

2 段数量化分析法

Two-step analysis with unified quantification

○岡本安晴

日本女子大学人間社会学部

1 はじめに

カテゴリ変数の数量化を分析法に依存せずに行うことを提案する。まず、(第1段) カテゴリ変数ごとに数量化を行い、その後(第2段)、数量化された変数に対して多変量解析を行う(図1)。第2段の多

変量解析に先立って数量化を行うので、数量化された変数の解釈は各カテゴリ変数ごとに第2段の多変量解析とは独立して行うことができる。

第2段の分析として主成分分析を行う場合には村上

(1999)を基に

したものとして岡本(2013)で報告した。岡本(2013)の図1と同じ結果はMFAによっても得ることができる(矢野、2014)。但し、MFAでは第1固有値 λ_1 のみが使われるが、本報告における数量化では有効な固有値すべてを使っている。

回帰分析と判別分析では、以下に報告するように数量化1類と2類による分析と同じグラフが得られた。第1段での数量化が第2段での分析と独立して行われるのにも関わらず、数量化と分析が同時に行われる数量化1類あるいは2類と同じグラフが得られることは、2段数量化分析法による分析結果と他の数量化法による分析結果との間の共通性を示すものであるとともに、第1段において得られる数量化を第2段での様々な分析にわたって共通に使用することの保証を与えるものであると考えることができる。なお、2段に分けて判別分析を行う方法は、趙・田中(2009)でも提案されているが、趙らの方法ではカテゴリ変数間の相関に応じて第1段の数量

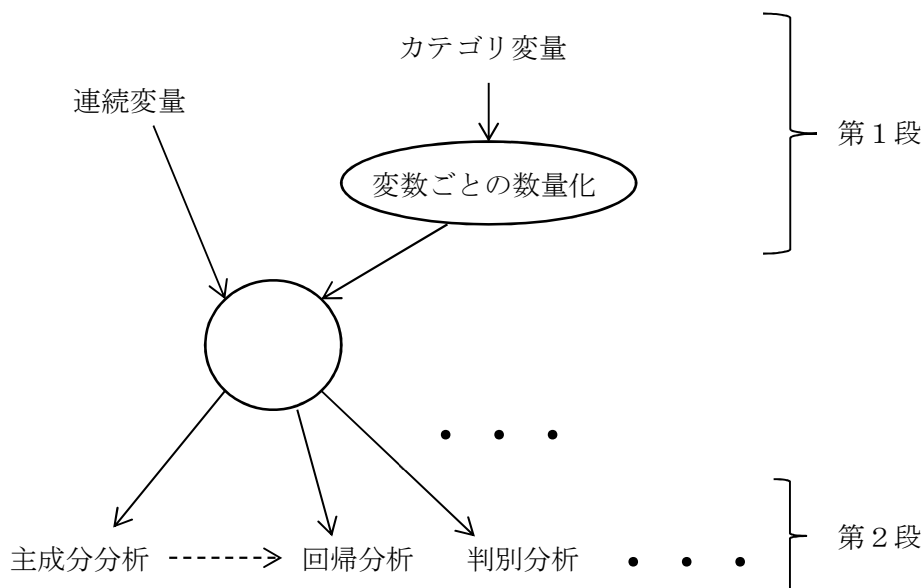


図1 2段数量化分析法

化の方法が選ばれている。しかし、本報告における数量化はカテゴリ変数単位で行われるもので、カテゴリ変数間の相関の強さとは独立して行うことができる。

第2段での分析が回帰分析あるいは判別分析の場合の分析例を示す前に、両分析に共通である第1段での数量化について簡単に説明する。

2 第1段：数量化

カテゴリ変数 k の指標行列 G_k (サイズ $n \times p_k$) に対して

$$H_k = G_k - \frac{1}{n} \mathbf{1}_n \mathbf{f}'_k, \quad \mathbf{f}'_k = \mathbf{1}'_n G_k$$

とおく。このとき、

$$\omega' \omega = 1$$

の制約のもとで分散

$$(H_k \omega)' (H_k \omega)$$

を最大にする ω は

$$H'_k H_k \omega = \lambda \omega$$

を満たすものとして求めることができる。このとき

$$(H_k \omega)' (H_k \omega) = \lambda$$

である。

$H'_k H_k$ の j 番目の固有値 λ_j (但し、 $\lambda_j \neq 0$ のもの)と固有ベクトル ω_j により与えられる数量化標準得点を

$$z_j = H_k \omega_j / \sqrt{\lambda_j / n}$$

とおき、 G_k の数量化得点を

$$G_k^z = \left(z_1^{(k)} \quad \dots \quad z_{p_k-1}^{(k)} \right)$$

で与える。

3 第2段：回帰分析の場合

連続型の従属変数 y に対して

$$y = c + \sum_{k,j} \beta_j^{(k)} z_j^{(k)} + \varepsilon \quad (1)$$

と回帰分析を行う。岩坪 (1987) のデータ (表 2.1) に対して式 (1) による回帰分析を行い、 y の予測値

$$\hat{y} = \hat{c} + \sum_{k,j} \hat{\beta}_j^{(k)} z_j^{(k)}$$

を求めると図2のようになる。ここで、 \hat{c} および $\hat{\beta}_j^{(k)}$ は最小2乗法による推定値である。図2は、岩坪 (1987) の数量化1類による分析結果 (図 2.1) と同じである。

なお、 $z_j^{(k)}$ などの独立変数間の相関が強いために $\beta_j^{(k)}$ の推定が不安定になるときは、まず主成分分析などを行って独立変数間の関係を整理してから回帰分析を行えばよい。

4 第2段：判別分析の場合

K 個のカテゴリ変数の数量化得点 G_k^z ($k = 1, \dots, K$)が与えられているとき、 g 個のグループ分けに対してFisherの線形判別関数を求める。第 h グループの i 番目の個体の数量化の値を

$$\mathbf{x}_{hi} = (x_{hi1} \ \dots \ x_{hip})'$$

とおく。ここで、 x_{hi1}, \dots, x_{hip} は、グループ h の i 番目の個体の $z_1^{(1)}, \dots, z_{p_k-1}^{(K)}$ の値を表す。各 $z_j^{(k)}$ の平均値は0であることに注意。 x_{hij} の線形結合

$$y_{hi} = \sum_j v_j x_{hij} = \mathbf{v}' \mathbf{x}_{hi}$$

において、 y_{hi} のグループ間変動とグループ内変動の比が最大になるように \mathbf{v} を求める。グループ間変動は

$$SS_{between} = \sum_{h,i} (\mathbf{v}' \bar{\mathbf{x}}_h - \mathbf{v}' \bar{\mathbf{x}}_{..})^2 = \mathbf{v}' \mathbf{A} \mathbf{v}$$

で与えられる。 $\bar{\mathbf{x}}_h$ と $\bar{\mathbf{x}}_{..}$ はグループ h における平均とデータ全体にわたる平均を表し、

$$\mathbf{A} = \sum_{h,i} \bar{\mathbf{x}}_h \bar{\mathbf{x}}_h'$$

である。グループ内変動は

$$SS_{within} = \sum_h \left\{ \sum_i (\mathbf{v}' \mathbf{x}_{hi} - \mathbf{v}' \bar{\mathbf{x}}_h)^2 \right\} = \mathbf{v}' \mathbf{W} \mathbf{v}$$

で与えられる。ここで、

$$\mathbf{W} = \sum_{h,i} (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)(\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)'$$

である。

グループ間変動とグループ内変動の比

$$Q = SS_{between} / SS_{within} = \mathbf{v}' \mathbf{A} \mathbf{v} / \mathbf{v}' \mathbf{W} \mathbf{v}$$

を最大にする \mathbf{v} を、条件

$$\mathbf{v}' \mathbf{W} \mathbf{v} = 1$$

のもとで求める。求める解は

$$\mathbf{A} \mathbf{v} = \lambda \mathbf{W} \mathbf{v}$$

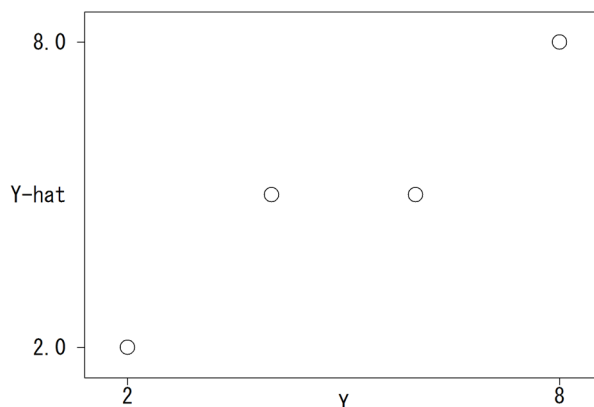


図2 (y, y-hat)の散布図

を満たす λ と \mathbf{v} の組として求められ、このとき

$$Q = \lambda$$

となる。

岩坪 (1987) のデータ (表 3.1) に対して、2 段数量化分析法による判別分析を行うと図 3 の結果を得る。判別関数を λ の大きさの順に判別関数 1、判別関数 2 と表し、それぞれの値を横軸と縦軸の座標値としてプロットしたものである。小文字の英字が各グループ A、B、C の個体の座標値による位置、大文字がグループの平均値を座標値とする点を表す。図 3 は、岩坪 (1987) の図 3.1 に示されている数量化 2 類による分析結果と同じである。

判別関数と数量化変数との関係はそれらの間の相関を表す構造行列によって解釈することができる。

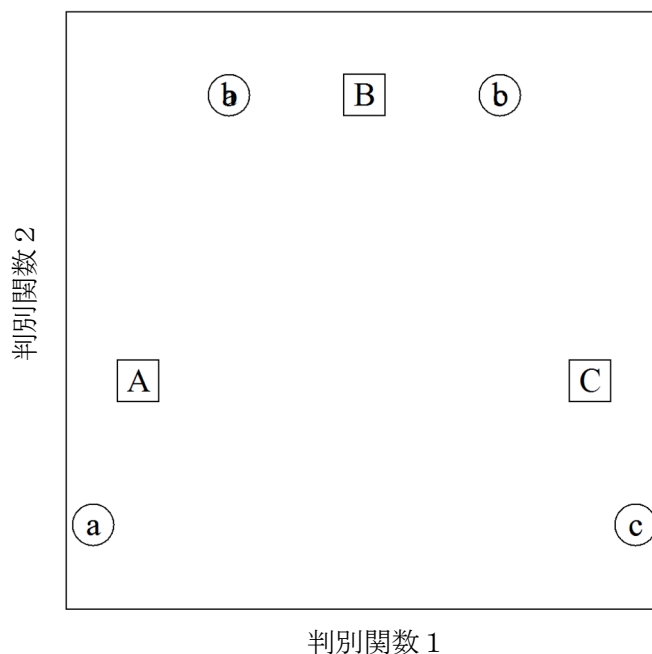


図 3 個体 (英小文字) とグループ平均値 (英大文字)

5 まとめ

カテゴリ変数が含まれる場合、第 1 段で変数ごとの数量化を行い、第 2 段において第 1 段で得られた共通の数量化変数に対して種々の分析を行う 2 段数量化分析法による分析例の結果が、他の分析法 (MFA、数量化 1 類、数量化 2 類) による解と同じものであったことは、2 段数量化分析法によって得られる情報が他の分析法によるものと共通のものであることを示唆するものである。第 1 段において、第 2 段における分析法とは独立して各カテゴリ変数の数量化が行われているので、これによって得られた数量化変数の意味付けは当該のカテゴリ変数との関係のみによって行われる。そのようにして意味づけられた数量化変数の所与のデータにおける機能あるいは構造が第 2 段における分析によって明らかにされる。

引用文献

- 岩坪秀一 (1987) 数量化法の基礎. 朝倉書店.
村上隆 (1999) 科学研究費補助金 (09610114) 研究成果報告書.
岡本安晴 (2013) 日本行動計量学会第 41 回大会抄録集、192-195.
趙雪艶・田中 豊 (2009). 計算機統計学、91-108.
矢野 環 (personal communication, 2014).