生活喜好態度無改變

- H_1 ：學期初及學期末，團員對社團生活喜好態度有改變

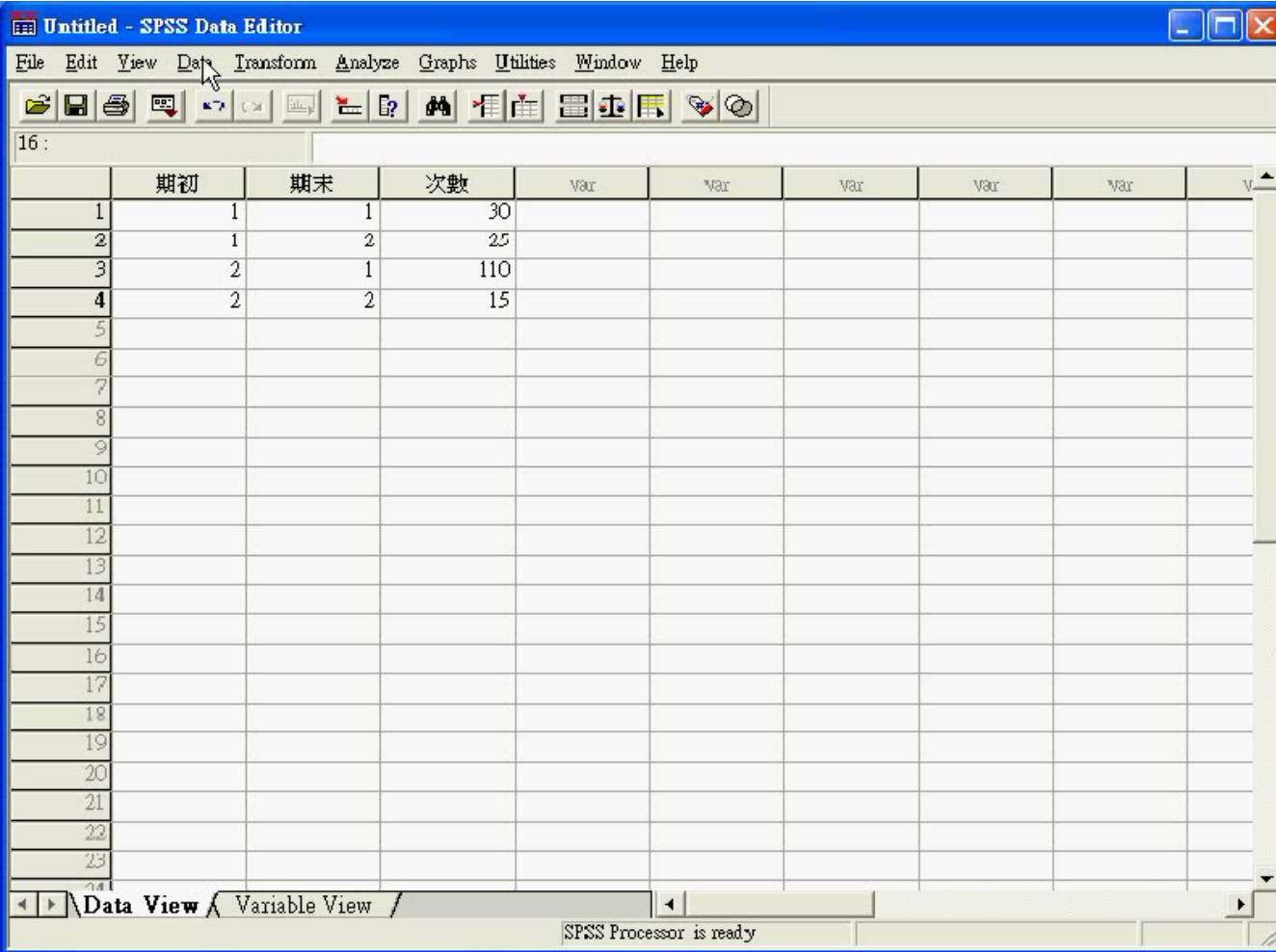
- 操作：

- 1. Analyze/Nonparametric Tests/2 Related Samples

- 2. 程式操作

- 3. 分析結果

範例七—程式操作



The screenshot shows the SPSS Data Editor window titled "Untitled - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The data grid shows a table with the following data:

	期初	期末	次數	var	var	var	var	var	var
1	1	1	30						
2	1	2	25						
3	2	1	110						
4	2	2	15						
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									

The status bar at the bottom indicates "SPSS Processor is ready". The window title bar shows "Data View / Variable View /".



範例七—分析結果

期初 & 期末

期初	期末	
	1	2
1	30	25
2	110	15

Test Statistics^b

	期初 & 期末
N	180
Chi-Square ^a	52.267
Asymp. Sig.	.000

a. Continuity Corrected

b. McNemar Test

Chi-sq = 52.267 , **Sig. = 0.000** < **0.05** , 拒絕 H_0 。

結論：學生在學期初及學期末，對學校生活喜好態度有顯著的改變。



範例八

- 探討手機換購品牌是否有顯著改變。建立假說如下：
 - H_0 ：手機換購品牌無顯著改變
 - H_1 ：手機換購品牌有顯著改變
- 操作：
 - 1. **Analyze/Nonparametric Tests/2 Related Samples**
 - 2. 程式操作
 - 3. 分析結果

範例八一程式操作

2004手機資料 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1: 編號 1

	編號	性別	所得	年齡	手機品牌	主要用途	居住區域	換購品牌	購買
1	1	1	130	42	1	2	3	2	
2	2	2	150	27	2	1	2	3	
3	3	1	140	45	1	2	2	2	
4	4	2	40	41	2	1	4	3	
5	5	1	80	63	2	1	2	3	
6	6	1	40	26	1	2	4	2	
7	7	1	100	22	1	2	2	3	
8	8	1	110	23	1	2	2	1	
9	9	2	20	32	2	3	3	3	
10	10	2	110	28	2	3	4	3	
11	11	2	100	24	2	1	4	3	
12	12	2	100	21	2	3	3	3	
13	13	2	50	60	3	1	4	4	
14	14	2	110	21	1	1	3	1	
15	15	2	30	20	1	1	2	2	
16	16	2	40	51	3	3	3	4	
17	17	2	60	35	2	3	2	3	
18	18	2	10	16	1	1	2	1	
19	19	2	10	27	2	3	4	3	
20	20	1	30	64	2	2	3	3	
21	21	2	90	44	3	3	4	4	
22	22	1	30	47	2	3	2	3	
23	23	2	110	49	1	1	2	4	
24	24	1	80	15	1	2	3	1	

Data View / Variable View

SPSS Processor is ready



範例八—分析結果

NPar Tests

Marginal Homogeneity Test

	目前使用之手機品牌 & 下次欲換購之手機品牌
Distinct Values	4
Off-Diagonal Cases	173
Observed MH Statistic	350.000
Mean MH Statistic	443.000
Std. Deviation of MH Statistic	7.746
Std. MH Statistic	-12.006
Asymp. Sig. (2-tailed)	.000

Sig. = 0.000 < 0.05，拒絕 H_0 。

結論：手機使用者未來在換購手機時，其換購的品牌與目前使用的品牌有顯著的改變。

結論(1/4)

(一)卡方檢定之共通特性

(1)適合度檢定、獨立性檢定及齊一性檢定均利用卡方分配檢定屬性資料(類別資料)的次數是否合乎某一特性。

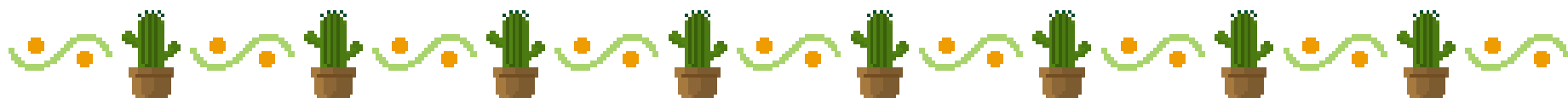
(2)若有任何一組之期望次數 $E_i < 5$ 時，則應將它與其他組合併直到每一組的期望次數均大於或等於5為止。

結論(2/4)

(二)三種檢定之個別特性

(1)適合度檢定用於檢定樣本資料是否來自於一特定之機率分配，其卡方分配之自由度為 $k-m-1$ ，其中 k 為資料之類別數， m 為所欲估計之參數個數。

(2)獨立性檢定用於檢定兩個屬性是否相互獨立，其卡方分配之自由度為 $(r-1) \times (c-1)$ ，其中 r 、 c 分別表兩屬性之等級數。



結論(3/4)

- (3) 齊一性檢定用於檢定兩個或兩個以上不同的母體是否具有相同的分配或相同的比例。其卡方分配之自由度為 $(r-1) \times (c-1)$ ，其中 r 表不同的母體數， c 表資料之類別數。
- (4) 齊一性檢定與獨立性檢定之檢定方式相似，所不同的是收集資料的方式，其中齊一性檢定是依不同的母體各自蒐集所需之樣本數，而獨立性檢定則是由蒐集到的資料來進行分類。

結論(4/4)

(5)在適合度檢定中，若欲檢定母體是否具有常態分配時，在樣本不夠大($n < 100$)時，需作修正，修正後之檢定值如下：

$$\chi^2 = \sum_{i=1}^K \frac{\left(|O_i - E_i| - \frac{1}{2} \right)^2}{E_i}$$

(6)齊一性檢定與獨立性檢定在 $r = c = 2$ 時，在樣本不夠大($n < 100$)時，需作修正，修正後之檢定值如下：

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{\left(|O_{ij} - E_{ij}| - \frac{1}{2} \right)^2}{E_{ij}}$$

(7)在(5)、(6)特性中，當樣本數夠大時，其結果影響不大，可不作修正。

SPSS統計軟體應用-卡方檢定(1/15)

(一) 單一變數類別資料分析：適合度檢定

統計軟體SPSS分析步驟：

步驟一：[資料] / [觀察值加權]。

步驟二：點選[依據...加權觀察值]，在次數變數中輸入次數或人數。

步驟三：[確定]

步驟四：[分析] / [無母數檢定] / [卡方分配]。

SPSS統計軟體應用-卡方檢定(2/15)

統計軟體SPSS分析步驟：

步驟五：選入欲分析之變數(非次數變數)

步驟六：在期望值中輸入欲檢定類別資料的比例。若比例相等，則選擇全部類別相等即可。

步驟七：[**確定**]。

SPSS統計軟體應用-卡方檢定(3/15)

例題

- ◎ 某飲料公司想調查消費者對其所生產之不同口味的冰淇淋之喜好是否有顯著地差異，於是隨機調查300位消費者，由消費者選擇一種最喜歡之口味，得其結果如下：

口味	檸檬	芒果	香草	巧克力	草莓
人數	50	70	45	60	75

請以顯著水準 $\alpha = 0.05$ 檢定消費者對其所生產之不同口味的冰淇淋之喜好是否有顯著差異？

SPSS統計軟體應用-卡方檢定(4/15)

解

假設 H_0 : 消費者對不同口味飲料之喜好沒有差異

將資料輸入如表下表所示部分，分別以1、2、3、4、5表檸檬、芒果、香草、巧克力及草莓五種口味，然後依單一類別資料SPSS分析步驟選擇『全部類別相等』進行分析，則可得其結果。

	口味	人數	Var	Var
1	1	50		
2	2	70		
3	3	45		
4	4	60		
5	5	75		
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

SPSS統計軟體應用-卡方檢定(5/15)

解

檢定統計量

	口味
卡方 ^a	10.833
自由度	4
漸近顯著性	.029

a. 0個格(.0%)的期望次數少於5
。最小的期望格次數為60.0。

由上表結果得知卡方值 $\chi_0^2 = 10.833$ ，顯著性 P 值 = 0.029 小於 α ，因此拒絕 H_0 ，消費者對於不同口味冰淇淋之喜好有顯著地差異。

SPSS統計軟體應用-卡方檢定(6/15)

(二) 兩變數類別資料分析：

(1) 齊一性檢定

統計軟體SPSS分析步驟：

步驟一：[資料] / [觀察值加權]。

步驟二：點選[依據...加權觀察值]，在次數變數中輸入次數或人數。

步驟三：[確定]

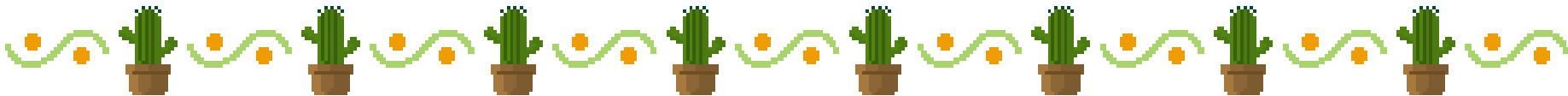
步驟四：[分析] / [描述統計] / [交叉表]，在列與直行中分別輸入欲分析之變數。

SPSS統計軟體應用-卡方檢定(7/15)

統計軟體SPSS分析步驟：

步驟五：在[**統計量**]中選擇卡方統計量。

步驟六： [**繼續**] / [**確定**]。



SPSS統計軟體應用-卡方檢定(8/15)

例題

◎ 假設隨機由甲、乙、丙、丁四個單位中各抽取100位員工作調查，得其對工作滿意度如下，請以 $\alpha = 0.05$ 檢定此四個單位員工之工作滿意度是否有顯著地差異？

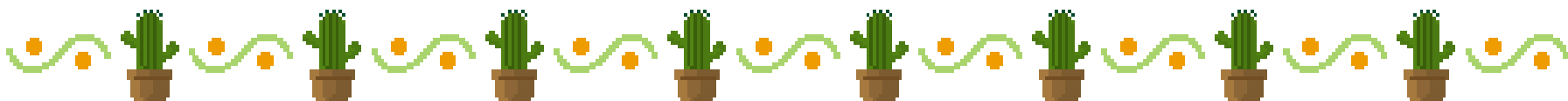
	好	普通	不好
甲	59	32	9
乙	48	44	8
丙	64	26	10
丁	54	42	4

SPSS統計軟體應用-卡方檢定(9/15)

解

假設 H_0 : 甲、乙、丙、丁四個單位員工滿意度相同

將資料輸入如表下表所示部分，分別以1、2、3、4在「單位」變數中分別註解甲、乙、丙、丁；亦以1、2、3在「意見」變數中分別註解好、普通、不好。然後依齊一性檢定分析步驟在「次數變數」中輸入人數，在列與直行中分別輸入「單位」與「意見」進行分析，則可得其結果。



SPSS統計軟體應用-卡方檢定(10/15)

解

承上頁：

	單位	意見	人數	Var
1	1.00	1.00	59.00	
2	1.00	2.00	32.00	
3	1.00	3.00	9.00	
4	2.00	1.00	48.00	
5	2.00	2.00	44.00	
6	2.00	3.00	8.00	
7	3.00	1.00	64.00	
8	3.00	2.00	26.00	
9	3.00	3.00	10.00	
10	4.00	1.00	54.00	
11	4.00	2.00	42.00	
12	4.00	3.00	4.00	
13				
14				
15				

SPSS統計軟體應用-卡方檢定(11/15)

解

單位 * 意見 交叉表

個數		意見			總和
		好	普通	不好	
單位	甲	59	32	9	100
	乙	48	44	8	100
	丙	64	26	10	100
	丁	54	42	4	100
總和		225	144	31	400

卡方檢定

	數值	自由度	漸近顯著性(雙尾)
Pearson卡方	11.180 ^a	6	.083
概似比	11.673	6	.070
線性對線性的關連	.242	1	.623
有效觀察值的個數	400		

a. 0格(.0%)的預期個數少於5。最小的預期個數為7.75。

上表卡方檢定中Pearson卡方值為11.180，顯著性 P 值=0.083大於 α 值，因此勉強接受 H_0 ，即甲、乙、丙、丁四個單位員工之工作滿意度無顯著地差異。

SPSS統計軟體應用-卡方檢定(12/15)

(2)獨立性檢定(離散型變數之相關性)：輸入方式與分析步
步均與齊一性檢定相同。

例題

◎隨機抽取36消費者作調查，得其性別與滿意度資料如下：

(女,滿意), (男,不滿意), (男,滿意), (女,滿意), (女,滿意), (男,滿意),
(女,滿意), (男,不滿意), (男,不滿意), (女,滿意), (男,滿意), (女,不滿意),
(女,滿意), (男,滿意), (男,不滿意), (女,不滿意), (男,滿意), (男,滿意),
(男,滿意), (女,滿意), (男,不滿意), (女,不滿意), (男,滿意), (女,滿意),
(男,滿意), (女,滿意), (男,不滿意), (男,不滿意), (女,不滿意),
(女,不滿意), (女,不滿意), (女,不滿意), (女,不滿意), (男,滿意), (男,滿意),
(男,滿意)。

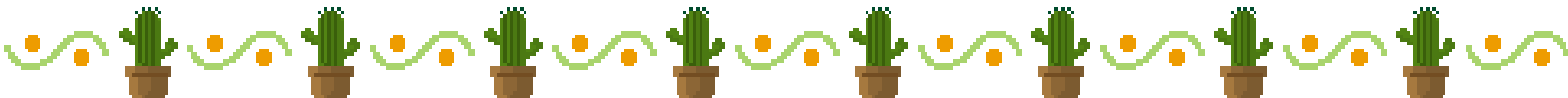
請以交叉表呈現此資料，並以卡方檢定檢定不同性別之民眾之滿意度是否有顯著地差異？($\alpha = 0.05$)

SPSS統計軟體應用-卡方檢定(13/15)

解

假設 H_0 : 男性與女性滿意度相同。

將資料輸入如表下表所示部分，分別以1、2在「性別」變數中分別註解男性、女性；亦以1、2在「滿意度」變數中分別註解滿意、不滿意。由於此資料是未進行整理之資料，不需輸入次數變數，所以直接由第四步驟開始，在列與直行中分別輸入「性別」與「滿意度」進行分析，則可得其結果。



SPSS統計軟體應用-卡方檢定(14/15)

解

承上頁：

	性別	滿意度	Var	Var
1	2.00	1.00		
2	1.00	2.00		
3	1.00	1.00		
4	2.00	1.00		
5	2.00	1.00		
6	1.00	1.00		
7	2.00	1.00		
8	1.00	2.00		
9	1.00	2.00		
10	2.00	1.00		
11	1.00	1.00		
12	2.00	2.00		
13	2.00	1.00		
14	1.00	1.00		
15	1.00	2.00		

SPSS統計軟體應用-卡方檢定(15/15)

解

性別 * 滿意度 交叉表

個數		滿意度		總和
		滿意	不滿意	
性別	男性	12	7	19
	女性	9	8	17
總和		21	15	36

卡方檢定

	數值	自由度	漸近顯著性(雙尾)	精確顯著性(雙尾)	精確顯著性(單尾)
Pearson卡方	.385 ^b	1	.535		
連續性校正 ^a	.080	1	.778		
概似比	.386	1	.535		
Fisher's精確檢定				.736	.389
線性對線性的關連	.375	1	.541		
有效觀察值的個數	36				

a. 只能計算 2x2 表格

b. 0格(.0%)的預期個數少於5。最小的預期個數為7.08。

由上表結果得知卡方值 $\chi_0^2 = 0.385$ ，顯著性P值=0.535大於 α 值，因此接受 H_0 ，即男性與女性之滿意度無顯著地差異，或稱性別與滿意度無顯著地相關性。