Ⅳ－3　交通事故件数の将来予測方法の比較

東京大学（生研）　正員　越　正毅
東京大学（生研）　正員　鹿島　茂
東京大学 大学院　学生員　○森　浩

1. はじめに

重回帰式を用いた将来値の推定法には、

(A) 各年度ごとの重回帰式を作り、それらの式から予測したい年度の重回帰式を推定し、説明変数の推定値を入れることによって被説明変数値を推定する方法（すなわち年度を層別因子として扱う方法）、

(B) 年度も説明変数として扱い、全年度のデータを使って一つの重回帰式を作り推定する方法、

の2つの方法が考えられる。(A)の方法は、各年度ごとに被説明変数を説明する構造が変化しているのではないかと考え、各年度ごとに偏回帰係数値を変化させる。(B)の方法は、被説明変数の推移は年度という説明変数の推移によって説明されると考えている。本研究は歩行者事故件数を例にとり、この2つの方法の比較検討を行った。また、(A)については偏回帰係数を直接推定する方法と、被説明変数および説明変数の共分散を用いて偏回帰係数を推定する方法の2通りの方法を検討した。

2. 重回帰式の予測法について

t年度における重回帰式を $\hat{y}_t = b_{t,0} + \sum_{i=1}^{p} b_{t,i} x_{t,i}$ とする。ここで、$t = 1, \cdots, m_0$ についてはこの式は既知であり、$b_{m,i}$ ($m > m_0$, $i = 0, \cdots, p$) を推定する。各 $b_{m,i}$ を推定するための方法として次の2つが考えられる。

(1) 過去の $b_{t,i}$ の変動状況から $b_{m,i}$ を推定する。それぞれの説明変数に対応する偏回帰係数ごとに時系列のグラフを描き、外挿する。

(2) 過去の各説明変数と被説明変数との偏差積和（あるいは共分散）の変動状況から $b_{m,i}$ を推定する。

(1)の方法は、各 $b_{m,i}$ が独立であるかのように扱い、(2)の方法は、共分散が各説明変数ごとに独立であるかのように扱うことになる。

$S_{m,ij}$ を m年における第i変数と第j変数の偏差積和、$S_{m,iy}$ を第i変数と被説明の偏差積和とすると $b_{m,i}$ は次のように書ける。

$$b_{m,i} = \frac{\det S_{m,i}}{\det S_m}, \quad \frac{d}{dt} b_{t,i}\Big|_{t=m} = b_{m,i}\, \mathrm{tr}\left(S_{m,i}^{-1} \frac{d}{dt} S_{t,i}^{-1}\Big|_{t=m} - S_m^{-1} \frac{d}{dt} S_t^{-1}\Big|_{t=m}\right)$$

ここで、
$$\begin{bmatrix} S_{m,11} & \cdots & S_{m,1p} \\ \vdots & & \vdots \\ S_{m,p1} & & S_{m,pp} \end{bmatrix} \qquad S_{m,i} = \begin{bmatrix} S_{m,11} & \cdots & S_{m,1y} & \cdots & S_{m,1p} \\ \vdots & & \vdots & & \vdots \\ S_{m,p1} & & S_{m,py} & & S_{m,pp} \end{bmatrix}$$

このように $b_{m,i}$ は $S_{m,ij}$ や $S_{m,iy}$ と複雑に係わりあっている。$b_{t,i}$ のグラフを単純に外挿する方法に対し、(1)との比較のために考案したのが(2)の方法である。本研究では、説明変数の将来値は既知であり、各 $S_{m,ij}$ も既知であるとしている。推定法としては、$S_{m,ij}$ を推定し上の式に代入して $b_{m,i}$ を求める。既知である $S_{m,ij}$ を用いているので、単に $b_{t,i}$ を外挿するよりも良いのではないかと考えられた。ただし、この方法は $b_{m,0} = \bar{y}_m - \sum_{i=1}^{p} b_{m,i} \bar{x}_{m,i}$ を求めるのに、推定すべき値の平均値を利用しなければならないという大きな欠点がある。しかし、各要素ごとの変動はバラバラであっても、平均の変化の様子が安定していれば比較的問題はないと考えられる。本研究で分析対象とした y_t は安定しているので(2)の方法も可能であると判断した。

3. 歩行者事故件数

従来の研究[1)]およびデータ入手の容易さ[2)]から説明変数および被説明変数を次のようにした。

y = 歩行者事故件数／道路面積［件／km^2］

x_1 = $\sqrt{\text{当量台キロ} \times \text{人口}}$／道路面積［｛（千台 km／12時間）・千人｝$^{\frac{1}{2}}$／$km^2$］

当量台キロ：走行台キロ×（1＋0.6×新規免許取得者／全免許所有者）

x_2 =（未改良道路延長／道路延長）×（DID人口／人口）

x_3 = $\sqrt{\text{防護柵延長} \times \text{道路照明基数}}$／幅員5.5m以上の道路延長［（km・基）$^{\frac{1}{2}}$／km］

データは、昭和46年から昭和50年までの沖縄を除く46都道府県についてのものである。重回帰式の偏回帰係数を表1に示す。

この式から53年における重回帰式を推定した。A－1の方法での$b_{53,1}$の推定方法を図1に示す。他の偏回帰係数も同様に行った。A－2の方法では$S_{53,iy}$をA－1の$b_{53,i}$と同じように求め、$S_{53,ij}$は53年の実際の値を用いた。

表1. 偏回帰係数の推移

		$b_{t,1}$	$b_{t,2}$	$b_{t,3}$	tの項	$b_{t,0}$	重相関係数
A	46	0.954	43.9	−6.47		−20.8	0.968
	47	0.888	41.0	−6.21		−18.3	0.968
	48	0.739	37.7	−4.83		−15.7	0.960
	49	0.628	37.6	−2.41		−16.5	0.958
	50	0.631	26.0	−2.35		−15.3	0.964
B	通年	0.790	36.0	−4.80	−4.26	186.6	0.956

昭和53年の推定重回帰式

A－1　$\hat{y} = 0.43x_1 + 20.2x_2 + 1.5x_3 - 12.0$

A－2　$\hat{y} = 0.65x_1 - 9.5x_2 - 4.9x_3 - 9.4$

B　　$\hat{y} = 0.79x_1 + 36.0x_2 - 4.8x_3 - 4.26 \times 53 + 186.6$

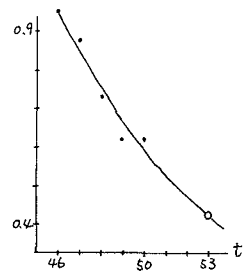

図1. $b_{t,1}$の推移

4. 結果

上の式に昭和53年の説明変数値（既知）を入れ、各県の推定値を出した結果を表2に示す。重回帰式を用いない分析法との比較として、単に事故件数のみを各県ごとに多項式で外挿した値も示す。ここでC－1,2,3はそれぞれ48～50、47～50、46～50年の値を2次、3次、4次多項式で外挿した値である。表を見てわかるように最も良いのはA－1であった。外挿による方法は53年のデータを用いてないので多少の真値とのずれは当然であるが、この場合はかなり大きく事故件数の予測には使えないと考えられる。53年の真値を使って偏回帰係数を求めたA－2よりも、真値を使わないで求めたA－1の方が良い結果になったのは予想外であった。実際の推定では$S_{m,ij}$を求めるために使う説明変数も、推定値になるのでこの場合よりさらに悪くなる。逆にA－1の方法はかなりすぐれた方法と言えよう。最も簡単なBの方法は、この場合には当てはまらない。46～50年のデータで重相関係数が0.956とかなり良く、53年のデータでも0.858となり一見良さそうだが53年のデータで46県中11県で事故件数が負になってしまうなどの欠点が見られた。

表2. 昭和53年の推定値

手法	件数(各県の計)	真値との相関
A－1	92080	0.956
A－2	84608	0.941
B	50429	0.858
C－1	163833	0.897
C－2	273053	0.628
C－3	199673	−0.001
真値	93459	

5. まとめ

将来予測を行うには年度を説明変数とする方法よりも、層別因子として扱う方法の方が優れていることが明らかとなった。特に説明構造が変化している場合には年度を説明変数として用いた式で推定することは問題が大きい。また、偏回帰係数の推定法としては、それぞれの係数を単純に外挿するという簡単な方法で十分実用的であると考えられる。

参考文献　1)昭和54年度交通事故発生状況の長期予測に関する調査研究報告書（内閣総理大臣官房交通安全対策室）第1分冊　2)同　第2分冊：交通安全対策基礎資料