用Excel解決農業研究資料統計分析的方法(三)－ χ^2檢定

農藝組 呂秀英 魏夢麗 呂椿棠

生物統計

一、前 言

絕大多數微軟作業系統的個人電腦用戶都擁有Excel試算表管理軟體，雖然它不是一個專用統計軟體，但只要我們精通它的資料管理及分析技巧，並且注意它在應用上的限制，可讓農業研究資料的初步探勘及分析的計算工作變得輕鬆起來。本系列報告在介紹如何使用Excel進行農業科學研究中資料分析的方法，前兩篇已介紹了敘述統計量、t檢定、F檢定的應用和分析，而本篇重點則在說明如何利用Excel來進行χ^2檢定。

二、χ^2檢定的意義及用途

從試驗中所獲得的性狀資料，可區分為兩類：量的變數與質的變數，量的變數如株高、體重、穗數、蟲數等，質的變數如喜歡等級程度、顏色、性別等。而針對數量變數的統計檢定方法，則分成兩類：計量值檢定及計數值檢定。計量值指的是連續變數的資料形式，如株高、產量等，這類資料往往存在有小數點，其檢定方法根據欲探討的統計量為何，又區分為變方檢定和平均值檢定(請參考前作技術服務季刊64期)。計數值則是不連續之離散變數的資料形式，通常是以個數呈現，如菌數、蟲數、人數等。而對於質的變數，當把它們轉換成計數值型態表示時，則可以進行統計分析，例如喜歡的人數、呈現各顏色的花朵數目、不同性別的個數等。χ^2檢定(唸成卡方檢定)就是專門用在處理這種資料形式以計數值呈現的統計顯著性測驗，它是利用不同處理間所產生的個數是否顯著差異之檢定比較，來決定所擬定的假說是否成立。計數值的檢定方法，除非將個數轉換成百分比則使用標準常態Z檢定法以外，一律採用χ^2檢定。

χ^2檢定的用途相當廣泛，一般統計教科書將其區分成四種應用場合：樣本變方與一個已知族群變方之比較、適合度檢定、關連表檢定、純度檢定，其間的意義、公式及其自由度間的差異整理如表一。除了單一變方檢定屬於計量值檢定法以外，後三者均屬於在計數值統計檢定上的應用範疇。由於變方之χ^2檢定的公式很簡單，且在農業試驗上較少有這方面的需求運用，故本篇將著重在計數值之χ^2檢定。

在一般統計教科書中，對於不同應用場合的計數值之χ^2檢定，往往會列出不同的χ^2檢定公式，其實它們都是由一個原始公式 $\chi^2=\sum^{k}\frac{(O-E)^2}{E}$ 推導衍生而來，O是實測值，E是理論(期望)值。在過去電腦尚未普遍而必須仰賴計算機甚至手算的時代裡，使用原始公式，欲求得最後的χ^2值，總要在計算機上敲打很多步驟，過程繁複容易出錯，於是一些統計學者便從該原始公式發展出各種簡式，以利人工

作者：呂研究員秀英
連絡電話：04-23302301-125

計算。但現在電腦當作計算工具，相當快捷且正確，其實未必一定得使用簡式，直接採用原始公式即可。

對於χ^2檢定，Execl並未提供如同敘述統計量、F和t檢定一樣之使用「工具」中的「資料分析」（技術服務季刊64期)，但提供以原始公式為基礎的CHITEST函數，可直接算出χ^2檢定的右尾機率值，來進行實測值與理論值(期望值)間差異的顯著性比較。只要活用此函數，就能應付χ^2檢定在計數值的所有不同應用場合，不過要特別注意χ^2檢定在不同應用場合中的自由度算法不盡相同，此將在以下各節中配合實例再加以說明。

三、χ^2檢定之求算步驟及結果解讀：適合度檢定

假設象牙色與紅色金魚草雜交之F_2後代所得各種顏色的株數為紅色:粉紅色:象牙色＝20:55:25，欲探討其遺傳分離比是否符合1:2:1之擬說，則利用CHITEST函數的分析步驟如下：

1. 首先根據擬說，將各組的期望值計算出來：適合度檢定的期望值求算方法，為總計乘以各組所佔比例。如本例總計=100，故三種顏色的期望值分別是25:50:25　　(即紅色期望值$=100\times\frac{1}{(1+2+1)}$，其他兩組同理類推)。
2. 將實測值和期望值以類似於兩向關連表的型態輸入到Excel工作表中：由於CHITEST函數對於自由度的計算，是採用兩向關連表形式的計算方式：(r-1)(c-1)，r為行數，c為欄數；因此資料輸入的格式務必遵守此原則，否則Excel內部在計算自由度時會出錯，進而造成檢定結果錯誤。如本例將實測值和期望值分置於不同欄位，各欄內逐行輸入各組顏色的資料，如圖一所示，此時雖然實測值(期望值)的欄數只有一行，但我們發現Excel不會讓(c-1)=(1-1)=0的情況發生，它自動將自由度計算成(3-1)=2，此與適合度檢定的實際自由度為(組數-1)=(3-1)＝2是一致的。

表一、χ^2檢定的四種不同應用場合之區分說明

資料形式	應用場合	意　義	原始公式	自由度[z]
計量值	變方檢定 (test of variance)	樣本變方(s^2)與一個已知族群變方(σ^2)之比較	$\chi^2=\frac{s^2}{\sigma^2}$	n-1
計數值	適合度檢定 (test for goodness of fit)	實測頻度(O)分布與理論頻度(E)分布(分布型態未定)是否一致(符合)的檢定	$\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$	k-1 或k-s-1
計數值	關連表檢定 (test of assouation table)	即獨立性檢定，測驗A、B二因子間是否有關連，又稱列聯表檢定(test of contingency table)	$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$	(r-1)(c-1)
計數值	純度檢定 (test of homogeneity)	計數資料的試驗，如取k個樣本，該等樣本是否純質	$\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}$	k-1

[z] n: 觀測值個數； k: 組數； s: 已知分布介量數,如常態分布 s=2； r,c: 關連表的行及欄數。

3. 然後於空白單格內插入CHITEST函數：其格式為CHITEST(Actual_range,Expected_range)，括號內的Actual_range指定實測值的範圍位址，而Expected_range指定期望值的範圍位址。如本例=CHITEST(B3:B5,C3:C5)。必須注意的是，CHITEST函數給出的是χ^2統計值的單尾機率值(計數值之χ^2檢定一律為單尾測驗，因對立擬說是不符合，無所謂誰大誰小)，而非χ^2統計值，因此只要檢視該機率值是否小於顯著水準0.05：倘小於0.05，即棄卻虛無擬說，表示顯著不符合假設的分離比；否則大於0.05，表示符合假設的分離比。如本例的機率值0.4724>0.05，故符合1:2:1分離比 (圖一)。
4. 上述步驟即已完成χ^2檢定，但若欲進一步得到χ^2統計量的實測值和臨界值，則可再利用CHIINV函數：其格式為CHIINV(probability,degrees_freedom)，括號內probability指定機率值或其範圍位址，而degrees_freedom輸入自由度。如本例=CHIINV(D8,2)，D8為前步驟由CHITEST函數所得機率值的所在位址，2為本例自由度，由此算出相對於機率值0.4724的χ^2實測值=1.5；倘欲得知在0.05顯著水準下自由度等於2的χ^2臨界值，同樣利用這個函數，但寫成=CHIINV(0.05,2)。χ^2檢定是否顯著，可比較χ^2實測值是否大於χ^2臨界值，也可直接查看其χ^2實測值的機率值是否小於0.05，兩者擇一即可，因此這些再利用CHIINV函數的步驟其實可省略，但考量在發表報告時，或許作者希望也能將χ^2實測值和臨界值一起列出以供參考，故本文在此特別補充介紹。

四、χ^2檢定之求算步驟及結果解讀：關連表檢定

假設一農藥防治試驗，欲檢定兩種農藥的防病效果有無差別，而調查試驗田間的罹病和正常株數，如圖二之實測值。這是一個2×2關連表檢定，虛無擬說為農藥與防病效果二因子間無關連。2×2關連表檢定是r×c關連表中最簡單的一種，其原理完全一樣的，自由度為(r-1)(c-1)。本文僅以2×2關連表檢定為例說明，但所描述的計算方式完全可推至r, c大於2以上的關連表。

關連表的χ^2檢定利用CHITEST函數的操作步驟與前述之適合度檢定相同，但因此處沒有已知的理論比例或分離比，故必須自己推估。由於關連表中各單格內理論值與邊界總計呈正比的既定關係，因此各單格內的理論值即等於其實測值所相對應的邊際和之乘積再除以總計。以本例而言，”甲”農藥之病株的理論值=313×47/564=26.08，”乙”農藥之病株=251×47/564=20.92，其他單格的理論值同理

	A	B	C	D
1	適合度檢定：			
2	F2	實測株數	理論株數	
3	紅色	20	25	
4	粉紅色	55	50	
5	象牙色	25	25	
6	total	100		
7				
8	Prob(卡方)	=CHITEST(B3:B5, C3:C5)=		0.4724
9				
10	卡方實測值	=CHIINV(D8,2)=		1.5
11	卡方臨界值	=CHIINV(0.05,2)=		5.9915

圖一、用Excel函數進行χ^2檢定－適合度檢定：以象牙色與紅色金魚草雜交之F_2後代是否符合1:2:1遺傳分離比為例說明。

類推。所推估的理論值和實測值，仍須維持兩向關連表資料形式輸入於工作表中。本例以CHITEST函數所得 χ^2 檢定的機率值等於0.3707>0.05，無法棄卻虛無擬說，表示農藥與防病效果間無顯著關連，即兩種農藥的防病效果無顯著差異。另外，可進而利用CHIINV函數反求該機率值下的 χ^2 實測值和臨界值(圖二)，此處該函數內所指定的自由度為(2-1)(2-1)=1。

	A	B	C	D
1	2x2關連表檢定：			
2	實測值			
3	農藥種類	病株	好株	計
4	甲	29	284	313
5	乙	18	233	251
6	計	47	517	564
8	理論值			
9	農藥種類	病株	好株	計
10	甲	26.08	286.92	313
11	乙	20.92	230.08	251
12	計	47	517	564
13				
14	Prob(卡方)	=CHITEST(B4:C5, B10:C11)=		
15				

圖二、用Excel函數進行 χ^2 檢定－關連表檢定：以檢定兩種農藥的防病效果有無差別為例說明。

	A	B	C	D
1	純度檢定：			
2	試區	實測值	理論值	
3	1	11	6	
4	2	4	6	
5	3	4	6	
6	4	5	6	
7	Total	24		
8				
9	Prob(卡方)	=CHITEST(B3:B6, C3:C6)=		0.129
10				
11	卡方實測值	=CHIINV(D9,3)=		5.667
12	卡方臨界值	=CHIINV(0.05,3)=		7.815

圖三、用Excel函數進行 χ^2 檢定－純度檢定：以檢定四個甘藍菜試區內所發現的尺蠖數之分布是否隨機為例說明。

五、χ^2 檢定之求算步驟及結果解讀：純度檢定

假設一個蟲害調查試驗，從四個甘藍菜試區發現尺蠖數各為11,4,4,5，欲探討尺蠖之分布是否隨機？此例的虛無擬說為隨機假設，故四個試區的理論值應該具等比例關係，即四個試區的理論值各為6,6,6,6(總計24/4=6)。然後將實測值和理論值並排輸入，以維持兩向關連表形式，如圖三所示，以避免CHITEST在求機率值過程中計算自由度時產生錯誤，但此時實測值(期望值)的欄數雖只有一行，Excel並不會讓(c-1)=(1-1)=0的情況發生，而自動將自由度計算成(4-1)=3，此與純度檢定的實際自由度等於(組數-1)=(4-1)=3是一致的。至於使用CHITEST函數的操作步驟與前述均相同，本例 χ^2 檢定的機率值等於0.1290>0.05，表示無法棄卻虛無擬說，即這四個試區的尺蠖發生呈隨機分布。至此，χ^2 檢定已經完成，但可進而利用CHIINV函數反求該機率值下的 χ^2 實測值及其臨界值(圖三)。