

## 回帰分析の実際

S. チャタジー, B. プライス著  
 佐和隆光・加納 悟訳  
 新曜社 250頁 2500円 1981年

多変量解析の諸手法の中で、最も広い場面で、また最も頻繁に応用されているのは(重)回帰分析であるといっ  
 てよいだろう。しかし、回帰分析を実際のデータに適用  
 して、何らかの有効な結論を打ち出そうとすれば、通常  
 の初等的な教科書には書かれていないような数多くの問  
 題が生じてくるのも事実である。

通常の重回帰分析においては、(1) モデルの線型性、  
 すなわち被説明変数の期待値が未知母数の線型結合で表  
 現できること、(2) 誤差の加法性、(3) 誤差の不偏性と  
 等分散性、(4) 誤差の独立性、(5) 誤差の正規性が仮定  
 される。しかし、現実のデータにおいては、これら基本  
 的諸仮定からのズレが見られるのが普通であろう。線型  
 モデルによる近似では期待するほどの精度が得られない  
 かも知れない。データの中にはいわゆる異常値(outlier)  
 が含まれていて、これを計算に含めるか含めないかで、  
 推定値が大きく変動するかも知れない。生のデータを用  
 いるよりは、何らかの変数変換をほどこしたほうがよい  
 かも知れない。誤差の分散は均一ではなく、通常最小  
 2乗解の“良さ”は保証されないかも知れない。誤差間  
 には系列相関があって、通常有意性検定の結果は信頼  
 できないかも知れない。説明変数間の多重共線性のため  
 に、係数の推定値の符号が“予想”とは異なるかも知れ  
 ない。また、説明変数の候補が沢山ある場合には、どれ  
 を取り込むべきか選択に窮するかも知れない。

本書の目的は、線型重回帰モデルの基本的諸仮定から  
 のズレをどのようにして検出し、どう修正を加えて  
 より有効なモデルを構築するかを説くこと、訳者たちの  
 言葉を借りれば、回帰分析を実際に行なううえでの“臨  
 床的診断能力”を読者に身につけさせることにある。そ  
 して著者たちは、理論的説明は最小限に押え、手ごろな  
 大きさで、また何らかの“問題”を含む実例への適用を  
 通じて、解析のプロセスやさまざまな手法を説明する。  
 通常最小2乗法とそれにもとづく検定、主成分、リッ  
 ジ(ridge)回帰、変数選択について、ベクトル・行列を  
 用いた簡潔な理論的説明が章末に付録としてまとめられ

ているものの、本文中には、煩雑な数式の展開は見られ  
 ず、きわめて読みやすい体裁をとっている。

章ごとに、標題と内容をまとめておこう。

- 1章 線型単回帰——回帰分析の目的と線型単回帰の  
理論の復習
- 2章 モデルのずれの検索と修正：単回帰——異常値  
と変数変換の問題
- 3章 重回帰モデル——重回帰分析の標準的理論
- 4章 定性的な説明変数——ダミー変数の取り扱い方
- 5章 加重最小2乗法——誤差分散の不均一性への対  
処法
- 6章 相関のある誤差の問題——系列相関への対処法
- 7章 共線性のあるデータの分析——多重共線性の影  
響と主成分分析による検出法
- 8章 回帰係数の片寄りのある推定——主成分回帰と  
リッジ回帰
- 9章 回帰式における変数選択——変数選択のための  
各種の規準とリッジ回帰にもとづく  
選択

データをグラフに表現する、特に各種の残差プロット  
 を眺めてモデルのズレを見出すという残差分析を重視し  
 ているのが本書の大きな特徴である。モデルを構築する  
 という過程においては、従来の仮説検定にもとづく形式  
 的方法よりも、このような形式にとらわれない(informal)  
 探索的方法のほうが有効であるというのが著者たち  
 の立場である。そして、この探索的データ解析(exploratory  
 data analysis)の立場から書かれた回帰分析の  
 テキストとしては、本書は、Daniel & Wood [1]と  
 並ぶきわめてすぐれたものと言えよう。従来のテキスト  
 (たとえば奥野他 [2], 佐和 [3])の副読本として、特  
 に実務家の方々に一読をおすすめしたい、というのが評  
 者の感想である。

最後に、細かい点ではあるが著者たちとは意見を異に  
 する点、明らかな誤りや誤植についてまとめておく。

- (1) 著者たちは、残差の分散が異なることを無視して、  
各種のプロットを行なっているが、等分散となるよ  
うに基準化することも考えられる。
- (2) 異常値が推定値に与える影響についての記述があ  
いまいである。この“影響”を示す指標については、  
Cook & Weisberg [4], Atkinson [5]等にく  
つつかの提案がある。
- (3) 著者たちはリッジ回帰を重要視しているがこれに  
は異論が多いただろう。評者は特別な場合を除けば  
リッジ回帰の機械的応用には反対である。詳細は  
Draper & Van Nostrand [6]を参照のこと。

- (4)  $C_p$  を用いた変数選択の方法は、提案者 Mallows [7] のやり方とは若干異なるように思われる。
- (5) 航空機事故の件数  $y$  をフライト数  $N$  に回帰させる例(2.10)において  $\sqrt{y}$  を被説明変数にするなら、 $\sqrt{N}$  を説明変数にすべきである。
- (6) その他、正誤表参照(\*は原著にある誤り)

参考文献

[1] Daniel, C. & Wood, F. S. (1980), *Fitting Equations to Data*, second edition, Wiley

[2] 奥野忠一他(1981), 「多変量解析法」, 改訂版, 日

科技連出版社

[3] 佐和隆光(1979), 「回帰分析」, 朝倉書店

[4] Cook, R. D. & Weisberg, S. (1980), *Technometrics*, Vol. 22, pp. 495-508

[5] Atkinson, A. C. (1981), *Biometrika*, Vol. 68, pp. 13-20

[6] Draper, N. R. & Van Nostrand, R. C. (1979), *Technometrics*, Vol. 21, pp. 451-466

[7] Mallows, C. L. (1973), *Technometrics*, Vol. 15, pp. 661-675. (大橋靖雄 東京大学)

ページ	行	誤	正
v	↑1	ドナルド・A・フィッシャー	ロナルド・A・フィッシャー
* 29	↑5	平均2乗誤差	?
* 74	↑3	重相関係数	重相関係数の2乗
* 101	↑5	$\pm \frac{ts/n}{b_1}$	$\pm \frac{ts}{b_1} - \left\{ 1/n + X_m^2 / \sum (x_i - \bar{x})^2 \right\}^{1/2}$ [正確には、 $E(Y)$ の信頼区間を“逆に”解けば良い。]
* 116	↓8	(1969)	(1971)
* 116	↑3	係数	説明変数の値
130	↑9	$r_i = 1$	$r_i = n_i$
155	↑6	4階,	4階, 6階,
175	↑8	売込み費	販売費用
192	↑5	$r_i$	$\lambda_i$
206	↑8	-0.4196	0.4196
* 237	↓7	591-612	621-625
238	↑2		このことは、そのとき消去された変数の回帰係数が、その推定値の標準偏差よりも小さいとき、縮小モデル…平均2乗誤差が小さくなることを意味している。
239	↓2		

書 評

地政学入門

河野 収著  
原書房 昭和58年

「地政学」という言葉は、(まだ小学生だった筆者は)戦前・戦中かすかに見聞したことがある。戦後は戦争のお先棒をかついだ学問として、日本においては忌避され、黙殺されてきた。しかし欧米、特に米国では国家戦

略の有力な理論として研究されている。こういうこともあって、数年前亜細亜大学の倉前盛通教授が、『悪の論理・ゲオポリティクとは何か』(日本工業新聞社 昭和52年)を出した時、子供の頃へのなつかしさもあってためらわずにこの本を買った。当時、あたかも米中復交、日中復交後の国際環境の変換期でもあり、また第一次石油危機の後であったので、中東をはじめとする今後の国際問題の見通しや情勢は如何と、多くの経済実務の担当者は、いろいろな書評でとりあげられたこともあり、この本を求めたといわれている。しかしまた、一部の識者は題名どおり、きわものみとて罪悪視しようとする人もあった。その後、国際環境は倉前教授の指摘どおり(新