

医薬安全性研究会

会報 No.18

Oct. 1985

目次

毒性学者のための統計学<最終回>	シャニー・C・カッド カロール・S・ウェイル	1		
初めて学ぶ統計学(1)-データを生かす視点-	吉村功	12		
第24回定例会出席者名簿	19	事務局だより	20

毒性学者のための統計学

シャニー・C・カッド

カロール・S・ウェイル

最終回

訳：加藤 正巳
トーア エイヨー(株)

優性致死分析

優性致死試験は本質的に、着床後から分娩前までの胎仔の致死限界点を研究し、標本単位の正確な確認と実験計画が重要なセクションである。それは相対的に答えられていない。ここでは、標本の大きさ(例えば1週間1用量あたり30から50の妊娠雌(5))であらば、平均の変動性および分布の性質は鋭敏な統計手法が使用できるようになる。

典型的に記録され解析される重要な変数を次に示す(それぞれのレベル/週): (a) 妊娠雌数, (b) 生存胎仔数/妊娠数, (c) 総着床数/妊娠数, (d) 早期胎仔死亡(早期吸収)/妊娠数, および(e) 後期胎仔死亡(後期吸収)/妊娠数。

これらのデータの解析には種々の手法が用いられてきたし、現在も用いられている。最も共通に用いられるのは、データを角変換(arc sine transform)によって変換した後、ANOVAを実施することである(51)。

ベータ二項(2, 70)およびポアソン分布(16)もまたこれらのデータの属性と考えられる。そして変換および特定の検定が、これらそれぞれの場合に用いられるよう提案されていく(それぞれの場合に、変換はデータの「分散を安定させる」の役に立つという特徴とともに)。十分な大きさの標本では、このセクションの前の方で明らかにされたように、Mann-WhitneyのU検定がここで使用に推奨されるべきである。より小さな標本では、Wilcoxonの順位和検定の使用が必要となるであろう。

角変換 (二項分布するようなデータにつかえる)

角変換した後、ANOVAを実施することもある。
(生存胎仔数/着床数、死亡胎仔数/着床数、等のデータ)

	1	2	...	n
対照	X_{11} / N_{11}	X_{12} / N_{12}	...	X_{1n} / N_{1n}
処理	X_{21} / N_{21}	X_{22} / N_{22}	...	X_{2n} / N_{2n}

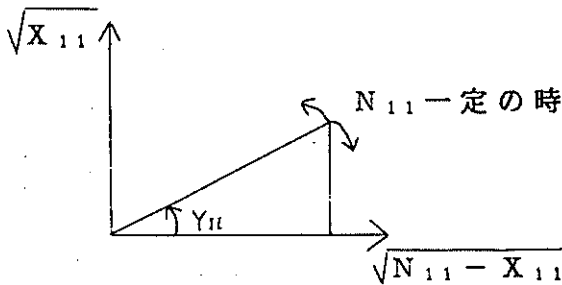
但し $X_{ij} \leq N_{ij}$

上のようなデータがある時、分散を安定させるため、

$$Y_{11} = \sin^{-1} \sqrt{X_{11} / N_{11}} \quad \text{とする。}$$

	1	2	...	n
対照	Y_{11}	Y_{12}	...	Y_{1n}
処理	Y_{21}	Y_{21}	...	Y_{2n}

但し、 $N_{11}, N_{12} \dots$ が違う時は基本的に問題あり(分散安定しないため)
受胎数は母体により違うので角変換を使う意味がない



$$\sin Y_{11} = \sqrt{X_{11}} / \sqrt{X_{11} + N_{11} - X_{11}}$$

- 単調な変換 X_{11} が大きくなると Y_{11} も大きくなる
- N_{11} が大きいと角度の分散小さくなる。
- 角変換した時の平均 → もとに戻したとき多少はずれるだろう。
- 実際には変換する必要はないのでは

○断定的な事はまだシュミレーションの量が少ないのでいえないが、最終的にはANOVAとt検定が良いらしいと言えそうである。今後、もう少し色々な場合についてやってみて最終的な結論をだしたい。

○角変換はかつて等分散性成り立つようなテクニックとして考えられたもので、サンプル大きい時にはあうが、実際的な乱数実験ではうまくいかない事がわかってきた。

ベータ二項分布

$$P\{X=x\} = \int_0^1 \binom{n}{x} P^x (1-P)^{n-x} \frac{P^{\varphi_1-1} (1-P)^{\varphi_2-1}}{B(\varphi_1, \varphi_2)} dP$$

$$= \frac{\binom{n}{x}}{B(\varphi_1, \varphi_2)} \int_0^1 P^{\varphi_1+x-1} (1-P)^{\varphi_2+n-x-1} dP$$

↓
ベータ関数

○一般の二項分布に比べると、母出現率 p が β 分布の形でゆらぐ。

○ベータ二項分布は、優性致死試験のデータの属性と考えられる。しかし、これに角変換を用いることはできない。

質問 二項分布、ポアソン分布に関する変換及び検定法とはどんなものがあるか。

答 二項分布→角変換
ポアソン分布→平方根変換 これらの変換を実施したのち通常の検定を行なう。
しかし、シュミレートすると必ずしも最適な方法とはいえないだう。

質問 Mann-WhitneyのU検定とWilcoxonの順位和検定を適用する際の違いがあるのかどうか。

答 Mann-WhitneyのU検定とWilcoxonの順位和検定は、まったく同じ式の別の表現である。

優性致死試験 (高木博司 監修 薬物の特殊毒性-1 胎児毒性と遺伝毒性 高野喜一ら著 南江堂 より)

化学物質の突然変異性検定のための試験法のひとつ

優性致死：精子または卵子どちらかに由来した遺伝物質に関して、ヘテロ接合型の受精卵(接合子)でも、発生の過程で死に至らしめる障害

常染色体の突然変異遺伝子の致死効果は通常それらに関してホモ接合型の場合に現われる。

優性致死の原因は配偶子の染色体の切断または再配列による受精卵の割球の染色体構成の異常、卵割の進行の異常と考えられている。

一般的な方法

被検物質を雄に単回投与又は7日間連続投与を行ない、無処置の雌と交配する。雌を更新しながら6週間(マウスの場合)* 交配を続ける。雌について妊娠12-13日ころに開腹、黄体数、子宮内の着床数、生存胎仔や死胚数(後期死胚、早期死胚)を求める。

* 精子形成期間に相当する

雌では、感受性が低い、卵の細胞質への影響、卵子の受精前死亡、子宮内膜へ与える障害、過剰排卵誘起の可能性など不都合な要因があるため雌のみに投与する事は一般に行われない。

吉村先生のシュミレーション

血液検査データ 10 個体について細胞を 100 個計数した時の出現数

0/100, 0/100, 1/100, ..., 1/100, 2/100 ... 10 個体

対照 0/100, 0/100, ..., 0/100, 1/100

処理 0/100, 0/100, ..., 1/100, 1/100

H_0 : $n p = 0.5$ (%) で二群とも同じ

H_1 : 対照 $n p = 0.5$ (%)

処理 $n p = 1.0$ (%) と違うばあい

上のふたつの条件でそれぞれ、t 検定、角変換したのち t 検定、順位和検定を 10000 回コンピューターにより乱数を発生させたデータについて実施した。

有意と判定された割合 (%)

	有意水準 5 %		有意水準 1 %	
	H_0	H_1	H_0	H_1
t 検定	5.86	36.3	0.87	15.2
角変換後 t 検定	6.31	37.2	1.24	17.6
順位和検定	6.28	37.4	1.21	18.6

○実際の有意水準は近似的な限界値を使っているのでピッタリにはならないが、まあまあ近い値になっている。

○検出力は角変換で大きいですが、有意水準くるっているのでこれがいいという根拠ない。

○実際の有意水準がちょうど 5 % 又は 1 % で検出力が大きいほうが良いしかし、有意水準をちょうどにすることができれば、どの方法も検出力は ほぼ同じになるだろう。

○へんな細工しなくとも(変換しなくとも)そのまま t 検定を実施してもかまわないのではないかと考えられる。

(訳) 養命酒製造(株) 中央研究所
小島 暁

食餌・気室分析

食餌試験においては、被験物質を飼料に混入することにより、動物に対する所定の用量を与えることができる。同様に、吸入試験においても、被験物質を試験動物の呼吸する空気に混入している。

両者において、被験物質が実際にどのくらい存在するのかを知るため、媒体としての飼料や空気をサンプル採取し、分析しなければならない。これらの投与された被験物質の濃度や、時間経過によってどのように変化するのかの正確な像を、持つことは、試験の基礎として必須のものである：

1. 統轄機関および厳密な実施規範は、分析された飼料および空気中の被験物質の濃度が設定水準の+10%であることを要求している。
2. 代謝系および修復系の過負荷を生じさせる著しいピーク濃度は、過度の急性作用を生じさせることになり、これは、慢性試験において被験物質の真の作用をあらわしていない見かけ上の成績を与えることになる。このような結果は、飼料や空気中の真の濃度を知っていなければ、誤解をまねくことになる。

サンプル採取の方法は、統計学的見地から“数”の問題ではなく、“幾何学”の問題である。従って、コンテナーの中身あるいは気室全体の空気を正確に、同時に、被験化合物の安定性にそって採取されなければならない。これらの方法や、最終的に集計されたデータの統計的解釈は、適当な専門家による十分な論議の後になされるべきである。代表例として、受容限界を超える値の同一性をともなう平均値や標準偏差などを求める計算が問題となる(8)。

広範に広がる値(はずれ値)を扱うことは、統計学の別の小さな分野を考えなくてはならない。ある時には、それが、はずれ値の統計的処理という主要な問題となりうる。Barnett と Lewisの間で、この問題について広く論議されている(4)。問題となるのは、予想されるものと明らかに異なる値であり、また、集計された一連のデータの他の値とも明らかに異なる値である。それらは、技術者の誤り、設備の誤動作、あるいは、見かけ上、動物の試験を失敗させようとする本能が原因として生じる。これらの原因の一つが、完全に、はずれ値と結びつけることができれば、この値を考察から削除することができる。しかし、原因とははずれ値を特定できなければ、かなり複雑な問題を抱えることになる。

このような値は、いろいろな問題の1つであろう。これは、試験全体にわたる基礎であるような特別な原因によることもありうる。あるいは、サンプルの誤差のある結果を画一的に集計することにもよりうる。あるいは、不注意な技術者の誤りなどによるであろう。後者の場合、はずれ値を削除し、残りのデータの考察から除外すべきである。しかし、明らかに統計学的には棄却すべきケースであることを、どのように見極めたらよいのであろうか。

データを棄却するには、様々な方法がある。その方法をいかに適切に利用するかは、データの分布の性格を理解することにかかっている。正規分布に対しては、Chauvenetの仮説(49)のような単純な方法が、よく採用されるであろう。平均からかけ離れた値の割合が $1/2N$ 以上ならば、他のデータとともに極めて慎重に考えねばならない。

質疑・応答

Q 1 : サンプル採取における“幾何学”とは？

A 1 : 飼料や空気の「どこの場所（三次元）から」サンプル採取をすべきかという意味であろう。

Q 2 : 飼料や空気の分析と“はずれ値の扱い”との関係は？

A 2 : 一般的な“はずれ値の扱い”の問題ではなく、この章では、あくまでも飼料や空気の分析結果における“はずれ値”について述べているのであろう。
つまり、食餌試験や吸入試験において、サンプルを分析した結果、はずれ値が多く見られた場合には、試験の管理が十分になされていないことを意味し、分析値の変動を抑えるよう慎重に試験されなければならないということが、述べられているのであろう。

Q 3 : Chauvenet の仮説の N とは？

A 3 : 例数か？

例ば、10 例とすれば $1/20$ 以上 か？

変異原性

グレラン製薬株式会社
研究開発本部
大林 繁夫

過去10年の間に変異原性試験のために幅広い方法が発展し、使用された。(有効な試験の概要は ref. 4 5 を参照) これらの試験は長期間動物全体を用いた試験に比し、興味ある物質が変異原性をもっているか、癌原性をもっているかを予測(決定ではない)するより早い、より簡単な方法として我々に与えられた。

このおびただしい数の試験(Ames, DNA repair, micronucleus, host-mediated, Chinese hamster ovary, sister chromatid exchange and Drosophila)の結果解析方法は新しく、非常に重要な問題である。現場の人達の中には統計的解析を成すことは不可能(不必要)と考え、ただ試験生物の突然変異発生の著しい増加を引き起こすかどうかをもとに陽性か陽性でないかを簡単に判断できると考えるものもいた。これはすぐに受け入れられるものではない。

現在我々の関心は多細胞生物が多くの複雑な役割の中の1つにどのような反応を起こすかということから、変異が一つの単細胞によって最も高頻度に現れる方法へと移っているため、我々が先に述べた事がらとは違った概念が検査されるのに必要である。より小さな試験生物を用いたこれらの方法で非常に多くの個体を取り扱うことができれば弱い変異原性も強い変異原性も両方検出することができる。

正しい統計解析を用い、その解析の結果を正しく用いるには変異を起こした生物の知識とこの知識から適切なモデルと仮説を発達させなければならない。我々は相互に影響し合う5つの要因のそれぞれを考慮してその過程を始める。

1. α , 第1種の過誤の確率(変異原性がないのにあるということで、Fisher's exact test のように初期に考えられた計画の p に等しい) ;
偽陽性
2. β , 第2種の過誤の確率(変異原性があるのにないということ) ; 偽陽性
3. Δ , 分析にてのぞまれる感度(1つの集団の突然変異でその10%の増加が検出できるというような)
4. σ , 生物系と偶然誤差の影響のバラツキ
5. n , これらのおのおのを達成するために必要なサンプルサイズ(我々は唯一等式の中のこの部分だけ自分の力で変えることができる)

$$n \text{ は } \frac{\sigma}{\alpha, \beta \text{ and } \Delta} \text{ に比例する。}$$

したがって、これは σ が大きくなれば n が大きくなり、 α , β , Δ の望ましいレベルが達成され、 α , β , Δ のすべてあるいはどれか1つが小さくなれば n は大きくなるということを示す。そして常に σ がどのくらいであるかを知っておかなければならない。バックグラウンドの変異レベルと自分の技術によるバラツキはどれくらいか？ 優秀な遺伝毒性学者あるいは一般毒性学者が認めるには対応する対照群が必須である。運よくこれらの試験方法で多数の n が簡単に得られたとしても、なおかつ後述べる他の複雑な問題が存在する。これらの方法の中の異なった方法で生じた混乱の1例はWeil (76)に紹介されたような今までの実験室内での比較で例証される。

これらの仮説と母集団の分布をもとにして、入力される変数(α , β or Δ)の1つをそのままに、1つを変更するために必要なコンピュータのバックグラウンドとともに新たな統計手法が提案された。特種な約束のもとに1組の統計手法が2つの論文の中でKatz (43, 44)によって提案された。彼は2つに分けて試験の統計値を提案した。 ϕ (実験群と対照群の個体の数を正確に算出できるときのため)と θ (おのおの群の生き残った個体の数を実際に算出しないで、そして被験物質が試験生物を殺すまでの間においてマイルドな毒性であることを予測できるときのため)これら2つの統計量はそれぞれ試験生物の

被験物質へのただ1回の暴露(1 dose の試験)をもとに式で示される。

$$\phi = \frac{a (M_E - 0.5) - Kb (M_C + 0.5)}{\sqrt{Kab (M_E + M_C)}}$$

これが得られたとき計算できる。a と b は対照群と試験群の数である。

N_C と N_E は生き残った微生物の数である。

$$K = N_E / N_C$$

M_E と M_C は試験群と対照群の突然変異の数である。

μ_e と μ_c は本当の(しかし未知である)突然変異の割合である。

(μ_c が小さいとき、 N_s は増加しなければならない)

2番目のケースも計算できる。(同じ意味の要素を用いて)

$$\theta = \frac{a (M_E - 0.5) - b (M_C + 0.5)}{\sqrt{ab (M_E + M_C)}}$$

両方のケースとも $\alpha = 0.05$ の信頼限界で、もし試験統計量(ϕ or θ) が1.64より小さければ $\mu_c = \mu_e$ を承認する。もしそれが1.64より大きいか等しければ、変異原性をもっていると結論する。($\alpha = 0.05$)

2番目のケースにおいて、(θ , 対照群と試験群で population size の分離した量を得られないとき)もしKが1.0から遠く離れたならば(もし被験物質が Markedly toxic であったならば)対照群の個体数よりも試験群の個体数を大きくとるべきである。(異なった生存頻度のときに使うそれぞれの比率の表は ref. 44に示される)もし違う α の値が欲しいならば、 θ と ϕ の表は ref. 42に示される。

この結果生存細胞に対する変異の比率が決定される。もし対照群の変異率が、被験物質によって変異率の減少が起きるほど高ければ、これらの試験統計量は両側の仮説を考慮しなければならないことを記憶すべきである。もし処理したいものが確実な信頼度で変異原性をもつことが確かめられるなら、(これは帰無仮説を否定することとは違う)、 β の値はおのこのケースにおいてあるいは試験において同様に調整してよい。

おわりに

ここで述べられている統計量 ϕ は二項分布の正規近似であり、すなわち二項分布を用いた解析方法であることがわかる。また θ は ϕ を求める式における K を1としたときの統計量であり、対照群と試験群で生存細菌数が等しいとしたとき(試験物質の毒性がほとんどない)と考えることができる。

上記の解釈において、セミの中でも「統計量としては ϕ だけで充分であり、 θ は必要ではないのではないか?」という質問を受けたが、私の考えとしては、 θ は ϕ の簡りやく法であり、 K を実際に求めることが難しいためわざわざ紹介したのだと思う。 K の求め方については、原論文を読んでもらいたいが、残念ながら私には方法論は分かったものの、理論的な解釈については理解できなかった。

初めて学ぶ統計学 (1)

— データを生かす視点 —

吉村 功

—— 吉村先生にはいつもいろいろとお世話いただいているのですが、さらに30分、時間をいただいてしまうということで、今回新しい企画をいれました。どうしてかという、私は今までの例会の中で司会をしながら内容についての質問をして少しずつでも話を理解しようとしてきたのですが、私があまり基礎的なことを聞いていると話が全然先へいなくなってしまうので、あるところで質問を打ち切ってしまうのでした。そういうことをくり返していたものですから、このごろではいつか統計なり確率なりの基礎的なことについて、もっときちとしたイメージを持ちたいと思うようになってきているのです。

例えば、さきほど先生が“確率変数 X が”という話をされましたが、これが実は私にはよく分っていないのです。しかし、そこを質問していくと初歩の話になってしまい会としての進行ができなくなるので、飛ばしてしましたのです。同じようなことで、私がいつも気になっているのは、確率変数、確率密度、それから二項分布、ポワソン分布、自由度などの言葉です。分ったようにはい、はいといっていますけれどもあまりよく分っていないのです。

例えば、二項分布というのもサイコロとかトランプでは教科書に書いてありますけれども、それが動物実験の場ではどういうモデルで、どこか二項分布になっているのかということか、あまりよく書いてありません。それです、と疑問になっているのです。そういうことを1項目ぐらいずつ先生に解説していただき、私だけでなく会員の方にも確率、統計の話がよく分かるようになればいいと考えたのです。

どういようになるか私にもよく分からないのですが、少しずつ吉村先生にお話をお願いしたいということで、こういうコーナーを設けてみました。今回は最初ということで、先生が朝日新聞で5回に渡って連載されたものを材料にして、「データを生かす視点」ということでお話をお願いしたいと思います。

吉村 今日少しお話し的なものにしましょう。定例会に出席されている方は統計学の専門家ではないと思いますが、そういう方が統計学を学ぶルートにはかなり異なる、二種類のものがあります。一つは、大学や高校などで学び始める人が大体そうですが、理論から入るというルートです。そこでは確率分布とはそもそも、という話から始まります。そして統計手法を現実にとどう使うかというところまでいかないうちに、講義が終わってしまいます。もう一つはいやおうなしにデータを処理しなくてはいけないということで、統計手法を使ってみる。しかしさ、ぱり何をやっているのか分からないので教科書を見て勉強するというルートです。このルートではまだ納得できないというところが、沢山あるままになりがちです。

ここにいらっしゃる皆さんはほとんどが後者だと思えますので、今日は後者の場合に引っかけるところを主にして話していこうと思えます。ここに新聞のコピー^{※1}がありまして、

『データを生かす視点』というのが 1, 2, 3, 4, 5 までありますが、今日は5番目の話をしようと思います。

I. 同じデータに対し、異なる手法がある

同じデータ、同じ問題に対して異なるいくつかの手法があるときに、どの手法を使えばいいのかということです。皆さん必ず一度はこれに引っ掛かっているのではないのでしょうか。例えば平均値の差の検定というと、t検定、ウエルチの検定、コックランの検定、ワイルコクソンの検定などいろいろなものがあります。自分の問題では、そのどれを使えばいいのかということです。これらの手法がすべて同じ結論に達するのであれば、好きなものを好みに応じて使えばいいわけです。ところが、方法によって結論が違ってくるので困るのです。その場合にどれを使ったらいいかというわけです。それを判断するには、なぜそのような違いが出てくるのかを心得ていたほうがよいようです。同じデータに対して違った手法が存在するのにはどういう場合があるのかです。

§1. 簡便法と効率のよい方法

第1番目は簡便法と効率のよい方法がある場合です。例えばデータの数が非常に少ないときに平均値はいくらか推定するのに、平均を計算するのと中央値を計算するとの二つの方法があります。正規分布の場合、中央値のほうが簡便法で、平均をとるほうが効率のよい方法です。データが5個くらいしかなかったら、5個を足して5で割るよりは、真ん中の値で平均はこれくらいだろうと見当をつけるほうが楽なのです。簡便法はちょっとした見当づけのためには非常に有効です。私も他人のデータを100と見てすぐに結論をくださなければならぬときに簡便法をよく使います。この場合にはあまり選択を迷わないと思います。

§2. 近似的な方法がいくつかある

第2番目は検定を行なうとき棄却限界値をきちり決めることができないため近似的な方法がいくつかできる場合です。例えば2x2の分割表のカイ二乗検定で連続修正をやるかどうかといったものです。あるいは、分散が違ふときの平均値の差の検定でウエルチの近似を使うか、コックランの近似を使うかです。棄却限界値の近似の仕方が違ふだけなのです。この場合にはどちらの近似がいいかということを見るのですが、世の中に通用してきたものは、大体近似が一長一短なことが多いので、細かい知識をもって場合わけをするより仕方ありません。しかし、この場合は結論がほとんど変わらないことが多いと思います。どちらかの近似がいいという場合には5年から10年モうちには片方はたんだん使われなくなります。

§3. 対立仮説が違ふ場合

第3番目は対立仮説が違ふ場合です。一番極端なものは片側検定をやるか両側検定をやるかということです。これはいまだ別の機会に話そうと思います。今日はこの話はしません。

注1) 「五話、朝日新聞名古屋版 昭和60年1月12日～2月9日(毎土曜日)

- ① 因果関係の証明 ② 「危険評価」の確率 ③ 計算機時代の心得
- ④ 表面からの内部推定 ⑤ 解析の自由度

§4. データをどこまで一つにまとめるか

第4番目にデータをどこまでひとまとめにするかによって方法が違ふ場合があります。例えば、多くの変数があるときに多変量解析でやるか一変量を別々にやるかです。二つの薬をそれぞれ共通のコントロールで実験をやるたびに全部ひくくめてやるか、薬を別々に実験しておいてコントロールだけは同じものを使っているときとみるわけです。こういう場合、多少結論が違、てくることがあります。これについても別の機会に議論したいと思ひます。

§5. ロバストな方法とセンシティブな方法

第5番目はロバストな方法、鈍感な方法とセンシティブな方法、敏感な方法とがある場合です。敏感な方法と鈍感な方法とでは敏感な方法のほうがいいと思ひれるかもしれませんが、実はそうではないのです。敏感な人間とつきあうのが都合がいい場合と鈍感な人間とつきあうほうがいい場合とがあるのと同じです。例えば人間の鼻で悪臭を検出しようという場合には敏感なほうがいいわけです。しかし毎日暮らしていてあまり敏感ですと、ノイローゼになってしまいます。こういう場合には、鈍感な人のほうがいいでしょう。統計手法もこれと同じように、時と場合によってどちらがいいかが変わります。敏感な方法はちょっとした条件のずれに対して、ものすごく悪影響が出やすいのです。例えばデータについて一個計算間違ひをしてちょっと値がずれただけに結論が10%と変わることがあります。これは必ずしも望ましい性質ではありません。これについてもいずか議論をしましょう。

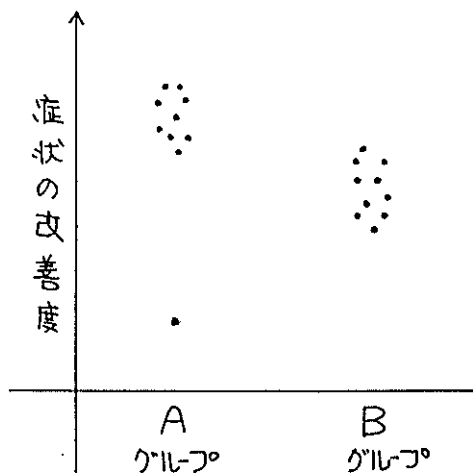
§6. 前提となる真の状態に応じて方法が違ふ

一番わかりやすいのが6番目で、前提となる真の状態に応じて方法が違ふ場合です。つまり、一つのデータが正規分布に従う場合にはこちらの方法がいいという場合です。分散が小さいときは検定がよく、分散が違ふときはウェルテ検定がよるいいというときもあります。つまりデータの性質がこちらのときはこちらのほうがよいという場合です。こういうときが一番わかりやすいのです。今日はこれに関係した話をしましょう。

II. 困難さを示す例

自然にはある種の“真の状態”というものがあつて、その中にはわれわれがすでに知っている部分と知らない部分とがあります。知らない部分のうちにも、今のデータで判断を下そうとしている部分と、今のデータでは分らない部分と、今のデータに影響しているか、実はよく分らない部分とがあります。ここで問題になるというのは、今のデータには影響しているけれども、本当はどうであるかがよく分らないという部分です。これ次第でデータを解析するための方法が違ふというのが一番困る場合です。そうしたときでも方法によって結論が違わなければ全然問題はないわけですが、方法によって結論が違ふという場合が困るのです。

例えばいま20人を二つに分け、片方にはAという薬、片方にはBという薬を与え、飲み方を教えて渡して、何日後かに症状がどうなったかを見たとします。そうしたところ、図1に示すように、例えばAグループの10人に関しては9人くらいは症状がよくなつたけれども、1人はちょっと悪いのが出てきた。Bグループにおいては極端にはなつたのはなくて、全体



としてAグループの9人に比べて少し下がっているという現象が出たとします。こういう場合に皆さんならどうするでしょうか。

このデータに対しては二つの場合が真の状態として考えられます。一つは患者が「食後30分後に飲みなさい」といっておいだにもかかわらず飲まなかった、言い付けを守らなかったという場合です。実際に私などももらった薬をまともに飲んだことは一度もないですから、真の状態としては十分ありうることです。

もしこれが真実ならAグループの最も低い人のデータは除かなければなりません。そしてAの1人がなくなるとすると、Aは9個のデータで、Bは10個になります。9個と10個では少し

不公平だから、Bの一番低いデータも捨てます。両方とも低いデータを捨てるのだから公平になるでしょう。同じ9個で調べると、どうみてもAグループのほうがいいわけですから、Aという薬がよいという結論に達します。ところがAの1人が本当に真面目に飲んだというのが真の状態のことも考えられます。このときには、どのデータも除いてはいけません。すべてが本当のデータになります。つまりAという薬は大体の人にはよく効くけれども、ごく少数の人にはものすごい悪影響があるということになるわけです。

有名なサリドマイドがそうなのです。臨床実験のデータを見ましたか、大抵の人にとっては特に副作用なくよい睡眠薬でした。ところが、10人とか20人の中で1人か2人だけは、めまいがするとか、頭痛がするとかの副作用を訴えていました。20人の中で1人かせいぜいでも2人ぐらいでした。ですからみんな特別なのだと思って無視してしまい、基本的には非常にいい薬であるという結論を出してしまたのです。ところがあとになってみたら、末梢神経炎という症状が次々と現われて問題になったわけです。

そのように大抵の人にはいいのだけれども、ごく少数の人には非常に悪いということは十分あります。そうすると結論として大勢の人に対してはBよりいいけれども、少数の人にはBより悪いという結論を出すべきデータであるということになります。問題はそのどちらが真の状態であるかですが、このデータだけからは、どちらが真の状態であるかを判断するのは不可能です。そうした場合は手法選択において最も困難な場合になります。

今のは非常に極端な例ですけれども、先ほどいって分散が違うかどうかとか、正規分布か指数分布かというのも基本的には同じことなのです。真の状態が二つあり得て、それが手法の選択に影響を及ぼす。そしてどちらかの手法を使うかによって結論が違う。それにもかかわらず、真の状態が未知のために手法を選べない。こういうことです。これはある意味では統計学の宿命の困難です。どんなデータ解析をやっても、この問題は解決できません。こういう現象に関しては基本的にはまず解決不能だと悟りを開かなければいけません。

では悟りを開けばあとは好き勝手なことをやっていこうかという、そうではありません。そんなことをすると大変なことが起ると、怖いわけです。こういう問題に対しては、ある種のセンスを身につけてそれに対処していくしかないと思います。私などは、偉そうなことを言う柄ではないのですけれども、自分の身をなるべく安全に置きたいですから、このようなことをいこうとしております。

Ⅲ.方法によって結論が違ふときの対処法

§1.安全側—石橋主義

今述べたような事情で、方法を選ばないのに方法によって結論が違ふというとき、現実には存在している対処法は五つあると思います。一つは安全側を考へて石橋を叩いて渡る主義です。つまり両方の方法を適用しておいて、世の中の人たちにとって安全な側に結論を出すというやり方です。

例えば「副作用がありそうだ」というときには、あるかもしれないという結論を出します。薬の効果などという場合には、もしかしたら効かないかもしれないというほうに結論を出します。そうすれば本当に効かない薬、あるいは本当に副作用をもつ薬を売り出すことはないでしょう。安全側であるかわりに積極的には何もしないこととなりますから損をします。

§2.両論併記—あなたまかせ主義

2番目は ああであるかもしれないし、こうであるかもしれないという両論併記です。先ほどのデータでいうならば、もしかしたら少数の人に非常に有害であるかもしれないし、もしかすると患者の1人が言いつけを守らなかつたデータかもしれないということになります。もし患者が言いつけを守らなかつたためだつたならば、この薬はいい薬である。もしみんなが真面目であつたならば、この薬は少数の人には非常によくない薬らしい。こういう結論を両方並べて書いておいて、あとはお読みになつた方に適当に考へてもらうというやり方です。これは両論併記のあなたまかせ主義です。

§3.主観的に判断

3番目はそんな無責任なことはいってはいけない、やはりデータ解析をする人間の責任においてすべてを処置しようというやり方です。私はどうみてもあの患者はルーズな患者だと思つたから、あの患者は嘘をついているに違ひないから、あのデータは切り捨てるということです。あるいはこの薬はやはりそういう特徴があるかもしれないだろう、だからこの薬は危険であると私の責任で判断しますということにもなります。これはある意味では主観的な判断です。ただしこの主観的な判断の場合でもそれなりの根拠がなければ危険です。希望的観測だけで判断されると非常に怖い。だから主観主義とはいつても客観的な根拠を見つける努力をしなければいけないと思います。

§4.ルールを事前に固定

ルールを事前に固定しておくというやり方もあります。例えばこれは平均値の差の検定しかやらないので、どんなデータが出てもそれしかやらない。あるいはウエルコフソンの検定しか絶対にやらないということあらかじめ決めておくわけです。そうして何かでてきてもそれと頑張りわけです。これは、純粹客観主義です。統計学にネーマン・フィッシャー論争というのがあります。ネーマン先生はどちらかということに近いことを主張しました。それに比べるとフィッシャー先生ははるかに主観主義に近いことを結果としては主張したことになります。

§6. 完全主義—再実験をする

それからももう少し優等生的なものは、そういうときは再実験して必ず新データをとる。わからないところはちゃんとやり直して調べなければいけないという完全主義です。

現実に行なわれているのは、こんなところだろうと思いますが、この五つの対処法の中のどれがよくてどれが悪いかということは、私にはいえません。状況に応じて違うと思います。先ほどいったような、薬をいよいよ売り出すというときには、危ないですから1番目の方式が重要だと思います。それに対して研究の初歩的な段階でこれからたくさん研究するのだというときには、なるべくていねいに情報を提供し、これからの処置にそなえるという2番目の方法がいいと思います。

また、一般にいろいろな状況を一番よく知っているのは実験者です。あの患者はどういう人間であるのかも知っていて、この間もいろいろ加減なことをやめたからこの例は間違いない、いろいろ加減だろうとか、ほかの人たちを見ていても特に異常は起こっていないのではないかとか、いうことを実験者がきちんと観察した上で、何か問題があっても責任はちゃんと取るぞというのであれば3番目でもいいと思います。わか社の趨勢からいうと、無責任に儲かることさえあればいいという雰囲気があるときであれば、逆に4番目でがんばらなければいけません。希望的観測ばかりで結論を出さずには危ないですから、このほうがいいと思います。

再実験をして新データをとるのが自分でできるものであれば、一番望ましいのは5番目です。そのときには前に起こった現象をちゃんと確かめる実験を組めばいいのです。自分でできるならこれが一番いいだろうと思います。データ解析家や実験担当者は、自由自在に誤りなく、このような判断で対処するというセンスを持ってほしいと私は思います。

IV. 他人のデータを読むとき

これはデータを解析する立場で考えたことですが、世の中にあるのは解析する立場だけではなく、他人の結論を読む立場というものもあります。皆さんもデータ解析をいつもするわけではなくて、他人のデータ解析を読む立場に立つことがきっとあると思います。そういう立場に立ったときに、今のような現象があるということを心得て下さい。統計手法そのものは非常に客観的なものでありながら、使い方次第でまったく同じデータに対して逆の結論を出せることがいくらでもあるのです。そしていくつかの手法があって結論を出したがるのが人情です。これを注意して下さい。

先ほどいった1から5までの選び方は基本的には優等生の話で、わけわけ凡人がいつも優等生的にふるまえるわけはありません。誘惑にかられて、自分の好きな結論を出すという可能性が十分あります。そういうことを相手に対して考えたいほうがいいと思います。ですから、相手にとってあまり都合の良い結論のときにはマユにツバをつけて話半分に聞く必要があります。どんなに客観的に見えたとしてもそうするというのが、他人のデータ解析をみるときの一つの心得だと思います。統計手法を使って客観的な結論を出したのだからといっても、今のように解析に自由度がある限り危ないのです。もちろん非常に極端な場合はどんな方式を使っても結論が同じになり、そのようなことをしなくてもよいのですけれども、先ほどの話のように1人だけ変なデータが出たという微妙な話も現実にはいろいろあります。ですから、あまり都合の良い結論というのは危ないと思ったほうがいいと思います。

話半分に聞くという態度ではなく、他人のデータは自分で吟味するか、他人に追試させたものでなければ信用しないという対処法もあります。全然異質の人間、どちらかといえばあそこには負けたくないという対立する人間が追試して、同じ結論が出るのなら信用してもいいと思います。さもなければ可能なかぎり自分で追試する努力をし、それまでは多少マユにツドをつけてみておくわけです。

その人間にとって都合の悪い結論が出ているのであれば、インチキしていることは考えられないわけですからいいのです。あまりにも都合の良いすぎる結論が出ているときだけは、データ解析がいろいろと私は思っています。

V. 統計学の有効性と限界

統計手法を使えば非常に客観的で、科学的で間違いがないということが一時期強く主張されました。戦前のデータ解析は主観的で、何々先生かうんと言うか言わないかで結論が決まるという非常に危険なものだ、と面がありますからこれに対する反省として客観主義が強く主張されたのは一理あります。推計学という言葉はそういう状況の中で存在意識を強調されました。それはある意味で実際は重要な役割を果たしました。戦後、薬効検定などがずいぶんうるさくかわるようになって、最近のようにある程度のレベルまで達したのには、そのような風潮の影響があると私は思います。

ところがこのレベルになると逆の部分も出てきています。つまり統計手法は万能であり、これは何々検定という統計手法でやりましたから間違いありません、ということで結論を無理に押しつけるという状況が生じてきているのです。実際の統計手法というものは、それほどの万能性はないのです。

統計手法あるいは推計学の有効性と限界をかなりは、きり意識しておいてもらわないと間違、た結論を出してしまうことになります。間違、た結論を世間が認めて、誤解する危険が出てきます。一時期ありましたが、味の素を食べると頭がよくなるというタイアの妙な結論を出してしまうということになるのです。注意していただきたいと思います。

—— 今日のお話を踏まえて、また次回も確率の基本的なところを話していただこうと思います。先生もお忙しいのでいつもというわけにはいかないかも知れませんが、どうも、こういう30分コーナー的なことは今後もできるだけ続けていきたいと思っております。

第24回定例会出席者名簿

日時：1985年10月12日

御出席頂いた先生方

場所：グリーンホテル 2F会議室
11:00~12:00 基礎講座
13:00~17:00 定例会

• 吉村 功 (名研大) • 中村隆太郎 (日本ダリ)
• 柳本武美 (統計数研) • 吉田秀雄 (大塚製薬)
• 大橋靖雄 (東大病院)

- | | | |
|--|--|--|
| 1 井野裕子 (日本ロシ) | 28 鶴田 (興和) | 55 田中光男 (田辺製薬) |
| 2 相馬義徳 (") | 29 本多一裕 (津村順天堂) | 56 坂巻政次 (臨牀科特研) |
| 3 内田 (") | 30 内山 武 (") | 57 樋口史郎 (わかもと製薬) |
| 4 植村昌平 (ヒラマ薬品) | 31 三上正秋 (豊 ^洋 ウツクス) | 58 瀬谷雅行 (") |
| 5 藤田 (") | 32 栗山敏治 (東洋醸造) | 59 山岡秀明 (住友化学工業) |
| 6 高木 悟 (ヒキスシリン) | 33 " (") | 60 高橋行雄 (日本ロシ) |
| 7 武政俊彦 (セリア新薬工業) | 34 佐野正樹 (生技研) | 61 下井信夫 |
| 8 中島信明 (残留農薬研) | 35 田中 健 (日本 ^天 糖 ^研) | 62 半田 淳 (日本化薬) |
| 9 渡辺敏彦 (科研製薬) | 36 倉元達郎 (生技研) | 63 今溝 裕 (東洋醸造) |
| 10 長谷文雄 (ケシラン製薬) | 37 小笠原定則 (肉 ^果 医師製薬) | 64 佐々木 (") |
| 11 中嶋康彦 (キッコマン) | 38 聳城 豊 (ライオン) | 65 永見俊之 (日本農薬) |
| 12 大導寺俊平 (科研製薬) | 39 関 康弘 (") | 66 山口龍一 (三菱 ^油 化薬品) |
| 13 阿部俊一 (シドリナ) | 40 桑山典之 (帝國 ^蔵 器製薬) | 67 吉田 剛 (パー ^ン カー ^{イン} ケル ^{イク}) |
| 14 高橋昌三 (日本 ^{カイ} ヤ) | 41 鈴木 稔 (") | 68 小島 暁 (養命酒製法) |
| 15 佐藤七平 (臨 ^験 医学研) | 42 藤倉元起 (日研 ^化 学) | 69 奥富康雄 |
| 16 金子泰久 (ア ^フ フ ^ロ ビ ^ョ ン ^フ レ ^ィ ド) | 43 内田英男 (大正製薬) | 70 佐藤喬俊 (サ ^ン ド薬品) |
| 17 早川 潤 (ホ ^ン チ ^{セン} ター) | 44 大沢浩一 (") | 71 茂田哲哉 (日本 ^ダ リ) |
| 18 富安泰山 (白糖 ^{マイ} ガ) | 45 瀬戸山 (") | 72 寺田 (安 ^評 セン ^{ター}) |
| 19 秦 正弘 (鳥居薬品) | 46 池田正己 (日本 ^ハ ッ ^カ) | 73 石村 (サ ^ン ド薬品) |
| 20 五島滋喜 (大 ^丸 薬品 ^工 業) | 47 横井義之 (鐘 ^紡) | 74 小林高之 (鳥居薬品) |
| 21 近藤 満 (天野製薬) | 48 後藤和則 (日本 ^曹 達) | 75 奥村紘二 (ヒ ^ュ マ ^ラ イ ^フ) |
| 22 松浦 (薬 ^術 科 ^研 院) | 49 斉藤太郎 (化 ^薬) | 76 " (") |
| 23 神谷 (") | 50 兼澤 敦 (湧 ^永 製薬) | 77 中島敏秀 (化 ^研 生 ^薬) |
| 24 福田武司 (日本 ^生 物 ^化 学 ^セ) | 51 河村 寿 (ア ^フ ス ^ヒ マ ^マ 研) | 78 松下 (三 ^薬) |
| 25 瀬川美秀 (セ ^リ ア ^新 薬 ^工 業) | 52 戸塚和男 (東 ^薬 品 ^工 業) | 79 武井 (セ ^リ ア ^新 薬) |
| 26 今井 郎夫 (動物 ^繁 殖 ^研) | 53 大塚 (") | 80 石塚修司 (エ ^ス エ ^ス 製薬) |
| 27 森脇 (興和) | 54 小山 (") | |

〔事務局だより〕

第24回定例会では、「統計以前の動物実験の諸問題」というテーマに関連して、QAUの役割について講演して頂いた。当会の今までのテーマからすると異色のものであると思うが、お話を聞いていると、生データをどうしたら正確にとれるかということに直結しているので、むしろ当会の原点に当たるようなテーマとして、興味深かった。それに、質保証というのは、どんな仕事にも言えることで、我々のような多種類の事務をこなす場合でもQAU的な観点をしっかり持つことか、業務を誤りなく進行させる決め手であると思った。普通は、忙しさに追われてバタバタしてしまい、後であれを忘れた、とかここを落した、という騒ぎになって、急いだ筈が却、て大幅に遅れるなどという失敗がよく起きる。いくら急かしくとも、同時に二つのことはできず仕事は一つ一つしか進まない、と思いつらないと到底急かしさに打 克つことはできないもののようなのである。

次回 第25回定例会(1986年1月11日(土))は

総評会館 です。

- 基礎講座 Methods and Concepts of Biometrics Applied to Teratology
- 定例会 標準値(正常範囲)の統計的決定法 大橋靖男
(東京大学 医学部)
- 30分コーナー 「確率変数って何?」 吉村 功
(名古屋大 工学部)

医薬安全性研究会 会報 No.18

昭和60年10月31日発行

編集・発行 (株)サイエティスト社

〒101 東京都千代田区神田駿河台3-2

山崎ビル ☎03(253)8992

振替 東京8-71335

印刷・製本 ナガノ印刷

1985©