

# あるコンビニの商品売り上げに関する統計的分析

2005MM021 伊藤 孝太

指導教員：木村 美善

## 1 はじめに

コンビニは24時間営業しており、食料品や雑貨、雑誌、ファーストフード（以下FFとする）など様々な商品が売られている。そこで、コンビニの商品群の中で手軽に食べられるFFで、どんなものが人気があり、どのくらい売り上げがあるのか、そして天候や値引きセールなどのイベント、夏休みなどの大型連休によって、また、新商品や季節商品の登場により売り上げがどのように変化するのかに興味を持った。私がアルバイトしていたコンビニのFFの売り上げデータを用いて、2007年6月～2007年8月のFFの売り上げについてクラスター分析法、重回帰分析法、ロバスト回帰法を用いて分析する（[1],[2],[3]参照）。

## 2 データについて

データは私がアルバイトしていたコンビニの2007年6月1日～2008年5月31日までの一年間のものであり、FFの一日の売り上げ高を目的変数 $y$ とし、説明変数として次の20個を用いた。 $x_1=CVS^1$ ,  $x_2=$ 入客数,  $x_3=$ 酒売り上げ高,  $x_4=$ たばこ売り上げ高,  $x_5=$ 湿度,  $x_6=$ 最高気温,  $x_7=$ 最低気温,  $x_8=$ 平均気温,  $x_9=$ 日照時間,  $x_{10}=$ 平均風速,  $x_{11}=$ 降水量,  $x_{12}=$ セール<sup>2</sup>,  $x_{13}=$ 火曜日,  $x_{14}=$ 水曜日,  $x_{15}=$ 木曜日,  $x_{16}=$ 金曜日,  $x_{17}=$ 土曜日,  $x_{18}=$ 日曜日,  $x_{19}=$ 晴れ,  $x_{20}=$ 曇りを説明変数とした（[4]参照）。FFのデータに関しては、 $X \cdot H$ セット<sup>3</sup>（ホットドック、チリドック、タルタルフィッシュサンド、バジルチキンサンド）、ホットドック、チリドック、タルタルフィッシュサンド、バジルチキンサンド、スパイシーチキン、カラアゲ、骨付きチキン、カリカリマン、チビタコ、 $X$ フライドポテト、ハッシュドポテト、ソフトクリーム（バニラ・ミックス・コーヒー）、パフェ（ $A \cdot B$ ）、ホットコーヒー、ハロハロ（ラムネ・メロン・宇治金時・レモン）、フルーツ氷（イチゴ・マンゴー）、タコタマ、ヒレカツサンド、 $X \cdot H$ セット（ヒレカツサンド）の35種類を用いた。図1のF1～F35はこの順である。

## 3 分析方法

### 3.1 クラスタ分析とは

類似度によって似たものに分類する方法（[3]参照）。

### 3.2 重回帰分析とは

目的変数 $y$ を $p$ 個の説明変数 $x_1, \dots, x_p$ で予測する分析法（[1]参照）。

### 3.3 ロバスト回帰法とは

仮定からのずれや外れ値からあまり影響を受けず、良さを少ししか損なわない回帰分析法であり、最小二乗法では検出できない外れ値の検出にも有効である（[2]参照）。

## 4 クラスタ分析の結果

全35種類のFFの売れ方について（縦：FF商品35種類、横：曜日、天候、セール、月の18変数）クラスター分析にかけた。図1を距離550辺りで切って、5群に分け、左から第1群、…、第5群とした。

**第1群**：骨付きチキン、ソフト（バニラ）でこの期間最も売り上げ個数が多かった群。

**第2群**：あまり人気がない商品で売り上げ個数が安定していない群。注文後、調理に時間がかかるため、人気がないようである。

**第3群**：この期間あまり売り上げ個数を伸ばせなかった群。第4群の商品に人気を集めたためであろう。

**第4群**：夏休みに入って売り上げ個数が伸びた群。

**第5群**：第1群より少し売り上げ個数が劣る群。

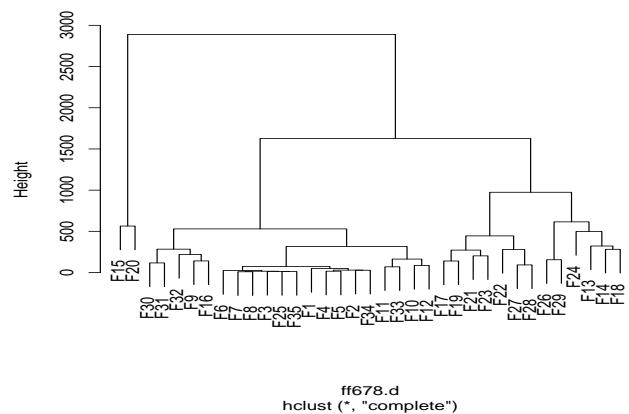


図1 FF全35種類の売り上げテンドログラム

### 4.1 考察

この期間は夏で気温が高く、コールド商品が人気を集めた。ハロハロ各種、ソフトクリーム（特にバニラ）、年間を通して人気のある骨付きチキンなどがよく売れた。

## 5 重回帰分析の結果

変数選択法（ステップワイズ法）にかけた結果、 $x_2=$ 入客数、 $x_3=$ 酒売り上げ高、 $x_7=$ 最低気温、 $x_9=$ 日照時間、 $x_{17}=$ 土曜日、 $x_{18}=$ 日曜日の合計6つの説明変数が残った。

重回帰式は

$$\hat{y} = -49.64937 + 0.05628x_2 + 0.26557x_3 + 0.92171x_7 + 0.78044x_9 + 14.00706x_{17} + 21.55303x_{18} \quad (4.1)$$

となる。この時の決定係数は0.72、自由度調整済決定係数は0.7002である。表1より、 $x_{17}=$ 土曜日、 $x_{18}=$ 日曜日の $p$ 値が0に近く、重回帰係数が大きな値であることから、FFの売り上げに影響を及ぼすことが分かる。休日は学生や親子連れの人が多いため、店に来るお客が増えている。

<sup>1</sup>FFの売り上げ高を除いた店全体の売り上げ。

<sup>2</sup>期間限定増量・値引きセールのこと。

<sup>3</sup> $X$ フライドポテトセット、ハッシュドポテトセットのこと。

そのため,FF を買って行く人も多くなっていると考えることができる。

表 1 6月～8月重回帰分析の結果

変数	係数	標準偏差	t 値	p 値
Intercept	-49.64937	11.58708	-4.285	$4.79 \times 10^{-5}$
$x_2$	0.05628	0.01857	3.031	0.00323
$x_3$	0.26557	0.18681	1.422	0.15880
$x_7$	0.92171	0.44062	2.092	0.03944
$x_9$	0.78044	0.31582	2.471	0.01546
$x_{17}$	14.00706	3.41280	4.104	$9.27 \times 10^{-5}$
$x_{18}$	21.55303	3.20673	6.721	$1.95 \times 10^{-9}$

### 5.1 月別による FF の売れ方の特徴

回帰係数を用いて,月別の FF の売れ方の特徴を見るためにクラスター分析を行った。スペースの都合上,結果のみを示す。

- ・第 1 群：8 月を表し,夏休みである。
- ・第 2 群：3 月,4 月を表し,春休みがある。
- ・第 3 群：5 月,7 月を表し,連休と夏休みがある。
- ・第 4 群：2 月,6 月,10 月を表し,祝日など休日が少ない。
- ・第 5 群：1 月,9 月,11 月,12 月を表し,冬休みや祝日などの休日が第 4 群よりある。

第 1 群,・・・,第 5 群に分けた結果,休日の日数により,それぞれ群分けした月同士で似たような売れ方をしていると考えられる。

## 6 ロバスト重回帰分析の結果

ここでは,最小二乗法では見つけられない外れ値の検出を行う。

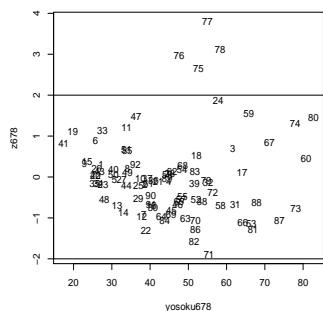


図 2 LS 残差プロット

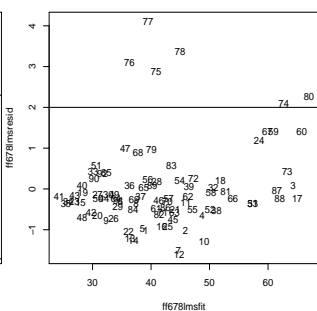


図 3 LMS 残差プロット

図 2 は最小二乗法による残差プロット図で,番号 75,76,77,78 が外れ値となっている。しかし,LMS 法(図 3)では,図 2 と違い,最小二乗法では検出できなかった外れ値 74,80 を検出できた。LMS 法を用いた重回帰式の結果は次のようになる。

$$\hat{y} = -28.03579 + 0.05839x_2 - 0.15010x_3 + 0.41159x_7 + 0.62259x_9 + 8.86444x_{17} + 16.43686x_{18} \quad (6.1)$$

番号 74,75,76,77,78,80 を削除後,再度回帰分析を行った。

重回帰式は

$$\hat{y} = -37.77891 + 0.07351x_2 - 0.15143x_3 + 0.22281x_7 + 0.37471x_9 + 17.80777x_{17} + 24.42304x_{18} \quad (6.2)$$

となる。そして,このときの決定係数は 0.833,自由度調整済決定係数 0.8203 である。ここで LMS 法で見つけた外れ値を見てみる。この期間の分析では,番号 74,75,76,77,78,80 を外れ値として検出できた。これらは順に,8月12日(日),13(月),14(火),15(水),16(木),18(土)を表し,すべてお盆の時期である。県外などから帰省する人や仕事休みの人が多く,店に訪れるお客が増えたことで外れ値となったのだろう。また,LS 法で見つけられなかった外れ値 74,80 を見てみる。番号 74:8月12(日)はお盆の始まりの時期である。番号 80:8月18(土)は,翌日の日曜日はお盆の終わりの時期である。日曜日は,県外などへ帰る人が多かったからか,土曜日にお客が増えたのだろう。そして,外れ値削除後の重回帰式を見てみると,(6.2)式より,土曜日,日曜日における重回帰係数が(4.1)式より少し大きくなった。このことから,FF は土曜日,日曜日に売り上げが見込めることが分かる。

## 7 まとめ

6月～8月の期間は一年の中でも一番売り上げ高があり,特にハロハロ各種,ソフトクリームなどのコールド商品が人気を集めた。冬季は中華まんなどのホット商品が人気を集める結果となった。一年間で常に人気のある骨付きチキンと,その期間の季節商品がよく売れることが分かった。よって,季節商品の売れ方で,FF の売り上げ高が変化している可能性があると考えられる。また,どの期間においても,土曜日,日曜日に FF の売り上げが見込める。そして,ロバスト回帰法による分析で,主に祝日,お盆,お正月の期間が外れ値として現れているので,休日や特別な期間においても売り上げを見込めることが分かった。ここでは載せていないが,9月～11月と12月～2月の期間において  $x_{12}$ =セールスの説明変数が残ったことから,主に冬季に増量・値引きセールスの効果があると考えられる。夏季などの気温が高い時期は,コールド商品が人気を集めるためか,あまり効果がないようである。

本研究では,クラスター分析法や,回帰分析法を中心に分析を行ったが,その他の分析手法を用いてもより詳しく分析できるだろう。また,曜日別での FF の売れ方について分析し,比較しても面白い結果が得られたであろう。

## 参考文献

- [1] Chatterjee,S.and Price,B.:Regression Analysis by Example,Wiley,1977(加納悟,佐和隆光訳:回帰分析の実際,新曜社,1981).
- [2] 神谷美紀・水戸藍・竹内亜希代:ロバスト回帰の研究,南山大学数理情報学部数理科学科卒業論文,2006.
- [3] 松下和哉:アイスクリームの販売に関する統計的分析,南山大学数理情報学部数理科学科卒業論文,2003.
- [4] 気象庁電子閲覧室過去の天気 <http://www.data.kishou.go.jp/>