

福山コンサルタント 正員 三宅秀隆

福山コンサルタント 正員 岸 敏司

1. はじめに

車の発生集中量等を線形モデル式で表現することが多い。今回、福岡市の発生集中モデルを作成するにあたり、見落しやちな問題点に遭遇したので、述べたい。現実には発生モデルを作成する際の作業としては、まず従属変数(y)と説明変数(x_i)との単相関係数をチェックし、相関度の高い変数を選択した後、最小自乗法で線形モデルを求め、重相関係数(R)あるいは F 検定でモデルの適合性を判定するという手順を使用している。ところが、重相関係数が高いのに説明変数に負の偏回帰係数(b_i)を生じ説明に困ることがしばしばある。そこで、線形モデルの構造内容をふりかえりつつ問題点を掲げてみたい。

2. 構造

重相関式の構造より気づくことは次の様なことがある。

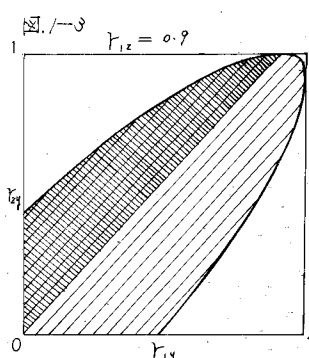
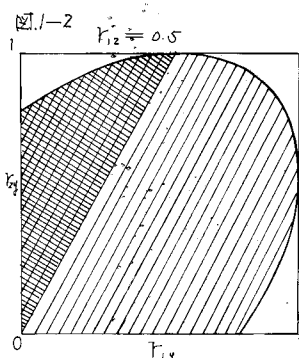
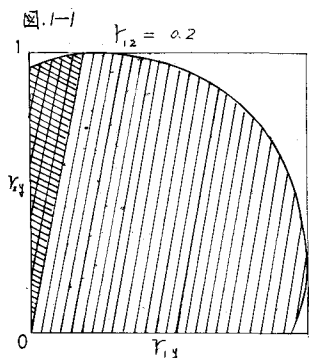
(i) 説明変数(x_i)を増せば、それが有効、無効にかかわらず重相関係数(R)は増大する。

(ii) 偏回帰係数(b_i)は従属変数(y)に与える影響を表わすものではない。

(i)についていえることは、重相関係数のみに着目していると問題の起こることが多いということであり、(ii)については、ややもすれば誤解しやすいことである。特に x 変数(x_1, x_2)の重回帰式をとりあげれば、

$$b'_i = \frac{1}{1 - r_{12}^2} (r_{iy} - r_{12} r_{2y}) \quad (i=1, 2)$$

ここに b'_i は標準偏回帰係数が成立する。そこで、 $r_{1y} > 0, r_{2y} > 0$ のもとに、 $b_1 < 0$ 、 r_{12} を変数としての範囲を求めれば次図が得られる。



これをみてわかることは、たとえ正の相関を有して($r_{1y}, r_{2y} > 0$)変数(x_1, x_2)を取りあげてもお互いの相関(r_{12})が強ければ強いほど、負の偏回帰係数(b_1 、 $b_2 < 0$)が構造上生じやす

く従って、偏回帰係数(b_1)によって従属変数に与える影響度を論じることには無意味(但し、 $b_2 \neq 0$ の時、 b_1 、 b_2 は、 y に与える影響を表わしたものであるといえる)といえる。すなわち、かりに説明変数間の相関度が低いものであることが望ましいといえる。

3. 発生集中モデル

そこで、以下のような構造上の問題点を考慮しながら、実際のモデル作成を行なってみる。取りあげたのは福岡市の乗用車発生集中量と人口関連指標であり、取りあげたデータ(福岡市内を19ゾーンに分割したものである。その結果単相関行列は次の通り求められた。

表. 1 要因相互間単相関

	居住人口	二次従業	商業	サービス業	その他	乗用車
居住人口	1.0000	0.8539	0.5609	0.7197	0.4816	0.5720
二次従業		1.0000	0.8250	0.9263	0.7571	0.8322
商業			1.0000	0.9649	0.9923	0.9918
サービス業				1.0000	0.9326	0.9747
その他					1.0000	0.9769
乗用車						1.0000

これによると、居住人口、二次従業人口と乗用車発生集中量の相関関係が低いのが目につく。そこでこの原因を探るため、これらの点をグラフ上にプロットしてみると図2-1~2の通り、1、2、3、4ゾーンの異なる傾向を示している。すなわち、1~4ゾーンのみで考えると明らかに負の相関関係があることがわかる。又、商業人口と発生集中量との相関も相関係数が高いが、グラフ上にプロットしてみると図2-3の通り1~4ゾーンと他のゾーンとは異なる構造を持っていると考えられ、これらを同一にして計算を行なうことが無意味といえる。

そこで、以上のようなデータの構造上の問題点を考慮し、取りあげたデータは都府部を除いたものとした。その結果単相関行列は、表2の通り求められた。

乗用車発生集中量と人口関連指標の相関図

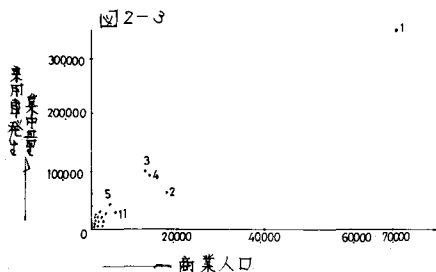
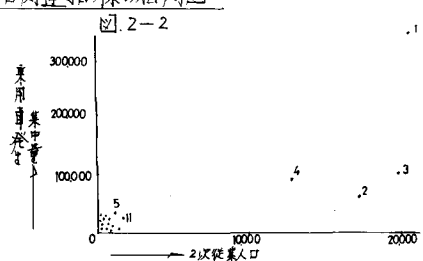
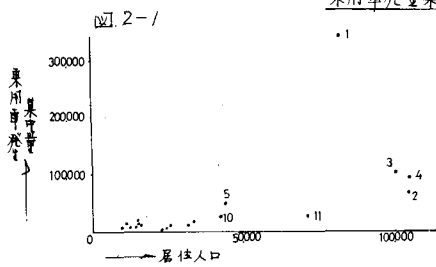


表. 2 要因相互間単相関

	居住人口	二次従業	商業	乗用車
居住人口	1.0000	0.6130	0.7352	0.8604
二次従業		1.0000	0.7418	0.9666
商業			1.0000	0.8338
乗用車				1.0000

ここで要因として居住人口、2次従業人口、商業を採用したモデル式を求めると、

$$y = 4889 + 0.203x_1 - 4.728x_2 + 7.186x_3$$

但し x_1 : 居住人口

x_2 : 2次従業人口

x_3 : 商業

重相関係数: 0.921

を得る。ここで求めた偏回帰係数の有意性の検定を行なう。この準備として、データより偏差平方和、積和行列 s_{ii} 、 s_{ij} の逆行列 s^{ii} を求め、回帰平方和 R 、残差分散 V_e を求めておく。そして検定には t 検定を用いて行なう。すなわち偏回帰係数 b_i が特定の値 b_{i0} をとるという帰無仮説 $H_0: b_i = b_{i0}$ を検定するに、

$$t = \frac{|b_i - b_{i0}|}{\sqrt{s^{ii} V_e}} \geq t_{(n-p-1; \alpha)}$$

α : 信頼度

が成り立てば、 H_0 を否定する。普通 $b_{i0} = 0$ を用いて b_i が0

か否かを検定する。その結果表、3に示す値となり b_1 は有意でなく、 b_2 は2%で、 b_3 は10%で有意となる。又、上記の計算で求めた指標の内2つを一組として求めた重相関係数の偏回帰係数の検定を行なった結果も示しておく。以上の考察により、居住人口、2次従業人口、商業人口の三指標より、居住人口が有意とならないで、

$$(\text{乗用車発生集中量}) = (\text{定数}) + b_1(\text{2次従業人口}) + b_2(\text{商業人口})$$

をモデル式として採用する。レベルはばうここで求めた。

$$y = 7239 - 5.268x_2 + 9.894x_3 \quad (\text{重相関係数 } 0.9145)$$

を用いて、他の地域、又将来の推計にも完全にあてはめることには、非常な危険が伴う。

なぜならば、ここで求めた偏回帰係数 b_i の95%信頼区間は、

$$b_2: -5.2683 \pm 3.5672$$

$$b_3: 9.8942 \pm 2.9996$$

であり、 x_2 はある値 x_2 に固定した時の集中発生量には図-3に示す如く非常な誤差を伴う。このため x_2 、 x_3 の変動域は各データの上限、下限内にととめるべきである。

4. おわりに

以上で求めた重回帰式は、福岡市について都市部を除くゾーン間でのモデル式であり、構造の異なる他の都市、又同じ福岡市においても将来の推計にこのまゝ用いるには疑問がある。以上問題点を述べたが、この解決については今後の課題であろう。

参考文献: 森田俊三「統計数理入門」／スミ 均、吉沢 正「多変量解析入門」

表3 偏回帰係数の有意性の検定

i	b_i	$t = b_i / \sqrt{s^{ii} V_e}$	判定
モデル式	$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$		
1	0.203	0.7972	—
2	-4.728	2.8827	5%
3	7.186	2.1765	5%
	$y = b_0 + b_1 x_1 + b_2 x_2$		
1	0.646	0.567	—
2	-2.925	1.512	20%
	$y = b_0 + b_1 x_1 + b_3 x_3$		
2	-5.268	3.218	1%
3	9.894	7.187	1%
	$y = b_0 + b_1 x_1 + b_3 x_3$		
1	0.908	1.576	20%
3	1.939	0.568	—

図3 発生集中量の予測限界

