

数量化理論Ⅱ類を用いたコンビニエンスストアの設置問題の検討

成蹊大学 *高嶋 成行 TAKASHIMA Masayuki

01001600

上田 徹 UEDA Tohru

1. 目的

企業が店舗を出店する際に、何を重要視して、どのような地域にたてているのかということを検討したい。そのために、昼間人口、夜間人口、公示地価、地域環境などのデータを説明変数とし、その地域のコンビニエンスの有無を目的変数として、数量化理論Ⅱ類を使い、次にどの地域に出店すればいいかを推測する。

2. 数量化理論Ⅱ類について

(1) 数量化理論Ⅱ類

判別分析において目的変数、説明変数に数量的な変数を与えるのではなく、名義的な尺度の変数を与えたものがある。ただし、本分析では、名義的な変数ばかりでなく数量的な変数も混在している場合を扱っている

(2) 問題の数式化

$$f = b_1 z_1 + b_2 z_2 + \dots + b_n z_n + c$$

ここで f は距離尺度の変数である。すなわち店舗が存在すればその地域は f の値は大きく、店舗がない地域ほど f の値は小さい。もちろんこのような f の値は、直接的には、観測できないので、対象地域の属性から合成される合成変数である。

(3) 本実験と数量化理論Ⅱ類について

本実験において0（店舗なし）と1（店舗あり）という目的変数と数値尺度、名義尺度の混在する9個の説明変数を与えて最適な係数 $b_1 \sim b_9$ を求め、その結果からセブンイレブンの設置場所を推測する。

3. 実験内容

(1) データ集計単位

市の中の町の丁目を1つのサンプル単位とする。

武蔵野市 - 51丁目
三鷹市 - 62丁目
田無市 - 39丁目 合計152丁目

(2) 説明変数

面積、人口密度、世帯数、昼間人口、昼間人口密度、夜間人口、昼間人口指数、公示地価
地域環境

(1:学校 2:その他のコンビニ 3:学校とその他のコンビニの重複 4:なし)

(3) 目的変数

セブンイレブンの有無を目的変数とする。有る、無いをカテゴリー変数として、有=1、無=0と表す。

(3) 実験内容

まず、データを推定用92丁目とデータ評価用60丁目の2群に分けた。推定用データを用いて、説明変数9個の中から、2~9個選んだ場合のすべての組み合わせの重回帰式を求める。推定用データを用いて、求めた的中率を判別的中率とする。判別的中率を求めた際に、同時に求まる各節名変数の係数を評価用データに用いて、評価用データに対しての的中率を求める。この的中率を正答率とする。以上の2つの観点から、もっとも望ましい説明変数の組み合わせを求める。

4. 結果と分析

(1) 正答率と判別的中率で見た組み合わせ人口密度、世帯数、昼間人口、昼間人口密度、夜間人口、昼間人口指数、地域環境の7つの説明変数であった。

(2) 説明変数の大事さの順位

下記の係数は、入力データを正規化しているため、標準回帰係数となっている。数値尺度の値は、係数*（最大値-最小値）。名義尺度の値は、（係数の最大値-係数の最小値）（レンジ）の値を用いる。

順位	係数	レンジ
1	世帯数	1.038
2	昼間人口	0.708
3	昼間人口密度	0.561
4	夜間人口	0.523
5	昼間人口指数	0.374
6	地域環境	0.373
7	人口密度	0.077

(3) 得られた重回帰式

$$Y = 0.077x_1 - 0.114x_2 + 0.097x_3 - 0.067x_4 + 0.065x_5$$

$$-0.052x_6 + 0.373x_7 + 0.104 \\ + 0.040x_{72} \\ + 0.208x_{73}$$

$$x_{7i} = \begin{cases} 1 & ; \text{当該地域の環境が } i \\ 0 & ; \text{その他} \end{cases}$$

となり $x_1 \sim x_7$ はそれぞれ、人口密度、世帯数、昼間人口、昼間人口密度、夜間人口、昼間人口指数、地域環境、公示地価である。

5. 問題点

今回用いた9つの説明変数は、相関関係の高い物が多いのでは??という問題点が考えられた。重回帰分析を用いた本研究において、説明変数の相関関係が高いと結果的に得られた答えの信用度が減ってくるので、まずは、説明変数の相関関係を調べてみた。そして、係数間の相関の高いものを取り除き、残った説明変数は、面積、昼間人口、人口密度、地域環境、公示地価の5つになった。この5つを用いてもう一度評価を行う。

6. 改善データを用いた計算

計算方法は、3.(3)と同様

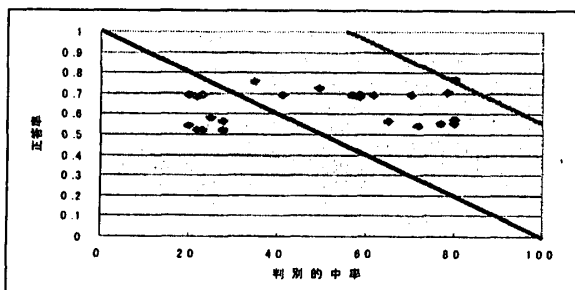
(1) 正答率で見た上位11モデル

順位	面積	人口密度	昼間人口	地域環境	公示地価	正答率
1	0	0	1	0	1	80
1	1	1	1	0	1	80
1	1	0	1	1	1	80
4	1	0	0	0	1	78.3
5	0	1	1	0	0	76.7
6	1	0	0	1	1	71.7
7	0	0	1	1	0	70
8	1	1	0	0	0	65
9	1	1	1	0	0	61.7
10	1	0	0	0	1	58.3
10	1	1	0	0	1	58.3
合計	8	5	6	3	7	

(2) 判別的中立で見た上位11モデル

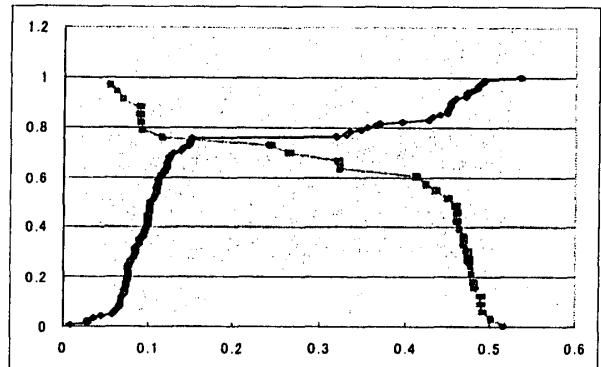
順位	面積	人口密度	昼間人口	地域環境	公示地価	判別の中率
1	1	0	1	1	1	0.768324
2	1	0	1	0	0	0.758242
3	0	1	0	1	0	0.725275
4	1	0	0	0	1	0.703297
5	0	1	0	0	1	0.692308
5	0	0	1	1	0	0.692308
5	0	0	0	1	1	0.692308
5	1	1	1	0	0	0.692308
5	1	1	0	1	0	0.692308
5	0	1	0	1	1	0.692308
5	1	1	1	1	0	0.692308
5	1	1	1	1	1	0.692308
合計	6	6	5	7	5	

(3) 二つの観点から見たグラフ評価



上記の図から正答率と判別の中率の大事さを5:5と見たときに最も望ましいとされた点は、判別の中率76.8%正答率80%であり、説明変数の組み合わせは、面積、昼間人口、地域環境、公示地価の組み合わせとなった。上記の説明変数の組み合わせをBEST組み合わせとする。

(4) 改善データ: BEST 組み合わせの計算結果



上図から右上がりのプロットがコンビニなし、右下がりのプロットがコンビニにありの地域を表している。縦軸が1に近いプロットほどコンビニエンスストアの存在に最もふさわしいという事を意味している。グラフから、コンビニエンスストアのない地域で最も設置すべきである地域は、吉祥寺本町4丁目が見えたと導かれた。

7. 改善データの計算結果と分析

(1) 説明変数の大事さの順位
方法は、4.(1)と同様

順位	係数	レンジ
1	地域環境	0.388
2	面積	0.169
3	公示地価	0.128
4	昼間人口	0.003

8. 最終的に得られた重回帰式と結論

$$Y = -0.0186x_1 + 0.0035x_2 + 0.3883x_{31} + 0.0247x_{32} + 0.2505x_{33} + 0.0134x_4 + 0.0895$$

$x_{3i} = 1$; 当該地域の環境が i
0; その他

となり $x_1 \sim x_4$ はそれぞれ、面積、昼間人口、地域環境、公示地価である。

9. まとめ

本検討では、購買力を示すデータが入っていないし、このほか出店に影響を及ぼすと思われる要因の検討も不十分であるが、その割には、そこそこの結果が得られた。