

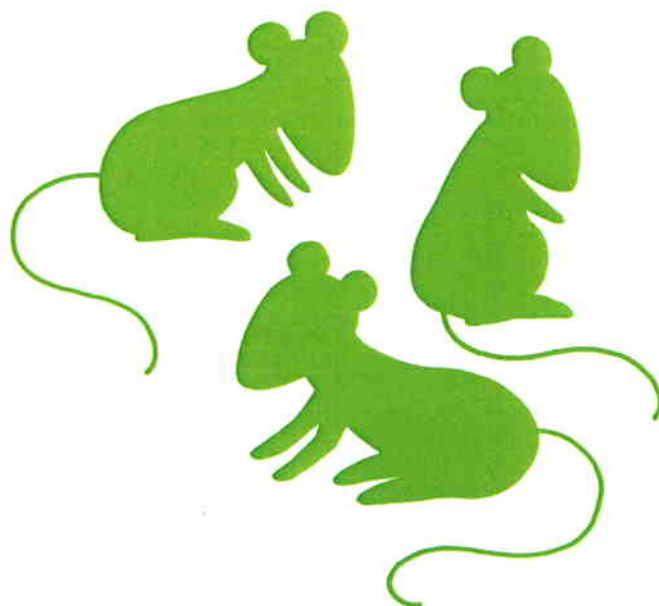
ISSN 0288-2906

# 医薬安全性研究会 会報

*Bulletin of Japanese Society for Biopharmaceutical Statistics.*

Sep.1996

No.42



サイエンティスト社

# 医薬安全性研究会 会報

第42号 1996年9月

## 目 次

### \* 医薬安全性研究会スケジュール

---

- 企業における統計教育の事例報告  
——薬理研究部門における——  
山岡 英明 ..... 1
  - ピアソン・ $x^2$ 、イエーツの補正、フィッシャーの  
直接確率計算法、どこがちがうのか  
——mid-*P*value の出番はあるか——  
岩崎 学 ..... 7
  - ラット体温に及ぼす解熱薬の影響  
——枝分かれ型経時測定データの統計解析(Dunnett 検定と Steel 検定)——  
杉山 公仁、佐々木秀雄 ..... 19
- 
- 事務局だより ..... 25

編 集

医薬安全性研究会

定価 1,030円

# 企業における統計教育の事例報告

－薬理研究部門における－

住友製薬株式会社 山岡秀明

## 1. はじめに

住友製薬総合研究所（以下「研究所」）は、1994年暮れ「生物統計ハンドブック」作成のプロジェクトを発足させ、翌年春それを完成し、以来研究所内での統計処理手法の統一を進めている。

筆者は当時、関連する住友化学工業株式会社生物環境科学研究所に所属していたが、依頼に応じて本プロジェクトに参加した。参考になればと思い、その活動概要を報告する。

## 2. 生物統計ハンドブックの作成

これまでは、実験データの統計処理手法および統計ソフトウェアの選択は個々の研究者にゆだねられていたが、研究所内でまず手法を統一することにした。

その方法として、適当な参考書（例えばピンク本「毒性・薬効データの統計解析」サイエンティスト社）を所内の標準とし、その中のどの検定を使えば良いかを勉強会等で徹底すればよい、とする意見が出た。一方、成書では実験の現場では使いにくい、また本には多様な検定手法が紹介してあるので不適切な検定を使う者が出て来る可能性があり、統制の維持が困難ではないか、等の反対意見も出された。

結局、現場の使い勝手を重視した社内用「生物統計ハンドブック」を作成することにした。内容的には、統計屋が勝手に作って研究者に無理強いするものではなく、実験現場でのデータの出かた（パターン）に応じて、どのような手法を使うべきかを分かりやすく示すことにした。また、研究者が日常的に統計処理に対して感じている具体的な問題意識等についても調査し、できるだけ多くの疑問に対して回答しておこうと考えた。

そこで、各研究グループから代表者に出してもらうことにした。各研究グループのリーダーは、代表として選ぶべき研究者はどんなレベルで、どのような素養の者が良いか、また本件のためにどれだけ拘束されるか、を気にした。そのため、本プロジェクトの趣旨、代表者が必要な理由及びかれらの役割、「ハンドブック」完成予定の努力目標を3カ月としていること、等を明示して、代表者の選出を依頼した。

一方、医薬開発部門には臨床統計専門家チームがあった。そこにも協力を要請し、本プロジェクトに加わってもらった。

計10名で活動を開始した。2～3週ごとにミーティングを開いた。

まず、各研究グループの実態調査をし、実験データのパターン、使っている統計手法(含検定)、パソコン機種と統計ソフト、問題点・疑問点等を洗い出した。

一方、同時にプロジェクトチームの主だった者が中心となり、検定手法の基本方針を検討した。多重比較の前にANOVAを入れた方が良いか否か? 多時点の解析では、時点の多重性を考慮すべきか否か? 等々。それらの問題は、統計学の専門家でも議論のあるところと聞く。従って、我々のレベルでの判断方法としては、「△△先生がこう言われた。」「じゃあ、そうしよう。」というやり方を採った。

当プロジェクト進行中、東京理科大の芳賀敏郎先生が研究所にご講演にお越しになる機会があり、その際、当ハンドブック(案)についてもご指導いただいた。先生には、非常に初期段階の草稿にもかかわらず、お目を通していただき、大変ありがたかった。とりわけ、「概ね問題ない」こと、及び「完璧でなくても、とにかくハンドブックをリリースすることが大事であり、それを活用しながら改善していけばよい。」とのアドバイスを頂いたことで、我々プロジェクトメンバーは大変気が楽になり、ハンドブック作成作業を一挙に加速して、とりあえず初版を完成させることができた。

「生物統計ハンドブック ver.1.0」完成後は、それをすべての研究グループに配布し、研修会・自習等で利用してもらい、統計処理手法の統一・徹底を計った。

### 3. 「生物統計ハンドブック ver. 1.0」

ハンドブックの構成を以下に述べる。

#### (1) 序論

「統計解析は意味のある変化を見つける道具であること。検定結果で真実が決まるのではないこと。絶対的な手法がないから今ある適当なものを注意しながら使おう。統計手法の基本的な選び方を研究所内で統一しよう。」など本ハンドブックの主旨、本書の構成と参考文献、および「スクリーニングテスト」の判断基準は別であること、等をここで述べた。

#### (2) データの種類

実験で発生するデータを、実数、スコア及び頻度の3種類に分類して説明した。

#### (3) データ種類別検定手法の選択

各データ種類ごとに、「対照群対1処置群の場合」、「対照群対多処置群の場合」等に分け、前者は各種1対1比較を、後者については各種多重比較をその特徴とともに解説した。各検定手法に関する記述には具体的な計算式までは含めず、手法を示すにとどまり、実際のデータ処理には既成のパソコンソフトを利用してもらうことにした。

事例として、対照群が複数ある場合の検定方法を、この章で示した。

#### (4) 実験パターン別検定手法の選択

実験データのパターンは事例に基づき、ほぼ次の組み合わせにまとめた。全部で13パターンとなった。(「1群多時点連続測定の場合」を加えた。)

{実数、スコア、頻度の別} × {2群または多群} × {1時点または多時点}

例えば、パターン△は、「実数データ/2群比較/多時点」の処理で、検定手法としては○○がある。事例としては「経時的体温変化の2群比較」があり、実際の実験グラフは図△のようになる、と言う風に、統計手法から事例までを対応させている。これが、意外に現場で受けているらしい。

#### (5) 利用できるソフトウェア

第5章で述べる。

#### (6) Q&A

各研究グループへの実態調査で洗い出した問題点・疑問点、等をまとめた。「データの基本的扱い方」、「S.D.、S.E.、バラツキ」、「検定」、「ノンパラデータ」などに分類し、ピンク本、地人書館の「毒性試験講座14」、等で調べ、できるかぎりの解説をした。

### 4. 生物統計ハンドブックの活用

当ハンドブックの効用をまとめると、次のようになる。

- (1) 「ハンドブック作成」という具体的目標を設定し、各研究グループの代表者の範囲で統計学の勉強ができた。
- (2) 研究所内の基本的共通ルールを構築した。
- (3) 研究者研修や自習用のテキストとして利用できる。
- (4) 議論のたたき台、または共通の土俵とすることができる。
- (5) 定期的な見直し(ハンドブック利用実態調査を含む)や改訂作業をすることにより、研究員の統計に対する意識を維持することができる。

## 5. 統計ソフトの選択

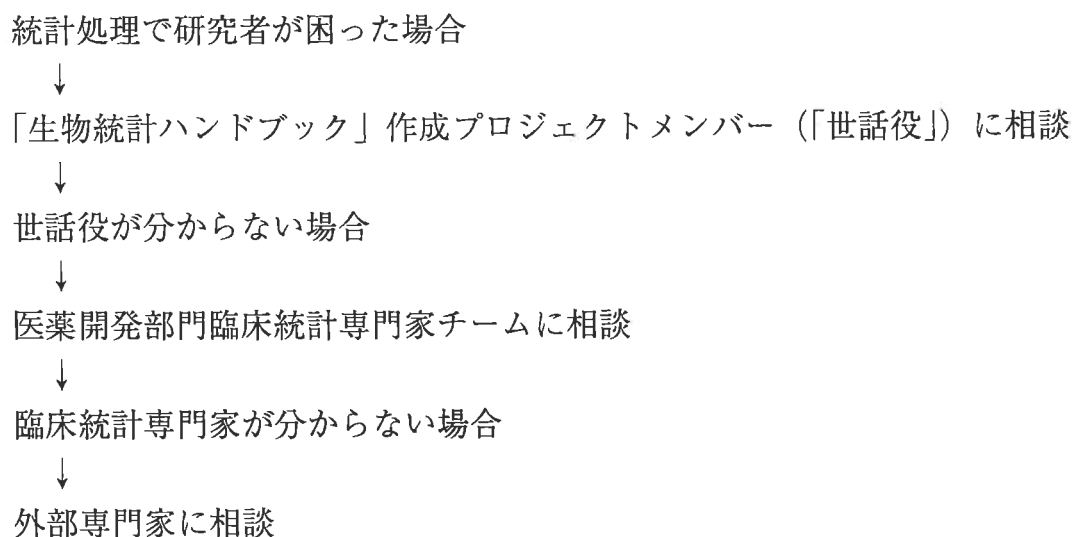
各研究グループの実態調査によって、パソコン及び統計処理プログラムの使い方が、バラバラであることが明らかになった。大きな問題として、マッキントッシュとウィンドウズマシンが混在しており、それぞれの機種でも、「Super ANOVA」、「Stat View」、「YUKMS統計ライブラリ」、「LOTUS1-2-3 自作マクロ」など、グループによって違う種類の統計ソフトが使われていた。

当プロジェクトチーム内でも、この事態にどう対応するか大いに議論した。これまで各研究グループで購入した統計ソフトを無駄にしないようにとの意見もあったが、ハンドブックによってどの検定手法を使えばよいか示しても、どのパソコンのどの統計ソフトならそれが出来るか、または出来ないか示さなければならない。グループによっては、推奨する検定手法が使えない場合も出てくる。簡単な検定でもソフトによってはp値が微妙に異なる場合がある。また、各統計ソフトの信頼性を担保するにはどうしたらよいか、等の問題があった。

統計ソフトを統一することにすればどうなるか？ まず、新規ソフトの購入費用が必要となる。でも、研究者にとっては、分からないことがあればグループを越えて相談できる仲間が増える。また、今後、我々プロジェクトメンバーが研究現場からの質問・問題を処理する場合でも、雑多な統計ソフトとかかわるよりも、あるひとつのソフトについて知っていれば、よりきめ細かに相談に乗ることが出来ると考えた。

結局、予算のメドがつくなら、今後のために統計ソフトを統一した方が良い、ということになった。

我々は、今回のプロジェクトが目標を達成した後の研究所内統計学的諸問題解決システムとして、次のような問題処理フローを試案した。



このプロセスを円滑に進めるためには、統計ソフトをそろえる必要があった。世の中には、SPSSとかUNIX-Sとか言った有名なものから、パソコン用ソフトまで、文字通り「ピンからキリまで」の統計ソフトがあったが、我々の臨床統計専門家は「SAS」という統計ソフトを利用していた。SASはかなり何でも出来る統計処理ソフトで、存知あげている統計学の先生方の多くも、それを利用されていると聞く。また、SASは世界的に多様な分野で利用されており、その信頼性は高く、Validationなしで利用できるとも聞いていた。従って、我々の研究所もそれを導入するのが良からうということになった。

だが、予算を獲得しなければならなかった。SASのライセンス料はかなり高額だった。

SASを導入できるまで「生物統計ハンドブック」は「絵に描いた餅」になろうとしていた。生物統計ハンドブックが推奨する検定が実際には出来ない研究グループも出てくるだろうし、世話役も研究者のこのような問題に十分対応出来ないであろうことは目に見えていた。SASの導入までもっと安くて買ってもらいやすいパソコン統計ソフトを入れて急場をしのぐ、という選択肢も検討した。しかしそれを購入して所内の一応の用が足りたら、かえってSASの購入許可が遠のいてしまうのではないかという矛盾の恐れがあった。

事態の救世主は、今溝 裕さん(旭化成)の「MultiStaff」だった。MultiStaffはLOTUS 1-2-3のアドイン統計ソフトとして開発され、既にかかなり使い込まれ、改良が加えられていて、なにより我々の生物統計ハンドブックが推奨するほとんどすべての検定手法を網羅していた。決定的なのは、それが無料で配布されていたことである。我々の研究所内ではLOTUS 1-2-3がかかなり行き渡っていたので、MultiStaffを導入する下地は十分であった。

我々は、SASを入れるまで、今溝さんの快諾のもとに「MultiStaff」を当面研究所内の標準統計ソフトとして利用することにした。

同時にSAS購入の予算申請をした。はじめはパイロット的に少数セットを導入することにし、主にプロジェクトチームメンバーが本当に使えるかどうかある程度評価できた後、残りの全研究グループ分を導入することにした。

## 6. 今後の方向

「SASは何でも出来るが、使いにくいらしい。」と、プロジェクトメンバーのひとりが聞きつけて来た。臨床統計の専門家は、「慣れれば、どうってことないよ」となだめたが、結局、不慣れな研究者や研究補助者でも迷うことなく使えるように「メニュー方式」でSASが使えるようにすることを、当プロジェクトチームの次の課題として検討することにした。それまではとりあえず、検定手法ごとに所定の統計処理が出来るSASプログラムをパソコンネットワークで研究者に供給し、研究者自らが実験データを「カット&ペースト」でプログラムの中に埋め込んで統計処理する方法を普及させることにした。

## 7. 最後に

我々の「生物統計ハンドブック」を生かすも殺すも、研究現場の当事者の意識にかかっていると思う。「ver.1.0」を完成させた以降も、それを枯れさせず研究所内で維持し育ててもらいたい。

以 上



ピアソン  $\chi^2$ , イエーツの補正, フィッシャーの  
 直接確率計算法, どこが違うのか  
 - mid-P value の出番はあるか -

成蹊大学工学部 岩崎 学

1. はじめに

N 人の被験者に関する比較実験を, 対照群には  $n_1$  人, 処置群には  $n_2$  人とランダムに振り分けて実施した結果が, 次のような  $2 \times 2$  の分割表にまとめられたとする.

表 1.1  $2 \times 2$  の分割表

	—	±<	計
対照群	a	b	$n_1$
処置群	c	d	$n_2$
計	$m_1$	$m_2$	N

ここでの解析の目的は

$$p_1 = \Pr\{- | \text{対照群}\}, p_2 = \Pr\{- | \text{処置群}\} \quad (1.1)$$

としたとき, 帰無仮説

$$H_0 : p_1 = p_2 (= p \text{ と置く}) \quad (1.2)$$

を

$$H_1 : p_1 > p_2 \text{ (あるいは } p_1 < p_2 \text{)} \text{ (片側仮説)} \quad (1.3)$$

もしくは

$$H_2 : p_1 \neq p_2 \text{ (両側仮説)} \quad (1.4)$$

に対して検定することにある.

たとえば, 具体的な数値例が表 1.2 のようになったとする.

表 1.2  $2 \times 2$  の分割表の数値例

	—	±<	計
対照群	8	2	10
処置群	4	6	10
計	12	8	20

神谷氏の計算では p 値が表 1.3 のようであったとのことである。

表 1.3  $\chi^2$  検定および Fisher の直接確率計算による P 値の違い

出現頻度				P 値	
8	2	4	6	$\chi^2$ 検定 (Yates の補正あり) 0.068 (0.171)	Fisher の直接確率計算 0.085

同じデータを手元の統計ソフトウェア (StatView) にかけてみると表 1.4 のような出力が得られる。

表 1.4 表 1.2 のデータに関する StatView の出力

分割表分析統計量 : 行、列

欠測値数	0
自由度	1
カイ 2 乗値	3.333
カイ 2 乗 p 値	.0679
G 2 乗値	3.452
G 2 乗 p 値	.0632
分割表分析係数	.378
ファイ	.408
カイ 2 乗値 (Yates 補正)	1.875
カイ 2 乗 p 値 (Yates 補正)	.1709
Fisher の直接法 p 値	.1698

ここでは p 値が計 4 種類出てきて、いったいどれを使えばよいのかが分からない。また、数値が表 1.3 とは異なるものもあり、ますます混乱してしまう。

本稿では、「カイ二乗 p 値」、「カイ二乗 p 値 (Yates 補正)」、「Fisher の直接法 p 値」の違いを、片側検定、両側検定の区別を交えて議論し、さらに、mid-P value についても簡単に触れる。なお、mid-P value についてさらに詳しくは岩崎 (1993) を参照されたい。

## 2. 検定統計量および p 値の定義

本節ではまず、それぞれの検定統計量の導出公式ならびに p 値の計算法を確認しておく。記号は表 1.1 のものを用いる。また、以下において、大文字の X は自由度 1 のカイ二乗分

すなわち、(1.2)の帰無仮説を(1.4)の両側仮説に対して検定していることに相当する。

数値例では、カイ二乗値 6.0 に基づく「正確な」p 値は、 $p = 0.5$  の場合、0.03125 である。それに対し、自由度 1 のカイ二乗分布を用いた近似では p 値は 0.0143 となる。

さて、次に、表 3.3 において実測値と同じ列和 ( $m_1 = m_2 = 3$ ) を持つ度数だけを抜き出すと表 3.5 のようになる。

表 3.5 表 3.3 において ( $m_1 = m_2 = 3$ ) となるものだけの抜粋

No.	a	b	c	d	確率( $p=0.5$ )	確率( $p=0.6$ )	確率( $p=0.7$ )	条件付確率
4	3	0	0	3	0.015625	0.013824	0.009261	0.05
7	2	1	1	2	0.140625	0.124416	0.083349	0.45
10	1	2	2	1	0.140625	0.124416	0.083349	0.45
13	0	3	3	0	0.015625	0.013824	0.009261	0.05

これらだけに限った「条件付確率」は、 $p$  の値によらず、0.05, 0.45, 0.45, 0.05 となることが確かめられる。この条件付確率が(2.6)の右辺の超幾何確率となるのである。Fisher の直接確率計算法では、このような  $m_1, m_2$  が実測値と同じになる度数だけを考慮し、条件付確率に基づいて  $p$  値の計算をする。このことから、Fisher の検定は「条件付検定」と呼ばれる。

表 3.1 の数値例では、(3, 0, 0, 3) の  $p$  値は 0.05 と計算される。この 0.05 という値は、上記のカイ二乗検定における  $p$  値と異なり、(0, 3, 3, 0) の確率を含んでいない。このことから、Fisher の検定は「片側検定」であり、(1.2)の帰無仮説を(1.3)の片側仮説に対して検定していることになる。

Fisher の検定を両側検定として用いる場合には、

- (1) 片側  $p$  値を 2 倍する
- (2) 確率が実測値以下の度数に対応する確率を全部足し上げる
- (3) カイ二乗値が実測値以下の度数に関する確率を全部足し上げる

のような方策が考えられる。 $n_1 = n_2$  の場合には両者は一致するが、一般にはそうではない。どちらの定義を用いるかは、ユーザーが決める必要がある。また、手持ちのソフトウェアがどちらの定義で計算しているかはチェックしておく必要があるだろう。

(注) StatView では (2) (あるいは (3)) の定義を用いている。ちなみに、表 1.4 では両側  $p$  値が、表 3.2 では片側  $p$  値が出力されている (!?)。

布に従う確率変数である。

カイ二乗統計量 (補正なし)

$$x = N(ad - bc)^2 / (n_1 n_2 m_1 m_2), \quad (2.1)$$

$$p \text{ 値} = \Pr \{ x \leq X \}. \quad (2.2)$$

カイ二乗統計量 (補正あり)

$$x' = N (|ad - bc| - N/2)^2 / (n_1 n_2 m_1 m_2), \quad (2.3)$$

$$p \text{ 値} = \Pr \{ x' \leq X \}. \quad (2.4)$$

Fisher の直接確率計算

$a^*$  を第 (1, 1) セルの実現値とし,

$$a^* \geq m_1 n_1 / N \quad (2.5)$$

のとき

$$p \text{ 値} = \sum_{a \geq a^*} \binom{n_1}{a} \binom{n_2}{c} / \binom{N}{m_1}. \quad (2.6)$$

(注) (2.5) の右辺は、分割表の行と列とが独立の場合の第 (1, 1) セルの期待値であり、不等号が逆向き ( $\leq$ ) のときは (2.6) における和は  $a \leq a^*$  となる。また、(2.6) の右辺は超幾何分布の確率である。

上記の各公式に表 1.2 の数値を代入すると次のようになる。

カイ二乗統計量 (補正なし)

$$x = 20 (8 \times 6 - 2 \times 4)^2 / (10 \times 10 \times 12 \times 8) = 3.333 \quad (2.7)$$

$$p \text{ 値} = \Pr \{ 3.333 \leq X \} = 0.0679 \quad (2.8)$$

カイ二乗統計量 (補正あり)

$$x' = 20 (|8 \times 6 - 2 \times 4| - 20/2)^2 / (10 \times 10 \times 12 \times 8) = 1.875, \quad (2.9)$$

$$p \text{ 値} = \Pr \{ 1.875 \leq X \} = 0.1709. \quad (2.10)$$

Fisher の直接確率計算

(2.5) の条件をチェックすると

$$8 \geq 12 \times 8 / 20 (= 4.8) \quad (2.11)$$

であるので、(2.6) より、 $a^* = 8$  以上の超幾何確率を加えて p 値を求めることになる。(2.6) の右辺の確率をすべての a について求めると表 2.1 のようになる。ここで、

$$a + b = 10, c + d = 10, a + c = 12, b + d = 8 \quad (2.12)$$

の条件があるため、a の値を定めれば他の b, c, d の値は一意的に決まってくることに注意されたい。

表 2.1 すべての a に関する超幾何確率

a	b	c	d	確率
10	0	2	8	0.000357
9	1	3	7	0.009526
8	2	4	6	0.075018
7	3	5	5	0.240057
6	4	6	4	0.350083
5	5	7	3	0.240057
4	6	8	2	0.075018
3	7	9	1	0.009526
2	8	10	0	0.000357

p 値は、表における上から三つまでの確率を加えて

$$0.000357 + 0.009526 + 0.075018 = 0.084901 \quad (2.13)$$

と求められる。

以上の計算結果から次の諸点を確認しておく。

- (1) 表 1.3 および表 1.4 の補正なしのカイ二乗 p 値は同じ (0.0679) である。
- (2) 表 1.3 および表 1.4 の補正ありのカイ二乗 p 値は同じ (0.1709) である。
- (3) 表 1.4 の Fisher の直接法 p 値 (0.1698) は表 1.3 の p 値 (0.085) の 2 倍である。
- (4) 補正ありのカイ二乗 p 値 (0.1709) は表 1.4 の Fisher の直接法 p 値 (0.1698) にほぼ等しく、したがって、表 1.3 の直接法 p 値 (0.085) の約 2 倍である。
- (5) 表 2.1 における超幾何確率は「左右対称」である。

### 3. カイ二乗と Fisher の直接確率法

本節では、前節で求めた p 値がどのような考えに基づくものであるかを考察する。まず、補正なしのカイ二乗と Fisher の直接確率計算法の違いを見る。議論を分かり易くするため、表 3.1 のようなさらに簡単な数値例を取り上げる。

表 3.1 簡単な数値例

	—	±<	計
対照群	3	0	3
処置群	0	3	3
計	3	3	6

この場合の解析結果は表 3.2 のようである。

表 3.2 表 3.1 のデータの解析結果

分割表分析統計量 : 行、列

欠測値数	0
自由度	1
カイ 2 乗値	6.000
カイ 2 乗 p 値	.0143
G 2 乗値	.
G 2 乗 p 値	.
分割表分析係数	.707
ファイ	1.000
カイ 2 乗値 (Yates補正)	2.667
カイ 2 乗 p 値 (Yates補正)	.1025
Fisherの直接法 p 値	.0500

各群に割り当てられる被験者数 (表 1.1 の記号では  $n_1$ ,  $n_2$ ) は実験計画で定められたものである。各群における「—」の度数は (1.2) の帰無仮説  $H_0: p_1 = p_2 (= p)$  の下では、それぞれパラメータ  $(n_1, p)$  および  $(n_2, p)$  の二項分布に従う。そこで、その二項分布に基づいて求めた確率と、それぞれの分割表におけるカイ二乗値をまとめると次ページの表 3.3 のようになる ( $p = 0.5, 0.6, 0.7$  の場合)。ちなみに、No. 1 および No. 16 では分母分子が 0 となってカイ二乗値が定義されない。

表 3.3 のカイ二乗値が同じとなる確率を加えて、カイ二乗値ごとに確率を計算したのが表 3.4 である (カイ二乗値が定義されない No. 1 および No. 16 のカイ二乗値は便宜上 0 として計算してある)。

表 3.3 表 3.1 のデータに関するすべての場合の確率

No.	a	b	c	d	$m_1$	$m_2$	確率(p=0.5)	確率(p=0.6)	確率(p=0.7)	カイ二乗
1	3	0	3	0	6	0	0.015625	0.046656	0.117649	—
2	3	0	2	1	5	1	0.046875	0.093312	0.151263	1.2
3	3	0	1	2	4	2	0.046875	0.062208	0.064827	3.0
4	3	0	0	3	3	3	0.015625	0.013824	0.009261	6.0
5	2	1	3	0	5	1	0.046875	0.093312	0.151263	1.2
6	2	1	2	1	4	2	0.140625	0.186624	0.194481	0.0
7	2	1	1	2	3	3	0.140625	0.124416	0.083349	0.67
8	2	1	0	3	2	4	0.046875	0.027648	0.011907	3.0
9	1	2	3	0	4	2	0.046875	0.062208	0.064827	3.0
10	1	2	2	1	3	3	0.140625	0.124416	0.083349	0.67
11	1	2	1	2	2	4	0.140625	0.082944	0.035721	0.0
12	1	2	0	3	1	5	0.046875	0.018432	0.005103	1.2
13	0	3	3	0	3	3	0.015625	0.013824	0.009261	6.0
14	0	3	2	1	2	4	0.046875	0.027648	0.011907	3.0
15	0	3	1	2	1	5	0.046875	0.018432	0.005103	1.2
16	0	3	0	3	0	6	0.015625	0.004096	0.000729	—

表 3.4 カイ二乗値ごとの確率

カイ二乗値	p = 0.5	p = 0.6	p = 0.7
0.0	0.31250	0.32032	0.34858
0.67	0.28125	0.24883	0.16669
1.2	0.18750	0.22348	0.31273
3.0	0.18750	0.17971	0.15346
6.0	0.03125	0.02764	0.01852

標本数  $N$  が大きいとき、表 3.4 の確率が自由度 1 のカイ二乗分布で近似されることから「カイ二乗値」の名前が付けられているのである（この例では標本数が少ないため近似はあまりよくはない）。

さて、表 3.1 の実測値 (3, 0, 0, 3) ではカイ二乗値は 6.0 であり、そのときの確率は、 $p = 0.5$  のとき 0.015625 であるが、カイ二乗値が 6.0 となる度数の組み合わせはもう一つある。それは No. 13 の (0, 3, 3, 0) であり、実測値 (3, 0, 0, 3) とは正反対の傾向を示すデータとなっている。このことは、カイ二乗値に基づく検定が「両側検定」であることを示している。

#### 4. Yates の補正

条件付検定における超幾何分布は、 $n_1, n_2$  があまりアンバランスでない限り、正規分布で近似される。図 4.1 は表 1.1 の例題に関する超幾何確率とその正規分布を表わしたものである。

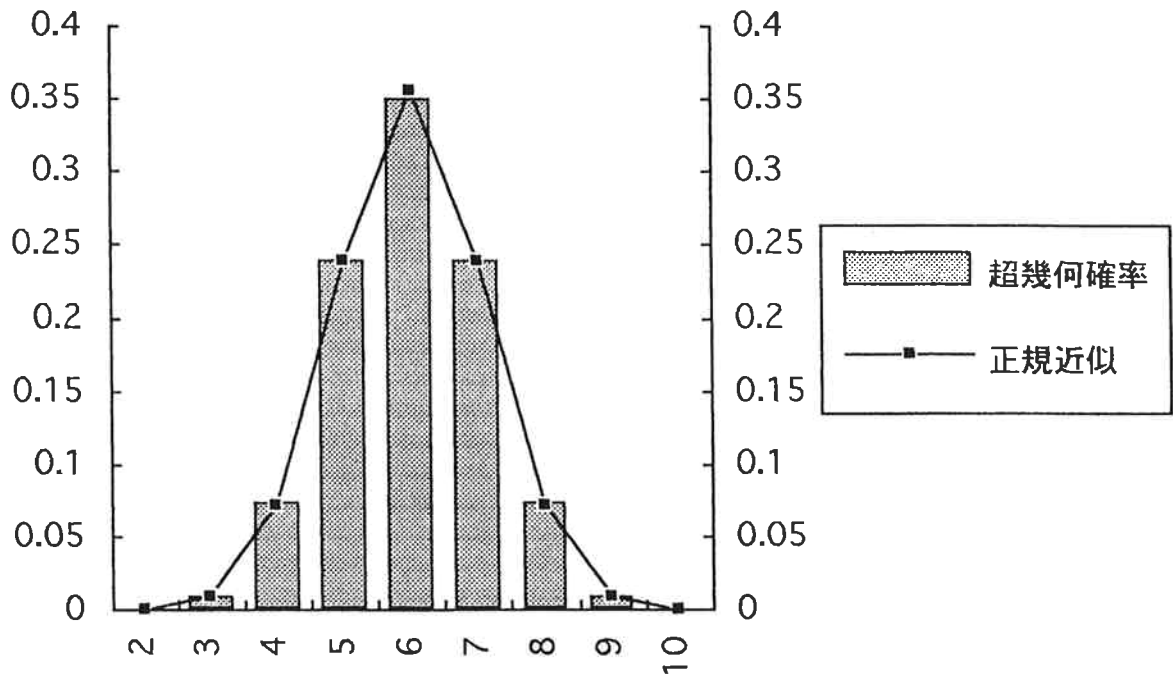


図 4.1 超幾何分布の正規近似の例

一般に、離散分布の確率を連続分布の積分で近似する場合、離散変量を  $X$  とし、連続変量を  $Z$  とするとき、

$$\Pr\{a \leq X\} \doteq \Pr\{a - 1/2 \leq Z\} \quad (4.1)$$

としなければならない(次ページの図 4.1 参照)。 $\Pr\{a \leq Z\}$  では図 4.1 における斜線部分が積分から外れてしまう。この補正を「連続補正」という。

条件付検定である Fisher の直接確率の計算をカイ二乗近似する場合、自由度 1 のカイ二乗分布は標準正規分布に従う確率変数の二乗の分布であるので、上記のような連続補正を施さないと、確率の過小評価になってしまう。その補正が Yates の補正であり、補正を施したカイ二乗  $p$  値は、したがって Fisher の直接確率計算における両側  $p$  値のよい近似となっている。第 2 節の最後の確認事項の (4) を思い起こしていただきたい。

すなわち、カイ二乗  $p$  値の計算において補正を施すか否かは、 $m_1, m_2$  を固定しない「無条件検定」か、それらを固定した「条件付検定」かの選択を意味していると考えられるので



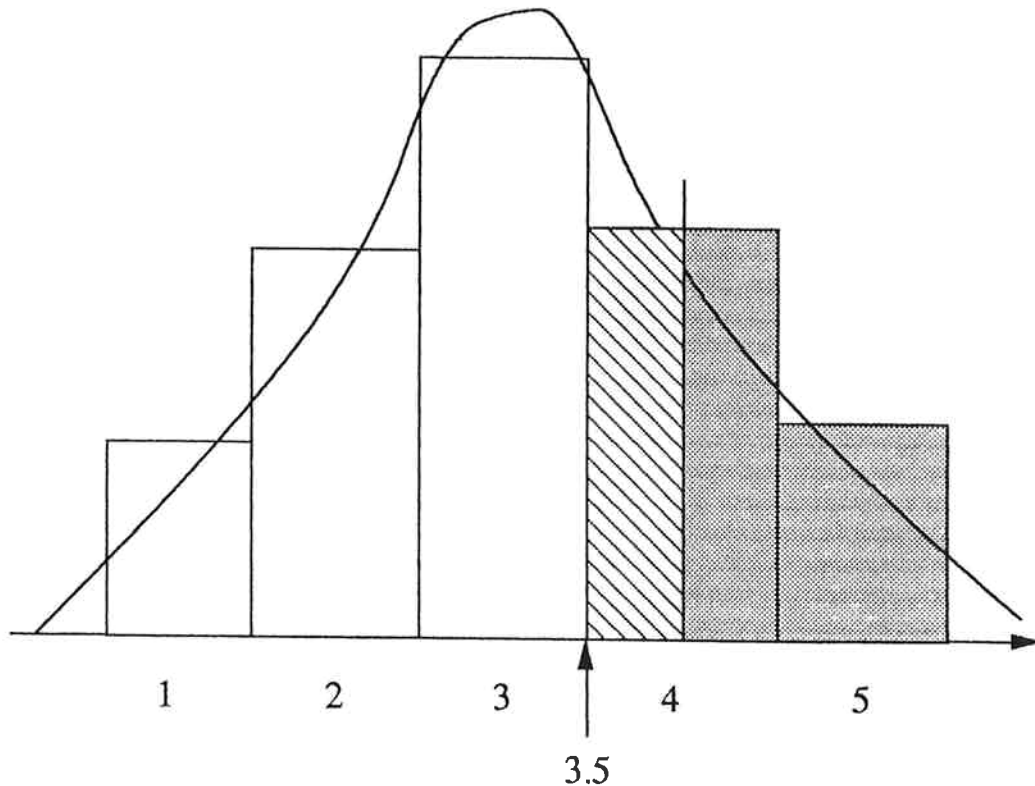


図 4.2 離散分布の連続分布による近似

ある。また、この際、片側検定か両側検定かの区別もきちんとしておく必要がある。

### 5. mid-P value とは

連続的な検定統計量  $T$  に基づく検定では、実現値  $t^*$  に対して、 $p$  値は

$$p \text{ 値} = \Pr \{ t^* \leq T \} \quad (5.1)$$

のように定義される。同じように、離散変量  $X$  の場合にも、実現値  $x^*$  に対して  $p$  値を

$$p \text{ 値} = \Pr \{ x^* \leq X \} \quad (5.2)$$

のように定義するのが自然であると考えられる。Fisher の直接確率法でもこのようにして  $p$  値を導出している。しかし、実務家の間では、『Fisher の  $p$  値は大きすぎて検定が有意にならない』との話がされることがある。

mid-P value は

$$\text{mid-P} = (1/2) \Pr \{ X = x^* \} + \Pr \{ x^* < X \} \quad (5.3)$$

と定義される (図 5.1 参照)。すなわち、実測値の確率を半分しか考慮しないのである。当然

$$\text{mid-P} < p \text{ 値} \quad (5.4)$$

であり、あらかじめ定められた有意水準  $\alpha$  に対し、 $p$  値では有意にならないが mid-P では有意になる、すなわち

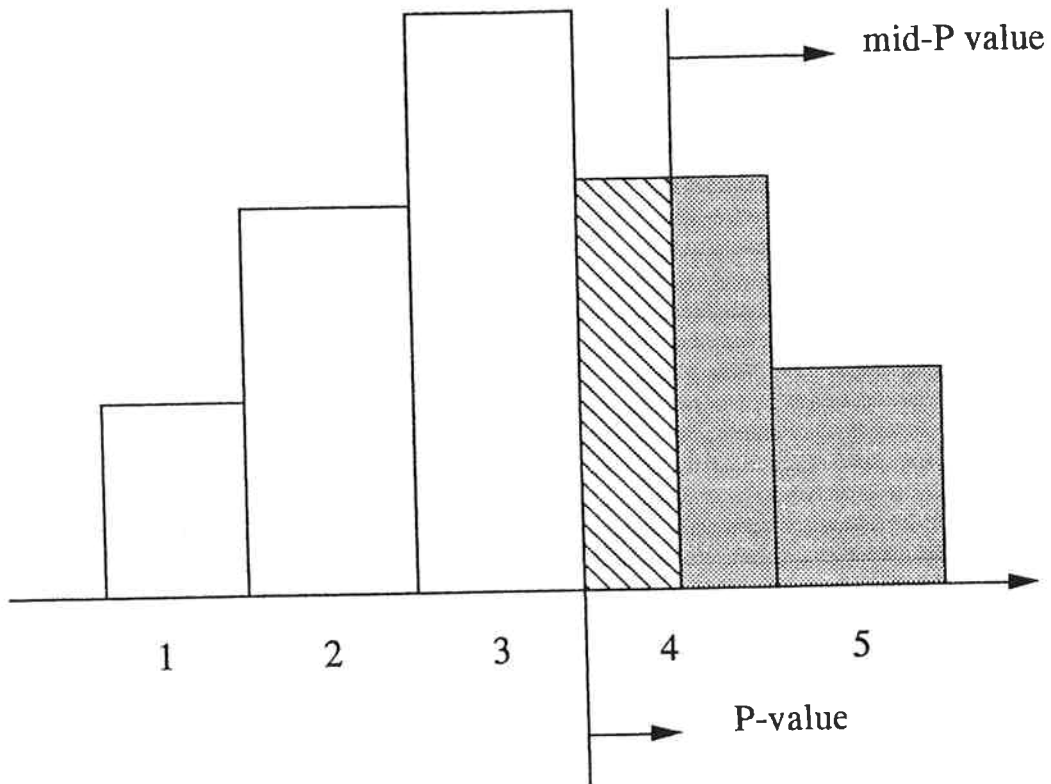


図 5.1 p 値と mid-P value との定義の違い

$$\text{mid-P value} < \alpha < \text{p 値} \quad (5.5)$$

となるケースも考えられる。表 2.1 の確率について p 値と mid-P value を計算すると表 5.1 のようになる。  $\alpha = 0.05$  で  $a = 8$  とするとそれに該当する。

表 5.1 表 2.1 の確率から求めた p 値と mid-P value

a	b	c	d	確率	p 値	mid-P value
10	0	2	8	0.000357	0.000357	0.000179
9	1	3	7	0.009526	0.009883	0.00512
8	2	4	6	0.075018	0.084901	0.047392
7	3	5	5	0.240057	0.324958	0.20493
6	4	6	4	0.350083	0.675042	0.5
5	5	7	3	0.240057	0.915099	0.79507
4	6	8	2	0.075018	0.990117	0.952608
3	7	9	1	0.009526	0.999643	0.99488
2	8	10	0	0.000357	1	0.999821

p 値には「第 1 種の誤りの確率」という意味があるが、mid-P value にはない。しかし、mid-P value を正当化する論理はいくつかある。特に、研究開発の初期あるいは中間段階では mid-P value の出番はあると考えられる。

とにかく、詳しくは岩崎 (1993) の解説を見られたい。

## 6. おわりに

各 p 値の違いについて分かり易く解説してきた (つもりである)。本稿で取り上げた以外に「尤度比カイ二乗」に基づく p 値というものもある (StatView の出力における G 2 乗値)。統計ソフトウェアで、どれがいいのか述べて並記するのは責任回避といわざるを得ないが、統計家の間でも決着がついていない問題であるため止むを得ないともいえる。

解析結果について最終的に責任を追うべき人が選択をすればよい (責任回避だ！)。

## 参考文献

岩崎 学 (1993) mid-P value : その考え方と特性. 応用統計学, 22, 67-80.



ラット体温に及ぼす解熱薬の影響

— 枝分かれ型経時測定データの統計解析 —

Q : 杉山 公仁 (昭和薬品化工)

A : 佐々木 秀雄 (旭化成工業)

解熱薬のラットの正常体温に及ぼす影響を調べるため、対照群および100、200、400mg/kg投与群の4群で各群のn数を12として実験を行った。解熱効果測定と異なり、発熱物質投与等の処置は行わない。体温はラットの直腸内に温度センサーを挿入し、15分間隔で、薬物投与後240分まで測定した。薬物は投与前体温を測定後経口投与した。投与前体温は投与前3回の体温測定値の平均値とした。

対照群に比べて体温低下を示す用量と体温低下が認められる投与後の経過時間を知るために、各個体の投与前体温値を薬物投与後の各測定時点の体温値から差し引いた体温変動値について、測定時点ごとに統計処理を行った。統計処理は4群の等分散性をバートレットで検定し、等分散の場合はダネット、等分散でない場合はスチールの多重比較検定を行った。ただし、投与前の体温値は別に統計処理を行い、等分散であり、分散分析で差が認められないことを確認した。以上のように実施した解析結果を報告したところ、測定時点により統計手法が異なるため体温低下を示す投与後時間を正確に評価できないのではないかとの指摘を受けました。薬物作用の発現時間を評価するにはいかにすればよろしいでしょうか。

《投与前データの統計解析結果》

-----  
Basic statistic  
-----

	対照群	100mg/kg	200mg/kg	400mg/kg
Mean	36.8775	36.89833	36.86917	36.91167
S. D.	0.45838	0.39726	0.4013	0.34491
S. E.	0.13232	0.11468	0.11585	0.09957

-----  
Bartlett test

alpha=0.0500

CHI-SQR cal	0.852419
CHI-SQR( 3, 0.0500)	7.814728
Upper probability	0.836892

-----  
1-way ANOVA with completely randomized design  
-----

Source	Sum of SQR	df	Mean SQR	Fcal	Probability
A	0.013517	3	0.004506	0.027815	0.993652
Error	7.12725	44	0.161983		
Total	7.140767	47			

=====

《投与後データの統計解析結果》

投与後時間	パートレット	対照群	100mg/kg	200mg/kg	400mg/kg
15'	—	0.04±0.06	0.06±0.07	-0.10±0.06	-0.26±0.06**
30'	—	0.06±0.06	-0.03±0.10	-0.19±0.09	-0.67±0.10**
45'	—	0.06±0.06	-0.12±0.08	-0.25±0.10*	-0.82±0.11**
60'	—	0.06±0.07	-0.16±0.09	-0.26±0.08*	-0.91±0.12**
75'	—	0.02±0.06	-0.14±0.09	-0.26±0.09	-1.02±0.13**
90'	p<0.05	0.02±0.07	-0.15±0.10	-0.24±0.07*	-1.16±0.15**
105'	p<0.05	0.03±0.09	-0.21±0.10	-0.28±0.09	-1.29±0.20**
120'	p<0.05	0.00±0.10	-0.26±0.10	-0.34±0.07*	-1.53±0.23**
135'	p<0.05	-0.03±0.10	-0.17±0.10	-0.32±0.11	-1.63±0.28**
150'	p<0.05	-0.03±0.10	-0.17±0.06	-0.39±0.11*	-1.72±0.33**
165'	p<0.05	-0.05±0.11	-0.23±0.07	-0.37±0.09*	-1.90±0.41**
180'	p<0.05	-0.07±0.11	-0.28±0.08	-0.39±0.08*	-2.07±0.46**
195'	p<0.05	-0.08±0.09	-0.25±0.09	-0.38±0.11*	-2.19±0.49**
210'	p<0.05	-0.08±0.09	-0.25±0.09	-0.30±0.12	-2.16±0.50**
225'	p<0.05	-0.01±0.10	-0.28±0.10	-0.28±0.12	-2.17±0.50**
240'	p<0.05	-0.11±0.12	-0.22±0.11	-0.30±0.12	-2.13±0.48**

\* : p<0.05, \*\* : p<0.01(ダネット), # : p<0.05, ## : p<0.01(スチール)

## 1. はじめに

ここに示されたデータの解析には、いくつかの課題が考えられる。

- ①等分散な場合と不等分散な場合とでの、パラメトリックな手法とノンパラメトリックな手法の使い分けの問題。
  - ・時点毎に異なる解析手法を用いてもよいか？
- ②用量反応関係に関する問題。
  - ・単調増加、減少が前提とできるか？
  - ・用量反応関係として捉えるか、対照からの差を考えるか？
- ③経時測定データの解析手法の問題。
  - ・時点毎に解析するか？
  - ・枝分かれ型モデルを当てはめるか？
  - ・成長曲線モデルを当てはめるか？
- ④薬理的、毒性学的観点からの閾値の問題。
  - ・問題となる閾値を越えるか？
- ⑤解析すべきデータの問題。
  - ・実測データを処理するか？
  - ・投与前の値を共変量とするか？
  - ・投与前からの変化量を求め解析データとするか？
- ⑥分散の大きさは作用発現と考えるべきという問題

当初の副題は「枝分かれ型経時測定データの統計解析」とされていたが、ここでは杉山氏の質問に従い、①の課題すなわち「Dunnett検定とSteel検定とが混在してもよいか」という課題に限局して検討する。副題の内容を期待された方にはお詫びいたします。

杉山氏の解析は

- ・まず、各時点毎に4群間の等分散性の検討を行い、
  - ・その後、等分散ならDunnett検定、不等分散ならSteel検定を行った。
- というものである。

このような決定樹的な統計手法の選択手順は、多くみられピンク本にも記載がある。一つのデータ群に対する解析手法を選択する手段としては、決定樹的方法論も実務者にとっては、それなりに価値のあるものである。

しかし、並列するデータ群に対して、同様にして解析手法を選択すると、パラとノンパラの混在といったような問題が発生する。実務者にとっては、かなり頻繁に直面する課題である。

## 2. 検討の目的

2群の計量データの場合の、かつての一般論は

- ・等分散なら t 検定
- ・不等分散なら Welch 検定

というものであった。

しかし現在では、2群の大きさがほぼ同じならば両者にはほとんど違いがないとされている（毒性試験データの解析：p 26、医薬安全性研究会会報34など）。

さらに、

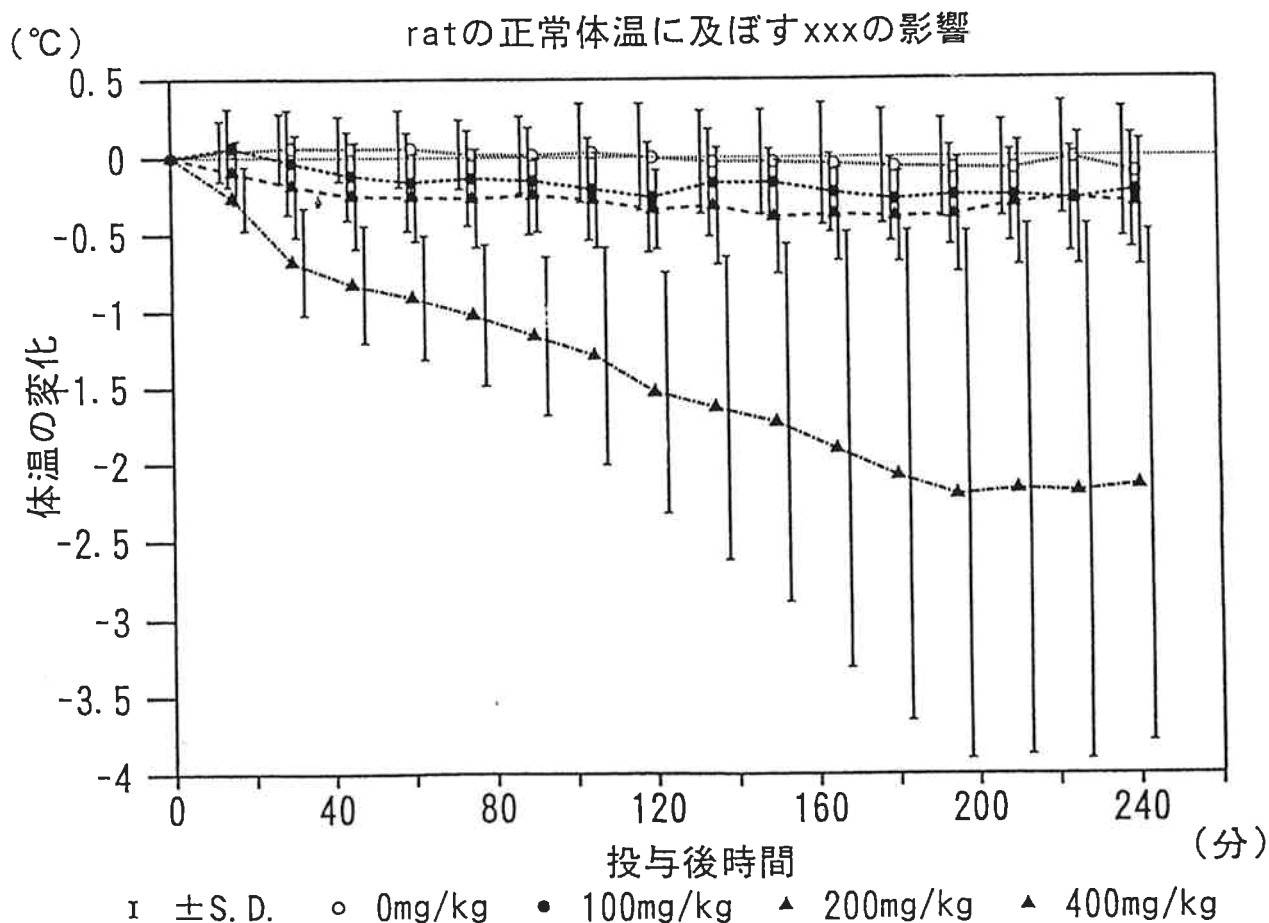
- ・群の数が多く、群の大きさがほぼ等しければ、等分散性を前提にした手法は、等分散性の崩れの影響をあまり受けない（毒性試験データの解析：p 33）。
- ・ノンパラ手法は同じ分布に従うことが前提（同：p 109）。

ともある。

そこで、ここに示された程度に、等分散性がくずれたデータで

- Dunnett 検定と Steel 検定の  $\alpha$  過誤はどの程度か？
- 検出力はどの程度か？
- 検定結果の一致度はどうか？

といったことについて、シミュレーションにより検定手法の性質を捉え、問題点の把握と整理を行った。





### 3. 検討結果

#### 1) 不等分散の場合の $\alpha$ 過誤

各4群の母平均は0とし、以下の母標準偏差をもつ乱数を発生させ、検定を行った。

#### ・Dunnett検定

以下の3種のケースについて、各々10万回の試行を行いDunnett検定の $\alpha$ 過誤を調べた。表中の数値は検定(両側5%)で有意となった回数である(以下同様)

	ケース1-1 $\sigma$ p<0.05	ケース1-2 $\sigma$ p<0.05	ケース1-3 $\sigma$ p<0.05
対照群	0.4	0.4	0.4
第1投与群	0.4 1880	0.4 296	0.4 5
第2投与群	0.4 1892	0.4 314	0.4 2
第3投与群	0.4 1928	0.8 5096	1.6 8829
FWE	4966	5536	8835

結果：ばらつきが大きくなると $\alpha$ 過誤は増大し、ばらつきの大きな群に過誤が集中した。

#### ・Steel検定

同様に以下の3種のケースについて、各々1万回の試行を行なった。

	ケース1-1 $\sigma$ p<0.05	ケース1-2 $\sigma$ p<0.05	ケース1-3 $\sigma$ p<0.05
対照群	0.4	0.4	0.4
第1投与群	0.4 165	0.4 165	0.4 165
第2投与群	0.4 154	0.4 154	0.4 154
第3投与群	0.4 173	0.8 217	1.6 305
FWE	436	494	587

結果：ばらつきが大きくなると $\alpha$ 過誤はやや増大し、ばらつきの大きな群に過誤が集中するが、その傾向はDunnett検定に比べ小さい。不等分散性に対しては頑健といえる？

#### ・両検定結果の一致性

Dunnett検定とSteel検定の結果の組み合わせ(1万回の試行結果)を以下に示す。

Dunnett	Steel	ケース1-1		ケース1-2		ケース1-3	
		p<0.05	ns	p<0.05	ns	p<0.05	ns
第1群	p<0.05	120	72	27	1	0	0
	ns	45	9763	138	9834	165	9835
第2群	p<0.05	102	71	13	5	0	0
	ns	52	9775	141	9841	154	9846
第3群	p<0.05	116	82	200	322	272	651
	ns	57	9745	17	9461	33	9044
FWE	p<0.05	309	180	249	307	299	624
	ns	127	9384	245	9199	288	8789

結果：検定結果の一致度は高くはない。等分散性からのズレが大きくなると一致度はさらに低下する。

2) 母平均の差が小さい場合の検出力

・Dunnett検定 (試行回数各1万回)

	ケース2-1			ケース2-2			ケース2-3		
	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05
対照群	0	0.4		0	0.4		0	0.4	
第1投与群	0.1	0.4	413	0.132	0.4	141	0.218	0.4	8
第2投与群	0.2	0.4	1206	0.265	0.4	687	0.436	0.4	271
第3投与群	0.3	0.4	2893	0.397	0.8	3204	0.654	1.6	3481
FWP	3406			3521			3591		

結果：第3投与群のばらつきが大きくなり等分散性が崩れても、全体としての検出力はほぼ一定である。しかし、母平均の大きな第3投与群の検出力は増加するが、母平均の小さい投与群の検出力は低下する。

・Steel検定

	ケース2-1			ケース2-2			ケース2-3		
	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05
対照群	0	0.4		0	0.4		0	0.4	
第1投与群	0.1	0.4	337	0.132	0.4	492	0.218	0.4	1205
第2投与群	0.2	0.4	1025	0.265	0.4	1855	0.436	0.4	5311
第3投与群	0.3	0.4	2436	0.397	0.8	1773	0.654	1.6	1570
FWP	2999			3241			6065		

結果：第3投与群はばらつきが大きくなると、検出力は低下する。この例では、 $\mu$ の増大よりも $\sigma$ の増大が、順位に与える影響が大きい。逆に、第1第2投与群の検出力は極端に増大した。しかしこれは当然である。というのも、Steel検定はセパレートランキング (対比する2群だけでの順位) をもとに統計量を求めるので、 $\sigma$ が不変のまま $\mu$ が大きくなれば、統計量が大きくなるからである。

・両検定結果の一致性。

Dunnett	Steel	ケース2-1		ケース2-2		ケース2-3	
		p<0.05	ns	p<0.05	ns	p<0.05	ns
第1群	p<0.05	258	155	122	19	8	0
	ns	79	9508	370	9489	1197	8795
第2群	p<0.05	839	367	644	43	267	4
	ns	186	8608	1211	8102	5044	4685
第3群	p<0.05	2202	691	1716	1488	1504	1977
	ns	234	6873	57	6739	66	6453
FWP	p<0.05	2677	729	2419	1102	2810	781
	ns	322	6272	824	5655	3255	3154

結果：不当分散の場合は明らかに両検定の結果には不整合が生じる。

3) 母平均の差が大きい場合の検出力

・Dunnett検定 (試行回数各1万回)

	ケース3-1			ケース3-2			ケース3-3		
	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05
対照群	0	0.4		0	0.4		0	0.4	
第1投与群	0.2	0.4	1199	0.265	0.4	706	0.436	0.4	288
第2投与群	0.4	0.4	5148	0.529	0.4	5158	0.872	0.4	5393
第3投与群	0.6	0.4	8880	0.794	0.8	8475	1.308	1.6	8207
FWP	9070			8887			8798		

結果：第3投与群のばらつきが大きくなり等分散性が崩れると、全体としての検出力はやや低下する。特に、母平均の小さな第1投与群の検出力はかなり低下する。

・Steel検定

	ケース3-1			ケース3-2			ケース3-3		
	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05	$\mu$	$\sigma$	p<0.05
対照群	0	0.4		0	0.4		0	0.4	
第1投与群	0.2	0.4	985	0.265	0.4	1847	0.436	0.4	5217
第2投与群	0.4	0.4	4468	0.529	0.4	7189	0.872	0.4	9937
第3投与群	0.6	0.4	8335	0.794	0.8	6588	1.308	1.6	5399
FWP	8623			8710			9955		

結果：2)と同様に、第3投与群のばらつきが大きくなると、検出力はかなり低下する。逆に、第1第2投与群の検出力は増大した。

・両検定結果の一致性。

Dunnett	Steel	ケース3-1		ケース3-2		ケース3-3	
		p<0.05	ns	p<0.05	ns	p<0.05	ns
第1群	p<0.05	839	360	655	51	288	0
	ns	146	8655	1192	8102	4929	4783
第2群	p<0.05	4194	954	5049	109	5393	0
	ns	274	4578	2140	2702	4544	63
第3群	p<0.05	8231	649	6561	1914	5369	2838
	ns	104	1016	27	1498	30	1763
FWP	p<0.05	8507	563	8263	624	8770	28
	ns	116	814	447	666	1185	17

結果：不当分散の場合は明らかに両検定の結果には不整合が生じる。

#### 4. まとめ

以上をまとめると、

- ・ Dunnett検定は、各群の大きさが等しくても、ある1群のばらつきがかなり大きい場合は（今回の例では4倍） $\alpha$ 過誤が増大する。
- ・ Dunnett検定は、ばらつきの大きな群で有意となりやすく、小さな群では有意となりにくい。
- ・ Steel検定は、セパレートランキングにより統計量を求めるため、ある群の統計量は、他の群のばらつきの大きさに影響されない。
- ・ Steel検定は順位に基づいて統計量を求めるため、ばらつきが大きくなると検出力が低下する。
- ・ 不等分散の場合は、両検定結果の一致度は低い。

よって、

- 不等分散のときはいつでもSteel検定を行うという手順は、適切ではない。
- 各群の例数が等しくても、極端に不等分散な場合は、Dunnett検定はやや危険である。
- 今後は、不等分散の場合に提案されている Tamhane の方法について検討したい。

以上

〔事務局だより〕

前号で予告した『佐久間 昭先生 特集号』を、会報から、モノグラフシリーズに変更し、単行本として刊行したために、会報の刊行スケジュールがずれ込んでしまいました。出版社の本能として、容量オーバーすると分かっているにもかかわらず、新しい企画が発生すると食い付いては、消化不良を起こしてしまいます。これからもその繰返しで進みそうです。

## 医薬安全性研究会 会報NO.42

*Bulletin of Japanese Society for Biopharmaceutical Statistics.*

---

1996年 9月30日 発行

定価1,030円 (本体1,000円)

編集・発行 (株)サイエンティスト社

〒101 東京都千代田区神田駿河台3-2 山崎ビル

TEL. 03 (3253) 8992

FAX. 03 (3255) 6847

振替 東京00180-1-71335

印刷・製本 平河工業社

# 医薬安全性研究会 会報

第42号 1996年9月

## ■ 目 次 ■

### \* 医薬安全性研究会スケジュール

---

- 企業における統計教育の事例報告  
——薬理研究部門における——  
山岡英明…………… 1
  - ピアソン・ $x^2$ 、イエーツの補正、フィッシャーの  
直接確率計算法、どこがちがうのか  
——mid-*P*valueの出番はあるか——  
岩崎学…………… 7
  - ラット体温に及ぼす解熱薬の影響  
——枝分かれ型経時測定データの統計解析(Dunnett検定とSteel検定)——  
杉山公仁、佐々木秀雄……………19
- 
- 事務局だより……………25

編 集

医薬安全性研究会

定価 1,030円