

生存時間解析を用いたヘアサロンにおける店販購入確率予測

中央大学大学院 理工学研究科 経営システム工学専攻 片桐 優帆



目次

1. 背景・目的
2. データ概要
3. 分析
4. 提案モデル
5. 結果・精度評価
6. 考察
7. まとめ・今後の課題

1. 背景・目的

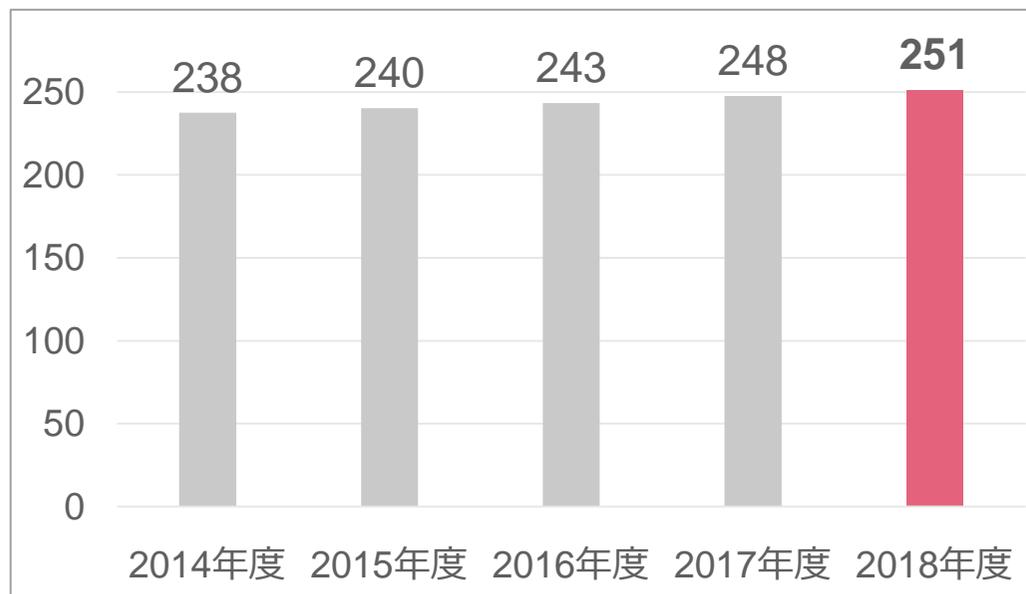


図1. 美容所の店舗数(千店舗)

近年、美容室の店舗数は**増加傾向**[1].

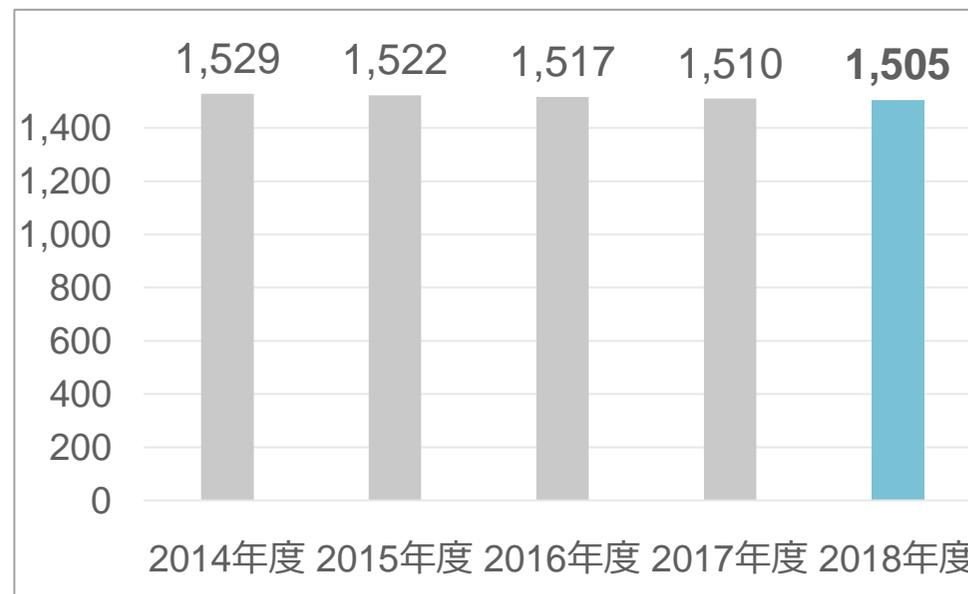


図2. 美容市場規模の推移(百万円)

サービス料金は低価格化し、市場規模は**減少**[1].

各店舗の経営環境は厳しく、集客や売上向上のための工夫が必要と考えられる。

1. 背景・目的

- 美容院の主な業務は、カットやカラーなどの施術提供。
- 時間を伴うサービスのため、現状の人的資源でこれまで以上の売上増加が難しい場合も少なくない。



時間を伴わないサービス「**店販商品**」に注目。

施術サービス以外に店販商品の購入を促すことは、売上拡大を目指す上で重要。

1. 背景・目的

目的

複数の店舗を持つヘアサロンチェーンを対象に、
店販商品の購入間隔を分析。

購入確率の予測を高い精度で行う。



- 効果的なレコメンド・ダイレクトメール(DM)送付タイミングを把握し、
購入意欲が高いタイミングに合わせてPR.
- 商品の在庫管理に活用.

2-1. データ概要

- 関東圏内11店舗のヘアサロンチェーンのID付きPOSデータ
 - データ期間：2015年7月～2017年6月の2年間
2015年7月～2017年3月を学習データ，2017年4月以降をテストデータ
 - データ概要：表1. 使用データの概要

表1. 使用データの概要

データ名称	内容
顧客情報データ	顧客ID, 初回来店年, 初回来店店舗ID, 郵便番号, DM送信可否, 性別, 年代
顧客情報付き会計履歴データ	顧客ID, 会計日時, 会計金額, 選択した施術メニュー, 来店店舗等の会計情報

2-2. 分析対象

対象店舗，対象顧客は以下の通りである．

- 店舗：大型店舗(3店舗)
- 顧客：2015年7月～2017年3月の店販購入回数が2回～20回の顧客(1978人)

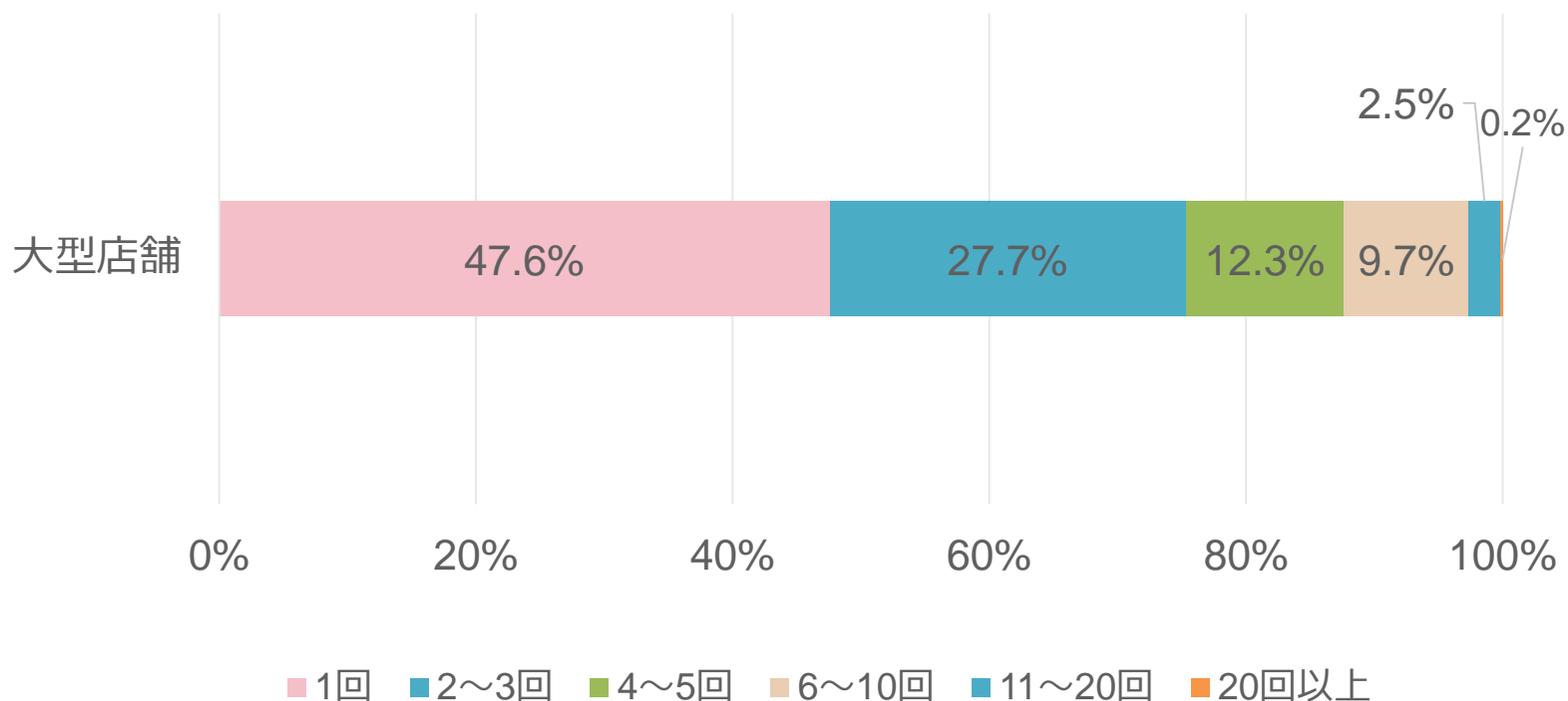
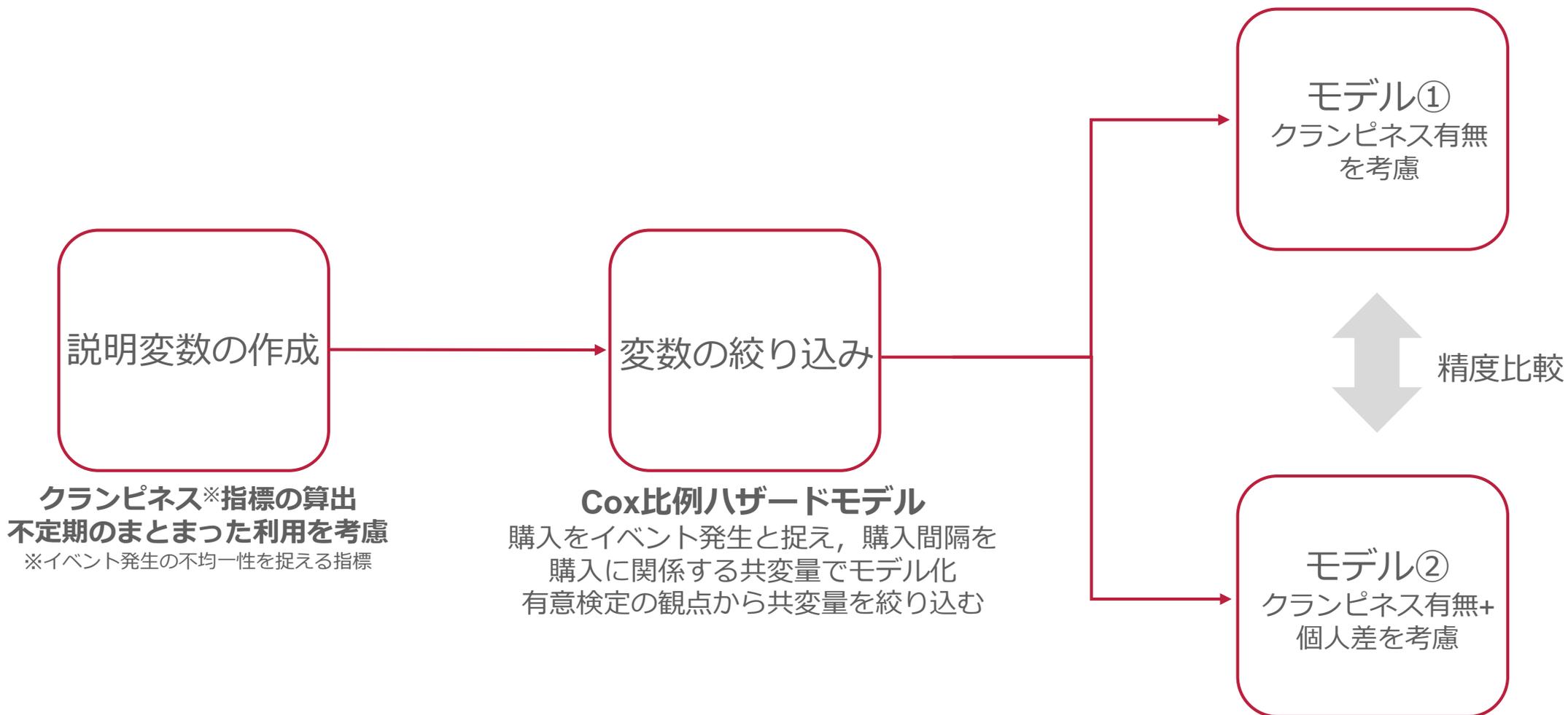


図3. 店販購入回数人数比

3-1. 分析の流れ



3-2. 説明変数の作成

店販購入間隔に影響を与えると考えられる、5項目をもとに説明変数を作成した。

- ① 前回, 2回前来店時の施術 (前回, 2回前来店時の施術内容)
- ② 施術選択確率 (各施術メニューをどの程度の確率で選択しているか)
- ③ 前回来店時の会計情報 (前来店時の使用金額, 店販売購入有無等)
- ④ ヘアサロンへのロイヤリティ (累積使用金額, 店販売購入回数等)
- ⑤ 個人属性 (年代, 性別等)

3-2. 説明変数の作成 <クランピネス指標>

本研究では、RFM指標の1つとしてクランピネス指標[5]を用いた。

- 顧客関係管理(CRM)手法の1つとして、最終購入日(Recency), 来店頻度(Frequency), 購入金額(Monetary)を顧客ごとに記録したRFM指標は簡潔で有用なデータとして広く活用されてきた[6,7].

→しかし、不規則のまとまった利用などRFM指標だけでは**識別できない購買パターンが存在.**

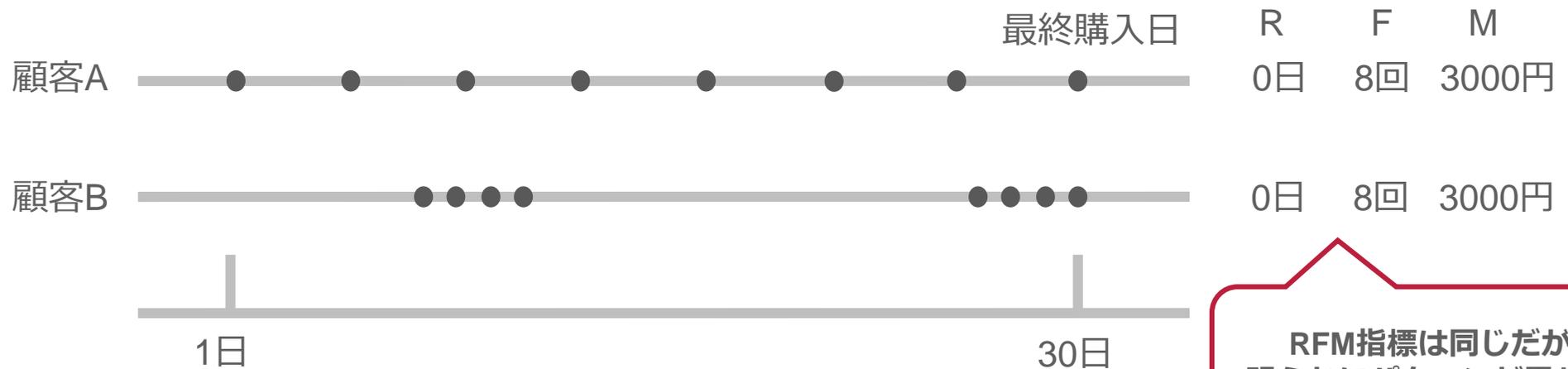


図4. RFM指標だけでは識別できない購買パターン

3-2. 説明変数の作成 <クランピネス指標>

- クランピネス(C)指標：イベントが均等な間隔に従わない度合い
イベント発生の不均一性を捉えることができる。



RFM指標だけでは識別できない顧客の購買行動を捉え、
購買パターンの異質性を考慮できる。

各顧客の店販商品購買行動(2015年7月～2017年3月)から
クランピネス有無を算出。
購買パターンの異質性をモデルに組み込む。

3-2. 説明変数の作成 <クランピネス指標>

[6] 中山(2016)

H 指標(H_p) : 消費者 p の購買タイミングの不均一性を表す指標

購買機会 N 回のうち n 回購買が行われた場合に, i 回目に購買が行われたタイミングを t_i とする.

さらに, t_i を基準化した x_i を

$$x_i = \begin{cases} t_1 & (i = 1) \\ t_i - t_{i-1} & (i = 2, 3, \dots, n) \\ N + 1 - t_n & (i = n + 1) \end{cases}$$

と定義したとき,

$$H_p = 1 + \frac{\sum_{i=1}^{n+1} \log(x_i)x_i}{\log(n+1)}$$

と表せる.

3-2. 説明変数の作成 <クランピネス指標>

次に購買機会 N 回, 購買回数 n の条件下でランダムな購買シミュレーションを M 回行い, その経験分布の $\alpha\%$ 分位点を境界値 H_0 とする.

クランピネス指標 C_p は,

$$C_p = \begin{cases} 1 & (H_p \geq H_0) \\ 0 & (\text{その他}) \end{cases}$$

と計算される.

※ $C_p = 1$: クランピネス有, $C_p = 0$: クランピネス無

3-2. 説明変数の作成 <クランピネス指標>

- 購買機会 $N = 640$ (2015年7月~2017年3月)
- 購買シミュレーションを $M = 20000$ 回
- $\alpha\%$ 分位点 = 0.10
として, 各顧客のクランピネス有無を算出した.

表2. クランピネス有無人数

クランピネス有り	クランピネス無し
147	1831

3-2. 説明変数の作成 <クランピネス指標>

- 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，データ処理，クランピネス指標の算出を行った。

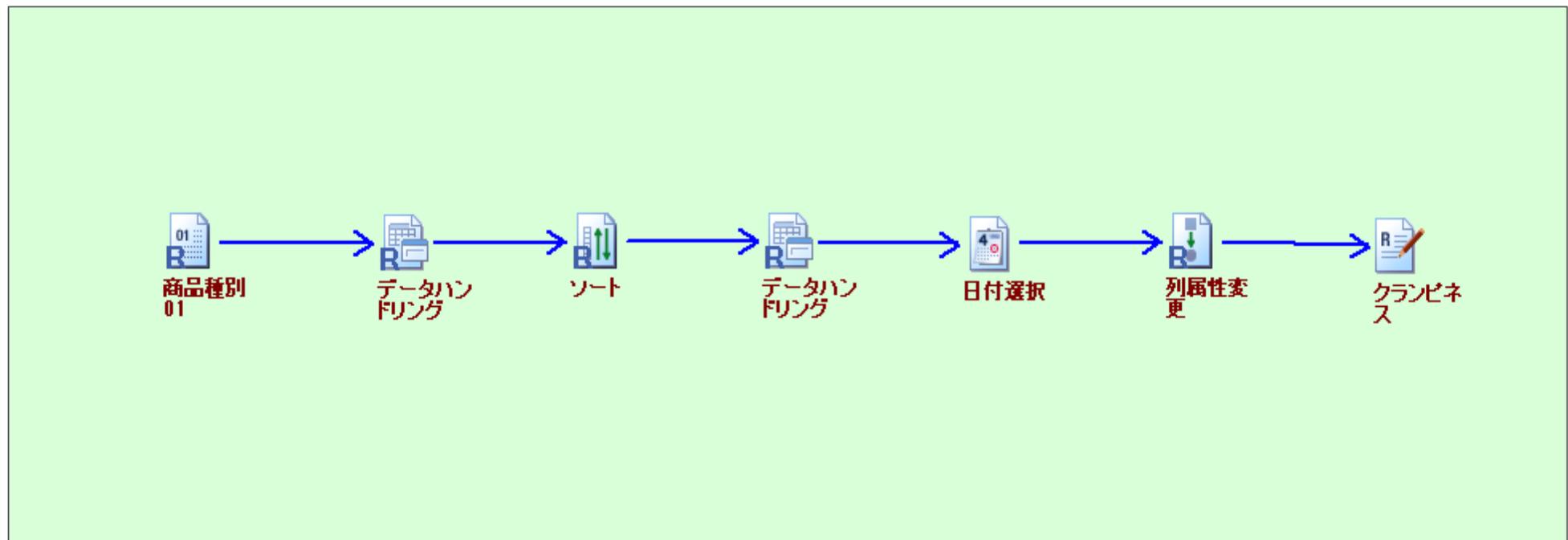


図5. VRPの例（クランピネスの算出）

3-2. 説明変数の作成

クランピネス指標を含め、全27変数を作成した。

表3. 変数一覧

項目	交互作用か	変数名	データ型
前回, 2回前来店時の施術	交互作用	お直し	0 or 1
		カット	0 or 1
		カラー	0 or 1
		クーポン	0 or 1
		ストレート	0 or 1
		トリートメント	0 or 1
		パーマ	0 or 1
		ブロー・シャンプー・ヘアセット	0 or 1
施術選択確率	交互作用でない	お直し	実数
		カット	実数
		カラー	実数
		クーポン	実数
		ストレート	実数
		トリートメント	実数
		パーマ	実数
		ブロー・シャンプー・ヘアセット	実数

3-2. 説明変数の作成

表3. 変数一覧

項目	交互作用か	変数名	データ型
前回来店時の会計情報	交互作用でない	使用金額	実数
		割引金額	実数
		店販購入有無	0 or 1
		指名有無	0 or 1
ロイヤリティ	交互作用でない	累積使用金額	実数
		期間内の累積店販購入金額	実数
		店販購入回数	実数
		クランピネスの有無	0 or 1
個人属性	交互作用	年代 性別	カテゴリ カテゴリ
	交互作用でない	DM送信可否	カテゴリ

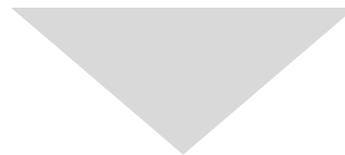
3-3. 説明変数の絞り込み

[8] 山口(2014)

➤ **Cox比例ハザードモデル**（ステップワイズ法）を用いて変数を絞り込む.

- 目的変数：非購入期間（日）
- 説明変数：作成した全変数
- モデル選択基準：ステップワイズ法

提案モデルで使用する学習データを用いて、**Cox比例ハザードモデル**で
パラメータ推定.



各係数の有意検定の観点からパラメータ推定結果を検証し、
変数を絞り込む.

3-3. 説明変数の絞り込み

- 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.

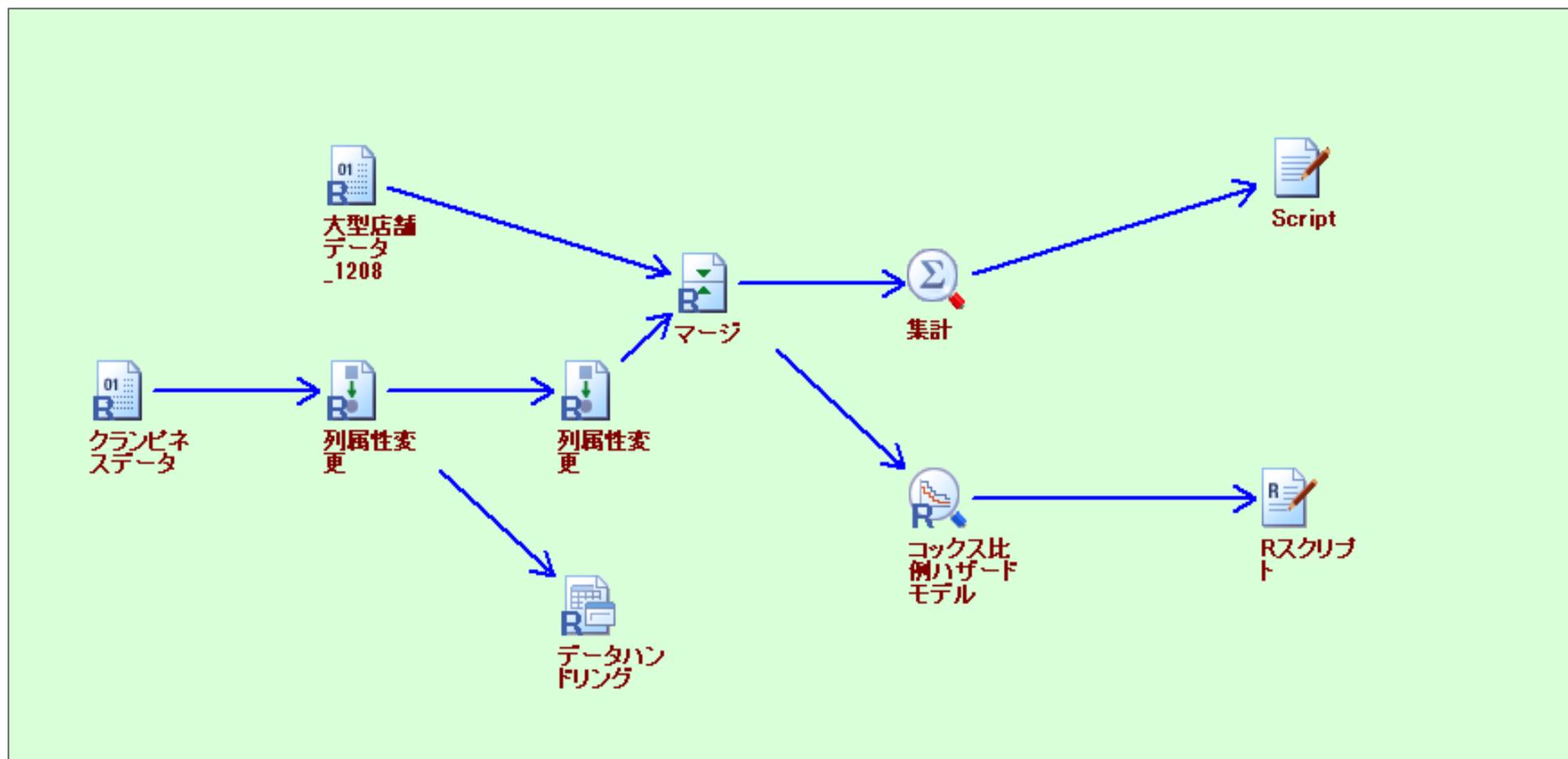


図6. VRPの例 (Cox比例ハザードモデル)

3-3. 説明変数の絞り込み

表4. Cox比例ハザードモデル結果

	coef	exp(coef)	se(coef)	z	Pr(> z)	
前来店時店販有無	1.01	2.73	0.03	32.49	2.00E-16	***
期間内の累積店販購入回数	0.21	1.24	0.01	30.48	2.00E-16	***
クランピネス有無	-0.98	0.38	0.08	-12.15	2.00E-16	***
期間内の累積金額_千円	0.00	1.00	0.00	-5.52	3.49E-08	***
ブロー.シャンプー.ヘアセット	0.38	1.46	0.07	5.23	1.74E-07	***
2回前カット	0.53	1.70	0.10	5.15	2.63E-07	***
前来店カット:2回前カット	-0.49	0.61	0.10	-4.76	1.92E-06	***
クーポン	-1.89	0.15	0.42	-4.45	8.47E-06	***
トリートメント	0.34	1.41	0.09	4.01	6.19E-05	***
DM送信可否許可	0.12	1.13	0.03	3.82	0.000132	***
パーマ	-0.47	0.62	0.12	-3.81	0.000141	***
前来店時指名有無	0.29	1.33	0.08	3.68	0.00023	***

有意検定の観点から、ロイヤリティ変数を
提案モデルの説明変数として採用。

4. 提案モデルの概要

➤ ワイブル分布を仮定した生存時間解析

生存時間解析：死亡や故障など、イベントが起こるまでの時間を分析する手法。

- 店販商品の購入（イベント）が起こるまでの時間： t
→ **ワイブル分布**に従うと仮定

- 確率密度関数：
$$f(t) = \left(\frac{m}{\eta}\right) \left(\frac{t}{\eta}\right)^{m-1} \exp\left\{-\left(\frac{t}{\eta}\right)^m\right\}$$

- ハザード関数：
$$h(t) = \frac{m}{\eta^m} t^{m-1}$$

- 生存関数：
$$S(t) = \exp\left\{-\left(\frac{t}{\eta}\right)^m\right\}$$

4. 提案モデルの概要

[9] 津田他(2006)

ここで,

$$\begin{aligned}h(t|X, \beta) &= mt^{m-1}\eta^{-m} \\ &= \frac{m}{\eta^m} t^{m-1} \\ &= m\lambda t^{m-1}\end{aligned}$$

$\lambda = \frac{1}{\eta^m}$ へ共変量の効果を加え, 以下のように定義する.

$$\lambda = \exp(\beta_0 + \sum_{i=1}^p x_i \beta_i)$$

よって, 尺度パラメータ η は以下.

$$\eta = \exp\left(-\frac{\beta_0 + \sum_{i=1}^p x_i \beta_i}{m}\right)$$

4. 提案モデルの概要

以上より, この場合の確率密度関数と生存関数は以下のように計算できる.

$$f(t) = mt^{m-1} \exp(X^T \beta) \exp\{-t^m \exp(X^T \beta)\}$$

$$S(t) = \exp\{-t^m \exp(X^T \beta)\}$$

$$(X^T \beta = \beta_0 + \sum_{i=1}^p x_i \beta_i)$$

4. 提案モデルの概要 <クランピネスごとの比較>

ここで、クランピネス有無ごとに期間内の累積店販購入回数，期間内の累積使用金額（千円単位），前来店時店販購入有無について集計を行った。

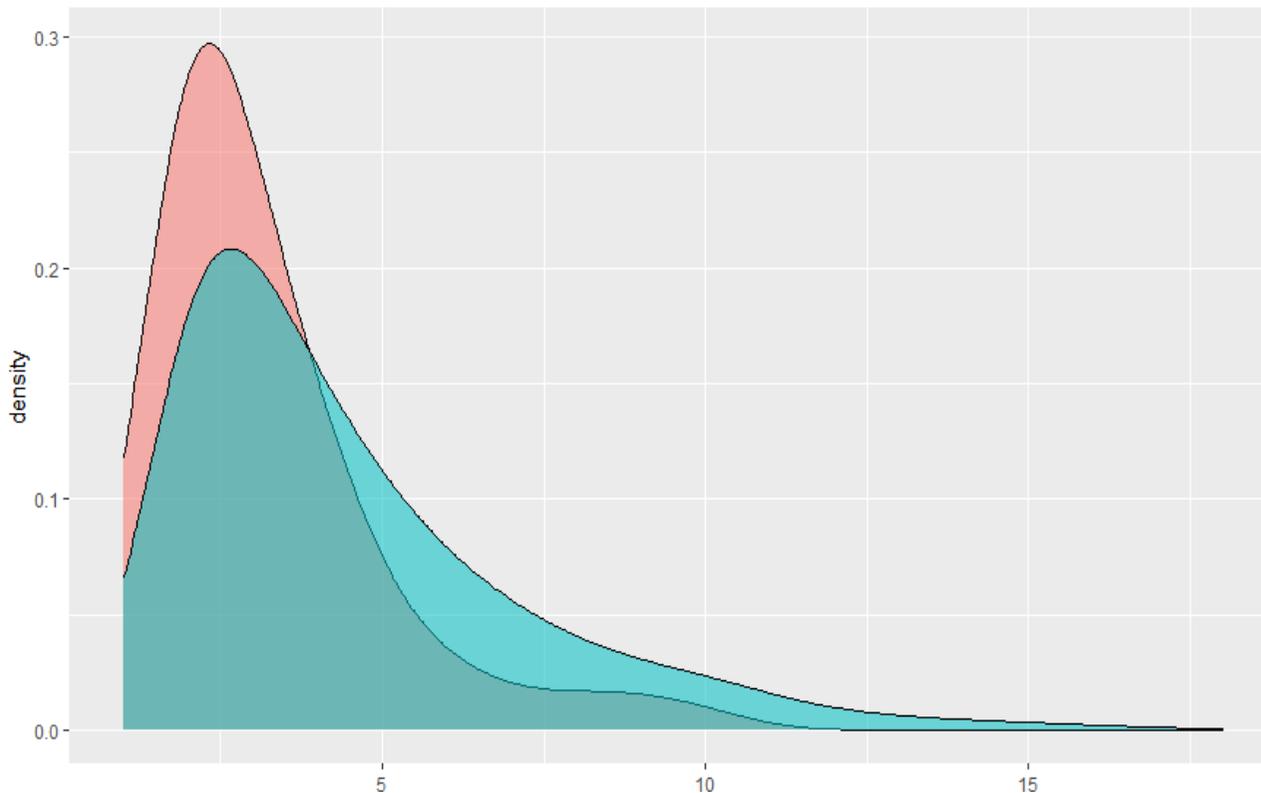


表5.期間内の累積店販購入回数 平均と分散

クランピネス	平均	分散
有り	91.36	5501.45
無し	139.92	9052.29

alpha
□ 0.5

クランピネス

■ クランピネスあり
■ クランピネスなし

図7. クランピネスごとの期間内の累積店販購入回数

4. 提案モデルの概要 <クランピネスごとの比較>

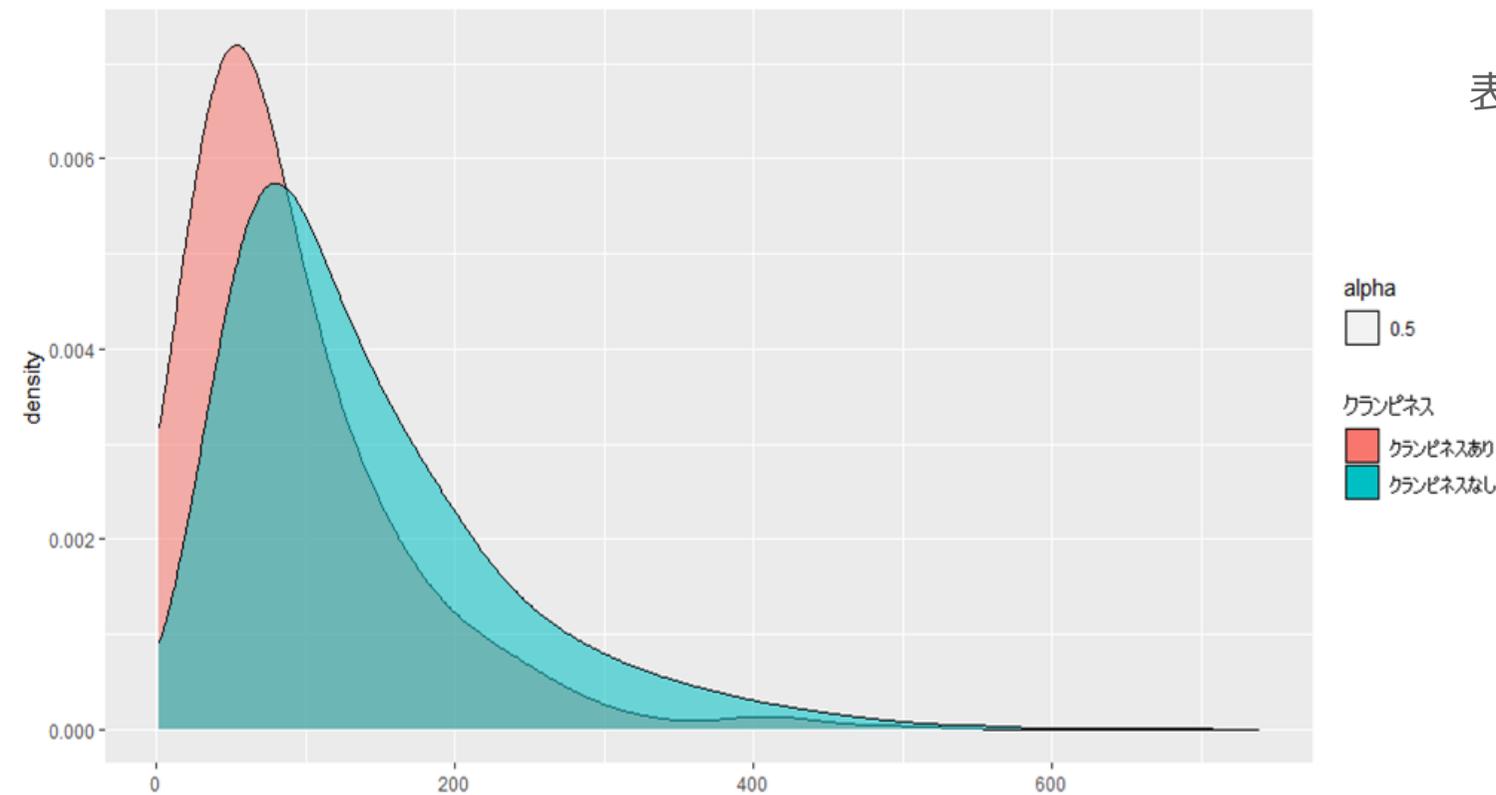


表6.期間内の累積使用金額 平均と分散

クランピネス	平均	分散
有り	3.21	2.97
無し	4.49	7.33

図8. クランピネスごとの期間内の累積使用金額 (千円単位)

4. 提案モデルの概要 <クランピネスごとの比較>

表7. クランピネスごとの前来店時店販売購入有無

	前来店時店販購入有り	前来店時店販購入無し
クランピネス有り	69.8%	30.2%
クランピネス無し	47.8%	52.2%

クランピネスの有無で各変数の傾向に差が見られた。
→クランピネスの有無をグループと考え、
各説明変数の係数をクランピネスの有無ごとに推定する。

4. 提案モデルの概要

本研究では2つの予測モデルを作成し、精度比較を行う。

① クランピネス有無を考慮したモデル
(個人差は考慮しない)

② クランピネス有無+
個人差を考慮したモデル
(切片 = 共通項 + 個人差項)

MCMCを用いてパラメータを推定(Rstan)

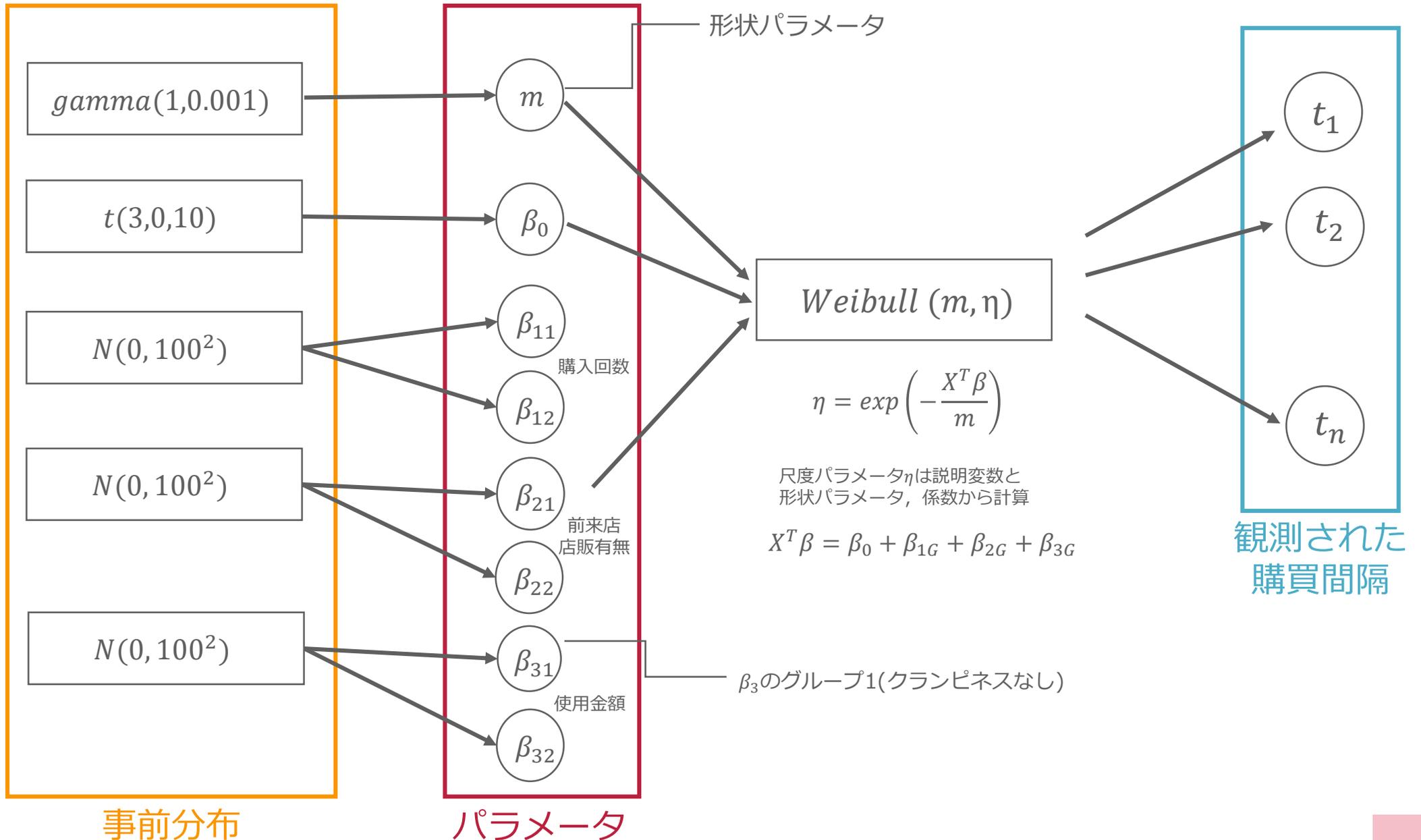
予測精度の比較

4. 提案モデルの概要

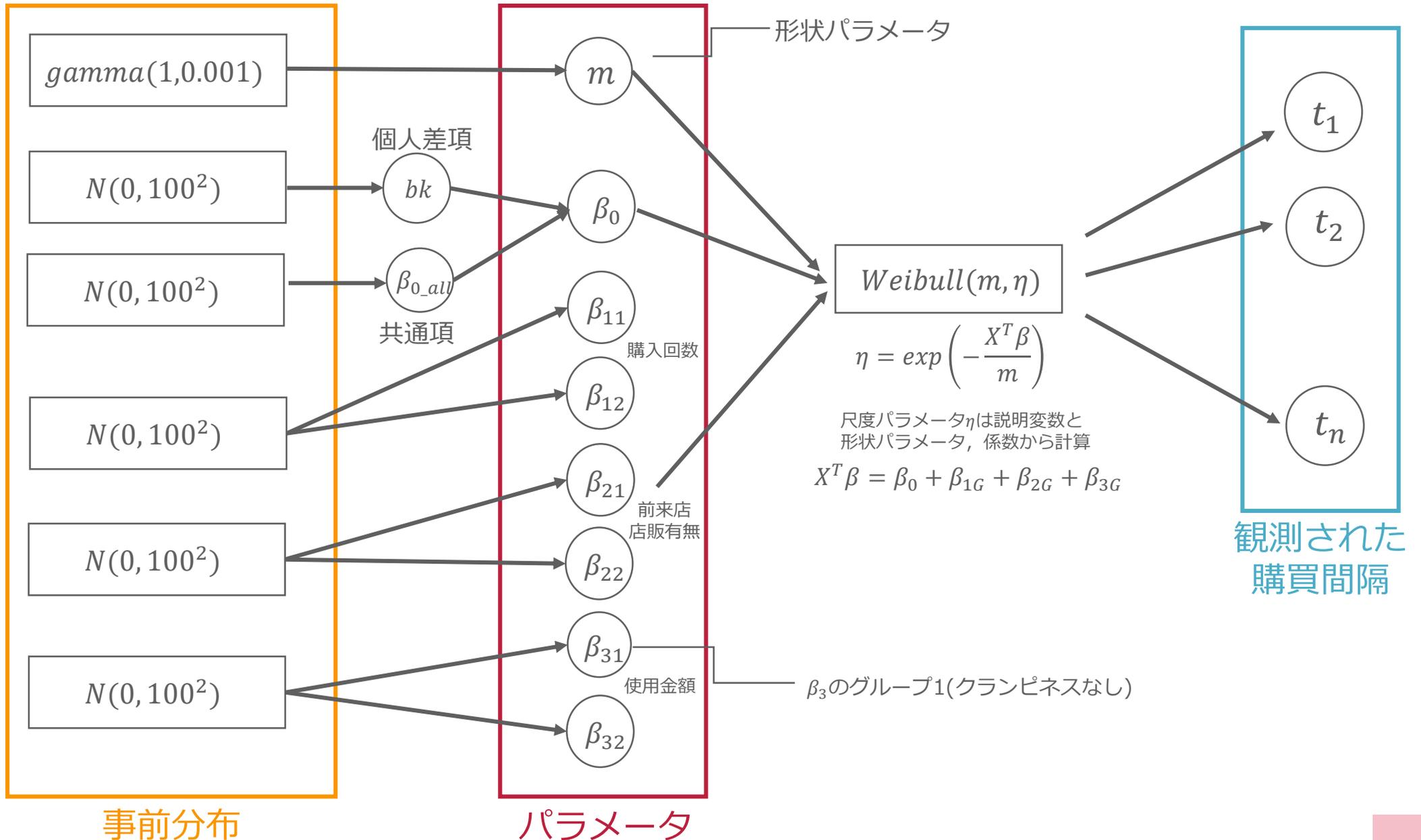
目的変数，説明変数は以下の通りである．

- 目的変数：非購入期間（日）
- 説明変数：
期間内の累積店販購入回数，前来店時店販有無，期間内の累積使用金額
（千円単位）
→ 順に係数の記号は β_1 , β_2 , β_3 とする．
- **グループ**：クランピネスの有無（グループ1：無，グループ2：有）

4. 提案モデルの概要 <モデル①>



4. 提案モデルの概要 <モデル②>



4. 提案モデルの概要

- 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)”により分析用データの作成，モデル作成，精度計算を行った。

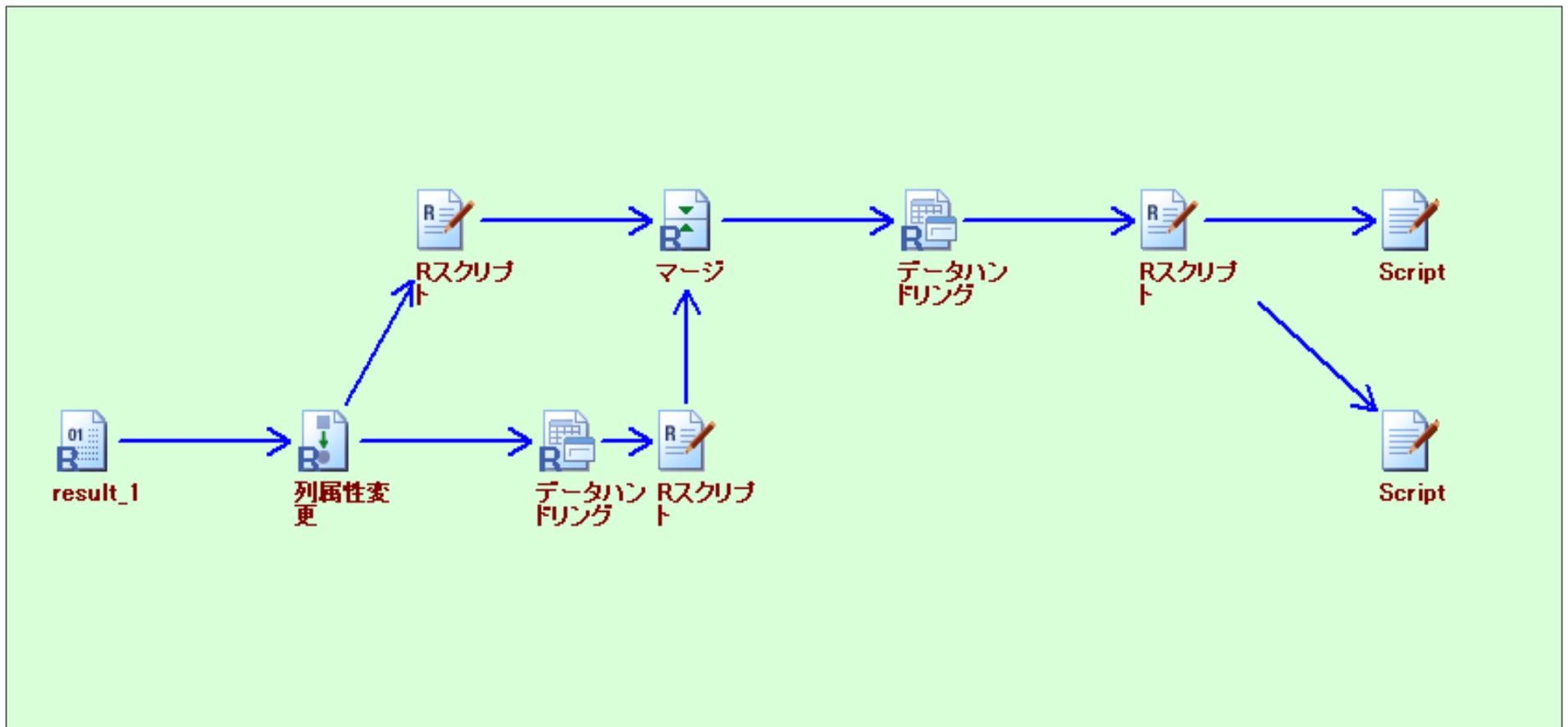


図9. VRPの例（ワイブル分布を仮定した生存時間解析）

5. 結果・精度評価

<精度評価>

- 学習データ（2015年7月～2017年3月）、テストデータ（2017年4月～）を用いて実際に購入がある日数での購入確率を計算。
- flg：予測購入確率50%以上の場合1, 50%未満の場合0
- 正解率：flg=1の数 / データ数

表8. 学習データでの正解率

	flg = 0	flg = 1	正解率
モデル①	5318	2496	0.319
モデル②	5054	2760	0.353

表9. テストデータでの正解率

	flg = 0	flg = 1	正解率
モデル①	0	560	0
モデル②	319	241	0.430

5. 結果・精度評価

<各モデルのパラメータ推定結果（事後平均）>

表10. モデル①推定結果（標準偏差）

クランピネス	形状 パラメータ	β_1	β_2	β_3
有り	1.561 (0.000)	0.189 (0.001)	-0.289 (0.002)	-0.003 (0.000)
無し	1.561 (0.000)	0.214 (0.000)	0.943 (0.000)	0.000 (0.000)

表11. モデル②推定結果（標準偏差）

クランピネス	形状 パラメータ	β_1	β_2	β_3
有り	2.831 (0.031)	-0.427 (0.149)	2.801 (0.348)	-0.017 (0.005)
無し	2.831 (0.031)	0.220 (0.017)	2.378 (0.049)	-0.010 (0.001)

個人差を考慮した場合と考慮しない場合で、

- クランピネス有り顧客の β_1 （累積店販購入回数）、 β_2 （前来店時店販有無）の符号が反対となった。
 - β_2 （前来店時店販有無）の値に差が見られた。

5. 結果・精度評価

<各モデルのパラメータ推定結果（事後平均）>

表10. モデル①推定結果（標準偏差）

クランピネス	形状 パラメータ	β_1	β_2	β_3
有り	1.561 (0.000)	0.189 (0.001)	-0.289 (0.002)	-0.003 (0.000)
無し	1.561 (0.000)	0.214 (0.000)	0.943 (0.000)	0.000 (0.000)

表11. モデル②推定結果（標準偏差）

クランピネス	形状 パラメータ	β_1	β_2	β_3
有り	2.831 (0.031)	-0.427 (0.149)	2.801 (0.348)	-0.017 (0.005)
無し	2.831 (0.031)	0.220 (0.017)	2.378 (0.049)	-0.010 (0.001)

標準偏差について、個人差を考慮したモデル②の方が、個人差を考慮していないモデル①より大きい値となった。

また、どちらのモデルにおいても \hat{R} は全パラメータで1.1以下となり、収束したと判断できる。

6. 考察

<パラメータ推定値に関する考察>

- 表8, 9より, 個人差を考慮したモデル②の方が高い確率で購入を予測できた.
- 表10より, 個人差を考慮していないモデル①では多くのパラメーターの標準偏差が0となった.
- 表11より, 個人差を考慮したモデル②では, 特にクランピネス有り顧客のパラメータにおいて標準偏差がモデル①の結果よりも大きい値となった.

モデル①：顧客の異質性を考慮しなかったため、パラメーター値は全顧客の平均に近くなったと考えられる。

モデル②：特にクランピネス有り顧客について、平均とは異なる購入行動をとる顧客が存在すると考えられる。

→ 顧客の異質性を考慮することで各個人が持つ性格や価値観による購買行動の違いを表現できた。消費者の異質性の導入がモデル精度を改善する上で重要。

6. 考察

<パラメータに関する考察（モデル②）>

- 表11より、クランピネス無し顧客の場合、 β_1 （期間中の累積店販購入回数）、 β_2 （前来店時店販有無）の符号が正と推定された。
- 一方クランピネス有り顧客の場合、 β_1 （累積店販購入回数）の符号は負と推定された。

クランピネス無し顧客：累積店販購入回数が多い顧客、前来店時店販有無が有りの顧客は、より短い間隔で店販商品を購入すると考えられる。

クランピネス有り顧客：店販購入回数について、一般的な顧客とは対照的な特徴を持つと考えられる。

7. まとめ・今後の課題

<まとめ>

- ワイブル分布を仮定した生存時間解析により、各顧客の店販購入間隔を分析した。
- 顧客の個人差を考慮することで、モデル精度を改善することができた。
- クランピネスの有無によって、購入間隔が短い顧客の特徴に差が見られた。

<今後の課題>

- グループ差について、クランピネス以外の要素も考慮する必要があると考えられる。
- 本研究で構築したモデルの精度は高いとは言えず、改善の余地があると考えられる。

参考文献

- [1] 厚生労働省, “平成30年度衛生行政報告例の概況”, 厚生労働省HP, 2020-10-31, https://www.mhlw.go.jp/toukei/saikin/hw/eisei_houkoku/18/, (2020-10-2参照)
- [2] 三富悠紀, “消費者の異質性を考慮した推薦商品カテゴリーの検討”, 東京大学経済学研究会, Vol. 60, pp.1-13 (2018)
- [3] Easy to type, “Stanで生存時間解析(Weibull 回帰)”, Hatena Blog, 2017-08-05, <https://ajhjhaf.hatenablog.com/entry/2017/08/05/194939>, (2020-10-2参照)
- [4] 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, Vol.54, No.2, pp.445-459 (2006)
- [5] Zhang Y, Bradlow ET, Small DS (2015) Predicting customer value using clumpiness : From RFM to RFMC. Marketing Science 34(2): 195208.
- [6] 中山 雄司, “顧客関係管理研究の新動向 -来店/購買間隔の不均一性を測るクランピネス指標-”, 甲南経営研究, Vol. 57, No.2, pp. 161-181 (2016)
- [7] 新美 潤一郎, 星野 崇宏, “RFMC分析におけるClumpiness指標の拡張と自社顧客の行動予測への応用 Clumpinessを活用した離脱時期と競合利用の予測手法の提案”, 日本行動計量学会, Vol.92, No.47, pp.27-40 (2020)

参考文献

- [8] 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・サイエンス学会論文誌, Vol.22, No.1, pp.13-29 (2014)
- [9] 津田 尚胤,, 貝戸 清之, 山本 浩司, 小林 潔司, “ワイブル劣化ハザードモデルのベイズ推計法”, 土木学会論文集 F , Vol.62, Np.3, pp. 473-491 (2006)
- [10] gaiaskyの技術メモ, “ベイズ生存時間分析(Weibull分布)”, Hatena Blog, 2018-11-11, <https://gaiasky.hatenablog.com/entry/2018/11/11/234023>, (2020-10-2参照)
- [11] バナナでもわかる話, “階層ベイズ生存解析を用いたwebサイトの訪問者分析に関するStanでの実装”, Hatena Blog, 2019-2-17, <https://www.bananarian.net/entry/2019/02/17/090000>, (2020-10-2参照)
- [12] Kazuki Yoshida, “Bayesian Survival Analysis 1: Weibull Model with Stan”, 2018-10-31, https://rstudio-pubs-static.s3.amazonaws.com/435225_07b4ab5afa824342a4680c9fb2de6098.html, (2020-10-2参照)

精度評価（補足）

$t = 217$ の購入確率

➤ 実際に購入がある日数での購入確率

顧客ID	間隔	説明変数・・・	shape	beta0	beta1	beta2	beta3	scale	非購入確率	購入確率	flg
141	217	・・・	2.830	-16.257	0.220	2.379	-0.010	321.446	0.43	0.57	1
141	158	・・・	2.830	-16.257	0.220	2.379	-0.010	133.872	0.21	0.79	1
141	91	・・・	2.830	-16.257	0.220	2.379	-0.010	129.253	0.45	0.55	1
146	158	・・・	2.830	-16.257	0.220	2.379	-0.010	129.253	0.61	0.39	0
146	80	・・・	2.830	-13.807	0.220	2.379	-0.010	152.147	0.60	0.40	0
146	154	・・・	2.830	-13.807	0.220	2.379	-0.010	67.361	0.50	0.50	1
146	86	・・・	2.830	-13.807	0.220	2.379	-0.010	194.002	0.43	0.57	1
146	131	・・・	2.830	-13.807	0.220	2.379	-0.010	86.059	0.45	0.55	1
330	301	・・・	2.830	-13.807	0.220	2.379	-0.010	224.851	0.25	0.75	1
621	112	・・・	2.830	-13.807	0.220	2.379	-0.010	96.989	0.48	0.52	1

パラメータ推定結果（一部）

表12. モデル①パラメータ推定結果

	mean	se_mean	sd	2.50%	25%	50%	75%	97.50%	n_eff	Rhat
shape	1.561	0.000	0.015	1.532	1.551	1.561	1.571	1.591	10930.216	1.000
β_0	-9.032	0.001	0.092	-9.214	-9.094	-9.032	-8.970	-8.849	10696.541	1.000
β_{11}	0.214	0.000	0.006	0.202	0.210	0.214	0.219	0.227	14104.340	1.000
β_{12}	0.189	0.001	0.046	0.097	0.157	0.189	0.220	0.278	6435.001	1.001
β_{21}	0.943	0.000	0.030	0.885	0.923	0.943	0.963	1.001	13929.030	1.000
β_{22}	-0.289	0.002	0.124	-0.530	-0.374	-0.288	-0.204	-0.047	4263.343	1.001
β_{31}	0.000	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	18458.544	1.000
β_{32}	-0.003	0.000	0.001	-0.006	-0.004	-0.003	-0.002	-0.001	11326.553	1.000

パラメータ推定結果（一部）

表13. モデル②パラメータ推定結果

	mean	se_mean	sd	2.50%	25%	50%	75%	97.50%	n_eff	Rhat
shape	2.831	0.001	0.031	2.770	2.811	2.832	2.853	2.892	1087.213	1.007
β_{11}	0.220	0.001	0.017	0.187	0.209	0.220	0.231	0.253	301.075	1.013
β_{12}	-0.427	0.013	0.149	-0.722	-0.525	-0.424	-0.325	-0.140	139.502	1.051
β_{21}	2.378	0.001	0.049	2.281	2.344	2.378	2.411	2.475	1212.088	1.004
β_{22}	2.801	0.026	0.348	2.118	2.562	2.799	3.037	3.482	181.471	1.012
β_{31}	-0.010	0.000	0.001	-0.011	-0.011	-0.010	-0.010	-0.009	261.664	1.014
β_{32}	-0.017	0.001	0.005	-0.027	-0.021	-0.018	-0.014	-0.006	99.784	1.060

トレースプロット (一部)

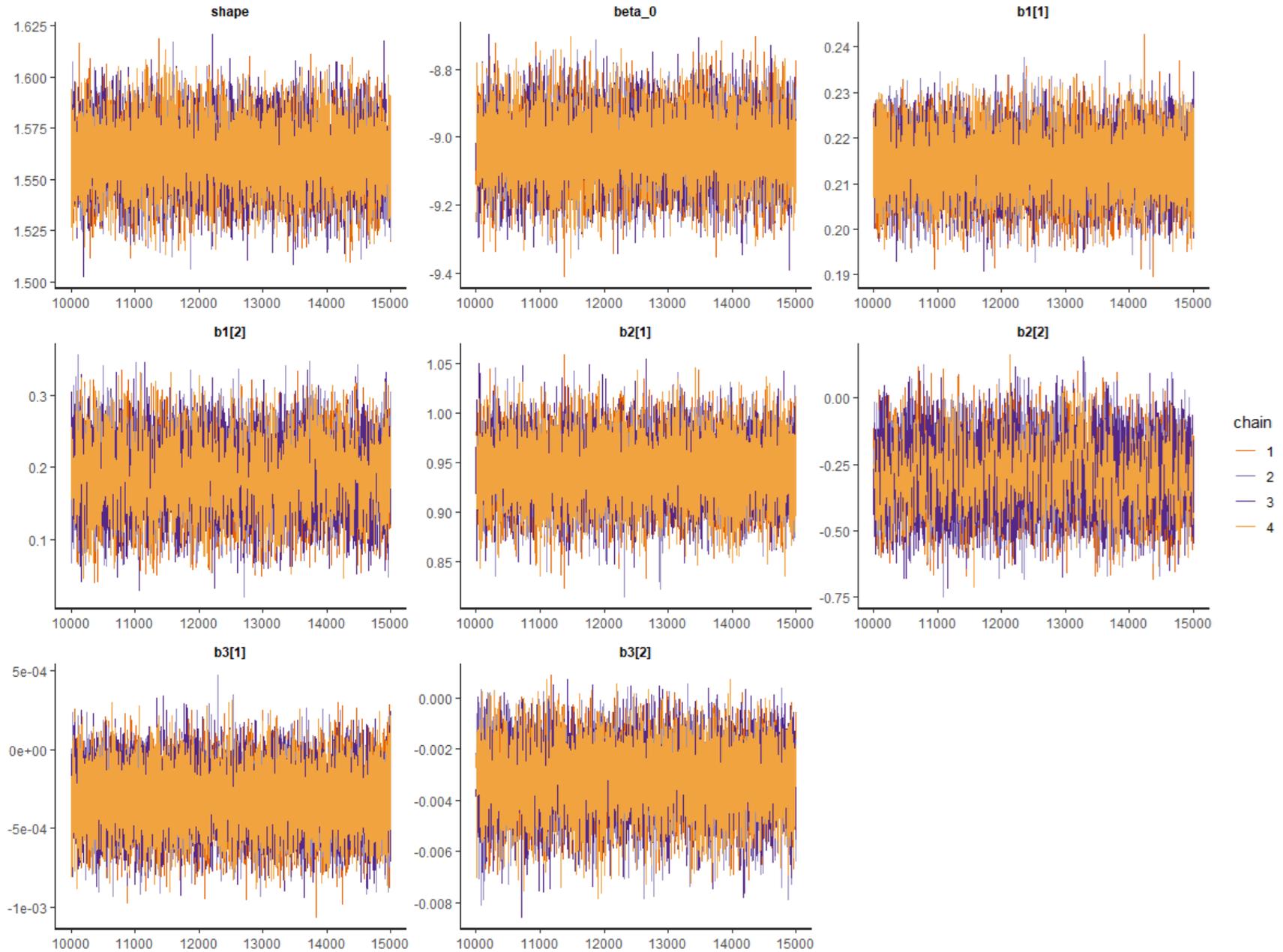


図10. モデル①のトレースプロット (一部)

トレースプロット (一部)

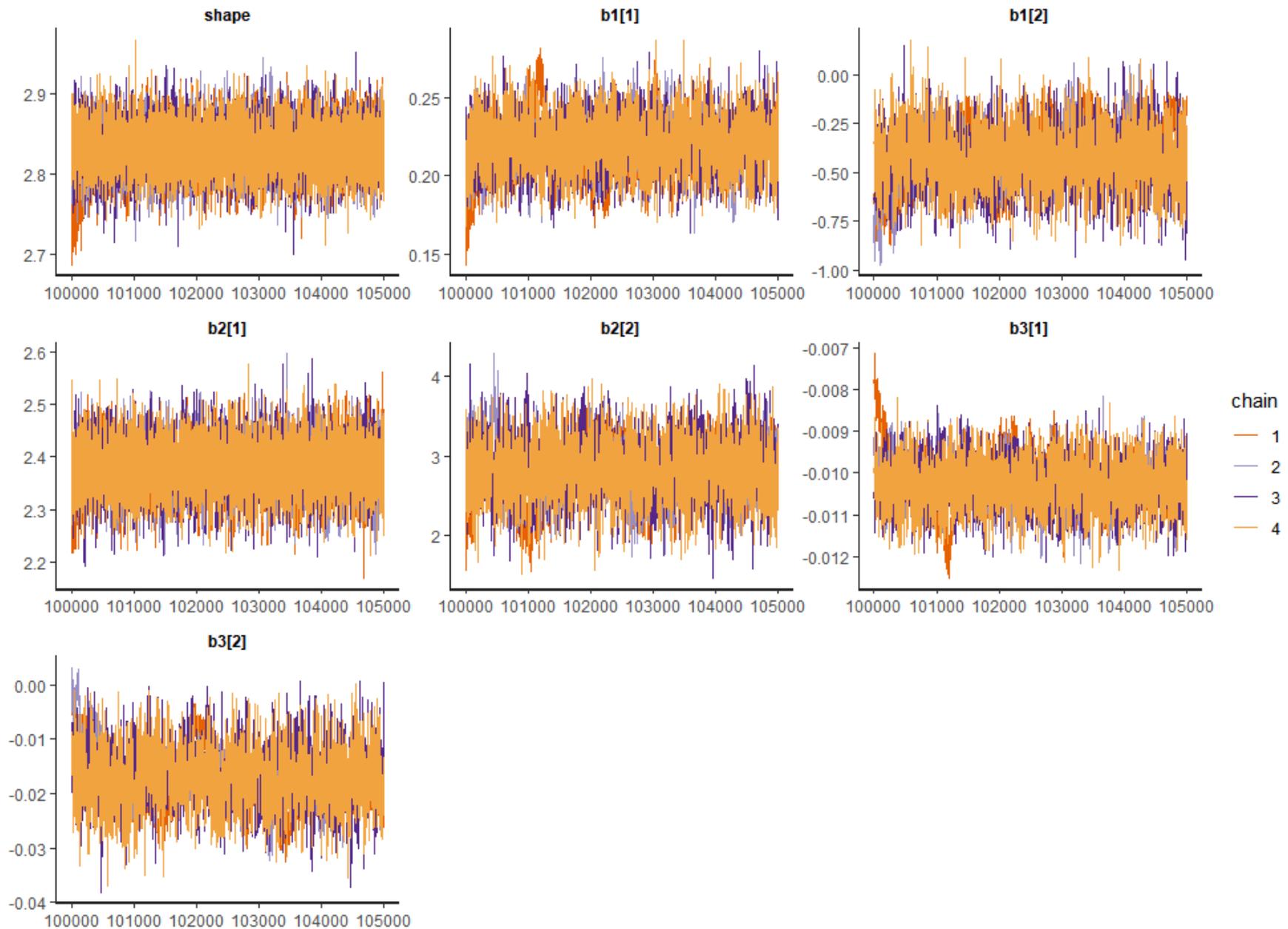


図11. モデル②のトレースプロット (一部)

事後分布 (一部)

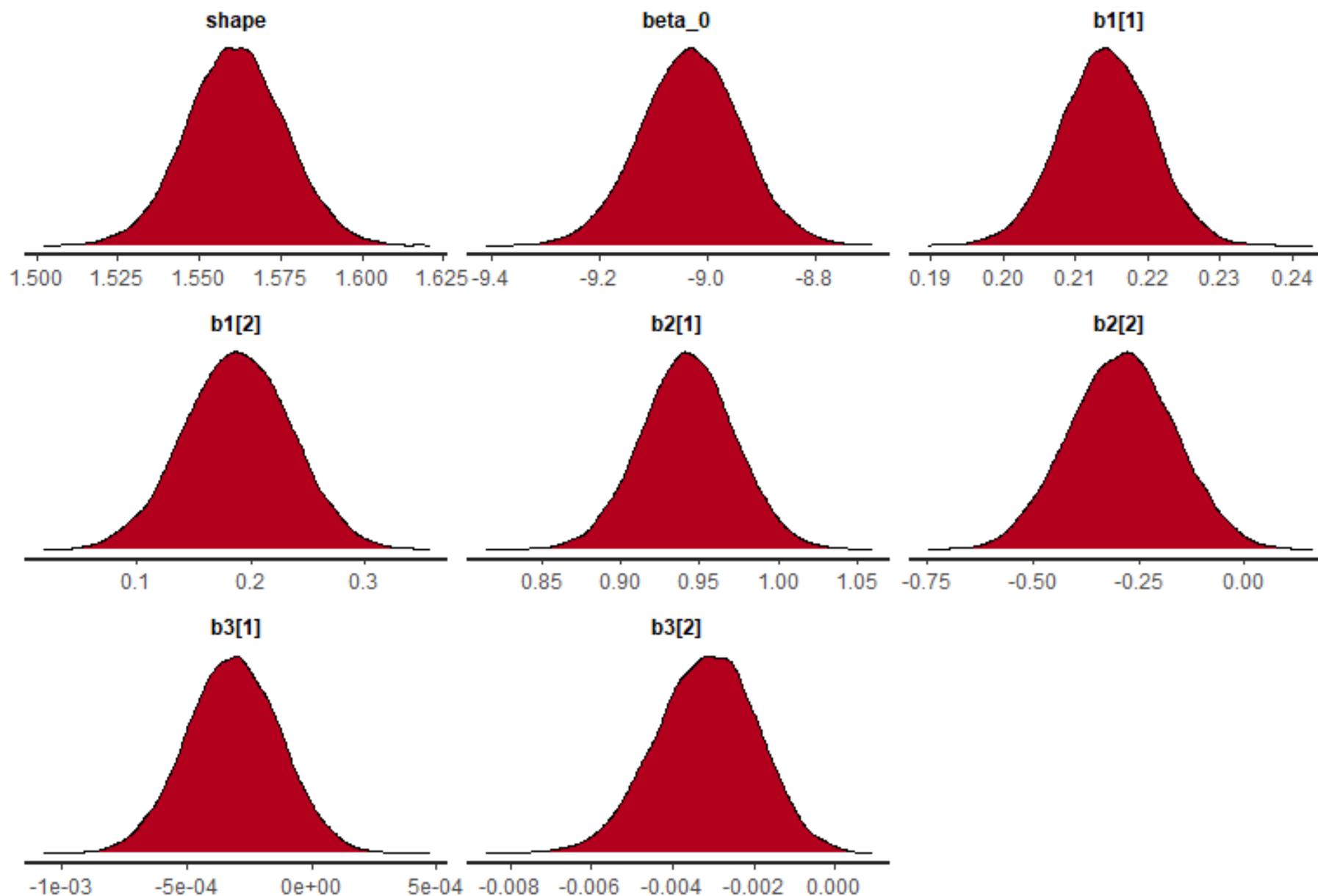


図12. モデル①の事後分布 (一部)

事後分布 (一部)

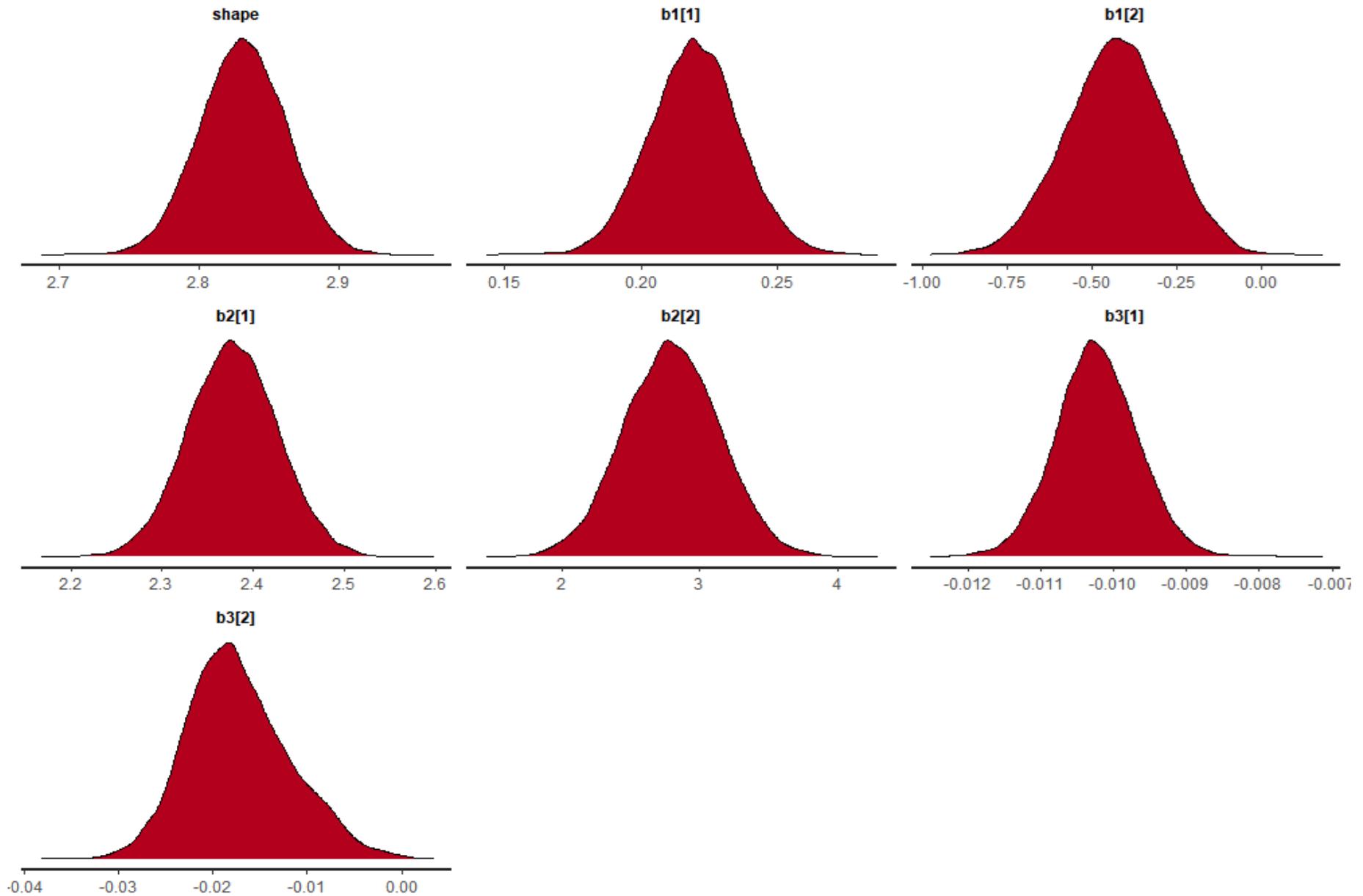


図13. モデル②の事後分布 (一部)