

平成 12 年度

学士学位論文

# POS データにおける標本相互相関に関する 研究

A Study on Sample Cross-Correlation in POS Data

1010371 岩城 広哲

指導教員 Ruck Thawonmas 助教授

2001 年 2 月 4 日

高知工科大学 情報システム工学科

# 要 旨

## POS データにおける標本相互相関に関する研究

岩城 広哲

流通業者にとって商品の発注を適正に行うことは、企業の競争力を左右する最大の課題である。従来は人間の経験則に基づいた数量の決定を行っていたため、在庫過小による販売機会の逸失や、在庫過大による経営の圧迫といった問題を抱えている。本研究はこうした問題点を解決するために過去の POS（販売時点情報管理）データを標本相互相関関数を用いて時系列解析を行い、商品間の因果関係を突き止め、将来の販売数量を最小限の在庫数量で実現しようとするものである。

**キーワード** POS, 標本相互相関関数, 時系列解析



# Abstract

## A Study on Sample Cross-Correlation in POS Data

Hiroaki Iwaki

In circulation business, proper goods ordering contributes to competition power of a certain company. Conventionally, decision of goods quantity is done according to human experience. However, this approach has severe problems. This is because sale chance is lost due to under stocking and company management is impaired by over stocking. This study attempts to solve these problems. Analysis of past POS data is conducted using sample cross-correlation function and results are given.

*key words* POS, Sample Cross-Correlation Functions, Time Series Analysis

# 目次

第 1 章	はじめに	1
第 2 章	流通	2
2.1	流通の概念	2
2.2	流通の情報化	2
2.2.1	情報化革命	2
2.2.2	POS システム	4
2.2.3	VAN	5
2.2.4	EOS	6
第 3 章	背景 (共同研究プロジェクト)	7
第 4 章	時系列	10
4.1	時系列解析	10
4.2	標本相関関数	11
4.2.1	標本自己相関関数	11
4.2.2	標本相互相関関数	11
4.3	IID ノイズ	12
第 5 章	実験	14
5.1	POS データからの Perl による目的データ抽出	14
5.1.1	POS データの形式	14
5.1.2	抽出過程	15
5.2	MATLAB による標本相互相関解析	17
第 6 章	結果	18

## 目次

6.1	掲載された結果グラフの説明 . . . . .	18
6.2	分類ごとの POS データ解析結果 . . . . .	19
6.2.1	POS __洗髪料における解析結果 . . . . .	19
6.2.2	POS __男性化粧品における解析結果 . . . . .	28
6.2.3	POS __化粧品における解析結果 . . . . .	31
6.2.4	POS __生理用品における解析結果 . . . . .	34
6.2.5	POS __ペットフードにおける解析結果 . . . . .	37
<b>第 7 章</b>	<b>結論</b>	<b>40</b>
	<b>謝辞</b>	<b>42</b>
	<b>参考文献</b>	<b>43</b>

# 目次

2.1	流通の役割 . . . . .	3
2.2	POS システム . . . . .	4
2.3	EOS システム . . . . .	6
3.1	共同研究パートナー . . . . .	7
3.2	共同研究テーマ概要 . . . . .	9
5.1	POS データ形式 . . . . .	14
5.2	POS データ抽出過程 . . . . .	15
5.3	解析商品組み合わせ . . . . .	16
6.1	B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	19
6.2	B1 の SACF . . . . .	20
6.3	B2 の SACF . . . . .	20
6.4	B1 と B2 間の SCCF . . . . .	20
6.5	B1 と B2 の散布図 . . . . .	21
6.6	B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	22
6.7	B1 の SACF . . . . .	23
6.8	B2 の SACF . . . . .	23
6.9	B1 と B2 間の SCCF . . . . .	23
6.10	B1 と B2 の散布図 . . . . .	24
6.11	B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	25
6.12	B1 の SACF . . . . .	26
6.13	B2 の SACF . . . . .	26
6.14	B1 と B2 間の SCCF . . . . .	26

## 図目次

6.15 B1 と B2 の散布図 . . . . .	27
6.16 B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	28
6.17 B1 の SACF . . . . .	29
6.18 B2 の SACF . . . . .	29
6.19 B1 と B2 間の SCCF . . . . .	29
6.20 B1 と B2 の散布図 . . . . .	30
6.21 B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	31
6.22 B1 の SACF . . . . .	32
6.23 B2 の SACF . . . . .	32
6.24 B1 と B2 間の SCCF . . . . .	32
6.25 B1 と B2 の散布図 . . . . .	33
6.26 B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	34
6.27 B1 の SACF . . . . .	35
6.28 B2 の SACF . . . . .	35
6.29 B1 と B2 間の SCCF . . . . .	35
6.30 B1 と B2 の散布図 . . . . .	36
6.31 B1 と B2 の売上グラフ (実線が B1, 波線が B2) . . . . .	37
6.32 B1 の SACF . . . . .	38
6.33 B2 の SACF . . . . .	38
6.34 B1 と B2 間の SCCF . . . . .	38
6.35 B1 と B2 の散布図 . . . . .	39

# 第 1 章

## はじめに

流通業者にとって商品の発注を適正に行うことは、企業の競争力を左右する最大の課題である。従来は人間の経験則に基づいた数量の決定を行っていたため、在庫過小による販売機会の逸失や、在庫過大による経営の圧迫といった問題を抱えている。こうした問題点を解決するために高知工科大学と(株)高知流通情報サービス及び、(株)サニーマートの3社と共同で「ニューラルネットワーク等による商品発注支援システム」の研究プロジェクトが発足した。本研究は上記プロジェクトの初期研究である。過去のPOS(販売時点情報管理)データを標本相互相関関数を用いて時系列解析を行い、商品間の因果関係を突き止め、将来の販売数量を最小限の在庫数量で実現しようとするものである。最終的には、本研究によって解析したPOS販売実績データ、及びEOS(自動受発注システム)発注実績データ、事象データをニューラルネットワークの自己学習機能を用いて適正発注ロジックを自己生成し、将来の商品の販売予測を行おうとするものである。

## 第 2 章

# 流通

### 2.1 流通の概念

私たちは、様々なものに囲まれて生活している。それらは、現在の社会では、大半を別の人が作り、それをお金を出して私たちが買い、使うというのが実状である。このようなことが可能になったのは、分業体制が進み、時間的、空間的に離れた場所で作られた物でも、お金との交換ができる貨幣経済の体制のおかげである [1]。このような体制で作った人（生産者）から使う人（消費者）のところへ物が流れていく過程を「流通」というのである。

具体的には、生産者・メーカーによって作られた製品を、卸売業者が買い取る。それを各地の小売業者に、販売し納品する小売業者は、仕入れた商品を消費者に販売する。これが経済活動の中で、卸売業者や小売業や物流業の受けもつ活動のすべてである。この商品の動く道筋を流通経路という。この流通チャンネルをひとつの線としてとらえ、生産者、メーカー側を川上、消費者側を川下といい、常に統合を繰り返している [2][3]。

### 2.2 流通の情報化

#### 2.2.1 情報化革命

現在、わが国では、流通業における情報化が急速な進展をみせている。いわゆる「流通情報革命」と呼ばれている状況がそれである。初めは流通企業によるコンピュータの導入というかたちであったが、利用は各種の計算業務が主で応用範囲は限られていた。流通活動そのものの情報システム化は、1970 年代後半から普及が始まる POS(Point Of Sales:販売時

## 2.2 流通の情報化

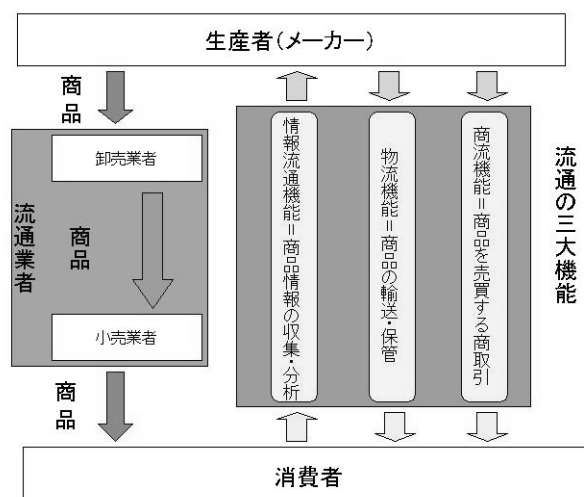


図 2.1 流通の役割

点情報)が口火になったといえる。それとともに、メーカーと卸売業者の間に VAN(Value Added Network:付加価値通信網)が構築され、商品情報や販売情報などが盛んに交換されるようになった。大規模小売店舗や FC(フランチャイズ・チェーン)、VC(ボランタリー・チェーン)の本部と卸売業者の間がコンピュータでつながれ、自動的に商品の補充発注が行える、いわゆる EOS(Electronic Ordering System:電子受発注システム)もすでに一般化しつつある。これはコンピュータの単体利用からネットワーク型システム化への展開であり、流通活動のあり方と流通システムを根底から変化させるものである。情報システム化は情報処理技術と情報伝達(通信)の双方の発展が結びつくことによって促進された。情報処理の技術の発展はハードウェアの発達とソフトウェアの進歩によって支えられている。ハードは大量化・高速化、すなわちスーパーコンピュータ化の方向と、小型化・軽便化という2つの方向で発展した。ワープロやパソコン、ワークステーションその他、各種端末の出現と普及は後者の例であるが、これらが相互に接続されることによりネットワーク化が進んだ。以下においてPOS及びVANの解説を行う。



## 2.2 流通の情報化

### 2.2.2 POS システム

POS とは自動読みとり装置を使い、単品ごとの売上の情報を、コンピュータを使って集積し、経営に役立てるシステムのことである。通産省の定義によれば、「光学式自動読み取り方式のレジスターにより単品別に収集した販売情報や、仕入れ・配送などの段階で発生する各種情報をコンピュータに送り、各部門がそれぞれの目的に応じて有効利用できるような情報に処理、加工、伝達するシステム」ということになる。その仕組みは図 2-2 のようになっている。商品に付いているバーコードを、レジスターと連動した POS 端末で読みと

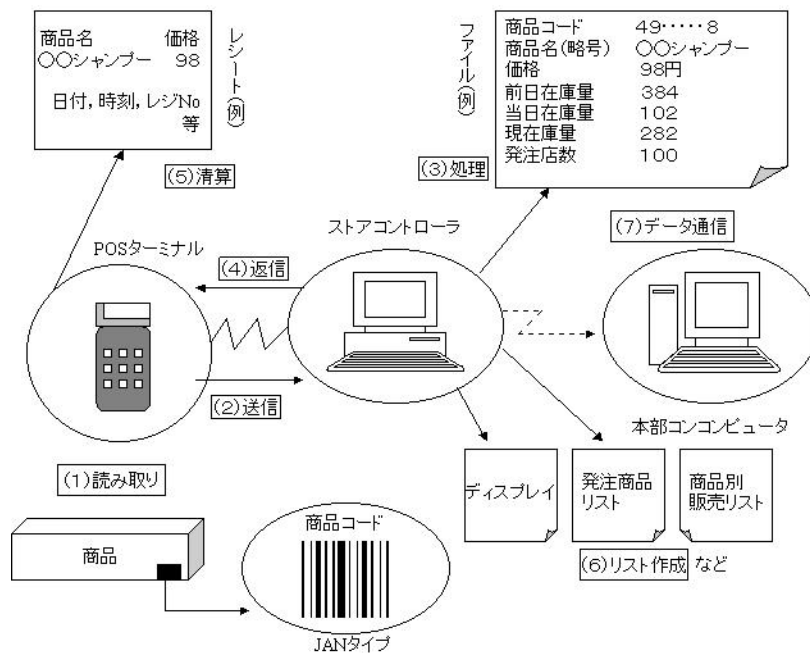


図 2.2 POS システム

る。店にはストア・コントローラというコンピュータが 1 台あって、商品の価格や商品名の一覧表（商品マスター）を持っている。POS 端末からバーコードが数字化されて送られてくると、ストア・コントローラはマスターを参照して商品名と価格を答え、POS 端末はレシートを印刷する。ストア・コントローラには、何がいつどれだけ売れたか、という情報が蓄積される。この情報はチェーン店の本部のコンピュータに集められて、売れ筋商品・死に

## 2.2 流通の情報化

筋商品が瞬時に把握される。死に筋商品（売れなくなった商品）はできるだけ早く新製品と交替させられるので、棚の効率的な利用が出来る。それ以上の販売データ活用は現在ではまだ難しいが、例えば次のようなものがある。弁当の売り切れる時刻が分かるので、あと何個用意すればいいか予測する手がかりが得られる。同時に買ってゆかれる商品の組み合わせが分かる。同時に売れる商品同士を近くに並べれば、消費者がその商品を買う用事を「思い出して」くれる確率が高くなるだろう。クレジットカードの信用情報のやりとり（そのカードが使用停止になっていないかのチェック）などにも使われる。バーコードにはJANコードと呼ばれる統一規格があって、国・メーカー・メーカーが決めた商品番号が書き込んである。しかし新製品を商品マスターに登録することは大変な労力なので、中小の小売店ではなかなか使いこなせないと言われる。生鮮食料品は、卸が包装するときバーコードのついたシールを貼っていることが多い。メーカーがバーコードを自分で付けることをソースマーキング、生鮮食料品のように流通業者が付けることをベンダーマーキングという。メーカーがJANコードに従ったソースマーキングを行うようになってから、バーコードによるPOSが急速に普及した。

### 2.2.3 VAN

VANは加入者の持つコンピュータの間で、専用線や電話回線を使って情報を交換することを言う。NTTの民営化と歩調を合わせて、民間のVAN運営会社がVANの管理サービスを行えるようになった。メーカーと卸、卸と小売店、あるいはそのすべての間の流通VANが注目を浴びている。POSやEOSを共通のものにしてみんなで使う、ということである。端末機やシステムソフトウェアへの二重投資を避けることができ、データの広範囲に渡る集計や比較が出来るようになる。これらのデータはのちに述べるリテールサポートのための情報としても貴重である。ひとつのメーカー、ひとつの卸が取引先との間で作るものもあるが、複数のメーカーが共同で、あるいは卸の組合が共同で作る「業界VAN」もあらわれている。

## 2.2 流通の情報化

### 2.2.4 EOS

EOSとは自動補充発注システムの略称で電話回線や専用回線を使って、コンピュータ同士のやりとりで製品の補充を注文することを指す。図2-3はVANを用いたEOSの一例である。チェーン店の本部に、各店のコンピュータから商品の発注データを集め、本部から

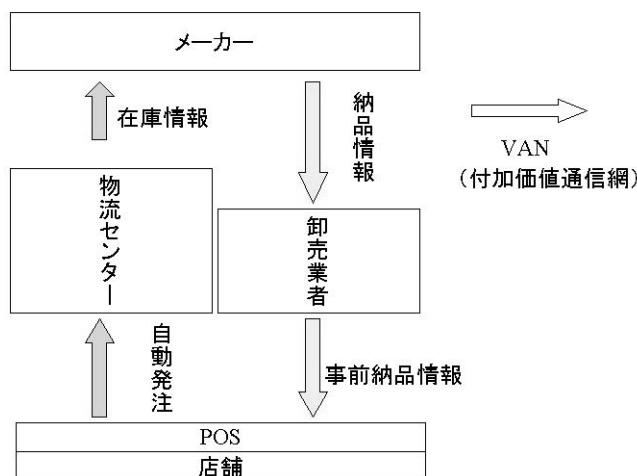


図 2.3 EOS システム

一括して注文することも多い。これによって人件費の削減と発注の正確化が期待される。現在のEOSはPOSと連動していて、POSのシステムの一部になっていることが多い。どの商品がいくら売れたか、というPOSデータを見ながら、そのシステムの中から発注ができる。しかし、EOSは異なる企業間のシステムでは、コンピュータでのデータのやり取りに関するルール、例えば商品名や発注するチェーン名・店舗などのコード類の手続きなどが社会的に統一されていないと、効率化を目指したはずのEOSが返って非効率になってしまう。こうした背景の元で1980年代末から、それまでの特定企業間の個別・専用的なシステムに代わって、だれもがアクセスできる統一的・社会的なオンライン・ネットワーク・システムが形成されつつある。すなわち、EDIが急速に普及し始めている。ここで、EDIとは、Electronic Data Interchangeの略であり、商取引を行おうとする企業同士が、標準化されたルールによってオンラインでコンピュータ相互のデータ交換を行うことである。

## 第3章

# 背景（共同研究プロジェクト）

本研究のバックグラウンドとして、現在高知工科大学と高知県内流通業者(株)高知流通情報サービス(KDIS)及び、同県内小売業者(株)サニーマートの3社で「ニューラルネットワーク等による商品発注支援システム」という共同研究プロジェクトが進行している。

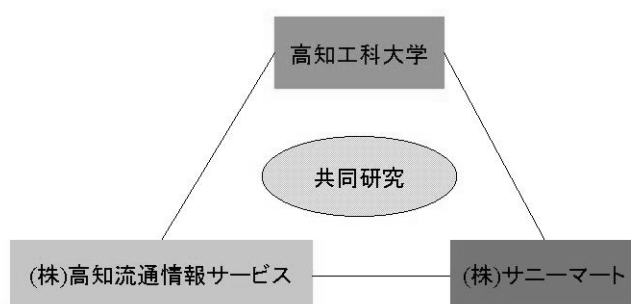


図 3.1 共同研究パートナー

高知流通情報サービスは、特別第2種電気通信事業者の認可を受け、地域密着型のVAN(付加価値通信網)会社として、地元企業の情報化の推進に大きな働きを努めてきている。そのネットワークは県内に広がっており、KDISのVANを経由して受発注情報を交換している企業は300社を超えている。また、近年のインターネットの急速な発達に対応し、インターネットを経由したデータ交換システム(Web-EDI)の開発も進めている。流通企業は他業界に比べてもIT化に早くから取り組んでおり、電子受発注システム(EOS)、

販売時点売上情報 (POS) などを積極的に進めてきている。特に店舗に関してはIT投資が積極的に行われており、店舗内LANはもちろん、店舗—本部間ネットワーク、イントラネット、ポータブル端末といったシステムがすでに装備されているのが現状である。

上記の共同研究プロジェクトは、こうした取り組みをベースに、店舗での最も重要な作業である発注作業の付加価値をより高めるためのシステムを構築するものである。生活者の多様化した購買ニーズに対応できる、フレンドリーな店舗運営を可能とする事により、企業戦略における重要なKFS (Key for Success : 成功要因) と位置付けられるものである。

流通業者にとって商品の発注を適正に行うことは、企業の競争力を左右する最大の課題である。従来は人間の経験則に基づいた数量の決定を行っていたため、在庫過小による販売機会の逸失や、在庫過大による経営の圧迫といった過大を抱えている。需要に応じた発注数量の最適化は、流通業におけるSCM (Supply Chain Management) のキーファクターであり、メーカ、卸売業者、小売業者をつなぐ最適化ネットワークの実現に大きく寄与するものである。「ニューラルネットワーク等による商品発注支援システム」は、流通業者における販売商品の適正発注数量を、ニューラルネットワークの機能により算出し、これをもって最大限の販売数量を最小限の在庫数量で実現しようとするものである。さらに、ニューラルネットワークの特徴である、自己学習機能を採用することにより、地域や商品、業界の特性に合わせた適正発注量ロジックを自己生成し、多様な市場のニーズに合わせた判断を、プログラミングレスで実現するものである。

過去の販売データ、発注データ、事象データを元に、仮想ルールによる発注ロジックを生成する。さらに、過去のデータを元にした事故学習 (仮想発注→実際の販売データによる政界情報による学習) を繰り返し、ロジックの精度を向上させる。また、現場担当者に実際に操作を行ってもらい、そのノウハウを自己学習の正解情報として学習させることで、地域や商品の特性に合わせたロジックの進化を目指す。

本研究は上記のプロジェクトの初期研究という位置付けとなっている。対象データを販売データ1つに絞り、膨大な商品に対して商品間の因果関係を突き止めるのが目的である。一方の商品が売れば、他方の商品にどのような影響が起こるのか、相互依存があり、同時に

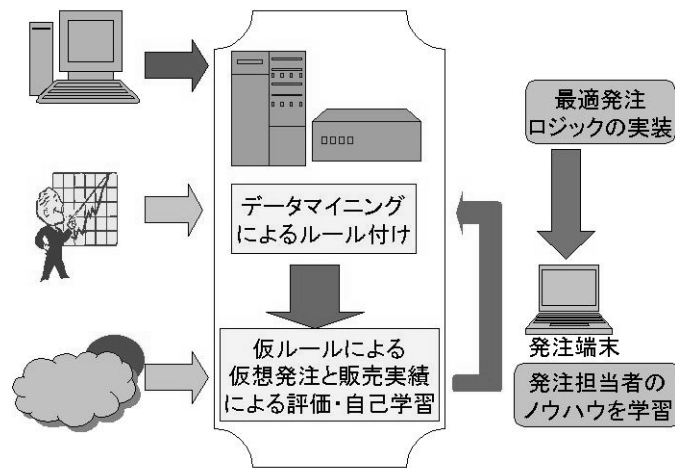


図 3.2 共同研究テーマ概要

売れるのが、それとも売れなくなるのか、といったものを時系列解析で検証して、将来のニューラルネットワークの自己学習のデータに役立たせるものである。

# 第 4 章

## 時系列

### 4.1 時系列解析

時間とともに不規則に変動する現象の記録が時系列である。つまり観測値  $x_t$  の集合であり、各観測値はある特定の時間  $t$  において記録されている。特に 2 つ以上の現象を同時に記録したものを多変量時系列と呼ぶ [5][6]。離散時間時系列とは、一定の時間間隔で観測する場合のように、観測時間の集合  $t_0$  が離散集合である時系列である。連続時間時系列とは、たとえば  $t_0 = [0, 1]$  のように、ある期間において観測値が連続的に記録される時系列である。時系列解析では、データに適する確立モデルを選ぶことが重要である。将来の観測値を完全には予測できないという性質を考慮すると、各観測値  $x_t$  はある確率変数  $X_t$  の実現値であると仮定することは自然である。確率変数の列  $X_1, X_2, \dots$ , についての完全な確率論的な時系列モデルは、確立ベクトル  $(X_1, X_2, \dots, X_n)'$ ,  $n = 1, 2, \dots$  のすべての同時分布、あるいは同じことであるがすべての確立

$$P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n], \quad -\infty < x_1, x_2, \dots, x_n < \infty, \quad n = 1, 2, \dots \quad (4.1)$$

を示したものである。このようにすべての同時分布を示すことは時系列解析においてまれである。なぜならば、一般に利用可能なデータから推定すべきパラメータが多すぎるからである。そのかわりに、同時分布の 1 次モーメントと 2 次モーメント、すなわち期待値  $EX_t$  と積の期待値  $E(X_{t+h}X_t)$ ,  $t = 1, 2, \dots, h = 0, 1, 2$  のみを示すことにする。そして、これらのモーメントのみに依存するような列  $X_t$  の性質に注目する。 $X_t$  のそのような性質を 2 次の性質と呼ぶ。すべての同時分布が多変量正規分布である特別な場合には、 $X_t$  の 2 次の性質は同時分布を完全に決定し、したがって列の確率論的な性質を完全に与える。実際に現

## 4.2 標本相関関数

れる時系列の多くは、ベクトル値をとる（多変量）時系列  $X_t$  の成分であり、その各成分系列  $X_{ti}$  内に依存性があるのみならず、異なる成分系列  $X_{ti}$  と  $X_{tj}$ 、 $i \neq j$  の間にも相互依存性があると考えるのがもっともよい。

## 4.2 標本相関関数

### 4.2.1 標本自己相関関数

時系列  $X_t, t = 0, \pm 1, \pm 2, \dots$  の統計適性質がすべての整数  $h$  について「時間をずらした」系列  $X_{t+h}, t = 0, \pm 1, \pm 2, \dots$  と同じならば、時系列は定常であるという。データにおける依存の程度を評価し、依存性を反映するモデルを選ぶための重要な道具の一つが標本自己相関関数（標準 ACF）である。標本自己相関関数を次のように定義する。

$x_1, x_2, \dots, x_n$  をある時系列の観測値とする。

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad (4.2)$$

を  $x_1, x_2, \dots, x_n$  の標本平均という。

$$\hat{\gamma} = n^{-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n \quad (4.3)$$

を標本自己共分散関数という。

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n \quad (4.4)$$

を標本自己相関関数という。

つまり標本自己相関関数は標本共分散関数を  $h = 0$  の標本共分散関数で割った、つまり正規化したものである。

### 4.2.2 標本相互相関関数

標本自己相関関数は対象となる時系列が1つであったが、標本相互相関関数（標本 CCF）は2つの、つまり多変量の時系列を扱うことが可能である。下記のように次に標本相互相関



### 4.3 IID ノイズ

関数を定義する。  $x_1, x_2, \dots, x_n$  及び  $y_1, y_2, \dots, y_n$  をある 2 つの時系列の観測値とする。

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad (4.5)$$

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad (4.6)$$

を  $x_1, x_2, \dots, x_n$  及び  $y_1, y_2, \dots, y_n$  の標本平均という。

$$\hat{\gamma} = n^{-1} \sum_{t=1}^{n-h} [x_{t+h} - \bar{x}](y_t - \bar{y}), \quad h \geq 0 \quad (4.7)$$

$$\hat{\gamma} = n^{-1} \sum_{t=-h+1}^n [(x_{t+h} - \bar{x})(y_t - \bar{y})], \quad h \leq 0 \quad (4.8)$$

を標本自己共分散関数という。

$$\hat{\rho}(h) = \hat{\gamma}(h) [\hat{\gamma}_x(0) \hat{\gamma}_y(0)]^{-1/2}, \quad |h| > 0 \quad (4.9)$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\sqrt{\hat{\gamma}_x(0) \hat{\gamma}_y(0)}} \quad |h| > 0 \quad (4.10)$$

を標本相互相関関数と言う。標本相互相関関数は標本自己相関関数の定義式にあらたに時系列を追加したものと考えてよい。ただし、標本相互相関関数の式では分母が  $x$  との双方のずれのない ( $h = 0$ ) 自分自身の標本共分散関数の積であることが重要である。

### 4.3 IID ノイズ

(式 4.1) の考えを導入して POS データの観測値  $x_1, x_2, \dots, x_n$  は確率変数  $X_1, X_2, \dots, X_n$  の実現値とする [10][11]。これの  $X_1, X_2, \dots, X_n, n$  個の確率変数が互いに独立、つまり、それぞれは無相関であることを任意の  $x_1, x_2, \dots, x_n$  に対して、

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) * \dots * P(X_n \leq x_n) \quad (4.11)$$

が成り立つことを定義する。これを、 $X_1, X_2, \dots, X_n$  は独立同一分布 (IID:Independent Identically Distributed) に従うと言う。

また、これら  $X_1, X_2, \dots, X_n$  を IID ノイズと言う。

### 4.3 IID ノイズ

特に、任意の  $h \geq 1$  と  $x, x_1, \dots, x_n$  に対して、

$$P[X_{n+h} \leq x | X_1 = x_1, \dots, X_n = x_n] = P[X_{n+h} \leq x] \quad (4.12)$$

であるから、 $X_1, X_2, \dots, X_n$  に関する知識は  $X_{n+h}$  の挙動を予測するのに役立たない。実際、 $X_1, X_2, \dots, X_n$  の値が与えられた時、平均二乗誤差

$$E[(X_{n+1} - f(X_1, \dots, X_n))^2] \quad (4.13)$$

を最小化する関数は恒久的に 0 である。この性質から IID ノイズは、予測の面では興味の無い確率過程である [12]。分散が有限の IID ノイズでは、 $h > 0$  の標本自己相関関数及び標本相互相関関数  $\rho(\hat{h})$  は、 $n$  が大きいと近似的に IID の  $N(0, \frac{1}{n})$  に従う。したがって、両相関係数の約 95 % は、 $\pm \frac{1.96}{\sqrt{n}}$  を信頼限界とする区間にはいるはずである。(ただし、1.96 は標準正規分布の 0.975 確定点である。) つまり、理想的には相関解析によって相関が表れるためにはデータの割合が信頼区間から 5 % 以上出ていることが重要となってくる。

[13]

# 第 5 章

## 実験

### 5.1 POS データからの Perl による目的データ抽出

#### 5.1.1 POS データの形式

POS(販売時点情報管理) データにはその地域, 企業, 商品特性によって構成形式がことなっている。本実験で解析に用いた POS データはサニーマートから頂いたものである。この POS データの構成は以下の図の通りである。

店コード	売上日	曜日コード	分類コード	JANコード	商品名	売上数量	売上金額	商品単価
001	990101	7	0432	123456	AAAA	12	2400	200
003	990213	6	0442	458123	BBBB	7	3500	500
026	990221	1	0201	643226	CCCC	3	750	250
001	990412	7	0322	256223	DDDD	5	4000	800
011	990422	2	0221	452456	EEEE	23	2300	100
026	990508	3	0432	122342	AAAA	11	2200	200
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

図 5.1 POS データ形式

店コード：店舗コード、入力桁数は 3 桁である。売上日：商品が実際に売れた日、入力桁数は 6 桁である。年号下 2 桁・月 2 桁・日 2 桁である。曜日コード：商品が売れた曜日、入力桁数は 1 桁である。日曜日を 1 とし、月曜日を 2、以下続く。分類コード：商品カテゴ

## 5.1 POS データからの Perl による目的データ抽出

リーコード、入力桁数は4桁である。JANコード：商品コードともいう。商品識別のコード、入力桁数は13桁である。商品名：半角カタカナで記述されている、入力桁数は25桁である。売上数量：実際に売れた数量、入力桁数は8桁である。売上金額：実際の売上金額、入力桁数は8桁である。商品単価：商品の単価、入力桁数は7桁である。

計9つのデータがカンマで区切られた、いわゆるCSV形式で構成されている。

### 5.1.2 抽出過程

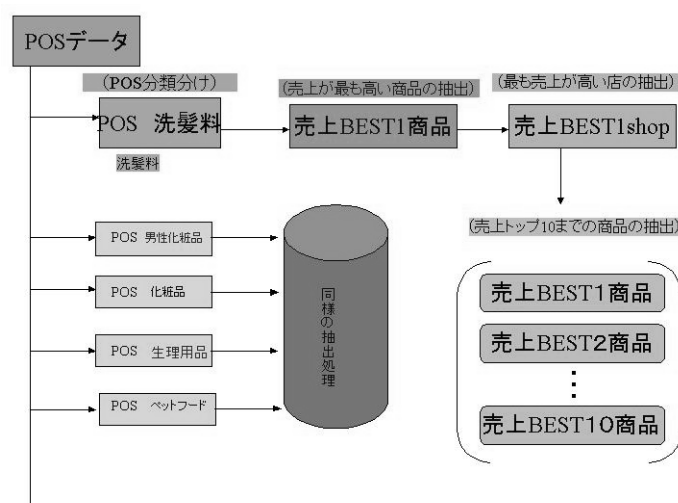


図 5.2 POS データ抽出過程

POS データをまずは分類コード毎に分解する、この作業は (株) サニーマートによって分解されていたので、実際の作業は分類分けされたそれぞれの POS データからの目的のデータを抽出することから始める。今回解析した POS データは洗髪料、男性化粧品、化粧品、生理用品、ペットフードの5つである。

作業手順は1にそれぞれの POS データの中で最も売上の高い商品のデータを取り出す。そしてその商品を最も多く売った店についてのデータを抽出する。その店の中で売上の高い商品を1位から10位までのデータを取り出す。本研究ではこの5つの分類それぞれ10商品のデータについて解析を行うことにする。

上の図のように BEST 1 商品を基準として BEST2 から BEST 10 までの標本 CCF を

## 5.1 POS データからの Perl による目的データ抽出

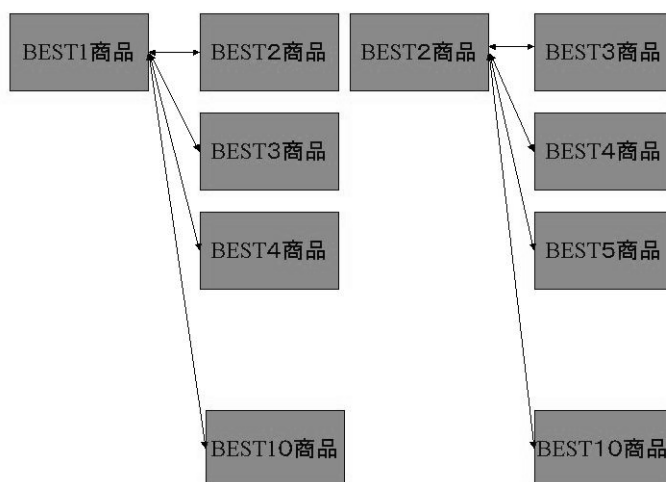


図 5.3 解析商品組み合わせ

求める。また、基準を逆にしての標本 CCF を行う。そして BEST2 を基準として BEST3 から BEST10 までの処理も同様に行う。1 つも分類で 34 通りの組み合わせがある。それを 5 つの分類についても解析するので計 170 通りの組み合わせがあることになる。本当は POS に収録されている商品全ての相関解析を行うのが最もよいのであるが、1 つの分類中での商品数は優に数百を超えており、全ての組み合わせは数十万通りとなる。また、サニーマートからデータを頂いた時期が遅く、全ての解析を完了することは不可能であったため、今回は売上の高い商品に限って解析をおこなった。

実際の POS データ抽出はプログラミング言語 Perl[7][8][9] でアルゴリズムを作成して処理を行った。EXCEL などの既存の計算ソフトでは、処理に対応しきれない。なぜならば、POS データは大容量の CSV ファイルであり、計算ソフトの限界を超えている。また、Perl は CSV 形式のファイル処理に強く、抽出処理が容易な点があり、今回は Perl を選び、プログラムを作成した。

## 5.2 MATLABによる標本相互相互相関解析

標本相互相関関数を行列演算ソフト MATLAB[4] でアルゴリズムを作成した。プログラムの内容は、以下の通り。始めに、2つの時系列における売上データを重ねてグラフ化する。

次に標本相互相関を求める。

1. 解析する2つの売上データそれぞれの平均を求める。
2. 標本共分散関数を求める。それぞれの時系列の平均との差分をとり、時間派パラメータをずらして平均を取る。
3. 標本相互相関関数を求める。標本共分散を2つの時系列  $h = 0$  で割り、正規化を行う。
4. 最も高い相関係数とそれが出た時のラグをもとめ、 $h$  ずらした時の2つの商品の散布図を作成する。
5. 信頼限界を求め、その信頼区間を標本相互相関関数グラフ中に加える。
6. 信頼区間からデータが出ている割合を求める。

同時に、標本相互相関関数の有効性を示すために標本自己相関関数も求める。また、BEST1の店だけではなく、あとランダムに選んだ、2つの店について、上記と同様の処理を行い、解析結果の正当性を検証する。

# 第 6 章

## 結果

### 6.1 掲載された結果グラフの説明

次に結果グラフ中の対象データ及びその略称を説明する。

#### 1. POS \_\_洗髪料

- BEST1 商品を B1, BEST2 商品を B 2 とする。
- 比較のために 3 つの店について解析しているが、それぞれの店を S1,S2,S3 とする。

#### 2. POS \_\_男性化粧品

- BEST1 商品を B1, BEST 2 商品を B 2, BEST 8 商品を B 8 とする。

#### 3. POS \_\_化粧品

- BEST1 商品を B1, BEST 2 商品を B 2, BEST 6 商品を B6 とする。

#### 4. POS \_\_生理用品

- BEST1 商品を B1, BEST 2 商品を B 2, BEST 8 商品を B8 とする。

#### 5. POS \_\_ペットフード

- BEST1 商品を B1, BEST 2 商品を B 2, BEST 8 商品を B8 とする。

## 6.2 分類ごとの POS データ解析結果

### 6.2 分類ごとの POS データ解析結果

#### 6.2.1 POS \_\_洗髪料における解析結果

- s1 についての解析結果

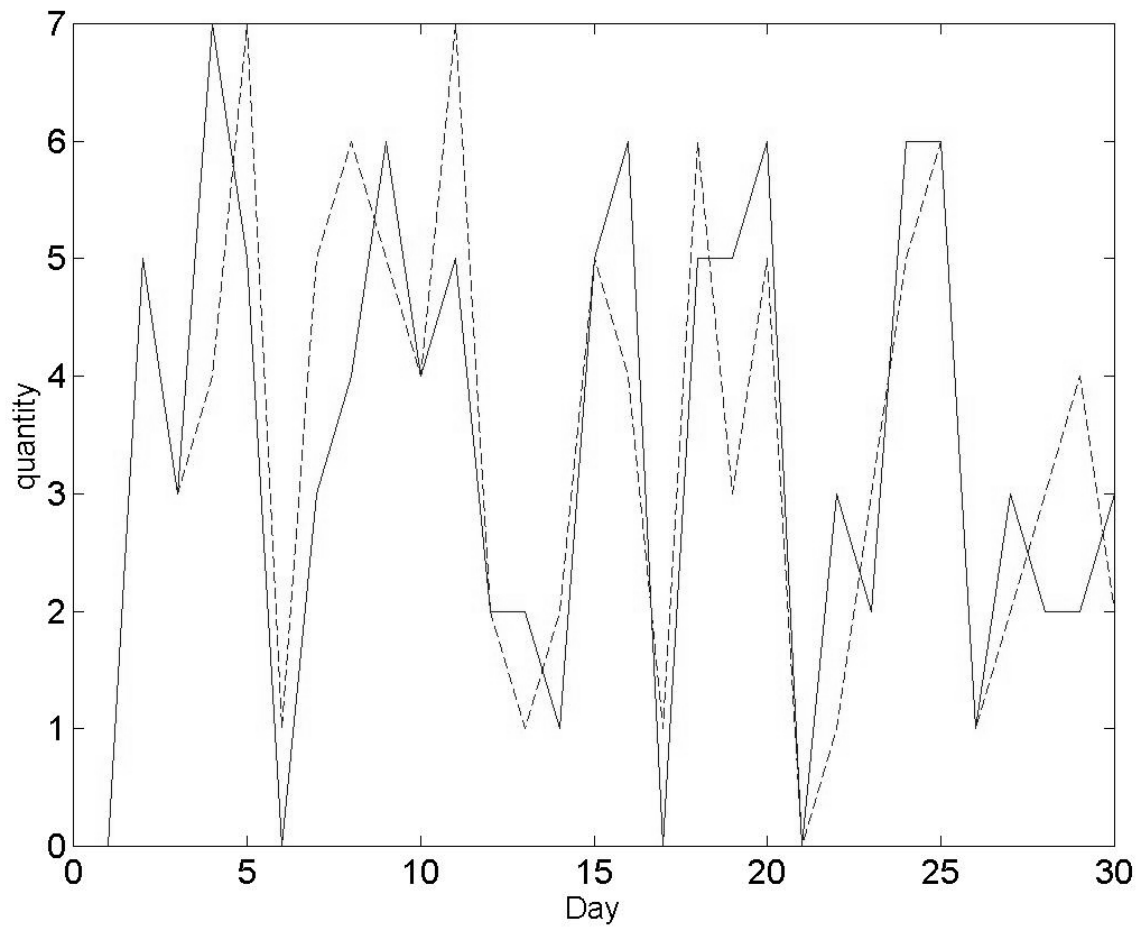


図 6.1 B1 と B2 の売上グラフ (実線が B1, 波線が B2)



## 6.2 分類ごとの POS データ解析結果

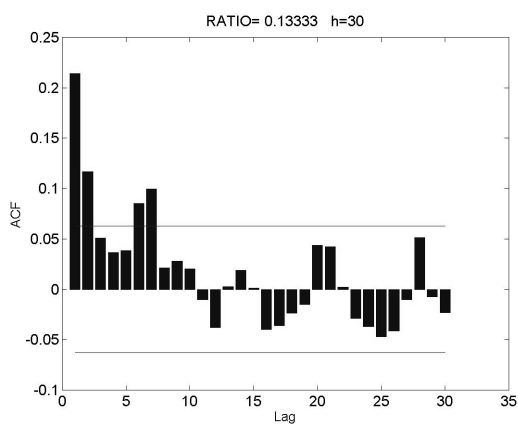


図 6.2 B1 の SACF

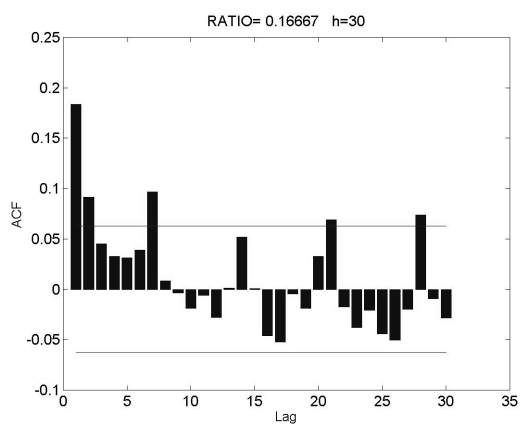


図 6.3 B2 の SACF

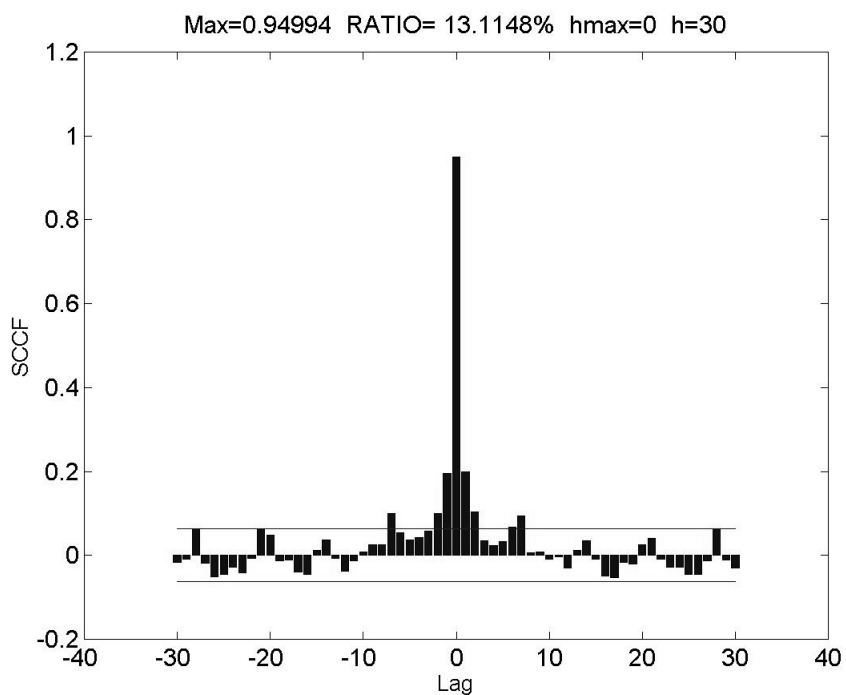


図 6.4 B1 と B2 間の SCCF

## 6.2 分類ごとの POS データ解析結果

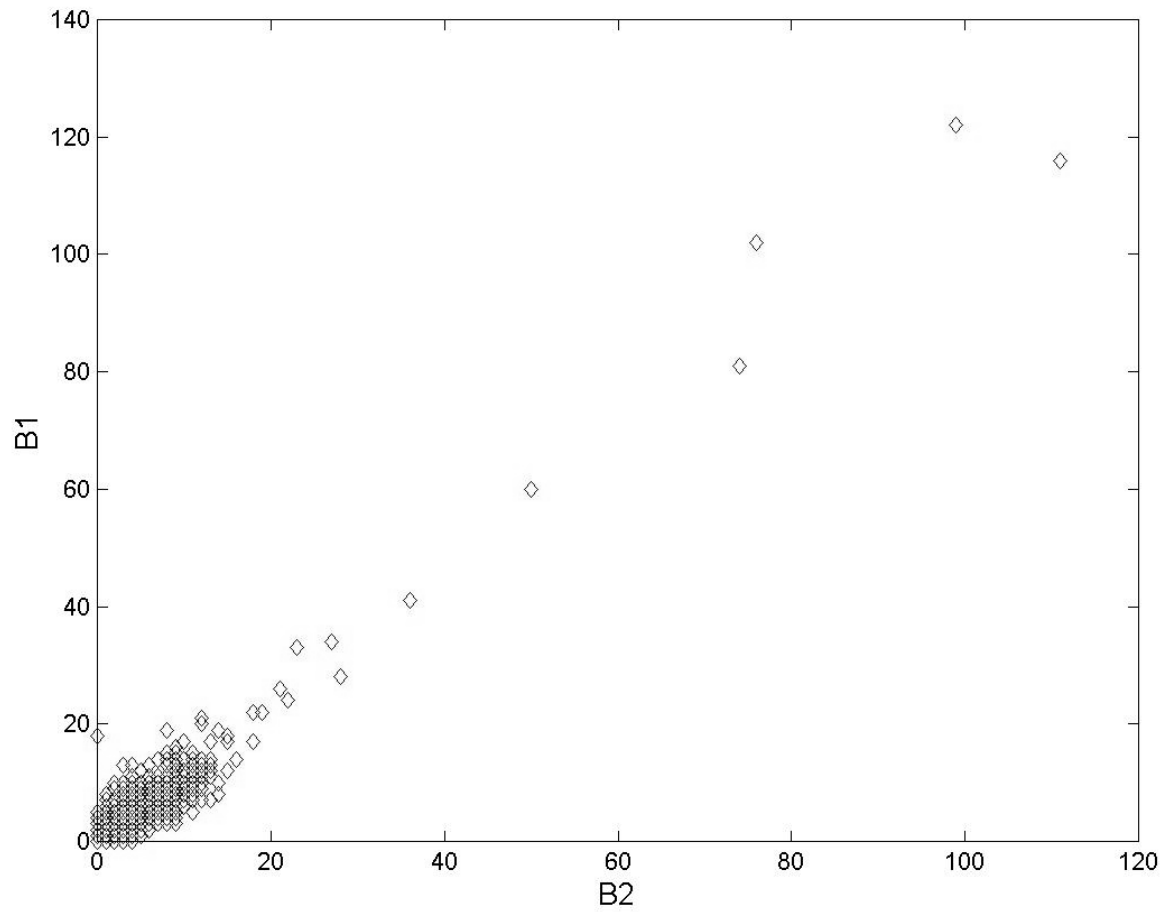


図 6.5 B1 と B2 の散布図

## 6.2 分類ごとの POS データ解析結果

- s2 についての解析結果

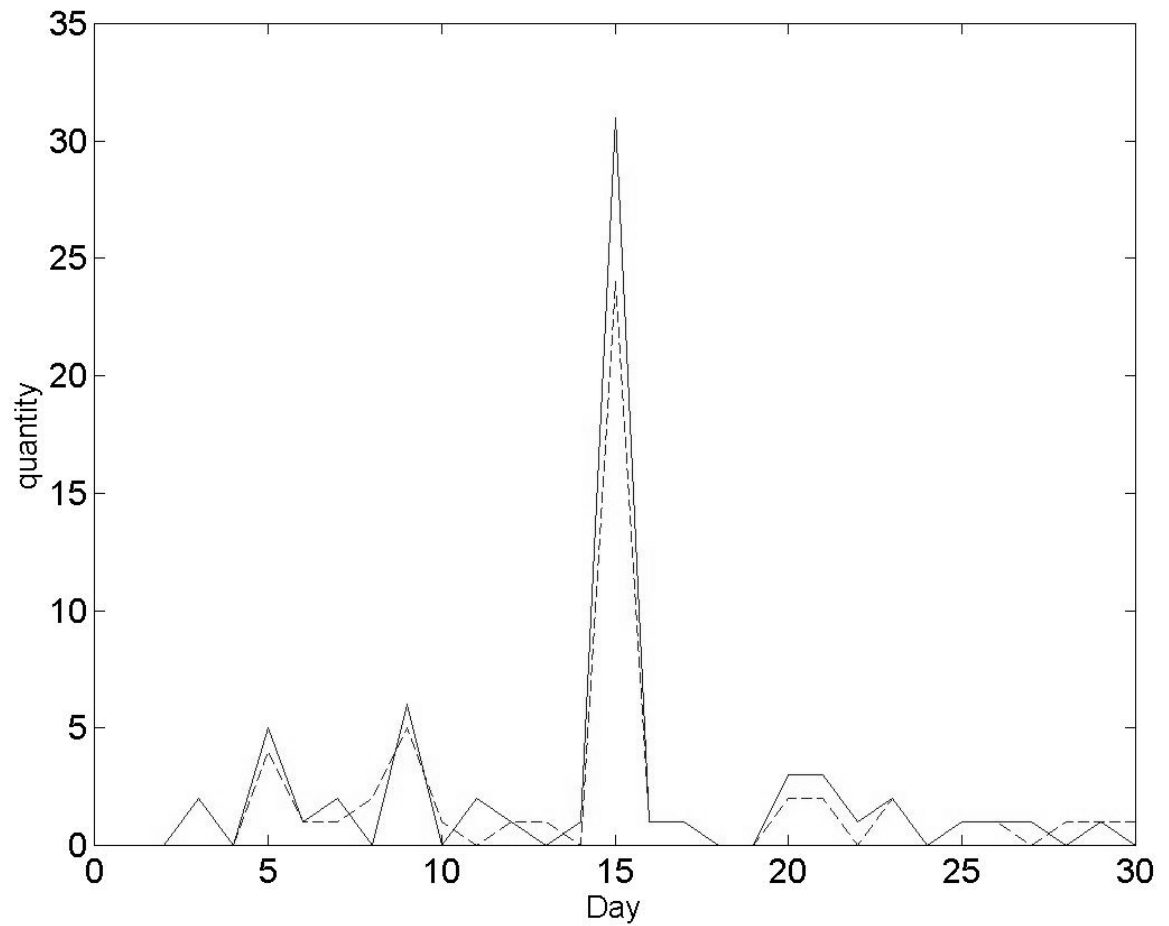


図 6.6 B1 と B2 の売上グラフ (実線が B1, 波線が B2)

## 6.2 分類ごとの POS データ解析結果

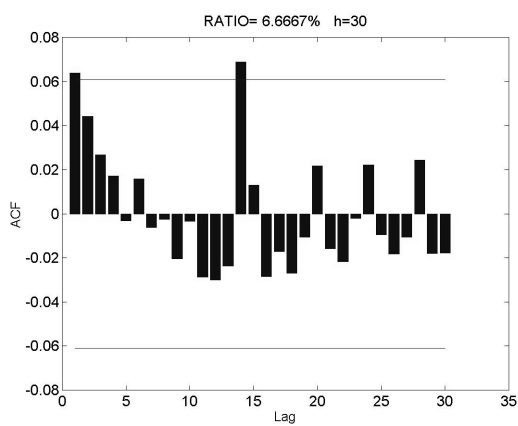


図 6.7 B1 の SACF

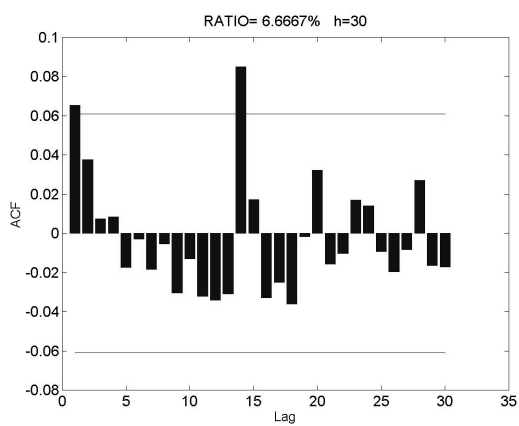


図 6.8 B2 の SACF

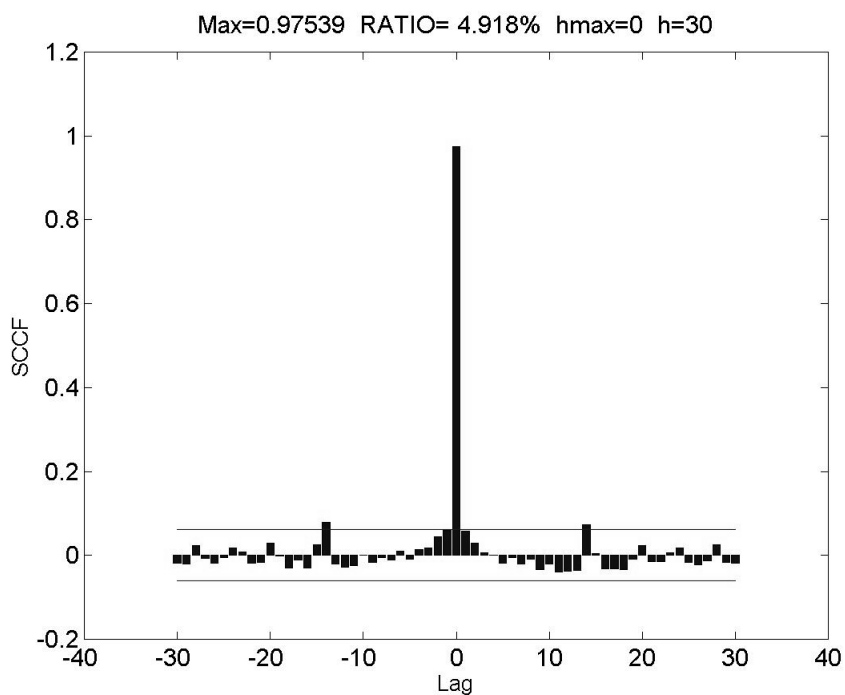


図 6.9 B1 と B2 間の SCCF

## 6.2 分類ごとの POS データ解析結果

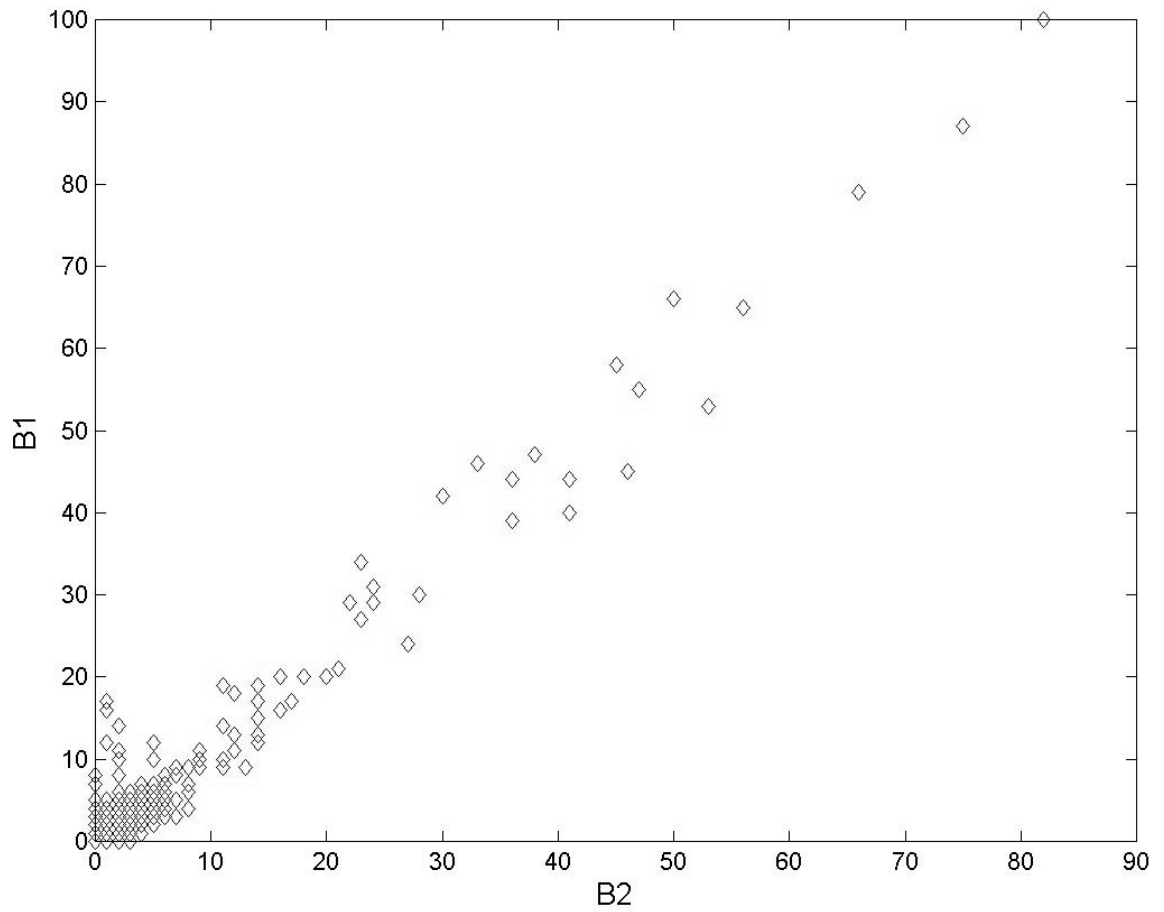


図 6.10 B1 と B2 の散布図

## 6.2 分類ごとの POS データ解析結果

- s3 についての解析結果

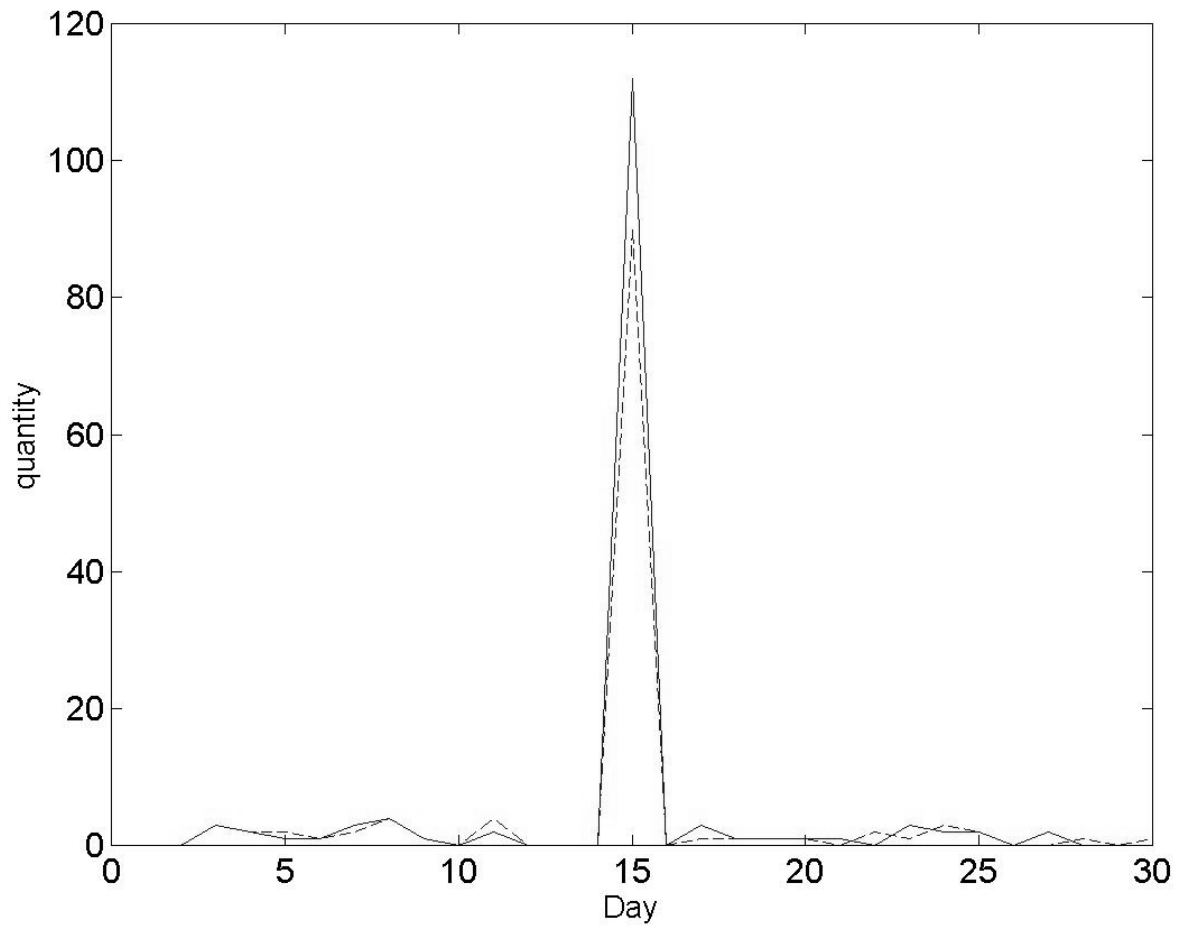


図 6.11 B1 と B2 の売上グラフ (実線が B1, 波線が B2)

## 6.2 分類ごとの POS データ解析結果

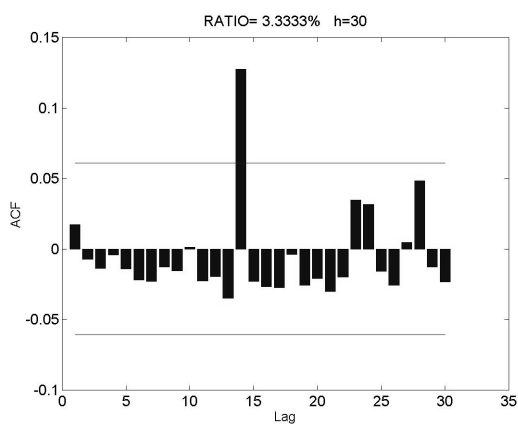


図 6.12 B1 の SACF

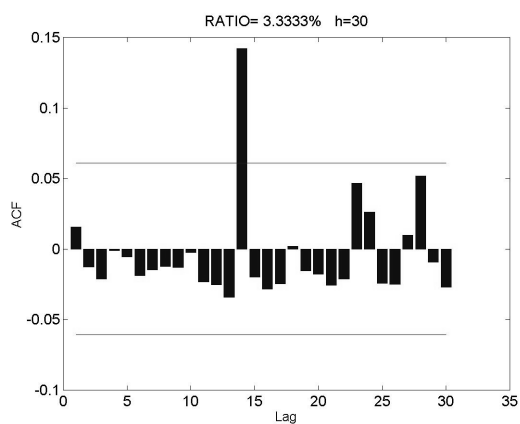


図 6.13 B2 の SACF

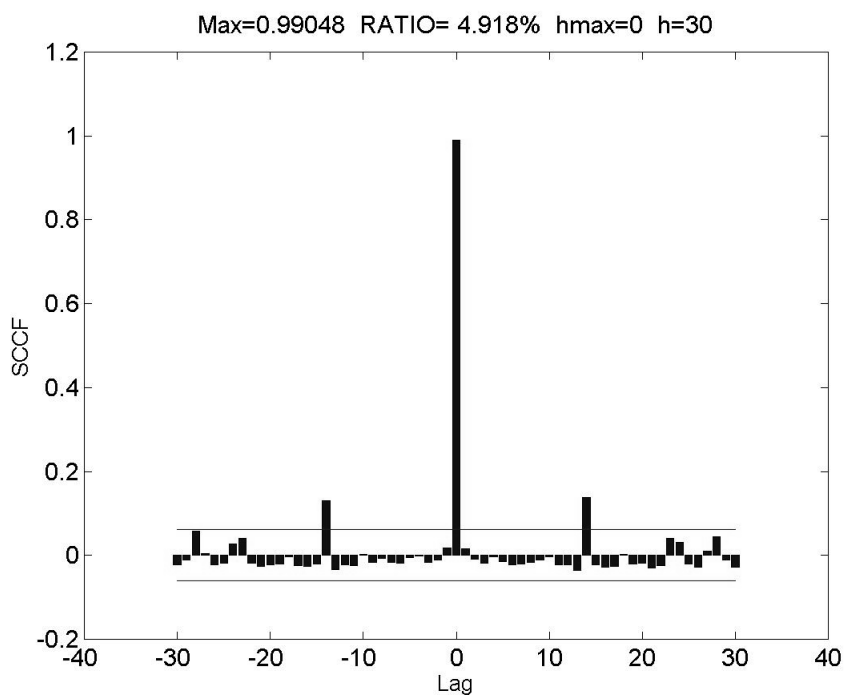


図 6.14 B1 と B2 間の SCCF

## 6.2 分類ごとの POS データ解析結果

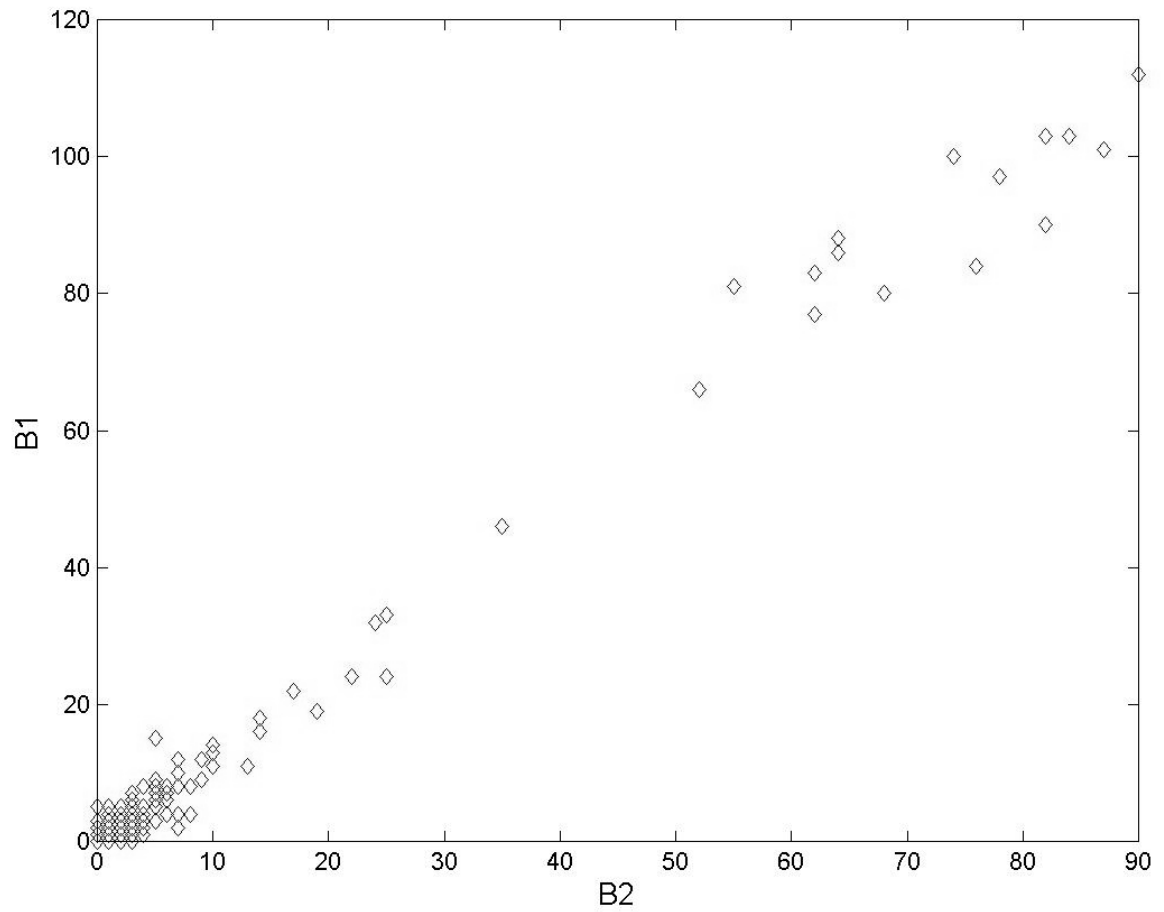


図 6.15 B1 と B2 の散布図



## 6.2 分類ごとの POS データ解析結果

### 6.2.2 POS \_\_男性化粧品における解析結果

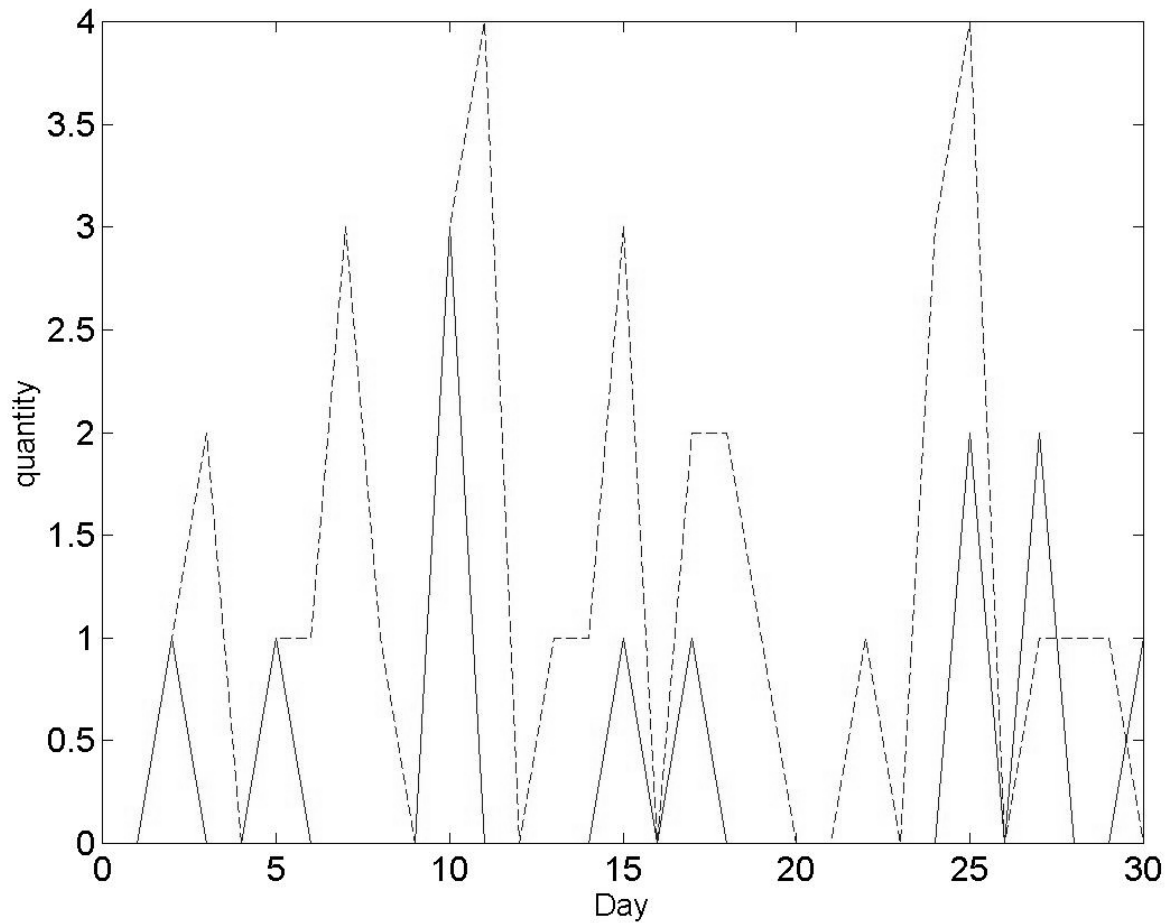


図 6.16 B1 と B2 の売上グラフ (実線が B1, 波線が B2)

## 6.2 分類ごとの POS データ解析結果

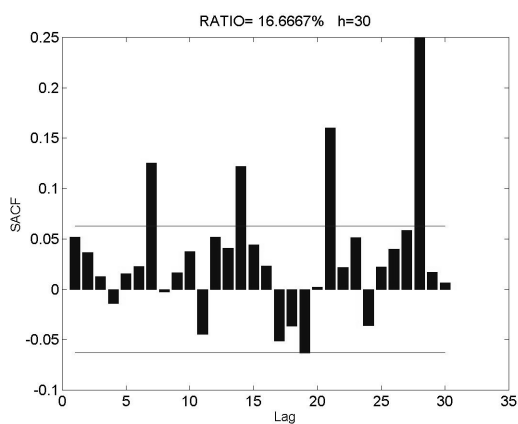


図 6.17 B1 の SACF

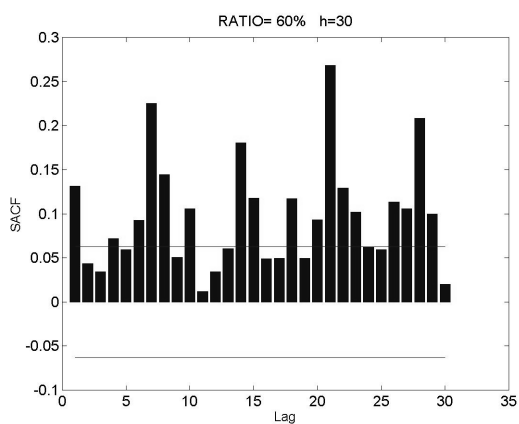


図 6.18 B2 の SACF

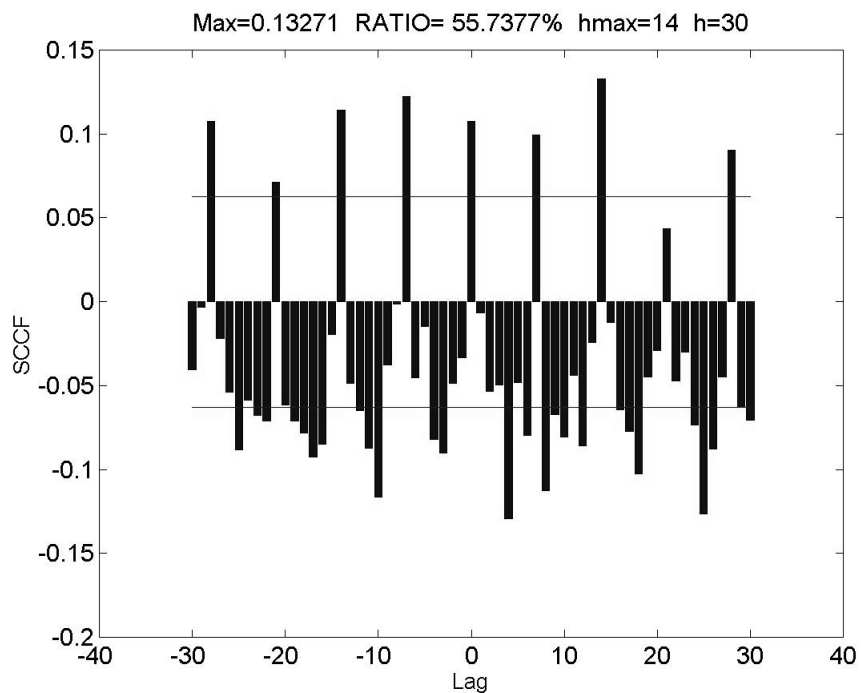


図 6.19 B1 と B2 間の SCCF

## 6.2 分類ごとの POS データ解析結果

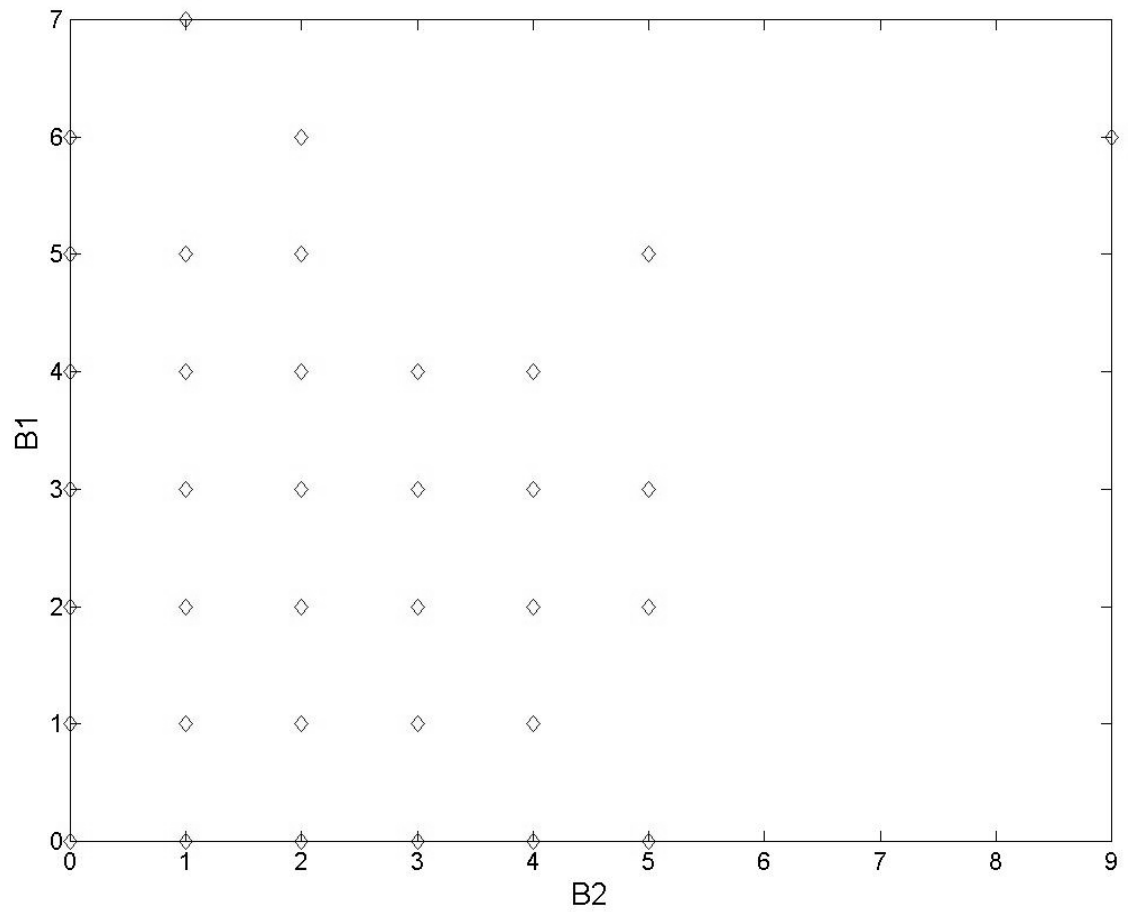


図 6.20 B1 と B2 の散布図

## 6.2 分類ごとの POS データ解析結果

### 6.2.3 POS \_\_化粧品における解析結果

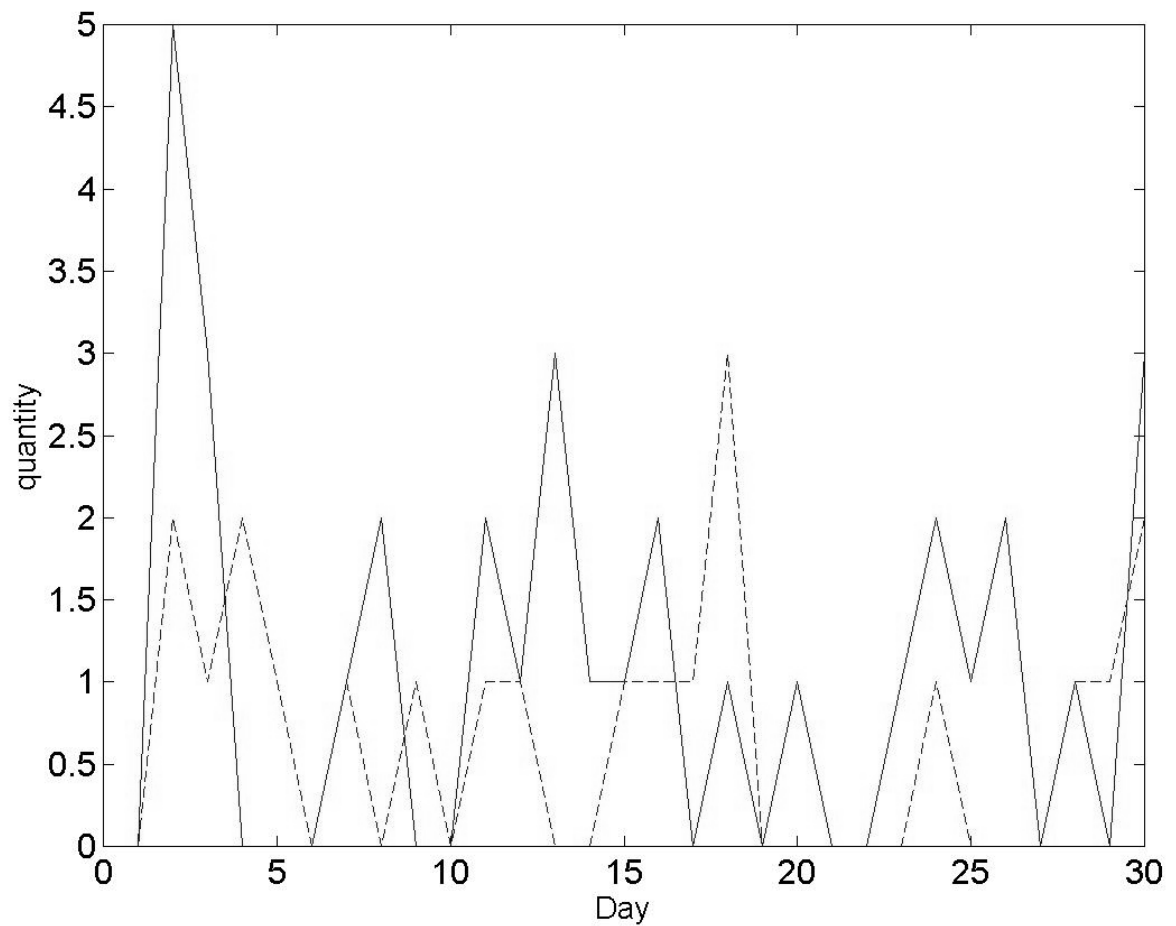


図 6.21 B1 と B2 の売上グラフ (実線が B1, 波線が B2)

## 6.2 分類ごとの POS データ解析結果

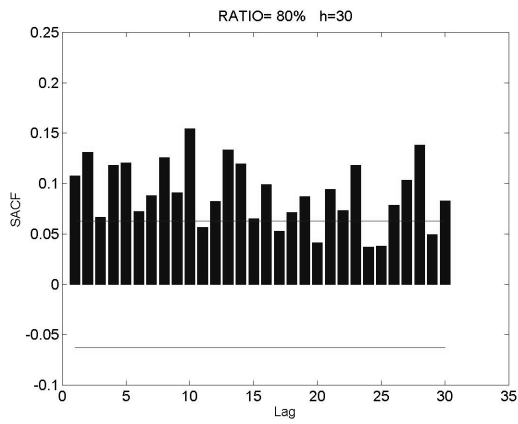


図 6.22 B1 の SACF

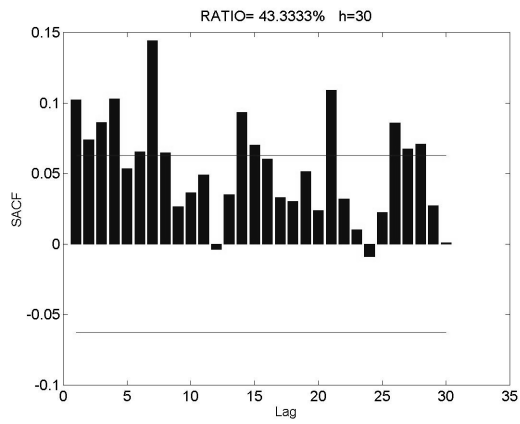


図 6.23 B2 の SACF

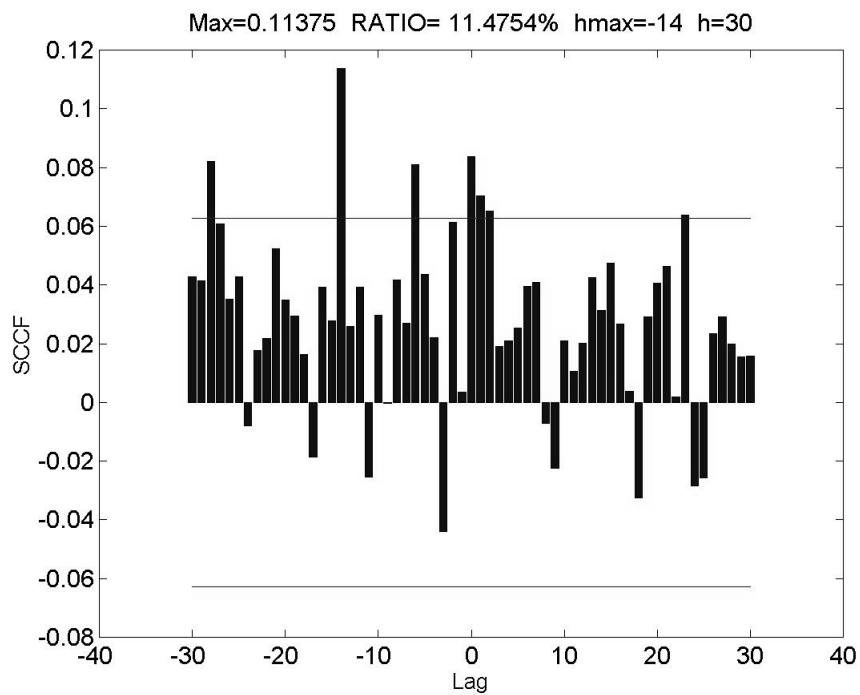


図 6.24 B1 と B2 間の SCCF

## 6.2 分類ごとの POS データ解析結果

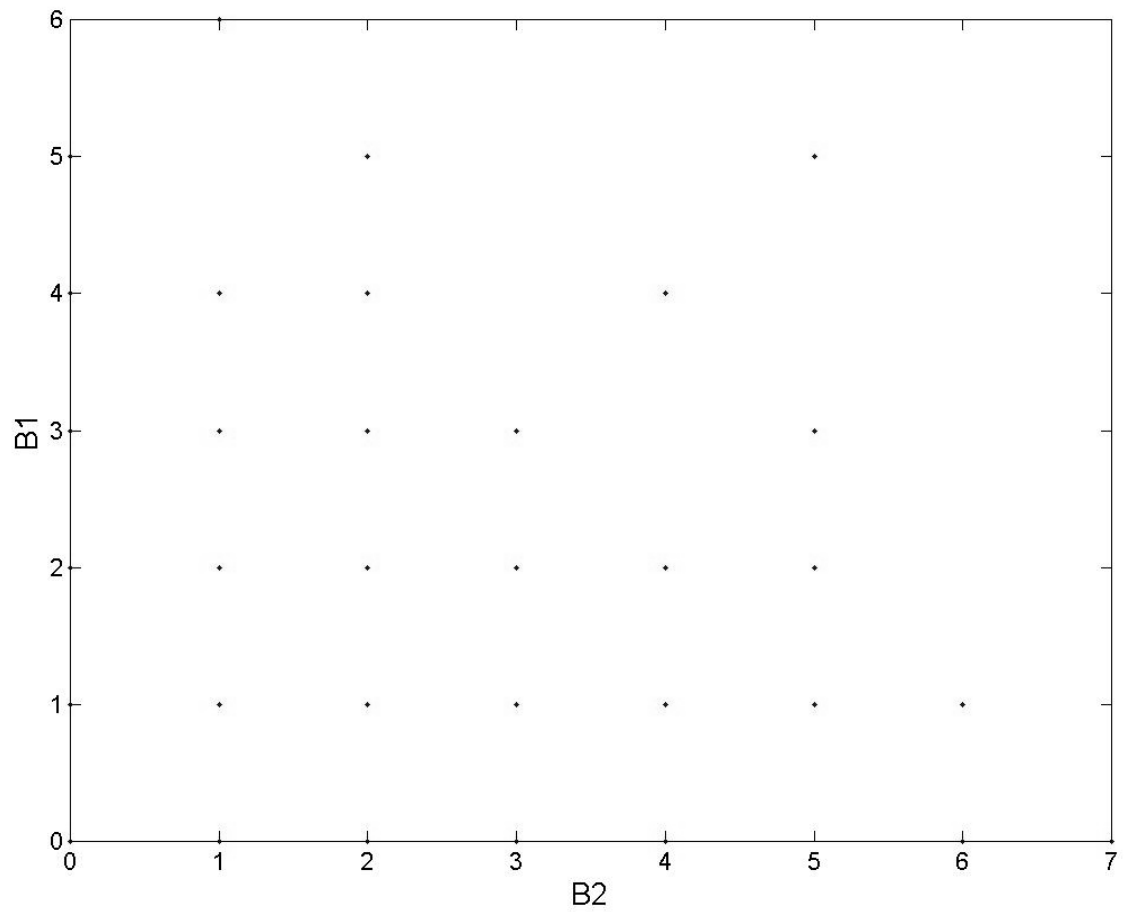


図 6.25 B1 と B2 の散布図

## 6.2 分類ごとの POS データ解析結果

### 6.2.4 POS \_\_生理用品における解析結果

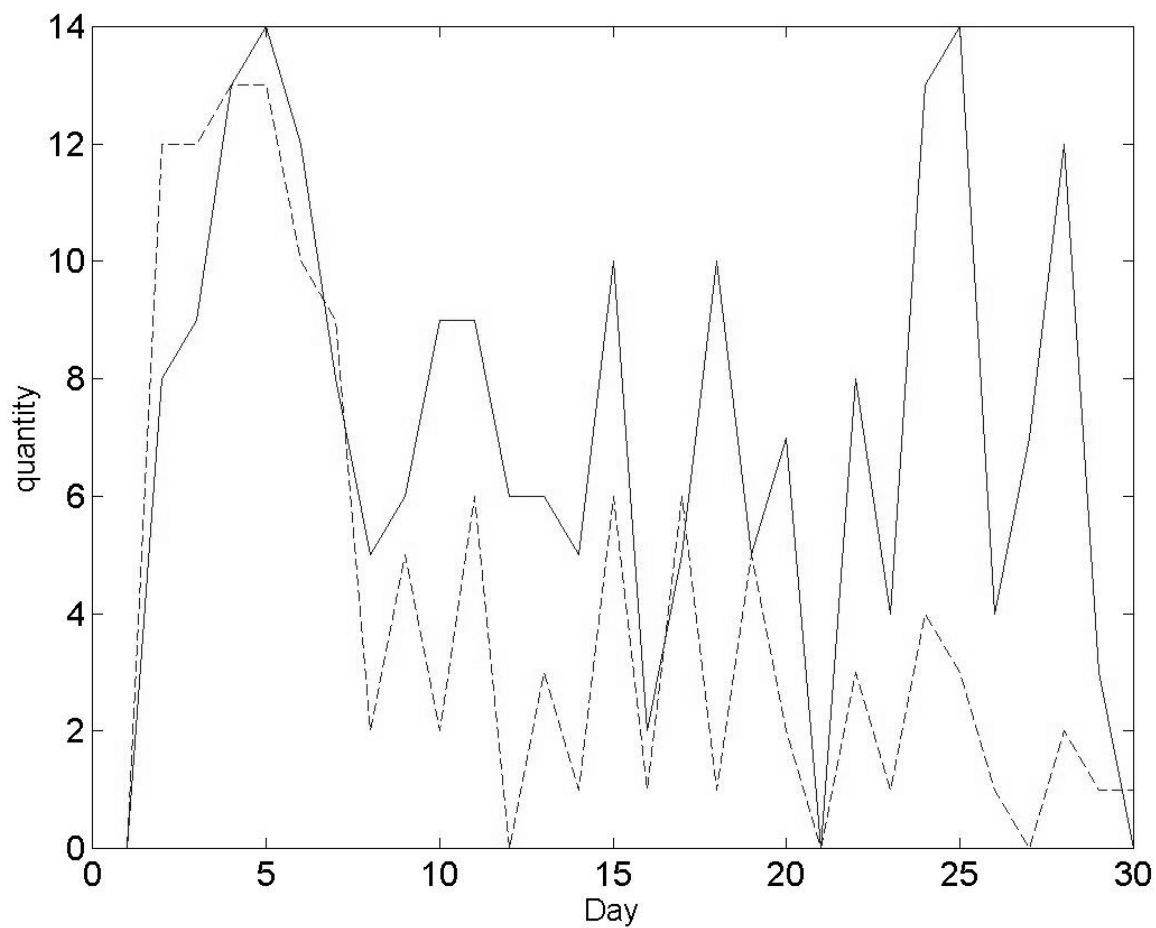


図 6.26 B1 と B2 の売上グラフ (実線が B1, 波線が B2)

## 6.2 分類ごとの POS データ解析結果

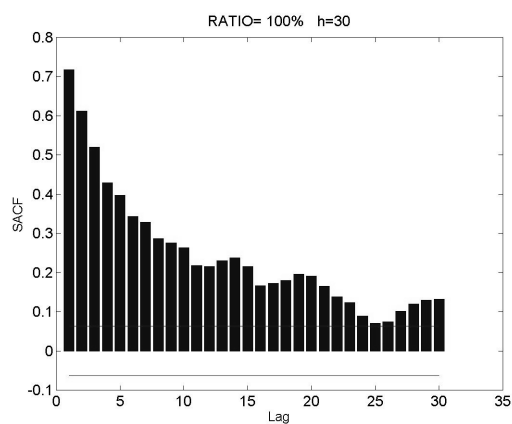


図 6.27 B1 の SACF

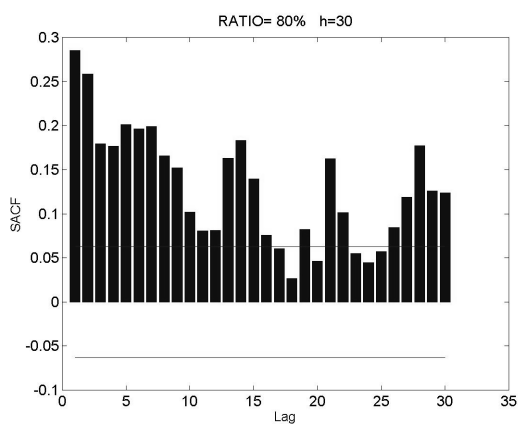


図 6.28 B2 の SACF

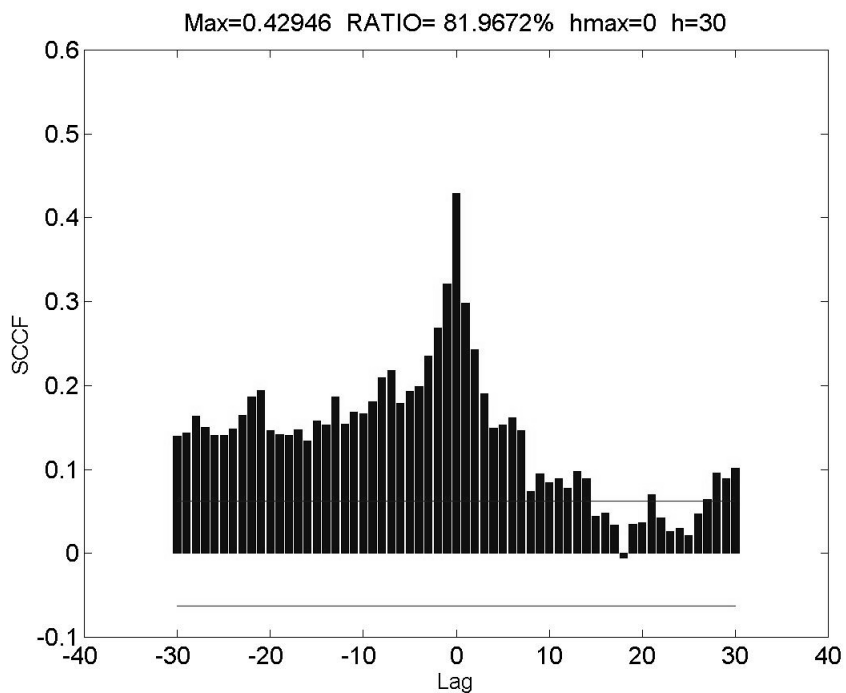


図 6.29 B1 と B2 間の SCCF



## 6.2 分類ごとの POS データ解析結果

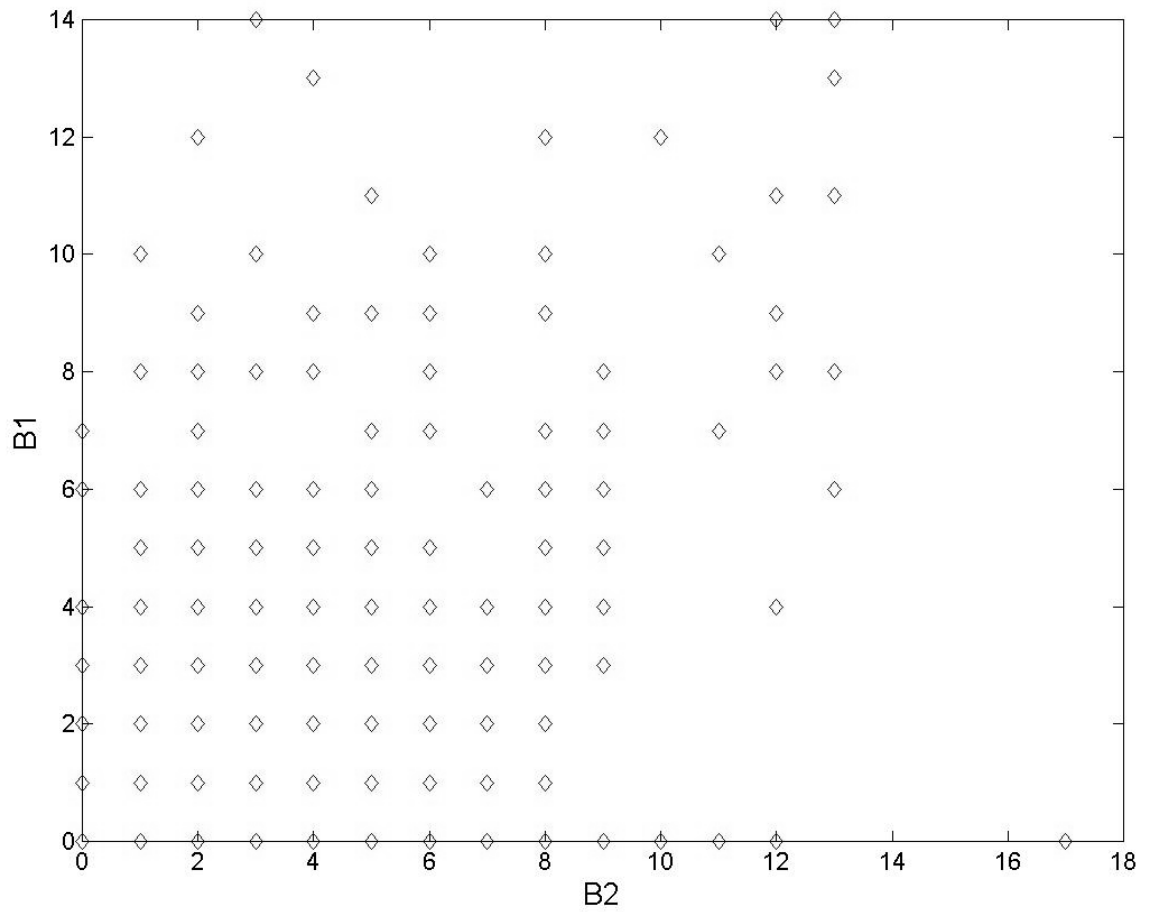


図 6.30 B1 と B2 の散布図

## 6.2 分類ごとの POS データ解析結果

### 6.2.5 POS \_\_ペットフードにおける解析結果

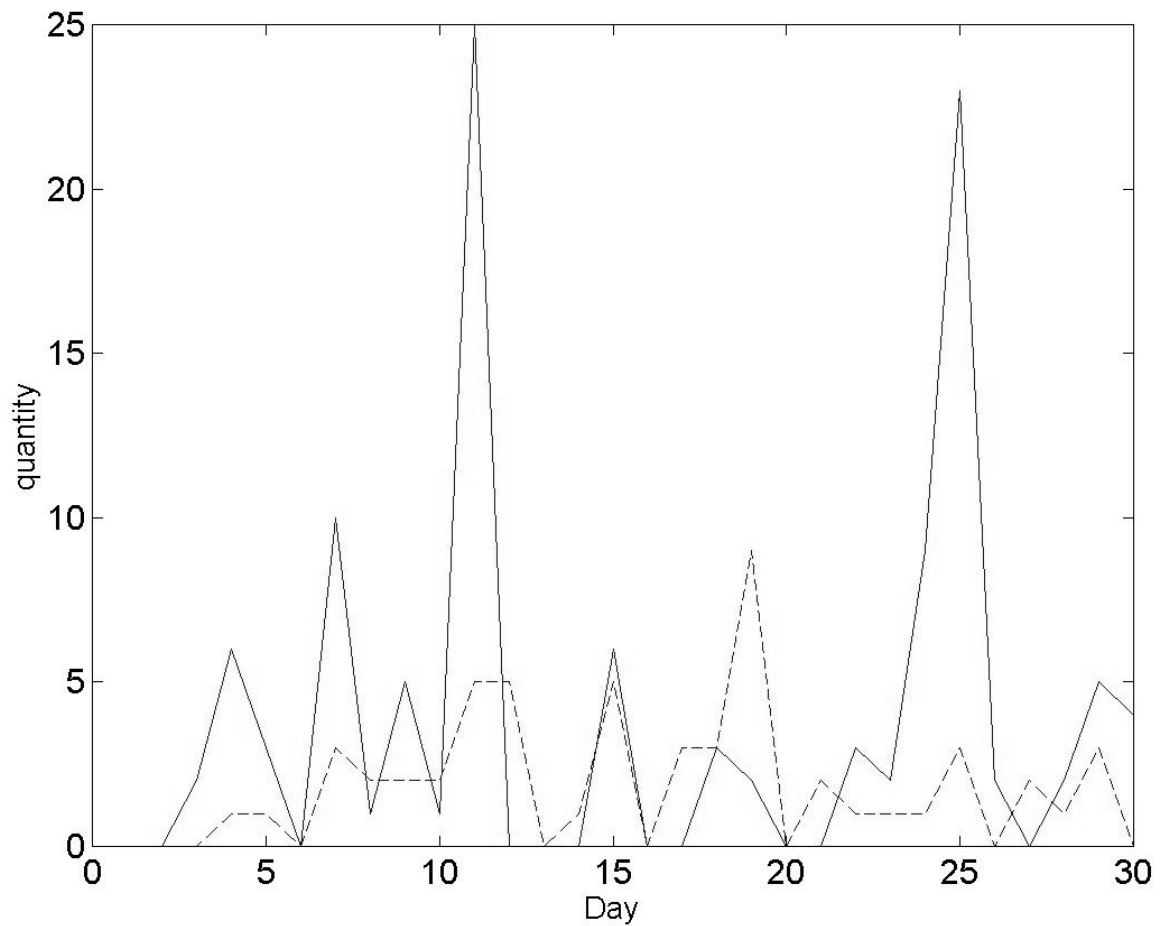


図 6.31 B1 と B2 の売上グラフ (実線が B1, 波線が B2)

## 6.2 分類ごとの POS データ解析結果

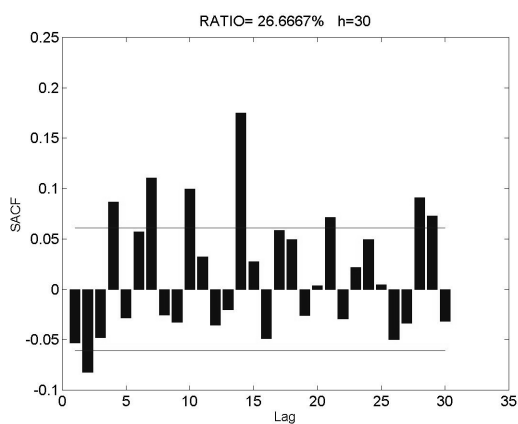


図 6.32 B1 の SACF

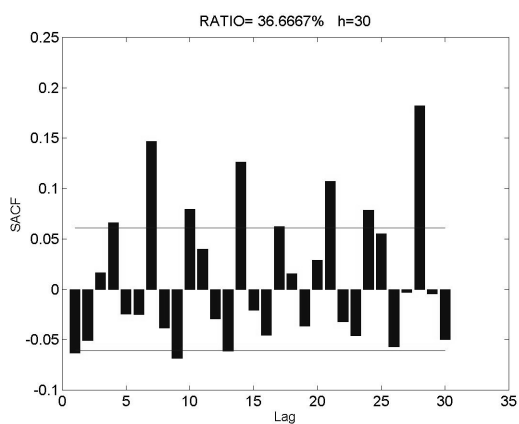


図 6.33 B2 の SACF

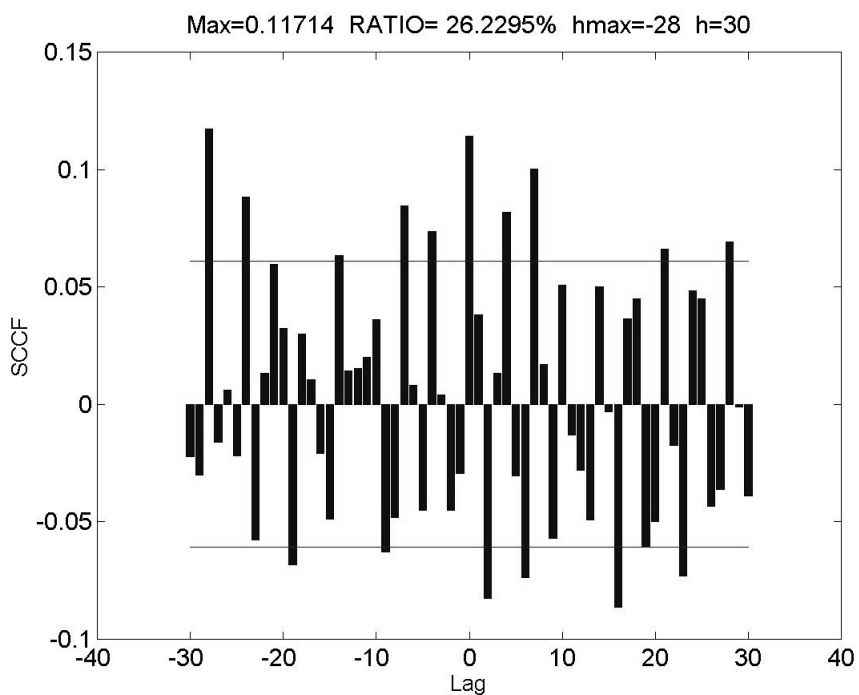
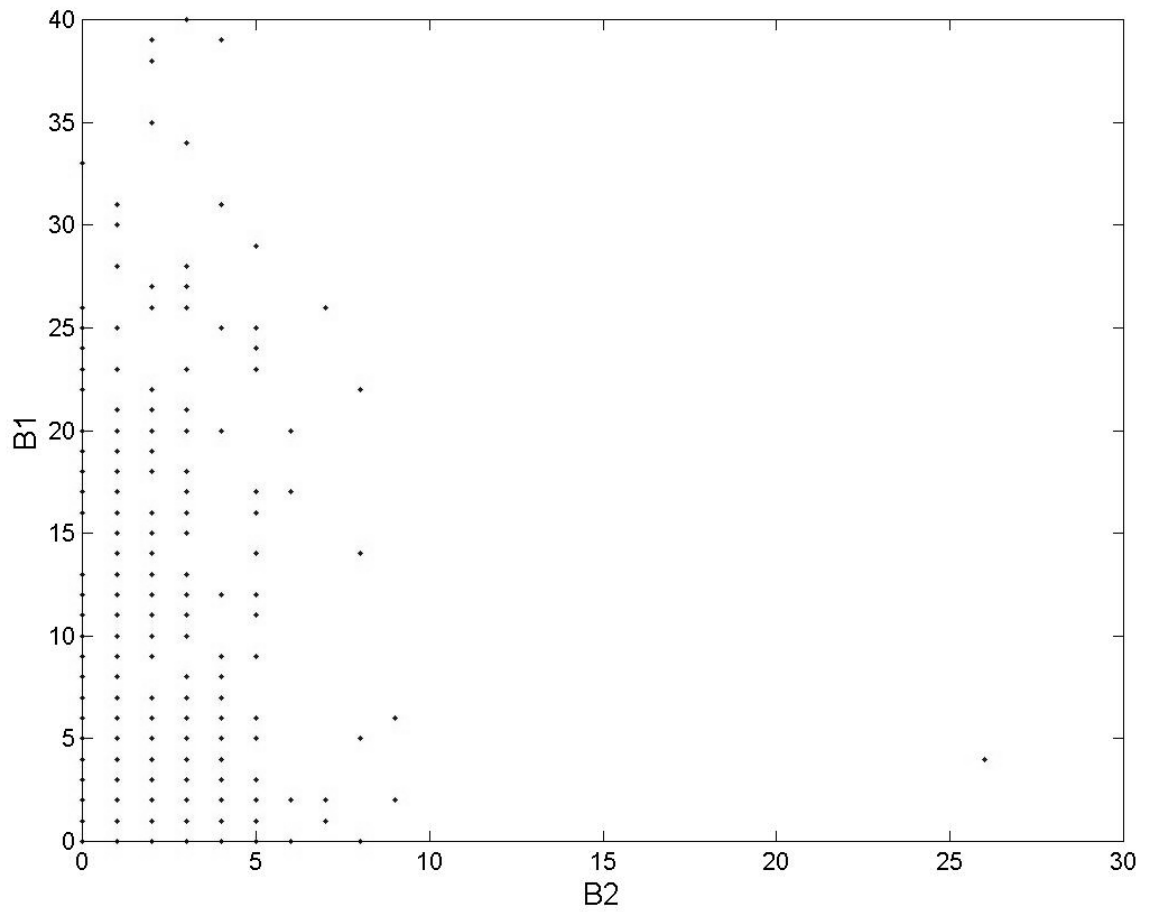


図 6.34 B1 と B2 間の SCCF

## 6.2 分類ごとの POS データ解析結果



# 第7章

## 結論

1. POS \_\_洗髪料洗髪料の BEST1 と BEST2 それぞれの SACF は高くない。したがってこれらの商品についてはこの手法では時系列解析には役に立たない。しかし、SCCF については、ラグが 0 の時、0.9 以上の高い相関係数が出た。ラグが 0 ということは 2 つの商品は同日にセットで買われているということである。また、ラグ=0 時の BEST1 と BEST2 の散布図でも強い正の相関が表れている。つまり、2 つの商品間には相互依存の関係が存在していることが判明した。ラグが 0 以外では信頼限界を超えていない割合が多く、時系列予測には役に立たない。他の 2 つの店の結果でもほぼ同様の相関が現れており、2 つの商品の相互相関は強い相関があるのを確認できた。
2. POS \_\_男性化粧品  
男性化粧品については、BEST1 の SACF はかなり低い値を出しているのに対し、BEST2 は割合が 60 % という高い相関が表れた。2 つの SCCF は 55 % と高いのではあるが、全体の係数が 0.15 以下であり、あまり予測に使えるという信頼は低いと思われる。散布図にかんしても無相関であることが確認できる。
3. POS \_\_化粧品  
化粧品に関しては、2 つの商品の SACF の割合に高い相関が表れている。しかし、SACF で見ると係数はかなり低い。この化粧品については SCCF よりも SACF での時系列解析を行ったほうが予測において有効であるとおもわれる。散布図で見ても無相関であることが確認できる。
4. POS \_\_生理用品生理用品に関しては、BEST1 商品の SACF が 100 % という高い数値がでた。また、SCCF に関しても、割合が 81 % と高かった。しかし散布図で見るとあ

まり相関があるとは思えない。しかし、SACF が 100 % という結果は、この生理用品は SACF での解析の方が有効であると思われる。つまり生理用品は購買周期に規則性があることが分かる。

#### 5. POS \_\_ペットフード

ペットフードに関しては、SACF 及び SCCF 双方低い相関であった。共同研究者である、(株) 高知流通情報サービスの考えでは、ペットフードは客が 1 度買ったものを買いつづける傾向にあり、他の商品に影響を及ぼさないのではないということであったが、本研究の結果によってそのことが実証できたと言えるだろう。

#### 6. まとめ

本研究の SCCF によって商品間の因果関係を突き止める事に有効な手段であることが確認できた。今後の課題としては、

- POS 全ての商品の SCCF を徹底的に解析する。(1 つの分類中には商品は数百あり、その組み合わせは数十万通りある。)
- 分類間を超えて SCCF を行う。
- 今回の得られた結果を以って販売予測に役立ててみる。

である。そうして、最終的には「ニューラルネットワーク等による発注支援システム」を完成させることである。そのために、その初期研究である、SCCF は時系列予測について強い味方となるであろう。

# 謝辞

担当指導員のラック助教授には、度々懇切なご指導とご助言をいただいた。解析した POS データは (株) サニーマートのご好意により提供されたものである。また、研究過程において (株) 高知流通情報サービスの方々には有益なアドバイス及びアイデアをいただいた。ここに上記の方々に対し、心からお礼申し上げる。

## 参考文献

- [1] 田島義博, 原田英生, “ゼミナール流通入門”, 日本経済新聞社, 1997
- [2] 大石展緒, 二宮政彦, “カオスの市場の販売予測”, 共立出版株式会社, 1998
- [3] 内村敬, “流通のしくみがわかる本”, ダイヤモンド社, 1997
- [4] 小林一行, “MATLAB ハンドブック”, 株式会社秀和システム, 1999
- [5] P.J. ブロックウェル, R.A. デービス, “入門時系列解析と予測”, CAP 出版株式会社, 2000
- [6] 赤池弘次, 北川源四朗, “時系列解析の実際 ”, 株式会社朝倉書店, 1995
- [7] 青柳龍也, “UNIX 短編シリーズ Perl”, クオリティ株式会社, 1997
- [8] Randal L.Schwartz, “初めての Perl”, ソフトバンク株式会社, 1995
- [9] 増井俊之, “Perl 書法”, 株式会社アスキー, 1993
- [10] 添田喬, 中溝高好, 大松繁, “信号処理の基礎よ応用”, 株式会社日新出版, 1979
- [11] 竹村彰通, “統計”, 共立出版株式会社, 1997
- [12] 柴田文明, “確率・統計”, 株式会社岩波書店, 1996
- [13] G.P. ウォズウェアス, J.G. ブライアン, “確率／統計の理論と応用”, ブレイン図書出版株式会社, 1986