

11-開-05

データベース構築促進及び技術開発に関する報告書

POSデータに対するデータマイニング手法群の
比較研究と実用化技術の開発

平成12年3月

財団法人 データベース振興センター

委託先 株式会社日経リサーチ



この事業は、競輪の補助金を受けて実施したものである。

報告書訂正

報告書191ページ

(3) Apriori と同時購買信頼度について

		商品 B	
		買っていない	買った
商品 A	買っていない	a	c
	買った	b	d

(誤)
 Aを買ったとき、Bを同時に買うとされる信頼度 $= d / (c + d)$
 Bを買ったとき、Aを同時に買うとされる信頼度 $= d / (b + d)$

(正) ↓
 $= d / (\underline{b} + d)$
 $= d / (\underline{c} + d)$

まず上記のように縦軸が商品Aの購買状況、横軸が商品Bの購買状況のマトリクスを描いてみる。

AもBも同時に買われていない件数 $= a$
 Aは買ったがBは買われなかった件数 $= b$
 Aは買われなかったがBは買った件数 $= c$
 AとBを同時に買った件数 $= d$ とすると、

(誤)
 Aを買ったとき、Bを同時に買うとされる信頼度 $= d / (c + d)$ 、
 Bを買ったとき、Aを同時に買うとされる信頼度 $= d / (b + d)$ となる。

(正) ↓
 $= d / (\underline{b} + d)$ 、
 $= d / (\underline{c} + d)$ となる。

商品数が少ない場合は、電卓1個でも併買の信頼度が計算できる。

例えばカレーと福神漬の関係ならば

		福神漬	
		買っていない	買った
カレー	買っていない	500	150
	買った	300	100

(誤)
 カレーを買ったら福神漬も買う信頼度 $= 100 / (100 + 150) = 0.40 (40\%)$
 福神漬を買ったらカレーも買う信頼度 $= 100 / (500 + 300) = 0.125 (12.5\%)$ となる。

(正) ↓
福神漬を買ったらカレーも買う信頼度 $= 100 / (100 + 150) = 0.40 (40\%)$
カレーを買ったら福神漬も買う信頼度 $= 100 / (\underline{300} + 100) = 0.25 (25\%)$ となる。



11-開-05

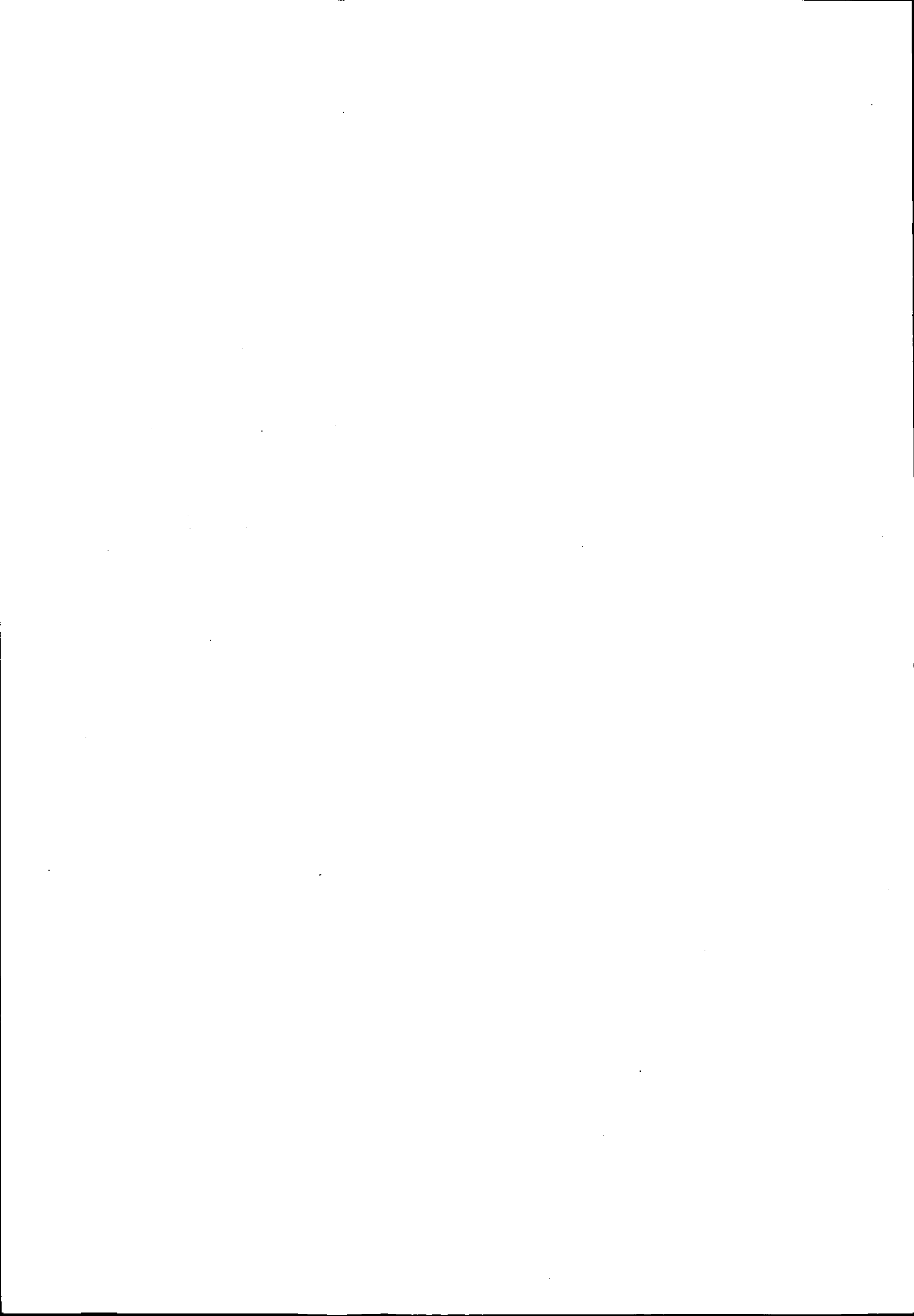
データベース構築促進及び技術開発に関する報告書

POSデータに対するデータマイニング手法群の
比較研究と実用化技術の開発

平成12年3月

財団法人 データベース振興センター

委託先 株式会社日経リサーチ



序

データベースは、わが国の情報化の進展上、重要な役割を果たすものと期待されている。今後、データベースの普及により、わが国において健全な高度情報化社会の形成が期待される。さらに海外に対して提供可能なデータベースの整備は、国際的な情報化への貢献および自由な情報流通の確保の観点からも必要である。現在わが国で流通しているデータベースの中でわが国独自のものは約半数であるが、わが国データベースサービスについてはバランスある情報産業の健全な発展を図るためには、今後もわが国独自のデータベースの構築およびデータベース関連技術の研究開発を強力に促進し、データベースの拡充を図る必要がある。

このような要請に応えるため、(財)データベース振興センターでは日本自転車振興会から機械工業振興資金の交付を受けて、データベースの構築および技術開発について民間企業、団体等に対して委託事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の発展の促進に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当って、当財団に学識経験者の方々に構成されるデータベース構築・技術開発促進委員会(委員長 東海大学教授 上條史彦氏)を設置している。

この「POSデータに対するデータマイニング手法群の比較研究と実用化技術の開発」は、平成11年度のデータベースの構築促進および技術開発促進事業として実施した課題の一つで、当財団が株式会社日経リサーチ に対して委託実施したものである。この成果が、データベースに興味をお持ちの方々や諸分野の皆様方のお役に立てば幸いである。

なお、平成11年度データベースの構築促進および技術開発促進事業で実施した課題は次表のとおりである。

平成12年3月

財団法人 データベース振興センター

平成11年度 データベース構築・技術開発促進事業委託課題一覧

区分	No.	課題名	企業名
一般	1	認定サービスにおけるデータベースの有効利用に関する調査研究	(株)シネジャーナルプロダクション
	2	ゲームソフトにおける知的財産権管理流通ビジネスに関する調査研究	京都リサーチパーク(株)
	3	地形データを効率よくデータベース化する技術開発	(株)セタ
	4	戦前期雑誌記事索引100万件の第一期データベース構築	(株)皓星社
	5	POSデータに対するデータマイニング手法群の比較研究と実用化技術の開発	(株)日経リサーチ
	6	北九州産業技術発展の歴史 記録の収集とデータベース化	(株)日鉄技術情報センター
地域振興	7	写真データベース構築	(株)琉球新報社
	8	バリアフリー施設データベースのプロトタイプ作成	(株)札幌ネクシス
	9	新居浜市機械産業における加工技術ノウハウ継承データベース構築	(株)四国インターネット
	10	インターネットによるラジオニュース公開型データベース構築	(株)山梨放送

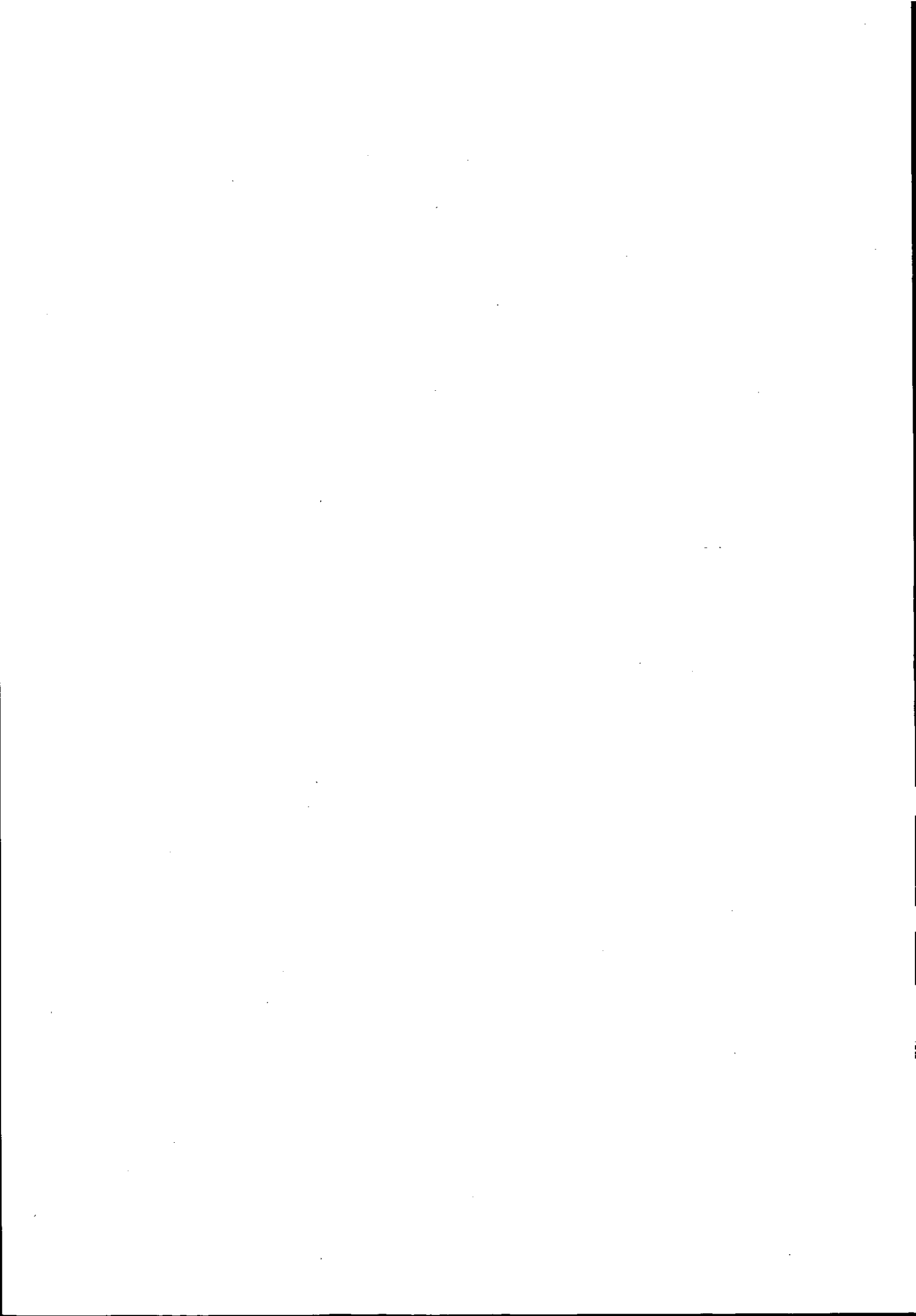
目 次

はじめに	1
1. データマイニングの可能性と概要	3
1. 1 データマイニングとは	3
1. 1. 1 データマイニングの定義	3
1. 1. 2 データマイニングの背景	4
1. 2 データマイニング経営の萌芽	6
1. 2. 1 英国の自動車保険	6
1. 2. 2 米国の流通業	6
1. 2. 3 日本の消費者金融大手	7
1. 2. 4 amazon. com社	7
1. 3 データマイニング実用化の前提 - 情報生態系	8
1. 3. 1 データマイニングに必要なもの	8
1. 3. 2 情報工場としての企業と情報生態系	8
1. 3. 3 データマイニングとマーケティングの現段階	10
1. 4 データマイニングの実行	13
1. 4. 1 データマイニング技術	13
1. 4. 2 データマイニング活用のポイント	14
2. 日本経済新聞社POSデータの概要	16
2. 1 日本経済新聞社のPOS情報	16
2. 2 NEEDS-SCAN	18
2. 3 NEEDS-SCAN/PANEL	21
3. 新製品立ち上がり売れ行き予測	26
3. 1 分析の目的	26
3. 2 分析の流れ	26
3. 3 データの定義とクリーニング	27
3. 3. 1 分析対象としたデータ	27
3. 3. 2 データの概要	28
3. 3. 3 データのクリーニング	30
3. 3. 4 ターゲットの選択	31
3. 4 事前の分析	32
3. 4. 1 成長性評価の指標	32
3. 4. 2 単純な相関分析	33
3. 4. 3 説明変数の評価と選択	34
3. 5 Decision Tree による分析	37
3. 6 まとめ	38

4. 新製品をヒットに導く要因は何か	40
4. 1. 目的	40
4. 1. 1 問題意識	40
4. 1. 2 分析の観点	40
4. 1. 3 分析結果の判断方法	41
4. 2 データ説明	42
4. 2. 1 データ概要	42
4. 2. 2 分析方法	42
4. 2. 3 使用データ詳細説明	42
4. 3 CHAID	46
4. 3. 1 大ヒットとなる要素を探る場合	46
4. 3. 2 失敗しないための要素を見つけ出す場合	50
4. 3. 3 結果の比較	56
4. 4 C5.0	57
4. 4. 1 大ヒットとなる要素を探る場合	57
4. 4. 2 失敗しないための要素を見つけ出す場合	61
4. 4. 3 結果の比較	72
4. 5 CHAID、C5.0による結果の比較	73
4. 5. 1 大ヒットとなる要素を探る場合	73
4. 5. 2 失敗しないための要素を見つけ出す場合	74
4. 5. 3 まとめ	75
5. ニューラルネットワークによる売れ行き要因分析	76
5. 1 使用データ	76
5. 2 データ変換	78
5. 3 分析手法	81
5. 4 分析結果	81
5. 5 まとめ	123
6. 牛乳販売量の予測モデル	125
6. 1 目的	125
6. 2 分析の方法	125
6. 3 分析データ	125
6. 4 予備解析	126
6. 4. 1 単純集計によるデータ概観	126
6. 4. 2 販売個数と販売金額の関係	128
6. 4. 3 販売個数の時系列変化	131
6. 4. 4 来店客数の変動	139
6. 4. 5 販売価格の変動	150
6. 5 マイニングツールによるモデル構成	150
6. 6 モデルの検討	159

7.	パネルデータを用いた購買行動の分析	161
7. 1	分析データの作成	161
7. 1. 1	パネルの選択	161
7. 1. 2	商品の選択	162
7. 1. 3	購買・非購買データの商品の定義と作成	163
7. 1. 4	商品属性の選択と作成	164
7. 1. 5	最終的なデータ形式	166
7. 2	分析	168
7. 2. 1	第1基準の作成	168
7. 2. 2	フェース (パネル属性) 分析	170
7. 2. 3	グループごとの購買行動分析	174
7. 3	手法の比較	179
7. 4	検討課題	186
8.	パネルデータによるバスケット分析	187
8. 1	分析方法概要	187
8. 1. 1	データ作成	187
8. 1. 2	分析手法	190
8. 2	データ分析	193
8. 2. 1	全体的な併買パターン	193
8. 2. 2	コーザルデータを追加した分析	197
8. 2. 3	バーチャルアイテムを活用した分析	199
8. 2. 4	改善率～相乗効果を上げるための指標	209
8. 3	多次元尺度法によるバスケット分析の簡便法	213
8. 3. 1	カレールーとの関連	213
8. 3. 2	全分類によるMDS	217
8. 3. 3	非対称MDS	219
8. 4	課題と発展	222
9.	まとめ	224
9. 1	マイニング手法と実行環境	224
9. 2	データマイニングを実施するための条件	225
9. 3	予備解析の重要性	226

参考文献・資料



はじめに

データベースには「蓄積する」側面と、それを「活用する」側面がある。前者はこれまでなかったデータを体系化して検索できるようにするという重要な意味を持つ。後者は溜め込んだ状態の“データ”を活性化して“情報”にまで育てあげる場面で重要である。

この委託課題は後者に属するもので、最近になって流行語の感もある「データマイニング」あるいは「KDD」(Knowledge Discovery in Databases)と呼ばれる技術の、POS データにおける実用化を目的とした。

データマイニングは大規模データベースから有用な情報(知識)を発見してビジネスに活用しようとするもので、Internet に象徴されるように情報通信分野の急速な発展によって、データが半自動的に大規模集積される事態の中で、それをどうやって活用したらよいか——という問題意識を背景にしている。POS データもまた日々刻々と蓄積され続けて大規模データベース化される典型的な例である。

日本経済新聞社は POS データの提供事業者でもある。今回の委託課題では全面的に協力をして頂き、実際のデータを使ったマイニング手法の研究をすることができた。データマイニングに関する参考書も出版されてはいるが、実際のデータを取り上げてマイニングしていく過程や失敗した例を説明している例は少ない。そこで、企業の抱える具体的問題を提示して、実際の POS データを使い、分析過程を省略せずに、結果として失敗した例も含めた「実例集」を公開することに意味があると考えた。企業内で実施され、発見された成功例はほとんどの場合に公表されないからである。

具体的な進め方として、以下のようなさまざまな立場(研究者、利用者)の専門家の方々に集まっていただき委員会を組織した。

氏名	所属
江原淳(委員長)	専修大学商学部 教授
片桐嘉信	日本経済新聞社
清水順也	ハウス食品株式会社
中林三平	株式会社金融エンジニアリング・グループ
茂木宏之	キッコーマン株式会社

また委員以外にも、立教大学教授の岡太彬訓氏には非対称多次元尺度構成法を適用する際にアドバイスをいただき、実際に岡太氏のプログラムを使って解析していただいた。

委員会は原則として毎月開催した。最初に、POS データのユーザーの立場から具体的問題について提言をいただき、その解決のためにはどのような手法が有効かを研

究者の立場からアドバイスする過程で、以下のようなテーマを選定した。

- 菓子新製品の発売初期で有望性やヒット性を判定するモデル
- バスケット分析による購買構造の探索
- 商品属性による購買パターンの探索
- 牛乳の日次販売量の予測モデル
- 焼き肉のたれの売れ行き要因

上記テーマの実際の分析作業は委員各位のほか、各分析スタッフが担当し、随時の小委員会で検討をしながら、毎月の委員会で分析結果を報告・議論した。その結果を持ち帰って再度分析するという過程を繰り返した。

データマイニングを実施するためにはソフトウェア(マイニング・ツール)を利用しなければ不可能である。現在ではいくつかのソフトウェアが提供されているが、この研究会で主に使ったマイニング・ツールは以下のソフトウェア会社製品である。

- SAS インスティテュートジャパン社、Enterprise Miner
- エス・ピー・エス・エス社、Clementine

マイニング・ツールのほかにも上記社製の各種ソフトウェアを利用して解析した。またこれらのソフトウェアの使用にあたって各社にはさまざまな協力をいただいた。今回の委託課題では、これらのツールの操作性に関しても報告をしている。

データマイニングには解析のセンスが重要であるが、ソフトウェア・ツールを効率的に操作する技術も求められる。また分析者が集中的に使える時間も必要であることが、今回の経験として得られた。限られた時間の中で作成した報告書なので、さらにマイニングを続ければ新しい「宝」が発掘される可能性がある。従って、今後とも継続して委託課題の成果を高めていく予定である。具体的にはPOSデータのユーザーへの事例紹介サービスや、ソフトウェア会社などが開催するセミナーなどへの参加によって、今回の成果を広く社会に還元するように努力していく。

1. データマイニングの可能性と概要

はじめに

データマイニングが流行している。その内実は全ての企業の全ての取り引き行為に関連しており、データマイニングなしにこれからの経営は考えられない。にもかかわらず、他の流行と同じく一過性のブームに終わってしまう恐れがある。

企業が、第三者や一般的な統計から学習する前に、自らの取り引きデータから学習する必要があるというデータマイニングの位置づけを中心に述べたい(*0)。具体的な手法の説明や使い方は、後述の文献などを参照されたい。

1. 1 データマイニングとは

1. 1. 1 データマイニングの定義

意思決定支援システム (Decision Support System, D S S) ・データウェアハウス (DWH) などと同様に、データマイニングは明確な定義をもたない。論者によって様々な意味が付加されている。多変量解析等でデータを様々な角度から分析することをこう呼ぶ人さえいる。

しかしながら、実務ではともかく、統計分析の領域ではそれは通常よく行われている作業にすぎない。単なる記述統計で何を騒いでいるのか？といった感想さえ聞かれる。

これらは、分析プロセスやデータ処理方法に偏った見方であると言わなければならない。データマイニングは環境認識をより正確に行い、顧客環境へのオファーをより効果的に行うための方法である。以下のような定義もある。

「意味のあるパターンやルールを発見するために、大量のデータを自動的にないし半自動的手段で分析し探索することがデータマイニングである。」(*1)つまり、

- ・ 目的は知識発見
- ・ 大量データを分析・探索する
- ・ 手作業でなくなんらかのツールを用いる

という特徴があるのである。ツールを用いなければ発見できないような複雑なパターンを目的としていること、組織内部の要約済みのデータでなく環境とのトランザクションデータのような大量データを対象としていることが分かる。単なる探索的データ解析 (EDA) や調査データのようなあらかじめ発見すべき知識が定まっている少量データの分析は、そう呼ぶべきかどうか議論はあるだろうが、少なくとも典型的なデータマイニングではない。

1. 1. 2 データマイニングの背景

データマイニングが注目されるようになったのには、様々な背景があるが、全てこの数年で成熟してきた条件である。従って様々なコンピューティングパラダイムがそうであったように、一時的流行で終わることは考えられず、どう呼ばれるかはともかくとして内容的には今後のデータ処理の核心的部分の一つがデータマイニングとなるであろう。

(1) 大量データ

企業間の取引の大半がコンピュータに入力されるか、ネットワーク上での取引となるか、いずれにせよコンピュータリダブルなデータを残す形となってきた。取引自体がオンライン処理の部分(O L T P)もますます増加している。

データが発生してもこれまでのように処理不可能でサマリーしてトランザクション自体は捨ててしまえば別に何もおこらない。しかしながら、取引で生じたデータはその取り引きについての様々な情報を含んでいる。

たとえば、取引先である小売業に、土日の販売促進策の提案を行う時、単なるアイデアや実験フィールドでの知見でなく、当該店舗のPOSデータに基づいていけば説得力も効果も違うであろう。ところが、取引先別に個店・日別・商品別にはデータが残されていないのが普通であった。POSでも、日別・店舗別・単品別のデータは、当初は全店合計の日別単品データや、個店別日別部門別金額データなどにサマライズされる事はあっても、何年も保存される事はなかった。データが取れる事とデータが利用できるようになってくる事との間はかなり隔たっていたのである。金融機関でも事情は全く同様である。

そこに光をあてたのが、データウェアハウス(DWH)である。磁気記憶装置の低価格化と、大量データを処理するための並列コンピュータと、それを可能とするオペレーティングシステムなどが現実化したのは1990年代後半である。物理的・技術的に、そのような設計をしておけば無理なく全てのトランザクションを蓄積する事ができるようになってまだ数年しかたっていないのである。データウェアハウスやその定型的部分であるデータマートのアーキテクチャは発展途上で、こちらも定義する事は困難であるが、大量のトランザクションデータを利用可能な形で保存・蓄積しておく事は共通している。

(2) データからの知識獲得圧力

チャネルを通じて不特定多数を対象にしていた時代と異なり、少なくとも直接の取引先については特定化されデータが残り個別のオファーが可能となってきた。ダイレクトマーケティングなど、最終消費者まで分かっている取り引きも徐々に増えてきている。企業間競争が、価格や製品だけでなく、標的市場の設定の仕方・それへのアクセス方法・マーケティングオファー全

体のあり方にまで及んできているといっても過言ではない。

ワントゥワンマーケティング、マス・カスタマイゼーションなどの考え方が広まりつつあり、関係性マーケティングとして顧客公衆とのリレーションシップ自体がマーケティングマネジメントの課題となりつつある。

このとき、現実の取り引きデータや顧客属性データからいかに効率的なマーケティングプランを生成するか（データベースマーケティング）、現在のビジネスのあり方自体をいかにデータの示すパターンに対応したものへと変革して行くか、といったことが決定的に重要となる。データを通じた環境認識能力・環境適応能力・知識獲得能力・その構造化能力が問題なのである。

換言すると、システムやネットワークやデータ自体から、それができたところでのそこからの環境認識能力・環境適応能力に競争上のフロンティアがうつりつつあり、データからの知識発見を具体的な顧客サービス・製品・販売方法・マーケティング計画などに活用することが課題となってきたのである。

このことは、データマイニング手法にも大きく影響せざるを得ない。経営上の意思決定につながる事がマイニングの目的であるので、分析手法の適切性と並んで、分析手法の伝達可能性・理解可能性が問われる。たとえばニューラルネットワークはもっとも強力なデータマイニング手法であるが、経営者は理解不可能なものにしたがって決断するわけには行かない。よりシンプルな決定木などが多用されるゆえんである。

(3) ソフトと分析環境の整備

並列マシンとデータウェアハウスの実現は、それを分析する環境の整備を促した。

これまでのデータ処理手法は、リアルタイムでOLAPとして行えるようになった。だが、複雑なパターンや知識を発見するには、OLAPツールでは不十分であり、不可能ですらある。

そこに、80年代の人工知能研究・非線形最適化手法の前進・知識情報処理の進展等をふまえたデータマイニングツールが商用のものとして出現した。コンピュータメーカー各社だけでなく、SAS、SPSSなど独立系ソフト会社もツールを販売している。

かつてはスーパーコンピュータと専門家チームを必要としたような分析が、安価なマシンと商用パッケージで簡単に行えるようになってきたのである。米国では、大手流通業・サービス業・消費財メーカーなどで、数十人規模の顧客分析部門をもち多数のデータマイニング専任者のいる企業がたくさん出てきている。②でみたように、それが経営の根幹に関わるからである。しかしながら、我が国では、少なくとも消費財関連では実務の中にデータマイニングプロセスがビルトインされている事例は存在しない。金融業や通販業の一部に例外的に見られるだけである。この問題点は後述する。

1. 2 データマイニング経営の萌芽

企業の環境との取り引き全てがデータとして残り、しかもその履歴データがDWHとして再利用可能となってきた。このことは、あらゆる経営意思決定やマーケティングアクションの効率性・効果が検証可能となってきたことを意味している。少なくとも同じ失敗は繰り返さなくて済むのである。それだけでなく、取引データの集合に隠れている交換(exchange)の要因を、データから類推していくことで、新たな交換の創造につながる知識発見とマーケティング技術開発とが可能となってきた。最もよくデータから学び実行できたものがもっともよい成果につながるのである。

そのような、トランザクションデータから次の交換の創造・促進を行う技術を中核としたビジネスが急成長している。いくつか例を挙げてみよう。

1. 2. 1 英国の自動車保険

自動車保険は他の保険と異なり毎年取り引きが必ず発生する。英国では、コンピュータ業界から1986年に参入した業者を中心に、10年で半分以上のシェアを電話による通販がとってしまった。情報公開で料率算定会に相当するデータが公開され、そのデータをデータマイニングし、何十%も安い個人別料率が可能となったからである。10年遅れでわが国でもこのプロセスが進行している。保険の他の分野でも、現在日本の全ての損保会社がC4.5など判別ツリーのパッケージを導入しているのは偶然ではない。(*2)

データマイニングで、新製品を作り出した事、それを新チャネルで販売した事、それに見合ったプランを実施した事など、データマイニングを契機にビジネスが一変してしまった先駆的な事例であろう。

1. 2. 2 米国の流通業

日本でこれから模倣しようとしている日曜雑貨・化粧品等H a B A (ヘルスアンドビューティエイズ) H B A (ヘルスアンドビューティケア)のサービスマーチャンドライザー(卸売業)と「トイザラス」などカテゴリーキラー(専門大店チェーン)が危機にある。食品スーパーがより粗利率の高いH a B Aを扱おうとした時、商品回転率の異なる、したがって意思決定方法の異なる商品についてマーチャンドライジングから販売促進・陳列までノウハウをもたなかった。そのノウハウを商品とセットで供給するサービスマーチャンドライザーは急成長し、コミュニティショッピングセンター(C S C)に不可欠のコンボス業態(食品スーパーにドラッグストアを併設した業態)をサポートしたのである。しかし、スーパー自体が情報処理能力をつけ、企業内でH a B Aのマーチャンドライジングからノウハウ開発・蓄積まで行えるようになり、サービス中心の卸売り業を不要にしてしまった。

専門品は深く広い品揃えが要求される事から、郊外型の大規模な専門品チェーンが急成長した。玩具の「トイザラス」はその典型である。しかしながら、「ウォルマート」などディスカウントストアが、それぞれ専門的情報処

理能力を身につけてそれらの業態の一部をを不要にしつつある。同程度の情報処理能力で同程度の売場を作ることができれば、価格競争力があるからである。化粧品・おもちゃ・家電等の専門品のマーチャンダイジングから販売ノウハウまで、従来経営組織内で獲得・維持されてきた知識に、それをもたない量販店でも情報処理能力とデータマイニング技術があればキャッチアップできることが明らかとなってしまった。データマイニングなど情報処理能力が業態間競争・企業間関係を変えてしまっているのである。

1. 2. 3 日本の消費者金融大手

赤字に苦しむ大手銀行を尻目に大手消費者金融会社は数百億の利益を上げ続けている。これは与信管理における情報処理能力の差である。消費者金融では申込書のデータをエントロピーによる判別ツリー等で分析し、焦げ付かない確率の高い消費者属性のパターンを識別している。単純な顧客属性別ポイント積み上げ制しかやっていない銀行とは雲泥の差である。少なくとも、対消費者への無担保融資の与信技術では、消費者金融業者は全ての銀行よりはるかに優れていると言わざるを得ない。データマイニングが業務効率を左右している典型といえよう。

1. 2. 4 amazon.com (アマゾン・ドット・コム) 社

アマゾンのCD・書籍通販の伸びはとどまるところを知らない。調達・物流システムの整備もすばらしいが、何か本を検索すると「その本を買った人が、他にどんな本を買っているか」書名を数冊提示し関連購買を喚起している。古いCDを買えば、その同じ年の他のヒット曲のCDを提示する。消費者の意思決定のパターンに合わせて、トランザクションデータを意思決定支援情報として要約した形で提示することで、次の交換を創造し促進しているのである。また、購買を中止しても、次回アクセス時に顧客識別技術により同一顧客の前回の中止された注文内容を提示する事すらしている。顧客が自己の購買履歴をどのようなスキームで振り返るかという点に絞って、技術的なあるいは運用上の困難さにもかかわらず優れたインタフェースを実現している。このインタフェースの優位性は、デザインの力ではなく、履歴データのマイニングの力である。

このように、情報技術が市場を変え競争条件を変えてしまうことは、よく知られてきている。しかしながら、サプライチェーンマネジメント(SCM)に典型のように、ネットワーク化による情報共有と中間組織の再設計・ネットワークの経済性の実現に注目が集まっており、肝心のデータをどのように処理していくべきかについては試行錯誤の域をでていない。現在わが国で急成長しているドラッグストア・ホームセンター・軽衣料等のチェーンは、限界はあるもののその両者がある程度は実現しているからこそ、量販店を凌駕できていることを忘れるべきではない。マイニングなどデータ処理技術とビジネスプロセス改善との統合こそが現在必要とされているのである。

1. 3 データマイニング実用化の前提－情報生態系

1. 3. 1 データマイニングに必要なもの

マーケティングリサーチは、環境の一部を刺激してその反応データから母集団推計することで環境の不確実性を削減してきた。未知事象についてのリサーチは、定性調査技法を中心に今後とも活用されていくであろう。しかしながら、情報化はマーケティングアクションの対象の特定化をもたらした。チャンネル政策などは一般的傾向から論じるのではなく、取引先別の費用対効果比から取引先別管理・セグメンテーション別管理へと移行せざるを得ない。ABC (Activity Based Costing、活動基準原価計算) とそれに基づく管理 ABM (Activity Based Management、活動基準原価管理) は、原価計算手法ではなくこの取引先別・顧客環境のセグメント別に利益管理していくための手法の一つとみなすことができる。

どのような管理手法も、データがなければ実用化できない。したがって、データ収集・処理方法が管理手法とセットで問題とされてきた。現在では、環境との全トランザクションを保存することも不可能ではない。そのような粗データの倉庫をDWHとして実現していれば、様々な管理手法・分析手法と結びつけることができる。

たとえば、流通業の物流センターのデータから絶対単品別のコストも計算し、DPP (Direct Product Profit) を求めてマーチャンダイジングに活かそうとしても、物流センターでは金額ベースのデータしかない、もしくは数量ベースがあっても配送単位が不明等々、そこでの作業でしかデータ入力されていないことが普通であった。このような個別システムをいくら統合しても、全体としてのビジネスプロセスは改善できない。組織は環境と情報を代謝する存在であり、やりとりした情報は全てデータウェアハウス化すること、その一部で用途の決まっているデータはデータマート化すること、意思決定や判断の基準となるデータは属人的に持っていないで履歴参照データとしてデータベース化すること…等が必要となる。

1. 3. 2 情報工場としての企業と情報生態系

W. H. インモン (W. H. Inmon) は、業務処理・管理・ビジネス知識の三局面を考えたコーポレート・インフォメーション・ファクトリー (CIF) を提唱し、企業は情報生態系としてCIFを構築・管理すべきだとしている (*3、図 1-1、2)。企業が日々外部の市場と直面しているのはマーケティングと財務であり、CIFとDWHのかなりの部分がマーケティング上の問題解決に使われるであろうことはいうまでもないであろう。自社のマーケティングアクションとその時の状況と成果とがDWH化され、その時の判断基準もCIFから獲得できれば、企業のマーケティング意思決定は飛躍的に改善する。少なくとも同じ失敗の反復は回避でき、DWHからの知識獲得 (KDD: Knowledge Discovery in Databases) で徐々に前進するこ

とが可能であるので。

DWHは成果管理からプロセス管理への移行にも不可欠である。各プロセスでのトランザクションの履歴がなければプロセス改善はできない。新製品の売り上げ記録は全ての企業が持っているであろう。しかし、全ての過去の新製品についてその導入計画・販促計画・トラッキング結果等を整理して持っている企業は極めて希である。そもそも何をやり何をやらなかったからその結果となったのか、未知のままマーケティング上では同じ失敗を繰り返しているのである。

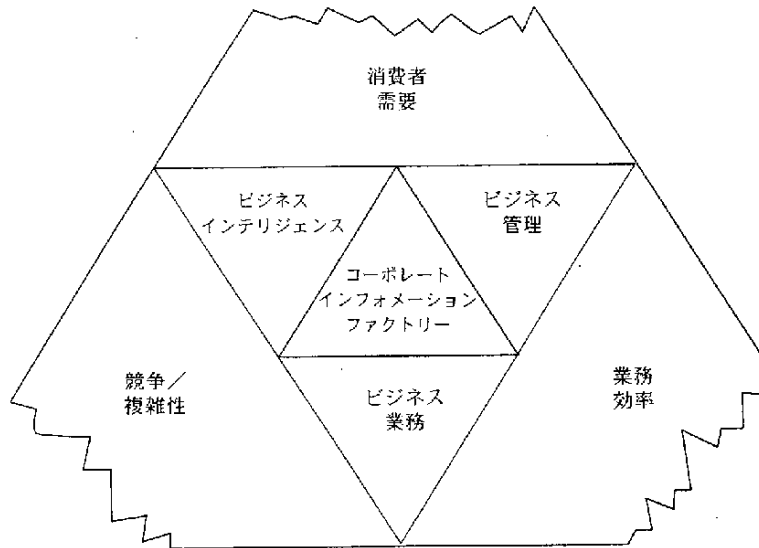


図 1-1 ビジネスとそこに必要な能力の中に CIFがある (出典*3)

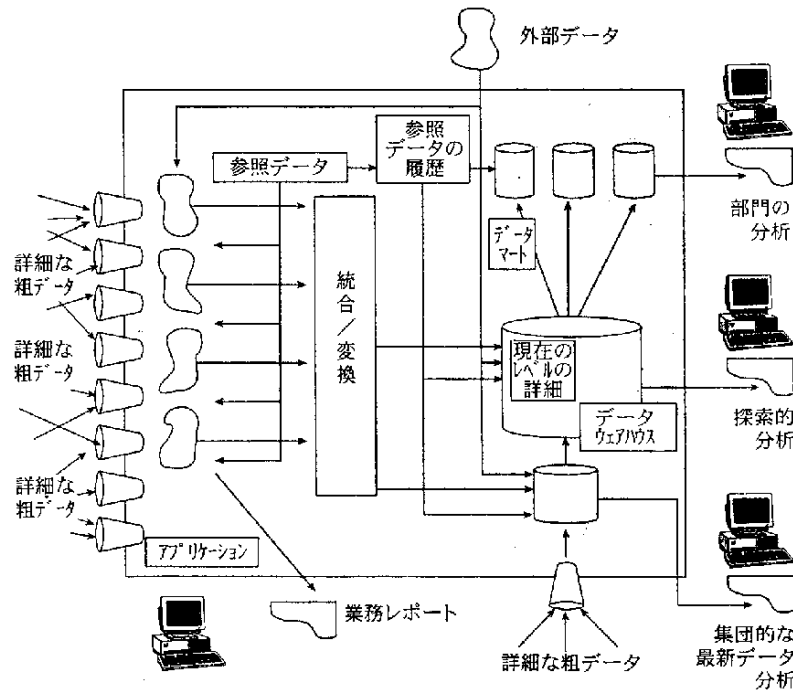


図 1-2 CIFの基本構造 (出典*3)

1. 3. 3 データマイニングとマーケティングの現段階

(1) マーケティングとSFA

よく知られているように、日本の「営業」は、マーケティングとセリングの融合されたものである。営業マンは取引先別のマーケターとなって個別提案し成果を達成する。このとき、SFA(Sales Force Automation)も米国とは別のマーケティング的性格を持たざるを得ない。SFAでは以下のような点が課題となる。

- ・ エリア管理・テリトリー管理
- ・ 取引先管理
- ・ 販売計画
- ・ 文書自動記入
- ・ チーム営業支援
- ・ 参照・引用資料データベース
- ・ 生産性分析ツール
- ・ 発送文書の一括処理
- ・ 電話での販売
- ・ 営業サイクルの管理
- ・ 営業方法
- ・ 予測

これらに加えて、日本の営業組織では取引先別のマーケティングプラン代行作成支援とか、営業成功事例交換とか、他業種成功事例とか、本来スタッフ部門で扱うべきデータを現場の営業マンがマーケティング上使用している。グループウェアを導入するとまず成功事例交換のBBS(ブリテン・ボード・システム)を始めたりするが、うまく動いている事例はほとんどない。属人的営業から、DWHとそこからの知識獲得を前提とした情報化営業として位置付いていないからである。取引先別・地域別・店舗別・顧客別等のマイクロマーケティングの枠組みなしに、属人的営業スタイルのまま情報活用するのは不可能に近い。

つまり、マーケティング機能を含んだ営業を維持するには、DWHの標準化とデータマイニングの一部の手順を標準的に装備せざるを得なくなってきたのである。

(2) ワンツウワンマーケティングとCRM

マーケティング効率が、全ての消費者に対して同一のマーケティング手段を採用するよりも、同質的な細分化市場ごとに別の手段をとるほうが高いことが多いことはよく知られてきた。顧客データベースの整備と個人別アクセス手段の充実に伴い、このマーケットセグメンテーションを超えて、一人一人の顧客に対して異なるマーケティング手段を採用するワンツウワンマーケティングの考え方が広まってきた。交換の意思決定者である顧客との関係

性を重視し、顧客内でのシェアを高め、顧客の生涯価値（Life Time Value、LTV）に注目する手法である。

DWHからの一人一人の購買履歴データ自体を顧客属性として扱い、その過去のデータから将来の購買確率を個人別に予測することがベースとなる。現実には、類似履歴・関係性ごとの対応となることが多く、市場細分化との区別が曖昧であるが、市場細分化は市場構造と消費者属性から作成されるのに対し、ワンツウワンマーケティングでは一人一人の購買履歴をベースにしているところが全く異なっている。態度でなく、行動レベルのデータ、調査でなく取引のデータからスタートしているのである。

つまり、市場細分化はマーケティング計画の課題であるが、ワンツウワンマーケティングはビジネスプロセスとDWHないしデータマイニングの課題なのである。より広義には、CRM (Customer Relationship Management) が課題となり、顧客と接触した全ての情報から顧客ごとの需要を予測し遺失需要を減らし、LTVを高めることが目標となる。ここでも、データマイニングやシミュレーションなど具体的な情報処理技術がポイントである。店頭、ダイレクトメール、コールセンター、eメールやwebなど、個人別にアクセスできる手段が増えるにつれてこのような技術の比重は増さざるを得ない。そして、長期的にはマーケティングのあり方自体にも影響してくるものと考えられる。

顧客との関係性を管理していくという中には、それを通じて組織が顧客と市場を学習していくという側面が存在する。ネットワークとデータベースと情報システムを通じた相互学習関係は、長期的には生産と消費のあり方に大きく影響してくるであろう。

もう一度言うと、ワンツウワンマーケティングやCRMでは、データマイニングは不可欠の業務プロセスであり、必要なのはその技術であって哲学ではないのである。

（3）データマイニングとプロモーションの変化

FSP (Frequent Shoppers Program) では、航空会社のマイレージサービスのように、顧客環境を自社にとっての価値で階層化し、よりよい顧客にはよりヘビーなプロモーションを行うことで顧客を固定化していく。これは、販売促進のあり方としては画期的である。

これまでチラシ特売が典型のように、誰にでも一律に販売促進費用を還元していた。むしろバーゲンハンターのように、来店頻度の少ない、または来ても特売品しか購入しない顧客に費用をつぎこみ、本来その原資を提供している優良顧客にはほとんど還元がないという状態であったといっても過言ではない。この状態は顧客識別のできない時代の産物であり、顧客のもたらず利益に応じてフェアに顧客を遇することができず単純に平等に扱うことしかできなかった時代の産物である。プロモーションは一回ごとの特売の成果のために行うのでなく、長期的な店舗愛顧やロイヤルティ形成、顧客のL

TV向上を目的として政策的に行わなければならない。

支払い手段をめぐるクレジットカード、デビットカードや電子マネーも、手段の利便性の提供だけでなく顧客識別による顧客ごとのプロモーションと組み合わせて考えていく必要がある。

ここで重要となるのは、マーケティングリサーチでなくマーケティングエクスペリメントである。顧客環境の一部に別のオファーを行い、その反応関数を特定化する。それによって顧客集合全体への費用対効果比をシミュレーションする事となる。本来価値のない空席を提供している航空会社のFFP（フリークエント・フライヤーズ・プログラム）とは異なり、FSPではプロモーション費用が非常に高い。データマイニングで費用対効果の高い方法を見いだした流通業者のみが長期的に提供でき、技術を持たずに方法だけ模倣するものにとってはほとんど利益のでない競争状態がすぐそこまで来ているのである。プロモーションの計画・実行・統制は、基本的にデータマイニングとならざるを得ない。

（４）WEBマーケティング

ECより広義に、WEB技術をマーケティングやマーケティングリサーチに活用する動きが急である。そこでは、次のような方法が行われている。

- ・ WEBサイトの顧客別カスタマイズ、前回履歴を反映したサイト活用
- ・ eメールやメーリングリストによる情報サービス、パブリシティ、ダイレクトメール
- ・ プッシュ技術
- ・ ユーザーのコミュニティ形成
- ・ WEB広告とそのカスタマイズ
- ・ トラフィック分析・ログ分析
- ・ データベースインテグレーション

（たとえば、前述の amazon.com のサイトでは、前回買い物かごに入れたがチェックアウトしていない商品を次回にも提示しているように、未処理情報・配送情報・顧客の評価情報・支払い情報等を全て顧客側から統合して見えるようにしている）

今後更に、インテリジェントエージェント・コラボレーティブフィルタリングなどの新技術が取り入れられていくであろう。しかし、バナー広告一つとってみても、情報検索プロセスでそれと無関係な広告が提示されてもほとんど効果がないといった消費者の情報処理プロセスや意思決定スキーマの基礎研究がほとんどない。現在のPCのウィンドウシステム・GUIのベースとなった基礎研究(*3)等を欠いたままでは、マーケティングツールとして不十分である。それ以前に、サプライサイドで作られたホームページが大半を占めている現在、アクセスログ分析に基づくより効果的なサイト設計が急務であろう。

この領域でも、トラフィックの分析に比して、顧客の選択パターン、クリックスルーパターン、週間アクセスバスケット、サイト間関連性分析など、顧客に関連した分析はほとんど手がつけられていない。代金決済が標準化されつつあるので、情報にふれるきっかけ、情報探索、購買プロセス・代金決済、アフターサービスまで、多様な機能間の顧客のパターンをいち早く見出し対応したものが果実を手に入れることとなろう。

1. 4 データマイニングの実行

1. 4. 1 データマイニング技術

これまで見てきたように、社会的にデータが爆発的に増えており、データ分析をふまえたビジネスプロセス改善で効率的・効果的になる領域は大変多い。しかしながら、そのデータは現実の顧客や市場を反映したものであるので、簡単な統計的モデルで記述できることはまれである。とりわけカテゴリカルなデータのモデル化では、説明変数が階層構造など何らかの構造を持っている場合手法がなかった。(せいぜい対数線形モデルに交互作用変数をいれることくらいしかできなかった。)

そのような複雑なモデル化を行う手法としてデータマイニングが脚光を浴びている。(※4)

- ① ニューラルネットワーク
- ② 決定木 (χ^2 による CHAID, ジニ係数による CART, エントロピーによる C5. 0 等)
- ③ アソシエーションルール
- ④ 記憶ベース推論
- ⑤ k-means 法クラスタ分析
- ⑥ リンク分析
- ⑦ 遺伝子アルゴリズム (GA)
- ⑧ テキストマイニングの諸手法
- ⑨ マーケットバスケット分析

など、様々の手法が提唱されている。基本的には記述統計量とそれによるモデル化であったり非線形の記述モデルである。多くは隣接分野からの借用である。

手法選択では、RDBとの親和性(OLAPとマイニングは連続的であるため)

- ・ユーザが得られた知識やルールをメンテナンスできること
- ・結果が自然言語で伝達可能であること
- ・欠損値やノイズ処理ができること
- ・得られたパターンが理解可能であること
- ・統計的有意性が評価できること

等に留意しなければならない。論文を書くためにではなく、経営上の意思決定のために分析しているのであるから。したがって、最も強力な手法であるニューラルネットワークが、理解可能性・伝達可能性が低く最も使いづらいのが現実である。

1. 4. 2 データマイニング活用のポイント

米国での金融業・流通業のように、数十人規模のデータマイニング・顧客モデリング専門部門をもっている企業では、すでにマーケティング計画のあらゆる分野でデータマイニングがインフラとなっている。計画と実行とが分離されている職務分掌では、計画を高度化する不可欠のプロセスなのである。

しかし、両者が渾然一体となっている日本企業では、テスト段階を超えているところはまれである。

これまでのところ、POSデータからのショッピングバスケット分析・スコアリングなど通販での顧客のセグメンテーション・FSPなどプロモーション設計・金融業におけるALMやリスク分析・クレーム発生頻度予測と事前対策・定性調査や購入者アンケートでの自由記入回答のテキストマイニング等々、様々の領域での断片的な事例が報告されている。

しかしながら、テスト的な分析はデータマイニングとは言えない。DWHからの知識発見と新たなビジネス上・マーケティング上のプランに活かされなければ、単なる統計的処理にすぎない。

わが国における活用の遅れの原因は、実は手法やソフトにあるのではなく、経営組織自体に内包されている。データマイニングは、よりよいビジネスルールを発見するために行われる。データからルールを知識発見し、より効果的なビジネスをするのである。既存ビジネスルールとそのために必要なメタデータが明らかでない、データに問いかけるべき課題をもてないのである。たとえば、消費者の意思決定しやすい陳列を作成するという課題に対し、消費者はどのようなサブカテゴリー内で購買しているか・陳列のスペース弾力性をどう扱うか・陳列場所の効果をどう扱うか等についてのビジネスルールがなければ、解は求まらない。そのルールがあって初めて、商品ごとの陳列フェイシング数を含むデータベーススキーマ・スペース弾力性というメタデータから最適陳列方法というビジネスルールが得られるのである。POSデータがあれば最適な陳列が作成できるのではなく、ビジネスルールをPOSデータで検証し新たなルールを加えて初めて作成できるのである。

わが国で問題なのはデータではなく、共有されたビジネスルールなき属人的管理水準であるといわなければならないであろう。今日のデータ状況は属人的処理の範囲を遙かに超えてしまっている。DWHとデータマイニングを中心にすえた顧客データ駆動型の組織設計から始めない限り、データマイニングに基づく創造的なマーケティング活動は望むべくもない。

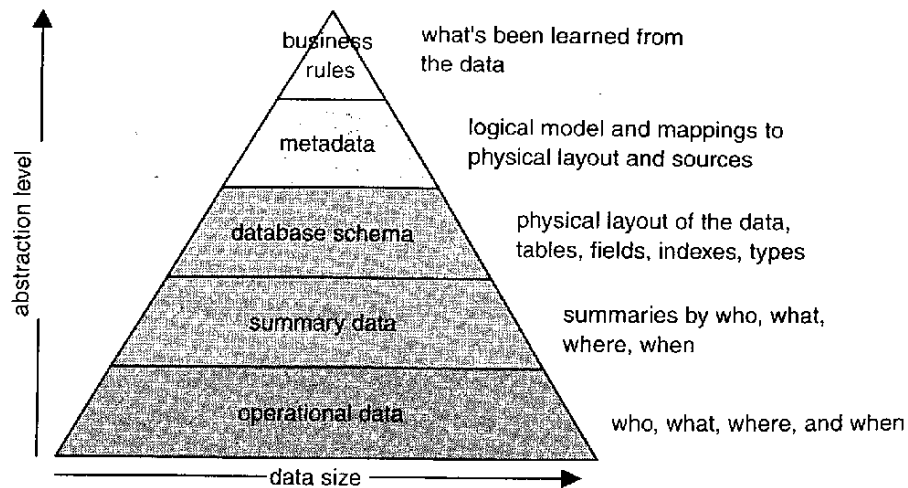


図 1-3 ビジネスルールとデータ (出典*4)

- (*0) 江原淳「マーケティング革新とデータマイニング」「品質管理」2000.2月号。本稿は、これを加筆・修正したものである。
- (*1) 上田和勇 「イギリスの保険マーケティング」, 1998.
- (*2) Inmon, W.H. et al., *Corporate Information Factory*, Wiley, 1998.
江原 淳他訳「コーポレートインフォメーションファクトリー」、海文堂、1999。
- (*3) Card, S.K. et al., *The Psychology of Human-Computer Interaction*, LEA, 1983.
- (*4) Berry, M.A. and L. Gordon, *Data Mining Techniques*, Wiley, 1997.
SAS・江原・佐藤訳「データマイニング手法」、海文堂、1999。
- (*5) Fayyad, U.M. et al., eds., *Advances in Knowledge Discovery and Data Mining*, AAAI, 1996.

2. 日本経済新聞社POSデータの概要

2.1 日本経済新聞社のPOS情報

日本経済新聞社は、1985年1月に米国のPOS情報サービス企業として最大手のIRI (Information Resources, Inc.) と技術提携して、日本で最初の商用POS情報サービス「NEEDS-SCAN」のデータ蓄積を開始した。

当時は、POSレジスターの導入店も非常に少なく、1987年初に収録していたのは6チェーン24店舗に過ぎなかったが、1990年初には25チェーン47店舗となり、1995年初には33チェーン103店舗、2000年1月現在で46チェーン235店舗と収録チェーン数・店舗数共に着実に増加している。

これらの235店舗には1日平均80万人の来店客があり、800万個もの商品が購入されていく。1年間このPOSデータを蓄積すると、実に6000億円のデータベースとなる。日本チェーンストア協会加盟の119チェーン、7376店舗の年商が約16兆6千億円であり、その3.6% (同店の食料品、日用雑貨、医薬・化粧品に限れば10兆5千億円の5.7%) に当たる購買が单品別に捕捉できる国内有数のPOSデータベースとなった。

また、この日本経済新聞社のオリジナル収録POSデータとは別に、財団法人流通システム開発センターが収録し、再販業者へ提供するRDS (流通POSデータサービス) POSデータも導入し、蓄積・データベース化している。RDS・POSは現在420店舗の中堅スーパーマーケットが中心で、420店の年商は3000億円となっている (図2-1)。

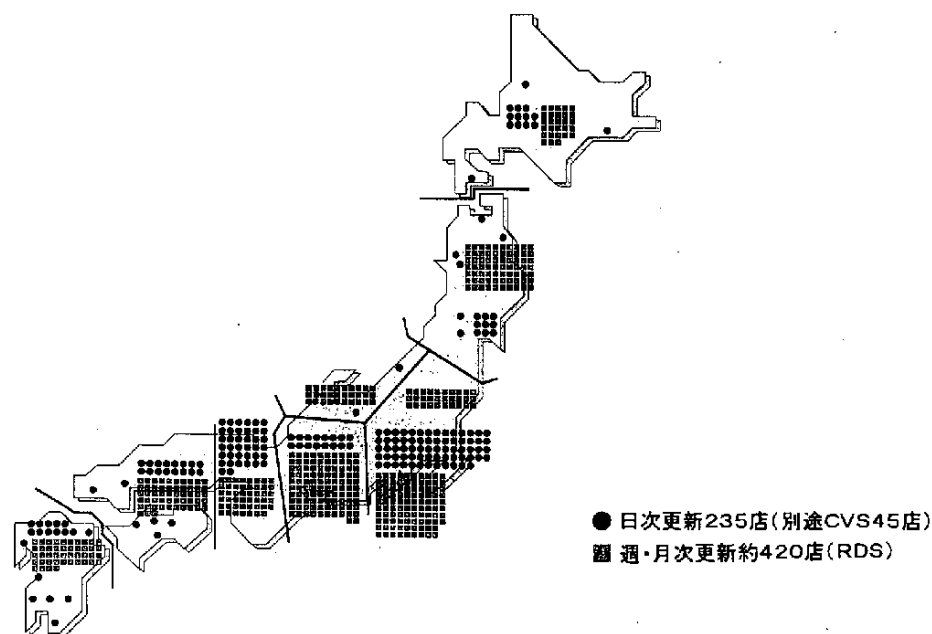


図2-1 日本経済新聞社POSデータ・RDS・POSデータ収録地域

これらのPOS情報では、「どこの店で」、「どの商品が」、「単価いくらで」、「何個購入されたのか」については99%以上の精度で読み取ることができる。

さらに、POSデータからは判明できない「誰が購入したのか」というデータについては、日本経済新聞社では前述したIRIとの提携で得たノウハウを活用して、購入者のプロフィールのわかるデータベース「NEEDS-SCAN/PANEL」を1987年から収録し続けている。パネラーとなっていた世帯にIDカードを渡し、指定の店舗での支払い時にそのIDカードをPOSレジスターで読み取り、その後買い物全品をスキャンして決済すると、その顧客の当日の購入品がすべて記録される。消費者の「顔」と「購入の場面」が見えるこのPANELデータは現在関東地区の2地域に設定され、2500世帯の協力のもとに13年を越えるデータの蓄積ができた。

もう一つPOSデータ収録で重要な点は、「商品マスター」の作成方法にある。日本経済新聞社では、この地道なマスター作りを外部のマスターに依存せず、すべてオリジナルの「日経商品マスター」として運用し続け、15年間で280万商品を収録している。この日経商品マスターは店頭第一主義を基本としている。上記235店舗の店頭で毎日登場する新製品（新しいJANコード番号）は1日平均で200~250アイテムになる。「バレンタインデー」直前や春・秋の棚替え、「クリスマス」「年末おせち」の季節には1日当たり1000種類以上の新製品の登場する日も珍しくない。1年間では9万種類にのぼる新製品が発売されては消えていく。この新製品は店頭調査員がそのデータの発生した店舗でJANコードを頼りに、一品ずつ発見し、正確な商品名称、内容量、分類（カテゴリー）を登録していく。POSデータの発生と商品マスターメンテナンスが一元化された、生きているシングル・ソースデータである。

過去から最新まで整合性のとれた日経商品マスターはカテゴリー毎の商品管理やデータ分析に欠かせないものである。また、15年間に10回のカテゴリー変更を実施してその時代に合致した分類方法を常に追求しており、2月1日現在、食品は152の大分類、その下に1192の小分類、家庭用品は59の大分類、その下に458の小分類に分けて、日経独自の方法で運用されている。

また、日経商品マスターは、店頭第一主義により他社にはなかなかできないPB（プライベート・ブランド）商品の収録や、海外からの輸入商品（ユニバーサル・プロダクツ・コード、通称UPCと呼ばれる米国、カナダの商品と、ヨーロッパ・アーティクル・ナンバー、通称EANと呼ばれる米国、カナダ以外の全世界からの商品がある）も収録している。また、単品別に、初めて販売された日と一番最近販売のあった日、同一JANコ

ードを別の商品に付番・使い回しに対応するための商品毎の世代管理など、他社のマスターにない特徴を備えている。

次に、今回POSデータのデータマイニング分析用に利用したそれぞれのデータについての詳しい説明にはいる。

2.2 NEEDS-SCAN

日本経済新聞社オリジナルのPOSデータベースNEEDS-SCANに現在収録されているのは、大手量販店46チェーン235店舗（この他にコンビニエンス・ストアが45店舗ある）であるが、最大の特長はすべての店舗毎に1日毎のデータをオンライン収集している点にある。他のデータと比較してPOSデータが優位なのは、早く、単品別に、店頭実売価格でデータを見ることが出来る点にある。また、日経では各量販店からのデータを中2日で収録しており、その日別を7日分合計して週次を、31日分を合計して月次を作成しているためデータの更新スピードが非常に速い。

更に、小売業とのデータ購入契約で、大半の店の個店別データを食品、家庭用品メーカーに限って提供することができる点も他社にはない強みである。

一方、収録を開始した店舗のPOSデータはそれ以後すべての期間にわたってデータベース化しており、ユーザーの要望によって全カテゴリーの全期間（最長で15年間）を提供することができる。

通常のメーカーのPOSデータの見方としては、まず全国の月別の指定カテゴリー内で自社製品の金額シェアの上下と売価を確認し、各地域毎に分解したり、より細かな週別データを検証することが一般的であろう。ただ、当研究会のためには主に日経が保有するPOSデータの最小単位である、個店別・日別データを活用するケースが多いので、改めてNEEDS-SCANより提供できる主要項目を再確認する。

表2-1 NEEDS-SCAN主要項目

販売個数	1つの固有のJAN(ジャパン・アークル・ナンバー)コードが1単位。500mlのビール1缶も、段ボール入り500mlの24缶パックも双方1個となる。
販売金額	税抜き合計販売金額
販売単価	「販売金額」÷「販売個数」で求める加重平均単価
販売数量	「販売個数」×「容量」。NEEDS-SCANには「グラム」「ミリリットル」「メートル」「個数・枚数」の4種類の容量がある。
通常価格	単一店舗での過去一定期間の中での最高価格。通常は同店舗での定番価格となる。特に最近では、メーカー側のオープン価格政策もあって、店毎に商品に対する定番価格のバラ付きが多くなる傾向にある。
特売個数	上記、店の定番価格未満で販売された合計販売個数。
特売金額	上記、店の定番価格未満で販売された合計販売金額。
値引率	上記、店の定番価格と実売価格の比率
金額シェア	指定商品(指定メーカー)の販売金額÷指定商品群(カテゴリー)全商品の合計販売金額×100
個数シェア	指定商品(指定メーカー)の販売個数÷指定商品群(カテゴリー)全商品の合計販売個数×100
数量シェア	指定商品(指定メーカー)の全販売容量÷指定商品群(カテゴリー)の全販売容量×100
特売シェア	店の定番価格未満で販売された全合計販売金額÷指定カテゴリーの全販売金額×100
客数	一定期間(日、週、月)でのPOSレジスター通過客数(レシート発行枚数)。同一人物が1日に2度レジスターを通過した場合は2人と計算する。
来店客千人当たり金額	一定期間の実販売金額÷同期間の客数×1000人(単位円)。大規模店と中規模店、2月と12月など、実際の金額での比較がしにくい際に特に有効な指数。レジスター通過客が一人平均いくら購入しているか(数値が小さくなるため1000倍している)がわかるので、店舗規模、営業日数の違いを標準化して比較が可能となる。
来店客千人当たり個数	一定期間の販売個数÷同期間の客数×1000人。単位個数。
来店客千人当たり数量	一定期間の販売数量÷同期間の客数×1000人。単位は「グラム」「ミリリットル」「メートル」「個数・枚数」のいずれか

出現店数	指定の商品コードが、指定の期(日、週、月)に1回以上POSレジスターを通過した延べ店舗数。逆に出現しない場合には、①本当に店頭におかれておらずデータに現れない場合、②実際には棚に置かれていたにもかかわらず1個も売れない場合、③商品が売れ過ぎて棚から欠品となり、補充が間に合わず結果として売れなかった場合、の3通りが考えられる。
カバー率	出現店数 ÷ 当該全店舗数 × 100

その他、「1日当たり販売個数」「1日当たり販売金額」「1日当たり販売数量」や店舗データとして、「店舗販売総金額」「店舗販売総個数」「来店客1人当たり販売金額(客単価)」「来店客1人当たり販売個数」「全品平均単価」「延べ営業日数」などの項目を出力することができる。

以上主要なNEEDS-SCANに収録されている出力可能項目を列挙したが、後半の各分析では主に、「販売個数」、「販売金額」、「販売単価」、「来店客千人当たり販売金額」、「金額シェア」を利用している。

2.3 NEEDS-SCAN/PANEL

通常のPOSデータでは取れない、消費者の「顔」と「購入の場面」が見えるIDカードを利用したパネルデータで、同系統のサービスとしては財団法人流通経済研究所と日本経済新聞社の2種類がある。

まずPANEL地域を設定するに当たり、その地域が独立した商圈を形成しているのが望ましい。次にその地区の圧倒的な一番店で、ストアロイヤリティの高い一次商圈内の世帯にパネラーになっていただき、IDカードを常に携帯し、買い物時には必ずレジスターに提示してもらう。そのためには多少のインセンティブも必要となる。

データ処理も重要で、ただ単なるPOSデータと異なりID番号毎のレシートデータの切り出しが可能な、情報システムのしっかりした大手量販店との連携が必要である。

また、パネラーの属性を継続的に把握しておくことが肝要で、日経では2500世帯の属性を1年に1度、調査票記入方式で必ず更新している(図2-2)。以下に主な属性項目を列記する。

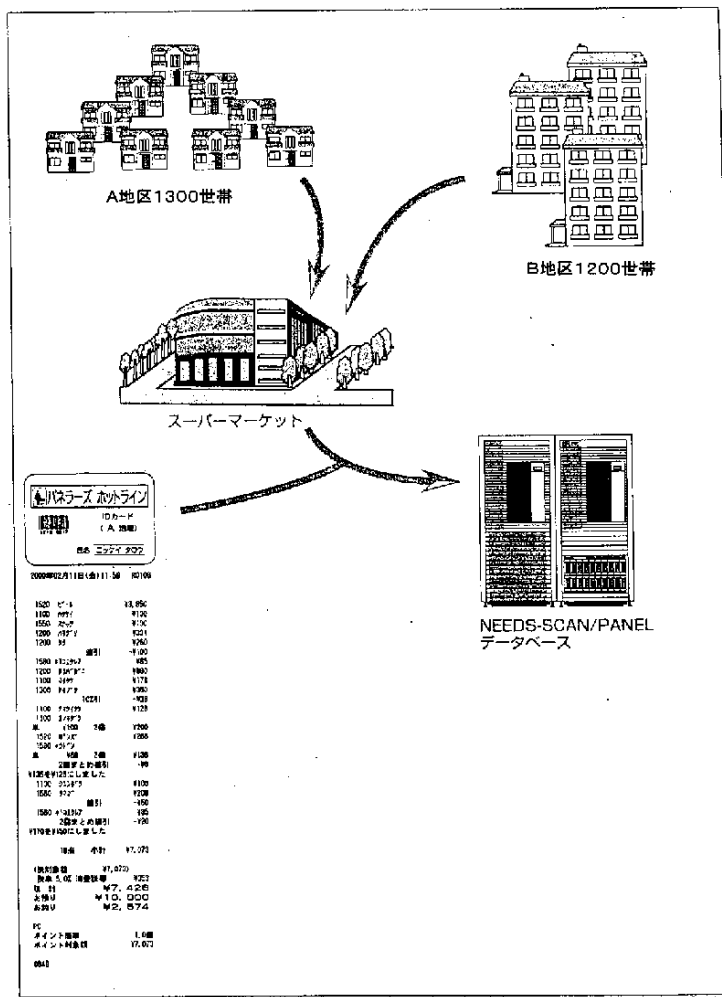


図2-2 NEEDS-SCAN/PANEL データの取得

表2-2 NEEDS-SCAN/PANEL 属性項目

①地域	⑥主婦の職業
01 A地区	01 無職
02 B地区	02 自営業
	03 事務職
②世帯人員	04 管理職
01 2人以下	05 専門職
02 3人	06 技能職
03 4人	07 サービス
04 5人	08 その他
05 6人以上	
	⑦主人の職業
③年収	01 管理職
01 300万円未満	02 専門職
02 ~400万円未満	03 事務職
03 ~500万円未満	04 技能職
04 ~600万円未満	05 販売サービス
05 ~800万円未満	06 自営業
06 ~1000万円未満	07 無職・その他
07 ~1000万円以上	
	⑧主婦の学歴
④主婦の年齢	01 中学卒
01 ~29歳	02 高校卒
02 30~34歳	03 短大卒
03 35~39歳	04 大学・大学院卒
04 40~44歳	05 主婦いない
05 45~49歳	
06 50歳以上	⑨主人の学歴
	01 中学卒
⑤主人の年齢	02 高校卒
01 ~29歳	03 短大卒
02 30~34歳	04 大学・大学院卒
03 35~39歳	05 主人いない
04 40~44歳	
05 45~49歳	
06 50歳以上	

このほか⑩子供の有無 ⑪子供の年齢 ⑫息子の有無 ⑬息子の年齢
 ⑭娘の有無 ⑮娘の年齢 ⑯世帯構成員の年齢別 ⑰住宅の保有状況 ⑱
 住宅の広さ ⑲ペットの有無 ⑳耐久消費財の保有状況 など30項目にわ
 たる調査を繰り返して実施している。

ただし、詳細な世帯毎の属性情報は個人のプライバシー問題とも絡むものであり、個人が特定できるような利用の方法やデータの提供は禁止し、すべて統計処理を施した集計結果のみが可能である。

このように、データベース化された「NEEDS-SCAN/PANEL」から、何がわかるのだろうか。1年間に2500世帯が該当店舗で買った約12万枚の全レシートから取り出せる情報を考えていただきたい。

①パネル別購入分析

全レシートの中から指定されたカテゴリーに属するすべての商品を時系列的に抜き出す。すなわち、あるカテゴリー内の商品をパネラーがいつ、何を、いくつ、いくらで買ったかの全購入履歴を出力すると、「購入間隔」「商品毎のブランド・スイッチ」状況が正確に把握できる。また、新製品の発売時点から継続的に見ていけば、その商品の購入世帯比率がわかり、細かく分解すれば、1回購入世帯が全世帯に占める「トライアル率」、2回以上購入の「リピート率」が算出できる。

②デモグラフィック分析

カテゴリー別に特定期間、蓄積したパネラーの購買データを、前述した世帯毎に把握している世帯属性で切り分けて集計する。その結果として、指定のカテゴリーや商品のヘビーユーザー像が浮かび上がり、競合状況、購入頻度の多い層、新製品の浸透層なども属性別に切ることができる。

③バスケット分析

PANELデータの最も原始データである、パネラーの購入毎の全レシート情報からは、まず平均「客単価」がわかる。また、商品別に見れば米国ウォルマート社の分析で有名になった「紙おむつ」と「ビール」の併買が多いという、組み合わせ購買や、全レシートにおける当該商品の出現する比率すなわち購入パネル率、特売時購入率、購入時平均値引率などの販売促進がどれだけ購買に寄与しているかがわかる。

更にロイヤリティが算出できる。PANELデータでの「シェア」は、当該商品の購入世帯も非購入世帯も含めた全世帯で購入された金額を、全世帯で購入されたその当該商品を含むカテゴリー全商品の合計金額で割って算出する。それに比較して日経の「ロイヤリティ」は当該商品の購入世帯だけを抽出して、その購入世帯の指定カテゴリーに占める当該商品のシェアとして求めている。いわば購入経験世帯内シェアである。

今回紹介した、これら「NEEDS-SCAN/PANEL」の各種分析は毎月第一土曜日の日経流通新聞に掲載を行っているのでご参照いただきたい。（記事見本を次ページに掲載。）

以上述べてきたように、NEEDS-SCAN/PANELデータベースは丁寧に蓄積されていることがご理解いただけると思う。

更に、分析する目的により、より多くの全パネラーを利用する場合から、極端に来店回数の多いパネラーを抽出して、例えば月間来店回数が月当たり8回以上のパネラーだけに絞り込んだ上で集計、分析するなどの機能を付与している。

このSCANデータとPANELデータを応用すれば、店頭販売実験も可能である。日経と大手食品メーカー、大手量販店の3社の共同実験として、「冷凍ショーケース内商品の販売効率化」を実施した。それまで冷凍ショーケースの中には商品がカテゴリー別に置かれていたものを、すべてメーカー別に置き直した結果、同じ冷凍ショーケース内の売り上げが約15%アップした。消費者側からのデータでは、同一メーカーの別のカテゴリー商品の同時購入が増えたことが販売額アップに貢献したことが判明した。

日経のデータに限らず、店頭実験はさまざまな小売業で実施されているようである。店舗を限定したプリテスト実験、本格発売前の1地域に絞った販売テスト、1チェーンと共同開発したダブルブランド販売実験など枚挙に暇がない。

1年間に登場する9万品の新製品の中には、前述の丁寧なテストを重ねた自信作から数打てばいくつかは当たる方式の出た所勝負型まで玉石混交だが、1年後まで生き残る商品は10%に満たない。

繰り返しになるが、POSデータはバーコードを基本単位としたデータベースであり、日経は可能な限り細かい単位（NEEDS-SCANは店別・日別、NEEDS-SCAN/PANELは店別・POSレジスター通過時間まで）でデータ収録を継続し、自ら合計してユーザーに提供している。一方、ユーザー側のデータの見方は逆に大きく見た上で疑問点があれば、次第に細かいものまで遡及するのが普通である。

この後、実際に分析事例が登場するが、なるべく細かいデータを使い、普段めったに実施しないような所まで立ち入ったつもりである。結果としてすべて満足いかないまでも、一つ二つのヒントにでもしていただければ、分析担当者の汗もさわやかな汗に変わるだろう。

3. 新製品立ち上がり売れ行き予測

3. 1 分析の目的

本章における分析は、商品を発売してから可能な限り早い段階での売れ行き情報に基づき、長期的な安定販売数量を予測する試みである。基礎となるデータは、日経の「NEEDS-SCAN」(POSデータ)を使用している。

新製品の発売時には、事前のマーケット調査などにより可能な限りの売れ行き予測が行われるのが普通である。しかし、限られた事前調査の情報からの確に発売後の結果を予測するのが困難であることは間違いがない。もし、タイムリーに供給されるPOSの売れ行きデータから将来の売れ行きが予測できるとすれば、生産・流通のコントロールに反映することにより、経営的なメリットは大きいものと考えられる。

データマイニングの領域において「予測」はかなり大きなウェイトを持つ。しかし、「予測」のために適切な手法というのが明確に存在するわけではない。利用の目的に応じて、例えば「大ヒット商品」と「小ヒット商品」に仕分けるための「クラシファイア」を利用することも可能であるし、GAMなどの適応型回帰モデルなどを利用することも可能である。

本章においては、同一データに対して、複数の(比較的ポピュラーな手法を)適用し、予測の精度を評価することも試みる。

3. 2 分析の流れ

分析を行うにあたっては、いわゆるKDDプロセスに準拠する。これは、特段珍しいものではなく、従来の統計分析にあたっても採用されている流れであるが、大規模で複雑なデータを分析対象とするデータマイニングの標準的なステップとして提唱されているものである^{*1}。

- a. Selection : 分析対象とするデータを選択する(サンプリング、データ項目の事前選択など)
- b. Preprocessing : データの事前加工(データクリーニング、時系列データに対する処理など)
- c. Transformation : データ変換および生成(本格的な変数選択や複合的な変数の生成)
- d. Data Mining : 各種の手法により、分析目標となるパターンを抽出する
- e. Evaluation : 分析結果の評価、検証および解釈を通じて、業務的に適用可能な「知識」に形式を整える

なお、分析には「SAS System」(SASインスティテュート・ジャパン)を利用した。

3. 3 データの定義とクリーニング

3. 3. 1 分析対象としたデータ

分析対象とする商品カテゴリとしては「スナック菓子」を選定した。選定の理由は、比較的頻繁に新商品が発売され、ある程度限られている時系列的なデータ観測期間中に適切なサンプルとなりうる商品が、相対的に多く発見できると考えたからである。

データは「NEEDS-SCAN」(POSデータ)より、次の要領で抽出を行った。

抽出期間：1997年10月より1999年11月まで(104週間)

抽出項目：下記の10項目とした

地域(9地域分類、全国計を含めて10分類)

日付(週単位のサマリー)

商品名

商品コード

カテゴリーコード(スナックの中の分類、10分類)

千人当り販売金額

販売金額(実額)

販売個数

平均単価(販売金額/販売個数より算出)

カバー率(調査対象店舗中、販売実績が計上された店舗の比率)

地域分類：全国

北海道

東北

北陸

関東外郭

首都圏

中京

近畿

中四国

九州

カテゴリ：ポテトチップス

ジャガ芋系スナック

サツマ芋系スナック

ポップコーン

ポップコーンの素

コーン系スナック

小麦系スナック

米系スナック

豆系スナック

その他スナック菓子

抽出件数：384, 599件

対象商品：1693商品

3.3.2 データの概要

上記のデータの内容について若干取りまとめておく。

(1) 商品のカテゴリー別構成

対象とする商品のカテゴリー別の構成を表3-1に示す。

表3-1 商品のカテゴリー別構成

商品カテゴリー	商品数	構成比
ポテトチップス	458	27.1%
ジャガイモ系スナック	132	7.8%
サツマイモ系スナック	27	1.6%
ポップコーン	113	6.7%
ポップコーンの素	8	0.5%
コーン系スナック	415	24.5%
小麦系スナック	363	21.4%
米系スナック	121	7.1%
豆系スナック	31	1.8%
その他スナック菓子	25	1.5%
合計	1693	100.0%

「ポテトチップス」、「コーン系スナック」、「小麦系スナック」というカテゴリーに商品が集中しており、この3カテゴリーで全体の73%の商品数を占めている。

(2) 平均的な販売個数

全国での週あたりの平均販売個数をまとめてみると次の通り。個数ベースで、最も売れているのは「豆系スナック」であり、全国で平均1860個/週程度が売れている。しかし、最大ヒット商品を持っているのは「ポテトチップス」であり、週5万個を超える商品が見られる。

表3-2 商品の平均販売個数

商品カテゴリー	平均販売個数	最大値
ポテトチップス	1,059	52,177
ジャガイモ系スナック	907	32,733
サツマイモ系スナック	489	6,493
ポップコーン	205	3,369
ポップコーンの素	290	4,840
コーン系スナック	658	15,176
小麦系スナック	652	17,969
米系スナック	191	6,944
豆系スナック	1,863	16,478
その他スナック菓子	107	3,326
全体	707	52,177

(3) 商品のライフサイクル

データを絞り込む前なので厳密なものではないが、商品のライフサイクルを簡単に取りまとめる。データの定義上、もっとも長いライフを持つ商品でも104週に限られる。また、観測打ち切り時に継続販売中のものも多く存在するため、明らかに短めのバイアスがかかっている。

表3-3 商品の平均ライフサイクル

商品カテゴリー	販売継続期間
ポテトチップス	71.6
ジャガイモ系スナック	70.8
サツマイモ系スナック	93.0
ポップコーン	76.7
ポップコーンの素	99.6
コーン系スナック	76.2
小麦系スナック	83.4
米系スナック	80.4
豆系スナック	65.7
その他スナック菓子	68.8
全体	76.5

「ポップコーンの素」、「サツマイモ系スナック」がやや長命商品が多い（もしくは新製品発売の件数が少ない）と推定されることを除けば、各カテゴリーとも比較的類似した商品ローテーションが行われていると推定される。

3. 3. 3 データのクリーニング

通常のデータクリーニング作業で必要となる「異常値」や「欠損値」は今回のデータには存在しなかった。したがって、分析の目的に沿うように分析サンプルの絞り込みに入る。

(1) 販売開始時期の認定

週間集計値であるため、いつ発売されたと考えるかという点については、ある程度定義していく必要がある。ここでは、週間の販売個数が全国で50個以上になった週を「販売開始週」と定義した。

このチェックにより、販売後一度も販売個数が50個を超えなかった商品は振るい落とされる(極めて、ローカルな商品、マイナーな商品は除外したことになる)。この段階で、残った商品は「769種類」であり、当初の1693商品が大幅に減少したことになる。

(2) 短期販売商品の除去

分析目的が、成長性の分析であるため、当初から期間限定商品(もしくは、販売不振のためなどにより、短期で販売を打ち切った商品)と推定されるものを除去する。除去の規準は、販売継続期間が13週(3ヶ月に相当)以下のものである。

このチェックにより、データ観察の最終週の13週以内に発売された商品も除去されることになる。このチェックを終えた段階で残った商品は「621種類」となる。当初の商品数に比較して約38%にまで減少しており、ある程度のボリュームをもって、比較的長期にわたって販売されている商品が少ないことが明らかである。

(3) 地域限定商品の検討

全国を9地域に分割したときに、商品の地域カバー率については、事前に検討しておく必要があるだろう。発売後14週間経過した時点で、9地域のうちのいくつの地域で販売実績が上がっているかをカウントしてみると、次のような分布を示す。

表3-4 販売エリアの広さ

販売地域数	商品数	構成比
1	289	46.5%
2	66	10.6%
3	38	6.1%
4	31	5.0%
5	16	2.6%
6	34	5.5%
7	28	4.5%
8	49	7.9%
9	70	11.3%
合計	621	100.0%

全体の半数近くの商品が、実は1地域でしか販売されていない。全国展開されている商品は、僅か70種類であり、全体の11%程度となっている。「地域」という概念は、前述のようになりに広い地方であるため、このような結果は十分に考えられる。したがって、単純に「地域限定商品」を除去するのではなく、成長性を予測する際に「地域の広がり」というファクターを考慮するという扱いの方が望ましいと判断する。

3. 3. 4 ターゲットの選択

以上の621商品の中から、今回の分析サンプルとなる商品を選び出す。まず、全体の商品を次のカテゴリに分ける。

- a. 通期販売：データ観察開始週から最終週まで継続して販売されていた
- b. 中途消滅：データ観察開始週には販売されていたが、途中で打ち切り
- c. 販売開始：データ観察開始週には販売されていなかった
 - c-1. 短期商品：観察期間中に販売開始し、中途消滅
 - c-2. 継続商品：観察期間中に販売開始、終了時にも継続

発売後のパターンを観察する必要があるため、上記のカテゴリcが分析対象となる。(ちなみに、c-1のカテゴリに属する商品も、少なくとも14週間の販売実績が存在することに注意)。

カテゴリ別の商品数は次のようになる。

表3-5 販売パターン別の商品数

商品カテゴリー	通期販売	中途消滅	販売開始	短期商品	継続商品
ポテトチップス	45	35	121	76	45
ジャガイモ系スナック	13	16	26	14	12
サツマイモ系スナック	4	0	3	1	2
ポップコーン	12	11	15	2	13
ポップコーンの素	2	0	0	0	0
コーン系スナック	51	28	76	39	37
小麦系スナック	41	13	67	30	37
米系スナック	15	1	9	5	4
豆系スナック	3	1	8	3	5
その他スナック菓子	1	1	3	1	2
全体	187	106	328	171	157

ターゲットとなる商品は、「328種類」である。

3. 4 事前の分析

3. 4. 1 成長性評価の指標

成長性を表現する指標として、当該商品の「全国シェア」を用いることとする。(このシェアは、個数ベースである)。シェアを評価基準とするのは、スナックに対する需要の季節変動を除去するためである。

まず、14週間経過時点での全国シェアの分布を見てみる。

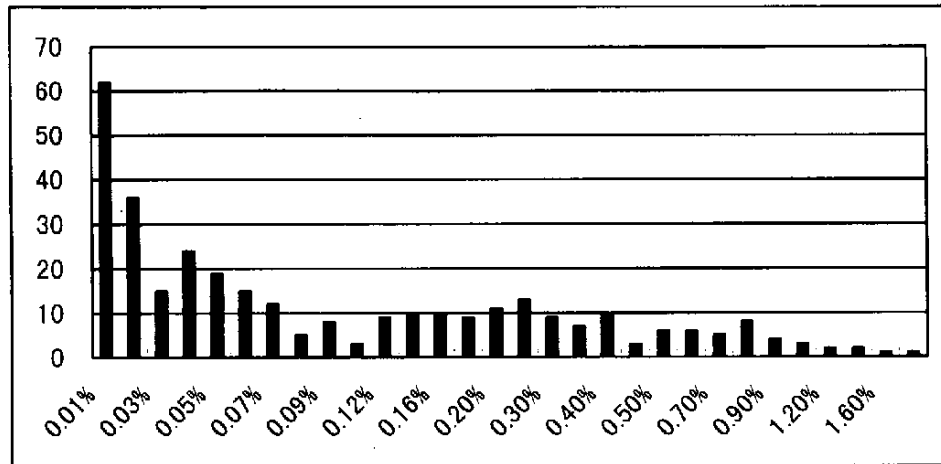


図3-1 全国シェアの分布

分析対象となる328商品の平均全国シェアは0.169%である。また、メディアン(中位値)は0.054%であり、最大シェアを持つ商品は1.61%までに達している。この時点で優勝劣敗はかなり明らかになっている。(なお、図3-1の横軸は不均等分割であることに注意)。成長性をシェアの数値ではなく、カテゴリとして表現する場合には、次のような基準で分割するのが妥当であろう。

- a. 優位：0.2%超の全国シェア (約25%の商品が該当)
- b. 中位：0.05%超0.2%以下の全国シェア (約25%の商品が該当)
- c. 劣位：0.05%以下の全国シェア (約50%の商品が該当)

ちなみに、地域展開の程度は当然のことながら、全国シェアの大小を支配する。高々2地域までの展開を「少地域」とし、3地域以上の展開を「多地域」として、上記のシェアカテゴリを見ると次のようになる。

表3-6 地域展開の広さと獲得シェアの関係

	少地域	構成比	多地域	構成比	合計	構成比
優位	31	17.3%	49	32.9%	80	24.4%
中位	42	23.5%	50	33.6%	92	28.0%
劣位	106	59.2%	50	33.6%	156	47.6%
合計	179	100.0%	149	100.0%	328	100.0%

3. 4. 2 単純な相関分析

発売開始後1週間目および2週間目のシェアと、推定ターゲットとする第14週目のシェアを比較してみよう。

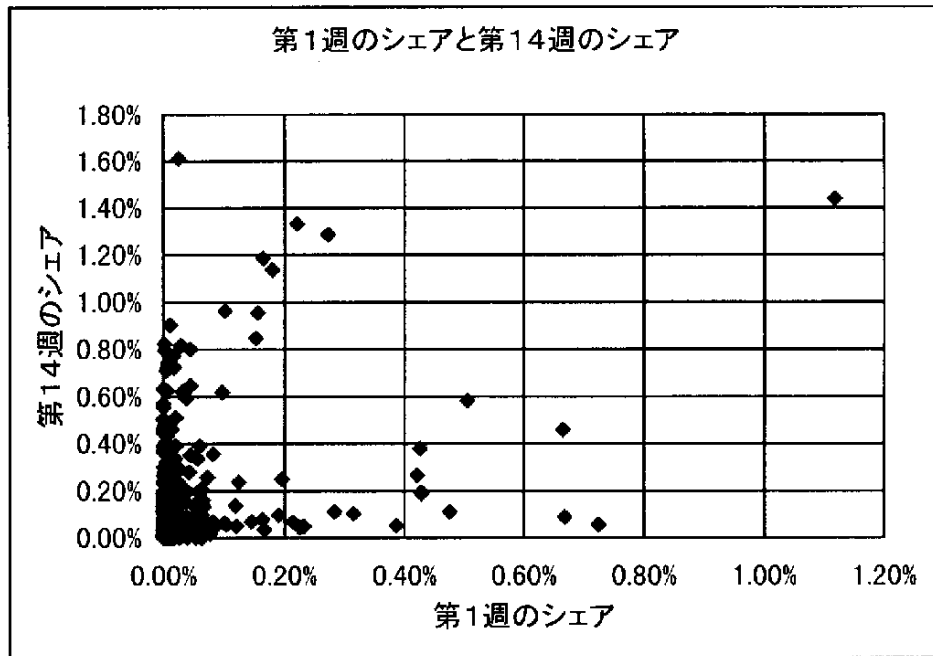


図3-2 第1週と第14週のシェアの関係

単純な線形回帰を行うと次のような結果となる。

$$\begin{aligned} \text{第14週シェア} &= 0.0014 + 0.6336 * \text{第1週シェア} \\ \text{決定係数} &= 0.079 \end{aligned}$$

つまり、第1週シェアでは第14週シェアの分散の8%程度しか説明できない。また、図3-2を目視しただけで明らかなように、一商品だけ非常に早期から大きなシェアを獲得し、その状態を14週目まで継続しているものがある(図の右上)。これが全体の傾向に大きく影響しており、この商品を除去した後に、同様の回帰を行ってみると、決定係数は0.029まで低下する。この水準は、ほぼ無相関と言ってもよいであろう。

次に、第2週のシェアと第14週のシェアとの関連を同様に分析してみよう。図3-3に示すように、かなり状況は安定してくる。しかし、まだ関連性は低く、単純な回帰分析では「予測」といえる水準は達成できそうもない。

線形回帰のパラメータは次の通り。

$$\begin{aligned} \text{第14週シェア} &= 0.0009 + 0.4571 * \text{第2週シェア} \\ \text{決定係数} &= 0.275 \end{aligned}$$

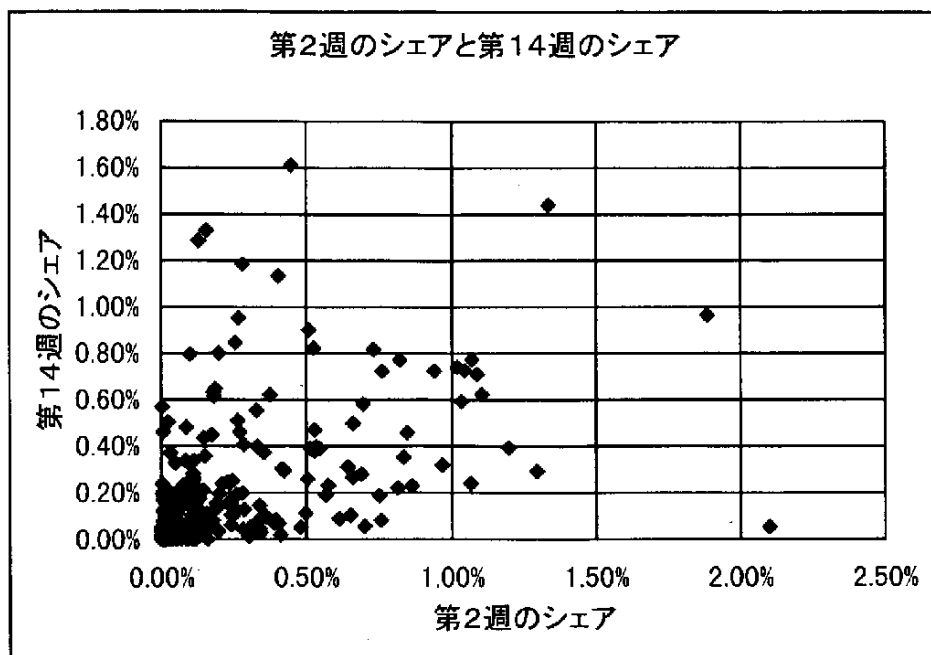


図3-3 第2週と第14週のシェアの関係

3. 4. 3 説明変数の評価と選択

次の事前分析として、利用可能な説明変数を評価してみる。この時に、若干特殊な手法を利用する。全ての商品を利用するのは可能であるが、「優位」商品と「劣位」商品を仕分ける要因を際立たせるために、「中位」商品を除きして考える。

説明変数の候補としては、第1週と第2週の双方に関する次のデータとする。

- a. 地域別のシェア（対全国シェアと対地域シェア）
- b. 地域別のカバー率
- c. 全国シェアの変動（第2週シェア－第1週シェア）
- d. 全国カバー率の変動（第2週シェア－第1週シェア）
- e. 販売中地域数

各変数の説明力を評価する尺度としては、カイ自乗値を利用する（これが大きいほど説明力がある）²。

結果を表3-7に示す。表中の変数は、カイ自乗値の大きい順に並んでいる。有意水準5%のところまでで25個の変数が選択された。非常に分かりやすい変数群が抽出されており、判別力もかなり高いことが推定される。

表3-7 各変数の説明力評価(その1)

評価対象変数	カテゴリー値	劣位商品		優位商品	
		件数	横%	件数	横%
1 2週全国カバー率 $\chi^2=125$ ($p=0.000$)	0.4-2.25	53	89.8	6	10.2
	2.25<-4.75	56	94.9	3	5.1
	4.75<-12.05	42	71.2	17	28.8
	12.05<-48.9	5	8.5	54	91.5
2 シェア増減 $\chi^2=124$ ($p=0.000$)	-0.00145-0.0002	55	93.2	4	6.8
	0.00003<-0.0001	56	94.9	3	5.1
	0.00012<-0.001	39	66.1	20	33.9
	0.001<-0.0178	6	10.2	53	89.8
3 2週全国シェア $\chi^2=124$ ($p=0.000$)	0-0.00013	55	93.2	4	6.8
	0.00014-0.0003	56	94.9	3	5.1
	0.0003-0.0018	39	66.1	20	33.9
	0.0019-0.0188	6	10.2	53	89.8
4 カバー率増減 $\chi^2=114$ ($p=0.000$)	-0.9	56	96.6	2	3.4
	0.4<-1.9	52	88.1	7	11.9
	1.9<-8.45	41	68.3	19	31.7
	8.45<-45.5	7	11.9	52	88.1
5 2週関東カバー率 $\chi^2=68$ ($p=0.000$)	3.4-3.4	37	94.9	2	5.1
	3.4<-6.9	18	72	7	28
	6.9<-25	14	45.2	17	54.8
	25<-82.1	0	0	32	100
6 2週関東シェア $\chi^2=61$ ($p=0.000$)	0-0.00001	28	90.3	3	9.7
	0.00002-0.00005	27	84.4	5	15.6
	0.00006-0.0003	12	37.5	20	62.5
	0.0004-0.0023	2	6.3	30	93.8
8 2週首都圏カバー率 $\chi^2=60$ ($p=0.000$)	1.4-4.5	31	83.8	6	16.2
	4.5<-7.5	34	89.5	4	10.5
	7.5<-19.1	22	55	18	45
	19.1<-62.5	4	10.8	33	89.2
9 2週首都圏シェア $\chi^2=59$ ($p=0.000$)	0-0.00007	32	84.2	6	15.8
	0.00008-0.0001	34	89.5	4	10.5
	0.0002-0.0008	20	52.6	18	47.4
	0.0009-0.007	5	13.2	33	86.8
10 2週東北シェア $\chi^2=51$ ($p=0.000$)	0-0.00076	20	87	3	13
	0.00077-0.0028	21	91.3	2	8.7
	0.0029-0.0156	7	29.2	17	70.8
	0.0156-0.0528	1	4.3	22	95.7
11 地域数増減 $\chi^2=61$ ($p=0.000$)	-1	2	100	0	0
	0	83	85.6	14	14.4
	1	35	77.8	10	22.2
	2	20	55.6	16	44.4
	3	8	40	12	60
	4	4	26.7	11	73.3
	5	3	33.3	6	66.7
	6	0	0	8	100
7	1	25	3	75	
12 2週販売地域数 $\chi^2=50$ ($p=0.000$)	1	53	91.4	5	8.6
	2	29	80.6	7	19.4
	3	26	74.3	9	25.7
	4	15	60	10	40
	5	7	33.3	14	66.7
	6	9	37.5	15	62.5
	7	8	57.1	6	42.9
	8	7	50	7	50
9	2	22.2	7	77.8	

表3-7 各変数の説明力評価(その2)

評価対象変数	カテゴリー値	劣位商品		優位商品	
		件数	横%	件数	横%
13 2週東北カバー率 $\chi^2=36$ ($p=0.000$)	7.1-14.3	24	85.7	4	14.3
	14.3<-28.6	11	68.8	5	31.3
	28.6<-75	13	50	13	50
	75<-100	1	4.3	22	95.7
14 2週中京カバー率 $\chi^2=34$ ($p=0.000$)	6.2-6.2	22	100	0	0
	6.2<-20	12	54.5	10	45.5
	20<-46.7	9	42.9	12	57.1
	46.7<-100	3	13.6	19	86.4
15 2週東北シェア $\chi^2=34$ ($p=0.000$)	0-0.00003	23	88.5	3	11.5
	0.00004-0.0001	21	80.8	5	19.2
	0.0002-0.0003	12	44.4	15	55.6
	0.0004-0.0016	5	19.2	21	80.8
16 2週九州カバー率 $\chi^2=32$ ($p=0.000$)	5.6-12.5	21	95.5	1	4.5
	12.5<-20	14	77.8	4	22.2
	20<-50	7	38.9	11	61.1
	50<-86.7	3	15.8	16	84.2
17 2週九州シェア $\chi^2=28$ ($p=0.000$)	0-0.00002	17	89.5	2	10.5
	0.00003-0.00008	16	84.2	3	15.8
	0.00009-0.0004	9	45	11	55
	0.00045-0.0028	3	15.8	16	84.2
18 2週東北カバー率 $\chi^2=28$ ($p=0.000$)	4.5-9.1	19	79.2	5	20.8
	9.1<-10	23	88.5	3	11.5
	10<-25	12	42.9	16	57.1
	25<-65	7	25.9	20	74.1
19 2週近畿カバー率 $\chi^2=23$ ($p=0.000$)	2.6-6.1	17	73.9	6	26.1
	6.1<-10.5	23	79.3	6	20.7
	10.5<-18.4	12	50	12	50
	18.4<-87.9	5	20	20	80
20 2週近畿シェア $\chi^2=22$ ($p=0.000$)	0-0.00004	18	72	7	28
	0.00005-0.0001	17	68	8	32
	0.0002-0.0005	18	69.2	8	30.8
	0.0005-0.0141	4	16	21	84
21 2週中京シェア $\chi^2=18$ ($p=0.000$)	0-0.000029	18	85.7	3	14.3
	0.0003-0.0001	13	59.1	9	40.9
	0.0001-0.0003	10	45.5	12	54.5
	0.0004-0.0018	5	22.7	17	77.3
22 2週中四国カバー率 $\chi^2=16$ ($p=0.001$)	4.5-9.1	10	47.6	11	52.4
	9.1<-15	15	68.2	7	31.8
	15<-50	7	33.3	14	66.7
	50<-81.8	2	9.5	19	90.5
23 2週中四国シェア $\chi^2=16$ ($p=0.001$)	0-0.00047	11	52.4	10	47.6
	0.00048-0.00283	14	66.7	7	33.3
	0.00284-0.0122	7	31.8	15	68.2
	0.0123-0.0502	2	9.5	19	90.5
24 1週東北カバー率 $\chi^2=15$ ($p=0.002$)	9.1<-10	15	83.3	3	16.7
	4.5-4.5	15	88.2	2	11.8
	4.5<-9.1	4	36.4	7	63.6
	10<-45.5	6	40	9	60
25 1週東北シェア $\chi^2=15$ ($p=0.002$)	0-0.00001	10	66.7	5	33.3
	0.00002-0.00003	15	100	0	0
	0.00004-0.0001	10	62.5	6	37.5
	0.0002-0.0011	5	33.3	10	66.7

3. 5 Decision Tree による分析

上記の分析にしたがい、変数を若干絞り込んだ上で、データマイニングを行う。マイニングにあたっては、Decision Tree による成長性クラシファイアを作成することとした。ツリー生成にあたっては、「C4. 5」が採用しているものと同様の、エントロピー最小化基準に基づくアルゴリズムを適用した³⁾。

生成されたツリーは次のような構造を持つ。

表 3-8 生成されたツリー

母集団 N=236 P=0.339	全国シェア増減 [0. 1%以上] N=87 P=0.793	全国カバー率 [10%未満] N=20 P=0.400		
		全国カバー率 [10%以上] N=67 P=0.910	東北シェア [0. 01%以上] N=21 P=1.000	
			東北シェア [0. 01%未満] N=46 P=0.870	首都圏カバー率 [7%未満] N=11 P=1.000
		全国シェア増減 [0. 1%未満] N=149 P=0.074	関東シェア [0. 05%未満] N=90 P=0.100	首都圏シェア [0. 01%未満] N=74 P=0.108
				首都圏シェア [0. 01%以上] N=16 P=0.063
		関東シェア [0. 05%以上] N=59 P=0.034	関東カバー率 [7%以上] N=11 P=0.182	関東カバー率 [7%未満] N=48 P=0.000

註) 「N」はサンプルサイズ、「P」優位商品に成長する確率を示す。

限られたサンプルを対象とするため、比較的浅いツリーとなっているが、結果は極めて明らかである。

次のような点が、優位商品としての地位を確立するための条件となる。

- a. 発売後2週目までに全国シェアを0.1%以上拡大する
- b. かつ、全国でのカバー率を10%以上に持つてくる

上記の2つの条件がクリアされた場合に、優位商品に成長する可能性は（劣位と比較して）90%程度になる。面白いのは、東北地方でのシェアであり、上記2条件に加えて、東北地方での全国シェアが0.01%以上になると優位商品への成長可能性が非常に高くなることが示されている。

一方、劣位商品になる条件は次のようになる。

- c. 1週目から2週目までの全国シェアが0.1%未満である
- d. かつ、関東地方でのシェアが0.05%未満である

これに店舗カバー率の低さが加わると、劣位商品になる可能性がきわめて高い。

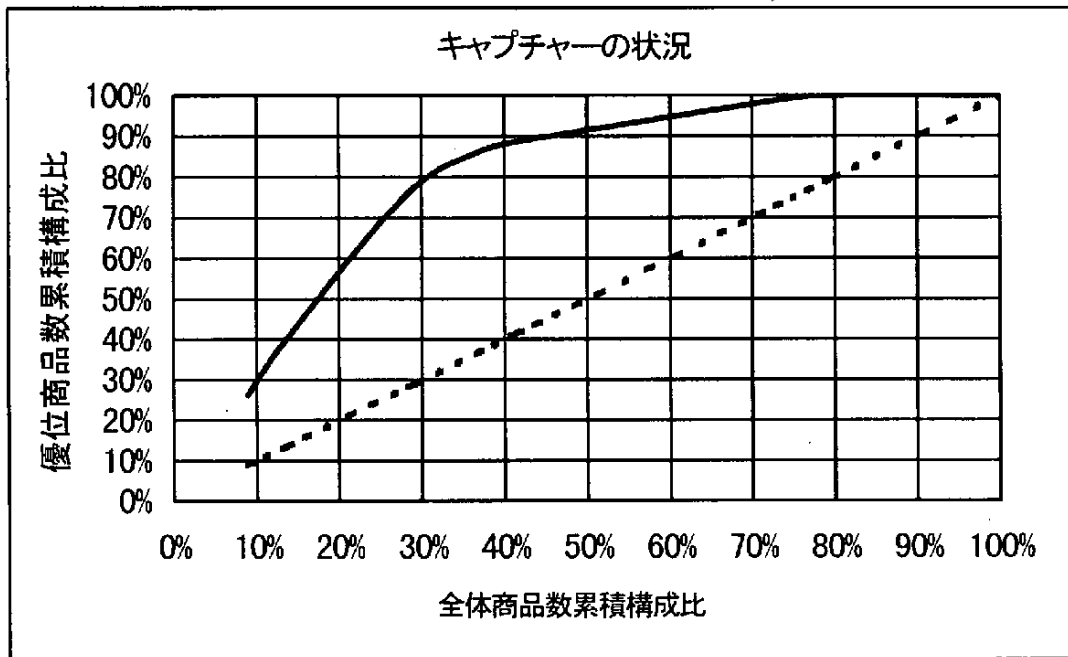


図3-4 判別効率の検証

なお生成されたツリーの判別効率を図3-4で見ると、推定で成長性が高いと見られる上位グループ50%をとると、その中に、実際に優位商品となったものの90%程度が含まれている。

3.6 まとめ

分析の結果については、ごく当たり前のものが示されているのではないかと考えられる。しかし、ツリーの各リーフの「P」の値が、平均値と比較して大きく Diverse していることから明らかなように、かなりうまく判別はできている。ツリーの安定性の検証はサンプルの制約により、充分に行われてはいないが、もし、今後このような分析を行う機会があれば、サンプルを拡大して検証すべきであろう。

*¹ Fayyad, Piatetsky-Shapiro, Smyth: "From Data Mining to Knowledge Discovery" in 'Advances in Knowledge Discovery and Data Mining', AAAI Press, 1996.

*² 金融エンジニアリング・グループが開発したSAS上で稼動するツール「CAT」を利用した。

*³ 金融エンジニアリング・グループが開発したSAS上で稼動するツール「CPA」を利用した。

4. 新製品をヒットに導く要因は何か

4. 1 目的

4. 1. 1 問題意識

現在の市場の特徴として、大量に生み出される新製品の数の多さがあげられるだろう。各メーカーとも、流行や消費者の好みにあわせ、次々と市場に新しい商品を投入してくる。

最近では、マスメディアから次々と流される情報にあわせるように、難しい名前の栄養素がパッケージに並んだり、海外からの目新しい素材などが、素早く商品に盛り込まれるようになってきている。

このように、季節や分類に関わらず数々の新製品が生まれているが、その大半の商品が一定期間の後、消えてゆくのが現状だ。消費者は一時、新製品に飛びついて、引き続き長期間にわたって、同じ商品を買いつづけてくれるわけではない。

今回、分析の対象とするポテトチップスを例にとってみても、市場の厳しさは歴然としている。1998年1年間で発売になった新製品のうち、翌年99年にも売り上げの上がった商品の比率、つまり新製品の生存率とでもいうべきものは、63.1%に過ぎない。半数近くの商品は、1年足らずで市場から消えていっている。

そして、消費者の嗜好が複雑多岐にわたる中、どんな商品がヒットを勝ち取るか、その予想ははるかに難しくなっている。ヒットの仕掛けとなる媒体も、マスメディアだけではなく、口コミの威力が大きくなっており、ヒットの理由もさまざま、その構造はますます複雑さを増している。

しかし、そんななかでも、過去の新製品を大量に振り返ってみれば、そこから売れる理由もしくは売れない理由など、何らかの共通点を見つけ出すことが出来るのではないだろうか。データマイニングのツールを利用することで、いったいどんな要素がヒットを生む要因となっているのか、その共通項ともいえるものを見つけ出してみたいと思う。

4. 1. 2 分析の観点

今回の分析において、ヒットを生み出す要因として想定するものは、メーカーが何らかの努力によって、対応し、操作しうるものに限定する。口コミや、情報番組での取り扱いなど、重要な要因と考えられるものでも、メーカーが事前に対応できないものは、運を天にまかせることしかできないからだ。

それに加え、実際の数値としてメーカー側が比較的簡単に、発売後早い段階で入手可能な要素に絞ることとする。リピート率やトライアル率といったデータは、入手可能であるものの、発売後からのタイムラグが大きく、瞬時に役立てることはできない。

そこで今回の分析では、要因として、各地域の販売価格、カバー率、個数シェアを使用することにする。なぜなら、販売価格はメーカー側である程度操作可能であること、そしてカバー率はまさにメーカーの営業力そのものであること、個数シェアは販売後の市場にまかせるしかない項目であるものの、比較的早い段階で入手でき、ヒットを判断するために欠かせない要素であるからだ。

また、要因を細かく探るために、ヒットの定義のレベルに2種類の考え方を設けることにする。ひとつは、大ヒットとなる理由。もうひとつは、ヒットとならずとも失敗はしないための理由である。大ヒットとなるために必要な条件と、失敗を避けるための条件とでは、異なるルールがあると予想できるからだ。

もちろんメーカー側としては、如何なる時も、大ヒットを目指して新製品を投入するわけではあるが、現実問題として、その可能性というのは極めて低い。となると、あるレベルを確保するための条件、失敗しないための条件を知ることが大切になってくるであろう。

4. 1. 3 分析結果の判断方法

以上の分析の結果、ヒットの要因となる要素が見つげ出された後、その結果は、新製品の売れ行きを伸ばすために役立てていくことが、次の課題として期待される。そこで、結果として見出された要因を、メーカー側の立場に立って、次のふたつの観点からみていくことにする。

(1) 発売前にメーカーが準備すべき条件

メーカー側では発売後3ヶ月というのを、新製品の売れ行きを判断する、ひとつの目安として考えている。そこで、発売後3ヶ月目の売れ行きを最大化するために、発売前には、どの要素に力点を置くことが必要なのか、その条件を見つけ出す。

(2) 発売後の判断基準となる項目

新製品発売後には、メーカーにはふたつの課題が設定される。ひとつには、今以上に売り上げを伸ばすこと。そしてもうひとつは、生産ラインの関係から販売予測をできる限り正確に立てることである。

そこで、過去の新製品の傾向と照らし合わせ、現在のカバー率やシェアでは3ヶ月後の売り上げはどの程度になるのか予測をたてる。また売り上げを伸ばすためには、どの要素が欠けているのか、問題点を明確にしていく。

4. 2 データ説明

4. 2. 1 データ概要

データソース： 日本経済新聞社のPOSデータベース
「NEEDS-SCAN」
対象地域： 全国 以下の9地域に分割
(北海道、東北、北陸、関東外郭、首都圏、中京、近畿、
中四国、九州)
対象店舗： 全国46チェーン231店舗のスーパー(2000年1月現在)
商品分類： ポテトチップス
対象期間： 97年11月～99年12月の週次データ
使用項目： 各地域別の平均価格、個数シェア、カバー率

4. 2. 2 分析方法

メーカーでは新製品のヒットを判断する材料として、発売後3ヶ月目のシェアをひとつの目安としている。そこで、「新製品発売後12週目の全国の個数シェア」を被説明変数(目的変数)として設定する。そして、それに影響を与えていると考えられる、各地域(全国を9地域に分割、集計値としての全国の計10地域)の発売後1、2週目の価格、個数シェア、カバー率の計60項目を説明変数とする。

以上のデータを商品ごとに作成し、分析ツールにかけ、第12週全国個数シェアの行方に影響を及ぼしている要素を探し出し、そのルールを見つけ出す。分析手法はCHAIDおよびC5.0。

なお、分析ソフトは、エス・ピー・エス・エス株式会社「Clementine」ならびに「AnswerTree」を用いた。

4. 2. 3 使用データ詳細説明

(1) 新製品の判断基準

97年11月から連続2週以上売り上げがゼロ(*1)であり、かつ初めて売り上げがたってから2週続けて全国いずれかの地域で売り上げがたった商品314アイテムのうち、12週後の値がとれる、299商品を新製品と判断し、今回の対象とする。

(*1) 日経のデータの特徴上、価格がゼロであるものを売り上げゼロと判断する。

(価格=販売金額÷販売個数、で算出している。個数シェアおよびカバー率は小数点第3位以下を省略して、値がゼロとなっている可能性があるが、価格の場合は、販売金額がゼロのときのみゼロとなる)

(2) データ加工方法

上記新製品において、いずれかの地域ではじめて売上げがたった週を第1週目とし、そこから各商品ごとに相対的に第1週目、第2週目と週をカウントする。

そして地域ごと（9地域と合計である全国の計10地域）に、発売後1、2週目の売上げデータを使用する。データ項目は平均価格、個数シェア、カバー率の3項目。10地域、2週分、3項目で、あわせて60の変数を各商品ごとに用意し、発売後12週目の全国の個数シェアに影響を与える説明変数とする。

(3) 各項目の詳細

(a) 説明変数として用いる、3つの項目の設定は以下の通りとする

○ 平均価格

日経のPOSデータでは、平均価格は以下のように求めている。そのため、平均価格ゼロというのは、売上げなしということを意味している。

$$\text{平均価格} = \text{販売金額} \div \text{販売個数}$$

今回の分析の中では、実際の数値ではなく、平均価格を以下の6つの分類に分割し、カテゴリ変数として使用する。ただし0は売上げなしを意味する。

1 : ~97円 2 : ~100円 3 : ~110円 4 : ~120円
5 : ~140円 6 : 140円~

○ 個数シェア

ここでは消費者が新製品を選んだ回数を問題としたいため、商品単価に影響を受ける金額シェアではなく、販売個数を採用し、個数シェアで見て行くことにする。また、実数値ではなく、シェアを採用することで、カテゴリ全体を常に100と見るため、季節要因を排除することができる。個数シェアの算出方法は以下の通り。

$$\text{個数シェア} = \frac{\text{対象商品の対象地域での販売個数}}{\text{対象分類全体の対象地域での販売個数}} \times 100$$

今回の分析の中では、実際の数値を使用する。

○ カバー率

カバー率とは、データ収録店舗のうち何店舗で実際にその商品が置いてあるかという比率で、算出方法は以下の通り。

(注) 正確にはカバー率とは、その商品が売れた店舗の比率であるが、売れない商品はすぐに売り場から排除されるため、売り場に置いてある比率と同義ととらえてよい。

$$\text{カバー率} = \text{該当商品が売れた店舗数} \div \text{対象地域の全店舗数} \times 100$$

なお今回の分析では、実際の数値を使用する。

(b) 被説明変数（目的変数）となる、発売後12週目の全国の個数シェアは、分析に応じて以下の2種類を用意する

○ 大ヒットとなる要素を探る場合

目的変数を以下のような、カテゴリ変数であらわす。

0 : 0.5%未満（ヒットせず）

1 : 0.5%以上（大ヒット）

○ 失敗しないための要素を見つけ出す場合

目的変数をおよそ3分の1ずつ分割し、以下のようなカテゴリ変数であらわす。

0 : 0%（Cランク）

1 : 0.01~0.2%未満（Bランク）

2 : 0.2%以上（Aランク）

○ 地域

使用する地域は以下の10地域で、下記の通り、各番号がそれぞれの地域を指す。

1・全国 2・北海道 3・東北 4・北陸 5・関東外郭 6・首都圏

7・中京 8・近畿 9・中四国 10・九州

(4) データの読み方

変数名はWが週カウントを表し、W1は第1週目となる。Pが平均価格、Cはカバー率、Sは個数シェアを表し、その後の数値は地域番号を意味する。

W1C2は北海道第1週目カバー率の意味となる。

表4-1 使用データ形式見本

JANコード	WIP1F	WIS1	WIC1	WIP2F	WIS2	WIC2	WIP3F	WIS3	WIC3
84114009940	6	0	0.5	0	0	0	0	0	0
84114009950	6	0	0.5	0	0	0	0	0	0
84114009980	6	0	1.9	0	0	0	0	0	0
84114033330	6	0.01	1.3	0	0	0	0	0	0
370003095120	2	0	0.4	0	0	0	0	0	0
728229678940	6	0	0.5	0	0	0	0	0	0
728229678970	6	0	0.5	0	0	0	0	0	0
4017100537004	6	0	0.5	0	0	0	0	0	0
4901002028893	6	0.01	1.4	0	0	0	0	0	0
4901002028916	6	0	1.4	0	0	0	0	0	0
4901002037987	2	0	0.5	0	0	0	0	0	0
4901002040628	2	0	0.5	0	0	0	0	0	0
4901002040635	2	0	0.5	0	0	0	0	0	0
4901002044985	4	0.01	2.1	0	0	0	0	0	0
4901002045111	1	0.05	0.5	0	0	0	1	0.42	4.5
4901002046378	2	0	0.4	0	0	0	0	0	0
4901002048440	2	0.02	1.3	2	0.09	7.1	0	0	0
4901002048457	2	0.01	0.9	0	0	0	0	0	0
4901330500450	1	0.01	0.9	0	0	0	0	0	0
4901330510213	1	0.01	0.9	0	0	0	0	0	0
4901330520359	1	0.01	0.9	0	0	0	0	0	0
4901330541682	1	0.04	6.6	2	0.13	7.7	0	0	0
4901330542122	5	0.01	0.5	0	0	0	5	0.16	5
4901330542276	1	0.08	0.4	0	0	0	0	0	0
4901330543105	6	0.01	0.4	6	0.08	7.1	0	0	0
4901330551292	2	0	0.9	0	0	0	0	0	0
4901330560102	5	0	1.3	0	0	0	0	0	0
4901335106220	6	0.01	0.9	0	0	0	0	0	0
4901335106336	6	0.04	1.4	0	0	0	0	0	0
4901335109092	1	0	0.5	0	0	0	0	0	0
4901335109108	1	0	0.5	0	0	0	0	0	0
4901335111026	6	0.02	2.8	0	0	0	0	0	0
4901335152029	6	0	0.5	0	0	0	0	0	0
4901510018300	2	0.01	0.5	0	0	0	0	0	0
4902402446706	3	0	0.5	0	0	0	0	0	0

4. 3 CHAID

4. 3. 1 大ヒットとなる要素を探る場合

目的変数を以下のような、カテゴリ変数であらわし、大ヒットとなる場合のルールを探す。フラグ設定は以下の通り。

0 : 0.5%未満 (ヒットせず)

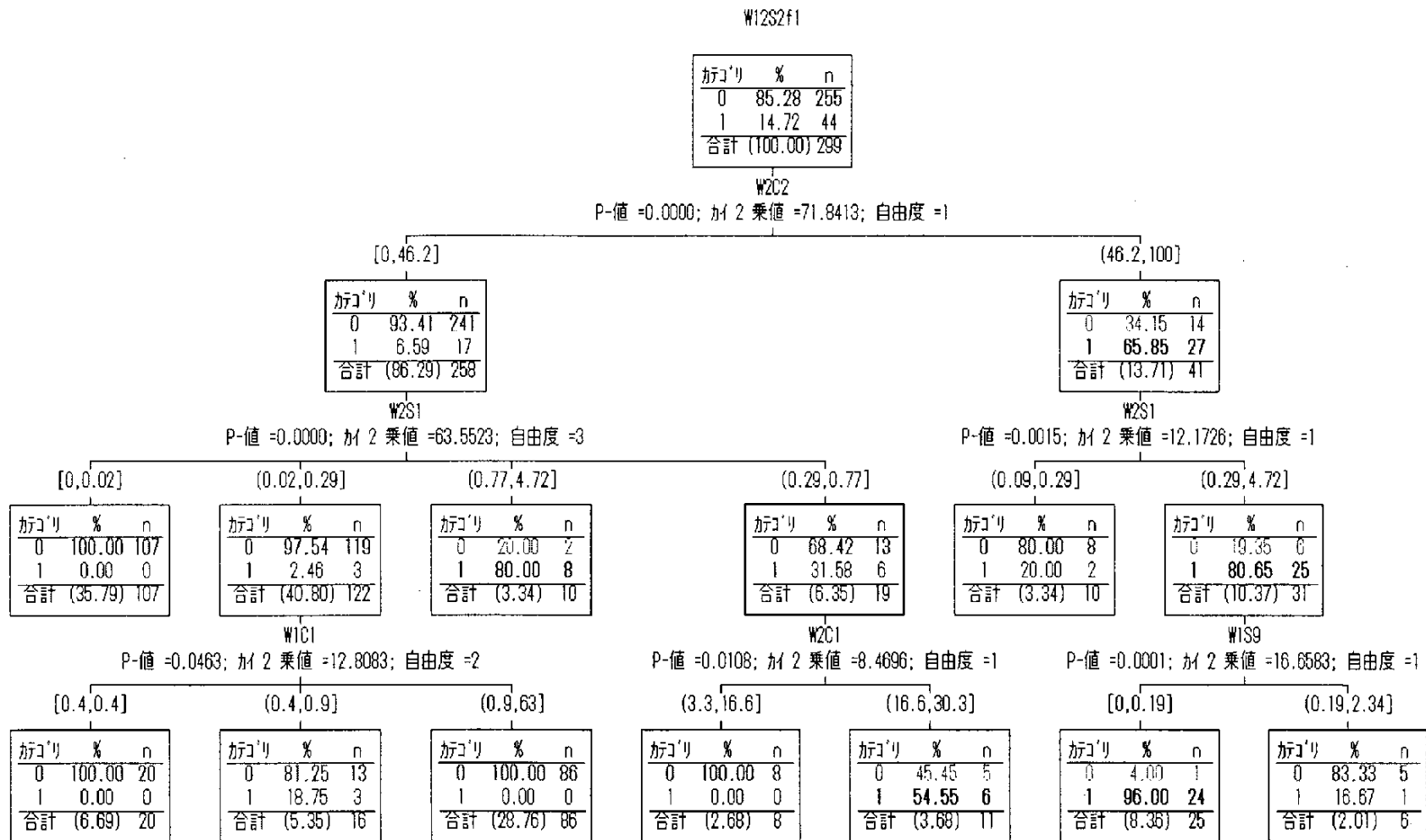
1 : 0.5%以上 (大ヒット)

(1) 目的変数が0となる比率の方が高い場合

- ・第2週北海道カバー率が46.2%以下、第2週全国シェアが0~0.29%
- ・第2週北海道カバー率が46.2%以下、第2週全国シェアが0.29~0.77%、かつ第2週全国カバー率が3.3~16.6%
- ・第2週北海道カバー率が46.2%超、第2週全国シェアが0.09~0.29%
- ・第2週北海道カバー率が46.2%超、第2週全国シェアが0.29~4.72%、かつ第1週中四国のカバー率が0.19~2.34%

(2) 目的変数が1となる比率の方が高い場合

- ・第2週北海道カバー率が46.2%以下、第2週全国シェアが0.77~4.72%
- ・第2週北海道カバー率が46.2%以下、第2週全国シェアが0.29~0.77%、かつ第2週全国カバー率が16.6~30.3%
- ・第2週北海道カバー率が46.2%超、第2週全国シェアが0.29~4.72%、かつ第1週中四国カバー率が0~0.19%



[0.4, 0.4]

カテゴリ	%	n
0	100.00	20
1	0.00	0
合計	(6.69)	20

(0.4, 0.9]

カテゴリ	%	n
0	81.25	13
1	18.75	3
合計	(5.35)	16

(0.9, 63]

カテゴリ	%	n
0	100.00	86
1	0.00	0
合計	(28.76)	86

(3.3, 16.6]

カテゴリ	%	n
0	100.00	8
1	0.00	0
合計	(2.68)	8

(16.6, 30.3]

カテゴリ	%	n
0	45.45	5
1	54.55	6
合計	(3.68)	11

[0, 0.19]

カテゴリ	%	n
0	4.00	1
1	96.00	24
合計	(8.35)	25

(0.19, 2.34]

カテゴリ	%	n
0	83.33	5
1	16.67	1
合計	(2.01)	6

図4-1 CHAID出力結果 (01の場合)

(3) 結果

表4-2 CHAID (フラグ01の場合)の主なルールで分割された対象商品の比率

第2週北海道カバー率

	46.2%以下	46.2%超
ヒットせず	93.4	34.2
ヒット	6.6	65.8

第2週北海道カバー率46.2%以下

第2週全国シェア

	0.77%以下	0.77~4.72%
ヒットせず	96.4	20
ヒット	3.6	80

第2週北海道カバー率46.2%超

第2週全国シェア

	0.29%以下	0.29~4.72%
ヒットせず	80	19.4
ヒット	20	80.6

○ 発売前にメーカーが準備すべき条件

発売後12週目の全国の個数シェア0.5%以上というヒットをおさめるためには、発売後第2週目までに、北海道におけるカバー率を46.2%超まであげる努力が求められる。北海道地区の半数近くの店舗で発売することで、全国レベルでの成功の確率が7割弱まで高まるという結果が出ている。

決して高いとは言えない確率ではあるものの、逆に、これを下回る場合、ヒットにならない確率は9割以上となってしまふ。

発売前の準備段階としては、北海道地区のカバー率を最低5割程度に維持すること、つまり半数近くの店舗に商品を置くことが、ヒットを生むために必要な条件となっているようだ。

○ 発売後の判断基準となる項目

発売してからのち、12週目の全国個数シェアを0.5%以上というヒットに導くためには、何が必要となってくるのか。

まず、第1のチェックポイントとなるのが、第2週北海道カバー率が46.2%を上回っているかどうか。これを上回っている場合、先にも述べたように、過去の事例ではこの時点で7割弱の商品はヒットとなっている。

逆に第2週北海道カバー率が46.2%を下回る場合は、それだけで9割がた失敗が確定してしまう。この場合でもヒットとなっている商品の条件というのは、同週全国シェアが0.77~4.72% (0.77%以上と同義と考えられる) と高い値を維持していることだ。その場合、12週目全国シェア0.5%以上を獲得できる可能性は8割と高い。

さらに、第1の条件をクリア、つまり第2週北海道カバー率が基準値を達成している場合を見ていく。第2の条件としてあがるのが、第2週全国シェアが0.29~4.72% (0.29%超と同義と考えられる) を確保できているかどうかだ。この場合、発売後12週目の全国シェアが0.5%以上となり、ヒット商品となる確率は8割を超える。

逆に第2週全国シェアが0.29%以下の場合は、第2週北海道カバー率が基準値をクリアしていても、12週目の全国シェアが0.5%を下回る可能性は8割になり、ヒットとならない確率はかなり高くなる。

以上のことからいえることは、第2週目の北海道のカバー率を半数程度、かつ同週の全国シェアを0.3%程度保っていれば、3ヶ月後のヒットはかなりの確率で保証されるということだ。

しかし、北海道カバー率が高くとも全国シェアが低い場合はヒットの確率は急激に下がっており、またカバー率が低くともシェアが高ければヒットの確率も高くなっている。結局は、2週目のシェアが高ければ12週目のシェアも高いということではあるものの、大ヒットを分ける要因として、第1にあがるものが、北海道のスーパーでのカバー率、浸透具合、であるというのはおもしろい。

4. 3. 2 失敗しないための要素を見つけ出す場合

対象アイテムを目的変数（発売後12週目全国個数シェア）の値でソートし、およそ3分割しABCとランク付けを行う。そしてCランクとならないルールを探し出す。フラグ設定は以下の通り。

0 : 0% (Cランク)

1 : 0.01~0.2%未満 (Bランク)

2 : 0.2%以上 (Aランク)

(1) 目的変数が0となる比率の方が高い場合

- ・第2週首都圏カバー率が9%以下、同週全国カバー率が0.4~0.9%
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が0.9~2.4%、同週関東外郭カバー率が0%
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が2.4~3.3%、かつ第1週首都圏価格帯が100円~140円まで
- ・第2週首都圏カバー率が9~22.1%、第2週全国カバー率が2.4~5.2%

(2) 目的変数が1となる比率の方が高い場合

- ・第2週首都圏カバー率が9%以下、同週全国カバー率が0.9~2.4%、同週関東外郭カバー率が0~13.8%
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が2.4~3.3%、かつ第1週首都圏価格帯が100円まで
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が3.3~9%、かつ第2週北海道シェアが4.13以下
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が9~61.6%、第2週東北カバー率が0~25%以下
- ・第2週首都圏カバー率が9~22.1%、第2週全国カバー率が5.2~16.6%

(3) 目的変数が2となる比率の方が高い場合

- ・第2週首都圏カバー率が9%以下、同週全国カバー率が0.9~2.4%、同週関東外郭カバー率が0~13.8%
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が3.3~9%、かつ第2週北海道シェアが4.13~8.25%
- ・第2週首都圏カバー率が9%以下、同週全国カバー率が9~61.6%、第2週東北カバー率が0%もしくは、25~75%
- ・第2週首都圏カバー率が9~22.1%、第2週全国カバー率が16.6~30.3%
- ・第2週首都圏カバー率が22.1~92.5%

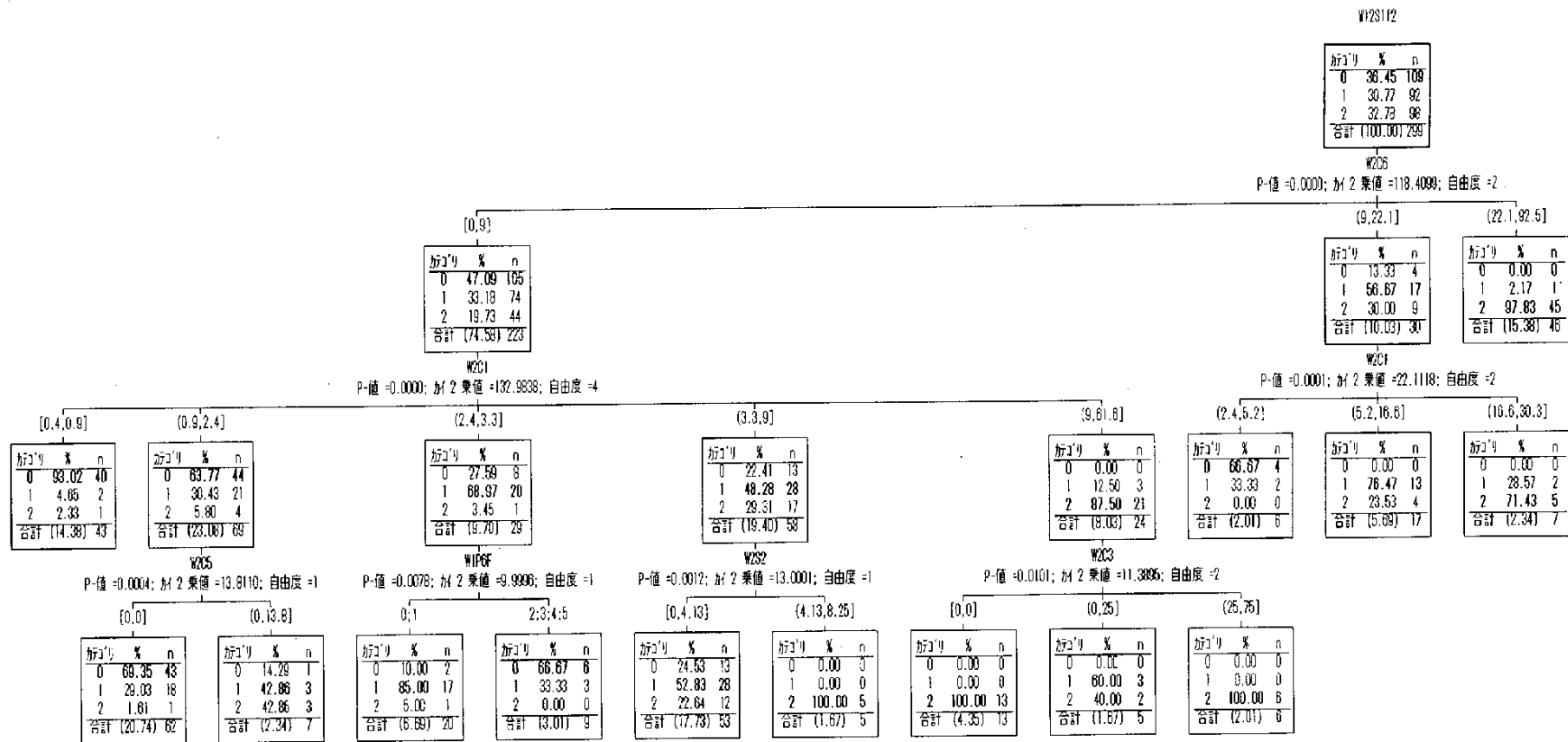


図 4-2 CHAID 出力結果 (フラグ 0 1 2 の場合)

(4) 結果

表4-3 CHAID (フラグ012の場合) の主なルールで分割された対象商品の比率

第2週首都圏カバー率

	9%以下	9~22.1%	22.1~92.5%	
C	47.1	13.3	0	
B	33.2	56.7	2.2	
A	19.7	30	97.8	

第2週首都圏カバー率9%以下

第2週全国カバー率

	0.4~0.9%	0.9~2.4%	2.4~3.3%	3.3~9%	9~61.6%	
C	93	63.8	27.6	22.4	0	
B	4.7	30.4	69	48.3	12.5	
A	2.3	5.8	3.5	29.3	87.5	

第2週首都圏カバー率9~22.1%

第2週全国カバー率

	2.4~5.2%	5.2~16.6%	16.6~30.3%	
C	66.7	0	0	
B	33.3	76.5	28.6	
A	0	23.5	71.4	

○ 発売前にメーカーが準備すべき条件

発売後12週目の全国の個数シェアがどのランクにつくことになるのか、その行方は、首都圏のカバー率によって大きく左右される。発売前の準備段階において、この項目を抑えておく必要がポイントとなりそうだ。

発売後、第2週目までに首都圏カバー率が22.1%超を確保できれば、12週目の全国シェアのAランク入りは、ほぼ100%近い確率で確実なものとなっている。これはそれほど高い目標ラインではないだろう。

首都圏でのカバー率が9~22.1%程度になると、Aランク入りの可能性は急激に下がる。しかし、全国ベースでのカバー率が5.2%超であれば、Cランクとなる可能性はまったくなく、シェアが0%になってしまう大失敗は避けられる。

首都圏でのカバー率が9%以下と想定される場合には、全国ベースでのカバー率の値によって、ヒットの度合いは変わってくる。なるべく高いカバー率を確保することが必要となるわけだが、第2週目までに全国カバー率が9%超であれば、失敗の可能性はゼロだ。しかし、2.4%を下回りそうな場合には、その時点で既に12週目の全国シェア0%という失敗の可能性が極めて高くなってしまふ。発売前からカバー率

を高めることの大切さが如実にあらわれている。2.4~9%では、半数以上がBランクになっており、Aランクよりは、Cランクに入る可能性の方が高くなるラインだ。

第2週目までに、首都圏でどれだけのスーパーで新商品をおいてもらうことができるか。これによって、12週後の全国レベルでのシェアの行方までも、ほぼ決まってしまうということだ。首都圏におけるカバー率22.1%超、この数字は発売前の入念な努力によって達成されなければならない。それを下回る場合には、全国ベースで9%超となるカバー率を維持したい。発売前から始まり、早い段階でのカバー率を高める努力の必要性は大きい。

○ 発売後の判断基準となる項目

発売後、12週目の全国の個数シェアがABC、どのランクに属することになるのか、その分かれ目となる項目は何だろうか。

第1のチェックポイントとしては、第2週目における首都圏のカバー率をみればよい。第2週首都圏カバー率が22.1~92.5%（22.1%超と同義と考えられる）であれば、過去の事例からいえば、ほぼすべての商品がAランクを獲得している。12週目全国シェアは、この時点ですでに安泰だ。またそれだけではなく、そのカバー率をクリアしていれば、Cランク、つまりシェアが0%になった商品は過去に例がない。第2週目に、首都圏で2割強のカバー率が達成されていれば、12週目の全国シェアが0.2%を達成し、Aランク入りすることはほぼ確実だ。

第2週首都圏カバー率が9~22.1%の場合には、Aランクが3割、Bランクが6割弱、Cランクが1割強となる。ただその場合でも、同週全国カバー率が5.2%超であれば、Cランクとなる可能性はなくなる。首都圏でのカバー率がそこそこでも、全国平均で5.2%を達成できていれば、シェア0%となる失敗は避けられると思っ
てよい。中でも16.6%超となっているものについては、7割以上の商品がAランクになっている。しかし5.2%を下回ると7割弱の商品がCランクに終わっている。第2週全国カバー率5.2%ラインが最低ラインとなっているようだ。

第2週首都圏カバー率が9%以下の場合には、Aランクが2割弱、Bランクが3割強、Cランクが5割弱となる。しかし、同週全国カバー率が9~61.6%（9%以上と同義と考えられる）であれば、Cランクを避けることができ、9割がたの商品がAランクに属している。第2週全国カバー率の低下とともにBランク、Cランクの比率が高まっていくことになるが、特にその分かれ目となるのが2.4%というラインだ。2.4%を下回ると急激にCランクとなる比率が高まる。つまり、首都圏でのカバー率が9%以下で、かつ全国カバー率も2.4%となるような状況では、発売後、第2週目という早い段階で、そもそも勝ち目がなく失敗の色が濃いと考えた方がいい。

以上のことからわかることは、ABCのいずれのランクに属するかの大きな決め手となるのはカバー率であるということだ。なかでも第2週目の首都圏におけるカバー率が与える影響は大きく、2割強の値を維持することが、Aランク獲得のための大切な要素となってくる。

逆にCランク入りの可能性が大きいのが、第2週首都圏カバー率が9～22.1%かつ同週全国カバー率が5.2%以下。または第2週首都圏カバー率が9%以下で、同週全国カバー率が2.4%以下の場合。このいずれかに属している時は、12週目全国シェアがゼロになることを覚悟した方がよさそうだ。

4. 3. 3 結果の比較

以上の結果を比較してわかることは、大ヒット商品となる場合（4. 3. 1の場合）と、失敗はせずにそこそこ売れる（4. 3. 2の事例）という場合には、ルールが違ってきているということだ。

ヒット商品となるか否か、ここでは発売後12週全国個数シェアが0.5%以上をおさめるかどうか、という観点に立った場合には、第2週目の北海道のカバー率と同週の全国シェアが大きく影響を与える要因とでている。

一般的に、ポテトチップスの新製品は北海道から順に投入されていく。北海道でのカバー率の高さ、つまり北海道地区でどれだけ新製品が浸透するか、認知されるかが、その後の新製品の行方を占う指標となっているというのは、このことを考えあわせると納得がいくだろう。そしてその後、当たり前ではあるが、全国シェアをある程度保つことが必要となってくる。

それに対し、そこそこ売れるかどうか、つまり発売後12週全国個数シェアをABCと3つにランクづけし、どこに属するかという観点に立った場合には、第2週目の首都圏のカバー率と同週の全国カバー率が影響を与えるとする。

このランク付けでは、Cランクにならないこと、つまり12週目のシェアがゼロにならないことをひとつの目標と考えると、そのためには、とにもかくにもカバー率をあげること、新製品を世の中に浸透させることの必要性がうかがえるだろう。なかでも、首都圏のスーパーでの浸透具合というのは、その後の売り上げに大きく関わってきている。

発売するからには、当然、ヒットをねらうわけではあるが、数々の新製品が生まれれば消えて行く中で、消えないレベルを保つことも必要といえる。ヒットをねらうには、まず北海道に力を入れる。そして、新商品の生き残りを考える上では、首都圏での浸透を考えるべきだということがわかってくる。

4. 4 C5. 0

4. 4. 1 大ヒットとなる要素を探る場合

目的変数を以下のような、カテゴリ変数であらわし、大ヒットとなる場合のルールを探す。フラグ設定は以下の通り。

0 : 0. 5未満 (ヒットせず)

1 : 0. 5以上 (大ヒット)

(1) 目的変数が0となるルール

- ・ 第2週目全国シェアが0. 27%以下
(確信度98. 7%)
- ・ 第2週目の全国シェアが0. 27%超で、第1週目中四国のカバー率が10%超、かつ第1週目九州のシェアが0. 48%以下
(確信度100%)

(2) 目的変数が1となるルール

- ・ 第2週目全国シェアが0. 27%超で、第1週目中四国カバー率が10%以下
(確信度79. 2%)
- ・ 第2週全国シェアが0. 27%超で、第1週中四国カバー率が10%超、かつ第1週九州シェアが0. 48%超
(確信度100%)

表4-4 C5. 0出力結果・ツリー (フラグ01の場合)

```
W2S1 =< 0.27 (237.0, 0.987) -> 0.0
W2S1 > 0.27
  W1C9 =< 10 (48.0, 0.792) -> 1.0
  W1C9 > 10
    W1S10 =< 0.48 (11.0, 1.0) -> 0.0
    W1S10 > 0.48 (3.0, 1.0) -> 1.0
```

表4-5 C5. 0出力結果・ルール (フラグ01の場合)

```
ルール: 0.0:
  ルール #1 : 0.0:
    if W2S1 =< 0.27
    then-> 0.0 (237.0, 0.987)

  ルール #2 : 0.0:
    if W2S1 > 0.27
    and W1C9 > 10
    and W1S10 =< 0.48
    then-> 0.0 (11.0, 1.0)

ルール: 1.0:
  ルール #1 : 1.0:
    if W2S1 > 0.27
    and W1C9 =< 10
    then-> 1.0 (48.0, 0.792)

  ルール #2 : 1.0:
    if W2S1 > 0.27
    and W1C9 > 10
    and W1S10 > 0.48
    then-> 1.0 (3.0, 1.0)

デフォルト : -> 0.0
```

(3) 結果

表4-6 C5.0 (フラグ01の場合)の主なルールで分割された対象商品の比率

第2週全国シェア

	0.27%以下	0.27%超
ヒットせず	100	17.7
ヒット	0	82.3

第2週全国シェア0.27%超

第1週中四国カバー率

	10%以下	10%超
ヒットせず	0	78.6
ヒット	100	21.4

○ 発売前にメーカーが準備すべき条件

この結果から判断すると、第2週目の全国シェアが12週目のヒットにもっとも影響を与える要因となっているため、発売前から操作するのは不可能であり、カバー率や価格など発売前から、力をいれるべき項目を特定することはできない。

○ 発売後の判断基準となる項目

12週目の全国の個数シェアが0.5%以上というヒットをおさめることができるかどうか。第1の分かれ目となるのが、第2週目の全国個数シェアが、0.27%超を確保できるかどうかだ。

確保できなかった場合、今回の対象アイテムの事例では、すべての商品が12週目全国シェア0.5%未満となっており、3ヶ月後にヒットとなる可能性はほとんどなくなる。つまり、発売後第2週目という、かなり早い段階において、全国個数シェアで0.3%程度確保できるかどうか、発売直後、いかに勢いにのるかがその後の行方を分けることとなる。

逆に確保できた場合だが、第1週中四国カバー率が10%以下の場合、ヒットとなると結果は示されている。カバー率がそれよりも高い、10%以上の場合では、78.6%の商品が過去には失敗となっている。

以上のことから言えることは、ヒット商品となるかどうかの分かれ目は、第2週目の全国個数シェアが0.27%超となっていることだ。中四国におけるカバー率の影響は奇妙な結果となっているが、新製品発売後1週目に中四国に商品が行き渡ってい

ない可能性が現実の市場では高いことが考えられる。そのため、第1週中四国カバー率の大小ではその後の全国レベルでのヒットを判断する要因とはなりえないと読むことも出来るのではないだろうか。

4. 4. 2 失敗しないための要素を見つけ出す場合

対象アイテムを目的変数（発売後12週目全国個数シェア）の値でソートし、およそ3分割しABCとランク付けを行う。そしてCランクとまらないルールを探し出す。フラグ設定は以下の通り。

- 0 : 0% (Cランク)
- 1 : 0.01~0.2%未満 (Bランク)
- 2 : 0.2%以上 (Aランク)

(1) 目的変数が0となるルール

- ・ 第2週全国カバー率が1.9%以下、第1週中四国シェアが0.3%以下
(確信度87.4%)
- ・ 第2週全国カバー率が1.9~7.6%、かつ第1週中京の価格帯によって下記のように別れる
 - 97~100円 (確信度100%) ただし第2週全国シェアが0.02%以下
 - 100~110円 (確信度100%)
 - 110~120円 (確信度75%)
- ・ 第2週全国カバー率が1.9~7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、かつ第2週東北シェアが0.09%以下の場合は第2週中四国の価格帯によって下記のように別れる
 - 100~110円 (確信度50%)
 - 140円~ (確信度100%)
- ・ 第2週全国カバー率が1.9~7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、かつ第2週中四国の価格帯も0、第1週関東外郭の価格帯が140円~
(確信度66.7%)

(2) 目的変数が1となるルール

- ・ 第2週全国カバー率が1.9%以下、第1週中四国シェアが0.3%超
- ・ 第2週全国カバー率が1.9~7.6%、かつ第1週中京の価格帯によって下記のように別れる
 - 97~100円 (確信度100%) ただし第2週全国シェア0.02%超

100～110円（確信度100%）

120～140円（確信度100%）

140円～（確信度100%）

- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなかった場合は第2週中四国の価格帯によって下記のように別れる

97～100円（確信度83.3%）

100～110円（確信度100%）ただし第1週全国カバー率2.3%超

110～120円（確信度100%）

- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、第2週中四国の価格帯も0、かつ第2週東北シェア0.09以下の場合、第1週東北の価格帯によって下記のように別れる

～97円（確信度100%）

97～100円（確信度50%）

120～140円（確信度100%）

- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、第2週中四国の価格帯も0、かつ第2週東北シェア0.09以下第1週東北の価格帯が0、第2週北海道シェアが4.03%以下の場合、第1週関東外郭の価格帯によって、下記のように分かれる。

売り上げなし（確信度60%）

～97円（確信度100%）

100～110円（確信度100%）

110～120円（確信度100%）

- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、第2週中四国の価格帯も0、かつ第2週東北シェア0.09超、第2週中京カバー率が6.7%以下

- ・第2週全国カバー率が7.6%超、第2週全国シェアが0.55%以下の場合、第2週北海道の価格帯によって下記のように分かれる

97～100円（確信度100%）

100～110円（確信度100%）

110～120円（確信度83.3%）ただし第2週中京カバー率が12.5%以下

(3) 目的変数が2となるルール

- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売

り上げがなく、第2週中四国の価格帯が～97円
(確信度100%)

- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、第2週中四国の価格帯も0、かつ第2週東北シェア0.09%以下、第1週東北の価格帯が0、第2週北海道シェアが4.03%以下、第1週関東外郭の価格帯が97～100円
- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、第2週中四国の価格帯も0、かつ第2週東北シェア0.09%以下、第1週東北の価格帯が0、第2週北海道シェアが4.03%超
- ・第2週全国カバー率が1.9～7.6%、かつ第1週中京の価格帯が0、つまり売り上げがなく、第2週中四国の価格帯も0、かつ第2週東北シェア0.09%超、第2週中京カバー率が6.7%超
- ・第2週全国カバー率が7.6%超、第2週全国シェアが0.55%以下の場合、第2週北海道の価格帯によって下記のように分かれる
売り上げなし (確信度85.7%)
～97円 (確信度100%)
110から120円 (確信度100%) ただし第2週中京カバー率が12.5%超
120～140円 (確信度100%)
140円～ (確信度66.7%)
- ・第2週全国カバー率が7.6%超、第2週全国シェアが0.55%超

表4-7 C5. 0出力結果・ツリー (フラグ012の場合)

```

W2C1 =< 7.6
  W2C1 =< 1.9
    W1S9 =< 0.3 (87.0, 0.874) -> 0.0
    W1S9 > 0.3 (5.0, 0.8) -> 1.0
  W2C1 > 1.9
    W1P7F 1.0 (3.0, 1.0) -> 0.0
    W1P7F 3.0 (0.0, 1.0) -> 1.0
    W1P7F 4.0 (4.0, 0.75) -> 0.0
    W1P7F 5.0 (1.0, 1.0) -> 1.0
    W1P7F 6.0 (2.0, 1.0) -> 1.0
    W1P7F 2.0
      W2S1 =< 0.02 (2.0, 1.0) -> 0.0
      W2S1 > 0.02 (2.0, 1.0) -> 1.0
    W1P7F 0.0
      W2P9F 1.0 (2.0, 1.0) -> 2.0
      W2P9F 2.0 (6.0, 0.833) -> 1.0
      W2P9F 4.0 (2.0, 1.0) -> 1.0
      W2P9F 5.0 (1.0, 1.0) -> 0.0
      W2P9F 6.0 (1.0, 1.0) -> 0.0
      W2P9F 3.0
        W1C1 =< 2.3 (3.0, 1.0) -> 1.0
        W1C1 > 2.3 (2.0, 1.0) -> 0.0
      W2P9F 0.0
        W2S3 =< 0.09
          W1P3F [1.0 5.0] (0.0, 1.0) -> 1.0
          W1P3F 2.0 (2.0, 0.5) -> 1.0
          W1P3F 3.0 (2.0, 0.5) -> 0.0
          W1P3F 4.0 (1.0, 1.0) -> 1.0
          W1P3F 6.0 (1.0, 1.0) -> 0.0
          W1P3F 0.0
            W2S2 =< 4.03
              W1P5F 0.0 (40.0, 0.6) -> 1.0
              W1P5F 1.0 (3.0, 1.0) -> 1.0
              W1P5F 2.0 (4.0, 0.75) -> 2.0
              W1P5F [3.0 4.0 5.0] (0.0, 1.0) -> 1.0
              W1P5F 6.0 (3.0, 0.667) -> 0.0
            W2S2 > 4.03 (3.0, 1.0) -> 2.0
          W2S3 > 0.09
            W2C7 =< 6.7 (16.0, 1.0) -> 1.0
            W2C7 > 6.7 (3.0, 0.667) -> 2.0
        W2C1 > 7.6
          W2S1 =< 0.55
            W2P2F 0.0 (28.0, 0.857) -> 2.0
            W2P2F 1.0 (5.0, 1.0) -> 2.0

```

表4-7 C5.0出力結果・ツリー(続き)

W2P2F 2.0 (5.0, 1.0) -> 1.0
W2P2F 3.0 (2.0, 1.0) -> 1.0
W2P2F 5.0 (4.0, 1.0) -> 2.0
W2P2F 6.0 (3.0, 0.667) -> 2.0
W2P2F 4.0
 W2C7 =< 12.5 (6.0, 0.833) -> 1.0
 W2C7 > 12.5 (4.0, 1.0) -> 2.0
W2S1 > 0.55 (41.0, 1.0) -> 2.0

表4-8 C5.0出力結果・ルール(フラグ012の場合)

ルール: 0.0:

ルール #1 : 0.0:
 if W2C1 <= 7.6
 and W1S9 <= 0.3
 then-> 0.0 (87.0, 0.874)

ルール #2 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 1.0
 then-> 0.0 (3.0, 1.0)

ルール #3 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 4.0
 then-> 0.0 (4.0, 0.75)

ルール #4 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 2.0
 and W2S1 <= 0.02
 then-> 0.0 (2.0, 1.0)

ルール #5 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 0.0
 and W2P9F == 5.0
 then-> 0.0 (1.0, 1.0)

ルール #6 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 0.0
 and W2P9F == 6.0
 then-> 0.0 (1.0, 1.0)

ルール #7 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 0.0
 and W2P9F == 3.0
 and W1C1 > 2.3
 then-> 0.0 (2.0, 1.0)

ルール #8 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 <= 0.09
 and W1P3F == 3.0
 then-> 0.0 (2.0, 0.5)

ルール #9 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 <= 0.09
 and W1P3F == 6.0
 then-> 0.0 (1.0, 1.0)

ルール #10 : 0.0:
 if W2C1 > 1.9
 and W2C1 <= 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 <= 0.09
 and W1P3F == 0.0
 and W2S2 <= 4.03
 and W1P5F == 6.0
 then-> 0.0 (3.0, 0.667)

表4-8 C5. 0出力結果・ルール (続き)

ルール: 1.0:

ルール #1 : 1.0:
 if W2C1 =< 7.6
 and W1S9 > 0.3
 then-> 1.0 (5.0, 0.8)

ルール #2 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 3.0
 then-> 1.0 (0.0, 1.0)

ルール #3 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 5.0
 then-> 1.0 (1.0, 1.0)

ルール #4 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 6.0
 then-> 1.0 (2.0, 1.0)

ルール #5 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 2.0
 and W2S1 > 0.02
 then-> 1.0 (2.0, 1.0)

ルール #6 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 2.0
 then-> 1.0 (6.0, 0.833)

ルール #7 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 4.0
 then-> 1.0 (2.0, 1.0)

ルール #8 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 3.0
 and W1C1 =< 2.3
 then-> 1.0 (3.0, 1.0)

ルール #9 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == [1.0 5.0]
 then-> 1.0 (0.0, 1.0)

ルール #10 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 2.0
 then-> 1.0 (2.0, 0.5)

ルール #11 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 4.0
 then-> 1.0 (1.0, 1.0)

表4-8 C5. 0出力結果・ルール(続き)

ルール #12 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 0.0
 and W2S2 =< 4.03
 and W1P5F == 0.0
 then-> 1.0 (40.0, 0.6)

ルール #13 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 0.0
 and W2S2 =< 4.03
 and W1P5F == 1.0
 then-> 1.0 (3.0, 1.0)

ルール #14 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 0.0
 and W2S2 =< 4.03
 and W1P5F == [3.0 4.0 5.0]
 then-> 1.0 (0.0, 1.0)

ルール #15 : 1.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 > 0.09
 and W2C7 =< 6.7
 then-> 1.0 (16.0, 1.0)

ルール #16 : 1.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 2.0
 then-> 1.0 (5.0, 1.0)

ルール #17 : 1.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 3.0
 then-> 1.0 (2.0, 1.0)

ルール #18 : 1.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 4.0
 and W2C7 =< 12.5
 then-> 1.0 (6.0, 0.833)

表4-8 C5.0出力結果・ルール(続き)

ルール: 2.0:

ルール #1 : 2.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 1.0
 then-> 2.0 (2.0, 1.0)

ルール #2 : 2.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 0.0
 and W2S2 =< 4.03
 and W1P5F == 2.0
 then-> 2.0 (4.0, 0.75)

ルール #3 : 2.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 =< 0.09
 and W1P3F == 0.0
 and W2S2 > 4.03
 then-> 2.0 (3.0, 1.0)

ルール #4 : 2.0:
 if W2C1 > 1.9
 and W2C1 =< 7.6
 and W1P7F == 0.0
 and W2P9F == 0.0
 and W2S3 > 0.09
 and W2C7 > 6.7
 then-> 2.0 (3.0, 0.667)

ルール #5 : 2.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 0.0
 then-> 2.0 (28.0, 0.857)

ルール #6 : 2.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 1.0
 then-> 2.0 (5.0, 1.0)

ルール #7 : 2.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 5.0
 then-> 2.0 (4.0, 1.0)

ルール #8 : 2.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 6.0
 then-> 2.0 (3.0, 0.667)

ルール #9 : 2.0:
 if W2C1 > 7.6
 and W2S1 =< 0.55
 and W2P2F == 4.0
 and W2C7 > 12.5
 then-> 2.0 (4.0, 1.0)

ルール #10 : 2.0:
 if W2C1 > 7.6
 and W2S1 > 0.55
 then-> 2.0 (41.0, 1.0)

デフォルト : -> 0.0

(4) 結果

表4-9 C5. 0 (フラグ012の場合) の主なルールで分割された対象商品の比率

第2週全国カバー率

	7.6%以下	7.6%超	
C	52.7		0
B	41.3		13.3
A	6		86.7

第2週全国カバー率7.6%以下

第2週全国カバー率

	1.9%以下	1.9%超	
C	94.6		17.4
B	5.4		71.6
A	0		11

第2週全国カバー率7.6%超

第2週全国シェア

	0.55%以下	0.55%超	
C	0		0
B	22.8		0
A	77.2		100

○ 発売前にメーカーが準備すべき条件

発売後12週目の全国の個数シェアがどのランクにつくことになるのか、その行方は、全国のカバー率を7.6%以上に維持することができるかどうかにかかっている。

これをクリアできれば、12週目の全国シェアAランク入りは、この時点で9割弱の確率で確実のものとなる。また、Cランクになる可能性はゼロとなり、シェア0%は回避したとあってよい。1割弱のスーパーに、新製品を確実に置くことが、ヒットを分ける大きな要素となっているのだ。

さらにはそれ以下であっても、1.9%ラインを確保できる保証があるならば、8割がたの商品がBランク以上になると期待することが出来る。Cランクの可能性は2割弱と低い。

逆に1.9%ラインを下回りそうな場合には、Cランクになる確率は94.5%だ。この時点で、そもそも勝ち目がないと思ったほうが良いだろう。

第2週目までに全国のカバー率を7.6%超確保する、全国の1割弱のスーパーに新製品をおいてもらう、このそれほど高くはないと思われる目標がAランク入りの大きな決め手となるのだ。逆に、最低でも1.9%ラインを下回らないことが、シェア

0%という失敗を回避するための絶対的に必要な条件となっている。

○ 発売後の判断基準となる項目

発売後、12週目の全国個数シェアがABC、どのランクに属することになるのか、その分かれ目はどこになるのだろうか。

第1のチェックポイントは、第2週全国カバー率が7.6%を上回っているかどうかにある。今回の対象アイテムでは、このラインを上回れば9割がたの商品が、発売後12週目の全国シェアAランクを獲得している。しかもCランクに属する可能性はゼロとなり、12週目シェア0%となる可能性はまったくなくなる。

さらにその中でも、第2週全国シェアが0.55%を超えていれば、100%の確率で、Aランクを獲得することができている。たとえ、それ以下であっても、Aランクの可能性は、まだ8割弱ある。カバー率の威力は大きい。

逆に7.6%を下回った場合、状況は大きく異なってくる。過去の事例では、半数以上の商品がCランク、つまり12週後にはシェア0%となっている。

さらに詳細を見ていくと、同項目が1.9%ラインを維持できるかどうか、次の分かれ目となっている。1.9%を超えていれば、Cランクの可能性は2割弱まで小さくなっている。しかし1.9%を下回っている場合には、95%近い商品がCランクとなる。第2週全国カバー率1.9%、これがCランクを回避するための、最低限のラインとなりそうだ。

以上のことから言えることは、ABCいずれのランクに属するかを決め手となるのは、カバー率だということだ。

全国のカバー率を発売2週間以内に7.6%以上に高めれば、ほぼAランク。また1.9%ラインを切ると、Cランク入りが確実となってしまふ。それほど難しい数値目標ではないと思われるが、これが12週目のシェアに及ぼす影響の大きさには驚かされる。

4. 4. 3 結果の比較

結果を比較してわかることは、ヒット商品となる場合（4. 4. 1の事例）と、失敗はせずにそこそこ売れる（4. 4. 2の事例）という場合では、CHAIDの場合と同様に、ルールが違ってきているということだ。

ヒット商品となるか否か、ここでは発売後12週目の全国個数シェアが0.5%以上をおさめるかどうか、という観点に立った場合には、第2週目の全国個数シェアが、大きく影響を与える要因とでている。

これは、発売後2週目でシェアを、ある程度維持できていない商品は、その後3カ月たってシェアを挽回できるチャンスはほとんどないということを意味している。発売直後のシェアの確保がその後に与える意味は大きいのだ。

それに対し、そこそこ売れるかどうか、つまり発売後12週目の全国個数シェアをABCと3つにランクづけし、どこに属するかという観点に立った場合には、第2週目の全国カバー率が影響を与えるとでている。

このランク付けでは、Cランクにならないこと、つまり12週目のシェアがゼロにならないことをひとつの目標としてとらえると、そのためには、やはりカバー率をあげることに、新製品を世の中に浸透させることが必要となっている。しかもその値は、全国平均のカバー率で1割弱という値でよいのだ。

ヒットをねらうには、やはり全国レベルでのシェアの確保が求められる。そして、新商品の生き残りを考える上では、カバー率、つまりいかに新製品を浸透させるかを考えるべきだということがわかってくる。

4. 5 CHAID、C5. 0による結果の比較

4. 5. 1 大ヒットとなる要素を探る場合

CHAIDとC5. 0、それぞれによるルール第1、2にくる主な分岐は以下の通りであった。

CHAID

第1分岐 第2週北海道カバー率

第2分岐 第2週全国シェア

C5. 0

第1分岐 第2週全国シェア

第1分岐となるのがCHAIDでは第2週北海道カバー率、C5. 0では第2週全国シェアである点を見ると、まったく異なる結果となったように見える。

しかしCHAIDの第2分岐は、いずれも第2週全国シェアとなっている。第2分岐を細かく見ていくと、4. 3. 1でも述べたように、第1分岐こそカバー率であるものの、第2分岐である第2週全国シェアの大小によって、ヒットか否かは大きく左右されている。

となると、大ヒットとなるかどうか、発売後12週目に全国個数シェア0. 5%超を獲得できるかどうかは第2週の全国シェアにかかっている、という点で、両ツールにおいて、同様の結果が出ているといえることができる。発売直後のシェアによって、12週目のシェア、ヒットの行方もほぼ決まってしまうと考えてよいようだ。

4. 5. 2 失敗しないための要素を見つけ出す場合

CHAIDとC5.0、それぞれによるルール第1、2にくる主な分岐は以下の通りであった。

CHAID

第1分岐 第2週首都圏カバー率
第2分岐 第2週全国カバー率

C5.0

第1分岐 第2週全国カバー率
第2分岐 第2週全国カバー率

第1分岐となるのが、CHAIDでは第2週首都圏カバー率、C5.0では第2週全国カバー率となっている。

第2分岐を見ると、CHAIDでは第2週全国カバー率、つまりC5.0の第1分岐がきている。C5.0では第1分岐と同じく第2週全国カバー率だ。

12週目の個数シェアがどのランクに属するかという点から見ると、両ツールともに、カバー率が影響の大きい要素としてあがっており、共通点がみられる。

さらにここで、この分析の目的が、失敗しないため、つまりCランクにならないためである点に着目する。

CHAIDでは、第1分岐である第2週首都圏カバー率が9%以下の場合に半数近くがCランクとなっており、かつ第2週全国カバー率が2.4%以下であると、急激にCランクとなる確率が高まっている。C5.0では、第2週全国カバー率が1.9%以下の場合に9割以上のアイテムがCランクとなっている。

両ツールの結果を比較すると、どうやら失敗しないため、発売後シェアが0%にならないためには、第2週までに全国カバー率2%前後を達成することが最低の条件となるようだ。

4. 5. 3 まとめ

以上のように、個別に見ていくと、CHAID、C5. 0の両ツールでは結果に差が出てくる。しかし、大ヒットを狙う場合のもっとも重要な指標は「第2週全国個数シェア」、失敗しないためには「第2週全国カバー率」と、両ツールにおいて同様の見解をえることができたと言えるだろう。

ツールによって、さまざまな違いが出た部分については、ツールの特徴の認識を深めるとともに、その原因の究明を今後の課題としていきたいと思う。

また、季節限定商品、地域限定商品など、今回、考慮にいれなかった要素についても分析のやり方をさらにつめていくことが必要となる。

今後、さらに他のカテゴリーについても、この分析をすすめ、新製品をヒットに導く構造を明らかにしていきたい。

5. ニューラルネットワークによる売れ行き要因分析

スーパーマーケットでは様々な商品が売られている。しかも、日々価格が変動している。昨日はA商品が198円、B商品が228円、今日はA商品が238円、B商品が188円といった具合である。毎日のように価格が変動する商品の間の競合状況はどうなっているのでしょうか？今まで経験や勘でいわれていたことがデータで裏付けられるでしょうか？

カテゴリー分析の第一歩として、POSデータを用いて分類内の商品別競合関係、価格が売りに上げに与える影響をデータマイニングの手法を使って調べた。

5.1 使用データ

日本経済新聞社のPOSデータベース「NEEDS-SCAN」から首都圏3店舗の毎日の売上データを用いた。3店舗はそれぞれ別のチェーンでA店は食品スーパー、B店、C店は総合スーパーである。この3店の1年間（1998年4月1日から1999年3月31日まで）の日次データを使った。今回分析したカテゴリーは「焼き肉のたれ」である。

表5-1 焼き肉のたれ商品一覧

商品名	変数名	A店		B店		C店	
		販売日数		販売日数		販売日数	
エバラ 焼肉のたれ 醤油味 瓶 300G	001165	342	○	330	○	295	○
エバラ 焼肉のたれ 甘口 瓶 300G	001172	294	○	28	×	0	
エバラ 焼肉のたれ 辛口 瓶 300G	001189	0		25	×	0	
エバラ 黄金の味 甘口フルーツタイプ瓶210G	002001	311	○	306	○	314	○
エバラ黄金の味中辛フルーツ&スパイス瓶210G	002018	344	○	302	○	320	○
エバラ 黄金の味 辛口スパイスタイプ瓶210G	002025	273	○	235	○	0	
エバラ 黄金の味 甘口フルーツタイプ瓶400G	002100	0		104	×	192	○
エバラ黄金の味中辛フルーツ&スパイス瓶400G	002117	0		260	○	264	○
エバラ 黄金の味 辛口 スパイス 瓶 400G	002124	0		0		197	○
エバラ おろしのたれ 焼肉のたれ 瓶 270G	002360	303	○	332	○	328	○
エバラ 焼肉のたれ 韓国風醤油味 瓶 180G	003183	0		0		260	○
エバラ 焼肉のたれスパイスにんにく味瓶180G	003206	0		0		230	○
エバラ 生姜焼のたれ 瓶 180G	003220	280	○	247	○	256	○
エバラ 粒ごまのたれ 豊潤 瓶 280G	003336	72	×	226	○	195	○
エバラ 焼肉帝王 うまくち醤油超特選瓶220G	003640	0		148	×	78	×
エバラ焼肉帝王コチュジャン醤油超特選瓶220G	003657	0		86	×	0	
オリバー 六甲 本場神戸のたれ 中辛瓶220G	812122	0		0		131	×
イカリ しょうが焼のたれ しょうゆプラ150G	394845	0		0		67	×
アサムラサキ生だれ焼き肉のたれ無添加プラ280G	015087	0		0		9	×
キッコーマン新撰焼肉黒だれやや甘口プラ400G	341489	0		0		178	○
キッコーマン 新撰焼肉 赤と黒 プラ 400G	342189	0		0		242	○

表5-1 焼き肉のたれ商品一覧(続き)

キッコーマン新撰焼肉本場韓国風もみだれ中辛240G	342226	0		11	×	24	×
サンエースれんが亭焼肉のたれしょうゆ味瓶190G	051005	0		0		47	×
日ハム 焼肉のたれ辛口ファミリーサイズ570G	223946	0		0		1	×
ハウスさっと逸品焼肉ドレすっきりレモン醤油300ML	480618	143	×	13	×	155	○
ハウスさっと逸品焼肉ドレまろやかごま醤油300ML	480748	0		0		172	○
富士甚 鄙 湯布院焼肉のたれこく味プラ300G	420437	0		0		53	×
富士甚 ヤキニク 塩だれ レモン 瓶150ML	420505	0		0		46	×
ブルドック しょうが焼用ソース 3人前100G	165466	6	×	0		0	
モランボンあっさり焼肉菜味おろしぼん酢プラ225G	330136	0		0		31	×
モランボンジャンコチュジャン味牛焼肉レトルト80G	330204	273	○	0		0	
モランボン 生姜焼きのたれ プラ 190G	340166	185	○	0		0	
盛田 麻布十番三幸園焼肉のたれ 230G	450021	0		109	×	0	
日本食研 晚餐館牛と野菜の鉄板鍋のたれプラ520G	137740	0		0		30	×
日本食研晚餐館生姜焼きのたれすり生姜入プラ210G	210375	0		0		31	×
日本食研 晚餐館 焼肉のたれ 甘口 210G	212171	0		95	×	0	
日本食研 焼肉のたれ 900G	212256	0		0		6	×
日本食研 晚餐館 焼肉のたれ 中辛 210G	212270	0		203	○	48	×
ダイショー焼肉一番甘口パーティーボトルプラ590G	010485	0		0		6	×
ダイショー カルビクイ韓国風焼肉のたれ紙P340G	010638	0		65	×	0	
ダイショー 焼肉のたれ 甘口 プラ 240G	018276	0		0		51	×
ダイショー 焼肉のたれ 中辛 プラ 240G	018283	0		0		36	×
ダイショー 焼肉のたれ 得用 PET1.2KG	018290	0		0		33	×
ダイショー 生姜焼のたれ プラ 170G	018320	0		0		54	×
ダイショー唐辛子味噌コチュジャンベース焼肉用50G	060015	78	×	0		0	
サンダイナー プルコギャン韓国焼肉たれ辛口190G	259629	0		10	×	0	
サンダイナー プルコギャン韓国焼肉たれ甘口190G	259636	0		7	×	0	
有紀 焼肉みそダレ サムジャン 130G	434200	0		0		45	×
キッコーマン 新撰焼肉 赤と黒 プラ 210G	467574	288	○	309	○	300	○
キッコーマン新撰焼肉黒ごましょうゆ味プラ210G	467727	6	×	0		7	×
キッコーマン新撰焼肉和風香味だれ青じそ下し210G	645101	76	×	0		91	×
キッコーマン 新撰焼肉 黒だれ プラ 210G	645118	247	○	299	○	249	○
キッコーマン新撰焼肉和風香味だれねぎ生姜210G	645125	67	×	0		172	○
キッコーマン 新撰焼肉 赤だれ プラ 210G	645132	104	×	0		0	
キッコーマン新撰焼肉ねぎしょうが味プラ210G	645453	4	×	13	×	9	×
柏原 金龍焼肉のたれ 210G	000100	0		0		40	×
サンダイナー ルコギャン韓国焼肉たれ中辛400G	000440	0		14	×	14	×
サンダイナーブルコギャン韓国焼肉たれ辛口400G	040231	201	○	0		0	
チョコー 焼肉のたれ 瓶 200ML	541068	0		0		68	×
徳山大阪鶴橋焼肉のタレみそ味手造り風瓶230G	200193	0		0		39	×
徳山大阪鶴橋焼肉のタレしょうゆ味手造り風瓶230G	200209	0		0		148	×
松屋栄 牛焼肉のたれ 専門店用 430G	070793	0		339	○	0	
松屋栄 専門店の味 焼肉のたれ 500ML	105129	0		225	○	0	

この1年間に登場した焼き肉のたれは64商品であるが、各店ごとに品揃えはかなり異なる(表5-1)。食品スーパーのA店では21商品、総合スーパーのB店では27商品、C店では46商品が販売された。表5-1の販売日数が0の商品はその店では1日も販売実績がなかったことを示す。販売実績があった商品を見ると、3店とも毎日売れた商品は一品もない。また、販売日数から定番商品とその他の商品に分かれていることがわかる(表5-2)。店の規模により取扱商品数は異なるが、定番商品の数は各店ともほぼ同じである。3店共通して250日以上販売実績があるのは5商品に過ぎない。

今回は定番3商品(「エバラ 黄金の味 甘口 フルーツタイプ 瓶210G」、
「エバラ おろしのたれ 焼肉のたれ 瓶 270G」、
「キッコーマン 新撰焼肉 赤と黒 プラ 210G」)の販売個数を被説明変数に、他の商品の値引率、来店客数、休日を説明変数にして分析した。

表5-2 販売日数別商品数

店舗名(営業日数)	A店(351)	B店(341)	C店(348)
販売実績250日以上	9	8	8
販売実績150日以上	3	5	10
販売実績150日未満	9	14	28
総商品数	21	27	46

5.2 データ変換

POSデータでは販売実績がない限り、データは発生しない。このため、販売実績がない日のデータはそのままでは欠損値となる。実際には棚にあって売れなかったためにデータが発生していないのであるから、欠損値にはすべてデータを代入した。欠損値データには販売個数0、価格は前販売日の価格を代入した。ただし、分析期間の始め、及び終わりに連続して販売実績がない場合は新製品、あるいは棚からはずされた死に筋商品として欠損値のままにした。さらに、販売日数が非常に少ない商品はスポット商品と考えられるので、欠損値のままにした(結果的にこのように販売日数が少ない商品は分析対象からはずした。表5-1の×印の商品)。

価格は商品によって大きく異なるため値引率を採用した。これは同じ50円引きでも200円の商品と500円の商品では重みが異なると考えたからである。値引率の計算に当たっては、分析対象の1年間の日々価格データから最高価格をもとめ、その最高価格から値引率を算出した。

分析した3商品の販売個数と値引率の関係を各店舗ごとに図5-1に示す。グラフの横軸は値引率、縦軸は販売個数である。どのデータも突出して売れている日(いわゆる特異点)がある。この特異点は分析に大きな影響を与えるが、

今回はPOSデータの実体をそのまま使うこととし、特異点に対して特別なデータ操作は行わなかった。

販売日数が少ない商品は分析の対象から外した。影響が少ないと考えられるからである。今回の分析では販売実績が150日以上ある商品だけを対象とした。表5-1の×印の商品を分析対象から外し、○印の商品だけを使って分析をした。

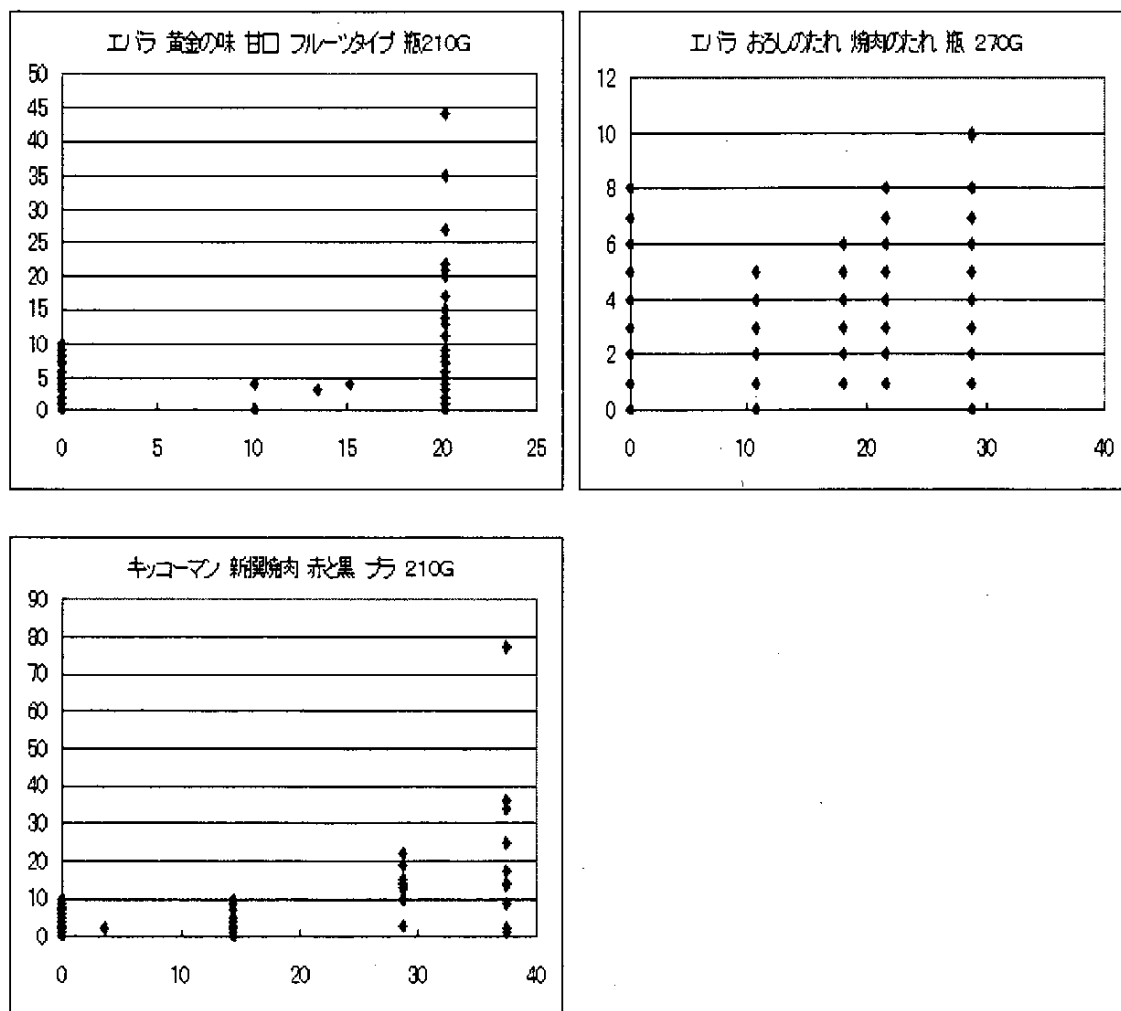


図5-1 A 販売個数と値引率 (A店、横軸：値引率、縦軸：販売個数)

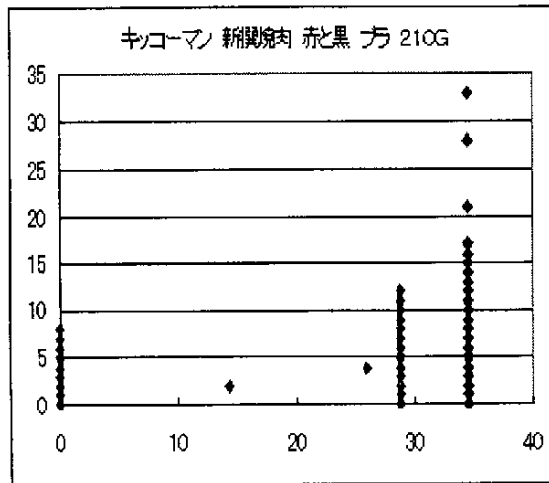
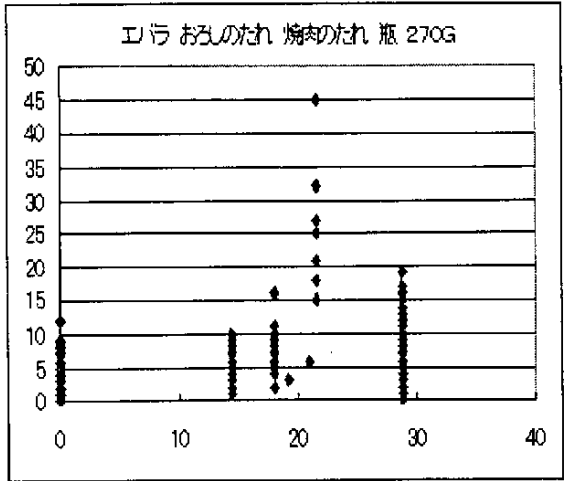
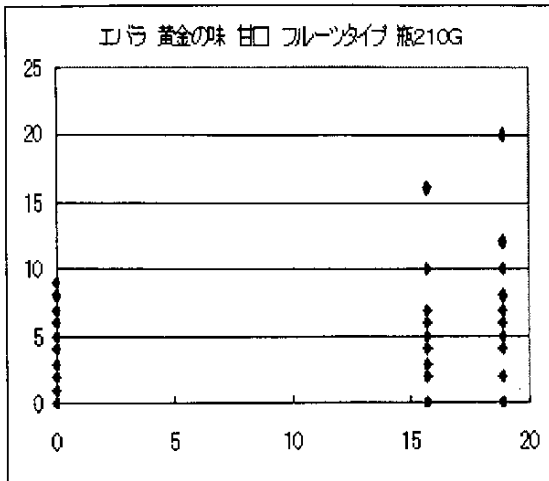
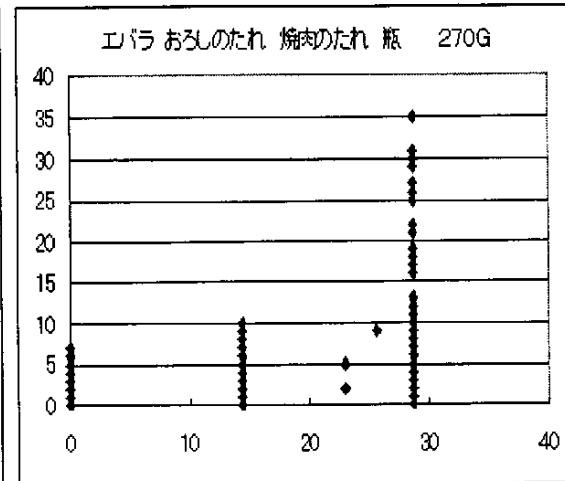
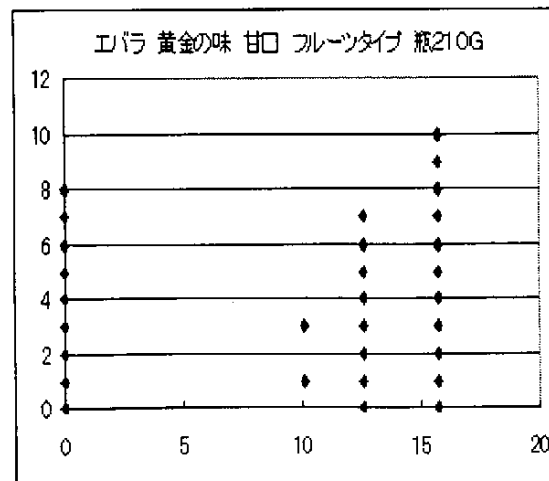


図5-1B 販売個数と値引率 (B店、横軸：値引率、縦軸：販売個数)



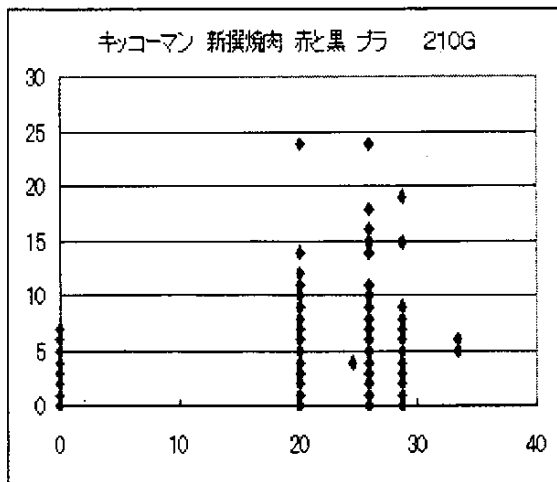


図5-1C 販売個数と値引率 (C店、横軸：値引率、縦軸：販売個数)

5. 3 分析手法

エス・ピー・エス・エス株式会社 (以下 SPSS 社) のデータマイニング用ソフト、「Clementine」のニューラルネットワークを使って分析した。また、ニューラルネットワークの結果を理解する助けとするために SPSS 社の「Answer Tree」を使って同様の分析を行い結果を比べた。

Clementine のニューラルネットワークによるデータマイニングでは1998年4月1日から1999年2月28日までの11カ月間のデータを学習用データとし、1999年3月のデータをそのモデルの検証に使った。ニューラルネットワークでは入力出力とも0から1の範囲の値を取るとき、最適な働きをする。Clementine のニューラルネットワークは、そのアルゴリズムの中で自動的にすべての入力出力の値を0から1の範囲に変換して分析を行う。しかし、事前に入力出力を0から1の範囲に変換した分析と元のデータをそのまま使った分析とでは結果が異なることがあった。このため、同一のデータで、データをそのまま使った分析と0から1の範囲に変換した分析の2通りの分析を行った。この報告書では2通りの分析で結果があまり大きく違わない時は一方の結果のみを掲載した。

Clementine では、自動的に枝刈りをするモードでニューラルネットワークを実行した。その他のオプションはデフォルトで実行した。Answer Tree では CHAID のアルゴリズムを使った。

なお、POSデータは本来、時系列データであるが、本分析では時系列データとしては取り扱わなかった。

5. 4 分析結果

定番3商品 (「エバラ 黄金の味 甘口 フルーツタイプ 瓶210G」、
「エバラ おろしのたれ 焼肉のたれ 瓶 270G」、
「キッコーマン 新撰焼肉 赤と黒 プラ 210G」) の販売個数を被説明変数に、他の商品の

値引率、来店客数、休日を説明変数にして分析した。

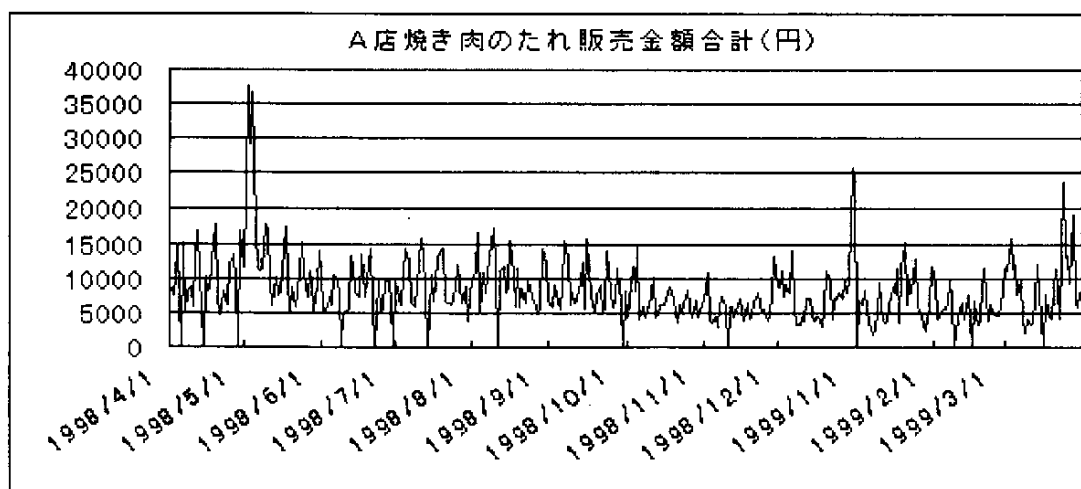
各商品には表5-1の変数名を与え、先頭にメーカー名の先頭のアルファベット1文字を追加した。末尾のKは個数、Pは値引率、Vは0から1に変換した個数、Xは0から1に変換した値引率、Custは来店客数、CustXは0から1に変換した来店客数、HoliFは休日フラグである。「エバラ 黄金の味 甘口 フルーツタイプ 瓶210G」を例にとると次のようになる。

E002001	エバラ 黄金の味 甘口 フルーツタイプ 瓶210G
E002001K	販売個数
E002001P	値引率
E002001V	0から1の範囲に変換した販売個数
E002001X	0から1の範囲に変換した値引率
Cust	来店客数
CustX	0から1の範囲に変換した来店客数
HoliF	休日フラグ (土日曜日、祝祭日を休日1他の曜日を0)

*先頭に\$Nが付くのはニューラルネットワークによる予測値

図表の変数はすべて上記の規則に従っている。具体的な商品名は表5-1を参照されたい。

始めにこの1年間における3店での焼き肉のたれの販売金額の推移を図5-2に示す。販売金額が0の日は休店日である。



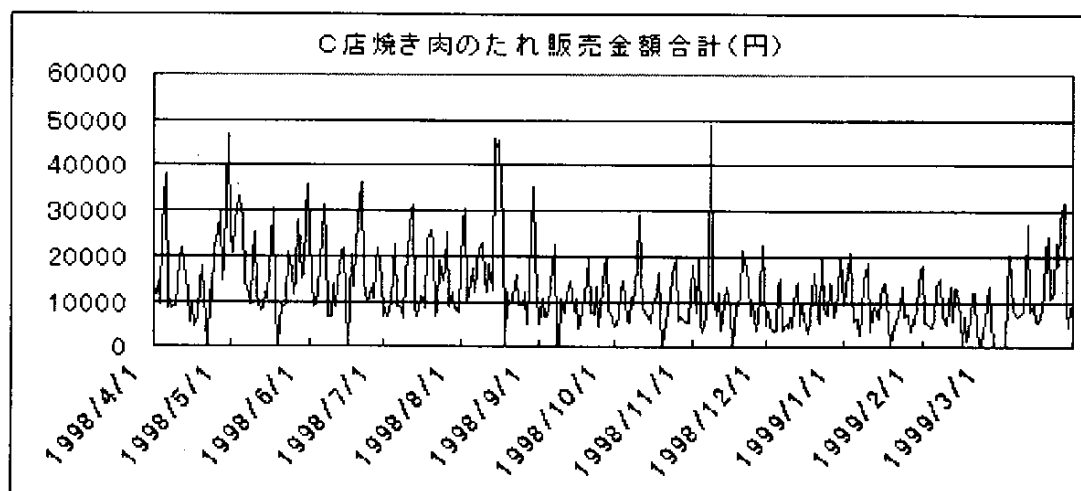
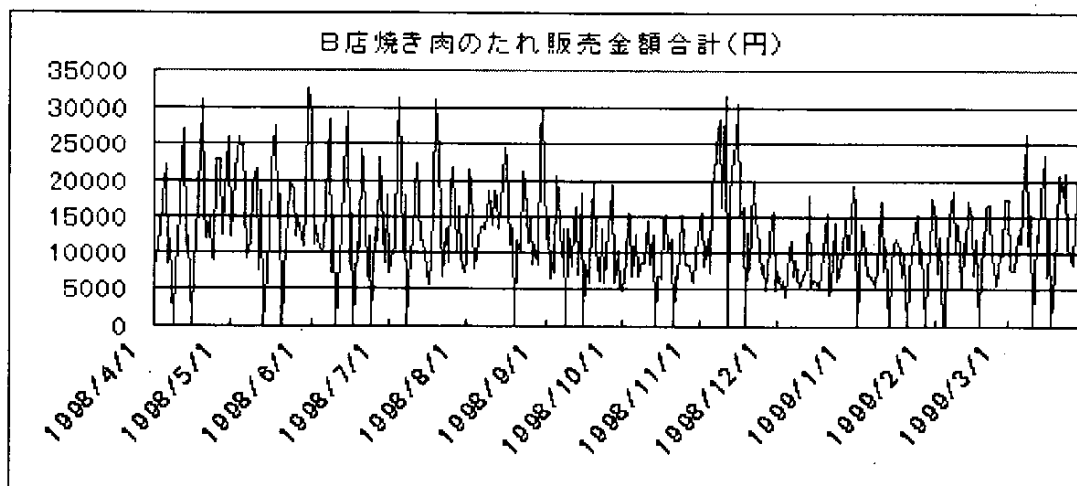


図5-2 「焼き肉のたれ」販売金額

趨勢として春から夏にかけて販売金額が大きく、秋から冬にかけて販売金額が小さい。しかし、それほど大きな差ではない。もう少し小さな周期で見ると土日曜日に販売金額が伸びる周期性を持っている。また、何日か特別にたくさん売れた日、いわゆる特異日を持っている。3店とも同じような販売金額のトレンドをしている。

「エバラ 黄金の味 甘口 フルーツタイプ 瓶210G」、「エバラ おろしのたれ 焼肉のたれ 瓶 270G」、「キッコーマン 新撰焼肉 赤と黒 プラ 210G」の3商品について、商品別にA店、B店、C店の3店舗でどのように売れているか、競合商品はあるか、についてデータマイニングをした結果を示す。以下のグラフでCountと日付の関係は以下の通り。

Count	1	60	120	180	240	300
日付(A店)	1998/4/1	1998/6/1	1998/8/5	1998/10/6	1998/12/6	1999/2/5
日付(B店)	1998/4/1	1998/6/2	1998/8/6	1998/10/8	1998/12/11	1999/2/15
日付(C店)	1998/4/1	1998/5/31	1998/8/1	1998/10/3	1998/12/4	1999/2/3

<エバラ 黄金の味 甘口 フルーツタイプ 瓶210G>

A店

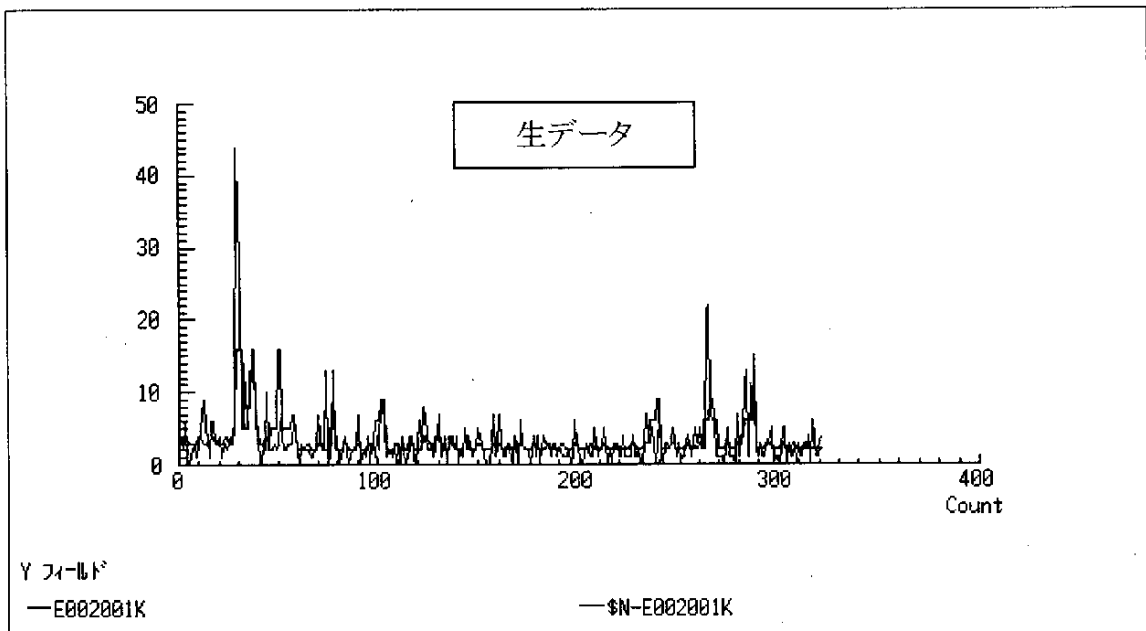
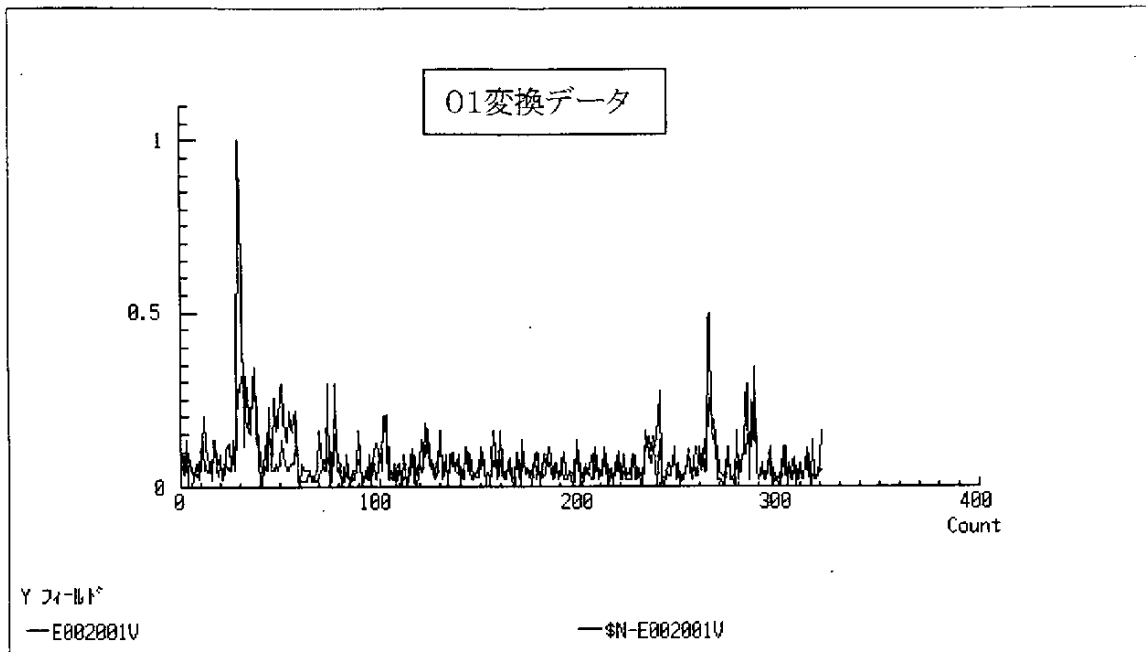


図5-3 ニューラルネットワークによる学習結果、実績値との比較 (A店)

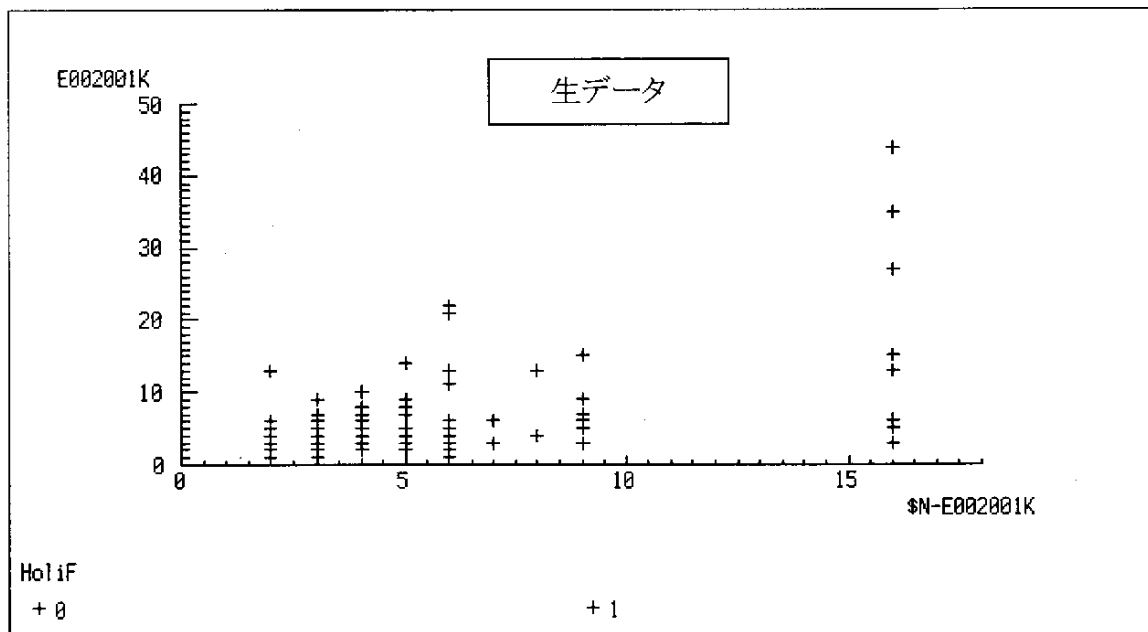
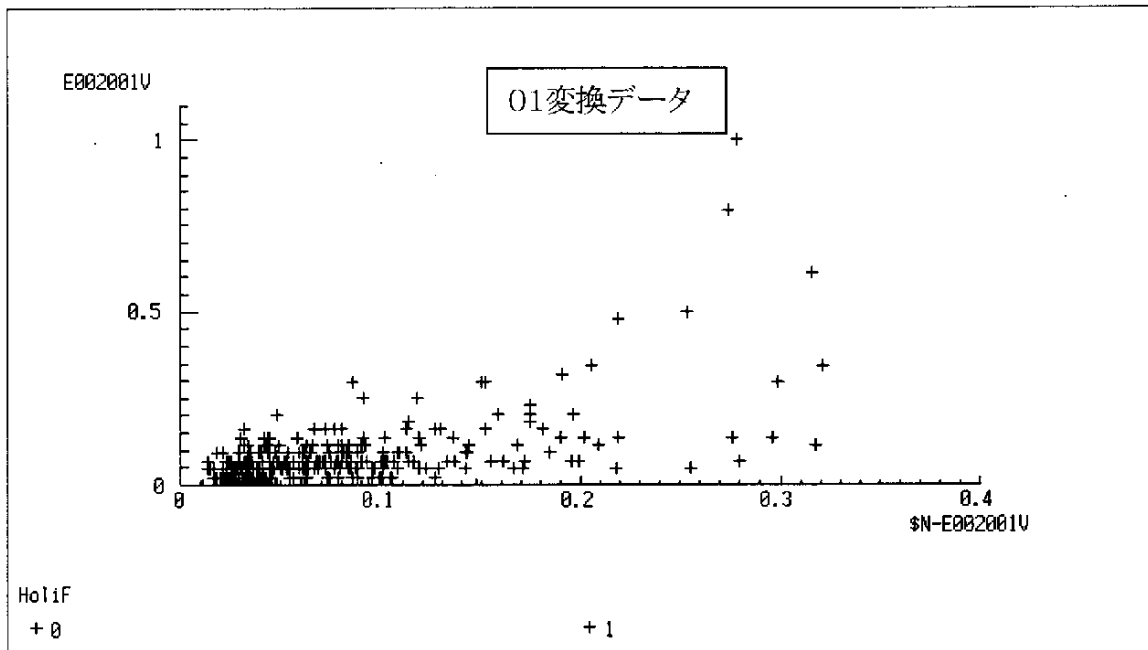


図5-4 ニューラルネットワークによる学習結果、実績値との比較 (A店)

表5-3 重要度分析 (01変換データ)

ニューラル・ネットワーク・アーキテクチャ: "E002001V"	
入力層	: 12 ニューロン
隠れ層 #1	: 12 ニューロン
出力層	: 1 ニューロン
予想正確度	: 95.47%

表5-4 精度分析 (01変換データ)

出力フィールドの結果: E002001V	
\$N-E002001V のエラー分析	
エラーの最小値	: -0.21169
エラーの最大	: 0.72126
エラーの平均値	: 0.0017071
絶対エラーの平均値	: 0.045162
標準偏差	: 0.078801

表5-3 重要度分析続き

入力の関係重要度	
CustX	: 0.08220
HoliF	: 0.05268
E002001X	: 0.04386
E002025X	: 0.02817
E002018X	: 0.02640
E002360X	: 0.02545
M330204X	: 0.02335
K467574X	: 0.01940
E001172X	: 0.01903
E001165X	: 0.01759
S040231X	: 0.01418
E003220X	: 0.00857

表5-4 精度分析続き

線形相関	: 0.61424
出現頻度	: 321

表5-5 重要度分析 (生データ)

ニューラル・ネットワーク・アーキテクチャ: "E002001K"	
入力層	: 6 ニューロン
隠れ層 #1	: 4 ニューロン
出力層	: 1 ニューロン
予想正確度	: 95.54%
入力の関係重要度	
E002001P	: 0.10589
HoliF	: 0.05614
E003220P	: 0.04624
M340166P	: 0.04072
E002360P	: 0.03222
E001172P	: 0.02101

表5-6 精度分析 (生データ)

出力フィールドの結果: E002001K	
\$N-E002001K のエラー分析	
エラーの最小値:	-13
エラーの最大値:	28
エラーの平均値:	-0.049844
絶対エラーの平均値:	1.9626
標準偏差:	3.3788
線形相関:	0.64197
出現頻度:	321

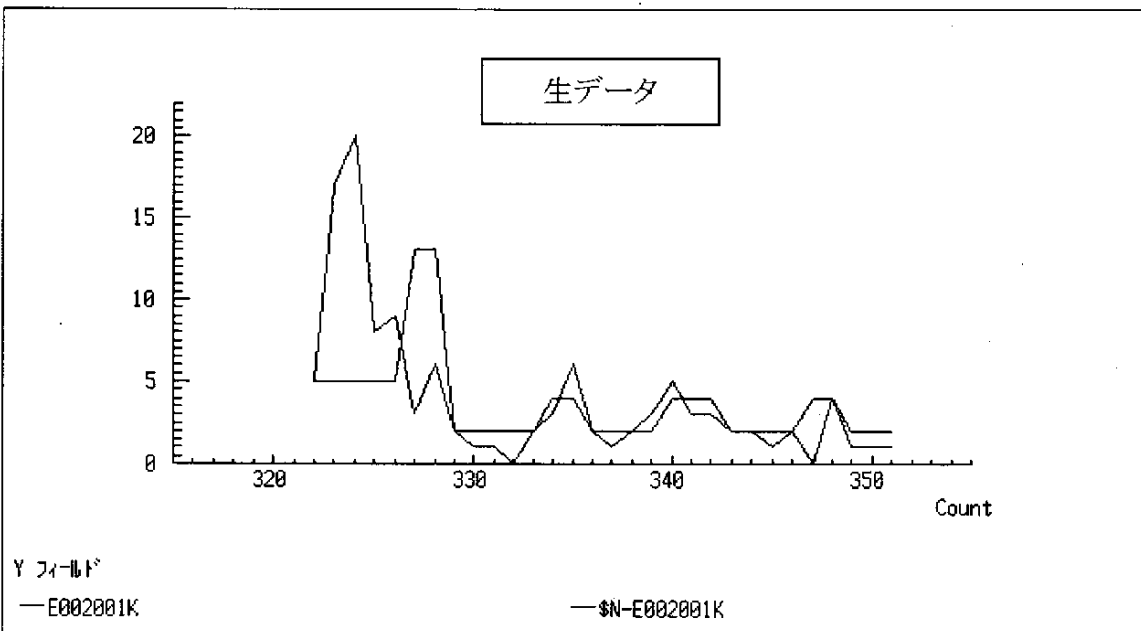
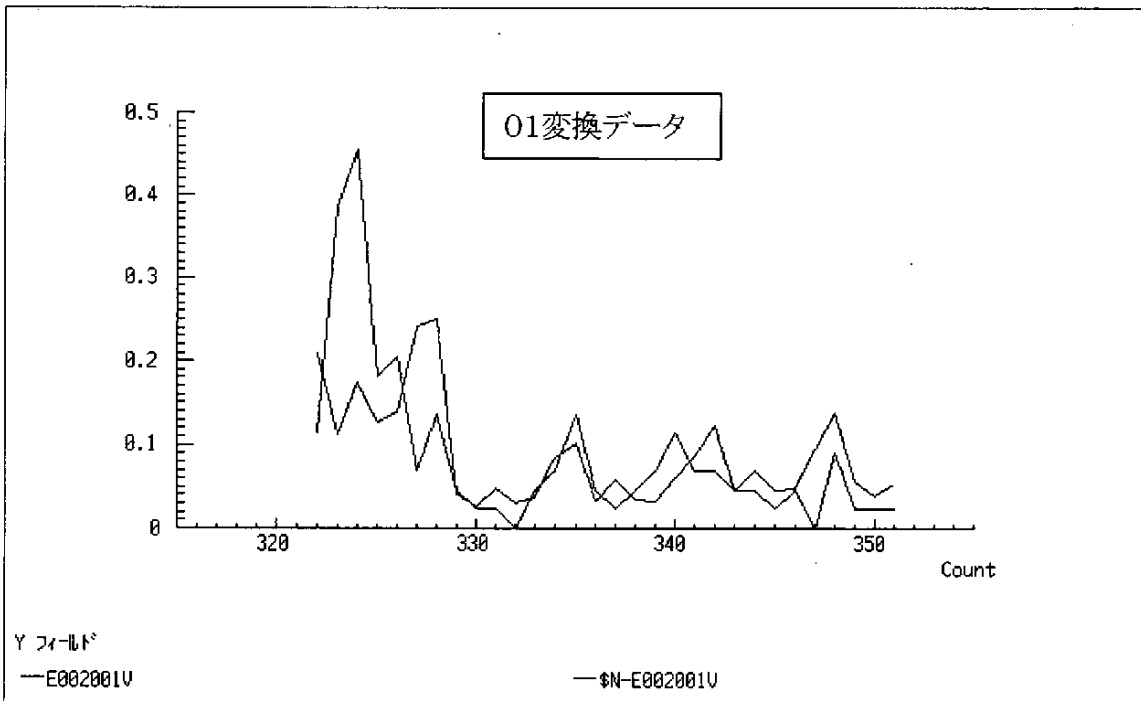


図5-5 検証 実績値との比較 (A店)

表5-7 精度分析 (左: 01変換データ、右: 生データ)

出力フィールドの結果: E002001V	
\$N-E002001V のエラー分析	
エラーの最小値:	-0.17359
エラーの最大値:	0.28041
エラーの平均値:	0.000072439

出力フィールドの結果: E002001K	
\$N-E002001K のエラー分析	
エラーの最小値:	-10
エラーの最大値:	15
エラーの平均値:	0.16667

表5-7 精度分析続き (左: 0.1変換データ、右: 生データ)

絶対エラーの平均値 :	0.055809
標準偏差 :	0.091030
線形相関 :	0.49267
出現頻度 :	30

絶対エラーの平均値 :	2.3667
標準偏差 :	4.4651
線形相関 :	0.34162
出現頻度 :	30

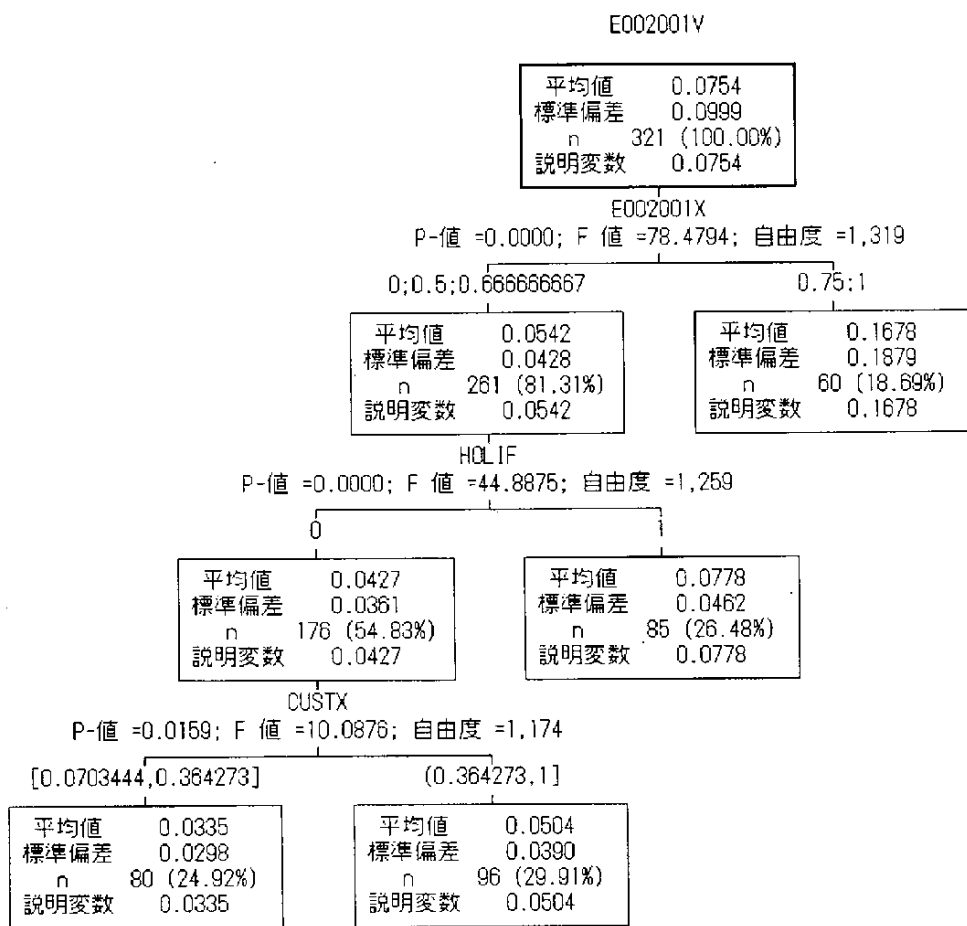


図5-6 Answer Tree (CHAID) による分析結果 (A店、0.1変換データ)

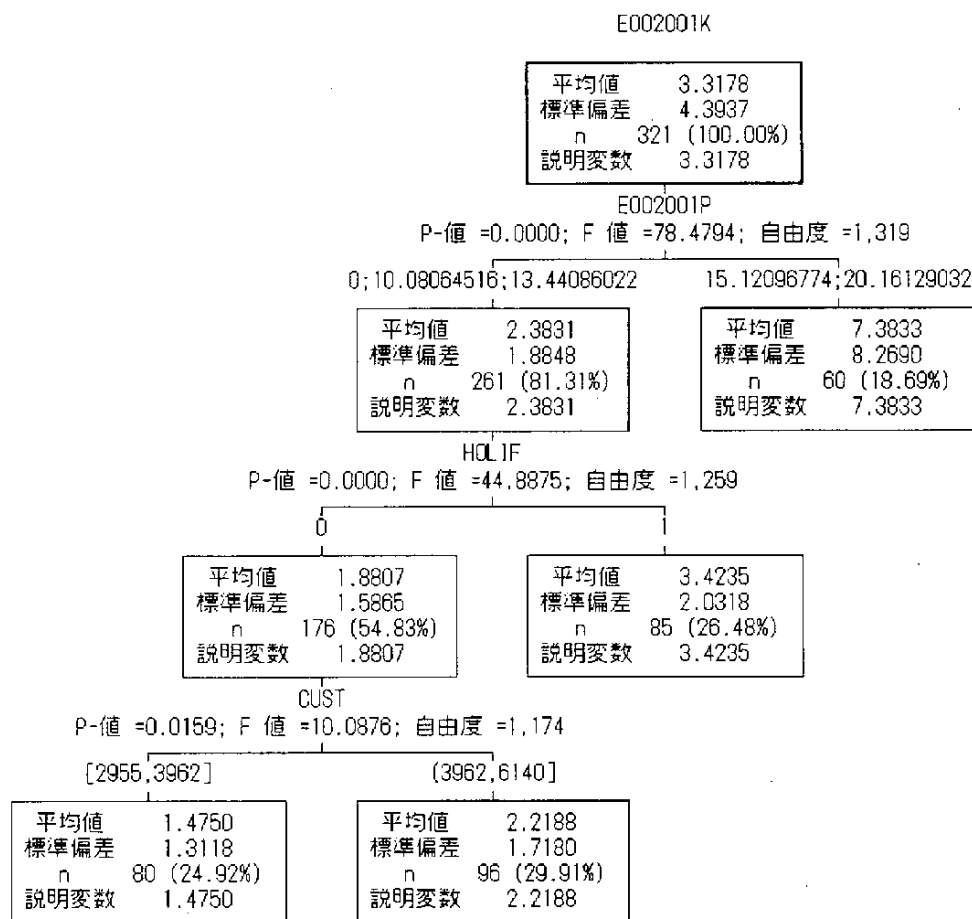


図5-7 Answer Tree (CHAID)による分析結果 (A店、生データ)

「エバラ 黄金の味 甘口 フルーツタイプ 瓶210G」のA店での販売動向を見ると、01変換データ、生データの両方とも突出した売り上げがあった日は予測できていないが、その他の日はだいたいの傾向をつかんでいる(図5-3、4)。重要度分析を見ると01変換データと生データでは違いが見られる(表5-3、5)。前者が来店客数や休日フラグの影響を大きく見ているのに対し、後者は自分自身の価格(値引率)が一番売り上げに関係するとしている。Answer Treeの分析結果を見ると(図5-6、7)、両者とも同じ結果が得られた。一番影響を与えるのは自分自身の価格(値引率)で、その後休日フラグ、来店客数である。

次に求めたモデルで3月の売上データに当てはめてみた(図5-5)。01変換データ、生データとも大きな違いはなく、そこそこ予測できている。

<エバラ 黄金の味 甘口 フルーツタイプ 瓶210G>

B店

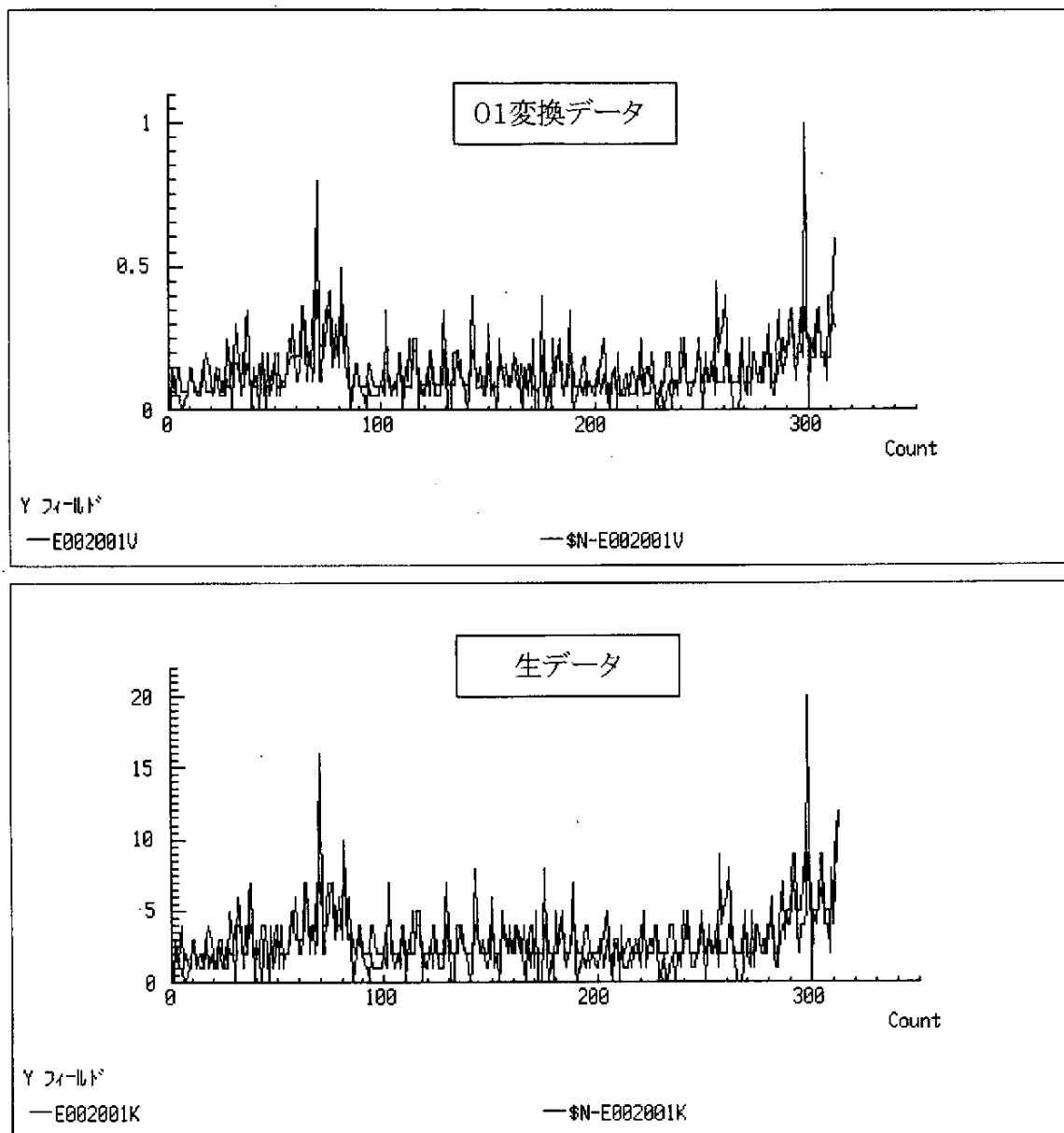


図5-8 ニューラルネットワークによる学習結果、実績値との比較 (B店)

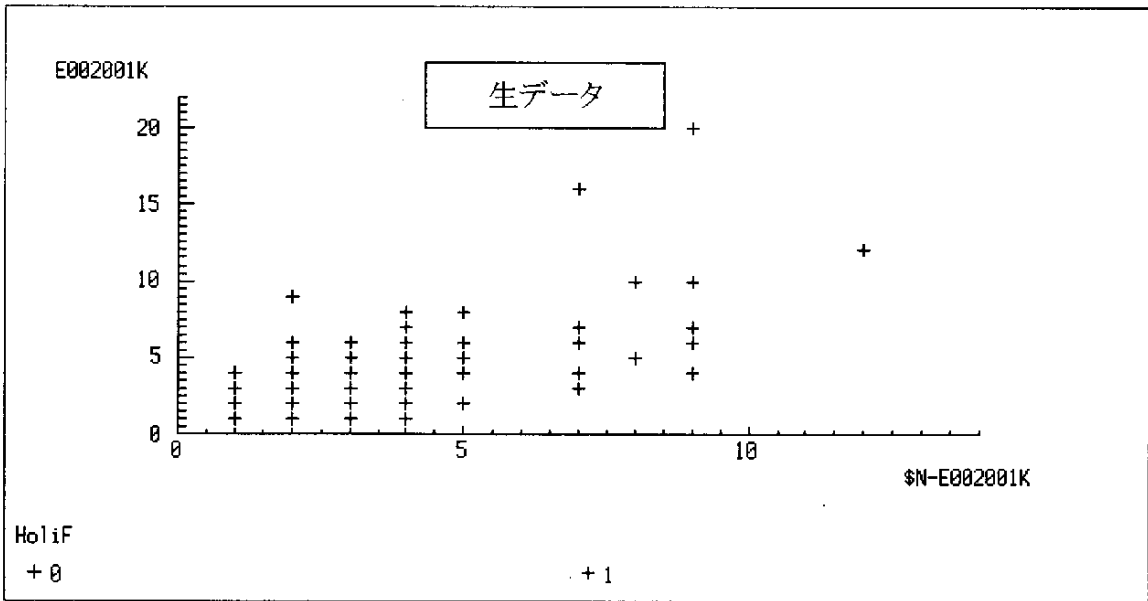
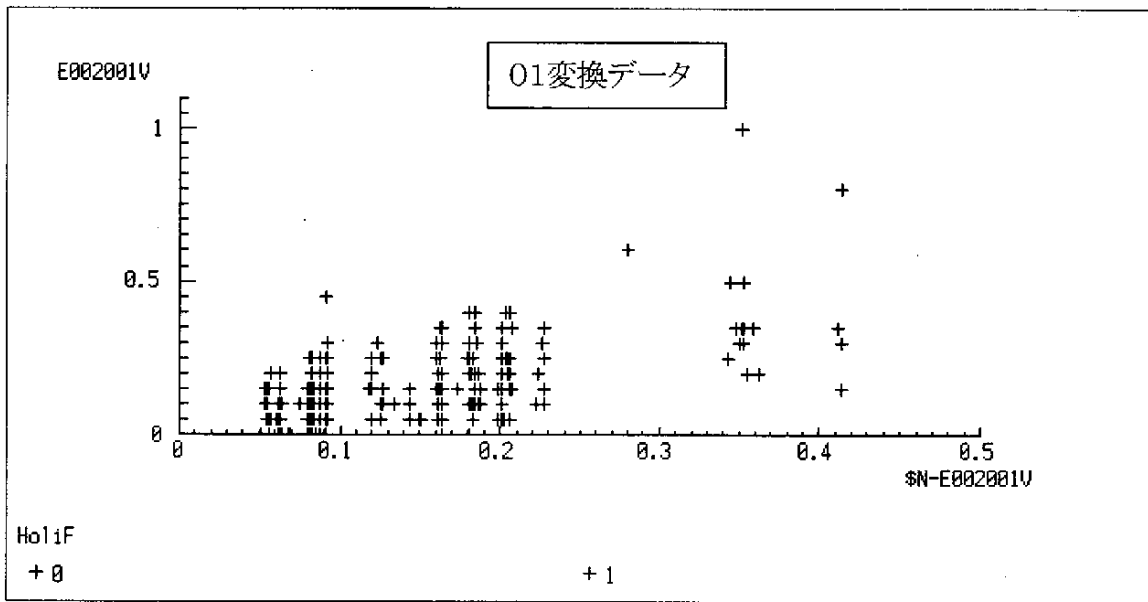


図5-9 ニューラルネットワークによる学習結果、実績値との比較 (B店)

表5-8 重要度分析 (01変換データ)

ニューラル・ネットワーク・アーキテクチャ: "E002001V"	
入力層	: 16 ニューロン
隠れ層 #1	: 10 ニューロン
出力層	: 1 ニューロン
予想正確度	: 93.20%

表5-9 精度分析 (01変換データ)

出力フィールドの結果: E002001V	
\$N-E002001V のエラー分析	
エラーの最小値	: -0.26444
エラーの最大値	: 0.64835
エラーの平均値	: 0.0068722
絶対エラーの平均値	: 0.067775
標準偏差	: 0.095096
線形相関	: 0.61568

表5-8 重要度分析 続き

入力の関係重要度	
E002018X	: 0.05745
E002025X	: 0.05021
E002117X	: 0.04015
K645118X	: 0.03518
M105129X	: 0.03273
HoliF	: 0.02797
K467574X	: 0.02596
E002360X	: 0.02357
E001165X	: 0.01438
E002001X	: 0.00751
M070793X	: 0.00572
N212270X	: 0.00543
E003220X	: 0.00473
E003336X	: 0.00458
CustX	: 0.00345

表5-9 精度分析 続き

線形相関	: 0.61568
出現頻度	: 312

表5-10 重要度分析 (生データ)

ニューラル・ネットワーク・アーキテクチャ: "E002001K"

入力層 : 15 ニューロン
 隠れ層 #1 : 5 ニューロン
 出力層 : 1 ニューロン

予想正確度 : 93.73%

入力の関係重要度

N212270P	: 0.15265
HoliF	: 0.10310
E002018P	: 0.06958
E002001P	: 0.06530
M105129P	: 0.05037
E003220P	: 0.02823
E003336P	: 0.02564
E002117P	: 0.02400
E001165P	: 0.01507
E002360P	: 0.01264
K467574P	: 0.01220
M070793P	: 0.00973
K645118P	: 0.00843
Cust	: 0.00723
E002025P	: 0.00655

表5-11 精度分析 (生データ)

出力フィールドの結果: E002001K

\$N-E002001K のエラー分析

エラーの最小値	: -5
エラーの最大値	: 11
エラーの平均値	: -0.26603
絶対エラーの平均値	: 1.3558
標準偏差	: 1.8729
線形相関	: 0.63402
出現頻度	: 312

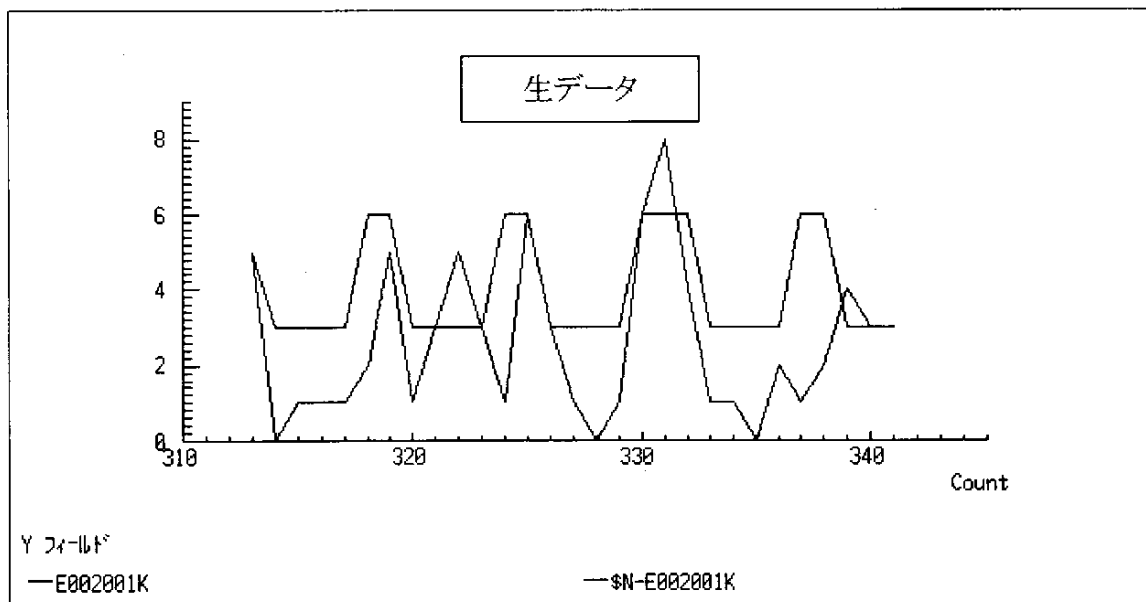
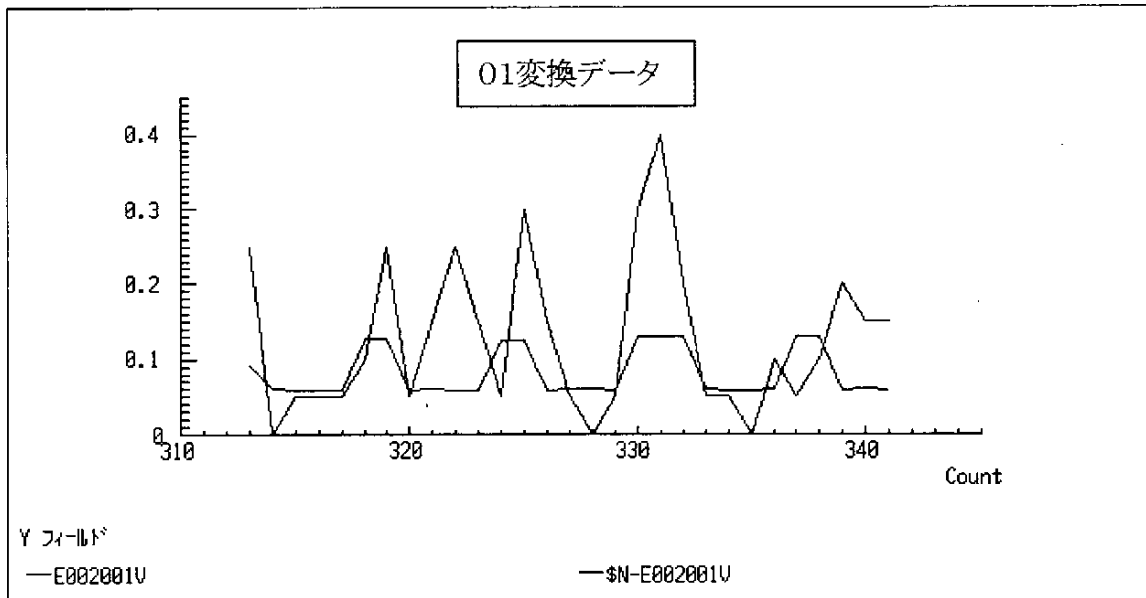


図5-10 検証 実績値との比較 (B店)

表5-12 精度分析 (左: 01変換データ、右: 生データ)

出力フィールドの結果: E002001V \$N-E002001V のエラー分析	出力フィールドの結果: E002001K \$N-E002001K のエラー分析
エラーの最小値 : -0.079368	エラーの最小値 : -5
エラーの最大値 : 0.27069	エラーの最大値 : 2
エラーの平均値 : 0.045695	エラーの平均値 : -1.4483
絶対エラーの平均値: 0.077902	絶対エラーの平均値: 1.7931
標準偏差 : 0.092901	標準偏差 : 1.8436
線形相関 : 0.48331	線形相関 : 0.49772
出現頻度 : 29	出現頻度 : 29

E002001V

平均値	0.1401
標準偏差	0.1207
n	312 (100.00%)
説明変数	0.1401

E002018X

P-値 =0.0000; F 値 =79.1039; 自由度 =1,310

0

0.8333333333;1

平均値	0.1158
標準偏差	0.0895
n	260 (83.33%)
説明変数	0.1158

平均値	0.2615
標準偏差	0.1734
n	52 (16.67%)
説明変数	0.2615

HOLIF

P-値 =0.0000; F 値 =51.5152; 自由度 =1,258

0

1

平均値	0.0895
標準偏差	0.0726
n	171 (54.81%)
説明変数	0.0895

平均値	0.1663
標準偏差	0.0973
n	89 (28.53%)
説明変数	0.1663

図5-11 Answer Tree (CHAID)による分析結果 (B店、01変換データ)

E002001K

平均値	2.8013
標準偏差	2.4136
n	312 (100.00%)
説明変数	2.8013

E002018P

P-値 =0.0000; F 値 =79.1039; 自由度 =1,310

0 15.72327044; 18.86792453

平均値	2.3154
標準偏差	1.7903
n	260 (83.33%)
説明変数	2.3154

平均値	5.2308
標準偏差	3.4676
n	52 (16.67%)
説明変数	5.2308

HOLIF

P-値 =0.0000; F 値 =51.5152; 自由度 =1,258

0 1

平均値	1.7895
標準偏差	1.4520
n	171 (54.81%)
説明変数	1.7895

平均値	3.3258
標準偏差	1.9469
n	89 (28.53%)
説明変数	3.3258

図5-12 Answer Tree (CHAID)による分析結果 (B店、生データ)

B店では0 1変換データ、生データの両方ともほぼ似たような販売動向を予測しており、両者ともそれほど正確ではない(図5-8、9)。重要度分析を見ると(表5-8、10)両者で大きく異なっている。0 1変換データの方は、「エバラ 黄金の味 中辛フルーツ&スパイス瓶210G」の値引率が一番影響が大きいですが、関係重要度自体は低く、作られたモデルの性能があまりよくないと考えられる。それに対して、生データの方は「日本食研 晩餐館 焼肉のたれ 中辛 210G」の値引率の影響がもっとも大きい。関係重要度も0 1変換データの「エバラ 黄金の味 中辛フルーツ&スパイス瓶210G」の値引率より大きい。Answer Tree の分析結果を見ると(図5-11、12)、両者で差はなく、0 1変換データの方を支持しているようだ。なぜ、このような食い違いが出たのかはわからない。

次に求めたモデルを3月の売上データに当てはめてみた(図5-10)。どちらもあまりいい当てはまりとはいえない。

<エバラ 黄金の味 甘口 フルーツタイプ 瓶210G>

C店

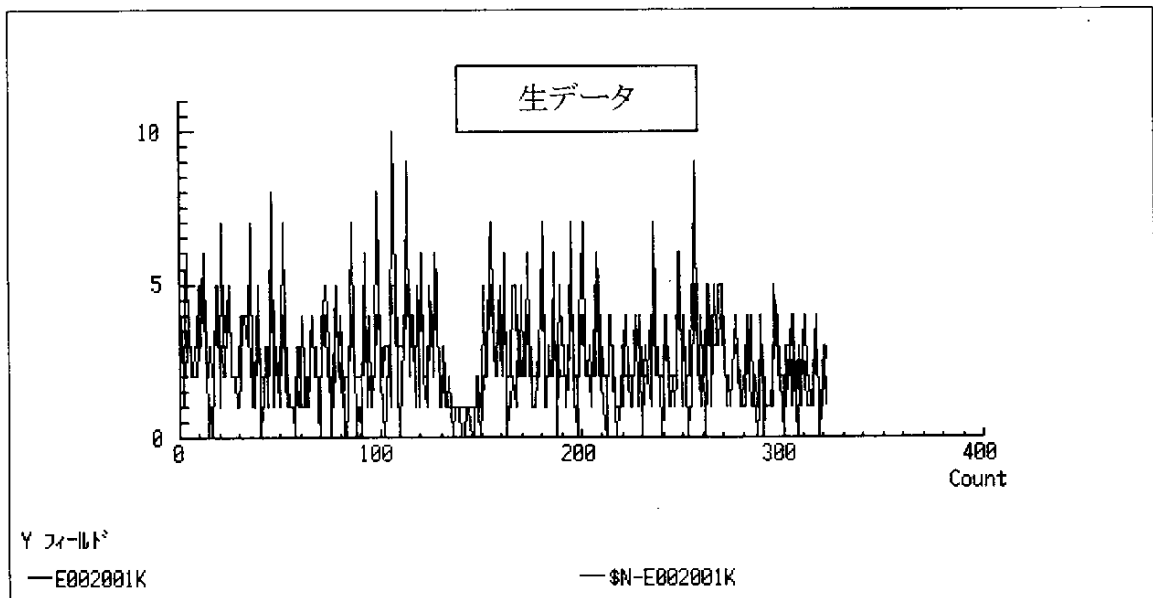
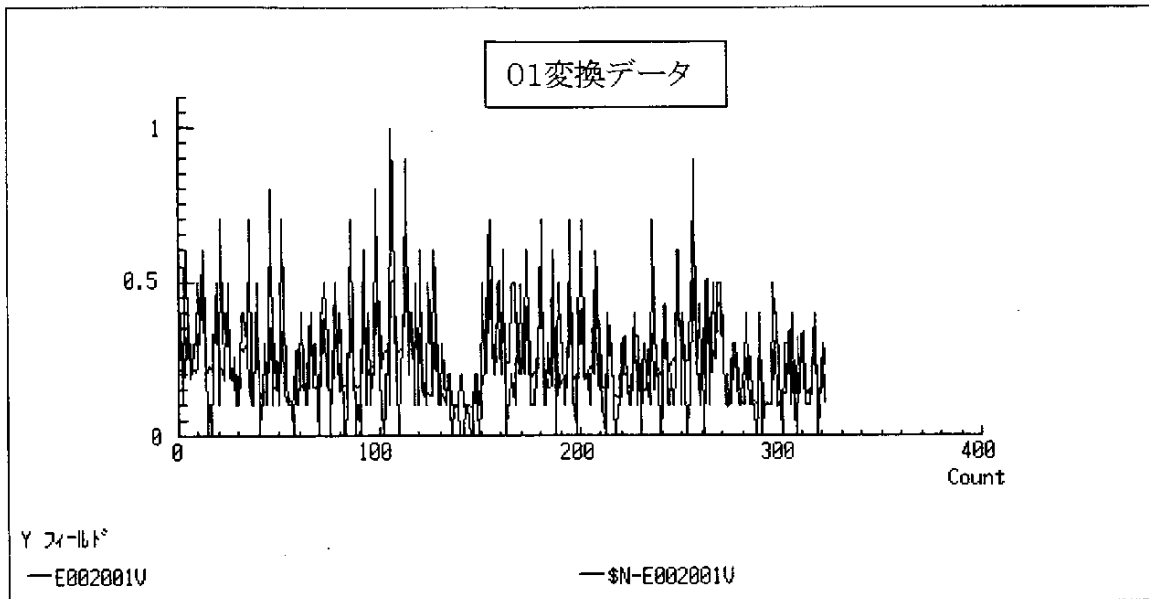


図5-13 ニューラルネットワークによる学習結果、実績値との比較 (C店)

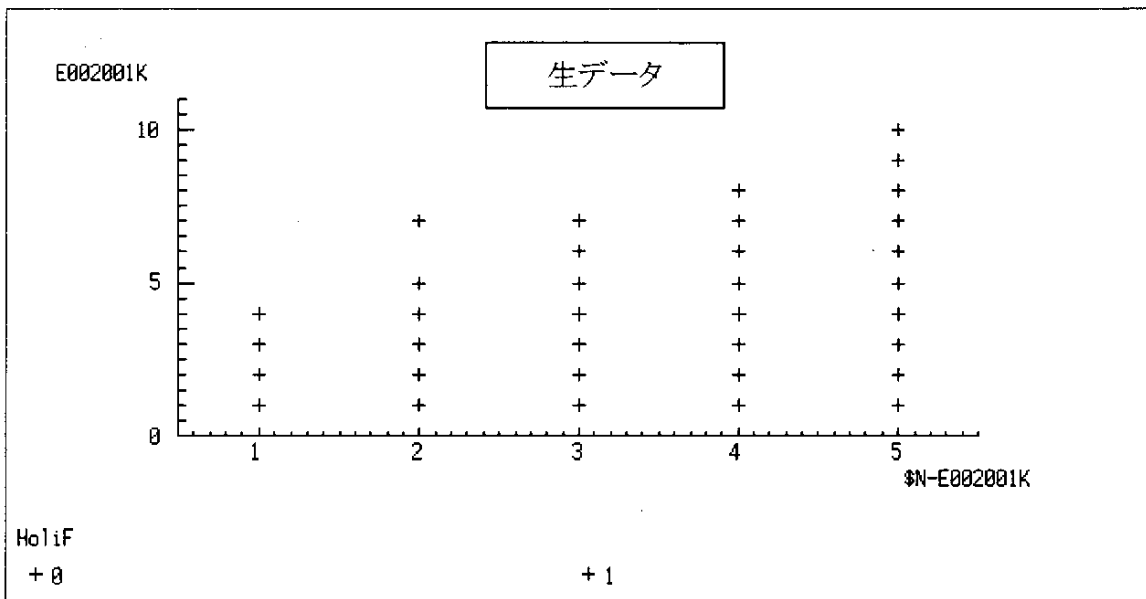
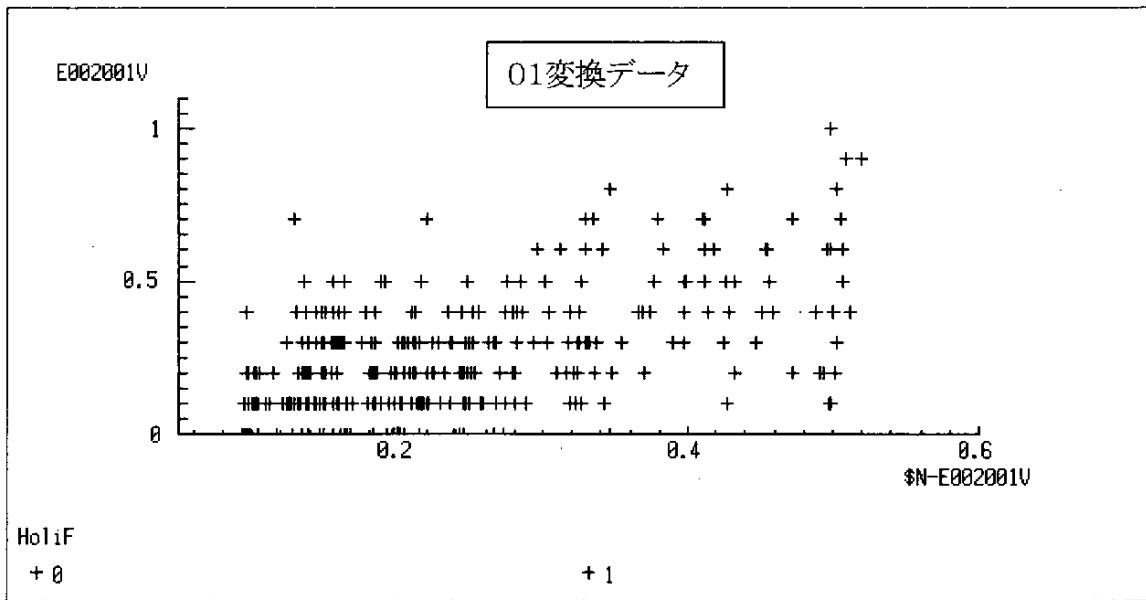


図5-14 ニューラルネットワークによる学習結果、実績値との比較 (C店)

表5-13 重要度分析 (01変換データ)

ニューラル・ネットワーク・アーキテクチャ: "E002001V"	
入力層	: 20 ニューロン
隠れ層 #1	: 16 ニューロン
出力層	: 1 ニューロン
予想正確度	: 87.57%

表5-14 精度分析 (01変換データ)

出力フィールドの結果: E002001V	
\$N-E002001V のエラー分析	
エラーの最小値	: -0.39808
エラーの最大値	: 0.57067
エラーの平均値	: 0.014024
絶対エラーの平均値	: 0.12366
標準偏差	: 0.15989
線形相関	: 0.56384
出現頻度	: 322

表5-13 重要度分析 (01変換データ) 続き

入力の関係重要度	
HoliF	: 0.18877
E003206X	: 0.06320
E002018X	: 0.06279
E002117X	: 0.06210
CustX	: 0.05058
K467574X	: 0.04262
K645125X	: 0.03627
E002124X	: 0.03171
K342189X	: 0.03165
K341489X	: 0.01879
E002001X	: 0.01507
E003183X	: 0.01316
E002360X	: 0.01167
E001165X	: 0.01150
E003336X	: 0.01062
H480748X	: 0.01016
K645118X	: 0.00950
E002100X	: 0.00940
E003220X	: 0.00612
H480618X	: 0.00585

表5-15 重要度分析 (生データ)

ニューラル・ネットワーク・アーキテクチャ: "E002001K"	
入力層	: 20 ニューロン
隠れ層 #1	: 14 ニューロン
出力層	: 1 ニューロン
予想正確度	: 88.38%
入力の関係重要度	
Cust	: 0.17314
HoliF	: 0.15193
H480618P	: 0.09433
E002117P	: 0.09273
K467574P	: 0.07402
K341489P	: 0.05054
E002100P	: 0.04383
E001165P	: 0.04201
K645118P	: 0.04099
E003336P	: 0.03823
E003183P	: 0.03758
E002001P	: 0.03111
E003220P	: 0.03017

表5-16 精度分析 (生データ)

出力フィールドの結果: E002001K	
\$N-E002001K のエラー分析	
エラーの最小値	: -4
エラーの最大値	: 5
エラーの平均値	: -0.037267
絶対エラーの平均値	: 1.1863
標準偏差	: 1.5524
線形相関	: 0.59751
出現頻度	: 322

表5-15 重要度分析(生データ) 続き

E002018P	: 0.02419
H480748P	: 0.02226
K342189P	: 0.01954
E002124P	: 0.00949
K645125P	: 0.00841
E002360P	: 0.00743
E003206P	: 0.00623

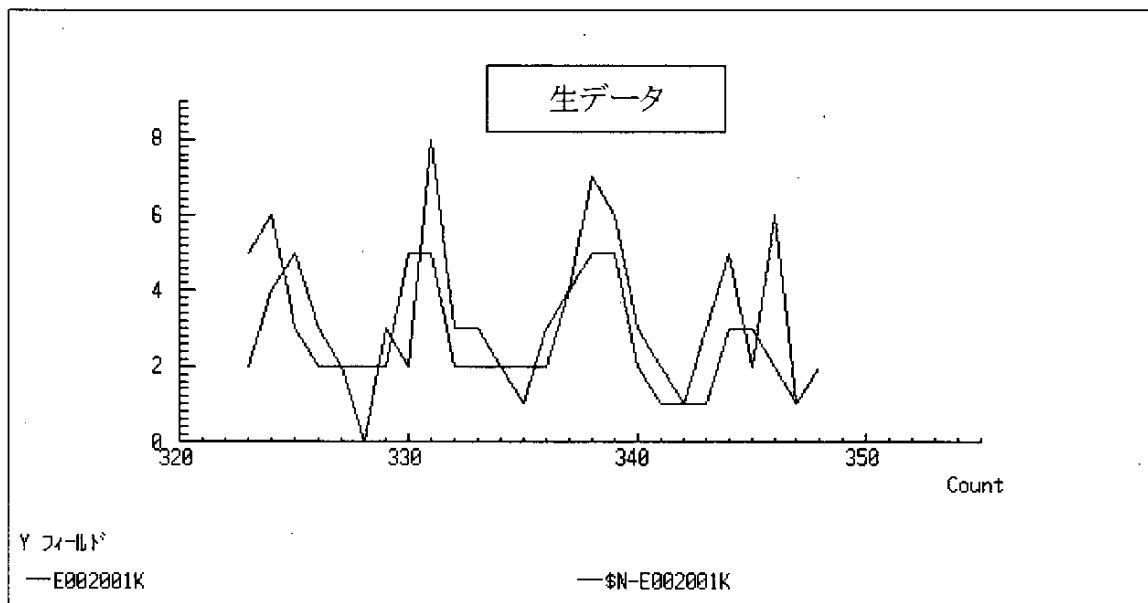
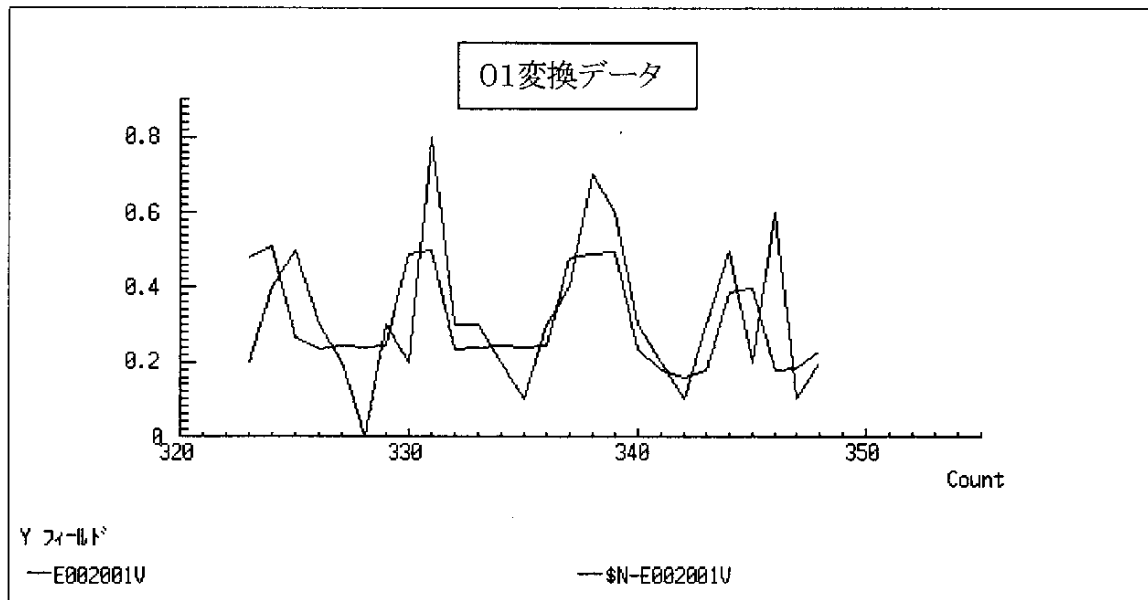


図5-15 検証 実績値との比較(C店)

表5-17 精度分析 (左: 0.1変換データ、右: 生データ)

出力フィルドの結果: E002001V \$N-E002001V のエラー分析		出力フィルドの結果: E002001K \$N-E002001K のエラー分析	
エラーの最小値	: -0.28818	エラーの最小値	: -3
エラーの最大値	: 0.41953	エラーの最大値	: 4
エラーの平均値	: 0.012463	エラーの平均値	: 0.42308
絶対エラーの平均値	: 0.13461	絶対エラーの平均値	: 1.3462
標準偏差	: 0.17170	標準偏差	: 1.7011
線形相関	: 0.50117	線形相関	: 0.54455
出現頻度	: 26	出現頻度	: 26

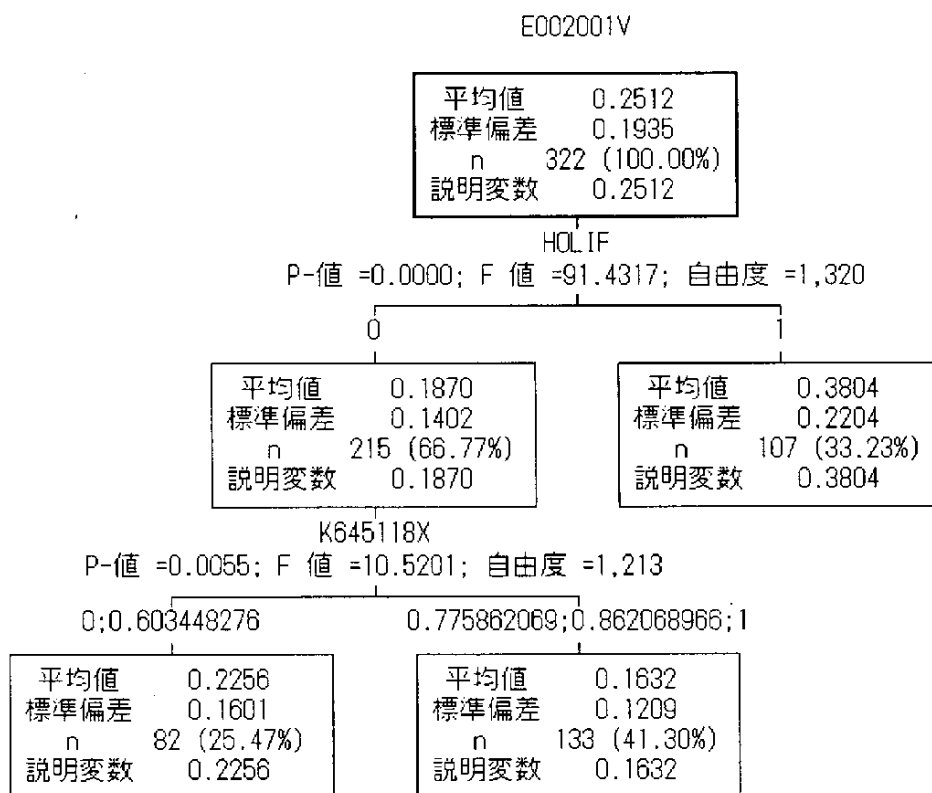


図5-16 Answer Tree (CHAID) による分析結果 (C店、0.1変換データ)

E002001K

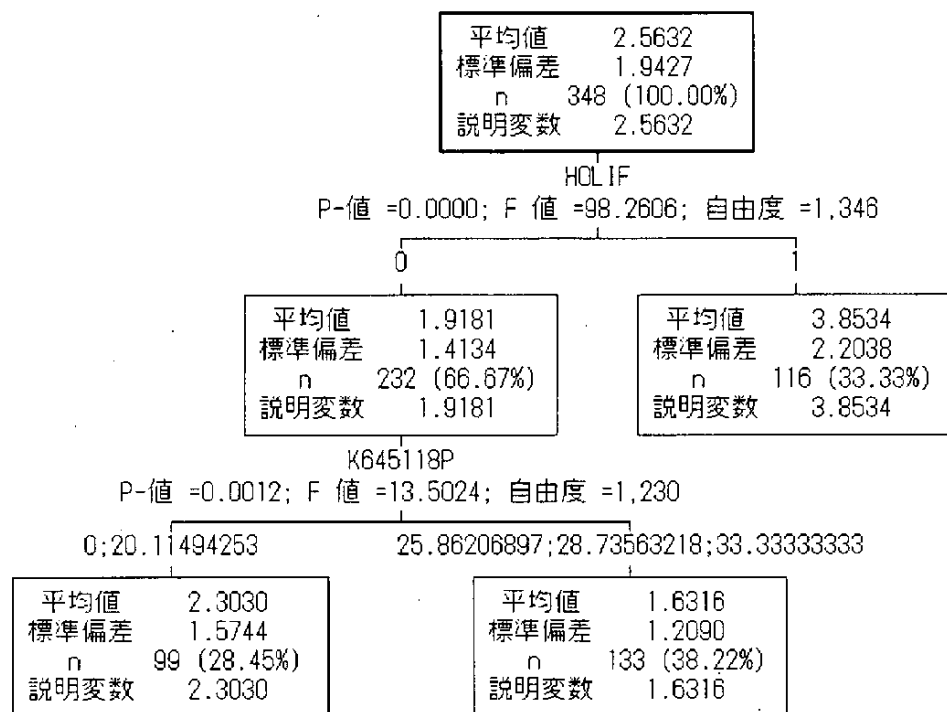


図5-17 Answer Tree (CHAID)による分析結果 (C店、生データ)

C店では0 1変換データ、生データの両方ともほぼ同じような販売動向を示している(図5-13、14)。大まかな傾向はつかんでいるが、販売個数が大きいときや0の時は追い切れていない。重要度分析を見ると(表5-13、15)、やはり大きく異なっている。0 1変換データの方は休日フラグが一番強い影響力を持っている。他の入力の関係重要度はあまり高くない。強い影響を与える説明変数は休日フラグだけである。一方、生データの方は来店客数、休日フラグの順に関係重要度が高く、続いて「ハウスさっと逸品焼肉ドレすっきりレモン醤油300ML」の値引率、「エバラ 黄金の味 中辛フルーツ&スパイス瓶400G」の値引率となっている。C店では自分自身の価格(値引率)は売り上げには関係ないようだ。Answer Treeの分析結果を見ると(図5-16、17)、休日フラグが一番の要因になっている。次に「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率が要因として挙げられている。休日フラグ以外の要因は分析手法によりまちまちである。

次に求めたモデルを3月の売上データに当てはめた(図5-15)。両者ともあまり当てはまりがいいとはいえない。

「エバラ おろしのたれ 焼肉のたれ 瓶 270G」 A店

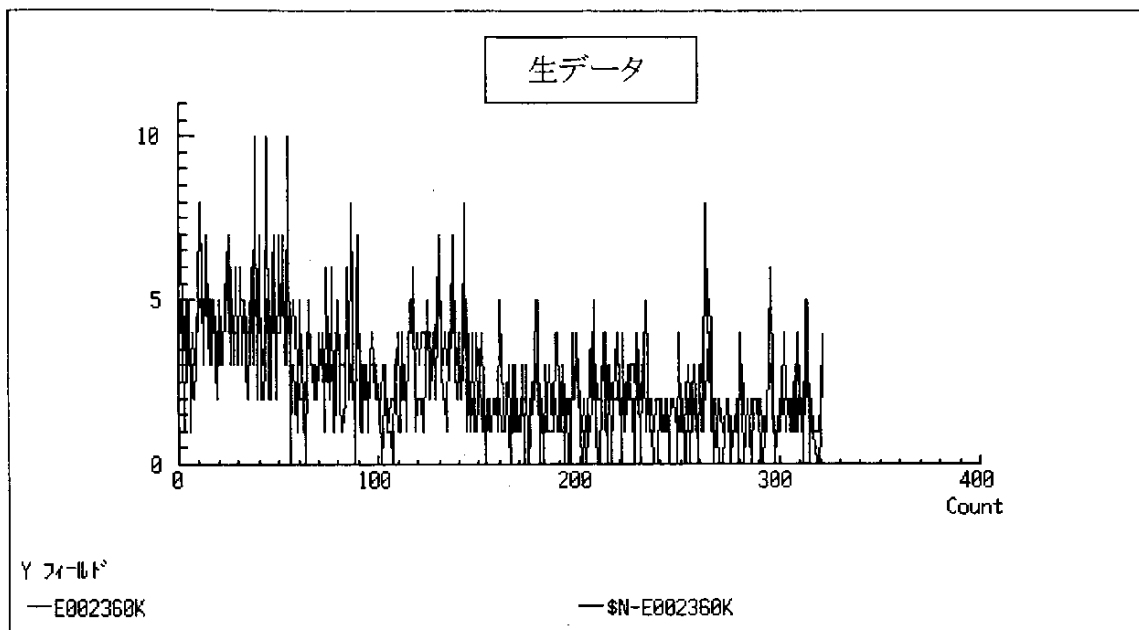


図5-18 ニューラルネットワークによる学習結果、実績値との比較 (A店)

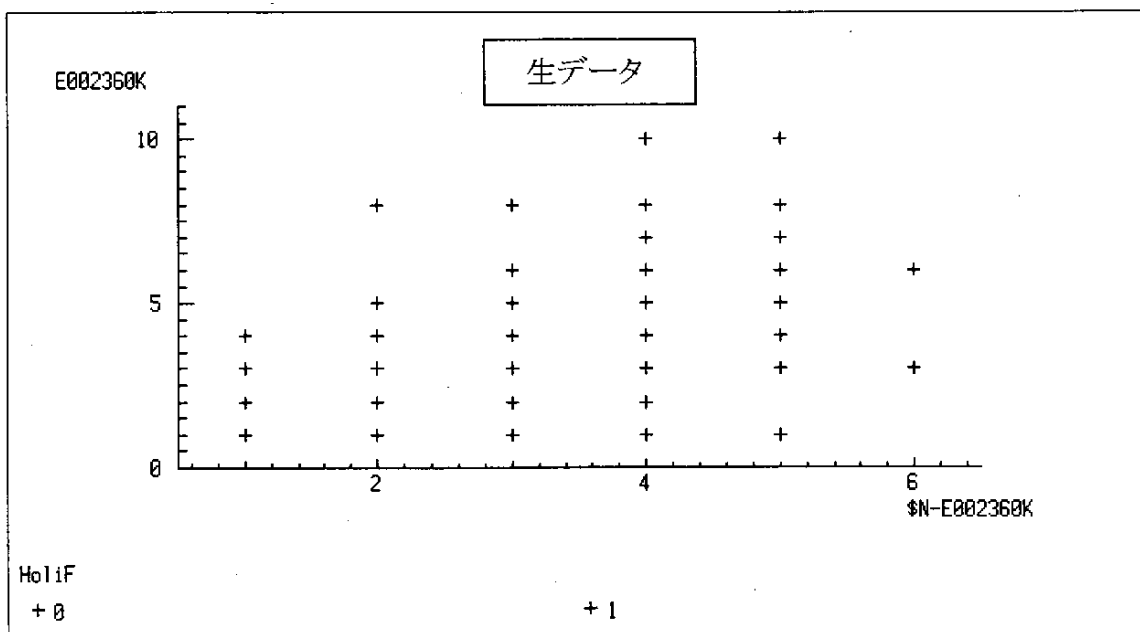


図5-19 ニューラルネットワークによる学習結果、実績値との比較 (A店)

表5-18 重要度分析 (生データ)

ニューラル・ネットワーク・アーキテクチャ: "E002360K"

入力層 : 12 ニューロン
 隠れ層 #1 : 10 ニューロン
 出力層 : 1 ニューロン

予想正確度 : 87.12%

入力の関係重要度

Cust	: 0.27719
E002360P	: 0.25602
E001165P	: 0.10482
M340166P	: 0.08978
HoliF	: 0.07636
S040231P	: 0.07096
E002018P	: 0.06793
E002025P	: 0.06086
E002001P	: 0.04820
K467574P	: 0.02783
K645118P	: 0.02162
E001172P	: 0.01691

表5-19 精度分析 (生データ)

出力フィールドの結果: E002360K

\$N-E002360K のエラー分析

エラーの最小値	: -4
エラーの最大値	: 6
エラーの平均値	: -0.12461
絶対エラーの平均値	: 1.2897
標準偏差	: 1.6668
線形相関	: 0.55885
出現頻度	: 321

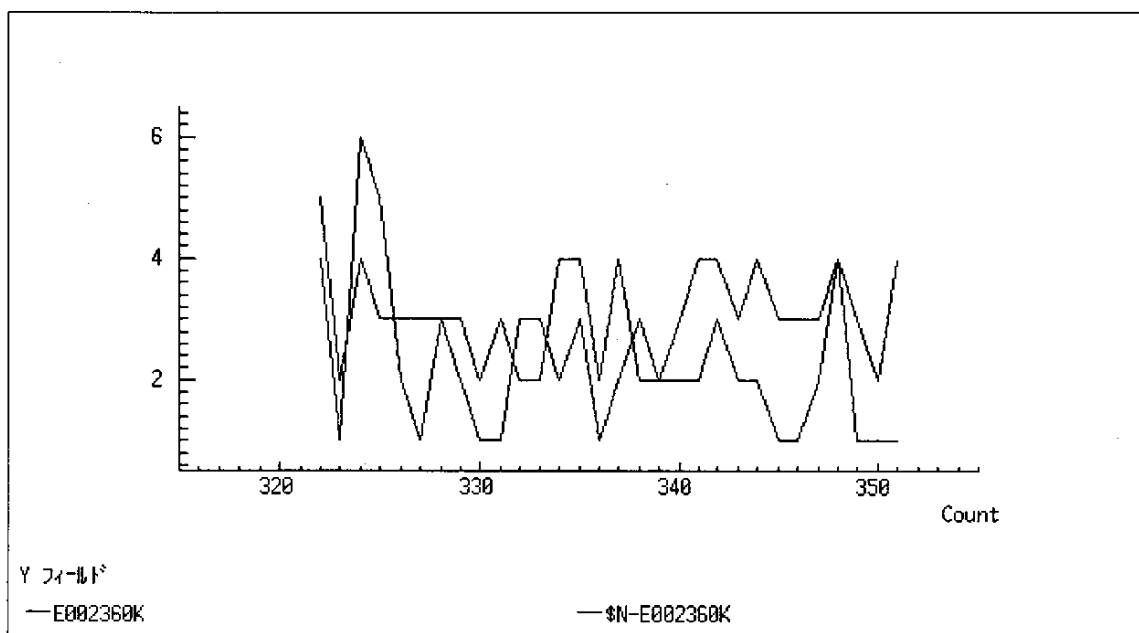


図5-20 検証 実績値との比較 (A店)

表5-20 精度分析 (生データ)

出力フィールドの結果: E002360K	
\$N-E002360K のエラー分析	
エラーの最小値	: -3
エラーの最大値	: 2
エラーの平均値:	-0.86667
絶対エラーの平均値:	1.3333
標準偏差	: 1.2521
線形相関	: 0.36088
出現頻度	: 30

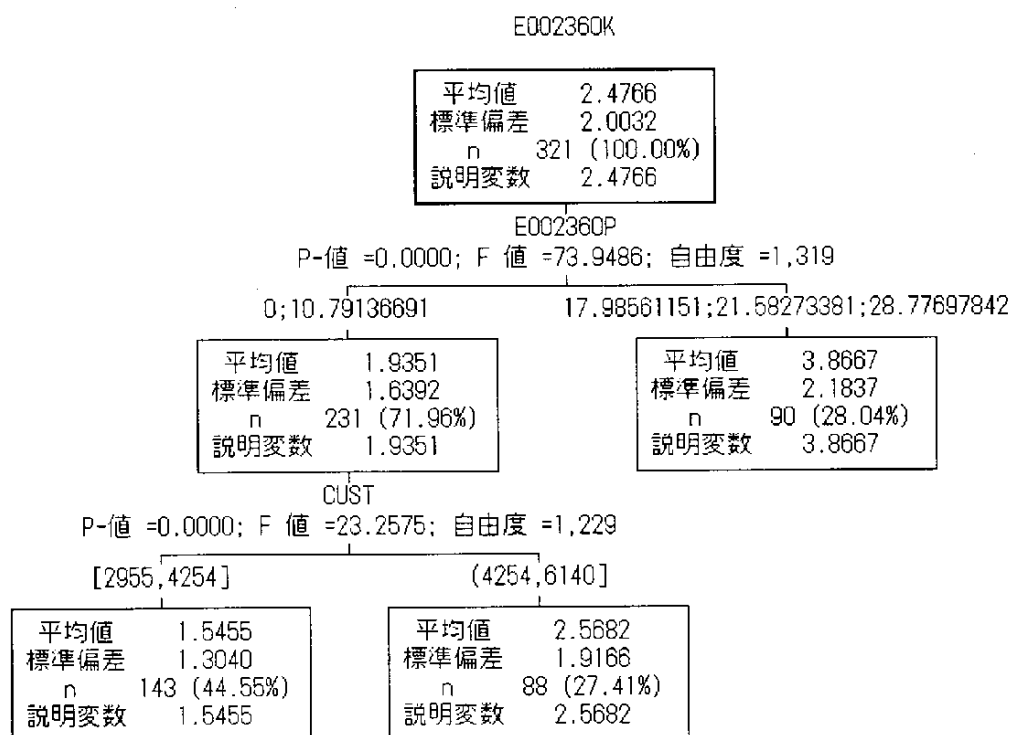


図5-21 Answer Tree (CHAID) による分析結果 (A店、生データ)

「エバラ おろしのたれ 焼肉のたれ 瓶 270G」のA店での販売動向を、生データを使って分析した(図5-18)。突出して売れた日やまったく売れなかった日は予測できていないが、日々の変動を丸くしたおおよその傾向はつかめている。重要度分析を見ると(表5-18)、来店客数、自分自身の値引率に影響を受ける事がわかる。Answer Tree の分析結果も同様(図5-21)で、自分自身の値引率、来店客数で分割されている。

次に求めたモデルを3月の売上データに当てはめてみた(図5-20)。あまりいい精度とはいえない。

「エバラ おろしのたれ 焼肉のたれ 瓶 270G」 B店

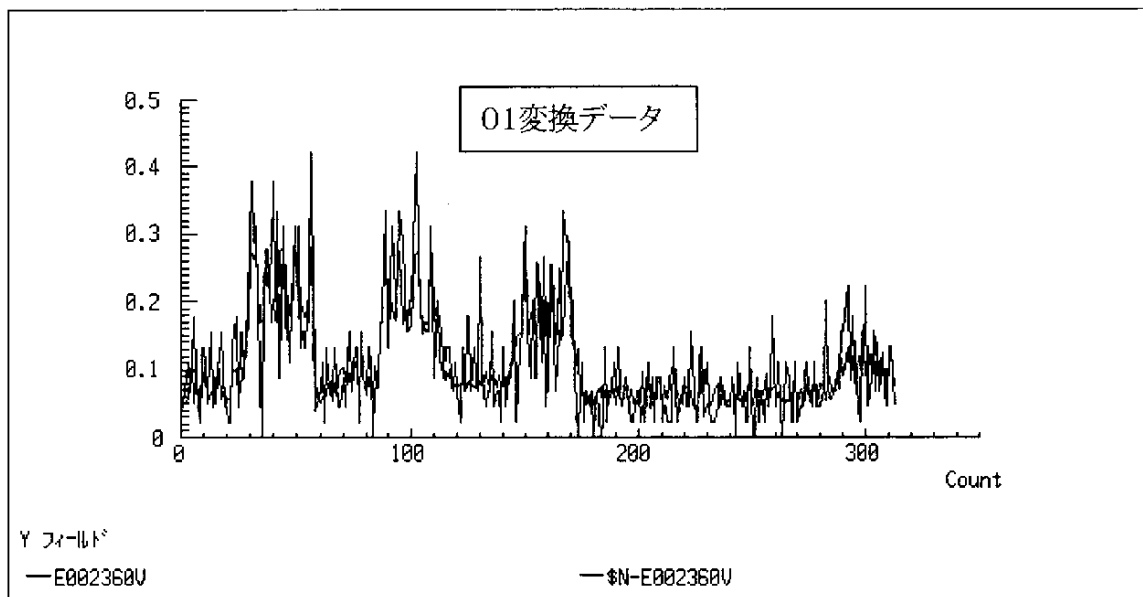


図5-22 ニューラルネットワークによる学習結果、実績値との比較 (B店)

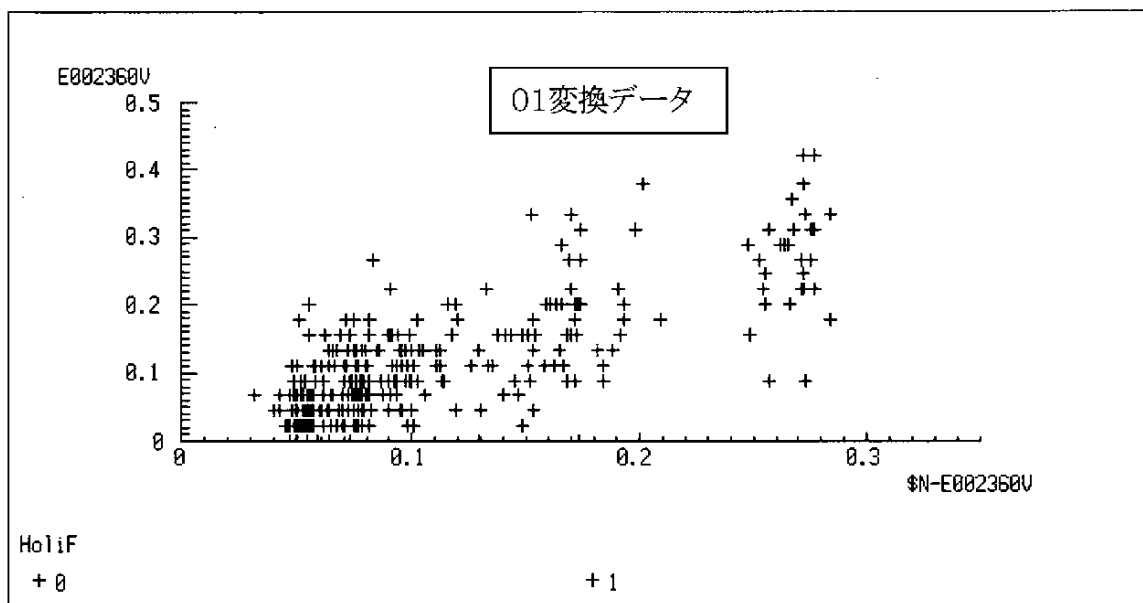


図5-23 ニューラルネットワークによる学習結果、実績値との比較 (B店)

表5-21 重要度分析 (01変換データ)

ニューラルネットワークアーキテクチャ: "E002360V"

入力層 : 14 ニューロン
 隠れ層 #1 : 14 ニューロン
 出力層 : 1 ニューロン

予想正確度 : 91.16%

入力の関係重要度

E002360X	: 0.25930
CustX	: 0.12325
N212270X	: 0.08998
K645118X	: 0.08412
M070793X	: 0.06800
K467574X	: 0.06670
E003220X	: 0.05306
E001165X	: 0.05123
E003336X	: 0.04954
E002117X	: 0.04759
M105129X	: 0.04577
E002001X	: 0.04070
HoliF	: 0.03112

表5-22 精度分析 (01変換データ)

出力フィールドの結果: E002360V

\$N-E002360V のエラー分析

エラーの最小値	: -0.18400
エラーの最大値	: 0.18341
エラーの平均値	: 0.0041215
絶対エラーの平均値	: 0.040671
標準偏差	: 0.054422
線形相関	: 0.76851
出現頻度	: 312

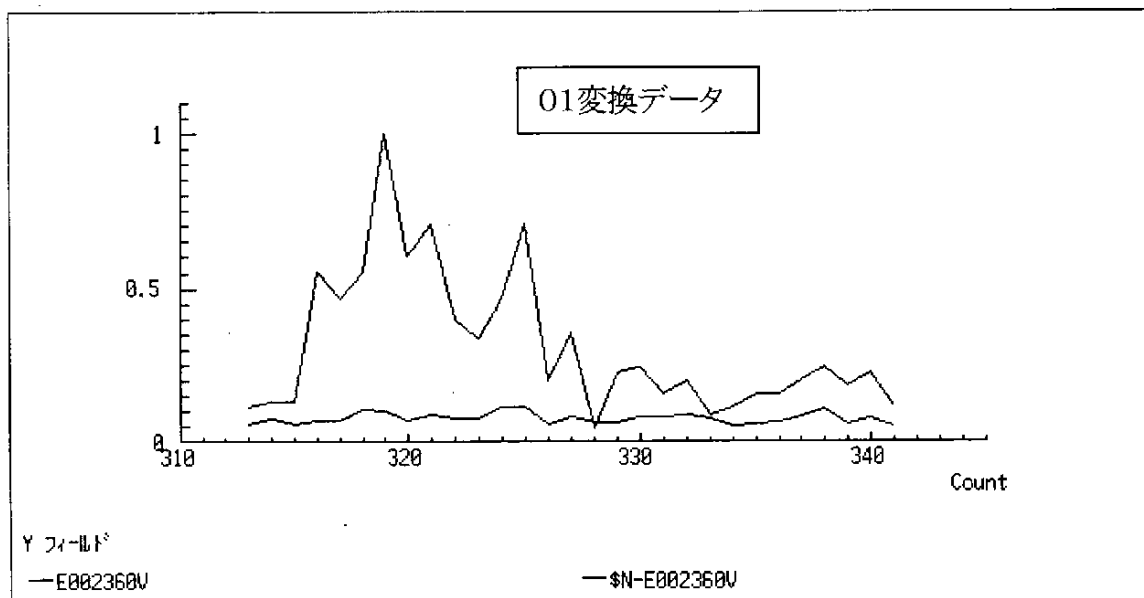


図5-24 検証 実績値との比較 (B店)

表5-23 精度分析 (B店、01変換データ)

出力フィールドの結果: E002360V	
\$N-E002360V のエラー分析	
エラーの最小値	: -0.019413
エラーの最大値	: 0.89766
エラーの平均値	: 0.23728
絶対エラーの平均値	: 0.23861
標準偏差	: 0.22184
線形相関	: 0.59016
出現頻度	: 29

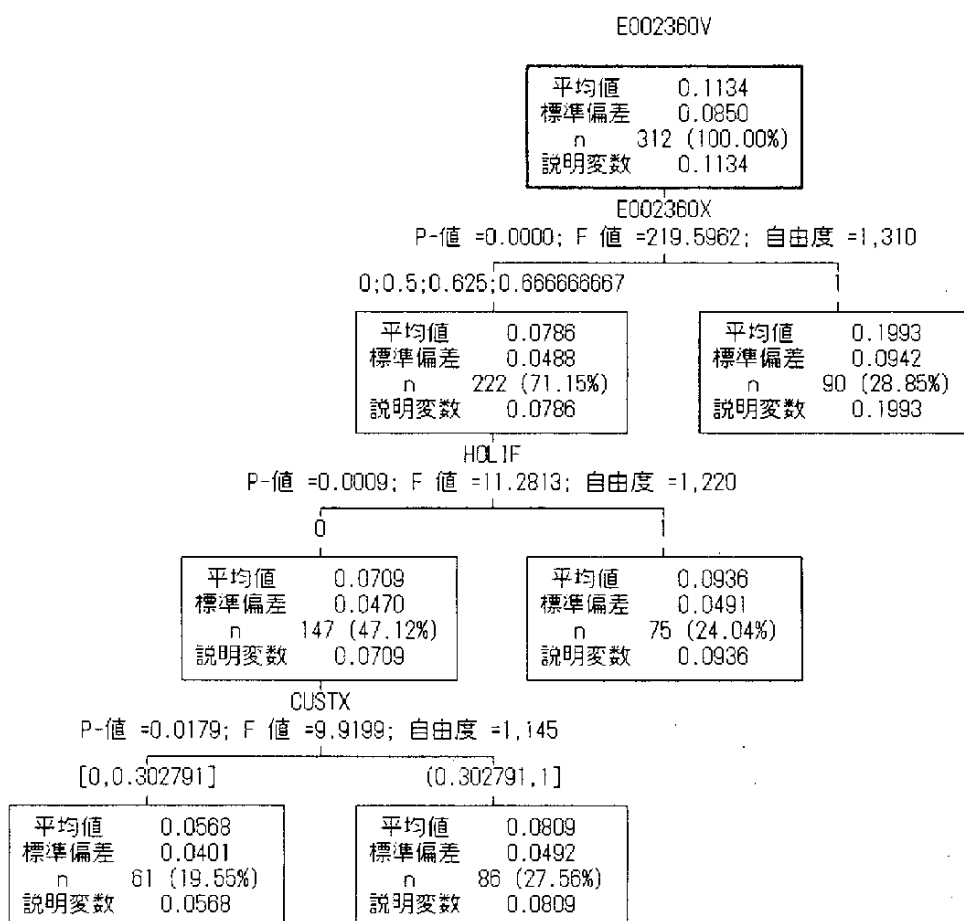


図5-25 Answer Tree (CHAID) による分析結果 (B店、01変換データ)

B店では01変換データを使った。販売動向を見ると(図5-22、23)、売上個数は相当複雑な動きをしているが、かなりフォローできている。重要度分析を見ると

(表5-21)、自分自身の値引率、来店客数が販売個数に大きな影響を与えているようだ。Answer Tree での分析結果も同様で(図5-25)、自分自身の価格、休日フラグ、来店客数の影響が大きい。

次に求めたモデルを3月の売上データに当てはめてみた(図5-24)。残念ながら、ほとんど当てはまらない。

「エバラ おろしのたれ 焼肉のたれ 瓶 270G」 C店

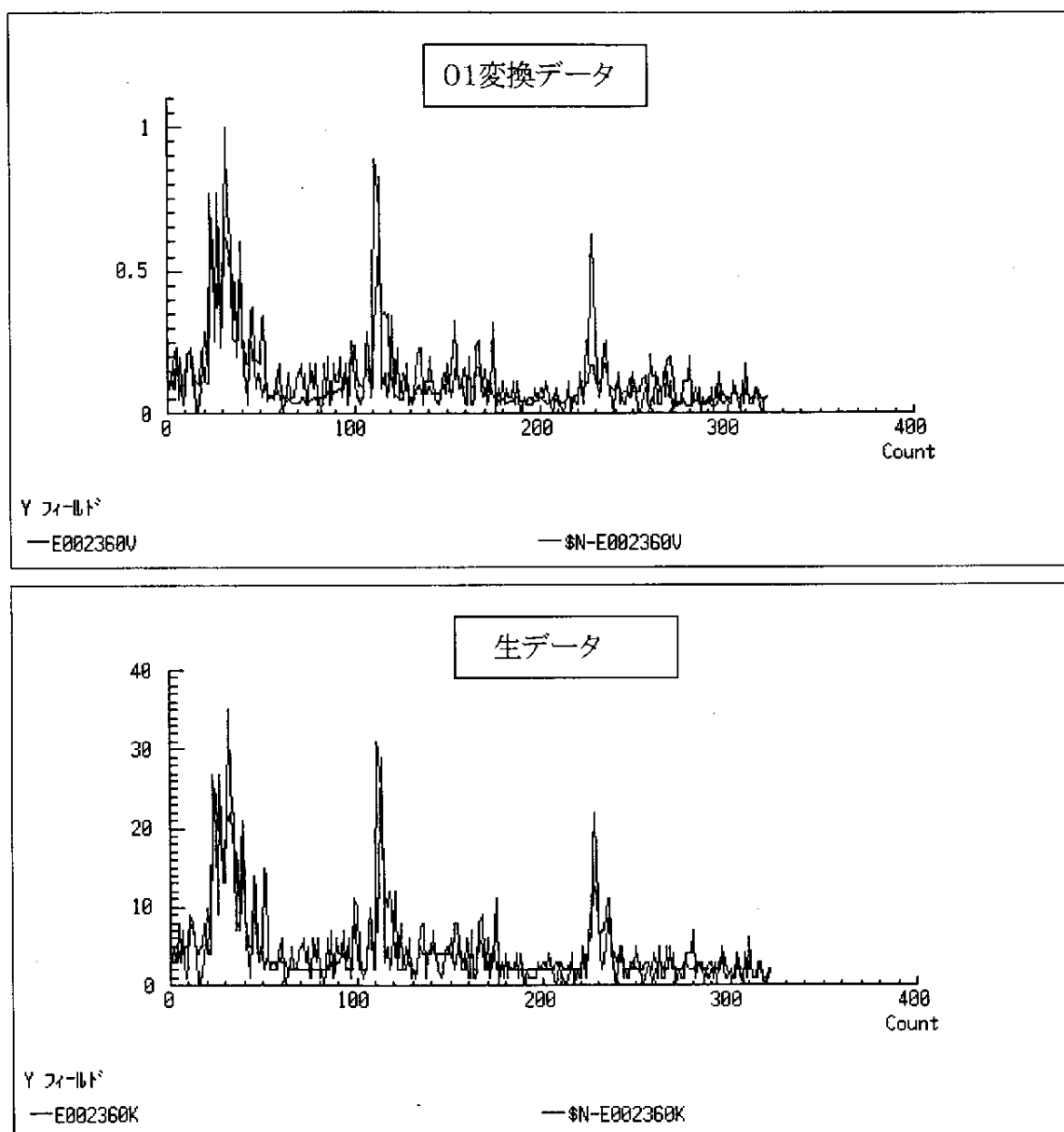


図5-26 ニューラルネットワークによる学習結果、実績値との比較 (C店)

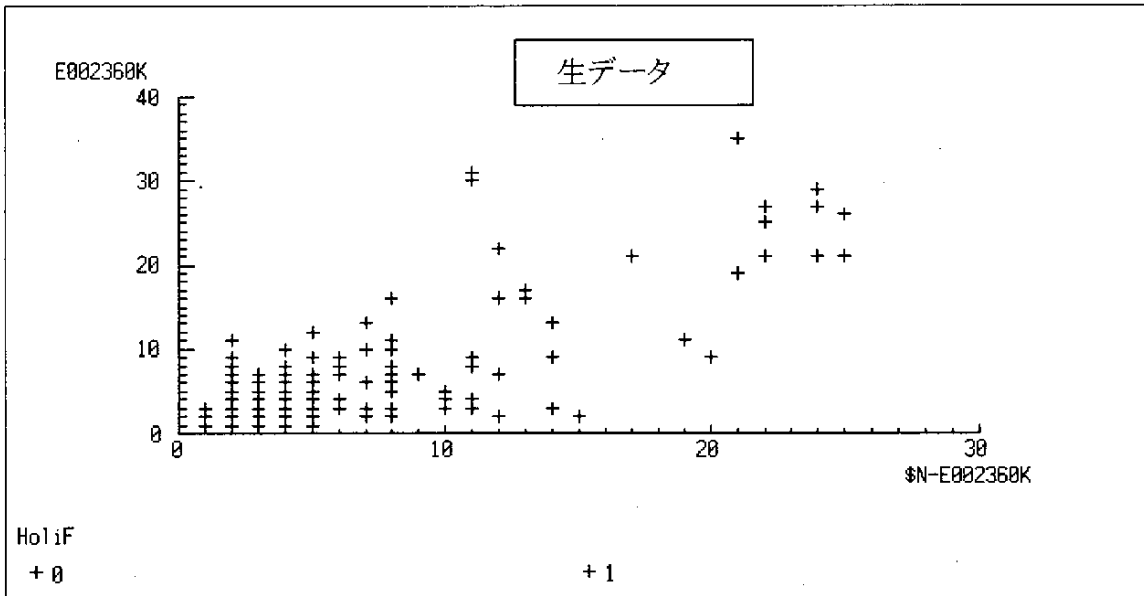
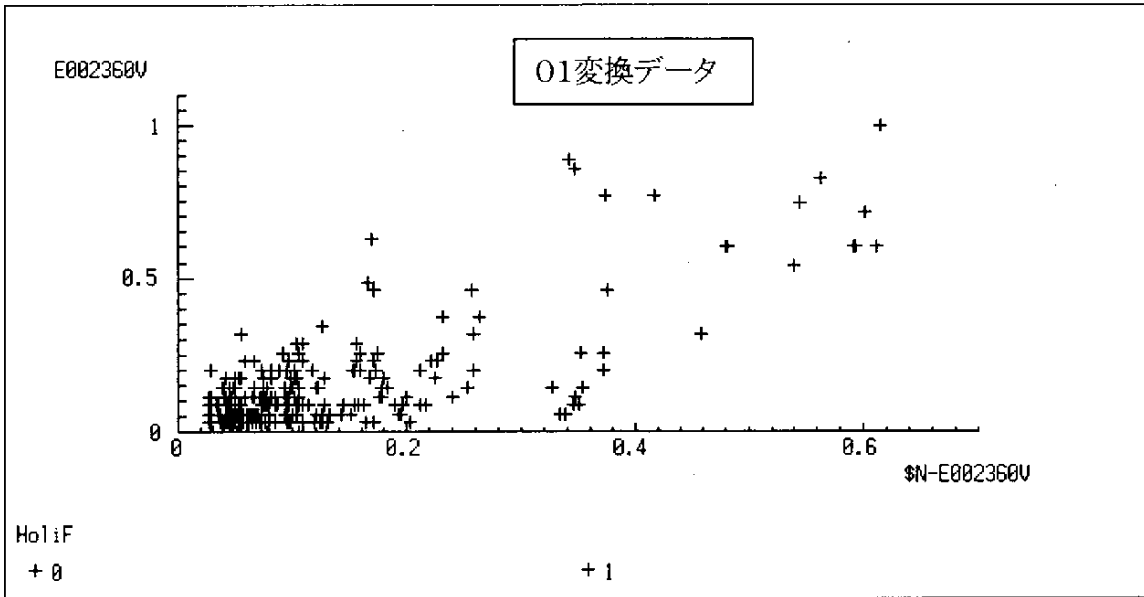


図5-27 ニューラルネットワークによる学習結果、実績値との比較 (C店)

表5-24 重要度分析 (01変換データ)

ニューラルネットワークアーキテクチャ: "E002360V"	
入力層	: 17 ニューロン
隠れ層 #1	: 13 ニューロン
出力層	: 1 ニューロン
予想正確度	: 93.33%

表5-25 精度分析 (01変換データ)

出力フィールドの結果: E002360V	
\$N-E002360V のエラー分析	
エラーの最小値	: -0.28121
エラーの最大値	: 0.54339
エラーの平均値	: 0.011431
絶対エラーの平均値	: 0.067022
標準偏差	: 0.10299

表5-24 重要度分析 続き

入力の関係重要度	
K467574X	: 0.15052
K645118X	: 0.14505
E002117X	: 0.09878
K645125X	: 0.09704
E002360X	: 0.08443
E003336X	: 0.08181
E002124X	: 0.05601
HoliF	: 0.05325
K341489X	: 0.05213
E002001X	: 0.03455
CustX	: 0.03194
E002018X	: 0.02816
H480748X	: 0.02266
K342189X	: 0.02107
E003183X	: 0.01989
E003206X	: 0.01571
H480618X	: 0.00995

表5-25 精度分析 続き

線形相関	: 0.75449
出現頻度	: 322

表5-26 重要度分析 (生データ)

ニューラルネットワークアーキテクチャ: "E002360K"	
入力層	: 20 ニューロン
隠れ層 #1	: 15 ニューロン
出力層	: 1 ニューロン
予想正確度	: 93.49%
入力の関係重要度	
K645118P	: 0.15347
K645125P	: 0.12498
HoliF	: 0.10023
E003336P	: 0.09603
E002100P	: 0.08220
E002360P	: 0.08146
K467574P	: 0.07193
K342189P	: 0.06697
Cust	: 0.06483
E002117P	: 0.06178
E003206P	: 0.05469
E001165P	: 0.05222
H480618P	: 0.04950
E002124P	: 0.04673
E003183P	: 0.04176
E003220P	: 0.02833
K341489P	: 0.02557
E002018P	: 0.02001
H480748P	: 0.01436
E002001P	: 0.01360

表5-27 精度分析 (生データ)

出力フィールドの結果: E002360K	
\$N-E002360K のエラー分析	
エラーの最小値	: -13
エラーの最大値	: 20
エラーの平均値	: -0.0093168
絶対エラーの平均値	: 2.0652
標準偏差	: 3.2836
線形相関	: 0.80172
出現頻度	: 322

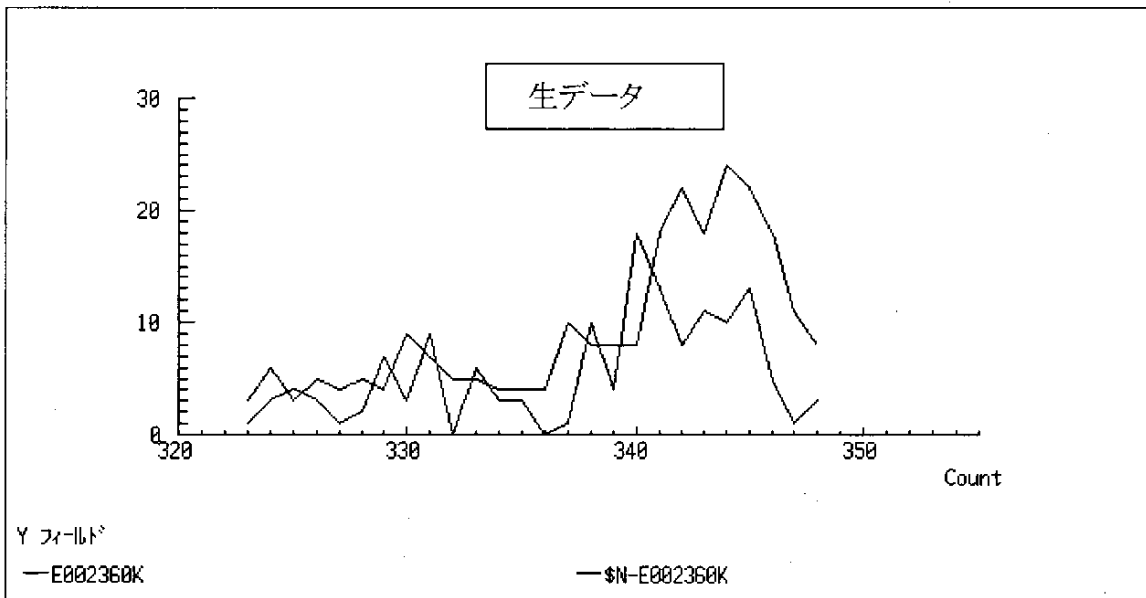
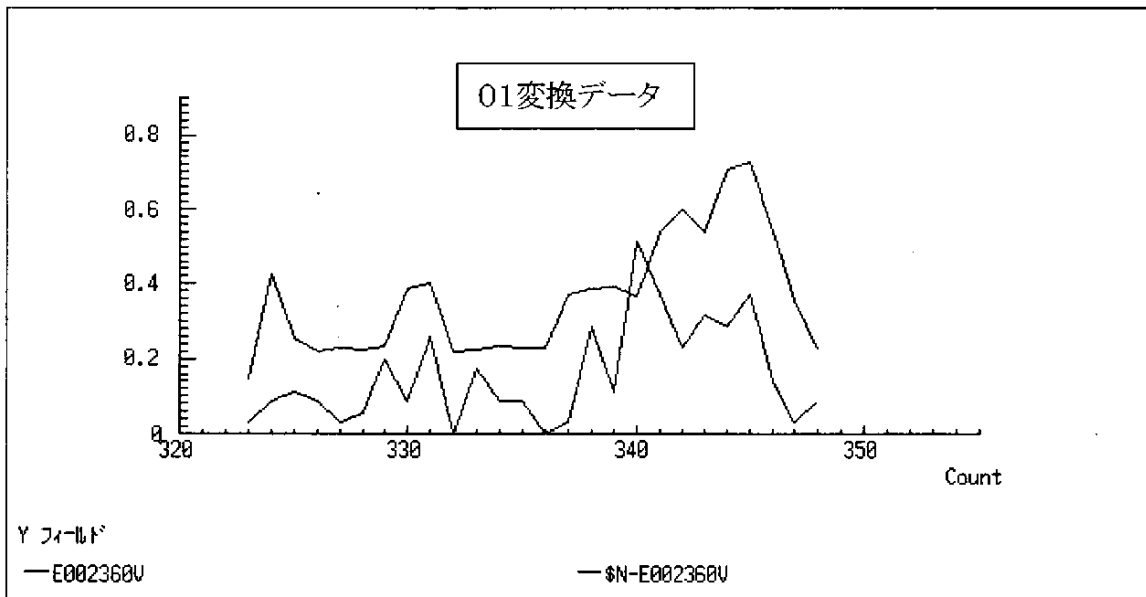


図5-28 検証 実績値との比較 (C店)

表5-28 精度分析 (左: 01変換データ、右: 生データ)

出力フィールドの結果: E002360V \$N-E002360V のエラー分析		出力フィールドの結果: E002360K \$N-E002360K のエラー分析	
エラーの最小値	: -0.41921	エラーの最小値	: -14
エラーの最大値	: 0.14564	エラーの最大値	: 10
エラーの平均値	: -0.20574	エラーの平均値	: -3.8846
絶対エラーの平均値	: 0.21694	絶対エラーの平均値	: 5.3462
標準偏差	: 0.12974	標準偏差	: 5.5809
線形相関	: 0.62883	線形相関	: 0.55555
出現頻度	: 26	出現頻度	: 26

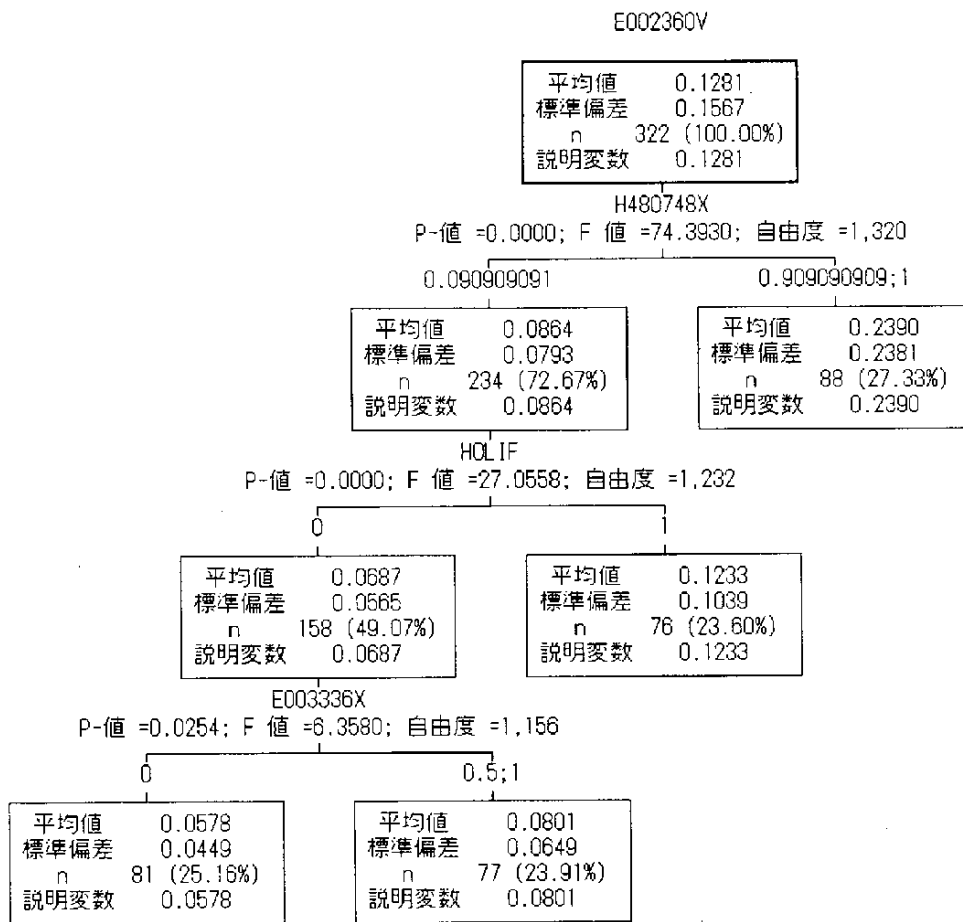


図5-29 Answer Tree (CHAID) による分析結果 (C店、01変換データ)

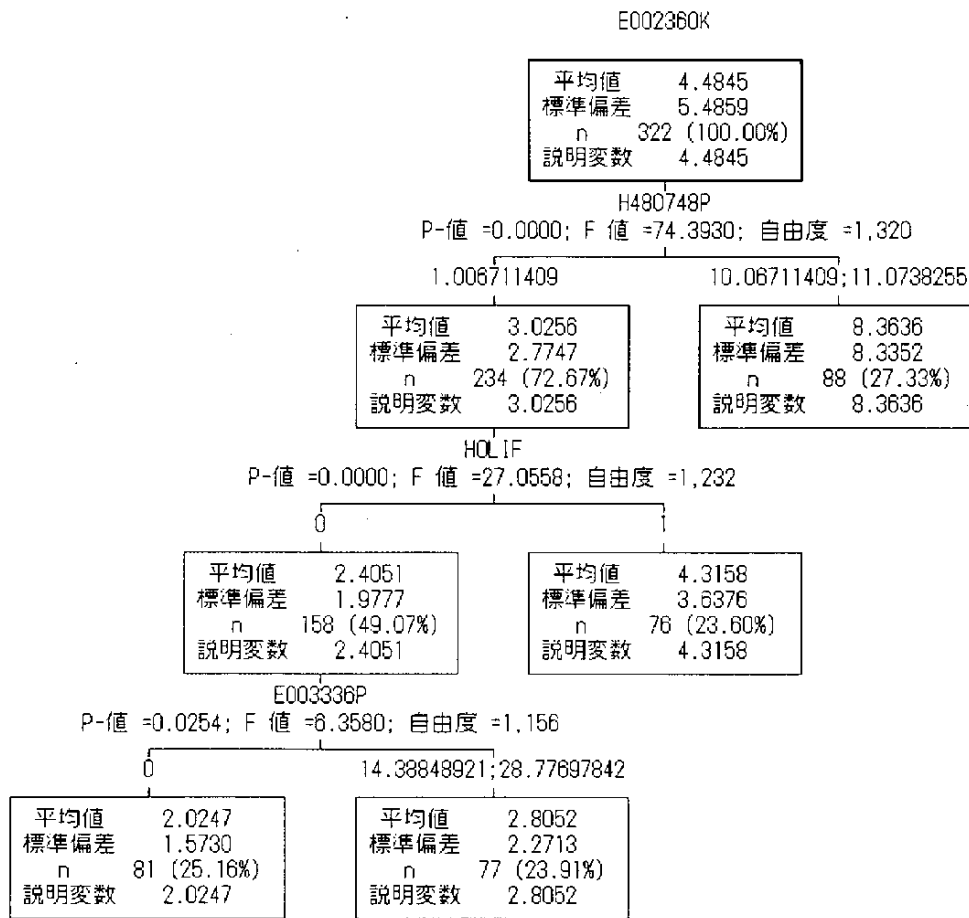


図5-30 Answer Tree (CHAID) による分析結果 (C店、生データ)

C店では0 1変換データ、生データの両方を使って販売動向を分析した(図5-26、27)。両方とも実績値をなだらかにしたような感じで、だいたいの傾向をつかんでいる。重要度分析を見ると(表5-25、26)、0 1変換データと生データでは大きく異なる。0 1変換データでは「キッコーマン 新撰焼肉 赤と黒 プラ 210G」、「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率に販売個数は大きな影響を受ける。一方、生データでは、「キッコーマン 新撰焼肉 黒だれ プラ 210G」、「キッコーマン 新撰焼肉和風香味だれねぎ生姜210G」の値引率の影響を受ける。Answer Treeの分析結果を見ると(図5-29、30)Clementineとは異なり、「ハウスさっと逸品焼肉ドレまろやかごま醤油300ML」、休日フラグ、「エバラ 粒ごものたれ 豊潤 瓶 280G」の影響を受ける。ただし、Answer Treeの結果では、ライバル商品が値引きをした方が、「エバラ おろしのたれ 焼肉のたれ 瓶 270G」の売り上げが伸びるといようにツリーが分割されており、常識とは異なった結果になっている。

次に求めたモデルを3月の売上データで検証してみた(図5-28)。あまり当てはまっているとは言い難い。

「キッコーマン 新撰焼肉 赤と黒 プラ 210G」 A店

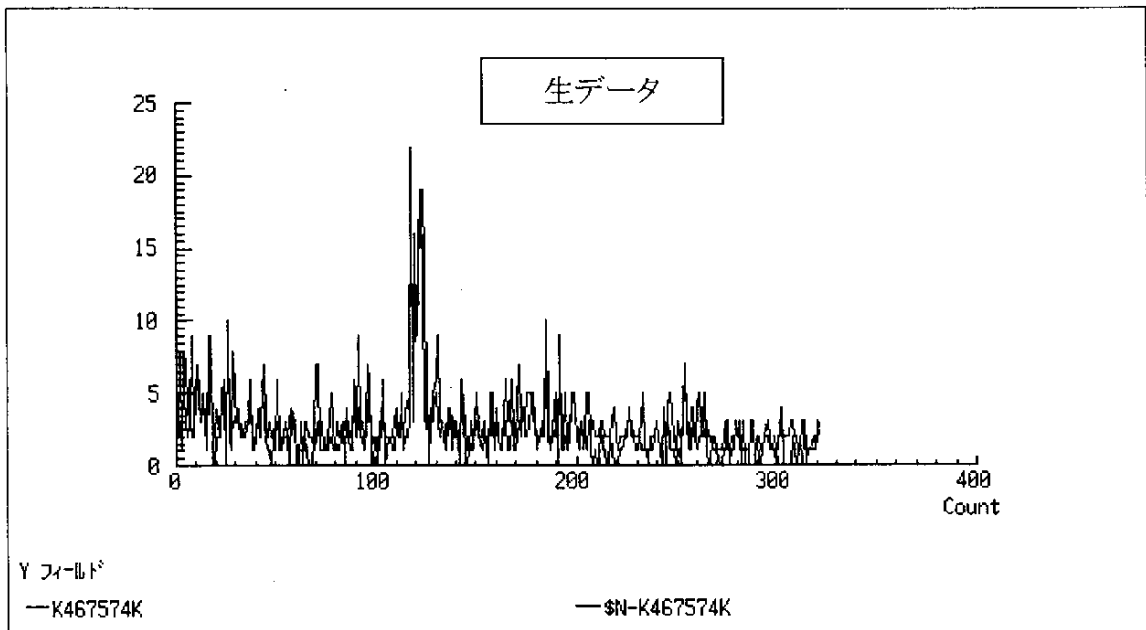


図5-31 ニューラルネットワークによる学習結果、実績値との比較 (A店)

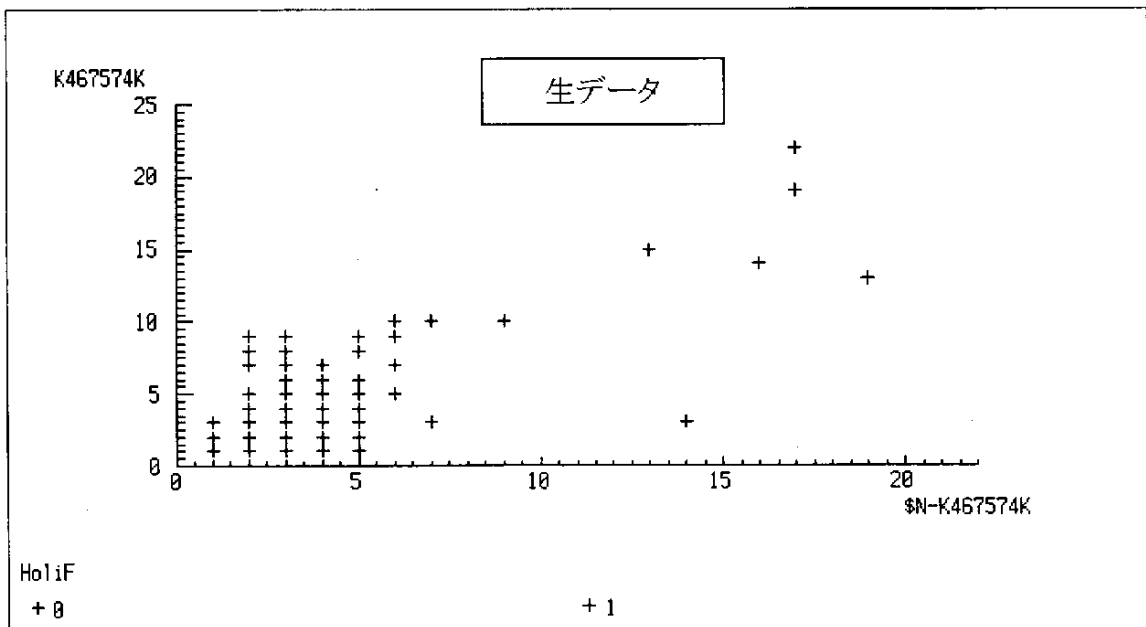


図5-32 ニューラルネットワークによる学習結果、実績値との比較 (A店)

表5-29 重要度分析 (生データ)

ニューラルネットワークアーキテクチャ: "K467574K"

入力層 : 10 ニューロン
 隠れ層 #1 : 8 ニューロン
 出力層 : 1 ニューロン

予想正確度 : 93.95%

入力の関係重要度

Cust	: 0.17237
E002360P	: 0.13638
K645118P	: 0.11444
M330204P	: 0.09901
K467574P	: 0.08857
M340166P	: 0.08251
HoliF	: 0.07533
S040231P	: 0.06269
E002025P	: 0.04875
E003220P	: 0.04031

表5-30 精度分析 (生データ)

出力フィールドの結果: K467574K

\$N-K467574K のエラー分析

エラーの最小値	: -11
エラーの最大値	: 7
エラーの平均値	: -0.13396
絶対エラーの平均値	: 1.4424
標準偏差	: 1.9727
線形相関	: 0.71663
出現頻度	: 321

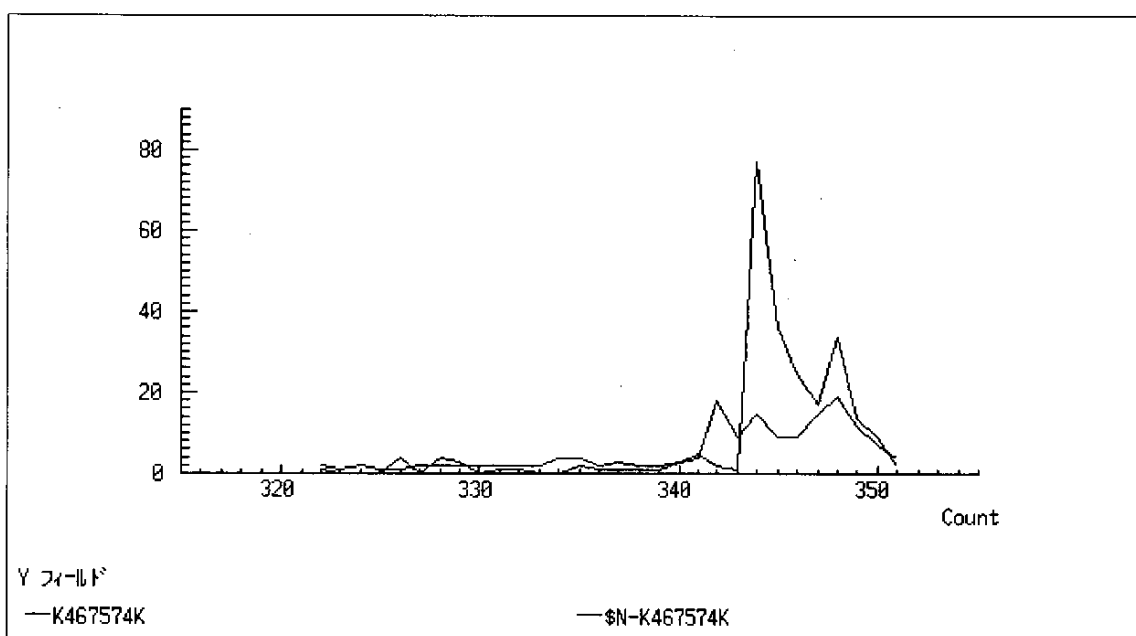


図5-33 検証 実績値との比較 (A店)

表5-31 精度分析 (生データ、A店)

出力フィールドの結果: K467574K	
\$N-K467574K のエラー分析	
エラーの最小値	: -16
エラーの最大値	: 62
エラーの平均値	: 2.7667
絶対エラーの平均値	: 6.1000
標準偏差	: 13.421
線形相関	: 0.64888
出現頻度	: 30

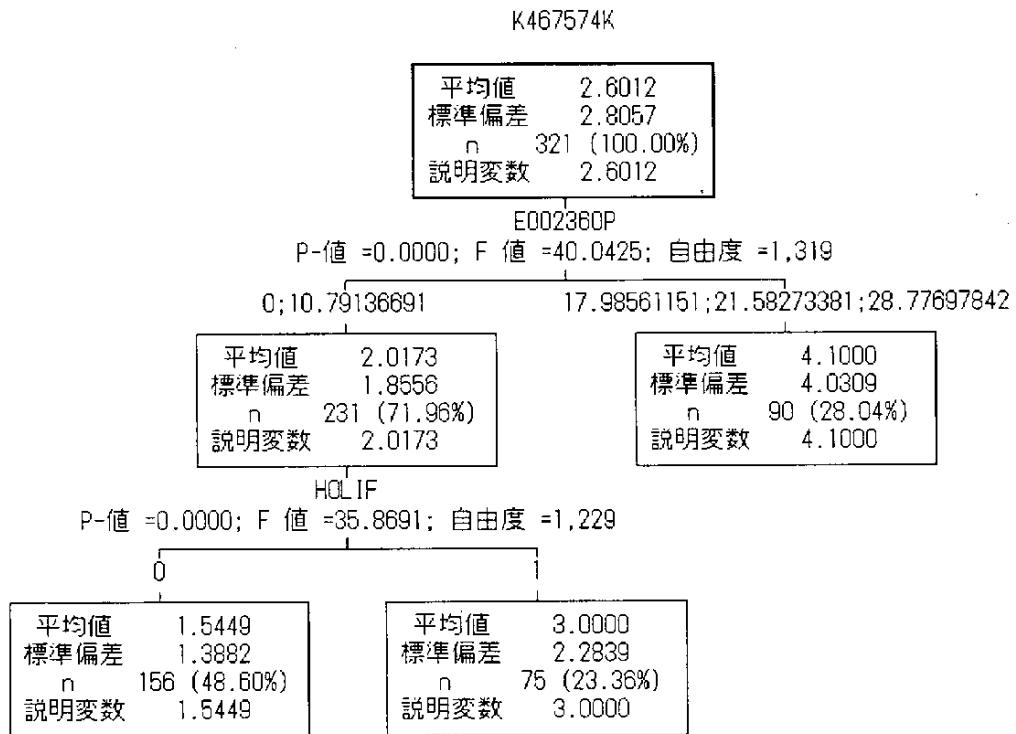


図5-34 Answer Tree (CHAID) による分析結果 (A店、生データ)

「キッコーマン 新撰焼肉 赤と黒 プラ 210G」のA店での販売動向を、生データを使って分析した(図5-31、32)。特異点も含めて比較的よくあっている。重要度分析を見ると(表5-29)、来店客数、「エバラ おろしのたれ 焼肉のたれ 瓶 270G」、「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率の影響が大きい。自分自身の値引率の影響は大きくない。Answer Tree の分析結果を見ると(図5-34)、「エバラ おろしのたれ 焼肉のたれ 瓶 270G」の値引率、休日フラグで販売個数を分割している。休日フラグと来店客数は似たような動きをするので、おおむね同じ結果になっていると考えてよい。ただし、ここでもライバル商

品が値引きをした方が売り上げが伸びるという結果になっている。

次に求めたモデルを3月の売上データで検証してみた(図5-33)。特異点があるため、結果はもうひとつである。

「キッコーマン 新撰焼肉 赤と黒 プラ 210G」 B店

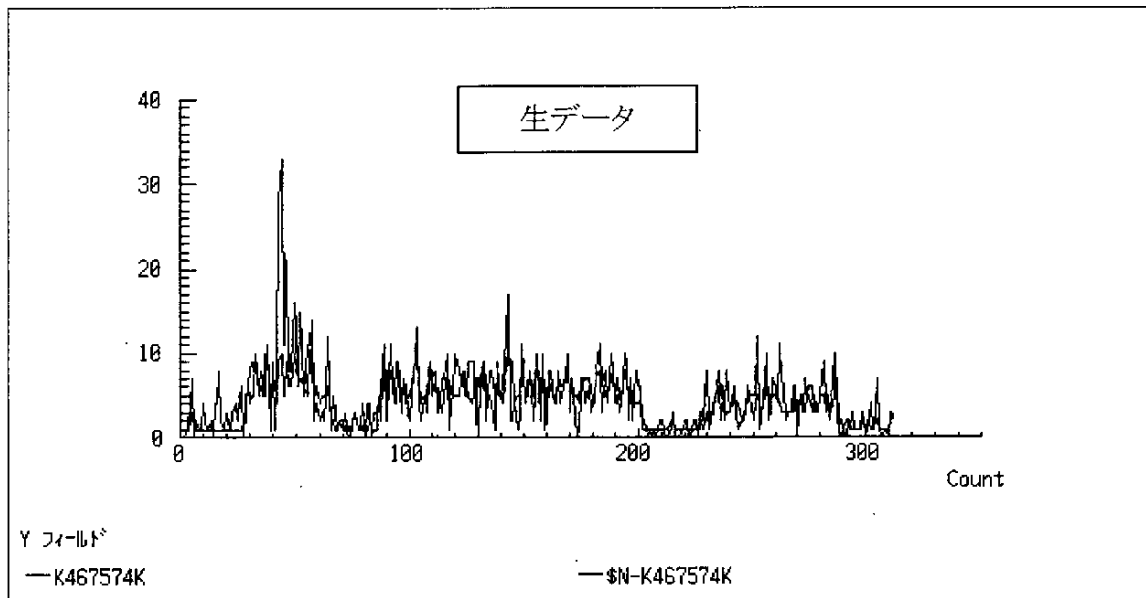


図5-35 ニューラルネットワークによる学習結果、実績値との比較 (B店)

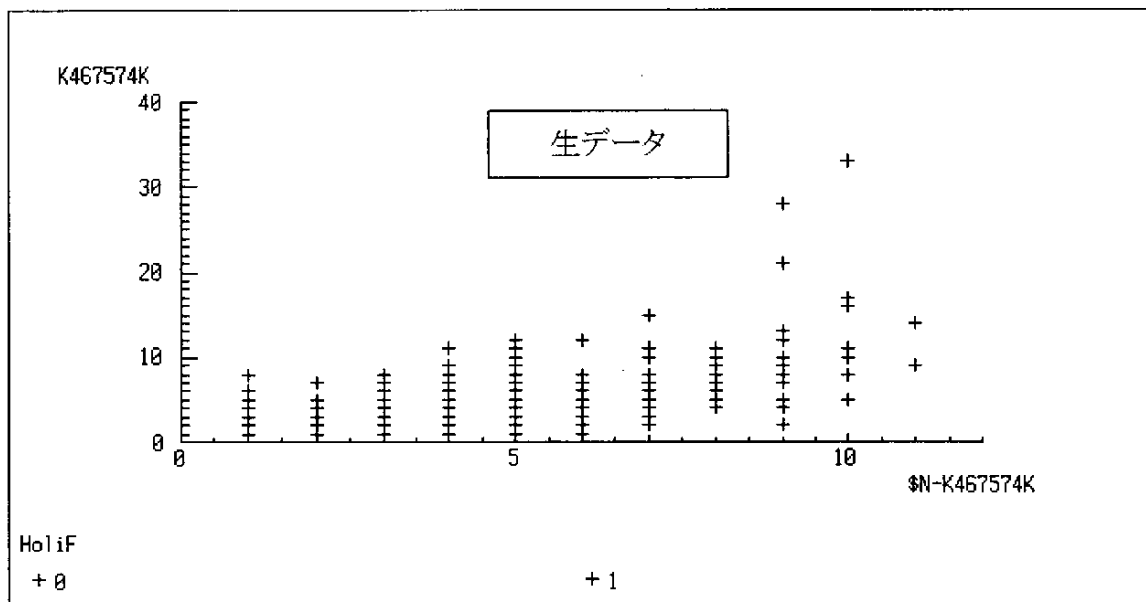


図5-36 ニューラルネットワークによる学習結果、実績値との比較 (B店)

表5-32 重要度分析 (生データ)

ニューラル・ネットワーク・アーキテクチャ: "K467574K"

入力層 : 15 ニューロン
 隠れ層 #1 : 12 ニューロン
 出力層 : 1 ニューロン

予想正確度 : 94.29%

入力の関係重要度

Cust	: 0.12344
K467574P	: 0.06865
HoliF	: 0.06190
E002117P	: 0.05710
M070793P	: 0.05380
K645118P	: 0.04067
E003220P	: 0.03725
E002360P	: 0.03483
E002025P	: 0.03221
M105129P	: 0.02919
E001165P	: 0.01981
E003336P	: 0.01334
E002001P	: 0.01009
N212270P	: 0.00389
E002018P	: 0.00231

表5-33 精度分析 (生データ)

出力フィールドの結果: K467574K
 \$N-K467574K のエラー分析

エラーの最小値	: -7
エラーの最大値	: 23
エラーの平均値	: 0.46154
絶対エラーの平均値	: 1.9615
標準偏差	: 3.0153
線形相関	: 0.67180
出現頻度	: 312

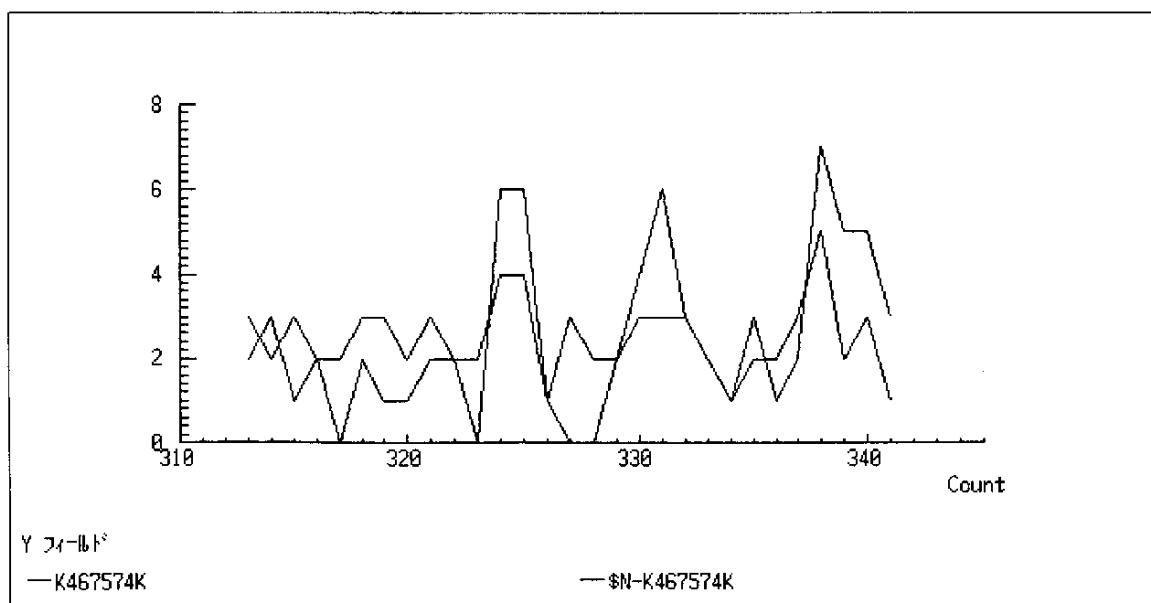


図5-37 検証 実績値との比較 (B店)

表5-34 精度分析 (生データ、B店)

出力フィールドの結果: K467574K	
\$N-K467574K のエラー分析	
エラーの最小値	: -3
エラーの最大値	: 3
エラーの平均値	: 0.13793
絶対エラーの平均値	: 1.3103
標準偏差	: 1.6415
線形相関	: 0.57463
出現頻度	: 29

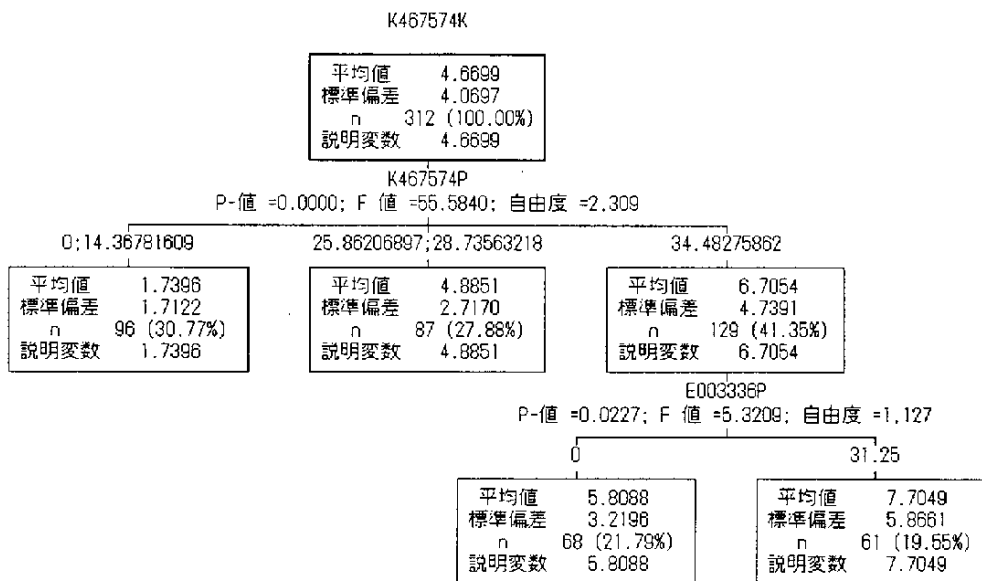


図5-38 Answer Tree (CHAID) による分析結果 (B店、生データ)

B店でも生データを使って販売動向を分析した (図5-35、36)。突出した売り上げがあった日は予測できていないが、概ねだいたいの傾向をつかんでいる。重要度分析を見ると (表5-32)、来店客数、次に自分自身の価格 (値引率) が強い影響力を持っている。Answer Tree の分析結果を見ると (図5-38)、自分自身の値引率、「エバラ 粒ごまのたれ 豊潤 瓶 280G」の値引率の影響が強い。ただし、後者は値引率が高いほど、「キッコーマン 新撰焼肉 赤と黒 プラ 210G」がよく売れる結果となっており常識と合わない。

次に求めたモデルを3月の売上データに当てはめてみた (図5-37)。結果は芳しくない。

「キッコーマン 新撰焼肉 赤と黒 プラ 210G」 C店

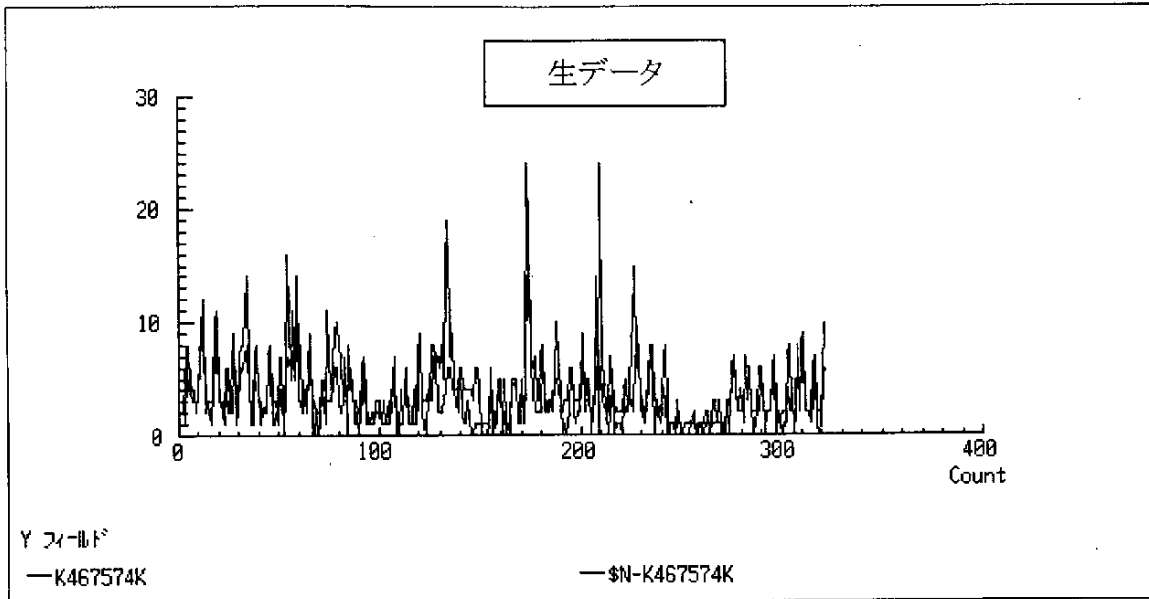


図5-39 ニューラルネットワークによる学習結果、実績値との比較 (C店)

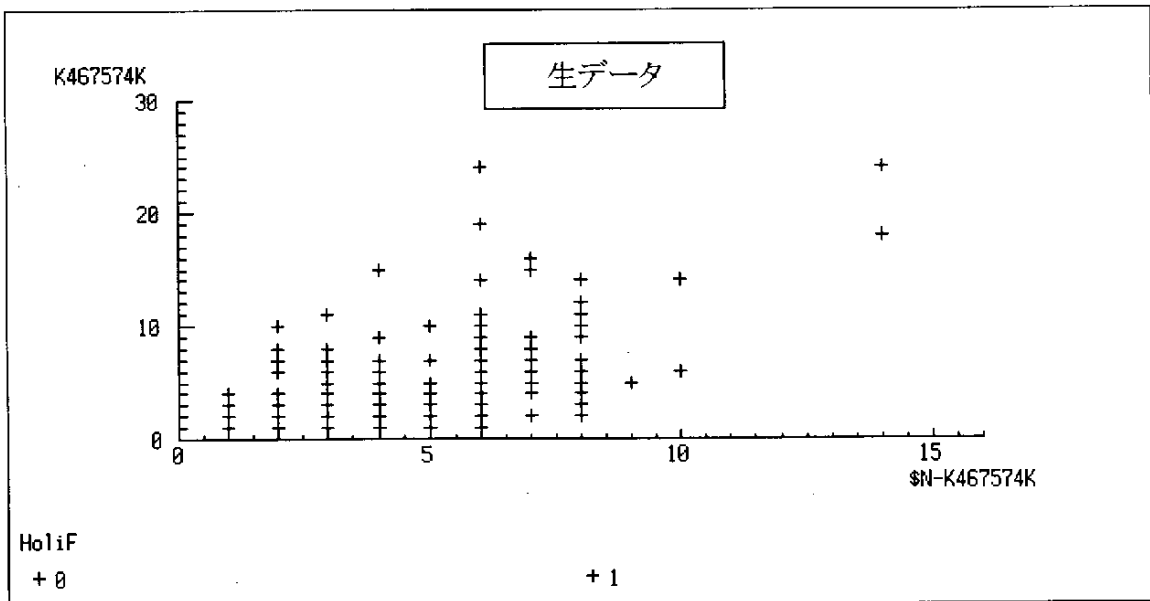


図5-40 ニューラルネットワークによる学習結果、実績値との比較 (C店)

表5-35 重要度分析 (生データ)

ニューラルネットワーク・アーキテクチャ: "K467574K"

入力層 : 10 ニューロン
 隠れ層 #1 : 15 ニューロン
 出力層 : 1 ニューロン

予想正確度 : 93.12%

入力の関係重要度

K341489P	: 0.25241
K467574P	: 0.23007
K342189P	: 0.18559
HoliF	: 0.12603
Cust	: 0.08663
H480618P	: 0.08257
E002018P	: 0.05083
E002001P	: 0.04721
E002117P	: 0.04013
E002124P	: 0.02390

表5-36 精度分析 (生データ)

出力フィールドの結果: K467574K

\$N-K467574K のエラー分析

エラーの最小値	: -6
エラーの最大値	: 18
エラーの平均値	: 0.0093168
絶対エラーの平均値	: 1.8789
標準偏差	: 2.7991
線形相関	: 0.63881
出現頻度	: 322

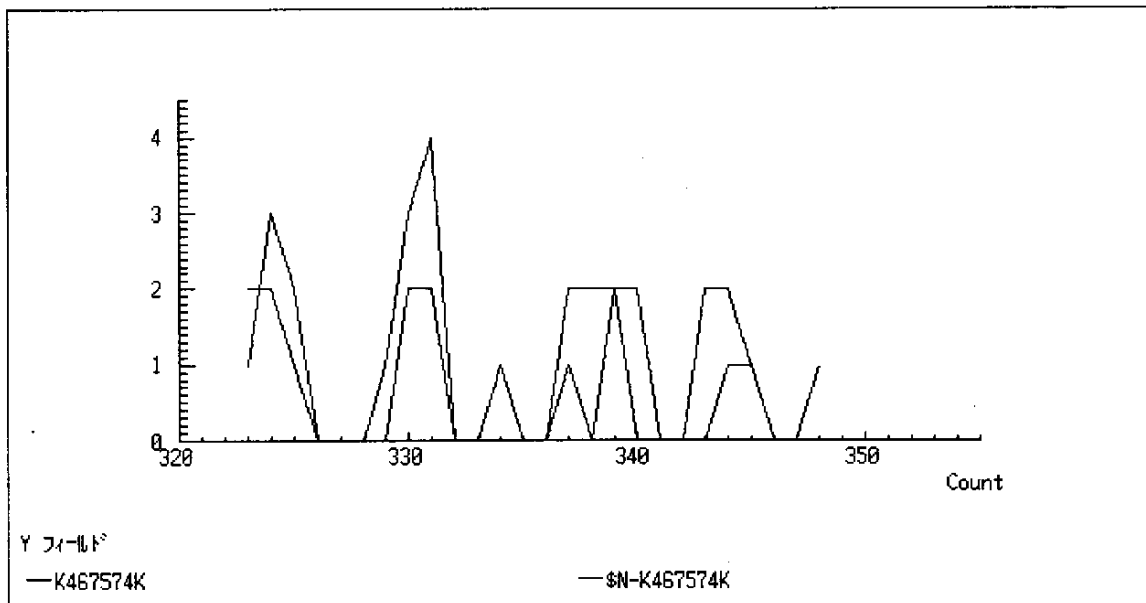


図5-41 検証 実績値との比較 (C店)

表5-37 精度分析 (生データ、C店)

出力ファイルの結果: K467574K	
\$N-K467574K のエラー分析	
エラーの最小値	: -2
エラーの最大値	: 2
エラーの平均値	: 0.34615
絶対エラーの平均値	: 0.65385
標準偏差	: 0.93562
線形相関	: 0.61535
出現頻度	: 26

K467574K

平均値	3.5497
標準偏差	3.6360
n	322 (100.00%)
説明変数	3.5497

HOLIF

P-値 =0.0000; F 値 =77.5424; 自由度 =1,320

0

平均値	2.4186
標準偏差	2.5212
n	215 (66.77%)
説明変数	2.4186

平均値	5.8224
標準偏差	4.4037
n	107 (33.23%)
説明変数	5.8224

K467574P

P-値 =0.0004; F 値 =16.3085; 自由度 =1,213

0

20.11494253; 24.42528736; 25.86206897; 28.73563218; 33.33333333

平均値	1.2157
標準偏差	1.0453
n	51 (15.84%)
説明変数	1.2157

平均値	2.7927
標準偏差	2.7234
n	164 (50.93%)
説明変数	2.7927

図5-42 Answer Tree (CHAID) による分析結果 (C店、生データ)

C店でも生データを使って販売動向を分析した(図5-39、40)。突出した売り上げがあった日は予測できていないが、概ねだいたいの傾向をつかんでいる。重要度分析を見ると(表5-35)、「キッコーマン 新撰焼肉 黒だれやや甘口プラ400G」、次に自分自身、「キッコーマン 新撰焼肉 赤と黒 プラ 400G」の価格(値引率)、休日フラグが強い影響力を持っている。Answer Treeの分析結果を見ると(図5-42)、休日フラグ、自分自身の値引率の影響が強い。

次に求めたモデルを3月の売上データに当てはめてみた(図5-41)。そこそこあっている。

5. 5 まとめ

ここまで行った分析結果について、店別に売り上げに影響を与える要因上位2~3を一覧表にまとめた(表5-38)。ただし、影響の度合いは考慮していない。また、本文中には書かなかった結果も一緒に記述する。

表5-38A エバラ 黄金の味 甘口 フルーツタイプ 瓶210G

店舗名	01変換データ	生データ
A店	来店客数 休日フラグ	自分自身の値引率 休日フラグ
B店	「エバラ 黄金の味 中辛フルーツ &スパイス瓶210G」の値引率 「エバラ 黄金の味 辛口 スパイ スタイブ 瓶210G」の値引率	「日本食研 晚餐館 焼肉のたれ 中辛 210G」の値引率 休日フラグ 「エバラ 黄金の味 中辛フルーツ &スパイス瓶210G」の値引率 自分自身の値引率
C店	休日フラグ 「エバラ 焼肉のたれ スパイスに んにく味 瓶180G」の値引率 「エバラ 黄金の味 中辛フルーツ &スパイス瓶210G」の値引率	来店客数 休日フラグ 「ハウスざっと逸品焼肉ドレすっき りレモン醤油300ML」の値引率

表5-38B エバラ おろしのたれ 焼肉のたれ 瓶 270G

店舗名	01変換データ	生データ
A店	来店客数 自分自身の値引率 「エバラ 焼肉のたれ 甘口 瓶 300G」の値引率	来店客数 自分自身の値引率 「エバラ 焼肉のたれ 醤油味 瓶 300G」の値引率
B店	自分自身の値引率 来店客数 「日本食研 晚餐館 焼肉のたれ 中辛 210G」の値引率	自分自身の値引率 来店客数 「日本食研 晚餐館 焼肉のたれ 中辛 210G」の値引率
C店	自分自身の値引率 「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率 「エバラ 黄金の味 中辛フルーツ &スパイス瓶400G」の値引率	「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率 「キッコーマン 新撰焼肉和風香味 だれねぎ生姜210G」の値引率 休日フラグ

表5-38C キッコーマン 新撰焼肉 赤と黒 プラ 210G

店舗名	01変換データ	生データ
A店	来店客数 「エバラ おろしのたれ 焼肉のたれ 瓶 270G」の値引率 自分自身の値引率	来店客数 「エバラ おろしのたれ 焼肉のたれ 瓶 270G」の値引率 「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率
B店	来店客数 自分自身の値引率 「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率	来店客数 自分自身の値引率 休日フラグ
C店	「キッコーマン 新撰焼肉 黒だれ プラ 210G」の値引率 「エバラ 黄金の味 中辛フルーツ &スパイス瓶400G」の値引率	「キッコーマン 新撰焼肉 黒だれ やや甘口プラ400G」の値引率 自分自身の値引率 「キッコーマン 新撰焼肉 赤と黒 プラ 400G」の値引率

Clementine のニューラルネットワークを使って売れ行きの要因を分析したが、やはりニューラルネットワークはわかりにくい。どんな分析にでも使えるので便利であるが、基本的にブラックボックスなので結果は出るが解釈が出来ないという状況になる。Clementine は重要度分析で影響を与える入力を数値で示してくれるが、プラスに働くのかマイナスに働くのかわからない。今回は Answer Tree の助けを借りた。

また、今回は入出力を01変換したり生データのまま使ったりしたが、必ずしも結果は一致しなかった。また、Answer Tree の結果も異なることがあった。このような場合の解釈の仕方が難しい。どの入力にもあまり関係がないから異なる結果になったのかもしれない。もう少し、詳しいマニュアルがあれば解決したかもしれない。

Answer Tree ではライバル商品が値下げをしたら、自分も売り上げが上がるという結果が何度か出た。このような場合はデータにもどって調べてみる必要がある。ライバル商品と同じ時期に値下げをしているかもしれない。

今回の分析ではなかなか明確な結果が出てこなかった。いろいろなデータマイニングソフトを使って結果を出すのは簡単であるが、その解釈が成功への鍵を握る。やはり、データの内容を見ることがデータマイニングでも重要である。

6. 牛乳販売量の予測モデル

6. 1 目的

POS データとして収集される牛乳の日次販売量（明日の売れ行き）を予測したい。比較的容易に入手できる価格データや気象データなどから、できるだけ安定して使える予測モデルを構成することを第一の目的とする。第二に予測手法として伝統的な重回帰分析モデル（線形モデル）とニューラルネットワークモデル（非線形モデル）の2種類を適用して、どちらの予測精度が良いかを検討する。第三の目的として、最近ソフトウェア各社が提供を始めている「データマイニング・ツール」のひとつである SAS 社の「Enterprise Miner Version 2.02」（以下、EM と略称）を使用して分析ツールの実用性の検討をする。

6. 2 分析の方法

POS データベースから分析用データを CSV 形式で取り出した状態から分析を開始する。まず、生データの吟味・予備解析を実施してモデル構成に有望な変数を探索しながらデータに対する理解を深めていく。

次に EM を使って、ニューラルネットワークと重回帰分析による予測モデルの構成と評価を実施する。

6. 3 分析データ

- 対象店舗：首都圏に所在するスーパーマーケット5店（A、B、C、D、Eで表示する）
- 対象期間：1998年4月1日から1999年11月までの1年8ヶ月間（20ヶ月）
- 対象商品：牛乳
- 測定変数：表6-1のとおり（このほか、予備解析の過程で既存変数から合成・加工した新しい変数については解析の節で説明する）

データの形式は多変量データ行列であり、データセットの1行は「ある店舗の、ある日に売れた、ある牛乳に関する、販売量」に相当する。すなわち、もっとも下位レベルである「ある日のレシート1枚」を1行とするデータを、商品ごとに集計して1日販売量を1行とするデータが得られていることになる。従って1日分のデータは牛乳商品の種類だけの行数となる。これが各店舗ごと

に得られている。

気象データはPOSシステムから収集されたものではなく外部機関から購入してPOSデータと連結編集した。

表6-1 分析データに含まれる20個の変数

P O S デ ー タ	1	店舗コード	名義尺度(A~E)	気 象 デ ー タ	10	平均気温	間隔尺度
	2	日付	順序尺度		11	最高気温	間隔尺度
	3	曜日	名義尺度(1~7)		12	最低気温	間隔尺度
	4	月	名義尺度(1~12)		13	平均差	間隔尺度
	5	商品名	名義尺度		14	最高差	間隔尺度
	6	販売個数	間隔尺度		15	最低差	間隔尺度
	7	販売金額	間隔尺度		16	湿度	間隔尺度
	8	価格	間隔尺度		17	降水量	間隔尺度
	9	来店人数	間隔尺度		18	日照時間	間隔尺度
					19	不快指数	間隔尺度
			20	天気	名義尺度		

6. 4 予備解析

6. 4. 1 単純集計によるデータ概観

各店舗の20ヶ月間(609日間)の営業日数を表6-2に示した。休店日を曜日別に見ると、定休日が月1~2回で、ほかに年間で2、3日の特別休店日があることが推察できる。なお12月だけは5店とも無休であった。

対象期間中に販売実績のあったすべての牛乳商品の種類は、店舗によって21種類から43種類と多様である。ただし、すべての種類が年間を通じて継続的に販売されているわけではない。

表6-2 分析対象期間(20ヶ月)における営業日数と商品種類

店舗	A	B	C	D	E
営業日数	587	588	584	571	588
休店日数	22	21	25	38	21
牛の乳商品種類	43	23	23	21	35

商品の容量を1%商品と、1%以外容量の商品に分けて販売個数を比較した結果を図6-1に示す。5店とも90%前後が1%牛乳で占めている。図6-2は同様に売上金額で比較した結果である。1%商品のシェアはさらに3ポイント程度上昇する。従って分析対象を1%商品だけに限定しても情報の損失はほとんどない。

表6-3は1%商品に限定してA店の月別の販売日数を集計した結果である。新商品や販売中止などで20ヶ月間を通じて継続的に販売されていない種類も多い。この傾向は他店でも同様である。

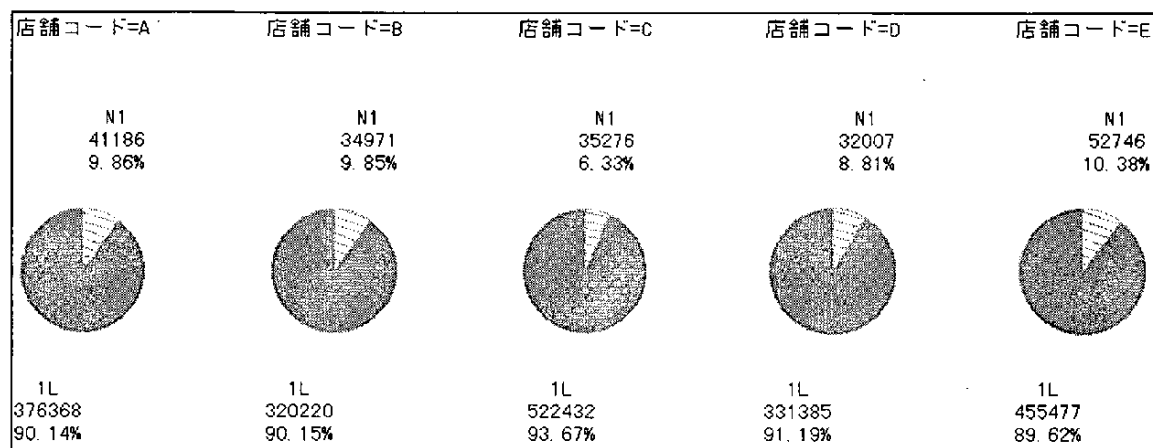


図6-1 店舗別・1%商品と1%以外商品の「販売個数」の比較

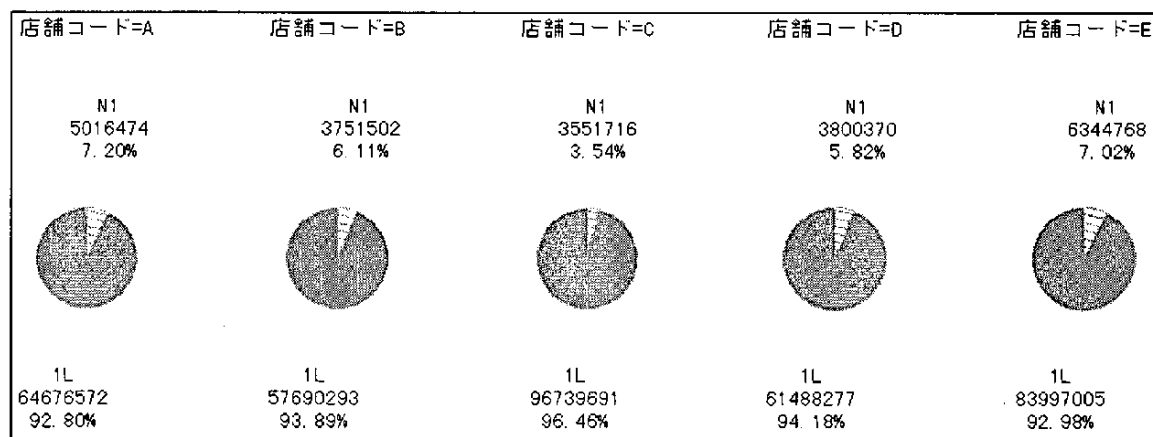


図6-2 店舗別・1%商品と1%以外商品の「販売金額」の比較

分析対象期間を通じて継続的に販売されている牛乳商品に限定した方がモデル構成に好都合なので、対象20ヶ月すべての月で販売実績のある商品に限定する。結果的には、およそ400日以上販売日数の商品が該当する。除外される商品の販売量は非常に少ないので、このように分析対象を限定しても、全体的傾向を把握する目的にとっては障害にはならないと考えられる。

この限定によって、A店の場合は図6-3のように7商品が該当することとなるが、実質的にはPB（プライベートブランド）の1商品が74%の圧倒的シェアを占めており、上位数種類の商品だけを分析対象にしても全体的傾向が把握できる。

表 6-3 A店における1%牛乳の月間・販売実績(日数)

	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	計
1 特濃4.3牛乳	29	30	28	31	31	29	29	25	31	30	27	31	29	30	29	31	31	29	28	29	587
2 PB-1 3.5牛乳 紙	29	30	28	31	31	29	29	25	31	30	27	31	29	30	29	31	31	29	28	29	587
3 PB-2 北海道3.7牛乳	29	30	28	31	31	29	29	25	31	30	25	31	28	30	29	31	31	29	28	29	584
4 PB-3 1.0低脂肪乳	29	30	28	31	31	29	29	25	31	29	27	31	29	30	29	31	31	29	28	29	586
5 PB-4 富良野沿線3.6	27	29	25	30	29	26	26	24	28	24	25	23	23	26	26	22	27	26	19	19	504
6 古谷 低温殺菌3.6牛	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	31	29	28	29	123
7 古谷 低温殺菌牛乳	0	0	0	0	0	0	23	25	31	30	27	30	29	30	29	24	0	0	0	0	278
8 古谷 特別仕立酪農	29	30	28	31	31	29	29	25	31	30	27	31	29	30	29	31	31	29	28	29	587
9 高梨 低温殺菌牛乳	29	30	28	31	31	29	29	16	0	0	0	0	0	0	0	0	0	0	0	0	223
10 小岩井 農場3.7牛乳	29	30	28	31	31	29	29	22	0	0	0	0	0	0	0	0	0	0	0	0	229
11 森永乳 ドライモン牛乳	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	31	31	29	27	26	172
12 榛名酪農 朝/牛乳物	0	0	0	0	0	0	0	0	0	0	0	0	0	12	28	31	31	29	28	29	188
13 雪印乳 3.5牛乳 紙	29	30	28	31	31	29	29	25	31	30	27	31	29	30	29	31	31	29	28	29	587
14 雪印乳 低脂肪乳 加	25	30	24	31	31	29	29	22	0	0	0	0	0	0	2	0	1	1	2	0	227
15 全酪 JM ハンドキリ 牛乳	29	30	28	31	31	29	28	24	31	30	27	31	29	10	0	0	0	0	0	0	388

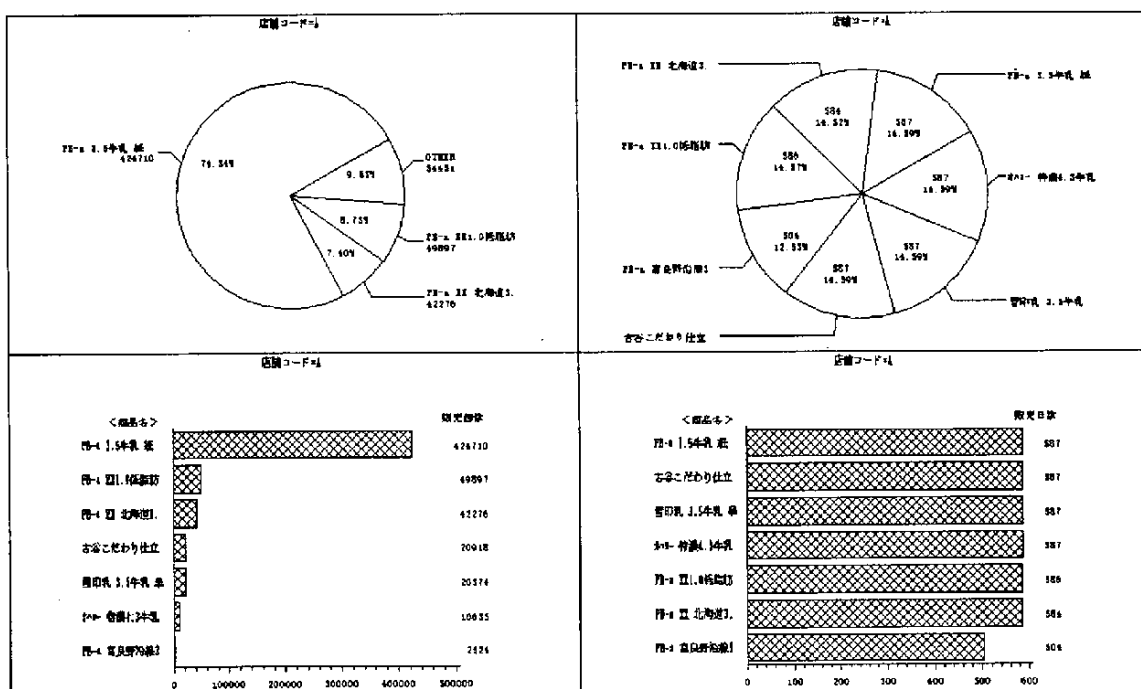


図 6-3 対象 20 ヶ月に継続販売された商品の販売個数シェアと販売日数 [A店]

上位の固定商品のシェアが高い傾向は他店でも同様であり。B店～E店については図 6-4 と図 6-5 に示した。

6. 4. 2 販売個数と販売金額の関係

予測目的は「日次販売量」であるが、この場合の指標として販売個数と販売

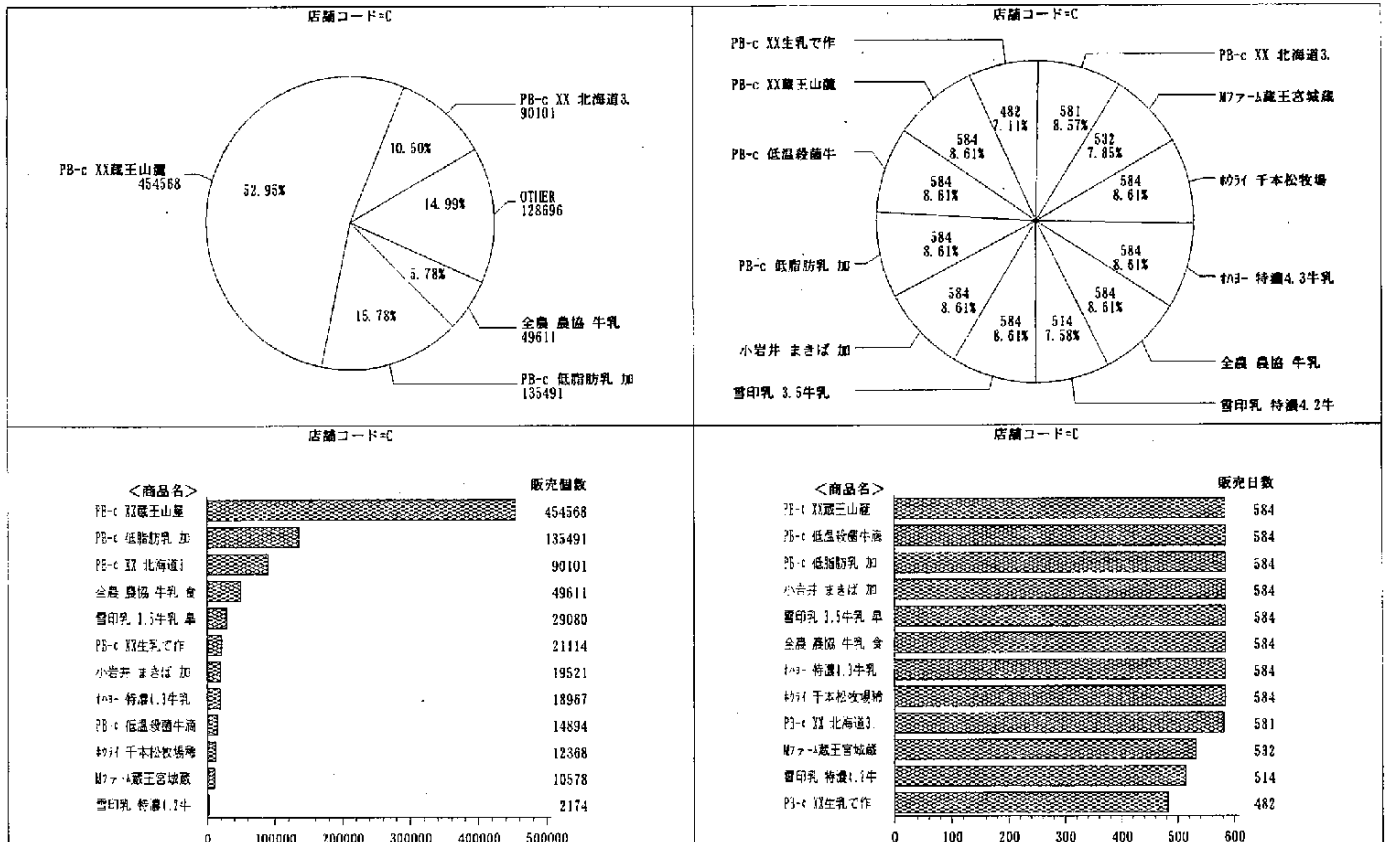
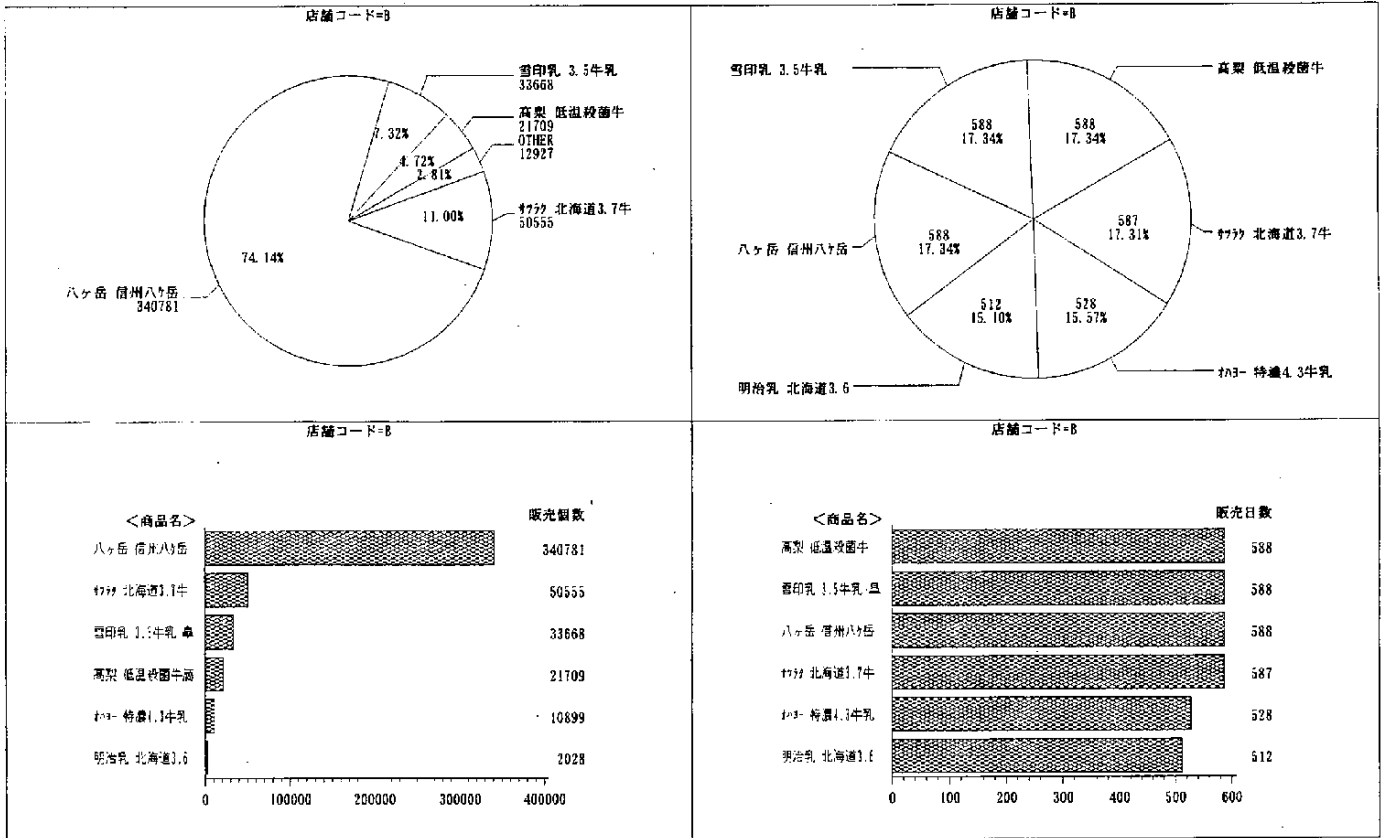


図6-4 対象20ヶ月間に継続販売された商品の販売個数シェアと販売日数 [B店、C店]

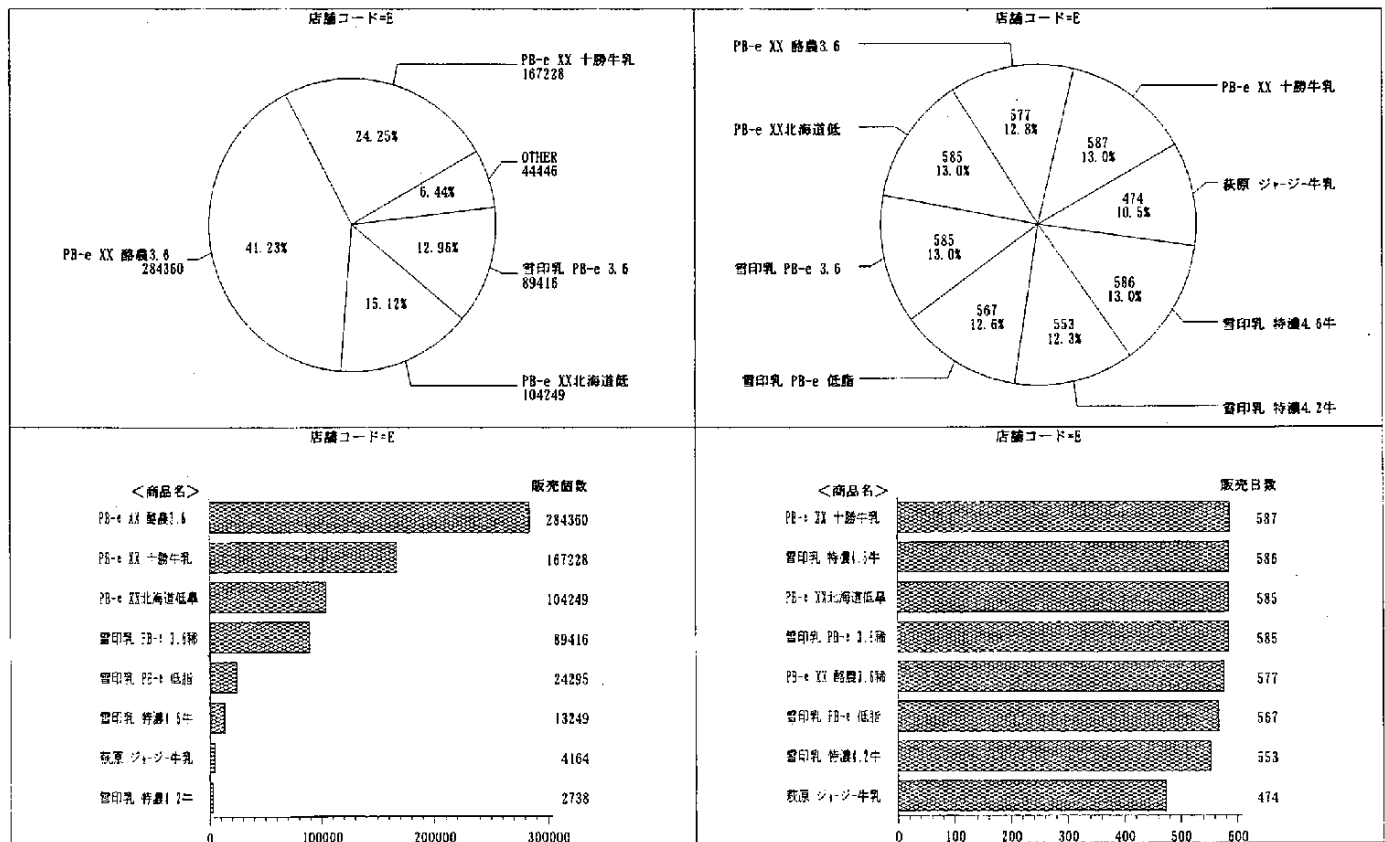
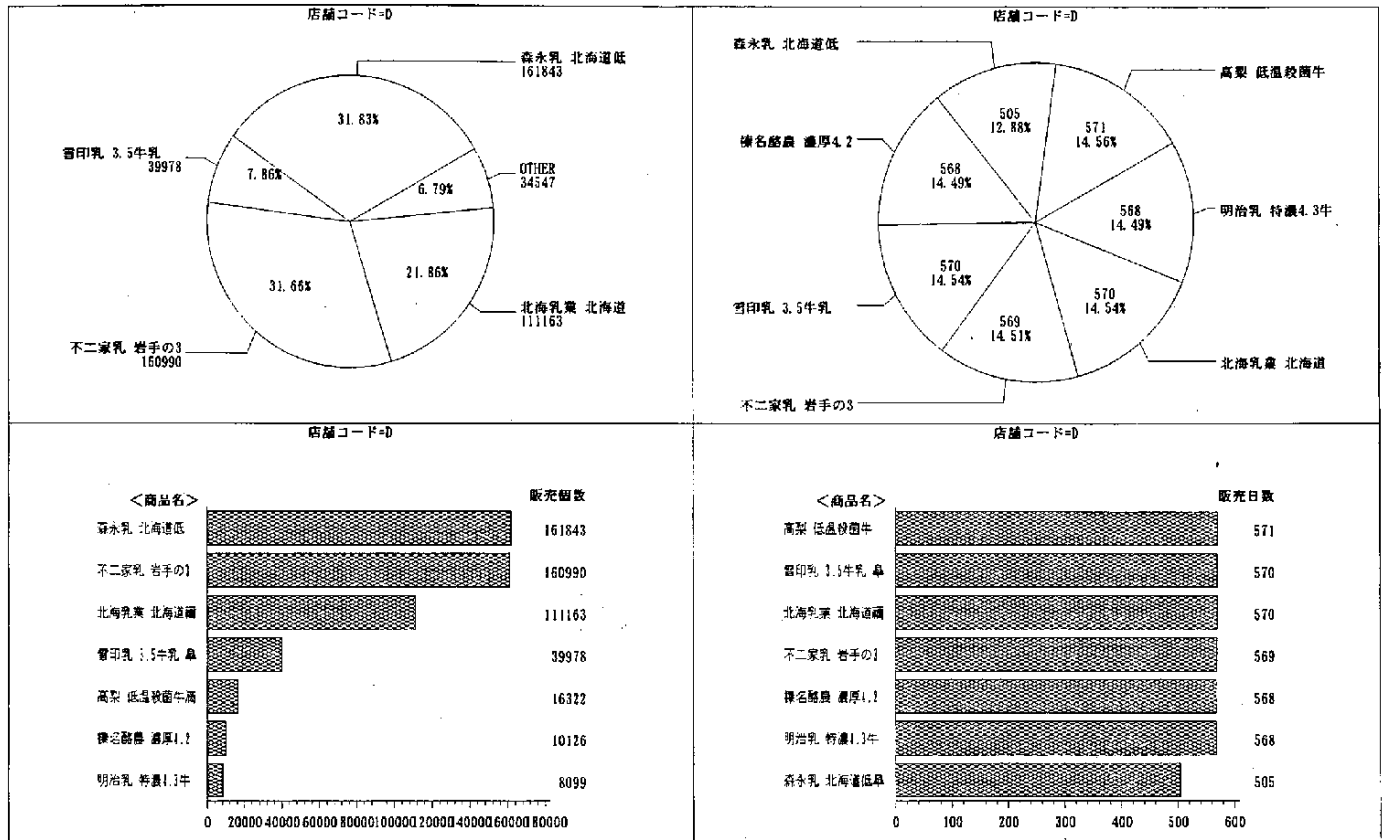


図6-5 対象20ヶ月間に継続販売された商品の販売個数シェアと販売日数 [D店、E店]

金額のどちらを予測するかという選択があり得る。図 6-6 に各店の販売個数と販売金額の関係を散布図で示した。B 店は価格戦略を 2 層に分割していると推察されるような分布を観察できるが、どの店も強い相関を示しているため、個数と金額のどちらを使用しても実質的には同じとみなしてよい。販売戦略上は個数が扱いやすいので、予測モデルの従属変数としては「販売個数」を採用した。

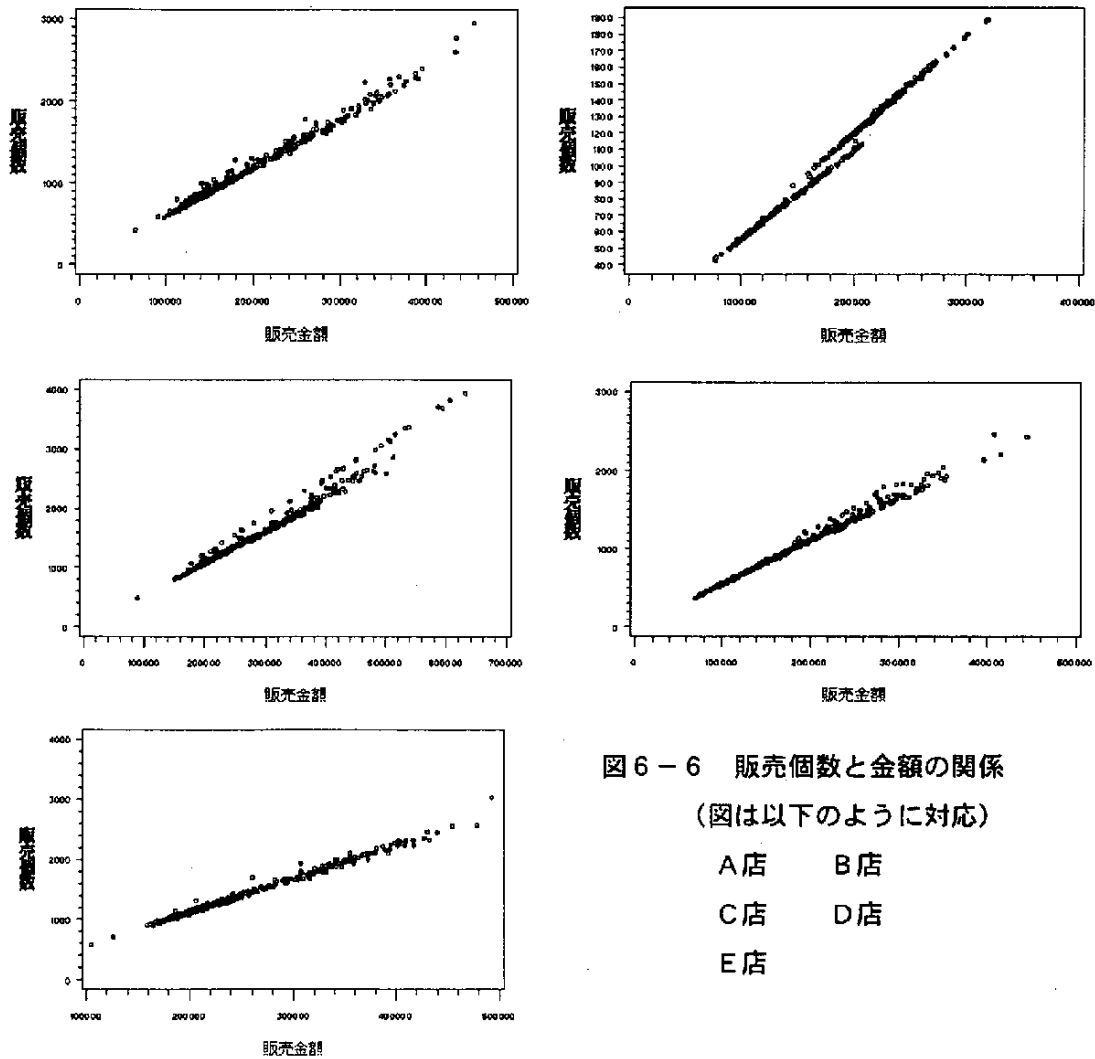


図 6-6 販売個数と金額の関係
 (図は以下のように対応)
 A店 B店
 C店 D店
 E店

6. 4. 3 販売個数の時系列変化

総個数の日次変動を折線グラフとヒストグラムで示した結果が図 6-7 (A~E) である。どの店の場合も販売個数には大きな変動がある。さらに飛び外れて大量に販売された数日を含んでいる。この傾向は商品別でも同じであり、特定の商品の問題ではない。図 6-8 (A~E) は商品別の日次変動を示す。

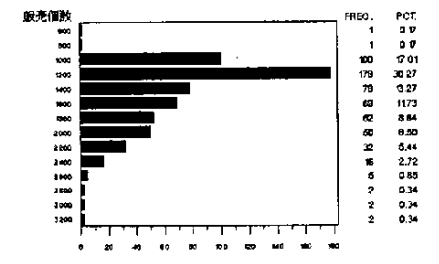
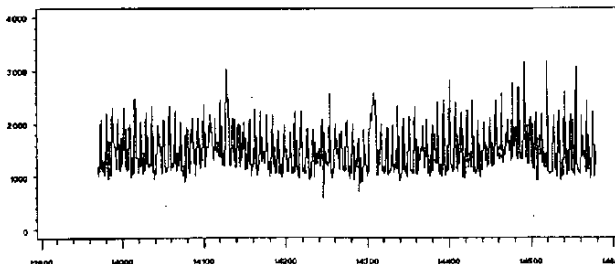
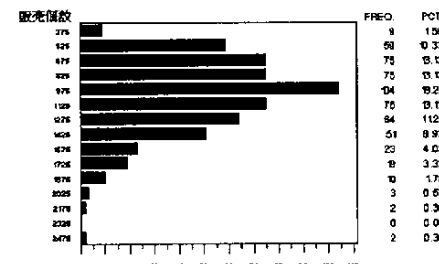
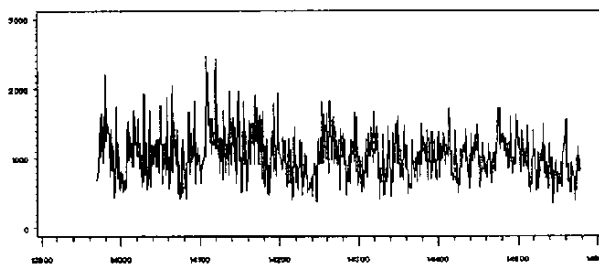
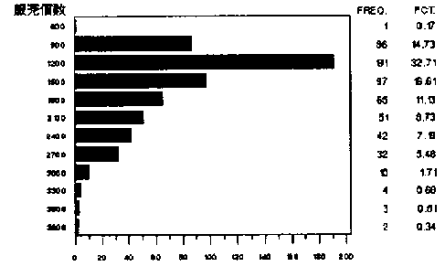
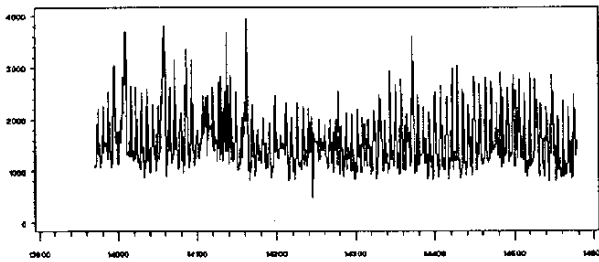
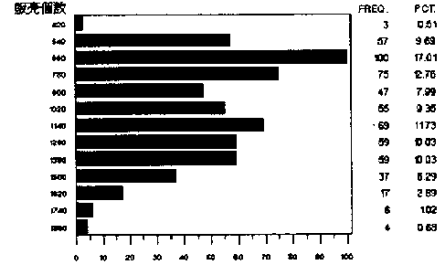
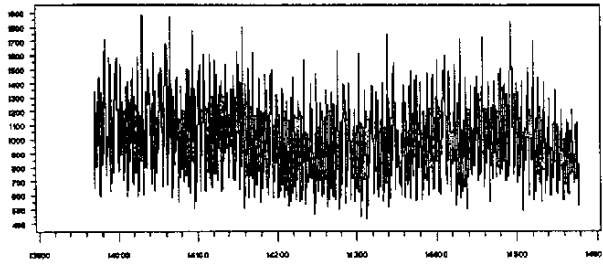
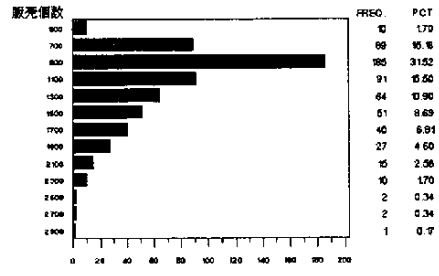
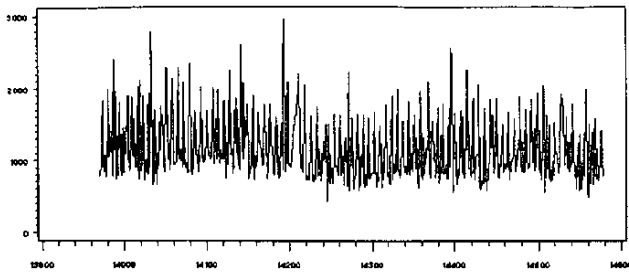


図6-7 販売総個数の日次系列変動 [上からA、B、C、D、E店]

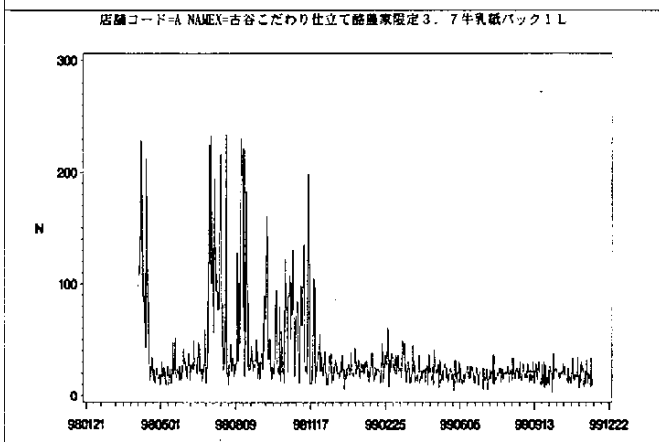
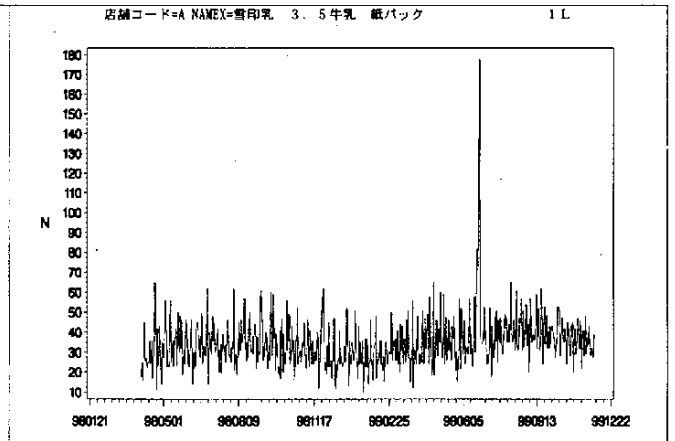
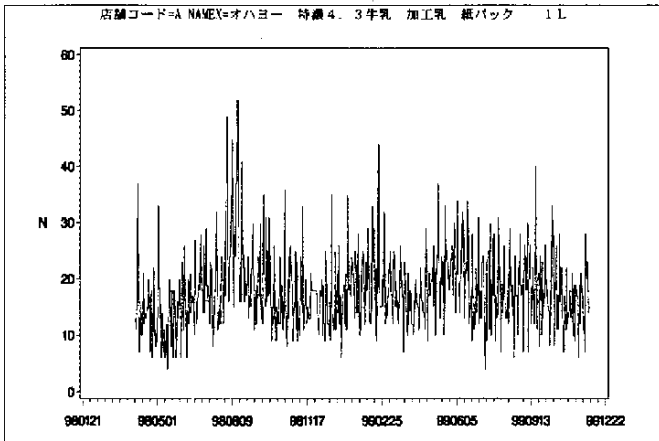
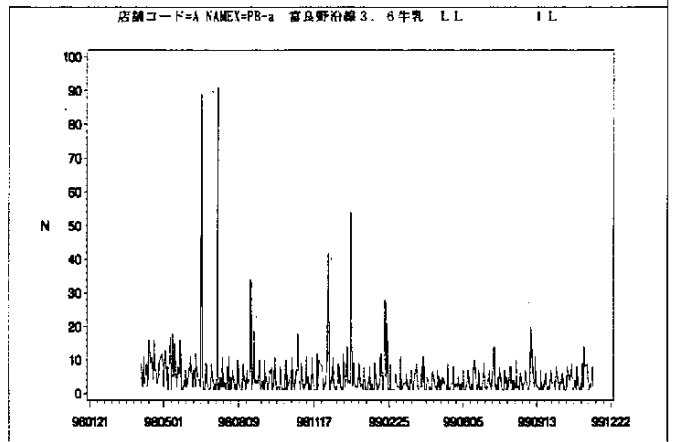
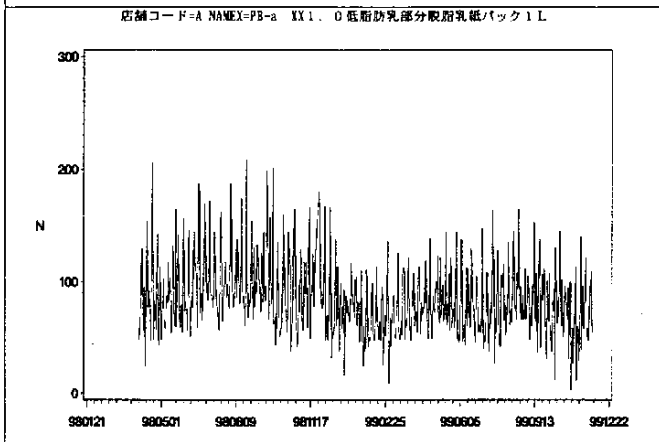
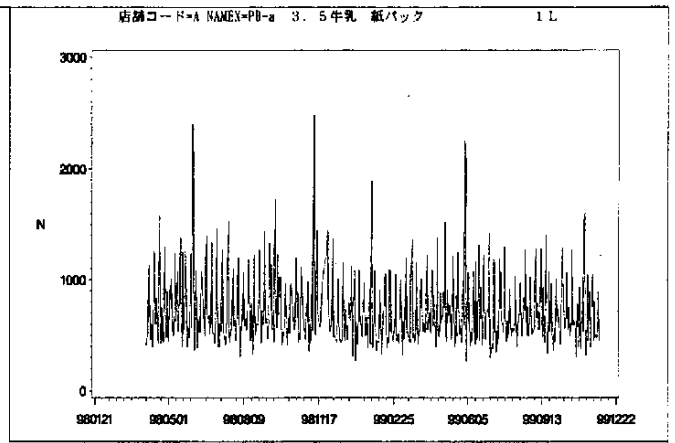
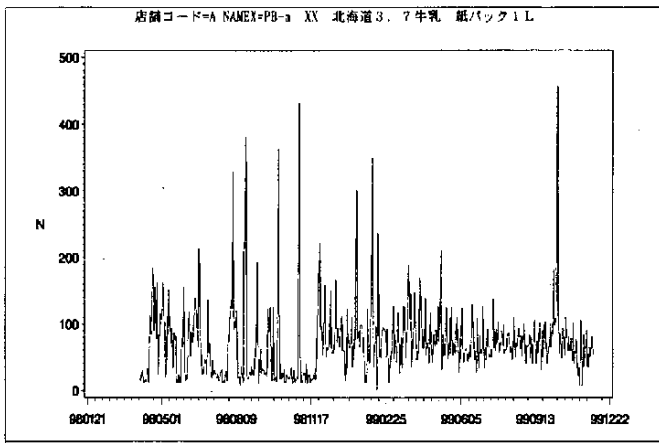


図6-8A 商品別の日次販売個数[A店]

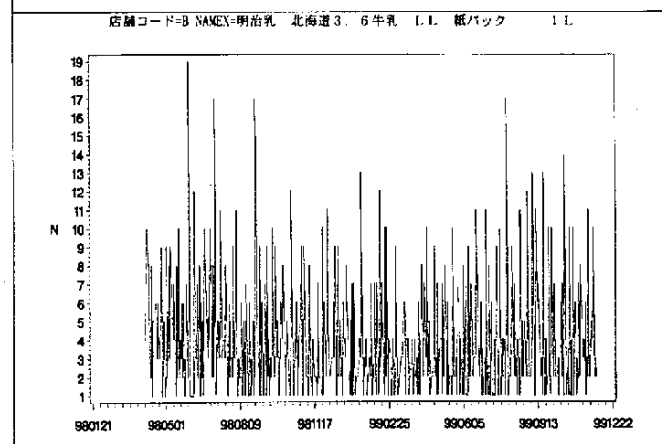
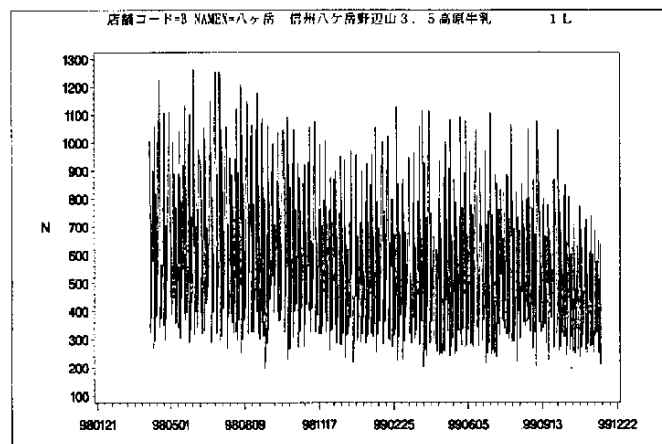
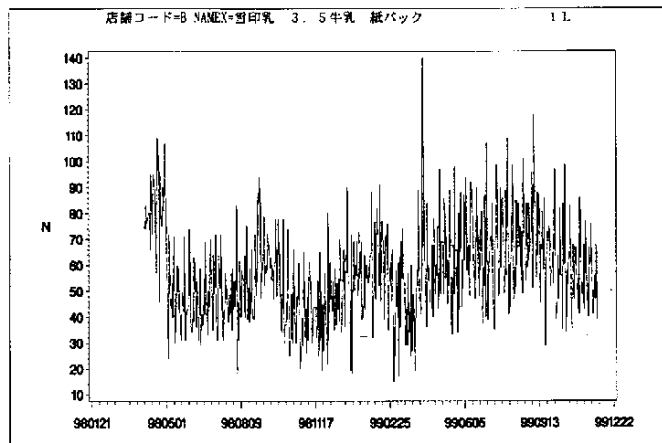
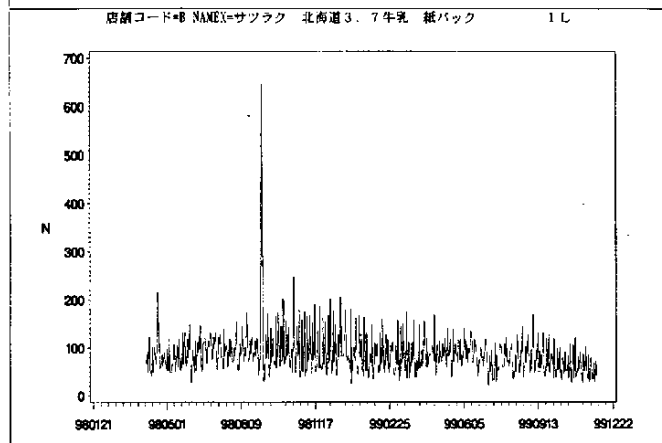
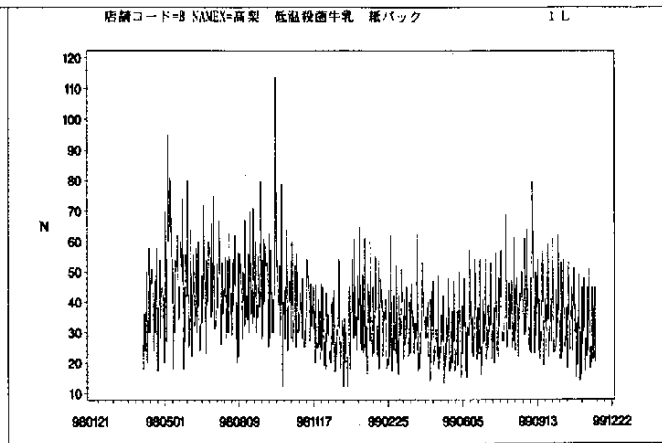
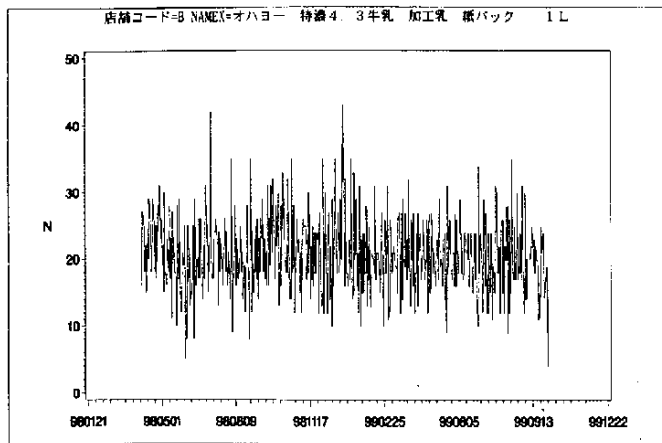


図6-8B 商品別の日次販売個数 [B店]

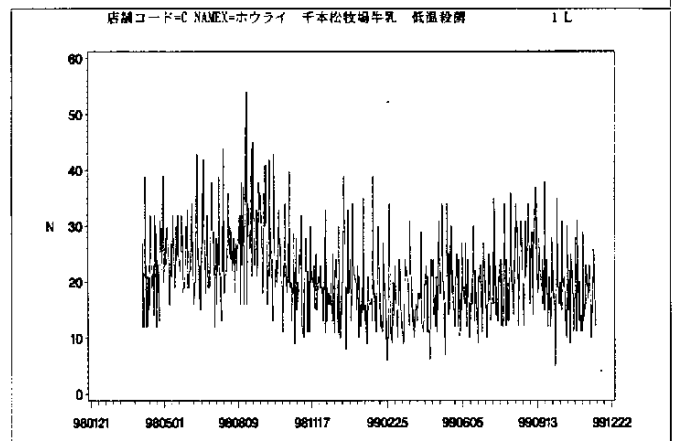
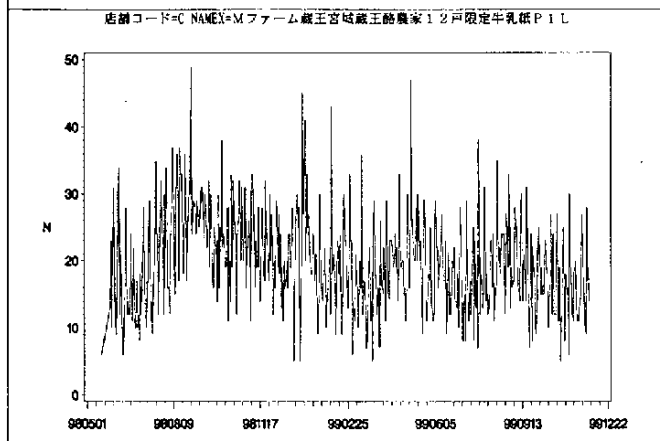
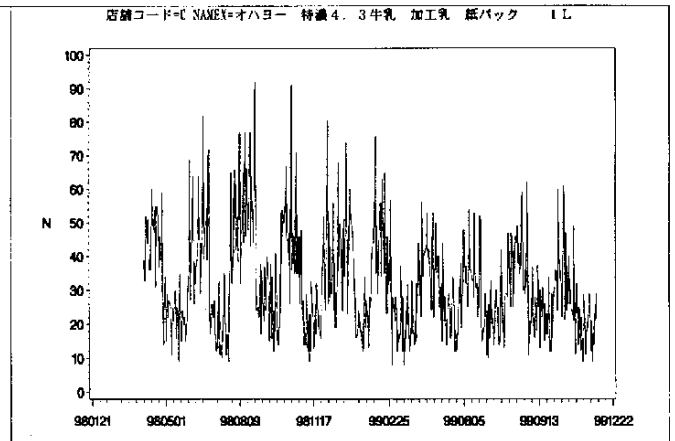
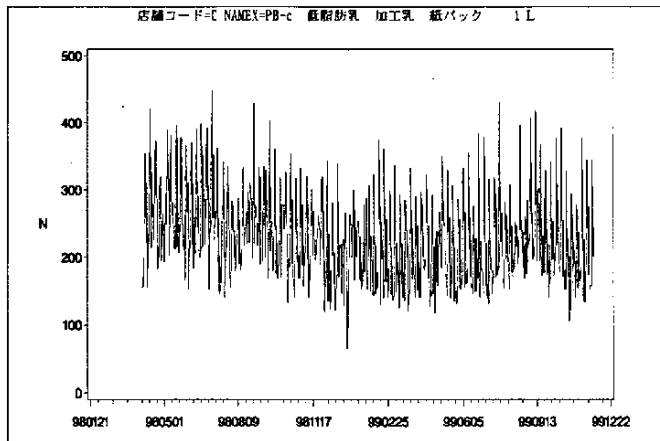
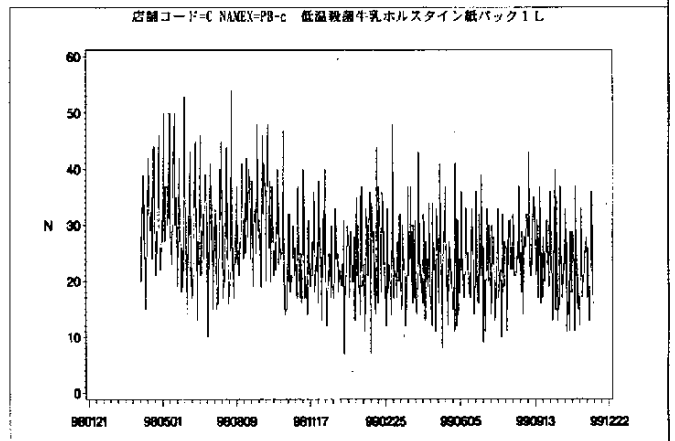
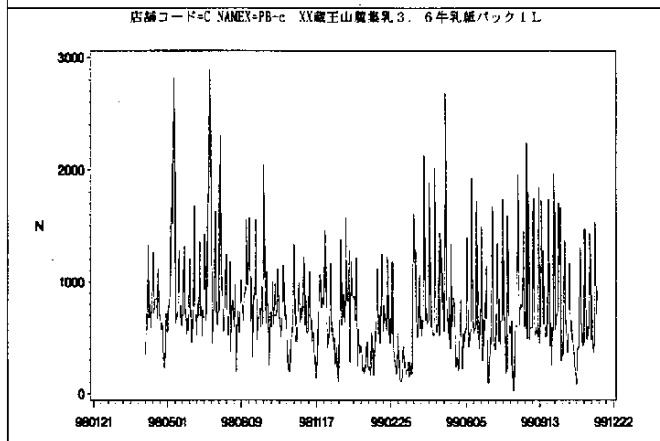
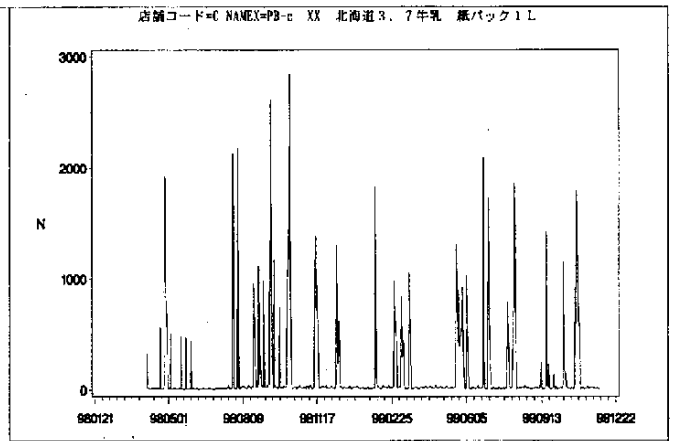
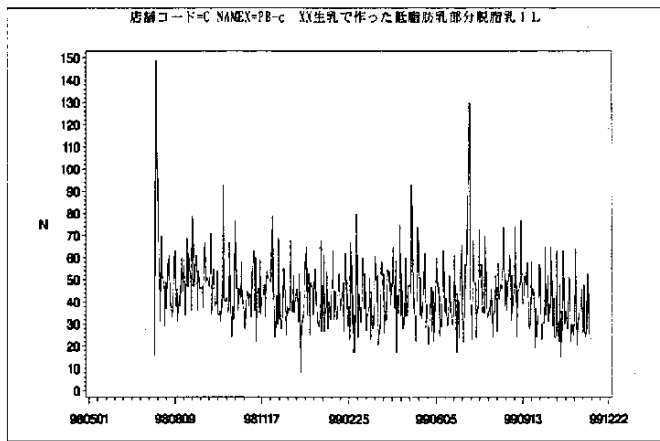


図6-8C 商品別の日次販売個数[C店]

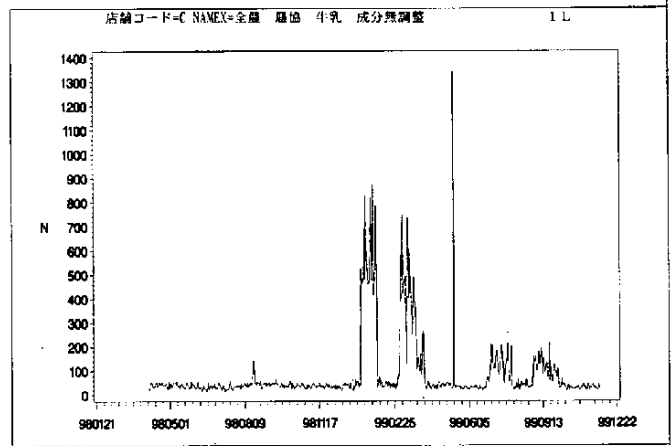
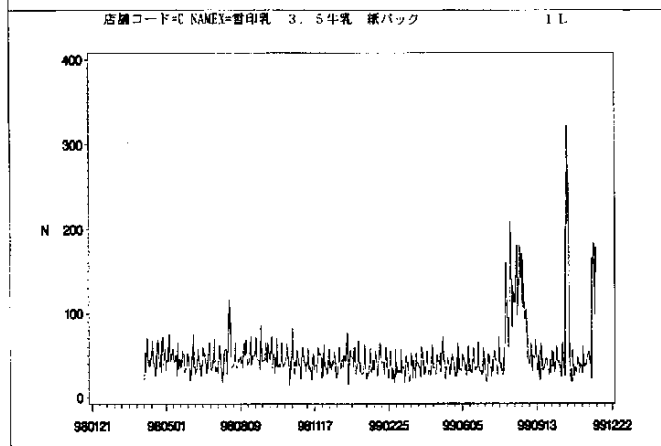
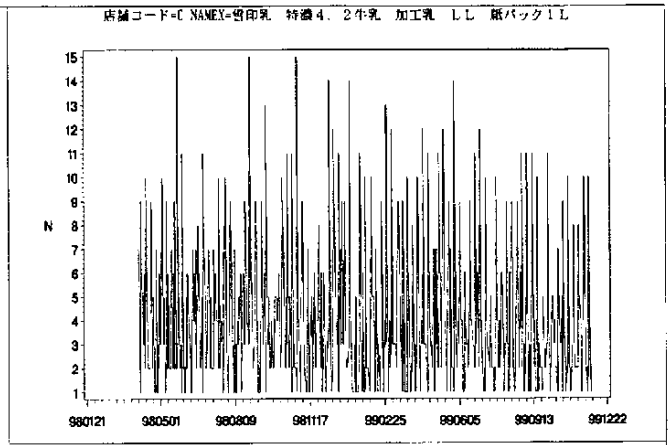
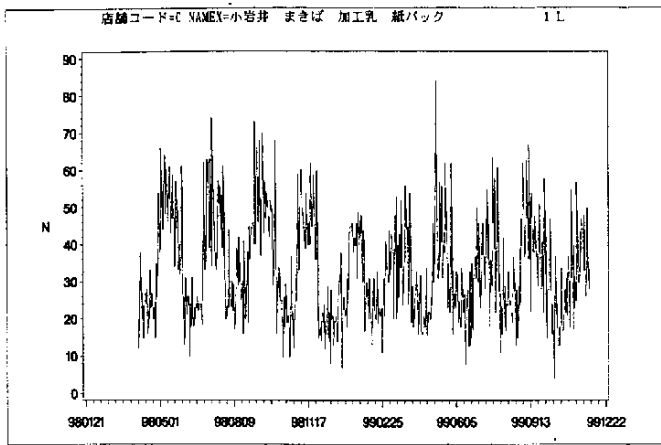


図6-8C 商品別の日次販売個数 [C店]

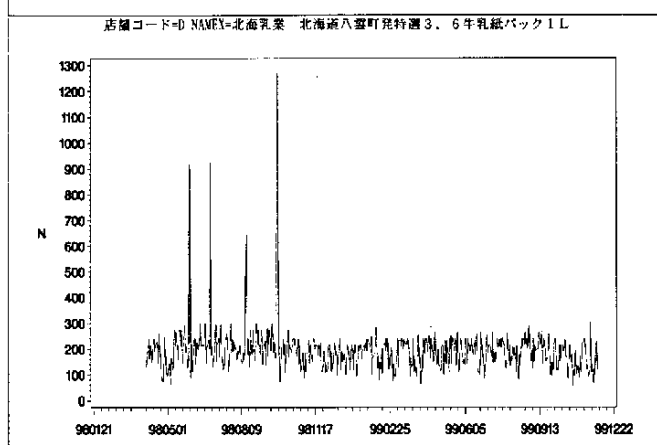
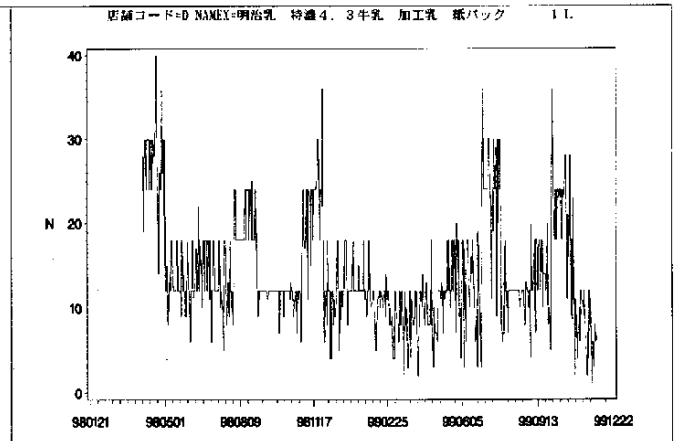
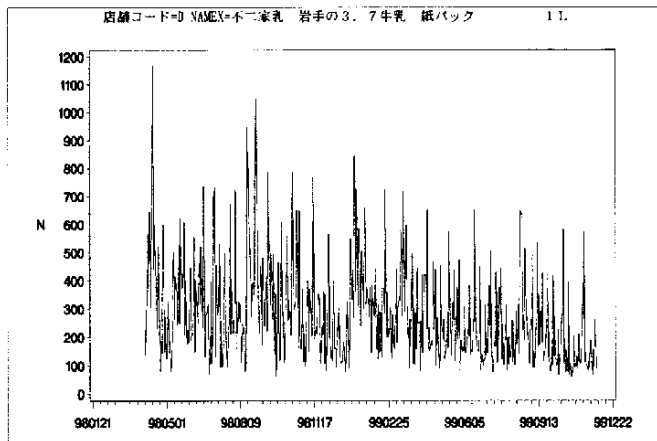
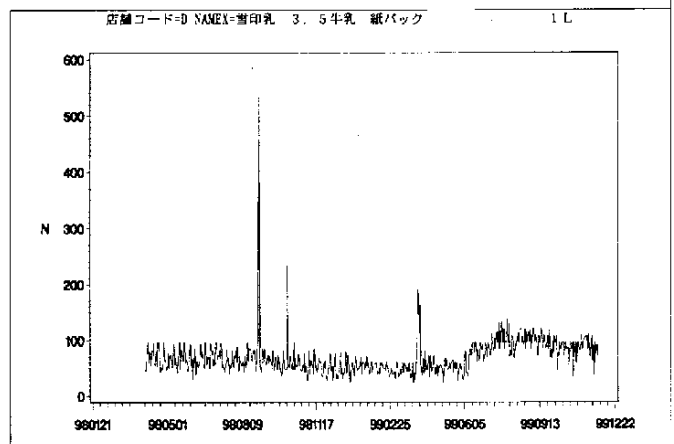
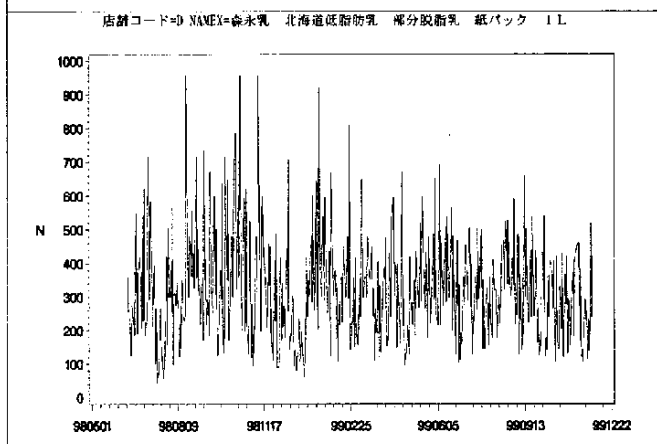
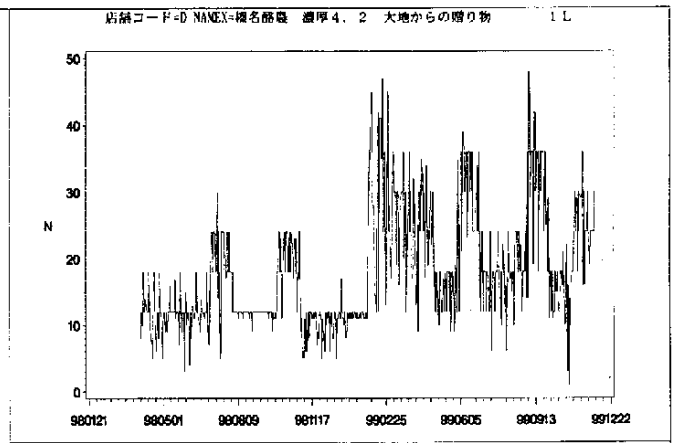
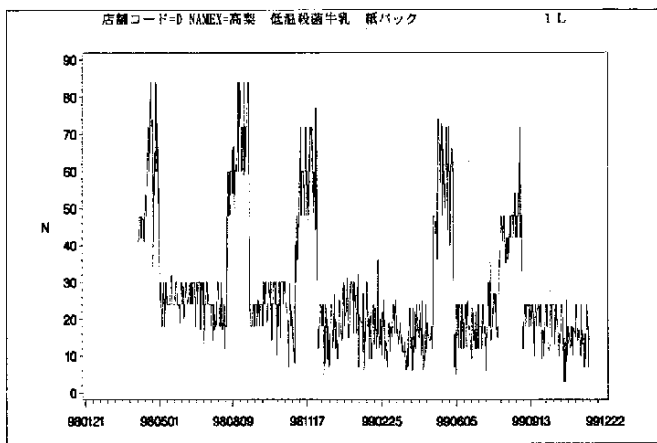


図6-8D 商品別の日次販売個数 [D店]

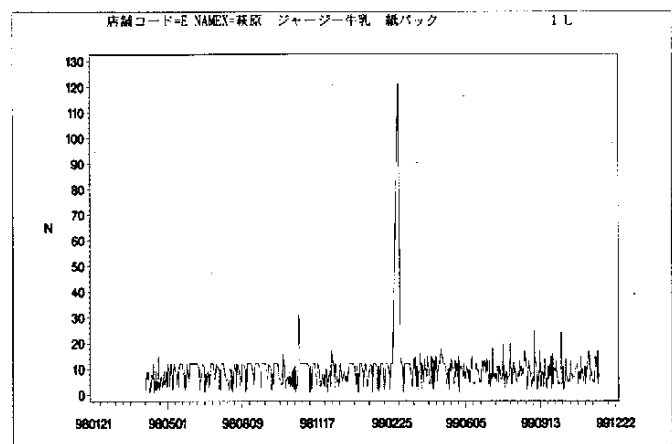
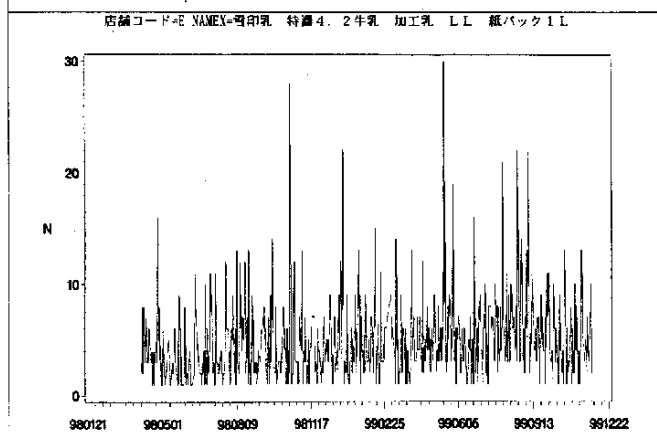
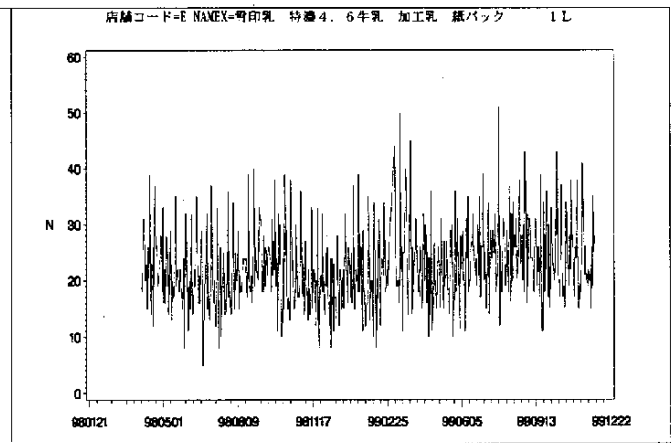
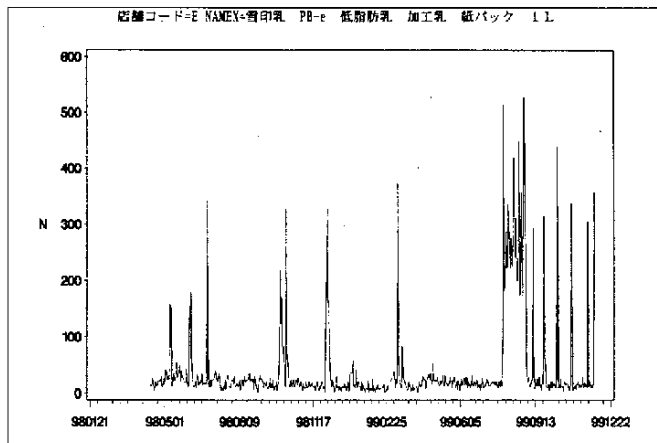
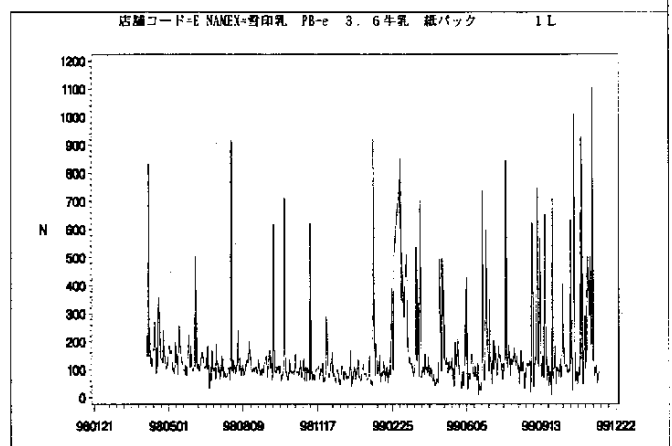
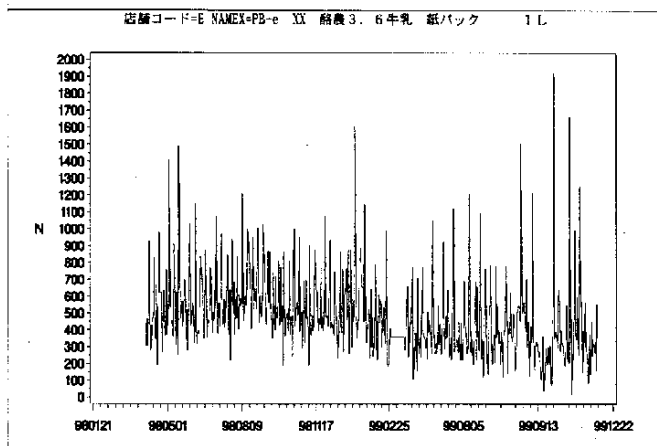
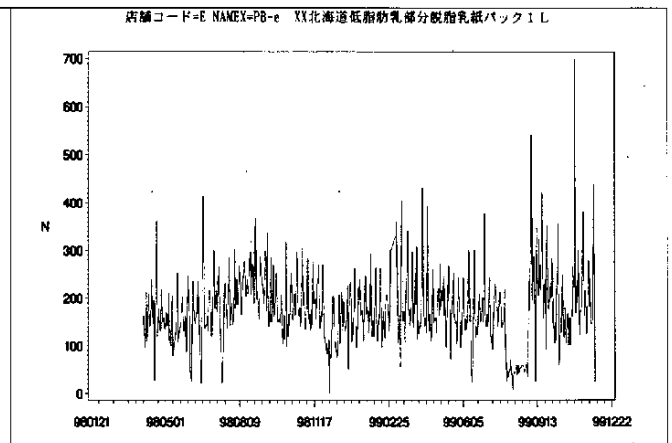
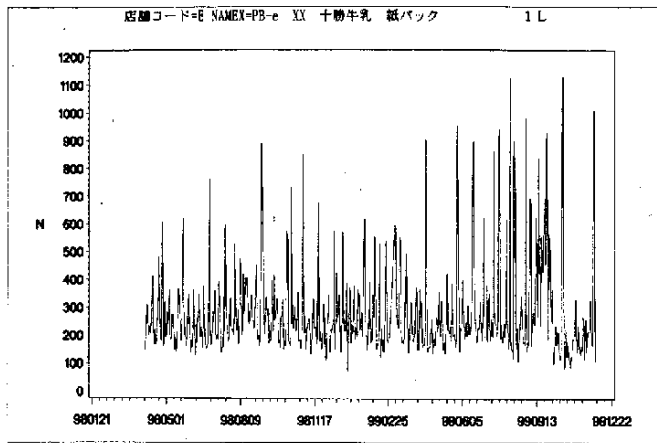


図6-8E 商品別の日次販売個数[E店]

6. 4. 4 来店客数の変動

販売個数の変動に影響する要因として、常識的に予想されるのが「来店客数」である。A店の販売個数（全種類合計）と来店客数の関係を図 6-9 に示す。明らかに強い相関があり、「客が増えれば牛乳が売れる」という単純な関係がある。

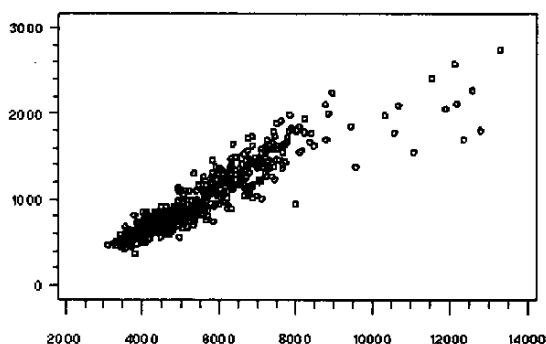


図 6 - 9 販売個数（全種類）と来店客数【A店】

しかしB店（図 6-10）では事情が異なる。同じ来店客数でも販売個数に大きな相違が観察される。さらに販売状況は2層構造を形成していることが分かる。これは図 6-6 で既に予想したようにB店が特殊な価格戦略を設定しているのではないかという背景と関係がありそうである。

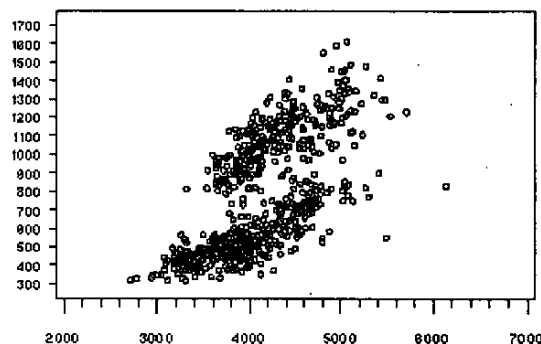


図 6 - 10 販売個数（全種類）と来店客数【B店】

またD店（図 6-11）の場合はA店のような強い相関はなく、来店客数が増えるに従って、販売個数の変動が大きくなっている。このような場合も単純に来店客数だけを要因として考えるのは不十分であることが示唆される。

来店客数と販売個数の関係が強くても、来店客数を販売個数の予測変数として使うことはできない。来店客数もまた販売個数と同じく予測したい「結果」であって原因変数ではないからである。来店客数を独立変数にしたところで、

見かけの適合はよくなっても、現実的には事前に来店客数が分からなければ予測モデルとしては一般的に役に立たないのである。

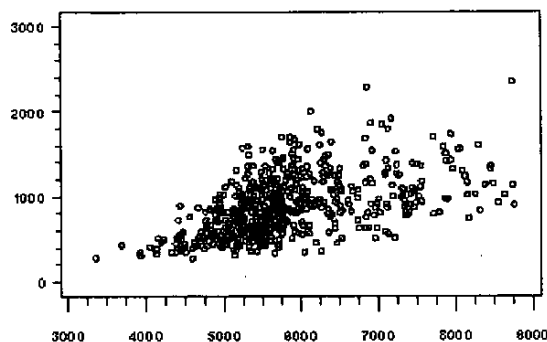


図6-11 販売個数(全種類)と来店客数[D店]

来店客数は少数の「飛び外れ値」を持っており、一般的な分析では除外される。A店の平均的な来店者数は5000人をやや上回る程度だが、最高では13281人の日がある。このような外れ値は異常値や誤集計ではなく、1年に何回かはそのような日が存在するのである。ここでは除外するかわりに「特異日フラグ」という新しい変数を作成して、特異日を説明するためのダミー変数を作成した。

飛びはずれ値の目安は箱ヒゲ図(図6-13)を作成して、四分位数の1.5倍を境界とした。その結果、表6-4のような特異日が特定された。大晦日もあるが、特定の曜日に決まっていたり人為的な操作をした戦略日であると想像できる。A店の場合、日次の来店者数のグラフに、境界線を引くと図6-12のようになる。

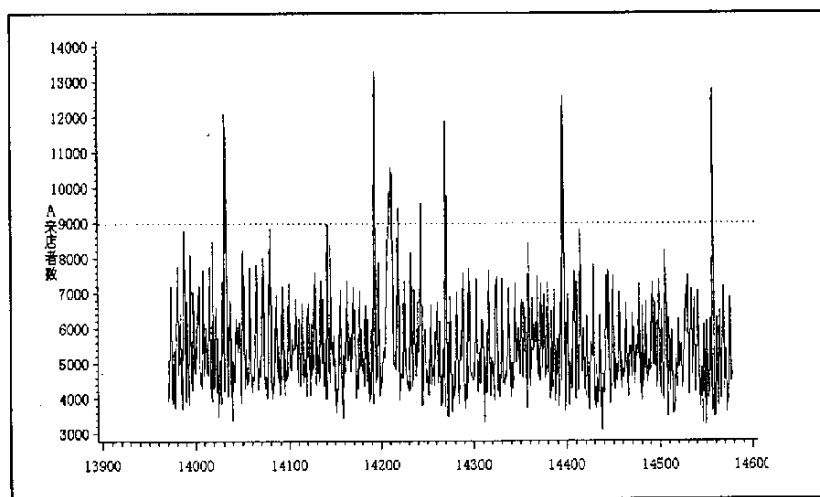


図6-12 A店の来店者・時系列の飛び外れ値境界線

図6.14は来店客数と販売個数の商品層別散布図。図6.15は商品種類ごとの散布図で、それぞれ回帰直線と95%信頼区間も表示してある。

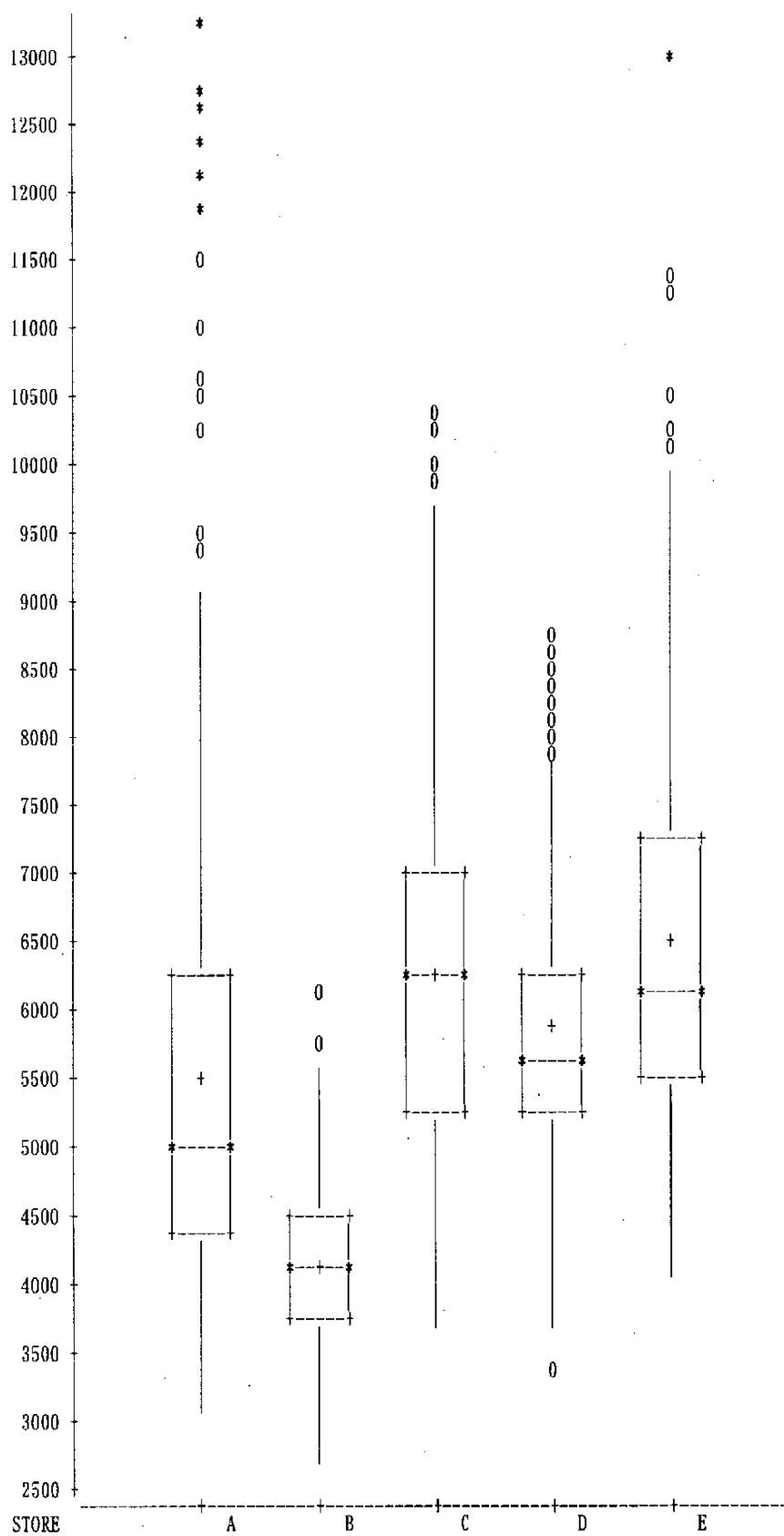


図 6 - 1 3 来店者数の店舗別箱ヒゲ図

表6-4 来店客数の「特異日」(特に多い日)

店舗コード=A						
OBS	DATE	WEEK	WEEKDAY	N_STORE	TENKI	OUTLIER
1	1998-06-02	火	3	12095	くもり	*
2	1998-06-03	水	4	10661	雨	*
3	1998-11-10	火	3	12165	くもり	*
4	1998-11-11	水	4	13281	晴れ	*
5	1998-11-28	土	7	10554	晴れ	*
6	1998-11-29	日	1	10302	晴れ	*
7	1998-12-06	日	1	9420	晴れ	*
8	1998-12-31	木	5	9539	晴れ	*
9	1999-01-26	火	3	11051	晴れ	*
10	1999-01-27	水	4	11873	くもり	*
11	1999-06-01	tue	3	11524	晴れ	*
12	1999-06-02	wed	4	12574	晴れ	*
13	1999-11-09	tue	3	12338	晴れ	*
14	1999-11-10	wed	4	12774	晴れ	*

店舗コード=B						
OBS	DATE	WEEK	WEEKDAY	N_STORE	TENKI	OUTLIER
15	1998-12-31	木	5	6140	晴れ	*
16	1999-08-25	wed	4	5711	晴れ	*

店舗コード=C						
OBS	DATE	WEEK	WEEKDAY	N_STORE	TENKI	OUTLIER
17	1998-05-10	日	1	9894	くもり	*
18	1998-06-28	日	1	9386	くもり	*
19	1998-10-08	木	5	10007	晴れ	*
20	1998-10-11	日	1	10397	晴れ	*
21	1998-10-12	月	2	9172	くもり	*
22	1998-12-31	木	5	10257	晴れ	*
23	1999-01-27	水	4	9890	くもり	*
24	1999-05-08	sat	7	9581	くもり	*
25	1999-07-24	sat	7	9350	晴れ	*
26	1999-08-26	thu	5	9143	くもり	*
27	1999-10-09	sat	7	9362	晴れ	*
28	1999-11-27	sat	7	9086	晴れ	*

店舗コード=D						
OBS	DATE	WEEK	WEEKDAY	N_STORE	TENKI	OUTLIER
29	1998-05-19	火	3	8269	くもり	*
30	1998-06-09	火	3	8176	くもり	*
31	1998-06-16	火	3	8497	くもり	*
32	1998-08-18	火	3	8733	くもり	*
33	1998-08-25	火	3	7939	晴れ	*
34	1998-09-08	火	3	8034	くもり	*
35	1998-10-20	火	3	8297	くもり	*
36	1998-10-27	火	3	7846	くもり	*
37	1999-01-19	火	3	7943	くもり	*
38	1999-01-26	火	3	7964	晴れ	*
39	1999-02-02	火	3	8454	くもり	*
40	1999-02-16	火	3	8153	晴れ	*
41	1999-02-23	火	3	7875	晴れ	*
42	1999-03-16	火	3	8063	晴れ	*
43	1999-04-13	tue	3	8459	晴れ	*
44	1999-04-20	tue	3	8554	晴れ	*
45	1999-04-25	sun	1	8092	晴れ	*
46	1999-05-11	tue	3	7900	くもり	*
47	1999-05-18	tue	3	8741	くもり	*
48	1999-05-25	tue	3	8378	晴れ	*
49	1999-06-08	tue	3	8164	晴れ	*
50	1999-06-15	tue	3	8637	くもり	*
51	1999-07-06	tue	3	7872	くもり	*
52	1999-09-07	tue	3	8317	くもり	*
53	1999-11-09	tue	3	8766	晴れ	*
54	1999-11-12	fri	6	8125	くもり	*
55	1999-11-13	sat	7	7879	晴れ	*

店舗コード=E						
OBS	DATE	WEEK	WEEKDAY	N_STORE	TENKI	OUTLIER
56	1998-12-31	木	5	11316	晴れ	*
57	1999-03-06	土	7	13055	晴れ	*
58	1999-03-07	日	1	11291	雨	*
59	1999-06-06	sun	1	10194	晴れ	*
60	1999-07-04	sun	1	10294	晴れ	*
61	1999-09-05	sun	1	10519	晴れ	*
62	1999-10-03	sun	1	10163	くもり	*

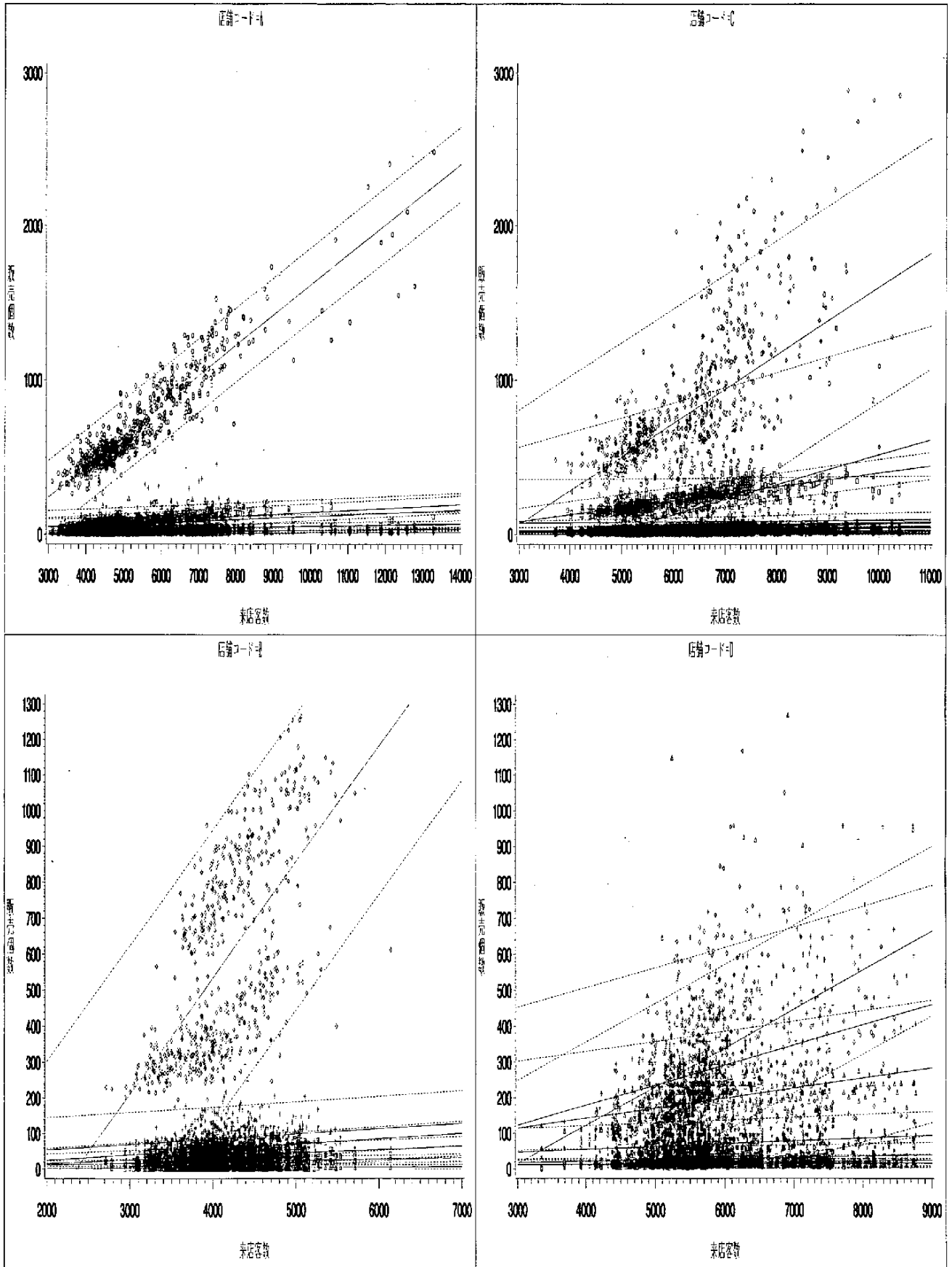
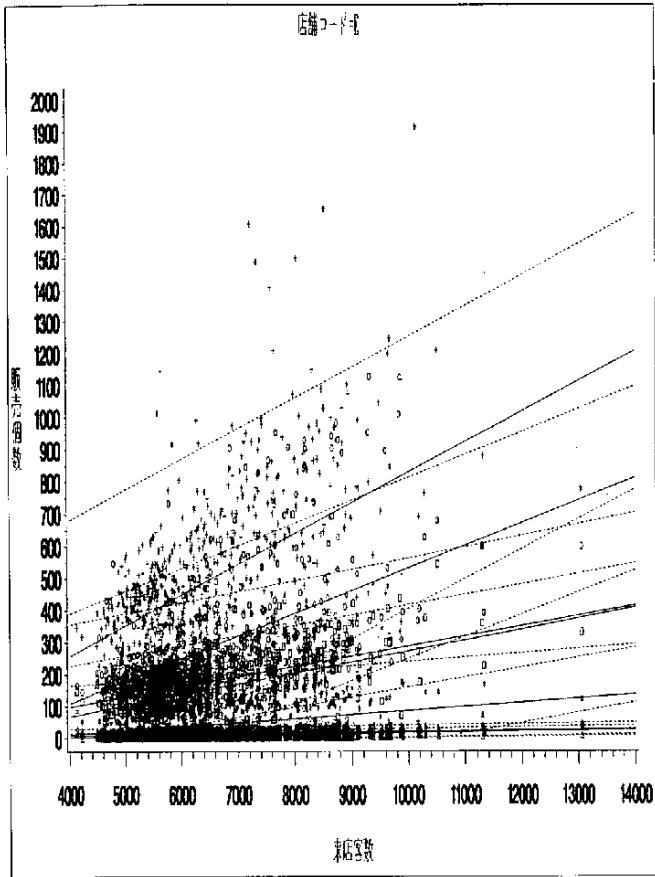


図6-14 来店客数と販売個数の商品層別散布図



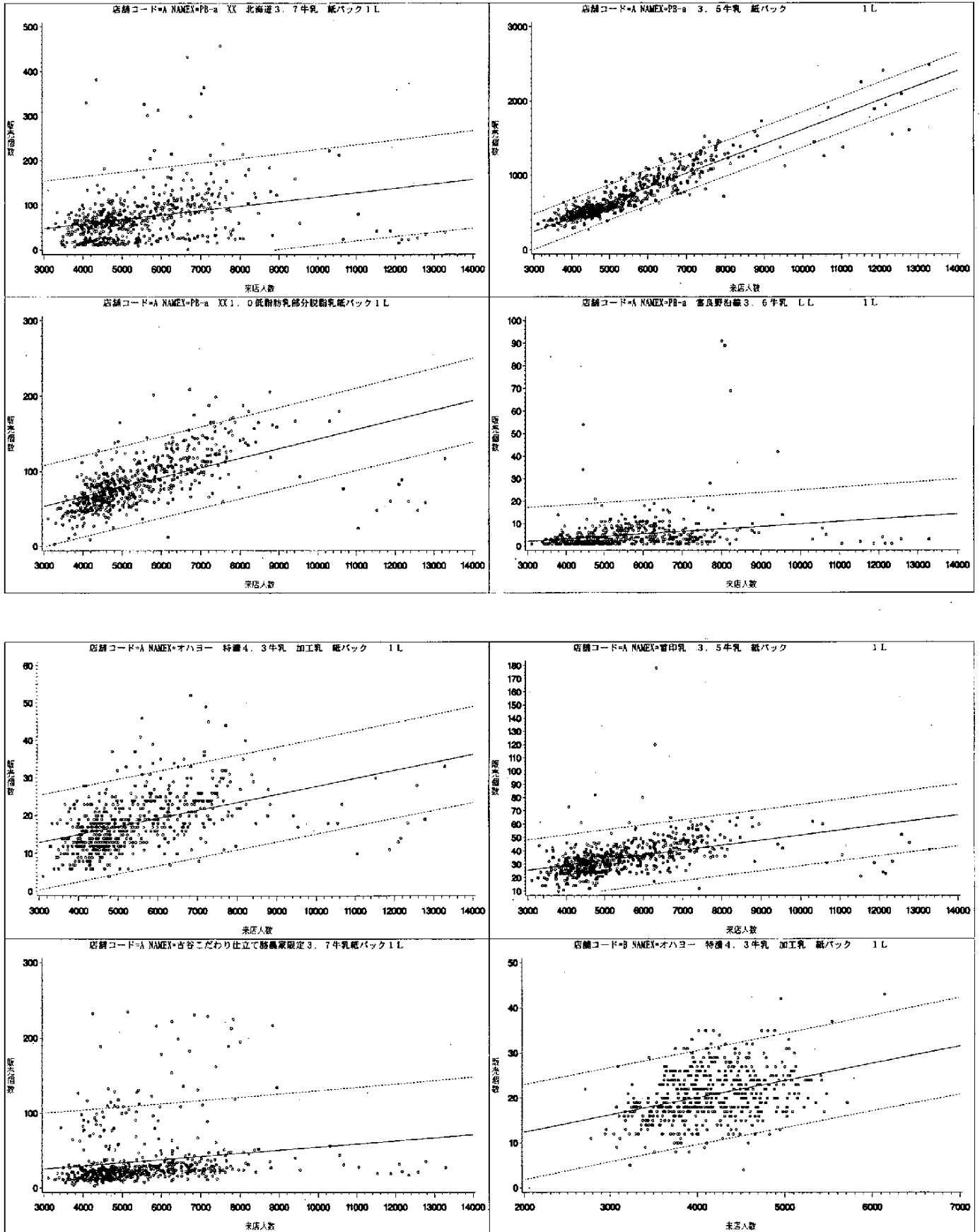


図6-15 来店客数と販売個数の各商品ごとの散布図

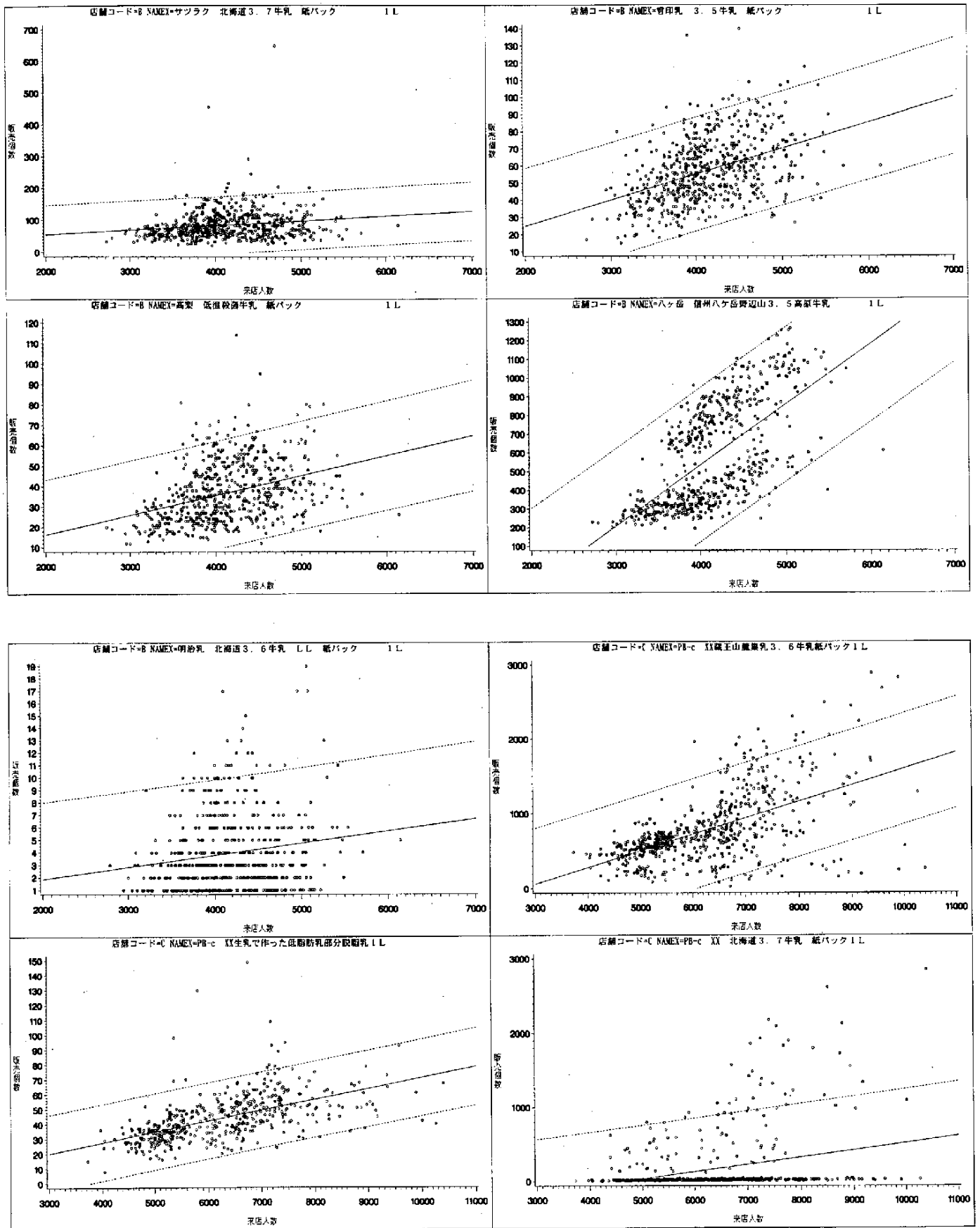


図6-15 来店客数と販売個数の各商品ごとの散布図

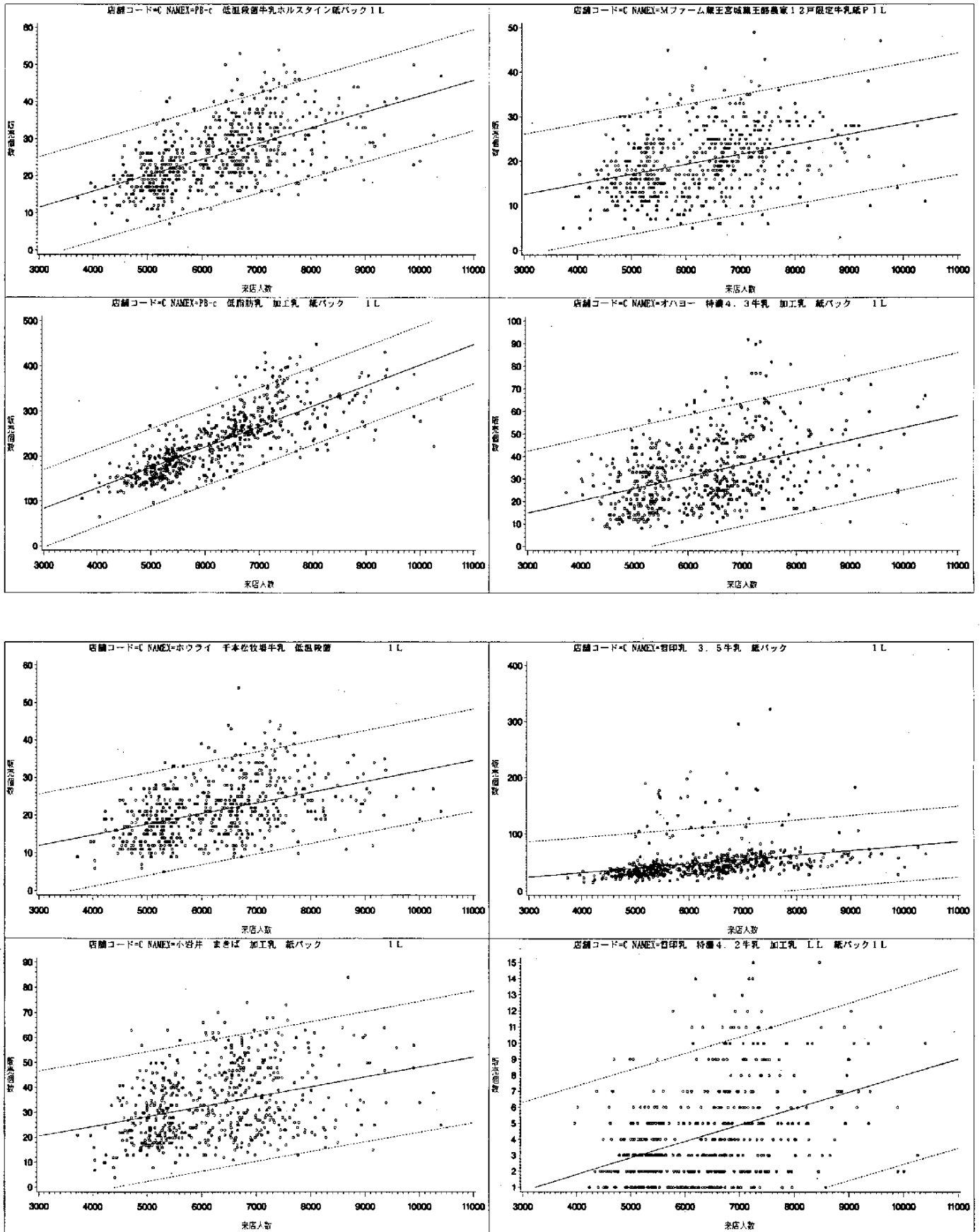


図6-15 来店客数と販売個数の各商品ごとの散布図

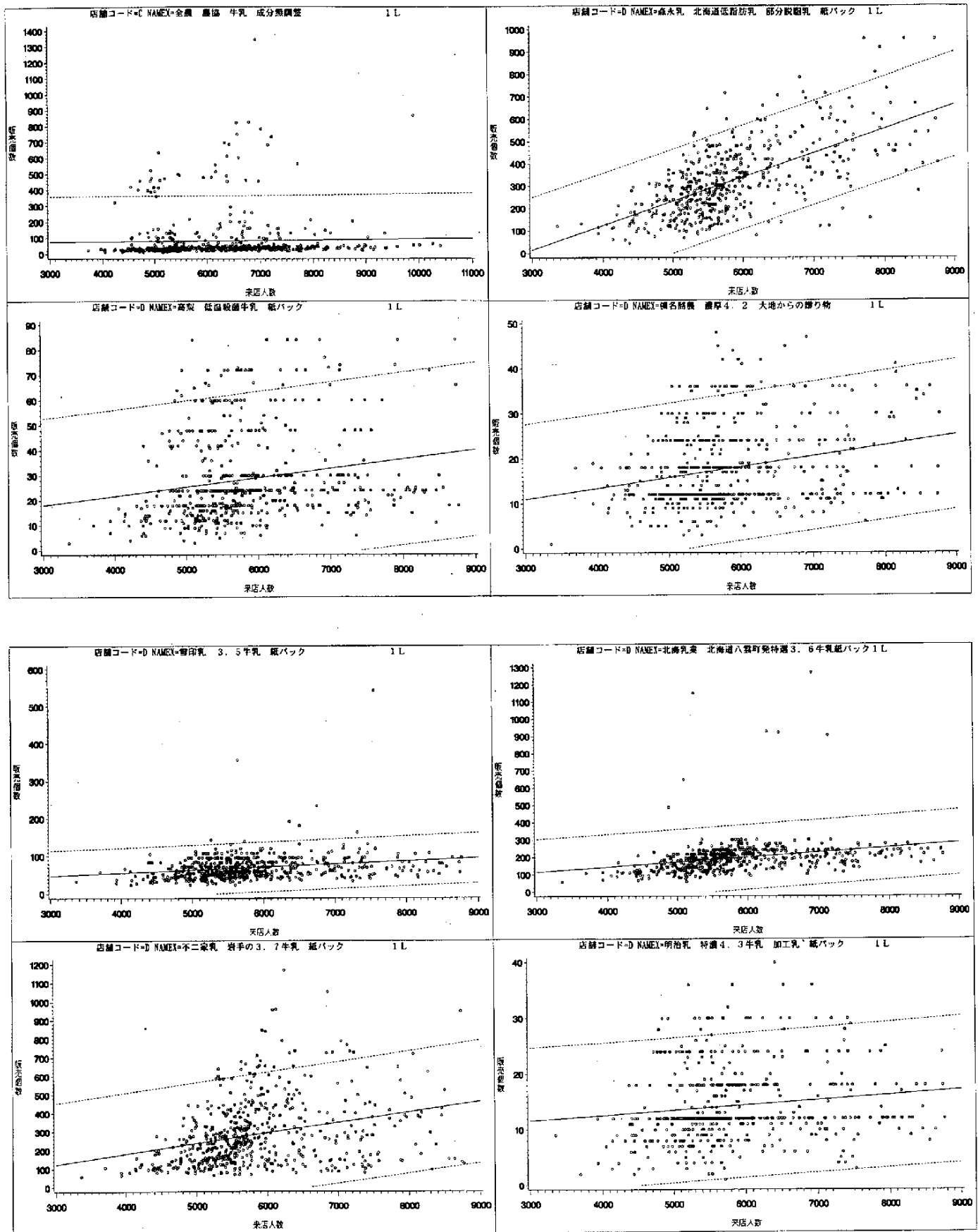


図6-15 来店客数と販売個数の各商品ごとの散布図

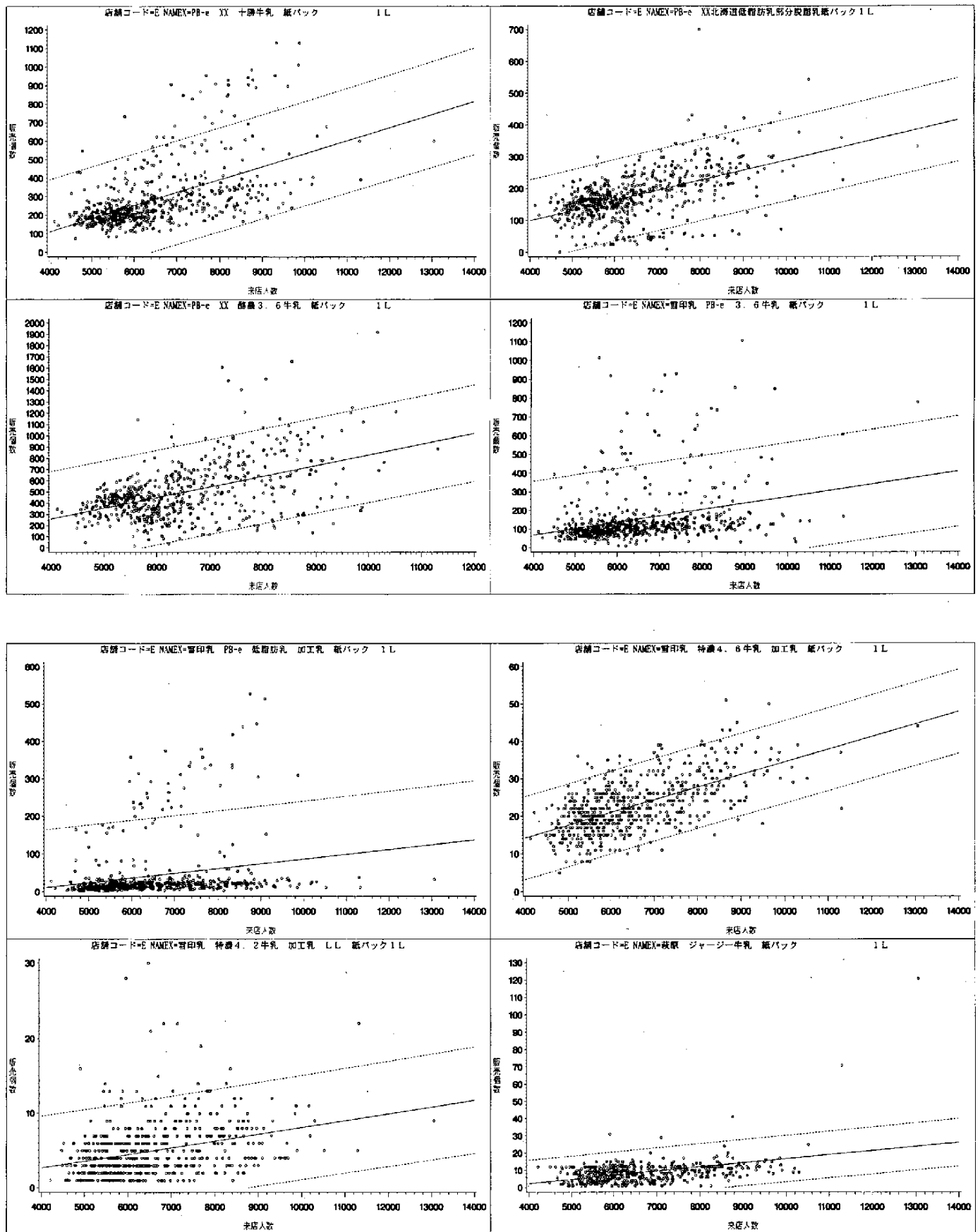


図6-15 来店客数と販売個数の各商品ごとの散布図

6. 4. 5. 販売価格の変動

販売個数に影響する変数として経験的に知られているのが価格である。図 6-16 は日次で販売価格がどのように変化（値引き）するかを示した時系列グラフで、店別・商品種類別に作成した。店によっても商品によっても価格設定が人為的に操作されていることが分かる。

6. 5 マイニングツールによるモデル構成

予備解析で検討した結果から予測変数を以下のように準備した。

- 気象関係の変数
- 特異日のダミー変数
- 休店前後を示すダミー変数
- 商品別の価格
- 月、曜日を示す変数

分析データの形式は、1日を1行とする形に変換した。すなわち、商品の価格を変数として扱った。このため営業日に販売されなかった商品は価格が欠損値になるので最高価格を代入した。通常の解析では欠損値に平均値を代入することが多いが、この場合は「値引きなし」という状態を示すために最高値とした。店頭の商品がないという状態なので、消費者にとっては選択肢として与えられていない状態に近くするという考え方である。EM では欠損値処理のオプションとして「最大値」を代入するオプションがないので、事前の DATA ステップで処理した。

EM によるモデル構成は A 店の場合を取り上げる。EM は図 6-17 のような GUI（グラフィカル・ユーザー・インターフェース）を装備しており、処理の手順（ワークフロー）を示し、パスに沿ってデータが流れて行くイメージで考えれば良い。分析者はプログラミングしないで、マウスで作業手順を作成すれば良い。この GUI の考え方は伝統的な構造的プログラミングをアイコンで抽象化したもので、伝統的思想が本質的には保存されている。たとえばダイクストラはプログラミングを「真珠の首飾り」のように作り上げるべきであり、決してスパゲッティのように（グチャグチャに）書くべきでないと言ったのだが、「真珠の首飾り」が「ノードのパスダイアグラム」として抽象化されたといえる。

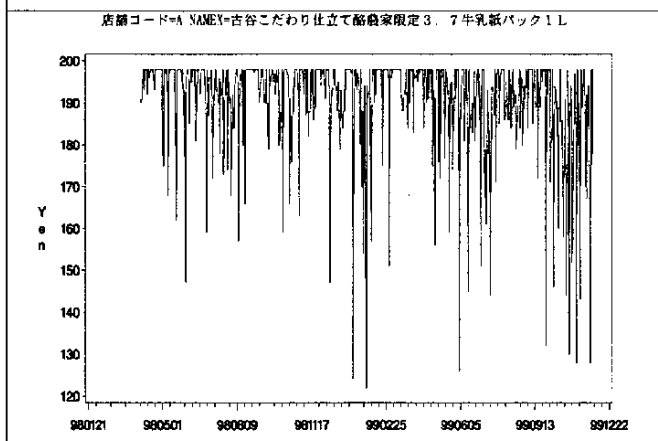
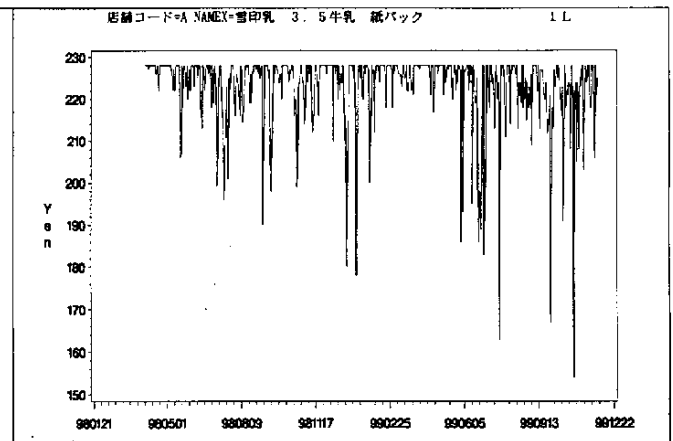
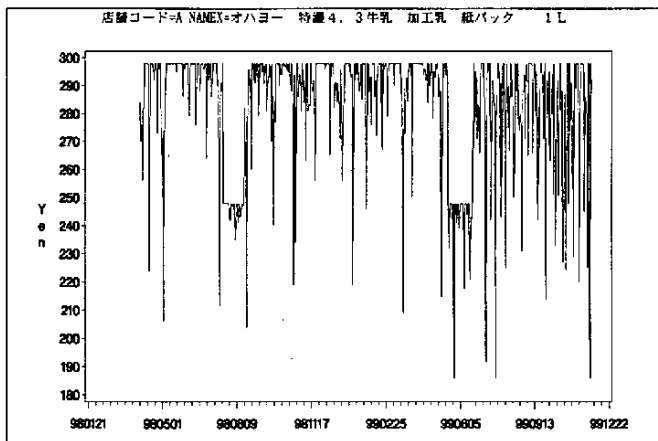
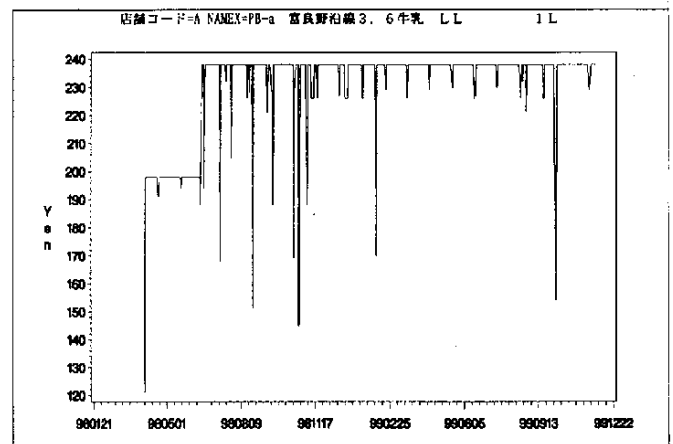
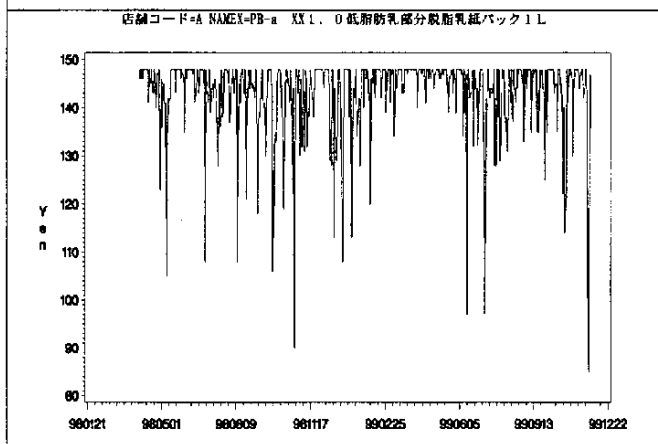
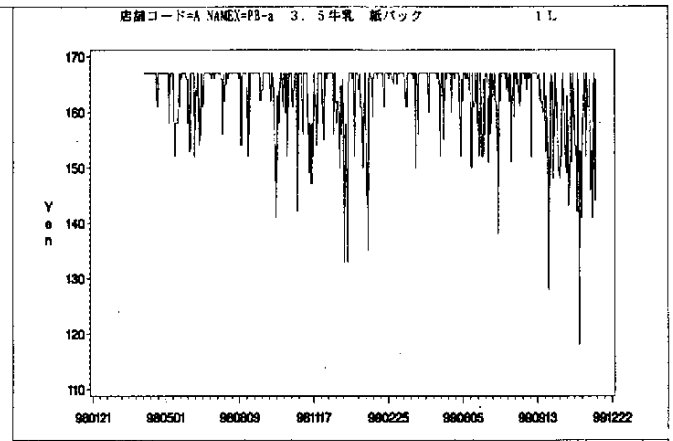
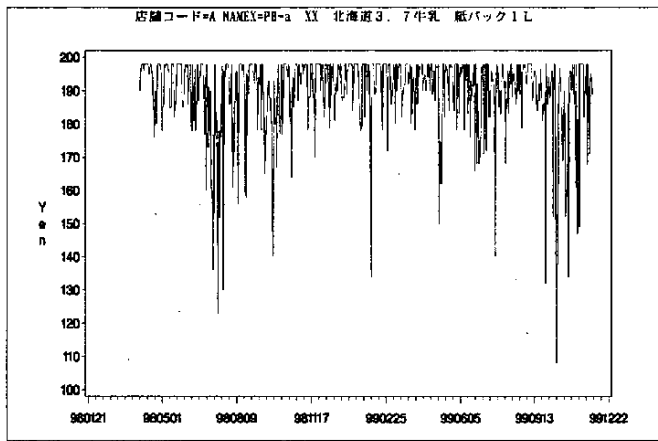


図6-16 販売価格の日次変化(商品別)[A店]

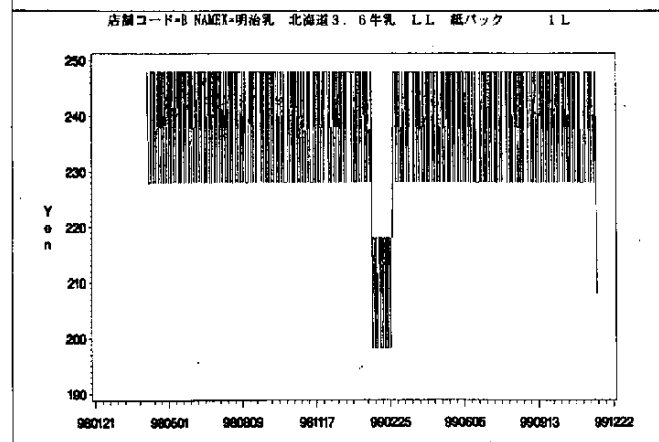
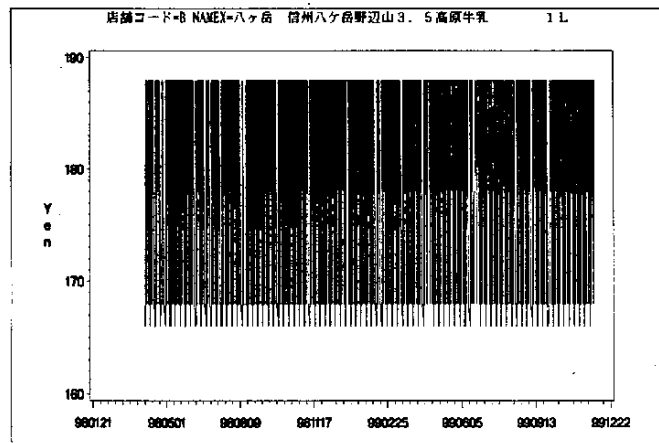
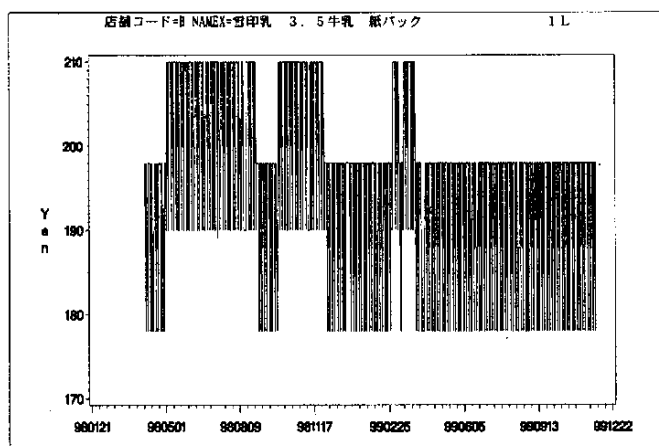
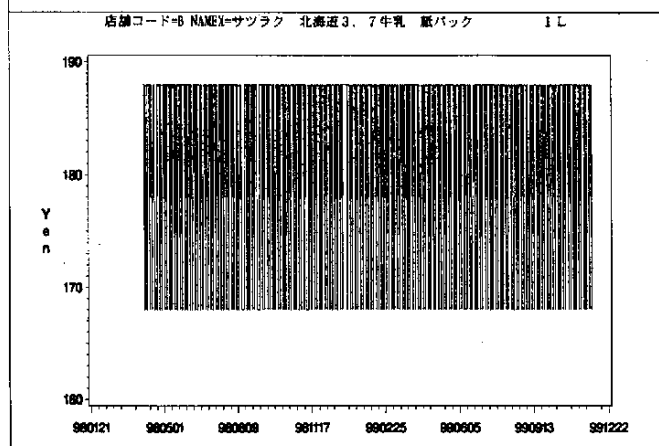
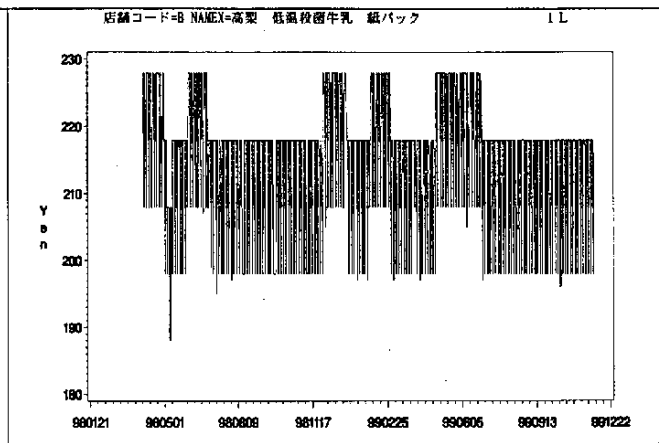
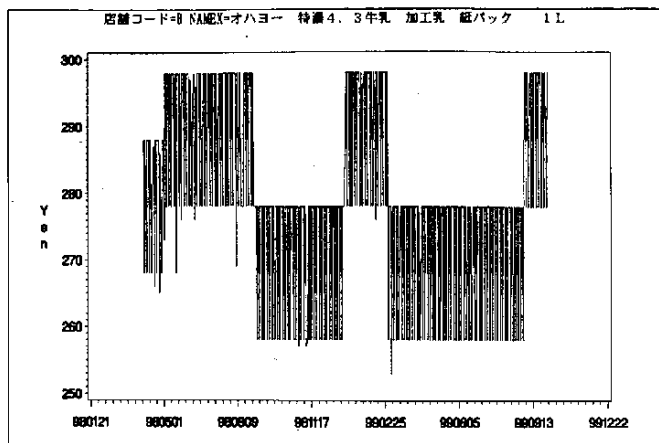


図6-16 販売価格の日次変化(商品別)[B店]

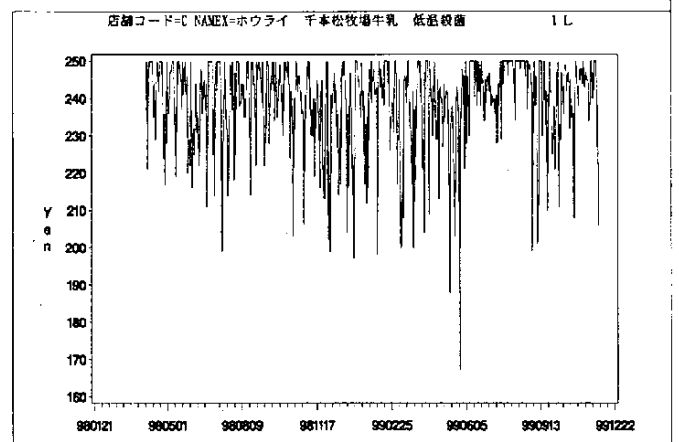
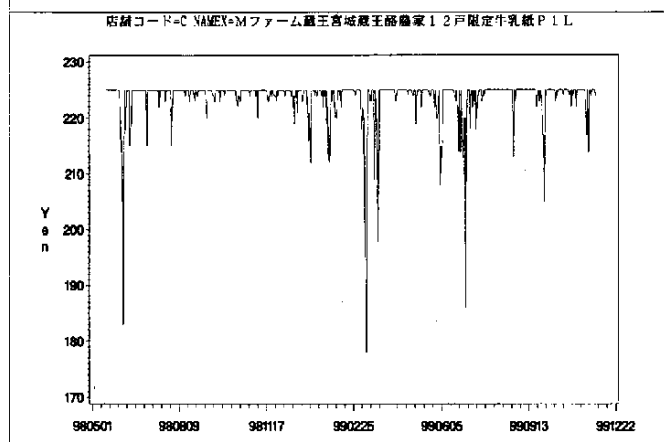
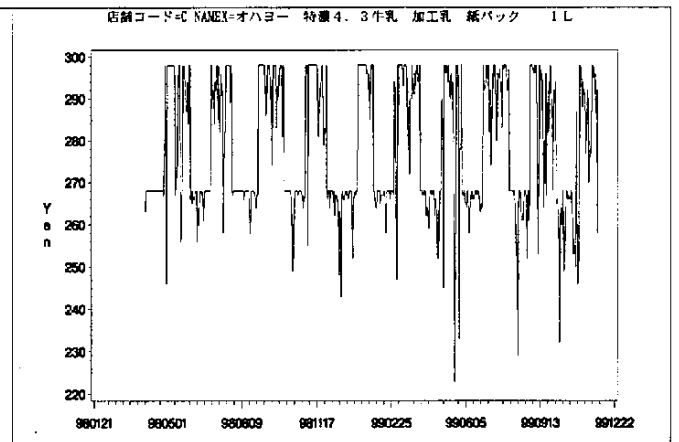
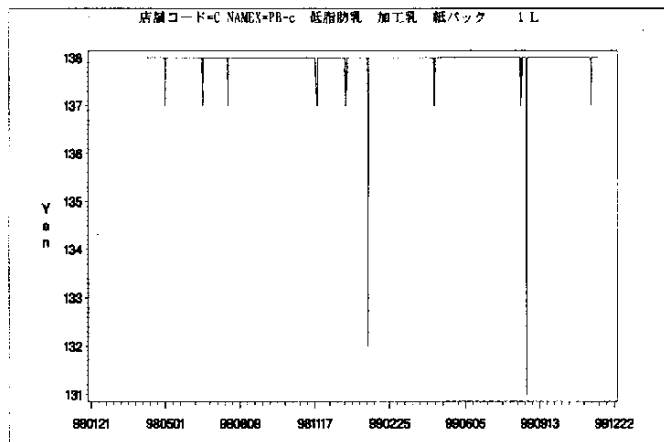
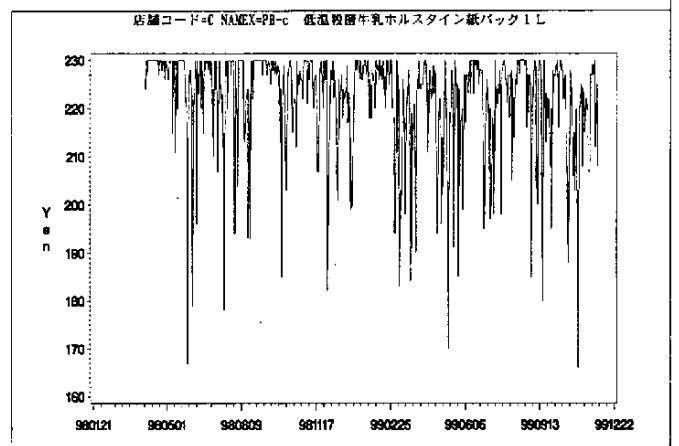
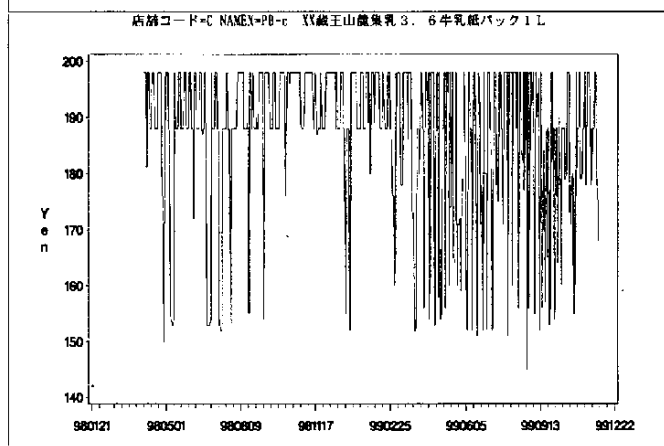
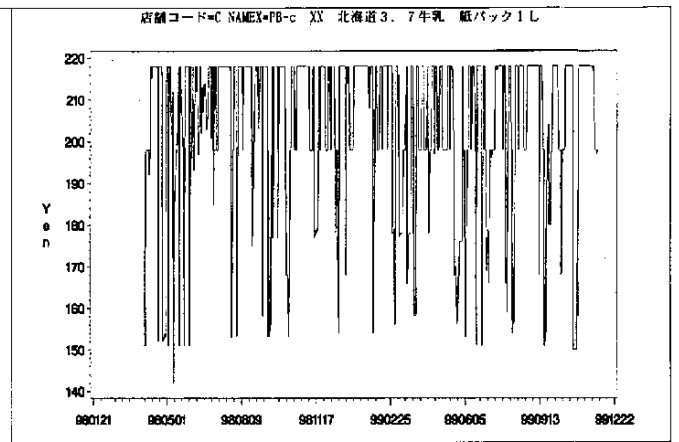
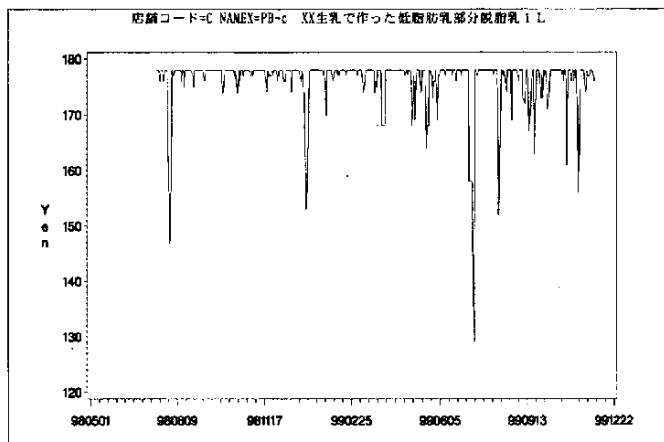


図6-16 販売価格の日次変化(商品別)[C店]

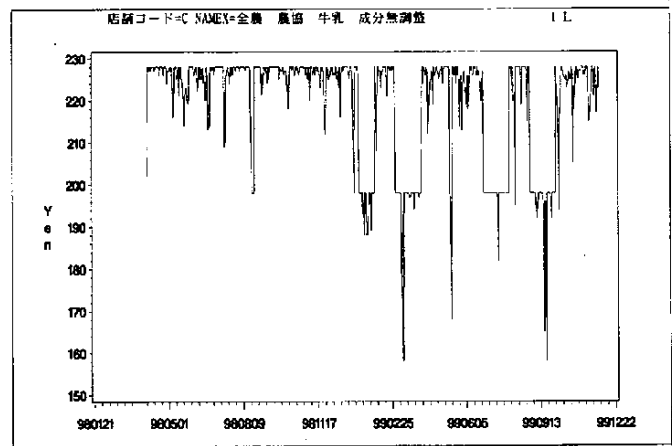
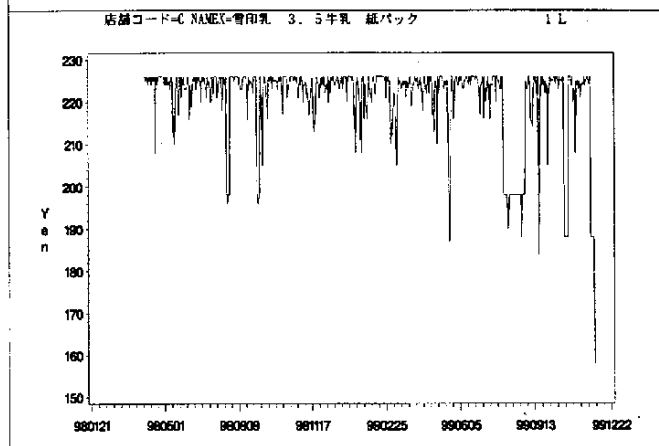
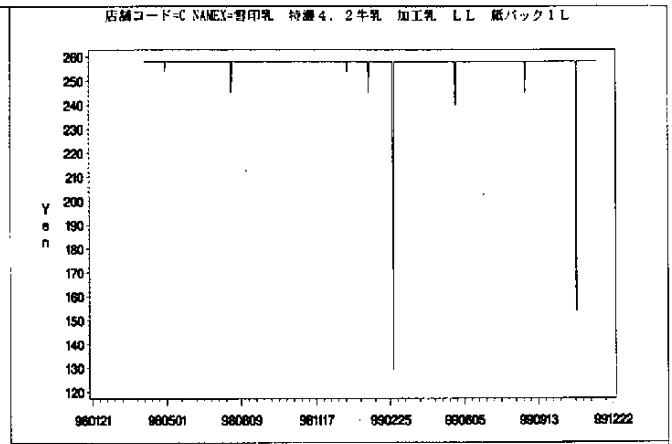
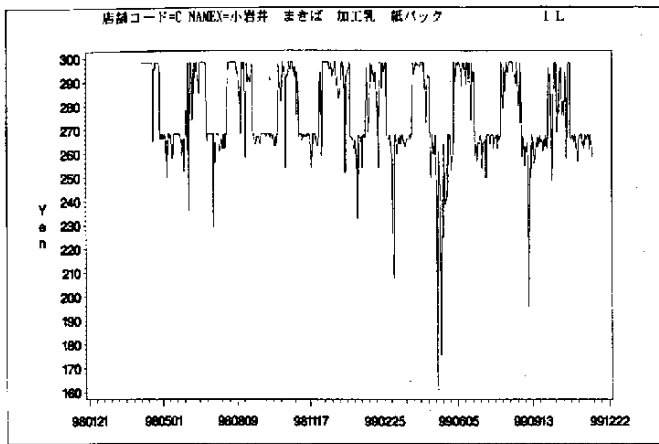


図6-16 販売価格の日次変化(商品別)[C店]

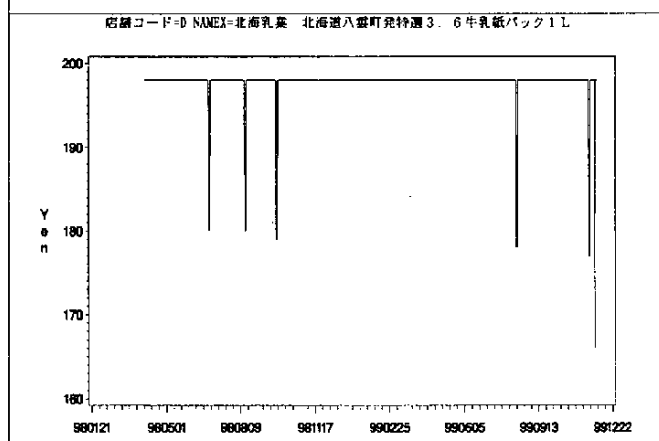
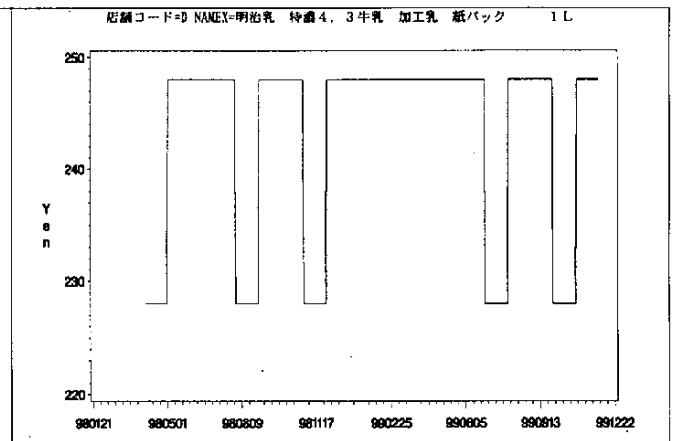
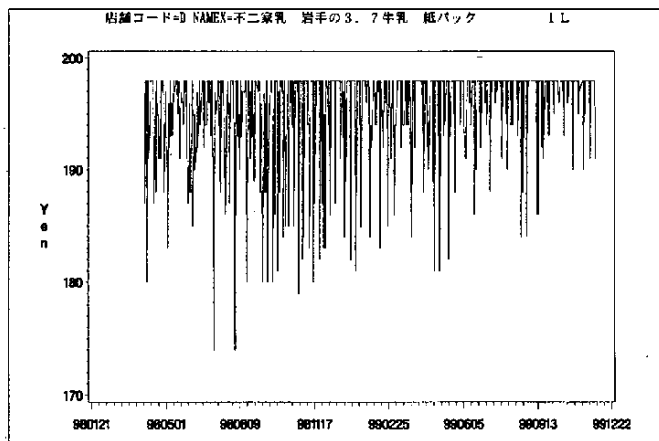
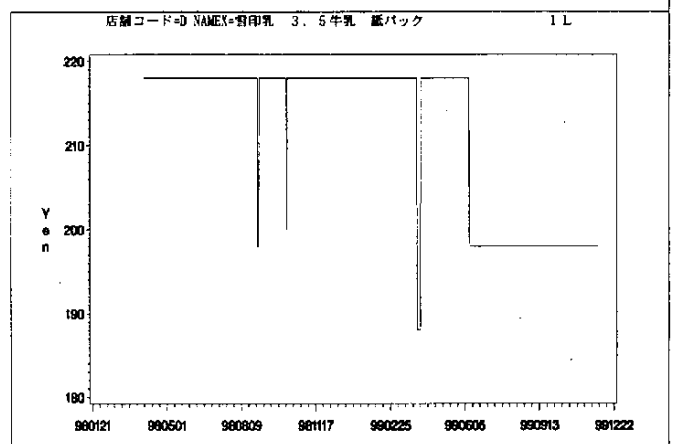
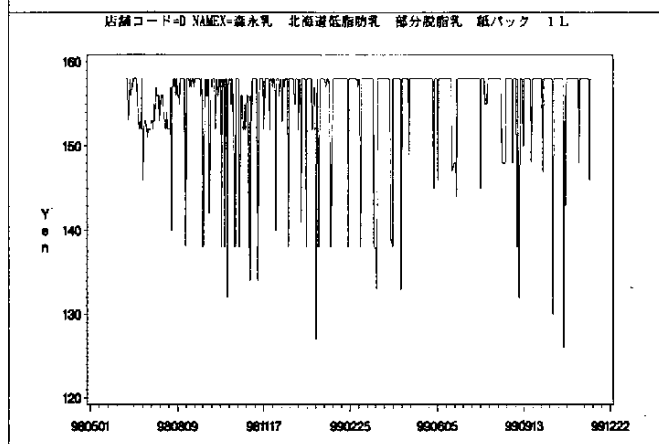
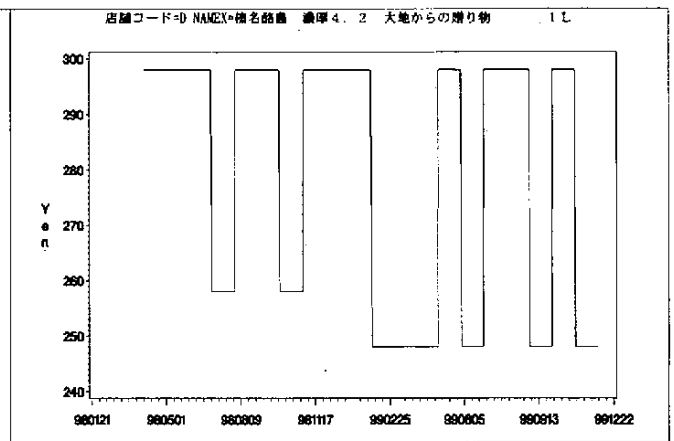
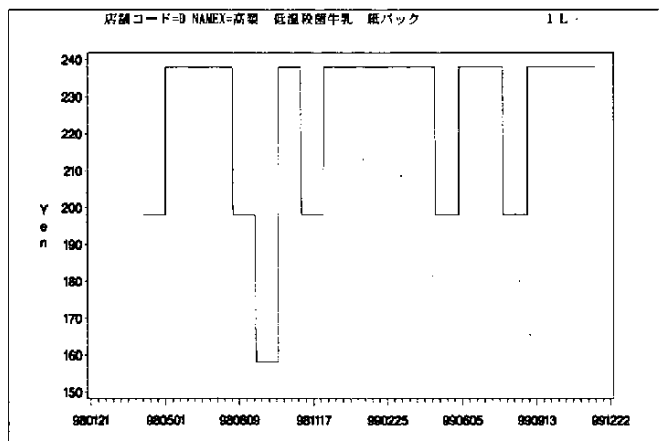


図6-16 販売価格の日次変化(商品別)[D店]

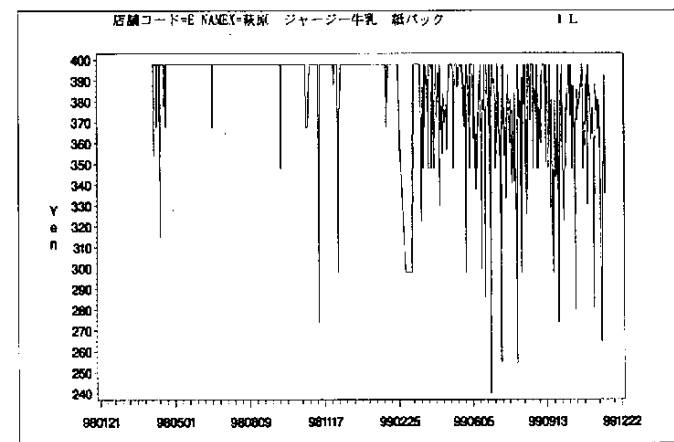
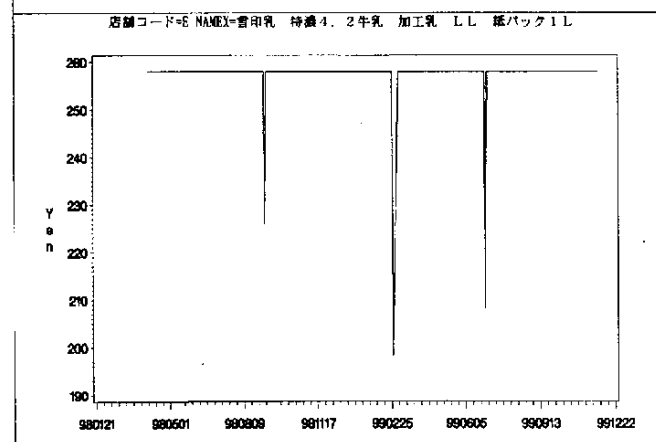
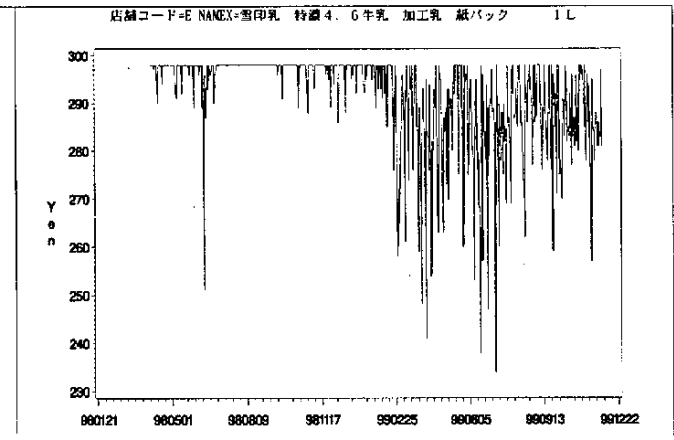
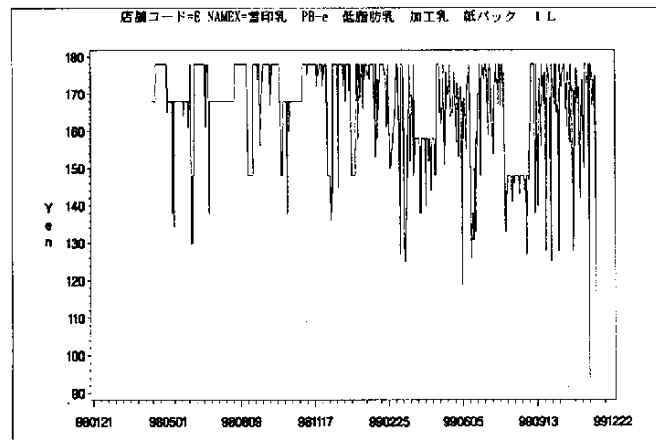
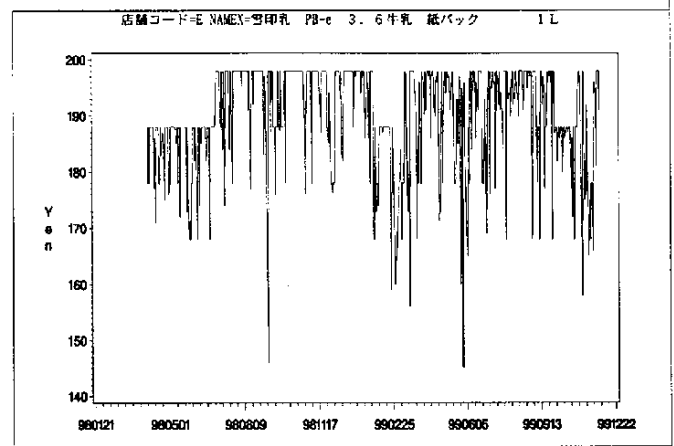
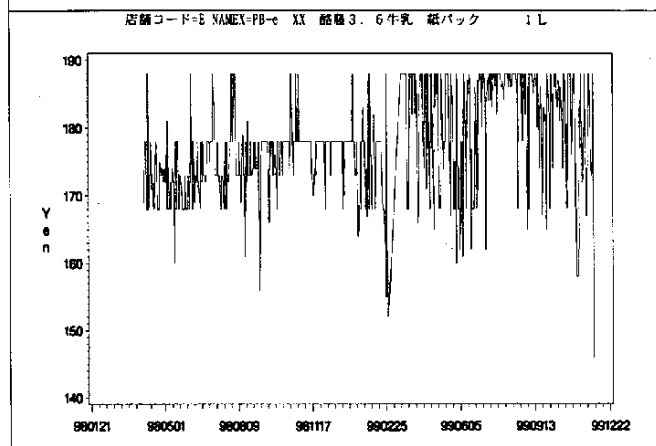
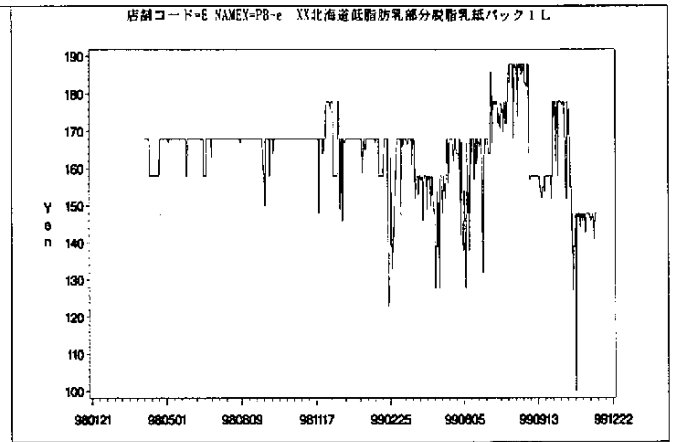
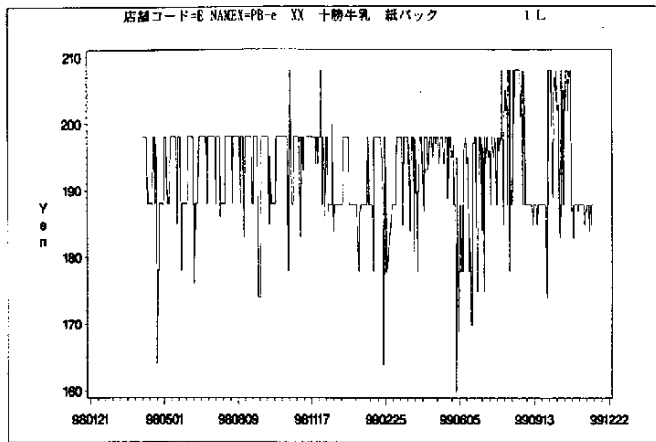


図6-16 販売価格の日次変化(商品別)[E店]

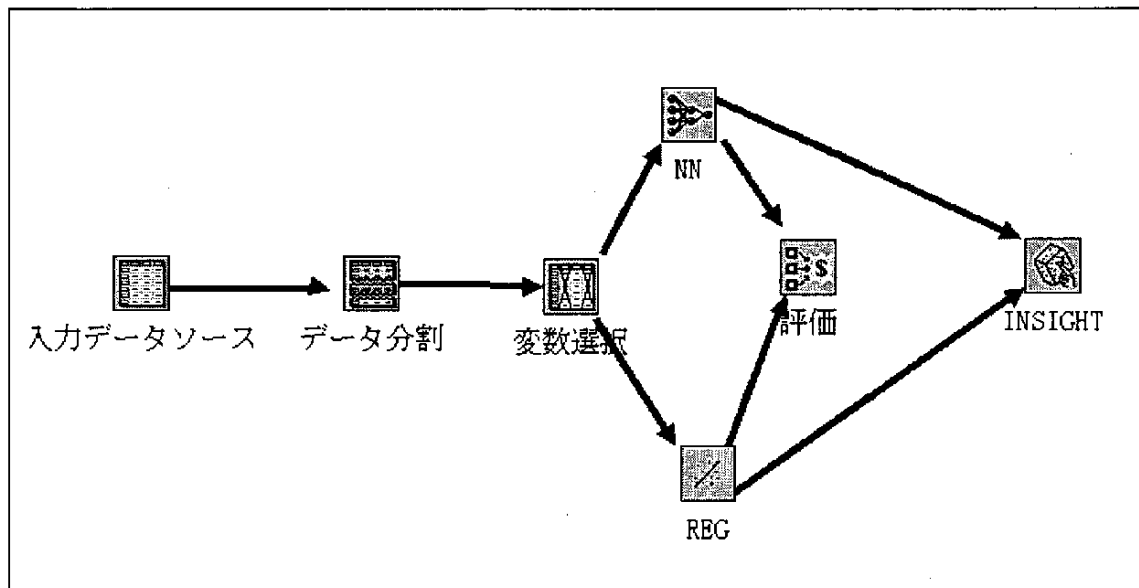


図6-17 EMのダイアグラム

「入力データソース」で分析データを定義する。次に「データ分割」で学習用、検証用、テスト用にデータセットを分けた。これは主に NN（ニューラルネットワークモデル）のための処理である。EM では分割方法として以下の3種類がラジオボタンで選択できる。

1. 単純無作為分割
2. 層別分割
3. ユーザー定義分割

多くの場合に単純無作為で指定した割合に分割する。ただ、この分析では 1998 年度のデータを学習と検証に使用し、そこで作ったモデルを 1999 年度データでテストするというアプローチを考えていたが、そのような指定は事前に DATA ステップで処理しておくか、EM の中でプログラミングする必要がある。ここでは単純無作為で分割した。1 年間に数日だけの特異日があることを考えると、分析データは 3 年分用意して各年度のデータを学習、検証、テストに使うべきかも知れない。

「変数選択」ノードでは重相関係数 2 乗を基準とした。ほかにカイ 2 乗値を基準にすることができる。量的変数の場合は CHAID と同様な基準で EM が「適切」にカテゴリ化する。しかし、いずれにせよ非線形な関係を考慮した変数選択はできないと思われるので、決定木によって有望な変数を見つけることも重要かと考えられる。

変数選択の結果、以下の変数が選択された。

1. 天気
2. 不快指数
3. 最高気温
4. 最高気温との差
5. 月
6. 曜日
7. 特異日
8. 商品4の価格
9. 商品5の価格
10. 商品7の価格

「ニューラルネットワーク」(NN)はデフォルト設定として、隠れ層を1つ、ニューロン数を3とした(図6-18)。

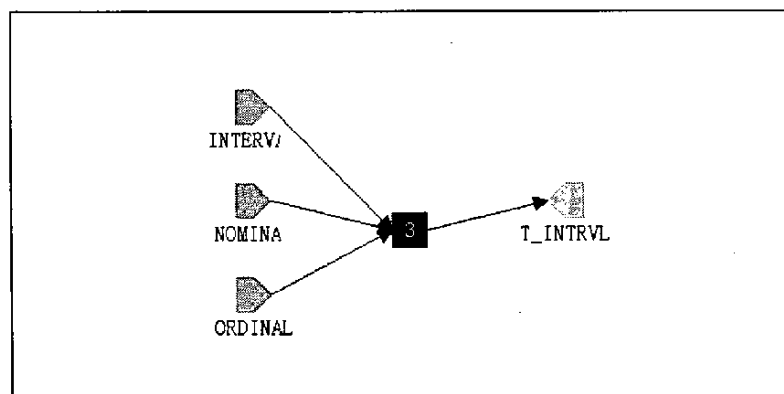


図6-18 ニューラルネットワークモデル

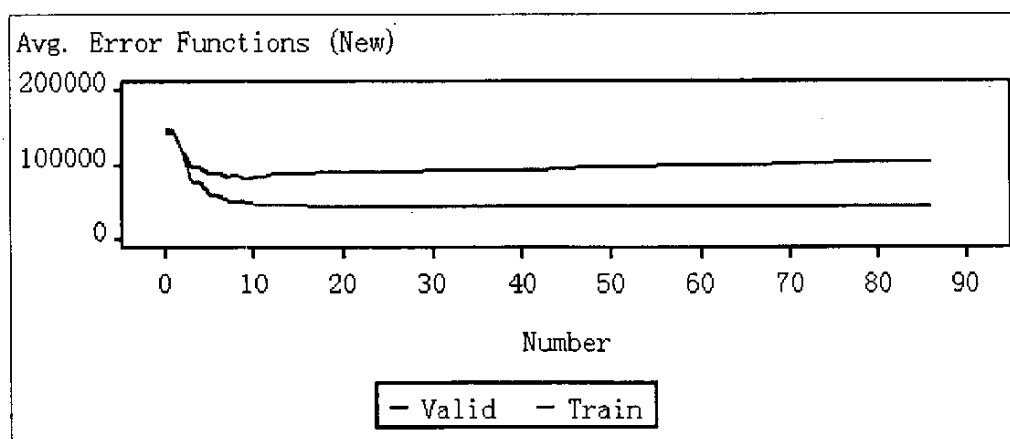


図6-19 ニューラルネットワークの実行結果(上線がValid、下線がTrain)

実行結果（図 6-19）から、反復 9 回で誤差関数が最適化したと判断されていることが分かる。

「重回帰分析」(REG) は伝統的な重回帰分析である。「評価」ノードにパスをつなぐことによって、NN と REG の両モデルの性能を評価することができる。

6. 6. モデルの検討

ニューラルネットワークモデルと重回帰分析モデルを表 6-5 の各種指標で比較すると、いずれも重回帰分析モデルの方が良好であった。

表 6-5 モデルの評価指標

	AIC	SBC	RMSE	RMSE(V)	RMSE(T)
NN	2651.9	2842	255.1	286.2	242.3
REG	2532.0	2591	211.1	258.2	209.9

この結果から重回帰分析モデルが良いという結論をするのは不十分かと考えられる。ニューラルネットワークモデルは本質的に非線型な場合に有効な方法であるから、変数選択の際に線形な基準を使ったことによる不利益が考えられる。またニューロンの数や隠れ層の数を変更して複数のモデルを検討すれば改善の余地が残っているからである。一方、重回帰分析モデルでは、このような試行の余地は少なく、最初の探索的データ解析から有望な変数を作り出すしれない。

「INSIGHT」ノードでは分析結果のデータセットを使って SAS/INSIGHT という探索的データ解析ツールを起動することができる。作成したモデルをテストデータに適用して、どの程度の予測精度になるか確認した。

予測値と実測値のプロットを、図 6-20 (NN) と図 6-21 (REG) に示した。残差プロットを、図 6-22 に NN モデル (左) と REG (右) に示した。散布図を見るとやはり重回帰モデルの方が良好である。相関係数の 2 乗 (寄与率) は、NN モデルが 0.581 であるのに対し、REG モデルは 0.687 であった。

またテストデータにおける実際の販売個数の差 (絶対値) は、NN モデルで平均 190 本、標準偏差 151 本で、REG モデルで平均 163 本、標準偏差 132 本であった。

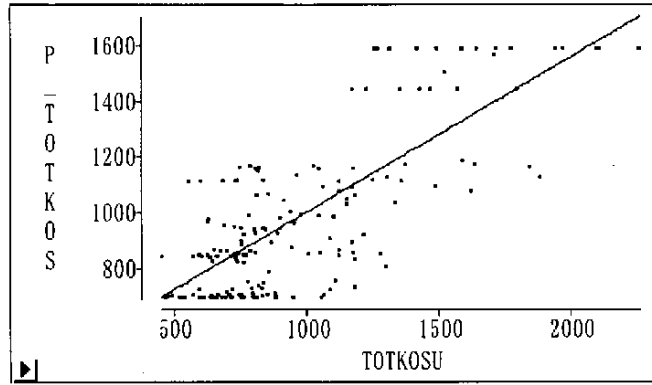


図 6 - 2 0 NN モデルの予測値と実測値プロット

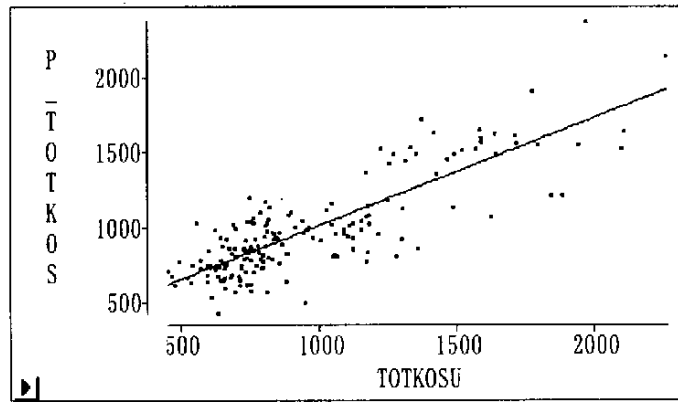


図 6 - 2 1 REG モデルの予測値と実測値プロット

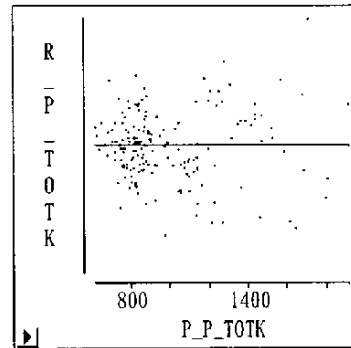
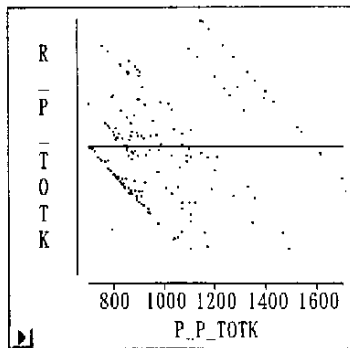


図 6 - 2 2 NN モデルの残差 (左) と REG モデルの残差 (右)

7. パネルデータを用いた購買行動の分析

消費者は、商品を買うとき、どんな点を比較・検討して購買を決定するのだろうか。価格や形、品質、ブランドなど購買に影響を与える要素は様々考えられる。ある人は価格を重視し、ある人はブランドを重視するだろう。また、Aというブランドは、値引きしなくても買うが、Bというブランドは値引きしなければ買わないといった購買行動も考えられる。ここでは、パネルデータを用いて消費者の購買行動を分析してみたい。分析は2段階に別れる。第1段階として、重視する商品属性（ここでいう商品属性とは、形やブランドのようなものだけでなく、価格のような可変的な要素も含める）によってパネルをグルーピングする、さらに第2段階として、グループごとに購買行動を探っていく。

分析には決定木と呼ばれる手法を適用する。「買った」商品と「買わない」商品の2種類のデータを作成し、2つを分類する項目を目的変数とする。説明変数には、商品属性を使う。

データは、カレールーの購買データを用いたが、そのままの形式では分析できない。様々な変換を施して分析用にデータを作成する必要がある。この作業は非常に複雑なものであり、まずそのことについて説明する。その後、分析結果について述べる。

データの作成と事前分析には、「SAS System」(SASインスティテュート・ジャパン)を使用した。分析には、主にエス・ピー・エス・エス社(以後、SPSS Inc.)が提供している「Clementine」の「C5.0」を利用した。また、手法の比較として同じくSPSS Inc 製の「AnswerTree」に組み込まれている「CHAID」と「C&RT」も試みている。

7. 1 分析データの作成

分析データの作成は、次の3つの作業からなる。

- ①分析対象(パネル・商品)の選択
- ②「買った」商品と「買わない」商品の定義と作成
- ③購買要因となる商品属性の選択と作成

7. 1. 1 パネルの選択

データは、まず1998年6月1日～1999年5月31日までの1年間におけるカレールーの購買データを対象にした。1年間にカレールーを購入したパネルは、1,162あったが、1パネルあたりの年間購買件数は、思ったより少ない。図7-1は、1パネルあたりの年間購買件数の分布である。10個以上購入しているパネルの割合は、全体の17%程度となってしまふ。しかし、分析には、ある程度の購買件数が必要とな

るので、分析対象を年間に10個以上購入したパネルに限定した。

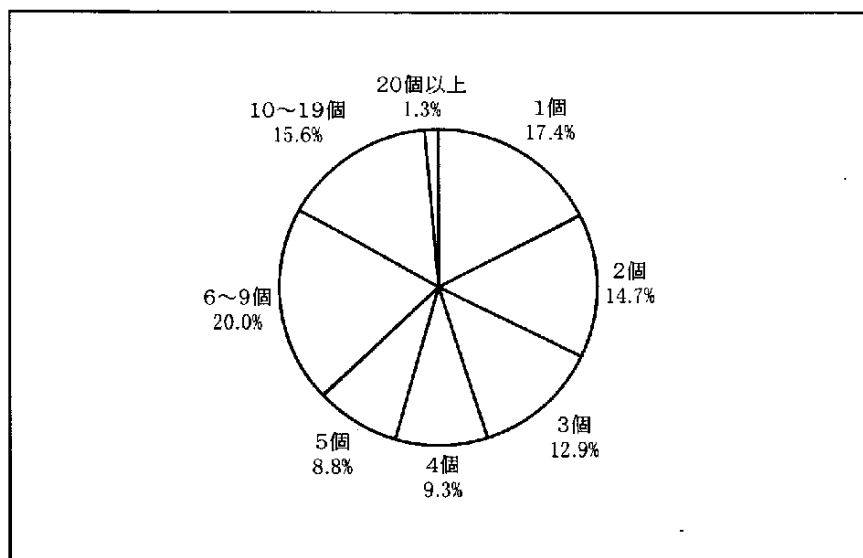


図7-1 1パネルあたり年間のカレールー購買件数の分布

7. 1. 2 商品の選択

1,162パネルが年間に購入した商品(カレールー)は59あった。商品のなかには季節限定的な商品や途中で店頭から外された商品も含まれている。そういった商品の購買パターンも興味深いだが、購買個数が少ないとデータ作成上問題があるため、ここでは年間通して販売された商品だけに絞った。販売月数の確認にはパネルデータではなく、店舗データを参考にした。1998年6月1日から1999年5月31日の間での販売開始日と販売終了日を割り出した結果、年間通して販売されたのは37商品だった。この中から販売個数の少ない商品や用途が特異な商品を除き、販売月数が11ヶ月の1商品を加え、最終的に30商品を選択した。

表7-1 販売月数別の商品数

販売月数	商品数
12ヶ月	37
11ヶ月	1
10ヶ月	4
9ヶ月以下	17
合計	59

最終的に分析に使用した購買データは、以下のとおり

店舗 : A点
商品 : カレールー (30 商品)
期間 : 1998年7月1日~1999年5月31日 (11ヶ月間)
パネル : カレールーを10個以上購入したパネル (146パネル)

7. 1. 3 購買・非購買データの商品の定義と作成

この分析では、買った商品と買わなかった商品をパネルごとに定義していく。あるパネルにとっての購入対象は一定期間に購入した商品のみと考え、その中で、1日ごとに買った商品と買わなかった商品に分けていく。

具体的な作成手順をデータ例と共に以下に示す。

- (1) あるパネルの11ヶ月間の購入履歴を抽出する (表7-2)
- (2) このパネルにとっての購入対象は、6商品と考える (表7-3)
- (3) 1日単位で6商品に対し買った、買わないの情報を付加する (表7-4)
1998年7月18日の購買行動は、「ハウス パーモンドカレー 中辛 250g」と「ハウス こくまるカレー 中辛 200g」を「買った」が、残りの4商品は「買わなかった」ことになる
- (4) このパネルがカレーを購入した日、すべてに(3)の処理をする
- (5) 146のパネルについて、(1)~(4)の手順を繰り返す

表7-2 あるパネルの11ヶ月間におけるカレー購買履歴

1998-07-18	パーモンドカレー	中辛2	250G
1998-07-18	こくまるカレー	中辛	200G
1998-09-09	ゴールデンカレー	中辛	240G
1998-11-25	ゴールデンカレー	中辛	240G
1998-12-20	ディナーカレー	中辛	100G
1998-12-20	ゴールデンカレー	中辛	100G
1998-12-20	こくまるカレー	中辛	100G
1999-02-04	ゴールデンカレー	中辛	240G
1999-02-04	PBカレー	中辛	240G
1999-04-19	ゴールデンカレー	中辛	240G

表7-3 あるパネルの11ヶ月間に購入したカレー（商品ベース）

バーモントカレー	中辛2	250G
こくまるカレー	中辛	200G
ゴールデンカレー	中辛	240G
ディナーカレー	中辛	100G
PBカレー	中辛	240G
ゴールデンカレー	中辛	240G

表7-4 あるパネルの1998年7月18日の購買行動

買った	バーモントカレー	中辛2	250G
買った	こくまるカレー	中辛	200G
買わない	ゴールデンカレー	中辛	240G
買わない	ディナーカレー	中辛	100G
買わない	PBカレー	中辛	240G
買わない	ゴールデンカレー	中辛	240G

7. 1. 4 商品属性の選択と作成

回帰分析や判別分析では、ある事象の要因としてモデルに組み込む項目を説明変数と呼んだり、独立変数と呼んだりする。ここでは、消費者が商品を選択するときに基準となるような商品属性を説明変数としてモデルに組み込む。具体的にカレーで考えれば、「メーカー」や「ブランド」、「辛さ」、「値引率」などが説明変数になるだろう。

また、説明変数のもつ特定の値をカテゴリー値と呼ぶことにする。「辛さ」を説明変数とするならば甘口や辛口といったことが、カテゴリー値にあたる。購買行動を分析する上で、どういった項目を説明変数に採用するのか。その説明変数の値をどうかategorizeするのかによって結果も異なってくる。

今回、説明変数として、「メーカー」、「辛さ」、「クラス」、「容量」、「値引率」の5属性（表7-5）を採用した。選択要因は、他にも考えられるがサンプルサイズ（この場合は、1パネルあたりの購買データ）が小さいので説明変数とカテゴリー値を限定した。ブランドなどは、購買要因として興味のある商品属性だが、「メーカー」と同値になることもあり、そうなるとどちらが要因か区別がつかなくなる。そこで、ブランドを任意で分類した「クラス」を作成した。「クラス」のカテゴリー分けは、昔から販売され馴染みのある「定番・一般」、価格が高めの「高級・本格」、ここ数年で発売されてきた「中級・後発」の3つにした。「メーカー」と「クラス」の組み合わせによってほしいブランドが規定できる。

「値引率」は、商品選択をするうえで非常に重要な基準となるが、説明変数として採用するには大きな障害がある。それは、買わなかった商品の「値引率」が分からないことである。そこで、店舗データを利用することにした。分析に使用したパネルデ

一夕は、A店でのご購買履歴であり、A店の店舗データから販売価格が分かる。

ただし、それでも「値引率」が判明しないときがある。A店に販売履歴がないときである。その場合、少し恣意的だが前後にある日付の販売価格のうち高い方を採用した。店舗データに販売履歴がないのは、その日にその商品が全く売れなかったか、店頭自体にその商品がおかれていなかったかのどちらかだ。どちらの理由によるものなのかはデータだけでは判断がつかない。

表7-5 購買行動分析で用いる説明変数とカテゴリー値

説明変数	メーカー	辛さ	クラス	容量	値引率
カテゴリー値	メーカーA	甘口	定番・一般	200g以上	注1)
	メーカーB	中辛	高級・本格	200g未満	
	メーカーC	辛口	中級・後発		
	PB				

注1) 値引率 = (販売価格 - 通常価格) / 通常価格 * 100

説明変数やカテゴリー値を決める際には、事前に行った因子分析の結果（バリマクス回転）が参考になる。ただし分析に使用したデータ形式は、購買行動分析での形式とは異なる。表7-6にあるように、オブザベーションは146パネル。項目は、30商品ごとの購買個数シェアを使用している。

表7-7は、因子分析の結果から得られた因子負荷行列である。各軸は、146パネルの市場構造が表われている。同一軸内で*がついている商品同士は、競合関係にあると考えられる。ただし、数値の符号が逆転している場合は、競合関係ではない。+の符号の付いている商品を買うパネルは、-の符号の商品は買わないという関係になる。

第1軸では、辛口の商品が競合商品として並んでいる。甘口の商品にはマイナスの符号がついているので、辛いカレーをよく購入するパネルは、甘いカレーは買わないということが分かる。第2軸では、比較的高級な商品が並び、第3軸では、「メーカー」によって商品が別れている。

因子分析の結果から、「辛さ」、「クラス」、「メーカー」が購買行動を分析するうえで必要な商品属性であることがわかる。

表7-6 因子分析で使用したデータの形式

パネルNo	商品1	商品2	商品30
1	4.0	3.0		0.0
2	0.0	10.5		0.0
⋮				
⋮				
146	3.5	0.0		20.0

・商品1の購買個数シェア = 商品1の購買個数 / 30商品の購買個数 * 100

表7-7 因子分析の結果 (バリマクス回転後の因子負荷行列)

第1軸	第2軸	第3軸	商品			
71 *	-9	1	メーカーC	中級・後発	辛口	200g
67 *	-11	10	PB		辛口	240g
54 *	-2	-2	メーカーA	定番・一般	辛口	240g
53 *	-1	-5	メーカーB	中級・後発	辛口	200g
24	-15	-3	PB		中辛	240g
-18	-12	-17	PB		甘口	240g
-36 *	-27 *	26 *	メーカーC	定番・一般	甘口 1	250g
-6	57 *	11	メーカーC	高級・本格	中辛	140g
-4	57 *	7	メーカーA	高級・本格	甘口	100g
-4	50 *	10	メーカーC	高級・本格	甘口	140g
-3	46 *	-6	メーカーA	高級・本格	辛口	100g
-4	39 *	-5	メーカーA	高級・本格	中辛	100g
-2	36 *	-5	メーカーC	高級・本格	辛口	140g
-4	28 *	18	メーカーB	高級・本格	中辛	155g
-1	17	4	メーカーC	中級・後発	中辛	100g
-15	-20	0	メーカーC	定番・一般	中辛 2	125g
-11	-36 *	12	メーカーC	定番・一般	中辛 2	250g
-12	-13	64 *	メーカーC	定番・一般	中辛 4	220g
-18	5	58 *	メーカーC	定番・一般	甘口 1	125g
-6	13	55 *	メーカーC	定番・一般	中辛 4	125g
24	-24	33 *	メーカーC	定番・一般	辛口 3	250g
8	-12	22	メーカーC	定番・一般	辛口 5	220g
-1	0	-6	メーカーA	定番・一般	中辛	100g
10	-10	-17	メーカーA	定番・一般	中辛	240g
-4	0	-18	メーカーA	中級・後発	中辛	200g
11	-17	-20	メーカーC	中級・後発	中辛	200g
-21	-6	-23	メーカーB	中級・後発	甘口	200g
3	-6	-25	メーカーB	中級・後発	中辛	200g
-5	2	-25	メーカーA	中級・後発	辛口	200g
-26 *	-9	-29 *	メーカーA	定番・一般	甘口	240g

・数値は、見やすいように100倍して、相対的に大きな絶対値に*を付けている。

7. 1. 5 最終的なデータ形式

購買行動分析は最終的に、表7-8のデータ形式で行なった。

表7-8 分析に用いる最終的なデータ形式

日付	購入	メーカー	辛さ	クラス	容量	値引率
1998/7/18	買わない	メーカーA	中辛	定番・一般	200g未満	-16.5
	買わない	メーカーA	中辛	高級・本格	200g以上	0
	買わない	メーカーA	中辛	定番・一般	200g以上	0
	買った	メーカーC	中辛	定番・一般	200g未満	-41.1
	買わない	メーカーC	中辛	中級・後発	200g以上	0
	買った	メーカーC	中辛	中級・後発	200g未満	-24.7
	買わない	P B	中辛	定番・一般	200g未満	0
1998/9/9	買った	メーカーA	中辛	定番・一般	200g未満	-16.5
	買わない	メーカーA	中辛	高級・本格	200g以上	0
	買わない	メーカーA	中辛	定番・一般	200g以上	0
	買わない	メーカーC	中辛	定番・一般	200g未満	-33.6
	買わない	メーカーC	中辛	中級・後発	200g以上	0
	買わない	メーカーC	中辛	中級・後発	200g未満	-10.1
	買わない	P B	中辛	定番・一般	200g未満	0
1998/11/25	買った	メーカーA	中辛	定番・一般	200g未満	-29.1
	買わない	メーカーA	中辛	高級・本格	200g以上	-26.1
	買わない	メーカーA	中辛	定番・一般	200g以上	0
	買わない	メーカーC	中辛	定番・一般	200g未満	-41.5
	買わない	メーカーC	中辛	中級・後発	200g以上	0
	買わない	メーカーC	中辛	中級・後発	200g未満	-10.1
	買わない	P B	中辛	定番・一般	200g未満	0
1998/12/20	買わない	メーカーA	中辛	定番・一般	200g未満	-31.6
	買った	メーカーA	中辛	高級・本格	200g以上	-29.9
	買った	メーカーA	中辛	定番・一般	200g以上	-5.1
	買わない	メーカーC	中辛	定番・一般	200g未満	-29.2
	買った	メーカーC	中辛	中級・後発	200g以上	-5.1
	買わない	メーカーC	中辛	中級・後発	200g未満	-13.1
	買わない	P B	中辛	定番・一般	200g未満	-5.4
1999/2/4	買った	メーカーA	中辛	定番・一般	200g未満	-29.1
	買わない	メーカーA	中辛	高級・本格	200g以上	0
	買わない	メーカーA	中辛	定番・一般	200g以上	0
	買わない	メーカーC	中辛	定番・一般	200g未満	-41.1
	買わない	メーカーC	中辛	中級・後発	200g以上	0
	買わない	メーカーC	中辛	中級・後発	200g未満	-24.7
	買った	P B	中辛	定番・一般	200g未満	0
1999/4/19	買った	メーカーA	中辛	定番・一般	200g未満	-32.9
	買わない	メーカーA	中辛	高級・本格	200g以上	-4.5
	買わない	メーカーA	中辛	定番・一般	200g以上	-5.1
	買わない	メーカーC	中辛	定番・一般	200g未満	-3.2
	買わない	メーカーC	中辛	中級・後発	200g以上	-3.4
	買わない	メーカーC	中辛	中級・後発	200g未満	-14.1
	買わない	P B	中辛	定番・一般	200g未満	-2.7

7. 2 分析

因子分析の結果から、購買行動に影響を与えている商品属性はだいたい分かる。ここから、商品属性の組み合わせ（特に値引率）による購入意向の違いを探るためにさらに決定木によって分析をすすめていく。前節までで作成したデータを用いて、分析は次の手順で行なう。

- (1) 146 のパネルごとに、購買に最も影響をあたえる商品属性（以後、第1基準と呼ぶ）を決める
- (2) 第1基準とフェース（パネル属性）との関連性をみる
- (3) 第1基準ごとにパネルをグルーピングする
- (4) グループごとに、パネルの購買データを合算して購買行動を分析する

7. 2. 1 第1基準の作成

パネルごとに異なる購買要因を調べるために、146 のパネルを個別に C5.0 で分析した。表7-9は、個別に C5.0 で分析した結果の一部（3パネル分）である。No. 1 のパネルの分析結果を例にとってツリーの見方を説明する。まず、最初に買うか買わないかの基準となっているのは値引率となっている（つまり、このパネルの第1基準は値引率）。そして、その境界は、5.1%になる。表7-9の(1)は、

値引率 \leq -5.1

となっているが、これは値引率が 5.1%以上のときを表わしている。値引率が、マイナスの数値で表現されているため、符号の意味が逆転しているので注意が必要である。

(1) の下にはメーカーが並んでいる。このような状態を枝分かれと呼ぶ。値引率が 5.1%以上のときは、メーカーによって購買パターンが異なっている。また、ツリーの深さを階層という言葉で表現する。このツリーは、6階層までできているが、あまり深い階層の枝分かれは信用できない。サンプルサイズが小さいうえ、ツリーを作成するとき設定する枝分かれ基準が非常に甘いものになっているからだ。

次に、(2) をみると、

値引率 $>$ -5.1 (19.0, 0.947) \rightarrow 買わない

となっている。値引率が 5.1%に満たない場合は、その商品は買わないと判定している。カッコ内の数字は、前部が出現件数で、後部がその比率となっている。この場合、「値引率が 5.1%に満たない商品がのべ 19 あり、そのうち 94.7%は買っていない」ことを

表わしている。

146 パネルの第1基準は、表7-10にあるように値引率が最も多かった。(パネル数は55) 次いで、辛さ、メーカー、クラスと続く。容量は5パネルしかなく、全く第1基準が見出せなかったパネルも18あった。理由は、①購買パターンがバラバラで、購入比率を改善するような要因が見つからない、②同一商品しか買っていないため比較できない、などが考えられる

表7-9 パネルごとの購買行動分析の結果(3パネル分)

オプション	
枝刈り度	: 5%
枝あたり最小レコード	: 1
No.1 買った商品: 10 買わない商品: 32	
値引き率	< -5.1 (1)
メーカー その他 (0.0, 1.0)	-> 買わない
メーカー PB (1.0, 1.0)	-> 買わない
メーカー メーカーC	
容量 200g以上 (1.0, 1.0)	-> 買う
容量 200g未満 (11.0, 0.818)	-> 買わない
メーカー メーカーA	
値引き率	< -29.1
値引き率	< -31.6
値引き率	< -32.9 (1.0, 1.0) -> 買う
値引き率	> -32.9 (1.0, 1.0) -> 買わない
値引き率	> -31.6 (3.0, 1.0) -> 買う
値引き率	> -29.1 (5.0, 0.6) -> 買わない
値引き率	> -5.1 (19.0, 0.947) -> 買わない (2)
No.2 買った商品: 11 買わない商品: 9	
辛さ 甘口 (10.0, 0.9)	-> 買わない
辛さ 中辛 (10.0, 1.0)	-> 買う
辛さ 辛口 (0.0, 1.0)	-> 買う
No.3 買った商品: 12 買わない商品: 24	
値引き率	< -21.7
辛さ 甘口 (0.0, 1.0)	-> 買う
辛さ 中辛 (12.0, 0.583)	-> 買う
辛さ 辛口	
クラス 高級・本格 (0.0, 1.0)	-> 買わない
クラス 中級・後発 (3.0, 0.667)	-> 買わない
クラス 定番・一般	
値引き率	< -33.6 (5.0, 0.8) -> 買わない
値引き率	> -33.6 (4.0, 0.75) -> 買う
値引き率	> -21.7 (12.0, 1.0) -> 買わない

表7-10 第1基準ごとのパネル数

第1基準	パネル数
メーカー	22
辛さ	31
クラス	20
容量	5
値引率	50
なし	18
合計	146

7. 2. 2 フェース (パネル属性) 分析

パネルデータの特徴の1つとして、フェースデータがあることが挙げられる。フェースには、「主婦の年齢」、「主婦の職業」、「世帯収入」、「住居形態」、「子供の人数」などがあり、世帯属性によって商品選択基準に差があるのかを分析できる。まずは、第1基準を表頭、フェースを表側にとったクロス集計を行い、第1基準とフェースとの関連性を調べてみた。

結果は、表7-11にある。件数が少ないためあまり顕著な結果は出ていないが、特徴として、

- ・主婦の年齢は、30代までの方が値引率の割合が高く、40代以上の方が少ない
- ・専業主婦は、値引率の割合が低い
- ・一戸建ての方が集合住宅より値引率の割合が高い

などが挙げられる。

さらに、隠れた結果がないかC5.0にかけて探してみる。ただし、パネル数が少ないため、第1基準が値引率のグループとそれ以外のグループとの差を分析するだけにとどめた。

結果(表7-12)は、「20代後半の家族人数」が第1基準となった。20代後半の家族がいるパネルは、80.0%値引き率グループに属する。一方、20代後半の家族がいないパネルでは、「主婦の職業」で枝が分かれしている。「主婦の職業：事務職、技能職、お勤めその他、自由業」のようなパネルは、78.9%が値引率グループになる。「専業主婦」の場合、さらに「世帯収入」でわかれ、1,000万円以上だと値引率グループが多くなる結果となった。

ところで、ツリーも枝分かれが多くなると、複雑になり分かりづらい。C5.0には、ツリー以外にルールと呼ばれる出力形式があるが、表7-13にあるように、「IF ~ THEN ...」の形式で表現される。ツリーが複雑なときだけでなく、例えば値引率グループの割合が非常に高くなるような組み合わせをダイレクトに知りたいときは有効である。

表7-11 第1基準とフェースのクロス集計表(1)

	パネル数	値引率	メーカー	辛さ	クラス	容量	なし
全体	146	50 34.2	22 15.1	31 21.2	20 13.7	5 3.4	18 12.3
●主人の職業 役員・管理職	41	17 41.5	9 22.0	7 17.1	4 9.8	2 4.9	2 4.9
事務職	10	3 30.0	1 10.0	3 30.0	0 0.0	1 10.0	2 20.0
専門研究職(弁護士など)	27	9 33.3	3 11.1	5 18.5	5 18.5	0 0.0	5 18.5
技能職(キーパンチャーなど)	22	6 27.3	4 18.2	4 18.2	4 18.2	0 0.0	4 18.2
販売・サービス職(スーパーの販売員等)	13	4 30.8	0 0.0	5 38.5	2 15.4	1 7.7	1 7.7
その他お勤め	5	2 40.0	1 20.0	2 40.0	0 0.0	0 0.0	0 0.0
自営業	19	5 26.3	4 21.1	5 26.3	3 15.8	1 5.3	1 5.3
自由業	1	1 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
無職	5	1 20.0	0 0.0	0 0.0	2 40.0	0 0.0	2 40.0
その他	0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
●主人の年齢 34才以下	5	1 20.0	1 20.0	2 40.0	1 20.0	0 0.0	0 0.0
30代後半	12	7 58.3	2 16.7	1 8.3	1 8.3	0 0.0	1 8.3
40代前半	30	9 30.0	2 6.7	8 26.7	6 20.0	1 3.3	4 13.3
40代後半	37	12 32.4	4 10.8	11 29.7	6 16.2	1 2.7	3 8.1
50才以上	58	19 32.8	13 22.4	8 13.8	6 10.3	3 5.2	9 15.5
●主婦の学歴 中学卒	5	1 20.0	0 0.0	2 40.0	0 0.0	1 20.0	1 20.0
高校卒	91	33 36.3	13 14.3	21 23.1	11 12.1	3 3.3	10 11.0
短大卒	38	12 31.6	7 18.4	6 15.8	7 18.4	1 2.6	5 13.2
大学・大学院卒	12	4 33.3	2 16.7	2 16.7	2 16.7	0 0.0	2 16.7
●主婦の職業 役員・管理職	0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
事務職	2	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	2 100.0
専門研究職(弁護士など)	14	7 50.0	1 7.1	2 14.3	1 7.1	1 7.1	2 14.3
技能職(キーパンチャーなど)	7	4 57.1	0 0.0	1 14.3	1 14.3	0 0.0	1 14.3
販売・サービス職(スーパーの販売員等)	26	7 26.9	1 3.8	9 34.6	3 11.5	1 3.8	5 19.2
その他お勤め	7	4 57.1	2 28.6	1 14.3	0 0.0	0 0.0	0 0.0
自営業	6	1 16.7	1 16.7	0 0.0	3 50.0	0 0.0	1 16.7
自由業	1	1 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
専業主婦	81	26 32.1	16 19.8	17 21.0	12 14.8	3 3.7	7 8.6
その他	2	0 0.0	1 50.0	1 50.0	0 0.0	0 0.0	0 0.0
●主婦の年齢 34才以下	12	5 41.7	2 16.7	3 25.0	1 8.3	0 0.0	1 8.3
30代後半	19	9 47.4	2 10.5	1 5.3	4 21.1	0 0.0	3 15.8
40代前半	38	11 28.9	5 13.2	11 28.9	7 18.4	1 2.6	3 7.9
40代後半	35	10 28.6	6 17.1	11 31.4	2 5.7	2 5.7	4 11.4
50才以上	42	15 35.7	7 16.7	5 11.9	6 14.3	2 4.8	7 16.7

表7-11 第1基準とフェースのクロス集計表(2)

	パネル数	値引率	メーカー	辛さ	クラス	容量	なし
全体	146	50 34.2	22 15.1	31 21.2	20 13.7	5 3.4	18 12.3
●世帯の収入 200万円未満	2	1 50.0	0 0.0	1 50.0	0 0.0	0 0.0	0 0.0
200~300万円未満	2	2 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
300~400万円未満	7	1 14.3	2 28.6	2 28.6	0 0.0	0 0.0	2 28.6
400~500万円未満	16	6 37.5	2 12.5	4 25.0	2 12.5	0 0.0	2 12.5
500~600万円未満	14	6 42.9	2 14.3	2 14.3	1 7.1	1 7.1	2 14.3
600~800万円未満	31	8 25.8	4 12.9	6 19.4	8 25.8	2 6.5	3 9.7
800~1000万円未満	33	8 24.2	5 15.2	8 24.2	7 21.2	0 0.0	5 15.2
1000~2000万円未満	39	18 46.2	6 15.4	8 20.5	2 5.1	1 2.6	4 10.3
2000万円以上	2	0 0.0	1 50.0	0 0.0	0 0.0	1 50.0	0 0.0
●住居形態 持家一戸建	54	21 38.9	8 14.8	10 18.5	6 11.1	3 5.6	6 11.1
持家集合住宅	57	16 28.1	11 19.3	10 17.5	11 19.3	1 1.8	8 14.0
借家一戸建	3	2 66.7	0 0.0	0 0.0	1 33.3	0 0.0	0 0.0
借家集合住宅	32	11 34.4	3 9.4	11 34.4	2 6.3	1 3.1	4 12.5
給与住宅(社宅、寮)	0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
その他(間借りなど)	0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
●居住年数 1年未満	8	5 62.5	0 0.0	2 25.0	1 12.5	0 0.0	0 0.0
1年~2年未満	6	0 0.0	1 16.7	2 33.3	2 33.3	0 0.0	1 16.7
2年~3年未満	2	1 50.0	0 0.0	1 50.0	0 0.0	0 0.0	0 0.0
3年~5年未満	13	5 38.5	2 15.4	2 15.4	2 15.4	1 7.7	1 7.7
5年~10年未満	20	9 45.0	1 5.0	3 15.0	3 15.0	0 0.0	4 20.0
10年以上	97	30 30.9	18 18.6	21 21.6	12 12.4	4 4.1	12 12.4
●住居の部屋数 2LDK以下	7	3 42.9	0 0.0	3 42.9	0 0.0	0 0.0	1 14.3
2LDKまたは3DK	38	15 39.5	6 15.8	10 26.3	3 7.9	1 2.6	3 7.9
3LDKまたは4DK	58	17 29.3	10 17.2	9 15.5	11 19.0	2 3.4	9 15.5
4LDKまたは5DK	26	10 38.5	3 11.5	6 23.1	5 19.2	0 0.0	2 7.7
5LDKまたは6DK以上	17	5 29.4	3 17.6	3 17.6	1 5.9	2 11.8	3 17.6
●家族人数 3人以下	14	6 42.9	1 7.1	2 14.3	1 7.1	1 7.1	3 21.4
4人	72	24 33.3	14 19.4	13 18.1	11 15.3	2 2.8	8 11.1
5人	43	15 34.9	2 4.7	12 27.9	7 16.3	2 4.7	5 11.6
6人以上	17	5 29.4	5 29.4	4 23.5	1 5.9	0 0.0	2 11.8
●子供の人数 0人	3	1 33.3	0 0.0	1 33.3	0 0.0	0 0.0	1 33.3
1人	13	6 46.2	1 7.7	1 7.7	1 7.7	1 7.7	3 23.1
2人	80	26 32.5	15 18.8	14 17.5	13 16.3	3 3.8	9 11.3
3人以上	50	17 34.0	6 12.0	15 30.0	6 12.0	1 2.0	5 10.0

表7-11 第1基準とフェースのクロス集計表(3)

	パネル数	値引率	メーカー	辛さ	クラス	容量	なし	
全体	146	50 34.2	22 15.1	31 21.2	20 13.7	5 3.4	18 12.3	
●家族人数(0~6才)	0人	124	40 32.3	20 16.1	27 21.8	16 12.9	5 4.0	16 12.9
	1人	17	7 41.2	1 5.9	4 23.5	3 17.6	0 0.0	2 11.8
	2人以上	5	3 60.0	1 20.0	0 0.0	1 20.0	0 0.0	0 0.0
●家族人数(7~12才)	0人	76	26 34.2	13 17.1	13 17.1	9 11.8	4 5.3	11 14.5
	1人	45	16 35.6	6 13.3	8 17.8	9 20.0	1 2.2	5 11.1
	2人以上	25	8 32.0	3 12.0	10 40.0	2 8.0	0 0.0	2 8.0
●家族人数(13~18才)	0人	70	24 34.3	11 15.7	14 20.0	9 12.9	2 2.9	10 14.3
	1人	44	16 36.4	5 11.4	10 22.7	7 15.9	2 4.5	4 9.1
	2人以上	32	10 31.3	6 18.8	7 21.9	4 12.5	1 3.1	4 12.5
●家族人数(19~24才)	0人	96	32 33.3	13 13.5	19 19.8	16 16.7	3 3.1	13 13.5
	1人	31	10 32.3	7 22.6	7 22.6	3 9.7	2 6.5	2 6.5
	2人以上	19	8 42.1	2 10.5	5 26.3	1 5.3	0 0.0	3 15.8
●家族人数(25~29才)	0人	127	47 37.0	17 13.4	29 22.8	15 11.8	3 2.4	16 12.6
	1人	14	3 21.4	3 21.4	2 14.3	4 28.6	1 7.1	1 7.1
	2人以上	5	0 0.0	2 40.0	0 0.0	1 20.0	1 20.0	1 20.0
●家族人数(30~39才)	0人	107	34 31.8	18 16.8	26 24.3	13 12.1	5 4.7	11 10.3
	1人	21	8 38.1	2 9.5	2 9.5	5 23.8	0 0.0	4 19.0
	2人以上	18	8 44.4	2 11.1	3 16.7	2 11.1	0 0.0	3 16.7
●家族人数(40才代)	0人	56	21 37.5	10 17.9	7 12.5	8 14.3	2 3.6	8 14.3
	1人	40	16 40.0	7 17.5	7 17.5	3 7.5	1 2.5	6 15.0
	2人以上	50	13 26.0	5 10.0	17 34.0	9 18.0	2 4.0	4 8.0
●家族人数(50才代)	0人	87	29 33.3	9 10.3	21 24.1	15 17.2	2 2.3	11 12.6
	1人	33	13 39.4	7 21.2	6 18.2	2 6.1	1 3.0	4 12.1
	2人以上	26	8 30.8	6 23.1	4 15.4	3 11.5	2 7.7	3 11.5
●家族人数(60才以上)	0人	117	40 34.2	17 14.5	28 23.9	16 13.7	4 3.4	12 10.3
	1人	19	6 31.6	3 15.8	3 15.8	3 15.8	1 5.3	3 15.8
	2人以上	10	4 40.0	2 20.0	0 0.0	1 10.0	0 0.0	3 30.0

表7-12 C5.0によるフェース分析 (ツリー)

オプション	
枝刈度	: 75%
枝あたり最小レコード	: 5
目的変数の件数 第1基準値引率: 55 第1基準その他 (値引き以外): 91	
20代後半の家族人数 <= 0	
主婦の職業 [その他, 自営業] (7.0, 0.857)	-> その他
主婦の職業 [事務, 技能, お勤めその他, 自由業] (19.0, 0.789)	-> 値引率
主婦の職業 専業主婦	
世帯収入 <= 1,000万円未満 (46.0, 0.717)	-> その他
世帯収入 > 1,000万円以上 (18.0, 0.667)	-> 値引率
主婦の職業 販売職・サービス職	
40代の家族人数 <= 1 (10.0, 0.6)	-> 値引率
40代の家族人数 > 1 (8.0, 1.0)	-> その他
20代後半の家族人数 > 0 (15.0, 0.8) -> 値引率	

表7-13 C5.0によるフェース分析 (ルール)

ルール: 買わない:	ルール: 買う:
ルール #1 : 0:	ルール #1 : 1:
if 20代後半の家族人数 <= 0	if 20代後半の家族人数 <= 0
and 主婦の職業 == [その他, 自由業]	and 主婦の職業 == [事務職, 技能職, お勤めその他, 自由業]
then-> その他 (7.0, 0.857)	then-> 値引率 (19.0, 0.789)
ルール #2 : 0:	ルール #2 : 1:
if 20代後半の家族人数 <= 0	if 20代後半の家族人数 <= 0
and 主婦の職業 == 専業主婦	and 主婦の職業 == 専業主婦
and 世帯収入 <= 1000万円未満	and 世帯収入 > 1,000万円以上
then-> その他 (46.0, 0.717)	then-> 値引率 (18.0, 0.667)
ルール #3 : 0:	ルール #3 : 1:
if 20代後半の家族人数 <= 0	if 20代後半の家族人数 <= 0
and 主婦の職業 == 販売職・サービス職	and 主婦の職業 == 販売職・サービス職
and 40代の家族人数 > 買う	and 40代の家族人数 <= 1
then-> その他 (8.0, 1.0)	then-> 値引率 (10.0, 0.6)
ルール #4 : 0:	デフォルト: -> 買わない
if 20代後半の家族人数 > 0	
then-> その他 (15.0, 0.8)	

7. 2. 3 グループごとの購買行動分析

次に、「値引率」、「メーカー」、「辛さ」、「クラス」、4つの志向ごとに購買行動がどう異なるかをみてみることにする。具体的には、パネルを第1基準から4つのグループに分け、各々のグループごとに決定木で分析をする。ここでは、パネルごとではなく4つのグループごとにデータを合算して行なう。

メーカーを第1基準にしたパネルは、メーカーによって購買パターンが異なるのだ

ろうか。例えば、「あるメーカーは、値引率が高くないと買わない」とか「あるメーカーでは、値引率は関係ない」といった結果がみられるのだろうか。

C5.0 のオプションは、「枝刈り度：75%」、「枝当たり最小レコード：25」とした。

(1) 値引率グループのツリー

このグループのツリー(表7-14)をみると、値引率で3分割(20.2%未満、20.2%~37.7%未満、37.7%以上)され、各々の枝によって見事に購買行動が異なっている。値引幅が小さければ買わないし、大きければ買う。20.2%~37.7%の値引率では、クラスや容量、メーカーによってパターンが異なってくる。

表7-14 C5.0による値引率グループの購買行動分析の結果(ツリー)

目的変数の件数	買った商品：636	買わない商品：1961
値引率 < -20.2		
値引率 < -36.7 (179.0, 0.698)		→ 買う
値引率 > -36.7		
クラス 定番・一般		
値引率 < -32.3		
メーカー メーカーA (30.0, 0.733)		→ 買う
メーカー その他 (96.0, 0.688)		→ 買わない
値引率 > -32.3 (483.0, 0.758)		→ 買わない
クラス [高級・本格、中級・後発]		
容量 200G未満 (181.0, 0.58)		→ 買う
容量 200G以上		
メーカー メーカーB (34.0, 0.618)		→ 買わない
メーカー その他 (91.0, 0.549)		→ 買う
値引率 > -20.2 (1503.0, 0.884)		→ 買わない

表7-15 C5.0による値引率グループの購買行動分析の結果(ルール)

<p>ルール: 買わない:</p> <p>ルール #1 : 0:</p> <p>if 値引率 > -36.7</p> <p>and 値引率 <= -20.2</p> <p>and クラス == 定番・一般</p> <p>and メーカー == [メーカー-B, P-B, メーカー-C]</p> <p>then-> 買わない (96.0, 0.688)</p> <p>ルール #2 : 0:</p> <p>if 値引率 > -32.3</p> <p>and 値引率 <= -20.2</p> <p>and クラス == 定番・一般</p> <p>then-> 買わない (483.0, 0.758)</p> <p>ルール #3 : 0:</p> <p>if 値引率 > -36.7</p> <p>and 値引率 <= -20.2</p> <p>and クラス == [高級・本格, 中級・後発]</p> <p>and 容量 = 200g以上</p> <p>and メーカー == メーカー-A</p> <p>then-> 買わない (34.0, 0.618)</p> <p>ルール #4 : 0:</p> <p>if 値引率 > -20.2</p> <p>then-> 買わない (1503.0, 0.884)</p>	<p>ルール #1 : 1:</p> <p>if 値引率 <= -20.2</p> <p>then-> 買う (179.0, 0.698)</p> <p>ルール #2 : 1:</p> <p>if 値引率 > -36.7</p> <p>and 値引率 <= -20.2</p> <p>and クラス == 定番・一般</p> <p>and メーカー == メーカー-A</p> <p>then-> 買う (30.0, 0.733)</p> <p>ルール #3 : 1:</p> <p>if 値引率 > -36.7</p> <p>and 値引率 <= -20.2</p> <p>and クラス == [高級・本格, 中級・後発]</p> <p>and 容量 = 200g未満</p> <p>then-> 買う (181.0, 0.58)</p> <p>ルール #4 : 1:</p> <p>if 値引率 > -36.7</p> <p>and 値引率 <= -20.2</p> <p>and クラス == [高級・本格, 中級・後発]</p> <p>and 容量 = 200g以上</p> <p>and メーカー == [メーカー-B, P-B, メーカー-C]</p> <p>then-> 買う (91.0, 0.549)</p>
<p>ルール: 買う:</p>	<p>デフォルト: -> 買わない</p>

(2) メーカーグループのツリー

メーカーグループのツリー(表7-16)は、値引率グループと異なりメーカーそのものが第1基準にはならなかった。よく考えてみると、このグループに属するパネル全員がある特定のメーカーに偏っていない限り、最初にメーカーで枝分かれすることはないだろう。このグループの第1基準がメーカーにならなかったということは、特定のメーカーに人気は集中してはいないことを表わしている。

表7-16 C5.0によるメーカーグループの購買行動分析の結果(ツリー)

<p>目的変数件数 買った商品: 306 買わない商品: 840</p> <p>容量 200g未満</p> <p>値引率 <= -32.9 (71.0, 0.521) -> 買う</p> <p>値引率 > -32.9 (864.0, 0.767) -> 買わない</p> <p>容量 200g以上</p> <p>メーカー メーカー-B (26.0, 0.731) -> 買う</p> <p>メーカー [P-B, メーカー-C] (105.0, 0.857) -> 買わない</p> <p>メーカー メーカー-A</p> <p>値引率 <= -4.7 (51.0, 0.588) -> 買う</p> <p>値引率 > -4.7 (29.0, 0.862) -> 買わない</p>
--

表7-17 C5.0によるメーカーグループの購買行動分析の結果(ルール)

<p>ルール: 買わない:</p> <p>ルール #1 : 0:</p> <p>if 容量 == 200g 未満 and 値引率 > -32.9 then-> 買わない (864.0, 0.767)</p> <p>ルール #2 : 0:</p> <p>if 容量 == 200g 以上 and メーカー == [P.B, メーカー-C] then-> 買わない (105.0, 0.857)</p> <p>ルール #3 : 0:</p> <p>if 容量 == 200g 以上 and メーカー == メーカー-A and 値引率 > -4.7 then-> 買わない (29.0, 0.862)</p> <p>ルール: 買う:</p>	<p>ルール #1 : 1:</p> <p>if 容量 == 200g 未満 and 値引率 <= -32.9 then-> 買う (71.0, 0.521)</p> <p>ルール #2 : 1:</p> <p>if 容量 == 200g 以上 and メーカー == メーカー-B then-> 買う (26.0, 0.731)</p> <p>ルール #3 : 1:</p> <p>if 容量 == 200g 以上 and メーカー == メーカー-A and 値引率 <= -4.7 then-> 買う (51.0, 0.588)</p> <p>デフォルト : -> 買わない</p>
---	--

(3) 辛さグループのツリー

このグループのツリー(表7-18)における1番の特徴は、階層が深くないと辛さがでてこないことである。分岐のパターンとしては、値引率グループのツリーに似ている。まず、第1基準、第2基準とも値引率になっていて、3分割できる。分岐の値もほぼ同じ21.7%と37.2%となっている。さらに、21.7%と37.2%の間の値引率では、クラスで購買行動が異なるところまで同じである。この結果からは、甘口派と辛口派では、価格感応性などにあまり差がないと考えられる。

辛さのカテゴリー分けには、もう少し異なるアプローチもあった。今回の辛さの基準は、商品名についている辛口や甘口のラベルをもとにカテゴリー分けをした。実際は、ある商品の甘口がもう一方の商品の辛口よりよっぽど辛いことがある。さらに厳密な分析を行なうためには、商品パッケージについている辛さ表示よりも実質的な辛さを基準に商品を選別する方法があり得る。

表7-18 C5.0による辛さグループの購買行動分析の結果(ツリー)

<p>目的変数件数 買った商品 : 430 買わない商品 : 921</p> <p>値引率 <= -21.7</p> <p>値引率 <= -37.2 (107.0, 0.617) -> 買う</p> <p>値引率 > -37.2</p> <p>クラス 定番・一般 (323.0, 0.656) -> 買わない</p> <p>クラス [高級・本格, 中級・後発]</p> <p>メーカー P.B (0.0, 1.0) -> 買わない</p> <p>メーカー [メーカー-A, メーカー-B]</p> <p>辛さ [甘口, 辛口] (55.0, 0.745) -> 買う</p> <p>辛さ 中辛 (49.0, 0.653) -> 買わない</p> <p>メーカー メーカー-C</p> <p>辛さ [甘口, 中辛] (45.0, 0.733) -> 買わない</p> <p>辛さ 中辛 (40.0, 0.525) -> 買う</p> <p>値引率 > -21.7 (732.0, 0.779) -> 買わない</p>

表7-19 C5.0による辛さグループの購買行動分析の結果 (ルール)

<pre> ルール: 買わない: ルール #1 : 0: if 値引率 > -37.2 and 値引率 <= -21.7 and クラス == 定番・一般 then-> 買わない (323.0, 0.656) ルール #2 : 0: if 値引率 > -37.2 and 値引率 <= -21.7 and クラス == [高級・本格, 中級・後発] and メーカー == [メーカーA, メーカーB] and 辛さ == 中辛 then-> 買わない (49.0, 0.653) ルール #3 : 0: if 値引率 > -37.2 and 値引率 <= -21.7 and クラス == [高級・本格, 中級・後発] and メーカー == メーカーC and 辛さ == [甘口, 辛口] then-> 買わない (45.0, 0.733) ルール #4 : 0: if 値引率 > -21.7 </pre>	<pre> then-> 買わない (732.0, 0.779) ルール: 買う: ルール #1 : 1: if 値引率 <= -21.7 then-> 買う (107.0, 0.617) ルール #2 : 1: if 値引率 > -37.2 and 値引率 <= -21.7 and クラス == [高級・本格, 中級・後発] and メーカー == [メーカーA, メーカーB] and 辛さ == [甘口, 辛口] then-> 買う (55.0, 0.745) ルール #3 : 1: if 値引率 > -37.2 and 値引率 <= -21.7 and クラス == [高級・本格, 中級・後発] and メーカー == メーカーC and 辛さ == 中辛 then-> 買う (40.0, 0.525) デフォルト: -> 買わない </pre>
--	---

(4) クラスグループのツリー

このグループの第1基準も値引率になっているが、他のグループと比べると、分岐の値は小さく比較的価格弾力性は低い。5.4%以上の値引きをすると、クラスによって購買行動が異なってくる。中級・後発クラスでは、辛さや値引率などの組み合わせによって購入比率が異なる。

表7-20 C5.0によるクラスグループの購買行動分析の結果 (ツリー)

<pre> 目的変数件数 買った商品: 273 買わない商品: 604 値引率 <= -5.4 クラス 定番・一般 (358.0, 0.735) -> 買わない クラス 高級・本格 (12.0, 0.833) -> 買う クラス 中級・後発 辛さ [甘口, 辛口] (123.0, 0.715) -> 買わない 辛さ 中辛 値引率 <= -14.6 (90.0, 0.711) -> 買う 値引率 > -14.6 (137.0, 0.613) -> 買わない 値引率 > -5.4 (157.0, 0.898) -> 買わない </pre>

表7-21 C5.0によるクラスグループの購買行動分析の結果(ルール)

ルール: 買わない:	and クラス = 中級・後発
ルール #1 : 0:	and 辛さ = 中辛
if 値引率 < -5.4	then-> 買わない (137.0, 0.613)
and クラス = 定番・一般	
then-> 買わない (358.0, 0.735)	
ルール #2 : 0:	ルール #4 : 0:
if 値引率 < -5.4	if 値引率 > -5.4
and クラス = 中級・後発	then-> 買わない (157.0, 0.898)
and 辛さ = [甘口, 辛口]	
then-> 買わない (123.0, 0.715)	
ルール #3 : 0:	ルール: 買う:
if 値引率 > -14.6	ルール #1 : 1:
and 値引率 < -5.4	if 値引率 < -5.4
	and クラス = 中級・後発
	and 辛さ = 中辛
	then-> 買う (90.0, 0.711)
	デフォルト: -> 買わない

(5) 分析全体の特徴

第1の特徴は、メーカーグループを除いて、値引率が第1基準となったことだ。これには、①クラスや辛さなどの属性は、パネルごとに購入志向は異なるが、購入比率と線形関係になりやすい値引率がどうしても第1基準になってしまう、②尺度の多い方が枝分かれがしやすい、などが考えられる。②に関しては、値引率をカテゴライズして再度分析してみる必要がある。

第2の特徴は、値引率と辛さグループは価格感応度が高く、クラスグループの価格感応度は相対的に低いことだ。

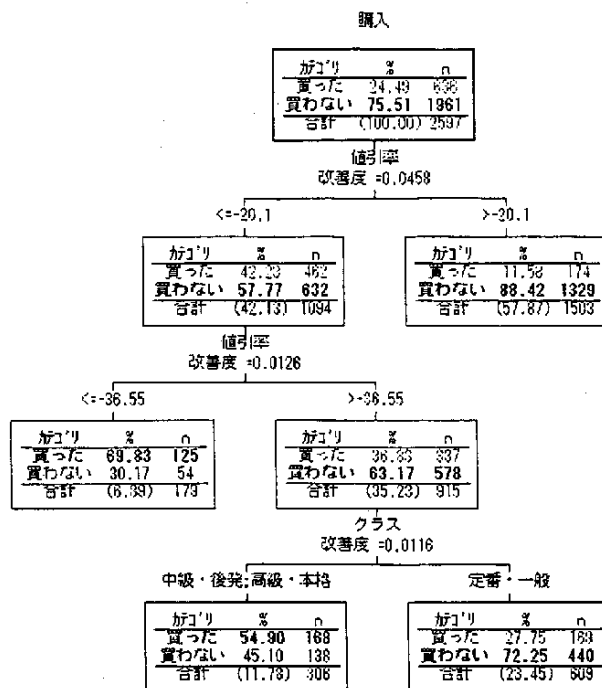
7.3 手法の比較

ここでは、グループごとのツリーをC&RTとCHAIDを用いて作成してみる。C&RTもCHAIDもC5.0と同様、決定木に分類される手法である。3つの手法の相違点は主に、分岐の指標に求められる。CHAIDは、カイ2乗値の有意確率が枝分かれの指標となる。C&RTは、Gini係数と呼ばれる指標によって枝を分ける。C5.0は、利得基準比となっている。CHAID、C&RTとも、ツリーを作成するときの停止基準は、「親ノード: 50」、「子ノード: 25」で設定した。

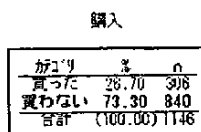
まず、C&RTとC5.0の4つツリーを比較してみると、値引率グループのツリー(C5.0: 表7-14、C&RT: 図7-2)とクラスグループのツリー(C5.0は、表7-20、C&RT: 図7-3)は、各々よく似ている。枝の生え方だけでなく、値引率の分岐の値もほぼ同じになっている。

C&RTのツリーは、かなり枝が刈られてしまっている。メーカーグループは、結局全ての枝が刈られてしまった。

1) 値引きグループ



2) メーカーグループ



3) 辛さグループ

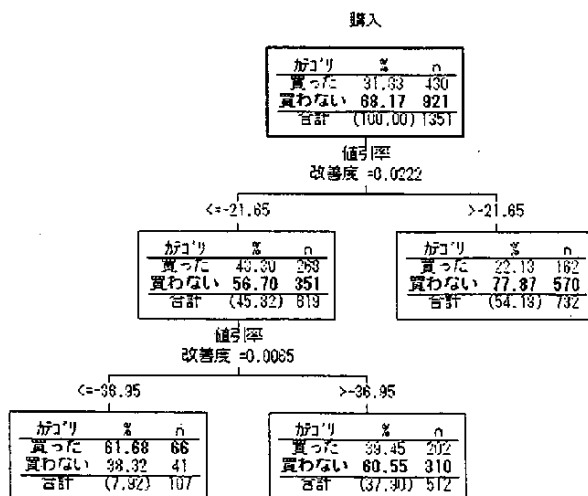


図7-2 C&RTによる購買行動分析の結果1

4) クラスグループ

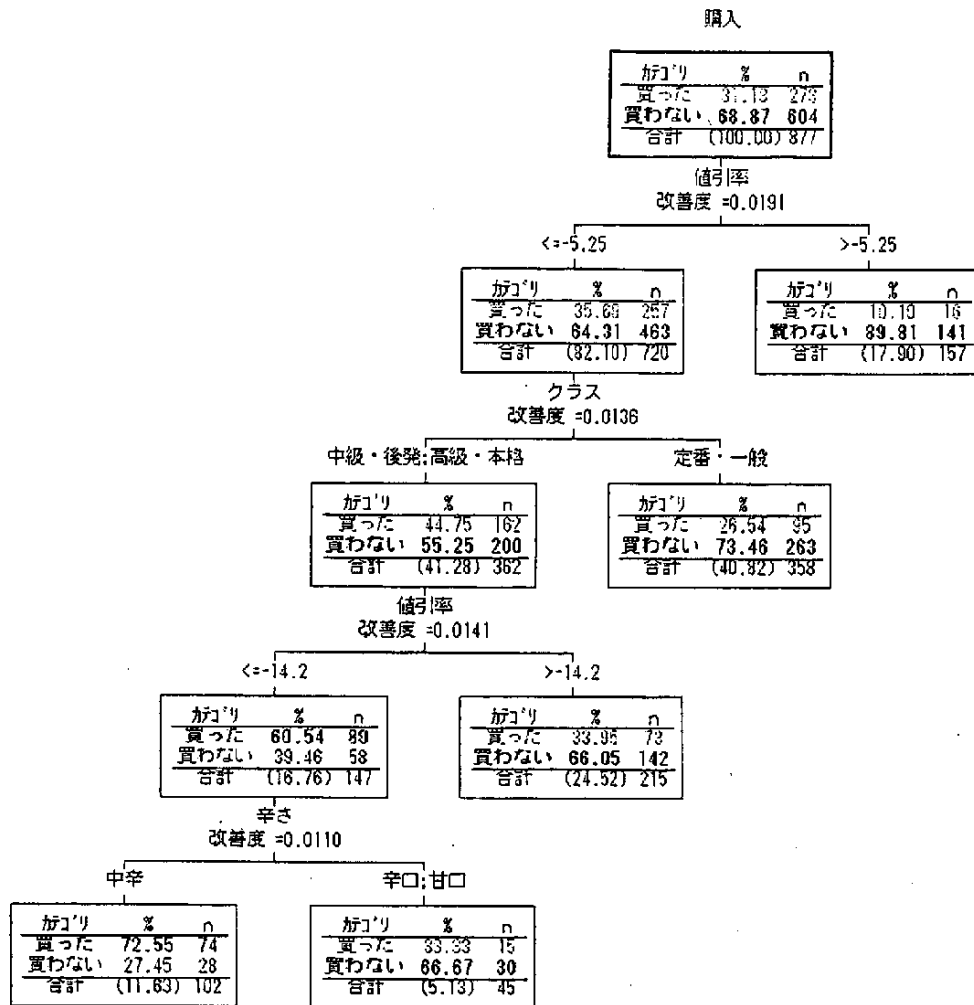


図7-3 C&RTによる購買行動分析の結果2

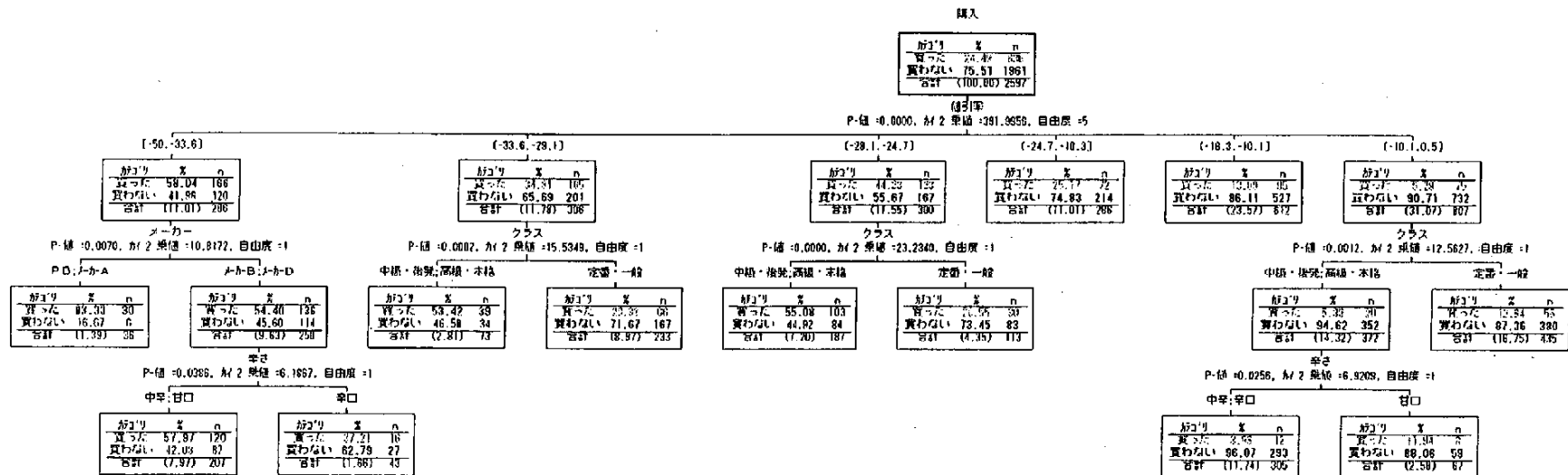


図 7-4 CHAID による値引率グループの購買行動分析の結果

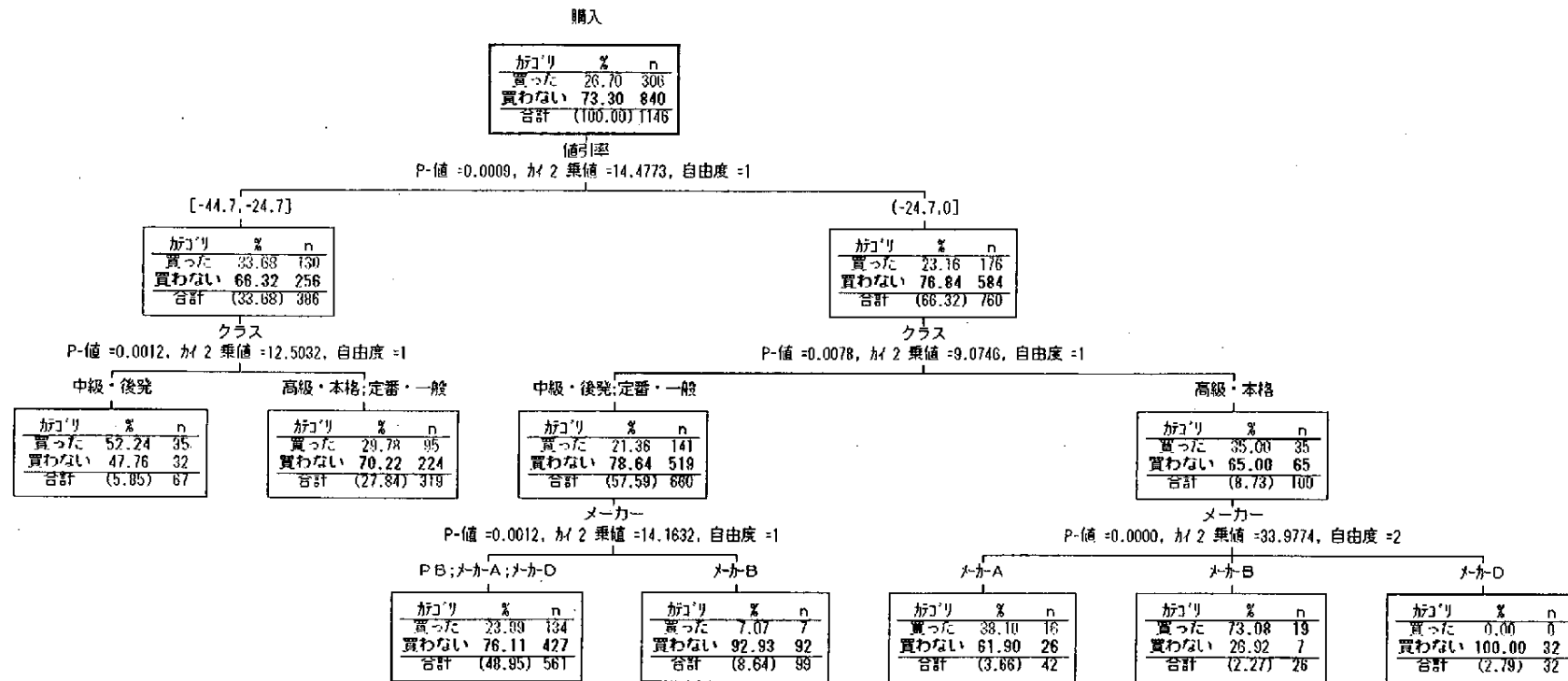


図7-5 CHAIDによるメーカーグループの購買行動分析の結果

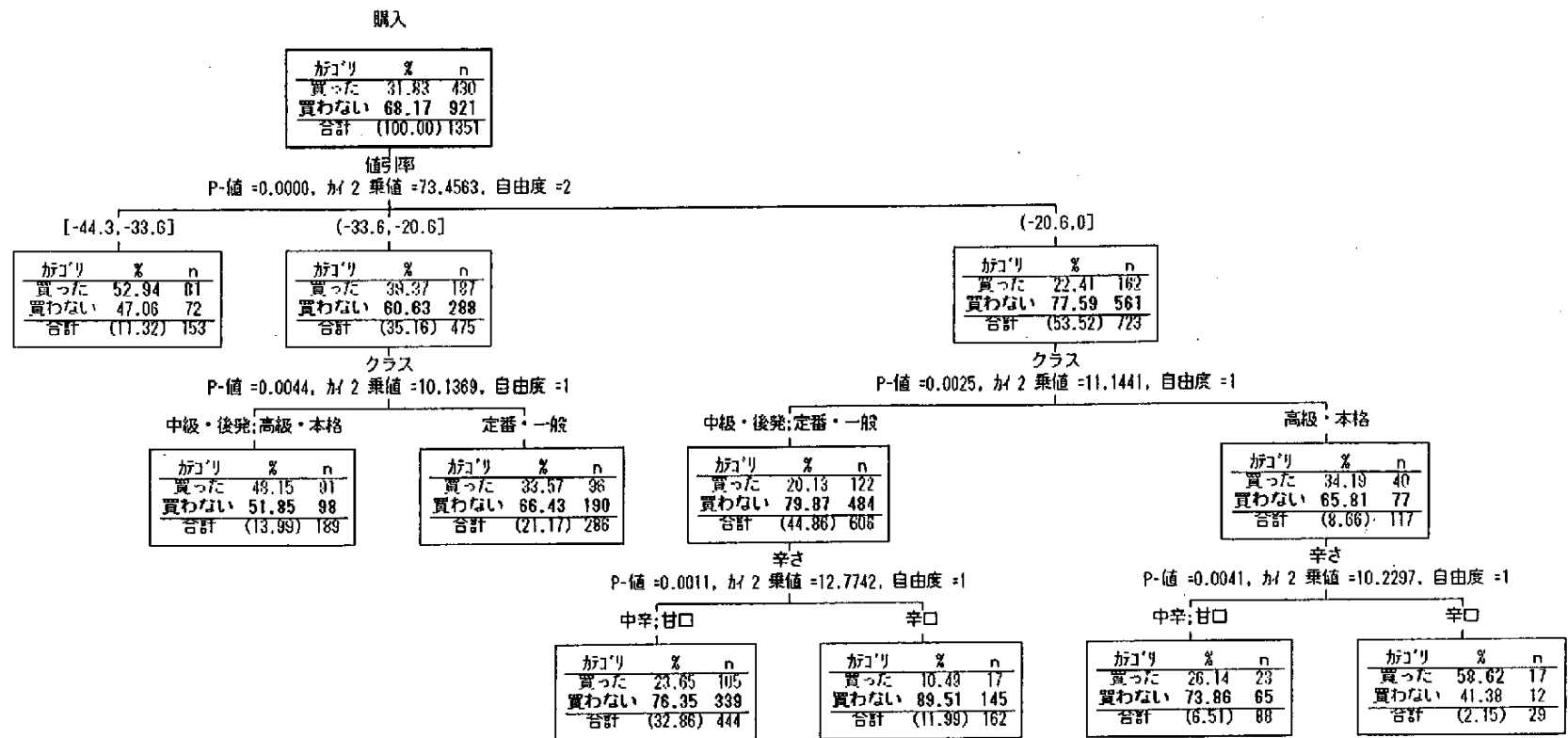


図7-6 CHAIDによる辛さグループの購買行動分析の結果

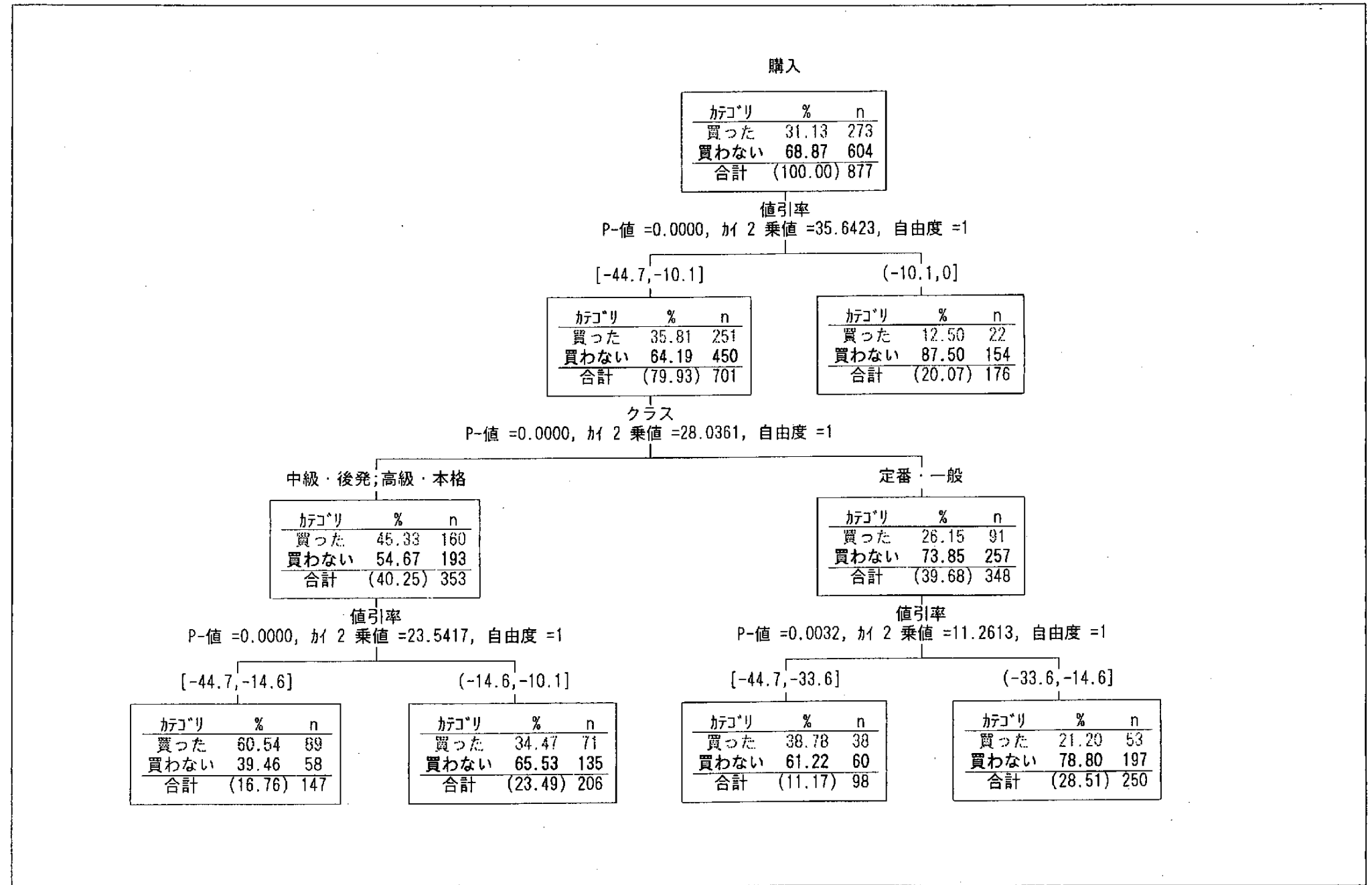


図7-7 CHAIDによるクラスグループの購買行動分析の結果

CHAIDとC5.0を比べると、辛さグループのツリー（C5.0：表7-18、CHAID：図7-6）とクラスグループ（C5.0：表7-20、CHAID：図7-7）のツリーに関しては、比較的似ている。CHAID自体の特徴としては、どのグループも第1分岐が値引率で、第2分岐がクラスになっている。

3つの手法を比較してみて、クラスグループのツリーは、どの手法をみてもかなり安定している。逆に、メーカーグループのツリーは、非常に不安定な結果となっている。

7.4 検討課題

複数の商品のうちどれを購入したいかといった質問は、アンケート調査ではよく行なわれる。この分析は、アンケートデータの代わりに購買データを用いて行なったものとなる。購買データの利点は、実際に買ったという事実に基づいていることだ。ある商品が好きだという気持ちとは裏腹に、異なる商品を買ってしまった経験は誰にでもあるだろう。一方、購買データの欠点は、データの形式が必ずしも分析に適合していないことである。また、比較したい情報がないこともある。そういった状況のなかで、分析はいくつかの工夫をしていった。それ故、課題も多い。以下、主な課題点について述べてみたいと思う。

(1) 値引率の推定の問題

買わなかった商品の値引き率の推定手順は、2段階に別れていた。①店舗データから同日付の販売価格を採用する。②同日に実販売履歴のない商品は、その前後に販売された価格のうち高い方を採用する。①は、適切な処理だろう。②に関しては考察の必要がある。②の処理をしないと、データは、欠損値になる。調べてみると、買わなかった商品データ（4,328）のうち、8.8%（382商品）が②の方法によって推定されていた。決定木の場合、説明変数が欠損値となっても分析は可能だが、値引率が欠損値のとき必ず買わないという状況になってしまう。

(2) 説明変数に単一のカテゴリー値しかない場合

例えば、甘口のカレーしか買っていない場合、第1基準に辛さは選ばれない。なぜなら、1つのカテゴリーしかない属性は、買ったときと買わなかったときの差を分析できないからである。だが、そのパネルの第1基準は辛さとも考えられる。

そのように考えるなら、今回のような分析手法によって第1基準を作成する前に、事前に購入商品の属性のうち単一カテゴリーがないか調べる必要がある。

8. パネルデータによるバスケット分析

分析目的～バスケット分析

アメリカのスーパーマーケット「ウォルマート」で、「紙おむつを買う人はビールを買う」という分析結果から、紙おむつとビールを同じ売場に隣接させて販売したところ、売上が上昇したという例がある。

もちろん、日本のスーパーマーケットで、紙おむつとビールを隣接させるなどということは、特設売場による販売ぐらいでしか例を見ない。しかし、一見関係のないように見える商品同士でも、1回の購買行動中で一緒に買われる（同じかごの中に入る）商品の組み合わせルールから新たな販売戦略に結びつくような発見をすることがある。このような同時購買（併買）パターン分析は、マーケットバスケット分析と呼ばれ、以前からデータマイニングの手法の一つとして実施されている。

様々な条件（店舗・客・プロモーション、他）の下で、どのような併買パターンがあるのかを知りたいときにバスケット分析は有効である。

分析結果は特に小売業にとって密接なデータとなる。店舗内のレイアウト計画／製品のバンドル販売／片方の商品しか買わなかったときに、もう片方のクーポンを提供するなど、さまざまな店内プロモーションへの指標となりえる。

また、メーカーサイドにとっても、自社製品をエンドに置くための小売への売場・棚割提案材料となる他、自社商品にない商品カテゴリーとの関連から新市場開拓にもつながる。

今回は、実際のパネルデータを適用して〔商品Aを購入するとき、同時に商品Bを購入する〕というルールを見いだすバスケット分析を行った事例報告を行う。ソースデータには大規模データである日本経済新聞社のパネル購買データ「NEEDS-SCAN/PANEL」を用いた。商品分類間および商品間での併買状況を発掘することを目的とし、多方向からの分析を試みた。

8. 1 分析方法概要

8. 1. 1 データ作成

(1) データの内容

出 所：日本経済新聞社の購買パネルデータベース

「NEEDS-SCAN/PANEL」（第2章参照）

対象店舗：首都圏AチェーンB店の買い物パネル

対象期間：1998年11月1日～1999年10月31日（1年分）

表8-1 バスケット分析ローデータ

IDナン バー	購入日	時間	購入点数	購買金額	店頭販売 価格	平均価 格	システム コード	世代 コード	スキャンコード	商品名	大分類番号	小分類番号
11051213	1998/10/3	10:42	1	298	298	298	1	1	4901003900112	石井 おべんとクン ミートボール 120GX3	33	12
11051213	1998/10/3	10:42	1	198	198	228	1	1	4901116002970	旭フーズ レンジでできたて カレー男爵6個180G	252	5
11051213	1998/10/3	10:42	1	148	148	148	1	1	4901670054224	サランラップ クックバーの紙カップ 8A 白40枚	704	32
11051213	1998/10/3	10:42	1	548	548	548	1	1	4901872887859	資生堂 ミュウ センターイン夜用羽付買得20個X2	608	1
11051213	1998/10/3	10:42	1	188	188	218	1	1	4901901270409	マルハ パールエース印 上白糖 1KG	111	1
11051213	1998/10/3	10:42	1	298	298	298	1	1	4901987200819	東洋アルミフードウェルペーパーホイル30CMX5M	704	3
11051213	1998/10/27	12:14	1	198	198	258	1	1	4901990044431	マルちゃん生ラーメンみそ3人前スープ付120GX3	29	1
11051213	1998/10/27	12:14	1	198	198	198	1	1	4902106970415	ミツカン ほんてり みりん風調味料 500ML	117	2
11051213	1998/10/27	12:14	1	148	148	253	1	1	4902402048696	ハウス バーモントカレー 甘口1 250G	131	1
11051213	1998/10/27	12:14	1	148	148	178	1	1	4902705048584	明治乳 無脂肪乳もっとCa 乳飲料 紙パック 1L	47	1
11051213	1998/10/27	12:14	1	178	178	178	1	1	4903110020097	山崎 テーブルロール 黒糖入り 増量 1+7個	162	3
11051213	1998/10/27	12:14	1	100	100	100	1	1	4903332010630	桑陽 ホーネン横浜名物シウマイポーク12個168G	27	1
11051213	1998/10/27	12:14	1	175	175	198	1	1	4970050033531	一正 うずら卵ボール 魚肉ねり製品 5個	24	1
11051213	1998/10/27	12:14	1	100	100	100	1	1	4971661980818	嘉平屋 はんぺん 2枚	23	2
11057116	1999/4/14	15:55	1	228	228	228	1	1	4901013011143	アイク 塩ゆでえだまめ 400G	251	3
11057116	1999/4/14	15:55	2	136	68	68	1	1	4901810306800	PB 烏龍茶 缶 340G	76	1
11057116	1999/4/14	15:55	1	118	118	125	1	1	4902102020367	コカコーラ ファンタ オレンジ PET 500ML	78	14
11057116	1999/4/14	15:55	1	115	115	125	1	1	4902102020374	コカコーラ ファンタ グレープ PET 500ML	78	14
11057116	1999/4/14	15:55	1	198	198	228	1	1	4902130320668	ニチレイ お弁当にグッド えびカツ 4個 120G	252	9
11057116	1999/4/14	15:55	1	100	100	100	1	1	4902410207061	フジ バターロール 5個	162	3
11057116	1999/4/14	15:55	1	350	350	398	1	1	4902705058811	明治乳ブルーージェ苺シャベット&苺ミルク50MLX8	262	1
11057116	1999/4/14	15:55	1	148	148	198	1	1	4903050213962	雪印乳 毎日骨太ヨーグルト 95GX3	45	3
11057116	1999/4/14	15:55	1	198	198	248	1	1	4970838001172	あぶくま 杉樽一本漬 きゅうりのしょうゆ漬200G	11	5
11057116	1999/4/14	15:55	1	580	580	580	1	1	4971423070016	井上 Fフーズ 北海するめ 2枚	202	2
11057116	1999/4/14	15:55	1	158	158	198	1	1	4974911880012	ダイサン水野 味鮮華 特天 えび 140G	24	1
11057116	1999/4/14	15:55	1	278	278	348	2	2	4946757400000	キッコーマン 新撰焼肉 赤と黒 プラ 210G	128	1
11057116	1999/4/16	14:24	2	196	98	102	1	1	3068320007397	エビアンエビアンNミネラルウォーターPET330ML	85	4
11057116	1999/4/16	14:24	1	278	278	278	1	1	4901002024246	SB 味付 あらびき塩こしょう 180G	125	3

※購買金額=店頭販売金額×購入点数を表す

※小分類は大分類番号3桁+小分類番号3桁で表す

(2) データの加工

ローデータは、表8-1にあるように、あるパネラーが1回の買い物中で購入した一商品分の購入行動が1レコードとなる。したがって、パネラーが1回の買い物で商品を5点購入していれば、ローデータのレコードは5件となる。今回抽出した1年分の購買ローデータのレコード数は1,359,254レコードにのぼった。

このローデータを、買い物1回単位の購買データ、つまりレシート単位でのデータに変換したものが表8-2となる。1レコードはIDナンバー11085217のパネラーの7月16日に買い物したレシートの内訳が1レコードとなる。(1日に2回買い物した場合は2レコード。)バスケット分析はこの表8-2の形式のデータを使って行う。レシートデータの収録項目は下記の通り。

- ・パネラーの ID ナンバー
- ・購入年月日
- ・1年間の購入回数の合計
- ・商品分類別購入フラグ →1回の買い物中で購入点数に関わらず、
買っていたら「1」、買っていなかったら「0」

表8-2 分析用レシートデータ

ID ナンバー	購入年月日	購入回数 合計	商品分類別購入フラグ※										
			X001_001	X001_002	X002_001	X076_004	X077_004	X078_014	X131_001	~	X301_002	X301_003	
11085217	1999/7/16	54	1	1	0	0	0	0	0	0	~	0	0
11085415	1999/8/31	64	0	0	1	0	0	0	0	0	~	0	0
11086914	1999/7/15	5	0	1	0	0	0	0	0	0	~	0	0
11087712	1999/6/29	57	0	1	1	0	1	0	0	0	~	0	0
11094810	1999/6/20	16	0	1	0	0	0	0	0	0	~	0	0
11099715	1999/6/20	82	0	1	0	0	0	0	0	0	~	0	0
11102115	1999/7/17	28	0	0	0	0	0	0	0	0	~	0	0
11106519	1999/8/8	4	0	1	1	0	0	0	0	0	~	0	0
11109912	1999/6/22	40	0	0	0	0	0	0	0	0	~	0	0
11110711	1999/8/22	19	0	1	0	0	0	0	0	0	~	0	1
11111612	1999/7/19	57	0	0	0	0	0	0	0	0	~	0	0
11112913	1999/8/22	23	0	0	0	0	0	0	0	0	~	0	1
11114217	1999/8/17	7	0	1	0	0	0	0	0	0	~	0	0
11115214	1999/8/1	21	0	1	0	0	1	0	0	0	~	0	0
11124315	1999/6/28	28	0	0	0	0	0	0	0	1	~	0	0

※X001_001は小分類を表す。ここでの小分類=表8-1の大分類番号3桁+小分類番号3桁で表す。

なお、表8-2の購入フラグは小分類単位である。小分類については第2章の日経POSデータの説明を参照いただきたいが、データは大分類単位、小分類単位、商品単位のいずれにも作成することが可能である。使う側のマーケティングテーマに合致したデータ作成を行うことが重要である。

(3) データの絞り込み

当初は1年間分のデータで分析を試みようとした。しかし、商品単位で分析をしようとする、約14万レシート×約12万商品という、非常に膨大なデータの作成、分析を行わなくてはならない。また、小分類単位にしても、スタンドアローンのWindows NTではデータが膨大すぎて、データを読む段階でのメモリー不足、ハングアップを引き起こす危険性がある。

そのため、今回は

- ①購入フラグを小分類単位にした
 - ②家庭用品分類を削除し、食品の分類に限定した
 - ③出現率の低い分類（年間12件未満の分類）を削除した
 - ④併買の相関を見る分析については3ヶ月分のデータに限定した
- 上記①～④の条件を絞り込みしたデータでの分析に切り替えた。

今回用いた1年間分および3ヶ月分のバスケット分析用データサイズは以下の通り。

表8-3 バスケット分析用データサイズ

	1年間データ	夏期3ヶ月データ
データ対象期間	1998年10月1日～ 1999年9月30日	1999年6月1日～ 8月31日
レコード数（＝レシートの数）	139,010レコード	34,866レコード
対象商品分類（小分類）の数※	940分類	736分類
データ容量	約103MB	約59MB

8. 1. 2 分析手法

(1) 使用ソフトウェア

データ加工、および分析の一部はSASインスティテュートジャパン株式会社「SAS System」、バスケット併買分析はエス・ピー・エス・エス株式会社「Clementine」を使用した。

(2) 分析手法

Clementineには、バスケット分析等で用いる相関アルゴリズムとして「GRI」と「Apriori」の2つがあるが、今回は「Apriori」を採用した。(*1)

(3) Apriori と同時購買信頼度について

同時購買の確率（以下、信頼度）の算出は下記のような簡単な手順による。

(1対1分類の場合)

		商品 B	
		買っていない	買った
商品 A	買っていない	a	c
	買った	b	d

Aを買ったとき、Bを同時に買うとされる信頼度 $=d/(c+d)$

Bを買ったとき、Aを同時に買うとされる信頼度 $=d/(b+d)$

まず上記のように縦軸が商品Aの購買状況、横軸が商品Bの購買状況のマトリクスを描いてみる。

AもBも同時に買われていない件数 $= a$
Aは買ったがBは買われなかった件数 $= b$
Aは買われなかったがBは買った件数 $= c$
AとBを同時に買った件数 $= d$ とすると、

Aを買ったとき、Bを同時に買うとされる信頼度 $= d/(c+d)$ 、
Bを買ったとき、Aを同時に買うとされる信頼度 $= d/(b+d)$ となる。
商品数が少ない場合は、電卓1個でも併買の信頼度が計算できる。

例えばカレーと福神漬の関係ならば

		福神漬	
		買っていない	買った
カレー	買っていない	500	150
	買った	300	100

カレーを買ったら福神漬も買う信頼度 $= 100 / (100+150) = 0.40$ (40%)
福神漬を買ったらカレーも買う信頼度 $= 100 / (500+300) = 0.125$ (12.5%)となる。

(*1) GRI は数値データとシンボルデータ (カテゴリー値、フラグ値) の両方を入力し、シンボルデータでルールが出力されるが、Apriori はシンボルデータのみ扱う。

今回のような入力変数、出力変数 (目的変数) がいずれもフラグ値の場合は、Apriori の方がモデル生成の時間が短時間で済む (GRI ではパソコンの容量が不足し、ハングアップしてしまうこともしばしばあった) ため、Apriori を用いた。

Clementine で Apriori を実施する際には、

- ・ 併買商品の組み合わせ数の上限
(Clementine 上では「最大ルール前提条件」という)
- ・ 出力された相関ルールの信頼度の下限 (「最小ルール正確度」という)
- ・ 併買商品の出現ボリュームの下限 (「最小ルール範囲」という)

を、目的に応じて設定することができる。

今回の分析では、Apriori の条件を

- ・ 最大ルール前提条件=最大 5 分類まで、
- ・ 最小ルール範囲=最小 0.2%以上 (約 70 件) は固定し、
- ・ 最小ルール正確度 (信頼度) の下限のみを出現状況に応じて調整した (最小は 5%以上)。

最大ルール前提条件については、1対1の関係だけでは信頼度の出力が少なかったことと、2つ以上の組み合わせから新たな相関ルールを見いだすため、最大数を5分類までとした。はじめは分類を1分類対1分類で見る方がわかりやすい。

これらの数値をどのくらいにするのが適当であるかは目的によって異なる。予測していた組み合わせが現れない場合は、条件設定を緩める必要がある。逆にいくら信頼度が高くても、その商品自体があまり売れていないものであれば、クローズアップしても期待するほどの売上効果は得られない。そのようなルールを重視する必要性は低いと判断する。

なお、Apriori の出力結果の見方は下記の通り。

目的商品A <= 併買商品 B(100,17.0%,0.25)の場合

① ② ③

①併買商品出現数：併買商品Bの出現数

(2商品、あるいは2分類以上の組み合わせの時は、その組み合わせの出現数)

②併買商品出現率：①の全レシート数中の割合

③同時購買の信頼度：前述(3)を参照

※同時購買件数の値は①×③となる。

解釈：Bを買ったときにAを買う現象は全レシートの17%にあたる100レシート中25%の信頼度(*2)で発生している

(*2)以降、本文中の信頼度の数値は、上記③の%値を100倍した値で表記する。

8. 2 データ分析

ここから実際に、夏期3ヶ月データを用いて、いくつかのテーマから分析を行った結果について報告する。

8. 2. 1 全体的な併買パターン

3ヶ月データの分類同士での併買の高さをみた(表8-4a)。

最も信頼度が高い併買は〔「中華そば用つゆ」を買った時、「生中華そば」を買う〕組み合わせで、併売された信頼度(以下%表示の数値はすべて信頼度を表す)は67.3%(*3)とかなり高い。

次いで〔「納豆」と「生鮮農産物」と「鶏卵」を買ったら「絹ごし豆腐・ソフト豆腐」を買う〕組み合わせ(信頼度62.6%)、「粒納豆」と「生鮮農産物」と「ハードヨーグルト」を買ったら〔「絹ごし豆腐・ソフト豆腐」を買う〕組み合わせ(同60%)が高い。

日配食品以外では、〔「紙パック入り果汁入り清涼飲料」と「紙パック入り無果汁清涼飲料」〕(同58.5%)と、飲料同士の組み合わせが挙げられている。

日配食品は出現頻度が高いため、結果的に上位に並ぶ結果になってしまった。そこで、日配食品以外の組み合わせを探すべく、出現頻度が極端に高い日配品7分類(絹ごし豆腐、粒納豆、一般牛乳、普通食パン、菓子パン、生鮮農産物、生鮮水産物)を目的商品指定から外し、再分析すると(表8-4b)、併買の組み合わせにいくつかのパターンが見られた。

(*3) Clementine・Apriori の出力表上の注意

ここでこれ以降 Apriori によって出力された併買パターンの表について一点お断りしておく。実際に各分類の出現件数をカウントすると、最も信頼度が高く出ている生中華そば-中華そば用つゆの生中華そば543レシート(出現率1.56%)、中華そば用つゆ98レシート(出現率0.28%)、同時併買数は66件である。

しかし、表8-3aの Clementine による出力では(105, 0.3%, 0.673)と出力され、信頼度の値は一致するが、併買商品出現数が、併買商品出現率を小数点第1位の値に揃えられた値から再計算された値になっている。そのため、それぞれの実際の出現数、出現率は異なっている。

表8-4a 3ヶ月データ併買状況(上位のもののみ抜粋)

生中華そば<中華そば用つゆ ②	(105:0.3%, 0.673)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & 鶏卵	(105:0.3%, 0.626)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & ハードヨーグルト	(70:0.2%, 0.6)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & はんぺん	(105:0.3%, 0.596)
菓子パン <= 蒸しパン & さつま揚げ・揚げかまぼこ	(70:0.2%, 0.595)
絹ごし豆腐・ソフト豆腐 <= 油揚げ & キムチ・朝鮮漬	(70:0.2%, 0.588)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮水産物 & 油揚げ	(139:0.4%, 0.587)
紙パック入り果汁入り清涼飲料 <= 紙パック入り無果汁清涼飲料 ③	(139:0.4%, 0.585)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 普通食パン & 鶏卵	(105:0.3%, 0.574)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生ちくわ・焼きちくわ	(209:0.6%, 0.568)
菓子パン <= デニッシュ・ペストリー & フルーツ入りヨーグルト	(174:0.5%, 0.554)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & 即席カップ中華そば	(70:0.2%, 0.553)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 一般牛乳 & 油揚げ	(105:0.3%, 0.551)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮水産物 & ウインナソーセージ	(105:0.3%, 0.549)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 冷凍お好み焼き・たこ焼き	(70:0.2%, 0.543)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮水産物 & 即席カップ中華そば	(70:0.2%, 0.542)
絹ごし豆腐・ソフト豆腐 <= 油揚げ & 鶏卵	(139:0.4%, 0.542)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮水産物 & その他漬物	(70:0.2%, 0.541)
粒納豆 <= 油揚げ & キムチ・朝鮮漬	(70:0.2%, 0.541)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 釜めし・炊き込みごはんの素	(70:0.2%, 0.541)
菓子パン <= デニッシュ・ペストリー & 紙パック入り果汁混合野菜ジュース	(70:0.2%, 0.539)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & 油揚げ	(139:0.4%, 0.538)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮水産物 & 鶏卵	(105:0.3%, 0.537)
絹ごし豆腐・ソフト豆腐 <= ロールパン・コッペパン & 油揚げ	(174:0.5%, 0.536)
菓子パン <= 普通食パン & 生鮮農産物 & デニッシュ・ペストリー	(70:0.2%, 0.535)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 普通食パン & 油揚げ	(139:0.4%, 0.532)
絹ごし豆腐・ソフト豆腐 <= 菓子パン & 粒納豆 & ハードヨーグルト	(70:0.2%, 0.532)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & チルド中華総菜の素	(70:0.2%, 0.532)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & ロールパン・コッペパン	(105:0.3%, 0.529)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & ホームタイプマルチアイス & 即席カップ中華そば	(105:0.3%, 0.527)
絹ごし豆腐・ソフト豆腐 <= 菓子パン & 粒納豆 & スライスロースハム	(70:0.2%, 0.526)
菓子パン <= デニッシュ・ペストリー & さつま揚げ・揚げかまぼこ	(105:0.3%, 0.526)
菓子パン <= 粒納豆 & ドーナツ	(139:0.4%, 0.525)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 味付けノリ・切りノリ	(70:0.2%, 0.524)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & ホームタイプマルチアイス & 油揚げ	(70:0.2%, 0.523)
菓子パン <= ウインナソーセージ & ドーナツ	(70:0.2%, 0.523)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 油揚げ	(663:1.9%, 0.523)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & ロールパン・コッペパン & ウインナソーセージ	(105:0.3%, 0.522)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & ウインナソーセージ	(105:0.3%, 0.521)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 紙パック入り果汁混合野菜ジュース	(139:0.4%, 0.52)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & だし入りみそ	(70:0.2%, 0.519)
菓子パン <= ポテトチップス & 蒸しパン	(105:0.3%, 0.518)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生中華そば	(105:0.3%, 0.518)
菓子パン <= 絹ごし豆腐・ソフト豆腐 & 普通食パン & デニッシュ・ペストリー	(105:0.3%, 0.515)
絹ごし豆腐・ソフト豆腐 <= 普通食パン & 一般牛乳 & 油揚げ	(70:0.2%, 0.514)
粒納豆 <= 絹ごし豆腐・ソフト豆腐 & ウインナソーセージ & 油揚げ	(70:0.2%, 0.514)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & ビスケット・クッキー	(70:0.2%, 0.514)
菓子パン <= 蒸しパン & 乳酸菌飲料	(70:0.2%, 0.514)
菓子パン <= デニッシュ・ペストリー & 冷凍パスタ	(70:0.2%, 0.514)
菓子パン <= デニッシュ・ペストリー & 冷凍米飯	(70:0.2%, 0.514)
菓子パン <= 一般牛乳 & 生鮮農産物 & デニッシュ・ペストリー	(70:0.2%, 0.513)
菓子パン <= 粒納豆 & 普通食パン & デニッシュ・ペストリー	(70:0.2%, 0.513)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮農産物 & その他漬物	(70:0.2%, 0.513)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 生鮮水産物 & フルーツ入りヨーグルト	(70:0.2%, 0.512)
絹ごし豆腐・ソフト豆腐 <= 油揚げ & マヨネーズ・マヨネーズタイプ調味料	(105:0.3%, 0.511)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 一般牛乳 & 生鮮農産物	(209:0.6%, 0.509)
絹ごし豆腐・ソフト豆腐 <= 即席カップ中華そば & 鶏卵	(105:0.3%, 0.509)
絹ごし豆腐・ソフト豆腐 <= 菓子パン & 粒納豆 & 油揚げ	(139:0.4%, 0.508)
菓子パン <= 蒸しパン & ドーナツ	(139:0.4%, 0.507)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & がんもどき	(70:0.2%, 0.507)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 冷凍フライドチキン	(70:0.2%, 0.507)
絹ごし豆腐・ソフト豆腐 <= 粒納豆 & 紙パック入り紅茶飲料	(70:0.2%, 0.507)
菓子パン <= 粒納豆 & 一般牛乳 & デニッシュ・ペストリー	(105:0.3%, 0.506)

表8-4b 3ヶ月データ併買状況 (除日配食品7分類)

缶入り無果汁炭酸飲料 <= 菓子パン & 缶入りコーラ飲料 ③	(70:0.2%, 0.466)
生中華そば<=めん用の具 ②	(105:0.3%, 0.452)
即席カップ中華そば <= 即席カップうどん ③	(174:0.5%, 0.449)
スパゲティ・ロングパスタ <= パスタソースミックス・ソース ②	(174:0.5%, 0.447)
即席カップ中華そば <= 粒納豆 & 即席カップ焼きそば ③	(139:0.4%, 0.434)
即席カップうどん <= 即席カップ日本そば ③	(139:0.4%, 0.418)
スパゲティ・ロングパスタ <= 生鮮農産物 & パスタソース缶詰 ②	(105:0.3%, 0.413)
缶入りコーラ飲料 <= 缶入り無果汁炭酸飲料 ③	(209:0.6%, 0.411)
カレールー<=ホームタイプマルチアイス & 福神漬 ②	(70:0.2%, 0.41)
即席カップ中華そば <= ホームタイプマルチアイス & 即席カップ焼きそば ③	(70:0.2%, 0.41)
即席カップ中華そば <= 絹ごし豆腐・ソフト豆腐 & 即席カップ焼きそば ③	(139:0.4%, 0.403)
ーロタイププロセスチーズ <= スモークプロセスチーズ ③	(70:0.2%, 0.402)
ホームタイプマルチアイス <= 普通食パン & 一般牛乳 & 即席カップ中華そば	(70:0.2%, 0.395)
ホームタイプマルチアイス <= 絹ごし豆腐・ソフト豆腐 & 冷凍お好み焼き・たこ焼き	(105:0.3%, 0.393)
生中華そば<=メンマ ②	(105:0.3%, 0.393)
スパゲティ・ロングパスタ <= 菓子パン & パスタソース缶詰 ②	(105:0.3%, 0.389)
紙パック入り果汁入り清涼飲料 <= 紙パック入りコーヒー飲料 & 紙パック入り紅茶飲料 ③	(70:0.2%, 0.386)
糸コンニャク・白滝 <= 焼き豆腐 ②	(139:0.4%, 0.386)
ホームタイプマルチアイス <= ハードヨーグルト & カレールー	(70:0.2%, 0.383)
即席カップ中華そば <= 即席カップ日本そば ③	(139:0.4%, 0.381)
ホームタイプマルチアイス <= 牛乳タイプ乳飲料 & 冷凍米飯	(70:0.2%, 0.38)
シュレッドナチュラルチーズ <= ピザソース・ピザトーストスプレッド ②	(105:0.3%, 0.379)
ポテトチップス<= 牛乳タイプ乳飲料 & 小麦系スナック ②③	(70:0.2%, 0.372)
ポテトチップス<= クッキー・ビスケット & ジャガイモ系スナック ②③	(70:0.2%, 0.371)
パスタソース缶詰 <= ウインナーソーセージ & スパゲティ・ロングパスタ ②	(105:0.3%, 0.365)
紙パック入りコーヒー飲料 <= 缶入りコーラ飲料 & 缶入り無果汁炭酸飲料 ③	(70:0.2%, 0.361)
ポテトチップス<= ビスケット・クッキー & コーン系スナック ③	(139:0.4%, 0.359)
即席カップ中華そば <= 一般牛乳 & 即席カップ焼きそば ③	(105:0.3%, 0.358)
ポテトチップス<= せんべい & コーン系スナック ③	(105:0.3%, 0.356)
紙パック入り果汁入り清涼飲料 <= 紙パック入り健康茶飲料 ③	(105:0.3%, 0.353)
スライスロースハム <= ロールパン・コッペパン & 生冷やし中華そば ②	(70:0.2%, 0.347)
ミカン缶詰<=絹ごし豆腐・ソフト豆腐 & バイナップル缶詰 ③	(105:0.3%, 0.347)
スパゲティ・ロングパスタ <= 普通食パン & パスタソース缶詰 ②	(105:0.3%, 0.344)
スパゲティ・ロングパスタ <= 粒納豆 & パスタソース缶詰 ②	(105:0.3%, 0.343)
ホームタイプマルチアイス <= 粒納豆 & 冷凍お好み焼き・たこ焼き	(70:0.2%, 0.343)
小麦系スナック<= ポテトチップス & 冷凍用ポリドリンク ②③	(70:0.2%, 0.342)
ウインナーソーセージ <= ロールパン・コッペパン & ペットボトル入り果汁入り清涼飲料 ②	(70:0.2%, 0.342)
ポテトチップス<= チョコレート菓子 & コーン系スナック ③	(139:0.4%, 0.341)
常温ゼリー<=菓子パン & その他常温デザート類 ③	(70:0.2%, 0.341)
スライスロースハム <= 一般牛乳 & 生冷やし中華そば ②	(174:0.5%, 0.339)
即席カップ中華そば <= 即席カップ焼きそば ③	(628:1.8%, 0.339)
ホームタイプマルチアイス <= 菓子パン & 普通食パン & 即席カップ中華そば	(70:0.2%, 0.338)
ビスケット・クッキー <= 粒納豆 & 菓子パイ ③	(70:0.2%, 0.338)
油揚げ <=粒納豆 & 生鮮水産物 & 木綿豆腐	(70:0.2%, 0.337)
即席カップ中華そば <= 生鮮農産物 & 即席カップ焼きそば ③	(105:0.3%, 0.337)
カレールー<=普通食パン & 福神漬 ②	(105:0.3%, 0.337)
ビスケット・クッキー <= 絹ごし豆腐・ソフト豆腐 & 菓子パイ ③	(105:0.3%, 0.337)
即席カップ日本そば <= 即席カップうどん ③	(174:0.5%, 0.335)
ポテトチップス<= チョコバー・粒チョコレート & コーン系スナック ③	(105:0.3%, 0.333)
ホームタイプマルチアイス <= 牛乳タイプ乳飲料 & ペットボトル入りスポーツ飲料	(70:0.2%, 0.333)
スパゲティ・ロングパスタ <= パスタソース缶詰 ②	(523:1.5%, 0.331)
ビスケット・クッキー <= チョコバー・粒チョコレート & 小麦系スナック ③	(105:0.3%, 0.33)
ポテトチップス<= ウインナーソーセージ & 小麦系スナック ③	(105:0.3%, 0.33)
スパゲティ・ロングパスタ <= 一般牛乳 & パスタソース缶詰 ②	(105:0.3%, 0.33)
チルドプリン<= 絹ごし豆腐・ソフト豆腐 & チルドコーヒーゼリー ③	(105:0.3%, 0.33)
ホームタイプマルチアイス <= 絹ごし豆腐・ソフト豆腐 & 粒納豆 & ペットボトルコーラ飲料	(70:0.2%, 0.329)
ホームタイプマルチアイス <= 絹ごし豆腐・ソフト豆腐 & 冷凍ミックス野菜	(70:0.2%, 0.329)
即席カップ中華そば <= 菓子パン & 即席カップ焼きそば ③	(174:0.5%, 0.327)
ホームタイプマルチアイス <= 粒納豆 & 生鮮水産物 & 即席カップ中華そば	(70:0.2%, 0.325)
ポテトチップス<= ホームタイプマルチアイス & コーン系スナック ③	(105:0.3%, 0.325)
ビスケット・クッキー <= 普通食パン & 菓子パイ ③	(70:0.2%, 0.325)
ビスケット・クッキー <= 菓子パン & 菓子パイ ③	(70:0.2%, 0.325)
チルドプリン<= 粒納豆 & チルドコーヒーゼリー ③	(70:0.2%, 0.32)

表8-4b 3ヶ月データ併買状況(除日配食品7分類) 続き

ビスケット・クッキー <= せんべい & チョコバー・粒チョコレート ③	(139:0.4%, 0.319)
ポテトチップス<= 普通食パン & チョコレートがけプレッツェル ③	(105:0.3%, 0.319)
カレールー<=菓子パン & 福神漬 ②	(105:0.3%, 0.319)
スライスロースハム <= デニッシュ・ペストリー & 生冷やし中華そば ②	(70:0.2%, 0.316)
ポテトチップス<= 小麦系スナック & ペットボトル入り果汁入り清涼飲料 ②③	(70:0.2%, 0.315)
ホームタイプマルチアイス <= 菓子パン & 冷凍お好み焼き・たこ焼き ②	(105:0.3%, 0.315)
ポテトチップス<= ビスケット・クッキー & チョコレートがけプレッツェル ③	(105:0.3%, 0.314)
ホームタイプマルチアイス <= その他ドライフルーツ	(70:0.2%, 0.314)
ビスケット・クッキー <= せんべい & 小麦系スナック ③	(139:0.4%, 0.313)
ポテトチップス<= チョコレート菓子 & 小麦系スナック ③	(139:0.4%, 0.312)
ホームタイプマルチアイス <= ビスケット・クッキー & 冷凍パスタ	(70:0.2%, 0.312)
ビスケット・クッキー <= ホームタイプマルチアイス & チョコレートがけプレッツェル ③	(70:0.2%, 0.311)
ホームタイプマルチアイス <= 即席カップ中華そば & コーヒー乳飲料	(70:0.2%, 0.311)
ホームタイプマルチアイス <= 絹ごし豆腐・ソフト豆腐 & 生鮮農産物 & ポテトチップス	(105:0.3%, 0.31)
ホームタイプマルチアイス <= 絹ごし豆腐・ソフト豆腐 & 一般牛乳 & ポテトチップス	(70:0.2%, 0.31)
ビスケット・クッキー <= せんべい & あられ・おかき ③	(70:0.2%, 0.31)
ホームタイプマルチアイス <= ハードヨーグルト & 冷凍パスタ	(70:0.2%, 0.31)
チルドプリン<= 生鮮農産物 & チルドコーヒーゼリー ③	(70:0.2%, 0.31)
フルーツ入りヨーグルト <= 生鮮農産物 & チルドコーヒーゼリー ③	(70:0.2%, 0.31)
ホームタイプマルチアイス <= 粒納豆 & スライスプロセスチーズ	(105:0.3%, 0.309)
ホームタイプマルチアイス <= ポテトチップス & 冷凍米飯	(70:0.2%, 0.308)
スパゲティ・ロングパスタ <= レトルトパスタソース ②	(139:0.4%, 0.308)
ウインナーソーセージ <= ロールパン・コッペパン & 生冷やし中華そば	(70:0.2%, 0.307)
カレールー<=福神漬 ②	(523:1.5%, 0.307)
スライスロースハム <= 調味料・食用油ギフトセット	(70:0.2%, 0.307)
ポテトチップス<= デニッシュ・ペストリー & 小麦系スナック ③	(70:0.2%, 0.306)
フルーツ入りヨーグルト <= 一般牛乳 & チルドコーヒーゼリー ③	(70:0.2%, 0.306)
ポテトチップス<= 菓子パン & コーン系スナック ③	(209:0.6%, 0.306)
ホームタイプマルチアイス <= 絹ごし豆腐・ソフト豆腐 & 生鮮水産物 & ポテトチップス	(105:0.3%, 0.305)
紙パック入りコーヒー飲料 <= 牛乳タイプ乳飲料 & 紙パック入り果汁100%飲料 ③	(105:0.3%, 0.305)

※ 出力条件:最小ルール範囲0.2%以上、信頼度30%以上、組み合わせルール数5分類以内

※ ○がこみの数字は下記のパターンの番号を示す。ただしパターン①は記していない。

※ 表8-4bは、非日配食品分類のうち、8-4a出現分は除く。

パターン① 日配品分類同士の組み合わせ

絹ごし豆腐、納豆、一般牛乳、生鮮農産物、生鮮水産物、普通食パン

パターン② メニュー型商品(主食+具、調味料あるいはつけ合わせ)の組み合わせ

生中華そば - 中華そば用つゆ - めん用の具、メンマ
 スパゲティ - 缶入りスパゲティソース - ウインナーソーセージ
 カレールー - 福神漬
 生冷やし中華そば - ハム
 (おやつ)スナック菓子 - 飲料
 せんべい - クッキー・ビスケット - スナック菓子 など

パターン③ 同類商品分類間での組み合わせ

缶入りコーラ飲料 - 缶入り無果汁炭酸飲料
 紙パック果汁入り清涼飲料 - 紙パック無果汁清涼飲料
 チーズスプレッド - 調理スプレッド
 カップ入り即席めん同士(中華そば-日本そば-うどん)
 他、プロセスチーズ同士、スナック菓子同士 など

8. 2. 2 コーザルデータを追加した分析

曜日、特売日、天候、チラシ、エンド陳列の有無など、売れ行きの要因データであるコーザルデータを追加することで、より深い分析が可能となる。

8. 2. 1 で用いた3ヶ月データに、曜日変数（月～日の各曜日分をフラグ化+土日）を「休日」フラグとして設定）、および月変数（6月、7月、8月）を加え、8. 2. 1 と同じ条件で Apriori で分析した結果が表8-5である。

表中のものは曜日変数が現れたものだけを掲載しているが、この表から以下、

- ①特に火曜日フラグの出現が多い、また8月の火曜日に「粒納豆」と「絹ごし豆腐」など、日配品が多く併買されている
→その店舗が火曜日を特売曜日にしていることがわかる
- ②7月の火曜日に「即席カップ焼きそば」を買った人の62.5%が「即席カップ中華そば」を購入している
- ③「ポテトチップ」と「コーン系スナック」、あるいは「小麦系スナック」は休日に購入される
- ④8月に「マルチタイプのホームサイズアイス」と冷凍食品の併買が多い。その中でさらに「冷凍お好み焼き・たこ焼き」は7月にも併買が見られる
- ⑤「カレールー」と「福神漬」は曜日に関係なく併買されている

等のルールを読みとることができる。

表8-5 日付データによる分析

絹ごし豆腐・ソフト豆腐<=AUGUST&TUE&粒納豆&ホームタイプマルチアイス ①	(70:0.2% 0.705)
絹ごし豆腐・ソフト豆腐<=AUGUST&TUE&粒納豆&生鮮農産物 ①	(70:0.2% 0.643)
絹ごし豆腐・ソフト豆腐<=AUGUST&粒納豆&生ちくわ・焼きちくわ	(70:0.2% 0.637)
絹ごし豆腐・ソフト豆腐<=AUGUST&TUE&粒納豆&油揚げ	(105:0.3% 0.636)
絹ごし豆腐・ソフト豆腐<=TUE&はんぺん	(105:0.3% 0.618)
絹ごし豆腐・ソフト豆腐<=TUE&粒納豆&ホームタイプマルチアイス	(209:0.6% 0.611)
絹ごし豆腐・ソフト豆腐<=AUGUST&TUE&粒納豆&普通食パン ①	(105:0.3% 0.609)
絹ごし豆腐・ソフト豆腐<=TUE&ホームタイプマルチアイス&油揚げ	(105:0.3% 0.602)
絹ごし豆腐・ソフト豆腐<=JULY&TUE&粒納豆&油揚げ	(70:0.2% 0.595)
絹ごし豆腐・ソフト豆腐<=AUGUST&ホームタイプマルチアイス&油揚げ	(105:0.3% 0.592)
粒納豆<=JUNE&TUE&油揚げ&木綿豆腐	(70:0.2% 0.575)
粒納豆<=JUNE&絹ごし豆腐・ソフト豆腐&TUE&木綿豆腐	(70:0.2% 0.554)
粒納豆<=JUNE&TUE&普通食パン&油揚げ	(105:0.3% 0.549)
粒納豆<=絹ごし豆腐・ソフト豆腐&TUE&普通食パン&油揚げ	(105:0.3% 0.535)
粒納豆<=JUNE&TUE&普通食パン&木綿豆腐	(70:0.2% 0.532)
粒納豆<=JUNE&絹ごし豆腐・ソフト豆腐&TUE&油揚げ	(174:0.5% 0.525)
一般牛乳<=菓子パン&SAT&ウインナソーセージ	(70:0.2% 0.414)
一般牛乳<=JULY&絹ごし豆腐・ソフト豆腐&菓子パン&ホームタイプマルチアイス	(70:0.2% 0.413)
一般牛乳<=HOLIDAY&ホームタイプマルチアイス&紙パック入りコーヒー飲料	(70:0.2% 0.395)
一般牛乳<=HOLIDAY&菓子パン&さつま揚げ・揚げかまぼこ	(70:0.2% 0.388)
一般牛乳<=HOLIDAY&普通食パン&鶏卵	(70:0.2% 0.372)
一般牛乳<=生鮮農産物&ホームタイプマルチアイス&THU	(105:0.3% 0.37)
一般牛乳<=菓子パン&WED&フルーツ入りヨーグルト	(70:0.2% 0.364)
一般牛乳<=HOLIDAY&粒納豆&鶏卵	(105:0.3% 0.363)
一般牛乳<=JUNE&絹ごし豆腐・ソフト豆腐&生鮮農産物&WED	(70:0.2% 0.36)

表 8-5 日付データによる分析 (続き)

缶入りコーラ飲料<=HOLIDAY&缶入り無果汁炭酸飲料	(70:0.2% 0.467)
缶入りコーラ飲料<=AUGUST&缶入り無果汁炭酸飲料	(70:0.2% 0.439)
缶入り無果汁炭酸飲料<=HOLIDAY&缶入りコーラ飲料	(105:0.3% 0.35)
缶入り無果汁炭酸飲料<=AUGUST&缶入りコーラ飲料	(105:0.3% 0.343)
缶入り無果汁炭酸飲料<=JUNE&缶入りコーラ飲料	(70:0.2% 0.3)
紙パック入り果汁入り清涼飲料<=THU&紙パック入り紅茶飲料	(70:0.2% 0.37)
紙パック入り果汁入り清涼飲料<=JUNE&紙パック入り紅茶飲料	(105:0.3% 0.31)
スパゲティ・ロングパスタ<=SUN&パスタソース缶詰	(105:0.3% 0.419)
スパゲティ・ロングパスタ<=HOLIDAY&パスタソース缶詰	(139:0.4% 0.4)
スパゲティ・ロングパスタ<=JUNE&パスタソース缶詰	(139:0.4% 0.373)
スパゲティ・ロングパスタ<=JULY&パスタソース缶詰	(209:0.6% 0.336)
スパゲティ・ロングパスタ<=FRI&パスタソース缶詰	(70:0.2% 0.325)
カレールー<=SUN&福神漬 ⑤	(70:0.2% 0.36)
カレールー<=MON&福神漬 ⑤	(70:0.2% 0.359)
カレールー<=FRI&福神漬 ⑤	(105:0.3% 0.333)
カレールー<=TUE&福神漬 ⑤	(105:0.3% 0.322)
カレールー<=JULY&福神漬 ⑤	(174:0.5% 0.319)
カレールー<=HOLIDAY&福神漬 ⑤	(174:0.5% 0.308)
カレールー<=JUNE&福神漬 ⑤	(139:0.4% 0.301)
即席カップ中華そば<=JULY&TUE&即席カップ焼きそば ②	(70:0.2% 0.625)
即席カップ中華そば<=JULY&即席カップ焼きそば	(244:0.7% 0.409)
即席カップ中華そば<=TUE&即席カップ焼きそば	(139:0.4% 0.404)
即席カップ中華そば<=SUN&即席カップ焼きそば	(105:0.3% 0.337)
即席カップ中華そば<=FRI&即席カップ焼きそば	(105:0.3% 0.327)
即席カップ中華そば<=AUGUST&即席カップ焼きそば	(174:0.5% 0.322)
即席カップ中華そば<=AUGUST&THU&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.321)
即席カップ中華そば<=HOLIDAY&即席カップ焼きそば	(174:0.5% 0.316)
即席カップ中華そば<=AUGUST&THU&冷凍米飯	(70:0.2% 0.303)
即席カップ中華そば<=AUGUST&THU&即席袋中華そば	(105:0.3% 0.302)
ポテトチップス<=JUNE&ビスケット・クッキー&小麦系スナック ③	(70:0.2% 0.357)
ポテトチップス<=AUGUST&HOLIDAY&小麦系スナック ③	(70:0.2% 0.333)
ポテトチップス<=JUNE&SUN&コーン系スナック ③	(70:0.2% 0.324)
ポテトチップス<=SAT&チョコレートがけプレッツェル	(70:0.2% 0.321)
ポテトチップス<=AUGUST&絹ごし豆腐・ソフト豆腐&コーン系スナック ③	(70:0.2% 0.319)
ポテトチップス<=SAT&小麦系スナック ③	(139:0.4% 0.315)
ポテトチップス<=JUNE&HOLIDAY&コーン系スナック ③	(105:0.3% 0.306)
ポテトチップス<=JUNE&菓子パン&コーン系スナック ③	(105:0.3% 0.303)
ホームタイプマルチアイス<=AUGUST&THU&冷凍米飯 ④	(70:0.2% 0.382)
ホームタイプマルチアイス<=THU&コーヒー乳飲料	(105:0.3% 0.344)
ホームタイプマルチアイス<=HOLIDAY&ポテトチップス&小麦系スナック	(70:0.2% 0.338)
ホームタイプマルチアイス<=SUN&レギュラーコーヒー(粉)	(70:0.2% 0.333)
ホームタイプマルチアイス<=AUGUST&冷凍お好み焼き・たこ焼き ④	(139:0.4% 0.331)
ホームタイプマルチアイス<=JULY&HOLIDAY&粒納豆&普通食パン	(70:0.2% 0.329)
ホームタイプマルチアイス<=JULY&菓子パン&普通食パン&一般牛乳	(70:0.2% 0.324)
ホームタイプマルチアイス<=AUGUST&SUN&ビスケット・クッキー	(70:0.2% 0.321)
ホームタイプマルチアイス<=THU&冷凍米飯	(105:0.3% 0.32)
ホームタイプマルチアイス<=HOLIDAY&ポテトチップス&ビスケット・クッキー	(105:0.3% 0.316)
ホームタイプマルチアイス<=AUGUST&絹ごし豆腐・ソフト豆腐&THU&即席カップ中華そば	(70:0.2% 0.315)
ホームタイプマルチアイス<=HOLIDAY&生鮮農産物&キムチ・朝鮮漬	(105:0.3% 0.315)
ホームタイプマルチアイス<=JULY&絹ごし豆腐・ソフト豆腐&粒納豆&即席カップ中華そば	(70:0.2% 0.315)
ホームタイプマルチアイス<=JULY&一般牛乳&せんべい	(70:0.2% 0.315)
ホームタイプマルチアイス<=一般牛乳&生鮮農産物&THU	(105:0.3% 0.312)
ホームタイプマルチアイス<=AUGUST&絹ごし豆腐・ソフト豆腐&冷凍米飯 ④	(105:0.3% 0.309)
ホームタイプマルチアイス<=AUGUST&冷凍中華そば ④	(105:0.3% 0.309)
ホームタイプマルチアイス<=AUGUST&絹ごし豆腐・ソフト豆腐&コーヒー乳飲料	(70:0.2% 0.308)
ホームタイプマルチアイス<=JULY&冷凍お好み焼き・たこ焼き ④	(139:0.4% 0.304)
ホームタイプマルチアイス<=AUGUST&トマト缶詰	(70:0.2% 0.303)
ホームタイプマルチアイス<=絹ごし豆腐・ソフト豆腐&THU&即席カップ中華そば	(105:0.3% 0.301)

※ 出力条件は表 8-4a と同じ。

※ ○がこみの数字は上記のルール番号を示す

8. 2. 3. バーチャルアイテムを活用した分析

次に特定の商品分類に特化し、分類そのものだけではなく、ブランド、味などの特徴を表す変数—バーチャルアイテムを設定し、併買パターンの違いを追ってみた。本分析では、カレールーを採用した。今回使用している夏期3ヶ月のレシートデータを見ると、「カレールーを買った」レシートの数は1397レシートだった。1日当たり15～6個売れていることになる。夏にカレールーはどのような購買パターンを示すだろうか。

(1) カレールーとの併買分類

まずは、「カレールー」分類全体について Apriori を行った。〔併買商品〇〇を買ったら同時にカレールーを買う〕、この併買商品の組み合わせで最も高い信頼度が出たのは、〔「ホームタイプマルチアイス」&「福神漬」を買ったら「カレールー」を買う〕組み合わせであった(表8-6)。信頼度は約41.0%と、マルチアイスと福神漬を買った人の半分近くがカレールーと一緒に買っていることがわかる。以下、「普通食パン」「一般牛乳」「粒納豆」など日配食品と「福神漬」を買ったときに「カレールー」を買う組み合わせが続く。福神漬以外では、「カレー粉」「ケチャップ」「マヨネーズ」など、調味料との併買が多い。

逆に福神漬からみたバスケット分析を行うと、最も高いのは〔「絹ごし豆腐」&「普通食パン」&「カレールー」〕との併買(18.3%)。以下、〔「カレールー」&「粒納豆」&「一般牛乳」〕、〔「カレールー」&「ドレッシング」〕との併買と続き、〔「カレールー」だけとの併買〕の場合でも信頼度は約11.5%。のきなみカレールーが入ったかごに福神漬と一緒に入っていることが伺える。

前述した「カレールー」と「福神漬」の信頼度、11.5%という数字を低いと見るならば、カレールーが買い置き商品であると考え、値引きあるいはまとめ買いをさせる方を練る。逆に高いと見るならば、カレールー自体の購買を上げるために、福神漬と一緒に(あるいはじゃがいも、人参、玉ねぎなども一緒に)カレーメニューを意識させる売場の設置を検討することなども可能である。これらの数字を戦略的に活用することが望ましい。

(2) バーチャルアイテムによる併買の違い

次にカレールーを単品単位にする。出現数の少ないブランドを削除し、39商品に対して、以下の4つのバーチャルアイテム変数を設定した。(採用商品と4変数の区分については表8-7参照。)

- ①ブランド別：12のブランド変数を作成した。
- ②容量別：200G以上と200G未満のグループに分割した。
- ③辛さ別：カレールーのパッケージ裏面に記載のある辛さ順位に基づいてグループを5分割した。
- ④クラス(高級志向)別：高級感のあるものと、子供用のカレールー、それ以外のカレールーを「一般」として3クラスに分類した。

表8-6 カレールーの併買状況

カレールー<=ホームタイプマルチアイス & 福神漬	(70:0.2% 0.41)	カレールー<=生鮮農産物 & 即席カップ焼きそば	(105:0.3% 0.135)
カレールー<=普通食パン & 福神漬	(105:0.3% 0.337)	カレールー<=一般牛乳 & コーン缶詰	(70:0.2% 0.134)
カレールー<=菓子パン & 福神漬	(105:0.3% 0.319)	カレールー<=その他パウダースライス	(105:0.3% 0.134)
カレールー<=福神漬	(523:1.5% 0.307)	カレールー<=普通食パン & 一般牛乳 & スライスロースハム	(105:0.3% 0.134)
カレールー<=一般牛乳 & 福神漬	(105:0.3% 0.303)	カレールー<=普通食パン & ポーション型菓子(ゼリーなど)	(105:0.3% 0.134)
カレールー<=生鮮農産物 & 福神漬	(70:0.2% 0.3)	カレールー<=粒納豆 & パスタソース缶詰	(105:0.3% 0.133)
カレールー<=絹ごし豆腐・ソフト豆腐 & 福神漬	(139:0.4% 0.279)	カレールー<=即席カップ中華そば & チョコバー・粒チョコレート	(105:0.3% 0.133)
カレールー<=粒納豆 & 福神漬	(105:0.3% 0.215)	カレールー<=生鮮農産物 & 濃口しょうゆ	(105:0.3% 0.133)
カレールー<=絹ごし豆腐・ソフト豆腐 & ホームタイプマルチアイス & スライスロースハム	(70:0.2% 0.197)	カレールー<=絹ごし豆腐・ソフト豆腐 & 濃口しょうゆ	(174:0.5% 0.132)
カレールー<=カレー粉	(70:0.2% 0.192)	カレールー<=粒納豆 & ペットボトル入りコーヒー飲料	(105:0.3% 0.132)
カレールー<=マヨネーズ・マヨネーズタイプ調味料 & ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.188)		
カレールー<=ビスケット・クッキー & 生揚げ・厚揚げ	(70:0.2% 0.176)		
カレールー<=絹ごし豆腐・ソフト豆腐 & シュレッドナチュラルチーズ	(70:0.2% 0.165)	福神漬 <=絹ごし豆腐・ソフト豆腐 & 普通食パン & カレールー	(70:0.2% 0.183)
カレールー<=ビスケット・クッキー & マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.165)	福神漬 <=絹ごし豆腐・ソフト豆腐 & 一般牛乳 & カレールー	(70:0.2% 0.172)
カレールー<=絹ごし豆腐・ソフト豆腐 & バター	(105:0.3% 0.16)	福神漬 <=カレールー & マヨネーズ・マヨネーズタイプ調味料	(105:0.3% 0.172)
カレールー<=粒納豆 & トマトケチャップ	(105:0.3% 0.16)	福神漬 <=絹ごし豆腐・ソフト豆腐 & ラッキョウ漬	(105:0.3% 0.17)
カレールー<=一般牛乳 & マヨネーズ・マヨネーズタイプ調味料	(174:0.5% 0.16)	福神漬 <=カレールー & せんべい	(105:0.3% 0.169)
カレールー<=鶏卵 & 生焼きそば	(70:0.2% 0.16)	福神漬 <=ハードヨーグルト & カレールー	(70:0.2% 0.16)
カレールー<=ホームタイプマルチアイス & トマトケチャップ	(105:0.3% 0.16)	福神漬 <=鶏卵 & カレールー	(105:0.3% 0.159)
カレールー<=一般牛乳 & ホームタイプマルチアイス & 即席カップ中華そば	(70:0.2% 0.158)	福神漬 <=シソ漬	(105:0.3% 0.158)
カレールー<=粒納豆 & ロールパン・コッペパン & ビスケット・クッキー	(70:0.2% 0.157)	福神漬 <=ホームタイプマルチアイス & カレールー	(209:0.6% 0.152)
カレールー<=即席カップ中華そば & チルドギョーザ	(70:0.2% 0.157)	福神漬 <=ロールパン・コッペパン & カレールー	(139:0.4% 0.151)
カレールー<=普通食パン & 冷凍用ポリドリング	(105:0.3% 0.157)	福神漬 <=油揚げ & カレールー	(70:0.2% 0.143)
カレールー<=絹ごし豆腐・ソフト豆腐 & ホームタイプマルチアイス & フルーツ入りヨーグルト	(105:0.3% 0.156)	福神漬 <=菓子パン & カレールー	(279:0.8% 0.14)
カレールー<=一般牛乳 & トマトケチャップ	(105:0.3% 0.155)	福神漬 <=ラッキョウ漬	(314:0.9% 0.139)
カレールー<=一般牛乳 & 生鮮水産物 & 鶏卵	(70:0.2% 0.155)	福神漬 <=デニッシュ・ベストリー & カレールー	(70:0.2% 0.139)
カレールー<=即席カップ中華そば & チルドプリン	(70:0.2% 0.154)	福神漬 <=粒納豆 & 普通食パン & せんべい	(70:0.2% 0.129)
カレールー<=ビスケット・クッキー & その他漬物	(70:0.2% 0.151)	福神漬 <=スライスロースハム & カレールー	(105:0.3% 0.128)
カレールー<=粒納豆 & 冷凍ハンバーグ	(70:0.2% 0.149)	福神漬 <=一般牛乳 & カレールー	(244:0.7% 0.127)
カレールー<=スライスロースハム & ベーコン	(70:0.2% 0.149)	福神漬 <=普通食パン & カレールー	(279:0.8% 0.126)
カレールー<=フルーツ入りヨーグルト & せんべい	(105:0.3% 0.148)	福神漬 <=ポテトチップス & カレールー	(105:0.3% 0.126)
カレールー<=ホームタイプマルチアイス & ふりかけ	(105:0.3% 0.146)	福神漬 <=カレールー & ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.123)
カレールー<=絹ごし豆腐・ソフト豆腐 & ホームタイプマルチアイス & 鶏卵	(70:0.2% 0.145)	福神漬 <=絹ごし豆腐・ソフト豆腐 & カレールー	(314:0.9% 0.118)
カレールー<=ウインナーソーセージ & マヨネーズ・マヨネーズタイプ調味料	(105:0.3% 0.143)	福神漬 <=絹ごし豆腐・ソフト豆腐 & ホームタイプマルチアイス & 鶏卵	(70:0.2% 0.118)
カレールー<=ウインナーソーセージ & 冷凍用ポリドリング	(70:0.2% 0.143)	福神漬 <=カレールー	(1395:4.0% 0.115)
カレールー<=絹ごし豆腐・ソフト豆腐 & 粒納豆 & チルドギョーザ	(70:0.2% 0.141)	福神漬 <=菓子パン & 菓子パイ	(70:0.2% 0.112)
カレールー<=スライスロースハム & 小麦系スナック	(70:0.2% 0.141)	福神漬 <=生鮮農産物 & カレールー	(244:0.7% 0.106)
カレールー<=ビスケット・クッキー & ポーション型菓子(ゼリーなど)	(70:0.2% 0.141)	福神漬 <=絹ごし豆腐・ソフト豆腐 & 粒納豆 & ペットボトル入りコーラ飲料	(70:0.2% 0.105)
カレールー<=絹ごし豆腐・ソフト豆腐 & 味付けノリ・切りノリ	(105:0.3% 0.139)	福神漬 <=ビスケット・クッキー & カレールー	(105:0.3% 0.103)
カレールー<=即席カップ中華そば & 濃口しょうゆ	(70:0.2% 0.137)	福神漬 <=絹ごし豆腐・ソフト豆腐 & 菓子パイ	(105:0.3% 0.101)
カレールー<=即席カップ中華そば & つゆの素・めんつゆ(濃縮)	(70:0.2% 0.137)	福神漬 <=絹ごし豆腐・ソフト豆腐 & 菓子パン & 紙パック入り果汁入り清涼飲料	(105:0.3% 0.1)
カレールー<=即席カップ中華そば & ビスケット・クッキー	(174:0.5% 0.136)	福神漬 <=その他漬物 & カレールー	(70:0.2% 0.1)
カレールー<=普通食パン & 一般牛乳 & 即席カップ中華そば	(70:0.2% 0.136)	福神漬 <=紙パック入りコーヒー飲料 & 小麦系スナック	(70:0.2% 0.1)
カレールー<=絹ごし豆腐・ソフト豆腐 & 粒納豆 & うるち米	(70:0.2% 0.135)	福神漬 <=普通食パン & 菓子パイ	(70:0.2% 0.1)
カレールー<=絹ごし豆腐・ソフト豆腐 & トマトケチャップ	(139:0.4% 0.135)	福神漬 <=普通食パン & バター	(70:0.2% 0.099)
カレールー<=粒納豆 & ピラフ・チャーハンの素(ドライ)	(70:0.2% 0.135)	福神漬 <=紙パック入りコーヒー飲料 & カレールー	(105:0.3% 0.098)

表8-7 カレールー変数一覧

変数A : 商品	商品名	出現数	変数B : ブランド	変数C : 辛さ	変数D : 容量	変数E : クラス
A1	SB ゴールデンカレー 中辛 240G	551	1	3	2	1
A2	SB ゴールデンカレー 甘口 240G	298	1	1	2	1
A3	SB ゴールデンカレー 辛口 240G	211	1	5	2	1
A4	SB ディナーカレー 中辛 100G	203	2	4	1	2
A5	SB ディナーカレー 甘口 100G	23	2	2	1	2
A6	SB ディナーカレー 辛口 100G	122	2	5	1	2
A7	SB ゴールデンカレー 中辛 100G	6	1	3	1	1
A8	SB あら挽きカレー 中辛 200G	26	3	4	2	1
A9	SB あら挽きカレー 辛口 200G	23	3	5	2	1
A10	SB 妖精たちの大好きなカレー 甘口 160G	35	4	1	1	3
A11	SB 妖精たちの大好きなカレー 中辛 160G	41	4	2	1	3
A12	SB あらびきカレー 中辛 160G	118	3	4	1	1
A13	SB あらびきカレー 辛口 160G	68	3	5	1	1
A14	SB カレーの王子さま 80G	31	5	1	1	3
A15	グリコ 熟カレー 甘口 200G	97	6	1	2	1
A16	グリコ 熟カレー 中辛 200G	228	6	3	2	1
A17	グリコ 熟カレー 辛口 200G	32	6	4	2	1
A18	グリコ 極 カレールウ 甘口 155G	12	7	1	1	2
A19	グリコ 極 カレールウ 中辛 155G	50	7	3	1	2
A20	グリコ 極 カレールウ 辛口 155G	1	7	5	1	2
A21	PB カレー 辛口 240G	37	8	5	2	1
A22	PB カレールウ 中辛 125GX3	1	8	3	1	1
A23	ハウス ジャワカレー 中辛4 125G	42	9	4	1	1
A24	ハウス ジャワカレー 中辛4 220G	110	9	4	2	1
A25	ハウス ジャワカレー 辛口5 220G	87	9	5	2	1
A26	ハウス パーモントカレー 甘口1 125G	50	10	1	1	1
A27	ハウス パーモントカレー 甘口1 250G	543	10	1	2	1
A28	ハウス パーモントカレー 辛口3 250G	216	10	3	2	1
A29	ハウス パーモントカレー 中辛2 125G	30	10	2	1	1
A30	ハウス パーモントカレー 中辛2 250G	613	10	2	2	1
A31	ハウス こくまるカレー 中辛 100G	116	11	3	1	1
A32	ハウス こくまるカレー 中辛 200G	813	11	3	2	1
A33	ハウス こくまるカレー 辛口 200G	365	11	4	2	1
A34	ハウス ザ・カレー 甘口 140G	54	12	1	1	2
A35	ハウス ザ・カレー 中辛 140G	137	12	2	1	2
A36	ハウス ザ・カレー 辛口 140G	32	12	3	1	2
A37	ハウス ジャワカレー スパイシーブレンド 220G	34	9	5	2	1
A38	PB カレー 甘口 240G	83	8	1	2	1
A39	PB カレー 中辛 240G	81	8	3	2	1

ブランド変数内訳

B1 ゴールデン	B7 極カレー
B2 ディナーカレー	B8 PBカレー
B3 あらびきカレー	B9 ジャワカレー
B4 妖精たち	B10 パーモントカレー
B5 カレー王子さま	B11 こくまるカレー
B6 熟カレー	B12 ハウス ザ・カレー

辛さ変数内訳

C1 甘口
C2 やや甘口
C3 中辛
C4 やや辛口
C5 辛口

容量変数内訳

D1 200G以上
D2 200G未満

クラス (高級志向) 変数内訳

E1 一般
E2 高級
E3 子供

Apriori 出力条件は

- ・最大ルール前提条件=最大5分類まで、
- ・最小ルール範囲=最小0.2%以上(約70件)、
- ・最小ルール正確度(信頼度)=5%以上とした。

(3) 考察

① ブランド別(表8-8)

もともとのレシート数に対し、ブランド別に変数を細分化した事で出現数がかなり小さいものになるため、上記の設定では出現数の多い5ブランドでしか、関連ルールが出てこなかった。

いずれのブランドも、福神漬との併買が高い。福神漬以外で特徴があるのは〔「パームントカレー」と「ビスケット・クッキー」〕の組み合わせ、および〔「こくまるカレー」で「チルドプリン」や「トマトケチャップ」〕との併買が見られていた。他にわずかな数ではあるが、〔「ゴールデンカレー」と「パイナップル缶詰」〕、〔「あらびきカレー」と「カレー粉」〕、〔「ディナーカレー」と「紙パック入り果汁100%飲料」〕など、ブランドごとにそれぞれ異なった分類の出現が見られた。

表8-8 カレールーブランド別併買状況

ゴールデン <= 普通食パン & 福神漬	(105:0.3%, 0.092)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & 冷凍ハンバーグ	(70:0.2%, 0.083)
ゴールデン <= ビスケット・クッキー & プレーンヨーグルト	(105:0.3%, 0.065)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & 福神漬	(139:0.4%, 0.064)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & バター	(105:0.3%, 0.064)
ゴールデン <= 生鮮水産物 & 味付けノリ・切りノリ	(70:0.2%, 0.064)
ゴールデン <= 福神漬	(523:1.5%, 0.059)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & 卵豆腐	(105:0.3%, 0.058)
ゴールデン <= 粒納豆 & ロールパン・コッペパン & ビスケット・クッキー	(70:0.2%, 0.057)
ゴールデン <= 一般牛乳 & 福神漬	(105:0.3%, 0.055)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & 乾燥くず切り・春雨	(70:0.2%, 0.054)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & パイナップル缶詰	(105:0.3%, 0.053)
ゴールデン <= 普通食パン & 焼き肉のたれ	(70:0.2%, 0.053)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & ウインナーソーセージ & 即席カップ中華そば	(70:0.2%, 0.052)
ゴールデン <= とろろコンブ	(70:0.2%, 0.052)
ゴールデン <= 絹ごし豆腐・ソフト豆腐 & 普通食パン & せんべい	(70:0.2%, 0.051)
ゴールデン <= 菓子パン & 福神漬	(105:0.3%, 0.05)
ディナーカレー <= 木綿豆腐&紙パック入り果汁100%飲料	(70:0.2%, 0.051)
あら挽きカレー <= 絹ごし豆腐・ソフト豆腐&普通食パン&マヨネーズ・マヨネーズタイプ調味料	(70:0.2%, 0.056)
あら挽きカレー <= カレー粉	(70:0.2%, 0.055)
熱カレー <= ホームタイプマルチアイス&福神漬	(70:0.2%, 0.077)
熱カレー <= 生鮮農産物&福神漬	(70:0.2%, 0.062)
熱カレー <= 普通食パン&バラエティ食パン	(105:0.3%, 0.057)

表8-8 カレールーブランド別併買状況(続き)

バーモント <= 生鮮農産物 & 福神漬	(70:0.2%, 0.162)
バーモント <= ホームタイプマルチアイス & 福神漬	(70:0.2%, 0.154)
バーモント <= 粒納豆 & 福神漬	(105:0.3%, 0.112)
バーモント <= 一般牛乳 & 福神漬	(105:0.3%, 0.11)
バーモント <= フルーツ入りヨーグルト & せんべい	(105:0.3%, 0.102)
バーモント <= 即席カップ中華そば & チルドギョーザ	(70:0.2%, 0.1)
バーモント <= 絹ごし豆腐・ソフト豆腐 & ホームタイプマルチアイス & スライスロースハム	(70:0.2%, 0.099)
バーモント <= ビスケット・クッキー & マヨネーズ・マヨネーズタイプ調味料	(70:0.2%, 0.094)
バーモント <= ビスケット・クッキー & その他漬物	(70:0.2%, 0.093)
バーモント <= 福神漬	(523:1.5%, 0.086)
バーモント <= 普通食パン & 福神漬	(105:0.3%, 0.082)
バーモント <= 絹ごし豆腐・ソフト豆腐 & 缶入りコーヒー飲料	(70:0.2%, 0.081)
バーモント <= ホームタイプマルチアイス & 麻婆の素	(105:0.3%, 0.079)
バーモント <= ウインナソーセージ & 冷凍用ポリドリンク	(70:0.2%, 0.078)
バーモント <= 即席カップ中華そば & 生焼きそば	(105:0.3%, 0.075)
バーモント <= 生鮮農産物 & 冷凍用ポリドリンク	(70:0.2%, 0.075)
バーモント <= 絹ごし豆腐・ソフト豆腐 & 福神漬	(139:0.4%, 0.071)
バーモント <= ビスケット・クッキー & 生焼きそば	(70:0.2%, 0.071)
バーモント <= 粒納豆 & チルドミートボール	(70:0.2%, 0.07)
バーモント <= 粒納豆 & 生鮮農産物 & ロールパン・コッペパン	(105:0.3%, 0.069)
バーモント <= ウインナソーセージ & 冷凍米飯	(105:0.3%, 0.069)
バーモント <= 即席カップ中華そば & 冷凍米飯	(105:0.3%, 0.069)
バーモント <= 即席カップ中華そば & 濃口しょうゆ	(70:0.2%, 0.068)
バーモント <= 即席カップ中華そば & つゆの素・めんつゆ(濃縮)	(70:0.2%, 0.068)
バーモント <= ロールパン・コッペパン & ドレッシング・ドレッシングタイプ調味料	(70:0.2%, 0.068)
バーモント <= ビスケット・クッキー & 生揚げ・厚揚げ	(70:0.2%, 0.068)
バーモント <= 牛乳タイプ乳飲料 & つゆの素・めんつゆ(濃縮)	(70:0.2%, 0.067)
バーモント <= 菓子パン & 福神漬	(105:0.3%, 0.067)
バーモント <= 鶏卵 & 生焼きそば	(70:0.2%, 0.067)
バーモント <= 薄力小麦粉	(244:0.7%, 0.066)
バーモント <= 普通食パン & 即席カップ焼きそば	(105:0.3%, 0.066)
バーモント <= 生鮮農産物 & ホームタイプマルチアイス & ロールパン・コッペパン	(70:0.2%, 0.066)
バーモント <= ホームタイプマルチアイス & トマトケチャップ	(105:0.3%, 0.064)
こくまろ <= 菓子パン & 福神漬	(105:0.3%, 0.126)
こくまろ <= ホームタイプマルチアイス & 福神漬	(70:0.2%, 0.115)
こくまろ <= 普通食パン & 福神漬	(105:0.3%, 0.102)
こくまろ <= 絹ごし豆腐・ソフト豆腐 & 菓子パン & 一般牛乳 & ホームタイプマルチアイス	(70:0.2%, 0.097)
こくまろ <= 即席カップ中華そば & チルドプリン	(70:0.2%, 0.09)
こくまろ <= 一般牛乳 & 福神漬	(105:0.3%, 0.083)
こくまろ <= 福神漬	(523:1.5%, 0.082)
こくまろ <= 絹ごし豆腐・ソフト豆腐 & 福神漬	(139:0.4%, 0.079)
こくまろ <= 普通食パン & 冷凍用ポリドリンク	(105:0.3%, 0.079)
こくまろ <= 菓子パン & トマトケチャップ	(105:0.3%, 0.076)
こくまろ <= スライスロースハム & チルドプリン	(70:0.2%, 0.074)
こくまろ <= ホームタイプマルチアイス & トマトケチャップ	(105:0.3%, 0.074)
こくまろ <= ホームタイプマルチアイス & 冷凍用ポリドリンク	(70:0.2%, 0.074)
こくまろ <= 生鮮水産物 & 冷凍コロッケ	(70:0.2%, 0.07)
こくまろ <= 絹ごし豆腐・ソフト豆腐 & 菓子パン & 小麦系スナック	(70:0.2%, 0.069)
こくまろ <= 菓子パン & たら焼き・三笠山	(70:0.2%, 0.069)
こくまろ <= 絹ごし豆腐・ソフト豆腐 & 一般牛乳 & 乳酸菌飲料	(70:0.2%, 0.068)
こくまろ <= 粒納豆 & 生鮮水産物 & その他漬物	(70:0.2%, 0.068)
こくまろ <= スライスロースハム & ベーコン	(70:0.2%, 0.068)
こくまろ <= ビスケット・クッキー & 生揚げ・厚揚げ	(70:0.2%, 0.068)
こくまろ <= 絹ごし豆腐・ソフト豆腐 & ホームタイプマルチアイス & フルーツ入りヨーグルト	(105:0.3%, 0.067)
こくまろ <= 菓子パン & 一般牛乳 & ホームタイプマルチアイス	(174:0.5%, 0.066)
こくまろ <= 粒納豆 & 福神漬	(105:0.3%, 0.065)
こくまろ <= 菓子パン & 濃口しょうゆ	(105:0.3%, 0.063)
こくまろ <= 絹ごし豆腐・ソフト豆腐 & ピラフ・チャーハンの素(ドライ)	(105:0.3%, 0.062)
こくまろ <= 菓子パン & 冷凍用ポリドリンク	(105:0.3%, 0.061)
こくまろ <= スライスロースハム & その他漬物	(105:0.3%, 0.06)
こくまろ <= 一般牛乳 & 生鮮水産物 & 鶏卵	(70:0.2%, 0.06)
こくまろ <= 菓子パン & ホームタイプマルチアイス & 即席カップ中華そば	(70:0.2%, 0.06)

②容量別（表8-9）

200G未満のグループで〔「ホームタイプマルチアイス」&「福神漬」〕との併買が信頼度33.3%と最も高い。一方200G以上のクラスでは〔「一般牛乳」&「福神漬」〕との併買組み合わせが12.8%。容量別では「カレー粉」がどちらのグループにも共通して上位に並んでいた。200G未満の商品はもともとの出現件数が多いことから信頼度の高い組み合わせが多い。

表8-9 カレールー容量別併買状況

200G以上<=一般牛乳&福神漬	(105:0.3% 0.128)
200G以上<=菓子パン&福神漬	(105:0.3% 0.101)
200G以上<=普通食パン&福神漬	(105:0.3% 0.092)
200G以上<=福神漬	(523:1.5% 0.084)
200G以上<=カレー粉	(70:0.2% 0.082)
200G以上<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.077)
200G以上<=生鮮農産物&福神漬	(70:0.2% 0.075)
200G以上<=絹ごし豆腐・ソフト豆腐&練りカラシ・マスタード	(70:0.2% 0.07)
200G以上<=ウインナーソーセージ&冷凍コロッケ	(70:0.2% 0.065)
200G以上<=ポテトチップス&ジャガ芋系スナック	(70:0.2% 0.065)
200G以上<=ビスケット・クッキー&冷凍パスタ	(70:0.2% 0.065)
200G以上<=粒納豆&福神漬	(105:0.3% 0.065)
200G以上<=即席カップ中華そば&チルドプリン	(70:0.2% 0.064)
200G以上<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.064)
200G以上<=木綿豆腐&紙パック入り果汁100%飲料	(70:0.2% 0.063)
200G以上<=普通食パン&一般みそ	(105:0.3% 0.062)
200G以上<=普通食パン&麦茶ティーバッグ	(70:0.2% 0.059)
200G以上<=その他漬物&ペットボトル入り果汁入り清涼飲料	(70:0.2% 0.057)
200G以上<=木綿豆腐&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.056)
200G以上<=絹ごし豆腐・ソフト豆腐&普通食パン&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.056)
200G以上<=普通食パン&バター	(70:0.2% 0.056)
200G以上<=ビスケット・クッキー&ポーション型菓子（ゼリーなど）	(70:0.2% 0.056)
200G以上<=生鮮農産物&チルドコーヒーゼリー	(70:0.2% 0.056)
200G以上<=ホームタイプマルチアイス&小麦系スナック	(174:0.5% 0.056)
200G以上<=ポップコーン	(105:0.3% 0.056)
200G以上<=粒納豆&冷凍ハンバーグ	(70:0.2% 0.054)
200G以上<=菓子パン&魚肉ソーセージ	(70:0.2% 0.054)
200G以上<=ホームタイプマルチアイス&紙パック入り果汁混合野菜ジュース	(70:0.2% 0.054)
200G以上<=粒納豆&一般みそ	(139:0.4% 0.054)
200G以上<=一般牛乳&一般みそ	(105:0.3% 0.054)
200G未満<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.333)
200G未満<=普通食パン&福神漬	(105:0.3% 0.255)
200G未満<=菓子パン&福神漬	(105:0.3% 0.244)
200G未満<=福神漬	(523:1.5% 0.241)
200G未満<=生鮮農産物&福神漬	(70:0.2% 0.237)
200G未満<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.236)
200G未満<=一般牛乳&福神漬	(105:0.3% 0.211)
200G未満<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&スライスロースハム	(70:0.2% 0.183)
200G未満<=ビスケット・クッキー&生揚げ・厚揚げ	(70:0.2% 0.176)
200G未満<=絹ごし豆腐・ソフト豆腐&シュレッドナチュラルチーズ	(70:0.2% 0.153)
200G未満<=粒納豆&福神漬	(105:0.3% 0.15)
200G未満<=スライスロースハム&ベーコン	(70:0.2% 0.149)
200G未満<=ビスケット・クッキー&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.141)
200G未満<=ビスケット・クッキー&その他漬物	(70:0.2% 0.14)
200G未満<=カレー粉	(70:0.2% 0.137)
200G未満<=一般牛乳&生鮮水産物&鶏卵	(70:0.2% 0.131)
200G未満<=ウインナーソーセージ&冷凍用ポリドリンク	(70:0.2% 0.13)
200G未満<=牛乳タイプ乳飲料&せんべい	(70:0.2% 0.128)
200G未満<=ホームタイプマルチアイス&トマトケチャップ	(105:0.3% 0.128)
200G未満<=粒納豆&トマトケチャップ	(105:0.3% 0.126)

表8-9 カレールー容量別併買状況(続き)

200G未満<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.333)
200G未満<=普通食パン&福神漬	(105:0.3% 0.255)
200G未満<=菓子パン&福神漬	(105:0.3% 0.244)
200G未満<=福神漬	(523:1.5% 0.241)
200G未満<=生鮮農産物&福神漬	(70:0.2% 0.237)
200G未満<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.236)
200G未満<=一般牛乳&福神漬	(105:0.3% 0.211)
200G未満<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&スライスロースハム	(70:0.2% 0.183)
200G未満<=ビスケット・クッキー&生揚げ・厚揚げ	(70:0.2% 0.176)
200G未満<=絹ごし豆腐・ソフト豆腐&シュレッドナチュラルチーズ	(70:0.2% 0.153)
200G未満<=粒納豆&福神漬	(105:0.3% 0.15)
200G未満<=スライスロースハム&ベーコン	(70:0.2% 0.149)
200G未満<=ビスケット・クッキー&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.141)
200G未満<=ビスケット・クッキー&その他漬物	(70:0.2% 0.14)
200G未満<=カレー粉	(70:0.2% 0.137)
200G未満<=一般牛乳&生鮮水産物&鶏卵	(70:0.2% 0.131)
200G未満<=ウインナーソーセージ&冷凍用ポリドリンク	(70:0.2% 0.13)
200G未満<=牛乳タイプ乳飲料&せんべい	(70:0.2% 0.128)
200G未満<=ホームタイプマルチアイス&トマトケチャップ	(105:0.3% 0.128)
200G未満<=粒納豆&トマトケチャップ	(105:0.3% 0.126)
200G未満<=紙パック入りコーヒー飲料&プレーンヨーグルト	(70:0.2% 0.111)
200G未満<=即席カップ中華そば&ふりかけ	(70:0.2% 0.111)
200G未満<=ビスケット・クッキー&豆菓子	(70:0.2% 0.111)
200G未満<=即席カップ中華そば&つゆの素・めんつゆ(濃縮)	(70:0.2% 0.11)
200G未満<=絹ごし豆腐・ソフト豆腐&味付けノリ・切りノリ	(105:0.3% 0.109)
200G未満<=絹ごし豆腐・ソフト豆腐&トマトケチャップ	(139:0.4% 0.109)
200G未満<=一般牛乳&チルドギョーザ	(139:0.4% 0.109)
200G未満<=菓子パン&トマトケチャップ	(105:0.3% 0.109)
200G未満<=絹ごし豆腐・ソフト豆腐&生鮮農産物&ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.108)
200G未満<=粒納豆&ピラフ・チャーハンの素(ドライ)	(70:0.2% 0.108)
200G未満<=ロールパン・コッペパン&ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.108)
200G未満<=ビスケット・クッキー&プレーンヨーグルト	(105:0.3% 0.108)
200G未満<=その他パウダースパイス	(105:0.3% 0.107)
200G未満<=普通食パン&焼き肉のたれ	(70:0.2% 0.107)
200G未満<=ビスケット・クッキー&生焼きそば	(70:0.2% 0.107)
200G未満<=鶏卵&生焼きそば	(70:0.2% 0.107)
200G未満<=ホームタイプマルチアイス&ふりかけ	(105:0.3% 0.107)
200G未満<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&鶏卵	(70:0.2% 0.105)
200G未満<=油揚げ&せんべい	(105:0.3% 0.105)
200G未満<=粒納豆&パスタソース缶詰	(105:0.3% 0.105)
200G未満<=一般牛乳&ホームタイプマルチアイス&即席カップ中華そば	(70:0.2% 0.105)
200G未満<=絹ごし豆腐・ソフト豆腐&一般牛乳&スライスロースハム	(139:0.4% 0.104)
200G未満<=絹ごし豆腐・ソフト豆腐&濃口しょうゆ	(174:0.5% 0.104)
200G未満<=紙パック入り果汁入り清涼飲料&チルドプリン	(70:0.2% 0.104)

③辛さ別 (表8-10)

いずれの辛さにおいても「福神漬」との併買は高いが、「甘口」「やや甘口」および「中辛」グループでは「冷凍ポリドリンク」や「チルドプリン」との併買が高い。また、「やや辛口」および「辛口」クラスだけに「カレー粉」や「その他パウダースパイス」との併買が見られた。辛さごとに特有の買い方があると言えよう。

表8-10 カレールー辛さ別併買状況

甘口<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.141)
甘口<=一般牛乳&福神漬	(105:0.3% 0.119)
甘口<=普通食パン&福神漬	(105:0.3% 0.102)
甘口<=菓子パン&福神漬	(105:0.3% 0.092)
甘口<=スライスロースハム&小麦系スナック	(70:0.2% 0.085)
甘口<=福神漬	(523:1.5% 0.08)
甘口<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.079)
甘口<=とろろコンブ	(70:0.2% 0.078)
甘口<=絹ごし豆腐・ソフト豆腐&普通食パン&せんべい	(70:0.2% 0.077)
甘口<=生鮮農産物&福神漬	(70:0.2% 0.075)
甘口<=せんべい&袋入りハードキャンデー	(70:0.2% 0.071)
甘口<=普通食パン&バター	(70:0.2% 0.07)
甘口<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&スライスロースハム	(70:0.2% 0.07)
甘口<=即席カップ中華そば&濃口しょうゆ	(70:0.2% 0.068)
甘口<=ハードヨーグルト&せんべい	(70:0.2% 0.068)
甘口<=フルーツ入りヨーグルト&せんべい	(105:0.3% 0.068)
甘口<=スライスロースハム&ベーコン	(70:0.2% 0.068)
甘口<=普通食パン&冷凍用ポリドリンク	(105:0.3% 0.067)
甘口<=紙パック入り果汁入り清涼飲料&チルドプリン	(70:0.2% 0.065)
甘口<=粒納豆&福神漬	(105:0.3% 0.065)
甘口<=鶏卵&せんべい	(70:0.2% 0.063)
甘口<=ポテトチップス&冷凍用ポリドリンク	(70:0.2% 0.063)
甘口<=生鮮農産物&冷凍用ポリドリンク	(70:0.2% 0.062)
甘口<=絹ごし豆腐・ソフト豆腐&一般牛乳&スライスロースハム	(139:0.4% 0.059)
甘口<=ホームタイプマルチアイス&ふりかけ	(105:0.3% 0.058)
甘口<=菓子パン&生鮮農産物&ポテトチップス	(70:0.2% 0.057)
甘口<=即席カップ中華そば&チルドギョーザ	(70:0.2% 0.057)
甘口<=牛乳タイプ乳飲料&ドーナツ	(70:0.2% 0.057)
甘口<=スライスロースハム&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.057)
甘口<=紙パック入りコーヒー飲料&その他冷凍野菜素材	(70:0.2% 0.056)
やや甘口<=生鮮農産物&福神漬	(70:0.2% 0.1)
やや甘口<=ビスケット・クッキー&その他漬物	(70:0.2% 0.093)
やや甘口<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.077)
やや甘口<=ビスケット・クッキー&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.071)
やや甘口<=生鮮農産物&ホームタイプマルチアイス&ロールパンコッペパン	(70:0.2% 0.066)
やや甘口<=粒納豆&福神漬	(105:0.3% 0.065)
やや甘口<=一般牛乳&福神漬	(105:0.3% 0.064)
やや甘口<=フルーツ入りヨーグルト&せんべい	(105:0.3% 0.057)
やや甘口<=菓子パン&生クリーム	(70:0.2% 0.056)
やや甘口<=菓子パン&生鮮水産物&ロールパンコッペパン	(70:0.2% 0.056)
やや甘口<=ホームタイプマルチアイス&麻婆の素	(105:0.3% 0.056)
やや甘口<=ビスケット・クッキー&生揚げ・厚揚げ	(70:0.2% 0.054)
やや甘口<=その他漬物&その他水産加工品	(70:0.2% 0.053)
やや甘口<=ウインナーソーセージ&冷凍用ポリドリンク	(70:0.2% 0.052)
やや甘口<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&ロールパンコッペパン	(105:0.3% 0.051)
やや甘口<=普通食パン&福神漬	(105:0.3% 0.051)
中辛<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.167)
中辛<=普通食パン&福神漬	(105:0.3% 0.153)
中辛<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.143)
中辛<=福神漬	(523:1.5% 0.136)

表8-10 カレールー辛さ別併買状況(続き)

中辛<=一般牛乳&福神漬	(105:0.3% 0.128)
中辛<=菓子パン&福神漬	(105:0.3% 0.126)
中辛<=即席カップ中華そば&チルドプリン	(70:0.2% 0.115)
中辛<=絹ごし豆腐・ソフト豆腐&シュレッドナチュラルチーズ	(70:0.2% 0.106)
中辛<=生鮮農産物&福神漬	(70:0.2% 0.1)
中辛<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&スライスロースハム	(70:0.2% 0.099)
中辛<=絹ごし豆腐・ソフト豆腐&生鮮農産物&ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.095)
中辛<=粒納豆&ピラフ・チャーハンの素(ドライ)	(70:0.2% 0.095)
中辛<=粒納豆&トマトケチャップ	(105:0.3% 0.092)
中辛<=一般牛乳&ホームタイプマルチアイス&即席カップ中華そば	(70:0.2% 0.092)
中辛<=普通食パン&冷凍用ポリドリンク	(105:0.3% 0.09)
中辛<=マヨネーズ・マヨネーズタイプ調味料&ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.087)
中辛<=即席カップ中華そば&チルドギョーザ	(70:0.2% 0.086)
中辛<=粒納豆&味付けノリ・切りノリ	(70:0.2% 0.085)
中辛<=絹ごし豆腐・ソフト豆腐&ウインナーソーセージ&油揚げ	(70:0.2% 0.083)
中辛<=絹ごし豆腐・ソフト豆腐&白砂糖	(70:0.2% 0.083)
中辛<=絹ごし豆腐・ソフト豆腐&菓子パン&一般牛乳&ホームタイプマルチアイス	(70:0.2% 0.083)
中辛<=スライスロースハム&ベーコン	(70:0.2% 0.081)
中辛<=ビスケット・クッキー&生揚げ・厚揚げ	(70:0.2% 0.081)
中辛<=スライスロースハム&その他漬物	(105:0.3% 0.08)
中辛<=鶏卵&生焼きそば	(70:0.2% 0.08)
やや辛口<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.128)
やや辛口<=カレー粉	(70:0.2% 0.11)
やや辛口<=菓子パン&福神漬	(105:0.3% 0.092)
やや辛口<=福神漬	(523:1.5% 0.082)
やや辛口<=一般牛乳&福神漬	(105:0.3% 0.073)
やや辛口<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.071)
やや辛口<=木綿豆腐&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.07)
やや辛口<=ビスケット・クッキー&菓子パイ	(105:0.3% 0.068)
やや辛口<=一般牛乳&麻婆の素	(105:0.3% 0.066)
やや辛口<=一般牛乳&冷凍コロッケ	(70:0.2% 0.065)
やや辛口<=菓子パン&ホームタイプマルチアイス&ポテトチップス	(105:0.3% 0.064)
やや辛口<=生鮮農産物&福神漬	(70:0.2% 0.062)
やや辛口<=即席カップ中華そば&スバゲティ・ロングパスタ	(70:0.2% 0.061)
やや辛口<=ビスケット・クッキー&生ちくわ・焼きちくわ	(70:0.2% 0.06)
やや辛口<=菓子パン&粒納豆&ビスケット・クッキー	(105:0.3% 0.058)
やや辛口<=粒納豆&一般牛乳&紙パック入りコーヒー飲料	(70:0.2% 0.057)
やや辛口<=粒納豆&ロールパン・コッペパン&ビスケット・クッキー	(70:0.2% 0.057)
やや辛口<=その他漬物&ペットボトル入り果汁入り清涼飲料	(70:0.2% 0.057)
やや辛口<=ホームタイプマルチアイス&フレーク	(105:0.3% 0.057)
やや辛口<=絹ごし豆腐・ソフト豆腐&普通食パン&	(70:0.2% 0.056)
やや辛口<=菓子パン&生鮮農産物&即席カップ中華そば	(70:0.2% 0.056)
やや辛口<=ビスケット・クッキー&ポーション型菓子(ゼリーなど)	(70:0.2% 0.056)
やや辛口<=普通食パン&濃口しょうゆ	(105:0.3% 0.055)
やや辛口<=普通食パン&ホームタイプマルチアイス&ポテトチップス	(70:0.2% 0.055)
やや辛口<=ポテトチップス&ハードヨーグルト	(105:0.3% 0.055)
やや辛口<=ビスケット・クッキー&フレーク	(105:0.3% 0.055)
やや辛口<=その他パウダースライス	(105:0.3% 0.054)
やや辛口<=絹ごし豆腐・ソフト豆腐&菓子パン&木綿豆腐	(70:0.2% 0.053)
やや辛口<=一般牛乳&マヨネーズ・マヨネーズタイプ調味料	(174:0.5% 0.053)
やや辛口<=紙パック入りコーヒー飲料&キムチ・朝鮮漬	(70:0.2% 0.053)
やや辛口<=即席カップ中華そば&マヨネーズ・マヨネーズタイプ調味料	(105:0.3% 0.053)
やや辛口<=菓子パン&生冷やし日本そば	(70:0.2% 0.053)
辛口<=普通食パン&福神漬	(105:0.3% 0.071)
辛口<=ビスケット・クッキー&ポーション型菓子(ゼリーなど)	(70:0.2% 0.056)
辛口<=カレー粉	(70:0.2% 0.055)
辛口<=ビスケット・クッキー&プレーンヨーグルト	(105:0.3% 0.054)

④ クラス別 (表8-11)

大半の商品が一般グループの商品に属するため、一般グループの結果はカレールー全体の数値 (表8-6) に近くなっている。高級グループではわずかなではあるが「紙パック入り果汁100%飲料」の併買が見られた。

表8-11 カレールークラス別併買状況

一般<=ホームタイプマルチアイス&福神漬	(70:0.2% 0.372)
一般<=普通食パン&福神漬	(105:0.3% 0.306)
一般<=生鮮農産物&福神漬	(70:0.2% 0.3)
一般<=一般牛乳&福神漬	(105:0.3% 0.284)
一般<=菓子パン&福神漬	(105:0.3% 0.277)
一般<=福神漬	(523:1.5% 0.274)
一般<=絹ごし豆腐・ソフト豆腐&福神漬	(139:0.4% 0.25)
一般<=粒納豆&福神漬	(105:0.3% 0.196)
一般<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&スライスロースハム	(70:0.2% 0.183)
一般<=カレールー粉	(70:0.2% 0.178)
一般<=ビスケット・クッキー&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.165)
一般<=絹ごし豆腐・ソフト豆腐&バター	(105:0.3% 0.149)
一般<=スライスロースハム&ベーコン	(70:0.2% 0.149)
一般<=ホームタイプマルチアイス&トマトケチャップ	(105:0.3% 0.149)
一般<=普通食パン&冷凍用ポリドリンク	(105:0.3% 0.146)
一般<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&フルーツ入りヨーグルト	(105:0.3% 0.144)
一般<=粒納豆&ロールパン・コッペパン&ビスケット・クッキー	(70:0.2% 0.143)
一般<=即席カップ中華そば&チルドギョーザ	(70:0.2% 0.143)
一般<=絹ごし豆腐・ソフト豆腐&シュレッドナチュラルチーズ	(70:0.2% 0.141)
一般<=即席カップ中華そば&チルドプリン	(70:0.2% 0.141)
一般<=ビスケット・クッキー&その他漬物	(70:0.2% 0.14)
一般<=一般牛乳&マヨネーズ・マヨネーズタイプ調味料	(174:0.5% 0.139)
一般<=マヨネーズ・マヨネーズタイプ調味料&ドレッシング・ドレッシングタイプ調味料	(70:0.2% 0.138)
一般<=ビスケット・クッキー&生揚げ・厚揚げ	(70:0.2% 0.135)
一般<=鶏卵&生焼きそば	(70:0.2% 0.133)
一般<=一般牛乳&生鮮水産物&鶏卵	(70:0.2% 0.131)
一般<=ウインナソーセージ&冷凍用ポリドリンク	(70:0.2% 0.13)
一般<=絹ごし豆腐・ソフト豆腐&味付けノリ・切りノリ	(105:0.3% 0.129)
一般<=絹ごし豆腐・ソフト豆腐&粒納豆&チルドギョーザ	(70:0.2% 0.127)
一般<=絹ごし豆腐・ソフト豆腐&生鮮農産物&生ちくわ・焼きちくわ	(70:0.2% 0.127)
一般<=普通食パン&バター	(70:0.2% 0.127)
一般<=ポテトチップス&冷凍用ポリドリンク	(70:0.2% 0.127)
一般<=絹ごし豆腐・ソフト豆腐&濃口しょうゆ	(174:0.5% 0.126)
一般<=粒納豆&トマトケチャップ	(105:0.3% 0.126)
一般<=ホームタイプマルチアイス&ふりかけ	(105:0.3% 0.126)
一般<=普通食パン&ポーシオン型菓子 (ゼリーなど)	(105:0.3% 0.125)
一般<=ウインナソーセージ&マヨネーズ・マヨネーズタイプ調味料	(105:0.3% 0.124)
一般<=一般牛乳&トマトケチャップ	(105:0.3% 0.124)
一般<=普通食パン&卵豆腐	(105:0.3% 0.124)
一般<=生鮮農産物&即席カップ焼きそば	(105:0.3% 0.124)
一般<=一般牛乳&チルドギョーザ	(139:0.4% 0.123)
一般<=サバ缶詰	(70:0.2% 0.123)
一般<=絹ごし豆腐・ソフト豆腐&粒納豆&うるち米	(70:0.2% 0.122)
一般<=粒納豆&味付けノリ・切りノリ	(70:0.2% 0.122)
一般<=粒納豆&ピラフ・チャーハンの素 (ドライ)	(70:0.2% 0.122)
一般<=粒納豆&豆腐子	(139:0.4% 0.122)
一般<=一般牛乳&コーン缶詰	(70:0.2% 0.122)
一般<=粒納豆&ロールパン・コッペパン&ウインナソーセージ	(105:0.3% 0.12)
一般<=絹ごし豆腐・ソフト豆腐&一般牛乳&スライスロースハム	(139:0.4% 0.119)
一般<=ウインナソーセージ&冷凍米飯	(105:0.3% 0.119)
一般<=一般牛乳&ふりかけ	(139:0.4% 0.119)
一般<=絹ごし豆腐・ソフト豆腐&ホームタイプマルチアイス&鶏卵	(70:0.2% 0.118)
一般<=即席カップ中華そば&ビスケット・クッキー	(174:0.5% 0.118)
一般<=ビスケット・クッキー&即席袋中華そば	(70:0.2% 0.118)

表8-11 カレールー辛さ別併買状況(続き)

一般<=紙パック入り果汁入り清涼飲料&チルドプリン	(70:0.2% 0.117)
一般<=とろろコンブ	(70:0.2% 0.117)
一般<=鶏卵&マヨネーズ・マヨネーズタイプ調味料	(70:0.2% 0.117)
一般<=一般牛乳&麻婆の素	(105:0.3% 0.116)
一般<=牛乳タイプ乳飲料&せんべい	(70:0.2% 0.116)
一般<=その他パウダースライス	(105:0.3% 0.116)
高級<=牛乳タイプ乳飲料&缶入りコーヒー飲料	(70:0.2% 0.054)
高級<=菓子パン&普通食パン&小麦系スナック	(70:0.2% 0.053)
高級<=スライスロースハム&そうめん	(105:0.3% 0.052)
高級<=木綿豆腐&紙パック入り果汁100%飲料	(105:0.3% 0.051)

以上①～④で傾向の見られた分類について表8-12にまとめた。カレールーでは特に辛さ別、ブランド別でそれぞれのグループごとに異なる併買分類が見られた。このように、従来の商品分類よりもさらに商品特性で区分することでより詳細なバスケットのの中身を知ることができる。

さらに分析を発展させるためには、併買分類でもバーチャルアイテムを設定するといいだろう。単品との併買を見ることで、例えば、[高級クラスのカレールーを買う人は豆腐も無農薬大豆の高級なものを買う]、などの詳細な購買パターンを見つけることが可能になるであろう。

8. 2. 4 改善率～相乗効果を上げるための指標

信頼度の場合は[商品Aを買うとき、商品Bも同時に購買する]、この商品Aと商品Bの出現数に結果が左右されてしまう。特にA、B間の出現数の差が大きいと、片側からみた結果ばかりが出力されることに注意しなければならない。ここではA、B双方にとってより効果をもたらす関係を見いだすため、改善率という指標を用いて併買関係を探ってみる。

改善率は理論上、「併売されるであろう」という予測値と実際の出現率との比である。実際値が予測値に対しどのくらい大きくなるかによって併買効果の大きさが見えてくる。式に書くと下記のようなになる。

$$\begin{aligned}
 \text{改善率} &= \frac{\text{同時購買率}}{\text{商品Aの出現率} \times \text{商品Bの出現率}} \\
 &= \frac{\frac{\text{同時購買件数}}{\text{全レシート数}}}{\frac{\text{商品A購買件数}}{\text{全レシート数}} \times \frac{\text{商品B購買件数}}{\text{全レシート数}}}
 \end{aligned}$$

表8-12 パーチャルアイテム別併買信頼性

	ホームタ イプマル チアイス &福神漬	普通食パ ン&福神 漬	菓子パン &福神漬	福神漬	一般牛乳 &福神漬	生鮮農産 物&福神 漬	絹ごし豆 腐・ソフ ト豆腐& 福神漬	粒納豆& 福神漬	カレー粉	マヨネー ズ	ビスケッ ト・クッ キー	冷凍用ポ リドリ ンク	チルドプ リン	トマトケ チャップ	その他パ ウダー ス	紙バック 入り果汁 100%飲料
カレー全体	41	33.7	31.9	30.7	30.3	30	27.9	21.5	19.2	18.8	17.6	15.7	15.4	13.5	13.4	12.5
(ブランド別)																
ゴールデン	-	9.2	5	5.9	5.5	-	6.4	-	-	-	6.5	-	-	-	-	-
パーモント	16.4	8.2	6.7	8.6	11	16.2	7.1	11.2	-	9.4	9.4	7.8	5.7	6.4	-	-
こくまろ	11.5	10.2	12.6	8.2	8.3	-	7.9	6.5	-	-	6.8	7.9	9	7.6	-	5.7
熟カレー	7.7	-	-	-	-	6.2	-	-	-	-	-	-	-	-	-	-
ディナーカレー	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5.1
あらびきカレー	-	-	-	-	-	-	-	-	5.5	5.6	-	-	-	-	-	-
(容量別)																
200G以上	7.7	9.2	10.1	8.4	12.8	7.5	6.4	6.5	8.2	5.6	6.5	-	6.4	5.2	-	5.3
200G未満	33.3	25.6	24.4	24.1	21.1	23.7	23.6	15	13.7	14.1	17.6	13	10.4	12.8	10.7	8
(クラス別)																
一般	37.2	30.6	27.7	27.4	28.4	30	25	19.6	17.8	16.5	16.5	14.6	14.1	14.9	11.6	10.2
高級	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	5.1
(辛さ別)																
甘口	14.1	10.2	9.2	8	11.9	7.5	7.9	6.5	-	5.7	-	6.7	6.5	-	-	-
やや甘口	7.7	5.1	-	-	6.4	10	-	6.5	-	7.1	9.3	5.2	-	-	-	-
中辛	16.3	15.7	12.6	13.6	12.8	10	14.3	7.5	-	8.7	8.1	9	11.5	9.2	-	5.7
やや辛口	12.8	5.1	9.2	8.2	7.3	6.2	7.1	-	11	7	6.8	-	-	-	5.4	-
辛口	-	7.1	-	-	-	-	-	-	5.5	-	5.6	-	-	-	-	-

※ 数字はすべてカレールー、および各パーチャルアイテムとの併買の信頼性(%)。-は、信頼性5.0%未満のもの。

※ マヨネーズ、ビスケット・クッキー、冷凍ポリドリンク、チルドプリン、トマトケチャップ、紙バック入り果汁100%飲料は組み合わせの内容に関係なく、その分類が登場したときの最も信頼性の高い数値を採用。

※ クラス別で、子供用のカレーでは、信頼性5.0%以上のものが出なかった。

改善率が1より大きい場合（＝通常の出現率＜併買による出現率）であれば、普段単品で買うより併買をすることによって、両方の商品の売上があがる相乗効果をもたらす可能性がある。

改善率について注記しておく、商品A、Bの出現率が小さいほど改善率の値は大きくなりやすい。逆に出現率が高いものは、改善率の数字は大きくなる（出現率50%の商品の改善率は2より大きくなる）。改善率は、率の大きさよりも特に出現数のあまり多くない分類の良好な併買関係を発見する指標として適したものと考えて欲しい。

実際に3ヶ月データで、小分類単位で改善率を出した結果は表8-13の通り。

最も改善率が高かったのは「洋風総菜缶詰」と「畜肉調理缶詰」であるが、「洋風総菜缶詰」自体の出現数がはもともと3ヶ月で13件と、あまり売れていない分類である。

そこで、同時件数が12件以上のものを追ってみると、もっとも改善率が高いのは〔「紙パック入り健康茶飲料」と「紙パック入り緑茶飲料」〕の同属性分類同士の組み合わせ（改善率114.2）。次いで〔「お好み焼きミックス」と「青ノリ」〕（同91.2）、〔「即席カップ入りうどん」と「即席カップ入り日本そば」〕の組み合わせ（同87.3）と続く。お好み焼きの具となる商品ををエンドで販売したり、飲料などは夏期にバンドル販売を行い、売上の相乗効果をあげるなどの販売戦略に役立ちそうだ。

また、〔「練乳・コンデンスミルクとかき氷シロップ・氷みつ」〕など、Aprioriで全体の概要を見たときには出力されなかった小規模分類同士の組み合わせが出てくる。改善率はメーカーサイドにとっては自社で製造している商品分類の併買効果を見るのに役に立つ指標となるだろう。

表8-13 3ヶ月データ全体の改善率

分類A	件数	分類B	件数	併買 件数	改善率	分類A	件数	分類B	件数	併買 件数	改善率
和風総菜缶詰	13	畜肉調理缶詰	43	7	436.7	その他駄菓子	154	米系スナック	72	18	56.6
イースト菌	21	強力小麦粉	46	9	324.9	ペットボトル入り果汁100%飲料	44	パウチ入り清涼飲料	86	6	55.3
もち米	29	乾燥小豆	33	8	291.5	その他駄菓子	154	水あめ	21	5	53.9
するめ・乾燥イカ	15	乾燥エビ	100	11	255.7	缶入り無果汁炭酸飲料	202	缶入りコーラ飲料	278	83	51.5
畜肉調理缶詰	43	赤貝缶詰	22	6	221.2	紙パック入り野菜ジュース	69	紙パック入り緑茶飲料	50	5	50.5
ゆで小豆・練りあん缶詰	42	白玉粉	45	11	203	赤貝缶詰	22	イワシ缶詰	220	7	50.4
その他即席米飯・無菌包装米飯	100	即席釜飯	19	11	201.9	滋養強壮ドリンク	118	瓶入り無果汁炭酸飲料	84	14	49.3
ケーキ用チョコレート	32	ケーキ用ナッツ・フルーツ	45	6	145.3	サンマ缶詰	79	マグロ缶詰	46	5	48
カツオ油漬缶詰	21	缶入り緑茶飲料	89	7	130.6	その他ドライフルーツ	70	海藻珍味	115	11	47.7
その他レギュラーアイス	24	アイスキャンデー	62	5	117.2	ラー油	58	七味トウガラシ	65	5	46.3
紙パック入り健康茶飲料	116	紙パック入り緑茶飲料	50	19	114.2	揚げめん	134	なると巻	51	9	45.9
ピラフ・チャーハンの素(ウエット)	48	その他アジア・西洋香辛・調合調味料	35	5	103.8	その他冷凍中華総菜	90	冷凍ミートボール	68	8	45.6
烏卵缶詰	13	マッシュルーム缶詰	131	5	102.4	その他ちくわ	125	チーズ入りちくわ	68	11	45.1
お好み焼き・たこ焼きミックス	117	するめ・乾燥イカ	15	5	99.4	すき焼き用割り下	56	焼き豆腐	140	10	44.5
即席ワンタン	68	ハンバーグソース	26	5	98.6	畜肉調理缶詰	43	イワシ缶詰	220	12	44.2
お好み焼き・たこ焼きミックス	117	青ノリ	49	15	91.2	カツオ油漬缶詰	21	缶入りコーヒー飲料	263	7	44.2
その他生タイプ即席袋めん	54	生タイプ即席袋うどん	57	8	90.6	紙パック入り健康茶飲料	116	紙パック入りウーロン茶飲料	62	9	43.6
即席カップうどん	167	即席カップ日本そば	134	56	87.3	乾燥ひじき	137	大豆煮豆	41	7	43.5
その他スポーツ飲料	80	パウチ入り清涼飲料	86	17	86.2	中華そば用つゆ	98	生中華そば	543	66	43.3
お好み焼きソース	59	青ノリ	49	7	84.4	その他冷凍和風総菜	156	ハンバーグソース	26	5	43
お好み焼き・たこ焼きミックス	117	お好み焼きソース	59	16	80.8	ちくわぶ	19	はんぺん	346	8	42.4
どんぶりのつゆ	27	その他チルドスナック半製品	84	5	76.9	豆乳飲料	63	調整豆乳	171	13	42.1
練乳・コンデンスミルク	67	ゆで小豆・練りあん缶詰	42	6	74.4	ノリつくだ煮瓶詰	126	その他漬物瓶詰	33	5	41.9
その他アジア・西洋香辛・調合調味料	35	チキンステーキ・チキンロール	68	5	73.3	風船ガム	200	水あめ	21	5	41.5
お好み焼き・たこ焼きミックス	117	その他水産乾物	29	7	71.9	タラ珍味	51	イカ珍味	203	12	40.4
その他畜産珍味	12	イカ珍味	203	5	71.6	かき氷用シロップ・米みつ	124	ゆで小豆・練りあん缶詰	42	6	40.2
練乳・コンデンスミルク	67	かき氷用シロップ・米みつ	124	17	71.4	アイスモナカ	22	カップアイス	198	5	40
ハンバーグソース	26	ステーキソース・バーベキューソース	94	5	71.3	冷凍エビフライ	26	その他チルド和風総菜半製品	168	5	39.9
ホットケーキミックス	85	ケーキシロップ	41	7	70	乾燥刻みコンブ	12	板コンニャク	520	7	39.1
その他冷凍和風総菜	156	冷凍エビフライ	26	8	68.8	おろしニンニク	85	おろしショウガ	205	19	38
コーンアイス	18	カップアイス	198	7	68.5	その他ガム	101	風船ガム	200	22	38
お好み焼き・たこ焼きミックス	117	その他農産乾物素材	22	5	67.7	ピザソース・ピザトーストスブレッド	116	シュレッドナチュラルチーズ	354	44	37.4
お好み焼きソース	59	乾燥エビ	100	11	65	その他ホールスパイス	65	カレー粉	73	5	36.7
生タイプ即席袋スパゲティ	87	卵料理の素	31	5	64.7	その他レギュラーアイス	24	カップアイス	198	5	36.7
ピラフ・チャーハンの素(ウエット)	48	チキンステーキ・チキンロール	68	6	64.1	紙パック入り無果汁清涼飲料	123	紙パック入り健康茶飲料	116	15	36.7
するめ・乾燥イカ	15	天かす・揚げ玉	221	6	63.1	缶入りウーロン茶飲料	76	缶入り緑茶飲料	89	7	36.1
乾燥エビ	100	青ノリ	49	8	56.9	めん用の具	104	焼き豚	243	26	35.9
お好み焼き・たこ焼きミックス	117	乾燥エビ	100	19	56.6	缶入り緑茶飲料	89	缶入りコーヒー飲料	263	24	35.8

※線で囲った部分は併買件数が12件以上の組み合わせ

8. 3 多次元尺度法によるバスケット分析の簡便法

バスケット分析の場合、あらゆる商品アイテムの併買パターンを調べる際には膨大な計算量を必要とする。そのため分析したいアイテムが多い場合は計算能力や計算時間の限界に達することがあり、実用的でなくなる可能性がある。

多次元尺度法 (Multidimensional Scaling、以下MDS) は対象間の距離 (非類似度) に基づいて、対象の親近性を視覚化可能な2~3次元 (平面ないし空間) に散布図で表現する手法である。今回のバスケット分析で得られている信頼度を、商品間の類似性とみなすことで、バスケット分析のデータはMDSでつかうことができる。

MDSでは一対の関係しか扱えないが、比較的多くの商品を扱うことができ実用的である。さらにMDSによって併買地図を描くことで、視覚的な関係を理解することができる。分析の初期段階でMDSで得られる知見は、バスケット分析を行う前のヒントにもなるだろう。

8. 3. 1 カレールーとの関連

目的: カレールーと信頼度の高かった分類を取り出し、関連性 (併売されているか) を平面上に表現する。さらに、他分類間の併買の類似性についても明らかにする。

データ: 1998年10月1日~1999年9月30日まで (1年分) のパネル購買データ

対象: 上記のデータを Clementine の Apriori で分析し、カレールーおよびカレールーとの信頼度が高い上位21分類と、その21分類を目的商品とした時に信頼度の高い分類12分類、計34分類を選出した。(表8-15、併買商品出現件数24件未満、同時併買件数5件未満の分類は削除している)

手順: SAS で表8-15の34分類について二分類間の信頼度を表8-14のマトリクスにする。信頼度による類似度行列は対称ではないので、下記の3パターンの対類似度行列を作成した上で分析をした。

- ・表8-14の対角線の右上三角部分の関係 (以下「上三角」とする。結果は図8-1)、
- ・表8-14の対角線の左下三角部分の関係 (以下「下三角」とする。結果は図8-2)、
- ・上下三角の信頼度の平均値でプロットしたもの (結果は図8-3)
なお、対角要素は含まない。

なお、分析には SAS のMDSプロシジャを使った。類似性としての信頼度は比率として測定されているものだが、今回は順序尺度として取り扱った。したがって、距離の「近い」「遠い」関係において関連性を解釈する。最適化アルゴリズムは「Kruskalの公式1」である。

表 8-14 二分類間の信頼度のマトリクス表

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32	V33	V34				
V1	100	13.54	1.47	0	0.1	0	0.1	0.1	0	0.1	0.1	0	0	0.1	0.1	0.69	0.2	0.2	0.49	10.3	1.57	0	0	1.18	0	0	0.59	0.2	0.2	0.29	0	0.69	0.1	0				
V2	7.14	100	2.43	0.05	0.21	0.05	0.57	0.1	0.05	0.05	0.21	0	0	0.16	0.47	0.57	0.62	0.88	0.31	28.8	1.76	0.05	0.62	1.34	0.36	0.05	0.1	0.16	0.21	0.88	0	0.83	0.05	0				
V3	4.34	13.58	100	0	0	0	0.87	0.29	0.29	0	0	0	0	0	0.29	0	0	0.29	0.29	7.23	1.45	0	0.58	0.58	0.87	0.29	0.58	0.29	0	1.16	0	1.45	0	0				
V4	0	1.89	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	1.89	0	11.32	3.77	0	0	1.89	0	0	0	0	0	0	0	1.89	0	0				
V5	0.89	3.57	0	0	100	0	0.89	0	0	0.89	0	0	0	0	0	0	0	0	0	13.39	2.68	0	0	0.89	0	0	0	0	0.89	0	0	0	0	0	0			
V6	0	1.69	0	0	0	100	0	0	0	0	0	0	0	0	0	0	1.69	1.69	1.69	13.56	3.39	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
V7	0.26	2.83	0.77	0	0.26	0	100	2.57	0	0	0.26	0	0	0	0.26	0.51	0.77	0.51	0.77	4.37	3.08	0.26	0.26	0.51	0	0.26	1.29	0	0.51	24.42	0.26	0	0	0.26				
V8	2.38	4.76	2.38	0	0	0	23.81	100	0	0	2.38	0	0	2.38	0	0	0	0	0	11.9	0	0	0	2.38	0	0	2.38	0	0	23.81	0	2.38	0	2.38	0	2.38		
V9	0	1.56	1.56	0	0	0	0	0	100	0	0	0	0	0	1.56	0	0	0	0	14.06	0	0	0	0	0	0	0	1.56	0	1.56	3.13	10.94	0	0	0			
V10	2.86	2.86	0	0	2.86	0	0	0	0	100	0	0	0	0	2.86	8.57	0	2.86	20	8.57	0	0	17.14	0	0	0	0	8.57	0	0	0	0	0	0	0			
V11	0.47	1.89	0	0	0	0	0.47	0.47	0	0	100	0.47	0	2.36	3.3	3.3	3.77	0.47	0	13.68	3.3	0	0.47	0.94	0	0	0.47	0	0	0.47	0.47	0	0	0.47	0			
V12	0	0	0	0	0	0	0	0	0	0	4.17	100	29.17	0	4.17	4.17	0	0	4.17	12.5	4.17	0	0	0	0	0	0	0	0	4.17	0	0	0	0	0			
V13	0	0	0	0	0	0	0	0	0	0	0	25.93	100	0	3.7	7.41	0	0	0	18.52	7.41	0	0	0	0	0	0	0	0	3.7	0	3.7	0	0	0			
V14	0.81	2.42	0	0	0	0	0	0.81	0	0	4.03	0	0	100	1.61	2.42	8.06	1.61	0.81	11.29	2.42	0	0	0.81	0	0	0	0	0	2.42	1.61	0.81	0	0	0			
V15	0.42	3.8	0.42	0	0	0	0.42	0	0.42	0	0.42	0.42	0.84	100	1.27	2.11	1.27	0	6.33	2.53	0.42	0	0.42	0	0	0.84	0	0.84	0.84	0	1.27	0.42	0	0	0			
V16	2.35	3.69	0	0	0	0	0.67	0	0	0.34	2.35	0.34	0.67	1.01	1.01	100	7.05	1.68	15.1	6.04	0	1.34	4.03	0	1.01	0	0.34	1.34	0.67	0	0.67	0	0	0	0			
V17	0.53	3.2	0	0	0	0.27	0.8	0	0	0.8	2.13	0	0	2.67	1.33	5.6	100	0	1.33	10.4	4.8	0	0.53	4	0	0	0.53	0	0.53	0.8	0.27	1.07	0	0	0			
V18	1.39	11.81	0.69	0.69	0	0.69	1.39	0	0	0	0.69	0	0	1.39	2.08	3.47	0	100	2.08	22.22	9.72	0	0.69	4.17	0.69	0.69	0.69	2.78	2.08	0.69	0	2.78	0	0.69	0			
V19	2.3	2.76	0.46	0	0	0.46	1.38	0	0	0.46	0	0.46	0	0.46	0	2.3	2.3	1.38	100	8.29	10.14	0	0	7.83	0	1.38	0	0.46	0.92	1.38	0	0.46	0	0	0			
V20	2.1	11.16	0.5	0.12	0.3	0.16	0.34	0.1	0.18	0.14	0.58	0.06	0.1	0.28	0.3	0.9	0.78	0.64	0.36	100	5.41	0.08	0.32	1.24	0.14	0.14	0.38	0.14	0.58	0.46	0.1	0.7	0.04	0.14	0			
V21	0.78	1.66	0.24	0.1	0.15	0.1	0.58	0	0	0.15	0.34	0.05	0.1	0.15	0.29	0.88	0.88	0.68	1.07	13.15	100	0.05	0.19	3.51	0.24	0.1	0.58	0.19	0.29	0.58	0	1.02	0.29	0.15	0			
V22	0	2.86	0	0	0	0	2.86	0	0	0	0	0	0	0	2.86	0	0	0	0	11.43	2.86	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
V23	0	11.88	1.98	0	0	0	0.99	0	0	0	0.99	0	0	0	0	3.96	1.98	0.99	0	15.84	3.96	0	100	26.73	0.99	0.99	0	0	2.97	0	0	0.99	0	0	0	0		
V24	1.79	3.87	0.3	0.15	0.15	0	0.3	0.15	0	0.89	0.3	0	0	0.15	0.15	1.79	2.24	0.89	2.53	9.24	10.73	0	4.02	100	0.15	0.89	1.04	2.38	2.09	0.6	0	0.6	0.15	0	0			
V25	0	10.61	4.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10.61	7.58	0	1.52	1.52	100	0	0	1.52	0	0	1.52	0	0	0	0	0		
V26	0	2.13	2.13	0	0	0	2.13	0	0	0	0	0	0	0	0	0	6.38	0	2.13	6.38	14.89	4.26	0	2.13	12.77	0	100	10.64	6.38	0	0	0	0	0	0	0		
V27	2.51	0.84	0.84	0	0	0	2.09	0.42	0	0	0.42	0	0	0	0.84	0	0.84	0.42	0	7.95	5.02	0	0	2.93	0	2.09	100	1.67	0.42	0.84	0	1.26	0	0	0	0		
V28	3.23	4.84	1.61	0	0	0	0	0	1.61	0	0	0	0	0	0	1.61	0	6.45	1.61	11.29	6.45	0	0	25.81	0	4.84	6.45	100	0	0	0	0	0	0	0	0		
V29	1.68	3.36	0	0	0.84	0	1.68	0	0	2.52	0	0	0	0	1.68	3.36	1.68	2.52	1.68	24.37	5.04	0	2.52	11.76	0.84	0	0.84	0	100	1.68	0	0.84	0	0	0	0		
V30	0.67	3.77	0.89	0	0	0	21.06	2.22	0.22	0	0.22	0.22	0.22	0.67	0.44	0.44	0.67	0.22	0.67	5.1	2.66	0	0	0.89	0	0	0.44	0	0.44	100	0	1.11	0	0.22	0	0		
V31	0	0	0	0	0	0	2.38	0	4.76	0	2.38	0	0	4.76	0	0	2.38	0	0	11.9	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0		
V32	0.99	2.25	0.7	0.14	0	0	0	0.14	0.99	0	0	0	0	0.14	0.14	0.42	0.28	0.56	0.56	0.14	4.93	2.96	0	0	0.56	0.14	0	0.42	0	0.14	0.7	0	100	0.14	0	0		
V33	1.82	1.82	0	0	0	0	0	0	0	0	0	0	0	0	1.82	0	0	0	0	3.64	10.91	0	0	1.82	0	0	0	0	0	0	0	0	1.82	100	0	0	0	
V34	0	0	0	0	0	0	2.22	2.22	0	0	2.22	0	0	0	0	0	0	0	2.22	0	15.56	6.67	0	0	0	0	0	0	0	0	0	0	2.22	0	0	0	100	0

※ 数値はA p r i o r iによる、各分類間の〔縦軸の分類を買った時、同時に横軸の分類を買う〕場合の信頼性(%)。

表8-15 MDS用選定分類一覧

NO	分類名	分類出現 件数	カレーとの 信頼度(%)	NO	分類名	分類出現 件数	カレーとの 信頼度(%)
V1	ラッキョウ漬 (福神漬)	1019	10.3	*V21	シチュールー	2054	13.15
*V2	福神漬	1934	28.8	*V22	その他電子レンジ食品	35	11.43
V3	シソ漬 (福神漬)	346	7.23	*V23	グリーンピース缶詰	101	15.84
*V4	ワカメ佃煮	53	11.32	V24	マッシュルーム缶詰 (シチュールー)	671	9.24
*V5	チルドコロッケ	112	13.39	V25	カツオ油漬缶詰 (福神漬)	66	10.61
*V6	チルド香辛料	59	13.56	*V26	アサリ缶詰	47	14.89
V7	乾燥エビ (カツオ削り粉)	389	4.37	V27	ホタテ缶詰 (アサリ缶詰)	239	7.95
*V8	カツオ削り粉	42	11.9	*V28	鳥卵缶詰	62	11.29
*V9	角砂糖	64	14.06	*V29	その他漬物瓶詰	119	24.37
*V10	調理用ワイン	35	20	V30	お好み焼きミックス (カツオ削り粉)	451	5.1
*V11	カレーパウダー	212	13.68	*V31	ドクダミ茶	42	11.9
*V12	カラシ粉	24	12.5	V32	クリーミングパウダー (角砂糖)	710	4.93
*V13	わさび粉	27	18.52	V33	袋入りバームクーヘン (シチュールー)	55	3.64
*V14	ニンニク粉	124	11.29	*V34	冷凍シーフードミックス	45	15.56
V15	ラー油 (カレーパウダー)	237	6.33				
*V16	その他ホールスパイス	298	15.1				
V17	その他パウダースパイス (その他ホールスパイス)	375	10.4				
*V18	アジア系調合調味料	144	22.22				
V19	デミグラスソース (シチュールー)	217	8.29				
V20	カレールー	4989	-				

※ * 印のついた分類は、カレーとの信頼度が高い併買分類、印のない分類は () の分類との信頼度が高い分類
※併買商品出現件数24未満、同時併買件数5未満の分類は含まない

図8-1を見ると、「カレールー」と「福神漬」との距離よりも「カレールー」と「シチュールー」との距離が近く、必ずしも信頼度の高さと同じ結果にはならない。

しかしながら、「カレールー」と信頼度の近かった「福神漬」「アジア系調合調味料」「その他漬物瓶詰」などが比較的近接している傾向が見られる。

「カレールー」以外では、「マッシュルーム缶詰」と「グリーンピース缶詰」が近接している。一方で「ホタテ」「アサリ」「鳥卵」の缶詰が近接しており、缶詰で2グループの形成が見られる。

同様に「香辛料」のグループ、「お好み焼きミックス」-「カツオ削り粉」-「乾燥エビ」というお好み焼きメニューグループも形成されており、メニュー型商品分類の購買行動が表現されている。

次に、下三角部分のデータも同様に見る(図8-2)と、図8-1が右回りに90度回転したような布置をとった。前述した缶詰グループやお好み焼きグループの分類の近接がここでも見られる。「カレールー」と最も近接しているのは「その他漬物瓶詰」。「カレールー」との信頼度があまり高くない「シソ漬」「ラー油」などの距離が近くなっている。これらの分類はそれぞれ「福神漬」や「カレーパウダー」との信頼度が高い分類であることが起因しているであろう。

さらに、上下三角の信頼度の平均値でプロットしたもの(図8-3)については、距離の順序関係は図8-1に近くなる。

いずれのプロットも「カレールー」と信頼度の高い分類についてはその順序は異なるが「カレールー」と近接している。また、「香辛料」「お好み焼きミックスとその材料」、「缶詰・瓶詰」間が近距離に布置される。

このように、併買分類から見たカレールーだけではなく、カレーから見た併買分類の情報も重要である。MDSによって双方向の関係を見ることで、メーカーにとっては自社の商品カテゴリーだけではなく、併買商品との相乗効果がねらえることを店に訴える材料となる。

8. 3. 2 全分類によるMDS

次に、1年分のデータで出現する食品940分類すべての同時購買信頼度のマトリクス表(940列×940行)を作成し、MDSを行った。

なお、ここでもカレールーとの関係を見ることにした。8. 3. 1でのマトリクス表の利用方法では、ある商品分類からは「併買分類〇〇を買ったらカレールーを買う」信頼度、また別の分類からは「カレールーを買ったら併買分類〇〇を買う」信頼度というように、条件の異なる値をひとつの散布図に図示していることになってしまう。そこで、カレールーを1番はじめの変数とし、以降を分類番号順に配置したマトリクス表にした。

940分類すべてをプロットしたものが図8-4。信頼度を順序尺度にしているためそれぞれの変数が等間隔にちらばり大きな円形になる。(これを間隔尺度にすると、牛乳や豆腐など、出現数の多い日配食品分類が外れ値となって飛び出た形になった。)

図8-4の円形からカレールーの位置周辺をクローズアップしたものが図8-5(上三角部分のデータ使用)、図8-6(下三角部分のデータ使用)である。

図8-5については、「福神漬」との距離が遠くなっている。また、他の分類についても説明可能な関連性が見いだせなかった。

一方、図8-6を見ると、「カレールー」と「福神漬」が近接している。これを見ると信頼度を反映しているようである。しかしながら、福神漬の次に信頼度の高かった「その他漬物瓶詰」や「シチュールー」などの分類は図の範囲外になってしまう。

この図8-6からは周辺分類の類似性に注目できる。例えば、図の左上側にチーズとヨーグルト、乳飲料など、通常スーパーで隣接されていることの多い乳製品売場の商品分類同士の距離が近い。同様の現象は豆腐、納豆類や「チルドシューマイ」「チルド中華まんじゅう」など中華系総菜にも見られる。一方で、同じ乳飲料でも、「フルーツ乳飲料」だけが図の右下に布置され、「調製豆乳」と近接している。図では「カレールー」を中心においたものだが、より全体を見渡した際に、新たな併買パターンを発見できそうだ。

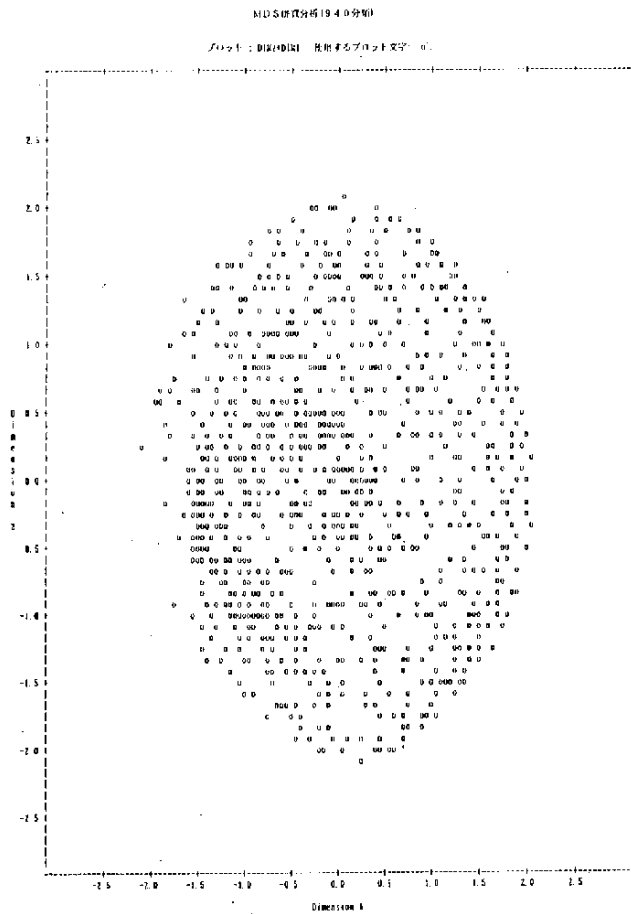


図8-4 MDSによる併売状況プロット (940変数全体)

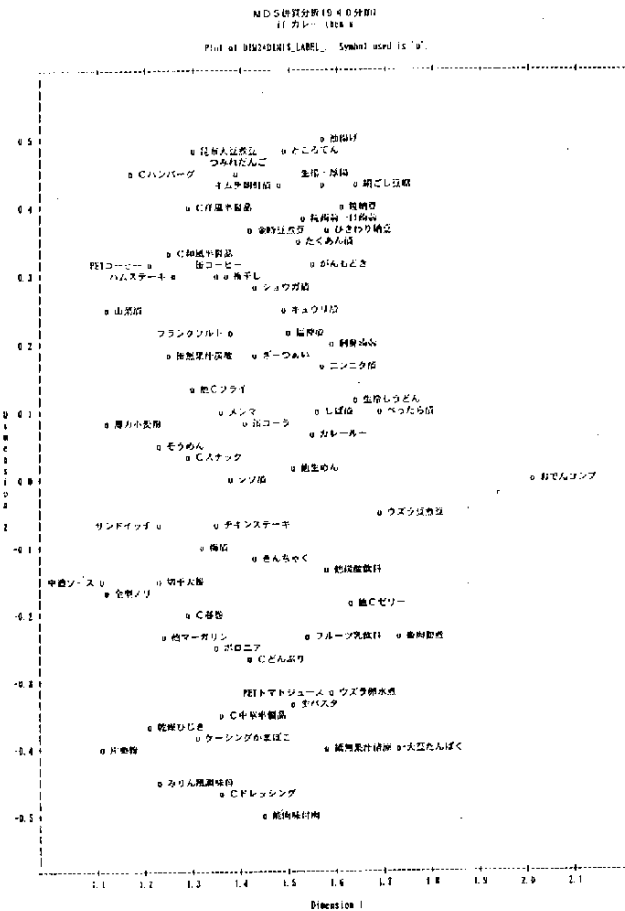


図8-5 MDSによる併売状況プロット (940変数カレールー近辺・上三角)

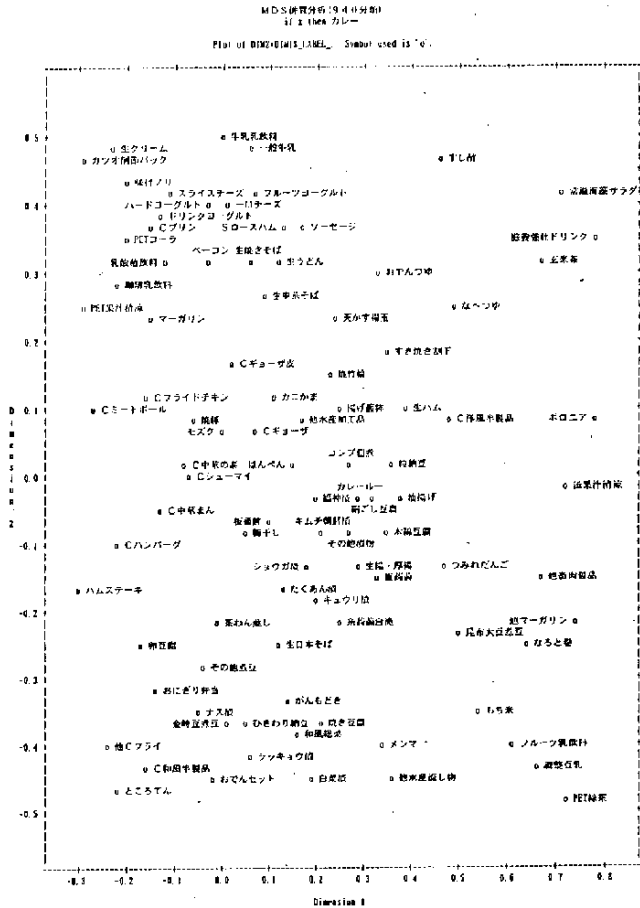


図8-6 MDSによる併売状況プロット (940変数カレールー近辺・下三角)

※図8-5、8-6はカレールーの位置が大体中央になるように位置を指定した。

8. 3. 3 非対称MDS (*4)

8. 3. 1で行った、非対称のマトリクス表(表8-14)を上三角、下三角と分ける、あるいは信頼度の平均をとったものでは、カレーから見た福神漬、あるいは福神漬から見たカレーのうち、片側からの類似性を布置しているため、解釈のできるものとできないものが現れた。そこで、非対称性をそのまま非対称の情報としてモデルに組みこんだ、非対称MDSでの分析を試みた。

今回のデータは表8-14のマトリクス表をそのまま用いた。なお、非対称MDSは研究途上にあり、SASなどの商用ソフトウェアではまだ実行できない。ここでの分析作成には立教大学教授、岡太彬訓氏のご協力をいただいた。この場を借りてお礼を申し上げる。

非対称MDSでは3次元解を採用した。

横軸 DIM1-縦軸 DIM2 (図8-7)

横軸 DIM2-縦軸 DIM3 (図8-8)

横軸 DIM1-縦軸 DIM3 (図8-9)

<非対称MDS・結果考察>

岡太-今泉氏のアルゴリズムによる非対称MDSでは、座標を中心とする円の大きさを非対称性の解釈をする(*5)。図8-7に関して実際に円を描いた図を図8-10に示す。今回は布置の解釈のみを述べておく。

いずれの図においても、カレールーがほぼ中央に布置された。

カレーを中心に置く処理を施してはいないが、カレーとの関連性の高い分類を選択していることが作用しているようだ。

・カレールーとの関係

図8-7を見てみる。Aprioriの結果では「福神漬」を買うときの「カレールー」を買う信頼度は「シチュール」のそれに比べて大きかった。にもかかわらず、図上では、「カレールー」-「福神漬」間の関係は「カレールー」-「シチュール」間よりも遠い位置に布置されている。カレールーとシチュールはこの34変数の中では互いにとって良好な関係を持っていると言える。

図8-8、8-9ではDIM2が横軸になっているので、図8-7と縦横の関係が入れ替わった形に類似している。「福神漬」よりも「アジア系調合調味料」が近接した。

(*4) 非対称MDS (多次元尺度法)

複数の対象の相互間の類似度や親近度が与えられた場合、それらの関係をできる限り満たすような対象の多次元空間の位置を推定し、対象間の親近関係や類似度・親近度のゆがみの関係をわかりやすく図示する方法。従来の多次元尺度法が対象間の対称な関係のみを問題にするのに対して、対象間の非対称な関係の背後にある複雑な構造をわかりやすく整理し表示する点に特徴がある。

(千野直仁氏「非対称多次元尺度構成法」より)

(*5) 上記「非対称多次元尺度構成法」を参照。

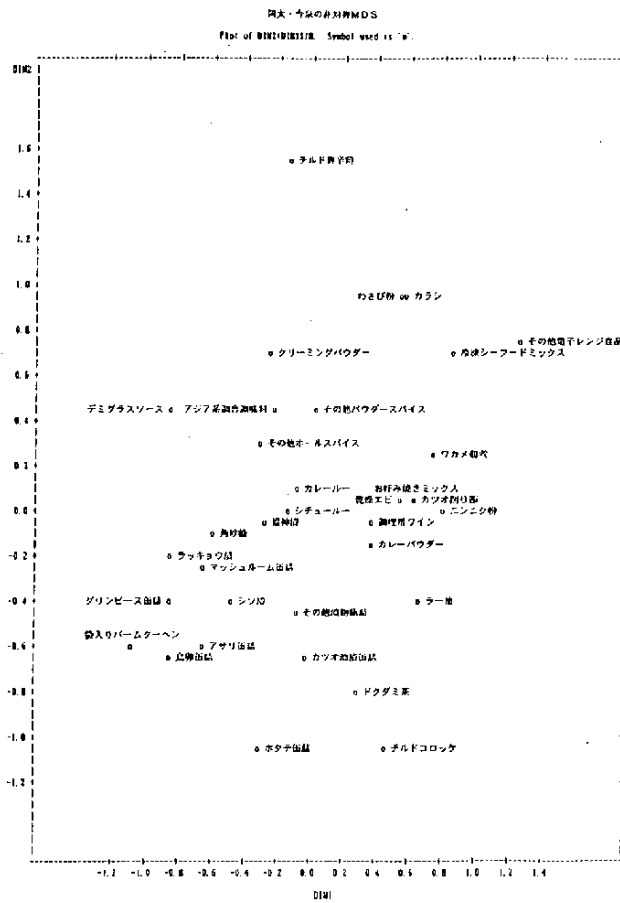


図8-7 非対称MDS・DIM1*DIM2

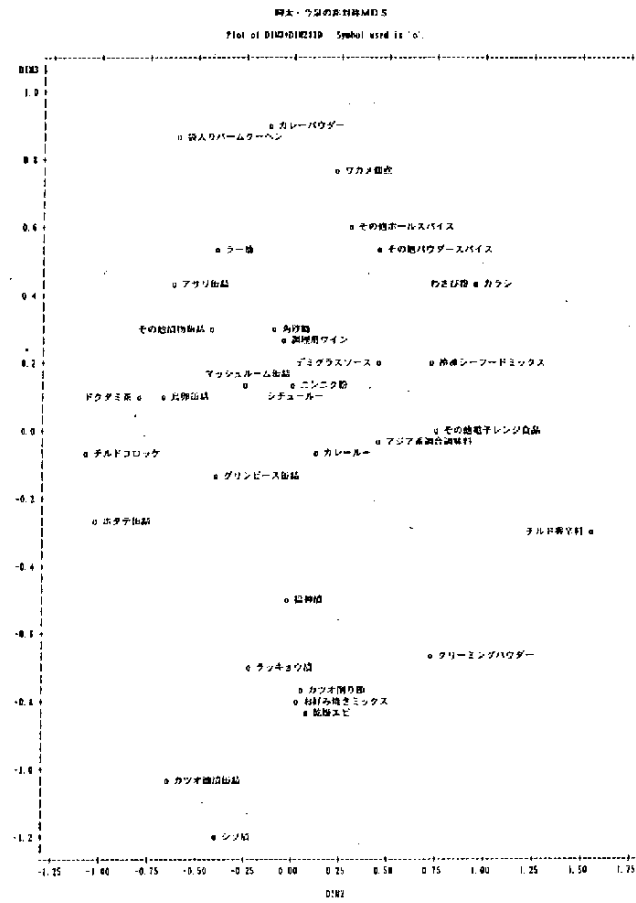


図8-8 非対称MDS・DIM2*DIM3

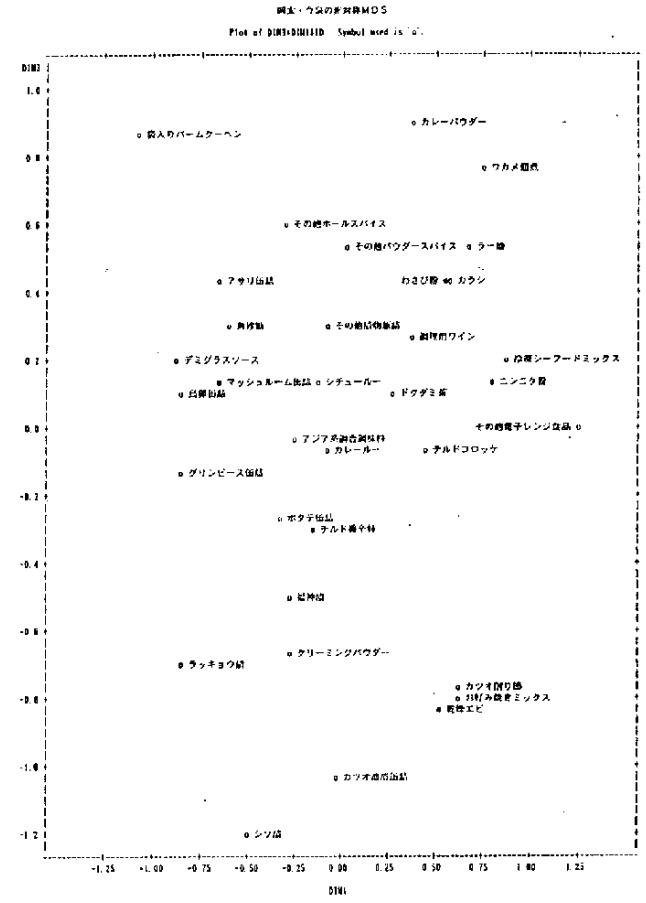


図8-9 非対称MDS・DIM1*DIM3



図8-10 非対称MDS・DIM1* DIM2

・周辺分類の関係

3つの散布図いずれでも「お好み焼きミックス」と「乾燥エビ」と「カツオ削り粉」が近接している。

表8-15のAprioriの結果では「カレールー」との信頼度は「カツオ削り粉」を除き、あまり高くはないものの、図8-7からは「お好み焼きミックス」、「乾燥エビ」、「カツオ削り粉」の順にカレーの近くに並んだ。この3分類は他の二軸から見ても近接しており、より高い購買の類似性を示している。「わさび粉」と「カラシ粉」、「その他ホールスパイス」と「その他パウダースパイス」も同様の傾向を示している。

さらに大きく捉えると、漬物類、缶詰・瓶詰、調味料類など、通常同一区画の売場に陳列されている分類のものが近接している。

結果的には図8-7の散布図がカレーとの関係および周辺分類の関係においても、もっとも信頼度に近い形になっている。DIM3はカレーとの関連を捉える軸ではなく、むしろ周辺分類の関係を解釈するのが適した軸であると言えよう。

MDSですべての分類を見る場合は、実際スーパーで売られているすべての商品分類の併買関係をプロットした時に出てくる併買の関連性を発見することで、売場陳列に役立つ情報となる。また、8.3.1のようにあらかじめ目的分類を限定する場合は、本来の目的としている分類との関係が明らかになる。例えば、あるメーカーが売り出している全カテゴリーだけをピックアップし、MDSで関係を見ることで、メーカー内で製造している商品の販売戦略の検討材料となる。MDSによる分析結果はバスケット分析の簡便法として、様々な利用法がありそうだ。

8.4 課題と発展

以上、バスケット分析について、多方面から試みた。分析の途中で様々な課題に直面し、新たな方向からの分析を試みる必要性が生じた。今後新たな結果を導くべく、最後にバスケット分析の課題と分析の発展形について記しておきたい。

(1) 大容量データベース対応

バスケット分析用のデータは、とにかく膨大な量である。とりわけPOSデータ、パネルデータなどのテラデータベースはそのハンドリングを容易にすることが大きな課題である。実際、約100MBのデータを使用していて、夜実行をスタートさせたが、分析が朝になっても終わらず、やがてハードウェアがハングアップしてしまうという状況にしばしば悩まされた。

分析をスムーズに行うためには、データアクセスおよび利用データのセグメントが容易にでき、かつ大容量データの分析に耐えられるシステム構築が求められる。

(2) 出現率の差を克服する

目的商品と併買商品、または併売される商品の組み合わせの信頼度を見る際に、双方の出現率の差が小さいほど、信頼度は有効なものになってくる。

日配品のような出現数の多い分類に結果がひきづられない変数設定を可能にすることが重要である。

そのためには出現率の高い分類は8. 2. 3で行ったような、新しい変数の設定によってブレイクダウンし、出現率の低い分類は同じような商品特性のものをまとめて変数とする、など出現率の値を近づけていくことで、さらに結果が見つけやすくなる。

(3) 発展・時系列バスケット分析

例えば、〔6月にA社の栄養補助ドリンクとB社の健康茶飲料を併買した〕人のそれ以降の飲料併買動向など、バスケット分析でも時系列にデータを見ていくことによって、リピート購入、ブランドスイッチなどの変遷などを明らかにできる。

(4) 発展・パネル属性データの活用

併買の相関ルールが見つかったら併買要因分析を行い、購入の背景を探ることもマーケティングにとっては重要である。8. 2. 2で月と曜日のコーザル変数を用いたが、さらにパネルの属性データを使い、〔カップ入り即席中華そばとペットボトル飲料を同時に購入するのは主婦の年齢が30代後半で6～12歳の子供がいる世帯である〕というような購買者層を把握することでより詳細な店舗マーケティングデータへの発展へ繋がるだろう。

9. まとめ

本報告書に収録した6つのデータマイニング事例（第3章から第8章）には、以下のような共通した特徴がある。

1. 企業にとっての具体的な問題・課題をとりあげている
2. 実際のPOS（PANEL）データを使用して分析している
3. 一般に普及している商用のデータマイニングツールを利用している
4. うまくいった結論だけではなく全ての分析過程・手順を示している

データマイニングは大規模データベースを扱いながら企業の具体的課題を解決するための方法だから、小さな仮想的データを使っても現実味が少ない。実際の問題解決の過程を示すことが本報告の主目的であった。従って、必ずしも見本となるような美しい分析事例にはなっていない部分もあるが、現実の問題を実際のデータを使って実行した記録となっている。

POSのような巨大なデータベースを対象にデータマイニングを実施できる機会は少ないので、その意味でも貴重な公表事例集だといえる。マイニングツールを使用して気づいた点も盛り込んでおり、これからデータマイニングに実際に取り組もうとする人々に何らかの参考となる情報を提供しているはずである。

今回のプロジェクトを振り返り、データマイニングに関する若干の印象をまとめておきたい。

9. 1 マイニング手法と実行環境

データマイニングで利用される分析手法は伝統的な多変量解析法も含まれるが、本報告書に収録した事例で使われた手法は、

- ニューラルネットワーク
- デシジョンツリー（CHAID、C5.0、C&RT）
- マーケット・バスケット分析
- MDS（多次元尺度構成法）
- 因子分析や重回帰分析など伝統的な線形多変量解析法

などである。これらはかなり高度な手法なので専門家でなければ詳細まで理解して使うことは困難であるが、現在のマイニングツールはGUIで実行できる環境を装備しているので、初心者でも比較的簡単に利用できる。

しかし利用環境が便利になってさえいけば、高度な手法を使いこなすことができるというわけではない。分析の途中でしばしば困難な壁にぶつかることがあるが、それが分析手法を適切に適用していないのが原因なのか、データそのものに原因があるのか、あるいはその複合した結果なのかを見極めることが難しい。

特にニューラルネットワークは計算過程がブラックボックスであり、予測に役立つだろうと了解はできても、何をやっているのか理解・納得するのは簡単でない。ニューラルネットワークに関する解説書がいくつも出版されており、一般論として学習する環境は整っているが、実際のデータマイニングで使っている SAS 社や SPSS 社が採用しているニューラルネットワークのアルゴリズムや、分析を成功させるためのガイドラインなどに関しては良いマニュアルが整備されていない。同じデータを同じ条件で別のニューラルネットワークのプログラムで実行すると結果が異なることがあり、どこか納得の行かない感覚が残ってしまう。特に日本語のドキュメントは操作説明書しかない。本報告では実例を示すところまで到達しているが、手法に関して理論的な解説を加えることまではしていない。

しかし、繰り返し強調するように、データマイニングを実施するソフトウェア環境が整備されていることは事実であり、SAS 社や SPSS 社の製品がなければ本報告書に示したような分析はほとんど不可能である。

9. 2 データマイニングを実施するための条件

企業における人材問題として「業務知識やデータ知識はあるが、統計学やデータ解析の知識・経験がない者」と「統計的データ解析の技術・知識はあるが POS データや業務の知識は乏しい者」という、いわば業務派（前者）と解析派（後者）が別々の個人となっていることが多いという話をよく耳にする。実際のデータ解析はかなり経験を積まないと簡単にはできないものであるが、SAS 社の Enterprise Miner や SPSS 社の Clementine などのマイニングツールは、業務派（データ解析の素人）であってもすぐに操作ができるようになる。この点に関してはソフトウェア会社の宣伝はそれほど過大ではないと考えられる。従って、それだからこそ入門後・中級レベルの解説マニュアルがぜひとも望まれる。すぐに操作はできるようになるので、ただちに実際の具体的な応用問題に直面してしまうのである。現状では、マニュアルには書いてないが使ってみて調べて初めて明らかになった——というような属人的ノウハウに依存する部分もある。

データマイニングを実施するための条件として、マイニングすべきデータが手元にあることは前提とすれば、

- 分析するシステム環境（ハードウェア、ソフトウェア、データウェア）
- 分析する分析者能力と、組織的な分析環境の整備

の両方が必要である。

現在のコンピュータは UNIX のワークステーションで十分な能力を発揮でき、性能の高いサーバーがあればほとんどの処理ができる。今回は Windows のクライアントだけで分析する結果となったが、バスケット分析のように大量の計算を必要とする場合は、クライアント→サーバー環境を用意しなければ現実的ではなく、しばしば処理

不能にさえなる。また実際の業務の場合も、分析者であるエンドユーザーが直接データベースにアクセスして必要なデータを分析できる環境にすべきである。

分析者は業務派と統計派の両面を備えた人材であることが望ましい。しかし現実にはなかなか難しい要請なので、この条件が整わない場合は、SAS社やSPSS社のデータマイニングツールを使いながら業務派が自分でデータマイニングを実施し解析派に成長していくことが次善の戦略ではないだろうか。

対象としている業務の知識やデータに対する知識がなければ良い分析は不可能である。分析結果に対して直感的に「おかしいのではないか」と感じることをするために業務知識が必要である。そうでなければ、コンピュータの出力結果を信じて誤った判断をすることさえあるだろう。

またデータマイニングは大量のデータをさまざまな角度から分析するので集中力の継続が重要である。分析中はデータの変数名はもちろんデータの性質について具体的なイメージが定着している。しかし2、3日中断するとそれまでの分析成果を取り戻すことが難しい。従ってデータマイニングに専念できる環境を組織的に容認するような配慮も必要である。また長期間の集中力維持のためには、統計的データ解析のための知力だけでなく、体力も必要である。これらは些末なことのように聞こえるかも知れないが、実際の応用場面では切実な要請となるであろう。

9.3 予備解析の重要性

データマイニングツールを使用すれば確かに初心者でも何らかの成果を出力することができる。しかしデータマイニングの過程で重要なのは分析手法やモデルを適用することではなく、そこにいたるまでの予備解析でデータに対する知見を蓄積することである。そのような知見によって、どんな分析モデルが適切かという判断力も養成されるのである。

実際に費やされる時間の大半も予備解析であり、極端な場合は最後の分析モデルの適用は瞬時に計算されてしまうこともある。予備解析では時間をかけて分析者がさまざまなことを試行錯誤しているのである。そのため、よい結果を得るには最終的には分析者に「データ解析のセンスの良さ」が備わっているか否かという属人性が重要な鍵であるという実感を持った。高価なマイニングツールさえ導入すれば、そこにデータを入れると「宝」がポンと出てくる——ということなど到底あり得ない。

予備解析はデータの整理のような単純作業から始まって、分布状態や外れ値の検討などをグラフィカルな表示ツールを利用しながら進める。その意味ではEDA（探索的データ解析）の精神を共有している。データマイニングが扱うデータは、統計理論が扱う「母集団からの確率標本」といえるような計画的データではないが、統計学の成果・技術をかなり応用していることも事実であり、統計理論家と応用実務家が相互に積極的に関与する態度、また理論家は手法の性質について広く啓蒙活動を展開し、実務家はもっと理論を勉強するという態度が今後とも求められるであろう。

参考文献・資料

- Berry, M.A. and L. Gordon, *Data Mining Techniques*, Wiley, 1997.
マイケル J.A. ベリー / ゴードン リノフ 著
SAS インスティテュート ジャパン / 江原 淳 / 佐藤 栄作 共訳 「データマイニング手法」、海文堂, 1999
- 江原 淳 「マーケティング革新とデータマイニング」 「品質管理」 2000.2 月号
- 上田和勇 「イギリスの保険マーケティング」, 1998
- Inmon, W.H. et al., *Corporate Information Factory*, Wiley, 1998.
W.H. インモン / クローディア・インホフ / ライアン・ソーサー 著
江原 淳 / 藤野明彦 / 松永賢次 / 本江 渉 訳 「コーポレート・インフォメーション・ファクトリー」 海文堂, 1999
- Card, S.K. et al., *The Psychology of Human-Computer Interaction*, LEA, 1983.
- Fayyad, U.M. et al., eds., *Advances in Knowledge Discovery and Data Mining*, AAAI, 1996
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth: "From Data Mining to Knowledge Discovery" in 'Advances in Knowledge Discovery and Data Mining', AAAI Press, 1996
- J. R. キンラン 著 古川康一 監訳 「AI によるデータ解析」 トッパン, 1995
- Okada, A. (1988). *Asymmetric multidimensional scaling of car switching data*. In W. Gaul & M. Schader (Eds.), *Data, expert knowledge and decisions* (pp. 279-290). Berlin: Springer-Verlag.
- 千野直仁 「非対称多次元尺度構成法」 現代数学社, 1997

————— 禁無断転載 —————

平成12年 3 月発行

発行 財団法人 データベース振興センター
東京都港区新橋2丁目13番8号
新橋東和ビル5階
TEL 03-3508-2430

委託先 株式会社 日経リサーチ
東京都千代田区神田司町2-2-7
パークサイド1ビル
TEL 03-5296-5131

印刷所 株式会社 廣栄社
東京都千代田区内神田1-3-5 広栄ビル
TEL 03-3295-3545



