

アプトン

調査分類データの解析法

池田 央
岡太彬 訓 訳

朝倉書店

アプトン

調査分類データの解析法

池 田 央
岡 太 彬 訓
訳

朝 倉 書 店

訳 者 序

社会調査やアンケート調査を経験した人は、だれしも与えられた質問項目（正しくはカテゴリー）に対して回答者の何人あるいは何パーセントくらいが‘はい’とか‘いいえ’とか答えてくれているだろうかということに関心をもつものである。そのような単純集計ばかりでなく、もう少し立入った関係に関心を寄せる人は、ある質問項目と別の質問項目との関連性に着目し、たとえば男女別や年齢別に質問項目の答え方がどう異なっているかとか、ある質問に‘はい’と答えたもののうち、何パーセントが別の項目で‘いいえ’と答えているか、とか、両者の関係を探ろうとするであろう。そういった質問間の関係の分析は通常クロス表といわれる2重分類表を作り、その関係をみることから始まる。しかし、一見簡単そうに見えるこのクロス表の分析も、関係する要素の数がふえてくるにしたがって複雑となり、またその関連性を表示する統計測度にも多種多様のものがあって、一筋縄にいかないものであることがわかってこよう。

ところで、この10年間くらいのあいだにこうしたクロス表の解析法についてはめざましい進歩があった。その一つが本書にまとめられている L. A. Goodman らを中心とする対数-線形モデルで、多変量にわたるクロス表を統一的に扱う一つのモデルを提供している。それは定量的変数を扱う分散分析の概念とも密接なつながりをもっているので、多くの人から関心をもたれている。

こうした新しい発展を紹介する書物が最近種々出版されるようになり（巻末文献補遺参照）、中でも本書は比較的読みやすい概説的入門書として手頃なものであろう。

翻訳は第1章から4章までと10章の一部を池田が担当し、5章から9章までと10章の残りを岡太が担当し、最後に二人で調整を行った。また邦語の定訳のない“Odds”，“Odds ratio”は“勝目”，“優比”と訳してみた。そのほ

か二，三の新しい訳語を当てた部分もある。

本書がクロス分類表の新しい解析法をわが国の読者に広めるのに役立てば，
この上ない幸せである。

1980 年 9 月

池 田 央
岡 太 彬 訓

序

本書は 1976 年エセックス大学(University of Essex)で行われた欧州政治研究協会(European Consortium for Political Research)の夏期講座での講義から誕生したものである。この講義のノートを用意している間に、クロス表データの解析法が 1970 年来めざましい早さで発展していることがわかり、またこの新しい方法についての入門書が一冊もないことがわかった。夏期講座に参加した人たちには私のノートが熱心に迎え入れられ、本書を出版しようという考えが生まれた。

本書の意図するところは、統計学者には分割表として、また社会学者にはクロス表として知られているクロス分類データの解析法の複雑さを紹介することである。本書が目標とする読者は、原則として、いままでデータを集めたことのある社会科学研究者で、それを今解析しなければならない人たちである。また過去に 2 元表の連関度を分析した経験をもつ多次元データを扱った人は、本書を学ぶことによって、このより精妙な分析をしてみたいと考えるようになると思われる。

本書には、たくさんの原論文の文献があげられている。また使用記号は実用的な範囲でできるだけ原論文の記号と対応するように用いた。基礎的な統計方法論に慣れていることだけが読者に要求されていることで、したがって本書を読めば研究者にもまたより進んだ統計コースを履習している学生にも、方法論のテキストとして等しく有用であることが明らかになるであろう。

1 章では、一般問題を概説し、統計的技法の必須部分を簡単に復習する。2 章から 4 章までは‘伝統的’解析法、とくに連関法についてより便利なレジメをし、同時にそれはつづく 5 章から 9 章までにわたってくわしく論ずる対称線形モデルの使用のいとぐちとなるものである。最後の章はパネルデータに関連した特殊問題を扱う。本書の主要テーマは、対数線形モデルの重要性が中心にあり、それに関連したクロス表データの解析技法を扱う。

式はできる限り用いないで、そこで記述されている技法の理解を助ける材料となる場合にのみ使用した。理論がむずかしくなってきたときには（読者にとってもまた著者にとっても）、それを省略し、原論文の文献をあげることと、簡単な要約を述べることで置きかえた。

本書の実現にあたっては多くの人に感謝しなければならない。Ivor Crewe 氏には、最初夏期講習を招聘して下さったことに対して、Jim Alt 氏には、計算機ルーチン ECTA を理解しこれを稼動可能にして下さった助力に対して、David Robertson 氏には、1975 年の国民投票のデータを揃えて下さったことに対して、また他の参加者には有益な多くの討議に対して、お礼を述べたい。とくに妻には、執筆中の生活上の犠牲に対して寛容であったことに対して、子どもたちにはあまり執筆の邪魔をしなかったことに対して、また Pat Reader さんには多量のタイプをしてくれたことに対して感謝したい。

本書に誤りが残されているとすればその責任は私のみにある。読者からの指摘があれば大変有難い。

1977 年 8 月

G. J. G. U.

現代人の統計

本にはいろいろの型がある。理論体系をしっかりと明確に書いた理論的な本、実用書的な本といった分類もその一つである。理論的に書いた本を読んでも、直ちに右から左へと役立てることは難しい。これは当然のことであると思う。こうした本は、基礎をきっちり身に付けるのに役立ち、ポテンシャルを蓄えるという機能を持つものであるが、問題意識の不明確な書き方のものでは、冷い空々しさのみが残る。実用書はその書く通りやって行けば何か出来るようになっていくが、これは読むだけではやはり解るものではなく、その通りやって行くうちに実用のコツを自ら覚えるという努力をしなければならぬ。統計学のようなもので、実用書的な書き方を希望する人も多いが、やはり本質を見失って、知らず知らずのうちに誤りを犯すことになる。しかし折衷というのは、二兎を追うことになってうまくない。理論的にしっかりと書くが、その視点を「役に立つ統計」というところに置くという行き方が望ましい。つまり役に立てるという考え方を軸にして、統計理論をしっかりと書き分けるという立場である。

これはなかなか難しいことであるが、いままでの統計学の書にあきたらなく思い、何とかしようという意欲のある新進気鋭の統計研究者が集ってこの叢書をまとめることになった。出来上がったものは、これまでにない新鮮な内容で随分ユニークなものとなったと思う。

しかし、本はあくまでも本であって、これをポテンシャルにして冷暖自知、工夫を凝らしてデータ解析に生かすのは読者諸氏にあると思う。その種子はこの叢書の中にあるので育てあげて欲しいものである。 [林 知己夫]

- | | | |
|----------------|------------------|----------|
| 1. 統計解析法の原理 | 鈴木義一郎著 | 価 2000 円 |
| 2. 多変量解析法 | 柳井 晴夫 著
高根 芳雄 | 価 2000 円 |
| 3. 社会調査法 | 杉山 明子著 | ☆ |
| 4. 意思決定の基礎 | 松原 望著 | 価 2000 円 |
| 5. 実験計画法の基礎 | 早川 毅著 | 価 2000 円 |
| 6. データ解析と管理技法 | 大隅 昇著 | 価 2500 円 |
| 7. 多元的データ分析の基礎 | 駒沢 勉著 | 価 2000 円 |

——☆印は続刊予定(全7巻)——

定価は昭和 55 年 9 月現在のものです。

人間科学の統計学

本シリーズは、米国において非常な好評を博し、多くの版を重ねている
“Series : Quantitative Applications in the Social Sciences” (SAGE
Publications)の日本版です。各種の数理的手法を、統計学や数学の専門
的な知識を必要としないで理解できるよう、簡潔にわかりやすく、また
きわめて实际的にまとめられております。多忙な現代人にとって待望の
シリーズの誕生です。

1. 多次元尺度法 クラスカル・ウィッシュ著 高根芳雄 訳 価 1400 円
2. 因果分析法 フッシャー著 広瀬弘忠 訳 価 1400 円
3. 生態学的推論 ラングバイン・リヒトマン著 長谷川政美 訳 価 1300 円

〔以下続刊〕

コーホート分析法 グレン著 藤田英典 訳

時系列解析—ダイナミックプロセスの分析 オストローム著 松原望 訳

多変量相関分析の方法 レビン著 柳井晴夫 訳

質的データの解析法Ⅰ—名義尺度データ レイノルズ著 中村隆 訳

質的データの解析法Ⅱ—順序尺度データ ヒルデブランド他著 大隅昇 訳

統計的検定 ヘンケル著 松原望・野上佳子 訳

多重指標分析 サリバン他著 広瀬弘忠・石塚智一 訳

探索的データ解析の方法 ハートウィグ他著 柳井晴夫・高木広文 訳

テストの信頼性と妥当性 カーマイン他著 水野欽司・野嶋栄一郎 訳

パネルデータの分析 マルクス著 水野欽司 訳

定価は昭和 55 年 9 月現在のものです。

東工大教授 日野 幹雄著
工学博士

統計ライブラリー **スペクトル解析**

A 5 判 312頁 価3800円 〒200

〔内容〕ランダム変動の表現・自己相関関数とスペクトル、相互相関とクロス・スペクトル、定常性、情報エントロピーとスペクトル、線型システムの理論、スペクトル計算の誤差理論、データ処理の手法、さらにすすんだスペクトルの概念。

京大助教授 佐和 隆光著
経済学博士

統計ライブラリー **回帰分析**

A 5 判 196頁 価2300円 〒200

統計手法のうち最も実用に供されることの多い回帰分析の方法を、豊富な実例と厳密な数学的証明をおりまぜて解説。〔内容〕回帰分析への誘い、ベクトルと行列、多変量正規分布、線形回帰モデル、仮説検定・区間推定・予測、説明変数の問題

名大教授 柏木 力著
医学博士

統計ライブラリー **医学統計解析**

A 5 判 248頁 価2800円 〒200

医学・生物学の領域においてその統計的方法はここ数年来の電子計算機の著しい普及によってますますその応用範囲を広げている。本書は臨床医学出身の著者が具体的な例題を中心に、きわめて明快平易に實際的にまとめあげた医学統計の入門書

後藤昌司・畠中駿逸・田崎武信訳
統計ライブラリー

コックス **二値データの解析**

—医学・生物学への応用—

A 5 判 240頁 価3000円 〒200

医学・薬学・生物学・農学系の研究者、技術者のための適用例が数多く収められ、さらに日本版にはプログラムも加えられ、実際に役立つようまとめられている。〔内容〕線形ロジスティックモデル、単一パラメータの解析、複雑な解析、他。

北大助教授 齋藤 堯幸著

統計ライブラリー **多次元尺度構成法**

A 5 判 250頁 価2700円 〒200

心理学、社会学、政治学をはじめとした社会科学から、医学、生物学を主とした自然科学に至るまで広い範囲にわたってその応用範囲を拡げている多次元尺度構成法について応用例を中心にしてきわめて實際的にかつ懇切丁寧にまとめられている

脇本和昌・後藤昌司・松原義弘著

多変量グラフ解析法

A 5 判 208頁 価2500円 〒200

多次元データのもつ意味を総合的に伝達また解析する手法として開発された顔形グラフ、体形グラフ、星座グラフ、木形グラフ、バイプロット法、非線形変換グラフ、樹形図等の多変量グラフ解析法を平易に解説。巻末にプログラムを付した。

大山 正・武藤真介・柳井晴夫著
現代の心理学

行動科学のための統計学

A 5 判 228頁 価2300円 〒200

心理学を中心に行動科学の領域において、その統計的方法はますます応用範囲を広げている。本書は具体的な例題を中心に、直観的・感覚的の把握ができ、現実の問題に的確な適用ができるよう細かい配慮がなされ極めて實際的にまとめられている

理学博士 寺田一彦 青山学院大教授・理博 鈴木栄一著

例解入門 **推測統計法**

A 5 判 264頁 価2500円 〒200

統計の考え方を初めての人にも分り易いよう、数式をなるべく用いずに解説した定評ある入門書。今回情報化時代に即応するよう改訂。〔内容〕母集団と標本、測定値の統計、正規分布、相関関係、標本調査、実験計画、情報と予測、情報と判別・他

定価は昭和 55 年 9 月現在のものです。

目 次

1. 序 論	1
1.1 はじめに	1
1.2 クロス表の作成	2
1.3 サンプル, 母集団, ランダム変動	3
1.4 正規分布	4
1.5 カイ 2 乗分布	5
1.6 仮説の検定	7
1.7 推定と期待値	8
1.8 カイ 2 乗による適合度検定	9
2. 2×2 表の連関と独立	13
2.1 2×2 の基本表	13
2.2 表の構造	14
2.3 A と B の独立性の検定	15
2.4 Fisher による厳密な検定	19
2.5 連関性の測度	21
2.6 Yule の Q	22
2.7 優 比	23
2.8 Guttman の対称, 非対称 λ 値	24
2.9 Goodman と Kruskal の τ 測度	25
2.10 X^2 に直接基づく測度	25
2.11 連関性測度の選定	26
3. $I \times J$ 表の連関と独立	27
3.1 独立性のためのカイ 2 乗検定	27
3.2 独立性の欠如個所の位置づけ	30

3.3	カイ 2 乗の分割	31
3.4	$I \times J$ 表の連関測度	33
3.5	λ_b を用いた連関性の測定	33
3.6	λ_b または λ を用いた関連性の測定	35
3.7	λ 測度の信頼区間	36
3.8	X^2 に基づく連関測度	37
3.9	Goodman と Kruskal の τ 測度	38
3.10	順位表についての連関測度	39
3.11	Goodman と Kruskal の γ	40
3.12	Kendall の τ	41
3.13	Somers の d	42
3.14	連関性の測度の比較	43
4.	多次元表の連関と独立性	44
4.1	はじめに	44
4.2	3 元表の記号	44
4.3	A, B, C の相互独立性	45
4.4	条件付独立性	46
4.5	C の A と B に対する多重独立性	47
4.6	Simpson のパラドックス	48
4.7	3 因子交互作用の解釈と定義	49
4.8	多元表の連関の分析	49
5.	2×2 表の対数線形モデル	51
5.1	はじめに	51
5.2	モデルを用いる目的	51
5.3	モデルの形	52
5.4	2×2 表の飽和モデル	54
5.5	2×2 飽和モデルのパラメタ推定	57

5.6	2×2 表の独立モデル	59
5.7	2×2 表のその他のモデル	61
5.8	階層モデル.....	62
5.9	モデルの再パラメタ化.....	63
6.	多元表の飽和モデル	65
6.1	多元表の分析.....	65
6.2	一般3元表.....	65
6.3	2×2×2 表.....	67
6.4	飽和モデルのパラメタの推定.....	68
6.5	パラメタの推定値の標準化.....	69
6.6	重要な λ の選択.....	70
6.7	より高次の多次元表.....	71
6.8	モデルの解釈.....	74
7.	多元表の不飽和モデル	78
7.1	λ と周辺和の関係.....	78
7.2	階層モデルの例.....	79
7.3	λ の有意性検定.....	80
7.4	5要因表の分析.....	83
7.5	最適モデル発見技法の検討.....	89
7.6	要因と反応の混合した場合.....	90
7.7	要因/反応が混合している例	91
7.8	多段階要因/反応モデル	94
7.9	多段階モデルの例.....	95
7.10	パス図による要約	95
8.	多元表分析のさらに進んだ技法	98
8.1	ふるい分け.....	98

8.2	残差の分析	100
8.3	対数線形モデルの乗算型	102
8.4	表の調整—モステラー化	104
8.5	その他の検定法と推定法	106
8.6	擬ベイズ推定値	109
8.7	混合モデル	111
9.	多分変数と不完備表	114
9.1	多分変数の場合の母数の解釈	114
9.2	指針としての飽和モデル	117
9.3	分離可能な不完備表	122
9.4	分離不能な不完備表	123
9.5	不完備表に対する準独立モデルおよび他のモデル	124
9.6	完備表の分割—セルの除去	126
9.7	完備表分析の別方法	129
9.8	部分的にカテゴリー化された表	131
10.	対称性・パネルデータ・変動モデル	132
10.1	はじめに	132
10.2	対称性	132
10.3	周辺同等性と準対称性	134
10.4	高次元表での対称性	137
10.5	パネル研究	137
10.6	移動者-停留者モデル	138
10.7	移動者-停留者モデルの失敗	140
10.8	特殊対数線形モデル	141
10.9	潜在構造モデル	145
10.10	メンバーの減少を伴うパネルデータの分析	148
10.11	移動者-停留者モデルの拡張と関連する考え	149

付 表	150
付表 1. 正規分布関数	150
付表 2. カイ 2 乗分布の上側確率点	151
参考文献	152
人名索引	159
事項索引	161

Ⅱ。序 論

1.1 はじめに

本書は、ある特殊なタイプのデータの分析に利用される方法について書かれたものである。そのデータとは、いろいろな特徴の組合せをもった人、場所、事柄などの個数を数え上げてできた**計数(counts)**である。そのようなデータは調査結果を要約したり質問表の結果を分析したりするときには必然的に生じてくるもので、計量的研究をめざす社会科学者にはごく見慣れたものである。

表 1.1 は仮想のものではあるが、英国民のスポーツへの興味について知ろうとして行った調査要約データを表す。

表 1.1 スポーツへの興味調査結果

回答のカテゴリー	人 数
45 歳をこえる男子で、テニスよりもクリケットの方が好きなもの	58
45 歳をこえる男子で、クリケットよりもテニスの方が好きなもの	24
45 歳およびそれ未満の男子で、テニスよりもクリケットの方が好きなもの	74
45 歳およびそれ未満の男子で、クリケットよりもテニスの方が好きなもの	38
45 歳をこえる女子で、テニスよりもクリケットの方が好きなもの	12
45 歳をこえる女子で、クリケットよりもテニスの方が好きなもの	86
45 歳およびそれ未満の女子で、テニスよりもクリケットの方が好きなもの	13
45 歳およびそれ未満の女子で、クリケットよりもテニスの方が好きなもの	95

ここで調査を受けた人は三つの基準、つまり性別、年齢、スポーツに対する好み、によって分類されている。これら三つの分類はすべて同時に行われているので、われわれはこれらのデータは**クロス分類**されたという。また、この調査を受けた各人は表にある八つのカテゴリーのどれかに属しているので、われわれは、これらのデータは**カテゴリカル データ**であるという。

表 1.1 の数をくわしくみるといろいろな特徴が解釈できるのに気がつく。そのうちもっとも明らかなのは、大多数の女性はテニスを好み(206 人中 181 人)、大多数の男性はクリケットを好む(194 人中 132 人)ということである。次にもっとも明らかな特徴は、おそらく、サンプルの中では、若い人が多いこと(45 歳以下が 55%)、また、45 歳をこえる人の中では女性が多いこと(180 人中 98 人で 54%)などであろう。そのほか、いろいろなちがいがまた見受けられる。本書の目的は、この種のデータに含まれる興味ある特徴をどうやってつかみ出すか、それに役立つ技法を解説することである。

1.2 クロス表の作成

クロス表は、SPSS (Statistical Package for the Social Sciences, 社会科学のための統計パッケージ; Nie ほか, 1975)のようなコンピュータパッケージの利用者には、耳慣れた言葉である。クロス表の形は表 1.1 のデータが、表 1.2 のように、より簡潔な形で表示することができることで示される。

表 1.2 表 1.1 のデータのクロス表

男				女			
好 み	46歳 以上	45歳 以下	計	好 み	46歳 以上	45歳 以下	計
クリケット	58	74	132	クリケット	12	13	25
テ ニ ス	24	38	62	テ ニ ス	86	95	181
計	82	112	194	計	98	108	206

表 1.2 の左右のそれぞれは、二つの下位表と考えられる。各下位表は年齢とスポーツの好みという二つの基準で作られている。そこで、われわれはこのような表を **2 元表**(two-way table)という。あるいはこの下位表は 2 次元に並べられた数の配列で容易に書かれるので、**2 次元配列**(two-dimensional array)とか **2 次元表**(two-dimensional table)ともよばれる。一方、完全な全体表は、容易に 2 次元に分解されて、表 1.2 のように示すことができる。表 1.2 は最初に性別で分けられたが、同じようにして、表 1.3 のように年齢で分けることもできる。

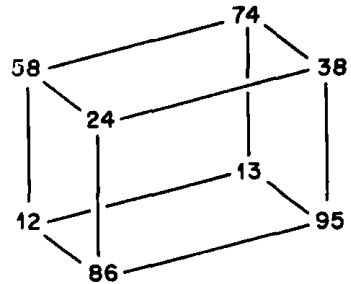
表 1.3 表 1.2 の別提示

46 歳以上				
好	み	男	女	計
ク	リ	58	12	70
テ	ニ	24	86	110
ス				
計		82	98	180

45 歳以下				
好	み	男	女	計
ク	リ	74	13	87
テ	ニ	38	95	133
ス				
計		112	108	220

表 1.2 も 1.3 もともに、表 1.4 に示される 3 次元表を二つの異なった方法で 2 次元表示したものである。実際、表 1.4 は表 1.1 の対称的な配列で、それはある一組の基準の間の相互連結をとりあげるため、他の相互連結がかくれてしまふことを避けている。その理由からこれは

表 1.4 表 1.2 の 3 次元配列



のちに、われわれは表 1.1 に示されるような 3 次元データをいかに分析するか学ぶことになる。また事実その技法は n 次元データ ($n > 3$) にも等しく適用できることがわかってこよう。

1.3 サンプル、母集団、ランダム変動

表 1.1 で仮想的な調査結果が示された。この結果が得られるまでの経過をたどると、調査チームに与えられた最初の指示にまで行きつくことになる。その人たちには、要約するとおそらく「外へでて、次のルールに従って 400 人の成人の被面接者を選び、インタビューしなさい…」という形の指示が与えられたにちがいない。もし、サンプリングのルールが慎重に考えられたものであるなら、このサンプルから得られた結果は、全体として母集団をよく反映していると期待されるだろう。ここでの母集団が何であるかは、サンプリングの指示に依存する。たとえば、もしすべてのインタビューが 1977 年 7 月 22 日金曜日の午前 10 時から 11 時までの間に、大ロンドン市で行われたものであるなら、そ