



Discussion Papers In Economics And Business

ID 付 POS データを用いた考慮集合の推測

ウィラワン・ドニ・ダハナ

中島 望

Discussion Paper 11-25

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

ID 付 POS データを用いた考慮集合の推測

ウィラワン・ドニ・ダハナ

中島 望

Discussion Paper 11-25

August 2011

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

ID 付 POS データを用いた考慮集合の推測

ウィラワン・ドニ・ダハナ†

中島 望‡

要旨：

考慮集合の概念は古くからマーケティング研究者の主な研究対象のひとつとされている。消費者へのアンケートなどによりその「中身」を解明しようとする研究は数多いが、消費者の購買履歴データを用いたものはわずかである。本研究では考慮集合と選択集合の形成過程をモデル化することにより ID 付 POS データから考慮集合の推測を試みる。考慮集合の情報からマーケティングへの有用なフィードバックの可能性を考察する。

JEL Classification Numbers: M11, M31, M37

キーワード：考慮集合、選択集合、選択モデル

† 大阪大学大学院経済学研究科 〒560-0043 豊中市待兼山 1 - 7

E-mail: dony@econ.osaka-u.ac.jp

‡ 大阪大学大学院経済学研究科 〒560-0043 豊中市待兼山 1 - 7

E-mail: nakajima@econ.osaka-u.ac.jp

1. はじめに

スキャナーデータの登場をきっかけに、POS に代表される情報システムの高度化は、売り場での単なる省力化だけではなく、マーケティングの方法にもいろいろと影響を与えてきた。さらに、顧客の囲い込みや情報収集を目的に発行された顧客カードが利用され、購買データは個人の ID とともに記録されることで (=ID 付 POS データ)、パネルデータとしての性格も備えるようになった。そのため、こうした購買データを用いて単に価格や店内販促への消費者反応を記述するだけでなく、消費者の情報処理の様子まで推測しようという研究が増えつつある。例えば、価格閾値の推測に関する研究[1]がある。本研究もこうした試みの 1 つであり、消費者が商品を購入する際にどのようなアイテムを検討したのか (=選択集合)、ID 付 POS データから推測を行ってみる。

消費者の選択集合がわかれば、選択の場面におけるマーケティング変数の効果が的確に把握できるだけでなく、最終的な選択に至るまでの諸局面で行われてきたマーケティング活動全般について有益な診断情報が得られる。例えば、大量の広告投入の目的のひとつとして、自社ブランドが消費者の考慮集合に入ること、とりわけ最初に想起されるブランドになることがよくあげられる。また当然のことながら、製品使用後の製品評価は考慮集合の形成と密接な関係にある。こうしたことから、直接には観測されない考慮集合・選択集合について、ID 付 POS データのような日常のオペレーションを通じて半ば自動的に収集されるデータから推測が可能であれば、小売業だけでなくメーカーにとってもマーケティングの実践にきわめて有益となるだろう。

消費者は商品を購入する際に認知している商品の全てを検討の対象とするわけではなく、通常、はるかに少ない商品を検討の対象として想起すると考えられている。ブランド・カテゴリライゼーション[2]の概念モデルでは、想起集合 (evoked set) と呼ばれる検討対象となる商品の集合が規定されており、製品カテゴリへの関与水準や情報処理能力・コストなどの制約から、その要素数は通常 2~8 程度とされている。このような選択時に検討対象となる商品の集合を指す用語として、想起集合以外にも考慮集合 (consideration set) や選択集合 (choice set) といったものもよく使われる。

消費者の購買履歴データの分析にあたって、こうした選択集合を組み込んだ選択モデルもいくつか提案されているが (例えば[3],[4],[5])、どうしても多数のアイテムを含んだ選択集合も扱わざるを得ず、計算負荷の面でも現実的ではない。それに対し本研究では、各消費者は少数の要素からなる考慮集合を持つものとし、考慮集合の要素自体が時間的に変化することは認めるものの、要素数は常に一定と仮定している。そのかわり実際の購買に際しては、考慮集合だけでなく店頭でのプロモーション効果も考慮して選択集合を構成し、そこから選択 (購買) を行うものとする。そうすることで、選択の際の代替案 (選択肢) の数を現実的な数に保ちながら、店頭での効果も考慮したモデル化が行える。

このようなモデル化のもたらすメリットはいくつかあげられる。まず、店頭にあるすべての商品を選択の対象とするのではなく、考慮集合と店頭効果による追加的な要素からな

る選択集合を検討対象とすることは現実的な状況を反映しており、それに基づいた選択モデルは、選択結果をよりの確に反映したものになるであろう。またモデルの推定結果からは、考慮集合に含まれるアイテム数によるセグメンテーションが可能であり、アイテム数の多寡による購買ブランドや価格の違い、あるいは考慮集合の構成内容に関する特徴など、ターゲット・マーケティングにとって有益な情報が得られるものと期待される。

本稿の次節以降の構成は次のとおりである。第2節で提案モデルの詳細を述べ、第3節でモデルの推定に用いるデータについて説明する。第4節ではパラメータの推定結果と推測した考慮集合についての分析結果を示す。最後に第5節でまとめを述べる。

2. 考慮集合の形成と選択モデル

本研究では消費者のブランド選択行動を2つの段階に分けている。第1段階は消費者が認知しているすべてのブランドを少数のブランドに絞り込む段階である。この段階はさらに消費者が記憶をベースに考慮集合を形成する段階と、店頭における外部探索の結果を追加した選択集合を形成する段階に分けている。第2段階は消費者が選択集合のなかから最も魅力的なものを選択する段階である。

図1

図1は提案したモデルの概要を示している。購買機会 t における消費者 h の考慮集合と選択集合をそれぞれ E_{ht}, C_{ht} と表す。ただし、考慮集合の要素数は時間とともに変化しないと仮定する。時刻 t で消費者は記憶に基づいて考慮集合 E_{ht} を形成するが、店頭での探索結果から必要に応じていくつかのアイテムを考慮集合に加えて選択集合 C_{ht} を形成する。そして、消費者は選択集合 C_{ht} のなかから効用値が最大なアイテム $b_{ht} \in C_{ht}$ を選択する。購買後、形成された選択集合は記憶内に保存され、次の購買機会に考慮集合 $E_{h(t+1)}$ として想起される。ただし、考慮集合の要素数は時間とともに変化しないと仮定しているため、選択集合 C_{ht} のなかから効用値の大きい順に一定の数のアイテムが次期の考慮集合 ($E_{h(t+1)}$) として想起される。

より具体的には以下のような形成の手続き(i)~(iii)とする。

(i) $k \in E_{ht} \Rightarrow k \in C_{ht}$.

(ii) $k = b_{ht} \Rightarrow k \in C_{ht}$.

(iii) $P_{kht} < \text{mean}(P_{kh1}, \dots, P_{kh(t-1)}) \times 0.8 \Rightarrow k \in C_{ht}$.

ただし、 P_{kht} 時点 t で消費者 h が直面したアイテム k の価格、 b_{ht} は選択されたアイテムである。(i)は考慮集合に属するアイテムはすべて選択集合に属するという仮定である。(ii)は選択されたアイテムは必ず選択集合に入るという仮定で、モデルの整合性のために必要な仮定である。(iii)は価格が過去の平均価格より2割以上安いアイテムは選択集合に入るという仮定を示している。

ところで、考慮集合の要素数に関する先行研究によると、いくつかの日用品について考慮集合のサイズが2～5という報告がある[6]。考慮集合のサイズは直接に観測できない潜在的な概念であるが、ここでは1～5と設定する。つまり、考慮集合のサイズによって市場を5つのセグメントに分割し、各セグメントへの所属確率を潜在クラスモデルを用いてモデル化する。なお計算に便利のように消費者 h の考慮集合のサイズが $l = 1, 2, \dots, 5$ である確率を $f_{hl} = \exp(\lambda_{hl}) / \sum_m \exp(\lambda_{hm})$ のように表しておく。

一方、選択集合のサイズは考慮集合のサイズに依存するので、ここからは選択集合を $C_{ht}^{(l)}$ と表す。選択集合 $C_{ht}^{(l)}$ が与えられたとき消費者 h は $C_{ht}^{(l)}$ のなかから効用値が最も大きいアイテムを選択すると仮定する。時点 t における消費者 h のアイテム $j \in C_{ht}^{(l)}$ の効用を次のように定義する。

$$(1) \quad u_{jht}^{(l)} = \alpha_{jh}^{(l)} + \beta_{1h}^{(l)} P_{jht} + \beta_{2h}^{(l)} BL_{jht} + \beta_{3h}^{(l)} ML_{jht} + e_{jht}^{(l)}$$

ただし、 BL_{jht}, ML_{jht} はそれぞれ時点 t における消費者 h のアイテム j に対するブランド・ロイヤルティとメーカー・ロイヤルティである。これらのロイヤルティ変数は特定のアイテムへの偏った購買行動を補足した変数であるが、ここでは[7]の定式化に倣っている。最後に e_{jht} は極値分布に従う誤差項である。ここで注意して欲しいのは(1)式の右辺の切片と係数 (α, β) は個人および所属するセグメントによって異なっている点である。

考慮集合のサイズが l という条件のもとで消費者 h がアイテム j を選択する確率を次のように表す。

$$(2) \quad \Pr_{ht}(j|l) = \frac{\exp(u_{jht}^{(l)})}{\sum_{k \in C_{ht}^{(l)}} \exp(u_{kht}^{(l)})}$$

また、アイテム j の条件なし選択確率を次のように表すことができる。

$$(3) \quad \Pr_{ht}(j) = \sum_l f_{hl} \Pr_{ht}(j|l)$$

最後に、全パラメータの集合を θ とおけば消費者 h の尤度関数を次のように求めることができる[8]。

$$L_h(\theta) = \sum_l \left[\frac{\exp(\lambda_{hl}) \prod_{t=1}^T \Pr_{ht}(b_{ht}|l)}{\sum_{k=1}^5 \exp(\lambda_{hk})} \right]$$

提案したモデルの説明力を確かめるために、ほかの2つのモデルとの比較を行う。ひと

つは消費者が購入したことのあるアイテムが常に選択集合に入っているモデルである。もうひとつはすべてのアイテムが選択集合に入っているという従来の選択モデルである。

- M1：提案モデル
- M2：消費者が購入したことのあるアイテムが選択集合とするモデル
- M3：分析対象となるアイテムの全体が選択集合とするモデル

3. データの概要

上述の提案したモデルを消費者の購買履歴データに適用し、考慮集合の推測を行う。使用するデータはカスタマー・コミュニケーションズ(株)より提供されたドラッグ・ストアのID付POSデータである。分析対象製品として歯磨きを選んだ。分析単位はブランドではなく、より詳細な単位であるアイテムを用いている。提供されたデータの制約から、ここでは同一ブランドの異なったアイテムを識別するためにブランド名の後ろに異なる番号を付けることにする。分析対象アイテムとして販売数量シェアが高い20個のものを選んだ。この20個のアイテムの合計販売シェアは52.7%である。表1は分析対象アイテムの平均販売価格とシェアを示している。

表1

分析対象顧客としては、データ期間中に歯磨きを10回以上購入した顧客を選んだ。この条件を満たしている顧客数は672人である。図2で示すように、対象顧客のうちおよそ9割は女性であり、年齢で見ると5割以上は40代から60代までの顧客である。また、大部分の顧客はデータ期間中に15回から20回ほど購入し、2~5種類の異なったアイテムを購入している。

図2

4. 分析結果

モデルの推定にはMCMC法のM-Hアルゴリズムを用いた。イタレーションの回数は2万回とし、そのうち最後の5千回をパラメータの推測に用いる。考慮集合の初期値として販売シェアの上位のアイテムを選んだ。表2は推定した3つのモデルの対数周辺尤度を示している。この表から、提案したモデル(M1)は他の2つのモデル(M2,M3)に比べてうまくデータを説明していることが分かる。全アイテムが選択集合とするM3の精度が最も悪く、消費者が全アイテムではなくより少数のアイテムのなかから選択行動を行なっているということを反映していると考えられる。また提案モデル(M1)がM2より優れていることに関して、消費者は常に購買経験のあるアイテムの集合から選択行動を行うのではなく、時には外部探索を行い店頭で魅力的なアイテムがあればその都度選択集合に加えているという行動を取っていると解釈できる。

表 2

次にセグメントへの所属確率のパラメータ λ_{hl} の推定結果を見てみる。表 3 は λ_{hl} の推定値を顧客について平均したものを表している。ただし、識別性の問題を避けるために $\lambda_5 = 0$ と固定している。 λ_{hl} の推定値を用いて顧客の各セグメントへの確率 f_{hl} を求めることができる。番号が 1 から 5 までの顧客の f_{hl} の値を表 4 で示す。この表を見ればわかるように、いずれか 1 つのセグメントへの所属確率が 1 に近い顧客がほとんどである（所属確率の最大値が 0.7 以下の顧客は 12 人）。従って、各顧客をこの所属確率が最大となるセグメントに振り分けても問題ないであろう。つまり、もし $f_{hk} = \max_l(f_{hl})$ であれば顧客 h は k 番目のセグメントに所属すると見なしてよい。このようにして顧客を 5 つのセグメントに分類し考慮集合のサイズの分布を描くと、図 3 のような分布が得られる。この図から歯磨きのカテゴリーにおいて多くの顧客は考慮集合のサイズが 3 または 4 であることが分かる。

表 3

表 4

図 3

表 5 は各セグメントの反応係数の事後平均を示している。これを見れば反応係数はセグメント毎に異なっていることが分かる。価格の係数は考慮集合のサイズが大きくなるに連れて小さくなる。これは考慮集合のサイズが大きいセグメントほど価格に対しより敏感であることを示している。一方、ブランド・ロイヤルティとメーカー・ロイヤルティについてはそういった単調な傾向が見られない。

表 5

これまではパラメータの推定値の議論をしたが、ここからは考慮集合の推測結果の分析について述べる。図 4 は対象アイテムがどれだけ顧客の考慮集合に入っているかを示す考慮集合シェアと実際の販売シェアを示している。図 4 を見ると考慮集合シェアと販売シェアはアイテムについて完全には連動していないことが分かる。つまり、販売シェアが低くても考慮集合シェアが比較的高いアイテムがあれば、またその逆のアイテムもある。前者の場合、顧客の考慮集合に入るという面では広告や製品そのものがうまくいっているものの、選択率が低いという面ではプロモーションなどの最後の一押しが足りないアイテムであると考えられよう。

図 4

次に、考慮集合のサイズ毎に購入アイテムのシェアを調べてみるといくつかの際立った違いが目につく。例えば、セグメント 1 に属している顧客はクリアクリーン 1 を多く購入しているが、セグメントに 2 に属している顧客はクリアクリーン 3 やデントヘルスの方をよく購入していたりする。図 5 から分かるように、他のセグメントについても購入するアイテムに違いがあることを見ることができる。

図 5

また、購入アイテムだけではなく、セグメントの間で考慮するアイテムに違いのあることもわかる。例えば、図 6 で示しているように、セグメント 2 ではアクアフレッシュ 2 があまり考慮されていないのに対して、セグメント 4 では考慮されている上位のアイテムになっている。さらに、この両セグメントの購入価格について見てみると、もうひとつの違いが見られる。図 7 で示しているように、セグメント 1 とセグメント 2 に属している顧客は主に 3 割引程度の価格水準で購入しているのに対して、セグメント 3、セグメント 4 とセグメント 5 に属している顧客はずっと幅広い価格掛け率で購入を行っている。このように、考慮集合のサイズによるセグメンテーションを行うことによって各セグメントの行動の違いを明確にすることができる。

図 6

図 7

5. まとめ

本稿では選択行動データから直接に観測できない考慮集合の推測を試みた。推測が可能なようにいくつかの強い仮定を置いたが、結果として提案したモデルは考慮集合を取り入れない従来のモデルと比較してデータの説明力の面において優れている。また、考慮集合のサイズをモデル化することにより、従来調査をしなければわからなかった情報を消費者の購買データから引き出すことができた。

このモデルを用いた分析により考慮集合の「中身」を把握することができたが、仮にその推測結果が消費者の真の考慮集合に近い結果だとすれば、そこから消費者行動に関する新しい知見を得ることができる。例えば、消費者が頻繁に考慮するアイテムはどのようなものなのか、頻繁に考慮されているアイテムは選択されているのかといった情報が得られる。また、市場を考慮集合のサイズに基づいてセグメント化することによって各セグメン

トの行動的な特性の違いを見ることができる。例えば本研究の結果にあったように、考慮集合のセグメントによって購入する商品、考慮する商品、購入する価格掛け率が異なっていることが分かる。

考慮集合に関する知識はまた、消費者の選択行動の記述だけではなく、マーケティングへの有益なフィードバックを提供できると考える。例えば、メーカーにとって広告した自社ブランドが消費者の購買場面において選択の対象とされているかという評価が可能である。また、選択の対象とされてはいるが購入率が低い商品があれば消費者の購買を促すプロモーションの施策にも有用なヒントが得られよう。メーカーのみならず、商品アソートメントの改善を図りたい小売業にとっても、考慮集合の情報は重要な示唆を与えてくれると考えられる。

最後に、本研究は考慮集合という潜在的な概念を可視化する試みであり、そのモデル化において考慮集合と選択集合の形成に関する厳しい仮定を置いている。これらの仮定についてもっと現実に近づけるようなものにする必要があると考え、これを今後の課題とする。

付録：モデルの推定方法（MCMC法）

1. 効用関数の係数 $\beta_h^{(l)}$ の推定：

- ① 初期値 $\alpha_{jh1}^{(l)} = \alpha_{M3}, \beta_{h1}^{(l)} = 0, E_1 = \{\text{販売数量で上位 } l \text{ 個のアイテム}\}$ を設定する。ただし、 α_{M3} は M3 の切片の推定値である。
- ② s 番目のイタレーションで $e_s \sim N(0,09)$ をサンプリングし、 $\beta_{h*}^{(l)} = \beta_{hs}^{(l)} + e_s$ とする。
- ③ $\phi(\beta_{h*}^{(l)}, \beta_{hs}^{(l)})$ の確率で $\beta_{h*}^{(l)}$ を採択する ($\beta_{h(s+1)}^{(l)} = \beta_{h*}^{(l)}$)。ただし、 $\pi(\beta_h^{(l)})$ を $\beta_h^{(l)}$ の事前分布、 $L(\beta_h^{(l)})$ を $\beta_h^{(l)}$ の尤度関数としたとき

$$\phi(\beta_{h*}^{(l)}, \beta_{hs}^{(l)}) = \frac{\pi(\beta_{h*}^{(l)})L(\beta_{h*}^{(l)})}{\pi(\beta_{hs}^{(l)})L(\beta_{hs}^{(l)})}$$

である。

- ④ $\alpha_{jh}^{(l)}$ の推定も同様の手続きで行う。

2. セグメントへの所属確率のパラメータ λ_{hl} の推定：

- ① 初期値 $\lambda_{hl1} = 0$ を設定する。
- ② s 番目のイタレーションで $\varepsilon_s \sim N(0,09)$ をサンプリングし、 $\lambda_{hl*} = \lambda_{hls} + \varepsilon_s$ とする。
- ③ $\phi(\lambda_{hl*}, \lambda_{hls})$ の確率で λ_{hl*} を採択する ($\lambda_{hl(s+1)} = \lambda_{hl*}$)。ただし、 $\pi(\lambda_{hl})$ を λ_{hl} の事前分布、 $L(\lambda_{hl})$ を λ_{hl} の尤度関数としたとき

$$\phi(\lambda_{hl*}, \lambda_{hls}) = \frac{\pi(\lambda_{hl*})L(\lambda_{hl*})}{\pi(\lambda_{hls})L(\lambda_{hls})}$$

である。

参考文献

- [1] N. Terui and W.D. Dahana: "Estimating Heterogeneous Price Thresholds," *Marketing Science*, Vol.25 no.4, 384-391, 2006.
- [2] J.E. Brisoux and M. Laroche: "A Proposed Consumer Strategy of Simplification for Categorizing Brands," In John D. Summey and R.D. Taylor Eds., *Evolving Marketing Thought for 1980*, Southern Marketing Association, 112-114, 1980.
- [3] M.J.I Gaudry and M.G. Dagenais: "The Dogit Model," *Transportation Research, Series B* 13B, 105-111, 1979.
- [4] J.D. Swait and M. Ben-Akiva: "Empirical Test of a Constrained Choice Discrete Model: Mode Choice in Sao Paulo, Brazil," *Transportation Research, Series B* 21B, 103-115, 1987.
- [5] R.L. Andrews and T.C. Srinivasan: "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data," *Journal of Marketing Research*, 32, 30-41, 1995.
- [6] J.R. Hauser and B. Wernerfelt: "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16, 393-408, 1990.
- [7] P. Guadagni and J. Little: "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, Vol. 2. no.3, 203-238, 1983.
- [8] W.A. Kamakura and G.J. Russell: "A Probabilistic Choice Model for Market Segmentation and Elasticity Structuring," *Journal of Marketing Research*, 26, 379-90, 1989.

図1 モデル概要

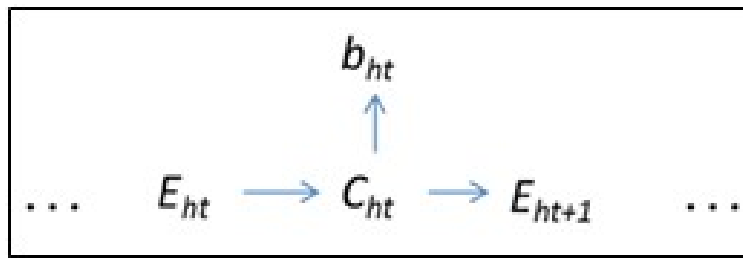


表1 対象製品の基本統計

アイテム	平均価格	数量シェア
アクアフレッシュ 1	125	3.59%
アクアフレッシュ 2	122	4.23%
シュミテクト	752	1.34%
ガードハロー	74	5.02%
クリアクリーン 1	175	3.45%
クリアクリーン 2	176	2.36%
クリアクリーン 3	176	4.17%
つぶ塩	303	0.97%
GUM1	272	0.79%
GUM2	257	1.35%
オーラ 2	280	1.33%
薬用クリアガード	101	9.77%
薬用メディカ	138	1.78%
デンタークリアMAX1	149	1.28%
デンタークリアMAX2	146	2.07%
デンタークリアMAX3	148	2.19%
デンターシステム 1	343	0.88%
デンターシステム 2	277	0.85%
デントヘルス	1234	1.81%
ホワイト&ホワイト	120	3.44%
	合計	52.68%

図2 分析対象顧客の特性

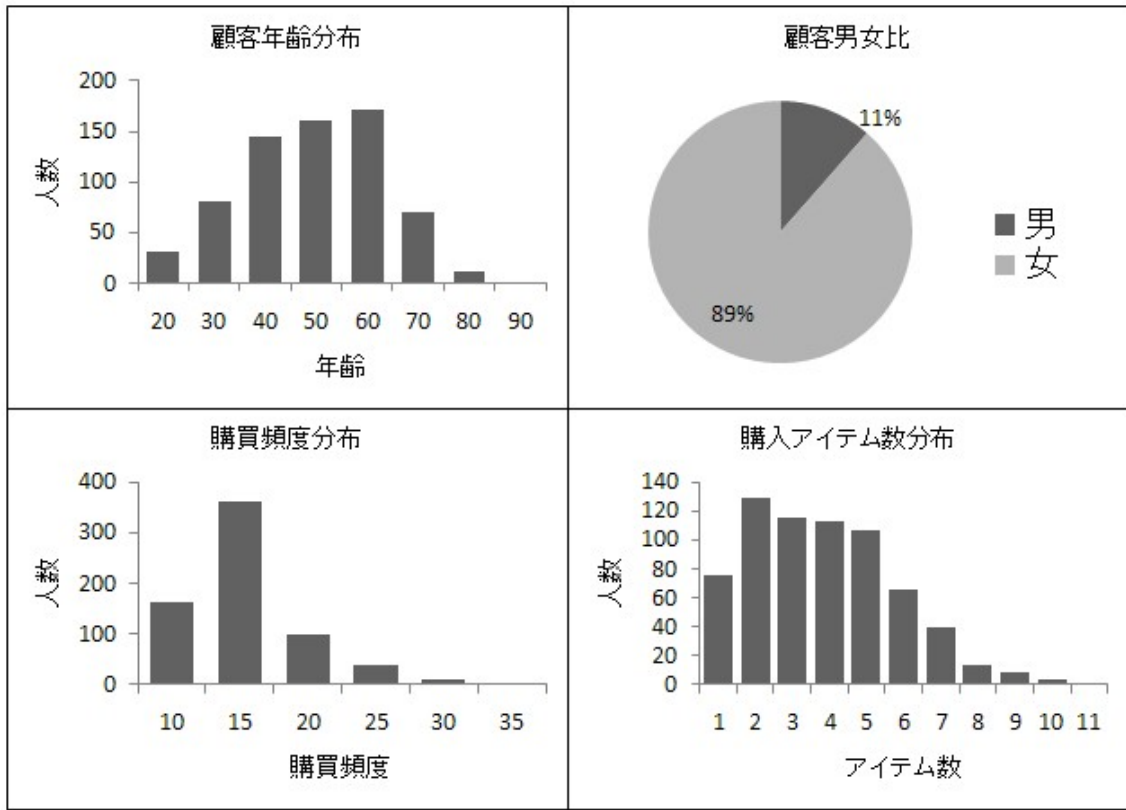


表2 比較モデルの対数周辺尤度

	対数周辺尤度
M1	-4,787
M2	-5,621
M3	-7,439

表3 λ の推定値

	事後平均	事後分散
λ_1	-10.969	17.345
λ_2	-5.315	5.481
λ_3	0.707	1.685
λ_4	3.765	2.677
λ_5	0	-

表 4 f_{hl} の推定値

顧客番号	f_1	f_2	f_3	f_4	f_5
1	0	0	0	1	0
2	0	0.001	0	0.999	0
3	1	0	0	0	0
4	0.015	0	0	0.985	0
5	0	0	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮

図3 考慮集合のサイズの分布

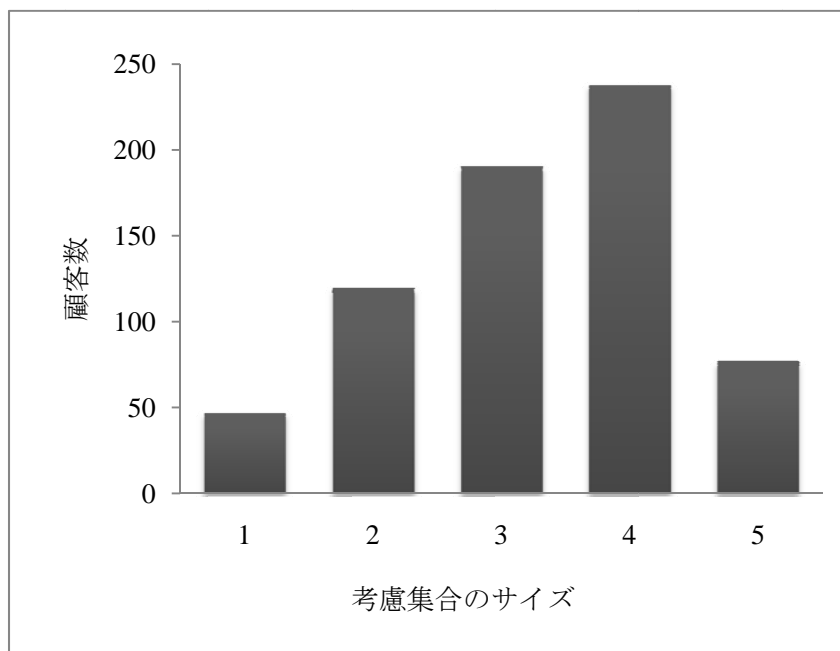


表 5 β の推定値

	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
価格	-0.068	-3.976	-5.258	-6.657	-7.460
BL	3.278	6.825	6.716	8.093	7.771
ML	1.671	2.610	2.736	1.972	2.051

図4 考慮集合と選択集合のシェア

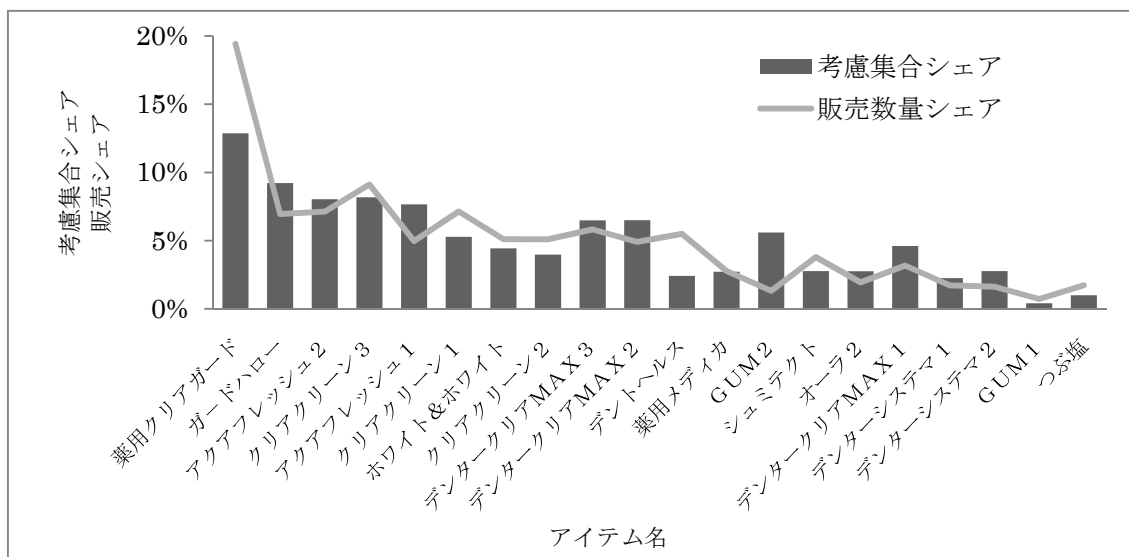


図5 各セグメントの購入アイテム

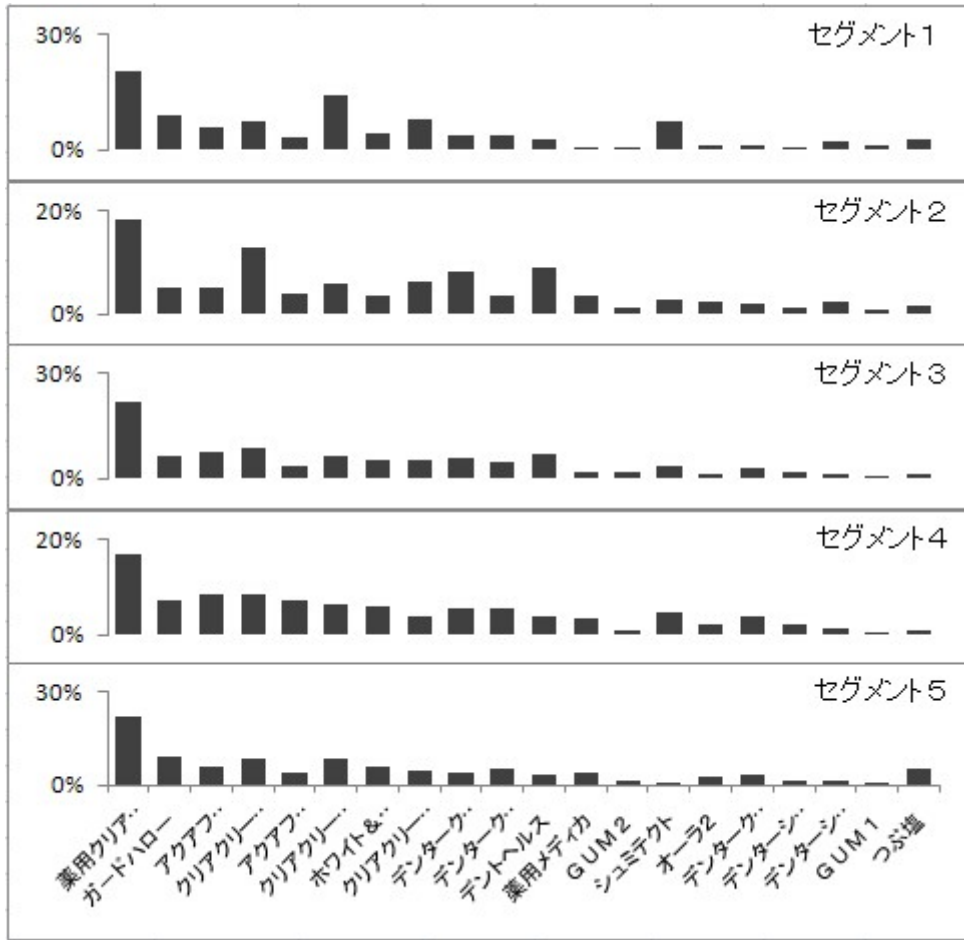


図6 各セグメントの考慮アイテム

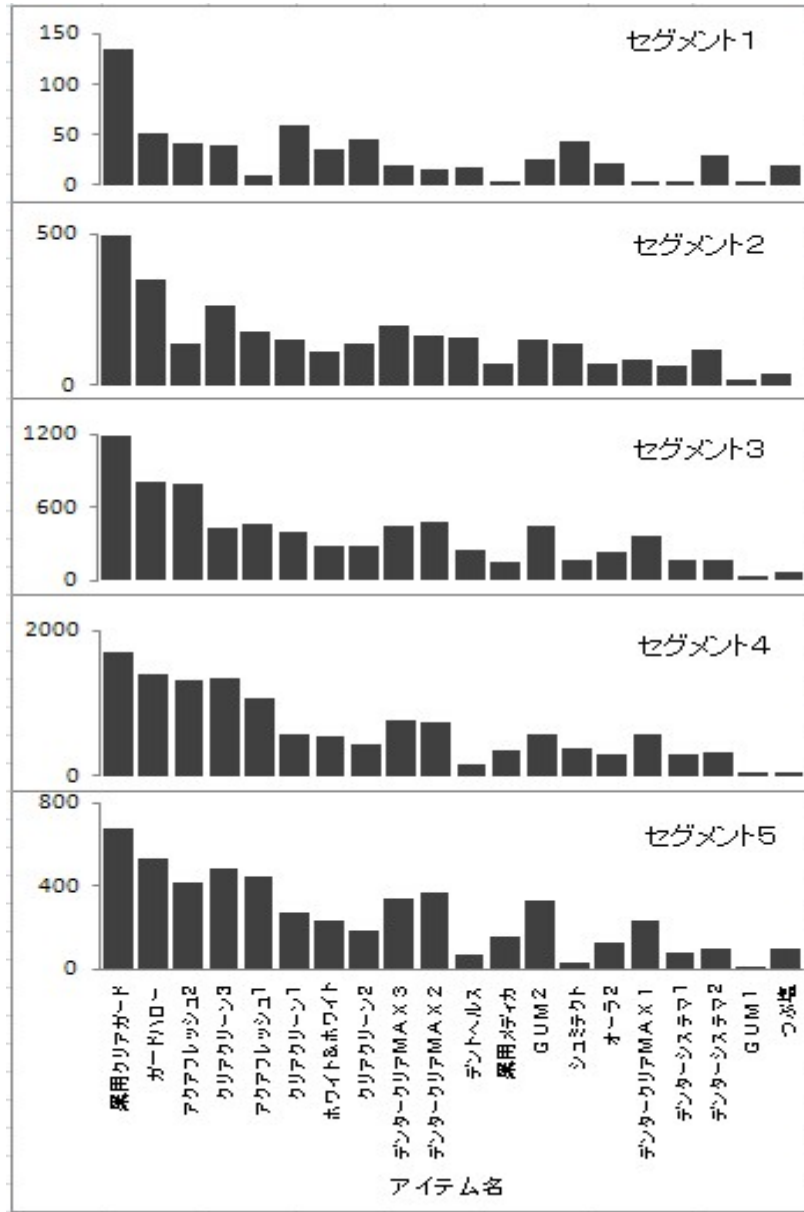
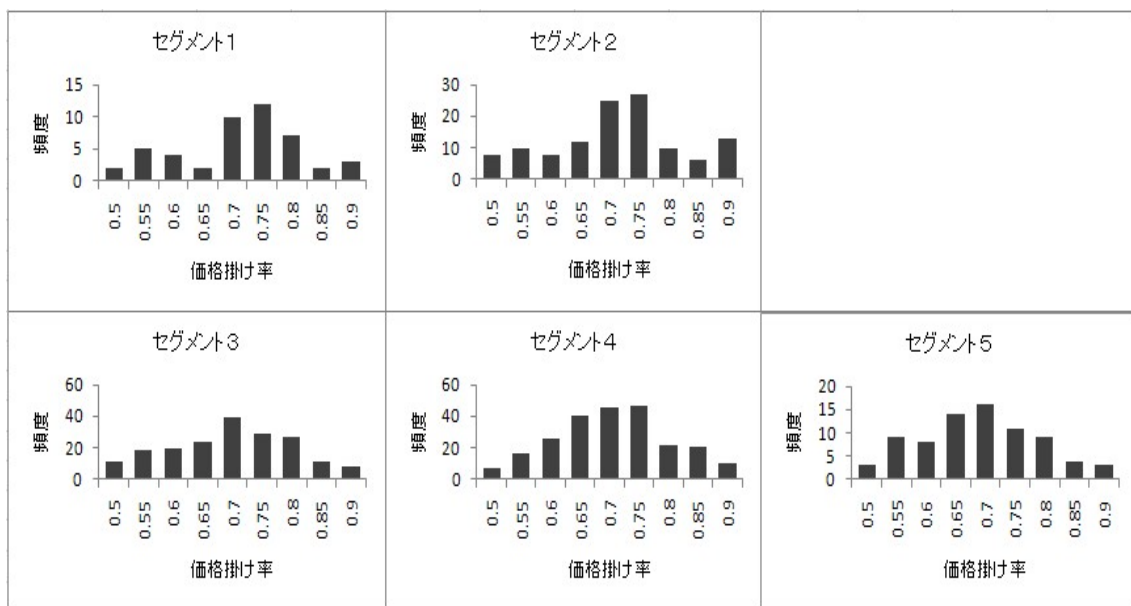


図7 各セグメントの購入価格掛け率



Inferring Consideration Set from Scanner Data

Wirawan Dony Dahana[†]

Nozomi Nakajima[‡]

Abstract:

Consideration set has been one of main research topics in marketing field for a long time. Using interview data, many studies have been conducted to investigate the nature of its content. However, only few studies tried to do so by using consumer purchase history. By modeling the process of consideration and choice set formation, in this research we try to elicit consumer consideration set from scanner data. Some managerial implications for marketing decisions derived from the information of consideration set are discussed.

JEL Classification Numbers: M11, M31, M37

Keywords: Consideration Set, Choice Set, Choice Model

[†] Wirawan Dony Dahana, Graduate School of Economics, Osaka University, 1-7 Machikaneyama, Toyonaka, Osaka, Japan. E-mail: dony@econ.osaka-u.ac.jp.

[‡] Nozomi Nakajima, Graduate School of Economics, Osaka University, 1-7 Machikaneyama, Toyonaka, Osaka, Japan. E-mail: nakajima@econ.osaka-u.ac.jp.