

第17章 類別資料的分析 應用統計學 四版

17 類別資料的分析

學習目的

1. 了解卡方分配的意義與性質。
2. 了解當碰到類別資料時，應如何來做統計檢定的工作。
3. 了解配合度檢定的意義與檢定方法。
4. 了解獨立性檢定的意義與檢定方法。
5. 了解齊一性檢定的意義與方法。
6. 利用Excel 來做卡方檢定。

林島玲 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

本章結構

```

graph TD
    A[卡方檢定] --> B[類別資料的整理]
    A --> C[配合度檢定]
    A --> D[獨立性檢定]
    A --> E[齊一性檢定]
    A --> F[Excel的使用]
    C --> G[配合度檢定的基本概念]
    C --> H[配合度檢定的方法]
    
```

林島玲 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

卡方檢定

○卡方檢定

卡方檢定主要用來檢定名目尺度資料與順序尺度資料。
卡方分配可做三種檢定：配合度檢定，獨立性檢定與齊一性檢定。

林島玲 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

類別資料的整理

表17.1 經濟發展看法的調查

兩岸關係的看法	人數
維持現狀	1,100
獨立	70
統一	30

林島玲 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

類別資料的整理

表17.2 列聯表-性別與喜好車型

性別\車型	中型車	小型車	迷你車
男	30	25	25
女	15	20	35

林島玲 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

○配合度檢定的意義

配合度檢定是利用樣本資料檢定母體分配是否為某一特定分配或理論分配的統計方法。

○多項實驗

- ① 包含 n 次相同的試行。
- ② 每一次試行結果為 k 個可能結果中的一個 ($k > 2$)。
- ③ 每一次試行是獨立的。
- ④ 每一次試行各個可能結果的機率是固定的。即若每一結果的機率為 $p_i (i=1, \dots, k)$ ，則每一次試行結果的機率均相同且 $p_1 + p_2 + \dots + p_k = 1$ 。

林島玲 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

○ 觀察次數與期望次數

抽樣實驗中得到的次數稱為觀察次數，表為 O_i ；期望次數是如果虛無假設為真時，預期會發生的次數，表為 E_i 。各類別的期望次數為：

$$E_i = np$$

n 為樣本數， p 為虛無假設為真時，各類別的比例或機率。

林惠玲 陳正倉著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

○ 配合度檢定統計量

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi_{k-1-m}^2$$

式中： O_i ：第 i 組的樣本觀察次數， E_i ：期望次數(或理論次數)， k ：組數。自由度為 $k-1-m$ ， m 為估計的參數個數。

○ 決策法則

① 若 $\chi^2 > \chi_{k-1-m, \alpha}^2$ ，則拒絕 H_0 。

② 若 $\chi^2 \leq \chi_{k-1-m, \alpha}^2$ ，則接受 H_0 。

林惠玲 陳正倉著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

母體變異數的區間估計

○ 卡方分配的性質

① 卡方分配為一定義在大於等於0(正數)範圍的右偏分配，不同的自由度決定不同的卡方分配。

② 卡方分配只有一個參數即自由度，表為 ν 。卡方分配的平均數與變異數為：

$$E(\chi^2) = \nu, V(\chi^2) = 2\nu$$

③ 卡方分配當自由度增加而逐漸對稱，當自由度趨近於無窮大時($\nu \rightarrow \infty$)，卡方分配會趨近於常態分配。

④ 設 $X \sim N(\mu, \sigma^2)$ ，令 $Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$ ，則 Z^2 為自由度1的卡方分配。

林惠玲 陳正倉著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

母體變異數的區間估計

圖17.1 卡方檢定的拒絕域與接受域

林惠玲 陳正倉著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

表17.3 旅客對高雄國際航空站的滿意度

旅客對高雄國際航空站滿意度	人數
非常滿意	157
滿意	463
尚可	118
不滿意	20
非常不滿意	43

林惠玲 陳正倉著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

圖17.2 旅客對高雄航空站滿意度的配合度檢定

林惠玲 陳正倉著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

表17.4 旅客滿意程度的觀察次數與期望次數

滿意度	O_i	機率 p_i	$E = np_i$	$(O - E)$	$(O - E)^2$	$\frac{(O - E)^2}{E}$
非常滿意	157	0.36	$801 \times 0.36 = 288$	-131	$(-131)^2 = 17,161$	59.51
滿意	463	0.38	$801 \times 0.38 = 304$	159	$(131)^2 = 25,281$	83.06
尚可	118	0.19	$801 \times 0.19 = 152$	-34	$(-34)^2 = 1,156$	7.6
不滿意	20	0.03	$801 \times 0.03 = 24$	-4	$(-4)^2 = 16$	0.67
非常不滿意	43	0.04	$801 \times 0.04 = 32$	11	$(11)^2 = 121$	3.78
						154.61

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

表17.5 旅客滿意程度的觀察次數與理論次數

	A	B	C	D	E	F	G	H	I	J	K
1	觀察次數					理論次數					
2	非常滿意	滿意	尚可	不滿意	非常不滿意	非常滿意	滿意	尚可	不滿意	非常不滿意	
3	157	463	118	20	43	288	304	152	24	32	

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

圖 17.3 χ^2 檢定對話方塊圖

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

表17.6 大學指考國文成績的分配

分數	0-10	10-20	20-30	30-40	40-50
人數	5	38	133	245	290
分數	50-60	60-70	70-80	80-90	90-100
人數	227	114	28	2	0

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

配合度檢定

表17.7 大學指考國文成績的觀察次數與理論次數

國文分數	人數 (O_i)	$Z = \frac{U - 44.07}{14.14}$	$Z = \frac{L - 44.07}{14.14}$	機率值 (p_i)	$E(np_i)$	$(O - E)^2$
0-10	5	-2.41	-∞	0.008	8.7	13.69
10-20	39	-1.70	-2.41	0.037	40.1	1.21
20-30	133	-1.00	-1.70	0.114	123.5	90.25
30-40	245	-0.29	-1.00	0.227	245.8	0.64
40-50	290	0.42	-0.29	0.270	300.0	5.76
50-60	227	1.13	0.42	0.208	225.3	2.89
60-70	114	1.83	1.13	0.096	104.0	100
70-80	28	2.54	1.83	0.028	30.3	5.29
80-90	2	3.25	2.54	0.005	5.4	16.00
90-100	0	∞	3.25	0.0006	0.6	

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

獨立性檢定

○ 獨立性檢定意義
獨立性檢定是用來檢定兩個屬性間是否獨立的統計方法。

○ 獨立性檢定的檢定統計量

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

其中： r ：列聯表中橫列的個數， c ：縱行的個數， O_{ij} ：樣本觀察次數， \hat{E}_{ij} ：估計的理論次數(下面的例子會說明其計算方法)。自由度為 $(r-1)(c-1)$ 。

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

獨立性檢定

○ 決策法則

① 若 $\chi^2 > \chi_{(c-1)(r-1), \alpha}^2$ ，則拒絕 H_0 。

② 若 $\chi^2 \leq \chi_{(c-1)(r-1), \alpha}^2$ ，則接受 H_0 。

林島玲 陳玉金著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

獨立性檢定

表17.8 城市與鄉村小學生近視與否

	城市	鄉村	合計
有近視	144	28	172
沒有近視	36	72	108
合計	180	100	280

林島玲 陳玉金著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

獨立性檢定

表17.9 城市與鄉村小學生近視與否

	城市	鄉村	合計
有近視	144 (111)	28 (61)	172
沒有近視	36 (69)	72 (39)	108
合計	180	100	280

林島玲 陳玉金著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

獨立性檢定

表17.10 廣告態度與年齡層

年齡	喜歡廣告	厭惡廣告	列總和
18歲以下	6	16	22
19-20歲	26	61	87
21-22歲	33	42	75
23-25歲	9	6	15
25歲以上	3	2	5
行總和	77	127	204

林島玲 陳玉金著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

獨立性檢定

表17.12 廣告態度與年齡層的理论次數與觀察次數

	喜歡廣告	厭惡廣告	列總和
18歲以下	6 (8.30)	16 (13.70)	22
19-20歲	26 (23.84)	61 (54.16)	87
21-22歲	33 (28.31)	42 (46.69)	75
23歲以上	12 (7.55)	8 (12.45)	20
行總和	77	127	204

林島玲 陳玉金著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

齊一性檢定

○ 齊一性檢定意義

齊一性檢定是檢定兩個或兩個以上母體的某一特性的分配(各類別的比例)是否齊一或相近。

○ 齊一性檢定統計量

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

其中： r ：橫列個數， c ：縱行個數， O_{ij} ：樣本觀察次數， \hat{E}_{ij} ：估計期望論次數，自由度為 $(r-1)(c-1)$ 。

○ 決策法則

① 若 $\chi^2 > \chi_{(c-1)(r-1), \alpha}^2$ ，則拒絕 H_0 。

② 若 $\chi^2 \leq \chi_{(c-1)(r-1), \alpha}^2$ ，則接受 H_0 。

林島玲 陳玉金著 雙葉書廊發行 2009

第17章 類別資料的分析 應用統計學 四版

齊一性檢定

表17.11 外語測驗成績人數分佈

分數	第一次	第二次	合計人數
80-100	10	7	17
60-79	21	23	44
40-59	11	9	20
0-39	4	4	8
合計人數	46	43	89

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

齊一性檢定

表17.13 外語能力考試成績分配

分數	第一次	第二次	合計人數
80-100	10 (8.79)	7 (8.21)	17
60-79	21 (22.74)	23 (21.26)	44
40-59	11 (10.34)	9 (9.66)	28
0-39	4 (4.13)	4 (3.87)	8
合計人數	$n_1 = 46$	$n_2 = 43$	89

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

齊一性檢定

表17.14 大學聯考成績的分配

大學聯考結果	高中入學方式		小計
	聯考生	非聯考生	
錄取	31	24	55
未錄取	19	17	36
小計	50	41	91

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

齊一性檢定

表17.15 大學聯考成績的次數分配

大學聯考結果	高中入學方式		小計
	聯考生	非聯考生	
錄取	31 (30.2)	24 (24.8)	55
未錄取	19 (19.8)	17 (16.2)	36
小計	50	41	91

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

EXCEL 的使用

表17.16 選舉人名單

	A	B	C	D	E	F
1	編號	姓名	性別	年齡	教育程度	政黨傾向
2	1	黃文章	男	33	大學	K
3	2	李麗娟	女	56	高中	M
4	3	陳文玲	女	36	研究所	K
5	4	徐小翠	女	62	國中	M
6	5	胡光雄	男	57	高中	K
7	6	朱天文	男	26	大學	K
8	7	鍾美紅	女	59	國中	M

林島峰 陳正倉著 雙葉書庫發行 2009

第17章 類別資料的分析 應用統計學 四版

EXCEL 的使用

圖17.4 資料來源

林島峰 陳正倉著 雙葉書庫發行 2009

EXCEL 的使用

表17.17 教育程度與政黨取向

計數 - 教育欄標籤	K	M	總計
大學	2		2
研究所	1		1
高中	1	1	2
國中		2	2
總計	4	3	7

林惠坤 陳正倉著 雙葉書廊發行 2009

EXCEL 的使用

圖17.5 分配資料於適當欄位



林惠坤 陳正倉著 雙葉書廊發行 2009