

アンケートデータを用いた各種グルーピング方法の比較

2003年11月

株式会社インタースコープ

天辰 次郎 jiro@interscope.co.jp

高田 徹 takata@interscope.co.jp

目次

1. はじめに

- 本講演の内容
- 会社説明

2. マーケティング・セグメンテーション論

- マーケティングにおけるセグメンテーション論の推移

3. クラスタ分析概要

- クラスタ分析とは?
 - 階層クラスタ分析の特徴
 - 非階層クラスタ分析の特徴
- 階層と非階層の使い分け

4. 階層クラスタ分析

- 距離選択方法
- クラスタ化法の特徴

5. 非階層(分割型)クラスタ分析

- 初期シードの選択
- クラスタ数の選択
- 元データの処理

6. 非階層、階層クラスタ分析のコンビネーション

7. 決定木 (Decision Tree)

8. 決定木とクラスタ分析のコンビネーション

9. まとめ

- 実務家にとって便利なセグメンテーション手法の使い分け

10. 事例

- アミノ形飲料市場におけるダイエット意識のセグメンテーション

株式会社インタースコープのご紹介

設立年月日:2000年3月9日

資本金: 3億1,445万円(資本準備金: 2億3,995万円)

- ・株主 平石郁生、山川義介他創業メンバー
エイパックス・グロービス・パートナーズ(AGP)、日本アジア投資(JAIC)、
JAFCO、SMBCキャピタル、新光インベストメント、東京中小企業投資育成、
新規事業投資

事業内容

- ・インターネットリサーチの企画実施および分析
- ・インターネットマーケティングおよび経営に関するコンサルティング
- ・One-to-One マーケティングシステムの開発および運用受託
- ・I.T.の有効活用による新しい調査手法および分析手法の研究開発
- ・I.T.関連の新規事業開発に関するコンサルティング

<http://www.interscope.co.jp/>

03-5728-6100(代表)

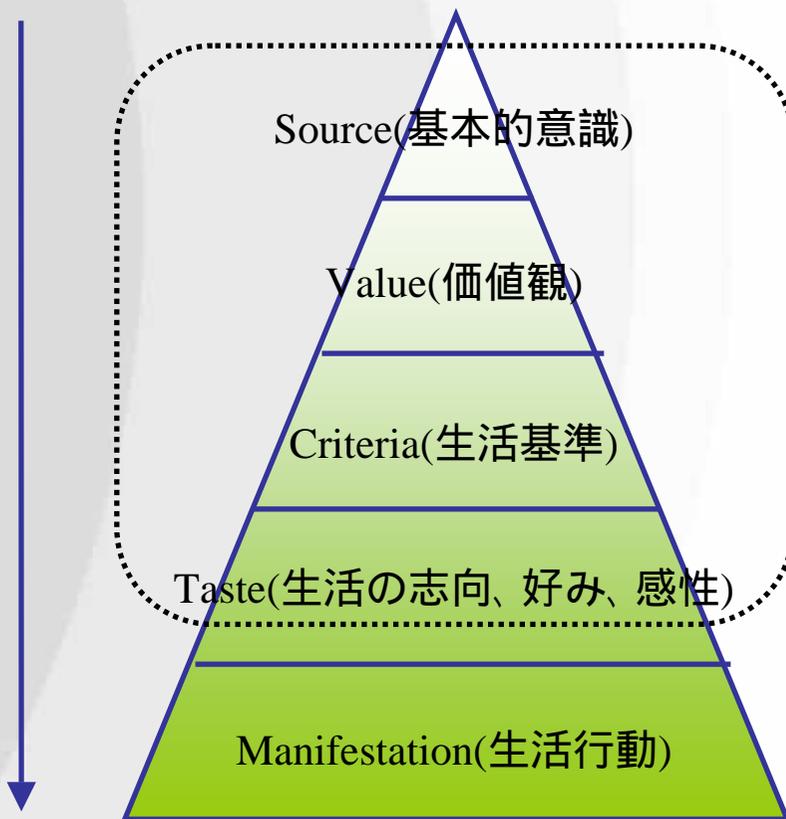
事業領域

1. インターネットリサーチの
企画・設計・実査・集計・解析・分析
2. 従来型リサーチの
企画・設計・実査・集計・解析・分析
3. テキストマイニング
4. Quick Scope の提供
(インターネットリサーチASP) の提供
5. モニターDBの構築および運用
6. マーケティング戦略立案コンサルティング



セグメンテーション論

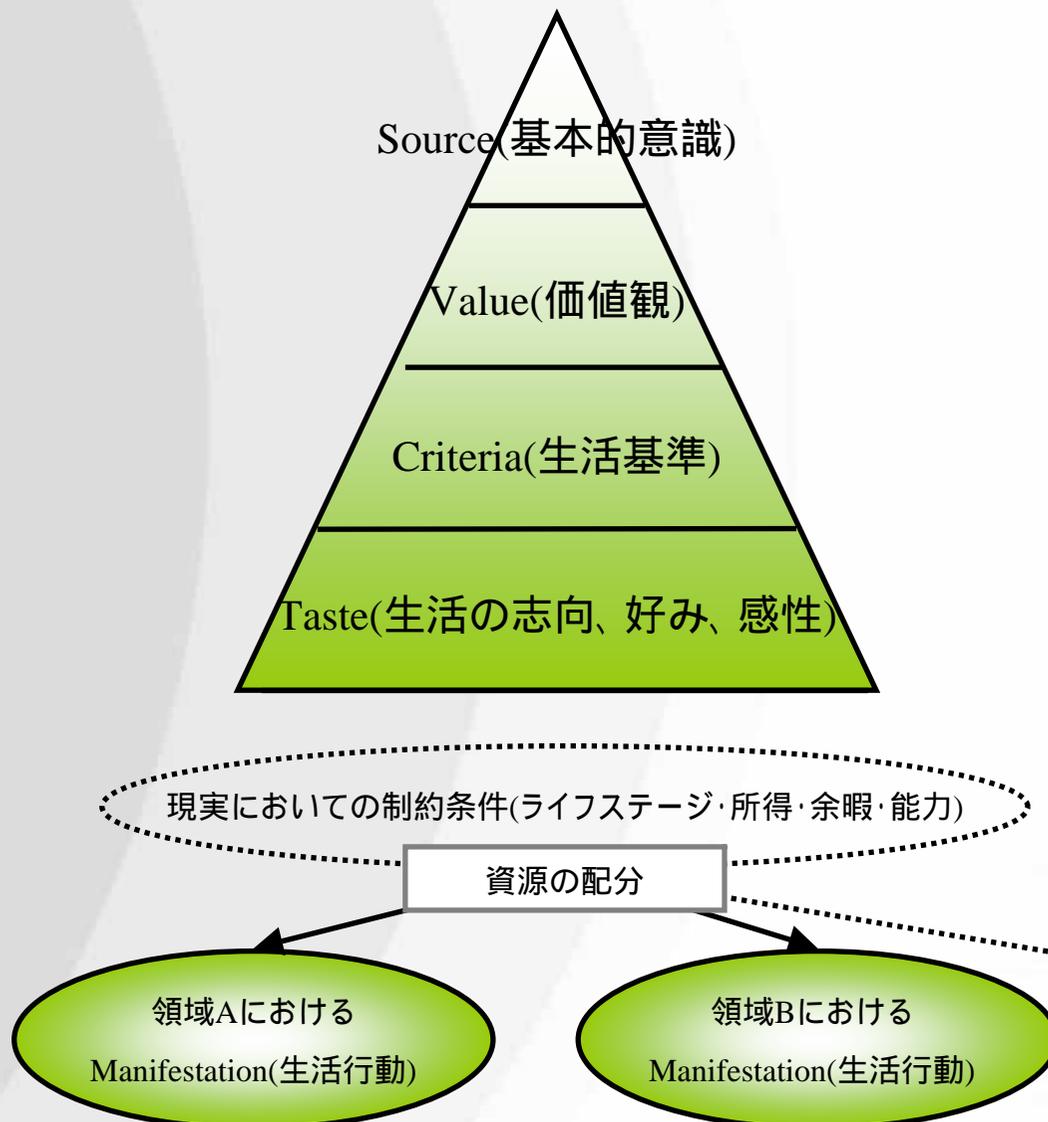
- 従来の価値観・ライフスタイルによるセグメンテーション
 - 従来の価値観・ライフスタイルによるセグメンテーションは、人間のManifestation(生活行動)を内部より規定していくSource(基本的意識)、Value(価値観)、Criteria(生活基準)、Taste(生活の志向、好み、感性)をいくつかのパターンに分類し、そこから生活行動パターンを予測・分類するという試みでありました。



ヤンケロビッチの価値観ヒエラルキー

「Source、Value、Criteria、Tasteといった上位概念が、Manifestationに強い影響を与える。」という仮説にもとづく。実際のManifestationが、流行などに左右されて変化し易いのに対して、その上位概念が変化するには時間がかかる。この性質を利用して、将来の消費者のニーズ・ウォンツを読み解くもしくは創造する際にその上位概念をいくつかのパターンにまとめ上げ、同じパターンを持った人々をグルーピング(セグメンテーション)することで、セグメント毎の戦略を考えていくことが非常に有用であると考えられていた。

価値観・ライフスタイルの現実



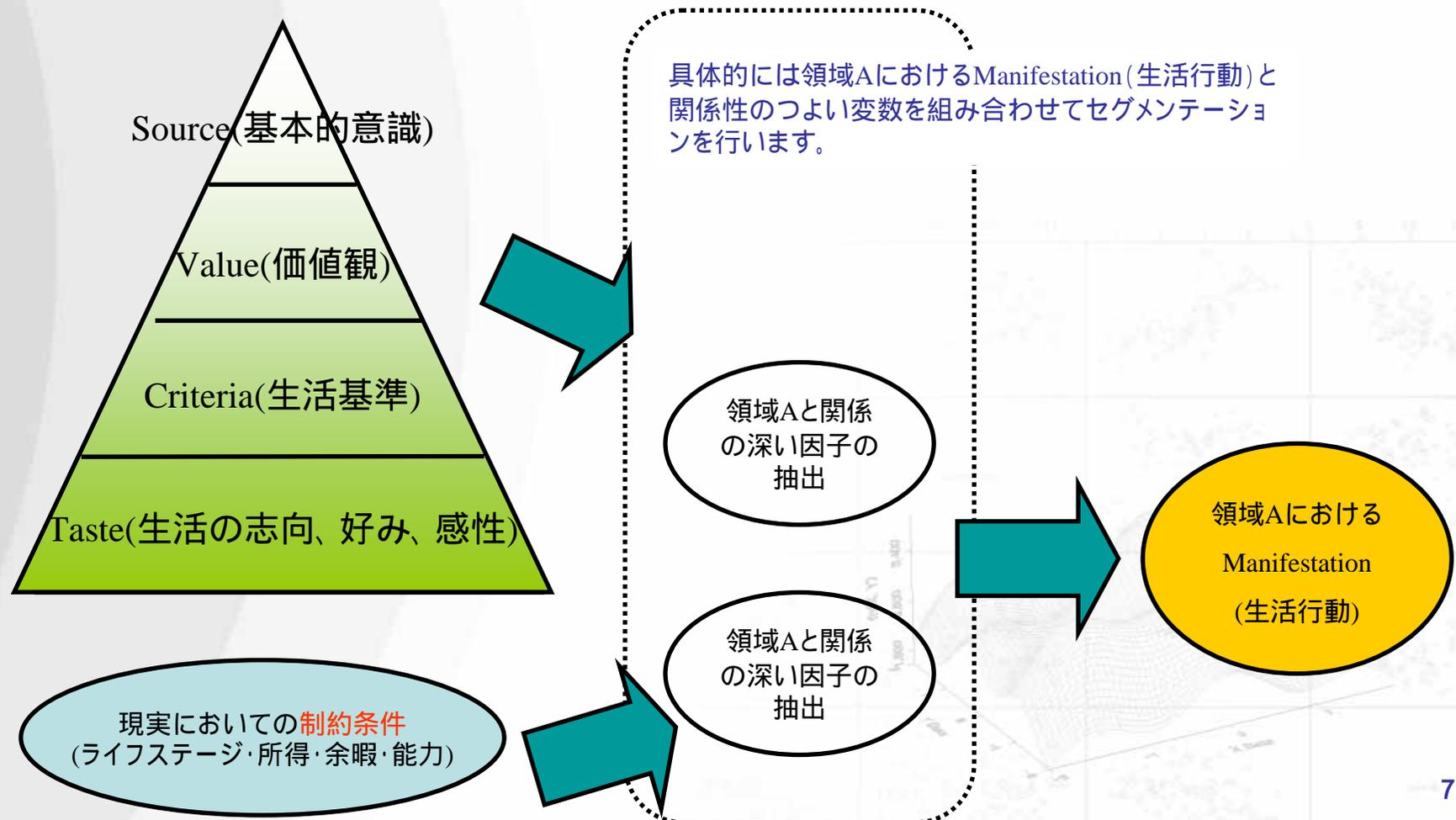
しかしながら、現実には Manifestation(生活行動)を叶える上で、所得やライフステージや持っている余暇や能力・スキルといった様々な制約条件があります。

一方で実現できる消費行動の数も現代では非常に多様化しており、ある分野にはお金をかける、ある分野はそこそこ、ある分野は見ないふりをするというような生活が迫られています。

セグメンテーションの考え方の中で、この制約条件は無視することはできず、また、Manifestation(生活行動)の領域を横断して、ひとくりに日本人を分類することはかなり難しくなります。

価値観・ライフスタイルの現実に合わせてソリューション

- 価値観・ライフスタイルの**現実**に合わせてソリューション
 - そこで、ある市場におけるターゲット別の戦略を構築する際には制約条件と価値観・ライフスタイルの両方を組み合わせたセグメンテーションが必要となり、また価値観・ライフスタイルについても、分析の対象となる領域と関係の深い因子を抽出して考えていくべきでしょう。



クラスター分析とは？

クラスター分析

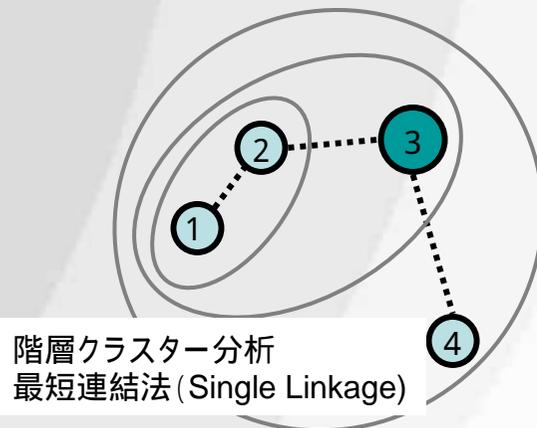
- 一群の対象のどれとどれが類似しているかを見つけ出すために用いられるさまざまな数学的方法の総称 (Cluster Analysis For Researchers, Romesburg, 1992より抜粋)
- 応用範囲が広く、多くの学問、ビジネスの分野で利用されるため、その方法もさまざま
- 最も一般的な区分方法は階層クラスター分析と非階層クラスター分析

階層クラスター分析

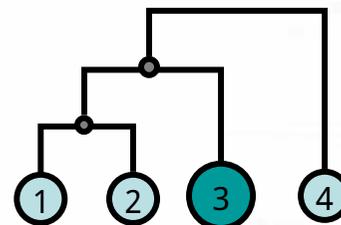
- 全対象間の距離を計算、最も近いものから順にすべての対象がひとつのクラスターになるまで結合を繰り返す
- デンドログラムを利用し、結合過程を観察できる
- 計算量が多い

非階層クラスター分析

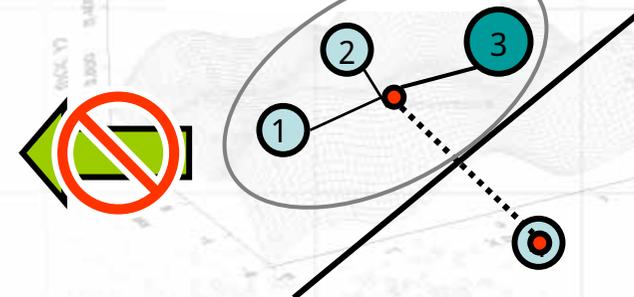
- 全対象を仮にk個のクラスターに分割、クラスターのシードを変化させ、最もよいk分割を探索する (分割最適化)
- 結合過程を観察できない(情報量が少ない)
- 計算量が少ない



デンドログラム



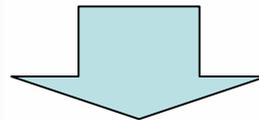
非階層クラスター分析
K-means法



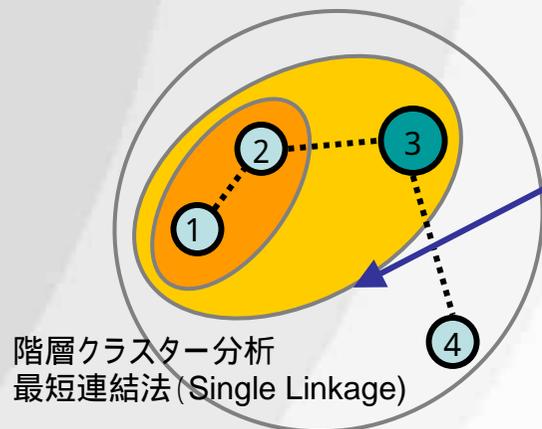
階層クラスター分析と非階層クラスター分析の使い分け

目的にあわせて使い分ける
(アンケートデータ・POSデータなどでの一般論)

- 階層型クラスター分析
 - 変数(ブランドなどの質問項目)の側をクラスター化したいとき
- 非階層型クラスター分析
 - 回答者(サンプル)の側をクラスター化したいとき

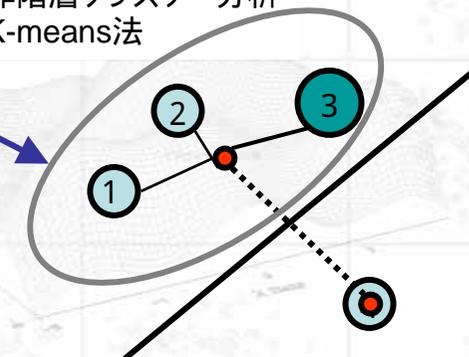


- サンプルの結合過程を観察しても、ほとんどなにも得られない
- ブランドなどの結合過程の観察からは、示唆が得られることが多い

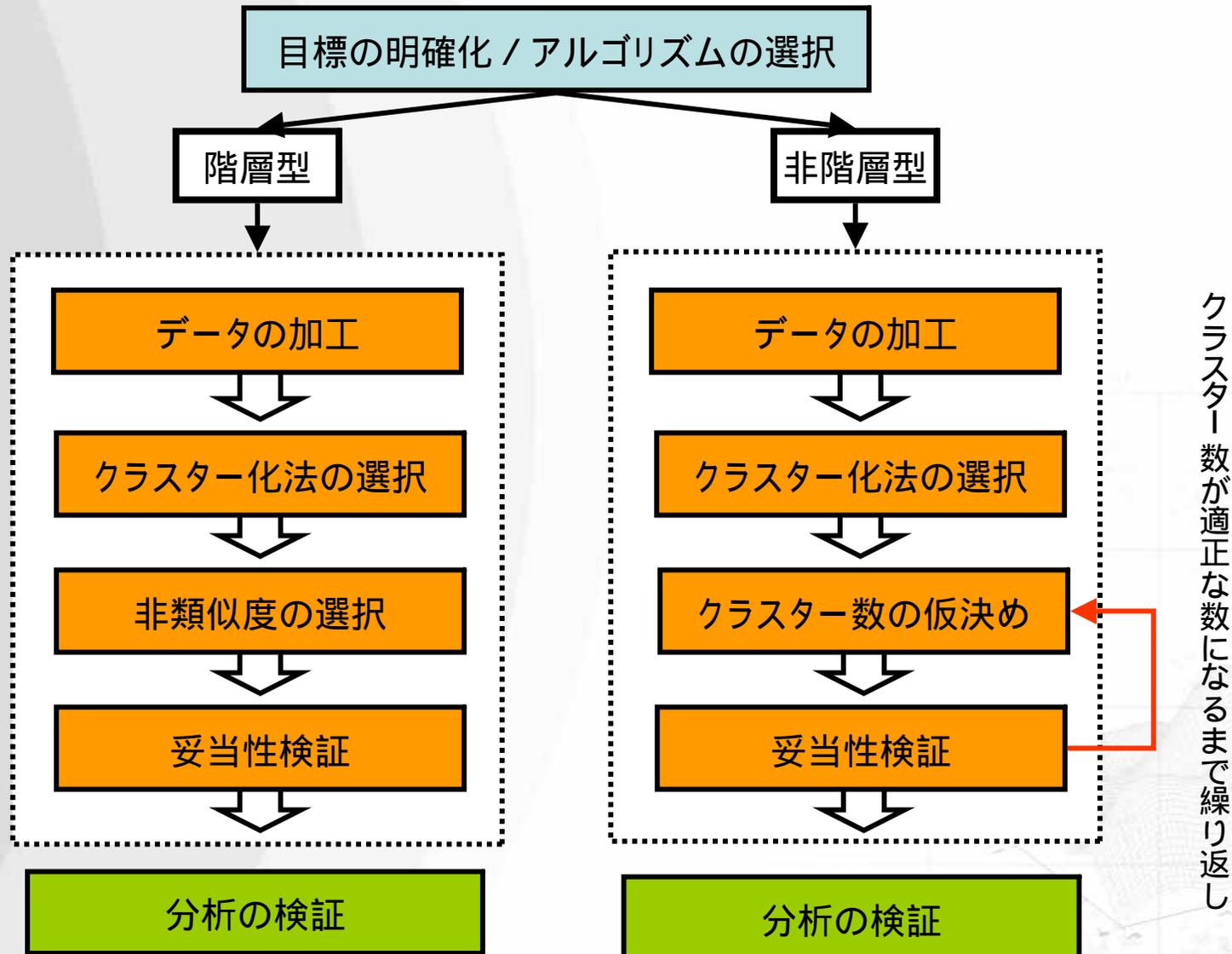


この情報で目的は達成できるか

非階層クラスター分析
K-means法



クラスター分析のステップ



階層クラスター分析 元データの処理

階層型 / 非階層型共通

どのようなデータをクラスター分析にかけるのか？

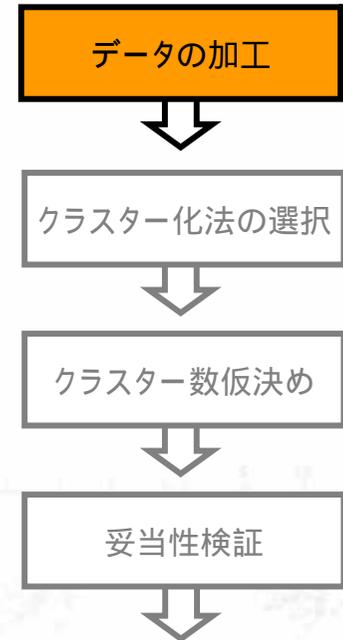
(アンケート分析、意識設問でのクラスター分析の場合)

- 5～7段階SD法
- 因子分析、主成分分析等で変数を分解 / 要約した変数
- その他: 二値変数

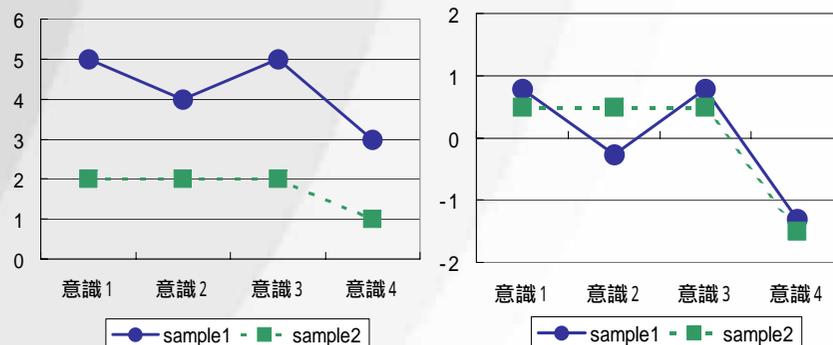
どのような処理をするのか？

一般的な変数の処理としての標準化をすることが必ずしも正しいとは限らない

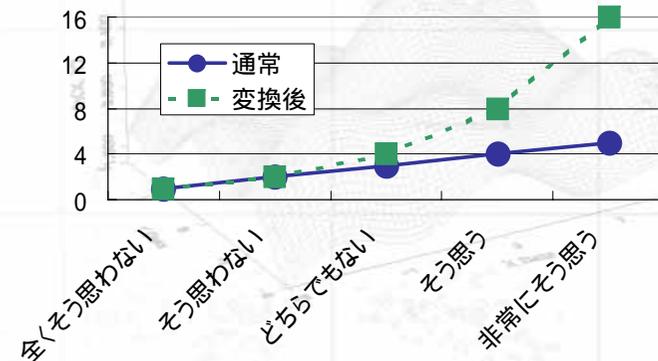
- 目的に合致した処理
- 分析に変数の選択 不要な変数の削除
- 変数の変換、標準化 結果のよりよい解釈のためにおこなう



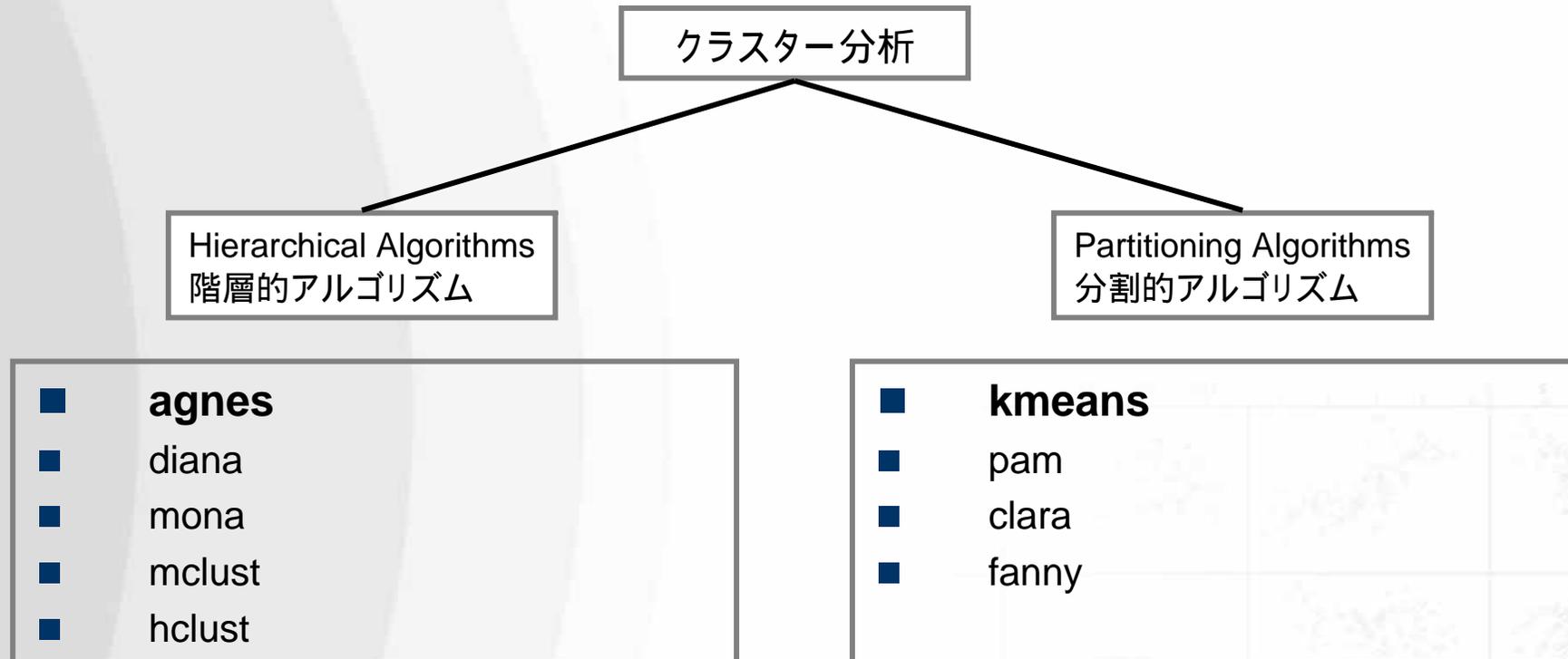
例) 意識の標準化



例) 5SDの変換



S-Plusにおけるクラスター分析の分類



S-PLUS Guide to Statistics for Windows
Vol.2 Chapter23 等参照

今回は各アルゴリズムの中で最も一般的な
S関数agnes,.kmeans、を用いた実例を紹介する

階層クラスター分析 クラスター化の方法の選択

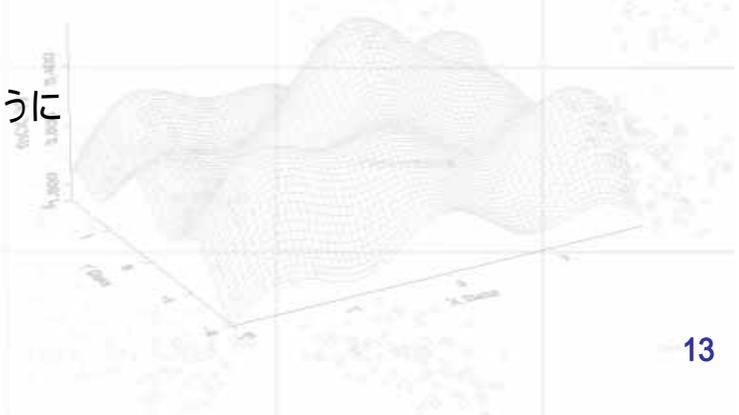
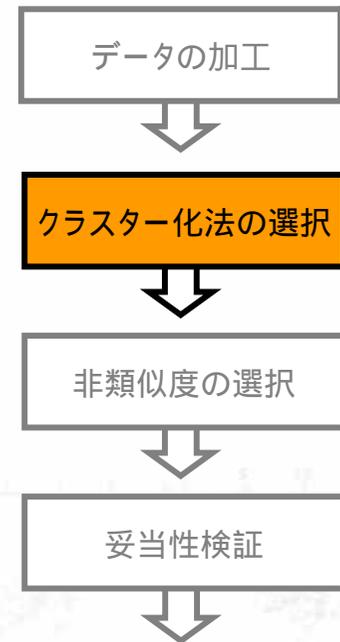
階層型

Agnesで利用できるクラスター化法は

- “average” (group average method) = 群平均法
- “single” (single linkage) = 単連結法 = 最短距離法
- “complete” (complete linkage) = 完全連結法 = 最長距離法
- “ward” (Ward’s method) = ウォード法
- “weighted” (weighted average linkage)
= 算術平均を用いた重み付き対グループ法

クラスター化法の説明と特徴

- **最短距離法**(Single Linkage clustering method)
クラスター間の非類似度 = クラスターの要素間で最小をとる非類似度
もっとも鎖効果、階層構造が現れやすいクラスター化法
- **群平均法**(UPGMA clustering method)
クラスター間の非類似度 = クラスターの要素間の非類似度の平均
鎖効果、拡散現象に対して強い
- **Ward法**(Ward’s clustering method)
クラスター内分散を最小化クラスター間分散を最大にするように
鎖効果、拡散現象に対して強い



階層クラスタ分析 特徴的なクラスタ化の方法

階層型

agnes(dissimilarity matrix,diss=T, method="ward")

最も利用されるクラスタ化法はWard法、群平均法

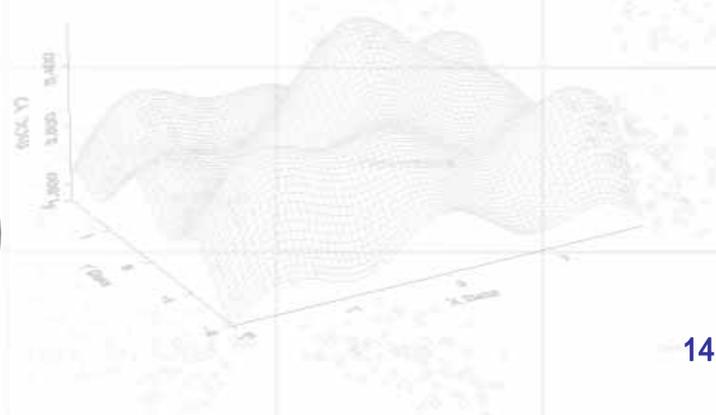
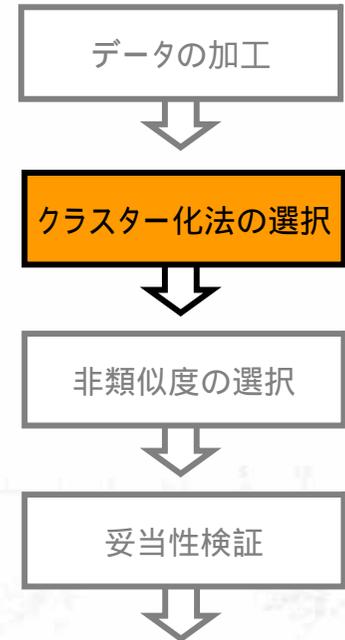
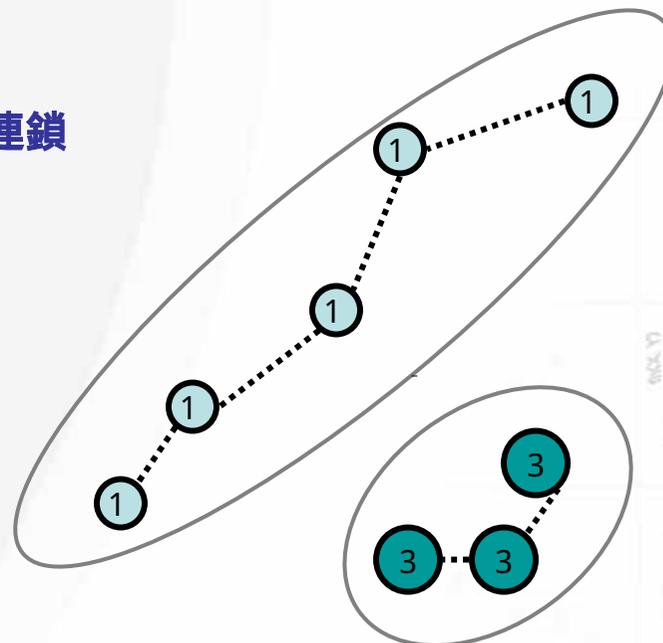
特徴的なのは最短距離法

- 群平均法、Ward法は連鎖に対して強い性質を持つ

群平均、ward法が最適なのか？

- 重要なのは実質科学的に意味があること
 - 連鎖という現象を検証する目的ならば最短距離法の利用価値も存在する

連鎖

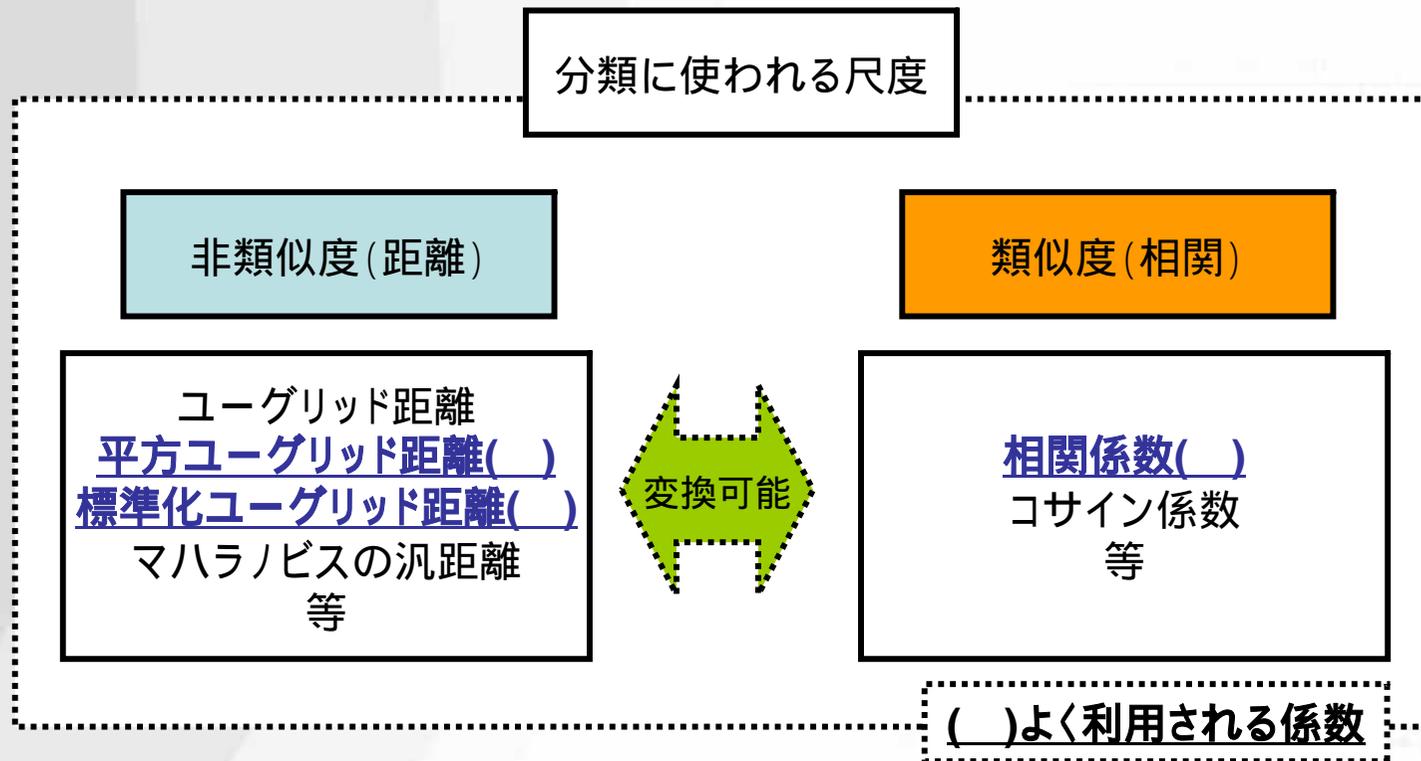


階層クラスタ分析 非類似度 / 類似度の種類

階層型

agnes(dissimilarity matrix,diss=T, method="ward")

- S-plusで距離を計算する関数はdist,daisy
- agnesではデータ行列以外に(非)類似度行列を利用することができる
- 距離は比尺度でも順序尺度でもよい



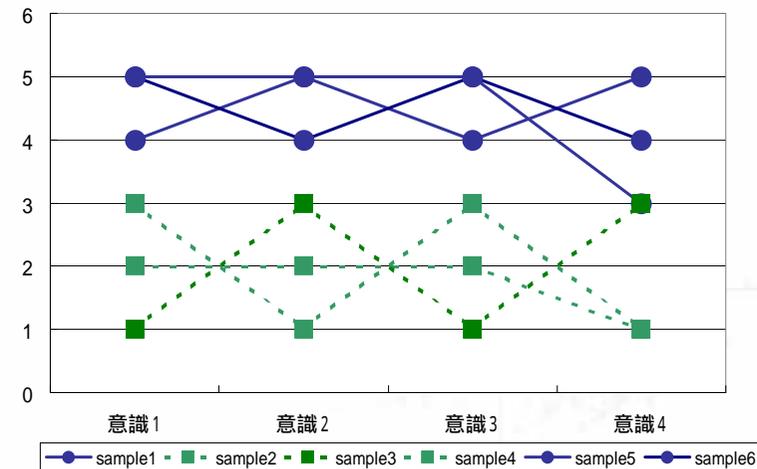
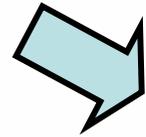
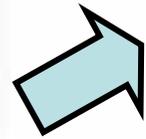
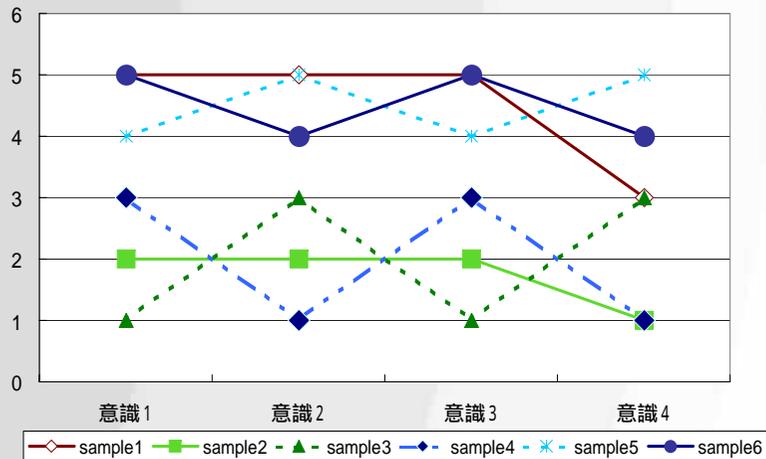
階層クラスター分析 非類似度の選択



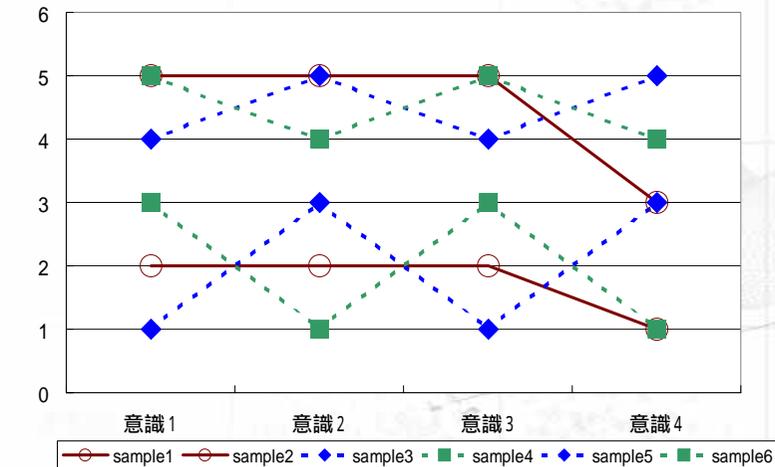
分析に使う非類似度は対象間のどのような関係を示しているのか？

「何を持って分類したいか」によって非類似度を使い分ける

例) 意識設問でクラスター分析を行う場合



意識の高低による分類



回答傾向による分類

非階層クラスター分析 分析の流れ

非階層型

```
kmeans(data, centers, iter.max=10)
```

クラスター化法の選択

- kmeansにおいて選択(操作)できるオプションは収束回数 (iter.max) のみ
- その他の属性の操作(距離等)は別の非階層クラスター分析(pam,fanny等)はS-PLUSのhelp等を参照)
- kmeansでデフォルトに採用されているのはユーグリッド距離
- クラスターの最小化は平方ユーグリッド距離の和を基準に行う。

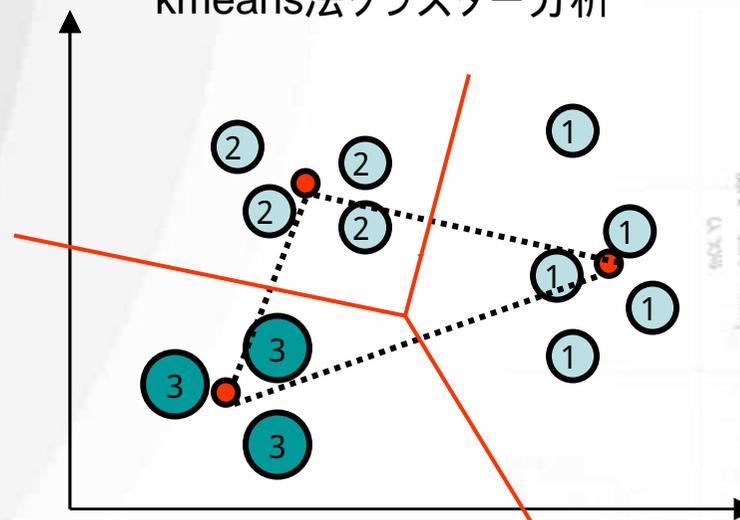
データの加工

クラスター化法の選択

クラスター数の仮決め

妥当性検証

kmeans法クラスター分析



非階層クラスター分析 分析の流れ

階層型

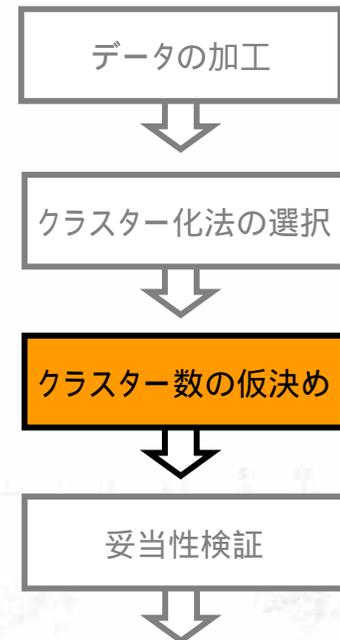
```
kmeans(data, centers, iter.max=10)
```

クラスター数の決定

- 通常実務では経験的に適正な数、もしくは仮説にのっとって適切なクラスター数を仮に決定し、後に検証する
- 説明変数数をクラスター数の上限とみなし、クラスター数を順に減少させてく探索的方法
- 必要なクラスター数を下減としてクラスター数を増やしていく探索的方法

補足

- クラスターが存在しない(対象で1つのクラスター)
 - クラスターが対象の数だけ存在する(すべての1クラスターサンプル)
- という可能性もあるので注意する必要がある



では何を持って最適なクラスターを判断するのか？

階層 / 非階層クラスター分析 クラスター数の選択

階層型 / 非階層型共通

妥当性検証

クラスター数の選択

- クラスターの数を機械的に選択する方法は存在しない。
(Cluster Analysis For Application, Michael R. Anderberg ,1988)

数値的方法

- 何らかの数値的な指標を元に判断する方法。たくさんの指標が提案されている。
- S-PLUSのkmeansのhelpでは以下のように、クラスターを増やした際のバラつきをもとに、クラスター数決定の参考にする方法を紹介している
When deciding on the number of clusters, Hartigan (1975, pp 90-91) suggests the following rough rule of thumb. If k is the result of kmeans with k groups and $kplus1$ is the result with $k+1$ groups, then it is justifiable to add the extra group when...

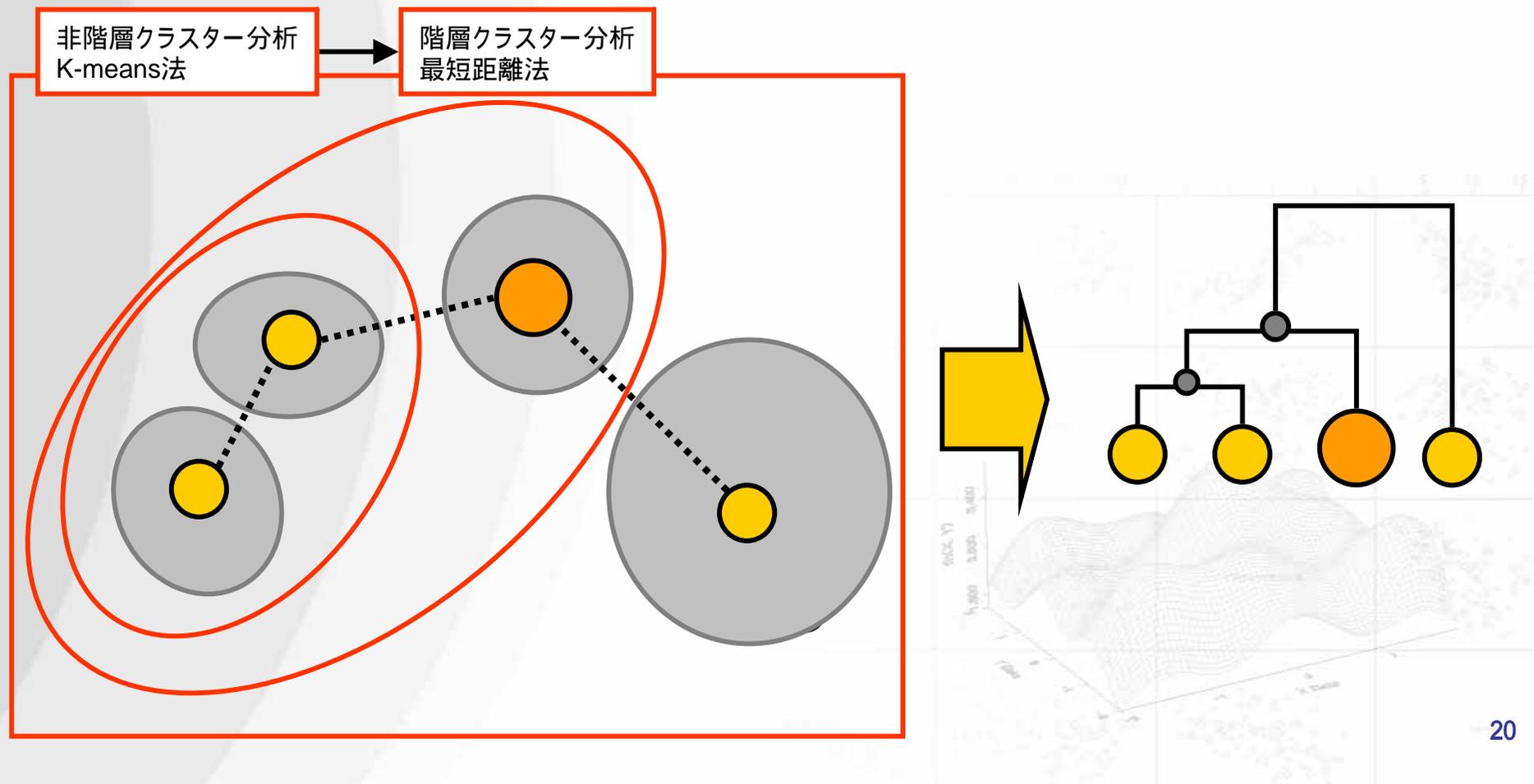
探索的方法

- 数値的方法はクラスターのまとまりのよさを表現する際にに有力なツールとなりうる。しかし、マーケティングにおいては、クラスター単体の切れ味がよいことよりも、「目的」に対して、いかに切れ味のよい指標となりうるかが重要である。下記の探索的方法によって最適なクラスターを選択する方法を提案する。
1. クロス表を用いて目的としている変数がよく説明できているかによって判断する
 2. 決定木等目的変数を考慮した分析手法で判断する(事例にて紹介)

階層、非階層クラスター分析のコンビネーション

KEY POINT

- 非階層クラスター分析を行った後、クラスターをサンプルとして階層クラスター分析にかけることで上位の結合過程を観察することができる。



事例～ダイエット飲料購入実態調査

調査目的	ダイエット飲料市場が活性化されている現在、ダイエット飲料に関わる消費者の意識と市場構造を明らかにし、ターゲットとするユーザのブランド選考理由を把握することを目的とする。
調査手法	インターネットweb定量調査(弊社自主調査)
対象エリア	予備調査:全国 本調査:首都圏
調査対象	「ダイエット飲料を飲んでいる20～49歳の男女」 <割付条件> 20～29歳: 男性:200s 女性:200s 30～39歳: 男性:200s 女性:200s 40～49歳: 男性:200s 女性:200s 計 1200S
調査内容	(1) 運動・食事・健康・ダイエットについての意識意識 (2) ダイエット飲料の飲用実態 (3) ダイエット飲料の評価
サンプル抽出方法	弊社「スコープNet」を利用
実査期間	2003年9月24日～2003年9月29日

分析ステップ

- 目的: ダイエット飲料A購入経験者の消費者像を以下の観点から浮き彫りにする
 - 属性(性別・年代など)
 - 健康・食事・ダイエット意識等

健康・食事・ダイエット意識質問による因子分析



抽出された因子による非階層クラスター分析



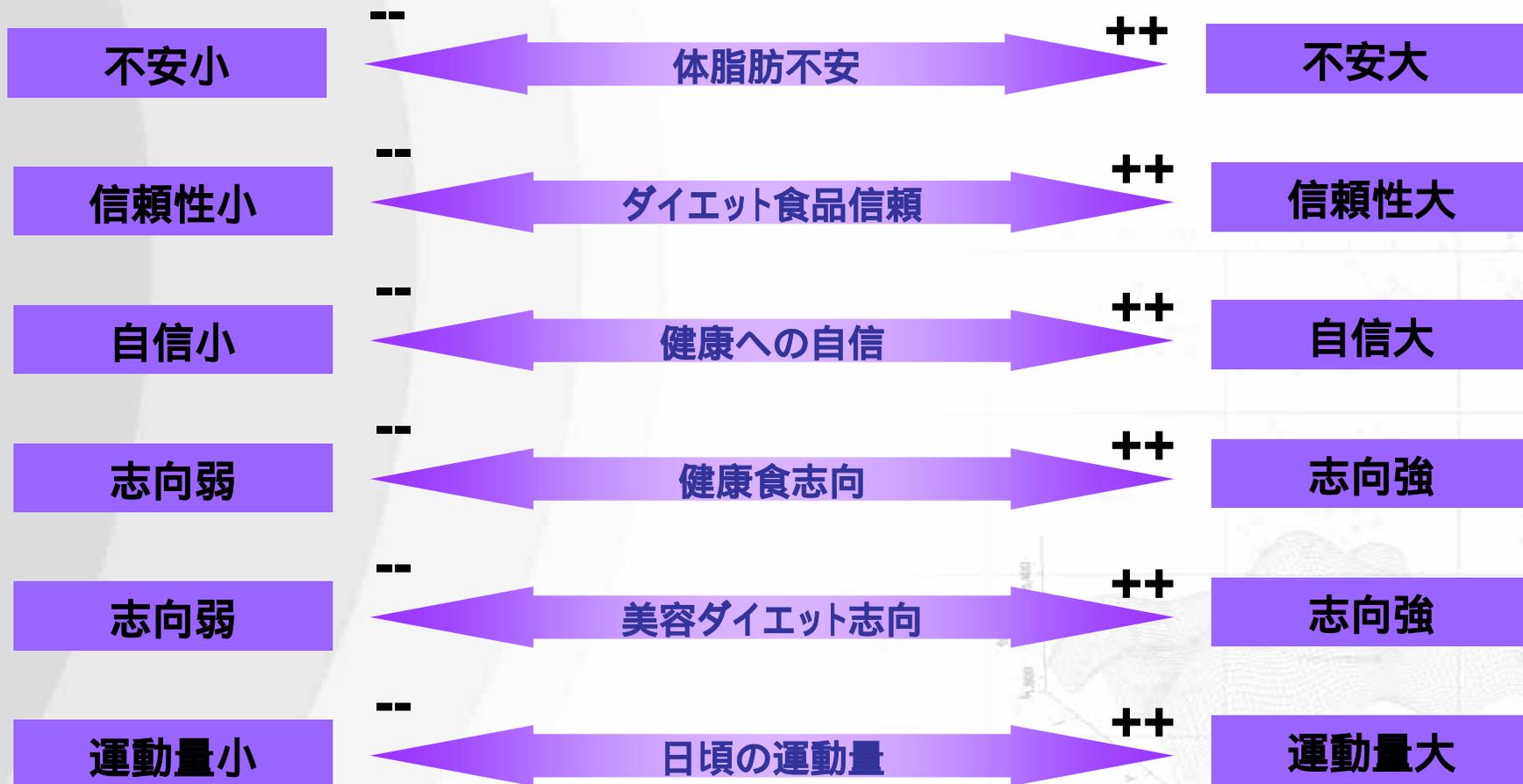
購入実態を目的変数にした決定木



消費者像のサマリー

因子の抽出

- 健康・食事・ダイエットなどに関連する質問が以下の6つの因子にまとめられました
 - 推定法: 最尤推定法
 - 回転方法: プロマックス回転



クラスター作成～複数パターンのクラスターを候補にする

- ブランドAの購入実態の違いをもっともよく説明するセグメントを作成するために、候補となる複数のクラスターを作成します。
 - データ形式: 因子得点をそのまま使う方法・因子得点をサンプル別に標準化したもの
 - クラスター数: 2クラスター～7クラスターまでを作成

データ形式(2パターン)

因子得点のまま

OR

サンプル内標準化得点

×

クラスター数(7パターン)

クラスター数2

↘

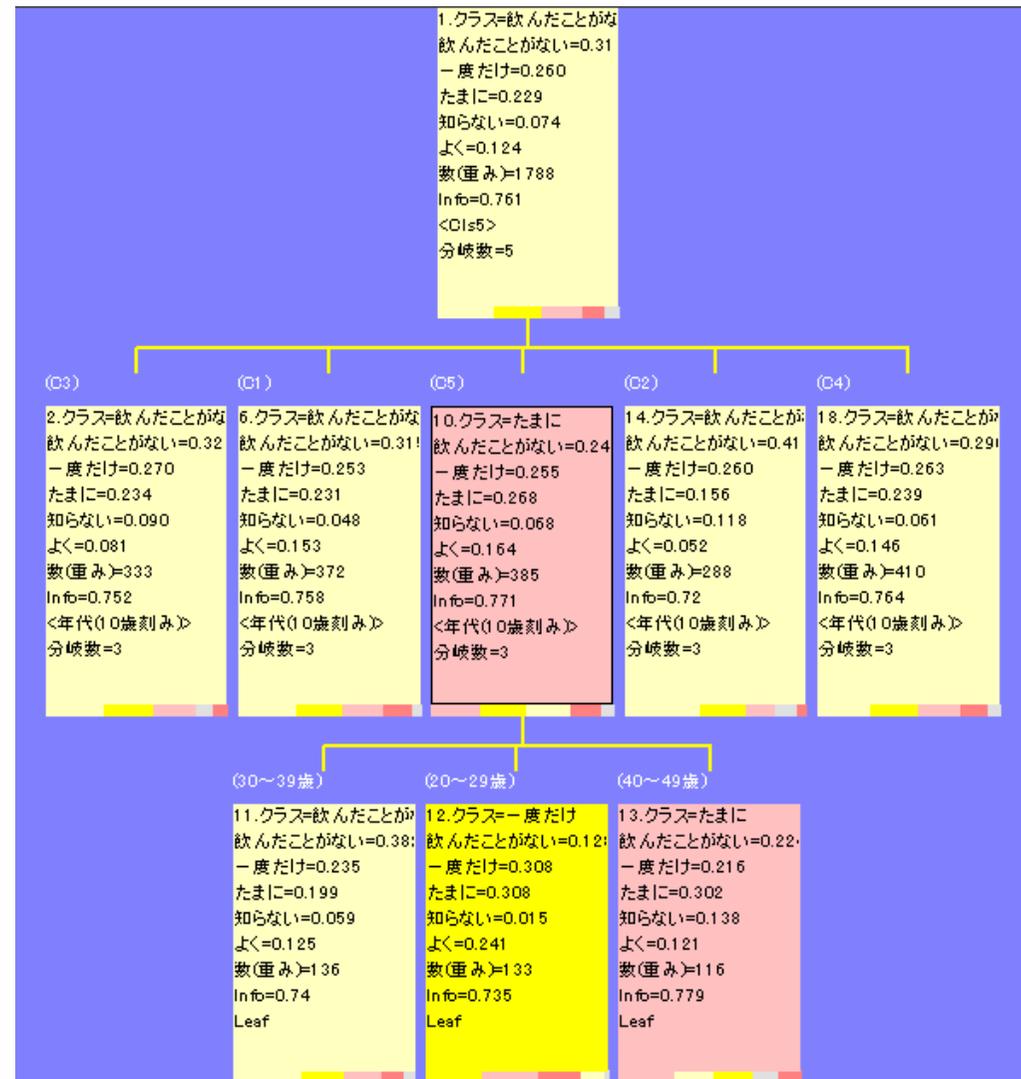
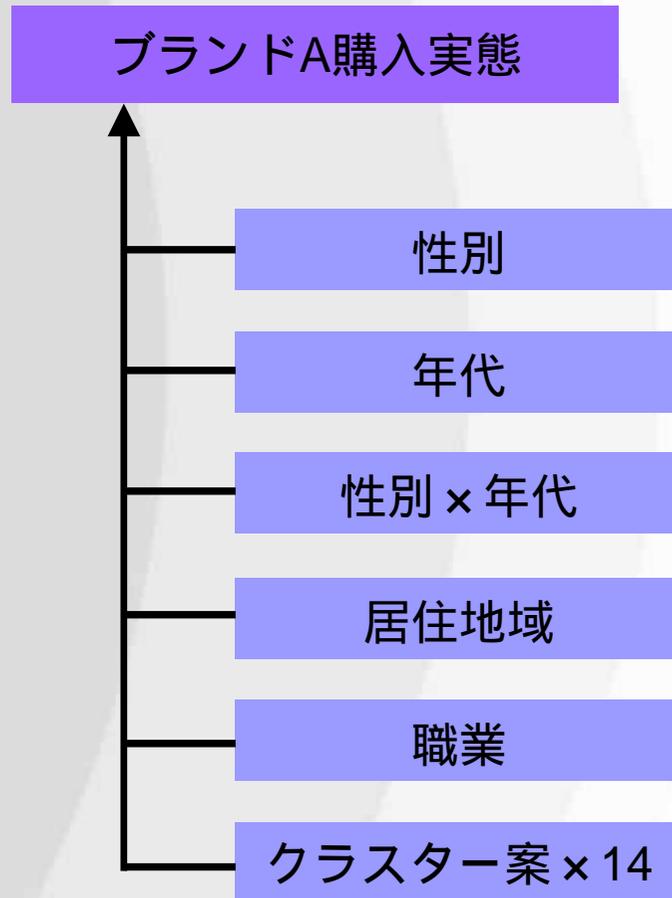
クラスター数7



全14通りのクラスターを準備

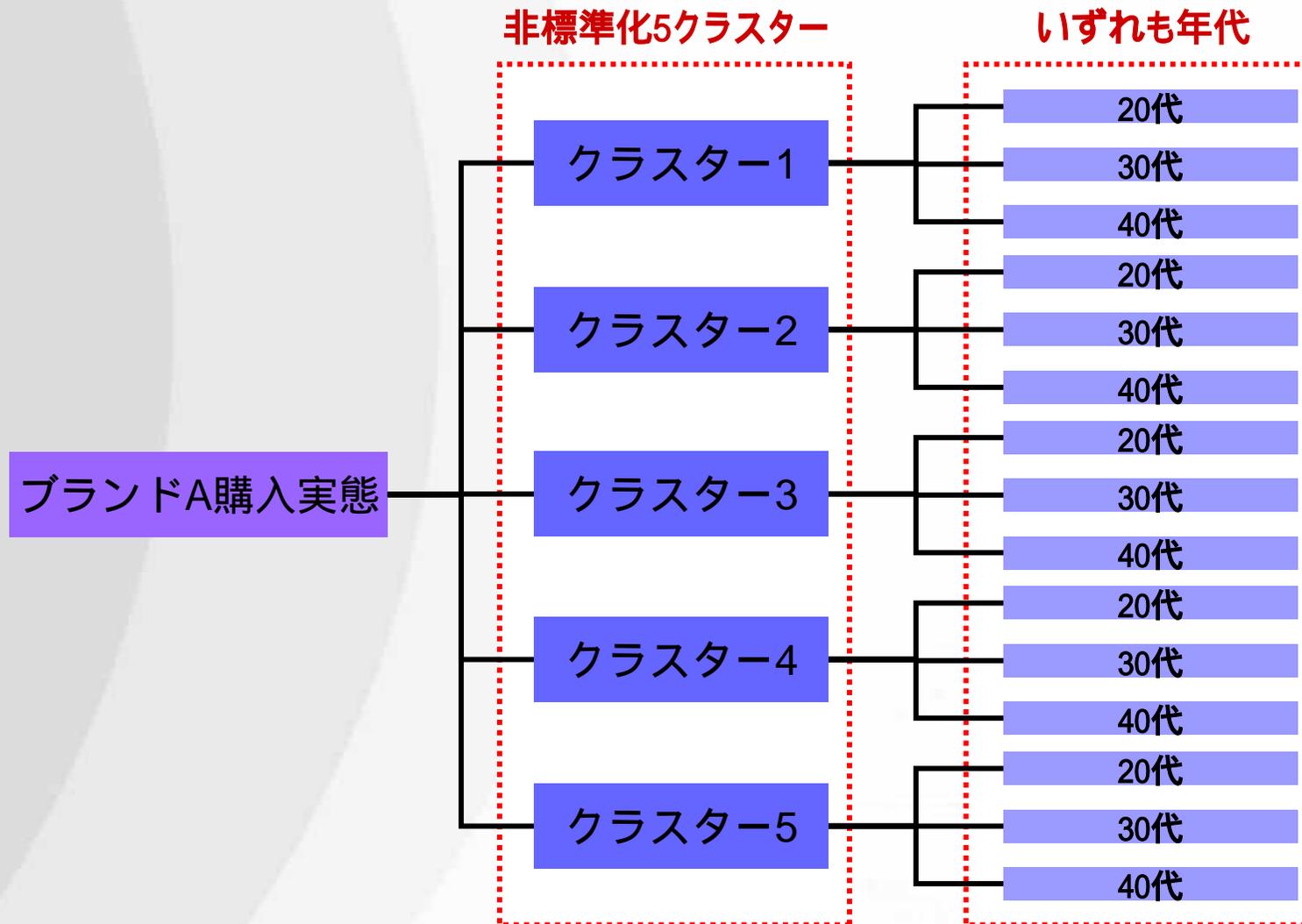
切れ味のよいセグメント軸の探索

- ブランドAの購入実態を目的変数としたDecision Treeを行い、切れ味のよいセグメントを探索します。



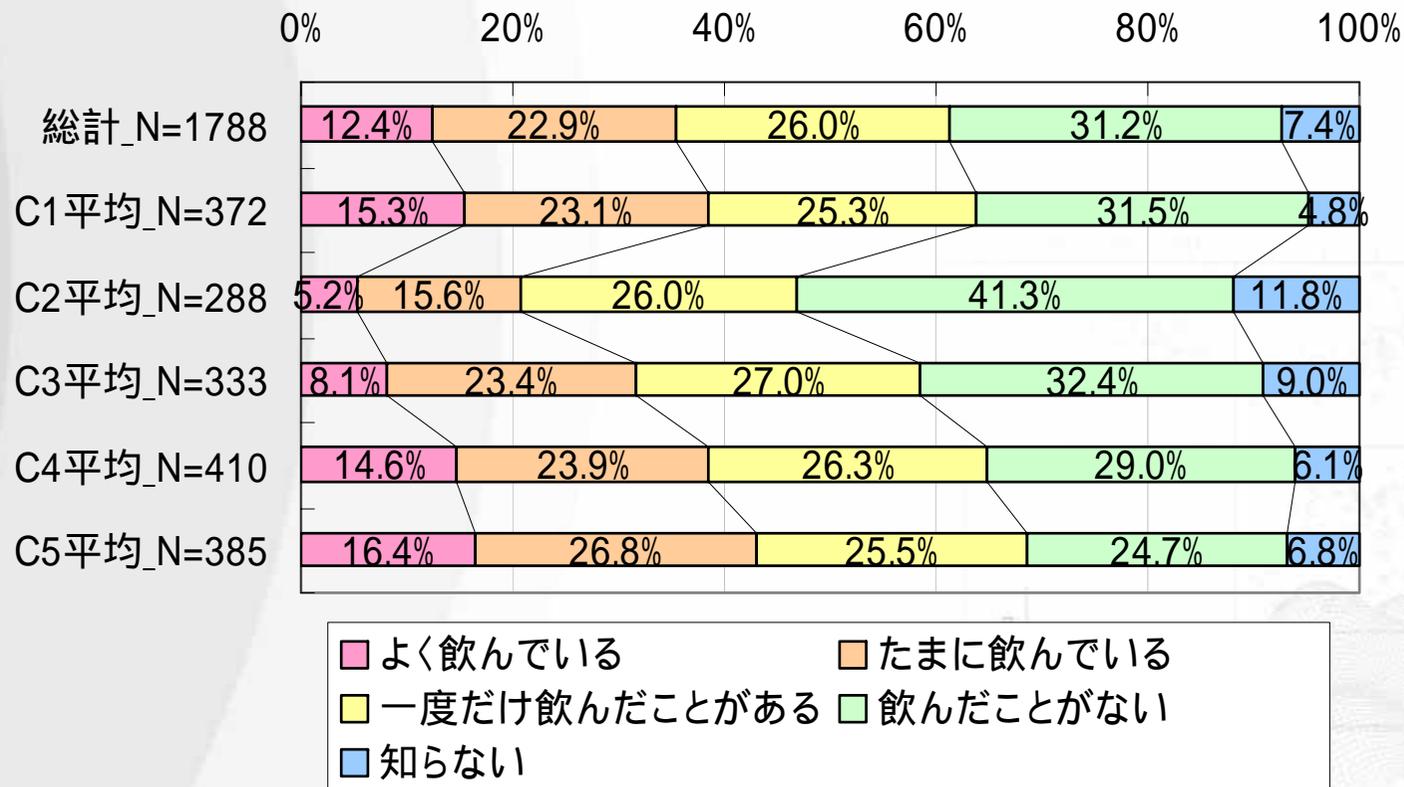
切れ味のよいセグメント軸の探索

- Decision Treeの分析結果として、1階層目が標準化していない場合のクラスター数5のパターンが選ばれた。次の階層でもいずれも年代によって分岐していた。



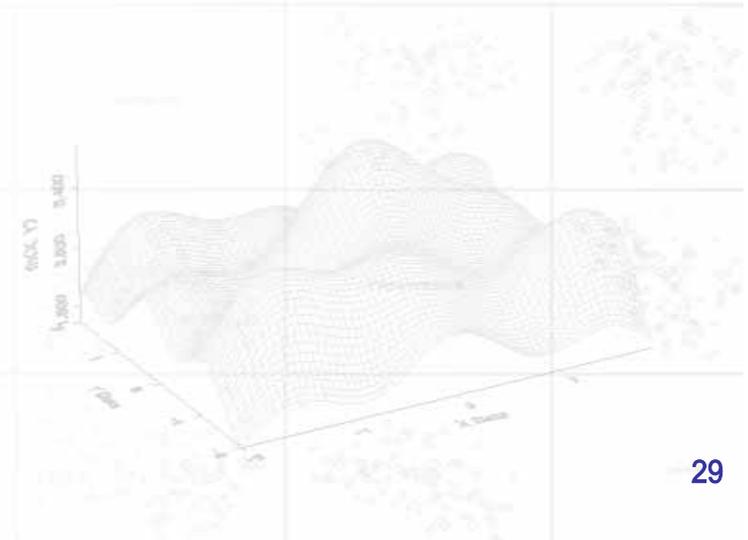
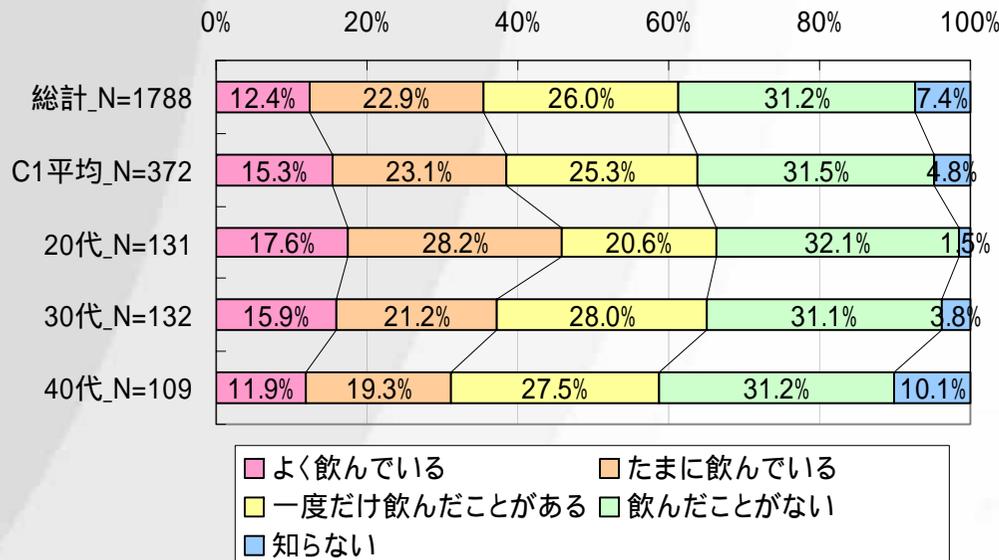
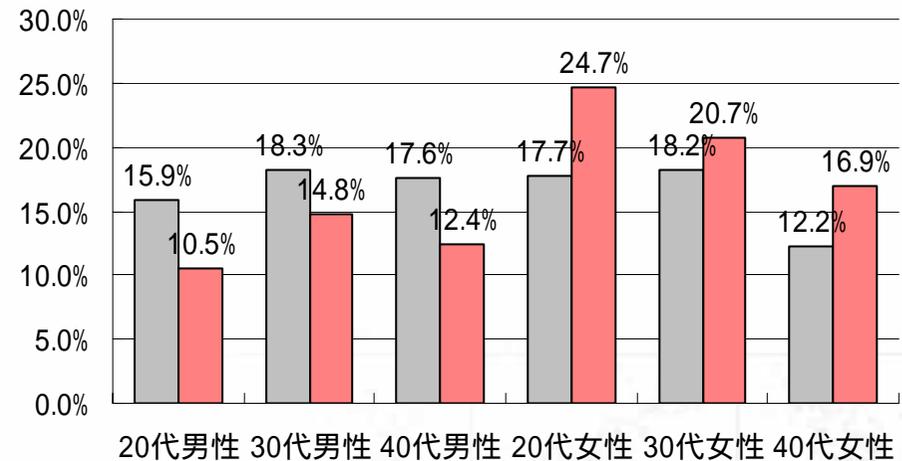
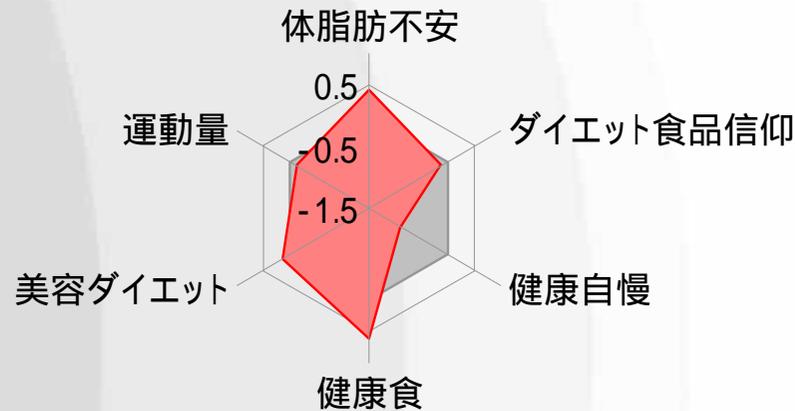
クラスターごとのブランドA購入実態

- Decision Treeの結果に基づき、非標準化因子スコアを用いたクラスター分析ごとに、ブランドAの購入意向の差を確認する。
- クラスター1・4・5はおおよそ平均並みだが、クラスター2はトライアル・リピート共に著しく低い。クラスター3はトライアル購入は平均並みだが、よく飲んでいるというスコアが低い。



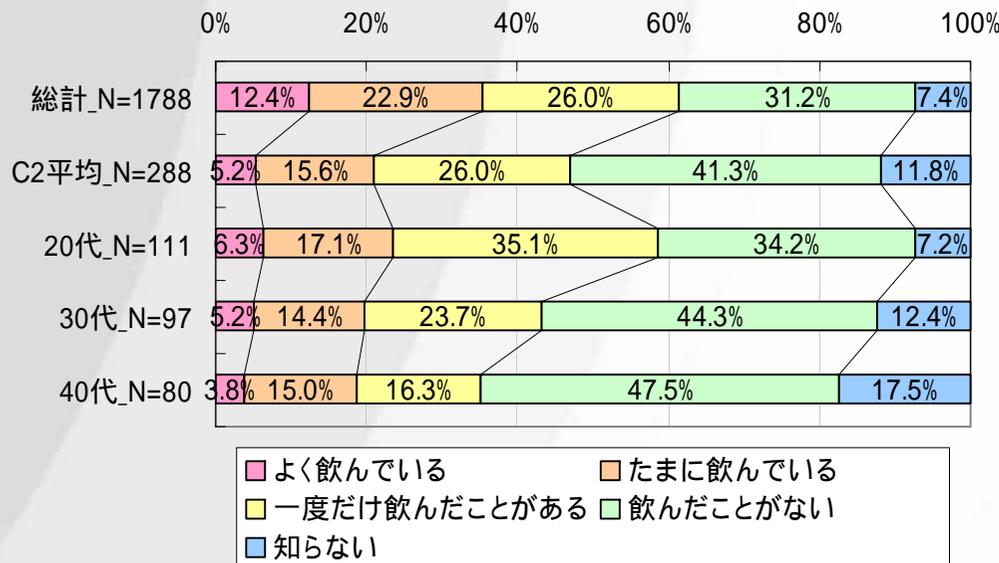
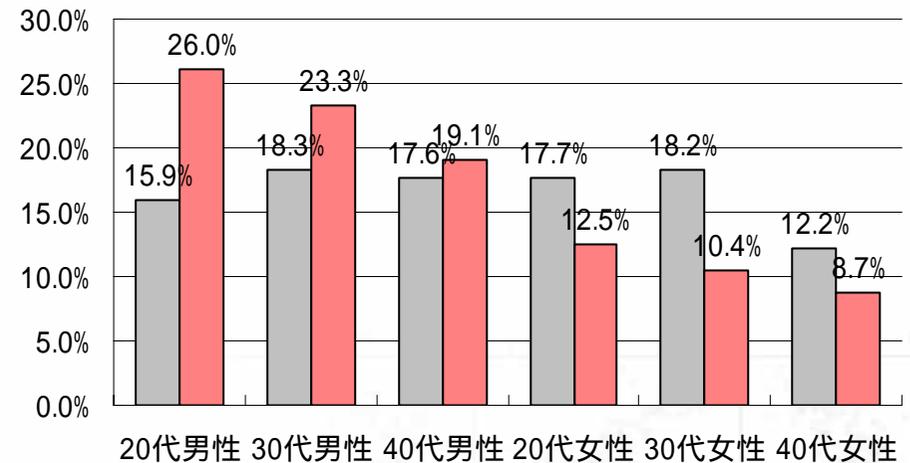
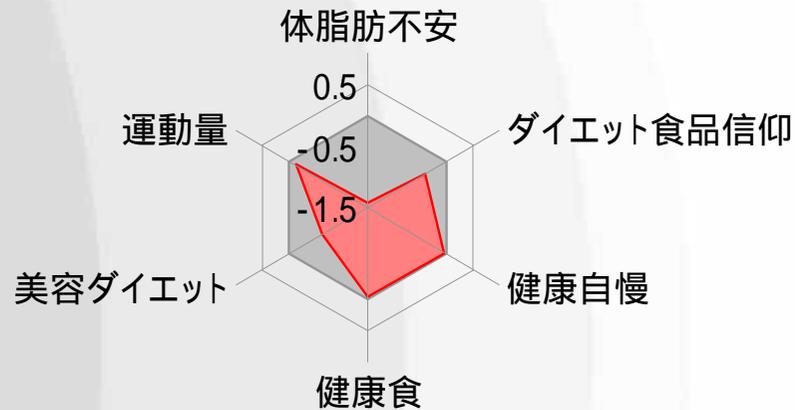
クラスター1

- 健康に対する不安が最も高いグループ。女性比率が全般的に高い。
- ブランドA購入意向は全般的に若干高い程度。歳が若いほど購入する傾向が高い。



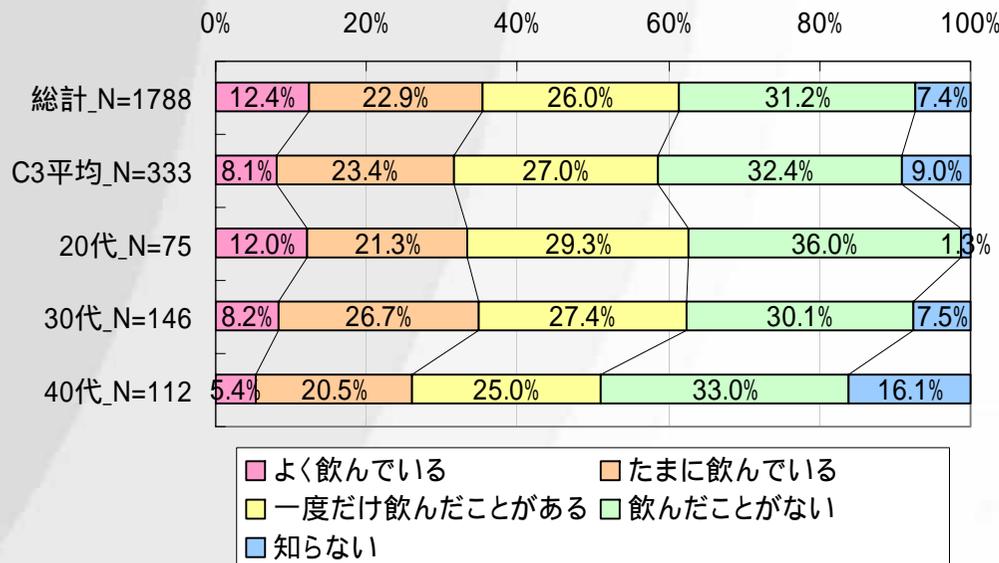
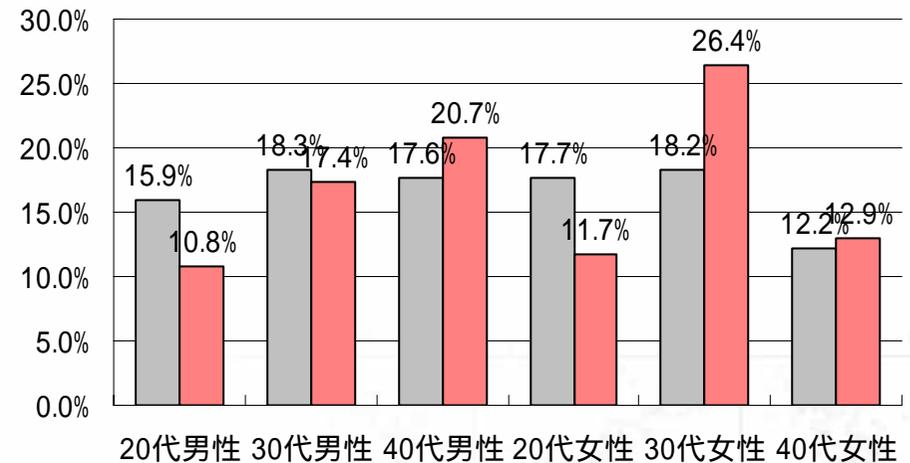
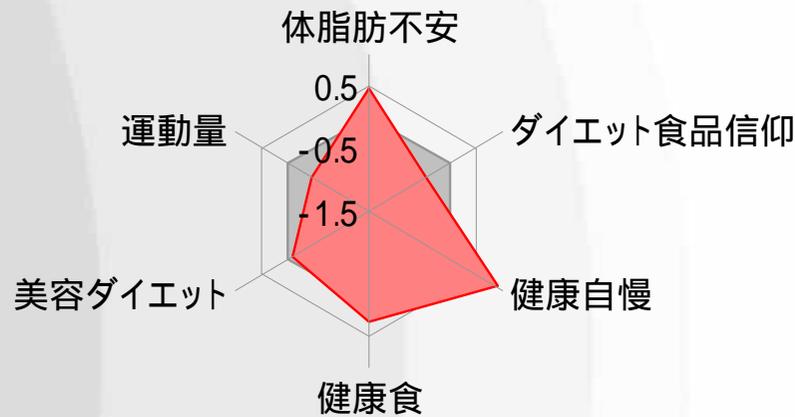
クラスター2

- 体脂肪に対する不安が少なく、ダイエットへの関心も低い。特に20～30代の男性比率が高い。
- ブランドAリピート率が非常に低い。それでも20代層では一度は飲んだことがある人が多い。



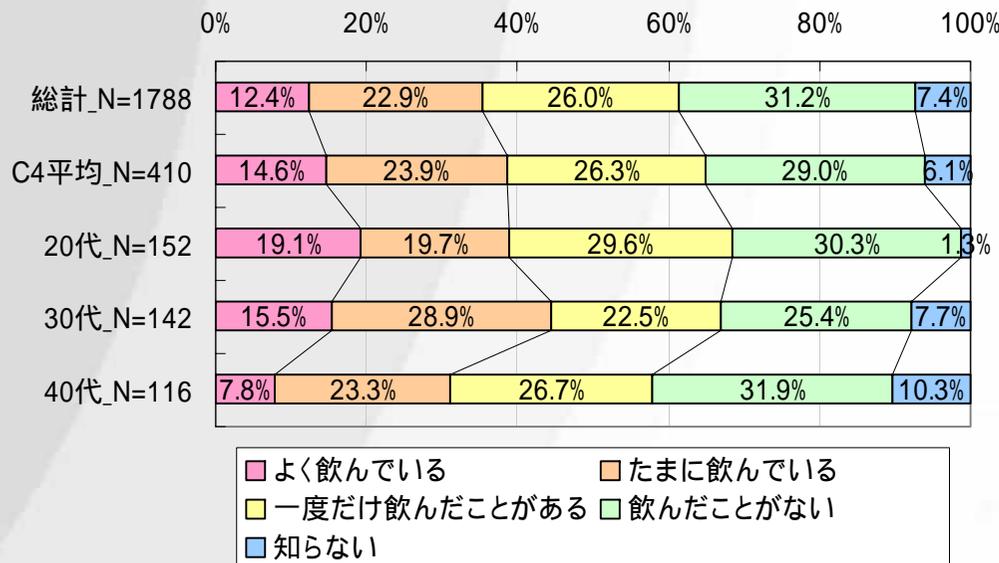
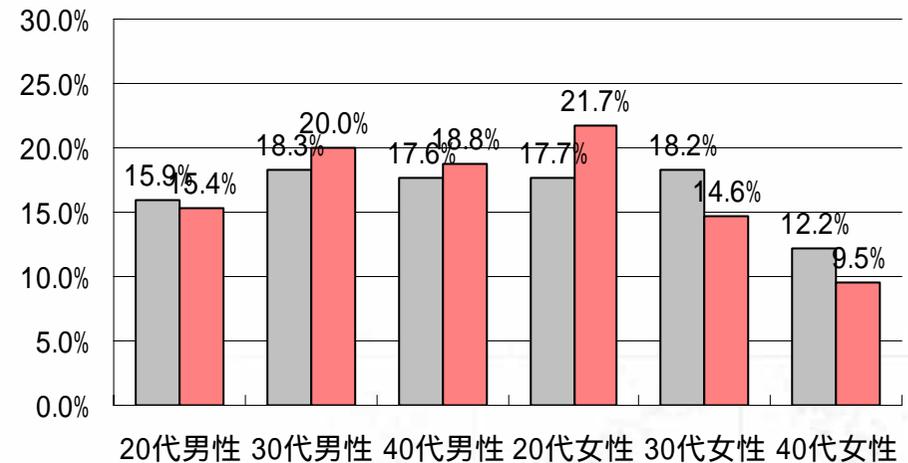
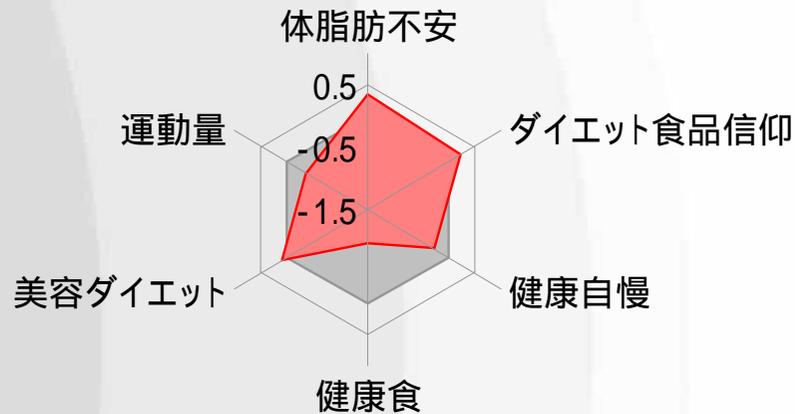
クラスター3

- 健康自信があるが、体脂肪に対する不安も高い。30代の女性比率が高い。
- ブランドA購入傾向傾向は若干低く。ブランドAのターゲットとはなっていない。



クラスター4

- 体脂肪への不安が大きく、ダイエット機能系食品の機能を信じている。食生活は乱れがち。
- 年代による差が比較的大きい。特に20代ではよく飲んでいる人の比率が20%近い。



■ よく飲んでいる ■ たまに飲んでいる
■ 一度だけ飲んだことがある ■ 飲んだことがない
■ 知らない



クラスター5

- 体脂肪への不安はないが、それ以外の意識が全般的に高い。
- ブランドA購入比率は全体で高く見えるが、20代のリピート率の高さが全体を引っ張っている。

