

---

# 消費者向けECサイトにおける 顧客の行動予測モデルに関する研究

---

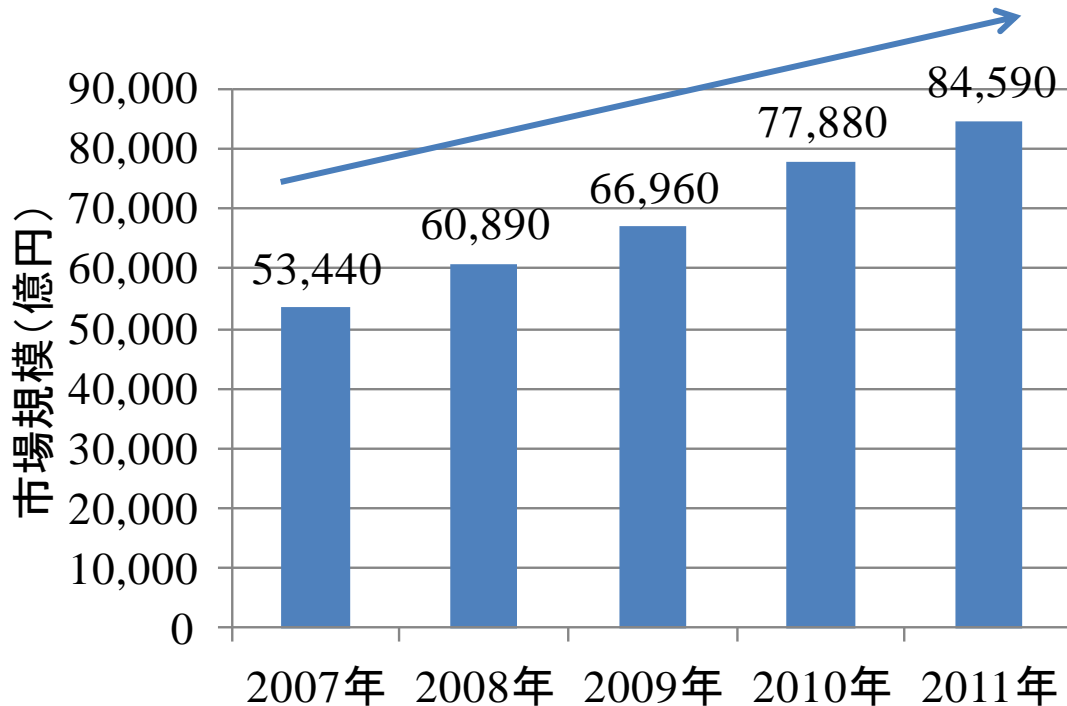
東京理科大学大学院  
工学研究科 経営工学専攻  
佐藤 翔太

# 構成

---

1. 研究背景
  2. 研究目的
  3. 分析モデル
  4. データ概要
  5. 分析結果
  6. まとめ
  7. 今後の課題
- 参考文献

# 研究背景



消費者向け電子商取引  
(EC: Electronic Commerce)  
サイトの  
市場規模は増加傾向[4].



ECサイトの増加.

図1: 消費者向けECサイト市場規模[4]

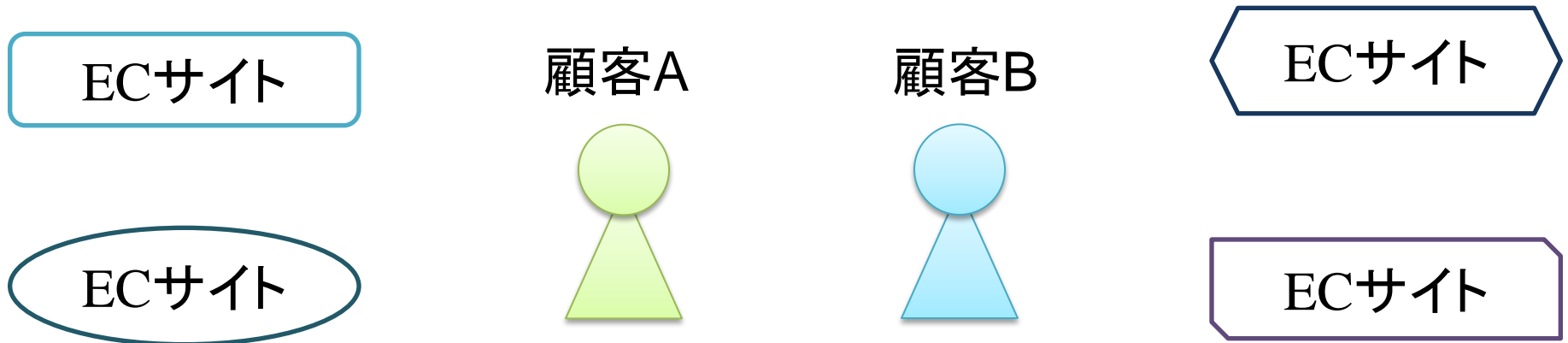
## 電子商取引:

インターネット上で行われる財またはサービスの販売または購入[3].

# 研究背景

▶ ECサイトの増加.

⇒ ECサイト間での競争激化, 顧客奪い合い



ECサイトのマーケティング担当者にとって

顧客により多くサイトに来てもらい, より多く購買してもらうことを狙う.  
顧客の将来における購買・来店予測の必要性.

# 先行研究

## ➤ ECサイトにおける顧客の行動予測に関する研究

### 来店予測

- Moe and Fader[6]  
—顧客が次にいつ来店するかを予測

### 購買予測

- Moe and Fader[7]
- Van den Poel and Bunckinx[9]  
—「どの来店で購入するか」を分析  
—顧客の次回来店時の購買有無を予測

## 問題点

- ✓ 次回来店が「1か月後」なのか「1時間後」なのかはわからない。
- ✓ 「どの来店で購入するか」を予測するため、顧客の来店を予測できない。

# 研究目的

## ➤ 本研究の着目点

- 「次回来店」ではなく、「将来のある日時」の行動を予測する
- 顧客の購買と来店を同時に予測する
  - 「明日」購買してくれそうな顧客の発見が可能
  - 購買意向が高まっている顧客の発見が可能

## 研究目的

- 顧客ごとに将来の購買・来店を予測するモデルの構築.  
例)「明日」購買するのか, 来店はするのか,
- 顧客の購買・来店傾向の把握.

# 顧客の購買・来店行動をモデル化

## ■本研究の購買・来店モデル

「購買機会」を「カレンダー上の日にち」ととらえる。  
t番目の「購買機会」はt番目の日にち。

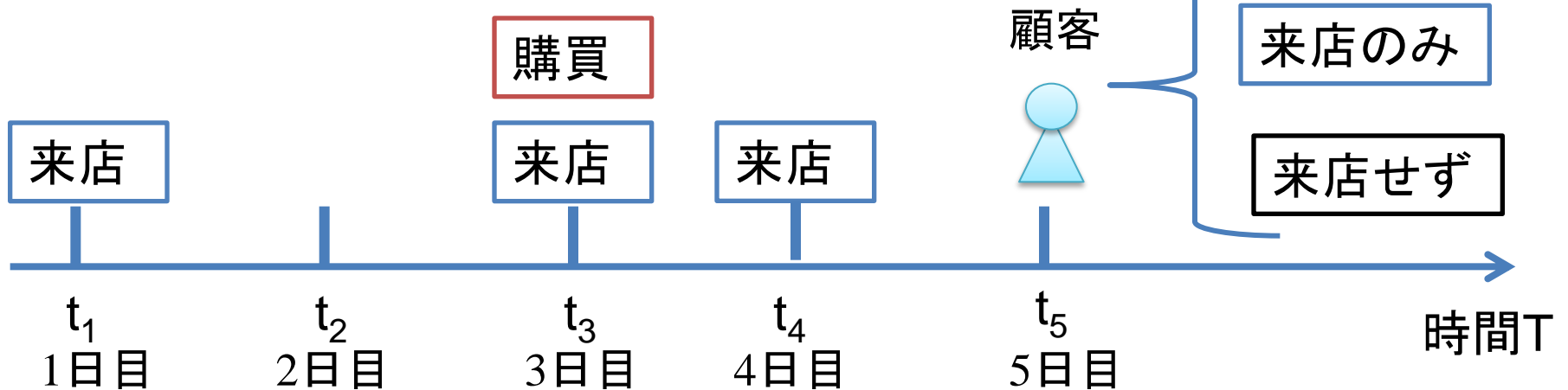


図2: 本研究モデルでの顧客行動

第t-1日までの行動から、第t日目(図2の例の場合 $t_5$ )において「購買」、「来店のみ」、「来店せず」のいずれになるかを予測.

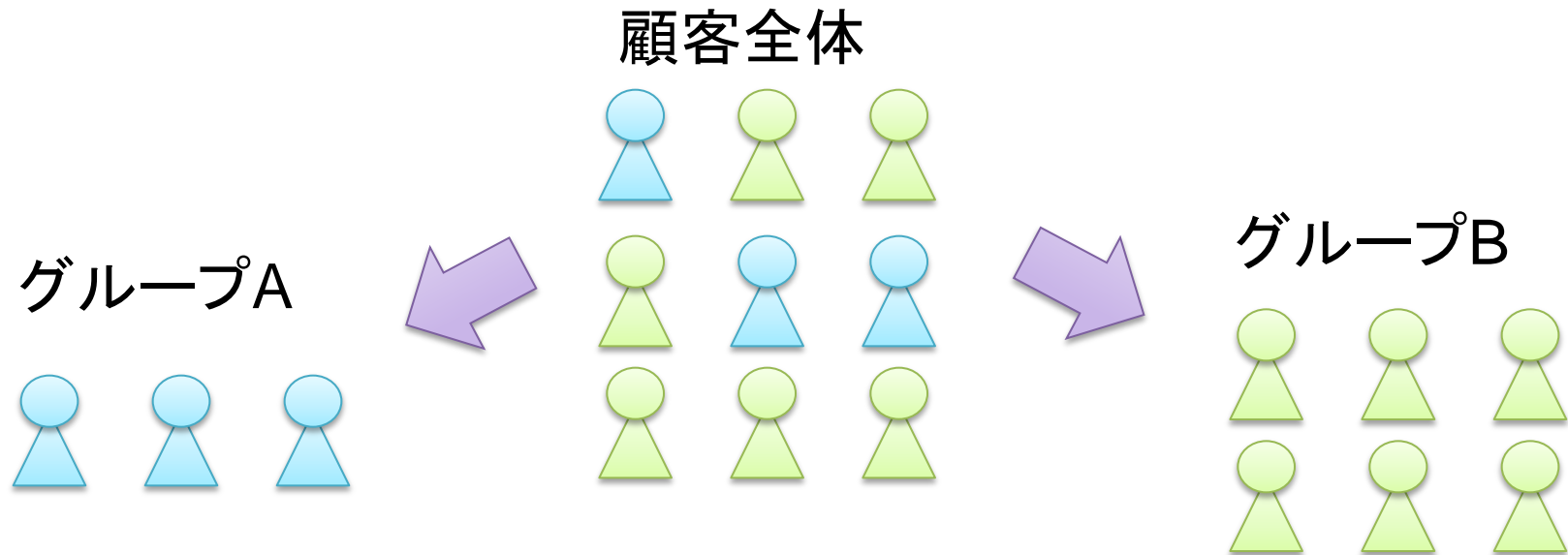
# 来店・購買行動をモデル化

- 第 $t$ 日に以下の3つの選択肢のうち、どの行動をとるかを予測する
    - ・「来店し購買する」
    - ・「来店はするが購買はしない」
    - ・「来店しない」
- 多項ロジットモデル[1]

- 顧客の異質性の考慮
    - ・購買や来店に対して顧客ごとに傾向は異なる可能性  
例) 休日に来店する顧客, 平日に来店する顧客
- 潜在クラスモデル[1]



# 潜在クラス多項ロジットモデルの例



## グループAの特徴

- ・平日に来店する傾向
- ・平日に購買する傾向

## グループBの特徴

- ・休日も平日も来店する傾向
- ・休日に購買する傾向

# 潜在クラス多項ロジットモデル

- 顧客*i*の第*t*日における購買・来店を考える(多項ロジットモデル).

$$V_{itsj} = x'_{it} \beta_{sj} \quad (1)$$

$$P_{its}(j) = \frac{\exp(V_{itj})}{\sum_{j=1}^3 \exp(V_{itj})} \quad (2)$$

セグメント, 選択肢ごとに異なる

$x_{it}$ : 顧客*i*の第*t*日における説明変数

$\beta_{sj}$ : セグメント*s*における選択肢*j*に対するパラメータベクトル

$P_{its}(j)$ : セグメント*s*における顧客*i*の第*t*日に選択肢*j*を選択する確率

*i*: *i*番目の顧客 ( $i=1, \dots, N$ )

*t*: *t*番目の日にち ( $t=1, \dots, T$ )

*j*: 選択肢 ( $j=1, \dots, 3$ )

$j=1$ : 「購買」,  $j=2$ : 「来店のみ」

$j=3$ : 「来店せず」

# 潜在クラス多項ロジットモデル

## ■ モデルの尤度関数

$$L_i = \sum_{s=1}^S r_s \prod_{j=1}^3 \prod_{t=1}^T (P_{its}(j))^{y_{ij}} \quad (3)$$

各クラスの割合

多項ロジットモデル

$$L = \prod_{i=1}^N L_i \quad (4)$$

(4)式の対数をとったものを最大とするパラメータを  
EMアルゴリズムを用いて推定(最大化問題、S+NUOPTを利用)

$y_{ijt}$ : 顧客*i*が第*t*日に選択肢*j*を  
選択していたら1, それ以外0

$r_s$ : セグメント*s*の割合を示すパラメータ

$L$ : 全体の尤度

# データ概要

- 提供: 経営科学系研究部会連合協議会
- 概要: ゴルフ用品を扱う消費者向けECサイトの  
アクセスログデータ, 購買データ, 会員データ
- 期間: 2010年8月1日から2011年3月10日までを利用

8月から1月末を学習用データ,  
2月から3月10日を検証用データとして利用

- 分析対象: 165人(5回以上購買をしている顧客)

※本研究では日次の顧客行動をモデル化するため  
データを日ごとに要約して分析に利用.

# データ要約

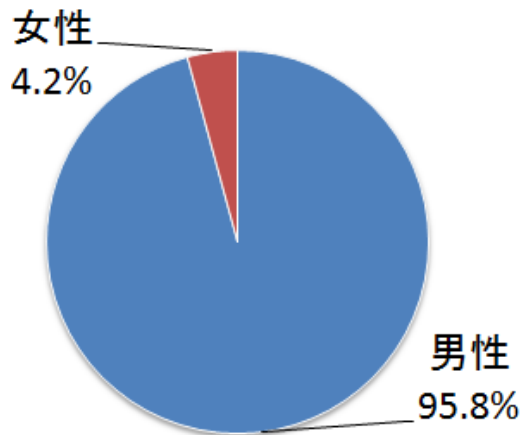


図3: 性別の割合

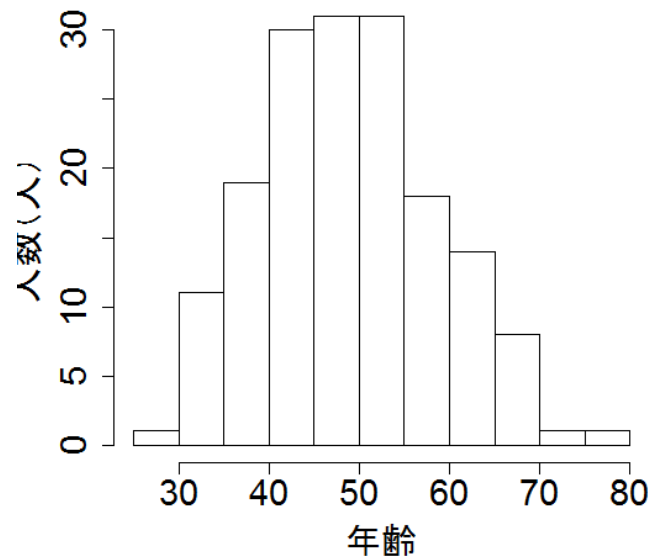


図4: 年齢の分布

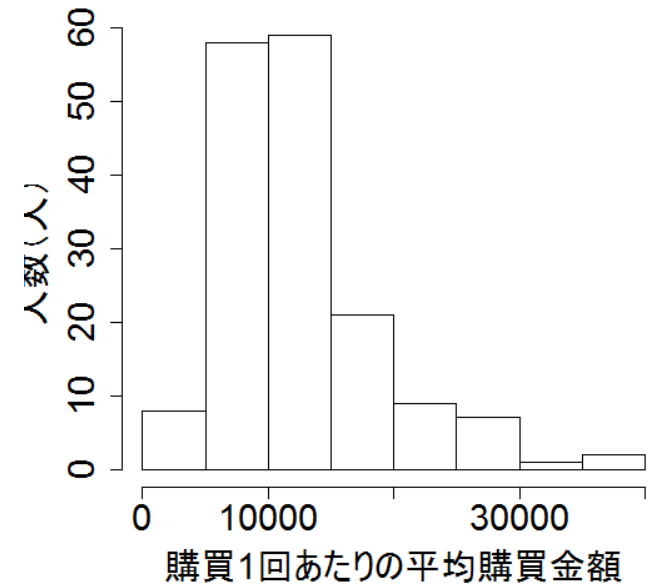


図5: 購買1回当たりの平均金額

- ✓ 全体165人中, 男性が9割以上. → 男性が多い.
- ✓ 年齢は40歳代~50歳代が中心. → 年配層がやや多い.
- ✓ 1回の購買における顧客の平均購買金額は10,000円前後.

# 説明変数

- Moe and Fader[6]では顧客の来店と購買履歴から購買を予測。  
 → 本研究では顧客*i*の第*t*日の購買・来店を以下の要因で予測する。

表1: 説明変数

変数名	説明
1週間累積滞在時間	( <i>t</i> -7)日から( <i>t</i> -1)日のサイト滞在時間の合計(時間)
1か月累積金額	( <i>t</i> -30)日から( <i>t</i> -1)日までの合計購買金額(万円)
土日祝日	<i>t</i> 日が土日祝日であったら1, それ以外0

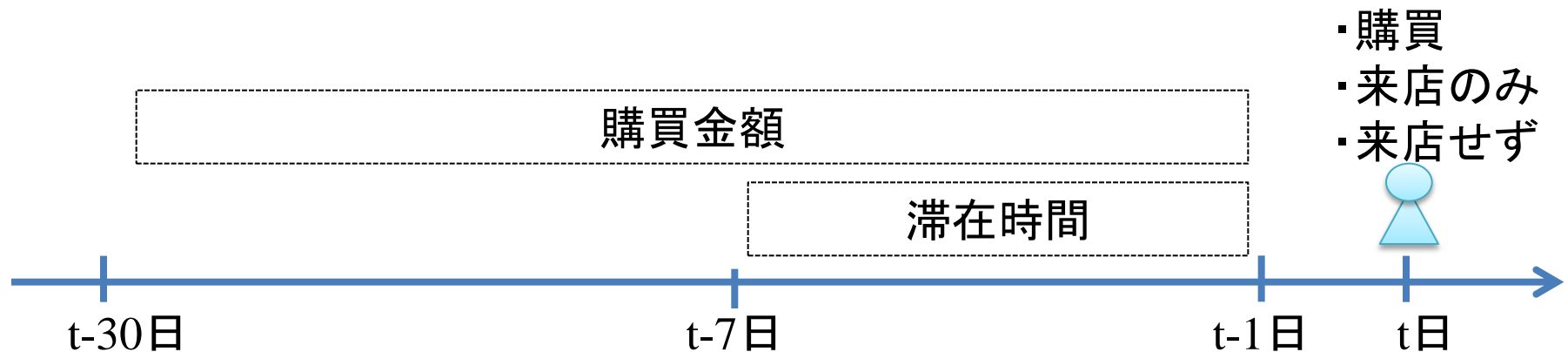


図6: 購買・来店と説明変数

# セグメント数の決定

- ✓ BIC(Bayesian Information Criteria)によってセグメント数を決定  
—BIC: 小さいほど良いモデル
- ✓ セグメント数6以上については、構成比がほぼ0のセグメントなど解釈が困難であったため、セグメント数1から5のBICを比較

表2: 各セグメントのBIC

セグメント数	BIC
1	44169.59
2	42056.17
3	41470.71
4	40809.64
5	<b>40634.96</b>

- ✓ セグメント数1から5では  
セグメント数5が最適なモデル  
→5セグメントを採用
- ✓ セグメント数1のモデルは  
通常が多項ロジットモデル  
→本研究モデルは通常が多項  
ロジットモデルより良いモデル

# セグメント別パラメータ推定結果

表3: パラメータ推定結果

		セグメント1	セグメント2	セグメント3
購買	切片	-3.429**	-3.839**	-1.964**
	土日祝日	0.388**	1.203**	-0.880**
	1か月購買金額	0.083**	0.007	-0.120
	1週間滞在時間	0.344**	0.319**	0.397**
来店のみ	切片	-0.623**	-1.632**	0.981**
	土日祝日	0.148**	0.809**	-0.716**
	1か月購買金額	0.036**	0.027	-0.071**
	1週間滞在時間	0.331**	0.323**	0.288**
セグメント割合		34.8%	19.0%	11.0%

「来店せず」のパラメータは識別性のために0に固定, \*\*:1%有意

「1週間滞在時間」はどのセグメントでも「来店のみ」, 「購買」に正の影響。

・直近1週間で滞在時間が長い時期は, 来店・購買する傾向。



# セグメント別パラメータ推定結果

表4: パラメータ推定結果

		セグメント4	セグメント5
購買	切片	-3.736**	-3.225**
	土日祝日	-0.205	-2.724**
	1か月購買金額	0.004	-0.001
	1週間滞在時間	0.554**	0.552**
来店のみ	切片	-2.271**	-0.337**
	土日祝日	-0.524**	-2.344**
	1か月購買金額	0.008	-0.031
	1週間滞在時間	0.736**	0.381**
セグメント割合		23.5%	11.7%

「来店せず」のパラメータは識別性のために0に固定, \*\*:1%有意

「1週間滞在時間」はどのセグメントでも「来店のみ」, 「購買」に正の影響.

・直近1週間で滞在時間が長い時期は, 来店・購買する傾向.

# セグメント別特徴

## ➤ セグメント1 割合 34.8%

- ・購買・来店のみとともに休日に行く傾向
- ・「1か月購買金額」が来店・購買で正の影響  
→購買したら近々また来店や購買を行う  
→ある一定時期に購買や来店を繰り返す傾向

## ➤ セグメント2 割合 19.0%

- ・購買・来店ともに休日に行く傾向が強い

## ➤ セグメント3 割合 11.0%

- ・購買・来店ともに休日に行く傾向が強い
- ・「来店のみ」に対する「切片」が正, 「1か月購買金額」が負  
→普段は比較的来店するが, 直近1か月に購買していると  
来店しなくなる傾向

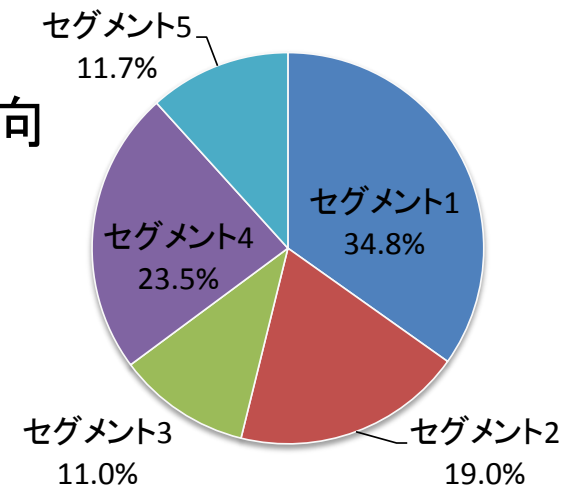


図6: セグメント別割合

# セグメント別特徴

## ➤ セグメント4 割合 23.5%

- ・購買は休日平日関係なし
- ・来店は平日に行う傾向
- ・「1週間累積時間」の来店への影響が5セグメント中最も強い

## ➤ セグメント5 割合 11.7%

- ・購買・来店ともに平日に行う傾向がとて強い
- ・「1週間累積時間」の購買への影響が他セグメントに比べ強い

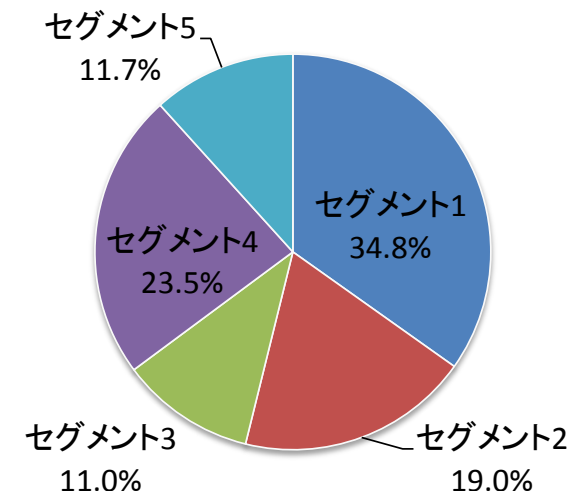


図7: セグメント別割合

# モデルの検証

- 得られたパラメータと検証用データによるモデルの検証
  - ・本研究モデル
  - ・通常の多項ロジットモデル(比較モデル)

表5:モデルの検証

	本研究モデル	比較モデル
的中率	71.9%	68.8%
BIC	40634.96	44169.59

選択確率が最大となる選択肢を選択すると予測的中率: 予測的中数 / 全予測数

顧客の異質性を考慮したモデルの方があてはまりが良い。

# まとめ

✓ 顧客の将来の来店・購買を予測するモデルを構築し、顧客の購買・来店につながる要因を確認した。

—セグメント4: 直近1週間で滞在時間が長いほど  
来店や購買をする傾向が強い

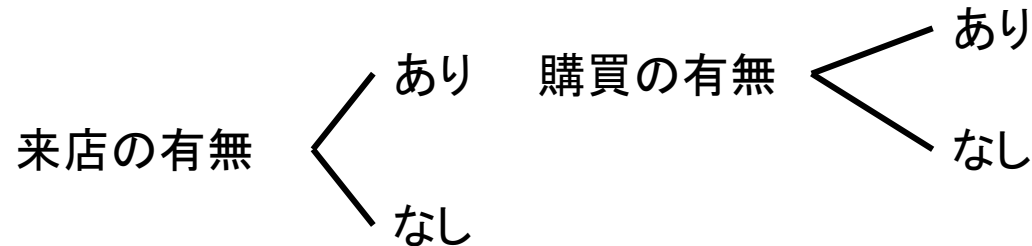
—セグメント5: 直近1週間で滞在時間が長いほど  
特に購買をする傾向が強い

例1) セグメント4は、直近1週間で滞在時間が長い時  
来店・購買をする傾向なため、衝動買いを促すアプローチ。

例2) セグメント5は、直近1週間で滞在時間が長い時、購買する  
傾向なため、滞在時間が長いと購買意向有の可能性。

# 今後の課題

- ✓ 説明変数の検討による予測精度の向上
- ✓ 「来店」がおこらないと「購買」がおこらないことを考慮  
— 入れ子ロジットモデル[2]



- ✓ 顧客個別のパラメータの推定

# 主要参考文献

- [1]阿部誠, 近藤文代:「マーケティングの科学—POSデータの解析—」, 朝倉書店(2005).
- [2]Bucklin, R. E., Gupta, S: “Brand choice, Purchase incidence and segmentation: An integrated modeling approach”, *Journal of Marketing Research*, 29, pp.201-205 (1992).
- [3]経済産業省:「平成21年度我が国情報経済社会における基盤整備(電子商取引に関する市場調査)報告書」, 経済産業省(2010).
- [4]経済産業省,  
( [http://www.meti.go.jp/policy/it\\_policy/statistics/outlook/bessi3H23EChoukokusho.pdf](http://www.meti.go.jp/policy/it_policy/statistics/outlook/bessi3H23EChoukokusho.pdf) ),  
最終閲覧日:2012/9/24
- [5]小西貞則, 越智義道, 大森裕浩:「計算機統計学の方法—ブートストラップ・EMアルゴリズム・MCMC—」, 朝倉書店(2010).
- [6]Moe, W. W., Fader, P S: “Capturing Evolving Visit Behavior in Clickstream Data”, *Journal of Interactive Marketing* 18(1), pp.5-19(2004).
- [7]Moe, W. W., Fader, P S: “Dynamic Conversion Behavior at E-Commerce Sites” , *Management Science* 50(3), pp.326-335(2004).
- [8]里村卓也:「Rで学ぶデータサイエンス13 マーケティング・モデル」, 共立出版(2010).
- [9]Van den Poel, D., Buckinx, W: “Predicting online-purchasing behaviour” , *European Journal of Operational Research* 166(2), pp.557-575(2005).

# Appendix



# 推定方法

(4)式の対数をとった $LogL$  を最大化するパラメータを推定.  
推定にはEMアルゴリズムを用いる.

## EMアルゴリズム[5][8]

1. 観測されない架空の変数を導入
  - 消費者のセグメントに対する所属変数 $z_{ns}$ を導入  
(消費者 $n$ がセグメント $s$ に所属: $z_{ns}=1$ , 所属しない: $z_{ns}=0$ )
2. 観測されたデータとパラメータの推定値から $z_s$ を推測 (E-step)
3. 推測された $z_s$ に基づいてパラメータの推定値を更新 (M-step)
4. 対数尤度の向上が基準以下になるまでE-step, M-stepを反復

# EMアルゴリズム

潜在変数 $z_{ns}$ の導入した対数尤度

$$\log L_c = \sum_{i=1}^N \sum_{s=1}^S \{z_{is} \log f_s(y_i | \beta_s; x_i) + z_{is} \log r_s\} \quad (5)$$

ただし,

$$f_s(y_i | \beta_s; x_i) = \prod_{j=1}^3 \prod_{t=1}^T (P_{its}(j))^{y_{ij}} \quad (6)$$

# EMアルゴリズム

## ➤ E-step

データとパラメータを既知として  $z_{is}$  の期待値を求める。

$$\hat{z}_{is} = \frac{r_s f_s(y_i | \beta_s; x_i)}{\sum_{s=1}^S r_s f_s(y_i | \beta_s; x_i)} \quad (7)$$

## ➤ M-step

(10)式を最大化するパラメータをラグランジュの未定乗数法を用いて求める。

$$E[\log L_c] = \sum_{i=1}^N \sum_{s=1}^S \hat{z}_{is} \log r_s + \sum_{n=1}^N \sum_{s=1}^S \hat{z}_{is} \log f_s(y_i | \beta_s; x_i) \quad (8)$$

$$0 \leq r_s \leq 1, \quad \sum_{s=1}^S r_s = 1, \quad (9)$$

# 予測的中率

モデルより、「購買」、「来店のみ」、「来店せず」それぞれの選択確率が算出できる。顧客は所属確率が最も高いセグメントに所属するとする。

$$\text{予測的中率} = \frac{a+e+i}{a+b+c+d+e+f+g+h+i} \quad (10)$$

表6: 予測・実測の行列

予測\実測	購買	来店のみ	来店せず
購買	a	d	g
来店のみ	b	e	h
来店せず	c	f	i