

多変量解析法（その3）

——公衆衛生学・疫学への適用例をめぐって——

柳 井 晴 夫 松 村 康 弘
木 村 信 子 高 木 廣 文

公 衆 衛 生

第47巻 第11号 別刷

1983年11月15日 発行

医学書院

多変量解析法（その3）

——公衆衛生学・疫学への適用例をめぐって——

柳井 晴夫* 松村 康弘**
木村 信子*** 高木 廣文****

はじめに

これまでの2回、多変量解析の手法を外的基準のない場合とある場合に分けて、それぞれの場合に含まれる手法の概要を数値例とともに説明してきたが、今回は、公衆衛生学・疫学分野における多変量解析の適用の現状と、その問題点について考察する。

計量診断と多変量解析

公衆衛生学に限定せず、広く医学一般の観点からみると、医学の分野で多変量解析が頻繁に用いられているのは、いわゆる問診や医療検査値にもとづく、疾病の自動診断、すなわち、計量診断における判別関数の利用であった。我が国で、判別関数を用いて最初に計量診断を行った研究は、1950年頃の鳥居ら¹⁾による胆石と胆道癌の判別であり、また判別すべき項目がカテゴリー項目に拡張された、数量化理論2類による計量診断の研究がある²⁾。さらに1950年から1960年代にかけて、変量が0—1データの場合、変量の独立性を仮定した尤度法による計量診断、尤度法に事前確率を考慮したベイズ確率による計量診断、さらには枝分かれ方式、多段判別方式による計量診断が行われ、その適用範囲は、小脳腫瘍、脳卒中、腎疾患、肝疾患、頭部外傷、虫垂炎等、内科的、外科的疾患に集中していた。そして、これらの研究の多くは高橋³⁾によって編集された「計量診断学」にまとめられている。しかし、1970年代に入ると計量診断の対象は内科学、外科学の疾患に限らず、産婦人科⁴⁾、精神神経科^{5,6)}、眼科⁷⁾等、あらゆる領域の疾患に拡張され、これ

らの内容は、古川ら⁸⁾によってまとめられている。

ところで、判別関数を用いて計量診断を行う場合、診断の対象となる病名が、分析以前に明確に定義されていなければならない。統計学的に言えば、病名は固定された概念ではなく、患者の呈するさまざまな病状を分類する為の一つの基準であると考えられる。さらには患者の臨床症状を表わす多くの臨床検査値の間には、かなり高い相関がみられるのが普通であるから、これを因子分析または、主成分分析にかけて、冗長な情報を排除しておくことも必要とされる。膠原病のように、複数臓器に異常をもたらす疾患の場合病名を診断する明確な基準が存在しないので、患者の呈する症状の組み合わせにクラスター分析の手法の適用⁹⁾により、患者をいくつかの疾患群に分類するといったように計量診断の研究にも外的基準のない分析法が有効となってきている。

リスクの推定と多変量解析

ところで、計量診断の研究の多くは、与えられた検査値を用いて、いかに高い精度の診断ができるかを意図していたものであったが、公衆衛生学の立場からすれば、診断の精度を上げるよりむしろ、疾病の発症に、どのような因子が関与しているかを明らかにすること、すなわち、その疾病の発症に関連するリスクファクター（危険因子）を明らかにすることが急務である。このような立場にたつ研究の嚆矢は、1948年から継続して実施されているフラミンガム研究（Framingham Study）であることはいままでもない。フラミンガム研究とは虚血性心疾患を含めた各種の循環器疾患に対する、各因子のリスクを評価するために行われた大規模な前向き調査で、この研究の過程で因子 $X_1, X_2, \dots, X_p$ から、ある疾病の生ずる確率を

$$P(X_1, X_2, \dots, X_p) = 1 / (1 + \exp(-f)) \quad (1)$$

$$\text{ただし } f = a_1 X_1 + a_2 X_2 + \dots + a_p X_p + b \quad (2)$$

で評価する多重ロジスティックモデルが生まれてい

* YANAI, Haruo : 千葉大学文学部行動科学科

** MATSUMURA, Yasuhiro : 東京大学医学部保健学科

*** KIMURA, Nobuko : 東京大学医学部保健学科

**** TAKAGI, Hirofumi : 聖路加看護大学看護学部
別刷請求先 : (〒113) 東京都文京区本郷7-3-1
東京大学医学部保健学科疫学教室 柳井晴夫

る。上記の式に含まれる重みつき合計点 f は、 $X_1, X_2, \dots, X_p$ が多次元正規分布に従う場合は、判別分析によって得られる判別関数に一致するが、多次元正規性の仮定が満足されない場合は、 $a_1, a_2, \dots, a_p, b$ の推定には最尤推定が必要とされる。なお (1) 式は $f \rightarrow -\infty$ のとき 0, $f \rightarrow +\infty$ のとき 1 になるもので、一変量の量反応関係を多変量に拡張したものになっている。

ここで、(2) 式における係数、 $a_1, a_2, \dots, a_p$ の意味づけについて考えてみよう。

a_j ($j=1, 2, \dots, p$) は、重回帰分析における偏回帰係数と同様に、 $X_1, X_2, \dots, X_p$ の内部相関に左右されるもので、 X_j が 2 群で差がなくても判別係数が正になったり、差がある場合に判別係数がゼロになることもある。また、判別係数 a_j は、変数 X_j が 1 単位増加した時の判別値の増加率を示すものであるが、日常の医療場面において、収縮期血圧を下げるような治療をすれば、拡張期血圧も下がる。さらにパン食を米食に切り替えるように栄養指導すれば、牛乳を飲む量も減少してくるといったように、単独の変数の変化率 (判別係数) の大きさだけに注目するのは正しくない。統計学的には、判別係数や偏回帰係数は、判別に ($X_1, X_2, \dots, X_p$) とした P 個の変数を用いるという条件下で、それぞれの変数が判別

の対象となる疾患にどの程度関連しているかを示すものである。上記のような観点から、判別係数の持つ意味を解釈していくことは、当該疾患の etiology を解明していく上で、きわめて有用である。

アメリカのフラミンガム調査に刺激されて、我が国でも 1960 年代から 1970 年代にかけて、さまざまな疾病のリスクを評価するための前向き調査が実施されるようになってきた。このうち、最も大規模なものは、平山¹⁰⁾による 40 歳以上の成人約 26 万人を対象としたもので、癌の各種部位に対する喫煙、食事摂取量などの生活習慣の各因子のリスクが求められている。しかし、この研究のデータの解析に多変量解析法は適用されていない。平山のような大規模なものではないが、Hirota¹¹⁾ による 6 年間の追跡調査に基づく脳卒中の前向き調査、鏡森¹²⁾ による 3 年間にわたる健康診断者の追跡による心疾患、脳卒中、循環器死亡の発生因子の分析、大和田¹³⁾ による農村婦人の貧血発生に関する諸要因の分析には、判別分析、多重ロジスティック分析、重回帰分析などが使われ、それぞれ疾病に対するリスクが定量的に評価されている。したがって、多変量解析の観点にたつて、非感染性の疾患に対するリスクを考える場合、それぞれの因子に固有のリスクが存在すると考えるよりは、調査対象に含めら

表 1 「日本公衆衛生雑誌」における多変量解析法の使用頻度

年 度	外的基準がある場合						外的基準がない場合						A	B	C	D
	重回 帰分 析	正 準 相 関 分 析	判 別 分 析	数 量 化 一 類	数 量 化 二 類	多 重 ロ ジ ス テ ィ ッ ク モ デ ル	主 成 分 分 析	因 子 分 析	数 量 化 三 類	ク ラ ス タ ー 分 析	手 法 の 数	多 変 量 解 析 法 の 数	総 論 文 数	多 変 量 解 析 法 の 数	多 変 量 解 析 法 の 数	多 変 量 解 析 法 の 数
1963-1972	1	0	1	0	0	0	0	3	0	0	5	5	—	—	—	—
1973	1	0	0	0	0	0	1	0	0	0	2	2	36	5.6	—	—
1974	0	0	1	0	0	0	0	0	0	0	1	1	36	2.8	—	—
1975	0	0	0	0	0	0	1	4	0	0	5	5	44	11.4	—	—
1976	2	0	0	0	0	0	0	0	0	0	2	2	49	4.1	—	—
1977	2	0	2	0	1	0	0	2	0	0	7	7	62	11.3	—	—
1978	3	1	1	1	3	0	1	2	1	1	14	11	57	19.3	—	—
1979	7	1	3	0	1	1	2	5	0	2	22	14	58	24.1	—	—
1980	2	0	1	0	0	0	1	1	0	2	7	5	51	9.8	—	—
1981	1	1	0	0	2	0	1	2	0	3	10	6	43	14.0	—	—
1982	0	0	2	0	0	0	0	1	0	0	3	2	38	5.3	—	—
1983*	1	0	0	0	0	0	1	0	0	1	3	3	19	15.8	—	—
計	20	3	11	1	7	1	8	20	1	9	81	63	493	12.8	—	—

(注) * (6 月号まで) (D は $B/D \times 100$)。

表 2 「American Journal of Epidemiology」における多変量解析法の使用頻度

年 度	外的基準				パス 解析	手 法 の 数	多 変 量 解 析 法 の 数	総 論 文 数	多 変 量 解 析 法 の 数	多 変 量 解 析 法 の 数
	あ り	な し	あ り	な し						
1971	1	1	0	0	0	2	2	117	1.7	—
1972	3	2	0	0	1	6	5	105	4.8	—
1973	3	0	1	0	0	4	3	89	3.4	—
1974	3	0	0	0	0	3	3	95	3.2	—
1975	4	0	0	0	0	4	4	103	3.9	—
1976	6	4	1	0	0	11	11	105	10.5	—
1977	3	1	0	0	0	4	4	99	4.0	—
1978	7	2	8	0	0	17	17	100	17.0	—
1979	6	4	4	1	0	16	14	109	12.8	—
1980	6	2	6	0	0	14	13	110	11.8	—
1981	14	2	10	0	0	26	24	153	15.7	—
1982	13	0	15	0	0	28	27	172	15.7	—
計	69	18	45	1	1	135	127	1357	9.4	—

表 3 臨床医学における多変量解析法の使用頻度 (日本 ME 学会, 日本行動計量学会, 日本臨床病理学会での発表論文による, 資料は文献 No. 15 より引用)

年 度	外的基準がある場合				外的基準がない場合				合 計
	重回帰分析	判別分析	数量化一類	数量化二類	多重ロジスティック分析	主成分分析	因子分析	クラスター分析	
1974	1	1	0	0	0	0	1	0	3
1975	0	1	0	2	0	3	1	1	9
1976	0	3	0	1	1	6	1	2	15
1977	3	1	0	2	0	4	1	1	13
1978	3	9	2	4	1	6	2	0	27
1979	2	4	0	2	1	9	3	0	21
計	9	19	2	11	3	28	9	3	88

れた, 相関の強い因子の全体を1つのセットとして考える, マルチプルリスクの概念¹⁴⁾が有用となってくるわけである。(この例については, 適用例1を参照のこと)

■学会誌にみる多変量解析の応用

次に公衆衛生, 疫学の分野で, どのような多変量解析の手法が, どのような研究テーマに結びつけられて使用されているかを, より詳細に検討してみた。検討のための資料として, 「日本公衆衛生雑誌」(以下日本公衛誌と略記する) の1963年1月号から1983年6月号までに掲載されている論文と, 1972年から1982年までに発行された「American Journal of Epidemiology」(以下, A.J.E. と略記) に掲載されている論文で, どのような多変量解析の手法が, どのような研究テーマに使用されているかを表1, 2, 4および5に示した。表1, 表2に共通してみられる特徴は, 日本公衛誌, A.J.E. においてともに, 多変量解析を使用した論文の割合が, 1970年代の後半になって著しく増加している点である。A.J.E. の場合, 平均の使用率は9.4%と, 日本公衛誌を下まわるが, 1978年以降で比較すると, A.J.E. の場合, 14.8%となり日本公衛誌の15.4%にはほぼ匹敵する。しかし, A.J.E. と日本公衛誌を比較すると, 使用された多変量解析の手法の種類が著しく異なる。すなわち, A.J.E. の場合, 外的基準のない場合の方法である主成分分析, クラスター分析は, わずかに1論文ずつ使われたのみで, 因子分析にいたっては1度も使用されていない。A.J.E. で用いられ

表 4 「日本公衆衛生雑誌」における多変量解析手法が使用された論文の内容

手 法 内 容	外的基準がある場合						外的基準がない場合				論 文 数	手 法 の 数
	重回帰分析	正準相関分析	判別分析	数量化一類	数量化二類	多重ロジスティックモデル	主成分分析	因子分析	数量化三類	クラスター分析		
環境要因	3	0	1	0	1	0	0	5	0	0	9	10
個体要因	8	1	2	0	1	0	1	4	0	0	14	17
食生活	1	0	0	0	0	0	3	4	0	5	9	13
疾病の疫学	7	1	8	0	3	1	3	6	1	1	22	31
医療行動	0	0	0	1	2	0	1	0	0	3	6	7
方法論	1	1	0	0	0	0	0	1	0	0	1	3

表 5 「American Journal of Epidemiology」における多変量解析が使用された論文の内容

手 法 内 容	外的基準			パ ス 解 析	計
	あ り		な し		
	重回帰分析	判別分析	多重ロジスティックモデル		
個体要因	47	5	1	1	55
疾病の疫学	20	13	38	0	72
方法論	2	0	6	0	8
計	69	18	45	1	135

た分析法としては, 重回帰分析が最も多いが, 年を追って, 多重ロジスティックモデルの使用頻度が高くなっていくのが顕著な傾向である。これに対し, 日本公衛誌の場合, 外的基準のない場合の手法が使用された回数の総計は, 38で, 全体の47%を占め, A.J.E. に比較すると, その割合は著しく高い。さらに A.J.E. で全く使用されていない因子分析が20の論文で適用されているのが目につく。また日本公衛誌の場合, A.J.E. と比較すると, 多重ロジスティックモデルが使用されたのは僅かに1論文²¹⁾にすぎない。これに対して, 我が国で開発された林知己夫による数量化理論のうち, とくに二類が7論文で使用されている。数量化二類で取り扱われる項目カテゴリー形式の名義尺度のデータに関しては, 判別分析は適用できないのは明らかであるが, 多重ロジスティックモ

デルには、カテゴリカルデータを説明変数として取り入れることができるので、A.J.E. のいくつかの論文では、このような形で、カテゴリカルデータが処理されている。

なお、表 3 に 1974 年から 1979 年の 5 年間の臨床医学に関する論文における多変量解析の利用状況を、古川ら¹⁵⁾の資料に基づき集計したものを示した。判別分析が多く使われている理由は、臨床医学に計量診断が含まれていることから当然のことであるが、主成分分析が多い理由は、集団検診における生化学検査や血液検査などの各検査の構造分析によるものである。

■分析手法とその内容

つづいて、分析の手法と研究のテーマの関連を示すクロス表(表 4、表 5)について考察しよう。

個体要因と環境要因に関する論文は、厳密には区別しにくいものであるが、悪臭¹⁶⁾、騒音¹⁷⁾、さらに疾病と環境要因の関連分析で、とくに環境に重点をおいたものは、環境要因として分類した。A.J.E. の場合、環境要因に分類される論文は皆無であって、疾病の疫学に分類されない多くの論文は個体要因に関した分析であった。日本公衛誌の場合、個体要因としては、体重、身長などの体格要因を取り扱ったものが数多くみられたが、A.J.E. の場合は、血液検査や生化学検査の検査値の 1 つ(例えば血圧〔収縮期、拡張期〕や HDL コレステロール)を基準変数として、その変動を他の検査成績で説明するために、重回帰分析を用いるといったタイプの研究が数多くみられている。(後述する適用例の例 2 を参照のこと)

また、疾病の疫学に分類される研究には、フラミンガム調査の開始以降、前向き調査に基づくリスクの推定に関する多くの研究が生まれている。これらの分析に用いられる手法の中心は、外的基準のある場合の手法であることはいうまでもないが、日本公衛誌の場合は、疾病の疫学に分類される研究に因子分析、主成分分析といった手法が多く使われている。このうち、地域単位(県、国)に基づく死亡率データの構造分析に、主成分分析や因子分析の適用される¹⁸⁾のは当然であるが、通常、判別分析や重回帰分析が適用される説明変数($X_1, X_2, \dots, X_p$)のデータにも因子分析や主成分分析を適用することも、決して無駄なことではない。外的基準のある場合の分析法によって導かれる結果は、外的基準変数の選び方に著しく依存するものであるから、判別分析に適用するデータの判別の基準(例えば、患者の病名をつける診断基準)が明確でないような場合には、説明変数に基づく因子分析によって得られた因子得点によって患者を分類し、診

表 6 分析に用いた対象者数

年齢階層		30歳代	40歳代	50歳代	60歳代
対象	胃がん群	129	200	200	200
	対 照 群	200	200	179	92
対の 照内 群訳	異常所見なし	100	66	18	7
	萎縮性胃炎	100	134	161	85

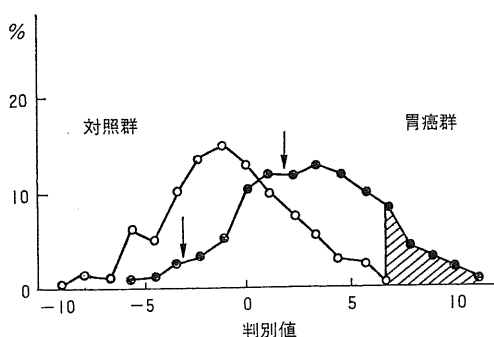


図 1 胃癌群と対象群の判別値の分布

断基準を再構成することも必要であろう。

疾病の疫学的研究として、外的基準のある場合とない場合の方法を同時に適用したものに、駒沢¹⁹⁾による動脈硬化症の分析、奥津ら²⁰⁾による更年期障害の自覚症状の分析等がある。

この他、表 4 で特筆されることは、食生活に關したデータに外的基準のない場合の多変量解析の手法が数多く使用されている点である。ここでは、このような研究の嚆矢とみられる研究の文獻^{22,23)}を掲げるにとどめよう。また、質問紙による健康調査法の作成にも、多変量解析法が使われている^{6,24)}。

■適用例

ここで、日本公衛誌と American Journal of Epidemiology に掲載された多変量解析使用論文のうち 2 つを紹介しよう。

〔例 1〕数量化二類を用いた胃がんの Risk factor に関する統計的解析²⁵⁾

データは 1967 年 1 月～1972 年 5 月までの 5 年 5 ヶ月間に、愛知県がんセンター内科消化器系専門外来を受診した胃がん患者、対照群(胃に異常所見の認められない患者、および萎縮性胃炎の患者)に対して、栗田英男によって行われた面接により、retrospective に聴取された疫学データを用いてある。各年齢階層の例数を揃えるため、男性のみが分析対象とされ、対照群は無作為に抽出されている。(表 6)

表 7 数量化二類によって得られた各カテゴリーの重み

項 目	カテゴリー	男 子			
		30歳代	40歳代	50歳代	60歳代
学 歴	義務教育	0.39	-0.28	-0.36	0.46
	その他	-0.30	0.28	0.78	-1.38
職 種	専門職	-0.76	-1.07	-0.50	-5.41
	販売業	0.48	0.41	0.89	-0.77
	運転手	0.36	0.59	-0.16	-0.29
	農業・漁業	0.24	-3.27	-3.83	-1.71
	技能職	-3.24	2.34	2.80	1.07
	単純作業	-0.03	-0.98	-0.80	1.32
職場環境	埃・熱	2.29	2.87	2.16	0.31
	肉体疲労	1.17	-0.14	1.08	0.92
	神経疲労	-2.76	-2.24	-3.41	-2.65
	特記なし	-4.12	-1.79	-0.37	-0.36
収入（月収）	5万円以下	0.24	3.58	-1.56	-0.45
	～6万円	3.42	2.37	1.30	-2.23
	～8万円	1.00	2.65	2.05	1.13
	～10万円	-3.59	-2.16	-1.47	0.09
	～12万円	-3.66	-0.22	-2.05	2.80
	～15万円	1.67	-0.81	-2.28	-0.20
	15万円以上	2.12	-2.13	0.73	-0.96
労働時間	8時間未満	1.89	-0.81	0.83	1.59
	8時間以上	-0.59	0.79	-0.21	-0.40
	不定	-0.61	-2.59	-0.29	-1.48
米飯量	パン	2.99	0.87	-0.63	-3.18
	7杯まで	-1.81	-0.53	-0.56	-1.05
	7杯をこえる	1.21	0.44	0.67	1.03
食事の速さ (*)	遅い	-1.70	-0.94	0.77	-1.49
	普通	1.35	0.03	-1.41	1.85
	速い	0.13	0.37	0.24	-0.25
食事の熱さ (**)	冷たい	-0.98	-4.64	-0.19	-0.50
	普通	0.74	1.48	0.08	0.56
	熱い	-0.20	-0.02	-0.04	-0.31
コーヒー摂取	なし	0.68	1.73	-0.29	-0.95
	ときどき	-1.00	-1.46	0.31	0.90
	毎日	-0.19	-1.50	0.45	2.17
牛乳摂取	なし	0.48	-1.17	0.39	-0.42
	ときどき	-1.02	-0.18	-0.10	0.46
	毎日	0.81	1.60	-0.55	0.24
高塩分食品摂取	なし	-0.28	-0.79	0.63	0.02
	普通	-1.07	-0.49	-2.50	-2.60
	多い	0.26	0.36	0.61	0.90

上記のデータに数量化二類が適用され、算出された各項目のカテゴリーに対する重みは、表7に与えられている。重みが正でその値が大きいほど、そのカテゴリーが胃がんに対してハイリスク因子であることを示す。しかし、この値は重回帰分析での偏回帰係数に相当するもので、すでに述べたように、各カテゴリー単独の胃がんに対する危険度を示すものでない。

判別値による胃がん群と対照群の2群の重心を分割点 (Cutting Point) としたとき、各年齢階層を通じての正診率は70%台であった。図1が判別値による分布で、2群がそれぞれ正規分布にしたがうものと仮定すると、50歳代では、判別値が胃がん群の平均値より1標準偏差大きい7.85以上ならば、胃がんになる確率（尤度）は92%と推測される。

ここで得られた表7の重みを巨視的な立場からみると、職業（とくに30代を除く技能職）、職場環境（塵埃、肉体、疲労）、高塩食品摂取、多量米飯摂取、食事の速さ、多量飲酒、高喫煙などが胃がんの高危険因子群として浮び上がってくることになる。つまり平易に言えば、米飯を1日7杯を超えて摂取し、食事の食べ方が速く、多量飲酒や高喫煙、さらに塩辛い物が好きという条件を多く満たす人ほど、胃がんの危険率が高まることを意味するものである。ところで牛乳を多量に飲むことが、胃がんのリスクファクターとして抑制的に作用するか否かは、本結果では必ずしも統一性がみられない。しかし、牛乳摂取と米飯摂取量の間には逆相関があり、牛乳摂取と果物摂取には正相関がみられることから、牛乳、パン食、果物といった西欧化された食生活のパターンが、胃がんのリスクファクターの1つのセット、すなわち、マルチプルなリスクとして抑制的に作用するものと考えれば、表7の結果も妥当性をもつものといえる。

〔例2〕血圧の規定要因に関する研究

ここで、A.J.E. 誌に掲載されているフラミンガム Offspring Study における血圧に関する研究を取り上げてみよう。

これは、血圧に影響を及ぼす要因がそれぞれ、どれ位血圧の変動を説明するかをみたも

飲 酒 量	な し	0.09	0.54	0.50	0.77
	ときどき	-1.14	-1.56	-1.76	-0.07
	毎日ビール	0.18	-1.94	1.11	-1.45
	180 ml/日	-1.53	0.16	-1.73	-0.88
	360 ml/日	2.43	-1.18	0.13	0.94
	540 ml/日未満	-0.32	3.61	3.01	0.73
	540 ml/日以上	5.25	5.42	3.05	-5.77
喫煙量 (本／日) × (喫煙年数)	な し	-2.00	-2.24	-4.65	-2.46
	100	0.87	-1.87	-2.80	-1.81
	200	-1.73	0.78	0.41	-3.96
	400	1.04	-1.95	-1.37	-0.60
	600	-0.43	0.78	-1.65	0.33
	800	4.47	2.64	3.12	-0.96
	1000以上	7.04	3.23	3.34	2.24
親族の胃癌	2人以上	4.03	-0.09	-1.91	2.50
	2人癌(***)	1.78	4.95	-3.17	5.62
	1人癌	-0.21	-0.14	-0.01	0.00
	1人胃癌	-0.53	0.29	0.71	-1.78
	な し	1.28	-0.26	0.60	-0.40

(*) 朝昼食10分以内、夕食15分以内……「速い」

朝昼食20分以上、夕食30分以上……「遅い」

その中間……「普通」

(**) 朝食のみそ汁の熱さを基準にして判定

(**) 1 人が胃癌，他の 1 人がその他の癌の場合

(文献 No. 25 より引用)

ので、血圧の変動を規定する説明変数として、両親の 25 年前の血圧、心拍数、体重、血清総タンパク、アルコール消費量、空腹時血糖、血清リン、ヘマトクリット値、血清中性脂肪をとりあげ、収縮期血圧 (SBP) を基準変数として男女別に重回帰分析を行い、表 8 のような結果が得られた。その結果、男子の場合、体重、心拍数、血清総タンパク、女子の場合、年齢、体重、心拍、ヘマトクリット値、及び母親の血圧が収縮期血圧に強く関係することが解った。なお、重相関係数の大きさから、上記の要因が SBP に影響する度合は、男子に比べて、女子の方がより強いことが示唆される。

■多変量解析法の適用をめぐる

ところで、公衆衛生学や疫学の研究に、多変量解析の手法の利用が、1970年代の後半になって加速度的に増加してきた最も大きな理由は、SPSS, BMD, さらに最近ではあるが、SAS といったプログラムパッケージが、容易に利用できることになったことであろう。さらに、最近の著しいマイコンの普及も多変量解析法を、相関係数の計算や平均値の差の検定のような“at home”な分析法に替えて行くことを可能にした。したがって、多変量

表 8 収縮期血圧 (SBP) に対する重回帰分析 (男女20歳から49歳までのデータによる)

説 明 変 数	男子(n=585)		女子(n=598)	
	標準化 回 係	F	標準化 回 係 数	F
心 拍 数	0.196	26.2	0.249	51.5
体 重	0.231	39.9	0.260	55.3
血清総タンパク	0.186	24.2	0.090	7.2
アルコール 消費量	0.107	8.6	0.063	3.9
年 齢	0.016	0.2	0.173	21.2
血清リン	-0.084	5.0	-0.108	9.0
空腹時血糖値	0.094	6.5	0.062	3.0
ヘマトクリッ ト値	0.097	6.9	0.121	13.1
血清中性脂肪	0.057	2.3	0.081	5.1
父親の SBP	0.097	5.7	0.129	13.8
母親の SBP	0.071	3.8	0.048	2.1
重相関係数の 平方	0.278		0.404	

$F > 3.84$ 5%で有意, $F > 6.63$ 1%
で有意 (文献 No. 26 Table 8 を翻訳し
たもの)

化石(アウストラロピテクス・アフリカーヌス)

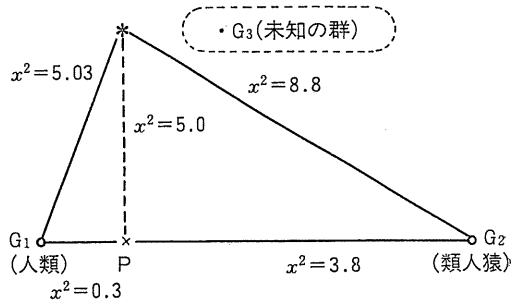


図 2 未知の群への判別 (文献 No. 27 による)

(注) 2群の判別分析のプログラムでは、通常 P と G_1 , P と G_2 の距離が比較されるが上記の例のように $*$ と G_1 , $*$ と G_2 を比較すると、化石が G_1 , G_2 の群に所属する可能性はともに小さくなる。

なお, χ^2 は, それぞれの距離 (マハラノビス汎距離) に対応するカイ 2 乗値.

解析の数学的論理を完全に把握していない場合でも、多変量解析のプログラムパッケージによって、自己のデータを分析できるようになったわけである。これはもちろん、喜ばしいことには相違ない。しかし、多変量解析の

適用にあたっては、適用すべき手法の選択や、結果の解釈に際して、必要とされる最低限の予備知識をもっていることが好ましいのはいうまでもない。また、多変量解析によって得られる結果は、すべて分析に用いる変数（または項目）間相互の相関係数やクロス集計表が基礎になって導かれるものであるから、それぞれの変数に関する分布（平均値、標準偏差など）はもちろんのこと、相関係数やクロス集計の結果についても十分注意する必要がある。

例えば、分析の対象となる変数群の相関が著しく低い場合を考えよう。相互に著しく相関の低い変数群を因子分析にかけたとしても、何も新しい因子は得られない。このような相関係数をもつ行列に主因子法—バリマックス回転を行ったとすれば、それぞれの変数の独自性が高いという、自明な結果が得られるのみである。このような場合、各変数間に曲線的相関が存在しないか、さらには、被験者をいくつかのグループに分類することによって、意味ある結果が得られるか否か等について探索的に分析していくことが必要とされよう。

また、分析にあたっては、既存のプログラムを用いる場合でも、そのプログラムのもつ限界について十分認識しておくことが必要である。

ここで、今日の世界の統計学会をリードする C.R. ラオの講演²⁰⁾から引用した一例を紹介しよう。

アフリカで発見された化石が、人類のものか、類人猿のものかを判定するために、既成の判別分析のコンピュータプログラムにかけたところ、その化石は、類人猿よりは人類により近いと判定されたとしよう。2次元空間における化石の位置は、図2の*にあったとすれば、この化石は、人類、および類人猿のグループの重心の位置のいずれにも遠いことになる。したがって、この化石は、その存在をこれから確認しなければならないような未知のグループ (G_0) に所属している可能性が強くなってくる。

上記の例から明らかなように、個別の研究テーマにびったり合う既成のプログラムというものは存在しない。したがって、既成のプログラムを使う場合でも、自分の研究テーマに応じて、必要とされる最低限の修正は施すべきであろう。

■おわりに

公衆衛生学や疫学の研究においては、前回の講座（その2—外的基準のある場合）で解説された因果関係の分析が重要である。因果関係の分析法は、これまでにいくつかの方法が提案されているが、すべてのデータに共通

した最適な因果分析の方法というものは存在しない。先に述べたリスクの推定にも、多重ロジスティックモデルを利用することは望ましいことには違いないが、この方法によると、時間の変数が十分に考慮されていないという、問題点があるので、時間のファクターを考慮した多変量解析法である指数ワイブルモデル²⁸⁾やコックスの比例ハザードモデル²⁹⁾を併用することが望まれる。

したがって、与えられたデータに最適な一つの手法というものは存在しないから、いくつかの手法を併用して、さまざまな角度からデータを分析し、研究者自身の視点から、それらの結果を再構成することが必要である。

文 献

- 1) 鳥居敏雄，他：医学生物学のための推計学，東大出版会，1954.
- 2) 崎野滋樹，巷野悟郎：数量化による予測の判定，統計研究報，1；1，1953.
- 3) 高橋正編：計量診断学，東大出版会，1970.
- 4) 久保武士，本多洋他：妊娠中毒症における周産期死亡の予測，産婦人科の世界，25；6，1973.
- 5) 斎藤陽一，繁井算男：精神科領域における計量診断，高橋正編，宮原英夫編：臨床診断とコンピュータ，産業図書，1972.
- 6) 青木繁伸，他：新しい質問紙健康調査票（TH-PI）作成のこころみ，行動計量学，2；1，1974.
- 7) 北沢克明：緑内障に関する諸問題，日本眼科学会雑誌，79；11，1975.
- 8) 古川俊之他編：計量診断，最新医学，33；1，1978.
- 9) 宮原英夫：全身性エリテマトーデスのクラスター分析，医用電子工学，13；3，1974.
- 10) 平山 雄：予防ガン学，新宿書房，1977.
- 11) Yasuo Hirota, et al. : A Multivariate analysis of risk factors for cerebrovascular diseases in Hisayama, Kusu Island, Behaviormetrika No. 2, 1975.
- 12) 鏡森定信：老人の循環器疾患死亡の予知に関する研究，日本公衛誌，20；8，1973.
- 13) 大和田里夫他：農村婦人における貧血発生の諸要因に関する疫学的検討，日本公衛誌，21；7，1974.
- 14) 古川俊之：コンピュータ診断，共立出版，1982.
- 15) 古川俊之，田中 博：多変量解析プログラムパッケージ入門，医学書院，1983.
- 16) 神谷明男他：悪臭に関する研究（第1報）悪臭環境の悪臭物質濃度と官能試験結果との関係，日本公衛誌，25；5，1978.
- 17) 小林章雄他：道路交通騒音に対する主婦のさわがしさの感じについて，日本公衛誌，26；11，1979.
- 18) 稲葉 裕他：わが国における部位別がん死亡率

- の地理的分布に関する因子分析, 日本公衛誌 26; 2, 1979.
- 19) 駒沢 勉: 数量化理論による動脈硬化性疾患の予後予測に関する研究, 日本公衛誌 25; 3, 1978.
 - 20) 奥津則子他: 更年期婦人の自覚的な訴えに関連する環境心理的要因, 日本公衛誌 28; 1, 1981.
 - 21) 高木廣文: 糖尿病長期治療患者に関する疫学研究(3): 多重ロジスティックモデルによる眼底網膜症のリスクファクターに関する研究, 日本公衛誌 26; 4, 1979.
 - 22) 豊川裕之他: わが国の食物摂取に関する研究(第1報) 全国規模の分析, 日本公衛誌 22; 10, 1975.
 - 23) 豊川裕之: 食生活と健康, 大修館, 1976.
 - 24) 柳井晴夫: 質問紙法によるコンピューター診断, 数理科学 8月号, 1983.
 - 25) 柳井晴夫, 栗田英男: 胃がん, 医学のあゆみ, 計量医学特集, 110; 13, 1979.
 - 26) R.J. Havlik et al.: Blood pressure aggregation in families, *American J. of Epidemiology*, 10; 3, 1979.
 - 27) C.R. Rao MULTIVARIATE ANALYSIS: Some reminiscence on its origin and development, *Technical Report 82-23, Center for Multivariate Analysis*, 1983. (柳井晴夫, 竹内啓訳ラオ, 多変量解析—その起源と発展に関する回想—, 応用統計学, 12巻3号, 1983)
 - 28) A.R. Dyer: An Analysis of the Relationship of Systolic Blood Pressure, Serum Cholesterol, and Smoking to 14-Year Mortality in the Chicago Peoples Gas Company Study 1. Total Mortality in Exponential Weibull Model *J. Chron. Dis.* 28: 1975.
 - 29) D.R. Cox: Regression models and life-tables; *J. of Roy. Statist. Soc.* 13; 34, 1972.

第一線研究者の最新兵器 コンピュータ自由自在

多変量解析プログラム入門

新刊

古川俊之 東京大学教授
田中 博 東京大学講師

本書は、今後ますます必要とされるコンピュータによる計量診断を利用者としての研究者の立場から解説している。著者らが扱った医学上の問題を例題として、分析のねらい、プログラムパッケージの使用法、結果の解釈を中心にただちに実践にとびこめるよう記述。失敗を避けるために「より深い理解のために」の項を設けた。

● 主要内容

コンピュータとは 重回帰分析 判別分析 主成分分析 因子分析 クラスター分析 数量化Ⅰ, Ⅱ類 数量化Ⅲ, Ⅳ類 多重ロジスティック, 多重ワイブル関数 多変量データと分布論 多変量解析プログラムパッケージ

● B5横 頁256 図152 写真18 1983
¥5,800 円300



医学書院

本社 1113-91 東京・文京・本郷5-24-3 ☎東京(03)811-1101(代) 振替東京7-96693
洋書部 1113 東京・文京・本郷1-28-36 鳳明ビル ☎東京(03)814-5931~5 振替東京1-53233