

運輸省港湾技術研究所 正会員 ○ 白石 悟
 北海道大学 正会員 佐藤 馨一
 北海道大学 正会員 山形 耕一

1. はじめに

重回帰モデルは、発生交通量回帰モデルなど、極めて多くの分野で用いられているが、どのような変数をモデルに取り入れるかという変数選択には問題が残されている。この問題に対して、残差平方和が小さくなるように変数選択を逐次加減して行なう方法が一般的である。これに対して、D. M. Allenは、重回帰モデルの変数選択法として予測平方和基準⁽¹⁾(PSS基準)を提案した。本研究は、このPSS基準を、発生交通量回帰モデルに適用し、PSS基準の有効性に関して、考察を加えたものである。なお、本研究に用いたデータは、すべて道央都市圏パーソントリップ調査(昭和47年)によって得られたものである。

2. 残差平方和と予測平方和

重回帰モデルは、 $y_i = X_i(\beta + \varepsilon) \dots\dots (1)$ で定式化される。ここで、 y_i :被説明変量、 X_i :説明変量、 ε :誤差項、 β :偏回帰係数を表す未知パラメータ。(1)式の β を決めるには、正規方程式を解けば良い。ところで回帰モデルの作成にあたっては、 X_i の内容は固定していないのが通常である。すなわち、入手可能な変数群から説明上必要最小限の変数群 $(x_1, x_2, \dots, x_p)$ を選択してなければならない。そのための方法としては、次のような考え方があつた。

(I) RSS基準 (2)式で、定義される残差平方和(RSS)の増減量に対して、(3)式によるF検定を行なうことによって変数の除去あるいは追加を行なう。

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots\dots (2) \quad \hat{y}_i: \text{実際値} \quad \hat{y}_i: \text{推定値}$$

$$F = (\Delta RSS / RSS) \times (N - k - 1) \dots\dots (3) \quad N: \text{データ組数} \quad k: \text{採用変数の数}$$

この方法では、①変数選択の打ち切り基準(危険率 α)の取り方の恣意性、②慣用的に用いられている $\alpha = 5\%$ 基準では、一般に変数を取り込みすぎる傾向にあることが指摘されている。

(II) PSS基準 予測平方和(PSS)は、(4)、(5)式によって定義される。

$$PSS = \sum_{i=1}^n (y_i - \hat{y}_i^*)^2 \dots\dots (4)$$

$$\hat{y}_i^* = b_{0x} + b_{1x}x_{i1} + b_{2x}x_{i2} + \dots\dots + b_{px}x_{ip} \dots\dots (5) \quad \hat{y}_i^*: \text{推定値}$$

係数 b_{ix} は第 i 番目の観測値 y_i を除いた残りの $(n-1)$ 組の観測データから求める偏回帰係数である。PSS基準では、式(6)あるいは、図1に示すようにPSSが極値を持つので、その時の変数を用いた回帰モデルを、決定回帰モデルとする。またPSSとRSSの間には、

$$E[PSS] \geq E[RSS] + 2(n-k-1)\sigma^2 \dots\dots (6)$$

σ^2 : 残差分散 なる関係がある。⁽²⁾

3. PSS基準の発生交通量回帰モデルへの適用

PSS基準を発生交通量回帰モデルに適用し、変数選択基準としてのPSS基準の有効性を検討した。PSS基準による方法においても、変数選択はPSSが小さくなる方向に逐次的に行なわなければならない。本研究では、①PSS変数増減法②PSS変数減増法③R-P法(変数選択はRSSによって行ない、打ち切り基準としてPSSの値を採用する方法)を検討の対象とした。

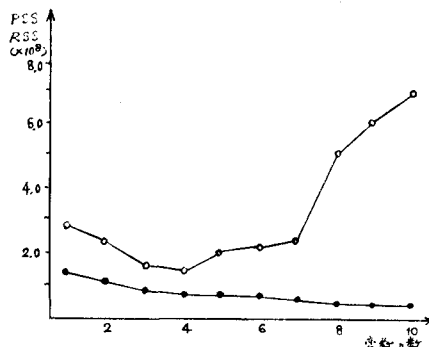


図1. PSSとRSS (観測目的発生交通量)

同時にRSS基準法として④ $F_{0.05}=2.5$, $F_{0.05}=2.0$ 法⑤ $\alpha=5\%$ の方法を行ない比較の対象とした。

4. 結果および考察

(1) PSSの小さな重回帰式の方が、偏回帰係数の安定性がある。(特定のゾーンを除くことによる偏回帰係数のばらつきが小さい。) (図2および表1は、RSSがほぼ等しく、PSSが異なるケースの例である。PSSの値は、特定のゾーンに大きく支配されていることがうかがえる。)

(2) いずれの交通目的においても、PSS基準の方が、一般に変数の少ない回帰式を得た。(表2)

(3) 表3は、全目的発生交通量の計算例である。この例では、PSS変数増減法では、十分にPSSの最小化が達成されていない。PSSの値からは、PSS変数減増法がすぐれているが、モデルの簡明性からはR-P法がすぐれている。PSSとRSSの値の順位相関は0.8~0.9ぐらいであり、PSS基準法とRSS基準法では、選択する変数が異なり、PSS最適の時に、RSSは必ずしも最適ではない。

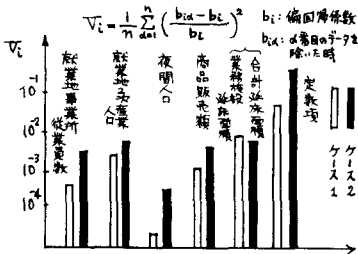


図2. 偏回帰係数の安定性

表1. 計算値比較例 (全目的発生)

	ケース 1	ケース 2
RSS	0.1852×10 ¹⁰	0.1853×10 ¹⁰
寄与率	99.06%	99.06%
PSS	0.5373×10 ¹⁰	0.9619×10 ¹⁰
予測率	97.28%	95.14%
ゾーン のPSSに 対する寄与率	111 112 712	56.7% 2.2% 10.7%
		73.7% 4.0% 2.5%

表2. 各方法による採用変数の数 (発生交通量)

	PSS基準					RSS基準				
	①	②	③	④	⑤	①	②	③	④	⑤
全目的	3	5	2		6					
通勤	4	8	4		7					
業務	7	7	2	6	7					
通学	4	4	4	6	4					
娯楽	4	4	2		3					
買物	1	3	1	6	3					
帰宅	1	4	1		6					
その他	1	6	1	6	5					

予測率 = 1 - PSS/総変動

表3. 各方法による発生交通量重回帰式 (全目的発生交通量)

	RSS(×10 ¹⁰) 寄与率	PSS(×10 ¹⁰) 予測率
① PSS変数増減法	0.5186 97.38	1.7236 91.28
② PSS変数減増法	0.1852 99.06	0.5373 97.28
③ R-P法	0.3597 98.18	0.8309 95.80
④⑤ RSS基準法	0.1642 99.17	1.0662 94.61

決定回帰式

$$① \hat{y} = 1.6793x_1 + 6.9104x_7 - 0.0652x_9 - 3400$$

$$② \hat{y} = 1.5851x_1 - 1.7079x_2 + 1.6774x_3 + 0.1529x_4 + 5.7610x_6 + 585$$

$$③ \hat{y} = 0.4037x_4 + 1.9006x_3 + 1725$$

$$④⑤ \hat{y} = 1.7766x_7 + 0.1446x_4 + 1.3755x_3 + 1.3371x_1 - 1.8050x_2 + 5.4494x_6 - 1079$$

x_1 : 就業地事業所従業員数 x_2 : 就業地2次産業従業員数
 x_3 : 夜間人口 x_4 : 商品販売額 (100万円) x_6 : 商業施設延床面積 (100m²)
 x_7 : 合計延床面積 (100m²)

5. 結論と今後の課題

(1) PSS基準による変数選択では、安定した回帰モデルが得られる傾向にあるため、質的あるいは量的変数の少ない短期の予測に、より適した方法である。ただし、妥当な変数を選択するかどうかに対しては、検討の余地がある。また、逐次選択法を三種の方法で行なったが、採用変数および採用変数の数に差異が生じ、いずれも一長一短があり、今後の検討を要する。

(2) RSS基準、PSS基準いずれの方法においても、現在の一時点のデータに対する回帰式では、予測対象の量的および質的変数に対して、大きな影響を受ける。少なくとも二時点以上のデータを用い、時間要素を加味した分析を行なう必要がある。

(3) 発生交通量を予測するには、用いる変数自体も予測しなければならないという二重構造的な問題がある。したがって、変数の予測における誤差および予測困難度を考慮した分析も必要である。

6. 参考文献

[1] David M. Allen 「The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction」 Technometrics Vol.16 No.1 (1974.2)

[2] 奥野 忠一 「統計変数解析法」 日科技連 (1976.3)

本研究は、北海道大学在学中に行なったものである。研究を進めるにあたっては、五十歳教授に御助言をいただいた。末筆ながら謝意を表します。