

〈総説〉

看護研究における多変量解析利用の現状と 多重ロジスティック・モデルの紹介

西川浩昭¹

A Note on Multivariate Analysis and Multiple Logistic Model Applied to the Nursing Studies

Nishikawa Hiroaki¹

Abstract

Recently, many kinds of multivariate analysis have been applied in various fields of science. Especially, these analyses are familiar methods in the studies of social medicine, such as in the epidemiology. However, in nursing field of our country, it is hard to find the studies in which those analyses was applied, whereas a large number of studies published in foreign country used those methods for analyzing their data.

The present study shows several reasons why those methods are not applied to the nursing study in our country. Since multiple logistic model is considered as a most suitable method for adjusting confounding factors, this method should be applied to the studies of social medicine or nursing science, as a useful tool.

Key Words : Multivariate analysis, Multiple logistic model, odds ratio, 2 × 2 table

はじめに

多変量解析が生まれた心理学の領域は言うに及ばず、現在では医学、特に社会医学の領域においても疾病などの要因分析を目的として多変量解析が多用されている。社会医学に近いところに位置していると言われている看護学の領域においてもその傾向が見られることが予想されるが、現実には状況を異にしている。以下ではその実態を明らかにし、今後利用が増加すると思われる手法に関して解説する。

看護領域における多変量解析の利用状況

1991～2000年の10年間に看護学領域の雑誌に原著として掲載された論文で用いられている多変量解析の手法を全て列挙し、数え上げた。対象とした雑誌は、看護系の和雑誌として「日本看護科学会雑誌」, 「看護研究」, 同じく洋雑誌として「Nursing Research」(以下「Nurs. Res.」と略), 比較対照とし

て領域的に近いと考えられる「日本公衆衛生学雑誌」(以下「公衛誌」)の4誌である。なお、ここで列挙の対象とした分析手法は3変量以上を同時に分析変数として扱っている分析法であり、単変量の基本統計量や2変数間の相関係数, 1変数をグループ変数として扱った統計的検定や1元配置の分散分析は対象外とした。これらの手法は、数値を扱う際には当然行うものであり、要因分析のための分析手法には属さないと考えたためである。

図1に過去10年間の年次別の多変量解析が用いられた論文の割合を示した。年次別に見た場合、論文数が少ないため、変動が大きいため、より明確にするために対象年次を1991～1995年(以下前半), 1996～2000年(以下後半)と2期に分けて、利用状況を図2に示した。看護系の和雑誌では、利用されている割合が変化していないか、減少しているのに対し、Nurs. Res.と公衛誌では前半と後半の間で10%程度の

1. 筑波大学医療技術短期大学部 〒305-8577 茨城県つくば市土台1-1
1. College of Medical Technology and Nursing, University of Tsukuba, Ibaraki.

増加が見られる。

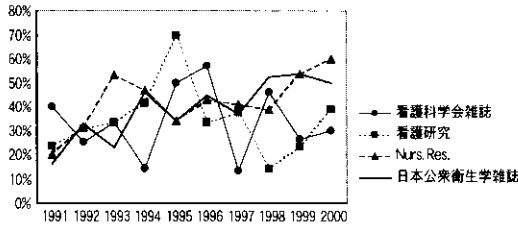


図1. 多変量解析の利用状況の変化

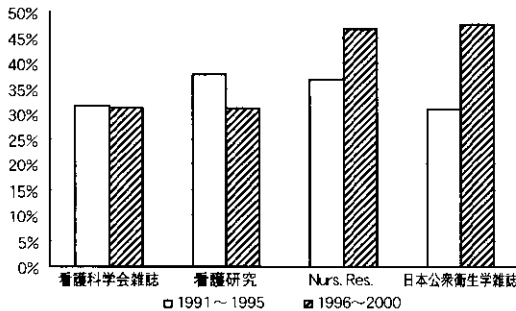


図2. 多変量解析の利用状況

表1. 手法別多変量解析利用状況

	看護学会雑誌	看護研究	Nurs. Res.	日本公衛誌
回帰分析				
重回帰分析	11	8	112	74
数量化Ⅰ類	2	3		2
一般線形モデル		1		8
判別分析				
判別分析	1		5	5
数量化Ⅱ類	1			7
要因分析				
主成分分析	2			11
因子分析	16	31	30	22
数量化Ⅱ類				9
正準相関分析			3	
共分散構造分析	2		7	7
非線形モデル				
ロジスティックモデル	2	1	14	88
比例ハザードモデル				15
その他				
クラスター分析				4
多次元尺度法		1		1
総論文数	102	122	427	600

表1に10年間に用いられた手法別の頻度を示した。全誌とも重回帰分析と因子分析の使用頻度が高くなっていったが、看護系の和雑誌2誌は他の2誌に比べて明らかに非線形モデル、なかでも特にロジスティック・モデルの頻度が低くなっていた。

多重ロジスティック・モデルの概要

前章の結果を踏まえて、以下では多重ロジスティック・モデルについて解説する。

多重ロジスティックモデルは、ある一つの目的とする変数に様々な要因が拘わっていると仮定される場合に、他の要因の影響を取り除いた、影響の程度を推定する方法であり、ロジスティック回帰分析とも呼ばれ回帰分析に分類される場合もある。また疫学の領域ではその目的のためにRelative Risk Functionと呼ばれることもある。方法自体はCornfieldにより1961年に発表された方法である¹⁾。その後、いくつかの改良が行われ^{2), 12-13)}、諸外国はもとより、我が国においても主に医学領域で用いられている^{4), 6-7), 9-11)}。また、近年では医学領域に限らず、様々な領域で用いられている⁸⁾。ただし一般の回帰分析と大きく異なる点は、複数の要因が存在する場合、それらに関して相加性ではなく、相乗性を仮定している点である。相加性と相乗性の違いは、ある要因Aが存在した場合50%増加(1.5倍)し、もう一つの要因Bが存在した場合100%増加(2倍)するとした場合、二つの要因A, Bが同時に存在した場合、150%(50%+100%)増加すると考える場合が相加性であり、3倍(1.5×2)になると考える場合が相乗性である。

2×2分割表

多重ロジスティック・モデルの解説に入る前に、多重ロジスティック・モデルで扱うデータのもとなっている2×2分割表について解説する。2×2分割表は表2, 3に示すような形式のもので、それほどめづらしいものではない。なお、要因の性質として

表2. 2×2分割表(I)

特性A	特性B		計
	有	無	
有	n_{11}	n_{12}	$n_{1\cdot}$
無	n_{21}	n_{22}	$n_{2\cdot}$
計	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

但し、 $n_{\cdot\cdot}$ の・はその添字の合計を示す。
ex. $n_{1\cdot} = n_{11} + n_{12}$

特性Aは先行因子すなわち原因と見なされるものを示し、特性Bは結果と見なされるものを示している。

正式には、表2で示した形の表示法を用いるが、以下では分かりやすくするために表3の記号(a, b, c, d, N)を用いて示す。

表3. 2×2分割表(II)

特性A	特性B		計
	有	無	
有	a	b	a + b
無	c	d	c + d
計	a + c	b + d	a + b + c + d = N

オッズ比

オッズ比(odds ratio 見込み比, cross-product ratio 交差積比とも呼ばれる)とは、1961年にCornfieldによって提唱された先行因子と結末的な特性たとえば、罹患、死亡などとの関連の大小を測る尺度である¹⁾。また、これはRelative Riskの近似値としてかなり良い値を示すという性質を有している²⁾。また、最近では正確にはオッズ比であるにも拘らず、relative riskと記載している報告もかなりみられる。

以下では、その算出法について表3に示した2×2分割表の記号を用いて説明する。

1) オッズ(odds)

オッズとは、ある先行因子(原因)存在下において結末的な特性が現れる可能性の尺度である。つまり特性Aが存在している(いない)ときに特性Bが発生する見込みである。

特性Aが存在しているときに特性Bが起こる確率は、 $a / (a + b)$ であり、Bが起きない確率は、 $b / (a + b)$ である。このときのオッズ O_A は、

$$O_A = \frac{a / (a + b)}{b / (a + b)} = a / b$$

他方、特性Aが存在していないときのオッズ $O_{\bar{A}}$ は、

$$O_{\bar{A}} = c / d$$

となる。

2) オッズ比(odds ratio, OR)

1)で算出した2つのオッズの比を用いることにより2つの特性間の関連性の尺度を作成することができる³⁾。しかしこの考えは古くからあり、1961年にCornfieldによって提唱される¹⁾以前にも存在していた。

Yule¹⁴⁾では、

$$Q = \frac{O_A - O_{\bar{A}}}{O_A + O_{\bar{A}}}$$

また、Yule¹⁵⁾による別種の尺度は、

$$Y = \frac{\sqrt{O_A} - \sqrt{O_{\bar{A}}}}{\sqrt{O_A} + \sqrt{O_{\bar{A}}}}$$

である。しかし現在、最も一般的に用いられているのは、1973年にFleissにより発表されたオッズを単純な比で対比させたものであり³⁾、

$$OR = \frac{O_A}{O_{\bar{A}}} = \frac{a/b}{c/d} = \frac{a d}{b c} \dots\dots\dots (1)$$

で推定される。この式の形式からオッズ比のことを交差積比(cross-product ratio)と呼ばれることもある。もし2つのオッズ O_A と $O_{\bar{A}}$ の値が等しければ、2つの特性は独立ないし無関連であり、オッズ比=1となる。

またこのオッズ比の標準誤差は近似的に、

$$SE|OR| = OR \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

で示される。この標準誤差は推定したオッズ比の精度を見るためには役立つが、オッズ比の有意性の検定には利用できない。母集団での真のオッズ比が1であるという仮説の検定にはカイ2乗検定を用いなくてはならない。

3) 修正オッズ比

2)の様に定義され、関連性の指標として広く利用されているオッズ比ではあるが、唯一難点がある。というのは(1)式にあるa, b, c, dのうち、b, cの値が0の場合にはオッズ比そのものが、またa, b, c, dのうちいずれか一つでも0の場合には標準誤差が推定出来なくなる。この点を含めて、GartとZweifel⁵⁾は、

$$OR' = \frac{(a + 0.5)(d + 0.5)}{(b + 0.5)(c + 0.5)}$$

$$SE\{OR'\} = OR' \sqrt{\frac{1}{a+0.5} + \frac{1}{b+0.5} + \frac{1}{c+0.5} + \frac{1}{d+0.5}}$$

がよりよい推定値であると示唆している。

4) 対数オッズ比

表2の様な2×2分割表が与えられた場合、特性Aが存在しているときのオッズ O_A は、以下の形に変形できる。

$$\begin{aligned} O_A &= \frac{a / (a + b)}{b / (a + b)} = \frac{a / (a + b)}{(a + b - a) / (a + b)} = \\ &= \frac{a / (a + b)}{1 - a / (a + b)} \end{aligned}$$

ここで、 $a / (a + b)$ は特性Aが存在しているときに特性Bが起こる確率であるから、これを p_A とすると、分母は $1 - p_A$ となるから、

$$O_A = p_A / (1 - p_A)$$

とすることができる。

他方、特性Aが存在していないときのオッズ $O_{\bar{A}}$ は、

$$O_{\bar{A}} = p_{\bar{A}} / (1 - p_{\bar{A}}) = (1 - p_A) / p_A$$

となる。

また、オッズとは確率の比であるから、0以上である。ここで、オッズが0となる場合を考えると $a = 0$ のときのみであり、この時は特性Aがあった場合には特性Bが起きないことを示している。また、オッズが0となるときは $b = 0$ のときであり、この場合には特性Aがあった場合には必ず特性Bが生じることを示している。この極端な2つの場合を除けば、 O_A は正で有限の値を取るようになる。このとき、 O_A の対数値を取ると、これは全ての値を取ることができる。

多重ロジスティック・モデルの理論

ある結末的な特性Bの生起が、 n 個の要因 x_1, x_2, \dots, x_n で説明づけられるとする。これは、言い替えばBが起こる確率 p_B が x_1, x_2, \dots, x_n の関数となることである。

$$p_B = f(x_1, x_2, \dots, x_n) \dots \dots \dots (2)$$

この仮定のもとで、 p_B のオッズを計算すると、

$$O(B) = p_B / (1 - p_B) \dots \dots \dots (3)$$

となる。

ここで、 $O(B)$ の対数オッズを考える。前の段落で説明した様に $O(B)$ の対数オッズは全ての実数値を取り得る。また、(2) で定義した様に p_B は x_1, x_2, \dots, x_n の関数であるから、(3) 式で定義される $O(B)$ の対数も x_1, x_2, \dots, x_n の関数となる。

$$\log(O(B)) = g(x_1, x_2, \dots, x_n) \dots \dots \dots (4)$$

この関数 $g(x_1, x_2, \dots, x_n)$ にどのような関数を当てはめるかで様々なモデルを作ることができる。いま、最も単純に g に x_1, x_2, \dots, x_n の一次線形結合を仮定する。

$$g(x_1, x_2, \dots, x_n) = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \dots \dots \dots (5)$$

数学的には、 a, b_1, b_2, \dots, b_n を求める方法に興味があるのだが、ここではさしたる問題ではないので省略する。

結果としては、

$$\log(p_B / (1 - p_B)) = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \dots \dots \dots (6)$$

となる。(5) 式の左辺は前述した様に全ての実数値を取るの、 x_i の係数 b_i は、正にも負にも成り得る。また、(6) 式を書き直すと、

$$p_B / (1 - p_B) = \exp(a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n) \dots \dots \dots (7)$$

つまり、ある特性の対数オッズがその特性に影響を与える(原因と成り得る)要因によって説明されるモデルが作成されたことになる。これが、Cornfieldにより提唱された多重ロジスティック・モデル(Multiple Logistic Model)である。また、このモデルを見れば一目瞭然の様に原因となる各要因が変化するに伴い、結末的な特性Bの生起が変化することが判る。

ロジスティック・オッズ

ロジスティック・オッズとは、多重ロジスティック・モデルにより算出されたオッズ(比)のことである。以下では、多重ロジスティック・モデルを用いたオッズ比の算出法について解説する。

(1) で仮定したように、ある結末的な特性が起こる確率はその原因となる要因の関数となる。そこで、今一つの要因 x_1 についてのみ検討する。 x_1 が x_{11} と x_{12} のときの特性Bの発生確率をそれぞれ、 $p_B(x_{11}), p_B(x_{12})$ とする。このとき、(6) 式により特性Bのオッズを求めると各々、

$$\begin{aligned} & p_B(x_{11}) / (1 - p_B(x_{11})) \\ &= \exp(a + b_1 x_{11} + b_2 x_2 + \dots + b_n x_n) \\ & p_B(x_{12}) / (1 - p_B(x_{12})) \\ &= \exp(a + b_1 x_{12} + b_2 x_2 + \dots + b_n x_n) \end{aligned}$$

となる。ここで、 x_1 以外の要因を固定する。つまり、 x_2, \dots, x_n は、どの様な値でも構わないが、等しいとする。そうしておいて、今求めたオッズを用いてオッズ比を算出する。

$$\begin{aligned} OR_B(x_{12}/x_{11}) &= \frac{p_B(x_{12}) / (1 - p_B(x_{12}))}{p_B(x_{11}) / (1 - p_B(x_{11}))} \\ &= \frac{\exp(a + b_1 x_{12} + b_2 x_2 + \dots + b_n x_n)}{\exp(a + b_1 x_{11} + b_2 x_2 + \dots + b_n x_n)} \\ &= \exp\{(a + b_1 x_{12} + b_2 x_2 + \dots + b_n x_n) - (a + b_1 x_{11} + b_2 x_2 + \dots + b_n x_n)\} \\ &= \exp\{(a - a) + b_1(x_{11} - x_{12}) + b_2(x_2 - x_2) + \dots + b_n(x_n - x_n)\} \\ &= \exp\{b_1(x_{11} - x_{12})\} \dots \dots \dots (8) \end{aligned}$$

というように要因の値とその要因が持つ係数の値のみによって定まっている。

(8) 式の結果を用いると任意の状態の任意の状態に対するオッズ比を算出することが出来る。ところが、ここに一つの疑問点が存在している。(8)式を見れば、判るように問題となるのは、実際には状態間の差である。つまり差さえ同じであれば、同じオッズ比が得られることになる。但し、他の要因は固定してあるため、他の要因の影響は排除されている。これは、(5)式で g に一次線形結合を仮定したため、ここにロジスティック・オッズ比の特徴がある。

また、(4)式で係数決定の際に得られる、係数 b_1 の標準誤差を $SE(b_1)$ とすると、要因 x_1 の2状態 x_{11} 、 x_{12} 間のオッズ比の $1 - \alpha$ 信頼限界は、

$$\exp \{ [b_1 \pm z(\alpha) \times SE(b_1)] (x_{11} - x_{12}) \}$$
 となる。なお、 $z(\alpha)$ は、正規分布の α 点の値 (例えば $\alpha = 0.05$ ならば 1.96) である。

おわりに

看護学の領域では、調査研究が大半を占めている。調査研究では要因間の関係を明らかにすることを目的としているが、結果と見なされる事象が単一要因で成り立っていることは稀であり、実際には様々な要因が複雑に絡み合っている。そうした関係を明確にするためには、それにふさわしい分析手法が必要であるにも拘わらず、実際にはあまり行われていない。そのことがかえって本来、探索的手法であるはずの質的研究を助長し、質的研究による Decision Making をさせていると思われる。計算機の能力が向上するに伴い、従来までならば大型計算機でしか不可能であった分析がパーソナル・コンピュータでも可能になり、身近になって来たのであるから、もっと様々な分析方法が適用されるべきである。

参考文献

- 1) Cornfield, J., Gordon, T. and Smith, W.W. : Quantal response curves for experimentally uncontrolled variables. Bull. Inst. Stat., 38 : 97-115, 1961.
- 2) Cornfield, J. : Dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure : a discriminant function analysis. Fedn. Proc. 21 (4) Pt. 11 : 58, 1962.
- 3) Fleiss, J.L. : Statistical methods for rates and proportions., 1973.
- 4) 廣田安夫 : 心筋梗塞. 医学のあゆみ, 110 : 889-895, 1979.
- 5) Gart, J.J. & Zweifel, J.R. : On the bias of various esti-

- mators of the logit and its variance, with application to quantal bioassay. Biometrika, 54 : 181-187, 1967.
- 6) 小林幸子, 他 : 中学生の愁訴出現に關する食生活因子について. 小児保健研究, 49 : 573-579, 1990.
- 7) 小町善男, 辻岡克彦, 飯田稔, 他 : 脳卒中. 医学のあゆみ, 110 : 879-888, 1979.
- 8) McCullagh, P. & Nelder, J.A. : Generalized Linear Models, 2nd Ed., 1989.
- 9) Smith, W.C.S., Tunstall-Pedoe, H., Crombie, I.K. & Tavendale, R. : Concomitants of excess coronary deaths : major risk factor and lifestyle findings from 10,359 men and women in the Scottish Heart Health Study. Scotland Med. J., 34 : 550-555, 1989.
- 10) 高木廣文 : 糖尿病長期治療患者に關する疫学研究 (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・ファクターに關する研究, 日本公衆衛生雑誌, 26 (4) : 185-194, 1979.
- 11) Tango, T. : Statistical model of changes in repeated multivariate measurements associated with the development of disease. Com. stat. Ana., 3 : 77-88, 1985.
- 12) Truett, J., Cornfield, J., & Kannel, W., : A multivariate analysis of risk of coronary heart disease in Framingham. J. Chron. Dis., 20 : 511-524, 1967.
- 13) Walker, S.H. & Duncan, D.B. : Estimation of the probability of an event as a function of several independent variables. Biometrika, 54 : 167-179, 1967.
- 14) Yule, G.U. : On the association of attributes in statistics. Philos. Trans. roy. Soc. Ser. A, 194 : 257, 1900.
- 15) Yule, G.U. : On the methods of measuring the association between two attributes. J. roy. statist. Soc., 75 : 579-642, 1912.