

実務での統計利用

株式会社マーケティング・エクセレンス マネージング・ディレクター
明治大学大学院 グローバル・ビジネス研究科 教授
戸谷圭子

1. よくある統計の落とし穴
2. 実務で使われている統計手法の紹介
3. 最低限必要と考えられる知識

よくある統計の落とし穴

よくある統計の落とし穴

問題の社会的背景

- インターネット調査の普及（低コスト・早い）
 - ✓ 結果数値のみが取り上げられ、誰も疑問を持たない
 - ✓ サンプルの代表性、回答の信頼性、集計方法、結論の出し方などに問題のあるケースが多数

 - ビッグデータの入手と分析可能性の向上
 - ✓ ツールを使えば、とりあえず答えは出てくる
 - ✓ システムデータの特徴を踏まえたデータクリーニングの必要性無視
(残高は0は、口座ありで0なのか、欠損なのか…etc)
- データ・数値の意味を正しく理解し、考察できる力が必要**

(例) 平均値に本当に差があるか？

自社満足度調査で、女性のほうが男性より満足度が高いと判明？

- 男性 …「4.00」(5点満点の平均)
- 女性 …「4.35」(5点満点の平均)

➡ どういうサンプリングか？

- ✓ 母集団を代表していない(昼間時間帯のみ電話調査・小サンプルetc.)

➡ どういう調査方法か？

- ✓ 偏りを招く調査(担当者が対面で来店中の顧客に調査→悪く言わないetc.)

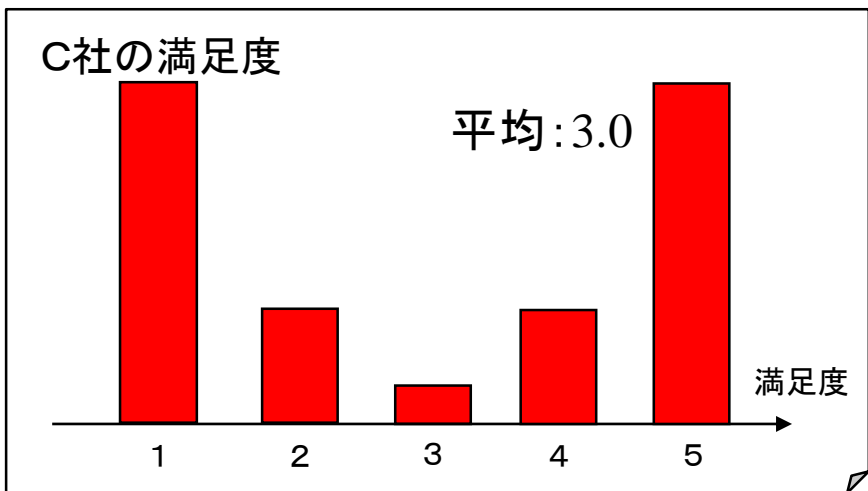
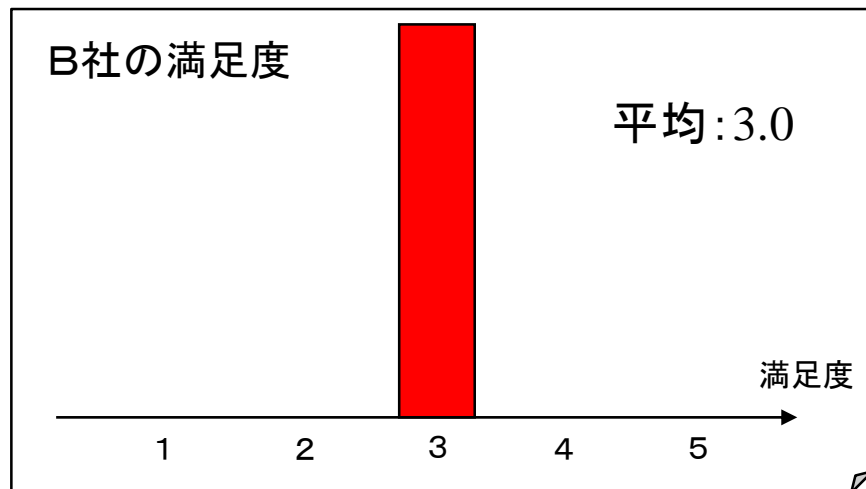
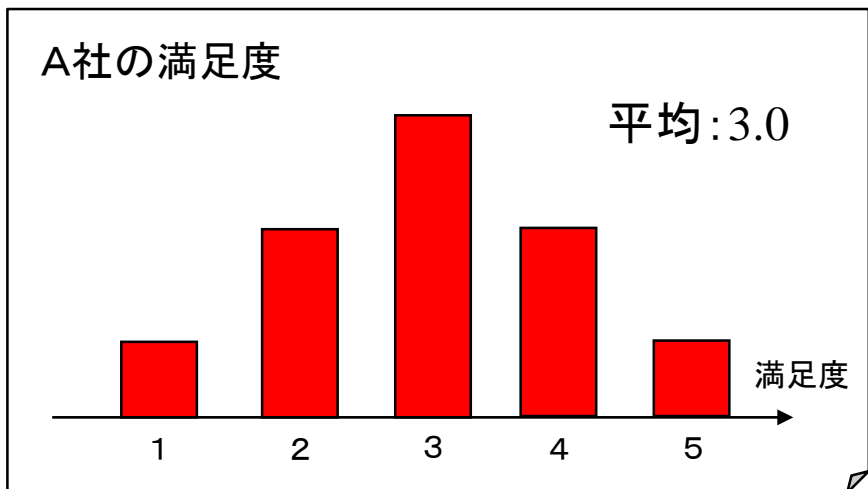
➡ どういう回答方法でどう集計したか？

- ✓ 偏りを招く尺度(「どちらでもない」を許さない偶数尺度回答→一般に満足の側に偏る)

➡ その差は偶然(測定誤差)ではないのか？

- ✓ **平均値の差の検定を行わずに平均の差を議論する(日常的)**

例えばこんなケースも



3社の顧客は同様に満足しているか？

平均値以外の

データの特徴を示す代表的な統計量の理解が必要

【最低限確認すべき基礎統計量】

- ・ データの中心をあらわす
 - 平均値
 - 中央値(メディアン) ……真中に位置する値
 - 最頻値(モード) ……出現頻度の最も高い値
- ・ 範囲をあらわす
 - 最大値
 - 最小値
- ・ ばらつきをあらわす
 - 分散
 - 標準偏差
 - 四分位点(25%, 50%, 75%) ……データを大きさ順に4分割する値

企業で一般的に使用されている統計分析手法

(一般的)

1. 基礎統計量(分布確認・平均値・外れ値など)
2. クロス集計
3. 相関分析
4. 因子分析
5. 重回帰分析・ロジスティック分析

(目的別)

6. 主成分分析
7. 因子分析
8. 判別分析
9. クラスタリング

企業のマーケティング分野で使用されている 統計分析手法

(一般的)

1. **連関分析**: 購買商品の組み合わせパターンを知る
2. **クラスタリング**: 顧客セグメントを分ける
3. **因子分析**: ブランドのポジショニングの軸を知る
4. **重回帰分析**: 因果関係を知る
5. **多次元尺度構成法 (知覚マップ)**: 顧客にとっての自社の位置を知る
6. **生存分析**: 顧客離脱の原因を知る
7. **共分散構造分析**: 顧客満足の構造を知る
8. **コンジョイント分析**: 製品・サービスを設計する

より専門的、また、ビッグデータの場合、手法はさらに多様化する
(決定木・ニューラルネット・ベイジアンネットなどなど)

マーケティング利用例

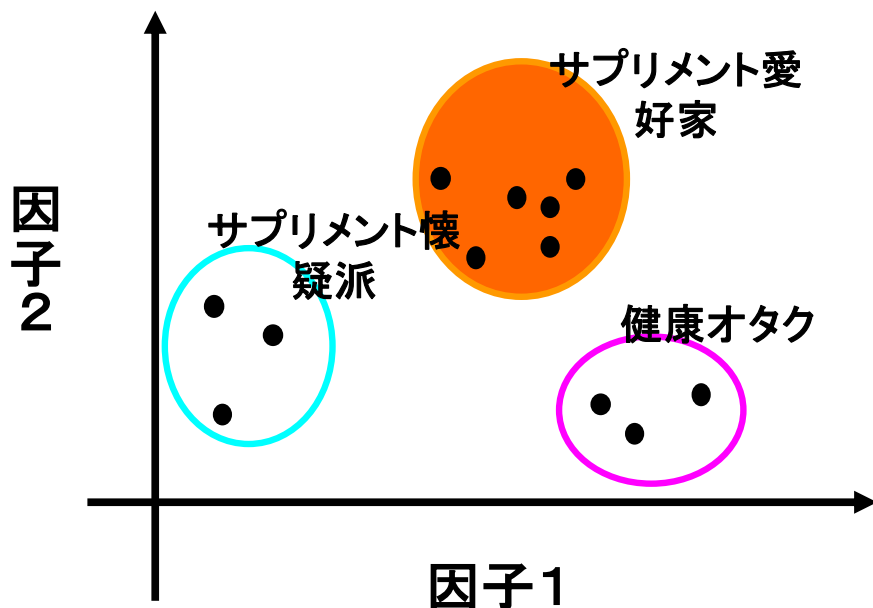
ダイレクトメールを送付する顧客をリストから選出する

因子分析

調査などで観測された回答傾向から、背後にある潜在的な共通要因を特定する手法。

クラスタリング

対象(ケース、変数)の中から互いに似たものを集め、個体間(サンプル間)の距離からいくつかのグループに分類する手法。顧客をセグメンテーションする場合によく用いる。



機械学習

決定木・ニューラルネットワーク・サポートベクターマシン

調査に回答していない顧客リストから、「サプリメント愛好家」セグメントに属する人を見つける

DMの内容を顧客に合わせて購入確率を上げる

連関分析

「商品A」を買った顧客は、「商品B」を買う確率が高い、というルールを抽出する分析方法。併買を促すための推奨に活用される。

| ルール | 条件 (…ならば) | 出力 (…である) | 条件付き確率 | 前提確率 (条件なし確率) |
|------|--------------|--------------|--------|------------------|
| ルール1 | サプリアを買った | サプリアを買う | 46.3% | 10.7% |
| ルール2 | サプリアEを買った | サプリアBを買う | 31.8% | 10.7% |



サプリアBのプロモーションを無作為に行えば1割程度の購買可能性があるが、サプリアAを購入したことのある顧客を対象にすれば4倍、サプリアEを購入したことのある顧客を対象にすれば3倍の購入確率になる。

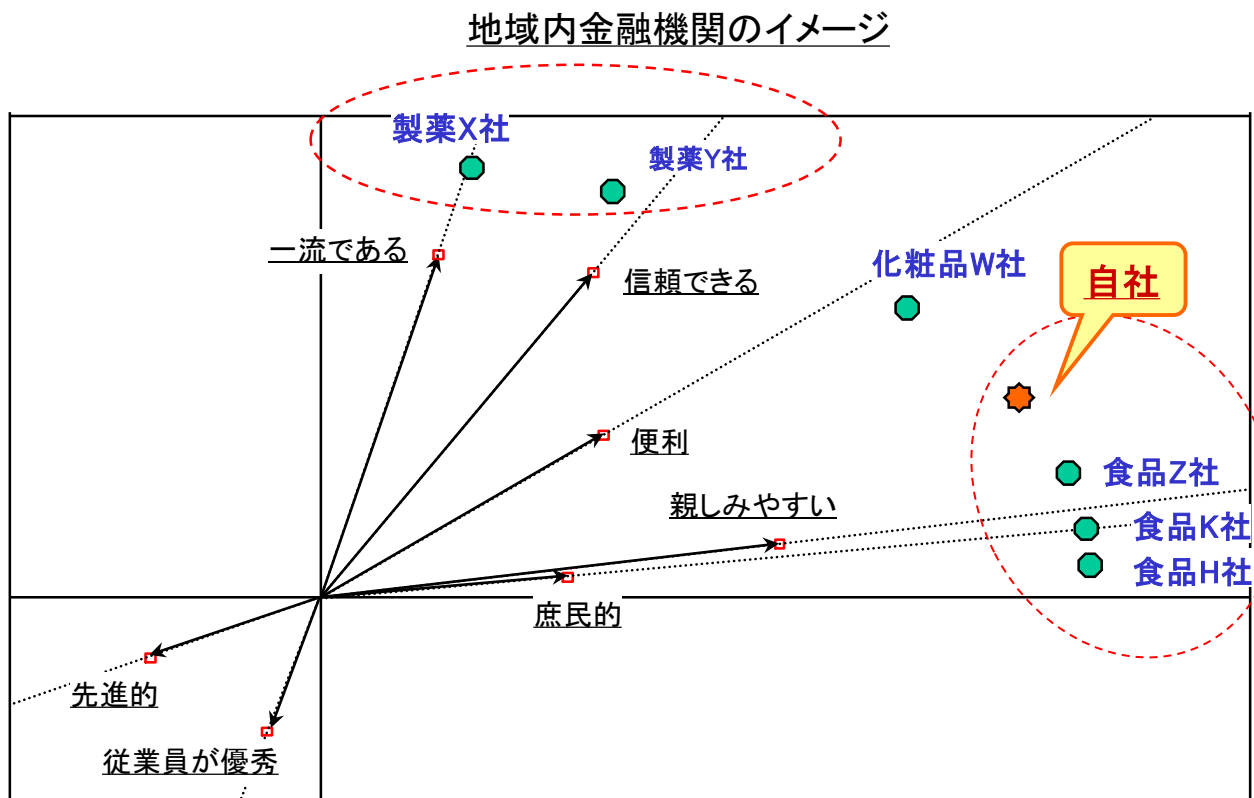
スーパーの棚割りなどでも多用

・「おむつ」を買った顧客は、「ビール」を買う確率が高い→隣同士の棚に置く

広告の宣伝コピーを作る - 多次元尺度構成法

< イメージのポジショニングマップ例 >

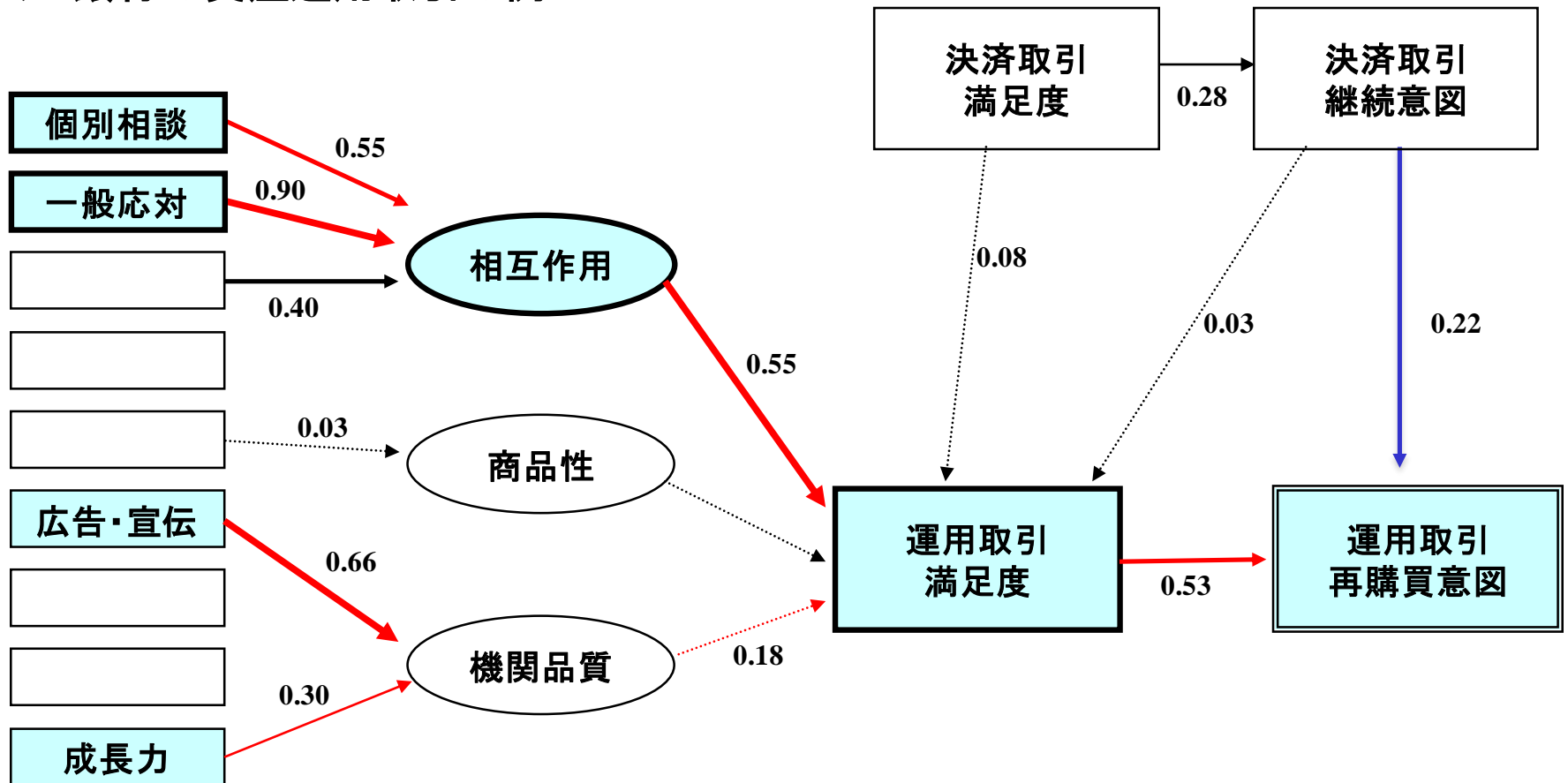
自社の認識（リサーチ前）・・・「製薬会社と競合」、「信頼されている」



顧客が自社と比較するのは、
製薬会社ではなく、食品会社であった。

自社のビジネス構造を再考する - 共分散構造分析

◆ 銀行の資産運用取引の例



出典: 戸谷圭子, Chance Management Processes in Financial Services - 金融サービスにおける再購買意図形成プロセスの研究 -, 2000/11/9 日本人工知能学会 第43回人工知能基礎論研究会

最低限必要と考えられる知識

最低限必要と考えられる知識

● 義務教育まで

インターネットやマスメディアで目にする世論調査などの統計値の確からしさを判断できる知識

- ✓ 分布・記述統計の基礎・・・そのデータの分布を表現するもの(平均値・分散・外れ値など、可視化による知覚の歪みも含めて)
- ✓ 標本調査(母集団)の考え方

● 高校卒業まで

社会の課題解決(因果関係や構造の理解)に統計がどのように活用できるかについての基礎知識

- ✓ システムデータの特徴や利用範囲
- ✓ 正規分布・確率分布
- ✓ 相関・検定