

第 16 章

卡方檢定

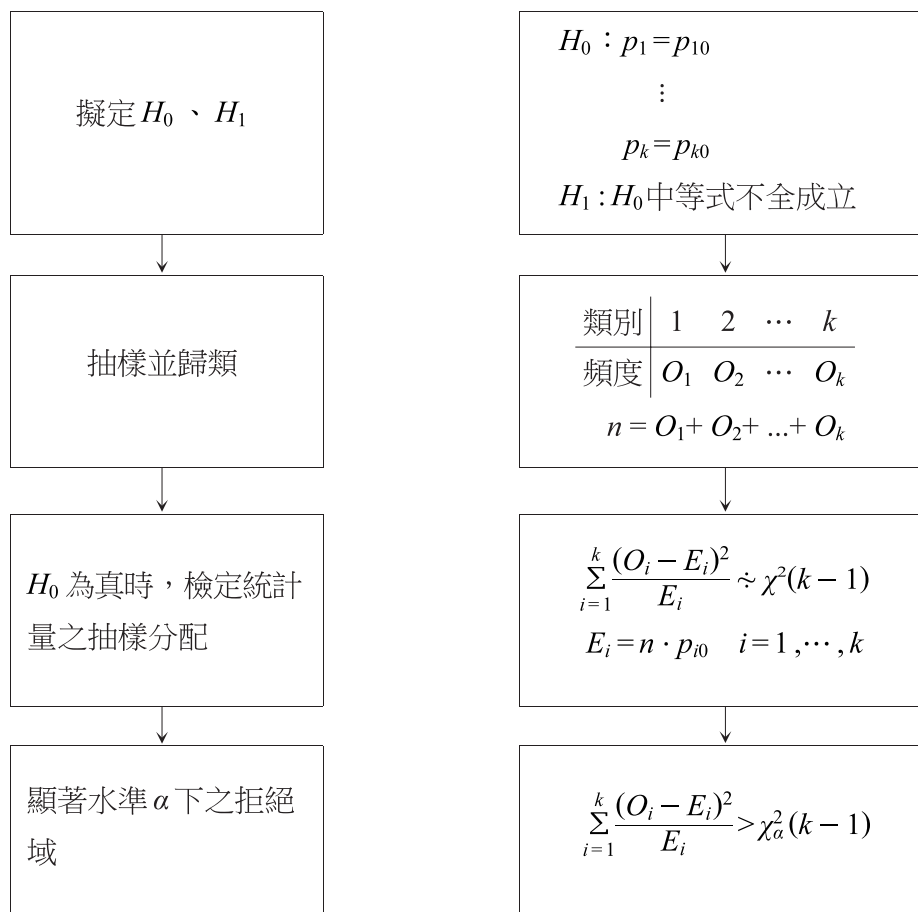


圖 16-1-1 多項母群體參數結構之檢定程序

注意事項：本章各節的卡方檢定中，各類別(或分類格子)中的期望值不可小於 5，當某一類別出現此現象時解決之道有二 (1)將該類別歸併至鄰近類別 (2) 增加抽樣樣本數。

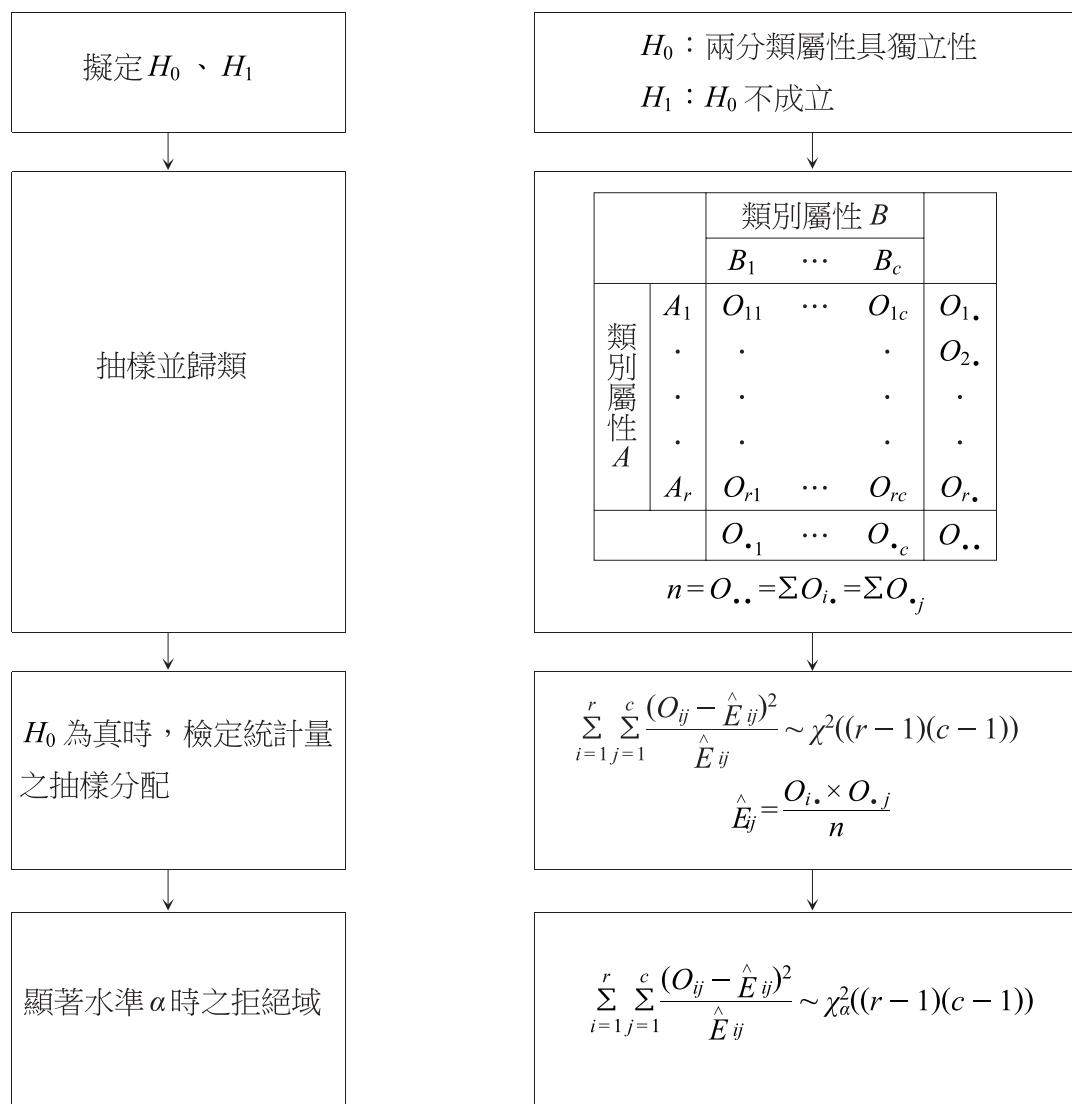


圖 16-2-1 兩隨機變數獨立之檢定程序

表 16-2-1 隨機樣本列聯表

		類別屬性 B				
		B_1	B_2	$\cdots$	B_c	
類別屬性 A	A_1	O_{11}	O_{12}	$\cdots$	O_{1c}	$O_{1\cdot}$
	A_2	O_{21}	O_{22}	$\cdots$	O_{2c}	$O_{2\cdot}$
	$\cdot$			$\cdot$		$\cdot$
	$\cdot$			$\cdot$		$\cdot$
	$\cdot$			$\cdot$		$\cdot$
	A_r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	$O_{r\cdot}$
		$O_{\cdot 1}$	$O_{\cdot 2}$	$\cdots$	$O_{\cdot c}$	$O_{\cdot \cdot}$

332 • 統計學習題解答

16.1 抽樣調查 600 位同學體育選修科目人數分佈如表 E-16-1-1

表 E-16-1-1

籃球	排球	高爾夫球	游泳	其他
155	56	148	152	89

在顯著水準 $\alpha=0.05$ 下，檢定大專同學選修體育的比例結構是否如 H_0 所描述。其中 p_1 表示選修籃球的比率， p_2 表示選修排球的比率， p_3 表示選修高爾夫球的比率， p_4 表示選修游泳的比率， p_5 表示選其他科目的比率。

$$H_0 : p_1 = 0.25$$

$$p_2 = 0.1$$

$$p_3 = 0.25$$

$$p_4 = 0.25$$

$$p_5 = 0.15$$

$$H_1 : H_0 \text{ 不成立}$$

解：

卡方檢定的試算表如表 E-16-1-2 所示。

表 E-16-1-2

	籃球	排球	高爾夫球	游泳	其他	
p_i	0.25	0.1	0.25	0.25	0.15	
O_i	155	56	148	152	89	600
E_i	150	60	150	150	90	
$O_i - E_i$	5	-4	-2	2	-1	
$\frac{(O_i - E_i)^2}{E_i}$	0.166667	0.266667	0.026667	0.026667	0.011111	0.497778

$\alpha=0.05$ ， $\chi^2_{0.05}(4)=9.49$ 。所以檢定的拒絕域為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \middle| \sum \frac{(O_i - E_i)^2}{E_i} > 9.49 \right\}$$

由試算表中得知檢定統計值為 $0.498 < 9.49$ ，所以檢定的結果為不拒絕 H_0 ，這表示大專生選修體育科目的結構並未明顯偏離 H_0 所描述的結構。

16.2 抽樣調查 98 位大專生統計學成績分佈如表 E-16-2-1 所示。

表 E-16-2-1

80 分以上	70-80	60-70	50-60	50 以下
18	40	30	6	4

在顯著水準 $\alpha=0.25$ 下，檢定大專生統計成績比例分佈是否如 H_0 所描述。其中 p_1 表示 80 分以上的比例， p_2 表示 70 分到 80 分之間的比例， p_3 表示 60 分到 70 分間的比例， p_4 表示 50 分到 60 分間的比例， p_5 表示 50 分以下的比例。

$$H_0 : p_1 = 0.2$$

$$p_2 = 0.4$$

$$p_3 = 0.3$$

$$p_4 = 0.06$$

$$p_5 = 0.04$$

$$H_1 : H_0 \text{ 不成立}$$

解：

卡方檢定的試算表如表 E-16-2-2 所示。

表 E-16-2-2

	80 分以上	70-80	60-70	50-60	50 以下	
p_i	0.2	0.4	0.3	0.06	0.04	
O_i	18	40	30	6	4	98
E_i	19.6	39.2	29.4	5.88	3.92	
$O_i - E_i$	-1.6	0.8	0.6	0.12	0.08	
$\frac{(O_i - E_i)^2}{E_i}$	0.1306	0.0163	0.0122	0.0024	0.0016	0.1633

試算表 E-16-2-2 中，50 分以下的組期望值為 $3.92 < 5$ ，所以必需予以併組，我們將它歸併到 50-60 組，重新擬定 H_0 及 H_1 並重新試算如表 E-16-2-3 所示。

$$H_0 : p_1 = 0.2$$

$$p_2 = 0.4$$

$$p_3 = 0.3$$

$$p_4 = 0.1$$

$$H_1 : H_0 \text{ 不成立}$$

表 E-16-2-3

	80 分以上	70-80	60-70	60 以下	
p_i	0.2	0.4	0.3	0.1	
O_i	18	40	30	10	98
E_i	19.6	39.2	29.4	9.8	98
$O_i - E_i$	-1.6	0.8	0.6	0.2	
$\frac{(O_i - E_i)^2}{E_i}$	0.1306	0.0163	0.0122	0.004	0.1631

$\alpha = 0.025$ ， $\chi_{0.025}^2(3) = 9.35$ 。所以檢定的拒絕域為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 9.35 \right\}$$

由試算表 E-16-2-3 中得知檢定值為 $0.1631 < 9.35$ ，所以檢定的結果為不拒絕 H_0 ，這表示大專生統計成績分佈的結構並未明顯偏離 H_0 所描述的比例結構。

16.3 抽樣調查 95 位高科技產業員工教育程度的分佈比例如表 E-16-3-1 所示。

表 E-16-3-1

國中	高中	大專	碩士	博士
4	12	59	12	8

在顯著水準 $\alpha=0.05$ 下，檢定高科技產業員工教育程度分佈比例是否如 H_0 所描述。其中 p_1 表示國中所佔的比例， p_2 表示高中所佔的比例， p_3 表示大專所佔的比例， p_4 表示碩士所佔的比例， p_5 表示博士所佔的比例。

$$H_0 : p_1 = 0.06$$

$$p_2 = 0.13$$

$$p_3 = 0.6$$

$$p_4 = 0.15$$

$$p_5 = 0.06$$

$$H_1 : H_0 \text{ 不成立}$$

解：

卡方檢定的試算表如表 E-16-3-2 所示。

表 E-16-3-2

	國中	高中	大專	碩士	博士	
p_i	0.06	0.13	0.6	0.15	0.06	
O_i	4	12	59	12	8	95
E_i	5.7	12.35	57	14.25	5.7	
$O_i - E_i$	-1.7	-0.35	2	-2.25	2.3	
$\frac{(O_i - E_i)^2}{E_i}$	0.5070	0.0099	0.0702	0.3553	0.9281	1.8704

$\alpha = 0.05$ ， $\chi^2_{0.05}(4) = 9.49$ 。所以檢定的拒絕域為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 9.49 \right\}$$

由試算表 E-16-3-2 中，得知檢定統計值為 $1.8704 < 9.49$ ，所以檢定的結果為不拒絕 H_0 ，這表示高科技產業員工教育程度比例分配並未明顯偏離 H_0 所描述。

16.4 抽樣調查 1062 位都會區民眾的年齡結構比例如表 E-16-4-1 所示。

表 E-16-4-1

18 歲以下	18-30 歲	30-45 歲	45-60 歲	60 歲以上
140	280	400	152	90

在顯著水準 $\alpha = 0.025$ 下，檢定都會區居民年齡分佈是否如 H_0 所描述。其中 p_1 表示 18 歲以下所佔的比例， p_2 表示 18-30 歲所佔的比例， p_3 表示 30-45 歲所佔的比例， p_4 表示 45-60 歲所佔的比例， p_5 表示 60 歲以上所佔的比例。

$$H_0 : p_1 = 0.16$$

$$p_2 = 0.25$$

$$p_3 = 0.35$$

$$p_4 = 0.15$$

$$p_5 = 0.09$$

$H_1 : H_0$ 不成立

解：

卡方檢定的試算表如表 E-16-4-2 所示。

表 E-16-4-2

	18 歲以下	18-30 歲	30-45 歲	45-60 歲	60 歲以上	
p_i	0.16	0.25	0.35	0.15	0.09	
O_i	140	280	400	152	90	1062
E_i	169.92	265.5	371.7	159.3	95.58	
$O_i - E_i$	-29.92	14.5	28.3	-7.3	-5.58	
$\frac{(O_i - E_i)^2}{E_i}$	5.2684	0.7919	2.1547	0.3345	0.3258	8.8753

$\alpha = 0.025$ ， $\chi_{0.025}^2(4) = 11.14$ 。所以檢定的拒絕域為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 11.14 \right\}$$

由試算表 E-16-4-2 中，得知檢定值為 $8.8753 < 11.14$ 。所以檢定結果為不拒絕 H_0 ，這表示都會區民眾的年齡結構並未明顯偏離 H_0 所描述。

16.5 抽樣調查 800 位某地區居民的宗教信仰分佈如表 E-16-5-1 所示。

表 E-16-5-1

基督教	佛教	其他
100	427	273

338 • 統計學習題解答

在顯著水準 $\alpha=0.05$ 下，檢定該地區居民宗教信仰比例結構是否如 H_0 所描述。其中 p_1 表示基督教徒所佔的比例， p_2 表示佛教徒所佔的比例， p_3 表示其他宗教所佔的比例。

$$H_0 : p_1=0.12$$

$$p_2=0.54$$

$$p_3=0.34$$

$$H_1 : H_0 \text{ 不成立}$$

解：

卡方檢定的試算表如表 E-16-5-2 所示。

表 E-16-5-2

	基督教	佛教	其他	
p_i	0.12	0.54	0.34	
O_i	100	427	273	800
E_i	96	432	272	800
$O_i - E_i$	4	-5	1	
$\frac{(O_i - E_i)^2}{E_i}$	0.166667	0.05787	0.003676	0.228214

$\alpha=0.05$ ， $\chi^2_{0.05}(2)=5.99$ 。所以檢定的拒絕域為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 5.99 \right\}$$

由試算表 E-16-5-2 中，得知檢定值為 $0.228214 < 5.99$ ，所以檢定結果為不拒絕 H_0 ，這表示該地區居民宗教信仰的結構比例並未明顯偏離 H_0 所描述。

- 16.6 欲檢定貸款信用狀況與所得是否相關，抽樣調查 102 位貸款人，以還款信用狀況與所得水準將他們歸類成表 E-16-6-1。

表 E-16-6-1

信用狀況	所得水準			
	高	中	低	
良好	18	22	12	52
不良	16	20	14	50
	34	42	26	102

在顯著水準 $\alpha=0.025$ 下，檢定信用狀況與所得水準是否相關。

解：

H_0 ：貸款信用狀況與所得高低具獨立性

H_1 ：貸款信用狀況與所得高低具相關性

列聯表獨立性檢定的試算表如表 E-16-6-2 及表 E-16-6-3 所示。

表 E-16-6-2

	高	中	低	
良好	17.33	21.41	13.25	52.00
不良	16.67	20.59	12.75	50.00
	34.00	42.00	26.00	102.00

表 E-16-6-3

	高	中	低	
良好	0.0256	0.0162	0.1188	0.1606
不良	0.0267	0.0168	0.1236	0.1670
				0.3276

$\alpha=0.025$ ， $\chi^2_{0.025}(2 \times 1) = \chi^2_{0.025}(2) = 7.38$ 。所以檢定的拒絕域 R 為

340 • 統計學習題解答

$$R = \left\{ \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \mid \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 7.38 \right\}$$

由試算表 E-16-6-3 得知，檢定值為 $0.3276 < 7.38$ 。所以檢定的結果為不拒絕 H_0 ，這表示貸款信用狀況與所得高低並不具顯著相關性。

- 16.7 欲檢定民眾對電視節目的偏好與教育程度是否相關，抽樣調查 120 位民眾依他們喜好的電視節目與教育程度歸類成表 E-16-7-1。

表 E-16-7-1

教育程度	節目型態			
	休閒	政治	經濟	
高	30	10	25	65
低	10	37	8	55
	40	47	33	120

在顯著水準 $\alpha=0.025$ 下，檢定節目型態偏好與教育程度是否相關。

解：

H_0 ：節目型態偏好與教育程度高低具獨立性

H_1 ：節目型態偏好與教育程度高低具相關性

獨立性檢定的列聯試算表如表 E-16-7-2 及表 E-16-7-3 所示。

表 E-16-7-2

	休閒	政治	經濟	
高	21.67	25.46	17.88	65.00
低	18.33	21.54	15.13	55.00
	40.00	47.00	33.00	120.00

表 E-16-7-3

	休閒	政治	經濟	
高	3.2051	9.3863	2.8400	15.4315
低	3.7879	11.0929	3.3564	18.2372
				33.6687

$\alpha = 0.025$ ， $\chi^2_{0.025}(2 \times 1) = \chi^2_{0.025}(2) = 7.38$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \mid \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 7.38 \right\}$$

由試算表 E-16-7-3 得知，檢定值為 $33.6687 > 7.38$ 。所以檢定的結果為拒絕 H_0 ，這表示電視節目的偏好與教育程度有顯著的相關性。

- 16.8 欲檢定民眾對於某法案贊成與否的態度傾向與性別是否相關，抽樣調查 160 位民眾並依其態度與性別歸類成表 E-16-8-1。

表 E-16-8-1

性別	態 度			
	贊成	反對	無意見	
男	38	12	26	76
女	24	10	50	84
	62	22	76	160

在顯著水準 $\alpha = 0.025$ 下，檢定對法案贊成與否的態度與性別是否相關。

解：

H_0 ：對法案所持的態度與性別相互獨立

H_1 ：對法案所持的態度與性別相關

獨立性檢定的列聯表試算如表 E-16-8-2 及表 E-16-8-3 所示。

表 E-16-8-2

	贊成	反對	無意見	
男	29.45	10.45	36.10	76.00
女	32.55	11.55	39.90	84.00
	62.00	22.00	76.00	160.00

表 E-16-8-3

	贊成	反對	無意見	
男	2.4823	0.2299	2.8258	5.5379
女	2.2459	0.2080	2.5566	5.0105
				10.5484

$\alpha = 0.025$ ， $\chi^2_{0.025}(2 \times 1) = \chi^2_{0.025}(2) = 7.38$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \mid \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 7.38 \right\}$$

由試算表 E-16-8-3 得知，檢定值為 $10.5484 > 7.38$ 。所以檢定的結果為拒絕 H_0 ，這表示對法案贊成與否的態度傾向與性別有顯著相關性。

16.9 某次民意調查中，共發出 1800 份問卷，受訪者年齡及是否願意接受訪查歸類成表 E-16-9-1。

表 E-16-9-1

是否願意接受 訪查	年 齡			
	18 歲以下	18-40 歲	40 歲以上	
是	120	150	180	450
否	400	600	350	1350
	520	750	530	1800

在顯著水準 $\alpha=0.05$ 下，檢定年齡與接受訪查意願是否相關。

解：

H_0 ：年齡與接受訪查意願相互獨立

H_1 ：年齡與接受訪查意願相關

獨立性檢定的列聯表試算如表 E-16-9-2 及表 E-16-9-3 所示。

表 E-16-9-2

	18 歲以下	18-40 歲	40 歲以上	
是	130	187.5	132.5	450
否	390	562.5	397.5	1350
	520	750	530	1800

表 E-16-9-3

	18 歲以下	18-40 歲	40 歲以上	
是	0.76923	7.50000	17.02830	25.29753
否	0.25641	2.50000	5.67610	8.43251
				33.73004

$\alpha=0.05$ ， $\chi_{0.05}^2(2 \times 1) = \chi_{0.05}^2(2) = 5.99$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \mid \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 5.99 \right\}$$

由試算表 E-16-9-3 得知，檢定值為 $33.73 > 5.99$ ，所以檢定的結果為拒絕 H_0 ，這表示是否接受訪查與受訪者年齡有顯著相關。

16.10 牙科醫生記錄 300 位病人的蛀牙數及是否有飯後刷牙習慣如表 E-16-10-1 所示。

344 • 統計學習題解答

表 E-16-10-1

刷牙習慣	蛀牙數目				
	1	2	3	4 以上	
有	45	33	24	12	114
無	18	36	57	75	186
	63	69	81	87	300

在顯著水準 $\alpha=0.05$ 下，檢定蛀牙數與刷牙習慣是否相關。

解：

H_0 ：蛀牙數目與刷牙習慣相互獨立

H_1 ：蛀牙數目與刷牙習慣相關

獨立性檢定列聯表試算如表 E-16-10-2 及表 E-16-10-3 所示。

表 E-16-10-2

	1	2	3	4 以上	
有	23.94	26.22	30.78	33.06	114
無	39.06	42.78	50.22	53.94	186
	63	69	81	87	300

表 E-16-10-3

	1	2	3	4 以上	
有	18.52647	1.753181	1.49345	13.41572	35.18881
無	11.35493	1.07453	0.915341	8.222536	21.56734
					56.75615

$\alpha=0.05$ ， $\chi^2_{0.05}(3 \times 1) = \chi^2_{0.05}(3) = 7.81$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \mid \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 7.81 \right\}$$

由試算表 E-16-10-3 得知，檢定值為 $56.756 > 7.81$ ，所以檢定結果為拒絕 H_0 ，

這表示蛀牙數目與飯後刷牙習慣有顯著相關。

16.11 抽樣調查 100 位大專畢業社會新鮮人第一份月薪的分佈如表 E-16-11-1。

表 E-16-11-1

薪水區間	人 數
26500-27000	3
26000-26500	5
25500-26000	8
25000-25500	15
24500-25000	22
24000-24500	23
23500-24000	24

在顯著水準 $\alpha=0.05$ 下，檢定月薪的分配為 $N(25000, 1000^2)$ 。

解：

以 X 表示月薪，則適合度檢定的 H_0 及 H_1 為

$$H_0 : X \sim N(25000, 1000^2)$$

$$H_1 : H_0 \text{ 不成立}$$

表 E-16-11-2

X	p_i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
26500-27000	0.067	3	6.68	-3.681	13.54772	2.027882
26000-26500	0.092	5	9.18	-4.185	17.51258	1.906691
25500-26000	0.150	8	14.99	-6.988	48.83532	3.258245
25000-25500	0.191	15	19.15	-4.146	17.19136	0.897897
24500-25000	0.191	22	19.15	2.854	8.143908	0.425353
24000-24500	0.150	23	14.99	8.012	64.1885	4.282595
23500-24000	0.159	24	15.89	8.134	66.16967	4.170657
	1.000	100	100			16.96932

適合度檢定的試算表如表 E-16-11-2 所示，其中第二欄 (p_i) 的計算為

$$\begin{aligned}
 P(X \geq 26500) &= P(Z \geq \frac{26500 - 25000}{1000}) = 0.067 \\
 P(26000 \leq X \leq 26500) &= P(\frac{26500 - 25000}{1000} \leq Z \leq \frac{265000 - 25000}{1000}) = 0.092 \\
 P(25500 \leq X \leq 26000) &= P(\frac{25500 - 25000}{1000} \leq Z \leq \frac{265000 - 25000}{1000}) = 0.15 \\
 P(25000 \leq X \leq 25500) &= P(\frac{25000 - 25000}{1000} \leq Z \leq \frac{25500 - 25000}{1000}) = 0.191 \\
 P(24500 \leq X \leq 25000) &= P(\frac{24500 - 25000}{1000} \leq Z \leq \frac{25000 - 25000}{1000}) = 0.191 \\
 P(24000 \leq X \leq 24500) &= P(\frac{24000 - 25000}{1000} \leq Z \leq \frac{24500 - 25000}{1000}) = 0.15 \\
 P(X \geq 24000) &= P(Z \geq \frac{24000 - 25000}{1000}) = 0.159
 \end{aligned}$$

表 E-16-11-2 中的第四欄 (期望值) 計算如下 (表中數字為電腦計算，經四捨五入後可能會有些許誤差)

$$\begin{aligned}
 E_1 &= n \cdot p_1 = 100 (0.067) = 6.68 \\
 E_2 &= n \cdot p_2 = 100 (0.092) = 9.2 \\
 E_3 &= n \cdot p_3 = 100 (0.15) = 15 \\
 E_4 &= n \cdot p_4 = 100 (0.191) = 19.1 \\
 E_5 &= n \cdot p_5 = 100 (0.191) = 19.1 \\
 E_6 &= n \cdot p_6 = 100 (0.15) = 15 \\
 E_7 &= n \cdot p_7 = 100 (0.159) = 15.9
 \end{aligned}$$

$\alpha = 0.05$ ， $\chi^2_{0.05}(7-1) = \chi^2_{0.05}(6) = 12.59$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 12.59 \right\}$$

由試算表 E-16-11-2 第 7 欄得知， χ^2 檢定值為 $16.969 > 12.59$ ，所以檢定結果為拒絕 H_0 ，這表示月薪分配顯著偏離 $N(25000, 1000^2)$ ，所以用 $N(25000, 1000^2)$ 來描述新鮮人第一份月薪的分佈並不適合。

16.12 同 11 題，在顯著水準 $\alpha=0.05$ 下，檢定月薪分配是否為常態分配。

解：

在第 11 題中常態分配的參數已知，而本題則是在參數未知下進行適合度檢定，所以我們要先估計其平均數及變異數（或標準差）。

表 E-16-12-1

X	人數			
26750	3	80250	2065	12792675
26250	5	131250	1565	12246125
25750	8	206000	1065	9073800
25250	15	378750	565	4788375
24750	22	544500	65	92950
24250	23	557750	-435	4352175
23750	24	570000	-935	20981400
		2468500		64327500
		24685		806.085

表 E-16-12-1 為估計平均數 ($\bar{X}$) 及變異數 (S^2) 的試算表，其中第一欄為表 E-16-11-1 各薪水區間的組中點。第三欄為第一欄與第二欄的乘積，並求得 $\bar{X} = \frac{2468500}{100} = 24685$ 。第四欄為第一欄與 $\bar{X}$ 的偏差。第五欄為第四欄的平方，並求得 $S = 806.085$ 。所以，適合度檢定的 H_0 及 H_1 為

$$H_0 : X \sim N(24685, 806^2)$$

$$H_1 : H_0 \text{ 不成立}$$

表 E-16-12-2

X	p_i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
26500-27000	0.012	3	1.22	1.78	3.18	2.6107
26000-26500	0.039	5	3.92	1.08	1.16	0.2953
25500-26000	0.105	8	10.46	-2.46	6.04	0.5779
25000-25500	0.192	15	19.20	-4.20	17.63	0.9182
24500-25000	0.243	22	24.28	-2.28	5.19	0.2138
24000-24500	0.212	23	21.15	1.85	3.42	0.1615
23500-24000	0.198	24	19.77	4.23	17.87	0.9040
	1.000	100	100			5.6815

適合度檢定的試算表如表 E-16-12-2 所示，其中第二欄 (p_i) 的計算如下：

$$P(X \geq 26500) = P(Z \geq \frac{26500 - 24685}{806}) = 0.012$$

$$P(26000 \leq X \leq 26500) = P(\frac{26000 - 24685}{806} \leq Z \leq \frac{26500 - 24685}{806}) = 0.039$$

$$P(25500 \leq X \leq 26000) = P(\frac{25500 - 24685}{806} \leq Z \leq \frac{26000 - 24685}{806}) = 0.105$$

$$P(25000 \leq X \leq 25500) = P(\frac{25000 - 24685}{806} \leq Z \leq \frac{25500 - 24685}{806}) = 0.192$$

$$P(24500 \leq X \leq 25000) = P(\frac{24500 - 24685}{806} \leq Z \leq \frac{25000 - 24685}{806}) = 0.243$$

$$P(24000 \leq X \leq 24500) = P(\frac{24000 - 24685}{806} \leq Z \leq \frac{24500 - 24685}{806}) = 0.212$$

$$P(X \leq 24000) = P(Z \leq \frac{24000 - 24685}{806}) = 0.198$$

由於在檢定過程中我們採用了兩個參數估計值，所以檢定統計量的自由度減 2， $\alpha = 0.05$ ， $\chi_{0.05}^2(7 - 1 - 2) = \chi_{0.05}^2(4) = 9.49$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 9.49 \right\}$$

由試算表 E-16-12-2 第 7 欄得知，檢定值為 $5.6815 < 9.49$ ，所以檢定的結果為不拒絕 H_0 ，這表示月薪分配並未明顯偏離 $N(24685, 806^2)$ 。

16.13 抽樣調查 1000 個家庭的年收入，將它們歸類如表 E-16-13-1。

表 E-16-13-1

150 萬以上	100-150 萬	80-100 萬	60-80 萬	60 萬以下
105	170	355	224	146

在顯著水準 $\alpha=0.05$ 下，欲檢定在經濟不景氣的衝擊下，家庭收入的結構比例是否仍如 H_0 所描述。其中 p_1 表示家庭年收入為 150 萬以上的比例， p_2 表示家庭年收入介於 100 萬到 150 萬間的比例， p_3 表示家庭年收入介於 80 萬到 100 萬間的比例， p_4 表示家庭年收入介於 60 萬到 80 萬間的比例， p_5 表示家庭年收入不到 60 萬的比例。

$$H_0 : p_1=0.1$$

$$p_2=0.2$$

$$p_3=0.4$$

$$p_4=0.18$$

$$p_5=0.12$$

$$H_1 : H_0 \text{ 不成立}$$

解：

卡方檢定的試算表如表 E-16-13-2 所示。

表 E-16-13-2

	150 萬以上	100-150 萬	80-100 萬	60-80 萬	60 萬以下	
p_i	0.1	0.2	0.4	0.18	0.12	
O_i	105	170	355	224	146	1000
E_i	100	200	400	180	120	1000
$O_i - E_i$	5	-30	-45	44	26	
$\frac{(O_i - E_i)^2}{E_i}$	0.2500	4.5000	5.0625	10.7556	5.6333	26.2014

350 · 統計學習題解答

$\alpha = 0.05$ ， $\chi^2_{0.05}(4) = 9.49$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 9.49 \right\}$$

由試算表 E-16-13-2，計算出卡方檢定統計值為

$$\begin{aligned} & \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(105 - 100)^2}{100} + \frac{(170 - 200)^2}{200} + \frac{(355 - 400)^2}{400} + \frac{(224 - 180)^2}{180} + \frac{(146 - 120)^2}{120} \\ &= 0.25 + 4.5 + 5.0625 + 10.7556 + 5.6333 \\ &= 26.2014 > 9.49 \end{aligned}$$

所以檢定的結果為拒絕 H_0 ，這表示家庭收入結構已明顯的打破如 H_0 所陳述的結構。我們更進一步檢視各組的 $\frac{(O_i - E_i)^2}{E_i}$ 值時發現，第一組（150 萬以上）數值甚小（0.25），這表示家庭所得在 150 萬以上者，在整體結構上的比例大致維持不變，其他四組中有兩組（年收入在 100 萬到 150 萬及 80 萬到 100 萬）比例結構向下降，另外兩組（年收入在 60 萬到 80 萬及年收入不及 60 萬）比例結構向上升。這顯示了，經濟不景氣對高收入家庭影響不大，中低收入家庭所受到的衝擊甚大。

- 16.14 欲檢定車禍受傷的嚴重性與繫安全帶是否相關，抽樣調查 160 起車禍並將其歸類成如表 E-16-14-1 所示。

表 E-16-14-1

	死亡	重傷	輕傷	
繫安全帶	12	20	32	64
未繫安全帶	44	32	20	96
	56	52	52	160

在顯著水準 $\alpha = 0.05$ 下，檢定車禍嚴重性與繫安全帶是否相關。

解：

H_0 ：車禍嚴重性與繫安全帶相互獨立

H_1 ：車禍嚴重性與繫安全帶具有相關性

列聯表中每一個格子的期望值 ($\hat{E}_i$) 及卡方檢定值試算表分別如表 E-16-14-2 及表 E-16-14-3 所示。

表 E-16-14-2

	死亡	重傷	輕傷	
繫安全帶	22.40	20.80	20.80	64.00
未繫安全帶	33.60	31.20	31.20	96.00
	56.00	52.00	52.00	160.00

表 E-16-14-3

	死亡	重傷	輕傷	
繫安全帶	4.8286	0.0308	6.0308	10.8901
未繫安全帶	3.2190	0.0205	4.0205	7.2601
				18.1502

$\alpha = 0.05$ ， $\chi^2_{0.05}(2 \times 1) = \chi^2_{0.05}(2) = 5.99$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \mid \sum \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 5.99 \right\}$$

由試算表 E-16-14-3 得知，檢定值為 $18.1501 > 5.99$ 。所以檢定的結果為拒絕 H_0 ，這表示車禍嚴重性與是否繫安全帶顯著相關。

16.15 抽樣調查 105 位社區民眾電子郵件地址 (E-mail address) 數目的分配如表 E-16-15-1 所示。

表 E-16-15-1

電子郵件地址數目	人數
0	6
1	10
2	18
3	20
4	24
5	18
6	9

在顯著水準 $\alpha=0.05$ 下，檢定民眾擁有電子郵件地址數量是否為卜瓦松分配。

解：

由於卜瓦松分配的參數（平均數）未知，所以進行檢定之前先估計其平均數，以 X 表示電子郵件地址數量，則

$$\bar{X} = \frac{0(6) + 1(10) + \cdots + 6(9)}{105} = 3.3$$

所以，適合度檢定的 H_0 及 H_1 可擬定為

$$H_0 : X \sim \text{Poisson}(3.3)$$

$$H_1 : H_0 \text{ 不成立}$$

表 E-16-15-2

X	p_i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
0	0.037	6	3.89	2.11	4.45	1.1428
1	0.122	10	12.82	-2.82	7.97	0.6213
2	0.201	18	21.13	-3.13	9.78	0.4627
3	0.221	20	23.21	-3.21	10.28	0.4428
4	0.182	24	19.12	4.88	23.84	1.2472
5	0.120	18	12.60	5.40	29.17	2.3153
≥ 6	0.117	9	12.24	-3.24	10.48	0.8567
	1.000	105	105			7.0888

適合度檢定的試算表如表 E-16-15-2，其中第二欄 (p_i) 的計算如下（小數點三位有效，所以計算中會有些許誤差）

$$\begin{aligned} p_1 &= P(X=0) = \frac{e^{3.3}(3.3)^0}{0!} = 0.037 \\ p_2 &= P(X=1) = \frac{e^{3.3}(3.3)^1}{1!} = 0.122 \\ p_3 &= P(X=2) = \frac{e^{3.3}(3.3)^2}{2!} = 0.201 \\ p_4 &= P(X=3) = \frac{e^{3.3}(3.3)^3}{3!} = 0.221 \\ p_5 &= P(X=4) = \frac{e^{3.3}(3.3)^4}{4!} = 0.182 \\ p_6 &= P(X=5) = \frac{e^{3.3}(3.3)^5}{5!} = 0.12 \\ p_7 &= P(X \geq 6) = 1 - P(0 \leq X \leq 5) = 0.117 \end{aligned}$$

由於檢定過程中我們使用了一個估計值（平均數），所以檢定統計量的自由度減 1。 $\alpha = 0.05$ ， $\chi^2_{0.05}(7 - 1 - 1) = \chi^2_{0.05}(5) = 11.07$ 。所以檢定的拒絕域 R 為

$$R = \left\{ \sum \frac{(O_i - E_i)^2}{E_i} \mid \sum \frac{(O_i - E_i)^2}{E_i} > 11.07 \right\}$$

由試算表 E-16-15-2 得知，檢定統計值為 $7.088 < 11.07$ ，所以檢定結果為不拒絕 H_0 ，這表示該社區民眾擁有電子郵件地址數量分配並未明顯偏離 Poisson(3.3)。

