

存続時間分析による美容院顧客の来店確率予測

小西 葉子[†]

(受付 2006年2月10日;改訂 2006年6月14日)

要 旨

本稿では、存続時間分析を利用して美容院顧客の来店行動を特定化し、来店確率の予測を行う。具体的には、ヘアカット、ヘアカラー、パーマ・ウェーブの各サービスを受ける顧客の来店意欲(intensity)をCoxモデルで推定し、次回来店者数の予測を行った。結果、カットの来店確率は個人の属性や髪に対する選好の差を考慮したCoxモデルで予測でき、カラーとパーマに関しては、来店確率の予測に個人の属性や質の考慮が必要なかった。来店確率を一定精度で予測できたことは、誰が、いつ、何を施術するのかという美容院の経営上最も重要な情報が得られることを意味する。これはマイクロベースの売上予測、薬剤の在庫管理、スタッフの人員配置などに利用可能である。応用例として、存続時間分析でもマーケティング分析でも本質的な問題である、打ち切りデータや店離れた客の判断も統計的根拠を持って行えることを示した。またダイレクトメールを送付する際は、ヘアカラーの割引券が売上増加に効果的である。

キーワード: Coxモデル, 美容院, 来店確率の予測。

1. はじめに

昨今、景気の低迷や情報技術の革新がもたらす消費者の嗜好の多様化により、企業競争は激化している。企業が成長を伴い存続するには新規顧客の開拓もさることながら、既存顧客を離反させず、企業へのロイヤリティ(忠誠・信頼度)や顧客価値を高めていくことが重要である。多くの産業で、各企業固有の特徴や問題を考慮した上で、マイクロマーケティングや個人に対応したone-to-oneマーケティングの導入、実施が進められている。山口他(2004)では、レストランの日々の売上高を、立地条件など店舗固有の情報により要因分解するモデルを提案している。矢田他(2000)では、ドラッグストアで新製品をプロモーションする際に、顧客の属性、購買履歴からダイレクトメールを送付する対象を選別し、その効果を検証している。

本稿では、美容院の各顧客の来店履歴データを用い、存続時間分析を利用して顧客の来店行動を特定化し、来店確率の予測を行う。具体的には、まずヘアカット(以下カット)、ヘアカラー(以下カラー)、パーマ・ウェーブ(以下パーマ)の各サービスを受ける顧客の来店意欲(intensity)をCoxモデルで推定する。被説明変数は、来店間隔(非来店期間)である。説明変数は年齢、性別、自宅から当該美容院までの距離といった属性変数に加え、顧客の髪に関する嗜好や当該美容院に対するロイヤリティや帰属度を表す変数として、専属担当者の有無や、前回何を施術したか、総使用金額も利用する。また、推定に利用する顧客のセグメンテーションはHughes(1994)のRFM(Recency, Frequency, and Monetary value)法を部分的に採用している。

[†] 一橋大学 経済研究所: 〒186-8603 東京都国立市中 2-1; konishi@ier.hit-u.ac.jp

来店確率の予測は、次回誰が、いつ、何を施術しに来店するかという美容院経営上最も重要な情報を教えてくれる。いつ、何名が、何を施術しに来店するかわかることは、1. 個人の来店行動をベースにした売上高予測、2. 薬剤の在庫管理、3. モデル内の担当者に関する情報より、スタッフの人員配置といった経営戦略に利用可能である。本稿では、この来店確率の予測を一定精度で行うことを目指す。

また美容院では、通常、割引券として使えるダイレクトメールの送付が顧客の再来店を促し、失客を防ぐ手段として利用されている。しかしながらその多くが、全ての顧客に一斉に同じ内容のダイレクトメールを送付するマスマーケティング型である。効果的なダイレクトメール送付を実現するためには、Jonker et al. (2004)、Tsai and Chiu (2004)で述べられているように送付対象となる顧客のセグメンテーションが重要である。さらに美容院に限定すると、送付のタイミングが非常に本質的な問題となる。例えば顧客がそろそろ髪を切りたくなり、来店意欲が高まっている時に送付できれば、来店周期の遅延防止や他店への流出を防ぐことが期待できる。本稿の存続時間分析に基づく来店確率の予測ができれば、各個人に対して次回来店のタイミングに合わせてダイレクトメールを送付するマーケティング戦略が行え、その予測精度が高いほど高い効果が期待される。よって経営戦略、マーケティング戦略の観点からも、来店確率の予測を行うことは大変意義がある。

本稿の構成は以下の通りである。次節の予備的分析では、当該美容院で得られるデータの特徴について簡単に触れた後、本稿の分析対象となる顧客の選定とそのセグメンテーション法の妥当性について検証する。第3節では、Coxモデルの説明と推定に使用するデータ及び推定結果について述べている。第4節では、推定結果に基づき、顧客の来店確率の予測を行い、ダイレクトマーケティングへの応用の可能性を示す。第5節では結論を述べている。

2. 予備的分析

第2.1節で対象となる美容院で得られるデータの紹介、第2.2節では、第3節以降の分析対象となる顧客を抽出して、セグメントに分類し、第2.3節でセグメンテーションの頑健性と識別性について検証する。

2.1 データについて

当該美容院は、2003年7月にオープンし、ビジネス街と各種商業施設に隣接し、交通の便も良い場所に立地している。分析期間は、オープン日から2004年9月までの416日間であり、期間中に来店した4639名の顧客の来店履歴を利用する。通常美容院では、初回来店時に、氏名、性別、年齢、住所、職業、髪に対する悩みや希望、趣味やその他の嗜好を尋ねるアンケートを行う。また、全ての客の精算情報を管理しており、来店日付、担当者、施術したサービスの種類、利用した割引率、支払い金額などの情報が日々蓄積されている。本稿ではこれらのデータの中から、性別や年齢などの属性変数と精算データを用いて分析する。当該美容院では、髪への施術以外にも、シャンプーやトリートメントの販売、メイクやネイルケアなど様々なサービスを提供している。施術数の割合は、42%がカット、22%がカラー、14%がパーマであり全体の約80%を占めている。よって、特にカット、カラー、パーマが当該美容院のサービスの中心と考え分析を行う。カットの平均客単価は4200円、カラーは6300円、パーマは6700円である。当該美容院は顧客の90%が女性であり、年齢は、最少年齢が12歳、最高齢が78歳で、平均は27歳である。距離は、アンケートに記入された住所と当該美容院までの距離で、平均が15km、メディアン5kmで、いずれにしても徒歩圏内とは言い難く、通勤先、通学先から通う顧客が多いことが推察される。

2.2 分析対象となる顧客の抽出

効果的なマーケティングの実施には、企業にとって価値のある顧客を抽出するセグメンテーション(分類)が必要不可欠である。Korgaonkar et al.(1985)と Akaah et al.(1995)では、顧客の企業への関心の高低がダイレクトマーケティングへの反応の差として現れると述べている。ここでは、第3節以降の分析に用いる顧客を抽出し、セグメントに分類する方法を述べる。まず前述の様に、当該美容院では、メディアン距離が5km、平均距離が15kmであり、徒歩で通っているとは考えにくく、大半が徒歩圏外に居住していることになる。実際、顧客の居住地域をプロットすると、ある特定の鉄道沿線に集中していることがわかった。よって、鉄道を利用して来店することを考慮して、推定には店舗からの距離が30km以内に居住する顧客を抽出する。なお、除去したサンプルの比率は5.5%である。

次に、本稿ではカット、カラー、パーマを施術した顧客の来店間隔(非来店期間)を分析対象として抽出する。その際サービス毎にモデルを推定する。もし、サービスで分類せずに、全ての来店行動を一つのモデルで推定すると、前回カットに来た客が、次回カットで来店する前にカラーやパーマの施術で来店した場合、この期間が何を意味するのか非常にわかりにくい。また、カットの来店間隔は髪が伸びる速度に依存し、ある程度規則的な周期があると予想できる。一方、価格もより高く、髪に対する個人の嗜好が反映されるカラーやパーマの来店意欲は、個人間で差があると考えられる。

また1回のみ来店した顧客をサンプルから除く。これは、オープン日の付近に来店し、その後来店しなかった客(非来店期間が非常に長い)と、観察終了時付近に来店した客(非来店期間が非常に短い)を除くことを目的としている。少なくとも2回以上来店した顧客が対象となる。実際、各サービスの来店間隔に違いがあるのか、図1で来店間隔の密度関数を推定した。カットは実線、カラーは破線、パーマは点線である。バンド幅選択は Sheather and Jones(1991)の方法を用いている。密度関数のピークは、カット、カラー、パーマの順に正の方向にシフトしている。また分散、尖度もカット、カラー、パーマの順で大きくなっており、その形状は異なる。

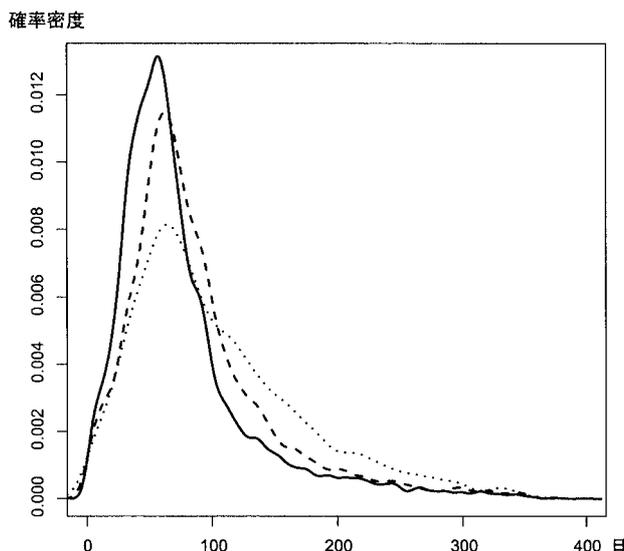


図1. 来店間隔の密度関数。カット: 実線, サンプル数 7909, バンド幅 4.5。カラー: 破線, サンプル数 3911, バンド幅 6.1。パーマ: 点線, サンプル数 1693, バンド幅 10.1。

る。これは各サービスで来店間隔に差があることを意味し、サービス毎の分析が望ましい。

さらに、当該美容院は、初回に30-35%、4回目までは20%、5回目以降は10%割引率を提供し、来店意欲を高める努力をしている。美容院は、5回以上来店すると定着した顧客で、ロイヤリティをもつ優良顧客とみなす。そこで、本稿では4回以下と5回以上でサンプルを二分し、それぞれ一般顧客と優良顧客と呼ぶことにする。ここでの回数は、サービスを限定せず全体で何回来たかである。

2.3 セグメンテーションの評価

前小節のセグメンテーション法を要約すると、1. 30 km 以遠に居住する顧客と1回のみ来店した顧客を除き、2. カット、カラー、パーマの施術毎にサンプルを分け、3. 来店回数で一般顧客と優良顧客を分類した。1, 2については前小節で分類の理由を述べた。3に関しては美容院業界の慣例や経営者の感覚に依るもので、科学的な根拠がない。そこで、来店回数が4回以下を一般顧客、5回以上を優良顧客と分類し、異なるセグメントと扱う妥当性を検討する。

Wedel and Kamakura(1998)では、セグメンテーションと分類されたセグメントが満たすべき様々な条件が示されている。ここでは、本稿のセグメンテーションと深く関連する、以下2つに焦点を当て検証していく。第一に頑健性が挙げられる。ある基準(本稿では来店回数)で分類された各セグメント内は homogeneity(均質)で、その性質は時間を通じて安定的であること、第二に、分類されたセグメント間では、違いが区別できる識別性があることである。これは、マーケティングのみならず統計解析においても非常に本質的な問題である。

そこで、一般顧客、優良顧客のセグメントが時間を通じて安定的か否かを調べるため、サンプル(416日間)を前半と後半に二分し各々の来店間隔の密度関数を推定した。推定方法は図1と同じである。図2の上から順にカット、カラー、パーマで、左から1列目が一般顧客、2列目が優良顧客の密度関数である。実線が各々の前半の密度関数、破線が後半の密度関数である。カット、カラー、パーマとも前半、後半で密度関数の形状は非常に似通っており、時間を通じてセグメント内の顧客構成が安定的であると言える。これは全期間のデータを用いて分析することの正当性の根拠となる。

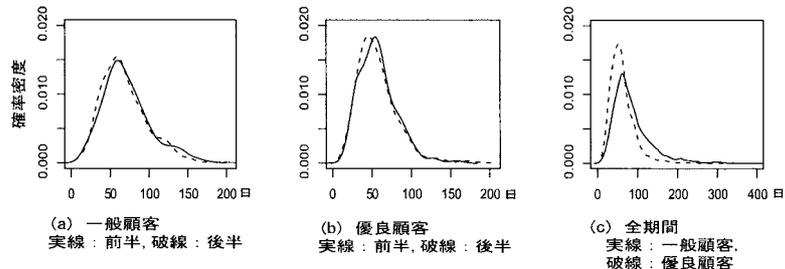
次に一般顧客と優良顧客の間の異質性についてみる。上記で時間に対する頑健性を確認したので、各サービスの全期間データを用いた密度関数を推定した。図2の左から3列目は実線が一般顧客、破線が優良顧客である。特筆すべきはカットの推定結果であろう。優良顧客の来店間隔の期待値は明らかに一般顧客よりも期間が短くまたその分散が小さい。裾の部分の形状も異なっている。比較するとカラーとパーマの密度関数は形状に大きな違いがないように見えるが、実際の相違を判断するには十分ではない。

そこで、数量的に判断するために、カプランマイヤー(以下 KM)法で非来店確率関数の推定を行い、一般、優良顧客で関数型に統計的に有意な差があるかをログランク検定で確認する。美容院への来店(イベント)を時間 T とすると、 T は非来店期間を表す確率変数で、その分布関数は $S(t) = \Pr\{T \geq t\}$ で表せる。 $S(t)$ を KM 法で推定すると、観察開始時には推定値 $\widehat{S}(t)_{KM} = 1$ で、その後顧客が来店する度、センサーされる度に非来店確率 $\widehat{S}(t)_{KM}$ が減少していく右下がりの階段関数となる。一般的には、 $t_i < t \leq t_{i+1}$ である $S(t)_{KM}$ の推定値は、以下の様になる。

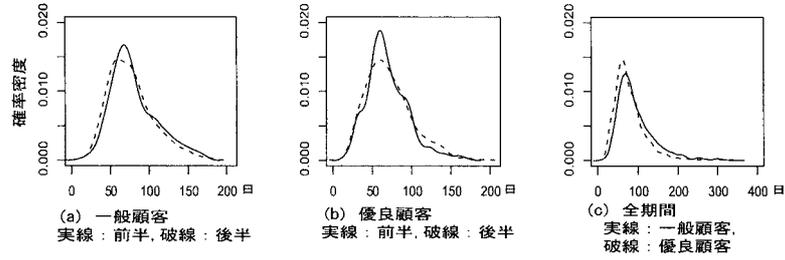
$$(2.1) \quad \widehat{S}(t)_{KM} = \widehat{S}(t_i)_{KM} \left(1 - \frac{R_i}{n_i}\right).$$

$\widehat{S}(t_i)_{KM}$ は $t > t_{i-1}$ での来店率の推定値で、 R_i は時間 t_i において来店した顧客数、 n_i は t_i で未だ非来店の顧客数である。図3はカット、カラー、パーマの非来店確率関数で、実線が一般顧客、破線が優良顧客である。視覚的にも一般顧客と優良顧客は異なり、全てのサービスで優良顧客が早く非来店確率が0に近づいている。またそれぞれが交差しておらず、比例性を満

(1) カット



(2) カラー



(3) パーマ

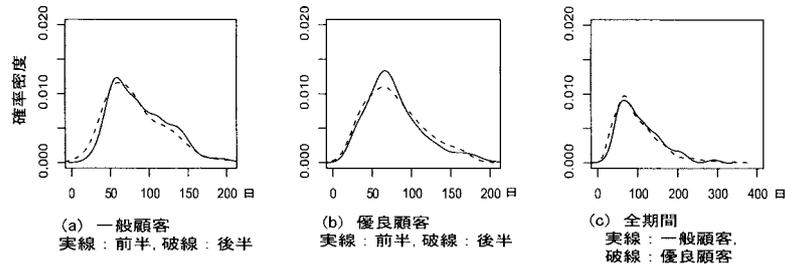


図 2. 来店間隔の密度関数・セグメントの頑健性と識別性の検証。

(1) カット

(2) カラー

(3) パーマ

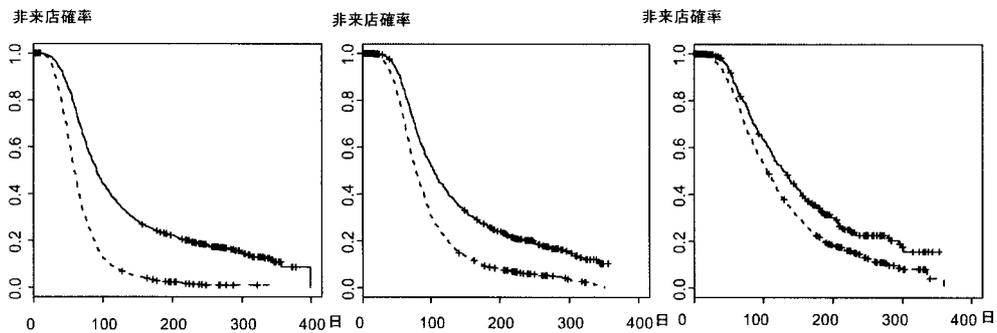


図 3. KM 推定による非来店確率関数(実線：一般顧客, 破線：優良顧客)。

たしているように見えるため、次節での Cox 比例ハザードモデルへの特定化の根拠の一つとなろう。この推定結果を用いてログランク検定を行う。ログランク検定の詳細は、Kalbfleisch and Prentice (1980) を参照されたい。帰無仮説の下では、二つのセグメントの非来店確率関数が等しいと仮定する。具体的には、観測期間を通じて、一般顧客と優良顧客の非来店確率の期待値が等しいとし、得られた検定統計量は自由度 1 のカイ二乗分布に従う。結果は、カット、カラー、パーマの統計量は順に 1001, 189, 24 でいずれも帰無仮説を棄却し、一般、優良顧客の関数型は有意に異なる。特にカットはその統計量が大きいため、設定した帰無仮説が大きく現実と異なることを示唆する。以上の結果より、経験則に基づいて、美容院で慣例的に行われている顧客のロイヤリティに関する評価があなたが間違っていないことは興味深い。また最善ではないものの、サービスの分類、ロイヤリティでの分類を行うことで各セグメントの均質性もある程度確保されたと考えたい。セグメント内の顧客の質の差は、以降の分析において属性等でコントロールしていく。

以上より、次節の実証分析では、カット、カラー、パーマについて、全期間のデータを用いて、来店回数によって分類した一般、優良顧客の各セグメントについて推定を行う。

3. 推定

ここでは、各サービスについて顧客の来店意欲の強度関数を Cox モデルで特定化し、推定を行う。次小節では推定モデルの説明、第 3.2 節では被説明変数と説明変数の説明、第 3.3 節では推定結果を示す。

3.1 推定モデル

Cox (1972) の Cox 比例ハザードモデルは、古くから医療や薬学分野において、治療や薬効が患者の生存率に与える影響をみるために利用されてきた。経済学においても企業の倒産や債券のデフォルト確率の推定に応用されている。本稿では、顧客の非来店期間を対象とし、再来店をイベントの生起と捉えるため、生存率関数のことを(非)来店確率関数、ハザード関数のことを強度関数(intensity 関数)と呼ぶこととする。前節と同様に、前回来店時から次に来店するまでの非来店期間を確率変数 T とし、その分布関数を非来店確率関数 $S(t)$ と表す。強度関数 $\lambda(t)$ を特定化すると、 $S(t)$ が特定化される。強度関数は時点 t までに来店イベントが生起していないという条件の下で、次のごく短い期間 $t + \Delta$ 後に来店イベントが生起する確率を表す。個人の強度関数と非来店確率関数は以下の式で表せる。

$$(3.1) \quad \lambda_i(t|x) = \lambda_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi}), \quad (i = 1, \dots, k).$$

$$(3.2) \quad S_i(t|x)_{\text{Cox}} = S_0(t)^{\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi})}.$$

$\beta = (\beta_1, \dots, \beta_m)'$ は未知パラメータ、 $x = (x_1, \dots, x_m)'$ は説明変数である。 $\lambda_0(t)$ は基準となる顧客の強度を示すベースライン強度関数であり、 $S_0(t)$ は基準顧客の非来店確率関数である。式(3.1)より、個人の強度関数は、 $\lambda_0(t)$ と時間に依存しない比例定数 $\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi})$ で表現され、部分尤度関数は(3.3)式で与えられる。 R は観察された来店者数、来店の時間は $t_1, \dots, t_i, \dots, t_R$ とし、 $K(t_i) = \{j: t_j \geq t_i\}$ は t_i の直前まで来店せずに、次の $t_j = t_i + \Delta$ に来店する顧客数である。ここでは、簡単化のために、各個人の来店時間は全て異なるとする。

$$(3.3) \quad L(\beta) = \prod_{i=1}^R \frac{\exp(\beta' x_i)}{\sum_{j \in K(t_i)} \exp(\beta' x_j)}.$$

推定では、Newton-Raphson 法などによって以下の対数尤度関数の最大値を見つけ推定値 $\hat{\beta}$

表 1. 推定に用いた説明変数一覧.

変数名	種類	説明
性別	属性変数	ダミー変数 男=0, 女=1
年齢	属性変数	平均値 27 歳, メディアン 25 歳
距離	属性変数	当該美容院までの距離 (単位: Km) 平均値 15km, メディアン 5km
割引率	美容院のサービスに関する変数	前回何 % 割引されたか
指名	美容室へのロイヤリティを表す変数	ダミー変数, 担当者を指名しない=0, 指名する=1
金額	美容室へのロイヤリティを表す変数	観察最終日までの使用金額 (単位: 千円)
前回カットのみ施術	個人の髪型, 嗜好に関する変数	カットの回帰に使用
前回カラーのみ施術	個人の髪型, 嗜好に関する変数	カラーの回帰に使用
前回パーマのみ施術	個人の髪型, 嗜好に関する変数	パーマの回帰に使用
前回カットとカラー施術	個人の髪型, 嗜好に関する変数	カットとカラーの回帰に使用
前回カットとパーマ施術	個人の髪型, 嗜好に関する変数	カットとパーマの回帰に使用
前回カラーとパーマ施術	個人の髪型, 嗜好に関する変数	カラーとパーマの回帰に使用
前回全て施術	個人の髪型, 嗜好に関する変数	全ての回帰に使用

を得る.

$$(3.4) \quad LL(\beta) = \sum_{i=1}^R \beta' x_i - \sum_{i=1}^R \log \left[\sum_{j \in K(t_i)} \exp(\beta' x_j) \right].$$

以上の定式化はイベントが同時に発生するタイデータが無い場合の説明である. 本稿では美容院顧客の来店行動が分析対象なのでタイデータが存在する. そのため尤度関数は発生順序の組み合わせの分だけ (3.3) 式よりも複雑になる. 本稿の推定結果はタイデータを考慮した, Breslow (1974) の近似法によるものである.

3.2 推定で用いるデータについて

被説明変数, 説明変数を紹介する. 顧客のセグメンテーションに関しては前節の説明の通りである. 被説明変数は, 前回来店日から次の来店日までの非来店期間である. 来店するとイベント発生とし, 期間中来店しない場合はセンサリングデータとして扱う. 観察終了日が 416 日であるため, その日がセンサリング(打ち切り)時点になるタイプ 1 センサーである. 例えば期間中 3 回来店した顧客は, 非来店期間を 3 つ持っており, 来店というイベントが 2 回と, 3 回目に来店した日から観察終了日までの非来店時間が打ち切りデータになる. 説明変数は, 表 1 で示す. 各モデルに含まれる説明変数は 9 個で, 性別, 年齢, 自宅から当該美容院までの距離, 前回の割引率, 担当者の指名の有無に関するダミー変数, 総使用金額, 前回受けたサービスに関するダミー変数 (3 種類) である. 総使用金額は観察終了時まで使用した総額である. 顧客の価値を正確に評価するために, 観察終了日までに使用した金額の総額を用いた. 平均を使わない理由は, 金額が低いサービスを多く利用する顧客の過小評価を避けるためである. 指名変数は予約時に専属担当者を指名するか否かのダミー変数であり, 美容院に対するロイヤリティや帰属度を表す代理変数と想定する.

3.3 推定結果

表 2 は一般顧客のカット, カラー, パーマの Cox 回帰モデルの推定結果である. 変数選択には最小の AIC を達成するまで, 必要, 不必要な変数を逐次的に抜き差しするステップワイス変数選択法を用いている. 表 2, 3 中の添字 a は, 当該変数が 5% 水準で有意であることを示し, b は当該変数がステップワイス変数選択法で選択されなかったことを意味する. また空欄はも

ともモデルに含まれない変数である(表1参照). カットのCoxモデルでは全変数が選択され有意であった. 年齢の係数は正値で0.006である. 例えば年齢20歳の顧客と21歳の顧客を比較するとCoxモデルの比例強度性の仮定より, $\exp(0.006 \times 21) / \exp(0.006 \times 20) = 1.006$ となり, 年齢が1歳上がると1.006倍, つまり0.6%来店意欲が高まり, 10歳高くなると6.1%強度(intensity)が上昇する. 同様に, 家からの距離, 前回の割引率, 指名の有無, 総使用金額が正値であった. 金額に関する変数についてみると, 前回割引率は, 割引を利用しなかった顧客と比較して10%割引を受けた場合, 19.7%来店意欲が高まり, また, 今まで美容院で支払った金額が1単位(1000円)高くなると, 3.8%来店意欲が高まる. 来店意欲の上昇は, 再来店までの期間が短縮することを意味する. 注意が必要なのは距離に関する係数で, 係数が0.728で, 居住地が美容院から1km離れると来店意欲が2.07倍以上, 2km, 3kmと離れると2.07の乗数倍で強度が高まっていく. 遠くに住む顧客ほど頻繁に来店することを意味し, 通常の距離と来店行動の関係とは異なる. これは前述の様に, 顧客がアンケートに記入した住所以外に, 通勤先や通学先なども一つの住所を持ち, 距離のデータが機能していないことを示唆する. 実際, 今回の分析では欠損値が多いため利用していないが, 職業について答えた顧客の62%が会社員, 19%が専門学校生と大学生であった. 距離の効果を正確に測るためには, アンケートに通勤, 通学先や美容院に通う際の最寄り駅などを記入してもらう工夫が必要であろう.

次に, 推定した係数が負値となったものは, 性別, 指名の有無を示すダミー変数, 前回に施術したサービスに関する3つのダミー変数であった. 性別は, 男性=0, 女性=1とするダミー変数なので, 男性の方が女性よりも来店間隔が短い. これは男性には短髪が多く, 髪全体の長さに対して1cm伸びた際の見た目のマージナルな変化が大きく, 髪型が崩れやすいことに関係すると考えられる. 指名変数は, 担当する美容師を指名しない場合=0, 指名する場合=1で, 顧客の美容院への帰属度やロイヤリティの代理変数の一つであると考え, 推定値は正値であると予想していた. しかし係数は負値であり, 特定の担当者を希望しない顧客の方が来店間隔が短いことになる. 専属担当者への指名は, 顧客が担当者の技術に満足し, また美容師も施術中の会話を通して顧客の髪型への嗜好や髪質も理解でき, それがより良いサービスの提供に繋がることが期待される. つまり顧客の髪質を捉えた質の高い施術によって, 型崩れするまでの期間と顧客の満足期間が長くなることにより, 来店間隔が長くなっている可能性がある. 指名変数は顧客のロイヤリティを表しているというよりは, 前回の施術への顧客の満足度と個々の美容師の技術の差を表す代理変数と捉える方が妥当かもしれない.

特筆すべきは, 前回の施術の種類を表すダミー変数の係数である. カラーとカットを行った客, カットとパーマを行った客, カット, カラー, パーマ全てを施術した客とも係数が負であった. これは, カットのみを施術した客と比較して, カラーやパーマを組み合わせた客の方が非来店期間が長くなることを意味する.

次に, カラーのモデルで有意になった変数は, 割引率, 指名ダミー, 使用金額, 前回全てを施術した顧客に関するダミー変数であった. パーマのモデルで有意になった変数は, 割引率, 指名ダミー, 使用金額, 前回カットとパーマを施術した顧客に関するダミー変数である. カットのモデルと比較して, 両モデルは来店確率に対して割引率の効果が大きい. また指名効果がカットのモデルと同様に負値であった. 使用金額についても, パーマのモデルが若干係数が小さいものの, カットのモデルと同様に正値で有意である. カラーのモデルでは前回3種類全てのサービスを行った客の来店間隔が長くなり, パーマのモデルでは, カットと組み合わせた客の来店意欲は組み合わせなかった顧客よりも約40%高かった.

表3は優良顧客に関する推定結果である. カットに関しては, 距離と指名ダミーが有意でなくなった以外は一般顧客と符号が同じである. 距離についてはカラー, パーマのモデルでも有意でない. カラーについては, 年齢, 割引率, 金額のみ有意であった. パーマは, 年齢, 割引

表 2. 一般顧客の Cox 回帰モデルの推定結果($n \leq 4$).

	カット		カラー		パーマ	
	$\hat{\beta}$	$exp(\hat{\beta})$	$\hat{\beta}$	$exp(\hat{\beta})$	$\hat{\beta}$	$exp(\hat{\beta})$
年齢	0.006 ^a	1.006	0.009	1.009	0.014	1.015
性別	-0.207 ^a	0.813	— ^b	—	-0.489	0.614
距離	0.728 ^a	2.070	—	—	—	—
割引率	0.018 ^a	1.018	0.028 ^a	1.028	0.030 ^a	1.030
指名	-0.523 ^a	0.593	-0.393 ^a	0.675	-0.407 ^a	0.666
金額	0.037 ^a	1.038	0.039 ^a	1.039	0.025 ^a	1.025
カットとカラー	-0.527 ^a	0.590	-0.197	0.821	—	—
カットとパーマ	-0.580 ^a	0.568	—	—	0.355 ^a	1.426
カラーとパーマ	—	—	—	—	—	—
全て	-0.999 ^a	0.368	-0.4579 ^a	0.633	—	—
サンプル数	3119		1279		479	
決定係数	0.114		0.120		0.104	
AIC	30818		10430		3159	

^a 5% 有意水準で帰無仮説 $\hat{\beta} = 0$ を棄却.

^b AIC 変数選択によって選択されなかった変数.

注 空欄は最初からモデルに含まれない変数.

表 3. 優良顧客の Cox 回帰モデルの推定結果($n \geq 5$).

	カット		カラー		パーマ	
	$\hat{\beta}$	$exp(\hat{\beta})$	$\hat{\beta}$	$exp(\hat{\beta})$	$\hat{\beta}$	$exp(\hat{\beta})$
年齢	0.0135 ^a	1.014	0.021 ^a	1.020	0.016 ^a	1.016
性別	-0.561 ^a	0.571	—	—	-0.370	0.691
距離	— ^b	—	—	—	—	—
割引率	0.007 ^a	1.007	0.014 ^a	1.010	0.011 ^a	1.011
指名	—	—	—	—	-0.310 ^a	0.733
金額	0.011 ^a	1.011	0.014 ^a	1.010	0.011 ^a	1.011
カットとカラー	-0.371 ^a	0.690	—	—	—	—
カットとパーマ	-0.330 ^a	0.717	—	—	0.225 ^a	1.252
カラーとパーマ	—	—	—	—	—	—
全て	-0.465 ^a	0.628	0.116	1.120	—	—
サンプル数	3408		1954		878	
決定係数	0.105		0.136		0.107	
AIC	45919		22434		7995	

^a 5% 有意水準で帰無仮説 $\hat{\beta} = 0$ を棄却.

^b AIC 変数選択によって選択されなかった変数.

注 空欄は最初からモデルに含まれない変数.

率, 指名, 金額, カットとパーマダミー変数が有意であった. カラーとパーマのモデルで, 一般顧客では非有意だった年齢が有意になっている. またカットのモデルでも年齢が来店意欲に与える効果が高くなっている. 年齢が高い顧客が髪に関心が高いというよりは, 割引率が低くなった状態で美容院に頻繁に通うのは費用がかさむため, 年齢が所得効果の一部を表していると考えられる. そのため, カットよりも単価が高いカラーとパーマでも有意になったと考えられる. 一般顧客との結果と比較すると, 距離が有意でなくなり, 割引率や使用金額の違いが来店意欲に与える影響は低くなる. また指名の有無がカットとカラーで有意でなくなった. 前述の様に, この変数が顧客の満足度や美容師の技術の差を表すならば, あり, なしといった二分

の情報では不十分であろう。顧客は満足が得られなければ担当を変更するので、誰が担当したかで顧客の来店間隔が異なると仮定する方がより現実を反映していると考えられる。

以上より、カット、カラー、パーマのモデルはそれぞれ異なり、各サービスで推定することが望ましい。同様に一般顧客と優良顧客についても違いがみられるため、来店回数での顧客のセグメンテーションには機能している。但し、推定結果全体を通じて、決定係数が十分高くないことに留意する必要がある。特にカットの一般顧客では、採用した全ての変数の係数が有意にも関わらず、決定係数が高くない。これは顧客の行動を特定化するのに未だ不足している変数があること、Cox モデルが十分ではない可能性を示唆している。

4. 予測

本節では、前節の推定結果に基づき、次小節では来店確率の予測、第 4.2 節では予測結果の考察を行い、存続時間分析のダイレクトマーケティングへの応用例を紹介し分析のまとめとしている。

4.1 来店確率の予測

前節の推定結果を用いて顧客の来店確率の予測を行う。前述のように、マーケティングでは、セグメンテーションの方法として Hughes (1994) の RFM (Recency, Frequency, and Monetary value) 法がある。これは、当該美容院に関心が高く、美容院にとって価値の高い顧客を、最近来店した顧客 (R), 来店頻度が高い顧客 (F), 使用金額が高い顧客 (M) であると考え、3 つの指標で顧客の価値を評価し分類する。RFM によるセグメンテーション法の有効性や限界、代替する方法を詳しく述べた論文に Fader et al. (2005) がある。既に、推定で来店回数 (F) を用い、顧客を分類し、総使用金額 (M) を説明変数に採用している。ここで、観察終了時から遡って直近の (R) 40 日間 (2004 年 8 月から 2004 年 9 月 10 日) に来店した顧客が次の 100 日以内に来店する確率を予測する。100 日間にした理由は他のサービスと比較して来店間隔が長いパーマのメディアンが 100 日強だからである (3.2) 式の非来店確率 $S_i(t|x)_{\text{Cox}}$ は (2.1) 式の KM 法をベースに Cox 回帰分析で得られた来店強度の情報を用いて計算する。具体的には、KM 法では、 $t_i < t \leq t_{i+1}$ の時、 t_i の直前の非来店率に $(1 - \frac{R_i}{n_i})$ を掛けて、 $t > t_i$ の非来店率を計算した。 $S_i(t|x)_{\text{Cox}}$ では、基準顧客の非来店確率関数 $S_0(t)$ を得る際、 $(1 - \frac{R_i}{n_i})$ の代わりに、 $(1 - \frac{R_i^*}{n_i^*})^{(1/R_i^*)}$ を用いる。 R_i^* は時点 t_i で来店した顧客の強度の大きさで、 $R_i^* = \exp(\beta' x_i)$ であり、 n_i^* は、 $n_i^* = \sum K_k = \sum \exp(\beta' x_k)$ で時点 t_i で未だ来店していない顧客の強度の大きさの合計を表す。よって基本顧客の非来店率は $\widehat{S_0}(t) = \widehat{S_0}(t_i) (1 - \frac{R_i^*}{n_i^*})^{(1/R_i^*)}$ となり (3.2) 式より、各個人の非来店率の予測値は以下の式で計算できる。

$$(4.1) \quad \widehat{S_i}(t|x)_{\text{Cox}} = \widehat{S_0}(t) \exp(\widehat{\beta}' x_i) = \left(\widehat{S_0}(t_i) \left(1 - \frac{\widehat{R}_i^*}{\widehat{n}_i^*} \right)^{(1/\widehat{R}_i^*)} \right)^{\exp(\widehat{\beta}' x_i)}$$

今、サンプルから観測終了期間直近の 40 日間に来店した l 人に対して、来店すれば 1、来店しなければ 0 の二項確率を考え (式 4.2)), 式 (4.1) で各個人の $t > 100$ となる来店率を計算すると、全体の来店確率が式 (4.3) で得られる。推定値 $\widehat{S}(t)$ は、漸近的に平均 $S(t)$ 、標準偏差 $SE(t)$ の正規分布に従うことより、予測の 95% 信頼区間を求めた。

$$(4.2) \quad E\{1_{\{T_i \leq t_i\}}\} = P\{1_{\{T_i \leq t_i\}}\} = 1 - S_i(t|x)_{\text{Cox}}$$

$$(4.3) \quad \text{来店確率} = \frac{1}{l} \sum_{i=1}^l \left(1 - \widehat{S_0}(t_i) \exp(\widehat{\beta}' x_i) \right)$$

$$(4.4) \quad \text{標準誤差} = \widehat{SE}(t) \frac{1}{\sqrt{t}} \sqrt{\left(1 - \widehat{S}_0(t_i)^{\exp(\hat{\beta}'x_i)}\right) \left(\widehat{S}_0(t_i)^{\exp(\hat{\beta}'x_i)}\right)}$$

結果、表 4 に示すようにカットのみ、真の来店率が Cox モデルに基づく予測で得られた信頼区間に入った。カラーとパーマは個人の属性や質をモデルに含めない KM 法で得た予測値の信頼区間に真の来店率が入るといった結果になった。KM 法では $t > 100$ の来店率の予測値を(2.1)式で求め、標準誤差を $\widehat{S}(t_j) \left[\sum_i \frac{R_i}{n_i(n_i - R_i)} \right]^{1/2}$ で計算し、Cox 回帰の場合と同様に信頼区間を求める。カットの一般顧客を例にすると 100 日間に来店する人数は、95%水準で(261 人, 295 人)の間に真の来店率が入ると予測される。実際真の来店人数は 262 人でこの区間に入るため、一定精度での来店確率の予測ができた。同様に、カットの優良顧客は Cox モデルで、カラーとパーマは KM 法で、来店確率の予測ができた。カラーとパーマの結果が属性変数を考慮しない KM 法を支持したことに留意しよう。特にカラーに関しては、一般、優良顧客も分ける必要がないという結果になった。図 4 はカラーの全顧客の KM 法と Cox 回帰によって得られる基準顧客の非来店確率関数 $\widehat{S}_0(t)$ である。実線が KM 推定、破線が Cox 回帰の結果である(2.1)式と $\widehat{S}_0(t)$ は属性変数で表現される個別の強度の重みの分だけ差が出るはずである。しかし、その影響はみてとれず、二つの曲線はほぼ等しい。Cox 回帰に基づく予測では、 $\widehat{S}_0(t)$ を用い、最終的には(4.1)式で個々の来店確率を予測するため、KM 法と比較して $\exp(\hat{\beta}'x_i)$ の分だけ上方予測する傾向にあり、信頼区間に真の来店率が入らなかった。推定では有意に係数が 0 でないのに、予測の観点からは結果が支持されない点については今後も検討を要する。

本稿では当初、カットはその周期に個人内、個人間で大きな差がないと考え、またカラーやパーマについては、個人間で髪質や髪型への嗜好が異なりその差が大きいと想定した。しかし結果は、カットの来店確率予測には個人の属性や美容院に対する選好の差を表す変数が必要で、カラー、パーマは、推定においていくつかの説明変数が選択されたが、予測の観点からはそれらの効果を考慮する必要がなかった。日本パーマメントウェーブ液工業組合(2005)によると、髪は 1 日平均 0.4 mm, 1ヶ月で 1.25 cm 伸びるとされるが、実際には季節や年齢、性別、遺伝など個人差が大きく、2 倍以上の個人差があるケースが報告されている。一方、日本ヘアカラー工業組合(2005)によると、カラーの薬剤は通常、効果が 2, 3ヶ月あるが、それよりも早く、3 週間程度で髪全体の色味が明るくなる退色現象が起きる。次いで髪が伸びると根元との色の違いが目立ってくる。パーマの薬剤の効果も 2, 3ヶ月であるが、カラーほど根元が伸びた時の見た目の変化が大きくないため来店間隔が長い。それ故、パーマについては、髪型維持に対する嗜好の差が、一般顧客と優良顧客の来店周期に差を生じさせると考えられる。つまり、カラーやパーマは個人の属性や髪に対する選好の差よりも、薬剤の効果が来店周期に与える影響が大きいと考えられ、特にカラーでは一般、優良顧客を分ける必要もないことからその傾向が強いといえる。

表 4. 予測結果。

サービス	分類	推定モデル	予測来店確率	標準誤差	95%信頼区間	真の来店率 (来店者数/対象者数)
カット	一般	Cox	0.729	0.027	(0.684,0.773)	0.686(262 人/382 人)
カット	優良	Cox	0.815	0.021	(0.774,0.855)	0.823(287 人/356 人)
カラー	全体	KM	0.616	0.026	(0.565,0.667)	0.647(226 人/349 人)
パーマ	一般	KM	0.367	0.063	(0.306,0.491)	0.393(22 人/57 人)
パーマ	優良	KM	0.531	0.052	(0.427,0.634)	0.449(40 人/89 人)

Cox: Cox 回帰モデルの推定に基づく予測
 KM: カプランマイヤー推定に基づく予測

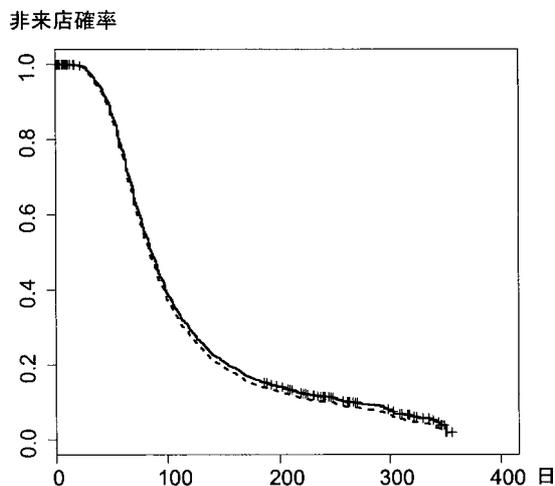


図 4. カラーの非来店確率関数(実線: KM 推定, 破線: Cox 回帰モデルによる推定: $\widehat{S}_0(t)$).

またここでは詳しくは述べないが, 来店行動は再発事象であるため, 再発事象を対象としたモデル(PWP モデル; Prentice et al., 1981)によって同様に推定, 予測を行った. 結果は, カットの優良顧客の一部を除いて, 信頼区間の中に真の来店率が入ることはなかった. よって, 本稿で対象とする美容院顧客の来店行動については, 各来店イベントが個人内でも個人間でも独立であると仮定する Cox モデルや KM 法の採用が改善の余地はあるものの, 妥当であったといえる.

4.2 応用と考察

ここでは, 存続時間分析によって得られた結果をどのように美容院経営, ダイレクトマーケティングに活かせるか考察し, 分析のまとめとする. 前小節で, 来店者数の予測が一定精度で行えた. これは, いつ, 何名が, 何を施術しにくるかを教えてくれる. この情報は, 美容院の売上予測や在庫管理, 人員管理にとって有益な情報となる. 同様に, 各個人についても, いつ, 何を施術しにくるかも知ることができる. これは, 各個人の次回来店タイミングに合わせたダイレクトメール送付というマーケティング戦略に利用可能である.

応用として, 個人の来店確率予測を利用して, 存続時間分析においても, ダイレクトメール送付においても本質的な問題となる, 顧客の離反について考察する. 非来店期間が長い顧客を未だ顧客であるとするか, 店離れ客と判断するかは非常に難しくナイーブな問題である. 大雑把には, 他地域在住者や, 半年, 1 年来店がなければ失客とみなしダイレクトメール送付対象から外すというやり方がある. しかし本稿のように通勤, 通学圏の広い地域の美容院では適切ではなく, 根拠なく 3ヶ月, 半年, 1 年と期間を区切るのも説得力に欠ける. そこで非来店確率関数を用いて, 各顧客について統計的根拠を持った評価を与えてみる. 具体的には, 各個人が来店しない確率が 5%になる日数を予測し, 実際非来店の期間の方が長い顧客は店離れ客と考え, 短ければ未だ顧客であるとしよう. 例えばカットの一般顧客については, 100 日間非来店だった顧客 689 人中, 27%の 184 人を離反客と判断できる. この情報を活かし, ダイレクトメール送付の際, これらの客を除くことでコストダウンもできるし, 失客用のダイレクトメールを送付することで再来店を促す戦略も採れる. 5%の値を 10%, 15%と変えることで, 失客

防止ダイレクトメールの送付も可能である。存続時間分析応用の利点は送付時期や対象を個人の来店行動に基づき統計的根拠をもって決定できることである。

最後に、分析を通じて得られた知見を示す。カットをする顧客は、カットのみの施術が約35%、カラーと同時に施術が約40%、パーマと同時が約18%である。一方、カラーとパーマの顧客は、単独で施術するのは10%以下で、カラーでは約70%、パーマでは約60%以上の顧客がカットを同時に施術する。またカラーとパーマを同時に施術する顧客は1%以下である。これは、髪の損傷と、お互いの薬剤の効き目が薄くなることを避けるため、美容師が同時施術を勧めないことに起因する。前小節の結果より、カラーは施術する顧客も多く、薬剤の持つ周期が来店周期に強く影響を与えていることがわかった。また変動係数を計算すると、3種のサービスの中で最も小さい。よってカットとカラーを同時に施術することは、売上が倍増し、カットの周期を規則的にする効果が期待される。またカラー、パーマを施術する際は、カットを同時に施術する顧客が多いことから、カットの割引券の送付は得策ではない。割引券送付によりカラー、パーマの施術者を増やすことは短期的な売上増加と、その後の来店行動に、より規則性を持たせる効果が期待できる。

5. おわりに

本稿では、カット、カラー、パーマを施術する顧客の来店間隔を用いて、Coxモデルによって来店意欲の強度関数を推定した。次に推定結果に基づき、次の来店確率予測を行った。結果は、カットの来店確率は個人の属性や質を考慮したCoxモデルで予測でき、カラーとパーマに関しては、個人の属性や質の考慮が必要ないことがわかった。来店確率の予測は、次回いつ、何名が、何を施術しに来店するかという美容院の経営上最も重要な情報を与えてくれる。これは、1. 個人の来店行動をベースにした売上高予測、2. 薬剤の在庫管理、3. スタッフの人員配置といった経営戦略に利用可能である。本稿ではこの来店確率の予測を一定精度で行うことができたという点で意義がある。またその他の発見としてダイレクトメールを利用して割引券を送付する際は、ヘアカラーの割引券が売上増加により効果的であることがわかった。

以上のように本稿では、美容院データのマイクロダイレクトマーケティングへの応用の可能性を探り、いくつかの知見を得ることができた。

最後に、推定結果の決定係数が十分高くなく、カラー、パーマではKM推定による予測が支持されたことを記す。美容院顧客の行動に属性が影響しないとは考え難く、この点で必要変数の不足やCoxモデルによる特定化が不十分である可能性は否めない。今後も、美容院顧客の来店行動を表現するのにより適したモデル構築をし、予測精度を高めていくことを課題とする。併せて、個人のタイミングに合わせたダイレクトメール送付が従来の一斉送付と比較して効果的か否か実験、検証するという新たな取組みも行っていきたい。

謝 辞

本稿の執筆にあたり、2004年11月に鹿児島で行われた科研費集会「時空間点過程の推測理論と応用」と同年12月に統計数理研究所で開催された「International Conference on “Recent Development of Statistical Modeling in Marketing—Latent Variable and Latent Structure Approach—」では数々の有益なコメント、アドバイスを頂きました。また、研究開始当初から数多くの議論、提案を頂いた横浜国立大学経済学研究科の永井圭二助教授、京都大学経済研究所の西山慶彦教授に記して感謝致します。また、編集委員、レフリーお二方には建設的なアドバイス、ご質問を頂き、それらを反映することで本稿の内容が大きく改良されました。

最後にこの研究で使用したデータの提供と多くのリクエストを快諾して下さった美容院

オーナー小浦洋氏, 美容業界に知識の浅い筆者に数多くのアドバイスやアイデアをくださった美容院スタッフ板橋恵子氏, 水本まゆこ氏, 宮田弓氏の惜しみない協力を感謝いたします。当然ながらこの原稿の誤謬は全て筆者の責任に帰すものである。

参 考 文 献

- Akaah, I. P., Korgaonkar, P. K. and Lund, D. (1995) Direct marketing attitudes, *Journal of Business Research*, **34**, 211–219.
- Breslow, N. (1974) Covariance analysis of censored survival data, *Biometrics*, **30**, 89–99.
- Cox, D. R. (1972) Regression models and life tables (with discussion), *Journal of the Royal Statistical Society Series B*, **34**, 187–220.
- Fader, P. S., Hardie, Bruce G. S. and Lee, K. L. (2005) RFM and CLV: Using Iso-value curves for customer base analysis, *Journal of Marketing Research*, **XLII**, 415–430.
- Hughes, A. M. (1994) *Strategic Database Marketing: The Master Plan for Starting and Managing a Profitable, Customer-Based Marketing Program*, Probus Pub Co., Chicago.
- Jonker, J. J., Piersma, N. and Van den Poel, D. (2004) Joint optimization of customer segmentation and marketing policy to maximize long-term profitability, *Expert Systems with Applications*, **27**, 159–168.
- Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*, John Wiley and Sons, New York.
- Korgaonkar, P. K., Lund, D. and Price, B. (1985) A structural equations approach toward examination of store attitude and store patronage behavior, *Journal of Retailing*, **61**, 39–60.
- 日本ヘアカラー工業組合 (2005) カラーリング剤について, <http://www.jhcia.org/haircolor/coloring/index.html>.
- 日本パーマメントウェーブ液工業組合 (2005) 毛髪に関する事項, <http://www.perm.or.jp/txt/index.html>.
- Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981) On the regression analysis of multivariate failure time data, *Biometrika*, **68**, 373–379.
- Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society Series B*, **53**, 683–690.
- Tsai, C. Y. and Chiu, C. C. (2004) A purchase-based market segmentation methodology, *Expert Systems with Applications*, **27**, 265–276.
- Wedel, M. and Kamakura, W. A. (1998) *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Boston, Dordrecht, London.
- 矢田勝彦, 加藤直樹, 羽室行信 (2000) 顧客の購買履歴からのデータマイニング, 発見科学とデータマイニング, *bit* 別冊(森下真一, 宮野 悟 編), 169–178.
- 山口 類, 土屋映子, 樋口知之 (2004) 状態空間モデルを用いた飲食店売上の要因分解, オペレーションズ・リサーチ, 2004年5月号, 316–324.

A Duration Analysis of Hair Salon Consumers' Behavior and Prediction of Revisit Rates

Yoko Konishi

Institute of Economic Research, Hitotsubashi University

In this study, we apply duration analysis to specify a model of hair salon customers' behavior regarding their visits and then predict their next revisit rates. We estimate the intensity function by adopting the Cox model for the hair salon data to examine which demographic, geographic, and/or behavioral variables influence each customer's attitude. Our target customers use three hair services; haircut, color, and permanent wave. These customers were divided into non-loyal and loyal customers using the RFM method. Estimation results showed that the intensity functions of three items have been specified by different models, and we found differences in intensity between non-loyal and loyal customers.

As a next step, we apply the Cox model estimation results to calculate the revisit rates for each customer within 100 days since the last visit from the interval prediction of his/her next revisit at 95% confidence interval level. We then compare the predicted rates with the actual rates of the customers' revisits. For haircut, loyal and non-loyal customers' actual revisit rates were within the predicted confidence interval based on the Cox model's results. However, hair color and permanent wave results based on the Cox model cannot predict the revisit rates well, but the predicted results based on the Kaplan-Meier estimation, which do not consider differences among individuals, yielded actual revisit rates within the prescribe confidence intervals. The prediction of revisit rates tell us when and how many customers will visit a salon and which services they will receive. The use of this information can help achieve an effective direct mailing strategy.