

マーケティング・リサーチ特論 ～潜在クラスロジットモデル～

2026年度1学期： 金曜3限
担当教員： 石垣 司

現代的なマーケティングの時代背景

消費価値観の変化に準じてマーケティングも進化

画一的価値観の市場
(1950-70年代の日本)

価値観の分化した市場
(1980-90年代の日本)

価値観の多様化した市場
(2000年-現在の日本)

十人一色の時代
大量消費, 機能に価値
マス・マーケティング

マスからの脱却
ライフスタイルの共有に価値
セグメンテーション・マーケティング

十人十色の時代
個人の異なる価値, 価値共創
one to oneマーケティング



「消費する大衆」から
「生活者」へ



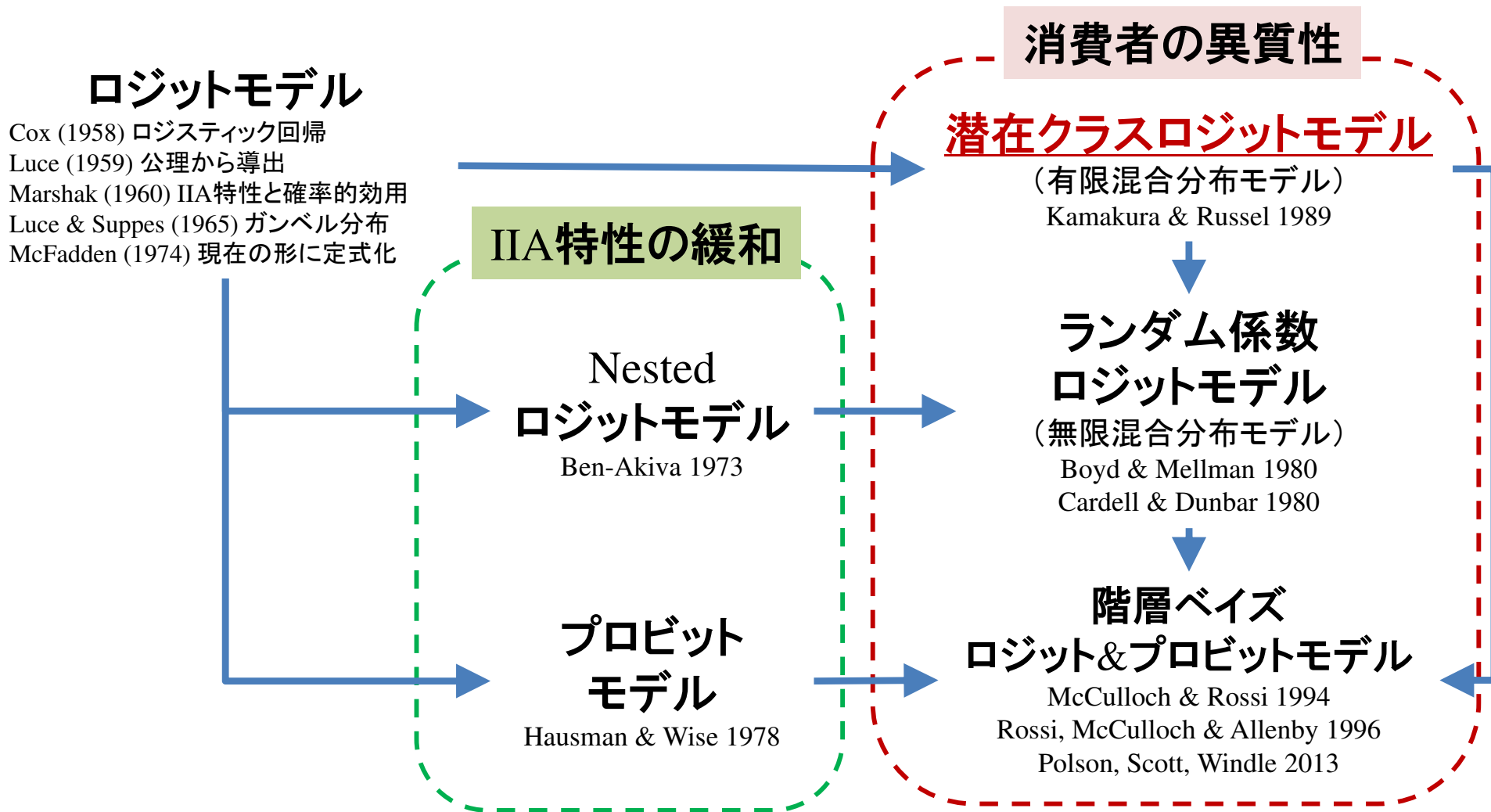
$$\text{効用: } U_{itj} = f(\mathbf{b}, x_{itj}) + e_{itj}$$

$$\text{効用: } U_{itj} = f(\mathbf{b}_s, x_{itj}) + e_{itj}$$

$$\text{効用: } U_{itj} = f(\mathbf{b}_i, x_{itj}) + e_{itj}$$

ブランド選択モデルの流れ

本講義で紹介する多項ブランド選択モデル



潜在クラスモデルとは？

観測データを少数のセグメントに分類・判別するための統計的モデリング法の総称

- 有限混合分布モデル
- 混合正規分布モデル (GMM: Gaussian Mixture Model)
- EM アルゴリズムによる推定

マーケティング・リサーチでは？

- 潜在クラスモデルはセグメンテーションの概念と相性がよい
- 選択行動に関する消費者セグメントをデータから自動的に見つけ出す
- 潜在クラスロジットモデル
 - 有限混合分布モデルと多項ロジットモデルを統合した統計モデル
 - 消費者の異質性を取り入れた離散選択モデル

有限混合分布モデル

複数の異なるセグメントからデータが発生している状態を統計的にモデリング

- セグメント: $s = 1, \dots, S$
- セグメント数 S の目安: $1 < S \ll$ “サンプルサイズ”

データ x_i を発生する確率密度関数

$$p(x_i | \theta) = \sum_{s=1}^S \pi_s f(x_i | \theta_s), \theta = \{\pi_1, \dots, \pi_S, \theta_1, \dots, \theta_S\}$$

- セグメント s 毎に異なるパラメータ $\{\pi_s, \theta_s\}$ をもつ
- パラメータの制約 $\pi_s \geq 0, \sum_{s=1}^S \pi_s = 1$

関数 f のモデリングで多様な分布を表現する

混合正規分布モデル(GMM)

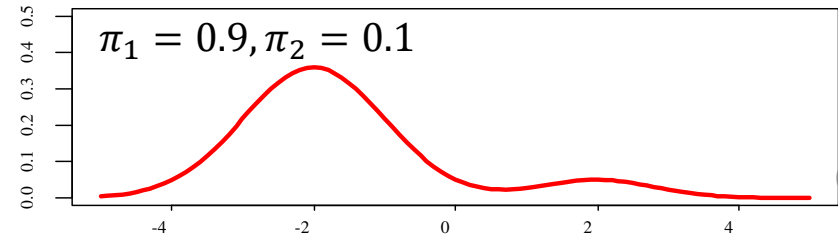
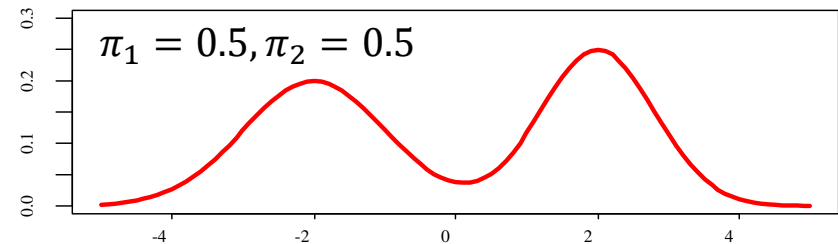
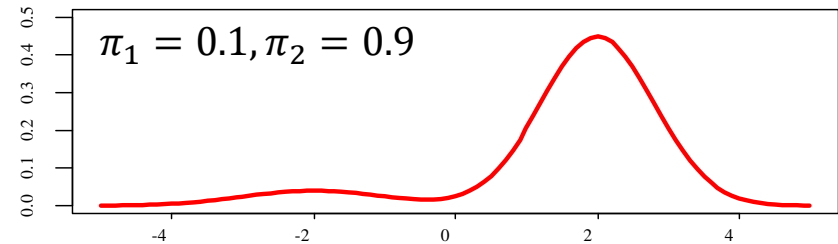
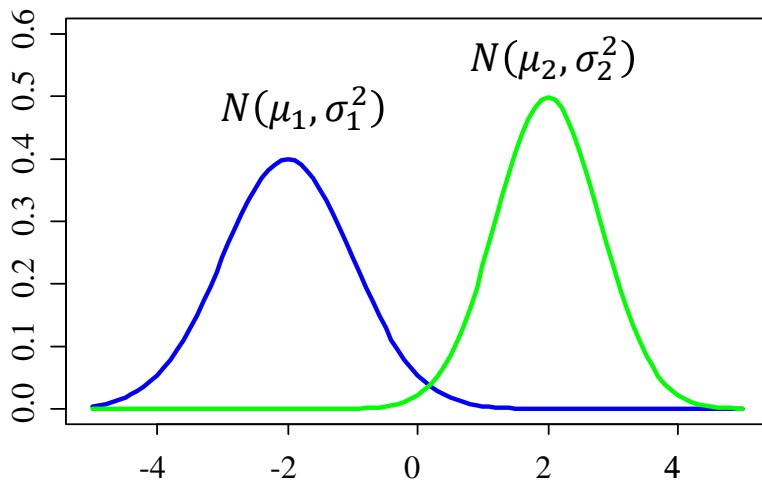
関数 f を正規分布(多変量正規分布)でモデリング

– S 個の正規分布の重み付き組み合わせで一つの分布を形成

$$p(x_i|\theta) = \sum_{s=1}^S \pi_s N(x_i|\theta_s)$$

混合正規分布

2つの正規分布($S = 2$)の例
($\theta_s = \{\mu_s, \sigma_s^2\}$)



多変量正規分布の場合、パラメータは平均ベクトルと分散共分散行列 $\theta_s = \{\mu_s, \Sigma_s\}$

GMMのセグメント所属確率

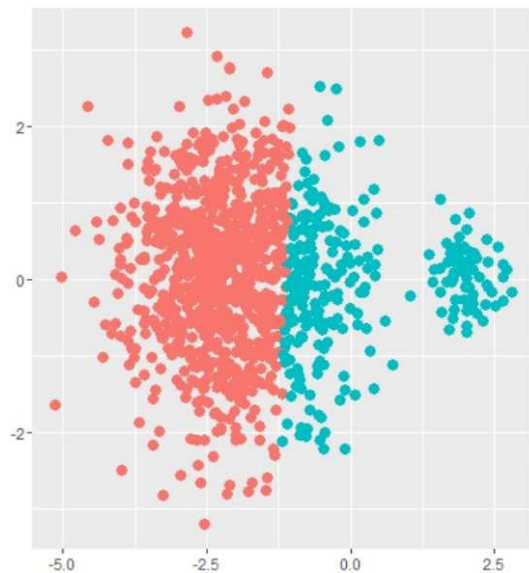
各データ x_i がセグメント s へ所属する確率

– 推定されたパラメータを利用して算出

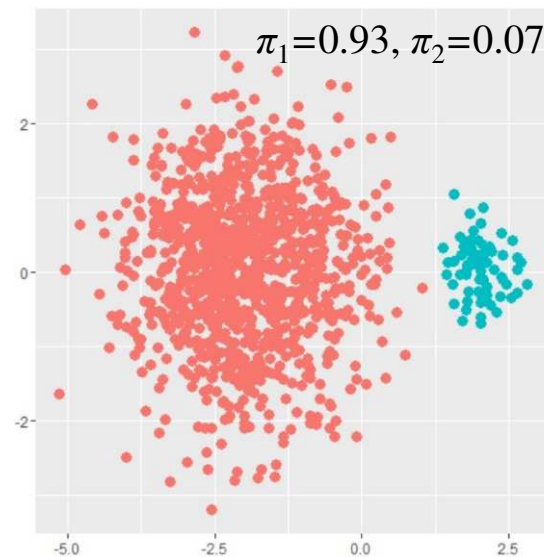
$$\Pr(i_s = s | x_i) = p_{is} = \frac{\pi_s N(x_i | \theta_s)}{\sum_{k=1}^S \pi_k N(x_i | \theta_k)}$$

クラスタリングの例: Imbalanced data

– k -means では失敗する例でもGMMでは成功



Kmeans
● 1
● 2



GMM
● 1
● 2

- library(mclust)
- gmm =
Mclust(data,G=2,ModelNames="EII")

GMMのパラメータ推定

GMMの尤度最大化

– 対数尤度: $\log L(\{x_i\}) = \sum_{i=1}^N \log \sum_{s=1}^S \pi_s N(x_i | \mu_s, \Sigma_s)$

– GMMの尤度最大化は不良設定問題

原因: 対数尤度関数の中の s に関する和の計算

⇒ ラグランジュ未定乗数法では陽な解は得られない

(各パラメータの推定値はパラメータ $\{\pi_s\}$ の値に依存する)

多変量混合正規分布のラグランジュ未定乗数法の解

$$\mu_k = \left\{ \sum_{i=1}^N \left(\frac{\pi_k N(x_i | \theta_k)}{\sum_{s=1}^S \pi_s N(x_i | \theta_s)} \right) \right\}^{-1} \sum_{i=1}^N \left(\frac{\pi_k N(x_i | \theta_k)}{\sum_{s=1}^S \pi_s N(x_i | \theta_s)} x_i \right)$$

$$\Sigma_k = \left\{ \sum_{i=1}^N \left(\frac{\pi_k N(x_i | \theta_k)}{\sum_{s=1}^S \pi_s N(x_i | \theta_s)} \right) \right\}^{-1} \sum_{i=1}^N \left(\frac{\pi_k N(x_i | \theta_k)}{\sum_{s=1}^S \pi_s N(x_i | \theta_s)} (x_i - \mu_k)(x_i - \mu_k)^T \right)$$

$$\pi_k = N^{-1} \sum_{i=1}^N \left(\frac{\pi_k N(x_i | \theta_k)}{\sum_{s=1}^S \pi_s N(x_i | \theta_s)} \right)$$

EMアルゴリズム

潜在変数を含む統計モデルの最尤推定法

- 反復計算。各反復で必ず尤度が増加
- 汎用性が高く、多様な応用成果あり
- 局所最適解には収束する(全体最適解に収束する保証は無い)

#メモ 潜在変数を含む統計モデルのパラメータ推定法の定番。理論的にも深い背景があり、この周辺を学習することで統計モデルとパラメータ推定に関する理解が深まる。

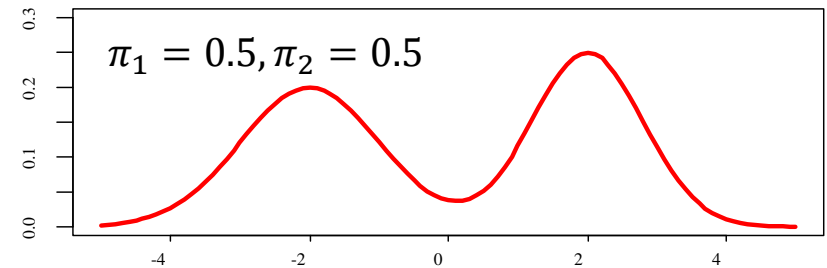
GMMのためのEMアルゴリズム

1. パラメータの初期値 $\theta^{(0)} = \{\mu_k^{(0)}, \Sigma_k^{(0)}, \pi_k^{(0)}\}$ を与える ($t = 0$)
2. E-step: $p_{ik}^{(t)} = \frac{\pi_k^{(t)} N(x_i | \mu_k^{(t)}, \Sigma_k^{(t)})}{\sum_{s=1}^S \pi_k^{(t)} N(x_i | \mu_k^{(t)}, \Sigma_k^{(t)})}$ の計算
3. M-step: $p_{ik}^{(t)}$ を用いて $\theta^{(t+1)} = \{\mu_k^{(t+1)}, \Sigma_k^{(t+1)}, \pi_k^{(t+1)}\}$ を更新
4. 収束していたら終了。していなければ, E-step に戻る

有限混合分布から発生するデータの解釈

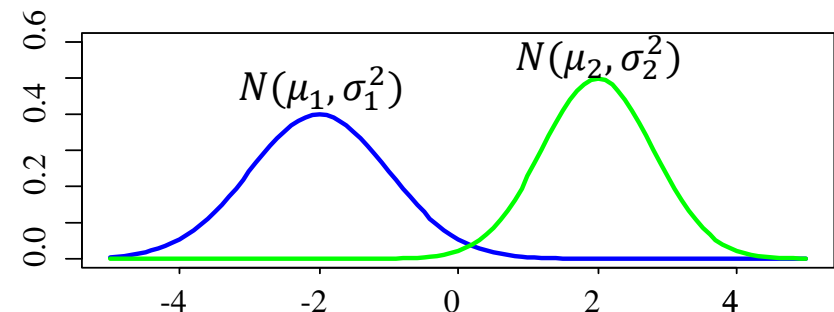
解釈1

- 多峰性の複雑な分布から各データ x_i が発生している



解釈2

- 各データ x_i は正規分布群の中の一つから発生していて、その所属確率は $\{\pi_s\}$ と尤度に依存する



マーケティング・リサーチでは？

- マーケティング・セグメンテーションの概念との整合性から、解釈2で理解するが多い

潜在クラスロジットモデル #1

S 個の多項ロジットモデルの有限混合分布

Kumakura & Russel, A Probabilistic Choice Model for Market Segmentation and Elasticity Structure, J. of Marketing Research 26(4) 1989

- 各セグメント s 毎に異なる反応係数 b_s を設定し、セグメント間での異なる選択行動を表現できる離散選択モデル
- 各セグメント s の反応係数 $b_{sp} = 0$ について検定ができる
機械学習的な分類・判別・クラスタリング等では検定は行えない



【抽出できる情報の例】

セグメントAはブランドAが200円でB・Cが180円の時のブランドAの選択確率30%

セグメントBはブランドAが200円でB・Cが180円の時のブランドAの選択確率75%

潜在クラスロジットモデル #2

消費者 i がセグメント s に所属するときの選択肢 j の効用の確定項

$$V_{itj|s} = b_{sj0} + b_{s1}x_{itj1} + \dots + b_{sP}x_{itjP}$$

消費者 i の時刻 t での選択肢 j の選択確率

$$\Pr(Y_{it} = j) = \sum_{s=1}^S \pi_s \frac{\exp(V_{itj|s})}{\sum_{l=1}^J \exp(V_{itl|s})}$$

消費者 i のセグメント s への所属確率

- 消費者 i が時刻 t で選択肢 j を選択したとき $y_{itj} = 1$, それ以外 $y_{itj} = 0$
- 消費者 i が s に所属のときの尤度関数: $L_{i|s} = \prod_t^T \prod_j^J \left(\frac{\exp(V_{itj|s})}{\sum_{l=1}^J \exp(V_{itl|s})} \right)^{y_{itj}}$
- 所属確率: $\Pr(i_s = s) = \frac{\pi_s L_{i|s}}{\sum_{k=1}^S \pi_k L_{i|k}}$

Rでの分析例 #1

“flexmix” パッケージ

- 有限混合分布をEMアルゴリズムで解くパッケージ

再掲：ケチャップの購買履歴データの分析

Jain DC, Vilcassim NJ, Chintagunta PK (1994). “A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data.” Journal of Business & Economic Statistics, 12(3), 317-328.

- ブランド：Heinz 28, Heinz 32, Heinz 41, Hunt’s 32

Heinz 28, 32 & 41 は同じブランドでサイズのみが異なる

米国ミズーリ州スプリングフィールドの300世帯で2年間に観測された
2,798 ケチャップの購買履歴(スキャンパネルデータ)

- 説明変数：

価格(prices of all brands in the product category)

ディスプレイ(special displays of brands in the store)

チラシ(newspaper feature advertisements)



Heinz: <https://www.heinz.com/products?condimentType%5B0%5D=ketchup>
Hunt's: <https://www.hunts.com/ketchup>

Rでの分析例 #2

セグメント数 2 での潜在クラスロジットモデルの R コードの例

```
library(mlogit)
library(flexmix)
data("Catsup", package = "mlogit")
Catsup$t = seq_len(nrow(Catsup))
vnames = c("display", "feature", "price")
Cdata = reshape(Catsup, idvar = c("id", "t"), times = c("heinz41", "heinz32", "heinz28", "hunts32"),
  timevar = "brand", varying = matrix(colnames(Catsup)[2:13], nrow = 3, byrow = TRUE),
  v.names = vnames, direction = "long")
Cdata$choice = with(Cdata, choice == brand)
Cdata = Cdata[, c("id", "choice", "t", vnames, "brand")]
Cdata$brand <- relevel(factor(Cdata$brand), "heinz32")
set.seed(1234)
m2 = flexmix(choice ~ display + feature + price + brand | id, model = FLXMRcondlogit(strata = ~ t), data = Cdata,
k = 2)
Summary(m2)
parameters(m2)
m2_results = refit(m2) #係数の検定結果を返す処理だが、計算時間がかかるので注意
summary(m2_results)
plot(m2_results)
```

Rでの分析例 #3

潜在クラスモデルの推定の特徴

- 推定結果はアルゴリズムの初期値に依存する
- セグメント数 S は分析者が与える必要がある

情報量規準によるセグメント数の決定

各セグメント数で異なる初期値で10回推定した結果の各指標の平均値

セグメント数	1	2	3	4	5	6	7	8	9
対数尤度	-2517	-2252	-2129	-2066	-2041	-2030	-2025	-2019	-2019
AIC	5047	4531	4298	4187	4150	4140	4137	4134	4136
BIC	5091	4626	4445	4385	4399	4430	4458	4486	4487

- 理論的にはセグメント数の増加で尤度も単調増加する
- 以下, BIC が最適なセグメント数 4 の結果を見ていく

#メモ ここでAICではなくBICを採用する理由。セグメント数8よりも4の方が解釈しやすそうだからというただの主観。ただし、AICとBICは指標としての意味が異なる点には注意

Rでの分析例 #3

セグメント数 $S = 4$ での分析結果

– セグメント間の係数の比較からセグメントの特徴を理解

```
$Comp.1
      Estimate Std. Error z value Pr(>|z|)
display    1.77215   0.20501   8.6441 < 2.2e-16 ***
feature    0.81248   0.26600   3.0545 0.002255 **
price     -1.40988   0.12901 -10.9286 < 2.2e-16 ***
brandheinz28 -0.50997   0.15108  -3.3754 0.000737 ***
brandheinz41 -1.32032   0.22269  -5.9289 3.05e-09 ***
brandhunts32 -3.90526   0.29083 -13.4280 < 2.2e-16 ***

$Comp.2
      Estimate Std. Error z value Pr(>|z|)
display    0.42737   0.25497   1.6761 0.09371 .
feature    1.43071   0.28160   5.0806 3.763e-07 ***
price     -2.54933   0.19501 -13.0726 < 2.2e-16 ***
brandheinz28 3.55475   0.31007  11.4645 < 2.2e-16 ***
brandheinz41 1.71078   0.32377   5.2840 1.264e-07 ***
brandhunts32 -0.89603   0.18404  -4.8687 1.123e-06 ***

$Comp.3
      Estimate Std. Error z value Pr(>|z|)
display    1.019516   0.357155   2.8545 0.004310 **
feature    1.604525   0.394020   4.0722 4.657e-05 ***
price     -0.054341   0.219566  -0.2475 0.804526
brandheinz28 2.172853   0.390777   5.5603 2.692e-08 ***
brandheinz41 1.335058   0.416480   3.2056 0.001348 **
brandhunts32 -0.722528   0.469274  -1.5397 0.123640

$Comp.4
      Estimate Std. Error z value Pr(>|z|)
display    0.49503   0.34705   1.4264 0.1537528
feature    1.10957   0.32921   3.3704 0.0007507 ***
price     -2.29902   0.21348 -10.7692 < 2.2e-16 ***
brandheinz28 0.82388   0.27402   3.0067 0.0026414 **
brandheinz41 0.25972   0.43072   0.6030 0.5465111
brandhunts32 0.38920   0.18463   2.1080 0.0350291 *
```

```
      prior size post>0 ratio
Comp.1 0.398 5260   9004 0.584
Comp.2 0.307 3096   8484 0.365
Comp.3 0.134 1276   4092 0.312
Comp.4 0.161 1560   7040 0.222
```

```
'log Lik.' -2066.434 (df=27)
AIC: 4186.868   BIC: 4384.588
```

これ以降のスライドでは各セグメントの
所属サイズ順にセグメント番号を振り直す

Rでの分析例 #3

セグメント数 $S = 4$ での分析結果

	所属比	Display	Feature	Price	Heinz28	Heinz41	Hunt's 32
セグメント1	40 %	1.77 *	0.81 *	-1.41 *	-0.51 *	-1.32 *	-3.91 *
セグメント2	31 %	0.42	1.43 *	-2.55 *	3.56 *	1.71 *	-0.90 *
セグメント3	16 %	0.50	1.11 *	-2.30 *	0.82 *	0.25	0.34 *
セグメント4	13 %	1.02 *	1.61 *	-0.05	2.17 *	1.34 *	-0.72

セグメント1：最も大きいセグメント。3種類すべてのマーケティング変数が有意に効果的。Heinz 32 のブランド価値が高い

セグメント2：2番目に大きいセグメント。チラシと値下げに有意に反応。Heinz 28 のブランド価値が高い

セグメント3：チラシと値下げに有意に反応。32oz で比較した場合、Heinz よりも Hunt's ブランドのブランド価値が高い唯一のセグメント

セグメント4：チラシと値下げに有意に反応。値下げの効果が薄い可能性のある唯一のセグメント