

医薬安全性研究会

会報 No. 28

May. 1989

目次

<前付>

*医薬安全性研究会スケジュール

Parallel-Line and Slope-Ratio Assay 「連載第4回」

Introduktion to Biological Assay /Four-Point Parallel-Line Assay

大阪市立環境科学研究所 野田 勉 訳 1

毒性試験に用いる統計処理法の模索

食品農医薬品安全性評価センター 小林 克己 9

チューキーの方法は群の大きさが違っていても使える

大鵬薬品工業株式会社 佐野 正樹 20

自由度についての補足

名古屋大学 吉村 功 24

第36回定例会出席者名簿 29

事務局だより 30

1989年医薬安全性研究会

これからのスケジュール

☆ 7月15日(土) 第39回定例会(総評会館)

<講演>実験データのグラフ表示 ー様々な事例ー

中里溥志(雪印乳業技術研究所)

☆10月21日(土) 第40回定例会(総評会館)!!!10周年記念大会!!!

☆11月17日(金)~18日(土) データ解析講習会(総評会館)

☆1990年1月20日(土) 第41回定例会(総評会館)

★実験データのグラフ表示 (中里溥志 著)

6月下旬刊 予定

★医学・薬学・生物学のための統計学 (吉村 功 著)

10月刊 予定 詳細は追ってお知らせします。

Practical Statistics for
Experimental Biologists

A. C. Wardlaw

10章 平行線検定と勾配比検定

Parallel-Line and Slope-Ratio Assays

第4回

野田 勉（大阪市環境科学研）訳

10.1 生物検定法について

10.1.1 要素とデザイン

生物検定すなわちバイオアッセイの目的は生体におよぼす効果を測定することにより生理活性物質の効力を検定することである。この種の分析には以下の4つの重要な要素がある。

- a) 未知検体の効力を検定すること。
- b) 効力が既知の同一物質があり、標準または基準として使用できること。
- c) 標準検体および未知検体の「標的」となる植物、動物、微生物、細胞または組織が手にはいること。
- d) 標準検体および未知検体の希釈率や投与量の違いにより生体の「反応」に差がでること。

Table 10.1に例をあげる。ここで「stimulus」（刺激）という用語は投与経路とともに生理活性物質のために用いた。

統計学的観点からは大部分のバイオアッセイは次の3つのいずれかに当てはまる。

- 1) 検量線法：9.4節で取り上げた Biuret法による蛋白定量法と同様の方法であり、ここでは説明を省略する。
- 2) 平行線検定法：標準検体の段階希釈によって用量－反応直線が得られ、未知検体の段階希釈によって標準検体の直線と平行な直線が得られる。このタイプの検定には4点法、6点法あるいは量－反応検定といった様々の形がある。これらについては後述する。
- 3) 勾配比検定法：標準検体と未知検体の用量－反応直線において平行性の代わりに、生理活性物質の濃度0の点を等しくする方法。この例は10.5節で

Table 10.1 A few examples of stimuli, test organisms and responses in biological assays

Substance	Stimulus		Response
	Mode of administration	Test organism or tissue	
Penicillin	Wells cut in an inoculated agar plate	<i>Bacillus subtilis</i>	Zone of inhibition of growth
Indolylacetic acid	Agar blocks	The growing tips of oat seedlings	Bending
Tetanus toxin	Intramuscular injection	Mouse	Death or paralysis
Histamine	Bathing the organ in a controlled-temperature-bath	Strips of guinea pig ileum	Muscle contraction recorded on a kymograph
DDT	Aerosol	House flies	Death
Serum complement	Mixing with target cells in a test tube	Antibody-coated erythrocytes	Haemolysis
Lipopolysaccharide (endotoxin)	Intravenous injection	Rabbits	Elevation in body temperature

述べる。

バイオアッセイには測定したい物質の生理活性を直接測定できるという利点がある。しかしながら、時にはより安価で、より迅速に測定できる化学的あるいは免疫学的方法で代用されることがある。例えば、インシュリンはラジオイムノアッセイ法（RIA）で簡単にしかも早く測定できる。ところが、RIA法では、活性のあるインシュリンと活性のないプロインシュリンを区別して定量できない。だから、インシュリンを研究している研究室ではRIA法と同時にバイオアッセイ法もテクニックとして持っていたいだろう。同様の例として、破傷風毒素と類毒素（toxoid：ホルマリン処理により毒性を除去）をあげることができる。両者とも抗体と反応はするが、筋肉麻痺をおこすのは毒素だけである。

10.1.2 標準物質

突然変異性も含めて生体が本来持っている可変性と、一定で再現性のある結果を得ることの難しさがバイオアッセイの最大の難点である。微生物を標的とするバイオアッセイでは、ある特定の系統の菌を用いることにより測定方法の標準化を行うことができる。例えば、ビタミンB12の測定には大腸菌NCTC8878を使用する（Hewitt, 1977）。しかしながら、こういった標準化法でもまだ十分ではない。

というのは、標準系の細菌を用いても加えたビタミンB12だけによる増殖量を測定できるわけではなく、培養液中のB12以外の様々な因子全部に影響されてくるわけである。そこから、バイオアッセイの哲学 (Philosophy of biological assay) が生じてくる。すなわち、1回に1群の (あるいは1個の) 標的生物を用いて、未知検体に対する反応と、それと平行して行った標準検体あるいは基準検体に対する反応とを比較することによって標準化は最高にその力を発揮する、ということである。標準検体と未知検体を無作為に分けることによって生物学的なバラツキを両者に等分でき、未知検体の活性を標準検体の活性の倍数あるいは分数として表すことができる。それ故、バイオアッセイにおける標準 (biological standard) とは試験系やその反応ではなくて、その活性を損なわずに保存されている生理活性物質を指している。例えば、ストレプトマイシンの国際標準品はアンプルになっており、その中に硫酸ストレプトマイシンが 175mg入っている。そして、その0.001282mgを1国際単位と決めている。メートルとかキログラムといった物理的標準と違って、今まで述べてきた生物学的標準は測定するときには必ず標準のうちのあるものを使用しなければならないということである。ゆえに、国際的な生物学的標準品は毎日の作業に必ず使われるわけではなく、実験者はルーチン用に個人的な標準品を用意しなければならない。このような個人的標準品は使用前にまず国際標準品で較正してその効力を国際単位として測定しておく必要がある。

10.2 4点平行線検定法

10.2.1 はじめに

平行線検定法の最も単純な形が4点法である。すなわち、標準検体の高用量および低用量を投与したときの反応と未知検体の高用量および低用量での反応とを比較することである。2本の平行な用量-反応直線を引き、それらの直線間の水平距離から未知検体の効力を測定する。

Bacillus subtilis NCTC 3610 の増殖阻止効果によるペニシリンの効力検定の例を示す。この菌を均一に分散させた寒天培地をペトリ皿中に入れて固め、その寒天培地にあけたウェルに段階希釈した一定量のペニシリンを入れる。培養中に抗生物質がそのウェルから寒天中へ拡散し阻止円を形成する。翌日、その阻止円の直径を測定する。方法に関する詳細はWardlaw (1982) を参照のこと。

実際に4点法を行う際には、まず、用量と反応の間に直線性を有する範囲を決めるために、広い範囲についてその系の用量-反応特性を調査しなければならない。Fig. 10.1aではペニシリン濃度に対する阻止円の直径が双曲線となることを

示した。ところが、Fig. 10.1bでは、ペニシリン濃度の常用対数を取ってプロットすると、この曲線が0.1から10units/mlの濃度範囲で直線となることを示している。この直線関係を用いると、9.4節で示した標準曲線による逆推定でペニシリンの未知検体の効力検定ができる。(b)図が(a)図と異なっている点はX軸に対数尺度を使ったことであり、そのために濃度の対数値を用いて計算をしている点だけである。

さて、ここにあるペニシリン溶液があり、その効力は検量線法でおおよそのところ決められているが、さらに正確な値を測定したいという場合には4点平行線検定法が有用である。

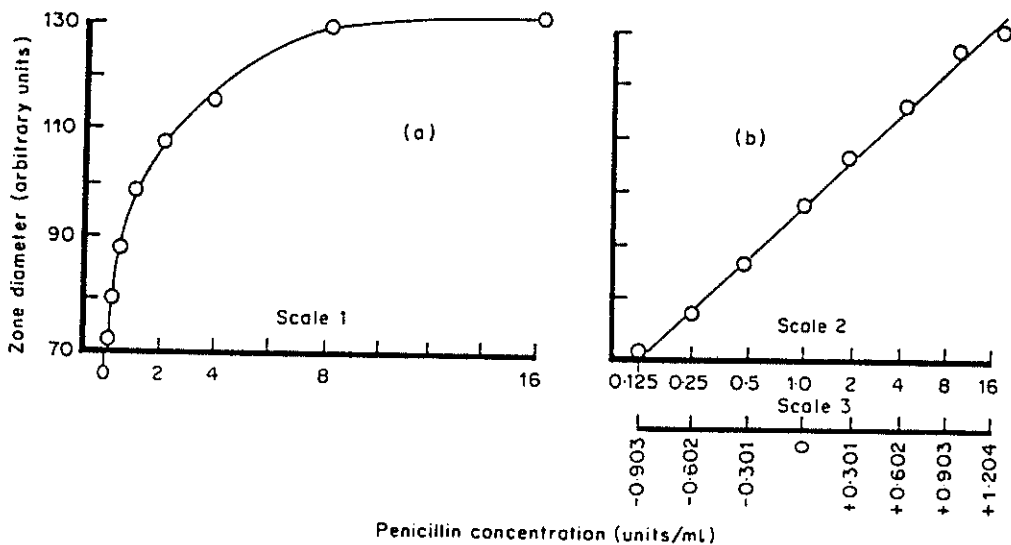


Fig. 10.1 Penicillin dose/response curves. In (a) the plot of antibiotic concentration on an arithmetic scale against the diameter of the zone of growth-inhibition gives an obvious curve; in (b) linearization is achieved by expressing the concentration on a logarithmic scale. This can be done in the two alternative ways illustrated by scales 2 and 3.

4点法の基本的なデザインは、直線を引くために適当な間隔をおいた2点をグラフ上に求めること、ただそれだけである。だから、あるペニシリン溶液(未知検体)の効力検定を行うためには、その溶液とペニシリン標準液(標準検体)についてそれぞれ2点ずつをグラフ上に求めればよい。両検体について高用量として原液、低用量として25%希釈液をそれぞれ試験すると、標準検体の用量-反応直線と未知検体の用量-反応直線とをつくるための4点が得られる。この2本の直線は平行であるはずである。しかしながら、もし万一未知検体に標準検体と全く同量のペニシリンが含まれていれば、この2本の直線は一致して1本になってしまう。もしそうでなければ、2本の直線はお互いに水平移動の関係にある。この2本の直線間の水平距離の長さや方向によって標準検体から未知検体の相対的

な効力を求めることができる。

さて、標準検体の高用量 (5.0 units/ml) および低用量 (1.25units/ml) と未知検体の高用量および低用量とをそれぞれ4回テストするという例で考えてみよう。ここで、未知検体の低用量は高用量の1/4の濃度とし、対称的な実験デザインとした。すなわち、4枚のペトリ皿を用意し、1枚のペトリ皿に4個のウェルをあけ、それらに標準検体の高用量、低用量、未知検体の高用量、低用量を入れた。統計学的観点からそれぞれのペトリ皿を無作為なブロック (randomized block) とするために、4種の溶液を前もって決めたランダムな順番でそれぞれのペトリ皿に入れた (Section 8.3)。全体の検定はこのような4個のブロックからなる。標準検体および未知検体それぞれの低用量と高用量をつくり、それぞれ S_L 、 S_H 、 U_L 、 U_H とし、その阻止円の直径を Table 10.2 に示した。この表には分散分析に必要なある用量における阻止円の直径の合計と、あるペトリ皿における阻止円の直径の合計はもちろんのこと、さらに、ある用量における阻止円の直径の平均値も示している。

Table 10.2 Sample results of a 4-point assay of penicillin

Plate	Zone diameter, arbitrary units (y) ^a				Plate total (Q)
	S_L	S_H	U_L	U_H	
A	83	101	79	95	358
B	85	103	81	96	365
C	82	98	80	94	354
D	87	101	78	98	364
Dose total (T)	337	403	318	383	1441
Dose mean (\bar{y})	84.25	100.75	79.5	95.75	

^aZone diameters were measured by placing each culture plate on an overhead projector and throwing an image, enlarged about 20 times, on to the laboratory wall. An ordinary ruler was then used to read the diameter of each zone image in millimetres (see Wardlaw, 1982).

10.2.2 グラフによる評価

グラフによる評価のためには Fig. 10.2 のように片対数紙を使用するのがもっともよい。すなわち、阻止円の直径として測定される「反応」には普通目盛を、抗生物質の用量には対数目盛を用いる。注意しなければならないのは、我々にはサンプル中の抗生物質の濃度がわからないのだから、抗生物質の用量を units/ml で表すわけにはいかないということである。そのかわりに、相対的な用量として表現する。例えば、標準検体あるいは未知検体の原液のことを相対用量 1.0 と呼

び、それらの4倍希釈液を相対用量0.25と表す。

2本の用量-反応直線を引くとき、ただ単にそれぞれの低用量および高用量の平均値を結ぶ線を引くだけではいけない。そのかわりに、4点全部を等しく評価するように最も良くフィットする2本の平行な直線を引かなければならない。というのは、真の用量-反応直線は平行であり、平行性からのズレは単に誤差のためであるとの仮説に基づいているからである。

この様にして2本の平行な最良近似直線を引いたのち、次のようにして未知検体の効力を計算する:横軸の1.0の点から上に垂線を引き未知検体の用量-反応直線との交点を求め、その点から水平に線を引き、標準検体の用量-反応直線との交点を求め、さらにそこからX軸に垂線をおろす、Fig.10.2の矢印の通りである。矢印がX軸に戻ってきたその点(本例ではおおよそ0.67である)が未知検体の相対的効力(R)である。相対的効力は次式で表される。

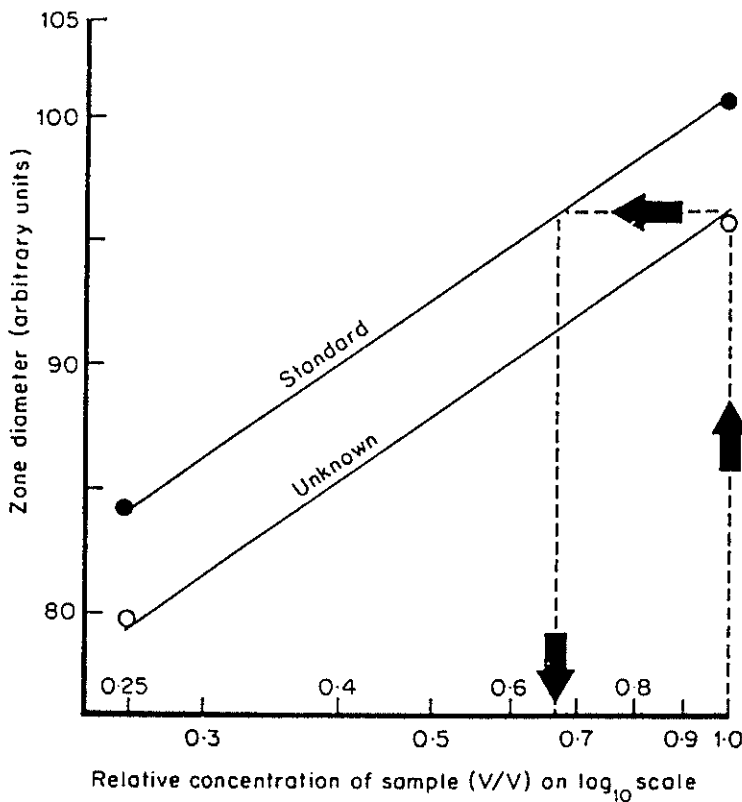


Fig. 10.2 Graphical plot of 4-point assay results from Table 10.2. The horizontal displacement of the two lines is $\log_{10}R$, the logarithm of the relative potency.

$$R = \frac{\text{標準検体の用量}}{\text{未知検体の用量}}$$

ここで、分子および分母の用量は生物学的に同じ強さの効果を示す用量のことである。この例のように1.0から矢印線で内挿するのはR比の分母を1.0として計算を容易にするためである。実際には、2本の線は平行だからその間に任意の水平線を引き、それらの交点からX軸へ垂線をおろし相対的な用量の比を計算することもできる。

このように、もし、未知検体の標準検体に対する相対的効力をグラフから0.67と読み取れば、そのペニシリンの効力は $5.0 \times 0.67 = 3.35 \text{ units/ml}$ と計算される。

データをグラフで分析することは、代数計算でのエラーのチェックに有効である。グラフによる計算でも、ある種の目的のためには十分に正確な値となる。しかし、グラフではブロック間の変動の大きさや平行性からのズレという観点からみた客観的な評価ができないという短所があるし、さらに効力の95%信頼限界も計算できない。これらの点を考察するために、次節以下ではまず分散分析を行い、検定の有効性を確認し、次に相対的効力と信頼限界を公式から計算してみよう。

[Q & A]

Q：一般的に2つの用量－反応直線は切片も勾配も異なるはずだ。生物検定法ではなぜ3つの型に分け得るのか。

A：異なる2つの物質の用量－反応直線の場合は確かに切片も勾配も異なるはずである。しかし、生物検定法で問題とするのは、測定したい物質があらかじめ決まっていて、その標準物質の効力と未知検体の効力とを比較することによって未知検体中の当該物質の量（あるいは効力）を測定するものであるから、全くの一般論とは別である。ただし、異なる2つの物質の場合でも、平行線検定法あるいは勾配比検定法の条件に合致すれば相対的効力（効力比）の測定は出来る。

なぜ生物検定法がこの3つの型に分け得るのかは不明。

意見：一般的には切片も勾配も異なるので、これらを合わせて評価する方法があれば一つにまとめられるはずだ。

Q：平行線生物検定法では低用量と高用量の2点だけで検定に十分なのか。

A：これは平行線生物検定法の一番簡単な例である。通常、直線性の検定には1本の用量-反応直線に対して3水準以上の用量が必要であるが、あらかじめ直線性や平行性が保証されているときには標準および未知検体の各2用量でもよいと言われている。

Q：「サンプル中の抗生物質の濃度は未知」とは？

A：サンプルが個体であればmgあるいはgとして、液体であればml等として測定できるのは当然である。しかし、これらの未知検体は一般的には粗抽出物の場合が多く、重量あるいは液量では測定出来ない。すなわち、ある一定重量（あるいは液量）の未知検体中に活性な抗生物質がどれだけ含まれているかは不明ということである。

Q：検量線法と平行線生物検定法とをどう使い分けるのか？

A：検量線法では単に作用の強さだけによって検定している。平行線生物検定法では直線の平行性を検定するというステップが入ることによって、未知検体に不純物による反応阻害や作用機序の異なる活性不純物の存在を平行性や直線性の検定から推定できる可能性がある。検量線法よりも不確定要素が多い場合に平行線生物検定法を用いるようである。

Q：「この様にして・・・」とは？

A：4点法においてグラフ法で直線を当てはめるときには標準検体、未知検体それぞれの2点を単に結んで2本の直線を引くだけでは実験誤差のために2本の直線は平行にならない。

4点法で検定するという事は、あらかじめ2本の用量-反応直線は平行であるとの保証がある場合であるので、平行な2本の直線が4点にできるだけ近似するように傾き、切片を調整する（ただし、視覚的に）必要がある。もし、明らかに平行でなければ実験に不備があるか、4点法で検定することに無理がある。

毒性試験に用いる統計 処理法の模索

財) 食品農医薬品安全性評価センター
小林 克己

科学の進歩に伴い数多くの化学物質が合成され世の中に送り出されている。それらは、医薬、食品添加物、化粧品等直接ヒトに用いられるもの、農薬や動物用薬品のように動植物を介してヒトに摂取される可能性のあるもの等様々な場面でヒトと接触を持つ。

これら化学物質は有効面(薬理薬効、生物活性)の評価と毒性面の評価がなされた後商品となるが、この段階での評価判定には生物統計手法が用いられる。

前者の薬理薬効試験に用いられる統計手法では多くの薬剤濃度が設定できました、繰り返しも多く、再試験が可能で試験期間も比較的短期である。さらに医薬、農薬では薬効の有無は対照群と比較しなくても一目瞭然である場合が多い。しかし動物用薬剤の中の成長促進剤等では対照群と比較して通常でも10%以内¹⁾の増加にすぎなく飼料効率改善効果もほぼ同程度である。このような評価試験では各群の分散もほぼ同程度を示すため、直ちに分散分析法(analysis of variance, ANOVA)やt検定等、いずれの統計手法を採用してもさほど問題はないと思われる^{1), 2), 3)}。しかし後者の毒性試験になると種々の問題が生じる。

一般的に毒性試験では対照群を含め4群を設定し試験を実施する。そして確実中毒量および無作用量を検索するのが常である。

今回受託試験施設がルーチン的に実施でき、毒性試験に適した統計処理法を模索し、いくつかの知見を得たので数種の定量データ(parametric data)を示し報告する。

一般的に統計処理の基礎には Student の t-検定^{4), 5), 6), 7), 8)} の様に対照群と被験物質を投与した1群間との比較検定、いわゆる2群間の検定(2つの平均値の差の検定)である。各群の動物数の同一化、等分散を持ったものが好ましいが、これらが異っていても検定可能な Aspin-Welch

検定⁹⁾ および Cochran の t-検定法^{10), 11)} 等の2群間の比較検定が最も多く現在使用されている。この場合、両群の分散の相違によって対照群との差が約2.8%減少していても5%水準で有意差を示し、5.6%減少していても同様の有意差を示す場合が時々見られる(Table 1)。我々は(有意水準値5%、 $p=0.05$)と対照群との差(指数)の

Table 1. Significant Differences by Student's Test and in Percents of Control Group

Sex	Dose level (ppm)	Erythrocyte counts, ($\times 10^6/\text{mm}^3$)
Male	0	8.23 \pm 0.17 (100.0)
	1,000	8.00 \pm 0.28* (97.2)
	10,000	7.91 \pm 0.21** (96.1)
	100,000	7.77 \pm 0.50* (94.4)

*: $p < 0.05$.

** : $p < 0.01$.

(): ln % of control group.

Ten rats in each group at 13 weeks after dosing.

どちらかを重要視すべきかである。これも薬量相関性の有無を含め今後重要な判断材料の規準となろう^{12), 13)}。ともあれ各ガイドライン^{14), 15)}は統計処理法を必要としているが手法については触れていない。そこで試験成績を正確に表現し検出力の高い方法で実施することが必要となるが、この検出力の強さは手法によって異なる為、解釈が大変むずかしい、その理由に少い群数の設定、検出力が高すぎて有意差が出すぎはしないか?、これら検出力の強弱をカバーするため生物学的有意差¹⁶⁾いわゆる背景値(background data)の集積が大切で、これらデータと比較することも大きな手助けとなる。

受託施設では試験依頼者の要望または受託者の推奨する統計手法によって試験結果の検定が実施される。しかし長期試験では、投薬後52~104週に進むにつれて各群間の分散値も極めて異なり動物数も同様に変化してくる。これらの事を考慮し分散値の大きな群が1ないし2群あっても一応に統計処理ができるように Chayne Gad^{9), 10), 11)}らによる決定樹法(decision tree, Fig. 1)やこれらを若干モディファイした方法が使用さ

れつつある^{17), 18), 19)}。

この決定樹法以前から分散分析法で有意差を示した場合の群間が有意差を持つかを検索するため、Duncan^{10), 20), 21), 22)}およびTurky^{9), 22), 23)}の多重範囲検定法 (Duncan's multiple range test, Turky's multiple range test) や Dunnett の多重比較検定法^{9), 24), 25), 26)} (Dunnett's multiple comparison test) 等で有意差を吟味してきた。これら分散分析法は3群以上の場合に用いられる検定法でANOVA実施前にBartlettの等分散検定^{9), 10)} (Bartlett's test) の必要が重要視されているようである。これはパラメトリックデータとして扱うか否かの第1関門ともなっている。これら多重範囲 (比較) 検定の利点は対照薬剤が入り多群となった場合、有意判定が各群平均値の肩口につけるアルファベットによって敏速に判断ができる特徴を持っている (Table 2)。特に動物用薬剤の薬効および獣医畜産関係の論文に多く用いられている^{27), 28), 29)}。しかし毒性試験では対照群との比較が主で Table 2 に示した有意差マークを用いることがほとんどない。

さてここで最近使用されてきつつある前述の決定樹法を若干モディファイ (Fig. 3) し片側検定によって種々検討した例を示す。

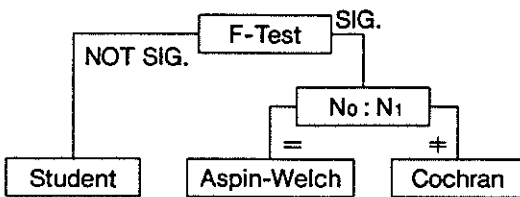


Fig. 2. Statistical Analysis of Among the Two Groups on the An-Pyo Center.

これらはいずれも二年間の長期毒性で、慢性毒性と癌原性試験を組合せた併合試験をラットで、癌原性試験をマウスを用いて実施した。両試験共対照群を含めて4群を設定し、パラメトリックデータの種類と検定項目はTable 3に示した。

調査項目は体重、飼料摂取量、飲水量、血液学検査、生化学検査、尿検査および臓器重量とその体重比等でラットの併合試験では雄雌合計680回、マウスの癌原性試験では雄雌合計434回の検定回数が必要となる (マウスの癌原性試験では上記の調査項目中実施しないものもある)。

この二試験の各測定項目を Fig. 3 の決定樹法を用いて実施した。まずBartlettの等分散検定で吟味 (P=0.05) した結果、ラットの併合試験は Table 4 にマウスの癌原性試験は Table 5 に各々示した。ラットの試験では680項目中186項目の27.3%が有意差を示し等分散でないことを示した。又体重の項目は最も等分散化していた。一方マウスの試験では434項目中114項目の26.2%が有意差を示した。中でも雌体重に高い率を示した。

これらの事からラット・マウス各二年間の試験に於るBartlettの等分散検定で有意差を示す割合は26~27%前後であった。従ってこの26~27%の統計量がノンパラメトリックデータとしてクラスカル・ワリスの順位検定 (Kruskal-Wallis Nonparametric Analysis of Variance)^{9), 10), 30)} を通らなくてはならない。この検定は周知の如く数の大きい値から並べて行うもので差の大小にかかわらず順位づけを行う検定でパラメトリック手法と比較すると説得性に欠ける様に思える。Bartlettの等分散検定で有意差を認める経時的変化はTable 6に示した。体重では

Table 2. Expression of the Significant Differences on Various Multiple Comparison or Range Test

Sex	Dose level (ppm)	Erythrocyte counts, ($\times 10^6/\text{mm}^3$)		
		Dunnett's	Duncan's	Turkey's
Male	0	8.23 ^a	8.23 ^a	8.23 ^a
	1,000	8.00 ^{ab}	8.00 ^{ab}	8.00 ^{ab}
	10,000	7.91 ^{ab}	7.91 ^b	7.91 ^{ab}
	100,000	7.77 ^b	7.77 ^b	7.77 ^b

Means with different superscript letters differ significantly ($p < 0.05$).

ラットで79~104週に渡り若干増加したがマウスで0~26週に最も多く、その後減少した。飼料摂取量ではラットおよびマウスで週が進むにつれて増加が認められた。血液学検査はラットおよびマウス共大きな変化が認められなかった。生化学検査はラットのみのデータから週が進むに従い増加の傾向を示した。臓器重量ではラットのみのデータではあるが試験期間を通して大きな変化を示さなかった。しかし臓器重量体重比は52週以降若干増加を示した。

従って52週以降の78および104週時になれば必ずいくつかの群で分散が極めて不均等となり Bartlett の等分散検定で有意差を示してしまう。これらの原因は高用量群の分散が大きい場合、また小さい場合、反面对照群の分散が極めて小さい場合等さまざまであった。さて前述のように26~27%のデータをノンパラメトリック扱いとして進めても良いのだろうか、Bartlett の等分散検定の重要性は毒性試験のように対照群の平均値に対して始めから数10%も減少する事が分かっておりまた分散値も投薬群で著しく変化することが当然である場合、この検定法が有用なのか大変疑問を持たざるを得ない。

ここでこの二つの試験の有意差検定の解析を Fig. 2 に示した t-検定(片側検定)、Bartlett の等

分散検定で有意差を示さなかった項目について Dunnett の多重比較検定、Bartlett の等分散検定をすることなく Dunnett の多重比較検定、最後に Fig. 3 に示した決定樹法の4手法によって試みた。有意差はいずれも5%水準値で吟味し、その結果をラットについて Table 7、マウスについて Table 8 に各々示した。

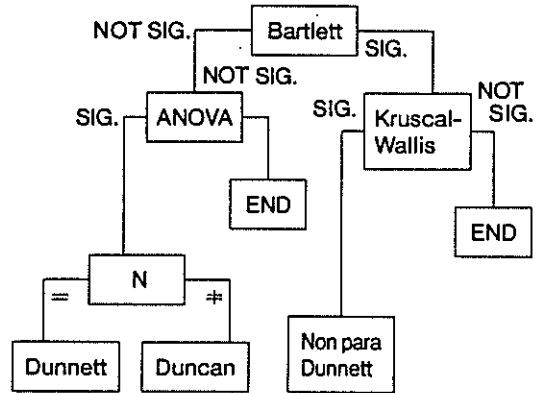


Fig. 3. Decision Tree of the An-pyo Center for Toxicological Study.

ラットの慢性および癌原性併合試験では、最高に有意差を示す可能性 2,040 個中、t-検定が 513 個 (25%)、Bartlett の等分散検定を通過後 Dunnett の多重比較検定法が 268 個 (13%)、Dunnnett の多重比較検定法のみが 320 個 (16%)、決定樹法が 308 個 (15%) であった。

Table 3. Number of Items on Statistical Analysis for 2 Years Chronic Feeding and Oncogenicity Studies in Rats and Carcinogenicity Studies in Mice

Items	80 rats/group × 4 groups			70 mice/group × 4 groups		
	Male	Female	Total	Male	Female	Total
Body weight	58	58	116	58	58	116
Food consumption	120	120	240	120	120	240
Water consumption	6	6	12	Not analysis		
Hematology	11 parameters × 4 times 44		88	10 parameters × 3 times 30		60
Blood chemistry	17 parameters × 4 times 68		136	Not analysis		
Urinalysis	2 parameters × 4 times 8		16	Not analysis		
Absolute organ wt.	20 (5×4times)	16 (4×4times)	36	5 (5×1time)	4 (4×1time)	9
Relative organ wt.	20 (5×4times)	16 (4×4times)	36	5 (5×1time)	4 (4×1time)	9
Total	344	336	680	218	216	434

Body weight was measured every 4 weeks from 26 weeks after dosing.

一方マウスの癌原性試験では最高に有意差を示す可能性1,302個中、t-検定が412個(32%)、Bartlettの等分散を通過後Dunnettの多重比較検定法が201個(15%)、Dunnettの多重比較検定法のみが282個(22%)そして決定樹法が213個(16%)であった。いずれの試験もt-検定が最も有意差を多く検出し、次いでDunnettの多重比較検定法、決定樹法そしてBartlettの等分散検定後Dunnettの多重比較検定法の順であった。しかしt-検定を除く3検定法間では大きな差が見られず、これら3検定法に比べてt-検定法は約1.5倍の有意差を示したことになり今回検討した4法中t-検定が最も高い検出力を示した。

毒性試験では二年間の内、後半以降にいずれかの群で分散値が著しくなる。動物数が各群いずれも10匹以上の場合、パラメトリックデータはパラメトリックデータとして扱いたいのが人情である。またBartlettの等分散検定で有意差が出る場合の多くは、ある1群に原因のある場合が多い。従ってBartlettの等分散検定を除外し分散分析法を実施しても、分散の大きい群は全平方和から全体の誤差項に吸収されてしまう。時にはこの方法でも著しく大きい分散を持つ群が混入することにより、低薬量群においても有意差水準がt-検定に比べて甘くなる。

分散分析法では全数値を用いて全平方和を作成し、各自由度で割った分散の誤差項を分母にし、各調査項目間の分散を分子にして算出し、その値(分散比)をF表を用いて判定する。この方法では数10年前のように分析機器の未発達時代の用手法時代や、測定に長時間を用いた時代等に於ける人為的測定誤差が誤差項に十分考慮されている。最近では血液分析法も多項目を短時間で処理し、体重、飼料等は640頭を各々1時間以内で測定可能で、もはや人為的疲労による誤差は極めて少い。また分散分析法は二元配置および多元配置法によって多項目を一度に吟味できるのも特徴であるが、毒性試験では多元配置法はなく、二元配置法が時おり見られる³¹⁾。

従って対照群、低薬量群、中薬量群および高薬量群は各々独立して存在していると考えられる。これらの事からt-検定系が被験物質の影響

を適確に検出する能力を持っていると推測できる。もし1群のみ分散値が極めて大きい、または小さい場合、また $n_1 \neq n_2$ の場合でも他への群への影響は全く無いのが特徴である。

前述の慢性毒性試験では定量データの内の約30%弱がBartlettの等分散検定で有意差を示し、ノンパラメトリック検定に進むが同データをFig. 2に示したF-検定で有意差($P < 0.05$)を示し、Aspin-WelchおよびCochranのt-検定を実施した率をTable 9に示した。Bartlettの等分散検定に比べて、体重はラットで7%、マウスで21%、飼料摂取量はラットで13%、マウスで17%と、Bartlettの等分散検定に比較して極めて少い頻度で有意差を示した。他の血液学検査および臓器重量などについても、Bartlettの等分散検定に比べ同等またはそれ以下で有意差を示すことから、t-検定系での検索の方が妥当のように思われる。

以上の事から毒性試験に用いられる公平且つ検出力の高い定量データの統計処理法はFig. 2に示したt-検定系と考える。この検定法で有意差の認められてしまう軽微な変化は背景データや対照群間との差で判断することが最適である。

本報告は安評センター所報、第2巻、1988に掲載したものをそのまま引用した。

財団法人

食品農医薬品安全性評価センター

(略称 安評センター)

〒437-12 静岡県磐田郡福田町塩新田字荒浜 582-2

TEL (0538)58-1266

FAX (0538)58-1393

Original Figure by Shayne Gad and Carrol S. Weil.

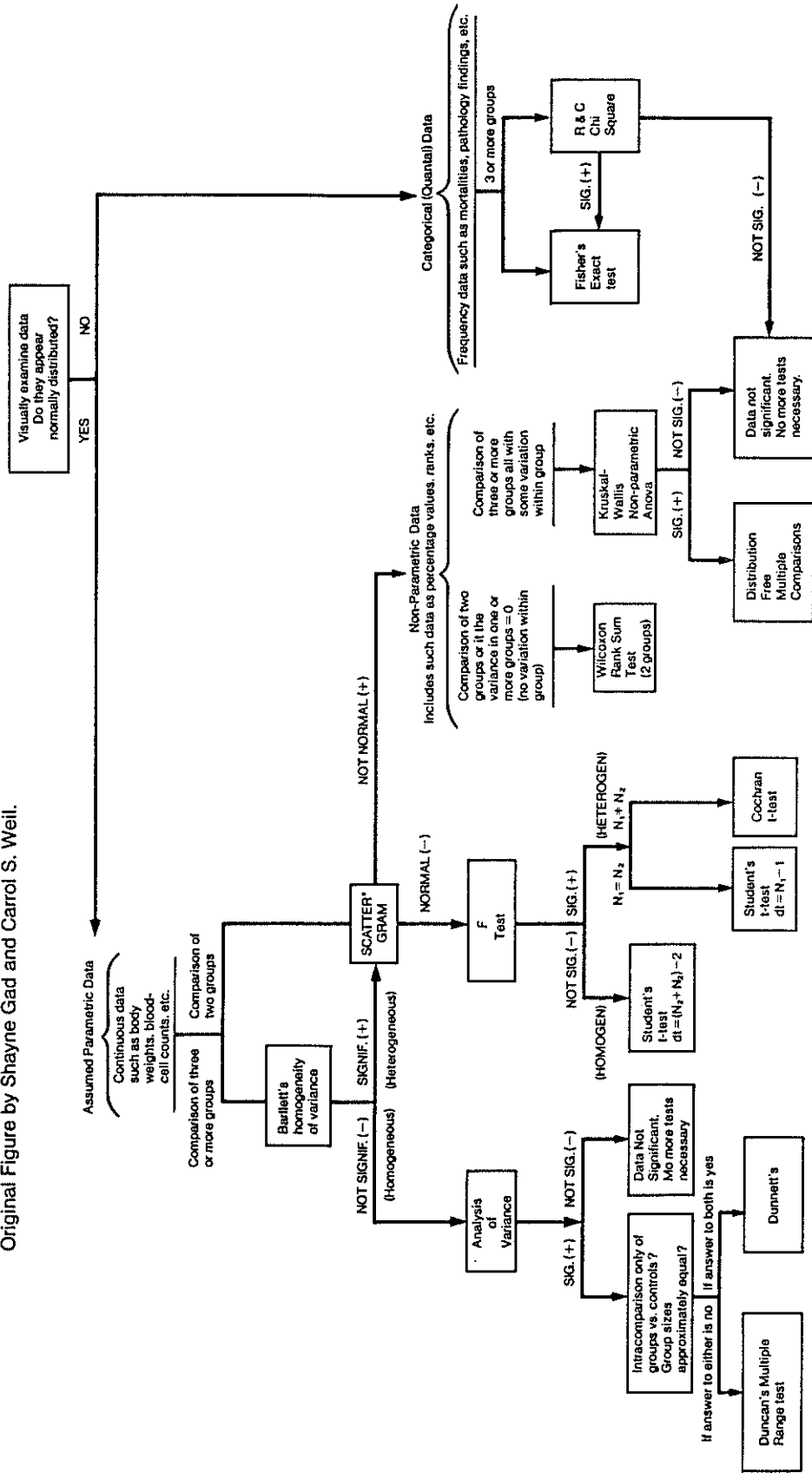


Fig. 1. Decision Tree for Selecting Hypothesis-testing Procedures.

* If plot does not clearly demonstrate lack of normality exact tests may be employed.
 - if continuous data, Kolmogorov Smirnov test.
 - if discontinuous data, Chi-Square Goodness-of-Fit test may be used.

Table 4. The Rates of Statistical Significance on Various Parameters for 2 Years Chronic Feeding and Oncogenicity Studies in Rats by Bartlett's Test

Test period :104 weeks

Animals :F344

No. of group :Four groups/sex

No. of animals :80/group/sex, a total 640 animals

Sacrificed and analysis : At 26, 52, 78 and 104 weeks after dosing

Items	Sex	Times of measurement	Significant at 5% level	
			Times	Rate (%)
Body weight *	Male	58	5	8.6
	Female		0	0.0
Food consumption**	Male	120	28	23.3
	Female		39	32.5
Water consumption	Male	6	1	16.6
	Female		1	16.6
Hematology	Male	11 parameters × 4, a total 44	14	31.8
	Female		16	36.3
Blood chemistry	Male	17 parameters × 4, a total 68	24	35.2
	Female		21	30.8
Urinalysis	Male	2 parameters × 4, a total 8	4	50.0
	Female		2	25.0
Absolute organ wt.	Male	5 organs × 4, a total 20	9	45.0
	Female	4 organs × 4, a total 16	3	18.7
Relative organ wt.	Male	5 organs × 4, a total 20	13	65.0
	Female	4 organs × 4, a total 16	6	37.5
	Male	344	98	28.4
	Female	336	88	26.1
Total		680	186	27.3

* Including of body weight gain.

**Including of any total consumption.

Table 5. The Rates of Statistical Significance on Various Parameters for 2 Years Carcinogenicity Studies in Mice by Bartlett's Test

Test period : 104 weeks
 Animals : B6C3F₁
 No. of group : Four groups/sex
 No. of animals : 70/group/sex, a total 560 animals
 Sacrificed and analysis : At 52, 78 and 104 weeks after dosing

Items	Sex	Times of measurement	Significant at 5% level	
			Times	Rate (%)
Body weight *	Male	58	4	6.8
	Female		34	58.6
Food consumption**	Male	120	21	17.5
	Female		21	17.5
Hematology	Male	10 parameters × 3, a total 30	15	50.0
	Female		11	36.6
Absolute organ wt.	Male	5 organs × 1, a total 5	2	40.0
	Female		4 organs × 1, a total 4	2
Relative organ wt.	Male	5 organs × 1, a total 5	2	40.0
	Female		4 organs × 1, a total 4	2
	Male	218	44	20.1
	Female	216	70	32.4
Total		434	114	26.2

* Including of body weight gain. **Including of any total consumption.

Table 6. Changes in Number of Significant Differences by Bartlett's Test During the Test Period

Items	Weeks into dosing							
	Rats				Mice			
	0-25 (26)	27-52 (52)	53-78 (78)	79-104 (104)	0-25 (26)	27-52 (52)	53-78 (78)	79-104 (104)
Body weight	0/60	0/16	0/16	5/16	21/60	8/16	7/16	2/16
Food consumption *	3/56	7/56	16/56	41/56	8/56	5/56	10/56	19/56
Hematology	3/20	8/20	9/20	9/20	—	8/20	9/20	9/20
Blood chemistry	7/34	7/34	16/34	15/34	—	—	—	—
Absolute organ wt.	1/11	4/11	4/11	3/11	—	—	—	4/9
Relative organ wt.	1/11	4/11	7/11	7/11	—	—	—	4/9

* Excluding of 0 week. — : Not tested. () : Sacrificed weeks after dosing.

Table 7. Comparison of Student't Test and Dannett's Multiple Comparison Test with Bartlett's Test, Dannett's Multiple Comparison Test without Bartlett's Test and Decision Tree for Number of Statistically Significance from Control Group on Various Parameters for 2 Years Chronic Feeding and Oncogenicity Studies in Rats

Items	Sex	Number of analysis	No. of significant differences ($p < 5\%$), (%)			
			Student	Dunnett #	Dunnett ##	Decision Tree
Body weight*	Male	$58 \times 3 = 174$	90 (52)	72 (41)	77 (44)	72 (41)
	Female		60 (34)	55 (32)	55 (32)	55 (32)
Food consumption**	Male	$120 \times 3 = 360$	72 (20)	29 (8)	32 (9)	29 (8)
	Female		78 (22)	28 (8)	45 (13)	28 (8)
Water consumption	Male	$6 \times 3 = 18$	4 (22)	1 (6)	2 (11)	2 (11)
	Female		2 (11)	2 (11)	2 (11)	2 (11)
Hematology	Male	$44 \times 3 = 132$	18 (14)	6 (5)	7 (5)	10 (8)
	Female		19 (14)	7 (5)	8 (6)	9 (7)
Blood chemistry	Male	$68 \times 3 = 204$	60 (29)	25 (12)	33 (16)	33 (16)
	Female		63 (31)	25 (12)	40 (20)	36 (18)
Urinalysis	Male	$8 \times 3 = 24$	3 (13)	0 (0)	3 (13)	1 (4)
	Female		1 (4)	0 (0)	0 (0)	0 (0)
Absolute organ wt.	Male	$20 \times 3 = 60$	8 (13)	2 (3)	2 (3)	2 (3)
	Female	$16 \times 3 = 48$	3 (6)	1 (2)	1 (2)	1 (2)
Relative organ wt.	Male	$20 \times 3 = 60$	20 (33)	8 (13)	9 (15)	9 (15)
	Female	$16 \times 3 = 48$	13 (27)	7 (15)	7 (15)	8 (17)
	Male	1,032	275 (27)	143 (14)	165 (16)	158 (15)
	Female	1,008	238 (24)	125 (12)	155 (15)	150 (15)
Total		2,040	513 (25)	268 (13)	320 (16)	308 (15)

Four dose levels including these studies.

* Including of body weight gain.

**Including of any total consumption.

Dunnett #: Dannett's Multiple Comparison Test with Bartlett's Test.

Dunnett ## ; Dannett's Multiple Comparison Test without Bartlett's Test.

Table 8. Comparison of Student's t Test and Dannett's Multiple Comparison Test with Bartlett's Test, Dannett's Multiple Comparison Test without Bartlett's Test and Decision Tree for Number of Statistically Significance from Control Group on Various Parameters for 2 Years Carcinogenicity Studies in Mice

Items	Sex	Number of analysis	No. of significant differences (p < 5%), (%)			
			Student	Dunnett #	Dunnett ##	Decision Tree
Body weight*	Male	57 × 3 = 174	127 (73)	104 (60)	113 (65)	107 (61)
	Female		98 (56)	21 (12)	73 (42)	25 (14)
Food consumption**	Male	120 × 3 = 360	76 (21)	39 (11)	44 (12)	40 (11)
	Female		67 (19)	24 (7)	31 (9)	24 (7)
Hematology	Male	30 × 3 = 90	19 (21)	2 (2)	5 (6)	4 (4)
	Female		10 (11)	2 (2)	4 (4)	4 (4)
Absolute organ wt.	Male	5 × 3 = 15	1 (7)	0 (0)	2 (13)	0 (0)
	Female	4 × 3 = 12	0 (0)	0 (0)	0 (0)	0 (0)
Relative organ wt.	Male	5 × 3 = 15	9 (60)	8 (53)	9 (60)	8 (53)
	Female	4 × 3 = 12	5 (42)	1 (7)	1 (7)	1 (7)
	Male	654	232 (35)	153 (23)	173 (26)	159 (24)
	Female	648	180 (28)	48 (7)	109 (17)	54 (8)
Total		1,302	412 (32)	201 (15)	282 (22)	213 (16)

Four dose levels including these studies.

* Including of body weight gain.

** Including of any total consumption.

Dunnett # ; Dannett's Multiple Comparison Test with Bartlett's Test.

Dunnett ## ; Dannett's Multiple Comparison Test without Bartlett's Test.

Table 9. Number of Significant Differences by F-Test from Fig.2

Items	Weeks into dosing							
	Rats				Mice			
	0-25 (26)	27-52 (52)	53-78 (78)	79-104 (104)	0-25 (26)	27-52 (52)	53-78 (78)	79-104 (104)
Body weight	23/318 (7%)				68/318 (21%)			
Food consumption *	90/672 (13%)				115/672 (17%)			
Hematology +	42/168	43/168	51/168	60/168	Not tested	14/60	18/60	29/60
Blood chemistry								
Absolute organ wt.	9/33	9/33	10/33	7/33	—	—	—	12/27
Relative organ wt.	4/33	8/33	12/33	14/33	—	—	—	13/27

* Excluding of 0 week. () : Sacrificed weeks after dosing.

参 考 文 献

- 1) H. F. Kling, M. W. Moeller, B. L. Damron, R. H. Harms, C. L. Quarles, L. M. Potter, W. L. Beane, B. C. Dilworth, E. J. Day and S. A. Edgar, Response to Flavomycin and 3-Nitro in Broiler Chicken Diets Containing Clopidol, *Poultry Science* 55: 694-699, 1976.
- 2) M. W. Stutz, S. L. Johnson, F. R. Judith and B. M. Miller, *In Vitro* and *In Vivo* Evaluations of the Antibiotic Eftromycin, *Poultry Science* 62: 1612-1618, 1983.
- 3) M. W. Stutz, S. L. Johnson, and F. R. Judith, Effects of Diet, Bacitracin, and Body Weight Restrictions on the Intestine of Broiler Chicks, *Poultry Science* 62: 1626-1632, 1983.
- 4) 柴田寛三、新・生物統計学序説、創文、東京、1983.
- 5) 日本製薬工業協会、医薬品評価委員会・基礎研究部会、実験動物技術者のための教育資料「統計」1984.
- 6) Swinscow T. D. V. 著・西村昂三監訳、大島邦夫訳、統計処理、共立出版、東京、1982.
- 7) Student, *Biometrika* 6, 1, 1908.
- 8) 佐久間昭、生物検定法、東京大学出版会、東京、1967.
- 9) 吉村功、毒性・薬効データの統計解析、サイエントリスト社、東京、1987.
- 10) Shayne Gad and Carrol S. Weil, *Statistics and Experimental Design for Toxicologist*, The Telford Press, USA, 1986.
- 11) Shayne C. Gad and S. W. Carrol, *Statistics for Toxicologists, Principles and Methods of Toxicology*, edited by A. Wallace Hayes, Raven Press, New York, 1982.
- 12) Jonckheere, A. R., A Clisaribution-Free K-Sample Test Against Ordered Alternatives, *Biometrika*, vol. 41: P133-145, 1954.
- 13) 高橋行雄、薬量相関の有意性について、Jonckheere test, 医薬安全性研究会、第11号、P12-16, 1983.
- 14) 「毒性に関する試験成績を作成するに当たっての指針」農林水産省農蚕園芸局植物防疫課、59農蚕第4200号、昭和60年1月28日.
- 15) 「医薬品の製造(輸入)承認申請に必要な毒性試験のガイドラインについて(その1)」厚生省薬務局、薬審第118号(昭和59年2月15日).
- 16) 新毒性試験法 一方法と評価一、180頁、リアライズ社、東京、1985.
- 17) 大高忠彦、加藤忠彦、葉山恵子、三好幸二、三和和、仲吉洋、Ipriflavone (TC-80)のラットにおける1年間経口投与毒性試験、*応用薬理*、31(1)、137-154, 1986.
- 18) 工藤南雄、大川哲司、中村満利子、丸山潔、伊海正徳、Urapidilの毒性学的研究(第2報)ラットにおける12ヵ月経口投与慢性毒性試験、*応用薬理*、33、(3)、473-500, 1987.
- 19) 井本精一、青木道子、新保幸太郎、山下和正、Proglumetacin maleateのイヌの経口投与による慢性毒性試験、*The Journal of Toxicological Sciences* Vol.12, Supplement I, 1~33, 1987.
- 20) Duncan, D. B., Multiple range test and multiple F test. *Biometrics*, 11: 1-42, 1955.
- 21) 柴田寛三、生物統計学講義、東京農業大学育種学研究室編、東京、1970.
- 22) 吉田実、畜産を中心とする実験計画法、養賢堂、東京、1980.
- 23) Tukey, J. W. [unpublished] *The Problem of Multiple Comparisons* Princeton University.
- 24) 小林克己、ダンネットの多重比較検定法について、*医薬安全性研究会*、第10号、P11~15, 1983.
- 25) Dunnett, C. W., New table for multiple comparisons with a control. *Biometrics*, september 482-491, 1964.
- 26) Dunnett, C. W., A multiple comparison procedure for comparing several treatment with control. *J. Am. stat. Assoc.*, 50; 1096~1121, 1955.
- 27) 奥村純市、木野勝敏、ニワトリヒナにおける飼料尿素及びクエン酸2アンモニウムの成長促進効果、*日本家禽学会誌*、Vol.21、(2)、49-56頁、1984.
- 28) 秋葉征夫、高橋和昭、松本達郎、鶏の肝臓脂質蓄積と血漿トランスアミナーゼ活性に対する四塩化炭素経口投与の影響、*日本家禽学会誌*、Vol.20 (5) 277-283頁、1983.
- 29) Chi. M. S, C. J. Mirocha, H. J. Kurtz, G. Weaver, F. Bates and W. Shimoda, Subacute Toxicity of T-2 Toxin in Broiler Chicks *Poultry Science* 56: 306-313, 1977.
- 30) 統計学集中講座、日本卒後教育センター、東京、1987.
- 31) 奥田教隆、高木司郎、堀添宏、中村忠男、津曲立身。(士) 2-[(*o*-(2-Thenyl) phenoxy)methyl] morpholine maleate (Y-8894) のラットにおける26週間経口投与慢性毒性試験、*応用薬理* 33 (6)、907-927, 1987.

Trial on Statistical Analysis for Toxicity Studies using Laboratory Animals

Katsumi Kobayashi
Chief, Data Processing & Reporting Group,
First Toxicology Research Department,
An-Pyo Center

There are many methods of statistical analysis for toxicity studies which have parametric data, but it is very difficult to select the appropriate method to apply on this analysis. Recently, Dr. Shayne Gad and Carrol S. Weil published a paper, titled "Decision Tree for Selecting Hypothesis-testing Procedures" in the *Statistics and Experimental Design for Toxicologists* in 1986. A few laboratories are using this decision tree in the references. These decision tree methods include an analysis of Bartlett's Test. If it shows a significant difference, these parametric data were to be used as the nonparametric data. The statistical analysis for impartiality and good detect capacities were studied by two long term tests for 2 years. They were the chronic feeding and oncogenicity studies in rats and the carcinogenicity studies in mice.

The method was as follows: The parameters used were for body weight, food consumption, water consumption, hematology, blood chemistry, urinalysis, absolute organ weight and relative organ weight, 116, 240, 12, 88, 136, 16, 36 and 36 respectively, a total of 680 parameters in 80 rats/group \times 4 group study. Another study, the parameters used were for body weight, food consumption, hematology, absolute organ weight and relative organ weight, 116, 240, 60, 9 and 9 respectively a total of 434 parameters in 70 mice/group \times 4 group study. These parameters were analyzed by Bartlett's Test for homogeneity of variance. Next, number of significant differences ($p=0.05$) were counted among the groups by the Student t-test, Dunnett's multiple range test with Bartlett's test, Dunnett's multiple range test without Bartlett's test and modified decision tree, and compared with the number of significant differences from the control group.

The rates of significant differences by Bartlett's test were 27.3% (186/680) in studies of rats, and 26.2% (114/434) in studies of mice. Especially, the food consumption, hematology, blood chemistry and relative organ weight showed an increase in significant differences due to aging. The rates of significant differences ($p=0.05$) were 25, 13, 16 and 15% in the Student t-test, Dunnett's multiple range test with Bartlett's test, Dunnett's multiple range test without Bartlett's test and modified decision tree method, in totality 2,040 parameters.

The rates of significant differences ($p=0.05$) in the Student t-test, Dunnett's multiple range test with Bartlett's test, Dunnett's multiple range test without Bartlett's test and modified decision tree method were 32, 15, 22 and 16%, respectively, a total of 1,302 parameters. Numbers of significant differences in the Student t-test were 1.5 times as many as compared with the other 3 statistical analysis methods. Since, the Student t-test does not accept the effects of the other 2 group's variance, the Student t-test is the most recommended. And, if F-test shows significant differences, the Aspin-Welch or Cochran method is recommended.

Supplementary, almost all analyzers including blood analyzer and measurements of body weight, food consumption and organ weight are built in a computer. Therefore, there seems to be no artificial errors due to exhaustion.

チューキーの方法は群の大きさが違っていても使える

佐野正樹*

毒性・薬効データの解析において、すべての群間の平均値の違いを検定するときは、各群の分散が等しいという条件下でT u k e y・チューキーの方法が使われます。ところで、チューキーの方法の有意限界値は各群の例数（大きさ）が一定の場合について作られたものであることから、その大きさが異なっているときには、S c h e f f é・シェフェの方法を使わざるを得ないと今まで考えられてきました。しかし、シェフェの方法はあらゆる対比について有意水準を確保できるように考えられた手法です。したがって、群間の平均値を比較するだけの場合は、シェフェの方法はチューキーの方法よりも検出力の点で劣るという問題があります。これに関連する議論は医薬安全性研究会の定例会でも折りにふれ行われてきました。

ところで、1989年1月9日から2日間にわたって、東京（広尾）の統計数理研究所で『毒性・薬効データの解析における多重比較方式の標準化研究会』が開催されました。この研究会において、チューキーの方法はアンバランス（群の大きさが異なる）データに対しても使えることが、名古屋大学松田真一先生から紹介されました。この拡張されたチューキー法は、やや保守的な結果を与えるものの、手順が平易であることや有意水準が十分正確にコントロールされる点で、非常に有用な方法と考えられています¹⁾。私達にも参考になる話題と思われまので、この研究会に出席されなかった方々にも、誌面を借りてお知らせいたします。なお、この方法の手順は、群の大きさが同じ場合に通称ピンク本のそれと一致しますが、念のためピンク本の記述形式にしたがって説明しておきますので、本文（p. 57～59）と対応させながらご確認下さい。

- 1) YOSEF HOCHBERG & AJIT C. TAMHANE (1987) : Multiple Comparison Procedures; 3.2.1 The Tukey-Kramer Procedure, p.91-93, John Wiley & Sons, Inc., New York.

* 徳島市川内町平石字夷野 2 2 4 - 2 〒771-01
大鵬薬品工業株式会社 安全性研究所 一般毒性研究室

§ 2.3.1 チューキーの多重比較

[基本手順]

この手順で扱うデータは次の形式のものである。

$$\begin{aligned} \text{第1群} &: x_{11}, x_{12}, \dots, x_{1n_1} \\ \text{第2群} &: x_{21}, x_{22}, \dots, x_{2n_2} \\ &\quad \cdot \cdot \cdot \cdot \cdot \cdot \\ \text{第a群} &: x_{a1}, x_{a2}, \dots, x_{an_a} \end{aligned}$$

データは1元配置と同じ形で、各群の大きさは必ずしも等しくなくてよい。ここで説明する手法は、群の間で母平均に差があるとして、どの群とどの群との間に差があるかを判定するためのものである。

手順1) 各群における平均と、全データ数 n を次の式で求める。

$$\begin{aligned} \bar{x}_1 &= (x_{11} + x_{12} + \dots + x_{1n_1}) / n_1 \\ \bar{x}_2 &= (x_{21} + x_{22} + \dots + x_{2n_2}) / n_2 \\ &\quad \cdot \cdot \cdot \cdot \cdot \cdot \\ \bar{x}_a &= (x_{a1} + x_{a2} + \dots + x_{an_a}) / n_a \\ n &= n_1 + n_2 + \dots + n_a \end{aligned}$$

手順2) 次の式で、平方和、自由度を計算し、不偏分散 V_E を求める。

$$\begin{aligned} S_E &= (x_{11} - \bar{x}_1)^2 + (x_{12} - \bar{x}_1)^2 + \dots + (x_{1n_1} - \bar{x}_1)^2 \\ &\quad + (x_{21} - \bar{x}_2)^2 + \dots + (x_{2n_2} - \bar{x}_2)^2 + \dots \\ &\quad + (x_{a1} - \bar{x}_a)^2 + (x_{a2} - \bar{x}_a)^2 + \dots + (x_{an_a} - \bar{x}_a)^2 \\ \nu_E &= n - a, \quad V_E = S_E / \nu_E \end{aligned}$$

手順3) 有意水準を定め、これを α とする。

手順4) 統計数値表のステューデント化した範囲のパーセント点の表で、 $n = a$, $\nu = \nu_E$ のときの限界値 $q(a, \nu_E, \alpha)$ を求める。

手順5)
$$|\bar{x}_i - \bar{x}_j| \geq \frac{q(a, \nu_E, \alpha)}{\sqrt{2}} \sqrt{V_E (1/n_i + 1/n_j)}$$

ならば、第 i 群の母平均と第 j 群の母平均とには、差があると判定する。

または、次式の t_{ij} が $q(a, \nu_E, \alpha) / \sqrt{2}$ 以上であれば差があると判定する。

$$t_{ij} = \frac{|\bar{x}_{i.} - \bar{x}_{j.}|}{\sqrt{V_E (1/n_i + 1/n_j)}}$$

[解説]

- i) 各群の分散が等しいことを前提にした手法である。
- ii) 各群の大きさが違っていても有意水準が保たれることは、
H a y t e r (1 9 8 4) : A proof of the conjecture that the
Tukey-Kramer multiple comparisons procedure is conservative,
Ann. Statist., 12, 61-75 によって示された。
- iii) 1 因子多群の実験で、1 元配置分散分析が有意になったとき、どの 2 群の
平均の差が有意であるかを評価するとき用いられることが多い。しかし、
チューキーの方法自身は事前の分散分析を前提にしていない。

[数値例]

例 1 - 3 で、第 1 群から第 5 群までを考えたとして、どの群間に有意な差があるかを調べよう。データは次の通りである。このデータでは実際は等分散が成り立っていないのであるが、それを無視して考えることにする。

第 1 群 : 無処置群	156 155 156 157 154 155 154 153 152 155
第 2 群 : 対照群	153 153 152 156 158 151 151 150 148 157
第 3 群 : A 1 群	158 152 152 152 151 151 157 147 155 146
第 4 群 : A 2 群	153 146 138 152 140 146 156 142 147 153
第 5 群 : A 3 群	137 139 141 141 143 133 147 144 151 156

手順 1) 各群における平均は次の通り。群の数は $a = 5$, 各群の大きさは $n_k = 10$ である。

$$\bar{x}_{1.} = 154.7, \quad \bar{x}_{2.} = 152.9, \quad \bar{x}_{3.} = 152.1, \quad \bar{x}_{4.} = 147.3, \quad \bar{x}_{5.} = 143.2$$

$$n = 10+10+10+10+10 = 50$$

手順 2) 平方和と自由度は、次の値となる。

$$S_E = (156-154.7)^2 + \dots + (155-154.7)^2 + (153-152.9)^2 + \dots$$

$$+ (156-143.2)^2 = 989.6$$

$$\nu_E = 50-5 = 45, \quad V_E = 989.6/45 = 21.99$$

手順3) 有意水準を0.05にする。α=0.05である。

手順4) n=5, ν=45のときの限界値は表にないので,
 $q(5, 40, 0.05) = 4.0391$, $q(5, 60, 0.05) = 3.9774$ より, 補間で
 求めると,

$$q(5, 45, 0.05) = \frac{120/40 - 120/45}{120/40 - 120/60} \cdot 3.9774 + \frac{120/45 - 120/60}{120/40 - 120/60} \cdot 4.0391$$

$$= 4.0185$$

手順5)

$$|\bar{x}_i - \bar{x}_j| \geq \frac{4.0185}{\sqrt{2}} \sqrt{21.99(1/10 + 1/10)} \approx 5.959$$

となる。

いろいろな群の組合せを「基本手順」の手順5)で示した t_{ij} の
 計算で調べると, 下に示すように無処置群とA2, A3群の間, および対
 照群とA3群の間, A1群とA3群の間とに有意な差があることになる。
 ただし, **の付いたときは, 有意水準1%でも有意である。

	対照群	A1群	A2群	A3群
無処置群	0.858	1.240	3.529**	5.484**
	対照群	0.381	2.670	4.625**
		A1群	2.289	4.244**
			A2群	1.955

以上

自由度についての補足

吉 村 功 (名古屋大学工学部)

1. 補足のきっかけ

医薬安全性研究会の会報 No.27 にあった、「自由度をめぐる6つの疑問」での自由度の説明は、どうもいま一つ、もの足りない。それは、数学的構造の説明を避けているためである。自分でも不満なので説明を追加する。

2. 自由度の区別

自由度については、分布族における“母数としての自由度”と、分散分析における“統計量の自由度”と、仮説検定における“仮説の自由度”とを、区別して説明した方がよさそうである。

3. 母数としての自由度

まず最初の、分布族の母数としての自由度については、前述の会報の説明で十分である。

4. 統計量の自由度

いま、 Y_1, Y_2, \dots, Y_n を、観測値とする。これについて、

$$\text{平方和 } S = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

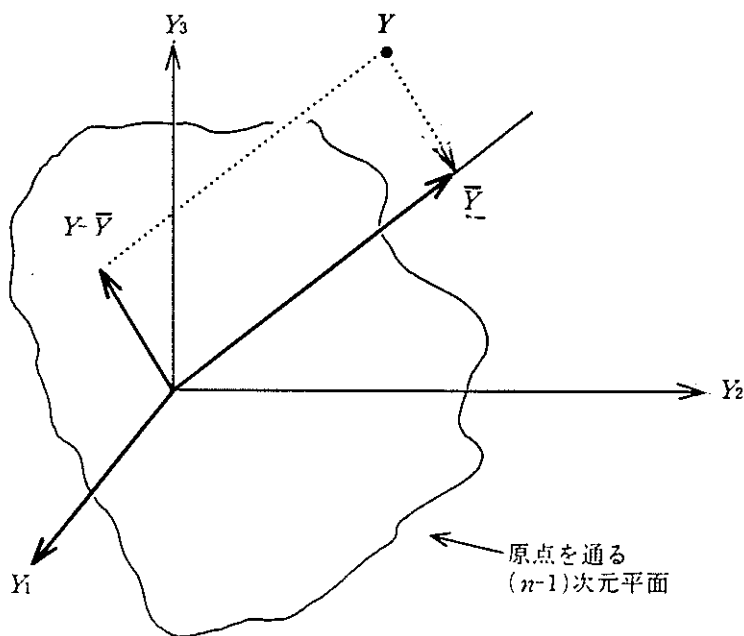
がどれだけ「自由に動けるか」考えよう。

Y_1, Y_2, \dots, Y_n を、 n 次元空間の座標 $Y = (Y_1, Y_2, \dots, Y_n)$ とみなそう。同じ空間に $\bar{Y} = (\bar{Y}, \bar{Y}, \dots, \bar{Y})$ という点をとると、これは、 $(1, 1, \dots, 1)$ というベクトルを含む直線へ Y から垂線を下ろした点になっている。そして、垂線をベクトル $Y - \bar{Y}$ とみなすと、このベクトルはかならずベクトル $(1, 1, \dots, 1)$ と直交する平面： $Y_1 + Y_2 + \dots + Y_n = 0$ に含まれる。すなわち、 Y からこの平面に垂線を下ろした点と原点との差が $Y - \bar{Y}$ である。観測値 Y が自由に動いても、 S はこの平面上の点の関数でしかなく、 S はこの $n-1$ 次元平面上の点 $(Y - \bar{Y})$ と原点との距離の二乗である。(この幾何学的関係を、イメージとして理解するには、 $n=3$ 次元空間で図を画いて考えるのがよい。)

さて、 Y_1, Y_2, \dots, Y_n を、互いに独立に $N(\mu, \sigma^2)$ に従う確率変数としよう。このとき、統計量

$$S = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

の分布のひろがりかたは、この平面の次元 $n-1$ によって決まる。(厳密にいうと σ^2 も影響するが、それはここでの議論で本質的でないから無視する。) もちろん次元が大きいほど、分布のひろがり大きい。だから、この次元の数を自由度と名づけて、“平方和 S の自由度は $n-1$ である”というのは、用語法として自然であり、合理的である。これが平方和という統計量に自由度なる概念を対応づける理由である。



計算すると S/σ^2 の分布はカイ二乗分布であり、カイ二乗分布は性質として、平均がその母数の値に等しく、分散もそれに比例する。母数の値は今の場合この次元数 $n-1$ に等しい。それゆえ、カイ二乗分布の母数を自由度という、上に述べた自由度の用語法とつじつまがあって都合がよい。

この説明の中で、統計量の分布のひろがり自由度という形で決まるための十分条件は、①観測変数の分布が正規分布であり、②統計量が、原点を通るある平面に下ろした垂線の足と原点との距離の二乗という形の式（2次の同次式）であることである。したがって、統計量が上の形の平方和 S でなくて、

$$S_R = \sum_{i=1}^n |Y_i - (\hat{\alpha} + \hat{\beta} x_i)|^2$$

$$\text{ただし、 } \hat{\alpha} = \bar{Y} + \hat{\beta} \bar{x}, \quad \hat{\beta} = \frac{\sum (Y_i - \bar{Y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

という形のものであっても、同じ様な議論で自由度が定義できる。この場合は、 Y から下ろした垂線が含まれる平面は、 $(1, 1, \dots, 1)$ と (x_1, x_2, \dots, x_n) の両方に直交する平面であり、その次元は $n-2$ である。従ってこの平方和 S_R の自由度は $n-2$ である。

もちろんこのとき、平方和の期待値は会報 No.27 で述べたように σ^2 に自由度をかけたものになっている。これを確かめるための式の計算はめんどうなので省略する。

5. 自由度の分解

一般の分散分析の書物を読んでいると、“自由度の分解”というのがあるが、読んでさっぱり分からない説明をしていることがある。これは次のように考えるのがよい。

平方和 S は、適当に Y の 1 次式 L_1, L_2, \dots, L_{n-1} を選ぶと、必ず次のように二乗和の形におおせる。

$$S = \sum_{i=1}^n (Y_i - \bar{Y})^2 = L_1^2 + L_2^2 + \dots + L_{n-1}^2$$

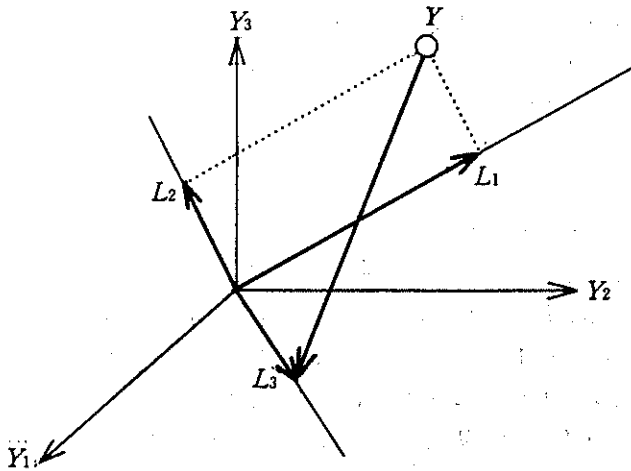
もちろん、 $n-1$ 個の 1 次式 L_1, L_2, \dots, L_{n-1} は、これが成立するようにうまく選ばなければならない。たとえば、 $n=2$ の場合は、

$$S = \left(\frac{1}{\sqrt{2}} Y_1 - \frac{1}{\sqrt{2}} Y_2 \right)^2$$

とすればよい。

上の式は、空間的にいうと、原点を通り互いに直交する直線を $n-1$ 個とって Y からそこに垂線をおろしたとき、 S が、各点の原点からの距離の二乗和になっていることを式で表したものである。すでに述べた自由度の定義によると、 $L_1^2, L_2^2, \dots, L_{n-1}^2$ はそれぞれ自由度 1 の平方和であるというのが合理的である。(二乗和も平方和も英語では Sum of Squares である。したがって日本語でも“平方和”で統一してよいのであるが、上の説明では、本当に二乗の和になっているというニュアンスを強めるために、二乗和という言い方をした。)

このような $n-1$ 個の 1 次式 L_1, L_2, \dots, L_{n-1} の選び方は、 n が 3 以上ならば無数に沢山ある。その中に、直面する現実問題で、ある仮説が真ならば誤差的なばらつきの大きさになり、仮説が偽ならば、誤差的なばらつきよりずっと大きな値になるようなものがある。しかもそのとき、個々の平方和ではなく、これらの平方和をいくつか加えたものを一つとして考えたほうが意味のあることがしばしばある。



$$S = S_1 + S_2 + \dots + S_q$$

↓ ↓ ↓

これらの一つ一つが上の L をいくつかあわせたもの。

その自由度は L をいくつか合わせたかで勘定する。

そのようなときには、その加えたものも平方和と呼ぶにふさわしい統計量である。そして“その自由度はそれぞれの自由度の和である”というふうに自由度の概念を定義すると、今までの説明がすべてそのまま使えて都合がよい。そこで一般の平方和においても、自由度をそう定義することが一般になされている。

分散分析で自由度の分解というのは、実は、自由度の分解ではなく、平方和 S をこのように現実問題で意味のある平方和に分解する（つまりいくつかの平方和を加えたものにする）ということである。その各平方和の自由度を全部加えると $n-1$ になるものだから、平方和の分解を自由度の分解という言い方で表現しているのである。“自由度の分解”という説明の仕方は、そういう背景を知っている人が使う分には簡単でよいのだが、一般の読者はそうでないことが多いから、一般には本質を誤解させるよくない説明である。

6. 仮説の自由度について

母数が、たとえば $\mu_1, \mu_2, \dots, \mu_p$ と p 個あって、それについて

$$H: \mu_1 = \mu_2 = \dots = \mu_p$$

という仮説があったとしよう。 $\mu = (\mu_1, \mu_2, \dots, \mu_p)$ を p 次元空間の点と見なすと、仮説によって、母数 μ の動きうる範囲は、 $(1, 1, \dots, 1)$ というベクトルの延長直線上に限られる。すなわち、仮説 H がなければ、 μ は p 次元空間の任意の点を動けたのに、仮説 H によって、母数の値が、 $p-1$ 次元分の自由度を制限され、動き得る範囲を1次元におとされたわけである。このとき、「この仮説の自由度は $p-1$ である」という。

この場合には、仮説によって母数の動き得る範囲が狭められる程度を自由度というのだから、用語法としては奇妙である。むしろ“不自由度”といたいくらいである。それなのに自由度というのは何故だろうか。

それは、検定統計量のためである。もし正規分布を前提にするならば、この仮説を検定するための統計量としては、自由度が $p-1$ の平方和を用いることになる。だから仮説に対しても、この不自由になる程度を自由度と名づけておくと、検定のときに、検定のための棄却限界値を、形式的に間違いなく選ぶことができ都合がよい。仮説の自由度の場合は、こういう意味でちょっと特殊な用語法になっている。

7. 分割表とカイ二乗検定

今までは、正規分布をイメージに画きながら、自由度を説明してきた。これに対して、“ 2×2 分割表の自由度は1である”というときの自由度はどうだろうか。

この場合も、自由度という用語の意味は上に説明したものと同じである。ただし、これらの場合には、議論が局所的な近似の世界で行われる。

たとえば、母集団 O_1 から大きさ n_1 の標本を選んだときにある性質を持った個体が X_1 だったとしよう。母集団 O_2 から大きさ n_2 の標本を選んだときにある性質を持った個体が X_2 だったとしよう。母出現率を π_1 、 π_2 とすると、その 2×2 分割表は次の形になる。

	性質あり	性質無し	計
母集団 O_1	X_1	$n_1 - X_1$	n_1
母集団 O_2	X_2	$n_2 - X_2$	n_2
計	$X_1 + X_2$	$n_1 - X_1 + n_2 - X_2$	$n_1 + n_2$

仮説として、 $H: \pi_1 = \pi_2$ を考えると、これは自由度が1である。仮説がなければ、 (π_1, π_2) は、 $0 \leq \pi_1, \pi_2 \leq 1$ という制約はあっても、一応2次元空間を自由に動けたのに、仮説が成り立つと、 $\pi_1 = \pi_2$ という直線上、つまり1次元空間に動きを制限されたからである。

この仮説に対する検定統計量は、よく知られたカイ二乗統計量であるが、これは n_1 、 n_2 の両方が十分大きいとき、正規分布を前提とした自由度1の平方和と同じようなふるまい、分布をする。したがってこの場合も、標本が十分大きいという前提、いわゆる“漸近理論”の枠組みでは、自由度という概念を今までと同じように理解して差し支えないのである。

8. 歴 史

私の文章のことを知った、後藤昌司氏(塩野義製薬)から、「全く同じ問題を数年前(1982年ごろ)に、人に尋ねられて調べ、アメリカでもこれが問題になっていて、論文があることを知り、Good, I. J. "What are degrees of freedom?", American Statistician, 27, 227-228, (1973) という文献を紹介したことがある。」と教えられた。洋の東西と時を問わずに、この問題が議論になっていることを知ってたいへん驚いた。

後藤氏からと同じ頃、増山元三郎氏(東京理科大学)からも、「Encyclopedea of Statistical Sciences の K. Pearson の箇所を調べてみたら予想通り、彼は“自由度の考え方を全く理解しなかった”と書いてあった。」というお便りをいただいた。Encyclopedea of Statistical Sciences は Degrees of Freedom の項しか調べていなかったのので、あわてて、R. A. Fisher, W. G. Gosset, K. Pearson の項を読みなおしてみたら、

K. Pearson の項では

「This led him in later years into many mathematical controversies in which he was wrong in the exact sense...」

とあり、R. A. Fisher の項では

「Fisher wrote to W. S. Gosset (Student) questioning his divisor ($n-1$) in the formula for the standard deviation. He then reformulated the problem in an entirely different and equally original way, in terms of the configuration of the sample in n -dimensional space, and showed that the use of the sample mean instead of the population mean was equivalent to reducing the dimensionality of the sample space by one; thus he recognized the concept of what he later called degrees of freedom.」
と書いてあった。

要するに、現在自由度という概念で示されているものは、最初は誰でも不思議に思えるようなものだった。だから、ピアソン氏もフィッシャー氏もそれをおかしなことだと思った。しかしフィッシャー氏はそれを数学的に吟味して、その内容が正しいことだと確信した。そしてやがてそれを自由度という概念に整理した。それをピアソン氏は受け入れなかった。ということのようである。

われわれ凡人が、自由度って分からないね、といっても不思議ではないし、洋の東西で同じ疑問が繰り返されているのも不思議はないのである。

これらを教えていただいた、増山氏と後藤氏に心から感謝したい。

1989.4.12

参考文献

Cramer, H. "Mathematical Methods of Statistics" p.234, p.381, (1946)
Walker, H. M. "Studies in the history of statistical method", The Williams & Wilkins Co. (1929),
(足利末男、辻博訳 "統計方法論史"、p. ix, 高城書店、(1959))
Tippet, L. H. C. "The Methods of Statistics", Wiley, pp.64-65, (1931)

正 誤 表 (会 報 No. 27, P.2)

行	誤	正
↓ 8	$\Sigma(X_i - \mu)^2$	$\Sigma(X_i - \bar{X})^2$
↓ 9	$\Sigma(X_i - \mu)^2 / n$	$\Sigma(X_i - \bar{X})^2 / n$

〔事務局だより〕

既に、小さなパンフレットでご承知のことと思いますが、今度当研究会ではテーマを絞ったモノグラフシリーズを刊行します。その第一弾『実験データのグラフ表示』（中里溥志著）が、原稿も完成し、6月末の発刊を目指して編集中です。今回は、ディスク・トップ・パブリッシングという新しい形式で組を行ないますが、今後、安全研の出版物に積極的に生かしていきたいと思います。また、秋には『医学・薬学・生物学のための統計学』を刊行すべく吉村先生が、鋭意執筆中です。ご期待下さい。随時、パンフレットなどで、ご案内致します。

(M. O.)

医薬安全性研究会

会報NO. 28

1989年5月31日 発行

定価1000円(本体971円)

編集・発行 (株)サイエンテスト社

〒101 東京都千代田区神田駿河台3-2 山崎ビル

☎ 03(253)8992

FAX 03(255)6847

振替 東京8-71335

印刷・製本 (有)ナガノ印刷