

文献紹介# 1 (RFM分析)

2015年2月
技術開発

0. 紹介文献

『RFMデータを用いた顧客生涯価値の算出』

東京大学 阿部誠 マーケティングジャーナル 2014 SUMMER (通巻133号)

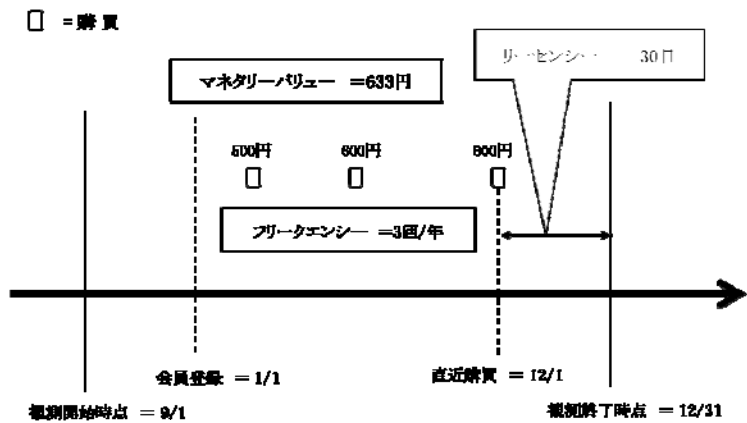
1. 紹介の目的

- 購買履歴データによる顧客セグメンテーション分析、LTV（ここではCLV）算出基準と予測、新規購買顧客の早期生存確率予測やCLV予測はその重要性は喧伝されているが、あまり進んでいないのが現状である。
- 購買履歴データ分析に際し、基本的指標であるRFM、及び、予測に関する考え方のベースを理解していることは重要。
- ◆顧客データ分析の基本的概念であるRFM分析についての理解を深めること、及び、RFM概念の限界と拡張の方向性（知りたいことは、だれが投資する価値があるのか？）を考えてもらうこと。
- ◆現在の業務に拡張してみると、既存クライアント別にRFM分析をおこなうことにより、維持のためにアプローチすべき段階、投資が推定できる。

2. 研究の目的

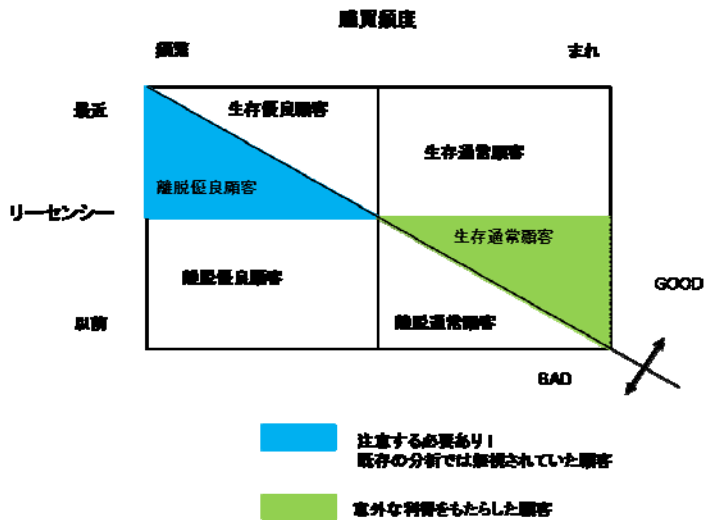
1) 顧客識別の手段 (RFM分析)

- 顧客を識別する目的は、CLV (カスタマー・ライフタイム・バリュー) 算出のため
CLVはLTV (ライフタイム・バリュー) ともいう
いずれにしても顧客生涯価値のこと
 - * CLVが大きければそれなりの投資を、少なければ投資しない
- 現在一般的に使用されているRFM分析は以下の3つ
 - * R=リセナー=直近購買時期、F=フリクエンシー=購買頻度
M=マネタリーバリュー=平均購買金額 (累積購買金額)
- R・Fは観測時点に左右される。また顧客の離脱に大きな影響を受ける、従って、顧客の離脱を把握することが重要。

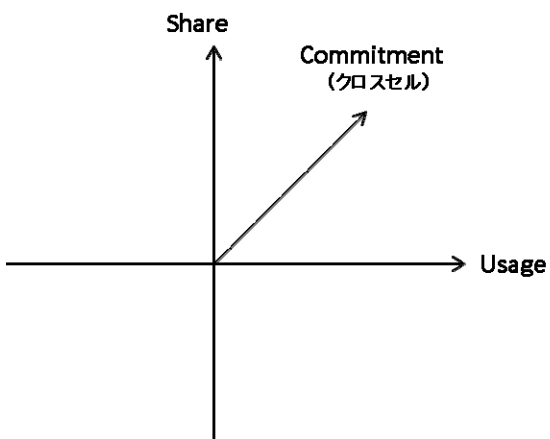


購買頻度

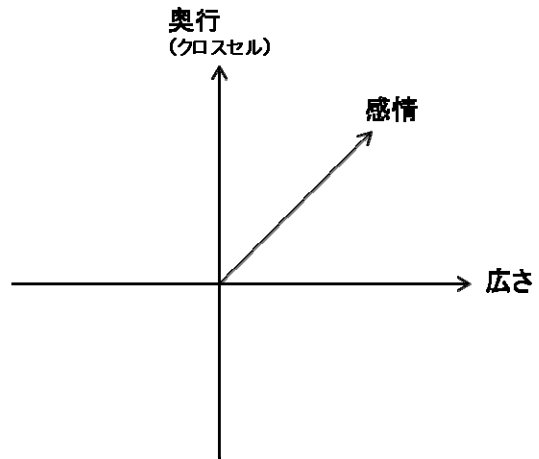
	頻繁	まれ	
最近	生存優良顧客	生存通常顧客	GOOD ↑ ↓ BAD
リセナー	離脱優良顧客	離脱通常顧客	
以前			



その他の考え方 1

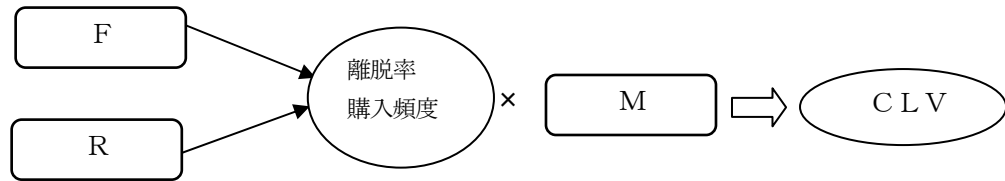


その他の考え方 2



2) 研究の目的

- (1) RとFの複雑な関係を紐解いて、顧客特性である購買頻度と離脱率を導き出し、それに購買金額を加味することによってCLVを算出する。



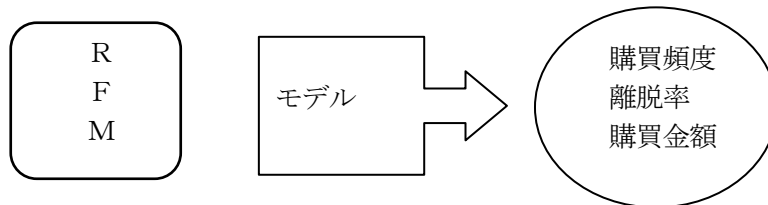
- (2) RとFとMの関係を理解する。

多頻度購買顧客の1回当たりの平均購買金額は低い傾向にある。

(Borle, Singh, and Jain 2008年)

だとすると、多頻度購買顧客のCLVは高いのか、低いのか？

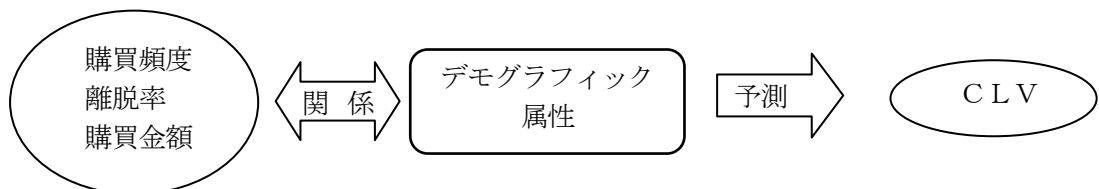
換言すれば、多頻度購買顧客は投資に値するか？



- (3) RFM指標を用いて、その顧客の未来の行動を予測しCLVを算出できるような購買行動モデルを構築し、規範的なインプリケーションを得ることにより、CLVを最大化するマーケティングアクションをマーケティングROIの観点から探る

※ROI：投資利益率 (Return On Investment)

- (4) CLVの概念を見込み客に拡張することによって新規顧客獲得のマーケティング戦略に役立てる。具体的には、既存顧客のデモグラフィック変数とRFMとの関係から予測する。



3) 分析データ

百貨店のポイントカードによって収集された顧客購買履歴データ

3. モデル構築の考え方（仮定）

1) モデルの仮定

(1) 消費者行動の仮定

仮定 1. 購買は、顧客の生存中、その顧客特有の確率でランダムに発生する

仮定 2. 離脱は、現時点までの生存期間に関係なく、その顧客特有の確率でランダムに起きる

仮定 3. 購買金額は、毎回異なるが、その顧客特有の分布に従う

* 上記はマーケティング・サイエンスの過去研究で検証されており、一般化されている

(2) 顧客の異質性に関する仮定

仮定 4. 購買頻度、離脱率、平均購買金額は顧客ごとに異なる

* 顧客の異質性を考慮することは、マーケティングでは不可欠

2) 購買回数、総購買金額、顧客生涯価値の導出

■顧客行動の3指標（RFM）を、仮定1～3に基づいた確率プロセスで表すことにより、その顧客の未来の購買行動を確率的に予測することが可能になる。

例えば、顧客 i の最終購買意向の生存期間は、パラメータ μ_i の指数分布に従う（仮定2 * 顧客 i の離脱は、パラメータ μ_i のハザード率に従う）

■さらに、最終購買から観測終了時点まで購買が発生しなかった顧客の場合、その顧客の購買頻度を考慮することにより、

(a) 生存にも拘わらず、単に次回購買が起きなかったのか、あるいは、

(b) 観測終了前のある時点で顧客が離脱したことによるのか

の可能性を確率的に表現できる。つまり、将来のある時点における生存確率を顧客別に求めることができる。

今回のモデルでは、以下の潜在変数を導入する。離脱状況 Z （離脱 = 0、生存 = 1）、離脱期間 y

3) モデルに使用する統計量

- a) 最終購買以降の期待生存期間
- b) 1年後の維持率
- c) 観測終了時点での生存確率
- d) 将来一定期間中の期待購買回数
- e) 将来一定期間中の期待総購買金額
- f) 顧客生涯価値（CLV）

4) 既存顧客の維持介入

■顧客維持のための投資額とその維持率との関係を仮定することによって、CLVを最大化する顧客維持投資レベルが算出できる。もし、顧客が離脱していた場合、投資レベル c によって一定の確率 $r(c)$ で呼び戻せると仮定すると、関数 $r(c)$ は指数曲線を仮定することによって、いくつかの質問に対し回答をベースに決定解析法で推定できる。

■維持介入戦術案は以下の2つ

- a. CLVの増加が最大になる投資レベルを顧客ごとに算出する（最適投資額の決定）
- b. 1人あたり金額 c を投資するに当たって、何日間購買がないと介入すべきかを顧客ごとに算出する（最適投資時期の決定）

5) 顧客デモグラフィック変数の組み込み

- 新規顧客の獲得をデモグラフィック要因によっておこなうためには、既存顧客の購買頻度、離脱率、購買金額のパラメータ値の違いをこれらの外部要因で説明することが有用となる。
- 本研究では、仮定4に従い顧客別パラメータ $\{\lambda_i, \mu_i, \eta_i\}$ をデモ要因で説明する回帰モデルを構築する。

6) 弾力性の導出

- 顧客行動の3指標(RFM)がCLVにどのような影響を与えるのか、さらに、デモグラフィック特性はCLVにどのような影響を与えるのか、を知るためには、CLVに対するRFMの弾力性と、デモグラフィック変数の弾力性を計算することによって明らかにできる。
※弾力性とは説明変数の変化量に対する目的変数の変化量。
(例、価格弾力性とは価格の変化量に対するシェアの変化量)
- 以下は、通常のLCVの公式から導かれた一般的な公式であり、パラメータの値に依存しない。
 - a. 購買頻度が1%増加するとCLVは1%増加する
 - b. 1回当たりの購買金額が1%増加するとCLVは1%増加する
 - c. 離脱率が1%減少すると、割引率によるが、CLVは1%未満しか増加しない
* 離脱率を減らす効果はCLVの増加に反映しない? (技術開発コメント)

7) モデルの推定

- RFM指標からその顧客行動特性を規定するパラメータ $\{\lambda_i, \mu_i, \eta_i\}$ の値を顧客ごとに求める。未来の購買行動やCLVは、その顧客特有のパラメータの値から予測できる。
- パラメータの推定には、マルコフチェーン・モンテカルロを用いたベイズ手法を使用した。上記モデルの設定により、事前分布と事後分布、尤度関数が求まるので、これをもとに、階層ベイズで個人ごとのパラメータを推定する。また、多変量ベイズ回帰モデルにより、CLVとデモグラフィック変数の関係を推定する。

4. 実証分析

1) 分析対象データ

- 2000年7月中に百貨店のFSPメンバーとなった顧客リストをフレームとして400名をランダム抽出
- 観測期間は2000年7月1日から2001年6月29日までの52週間
 - * 最初の26週データを推定用、残り26週を検証用としている
- 性・年齢・住所がデモ情報
 - * 住所と店舗の距離は必ずしもアクセス容易性とは関係していないため、訪店頻度に占める食品購入回数の割合をアクセスのしやすさを表す説明変数としてモデルに組み込む (Food)

表. 百貨店 FPS データの記述統計

	平均	標準偏差	最少	最大
購入回数	16.0	16.8	0	10.1
観察期間 (日数)	171.2	8.8	15.1	18.1
リセンサー (日数)	24.9	42.8	0	181
平均購入金額 ($\times 10^5$ 円)	0.07	0.12	0.0022	1.83
Food	0.79	0.27	0	1
年齢	52.7	14.6	22	87
性別 (男性=0、女性=1)	0.93	0.25	0	1

2) モデルの検証

- 推定期間の顧客購買データ (RFM指標) と顧客デモグラフィック情報 (性別、年齢、Food) を使ってモデルを推定、顧客別のパラメータは、行動特性の3要素、購買頻度、離脱率、購買金額。顧客共通パラメータは、デモグラフィック変数と3要素との関係を表す回帰係数。モデルが十分な制度を持つことは、推定されたパラメータを使って検証期間の購買行動を予測することで確認された。

3) 既存顧客に関する知見

(1) 顧客別パラメータとCLV

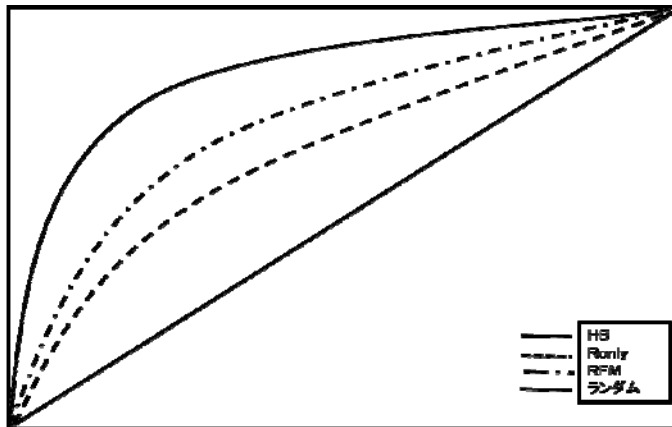
- 推定されたパラメータを見ると期待生存期間は10.0年で、最長24.7年から最短1.3年まで幅がある。また各パラメータとも顧客によって大きな差があることがわかる。CLVを見ると4万円から1,020万円まで大きな差があり、平均は69万円である。

表. パラメータの推定値

	購買頻度	離脱率	購買金額 ($\times 10^5$ 円)	最終購買 以降の期 待生存時 間	1年後の 維持率	観察終了 時点での 生存率	検証期間 中の期待 購買回数	検証期間 中の総購 買金額 ($\times 10^5$ 円)	CLV (\times 10^5 円)
平均	0.66	0.00564	0.038	10.0	0.823	0.929	16.0	0.74	6.9
最少	0.07	0.00165	0.007	1.3	0.454	0.182	0.5	0.01	0.4
最大	3.78	0.04713	0.207	24.7	0.926	1.000	96.1	9.61	102.0

■下図のHB（階層ベイズ）はCLVの高い順に並び替えx軸に顧客の数を、y軸にその累積の生涯価値をプロットしたものである。下図にはこの他にリセンサーの短い順、RFMのランキングの高い順、ランダムに並び替えたものを図示している。これを見ると多くの企業で、リセンサーのみの基準で顧客をランキングしているが、CLVはうまく識別されていない事がこの図から視える。リセンサーのほかに、FとMを加えることにより単純ランキングの単純平均でも識別度が上がることがわかる。ここで、注意しなければならないことは、RFM指標はCLVのランキングを予測する上では有用でも、CLVの絶対的な値はモデルを使用しないと算出できないということである。

図. ゲインチャート



(2) 顧客ベースと顧客資産

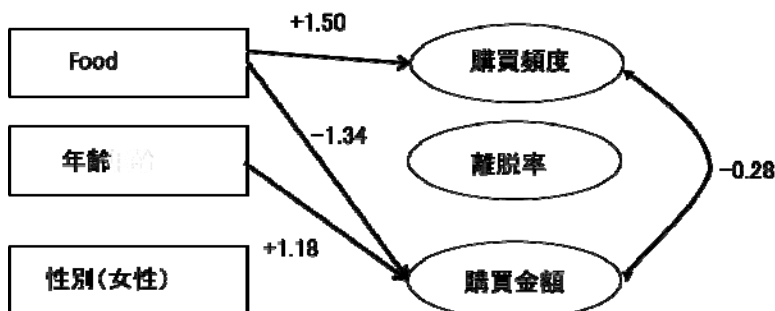
■推定機関直後実際に生存（購入）している顧客の期待数と顧客からもたらされる生涯価値の期待総額（顧客資産）はパラメータの推定値から算出できる。その時点で生存している顧客がベースとなるので $400 \times 0.929 = 371.6$ 人がベースとなる。生存率が低いと登録顧客数と生存顧客数に大きな差が出る。CLVと生存率を顧客ごとに掛け合わせ合計したものが顧客資産になる。今回のデータから計算すると2億7258万円になる。この顧客資産は長期的な視点で見ると重要な指標となる

(3) 顧客共通パラメータの解釈

■3行動特性とデモグラフィック変数の関係を見ると下図のようになる。有意なものだけに値を表示している。一番影響力の強い説明変数は店舗へのアクセスの容易さを表したFoodで、購買頻度に正、購買金額に対して負となった。食品購入者は1回当たりの購買金額は低いが、購入回数は多い。また、年齢は購買金額に正となった。年齢が高いほど購買金額が高いことを示している。

■次に、3行動特性の相関を見ると、頻度と購入金額の相関が負になった。これはFoodの影響が、頻度と金額に対して逆方向に働いている事実と整合性がある。

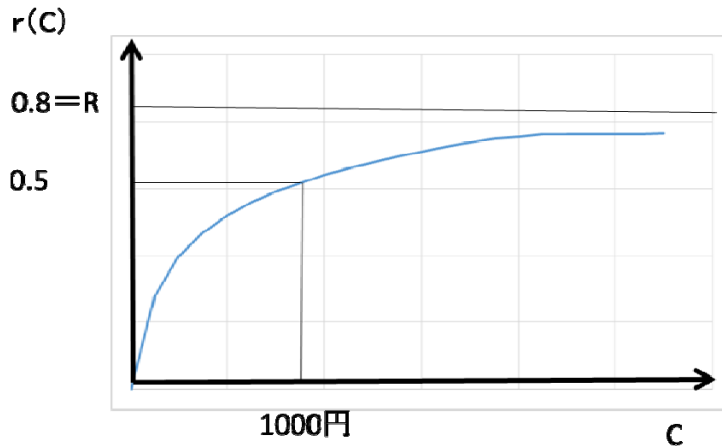
図. 顧客共通パラメータの関係



(4) 既存顧客の維持介入

■顧客が離脱した場合投資レベル c によって確率 $r(c)$ で呼び戻せる下図のような指数曲線を考える。下図の解釈は 1000 円のクーポン(投資)を送付することによって、離脱顧客の半数呼び戻せるが、どんな高額なクーポンを送付しても救出率は 0.8 で頭打ちになるというものである。

図. 維持介入投資金額 c と救出率 $r(c)$ の指数関数



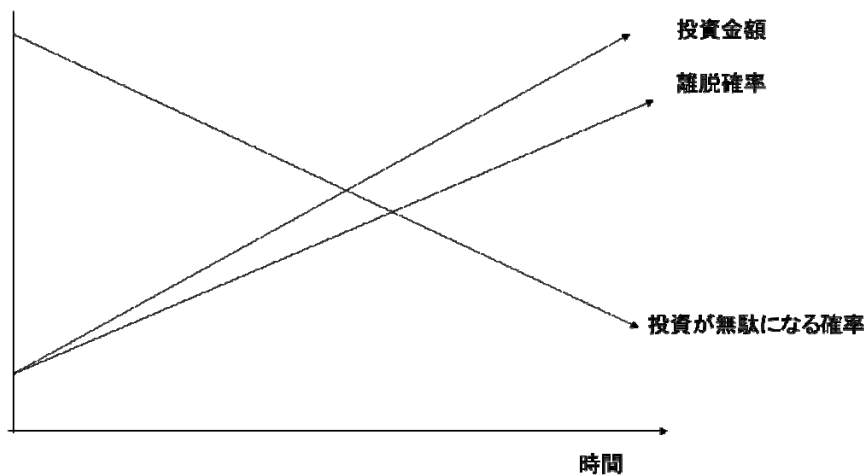
■維持介入のアクションの結果は下表ようになる。介入レベルの平均は 500 円である。400 人中 74 人の顧客は介入すべきでないとなった。

■介入のロジックは、離脱確率が低いと介入投資が無駄になってしまう。離脱確率は時間とともに増加する。しかし待ちすぎると、離脱確率が上がり 500 円での投資では不十分である。一般的に言って CLV が高い顧客には早い時点での介入が効果的である。

表. 顧客別の維持介入アクション

	最適介入レベル (円)	介入までに日数 (日)	観測終了時点での生存確率	CLV ($\times 10^5$ 円)
平均	500	32.6	0.929	6.9
最少	0	3.5	0.182	0.4
最大	1500	120.4	1.000	102.0

図. 介入のロジック



3) 新規顧客獲得に関する知見

- デモグラフィック情報として用いられた、性別、年齢、Food の 3 説明変数が CLV にどのような影響を与えるのかを見たものが下図である。これは 1 つの変数のみを変化させて、残りの 2 つの変数を平均値に固定した場合、CLV がどのように変化するかをプロットしたものである。例えば、説明変数が Food のグラフでは、年齢と性別を固定した場合、食品を購入する顧客の割合が変わるにつれて CLV がどう変化するかを表す。垂直の点線は Food の平均値である。Food と年齢の増加はどちらも CLV を増加させるが、性別の変化は影響が少ない。
- CLV の要因を、下から購入金額、購入頻度、生存期間 (離脱率の逆数) の順に 3 つを積み上げた形になっている。食品を購入する顧客の割合を増やすと、生存期間、購入期間は向上し、1 回当たりの購入金額は低下するが正味の CLV は増加する。高齢者の顧客の割合を増加させると 1 回当たりの購入金額は増加するが、生存期間と購入頻度は変化しない、正味の CLV は増加する。女性客の割合を増やしても 3 要素とも変化はなく、CLV も変化しない。

図. CLV に対する Food の影響度

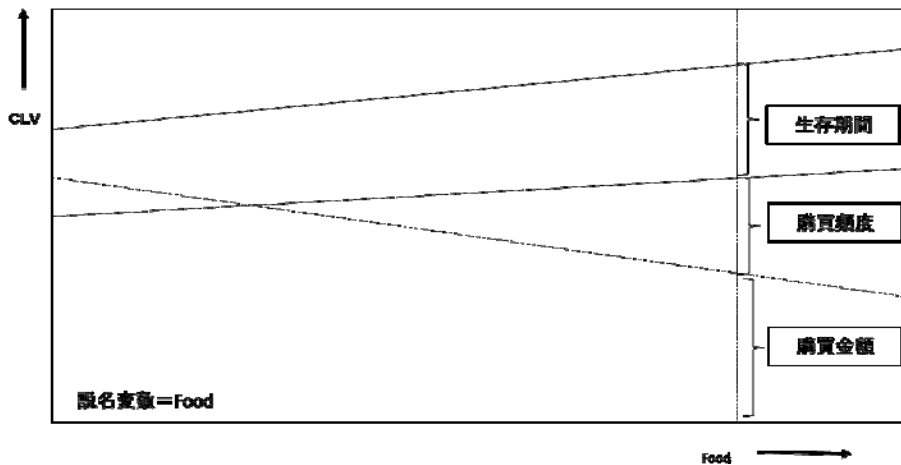


図. CLV に対する年齢の影響度

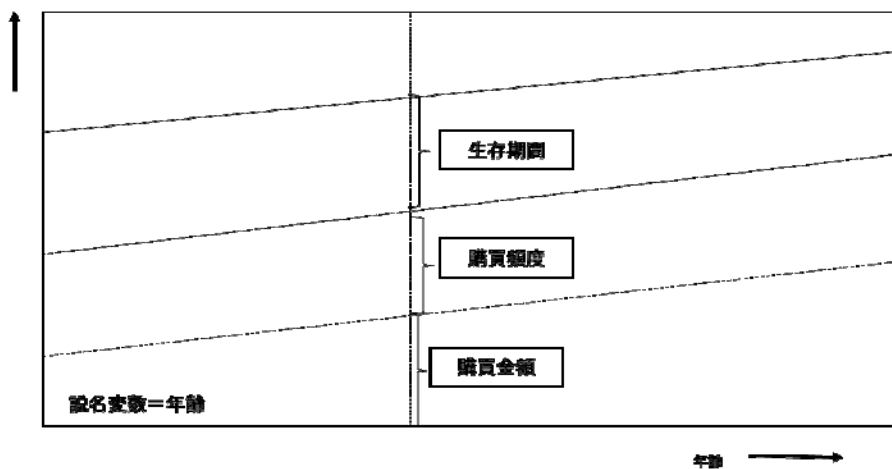
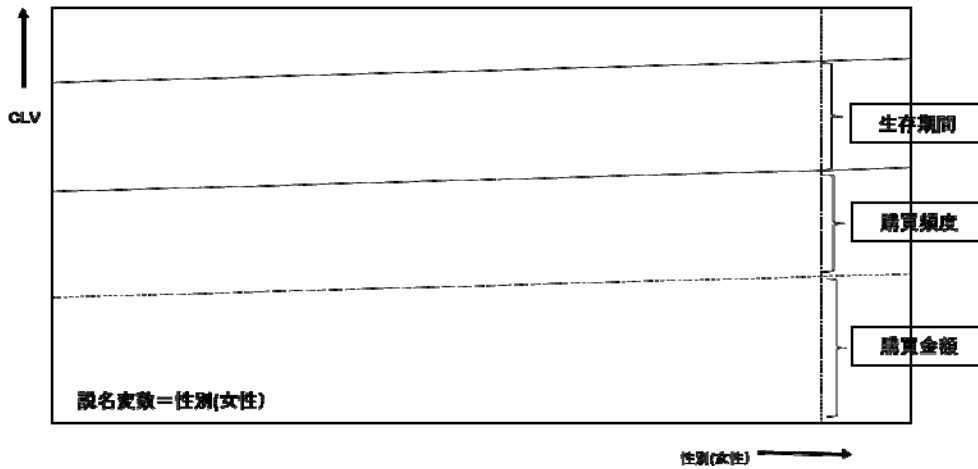


図. CLVに対する性別(女性)の影響度



5. MRSでの活用の方向性

1) 階層ベイズの導入

■階層ベイズを使えるソフトウェアは限られており、高価なものが多い。今後フリーのソフトを探して、階層ベイズに対応できるようにしていきたい。ソフトの候補としては、Rの中にあるベイズ対応のパッケージ、と WinBUGS である。

また、階層ベイズの細かい設定等については、阿部先生の「マーケティングの科学—POS データの解析」で学んでいきたい。

2) 顧客データ分析に対する提案

■MRSが顧客データの分析を提案する上で、本論文で述べられているように、顧客を「リセシー」のみでグループ化するのではなく、顧客をRFMそれぞれにランキングをして、その平均値でグループ化することの方が、顧客の生涯価値を識別する上で有用であるとされており、このRFMのランキングの平均値による顧客のグループ化を提案していくことが必要である。

以上