

第 11 章 クラスタ分析 —観測対象をグルーピングする

1-1 クラスタ分析の概要

❖ クラスタ分析 (cluster analysis) とは？

ある集団内の個人の属性や特性 (変量) に基づいて、似た者同士をいくつかのグループ (クラスタ) に分類する多変量解析法の 1 つ。特に複数の変数により個人を測定し、その特徴に基づいた個人の分類を行なう場合に有効な手法。

例：ある学習集団の英語能力を「読解」「語彙」「リスニング」の 3 つを用いて測定し、英語能力の傾向を調べる場合。

→ 集団内の学習者を個別に見ると、個人によって得点のパターンが異なることが分かる。

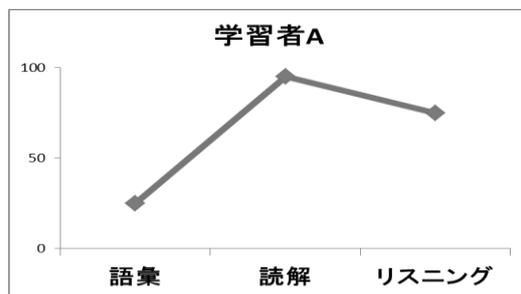


図 1 学習者 A の得点パターン

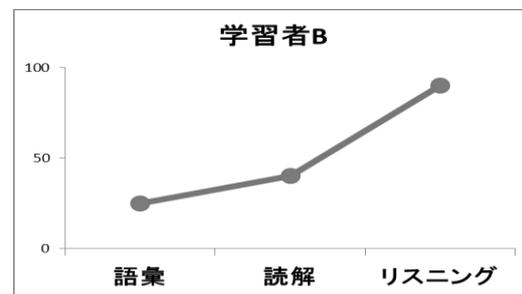


図 2 学習者 B の得点パターン

得点パターンを集団内の個人ごとに調べれば、学習者個人の英語能力について詳細な情報が得られるが、学習者の数が多ければ多いほど個人を検証するのは困難な上、全体としての傾向が見えにくくなる。

→ クラスタ分析を用いることで、似た得点パターンを示す学習者同士を「読解と語彙の得点は高いがリスニングの得点が低い群」や「語彙と読解の得点が低いリスニングの得点だけは高い群」のようにクラスタに分類し、個人差を踏まえた上での集団内の傾向を把握できる。

❖ 判断分析とクラスタ分析

- 分類するグループとその数を事前に想定するかどうかという点でクラスタ分析と判断分析 (discriminant analysis) とは異なる。
- 判断分析：ある変数をもとに、予め想定されたカテゴリーやグループにサンプルが分類されるのかどうかを判別する。説明変数と基準変数が明確に定められていて、説明変数で基準変数が予測可能かどうかを検証する。
- クラスタ分析：予め分類するグループを想定しないため分類基準の変数も存在しない。

1-2 使用可能なデータの種類

❖ クラスタ分析に使用できるデータ

- 相関係数が厳密に算出できる間隔尺度や比率尺度が望ましいが、順序尺度や2値の名義尺度 (e.g., 1, 0 データ) も使用可能。
- 得点幅や単位が異なるデータ (e.g., 身長と体重、年齢とテスト得点) を同時に投入してクラスタ分析を行なうことも可能。
⇒ この場合には、データの平均 0, 分散 1 への標準化を考慮する必要がある。標準偏差が大きい変数ほどクラスタの分類に大きく影響し、標準化の有無によって分類結果が異なる可能性がある。
- 標準化すべきかどうかの判断については厳密な基準はない。

❖ 分析の対象

サンプルクラスタ (Sample clustering)	観測された変数を元にサンプルとなる学習者をクラスタに分類する場合に用いる。最低1つの変数を用いることが条件
変数クラスタ (Variable clustering)	学習者から得られた複数の変数情報を元に、それらの変数自体をクラスタに分ける。変数間の相関係数を元に変数そのものを分類する点で、因子分析と類似した分析。最低3つの変数が必要。

- クラスタ分析では、被験者などのデータを総称して「個体」と呼ぶ。

1-3 クラスタ化

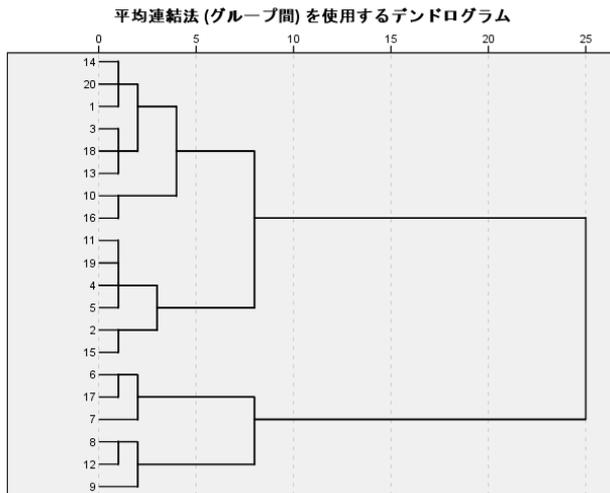
❖ クラスタ分析の手法

非階層的クラスタ分析 (Non-hierarchical cluster analysis)	あらかじめいくつかのクラスタに分けるかを決め、各個体がそれぞれのクラスタに当てはまるように計算を進める。
階層的クラスタ分析 (Hierarchical cluster analysis)	分析の前にはどんな傾向のグループがあるかわからないので、分析をしてその傾向を探し出す。

- クラスタ分析では似ているクラスタ同士を順に結合していき、最終的には1つのクラスタにまとまるまで計算が行なわれる。
- その際、非類似度の距離と、クラスタ間の距離が測定される。
- 非類似度の距離が遠い (値が大きい) ほど似ていないクラスタ同士が結合していると考えられる。クラスタ間の距離は、クラスタ同士を結合する基準となる。
※ 非類似度の計算法については、p5 参照。

1-4 樹形図

- クラスタ分析の結果は、デンドログラムという形で出力される。
- これは個別の傾向を持つ個体が、段階を追って最終的に1つの集団 (クラスタ) に集約されるまでの過程を図として表したもの。
- 個体同士が結合してクラスタを形成している距離が短い (結合距離の数値が低い) 地点で結合するほど、似た個体であるといえる。逆に、距離が長いところで結びついている個体同士は共通点が少ない。



縦方向には各個体が並べられている。
 横軸には 0 から 25 までの結合距離が
 示されている。
 各個体から伸びる水平線が垂直に曲
 がり他の水平線と結びつくと、クラス
 タが結合したことを示す。

図 3 デンドログラムの例図 1 学習者 A の得点パタ

- クラスタ分析は探索的な分析方法で、いくつかのクラスタに分類できそうかの判断には明確な基準がなく分析者の判断に委ねられる。
- ❖ クラスタ数の決定
- クラスタ数を決定するには、カットオフポイントを横軸のどこかに設定しクラスタを樹形図から切り取る必要がある。カットオフポイントの目安として、「結合距離が大きくなるとはねあがる、横の線が長くなる場所を探す」(磯田, 2004) がある。