的標準也一再降低，依美國職業安全衛生局（NIOSH,National Institute for Occupational Safety and Health ）原訂暴露標準為1ppm，在1992年更修訂為0.75ppm，更需提醒工作人員在使用此消毒劑時多加注意，不只注意自己安全，處理過器械用物亦要沖洗或排氣完全才可給病人使用。

## 參考文獻


1.Favero MS: Dialysis associated disease and their control. In:Benett JV d Brochman, PS(eds):Hospital Infection, 2nd ed. Boston, Little Brown, 1986:267－282

2.Rutala WA: APIC Guideline for selection and use of disinfectants. Am J Infect Control 1990;18:99－117

3.盧光舜：消毒學（再版）台北南山堂出版社，1985:37－88


流行病學暨生物統計專欄(九)

# 類別變項的統計方法


林明瀅

台北榮民總醫院感染管制委員會


## 前　言

本期介紹的統計方法為二個類別變項間的統計方法，簡單的説類別變項就是資料非連續的，例如在臨床上我們在紀錄有多少人使用呼吸器，資料只會有7人使用呼吸器，而不會有7.8人使用呼吸器。而類別變項的統計方法就如探討住院病人有無使用導尿管、與得到院內泌尿道感染是否有關係，此時使用導尿管屬於類別變項，其資料只有『有』或『沒有』二類，而院內泌尿道感染也屬於類別變項，其資料是『已經得到院內泌尿道感染』或『沒有得到院內泌尿道感染』，即變項的分類不同，只代表兩者不一樣，且彼此間沒有次序間的差異，探討這二個條件間是否有關聯存在，又稱關聯檢定或獨立性檢定，在臨床上常用的統計方法有 1.卡方檢定［Chi－square($x^2$)test］ 2.Yate's correct test 3.費歇恰當檢定Fisher's exact test) 4. McNemar test 5.波以松檢定(Poisson test）。

## 壹、卡方檢定（$x^2$－test）

卡方檢定（$x^2$－test，或稱為Chi－square "Chi"讀音與中文『開』同）是最常被廣泛應用的統計方法之一，卻也是最常被誤用與濫用的統計方法，其目的在於檢定在A變項的分類中屬某一項，是否在B變項的分類中較易屬於某一項，即說A，B二個類別變項有關聯性，或此二個類別變項的檢定具有統計上的意義，反之，則此二個變項沒有關聯性，或不具統計學上的意義。

一、適用條件：

1.類別變項—類別變項

2.兩組資料互相獨立

3.樣本是從母群體中隨機抽樣

4.類別變項，A變項可區分為c組（c＞＝2），B變項可分為r組（r＞＝2）

5.其資料可安排成r x c的表格

二、假設檢定：有無曝露於某危險因子的得病機率相同

三、檢定程序：

1.檢定每一細格之期望值是否過小（＜5）？

2.檢定有無曝露於某危險因子的得病機率是否相同？

四、檢定統計量：

| | | |
|---|---|---|
| a | b | a＋b |
| c | d | c＋d |
| a＋c | b＋d | n |

a,b,c,d為4個細格的數值

N為總樣本數

1)$X^2=\Sigma((0-E)/E)^2$或

2)自由度＝1時，

$$X^2=\frac{N(ad-bd)^2}{(a+b)(c+d)(a+c)(b+d)}$$

五、範例：某一研究者，欲探討有無使用導尿管與院內泌尿道感染是否有關聯，其資料如下：

| | | 使用導尿管 | | 合　計 |
|---|---|---|---|---|
| | | 是 | 否 | |
| 得泌尿道感染 | 是 | 3 | 16 | 19 |
| | 否 | 40 | 58 | 98 |
| 合　計 | | 43 | 74 | 117 |

計算過程

1)Ho: 有無使用導尿管與得泌尿道感染無關

H1:使用導尿管與泌尿道感染有關

P1＝P2,P1＝使用者生病的機率,P2＝未使用者生病的機率

2)計算每個格子的期望值

$E11=19/117\times43/17\times117=6.98$

$E12=19/117\times74/17\times117=12.02$

$E21=98/117\times43/17\times117=36.02$

$E22=98/117\times74/17\times117=61.98$

E11：代表使用導尿管並得到泌尿道感染。

E12：沒有使用導尿管並得到泌尿道感染。

E21：使用導尿管且沒有得到泌尿道感染。

E22：沒有使用導尿管且沒有得到泌尿道感染。

是否使用導尿管與得到泌尿道感染觀察值與預期值。

| | | 是否使用導尿管 | | 合　計 |
|---|---|---|---|---|
| | | 是 | 否 | |
| 患　病 | 是 | 3(6.98) | 16(12.02) | 19 |
| | 否 | 40(36.02) | 58(61.98) | 98 |
| 合　計 | | 43 | 74 | 117 |

3)計算$X^2$值［$X^2=\Sigma((0-E)E)^2$］

$$X^2=\frac{(3-6.98)^2}{6.98}+\frac{(16-12.02)^2}{12.02}+$$

$$\frac{(40-36.02)^2}{36.02}+\frac{(58-61.98)^2}{61.98}$$

$$=4.28$$

六、檢定結果

$X^295(1)=3.84$，$X^299(1)=6.63$　因此當$X^2=4.28$時，P在0.05與0.01之間，即$0.01<P<0.05$，如果選擇$\alpha=0.05$，就是拒絕Ho，而相信使用導尿管與泌尿道感染有關聯；若選擇$\alpha=0.01$，則接受無差異之假設，就認為使用導尿管與泌尿道感染無關。

七、決策原則：

1. 統計學家曾建議，當$N<20$，或$20<N<40$而其中有超過20%的期望次數小於5時，不要使用$X^2$檢定。
2. 當我們計算$X^2$值以後，並不需由研究者去計算其機率，只要查由數理統計家所計算$X^2$的機率表即可。在查表之前先要計算自由度，如範例$X^2$值大於3.84，$P<0.05$，小於6.63，$p>0.01$。$X^2$值愈大，表示觀察值與預期值相差愈大，兩者可能從同一母全體中抽出之機率也愈小。
3. 多項類別變項的卡方檢定其計算方法與二項類別變項相同，只是自由度會因變項增加而增大。
4. 計算自由度＝（A變項之項數－1）x（B變項之項數－1）。$X^2$表中最左邊之一縱列代表自由度，最上一橫列代表面積或機率，二者之交叉點即代表臨界值。

卡方分佈表

| 自由度\臨界值 | 95% | 99% |
|---|---|---|
| 1 | 3.84 | 6.63 |
| 2 | 5.99 | 9.21 |
| 3 | 7.82 | 11.35 |
| 4 | 9.49 | 13.28 |
| 5 | 11.07 | 15.09 |
| 6 | 12.59 | 16.81 |
| 7 | 14.07 | 18.47 |
| 8 | 15.51 | 10.09 |
| 9 | 16.92 | 21.67 |

5. 卡方檢定本身是在探究二個類別類項間是否『相關』。如果檢定結果『顯著』或『推翻無差異的假設』，則到底是A變項引起B變項或B變項引起A變項，或互為因果，或只是具統計上的關連但完全沒有因果關係，則需由研究者依研究法上的邏輯加以推定。亦即卡方檢定本身不假設因果關係的方向。
6. $X^2$值受總樣本數影響甚大，當每個格次的百分比保持恆定，而增大總樣本數時，檢定結果可能就從不顯著轉變為顯著。因此如何決定適當的樣本數是卡方檢定中最難的課題。

八、卡方檢定的濫用

卡方檢定計算方便，瞭解容易，是檢定二個類別變項間關係的好方法，但也常遭到不當的濫用。最常見的濫用是任意將等距或等比變項化為類別變項，然後以卡方檢定處理之。甚至也有將二個等距／等比變項均化為類別變項，做卡方檢定。此種做法犯了二個錯誤：一是

如果是等距／等比資料，因其資料含有的訊息最豐富，將其分為若干組，則丟棄了很多資料，迫使利用敏感度較低的統計方法，易趨向於接受無效的假設。分組不同，卡方檢定的結果就不相同，可能在一種分組方法下是接受假設而另一種分組法卻是推翻假設。

## 貳、Yate's corrected $X^2$ test

一、適用條件：1. 在2x2的卡方檢定中（亦即d.f.＝1），最小的期望值小於5時，$X^2$值分佈與$X^2$值機率分佈表並不完全符合，而有稍微的誤差，對誤差的修正稱為Yate's correction。亦即自由度為1時，

$$X^2=\frac{(|0-E|-0.5)^2}{E}$$ 做了修正$X^2$值比不修正小了一點，使之更接近自由度＝1的$X^2$分佈。

2. 在多項類別的卡方檢定，若總數界於20至40之間時，且有超過20%的期望值小於5時，一定要使用Yate's correction。

二、假設檢定：有無曝露於某危險因子的得病機率相同

三、檢定程序：

1. 檢定每一細格之期望值是否過小（＜5）？

2. 檢定有無曝露於某危險因子的得病機率是否相同？

四、檢定統計量：

1) $$X^2=\frac{(|0-E|-0.5)^2}{E}$$

2) 當自由度為1時，

$$X^2=\frac{N(|ad-bd|-N/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

五、範例：某研究者研究26位男性冠狀心臟病患者，其插心導管與院內心內膜炎的關係，其資料如下

| | | 使用心導管 | | 合　計 |
|---|---|---|---|---|
| | | 是 | 否 | |
| 得心內膜炎 | 是 | 13 | 1 | 14 |
| | 否 | 5 | 7 | 12 |
| 合計 | | 18 | 8 | 26 |

計算過程（代入公式）

$$X^2=\frac{26x(|13x7-5x1|-26/2)^2}{14x12x18x8}$$
$$=5.73$$

六、檢定結果

$X^2 95(1)=3.84$，$X^2 99(1)=6.63$因此當$X^2=5.73$時，P在0.05與0.01之間，即$0.01<P<0.05$，如果研究者選擇$\alpha=0.05$，就是拒絕Ho，而相信使用心導管與心內膜炎有關聯；若選擇$\alpha=0.01$，則接受無差異之假設，即使心導管與心內膜炎無關。

七、決策原則：

1. 校正後的值將比未校正的值小，亦即較不容易拒絕Ho，是比較保守的做法但卻能得到較可靠的結論；一般$X^2$值若在推翻與接受的臨界邊緣上，修正後則為『接受無效假設』。

2. 理論上說，自由度為1時應一律做Yate's修正，但如O－E的絕對值很大，則減或不減0.5對$X^2$並不產生影響。

（待續）