

あるホームセンターにおける売れ筋商品と同時購入の多い商品の 売上分析

2007MI007 安藤 幹也 2007MI247 東崎 祥也

指導教員：松田眞一

1 はじめに

現在あるホームセンターでは、売れ筋や同時購入の多い商品について商品の売上げの特徴までは把握されていなかった。これにより売れ筋や同時購入の多い商品がいつ売れるのか分からず、急激に売れる時期に在庫数が足りずに欠品が発生してしまう可能性があると考えられる。欠品が起これば店側には本来得るはずであった利益が得られなくなってしまう。その商品が売れ筋商品であった場合、欠品による損失はより大きくなってしまふ。本研究では、こうした売れ筋商品の販売傾向を部門ごとに特徴を把握し、欠品を防止を目的とする。

2 データについて

データはあるホームセンターに関して A~F 店舗 (A 店舗は山に囲まれている店、B 店舗は海に面している店、C 店舗は太平洋側にある店、D 店舗は都会に立地している店、E 店舗は田舎に立地している店、F 店舗は日本海側にある店) の 6 店舗における 2007 年、2008 年、2009 年の各 52 週の売れ筋かつ同時購入の多い商品の販売数と特売数のデータを使用する。

日用品、カー用品、食品の 3 部門の商品のみ 3 年分のデータを、その他の部門については 2009 年度のデータのみを使用した。これは同じ商品でも年が違つと説明できる分析が変わつて一貫性はないが、部門ごとの特徴をつかむためには 2009 年度のみで十分だと考えたためである。

また各商品ごとの年間の欠品回数についてのデータも使用した。欠品は 6 店舗で合計 2 回以上起こつたものを欠品の起こつた商品として扱つた。

3 分析の流れ

1. 売れ筋と同時購入の関係から部門を 3 つに絞り分析方法を考える。
2. 周期性の確認し周期変動のある商品を除く。
3. 周期性の見られなかつたものに対しポアソン分布に従っているか確認する。
4. ポアソン分布に従つてないものに対してヒストグラムで今後の方向性を確認する。
5. 時系列分析で前の週の影響があるか確認する。
6. 分類ごとにまとめた部門を商品ごとに分けて周期性から見直す。
7. 全 27 部門の多くを周期性から見直しそれぞれの部門の特徴をつかむ。
8. 欠品の防止策を提案する。

4 分析方法

全 27 部門を分類ごとに一つ一つ分析していくのは莫大な時間を用するため、はじめに売れ筋でかつ同時購入の

商品数が多い部門 (日用品)、少ない部門 (カー用品)、またおよそ真ん中の値であると思われる部門 (食品) の 3 部門に絞つて考えた。まずこれら 3 部門についての売上傾向をつかむことで、その結果を他の部門に対しても応用して分析が行えると考えた。なお、カー用品のみ商品ごとに分けて調査した。理由はカー用品は分類ごとだと 2 種類にしか分けられなかつたからである。

まずはじめに、商品が購入される要因が無作為であるかどうかを調べるためにポアソン分布の適合度検定を行った。その際商品に周期性があると季節による変動が生じ、商品は無作為に購入されているとは考えづらいため、周期性のある商品は外して考えた。

次にポアソン分布で説明できなかつた商品に対して、売上の傾向をつかむためにヒストグラムを描いた。その結果、前の週の売上の影響を受けているのではないかと考え、時系列分析 (ARIMA モデル) で説明できるか調査した。

5 周期性の確認

ポアソン分布の確認をするために 3 部門の商品に対し、分類ごとに周期性の確認をすることにした。周期性が見られるものは、商品の売上数に規則性があり、無作為に売れているとは考えづらいため、周期性のある商品は除いた。ここで周期性のあるものの中には、年周期のものや半年周期、3 か月周期といったものもあつたが、これら全てを除いた。

6 ポアソン分布

6.1 ポアソン分布とは

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots$$

を確率関数として持つときパラメータ λ のポアソン分布という。 $P_o(\lambda)$ と表す。ポアソン分布の平均と分散は、 $E(X) = Var(X) = \lambda$ となる。これにあてはまれば、商品の売上は無作為だといえる。(白旗 [3] 参照)

6.1.1 カイ 2 乗分布とは

X_1, \dots, X_n を互いに独立で、かつ標準正規分布 $N(0, 1)$ に従う確率変数とする。このとき

$$X = \sum_{i=1}^n X_i^2$$

の分布を自由度 n のカイ 2 乗分布とよび、 χ_n^2 と表す。(白旗 [3] 参照)

6.1.2 カイ 2 乗検定について

検定統計量 χ^2 は

$$\chi^2 = \sum_{i=1}^k \frac{X_i^2}{np_i} - n$$

で表すことができる。

検定統計量 χ^2 は n が大きければ近似的に自由度 $k-r-1$ のカイ 2 乗分布 χ_{k-r-1}^2 に従う。

ここで r はパラメータ の次元、すなわち p_i を決めるパラメータ数である。これより、

$$R = \{ \chi^2 > \chi_{k-r-1}^2() \}$$

を棄却域とする検定は近似的に有意水準 となる。これをカイ 2 乗検定という。(白旗 [3] 参照)

6.2 ポアソン分布の適合度検定を行う前に

特売の週を加えると特売だから買っているという要因が考えられ、無作為で買っているとは考えづらいため、あらかじめ特売時の売上を抜いたデータを用いて検定した。また有意水準は 5 % で考えた。また、自由度はすべて $k-2$ となった。

7 ヒストグラム

ポアソン分布でなければ無作為ではなく何らかの理由で顧客は商品を購入されていると考えられる。その理由が何なのか模索するためにカイ 2 乗値の上側の確率が 0.050 より大きく下回っている商品のヒストグラムを見ていった。理由はカイ 2 乗値の上側の確率が大きく下回っている商品は、より無作為ではない何かの関係しているといえるからである。

また特売の週を加えると特売という理由だけで売れている日、売れていない日をはっきりすると考えられるので特売の週は除いた。

7.1 ヒストグラムの結果

二つの山になる分布が多かった。つまり、売れている日と売れていない日をはっきりしている。

前述のとおり、一つ前の売上が現在の売上に影響を与えているのではないかと考え、時系列分析 (ARIMA モデル) で解析をすることにした。

8 時系列分析

8.1 時系列分析とは

過去のデータを解析して、将来のデータを予測する方法である。時系列分析において時系列データは傾向と無作為なエラーの組み合わせであると考えられる。長期増加や長期減少、季節性等の傾向を理解することによって、傾向がエラーから分離され将来を推定することが可能になる。(田中 [1] 参照)

$$y_t = \sum_{i=1}^p a_i y_{t-i} + e_t + \sum_{j=1}^q b_j e_{t-j}$$

を自己回帰平均 (ARMA) モデルと呼び、この ARMA モデルに差分 (d) を加えたものを自己回帰和分移動平均 (ARIMA) モデルと呼び、 $ARIMA(p,d,q)$ と表す。今回は標準的な ARIMA モデルを使用した。(金 [2] 参照)

8.1.1 判別基準

ARIMA モデルを行う際に、データは特売も含んだ 52 週のデータを使用した。モデルの選択で自己回帰部分の次数が 0 となったものは前の週の影響を加味していないため除いた。また、前の週だけの影響を受けているとは考えづらい年周期も除いた。後ろ 12 週の予測値と、実際の売上値を比較した結果、12 週のデータがすべて予測値に近い値になったデータはなかったので 12 週全てを予測することはできなかった。そこで翌週 (41 週目) のみを予測できているものがあるのかを調べることにした。41 週目が予測できているか判断するための基準として、前 40 週の売上データから標準偏差を求め、その標準偏差の値より予測値と実測値の差が小さいものを予測できていると判断した。

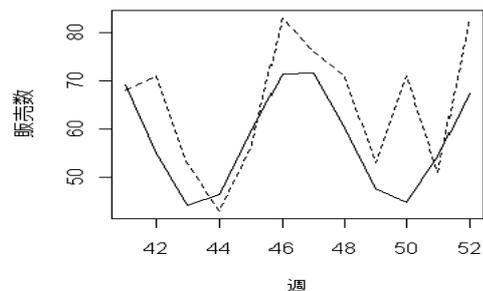


図 1 時系列分析で第 41 週が予測できた例

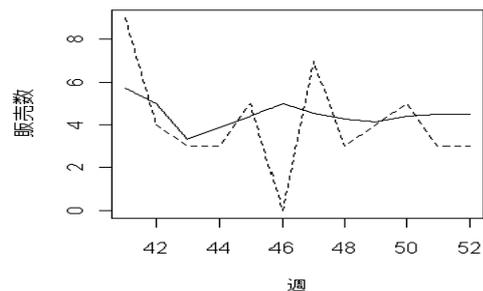


図 2 時系列分析で第 41 週が予想できなかった例

9 3つの部門の傾向と考察

初めに 3 つの部門を分類ごとと商品ごとに分析をした結果、下記の表のような結果になった。その結果、商品ごとのほうが説明できる数が増えたので商品ごとに前述の方法を繰り返して分析し、部門を広げていくことにした。

また、今までは各商品ごとに3年間のデータを扱ってきたが、これからは2009年のデータのみを分析していく。

9.1 分類ごとの3つの部門の考察

カー用品はすでに分類ごとではなく、商品ごとに分析し、約40%ポアソン分布で説明できている。そのため一定の確率で一定量売れる商品が多いと考えられる。つまり、約40%の商品は時期などに関係なく、必要時に各個人で買われているのではないかと考えられる。また年周期がほとんど見られなかったのではやはりカー用品は時期による影響が少ないと考えられる。

また約25%が時系列分析で説明できている。そのため前の週の売上データからの予測が可能である商品が存在する。この理由として特売に入る間隔が長く、購入機会が少ないため特売期間に買い、それ以外は買わない商品があると考えられる。

日用品は約80%がまだ説明できていない。

日用品は競合店やCM、広告などによる宣伝など入手することの難しいデータが販売数の変動幅に影響を大きく与えているため、特徴が把握しにくいのではないかと考えられる。ただし、今後商品ごとに分けたら何かしらの特徴が見つかることも考えられる。

食品は年周期が27%、周期性が31%と合計約60%が周期で説明できた。つまり食品は短いサイクルや年サイクルで売れる商品が多いと考えられる。例えばガムの場合、1月から3月に販売数が多いデータがある。この季節は受験シーズンで、眠気覚ましとしてガムを購入する顧客が増えるため周期性で説明できる数が多かったと考えられる。

表1 まとめ(分類ごと)

| | カー用品 | 日用品 | 食品 |
|------------|----------|----------|---------|
| 単位(個()内は%) | | | |
| 対象の数 | 108 | 54 | 30 |
| ポアソン分布 | 48(44.4) | 3(5.6) | 4(13.3) |
| 年周期 | 1(0.9) | 0(0) | 1(3.3) |
| 時系列分析 | 26(24.1) | 3(5.6) | 9(30) |
| 周期性 | 14(13) | 6(11.1) | 12(40) |
| 未説明 | 19(17.6) | 42(77.8) | 4(13.3) |

表2 まとめ(商品ごと)

| | 日用品 | 食品 |
|------------|----------|----------|
| 単位(個()内は%) | | |
| 対象の数 | 240 | 42 |
| ポアソン分布 | 39(16.3) | 10(23.8) |
| 年周期 | 11(4.6) | 11(26.2) |
| 時系列分析 | 23(9.6) | 2(4.8) |
| 周期性 | 58(24.2) | 10(23.8) |
| 未説明 | 42(45.3) | 9(21.4) |

10 商品別のデータ特徴の考察

部門別にどのような特徴が見られるか考察いく。表の見かたは 対:対象の数、ポ:ポアソン、時:時系列、周:周期性、年:年周期、未:未説明である。単位は個数で()内は%である。

10.1 部門ごとについて

部門ごとに何で説明できるかを調査した結果、説明できる分析方法が偏る傾向が多いことが分かった。つまり部門ごとに特徴が出るのが分かった。

10.1.1 ポアソン分布

HBC、カー用品、金物、スポーツ用品・玩具、工具、文具、塗料、ペット用品、寝具、家具収納はポアソン分布で説明できる割合が多かった。また、金物、スポーツ用品・玩具、工具、文具、塗料、ペット用品は50%以上がポアソン分布で説明できた。ポアソン分布で説明できた例として文具がある。文具は約60%がポアソン分布で説明出来ている。この理由はマジックやコピー用紙など消耗品で、必要時に各個人で購入される商品が多かったと考えられる。特にコピー用紙は、学校や会社において大量に消費されるものであり、一年を通して安定して売れる商品といえる。

10.1.2 周期

レジャー用品、冷暖房、日用品は周期性や年周期による季節変動で説明できる割合が多かった。また、レジャー用品、冷暖房は50%以上が周期で説明できた。例えば冷暖房は年周期も合わせると約60%が周期性で説明出来ているポリタンクと灯油ポンプが全て年周期で説明できた。これらは冬の時期に使われる商品であるので冬に多く売れていると考えられる。つまり季節の変動が売上に大きく影響を与えている。レジャー用品は約60%が周期性と年周期で説明できている。例えば、キャンプは夏のシーズンに行われるイベントなので季節変動が起きていると考えられる。ポアソン分布で説明できる商品の数が0である理由も主にこの理由であると考えられる。

10.1.3 時系列分析・ポアソン分布・周期

食品、家庭電器、木材は時系列分析・ポアソン分布・周期で説明できた。また、時系列分析で説明できる数が50%を超える部門はなかった。例えば食品は約25%がポアソン分布で説明出来ているので必要時に買われる商品があると考えられる。また約25%が時系列分析で説明出来ており、米が全て時系列分析となっている。米は値段が高いので特売の影響を受けており、特売の日にたくさん購入される顧客が多いからだと考えられる。さらに約30%が周期で説明出来ている。商品ではガムが周期で多く説明できている。受験シーズンの1月から3月に販売数が多いので眠気覚ましにガムを購入される顧客が増えるためだと考えられる。

10.1.4 売れ筋かつ同時購入の多い商品の数による特徴

売れ筋かつ同時購入の多い商品数が少ない部門はポアソン分布で説明できる割合の多い部門が多く見られた。無作為に売れている商品が多いということが分かった。

また、売れ筋かつ同時購入の多い商品数がおおよそ真ん中の部門はポアソン分布で説明できる部門とポアソン分布、時系列分析、周期性どれも平均的に説明ができる部門が多いことが分かった。

未説明の割合が約 40%を超えたものは日用品、HBC、寝具、家具収納があった。これらは売れ筋かつ同時購入の多い商品が比較的多い部門である。ここからも売れ筋かつ同時購入の多い商品が多い部門は未説明が多くなることが分かった。

また未説明となり何で説明できるか分からなかった商品のヒストグラムを見てみると二つの山になる分布が多かった。つまり、売れている日と売れていない日がはっきりしている。しかし、時系列分析でも説明できないので、売れている日と売れていない日がはっきりしているが週の影響ではなく、二つの山になっていると言える。競合店や CM、広告などによる宣伝などの影響が関係しているのではないかと考えられる。また年周期ではない複雑な周期の組み合わせで未説明のように見えている可能性も考えられる。

10.2 欠品防止策の提案

欠品の起こった商品の例として、木炭が店舗 B で多くの欠品が起こっている。また店舗 B 以外では木炭 3kg が店舗 A、C で欠品がおこり、木炭 6kg が店舗 C、D で起こっている。店舗 B の木炭は 6kg が年周期で説明できて、3kg が未説明になった。その他の店舗で欠品が起こったものは木炭 3kg が周期で 2 つ説明できて、木炭 6kg が年周期が 1 つと未説明が 1 つである。木炭の欠品の対策は海に近い店舗 B と他店舗で在庫の取り方を変えて、店舗 B では多く在庫を取る。また夏の時期に多く在庫を取る必要がある。

10.2.1 ポアソン分布

ポアソン分布で説明できる商品は無作為に売れている。欠品が生じた商品はいつ売れるのか分からないため、常に在庫を商品の売上の平均の数を取る必要がある。現在の発注ルーチンでは店舗ごとに直近平均を用いている。それには区分があり、欠品が生じるようなら区分を見直す必要がある。なお、発注ロジックについては大竹 [4] 参照。

10.2.2 周期

周期性が見られる商品は売上の平均どおり在庫を取っていればよいわけではなく、その商品や店舗ごとに販売数が増加する時期の在庫を多め取っておく必要がある。

10.2.3 時系列分析

時系列分析で説明できた商品に関しては前の週のデータから今週の売上が予測可能なので前の週の売上データから求めた予測値の値分在庫を取っていれば欠品は防げると考えた。

10.2.4 未説明

未説明の商品はその商品の特徴や売上の傾向、欠品が起こった店舗などを考慮して各商品にあった在庫の決定を行わなければならない。競合店や CM、広告などによる宣伝などのデータを手に入ればより精度の高い分析ができる可能性がある。

11 おわりに

本研究を終えて、各部門ごとの特徴をつかむことができた。分析結果ごとに欠品の対策を立てることができた。

表 3 商品ごとの分析結果

| | HBC | 食品 | 金物 | 工具 |
|---|----------|----------|-----------|----------|
| 対 | 90 | 42 | 51 | 36 |
| ポ | 27(30) | 10(23.8) | 27(52.9) | 22(61.1) |
| 年 | 0(0) | 2(4.8) | 1(2) | 0(0) |
| 時 | 5(5.6) | 10(23.8) | 7(13.7) | 2(5.6) |
| 周 | 19(21.1) | 11(26.2) | 9(17.6) | 5(13.8) |
| 未 | 39(43.3) | 9(21.4) | 7(13.7) | 7(19.4) |
| | 塗料 | 木材 | カー用品 | スポーツ |
| 対 | 82 | 54 | 108 | 18 |
| ポ | 48(58.5) | 9(16.7) | 48(44.4) | 10(55.6) |
| 年 | 4(4.9) | 0(0) | 1(0.9) | 1(5.5) |
| 時 | 5(6.1) | 13(24.1) | 26(24.1) | 3(16.7) |
| 周 | 17(20.7) | 15(27.8) | 14(13) | 3(16.7) |
| 未 | 8(9.8) | 17(31.4) | 19(17.6) | 1(5.5) |
| | レジャー | ペット | 日用品 | 文具 |
| 対 | 30 | 48 | 240 | 50 |
| ポ | 0(0) | 35(72.9) | 39(16.3) | 28(56) |
| 年 | 3(10) | 2(4.2) | 11(4.6) | 5(10) |
| 時 | 8(26.7) | 1(2.1) | 23(9.6) | 0(0) |
| 周 | 15(50) | 6(12.5) | 58(24.2) | 7(14) |
| 未 | 4(13.3) | 4(8.3) | 109(45.4) | 10(20) |
| | 家具収納 | 家庭電器 | 冷暖房 | 寝具 |
| 対 | 48 | 66 | 24 | 54 |
| ポ | 19(39.6) | 16(24.2) | 2(8.3) | 24(44.4) |
| 年 | 9(18.8) | 8(12.1) | 12(50) | 0(0) |
| 時 | 0(0) | 18(27.3) | 5(20.8) | 4(7.4) |
| 周 | 0(0) | 16(24.2) | 3(12.5) | 7(13) |
| 未 | 20(41.7) | 8(12.1) | 2(8.3) | 19(35.2) |

参考文献

- [1] 田中孝文：『R による時系列分析入門』，シーエーピー出版,2008.
- [2] 金明哲：『R によるデータサイエンス』，森北出版株式会社,1992
- [3] 白旗慎吾：『統計解析入門』，共立出版株式会社,1992
- [4] 大竹聡司：卒業論文『あるホームセンターの商品の自動発注ロジックの性能の分析』，南山大学数理工学系卒業論文,2010