

最適線形判別関数

日科技連出版社 212頁 2010年刊 定価2,800円

新村先生と初めてお話をさせていただいたのは、1983年アムステルダムで開催されたMEDINFO83（世界医療情報学会）であった。そのときまでは遠くから先生のご発表を聞きながら、SASと数理計画を使いながらの研究にものすごく詳しい先生という印象であった。「この歳でつい最近が初めての国際会議でした。あなたのように大学院生のときから経験できるとはうらやましい。頑張ってください」と言われたことを今も覚えている。その後、先生が研究の原点とされている大阪府立成人病センターの先生方とも縁の深い大阪大学に奉職できたのも何かのご縁かもしれない。特に、地域がん登録の研究でずっとご指導いただいていた大島先生と新村先生が研究をされておられたとは、つくづく世間は狭いと思った次第である。

本書はソフトカバーでしかも数式がほとんどなく212頁とコンパクトであり、初学者向けの本のように見える。確かに最初の数ページの導入は大変読みやすく、統計ユーザーとして自分が以前から疑問に思っていたことについて、同じ目線で疑問が掲げられ、それについて実際に検証してきたことを紹介していくという話になっている。しかし、実にポイントをついたまとめ方、話の進め方であつという間に本論に入っていく。正直、数式をもっと入れて解説してほしい、もし解説本を出して下さったら（おそらく3倍ほどの厚さにはなると思うが）ぜひ欲しいと思った。

判別分析は意思決定分析なのになぜそういう観点から解説されていないのだろう、多重ロジスティックモデルでのフィッシャーの仮説についてその場しのぎ的な検討の仕方になっているのではないか、判別関数の係数について信頼区間はどうかになっているのだろう、変数選択の総当たり法などなど、統計ユーザーが「判別」という課題に直面したときに疑問に思うことについて、一つ一つ実証をもって解説、方法の紹介をしてくれている。その意味で、書かれている内容は統計ユーザーとして判別分析を利用している人間にとって「そう考えていいのか!」と驚くことばかりで衝撃的

であった。

本の構成については、判別の方法論と特にモデル選択法と評価法とが並列して述べてある点、大変わかりやすい。前半にて簡単に理論が紹介された後、アイリスデータ（Fisherの仮説を満たす）、CPDデータ（19個の説明変数に多重共線性のある）、銀行紙幣データ（線形分離可能なデータ、すなわち最小誤分類数(MNM)=0）、学生データ（一般位置にないデータ）などを挙げて実証研究が詳しく記載されている。これも大変わかりやすい。しかし、これだけ広範囲な判別分析の方法論を比較実証で進める場合、もう一冊の「解説本」が欲しいと思うのは私だけではないと思う。また、判別の有効性をどう評価するかという評価法、モデル選択法についても実例を挙げての話でありわかりやすい一方、数式での解説が欲しいとも思った。さらに、サポートベクターマシン、整数計画法、BigMの導入、樹木図分析についても誤分類確率との関係をさらっと述べてあるにとどまっている。ぜひ、新村先生の筆による解説本を期待したい。100重交差検証法についても、これならばぜひ検証してみたいと思った。付録Cとして一緒に検証実験をしませんかというお誘いの部分があるが、ぜひ、がん登録をはじめとする様々な医学データで検討してみたい。手を挙げさせていただくものである。

一読して、そうかとすべてがわかる本とは言い難いが、この本で紹介された世界を見たい!と読者に思わせるには十分であり、日頃のデータ処理において忸怩たる思いを持って既存のプログラムを見つめてきた統計ユーザーには大変勇気を与えてくれる本である。まだまだ面白いことができそうだと、励ましてくれる稀有な本といえる。

先生のJMP関連の著書など一連の啓蒙書における実証主義という誠にも有難い構成がここでも生かされている。いよいよ我が国でもデータベースの公開へ動きがでてきている。判別分析は大変有意義な結果を導き出す強い武器となるものと期待される。（大野ゆう子）