

あるホームセンターにおける季節商品の販売数分析

2006MI100 三品光平 2007MI001 安部将成

指導教員：松田真一

1 はじめに

本研究では、あるホームセンターの季節商品の販売数予測を目的とする。その季節商品には殺虫剤の販売数を用いる。販売数予測モデルを作るにあたって、気象の影響や広告などの影響度を統計的に分析する。

2 データについて

2.1 商品データについて

夏季商品データは店舗A～F（店舗A～Dは太平洋側、店舗E～Fは日本海側）の6店舗における2005年度～2010年度（3月～10月）の週別の販売数を用いる。ただし、年度により売り始めの時期と売り終わりの時期が異なるので各年のデータ数を示す。2005年度23週、2006年度27週、2007年度34週、2008年度33週、2009年度28週、2010年度34週、合計179個のデータを用いる。6店舗すべて同時期に同じ季節商品を販売しているので店舗差はない。商品の種類は46種類あり、それを14分類に分けてその分類ごとの合計を求め、その合計したデータを使用する。

冬季商品データは店舗は夏季商品と同様の店舗で2006年度～2009年度（9月末～2月初め）各年21週を1年分のデータとし合計84個の週別の販売数を用いる。

2.2 気象データについて

気象データ [1] は、降水量・平均気温・平均風速・日照時間・最高気温を用いる。商品データの期間に合わせてデータをとり、降水量・日照時間は1日の合計の降水量・日照時間の週平均をとり、平均気温・平均風速はそれぞれの週平均をとり、最高気温は1日の最高気温の週平均をとる。

2.3 広告ダミー変数について

ある週の特売価格区分データがある。その特売の価格は、超目玉・大目玉・紹介のみ・特売なしの4つである。これらの特売価格区分に対しダミー変数を用いて3列の行列を作成する。

3 分析方法

主に重回帰分析を用いて分析を行う。季節商品の販売数を予測するために、気象データ・広告ダミー変数を用いて販売数にどのような影響があるか調べる。そして販売数の予測モデルを作成する。この際季節変動も考慮する。

本研究では、自由度調整済み決定係数を用い、以後表記は決定係数と省略する。

4 重回帰分析

重回帰分析とは、ある変数 y (目的変数) と、それに影響を及ぼすと考えられる他の変数 x_1, x_2, \dots, x_p (説明変数)

に関するデータに基づいて、

$$y_i = a_1x_1 + a_2x_2 + \dots + a_px_p + e_i \quad (i = 1, 2, \dots, p)$$
という線形回帰モデルを仮定して、回帰係数 a_1, a_2, \dots, a_p を推定し、推定された回帰係数を用いて、説明変数 x_1, x_2, \dots, x_p からある変数 y を予測する方法である。

今回の分析では予測させたい販売数を目的変数とし、その販売数を説明するための説明変数を気象データ・広告ダミー変数として重回帰分析を行う。少しでも予測精度を良くするため、決定係数が大きくなるよう検討していく。

5 季節変動

5.1 季節変動とは

季節変動とは、時系列データ Y をトレンド T 、季節変動 S 、不規則変動 I の3要素に分解した内の1つであり、決まった期間を周期とする定期的な波動を表している。また、夏季商品の場合3月～10月のみのデータであることと、各年度のデータ数が違うことから、全ての年度で重なっている期間を抽出してそれを23週で1周期としている。そして、それによって求められた季節変動を前半と後半で分けそれぞれで単回帰分析を行うことで重なっていない期間の欠測値を補完する。単回帰の予測値が負になる場合は0とした。冬季商品の場合は1年21週で統一されているのでそれを1周期とした。

$$Y(t) = T(t) + S(t) + I(t)$$

5.2 季節調整法

今回、季節変動を抽出するのに統計ソフトRにある `decompose()` 関数を用いた。この関数は季節変動の抽出に季節調整法を用いている。 $Y(t)$ の3要素への分解は、まず、 $Y(t)$ の中心化移動平均をとることでトレンドを抽出し、それを用いて $SI(t) = Y(t) - T(t) = S(t) + I(t)$ を得る。そして、この $SI(t)$ を23週周期（冬季では21週周期）で平均をとることによって季節変動を抽出することができる。

6 統一モデル

まず部門ごとかつ店舗ごとにおける重回帰モデルを作ったが、店舗が多くなるほど作成するモデルの数も増える。そこで、店舗を統一させ部門ごとのみにおけるモデルを作成することにする。夏季商品の場合、各店舗における販売数を1列に並べ145個×6店舗つまり1列870個になるようなデータ列を作る。これに対応させた、気象データや特売フラグデータ、季節変動の説明変数の行列を作成する。これらの行列をこれまでと同じように重回帰分析を行い全店舗で通用するようなモデルを作成する。各店舗における販売数を1列に並べる場合、店舗の差により販売数の大小が異なるのでそれを考慮して重回帰分析

を行わなければならない。そこで、基準化を用いて統一モデルを作成する。

6.1 基準化

基準化とは、あるデータをそのデータを含む集団の総和で割り、その和が1となる手法を指す。基準化をすると、ばらつきのあるデータを分析する場合に傾向がわかりやすくなる。そして、各店舗における販売数を1列に並べるとき、店舗ごとで基準化を行ってから1列に組み立てる。つまり、店舗Aにおける販売数を店舗Aの1年ごとの販売数の総和で割り145個のデータを作成する。同様の計算を残り5店舗に行い6店舗の列を縦1列にまとめる。店舗ごとで基準化を行う理由は、店舗ごとに販売数の差があるのでその差をなくすためである。基準化をすることにより店舗に関係なく販売数の傾向がわかる。各店舗ごとの予測モデルをみる場合、その店舗の総販売数を回帰係数にかけると通常モデルとして使うことができる。

7 夏季商品の検証について

7.1 分析方法

重回帰分析の統一モデルを用い、店舗差の表し方については基準化を用いる。

7.2 予測値の比率から販売数への復元

重回帰分析で予測モデルを作成する際、目的変数を販売数ではなく1年ごとに基準化した比率の値で置いているので、販売数の値に復元する必要がある。基準化したデータは、1年間の総販売数で割ることで比率の値に置き換えているので、比率の値で算出された予測値にその年の1年間の総販売数をかけることにより比率の値から販売数の値に復元することができる。しかし、かけなければならない総販売数は予測する年の実際の販売数の和なので予測する段階ではわかっていない。そのため復元するために1年間の総販売数を求める必要がある。

7.3 1年間総販売数の例年平均比率補正について

復元するためにかける1年間の総販売数を例年の平均販売数と今年の販売数の比率を用いて算出し、下記の式の様にして比率の予測値を販売数の予測値に復元する。ただし、販売数が少ない最初の7週間では1年間の例年平均総販売数をかけて復元する。

$$\begin{aligned} \text{販売数の予測値} &= \text{比率の予測値} \\ &\times \frac{\text{前の週までの販売数の和}}{\text{例年平均の前の週までの販売数の和}} \\ &\times \text{例年平均の1年間の総販売数} \end{aligned}$$

図1の1年間の例年平均総販売数をかけた場合のグラフをみると実測値と大きく離れていることがわかる。特に今年度は気温が異常に高かったことで販売数が例年よりも少ない傾向にあるので、その差がより顕著に表れている。それに比べて例年比で補正をかけたグラフは販

売数の山なりがうまく補正されており実際の1年間の総販売数をかけた場合とほとんど同じグラフとなっている。よって、例年平均販売数と今年の販売数の比率を用いて1年間の総販売数に補正をかけていく方法で年毎の変動を表すことができるといえる。

- A 販売数の実測値
- B 例年平均比率補正を用いて復元
- C 例年の1年間の総販売数を用いて復元
- D 実際の1年間の総販売数を用いて復元

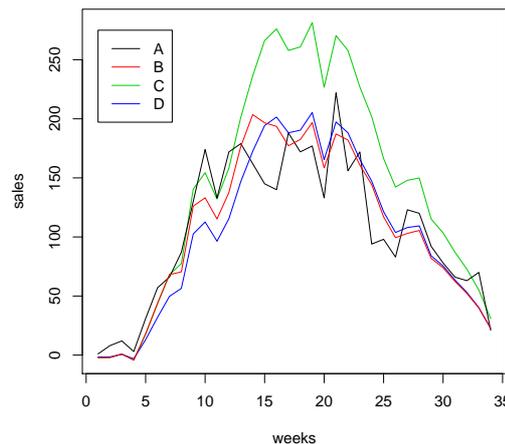


図1 各復元グラフ

実際 = 実際の1年間の総販売数
例年比 = 例年平均比率補正

表1 分類ごとの平均決定係数

	実際	例年比		実際	例年比
分類01	0.818	0.814	分類08	0.206	0.180
分類02	0.814	0.787	分類09	0.702	0.682
分類03	0.450	0.397	分類10	0.690	0.633
分類04	0.814	0.809	分類11	0.641	0.593
分類05	0.343	0.312	分類12	0.667	0.613
分類06	0.511	0.475	分類13	0.699	0.579
分類07	0.813	0.800	分類14	0.681	0.635

表1の実際の1年間の総販売数をかけた場合と例年平均比率補正を用いた場合の分類ごとの平均決定係数を比較すると、例年平均比率補正の決定係数が若干小さくなっているが、依然高い水準を維持できているといえる。これは、例年の平均販売数と今年の販売数との比率を用いた1年間の総販売数の補正がうまく働いている結果だといえる。

8 グラフの当てはまり具合

重回帰分析において決定係数は予測精度の指標である。しかし、その値がいくつなら良いかは、目的によって異なってくる。

ホームセンター側の最終目標は予測を元に販売数の上昇・ピーク・下降のタイミングを求める事であるため、予測値のグラフがそれらがわかる様な形になっているかが予測できているかいないかの判断基準となる。そこで各決定係数のグラフをみていき、どのくらいの決定係数ならば実用的なのかを確認していく。

各店舗各分類での予測値のグラフの当てはまり具合とその決定係数をみたとこ、まず、決定係数が0.7を超えているグラフは全体的に安定した予測ができています。次に、決定係数が0.5~0.7のグラフは実際のグラフの大まかな輪郭は合っており、販売数の上昇・ピーク・下降を捉えた予測ができていたといえるが、所々で予測が外れている箇所がある。最後に、決定係数が0.5未満のグラフ、特に0.4くらいのグラフから予測が外れる個所が多くなり、0.4よりも下だと予測がほとんどできていない。以上のことから、販売数の上昇・ピーク・下降がわかるのは決定係数が0.5前後までだとわかった。そこで、表1の例年平均比率補正の決定係数をみると14分類中10分類で決定係数が0.5よりも大きい値となっており、本研究の決定係数の基準を超えているといえる。

実測値と予測値のグラフを例として図2に示す。

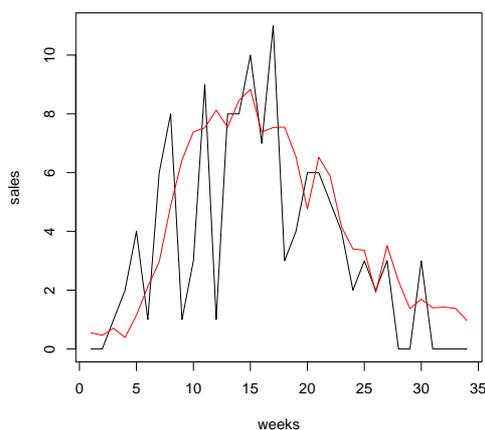


図2 店舗A 分類06 (決定係数: 0.491)

9 決定係数と販売数について

決定係数を各店舗、各分類ごとの平均値でみたとこ決定係数が大きいものや小さいものにわかれていることがわかった。グラフと決定係数を比較したとき決定係数が低いグラフは販売数が少ない傾向があるようだった。そこで、決定係数と販売数とを比較した。

まず、各店舗ごとの平均値をみてみると、店舗Eのみ決定係数が0.486と0.5よりも小さな値となっており次に小さい値の店舗Cの決定係数0.561とも大きな差があることがわかる。そこで、店舗Eの総販売数の店舗ごとの平均をみてみると、販売数が6店舗中最も少なく、その差も大きく販売数の総平均が625個なのに対して319個と約半分の販売数しかない。一方、他の店舗の平均決定係数は0.561から0.647と店舗Eよりも決定係数が約0.1

も大きく、総販売数の平均も472個から957個と幅はあるがどの店舗も319個の店舗Eとは販売数の差があるのは明らかである。

次に、各分類ごとの平均値をみてみると、分類によって決定係数に差があることがわかる。決定係数が大きい分類を挙げると分類01、分類02、分類04、分類07で、平均決定係数が0.8前後と大きな値になっている。そこで、それらの分類の総販売数をみてみると各分類の総販売数の平均が1000個よりも多い値になっており、他の分類よりも格段に販売数が多い。逆に、決定係数が小さい分類を挙げると分類03、分類05、分類06、分類08で、決定係数は0.5よりも小さな値となっている。それらの分類の総販売数をみてみると各分類の総販売数の平均が100個強と少ない値になっており、特に総販売数平均が100よりも小さい分類05、分類08の決定係数は0.312、0.180と非常に小さな値になっている。

最後に、各店舗各分類毎の決定係数と販売数の相関係数と相関図をみてみる。まず相関係数は0.590となっており正の相関がある。実際の相関図、図3をみてみると直線ではなく曲線を描いている様子にみえ、ある一定の販売数で急激に決定係数が大きくなる傾向にあると考えられる。

以上のことから各店舗各分類での決定係数の大小の差には販売数が関係していることが考えられる。

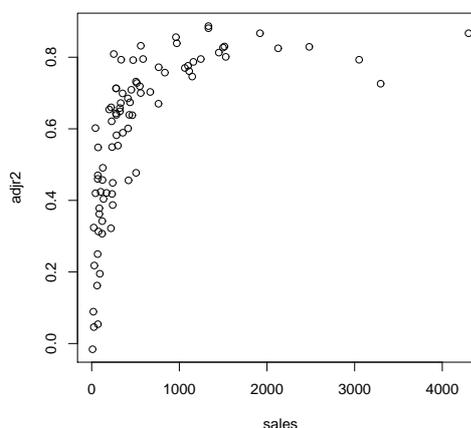


図3 相関図(夏季)

10 商品ごとの予測について

本研究では分類ごとで分析を行い予測値を算出しているが本来求めたいのは商品ごとの予測値である。そこで、商品ごとの予測ができそうかどうかみていく。

決定係数が特に良かった分類01、分類02、分類04、分類07の商品数をみてみると、分類01は3個、分類02は4個、分類04は10個、分類07は9個となっており、分類01、分類02は決定係数が高い上に商品数が少ないので商品ごとでの予測ができそうである。そこで、分類01を商品ごとで分析を行った。その結果、表2の様になった。各商品の内容をみてみると、商品01は高い水準で予

測ができていますが、商品 02、商品 03 は低い水準となっている。商品の内容についてみると、3 商品とも類似商品であり、商品 01 はお買い得パック、商品 02 と商品 03 はメーカー違いとなっている。そのため、商品 01 は他の類似商品の影響が少なくなり、一方、商品 02 と商品 03 では商品どうしの競合が起こり、売れ方が安定しないのではないかと考えられる。以上のことから、商品ごとで予測を行う場合、商品の内容によって決定係数に差が生じそうである。

表 2 商品ごとの平均決定係数

分類 01	商品 01	商品 02	商品 03
店舗 A	0.775	0.019	0.451
店舗 B	0.572	0.019	0.324
店舗 C	0.708	0.464	0.509
店舗 D	0.844	0.200	0.430
店舗 E	0.781	0.032	0.436
店舗 F	0.774	0.509	0.366
平均	0.742	0.207	0.419

11 冬季商品の分析

11.1 分析方法

夏季商品と同様の分析方法で分析を行った。

11.2 検証結果

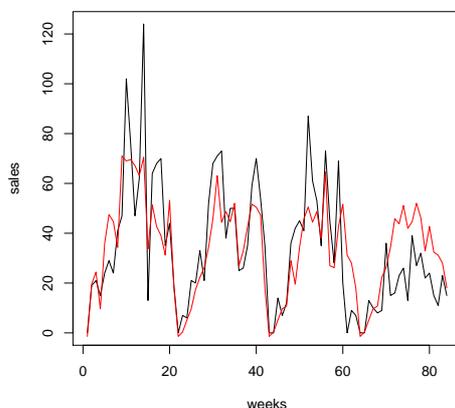


図 4 店舗 A 分類 11 (店舗別モデル)

図 4 の様に、販売数が極端に少なくなる年がありその減少幅を気象データ、広告データだけでは表せていない。この原因として、冬季商品は類似商品の新商品が多いことから今回の分析商品外の新商品による販売数のシェア争いが考えられる。

未来に関する予測精度を検証したとき、4 年分のデータの内 3 年分のデータから予測モデルを作成し、次の 1 年分のデータで予測精度の検証を行ったが 4 年分のデータで回帰分析を行ったときの決定係数から大きく下がってしまった。この原因としてデータ数の不足が考えられる。

夏季商品と同様に正の相関があり販売数が多いほど決定係数が高い傾向がある。しかし夏季商品の様に販売数の値が大きくないので販売数が一定の値を超えたときに急激に決定係数が高くなるという傾向があるとまではいえない。

以上のことから夏季商品と同様の分析方法でもある程度は予測できる可能性はあるが、データ数の不足や、データ数が数年分必要であるにも関わらず、新商品の登場により商品の種類が安定しない分類があることから、現時点ではそれ程の予測精度は出せず、また分類に関しても限定されてしまうと考えられる。

12 まとめ

本研究で行った大まかなことは、重回帰分析を用いた予測モデルの作成と、その予測モデルの検証作業である。まず、夏季商品について予測モデルの作成、検証を行った。予測モデルの検証の結果、低い水準の分類もあるが、高い水準で予測できている分類もいくつかあり、全体的にも比較的高い精度の予測モデルが作成できたといえる。しかし、商品ごとで分析を行う場合、商品の内容によって、予測精度に差が生じそうである。

次に、夏季商品での予測がうまくいったことから、冬季商品でも同様の方法で分析を行い販売数の予測ができないかと、分析し検証を行ったのだが、夏季商品ほどの予測精度は出なかった。

以上のことから、冬季商品に関しては不十分な結果となったが、夏季商品に関しては、ホームセンター側の最終目標である商品ごとの上昇・ピーク・下降の予測につながる結果が出せたのではないと思われる。

13 おわりに

今回の夏季商品の検証では未来の気象データがわかっているものとして予測モデルに当てはめ、検証を行ったが実際では、気象予報のデータを使用することになる。そのため、今回の研究で算出した予測精度から下がることが考えられる。

冬季商品の結果の原因として、類似の新商品により販売数のシェアを奪われた可能性などがあるが、このことは夏季商品（殺虫剤）はロングセラーが多いのに対して冬季商品（防寒具）は分類によっては商品の入れ替わりが激しいことを考えずに、同様の分析方法を行った結果である。つまり、夏季商品でうまくいったからといって冬季商品でもうまくいくわけではなく、季節ごとだけではなく分類、商品ごとの特性に合わせた新たな分析方法を加える余地があるといえる。

参考文献

- [1] 気象庁：気象統計情報, <http://www.jma.go.jp/> .
- [2] 金明哲：『R によるデータサイエンス』, 森北出版株式会社, 東京, 2007.
- [3] 田中孝文：『R による時系列分析入門』, シーエーピー出版株式会社, 東京, 2008.