十二、適合度和獨立性檢定

Chapter 12 Tests of goodness of fit and independence

追尋那道光 李明聰

目錄

學習目標

知識(認知)

1.可以清楚陳述特定研究變數適合度檢定的意涵。

2.可以清楚陳述兩個研究變數之間獨立性檢定的意涵。

3.可以說明各種狀況下，適合度檢定的程序和標準。

4.可以說明各種狀況下，兩個母體獨立性假設檢定的程序和標準。

5.評價各種情境下，適合度假設檢定和獨立性假設檢定的使用價值。

技能

1.能夠計算各種情境下的檢定統計值。

2.能夠利用檢定統計值與臨界值的比較，提出統計推論。

3.綜合所學，能夠於實務領域中，依據特定情境的需求進行假設檢定程序。

態度(情意)

1.意識到在日常生活或未來工作環境中，適合度檢定和獨立性檢定的重要性。

2.在各種情境下，依循假設檢定的程序，接受統計推論所傳達的意涵。

上課時間需求：3 小時

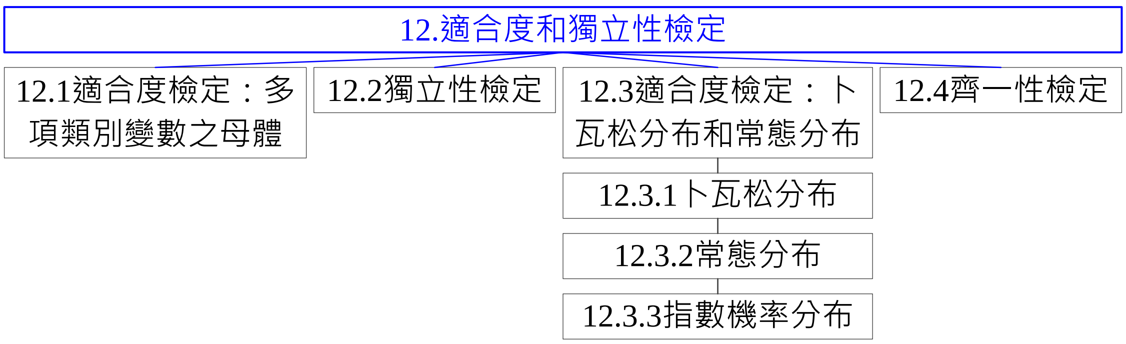
在社會科學領域中，任何一個研究議題的目標變數，皆可能受到很多其他變數的影響，任何其他變數對目標變數的影響之訊息，可來自於相關學術文獻或研究者的觀察判斷。例如：綠建築旅館推廣研究中，設定遊客對綠建築綠館住宿與否(名目尺度或順序尺度)為目標變數，遊客的性別(名目尺度或順序尺度)可能會影響遊客的住宿意願。因此，研究過程中欲探索【遊客對綠建築旅館住宿與否】與【遊客的性別】是否相互獨立。

此時，需要建立獨立性檢定假設：

虛無假設(null hypothesis) H_0: 遊客的性別與遊客對綠建築旅館住宿與否為相互獨立

對立假設(alternative hypothesis) H_1: 遊客的性別與遊客對綠建築旅館住宿與否為非相互獨立

設立顯著水準 α 進行卡方檢定(Chi-square testing)，以驗證其研究假設。



第 12 章節結構圖

12.1 適合度檢定：多項類別變數之母體

適合度檢定或適宜性檢定(goodness of fit test)是利用樣本資料檢定(驗證)母體的分布是否符合特定機率(比率)分布組合(mix)。

在多項母體(multinomial population)屬於母體中每一個基本元素皆可歸類於組別或類別型態(nominal scale)的情況。每次試驗皆屬於相互獨立。

多項類別變數(多項式)之母體的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合特定多項式機率分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定多項式機率分布。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)。

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)。

k：類別的數量。

$v = k - 1$：自由度。

E.若檢定統計值 $\chi^2 \le$ 臨界值$\chi^2_{\alpha,k-1}$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體特定變數符合特定多項式機率分布。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,k-1}$，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定多項式機率分布。

範例 12.1　深水大學石尖校區有 A、B 和 C 三家學生餐廳，該校區方圓 5 公里之內沒有其他餐廳可供學生選擇，故到該校區上課的學生只有三家餐廳可供選擇。經過一年的營運學生選擇餐廳皆已進入成熟的階段，三家餐廳市場佔有率分別為 30、35 和 35 %。現今 A 餐廳覺得其業績最差，提出用餐送飲料的行銷活動，企圖拉高其業績和佔有率。請觀光系學生進行市場調查，隨機抽取 300 位學生，詢問其前一天中餐到哪一家餐廳用餐，發現 A 餐廳 100 人；B 餐廳 105 人；C 餐廳 95 人。試評估此行銷活動是否改變原來的市場佔有率？(顯著水準 $\alpha = 0.05$)

題解：A 餐廳出現頻率 $f_A = 100$，B 餐廳出現頻率 $f_B = 105$，C 餐廳出現頻率 $f_C = 95$；A 餐廳人數期望值 $e_A = 300 \times 0.30 = 90$，B 餐廳人數期望值 $e_B = 300 \times 0.35 = 105$，C 餐廳人數期望值 $e_C = 300 \times 0.35 = 105$。餐廳數量 $k = 3$。

A.設定顯著水準 $\alpha = 0.05$，臨界值$\chi^2_{\alpha,k-1} = \chi^2_{0.05,3-1} = \chi^2_{0.05,2} = 5.9915$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: $p_A = 0.30, p_B = 0.35, p_C = 0.35$。

C.對立假設(alternative hypothesis) H_1: 母體的比率分布非 $p_A = 0.30, p_B = 0.35, p_C = 0.35$。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(100-90)^2}{90} + \frac{(105-105)^2}{105} + \frac{(95-105)^2}{105} = 1.1111 + 0 + 0.9524 = 2.0635$$

E.檢定統計值 $\chi^2 = 2.0635 <$ 臨界值$\chi^2_{\alpha,k-1} = 5.9915$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: $p_A = 0.30, p_B = 0.35, p_C = 0.35$。因此，A 餐廳提出用餐送飲料的行銷活動，沒有顯著的改變原來的市場佔有率。

練習 12.1　深水大學有 A、B、C 和 D 四家學生餐廳，該校區方圓 5 公里之內沒有其他餐廳可供學生選擇，故到該校區上課的學生只有四家餐廳可供選擇。經過一年的營運學生選擇餐廳皆已進入成熟的階段，四家餐廳市場佔有率依序分別為 0.25、0.25、0.25 和 0.25。現今 A 餐廳提出用餐送免費飲料的行銷活動，企圖拉高其業績和佔有率。請觀光系學生進行市場調查，隨機抽取 300 位學生，詢問其前一天中餐到哪一家餐廳用餐，發現 A 餐廳 95 人；B 餐廳 75 人；C 餐廳 60 人；D 餐廳 70 人。試評估此行銷活動是否改變原來的市場佔有率？(顯著水準 $\alpha = 0.05$)

題解：A 餐廳出現頻率 f_A = 95，B 餐廳出現頻率 f_B = 75，C 餐廳出現頻率 f_C = 60，D 餐廳出現頻率 f_D = 70；A 餐廳人數期望值 $e_A = 300 \times 0.25 = 75$，B 餐廳人數期望值 $e_B = 300 \times 0.25 = 75$，C 餐廳人數期望值 $e_C = 300 \times 0.25 = 75$，D 餐廳人數期望值 $e_D = 300 \times 0.25 = 75$。餐廳數量 $k = 4$。

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,k-1} = \chi^2_{0.05,4-1} = \chi^2_{0.05,3}$ = 7.8147[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: $p_A = 0.25, p_B = 0.25, p_C = 0.25, p_D = 0.25$。

C.對立假設(alternative hypothesis) H_1: 母體的比率分布非 $p_A = 0.25, p_B = 0.25, p_C = 0.25, p_D = 0.25$。

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i} = \frac{(95-75)^2}{75} + \frac{(75-75)^2}{75} + \frac{(60-75)^2}{75} + \frac{(70-75)^2}{75} = 5.3333 + 0 + 3.0000 + 0.3333 = 8.6667$

E.檢定統計值 χ^2 = 8.6667 > 臨界值$\chi^2_{\alpha,k-1}$ = 7.8147，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 母體的比率分布非 $p_A = 0.25, p_B = 0.25, p_C = 0.25, p_D = 0.25$。因此，A 餐廳提出用餐送飲料的行銷活動，有顯著的改變原來的市場佔有率。

練習 12.2 深水大學有 A、B、C、D 和 E 五家學生餐廳，該校區方圓 5 公里之內沒有其他餐廳可供學生選擇，故到該校區上課的學生只有五家餐廳可供選擇。經過一年的營運學生選擇餐廳皆已進入成熟的階段，五家餐廳市場佔有率依序分別為 0.15、0.15、0.25、0.20 和 0.25。現今 A 餐廳提出用餐送免費飲料的行銷活動，企圖拉高其業績和佔有率。請觀光系學生進行市場調查，隨機抽取 300 位學生，詢問其前一天中餐到哪一家餐廳用餐，發現 A 餐廳 65 人；B 餐廳 55 人；C 餐廳 60 人；D 餐廳 50 人；E 餐廳 70 人。試評估此行銷活動是否改變原來的市場佔有率？(顯著水準 α = 0.05)

題解：A 餐廳出現頻率 f_A = 65，B 餐廳出現頻率 f_B = 55，C 餐廳出現頻率 f_C = 60，D 餐廳出現頻率 f_D = 50，E 餐廳出現頻率 f_E = 70；A 餐廳人數期望值 $e_A = 300 \times 0.15 = 45$，B 餐廳人數期望值 $e_B = 300 \times 0.15 = 45$，C 餐廳人數期望值 $e_C = 300 \times 0.25 = 75$，D 餐廳人數期望值 $e_D = 300 \times 0.20 = 60$，E 餐廳人數期望值 $e_E = 300 \times 0.25 = 75$。餐廳數量 $k = 5$。

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,k-1} = \chi^2_{0.05,5-1} = \chi^2_{0.05,4}$ = 9.4877[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: $p_A = 0.15, p_B = 0.15, p_C = 0.25, p_D = 0.20, p_E = 0.25$。

C.對立假設(alternative hypothesis) H_1: 母體的比率分布非 $p_A = 0.15, p_B = 0.15, p_C = 0.25, p_D = 0.20, p_E = 0.25$。

D.計算檢定統計值：卡方值

檢定統計值$\chi^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i} = \frac{(65-45)^2}{45} + \frac{(55-45)^2}{45} + \frac{(60-75)^2}{75} + \frac{(50-60)^2}{60} + \frac{(70-75)^2}{75} = 8.8889 + 2.2222 + 3.0000 + 1.6667 + 0.3333 = 16.1111$

E.檢定統計值 χ^2 = 16.1111 > 臨界值$\chi^2_{\alpha,k-1}$ = 9.4877，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis)H_1: 母體的比率分布非 $p_A = 0.15, p_B = 0.15, p_C = 0.25, p_D = 0.20, p_E = 0.25$。因此，A 餐廳提出用餐送飲料的行銷活動，有顯著的改變原來的市場佔有率。

12.2 獨立性檢定

獨立性檢定(test of independence)或列聯表檢定是欲檢定單獨一個母體中兩個研究變數(屬性)之間的關係是否獨立之統計方法。

利用列聯表(contingency table, cross tabulation)的敘述統計表，進而利用卡方分析(Chi-square test)進行獨立性檢定(test of independence)。

列聯表

觀測次數 f_{ij}		欄 column						
		1	2	⋯	j	⋯	c	列合計
列 row	1	f_{11}	f_{12}	⋯	f_{1j}	⋯	f_{1c}	R_1
	2	f_{21}	f_{22}	⋯	f_{2j}	⋯	f_{2c}	R_2
	⋮	⋮	⋮		⋮		⋮	
	i	f_{i1}	f_{i2}	⋯	f_{ij}	⋯	f_{ic}	R_i
	⋮	⋮	⋮		⋮		⋮	
	r	f_{r1}	f_{r2}	⋯	f_{rj}	⋯	f_{rc}	R_r
	欄合計	C_1	C_2		C_j		Cc	n

獨立性檢定假設程序：

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: A 變數與 B 變數為相互獨立。

C.對立假設(alternative hypothesis) H_1: A 變數與 B 變數為非相互獨立。

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

其中 f_{ij}：在列聯表(contingency table)中，第 i 列及第 j 行(欄)的觀察次數(frequency)。

$e_{ij}=\frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}=\frac{\left(\sum_{j=1}^{c}f_{ij}\right)\times\left(\sum_{i=1}^{r}f_{ij}\right)}{\sum_{j=1}^{c}\sum_{i=1}^{r}f_{ij}}$，在列聯表(contingency table)中，第 i 列及第 j 行(欄)的期望次數(expected frequency)。

r：在列聯表中，列的數量。

c：在列聯表中，行(欄)的數量。

在列聯表中若所有行列方格的期望次數皆大於或等於 5($e_{ij} \geq 5$)，此一統計值符合自由度 v = (列數 $r-1$) × (欄數 $c-1$)的卡方分布。

D.若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: A 變數與 B 變數為相互獨立。

E.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: A 變數與 B 變數為非相互獨立。

範例 12.2 高美大學深水校區有 A、B 和 C 三家學生餐廳，該校區方圓 5 公里之內沒有其他餐廳可供學生選擇，故到該校區上課的學生只有三家餐廳可供選擇。經過一年的營運學生選擇餐廳皆已進入成熟的階段。欲瞭解學生性別與選擇餐廳時的偏好是否相互獨立。請觀光系學生進行市場調查，隨機抽取 300 位學生，詢問其前一天到哪一家餐廳用餐和其性別，調查結果如下表所示。試評估學生性別與選擇餐廳時的偏好是否相互獨立？(顯著水準 α = 0.05)

觀察次數		優先餐廳選擇			
		A	B	C	總和
學生性別	男生	40	35	50	125

觀察次數		優先餐廳選擇			
		A	B	C	總和
	女生	40	60	75	175
	總和	80	95	125	300

題解：期望次數表，利用期望次數 $e_{ij} = \frac{(第i列和)\times(第j行和)}{樣本數量}$，計算獲得

期望次數		優先餐廳選擇			
		A	B	C	總和
學生性別	男生	33.33	39.58	52.08	125
	女生	46.67	55.42	72.92	175
	總和	80	95	125	300

A.設定顯著水準 $\alpha = 0.05$，列數 $r = 2$，行數(欄數)$c = 3$，自由度 $v = (r - 1)\times(c - 1) = (2 - 1)\times(3 - 1) = 2$，臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = \chi^2_{0.05,(2-1)\times(3-1)} = \chi^2_{0.05,2} = 5.9915$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 對餐廳選擇的偏好與學生性別相互獨立。

C.對立假設(alternative hypothesis) H_1: 對餐廳選擇的偏好與學生性別並非相互獨立。

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(40-33.33)^2}{33.33} + \frac{(35-39.58)^2}{39.58} + \frac{(50-52.08)^2}{52.08} + \frac{(40-46.67)^2}{46.67} + \frac{(60-55.42)^2}{55.42} + \frac{(75-72.92)^2}{72.92} = 1.3333 + 0.5307 + 0.0833 + 0.9524 + 0.3791 + 0.0595 = 3.3383$

E.檢定統計值 $\chi^2 = 3.3383 < $ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = 5.9915$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 對餐廳選擇的偏好與學生性別相互獨立。因此，對餐廳選擇的偏好與學生性別相互獨立，對餐廳選擇的偏好與學生性別兩個變數之間沒有相關性存在。

練習 12.3　高美大學深水校區有 A、B、C 和 D 四家學生餐廳，該校區方圓 5 公里之內沒有其他餐廳可供學生選擇，故到該校區上課的學生只有四家餐廳可供選擇。經過一年的營運學生選擇餐廳皆已進入成熟的階段。欲瞭解學生性別與選擇餐廳時的偏好是否相互獨立。請觀光系學生進行市場調查，隨機抽取 300 位學生，詢問其前一天到哪一家餐廳用餐和其性別，調查結果如下表所示。試評估學生性別與選擇餐廳時的偏好是否相互獨立？(顯著水準 $\alpha = 0.05$)

觀察次數		優先餐廳選擇				
		A	B	C	D	總和
學生性別	男生	20	30	20	45	115
	女生	35	50	25	75	185
	總和	55	80	45	120	300

題解：期望次數表，利用 $e_{ij} = \frac{(第i列和)\times(第j行和)}{樣本數量}$ 計算獲得

期望次數		優先餐廳選擇				
		A	B	C	D	總和
學生性別	男生	21.08	30.67	17.25	46.00	115
	女生	33.92	49.33	27.75	74.00	185

總和	55	80	45	120	300

A.設定顯著水準 $\alpha = 0.05$，列數 $r = 2$，行數(欄數)$c = 4$，自由度 $v = (r-1)\times(c-1) = (2-1)\times(4-1) = 3$，臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = \chi^2_{0.05,(2-1)\times(4-1)} = \chi^2_{0.05,3} = 7.8147$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 對餐廳選擇的偏好與學生性別相互獨立。

C.對立假設(alternative hypothesis) H_1: 對餐廳選擇的偏好與學生性別並非相互獨立。

D.計算**檢定統計值**：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(20-21.08)^2}{21.08} + \frac{(30-30.67)^2}{30.67} + \frac{(20-17.25)^2}{17.25} + \frac{(45-46.00)^2}{46.00} + \frac{(35-33.92)^2}{33.92} + \frac{(50-49.33)^2}{49.33} + \frac{(25-27.75)^2}{27.75} + \frac{(75-74.00)^2}{74.00} = 0.0557 + 0.0145 + 0.4384 + 0.0217 + 0.0346 + 0.0090 + 0.2725 + 0.0135 = 0.8600$

E.檢定統計值 $\chi^2 = 0.8600 < $ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = 7.8147$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 對餐廳選擇的偏好與學生性別相互獨立。因此，對餐廳選擇的偏好與學生性別相互獨立，對餐廳選擇的偏好與學生性別兩個變數之間沒有相關性存在。

練習 12.4 高高餐廳有 A、B 和 C 三家魚肉供應商，在上個月經餐廳驗收人員抽檢資料如下表所示。以顯著水準 $\alpha = 0.05$ 檢定供應商與魚肉品質的獨立性。依據您的分析，對採購部門有何建議？

	魚肉品質		
供應商	優良	小瑕疵	嚴重瑕疵
A	20	9	9
B	40	6	5
C	35	5	5

題解：期望次數表，利用期望次數 $e_{ij} = \frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}$ 計算獲得

期望次數		魚肉品質			
		優良	小瑕疵	嚴重瑕疵	總和
供應商	A	26.94	5.67	5.39	38
	B	36.16	7.61	7.23	51
	C	31.90	6.72	6.38	45
	總和	95	20	19	134

A.設定顯著水準 $\alpha = 0.05$，列數 $r = 3$，行數(欄數)$c = 3$，自由度 $v = (r-1)\times(c-1) = (3-1)\times(3-1) = 4$，臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = \chi^2_{0.05,(3-1)\times(3-1)} = \chi^2_{0.05,4} = 9.4877$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 供應商與魚肉品質相互獨立。

C.對立假設(alternative hypothesis) H_1: 供應商與魚肉品質非相互獨立。

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(20-26.94)^2}{26.94} + \frac{(9-5.67)^2}{5.67} + \frac{(9-5.39)^2}{5.39} + \frac{(40-36.16)^2}{36.16} + \frac{(6-7.61)^2}{7.61} + \frac{(5-7.23)^2}{7.23} + \frac{(35-31.90)^2}{31.90} + \frac{(5-6.72)^2}{6.72} + \frac{(5-6.38)^2}{6.38} = 1.7879 + 1.9532 + 2.4213 + 0.4085 + 0.3414 + 0.6885 + 0.3006 + 0.4386 + 0.2987 = 8.6389$

E.檢定統計值 $\chi^2 = 8.6389 < $ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = 9.4877$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 供應商與魚肉品質相互獨立。因此，供應商與魚肉品質相互獨立，不同供應商對其供應的魚肉品質兩個變數之間沒有相關性存在。無法透過選擇供應商，而獲得魚肉品質比較好的原料，故，向哪一家供應商定貨，對魚肉品質沒有顯著的相關影響。

12.3 適合度檢定：卜瓦松分布和常態分布

適合性檢定或適宜性檢定(goodness of fit test)可應用於任何型態的機率分布。在母體屬於卜瓦松分布、常態分布和指數分布，皆使用卡方分布進行適合度檢定。

12.3.1 卜瓦松分布

母體(population)中特定變數分布屬於卜瓦松分布(在單位時間或空間內 n 個事件中所發生成功的次數 x 很少。)的適合度檢定。

卜瓦松機率函數(Poisson distribution function)

設在 n 次事件(樣本大小)中成功次數為隨機變數以 X 代表之，$x = 0, 1, 2, ...$，其平均值或期望值設為 $\mu = n \times p$，隨機變數 X 之機率分布公式。

$$f(x) = P_x = P(x) = e^{-\mu} \times \frac{\mu^x}{x!}$$

其中 $e = 2.71828$ 自然(對)數

卜瓦松分布母體的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合卜瓦松分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合卜瓦松分布。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)。

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)。

k：類別的數量。

p：欲利用樣本資料估算的母體參數(parameter)之個數。

自由度 $v = k - p - 1$。

E.若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,k-p-1}$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體特定變數符合卜瓦松分布。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,k-p-1}$，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 母體特定變數未符合卜瓦松分布。

範例 12.3 高雄清清連鎖 24 小時速食餐廳欲瞭解深夜時段在特定餐廳，每 10 分鐘內到達餐廳的人數分布是否符合卜瓦松分布，若符合卜瓦松分布時，可以提出符合卜瓦松分布的人力需求，以供管理階層進行服務員排班使用。請本系學生進行市場調查，在三個星期內於深夜時段隨機抽取 100 個 10 分鐘時段，利用計數器計算該 10 分鐘內，到達該餐廳的消費者人數，調查結果如下表所示。試評估每 10 分鐘內到達該餐廳人數是否符合卜瓦松分布？(顯著水準 $\alpha = 0.05$)

10 分鐘內消費者人數	觀察次數
0	2
1	5
2	6
3	17
4	25
5	18

10 分鐘內消費者人數	觀察次數
6	12
7	10
8	5
合計	100

題解：在 10 分鐘之內蒞臨該餐廳的消費者期望人數 μ(單位：人)，利用樣本估算

10 分鐘內消費者人數	觀察次數	人×次
0	2	0
1	5	5
2	6	12
3	17	51
4	25	100
5	18	90
6	12	72
7	10	70
8	5	40
合計	100	440

樣本平均值$\bar{x} = \frac{440}{100} = 4.4$ 人，以此數值代替母體平均人數 μ。利用 $f(x) = P_x = P(x) = e^{-\mu} \times \frac{\mu^x}{x!}$計算卜瓦松機率(或使用 Excel 軟體 POISSON.DIST 函數)與期望(觀察)次數(卜瓦松機率 × 全部觀察次數)。

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	100 次 10 分鐘期望(觀察)次數 = 100 × $f(x)$
0	0.0123	1.2277
1	0.0540	5.4020
2	0.1188	11.8845
3	0.1743	17.4305
4	0.1917	19.1736
5	0.1687	16.8728
6	0.1237	12.3734
7	0.0778	7.7775
8	0.0428	4.2776
9	0.0209	2.0913
10(含)以上	0.0149	1.4890
合計	1.0000	100.0000

欲利用卡方分布進行適合度檢定時，每一個類別(組別)的期望(觀察)次數必須至少 5。因此，必須透過鄰近(相似)類別(組別)合併的方式，達到期望次數至少 5 的條件。為了確保【10 分鐘內消費者人數】最高人數的期望次數可以達到 5 以上一點點(期望次數若太高，可能還要評估是否增列最高人數，以達到最佳檢定效益)，雖然【10 分鐘內消費者人數】觀察最多只有 8 人，在計算期望次數時，將【10 分鐘內消費者人數】多計算幾個人數，以了解期望次數的分布狀況，方便後續進行類別(組別)合併的程序。

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	100 次 10 分鐘期望(觀察)次數 = 100 × $f(x)$
0 或 1	0.0663	6.6298
2	0.1188	11.8845
3	0.1743	17.4305
4	0.1917	19.1736
5	0.1687	16.8728
6	0.1237	12.3734
7	0.0778	7.7775
8(含)以上	0.0786	7.8579

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	100 次 10 分鐘期望(觀察)次數 = 100 × $f(x)$
合計	1.0000	100.0000

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,v}$ = $\chi^2_{\alpha,k-p-1}$ = $\chi^2_{0.05,8-1-1}$ = $\chi^2_{0.05,6}$ = 12.5916[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 母體符合卜瓦松分布。

C.對立假設(alternative hypothesis) H_1: 母體未符合卜瓦松分布。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i} = 6.4765$$

10 分鐘內消費者人數 X	觀察次數 f_i	100 次 10 分鐘期望(觀察)次數 e_i = 100×$f(x)$	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
0 或 1	7	6.6298	0.3702	0.1370	0.0207
2	6	11.8845	–5.8845	34.6273	2.9137
3	17	17.4305	–0.4305	0.1853	0.0106
4	25	19.1736	5.8264	33.9469	1.7705
5	18	16.8728	1.1272	1.2706	0.0753
6	12	12.3734	–0.3734	0.1394	0.0113
7	10	7.7775	2.2225	4.9395	0.6351
8(含)以上	5	7.8579	–2.8579	8.1676	1.0394
合計	100	100.0000			6.4765

k = 8 類別的數量，p = 1 欲利用樣本資料估算的母體參數(parameter)之個數(每 10 分鐘內到達該餐廳人數)變數，故，自由度 $v = k - p - 1 = 8 - 1 - 1 = 6$。

E.檢定統計值 χ^2 = 6.4765 < 臨界值$\chi^2_{\alpha,v}$ = 12.5916，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體符合卜瓦松分布。每 10 分鐘內到達該餐廳人數分布型態符合卜瓦松分布。

練習 12.5 高雄鈺鈺連鎖 24 小時速食餐廳欲瞭解深夜時段在特定餐廳，每 10 分鐘內到達餐廳的人數分布是否符合卜瓦松分布，若符合卜瓦松分布時，可以提出符合卜瓦松分布的人力需求，以供管理階層排班使用。請本系學生進行市場調查，在四個星期內於深夜時段隨機抽取 120 個 10 分鐘時段，利用計數器計算該 10 分鐘內，到達該餐廳的消費者人數，調查結果如下表所示。試評估每 10 分鐘內到達該餐廳人數是否符合卜瓦松分布？(顯著水準 α = 0.05)

10 分鐘區間內消費者人數	觀察次數
0	2
1	5
2	6
3	17
4	22
5	20
6	14
7	12
8	9
9	8
10	5
合計	120

題解：在 10 分鐘之內蒞臨該餐廳的消費者期望人數 μ，利用樣本估算

10 分鐘內消費者人數	觀察次數	人次
0	2	0
1	5	5
2	6	12
3	17	51
4	22	88
5	20	100
6	14	84
7	12	84
8	9	72
9	8	72
10	5	50
合計	120	618

樣本平均值$\bar{x} = \frac{618}{120} = 5.1500$，以此數值代替母體平均人數 μ。利用 $f(x) = P_x = P(x) = e^{-\mu} \times \frac{\mu^x}{x!}$計算(或使用 Excel 軟體 POISSON.DIST 函數)卜瓦松機率和期望觀察次數(卜瓦松機率 × 全部觀察次數)。

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	120 次 10 分鐘期望(觀察)次數 = 120 × $f(x)$
0	0.0058	0.6959
1	0.0299	3.5840
2	0.0769	9.2289
3	0.1320	15.8429
4	0.1700	20.3978
5	0.1751	21.0097
6	0.1503	18.0333
7	0.1106	13.2674
8	0.0712	8.5409
9	0.0407	4.8873
10	0.0210	2.5169
11	0.0098	1.1784
12	0.0042	0.5057
13	0.0017	0.2003
14	0.0006	0.0737
15	0.0002	0.0253
16	0.0001	0.0081
合計	1.0000	120.0000

欲利用卡方分布進行適合度檢定時，每一個類別(組別)的期望(觀察)次數必須至少 5。因此，必須透過鄰近(相似)類別(組別)合併的方式，達到期望次數至少 5 的條件。

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	120 次 10 分鐘期望(觀察)次數 = 120×$f(x)$
0, 1 或 2	0.1126	13.5088
3	0.1320	15.8429
4	0.1700	20.3978
5	0.1751	21.0097
6	0.1503	18.0333
7	0.1106	13.2674
8	0.0712	8.5409
9(含)以上	0.0783	9.3958
合計	1.0000	120.0000

A.設定顯著水準 $\alpha = 0.05$，臨界值$\chi^2_{\alpha,v} = \chi^2_{\alpha,k-p-1} = \chi^2_{0.05,8-1-1} = \chi^2_{0.05,6} = 12.5916$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 母體符合卜瓦松分布。

C.對立假設(alternative hypothesis) H_1: 母體未符合卜瓦松分布。

D.計算檢定統計值─卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i} = 2.7084$

10 分鐘內消費者人數 X	觀察次數 f_i	120 次 10 分鐘期望(觀察)次數 $e_i = 100\times f(x)$	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
0, 1 或 2	13	13.5088	-0.5088	0.2589	0.0192
3	17	15.8429	1.1571	1.3388	0.0845
4	22	20.3978	1.6022	2.5672	0.1259
5	20	21.0097	-1.0097	1.0195	0.0485
6	14	18.0333	-4.0333	16.2676	0.9021
7	12	13.2674	-1.2674	1.6062	0.1211
8	9	8.5409	0.4591	0.2108	0.0247
9(含)以上	13	9.3958	3.6042	12.9901	1.3825
合計	120	120.0000			2.7084

$k = 8$ 類別的數量，$p = 1$ 欲利用樣本資料估算的母體參數(parameter)之個數(每 10 分鐘內到達該餐廳人數)變數，自由度 $v = k - p - 1 = 8 - 1 - 1 = 6$。

E.檢定統計值 $\chi^2 = 2.7084$ < 臨界值$\chi^2_{\alpha,v} = 12.5916$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體符合卜瓦松分布。每 10 分鐘內到達該餐廳人數分布型態符合卜瓦松分布。

練習 12.6 高雄鮮味連鎖 24 小時速食餐廳欲瞭解深夜時段在特定餐廳，每 10 分鐘內到達餐廳的人數分布是否符合卜瓦松分布，若符合卜瓦松分布時，可以提出符合卜瓦松分布的人力需求，以供管理階層排班使用。請本系學生進行市場調查，在五個星期內於深夜時段隨機抽取 110 個 10 分鐘時段，利用計數器計算該 10 分鐘內，到達該餐廳的消費者人數，調查結果如下表所示。試評估每 10 分鐘內到達該餐廳人數是否符合卜瓦松分布？(顯著水準 $\alpha = 0.05$)

10 分鐘區間內消費者人數	觀察次數
0	3
1	5
2	6
3	17
4	22
5	20
6	14
7	12
8	9
9	2
10	0
合計	110

題解：在 10 分鐘之內蒞臨該餐廳的消費者期望人數 μ，利用樣本估算

10 分鐘內消費者人數	觀察次數	人次
0	3	0
1	5	5
2	6	12
3	17	51
4	22	88
5	20	100
6	14	84
7	12	84

10 分鐘內消費者人數	觀察次數	人次
8	9	72
9	2	18
10	0	0
合計	110	514

樣本平均值$\bar{x} = \frac{514}{110} = 4.6727$，以此數值代替母體平均人數 μ。利用 $f(x) = P_x = P(x) = e^{-\mu} \times \frac{\mu^x}{x!}$計算卜瓦松機率(或使用 Excel 軟體 POISSON.DIST 函數)與期望(觀察)次數(卜瓦松機率 × 全部觀察次數)。

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	110 次 10 分鐘期望(觀察)次數 = 110 × $f(x)$
0	0.0093	1.0281
1	0.0437	4.8042
2	0.1020	11.2244
3	0.1589	17.4829
4	0.1857	20.4232
5	0.1735	19.0864
6	0.1351	14.8643
7	0.0902	9.9224
8	0.0527	5.7956
9	0.0274	3.0090
10	0.0128	1.4060
11	0.0054	0.5973
12	0.0021	0.2326
13	0.0008	0.0836
14	0.0003	0.0279
15	0.0001	0.0087
16	0.0000	0.0025
合計	1.0000	110.0000

欲利用卡方分布進行適合度檢定時，每一個類別(組別)的期望(觀察)次數必須至少 5。因此，必須透過鄰近(相似)類別(組別)合併的方式，達到期望次數至少 5 或稍微高於 5 的卡方檢定條件。

10 分鐘內消費者人數 X	卜瓦松機率 $f(x)$	110 次 10 分鐘期望(觀察)次數 = 110×$f(x)$
0 或 1	0.0530	5.8324
2	0.1020	11.2244
3	0.1589	17.4829
4	0.1857	20.4232
5	0.1735	19.0864
6	0.1351	14.8643
7	0.0902	9.9224
8	0.0527	5.7956
9(含)以上	0.0488	5.3676
合計	1.0000	110.0000

A.設定顯著水準 $\alpha = 0.05$，臨界值$\chi^2_{\alpha,v} = \chi^2_{\alpha,k-p-1} = \chi^2_{0.05,9-1-1} = \chi^2_{0.05,7} = 14.0671$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 母體符合卜瓦松分布。

C.對立假設(alternative hypothesis) H_1: 母體未符合卜瓦松分布。

D.計算檢定統計值—卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = 7.7860$$

10 分鐘內消費者人數 X	觀察次數 f_i	110 次 10 分鐘期望(觀察)次數 $e_i = 100\times f(x)$	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
0 或 1	8	5.8324	2.1676	4.6986	0.8056
2	6	11.2244	-5.2244	27.2946	2.4317
3	17	17.4829	-0.4829	0.2332	0.0133
4	22	20.4232	1.5768	2.4863	0.1217
5	20	19.0864	0.9136	0.8347	0.0437
6	14	14.8643	-0.8643	0.7469	0.0503
7	12	9.9224	2.0776	4.3165	0.4350
8	9	5.7956	3.2044	10.2684	1.7718
9(含)以上	2	5.3676	-3.3676	11.3408	2.1128
合計	110	110.0000			7.7860

$k = 9$ 類別的數量，$p = 1$ 欲利用樣本資料估算的母體參數(parameter)之個數(每 10 分鐘內到達該餐廳人數)變數，自由度 $v = k - p - 1 = 9 - 1 - 1 = 7$。

E.檢定統計值 $\chi^2 = 7.7860$ < 臨界值$\chi^2_{\alpha,v} = 14.0671$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體符合卜瓦松分布。每 10 分鐘內到達該餐廳人數分布型態符合卜瓦松分布。

12.3.2 常態分布

母體(population)中特定變數分布屬於常態分布(Normal distribution)與否的適合度檢定。

常態分布母體的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合特定平均值和標準(偏)差的常態分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定平均值和標準(偏)差的常態分布。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)。

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)。

k：類別的數量。

$v = k - p - 1$：自由度。

p：欲利用樣本資料估算的母體參數(parameter)之個數：平均值和標準(偏)差，故 $p = 2$。

E.若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,k-p-1}$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體特定變數符合特定平均值和標準(偏)差的常態分布。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,k-p-1}$，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定平均值和標準(偏)差的常態分布。

範例 12.4　餐旅系修統計學有 200 位學生，欲瞭解學生學習成效是否符合常態分布，以擬定相關篩選淘汰策略。現隨機抽取 50 位學生，進行統計學小考，考試成績如下表所示。試評估學生學習統計學成效(考試成績)是否符合常態分布？(顯著水準 $\alpha = 0.05$)

12	80	68	65	45	90	88	85	55	85
50	60	78	85	59	56	55	64	64	75
60	75	77	82	60	88	66	62	53	76
50	85	74	81	68	66	35	59	85	77

80	95	56	77	62	77	56	58	56	82

題解：利用樣本平均值和標準(偏)差作為母體平均值和標準(偏)差的點估計值。樣本平均值$\bar{x}$ = 67.94，樣本標準(偏)差 S = 15.68，樣本數量 n = 50。

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,v} = \chi^2_{\alpha,k-p-1} = \chi^2_{0.05,10-2-1} = \chi^2_{0.05,7}$ = 14.0671[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 修統計學學生的考試分數符合平均值 μ = 67.94 和標準(偏)差 σ = 15.68 的常態分布。

C.對立假設(alternative hypothesis) H_1: 修統計學學生的考試分數不符合平均值 μ = 67.94 和標準(偏)差 σ = 15.68 的常態分布。

欲利用卡方分布進行適合度檢定時，每一個類別(組別、層級)的期望次數必須至少 5。因此，在常態分布(ratio scale or interval scale)的檢定過程中，可以利用標準化常態分布表進行分數組數的劃分，以達到每個分數組數皆有期望次數至少 5 的條件。期望次數可以達到 5 的類別(組別、層級)數量即可，不宜提高期望次數而減少類別(組別、層級)數量，如此會降低假設檢定的效益(準確性)。樣本數量 45~49，$\frac{\text{樣本數量 45}}{\text{期望次數 5}}$ = 9，$\frac{\text{樣本數量 49}}{\text{期望次數 5}}$ = 9.8，分成 9 類別；樣本數量 50~54，$\frac{\text{樣本數量 50}}{\text{期望次數 5}}$ = 10，$\frac{\text{樣本數量 54}}{\text{期望次數 5}}$ = 10.8，分成 10 類別；樣本數量 55~59，$\frac{\text{樣本數量 55}}{\text{期望次數 5}}$ = 11，$\frac{\text{樣本數量 59}}{\text{期望次數 5}}$ = 11.8，分成 11 類別；樣本數量 60~64，$\frac{60}{5}$ = 12，$\frac{64}{5}$ = 12.8，分成 12 類別；樣本數量 65~69，$\frac{65}{5}$ = 13，$\frac{69}{5}$ = 13.8，分成 13 類別；其他樣本數量依此類推。

在樣本數量 n = 50 情況下，$\frac{\text{樣本數量 50}}{\text{期望次數 5}}$ = 10，分割為 10 個分數組數，期望次數為$\frac{\text{樣本數量 50}}{\text{分數組數 10}}$ = 5.0，達到稍微高於或等於 5 的卡方檢定條件。若在樣本數量 n = 51 情況，分割為 10 個分數組數，期望次數為$\frac{\text{樣本數量 51}}{\text{分數組數 10}}$ = 5.1，達到稍微高於或等於 5 的卡方檢定條件。若樣本數量 53，分割為 10 個分數組數，期望次數$\frac{\text{樣本數量 53}}{\text{分數組數 10}}$ = 5.3，達到稍微高於或等於 5 的卡方檢定條件。從機率 0 到 1 區間中，切成 10 等分，每一個等分機率為$\frac{\text{全部機率 1}}{\text{10 等分}}$ = 0.10，依序其對應的標準化 Z 值(利用 Excel 軟體 NORM.S.INV 函數)如下表所示。再由已知標準化 Z 值($z_i = \frac{x_i-\bar{x}}{S}$)、平均值和標準差數值，運算出其成績分數界限($x_i = z_i \times S + \bar{x}$)分別為：

機率 p	z	成績分數界限
0.1	-1.2816	47.85
0.2	-0.8416	54.74
0.3	-0.5244	59.72
0.4	-0.2533	63.97
0.5	0.0000	67.94
0.6	0.2533	71.91
0.7	0.5244	76.16
0.8	0.8416	81.14
0.9	1.2816	88.03

D.計算檢定統計值—卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = 13.6$$

50 位樣本同學的成績在各分數組別的觀察次數「善用 Excel 軟體中 COUNTIF(range, criteria "<"&A5)函數精準計次」與期望次數($\frac{\text{樣本數量 50}}{\text{分數組數 10}}$ = 5.0)。如下表所示，依序計算檢定統計值。

成績分數區間	觀察次數 f_i	期望次數 e_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
47.85 以下	3	5	-2	4	0.8
47.85~54.74	3	5	-2	4	0.8
54.74~59.72	9	5	4	16	3.2
59.72~63.97	5	5	0	0	0
63.97~67.94	5	5	0	0	0
67.94~71.91	2	5	-3	9	1.8
71.91~76.16	4	5	-1	1	0.2
76.16~81.14	8	5	3	9	1.8
81.14~88.03	9	5	4	16	3.2
88.03 以上	2	5	-3	9	1.8
合計			0	68	13.6

k = 10 成績分數組數的數量，p = 2 欲利用樣本資料估算的母體參數(parameter)之個數【平均值和標準(偏)差】變數，自由度 $v = k - p - 1 = 10 - 2 - 1 = 7$。

E.檢定統計值 $\chi^2 = 13.6$ < 臨界值$\chi^2_{\alpha,v} = 14.0671$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 修統計學學生的考試分數符合平均值 μ = 67.94 和標準(偏)差 σ = 15.68 的常態分布。

練習 12.7 餐旅系修統計學有 200 位學生，欲瞭解學生學習成效是否符合常態分布，以擬定相關篩選淘汰策略。現隨機抽取 40 位學生，進行統計學小考，考試成績如下表所示。試評估學生學習統計學成效(考試成績)是否符合常態分布？(顯著水準 $\alpha = 0.05$)

12	80	68	65	90	88	85	55
50	60	78	85	56	55	64	64
60	75	77	82	88	66	62	53
50	85	74	81	66	35	59	85
80	95	56	77	77	56	58	56

題解：利用樣本平均值和標準(偏)差作為母體平均值和標準(偏)差的點估計值。樣本平均值$\bar{x}$ = 67.7000，樣本標準(偏)差 S = 16.5269，樣本數量 n = 40。

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,v} = \chi^2_{\alpha,k-p-1} = \chi^2_{0.05,8-2-1} = \chi^2_{0.05,5}$ = 11.0705[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 修統計學學生的考試分數符合平均值 μ = 67.7000 和標準(偏)差 σ = 16.5269 的常態分布。

C.對立假設(alternative hypothesis) H_1: 修統計學學生的考試分數不符合平均值 μ = 67.7000 和標準(偏)差 σ = 16.5269 的常態分布。

欲利用卡方分布進行適合度檢定時，每一個類別(組別、層級)的期望次數必須至少 5。因此，在常態分布(ratio scale or interval scale)的檢定過程中，可以利用標準化常態分布表進行分數組數的劃分，以達到每個分數組數皆有期望次數至少 5 的條件。期望次數可以達到 5 的類別(組別、層級)數量即可，不宜提高期望次數而減少類別(組別、層級)數量，如此會降低假設檢定的效益(準確性)。樣本數量 45~49，分成 9 類別；樣本數量 50~54，分成 10 類別；樣本數量 55~59，分成 11 類別；樣本數量 60~64，分成 12 類別；樣本數量 65~69，分成 13 類別；其他樣本數量依此類推。

在樣本數量 n = 40 情況下，$\frac{\text{樣本數量 40}}{\text{期望次數 5}} = 8$，分割為 8 個分數組數，期望次數為$\frac{\text{樣本數量 40}}{\text{分數組數 10}} = 5.0$，達到稍微高於或等於 5 的卡方檢定條件。從機率 0 到 1 區間中，切成 8 等分，每一個等分機率為$\frac{\text{全部機率 1}}{\text{8 等分}}$

= 0.125，依序其對應的標準化 Z 值(利用 Excel 軟體 NORM.S.INV 函數)如下表所示。再由已知標準化 Z 值($z_i = \frac{x_i-\bar{x}}{S}$)、平均值和標準差數值，運算出其成績分數界限($x_i = z_i \times S + \bar{x}$)分別：

機率 p	z	成績分數界限
0.125	-1.1505	48.7
0.250	-0.6745	56.6
0.375	-0.3187	62.4
0.500	0.0000	67.7
0.625	0.3187	73.0
0.750	0.6745	78.8
0.875	1.1505	86.7

D.計算檢定統計值─卡方值

$$檢定統計值\ \chi^2 = \sum_{i=1}^{k} \frac{(f_i-e_i)^2}{e_i} = 10.4$$

40 位樣本同學的成績在各分數組別的觀察次數「善用 Excel 軟體中 COUNTIF(range, criteria "<"&A5)函數精準計次」與期望次數($\frac{樣本數量\ 40}{分數組數\ 8}$ = 5.0)。如下表所示，依序計算檢定統計值。

成績分數區間	觀察次數 f_i	期望次數 e_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
48.7 以下	2	5	-3	9	1.8
48.7~56.6	9	5	4	16	3.2
56.6~62.4	5	5	0	0	0
62.4~67.7	5	5	0	0	0
67.7~73.0	1	5	-4	16	3.2
73.0~78.8	6	5	1	1	0.2
78.8~86.7	8	5	3	9	1.8
86.7 以上	4	5	-1	1	0.2
合計	40	40	0	52	10.4

$k = 8$ 成績分數組數的數量，$p = 2$ 欲利用樣本資料估算的母體參數(parameter)之個數(平均值和標準(偏)差)變數，自由度 $v = k - p - 1 = 8 - 2 - 1 = 5$。

E.檢定統計值 $\chi^2 = 10.4$ < 臨界值$\chi^2_{\alpha,v}$ = 11.0705，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 修統計學學生的考試分數符合平均值 $\mu = 67.7000$ 和標準(偏)差 $\sigma = 16.5269$ 的常態分布。

練習 12.8 觀光系修統計學有 200 位學生，欲瞭解學生學習成效是否符合常態分布，以擬定相關篩選淘汰策略。現隨機抽取 35 位學生，進行統計學小考，考試成績如下表所示。試評估學生學習統計學成效(考試成績)是否符合常態分布？(顯著水準 $\alpha = 0.05$)

12	80	68	65	90	88	85
95	90	98	85	10	98	64
60	75	77	82	88	66	92
50	85	94	81	96	95	59
80	95	56	97	88	96	58

題解：利用樣本平均值和標準(偏)差作為母體平均值和標準(偏)差的點估計值。樣本平均值$\bar{x}$ = 77.0857，樣本標準(偏)差 S = 21.5698，樣本數量 n = 35。

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,v} = \chi^2_{\alpha,k-p-1} = \chi^2_{0.05,7-2-1} = \chi^2_{0.05,4}$ = 9.4877[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 修統計學學生的考試分數符合平均值 μ = 77.0857 和標準(偏)差 σ = 21.5698 的常態分布。

C.對立假設(alternative hypothesis) H_1: 修統計學學生的考試分數不符合平均值 μ = 77.0857 和標準(偏)差 σ = 21.5698 的常態分布。

欲利用卡方分布進行適合度檢定時，每一個類別(組別、層級)的期望次數必須至少 5。因此，在常態分布(ratio scale or interval scale)的檢定過程中，可以利用標準化常態分布表進行分數層級的劃分，以達到每個分數層級皆有期望次數至少 5 的條件。

在樣本數量 n = 35 情況下，$\frac{\text{樣本數量 35}}{\text{期望次數 5}}$ = 7，分割為 7 個分數組數，期望次數為$\frac{\text{樣本數量 35}}{\text{分數組數 7}}$ = 5.0，達到稍微高於或等於 5 的卡方檢定條件。從機率 0 到 1 區間中，切成 7 等分，每一個等分機率為 0.1429，依序其對應的標準化 Z 值(利用 Excel 軟體 NORM.S.INV 函數)如下表所示。再由已知標準化 Z 值($z_i = \frac{x_i-\bar{x}}{s}$)、平均值和標準差數值，運算出其成績分數界限($x_i = z_i \times S + \bar{x}$)分別：

機率 p	z	成績分數界限
0.1429	-1.0676	54.06
0.2857	-0.5659	64.88
0.4286	-0.1800	73.20
0.5714	0.1800	80.97
0.7143	0.5659	89.29
0.8571	1.0676	100.11

D.計算檢定統計值─卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i} = 18.4$$

35 位樣本同學的成績在各分數組別的觀察次數「善用 Excel 軟體中 COUNTIF(range, criteria "<"&A5)函數精準計次」與期望次數。如下表所示，依序計算檢定統計值。

成績分數區間	觀察次數 f_i	期望次數 e_i	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
54.06 以下	3	5	-2	4	0.8
54.06~64.88	5	5	0	0	0.0
64.88~73.20	3	5	-2	4	0.8
73.20~80.97	4	5	-1	1	0.2
80.97~89.29	8	5	3	9	1.8
89.29~100.11	12	5	7	49	9.8
100.11 以上	0	5	-5	25	5.0
合計	35	35	0	92	18.4

k = 7 成績分數組數的數量，p = 2 欲利用樣本資料估算的母體參數(parameter)之個數[平均值和標準(偏)差]變數。

E.檢定統計值 χ^2 = 18.4 ＞ 臨界值$\chi^2_{\alpha,v}$ = 9.4877，檢定統計值不位於虛無假設接受區域內，接受對立假設(alternative hypothesis) H_1: 修統計學學生的考試分數不符合平均值 μ = 77.0857 和標準(偏)差 σ = 21.5698 的常態分布。

練習 12.9 假設以下 20 筆資料是某班級同學統計學期末考分數：17, 18, 22, 27, 30, 30, 43, 46, 54, 63, 66, 71, 75, 82, 82, 88, 91, 93, 97, 99。請以顯著水準 0.05 檢定這 20 位學生統計學期末考分數是否為常態分布？(此 20 筆之平均數 $\bar{X}$ = 59.7，標準差 S = 28.6) (99 普考統計學概要)

12.3.3 指數機率分布【選擇教材】

母體(population)特定變數屬於指數(機率)分布(Exponential probability distribution)的適合度檢定。

母體特定變數屬於指數(機率)分布的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合特定速率參數 λ 的指數機率分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定速率參數 λ 的指數機率分布。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)。

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)。

k：類別的數量。

$v = k - p - 1$：自由度。

p：欲利用樣本資料估算的母體參數(parameter)之個數：速率參數 λ，故 $p = 1$。

E.若檢定統計值 $\chi^2 \le$ 臨界值 $\chi^2_{\alpha,k-p-1}$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 母體特定變數符合特定速率參數 λ 的指數機率分布。

F.若檢定統計值 $\chi^2 >$ 臨界值 $\chi^2_{\alpha,k-p-1}$，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定速率參數 λ 的指數機率分布。

範例 12.5　甜蜜廠牌新型燈管一批隨機抽出 200 支測試其壽命，請以顯著水準 $\alpha = 0.05$，檢定該批燈管之壽命是否符合平均壽命 250 天之指數分布。

壽命	0~100 天	100~200 天	200~300 天	300~∞
支數	52	53	70	25

題解：

A.設定顯著水準 $\alpha = 0.05$，臨界值 $\chi^2_{\alpha,v} = \chi^2_{\alpha,k-p-1} = \chi^2_{0.05,4-1-1} = \chi^2_{0.05,2} = 5.9915$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 該批燈管壽命符合速率參數 $\lambda = \frac{1}{250}$ (件/天)(平均值 250 天)的指數分布

C.對立假設(alternative hypothesis) H_1: 該批燈管壽命不符合速率參數 $\lambda = \frac{1}{250}$ (件/天)(平均值 250 天)的指數分布

D.計算檢定統計值—卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(52-65.94)^2}{65.94} + \frac{(53-44.20)^2}{44.20} + \frac{(70-22.63)^2}{22.63} + \frac{(25-60.24)^2}{60.24} = 80.3294$

壽命	0~100 天	100~200 天	200~300 天	300~∞天
支數 f_i	52	53	70	25
期望次數 e_i	65.94	44.20	29.63	60.24

速率參數 $\lambda = \frac{1}{250}$ (件/天)單位時間，以隨機變數 X 代表燈管的壽命(單位：天)

$P(X \le 100) = 1 - P(X \ge 100) = 1 - e^{-\frac{1}{250}\times 100} = 0.3297$(亦可使用 Excel 軟體 EXPON.DIST 函數運算)；期望次數 $e_1 = P(X \le 100) \times n = 0.3297 \times 200 = 65.94$。

$P(100 \le X \le 200) = P(X \ge 100) - P(X \ge 200) = e^{-\frac{1}{250}\times 100} - e^{-\frac{1}{250}\times 200} = 0.6703 - 0.4493 = 0.2210$；期望次數 $e_2 = P(100 \le X \le 200) \times n = 0.2210 \times 200 = 44.20$。

$P(200 \le X \le 300) = P(X \ge 200) - P(X \ge 300) = e^{-\frac{1}{250}\times 200} - e^{-\frac{1}{250}\times 300} = 0.4493 - 0.3012 = 0.1481$；期望次數 $e_3 = P(200 \le X \le 300) \times n = 0.1481 \times 200 = 29.63$。

$P(X \ge 300) = e^{-\frac{1}{250}\times 300} = 0.3012$；期望次數 $e_4 = P(X \ge 300) \times n = 0.3012 \times 200 = 60.24$。

期望次數都有達到 5 以上，因此可以使用卡方分布進行適合度檢定。唯，期望次數都高於 5 以上很多，因此檢定的精準度就會下降，此題目欲提升檢定精準度就要增加類別數量，只要達到期望次數都高於 5 即可。

k = 4 成績分數區間的數量，p = 1 欲利用樣本資料估算的母體參數(parameter)之個數(速率參數 λ)變數，自由度 $v = k - p - 1 = 4 - 1 - 1 = 2$。

E.檢定統計值 $\chi^2 = 80.3294$ > 臨界值$\chi^2_{\alpha,v} = 5.9915$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]，接受對立假設(alternative hypothesis) H_1: 該批燈管壽命不符合速率參數 $\lambda = \frac{1}{250}$ (件/天)(平均值 250 天)的指數分布。

12.4 齊一性檢定

齊一性檢定或均質性檢定(Test of homogeneity)為檢定兩個或兩個以上不同的母體是否具有相同機率分布或相同的比例之檢定方式。齊一性檢定由各母體中分別抽選出代表樣本，之後依據類別區分為一個多項列聯表。依據樣本所得到的觀察次數檢定各母體的比率是否相同。

齊一性檢定與獨立性檢定的程序相當類似，唯兩者資料蒐集的方式不同，獨立性檢定的蒐集資料時，並非事先決定好每一個母體的樣本數量，僅決定好全部樣本數量，在列聯表中各列或各行的總和之數值是隨機產生；齊一性檢定，在列聯表中各列或各行的總和之數值是事先設定好的，即事先決定好來自每一個母體的樣本數量。

齊一性檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 不同母體特定變數的分布相同。

C.對立假設(alternative hypothesis) H_1: 不同母體特定變數的分布不相同。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

其中 o_{ij}：在列聯表中第 i 列和第 j 行的樣本觀察次數(observation)

e_{ij}：在列聯表中第 i 列和第 j 行的期望次數(expected frequency)

r：橫列的數量(row)、母體數量

c：縱行的數量(column)、等級數量、類別區分等級數量

自由度：$df = (r - 1) \times (c - 1)$

E.若檢定統計值 $\chi^2 \le$ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 不同母體特定變數的分布相同。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$，檢定統計值不位於虛無假設接受區域內，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 不同母體特定變數的分布不相同。

範例 12.6 高雄市 A、B 和 C 三家咖啡速食餐廳，欲瞭解消費者對三家是否具有一樣的滿意度分布型態(比例)。於上週在三家咖啡速食餐廳中，分別隨機抽取 250 位消費者進行滿意度調查，三家咖啡速食餐廳的結果如下表所示。試評估 A、B 和 C 三家咖啡速食餐廳的滿意度分布型態(比例)是否相同？(顯著水準 α = 0.05)

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
A	23	56	50	70	51	250

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
B	32	52	45	76	45	250
C	41	45	38	71	55	250

題解：

A.設定顯著水準 α = 0.05，臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = \chi^2_{0.05,(3-1)\times(5-1)} = \chi^2_{0.05,8}$ = 15.5073[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: A、B 和 C 三家咖啡速食餐廳滿意度分布型態(比例)相同。

C.對立假設(alternative hypothesis) H_1: A、B 和 C 三家咖啡速食餐廳滿意度至少有兩家分布型態(比例)不相同。

D.計算檢定統計值：卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
A	23	56	50	70	51	250
B	32	52	45	76	45	250
C	41	45	38	71	55	250
合計	96	153	133	217	151	

期望次數表：利用期望值 $e_{ij} = \frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}$，計算獲得如下表

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
A	32.00	51.00	44.33	72.33	50.33	250
B	32.00	51.00	44.33	72.33	50.33	250
C	32.00	51.00	44.33	72.33	50.33	250
合計	96	153	133	217	151	

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \frac{(23-32.0)^2}{32.0} + \frac{(32-32.0)^2}{32.0} + \frac{(41-32.0)^2}{32.0} + \frac{(56-51.0)^2}{51.0} + \frac{(52-51.0)^2}{51.0} + \frac{(45-51.0)^2}{51.0} + \frac{(50-44.3)^2}{44.3} + \frac{(45-44.3)^2}{44.3} + \frac{(38-44.3)^2}{44.3} + \frac{(70-72.3)^2}{72.3} + \frac{(76-72.3)^2}{72.3} + \frac{(71-72.3)^2}{72.3} + \frac{(51-50.3)^2}{50.3} + \frac{(45-50.3)^2}{50.3} + \frac{(55-50.3)^2}{50.3} = 9.2096$$

E.檢定統計值 χ^2 = 9.2096 < 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$ = 15.5073，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: A、B 和 C 三家咖啡速食餐廳滿意度分布型態(比例)相同。在顯著水準 α = 0.05 的條件下，消費者對於 A、B 和 C 三家咖啡速食餐廳滿意度分布型態(比例)相同，彼此之間沒有顯著性地差異存在。

練習 12.10 現有一個骰子外表有 1~6 個點(面)，連續隨機擲 100 次，產生的點數敘述統計如下表所示。試以顯著水準 α = 0.05 檢定出現 1, 2, 3, 4, 5, 6 點機率是否相同？

X	1	2	3	4	5	6	合計
觀測值	19	14	22	17	8	20	100

題解：

A.設定顯著水準 α = 0.05，$\chi^2_{\alpha,(c-1)} = \chi^2_{0.05,(6-1)} = \chi^2_{0.05,5}$ = 11.0705[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 出現 1, 2, 3, 4, 5, 6 點機率相同。

C.對立假設(alternative hypothesis) H_1: 出現 1, 2, 3, 4, 5, 6 點機率不相同。

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = 7.6400$

X	1	2	3	4	5	6	合計
觀測值	19	14	22	17	8	20	100
期望值	16.6667	16.6667	16.6667	16.6667	16.6667	16.6667	100.0000
$\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$	0.3267	0.4267	1.7067	0.0067	4.5067	0.6667	7.6400

期望值 $= \frac{\text{樣本數量}}{6\text{ 面}} = \frac{100}{6} = 16.6667$(在一個公平的骰子中，六面，每一面出現機率會相等。100 次的投擲中，每一面出現的平均次數 16.6667 即是期望值)；自由度 $df = c - 1 = 6 - 1 = 5$

E.檢定統計值 $\chi^2 = 7.6400 < $ 臨界值$\chi^2_{\alpha,(c-1)} = 11.0705$，檢定統計值位於虛無假設接受區域內，接受虛無假設(null hypothesis) H_0: 出現 1, 2, 3, 4, 5, 6 點機率相同。在顯著水準 $\alpha = 0.05$ 的條件下，投擲該骰子出現 1, 2, 3, 4, 5, 6 點機率相同，彼此之間沒有顯著性地差異存在。

練習 12.11 旗山現有 A、B 和 C 三家咖啡速食餐廳，欲瞭解消費者對三家是否具有一樣的滿意度分布型態(比例)。於上週在三家咖啡速食餐廳中，分別隨機抽取數百位消費者進行滿意度調查，三家咖啡速食餐廳的結果如下表所示。試評估 A、B 和 C 三家咖啡速食餐廳的滿意度分布型態(比例)是否相同？(顯著水準 $\alpha = 0.05$)

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
A	80	80	50	70	50	330
B	32	52	45	76	55	260
C	41	45	38	71	95	290
合計	153	177	133	217	200	880

題解：

A.設定顯著水準 $\alpha = 0.05$，臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = \chi^2_{0.05,(3-1)\times(5-1)} = \chi^2_{0.05,8} = 15.5073$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: A、B 和 C 三家咖啡速食餐廳滿意度分布型態(比例)相同。

C.對立假設(alternative hypothesis) H_1: A、B 和 C 三家咖啡速食餐廳滿意度至少有兩家分布型態(比例)不相同。

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
A	80	80	50	70	50	330
B	32	52	45	76	55	260
C	41	45	38	71	95	290
合計	153	177	133	217	200	880

期望次數表：利用期望值 $e_{ij} = \frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}$，計算獲得如下表

餐廳	非常滿意	滿意	普通	不滿意	非常不滿意	合計
A	57.375	66.375	49.875	81.375	75.000	330
B	45.205	52.295	39.295	64.114	59.091	260
C	50.420	58.330	43.830	71.511	65.909	290
合計	153	177	133	217	200	880

檢定統計值 $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \frac{(80-57.375)^2}{57.375} + \frac{(80-66.375)^2}{66.375} + \frac{(50-49.875)^2}{49.875} + \frac{(70-81.375)^2}{81.375} + \frac{(50-75.000)^2}{75.000} + \frac{(32-45.205)^2}{45.205} + \frac{(52-52.295)^2}{52.295} + \frac{(45-39.295)^2}{39.295} + \frac{(76-64.114)^2}{64.114} + \frac{(55-59.091)^2}{59.091} + \frac{(41-50.420)^2}{50.420} + \frac{(45-58.330)^2}{58.330} + \frac{(38-43.830)^2}{43.830} + \frac{(71-71.511)^2}{71.511} + \frac{(95-65.909)^2}{65.909} = 47.2415$

E.檢定統計值 $\chi^2 = 47.2415$ > 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)} = 15.5073$，檢定統計值不位於虛無假設接受區域內，接受對立假設(alternative hypothesis) H_1: A、B 和 C 三家咖啡速食餐廳滿意度至少有兩家分布型態(比例)不相同。在顯著水準 $\alpha = 0.05$ 的條件下，消費者對於 A、B 和 C 三家咖啡速食餐廳滿意度分布型態(比例)不全部相同，彼此之間有顯著性地差異存在。

討論議題

1.非同步教學學習者之間議題討論：獨立性檢定應用

第一回合請於 D+3 日中午 12 點以前從「議題討論」區【張貼】標題：「應用情境」，本文：請針對第 12 章學習的【獨立性檢定】單元課程內容，透過老師線上講解、上課練習與平常考之後，對【獨立性檢定】的意涵更加了解。請具體陳述現在或未來最想運用到的情境(20 個字以上詮釋)。期望可以透過學習運用效益的交流，相互激勵，提升學習效益。

待有 25 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，請詳細檢視其他同學的回應內容。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳詮釋」，本文：哪一位同學詮釋得最好，其理由(25 個字以上詮釋)。透過同學之間的討論分享，可以提升學習效益。加油。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

2.非同步教學學習者之間議題討論：適合度檢定

第一回合請於 D+3 日中午 12 點以前從「議題討論」區【張貼】標題：「應用情境」，本文：請針對第 12 章學習的【適合度檢定】單元課程內容，透過老師線上講解、上課練習與平常考之後，對【適合度檢定】的意涵更加了解。請具體陳述您現在或未來最想運用到的情境(10 個字以上)，需要指定是哪一種分布。期望可以透過學習運用效益的交流，相互激勵，提升學習效益。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，請詳細檢視其他同學的回應內容。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳詮釋」，本文：哪一位同學詮釋得最好，其理由(10 個字以上)。透過同學之間的討論分享，可以提升學習效益。加油。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

3.非同步教學學習者之間議題討論：齊一性檢定

第一回合請於 D+3 日中午 12 點以前從「議題討論」區【張貼】標題：「應用情境」，本文：請針對第 12 章學習的【齊一性檢定】單元課程內容，透過老師線上講解、上課練習與平常考之後，對【齊一性檢定】的意涵更加了解。請具體陳述您現在或未來最想運用到的情境(10 個字以上)。期望可以透過學習運用效益的交流，相互激勵，提升學習效益。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，請詳細檢視其他同學的回應內容。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳詮釋」，本文：哪一位同學詮釋得最好，其理由(10 個字以上)。透過同學之間的討論分享，可以提升學習效益。加油。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

4.師生在非同步教學討論議題：D 日閉書平常考心得分享

第一回合請於 D 日早上 11：10 以前從「議題討論」區【張貼】標題：「準備心得」，本文：在此之前所有考試、評量和練習都是開書方式進行，在 D 日閉書複習平常考後，分析一下自己的應考心得與未來精進作為。分享一下自己比較得心應手的單元項目，與其解題的技巧，自己比較生疏的單元，未來的精進作為(35 個字以上)。期望可以透過學習經驗的交流，相互激勵，提升學習效益。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，檢視其他同學的回應內容。自己靜下心來，思考一下。第二回合【張貼】標題：「最大收獲」，本文：透過此議題討論給自己最大的收穫，請具體詮釋(35 個字以上)。透過同學之間的討論分享，可以提升學習效益。加油。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

重點整理

Excel 函數彙整

Excel 函數	統計功能	輸入資料	輸出資料
COUNTIF	計算符合條件的儲存格數量	區間、條件("<"&D5)	符合條件的儲存格數量

多項類別變數(多項式)之母體的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合特定多項式機率分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定多項式機率分布。

D.計算**檢定統計值**：卡方值

$$\text{卡方值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)

k：類別的數量

$v = k - 1$：自由度

E.若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,k-1}$，接受虛無假設(null hypothesis) H_0。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,k-1}$，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1。

獨立性檢定假設：

虛無假設(null hypothesis) H_0: A 變數與 B 變數為相互獨立

對立假設(alternative hypothesis) H_1: A 變數與 B 變數為非相互獨立

$$\text{檢定統計值：卡方值 } \chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

其中 f_{ij}：在列聯表(contingency table)中，第 i 列及第 j 行(欄)的觀察次數(frequency)

$e_{ij} = \frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}$，在列聯表(contingency table)中，第 i 列及第 j 行(欄)的期望次數(expected frequency)

n：在列聯表中，列的數量

m：在列聯表中，行(欄)的數量

在列聯表中若所有方格的期望次數皆大於或等於 5($e_{ij} \geq 5$)，此一統計值符合自由度 $v = (n-1)\times(m-1)$ 的卡方分布。

若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,(n-1)\times(m-1)}$，接受虛無假設(null hypothesis) H_0。

若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,(n-1)\times(m-1)}$，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1。

卜瓦松分布母體的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合卜瓦松分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合卜瓦松分布。

D.計算檢定統計值：卡方值

$$\text{卡方值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)

k：類別的數量

$v = k - p - 1$：自由度

p：欲利用樣本資料估算的母體參數(parameter)之個數

E.若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,k-p-1}$，接受虛無假設(null hypothesis) H_0: 母體特定變數符合卜瓦松分布。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,k-p-1}$，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1: 母體特定變數未符合卜瓦松分布。

常態分布母體的適合度檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 母體特定變數符合特定平均值和標準(偏)差的常態分布。

C.對立假設(alternative hypothesis) H_1: 母體特定變數未符合特定平均值和標準(偏)差的常態分布。

D.計算檢定統計值：卡方值

$$\text{卡方值 } \chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

其中 f_i：類別 i 的出現次數(frequency)

e_i：虛無假設成立時，類別 i 的期望次數(expected frequency)

k：類別的數量

$v = k - p - 1$：自由度

p：欲利用樣本資料估算的母體參數(parameter)之個數：平均值和標準(偏)差，故 $p = 2$

E.若檢定統計值 $\chi^2 \leq$ 臨界值$\chi^2_{\alpha,k-p-1}$，接受虛無假設(null hypothesis) H_0。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,k-p-1}$，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1。

齊一性檢定程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: 不同母體特定變數的分布相同。

C.對立假設(alternative hypothesis) H_1: 不同母體特定變數的分布不相同。

D.計算檢定統計值：卡方值

$$\text{卡方值 } \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

其中 o_{ij}：在列聯表中第 i 列和第 j 行的樣本觀察次數(observation)

e_{ij}：在列聯表中第 i 列和第 j 行的期望次數(expected frequency)

r：橫列的數量(row)、母體數量

c：縱行的數量(column)、等級數量、類別區分等級數量

$df = (r - 1) \times (c - 1)$：自由度

E.若檢定統計值 $\chi^2 \le$ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$，接受虛無假設(null hypothesis) H_0。

F.若檢定統計值 $\chi^2 >$ 臨界值$\chi^2_{\alpha,(r-1)\times(c-1)}$，拒絕虛無假設(null hypothesis) H_0，接受對立假設(alternative hypothesis) H_1。

關鍵詞彙解釋

適合度檢定(Goodness of Fit)

又稱為配合度檢定。為利用隨機獲得的樣本資料檢驗母體分布是否符合特定分布或理論方布的一種程序和統計方法。

獨立性檢定(test of independence)

獨立性檢定(test of independence)或列聯表檢定是欲檢定兩個研究變數(屬性)之間的關係是否獨立之統計方法。

齊一性檢定(Test of homogeneity)

齊一性檢定或均質性檢定(Test of homogeneity)為檢定兩個或兩個以上不同的母體，共有的特定變數(特性或屬性)是否具有相同機率分布或相同的比例之檢定方式。

列聯表(contingency table)

以列表方式呈現兩個或兩個以上變數或屬性共同出現的頻率分布。或者是將兩個變數或屬性的不同數值區間置於行和列的位置，在表格中填入兩個變數組合數值的出現頻率的表格。

補充教材

Phi 相關係數 ϕ

Phi coefficient 是在透過卡方檢定推論拒絕虛無假設：兩個變數之間屬於獨立之後，代表 2 個二項(二元)變數之間無法排除具有相關性的情況下，評量 2 個二項(二元)變數之間的相關程度。

若有兩個二項變數 X 和 Y 可以分別標示為

頻率 f_{ij}	X_0	X_1	合計
Y_0	A	B	A+B
Y_1	C	D	C+D
合計	A+C	B+D	n = A+B+C+D

其中，A、B、C 和 D 分別代表其出現的頻率(次數)。

Phi 相關係數的基本概念：X 和 Y 兩個二元變數觀察值的出現頻率，若大部分位於 2×2 列聯表「對角線」(從左上到右下線)欄位，亦即若觀察值的出現頻率大部分位於 A = (X_0,Y_0)和 D = (X_1,Y_1)兩個欄位，X 和 Y 兩個變數代表正向相關。反之，若 X 和 Y 兩個二元變數觀察值的出現頻率大部分位於「非對角線」

(非從左上到右下線)欄位，亦即若觀察值的出現頻率大部分位於 B = (X_1,Y_0)和 C = (X_0,Y_1)兩個欄位，X 和 Y 兩個變數代表負向相關。

$$\text{Phi 相關係數 } \phi = \frac{\mathrm{A}\times\mathrm{D} - \mathrm{B}\times\mathrm{C}}{\sqrt{(\mathrm{A+B})\times(\mathrm{C+D})\times(\mathrm{A+C})\times(\mathrm{B+D})}}$$

Phi 相關係數 ϕ 與卡方值 χ^2 的關係如下：

$$\phi^2 = \frac{\chi^2}{\mathrm{A+B+C+D}} = \frac{\chi^2}{n} \text{ 或 } \chi^2 = \phi^2 \times (\mathrm{A+B+C+D}) = \phi^2 \times n = \frac{(\mathrm{A}\times\mathrm{D} - \mathrm{B}\times\mathrm{C})^2\times n}{(\mathrm{A+B})\times(\mathrm{C+D})\times(\mathrm{A+C})\times(\mathrm{B+D})} = \frac{[(\mathrm{A}\times\mathrm{D})^2 - 2\times(\mathrm{A}\times\mathrm{D})\times(\mathrm{B}\times\mathrm{C})+(\mathrm{B}\times\mathrm{C})^2]\times n}{(\mathrm{A+B})\times(\mathrm{C+D})\times(\mathrm{A+C})\times(\mathrm{B+D})}$$

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

$$\text{卡方值 } \chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

期望值 e_{ij}	X_0	X_1	合計
Y_0	$\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}$	$\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}$	A+B
Y_1	$\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}$	$\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}$	C+D
合計	A+C	B+D	n = A+B+C+D

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{\left[\mathrm{A}-\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}\right]^2}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}} + \frac{\left[\mathrm{B}-\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}\right]^2}{\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}} + \frac{\left[\mathrm{C}-\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}\right]^2}{\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}} + \frac{\left[\mathrm{D}-\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}\right]^2}{\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}} =$$

$$\frac{\left[\mathrm{A}-\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}\right]^2\times(\mathrm{C+D})\times(\mathrm{B+D})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}\times(\mathrm{C+D})\times(\mathrm{B+D})} + \frac{\left[\mathrm{B}-\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}\right]^2\times(\mathrm{C+D})\times(\mathrm{A+C})}{\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}\times(\mathrm{C+D})\times(\mathrm{A+C})} + \frac{\left[\mathrm{C}-\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}\right]^2\times(\mathrm{A+B})\times(\mathrm{B+D})}{\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}\times(\mathrm{A+B})\times(\mathrm{B+D})} +$$

$$\frac{\left[\mathrm{D}-\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}\right]^2\times(\mathrm{A+B})\times(\mathrm{A+C})}{\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}\times(\mathrm{A+B})\times(\mathrm{A+C})} = \frac{\left[\mathrm{A}-\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}\right]^2\times(\mathrm{C+D})\times(\mathrm{B+D})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} + \frac{\left[\mathrm{B}-\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}\right]^2\times(\mathrm{C+D})\times(\mathrm{A+C})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\left[\mathrm{C}-\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}\right]^2\times(\mathrm{A+B})\times(\mathrm{B+D})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} + \frac{\left[\mathrm{D}-\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}\right]^2\times(\mathrm{A+B})\times(\mathrm{A+C})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} = \frac{\left[\mathrm{A}^2-2\times\mathrm{A}\times\frac{(\mathrm{A+B})\times(\mathrm{A+C})}{n}+\frac{(\mathrm{A+B})^2\times(\mathrm{A+C})^2}{n^2}\right]\times(\mathrm{C+D})\times(\mathrm{B+D})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\left[\mathrm{B}^2-2\times\mathrm{B}\times\frac{(\mathrm{A+B})\times(\mathrm{B+D})}{n}+\frac{(\mathrm{A+B})^2\times(\mathrm{B+D})^2}{n^2}\right]\times(\mathrm{C+D})\times(\mathrm{A+C})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} + \frac{\left[\mathrm{C}^2-2\times\mathrm{C}\times\frac{(\mathrm{C+D})\times(\mathrm{A+C})}{n}+\frac{(\mathrm{C+D})^2\times(\mathrm{A+C})^2}{n^2}\right]\times(\mathrm{A+B})\times(\mathrm{B+D})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\left[\mathrm{D}^2-2\times\mathrm{D}\times\frac{(\mathrm{C+D})\times(\mathrm{B+D})}{n}+\frac{(\mathrm{C+D})^2\times(\mathrm{B+D})^2}{n^2}\right]\times(\mathrm{A+B})\times(\mathrm{A+C})}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} =$$

$$\frac{\mathrm{A}^2\times(\mathrm{C+D})\times(\mathrm{B+D})-2\times\mathrm{A}\times\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}+\frac{(\mathrm{A+B})^2\times(\mathrm{A+C})^2\times(\mathrm{C+D})\times(\mathrm{B+D})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\mathrm{B}^2\times(\mathrm{C+D})\times(\mathrm{A+C})-2\times\mathrm{B}\times\frac{(\mathrm{A+B})\times(\mathrm{B+D})\times(\mathrm{C+D})\times(\mathrm{A+C})}{n}+\frac{(\mathrm{A+B})^2\times(\mathrm{B+D})^2\times(\mathrm{C+D})\times(\mathrm{A+C})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\mathrm{C}^2\times(\mathrm{A+B})\times(\mathrm{B+D})-2\times\mathrm{C}\times\frac{(\mathrm{C+D})\times(\mathrm{A+C})\times(\mathrm{A+B})\times(\mathrm{B+D})}{n}+\frac{(\mathrm{C+D})^2\times(\mathrm{A+C})^2\times(\mathrm{A+B})\times(\mathrm{B+D})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\mathrm{D}^2\times(\mathrm{A+B})\times(\mathrm{A+C})-2\times\mathrm{D}\times\frac{(\mathrm{C+D})\times(\mathrm{B+D})\times(\mathrm{A+B})\times(\mathrm{A+C})}{n}+\frac{(\mathrm{C+D})^2\times(\mathrm{B+D})^2\times(\mathrm{A+B})\times(\mathrm{A+C})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} =$$

$$\frac{\frac{\mathrm{A}^2\times(\mathrm{C+D})\times(\mathrm{B+D})\times n^2}{n^2}-\frac{2\times\mathrm{A}\times(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})\times n}{n^2}+\frac{(\mathrm{A+B})^2\times(\mathrm{A+C})^2\times(\mathrm{C+D})\times(\mathrm{B+D})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\frac{\mathrm{B}^2\times(\mathrm{C+D})\times(\mathrm{A+C})\times n^2}{n^2}-\frac{2\times\mathrm{B}\times(\mathrm{A+B})\times(\mathrm{B+D})\times(\mathrm{C+D})\times(\mathrm{A+C})\times n}{n^2}+\frac{(\mathrm{A+B})^2\times(\mathrm{B+D})^2\times(\mathrm{C+D})\times(\mathrm{A+C})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\frac{\mathrm{C}^2\times(\mathrm{A+B})\times(\mathrm{B+D})\times n^2}{n^2}-\frac{2\times\mathrm{C}\times(\mathrm{C+D})\times(\mathrm{A+C})\times(\mathrm{A+B})\times(\mathrm{B+D})\times n}{n^2}+\frac{(\mathrm{C+D})^2\times(\mathrm{A+C})^2\times(\mathrm{A+B})\times(\mathrm{B+D})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} +$$

$$\frac{\frac{\mathrm{D}^2\times(\mathrm{A+B})\times(\mathrm{A+C})\times n^2}{n^2}-\frac{2\times\mathrm{D}\times(\mathrm{C+D})\times(\mathrm{B+D})\times(\mathrm{A+B})\times(\mathrm{A+C})\times n}{n^2}+\frac{(\mathrm{C+D})^2\times(\mathrm{B+D})^2\times(\mathrm{A+B})\times(\mathrm{A+C})}{n^2}}{\frac{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}} =$$

$$\frac{\frac{\mathrm{A}^2\times(\mathrm{C+D})\times(\mathrm{B+D})\times n^2}{n}-\frac{2\times\mathrm{A}\times(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})\times n}{n}+\frac{(\mathrm{A+B})^2\times(\mathrm{A+C})^2\times(\mathrm{C+D})\times(\mathrm{B+D})}{n}}{(\mathrm{A+B})\times(\mathrm{A+C})\times(\mathrm{C+D})\times(\mathrm{B+D})} +$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}-\frac{2\times B\times(A+B)\times(B+D)\times(C+D)\times(A+C)\times n}{n}+\frac{(A+B)^2\times(B+D)^2\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}-\frac{2\times C\times(C+D)\times(A+C)\times(A+B)\times(B+D)\times n}{n}+\frac{(C+D)^2\times(A+C)^2\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}-\frac{2\times D\times(C+D)\times(B+D)\times(A+B)\times(A+C)\times n}{n}+\frac{(C+D)^2\times(B+D)^2\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=$$

$$\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times(A+B+C+D)\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times n}{n}+\frac{(A+B)^2\times(A+C)^2\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{(A+B)^2\times(B+D)^2\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{(C+D)^2\times(A+C)^2\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{(C+D)^2\times(B+D)^2\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times n\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times n}{n}+\frac{(A+B)^2\times(A+C)^2\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{(A+B)^2\times(B+D)^2\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{(C+D)^2\times(A+C)^2\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{(C+D)^2\times(B+D)^2\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times n^2}{n}+\frac{(A+B)^2\times(A+C)^2\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{(A+B)^2\times(B+D)^2\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{(C+D)^2\times(A+C)^2\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{(C+D)^2\times(B+D)^2\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times n^2}{n}+\frac{(A+B)\times(A+C)\times(A+B)\times(A+C)\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}$$

$$+\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{(A+B)\times(B+D)\times(A+B)\times(B+D)\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{(C+D)\times(A+C)\times(C+D)\times(A+C)\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{(C+D)\times(B+D)\times(C+D)\times(B+D)\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=$$

$$\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times n^2}{n}+\frac{\left(A^2+AB+AC+BC\right)\times(A+B)\times(A+C)\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{\left(B^2+AB+AD+BD\right)\times(A+B)\times(B+D)\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{\left(C^2+AD+AC+DC\right)\times(C+D)\times(A+C)\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{\left(D^2+BC+BD+CD\right)\times(B+D)\times(C+D)\times(B+D)\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=$$

$$\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times(A+B+C+D)^2}{n}+\frac{\left(A^2+AB+AC+BC\right)\times(A+B)\times(A+C)\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{\left(B^2+AB+AD+BD\right)\times(A+B)\times(B+D)\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{\left(C^2+AD+AC+DC\right)\times(C+D)\times(A+C)\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{\left(D^2+BC+BD+CD\right)\times(B+D)\times(C+D)\times(B+D)\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=$$

$$\frac{\frac{A^2\times(C+D)\times(B+D)\times n^2}{n}-\frac{2\times(A+B)\times(A+C)\times(C+D)\times(B+D)\times\left(A^2+B^2+C^2+D^2+2AB+2AC+2AD+2BC+2BD+2CD\right)}{n}+\frac{\left(A^2+AB+AC+BC\right)\times(A+B)\times(A+C)\times(C+D)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{B^2\times(C+D)\times(A+C)\times n^2}{n}+\frac{\left(B^2+AB+AD+BD\right)\times(A+B)\times(B+D)\times(C+D)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+\frac{\frac{C^2\times(A+B)\times(B+D)\times n^2}{n}+\frac{\left(C^2+AD+AC+DC\right)\times(C+D)\times(A+C)\times(A+B)\times(B+D)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}+$$

$$\frac{\frac{D^2\times(A+B)\times(A+C)\times n^2}{n}+\frac{\left(D^2+BC+BD+CD\right)\times(B+D)\times(C+D)\times(B+D)\times(A+B)\times(A+C)}{n}}{(A+B)\times(A+C)\times(C+D)\times(B+D)}=\text{待續}$$

範例 12.7　欲調查高雄深水大學餐旅管理系學生擁有中餐丙級證照和飲料調製丙級證照的狀況，隨機抽查數位學生資料分布如下表所示：請計算 Phi 相關係數 ϕ 和卡方值 χ^2。

觀測人數 f_{ij}	中餐丙級無	中餐丙級有	合計
飲料調製丙級無	30	20	50
飲料調製丙級有	10	80	90
合計	40	100	140

題解：期望次數表：利用期望值 $e_{ij} = \frac{(第i列和)\times(第j行和)}{樣本數量}$，計算獲得

期望人數 e_{ij}	中餐丙級無	中餐丙級有	合計
飲料調製丙級無	14.2857	35.7143	50
飲料調製丙級有	25.7143	64.2857	90
合計	40	100	140

A.設定顯著水準 $\alpha = 0.05$，$n = 2$ 列數，$m = 2$ 行數(欄數)，自由度 $v = (n-1)\times(m-1) = (2-1)\times(2-1) = 1$，臨界值 $\chi^2_{\alpha,(n-1)\times(m-1)} = \chi^2_{0.05,(2-1)\times(2-1)} = \chi^2_{0.05,1} = 3.8415$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 學生擁有中餐丙級和飲料調製丙級證照相互獨立

C.對立假設(alternative hypothesis) H_1: 學生擁有中餐丙級和飲料調製丙級證照並非相互獨立

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(30-14.2857)^2}{14.2857} + \frac{(20-35.7143)^2}{35.7143} + \frac{(10-25.7143)^2}{25.7143} + \frac{(80-64.2857)^2}{64.2857} = 37.6444$

E.檢定統計值 $\chi^2 = 37.6444 > $ 臨界值 $\chi^2_{\alpha,(n-1)\times(m-1)} = \chi^2_{0.05,(2-1)\times(2-1)} = \chi^2_{0.05,1} = 3.8415$，檢定統計值位於虛無假設拒絕區域內，接受對立假設(alternative hypothesis) H_1: 學生擁有中餐丙級和飲料調製丙級證照並非相互獨立。

Phi 相關係數 $\phi = \frac{A\times D - B\times C}{\sqrt{(A+B)\times(C+D)\times(A+C)\times(B+D)}} = \frac{30\times 80 - 20\times 10}{\sqrt{(30+20)\times(10+80)\times(30+10)\times(20+80)}} = \frac{2400 - 200}{\sqrt{50\times 90\times 40\times 100}} = \frac{2200}{\sqrt{18000000}} = \frac{2200}{4242.6407} = 0.5185$

卡方值 $\chi^2 = \phi^2 \times n = 0.5185^2 \times 140 = 0.2689 \times 140 = 37.6444$

答案：Phi 相關係數 $\phi = 0.5185$；卡方值 $\chi^2 = 37.6444$

列聯係數

Contingency coefficient $C = \sqrt{\frac{\chi^2}{n+\chi^2}} = \sqrt{\frac{\phi^2}{1+\phi^2}}$，因為 $\phi^2 = \frac{\chi^2}{n}$，其中 n 是全部樣本數

列聯係數(Contingency coefficient or coefficient of contingency)是使用於透過卡方檢定推論拒絕虛無假設：兩個變數之間屬於獨立之後，代表 2 個名目變數(nominal variable)之間無法排除具有相關性的情況下，該 2 個名目變數可能選項數量相等(列聯表中欄和列的數量相同，例如：3×3 或 4×4)同時可能選項數量是 3 或 3 個以上時，評量 2 個名目變數之間的相關程度。

當兩個變數沒有相關時，Contingency coefficient $C = 0$。Contingency coefficient 不會有負值的情況。另，當兩個變數之間有完美的相關時，Contingency coefficient C 會趨近於 1。Contingency coefficient C 不會超過 1，Contingency coefficient C 呈現的區間範圍比 ϕ 更受限。故，Contingency coefficient C 數值區間介於 0 到 1 之間。

範例 12.8　　欲調查高雄深水高中三個年級學生家庭經濟的分布狀況，隨機抽查數位學生資料分布如下表所示：請計算列聯係數。

觀測人數 f_{ij}	一年級	二年級	三年級	合計
低收入	30	20	20	70
中收入	50	20	50	120
高收入	10	80	80	170
合計	90	120	150	360

題解：期望次數表：利用期望值 $e_{ij} = \frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}$，計算獲得

期望人數 e_{ij}	一年級	二年級	三年級	合計
低收入	17.5000	23.3333	29.1667	70
中收入	30.0000	40.0000	50.0000	120
高收入	42.5000	56.6667	70.8333	170
合計	90	120	150	360

A.設定顯著水準 $\alpha = 0.05$，$n = 3$ 列數，$m = 3$ 行數(欄數)，自由度 $v = (n-1)\times(m-1) = (3-1)\times(3-1) = 4$，臨界值$\chi^2_{\alpha,(n-1)\times(m-1)} = \chi^2_{0.05,(3-1)\times(3-1)} = \chi^2_{0.05,4} = 9.4877$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 學生年級和經濟狀況相互獨立

C.對立假設(alternative hypothesis) H_1: 學生年級和經濟狀況並非相互獨立

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(30-17.5000)^2}{17.5000} + \frac{(20-23.3333)^2}{23.3333} + \frac{(20-29.1667)^2}{29.1667} + \frac{(50-30.0000)^2}{30.0000} + \frac{(20-40.0000)^2}{40.0000} + \frac{(50-50.0000)^2}{50.0000} + \frac{(10-42.5000)^2}{42.5000} + \frac{(80-56.6667)^2}{56.6667} + \frac{(80-70.8333)^2}{70.8333} = 71.2661$

E.檢定統計值 $\chi^2 = 71.2661 > $ 臨界值$\chi^2_{\alpha,(n-1)\times(m-1)} = \chi^2_{0.05,(3-1)\times(3-1)} = \chi^2_{0.05,4} = 9.4877$，檢定統計值位於虛無假設拒絕區域內，接受對立假設(alternative hypothesis) H_1: 學生年級和經濟狀況並非相互獨立。

$$\text{Contingency coefficient } C = \sqrt{\frac{\chi^2}{n+\chi^2}} = \sqrt{\frac{71.2661}{360+71.2661}} = 0.4065$$

答案：列聯係數 = 0.4065

Cramer’s V 係數

Cramer’s V 係數是依據 Phi 係數進行修正，適用於大於 2×2 表格。透過卡方檢定推論拒絕虛無假設：兩個變數之間屬於獨立之後，代表 2 個名目變數(nominal variable)之間無法排除具有相關性的情況下，該 2 個名目變數可能選項數量不相等(列聯表中欄和列的數量不相同，例如：2×3 或 3×4)，評量 2 個名目變數之間的相關程度。

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{n\times(k-1)}}$$

其中，k 代表行數或列數數值較小的數值。

範例 12.9 欲調查高雄深水大學學生性別與家庭經濟的分布狀況，隨機抽查數位學生資料分布如下表所示：請計算 Cramer’s V。

觀測人數 f_{ij}	男生	女生	合計
低收入	30	20	50
中收入	50	20	70
高收入	10	80	90
合計	90	120	210

題解：期望次數表：利用期望值 $e_{ij} = \frac{(\text{第}i\text{列和})\times(\text{第}j\text{行和})}{\text{樣本數量}}$，計算獲得

期望人數 e_{ij}	男生	女生	合計
低收入	21.4286	28.5714	50
中收入	30.0000	40.0000	70
高收入	38.5714	51.4286	90
合計	90	120	210

A.設定顯著水準 $\alpha = 0.05$，$n = 3$ 列數，$m = 2$ 行數(欄數)，自由度 $v = (n-1)\times(m-1) = (3-1)\times(2-1) = 2$，臨界值$\chi^2_{\alpha,(n-1)\times(m-1)} = \chi^2_{0.05,(3-1)\times(2-1)} = \chi^2_{0.05,2} = 5.9915$[使用 Excel 軟體 CHISQ.INV.RT(右尾機率)函數查詢獲得]。

B.虛無假設(null hypothesis) H_0: 學生性別和經濟狀況相互獨立

C.對立假設(alternative hypothesis) H_1: 學生性別和經濟狀況並非相互獨立

D.計算檢定統計值：卡方值

檢定統計值 $\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(30-21.4286)^2}{21.4286} + \frac{(20-28.5714)^2}{28.5714} + \frac{(50-30.0000)^2}{30.0000} + \frac{(20-40.0000)^2}{40.0000} + \frac{(10-38.5714)^2}{38.5714} + \frac{(51-51.4286)^2}{51.4286} = 66.3704$

E.檢定統計值 $\chi^2 = 66.3704 > $ 臨界值$\chi^2_{\alpha,(n-1)\times(m-1)} = 5.9915$，檢定統計值位於虛無假設拒絕區域內，接受對立假設(alternative hypothesis) H_1: 學生性別和經濟狀況並非相互獨立。

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{n\times(k-1)}} = \sqrt{\frac{66.3704}{210\times(2-1)}} = 0.5622$$

答案：Cramer's V = 0.5622

葉氏校正 Yate's correction

當樣本數量不夠大時，需要使用葉氏(連續化)校正(Yate's continuity correction or Yate's correction for continuity)方式，以觀測次數 f_{ij} 與期望次數 e_{ij} 之差的絕對值減去 0.5 後平方再除以期望次數 e_{ij} 之全部和，獲得校正的卡方值

$$\text{檢定統計值 } \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(|f_{ij}-e_{ij}|-0.5)^2}{e_{ij}}$$

費雪精確性檢定(Fisher exact test)

使用於樣本數量太少($n < 20$)

McNemar 檢定(McNemar's test)