

マーケティング研究と統計的手法

慶應義塾大学商学部教授

高橋 郁夫

はじめに

マーケティング研究は、売手と買手の取引や相互作用プロセスとして表出される複雑な現象を対象とする。しかも、これはミクロおよびマクロの両面から捉えることができる。例えば、ある TV コマーシャルが消費者の購買意欲に及ぼす影響は、ミクロ的なマーケティング現象として、他方、広告一般が企業間競争や物価に及ぼす影響は、マクロ的なそれとして、それぞれ認識可能である。マーケティング研究の目的は、このようなマーケティング現象を記述・説明し、さらには予測したり、あるいは、企業等の意思決定主体の目標に沿うようにそれを統制したりすることにある。したがって、こうした作業の究極的目的は、マーケティング理論の構築と、マーケティング管理にとって有用な知識の獲得にある。以下では、そのようなマーケティング研究において有益な統計的手法とそのためのデータについて概観すると共に、その役割や限界について整理する。

第1節 マーケティング研究における統計的手法の役割

本節では、上述のようなマーケティング研究の目的を達成する上で、統計的な手法が如何なる役割を果たしているかについてまず簡潔に述べることにしよう。

第1に、マーケティング現象の記述とは、その克明な描写を意味する。それは研究対象となった現象が、どのようなものであるかを客観的事実として把握することである。つまり、記述の目的は、その現象に対する総合的理解を深めることであり、その副産物として、現象の発生メカニズムや因果構造に関する仮説が得られることもある。記述は文書（言語モデル）の形で行うことも可能であるが、その現象の特性を変数で捉え、評点、度数、比率、平均、分散等の数量として示すこともできる。

第2に、説明とは、研究対象となったマーケティング現象の構造や発生メカニズムを、

原因を特定した上で客観的に明らかにすることである。そのためには、対象となる現象を被説明変数（基準変数、または、従属変数）として捉え、その現象の原因と仮定される説明変数（予測変数、または、独立変数）との関係を統計的に明示する必要がある。これらの変数間の関係を示した数式（例えば、回帰式）は説明モデルと呼ばれる⁽¹⁾。以上のように記述と説明は、マーケティング現象が「どのようになっているか。また、それはなぜか。」という問いに答えるものである。

第3に、予測は、言うまでもなく、対象とするマーケティング現象の予測である。そのためには、説明段階で得られた統計モデル、時系列データを用いた予測のためのモデル、さらには、シミュレーション・モデルなどを利用する。例えば、既に得られている説明モデル（ここでは、回帰式）を用いる場合、独立変数に予測すべき時点のデータを代入することで、従属変数の値が予測できる。ただし、予測は単にマーケティング現象の未来変化を知るためだけでなく、導かれた説明モデルの時間的ないしは空間的妥当性をテストすることによって、より説明力の高いモデルを得るためにも行われる。ここで、説明を主たる目的とする場合には、できるだけ少数の独立変数で現象を説明することが望ましいのに対し、単に予測力の向上をめざす場合は、個々の独立変数の説明力よりもそれら全体としての予測精度の方が重視されるため、必ずしもそれらが少数である必要はない。

最後に、統制はマーケティング現象が「どうあるべきか。」という問いに答える。つまり、統制を目的とする研究は、マーケティング現象を最も有効に統制し得るような戦略を意思決定者（通常は、企業のマーケティング担当者）に与えるためのものである。記述、説明、予測が、マーケティングの基礎理論の構築を目指すとするれば、統制はマーケティング管理上の知識を追求するきわめて応用的なものと言える。ただし、統制は記述・説明・予測と独立に存在するのではなく、むしろ、そうした作業から得られた客観的知識に基づいてなされるべきである。

第2節 各種データの性質と入手法

統計解析に用いることのできるデータは様々である。しかし、統計解析にそのまま用いることができるのは、そうしたデータの一部にしか過ぎない。しかし、そのままでは統計解析の対象とならない質的データでも、量的データに変換して統計解析に用いたり、それがカテゴリカル・データであれば、そのための解析手法を用いたりすることもできる。さら

に、内容分析などの質的データの分析プロセスには、キーワードの集計という形でデータの数量的取扱が含まれている。本節では、各種データの特性と入手方法、それらに対する統計的手法の適用の問題について概観する。

1. 一次データ

調査者の何らかの意図に基づいて新たに収集されるデータを一次データと言う。一次データを入手するための方法としては、以下のような方法がある。ただし、収集されたデータが、すべてそのままの形で統計解析の対象となるとは限らない。質的なデータについては、何らかの方法で数量的なデータに変換したり、量的データであっても加工したりしてから統計解析手法を適用する場合もある。

実査（サーベイ）

一般にアンケート調査と呼ばれる。5点尺度などで測定された各質問項目は、それぞれ変数を意味する。多くの回答者から得られた諸変数についてのデータは数量データとして扱われ、統計解析の対象となる。ただし、性別や職業などカテゴリカルなデータは質的データであるため、分散分析などカテゴリカル・データを扱うための分析技法を採用したり、ダミー変数に変換して回帰分析に掛けたりするなどの工夫が必要である。また、自由回答による記述は、そのままでは量的分析の対象とはならないため、質的データのところで後述するような分析方法を取る必要がある。

深層面接およびグループ・インタビュー調査

新製品開発のヒントや隠れた消費者ニーズを把握するために、しばしばこの種の調査方法が用いられる。しかし、これらの方法で集められた情報は質的データであるため、統計解析の対象とはなりにくい。

実査（サーベイ）における自由回答も含め、定性的調査によって得られたデータをいかに客観的な情報に変換するかについては後述する。

観察

調査対象（一般には消費者）の行動を捉えデータ化する方法。その例としては、新規店舗の出店可能性を探るための交通量調査、小売店内の顧客の通行を把握し売場作りための知見を得る顧客動線調査、広告物や店内陳列物に対する消費者の視線の動きを記録するアイカメラによる方法などが挙げられる。小売店に蓄積されるPOSデータも一種の観察データである。これらはいずれも客観的に記録されたデータではあるが、消費者別、店舗別、

あるいは製品別といった具合に集計し直し、それらの標本を一定数確保しなければ、統計的分析の適用は困難である。

実験

実験室的な環境の中でデータを収集する方法と、実際の市場においてデータを収集するテスト・マーケティングや擬似実験的な方法とが存在する。まず、実験室的な方法の例としては、異なるパターンの広告を提示してその反応を比較する方法が考えられる。また、擬似実験的方法としては、異なる地域や店舗において異なるマーケティング対応（異なる製品の販売、異なる価格や広告の提示等）を行って、その結果を見る方法がある。この種の方法では、実験的刺激のあるなしが、0 ないしは1 のダミー変数として把握され、それが結果（売上や消費者の好意的反応度など）に如何なる影響をもつかを統計的に明らかにしようという発想が基本にある。

2 . 二次データ

これに対し、他者によって収集されたデータを事後的に借用して統計解析に用いる場合があるが、これを二次データと呼ぶ。一般に調査予算には限界があり調査の重複を回避するためにも、まず社内に蓄積されたデータや外部の研究機関・組織・マスコミ等によって公表されたデータを有効に活用しようとするのが自然である。具体的には、国勢調査や商業統計等の公的機関による調査データ、マスコミ・広告代理店・金融機関・保険会社等によるレポート、新聞ないしは雑誌記事、各種統計データベース、インターネットのホームページ等からの情報が利用可能である。

3 . 量的データ

統計解析に用いられるデータは、量的データがその中心となる。量的データとは、一般に加算が可能なデータのこと、後述する間隔尺度と比尺度がこれに含まれる。他方、質的データには、まず、名義尺度および序数尺度で測定されたものと、次に、言葉や文章（テキスト）として記録されたものが含まれる。名義尺度と序数尺度による変数は、量的変数に適用されるものとは異なるカテゴリカル・データのための解析手法が開発されてきた。また、類似性や選好など量的に把握することが困難な対象を、データ入手の段階から序数的な量として回答させ解析するための手法（例、多次元尺度構成法、コンジョイント分析）も存在する。

一般にデータはその測定の水準によって次の4つのタイプ⁽²⁾に大別され、特性の違いによって適応可能な解析技法も異なることが多い。

名義尺度

学籍番号や銀行の口座番号のように個々の対象を単に識別するための数値をもつものを指す。許容される統計量としては、事例数、最頻値などがある。性別や職業のようにそのままでは数値として取り扱えない質的データは、カテゴリカル・データと呼ばれ、性別を男なら0、女なら1という具合にダミー変数に変換することによって数量データとしての利用可能性が開ける。数学的性質としては同一性を示すことができるため、次のことが言える。

$$A=B, B=C \quad \text{ならば} \quad A=C$$

序数尺度

順位、優劣、程度等によって対象の序列を表現した数値。中央値、パーセンタイル、順位相関係数等の統計量が許容される。厳密には値の順序しか意味をもっておらず、数学的には同一性および順位性を示すことができる。ここで、順位性については次のことが言える。

$$A > B, B > C \quad \text{ならば} \quad A > C$$

$$A > B, B = C \quad \text{ならば} \quad A > C$$

間隔尺度

絶対原点はもたないが、尺度上の水準間の距離が等しくなるように数値が割り当てられた尺度。具体的には温度や年号のように間隔尺度同士の乗算・除算は意味をなさない。例えば、気温30度と15度の差は、20度と5度の差に等しいとは言えるが、気温30度は15度の2倍暑いとは言えない。間隔尺度においては、平均、標準偏差、ピアソンの積率相関係数等を算出することができる。また、次の比尺度と同様、量的データとして回帰分析等の多変量解析技法がそのまま適用できる。数学的性質としては、同一性、順位性、加算性が成り立つ。ここで加算性とは、以下のことである。

$$A = C, B = D \quad \text{ならば} \quad A + B = C + D$$

比尺度

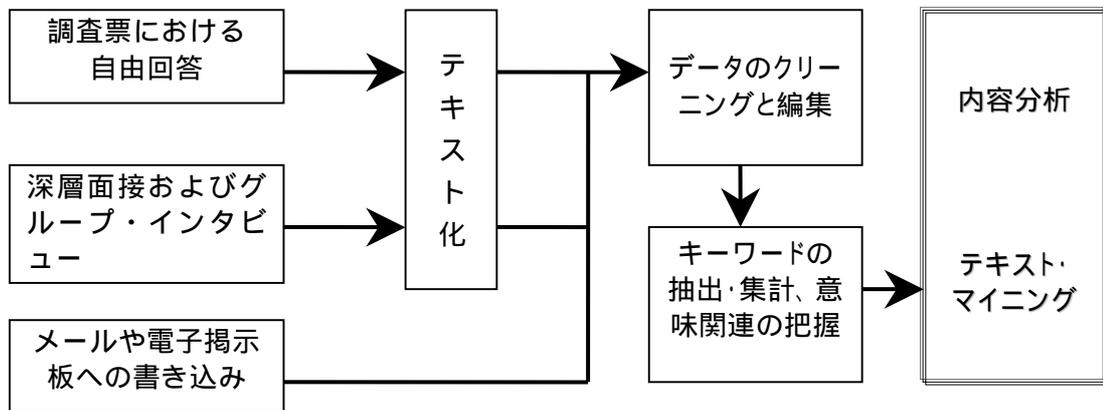
間隔尺度とは異なり絶対原点をもつなど、上記3つの尺度よりも測定水準のレベルが高いため、数学的には四則演算がすべて可能である。具体例としては、重さ、長さ、時間などである。したがって、対象を識別・分類することができるだけでなく、間隔・差異・平

均・比率を論じることも可能である。

4. 質的データ

名義尺度や序数尺度によって表現されたデータに加え、インタビューにおけるコメントや調査票における自由回答項目など、そのままでは数量として取り扱うことのできないテキスト・データを総称して質的データという。しかし、テキスト・データでも、その中に含まれるキーワード数を数えることによって、対象の性質を数量的に把握することも可能である。このようにして抽出されたキーワードの意味内容や関連性についても明らかにすることで、対象の深い理解を得ようとするのが内容分析と呼ばれる手法である。また、こうした作業をコンピュータによって自動的かつ迅速に行うための方法がテキスト・マイニングであり、最近ではウェブ調査による自由回答や電子掲示板に書き込まれたコメントなどに次第に適用されつつある。図1はテキスト・データと、その分析方法の関係について示したものである。内容分析やテキスト・マイニングは統計的手法とは異なるが、対象の客観的な記述や理解を得るための分析方法として用いることができる。

図1 テキスト・データとその分析方法



なお、質的データは、以下のような有用性をもっている。

調査者が予期しないような要因についても情報が入手できる可能性がある。

量的データの分析結果に対する理解が質的データの吟味によって深まることがある。

対象を記述・理解することで、新たな研究仮説を提供する。

特に常識的な仮説に対する反証事例について調べる場合、それが誤差によるものかあるいはその仮説を覆す必要があるのかの判断材料を与えてくれる可能性がある。

他方で、質的データおよびそうしたデータを入手するための定性調査には、以下のような限界や問題点があると考えられる。

グループ・インタビューや深層面接は、調査者の力量により分析結果が影響されがちである。

通常この種の調査においては、サンプル数が少ないため分析結果の一般性は低い。にもかかわらず、探索的分析結果や特異な事例に基づく解釈を、一般的なものとしてしまう恐れがある。

回答者の内面に踏み込める可能性は高まるが、反対に、プライバシー等の倫理上の問題が生じ易い。

5 . データにおける時間的性質

最後に、データの時間的性質について考えてみよう。データは、ある1時点で収集された横断的データと複数の時点で得られた縦断的データとに大別できる。横断的データは、ある変数について同一時点における個々の標本の値を示したデータであり、具体的には、ある時点で多数の標本から得られた質問紙調査における回答や、ある年度の商業統計の集計結果を都道府県別に示したものなどがこれに該当する。このタイプのデータによる変数間の関係は、基本的に相関に基づいており、因果関係を確定するための能力が弱い。しかし、縦断的データに比べ、データの入手が容易であることや予測よりも記述・説明を重視するといった理由から、マーケティング研究の分野では、横断的データを取り扱う分析技法の方が一般によく用いられてきた。

他方、縦断的データには、時間を規則的に追った時系列データや、同じ標本について異なる複数の時点で得られた反復測定データなどがある。特に後者は、同じ対象に対し定期的に行われる調査や複数回にわたる観察等によって得られる。縦断的データは、これまで、計量経済学、医学、教育学等の分野で用いられることが多く、時系列分析（伝達関数モデル等）コーホート分析、三層データを用いた因子分析など、そのための手法の研究も進んできている。この種のデータは、同一対象の時間的変化を追えるため、その変化の予測や、諸変数間に存在する因果関係の安定性の吟味を可能にする。今後、マーケティング分野でもスキャンパネル・データ等を活用した分析事例が増えてくるものと思われる。

第3節 統計解析手法の概観

1. 記述統計学と推測統計学

量的データを用いて、そのデータが表す統計集団の性質をできるかぎり簡単明瞭に記述・要約ないしは説明することを目的とするもの記述統計学と言う。具体的には、様々な統計表、グラフ、計算値などを使い、分布の代表性・散布度・形状を記述したり、要因間の関連性を示したりする。

これに対し、実験や観察により無作為に得られた数量データをもとにして、それらが抽出された母集団に関して推測を行うことを目的とした統計学を推測統計学と言う。現実の世界では、母集団全体の調査は時間的・費用的・技術的な理由で不可能な場合が多い。そのため、できるだけ母集団を代表するような標本を抽出することで、全体を調査したときと同様の結論を得るための方法が確立されてきた。標本（あるいは、サンプル）は母集団から抽出されたものであり、母集団のほんの一部に過ぎないため、そこから導出された推定には常に不確実性がつきまとう。したがって、この不確実性をコントロールしながら、できるだけ正確に母集団の特性を明らかにしようとするのが推測統計学である。なお、統計的推測には、仮説検定と推定という二つの柱がある⁽³⁾。

2. 多変量解析

調査などによって集められたデータは、通常、単純集計やクロス表分析によってその基本的な性質が調べられる。しかし、マーケティング現象はきわめて複雑であるため、データの記述だけでは、その現象について十分な説明や理解を得ることは困難である。例えば、ある製品の売上は、単に、製品の特性・品質だけでなく、価格、広告、販売経路など多くの要因によって説明可能である。しかも、それらの要因は、多くの場合、複合的な作用を及ぼしている。したがって、このようなマーケティングの現象を統計的に解明するには、ある1つの原因変数だけで説明を行うのではなく、複数の変数と、場合によってはそれらの変数間の交互作用をも同時に考慮する必要がある。加えて、全数調査ではなく限られた標本に基づいてその分析を行わなければならないことが多い。

多変量解析は、そうした目的に合致するものとしてマーケティングの分野でも広く用いられている。多変量解析の技法は多岐にわたり、それらすべてを1つの枠組みの下で類型

化するのにはなはだ困難である。以下に述べる多変量解析の諸技法は、データの記述・要約という面では記述統計学的側面を、また、分析結果に関し各種の検定が付随しているという点では推測統計学的側面をもったものと言える。

表1は、主に横断的データを取り扱うための代表的多変量解析技法について整理したものである。まず、被説明変数の解析技法のグループで、これらはいずれもマーケティング現象の解明にあたり、原因（説明変数ないしは予測変数）と結果（被説明変数ないしは基

表1 多変量解析の諸技法

被説明変数の有無	分析枠組み	分析技法名	分析の目的
あり： 被説明変数の解析技法	単一の被説明変数と説明変数群間の関係分析	クロス表分析 分散・共分散分析 重回帰分析 2群の判別分析 CHAID分析 コンジョイント分析	相関関係の統計的明示 要因効果の判定 要因効果の判定、被説明変数の予測 要因効果の判定、判別・分類 要因効果の判定、判別・分類 要因効果（重要度と効用値）の判定
	被説明変数群と説明変数群間の関係分析	多変量分散・共分散分析 正準相関分析 重判別分析 共分散構造分析	要因効果の判定 2組の変数群間の関連性の統計的明示 要因効果の判定、判別・分類 因果モデルのテスト
なし： 変数間の相互依存性に基づく解析技法	相互依存変数の分析	主成分分析 因子分析 共分散構造分析	変数の要約による変数の削減 潜在因子の抽出による変数の削減 確認的因子分析等によるモデルのテスト
	標本間の類似性の分析	クラスター分析 多次元尺度構成法（MDS）	標本の分類 次元の抽出

（出典）柳井晴夫(1994)『多変量データ解析法』朝倉書店、10頁および N.K. Malhotra (1999), *Marketing Research: An Applied Orientation*, Third ed., Prentice-Hall International, Inc. p.436 などをもとに筆者が作成。

準変数)とに峻別した上で、両者の関係を統計的に明らかにしたり、結果の予測を行ったりするものである。次に、変数間の相互依存性に基づく解析技法は、このような被説明変数をもたず、一組の変数群における相関や標本間の類似性に基づいて、変数群の要約や潜在因子の抽出を行ったり、あるいは、標本の分類を行ったりするものである。

第4節 統計的手法の限界と課題

前節まで、マーケティング研究における統計的手法の役割や適用法について述べてきた。統計的手法には様々な有用性と可能性が秘められていることは言うまでもないが、逆に、以下に示すような限界と課題があることもあらかじめ認識しておくことが重要である。

1. 統計的説明のための理論的仮説の必要性

例えば、無作為抽出によって選ばれたアメリカの消費者 2000 人を 2 度調べたところ、1 週間に 85% の人が鶏肉を食べたとする。この事実は単なる記述である。この事実をいくら吟味したところで、なぜアメリカ人が鶏肉好きなのかを説明したり、あるいは、今後、その割合がどのように変化するかについて予測したりすることはできない⁽⁴⁾。例えば、価格が安い(経済性)、健康に良い(健康に対する意識)、特定の人種によってよく消費される(デモグラフィック要因)といった原因と、鶏肉の消費という結果との関係が、何らかの論理に基づいて仮定され、それが消費者データによって確認されたとき、初めて統計的説明がなされたことになる。つまり、統計的説明の背景には何らかの仮説が必要なのである。もしそれがマーケティング理論研究の一環として行われているのであれば、その仮説は理論ないしはそれに基づく先行仮説から導かれたものであるか、あるいは、理論との関連性が想定されたものである必要がある。いずれにしても、調査者によって作成される調査票には、理論的仮説を構成する変数に関する設問が必ず含まれているはずであり、逆説的に言えば、説明の枠組みは調査者の発想に限定されていることになる。

なお、上述の例に戻れば、回帰分析などによって原因と結果の関係が推定された時、時間的、空間的に任意の新たなサンプルについての予測も可能になる。もし、結果(鶏肉の消費量)に関するデータのみが、仮に時系列的に揃っていたとしても、それだけでは消費量それ自体のトレンドで消費量を予測したに過ぎないことになってしまう。

2. 相関と因果関係をめぐる問題

一般的に多変量解析の諸技法は相関をベースにしており、必ずしもそれだけで因果関係を確証するものではない。近年開発された共分散構造分析は、因果関係の吟味に有効であるとされ、多変量解析の第2世代と呼んで、それまでの諸技法と区別する向きもある。しかし、それはそれとして理論構築の上で最も重要な因果関係の確証には、一般に次のような条件をクリアする必要がある⁽⁵⁾。

予測可能な共変動

変数間の関係が理論や仮説が想定した方向（正ないしは負）でなければならない。このことは、相関分析をすれば、容易にテスト可能である。ここで、変数間の相関を示しただけでは、因果関係の存在を示したことにはならないとの批判もあるが、逆に、相関の欠如は因果関係の欠如の有力な証拠にもなるのである。

事象の適切な時間的発生順序

因果関係が成立するためには、従属変数（被説明変数）の前に独立変数（説明変数）が発生している必要がある。

擬似的関係（他に考えられうる原因）の排除

今、仮に要因 A が要因 B の原因であることを示そうとするなら、A 以外の要因では B が発生しないことを示さなければならない。ただし、A 以外の潜在的要因を完璧に排除することは、複雑な社会現象を扱う場合、不可能に近い。したがって、統計的アプローチを取る場合には、実際には A によってどこまで説明できるかということと、A および B に共通の原因 C が存在しないかどうか議論されることが多い。

以上のような条件を認識しつつ、統計的手法を適用することは、科学的マーケティング研究を実践する上できわめて重要である。しかしながら、単に要因間の関係を統計的に示すだけは、科学的で理論的な研究であるとは言えない。なぜなら、要因 A が要因 B を引き起こすことが明らかになったとしても、それが経験的テストのために理論より導かれた仮説によって裏打ちされているか、あるいは、少なくとも理論との関連性が想定されていないければ、理論の構築を目指す研究とはならないからである。つまり、統計的手法によって浮かび上がった現象の説明には、理論的裏付けが必要である。そのため、統計的手法を適切に利用するためには、その現象を理解するための理論的考察が不可欠なのである。

3. 総合的判定の必要性

統計的手法は万能ではない。同じデータを分析しても、用いた分析技法、計算上のオプションの採用、さらには、結果の解釈の仕方などで、それが意味するところが変わってくることも多い。したがって、一度の分析結果によって仮説の適否を決め付けるのは無理がある。つまり、仮説が採択された場合は、暫定的にそれを保持することによって、真理に一步前進したと考え、また、棄却された場合には、仮説に問題があるのか、データに問題があるのかについて吟味する必要がある。また、常識的な仮説が棄却された場合には、むしろ、そこから何か新しい理論の糸口が見つかるかと信じて、分析を継続するというのも重要である。さらに、理論体系が未成熟なマーケティング研究領域においては、発見の文脈に力を注ぐことも重要である。そのためには、質的データとの補完的役割を前提に総合的な判定を行うことが重要である。

おわりに 要約と課題

統計的手法を用いた分析には、面白い発見がないとよく指摘される。確かに、常識的な理論を手持ちのデータに当てはめてテストしても、その理論仮説の確からしさが、確認されるだけで面白みはないかもしれない。しかし、未開拓領域の多いマーケティングの研究では、手探りの状態でその現象を記述・要約したりするという探索的研究も多い。正当化の文脈はもとより、発見の文脈において統計的手法を用いるというこのような作業は、複雑な現象の中に重複を見つけ出すことによって単純な構造としてそれを理解するという Simon の主張⁽⁶⁾にも通じている。したがって、仮説の構築や分析結果の解釈など質的データの補完的役割を加味することによって、統計的手法は、実り豊かな知見を今後も提供し続けるものと期待される。

(付記) 本稿は「マーケティング研究と統計的手法」『三色旗』第 634 号(慶應義塾大学通信教育部)14-19 頁(2001)を大幅に加筆修正したものである。

(1) ただし、説明の方法にはいくつかのパターンがある。Malhotra(1999), *Marketing Research: An Applied Orientation*, Third ed., Prentice-Hall International, Inc., p.51 が示すように、以下のモデルは皆、同じことを意味している。

言語モデル

「消費者はまずある百貨店の存在に気づく。すると、その消費者は自分の選択基準を構成する要因を評価することによって、その百貨店について理解する。そうした評価に基づいて、その消費者はその店舗に対する選好を形成し、もし、その選好が十分に強いものであれば、その消費者はその百貨店を愛顧するであろう。」

図式モデル



数学的モデル

$$y = a_0 + a_i x_i$$

ただし、 y = 選好の程度

a_0, a_i = 統計的に推定されるモデル・パラメータ

x_i = 選択基準を構成する店舗愛顧要因

- (2) ここでの記述は、次の文献を参考にしている。池田央(1993)『心理測定法』(財)放送大学教育振興協会、20-21頁。
- (3) 東京大学教養学部統計学教室編(1994)『人文・社会科学の統計学』東京大学出版会、7頁。
- (4) この事例は、以下の文献による。ポビー・カルダー(2001)『消費者理解』Dawn Iacobucci (ed.)(2001), *Kellogg on Marketing* / 奥村昭博・岸本義之訳『マーケティング戦略論 ノースウェスタン大学大学院ケロッグ・スクール』ダイヤモンド社、194-211頁。
- (5) 例えば、G.W. Bohrnstedt and D. Knoke (1982) 著 / 海野道郎・中村隆訳(1990)『社会統計学』ハーベスト社、341-344頁。
- (6) Herbert A. Simon 著(1996) / 秋葉元吉・吉原英樹訳(1999)『システムの科学 第3版』パーソナルメディア株式会社、219-256頁。