

# 制約つき比例ハザードモデルを用いた ヘアサロンの再来店状況分析

西村 直樹, 小林 健, 吉住 宗朔

## 1. はじめに

近年の情報技術の発展に伴い、多くの企業にて事業運営に関する大量かつ多様なデータを容易に取得できるようになった。企業が成長を伴い存続するためには、現在の経営状況を定量的に把握してその状況に合った施策を適時に行うことが重要である。

本研究では、ヘアサロンチェーンの複数の店舗における顧客の来店履歴データを用いて、ヘアサロンの売上構成の大部分を占める既存顧客の再来店状況を分析する。これにより各店舗の再来店状況の差異を捉えることができれば、店舗が再来店顧客を増加させるための知見を得ることができる。

顧客の再来店に与える影響を推定する先行研究では、比例ハザードモデル [1] を適用した研究が数多く行われている [2-4]。小西 [2] は比例ハザードモデルを用いてヘアサロンの顧客の再来店に影響を与える要因の分析を行っている。Van den Poel and Lariviere [3] は金融機関の顧客離反に影響を与える要因について分析を行っている。Kapoor et al. [4] はウェブサービスにおける顧客の再来訪時間の推定を行っている。

本研究では、顧客の再来店状況を分析するため、比例ハザードモデルを用いて顧客の累積来店回数が再来店間隔に与える影響を分析する。累積来店回数を説明変数に用いる場合、通常一つの量的変数として扱われるが、この場合累積来店回数とハザード比の間に線形関係の仮定してしまう。一方、累積来店回数別のダミー変数に分解して説明変数に用いると、累積来店回

数の影響の非線形性を表現することが可能となる。しかし各店舗についてその時々（たとえば1カ月ごと）で分析を行う場合、店舗それぞれで利用可能な来店履歴データは少数に限られる。そのため累積来店回数をダミー変数として個別の店舗でモデルのパラメータを推定した際、モデルがデータに対して過剰適合し、推定結果の誤った解釈や予測性能の低下につながる恐れがある。

過剰適合抑制のため、マーケティングにおける顧客セグメンテーションの方法であるRFM分析 [5] の最終来店 (Recency) と対象期間内の累積来店回数 (Frequency) に関する形状制約付きの推定手法が提案されている [6, 7]。さらに、比例ハザードモデルについても偏回帰係数に対して制約条件を課す方法 [8] が提案されている。

そこで本研究では、比例ハザードモデルのパラメータの推定における過剰適合抑制のため、累積来店回数が多くなるほど再来店率が高くなる「単調性」と、累積来店回数が多くなるほど再来店率の増分が減速する「凹性」を仮定し、これらの性質を制約条件として課したモデルを提案する。

実際のデータを用いて提案モデルのパラメータの推定を行った結果、各店舗の各顧客の再来店率の予測性能が制約条件を課さない場合に比べて向上し、特にデータの少ない小規模店舗で予測性能が向上することが確認された。さらに、提案手法により推定されたモデルの偏回帰係数を主成分分析により可視化することで、各店舗のその時々再来店状況の変化を捉えることができた。

本論文の構成は以下のとおりである。2節では、本研究で扱うデータの概要とヘアサロンにおける累積来店回数と再来店率の関係について述べる。3節では、ヘアサロンにおける顧客の再来店間隔を、累積来店回数に基づく制約条件を課した比例ハザードモデルにより推定する方法を提案する。4節では、提案手法の有効

にしむら なおき, よしづみ しゅうさく  
株式会社リクルートライフスタイル  
〒100-6640 東京都千代田区丸の内 1-9-2  
グラントウキョウサウスタワー  
こばやし けん

株式会社富士通研究所  
〒211-8588 神奈川県川崎市中原区上小田中 4-1-1  
受付 18.7.25 採択 18.11.2

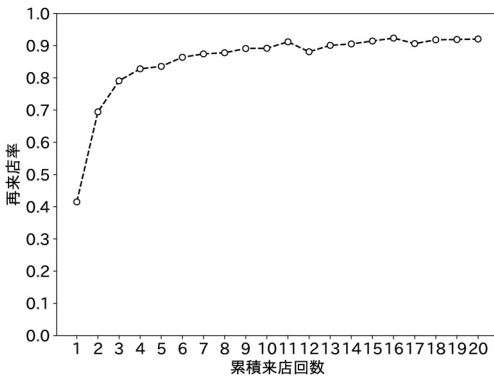


図1 顧客の累積来店回数と再来店率の関係

性を検証し、5節では、提案手法によって得られた顧客の再来店状況の推移を主成分分析を用いて可視化し、各店舗の状況を比較する。最後に6節でまとめと今後の課題を述べる。

## 2. ヘアサロンの再来店の特徴

本節では、本研究で扱うデータの概要とヘアサロンにおける顧客の累積来店回数と再来店率の関係について述べる。

### 2.1 データ概要

本研究では経営科学系研究部会連合協議会主催、平成29年度データ解析コンペティションで提供されたヘアサロンチェーンのデータを利用した。

提供された2015年7月から2017年6月までの2年間のデータのうち、新規店舗や長期間の休業期間があった店舗を除いた10店舗のPOSデータを分析対象とした。このPOSデータのうち、会計履歴と顧客のデモグラフィックをまとめたデータ（顧客マスタ）を用いて分析を行う。会計履歴では会計日、顧客ID、売上、各顧客の累積来店回数などがまとまっている。顧客マスタでは各顧客の郵便番号、性別、年代などがまとまっている。

### 2.2 累積来店回数と再来店率の関係

ここでは累積来店回数と再来店の関係について分析する。データ期間の前半2015年7月1日から2016年6月30日の1年間に上述の10店舗に来店した顧客を対象として、翌365日以内の来店を再来店顧客と定義し、顧客の累積来店回数と再来店率の関係について集計を行った。集計した結果を図1にまとめる。

図1から累積来店回数と再来店率の関係について以下二つの傾向があることがわかる：

1. 単調性：累積来店回数が多い顧客ほど再来店率は高い。

2. 凹性：累積来店回数が多くなるにつれて再来店率の増分は減減する。

次節では、このような単調性と凹性を考慮して累積来店回数が再来店に与える影響を分析するための制約つき比例ハザードモデルについて説明する。

## 3. 制約つき比例ハザードモデル

本節ではDeng et al. [8]に従って本研究で提案する制約つき比例ハザードモデルとそのパラメータの推定方法について述べる。

### 3.1 比例ハザードモデル

ここでは店舗ごとの再来店傾向を顧客の再来店間隔をもとに分析するため、比例ハザードモデル [1] を用いて再来店間隔をモデル化する。まず来店時から次に来店するまでの期間（来店間隔）を確率変数  $T$  とする。来店時から時点  $t$  まで来店しなかった顧客が次の瞬間に来店する確率（ハザード率）を  $\lambda(t)$  と表す。すなわち

$$\lambda(t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \Pr\{t \leq T \leq t + \Delta \mid T \geq t\}$$

である。ここで  $\Pr\{A \mid B\}$  は事象  $B$  が生じたもとで事象  $A$  が生起する条件つき確率を表す。比例ハザードモデルではハザード率に対して以下の形を仮定する：

$$\lambda(t \mid \mathbf{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}). \quad (1)$$

ここで  $\mathbf{x} = (x_1, x_2, \dots, x_M)^\top$  は  $M$  個の説明変数であり、 $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_M)^\top$  は対応する偏回帰係数を表す。また  $\lambda_0(t)$  は説明変数  $\mathbf{x}$  に依存しないハザード率を表し、ベースラインハザードとよばれる。なおハザード率とベースラインハザードの比の対数は

$$\log \left( \frac{\lambda(t \mid \mathbf{x})}{\lambda_0(t)} \right) = \boldsymbol{\beta}^\top \mathbf{x} \quad (2)$$

であり、 $\mathbf{x}$  に関する線形関数であることに注意する。なお以降では簡単のため、式 (2) を対数ハザード比とよぶことにする。

本研究では顧客の累積来店回数が来店間隔に与える影響を分析するため、比例ハザードモデルの説明変数には累積来店回数と各顧客の性別や居住地といった顧客属性を用いる。ここで顧客属性を説明変数として用いるのは、累積来店回数だけでなく顧客属性も来店間隔に影響を与えているという仮定に基づく。

累積来店回数は通常一つの量的変数として扱われるが、この場合累積来店回数と対数ハザード比との間に線形関係を仮定してしまう。しかし2節で累積来店回数と再来店率の間に単調性と凹性という非線形の関

係が見られたように、累積来店回数と対数ハザード比の間にも同様の非線形の関係があると考えられる。そこで本研究ではその非線形な関係を捉えるため、累積来店回数は量的変数としてではなく、回数別のダミー変数として説明変数に用いる。いま説明変数の添え字集合を、 $D = \{1, 2, \dots, K\}$ ,  $U = \{K + 1, K + 2, \dots, M\}$  (ただし  $1 \leq K < M$ ) と二つに分割し、累積来店回数に関するダミー変数を  $\mathbf{x}_D := (x_k)_{k \in D}$ 、顧客属性に関する説明変数を  $\mathbf{x}_U := (x_k)_{k \in U}$  とする。そして説明変数全体を  $\mathbf{x} = (\mathbf{x}_D^\top, \mathbf{x}_U^\top)^\top$  と表す。ここで  $\mathbf{x}_D$  は

$$x_k = \begin{cases} 1 & (\text{累積来店回数が } k \text{ の場合}) \\ 0 & (\text{そうでない場合}) \end{cases} \quad (k \in D)$$

とする。説明変数  $\mathbf{x} = (\mathbf{x}_D^\top, \mathbf{x}_U^\top)^\top$  に対応する偏回帰係数を  $\boldsymbol{\beta} = (\boldsymbol{\beta}_D^\top, \boldsymbol{\beta}_U^\top)^\top$  と表す。

累積来店回数のダミー変数  $x_k$  ( $k \in D$ ) に対応する偏回帰係数  $\beta_k$  ( $k \in D$ ) に着目すると、 $\beta_k$  の値が大きければ累積来店回数  $k$  回の顧客の来店間隔が短い (すなわち再来店しやすい) ことを表す。以降では  $\beta_k$  ( $k \in D$ ) の指数をとった値  $\exp(\beta_k)$  ( $k \in D$ ) を累積来店回数が  $k$  回の顧客の定着度とよぶことにする。

### 3.2 比例ハザードモデルの部分尤度関数

ここでは比例ハザードモデルのパラメータの推定に用いる部分尤度関数について説明する。ある一つの店舗に着目し、データの期間内で観測された  $N$  個の来店に関する標本の集合を  $\{(t^i, \delta^i, \mathbf{x}^i)\}_{i=1}^N$  と表す。ここで  $t^i$  は当該の来店から同一の顧客がデータ期間内に再来店するまでの期間 (来店間隔) を表す。ただしデータ期間内に再来店がなかった場合 (すなわち打ち切りデータの場合)、 $t^i$  は来店時からデータ期間の最後の時点までの期間とする。 $\delta^i$  はデータ期間内に再来店があったか否かを表し、データ期間内に再来店があった場合 1、そうでなければ 0 とする。 $\mathbf{x}^i = ((\mathbf{x}_D^i)^\top, (\mathbf{x}_U^i)^\top)^\top$  はその標本に対応する説明変数である。また各標本の来店間隔  $t^i$  に対し、来店間隔が  $t^i$  以上となる標本の集合 (リスク集合) を  $K(t^i)$  と表す。すなわち

$$K(t^i) = \{j \mid t^j \geq t^i\}$$

である。

いま来店間隔  $t^i$  ( $i = 1, 2, \dots, N$ ) がすべて異なる値で  $t^1 < t^2 < \dots < t^N$  と昇順で並べられているものとする。このときリスク集合  $K(t^i)$  が与えられたもとの説明変数  $\mathbf{x}^i$  の標本の来店間隔が  $t^i$  となる条件つき確率は

$$\frac{\lambda_0(t^i) \exp(\boldsymbol{\beta}^\top \mathbf{x}^i)}{\sum_{j \in K(t^i)} \lambda_0(t^j) \exp(\boldsymbol{\beta}^\top \mathbf{x}^j)} = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}^i)}{\sum_{j \in K(t^i)} \exp(\boldsymbol{\beta}^\top \mathbf{x}^j)} \quad (3)$$

と表される。比例ハザードモデル (1) の部分尤度関数はデータ期間内で再来店のある標本 (すなわち  $\delta^i = 1$  となる標本) について式 (3) を掛け合わせた

$$\prod_{i: \delta^i = 1} \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}^i)}{\sum_{j \in K(t^i)} \exp(\boldsymbol{\beta}^\top \mathbf{x}^j)} \quad (4)$$

として定義される。したがって比例ハザードモデルの対数部分尤度関数は式 (4) の対数をとって以下のようになる：

$$\begin{aligned} LL(\boldsymbol{\beta}) &:= \log \left[ \prod_{i: \delta^i = 1} \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}^i)}{\sum_{j \in K(t^i)} \exp(\boldsymbol{\beta}^\top \mathbf{x}^j)} \right] \\ &= \sum_{i: \delta^i = 1} \boldsymbol{\beta}^\top \mathbf{x}^i - \sum_{i: \delta^i = 1} \log \left[ \sum_{j \in K(t^i)} \exp(\boldsymbol{\beta}^\top \mathbf{x}^j) \right]. \end{aligned}$$

比例ハザードモデルの偏回帰係数の推定は、対数部分尤度関数  $LL(\boldsymbol{\beta})$  を最大化する最適化問題の求解により行われる。

### 3.3 制約つき比例ハザードモデルのパラメータ推定

2節で述べたとおり、顧客の累積来店回数と再来店率の間には単調性と凹性の関係があることが確認された。これら単調性、凹性の関係は、累積来店回数のダミー変数に対応する偏回帰係数  $\boldsymbol{\beta}_D$  についても同様に満たされると考えられる。そこで本研究では偏回帰係数  $\boldsymbol{\beta}_D$  に対して、累積来店回数に関する単調性と凹性の制約を課して推定を行う。

まず偏回帰係数  $\beta_k$  ( $k \in D$ ) が累積来店回数  $k$  に対して単調増加する性質 (単調性) は以下の制約として表される：

$$\beta_k \leq \beta_{k+1} \quad (k = 1, 2, \dots, K-1). \quad (5)$$

同様に、偏回帰係数  $\beta_k$  ( $k \in D$ ) の増分が累積来店回数  $k$  に対して単調減少する性質 (凹性) は以下の制約として表される：

$$\beta_{k+1} - \beta_k \geq \beta_{k+2} - \beta_{k+1} \quad (k = 1, 2, \dots, K-2). \quad (6)$$

これらの制約を課して対数部分尤度関数を最大化する最適化問題は以下のようになる：

$$\begin{aligned} &\text{最大化} && LL(\boldsymbol{\beta}) \\ &\text{制約条件} && \text{式 (5), (6)}. \end{aligned} \quad (7)$$

表 1 各手法で計算した平均対数部分尤度の平均値

	店舗									
	A	B	C	D	E	F	G	H	K	L
線形	-5.3457	-5.4724	-4.6284	-5.0375	-4.7545	-4.6068	-5.0867	-4.3243	-4.9772	-4.1188
制約なし	-5.3279	-5.4417	-4.5751	-4.9884	-4.7071	-4.5711	-4.8515	-4.2103	-4.9259	-4.0781
単調性	-5.3257	-5.4397	-4.5736	-4.9869	-4.7047	-4.5682	-4.8490	-4.2081	-4.9240	-4.0749
単調性+凹性	<b>-5.3253</b>	<b>-5.4395</b>	<b>-4.5731</b>	<b>-4.9865</b>	<b>-4.7042</b>	<b>-4.5676</b>	<b>-4.8484</b>	<b>-4.2075</b>	<b>-4.9236</b>	<b>-4.0740</b>

問題 (7) は最大化する目的関数が凹関数で実行可能領域が凸集合となる凸最適化問題であり、逐次二次計画法 [9] などを用いれば大域的最適解を得ることができる。

なお本研究では来店間隔を月単位で計算しているため、パラメータの推定に用いる標本の集合は来店間隔の等しい標本 (タイデータ) を含む。そこで実際にパラメータを推定する際は対数部分尤度関数  $LL(\beta)$  に Breslow の近似 [10] を適用した。

#### 4. 予測性能の検証

本節では、3 節で説明した制約つき比例ハザードモデルを適用し、その予測性能を検証する。

##### 4.1 検証の設定

提供された 2015 年 7 月 1 日から 2017 年 6 月 30 日までの 2 年間のデータを用いて、2015 年 7 月 1 日から 2016 年 6 月 30 日、2015 年 8 月 1 日から 2016 年 7 月 31 日というように、1 カ月ずつ期間をずらした 12 カ月間の 12 データセットを作成した。検証においては、それぞれのデータセットの中でデータをランダムに 3 等分し、一つをテストデータ、残りを学習用データとして検証を 3 回繰り返す 3 分割交差検証を行った。評価指標としてテストデータに対する平均対数部分尤度を用いる。ここで平均対数部分尤度は対数部分尤度を標本数で割った値である。テストデータに対する平均対数部分尤度が高いモデルはパラメータの推定に用いなかった未知データに対して当てはまりがよいことを表し、テストデータに対する平均対数部分尤度は比例ハザードモデルの汎化性能を評価する指標として用いられる [11]。

比例ハザードモデルでは説明変数として以下のものを用いた：

- ・顧客の累積来店回数：累積来店回数 1 回目から 19 回目について当該累積来店回数であれば 1、そうでなければ 0 とするダミー変数；
- ・性別：男性であれば 1、女性であれば 0；
- ・年代：「30 代」、「40 代」、「50 代以降」についてそれぞれ当該年代であれば 1、そうでなければ 0 と

するダミー変数；

- ・前回来店時売上：平均 0、分散 1 になるように正規化した連続変数；
- ・店舗との距離：店舗と顧客が住所登録した郵便番号の座標間の距離が「2 km 以上 5 km 未満」、「5 km 以上」についてそれぞれ当該距離であれば 1、そうでなければ 0 とするダミー変数；
- ・来店月：1 月から 11 月について当該来店月であれば 1、そうでなければ 0 とするダミー変数。

なお顧客属性を表すダミー変数である性別、年代、店舗との距離、来店月について、今回使用するデータセットをすべてまとめたデータに対してそれぞれログランク検定 [12] を行ったところ、これらの説明変数を含めた場合と含めなかった場合の差はいずれも有意水準 1% のもとで 0 ではなかった。

問題 (7) は python ライブラリ `scipy` の `optimize.minimize` 関数<sup>1</sup> を用いて逐次二次計画法 (`method='SLSQP'`) により求解した。

##### 4.2 単調性制約と凹性制約の有効性

ここでは本研究で提案する制約つき比例ハザードモデルの有効性を検証する。各店舗に対して、累積来店回数を一つの量的変数として扱った場合 (線形)、ダミー変数として扱い制約を課さなかった場合 (制約なし)、制約を課した場合 (単調性、単調性+凹性) それぞれの手法を全 12 個のデータセットに対して適用し、各データセットで 3 分割交差検証を行った。得られた平均対数部分尤度 (12 個 × 3 回 = 36 個の平均対数部分尤度) の算術平均を店舗別にまとめた結果を表 1 に示す。

表 1 を見ると、今回検証に利用したすべての店舗について累積来店回数を量的変数として扱うよりもダミー変数として扱った場合のほうが平均対数部分尤度の平均値は高いことがわかる。また制約なしの比例ハザードモデルに比べて単調性制約つきの比例ハザードモデルのほうが、さらに単調性制約つき比例ハザードモデルに比べて単調性+凹性制約つき比例ハザードモデル

<sup>1</sup> <https://docs.scipy.org/doc/scipy/reference/optimize.minimize-slsqp.html>

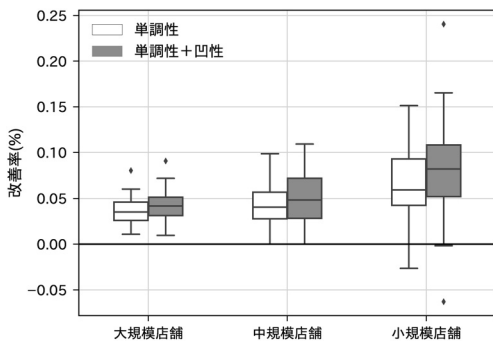


図2 店舗の規模別の平均対数部分尤度の改善率の比較

のほうが平均対数部分尤度の平均値は高い。この結果より、累積来店回数と再来店率に対して非線形に加えて単調性と凹性を仮定したことで予測性能が改善したと考えることができる。

続いて店舗の来客数ごとに大規模店舗 (A, B, D), 中規模店舗 (C, E, G, K), 小規模店舗 (F, H, L) と3区分に層別して、制約なし、単調性制約、単調性+凹性制約の比例ハザードモデルの比較を行う。ここでは表1と同様、各店舗に対して全12個のデータセットそれぞれで3分割交差検証を行って得られた各平均対数部分尤度に対して改善率を計算し、それらを店舗の各規模で層別して箱ひげ図を描く。ここで改善率は制約つきモデルの平均対数部分尤度を  $MLL_a$ 、制約なしモデルの平均対数部分尤度を  $MLL_b$  として、

$$\text{改善率 (\%)} = \frac{MLL_a - MLL_b}{MLL_b} \times 100$$

と定義した。店舗規模別の改善率をまとめた箱ひげ図を図2に示す。図2からは、いずれの店舗規模でも単調性制約つきモデルと比較して単調性+凹性制約つきモデルのほうが改善率の中央値が大きい値をとっていることがわかる。これら二つのモデルの改善率の結果に対してウィルクソンの符号順位検定 [13] を行ったところ、いずれの店舗規模においても有意水準1%のもとで改善率の差は0ではないことが確認された。また図2からは店舗の規模が小さくなるほど制約なしのモデルと単調性制約を課したモデル、制約なしモデルと単調性+凹性制約を課したモデルの改善率の差が大きくなるのがわかり、またその差は単調性+凹性制約つきモデルのほうが大きいことがわかる。

図3の小規模店舗Lの定着度を見ると、単調性制約を加えることで累積来店回数間の定着度の逆転が補正され、また凹性制約を加えることで定着度の変動が抑えられることがわかる。一方、図4の大規模店舗Aを見ると、小規模店舗と同程度に単調性制約により累積

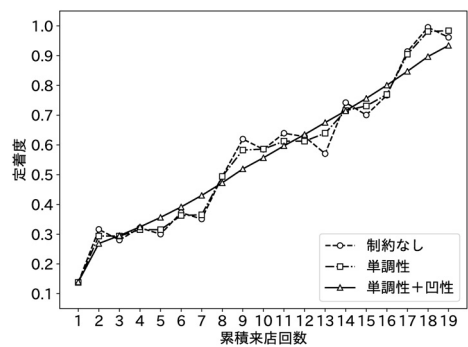


図3 小規模店舗Lでの定着度の比較

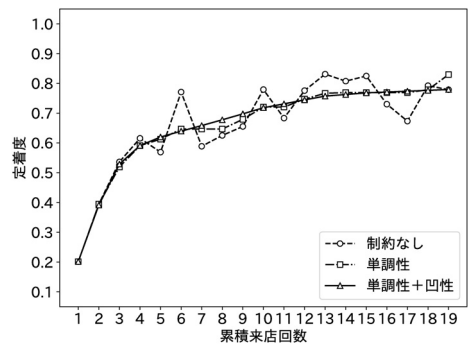


図4 大規模店舗Aでの定着度の比較

来店回数に対する定着度の逆転が補正されたが、単調性制約を満たした時点で凹性制約もおおむね満たしており、小規模店舗より凹性制約による補正幅が小さいことがわかる。このような結果は、来客数の少ない小規模店舗では過剰適合が生じやすく、単調性制約に加えて凹性制約を課すことの効果が大きく表れていたのが要因であると考えられる。

## 5. 主成分分析による店舗状況推移の可視化

本節では、提案モデルを用いて推定した各月の店舗ごとの定着度に対して主成分分析を適用することで状況推移を可視化し、各店舗の状況について考察を行う。

### 5.1 定着度に対する主成分分析の適用

3節で述べた提案モデルでは、各店舗に対して1回来店の顧客、2回来店の顧客など、それぞれの累積来店回数に対して定着度の推定を行った。しかし、累積来店回数による顧客の再来店の差異を細かく考察したい場合、各店舗の定着度が高次元となり解釈が困難となる。

そこで本研究では、各店舗の定着度に対して主成分分析を適用することで低次元の合成変数を作成し、状況を解釈しやすくすることを考える。具体的には図5のように、各店舗に対して推定した19次元の定着度か

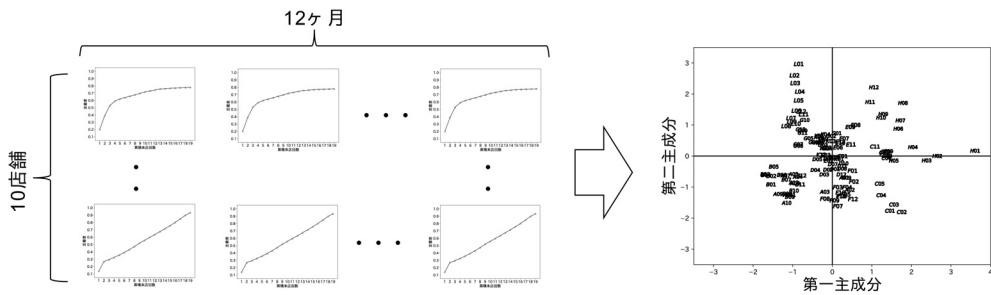


図5 各月、各店舗の定着度の主成分分析による可視化の概念図

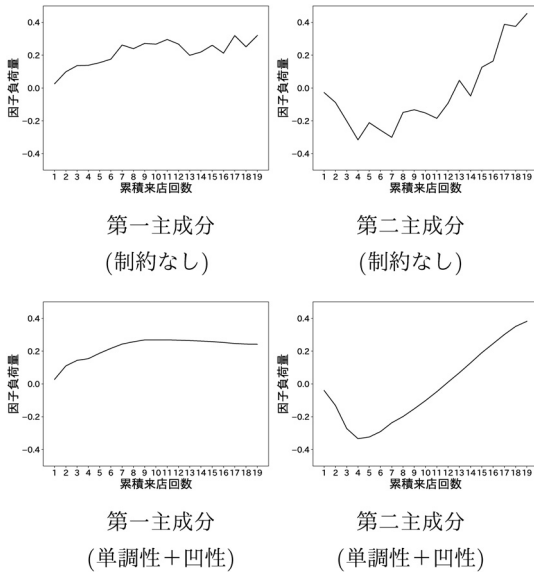


図6 主成分分析の因子負荷量

ら2次元の合成変数を作成することで各店舗を2次元平面上へと射影する。

2016年7月から2017年6月までの店舗ごとに推定された定着度に対して主成分分析を適用し、第一主成分と第二主成分の因子負荷量と寄与率をそれぞれ図6、表2にまとめる。図6より、単調性+凹性制約を課した場合は制約なしの場合と比較して累積来店回数に対する因子負荷量が平滑化されていることがわかる。また表2より、単調性+凹性制約を課した場合は累積寄与率が増加しているとわかる。

得られた主成分について考察する。第一主成分は各累積来店回数でそれぞれ同程度の値をとり、符号がすべて正であることから各累積来店回数の顧客の「平均的な定着度」の大きさと解釈できる。また第二主成分は累積来店回数が少ない場合には負の値をとり、累積来店回数が多い場合には正の値をとることから、累積来店回数が多い顧客と少ない顧客の「定着度の乖離」の

表2 寄与率の比較

	寄与率		
	第一主成分	第二主成分	累積
制約なし	0.596	0.106	0.702
単調性+凹性	0.845	0.116	0.961

大きさと解釈できる。第二主成分の値が他店舗と比較して相対的に大きい店舗は、累積来店回数が少ない顧客の定着度が累積来店回数が多い顧客に比べて低いことを表し、累積来店回数の少ない顧客について再来店率改善の余地があると考えられる。一方で相対的に小さい店舗は累積来店回数の多い顧客について再来店率改善の余地があると考えられる。

作成した因子負荷量を用いて、2016年7月から2017年6月の各月で各店舗の定着度に対して主成分得点を算出し、定着度推移の可視化を行った。紙面の都合上、図7では2016年7月を1カ月目として1~4, 4~8, 8~12カ月目にあたる2016年7月~10月、2016年10月~2017年2月、2017年2月~6月についてのみ抽出し掲載する。図中では年月を数字で表しており、期間の始まりである2016年7月を01とし、期間の終わりである2017年6月を12としている。

### 5.2 各店舗状況推移に対する考察

図7の定着度推移に基づいて、推移の特徴的な店舗に対して考察を行う。

■店舗A 2016年7月には第一主成分(平均的な定着度)の値が他店舗と比較して高かったが、2017年2月まで減少し全店舗で下位3番目まで減少した。平均的な定着度はその後増加するもの他店舗と比較すると依然として低い値であり、12カ月間を通して顧客全体の平均的な定着度は悪化したと解釈できる。また第二主成分(定着度の乖離)については12カ月間を通して他店舗と比較して低い値をとっており、相対的に累積来店回数の多い顧客の再来店率に課題があると解釈できる。店舗Aが累積来店回数の多い顧客の再来店率

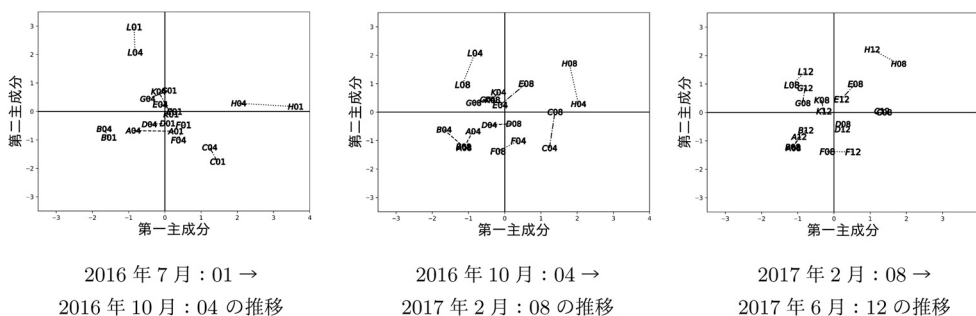


図7 各店舗の因子負荷量の推移

の改善を行うためには、12カ月を通して他店舗と比較して平均的な定着度が高く、かつ定着度の乖離の値が12カ月間で負の値から正の値に推移し累積来店回数の多い顧客の再来店率に対する改善が行われたと思われる店舗Cのオペレーション改善が参考になると示唆される。

■店舗L 第一主成分の値は他店舗と比較して小さい値であり、期間全体を通じてほぼ同程度の値をとっている。一方、第二主成分の値は2016年6月から2017年2月にかけて低下していたものの、それ以降2017年6月までは増加に転じている。このことから店舗Lは他店舗と比較して直近では新規顧客の再来店率が改善の傾向にあると解釈できる。店舗Lがさらに新規顧客の再来店率改善を行うためには、12カ月目において平均的な定着度が高い値をもち、定着度の乖離の値が低い店舗D、Fなどにおける顧客に対するオペレーションが参考になると示唆される。

店舗状況を表すデータが今後更新された場合にも、同じ因子負荷量を用いて同じ軸での解釈が可能になり、状況の変化を連続的に把握することができる。

## 6. おわりに

本研究では、複数の店舗をもつヘアサロンを対象とし、累積来店回数と再来店率に関する単調性と凹性を考慮した比例ハザードモデルにより再来店状況の分析を行った。ヘアサロンチェーンの実際のデータを用いた分析の結果、累積来店回数と再来店率の関係を考慮し制約条件を課すことで比例ハザードモデルの予測性能が向上し、特に来客数が少ない小規模店舗についてその効果が大きいという結果を得ることができた。

また、各店舗の顧客再来店の状況推移を把握するため、各月のデータを用いて比例ハザードモデルにより推定した定着度を主成分分析により可視化した。これにより各店舗における顧客の再来店状況の推移を捉え

ることが可能となり、それぞれの店舗における再来店率改善のための指針を得ることができた。

今後の課題としては、累積来店回数や顧客属性だけでなく会計情報などを用いてより詳細に再来店状況を分析することが挙げられる。また本研究では各店舗、時期による再来店状況の異質性を考慮するため、店舗別、月別で層別してモデルを構築している。先行研究ではこれらの異質性を考慮するために潜在クラスモデルや階層バイズモデルなどのアプローチ [14, 15] が提案されており、これらの手法を組み合わせた分析も今後の課題である。

謝辞 本論文の執筆にあたり、筑波大学高野祐一准教授から多くの有用な助言をいただきました。また、株式会社リクルートライフスタイル大原一輝氏から統計的検定に関する助言をいただきました。最後になりましたが、データを提供いただいた経営科学系研究部会連合協議会の皆様と提供者の方々にお礼申し上げます。

## 参考文献

- [1] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society, Series B (Methodological)*, pp. 187–220, 1972.
- [2] 小西葉子, "存続時間分析による美容院顧客の来店確率予測," *統計数理*, **54**, pp. 445–459, 2006.
- [3] D. Van den Poel and B. Lariviere, "Customer attrition analysis for financial services using proportional hazard models," *European Journal of Operational Research*, **157**, pp. 196–217, 2004.
- [4] K. Kapoor, M. Sun, J. Srivastava and T. Ye, "A hazard based approach to user return time prediction," In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1719–1728, 2014.
- [5] A. M. Hughes, *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-based Marketing Program*, McGraw-Hill, 2000.
- [6] J. Iwanaga, N. Nishimura, N. Sukegawa and Y.

- Takano, “Estimating product-choice probabilities from recency and frequency of page views,” *Knowledge-Based Systems*, **99**, pp. 157–167, 2016.
- [7] N. Nishimura, N. Sukegawa, Y. Takano and J. Iwanaga, “A latent-class model for estimating product-choice probabilities from clickstream data,” *Information Sciences*, **429**, pp. 406–420, 2018.
- [8] L. Deng, J. Ding, Y. Liu and C. Wei, “Regression analysis for the proportional hazards model with parameter constraints under case-cohort design,” *Computational Statistics & Data Analysis*, **117**, pp. 194–206, 2018.
- [9] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2006.
- [10] N. Breslow, “Covariance analysis of censored survival data,” *Biometrics*, **30**, pp. 89–99, 1974.
- [11] T. Iwata, K. Saito and T. Yamada, “Recommendation method for improving customer lifetime value,” *IEEE Transactions on Knowledge and Data Engineering*, **20**, pp. 1254–1263, 2008.
- [12] N. Mantel, “Evaluation of survival data and two new rank order statistics arising in its consideration,” *Cancer Chemotherapy Reports*, **50**, pp. 163–170, 1966.
- [13] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics Bulletin*, **1**, pp. 80–83, 1945.
- [14] 山口景子, “頻度の時間変化を考慮した階層ベイズモデルによるウェブサイト訪問行動の分析,” *マーケティング・サイエンス*, **22**, pp. 13–29, 2014.
- [15] 猪狩良介, 星野崇宏, “Online–offline チャネルにおける消費者の購買間隔と購買金額の同時モデリング,” *オペレーションズ・リサーチ: 経営の科学*, **61**(9), pp. 589–599, 2016.