

Analyses des données とはどんなことか

統 数 研 林 知 己 夫

いまさら数量化の話でもあるまいというところであるが、数量化の方法がいろいろのものに応用されだし他の方法と組み合わせられて使われだしてくると、意外に深いことが露程されてくるのが解ってきた。数量化の方法をただデータの解析に使ってみて応用例を示すという時代は過ぎ去り、本当の現象解析のまにまに使われ出してきた訳である。こうした動きが、ヨーロッパ、特にフランス・イタリアにも盛んになってきている。1973年以來のことである。この方法を *Analyses des données* (データ解析) というが、英語の *data analyses* とは全く様相を異にする。日本流に意識すれば、数量化と言うのが一番当っている。

昨年 *Analyse des données et Informatique* という国際会議が(パリ近くのIRIA(L Institut de Recherche d'Informatique et d'Automatique)がE Diday, L. Lebart, J. P. Page, R. Tomassone を組織委員会として)開催された。また、この一派の人達が今年の7月イタリアのナポリで、*Data Analysis and Informatics, Meeting the French School* という会合を開き、*analyse des données* の研究発表をやり、また講習会を開いた。この方面の主な人物を参考に書いてみると、仏ではE. Diday, Y. Escoufier, J. P. Felon, M. Jambu, L. Lebart, Y. Lechevallier, A. Morineau, G. Saporta, R. Tomassone, J. P. Bouroche という人達で、30才の後半から40才の前半位で大体同じ年齢層である。伊では、V. Amato, A. Gili, N. Lauro, E. Marubini, A. Rizzi などが出ているが、このほか、数量化の主の様なA. Herzel (1974年以降の発表、*quantificazione* という言葉を盛んに使っている)がいる。

さてこの一派の総師はJ. P. Benzecri (ベンゼクリと読む。バンゼクリではない) パリ第6大学の教授である。日本の人々にはBenzecriの名は珍らしかろうが、フランスでの数量化というよりパターン分類(Ⅲ類)の元祖である。correspondanceの名の下に1973年頃から精力的に仕事をし、本を書くやら雑誌を出すやら大変な活躍ぶりである。これまで私もいくつか本や雑誌をみていたが幻の人物であった。つまり国際会議など名前は出ているが本人は一向出席しない。一時は「寄せ名」ではないかと疑ったが、一昨年実在の人物とわかり、会いたいと思っていた。

パリ大学統計学部には二人の教授がおり一人はDugue (先年統数研に二ヶ月ばかりいた)で、一人はBenzecriである。一方は全くの数理統計学、他方はコレスポンドランス(データ解析志向という訳である)。

Dugue 教授は紳士で、Benzecri 教授は見るからに変わり者で、全貌正に教祖そのものであった。弟子は大学内外で200人を下らず頂点の張りの中にもこもり、弟子がいろいろ活躍するという形である。弟子は各地におり、アフリカ諸国にも信奉者がいるという。彼は統計学者か、数学者かといろいろの人に尋ねてみると、Benzecri est Benzecri という訳で笑話に終る。一人のやっていることは役に立たぬし、一人の方は狭すぎてどうにもならぬと不幸をかこつ人が多い。

ベンゼクリ教授の統計学に対する考え方が面白い。彼が *analyse des correspondances* を考え

た動機は、数量化の発生と甚だ似たところがある。従来の数理統計学への反旗であった。彼の研究の基盤としての「プリンシプル」をあげているが、その冒頭に、統計は確率ではない、数理統計学は気取った理論をいろいろやってきたが、そんな役に立たぬものを忘れることがよりよいことだ、という強烈な宣言がある。次に仮説検定の不毛性を述べ、大事なことは、合理的な仮説をたてるのに役立つような分析でなくてはならぬと述べている。そして、データの中から *renaissance des formes* することだという。外的基準のない場合、データの山から霧をつきぬけて妥当な情報を浮かび上らせることだという私の主張に近い。私としてはその類似性に驚いているが、根本の行き方・思想になにかの差があり、データ解析の発展方向・拡がりに違いが出ているのである。

少し蛇足を加えてみたい。数量化に関して R. A. Fisher が最初に考えたというようなことが書いてあるものもあるが、これは形態の一部が似ているに過ぎないので、思想は違う。Fisher のは実験計画法を中心とする統計学、欠測値の推定、有意性検定というものでありそれ以後の発展が全くないので見ても狙いの差が解る。Guttman のはスケール理論、Hayashi のは、最適尺度化・統計数理に基づくデータ解析の考え方、Benzecri のは現象理解のための *correspondance* (対応, 交感) という考え方によるもので、諸方法の発展の経過やデータによる現象解析のあり方に特色が出ているのである。ベンゼクリ氏の狙いは、森羅万象、*correspondance* による理解と言おうか。

話をもとへ戻そう。統計学者として有名なアメリカの Tukey 教授は、*Analyse des donnees* を評して、*exploratory data analysis* と呼ぶその本質をよくつかんでいるのはさすがである。私はこの後に *based on the idea of scaling* とか、*of qualitative variable* とかを加えたい気がする。

Tukey は、また、*Analysis of variance* に通じると述べている。しかし、それはあくまでも出来上った数理的形式のことであって、それを導いた考え方に言及していなかったのは惜しい。結果を形式的に見れば同じでも、数理統計という分散分析からは、*exploratory* というべきこうした方法は生れるよしもなかった。別のアイデアが必要であったのだ。

フランスのデータ解析は私の言う「統計数理」の方法、狭く言えば「数量化の方法」と同じ狙いを持っているので、私は大変喜ばしく感じている。口はばつたい言い方であるが、私や私どもの一部が主張してきたことが「やっとなんと解ってきたな」という感じで思わずニヤニヤさせられる。これは、Tukey の話にも通ずるものがある。彼の統計学批判は多くの人に読まれているが、どこまで真意が理解されているだろうか。最近も、*exploratory* をめぐって伝統的数理統計学者を批判しているものを見たが、やはり同感を禁じ得なかった。

彼は、測定値が数量であるし、絶対的なものと決めてかかるが、既存の統計量にこだわらず素直にデータを眺め、意味ありげな(妥当性のありそうな)クラスター化を考えるというパタン認識的な話もうまい所を突いている。しかし、以前から感じていたことであるが、彼の狙い(いまやろうとしていること)のある部分は数量化の考えで処理できるし、すでにそれで解決しているものもある。

いずれにせよ、統計学はパタン認識を通して脱皮しなければ、どうにもならぬ所にきていると思う。Tukey は物理学方面のデータしか取扱っていないので、測定値の任意性「数量を目的に応じ与える」

という考え方がなく、scalingの考え方は理解していないようである。全く惜しいことだと思う。あれだけの人物がどうしてそこが解らぬか、学問伝統の固定観念の強さには恐怖さえ感じる。「フランスのデータ解析は、そこにあるデータの解析で、データによる現象解析でないのが弱味だ、これには統計的アイデアが不可欠なのだ」と、Didayを始めフランス派の人に話して、共感を得た。何とかしたい、して欲しいとの意向を洩らしていた。フランスでは統計関係の協力が無いのである。この理由はフランスの統計学者がデータ解析に興味を示さぬところにある。狭い Benzecri にのみ頼らざるを得ない点が弱い。Benzecri 一派がいつ目をさますか、両者の対話がいつ成立するか、第三の流派がいつ出来るかにかかっている。

この様に弱さはあるものの、あれだけ多くの人々が熱を入れて研究を続けていること、一つの教室で Benzecri 一派だけで年間四冊の雑誌を出し、本を書き、多くの門弟を養い、精力的な活動を行っていることは、何としても驚くべきことである。たいしたことはない和高を括っていると、いつの間にか目ざめて新鮮な方法を展開してくることは予想に難くない。あれだけの熱気とエネルギーがあれば、大きな超克が期待されよう。それを傍観者的に見て、その成果を学ぶなどというのは、我々の先輩のしでかした誤謬と同じだと思わねばならない。

我々はいかにすべきか、自信をもつことが大事で、林業統計研究会のメンバーが真に重要である問題、すなわち視野を広く持ち日本の林業にとって何が最も大事であるかを考え自らのポテンシャルを活用し、たとえ問題が難しくともそれに迫ることを試みるべきである。この原型を諸外国に求めてもある筈はない。データによる現象解析を志向し、その体験や本来の意味の共同研究を通して上記の問題解決に対して基礎から考え直していかねばならない。そして得たより高い次元の立場を定め、そこから見える新しい地平線に突進することである。そうしたことが世界の学界の先駆的寄与になるという自信を持つべきである。右顧左弁することなく、自分の目を信じてよい。これには飽くことない情熱が必要である。