

《新書紹介》

Donald R. McNeil: *Interactive Data Analysis, A Practical Primer*

New York, John Wiley & Sons, 1977 ix+186pp.

鈴木 啓 祐

この書物の表題となっているインタラクティブ・データ・アナリシスという語は、データ、すなわち、統計的資料の分析をコンピューターとの対話の下に前進させ、データのもつ性質をより明確に取り出すための、いわば、「対話的データ分析手法」を意味する。

コンピューターによる通常のデータ解析では、一般にデータ解析者自身が、データの性質を、統計的、あるいは、理論的に検討して、データ解析の手法をあらかじめ用意し、その後、その手法に沿ってデータの分析がおこなわれる。そして、もし、その分析結果が不十分であったならば、あらためて、解析者自身が解析手法を変更し、再び新たな解析手法を作り、その新しい手法によって、再度データを分析することになる（データ解析の内容が複雑な場合には、このような分析方法は、今後も、存続するであろう）。

これに対して、インタラクティブ・データ・アナリシスでは、データが与えられると、まず、そのデータを分析する方向が解析者によって決定される。この分析方向だけは、解析者が決定しなければならない。たとえば、与えられたデータに対して度数分布に関する分析をおこなうか、時系列解析に関する分析をおこなうかという点について決定しなければならない。次いで、データが最も簡単な形(分析方法)でコンピューターによって分析される。その分析方法は、コンピューターのプログラムの中に、すでに、内蔵されているのである。その分析によって得られた結果は解析者によって判定され、それが満足な分析結果であると判断されたならば、分析作業は、この段階で完了する。これに反して、もし、その結果が不満であれば、データ解析者は、コンピューターに不満の意思を表明するかわりに、第1段階の分析とはやや異なった形の分析方法、あるいは、第1段階の分析に、それとは別の分析手法をつけ加えた方法でデータの分析をするように、コンピューターに指示する。そして、その指示によ

って得られる結果を検討する。その後は、上記と同様の手順によって、順次、コンピューターのプログラムの可能なかぎり（その限度はコンピューター・プログラムに内蔵された内容によって決定される）、種々の手法によるデータ解析を展開して行くのである。

データの解析者は、このようなコンピューターとの対話的分析を進める——すなわち、コンピューターに、「yes (それは良好である)」、あるいは、「no (それは、まだ、不満であり、さらに分析を進めよ)」という指示を与える——ことによって、即時的といえるほど、かなり短時間に、データの性質を明確にすることが可能となり——たとえば、データは、そのままの形で用いるべきか、あるいは、対数の形にして用いるべきか、あるいは、モデルとしては、どのような形のモデルが適当であるか等の点を明らかにすることが可能である——、かなり満足し得るデータ解析が可能となるのである。

いうまでもなく、この分析で得られた結果そのものを最終的なデータ解析結果として用いることも許されれば、その結果を基礎として、プログラムに内蔵されている分析方法以外の分析方法を解析者があらためて作り、その分析方法によって得られた結果を最終的解析結果として採用することも許される。しかし、いずれにせよ、このコンピューターとの対話的分析により、データの解析はデータの性質の十分な検討の下に実行され得るようになるのである。

本書では、このような方法による各種のデータ分析技術が解説され、その分析のためのプログラムも具体的に提示されている。

著者（の序文）によれば、広義の統計学は、(1)探索的データ解析 (exploratory data analysis), (2)統計的推論 (statistical inference), ならびに(3)統計的モデル・ビルディング (stochastic model building) の3つの部分からなり、この書物で取扱われる部分は、

これらのうちの第1の部分に相当し、この部分が統計的解析において、基本的に重要な部分であると彼は主張している。

本書の内容は、7つの章に分けられ、各章は、下記のような内容をもっている。

第1章『Display (表示)』においては、データの量的位置、あるいは、量的ひろがり、ならびに、度数分布の解析とその表示法が示されている。ここでは、各種の興味あるデータ表示法が挙げられているが、そのうちでも、特に、「オーストラリアは島であるか、大陸であるか」という問題を、インターラクティブ・データ・アナリシスによって解決して行く事例が印象的であった。

この問題を解く場合、はじめ世界中の海にかこまれた各種の陸地の面積“island”の度数分布が求められる。この段階では、よい結論が得られないため、第2に、陸地の面積の平方根“(island)^{1/2}”の度数分布が求められる。これでも、まだよい結果が得られず、さらに、コンピューターに、陸地の面積の対数“log(island)”の度数分布を作成するように指示する。このとき、変量“log(island)”の度数分布には、2つの度数の群が出現し、——小さな“log(island)”の領域に1つの度数の山が、また大きな“log(island)”の領域に別の度数の山が現われた——、オーストラリアの“log(island)”は、大きな“log(island)”をもつ度数の群の中に含まれることが知られた。この結果、『オーストラリア』は大陸である」という結論を得るのである。

第2章『Comparison (比較)』では、第1章の手法——そこでは、データの量的位置、あるいは、量的ひろがりの表示(メディアンと第1第3四分位数間の距離による表示)が可能である——を基礎として、2種以上のデータの比較に関する分析手法が挙げられている。

第3章『Relations (関係)』では、2種のデータを種々の値に変換して、それらの間の関係がどのような形で示されるかを解析して行く手法が述べられている。

第4章『Assays (検定法)』では、実験計画の手法をはじめとして、potency(効力)——薬品の効力——の測定方法が説明されている。

第5章『Tables (表)』においては、分割表の分析方法が挙げられている。ここでは、John Tukeyによって示唆され、この書物の著者、マックニール(McNeil)が“Median Polish”と名づけた方法による分析手法が挙げられている。

第6章『Smoothing (平滑化)』では、まずデータとして与えられた統計系列の平滑化の方法がとり上げられる。ここでは、“running median(これは、「移動メディアン法」と訳されるべきものである)”や“3RSR”という方法が紹介される。これらの方法においては、“end point rule”と名づけられた計算手法も導入されている。この方法による平滑化は、従来開発されている「移動平均法」とは、やや異なっていて、興味深い。

次いで、多重回帰分析がとり上げられる。この方法においては、通常の最小2乗法とは異った方法——いわば、「段階的多重回帰分析法」とでも名づけられる方法(これは、変数増減法のようなステップワイズに結果を求めて行く回帰分析の方法とも異なっている)——が用いられる。

最後の第7章『Fitting (あてはめ)』においては、これまでの各章で取扱われた各種の手法の基本的部分の構造や構成原理が詳述されている。

この種の内容をもつ書物としては、マックニールも指摘するように、John Tukey: *Exploratory Data Analysis*, Reading, Massachusetts, Addison-Wesley, 1977, xvi+688pp. が出版されているが、この書物はこの種の分析手法を詳細に解説した、きわめて大きな書物である。これに反して、ここに挙げたマックニールの *Interactive Data Analysis* は、本文が186頁という比較的小さな書物であり、量的な点から見ても、手軽に読める書物である。しかも、本書の説明方法は、かなり工夫され、原著者が述べているように、簡明な説明を目的として、事例を豊富に用い、その事例に基づいて説明をおこなうという方法が採用され、このことによつて、著者のいう“Interactive Data Analysis”あるいは、「対話的データ分析」というものがどのようなものであるかということ、この書物からきわめて容易に理解することができる。