

—新規開発薬剤が農業現場で普及・貢献するまでのプロセスにおける諸問題とその対応—

2) 登録申請データの収集と評価における諸問題とその対応

① 試験実施とデータ取りまとめの場面

効果比較試験における閉検定手順 (Closed testing procedure) の適用のすすめ

佐賀県果樹試験場 田代暢哉

はじめに

実際の防除技術開発の場面では、2群の比較、すなわち、無処理と試験薬剤の比較、あるいは対照薬剤と試験薬剤の比較だけということは少なく、無処理、対照薬剤、試験薬剤 A、試験薬剤 B、試験薬剤 C・・・という試験区の構成になることが多い。このような場合にはどのような比較を行って、試験薬剤の効果を評価すればよいのだろうか。頻度データについても多重比較検定を行うことは可能である。しかし、多重比較検定には①検出力の低下、②順序関係のある検定には不適という問題がある。

そこで、今回のワークショップでは3群以上のデータの比較を行う場合に用いる「閉検定手順 (Closed testing procedure)」について紹介する。

1. 無処理、対照薬剤、試験薬剤の各比較の意味

薬剤の効果や防除手段の有用性を評価するための試験を構成する要素には、試験薬剤の他に必ず無処理と対照薬剤が含まれており、それぞれの病害虫に応じた設計に基づいて試験がなされることになる。この場合、ただ単にすべての区の比較を行うわけではない。当然、そこには目的がある。まずは、試験が成立しているのかどうか、次に試験薬剤の効果があるかどうか、さらに対照薬剤と比較した場合にどうなのか、加えて、対照薬剤よりも効果が高い薬剤がいくつもあった場合にそれらの薬剤間での優劣はどうかといったことである。

以上のことはごく当たり前のことではあるが、防除効果の比較試験と言えは多重比較の a, b, c・・・の符号を付けるだけで順序だてた考察がなされていない例が多すぎるように思われる。試験を構成する要素とその比較の意味を整理し、理解しておくことが大切である。以下に、各比較の意味について説明する。

1) 対照薬剤と無処理の比較 (絶対条件)

対照薬剤として用いられるのは慣行として現場で使用されているものが大部分である。このため、試験が適切な設計のもとに適切に実施されてさえいれば、無処理と対照薬剤との間に有意な差があるのは当たり前で、逆に、有意でない場合には試験が適切に実施されなかった可能性が高いと考えられ、試験が成立しているとはいえない。このような条件下の試験では他の比較、例えば無処理と試験薬剤との比較や対照薬剤と試験薬剤との比較を行っても意味がないことになる。すなわち、対照薬剤と無処理の比較の目的は、試験が成立しているかどうかを確認することであり、「試験の感度をみるための比較」とも呼ばれる。本比較で有意差がみられるということは試験薬剤の効果を評価するための絶対条件である。

## 2) 試験薬剤と無処理の比較(必要条件)

試験薬剤に効果があることを言うためには、まずは少なくとも無処理よりも発生が少ないなどの有意な違いがあることを示さなければならない。これを証明するのが試験薬剤と無処理の比較である。試験薬剤の効果を主張するための最低限、必要な条件であるといえる。「試験薬剤の有効性をみるための比較」とも呼ばれる。

## 3) 試験薬剤と対照薬剤の比較(十分条件)

既存の対照薬剤よりもまさっているのか、同等なのか、あるいは劣っているのかを明らかにするための比較である。有意にまさっていることが証明できれば、試験薬剤がすぐれていることを強く主張できることになる。「試験薬剤の優位性をみるための比較」とも呼ばれる。

以上の三つの組み合わせの比較には有意になりやすい順序関係がある。すなわち、(対照薬剤と無処理の比較) > (試験薬剤と無処理の比較) > (試験薬剤と対照薬剤) の比較の順である。また、この順序は試験薬剤の有効性を証明するためのステップでもある。そこで、各ステップの仮説を確認しながら、順序立てて検定を行うことによって試験薬剤の有効性が評価されることになる。この手順を「閉検定手順(Closed testing procedure)」という。

## 2. 閉検定手順のステップ

<ステップ 1> 有意水準  $\alpha$  で対照薬剤と無処理の比較を行う。有意差が認められればステップ 2 に進み、試験薬剤の有効性を確認する。もし、有意でなければ、試験が適切な設計および条件のもとで行われておらず、試験は成立していないとして解析を終了する。この場合、試験が成立していないので試験薬剤の効果は判定不能ということになる。

<ステップ 2> 有意水準  $\alpha$  で試験薬剤と無処理の比較を行う。有意差が認められればステップ 3 に進み、試験薬剤の有効性を対照薬剤と比較する。もし、無処理に対して有意でなければ、試験薬剤の効果はなかったことになり、解析は終了となる。

<ステップ 3> 有意水準  $\alpha$  で試験薬剤と対照薬剤の比較を行い、効果を判定する。

以上が閉検定手順である。ステップ 3 までいくと、すべての群間比較を検定することになるが、同時に 3 回検定を行っているわけではないので、多重性の問題が起きることはなく、 $\alpha$  エラーの大きさを有意水準以下に抑えることができる。多重性の問題が起きるのは同時に 3 群以上を比較する場合である。閉検定手順は同時比較ではないので a, b, c... の符号を付けることはできないので注意が必要である。本手順に従うと、多重検定を行うよりも高感度に有意差を検出できる可能性が高くなる。

## 3. 閉検定手順の実際

表に  $\chi^2$  検定を用い、閉検定手順にしたがって行った薬剤の効果判定の実際を示した。試験薬剤の数が増えた場合には同様に<ステップ 2>、<ステップ 3>をそれぞれの薬剤について繰り返して判定すればよい。

検定手法は頻度データ(カテゴリーデータ)の場合、 $\chi^2$  検定の他に Fisher の正確確率検定、Mann-Whitney の  $U$  検定、Cochran-Mantel-Haenszel 検定などを用いる。なお、データ数やそのばらつきに問題がないような場合に頻度データをパーセントデータに変換して検定するときには、 $t$  検定や対応のある  $t$  検定を用いればよいが、おすすめはできない。

表 閉検定手順による試験薬剤の効果判定の一事例

対象病害虫名	カンキツ褐色腐敗病(接種試験)																											
1. 試験目的	防除効果および薬害の検討																											
2. 試験方法	佐賀県小城市小城町晴気 一般農家圃場																											
試験地場所	上野早生・16年生																											
供試品種、樹齢	1区3樹																											
試験の規模	甚発生(接種), 自然発病は試験樹では認められなかった。																											
対象病害虫発生状況	本試験に影響を及ぼす薬剤の散布はなかった。																											
試験開始前の薬剤散布	10月2日に動力噴霧機を用いて枝葉から薬液が滴り落ちる程度に十分量(10リットル/樹)を散布した。使用ノズルは強力キリナシプラ2頭口, 散布圧力は1.0MPaである。																											
処理年月日・量・方法																												
接種方法	散布10日後(10月12日), 15日後(10月17日), 22日後(10月24日)に各樹から15果を採取し, <i>Phytophthora palmivora</i> の遊走子懸濁液( $10^5$ 個/ml)に浸漬したサラシ片(1cm×1cm)を1果実あたり4箇所貼り付けることによって接種した。																											
試験期間中の気象の概要	<table border="1"> <thead> <tr> <th></th> <th>10月2日 (散布日)</th> <th>10月12日 (10日後)</th> <th>10月17日 (15日後)</th> <th>10月24日 (22日後)</th> </tr> </thead> <tbody> <tr> <td>降雨日数(日)</td> <td>1</td> <td>0</td> <td>0</td> <td>3</td> </tr> <tr> <td>降雨量(mm)</td> <td>2.0</td> <td>0</td> <td>0</td> <td>49.0</td> </tr> </tbody> </table>		10月2日 (散布日)	10月12日 (10日後)	10月17日 (15日後)	10月24日 (22日後)	降雨日数(日)	1	0	0	3	降雨量(mm)	2.0	0	0	49.0												
	10月2日 (散布日)	10月12日 (10日後)	10月17日 (15日後)	10月24日 (22日後)																								
降雨日数(日)	1	0	0	3																								
降雨量(mm)	2.0	0	0	49.0																								
調査月日・方法	接種果実は28℃温室条件下で保持した。接種7～10日後に接種部位の発病の有無を調査し, 発病部位率を算出した。なお, 接種3日目までは感染を促すためにサラシ片が乾かないように適宜ハンドスプレーを用いて滅菌蒸留水を噴霧して濡れた状態を保った。																											
3. 試験成績	<table border="1"> <thead> <tr> <th rowspan="2">供試薬剤</th> <th rowspan="2">希釈倍数</th> <th colspan="3">発病部位数/接種部位数(防除価)</th> <th rowspan="2">薬害</th> </tr> <tr> <th>散布10日後</th> <th>散布15日後</th> <th>散布22日後</th> </tr> </thead> <tbody> <tr> <td>(試験薬剤)ABCフロアブル</td> <td>5,000</td> <td>0/60 (100)</td> <td>2/60 (96.7)</td> <td>46/60 (23.3)</td> <td>—</td> </tr> <tr> <td>(対照薬剤)XYZ水和剤</td> <td>400</td> <td>1/60 (98.2)</td> <td>0/60 (100)</td> <td>8/60 (86.7)</td> <td>—</td> </tr> <tr> <td>無処理</td> <td>—</td> <td>57/60</td> <td>60/60</td> <td>60/60</td> <td></td> </tr> </tbody> </table>	供試薬剤	希釈倍数	発病部位数/接種部位数(防除価)			薬害	散布10日後	散布15日後	散布22日後	(試験薬剤)ABCフロアブル	5,000	0/60 (100)	2/60 (96.7)	46/60 (23.3)	—	(対照薬剤)XYZ水和剤	400	1/60 (98.2)	0/60 (100)	8/60 (86.7)	—	無処理	—	57/60	60/60	60/60	
供試薬剤	希釈倍数			発病部位数/接種部位数(防除価)				薬害																				
		散布10日後	散布15日後	散布22日後																								
(試験薬剤)ABCフロアブル	5,000	0/60 (100)	2/60 (96.7)	46/60 (23.3)	—																							
(対照薬剤)XYZ水和剤	400	1/60 (98.2)	0/60 (100)	8/60 (86.7)	—																							
無処理	—	57/60	60/60	60/60																								

## 4. 考察

## ステップ1:&lt;対照薬剤の効果と試験の感度&gt;

散布10日後, 同15日後および同22日後の無処理区と対照薬剤区における発病状況を $\chi^2$ 検定(Yatesの補正)によって比較するとp値(データの差が偶然によって生じる確率)はそれぞれ0.00001以下になり, 対照薬剤区では有意(99%水準)に発病が少なかったため本試験は成立していると判断し, 以下の判定を行った。

## ABCフロアブル5,000倍の評価

## ステップ2:&lt;無処理との比較&gt;

散布10日後, 同15日後および同22日後の本剤区と無処理区における発病状況を $\chi^2$ 検定(Yatesの補正)によって比較するとすべて $p < 0.0001$ となり, 本剤区では有意(99%水準)に発生が少なく, 本剤の有効性が認められた。

## ステップ3:&lt;対照薬剤との比較&gt;

本剤と対照薬剤の効果をもとに $\chi^2$ 検定(Yatesの補正)によって比較すると散布10日後および同15日後のp値はそれぞれ0.315および0.153で, 本剤と対照薬剤間に有意差は認められなかった。しかし, 散布22日後のp値は0.00001以下で, 本剤は対照薬剤に有意(99%水準)に劣った。

以上のように, 本剤は散布15日後までは対照薬剤と同等のすぐれた効果を示したが, 散布22日後になると対照薬剤に残効面で劣ると判断された。薬害の発生は認められなかった。

## &lt;参考&gt;

## Cochran-Mantel-Haenszel 検定

## 1. 反復して得られた頻度データの統計処理

病害虫防除技術開発に関する試験を実施する場合、種々の要因によって試験圃場内の場所の違いによって病害虫の発生がばらつくことはよくあることである。圃場試験より規模は小さいが、枠試験やポット試験でも病害虫の発生は多少の差はあれ、ばらつくことが多い。そこで、ばらつきの問題を少しでも小さくするために圃場をいくつかのブロックに分けて、その中に一連の処理区を設けることが行われる。このため、同一の処理が圃場内に複数存在することになり、それぞれのブロックで得られたデータをまとめて評価することが必要になってくる。このための方法として、分散分析と多重比較が用いられる。

しかし、これらの手法が適用できるのは収量や草丈、病斑数、虫数などの量的変数に限られる。発病率や発生率、被害率など、日頃よく使用するカテゴリー変数の場合にはどのような手法を用いればよいのだろうか。これまで一般にはカテゴリー変数を量的変数(パーセントデータ)に変換し、その値に基づいた分散分析や多重比較が行われている。この場合、調査個体数が数百個以上あって、なおかつ各反復および各処理の調査個体数がほぼ同一であるならば、変換したデータをこれらの解析に用いることは可能であるとされている。しかし、実際の試験の実施にあたってはこのような条件がいつも満たされているわけではない。また、頻度データ(発病数と無発病数など)をパーセントに変換することによってデータの持つ量的な情報が失われてしまうという問題が生じてくる。

## 2. パーセント変換データに基づく検定の問題点

例えば、第1表に示したカンキツかいよう病に対する薬剤試験のデータについて考えてみる。1区1樹3反復で試験を実施し、発病果率を算出した。そこで、これらのデータを  $\arcsin\sqrt{\quad}$  変換して「対応のある  $t$  検定」を行ったところ、 $p$  値は 0.03132 になり 0.05 よりも小さいことから「薬剤間に発病率の差はない」という帰無仮説は棄却され、95%の確率で薬剤間に効果差が認められると結論された(第1図参照)。

【第1表】カンキツかいよう病に対する各種薬剤の防除効果

供試薬剤	反復ごとの各樹の発病果率 (%)			平均発病果率 (%)
	1	2	3	
薬剤 A	36.1	30.0	31.7	32.6
薬剤 B	26.7	23.1	25.8	25.2

【注】供試品種はワシントンネーブルで、6月中旬に果実発病を認めてから各種薬剤を約3週間おきに4回散布後、10月上旬に果実発病の有無を調査した。試験は1区・1樹・3反復で実施した。

	A	B	C	D	E	F	G
1	反復	薬剤Aの発病率	薬剤Bの発病率	薬剤Aの発病率- 薬剤Bの発病率			
2	1	36.9	31.1	5.81700	差の平均	4.68033	
3	2	33.2	28.7	4.48500	差の標準偏差	1.46937	
4	3	34.3	30.5	3.73900	件数	3	
5					自由度	2	
6					t値	5.51705	
7					p値	0.03132	

【第1図】 第1表のデータに基づいた「対応のあるt検定」の結果

【注】発病率は第1表に示す実際の発病率データを arcsin $\sqrt{\quad}$ 変換した値, 得られたp値は0.05よりも小さいので「2群の平均値は等しい」という帰無仮説は棄却され, 「薬剤間に効果差がある」ことがわかる。

ところで, 第1表に示す発病率を算出するためには当然, 無発病個体数と発病個体数のデータが必要である。では, 実際にこれらのデータがどうなっていたのかというと, 第2図に示すように各反復間には調査個体数に大きなばらつきを生じていた。カンキツ類では着果量に大きな樹体間差を生じるのは一般によくあることで, このためばらつきができるだけ少なくなるように事前に供試樹を設定していてもこのようなことが起きてしまう。カンキツ類の場合に限らず, 試験を実施して行く過程で当初予定していた調査個体数が確保できなくなることは種々の要因の影響を受ける病害虫防除に関する試験ではどうしてもあり得ることである。データ数が多ければその結果の信頼性は高くなり, 逆にデータ数が少なれば信頼性は低いだろうということは直感的にわかることで, データの数をまったく考慮しないパーセントに変換された値に基づく検定には不安が付きまとう。このため, データ数とそのばらつきの問題は処理間の評価を行う上で十分に配慮されなければいけない点である。しかし, それにもかかわらず, これまでパーセントデータのみでの評価が行われてきたというのが現状である。

	A	B	C	D	E	F	G	H	I
1		薬剤A				薬剤B			
2		無発病	発病	調査	発病率	無発病	発病	調査	発病率
3		個体数	個体数	個体数		個体数	個体数	個体数	
4	反復1	108	61	169	36.1	22	8	30	26.7
5	反復2	56	24	80	30.0	150	45	195	23.1
6	反復3	43	20	63	31.7	98	34	132	25.8
7	計	207	105	312		270	87	357	
8	平均				32.6				25.2

【第2図】 カンキツかいよう病防除試験の各試験区における発病果実数および無発病果実数

【注】着果量には大きな樹間差がみられたので, それぞれの樹に着果していたすべての果実を調査対象にした。

### 3. 複数の頻度データの統合評価

では、第3図に示すような反復試験から得られた複数の頻度データ(4分表データ)をどのようにしてまとめて判定を下げばよいのだろうか。4分表データなので $\chi^2$ 検定を行うことから、単純に考えれば $\chi^2$ 値をまとめればよいのではないかと考えるのが普通である。しかし、第3図に示すように、反復間で調査個体数が大きく異なっている。このため、単純に平均することには問題がありそうだということになる。ランダムサンプリングがきちんと行われているとすると、やはり調査個体数が多いデータは少ないデータよりも事象をより正確に反映している、つまり、信頼性が高いと考えられる。このため、調査個体数が多いデータを少ないデータよりもより重視すべきだということになり、このための手法としてデータ数に応じて重み付けをすることになる。このような考え方に基づいて複数の頻度データを統合して判定するための手法が Cochran-Mantel-Haenszel 検定である。

		反復1			反復2			反復3		
		発病	無発病	計	発病	無発病	計	発病	無発病	計
薬剤A		61	108	169	24	56	80	20	43	63
薬剤B		8	22	30	45	150	195	34	98	132
計		69	130	199	69	206	275	54	141	195

### 4. Cochran-Mantel-Haenszel 検定の手順

本検定法の手順についてはいくつか示されているが、なかでも青木繁伸「Cochran-Mantel-Haenszel 検定の計算手順と Excel ワークシート」(<http://aoki2.si.gunma-u.ac.jp/lecture/mb-arc/Cochran-Mantel-Haenszel.xls> として2006年11月18日にダウンロード)がわかりやすく参考になる。以下に青木による検定手順と Excel ワークシートにそれぞれ若干の修正を加えたものを示す(第4図参照)。なお、反復が多くなる場合には行を増やしていくことによって対応すればよい。

#### Cochran-Mantel-Haenszel 検定の計算手順

- 合計例数  $n_i$  を計算  $n_i = n_{i1} + n_{i2}$
- 割合の差  $= p_{i1} - p_{i2}$  を計算
- 二群を込みにした割合  $\bar{p}_i$  を計算  $\bar{p}_i = (n_{i1} + n_{i2}) / n_i$
- $\bar{q}_i$  を計算  $\bar{q}_i = 1 - \bar{p}_i$
- 標準差(standardized difference)  $d_i$  を計算  $d_i = (p_{i1} - p_{i2}) / (\bar{p}_i) / (\bar{q}_i)$
- 重み  $w_i = \bar{p}_i * \bar{q}_i * n_{i1} * n_{i2} / n_i$  を計算
- 重みの合計  $\sum w_i$  を計算
- 標準差と重みとの積  $d_i * w_i$  を計算
- その合計  $\sum d_i * w_i$  を計算
- 標準差の二乗と重みとの積  $d_i^2 * w_i$  (各研究における連続性の補正をしないカイ二乗統計量) を計算
- その合計  $\sum d_i^2 * w_i$  を計算
- 全体的な標準差の推定値  $\bar{d} = \sum d_i * w_i / \sum w_i$  を計算
- その標準誤差  $SE(\bar{d}) = 1 / \sqrt{\sum w_i}$  を計算
- カイ二乗値 total は  $\sum d_i^2 * w_i$  (11に同じ)
- カイ二乗値 assoc は  $(\bar{d} / SE(\bar{d}))^2$
- カイ二乗値 homog は カイ二乗値 total からカイ二乗値 assoc を引いたもの
- カイ二乗値 assoc の自由度は「1」で、標準差の有意性の検定に使用
- カイ二乗値 total の自由度は「研究(反復)の数」で、第4図の場合は3
- カイ二乗値 homog の自由度は「カイ二乗値 total の自由度」から1を引いたもので、標準差の均等性の検定を行う検定に使用
- それぞれの  $p$  値は  $\text{chidist}(\text{カイ二乗値}, \text{自由度})$  により計算

第4図に示した計算例の解釈は、「 $\chi^2_{assoc}$ から得られる  $p$  値が 0.0754 と 0.05 よりも大きいので全体としての標準差は有意なものではない。つまり、割合に差は認められない。」ということになる。

以上の結果は、第1図に示したパーセントデータに基づいた「対応のある  $t$  検定」から得られた「有意差あり」の結論とは異なっている。このように、サンプル数を考慮しないパーセントデータでの検定では真実がみえてこないことがあり、注意が必要なことを示している。

12	薬剤A				薬剤B				合計調査 個体数 nt	発病割 合の差 pi1-pi2	二群を込み にした発病 割合 pi_bar	qi_bar=1 -pi_bar	標準差 di	重み wi	標準差と 重みの積 di*wi	標準差の二乗 と重みの積 di^2*wi
	無発病 個体数 nh1	発病 個体数 ni1	調査 個体数 nil	発病割合 pi1	無発病 個体数 nh2	発病 個体数 ni2	調査 個体数 ni2	発病割合 pi2								
14	108	61	169	0.361	22	8	30	0.267	199	0.094	0.347	0.653	0.41623	5.77087	2.40201	0.99979
15	56	24	80	0.300	150	45	195	0.231	275	0.069	0.251	0.749	0.36834	10.66210	3.92727	1.44657
16	43	20	63	0.317	98	34	132	0.258	195	0.060	0.277	0.723	0.29907	8.53932	2.55385	0.76378
17																
18	全体的な標準差の推定値 d_bar				0.3557											
19	全体的な標準差の推定値の標準誤差 SE[d_bar]				0.2001											
20																
21					X二乗値				自由度				p値			
22	X二乗値total				3.21013				3				0.36035			
23	X二乗値homoe				0.05023				2				0.97520			
24	X二乗値assoc				3.15990				1				0.07547			

【第4図】 Excelによる Cochran-Mantel-Haenszel 検定の実際

### 5. データ数と検定結果との関係

第5図は第2図に示した結果とパーセントデータは同一であるが、各処理および反復で調査個体数を同一(この場合はすべて100)にした場合の検定結果である。全調査個体数は第2図の場合の半分以下と少ないが、 $p$  値は0.04755で、第4図に示した 0.07547 よりも大幅に小さくなっており、0.05 よりも小さいことから 95%の確率で薬剤間に効果差があると判定されている。調査個体数をできるだけそろえることは検出力を高めるために必要なことで、事前の試験設計の段階で十分に配慮しなければならないことである。

一方で、調査個体数をそれぞれ200に増やすと  $p$  値は 0.00363 (計算過程は省略)と小さくなり、99%水準で有意差がみられるようになる。このように調査個体数は検定結果に大きく影響することから、試験の実施にあたっては適切な調査個体数の設定が重要である。

32	薬剤A				薬剤B											
	無発病 個体数	発病 個体数	調査 個体数	発病率	無発病 個体数	発病 個体数	調査 個体数	発病率								
34 反復1	64	36	100	36.0	73	27	100	27.0								
35 反復2	70	30	100	30.0	77	23	100	23.0								
36 反復3	68	32	100	32.0	74	26	100	26.0								
37 計	202	98	300		224	76	300									
38 平均				32.7				25.3								
39																
40																
41	薬剤A				薬剤B				合計調査 個体数	発病割 合の差	二群を込み にした発病割 合 $\bar{p}_i$	$q_i - \bar{p}_i$	標準差 di	重み wi	標準差と 重みの積 di*wi	標準差の二乗 と重みの積 di^2*wi
	無発病 個体数	発病 個体数	調査 個体数	発病割合	無発病 個体数	発病 個体数	調査 個体数	発病割合								
42																
43 反復1	64	36	100	0.360	73	27	100	0.270	200	0.090	0.315	0.685	0.41710	10.78875	4.50000	1.87696
44 反復2	70	30	100	0.300	77	23	100	0.230	200	0.070	0.265	0.735	0.35939	9.73875	3.50000	1.25786
45 反復3	68	32	100	0.320	74	26	100	0.260	200	0.060	0.290	0.710	0.29140	10.29500	3.00000	0.87421
46																
47	全体Bの標準差の推定値				0.3569											
48	全体Bの標準差の推定値の標準誤差				0.1801											
49																
50																
51	x 二乗値total		4.00903		自由度		3		p値		0.26049					
52	x 二乗値homog		0.06332		自由度		2		p値		0.95919					
53	x 二乗値assoc		3.92570		自由度		1		p値		0.04755					

【第5図】各処理および反復で調査個体数が同一であるデータに Cochran-Mantel-Haenszel 検定を適用した事例

〔注〕発病果率は【第1図】の場合と同一であるが、調査個体数を各処理および反復でそれぞれ100果実にそろえている。

### 6. Cochran-Mantel-Haenszel 検定の適用場面

本検定法は複数の4分表を統合評価するものである。このため、今回、紹介したように試験圃場が複数のブロックに分けられて、その中に適切に各処理区が配置されている場合に適用できるのはもちろんのこと、同一の設計に基づいて実施された複数場所のデータや異なった年次のデータを統合評価する場合、すなわちメタ・アナリシスを行う場合にも利用できる。

なお、当然のことであるが、反復なしの、すなわちブロック化されずに、例えば「1処理3樹、反復なし」で試験が実施されているような場合には本検定法を用いることはできない。このような場合は、1処理あたりの供試樹(この場合は3樹)のデータをまとめて $\chi^2$ 検定を行うことになる。

ところが、実際に第1図のデータについて $\chi^2$ 検定を行うと「99%水準で有意である」という結果になり、Cochran-Mantel-Haenszel 検定の場合とは異なった結果になっている(第6図参照)。この原因として、各調査樹における調査個体数のばらつきの問題がある。つまり、全部をこみにしてしまうと発病の樹体間差が考慮されなくなってしまう、調査数が多い樹のデータに全体が影響されてしまう。第6図に示している例でみると、処理間の発病率の差は全体をこみにすることによって表のデータに比べて7.4%から9.3%へとひらくため、差が検出されやすくなっている。調査個体数が各樹で同じならばこのような問題は起きないので、処理区全体をこみにして $\chi^2$ 検定を行う場合もできるだけ各調査対象の調査数をそろえることが重要になってくる。第6図の場合、調査数を3樹とも100果とした場合の $\chi^2$ 検定ではp値は0.04778となって95%水準で有意ではあるが、そろえる前の0.00809よりも大幅に大きくなって第5図に示した結果とほぼ同じになっている。

以上のように同一処理区内で調査個体数にばらつきが大きく、さらに傾向が異なっている場合に、区全体をこみにして $\chi^2$ 検定を行うとほんとうの結果がみえてこないもので注意が必要である。

	A	B	C	D	E	F	G	H	I
1	実測値								
2		発病果数	健全果数	計	発病率(%)				
3	薬剤A	105	207	312	33.7				
4	薬剤B	87	270	357	24.4				
5	計	192	477	669					
6	期待値								
7									
8		89.542601	222.4574						
9		102.4574	254.5426						
10	< Yatesの補正 >								
11	観測値と期待値の差								
12		15.457399	-15.4574			n		669	
13		-15.4574	15.457399			ad-bc		10341	
14						n/2		334.5	
15	(観測値と期待値の差)の二乗/期待値								
16		2.6883521	1.0740537			分子部分		6.7E+10	
17		2.3320052	0.8386688			分母部分		1.02E+10	
18									
19	$\chi^2$ 値			7.01308		補正済 $\chi^2$ 値		6.56671	
20	p: $\chi^2$ 分布の片側確率の値			0.00809		p: $\chi^2$ 分布の片側確率の値		0.01039	

【第6図】第1図に示す3反復のデータを合計した場合の $\chi^2$ 検定結果

【注】p値は0.01よりも小さいので99%の確率で「2群の平均値は等しい」という帰無仮説は棄却され、「薬剤間に効果差がある」ことがわかる。



データの大きさと各カテゴリーのばらつきによって同一パーセントのデータであっても検定結果は大きく異なってくる。このため、頻度データをパーセントデータに変換して検定が行われている場合には実際のデータをよくみる必要がある。データ数やそのばらつきに問題がないような場合であっても基本的には今回示した Cochran-Mantel-Haenszel 検定を適用すべきである。具体的なデータがない場合、パーセントデータとそれに基づく検定結果で効果差を判断することには問題があることを強調しておきたい

ところで、今回示したデータから薬剤Aと薬剤Bとの効果差はあると言えるのだろうか。検定結果は得られたデータの量とそのばらつきおよびそれを判定するための検定手法によって異なってくる。あくまでも、その時に得られたデータから最適な検定手法を用いて判断されるべきものであり、同じような傾向を示しているデータであっても場合によっては検定結果に違いがみられることになる。このため、現場に役立つ正確な情報を得る、あるいは提示するためには1回、あるいは1か所の試験データだけでは不十分で、複数のデータを統合評価して判断していくことが大切であり、そこにメタ・アナリシスの意義がある。

なお、処理間でみられる効果差はその値が農業現場で意味のあるものなのかどうかということが最も重要な判断基準であることは言うまでもない。有意差が認められるから意味があるとは言えないわけで、データを見る場合には常にこのことを考えることが大切である。

Mann-Whitney (マン・ホイットニー)の  $U$  検定

病虫害防除技術の有用性を判定する際に、多くの場合は頻度データ、すなわち、発生の有無のデータで評価できる。しかし、発生率が同じような値をとる場合、発病程度別、あるいは被害程度別に調査したデータを比較することは、処理間の差を明らかにする上で有用なことが多い。このため、発病(発生、あるいは寄生)状況をいくつかのカテゴリー(例えば、無、少、中、多、甚の5段階や病斑数0、同1~3、同4~11、同11~30、31以上の5段階など)に分類した発病程度別のデータがとられることになる。これらのデータは順序カテゴリーデータ(大小関係のある分類データ)と呼ばれる。

では、順序カテゴリーデータを用いて2群、あるいはそれ以上の多群の比較を行うにはどうしたらよいのだろうか。2群の平均値の差を検定する場合には、 $t$  検定が用いられるのが一般的である。しかし、 $t$  検定を使うことができるのは変数が間隔尺度や比例尺度のような連続尺度の場合で、さらにそれらが正規分布していることが前提である。一方、順序カテゴリーデータにおけるデータの順序関係は連続したものではなく、とびとびになっているので、平均値を求めたとしてもそれには意味がないため平均値を用いた検定はできない。

このため、順序カテゴリーデータを比較する場合には、「中央値の差の検定」手法である「Mann-Whitney (マン・ホイットニー)の  $U$  検定」を用いなければならない。 $t$  検定で用いる統計量が平均値と標準偏差であるのに対して、この手法では、変数を小さい順から並べたときの順位を統計量として解析を行うことになる。このように本法はグレード分布の中心位置の「ずれ」を検出する手法で、平均値の「ずれ」を検出するものではない点に注意が必要である。

## 1. 順位と順位和

順位はデータを小さい順から並べていったときの順番のことである。第1図の例では、まず最初の階級である発病程度0の順位を1として、これを最小順位とする。発病程度0のカテゴリーには全部で145個体があるので、このカテゴリーにおける最大順位は $1+145-1=145$ になる。次の階級値である発病程度1のカテゴリーの最小順位は一つ前の最大順位に1を足せばよいので、 $145+1=146$ になる。これらの操作を繰り返していくことによって、各カテゴリーの最小順位、最大順位を求めることができる。なお、各カテゴリーは同じ値の集団であるが、その平均順位として各階級の最小順位と最大順位とを足して2で割った値を用いるのが慣例になっている。この値に各カテゴリーの個体数を乗じたものが各グレードの順位和で、それらを合計したものが全体の順位和になる。

	A	B	C	D	E	F	G	H	
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
		発病程度別個体数					合計		
	供試薬剤	0	1	2	3	4			
	A剤	145	73	18	14	3	253		
	累積度数	145	218	236	250	253			
	累積相対度数	0.573	0.862	0.933	0.988	1			
	最小順位	1	146	219	237	251			
	最大順位	145	218	236	250	253			
	平均順位	73	182	227.5	243.5	252			
	各グレードの 順位和	10585	13286	4095	3409	756			
	全体の順位和						32131		

第1図 順序カテゴリーデータにおける順位和の算出

## 2. U検定の考え方

順序カテゴリーデータによって構成されている二つの群がある(第2図上段の表)。この二つの群の各々のデータには対応がないので、それぞれの組み合わせのデータから差を求めることはできない。このため、一標本に還元することができないので、これらの2群の標本に含まれるデータをまとめて、値の小さなほうから大きなほうへと順位をつけ、一つの群の順位和を  $R1$ 、もうひとつの群の順位和を  $R2$  とする。

このとき、両群のグレード分布の中心位置が同じであるならば、両群の順位和(各群の例数が異なる場合は順位平均)は等しくなる( $R1 = R2$ )。これに対して、中心位置がずれていると両群の順位和(各群の例数が異なる場合は順位平均)は当然異なる( $R1 \neq R2$ )はずである。そこで、このことを利用して、両群のグレード分布の中心位置が等しいかどうかを調べる手法が U検定である。

## 3. U検定の手順

第2図にしたがってU検定の手順を示す。

①A剤群(全個体数  $n1=253$ )とB剤群(全個体数  $n2=281$ )とをあわせて全体の発病程度別個体数をみると、7行目のようになる。発病程度0のカテゴリーの合計は  $145+151=296$  になるので、このカテゴリーの最小順位は1、最大順位は  $1+296-1=296$ (C9)である。次の発病程度1のカテゴリーの最小順位は発病程度0の最大順位296に1を加えた297になる(D8)。この手順を繰り返すことによって各発病程度の最小順位と最大順位を求める。

②さらに、各カテゴリーの平均順位を求めるために同じ列の最大順位と最小順位を足して2で割ると、10行目のようになる。

③各カテゴリーの平均順位に各カテゴリーのそれぞれの個体数を乗じることによってA剤順位の小計、B剤順位の小計を算出すると、11行および12行のようになる。

④それぞれの群の順位和、すなわちA剤の順位合計  $R1(=SUM(C11:G11))$ とB剤の順位合計  $R2(=SUM(C12:G12))$ を求め。

⑤ $R1+R2=(n1+n2)(n1+n2+1)/2$  になることを確認する(検算)。すなわち、それぞれの順位和の合計と全体の順位和が同じ値になっているかどうかを調べる。左右が同じ値であれば、これまでの計算が正しかったことが確認される。

⑥U統計量は以下のように定義されている。

$$U1=n1*n2+n1(n1+1)/2-R1$$

$$U2=n1*n2+n2(n2+1)/2-R2$$

このため、 $U1=39081$ 、 $U2=32012$ になる。U検定では、小さいほう(=MIN(C22:C23))の値を  $U_{cal}$  として用いる。

⑦ $U_{cal}$  の分布は平均値  $\mu$ 、標準偏差  $\sigma$  の正規分布に近似することがわかっており、それぞれは以下の式で定義される。

$$\text{平均値 } \mu = (n1*n2)/2$$

$$\text{標準偏差 } \sigma = \sqrt{\frac{n1*n2(n1+n2+1)}{12}}$$

⑧正規分布に近似している  $U$  の分布を平均値と標準偏差を用いて  $z = (U_{cal} - \text{平均値}) / \text{標準偏差}$  という変換をして、標準化を行う。すなわち、

$$z = \frac{|U_{cal} - (n1 * n2 / 2)|}{\sqrt{n1 * n2 (n1 + n2 + 1) / 12}}$$

なので、 $z = 1.9853$  ( $C30 = \text{ABS}(C27) / \text{SQRT}(C28)$ ) になる。この値は標準正規分布における  $X$  軸の値を示しており、 $z$  が 1.9853 以上になる確率  $p$  を求める。 $p = 1 - \text{NORMSDIST}(z)$  になるので ( $C32$ )、これと有意水準 5% あるいは 1% と比較して検定を行う。この場合は、 $p = 0.023555$  になり、 $p < 0.05$  なので、有意水準 5% で帰無仮説 (両群の順位平均は等しい) が棄却され、薬剤間の発病には有意な差がある、すなわち薬剤間で効果差があるといえる。

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4											
5											
6											
7	合計		296	123	53	38	24				
8	最小順位		1	297	420	473	511				
9	最大順位		296	419	472	510	534				
10	平均順位		148.5	358	446	491.5	522.5				
11	A剤順位の小計		21533	26134	8028	6881	1567.5	64143			
12	B剤順位の小計		22424	17900	15610	11796	10973	78702			
13											
14											
15	A剤の調査個体数 n1		253					A剤順位合計 R1	64143		
16	B剤の調査個体数 n2		281					B剤順位合計 R2	78702		
17	n1 * n2		71093								
18	n1 + n2		534					検算			
19	n1 + n2 + 1		535					R1 + R2	142845		
20								(n1 + n2 * (n1 + n2 + 1)) / 2	142845		
21											
22	U1		39081								
23	U2		32012								
24											
25	Ucal (U1, U2の小さい方の値)		32012								
26											
27	Ucal - (n1 * n2 / 2)		-3535								
28	n1 * n2 * (n1 + n2 + 1) / 12		3169562.917								
29											
30	z		1.98530985								
31											
32	p		0.023555								
33											

  

供試薬剤	発病程度別個体数				合計	発病率	発病度
	0	1	2	3	4		
A剤	145	73	18	14	3	253	42.7
B剤	151	50	35	24	21	281	46.3

  

第2図 ExcelによるU検定の手順

4. U検定と $\chi^2$ 検定の組み合わせによるデータの比較

順序カテゴリー分類による2群を比較する場合、今回説明した「Mann-WhitneyのU検定」を行う。この検定はあくまでも全体の差があるのかどうかを検定するものなので、さらに、例えば発病程度2以上の実被害を生じているものの割合や発病程度3以上の重症個体数の割合などの比較を2x2分割表の $\chi^2$ 検定で求めることによって、どの発病程度での差が大きいのかどうかを見出すことができる(表)。表に示す事例では、U検定で薬剤間に効果差がみられるが、発病割合の比較では薬剤間に $\chi^2$ 検定による有意な差は認められない。そこで、どこの割合の差が大きいのかをみると、発病程度

2以上の割合および発病程度3以上の割合については有意差のあることがわかる。すなわち、薬剤間で発病割合の差は認められないが、発病程度2以上の発病が激しいものについては差があることがわかる。

表 U検定と $\chi^2$ 検定の組み合わせによる薬剤の評価

供試 薬剤	発病程度別個体数					合計	発病率	発病度	U検定	$\chi^2$ 検定		
	0	1	2	3	4					全体の発病 割合	発病程度2 以上の割合	発病程度3 以上の割合
A剤	145	73	18	14	3	253	42.7	12.3	P=0.0235	P=0.4065	P<0.01	P<0.01
B剤	151	50	35	24	21	281	46.3	21.5				



これまで、病虫害防除技術開発の分野では、順序カテゴリーデータによる評価を行う場合、複数の現象(発病程度が甚, 多, 中, 少, 無・・・など)に応じて点数を付け、それぞれの現象が出現する頻度とかけあわせてスコア化して、発病度や発生度を算出して比較する手法がとられてきた。しかし、この場合はサンプルサイズが検定結果に反映しないという問題があり、さらに発病度や発生度というものの意味するところがいったい、なんであるのかははっきりしないことが多く、実用的な意味がどれだけあるのかという疑問があり、単なる数字の比較になっているようにも思われる。

今後、順序カテゴリーを現場における実用的な評価と同じようにすることによって、より現場に即した防除技術の評価を行っていくことが大切である。すなわち、Aという技術では対象病害(虫)の発生頻度は高いが実際の被害は問題にならないとか、逆に、B技術では発生はA技術よりも少ないけれども実際の被害は多くなって実用的ではないとかの評価を行うべきであり、「Mann-WhitneyのU検定」と「 $\chi^2$ 検定」とを組み合わせることによってこのことが可能になる。