

選択モデルの比較に関する一考察

鳥取大学工学部

正会員 奥山 育英

セントラルコンサルタント

正会員 ○細谷 涼子

鳥取大学大学院

学生員 高梨 誠

1. はじめに

1970年代より交通計画の交通機関選択を中心としてロジットモデルやロジットモデルを改良したモデルが多く利用されるようになった。しかしロジットモデルはプロビットモデルに代わるモデルとして開発された。このようなことをふまえて本研究では、選択モデルの中から同一のデータフォームで解析方法のみの異なるプロビットモデル、ロジットモデルおよび判別分析について精度や特徴の考察を行う。

2. 本研究で扱う選択モデル

本研究では前項に挙げた三つのモデルの選択可能な選択肢が二つのみの場合を扱う。二つの選択肢AとBがあるとき、選択肢Aに対して効用 V_i を持つ i 番目のサンプルが選択肢Aを選択する確率 P_{Ai} はプロビットモデルでは正規分布、ロジットモデルではワイブル分布を用いて、

$$[\text{プロビット}] \quad P_{Ai} = \frac{1}{\sqrt{2\pi}} \int_{-V_i}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz$$

$$[\text{ロジット}] \quad P_{Ai} = \frac{1}{1 + \exp(-V_i)}$$

と表わされる。ここで、効用 V_i は特性変数 x_{ki} ($k=1, 2, \dots, K$) に対して線形であり、

$$V_i = \sum_{k=1}^K \theta_k \cdot x_{ki} \quad (\theta_k: \text{未知のパラメータ})$$

とする。但し $\theta_1 \cdot x_{1i}$ は定数項にあたり、常に $x_{1i} = 1$ の選択肢固有ダミー変数である。プロビットモデル分析、ロジットモデル分析では実際の選択結果と選択確率を用いて最尤法により効用関数 $V(\theta)$ を推定する。

また判別分析は実際に選択肢Aを選択したグループと選択肢Bを選択したグループの効用 V_i ができるだけ分かれるように、つまり相関比が最大となるように効用関数を推定する。

3. 比較の方法

島中の研究にあるデータ¹⁾ (データA)とDaganzoの研究にあるデータ²⁾ (データB)の二種類のデータを用いて以下の手順で比較を行う。

- (1) 三つの分析方法により効用関数 $V(\theta)$ を推定し、プロビット、ロジットに関しては対数尤度も求める。但し、データBは選択肢固有ダミー変数以外の特性変数が一つのため、判別分析は行えないので、相関比のみを求める。
- (2) 相関比 P^2 を求める。相関比は選択結果ごとの効用 V_i の分離度を示す指標であり、1に近いほどグループ間の分離が明確であることを示す。
- (3) しきい値を用いた的中率 R_s を求める。これは効用のうちある値をしきい値として各サンプルの選択結果を推定する。この推定と実際の選択結果との一致具合を示す変数を S_i とし、

$$S_i = \begin{cases} 1 & (\text{推定と結果が一致}) \\ 0 & (\text{推定と結果が不一致}) \end{cases}$$

とすると、 R_s は

$$R_s = \frac{1}{N} \sum_{i=1}^N S_i$$

と定義する。

- (4) 確率を用いた的中率 R_R を求める。これはまず同じような効用をとるサンプルを一つのグループとして、全サンプルを M 個のグループに分ける。つぎに各グループの選択肢Aを選択するサンプル数を推定する(T_{Am})。実際に選択肢Aを選択したサンプル数を N_{Am} とすると、 R_R は次式で表わされる。

$$R_R = \frac{N - H}{N}$$

$$H = \sum_{m=1}^M |N_{Am} - T_{Am}|$$

4. 計算結果と考察

計算結果を以下に示す(表-1, 表-2)。尚、データBについて相関比は、0.4661であった。

表-1 データAの分析結果

		プロビット	ロジット	判別
パラメータ 推定値	θ_1	0.4055 (28.0)	0.7363 (27.0)	—
	θ_2	1.350 (93.1)	2.561 (93.8)	-8.731 (17.4)
	θ_3	-0.0145 (-1)	-0.0273 (-1)	0.501 (-1)
	θ_4	0.0994 (6.9)	0.1940 (7.1)	-3.218 (6.4)
対数尤度 L		-56.870	-55.860	—
相関比 P^2		0.2284	0.2286	0.3238
的中率	R_s	0.9488	0.9458	0.9337
	R_R	0.9508	0.9529	—

※()内の数値は $\theta_3 = -1$ としたときの
パラメータ間の比

表-2 データBの分析結果

		プロビット	ロジット
パラメータ 推定値	θ_1	-2.282 (-6.3)	-4.110 (-6.2)
	θ_2	0.3626 (1)	0.664 (1)
対数尤度 L		-103.213	-103.629
相関比 P^2		0.3220	0.3220
的中率	R_s	0.8991	0.8991
	R_R	0.9518	0.9427

※()内の数値は $\theta_2 = 1$ としたときの
パラメータ間の比

これらの結果をみると、パラメータ推定値そのものはモデルによりかなり異なるが、それぞれのモデルにおいてパラメータ間の比を計算すると、プロビットとロジットでは同じような比になっている。また、対数尤度、相関比、的中率というモデルの精度を表す指標もモデル間でほとんど差がみられ

ない。

これは、プロビットモデルとロジットモデルの違いが効用関数の誤差項の分布形の違いのみでその分布形も類似しており、なおかつ分析方法も変わらないためと考えられる。

ところで判別分析は、解析方法が他の二モデルと全く異なるにもかかわらず他の二モデルと近似的中率を示していたことから、効用関数の誤差項を考慮しない判別分析も場合により有効であるといえる。

5. おわりに

二種のデータを用いてプロビットモデルとロジットモデルを比較した結果、精度にほとんど差がないことが分かった。従って、両者を比べた場合二項選択モデルではどちらも容易に扱えるが、多項選択モデルになるとプロビットモデルの扱いが複雑になることを考えるならば、一般にプロビットよりロジットを用いるのが妥当であろう。

また確率を用いた的中率 R_R はしきい値を用いた的中率 R_s より向上するため、特に同じ効用値をとるサンプルがある程度以上あるデータではモデルの精度をみる指標として、あるいは現象予測にこのような方法を利用するとより現実 に即した判断や予測ができると考えられる。

このことは判別分析についても言え、本来判別分析における的中率の計算や判別分析による予測では確率は考えないが、サンプル数の多いデータでは各選択肢ごとの効用の分布より効用に対する選択確率を求め、プロビットモデルやロジットモデルのように的中率、予測に適用し得るであろう。

参考文献

- 1) 畠中康行^等：高速道路流入部におけるギャップアクセプタンス挙動の解析と加速車線長の決定法，第42回土木学会中国四国支部研究発表会講演概要集，pp.432-433，1990
- 2) Carlos F.Daganzo：Estimation of Gap Acceptance Parameters within and across the Population from Direct Roadside Observation，Transportation Research，vol.15B，pp.1-15，1981