

ゴルフECサイトにおけるリピート顧客 の閲覧行動分析

中央大学 理工学研究科
経営システム工学専攻
蘇 悦

目次

- 01 | 背景
- 02 | 目的
- 03 | データ概要
- 04 | 分析
- 05 | 結果と考察
- 06 | まとめと今後の課題

1 背景

- インターネットの急激な普及
 - オンラインショッピングの利用者率UP
- 国内におけるEC市場の拡大
 - ECサイト間の競争は激化
- 昨今の顧客獲得競争を勝ち抜くため,ECサイト市場においてはリピート顧客の獲得に注目
 - 新規顧客に対して次回以降の購買を促すようなマーケティング施策が求められている



1 背景

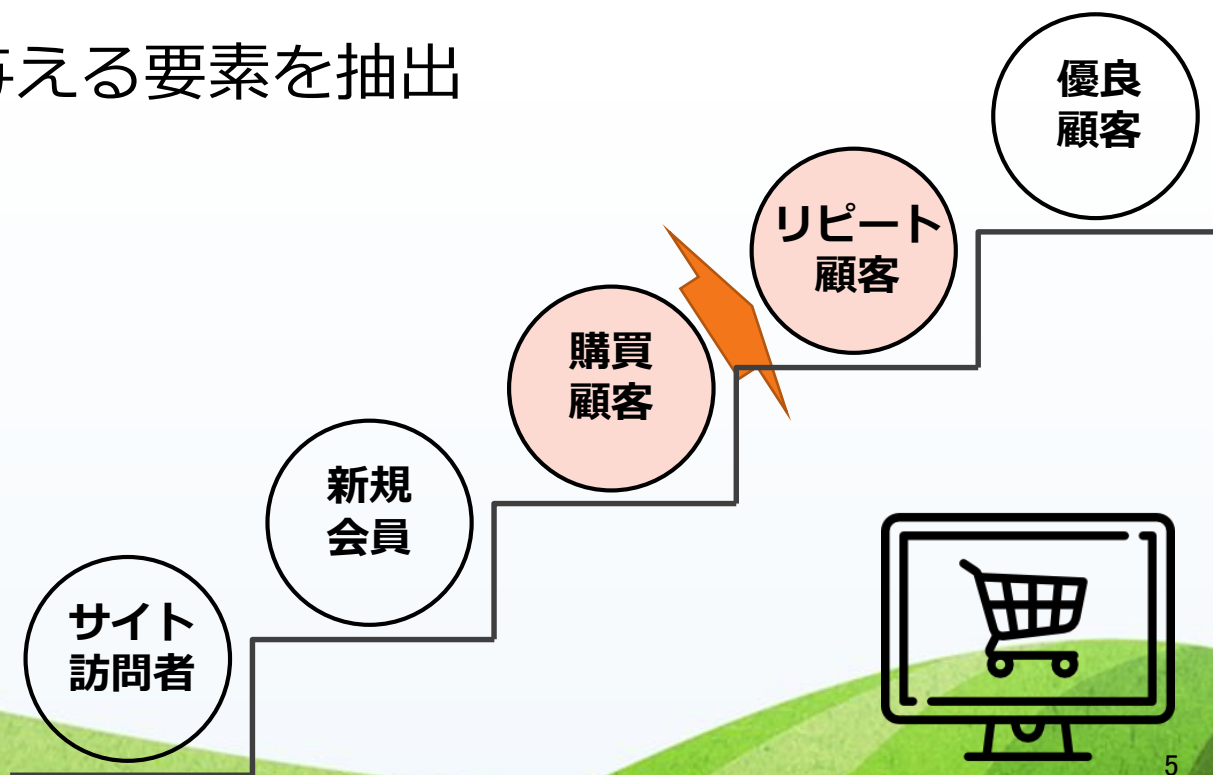
- アクセスログとは、アクセスによるWebサーバーの動作記録のこと。アクセスしたユーザーの様々な情報(IPアドレス、日次、アクセスされたファイル、リンク元のURLなど)が、ログデータとして記録される。*
- アクセスログと会員登録情報を組み合わせることで、購買に至るまでのプロセスを分析することができる。従来こうした購買前行動データは取得できなかったが、ECサイトではアクセスログデータにより、これらを把握することができる。



2 目的

- ECサイトにおいて初回購買後の1年間に再購買した顧客をリピート顧客と定義し、リピート顧客の閲覧行動を分析する。

- 初回購買日まで探索行動の特徴を明らかにする
- リピートに影響を与える要素を抽出



3 データ概要

- サイト：ゴルフ用品を取り扱うECサイト
- 使用データ：

アクセスログデータ

- 閲覧ページ
- 閲覧時間
- セッション回数
- ...

会員登録情報データ

- 登録日
- 性別
- 生年月日
- 都道府県
- ...



3-1 分析対象

会員登録

初回購買

再購買



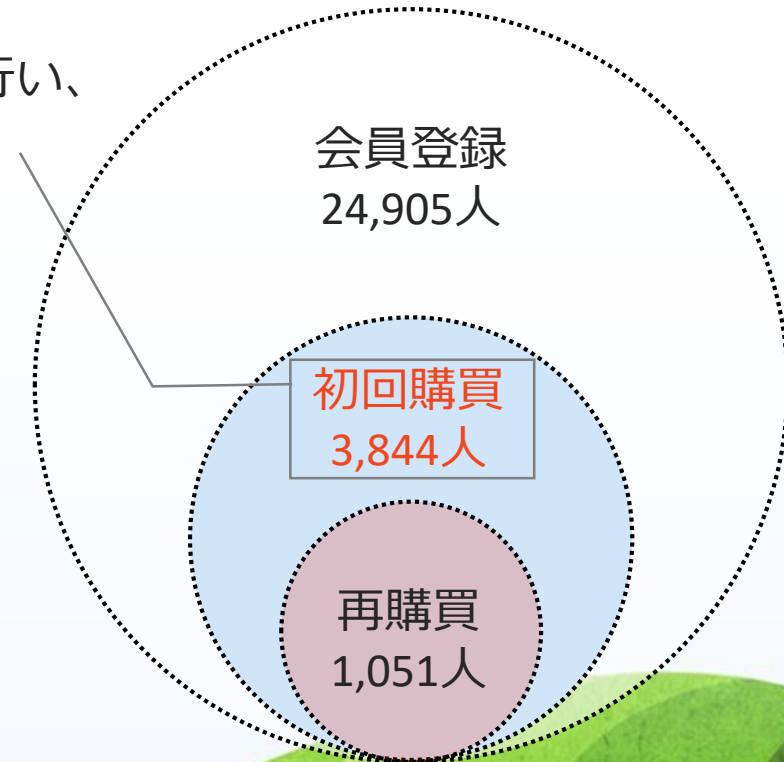
2017年 6月

7月

8月

2018年 7月

- 分析対象：
2017年6月1日～6月30日に会員登録を行い、
1ヶ月(30日)以内に再購買を行った顧客
- 分析対象期間：
2017年6月1日～2018年7月31日



3-2 基礎集計

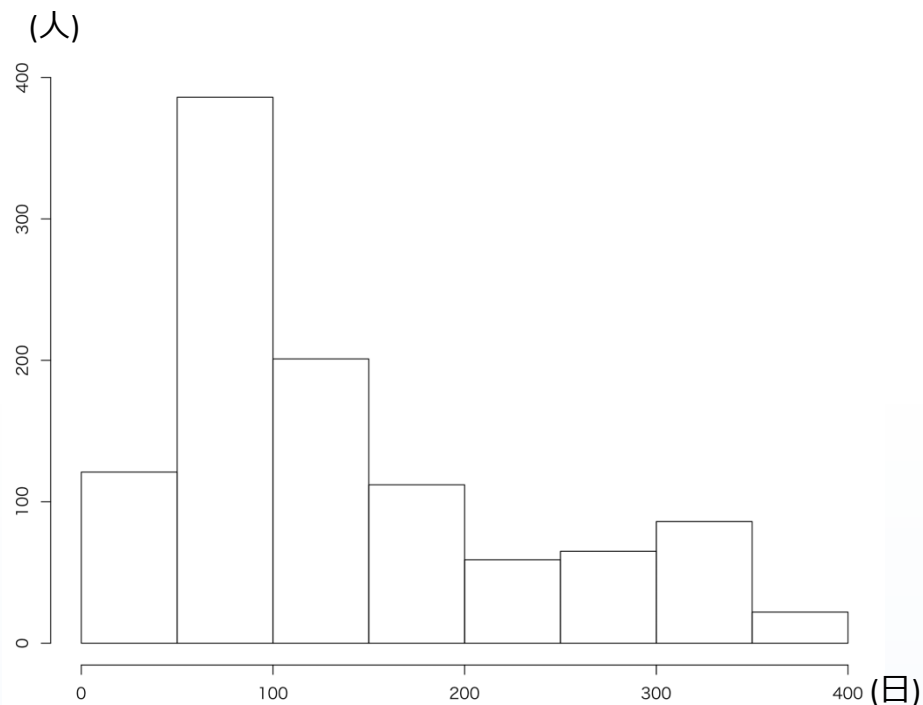


図1初回購買から再購買までの日数
(対象：1,051人)

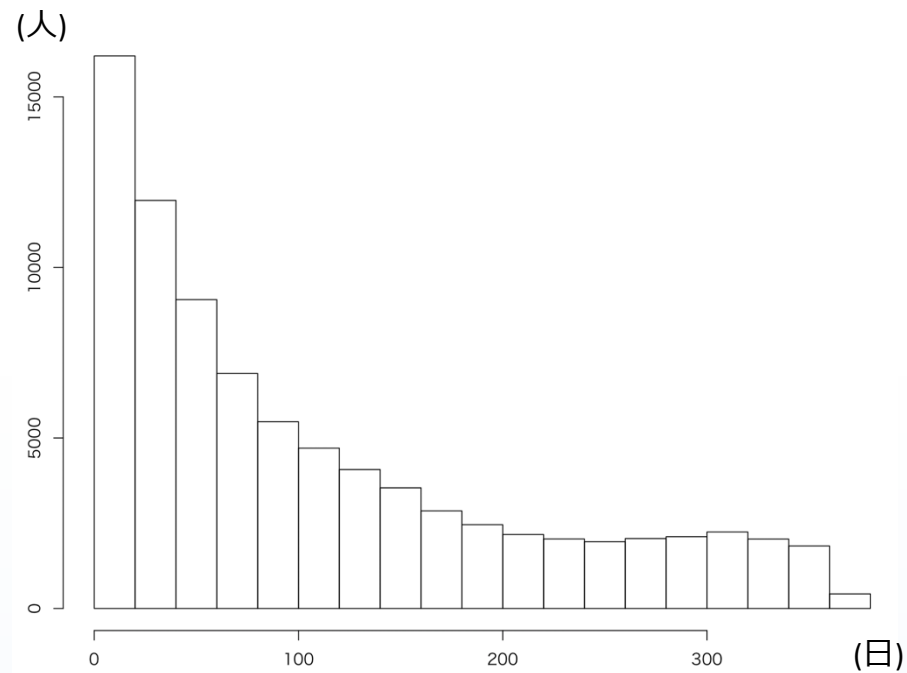
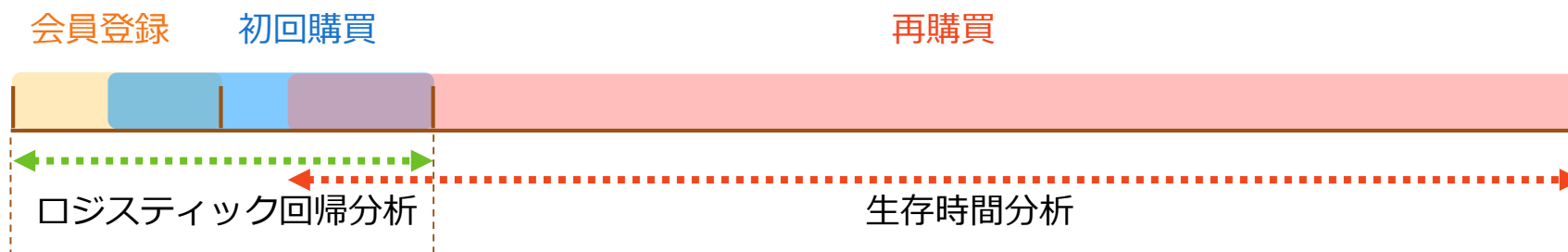


図2 初回購買からアクセスした日までの日数
(対象：3,844人)

4 分析



- ⊕ **01.ロジスティック回帰分析**
会員登録から初回購買までの閲覧行動を用いて再購買の要因を抽出
- ⊕ **02.決定木分析**
ロジスティック回帰分析で選択された変数を用いて要因を分類
- ⊕ **03.生存時間分析**
初回購買から再購買までの閲覧行動と再購買発生の関係を検討

4 分析① ロジスティック回帰分析

- 二項ロジスティック回帰分析では、目的変数を y として、 p 個の説明変数 x_1, \dots, x_p に対して顧客 i の再購買確率 p_i を下のロジスティック関数で表す。

$$p_i = \frac{\exp\{\beta_0 + \sum_{i=1}^n \beta_i x_i\}}{1 + \exp\{\beta_0 + \sum_{i=1}^n \beta_i x_i\}}$$

x_j : 再購買に影響を与える要因

β_j : 各説明変数に対するパラメータ

- 目的変数 : リピート顧客(=1), 離反顧客(=0)
- 説明変数 : 顧客の会員情報、初回購買日までの閲覧行動情報
- 変数選択 : ステップワイズ法

4 分析② 決定木分析

- 決定木は、条件分岐によってグループを分割して分類する手法である。
- 決定木では、不純度という指標をもとに分割します。ここで、決定木の分類クラス数を k とし、分岐ノード t で i 番目の分類クラスのデータに選ばれる確率を P として、*Gini*係数をもとに分割する変数とそのしきい値を決定

$$\begin{aligned}GI(t) &= \sum_{j \neq i} P(C_i|t)P(C_j|t) \\ &= \sum_{i=1}^K P(C_i|t)(1 - P(C_i|t)) \\ &= 1 - \sum_{i=1}^K P^2(C_i|t)\end{aligned}$$

4 分析③ 生存時間分析

- イベント (event) が起きるまでの時間とイベントとの間の関係に焦点を当てる分析方法である。
- 生存時間分析は、工学分野においては機械システムや製品の故障などを、医学分野においては疾患の病気の再発や死亡などを対象とした研究分野である。
- ここでは、初回購買から最購買が起きるまでの時間と再購買との間の関係を明らかにしたい。
- ノンパラメトリックモデル
 - 目的変数：初回購買からの日数
 - イベント：最購買の発生
 - 比較群：決定木で分類された変数

4 分析①②

株式会社NTTデータ数理システムから提供していただいた
“**Visual R Platform (VRP)**”により分析を行った。

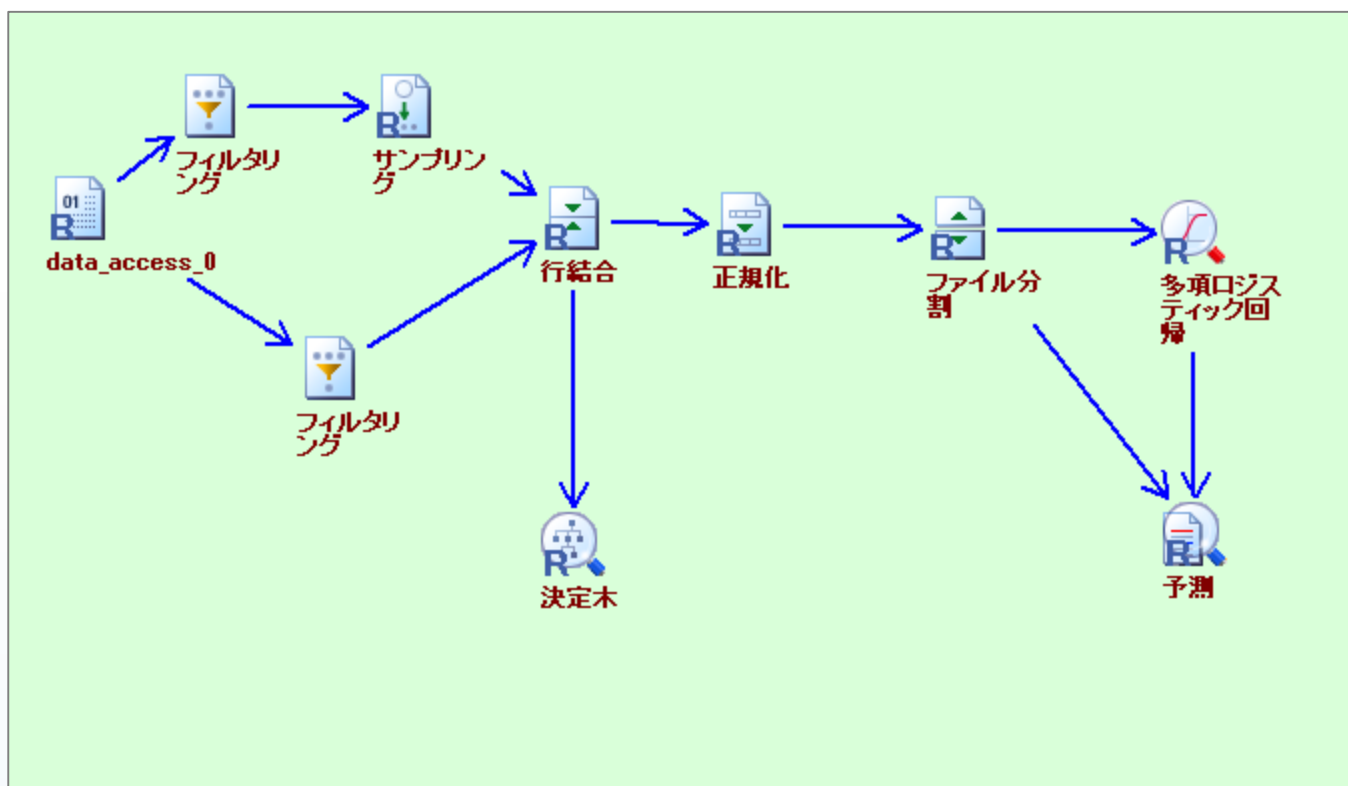


図1 VRPの例(ロジスティック回帰分析,決定木分析)

4 分析③

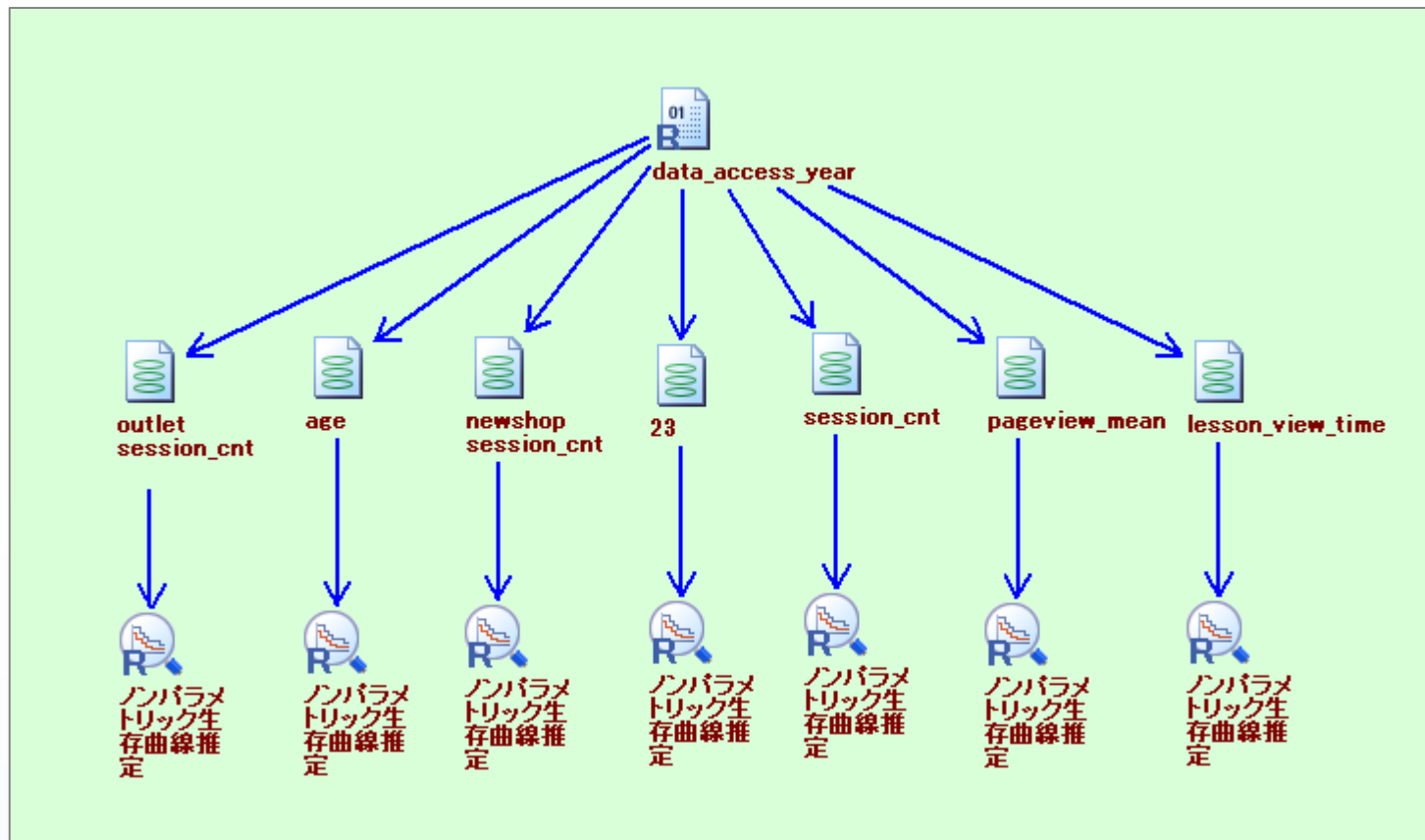


図2 VRPの例(生存時間分析)

5-1 ロジスティック回帰分析-結果と考察

表1 選択された偏回帰係数の推定値

変数名	回帰係数	標準誤差	Wald	p
切片	-0.2862	0.2569	1.2412	0.2652
年齢	0.2222	0.0548	16.4610	0.0000
会員登録日と初回購入日一致するかどうか	-0.2487	0.1424	3.0527	0.0806
セッション回数	-0.0510	0.0193	6.9609	0.0083
平均ページビュー数	0.2272	0.0593	14.6854	0.0001
パソコンで閲覧	0.2559	0.1732	2.1826	0.1396
スマホで閲覧	0.3537	0.1921	3.3895	0.0656
ゴルフ場予約ページのセッション回数	0.0758	0.0317	5.6929	0.0170
動画ページのセッション回数	1.7229	0.6410	7.2244	0.0072
動画ページの閲覧時間	-0.2721	0.1031	6.9651	0.0083
ニュースページのセッション回数	0.2474	0.1043	5.6292	0.0177
ニュースページの閲覧時間	-0.3786	0.1086	12.1396	0.0005
スタイルページのセッション回数	0.4161	0.2640	2.4838	0.1150
商品ページのセッション回数	0.1439	0.0277	26.9559	0.0000
中古品ページのセッション回数	0.1076	0.0324	11.0074	0.0009
ギアのセッション回数	-0.1438	0.0544	6.9991	0.0082
レッスンページの閲覧時間	0.2864	0.1161	6.0794	0.0137
アウトレットページのセッション回数	0.0967	0.0385	6.3181	0.0120
ショッピングカードページのセッション回数	-0.2164	0.0444	23.8084	0.0000
9時の閲覧時間	0.1588	0.0739	4.6126	0.0317
15分時の閲覧時間	0.1488	0.0637	5.4656	0.0194
23時の閲覧時間	0.1446	0.0678	4.5453	0.0330
21時の閲覧時間	0.1079	0.0645	2.8006	0.0942
2時の閲覧時間	-0.1576	0.0615	6.5721	0.0104
アウトレットページの閲覧時間	0.1922	0.1044	3.3870	0.0657

考察：

リピートしやすい要因

- 年齢層の高い
- 動画、ニュースなど多くアクセスするが、閲覧時間が少ない
- レッスンを長い時間に閲覧している
- 9時、15時、23時の時間帯に閲覧している

リピートしにくい要因

- ショッピングカードのセッション回数が高い
- 2時に閲覧している

5-2 決定木分析-結果

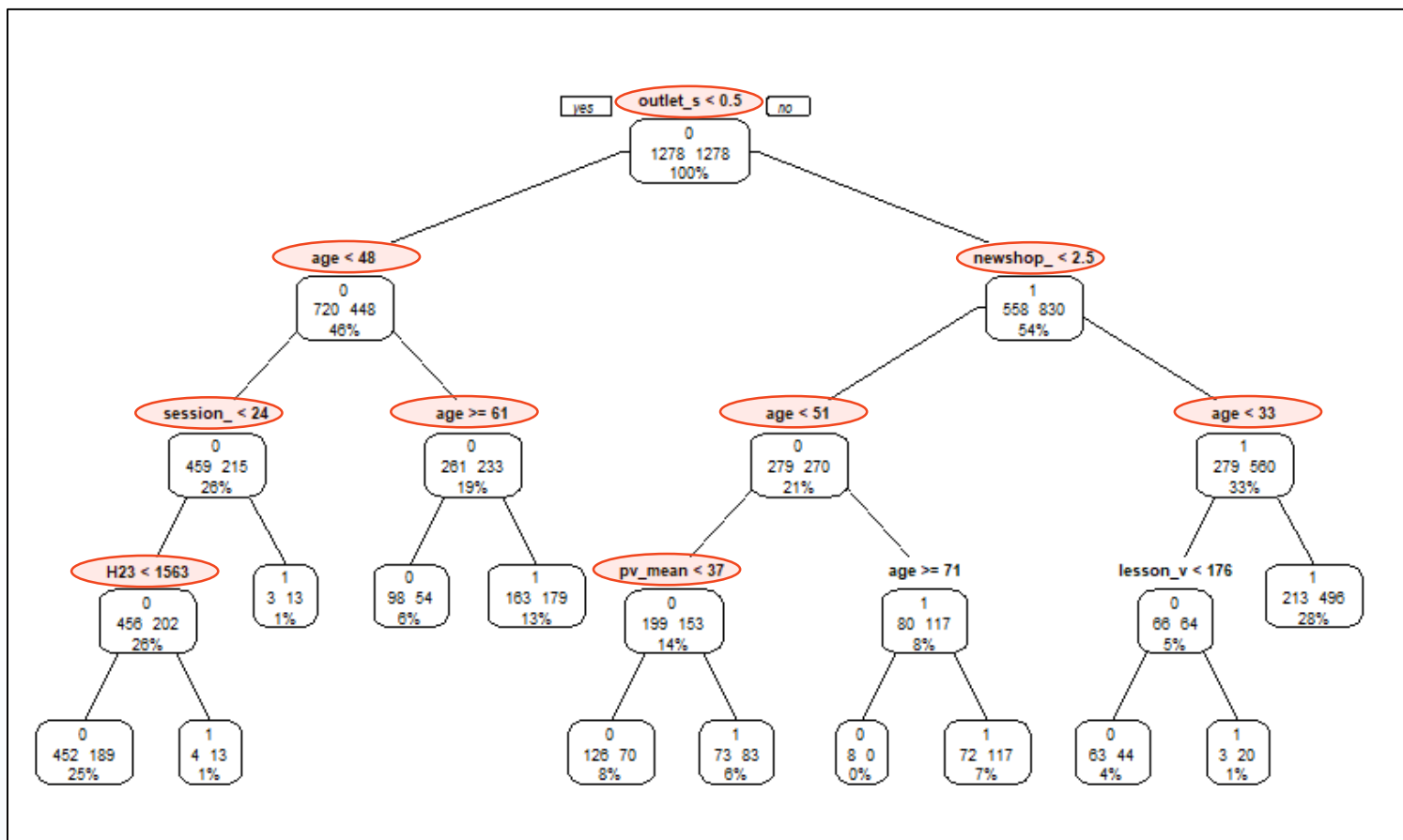


図3 VRPによるRの出力画面

5-2 決定木分析-生存時間分析

- ロジスティック回帰分析で有意(p値が0.05以下)になった変数を用いて、決定木分析を行った結果、分割前に10%以上の値が含まれている分岐条件は以下となる。
 - 年齢(0~32歳/33~47歳/48~50歳/51~60歳/61歳以上)
 - 総セッション回数(0~23/24以上)
 - 23時の閲覧時間(0~1562秒/1563秒以上)
 - 平均ページビュー数(0~36ページ/37ページ以上)
 - アウトレットページのセッション回数 (0回/1回以上)
 - 商品ページのセッション回数(0~2回/3回以上)
- この結果を条件として生存時間分析を行った。

5-3 生存時間分析—結果

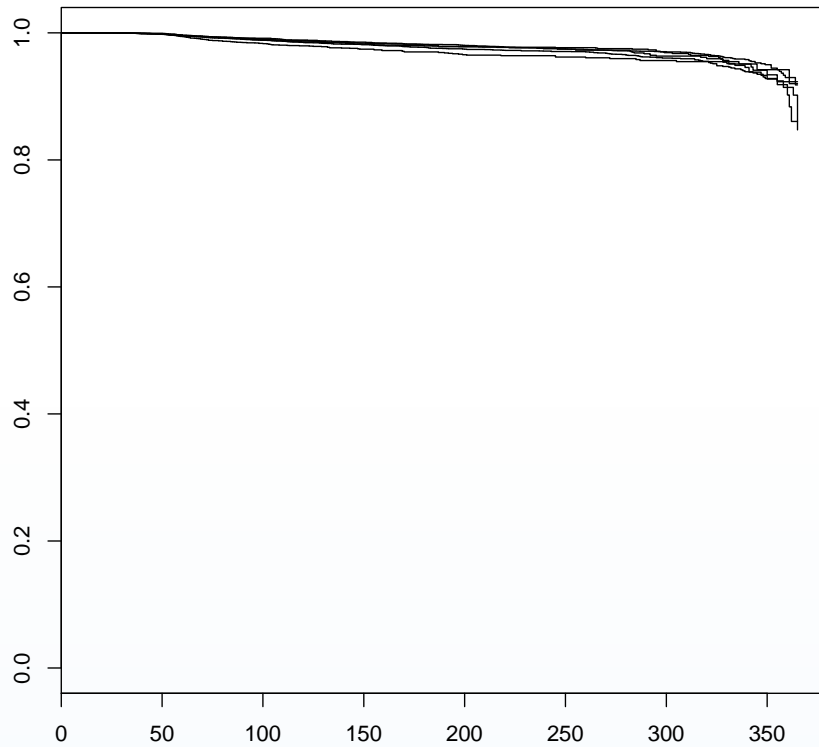


図4 年齢の群ごとの生存曲線

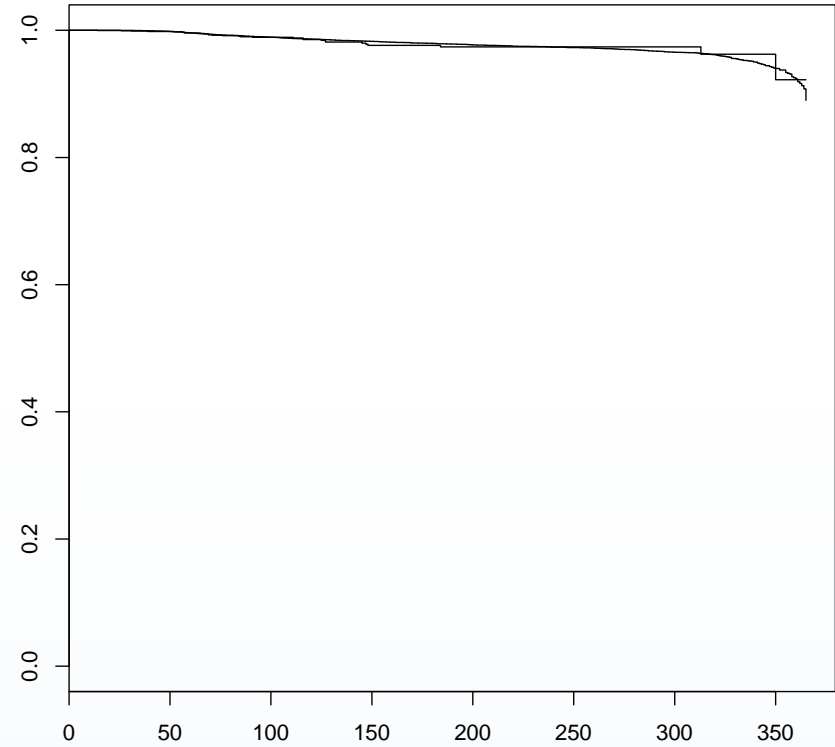


図5 レッスンページの閲覧時間の群ごとの生存曲線

年齢とレッスンページの閲覧時間の各群はあまり差がない。

5-3 生存時間分析—結果

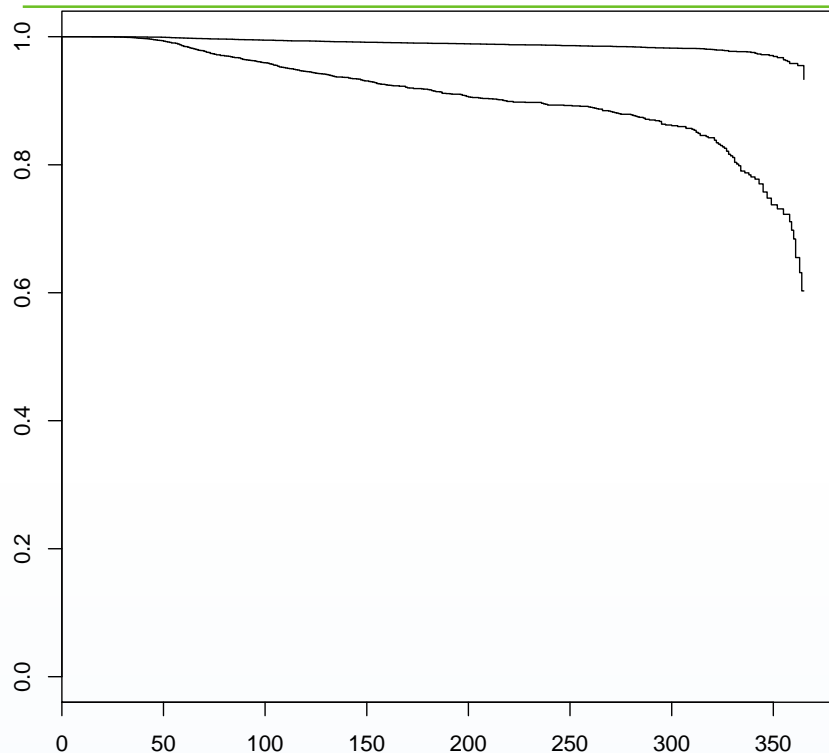


図6 アウトレットページのセッション回数
の群ごとの生存曲線

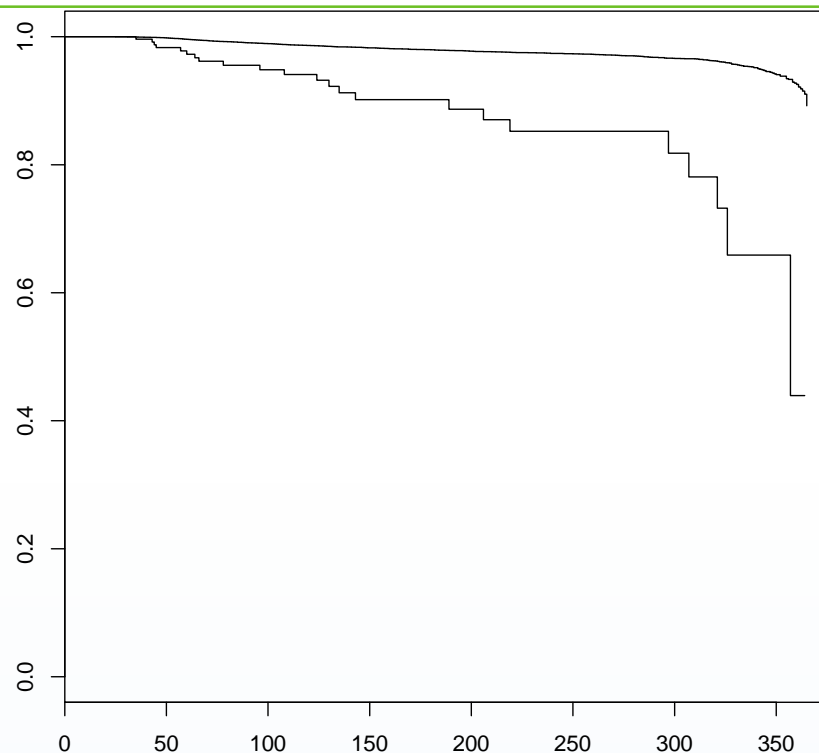


図7 23:00~23:59時間帯の閲覧時間
の群ごとの生存曲線

セッション内アウトレットページのアクセス有無の指標について、生存曲線では1年時点で3割ほどの差が示された。

23:00~23:59時間帯に合計閲覧時間が26分を超える顧客が明確な差があり、特に10月以後5割程度の再購買率の差が示された。

5-3 生存時間分析—結果

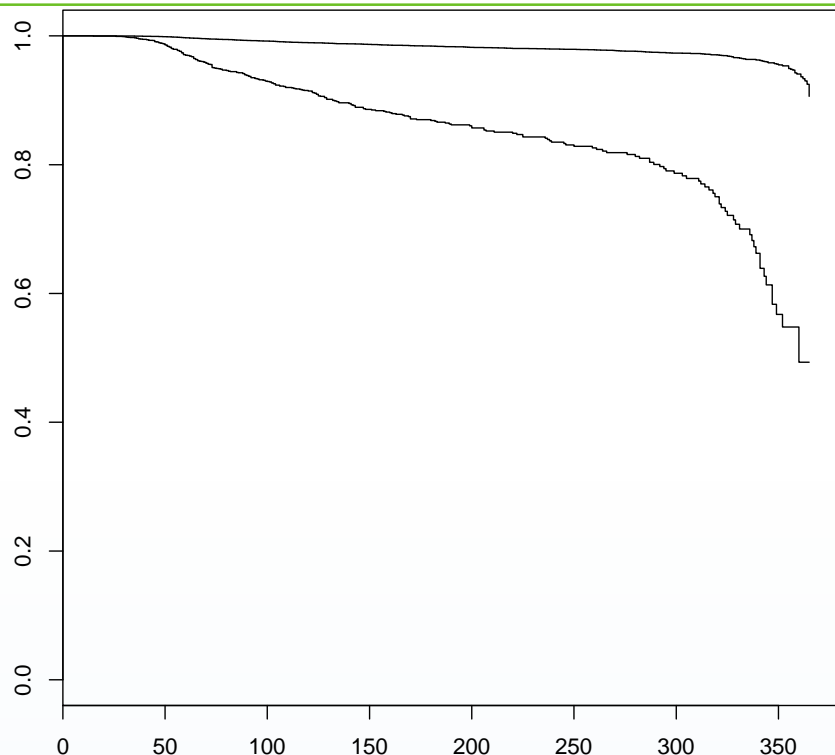


図8 商品ページのセッション回数の群ごとの生存曲線

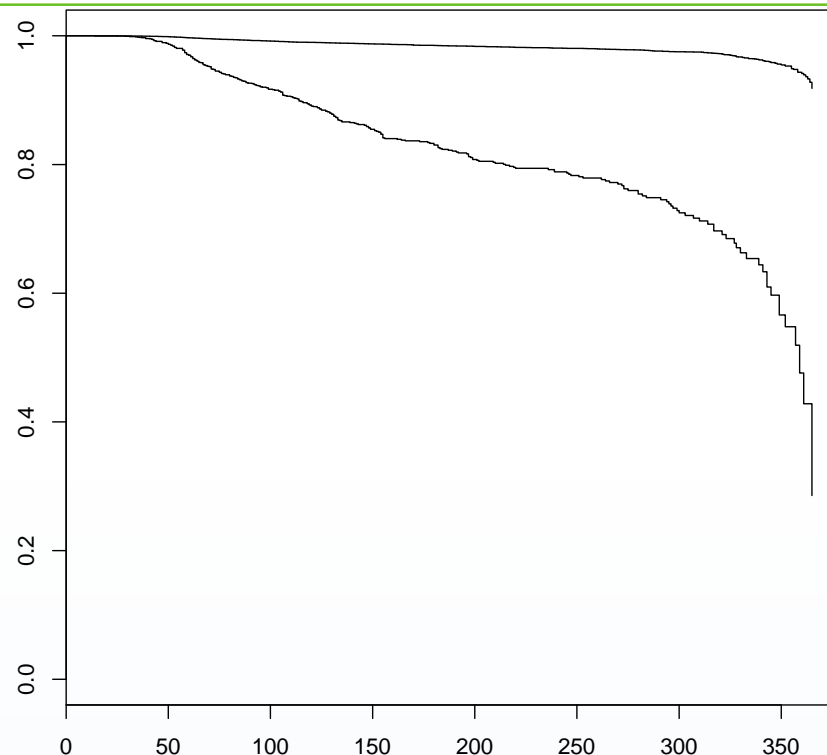


図9 平均ページビュー数の群ごとの生存曲線

商品ページのセッション回数2回以下と3回以上の指標について、生存曲線では1年時点で4割ほどの差が示された。

平均ページビュー数が27分を超える顧客が明確な差があり、1年時点で6割以上の再購買率の差が示された。

5-3 生存時間分析—考察

決定木で選択された指標の中に、

年齢(0~32歳/33~47歳/48~50歳/51~60歳/61歳以上)

総セッション回数(0~23/24以上)

の生存曲線ではっきりした傾向が示されず、

23時の閲覧時間(0~1562秒/1563秒以上)

平均ページビュー数(0~36ページ/37ページ以上)

アウトレットページのセッション回数 (0回/1回以上)

商品ページのセッション回数(0~2回/3回以上)

の生存曲線で明確な差が示された。

5 全体の考察

- ゴルフECサイトの特徴として、年齢層は高いが、年齢によるリピートの差はあまりない
- 商品ページをよくアクセスし、アウトレットページをアクセスしたことがある顧客は再購買しやすい傾向がある
- 23時によく閲覧している顧客、リピートしやすい

6 まとめ

- ECサイトにおいて初回購買後の1年間に再購買した顧客をリピート顧客と定義し、リピート顧客の閲覧行動を分析した。
- ✓ 初回購買日まで探索行動の特徴を明らかにした
- ✓ リピートに影響を与える要素を抽出した

今後の課題

- 分析対象期間を延ばすことで、対象者を増やし、季節性も考えた上で、より詳しい閲覧行動特徴を特定できると考える。
- 分析結果を用いた施策について、さらなる分析が必要である。

参考文献

- 株式会社 NTTデータ数理システム, Visual R Platformマニュアル, 2019
- 金明哲, 「Rによるデータサイエンス 第2版」, 森北出版, 2017
- 佐藤由将, 大竹恒平, 生田目崇, “ECサイトにおけるリピート顧客の特徴の分析,” 情報処理学会第79回全国大会講演論文集, pp. 507-508, 2017