

# パネルの特徴抽出によるシェア予測

森村 英典, 上野 慶子, 吉永 絵里

## 1. はじめに

日本女子大学の家政学部にあった家政理学科数学専攻の最後の卒業生のうち森村研究室に配属された7名(上野慶子, 大場理子, 岡田正美, 門脇志乃ぶ, 小林真由子, 谷村磨未子, 吉永絵理)は, その卒業論文の主要テーマとして, スキャンパネルデータの解析を取り上げ, MS部会のデータ解析コンペに参加した。しかし, 学部4年生のマーケティング・サイエンスについての基礎的な知識を全く持たない集団であるので, 課題として与えられた, 適切なモデルを考えてシェアを予測するという事に重点を置くよりも, むしろ提供されたデータから, 何らかの特徴的な知見を引き出すことは出来ないかという方向で解析を進めることにした。提供されたデータは, フロッピー・ディスクの形で手渡されたので, コンピュータ上でデータ・ハンドリングを行う方法の習得から始めなければならなかった。解析にはSASを利用した。まず,

- (1) 価格掛け率の変動状況の追跡,
- (2) エンドとチラシその他購買行動に影響を及ぼすと考えられる要因の抽出,
- (3) パネル毎の特徴の発見,

の3つを中心テーマとする3グループに分かれて, 分析作業を行った。本稿では, それぞれのグループにおいて得られたいくつかの知見を報告した後, パネルの

もりむら ひでのり 日本女子大学理理学部数物科学科〒112 東京都文京区目白台 2-8-1

うえの けいこ (株) 帝国アータバンク

よしなが りえ 興銀システム開発(株)

層別に基づく判別分析を利用したシェア予測の結果について述べる。

実はこの作業を始める前に, 予測自体は, 提供されたデータの特徴から, 前年の結果をそのまま予測値とするというような極めて乱暴な「予測」が案外いい線を行っているのではないかと、という予想をしたことも予測モデルに重点を置かなかつた理由の1つであつた。そして, 結果的にはこの予想は正しかつたようであるが, それではいささか寂しいので, あえて層別した判別分析を利用してみたのである。結果については若干の考察を付け加える。

## 2. 価格掛け率について

各銘柄に対する価格掛け率は, 図1に示すように, 2~3日, 時としてはわずかに1日しかある値を持続しない場合もあるが, 通常は1ヶ月程度の期間, ある値のまま推移することが多い。また, ゴールドブレンドの150gと100gのように同種の銘柄間では, 一方が安い値段で売られている時期には, 他方は定価で販売されている, といったはっきりした関連を持ちながら価格掛け率が変動している様子も見取れる。

各銘柄についての価格掛け率の変動を調べるうちに, 「価格入れ換え日」とでも称すべき特定の日がほぼ月1回程度の割合で生じており, この日に数銘柄の価格掛け率が一斉に変えられていることに気がついたので, この日の間を「月」とみなして解析を進めることが適当である, との知見が得られた。

前述したような同種の商品の間の価格掛け率の関係も, 販売政策という面から見れば極めて当然の事態であるけれども, そのような「人工的な」あるいは「管理的な」要因を無視してデータの解析を進めることは, いささか危険を伴うことではあるまいか。このことも, データ解析のための「月」を暦の上の月とは

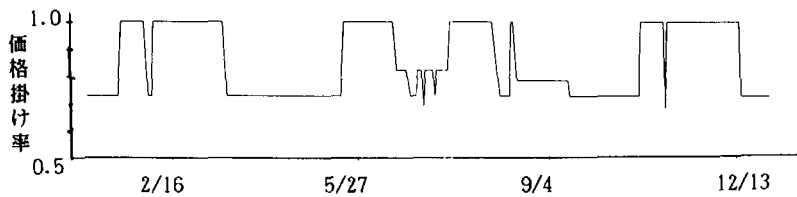


図 1: 価格掛け率の時間的変動 (銘柄 1)

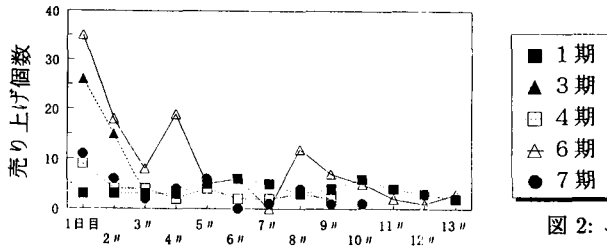


図 2: エンドに置いた日の売れ行き

別に定義することの妥当性を支持しているように思われる。

### 3. エンド等の要因

チラシが出されている日は、ある銘柄のいわば「激安日」に当たるので、当然、その銘柄の販売量は多くなっている。しかし、これを価格のためと見るか、チラシのためと見るか、その決め手を見つけられるほどのデータ量は無いので、確認は出来ていない。

ある銘柄がエンドに置かれているとき、その価格掛け率は低い時が多いので、たとえ売れ行きが好調でも、それをエンドだけの効果と見ることもやや困難である。そのような、数々の疑問は承知の上で、各要因の売り上げに及ぼす影響を分散分析にかけてみたところ、エンドを含めてほとんどの要因の影響が有意であるとの判定はなされなかった。わずかに、スプレードライと呼ばれる粒状のコーヒーが、水にも溶けるといふ宣伝の効果からか、夏場に売り上げが上がることだけが有意と判定された。

販売促進の手法として重視されているエンドが有意な効果を持たないという結果はいささか意外であった。そこで、連続してエンドに置かれている期間を「期間」と名づけて、各期間のうちの何日目にとどのくらい売れているかをグラフに書いてみると、図2に示すように、「期間」の間ずっとコンスタントに売れているわけではなく、むしろ、第1日目に購買が集中しているという特徴が見られた。分散分析は1日1日の売り上げとエンドの有無とをデータとして入力しているため、売り上げの伸びている日数が、さほどな

い日数に較べて少なく、従って効果はないと判断される結果を導いたものと考えられる。この事実をどのように解釈すべきか、いろいろな議論もあろうが、消費者行動としては注目してよい事柄の1つではなからうか。そして、販売促進担当者は、販促活動をしたとき、おそらくその第1日目に現場を見てその効果を確認しようとするであろうから、彼の目には、エンドの効果が現実以上に大きく写ってしまいがちになるのではなからうか。

その他、天候と気温のデータともつき合わせてみたが、半ば予想される通り、ほとんど購買量との間に関連は見られなかった。

### 4. パネル毎の特徴

「銘柄5」(2カップ),「銘柄6」(カプチーノ)及び「銘柄10」(30グラム)の3銘柄は、「特殊銘柄」と総称するが、これらを買う人は、その特殊な目的に合致して購買をするためブランドロイヤルティが極めて高く、価格の如何に拘わらず、その銘柄を購買するという特性を持つことが予想される。事実、銘柄5の常時購買者は1人、銘柄6のそれは5人という程度の少数ではあるが、これらの銘柄は極めて少量が1単位であるためもあって、当然購買数は多くなり、ごく少数の購買パネルが全購買数の大半を購入していることが判明した。

この人々を除外して、全パネルを購入数量によって5クラスに分類し、それぞれのクラスに入るパネルを個々に見た。そして、価格の安い時に購入しているか、比較的価格に無頓着に購入しているかで、概略の色分

けをしたところ、購入量の多い層ほどブランドロイヤルティは高い傾向があることが見られた。

更に、より詳細に各パネルの特徴を明確にするクラス分けとして、「フリーズ・ドライ派」、「スプレー・ドライ（パウダー）派」そして「特殊銘柄派」を核として、それらの中間層を適宜定義しながら分類することを試みた。図3にその分類のクラスを示すが、たとえばFというクラス(1)は、フリーズ・ドライの銘柄1, 2, 4, 8及び11以外の銘柄は全く買っていないパネルから成る層であり、F'というクラス(4)は、それらを50%以上買っている層を意味している。また、その中でも銘柄1と2だけに集中している「ゴールドブレンド」派は(2)と(3)として独立させている。かくして、全部で12のクラスに分類したが、200世帯くらいは「銘柄11」(その他)の存在がかなり邪魔をしていて、必ずしも明確な分類が出来たとは言いかねる。

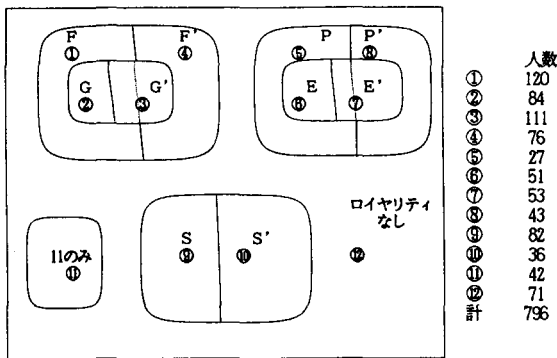


図3: パネルのクラス分け

いずれにしても、ある程度のブランドロイヤルティはほとんどすべてのパネルが持っており、いわゆる想起集合の中で実際の購買品を決めているらしい。それでは、想起集合の中での決め手は価格であろうか。買おうと思っている銘柄の中で、その日のお買い得品を購入するという消費者行動が一般的であると解釈してもよいのであろうか。現実のパネルを調べてみると、それぞれのクラスの銘柄の中で、一番低い価格掛け率の銘柄を購入しているパネルも実在する反面、全く価格掛け率には囚われない行動を示すパネルも実在している。

## 5. 判別分析と購買予測

上記の分類によるクラスの中の各パネルが、それぞれの購買時において、実際にどの銘柄を購入するかを

予測するために、一般的な手法として、判別関数の利用が考えられる。

そこで、たとえば1~9月のデータに基づいて判別関数の係数を定め、10~12月のデータについて判別を行ったときの誤答率と、それらの時期において、最も安い銘柄を購入すると見なした場合の誤答率とを比較して、どちらの方法がより正確に、購入銘柄を予測しているかを調べた。その結果の一例を表1に示すが、双方とも個々のデータに関していうと、かなりの誤答になっているものの、判別関数の方が成績が良い。これは、必ずしも、安い方を購入するという行動を消費者がしていないという事実を物語っている。

表1: グループGの購買予測の比較(93年度のデータ)

ブランド	実績値	判別関数予測	安値購買予測
1	73	101	36
2	105	155	57
3	1	0	0
4	22	0	0
5	0	0	0
6	9	0	0
7	1	0	0
8	4	0	0
9	2	0	0
11	62	23	186
誤答率		0.373	0.516

なお、判別関数を決定するに当たって入力データとして利用した変数はつぎの4種44変数である:(各銘柄ごとの)価格掛け率、エンドの有無、チラシの有無、過去の購買実績。そして、これらの判別は4節で述べた12のクラス毎に別個の関数を当てはめた結果であることを付け加えておく。

## 6. シェア予測

コンペのテーマは、インスタントコーヒーの各銘柄のシェア予測である。しかし、パネル全体は93年と94年とで変化がない。つまり、同一の消費者群がインスタントコーヒーという嗜好品を購入するという状況であるし、少なくともこれらのパネル達は、このスーパーでの買い物をメインにしているらしいということを考慮すると、最初に述べたように、シェア自体には年度による差はほとんど無いと予想できそうである。言い換えると、定常状態に達している、と見なせそうであろう。

そのように考えてしまえば、93年の実績シェアを94年の予測値とするという「方法」も1つのオルタナテ

イブであることは間違いない。そして、我々の採用した層別に基づく判別関数の方法を含め、他のグループの方法も、94年の変数値に93年のそれとの大きな差がない限り、シェア予測という見地からは50歩100歩であろう。

我々の採用した「予測」方法は、次のようである。5節で述べたように、すべてのパネルを、購買銘柄の特徴によって12のクラスに分類し、各クラス毎に判別関数を決定してある。94年度のデータには、各パネルの購買日を与えられているから、それらの購買機会毎に、どの銘柄を購入したかを、判別関数によって決定する。こうして、各パネルの購買銘柄が「予測」されることになる。これらの結果を集計し、全購買数で除してシェアを計算したのが、表2である。

表2: 94年1~7月のシェア予測

ブランド	実績値	判別関数予測	93年度シェア
1	14.9	14.8	14.2
2	11.5	14.8	11.8
3	10.0	10.6	11.0
4	9.2	6.3	10.3
5	0.2	0.5	0.2
6	8.6	9.7	8.3
7	4.3	4.5	5.3
8	2.5	4.4	3.6
9	5.4	5.1	3.1
10	1.4	4.3	2.6
11	31.9	24.8	29.6
不一致係数		0.116	0.051

なお、銘柄5はこのスーパーマーケットから撤去された商品であるため、これを買ったと「予測」される場合は、2月以降銘柄11に繰り入れて予測値とした。また、コンペの判定基準として採用されたタイルの不一致係数の値を、93年実績値に対するその値とともに示している。

## 7. 考察

最初、渡された94年のデータには、12月13日までのデータが含まれていた。そのため、この日までの購買機会について上述の「予測」を行った。そして、多分手違いからと思われるが、現実の購買銘柄も併せて記入されていたため、「予測」の結果を実績と比較することが出来たのである。その結果、銘柄11のシェアが極めて大きくて、かなりの乖離を示していた。

ところが、購買数量を時系列的に眺めると10月以降特に銘柄11が異常に伸びていることに気がついた。これは何かデータ自体に異変があると思われたので

問い合わせた結果、バーコードのミスで新発売の牛乳が紛れていたことが判明した。このため、7月までのデータで予測をし直すことになった。

今回取り扱ったデータはかなり大量のものであり、計算機なしにはとても取り扱えない。それだけに、データの不良に気がつく手段をいろいろと考えておかないといけない、ということを実感した。また、銘柄11というものの存在が解析の邪魔をしたこともしばしば感じたところである。もちろん、銘柄数が多すぎてデータの総量を絞りたいという気も一方ではあるが、「その他」という集団が30%を超える量を占めることは、1/3の知識を放棄するにも等しいことであるから、不便さには少々目をつぶっても、「その他」という括り方をするのは出来るだけ後にしたい。モデルによっては、そしてデータによっては、例えば銘柄1と2とは一緒でも構わなかったかもしれないが、それはやはり解析してみたらでないと何とも言えないことであろう。従って、データ解析は出来るだけ源泉で、という態度はいつでも鉄則の1つではなからうか。

また、前節でも触れたことではあるが、今回の対象が定常的なものであっただけに、結果的には「何もしない」という方法が最も「当たって」しまった。我々の採用した「予測」方法は、パネルの特徴を十分に活用したにも拘わらず、かえって誤差が大きかった。この事実は「足し算の相殺効果」を改めて強く印象づけるものであろう。もともと、5節でも述べたように、個々のパネルごとに見れば30%も誤って予測するのに、シェアとなると誤差はあまり目立たなくなっている。これも、それぞれの誤差が互いに相殺する結果に他ならない。同様のことが、ミクロの予測の積み上げとマクロの予測との間に言えるであろう。ここでの予測はミクロ・マクロというには相応しくはないが、分割しすぎる結果、かえって誤差を大きくしたという解釈は、さほどの外れでもないと思う。

今回の結果は、「何もしない」方法に軍配が上がったとしても、それがいつも良い筈はない。「予測」の決め手は「定常なもの発見」であるから、もともと定常な対象に対して有効であったというに過ぎない。今回、いくつものグループで試みられた様々な方法は、いずれも魅力ある方法である。それらの方法こそが有効な場はどのような場であるのか、今後のますますの研究が待たれるところである。