

マーケティング・データ分析のためのウェブ・ベース・システムの開発

Development of Web-based System for Marketing Data Analysis

高萩栄一郎[†] 生田目崇[†]
Eiichiro TAKAHAGI[†] Takashi NAMATAME[†]

[†] 専修大学 商学部
[†] School of Commerce, Senshu University

要旨:

これまで開発された、多くの学習用データ分析ツールは、スタンドアロン式プログラムもしくは、多くの人
が利用できるウェブ・ベースであっても、オリジナルのプログラムによって作成されたものである。また、R
言語に代表されるような汎用で高性能なデータ分析ソフトウェアもあるものの、数行とはいえプログラミング
の必要があるため、いまだに技術獲得において障壁は高い。本研究では、データ分析のためのフリーソフ
トウェアである R 言語をベースにしたウェブブラウザで利用できるマーケティング・データ分析のための学
習用システムを開発した。本システムは、複雑なプログラミングを必要としない高性能な分析ツールとして
利用できる。

Abstract:

Many existed education-oriented data analysis tools are stand-alone type system or original program even if web-based
system which can be used by many people. Even there are free charge, wide use and high performance data analysis
software like R language, it is difficult to use them efficiently. In this study, we develop a web-based educational
system for marketing data analysis based on R language. Also this system can be used as high performance one without
complex programming.

1. はじめに

本研究では、経済学・経営学・商学を専攻する学生
を念頭に、マーケティング分野のデータ分析技術習得
のためのソフトウェアの開発を行った。多くの場合、
学生は数式やプログラミングには慣れてはいないが、
分析に必要なデータの用意や、結果の解釈などにつ
いては学びたい意識が高い。したがって、実際の効率
の良い実際の分析システムを提供することで、こう
した学生のニーズに応えることができる。

これまでの多くのシステムでは、スプレッド・シ
ート上で実行できるプログラムや、ウェブ・シ
ステムであっても、独自の cgi などを利用したシ
ステムが多かった (たとえば[5, 9])。こうしたシ
ステムでは、広く知られている定型的なモデル
を利用することはできても、拡張しようとする
と、一からプログラムを作成しなければならない
など、開発工数の手間がかかる。また、こう
したソフトウェアのグラフ機能はあまり使い
勝手がよくないため、結果を効果的に見せよ
うとすると困難が伴うこともある。本研究で
対象としているマーケティング分野では、一
見、通常集計分析や多変量解析のモデルに
見えても、実際には独自の工夫やアルゴ
リズムが必要な場合も多い。こうした分析
を通常ソフトウェアで行う場合、実装が
相当困難となる。

近年では、R 言語に代表される、高性能で
無料のソフトウェアも広く用いられている。
こうしたソフトウェアの多くはオープン・
ソース・ソフトウェア (OSS) であり、
世界中の開発者が数多くの分析プログラム
を作成・公開しており、最新の分析手法
であっても手軽

に入手し利用することができる。ただし、
開発の手間が省ける一方で、プログラ
ミングは必要であり、初学者が分析に
使う場合は障壁は高いと言わざるを得
ない。

したがって、社会科学系の学生が
こうしたシステムを利用するとして
も教育効果はいずれも一長一短
であると言わざるを得ない。そ
こで本研究では、統計ソフト
の高機能さと、ウェブブラウザ
上での入出力の容易さを兼ね
備えたシステムの開発を行
った。

2. システムの概要

本研究では、データ整形については Microsoft Excel、
分析インターフェイスとしてウェブブラウザを利用し、
分析エンジンとして R を使った計算やグラフ作成を行
うシステムを開発した。データの入出力、分析ステ
ップの概要は以下の通りである。

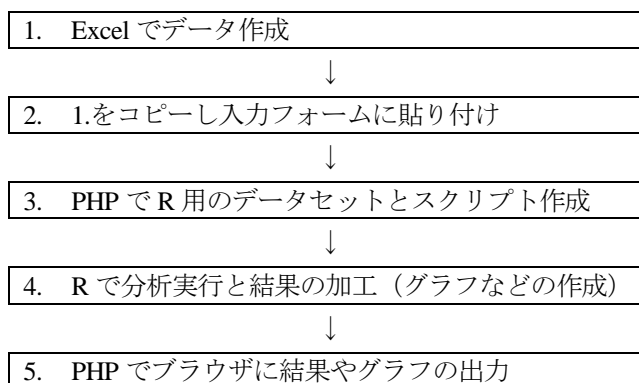


図1 システムの概要

本システムの動作環境としては、Apacheなどのウェブ・サーバが起動するPCに統計処理言語のRをインストールしておく必要がある¹。

分析用データについては、Microsoft Excel に代表されるスプレッドシート・ソフトウェアのワークシートで所定の形式のデータを作成する。

次に、ワークシートのデータをのうち必要な部分をクリップボードにコピーし、ブラウザの入力フォームに貼り付ける。学生向けの分析では、それほど大規模なデータを使うことはない想定されるため、クリップボードへのコピーで十分であると判断した。さらに、実際のデータを概観する場合はスプレッド・シートを利用することの方が、理解をしやすく、データの準備や後のレポート作成などを考えると、スプレッド・シートを利用の方が利便性は高い。

入力フォームでは、データのみテキストで入力しただけの状態であるため、これを、R の分析用にデータを格納する必要がある。さらに、実際に分析するためにはR のスクリプトを生成しなければならない。これらを行うのが第 3 ステップである。このステップはPHP によってデータを格納しながら、分析オプションの設定などを含めて、R のスクリプトを生成する。

そして、それらをR のソースファイルとして実行し、結果を保存する。なお、グラフを作成するためのデータの取り出しやグラフ作成を行い、結果のファイルから必要なデータを取り出し、グラフとともに出力画面に返す。

このようにすることで、分析実行者がデータ生成やR のスクリプトについて大きな負担なく、ウェブブラウザを利用して、R による分析を行うことができる。

また、自分でR を習得したい場合の情報として、実行しているR のスクリプトと入出力データを参照できるようにしている。

3. システムの特徴

前節で、本システムの特徴についてその概要を論じた。今回作成したシステムの特徴として以下のようなものにまとめることができる。

網羅性：R 言語は、統計処理言語である S 言語のフリー版クローンとして 1996 年にニュージーランドの R. Ihaka と R. Gentleman により開発されて以来、OSS であるという特徴を最大限に生かし、数多くのディストリビュータが多くの分析手法のパッケージを開発し、これらを自由に利用することができる。これらには統計関数や多変量解析のみでなく、オペレーションズ・リサーチ手法、機械学習やデータ・マイニング手法、テキスト処理やテキスト・マイニング手法も含まれる。

拡張性：網羅性とも関係するが、OSS であるため、自らが新たな分析手法を作ることができるプラットフォームとなっている。R は統計処理言語であり、行列データを扱いやすく、さらに、最適化もしくは求解手法についても各種そろっている。

精緻性：多変量解析などの伝統的な分析については、単にパラメータを推定するだけでなく、各種のモデル選択やモデル評価のための情報を得ることができる。たとえば回帰分析におけるパラメータの有意検定や、主成分分析の寄与率などである。そのほかにも従来計算が困難であった、一般化線形モデルにおけるパラメータの分散の近似値などもパッケージに含まれ、複雑なモデルにおいても適切なモデルの評価が可能である。

操作容易性：上記のようにデータ分析手法を網羅している言語であるが、実際の分析には、適切な形式のデータとR のスクリプトを適切に書いて実行する必要がある。また結果はすべて数値として得られるため、適切に集計・グラフ化するためにはさらにスクリプトを書かなければならない。こうした事項の習得はデータ分析初学者にとっては容易であるとは言えない。したがって、システム構築者がこうした入出力の制御については事前に仕込むことで、利用者は分析を容易にすることができる。

4. システムの詳細

本システムは試作段階であり、現状では十分な分析手法を含んでいるわけではない。マーケティング分野のデータの分析を念頭と置いているため、マーケティング分野で利用されるような集計や、分析モデルを中心に実装している。現状で実装している手法は表 1 および図 2 の通りである。

なお、すべての分析手法について、共通して「R のソースファイル」、「R のデータファイル」、「R の分析結果のファイル」についてそれぞれテキスト・ファイルの形式で得られる。

表 1 実装されている手法

分析目的	分析手法
予測, 因果関係	回帰分析
反応分析	ロジスティック 回帰分析
選択モデル	多項ロジット・モデル
知覚マップ	因子分析 (主成分分析を含む)
販売 3 評価	バレート分析
顧客評価	RFM 分析
顧客評価の応用	RFM 分析 (ファジィ)

以下で各分析手法について概説し、実際の分析手順について実際の分析手順について概観する。なお、各分析手法の入出力項目についても記述する。実際の利用イメージは付録を参照されたい。

¹ ただし、USB で起動できるシステムも提供しているため、実際にはこれらのソフトを別途インストールする必要性は必ずしもない。

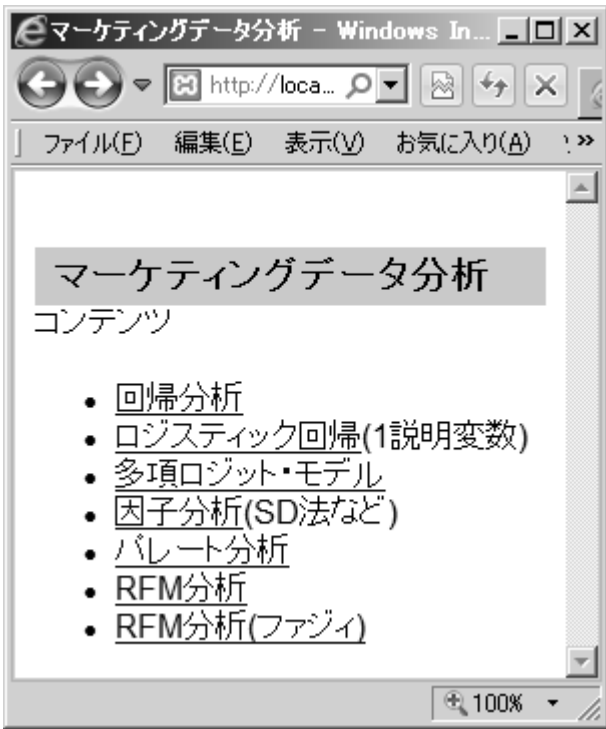


図2 メインメニュー

4.1. 回帰分析

回帰分析は、一つもしくは複数の説明変数が、一つの目的変数へどのように影響を与えるかを分析する統計的手法であり、一次関数の形式でパラメータを推定する [3]。説明変数を x_1, x_2, \dots, x_p とし、目的変数を y とするとき

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + \varepsilon$$

と表される線形関係において、観測される y と予測値の残差 ε の平方和を最小にするようにパラメータ $b_0, b_1, b_2, \dots, b_p$ を求める手法である。

入力データとしては、説明変数は各列に並べるような行列形式であり、目的変数は列ベクトルとしてそれぞれ別々に入力する。また、変数名を指定することができる。説明変数についてはモデルに含めるかどうかについてチェックボックスで指定することができる。

また、学習目的であるが、最適なモデルを試行錯誤で見つけることができるようにするために、AIC (Akaike Information Criteria) の値を出力するようにしている。

4.2. ロジスティック回帰分析

ロジスティック回帰分析は、ある変数が反応の有無に影響を与えるような場合に、反応の確率を S 字カーブで求める。説明変数 x について、反応確率 p を、

$$p = \frac{a}{1 + b \times \exp\{-cx\}}$$

として与える²。なお、本システムでは説明変数は 1 つに限っている。

表2 回帰分析の入出力項目

入力項目	
説明変数数	p (最大 10)
データ数	n
説明変数	$n \times p$ 行列
被説明変数	n 次元ベクトル
変数名	p 次元ベクトル (オプション)
使用する変数	チェックボックス
出力項目	
R の出力結果	モデル式 残差情報 係数情報 統計情報 AIC
データ	入力したデータ 被説明変数の予測値
相関係数	$(p + 1) \times (p + 1)$ 行列
変数間散布図	多変量連関図形式
予測式のグラフ	説明変数が 1 つの場合のみ
その他	R のソースファイル R のデータファイル R の分析結果のファイル

表3 ロジスティック回帰分析の入出力項目

入力項目	
データ数	n
説明変数	n 次元ベクトル
被説明変数	n 次元ベクトル
出力項目	
R の出力結果	モデル式 係数情報 残差情報
データ	入力したデータ 被説明変数の予測値
予測式のグラフ	散布図と予測式
その他	R のソースファイル R のデータファイル R の分析結果のファイル

4.3. 多項ロジット・モデル

多項ロジット・モデルは、離散選択問題における確率選択モデルの一つである [2]。複数の対象から一つを選択する問題において、各対象の効用を比較し、最も効用の大きな対象を選択するという仮定のもと、同時選択確率である尤度を最大にするようなパラメータを求める。回答者 i の対象 j に対する効用を U_{ij} とする。 U_{ij} はモデルで表される、確定的効用 V_{ij} と確率的効用 ε_{ij} により構成される。確定的効用は変数 x_{ik} を用いて、

² ロジスティック回帰分析には他のモデル表現もある。

$$V_{ij} = \sum_k b_k x_k$$

と表すことができ、ランダム効果である確率的効用を二重指数分布に従うと仮定する。したがって効用は、

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

と表すことができるが、このとき回答者 i の対象 j の選択確率は、

$$\Pr\{U_{ij} > U_{il} \mid \forall l \neq j\} = \frac{\exp\{V_{ij}\}}{\sum_k \exp\{V_{ik}\}}$$

と求めることができる。そして、全データに対する尤度を元にパラメータ b_k を求める。

本システムでは、多項ロジット・モデルについては濱岡 [1] のコードを元にした R のスクリプト [10] を元にしてデータの入出力部分を追加している。

表 4 多項ロジット・モデルの入出力項目

入力項目	
属性数	m (最大 10)
ブランド数	r (最大 10)
データ数	n
評価値データ	$(m \times r) \times n$ 行列
選択したブランド	n 次元ベクトル
使用する属性 ブランド定数	チェックボックス
出力項目	
最適パラメータ	係数情報 対数尤度 AIC
データ	入力したデータ
その他	R のソースファイル R のデータファイル R の分析結果のファイル

4.4. 因子分析

因子分析は、商品評価におけるアンケート回答などをもとに、複数の評価項目（変量）の反応のうち共通する項目をまとめることで、次元の縮約をしようとする分析である。マーケティングの分野では、知覚マップの作成や、商品評価データ行列から、製品セグメントを作成するなどといった目的に利用される。

求解方法としては、R のオプションのうちのいくつかを選択できるようになっている(表 5 を参照)。また、因子分析とともによく用いられる主成分分析も因子分析のメニューから分析可能である。

4.5. パレートの分析

パレート分析は ABC 分析ともいわれ、評価対象ごとの売上や利益などを集計し、上位から並び替えてランク付けをしようとする集計分析手法である。そのために、集計対象の変量のアイテムごとに数値を合計し、数値の大きい順に並び替える。そして、それぞれのア

アイテムの構成比率および累積構成比率を求め、アイテムごとの合計を縦棒グラフ、累積構成比率を折れ線グラフとする複合グラフを作成する。そして、たとえば累積構成比率が 70% 以下のアイテムを A ランク、それ以上で 90% 以下のアイテムを B ランク、それ以外を C ランクというように各アイテムをランク付けする。

表 5 因子分析の入出力項目

入力項目	
分析法	最尤法, 主因子法, 主成分分析
軸の回転	回転なし, VariMax 回転 (直交回転) ProMax 回転 (斜交回転)
因子数 (グラフ化する 主成分数) :	m
項目数	p
データ数	n
データ	$n \times p$ 行列
項目名	p 次元ベクトル (オプション)
データ名	n 次元ベクトル (オプション)
出力項目	
入力したデータ	$n \times p$ 行列
因子負荷量	m 個の棒グラフ 負荷量の表
因子負荷量と 因子得点の散布図	因子得点の散布図と 負荷量の矢線グラフ 因子得点の表
R の出力結果	因子負荷量 回転情報 統計情報 相関係数行列 因子得点
その他	R のソースファイル R のデータファイル R の分析結果のファイル

表 6 パレートの分析の入出力項目

入力項目	
データ項目数	p (最大 10)
データ数	n
データ	$p \times n$ 行列
項目名	オプション
集計の基準となる項目	列番号
集計対象の項目	列番号
グループ分けの基準	A ランクと B ランク の累積構成比率
出力項目	
対象別集計結果	数量 構成比率 累積構成比率 グループ
グラフ	ソート後の各対象の数量 累積構成比率の推移
その他	R のソースファイル 結果のファイル

4.6. RFM 分析

RFM 分析は、顧客ごとの購買履歴を多次元的に集計し、3つの視点から顧客の優良度を集計しようという方法である。ここでいう3つの視点とは、1) 基準日からさかのぼった直近購買日 (Recency)、2) 基準日からさかのぼった一定期間における購買回数 (Frequency)、および3) 累積購買金額 (Monetary Value) であり、これらの頭文字をつないで、RFM 分析と呼ばれる。本システムにおいては、各顧客について、RFM のそれぞれの値を集計する。

入力データとしては ID 付 POS データをイメージしており、購買単品ごとの「購買年月日」、「顧客識別番号」、「購買識別番号 (レシート番号)」、「購買個数 (単品ごとの購買金額を算出)」が必要となる。また、それぞれの評価対象をセグメント化するために、各評価項目についてしきい値を設け少数のグループ化をおこなうことも一般的である。

表7 RFM 分析の入出力項目

入力項目	
データ項目数	p (最大 10)
データ数	n
データ	$p \times n$ 行列
基準の項目	列番号 (顧客 ID など)
R, F の項目	列番号 (購買日)
M の項目	列番号 (購買金額)
基準の日付	yyyy/mm/dd
R, F, M の基準	各ランクのしきい値
抽出条件	すべて 非補償型 補償型
出力項目	
各グループの集計値	顧客数 平均日数 平均回数 平均金額
抽出された各顧客の集計情報	顧客 ID 直近購買までの日数 購買回数 購買金額 RFM 各スコア 総合点
RFM 各データの累積出現数	折れ線グラフ
RFM 各データのスコア	ヒストグラム
その他	R のソースファイル 結果のファイル

4.7. RFM 分析 (ファジィ)

RFM 分析 (ファジィ) は、RFM 各スコアの境界値の日数、回数、金額を元に、RFM 各スコアを線型で補間して、RFM 値を整数値ではなく実数値とする。抽出は、ショケ積分 (菅野、室伏 [4]) で計算した総合点をしきい値とする。抽出条件はファジィ測度で与え、与え方により最大値、最小値、加重和での総合評価を

含む RFM スコアの相互作用を考慮した総合点の計算が可能になる。それ以外は RFM 分析と同様である。

5. 現状の利用上の制限

本システムは基本的に利用の制限はないが、現状では以下のような注意点がある。

まず、データ量については、分析に R を用いているため、R を実行する計算機のメモリ量に依存する。R においても大規模データを利用するためのパッケージ (たとえば ff [6] や bigmemory [7]) などもあるが、学生教育目的であれば、現状で大きな問題にはならない。

また、本システムはセキュリティに関しては現状では一切考慮をしていない。したがって、データの送受信の際に、盗聴される可能性もある。特に、企業が情報保護すべきデータを利用するときは別途セキュリティ保護をするような措置が必要である。また、表示用のデータやグラフが一時的にサーバー上に残り、URL を推測すれば、他の人が参照可能な状態になっている。

ウェブ・ベースで利用することを想定しているが、まだ手元に R を実行できるウェブ・サーバがない。将来的には cgi を自由に変更できるレンタル・サーバでの運用などを視野に入れてはいるが、現状では、スタンドアロン PC 上でしか利用できない。

6. まとめ

本研究では、OSS である R をエンジンとした、マーケティング・データ分析のためのウェブ・ベース・システムを開発した。

マーケティング・データ解析では従来の汎用的な分析手法をそのまま利用することももあるが、独自の構造を持った手法も多い。そういった場合でも手軽に利用できる環境を提供することは、本分野における研究・教育において、価値があると考えている。

現段階ではまだ試作段階であるため、利用できる手法は少ないが、仕組みを変えることなく、容易に拡張可能である。また、利便性についてはユーザー・テストを行っていないため、改善の余地があると思われる。

※本稿は、平成 23 年度専修大学情報科学研究所共同研究の研究成果である。

参考文献

- [1] 濱岡豊. 第 8 章 ブランド選択行動の分析, 安厚志他 (編), 経済・経営のための統計学, 有斐閣, 2005.
- [2] 木島正明, 中川慶一郎, 生田目崇 (編著), マーケティング・データ解析, 朝倉書店, 2003.
- [3] 中村永友, 多次元データ解析法, 共立出版, 2009.
- [4] 菅野道夫, 室伏俊明, ファジィ測度 (講座ファジィ 3), 日刊工業新聞社, 1993.
- [5] 高萩栄一郎, 生田目崇, 奥瀬喜之, インターネットで学ぶ社会科学系の数学 [第 4 版], ムイスイ出版, 2010.
- [6] Package "ff", <http://cran.r-project.org/web/packages/ff/ff.pdf#search=truecluster+ff>

- [7] Bigmemory Project, <http://www.bigmemory.org>
- [8] RjpWiki ウェブサイト,
<http://www.okada.jp.org/RWiki/>
- [9] Black-Box --- data analysis on WWW,
<http://aoki2.si.gunma-u.ac.jp/BlackBox/BlackBox.html>
- [10] R—多項ロジットモデル,
<http://aoki2.si.gunma-u.ac.jp/R/mlm.html>

付録 利用例

以下にいくつかの分析手順と結果例示する。

回帰分析：

データは表 A1 のような形式である。ここで、販売量を目的変数、気温、湿度、来客数を説明変数として、回帰分析を行う。データは木島ら [1] で提供されているものを利用した。

表 A1 回帰分析用データ

日付	販売量	気温	湿度	来店客数
Date	Sales amount	Temperature	Humidity	No. of customer
7月1日	112	26.00	93.12	1179
7月2日	29	27.60	85.99	1787
7月3日	47	34.99	79.20	1997
7月4日	105	33.63	86.52	1150
7月5日	107	30.09	70.32	1641
7月6日	116	30.51	98.80	1660
7月7日	104	31.14	72.39	1067
⋮	⋮	⋮	⋮	⋮
9月30日	97	29.94	93.55	1288

このデータを入力フォームにコピー&ペーストする。

図 A1 回帰分析の入力フォーム

データをサーバに送信すると、分析が実行され、図 A2～A4 のような分析結果が返される。なお、図 A2 は R の実行結果、図 A3 は入力データと目的変数の予測値の表、図 A4 は散布図である。

```

線形回帰分析

メッセージ

結果(テキスト)

Call:
lm(formula = y ~ Temperature + Humidity + No._of_, data = lmd)

Residuals:
    Min       1Q   Median       3Q      Max
-72.172 -17.265  -0.252  19.316  63.947

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -220.95907   50.11612  -4.409 2.93e-05 ***
Temperature    1.65568    1.15396    1.435 0.1549
Humidity       2.67519    0.34744    7.700 1.89e-11 ***
No._of_        0.02596    0.01116    2.325 0.0224 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30.5 on 88 degrees of freedom
Multiple R-squared:  0.4262,    Adjusted R-squared:  0.4067
F-statistic: 21.79 on 3 and 88 DF,  p-value: 1.214e-10

AIC: [1] 895.8423
    
```

図 A2 回帰分析の結果出力 1

入力したデータ

表1: 入力したデータと y の予測値

	y	Temperature	Humidity	No._of_	y_Predicted_Value
1	112	26	93.12	1179	101.812306545665
2	29	27.6	85.99	1787	101.172411098318
3	47	34.99	79.2	1997	100.695438621086
4	105	33.63	86.52	1150	96.0359695934005
5	107	30.09	70.32	1641	59.5842692563062
6	116	30.51	98.8	1660	136.962442638676
7	104	31.14	72.39	1067	51.95796423582
8	128	30.21	71.13	1722	64.0528122366225
9	119	27.35	96.85	1134	112.857638531096

図 A3 回帰分析の結果出力 2

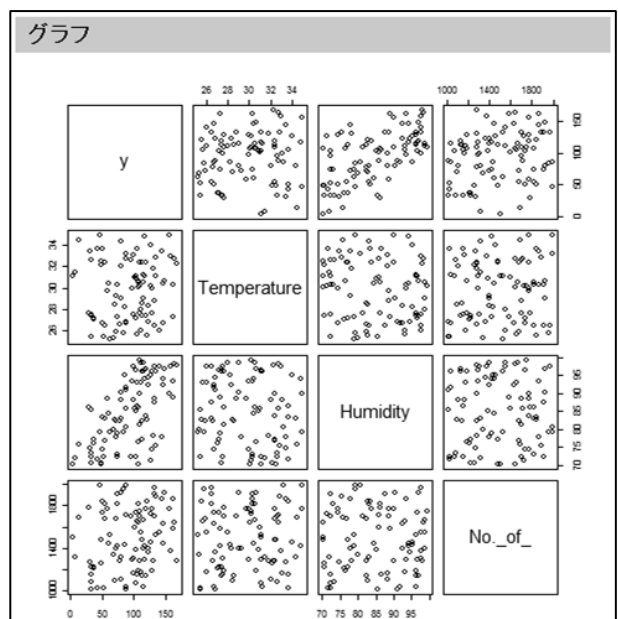


図 A4 回帰分析の結果出力 3

因子分析：

因子分析では、表 A2 のような複数の変数を持つデータを用いる。本データは木島ら [1] で提供されているものを用いた。

表 A2 回帰分析用データ

顧客 ID C_ID	しょうゆ soy source	味噌 miso	ソース source	マヨネーズ mayonnaise
1	8.62	7.37	1.46	1.16
2	9.82	7.25	1.97	0.66
3	6.76	5.16	1.90	1.07
4	9.98	8.36	1.48	0.74
5	8.88	5.98	1.35	0.76
6	10.15	5.68	1.90	0.94
7	7.36	7.43	1.74	0.75
:	:	:	:	:
100	8.54	9.27	1.74	0.89

このデータを入力フォームにコピー&ペーストする。

因子分析(SD法などで利用)・主成分分析を行う

分析法: 最尤法 主因子法 主成分分析 (因子数を求めるとき)
 回転法: 回転なし VariMax回転(直交回転) ProMax回転(斜交回転) 最尤法選択時のみ

因子得点: 回帰法
 項目数:
 データ数:
 因子数(グラフ化する主成分数):
 データ:

8.62	7.37	1.46	1.16
9.82	7.25	1.97	0.66
6.76	5.16	1.90	1.07
9.98	8.36	1.48	0.74
8.88	5.98	1.35	0.76

 項目名:

soy	miso	source	mayonnaise
-----	------	--------	------------

 データ名:

図 A5 因子分析の入力フォーム

データをサーバに送信すると、分析が実行され、図 A6~A8 のような分析結果が返される。なお、図 A6 は入力データ、図 A7 は得られた因子負荷量のグラフ、図 A8 は因子負荷量(矢線)と因子得点(数値)の散布図である。

因子分析

メッセージ

入力したデータ

表1: 入力したデータ

	soy	miso	source	mayonnaise
1	8.62	7.37	1.46	1.16
2	9.82	7.25	1.97	0.66
3	6.76	5.16	1.9	1.07
4	9.98	8.36	1.48	0.74
5	8.88	5.98	1.35	0.76
6	10.15	5.68	1.9	0.94
7	7.36	7.43	1.74	0.75
8	10.33	6.94	1.38	1.09

図 A6 因子分析の結果出力 1

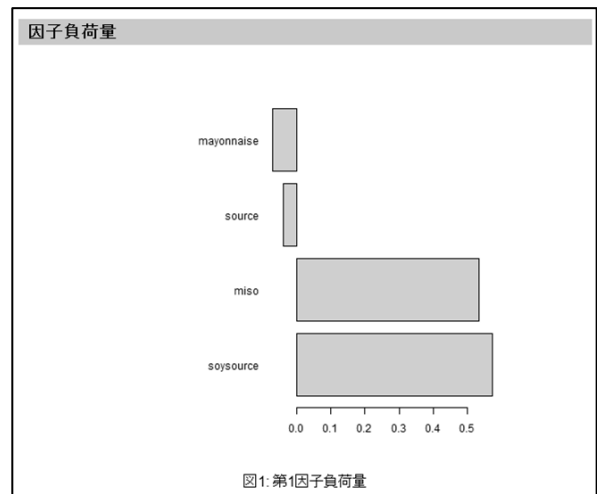


図 A7 因子分析の結果出力 2

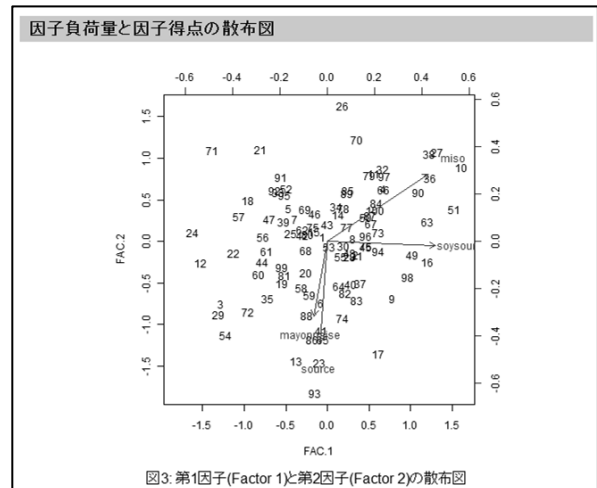


図 A8 因子分析の結果出力 3

パレート分析：

パレート分析では、(ID付) POS データのようなデータ形式を想定している。ただし、パレート分析においては、日付データは必ずしも必要ない。

表 A3 パレート分析用データ

日付 DATE	顧客 ID C_ID	商品 ID P_ID	数量 Units	単価 Price	金額 Amount
2012/8/10	A101	QW-632	2	330	660
2012/8/10	A101	PS-836	1	650	650
2012/8/10	A101	RS-232	2	760	1520
2012/8/10	A102	PS-836	3	650	1950
2012/8/10	A102	QW-632	1	330	330
2012/8/10	B302	NN-653	2	410	820
2012/8/10	B302	QW-632	4	330	1320
2012/8/10	B505	QW-632	2	330	660
2012/8/10	B505	TT-986	1	815	815
2012/8/11	B302	QW-632	2	330	660
2012/8/11	A101	PS-836	5	650	3250
:	:	:	:	:	:
2012/8/17	B505	QW-632	7	330	2310

このデータを入力フォームにコピー&ペーストする。

パレート分析

データ項目数(最大10):

データ数:

データ:

2012/8/10	A101	QW-632	2	330	660
2012/8/10	A101	PS-836	1	650	650
2012/8/10	A101	RS-232	2	760	1520

項目名:

DATE	C_ID	P_ID	Units	Price	Amount

集計の基準となる項目:

集計対象の項目:

グループ分けの基準

グループ名	累積の構成比率
A	<input type="text" value="70"/> %
B	<input type="text" value="90"/> %
C	100 %

図 A9 パレート分析の入力フォーム

データをサーバに送信すると、分析が実行され、図 A10, A11 のような分析結果が返される。なお、図 A11 は集計結果と構成比率、図 A11 はパレート図である。

パレート分析

メッセージ

分析結果

表1: 各項目の合計(頻度)とグループ

	Amount	構成比率	累積構成比率	グループ
QW-632	16170	26.79%	26.79%	A
RS-232	15200	25.19%	51.98%	A
PS-836	11050	18.31%	70.29%	B
QR-803	5733	9.50%	79.79%	B
NN-653	5330	8.83%	88.62%	B
TT-986	4890	8.10%	96.72%	C
WE-999	1980	3.28%	100.00%	C

図 A10 パレート分析の出力結果 1

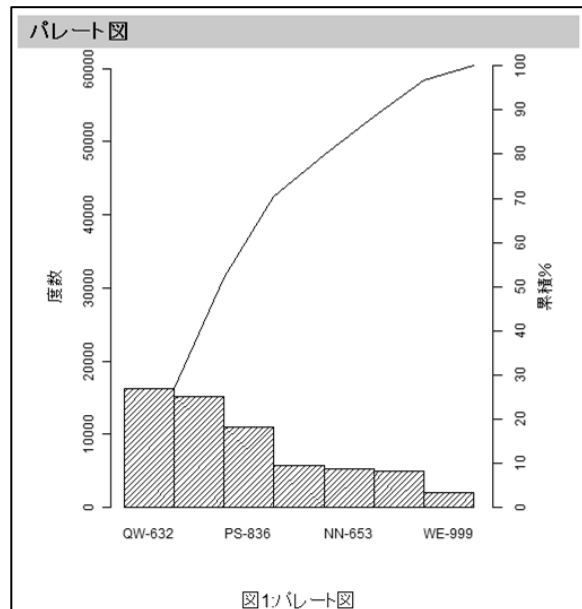


図 A11 パレート分析の出力結果 2

RFM 分析 :

RFM 分析はパレート分析と同様の、ID 付 POS データの形式を想定している。すなわち、単品(カテゴリ)ごとの購買単価、個数、金額とともに、顧客番号と購買日を含むデータである。入力については、前述のデータの他、集計列および、Recency を算出するための基準日の入力が必要である。

RFM 分析

データ項目数(最大10):

データ数:

データ:

2012/8/10	A101	QW-632	2
330	660		
2012/8/10	A101	PS-836	1
650	650		
2012/8/10	A101	RS-232	2

基準の項目(顧客IDなど):

R, Fの項目(購買日):

Mの項目(購買金額):

基準の日付

図 A12 RFM 分析の入力フォーム 1

また、図 A13 のように、RFM それぞれのランクの基準値 (5 段階のうち上位 4 段階のしきい値) についても入力する。R は上位ランクほど数値を小さく、F と M については上位ランクほど数値を大きくする。さらに、サンプルの抽出条件について選択する。

Mのスコアの基準

スコア	基準
5	<input type="text" value="8000"/> 円以上
4	<input type="text" value="6000"/> 円以上
3	<input type="text" value="4000"/> 円以上
2	<input type="text" value="2000"/> 円以上

抽出条件

- すべて
- 条件(非補償型)
 - R: 以上かつ
 - F: 以上かつ
 - M: 以上
- 線型(補償型)

総合点計算式: Rスコア + Fスコア + Mスコア

総合点 以上

図 A13 RFM 分析の入力フォーム 2

結果は図 A14 ~図 A16 のように得られる。図 A14 は、各グループ (ランクの組み合わせ) に関する集計情報、図 A15 は抽出された顧客の RFM の各集計値とスコア、図 A16 は各項目の集計値を並び替えた時の累積出現割合のグラフである。

得点での各グループの集計値

表1:RFMによるグループの集計情報

グループ	顧客数	平均日数	平均回数	平均金額
R=5 F=5 M=5	1	1.00	6.00	13134.00
R=4 F=5 M=5	1	2.00	12.00	16240.00
R=3 F=5 M=5	1	3.00	7.00	11499.00
R=3 F=2 M=3	1	3.00	2.00	4290.00
R=1 F=4 M=2	1	6.00	4.00	3460.00
R=1 F=1 M=5	1	6.00	1.00	11400.00
R=1 F=1 M=1	1	6.00	1.00	330.00

図 A14 RFM 分析の結果出力 1

顧客の抽出

表2:各顧客の集計情報

順位	顧客ID	日数	回数	金額	Rスコア	Fスコア	Mスコア	総合点
1	B505	1	6	13134	5	5	5	15
2	A101	2	12	16240	4	5	5	14
3	A102	3	7	11499	3	5	5	13
4	D981	3	2	4290	3	2	3	8
5	B302	6	4	3460	1	4	2	7
6	O608	6	1	11400	1	1	5	7
7	A303	6	1	330	1	1	1	3

抽出顧客数: 7件

図 A15 RFM 分析の結果出力 2

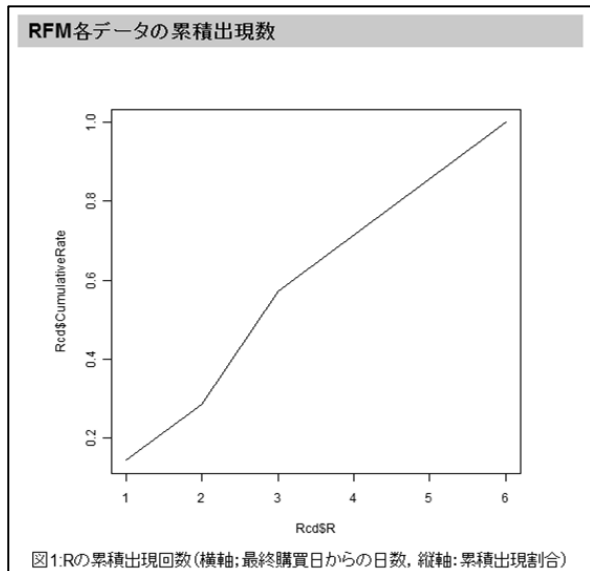


図 A16 RFM 分析の結果出力 3

この他、各クラスの人数のグラフや、総合スコアの分布のヒストグラムが表示される。