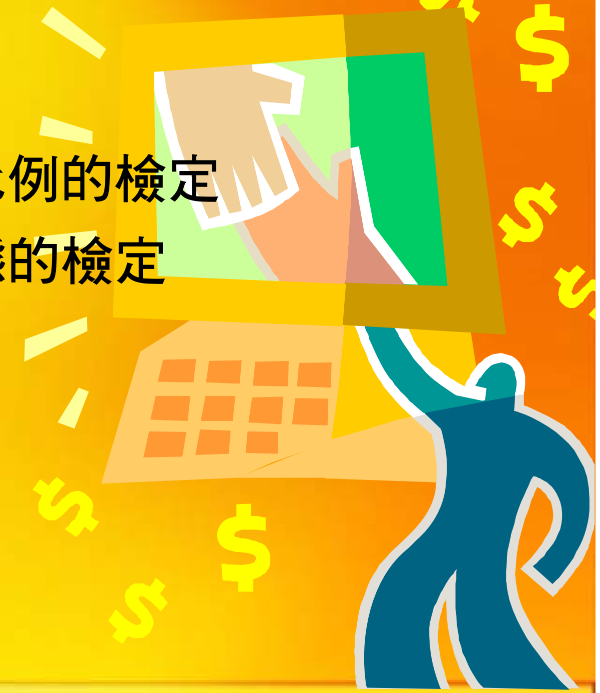
統計學

第十二章 點計(類別)資料分析－卡方檢定的應用



書號：512282

編著 江建良

普林斯頓國際有限公司

12-1 基本概念

類別資料概述：

類別資料 (categorical data)，是指根據觀察值 (即個體) 的屬性，按照不同類別加以分類 (分組)，而各分類 (分組) 的個數乃依點計方式計數而得，故又稱為點計資料 (count data)。

類別資料的種類：

單分類列聯表：

單分類列聯表 (one-way contingency table)，係指統計資料只依單一標準加以分類 。

雙分類列聯表：

雙分類列聯表 (two-way contingency table)，係指統計資料依兩標準加以分類 。

12-1 基本概念

- **卡方檢定的模式：**

類別	1	2		k	合計
觀察次數 o_i	o_1	o_2	…	o_k	n
H_0 下之期望次數 e_i	e_1	e_2	…	e_k	n

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \cdots + \frac{(o_k - e_k)^2}{e_k} = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

當 $X^2 > X^2_{(1-\alpha\,,f)}$ 時，表示樣本觀察次數與期望次數差異大，即樣本資料與 H_0 下之母體資料差異大，故放棄 H_0 之假設。由於放棄域列於 X^2 分配的右尾，所以一律採右尾檢定。

12-1 基本概念

- **常見的卡方檢定類型：**

(1) 多項式母體比例的檢定。
(2) 母體分配型態的檢定。
適合度檢定。

(3) 獨立性檢定。
(4) 齊一性檢定。
列聯表檢定。

12-2 適合度檢定(一) — 多項式母體比例的檢定

例題 12-1

設甲工廠生產的產品分 A、B、C 三級，所占之比例為 2：1：1，今隨機抽取 8 個，其中 A 級 5 個，B 級 2 個，C 級 1 個之機率為何？

解： $P_1=\dfrac{1}{2}, P_2=\dfrac{1}{4}, P_3=\dfrac{1}{4}$

$$P=C^{8}_{5,2,1}\left(\frac{1}{2}\right)^5\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^1=\frac{8!}{5!2!1!}\left(\frac{1}{2}\right)^5\left(\frac{1}{4}\right)^2\left(\frac{1}{4}\right)^1=\frac{21}{256}$$

12-2 適合度檢定(一) － 多項式母體比例的檢定

卡方檢定的步驟：

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

① 若 X^2 值落在放棄域內，則差異顯著，放棄 H_0，接受 H_1。

② 若 X^2 值落在接受域內，則差異不顯著，接受 H_0。

例題 12-2

擲一個銅板 50 次，其資料如下：

銅　板	正　面	反　面
次　數	30	20

試檢定出現正反面的機率是否相等？($\alpha = 0.05$)

解：(a) 假設：$\begin{cases} H_0 : P_1 = P_2 = \dfrac{1}{2} \\ H_1 : \text{至少有一個不相等} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{(0.95\,,\,2-1)} = 3.84$

(d) 計算：$e_1 = nP_1 = 50 \times \dfrac{1}{2} = 25$

$$e_2 = nP_2 = 50 \times \frac{1}{2} = 25$$

$$\chi^2 = \sum_{i=1}^{2} \frac{(o_i - e_i)^2}{e_i} = \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} = 2$$

(e) 判斷：$\because 2 < 3.84$，差異不顯著，$\therefore$ 接受 H_0，即該銅板爲一公正的銅板，正反面出現的機率均爲 $\dfrac{1}{2}$。

例題 12-3

依孟德爾研究豌豆的外型，得出圓而黃、圓而綠、皺而黃、皺而綠之遺傳比例為 9：3：3：1，茲觀察 200 棵豌豆外型，其資料如下：

豌　豆	圓而黃	圓而綠	皺而黃	皺而綠
數　目	110	35	40	15

試檢定此豌豆實驗是否符合遺傳理論？($\alpha = 0.05$)

解：(a) 假設：
$$\begin{cases} H_0 : P_1 = \frac{9}{16}, P_2 = \frac{3}{16}, P_3 = \frac{3}{16}, P_4 = \frac{1}{16} \\ H_1 : \text{至少有一個不相等} \end{cases}$$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{(0.95,4-1)} = 7.81$

(d) 計算：$e_1 = nP_1 = 200 \times \frac{9}{16} = 112.5$

$$e_2 = nP_2 = 200 \times \frac{3}{16} = 37.5$$

$$e_3 = nP_3 = 200 \times \frac{3}{16} = 37.5$$

$$e_4 = nP_4 = 200 \times \frac{1}{16} = 12.5$$

$$\chi^2 = \sum_{k=1}^{4} \frac{(o_i - e_i)^2}{e_i} = \frac{(110-112.5)^2}{112.5} + \frac{(35-37.5)^2}{37.5}$$

$$+ \frac{(40-37.5)^2}{37.5} + \frac{(15-12.5)^2}{12.5} = 0.89$$

(e) 判斷：$\because 0.89 < 7.81$，落在接受域，差異不顯著，$\therefore$ 接受 H_0，即該實驗符合遺傳理論。

12-3 適合度檢定(二) － 母體分配型態的檢定

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

① 若 X^2 值落在放棄域，則差異顯著，放棄 H_0，接受 H_1。

② 若 X^2 值不落在放棄域，則差異不顯著，接受 H_0。

例題 12-4

擲一個骰子 100 次，其出現點數次數之資料如下：

點　數	1	2	3	4	5	6
次　數	20	24	10	15	14	17

試檢定上述資料是否服從於分立均等分配？($\alpha = 0.05$)

解： (a) 假設：$\begin{cases} H_0: \text{母體服從於分立均等分配} \\ \quad (\text{即 } P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = \frac{1}{6}) \\ H_1: \text{母體不服從於分立均等分配} (P_i \text{ 不全相等}) \end{cases}$

(b) 顯著水準： $\alpha = 0.05$

(c) 放棄域： $\chi^2 > \chi^2_{(1-\alpha\,,\,k-1)} = \chi^2_{(0.95\,,\,5)} = 11.07$

(d) 計算： $e_1 = nP_1 = 100 \times \frac{1}{6} = 16.67 \; (i = 1, 2, \cdots, 6)$

$$\chi^2 = \sum_{i=1}^{6} \frac{(o_i - e_i)^2}{e_i} = \frac{(20-16.67)^2}{16.67} + \frac{(24-16.67)^2}{16.67} + \frac{(10-16.67)^2}{16.67}$$

$$+ \frac{(15-16.67)^2}{16.67} + \frac{(14-16.67)^2}{16.67} + \frac{(17-16.67)^2}{16.67} = 7.16$$

(e) 判斷：$\because 7.16 < 11.07$，落在接受域，差異不顯著，$\therefore$ 接受 H_0，即資料分配服從於分立均等分配，該骰子爲公正的骰子。

例題 12-5

設有一副牌以放回方式抽出三張牌，每次出現紅桃之張數為 X，重複試行 64 次之結果如下：

X	0	1	2	3
次數 (o_i)	21	31	12	0

試檢定上述資料是否符合 $n=3, p=0.25$ 之二項分配？$(\alpha = 0.05)$

解：(a) 假設：$\begin{cases} H_0：\text{服從於二項分配}\ (n=3\ ,\ P=0.25) \\ H_1：\text{不服從於二項分配}\ (n=3\ ,\ P=0.25) \end{cases}$

(b) 顯著水準：$\alpha = 0.01$

(c) 放棄域：$\chi^2 > \chi^2_{(1-\alpha\ ,\ k-1)} = \chi^2_{(0.99\ ,\ 3-1)} = 9.21$【註：$f = 3-1$，見下列說明】

(d) 計算：

① $x=0$ 時，$P_1 = C_0^3\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^3 = 0.4219$

$$e_1 = nP_1 = 64(0.4219) = 27$$

② $x=1$ 時，$P_2 = C_1^3\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2 = 0.4219$

$$e_2 = nP_2 = 64(0.4219) = 27$$

③ $x=2$ 時，$P_3 = C_2^3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1 = 0.1406$

$$e_3 = nP_3 = 64(0.1406) = 9$$

④ $x=3$ 時，$P_4 = C_3^3\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^0 = 0.0156$

$$e_4 = nP_4 = 64(0.0156) = 1$$

$\because e_4 = 1 < 5$，$\therefore$ 必須與鄰近一組合併，共 3 組，

$\therefore 3-1=2$，如下表：

X	0	1	2	3
o_i	21	31	12	0
			12	
e_i	27	27	9	1
			10	

$$\chi^2 = \sum_{i=0}^{3}\frac{(o_i-e_i)^2}{e_i} = \frac{(21-27)^2}{27} + \frac{(31-27)^2}{27} + \frac{(12-10)^2}{10} = 2.33$$

(e) 判斷：$\because 2.33 < 9.21$，落在接受域，$\therefore$ 接受 H_0，即該資料服從於二項分配 $(n=3\,,\,P=0.25)$。

例題 12-6

觀察台中市每日發生機車事故的次數，以 50 天為一期，其記載之資料如下：

事故次數	0	1	2	3	4	5	6 或以上
天數	25	15	8	0	1	1	0

試檢定資料是否適合卜瓦松分配？$(\alpha = 0.05)$

解：(a) 假設：$\begin{cases} H_0：此資料適合卜瓦松分配 \\ H_1：此資料不適合卜瓦松分配 \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{(0.95,\,1)} = 3.84$ （$f = 1$，見以下說明 ）

(d) 計算：卜瓦松分配爲 $P(x)=\dfrac{\mu^x e^{-\mu}}{x!}$ ，$\because \mu$ 未知，以 $\bar{x}$ 代替，

$$\bar{x}=\frac{(0)(25)+(1)(15)+(2)(8)+(3)(0)+(4)(1)+(5)(1)}{50}=0.8$$

$$\therefore P(x)=\frac{(0.8)^x e^{-0.8}}{x!}$$

① $P_1=P(x=0)=\dfrac{(0.8)^0 e^{-0.8}}{0!}=0.4493$ （ 查表 ）

$$e_1=nP_1=50(0.4493)=22.47$$

② $P_2=P(x=1)=\dfrac{(0.8)^1 e^{-0.8}}{1!}=0.3595$

$$e_2=nP_2=50(0.3595)=17.97$$

③ $P_3 = P(x=2) = \dfrac{(0.8)^2 e^{-0.8}}{2!} = 0.1438$

$e_3 = nP_3 = 50(0.1438) = 7.19$

④ $P_4 = P(x=3) = \dfrac{(0.8)^3 e^{-0.8}}{3!} = 0.0383$

$e_4 = nP_4 = 50(0.0383) = 1.92$

⑤ $P_5 = P(x=4) = \dfrac{(0.8)^4 e^{-0.8}}{4!} = 0.0077$

$e_5 = nP_5 = 50(0.0077) = 0.39$

⑥ $P_6 = P(x=5) = \dfrac{(0.8)^5 e^{-0.8}}{5!} = 0.0012$

$e_6 = nP_6 = 50(0.0012) = 0.06$

⑦ x 爲 6 或以上，機率爲 0.0002

$e_7 = nP_7 = 50(0.0002) = 0.00$

$\because e_4$ 、 e_5 皆小於 5，與鄰近合併一組，如下表：

X	0	1	2	3	4	5	6 或以上
o_i	25	15	8	0	1	1	0
					10		
e_i	22.47	17.97	7.19	1.92	0.39	0.06	0.00
					9.56		

由上表知共有 3 組，即 $k=3$，又 $\because$ 以一個估計統計量 $\bar{x}$ 代替 μ

$\therefore f=k-1-m=3-1-1=1$

$$\chi^2=\sum_{x=0}^{2}\frac{(o_i-e_i)^2}{e_i}=\frac{(25-22.47)^2}{22.47}+\frac{(15-17.97)^2}{17.97}+\frac{(10-9.56)^2}{9.56}=0.796$$

(e) 判斷：$\because 0.796<3.84$，不落在放棄域，$\therefore$ 接受 H_0，即此資料適合於卜瓦松分配。

例題 12-7

某單位進行一項 300 人的大規模測驗，其成績如下：

成　績	人　數
50 以下	0
50~60	24
60~70	64
70~80	120
80~90	73
90~100	19
100 以上	0

該單位宣稱此資料適合於 $\mu=75, \sigma=15$ 的常態分配，試檢定之 $(\alpha=0.05)$。

解：(a) 假設：$\begin{cases} H_0：此資料適合於\ \mu = 75,\sigma = 15\ 的常態分配 \\ H_1：此資料不適合於\ \mu = 75,\sigma = 15\ 的常態分配 \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{(0.95,6)} = 12.59\ (f = k - 1 = 7 - 1 = 6)$

(d) 計算：

①

成績 (X)	人數 (o_i)	$Z = \dfrac{\ell - 75}{15}$	機率 (P_i)	期望次數 $e_i = 300 \times P_i$
50 以下	0	-1.67 以下	0.0475	14.25
50~60	24	-1.67~-1.00	0.1112	33.36
60~70	64	-1.00~-0.33	0.2120	63.60
70~80	120	-0.33~0.33	0.2586	77.58
80~90	73	0.33~1.00	0.2120	63.60
90~100	19	1.00~1.67	0.1112	33.36
100 以上	0	1.67 以上	0.0475	14.25

(i) $Z=\dfrac{50-75}{15}=-1.67$

(ii) $Z=\dfrac{60-75}{15}=-1.00$

(iii) $P(50<X<60)=P(Z<-1.00)-P(Z<-1.67)$

$=0.1587-0.0457$

$=0.1112\cdots$ 其他類推

② $\chi^2=\sum_{i=1}^{7}\dfrac{(o_i-e_i)^2}{e_i}=\dfrac{(0-14.25)^2}{14.25}+\dfrac{(24-33.36)^2}{33.36}+\cdots+\dfrac{(0-14.25)^2}{14.25}$

$=61.89$

(e) 判斷：$\because 61.89>12.59$，落在放棄域，$\therefore$ 放棄 H_0，即測驗成績並不服從於 $\mu=75,\sigma=15$ 的常態分配。

例題 12-8

設 90 名學生之數學成績如下 (設已知該資料 $\overline{X}=39$, $\hat{S}=14$)：

組別（分數）	人數 (o_i)
$X<20$	10
$20\le X<30$	14
$30\le X<40$	21
$40\le X<50$	20
$50\le X<60$	15
$60\le X<70$	10

試檢定此資料是否適合於常態分配？($\alpha=0.05$)

解：(a) 假設：$\begin{cases} H_0：此資料適合於常態分配 \\ H_1：此資料不適合於常態分配 \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{(1-\alpha\,,\,k-1-m)} = \chi^2_{(0.95\,,\,6-1-2)} = 7.81$

(d) 計算：由於 μ 與 σ 未知，利用樣本資料得知 $\overline{X} = 39\,,\hat{S} = 14$，代替未知母數，$\therefore m = 2$，自由度 $f = k-1-m = 6-1-2 = 3$。

①

成績 (X)	人數 (o_i)	$Z = \frac{\ell - 39}{14}$	機率 (P_i)	期望次數 $(e_i = 90 \times P_i)$
$X < 20$	10	-1.36 以下	0.0869	7.821
$20 \le X < 30$	14	-1.36~-0.64	0.1742	15.678
$30 \le X < 40$	21	-0.64~-0.07	0.2668	24.012
$40 \le X < 50$	20	0.07~0.79	0.2573	23.157
$50 \le X < 60$	15	0.79~1.50	0.1480	13.32
$60 \le X$	10	1.5 以上	0.0668	6.012

② $\chi^2 = \sum_{k=1}^{6} \frac{(o_i - e_i)^2}{e_i} = \frac{(10-7.821)^2}{7.821} + \frac{(14-15.678)^2}{15.678} + \frac{(21-24.012)^2}{24.012}$

$+ \frac{(20-23.157)^2}{23.157} + \frac{(15-13.32)^2}{13.32} + \frac{(10-6.012)^2}{6.012} = 4.452$

(e) 判斷：$\because 4.452 < 7.81$，落在接受域，差異不顯著，$\therefore$ 接受 H_0，即該資料適合於常態分配。

12-4 列聯表檢定(一) － 獨立性檢定

- **卡方檢定的步驟：**

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

① 若 X^2 值落在放棄域內，則差異顯著，放棄 H_0，接受 H_1。

② 若 X^2 值落在接受域，則差異不顯著，接受 H_0。

例題 12-9

設 50 名學生之學業與操行成績如下表所示：

學業＼操行	甲　等	乙　等
A　等	18	12
B　等	10	10

試檢定學業與操行成績是否有關？($\alpha = 0.05$)

解：(a) 假設：$\begin{cases} H_0: \text{學業與操行無關} \\ H_1: \text{學業與操行有關} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{[(1-\alpha),(r-1)(c-1)]} = \chi^2_{(0.95,1)} = 3.84$

(d) 計算：

學業＼操行	甲　等	乙　等	合　計
A　等	18(16.8)①	12(13.2)	30
B　等	10(11.2)	10(8.8)	20
合　計	28	22	50

① $\dfrac{28\times 30}{50}=16.8\,,\ \dfrac{22\times 30}{50}=13.2$，其他類推

② 不考慮修正，$\because e_{ij}$ 均未小於 5

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

$$=\frac{(18-16.8)^2}{16.8}+\frac{(12-13.2)^2}{13.2}+\frac{(10-11.2)^2}{11.2}+\frac{(10-8.8)^2}{8.8}$$

$$=0.487$$

③ 考慮修正：在實務，$5\le e_{ij}\le 10$，予以修正

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

$$=\frac{(|18-16.8|-0.5)^2}{16.8}+\frac{(|12-13.2|-0.5)^2}{13.2}+\frac{(|10-11.2|-0.5)^2}{11.2}$$

$$+\frac{(|10-8.8|-0.5)^2}{8.8}=0.166$$

(e) 判斷：$\because$ 不考慮修正或考慮修正，兩者之值均小於 3.84，落在接受域，$\therefore$ 接受 H_0，即學業與操行無關。

例題 12-10

研究著想了解教育程度與社經地位是否有關，乃調查 300 人，其資料如下：

教育程度＼社經地位	低	中	高
大　學	20	24	40
高　中	16	20	30
國　中	40	50	60

試檢定教育程度與社經地位是否有關？($\alpha = 0.05$)

解：(a) 假設：$\begin{cases} H_0：\text{教育程度與社經地位無關} \\ H_1：\text{教育程度與社經地位有關} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{[(1-\alpha),(r-1)(c-1)]} = \chi^2_{(0.95,4)} = 9.49$

(d) 計算：

教育程度＼社經地位	低	中	高	合　計
大　學	20(21.3)	24(26.3)	40(36.4)	84
高　中	16(16.7)	20(20.7)	30(28.6)	66
國　中	40(38)	50(47)	60(65)	150
合　計	76	94	130	300

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

$$= \frac{(20-21.3)^2}{21.3} + \frac{(24-26.3)^2}{26.3} + \cdots + \frac{(60-65)^2}{65} = 1.44$$

(e) 判斷：$\because 1.44 < 9.49$，落在接受域，差異不顯著，$\therefore$ 接受 H_0，即教育程度與社經地位無關。

例題 12-11

茲調查 500 人，詢問“婚前可否性行為”的看法，其資料如下：

性別 \ 態度	贊　成	反　對
男	100	125
女	140	135

試檢定性別不同對婚前性行為的看法是否相同？($\alpha = 0.05$)

(a) 卡方檢定。

(b) 將男、女性別視為二個母體，以雙尾常態分配檢定男性對“贊成”的看法之比例與女性對“贊成”的看法之比例是否相同？

解：(a) 卡方檢定：

① 假設：$\begin{cases} H_0：性別與態度無關 \\ H_1：性別與態度有關 \end{cases}$

② 顯著水準：$\alpha = 0.05$

③ 放棄域：$\chi^2 > \chi^2_{(0.95,1)} = 3.84$

④ 計算：

態度 / 性別	贊　成	反　對	合　計
男	100(108)	125(117)	225
女	140(132)	135(143)	275
合　計	240	260	500

$$\chi^2 = \frac{(100-108)^2}{108} + \frac{(125-117)^2}{117} + \frac{(140-132)^2}{132} + \frac{(135-143)^2}{143}$$
$$= 2.07$$

⑤ 判斷：$\because 2.07 < 3.84$，落在接受域，差異不顯著，$\therefore$ 接受 H_0，即性別與態度無關。

(b) 兩母體比例差之檢定（可參考 9-4 節之說明）：

① 假設：$\begin{cases} H_0: P_1 = P_2 \\ H_1: P_1 \neq P_2 \end{cases}$

② 顯著水準：$\alpha = 0.05$

③ 放棄域：$Z < -Z_{0.975} = -1.96$ 及 $Z > Z_{0.975} = 1.96$

④ 計算：

$$\hat{P}_1 = \frac{x_1}{n_1} = \frac{100}{225}$$

$$\hat{P}_2 = \frac{x_2}{n_2} = \frac{140}{275}$$

$$\hat{P} = \frac{100+140}{225+275} = \frac{240}{500}$$

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\frac{100}{225} - \frac{140}{275}}{\sqrt{\left(\frac{240}{500}\right)\left(1-\frac{240}{500}\right)\left(\frac{1}{225}+\frac{1}{275}\right)}} = -1.44$$

⑤ 判斷：$\because -1.44 > -1.96$，落在接受域，$\therefore$ 接受 H_0，與 (a) 之方法同。我們亦可由 $\chi^2_{(1)} = Z^2$ 性質知，$2.07 = (1.44)^2$，兩者相同。

12-5 列聯表檢定(二) － 齊一性檢定

獨立性檢定與齊一性檢定之比較：

1. 一般常將齊一性檢定視為獨立性檢定的延伸，其所使用的檢定統計量、公式、判斷準則皆相同。

2. 獨立性檢定只從一母體分配中抽取一組隨機樣本，而使用不同變數對該資料加以分類；齊一性檢定則可視為由不同母體分別各抽取一組樣本而相互比較。

3. 獨立性檢定是檢定不同變數的分類是否相互獨立；而齊一性檢定則是檢定不同的隨機樣本在各個母體的比例是否一致。

因為齊一性檢定與獨立性檢定所使用的檢定統計量相同，本書不另外說明檢定的步驟 。

例題 12-12

今自各行業各隨機抽樣 100 人，依性別不同區分之各行業人數如下：

性別＼行業	服務業	製造業	農　業	礦　業
男	40	70	65	80
女	60	30	35	20
合　計	100	100	100	100

試檢定四種行業就業性別的比例是否相同？$(\alpha = 0.05)$

解：(a) 假設：$\begin{cases} H_0: \text{四種行業就業性別比例相同 (或性別與從事行業無關)} \\ H_1: \text{四種行業就業性別比例不同 (或性別與從事行業有關)} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{[0.95,(4-1)(2-1)]} = \chi^2_{(0.95,3)} = 7.81$

(d) 計算：本題乃先確立各行業抽出 100 人，邊際次數 (行) 固定。我們可將男與女兩組視爲不同的隨機樣本，以檢定該兩組是否在四種不同行業之比例齊一。

性別＼行業	服務業	製造業	農　業	礦　業	合　計
男	40(63.8)	70(63.8)	65(63.8)	80(63.8)	255
女	60(36.3)	30(36.3)	35(36.3)	20(36.3)	145
合　計	100	100	100	100	400

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

$$= \frac{(40-63.8)^2}{63.8} + \frac{(70-63.8)^2}{63.8} + \cdots + \frac{(20-36.3)^2}{36.3} = 37.60$$

(e) 判斷：$\because 37.60 > 7.81$，落在放棄域，差異顯著，$\therefore$ 放棄 H_0，即表示四種行業之就業性別比例不同，換言之，性別與所從事的行業有關。

例題 12-13

今以 300 位家長，250 位教師，300 位心理學家，360 名學生，詢問“若學生成績退步，是否應受到老師處罰？”，其贊成、無意見、反對之資料如下：

態度 / 對象	贊　成	無意見	反　對	合　計
家　長	128	102	70	300
教　師	106	100	44	250
心理學家	54	120	126	300
學　生	38	110	212	360
合　計	326	432	452	1,210

試檢定調查對象對此問題贊成與否比例是否相同？($\alpha = 0.05$)

解：(a) 假設：$\begin{cases} H_0：\text{調查對象贊成比例相同} \\ H_1：\text{調查對象贊成比例不同} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$\chi^2 > \chi^2_{(0.95,6)} = 12.59$

(d) 計算：$f = (r-1)(c-1) = (3-1)(4-1) = 6$

態度 / 對象	贊　成	無意見	反　對	合　計
家　長	128(80.8)	102(107.1)	70(112.1)	300
教　師	106(67.4)	100(89.3)	44(93.4)	250
心理學家	54(80.8)	120(107.1)	126(112.1)	300
學　生	38(97.0)	110(128.5)	212(134.5)	360
合　計	326	432	452	1,210

$$\chi^2 = \frac{(128-80.8)^2}{80.8} + \frac{(102-107.1)^2}{107.1} + \cdots + \frac{(212-134.5)^2}{134.5}$$
$$= 188.58$$

(e) 判斷：$\because 188.58 > 12.59$，落在放棄域，差異顯著，$\therefore$ 放棄 H_0，即四種不同對象對此問題贊成比例不同。

12-6 K-S 檢定法(選讀)

適合度檢定 - K-S檢定法(單一樣本情況)：

$$D = \max | F(X) - S(X) |$$

式中：$F(X)$ 表示理論分配各階段的累積機率。

$S(X)$ 表示實際分配各階段的累積機率。

$|F(X)-S(X)|$ 表示各階段理論分配與實際分配累積機率之差的絕對值，而絕對值中的最大值即為 D。

1. 當 $D > D_{\alpha/2}$ 時，落在放棄域，差異顯著，放棄 H_0，即實際次數分配不服從於理論分配。
2. 當 $D < D_{\alpha/2}$ 時，落在放棄域，差異不顯著，放棄 H_0，即實際次數分配服從於理論分配。

例題 12-14

為了解“是否廢除大學聯考”，今隨機抽取 100 位家長調查，其中非常贊成 37 人，贊成 16 人，沒有意見 10 人，不贊成 12 人，非常不贊成 25 人，設 $\alpha = 0.05$ 下，家長態度有無顯著差異？

“是否廢除大學聯考”

非常不贊成	—	—	—	—	—	非常贊成
	1	2	3	4	5	

解： 本題之資料爲順序資料（非常贊成>贊成>沒有意見>不贊成>非常不贊成，5>4>3>2>1)，適用 K-S 檢定法。

(a) 假設：$\begin{cases} H_0: \text{態度無顯著差異} \\ H_1: \text{態度有顯著差異} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$D > D_{0.025} = 0.1340 \quad (n = 100)$

(d) 計算：$n = 100$ 人，分 5 個評等，理論上每個評等爲 $100 \times \frac{1}{5} = 20$（人），$S(X)$ 與 $F(X)$ 均爲“向下累加次數”的各階段機率。

評定等級	人數	$S(X)$	$F(X)$	$\|F(X) - S(X)\|$
1	25	25/100	20/100	5/100
2	12	37/100	40/100	3/100
3	10	47/100	60/100	13/100
4	16	63/100	80/100	$17/100 \leftarrow D$ (最大值)
5	37	100/100	100/100	0

$$D = \max|F(X) - S(X)| = \frac{17}{100} = 0.17$$

(e) 判斷：$\because 0.17 > 0.1340$，落在放棄域，差異顯著，$\therefore$ 放棄 H_0，即態度上有顯著差異。

例題 12-15

試以 K-S 檢定法檢定本章例題 12-4。

解：(a) 假設：$\begin{cases} H_0：母體服從於分立均等分配\left(P_i=\dfrac{1}{6}\right) \\ H_1：母體不服從於分立均等分配\left(P_i=\dfrac{1}{6}\right) \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 放棄域：$D > D_{0.025} = 0.1340 \quad (n = 100)$

(d) 計算：

點數	1	2	3	4	5	6
次數	20	24	10	15	14	17
$S(X)$	$\frac{20}{100}$	$\frac{44}{100}$	$\frac{54}{100}$	$\frac{69}{100}$	$\frac{83}{100}$	$\frac{100}{100}$
$F(X)$	$\frac{100\times\frac{1}{6}}{100}$	$\frac{100\times\frac{2}{6}}{100}$	$\frac{100\times\frac{3}{6}}{100}$	$\frac{100\times\frac{4}{6}}{100}$	$\frac{100\times\frac{5}{6}}{100}$	$\frac{100\times\frac{6}{6}}{100}$
$\lvert F(X)-S(X)\rvert$	$\frac{\frac{20}{6}}{100}$	$\frac{\frac{64}{6}}{100}$ $\uparrow$ D	$\frac{\frac{24}{6}}{100}$	$\frac{\frac{14}{6}}{100}$	$\frac{\frac{2}{6}}{100}$	0

$$D=\max\lvert F(X)-S(X)\rvert=\frac{\frac{64}{6}}{100}=0.107$$

(e) 判斷：$\because \frac{\frac{64}{6}}{100}=0.107<0.1340$，落在接受域，差異不顯著，$\therefore$ 接受 H_0，

即該資料服從於分立均等分配。

例題 12-16

試以 K-S 檢定法檢定本章例題 12-7。

解：(a) 假設：$\begin{cases} H_0：此資料適合\ \mu=75,\sigma=15\ 的常態分配 \\ H_1：此資料不適合\ \mu=75,\sigma=15\ 的常態分配 \end{cases}$

(b) 顯著水準：$\alpha=0.05$

(c) 放棄域：$D>D_{0.025}=\dfrac{1.36}{\sqrt{300}}=0.079$ （$n>100,\alpha=0.025$，代入 $\dfrac{1.36}{\sqrt{n}}$，查表 ）。

(d) 計算：

		$Z_0 = \dfrac{\ell - 75}{15}$	$S(X)$	$F(X) = P(Z<Z_0)$	$\lvert F(X) - S(X) \rvert$
50 以下	0	−1.67 以下	0	0.0475	0.0475
50~60	24	−1.00	0.0800	0.1587	0.0787
60~70	64	−0.33	0.2933	0.3707	0.0774
70~80	120	0.33	0.6933	0.6293	0.064
80~90	73	1.00	0.9367	0.8413	$0.0954 \rightarrow D$
90~100	19	1.67	1	1	0

$$D = \max \lvert F(X) - S(X) \rvert = 0.0954$$

(e) 判斷：$\because 0.0954 > 0.079$，落在放棄域，$\therefore$ 放棄 H_0，即測驗成績並不服從於常態分配。

12-6 K-S 檢定法(選讀)

獨立性或齊一性檢定 – K-S 檢定法 (兩種獨立樣本情況)：

1. 先計算兩組樣本各階段一累積機率 $S_1(X)$ 與 $S_2(X)$。
2. 計算各階段的累積機率之差為 $S_1(X) - S_2(X)$。
3. 取統計量 D_+ 或 D_-，或 D，如下：

$$D_+ = \max\,[S_1(X) - S_2(X)]\ \text{(右尾檢定)}$$

$$D_- = \max\,[S_2(X) - S_1(X)]\ \text{(左尾檢定)}$$

$$D\ = \max\,|S_1(X) - S_2(X)|\ \text{(雙尾檢定)}$$

① 當 $2 \le m \le n \le 12$ 或 $m + n \le 16$，由 K-S 兩樣本檢定表的 mnD 欄中，可查出機率值 P，當 $P \le \alpha$，則放棄 H_0。

② 當 $9 \le m = n \le 20$ 可查表在 α 下之臨界值 mnD 或 mnD_+ 或 mnD_-，作成結論。

③ 當 m , n 較大或 $m + n > 40$，可查另一附表，作成結論。

例題 12-17

下表是男生與女生各 10 名的統計學成績，則男女統計學成績是否有顯著差異？(設 $\alpha = 0.05$，以 K-S 檢定法檢定)。

分數	50~54	55~59	60~64	65~69	70~74	75~79	80~84
男生	0	2	1	4	1	2	0
女生	1	1	0	1	2	4	1

解：(a) 假設：$\begin{cases} H_0: \text{男女生統計學成績無顯著差異} \\ H_1: \text{男女生統計學成績有顯著差異} \end{cases}$

(b) 顯著水準：$\alpha = 0.05$

(c) 計算：

分　　數	50~54	55~59	60~64	65~69	70~74	75~79	80~84
男生 $S_1(X)$	0/10	2/10	3/10	7/10	8/10	1	1
女生 $S_2(X)$	1/10	2/10	2/10	3/10	5/10	9/10	1
$\lvert S_1(X) - S_2(X) \rvert$	1/10	0	1/10	4/10	3/10	1/10	0

$$D = |S_1(X) - S_2(X)| = \frac{4}{10} = 0.4$$

(e) 判斷：由於 $m = n = 10$，查表得 $mnD = \dfrac{70}{100} = 0.7$，$\because 0.4 < 0.7$，接受 H_0，即男女生統計學成績無顯著差異。

例題 12-18

承上例，若 50 名男生，60 女生，統計學成績如下表所示：

分數	50~54	55~59	60~64	65~69	70~74	75~79	80~84
男生	5	5	9	10	9	5	5
女生	5	4	6	5	13	9	18

則男女生統計學成績是否有顯著差異？(設 $\alpha = 0.05$，以 K-S 檢定法檢定)

解：(a) 假設：$\begin{cases} H_0：男女生統計學成績無顯著差異 \\ H_1：男女生統計學成績有顯著差異 \end{cases}$

(b) 顯著水準： $\alpha = 0.05$

(c) 計算：

分　數	50~54	55~59	60~64	65~69	70~74	75~79	80~84
男生 $S_1(X)$	5/50	12/50	21/50	31/50	40/50	45/50	1
女生 $S_2(X)$	5/60	9/60	15/60	20/60	33/60	42/60	1
$\lvert S_1(X) - S_2(X) \rvert$	0.017	0.090	0.17	0.287	0.25	0.2	0

$$D = \lvert S_1(X) - S_2(X) \rvert = 0.287$$

(e) 判斷：由於

$$D_\alpha = D_{0.05} = 1.36\sqrt{\frac{N(=m+n)}{mn}} = 1.36\sqrt{\frac{50+60}{(50)(60)}} = 0.26$$

$\because 0.287 > 0.26$ ， $\therefore$ 差異顯著，放棄 H_0 ，即男女生統計學成績有顯著差異。