第七章

卡方分布

（*Chi-square distribution*）

章前摘要

卡方分布（Chi-square distribution），在推論統計學中，其觀念並不難，但由於計算時算式較長，很容易造成計算上的失誤，由於卡方分布是考試中的熱門考題，故須注意。

重點提要

㈠卡方分布（Chi-square distribution）：

檢定的時候，當資料是屬於名目（nominal）時，而要檢驗一個自變項對應變項的效果為何，就需要使用到卡方分布。卡方分布大約是在1990年首先由Pearson提出，由常態分布中所變化出來的，卡方值就是標準常態分布變值 Z 的平方所得到，其公式如下：

$$Z^2 = \frac{(x-\mu)^2}{\sigma^2} \quad \text{or} \quad Z^2 = \frac{n(\overline{x}-\mu)^2}{\sigma^2}$$

上述公式中，樣品的均值為 $\overline{x}$ ，母群的平均值為 μ ，母群的變方為 σ^2 ，假若由常態分布母群裡面抽樣出 n 個樣本，並把每一個樣本 x_i ，帶入上述公式，並求其總和，可得到：

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^2} = \frac{\sum (x_i-\mu)^2}{\sigma^2}$$

上式Pearson稱自由度為 $df = n$ 的卡方值，其卡方值的公式可表示如下：

$$\chi^2_{(n)} = \frac{\sum (x_i-\mu)^2}{\sigma^2}$$

若是由 n 個樣品資料，可以得到自由度為 $(n-1)$ 的卡方值，其公式如下：

$$\chi^2_{(n-1)} = \sum Z_i^2 = \frac{\sum (x_i-\overline{x})^2}{\sigma^2}$$

因此可以說，卡方值為 Z 分數的平方和。

㈡獨立性檢定（test of independence）：

在同一個母群中兩個變數之間是不是獨立，就是兩變數之間的關聯性檢定（test of association），由下列的 $r \times c$ 列連表（contingency table）：

	1	2	⋯	c	總和
1	a_{11}	a_{12}	⋯	a_{1c}	R_1
2	a_{21}	a_{22}	⋯	a_{2c}	R_2
⋮	⋮	⋮	⋮	⋮	⋮

r	a_{r1}	a_{r2}	…	a_{rc}	R_r
總和	C_1	C_2	…	C_c	n

卡方值計算公式：

$$\chi^2=\sum\frac{(O-E)^2}{E}\text{ ， }\chi^2=n\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{a_{ij}^2}{R_iC_j}-1\right)$$

O：每一方格的觀測值

E：每一方格的期望次數

使用以上兩公式所計算後所得到的卡方值結果都是相同，這兩個公式可以互用。其自由度爲：$(r-1)\times(c-1)$，以及查詢卡方表後，便可得進行獨立性檢定。

㈢同質性檢定（test of homogeneity）：

不一樣的母群的同一變數是否一致時，可利用同質性檢定得知，其檢定方式與獨立性檢定（test of independence）類似。

㈣卡方分布的應用：

變方相等性檢定（test of equal variance）：利用Bartlett B 值計算公式，作爲變方相等性檢定，其公式如下：

$$B=\frac{\{N\ln S^2-\sum f_i\log S_i^2\}}{C}$$

$$N=\sum f_i$$

$$C=1+\frac{1}{3(k-1)}\left\{\sum_{i=1}^{k}\frac{1}{f_i}-\frac{1}{N}\right\}$$

上述公式中，N 爲總自由度，利用 B 值與查表後所獲得 χ_α^2 值，便可以檢定其兩族群或是兩者以上族群的變方是否相等。

㈤適合性檢定（test of goodness of fit）：

這個檢定主要的方式，在比較觀測次數與期望次數之間的差異，假使差異夠大，表示樣品所顯示的結果，足夠推翻虛無假設，其公式如下：

$$\chi_{(n-1)}^2=\sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i}$$

上式中，實測值爲 O_i，期望次數爲 E_i，自由度爲 $(n-1)$ 之卡方分布。

㈥McNemar改變檢定（McNemar change test）：

精選題庫

➲經典題型

範題 1

今天有三種治療骨質酥鬆症的藥劑，分別給受試者使用後其結果如下圖，

藥劑名稱	A	B	C
有改善（單位：人）	48	56	34
無改善（單位：人）	32	30	58

請問這三種治療骨質酥鬆症的藥劑有沒有差異？

【解析】

畫出列聯表：

藥劑名稱	A	B	C	總和
有改善（單位：人）	48	56	34	138
無改善（單位：人）	32	30	58	120
總和	80	86	92	258

藥劑A有改善的期望次數：$E=\dfrac{(138\times 80)}{258}=42.79$

藥劑B有改善的期望次數：$E=\dfrac{(138\times 86)}{258}=46$

藥劑C有改善的期望次數：$E=\dfrac{(138\times 92)}{258}=49.21$

藥劑A無改善的期望次數：$E=\dfrac{(120\times 80)}{258}=37.2$

藥劑B無改善的期望次數：$E=\dfrac{(120\times 86)}{258}=40$

藥劑C無改善的期望次數：$E=\dfrac{(120\times 92)}{258}=42.79$

計算卡方值：

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

$$= \frac{(48-42.79)^2}{42.79} + \frac{(56-46)^2}{46} + \frac{(34-49.21)^2}{49.21} + \frac{(32-37.2)^2}{37.2}$$

$$+ \frac{(30-40)^2}{40} + \frac{(58-42.79)^2}{42.79}$$

$$= 16.14$$

$\alpha = 0.05$，自由度為：$(3-1)\times(2-1)=2$，$\chi^2_\alpha = 5.991$

H_0：三種治療骨質酥鬆症的藥劑沒有差異

H_1：三種治療骨質酥鬆症的藥劑有差異

$\chi^2 > \chi^2_\alpha$，拒絕H_0，接受H_1，所以三種治療骨質酥鬆症的藥劑有差異，但是犯第一類型錯誤（type I error）的機率可能有5%。

範題 2

某一位醫生想了解血型B型、AB型與O型與偏頭痛程度（高、中、低）是否有關聯，利用工作時，經病患同意後取得病患的投醫資料，其資料如下表：

	B	AB	O
高	64	21	7
中	52	75	60
低	10	14	42

請問血型B型、AB型、O型與偏頭痛程度（高、中、低）有沒有關連？

【解析】

畫出列聯表：

	B	AB	O	總和
高	64	21	7	92
中	52	75	60	187
低	10	14	42	66
總和	126	110	109	345

計算卡方值：

範題 12

卡方檢定是用來處理哪種類型的資料？雖然卡方檢定的定義公式只有一種，但是卻有不同的用途，請舉例卡方檢定的多種用法？

【解析】

卡方檢定主要是用於等距變項或是比例變項的資料。

㈠配適度檢定（goodness of fit test）：

卡方檢定可用於檢定對某件事物的機率分布是否是真還是不真，這個檢定就稱作是配適度檢定。例如，新開發的農藥殺蟲效果，是不是與藥商所說的符合。

㈡獨立性檢定（test of independence）：

卡方檢定可以用於檢定同一個母群中的兩個變數之間，彼此是不是無關、是否獨立，這就稱作是獨立性檢定。例如，男女性別的差異，與看事物看法的觀點是否獨立。

㈢同質性檢定（test of homogeneity）：

卡方檢定可用於檢定不同的樣品資料，是不是都來自同一個母群，此種卡方檢定，就稱作是同質性檢定。例如，三種不同廠牌的維骨力，對於治療退化性關節炎的效果是否相同。

㈣改變的顯著性檢定（test of significance of change）：

二樣品資料的取得，二者之間彼此是具有連帶關係，並不是獨立取得，假如要比較測驗此二樣品資料是否有差異，就稱為改變的顯著性檢定。

➲歷屆試題

範題 1

某一教學醫院中所有15-45歲被診斷為原發性血栓症的婦女，依據其所使用的避孕方法被分類如下，請問血栓症婦女所使用的各種避孕方法是否不同？（ $\alpha = 0.05$ ）列出虛無及對立假設，並進行檢定。

避孕方法	婦女人數
口服避孕藥	40
子宮內避孕器	25
輸卵管結紮	20
無	15
	100

（84年普考）

【解析】

此為卡方分布，計算其卡方值：

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

$$E = \frac{(40+25+20+15)}{4} = 25$$

$$\chi^2 = \frac{(40-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(15-25)^2}{25} = 14$$

自由度為：$n-1=4-1=3$，$\alpha = 0.05$，查表可得：$\chi^2_\alpha = 7.815$

H_0：表示相同

H_1：表示不同

$\chi^2 > \chi^2_\alpha$，拒絕H_0，接受H_1，故對於血栓塞婦女而言，所使用的各種避孕方法皆不同。

範題2

在1995年夏季之二個午後，德基水庫的五個水深之含氧量（ppm）如下表，試檢驗含量是否因日期而有顯著區別？

日期	水深（m）				
	0	2	4	6	8
7月1日	4.8	4.6	4.6	4.4	4.0
7月5日	5.1	5.0	4.8	4.4	3.9

（85年普考）

【解析】

使用卡方分布，計算其χ^2值

列出列聯表

日期	水深（m）					總和
	0	2	4	6	8	
7月1日	4.8	4.6	4.6	4.4	4.0	22.4
7月5日	5.1	5.0	4.8	4.4	3.9	23.2
總和	9.9	9.6	9.4	8.8	7.9	45.6

計算每個觀察次數對應的期望次數：

日期為7月1日且水深為0（m）的期望次數：$E=\dfrac{(22.4\times 9.9)}{45.6}=4.863$

日期為7月1日且水深為2（m）的期望次數：$E=\dfrac{(22.4\times 9.6)}{45.6}=4.716$

日期為7月1日且水深為4（m）的期望次數：$E=\dfrac{(22.4\times 9.4)}{45.6}=4.617$

日期為7月1日且水深為6（m）的期望次數：$E=\dfrac{(22.4\times 8.8)}{45.6}=4.323$

日期為7月1日且水深為8（m）的期望次數：$E=\dfrac{(22.4\times 7.9)}{45.6}=3.881$

日期為7月5日且水深為0（m）的期望次數：$E=\dfrac{(23.2\times 9.9)}{45.6}=5.037$

日期為7月5日且水深為2（m）的期望次數：$E=\dfrac{(23.2\times 9.6)}{45.6}=4.884$

日期為7月5日且水深為4（m）的期望次數：$E=\dfrac{(23.2\times 9.4)}{45.6}=4.782$

日期為7月5日且水深為6（m）的期望次數：$E=\dfrac{(23.2\times 8.8)}{45.6}=4.477$

日期為7月5日且水深為8（m）的期望次數：$E=\dfrac{(23.2\times 7.9)}{45.6}=4.019$

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

$$=\frac{(4.8-4.863)^2}{4.863}+\frac{(4.6-4.716)^2}{4.716}+\frac{(4.6-4.617)^2}{4.617}+\frac{(4.4-4.323)^2}{4.323}$$

$$+\frac{(4-3.881)^2}{3.881}+\frac{(5.1-5.037)^2}{5.037}+\frac{(5-4.884)^2}{4.884}+\frac{(4.8-4.782)^2}{4.782}$$