

有限混合多変量回帰モデルによる オークションデータの分析 —購入後の使用目的別クラスの探索—

久保 沙織, 豊田 秀樹

1. はじめに

近年、ネットオークションの普及によって、日本でもオークションは身近なものとなりつつある。オークションの場で取引される物品は、骨董品や自動車、不動産から権利に至るまで多岐にわたる。オークションは主として次の2つの形式に大別される。一つは、入札者がリアルタイムで相互に提示価格を知ることができる公開入札方式であり、もう一つは、自分以外の入札者の提示価格が公開されない封印入札方式である。

ここで、オークションでの購買行動と、一般的な購買行動との相違について考えてみよう。オークションにおいては、買い手が自ら値段を決定するという点が特徴的である。すなわち、オークションでの購買行動と、一般的な購買行動との大きな違いは、価格の決定権が前者では買い手にあるのに対して、後者では売り手にあるという点にある。通常の購買行動では、あらかじめ価格が提示されている商品の中から購入対象物を選択する。このとき、消費者が、ある価格をつけられた商品に対して、その対価を払って購入するということは、その商品にそれだけの価値があると判断するからである[4]。よって、その価格が満足するものでなければ、購買決定には至らない。その一方で、オークションでは、複数の商品の中から欲しいと思った商品に対して、買い手自らが値付けを行う。つまり、オークションには、本当に購入の意思のある者のみが参

加するのであり、個々の参加者の商品に対する価値評価が、値段という形で直接示されていると捉えることができる。

そこで、買い手ごとの値付けのメカニズムの違いに注目して、オークション参加者のセグメンテーションを行うことを本論文の目的とする¹。消費者ニーズの多様化に伴い、マーケティングの理論、実践両側面において、マーケット・セグメンテーションは中心的な概念となってきた[9]。

マーケティング・セグメンテーションのための基準として、「地理的変数」、「人口動態的変数」、「サイコグラフィック変数」、「行動上の変数」という4つの変数が考えられる[5]。中でも、行動によるセグメンテーションは、製品に対する知識、態度、使用形態、反応に基づいて買い手の分類を行うもので、より具体的には、機会によるセグメンテーション、ベネフィットによるセグメンテーションなどがある。前者は買い手が実際に購入する機会や購入した製品を使用する機会によって分類するというもので、企業が製品の使用方法を確立するときに役立つ。一方後者は、消費者が製品に求めるベネフィットによる分類であり、セグメントごとの明確な違いを発見するために高い効果が期待できるセグメントの方法である。このように顧客の特徴をとらえたセグメンテーションを行うことは、多種多様なニーズを持つ顧客に対する有効なアプローチのためにも、また、セグメントごとのより詳細な分析のための準備としても有用であると考えられる。

セグメンテーションにおいては、事前に与えられて

くぼ さおり

早稲田大学 大学院文学研究科

とよだ ひでき

早稲田大学 文学学術院

〒162-8644 新宿区戸山1-24-1

受付 08.7.15 採択 08.12.5

¹ 分析に使用したデータは業者間入札のデータであるため、厳密にはオークションの参加者がすなわち最終消費者ではないが、本論文では買い手としてのオークションの参加者が最終消費者であると仮定して分析を行った。

いる観測変数に基づいてクラス分けをすることももちろん可能である。しかしながら、直接的には観測できなくとも存在が仮定できる概念や背景となる因子がある場合、それらを潜在変数としてモデルに組み込むことで、実際に観測される複数個の変数間の複雑な関係を効果的に説明するモデル構築が可能となる。この種の統計モデルは潜在構造モデルとして知られており、近年の分析ソフトウェアの充実により、マーケティング分析や社会構造分析、医療分析など多くの分野で、その適用が広がっている[8]。渡辺[8]では、マーケティングの分野でよく用いられる潜在変数として、消費者の「嗜好」、「態度」、「購買意欲」を挙げている。潜在構造モデルは、分類基準が不明確な状態での判別分析と考えることができる[7]。そのため、顧客に関する事前情報が少ない状況でのデータ解析においても適用可能である。

本論文では、各オークション参加者が提示した価格という観測可能な変数の背後に、「購入後の使用目的」という潜在的な行動上の変数を仮定することにより、オークションデータに対して潜在構造モデルを適用した分析例を示す。次節以降では、まず分析対象となったデータについて概説し、モデルの理論的説明を行う。その後、分析結果の解釈を行い、最後に本分析の有用性について考察するとともに、今後の課題を述べてまとめとする。

2. 方法

2.1 分析対象

分析対象としたデータは平成19年度データ解析コンペティションで提供されたデータであり、リースアップ車を中心とした入札会において取引された、2005年6月22日～2007年6月12日の約2年間分の記録であった。オークションの形式は封印入札方式である。オークションデータは、主として、入札会の開催および入札の対象となった車に関する情報と、落札価格や札数などの落札情報からなり、どのような車がいくらで落札されたかがわかる。出品された車がオブザベーションとなっており、落札者のバイヤーコード以外は、買い手の属性に関する情報は全く含まれていなかった。

分析に先立ち、新たな変数を作成するとともに、データのクレンジングを行った。新たに作成した変数は以下の4つである。1つ目は、「落札金額」を「新車価格」で除し100倍した「対新車価格比」であり、2つ目は「(オークション)開催日」の開催月から「初

年度登録日」の月を引いた値を車の「使用月数」とした。3つ目として、それぞれ0(なし)と1(あり)のダミー変数で表されている「ヤブレ」「穴」「亀裂」の3種類のデータを足し上げて、0から3までの値をとる「損傷」というカテゴリカル変数にまとめた。4つ目は、連続変数として与えられた各車の「排気量」をカテゴリ化することで、車の形態を把握できるようにした。具体的には、660cc以下を0(軽自動車)、661cc以上1,500cc以下を1(小型車)、1,501cc以上2,000cc以下を2(中型車)、2,001cc以上2,500cc以下を3(ミドルクラス車)、2,501cc以上を4(大型車)という5段階とした。この5値のカテゴリカル変数を「排気量クラス」とする。

データクレンジングではまず、「落札金額」、「新車価格」、「走行距離²」のそれぞれに関して、欠損値または0を含むオブザベーションおよび、「距離区分³」をもとに、走行メーターを交換している車とその可能性のある車は削除した。また、買い手の値付けのメカニズムに注目するため、落札された車および流札となった車を分析対象として含め、無札であった車は除外した。左ハンドル、Wタイヤ、サンルーフ、冷凍冷蔵装置のいずれか一つでも装備されている車も、ごく少数しか存在しなかったことと、海外メーカーもしくは輸送用など用途の限定された車であるという理由から削除した。さらに、「形状」が全データの5%に満たないもの、および「車種」が同じく0.5%に満たないものを除いた。この時点までで残ったオブザベーションのうち、「対新車価格比」が上下5%に入る車は、オークションの一般的な相場に比べて不当に高いあるいは低い値付けをされているものと考え、外れ値と見なして切り捨て、残りの90%のデータを使用した。最終的に分析対象としたデータは1,556<「対新車価格比」<41.647であり、オブザベーション数は47,951であった。

最後に、「メーカー」および「形状」の内容を確認したところ、前者は「いすゞ」、「スズキ」、「ダイハツ」、「トヨタ」、「ニッサン」、「ホンダ」、「マツダ」、「ミツビシ」の8種類、後者は「BV(ボンネットバン)」

² 距離メーターの示す総走行距離

³ 距離メーターに関する4値のカテゴリカル変数。その内容は以下の通り。1:走行メーターに特段問題がないもの、2:タコグラフ、走行メーターが一体式のもの、3:走行メーターが交換されているもの、4:不明(走行メーター交換の可能性のあるもの等)

「BW (ボンネットワゴン)」, 「CT (キャブトラック)」, 「CV (キャブバン)」, 「HB (ハッチバック)」, 「SD (セダン)」の6種類であった。これらの名義変数をそれぞれ8値と6値のカテゴリカル変数に変換した。

2.2 有限混合多変量回帰モデル

本論文においては、「対新車価格比」を基準変数とした重回帰分析に潜在構造モデルを導入することで、購入後の車の使用目的という観点から落札者のセグメンテーションを行った。車のスペックやメーカーなどの違いによって、車ごとに妥当な値段の幅は異なるため、落札価格そのものではなく新車価格を基準として比較した「対新車価格比」を基準変数とした。

予測変数は、連続変数である「使用月数」、ダミー変数である「PW」(パワーウィンドウの有無), 「NAVI」(ナビゲーションシステムの有無), 「抹消区分」(一次抹消登録の有無), カテゴリカル変数である「損傷」の5つである。ただし、分析においては「損傷」という変数を連続変数として扱ったことに注意されたい。オークションにおける入札金額には、出品された車に対する落札者の価値評価が反映されているという前提のもとに、価格決定に際して重視する要因の差異によって落札者のセグメンテーションが可能になるものと予測される。そのため、本オークションにおける取引の対象がリースアップ車であることを考慮して、中古車選びの指標となるような変数を予測変数とした。また、これらの5変数間の相関は、ほとんどないことがあらかじめ確認された。

潜在構造モデルは、その潜在的なクラスをモデル化するために有限混合分布を仮定するため、有限混合モデルとも呼ばれる。そこで、重回帰分析と潜在構造モデルの手法とを組み合わせた本モデルを、有限混合多変量回帰モデルと定義する。潜在クラスの違いによって、重回帰分析の切片、重回帰係数、そして誤差項が異なると仮定するとき、有限混合多変量回帰モデルは以下のように表現される。

$$y_{ic} = \alpha_{ic} + \beta_{1c}x_{1i} + \beta_{2c}x_{2i} + \beta_{3c}x_{3i} + \beta_{4c}x_{4i} + \beta_{5c}x_{5i} + e_{ic} \quad (1)$$

i はオブザベーションの添え字であり、ここではオークションに出品された個々の車を表す。 α_{ic} は切片、 β は重回帰係数、そして e_{ic} は正規分布 $N(0, \sigma_{ic}^2)$ に従う誤差項である。 x_{1i}, \dots, x_{5i} は、それぞれ5つの予測変数に対応している。添え字 c によって潜在クラスからの影響が示されている。重回帰係数の違いに、落

札者がどの要因を重視するかの差異が表れていると解釈する。さらに、予測変数の値がすべて0(使用月数が0カ月、パワーウィンドウなし、ナビゲーションシステムなし、抹消登録なし、損傷なし)である車に対するオークション参加者の価値評価を表す切片や、落札金額のランダムな散らばりに左右される誤差がクラスごとに異なることを想定してモデル化を行った。以上の有限混合多変量回帰モデルをパス図で表すと図1のようになる。

本モデルの適用に先立って、注意すべき点について以下に2点述べる。1点目は、モデルの識別に関する問題である。有限混合モデルにおけるラベル交換問題については[6]をはじめ多くの文献で議論がなされているが、有限混合多変量回帰モデルには、ラベル交換問題以外に特別な識別問題があることが指摘されている[2][3]。すなわち、クラス数が重回帰モデルで用いられる予測変数の数よりも少ない場合、かつそのときに限って識別可能であるということである。本分析における予測変数の数は5個であるため、クラス数は4以下でないとモデルは識別されることがわかる。

2点目として、推定の初期値依存性に留意する必要がある。有限混合モデルでは与える初期値によって推定値が異なる可能性があり、局所解や不適解に陥りやすいことが一般に知られている。さらに、本モデルは実質科学的な解釈可能性を重視しており、重回帰係数に加えて切片と誤差項も潜在クラスごとに異なると仮定した制約の緩いモデルである。そのため、母数の数が多く、クラス数が増えるにつれて推定が困難になることが予測される。これらの理由により、異なる初期値を用いて複数回推定するなどの工夫が必要となる。本分析では、発生させた3,000組の乱数を初期値とし

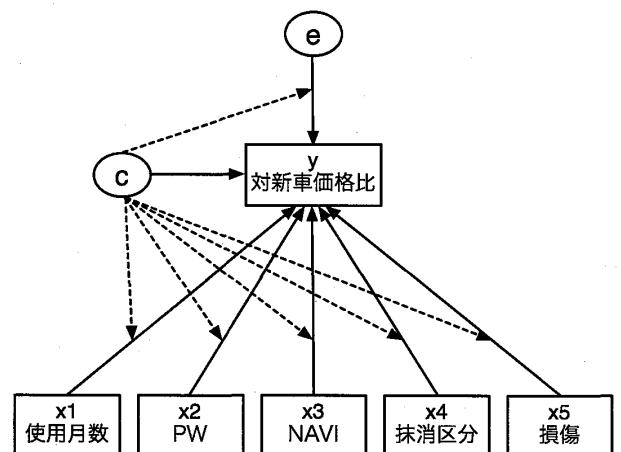


図1 モデルのパス図

て、推定結果の中から対数尤度の値が最も高いものを最終的な推定値とした。

母数の推定には、EM アルゴリズム⁴による最尤推定を適用した。EM アルゴリズムは、すべてのパラメータを同時に最適化するという複雑な作業をする代わりに、観測されない架空の変数を導入することによってパラメータをサブセットに分け、それぞれの最尤推定値が独立に容易に計算できる場合に用いられる[1]。有限混合モデルにおいては、各オブザベーションがどのクラスに属するのかわかりず、オブザベーション i が潜在クラス c に所属するかどうかを表す変数 $c_i = [c_{i1} \dots c_{ic} \dots c_{iC}]$ を、観測されない架空の変数として扱うこととなる。ここで、 i が潜在クラス c に所属する場合は $c_{ic}=1$ であり、そうでない場合は $c_{ic}=0$ である。EM アルゴリズムでは、観測されたデータと現時点でのパラメータの推定値から c_i を推定して、その c_i に基づいてパラメータの推定値を更新するという反復作業を、尤度の向上が見られなくなるまで継続する。 c_i を欠測値として扱い、EM アルゴリズムを適用することで、すべてのオブザベーションに対して、オブザベーション i が潜在クラス c に所属する事後確率が求められる。こうして計算された事後確率をもとに、各オブザベーションはその事後確率がもっとも大きかったクラスに分類される。分析は Mplus (ver. 4.1) を用いて行った。

上記のモデルによって、各オブザベーションのクラス分けを行った後に、「メーカー」、「形状」、「排気量クラス」の3つのカテゴリカル変数を用いて、各水準の構成比率をクラスごとに確認した。これにより、それぞれの潜在クラスに所属する車の特徴を知ることができる。

2.3 交差妥当化

分析にあたって、データクレンジング後のデータを、2005年6月～2006年6月の前半データ(22,881オブザベーション)と、2006年7月～2007年6月の後半データ(25,070オブザベーション)とに分割した。オーバーフィッティングを防ぎ、分析結果の頑健性を検討するために、まず前半データを用いてクラス数を2クラスから順に、理論的に識別可能な最大クラス数である4クラスまで増やし、情報量基準を参照すること

でクラス数の決定を行った。次に、後半データに対しても同様にクラス数2から4までで分析を行い、情報量基準に関して同様の結果が得られることを確認した。そして最も当てはまりのよかったクラス数の分析結果について推定値の比較検討を行ったところ、2つのデータセット間で大きな相違は見られず、推定の安定性が示されたため、次節では後半データに対する推定結果を用いて解釈を進めることとする。

3. 結果と考察

3.1 クラス数の決定

はじめに適切なクラス数を決定するために、情報量基準の値を参照した。

表1より、前半データ、後半データともにクラス数4の場合でAICおよびBICの値が最も小さかった。よって、クラス数は4が適当であると判断した。

3.2 有限混合多変量回帰モデルの適用

潜在クラス数を4に指定して、後半データを用いて有限混合多変量回帰モデルを適用した推定結果を表2に示した。()の中は z 値であり、その絶対値が1.96以上であれば推定値が有意であることを意味する。なお、 R^2 は自由度修正済み決定係数である。

表3には、各クラスに分類された車が獲得した札数について10区切りで度数表を作成した。札の多く集まった車は、人気のある車と見なすことができ、一方で札数の少ない車はあまり人気のない車であると考えられる。

z 値より、クラス1における損傷からの係数、クラス2におけるNAVIからの係数、クラス4における抹消区分と損傷からの係数以外は、すべて有意であった。標準化しないデータを用いて分析を行っているので、推定された係数は、当該変数以外の変数の値が同じならば、予測変数の値が1単位大きくなったとき基準変数の値が何単位変化するかという解釈が可能である。例えば、クラス1に注目するとき、その他4つの変数の条件が同じならば、パワーウィンドウが付いていることは、落札金額を新車価格比で0.46%高めることが読み取れる。以下で、有意性が認められた推定

表1 情報量基準の比較

クラス数	前半データ		後半データ	
	AIC	BIC	AIC	BIC
C=2	159747.6	159868.1	177327.8	177402.1
C=3	158963.7	159148.6	176620.9	176734.8
C=4	158437.6	158686.8	176102.2	176255.7

⁴ 有限混合モデルにおけるEMアルゴリズムによる推定の、詳細な理論および計算方法については文献[1]を参照されたい。

値をクラスごとに比較検討し、各潜在クラスに関する解釈を行った。

クラス1は所属割合が小さく、NAVIが装備されていることで落札金額が新車価格比で32.30%高くなることが示され、NAVIからの係数が極端に大きいことが特徴である。また、使用月数、抹消区分、損傷からの影響はそれぞれ-0.05、-0.64、-0.13であり、各係数が有意であった他のクラスに比較して最も小さかった。使用月数が長くても、抹消済みでも、損傷が多くても、対新車価格比はほとんど下がらない。これらのことから、実際にその車を使用する意図はほとんどなく、解体後に部品単位で使用するクラス、あるいは海外への輸出のために購入するクラスであると考えられる。ナビゲーションシステムを取り外し、他の車に取り付けて再利用するなど、部品としての需要を想定するとNAVIからの影響が大きい理由も説明できるだろう。表3より、クラス1に分類された車のほとんどが札数1~10のカテゴリに集中していることから、買い手は長年使用され傷も多く抹消済みであるという理由から人気の少ない車にターゲットを絞って、競争率の低い状態で落札しようとしていることがわかる。

次に、クラス2は、最も所属割合が大きく、過半数がこのクラスに分類されている。使用月数が1カ月長い、抹消済みである、損傷が1つ増えるという条件によって、それぞれ落札金額は新車価格比で0.25%、1.85%、0.45%低くなる。クラス1とは対照的に、こ

れら3つの変数がそろって対新車価格比に大きな負の影響を与えている。抹消済みであると落札価格が下がるということは、購入後すぐに乗りたいという意思の表れであり、しかもできる限り新車に近い状態を望んでいる。つまり、一般向けのレンタカーや個人の利用者のために、自家用としての使用目的で購入されるものと予測される。表3から、クラス2は、札数のばらつきが大きいことが見て取れる。札数が51~60と非常に人気の集中した車もこのクラスには2台含まれていた。

そしてクラス3とクラス4は、PWからの影響が比較的大きく、最低限の装備が求められていると解釈でき、共に事業用としての使用を目的としているものと考えられる。その上で、これら2つのクラスの差異をその他の変数が対新車価格比に与える影響の違いから検討する。クラス3には、全体の約4分の1が所属しており、抹消済みであることは、その他の4つの変数の値が同じならば落札金額を新車価格比で1.91%下げる。購入後に車検を取り直すことなくそのまま車を使用するためには、抹消済みであるか否かということがまず重要である。使用月数および損傷からの回帰係数はそれぞれ-0.13、-0.30であり、クラス2に比較してわずかに小さかった。これは、事業用として用いられる車では、自家用としての使用に比較すると、これらの要因に対する要求が緩められるためだと考えられる。以上の考察から、クラス3は、より状態が良く、すぐにでも使えるような事業用の車を求めるクラスとした。

クラス4は、所属割合が最も小さく、全体の3.6%がこのクラスに分類された。NAVIからの係数は-14.04と負の値であることから、装備されていると通常は価値が上がることが予測されるナビゲーションシステムの搭載によって対新車価格比が引き下げられる。クラス4に分類された車の内訳を観察すると、すべてナビゲーションシステムのない車であった。また、使用月数が1カ月長くなっても、落札金額は新車価格比で0.07%しか下がらない。長年使用され、ナビゲーションシステムが搭載されていないほど古い車でもかまわないという、小さなマーケティングセグメントの需要を表している。表3を見ると、札数11~30のカテゴリにおける度数が多く、他のクラスに比較して常に競争率の高い状態にあることがわかる。

3.3 各クラスの構成比率

続いて、車の形態に関する特徴を表す3つのカテゴ

表2 重回帰係数と決定係数

潜在クラス	クラス1	クラス2	クラス3	クラス4
所属割合	10.5%	59.1%	26.8%	3.6%
切片	8.64 (10.489)	39.71 (100.519)	22.27 (21.354)	41.30 (54.475)
使用月数	-0.05 (-8.412)	-0.25 (-63.928)	-0.13 (-13.693)	-0.07 (-4.972)
PW	0.46 (3.835)	0.67 (4.413)	0.93 (5.679)	0.74 (2.961)
NAVI	32.30 (41.681)	-0.61 (-1.436)	16.33 (19.301)	-14.04 (-11.048)
抹消区分	-0.64 (-2.473)	-1.85 (-8.942)	-1.91 (-6.241)	-0.39 (-1.366)
損傷	-0.13 (-1.890)	-0.45 (-5.296)	-0.30 (-3.444)	0.03 (0.260)
R ²	0.895	0.411	0.587	0.522

表3 札数の分布

札数	1~10	11~20	21~30	31~40	41~50	51~60
クラス1	2443	98	37	5	2	0
クラス2	3699	7849	2770	454	26	2
クラス3	4299	2180	209	26	4	0
クラス4	61	421	342	63	6	0

表4 カテゴリー変数の構成比率

	全体	クラス1	クラス2	クラス3	クラス4
メーカー					
いすゞ	0.1	0.0	0.1	0.0	0.1
スズキ	7.0	9.3	6.5	7.5	5.0
ダイハツ	9.0	14.7	6.5	12.6	4.6
トヨタ	37.4	10.4	48.3	18.7	79.3
ニッサン	22.8	27.5	17.9	34.5	2.1
ホンダ	6.7	5.0	7.4	6.1	4.8
マツダ	8.6	11.4	7.7	10.4	0.9
三菱	8.4	21.7	5.6	10.1	3.1
形状					
BV	37.8	49.9	31.2	51.5	9.4
BW	11.5	6.1	14.3	7.2	12.8
CT	1.2	0.4	1.6	0.4	3.0
CV	19.5	17.9	20.0	16.9	35.7
HB	11.2	17.0	9.5	13.8	2.9
SD	18.8	8.7	23.4	10.2	36.2
排気量クラス					
軽	23.7	41.5	18.7	29.1	14.0
小型	47.9	25.7	54.7	41.3	49.8
中型	17.0	20.6	15.7	19.1	12.0
ミドル	6.8	10.4	5.4	8.1	10.6
大型	4.6	1.9	5.5	2.4	13.5

リカル変数に注目して、各水準の構成比率をクラスごとに求めた。その結果は表4の通りである。単位は%であり、クラスごとの構成比率の比較基準として、「全体」の列には後半データ 25,070 オブザベーション全体における各水準の占める割合を示した。

メーカーの構成比率に注目すると、クラス1は、ニッサンと三菱の車でほぼ半数を占めている。特に、全体では8.4%しか含まれていない三菱の割合がこのクラスでは21.7%にも上っており、一方で全体の構成比率に比較してトヨタの車が非常に少ないことが特徴的である。部品単位での利用や海外への再販売のためには、できるだけ人気車を避け、コストを低く抑えようという意図があるのかもしれない。クラス2では、トヨタの割合が48.3%と高かった。クラス2ができるだけ新車に近い状態の自家用車のニーズを反映しているということからも、新車市場と同様トヨタの人気が高いことは納得できる。クラス3はトヨタとニッサンで過半数を占めている。ニッサンの割合は34.5%と全体での割合よりも11.7ポイント高く、他クラスと比較しても最も高い。自家用としての利用目的に比較して、事業用での使用となると技術面を武器としているニッサンの構成比率が顕著に大きくなるのは興味深い。クラス4においては、トヨタの占める割合が79.3%と圧倒的に高かった。

形状に関しては、クラス1では、BVの割合が49.9%であり、全体の場合と比較しても高かった。クラス2は、BVに加えてCVやSDの割合も高く、これらを合わせると約75%であった。ボンネットバン、

セダンは乗用車の一般的な形状であり、この結果もまた、自家用車として購入されていることを支持するものであった。クラス3で最も構成比率が大きかったのはBVで51.5%となっていることから、事業用として多く利用される車としてボンネットバンの需要が高いことがうかがえる。クラス4では、SDの割合が36.2%と最も高く、次いでCVが35.7%であり、これらは全体としての構成比率と比較しても高いといえる。

排気量クラス別の構成比率を見てみると、クラス1では、全体および他クラスと比較して軽自動車の割合が高く、41.5%であった。海外への輸送の都合上などから、コンパクトな軽自動車好まれるのかもしれない。自家用としての使用目的で購入されるであろうと考えられるクラス2においては、小型車の占める割合が54.7%と高かった。自動車市場全体としても生産需要とともに最も高いと予想される小型車は、典型的な自家用乗用車のイメージにも合致する。クラス3の構成比率は、全体の場合とほとんど大きな違いは見られなかった。軽自動車と小型車の割合が高く、事業用としてのニーズを表すものであると考えられる。クラス4においては、大型車の占める割合が13.5%であり、他クラスよりも高いことが特徴的であった。バスやトラックといった排気量の大きい大型車を含む特殊車がこのクラスに分類されやすいことがわかった。

4. まとめ

本分析では、車のオークションデータにおける落札価格を、出品された車に対する買い手の価値評価と捉えることで、使用目的という潜在変数を仮定した潜在構造モデルによって、買い手のニーズを探索した。本データのように、多くの変数によって車に関する情報が規定されている場合には、少数の特定の変数にのみ注目して多母集団を仮定することは情報の損失が著しく、限られた一面からしかデータを説明できないことが懸念される。そもそも多母集団分析では、母集団の数だけ同様のモデルを構成して母数の違い等を比較することになるため、仮定する母集団の数は通常2,3程度であり、それ以上になると互いに比較することが困難になる。よって、例えば数十もの種類があるメーカーを多母集団として設定することはできない。そこで潜在構造モデルを導入することで、車種やメーカー等既存のカテゴリによる多母集団分析では十分に捉えられなかった顧客の異質性をも明らかにすることができる。

特筆すべきは、購入対象となる商品に買い手自らが値付けを行うという、特殊なマーケティングの形態であるオークションの場においては、その値段を直接的な価値評価と見なすことが可能であるという仮説を分析の出発点としている点である。これにより、落札者の属性に関する変数を必要とすることなく、使用目的という買い手の嗜好を表す潜在的な変数を仮定することを可能にした。落札金額をもとに計算された対新車価格比を基準変数として、有限混合多変量回帰モデルを適用し、その結果から、潜在クラスは車の使用目的による分類として実際に解釈可能なものであることが示唆された。

さらに、有限混合多変量回帰モデルによる分析結果をもとに、3つのカテゴリカル変数を用いて各潜在クラスに分類された車の形態的特徴を明らかにした。3つの変数それぞれにおける、カテゴリごとの構成比率を比較検討することで、有限混合多変量回帰モデルで求められたクラスごとの解釈を裏付け、あるいは深めることができた。

購入する際の使用目的の違いによって、当然欲しい車に求める機能や状態も異なってくるはずである。オークション参加者のニーズに注目したマーケット・セグメンテーションは、オークション主催者にとって、買い手がどのような特徴を持ったセグメントから構成されているのかを正しく知り、それぞれのニーズに応じた有効なアプローチを検討していくための第一歩となる。本モデルの適用によってオークションにおける買い手の値付けメカニズムは大きく4種類に分類することが明らかになった。これら4つのクラスの特徴に応じて、「いますぐ乗れる新車同様の車をお探しの方はこちらへ」「海外への輸出車をお探しの方はこちらへ」といったように、セグメントごとにおすすめの車をピックアップしておくことで、オークション参加者がよりスムーズに目当ての車を検索するのに役立つだろう。

最後に、有限混合モデルの実用性に関して、以下の2点を課題として挙げる。1つは、有限混合モデルの中でも多変量回帰モデルに特有の識別問題についてである。有限混合多変量回帰モデルでは、重回帰モデルで用いられる予測変数の数よりもクラス数が少ない場合、かつそのときに限って識別可能である。本分析では予測変数の数を5個としたため、分析を実行する以前に、モデルが識別されるためにはクラス数は4以下でなくてはならないという条件が課される。モデルの

理論上の制約により、仮定できるクラス数の上限が予測変数の数に依存して決まってしまうのである。さらに文献[2][3]によると、予測変数の値が変化に乏しい場合、すなわち、マーケティング・リサーチにおいてしばしば見られるように、ダミー変数や水準数の少ないカテゴリカル変数が予測変数として用いられる場合には、上述の識別条件を満たしていてもなおモデルが識別されないこともあるという。

このように、有限混合多変量回帰モデルは識別条件の非常に厳しいモデルであるため、分析者があらかじめ立てた仮説にしたがって自由にクラス数を増やすことができない可能性がある。本分析では、理論的に識別可能な最大クラス数である4クラスで推定が可能であり、かつ、2クラス、3クラスの場合に比較して情報量基準が小さかったことから、最終的に4クラスが最適であると判断した。しかしながら、本モデルの汎用性を高めるためには、識別条件を緩め、クラス数決定に関する選択の幅を広げる必要があるだろう。理論上の制約に縛られることなくクラス数を増やすことができれば、より詳細なセグメンテーションが可能になることが期待される。

もう1つは、潜在構造モデルにおける初期値依存性である。本分析では3,000個もの乱数を発生させることで初期値として用いているが、クラス数を増やすとさらに多くの初期値を用意しないと収束は一層困難を極める。あらかじめ適切な初期値を知ることが難しいが、より早い時点でそれを見つけることができれば、計算時間が短縮できる。大規模なデータの中から、限られた時間で有益な情報を見つけ出すことを目的とするデータマイニングにおいて、いかに短い時間でより正確な解が得られるかということは、非常に重要な課題である。

謝辞 早稲田大学マーケティング・コミュニケーション研究所 CRM 研究会の先生方をはじめ、「平成19年度データ解析コンペティション」の審査員の先生方、ならびにコンペに参加された皆様には、本分析に対して有用なコメントを多数いただきました。ここに記して深謝申し上げます。また、データを提供して下さったシステム・ロケーション株式会社の関係者の皆様には、心より感謝申し上げます。

参考文献

[1] 阿部誠, 近藤文代: 消費者の異質性, 「マーケティング

- の科学—POSデータの解析—, 予測と発見の科学3, 第8章, 朝倉書店 (2005).
- [2] Frühwirth-Schnatter, S.: Finite Mixture of Regression Models, *Finite Mixture and Markov Switching Models*, chapter 8, Springer, New York (2006).
- [3] Henning, C.: "Identifiability of Models for Clusterwise Linear Regression," *Journal of Classification*, Vol. 17, pp. 273-296 (2000).
- [4] 小嶋外弘: 「価格の心理」, ダイヤモンド社 (1986).
- [5] コトラー, P.: 「コトラーのマーケティング入門 (Marketing an introduction) 第4版」, 恩藏直人 (監修), 月谷真紀 (訳), ピアソン・エデュケーション (1999).
- [6] Stephens, M.: "Dealing with Label Switching in Mixture Models," *Journal of the Royal Statistical Society, Ser. B*, Vol. 62, No. 4, pp. 795-809 (2000).
- [7] 豊田秀樹: 「共分散構造分析〈応用編〉—構造方程式モデリング—」, 朝倉書店 (2000).
- [8] 渡辺美智子: 因果と構造を把握するための統計手法—潜在クラス分析法—, 「マーケティングの数理モデル」, 経営科学のニューフロンティア6, 第3章, 朝倉書店 (2001).
- [9] Wedel, M. and Kumakura, W. A.: *Market Segmentation-Conceptual and Methodological Foundations*, International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston (2000).