

感染率與感染密度的統計檢定

林明滢

台北榮民總醫院 感染管制室

要證明感染率或感染密度在正常可接受範圍，不能憑個人的感覺，必須要有評斷的方法，比較感染率在可接受的範圍內或已下降時，採用『卡方統計檢定』，而對感染密度則是採用『二個樣本比率差別統計檢定』，本文將說明此二統計方法的適用條件，計算公式、於 Excel 設定公式之方法及統計檢定結果的解讀。希望藉由簡單的公式設定，使大家於平日的工作上可以更得心應手。

前 言

醫院評鑑條文 3.7.4.2，為各醫院必評項目，此項目若未符合 C 以上的標準，醫院評鑑將被認定為『未通過』；其內容為『定期對院內感染的發生及其動向，開會檢討分析，並訂定改善方案-C 的第 3 項規定：感染管制指標能回饋相關醫護人員，以致力於改善措施。A 的標準：根據修訂之方案，改善院內感染，著有成效』，要了解院內感染的動向，一般呈現的指標是採用感染率(感染人次/入(出)院人數 $\times 100\%$)或感染密度(感染人次/住院人日數)，尤其是入(出)院人數較少的病房單位或長期照護機構要呈現感染指標時，選用感染密度會較感染率合適。要證明感染率或感染密度在正常可接受範圍，或是經由院內各單位努力進行改善方案後感染率或感染密度真的有成效(下降)，不能憑個人的感覺，必須要有評斷的方法，欲證明感染率的與過去比較，是在可接受的範圍內或已下降，是採用『卡方統計檢定』，而對感染密度則是『二個樣本比率差別統計檢定』。

可能有讀者會認為感染密度為什麼不可以採用『卡方統計檢定』呢？因為感染密度的分子是『感染人次』，而分母的部份則已考量時間因素為『住院人日數』，並不符合『卡方統計檢定』的機率分佈模式，若逕使用卡方統計檢定去計算，所得出的 p 值是有偏差的，容易導致結論錯誤，故無法採用，而比較二時期感染密度的異同，則是符合『二個樣本比率差別統計檢定』的機率分佈，本文將介紹推論統計方法中『卡方統計檢定』及『二個樣本比率差別統計檢定』的適用條件，計算公式及於 Excel 設定公式之方法，以供讀者參考。

卡方統計檢定-感染率檢定

卡方檢定(X²-test,或稱為 Chi-Square test "Chi"讀音與中文『開』同)是最常被廣泛應用的統計方法之一，適用於探究二個類別變項間的關係，只檢定兩個類別變項的相關，而不指出何者為自變項，何者為依變項。何謂類別變項？例如臨床上我們記錄有多少病人使用導尿管，資料只會有 5 人使用導尿管，而不會有 5.4 人使用導尿管。以大家熟悉的項目說明，類別變項的統計方法就如探討住院病人有無使用導尿管、與得到院內尿路感染是否有關係，此時使用導尿管屬於類別變項，其資料只有『有』或『沒有』二類，而院內尿路感染也屬於類別變項，其資料是『已經得』或『未得到』院內尿路感染，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。

採用卡方統計檢定，必須符合以下條件：

(1)2 個變項皆是類別變項；(2)樣本是從母群體中隨機抽樣。(3)兩組資料互相獨立，即是所有調查病人只有使用導尿管或未使用導尿管 2 種情形之一，不會有同時使用導尿管又未使用導尿管的情形出現。我們較常碰到資料不互相獨立的情況，例如比較參加研討會的人員其前測與後測成績，要探討在職教育的成效，是以每人自己前後測的結果互相比較，此時所取得的資料就屬於『不互相獨立』，若要比較資料不互相獨立的 2 個類別變項的相關聯的統計，所選用的統計方法為『McNemar's test for paired sample』，限於篇幅本文不介紹。(4)類別變項，A 變項可區分為 c 組($c \geq 2$)，B 變項可分為 r 組($r \geq 2$)，其資料可安排成 $r \times c$ 的表格。(5)每一細格之期望值必須大於 5。以 2×2 表格為例，其中某一格的期望值小於 5，因預期值很小而達不到 5，則因機會引起差別的可能性增大(不是真正的差別機會增大)。亦不符合卡方檢定的機率分佈，而需改採用其他的統計方法來檢定(Fisher's exact test)。

以 2×2 的卡方統計檢定說明，其假設為『不論有無曝露於某危險因子，此 2 群人的得病機率相同』，承前例說明，即是使用導尿管與未使用導尿管的 2 群病人，得到院內尿路感染的機率相同。檢定前先計算所有細格期望值都大於 5 時，而後算出卡方值後，再由統計書籍依自由度不同，查出其 p 值。由於現在很多電腦軟體都可直接提供卡方檢定的 p 值，甚至最普遍的 Excel 軟體都可提供，可省去查書的麻煩。

卡方統計檢定的程序

一、每個細格的期望值計算

有危險因素且有疾病的期望值: $(a+c) \times (a+b) / N$

有危險因素且無疾病的期望值: $(b+d) \times (a+b) / N$

無危險因素且有疾病的期望值: $(a+c) \times (c+d) / N$

無危險因素且無疾病的期望值: $(b+d) \times (c+d) / N$

二、卡方值計算公式

$(O-E)^2$

$X^2 = \sum \left\{ \frac{---}{E} \right\}$ (O:觀察值;E:期望值)

E

計算單一細格觀察值減去期望值，再除以期望值，每個細格計算出的數值之總合，即是卡方值。算出卡方值後，再由統計書籍依自由度不同，查出其 p 值。

範例-Excel 公式的設定

以表二為例，探討某醫院使用導尿管與院內尿路感染的關聯性，第 1 步驟是要計算各情況的期望值，例如使用導尿管且有院內尿路感染期望值，依比率原則，使用導尿管的 70 位佔所有病人 200 位的比率，再乘 30 位院內尿路感染，故計算使用導尿管有院內尿路感染的期望值， $30 \times 70 \div 200 = 10.5$ ，[公式 $(a+c) \times (a+b) / N$]，依序將另 3 格的期望值算出，分別為 59.5、19.5、110.5(圖一)，4 個格子的期望值皆大於 5，符合卡方統計檢定的條件；第 2 步驟於 Excel 軟體中，可利用現有函數指令 CHITEST()，其功能為可自動傳回卡方獨立性統計結果(p 值)，分別輸入觀察值與期望值的範圍，電腦即可算出 p 值 0.001847(圖二)。

三、統計檢定結果的解讀

檢定結果的 p 值 0.001847，代表『使用導尿管與未使用導尿管而得院內尿路感染的機會相同』，其發生的可能機率只有 0.001847，要我們認同『不論是否使用導尿管得院內感染的機會一樣』，我們的把握度只有 0.18%，由於機率太小了，因此我們採反方意見，即是拒絕『不論是否使用導尿管得院內感染的機會一樣』的論點，由使用導尿管得院內尿路感染率為 25.7%比未使用的 9.23%高，故可判定『使用導尿管較容易得院內尿路感染』。

四、卡方統計檢定的決策原則

1.統計學家曾建議，當 2 x 2 表格中，有一細格的期望值小於 5 時，或是 2x3 以上的類別變項比較時，其中有超過 20%的細格的期望值小於 5，就不適合採用『卡方統計檢定』，而是要改採用『費氏準確統計檢定』(Fisher's exact test)。2.當我們計算卡方值後，利用現成軟體的函數得出機率。p 值愈小，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就愈小，通常採用 p<0.05，我們即判定有統計上顯著差異。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。應用函數指令 CHITEST()，即使是 5x8 的表格或更多分層的表格亦可自動計算出 p 值。

二個樣本比率差別統計檢定

卡方檢定只考量病人是否得院內感染，並未進一步去思量暴露危險因素時間長短的影響，是否住的愈久或使用某侵入性導管天數較長時，其得到感染的機會是否會相同。在院內感染監測，我們會再計算另一個指標，每千人日感染密度(incidence density)，感染密度的定義，在研究族群中於特定期間累積所有人日數所發生不良事件的比值。此項指標其分子為感染人次，分母則為住院人日數或使用侵入性導管病人的導管使用天數的總合。當要比較當月感染密度是否比前 6 個月的感染密度高時，由於是比較 2 群人發生不良事件的比值，其符合二項分佈的統計機率，可採用『2 個樣本比率差別統計檢定』，應用此統計方法必須符合以下條件，1)必須要變異數(V1; npq) ≥ 5 才可使用此檢定方法，如果變異數小於 5，則是採用精準統計檢測(exact test)；2)必須事先知道兩組資料的人日數。

二個樣本比率差別統計檢定的程序

一、計算群突發期的感染個案數期望值(E1)

$$E1 = \frac{a_1 + a_2}{t_1 + t_2} \times t_1$$

二、計算 2 組資料的變異數(V1，變異數平方根即是標準差 SD)

$$V1 = \frac{a_1 + a_2}{t_1 + t_2} \times \left[\frac{a_1}{t_1} \times \frac{t_1}{t_1 + t_2} + \frac{a_2}{t_2} \times \frac{t_2}{t_1 + t_2} \right]$$

(^2 代表 2 次方)

三、計算標準常態值(Z 值)

$$Z = \frac{|a1 - E1| - 0.5}{SD}$$

(0.5 為校正係數)

四、以 Z 值去查標準常態分佈表所對應的 p 值

Z 值檢核範圍：

1) Z 值 ≤ 1.64 → 在可接受範圍, 2) 1.64 < Z 值 ≤ 1.96 → 瀕臨群突發臨界值, 3) Z 值 > 1.96 → 可能已發生群突發。

範例-Excel 公式的設定

以表四為例，探討某醫院群突發期的感染密度是否比基準期高，以機率的概念檢解釋計算公式，第 1 步驟是要計算群突發期的期望值，依照群突發期人日數佔所有期間人日數(群突發期及基準期)的百分比，再乘以所有感染個案數，即是群突發期的住院人日數 1,135 佔所有人日數 6,832 的比率，再乘 37 位院內感染，故群突發期的期望值 $1,135 \div 6,832 \times 37 = 6.15 [(a1+a2) \times t1 \div (t1+t2)]$ ，第 2 步驟則是計算變異數 $(np(1-p))$ ，即是總感染個案數(37) x 群突發期住院人日數為全部人日數的比率 $(1,135/6,832, p \rightarrow 0.1661) \times$ 基準期住院人日數為全部人日數的比率 $(1-p \rightarrow 0.8339)$ ，故變異數為 $37 \times 0.1661 \times 0.8339 = 5.1256 [(a1+a2) \times t1 / (t1+t2) \times t2 / (t1+t2)]$ ， $5.1256 > 5$ 符合此統計檢定的條件，第 3 步驟是計算標準差，即將變異數開根號， $\sqrt{5.1256} = 2.2640$ ，在 Excel 軟體利用開根號函數指令 SQRT()，第 4 步驟計算標準常態值(Z 值)，即是觀察值 12 減去期望值 6.15 取絕對值，再減去 0.5 校正係數，最後除以標準差 2.2640，即可得到 Z 值 $2.3645 [(|a1 - E1| - 0.5) / SD]$ ，在 Excel 軟體利用絕對值函數指令 ABS()，第 5 步驟，利用 Excel 軟體利用標準常態累加分配函數指令 NORMSDIST()，得出累加分配機率為 0.9909，再以 1 減去 0.9909，得機率為 0.00903。

五、統計檢定結果的解讀

檢定結果的 p 值 0.00903，小於 0.05，代表群突發期的感染密度每千人日 10.6 例，與基準期的每千人日 4.4 例，有統計上顯著差異，故可判定『群突發確實發生』。

結語

病人的院內感染監測，不論病房別或科別每個月都需要判定與前幾個月的感染率(密度)是否有差異，希望本文介紹藉由 Excel 軟體，簡單的公式設定，使大家於平日的工作上可以更得心應手。至於期望值小於 5 或變異數小於 5，不適用『卡方統計檢定』或『二個樣本比率差別統計檢定』，而改採用 Fisher's exact test 或 exact test，因其計算公式較為繁複，未來有機會，再另文與讀者分享。

表一 卡方統計檢定 2 x 2 表格

危險因素 \ 疾病	有疾病	無疾病	合計
有危險因素	a	b	a+b
無危險因素	c	d	c+d
合計	a+c	b+d	N

a,b,c,d 為 4 個細格的數值，
N (a+b+c+d) 為總樣本數

表二 使用導尿管與院內尿路感染卡方統計檢定比較

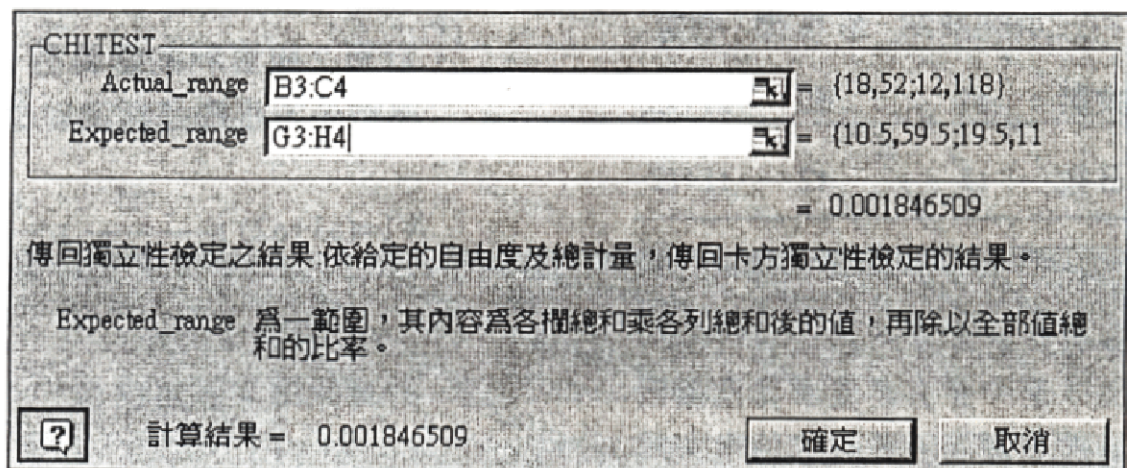
導尿管 (危險因素)	院內尿路感染	無院內尿路感染	合計
使用導尿管	18(a)	52(b)	70(a+b)
未使用導尿管	12(c)	118(d)	130(c+d)
合計	30(a+c)	170(b+d)	200(a+b+c+d)

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	E	F	G	H	I
1	觀察值				期望值			
2		院內尿路感染	無院內尿路感染	合計		院內尿路感染	無院內尿路感染	合計
3	使用導尿管	18	52	70	使用導尿管	10.50	59.50	70
4	未使用導尿管	12	118	130	未使用導尿管	19.50	110.50	130
5	合計	30	170	200	合計	30	170	200
6	每細格期望值皆大於5時							
7	CHI TEST (p值)	0.001847						

The formula bar shows: = \$D3 * B\$5 / \$D\$5

圖一 於 excel 軟體建立觀察值及期望值的表格範圍



圖二 應用函數指令 CHITEST()，並分別輸入觀察值與期望值的範圍

表三 2 個樣本比率差別統計檢定

	群突發期	基準期	合計
感染個案數	a1	a2	a1+a2
住院人日數	t1	t2	t1+t2

	A	B	C	D	E
1			群突發期	基準期	合計
2		感染個案	12	25	37
3		住院人日數	1135	5697	6832
4					
5	E1(期望值)=	$(a_1+a_2) \cdot t_1 / (t_1+t_2)$	$=(C_2+D_2) \cdot$		
6	V1(變異數)=	$(a_1+a_2) \cdot t_1 / (t_1+t_2) \cdot t_2 / (t_1+t_2) > 5$	5.1256		
7	SD(標準差)=	$SQRT(V_1)$	2.2640		
8	Z(常態分佈)=	$(ABS(a_1-E_1)-0.5)/SD$	2.3645		
9	p 值	$1-NORMSDIST(Z)$	0.00903		

圖三 於 excel 軟體建立二個樣本比率差別統計檢定的程序

表四 群突發期與基準期感染密度

	群突發期	基準期	合計
感染個案數	12(a ₁)	25(a ₂)	37(a ₁ +a ₂)
住院人日數	1,135(t ₁)	5,697(t ₂)	6,832(t ₁ +t ₂)
感染密度	0.0106(a ₁ /t ₁)	0.0044(a ₂ /t ₂)	

參考文獻

1. Rosner B: Fundamentals of Biostatistics, 3rd ed. Boston: PWS-KENT Publishing company, 1990:318-97.

2. 王文中: Excel 於資料分析與統計學上的應用。台北: 博碩文化。1997:418-70。