

統計十話

第8話 欠測値の問題

林 知己夫

文部省統計数理研究所

このごろ、統計学においては、データをとり方の問題が重くみられないように見える。データの分析のほうに力がかかっているのである。統計学はデータをどうとり、どう分析するかにかかっているのです。とられたデータの性格と分析とが緊密な関係になくはならないのにもかかわらず、この間の関連性が途切れているのである。実験でいえば、実験の計画性の問題であるが、これが専門化されて実験計画法としてこまかい問題が精密に研究されているが、現実の実験の計画法というソフトな部分が閑却されているのである。社会調査となると今日でもっとも現実的で大事な問題である「現状に適した標本計画法」がまったく研究されていないし、社会調査法についての技法の研究も皆無である。これらは40年間も研究が凍結されたままであるといえるのである。

これらは、一人でよくできるものではなく、チームが多年月倦まず撓まず研究を続けていかなければできるものではない。しかし、これができる研究環境や意欲のある研究者がなくなってしまったのである。

今日では、業績評価の時代——論文を書くこと——なので、じっくりした研究は損であり、手軽で一人でできる研究がいたるところで行われているのである。これに都合がよかったのがマイコンの普及である。相乗効果で事態はますます悪くなっていき、紙屑のよう

な論文の山を作り自然環境を破壊しているのである。

この要請にマッチしたのがモデル化である。医学のような領域のモデル化はごく局限された容易な——人間そのものが捨象された——場合でのみ成立するであろうが、これはトリヴィアルなものである。モデルがわかれば、おしまいで、そのパラメータの推定など大きな問題ではない。モデルがつくれたら大半は終りである。こうした事情であるが、モデル化は理論化にみえ、精密にみえるがゆえに珍重されているが、複雑な現象に対しては玩具の鉄砲のようなもので、実戦に役立つものではない。現象理解のための単なる教育手段であっていわばティーチング・マシンである。ここにおいてもパラメータの推定にはデータが用いられるが、そのデータがどんな性格のものか検討することをしない。

データに基づいて議論することは、今日では、「データの科学」(data science)の領域の問題となっている。これは、統計学・データ解析、行動計量といった領域の諸方法を止揚し、総合概念としての「データの科学」——データによってももの本質を理解・解明しようとするを第一義とする——のもとに、諸方法はいかにあるべきかを考え、諸方法を発展させようとするものである。いわば原点にかえって考えるということである。

データをとるという場合必ず欠測値が生ず

る。社会調査であれば、調査不能である。これを検討することなく無視して、手にあるデータのみを解析したら大きな誤りを犯してしまうことになる。この問題について考えてみたい。これは、統計的検定というような狭い問題ではなく、調査対象集団(ユニヴァース)とは何か、母集団(ポピュレーション)とは何か、標本(サンプル)とは何かにかかわる問題である(これについては第1話参照)。

ここでは実験・調査における欠測値について述べよう。「測定できない」ということは質的に異なったものを含意している可能性がある。計画がいくら緻密でも、欠測値の多く出るようなものは意味がない。欠測値のよく出るのは動物実験(シビアなことを試みる)であり、疫学(実験疫学)調査であり、臨床(長期においては必ず重要な問題となる)試験などの場合である。

長期間フォローアップすれば脱落(ドロップ・アウト)例が多く起り、欠測値となるのである。これをどう少なくするか、その性格をどう把握するかが大事な問題である。

大昔のことであるが、私はこれを体験したことがある。結核治療のための薬の吞ませ方の比較研究である。大規模な研究で、多くの結核療養所を対象にし、病態やその他の属性(性・年齢など)の等しい、いくつかの等質なグループを作成する。これに各種の「薬の吞ませ方」を定め、グループにランダムに割り付ける。ここまでは誰しも行っているところである。さて外的基準としては一応「喀咳の陰性化」が考えられた。

ここで調査がはじまるわけであるが、これを長く続けなければ意味がない。そのうちいろいろの理由で脱落がはじまる。これは、医療の現場である限り止むを得ないものである。さて、何か月たったところで、結果はどうかということでデータの分析がはじまる。そうしたところ、どうも、これまで考えられてい

たことと異なった結果が出てきたというので、私に相談に来られた。よく話を聞いてみると計画は実にきちんとしていたのである。問題は脱落例にあるように思えてきた。そこで、脱落例の推定、つまり欠測値の推定を考えたのである。時点 t までのデータはあったが、時点 $t+1$ のとき脱落したとしよう。つまり、 t という時点までの諸要因から $t+1$ という時点における外的基準の推定である。要因のなかには、性・年齢という属性、病態の要因、 t までの外的基準(+、±、-)のパターンのようなものが考えられる。これらは、質的要因であり、カテゴリカルデータとして表現されるのである。そこで数量化II類(統計十話 その2参照)を用い、 t までの諸要因から脱落例の $t+1$ の外的基準を推定することになる。

このとき、 $t-1$ までのデータを用い、 t のデータを推定することを考える。このなかには当然脱落例は含まれている。これによって脱落例の t 時点の推定がどの程度的中しているかを確かめてみる。あるいはこのとき、脱落例が t のデータをもっている、除外し、残りのもので $t-1$ までのデータを用い、 t のデータの推定を行う。このデータを用い、脱落例の t の外的基準の推定を行い、どの程度的中しているかをみるのである。なおこのとき、+±-そのものずばりを推定するということはまず難しいであろう。そこで諸要因のパターンから+になる確率 $P_+(t)$ 、±になる確率 $P_{\pm}(t)$ 、-になる確率 $P_-(t)$ を推定するのである。脱落例が n であるとき、それが示す+±-の人数は $n P_+(t)$ 、 $n P_{\pm}(t)$ 、 $n P_-(t)$ で推定されることになる。これが実際のデータとどう一致しているかで方法の妥当性を検討しておくことになる。

いずれにせよ、上述の二つの方法により推定の安定性を検討しておくことが大事である。こうしていよいよ本番に入り、 t まで諸要因から脱落例の t の外的基準 $P_+(t+1)$ 、

$P_{\pm}(t+1)$, $P_{-}(t+1)$ を推定するのである。こうすることによって欠測値の推定が一応可能になる。

このような考え方によって欠測値を推定し、データをまとめてみると医学のこれまでの予想と一致する結果が、はっきり出てきたのである。これに携わった人々は、欠測値のこわさを十分認識されたのである。その理由となると追及調査ができるわけのものではないので、本当のことは確かめようはないが、脱落理由にはいろいろあるが、つぎのようなことが予想された。外的基準がよくなったので自ら退院して行く——つまり残っているものは悪性のもののみが残り、なかなか薬が利かないという結果を招来する——、あるいは外的基準がよくなるので嫌になり他の病院に移る——つまり残っているものは良性のものが多く残り、薬の利きがよく出るという結果を招来する——ということが考えられ、さらにパーソナリティに基づく行動が絡むということになろう。この場合、この二つのことのために、よいと考えられる「薬の吞ませ方」のほうが表面的に悪い結果になり、それほどでないという吞ませ方のほうがよい結果を示したということになったのであろうと想像されるが、これは確かめようのない単なる想像でとるに足らぬことであるが、今後実験計画をするとき頭の片隅に入れてよしさやかな知見であろうかと思う。大事なことは、欠測値の推定ということにつきる。

また、アメリカの長期臨床試験に関する研究をみたことがある。糖尿病治療のための8年にわたる大サンプルの数グループの調査である。私の常識では、実験計画からして杜撰で、治療という観点からみたときまことに現実離れしたものである。しかしこれは不問にしよう。このように長期間では脱落例が多く出るように思われるが、その措置は全然とら

れておらず、質問したところ、脱落例に差はないとべなく答えただけで、データはまったく示されていなかった。これだけの大調査で、こうした考えの人たち——高名な人が代表者で、有名な研究者も多く含まれている！——がやっていたのでは信用できないと思われた。また、実験計画の薬の使い方も私から見ると現実離れしたものであった。アメリカのこうした大規模調査にも、こうしたものがあることに注意されたいのである。実験の計画と調査の実施管理と脱落例の処理である。データのこまかい検定論など枝葉末節のことで、データの妥当性を裏付けるものではない。

その調査結果が滑稽なものであった。多くのグループは発癌などで中止になり、残ったのはインシュリン定量投与グループと何もしないグループであった。しかも結論は、インシュリン投与に延命効果はないということである。

私は、この結果がどうのこうのということではできないが、調査実施の管理の厳格な条件監視と脱落例の処理なく出した結論は虚構そのもので、科学的研究に値するものではないと思う。調査実施の管理とは、薬を吞むということであれば本当にその指示どおり吞んでいるかということであり、注射なら身体はどこかが具合悪い場合でも注射しているかどうか、あるいは他の方法がとられているか、生活規律はすべてのグループで正しく守られているかなど調べる項目は実に多いのである。これらを抜きにして大規模の長期実験・調査は考えられないものである。こうした疫学研究は、正しく使えば役に立つが不用意に行くと害毒を及ぼすのみである。

社会調査では調査不能ということになるが、これは別様の扱いが必要になるが、この処理も欠測値同様大事であるということにとどめておこう。