

# 医薬安全性研究会

*Bulletin of Japanese Society for Biopharmaceutical Statistics.*

会報 No. 37

July 1993

## 目次

〈前 付〉

\* 医薬安全性研究会スケジュール

---

〈基礎講座〉〔連載第1回〕

Practical Statistics for Medical Research 第5章 立案研究/Q & A

〔訳〕小西寿美恵（日本たばこ）……1

〔Q & A〕No. 106 カイ二乗適合度検定（分割表）についての質問

Q：大林繁夫（グレラン製薬）

A：稲葉太一（神戸大学）……13

検定における検出力の考え方とその計算方法について

稲葉太一（神戸大学）……15

毒性試験における施設内ヒストリカルデータを使うための条件

芳賀敏郎（東京理科大学）……22

ゲッ菌類を用いた毒性試験から得られる定量データに使用される統計処理法の動向

小林克己、大橋信之、大河原一弘（安評センター）……32

1993~1994年 医薬安全性研究会  
これからのスケジュール

- ☆ 1993年12月3日（金）～4日（土）……………データ解析講習会（総評会館）
  - ☆ 1994年1月29日（土）……………第57回定例会（総評会館）
  - ☆ 1994年4月16日（土）……………第58回定例会（総評会館）
- 

★ 秋のデータ解析講習会・日程変更のお知らせ

第54回定例会のご案内に、秋のデータ解析講習会を11月12日（金）～13（土）とお知らせしたのですが、講師スケジュールの都合で、12月3日（金）～4日（土）に変更させて頂きました。各位のスケジュールの書き換えをよろしくお願い申し上げます。

# BIostatISTICS IN CANCER RISK ASSESSMENT

Edited by D. G. Hoel (NIEHS) and T. Yanagawa (Kyushu Univ.)

A4 Size 176pp. ¥ 4,000 (¥ 310YEN)

---

This monograph is a set of selected papers from the two proceedings of the conferences that we organized in 1984 and 1989 in Japan. These proceedings are:

Environmental Risk Assessment and Statistical Methods,

*Environmental Health Perspectives*, Vol 63, 1985

The Third Japan-US Conference on Biostatistics in the Study of Human Cancer

*Environmental Health Perspectives*, Vol 87, 1989

The papers selected in this monograph cover important and essential topics relevant to a range of human health studies. After the publication of these proceedings, we have been asked by many mathematical scientists to make them more accessible to the statistical community so that they may be used for the purpose of training and education of Biostatistics.

## 〈Part of Contents〉

"Statistical Studies in Genetic Toxicology : A Perspective from the U.S. National Toxicology Program" by B. H. Margolin

"Multistage Models of Carcinogenesis" by P. Armitage

"Statistical Analysis of Disease Onset and Lifetime Data from Tumorigenicity Experiments" by S. W. Lagakos and L. M. Ryan

"Use of Historical Controls in Animal Experiments" by T. Yanagawa and D. G. Hoel

"Confidence Intervals for Effect Parameters Common in Cancer Epidemiology" by T. Sato

"Statistical Analysis of  $K2 \times 2$  Tables : A Comparative Study of Estimators/Test Statistics for Association and Homogeneity" by T. W. O'Gorman, R. F. Woolson, M. P. Jones, and Jon H. Lemke

"Relative Risk Regression Analysis of Epidemiologic Data" by R. L. Prentice

"Empirical Bayes Methods for Smoothing Data and for Simultaneous Estimation of Many Parameters" by T. Yanagimoto and N. Kashiwagi

"A Nonparametric Method for Estimating Interaction Effect of Age and Period on Mortality" by M. Ohtaki, D. Kim, and M. Munaka

"An Index for Cancer Clustering" by T. Tango

.....

---

Orders should be directed to :Scientist Inc.

Yamazaki - bldg. 3 - 2

Kanda Surugadai, Chiyodaku,

Tokyo, 101 Japan

# INTERNATIONAL BIostatISTICS CONFERENCE IN THE STUDY OF TOXICOLOGY

B5size 135pp. ¥ 3000.

---

## CONTENTS

BIOSUTATISTICAL ISSUES IN THE DESIGN AND ANALYSIS OF ANIMAL CARCINOGENICITY EXPERIMENTS .....	Christopher J. Portier
CARCINOGENICITY TESTING : STATISTICAL ISSUES .....	Yasuo Ohashi
THREE STATE MODELS FOR RODENT CARCINOGENICITY EXPERIMENTS	
	Louise Lyan and Jane Lindsey
A NOTE ON THE USE OF HISTORICAL CONTROLS ..IsaoYoshimura and Kazuhiko Matsumoto	
DATA-BASED SCORES IN EXACT TREND TESTS FOR FETAL SURVIVAL .....	
	Keith A. Soper and Robert L. Clark
STATISTICAL METHODS FOR THE BETA-BINOMIAL MODEL IN TERATOLOGY	
	Eiji Yamamoto and Takemi Yanagimoto
MULTIPLE COMPARISONS IN LONG-TERM TOXICOLOGICAL STUDIES .....	Ludwig Hothorn
MULTIPLE COMPARISON AMONG GROUPS OF GROWTH CURVES.....	Toshinari Kamakura
STATISTICAL EVALUATION OF MUTAGENICITY TEST DATA : RECOMMENDATIONS OF THE U. K. ENVIRONMENTAL MUTAGEN SOCIETY.(UKEMS).....	David J. Kirkland
STATISTICAL ANALYSIS OF DATA IN MUTAGENICITY ASSAYS : IN THE CASE OF RODENT MICRONUCLEUS ASSAY.....	
	M. Hayashi, S.Hashimoto, Y. Sakamoto, C. Hamada, T. Sofuni and I. Yoshimura
BIostatISTICAL ISSUES IN THE DESIGN AND ANALYSIS OF MULTIPLE OR REPEATED GENOTOXICITY ASSAYS.....	Lutz Edler
NONPARAMETRIC REGRESSION ANALYSIS OF THE AMES MUTAGENICITY ASSAY	
	John B. Cologne and Norman E. Breslow
DATA MANAGEMENT FOR TOXICOLOGICAL STUDY .....	Ikuro Horii
STATISTICAL APPROACHES FOR STUDY OF THE INDUCTION KINETICS FOR SPECIFIC- LOCUS MUTATION IN <i>Neurospora crassa</i> .....	Frederick J. de Serres
POTENCY MEASURES IN MUTAGENICITY RESEARCH .....	Barry H.Margolin
STATISTICAL ISSUES ON THE NO-OBSERVED-ADVERSE-EFFECT LEVEL IN CATEGORICAL RESPONSE.....	Takashi Yanagawa, Yasuhiro Kikuchi and Kenneth G.Brown
PHYSIOLOGICALLY-BASED PHARMACOKINETICS IN CANCER RISK ASSESSMENT .....	
	Melvin E. Andersen

.....

---

Order should be addressed to : Scientist Inc.Yamazaki bldg. 3 - 2 Kanda Surugadai, Chiyoda - ku  
Tokyo 101, JAPAN

**こうすれば内容豊かな質問になる！**

## 医薬安全性研究会 質問募集

医薬安全性研究会では現場での疑問をもとに、より興味深い議論を行いたいと考えています。しかし、質問の仕方がうまくないために答えが面白くなくなっている場合があります。そこで次のように質問を募集します。

### 〔要 項〕

#### 〈答えにくい質問が多い〉

- 1) 抽象的な質問……「Q：分散分析の位置づけは？」  
「A：野球におけるキャッチボールのようなものだ。」
- 2) 状況説明が一般的……「Q：併用効果を調べるために、1. 対照群 2. Aを投与 3. Bを投与 4. AとBを投与という4群の実験をしたときの検定方法は？」  
「A：単純な2元配置実験なので2元配置実験のデータ解析をするといよい。」

#### 〈質問は具体的に〉

- 1) 質問の状況をできるだけ明らかに
  - 群の大きさ、群の数、動物種／● 測定変数は何か(血圧、時間、白血球数……)／
  - 測定値の例／● 測定条件／● 実験の目的 etc.
- 2) 具体的な説明が必要な理由
  - 質問者が気がついてないことで大切なことがある。  
「イヌの実験で実験者は一元配置分散分析だと考えていたのが二元配置型の実験であった。」
  - 前提とする分布等を知ることにより、より適切な解答となる。

#### 〈公開の原則〉

- 質問は医薬安全性研究会で公表し、討論することを原則とする。

#### 〈公表してほしくない部分への配慮〉

- 公表してほしくない部分は、質問の際に付記してあれば尊重する。  
質問者、対象動物種、疾患名、具体的な名前(測定変数、値の大きさ)

#### 〔質問送付方法〕

- \* 質問は随時受付けています。
- \* 裏面の質問用紙をコピーして御記入の上、右記宛お送り下さい。

#### 医薬安全性研究会事務局

〒101 東京都千代田区神田駿河台3-2 山崎ビル  
TEL 03<253>8992/FAX 03<255>6847  
/振替 東京8-71335

株式会社 サイエコディスト社 内

# 質問用紙

No.

御氏名		会員No	非会員	年 月 日
<p>&lt;具体的状況&gt;</p>				
<p>&lt;質問内容&gt;</p>				

(コピーしてお使い下さい)

<基礎講座> [連載第一回]

# Practical Statistics for Medical Research

## Chapter 5 Designing research

by Douglas G. Altman

訳 小西 寿美恵  
(日本たばこ産業)

### 5 立案研究

研究立案ほど無視される臨床研究の局面はおそらくない。熱心な若い研究者たちは医学統計学の講義に出席し、「P値」を算出するための多くの方法を見つけ出す。適切な臨床研究計画を構築する方法をめったに学ばない。しかし入念な研究計画はすぐれた臨床研究の基本である。

Noller and Melton(1986)

研究を誤る何千本もの道があるが、研究を適切に行うための道はほんの少ししかない。

Sackett(1986)

#### 5.1 序

あらゆる医学研究は一つあるいはより多くの目的について行われるが、目的により研究案あるいは研究計画を決めておくべきである。いくつかの事例では研究を先へ進めるための明確で良い方法があるが、妥当な研究計画の方法の中から選択することが多い。計画の統計学的な観点はおもに研究の構造及び測定の種類とそれらの頻度を含めてデータ収集の全ての局面に関係する。この章で扱われる一般論の多くは臨床試験に適用されるけれども、これら臨床試験には多くの特徴があり、詳細は第15章で議論される。

研究は「観察的研究」と「実験的研究」に大別できる。観察的研究では我々はいっあるいはそれ以上の被験者の群を対象に情報を集めるが、彼らに影響することはなにもしない。観察的研究は、被験者が集められ引き続いて起こる出来事についてデータを集める場合は「前向き研究」であり、過去の事柄について情報を集める場合は「後向き研究」ともいえる。観察的研究は人口調査、調査、患者-対照研究、ホート研究を含むが、それらは5.9～12節で詳細を述べる。

実験的研究は研究者がすべてあるいは一部の個体に発生する事象を左右(コントロール)することから成る。この章の重点は臨床研究であるが、人、動物及び実験

室サンプルの調査でも類似の問題が生じる。5.4～5.8節で実験的研究の立案について詳述する。

ほとんどの研究はかなり単純な問題に答えることがねらいであるからといって、かなり単純な計画になるとはかぎらない。重要なポイントは研究の目標に研究計画を合わせることである。十分な計画なしには研究者が意味のある結論を得ることは期待できない。研究立案に際してのいくつかの重要な一般原則はこの章の後半で述べる。

ほとんどの研究では、我々は研究から得た結果を一般の母集団に外挿することを期待する。この点において特に注意を要する2つの見地がある。第1は調査されるサンプルの代表性で、これは特に観察的研究にあてはまる。第2に、比較される複数の群は直接の研究対象となる主要点以外は可能な限り同じようにすべきである。これはとりわけ臨床試験のような実験的研究にあてはまり、また患者-対象研究の様な多くの観察的研究にもあてはまる。のちにこの問題にふれる。

統計学の医学に対する貢献の最も重要な面はおそらく研究計画である。そのために、50年以上のあいだ統計学者は医学研究者に、解析段階よりも彼らの研究立案の段階で相談することを促している。よい研究から得たデータは多くの方法で解析出来るが、研究立案の不備による問題を埋め合せ出来る適当な解析法はないのである。

## 5.2 研究計画の種類

研究計画は数種類の方法で分類できるがその内のいくつかを以下に示す。

1. 観察的なものか実験的なものか
2. 前向き（未来予測）か後向き（過去回顧）か
3. 縦断面（連続的）的なものか横断面（断片的）的なものか

これらの言葉はのちに説明する。1.の分類法は、他がデータの収集方法を詳述するのに対し研究の目的に関係する。これらの項目の組合せは全てが可能ではないが、大部分は可能である。

### 5.2.1 観察的なものか実験的なものか

「観察的研究」では研究者は研究対象の属性あるいは測定値に関する情報を収集するがその結果に影響を与えない。この一例は小さい子供の間における難聴の有病割合を発見するための研究であろう。観察的研究は調査と大多数の疫学的研究を含む。対照的に「実験的研究」では研究者は事象に対して故意に作用を及ぼしその干渉効果を調べる。実験的研究には臨床試験と動物及び実験室での研究が含まれる。一般に、より効果的な推論は観察的な研究よりも実験的な研究から生み出される。実験的研究は、普通は群間を比較することによって実施される。ただし観察的研究でも比較される場合があるが、たいていは本質的には記述的なものである。

### 5.2.2 前向き（未来予測）か後向き（過去回顧）か

研究の開始時点以降にデータが収集される「前向きの研究」と、過去の事象に基づき現存する病歴（病院の記録）のような情報源あるいは面接からデータが得られる「後向きの研究」の間には明かな違いがある。実験的研究は前向きの研究であるが、観察的研究は前向きの研究でもあり後向きの研究でもあり得る。もちろん後向きの研究のデータは例えば術式の違う乳房切除術のように異なる処理を比較するた



めに求められるが、このような研究は、あらかじめ明記された研究すなわち標準化された条件下で行われた研究ではないので実験とはいえないであろう。後向きの研究は患者-対照研究を含む（5.10節を参照）。

### 5.2.3 縦断面（連続的）的なものか横断面（断片的）的なものか

「縦断面的研究」はふつうは干渉に関して時間を越えて（経時的な）変化を調べることである。観察は1回以上を行うが、解析でその全てを使わないかも知れない。臨床試験は、ある時期に開始した処理がのちの結果に及ぼす作用に目が向けられるので縦断面的研究である。「横断面的研究」はたった1回だけ観察された個体の研究である。ほとんどの調査は、参照範囲を構築するための研究であるので横断面的である。観察的研究は縦断面的であり横断面的でもあり得るが、実験的研究は普通は縦断面的である。

各被験者はたった1回しか観察されなくても、そのデータが経時的な変化を詳しく説明するために使われるような模擬縦断面的な横断面的研究もある。その例は横断面的に子供たちの成長グラフを得る目的の研究や月経周期中のホルモンレベルの研究である（5.13参照）。

### 5.2.4 相互関係の要約

図5.1はいままで述べた計画分類の間のごく一般的な組合せをまとめたものである。ほとんどすべて前向きの研究であり縦断面的である実験的研究と、前向きのでもあり後向きのでもあり、また横断面的でもあり縦断面的でもある観察的研究には明かな違いがある。この理由のため、実験的研究と観察的研究についてはこの章のちで個々に検討する。

研究計画のもっと複雑な分類を組み立てることも可能である（Bailarら、1984）が、図5.1がほとんどの調査研究の主な特徴を詳述している。

これまでの立案の議論は広範な問題に関連付けられている。

後続の節では、標本から母集団への統計学的推測の重要な2つの側面、すなわちサンプルの代表性ならびに見いだされた関連性の解釈について特に強調し、立案の細部を検討する。

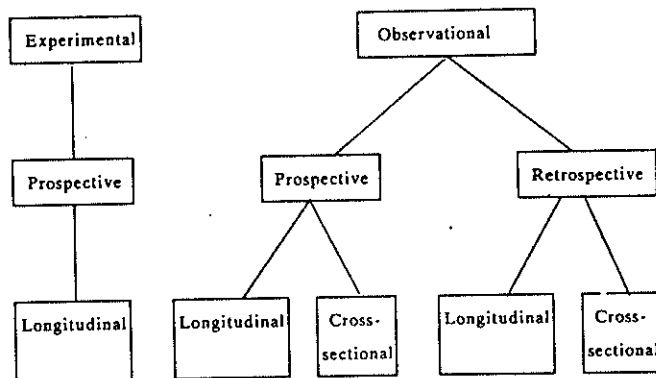


Figure 5.1 Types of research design.

### 5.2.5 対照

どんな実験でも実験の操作（処理）が適用されていない比較群あるいは対照群をおくことが重要である。通常「できたての処方患者に試して何が起こるか見てみましょう！」という提案は、科学的あるいは倫理的に認められない。比較できるように通常に処方（あるいはなんらかの違った方法を処方）をされた対照群をもつことがはるかに良いのである。もし我々が、妊娠期間中に母親が胎動を数えることの有益性を評価したいなら、我々はまた母親が胎動を数えない対照群も作るべきである。これは新しい治療あるいは医療手順の評価の不可欠要因である。

また観察的研究においても対照群を作ることが賢明である。もし我々が可視ディスプレイターミナル（VDTs）のユーザーに、彼らに目の痛みや腰痛があるかどうかを尋ねるのなら、我々はまたVDTsを使用していない比較できる従業員に同じ質問をすべきである。

どんな場合でも対照群の存在は研究の結果から作られるであろう推論を強化する。しかしながらもっとのちで述べるように観察的研究におけるふさわしい対照群の選択は簡単ではない。

### 5.3 変動要因

3章と4章は、いずれも統計学的アプローチに関して、変動の重要性についてのコメントから始めた。たとえそれがタバコや抗生物質の様なものであっても、ある種の刺激物に対する行動あるいは反応は変動するのがふつうである。3章に記載した様にいくつかの変動要因は、わかっているかもしれないし、想像できるかもしれない、がその多くは説明出来ない。例えば我々は出生児体重に影響を与えるいくつかの変動要因、すなわち妊娠期間、胎児の性別、出産経歴、母体の喫煙、海拔の高さ等を知っているが、そのような情報を取り入れている統計学的モデルは出生児体重での変動の4分の1についてだけを説明する。変動に影響を与えているがまだ確認されていない他の要因が確かにあるが、なんらかの重要な要因が未確認のままであるということはまずありえない。それ故、観察された大部分の変動は説明できないと考えざるをえないので、我々はそれらをランダム変動と呼ぶ。大部分の臨床的な測定値では、このランダム変動が相当重ある。体温のような一部のものについては相対的にわずかな変動であるが、出生児体重、血圧、あるいは多くの血清成分等は大きく変動する。我々がある臨床測定値に関して比較する群の研究を計画するときは、この自然変動の存在を肝に銘じなくてはいけない。我々は、このでたらめな変動のことを研究対象のシグナルあるいは効果を検出するためのバックグラウンドノイズと考えることが出来る。他の分野で使われる”ノイズに対するシグナル”比の概念になぞらえるのが良いのである。もしも測定結果が大きく変動するならば、研究対象の一貫した効果を検出するために我々は大規模な研究を必要とするであろう。考えられるもう一つの設計は、被験者の個体変動を研究することによって、被験者間の変動をベースラインのレベルから除外することである。

さらに個体は、直接研究されないが研究対象の変数に影響を与えるかも知れない別の要因に関しても同じように変動を示すであろう。実験計画の原則の多くは、我々が研究対象とする変動に焦点を合わせることが出来るように、研究対象としない変動をコントロールするための試みを目指している。研究計画に関連したこの一般的な変動性の2つの帰結は、以下の様である。

1. ある集団の代表的な標本をつくるためには注意が必要とされる。
2. 比較研究では既知の変動要因に関して類似の群を作ることに注意が必要とされる。

加えて、研究対象の測定値が大きく変動する時に、信頼出来る結果を得るためには多くの標本が必要なことを我々は心にとどめておくべきである。

これらの問題についてはまず実験的研究、続いて観察的研究についての順でのちに考察する。しかしながら比較研究のためのふさわしい標本のサイズを計算するための方法は、臨床試験を計画する際に最も頻繁に使われるので15章で示す。

実験計画の様々なタイプの詳細を調べる前に、問題点の多くが目には浮かぶような実際の研究を考慮することが助けになるであろう。

#### 5.4 実験：両腕の血圧は等しいか？

血圧は特に変動する測定値である。個体間でかなり変化するだけでなく、個体変動に関しては部分的な解釈は可能であるが各個体ごとに経時的に大きく変化する。日ごとの変動と同様に日内変動（サーカディアン変動）に注意すべきである。しかも血圧は測定することが難しい。近年では血圧を連続的に記録できるように小型テープレコーダーに接続した腕の留置カテーテルを用いた新しい技術が発展してきている。この携帯用装置による血圧測定は24時間のデータを示すのでより多くの情報を与えるものであり、また血圧を直接に観察エラーなしに測定できるのでより精密となり得る。多くの人々は、とりわけ間接的（例えば非侵入式）に測定する携帯用の記録計の新しい方法を評価するために、この動脈内カテーテル技術が”まことに優れた標準”であるとみなしている。変動性が言及されたので2つの装置を使って同時に測定値を得ることが重要であり、それ故両腕を使うことが重要である。そこで左右の腕の間に血圧の一貫した差があるかも知れないという疑問が生じる。

この疑問に答えるための研究が Gould らによって詳述された。計画は次の様であった。血圧を測定するために使われた装置は観測者の偏りを除くために設計された「ランダムゼロ」血圧計であった（観測された読みは測定終了時まで解らない他の任意値を自動的に加えられる）。腕帯が両腕に取り付けられ、両方が同じ血圧計に接続された。電動エアポンプが2つの腕帯への圧力を等しくするために使われた。明らかなことであるが、片方の腕に1人つまり2人の観測者が必要であった。特別な血圧計の使用にもかかわらず、観測者間の一貫した差がある場合においては一方の腕だけを1人の観測者が測定しないことが重要であった。それ故、観測者はおのおのの左腕と右腕の観察の2分の1づつを担当すべきであり、観測者が各患者の両腕を測定することが妥当であると感じられた（本質的ではないが）。ちょっとした違いがあるかも知れないが、同様な議論が2つの腕帯についてもあてはまった。それ故、それぞれの腕帯は両腕に同等に用いられるべきなので、両腕の腕帯を交換して再度各患者について測定された。血圧の既知の変動要因を考え、おのおのの観測者が各患者の腕ごとに各腕帯を使って2つの測定値を取り、その結果患者当たり16測定値をとることに決定した。最後に、患者の血圧に関しては一連の測定を通して一貫して変化する傾向があるかも知れない。それ故、腕帯を腕に適用する順序と観測者が左右の腕を測定する順序は無作為化によって変化させた。無作為化の詳しい解説は5.7節で述べられる。観測者は互いに結果の連絡をすることはなかった。

調査は本態性高血圧症の91人の被験者について実施された。上述の計画が、左右の腕の血圧の比較を可能な限り正しくするために適用された。加えて各被験者の

腕の円周が記録され、被験者の測定順の記録がとられた。この研究は、実験の計画の多くの主要な点を例証するが、そのいくつかについてもう少し詳しいことは5.5節で考察したい。

#### 観測者の数

患者ひとりあたり2人の観測者が必要であったが、観測者間の差を定量化できるので2人を必要としない場合でも、複数の観測者を使うことがしばしばよい結果を生む(14.2章参照)。

#### 繰り返し測定

研究対象の結果の評価について、より高い精度が与えられるためには実験条件のそれぞれの組合せでは1回以上読み取ることが望ましい。しかしながら各繰り返し測定は独立してはならない。腕の比較研究において、使用した装置では観測者が前の測定値が何であったかを知らなかった所以他们は独立していた。

#### つりあいのとれた計画

実験ファクターの組合せ毎に同じ回数だけ観測することが不可欠ではないが、上述の実験のように、もし全てのつりあいがとってあれば解析はとてもシンプルである。

#### 無作為化

各患者の両腕に割付ける観測者と腕帯の順序は無作為に決められた。

無作為化は実験計画の大切な要素のひとつである。

#### 共変量

結果に影響を与えたかも知れないものとして、時に実験処置とはかかわりのない主要点(共変量)を記録する必要がある。それは環境の温度のように観察ごとに変動するかも知れないが、年齢のように被験者ごとに変動するかも知れない。この研究では腕帯の接着状態に影響を与えるものとして腕の外周が共変量となる可能性が考慮された。腕の外周は観察ごとの変量ではないが被験者内での変量(すなわち両腕間)であり、上述の2例の間のものである。他の潜在的な共変量は観察の順序であった。記録された血圧は繰り返し測定によって低下するであろうと予想されたことから、設計は無作為化され、つりあいが取られた。しかしながら精度を高めるためには、解析の際に測定順を考慮することも可能である。

#### 標本の大きさ

両腕の差を正確に評価するために多くの標本が用いられた。

## Q & A

### 5. 1 序

Q (大野) : 5. 9 ~ 12 節のところでは聞けば良いのでしょうけれども、患者-対照研究とコホート研究ということについて、分かれば教えて下さい。

A (小西) : 私もよく分からなかったので調べたのですが、「患者-対照研究」は結果に基づいて調査の対象を選んで患者群と対照群を作り、結果に対応する原因因子の曝露に関する情報を集めて解析します。ですから比較的出現頻度の低いものでも陽性症例を集めるのにそれほど大きな集団を観察しなくても良いと言われています。それから結果と原因がある程度わかって調査しますので、期間とお金がかかりません。それに対して「コホート研究」の方は原因因子への曝露の有無ということで対象者を選び、患者群と対照群ということではなく処置群と非処置群、つまり原因に処置されているかいないかということで群を作り、後から得られる結果について情報を集めます。ですから出現頻度の低い事例の場合は、沢山の対象者を調査しなくてはならないのでお金もかかり、また必ず今よりも後のことについての情報を集めて（前向き研究）いきますので時間もかかります。

例にするには実験データですので少し変ですが、発癌実験のデータを例にこの2つの研究の数字上の違いを私の理解した範囲で説明します。マウスの皮膚発癌の実験では、DMBAというものを1回塗布した後にTPAを塗布し続けると約60~90%の動物の皮膚に腫瘍ができてきます（仮に60%とします）が、対照の動物にアセトンで1回塗布してTPAを塗布し続けても、便宜的にゼロとしますがほとんど腫瘍はできません。これを「患者-対照研究」という観点にたつと、最初から結果である皮膚腫瘍を有した60匹のマウスがDMBAを塗布されていたか（60匹中60匹）、および対応する60匹の皮膚腫瘍を有さなかったマウスはDMBAを塗布されていたかどうかということからDMBAと腫瘍発生の間に関係があるかないかを解析します。一方「コホート研究」では例えば100匹のマウスにDMBAを、別の100匹にはアセトンをそれぞれ塗布し、つまりDMBAという原因の有無に対してその結果皮膚腫瘍が出たのが100匹中60匹、出なかったのが100匹中0匹ということでDMBAと腫瘍発生間の関係を解析します。ですから病院の調査例で置き換えますと、私はたばこ会社の人間ですが、肺癌と喫煙の関

係の調査では、患者－対照研究ですとたばこと肺癌の関係がより強く出るといことなる  
と思います。

Q (大野) : 何かコメントがあるようなのでお願いします。

C (橋本) : 「患者－対照研究」も「コホート研究」もやはり疫学的研究の方法として使われ  
ることが普通です。今言われたように、肺癌と喫煙の問題について言えば、肺癌患者と肺  
癌患者でない人の間で喫煙状況を調査するのが「患者－対照研究」で、これに対したばこ  
を吸っている人と吸っていない人を、例えば10年位追いかけてどの位肺癌が出てくるかを  
調査するのが「コホート研究」です。

#### 5. 1 序 ↑5

Q (大野) : 「統計学の医学に対する貢献の最も重要な面は、恐らく研究計画である」とある  
のですが我々の医薬安全性研究会では必ずしも研究計画というところはやっていないよう  
な気がするのですが、日本ではこれが当てはまっているのでしょうか？どなたかコメント  
お願いします。

C (大橋) : 毒性研究の場合はガイドラインがある程度しっかりしていて、実験することは  
ルーティン化しています。但しデータ解析の方は相当難しく、解析方法が整理されていな  
かったのてこういう研究会が出来てきた訳です。それに比べ、臨床研究や疫学研究では、  
個々の問題に応じたルーティン化した方法というのがないのです。まず仮説に応じて計画  
を相当変えなければいけないことに加え規模も全然違うのです。例えば臨床研究一つする  
のに大体50億円以上かかるのですが、5年位かけて研究が終わって見たら何も出なかつ  
たという事も起こり得るのです。その大変良い経験をアメリカは初期の頃にしています。  
それはUGDP (ユニバーシティ グループ ダイアビテス プログラム) という糖尿病  
の比較研究ですが、インシュリンと経口血糖剤についてプラセボを加えて約7千人位を対  
象に1970年位から調査されたのですが、基本的には失敗に終わったのです。失敗の原  
因は経口血糖剤のドーズセッティングを間違ったこと、データ管理体制が不十分だったた  
めに後からデータの信頼性に疑問が出たことです。解析方法についても疑惑が生じ、何年  
もかけて全てを再解析しなおしています。解析の結果、ある経口血糖剤の心筋梗塞での死  
亡率がプラセボより高くなったため関係者は混乱し、また発表の仕方を十分に計画して  
いなかったのて途中で止めたしまったのです。ところがその事実がマスコミに流れてしま

ったので、「今飲んでいる薬は毒ではないか」という話になって開業医をはじめ世界中が大騒ぎになってしまったのです。単に解析だけでなく全体の発表をどこまでするのかというような一連の研究全体の仕組みが考えられずにスタートしたきらいがあって、現在この研究は反面教師とされています。その頃から特にパブリケーションデータも含めて、やはりきちんとしなければいけないということになったのです。1970年代後半から80年代にかけてデータコーディネーションのあり方がNIH等でも相当議論されるようになってきたのですが、そういう事を踏まえているということです。

それから、この著者はイギリス人でしょう。やはりブリティッシュ メディカル リサーチ カウンシルの貢献は大きいですし、なんと言っても「ランダム」を最初に行ったところですからその意識も強いのです。イギリスだと研究計画という点で有名な肺癌の研究に、ビートドールの内科医が20年位追いかけた例があります。たばこの害について年齢とパッケージでどれくらい変化するかというすばらしいデータは、きっちりした研究計画をたてるという態勢になってはじめて出来たのです。

あともっと良い例は広島放射線衛生研究所です。広島には原爆が落ちましたが日本は焦土になって何も出来なかったのです。アメリカはいち早く1949年頃から乗り込んできて放射線衛生研究所を作って現在に到るまで被爆者を追いかけています。当時は解析法がわからなかったけれど、とりあえずきっちりした体制でデータ集めを始めたのです。そして後から解析法が出来てきたのですが、このデータ集めの体制が研究の仕組みを作ることに大変大きく貢献したのです。基本的には現在の放射線被曝量の標準データは広島、長崎から出ているのです。

これらの背景を踏まえて書かれているのです。

### 5. 2. 3 縦断面的なものか横断面的なものか

Q (大野) : 5. 2. 3の下から7行目の「参照範囲」という言葉が引っかかったのですが、大橋先生これはなんですか？

A (大橋) : 正常範囲のことですが、最近は正常範囲というのは流行らないから「参照範囲」というのです。特にクリニカル テクノロジーをしている人は、正常という言葉は曖昧で人によって意味が全然違うので「正常範囲」とは言わずに「参照範囲」と言います。「参照」と言うことは鑑別のためとか、モニタリングのためとか何かを意識しなければいけない訳

です。各々の目的に応じて使うレンジは変えた方が良くはないかと言ったことを強調する意味で「正常」と言う言葉を避ける傾向があるのです。

Q (大野) : なるほど。新しい言葉なのですね。

A (大橋) : いいえ10年位前からありますから、大体定着していると思います。日本ではまだ厚生省のガイドラインの訳の中にも出て来ていますが、そのうち変わると思います。

Q (大野) : それから下から3行目の「各被験者はたった1回しか観察されなくても、云々・・子供の成長グラフを得る目的の研究・・」と言うところの意味が良くわからないのですが。

A (小西) : ネズミの成長曲線を得るのに1匹の動物を何年も何年もかけて観察し続けるのは大変ですから、普通は各週令のネズミを集めてきて各々の大きさ(ねずみの場合は通常は体重)を測って一つの成長グラフを作りますが、これを人間の子供(体重、身長他)でするということだと思います。

Q (大野) : 観察は実は1回しかしていないということですか?

A (小西) : はい、各個体に対しては1回なのですが、それがまるでずっと追ってきたようなグラフを作るということです。

### 5. 3 変動要因

Q (大野) : ↑10行目の「ノイズに対するシグナル比の概念になぞらえるのが良いのである」とありますが「ノイズに対するシグナル比」と言うのはSN比と言われているものですよね。何をどうしているのかがちょっと気になるのですが。

A (小西) : 結果の中にランダムな変動が沢山含まれていると、自分の目的とする対象効果を検出することが出来なくなるので、その部分をノイズとすることだと思いますが、「ノイズに対するシグナル比」を具体的にうまく説明出来ません。

Q (大野) : ここのところのコメントを頂ければと思いますが、どなたかお願いします。

A (松田) : ノイズと言うものは誤差の変動でして、それに対して見たいシグナルと言うのは例えば効き目が平均的に変化している場合は平均の差異ですね。シグナルとしてしっかり見たいものとそれに対する誤差の変動があるわけですから、それと比べて良くわかるようにしたいと言うことです。ですから普通のSN比と言っているのが、この場合にはそのような測定値における誤差の変動と本当の平均的な動きの差異と言うことだと思います。



Q (大野) : その比をとると言うことですか？

A (松田) : 比をとるというのはちょっと語弊があると思いますが、普通の場合検定を行う時には誤差の分散変動に対して真の差異を見いだしますよね。具体的には比をとって話をすることもあってしょうけれども、そればかりとは限りません。比をとる方で話をすると、例えばF検定だとその概念ですから真の変動、見たいものの変動と誤差の変動を比でとっています。F検定が一番わかりやすいと思います。

A (大野) : 一番このアナロジーが効くと言うことですね。

#### 5. 4 実験：両腕の血圧は等しいか？

Q (大野) : 「ランダムゼロ」血圧計とはどういうものなのですか？

A (大橋) : 任意値をかさ上げてあって、読む人は本当の値が分からないようになっている血圧計です。

C (橋本) : 血圧値の場合、測定者は大体これ位ということを知っていますし、測定される側も大体の自分の血圧の見当がつかますね。例えば収縮期で、人で300なんていう数字が出たら間違いだと誰もが思うわけですが、こういうことはある種の偏りを生じるのです。

Q (大野) : どんな偏りですか？

A (橋本) : 意識的か無意識かわからないけれど、大体もっともらしそうな方に値を読んでもうという偏りです。それを避ける一番良い方法は値の意味をわからなくしてしまえば良い、それを使ったのが「ランダムゼロ」ということです。

Q (大野) : 普通の健康診断ではしないのですか？

A (橋本) : 健康診断はその人のためにしているのであって研究ではないのでしません。ただWHOなどで国際共同研究として行われている循環器官系の研究データにする場合にはこれを使っています。

Q (大橋) : 日本にあるのですか？

A (橋本) : 売っているかどうかは知りませんが、日本も共同研究している人は使っていますよ、先生。

Q (大野) : そういうものですか。わかりました。

Q (小西) : その任意値というのも毎回毎回バラバラの値がかさ上げされるのですね。

A (大橋) : データを集める人が後で下駄を引くことによって、初めて真の値がわかるという

機構になっているようです。大規模な疫学研究とかランダムトライではこの装置を使うことがインターナショナルには大体必須になっているようですが、日本の臨床研究では使っていません。各病院にこんな装置を配れないので、病院にある装置を使っています。

とにかく血圧というものは患者さんが前の値に引きずられたりするものなのです。

Q (大野) : それは相当偏りができますか？

A (大橋) : それはありますよ。血圧を測ることを考えたら、看護婦さんの顔だけでも変わってしまうでしょう？

Q (大野) : よーくわかりました。

#### 5. 4 ー 共変量

Q (大野) : 「共変量」のところですが、ノンエクスペリメンタル フューチャーズがコバリエントであるということが突然わからなくなったのですが、「共変量」というものは簡単に言うてどういうものなのかどなたか教えていただけますか？

A (大橋) : 実験的には決められない被験者の男女差とか、身長などは実験の際に結果に影響与える可能性があるのを記録しますね。それが一方に偏ると評価に偏りができるかも知れないので補正の必要があるかも知れないし、それを誤差と見なされると誤差が大きくなるからそれをアジャストすれば結果の効率や検定あるいは推定の効率が上がる可能性があるわけですね。効率を上げることと偏りをなくすためにはこういう「共変量」を利用するのが一般的です。

Q (大野) : その場合共変量は身長等になるのですね。

A (大橋) : 他に性、温度等実験にコントロールされないけれど観察される変数は何でもそうです。

Q (大野) : 実験の主要な観測値ではないけれども、実験に伴って観測されてしまうから共変ということなのですね。

A (大橋) : 投薬する、しないで予後を見る癌の臨床試験の場合では、乳癌ならホルモン感受性はどうかであったか、リンパ節の転移は何個あったか、腫瘍の大きさが何センチあったか、遺伝子の発現はあったかというような結果に大きく影響する可能性がある類のものもあります。けれどもそれを実験がコントロールするわけではないというものなのです。

Q (大野) : よく読むとそう書いてありました。では今日はこれで終了します。

カイ二乗適合度検定（分割表）についての疑問

Q：大林繁夫（グレラン製薬）

A：稲葉太一（神戸大学）

Q：まず、次のようなデータが与えられたとする。

表1 改善度

処置	改善なし	軽度改善	著明改善	総数
被験薬	$f_{11}$	$f_{12}$	$f_{13}$	$f_{1.}$
偽薬	$f_{21}$	$f_{22}$	$f_{23}$	$f_{2.}$
総数	$f_{.1}$	$f_{.2}$	$f_{.3}$	$f_{..}$

このとき、各周辺の総数( $f_{.1}$ ,  $f_{.2}$ ,  $f_{.3}$ ,  $f_{1.}$ ,  $f_{2.}$ ,  $f_{..}$ )を固定して考えると、6個のセルの度数は多項超幾何分布に従うと考えられる。ここで、質問がある。

Q1： $f_{1.}$ ,  $f_{2.}$ ,  $f_{..}$  はあらかじめ与えられるので固定すると考えてよいが、 $f_{11}$ ,  $f_{12}$ ,  $f_{13}$  については結果として決まる数なので固定して考えるのはどうか。固定するならば、動物実験などで偽薬の改善なしが100%と分かっているときなどはむしろ偽薬の度数  $f_{21}$ ,  $f_{22}$ ,  $f_{23}$  の方ではないか。

Q2：もし結果が異なるのなら、どちらの方法が検出力が高いか。

A：この問題について、次の手順で説明することにする。

- 1) データの分布と仮説
- 2) 検定方式の仕組み（なぜ $\chi^2$ 検定がよいのか）
- 3) Q1とQ2に対する解答
- 4) 関連する話題（検出力について）

## 1. データの分布と仮説

### (1) データの分布

以下の5つの場合を考える。

- ① 周辺度数をすべて条件付きにしない。・・・ 6つのポアソン分布
- ②  $f_{..}$  を条件付きにする。・・・ 6項分布
- ③  $f_{1.}$ ,  $f_{2.}$ ,  $f_{..}$  を条件付きにする。・・・ ( $f_{11}, f_{12}, f_{13}$ ), ( $f_{21}, f_{22}, f_{23}$ )  
が各々3項分布
- ④ 周辺度数をすべて条件付きにする。・・・ 多項超幾何分布
- ⑤  $f_{21}$ ,  $f_{22}$ ,  $f_{23}$  を条件付きにする。
  - i)  $f_{1.}$  を条件付きにする。・・・ 3項分布
  - ii)  $f_{1.}$  を条件付きにしない。・・・ 3つのポアソン分布

(注)「条件付きにする」とは「固定する」ことである。

### (2) 仮説

仮説は、上のどの場合も次の仮説になる。

帰無仮説  $H_0$ : 行と列が独立（被験薬は効かない）

対立仮説  $H_1$ : 行と列が独立でない（被験薬は効く）

2. 検定方式の仕組み（なぜ $\chi^2$ 検定がよいのか）

まず、次の2つの検定方式について述べる。

①分散検定（これから、通常の $\chi^2$ 検定が導かれる。）

これは、ポアソン分布において平均と分散が同じことから提案された方式である。

②尤度比検定

これは、帰無仮説、対立仮説での尤度の最大値の比で検定する方式である。

これら2つの検定（分散検定と尤度比検定）はデータ数が大きいとき同じ検定になる。したがって、1の(1)で分類したすべての場合に対して、通常の $\chi^2$ 検定が導かれる。

3. Q1とQ2に対する解答

A1：周辺度数や偽薬の度数を固定してもしなくても、同じ仮説、同じ検定方式が導出される。（データの分布は異なる。）また、偽薬の度数を固定するかどうかは次のように考えればよい。もし偽薬の実験をするならば、通常の $\chi^2$ 検定をすればよい。それに対して、過去のデータ（ヒストリカルコントロール）を用いて偽薬の実験をしないときは、過去のデータから推定された改善なし、軽度改善、著明改善の比率データを用い、3項分布（ $f_{11}, f_{12}, f_{13}$ ）の母比率の検定をすればよい。つまり、過去のデータについての信頼性が、まず評価されるべきであろう。

A2：信頼できる理由があって度数を固定するならば、検出力は高くなる。

4. 関連する話題（検出力について）

Q3：帰無仮説 $H_0$ の下で、改善なしの比率 $p_{.1}$ の推定量として、 $p_{.1.} = f_{.1}/f_{..}$ と  
 $p_{.1.} = f_{21}/f_{2.}$ のどちらがよいか？

A3：帰無仮説 $H_0$ の下では、被験薬のデータも偽薬のデータも同じと考えられるので、偽薬のデータのみを用いた $p_{.1.} = f_{21}/f_{2.}$ よりも、すべてのデータを用いた $p_{.1.} = f_{.1}/f_{..}$ の方が（推定のばらつきが小さいという意味で）望ましい。

Q4：累積 $\chi^2$ 検定と $\chi^2$ 検定の違いは？

A4：まず、表1のデータの場合、列の要因は順序が意味を持つ。つまり、傾向のある対立仮説を考えるべきである。よって、通常の $\chi^2$ 検定よりも、累積 $\chi^2$ 検定（または、最大 $\chi^2$ 検定）を用いるべきである。

Q5：Wilcoxonの順位和検定と累積 $\chi^2$ 検定の違いは？

A5：累積 $\chi^2$ 検定は、いろんな傾向に対してまんべんなく検出する。また、傾向として、直線性のみを考えているときはWilcoxonの順位和検定が望ましい。

最後に、補足として最大 $\chi^2$ 検定にもふれておく。最大 $\chi^2$ 検定は、もちろん傾向のある場合に適応される方法であるが、累積 $\chi^2$ 検定と比べて両端での変化に対して（両端での変化を見つける意味で）検出力が高い。

コメント：臨床の分野では患者がみんな違うので、モデルは考えられないのでデザインベーストに考える。つまり、帰無仮説の下では被験薬も偽薬も単なる背番号であり、いま起こったデータはたまたまカードをきった結果だと考えるのである。この場合、周辺度数を固定するのはきわめて自然で、分布は超幾何分布となる。このようなときでも $\chi^2$ 検定は、望ましいのである。

コメント：Wilcoxonの順位和検定は、対立仮説がロケーションシフトで固定点でカットして分類したような場合、累積 $\chi^2$ 検定より検出力が高いことがある。

0. 仮説検定の例

はじめに、仮説検定の例を2つあげる。

例1. 降圧剤Aは、効くか？

最高血圧の母平均が160 mmHgの高血圧患者6人に降圧剤Aを投与した。投与後の血圧の結果が表1にある。この降圧剤Aは、効くと考えて良いか？（ただし、血圧の測定値は標準偏差で5 mmHgのバラツキがあるとする。）

表1 降圧剤Aの投与後のデータ(単位はmmHg)  
158, 161, 163, 153, 151, 162

(解析)母平均が $\mu$ 、母分散が $\sigma^2$ の正規母集団から、 $n$ 個のデータ $x_1, \dots, x_n$ を取ってきたと考える。ここで、血圧の母平均が下がったかどうかを知りたいから、

帰無仮説  $H_0: \mu = \mu_0$  ( $\mu_0 = 160.0$ )  
対立仮説  $H_1: \mu < \mu_0$

とする。統計量は、

$$u_0 = \frac{x - \mu_0}{\sigma / \sqrt{n}}$$

を用い、有意水準 $\alpha = 0.05$ とする。 $u_0$ が棄却域

$$R: u_0 \leq -u(0.10) = -1.645$$

に落ちれば、帰無仮説 $H_0$ を棄却すればよい。

表1のデータによる平均値 $\bar{x}$ は158.0であるから、統計量 $u_0$ は次のようになる。

$$u_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{158.0 - 160.0}{5.0 / \sqrt{6}} = -0.980$$

よって

$$u_0 = -0.980 > -u(0.10) = -1.645$$

を得るので、帰無仮説 $H_0$ は棄却されない。すなわち、降圧剤Aは効くとは言えない。

例2. 降圧剤Bは、効くか？

最高血圧の母平均が160 mmHgの高血圧患者6人に降圧剤Bを投与した。投与後の血圧の結果が表2にある。この降圧剤Bは、効くと考えて良いか？（ただし、血圧の測定値は標準偏差で5 mmHgのバラツキがあるとする。）

表2 降圧剤Bの投与後のデータ(単位はmmHg)  
153, 156, 158, 148, 146, 157

(解析)例1と同じ状況なので、帰無仮説 $H_0$ 、対立仮説 $H_1$ 、統計量 $u_0$ 、棄却域 $R$ は全く同じでよい。

表2のデータによる平均値 $\bar{x}$ は153.0であるから、統計量 $u_0$ は次のようになる。

$$u_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{153.0 - 160.0}{5.0 / \sqrt{6}} = -3.429$$

よって

$$u_0 = -3.429 < -u(0.10) = -1.645$$

を得るので、帰無仮説 $H_0$ は棄却される。すなわち、降圧剤Bは効くと言える。

1. 検出力とは

検定には、次の二つの誤りがある。

第一種の誤り・・・ $H_0$ が正しいのに、 $H_0$ を棄却する誤り ( $\alpha$ )

第二種の誤り・・・ $H_1$ が正しいのに、 $H_0$ を棄却しない誤り ( $\beta$ )

$H_1$ が正しいとき、正しく $H_0$ を棄却する確率は"検出力"と呼ばれる。つまり、 $1 - \beta$ である。

ここで例1における検出力を考えてみる。

帰無仮説  $H_0: \mu = \mu_0 (\mu_0 = 160.0)$

対立仮説  $H_1: \mu < \mu_0$

において、「対立仮説が正しいときに、それを検出する確率」が検出力である。しかし、対立仮説は複数の値があり得るので、検出力を計算するにはそのうちの一つを固定しなければならない。つまり、

$H_{1'}: \mu = \mu_1 (< \mu_0)$  ( $\mu_1$ は具体的に与える。)

の時の、 $H_0$ を棄却する確率を計算すればよいのである。よって、

$$Pr \{ \bar{x} \leq \mu_0 - u(2\alpha) \frac{\sigma}{\sqrt{n}} \mid \mu = \mu_1 \} = 1 - \beta \quad (= \text{検出力}).$$

例3. 例1の状況で検出力を考えてみる。たとえば、 $\mu_1 = 155.0$ とすると、

$H_{1'}: \mu = 155.0$

である。 $n=6$ ,  $\sigma=5.0$ ,  $\alpha=0.05$ のとき、棄却域Rは

$$R: \bar{x} \leq 160.0 - 1.645 \times 5.0 / \sqrt{6} = 156.64$$

となる。よって、検出力は

$$1 - \beta = Pr \{ \bar{x} \leq 156.64 \mid \mu = 155.0 \} = Pr \{ u_0 \leq \frac{156.64 - 155.0}{5.0 / \sqrt{6}} = 0.803 \} = 0.789 \quad (\text{約} 79\%)$$

とわかる。

例4. たとえば、薬効評価I (1977 佐久間昭著)によれば、対立仮説を固定した後、必要と考える検出力を保証するような検定の標本の大きさは、次の式で与えられる。(片側検定の場合)

$$n \geq \left( \frac{\sigma}{\mu_1 - \mu_0} \right)^2 (u(2\alpha) + u(2\beta))^2$$

よって、例1の状況で検出力が90%以上になるような検定の標本の大きさは、

$$n \geq \left( \frac{5.0}{155.0 - 160.0} \right)^2 (u(0.10) + u(0.20))^2 = 8.57 \rightarrow 9$$

となり、9以上とわかる。

例5. 例3の状況で標本の大きさnを9として検出力を考えてみる。対立仮説

$H_{1'}: \mu = 155.0$ ,  $n=9$ ,  $\sigma=5.0$ ,  $\alpha=0.05$ のとき、棄却域Rは

$$R: \bar{x} \leq 160.0 - 1.645 \times 5.0 / \sqrt{9} = 157.26$$

となる。よって、検出力は

$$1 - \beta = Pr \{ \bar{x} \leq 157.26 \mid \mu = 155.0 \} = Pr \{ u_0 \leq 1.36 \} = 0.913 \quad (\text{約} 91\%)$$

となり、90%以上になっていることが確認できた。

(まとめと若干の補足)

- ① nが6から9へと大きくなると、検出力は高くなる。
- ②  $\mu_1$ が155.0でなくもっと小さくなると、検出力は高くなる。
- ③  $\sigma$ が5.0でなくもっと小さいときも、検出力は高くなる。

2. 非心度、非心分布とは

非心度とは、標本の大きさ  $n$ 、対立仮説を表す  $\mu_1$ 、標準偏差  $\sigma$  の三つをまとめて扱うためのありがたい概念である。今、例1の検定(に用いられた)統計量は、

$u_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$  であった。ここで、対立仮説  $H_1'$  :  $\mu = \mu_1 (< \mu_0)$  が成り立っているとすると、

$$u_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$$

$\downarrow$   $N(0, 1)$                    $\downarrow$   $\lambda$  とおく

である。よって、 $u_1 = \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}}$  とおくと、対立仮説  $H_1'$  の下で  $u_1$  は標準正規分布

$N(0, 1)$  に従う。つまり、 $u_0$  は対立仮説  $H_1'$  の下で、 $N(0, 1)$  から平均がずれている。このずれを、非心度

$$\lambda = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$$

という。また、このような  $u_0$  の分布を非心分布という。

以下で、四つの場合を考える。

- ① [例1の状況] 平均値の検定(分散が既知の場合)  
 $u_0$  の分布は、 $N(\lambda, 1)$  となる。(正規分布のまま!)  
 [つまり、非心正規分布 = 正規分布?!]

- ② 平均値の検定(分散が未知の場合) [例1とほぼ同じ状況]

仮説は、例1と同じとする。検定統計量は、 $t_0 = \frac{\bar{x} - \mu_0}{\sqrt{V/n}}$  である。

ただし、 $V = \sum (x_i - \bar{x})^2 / (n-1)$ 。ここで、 $H_1'$  ( $\mu = \mu_1$ ) の下で

$$t_0 = \frac{1}{\sqrt{V/\sigma^2}} \left\{ \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right\}$$

$\downarrow$   $N(0, 1)$                    $\downarrow$  非心度  $\lambda$

は(自由度  $(n-1)$ 、非心度  $\lambda$  の)非心  $t$  分布に従うという。

(解説) 対立仮説  $H_1'$  ( $\mu = \mu_1$ ) の下で、検定に用いる統計量の分布を考えると必然的に非心  $t$  分布が出てくる。

- ③  $n$  個の正規母集団  $N(\mu_i, \sigma^2)$  (分散  $\sigma^2$  が既知) に対して

帰無仮説  $H_0: \mu_1 = \dots = \mu_n$  ( $\lambda = 0$ )

対立仮説  $H_1: H_0$  でない ( $\lambda = \lambda_1 > 0$ )

を考える。ただし、 $\lambda = \sum (\mu_i - \bar{\mu})^2 / \sigma^2$ 。

$H_0$  の下で  $\chi^2 = S/\sigma^2 = \sum (x_i - \bar{x})^2 / \sigma^2$  は自由度  $(n-1)$  の  $\chi^2$  分布。

$H_1$  の下で  $\chi^2 = S/\sigma^2$  は(自由度  $(n-1)$ 、非心度  $\lambda_1$  の)非心  $\chi^2$  分布。

- ④  $n$  個の正規母集団  $N(\mu_i, \sigma^2)$  (分散  $\sigma^2$  は共通だが未知) に対して

帰無仮説  $H_0: \mu_1 = \dots = \mu_n$  ( $\lambda = 0$ )

対立仮説  $H_1: H_0$  でない ( $\lambda = \lambda_1 > 0$ )

を考える。ただし、 $\lambda = \sum (\mu_i - \bar{\mu})^2 / \sigma^2$ 、 $V_1 = \sum (x_i - \bar{x})^2 / (n-1)$ 、 $V_2$  は自由度  $\nu_2$  の  $\sigma^2$  の不偏分散、 $V_1$  は  $V_2$  と独立とする。

$H_0$  の下で  $F = V_1/V_2$  は自由度  $(n-1, \nu_2)$  の  $F$  分布。

$H_1$  の下で  $F = V_1/V_2$  は(自由度  $(n-1, \nu_2)$ 、非心度  $\lambda_1$  の)非心  $F$  分布。

3. 検出力の計算方法は？（非心度を利用しよう！）

検出力の計算をやろうと思ったとき、まず最初にやる事は興味のある対立仮説を一つ固定する事である。

例4'. 次の仮説検定で90%の検出力を保証するには標本の大きさnをいくつにすればよいか？ 有意水準は5%とし、分散は未知とする。

帰無仮説  $H_0: \mu = \mu_0 (\mu_0 = 160.0)$

対立仮説  $H_1: \mu < \mu_0$  [ $H_1'$ :  $\mu = \mu_1 (< \mu_0)$ ]

ここで、 $\mu_1 = \mu_0 - 1.5\sigma$ （すなわち、 $(\mu_1 - \mu_0) / \sigma = -1.5$ ）のような決め方をする必要がある。

（こう決めずに、単に $\mu_1 = 155.0$ などと決めると、分散 $\sigma^2$ が未知なので破綻する。）

(1) 近似解法

今、分かっている事をまとめておく。

① 分散既知の場合

$$n = \left( \frac{u(2\alpha) + u(2\beta)}{\delta/\sigma} \right)^2 \quad \text{ただし、} \delta = \mu_1 - \mu_0 \text{ とおく。}$$

② 分散未知の場合

$$n \approx \left( \frac{u(2\alpha) + u(2\beta)}{\delta/\sigma} \right)^2 + \frac{u(2\alpha)^2}{2}$$

[参考：統計数値表（日本規格協会 1972）解説p174-5 例5]

(解) 今、 $\delta/\sigma = -1.5$  ゆえ  $n = \left( \frac{1.645 + 1.282}{-1.5} \right)^2 + \frac{1.645^2}{2} = 5.160 \rightarrow 6 \dots$  [答]

(2) 統計数値表を用いた方法（ $\sigma^2$ が未知の場合の厳密な解法）

(解)  $\delta/\sigma = -1.5$  である。今、片側検定で $\alpha = 0.05$  ゆえ、 $\alpha = 0.10$ （両側検定）と考えてよい。一方検出力は、 $1 - \beta = 0.90$ 。

(ア) まず、(1)の近似解法②を用いて目星をつける。（ $n=5$ か6位だろう。）

(イ) とりあえず、 $n=5$  としてみる。すると自由度は、 $\nu = n - 1 = 4$ 。

(ウ) ここで表1 統計数値表(抜粋)より、自由度4の時は $\lambda = 3.600$ 。

(エ) ゆえに、 $n = \left( \frac{\lambda}{\delta/\sigma} \right)^2 = \left( \frac{3.600}{-1.5} \right)^2 = 5.76 \rightarrow 6 > 5$  だから、ダメである。

(オ)  $n=6$  として(イ)~(エ)の真似をすると、 $\nu = 5$ 。数表より、 $\lambda = 3.432$ 、 $n = 5.234 \rightarrow 6 \leq 6$  となりOK。

このようにして、 $n \leq n$ となるまでnを大きくしていけば解を見つけることができる。

表1 統計数値表(抜粋；数表p330)

$\nu \setminus 1 - \beta$	...	.90	...
3		3.928	
4		3.600	
5		3.432	
6		3.330	



4. 1～3にたいするコメント

吉村氏① 例数設計の観点からは、 $(\mu_1 - \mu_0) / \sigma$ を固定する考え方が必要である。  
 $\mu_1$ を固定するだけではnは求まらない。

② ①のように対立仮説を固定したときに、あるnでの実験を行った後に自分の実験の検出力がどのくらいであったかについて、検出力の区間推定ができるはずである。

大橋氏① biostatistics の分野では、少なくとも2標本問題を考える。そのときは、必要な例数は2倍になる。また、2標本問題は、例数が非常に多いことが多く（例えば300例など）大ざっぱな解析でよいのでノンバラで検定する。

② 血圧のような臨床データは、母平均や母分散を考えるようなモデルベースな話は適当でなく、患者の個人差などを考慮したランダムイゼーションでモデルを作るデザインベースな話で議論するのが普通である。

③ 非心t分布を考えるのは、bioavailability（各個人で前値との差を見る実験）のときぐらいである（少数例であるから）。ただし、平均によって変動幅が影響を受けるので簡単ではない（regression to mean）。

5. シミュレーションの実施

検出力は、以上で述べたようなやり方で計算によっていつも求めることができるとは限らない。複雑な状況下ではよくシミュレーションが行われ、検出力の推定がなされる。

(1) 乱数の発生

今回のシミュレーションでは、pc9801の内部乱数（周期は $2^{31}$ ）を用いた。シミュレーションに用いる乱数は、少なくとも使用する個数よりも十分長い周期をもっていることが必要である。（参考：Knuth著「準数値算法／乱数」サイエンス社）

(2) 対立仮説の設定

シミュレーションを試みるときも、検出力を求めるときのように対立仮説を決める必要がある。以下の5種類の仮説（1）は帰無仮説、あとは対立仮説）を考える。

表2 各群の母比率

	$\pi_1$	$\pi_2$	$\pi_3$
1) $H_0$	0.400	0.400	0.400
2) $H_1$	0.200	0.400	0.600
3) $H_1$	0.285	0.285	0.630
4) $H_1$	0.200	0.400	0.400
5) $H_1$	0.200	0.400	0.200

ただし、 $\pi_1$ 、 $\pi_2$ 、 $\pi_3$ はそれぞれ第1群、第2群、第3群の母比率である。

### (3) 3つの手法の比較

傾向を考えにいた手法と傾向を考えない手法の違いを見るために①コックラン・アーミティジの傾向検定（ピンク本 p. 67）、②2×3分割表の $\chi^2$ 検定（ピンク本 p. 78）の比較を行う。また、傾向を考えにいたノンパラメトリックな手法として③ヨンキーの傾向検定（ピンク本 p. 53）を考える。なお、検定の有意水準は5%とする。

### (4) データの作成

(1)で述べた乱数（0から1までの一様乱数）を用いて、各群でその母比率の2項分布に従う2項乱数を作る。（10個の一様乱数のうち何個が $\pi_1$ より小さいかを数えるとできる）

### (5) 統計量の計算

(4)で作ったデータから(3)の3つの手法の統計量を計算する。

### (6) 10000回の反復

(4)、(5)を10000回繰り返して、そのうち何回帰無仮説が棄却されたかを数える。この回数を10000で割って検出力の推定を行う。結果を表3に示す。

表3 シミュレーションの結果

	母 比 率			回 数		
	$\pi_1$	$\pi_2$	$\pi_3$	①	②	③
1) $H_0$	0.400	0.400	0.400	438	412	535
2) $H_1$	0.200	0.400	0.600	4226	3171	6119
3) $H_1$	0.285	0.285	0.630	3243	3192	5025
4) $H_1$	0.200	0.400	0.400	1320	1331	2288
5) $H_1$	0.200	0.400	0.200	261	1597	308

## 6. シミュレーションの結果の解釈

- (1) 帰無仮説 $H_0$ の下では、3つの手法とも10000回の内、500回程度（5%程度）となっている。
- (2) ①と②を比べると2)の対立仮説に対して傾向を考慮にいた①のコックラン・アーミティジの傾向検定の検出力が高い。
- (3) 5)のような対立仮説に対しては、逆に①と③はよくないが②分割表の $\chi^2$ 検定はそこそこ検出している。

## 7. まとめ

もし、対立仮説として考える仮説が必ず傾向があるもののみであるならば、それに応じたコックラン・アーミティジの傾向検定やヨンキーの傾向検定を用いるとよい。また、必ずしも傾向があるとは限らないときは、ふつうの $\chi^2$ 検定を用いるのがよい。このようなやり方の事を、「検出力の高い手法を使う」と表現するのである。

Appendix

I)  $\bar{x}$  の分布は？ 標準化すればどうなる？

命題 A 1.  $x_1, \dots, x_n$  の期待値を各々  $\mu_1, \dots, \mu_n$  とすると、 $\sum x_i$  の期待値は  $\sum \mu_i$  である。また、 $x_1, \dots, x_n$  が互いに独立な場合、各々の分散を  $\sigma_1^2, \dots, \sigma_n^2$  とすると、 $\sum x_i$  の分散は  $\sum \sigma_i^2$  となる。

したがって、 $x_1, \dots, x_n$  が互いに独立に母平均  $\mu$ 、母分散  $\sigma^2$  の正規分布に従うとき、 $\bar{x}$  は母平均  $\mu$ 、母分散  $\sigma^2/n$  の正規分布に従う。よって、これを標準化（平均を引いて、標準偏差で割る）した

$$u = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

は標準正規分布（母平均 0、母分散 1 の正規分布）に従う。もちろん、各  $x_i$  を標準化した

$$u_i = \frac{x_i - \mu}{\sigma} \quad (i=1, \dots, n)$$

は互いに独立な標準正規分布に従う。（つまり、 $E(u_i) = 0, E(u_i^2) = 1$ ）

II) 不偏分散  $V$  はなぜ不偏？

定義 A 2. 互いに独立な標準正規分布  $z_1, \dots, z_n$  の二乗和  $\chi^2 = \sum z_i^2$  はカイ二乗分布に従う。

命題 A 3.  $E(\chi^2_n) = n$  ( $\because E(z_i^2) = 1$  と命題 A 1)

$$\text{命題 A 4. } \sum \left( \frac{x_i - \mu}{\sigma} \right)^2 = \sum \left( \frac{x_i - \bar{x}}{\sigma} \right)^2 + \left( \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right)^2$$

$$\begin{array}{ccc} \downarrow & \downarrow & \downarrow \\ \chi^2(n) & \chi^2(n-1) & \chi^2(1) \end{array}$$

( $\because$  I, 定義 A 2 と定理 A 5)

定理 A 5.  $\sum (x_i - \bar{x})^2$  と  $\bar{x}$  は独立 ( $\because$  略)

したがって、 $V = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{\sigma^2}{n-1} \chi^2(n-1)$  ゆえ、 $E(V) = \sigma^2$ .

III)  $t = \frac{\bar{x} - \mu}{\sqrt{V/n}}$  はなぜ自由度  $(n-1)$  の  $t$  分布？

定義 A 6. 標準正規分布  $z$  と自由度  $(n-1)$  のカイ二乗分布  $\chi^2(n-1)$  が互いに独立なとき、 $t = \frac{z}{\sqrt{\chi^2(n-1)/n-1}}$  は自由度  $(n-1)$  の  $t$  分布に従う。

まず、定理 A 5 より、 $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$  と  $\chi^2(n-1) = (n-1) \frac{V}{\sigma^2}$  は独立になる。よって、

て、 $t = \frac{z}{\sqrt{\chi^2(n-1)/n-1}} = \frac{\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}{\sqrt{V/\sigma^2}} = \frac{\bar{x} - \mu}{\sqrt{V/n}}$  は、定義 A 6 より自由度  $(n-1)$  の  $t$  分布に従う。

# 毒性試験における施設内 ヒストリカルデータを使うための条件

東京理科大学  
工学部 経営工学科  
芳賀敏郎

## § 1 なぜヒストリカルデータか

複数個 ( $k$  個) の新薬をコントロールと比較するために、次のようなデータがとられる。

水準	データ	平均
コントロール	$x_{01}, x_{02}, \dots, x_{0n_0}$	$\bar{x}_0$
新薬 1	$x_{11}, x_{12}, \dots, x_{1n_1}$	$\bar{x}_1$
...	...	...
新薬 $i$	$x_{i1}, x_{i2}, \dots, x_{in_i}$	$\bar{x}_i$
...	...	...
新薬 $k$	$x_{k1}, x_{k2}, \dots, x_{kn_k}$	$\bar{x}_k$

コントロールとして、毒性試験では無処理や溶媒対照が、薬効評価では従来の薬やプラセボが用いられる。このデータから、新薬  $i$  とコントロールの差  $\bar{x}_i - \bar{x}_0$  の有意性を検定する。このとき、一組のコントロールデータを  $k$  個の新薬との比較に用いることは当然のこととされる。このような実験を別の薬に対して再度実施するとき、新薬  $k+1, \dots$

と考え、コントロールについての実験を省略することは通常行なわれない。一回の実験に用いる動物は、できるだけ均質な集団を用いるように努力しているのに対し、実験の時期が異なると、実験動物の均質性が保証されない可能性があるからである。

動物愛護の立場から動物匹数を最少限にしたいという強い要望に応えるなどの理由で、コントロールを省略したり、コントロールの繰返し数  $n_0$  を減少する工夫が必要となる。そのために、従来の実験におけるコントロールデータ (ヒストリカルデータ, 背景データ) を利用することが考えられ、それを利用できる条件を明らかにし、さらに、その条件を満たすためにどのような改善が必要かを明確にすることは極めて有用であると考えられる。

ヒストリカルデータを利用する別の目的は、今回の実験の環境条件が従来の実験と特に変わっていないことを確かめるために、コントロールをヒストリカルデータと比較することがある。すなわち、実験環境を管理するために、コントロールデータの管理図を作成することが役に立つ。<sup>#1</sup>

§ 2 母平均の最良推定.

第  $i$  回目の実験におけるコントロールの第  $j$  番目の観測値を  $x_{ij}$  とする. その期待値と標準偏差を  $\mu_i, \sigma_w$  で表すことにする.

$$\begin{aligned} E[x_{ij}] &= \mu_i & (1) \\ V[x_{ij}] &= \sigma_w^2 \end{aligned}$$

第  $i$  回目の実験のコントロールの期待値  $\mu_i$  の期待値と標準偏差を  $\mu, \sigma_B$  で表す.

$$\begin{aligned} E[\mu_i] &= \mu & (2) \\ V[\mu_i] &= \sigma_B^2 \end{aligned}$$

以上の関係を図 1 に示す.

第  $i$  回目の実験のコントロールの期待値  $\mu_i$  を推定する方法を 3 つ示し, それらの良さを比較する.  $\mu_i$  の推定値  $\hat{\mu}_i$  の添字 (0), (1) の意味は方法 (iii) で明らかになる.

(i) 最も一般的な  $\mu_i$  の推定値としては, 第  $i$  回目の実験におけるコントロールの観測値  $x_{ij}$  の平均

$$\hat{\mu}_i(0) = \bar{x}_i = \sum_{j=1}^n x_{ij} / n \quad (3)$$

が用いられる.  $\hat{\mu}_i(0)$  の期待値と分散は

$$\begin{aligned} E[\hat{\mu}_i(0)] &= \mu_i & (4) \\ V[\hat{\mu}_i(0)] &= \sigma_w^2 / n \end{aligned}$$

となる. すなわち, 推定値の期待値が真値に一致するので偏り (bias) がない. 推定値の良さを測る量として「平均平方誤差, mse (mean square error)」がある. これは, 真値からの平均 2 乗距離であって, 偏りの 2 乗と分散の和で表される.

$$\begin{aligned} mse[\hat{\mu}_i(0)] &= E\{[E[\hat{\mu}_i(0)] - \mu_i]^2\} + V[\hat{\mu}_i(0)] \\ &= \sigma_w^2 / n \end{aligned} \quad (5)$$

(ii) もし, コントロール全体の期待値  $\mu$  がヒストリカルデータから十分な精度で求められているならば, それを  $\mu_i$  の推定値として用いることが考えられる.

$$\hat{\mu}_i(1) = \mu \quad (6)$$

この推定値の期待値, 分散および平均平方誤差は

$$E[\hat{\mu}_i(1)] = \mu \quad (7)$$

$$V[\hat{\mu}_i(1)] = 0$$

$$mse[\hat{\mu}_i(1)] = E\{[\mu - \mu_i]^2\} = \sigma_B^2 \quad (8)$$

となる.

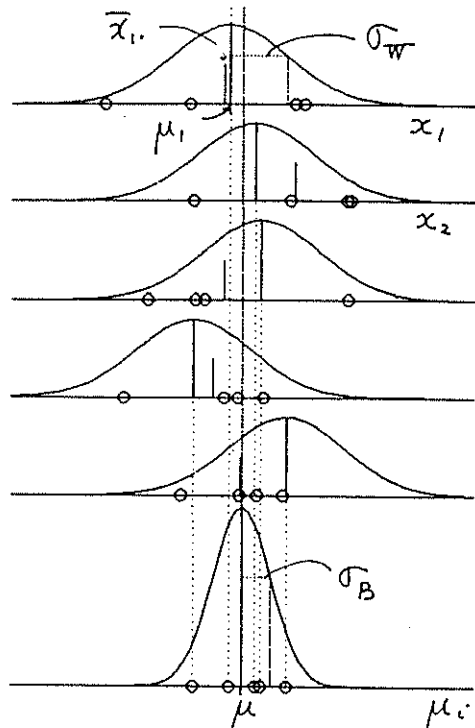


図 1 データ  $x_{ij}$  の構造

二つの推定値  $\hat{\mu}_{i(0)}$ ,  $\hat{\mu}_{i(1)}$  の良さを平均平方誤差で比較すると,  $\sigma_W^2/n$  と  $\sigma_B^2$  の何れが小さいかで優劣が決まる。

(iii) 上に挙げた2つの方法の中間的方法として,  $\bar{x}_i$ ,  $\mu$  の加重平均を用いることが考えられる。

$$\hat{\mu}_{i(a)} = a\mu + b\bar{x}_i, \quad b = 1 - a \quad (9)$$

$a$  は, ヒストリカルデータを利用する割合で, 0から1の間の値をとる。(i), (ii) の推定値  $\hat{\mu}_{i(0)}$ ,  $\hat{\mu}_{i(1)}$  はそれぞれ,  $a=0$ ,  $1$  としたものに相当する。 $\hat{\mu}_{i(a)}$  の期待値, 分散および平均平方誤差は

$$E[\hat{\mu}_{i(a)}] = a(\mu - \mu_i) \quad (10)$$

$$V[\hat{\mu}_{i(a)}] = b^2 \sigma_W^2/n$$

$$mse[\hat{\mu}_{i(a)}] = a^2 \sigma_B^2 + b^2 \sigma_W^2/n \quad (11)$$

となる。

ここで, 二つの標準偏差  $\sigma_B$ ,  $\sigma_W$  の比を

$$f = \sigma_B/\sigma_W \quad (12)$$

とする。 $f$  と  $n$  が与えられているとき,  $mse[\hat{\mu}_{i(a)}]$  を最小にする  $a$  を導くと,

$$a = 1/(1 + nf^2) \quad (13)$$

となり, 推定式は

$$\hat{\mu}_{i(a)} = (\mu + nf^2 \bar{x}_i)/(1 + nf^2) \quad (9')$$

となる。(注: 最尤法からも同じ式が導かれる)。この推定値の平均平方誤差は

$$\begin{aligned} mse[\hat{\mu}_{i(a)}] &= (1-a) \cdot \sigma_W^2/n \\ &= b \cdot \sigma_W^2/n \end{aligned} \quad (14)$$

となる。この式は,  $nf^2$  が小さいほどヒストリカルデータを利用する割合  $a$  が大きくなり, それにつれて平均平方誤差が小さくなることを示している。

横軸に  $\sqrt{nf}$  を縦軸に  $a$  をとったグラフを図2に示す。

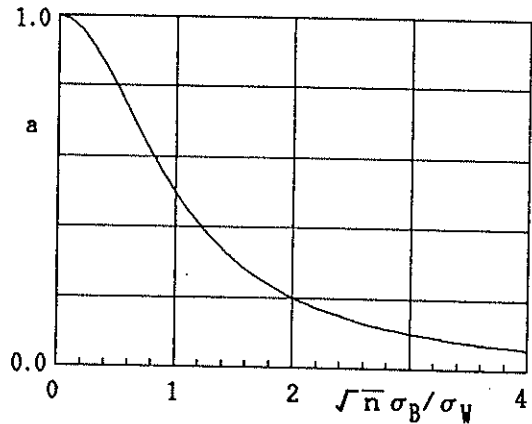


図2  $\sqrt{nf}$  と  $a$  との関係

ヒストリカルデータを利用した  $\hat{\mu}_{i(a)}$  を用いることにより、推定精度を向上させることが分った。そこで、 $n$  個の測定値から求めた  $\hat{\mu}_{i(0)}$  と、 $n^*$  個の測定値から求めた  $\hat{\mu}_{i(a)}$  とが同じ推定精度である条件を求めると

$$\sigma_W^2/n = \{n^* f^2 / (1 + n^* f^2)\} \sigma_W^2 / n^* \quad (15)$$

$$n = 1/f^2 + n^*$$

となり、

$$n^* = n - 1/f^2 \quad (16)$$

の関係が得られる。これは、 $\hat{\mu}_{i(a)}$  を用いることにより、 $1/f^2$  だけ測定値の個数を減らすことが可能であることを示している。この数は  $n$  によらず、 $f$  だけで決まるのは興味深い。また、 $n$  が  $1/f^2$  より小さいときは、 $n=0$  として（コントロールの実験を省略して）方法(ii)の  $\hat{\mu}_{i(1)}$  を用いた方がよい。

測定値の省略可能個数  $1/f^2$  と  $f$  の関係を次に示す。

$f$	.05	.10	.15	.20	.25	.30	.35	.40	.408	~	.447	~	.500	~	.577	~	.707	~	1	~	0
$1/f^2$	400	100	44	25	16	11	8	6	5		4		3		2		1		0		0

これから、 $n$  が小さいときはヒストリカルデータを利用する効果はあるが、 $n$  が大きいとき、その相対メリットは小さくなる。

### 3 $\sigma_B/\sigma_W$ の推定方法

前節に述べたように、ヒストリカルデータを有効に利用できるか否かは  $f = \sigma_B/\sigma_W$  の大ききで決まる。したがって、 $\sigma_B/\sigma_W$  の大ききを正しく把握することが必要である。

次のようなデータ（実験計画法では、「変量模型の1元配置データ」と呼ばれる）から、 $\sigma_B$ 、 $\sigma_W$  を推定することができる。

	データ	平均	総平均
コントロール1	$x_{11}, x_{12}, \dots, x_{1n_1}$	$\bar{x}_1$	
...			
コントロールi	$x_{i1}, x_{i2}, \dots, x_{in_i}$	$\bar{x}_i$	
...			
コントロールk	$x_{k1}, x_{k2}, \dots, x_{kn_k}$	$\bar{x}_k$	$\bar{x}..$

総平方和  $S_T$  を、水準間の平方和  $S_B$  と水準内の平方和  $S_W$  に分解する。同様に、自由度  $\phi_T$  も2つの成分に分解される（解析の詳しい過程は実験計画法の教科書<sup>\*2</sup>に譲る）。

$$S_T = \sum_i \sum_j (x_{ij} - \bar{x}..)^2 \quad \phi_T = N - 1 \quad N = \sum_i n_i$$

$$S_B = \sum_i n_i (\bar{x}_i - \bar{x}..)^2 \quad \phi_B = k - 1 \quad (17)$$

$$S_W = \sum_i \sum_j (x_{ij} - \bar{x}_i.)^2 \quad \phi_W = N - k$$

これから表1の分散分析表が得られる。

表1 分散分析表

要因	平方和S	自由度φ	分散V	分散比F
水準間	$S_B$	$\phi_B$	$V_B = S_B / \phi_B$	$F = V_B / V_W$
水準内	$S_W$	$\phi_W$	$V_W = S_W / \phi_W$	
全体	$S_T$	$\phi_T$		

2つの分散の期待値は、各回の繰返し数 $n_i$ が等しく、 $n$ であるとき

$$E[V_B] = \sigma_W^2 + n\sigma_B^2 \quad (18)$$

$$E[V_W] = \sigma_W^2$$

となり、これから、 $\sigma_W^2$ 、 $\sigma_B^2$ は次式で推定される。

$$\hat{\sigma}_W^2 = V_W \quad (19)$$

$$\hat{\sigma}_B^2 = (V_B - V_W) / n$$

$f = \sigma_B / \sigma_W$ の点推定値は

$$\hat{f} = \sqrt{(F-1)/n} \quad (F > 1) \quad (20)$$

$$= 0 \quad (F \leq 1)$$

となる。

この推定値にはサンプリング誤差が含まれるので、区間推定をする必要がある。信頼率が $1 - \alpha$ の信頼区間は

$$\sqrt{\frac{F}{F(\phi_B, \phi_W; \alpha/2)} - 1} < \sigma_B / \sigma_W < \sqrt{\frac{F}{F(\phi_B, \phi_W; 1 - \alpha/2)} - 1} \quad (21)$$

となる。

ここに、 $F(\phi_B, \phi_W; \alpha/2)$ 、 $F(\phi_B, \phi_W; 1 - \alpha/2)$ は、自由度が $(\phi_B, \phi_W)$ のF分布の上側および下側確率が $\alpha/2$ となる値である。信頼区間の幅は、ヒストリカルデータの組数 $k$ と繰返し数 $n$ によって変わる。信頼区間の幅が広過ぎるときは、 $k$ や $n$ を増やす必要がある。 $n$ を増やすより $k$ を増やす方が区間の幅を狭めるのに効果的である。

繰返し数が異なるときは、 $n$ は、次の式で表される $n_i$ の平均 $\bar{n}$ で置き換えられる。

$$\begin{aligned} \bar{n} &= (\sum n_i - \sum n_i^2 / \sum n_i) / (k - 1) \\ &= \bar{n} (1 - (s_n / \bar{n})^2 / k) \end{aligned} \quad (22)$$

ここに、 $\bar{n}$ 、 $s_n$ は、それぞれ、 $n$ の単純平均と標準偏差である。



4. 実際のデータによる検証

(1)  $\sigma_B/\sigma_W$  が小さい場合 (小核試験データ)

染色体異常誘発性を指標とする変異原性試験の一つに「小核試験」がある<sup>\*3</sup>。この試験の内容の具体的説明は文献3に譲ることにする。同文献には、背景データ (コントロールデータ) の意義やそれを使って実験を管理する方法について詳しく述べられている。それによれば、実験条件を十分管理すれば、背景データはポアソン分布に従い、実験間のバラツキ $\sigma_B$  は十分小さい。同書の著者から生のデータ (3種類) の提供を受けた。データの一部を表2に示す。前節では、データが連続量であることを仮定していたが、このデータは、0, 1, ..., の整数値のみをとる。

表2 データ (一部)

1	0	3	3	1	3
2	1	5	5	2	3
0	1	4	1	3	1
2	1	3	1	4	2
0	4	1	0	3	2
0	3	0	2	4	2
3	2	2	4	2	3
2	0	1	3	1	2
3	2	4	2	3	3
1	2	2	7	0	6
1	2	2	0	1	0
1	4	4	0	1	
2	2	2	2	4	2

最初の種類のデータ (67組, 総観測値の数 383) について、平均値と分散を求めた。分散 (2.166) と平均値 (2.037) の差は小さく、ポアソン分布に従っていると考えられる。観測値の度数分布についてカイ2乗検定を適用しても、 $\chi^2 = 3.66$  で有意とならない。(2番目と3番目のデータについての $\chi^2$  は、それぞれ、1.53, 3.16であった)。

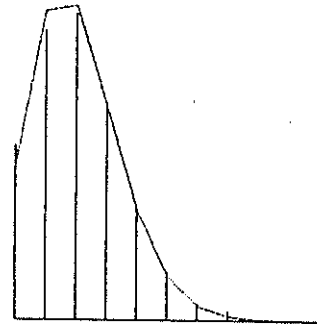


図3 ポアソン分布の当てはめ

3種類のデータについて、 $f = \sigma_B/\sigma_W$  を推定した結果を表3に示す。前節に説明した方法に加え、 $\sigma_W^2$  の推定値として、 $V_W$  の代わりに平均 $\bar{x}..$ を用いて計算した結果を ( ) 内に示す。 $f = \sigma_B/\sigma_W$  は0.2~0.4前後であって、 $n=6$ に対して $n^{\ddagger}=0$ にすることが可能である。

表3

データ種類番号	I	II	III
データの組数	67	32	15
観測値の総数	383	133	59
平均個数 $\bar{n}$	5.72	4.15	3.92
群間分散 $V_B$	2.491	2.341	4.503
群内分散 $V_W$ (平均)	2.098(2.037)	1.294(1.504)	3.375(4.356)
分散比 $V_B/V_W$	1.187(1.223)	1.810(1.557)	1.334(1.034)
f の点推定値	0.181(0.197)	0.442(0.366)	0.292(0.093)
f の信頼区間	0.000 ~ 0.368	0.122 ~ 0.757	0.000 ~ 0.809

(2)  $\sigma_B/\sigma_W$  が比較的大きい場合 (犬の血液検査データ)

A社からビーブル犬の投与開始前, 5~6ヶ月齢の血清の生化学検査データの提供を受けた。雌犬のデータについて,  $k, \tilde{n}, \sigma_B/\sigma_W$  の点推定値と区間推定値を, 雄犬のデータについては点推定値のみを表4に示す。ひずみ (skewness) が1以上の項目については自然対数変換後についても計算してある。

表4 血清の生化学データについての解析結果

項目名	組数	平均 繰返し数	$\sigma_B/\sigma_W$			雄の 点推定値
			下限	点推定値	上限	
ALX	10	14.52	0.425	0.695	1.340	0.886
ALPHA1	10	14.52	0.477	0.765	1.465	0.529
ALPHA1 (ln)	10	14.52	0.513	0.815	1.554	
ALPHA2	10	14.52	0.462	0.744	1.427	0.678
ALPHA2 (ln)	10	14.52	0.497	0.792	1.512	
GAMMA	10	14.52	0.562	0.883	1.674	
BETA	10	14.52	0.000	0.001	0.405	0.747
AG	11	14.31	0.584	0.899	1.643	1.049
ALB	11	14.31	0.570	0.879	1.608	1.120
TP	17	12.40	0.838	1.177	1.840	1.070
GOT	17	12.40	0.334	0.528	0.877	0.767
GPT	17	12.40	0.256	0.439	0.753	0.712
ALP	17	12.40	0.407	0.616	1.004	0.863
ALP (ln)	17	12.40	0.475	0.702	1.128	
LDH	17	12.40	0.632	0.905	1.430	0.780
LDH (ln)	17	12.40	0.646	0.922	1.456	0.600
CH-E	8	14.11	0.212	0.456	1.050	
BUN	17	12.35	0.404	0.613	1.000	0.775
CREA	17	12.40	0.823	1.157	1.809	0.821
T-CHOL	17	12.40	0.543	0.788	1.256	0.549
TG	8	15.73	0.409	0.700	1.503	0.693
TG (ln)	8	15.73	0.478	0.797	1.694	0.757
GLUCO	17	12.40	0.856	1.201	1.875	1.104
CA	17	12.40	0.987	1.376	2.141	1.497
CL	16	13.35	1.738	2.410	3.781	2.064
K	16	12.95	0.606	0.876	1.409	0.715
NA	16	12.95	1.173	1.639	2.584	1.209
P	17	12.34	0.645	0.921	1.455	1.116

図4は $\sigma_B/\sigma_W$ が小さい項目 (GPT) と大きい項目 (Na) について, 箱ひげ図を示したものである。GPTはひずみが0.73で, それ程大きな値ではないが, 対数変換後の箱ひげ図も示す。

fの値が大きい項目は, 箱ひげ図から見ても, ヒストリカルデータを利用する効果は少ないと考えられる。

图4

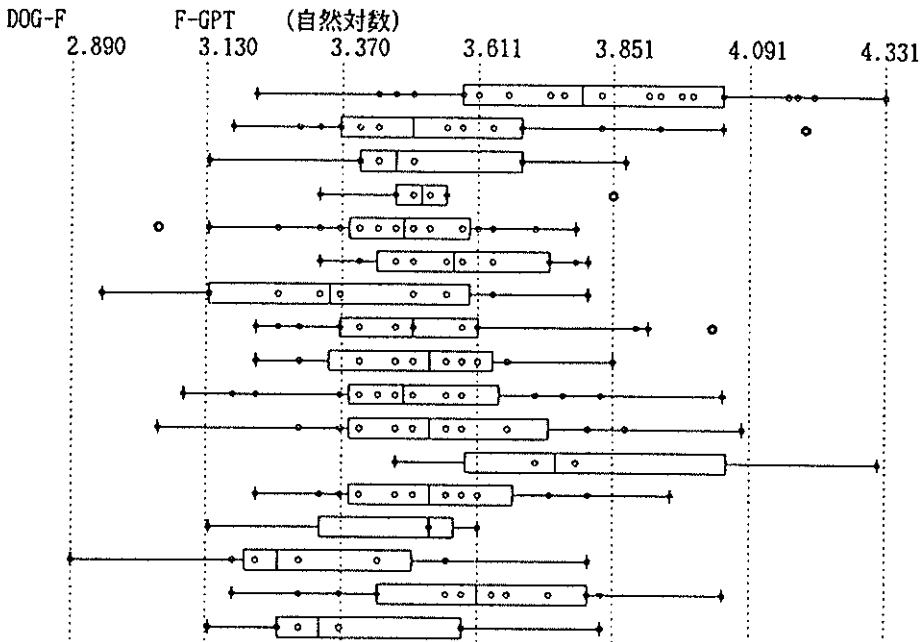
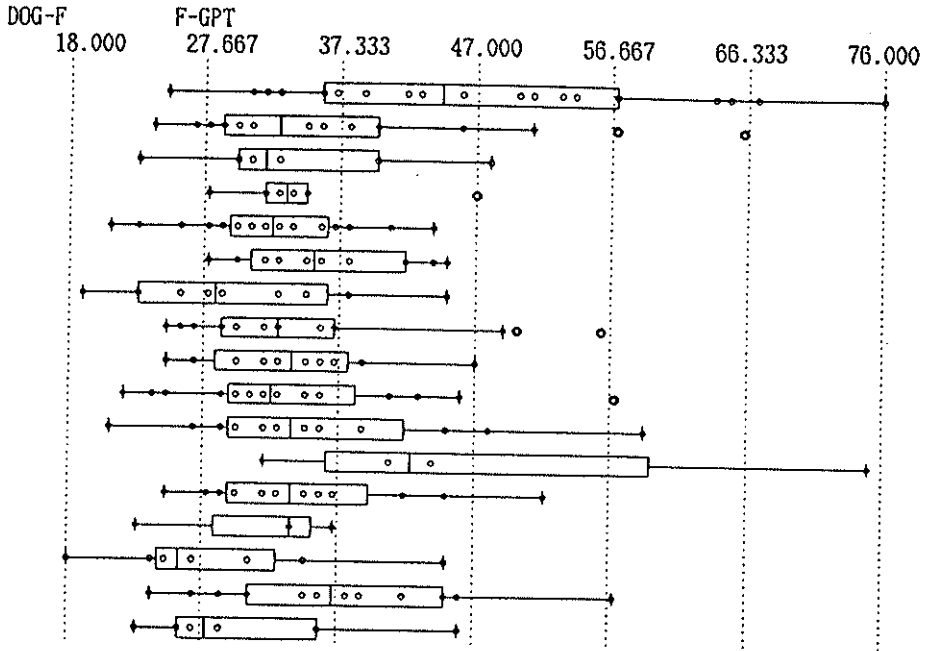
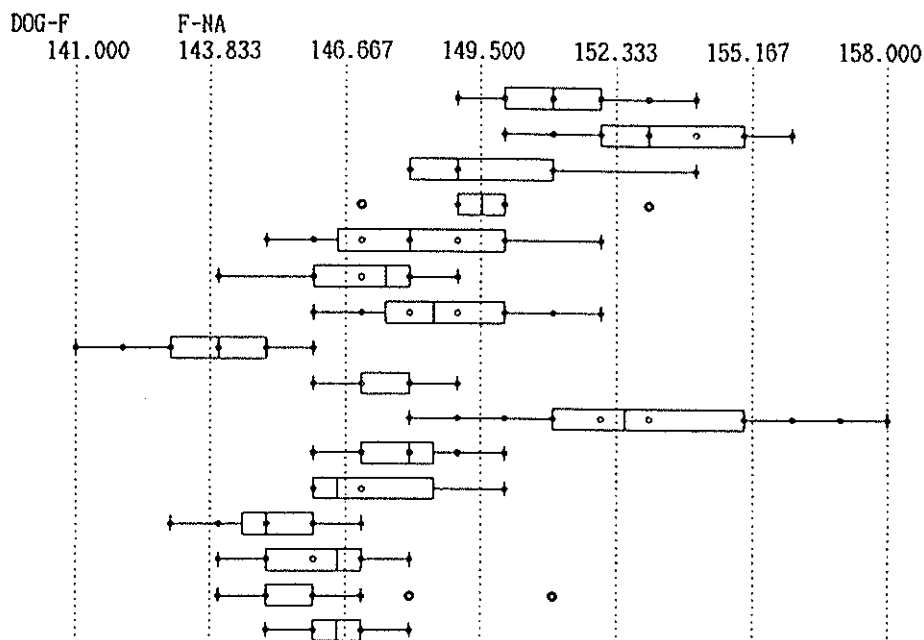


図4 (続き)



B社からも同様の血清生化学検査および血液検査データの提供を受けて解析した。血液データについての結果を表5に示す。

組数と、平均繰り返し数 $\bar{n}$ は全項目共通で、それぞれ、雄は11組、9.57で、雌は12組、8.32、である。fの95%信頼区間は省略する。

項目によって分散の比が大きく変化している。雄と雌の差が幾分ある項目もあるが、傾向はおおむね一致している。

項目によってはヒストリカルデータを利用することができるであろう。

表5 血液データのf

項目	雄	雌
RBC	0.56	0.58
HCT	0.75	1.07
HB	0.82	0.83
MCH	0.81	1.02
MCHC	1.16	2.49
MCV	0.75	1.87
PLT	0.33	0.20
WBC	0.47	0.41

#### 5. 今後の課題.

今回、3つの施設から生データの提供を受け、解析した結果を報告した。これは、研究の入口であって、この方法の妥当性について、数理統計の専門家に検証して頂くと共に、日常このようなデータに接しておられる専門家のご意見を受けて、方法を確立するための努力が必要である。

残された課題について、思い付くままに羅列する。

○ コントロールとして、提案する推定値  $\hat{\mu}_{i(a)}$  を利用する場合、自由度をどう扱ったらよいか。

○ 誤差分散の推定にもヒストリカルデータが使えるか？検定の自由度を稼ぐことができるかもしれない。

○ この報告は、 $\mu$ が既知である（正確に分っている）としているが、現実には $\mu$ の推定値にも誤差が含まれる。ここで提案する  $\hat{\mu}_{i(a)}$  に、 $\mu$ の推定値の誤差の考慮を加えるとどうなるか。 $\mu$ を推定するためには、何組位のヒストリカルデータが必要か。

○  $f$ の推定値は、前節に示した95%信頼区間の幅からも分るように、推定誤差が大きい。 $f$ を十分な精度で推定するためには、どの位の組数 $k$ と測定回数 $n$ が必要か。

○ 現実のデータは外れ値を含むことが多い。ここに提案する方法を適用するにあたって、外れ値をどのように取り扱うか。

○ 多種多様のデータについて解析し、ヒストリカルデータの利用が有効な分野を探索する。

○ 試験動物の均質化、飼育条件・検査方法の標準化、などについて、研究を進めることにより、 $\sigma_B$ を小さくする方法を見出す。 $\sigma_B$ が小さくなれば、ヒストリカルデータの利用範囲を拡大する。

○ ヒストリカルデータに影響を与えていると思われる要因（たとえば、動物の月齢、ブリーダー、餌の種類、投与経路（静注、経口など）、採血から測定までの時間など）をコントロールできないならば、それらを説明変数とする重回帰式を求めて、それら要因の影響を調整する。回帰残差の標準偏差が $\sigma_B$ となり、適切な要因を見出し、重相関係数が大きい重回帰式を求めることができれば、調整前の標準偏差に比べて小さくなる。

## 文献

1. M.Hayashi, I.Yoshimura, T.Sofuni & M.Ishidate, Jr. (1989), "A procedure for data analysis of the rodent micronucleus test involving a historical control", *Environmental and Molecular Mutagenesis*, Vol.13, pp.347-356.
2. 森口繁一編 (1989), 「新編 統計的方法 改訂版, 第8章」, 日本規格協会.
3. 林 真 (1991), 「医薬安全性研究会モノグラフシリーズ No. 2, 小核試験—実験法からデータの評価まで」, サイエンティスト社.
4. 吉村 功 (1989), "背景対照群を含める動物実験データの解析法", 応用統計学会 第 回研究発表会.

# ゲッ歯類を用いた毒性試験から得られる定量データに使用される統計処理法の動向

(財) 食品農医薬品安全性評価センター  
資料評価情報部

小林克己 大橋信之 大河原一弘

要約：農薬、医薬品および新規化学物質等に対する毒性試験から得られる定量データに対する統計処理をt-検定（スチュウデントt検定、アスピルウェルチt検定等）、決定樹方式多重比較検定（等分散検定、分散検定、パラメトリックおよびノンパラメトリック多重比較検定を含む）、または他の検定法のいずれを使用しているのか受託機関の一つである

(財) 食品農医薬品安全性評価センターおよび内外の五雑誌について調査した。t-検定は国内1雑誌のみで使用され、他の3雑誌および安評センターでは、t-検定に代って決定樹方式多重比較検定が近年大半を占めてきている。毒性試験の専門誌であるNTP 報告ではダネットの多重比較検定に代ってノンパラの多重比較検定法を使用していることに注視したい。t-検定やパラメトリック多重比較検定に比べて、ノンパラメトリックの順位検定は平均値の差の検定ではなく平均順位の差の検定の為、飛び離れた変量等が含まれると有意の判定に難解な場合ある。しかしながら、今後3群以上を設定した毒性試験に対しては、決定樹方式多重比較検定の使用が益々多くなるう。

## 1. はじめに

農薬、医薬品および新規化学物質等の開発に必須の試験として、実験動物を用いた毒性（安全性）試験が含まれている。これら毒性試験に使用される統計手法は各ガイドライン<sup>1-3)</sup>に明確化されてはならず、試験責任者が何らかの方法を選択している。一般的にゲッ歯類を用いた毒性試験では対照群、低用量群、中用量群および高用量群の四群が設定され、2-4週間の短期間の反復投与試験から2年間の癌原性試験を含めた長期の反復投与の試験までである。これら毒性試験から得られる定量データ（体重、飼料摂取量、飼料効率、血液・生化学検査値、尿検査値および臓器重量検査値）の解析には、大別してt-検定および決定樹方式多重比較検定<sup>4-9)</sup>が使用されている。毒性試験では、薬剤の投薬の影響によって投与群が対照群に対して各データの分散および平均値が著しく変動したり、動物数も大きく変化することがある。従ってこれら変動要因を十分に考察へ反映できる統計手法が望まれる。

本報告では、毒性試験に使用されている統計手法について農薬、医薬品および新規化学物質の申請窓口別に分け、受託機関である(財)食品農医薬品安全性評価センター（以下安評センター）の統計手法の動向を12年間に渡り調査し、さらに、国内で発行されている3種類の雑誌および米国で発行されている2種類の雑誌に記載されている同動向を調査した。しかし、我が国の農薬の毒性試験報告は要約のみの公表であり、従って統計手法については不明であった。

## 2. 調査材料および方法

### 1) 安評センターの動向

当センターで試験計画書を締結したものについて農薬、医薬品および新規化学物質に分類し、1980から1992年の12年間の動向をt-検定および決定樹方式多重比較検定に分けて調査した。調査対象試験は、2週間から2年間の対照群を含めて3群以上（ほとんど4群）を用いたゲッ歯類の反復投与試験および生殖・発生毒性試験・催奇形性試験について調査した。

### 2) 投稿雑誌掲載論文

下記の国内および米国発行誌に掲載された論文中でt-検定または決定樹方式多重比較検定を年代別に調査した。該当試験は、1)と同様の試験に加え、対照群を含めて3群以上を用いた薬剤の影響試験についても調査した。NTP誌以外の年代別分類は原稿受諾年によった。

#### a. 国内発行誌

- 日本癌学会誌、Jpn. J. Cancer Res.、1988-1992年
- 医薬品研究、IYAKUHIN KENKYU、1988-1992年
- 応用薬理、PHARMACOMETRICS、1988-1992年

#### b. 米国発行誌

- Toxicology and Applied Pharmacology, 1987-1992年 (Vol. 17, No. 1 まで)、本誌は農薬・医薬品等の薬剤全般に対する毒性試験、癌原性試験、薬理試験から *in vitro* まで広い分野に渡って掲載されており、供試動物もゲッ歯類、犬、魚類、鳥類と多種に渡っている。
- NATIONAL TOXICOLOGY PROGRAM (NTP), PUBLIC HEALTH SERVICE / National Institutes of Health, 1987-1992 (一部欠損ナンバーあり、1992年はTR329号の一冊のみ)、既存化学物質に対する毒性試験報告の専門誌で、この組織は米国内ラボで実施した急性、4週間、13週間の予備試験および2年間の慢性毒性試験/癌原性試験等に加えて *in vitro* の試験までが一冊となって薬剤毎に報告書化し各国へ配布されている。調査は発行年によった。

## 3. 調査結果

### 1) 安評センターの動向

当センターで1980-1992年間に試験計画書の締結した内、t-検定および決定樹方式多重比較検定に分けて調査した結果を表-1に示した。

農薬に対する毒性試験では、1980-1986年の間ほとんどt-検定が使用されてきたが、1987年から決定樹方式多重比較検定が使用され始め、その後少しずつ増加し1991および1992年の使用率は100 および67% となった。

医薬品では1982年を除き1981-1985年の間t-検定の使用率が100%で、その後t-検定が暫時減少し、その反面1986-1992年の間決定樹方式多重比較検定の使用率が増加し始め、1992年の使用率は57%となった。

新規化学物質について通産・厚生両者によって作成された毒性試験のガイドラインは1986年に公表され、当センターでは1986年からそれに沿った試験が行われた。表中ではその他で記載されているが、その中には食品添加物等の試験も若干含まれる。1986および19887年の間のt-検定の使用率は100%、1989年のt-検定の使用率は57%で、その後暫時減少し1992年の使用率は8%となった。

これら農薬、医薬品および新規化学物質等全体の動向は、1980-1983年の間t-検定の使用率が80-100%とほとんどを占めていたが、その後減少し始め1992年はその使用率が31%にまで減少し、代って決定樹方式多重比較検定が7割となった。

## 2) 国内発行誌の動向

国内発行誌の動向は、t-検定および決定樹方式多重比較検定に分けて調査した結果を表-2に示した。

日本癌学会誌は1988-1992年の間ほとんどt-検定が使用されている。

医薬品研究誌は1988-1990年の間t-検定の使用率が33-78%であったが、その後の2年間は決定樹方式多重比較検定の使用率が100%となった。

応用薬理誌は1988-1990年の間t-検定の使用率が72-80%であったが、その後減少して1991および1992年はt-検定の使用率が48および50%となった。

## 3) 米国発行誌の動向

米国発行誌の動向は表-3に示した。

Toxicology and Applied Pharmacology誌はt-検定の使用率が1987年に42%で、その後減少し続け1981年の使用率が17%、1992年では使用されなかった。その半面決定樹方式多重比較検定の使用率は1987年の58%から年々増加し、1991年の使用率が83%、1992年が100%となった。NATIONAL TOXICOLOGY PROGRAM (NTP) の報告では1987-1989年に渡り、主に体重と一部の臓器重量の値に対してほとんどダンネットの多重比較検定<sup>10)</sup>が用いられており、その後減少し1990年の使用率は約60%で1991および1992年では使用されなくなった。1989年から1992年に渡りDunn<sup>11)</sup>およびShirley<sup>12)</sup>のノンパラメトリック多重比較検定法で対照群と各投薬群間の差を吟味している。これら検定法の使用率は1989年が12%、1990年が43%で1991年以降は100%となった。NTPの報告は、他の雑誌に見られるt-検定を一切使



用しないで分散分析を含めたダンネットの多重比較検定を使用してきたが、近年DunnおよびShirleyのノンパラメトリック多重比較検定法で対照群と各投薬群間の差を吟味していることに注視したい。

### 3. 考察

長期毒性試験から得られる定量データの種類と検定対象を表-4に示した。これら多項目で大量の検定回数を速やかに処理する為に、Shayne and Carrol(1986)によって決定樹方式による自動判別法<sup>4)</sup>が提唱された。すなわち2群の設定時は、分散比の検定後、分布の状況と標本数の差異によって3法のt-検定を用いる。3群以上の多群を設定した場合は、パートレットの等分散検定による分散の均一性の検定により、分散が均一な場合は分散分析を実施する。群間に有意差が認められた場合、各群の標本数が同一であるならばダンネットの多重比較検定法により、同一でなければダンカンの多重範囲検定法によって群間の平均値の差を比較する。分散が均一でない場合はクラスカル・ワリスの順位検定で群間に有意差を認めた場合、ノンパラメトリックの多重比較検定を実施する(図-1)。

一般的に薬物の各種影響試験および特に毒性試験は通常3群以上を設定する。しかし、これら試験報告によってはt-検定の使用も多く見られることから多重比較検定を含め、どのような方法でこれらに対処しているのか調査した結果、農薬、医薬品および新規化学物質等の毒性試験を受託している安評センターの12年間の動向としては、1980-1988年においてはt-検定を多く使用し、その後1991-1992年の2年間ではt-検定が減少して44%に対して決定樹方式多重比較検定が66%と増加した。使用被験物質の分類別に見ると、1992年は決定樹方式多重比較検定の使用率が化審法の反復投与試験で90%以上で、次ぎに農薬の67%、医薬の57%と半数以上を占めている。このことから、今後は決定樹方式多重比較検定の使用の増加が示唆される。

近年の国内発行誌は、日本癌学会誌がt-検定を多用している。逆に医薬品研究はここ2年間で全て決定樹方式多重比較検定を使用している。応用薬理誌ではt-検定と決定樹方式多重比較検定の使用率が同率で、今後日本癌学会誌を除き決定樹方式多重比較検定の使用が大半を占めると考えられる。

Toxicology and Applied Pharmacology誌では、近年ほとんど決定樹方式多重比較検定の使用が主流となっている。しかし、本誌の中には下記の検定法の使用が散見される。

①Newman-Keulsの多重範囲検定法。

②分散分析法で有意差を示した場合は、t-検定で2群間の比較を行なう。

③パートレットの等分散検定後、均一性を示した場合はt-検定で2群間の比較を行なう。

また、均一性を示さない場合は順位検定あるいはSnedecor and CochranのQ検定。

NTPの報告では、1990年以前が全てダンネットの多重比較検定法を使用してきたが1991年からノンパラメトリックの多重検定が使用されている。これら定量データを始めからノン

パラメトリックの順位検定で吟味することの理由には、長期毒性試験から得られる定量データの一部は標本数、分散および平均値等が群間で著しく変化する為、一般的に決定樹方式で吟味しているがこれらの統計手法流れの繁雑性を避けること又、群間の一様性がある場合でもパラメトリック検定とノンパラメトリック検定を比べた場合その検出力に極めて大きい差がない為と思われる。さらに癌原性試験では最終的に病理組織学所見が重要視されることも理由となろう。小林<sup>13)</sup>によると2年間の長期毒性試験では、定量データの約30%がパートレットの等分散検定で有意差を示しノンパラメトリックの順位検定で吟味しなくてはならないと報告している。

長期の毒性試験から得られる定量データの内、特に52週齢以降のデータには加齢による個体差の増強が加わり、また薬剤の影響に群間個体数の差異が生じ(表-5)時には著しく外れた測定値も見られることから、各群間の分散が著しく異なる。

まれに決定樹方式多重比較検定で実施した為に2群間の平均値が同一であるが有意差を示す場合がある(表-6)。この理由は投薬群に飛び離れた変量があることからパートレットの等分散検定では一様性を示めさず順位検定に進んだことが一因となった為である。

製薬協のアンケート調査<sup>14)</sup>による試験施設間の調査結果によると近年t-検定が22-35施設から3-5施設に激減し、逆に決定樹方式多重比較検定が7-9施設から23-26施設に増加したこと、またt-検定は生物学的に問題とすべきでないような差をも有意と判定することがあり成績の評価を混乱させることもあると報告している。

今後、これら毒性試験の統計処理法は決定樹方式多重比較検定の使用が多くなる傾向にある。しかしながら、t-検定およびノンパラメトリックの順位検定の応用の利点も考えたい。

#### 4. 参考文献

- 1) 厚生省薬務局審査課：医薬品の製造（輸入）承認に必要な毒性試験のガイドラインについて、厚生省（薬審第24号）、1989。
- 2) 農林水産省農蚕園芸局物防疫課：農薬の登録申請に係る毒性試験成績取り扱いについて、59農蚕第4200号、1985。
- 3) 通商産業省：新規化学物質に係わる試験の方法について、環保業第700号、薬発第1039号、61基局第1014号、1986年。
- 4) Shayne Gad and Carrol S. Weil: Statistics and Experimental Design for Toxicologists. New Jersey: The Telford Press, INC., 1986.
- 5) 吉村 功（編）：毒性・薬効データの統計解析。サイエンティスト社、1987。
- 6) Colquhoun, D.: Lectures on Biostatistics. Oxford University Press, 1971: Iyaku journal Co. Ltd., Tokyo.
- 7) 佐久間 昭：薬効評価—計画と評価— I および II、東京大学出版会、1981。
- 8) 吉田 実：畜産を中心とする実験計画法、養賢堂、東京、1980。
- 9) 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、日本製薬工業協会、1984。
- 10) Dunnett, C. W.: A multiple comparison procedure for comparing several treatments with a control, J. Am. Stat. Assoc. 50, 1096-1122, 1955.
- 11) Dunn, O. J.: Multiple comparisons using rank sums, Technometrics 6, 241-252, 1964.
- 12) Shirley, E.: A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment, Biometrics 33, 386-389, 1977.
- 13) 小林 克己：毒性試験に用いる統計処理法の模索、医薬安全性研究会、No. 28、9-12。1989。
- 14) 日本製薬工業会医薬品評価委員会、基礎部会第5分科会：生殖・発生毒性試験における統計処理及びデータの取扱いについて、医薬品研究、23 (5) 656-663、1992。

---

財団法人 食品農医薬品安全性評価センター

〒437-12 静岡県磐田郡福田町塩新田字荒浜582-2、

電話 0538-58-1266、ファックス 0538-58-1393。

Table 1. Changes in Rodents Toxicological Study Analyzing t-tests or Multiple Comparison/  
Range Test Designed Three Groups or More in the An-Pyo Center

Year*	Percentages (Numbers of study)													
	Pharmaceutical				Pesticides				Others				Total	
	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT
1980	-Z(0)	-Z(0)	100Z(15)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	100Z(15)	0Z(0)	100Z(15)	0Z(0)	
1981	100Z(15)	0Z(0)	100Z(18)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	100Z(33)	0Z(0)	100Z(33)	0Z(0)	
1982	50Z(4)	50Z(4)	100Z(12)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	80Z(16)	20Z(4)	20Z(4)	20Z(4)	
1983	100Z(12)	0Z(0)	100Z(9)	0Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	100Z(21)	0Z(0)	100Z(21)	0Z(0)	
1984	100Z(7)	0Z(0)	86Z(12)	14Z(2)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	90Z(19)	10Z(2)	10Z(2)	10Z(2)	
1985	100Z(24)	0Z(0)	71Z(5)	29Z(2)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	94Z(29)	6Z(2)	6Z(2)	6Z(2)	
1986	72Z(13)	28Z(5)	100Z(7)	0Z(0)	100Z(1)	100Z(1)	100Z(1)	0Z(0)	0Z(0)	77Z(20)	23Z(6)	23Z(6)	23Z(6)	
1987	79Z(11)	21Z(3)	50Z(1)	50Z(1)	100Z(7)	0Z(0)	100Z(7)	0Z(0)	0Z(0)	83Z(19)	17Z(4)	17Z(4)	17Z(4)	
1988	95Z(19)	5Z(1)	60Z(9)	40Z(6)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	-Z(0)	80Z(28)	20Z(7)	20Z(7)	20Z(7)	
1989	93Z(13)	7Z(1)	64Z(9)	36Z(5)	57Z(4)	43Z(3)	57Z(4)	43Z(3)	43Z(3)	74Z(26)	26Z(9)	26Z(9)	26Z(9)	
1990	85Z(11)	15Z(2)	29Z(2)	71Z(5)	50Z(1)	50Z(1)	50Z(1)	50Z(1)	50Z(1)	64Z(14)	36Z(8)	36Z(8)	36Z(8)	
1991	85Z(22)	15Z(4)	0Z(0)	100Z(9)	22Z(2)	78Z(7)	22Z(2)	78Z(7)	78Z(7)	55Z(24)	45Z(20)	45Z(20)	45Z(20)	
1992	43Z(9)	57Z(12)	33Z(1)	67Z(2)	8Z(1)	92Z(11)	8Z(1)	92Z(11)	92Z(11)	31Z(11)	69Z(25)	69Z(25)	69Z(25)	

\* Year the study protocol concluded.  
MC/RT: Multiple comparison/range test.

Table 2. Changes in Rodents Toxicological Study Analyzing t-tests or Multiple Comparison/Range Test Designed Three Groups or More in the published Japanese Journals

Year*	Percentages (Nos. of study)					
	Jpn.J. Cancer Res.		IYAKUJIN KENKYU		PHARMACOMETRICS	
	t-tests	MC/RT	t-tests	MC/RT	t-tests	MC/RT
1988	100%( 2)	0%( 0)	57%( 8)	43%( 6)	78%( 7)	22%( 2)
1989	100%( 6)	0%( 0)	33%( 2)	66%( 4)	80%(20)	20%( 5)
1990	90%( 9)	10%( 1)	78%( 7)	22%( 2)	72%(18)	28%( 6)
1991	91%(10)	9%( 1)	0%( 0)	100%( 7)	48%(15)	52%(16)
1992	100%( 5)	0%( 0)	0%( 0)	100%( 1)	50%( 5)	50%( 5)

\* Year the manuscripts received.

MC/RT: Multiple comparison/range test.

Table 3. Changes in Rodents Toxicological Study Analyzing t-tests, Multiple Comparison/Range Test or Other Methods Designed Three Groups or More in the published American Journals

Year*	Percentages (Nos. of study)			
	Toxicol. Appl. Pharmacol		NTP	
	t-tests	MC/RT	Dunnett <sup>1)</sup>	Nonparametric <sup>2)</sup>
1987	42%(21)	58%(29)	100%( 5)	0%( 0)
1988	24%(12)	76%(39)	100%(10)	0%( 0)
1989	27%( 9)	73%(24)	88%( 7)	12%( 1)
1990	23%(10)	76%(34)	57%( 4)	43%( 3)
1991	17%( 9)	83%(43)	0%( 0)	100%( 5)
1992	0%( 0)	100%(17)	0%( 0)	100%( 1) <sup>3)</sup>

\* Year the manuscripts received by Toxicol. Appl. Pharmacol and the published by NTP.

NTP; National Toxicology Program/NIH.

<sup>1)</sup>Dunnett's t Test (Dunnett, 1955).

<sup>2)</sup>Nonparametric multiple comparison procedures of Dunn (1964) or Shirley (1977).

<sup>3)</sup>NTP, TR329 only.

MC/RT: Multiple comparison/range test.

Table 4. Statistical analyzed items and times of combined chronic feeding and oncogenicity study using rats for 2 years

Items of analysis	No. of analysis	Remarks
1. Body weight (66+8)x2*	148	Every week from 0 to 26 weeks, bi-weekly thereafter and 8 times of weight gains calculation during dosing period
2. Food consumption (104+8)x2*	224	Every week and 8 times of total consumption calculation during dosing period
3. Food efficiency (52+4)x2*	112	Every week and 4 times of mean food efficiency calculation until week 52
4. Hematological examination (including of leucocyte and coagulation parameters) 16 items x 4 times x 2*	128	Analysis at week 26, 52, 78 and 104 during dosing period
5. Blood chemistry examination 20 items x 4 times x 2*	160	Ditto
6. Urinalysis 2 items x 4 times x 2*	16	Ditto
7. Absolute and relative organ weights 5 organs x 4 times x 2 items x 2*	80	Ditto
Total	868	

\*: Sexes.

Table 5. Changes in number of surviving male rats fed the diet containing test substance on 10 oncogenicity studies

Dose level	No. of study	Surviving number on weeks during dosing period											
		0	→	→	52	58	65	71	78	84	91	97	104
Control	10	50	→	→	50±1	49±1	49±1	49±1	48±1	47±1	45±2	44±2	39±5
Low	10	50	→	→	50±0	50±0	49±1	48±1	47±2	46±3	44±4	41±3	
Intermediate	10	50	→	→	50±0	50±0	50±1	49±1	48±1	47±2	44±2	40±3	
High	10	50	→	→	50±0	49±1	49±1	46±4	43±8	39±11	35±14	26±16	

Values are expressed as mean±standard deviation.



Table 6. The Examples for Nonparametric Dunnett' Test  
 -- Creatinine levels ( $\mu$ /dl) of F344 rats fed the diet  
 containing test substance for 52 weeks --

Treatment	No. of animals	Nonparametric Dunnett' test	Mean ranks	Sum of ranks
None	20	0.69 $\pm$ 0.07	52.2	1043
Low	18	0.65 $\pm$ 0.09	41.2	751.5
Intermediate	20	0.66 $\pm$ 0.10	41.9	837
High	20	0.69 $\pm$ 0.54**	22.5	449.5

Mean $\pm$ standard deviation.

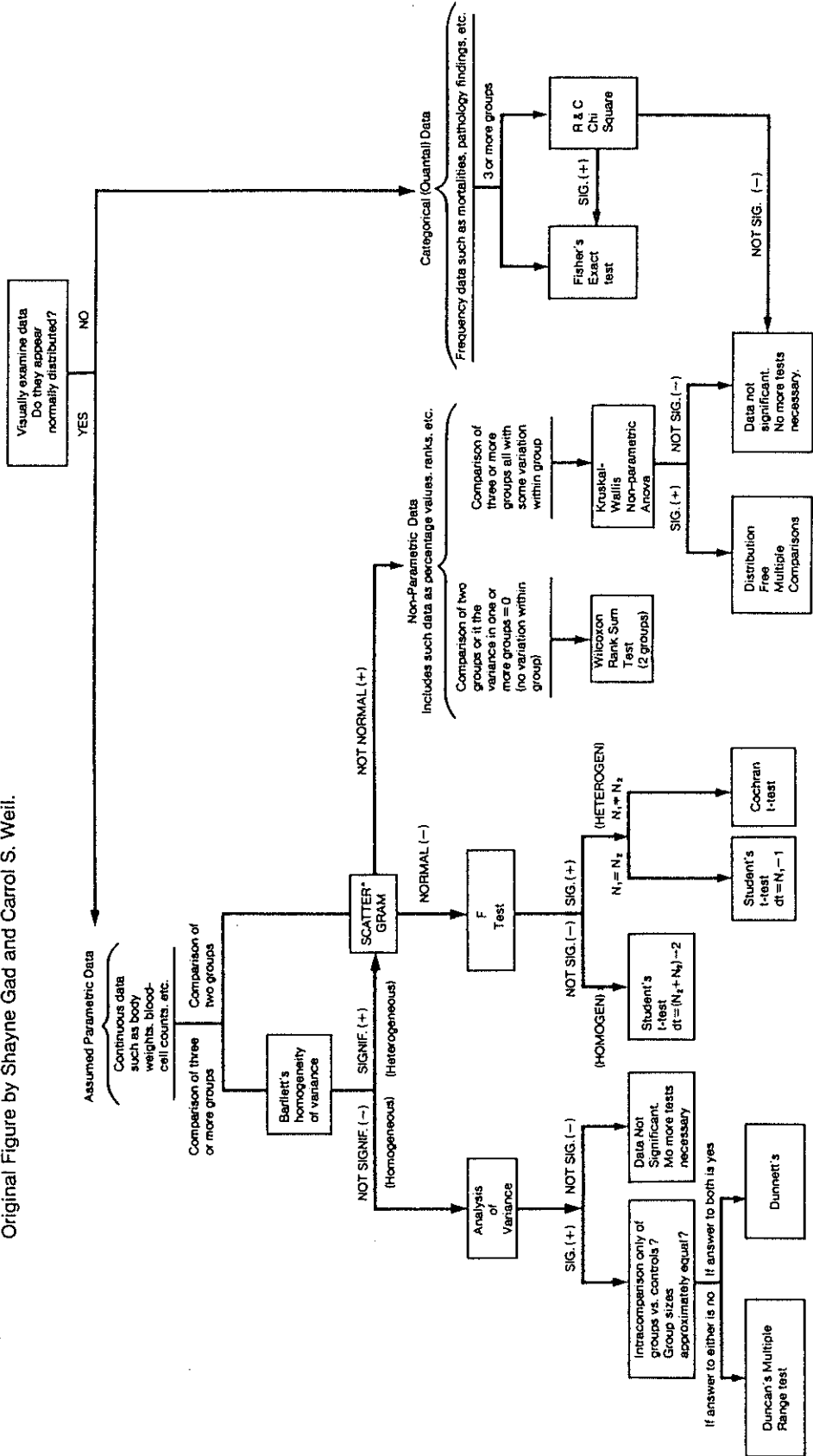


Fig. 1. Decision Tree for Selecting Hypothesis-testing Procedures.

\* 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.

[事務局だより]

ようやく37号をお届けすることができました。今年は忙しく、いろいろな仕事に追われているのに、本職の出版物は雑誌も書籍もほとんど刊行できずにきてしまいました。ともあれ、今号には報告を多数載せることができ、充実した会報にできたとおもいます。

私事になりますが、今年の4月より、理科大の統計セミナーにもぐりこみ『現代数理統計学』(竹村彰通著、創文社)の輪講に参加しています。少しだけ統計の基礎概念や統計的決定理論の枠組みを理解できたように思います。それを研究会活動に活かせればと願っています。もう一ランク上の研究会を目指して。(M.O.)

## 医薬安全性研究会 会報NO. 37

*Bulletin of Japanese Society for Biopharmaceutical Statistics.*

---

1993年7月31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 東京8-71335

印刷・製本 (有)ナガノ印刷