

医薬安全性研究会

Bulletin of Japanese Society for Biopharmaceutical Statistics.

会報 No. 34

Jul. 1991

目次

〈前 付〉

* 医薬安全性研究会スケジュール

癌原性試験の統計解析	菅野 純 (東京医科歯科大学).....	1
〈癌原性試験の統計解析〉 対しての議論	5
[基礎講座] (連載第2回)		
Analysis of 2×2 Tables	大橋信之 (安評センター).....	19
<hr/>		
〈特 集〉 十周年記念定例会		
決定樹方式での統計手法の利用の注意点	永田 靖 (熊本大学).....	24
中毒情報活動とデータ処理	新谷 茂 (中毒情報センター).....	38
薬の安全性をめぐる統計的問題	竹内 啓 (東京大学).....	41
製薬メーカーにおける		
臨床統計部の歩みとその将来	魚井 徹 (山之内製薬).....	43
変異原性試験における統計解析の実際	林 真 (国立衛生試験所).....	46
〈決定樹方式での統計手法の利用の注意点〉 質疑	48
<hr/>		
第43・44・45回定例会出席者名簿	64

1991年医薬安全性研究会

これからのスケジュール

- ☆ 1991年10月25日(金)～26日(土)……………データ解析講習会(総評会館)
 - ☆ 1991年11月16日(土)……………統計学会60周年記念共催・第48回定例会(専修大学)
 - *統計学者のための毒性試験入門
 - *薬剤予測学について 杉山雄一(東京大学薬学部)
 - *背景データの解析 芳賀敏郎(東京理科大学工学部)・松本一彦(東洋醸造)
 - ☆ 1992年1月25日(土)……………第49回定例会(総評会館)
 - ☆ 1992年4月11日(土)……………第50回定例会(総評会館)
-

★『医薬統計学—医薬品開発のための統計学』

今、基礎講座で取り上げている“Biopharmaceutical Statistics for Drug Development”の完訳が、刊行の予定です！鋭意編集中。

中上節夫・森川敏彦 監訳・KR研究会 訳 B5判 約600頁 予定価格9,300円

好評予約募集中。

- ★ 次回定例会は、統計学会60周年記念事業に共催して行なうため、日程を繰り下げ、11月16日となりました。それに伴い、データ解析講習会を繰り上げましたので、ご了承下さい。

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 $K_2 \times 2$ 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

Proceedings of the First Workshop on Computational Statistics

Theory and Applications in Computational Statistics

Edited by Ch. Asano, N. Niki and T. Okazaki (Kyushu Univ.)
B5 Size 290pp. ¥ 5,000YEN (¥ 310)

The effects and implications of the computerization of society are vast and universal. The impacts on statistics have been bearing a large amount and wide variety of fruits: new concepts, new methods and statistical softwares for creative problem-solving. *Computational Statistics* has been developing in these circumstances into a major research area in information science and technology.

This book contains research papers meant for presentation at the First Workshop on Computational Statistics held at Kyushu University, Fukuoka, in February 1991. Those papers cover a wide scope of topics in this promising research field.

Part of Contents

- "An Algorithm for Linear Structural Equation Model" by H. Ke and Ch. Asano
- "Study on Efficiency of Maximum Likelihood Estimates from Incomplete Data in Latent Class Analysis" by M. Watanabe
- "On the Latent Markov Chain Model and the Maximum Likelihood Estimation of the Parameters" by N. Eshima
- "Pooling of Data and Estimation after a Preliminary Test" by T. Okazaki
- "Strong Uniform Consistency of Regression Function Estimators When Errors are Dependent" by K. Liu *et al.*
- "Robust Estimation Using Iteratively Reweighted Least Squares Technique" by K. Yamaguchi
- "An Improved Algorithm of Interaction Spline Fitting" by X. Chen *et al.*
- "Collapsibility of Parameters in Logit Models" by Z. Liang and Z. Geng
- "An Overview of Analysis of Ordinal Data" by H. Uesaka
- "Normal Orthant Probability and its Application in Order Restricted Inference" by H.-J. Sun
- "Change of Bases of Symmetric Polynomials of Several Systems of Quantities and their Applications to Distribution Theory of Multivariate Statistics" by S. Nakagawa and N. Niki
- "Random Sequences and Runs Test" by M. Ichimura
- "The Z-Chain Structure—A Data Storage Form for Map Informations" by D. Hu *et al.*
- "Recent Developments in Statistical Methods Applied to Survey Research" by H. Ke
- "A Study on Sampling Inspection Plans with Adjustment—ISO 3951—" by H. Kimura
- "CAT: A Statistical System for Analysis of Categorical Data" by Z. Geng *et al.*

.....

Orders should be addressed to: Scientist Inc.
Yamazaki-bldg. 3-2
Kanda Surugadai, Chiyoda-ku
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	

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
〈具体的状況〉				
〈質問内容〉				

(コピーしてお使い下さい)

癌原性試験の統計解析 抄録：

菅 野 純（東京医科歯科大学）

癌原性試験における統計処理の特殊性を 表1 にまとめた。まず、ある処置によっておこる病変の母集団を想定すると、これは未知である。何匹かの動物を用いて実験をすることが、未知の母集団からサンプルを抽出する行為と考えられる。サンプルの構成要素は動物であるが、その中に種々の病変が生じてくる。表1ではA-Zで表したが、具体的には肝臓Adenoma、肝硬変、肝炎、肺Adenoma、肺炎、等等、腫瘍性病変に限らず、変性、炎症等をふくめた病変という要素が無数に含まれている。次に、病理医が剖検、切り出し、検鏡、あるいは血清検査結果を参照しつつ所見を取り、病名を記録して行く。この、神のみぞ知る「病変」→「所見」変換(①)時にどうしても病理医の力量や、その時々興味に影響されて、かなりの病変が無視される。言い替えると「情報が失われる」。

この時、病理医あるいは実験者は直観的に「あ、これは何かあるな」とか「これは何もないな」とか、何か思っているはずであり、これが最後まで大きな意味をもつ。

次に、統計処理出来るように「所見」を「数値」に変換する作業にはいる(②)。すなわち肝のAdenomaのあった動物が何匹、腹膜中皮腫のあった動物が何匹、等等、数値化する。もちろん、付随した情報、例えば、第何日目の所見か？、病変が大きい小さいか？、等等、数値化し得る項目は多いが、無意識、あるいは故意に、項目の取舍選択が行われている。大幅に情報が削られているのが現状であろう。この時、情報の取舍選択と、統計処理方法の選択とは密接に関係している。Fisherの直接確率計算法をやろうと思えば、初め動物の数と担癌動物数だけを用意すれば済むし、途中死亡を考慮するときは、時間的経緯に関するデータが必要になる、といった具合である。

次に、選んだ方法で統計計算をし、 $P < 0.05$ あるいは $P < 0.01$ 等の水準で検定する(③)ことになるが、 $P \neq 0$ という点で曖昧さがある。

次は、ある実験で1項目だけ、即ち「フェノバルビタール投与が肝臓のAdenomaを増加させない」という帰無仮説が $P < 0.01$ で棄却された」とする。実験者が「フェノバルビタールが肝adenomaを増やした」と思った瞬間、彼は「数値」を「所見」に再変換したことになり(④)、「フェノバルビタールは肝adenomaだけを増やす」と報告した瞬間、「数値」を「病変」にまで再変換した事になる(⑤)。

表1の①～⑤の各段階において、各々に問題が生じうる。

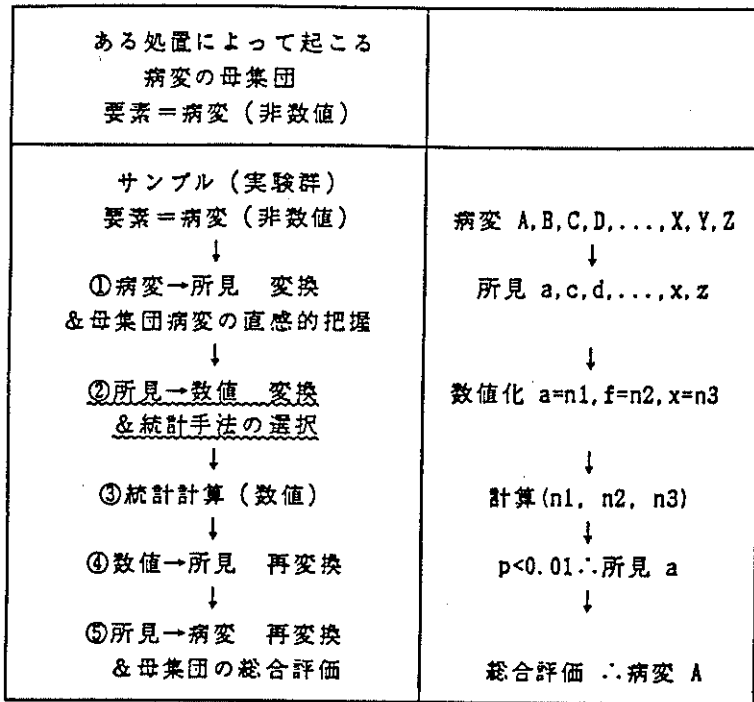
①では純然たる病理学の問題として、正確な所見がとれたかという問題、母集団の印象をどれくらいとらえられたかという問題が生じる。②の段階では、2つの事態が想定される。病理の用意した情報量が、統計学が必要としている情報量より足りない場合と、その逆の場合である。③の段階では、純然たる統計学の問題としての偽陽性、偽陰性の発生の問題がある*。④⑤は①②の逆操作であるが、エッセンスのような少数の情報で未知の母集団全体を代表する所見として提示されるために全く見当違いの結果が出ている危険性もなくはない、という問題である。

今回の(今回も)私のお話は②の_____の部分を中心に進めて行きたいと考えます。

*有意差 $P < 0.05$ で、20項目、30項目独立に有意差検定した時、どれか一つに偶然、有意差が認められてしまう確率はそれぞれ64%、79%と計算される。つまり、単純に30項目を $P < 0.05$ で検出していけば、全然有意差のないはずの実験10回のうち7、8回はどれか1項目に有意差が認められる勘定になる。故に、 $P < 0.05$ で有意差が認められた場合に、それが単なる偶然かどうかを見極めなければならないという事態が頻繁におこる事になる。それに対処するためには、その際に役に立つような緻密なデータ採取を日常的に行うことが望ましいと言うわけである。

<表 1 >

癌原性試験における統計処理の特殊性



- ① 病理学の問題：
病変 → 所見 変換 = 正確な病理所見が取れたか
母集団の病変全体のイメージを直感できたか
- ② 病理学と統計学の間で生じる問題：
所見 → 数値 変換 & 統計手法の選択 = 統計に合うように
情報をトリミング
= 情報の切捨て
その統計手法ではどれだけの情報を切り捨てることになるか
・統計が扱える情報量 > 病理が用意した情報量
情報の切捨てすぎ・病理の統計手法に対する無知
・統計が扱える情報量 < 病理が用意した情報量
病理の抱くイメージを統計手法が処理しきれない
- ③ 統計学の問題：統計学自体による偽陽性・偽陰性の発生
- ④ 病理学と統計学の間で生じる問題：
数値 → 所見 変換 = トリミングされた情報についての結果
のみがえられる
- ⑤ 総合判断の段階で生じる問題：
トリミングされた所見 → 母集団の病変 = 直感との対比が必要
= 総合判断

蛇足的付録：

例 1-1：

1箱100粒入りであることがわかっているキャンディーが一箱与えられたとする。実はそのうち95個は白で5個は赤であるが、あなたには知らされていない。もちろん、外からは、中身が見えない。

無作為に4個取り出して、それについて赤白を調べ、また箱に戻す。この操作だけが許されているとすると、これを何回か繰り返し、データを蓄積するうちに、

白：赤=97：3	± 誤差 5
白：赤=94：6	± 誤差 2
白：赤=95：5	± 誤差 0.9
白：赤=95：5	± 誤差 0.4

と、次第に誤差が小さくなる。キャンディーの個数が整数値であることを考えれば、無作為抽出を無限に繰り返さなくても、

白=95個、赤=5個
であることが分かる。

例 1-2：

ここは、菓子製造工場である。1箱正確に100粒入りのキャンディーが次々にベルトに乗って流れてくる。実は、全ての箱には正確に95個の白いキャンディーと5個の赤いキャンディーが詰まっているが、あなたには知らされていない。

ベルトを流れてきた箱から一回だけ無作為に4個取り出して、取り出したものについて赤白を調べ、また同じ箱に戻す。次に流れてきた箱についてまた一回だけ同じ操作をする。この操作だけが許されているとすると、何箱かについてのデータを蓄積するうちに、

白：赤=97：3	± 誤差 5
白：赤=94：6	± 誤差 2
白：赤=95：5	± 誤差 0.9
白：赤=95：5	± 誤差 0.4

と、次第に誤差が小さくなる。キャンディーの個数が整数値であることを考えれば、無限個の箱について繰り返さなくても、この工場で箱詰めされた箱は

白=95個、赤=5個
であることが分かる。

例 1-3：

ここは、別の菓子工場である。1箱正確に100粒入りのキャンディーが次々にベルトに乗って流れてくる。しかし、この工場では、白と赤とを適当に混ぜている。だいたい白が95個前後になるようにしてあるが、±2個ぐらいの誤差はよくある。このことを、あなたは知らされていない。

ベルトを流れてきた箱から一回だけ無作為に4個取り出して、赤白を調べ、また同じ箱に戻す。次に流れてきた箱についてまた一回だけ同じ操作をする。この操作だけがあなたに許されているとすると、何箱かについてのデータを蓄積するうちに、

白：赤=97：3	± 誤差 5
白：赤=94：6	± 誤差 2
白：赤=95：5	± 誤差 1.5
白：赤=95：5	± 誤差 1.4

と、次第に誤差が小さくなるが、ある値以下には決して小さくならないことに気がつくであろう。
 1000箱調べても1万箱調べても、
 白=95±2個、(赤=5±2個)
 であることが分かる。これは、キャンディーを詰める際の誤差に他ならない。

例2:

ここは、賭博場である。プロの壱振り士がいて、丁・半をかなり自在に振り分けるらしい。

胴元は、最近どうも儲けが減っておかしいと思っていた所へ、顧客の一人から「八百長」の苦情があった。そこで、壱振りが怪しいとにらんだ胴元は手下に「100回のうちで丁と半が何回ずつでたか数えろ」と命令した。結果は、

丁=46、半=54であった。統計計算の結果は「八百長とはいえない」であった。

次の日、同じ客から「やっぱり八百長だ」と苦情があったので、手下に「今度は1000回数えて来い」と命令した。結果、丁=512、半=488で、やはり「八百長とはいえない」と結論した。

翌々日、その顧客から「なにいつてんだ!ふざけんじゃねー、八百長だ!」といわれた胴元は、手下に「丁と半とがどんな順番で出たか50回分記録して来い」と命令した。

結果は、

丁丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半
 丁半丁半半丁半

(丁=23、半=27)

であった(規則性があれば八百長がわかると思ったのだが)。

丁と半の出方が本当にランダムかを検定するには、run検定を行えばよい。

丁丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁半丁
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

半半丁丁半半半丁丁丁半半半半半半半半半半半半半半半半半半半半半半半半半
 16 17 18 19 20 21 22 23 24 25 26 27 28

runの数=28、丁の数=23、半の数=27

1標本run検定 (丁>20、半>20であるから)

$$Z = \frac{r - \left(\frac{2n_1n_2}{n_1+n_2} + 1 \right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}}$$

故に Z=1.20

正規分布表から 両側 P=0.1151 x 2=0.23

故に、丁と半はランダムであったという帰無仮説は棄却されない。

そこで、胴元は「八百長なし」と判断せざるを得なかった。

ところが、儲けは減るし、複数の顧客から「もっとよく調べろ」と怒鳴り込まれるし、途方に暮れた胴元であったが、手下に、「気付かれないように天井裏から、客一人一人について丁半のどちらにいくら張ったかを、50回分正確に記録して来い」と命じた。

その結果、客Dが高額を賭けたときには必ず客Dが勝ち、客Dが少額を賭けたときには客Dが必ず負けることがわかった。この関係は、検定の結果、統計学的に有意であったので、壱振り士と客Dをつるし上げたところ八百長を自白したのであった。

〈癌原性試験の統計解析〉 対しての議論

吉村(名大) : Petoの方法では他の病気で死んだか、またはある腫瘍で死んだかということ
を区別するのでしょうか(表1、表2参照)。

菅野 : はい。肝癌のようにゆっくり育つ腫瘍の場合は、肝癌で死んだ動物と他の原因で死
んだ動物に見つかった肝癌とも区別します。しかし、「他の病気」の発生に関して興味があ
れば、「他の病気」は意味をもちません。

表1 PetoのContext of Observation : 腫瘍の観察時の状況

CONTEXT 0 (状況0) = SCHEDULED SACRIFICE CONTEXT	(計画屠殺状況)
CONTEXT 1 (状況1) = OTHER INCIDENTAL CONTEXT	(その他の偶然発見状況)
CONTEXT 2 (状況2) = PROBABLY INCIDENTAL CONTEXT	(多分偶然発見状況)
CONTEXT 3 (状況3) = PROBABLY FATAL CONTEXT	(多分致死状況)
CONTEXT 4 (状況4) = FATAL CONTEXT	(致死状況)
CONTEXT 5 (状況5) = MORTALITY-INDEPENDENT CONTEXT	(死亡に無関係な状況)

前提：
◎途中死亡の起こる率は群の違いや担癌の有無に影響されない。

表2

補足： 1. 1種類の腫瘍が多発したとき → 通常、total で1つの腫瘍と考える 2. 状況3・4の腫瘍 → 1匹に1個 3. 状況3・4の腫瘍が見あたらない場合 → 「無し」のままにする 4. 良性腫瘍でも重要器官を傷害し致死性であれば → 状況3または4 5. 腫瘍の組織学的悪性度を考慮して判定して良いのは状況2と状況3との間で迷ったときのみであ る。 6. 状況5は触診で判定してはいけない。 7. 皮膚腫瘍でも、剖検時にしか存在をチェックしない場合は内臓の腫瘍と同様に扱う。状況5は用 いてはいけない。

吉村 : 最後に剖検して腫瘍が見つかるという場合と、途中他の病気で死んでしまって剖検
して見つかる場合を区別するのですか。

菅野 : それは区別しません。Context 0,1,2ということでは区別しますがPrevalence method
で行うという、計算の中では区別していません。但し、インターバルを独立させるという
点では区別されます。担癌の有無にかかわらず、途中死亡には差がないという大前提がこ
こでも入ってきています。

大橋(東大) : その腫瘍で死んだ、死なないという判定は病理での判定ですか。そうすると、

病理学者による差というのは大きいと思いますが、それが解析に与える影響というのはどうなっているでしょうか（表3参照）。

表3

PETOの検定	Context 5	Context 3 + 4		Context 0 + 1 + 2	
腫瘍の種類	① 表在性腫瘍 表在腫	② 急速に増殖し 動物を殺す 腫瘍 (急性腫瘍)	③ 緩徐に 動物を腫 (緩徐)	増殖し 殺す 瘍 腫瘍)	④ 非致死性 腫瘍 不定(偶然)
理想の観測点 現実の観測法 理想との隔たり	発生日 外表観察 = 0	発生日 剖検 = 0 (致死)	発生 剖 不定(致)	日 検 死偶然)	発生日 剖検 不定(偶然)
	Onset rate method	Death rate method		Prevalence method	

状況5（表在性腫瘍）の場合：発生率法(ONSET RATE METHOD)
 仮定 = 処置が腫瘍発生に影響を与えない。
 本日中に腫瘍が発生する確率はどの群でも等しい = 分母は腫瘍の生じていない生存匹数

状況3+4（致死性腫瘍）の場合：死亡率法(DEATH RATE METHOD)
 仮定 = 処置が腫瘍発生に影響を与えない。
 本日中に腫瘍で死亡する確率はどの群でも等しい = 分母は生存匹数

状況0+1+2（非致死性 or 偶然発見腫瘍）の場合：有病率法(PREVALENCE METHOD)
 仮定 = 処置が腫瘍発生に影響を与えない。
 過当に短い期間に中間死した動物に、解剖によって注目する腫瘍が見つかる
 確率はどの群でも等しい = 分母は剖検匹数

菅野：当然あります。ですからContext を区別しないときよりも、ある程度の実力をもった病理学者が多少間違っていたとしてもContext を区別したほうがより正しいのだという考え方があります。明らかにそうであると言い切れる病変と曖昧だという病変があるわけですが、それをPetoはContext 0,1,2,3,4の、2と3に押し込めています。心配ならば曖昧な部分をすべて、どちらか片方に押し込んだときに結果がどうなるか、ためし算をしてみなさいということも言っています。それで判定が変わるようだったら問題だし、それでも変わらないくらいははっきり分かれていればそれでいいということです。病理診断が主観だと言われればそのとおりだと思います。

大橋：その曖昧さを考慮した統計手法などが出ているようですが、それは実際の場のほうには入ってきているのでしょうか。例えば論文の研究レベルで、曖昧さを考慮したPetoの検定法のmodificationなどは。

菅野：現場ではまだPetoが普及していない状況かと思います。つまりFisher主体ということです。Petoのmodificationは勉強不足で知りません。

逆にNTP(National Toxicology Program)ではこういう分類はやめてしまえという論文が出ています。それはなぜかと言うと、一つにはアメリカで行われている実験ではまともに病理医が見ている実験がそれ程ないためではないかと思うのですが、日本とかヨーロッパのほうはずっとまじめに見ているのではないのでしょうか。そういう意味でNTPの方法はむしろ過去のデータを利用したりするときに、ほとんどエッセンスしか残っていない場合にはそうしなさいということであって、決してそれがいいのだという問題ではないと思います。

大橋：実験のやり方で途中解剖という考え方がありますが、それによって情報量は増えるのでしょうか。ある種の癌にはいいとか悪いとか、どういう癌にはそれがいいのかという観点からではいかがでしょうか。

菅野：情報は増えるのではないのでしょうか。ただし、Cost performanceの問題が大きいと思います。単なる経費のみならず、知りたい点にどれだけ答えてくれるか、という点も問題です。そういう意味で、癌原性試験で途中解剖を行うべきかということはわからないです。ただPetoの検定にはそれも組み込めますが。発癌作用の弱い物質ですと微細な前駆病変とか関連病変の解析がどうしても絡むのですが、そういう場合に途中解剖した方がいいのかも知れません。ただし、それを一律に推進してもよいのかどうかという判断は大変難しいことだと思います。

滝沢(日本ロシユ)：Contextのほうでは段階的に分けながら、解析のほうではそれをまとめて処理してしまうことには何か理由があるのですか(表4参照)。

菅野：一つには先程言ったように、自信の持てない領域をとりあえずどちらかにスパッと分ける。つまり「あなたを信用しますからContext 2と3で分けなさい」としておいて、微妙な結果が出たときに確信の持てるデータだけでもう一回計算してみなさいというときの便宜のためだと思います。

ためしにContext 0+1と2+3+4、または0+1+2+3と4で解析するというをやってもいいですよということが書いてあります。そのために2と3がつくってあるとも考えられます。

滝沢：標準的なPetoの計算の方法ではそれは決まっているわけです。たとえば0+1とか、3+4とか。初めからそういうふうに分けてはいけないのですか。

菅野：そこは良心的に構えるかどうかの違いだと思うのです。ContextをFatalとincidentalの2つに分けると言う、だれしも、それは「きつい」と考えるでしょう。Fatalとincidentalと不明の3つに分けると、そのうち不明が増加するでしょう。Fatalと思われるものだけFatalとし、すなわち、non-fatal、すなわち、incidentalが増加してしま

表4

まとめ1. PETOの検定： 病理学的な所見を最大限に利用した統計手法
殆どどの実験の評価が問題なく行われている

発見状況	Context 5	Context 3 + 4		Context 0 + 1 + 2	
腫瘍の種類	表在性 腫瘍	急速に増殖し 動物を殺す 腫瘍	緩徐に し動物 殺す	増殖 を 腫瘍	非致死性 腫瘍
発見日	= 発生日	⇒ 発生日	死亡日	偶然	偶然
	Onset rate method	Death rate method		Prevalence method	

	Non-adjusted method	Age-adjusted Death rate method	Age-adjusted Prevalence method
群間比較	分割表 Fisher's exact test	Kaplan-Meier ☒ Log-rank test	interval を設ける Log-rank test
用量相関	分割表 Trend test	Context of observationに基づく表 Peto's Method	

大前提 . 中間死の起こる率は群、担癌の有無によらない。
 . Context 3 + 4 は②急性腫瘍として計算される。
 . Context 0 + 1 + 2 ④非致死性腫瘍として計算される。
 多い間違い = Prevalence method の分母を生存匹数にしてしまう。
 = prevalence method の分母から context 3, 4 を除くのを忘れる。

うでしょう。こういうBiasを避ける目的があろうかと思えます。2と3との区別を常にさせるということは、病理医をさばらせないためでもあるのでしょうか。

滝沢: Petoの方法を、たとえば会社などでやろうとしたとき、Context に分けるというのはかなり大変です。そこが混乱しているように思うのです。

吉村: Context の0は全然問題なく、4もわりとはっきりしています。偶然発見の1も問題はありません。結局2と3が非常に曖昧なわけです。だからそのところを病理学者が「おれだ」と言って判断する形になるのですか。

菅野: そうです。

吉村: それに対しアメリカなどではその辺は判断しないということですか。

菅野: そうだと思います。ある、なしだけでいっているのだと思います。Petoは2と3に入る状況は心配する程は多くないといっているわけです。

吉村: 純粹に数学的な理論的な議論をすると、曖昧なものは排除してしまったほうがいいということが出てくるのです。ただ今の場合には、時系列的に変化しているものだから、単純に排除するとしてもどこのインターバルから排除するかという問題が出てくるから、

かなり丁寧に議論してみないとどちらのほうで統計的検定として性能がいいかというのはわかりません。

大橋：今までにいろいろな物質についてのデータがあるわけですが、Context 1,2と3,4の間に差があるとして、それで本当にある種の情報量が増えているのかということ、数学モデルで議論するのはすごく難しい話です。過去のデータがどんな発癌物質で出てくるのかを見て整理していくと、ある程度モデルというのが出来てくる可能性がある。そういうモデルに基づいて、分けたほうがパワー（検出力）が上がるのかどうかという議論をしないといけないのです。

経験に基づいた、あるいは数学的考察に基づいたPetoの検定に対する正当化というのはあるのですか。

菅野：5,000匹程度のデータで解析した例が示されています。それは下垂体腫瘍のデータについてですが、Fisherではうまく解析できず、Petoでうまく行ったということと、Context 2と3は思いのほか少なかったという事が示されています。しかし数学的考察についてはわかりません。

大橋：発癌物質でどういうタイプが出てくるのか大変関心があるわけです。全体として見ると担癌の割合が変わらなくても、Context のdistributionがどうかということによって手法の良さが言われるべきだと思うのです。それで曖昧なことによる、あるいはmisclassificationによる情報量のロスとかパワーのロスという議論がそこに数学的に展開出来ると思うのです。

菅野：一応、さっき申し上げた5000匹の下垂体腫瘍の例で、Petoはそれを示したことにしているのだと思います。ただし、病理組織学的に同じ肝癌のContext をどれにするのかというときの問題と、甲状腺の腺腫のような完全に非致死性のものが他の病気と絡んだ時とちょっと意味合いが違うと思います。非致死性の腫瘍だけの場合と、致死、非致死の両方起き得る腫瘍の場合とでパワーが全然違うのではないのでしょうか。ただ非致死性の腫瘍に関しては、この方法がベストであろうということでしょうか？

大橋：それはベストではないです。理論的にはhazard rate が時間によらず一定であるということが、ある程度成り立てばログランク検定に基づく方法が漸近的に一番いいでしょう。ただそれが成り立つかどうかはわかりません。しかし、hazard rate が一定ではなくてそれが変わるというのは、多分あまりないだろうと思います。それでログランクでいいと思うのです。それは人間の癌とかマウスを扱うときと同じです。大体の状況で、ログランクというのは結構うまくいくのです。人間の場合は非常にたくさんデータがありますから、それで見ると危険率、死亡率の比というのは似たような感じでいって、それが極端にクロスすると変わってしまうということはないだろうと思います。そういうことでログランクを使うのです。

また計算が一番やりやすいということです。それから時間を直接解析に入れるというやり方は、実験の現場でやることは出来ないくらい複雑だし、モデルの妥当性はチェックが不可能だから、どうしてもノンパラメトリックな方法を使わざるを得ない。そこでログランクになっているのです。

浜田（武田）：癌原性試験の解析でtime adjusted の方法と、time adjusted でない方法があります。Fisherでやるのか、Petoでやるのかということなのですが、最初に生存率に差があるかどうか調べ、差がなければFisherでもいいのではないかとということが書いてあるのを時々見かけるのですが、その点についてはどうお考えでしょうか（表5参照）。

表5 腫瘍分類 + 生存・剖検匹数による補正→

	Non-adjusted method	Age-adjusted Death rate method	Age-adjusted Prevalence method
群間比較	分割表 カイ自乗検定 Fisher's exact test	Kaplan-Meier 図 分割表の複合化 Log-rank test (Mantel-Haenszel)	interval を設ける 分割表の複合化 Log-rank test (Mantel-Haenszel)
用量相関	分割表 Cochran-Armitage test for trend	Context of observationに基づく Peto's Method	

菅野：大丈夫なことが多いけれども、大丈夫ではない場合も考えられるというのが正しいのではないのでしょうか。つまり死んでいく動物がいて、その死んでしまった動物が計算の分母から削られているわけです。だから見かけ上同じように分母から削られていけばいいですが、腫瘍の出来る時間が早い遅いという差があった場合には分母が違うところで計算していることになりまますから正確ではないです。

浜田：いつもPetoを使っていたほうが良いということですか。Fisherを使わないでPetoだけで統一してしまったほうがよいというお考えでしょうか。

菅野：Petoの方が良いと思いますが、Petoだけかと言われると……。少なくともPetoの解析の見方をして、同じ生存カーブでも腫瘍の発生時期がずれているとか、そういうものが見えるかどうかを考えてくださいとしか言いようがないです。Petoさえやっていたらいいという考え方もありますが、それは非常に危険な発想ではないかと思います。特にComputer化してしまった場合、Context 分類に問題があっても盲目的にデータを入力することが考えられますが、盲目的にPetoをやることは、盲目的にFisherをやるより、ある意味では危険だと思います。

吉村：生存率の場合ですとFisherよりもPetoの考え方のほうを優先してもいいのではないかという気が少しするのです。というのは、これがショートタームの試験なら問題がない

のですけれど、ロングタームですから、初期の段階に何か偶発的なことで死んでしまった例がいくつか出ると期間が長いということが影響して、多分生存率曲線ではものすごく影響を受けてしまうのです。だからこういう長期間の試験の場合には、Fisherの最終生存率が違うか違わないかというのは別の情報として使ってもいいけれども、少なくとも生存曲線の評価においては初期の偶発的な死亡で左右されることはないほうがいいのではないかという気がします。ことに生物というのは個体差が非常に大きいから、ある意味では conditional にやる議論なのです。つまりスタートにおいては全く同じように割り付けたって、個体差の影響が早く出たり遅く出たりということだったとき、それはもう動かせないものとして、途中で死んだものと前提してやっていくようになるのです。だから生存率が仮に同じだったとしても基本的には生存時間解析はこういうPeto的な議論をしたほうがいいと思っているのです。

浜田：癌原性試験は普通4群とか5群で各群50匹ぐらいずつで行いますが、そのときに最終的な解析として、たとえば対照群と最高用量群などとの対比較でPetoの方法をやるとか、あるいはtrend を見るとかいろいろなものがあると思うのですけれど、最終的に、この化合物に腫瘍を増加させる傾向があるかないかという判断をするには、対比較よりもtrend で見たほうが良いという考えかどうかお伺いしたいのです。

菅野：最終判定の話はまた別だと思うのです。群間比較で出てtrend でももっと出たという場合の判定の仕方と、群間では出たけれどtrend は全く出ないというときや、群間では全く出ないけれどもtrend にするとはっきり出たというときの判定とは全部違うと思います。

浜田：併用して両方見ていくのがいいでしょうか。

菅野：はい。差が出たのはどうしてかということ考えた上で判定してくださいという意味です。trend で出た、群間が出たということだけにかかわってしまっただけではいけないと思います。群間だけでは駄目でtrend だけでは駄目というのが正しいと思います。

浜田：trend をやる時に両側検定と片側検定があるのですが。

菅野：trend は一応片側でやることになっています。ですからtrend が使えると判断したら片側でいいのです。極端に言えば実験をやっている、これは抑制に働いていると判断したら反対側の片側でやればいいのです。片側が正しいと思います。

大橋：萎縮してしまって、腫瘍発生率が高用量で落ちるという例が甲状腺でありましたが、他の臓器でもよくあることなのですか（表6参照）。

菅野：ホルモン臓器では多いかと思います。全身状態をあまり変えずにその臓器だけつぶしてしまうというのは私は甲状腺でしか見たことがないのですが。

大橋：Ames試験などで細胞毒性で細胞を殺してしまって復帰変異が起きないということと同じですね。そういうのはそれなりに検定が考えられているのですが、先生がおっしゃられ

表 6

3. 動物全体の状態の把握：

- a. 極端な体重増加抑制状態では腫瘍発生は抑制される。
- b. ホルモン減少状態でのホルモン依存性臓器の腫瘍発生は抑制される。
- c. ホルモン増加状態でのホルモン依存性臓器の腫瘍発生は促進される。

判断材料：体重曲線・ホルモン依存性臓器の異常な萎縮、肥大、等

上記の事態が起こっていると思われる時は、必要に応じて追加実験を行い、その事を確認・考慮した上で評価を慎重に行う。

たように“見ればわかる”というところがあるのです。

腫瘍は病理で見ればわかります。“これは萎縮してしまっているのだ”というのがわかります。Ames試験も見る人に言わせれば、死んでいるのか変異原性なしなのかわかるというわけです。その辺の情報をうまく取り入れるということは、実は統計家のほうはあまり考えてないわけです。それを統計学のフレームワークでどう扱えばいいのかを考えていないのではないのでしょうか。一種の後知恵をやらなければいけない状況なのかもしれないけど、全体を覆う戦略はもっていないのです。

それから人間の前立腺癌みたいに大体に出てくるといような癌はないですか。大体100%近く見れば出てしまうという癌はありませんか。

菅野：いわゆる良性腫瘍が多いですが、あります。

大橋：そういうものに関してログランク検定でいいのかという問題があると思います。つまりみんな最後には出てしまうというものだったら、むしろ進行のスピードの問題でしょう。そういうのと最終的な率などが問題になるのはちょっと違って、最終的にみんな出てしまうものにログランクでやるのはあまり本筋ではないし、やはり進行のスピードを見なければいけないと思うのです。

みんな出てしまうのはやはり問題だと思うのです。スピードを議論すべきだと思います。ログランクはそれに対してパワーは弱いのです。

吉村：それから面積をみる時の問題なのですが、増山元三郎先生がよく言うのですが、黄変米のときに肝臓をプレパレートで調べたときに、最初は差がなかったけれど、もう一回面積を広げてやったら明らかに有意差が出てしまったという例があります。検出力を上げるという意味では組織をなるべく丁寧に見てやるほうがいだろうという気がします（表7参照）。

ただそれが大橋先生が言われたように、大抵のものに少し出ているのだけれども、平気でみんな生きているのだということがもし仮にあったとすると、丁寧に調べれば全部出て

しまうため、それならば検索面積あたりの量みたいなののでインデックスを作ったほうが影響量の大きさは評価出来ると思います。そのときには有意差ありということが、すぐ死亡率のように、これだけの差がこれだけの致命的なダメージですよという言い方は出来なくなって、たとえばインデックスで影響があるということしか言えません。

だからガイドラインみたいなもののときに、それをどう考慮に入れるかという議論になるだろうと思うのです。そうするとサンプルから数値に変換するところでは、多分統計学者と、病理学者と、ガイドラインを決定する行政側とが相談して決めて、ただ基準だけは明瞭にしておいて、後は社会的にルールとして“大体その辺でいきましょう”ということを議論すればいいのではないかと思います。

表 7

問題点：
1. 標本の面積と顕微鏡的病変の検出率
2. 肉眼的に確認できた病変を切り出して標本で確認した場合と標本中に偶然見つかった病変は同等か？
3. 小さな臓器の小さな病変と、大きな臓器の小さな病変は同じ様に有意差検定にかかるか？
4. 発癌率を左右する可能性の大きい非腫瘍性病変（変性・壊死）の解釈
未解決 研究者の経験と認識

滝沢：精巣の間細胞腫のようにほとんど100%出るものについて、スピードを比較する方法はあるのですか。

菅野：途中解剖しかないでしょう。

大橋：ある特定の臓器にだけ腫瘍を発生させる臓器特異性というのは多くの発癌物質にあると思うのですが、それは非常に強いものなのでしょうか。つまり1つ、2つの臓器にだけ腫瘍が出るものなのか、もうちょっと4,5,6の臓器に出るものなのか、あるいは大抵の発癌物質は強ければどこへでも出てしまうのか。

菅野：経験上は1,2臓器だと思います。いわゆる古典的で強力なcomplete carcinogenであるメチルコラントレン、ニトロソウレアの強いものとか、DMBAは大量にやれば多臓器に発癌作用を及ぼしているのですが、それでも結局一番早く出て一番早く動物を殺す腫瘍しか、生じてくるのを観測できないことが多いようです。その裏付けの1つの例として濃度を下げていくと濃度によってtarget organが移動するということがあります。

大橋：だからやっかいなのですね。

菅野：最近弱いプロモーター的なもの、変異原性もないようなものが検定に引っかかったりします。そういった場合でもtarget organは少ないのが普通ではないでしょうか。

大橋：ぼけてくる感じになっていろいろなところに出てくるのではないですか。

菅野：弱いから分散するという事はないように思います。特に、薬物代謝が関与して発癌作用を現す物質の場合は、target organは、はっきり決まることが多いと思います。

大橋：特異性が高いと多臓器をまとめるという話が非常にやっかいになる。

菅野：大抵のこういうものでは多臓器をまとめると有意差は減る方向にいくのだと思うのです。

大橋：もちろんそうですけど。みんなが同じように動いたらまとめるということが成り立つわけです。第一種の過誤を押しえつつパワーを高めることがそれで出来るのだけど、非常に特異的に出てくるとなると、多臓器をどう扱うかというのは非常にやっかいで、統計上は問題になるのです。少し論文などで多臓器をどう扱うか、その多重性をどうやるかという提案は出ているけれど、まだまだでしょうね。

林（国立衛試）：データを最終的に評価するような場合に、今のお話を聞いていますと、実験の実施者の直感みたいなものが最終的にはかなり影響を及ぼしているのだと思うのですけれど、検定だとか、ヒストリカルなデータだとか、その辺を使って最終的な結論に導くストラテジーみたいなものはあるのでしょうか（表8、表9参照）。

菅野：そういうご質問があると、いつもケースバイケースだということにしています。化合物の系統が似ているのだったらそれなりの方式は立つかもしれないのですが、系統の全く違う化合物をやりますと、全然違うところから問題が沸いてくるのでケースバイケースとしか言えないと感じております。

林：たとえば検定でP値が0.05ぐらいの場合だったらもう少しヒストリカルなものを使っていろいろ考えてみるとか、P値が0.01ぐらいだったら検定の結果だけを全面に打ち出して話をまとめてしまうとか、そういうふうなこともあるのですか。

菅野：一般的にはそう言っていないのではないのでしょうか。0.01だったらかなり本当のことが起こっていると信用していいと思います。

表 8

まとめ
2. 有意差検定の結果が微妙な場合、あるいは病理的イメージに合致しない場合 前提・関連性病変
病変を数値へ変換する際に大切な所見を切り捨てていないか？
3. 最終評価には病理学的な経験に基づいた直感が大事である。 Historical control の把握（文献上のデータ・各研究施設固有のデータ） 動物全体の状態の把握 今までの報告との対比 動物種・用量設定・投与期間・投与方法

表9 発癌性試験のために統計学者が用意している手法を
運用するために必要な情報

- | | |
|------------------|------------------|
| 1. 発見され方による腫瘍の分類 | (統計学的腫瘍分類) |
| 2. その時々生存動物数 | (Age adjustment) |
| 3. 用量設定のデータ | (傾向性検定) |

動物一匹一匹について

1. 剖検日はいつか
2. 腫瘍の組織分類はなにか
3. 腫瘍がどの様にして発見されたか (発見の状況)
動物の死因は何か

というデータを 剖検時肉眼所見 + 病理組織所見 をもとに揃えなければなら
ない。

しかし $P=0.01$ だから絶対信用しなさいというのも無謀だし、それはP値あつての標本で
なくて、標本あつてのP値だということでは切り抜けられないのではないかと思います。
標本が間違っていればP値も狂うし、切り出し方が間違っていればP値も当然それに影響さ
れて狂ってきます。

ですから何かわからないことがあつたら生のデータに戻るしかありません。それだけ戻れ
る状況をどこまで用意しておくかが、1つの問題となるでしょう。

松本: historical controlの利用の仕方なのですが、腫瘍発生率のレンジをとって、た
とえばデータがこの幅に入ればいいけど、入らなかつたら駄目だというような話があつた
のですが、そういう意味ではないと思うのです。実際には衛試にしても先生のところにし
ても、どういうhistorical dataの利用の仕方をなさっているのかをお聞きしたいのです
が。たとえばレンジでなければ、 $\text{平均値} \pm 2\text{SD}$ や 3SD を使つてみたり (表10参照)。

表10

Historical control (文献上のデータ・各施設固有のデータ):

自然発生病変の種類・頻度・頻度の実験間での変動幅

対照群が本当に陰性対照になっているか? (自然発生率の高い腫瘍の場合)

今までにこんな腫瘍を見た事があるか? (非常に希な腫瘍の場合)

菅野: そんな厳密なものでは決してありません。

松本: そうすると、大体レンジがこのくらいで、だからここから少しでも外ずれたら駄目
ということではなくて、大ざっぱなものでしょうか。

菅野：はい。子宮の内膜のポリープだったら普通は1割から4割の間となるのだということです。ただ私の施設ではこんなにはなりませんということであれば、それはその施設でとり直さなければいけないのです。年々明らかに下がってきているということだったら、それはなぜかということを考えなければいけないですけど、それはやはり大ざっぱなことになるのです。

吉村：レンジだと件数が多いときに幅がものすごく大きくなるから、多少標準化した方がいいと思います。

小林(安評センター)：Historical data との関係で問題となるのは低頻度の発生率の癌で、非常にまれに出る癌の検定が非常に難しいのです。私のほうで考えているのは、in vitro の試験の、変異原性試験で使うKastenbaumを低頻度で出る腫瘍に使ってもいいのではないかとということです。

大橋：頻度が小さいものと近似のテクニカルの問題が出てきます。 χ^2 などは絶対駄目だし、exact testをやる必要は必ずあるし、そのとき非常にまれなものだったら、historical controlというのは使えるのかどうか。それは多分危ない。計算上の問題、それは解決出来ていると思っています。 χ^2 近似はよくないです。exact testにならないと駄目です。

滝沢：ad hoc runを使うのは、腫瘍発生率は加齢とともに増加するはずだから、逆転している区間を一緒にするわけですが、それは一般的に他の統計でもやるのですか。例えば、ウィリアムズの検定でもそのようなことをしていましたが、それについては何も問題はないのでしょうか。

吉村：多分単調に増加するであろうと思われる中で、実際のデータがそうでないときに、逆転したところはプールして平均で置き換えるという操作をするということは、ある意味での合理性はもっているのですけれど、統計的な検定にかけようというときには、それを考慮に入れた棄却限界値を設定しないと第一種の過誤がものすごく狂うのです。

たとえばウィリアムズにしる、パーソロミューにしる、そういうことを統計的な検定として主張されている人は、第一種の過誤が公称有意水準と同じになるように必ず調整しているのです。

ところがPetoの場合そういうことをやっているかというのは非常に疑問です。だからPetoがこういうad hoc runでやるということが本当に合理的かどうかというのにはちょっと疑問があるのですが。

菅野：prevalence method でグラフを描くと、それで比較検討できる腫瘍に関してはかなりきれいにいくものですから、使ってしまえということだと思えるのですけれど。

吉村：結局、差があるかないかという種類の議論に使うわけです。するとそういうやり方をするのが第一種の過誤を大きくするのか小さくするのか、あるいは検出力を大きくする

のか小さくするのがよくわからないのです。

だから見かけはきれいになるのですが、何をやっているかわからないことがありますから、その辺の吟味が出来ているかはわからないのです。だから無条件にきれいになるからいいとは言えないという気はするのです。

菅野：中間屠殺とか、ある一時期ある群だけに肺炎が出て10匹まとまって死んだというときには、それを別に1区画設けなさいということはPetoも言っています。特殊なことが起こった場合は、無理してad hoc runにするのは危険だということを少しは考慮しているようです。

滝沢：ログランク検定はノーマルな健康で癌のないネズミも担癌のネズミも死ぬ確率は同じであることが大前提となっているのですが、この統計的検定でどうしてそれがわかるのですか。どこを見ると、どうしてそれが大前提になっているということがわかるのでしょうか。

吉村：それは多分手法を誘導するときに、他のことで死ぬ割合は同じであることを、モデルのどこかで前提にして入れ込んであるはずなのです。そして手法の正当化をしているはずなのです。

大橋：prevalenceの場合、センサリングが両群について同じであるという仮定がどこに入っているかという、ノンパラの分散を計算するところで、permutation distributionを使っているところです。帰無仮説のもとでの分布をpermutation distributionで仮定しています。permutation ということはセンサリングも同等であるという仮説です。

大橋：それがくずれたときに何が起きるかという困るのです。たとえば癌になる前に毒性でみんな死んでしまったらどうなるか。癌は全然出現しないでしょう。発癌性がものすごく強い化合物であっても毒性を抑制できないために早く死んでしまって情報が出てこない。だからセンサリングが同じということは、死ぬというのが投与群と対照群とで同じであれば発癌性が等しいかどうかの検定はできますが、極論をいうとそもそもセンサリングのパターンが全然違うようなときには、検定自体全くできない。つまり、片方はあるところでみんな屠殺してしまうというのと、片方はずっと経過を見るというやり方で見たら、それでprevalence法でも何でもやったとしてもナンセンスでしょう。

菅野：ログランク検定は剖検に落ちてくる機会が均等であるという大前提でしか計算してないわけです。そういう意味で剖検にかかる頻度まで変わってきた場合には、別のことをしなければいけない。具体的にどういう状況が想定されるかという、非常に難しいのですが、最初の用量設定の段階では全然毒性が出てこなかったものが、ある慢性毒性が途中で突然ある群にだけ出た場合にどうするかとか、またその出方も甲状腺に腫瘍があってT4が減少している動物だけを選択的に殺すようなことが起こるのかなどです。そういうことが観測された場合は、今の統計は使えないことになります。

奥村（ヒューマンライフ）：直感的には差があるのだけれど検定には出ないような場合はやはり差はないと言っていいのですか。

菅野：本当に直感が正しければ、そういうことはあまりないでしょう。そうだった場合は直感が正しいのか、データのとり方が正しいか、もう一回洗い直して見るべきだということでしょう。

奥村：そういうことは結構あるわけですか。

菅野：余り多くはないでしょうが、時々あります。

(1990年4月7日)

基礎講座

第2版

第2回 Analysis of 2 × 2 Tables

大橋 信之（安評センター）訳

表2にリウマチ性関節炎患者への無作為化、二重盲検法、偽薬対照群、臨床試験から得たデータを示す。患者は効果を3段階：効果なし、いくらかの効果、著しい効果、で評価した。重要な説明変数は処置（試験薬vs偽薬）、性、年齢である。

もし応答変数を、ひとつの重要な説明変数として、2分的変化および考慮する処置とすれば（試験薬／偽薬、改善あり／改善なし）、データは次の表のように示される。

Treatment	Improvement		Total
	None	Some or marked	
Test drug	$n_{11} = 13$	$n_{12} = 28$	$n_{1+} = 41$
Placebo	$n_{21} = 29$	$n_{22} = 14$	$n_{2+} = 43$
Total	$n_{+1} = 42$	$n_{+2} = 42$	$n = 84$

(1)

いく人かの患者はプロトコル違反のため、それぞれのグループから排除したけれども、行の周辺和（ n_{1+} 、 n_{2+} ）は処置の配分により定められた。列の周辺和（ n_{+1} 、 n_{+2} ）もまた患者への処置の違いの無い、帰無仮説 H_0 のもと定められた。 n_{+1} は効果なし、 n_{+2} はいくらかあるいは著しい効果の和を示す。全ての周辺和を計算し与えたとき、無作為化により意味する確率モデルは超幾何分布によって与えられる。すなわち

$$\Pr(n_{ij} | H_0) = \frac{n_{1+}! n_{2+}! n_{+1}! n_{+2}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!} \quad (2)$$

Table 2 Rheumatoid Arthritis Data

Test drug treatment				Placebo treatment			
Pt. #	Sex	Age	Improvement ^a	Pt. #	Sex	Age	Improvement ^a
57	M	27	1	9	M	37	0
46	M	29	0	14	M	44	0
77	M	30	0	73	M	50	0
17	M	32	2	74	M	51	0
36	M	46	2	25	M	52	0
23	M	58	2	18	M	53	0
75	M	59	0	21	M	59	0
39	M	59	2	52	M	59	0
33	M	63	0	45	M	62	0
55	M	63	0	41	M	62	0
30	M	64	0	8	M	63	2
5	M	64	1	80	F	23	0
63	M	69	0	12	F	30	0
83	M	70	2	29	F	30	0
66	F	23	0	50	F	31	1
40	F	32	0	38	F	32	0
6	F	37	1	35	F	33	2
7	F	41	0	51	F	37	0
72	F	41	2	54	F	44	0
37	F	48	0	76	F	45	0
82	F	48	2	16	F	46	0
53	F	55	2	69	F	48	0
79	F	55	2	31	F	49	0
26	F	56	2	20	F	51	0
28	F	57	2	68	F	53	0
60	F	57	2	81	F	54	0
22	F	57	2	4	F	54	0
27	F	58	0	78	F	54	2
2	F	59	2	70	F	55	2
59	F	59	2	49	F	57	0
62	F	60	2	10	F	57	1
84	F	61	2	47	F	58	1

Table 2 (continued)

Test drug treatment				Placebo treatment			
Pt. #	Sex	Age	Improvement ^a	Pt. #	Sex	Age	Improvement ^a
64	F	62	1	44	F	59	1
34	F	62	2	24	F	59	2
58	F	66	2	48	F	61	0
13	F	67	2	19	F	63	1
61	F	68	1	3	F	64	0
65	F	68	2	67	F	65	2
11	F	69	0	32	F	66	0
56	F	69	1	42	F	66	0
43	F	70	1	15	F	66	1
				71	F	68	1
				1	F	74	2

^aImprovement: 0, none; 1, some; 2, marked.

n_{11} の期待値は

$$E\{n_{11} | H_0\} = \frac{n_1 + n_{+1}}{n} = m_{11} = 20.50 \quad (3)$$

分散は

$$V\{n_{11} | H_0\} = \frac{n_1 + n_2 + n_{+1} + n_{+2}}{n^2(n-1)} = v_{11} = 5.31 \quad (4)$$

もし標本数が十分大きければ、 n_{11} は、おおよそ中心極限理論 (Hannan and Harkness, 1963; Puri and Sen, 1971; Plackett, 1981) による正規分布を示す。

$$Q = \frac{(n_{11} - m_{11})^2}{v_{11}} \quad (5)$$

また自由度 (d.f.) 1 のカイ 2 乗分布を持つ。 Q 値は計算に使われた 4 つのセルに依存しない。 なぜなら ij 番目のセルは $n_{ij} - m_{ij} = \pm (n_{11} - m_{11})$ および $v_{ij} = v_{11}$ 、 m_{ij} と v_{ij} は、 H_0 下の期待値および n_{ij} の分散である。 統計値 Q は、また

$$Q = \frac{\{(n_{1+}n_{2+}/n)(p_{11} - p_{21})\}^2}{v_{11}} \quad (6)$$

と示される。

$p_{i1} = (n_{i1}/n_{i+})$ は i 番目のグループにおける効果なしの患者の割合を示す。 この表現は Q 値の大きさが、改善のみられない患者の比率を処置間の違いで示している。 この意味で、 $(p_{11} - p_{21})$ は処置と反応の関連を表している。

無作為カイ 2 乗 Q と Pearson カイ 2 乗 Q_p の関係は

$$Q_p = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - m_{ij})^2}{m_{ij}} = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1+}n_{2+}n_{+1}n_{+2}} \quad (7)$$

2 x 2 表に関しては $Q = [(n - 1) / n] Q_p$ である。 このように大きい標本においては 2 つの統計値は、ほとんど等しい。 この例では $Q = 10.59$ ($p < 0.01$) と $Q_p = 10.72$ ($p < 0.01$)。

2 x 2 分割表のための、Q または Q_p のようなカイ 2 乗検定の適合のための一般的ルールは、期待値 $m_{ij} = n_{i+}n_{+j}/n$ が全てのセルにおいて 5.0 を越えるべきである。 この例においても正しい。 セルの数が小さい場合、カイ 2 乗検定より適切は方法はフィッシャーの正確検定である。 両側フィッシャー正確検定の p 値は、観察表としての同一の周辺和にともなう、全ての表に基づいている。(2) 式を用いて、それぞれの表の確率を計算し、そして観察表と同様または、小さい表の確率を合計する。 片側フィッシャー正確検定では、片側ひとつ (ここでは、試験薬のより好ましい反応の表) によって特徴づけられた観察表の強い確率表の合計である。 この例では、両側フィッシャー正確検定の p 値は 0.002 であり、片側フィッシャー正確検定では 0.001 である。 標本数が 2 つの処置群においてほぼ等しい場合は、この例のように両側検定の p 値は片側検定の p 値の約 2 倍となる。 標本数が全く等しければ、この近似は等しい。 ところが標本数が異なる場合は、フィッシャー正確検定は対称的ではなく、片側 p 値の 2 倍より両側 p 値は、著しく小さくなる。

いく人かの著者 (Harber, 1986; Overall and Starbuck, 1983; Salama et al., 1984) はフィッシャー正確検定より強力な正確検定の使用を主唱した。

しかしながら、他の検定は、それぞれの処置に2項分布を仮定しているが、フィッシャー正確検定は無作為化から直接の利点がある。この仮定の正当性は、それぞれグループの患者は、ある大きな母集団からの無作為標本と同等でなければならない。2項分布の仮定が正しいという状態に関して、Upton (1982) は、Qに関するカイ2乗の近似は、Qpより良好であると提案している。

連続修正は、無作為化法およびPearsonカイ2乗法、両方において結果が、より正確検定の結果と等しくなるように適用される。連続修正無作為化統計量は

$$Q_C = \frac{\{ |n_{11} - m_{11}| - 0.5 \}^2}{v_{11}} \quad (8)$$

連続修正Pearson統計量 (Yates, 1934) は

$$Q_{PC} = \sum_{i=1}^2 \sum_{j=1}^2 \frac{\{ |n_{ij} - m_{ij}| - 0.5 \}^2}{m_{ij}} \quad (9)$$

この例では、連続修正Pearson統計量は、 $Q_{pc} = 9.34$ ($p = 0.002$)、修正なしの統計量より、両側フィッシャー正確検定により近似した。

Fleiss (1981) は、Pearsonカイ2乗統計量は、つねに連続性に関して修正を勧めた。しかしこの点に関しては、いくつか論争がある (Grizzle, 1967; Conner, 1974)。この論争に関するひとつの知見は、評価がそれぞれの処置において無作為化による超幾何分布または2項分布に基づいているかによる。その他、非対称の場合は、連続修正カイ2乗統計量の片側p値は、片側フィッシャー正確検定のp値とよく一致する。しかし両側検定の値 (2倍) は、両側フィッシャー正確検定のp値より大きい値となる。

要約すれば、臨床試験における2x2表の解析に関して、フィッシャー正確検定は常に適用でき、コンピュータにより簡単に得られる。近似については (すなわち、全カイ2乗検定、連続修正ある、なし)、標本数が十分に大きくなければ (例えば全 $m_{ij} \geq 10$)、慎重に、特に非対称の場合において使われるべきである。

基礎講座 “Biopharmaceutical Statistics for Drug Development”

Chap. 9 : Clinical Efficacy Trials with Categorical Data

第2回 Analysis of 2 × 2 Tables

Q & A

Q : χ^2 統計量の片側 p 値(テキスト p.409, 1.27)とは具体的にどのように計算するのか?

A : χ^2 統計量の両側 p 値を半分にした値を、片側 p 値として解釈する。

Q : 2 × 2 分割表の解析を行なう場合、
確率モデルとして 2 項分布を適用すべきか、超幾何分布を適用すべきか、という判断はどのように行なうのか?

A : 2 × 2 分割表の確率モデルとして、
通常以下の 3 つのモデルを考える。

- 1) 多項分布 (4 項分布) モデル
- 2) 2 つの 2 項分布モデル
- 3) 超幾何分布モデル

あらかじめ定めた被験者 n 人を 2 つの属性 A, B について分類する場合などは多項分布モデルを考える場合である。2 つの 2 項分布モデルを考える場合としては、ケース・コントロール研究のように、疾病ありの人(case群)を n_{+1} 人、疾病なしの人(control群)を n_{+2} 人集めて、それぞれのグループで過去の曝露状況により B1, B2 に分類する場合などが考えられる。超幾何分布モデルを考える場合としては、あらかじめ疾病ありの人

	B 1	B 2	計
A 1	n_{11}	n_{12}	n_{1+}
A 2	n_{21}	n_{22}	n_{2+}
計	n_{+1}	n_{+2}	n

と疾病なしの人がそれぞれ何人いるとわかっていて、このような被験者を1つは検査値によりA1, A2に分類し、もう1つは主観的な判断でB1, B2に分けるといような場合が考えられる。テキストの χ^2 統計量Qによる検定は、3)の場合には理論的前提を満たした検定方法であるが、1)や2)の場合は必ずしもそうではない。しかしFisherの直接確率による方法は1)や2)の場合についても、周辺和を与えた場合の条件付き確率として計算できるので、3つの場合全てについて統一的な方法で求めることができる。

Q：一般の $s \times r$ の分割表でセルの度数が5以下になるような場合、解析方法はFisher流の直接確率を計算する方法と、 χ^2 統計量による近似的な方法のいずれを用いるか？

A：Fisherの直接確率を計算するのに、 2×2 分割表の場合には周辺和を固定すると動きうるのは1つのセルの度数だけなので、観察された事象より極端な場合というのを定義することは容易だが、 2×2 以上の分割表の場合には2以上の自由度となり、極端な場合というのを定義することは難しい。そのような場合、極端さの基準として例えばピアソンの χ^2 統計量を用いて評価し、p値の計算はFisher流に極端な場合を全て足し合わせるとい方法が考えられ、最近の統計パッケージでもこの方法が用いられている。

決定樹方式での統計手法の利用の注意点

熊本大学工学部 永田 靖

1. はじめに

次頁の図1は「毒性・薬効データの統計解析」（以後“ピンクの本”）の p.24 にのっている決定樹である。得られたデータに基づいて、いくつかの△を通る際に右下へ行くか左下へ行くか判断して一番下の目的とする検定へ至る道筋が示されている。ここで△のなかには「群の数は？」とか「因子の数は？」というようにデータから（または、データをとる前に）明かに判断のつくものと、「等分散か？」などのようにデータを用いて検定して判定するものがある。本稿では後者の△を考える。

目的とする検定（これを“本検定”とよぼう）の前段階で△において行う検定を“予備検定”とよぶ。

本稿での要点はただ一つ：

Question: “予備検定を行う際、本検定と同様に有意水準を5%として検定してもよいか？”

Answer: “よくない場合がしばしばあって注意が必要である。”

2. 平均値の差の検定

図1において○で囲んだ部分を例にとって詳しく考えていこう。

以下はある研究室での会話：

A氏: 「先輩、なにやってるんですか？」とB氏のデスクをのぞきこむ。

B氏: 「I薬とII薬に差があるかどうかみるためこのデータの解析をするねん。

I薬 x : 2.15, 1.70, 1.45, 1.88, 1.59, 1.99 ($m=6$)

II薬 y : 2.03, 2.04, 1.84, 1.93, 2.25, 2.17, 2.01

2.09, 2.12, 1.90, 1.76, 1.88, 1.97 ($n=13$)

このデータ解析の目的はI薬とII薬における母平均 μ_x 、 μ_y に

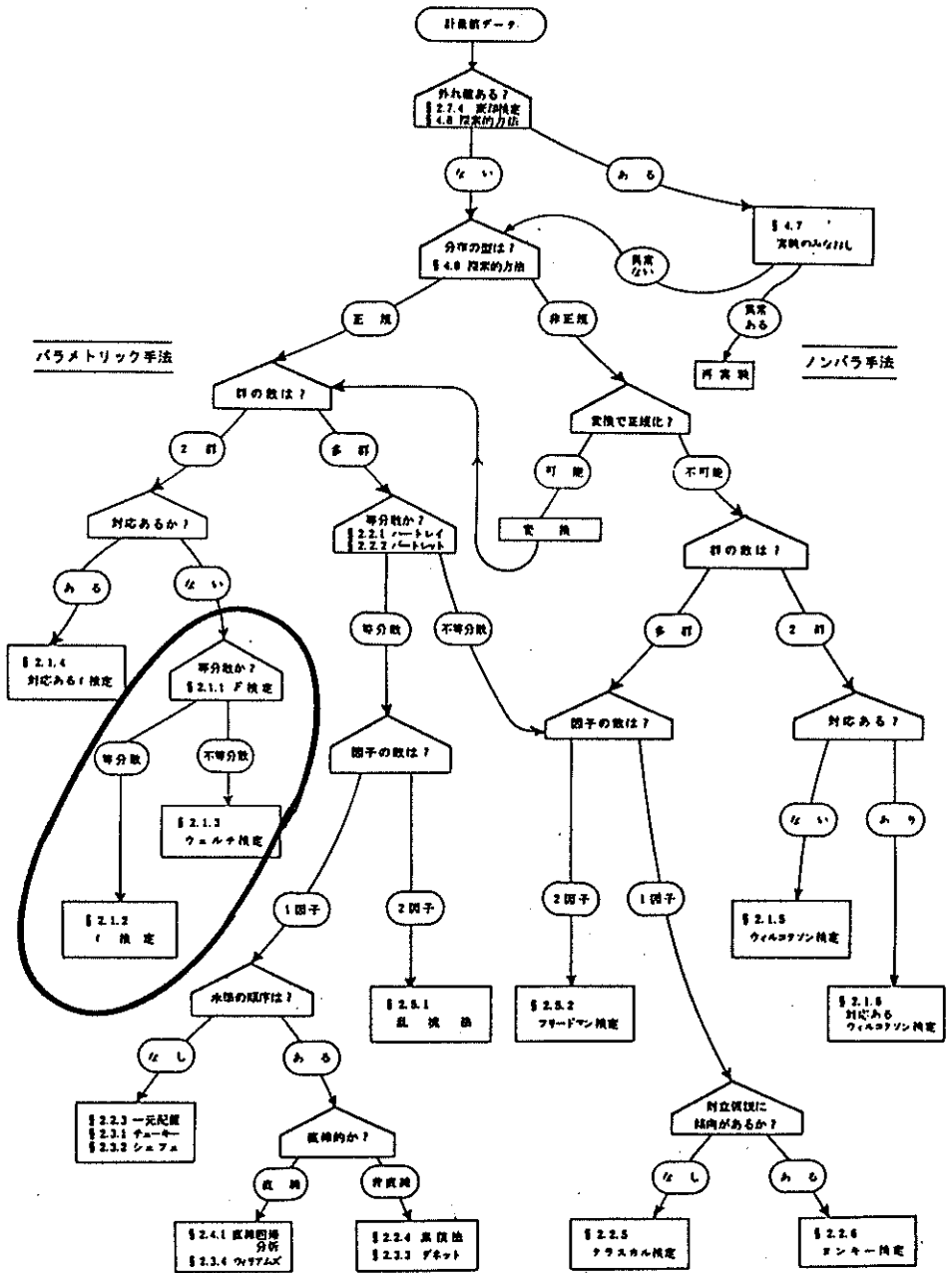


図1-6 計量値データの決定樹

図・1 (各分岐には他の手法も考えられるが、ここでは、本書でとりあげた手法のみを入れてある)。

ついて

$$H_0: \mu_x = \mu_y, \quad H_1: \mu_x \neq \mu_y$$

の検定やから、まず、ピンクの本の p.28 の「2群の分散比のF検定」をすればいいんや。

①等分散の検定

データから $S_x=0.3413$ 、 $V_x=S_x/(m-1)=0.06826$ 、 $S_y=0.2299$ 、 $V_y=S_y/(n-1)=0.01916$ となるから

$$F_0=V_x/V_y=3.563 < F(5, 12, 0.025)=3.891$$

となるわ。だから“有意水準5%で有意でない”となるんで等分散とみなして次にすすもっ！ピンクの本の p.30 の

②等分散とみなしたt検定

$$\bar{x}=1.793, \bar{y}=1.999, V=(S_x+S_y)/(m+n-2)=0.0336 \text{ より}$$

$$|t_0| = |\bar{x} - \bar{y}| / \sqrt{V(1/m+1/n)} = 2.277 > t(17, 0.025) = 2.110$$

($t(\cdot, 0.025)$ はt分布の上側2.5%点)となるから、有意水準5%で有意や。つまり μ_x と μ_y は異なることになるわ。いちようあがり！」

A氏: 「もし①で有意となったらどうするんですか？」

B氏: 「そんなときはな、ウェルチの検定をやったらいいねん。」

A氏: 「それ、どういうふうにするんですか？」

B氏: 「ピンクの本の p.32 のやつや。まねてやってみると

$$|t_0| = |\bar{x} - \bar{y}| / \sqrt{V_x/m+V_y/n} = 1.817$$

となるわ。自由度 ν の計算がちょっと厄介やな。

$$\nu = (V_x/m+V_y/n)^2 / ((V_x/m)^2/(m-1) + (V_y/n)^2/(n-1)) = 6.3$$

となって、少数自由度やから補間してみると

$$t(6.3, 0.025) = 0.7 \times t(6, 0.025) + 0.3 \times t(7, 0.025)$$

$$= 0.7 \times 2.447 + 0.3 \times 2.365 = 2.422$$

となるわ。だから

$$|t_0| = 1.817 < t(6.3, 0.025) = 2.422$$

となり、有意でない。つまり、 μ_x と μ_y とは異なるとはいえない
・・・」

A氏：「さっきのt検定のとくと結論が違いますね。」

B氏：「ほんまやなあ。計算間違いやろか・・・計算はあつとるなあ・・・」

A氏：「・・・」

B氏：「おまえが余計なこときくさかいに、わけわからんよーになったやないか」

A氏：「そんな・・・」と、他の参考書を調べまくる。「ああ、ここにこんなこと書いてありますよ。“ウェルチの検定はt検定より検出力が低い”・・・つまり有意になりにくいということじゃないですか。」

B氏：急に明るい顔になって「そうや、そうや、そうなんや。そやからウェルチの検定をやったら、さっき、有意にならんかったんや。おまえ、たまにはえーことゆーなあ。報告書には“有意差あり”って書いとこっ！」

A氏：B氏が報告書を書いている間、ピンクの本の p.28~p.34 に目を通す。「先輩はさっき①のF検定で、有意水準を5%で検定しましたよね。でも、ピンクの本の P.28 の [解説] iii)には、“有意水準は0.05より0.25くらいに大きくしたほうがよい”って書いてありますよ。だから①で $F_0=3.563$ と比べる値は P.28 の手順4から0.25の1/2の0.125を α にして $F(5,12,0.125)$ を調べると・・・」

B氏：「そんな上側12.5%点のF表なんてあらへんで。10%点やったらピンクの本の p.250 から $F(5,12,0.10)=2.394$ やけどな。」

A氏：「私のポケコン (SHARP PC-E500) なら求まります。えーと、・・・ $F(5,12,0.125)=2.178$ となります。」

B氏：「いっちょまえに道具だけはええもん持つとるな・・・。でもどっちにしても、 $F_0=3.563$ はこれらより大きいか。つまり、有意水準20%でも25%でも有意とゆーことやな。そしたら、分散を等しいとみなしたt検定やのうて、ウェルチの検定にすすまなあかんということか？」

A氏：「そうなりますよね。」

B氏：「そしたら、ウェルチの検定結果では有意じゃなかったから、報告書にも“有意じゃない”と書くんか？」

A氏：「そうですね。“有意じゃない”じゃなくて、“有意差がない”です、正しくは・・・」

B氏：「そやけど、検定するとき有意水準は5%か1%と相場は決っているで。なんで25%にせんとあかんのや。」

A氏：「先月、統計手法入門の講習会で検定を習ったとき、講師が“検定した後は、有意になったときのみアクションをとる”ってしきりに強調してました。いま先輩のやった解析では、①のF検定で5%有意にならなかったんで分散が等しいとみなして②へすすんでしょ。これは有意でないのにアクションをとったということにならないんですか。」

B氏：「そーいわれればそーやな。けど、5%有意やったら差があるとみなし、5%有意でなかったら等しいとみなす、なんてことはみんなやとんで一。」

A氏：「でも、このあいだの講師は α とか β とか言って図を書いて、“検定では有意になったときは『異なる』と断定できるが、有意にならなかったら『異なるとはいえない』という消極的な結論をせざるを得ない”って言っていました。」

B氏：「『異なるとはいえない』ってゆーのは“等しい”とゆーことやろ。」

A氏：「そうですか！それじゃあ、このあいだ、私が半年前にお貸した2万円を返して下さいと言ったとき、先輩は“返さんとはいわへんで”といわれましたけれど、返していただけるのですね！」

B氏：「いや、それはやっぱり違うわ。 α と β についてよう説明したろ。(図2を書きながら・・・) 真の状態、これは神様しか分からへんけど、 H_0 か H_1 や。このどっちかを判断するためデータをとって検定するんや。検定で有意となったら H_1 を採択するんやから、

		真の状態	
		H_0	H_1
検 定 に 判 よ 断 る	H_0	正	第2種の 誤り β
	H_1	第1種の 誤り α	正

図2

真の状態がもし H_0 なら検定結果は誤りとなって、これを第1種の誤りとゆーわけや。けど、その誤りの確率は α で、こいつは前もって有意水準として例えば5%と小さい値に設定してるから、あんまり問題ないわけや。そやから、有意なら断定してよいとなるんや。けど、検定で有意でないときは H_0 を一応採択するのやけど、もし真の状態が H_1 なら検定結果が誤りとなって、これを第2種の誤りとゆーて、この確率は β や。この β は0から $1 - \alpha$ くらいの値をとって、どんな値になるか分からへん。 $\alpha = 0.05$ やったら β は0.95くらいにもなるかもしらん。つまり、誤りの確率がかなり大きくなりうるわけや。そやから、有意でないときは H_0 を採択するゆーても断定できひんわけや。異なっている可能性も十分あるから“異なるとはいえない”とゆーいいかたをするんや。」

A氏：「だったら、“返さないとはいわない”ということは返してもらえない可能性も十分あるということですか？」

B氏：「論理的にはそーゆーことになるなー」

A氏：「・・・」

B氏：「そや、分かったで！」

A氏：「何がですか？」

B氏：「①の検定で有意水準を25%にするという理由や！この検定では有意となってもならなくても次の本検定にすまなあかん。そやから、①では有意となったときだけ断定できるような形にするのは不公平や。有意にならへんかったときのことも考えてやらなあかん。つまり、 β を小さくしたらなあかんのや。そのためには α を大きくせなしゃあない。 α を0.25とすると、 β は0~0.75くらいとなって、少しは小さくなりよる。そや、そーなんや。」

A氏：「それなら、 α を0.50とすると、 β は0~0.50くらいとなって、有意であるときと有意でないときとが公平になるんじゃないですか。」

B氏：「そーいわれるとそーやな。」

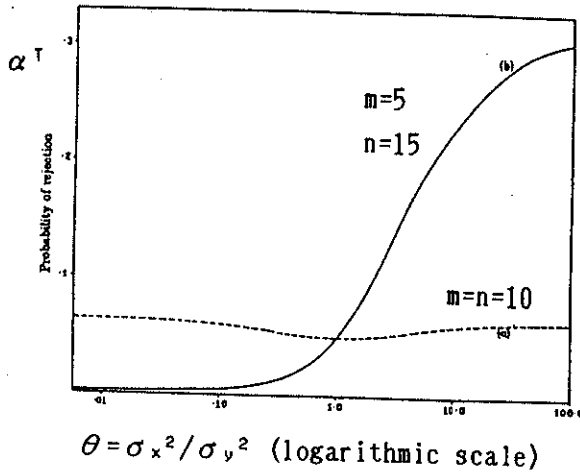
A氏：「それに、ウェルチの検定というのは分散が等しくないときに使うと書いてありますけど、分散が等しいときに使ってもいいんでしょ。検出力が悪いつてあったけど、t検定とそんなに違いがあるんですか？大きな違いがないのなら、①の等分散の検定なんかやらないで、最初からウェルチの検定をやればいいんじゃないですか？」

B氏：「・・・おまえ、なかなかスルドイなー」

3. 2節の解説

t検定とウェルチの検定について整理しよう。以下では α_0 、 α 、 α^T を区別して用いる。その区別を頭に入れながら読まないで混乱するから注意すること。

t検定は $\sigma_x^2 = \sigma_y^2$ を前提として構成されている。これを $\sigma_x^2 \neq \sigma_y^2$ のとき用いるとどうなるかということ、 $m \neq n$ なら $\sigma_x^2 \neq \sigma_y^2$ であってもほとんど問題ないが、 m と n とがかなり異なると正しい検定とはならない、ということが知られている。図3 (Welch(1937)より引用)を見てみよう。横軸は $\theta = \sigma_x^2 / \sigma_y^2$ 、縦軸は“本当の有意水準 α^T ”である。ここで、“本当の有意水準 α^T ”とは何かを理解することがとても重要である。我



図・3

々は有意水準を $\alpha=0.05$ と設定して数値表から $t(\cdot, 0.025)$ を求めるが、この値は t 検定の場合 $\sigma_x^2 = \sigma_y^2$ を前提にして “ $H_0: \mu_x = \mu_y$ が正しいのにこれを誤って棄却する確率 α^T ” が $\alpha=0.05$ と一致するように計算されたものである。したがって、 $\sigma_x^2 \neq \sigma_y^2$ ならこの α^T は $\alpha=0.05$ に一致するとは限らない。 α^T を計算するときの分布の形が変わってしまうからである。この α^T のことを “本当の有意水準” とよぶ。検定が正しいものであるためには $\alpha^T = \alpha = 0.05$ とならないといけない。図3より (a) $m=n=10$ のときは $\theta = \sigma_x^2 / \sigma_y^2$ の値が何であつても $\alpha^T = \alpha = 0.05$ である。一方、(b) $m=5, n=15$ のときは θ の値によつて α^T は大きく変化する。 $\theta=1$ のとき (つまり $\sigma_x^2 = \sigma_y^2$ のとき) は $\alpha^T = 0.05$ だが、 $\theta \ll 1$ または $\theta \gg 1$ のときは 0.05 と大きく食い違ふ。だから、~~~~部^{~~~~}に述べたことがいえる。($m \neq n$ なら t 検定は分散の違いによる影響をほとんど受けないから、 “ $m \neq n$ なら t 検定は分散の違いに対してロバスト (頑健) である” などといったりする。ピンクの本の p.31 の [解説] i) にはこのことが述べられている。(正規性についても同様の考え方でよい。)

つぎに、ウェルチの検定について述べよう。これは $\sigma_x^2 \neq \sigma_y^2$ であつても $\alpha^T = \alpha = 0.05$ となるように構成されているから、 σ_x^2, σ_y^2 がどうなつていても用いて構わない。($\sigma_x^2 = \sigma_y^2$ であつてももちろんよい。) ところが t 検定と比べると、 m と n とがかなり異なるときは検出力が若

干劣るという性質がある。ただし、t 検定が正しい検定手法であるのは（特に、m と n とがかなり異なるときは）、 $\sigma_x^2 = \sigma_y^2$ の場合に限られるのだから、この両者の比較は等分散を前提にしての話である。（この点については参考文献の間 17 を参照せよ。）

以上のことをまとめると図 4（t 検定を“t”、ウェルチの検定を“w”で表す）となる。図 4 より、どの場合でもウェルチの検定を用い

	$\sigma_x^2 = \sigma_y^2$ のとき	$\sigma_x^2 \neq \sigma_y^2$ のとき
m=n のとき	t と w : 引き分け 両者 : 正	t と w : 引き分け 両者 : 正
m≠n のとき	t の勝ち w でも正	w : 正 t : 誤り

図 4

てよい。しかし、t 検定が優る場合に目を向けて、 $H_0 : \sigma_x^2 = \sigma_y^2$ かどうかを①で予備検定して、 H_0 が採択されるなら少しでもよいもの（t 検定）を用いようという発想で 2 節の手順が出来上がった。果してこれが本当によりよい手法なのであろうか。①で予備検定することより、 σ_x^2 と σ_y^2 とがかなり異なるときはそれを検出できるから、図 3 で $\theta = \sigma_x^2 / \sigma_y^2 = 100$ などのときに t 検定を用いて本当の有意水準 α^T が 0.3 くらいになってしまうということはなくなるだろう。では、2 節の手順に基づいて、 α^T を求めたものをみてみよう。 α^T は $H_0 : \mu_x = \mu_y$ が成り立っているのにこれを誤って棄却してしまう確率だったから、 H_0 の下で

$$\alpha^T = \text{Pr} \{ \text{“①の F 検定で有意でない” かつ “t 検定で有意”} \} + \text{Pr} \{ \text{“①の F 検定で有意” かつ “ウェルチの検定で有意”} \}$$

である。Gans(1981) は①の F 検定の有意水準 α_0 をいろいろとって、シ

表・1 平均値の差の検定の α^T

(m, n)	α_0	$k = \sigma_1/\sigma_2$						
		0.25	0.50	0.67	1.00	1.50	2.00	4.00
(5, 5)	0	0.06	0.06	0.05	0.05	0.05	0.06	0.07
	0.05	0.07	0.05	0.06	0.04	0.05	0.05	0.05
	0.20	0.06	0.05	0.05	0.06	0.04	0.04	0.06
	1.00	0.06	0.04	0.04	0.05	0.04	0.06	0.05
(5, 10)	0	0.02	0.02	0.03	0.06	0.09	0.11	0.16
	0.05	0.03	0.03	0.04	0.05	0.08	0.09	0.07
	0.20	0.05	0.04	0.03	0.05	0.06	0.07	0.04
	1.00	0.05	0.04	0.05	0.06	0.05	0.06	0.05
(5, 20)	0	0.00	0.00	0.01	0.05	0.12	0.18	0.32
	0.05	0.04	0.03	0.03	0.07	0.12	0.11	0.07
	0.20	0.05	0.04	0.04	0.05	0.08	0.08	0.06
	1.00	0.04	0.05	0.05	0.07	0.07	0.07	0.06
(10, 10)	0	0.07	0.06	0.06	0.05	0.05	0.07	0.08
	0.05	0.03	0.05	0.05	0.05	0.04	0.05	0.04
	0.20	0.05	0.05	0.04	0.04	0.04	0.06	0.05
	1.00	0.05	0.05	0.04	0.05	0.05	0.05	0.05

ユミレーションにより α^T を求めている。その一部を表1に紹介する。
 $\alpha_0=0$ なら①のF検定で有意になることはないののでいつもt検定を用いた場合となる。このときは図3と同様の結果になっている。(m=n なら $\alpha^T=0.05$ だが、m≠n のときは α^T は 0.05 からずれる。) $\alpha_0=1$ なら①のF検定で常に有意なので、いつもウェルチの検定を用いることに対応し、このときはいつも $\alpha^T = \alpha = 0.05$ である。 $\alpha_0=0.05$ とすると、これは2節でB氏が最初に行った解析法に対応し、(m,n)=(5,20) のときをみると α^T は 0.05 より結構ずれている。だから、m≠n なら①のF検定で $\alpha_0=0.05$ とした手法はよくない。 $\alpha_0=0.20$ ならそのずれは少なくなっている。

以上より平均値の差の検定においては、m≠n ならどうやろうと問題はない。また、“ウェルチの検定1本でやれ” といってもよい。

一方、決定樹の②において等分散性についての予備検定を行うなら $\alpha_0=0.05$ を用いると、 $H_1: \sigma_x^2 \neq \sigma_y^2$ であるのにこれを検出できずに $H_0: \sigma_x^2 = \sigma_y^2$ と思ってt検定へと至る可能性が高く、m≠n のときは

この影響がt検定において大きく現れるので問題となる。 α_0 を大きくするとこの影響は及びにくくなってそこそこ無難となるが、必ずしもよりよい手法となるわけではない。このように個々の検定を別個に考えれば、それぞれの前提が成り立った下では正しい手法であっても、その前提をチェックして手法を選びながら解析するという流れ全体で考えると割り切ったものとはならない。

(補注) 2節においてウェルチの検定では有意でなく、t検定では有意であったが、その理由を表1から考えてみよう。2節の例では、 $(m, n) = (6, 13)$ 、 $k = \hat{\sigma}_x / \hat{\sigma}_y = \sqrt{V_x / V_y} = 1.89$ だったから表1の $(m, n) = (5, 10)$ 、 $k = 2$ の欄を見ることにすると、対応する α^T は、 $\alpha_0 = 0$ のとき $\alpha^T = 0.11$ 、 $\alpha_0 = 1$ のとき $\alpha^T = 0.06$ となっている。このことより、t検定の方が検出力が高いというのではなく、単に α^T が大きくて誤って有意になり易いだけだったのかもしれない・・・

4. 一般論

3節の内容を少し一般的に考えよう。それは次の図5にそった形となる。ここで本検定①と②とは同じ仮説を検定するための2種類の手法

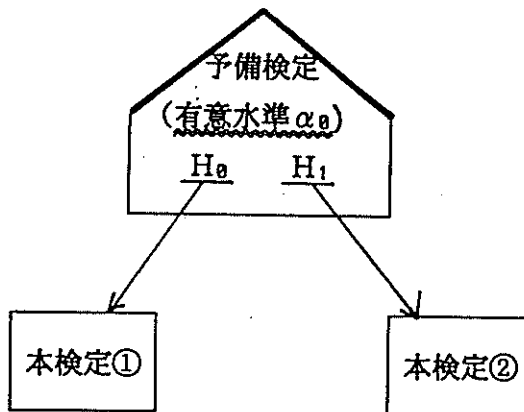


図5

(例えば、t検定とウェルチの検定)であるが、一般に次のようになっている。(以下で H_0 と H_1 とは予備検定における仮説である。)

本検定①： H_0 の下で正しい手法で②よりよい。

本検定②： H_0 , H_1 のどちらの下でも正しい手法。

このとき、本検定①が H_1 の下でロバストであるならあまり問題はないが（例えば3節で $m=n$ のときは $H_1: \sigma_x^2 \neq \sigma_y^2$ の下でも検定はロバストだった）、ロバストでないなら $\alpha_0=0.05$ だと問題である。

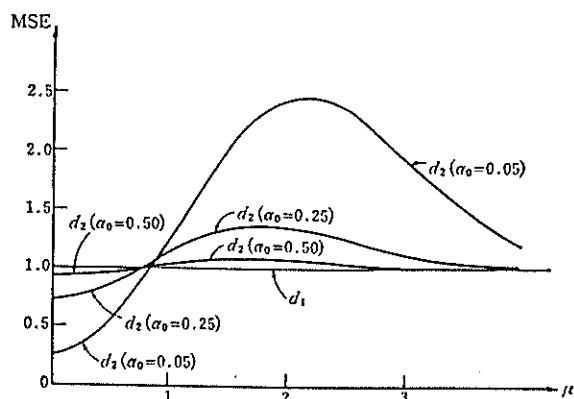
このような決定樹（の一部）はいろいろなところででてくる。分散分析では分散分析表を書いたあと、ある要因を無視して誤差項にプールするかどうか検討することがよくあるが、これは図5に全く対応する。

（詳しくは参考文献[4]を見られたい。）また、“データが正規分布にしたがっている”かどうか検討し、正規分布に基づく手法かノンパラの手法を用いるかに分かれる場合も図5に対応する。この場合、 α^T がどうなるか述べている文献については知らない。（正規分布に基づく手法のロバスト性やノンパラの手法の効率についての研究からそこそこ類推できると思えるが・・・）

いずれにしても目的が検定の場合は、図5の手順が単独の①または②よりよい手法になっているかどうかの判断が難しいと思われる。それなら②を常に用いればとなるのだが、図5にそった手順を用いたい場合もしばしばある。このことを例示するため推定について考えてみよう。図5で“本検定”を“推定”におきかえる。話を最も単純にするため、 $\bar{x} \sim N(\mu, 1^2)$ として μ の点推定を議論する。通常は $d_1 = \bar{x}$ で推定するが、図5にそって予備検定での仮説を $H_0: \mu=0$, $H_1: \mu \neq 0$ とし、 H_0 の下では $\hat{\mu}=0$ と推定し、 H_1 の下では $\hat{\mu}=\bar{x}$ と推定するものとする。 H_0 , H_1 の検定は、 $|\bar{x}| \geq u(\alpha_0/2)$ ($N(0, 1^2)$ の上側 $\alpha_0/2$ 点) なら有意水準 α_0 で H_0 を棄却するというものだから、ここで述べた推定方式は

$$d_2 = \begin{cases} 0 & (|\bar{x}| < u(\alpha_0/2) \text{ のとき}) \\ \bar{x} & (|\bar{x}| \geq u(\alpha_0/2) \text{ のとき}) \end{cases}$$

となる。点推定のよさの規準として平均2乗誤差 (MSE) $E(d - \mu)^2$ を用いることにすると、これは μ により変化し、その様子は図6のようになる。（ $\mu < 0$ のときは $\mu > 0$ と $\mu=0$ について対称となるから省略。） $d_1 = \bar{x}$ のMSEは μ によらず1であり、 d_2 のMSEがこれより下にあれ



図・6 推定量 d_1, d_2 の MSE 曲線

ばよい推定、上にあれば悪い推定である。 $\alpha_0=0.05$ とすると μ が 0 に近いとよいが、 μ が 2~3 くらいだとよくない。 α_0 を大きくすると d_1 の MSE に近づく。

図6をみただけでは α_0 のどの値がよいとはなんともいえない。 μ が 0 に近い確信があれば $\alpha_0=0.05$ とすればよいし、その確信がなければ $\alpha_0=0.05$ とするのは危険だからである。次に、この話を重回帰分析に一気に飛躍させてみる。モデル

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e = \eta + e$$

において、 η を推定したいとしよう。ここで説明変数 x_1 を落とせるかどうか検討することにする。そのためには $H_0: \beta_1=0$ 、 $H_1: \beta_1 \neq 0$ を有意水準 α_0 で t 検定すればよいから、 η の推定方式は

$$\hat{\eta} = \begin{cases} \hat{\beta}_0 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p & (\text{t 検定で有意でないとき}) \\ \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p & (\text{t 検定で有意のとき}) \end{cases}$$

となり、先の簡単な例と全く対応する。MSEの様子も図6に同様と考えてよい。さて、いまの場合は α_0 の選び方に少し説明がつく。基本的にモデルは簡単な方がよいと考えられるから、落としてもあまり推定精度に影響がない変数なら落としてしまえばよい。その影響の程度を図6でみると、 $\alpha_0=0.25 \sim 0.50$ くらいが適当と考えられる。このように説明変数を落としてよいかどうかという検討はモデル選択の問題であり、これは様々な規準、様々な観点から研究されている。そしてその内容は

“ α_0 をどう選べばよいか”にほぼ集約される。規準が違えばその規準の下での最適な α_0 の値も異なるが、たいていの場合 $\alpha_0 \gg 0.05$ である。

統計解析はデータから出発し、どんな形であるにしろモデル選択を行ってモデルを同定し、解析結果へ至るものである。そして、解析過程である決定樹において \square を通るときはモデルの同定を行っていると考えべきであり、 $\alpha_0=0.05$ を用いるのは不適當な場合がしばしばある。

参考文献

- [1] Welch, B. L. (1937). The significance of difference between two means when the population variances are unequal. *Biometrika*, 29, 350-362.
- [2] 近藤良夫、安藤貞一（編）(1967). 統計的方法百問百答, 日科技連出版社
- [3] Gans, D. J. (1981). Use of a preliminary test in comparing two sample means. *Commun. Statist.*, B, 10, 163-174.
- [4] 永田靖 (1988). 実験計画法をめぐる諸問題 —プーリング、逐次検定—, *品質*, 18, 196-204.

中毒情報活動とデータ処理

(財)日本中毒情報センター 新谷茂

今日、私達の身の廻りには、おびただしい数の化学物質や化学製品が使われていて、人々の生活を豊にしているが、その反面それらによる急性中毒例が増えつつある。厚生省統計情報部の人口動態統計によれば、毎年約6,000名の化学物質による不慮の事故死がみられている。それは交通事故死の約60%にせまっている。

日本中毒情報センターの事業活動は、電話を媒体として種々の急性中毒に関する問い合わせに対して、適切な対応方法や処置のための情報を提供することである。

当センターは活動を開始して3年が経過したばかりで、まだその経験は諸外国の30数年に比べると僅かである。

今回は現在までに得た活動結果のうちの受信状況に関する地検について紹介する。

I. データ処理のための前提条件

1) 事業内容: 下記の事業のうちの問い合わせ応答に関する要因解析結果に限定。

1. 中毒情報の問い合わせに対する回答
2. 中毒防止に関する講演会の開催等の啓蒙教育活動
3. 中毒情報に関する資料の収集と整備
4. 中毒症例の収集と解析、中毒に関する統計の作成
5. 国内外の毒性情報関連機関との連絡調整
6. 中毒に関する教育、研究の支援

2) 日本中毒情報センターの取り扱う対象:

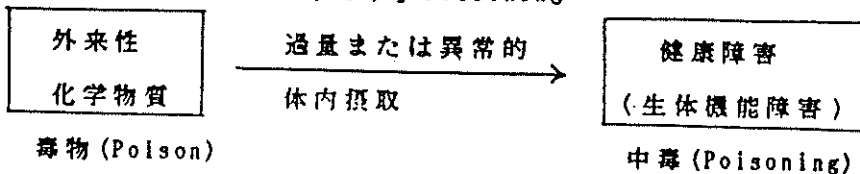
過量または異常摂取による化学性中毒(含む自然毒)

(除外例) 細菌性食中毒(伝染性、非伝染性、毒素型)

連用慢性中毒(麻薬、ひ素、薬物など)

妊娠中毒

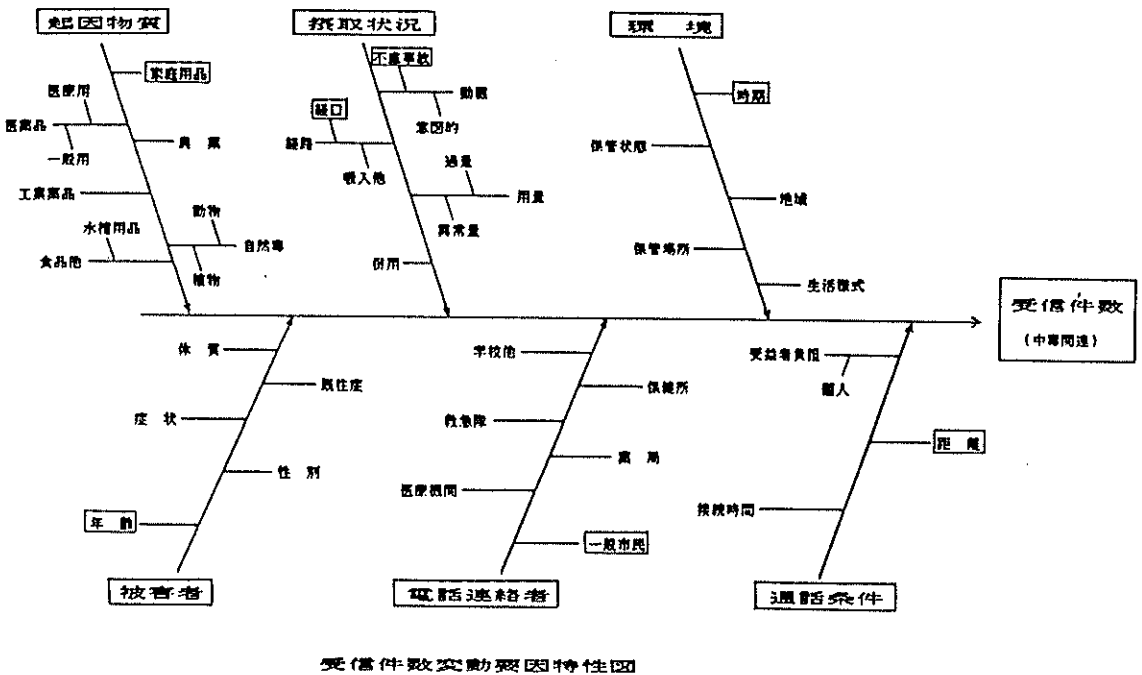
(参考) 中毒の定義: 「毒あたり」Poisoning



- 3) 統計量: ① 受付登録データ a. 総問い合わせ件数 (応答対象外も含む)
 b. 受信件数 (中毒関連) -- 要因別構成比
 c. 要受診件数
 d. 受信時無症状件数 (受信時無症率)
 e. 該当資料整備件数 (回答か率)

② 追跡症例データ: 症例数、摂取量、搬入時間、転帰成績、
 入院 (死亡) 日数

4) 受信件数変動要因



II. 要因解析結果: 以下の各種統計成績について紹介する。

- 1) 総問い合わせ件数: 1986.9-1989.9、107,780件
- 2) 受信件数: ① 起因物質 (大分類) 別件数
 ② 電話連絡者別都道府県別件数
 ③ 起因物質 (大分類) 別月間件数
 ④ 起因物質別被害者年齢分布
 ⑤ 起因物質別摂取経路別件数
 ⑥ 事故発生動機別件数

3) その他: ①起因物質別電話連絡者別受信時無症状件数

②起因物質別要受診件数

4) タバコに関する要因分析結果:

年齢別、月間推移、曜日別変化、時刻別変化、用量-反応関係

Ⅲ. 血中濃度と中毒症状の関係: アミノフェン、イブプロフェン

Ⅳ. 知見総括: 高頻度受信件数要因を挙げると、次のようになる。

1) 一般市民

2) 家庭用品、特にタバコ、家庭用殺虫剤(蚊取りマット)

3) 関東、近畿、中部地方

4) 乳幼児(5歳未満)

5) 経口摂取

6) 不慮の事故

参考文献

1) 新谷 茂、吉岡敏治、後藤京子、石沢淳子、山田容子、磯村千鶴: 中毒研究
2、305-319(1989)

2) Smolinske SC, Spoerke DG, Spiller SK, Wruk KM, Kulig K, Rumack BH:
Human Toxicol., 7, 27-31 (1988)

3) Volans GN et al (NPIS Monitoring Group): 1, 7-23 (1981)

4) Poisindex: 1988, 11

5) Litovitz TL, Schmitz BF, Matyunas N, Martin TG: Am J Emerg Med 6, 479-
515 (1988)

薬の安全性をめぐる統計的問題

竹内 啓 (東京 大学)

薬の安全性をめぐる統計的問題については具体的な事例を扱ったことはないので、実感を持ってお話することはできない。統計家の観点から、いくつかの方法論的問題を指摘したい。

かつてFinneyが来日したとき強調したことは「薬の害については、予め計画した実験はできない」ということであった。すなわち直接的なデータはすべて事後的観測によらざるを得ない後向きデータである。そして後向きデータについてはつねに因果関係の解決に疑問が生ずる余地がある。

その上薬害に関するデータについては、なお通常の疫学的データ以上の困難がともなう。第一に標本数は一般に余り大きくなく、場合によれば極めて小さい。第二に観察対象は多くの場合病人であって、安定した母集団がその標本を表しているとはいえない。

このような事情から、薬の安全性に関するデータについて、厳密な統計的方法を適用するには一般に大きな限界があるといわねばならない。

しかし一方薬にかかわるデータは、一般の疫学的データより有利な面もある。それは対象を精密に観察し、あるいは少なくとも事後には十分管理することが可能であるという点である。したがって、病理学的等の実体的な科学の論理に基づく分析が可能になる。厳密な結論はそれによってしか得られないであろう。

もう一つの方法は動物実験の利用である。ただしそれについては常に「外挿の問題」が伴うので、その結果の人間への応用に当たっては、実体統計的な十分な基礎づけが必要である。

今後の発展の方向としては、病理学等に基づく観察の結果を定量化するとともに、症状の進行過程のダイナミックな過程をモデル化して、単に結果の二元分類(有無、

生死等)、あるいは少数の目盛りからなる順序尺度(+++, ++, +, 0, -, 等)による分類でなく、より多くの情報を含んだデータを対象とする統計的分析の方法を開発する必要がある。これは薬の安全問題に限らず、臨床医学全般についていえることである。その際一方では臨床診断の規格化、統計学的には非定常系列ないし非定常確率過程のモデルとその解析法を確立する必要がある。統計学的に重要な課題は、第一に安定的な集団における平均や標準偏差、相関係数、あるいは定常時系列データに対するスペクトルのようなデータの性質を適切に表現する統計量を作り出すことである。

しかし当面の現実的な課題としては、精密複雑な数学的モデル、統計的方法を開発するよりも、臨床医学の多くの現場、すなわち病院や診療所での毎日の医療行為の中から得られている膨大な情報を適切な形で集め、整理して蓄積すること、つまりデータベース化することであると思う。

そのためには多くの制度的障害を克服しなければならないと思われるが、技術的にもデータの記述の標準化、規格化、コード化、そしてその表示や、場合によってはプライバシー保護のための手段など、多くの研究すべき課題がふくまれていると思われる。その中で統計家が協力できる分野も少なくないと思われる。また逆に統計家が臨床家から学ぶべきことも少なくないであろう。

数学的手法の開発や普及よりも、統計家と臨床家とのより深い相互理解のための討論が必要である。

山之内製薬臨床統計部

魚井 徹

1989/10/21

医薬安全性研究会

(1) 薬効評価の歴史

製薬企業における臨床統計業務の変遷は、種々の出来事を契機とした医薬品に関する社会的認識の変化および、薬効評価法の発達と密接に関わっている。以下には、薬効評価に関係の深い出来事を概観する。

1937年、ジエチレングリコールを使用したサルファ剤のエリキシルにより、107人の死亡者が出、後に発明者が自殺するという事件が発生した。翌38年には、食品薬品化粧品法(FDC Act)が米国において発効し、医薬品は安全であるべきことが示された。1948年には、ニュールンベルクの10ヶ条および、世界医師会のジュネーブ宣言が出され、人試験の基本的な倫理基準が示された。1961年11月18日、サリドマイドの奇形に関してレンツ報告がなされた。1962年には、米国で、KEFAUVER-HARRIS修正法が成立し、有効性の確認に関する基本的な考え方が示された。この法律は、米国内のみならず、その後の全世界の医薬品開発法の基礎となっている。1963年には、WHOの第16会総会において、医薬品の安全性に関する決議が行われ、以後の副作用モニター制度に道が開かれた。1964年には、第18回世界医師会総会において、ヘルシンキ宣言が出され、現在の倫理要綱の基礎が作られた。また、この年、米国において、再評価が指示されている。1977年より、米国においてGCP関連の法律が順次定められている。

日本においては、1961年に国民皆保険制度がスタートした。新薬の製造承認が基本的に現在の基準となったのは、1967年の医薬品製造承認等に関する基本方針が厚生省によって示されたときである。1970年には、スモンの第1号提訴がなされている。1971年より、日本においても医薬品の再評価が始まった。1979年には、薬事法が改正され、はじめて治験薬が法的に認められることとなった。1983年、中央薬事審議会委員の大幅入れ替えが行われ、その後、承認申請資料の統計学的検討がより厳密に行われることになる。1985年には、日本のGCP案が示され、1989年最終版が出された。

これらの事項のうち、国内の臨床統計業務に特に関わりの深いものは、KEFAUVER-HARRIS修正法、基本方針、中央薬事審議会委員入れ替え、GCPであろう。その

ほか、1985年より開始されたMOSS協議と関連した、国際ハーモナイゼーションの影響が大きい。現在、GCPを補うマニュアル、一般ガイドライン、統計ガイドラインが順次公表される予定であり、これらの影響も大きいと考えられる。

(2) 企業における臨床統計部門の成立

基本方針が出されるまでには、臨床統計に関する部門は、1社のみであったと考えられる。基本方針により、比較試験さらには、二重盲検比較試験が必要とされ、データの統計解析の必要性から、いくつかの会社で担当部門または、担当者が定められた。その後、外資系の企業において、本国との水準統一・データ交換の必要性を中心に制度化が進められてきた。さらに、審査における統計学的議論の増加を契機として各社に組織が作成され、現在では、GCP、一般および統計ガイドラインに対応するために組織化が進められている会社も少なくない。

(3) 臨床統計に関する研究会

臨床統計の必要性を示すものとして、現在行われている研究会等を挙げる。日本科学技術連盟の研究会は、1970年より開催されており、1972年より開始されたセミナーには年100名程度が出席者している。製薬協では1983年に統計の分科会が結成され、現在は約50社の参加の下に、最新の事項でかつ、実的な問題について研究している。その他、MIA、関西SDA研究会、KANSAS、医学統計集談会、KR研究会、原宿ミーティング、SASの各種研究会、薬効評価研究会、等の各研究会がある。

定常的な研究会のほかに、長期臨床試験シンポジウム、KICADIS、リウマチの薬効評価に関するシンポジウム、バイオメトリックスおよびサテライトシンポジウム、統計学会の薬効評価に関連する特別セッション、生存科学シンポジウム並びにそのサテライトシンポジウムその他の会合も多く開催されてきており、その都度話題を提供してきた。

(4) 臨床統計業務の公共性。

臨床統計業務が盛んになるに連れて、その公共性が徐々に認識されてきている。臨床統計業務の公共性は、2面より成り立っている。一つは、統計学の公共性であり、他方は、医薬品の公共性である。統計学は、直接決定の枠組みを採らなくても、決定に深く関与しており、統計解析および、統計学に関するあらゆる議論は、その社会的影響から逃れることはできない。他方、医薬品は人々の健康、生命に関わるために、正確な判断並びに情報提供が求められている。さらに、新薬は製造承認を得てはじめて商品となることから、行

政並びに企業へのインパクトが大きい。また、保険制度に基づく経済的な側面も有する。

(5) 臨床統計業務の広がり

現在の臨床統計業務は、単にデータ解析のみならず、試験（調査）の計画の技術としての試験の計画・調査表の設計、試験実施の方法論、試験全般にわたる品質管理、データクリーニング、入力・入力の確認、データ加工、解析、解析報告、論文の吟味およびこれらの業務の設計などに関係している。これらの業務を品質保証の下に効率よく実施するためには、専門的知識・ノウハウが非常に必要とされ、業務の専門化が必要である。業務を専門化することにより、上記の公共性の条件を満たすことが容易になると考えられる。

(6) 臨床統計業務の今後の課題

現在でも、統計学を事後的なデータ解析の技術であるとする者は多い。今後、試験計画および最適な情報の抽出といった考え方を推し進めることにより、正確な薬効評価を目指すべきであろう。調査項目の設計、作成など、調査技法の開発応用の推進も期待される。また、新薬開発における戦略的な利用法の開発も必要となろう。データ処理に関しては、単に解析するだけでなく、O A化に合わせて解析と文書処理の一体運用が望まれるが、そのためには、データベース設計、文書標準型式の研究、統計情報のメタデータベース化等の技術開発が必要である。さらに、A I技術を応用し、操作的な業務を合理化し、創造的業務の支援をすることも必要である。G C P施行下においては、データの入力品質のみならず、元データの品質測定と品質保証のための工程の設計などの技術も必要となる。解析手法としては、多重の推論に対する統計学的調整の問題、時系列測定の問題、順序分類データ解析の問題、同等性検証の問題等がある。

情報化が進むにつれて、特にプライバシー、情報の利用権限等に関しても積極的に取り組む必要がある。情報の共有とともに、画一的なデータ処理、解析を指向する傾向が見られるが、情報は資産であることを念頭において、その利用権限に配慮することも必要である。今後、臨床統計の枠内にとどまらず、諸規範の充足・会社と社会バランスの上での計画、決定、情報利用の業務として考える必要がある。公開の討論を前提とし、適切な競争原理の下に健全な発達が望まれる。

変異原性試験は化学物質の安全性を評価するための試験法の一つとして用いられている。厚生省の医薬品の安全性評価のためのガイドライン、OECDをはじめ各国のガイドラインでも変異原性試験の実施が義務づけられている。現在は、どのような変化を指標とするか、また何を実験材料として試験するかによって非常に多くの試験法が考案され、かつ実施されている。それらのうちで最も一般的に用いられているのが、サルモネラ菌を用いて遺伝子の復帰突然変異を検出する系（エイムス試験）、培養細胞を用いて染色体異常を検出する系、マウスの骨髄を用い染色体異常に起因する小核の出現頻度をみる系（小核試験）である。

それぞれのガイドライン（または解説書）にはこれら試験法の解説をはじめ具体的な実験手技等が記載されているが、実験結果の評価に関しては“適切な統計学的手法を用いて評価すること”等の表現があるのみで具体的な記述はなされていない。現状はそれぞれの試験法が開発された当初に用いられた方法が伝統的に踏襲されており、改良、改変はほとんどなされていない。近年、統計の専門家による評価手法の論文がかなりの数報告されているが、実際に変異原性試験を実施している者にとって、それらを読みこなす自分達のデータに適用するには難しすぎるようである。このように生物、毒性学分野と統計学分野のギャップはまだまだ大きいようであり、お互いにもっと協力しあう必要がある。

ギャップの原因が、それぞれの分野の研究者が互いに相手の分野を充分理解していない（専門家になる必要はない）点にあることは異論の無いことと思う。生物分野側の問題点の1つとしては、実験

を終了してから結果の判定，評価に困って初めて統計学者に救いを求める点にあると考えられる。救いを求めるならば実験を開始する以前に充分相談し，可能な範囲で最良の実験計画を立てる試みをすべきではなかろうか。特にガイドライン等を作製する場合にも統計学者の参画を要請し，評価の部分のみでなくプロトコールにも統計学的知識が盛り込まれることが望まれる。

次に統計分野側の問題点の1つとして考えられることは（私がそう思っているだけかも知れないが），実験データを信用し過ぎておられるのでは無いかと言うことである。確かに与えられた数字しかないので信用するよりしかたないのかも知れないが，データ自身をもっと疑っても良いのでは無かろうか。信頼性の低いデータをいかに処理しても正当な判定，評価を下すことは不可能である。この点は即生物分野側にも重要な点であり，もっと信頼性の高いデータを得るよう努力すべきであろう。

今回はエイムス試験及び染色体試験結果の評価法を概説した後，マウス骨髄を用いる小核試験結果を評価するために背景データとトレンドテストを組み合わせた方法を紹介し，議論のたたき台にしたい。

「決定樹方式での統計手法の注意点」質疑

大野（サイエンティスト）：6ページの“この確率は β や”というところで、 β が0から1- α まで変化するというところをもうちょっと教えていただきたいと思うのですが。

永田：いま二標本で話をしましたけど、これは一標本でも同じです。また、シナリオの中では分散の話でしましたけれども、平均の話にしても同じです。 $H_0: \mu = \mu_0$ に対して検定すると言ったときに、帰無仮説の下での分布というのは図1のAで、このとき R_1 、 R_2 を棄却域として統計検定量の値が U_1 より小さいか U_2 より大きいなら有意だと判断するというわけで

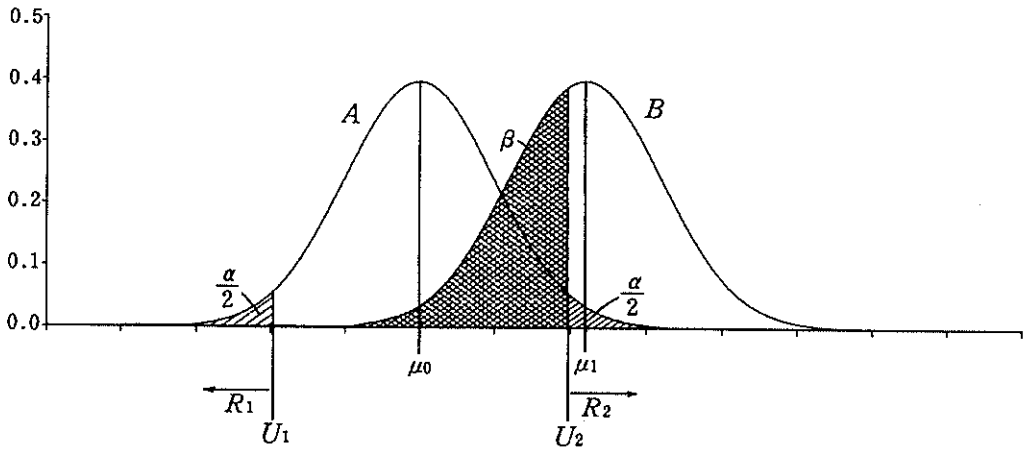


図 1

した。この二つの領域の確率を合わせると α です。けれどこれは帰無仮説が正しいとしたときで、帰無仮説が正しくない場合、そのときに β を考えるわけです。帰無仮説が正しくない場合、例えば対立仮説としてBが正しいとします。つまり本当の μ は μ_0 じゃなくて、 μ_1 であったとします。Bは H_1 の下での分布を表し、Aは H_0 の下でのものです。そうすると、このときはもし検定統計量の値が U_2 より大きい領域にあればAを棄却してBを採択するということから、正しい判断となるわけです。しかし U_1 と U_2 の間にはBが正しいのにAを採択してしまう。そのとき誤りとなって、その確率が β ですね。そうすると H_0 を採択する誤りというのは棄却域に落ちないわけだから、この範囲に検定統計量の値が計算されたときになります。いまはBの分布が正しいのだから、 β というのは、Bについて U_1 と U_2 の間の領域ということになります。Bの下では β はこれくらいの確率だけでも、母平均 μ がも

っと大きいところにきていたとしたら、分布が右にずれて例えば図2のCとなります。βというのは U_1 と U_2 の間の領域の確率なので対立仮説の下での分布がAの分布より大きく離れてきたら確率βは少なくなる。もっと離れたら0にどんどん近くなる。だからβというのは μ_0 と μ_1 との差によって決まる。もし μ_0 と μ_1 の差がギリギリに近づいてきたとしたら、 H_0 の下での分布と H_1 の下での分布がほとんど重なってしまうので、左側が $\alpha/2$ 、右側が $\alpha/2$ 、残り $1-\alpha$ となり、 $\beta \approx 1-\alpha$ となってしまうというわけです。

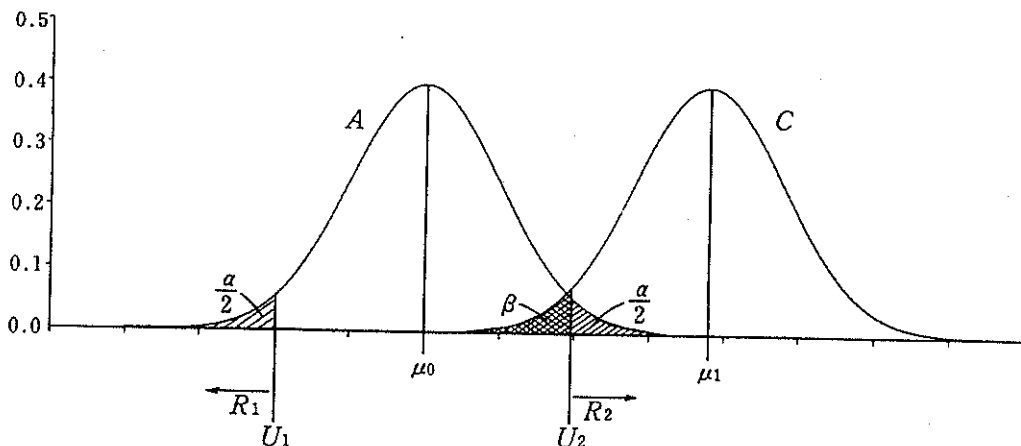


図 2

大野：つまり H_1 の分布の位置によってβも変わるということですか。

永田：その一言だけをいった方が簡単でよかったのでしょうか。

大野：その場合、 μ_0 と μ_1 の差によってβが決まると言われたのですが、 μ_1 というのはわからないわけですね、そうすると μ_0 と μ_1 の差というのは意味をもっているのですか。

永田：例えば μ_0 と μ_1 の差が2ぐらいのとき、2と言っても標準偏差で標準化しないと意味がないですけど、2ぐらいの距離のときにβの確率を0.1以下にしたいという要望を出して、それを満たすようにデータ数をいくつとればよいかというふうにするわけですね。そのために検出力曲線とか検出力を計算するわけですね。この距離は当然わかりませんが、遠く離れていたらβは小さいし、近かったらβは大きい。それではそれがある程度小さいときに、この違いが本検定に対し出来るだけ影響しないようにしようというのが私の話の目的です。

竹内（東大）：統計家としてではなく質問致します。まず、等分散の検定をして有意になって、ウェルチの検定をして有意差がないということになったら、それによって「その二つの薬には差があるということはいえない」と言ってよいのでしょうか。

永田：私はそういう判定をしると言われたことがないので考えたことがないのですが、竹内

先生の言われていることはこういうことではないかと思うのです。つまり違う二つの分散のもとで平均値の検定をする意味があるかということでしょうか。

竹内：逆に言いますと、二つの薬の効果の平均はほぼ同じだけど、バラツキが非常に違うというときに、薬の効果に差がないと言っているのですか。

永田：効果というのをどういうふうに定義するかわからないのですけれど…。二つの分散が違うときに平均値の検定に意味があるかということですが、これは普通の教科書には書いてありませんが、竹内先生と大橋先生の「統計的推測－2標本問題」という本にはこういうふうに書かれています。「 μ_1 と μ_2 が等しくて σ_1^2 と σ_2^2 は違う」という仮説は余り意味がない。なぜなら本当に知りたいことは二つの分布が一致するということを帰無仮説と置くべきで、それは μ_1 と μ_2 が等しく且つ σ_1^2 と σ_2^2 は等しい。対立仮説としてはこれの否定だから、 μ_1 と μ_2 、 σ_1^2 と σ_2^2 は両方とも等しくないか又は片方が等しくないということで、ここで考えている帰無仮説というものは、この対立仮説の一部に含まれてしまう。したがってこういう仮説は意味がない」というふうに書かれているわけで、確かにそうだと思うのです。「しかし……」と続いていまして、「 $\mu_1 - \mu_2$ の信頼区間を構成したい場面がしばしばあって、その場合は σ_1 と σ_2 の仮定なしに構成したいという場合がある。そして σ_1 と σ_2 が等しかろうと等しくなかりょうが $\mu_1 - \mu_2$ の信頼区間を構成する問題は、結局は検定の話と同じになるので、これこれの議論はやっておいた方がよい」というふうにこの本には書かれているわけです。

それで私の考え方なのですけれど、 μ と σ は質の違うパラメータでして、分布という概念とある程度離れて、母平均というものがそれだけで意味をもっているパラメータだということがよくあると思います。そのときは σ^2 が等しい、等しくない、のどちらにしても意味があると認めるならば検定したい。特にここでは話ができませんが二つの回帰式があって、その回帰係数が等しいかどうかを検定したいというときには、もうそのうしろにある誤差分布というのは余り興味がなく、それよりも表に出ているパラメータの検定をしたいということになるので、特別な意味をもっている状況がよくあると思うわけです。だからこういう仮説をおいてもいいのではないかと思います。

竹内：私は統計的ではなくて実際問題として伺いたいのですけど、「1薬と2薬に差があるかどうかをみるために次のデータを解析する」と書いてあって、そして「そのデータ解析の目的は、 μ_1 と μ_2 が等しいかどうかということを検定することだ」と書いてありますけれども、二つの薬に差があるかどうかということと、 $\mu_1 \neq \mu_2$ かということと同じなのでしょうか。

永田：それは私には判断しかねますが、バラツキが違うということは当然何か違うわけですから、何か違うという結論は併記しておくべきだと思います。

竹内：つまり検定問題について考えるとこういうふうには言えないのです。 $\mu_1 = \mu_2$ の仮説を検討するということは、「二つの薬の平均的効果のうちどっちがいいか」ということが言えるかですが、「その仮説が棄却されないということは、平均的にどっちがいいということとは言えない」ということです。そこまでの答えが出るわけです。次に、しかしそれは全然差がないということではなくて、バラツキが非常に違えば何らかメカニズムは違うはずで、そうするとこの薬はある特定の人には非常に効くけど、ある特定の人には効かないかひょっとしたら害があるかも知れない。そうすると安全性の問題について言ったら、 σ の大きなものというのは危険な可能性があるわけです。あるときには害を起こす危険性があるわけで、そういうときには同じ μ であっても σ の小さいほうがいいかも知れない。その逆に、もうとても望みがないときには、どっちみち平均的な薬では駄目なのだからひょっとしたらうまくいってホームランになるかも知れないというような薬のほうがいいかもしれないということもあるかもしれないわけで、そうするとそれはやっぱり意味が違うかも知れない。

ですから私が思いますに、永田さんがお書きになったものは、このロジックの範囲では大変結構だと思うのですけれども、2つの σ の検定もやっぱりつけておいて、その場合の α は25%にする必要はなく、5%とか1%でやって、やっぱり σ が著しく違う可能性があるということは注意としてはつけておいたほうがいだろうと一般的には思うわけです。薬に差があるかどうかという命題を統計的な命題に翻訳するところは、やはりかなり注意が必要なのではないか。これは統計家が決めることではなくて、薬の実際の問題をよくご存じの方が決めることであって、そういうことについては統計家を余り信用してはいけないうし、統計の本に書いてあることも信用しないほうがいいし、「統計的推測—2標本問題」なんかに書いてあることももちろん余り信用しないほうがいいというふうには考えています。

松本（東洋醸造）：竹内先生の話が安全性の問題に及んだのですが、動物を使つての毒性試験というのは3段階か4段階に用量を設定して高い用量を投与するわけです。一番高い用量は死ぬか死なないかのところに設定するため毒性が必ず出るところです。ということは対照群と比べると必ず分散は大きくなる。バラツキが大きくなります。発癌性なんかをやりますと甲状腺なり、下垂体なんていうのは100倍近い重量の差が出ます。そういうものの分散は大きくなります。そうすると σ_1^2 と σ_2^2 が違ってくるというのは、しょっちゅう起こることだと思うのです。竹内先生は、「それはそれで分散の検定で出しておいて、分散はこれだけ違ったとして、平均値の検定はそのままウェルチでもシェフェでもtでも何でもいですが二標本検定をする」とおっしゃっている。

竹内：そういう場合はおそらく μ ではなくて、 $\mu + \text{何}\sigma$ かが問題になるかもしれないし、あるいはあるところを越える確率というのが問題かもしれない。そうすると μ と σ が両方絡

んでくると思うのです。ですからその場合 μ だけの検定をしたり、 σ だけの検定をすることは危険で、両方総合的に判断しなければいけないということが一般的に多いだろうと思います。具体的なケースを伺わないとそのときにどういう行動になるかわかりませんが、特に人間に対する実験だと、死んでしまうまでやるのは危険で、はるか手前でやるわけですから、そうすると μ はほぼ同じだけれども、従って平均的な人は余り関係ないけれども、一方は σ が10倍であるから、こちらのほうはたとえば 1σ のところは危険で15%なり16%人は死ぬけれど、他方は 10σ だから1億人のうち1人しか死なないという差があった場合、 μ が同じだからと言って、検定で採用するととんでもないことの構造がありうると思うのです。それはどういうときにどうなるかは具体的にはわかりません。

だから何か統計学の教科書というのは一貫してばかり書いてあるし、悪いくせだと思っています。

滝沢（日本ロシュ）：今の竹内先生が出されたことのコメントですが、最近経験したことなのですが、農林水産省に申請したものの中に平均値が等しいが有意差があるものがあつたのです。これは順位検定のほうを採用したからそういうことが起こつたのですが、それで農林水産省の係官から「何だかおかしいのではないか」というコメントがあつて、その計算の手順をていねいに説明するといったこともありました。その場合、一つのデータが外れ値だつたわけです。でも研究者の人と議論して、その値はおかしい値ではなくて、どうしても削除はしたくないのでちゃんと入れたということです。いつもそういうデータがあるので、基本的に順位検定を使いたいということで、順位検定で行いました。

それから次に、先ほどの β が0から $1-\alpha$ までということで、それは H_1 が正しいときの第2種の誤りの可能性の割合だと言われました。それはそれでわかつたのですけれども、いつも迷うこと、あるいは実験者の人といろいろ議論して言われることなのですけれども、実験者にとって α とか β は何となくわかりづらい。Fisher流の考え方で、いろいろ実際に実験してその結果より H_0 の下で、つまり平均値が等しいと見たときに確率計算をすると有意確率が出てきます。それが5%以下だったら有意差がないのだというのはわりとわかりやすいし、最近はいろんな統計パッケージで有意確率が出てきます。それで有意確率をそう判断するわけです。つまり両方に差がない場合の確率だと。そうするとそれでは有意確率が例えば5%だったとすると残りの95%は何なのだ。つまりそれは有意確率が等しい確率ならば、残りの95%はというのは両群が等しくない確率なのではないだろうかと言われるのです。何か違うような感じもするのですが、もっともなような感じもして。何かこれをうまく説明して下さいという質問なのですが。つまり有意確率は何かということと、それから $1-$ （有意確率）は何かということです。

永田：1 - (有意確率) というのは余り意味がないというか……。正しい H_0 を棄却する確率が有意水準です。棄却域をたとえば5%と決めたら臨界値が求まり、たとえば1.96となります。たとえば2.00という検定統計量の値が出たとします。そうするとこれは5%で有意ということで正しい H_0 を捨てる確率は5%となり、これは有意水準です。有意確率というのは2.00と出てきたときに、この2.00よりも大きくなる確率、両側検定を考えているなら、その部分と対称に-2.00とった部分を両方合わせたのが有意確率と呼ばれるわけです。だから限界水準とでも言うのでしょうか。有意水準何%でギリギリ有意になるかと言ってもいいと思うのですが。検定統計量が与えられた後で、そこからさきの部分の確率を計算するというのが有意確率。定義はそうです。それをどう判断するかというのはピンク(毒性・薬効 データの統計解析)の本にも確か書いてあったと思いますが、この辺は吉村先生に答えたいと思います。

吉村(名大)：その有意確率の話はこの本で言うところ、有意確率の説明のところ、「有意確率の意味と利用」(P.198)ということが書いてあるので、ある意味ではここを議論していただけないのですが。有意確率というのは、基本的には検定統計量を別の形で言い直した話だと考えておきまして、あまりそれを「確率だ、確率だ」というふうには思わないほうがいいと思っています。検定統計量がどれくらい極端な値になっているかということを検定統計量の数字だけ見たのではわからないから、それを確率という形で表現してあるだけです。つまりどれくらい極端な値であるかということ、帰無仮説が正しかったならば、「そんな極端な値が出る確率はこんなに小さいのだよ」というような数値でその検定統計量の値の大きさ、つまり帰無仮説がどれくらい、多分信用できないだろうということの大きさを表したそういう量なのであって、それに我々がサイコロをふるときのような意味での確率をイメージはすべきではないと思っています。どうでしょうか。

滝沢：数学的にはそれでいいのかもしれませんが、普通実験やっている人とか、余りそういうことに慣れない人たちには、何となくわからないんです。例えば今吉村先生は「帰無仮説が信用できない確率だ」といわれましたが、それでは「信用できる確率は95%なのかな」などと思ってしまう。多分それは分かりやすく説明されているのかもしれませんが、逆に混乱するような感じもします。

竹内：「 H_0 が信用できない確率が5%」というふうな言い方をされるからそのようになるわけで、つまり H_0 が信用できる確率は5%なのです。信用できる確率は5%ですから95%のほうが信用できない確率、つまり他の方が多分正しいだろうと思う、一種の確率です。但しそれは厳密にやるとベイズ流になってしまいます。ベイズ流になってしまうと、またprior probability (事前確率) が絡んできて、実はposterior probability (事後確率) とが必ず

しも一致しないのでいろいろな妙なことが起こるのですけれども、直感的に言えば、信用できる確率が1%で小さいから信用しないことにしようということなのでそれからそれでいいのではないですか。信用できないということが1%というのは、信用できなくはないのだから信用できるということになって、それはおかしいでしょう。だから信用できない確率ではなくて信用できる確率が1%。従って信用できない方が95%とかで、それでいいのではないのでしょうか。

滝沢：わかりました。では信用できるということは、つまり平均が等しいということ信用できるのが1%だということですね。そうするとそこで飛躍かもしれないけど、実験者の人たちは、平均が等しくない確率が99%なのかなと思ってしまうのではないのでしょうか。

竹内：それはよくないのですけれど、つまり平均が等しくないというほうの信頼性が例えば99%ぐらいだと思うぐらいならかまわないと思います。「平均が等しくない確率が」というとまたそこに大分問題が起こって、仮にベイズ流でやってもそこは随分数値的にも違ってくることはありますから。

でもさっきおっしゃった一番はじめの言い方は、信用出来ない確率が1%じゃなくて信用出来る確率が1%である。それはそれだけで後には「それから先は難しいのだから他のを読んでくれ」と言って逃げたほうがいいと思います。

浜田（武田）：直接に対応するかどうかよくわからないのですが、毒性試験で大動物などだと1群3匹とかいうことでnの数が非常に少ないのですけれど、そういうようなデータでも対照群のデータだけ集めてきて、ヒストリカルコントロールをつくと1群50から100ぐらいのヒストリカルコントロールが出来ます。それと今新しい試験をやったときで処理した群の平均値の差を比較したいなどという問題が出ます。つまり3対100だとか3対50だとかいう非常にアンバランスな群の平均値を比較するというような状況が生じ得る可能性があるのですが、そういった場合にどういった検定を用いたほうがいいのかご意見を伺いたいのですが。

永田：ヒストリカルコントロールという話なんです、これも同じ問題でよくわかりません。今の答えになるような研究はしっかりやられてないと思います。ヒストリカルコントロールについては私はそれ程詳しくなくて、柳川先生たちが一生懸命やっておられる論文をたまにちらっと見るぐらいなんです。ヒストリカルコントロールのデータがあって、現実験でもコントロールの実験をやった。そのときにやりたいことというのは、前のデータと今度のデータの両方共何の処置もしていないのだから、同じ種類のデータだと思ってこれを合併させてやりたい。合併させるとサンプルサイズが上がるし、または現実験のコントロールのデータについて、動物を殺さなくても数が減らせますから合併できるものはさ

せたいということで、図3のような決定樹になる。ヒストリカルコントロールの過去の蓄積のデータの分布と現実験のコントロールのデータの分布をFとGとして、この分布が等しければヒストリカルコントロールのデータを取り込んで解析してやればよいし、違うのであれば取り込まずに解析をしたほうがいだろうということになる。

ヒストリカルコントロールのデータ～分布F
 現実験のコントロールのデータ～分布G

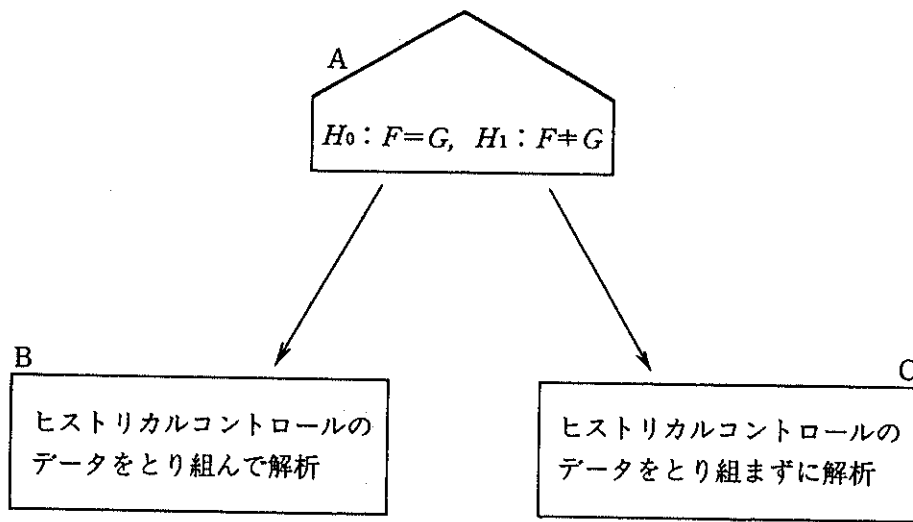


図 3

そうするとどうい問題点が出てくるかと言うと、今言われましたように取り込んでやるとコントロールのデータ数がとてもふえて、処置をしたところのデータというのは数が少ないからアンバランスになります。アンバランスになると何が問題になるかと言うと、 $H_0: F = G$ であるのだったら問題がないのだけれども、 $H_1: F \neq G$ であるのに等しいとみなして図3のBへきているときに問題となります。そのときにたとえば分散が違った、それを見逃して等しいとしてBへ来てしまったとします。そうすると分散が違ってサンプルサイズがかなり違うわけですから、それはさっきの話を延長で当然検定に影響されます。 α_T が大きくなったり小さくなったりします。二標本の検定でもそれから一元配置の分散分析などにおいても、標本数がアンバランスの場合に分散が大きく違ってくると検定に影響を受けるということが知られています。それから、母平均が違っているのに等しいとみなして、Bへ来ている状況も考えられるわけで、そのときにはバイアスの影響も当然入ってきます。こういうことに関して結局はやはり二つのデータを合併するかどうかというのを形式的に

5%の検定でやるのだったらダメだろうとは言えるけれども、具体的にこういうのがどの程度の影響を及ぼしてくるかというのは、文献に出ているのかもしれないけど、ぼくは知らないです。

竹内：今質問をした方にもう一度お伺いしたいのですが、実験の目的はということなのですか。つまりヒストリカルコントロールと現在の実験データを比べようということですか。それとも現在の状況のもとで現実実験のコントロールと対照群と比較して、その差をみるときにただ現在の実験のコントロールの数が少ないから過去のヒストリカルコントロールの分も加えてやろうということなんですか。その問題の状況をはっきり伺いたい。

浜田：当該試験の処理と、過去のヒストリカルコントロールのデータの平均値の比較をしたという話です。

竹内：当該試験の処理群だけでこの中にはコントロールはないわけね。

浜田：当該試験のコントロールをヒストリカルコントロールに合わせてもいいとは思っていますけれども。

竹内：だからその合わせるか合わせないかが問題なのか……。

浜田：合わせるか合わせないかの問題ではないです。

竹内：問題ではないわけですね。今お答えになったのはそれを合わせるか合わせないかの問題だというふうに質問を解釈してしまったようだったので、だから質問の主旨がそうかということちょっと伺いたかったのです。

つまり当該処理群とヒストリカルコントロールを比較する問題。

浜田：ヒストリカルコントロールが全体で50ぐらいあるときに当該試験のコントロールの3匹を加えても大きな変化はないので、使うか使わないかというのは余り本質的な問題ではないと思うのですけれど。要するに過去に集めた大きなデータと現在の処理の3匹ぐらいの少ないデータを比較するという問題……。

永田：過去のデータが現在の実験したコントロールと合致しているかどうかというのを図3のAのような形で検討しなければいけないと思うのですが……。

浜田：合致しているという条件のもとで、そのときに処理群と比較するときに、本当にアスピルウェルチの検定だとかt検定を用いることが妥当であるかどうかおたずねしたいのです。

永田：そういうのはもう普通の二標本のt検定ですから先程の話のとおりだと思います。データがアンバランスになっているから注意がいる、分散が違う可能性があれば、先程の決定樹のようなやり方だと影響される可能性があるという答えでいいのではないかなと思うです。

松本：その答えになるかどうか分かりませんが、当該コントロールと検定したり、ヒストリカルコントロールと検定したりするときには、おそらく両方の検定をすることは余り意味がないし、その3例が本当に異常かどうかというのを、2SDか3SDのどれかわからないけど、標準偏差のある範囲の中で判断したほうがぼくはいいと思います。

それはもう検定の論理ではないと思うのです。3例とか4例のデータは毒性の犬の実験ではたくさんあるのですけれど、もしそれが10例ぐらいであれば1群のデータの検定を行えばよいのではないのでしょうか。

吉村：ヒストリカルコントロールの問題については、まず今言われたのは大動物の場合には処置群も対照群も3匹程度ということです。だからその3匹程度の処置群と対照群で、例えば差の検定みたいのをまずやることに意味があるかということが一つの問題です。それに対しては、3匹程度で検定という論理で結論を出すことはそれ程意味があるとは思えません、余りにも検出力が低すぎて。だから今松本さんが言われたのは、その場合やはり検定と言う論理で形式的に議論をするのはよくないだろうというのが一つの問題です。そうすると過去において同じような実験をたくさんやっているから、コントロール群だけは非常にたくさんデータがある。だからそのコントロール群については過去のデータを全部使って、処置群に関しては今現在の実験データを使えばいいのではないかというのがもう一つの問題になるわけです。その場合に今永田先生が言われたことは、そもそも過去のコントロール群と現在のコントロール群とを同じものと完全にみなしていいのだろうかという問題を言われたわけです。

実はヒストリカルコントロールを使うことの最大の問題はそれなのでして、現在の対照群と過去の対照群が100%同じであるならば、現在の実験に対照群をおく必要はないわけです。それにもかかわらず現在の実験に対照群をおくというのは、やはりそれは必ずしも100%等しいという前提が満たされないからです。そうだとすると、現在の対照群と過去の対照群を同じものとみなしていいかということの検討がどうしても必要になる。それをめぐって今ヒストリカルコントロールの使用ということに関する研究がなされているわけです。いろんな提案があり、いろんな議論があります。

それでもし過去の対照群と現在の対照群が100%同じということが保証されたという前提であるならば、それは非常にサイズの大きい対照群とものすごくサイズの小さい処置群との差を問題にする話になって、平均の検定だけに関して言うと、今日の永田先生の今までの議論がそのまま利用出来るということになります。

竹内：前の部分は吉村先生のおっしゃたとおりでありますけれども、一番最後のところを申し上げますと、おそらく平均の差の検定のときに、片方のサイズが非常に小さくて、しかも処置群

は分散が非常に大きくなる。もし仮説からずれたときに分散が非常に大きくなる可能性があるときには、分散の推定に関してはその部分を使わないで、コントロール群だけから分散を計算したほうが良いと思います。それで少しだけ係数を注意すれば後は同じようにできますので。

浜田：この機会に是非伺っておきたいのですが、分布が正規分布からかなり離れているときには、要するにウィルコクソン検定を使うと分布の形を全部表現できるわけです。50とか100あればかなり、ヒストリカルコントロールデータにすると大体分布の形状がわかって、どれくらい正規分布から離れているかどうかということもわかるので、分布の形を利用するという意味で、ウィルコクソン検定のほうが良いのではないかなというふうに直感的に考えています。

竹内：ウィルコクソン検定は、処置群のほうもサンプルサイズが余り小さくないならいいと思うのですが、さっきおっしゃったように3ぐらいだとパワーがよくないと思うのです。10ぐらい以上でないといけませんね。やはり片方が3とかだと、つまり順位統計量の個数が少ないですから非常に離れても、一番大きくても n しかないわけです。だけど例えばそれが 3σ も 5σ も離れていれば、ふつうの意味の t 検定をやれば一度に決まってしまうから。

もう一つは、実は正規分布から離れていても t 検定は少なくとも α のサイズに関してはかなりロバストになります。その場合 α に対してはかなりロバストですからそれほど悪くないと思います。パワーが悪くなる可能性があるわけですが、どっちみち片方のサンプルサイズが非常に小さいとウィルコクソンでもパワーは非常に小さくなりますからそれは仕方がないので、先ほどどなたかがおっしゃったように片方の群が小さいときは、むしろアウト라이어検定的な発想で「もとの分布と同じ分布からの標本とみなし得るか」という意味の仮説の検定と考えてやったほうがよいと思います。そういう意味で私は分散の推定に処置群のほうを使わないで、ふつうの t 型の検定をやるのがいいだろうと思っています。

永田：以前に同じようなことを話したときに、「分散の検定で不等分散となったら t 検定を用いる、そうでなかったらウィルコクソンを用いる」と大阪の研究会のときにどなたかが言われました。つまり「分散が違っているときは分布が違うからノンパラを用いる。」でもそれは全然よくなくて、ウィルコクソンの検定というのも2つの分布の型が同じであるという前提で違いを計っているんで、分散が違うときにウィルコクソンを用いても、やはり t 検定と同じように α_T に影響が及ぶということがあります。だからウェルチの代わりにウィルコクソンを用いるということはいい手法ではないということなのです。

それからこれも答えのない話なので避けて通ってきたのですけれど、決定樹のところで「分布の型は正規分布か」というのもあります。これも答えがあるわけではなくて、問題点を言うだけの話ですが、「正規分布である」という帰無仮説に対して正規性の検定をやって、正規分布であるとみなせるのだったら正規分布に基づく手法を用いよう、そうでなかったらノンパラの手法でいこうという決定樹です。これも検定をやったときに、または推定をやったときにどういう影響が出てくるかということを経験的に定量的に書いてある文献というのを知らないのです。正規性の検定というのは、いろいろあって一体どれを使ったらいいのかということがまず1点。正規確率紙が一番いいという話もありますが、本当にそうかわからないし、それではそれをどのように用いるか。さっきの言葉を使うと、この予備検定での α_0 をどう設定するか。正規確率紙で言うと、どれぐらいで直線とみなすかです。そういうのも、どうすればいいのかという議論がない。

それではどういう話になるかという、先程竹内先生がちょっと言われたことかもしれませんが、正規分布に基づく手法であっても、正規分布の仮定がある程度くずれていても大体正しいものだということがあります。これはいろいろ調べられていていろいろなことがわかっている。それからノンパラの手法を用いても、それ程正規分布に比べて見劣りがしないかどうかということもあり、それは効率という問題で調べられている。この辺の問題を合わせてある程度判断は出来るかもしれないですが、そういう判断をきちっとやりたいというのがどうも吉村先生の前からの話です。吉村先生のことばを使うと、「パラとノンパラの使い分けをやりたい」というふうにおっしゃっているので、この辺の問題が先程から話が出てこないと思われる方は、まだ答えとしてきちっとしたのがないのであって、吉村先生がこれから検討していただけるんだろうというふうに思います。

丹後（国立公衆衛生院）：ロバストネスのことに関して話したいのですが、いわゆる統計家の間でのロバストネスと、データを使うユーザーにおけるロバストネスの意味がかなり食い違ってきておまして、統計的に、「これはロバストなんだから使っていていいんだ」と言いますと、大抵ユーザーは何でもかんでも使ってしまうという傾向が非常に多いわけです。ロバストネスというのは非常に慎重に言わなければいけないので、つまりどの程度だったらロバストかということが非常にむずかしいわけです。従って私の方針というのは、余りロバストネスということを強調することは決してしないで、やはり正規性かどうかということに関しても決定樹でする云々ということよりは、ヒストグラムを見て大体判断できるようなことで進んで十分ではないかということを経験的に議論しています。ですからロバストネス云々ということを強調し出すと、おそらく何でもかんでも使ってしまうことが多いので、その辺はこういう席で特にロバストネスということに関しては慎重にならなければいけない

のではないかというコメントをしておきたいと思います。

大橋(東大): 上げ足をとるようですみません。ヒストグラムを使えということなんですけど、私はこの講習会でも他の講習会でもヒストグラムは一切使うなと言っているのです。なぜかという、日本では品質管理教育がすごく徹底してしまっていて、ヒストグラムの書き方という流れ図があるわけです。そして標準偏差を計算していく。あれをやられると困るのです、バイオのデータで。まずアウトライアーが入るようなデータで標準偏差をバラツキの尺度とされたら大変困る。ですから幹葉表示を使った絵を描けということなんです。ヒストグラムよりもデータを見るということを重視すべきで、ヒストグラムは特に上があいたりするのですね、プールしたりするやり方もあるから。あれはよろしくないということが一つ。それからロバストネスという言葉はこの講習会でも言っているのですが二つあって、「正しさ」と永田先生がおっしゃった、統計だとバリディティ (validity) というのかな、ちゃんと第1種の過誤が守られるという意味のロバストネスと、パワー、効率のロバストネスと2通りあります。今の正規性の場合、バリディティの問題は余りないと思います。パワーの問題が出てくる。パワーで議論しなければいけないので、他の議論がもうちょっとむずかしくなるかもしれないなと思っています。

ですから今お話にあったような予備検定のように割にすっきりしたリコメンデーションが出にくいのかなという気がしています。ですからそれだけむずかしくてチャレンジなかもしれないと思っています。

竹内: 大橋さんのおっしゃっていることにちょっと補足したいのですが、さっき丹後さんがおっしゃっていたロバストネスという言葉は確かに注意して使ったほうがいいです。というのは、ある仮定がとれたときにはロバストであるという方法は他の仮定がずれたときにはダメになる。例えば「正規分布から形がずれても」というときには、正規分布から形がずれてくる。しかしそれが正確に同じ形のままで x の分布と y の分布がずれているときにはパワーがどうという議論があるわけですが、例えば分布の形は違うけど真ん中だけあっているということから検定する問題だったならばそれはまずくなる場合があります。よく統計の数学家で、ロバストネスのことをやっている人は、ときどきある仮定の特定の部分だけをバカに詳しくやって、他のことは全然忘れてる。それでこれはロバストネスだよということを言ってきたりする。これはやめたほうがいいのですが、それ以外のことを除けばロバストネスというのはやはり大事な概念だと思います。それからもう一つ正規分布を仮定した方法は、大体バリディティに関してはロバストネスだというのは、今大橋さんのおっしゃったとおりで、もう少し言いますと実はバリディティに関してコンサーバティブということなのです。つまり大体5%水準ということで検定をすると、実際に仮説が正しいと

きに棄却される確率は、3%になったり2%になったりするのです。特にコーシー分布みたいに極端に正規分布からはずれた分布を想定すると非常に小さくなる。小さくなるということは、つまりパワーが小さくなるということを意味するわけです。帰無仮説のもとでも小さくなる。対立仮説のもとではもっと小さくなるのです。

というわけですから、つまり正規分布かどうかわからないというときに、仮説を検定して有意差が出たら、これは有意差があるということについてはかなり確信をもっていいのですが、有意差が出なかったときに、ひょっとすると他の検定をやると有意になるかもしれないということのほうを注意したほうがいい。つまり注意の方向についてちょっと一言。

五十嵐（エーザイ）：ちょっと反論するみたいなのですが、基本的に重要なことは、今言おうとしたことは実は最初に竹内先生から言われてしまったのですけれども、基本的に薬効の試験を考えるときは集団はホモジニアスではないということなのです。つまり効く人もあれば効かない人もいるということが非常に重要なのです。従っていわゆる μ で見るということは非常に危険なのです。だからむしろ、バラツキの大きさを検討するというのは非常に重要だと考えておまして、特に時点を追って、例えば計量値を観測していく場合には、人との交互作用が非常に大きいのです。つまりそれはバラツキに現れてくるわけですから、ある意味で μ で見るということは、効かない人も効く人も要するに一緒にしてしまって0にしているという見方になってしまうので、むしろぼくはバラツキの大きさを検討するというのを十分にやって、それこそが本当の検定であって、 μ は次に来るべきものであろうというふうに考えております。

岩崎（防衛大）：ちょっと伺いたいのですが、分散が違うというときにその違い方にはいろいろあると思うのです。例えば分散をそろえる努力をしたときに違う場合と、いいかげんにデータを取ってきたら違ってしまったというような感じと両方あると思うのです。今の話はどちらの場合かということをお聞きしたい。ですから実験をやるとき一生懸命やった上で尚且つ違ってしまう場合と、ラフにデータを取ってきて違ってしまうという場合があります。だからそうした場合に同じでいいのかどうかということをお聞きしたいのです。

永田：こういう問題というのは統計の論文には必ず例えば「 $\beta_1 = 0$ であろうと疑われている状況ではこれこれの手順を用いる」と書いてあるわけです。それで今日お話したような議論が入ってくる。だからそういうことが書いてあるということは、何らかの先見的情報とかまたはそういう努力をしたとかいうことで、こういう仮説が多分成り立っているのではないかというような状況のもとにこれこれの手順が用いられると考えられているわけです。だけどそれをどういうふうに手順の中に反映させるかということとはよくわからないの

です。例えばプレイヤーみたいなもので、この帰無仮説がある程度の確率で成り立っているだろう、またはそういう努力をした、ということをごここに反映させるようにするというようなことはなかなか客観的に出来ないで、努力した場合もそうでない場合も同様に扱ってやっているということになってしまうと思います。

岩崎：バラツキにいろいろ種類があって、単にバラついているときと努力をしたのだけでも尚且つバラついているときがあって、多分後者のほうがバラツキの真の定義だと思うのです。だからさっき正規分布じゃないときはどうするのだと言われたとき、では正規分布になるように実験なり何なりを一生懸命組んでやってその結果正規分布にならなかったのか、その辺が実は知りたくてたまには統計のほうから反論しなければいけないと思っていますのですけど。

竹内：こういうことだと思うんです。仮説のもとでの μ と σ がありますね。対立仮説のもとでの μ と σ は、例えば対立仮説のもとで μ だけずれて σ は同じだという、あれは統計家が全く勝手にでっち上げたケースで、そういうことはほとんどありえないと思います。 μ もずれば σ もずれるに決まっていると私は思っています。一生懸命実験しての話です。でたらめにやれば σ はもちろん勝手に変わりますけれど、そういうことではなくて。特に薬の場合に、例えば標準的な薬と別な薬があるとき、恐らく薬の効果に差があれば、もし μ がずれるようなことがあれば当然 σ もずれるだろうと私は思うのです。それは例えば正常な人と病気の人との検定をしても、病気の人と正常の人のグループを考えたら当然 μ だけではなくて σ もずれるのは当たり前で、病気になると正確に血圧が正常な人より何ミリ上がって σ が同じだなんてバカなことはあり得ないです。だから σ は必ず大きくなるだろうといったら、「いや、そうでない場合がある」とうことをお医者さんのどなたかに教わったことがあって、病気のほうがかえって σ は小さくなるケースがあるのだという話を伺ったことがあるんです。それは極端な例を上げれば確かにそうで、死んでしまえば血圧は0になりますから、 σ は0であるということになるかもしれないのですけれども。

私から言いたいことは、今の話はつまり σ が変わるというのは当然であり、それは恐らくそういう意味で変わるだろうということですよ。

五十嵐： σ が変わったときには、当然集団はホモジニアスじゃないですね。ということは μ を計算すべきではないということですよ。

竹内：それはもとから、つまりコントロール群でももちろん集団は単一ではないわけです。だけどコントロール群で単一ではないにもかかわらずコントロール群に差が出ないのは効果がないからで、つまり誰にも効果がない。つまりさっきおっしゃった交互作用だと、まさにその通りだと思います。つまり薬と個人の交互作用があるわけです。何も効果がないも

のは誰にやっても同じです。

そこで今おっしゃった通りだと思うのでして、ただ問題は恐らくそこで、例えば個人の差というのを変量とみなすか母数とみなすか、ランダムなバラツキとみなすか、それともそれは何らかの意味で意味のある差とみるかということによって差があるので、これは統計家として言うべきことはないのだけれど、重要なことはその差が何によるものなのかを見つけ出すような何らかの手がかりを見つけることだと思います。つまりどういう人にはどういう性質をもっていて、従ってどういう薬が効くのだというようなロジックを見つけることだと思うわけです。そうでなければいくら差があってもランダムしか使えないので、そうすれば σ の差としか現れないで、本当はその差がある集団がいくつかあるはずなのに、それは単に σ の差としか出てこないということになっているというのがデータの上の話じゃないかと思えます。

★★★★★★★ 第43回定例会出席者名簿 ★★★★★★★

日時：1990年7月14日（土）

11：00～12：00 基礎講座

場所：総評会館

（途中昼食）

13：00～17：00 定例会

御出席頂いた先生方

- *吉村 功 (名古屋大学 工学部)
- *大橋靖雄 (東京大学 医学部)
- *鎌倉稔成 (中央大学 理工学部)
- *坂本 豊 (武田分析研究所)
- *浜田知久馬 (武田薬品工業)

- | | | | |
|----------|------------------|----------|---------------------|
| 1 栗原律子 | エーザイ ㈱ | 36 桑山典之 | 帝国臓器製薬 ㈱ |
| 2 山縣清壮 | 呉羽化学工業 ㈱ | 37 佐藤敦子 | 帝国臓器製薬 ㈱ |
| 3 大槻成章 | 呉羽化学工業 ㈱ | 38 渡辺敏彦 | 科研製薬 ㈱ |
| 4 青木敏春 | 呉羽化学工業 ㈱ | 39 田村博信 | 日本新薬 ㈱ |
| 5 川瀬芳克 | 愛知県総合保健センター | 40 五十嵐俊二 | エーザイ ㈱ |
| 6 今井節夫 | 財) 動物繁殖研究所 | 41 飯島 治 | ㈱ ツムラ |
| 7 佐々木秀雄 | 東洋醸造 ㈱ | 42 佐賀大三郎 | テイカ製薬 ㈱ |
| 8 早川保章 | 日本特殊農業製造 ㈱ | 43 下井信夫 | ユックムス ㈱ |
| 9 小林克己 | 財) 安評センター | 44 松本一彦 | 東洋醸造 ㈱ |
| 10 池田陽一 | ㈱ ミドリ十字 | 45 高橋行雄 | 日本ロシュ ㈱ |
| 11 吉田秀信 | 北陸製薬 ㈱ | 46 長谷文雄 | 日本ルセル ㈱ |
| 12 本川 裕 | キリンビール ㈱ | 47 滝沢 毅 | 日本ロシュ ㈱ |
| 13 清水辰巳 | キリンビール ㈱ | 48 今溝 裕 | 東洋醸造 ㈱ |
| 14 窪田 | 日本レダリー ㈱ | 49 大塚芳正 | 日本ロシュ ㈱ |
| 15 木住野 | 日本レダリー ㈱ | 50 中里博志 | 雪印乳業 ㈱ |
| 16 三浦哲朗 | 山之内製薬 ㈱ | 51 松本道治 | 日本バイオアッセイ研究センター |
| 17 登城 豊 | ライオン ㈱ | 52 久勝 諭 | アイ・シー・アイ・ジャパン ㈱ |
| 18 関 康弘 | ライオン ㈱ | 53 仲辻堅造 | 明治製菓 ㈱ |
| 19 松林哲夫 | 昭和薬科大学 | 54 光森達博 | 鐘紡 ㈱ |
| 20 渡部博之 | ㈱第一ラジオアイソトープ研究所 | 55 和気洋一 | 中外製薬 ㈱ |
| 21 荒川和仁 | ヘキストジャパン ㈱ | 56 佐藤 景 | 小玉 ㈱ |
| 22 斉藤富貴子 | スミスクリンピーチャム製薬 | 57 安藤誠人 | 明治製菓 ㈱ |
| 23 渡辺 徹 | クミアイ化学工業 ㈱ | 58 滝沢光男 | ヘキストジャパン ㈱ |
| 24 五十嵐幸信 | 住友製薬研究所 | 59 七田 修 | 東菱薬品工業 ㈱ 青梅研究所 |
| 25 大治政夫 | ㈱ 富士生物化学研究所 | 60 塚越普佐雄 | ㈱ 大塚製薬工場 |
| 26 荒木明宏 | 日本バイオアッセイ研究センター | 61 柴原俊一 | 大塚製薬 ㈱ |
| 27 加藤典代 | 日本バイオアッセイ研究センター | 62 荒野龍昭 | ワーナーランバート ㈱ |
| 28 土門総一 | 第一製薬 ㈱ | 63 小野 宏 | 財) 食品薬品安全センター-秦野研究所 |
| 29 三郎丸清 | 大塚製薬 ㈱ | 64 山田雅之 | ㈱ 富士テレビオ |
| 30 林 貢平 | 藤沢薬品工業 ㈱ | 65 佐々木 貴 | ファルミタリアカルロエルバ ㈱ |
| 31 平井 明 | クミアイ化学工業 ㈱ | 66 赤木健秀 | ファルミタリアカルロエルバ ㈱ |
| 32 五井康治 | センチュリ リサーチ センタ ㈱ | 67 田村友子 | ファルミタリアカルロエルバ ㈱ |
| 33 山岡秀明 | 住友化学工業 ㈱ | 68 坂原久子 | 協和醸酵 ㈱ |
| 34 竹内鉄雄 | 佐藤製薬 ㈱ | 69 成田 強 | エーザイ ㈱ |
| 35 池田高志 | 帝国臓器製薬 ㈱ | 70 野田義寛 | テルモ ㈱ |

- 71 藤原信之 ベーリンガー・マンハイム東宝 ㈱
- 72 中濱多恵子 KRI International. Inc.
- 73 小林紀彦 バイエル薬品 ㈱
- 74 山下久美 和光堂 ㈱
- 75 近藤専治 エーザイ ㈱
- 76 小山利之 東菱薬品工業 ㈱
- 77 桑原 孝 ㈱ 大塚製薬工場
- 78 有賀 日本バイオリサーチセンター
- 79 山本 日本バイオリサーチセンター
- 80 石井隆太郎 スミスクラインピーチャム製薬㈱
- 81 盛 昭雄 東レ ㈱
- 82 平河 威 三井製薬工業 ㈱
- 83 大脇康雄 セローノジャパン ㈱
- 84 恩田威俊 第一製薬 ㈱
- 85 直 弘 財) 実験動物中央研究所付属前臨
- 86 穂本雅子 三笠製薬 ㈱
- 87 山田英樹 ㈱ 興和
- 88 小熊義宏 ㈱ ビーエムエル
- 89 渡辺 進 高砂香料 ㈱
- 90 若森誠一 旭化成工業 ㈱
- 91 川口泰史 コニカ ㈱
- 92 若田明裕 山之内製薬 ㈱
- 93 中嶋 圓 財) 安評センター
- 94 佐藤精一 日本たばこ産業 ㈱
- 95 玉井功一 ㈱ 保健科学研究所
- 96 小沢重成 キッセイ薬品工業 ㈱
- 97 守田伸子 サンド薬品 ㈱
- 98 高田昌樹 日本ロシュ ㈱
- 99 三浦邦彦 ㈱ 日本ハイボックス
- 100 芦沢健拓 財) 化学品検査協会
- 101 栗原泰蔵 大塚化学 ㈱
- 102 池見直起 摂南大学薬物安全科学研
- 103 川東正美 持田製薬 ㈱
- 104 村野弘行 ㈱ 保健科学研究所
- 105 神谷達郎 エスエス製薬 ㈱
- 106 高橋正美 エスエス製薬 ㈱
- 107 石塚修司 メレルダウ製薬 ㈱
- 108 竹下修史 日産化学安全性評価グループ
- 109 中島和子 日本生物科学研究所
- 110 三内貞子 鐘紡 ㈱
- 111 鶴亀隆之 ㈱ 三菱化成安全科学研究所
- 112 城戸昭彦
- 113 山田 弘
- 114 諫山真樹
- 115 加藤秀成
- 116 竹中栄二
- 117 林 達也
- 118 萩原由恵
- 119 田嶋直子
- 120 萩原孝一
- 121 林 隆志
- 122 永見俊之
- 123 近藤正秀
- 124 坂本京子
- 125 佐藤 昇
- 126 大石正平
- 127 緒方秀俊
- 128 温井一彦
- 129 渡部知亜紀
- 130 本田純久
- 131 奥村大地
- 132 武田豊雄
- 133 石川光一
- 134 河上喜之
- 135 伊藤博之
- 136 渡辺三恵
- 137 清水千磋子
- 138 服部秀樹
- 139 池田正己
- 140 川野泰司
- 141 正木文夫
- 142 矢島 勉
- 143 松葉克子
- 144 久保欣也
- 145 岡本佳弘
- 146 大林繁夫
- 147 林 悦生
- 148 木船義久
- 149 東宮秀夫
- 140 竹原 功
- 141 佐野由明
- 142 越智誠支
- 143 小林高之
- 144 和田武夫
- ファイザー製薬 ㈱
- 中外製薬 ㈱
- サントリー ㈱
- ダイセル化学工業 ㈱
- ㈱ メニコン
- ㈱ メニコン
- 三井石油化学工業
- ㈱ イナリサーチ
- 日本農業 ㈱
- 日本チバガイギー ㈱
- 食品薬品安全センター
- ファイザー製薬 ㈱
- 科研製薬 ㈱
- ㈱ パナファームラボラトリーズ
- 住友製薬 ㈱
- 大正製薬 ㈱
- 東京大学
- ㈱ ヒューマンライフ
- ㈱ 三菱化成安全科学研究所
- 東京大学医学系大学院
- 実中験研
- ノボ薬品 ㈱
- 財) 残留農業研究所
- 日本たばこ安全性研究所
- 財) 日本食品分析センター
- ㈱ 日本ハイボックス
- 資生堂 ㈱
- 富士生物科学研究所
- 持田製薬 ㈱
- ㈱ エス アール エル
- キリンビール ㈱
- 大鵬薬品工業 ㈱
- グレラン製薬 ㈱
- 雪印乳業 ㈱
- 武田薬品工業 ㈱
- 住友製薬 ㈱
- ㈱ 新薬開発研究所
- 保健科学研究所
- 鳥居薬品㈱
- 武田製薬

★★★★★★ 第44回定例会出席者名簿 ★★★★★★★

日時：1990年10月20日（土）

11：00～12：00 基礎講座

場所：総評会館

（途中昼食）

13：00～17：00 定例会

御出席頂いた先生方

*吉村 功 (名古屋大学 工学部)
 *大橋靖雄 (東京大学 医学部)
 *岩崎 学 (防衛大学校)
 *橋本修二 (国立公衆衛生院)
 *栗原 久 (群馬大学 医学部)

1 本川 裕	キリンビール ㈱	36 三郎丸 清	大場製薬 ㈱
2 佐藤 景	小王 ㈱	37 小原直樹	大場製薬 ㈱
3 栗原律子	エーザイ ㈱	38 山田英樹	興和 ㈱
4 平河 威	三井製薬工業 ㈱	39 清水達也	㈱ イナリサーチ
5 荒川和仁	ヘキスト ジャパン ㈱	40 里見恵一	日本実験医学研究所
6 川野泰司	㈱ 資生堂研究所	41 山田和幸	日本実験医学研究所
7 中辻堅造	明治製菓 ㈱	42 土門総二	第一製薬 ㈱
8 松本正人	明治製菓 ㈱	43 山縣清壮	呉羽化学工業 ㈱
9 田村幸資	雪印乳業 ㈱	44 正木文夫	富士生物科学研究所
10 安井忠良	雪印乳業 ㈱	45 川東正美	摂南大学薬物安全科学研究所
11 高田昌樹	日本ロシュ ㈱	46 加藤秀成	サントリー ㈱
12 恩田威俊	第一製薬	47 田中徹也	東京都日野市
13 野木由基子	雪印乳業 ㈱	48 戸塚和男	東菱薬品工業 ㈱
14 五十嵐幸信	住友製薬 ㈱	49 荒野龍昭	ワーナー・ランバード ㈱
15 大石正平	科研製薬 ㈱	50 矢島 勉	持田製薬 ㈱
16 池見直起	大塚化学 ㈱	51 鶴亀隆之	鐘紡 ㈱
17 松林哲夫	昭和薬科大学	52 佐々木秀雄	東洋醸造 ㈱
18 今井節夫	財)動物繁殖研究所	53 坂内堅二	呉羽化学工業 ㈱
19 杉山公仁	昭和薬品化工 ㈱	54 桑原 孝	㈱ 大塚製薬工場
20 大治政夫	㈱ 富士生物科学研究所	55 近藤正秀	日本チバガイギー ㈱
21 池田高志	帝国臓器製薬 ㈱	56 増山 剛	味の素 ㈱
22 斎藤富貴子	ミスクリン・ビナム製薬 ㈱	57 佐々木 貴	ファルミタリア カルロエルバ ㈱
23 岡山佳弘	大鵬薬品工業 ㈱	58 赤木健秀	ファルミタリア カルロエルバ ㈱
24 田村博信	日本新薬 ㈱	59 滝沢 毅	日本ロシュ ㈱
25 池田陽一	㈱ミドリ十字	60 木村 滋	日本ロシュ ㈱
26 東野浩司	日本製薬 ㈱	61 相馬義徳	日本ロシュ ㈱
27 五十嵐俊二	エーザイ ㈱	62 井野裕子	日本ロシュ ㈱
28 馬場百合子	キリンビール ㈱	63 久保田 潔	㈱ ツムラ
29 光森達博	鐘紡 ㈱	64 諫山直樹	中外製薬 ㈱
30 三浦哲明	山之内製薬 ㈱	65 代谷 務	小野薬品 ㈱
31 水井信夫	日清製粉 ㈱	66 木船義久	武田薬品工業 ㈱
32 吉田秀信	北陸製薬 ㈱	67 松本一彦	東洋醸造 ㈱
33 栗原泰蔵	化学品検査協会	68 佐野正樹	大鵬薬品工業 ㈱
34 岡本博夫	森下製薬 ㈱	69 高橋行雄	日本ロシュ ㈱
35 山田恭史	㈱ 日本バイオリサーチセンター	70 下井信夫	㈱ ユックムス

- | | | | |
|---------|--------------|-----------|--------------------|
| 71 半田 淳 | 日本化薬 (株) | 89 桑山典之 | 帝国臓器製薬 (株) |
| 72 長谷文雄 | 日本ルセル (株) | 90 山下久美 | 和光堂 (株) |
| 73 今溝 裕 | 東洋醸造 (株) | 91 伊賀由則 | 雪印乳業 (株) |
| 74 柴野隆司 | (株) イナリサーチ | 92 安田広行 | 住友製薬 (株) |
| 75 林 真 | 国立衛生試験場 | 93 竹中栄二 | ダイセル化学工業 |
| 76 山岡秀明 | 住友化学工業 (株) | 94 小西寿美恵 | 日本たばこ産業 |
| 77 大塚芳正 | 日本ロシュ (株) | 95 河村 寿 | ブリストル・マイヤーズ研究所 (株) |
| 78 中里溥志 | 雪印乳業 (株) | 96 河上喜之 | 実中研付属前臨床医学研究所 |
| 79 三内貞子 | 日本生物科学研究所 | 97 石塚修司 | エスエス製薬 (株) |
| 80 松浦哲郎 | 摂南大学 | 98 越智誠支 | 日本新薬 (株) |
| 81 高橋正美 | エスエス製薬 (株) | 99 小栗健司 | 東洋醸造 (株) |
| 82 山下和昭 | 森下製薬 (株) | 100 瀬川光久 | (株) ミドリ十字 |
| 83 池田正己 | (株) 日本ハイボックス | 101 坂原久子 | 協和醗酵 |
| 84 永見俊之 | 日本農薬 (株) | 102 渡邊孝美 | 東洋醸造 (株) |
| 85 岸 洋文 | 日本化薬 (株) | 103 永井弘充 | (株) 三菱化成安全科学研究所 |
| 86 中沢隆弘 | エーザイ (株) | 104 井之川芳之 | スミスクラインビーチャム製薬 (株) |
| 87 且 喜夫 | 山之内製薬 (株) | 105 菅野 純 | 東京医科歯科大学 |
| 88 本田純久 | 東京大学 | 106 安藤正一 | 東京理科大学 |

★★★★★★★ 第45回定例会出席者名簿 ★★★★★★★

日時：1991年1月26日（土）

11：00～12：00 基礎講座

場所：総評会館

（途中昼食）

13：00～17：00 定例会

御出席頂いた先生方

- | | |
|--------------------|--------------------|
| *吉村 功 (名古屋大学 工学部) | *田中憲徳 (食品薬品安全センター) |
| *大橋靖雄 (東京大学 医学部) | *渡辺正己 (横浜市立大学 医学部) |
| *芳賀敏郎 (東京理科大学 工学部) | *岩崎 学 (防衛大学校) |
| *鎌倉稔成 (中央大学 理工学部) | |

- | | |
|------------------------------|------------------------------|
| 1 三郎丸 清 大塚製薬 (株) 徳島研究所 | 36 北埜貴士 ロート製薬 (株) 第一製品部 |
| 2 大井明英 大塚製薬 (株) 徳島研究所 | 37 佐々木秀雄 東洋醸造 開発管理部 |
| 3 大脇康雄 セローノジャパン (株) 医薬開発部 | 38 小栗健司 東洋醸造 開発管理部 |
| 4 五十嵐幸信 住友製薬 (株) 研究室 | 39 島野 (株) 第一ラボ/イトブ研究所薬事部 |
| 5 本川 裕 キリンビール (株) 医薬事業開発部 | 40 田村博信 日本新薬 (株) 中央研究所 |
| 6 矢島 勉 持田製薬 (株) 安全性研究所 | 41 高原利夫 藤本製薬 研究部 |
| 7 聳城 豊 ライオン (株) 生物科学研究所 | 42 北野英作 藤本製薬 研究部 |
| 8 村野弘行 持田製薬 (株) 安全性研究所 | 43 野田 勉 大阪市立環境科学研究所 |
| 9 佐藤 景 小玉 (株) 開発事業部開発課 | 44 山岡秀明 住友化学工業 (株) |
| 10 高田昌樹 日本ロシュ (株) 医薬品本部 | 45 松村浩二 日本新薬 (株) 中央研究所薬理一課 |
| 11 岡本博夫 森下製薬 (株) 東京支店開発部 | 46 池田陽一 (株) ミドリ十字安全性研究所 |
| 12 長谷川二郎 エーザイ (株) 臨床薬理センター | 47 野口奈佳恵 ベーリンガーマンハイム東宝 (株) |
| 13 栗原律子 エーザイ (株) 臨床薬理センター | 48 佐々木 貴 ファルミタリアカルロエルバ (株) |
| 14 松本正人 明治製薬 (株) | 49 赤木健秀 ファルミタリアカルロエルバ (株) |
| 15 清水辰巳 キリンビール (株) 医薬事業開発部 | 50 田村友子 ファルミタリアカルロエルバ (株) |
| 16 安田広行 住友製薬 (株) 医薬開発部 | 51 田中憲夫 日本レダグリー (株) 生物研究所 |
| 17 東宮秀夫 住友製薬 (株) 医薬開発部 | 52 牧之瀬俊彦 日本レダグリー (株) 情報システム室 |
| 18 恩田威俊 第一製薬 (株) | 53 杉山公仁 昭和薬品化工 (株) 研究所 |
| 19 渡辺敏彦 科研製薬 (株) 開発部統計解析G | 54 旦 喜夫 山之内製薬 (株) 臨床統計部 |
| 20 玉井康治 センチリ-リサーチ- (株) 統計開発部 | 55 山下久美 和光堂 (株) 第2研究所 |
| 21 三浦哲朗 山之内製薬 (株) 薬事監査部QA室 | 56 本田純久 東京大学 医学部 保健学科 |
| 22 岩崎 学 防衛大学校 数学物理学教室 | 57 川野泰司 (株) 資生堂研究所 第2安全性G |
| 23 山下和昭 森下製薬 (株) 医薬情報部 | 58 池見直起 大塚化学 (株) 鳴門研究所 |
| 24 盛 昭雄 東レ (株) 医薬開発部 | 59 桑山典之 帝国臓器製薬 (株) 安全性研究部 |
| 25 光森達博 カネボウ (株) 薬品研究所 | 60 土門総二 第一製薬 (株) 中央研究所 |
| 26 池田高志 帝国臓器製薬 (株) 安全性研究所 | 61 佐藤 昇 ファイザー製薬 (株) 医薬開発部統括 |
| 27 大石正平 科研製薬 (株) 統計解析室 | 62 萩原雄二 昭和電工 (株) 生化学研究所 |
| 28 大木文夫 富士生物科学研究所データ処理室 | 63 甲斐倫明 東京大学医学部放射線健康管理学 |
| 29 荒川和仁 ヘキストジャパン (株) 医学統計室 | 64 松本一彦 東洋醸造 (株) 医薬品研究所 |
| 30 富安泰山 ファイザー製薬 (株) | 65 林 真 国立衛生試験所 |
| 31 村岡仁之 ファイザー製薬 (株) | 66 佐野正樹 大鵬薬品工業 (株) 創業センター |
| 32 今井節夫 財) 動物繁殖研究所 コンピュータ室 | 67 高橋行雄 日本ロシュ (株) 統計解析室 |
| 33 大町賢吾 天藤製薬 (株) 研究部研究科 | 68 下井信雄 (株) ユックムス |
| 34 坂口信樹 参天製薬 (株) 中央研究所安全性研 | 69 長谷文雄 日本ルセル (株) 研究開発本部 |
| 35 山下哲司 ロート製薬 (株) 第一製品部 | 70 今溝 裕 東洋醸造 (株) リサーチセンター |

- | | | | |
|---------|-------------------|-----------|------------------|
| 71 柴野隆司 | (株) イナリサーチ 安全性研究部 | 95 石塚修司 | エスエス製薬 (株) |
| 72 大塚方正 | 日本ロシュ (株) 開発本部 | 96 稲葉太一 | 神戸大学教養部数学科 |
| 73 三内貞子 | 日本生物科学研究所 | 97 松浦哲郎 | 摂南大学薬物安全科学研究所 |
| 74 中里薄志 | 雪印乳業 (株) | 98 渡辺 徹 | クミアイ化学工業 (株) |
| 75 清水達也 | (株) イナリサーチ | 99 黒田隆英 | 湯浅商事 (株) |
| 76 川東正美 | 摂南大学薬物安全科学研究所 | 100 大林繁夫 | グレラン製薬 (株) |
| 77 川瀬芳克 | 愛知県総合保健センター | 101 五十嵐真一 | 中外製薬 (株) |
| 78 松本道治 | 中央日本バイotech研究センター | 102 永見俊之 | 日本農薬 (株) |
| 79 鶴亀隆之 | 鐘紡 | 103 山縣周三 | 呉羽化学工業 |
| 80 伏見恵文 | 厚生省統計情報部 | 104 高橋朋彦 | 呉羽化学工業 |
| 81 清水時彦 | 厚生省統計情報部 | 105 北村知宏 | 山之内製薬 (株) |
| 82 滝沢 博 | 協和醗酵 (株) | 106 山田雅之 | 富士レビオ (株) |
| 83 山田英樹 | 興和 (株) | 107 西山祐二 | カストルメイ-スクイ研究所 |
| 84 諫山真樹 | 中外製薬 (株) | 108 稲村修司 | 成和実験動物 |
| 85 木船義久 | 武田薬品工業 (株) | 109 河上喜之 | 実中研付属前臨床医学研究所 |
| 86 玉田輝巳 | コンアルコン (株) | 110 石垣智子 | 東京大学 医学部 |
| 87 本多紗絵 | (株) エスアールエル | 111 畠山邦彦 | 藤沢薬品工業 (株) 安全研 |
| 88 小林 | 雪印乳業 (株) | 112 松尾朱美 | ファイザー製薬 |
| 89 田代勝文 | 雪印乳業 (株) | 113 加藤秀成 | サントリー (株) 医薬センター |
| 90 玉井功一 | (株) 保健科学研究所 | 114 五十嵐俊二 | ユーザイ |
| 91 金子泰久 | アッゴソフ-マッシュウイカリミッド | 115 宮本政樹 | 日本チバガイギー (株) |
| 92 松井 一 | アッゴソフ-マッシュウイカリミッド | 116 安藤正一 | 東京理科大学 |
| 93 吉田和彦 | ヘキストジャパン (株) | 117 滝沢 毅 | 日本ロシュ (株) 統計解析室 |
| 94 山森 芬 | サントリー (株) 医薬センター | | |

〔事務局だより〕

5月23日～5月25日、既報のように「毒性研究のための計量生物国際会議」が東京大学山上会館で行なわれました。会議では文字通り活発な質疑が行なわれ、大いに盛り上がったのではないかと思います。会議の総括と opening address などは次号に掲載致します。また、中味については、abstracts を参照して下さい。充実した中味です。小社にとって国際会議は大イベントでしたが、その夢中の体験を、これから大いに展開していきたいものです。

今号では、二本のディスカッションを掲載できました。いずれも興味深い討論が次々に生起しているのではないかと思います。ほんとうはこの種の議論を本会報にどんどん載せたいのですが、テープ起こし原稿の整理は、難物で、毎回、非常に苦勞しています。言葉を正確に聞きとること、話をつなげたり、指示代名詞の指示するものを確定したりと、編集技術が試されます。それだけにまとめられたときの原稿は刺激に充ちていると感ぜられるのです。(M.O.)

〔お詫びと訂正〕

*前号33号の「Q & A Williamsの多重比較法について」の中に誤りがありましたので訂正致します。p27 ↓ 16 $t_i = \dots$ の数式の右辺、 $(\mu_i - \mu_i)$ は、 $(\mu_i - \mu_1)$ の誤りでした。

医薬安全性研究会 会報NO. 34

Bulletin of Japanese Society for Biopharmaceutical Statistics.

1991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