

「クラスター分析 (1): とにかくやってみる」

■クラスター分析の目的と魅力 [テキスト pp.148-151]

- ・ 回答が似ている人々を集めてグループ分けしたいというのは自然な発想。
- ・ 2 変数くらいであれば、よく考えれば適切に分類できる。
- ・ 3 変数以上になると、誰と誰が似ているのか、統計的判断に頼る必要が出てくる。

↓

・ **クラスター分析** (cluster analysis)

複数の量的変数を見渡して、人々 (ケース) を適切なグループに分類する。

例) 複数の科目の勉強時間の変数から考えると、A 君と C 君は文系重視型で、
B・D・F 君は全体バランス型……。

- ・ 魅力 ①主観になりがちな分類が客観的な統計で保証されること。
②自分では思いつかないような分類が探索できること。

・ 階層的クラスター分析

似たケースを徐々に結合して行って、少数のグループにまとめていく。通常はこちら。

非階層的クラスター分析

でたらめな分類から始めて、徐々により適切なグループ分けに近づけていく。

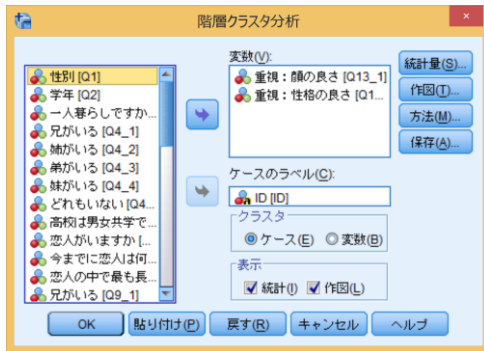
作業課題①

- (1) テキスト p. 259 の Web ページから「クラスター分析のデータ」をダウンロードする
 - (2) テキスト p. 151 に書かれているとおり、V3~V10 を人口比に応じた変数に変換しよう
 - (3) テキストの設定どおりにクラスター分析を実行して、結果を確認しよう。
-

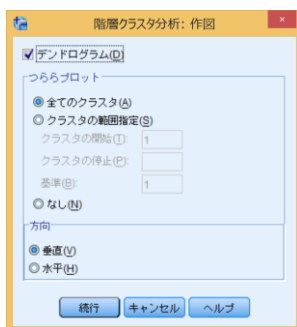
■SPSS でやってみよう [テキスト pp.151-164]

・ SPSS での回帰分析の操作

- ①メニューから、分析→分類→階層クラスタ
- ②分類のために用いる変数群をすべて [変数] 欄へ
(②' 結果をケース番号以外で示す場合は、ラベルの変数を [ケースのラベル] 欄へ)



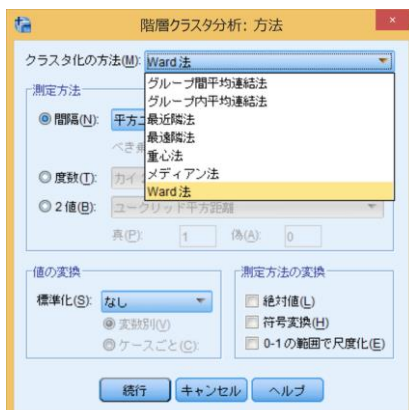
- ③ **作図** ボタンを押して、一番上の [デンドログラム] にチェックして **続行**



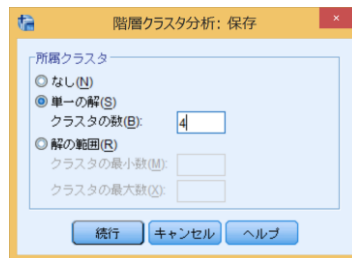
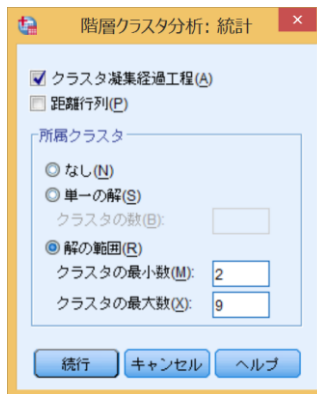
- ④ **方法** ボタンを押して、[クラスタ化の方法] を選択して **続行**

平均連結 [グループ間平均連結法] …… 近くのをどんどん吸収
完全連結 [最速隣法] …… グループが大きくなり過ぎないように
ウォード法 [Ward 法] …… 重心が動きすぎないようにバランスよく

- (④' ばらばらの単位の変数群を分析するときは、[標準化] で [Z 得点] を選択)



- ⑤ 各ケースがどのクラスターに分類されたかを、知りたいときは、**統計量** ボタンを押して、所属クラスターに自分が採用するクラスター数を入力する
(⑤' さらにその内容をデータに付け加えたい場合には、**保存** ボタンを押して、採用するクラスター数を入力する)



- ⑥ 元の窓で **OK** ボタン

作業課題②

(1) 「学生の恋愛観に関する調査」の1年生データを基にして、理想の恋人像(9変数)が似ている人々をクラスター分析で分類する。まず、直感的にどのような分類ができそうか想像してみよう。

(2) ウォード法、平方ユークリッド距離の設定でクラスター分析を実行し、デンドログラムから、適切なクラスター数を決定する。

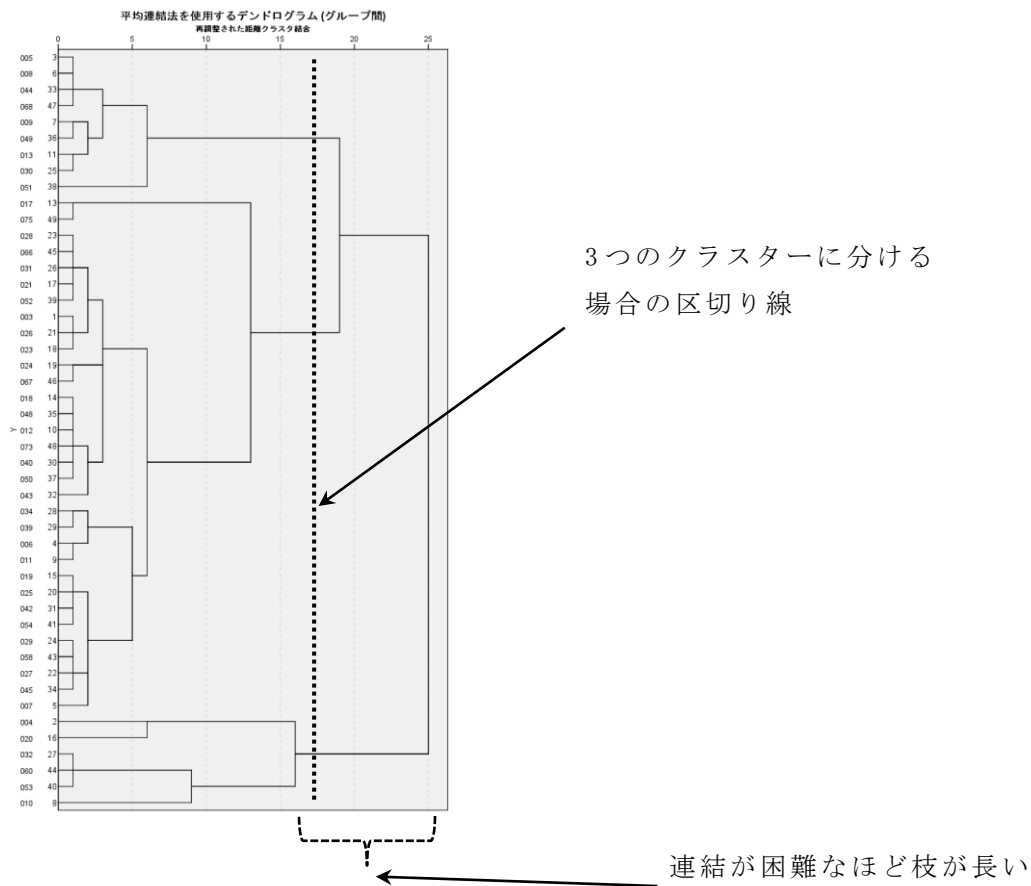
(3) 各変数の平均点を比較してクラスターの特徴を読み取り、クラスターに名称を付ける。

文章課題

当初想像した分類とクラスター分析の分類がどのように異なったか結果を記述した上で、この結果について自由な視点から考察しなさい。

■結果の読み取りとクラスターの特徴の把握

①デンドログラムからクラスター数を決定



②各変数の平均点をクラスター間で比較して、クラスターの特徴を把握

	クラスター1 (型)	クラスター2 (型)	クラスター3 (型)
顔の良さ			
スタイル			
ファッション			
頭の良さ			
性格の良さ			
ユーモア			
将来性			
趣味が合うか			
価値観が合うか			
分類ケース数			

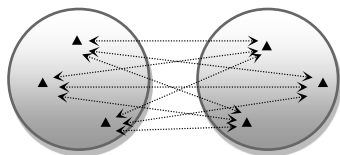
③分類結果をもとにして(変数として保存して)、他の変数との関連を調べるとおもしろい

「クラスター分析 (2) : 連結の種類を理解 + クラスターの有効利用」

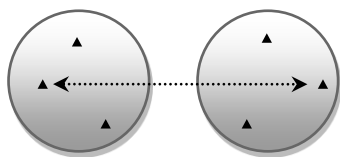
■ 連結の種類を理解 [テキスト pp.156-158]

- ・ クラスター分析は、回答の距離が近いものを「連結」してクラスターにまとめている。
- ・ 距離の測り方が決まっても、連結を進め方については複数の考え方がある。
- ・ 代表的な連結方法は 3 つ。
- ・ 分析目的に合う連結方法を理論的に選ぶのが理想だが、
実際的には複数の方法を試して納得のいく結果を選ぶとよい。

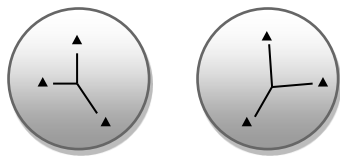
	形式的な定義	ねらい	欠点
平均連結 [グループ間平均連結法] (average linkage)	距離の平均値が一番小さいグループ同士を連結していく	近くのをどんどん吸収しておおらかな分類を考える	1 ケースずつ吸収していくような無意味なデンドログラムになることがある
完全連結 [最遠隣法] (complete linkage; furthest neighbor)	同じグループ内での一番遠いケース同士の距離が最小になるように連結していく	クラスターが広がり過ぎて類似性が損なわれないように注意する	周辺的なケースに引き寄せられて外れ値を重視した分類になることがある
ウォード法 [Ward 法] (Ward's method)	グループ内での重心 (平均的位置) のずれが最小になるように連結していく	分類にまとめることによって失ってしまう各ケースの個別情報の量を最小にする	無難で妥当だが意外性・発見性のない分類になることがある



平均連結 [グループ間平均連結法]

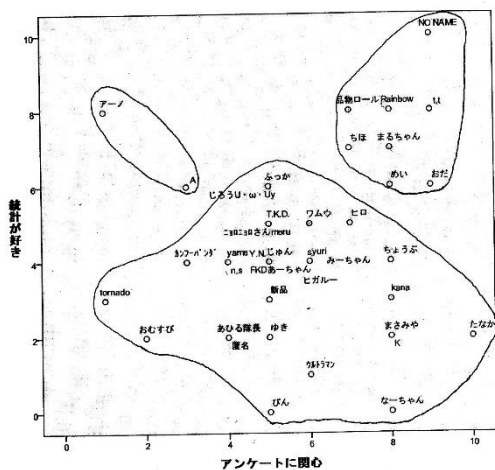
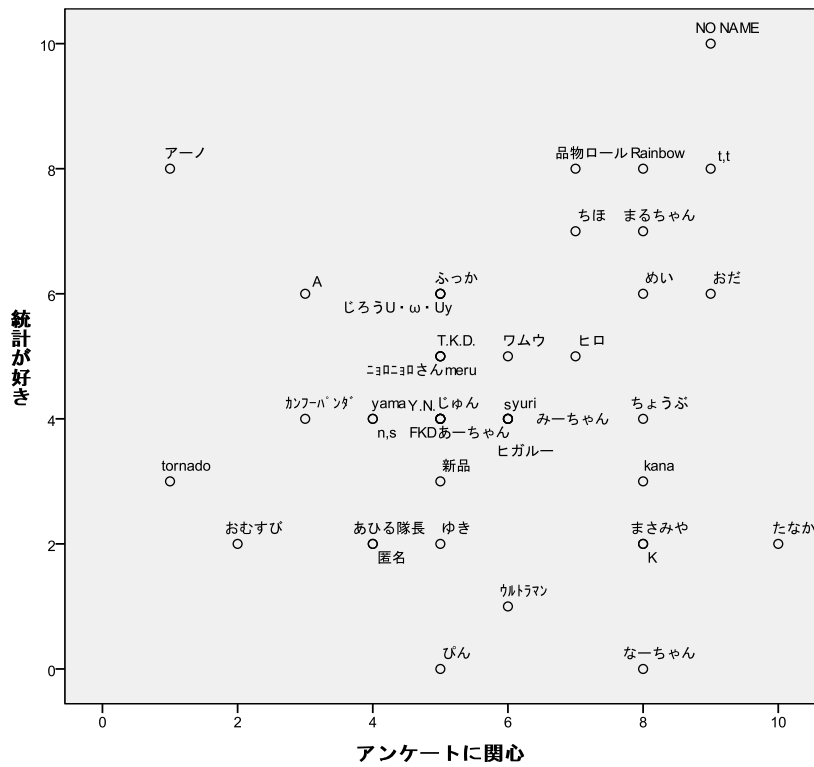


完全連結 [最遠接法]

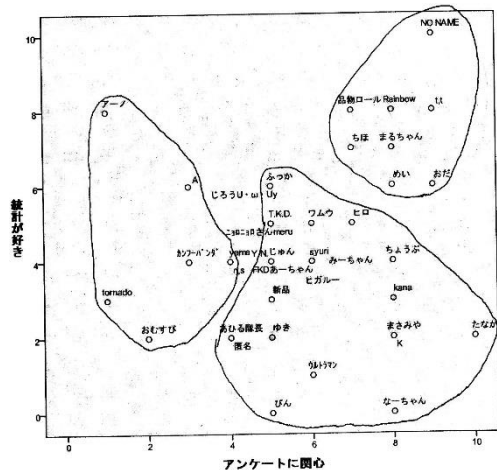


ウォード法 [Ward 法]

- ・ 本当は距離の測り方についても、複数の考え方があるが、ここではそこまで考えない。
(初期設定の「平方ユークリッド距離」のまま)



a 平均連結（グループ間平均連結法）の例



b 完全連結（最遠隣法）の例

作業課題①

(1) 「学生の恋愛観に関する調査」のフルデータ（2年生以上や関大生以外を含む）を用いて、理想の恋人像（9変数）が似ている人々をクラスター分析で分類する。前回と同じ「ウォード法」「標準化なし」でデンドログラムを析出しなさい。

(2) 連結方法を「グループ間平均連結法」「最遠隣法」に変更して、それぞれのデンドログラムを析出しなさい。

(3) 3つのデンドログラムを比較して、どの連結方法で、クラスターの数をいくつにすべきだと思うか、自分の意見を説明しなさい。

■ クラスターの有効利用

- ・ クラスター分析は、どんな分類が析出されるかということ自体もおもしろいが……
- ・ クラスターと他の変数の関係を調べると有意義。
- ・ あたかも最初からあった質問項目のように扱えばよい。

例) 理想の恋人として何を重視するか、4つのクラスターを析出したとする。

⇒ 「Qxx 理想の恋人を考えるととき何を重視するタイプですか。1つ選んでください。

1 容姿重視型 2 性格重視型 3 ○○型 4 ××型 」

という質問項目があったらと思って分析すればよい。

⇒ 他の変数とのクロス表

クラスターごとに他の変数の平均値を比較

ダミー変数として回帰分析に利用 など

作業課題②

(1) 理想の恋人像について、先ほど採用した連結方法・クラスター数で、各クラスターの特徴(各項目の平均値)を読み取りなさい。「○○タイプ」などの名前を付けてSPSSの「値ラベル」に入力すること。

(2) 各クラスターへの所属率が男女で異なるのか、分析しなさい。

(3) どのクラスターで「今までの恋人の人数」が多いのか(少ないのか)、分析しなさい。

文章課題

理想の恋人像のクラスターを有効に利用する分析関心を考えて、分析を実行しなさい。結果を1つの表またはグラフにまとめたうえで、

- ① 目的(何を知りたいのか)
- ② 方法(具体的にどの質問項目でクロス表を作る、平均値を比較するなど)
- ③ 結果(図表をどう読めばよいのか客観的に)
- ④ 考察(注目すべき点、意外だった点など)

をなるべく短い文章で説明しなさい。

■ (参考) 非階層的クラスター分析 [テキスト pp.164-165]

- ・ ケース数が多い場合は計算が膨大で、デンドログラムが読み取り困難なことがある。
- ・ その場合、階層的クラスター分析ではなく「非階層的クラスター分析」を使う。

あらかじめクラスターの数を指定して、ランダムなケース配分から開始

→ 1ケースずつ所属を移動させて、分類結果がよくなるなら採用、を繰り返す

→ いくら移動させても、分類がよくなることがなくなった時点で分析を止める

→ 分析結果は、クラスターごとに各変数の平均値を一覧にする

- ・ SPSSでは「分析 → 分類 → 大規模ファイルのクラスタ」