卒業研究論文

気象，曜日による食品の売り上げの変化について

学籍番号　01D8103002E

鈴木　涼

中央大学理工学部情報工学科　田口研究室

2005 年 3 月

あらまし

食品の売り上げに影響を与える要因は，多数あると考えられる．本研究では，多数ある要因の中で気象，曜日の 2 つを考え，食品の売り上げとの関係を調べる．食品の売り上げは，福岡県にあるスーパーの ID 付き POS データを用い，気象データは気象庁が公開しているデータを用いる. 気象データのうち, 最高気温, 降水量, 日照時間に着目する. 曜日のうち,来店者が増えると予測される土曜，休日（祝日も含む），さらに曜日ごとの来店者数から特徴ある曜日を選び出し，これらの曜日に着目する．気象，曜日を独立変数，食品の売り上げ個数を従属変数として重回帰分析を行う. さらに変数ごとの標準化回帰係数から食品の売り上げと気象，曜日の関係を数値的に捉える.

キーワード：重回帰分析，標準化回帰係数

目次

第 1 章　はじめに

現在，少子化による 1 世帯あたりの消費支出が減り，さらに長引く不況の影響から消費活動が減少傾向にある．そのため，商品の売り上げを予測することは，店側にとって重要な課題である．売り上げを予測することができれば，売り場面積や，発注量などを調整することにより，利益を最大にすることができる．そこで，売り上げに影響を及ぼす要因を見つけ出し，その要因がどの程度影響を与えているのかを分析することが重要である．本研究では，商品の売り上げを知るための有用なデータである，ID 付き POS データを用いる．ID 付き POS データは，顧客ごとの詳細な購買行動が記録されており，多くのマーケティング戦略の基盤となる研究に使われている．

また，「暑い日に冷たいものが売れる」，「日曜日は家族連れが多いので子供が欲しがる商品が売れる」など，気象，曜日，商品の特売日，顧客の嗜好などが，商品の売り上げに影響を与える要因として考えられる．

本研究では，気象と曜日に着目し，スーパーで扱っている商品のうち食品を分析対象とする．気象，曜日が食品の売り上げにどの程度影響しているかを分析する．

第 2 章　使用データ

本研究では，商品の売り上げを調べるために ID 付き POS データを，気象を調べるために気象データを用いる．なお，気象データは気象庁がホームページで公開しているものを用いる．

2.1　ID 付き POS データ

POS とは「Point of Sales」の略称であり，レジで発生した単品レベルの販売記録を，リアルタイムで仕入れ･在庫情報等に反映させることをコンピュータ上で行うシステムである．本研究では，各商品の売り上げを調べるため，福岡県にあるスーパーの 2000 年 5 月 1 日から同年 9 月 30 日までの 1 日ごとの ID 付き POS データを用いる．このデータは，誰が，いつ，何を，何個買ったかが記録されており，以下に示す 16 の項目に分類されている．

- 顧客番号
- 生年月日
- 性別（1：男性，2：女性）
- 郵便番号
- 住所 1（県名）
- 住所 2（町名，字など）
- 住所 3（町名，字など）
- 購入年月日
- ラインコード
- ライン名
- クラスコード
- クラス名
- 商品コード
- 商品名
- 購入金額
- 購入点数

ライン名，クラス名は商品の分類名であり，ライン名がより大きな分類となっている．本研究では，ラインごとの食品の売り上げに着目する．ラインは全部で 42 種類あり，そのうち食品関連のラインは 22 種類である．表 2.1 に食品関連のライン名と，それに属するクラス名の一部を示す．

表 2.1　食品関連のライン名とそれに属するクラス名の一部

ライン名	属するクラス名
アイス	コウキユウアイス，ノベルテイ
エンカン	エンカン，カイソウ，サケ・マス
カコウヒン	ハム，カコウヒン
ギユウニク	ギユウナマシヨク，ナマ・ホルモン
クダモノ	リンゴ，ブドウ，ナシ
ケイニク	トリニク，ヤキトリ
コメ	コメ，モチゴメ
シコウヒン	インスタントコーヒー，カツプソクセキメン
センギヨ	エビ，カニ，サシミ
ソウザイ S	ソウザイ，チキン，フライ
タマゴ	ツウジヨウラン，トクシユラン
チヨウミリヨウ	チヨウミリヨウ，カレー・シチユールー
ニユウセイヒン	ギユウニユウ，ヨーグルト
ネリモノニツパイ	ネリモノ，ハムソーセージ，メン
ノウカイサンカンブ	ノリ，カンメン，コナルイ
パン	シヨクパン，カシパン
フクロガシ	スナツク，チヨコレート，ビスケツト
ブタニク	ブタスライス，ブタニク
ベイハン	オニギリ，ベントウ
ミズモノニツパイ	アゲ，トウフ，コンニヤク，ツケモノ
ヤサイ	ヤサイ，キセツルイ
レイシヨク	オカズ，チユウカ

2.2　気象データ

気象庁のホームページ[4]では，前日までの各観測地点における気象データを公開している．気象データの値は観測地点によって，気象台・測候所で観測された値と，アメダスで観測された値に分類されている．また，ホームページ閲覧者が指定した観測期間（1 日の毎時の値，1 ヶ月間の毎日の値，1 年間の毎月の値，平均値（日別）の 4 種類）ごとに表示できる．本研究では，福岡県飯塚市の気象台・測候所で観測された値を用いる．このデータは

・ 平均現地気圧[hPa]
・ 平均海面気圧[hPa]
・ 平均気温[℃]
・ 最高気温[℃]
・ 最低気温[℃]
・ 平均相対湿度[%]
・ 最小相対湿度[%]
・ 平均風速[m/s]
・ 最大風速[m/s]
・ 最大風速を記録したときの風向
・ 最大瞬間風速[m/s]
・ 最大瞬間風速を記録したときの風向
・ 日照時間[時間]
・ 降水量[mm]
・ 最大 1 時間降水量[mm]
・ 最大 10 分間降水量[mm]
・ 天気概況（昼）

の 17 項目について 1 日ごとの値が記録されている．さらに

・ 平均現地気圧[hPa]
・ 平均海面気圧[hPa]
・ 平均気温[℃]
・ 平均相対湿度[%]
・ 平均風速[m/s]
・ 最多風向
・ 日照時間[時間]
・ 降水量[mm]

の 8 項目について，月を上旬（1 日～10 日），中旬（11 日～20 日），下旬（21 日～月の末日）の 3 期に分けた 1 期ごとと，その月を通しての観測値が記録されている．最高気温，最低気温の算出方法を表 2.2 に示す. 本研究では, 食品の売り上げに関連する気象要素として, 最高気温，降水量，日照時間を考慮する．

表 2.2　各データの算出方法

日最高気温	任意の時分の観測値で最も高い値
日最適気温	任意の時分の観測値で最も低い値

第 3 章　重回帰分析

食品の売り上げと，食品の売り上げに影響を与えていると考えられる要素を調べる．また関係がある場合，どの要素がどの程度影響しているのかを調べる手法として，重回帰分析を用いる．

この章の記述は主に[1，3，5]による．

3.1　重回帰分析とは

回帰分析の目的は，複数の変数のデータがあるとき，あるひとつの変数を他の変数で説明するような式を求めることである．そのような式のことを回帰方程式，回帰モデルなどと呼ぶ．説明される変数をYで表し，従属変数，非説明変数，内生変数などと呼ぶ．説明する変数をXで表し，独立変数，説明変数，外生変数などと呼ぶ．本論文では以下Yを従属変数，Xを独立変数と表記する．なお，質的変数を独立変数とするとき，ダミー変数を用いる．ダミー変数とは，値が 0 か 1 しかとらない変数である．回帰方程式とはYをXで説明する式であるが, YがXの線形関数である場合を線形回帰, それ以外の場合を非線形回帰と呼ぶ. ここでは線形回帰についてのみ考える．また従属変数が 1 つの場合を単回帰分析，または単純回帰分析といい，2 つ以上の従属変数を考える場合を重回帰分析という．

ある従属変数Yをk個の独立変数X_j（$j=1,2,\cdots,k$）で説明する場合を考える．母集団において

$$Y_i=\beta_1+\beta_2X_{2i}+\cdots+\beta_kX_{ki}+u_i \qquad (i=1,2,\cdots,n) \tag{3.1}$$

というモデルを考える．このモデルを母回帰方程式と呼び，$\beta_1,\beta_2,\cdots,\beta_k$を母回帰係数と呼ぶ．これは母集団の値であるから一般にはわからない．これについて推定，検定するのが回帰分析である．またu_iは誤差項と呼ばれ，次の 3 つの条件を満たす確率変数とする．

- 期待値は 0 である
- 分散σ^2は一定である
- 異なった誤差項は無相関である

このことは既に定まった値$X_{2i},X_{3i},\cdots,X_{ki}$に対応して変数$Y$が誤差$u_i$を含んで$Y_i$という値をとるが，その確率変数のとりうる値の期待値が$\beta_1+\beta_2X_{2i}+\cdots+\beta_kX_{ki}$であることを示す．

次にk個の未知の母回帰係数$\beta_1,\beta_2,\cdots,\beta_k$を推定する．（3.1）式から誤差項$u_i$は

$$u_i=Y_i-(\beta_1+\beta_2X_{2i}+\cdots+\beta_kX_{ki})$$

と表せる．ここでこの誤差項u_iを最小にすることを考える．まずu_iの符号の影響を取り除くために

$$S=\sum_{i=1}^{n}u_i{}^2=\sum_{i=1}^{n}\{y_i-(\beta_1+\beta_2X_{2i}+\cdots+X_{ki})\}^2$$

として，Sを最小にする$\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_k$を求めそれらを$\beta_1,\beta_2,\cdots,\beta_k$の推定量とする．この考

え方に基づく推定方法を最小 2 乗法と呼び，$\hat{\beta}_i$ を β_i の最小 2 乗推定量という．$\hat{\beta}_i$ は 1 次の編微分を 0 と置いた k 個の連立方程式

$$\partial S / \partial \beta_1 = 0, \partial S / \partial \beta_2 = 0, \cdots, \partial S / \partial \beta_k = 0$$

を解くことによって得られる．$\hat{\beta}_i$ のことを標本回帰係数という．また

$$Y = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$$

を標本重回帰方程式という．各 i の期待値 $E(Y_i)$ は

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$$

で推定される．これを回帰値という．

次に各独立変数がどれだけ従属変数に影響を与えているかを判断する場合を考える．上で求めた各独立変数の標本回帰係数を比較しても，その独立変数が従属変数に影響を与えている割合の指標にはならない．仮にある独立変数の単位を cm として，回帰式を求めたとする．つぎに単位を mm に変えて回帰式を求めると，その独立変数の標本回帰係数の値は 1/10 になる．また，そもそも単位の違うもの同士を比較しても（例えば独立変数 A の単位が cm，独立変数 B の単位が個の場合など）意味がない．そこで各変数の値を，それぞれの標本平均 $\overline{Y}, \overline{X}_2, \overline{X}_3, \cdots, \overline{X}_k$，標本標準偏差 $s_y, s_{x2}, s_{x3}, \cdots, s_{xk}$ を用いて標準化し

$$(Y_i - \overline{Y})/s_y, (X_{2i} - \overline{X}_2)/s_{x2}, (X_{3i} - \overline{X}_3)/s_{x3}, \cdots, (X_{ki} - \overline{X}_k)/s_{xk}$$

について重回帰分析を行う．これによって求められた回帰係数のことを標準化回帰係数という．

3.2 回帰式の有効性

3.2.1 推定値の標準誤差

回帰式（値）の当てはまりの良さを評価する方法がいくつかある．本節では推定値の標準誤差について述べる．

実測値 Y_i の，回帰値 $\hat{Y}_i$ からのずれを回帰残差 e_i といい

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki} \qquad (3.2)$$

から求める．回帰残差は X で説明されずに残った分である．また回帰残差 e_i は以下の条件を満たす．

$$\sum_{i=1}^{n} e_i = 0,\ \sum_{i=1}^{n} e_i X_{2i} = 0,\ \sum_{i=1}^{n} e_i X_{3i} = 0,\ \cdots,\ \sum_{i=1}^{n} e_i X_{ki} = 0 \qquad (3.2)$$

いま u_i の分散 σ^2 を

$$s^2 = \sum_{i=1}^{n} {e_i}^2 / (n-k) \qquad (3.3)$$

で推定する．これは残差の平方和 $\sum {e_i}^2$ を自由度 $(n-k)$ で割ったものである．なお，回帰残差 e_i は（3.2）の k 個の条件をみたすために制限が加わり，自由度が k 失われている．このと

き s を推定値の標準誤差といい，$s.e.$ で表す．$s.e.$ が小さいほど回帰式はよく適合していることを示す．

3.2.2　決定係数（寄与率），自由度調整済み寄与率

本節では, 回帰式の当てはまりの良さをはかる基準のひとつである決定係数（寄与率）と, それを補正した自由度調整済み寄与率について述べる.

Y_i のばらつきの総和（変動）は $\sum_{i=1}^{n}(Y_i-\bar{Y})^2$ で表される（$\bar{Y}$ は Y_i の標本平均とする）．これを X の回帰方程式で説明できる変動と，説明できない変動の 2 つに分けると

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2+\sum_{i=1}^{n}{e_i}^2$$

となる（前節で述べたように, 回帰残差 e_i は X で説明されなかった部分である）. このうち X の回　帰方程式で説明できる変動の割合のことを決定係数 η^2，または寄与率といい以下の式より求める.

$$\eta^2=\frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}=1-\frac{\sum_{i=1}^{n}{e_i}^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})} \qquad (3.4)$$

また決定係数 η^2 の，正の平方根のことを重相関係数といい R で表す.

η^2 は 0 から 1 の間の値をとり，値が大きいほど回帰式の当てはまりが良いことを示す．しかし決定係数は独立変数の数を増やすと，その変数が有用なものではない場合でも，前より高い値になってしまう欠点がある．そこで新たに独立変数を増やしたときに，その変数が有用でない場合，決定係数の値が下がるように自由度で（3.4）式を補正したものを自由度調整済み寄与率 R^{*2} という．R^{*2} は

$$R^{*2}=1-\frac{n-1}{n-k-1}(1-\eta^2)$$

から求める.

3.3　重回帰分析における有意性の検定

3.3.1　仮説検定

仮説検定の目的は，母集団について仮定された命題を，標本に基づいて検証することである．まず帰無仮説と，対立仮説の 2 つの仮説をたてる．帰無仮説のことを H_0，対立仮説のことを H_1 と表す．この 2 つの仮説は互いに否定の関係にある．帰無仮説に否定したい仮説をたてる．この帰無仮説の有意性についての検証を行う．検証を行った結果，帰無仮説が有意でないと判断されたとき，帰無仮説を棄却するといい，逆に有意であると判断されたときは，帰無仮説を採択するという．このときの判断基準のことを，有意水準 α という．有意水

準αは任意に定める．つぎに片側検定を行うか，両側検定を行うかを決める．一般に，母数の値がある目標値と等しいかどうかだけを調べる場合，両側検定を行う．母数の大きさが理論的，経験的に予測される場合，片側検定を行う．つぎに検定統計量をもとめる．検定統計量とは，検定に用いる統計量のことであり，検定方法ごとに違う．検定統計量がある分布x（検定統計量ごとに違う）に従うことを利用し，求めた検定統計量の分布xにおける起こりうる確率（これを有意確率といい，pで表す）と有意水準αを比較し，pがα以下のとき帰無仮説を棄却し，対立仮説は有意であると判断する．それ以外のときは帰無仮説を採択し，対立仮説は有意ではないと判断する．有意確率は，統計数値表から得ることができる．

3.3.2　重回帰方程式の検定

重回帰方程式の有意性の検定を行う．まず，$X_2, X_3, \cdots, X_k$のすべてがYを説明しないという帰無仮説H_0を

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

とし，$X_2, X_3, \cdots, X_k$のどれかひとつはYを説明しているという対立仮説H_1を

$$H_1 : \beta_2, \beta_3, \cdots, \beta_k \text{ の少なくともひとつが 0 ではない}$$

として，有意水準αを定め両側のF検定を行う．検定統計量Fは

$$F = \frac{\left(\sum_{i=1}^{n} (Y_i - \bar{Y})^2 - \sum_{i=1}^{n} e_i^{\,2} \right) \Big/ k-1}{\sum_{i=1}^{n} e_i^{\,2} \Big/ n-k}$$

で算出する．帰無仮説H_0が正しい場合，Fは自由度$(k-1, n-k)$のF分布$F\ (k-1, n-k)$に従うので，有意水準αとFのF分布$F\,(k-1, n-k)$における有意確率pを比較し，pがα以下ならば重回帰方程式は有意であり．それ以外のとき重回帰方程式は有意ではない．

3.3.3　標本回帰係数の検定

本節では標本回帰係数 1 つ 1 つに対する検定の方法を述べる．

（3.3）式で求めた誤差項の分散σ^2の推定値s^2から，$\hat{\beta}_i$の標準誤差$s.e.(\hat{\beta}_i)$を求める．いまs_{ij}を

$$s_{ij} = \sum_{p=1}^{n} (X_{pi} - \bar{X}_i)(X_{pj} - \bar{X}_j)$$

とおく．s_{ij}をij要素としてもつ行列のことを，独立変数X_i, X_j間の分散共分散行列という．s_{ij}と区別するために，（3.3）式で求めたs^2をV_eで表す．このとき，独立変数間の分散共分散行列をsとし，さらにsの逆行列のii要素をs^{ii}とするとき，$\hat{\beta}_i$の標準誤差$s.e.(\hat{\beta}_i)$は

$$s.e.(\hat{\beta}_i) = \sqrt{s^{ii} V_e}$$

である．この値を用いて標本回帰係数の検定を行う．帰無仮説H_0を

$$H_0 : \beta_i = 0$$

とし，対立仮説 H_1 を

$$H_1 : \beta_i \neq 0$$

として，有意水準 α で両側の t 検定を行う．検定統計量 t_0 は

$$t_0 = \frac{\hat{\beta}_i}{s.e.(\hat{\beta}_i)}$$

から求められる．t_0 は自由度 $(n-k-1)$ の t 分布 $t(n-k-1)$ に従うので，t_0 の t 分布 $t(n-k-1)$ における有意確立 p を求め，p が有意水準 α 以下ならば回帰係数 $\hat{\beta}_i$ は有意であり，それ以外のときは有意ではない．また t 分布と F 分布の関係から

$$F = \frac{\hat{\beta}_i^{\,2}}{(s.e.(\hat{\beta}_i))^2} \qquad (3.5)$$

となるので，求めた検定統計量 F の F 分布 $(1, n-k-1)$ における有意確立 p と有意水準 α を比較し，p が α 以下ならば回帰係数 $\hat{\beta}_i$ は有意であり，それ以外のときは有意ではない．

3.4　変数選択法

重回帰方程式の精度は，用いた独立変数によって定まる．従属変数をまったく説明していない独立変数のみで回帰式を求めても，その回帰式は意味を持たない．また，仮に求めた回帰式がひとつの重要な独立変数を含んでいたとしても，それ以外の重要な独立変数を含んでいなければ精度はよいものではない．重回帰分析において，従属変数を良く説明している変数を選択し，そうではない変数は選択しないということは，重要な課題のひとつである．独立変数の選択が不適切なとき，次のような問題が起こる．

①回帰式に無駄な独立変数（真の回帰係数が 0）が含まれているとき，標本回帰係数 $\hat{\beta}_i$ や回帰値 $\hat{Y}$ は不偏であるが，誤差項の分散推定値 V_e の自由度が小さくなるため，推定精度が悪くなる．

②必要な独立変数（真の回帰係数が 0 でない）が回帰式からもれているとき，$\hat{\beta}_i$ や $\hat{Y}$ は偏りを持ち，誤差項の分散推定値 V_e は過大評価になる．

③独立変数の中に，互いに相関の高い変数が含まれる場合には，分散共分散行列の行列式がほとんど 0 になるため，逆行列の要素の値が大きくなり，$\hat{\beta}_i$ の推定精度が悪くなる．また各独立変数と従属変数の相関係数の符号と，標本回帰係数の符号が一致しない場合が生ずる．ある独立変数で予測しすぎた部分を，他の変数で打ち消している場合がある．このような場合，多重共線性の問題がある．

このような問題を避けるため，変数の選択が重要になる．理論的基盤があるときはそれに従

って変数を選択すればよい．理論的基盤がないときは，統計的方法を用いて変数を選択する方法がある．統計的手法には

- 総当り法：すべての説明変数について回帰式を作成しどの回帰式が良いかを検討する方法.
- 逐次変数選択法：各回帰係数の有意性に基づいて，有効な変数と不要な変数を振り分ける方法．変数増加法，変数減少法，変数増減法，変数減増法の 4 つの方法がある．

などがある．本研究では変数増減法を用いる．変数増減法は，（3.5）式から求められる回帰係数ごとの F 値によって, 変数の取捨が判断される．まず, 変数を取り込む F 値の限界値 F_{in} と，変数を除去する限界値 F_{out} を定めておく． F_{in} は重要な変数を落とさないことに重点を置くなら小さい値，無駄な変数を取り込まないことに重点を置くなら大きい値を指定する．経験的に 2.0 を F_{in}，2.0 未満を F_{out} と置くことが多い．手順は次のようになる．

手順 1　どの変数も選ばれていない状態から始める．

手順 2　もし，すべての変数が含まれていれば，取り込むべき変数はないという情報をもって手順 3 に進む．すべての変数が含まれていなければ，残りの変数を順番に 1 つずつ採用してみて，回帰係数の検定のための F 値を計算し，その値が最大となる変数を選ぶ．その F 値が指定された F_{in} より大きければその変数を取りこんで手順 3 に進む． F_{in} より小さければ，取りこむべき変数はないという情報をもって手順 3 に進む.

手順 3　回帰式に含まれている変数について，回帰係数の検定のための F 値を計算し，F 値が最小となる変数を選ぶ．その F 値が指定された F_{out} より大きいとき，取りこむべき変数がないという情報があれば終了する．そうでなければ，どの変数も落とさず手順 4 に進む． F 値が F_{out} より小さいとき，その変数を落とし（取りこむべき変数がないという情報があれば，それを取り消して)，再び手順 3 に戻る．

手順 4　すべての変数が取りこまれていれば終了し，そうでなければ手順 2 に戻る．

第 4 章　分析結果

食品の売り上げに影響を与えていると考えられる要素を述べる．それらを独立変数とし，各ラインの売り上げ個数を来店者数で割った値を従属変数として重回帰分析を行う．

4.1　独立変数の決定

食品の売り上げに関連があると考えられる気象と曜日について調べる．

まず，気象要素について，最高気温[℃]，前日の最高気温と当日の最高気温の差[℃]，前 7 日間の最高気温の平均と当日の最高気温の差[℃]，前 3 日間の最高気温の平均と当日の最高気温の差[℃]，降水量[mm]，日照時間[時間]を調べる．気温について図 4.1，図 4.2 に，降水量と日照時間について図 4.3 に示す．期間は，2000 年 5 月 1 日～同年 9 月 30 日のうち，データがある 150 日間である．

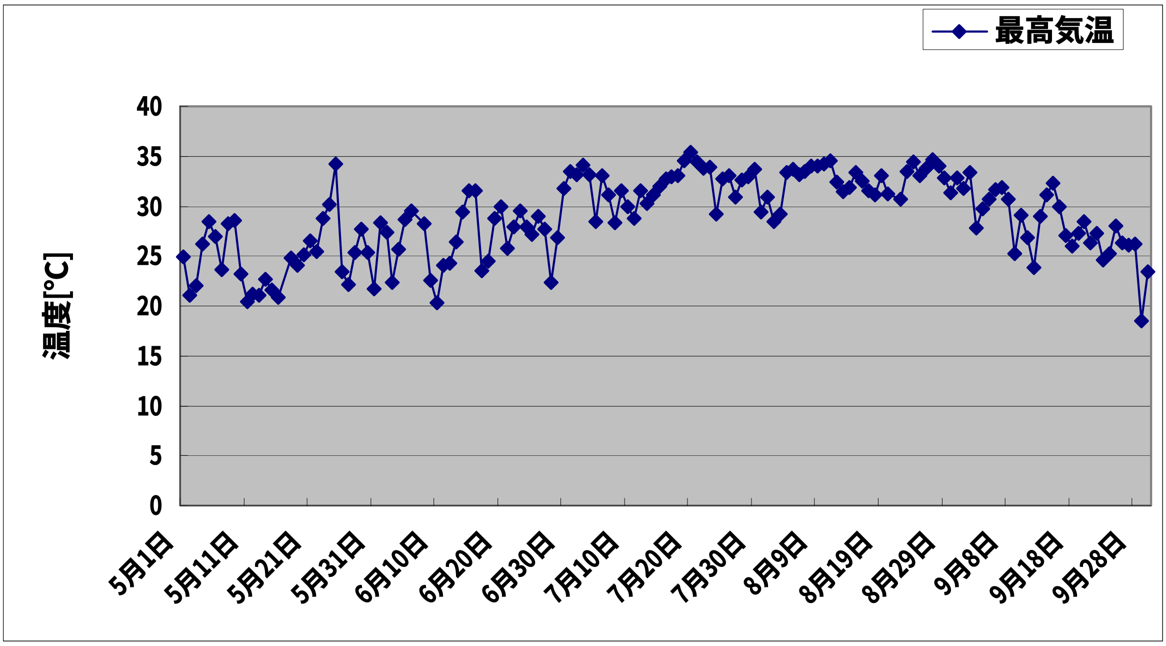

図 4.1　最高気温の推移

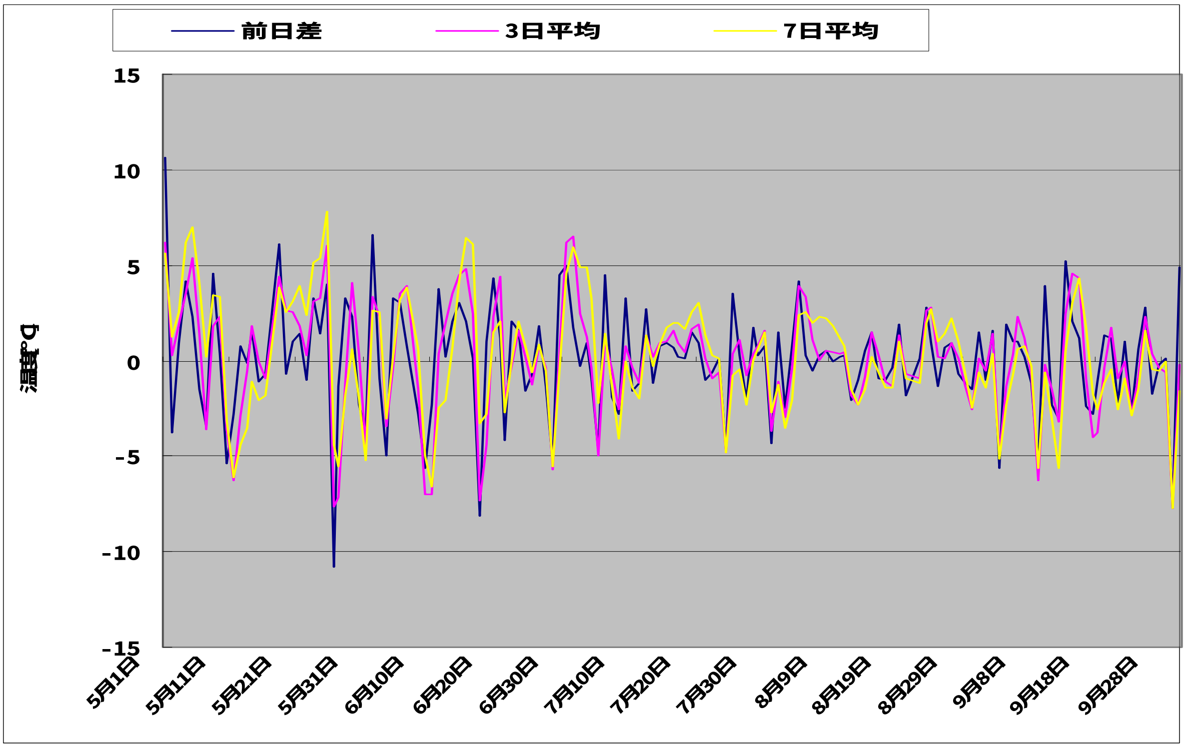

図 4.2　最高気温以外の気温要素

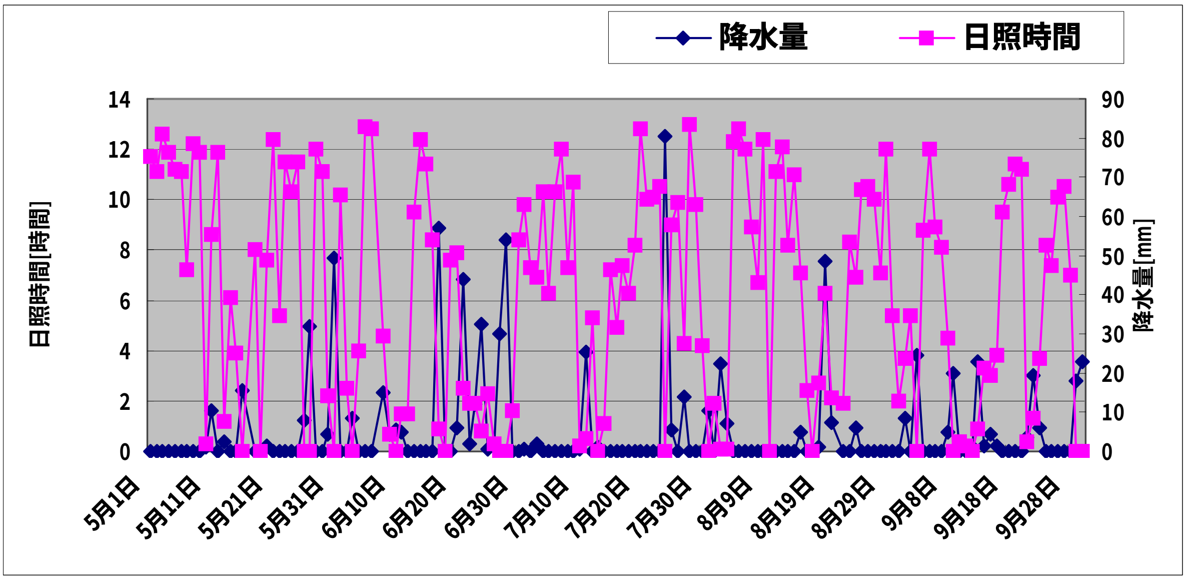

図 4.3　降水量と日照時間

次に，曜日ごとの来店者数を調べる．図 4.4 に分析対象期間の曜日ごとの総来店者数を，図 4.5 に平均来店者数を示す．2 つのグラフから，日曜日と火曜日の来店者数が多いことがわかる．一般に土日，祝日は来店者数が増えることが予測される．しかし，土曜については木曜，金曜と差は見られなかった．また，祝日に関しても，祝日ではない同じ曜日の日と来店者数を比較したが，差はみられなかった．

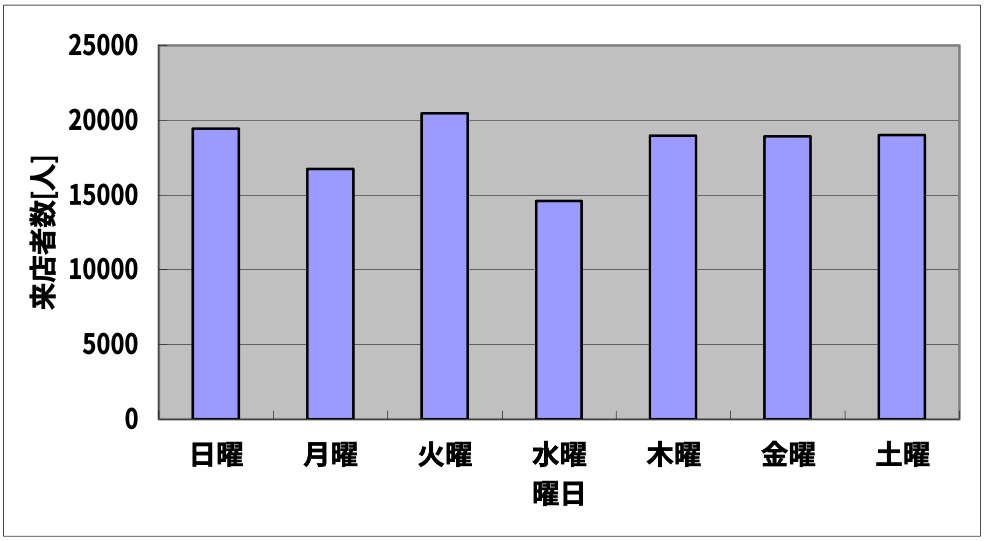

図 4.4　曜日ごとの総来店者数

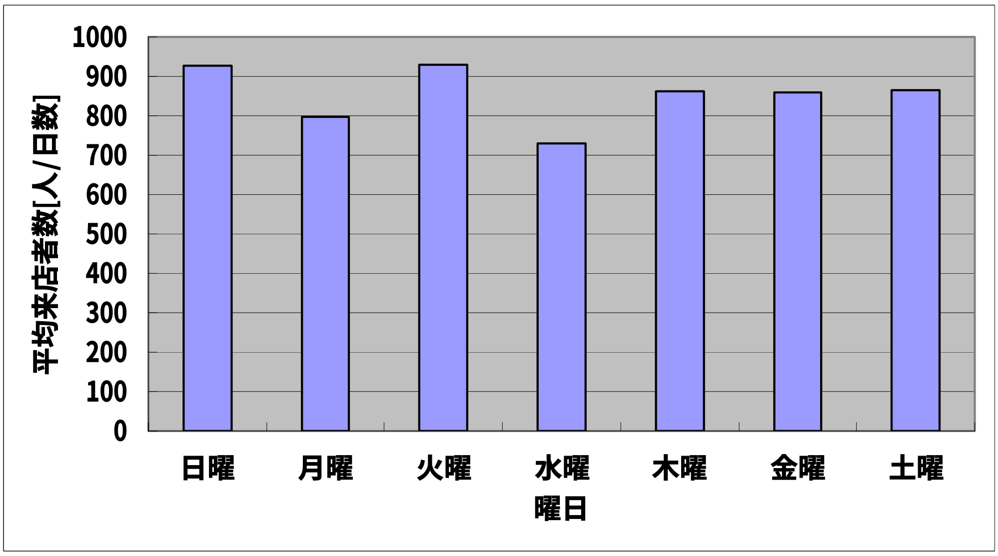

図 4.5　曜日ごとの平均来店者数

気象に関しては，最高気温，前日の最高気温と当日の最高気温の差，前 7 日間における 1 日ごとの最高気温の平均と当日の最高気温の差, 前 3 日間における 1 日ごとの最高気温の平均と当日の最高気温の差，降水量，日照時間を変数とする．曜日は，平均来店者数の多い日曜，火曜をそれぞれ変数とする．また，土曜・祝日は，月曜，水曜，木曜，金曜と来店する客層が異なると考え, ひとつの変数とする. このとき, 曜日を示す変数はダミー変数とする.

まず，3.4 節で述べた問題を回避するため．変数の相関係数を求めた．各変数の相関係数を表 4.1 に示す（ただし，表 4.1 において「前日差」は前日の最高気温と当日の最高気温の差を,「前 7 平均」は前 7 日間における 1 日ごとの最高気温の平均と当日の最高気温の差を,「前 3 平均」は前 3 日間における 1 日ごとの最高気温の平均と当日の最高気温の差をそれぞれ示している）.

最高気温の前日との差，前 7 日間における 1 日ごとの最高気温の平均との差，前 3 日間における 1 日ごとの最高気温の平均との差の相関が高い. そこで最高気温の前 7 日間平均との差，最高気温の前 3 日間平均との差の 2 変数を除外した，7 変数を独立変数とする.

表 4.1　各変数間の相関係数

	最高気温	降水量	日照時間	前日差	前 7 平均	前 3 平均	日曜
日照時間	0.3601	-0.4468					
前日差	0.3015	-0.3836	0.4131				
前 7 平均	0.4513	-0.4216	0.4873	0.7654			
前 3 平均	0.5052	-0.4079	0.5529	0.6098	0.8709		
日曜	-0.0139	-0.1138	0.0153	0.0469	-0.0862	-0.0446	
火曜	-0.0015	-0.0364	0.0836	-0.0021	0.0668	0.0159	-0.1673
土・祝日	-0.0748	0.1375	0.0187	-0.0894	-0.0403	0.0307	-0.1890

4.2　回帰係数の導出

ラインごとに重回帰分析を行い，回帰係数を導出する．ラインごとの 1 日の売り上げ個数をその日の来店者数で割った値を従属変数とし，4.1 節で定めた 7 変数を独立変数とする．以下，簡単のため独立変数に番号を割り当てる．変数番号と変数の対応を表 4.2 に示す．

全ラインで，変数選択法を用いない場合，変数選択法を用いた場合のそれぞれを重回帰分析した結果の自由度調整済み寄与率を表 4.3 に示す．

表 4.2　変数番号と変数の対応表

変数 1	最高気温
変数 2	降水量
変数 3	日照時間
変数 4	当日の最高気温と前日の最高気温の差
変数 5	日曜日
変数 6	火曜日
変数 7	土曜日・祝日

表 4.3　全ラインの自由度調整済み寄与率

ライン名	変数選択法無	変数選択法有	ライン名	変数選択法無	変数選択法有
ｱｲｽ	0.294729	0.305669	ﾁﾖｳﾐﾘﾖｳ	0.127981	0.139592
ｴﾝｶﾝ	0.071826	0.092703	ﾆﾕｳｾｲﾋﾝ	0.030630	0.045463
ｶｺｳﾋﾝ	0.124452	0.119328	ﾈﾘﾓﾉﾆﾂﾊﾟｲ	0.303684	0.302799
ｷﾞﾕｳﾆｸ	0.187692	0.200167	ﾉｳｶｲｻﾝｶﾝﾌﾞ	0.013849	0.030238
ｸﾀﾞﾓﾉ	0.254548	0.262883	ﾊﾟﾝ	0.054439	0.066370
ｹｲﾆｸ	0.064732	0.068936	ﾌｸﾛｶﾞｼ	0.386657	0.390664
ｺﾒ	-0.013050		ﾌﾞﾀﾆｸ	0.151256	0.164894
ｼｺｳﾋﾝ	0.227063	0.233718	ﾍﾞｲﾊﾝ	0.283876	0.290589
ｾﾝｷﾞｮ	0.131076	0.138726	ﾐｽﾞﾓﾉﾆﾂﾊﾟｲ	0.107409	0.119892
ｿｳｻﾞｲS	0.069526	0.087581	ヤサイ	0.644051	0.649611
タマゴ	0.426616	0.423279	レイショク	0.788799	0.792190

変数選択法の有無に関わらず自由度調整済み寄与率が 0.6 以上となった，ヤサイ，レイショクの 2 つのラインについて結果の詳細と，実データとの比較を以下に示す．3.1 節で述べた，各独立変数の従属変数に与える影響の度合いを示す標準化回帰係数の値に着目する．なお，コメに関しては変数選択法を用いなかった場合，自由度調整済み寄与率が負の値になっており，また変数選択法を用いた場合回帰方程式が求まらない．これは，設定した 7 つの独立変数が，コメの売り上げにほとんど影響していないことを示す．

・ヤサイ

ヤサイの重回帰分析結果を表 4.4，表 4.5 に示す．表 4.4 の標準化回帰係数の値を見ると変数 6 の係数の値が特に高く，変数 1 の係数の値が負の方向に高い．変数 1 と変数 6 の回帰係数が，有意水準 0.05 で有意である．回帰方程式の有意確率は 0 に近く，回帰方程式は有意である．変数選択法を用いた場合，変数 1，変数 2，変数 6，変数 7 が選ばれる．この 4 変数の中でも変数 6 の標準化回帰係数の値は特に高い値となっている．変数 1 は最高気温を表す変数，変数 2 は降水量を示す変数，変数 6 は火曜日を示すダミー変数，変数 7 は土曜日・祝日を示すダミー変数である．よって重回帰分析の結果から，ヤサイの売り上げは火曜日に著しく増加するといえる. また最高気温の高い日, 雨の降っている日に減少し, 土曜日, 祝日に少し増加するといえる．ヤサイの売り上げ個数を来店者数で割った値（以下，ラインごとの売り上げ個数を来店者数で割った値のことを, そのラインの平均売り上げ個数と表記する）と最高気温との比較を図 4.6 に，降水量との比較を図 4.7 に示す．図 4.6 から，最高気温とヤサイの売り上げが反比例していることがわかる. 平均売り上げ個数が著しく高い日があるが，多くが火曜日である．図 4.7 から，降水量と平均売り上げ個数が反比例していることがわかる. 変数選択法を用いた回帰方程式より求まるヤサイの予測平均売り上げ個数と, 実測平均売り上げ個数を図 4.8 に示す．図中に $y=x$ の直線を挿入してある．データが，この直線上付近に散布しているとき，予測値の精度がいいことをあらわす．以下，すべての実測値と予測値の比較の図に，$y=x$ の直線を挿入する．実測値に比べ予測値が大きい日は, 火曜日であるがヤサイの安売りをしていない日である. 曜日ごとのヤサイの平均売り上げ個数の合計を，曜日数で割った値を表 4.6 に示す．表 4.6 からも，火曜日，土曜日・祝日に売り上げが増加することがわかる．

表 4.4　ヤサイの重回帰分析結果（変数選択法を用いなかった場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	1.94720		16.72052	2.29E-35
変数 1	-0.02353	-0.31527	-5.87276	2.91E-08
変数 2	-0.00212	-0.08930	-1.55786	1.21E-01
変数 3	0.00317	0.04788	0.80462	4.22E-01
変数 4	-0.00026	-0.00255	-0.04536	9.63E-01
変数 5	-0.01135	-0.01319	-0.25731	7.97E-01
変数 6	0.63841	0.75627	14.72063	2.29E-30
変数 7	0.05677	0.07303	1.40205	1.63E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.64405	0.17877	39.51416	2.27E-30

表 4.5　ヤサイの重回帰分析結果（変数選択法を用いた場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	1.93727		17.26009	5.58E-37
変数 6	0.64453	0.76352	15.44123	2.01E-32
変数 1	-0.02256	-0.30230	-6.00605	1.46E-08
変数 2	-0.00251	-0.10565	-2.08517	3.88E-02
変数 7	0.06318	0.08127	1.62799	1.05E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.64961	0.17736	70.06028	6.48E-33

表 4.6　曜日ごとの平均売り上げ個数

日曜	月曜	火曜	水曜	木曜	金曜	土曜・祝日
1.27490	1.28907	1.91903	1.21670	1.27597	1.28614	1.33972

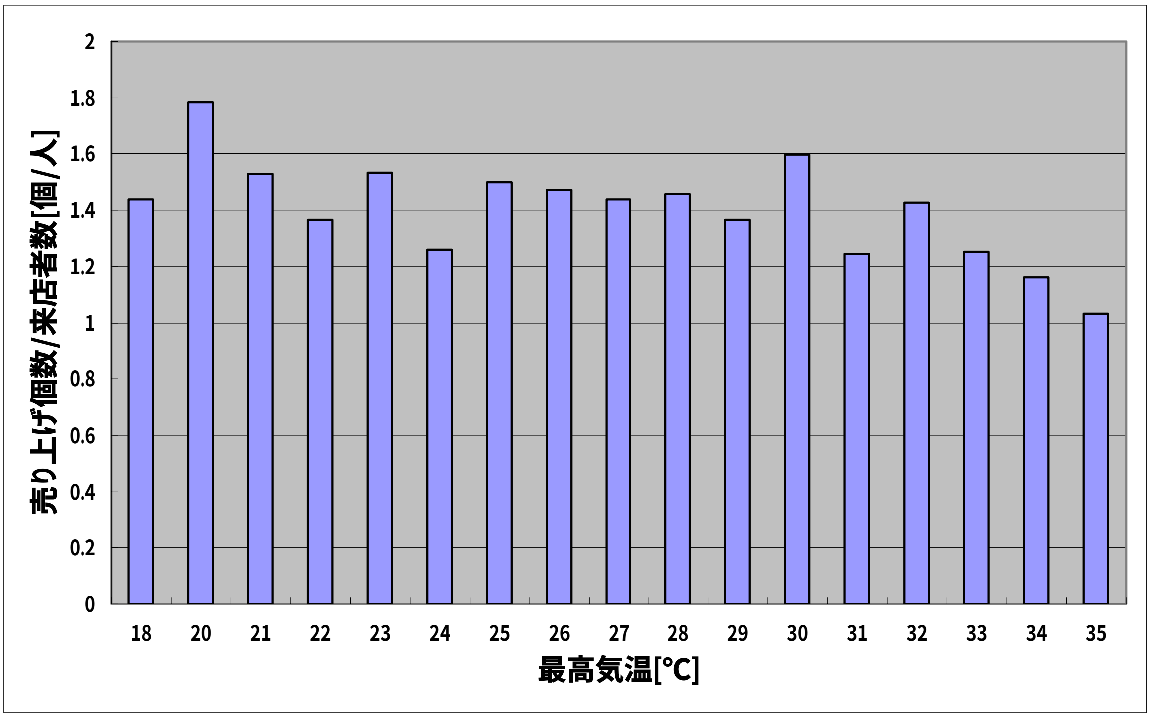

図 4.6　ヤサイの平均売り上げ個数と最高気温

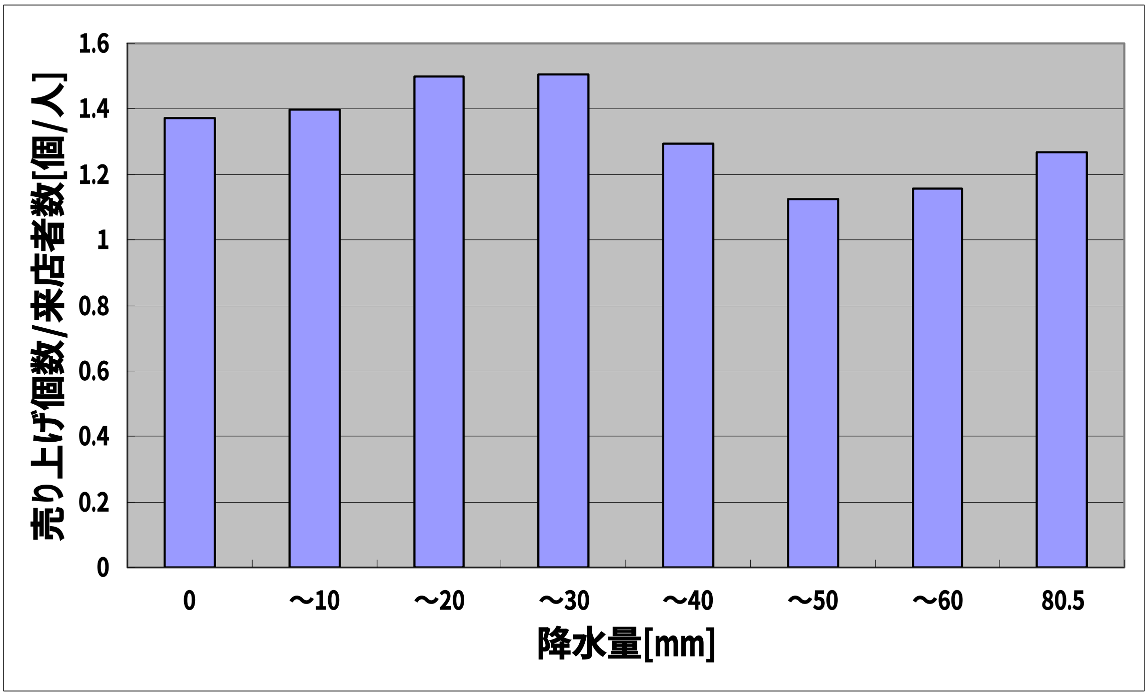

図 4.7　ヤサイの平均売り上げ個数と降水量

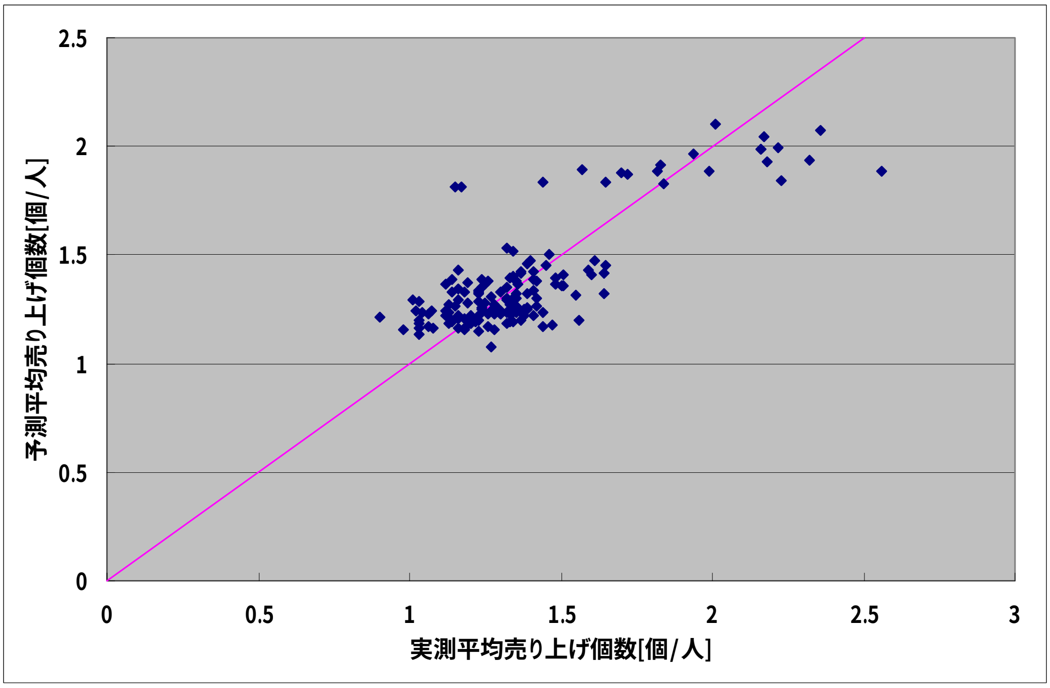

図 4.8　ヤサイの平均売り上げ個数の予測値と実測値

・レイシヨク

レイシヨクの重回帰分析の結果を表 4.7，表 4.8 に示す．標準化回帰係数は変数 5，変数 6 の係数が高い値となっている．変数 5，変数 6 の回帰係数が，有意水準 0.05 で有意である．回帰方程式の有意確率は 0 に近く，回帰方程式は有意である．変数選択法を用いた場合，変数 5，変数 6，変数 7 が選ばれ，この中で変数 5，変数 6 の標準化回帰係数が高い値となっている．変数 5，変数 6，変数 7 はそれぞれ日曜日，火曜日，土曜日・祝日を示すダミー変数である．よって重回帰分析の結果から，レイシヨクの売り上げは日曜日と火曜日に著しく増加するといえる．変数選択法を用いた回帰方程式より求まるレイシヨクの予測平均売り上げ個数と，実測平均売り上げ個数を図 4.9 に示す．実測平均売り上げ個数が著しく高くなっている日は，多くが日曜日と火曜日である．影響の強い 2 つの変数がともにダミー変数であるために，予測値が 0.18 付近，0.6，0.7 の大きく 3 つにわかれてしまった（ともに 1 であることはないので，2 つのダミー変数の組み合わせは 3 通りになる）．そのため，自由度調整済み寄与率は高いが，予測精度はよくない．曜日ごとのレイシヨクの平均売り上げ個数の合計を，曜日数で割った値を表 4.9 に示す．表 4.9 からも，日曜日と火曜日に売り上げが著しく増加することがわかる．

表 4.7　レイシヨクの重回帰分析結果（変数選択法を用いなかった場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	0.18260		2.41752	1.68E-02
変数 1	-0.00083	-0.01319	-0.31905	7.50E-01
変数 2	0.00022	0.01123	0.25446	7.99E-01
変数 3	-0.00238	-0.04262	-0.92972	3.54E-01
変数 4	0.00197	0.02290	0.52918	5.97E-01
変数 5	0.55379	0.76416	19.35568	1.17E-41
変数 6	0.45736	0.64343	16.25912	3.14E-34
変数 7	0.04270	0.06523	1.62578	1.06E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.78879	0.11595	80.49826	2.9E-46

表 4.8　レイシヨクの重回帰分析結果（変数選択法を用いた場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	0.14608		11.35993	7.96E-22
変数 5	0.55254	0.76243	19.59286	1.05E-42
変数 6	0.45410	0.63885	16.40009	6.23E-35
変数 7	0.04169	0.06370	1.62881	1.05E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.79219	0.11501	190.33390	3.04E-50

表 4.9　曜日ごとの平均売り上げ個数

日曜	月曜	火曜	水曜	木曜	金曜	土曜・祝日
0.69862	0.13733	0.60018	0.13091	0.15879	0.15696	0.18777

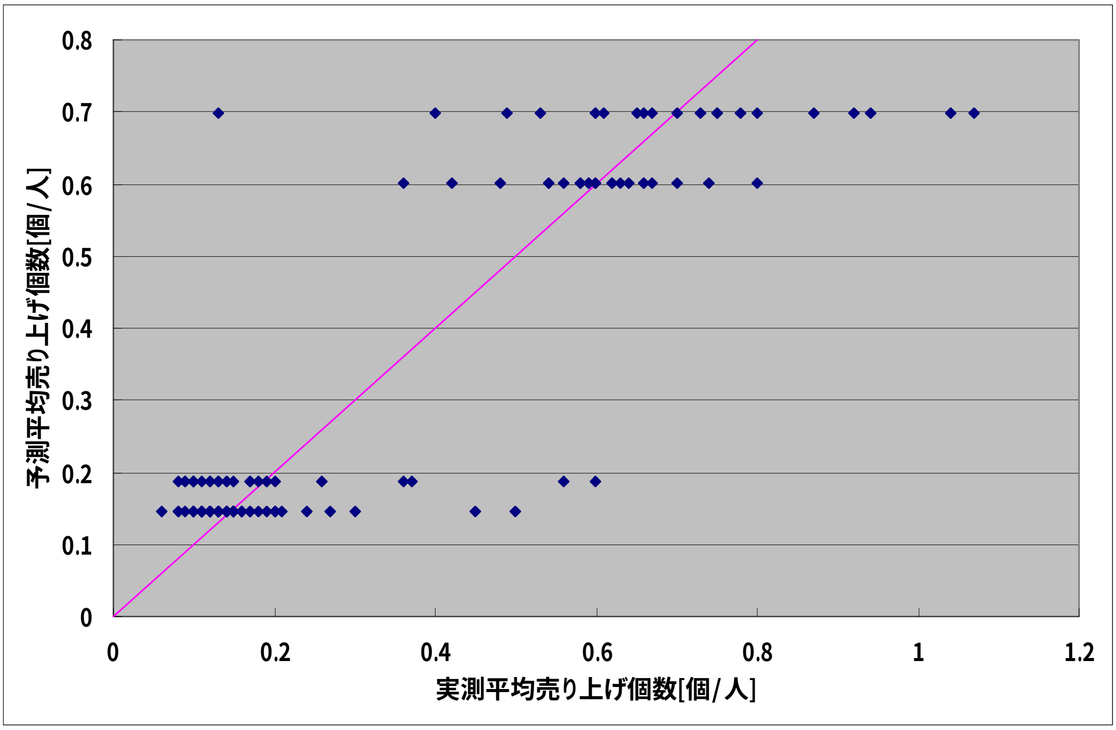

図 4.9　レイシヨクの平均売り上げ個数の実測値と予測値

求めた回帰方程式，実データともに，ヤサイの売り上げは火曜日に著しく増加し，レイシヨクの売り上げは日曜日と火曜日に著しく増加することを示している．毎週火曜日は，ヤサイとレイシヨクの一部商品が安くなっており，毎週日曜日にもレイシヨクの一部商品が安くなっている．また，火曜日，日曜日は他の曜日に比べ来店者数が多いことから，対象店舗におけるヤサイとレイシヨクの需要が高いことが予測される．レイシヨクの結果から，自由度調整済み寄与率が高くても，予測精度が良くないことがあることがわかった．自由度調整済み寄与率は，求めた回帰方程式上に実測値がない場合，値が低くなる．しかし，自由度調整済み寄与率が低くても，影響を及ぼしている要因を見つけ出すことは可能であると考えられる．そこで，自由度調整済み寄与率が 0.3 以上の食品である，ネリモノニツパイ，フクロガシの結果もみてみる．なお，アイス，タマゴについては次節で述べる．

・ネリモノニツパイ

ネリモノニツパイの，変数選択法を用いた重回帰分析の結果を表 4.10 に示す．自由度調整済み寄与率は 0.30279 であり，高い値ではない．変数 1，変数 6，変数 7 の回帰係数が，有意水準 0.05 で有意である. また回帰方程式の有意確率も 0 に近い値であり, 有意水準 0.05 で有意である．変数 1 は最高気温，変数 5 は日曜日，変数 6 は火曜日，変数 7 は土曜日・祝日を示す変数である．標準化回帰係数の値は変数 1 の係数が負の方向に高い．以上からネリモノニツパイの売り上げは最高気温の高い日に減少し，日曜日，火曜日，土曜日・祝日に少し増加するといえる．ネリモノニツパイの平均売り上げ個数と最高気温の比較を図 4.10 に示す．図 4.10 からも，ネリモノニツパイの売り上げが最高気温に反比例していることがわかる．回帰方程式より求まるネリモノニツパイの予測平均売り上げ個数と，実測平均売り上げ個数を図 4.11 に示す. 実測値が 0.4 付近と 0.6 付近で予測値が大きく実測値と異なっていることが，寄与率を下げた原因であると考えられる．曜日ごとのネリモノニツパイの平均売り上げ個数の合計を, 曜日数で割った値を表 4.11 に示す. 表 4.11 からも日曜日, 火曜日, 土曜日・祝日に売り上げが増加することがわかる.

表 4.10　変数選択法を用いたネリモノニツパイの重回帰分析結果

	回帰係数	標準化係数	t 値	有意確率
定数項	0.76888		19.45658	2.91E-42
変数 1	-0.00899	-0.46340	-6.75080	3.29E-10
変数 7	0.04633	0.22938	3.19123	1.73E-03
変数 6	0.04716	0.21499	3.01236	3.06E-03
変数 5	0.04476	0.20014	2.80648	5.69E-03
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.30279	0.06501	17.17796	1.49E-11

表 4.11　曜日ごとの平均売り上げ個数

日曜	月曜	火曜	水曜	木曜	金曜	土曜・祝日
0.55497	0.53385	0.55625	0.48854	0.50211	0.52892	0.56104

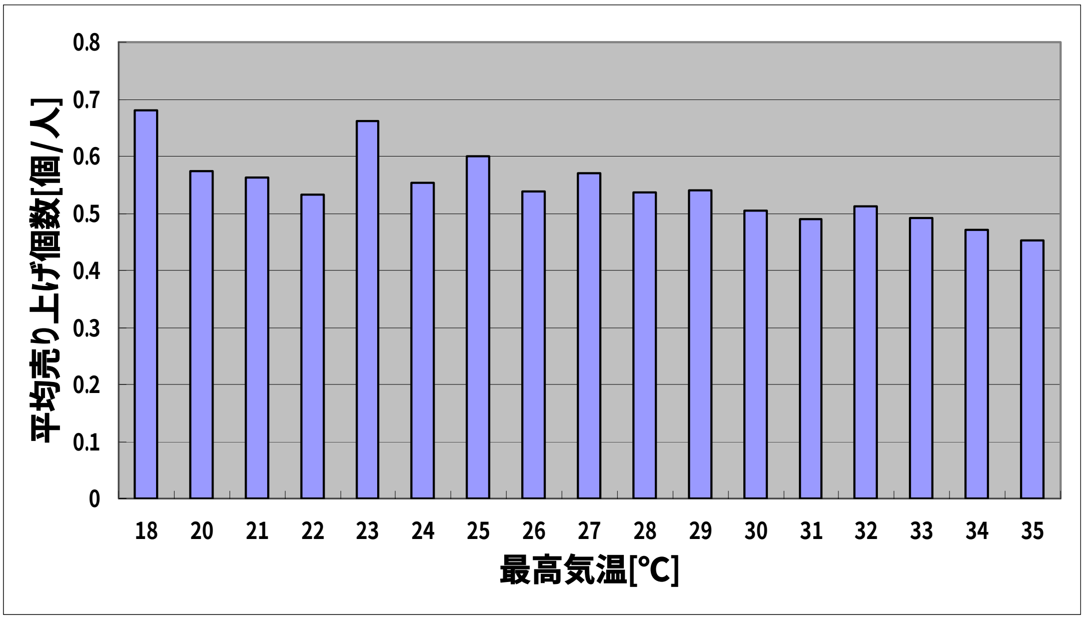

図 4.10　ネリモノニツパイの平均売り上げ個数と最高気温

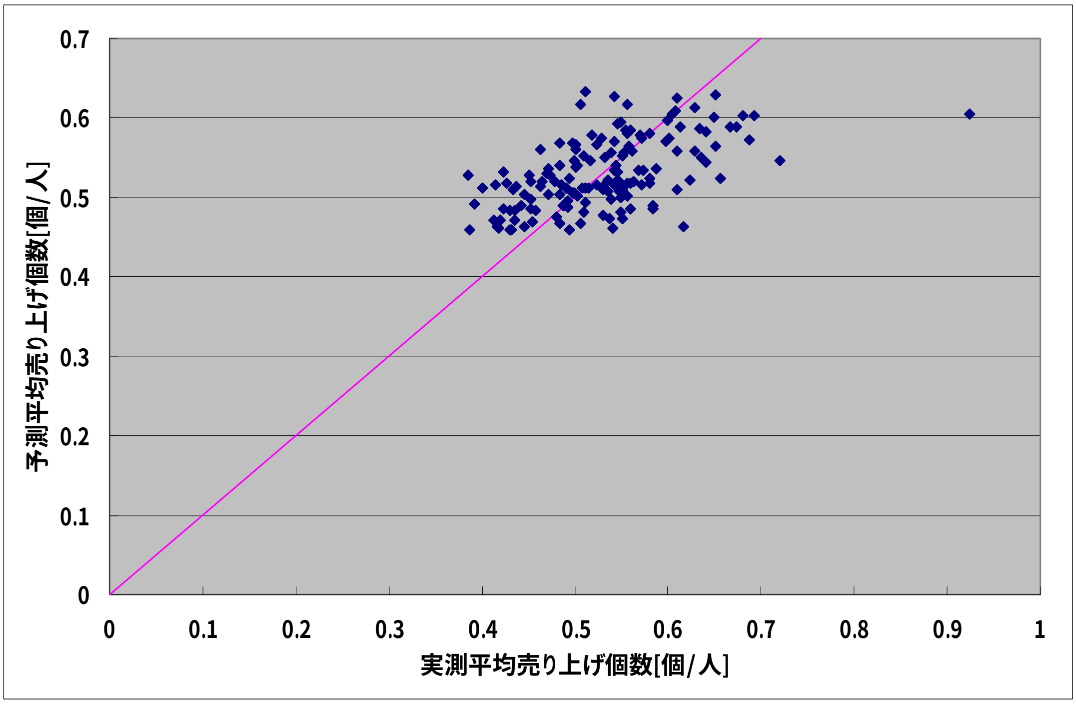

図 4.11　ネリモノニツパイの平均売り上げ個数の実測値と予測値

・フクロガシ

フクロガシの，変数選択法を用いた重回帰分析の結果を表 4.12 に示す．自由度調整済み寄与率は 0.39066 であり，高い値ではない．選択されたすべての変数が有意水準 0.05 で有意である．回帰方程式も，有意水準 0.05 で有意である．標準化回帰係数の値は，変数 1 の係数が負の方向に高く，変数 7 と変数 4 の係数が正の方向に高くなっている．変数 1 は最高気温，変数 4 は当日の最高気温と前日の最高気温の差，変数 7 は土曜日・祝日を示す変数である．以上から，フクロガシの売り上げは最高気温の高い日に減少し，前日よりも暑い日，土曜日・祝日に少し増加するといえる．フクロガシの平均売り上げ個数と最高気温との比較を図 4.12 に，前日の最高気温と当日の最高気温の差との比較を図 4.13 に示す．図 4.12 からも，フクロガシの売り上げと最高気温が反比例していることがわかる．図 4.13 からは，フクロガシの売り上げと当日の最高気温と前日の最高気温の差が比例しているとはいえない．回帰方程式より求まるフクロガシの予測平均売り上げ個数と，実測平均売り上げ個数を図 4.14 に示す．実測値が 0.4 と 0.9 付近で，予測精度が悪くなっていることが，寄与率を下げている原因と考えられる．曜日ごとのフクロガシの平均売り上げ個数の合計を，曜日数で割った値を表 4.13 に示す．表 4.13 からも，売り上げが土曜日・祝日に増加することがわかる．

表 4.12　変数選択法を用いたフクロガシの重回帰分析結果

	回帰係数	標準化係数	t 値	有意確率
定数項	1.18212		17.79641	2.14E-38
変数 1	-0.01738	-0.51472	-7.66462	2.3E-12
変数 7	0.13265	0.37713	5.86637	2.86E-08
変数 4	0.01100	0.23805	3.54048	5.36E-04
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.39066	0.10584	32.84279	2.77E-16

表 4.13　曜日ごとの平均売り上げ個数

日曜	月曜	火曜	水曜	木曜	金曜	土曜・祝日
0.65729	0.68155	0.68362	0.66399	0.67707	0.71055	0.81776

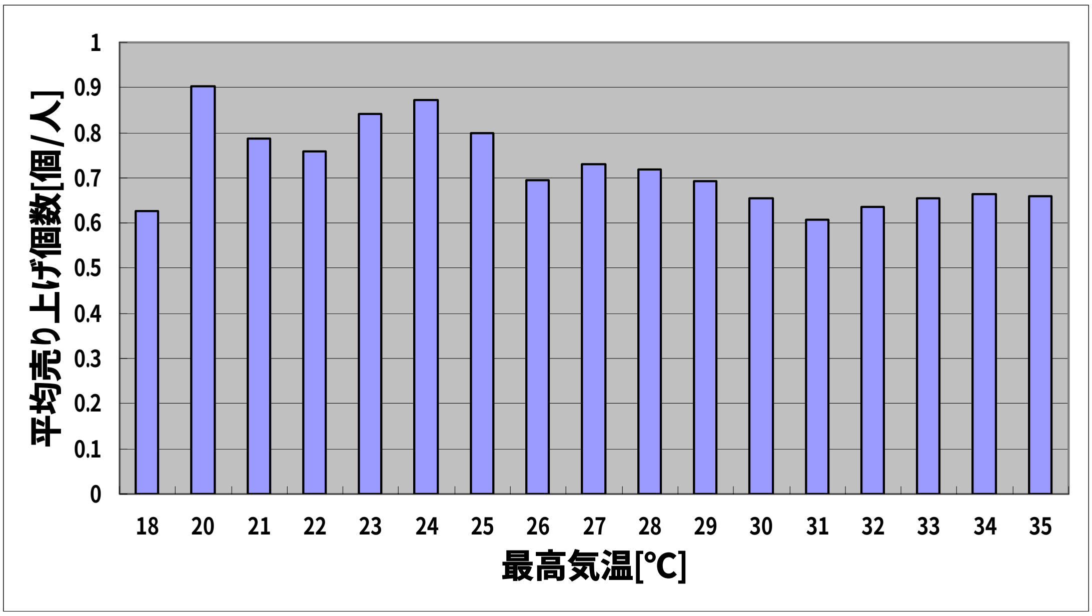

図 4.12　フクロガシの平均売り上げ個数と最高気温

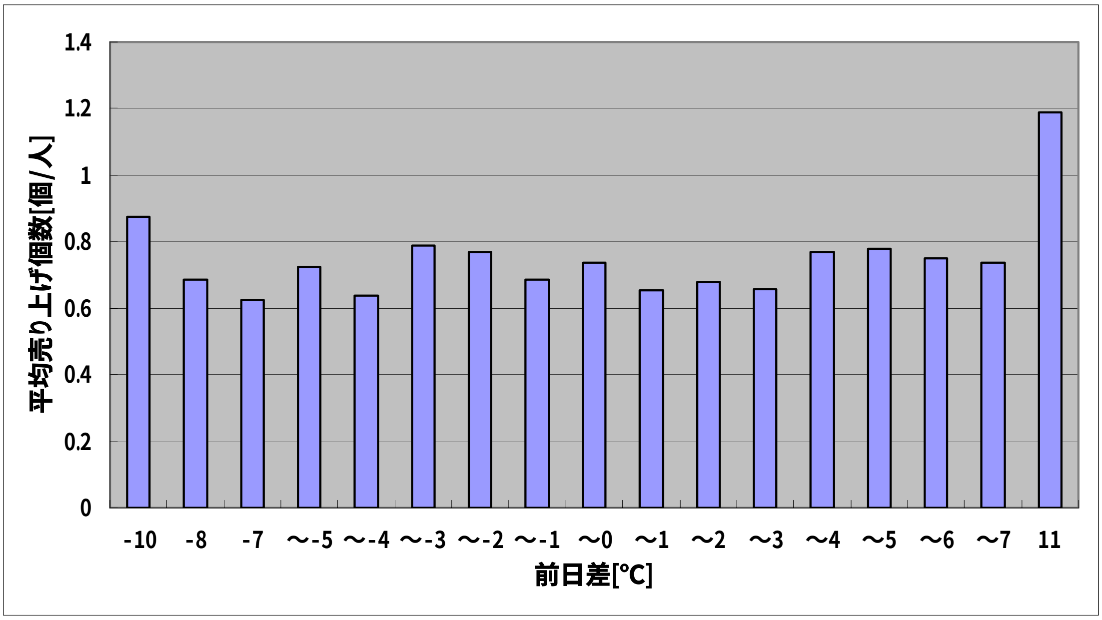

図 4.13　フクロガシの平均売り上げ個数と最高気温の前日差

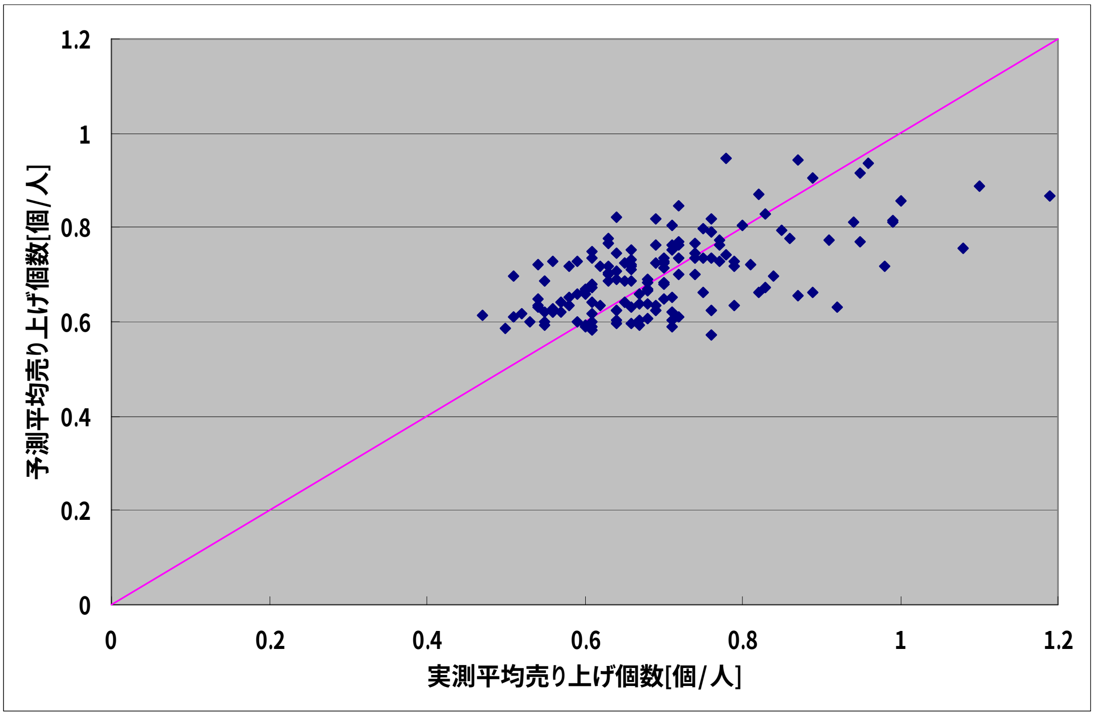

図 4.14　フクロガシの平均売り上げ個数の実測値と予測値

4.3　特売日の定義

4.2 節のヤサイ，レイシヨクの結果より，安売りの日が強く食品の売り上げに影響していることがわかる．特定の商品が安く売られており，そのラインの売り上げが著しく増加している日がある場合，その日を特売日とする．特売日を定義できるラインは，アイス，タマゴの 2 種類である．以下，アイスとタマゴの重回帰分析の結果を示す．

・アイス

アイスの，特売日を考慮しない重回帰分析の結果を表 4.14，表 4.15 に示す．表 4.14 の標準化回帰係数の値を見ると，変数 7 の係数が一番高く，次に変数 1 の値が高くなっている．これは表 4.15 においても同様である．選択される変数は変数 1，変数 5，変数 7 であり，それぞれ最高気温，日曜日，土曜日・祝日を示す変数である．変数 1 と変数 7 の回帰係数が有意水準 0.05 で有意である．また回帰方程式の有意確率は 0 に近く，回帰方程式は有意水準 0.05 で有意である．しかし自由度調整済み寄与率は変数選択法を用いない場合は 0.2947，変数選択法を用いた場合は 0.3056 となっており高い数値ではない．

表 4.14　アイスの重回帰分析結果（変数選択法を用いなかった場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	-0.08349		-2.09235	3.81E-02
変数 1	0.00481	0.26525	3.51027	6.00E-04
変数 2	-0.00002	-0.00439	-0.05439	9.56E-01
変数 3	0.00131	0.08127	0.97014	3.33E-01
変数 4	-0.00169	-0.06793	-0.85876	3.91E-01
変数 5	0.02380	0.11360	1.57464	1.17E-01
変数 6	-0.01074	-0.05228	-0.72299	4.70E-01
変数 7	0.09537	0.50404	6.87409	1.82E-10
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式式	0.29472	0.06125	9.89520	5.04E-10

表 4.15　アイスの重回帰分析結果（変数選択法を用いた場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	-0.08437		-2.29730	2.30E-02
変数 7	0.09919	0.52419	7.51709	5.21E-12
変数 1	0.00503	0.27694	4.04393	8.48E-05
変数 5	0.02616	0.12487	1.79554	7.46E-02
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.30566	0.06078	22.86499	3.42E-12

次に，アイスの特売日を新たに変数 8（特売日を 1，それ以外の日を 0 とするダミー変数）として重回帰分析を行った結果を，表 4.16，表 4.17 に示す．アイスの一部商品が安売りされており，平均売り上げ個数が 0.15 以上の日を特売日とした．表 4.14 の結果と表 4.16 の結果を比較すると，表 4.16 の自由度調整済み寄与率が非常に高い値となっていることがわかる．また，新たに設定した独立変数である変数 8 の標準化回帰係数の値が高い．変数 7 の標準化回帰係数が大幅に減っているのは，アイスの特売日が土曜日に多いことが原因である．また変数 3 の標準化回帰係数が高くなり，変数 5 の係数が低くなっている．表 4.15 の結果と表 4.17 の結果を比較すると，表 4.17 では新たに変数 3，変数 6，変数 8 が選ばれ，変数 5，変数 7 が除外されている．変数 1 は最高気温，変数 3 は日照時間，変数 6 は火曜日，変数 8 は特売日を示す変数である．以上からアイスの売り上げは，最高気温が高い日，日照時間が長い日に増加し，火曜日に少し減少するといえる．特売日を除いたアイスの平均売り上げ個数と最高気温との比較を図 4.15 に，日照時間との比較を図 4.16 に示す．図 4.15 から，アイスの売り上げと最高気温が比例していることがわかる．変数選択法を用いた回帰方程式より求まるアイスの予測平均売り上げ個数と，実測平均売り上げ個数を図 4.17 に示す．実測値の値が 0.16 付近で，予測値が大きく実測値を上回っている日があるが，多くの日が 9 月中の特売日である．曜日ごとのアイスの平均売り上げ個数の合計（特売日は除く）を，曜日数で割った値を表 4.18 に示す．表 4.18 からも，火曜日に少し売り上げが減少することがわかる．

表 4.16　特売日を考慮したアイスの重回帰分析結果（変数選択法を用いなかった場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	-0.04136		-1.65780	9.95E-02
変数 1	0.00311	0.17123	3.61617	4.15E-04
変数 2	0.00000	0.00034	0.00691	9.94E-01
変数 3	0.00216	0.13443	2.57735	1.09E-02
変数 4	-0.00056	-0.02272	-0.46145	6.45E-01
変数 5	0.00691	0.03300	0.73135	4.65E-01
変数 6	-0.00977	-0.04756	-1.05865	2.91E-01
変数 7	0.01263	0.06678	1.23645	2.18E-01
変数 8	0.18844	0.77763	15.06612	3.62E-31
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.72784	0.03805	50.81118	7.68E-38

表 4.17　特売日を考慮したアイスの重回帰分析結果（変数選択法を用いた場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	-0.03144		-1.37039	1.72E-01
変数 8	0.19728	0.81411	18.97053	4.08E-41
変数 1	0.00286	0.15766	3.45949	7.11E-04
変数 3	0.00216	0.13391	2.92931	3.94E-03
変数 6	-0.01253	-0.06095	-1.41734	1.58E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.73161	0.03778	102.54240	2.92E-41

表 4.18　曜日ごとの平均売り上げ個数

日曜	月曜	火曜	水曜	木曜	金曜	土曜・祝日
0.07647	0.05863	0.05416	0.05870	0.05951	0.07078	0.06228

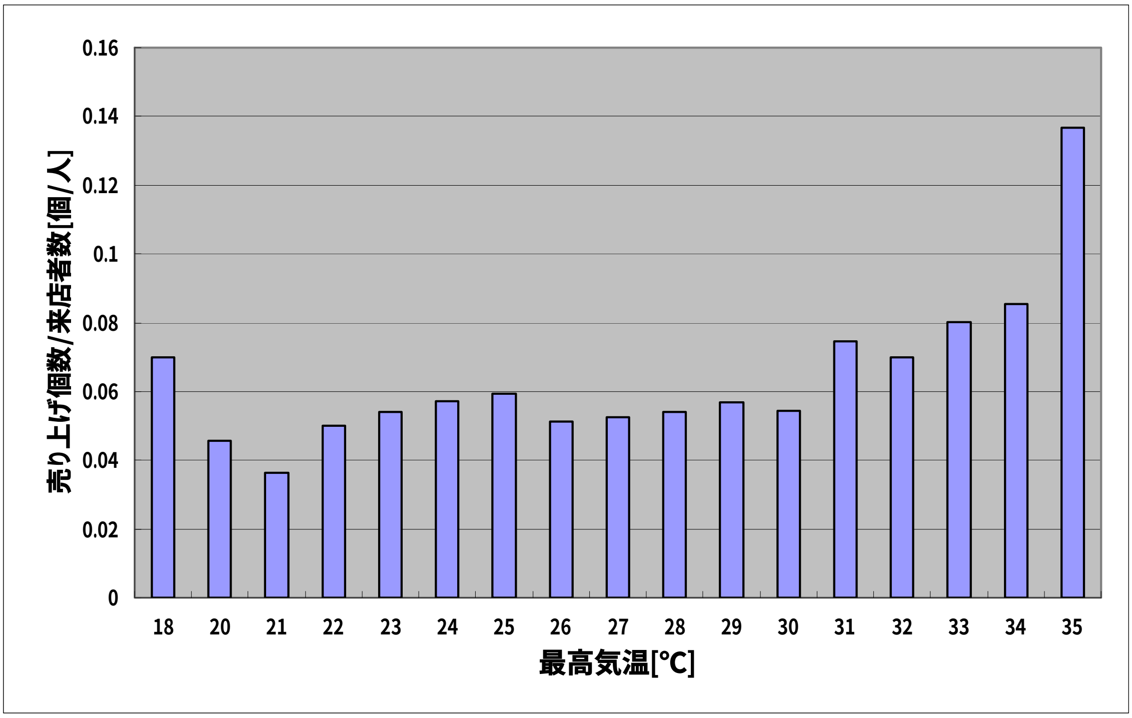

図 4.15　アイスの平均売り上げ個数と最高気温

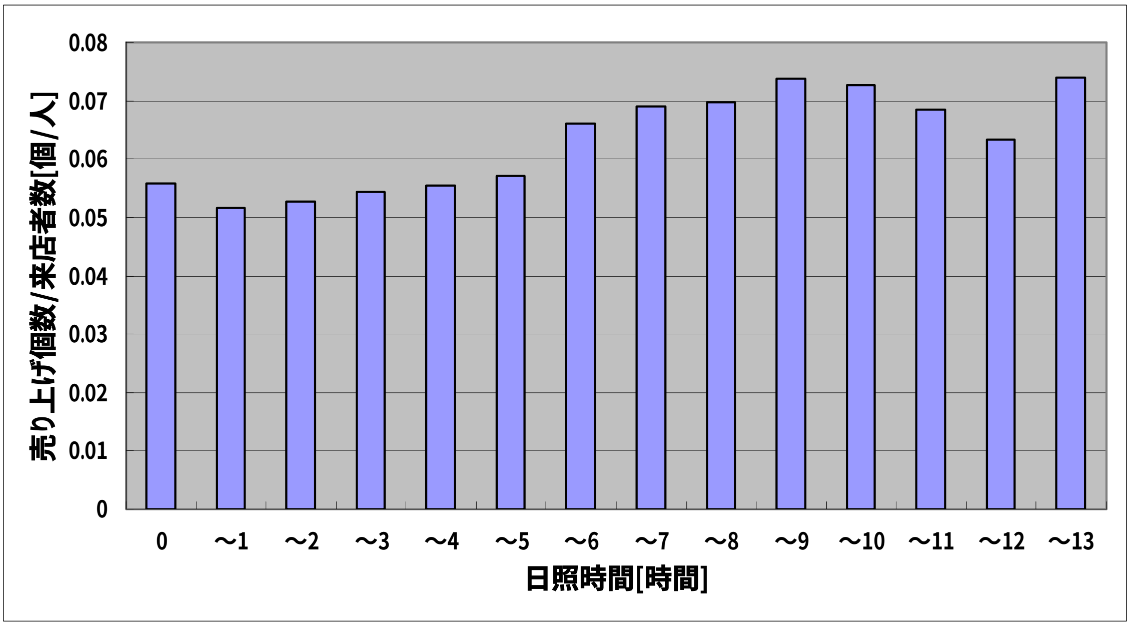

図 4.16　アイスの平均売り上げ個数と日照時間

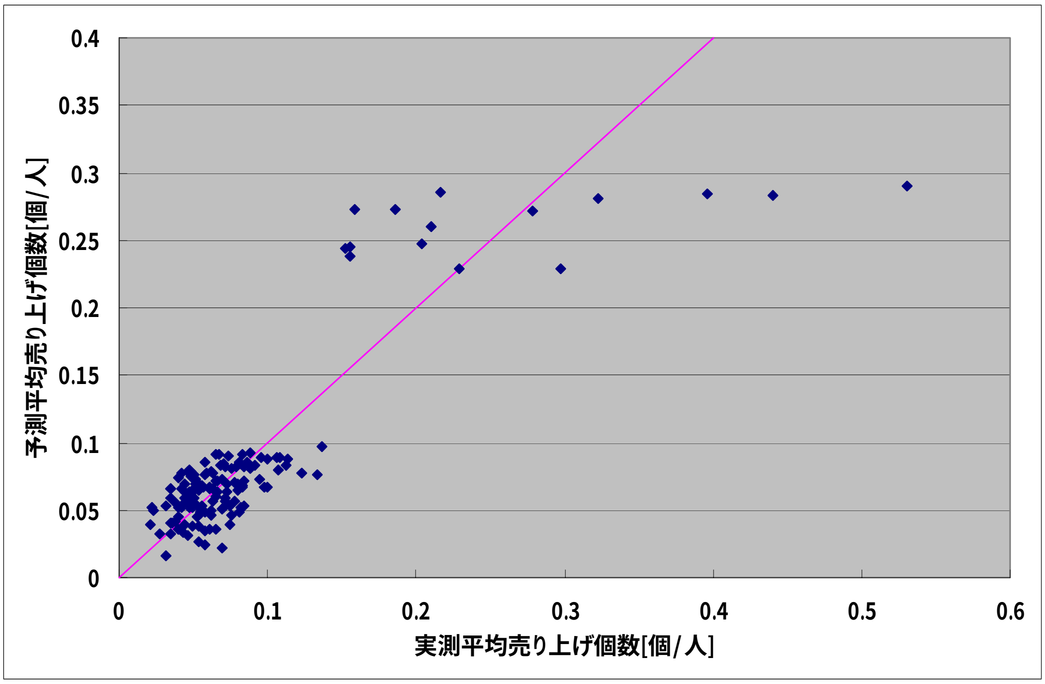

図 4.17　アイスの平均売り上げ個数の実測値と予測値

・タマゴ

タマゴの，特売日を考慮しない重回帰分析の結果を表 4.19，表 4.20 に示す．表 4.19 をみると変数 5 の標準化回帰係数が高い．変数選択法を用いると，変数 5 だけが選ばれる．変数 5 は日曜日を示すダミー変数である．回帰方程式の有意確率は 0 に近く，有意水準 0.05 で有意である．以上からタマゴの売り上げは，日曜日に著しく増加すると考えられるが，自由度調整済み寄与率が 0.42661 であり，高い値ではない．また，ダミー変数である変数 5 のみ選択されるので，予測値が 2 つの値しかとらない．そのため予測精度は良くない．

表 4.19　タマゴの重回帰分析結果（変数選択法を用いなかった場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	0.08502		3.56299	5.00E-04
変数 1	-0.00060	-0.05013	-0.73568	4.63E-01
変数 2	-0.00003	-0.00869	-0.11947	9.05E-01
変数 3	-0.00119	-0.11168	-1.47847	1.41E-01
変数 4	0.00211	0.12837	1.79991	7.39E-02
変数 5	0.09380	0.67508	10.37787	3.98E-19
変数 6	0.01267	0.09298	1.42608	1.56E-01
変数 7	0.00794	0.06333	0.95792	3.39E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.42661	0.03663	16.83728	4.56E-16

表 4.20　タマゴの重回帰分析結果（変数選択法を用いた場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	0.06384		19.73656	2.72E-43
変数 5	0.09081	0.65356	10.50511	1.24E-19
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.42327	0.03673	110.35730	1.24E-19

次に，特売日を考慮したタマゴの重回帰分析の結果を表 4.21，表 4.22 に示す．タマゴの一部商品が安売りされており，平均売り上げ個数が 0.1 以上の日を特売日とした．自由度調整済み寄与率が，表 4.21，表 4.22 ともに 0.8 以上の高い値となっている．標準化回帰係数は変数 8 の値が非常に高い．表 4.20 と表 4.22 を比較すると，変数 1 と変数 8 が新たに選ばれ，変数 5 の標準化回帰係数の値が減少している．これは，タマゴの特売日が日曜日に多いことが原因である. 変数 1 は最高気温, 変数 5 は日曜日, 変数 8 は特売日を示す変数である. 以上から，タマゴの売り上げは日曜日に少し増加し，最高気温の高い日に少し減少するといえる. 特売日を除いたタマゴの平均売り上げ個数と最高気温の比較を図 4.15 に示す. 図 4.15 からは，タマゴの売り上げと最高気温が反比例しているとはいえない．変数選択法を用いた回帰方程式より求まるアイスの予測平均売り上げ個数と，実測平均売り上げ個数を図 4.16 に示す. ダミー変数である変数 8 の影響が強すぎ, 予測値が大きく 2 つにわかれてしまった. そのため，自由度調整済み寄与率は高い値であるが，予測精度は良くない．曜日ごとのタマゴの平均売り上げ個数の合計（特売日は除く）を，曜日数で割った値を表 4.23 に示す．表 4.23 からも，日曜日に少し売り上げが増加することがわかる．

表 4.21　特売日を考慮したタマゴの重回帰分析結果（変数選択法を用いなかった場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	0.06560		6.17412	6.69E-09
変数 1	-0.00036	-0.02962	-0.97871	3.29E-01
変数 2	0.00006	0.01714	0.53037	5.96E-01
変数 3	-0.00035	-0.03284	-0.97443	3.31E-01
変数 4	0.00054	0.03283	1.02883	3.05E-01
変数 5	0.01011	0.07276	1.90415	5.89E-02
変数 6	0.00328	0.02409	0.82807	4.09E-01
変数 7	0.00500	0.03989	1.35827	1.76E-01
変数 8	0.11254	0.89675	24.06851	8.68E-52
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.88696	0.01626	147.14260	1.66E-64

表 4.22　特売日を考慮したタマゴの重回帰分析結果（変数選択法を用いた場合）

	回帰係数	標準化回帰係数	t 値	有意確率
定数項	0.06869		7.08933	5.34E-11
変数 8	0.11348	0.90423	24.72589	4.71E-54
変数 5	0.00767	0.05521	1.51009	1.33E-01
変数 1	-0.00047	-0.03915	-1.42656	1.55E-01
	自由度調整済み寄与率	推定値の標準誤差	F 値	有意確率
回帰方程式	0.88788	0.01619	394.33480	8.74E-70

表 4.23　曜日ごとの平均売り上げ個数

日曜	月曜	火曜	水曜	木曜	金曜	土曜・祝日
0.07203	0.05315	0.05556	0.05246	0.05550	0.05406	0.05715

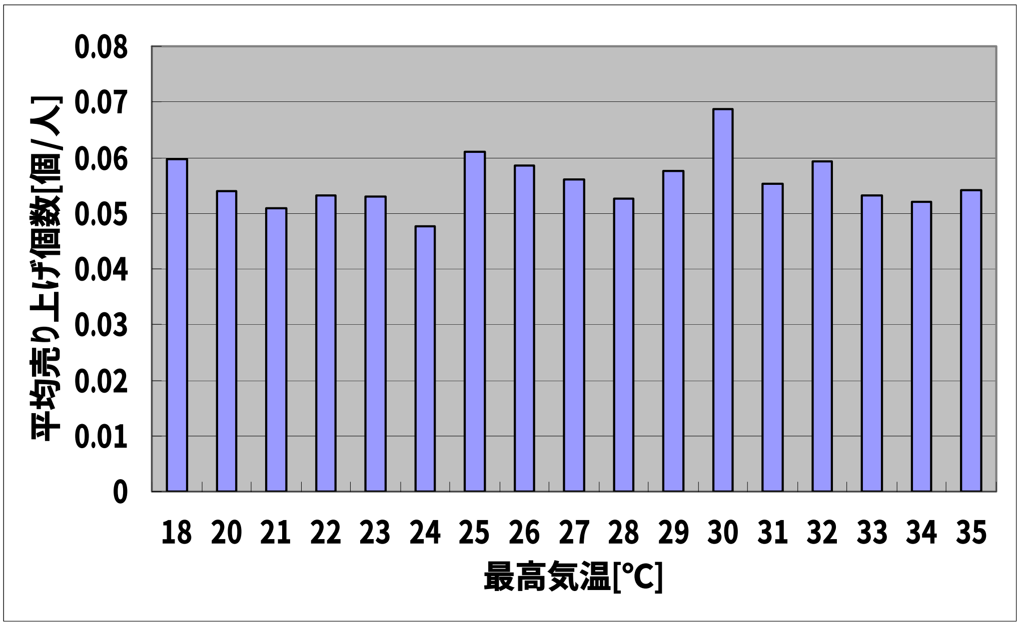

図 4.15　タマゴの平均売り上げ個数と最高気温

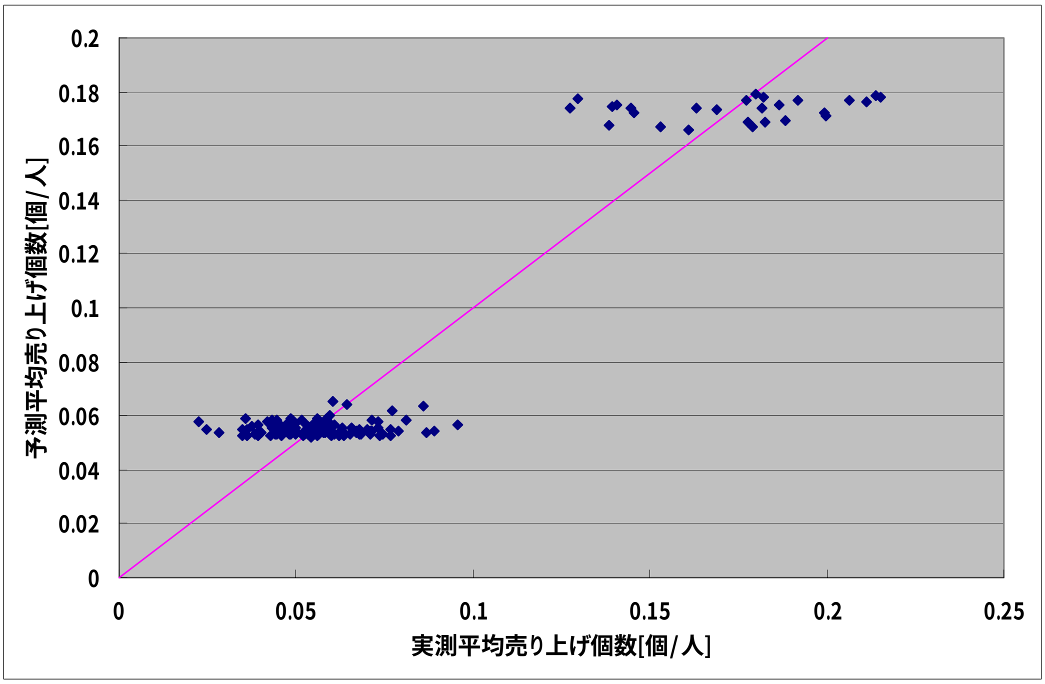

図 4.16　タマゴの平均売り上げ個数の実測値と予測値

第 5 章　おわりに

本研究では曜日，気象が食品の売り上げに影響していると考え，曜日，気象と食品の売り上げの関係を数値的に捉えるため，重回帰分析を行った．その結果，食品関連の全ライン22 種類のうち，ヤサイ，レイシヨクについては設定した独立変数により，高い自由度調整済み寄与率が得ることが出来たが，残りのラインについては出来なかった．原因として，設定した独立変数以外で，売り上げにより影響を及ぼす要素がある，またはラインに属するクラスが多いことなどが考えられる．ヤサイの重回帰分析の結果から売り上げに最高気温と降水量が影響していること，安売りが行われている曜日の影響が，気象の影響よりも大きいことがわかった．また，自由度調整済み寄与率は高い値ではなかったが，ネリモノニツパイ，フクロガシの結果からも，売り上げに最高気温が影響していることがわかった．次に，食品の売り上げに安売りが行われている日の影響が強いことがわかったので特売日を定義した．特売日を設定したアイス，タマゴの自由度調整済み寄与率は，特売日を設定しない場合の自由度調整済み寄与率に比べ，高い値となっており，それぞれの特売日を示す変数の標準化回帰係数の値も非常に高い値となっている．アイス，タマゴで，特売日以外に売り上げに影響している変数は，最高気温，日照時間である．アイスの売り上げと最高気温，日照時間が比例していることは，実データからも読み取れた．

今後の課題としては，独立変数の再考，また，対象食品の分類をより細かくし，その分類に対し特売日を設定するなどがあげられる．

謝辞

本研究を進めるにあたり，多くのご指導をいただいた中央大学理工学部情報工学科の田口東教授に深く感謝の意を表します．また，研究を進めるうえで，多くのご助言，ご協力をいただいた鳥海重喜氏をはじめ，田口研究室の皆さんに深く感謝します．

最後に，研究の方向性から有効に利用することが出来なかったが，貴重なデータを提供して下さった財団法人流通システム開発センターの銅直氏，株式会社日本ユニシス情報システムの高久氏に，深くお詫び申し上げるとともに，感謝致します．

参考文献

[1] 飯塚悦功，久米均，**回帰分析**，シリーズ入門統計的方法 2，岩波書店，東京，1999.

[2] 内田治，**すぐわかる SPSS によるアンケートの調査・集計・解析**，東京図書，東京，1997.

[3] 東京大学教養学部統計学教室編，**統計学入門**，基礎統計学Ⅰ，東京大学出版会，東京，1991.

[4] 気象庁ホームページ，**電子閲覧室**，http://www.jma.go.jp/JMA_HP/jma/

[5] 統計学学習ノート，**重回帰分析**，http://aoki2.si.gunma-u.ac.jp/lecture/