北海道教育大学学術リポジトリ
Hokkaido University of Education Repository

国立大学法人 北海道教育大学
HOKKAIDO UNIVERSITY OF EDUCATION

Title	数量化理論第3類のスコア間の関係 ： カテゴリカルデータとアイテム・カテゴリーデータの場合
Author(s)	坪内，昭夫
Citation	北海道教育大学紀要．第二部．A，数学・物理学・化学・工学編，30(1)：13-18
Issue Date	1979-09
URL	http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/6029
Rights	

Hokkaido University of Education

北海道教育大学紀要（第２部Ａ）第30巻 第１号
Journal of Hokkaido University of Education (Section II A) Vol. 30, No. 1

昭和54年９月
September, 1979

数量化理論第３類のスコア間の関係
――カテゴリカルデータとアイテム・カテゴリーデータの場合――

坪　内　昭　夫
北海道教育大学旭川分校数学教室

On Relations between the Scores of Quantification by the Theory of Quantification (Method of Case 3) ; in Case of Categorical Data and Item-Category Data

Akio TSUBOUCHI
Mathematics Laboratory, Asahikawa College, Hokkaido University of Education, Asahikawa 070

Abstract

We apply the theory of quantification (method of case 3) to categorical data. Comparing with the case of item-category data, we show the relation of category score to sample score and to grouping score. And we describe how these are changed by means of a linear transformation of category score.

1　はじめに

アイテム・カテゴリーデータに，数量化理論第３類を適用することによって，得られるカテゴリースコアとグルーピングスコアとの間に正比例関係があること，および，そのことによってグルーピングスコアが分析の中心的な役割を果すことについては前に述べたところである（参考文献(2)）.

本稿では，アイテム・カテゴリーデータの場合と対比しつつ，より一般なカテゴリカルデータに数量化理論第３類を適用したときに，得られる諸スコア（カテゴリースコア，サンプルスコア，グルーピングスコア）の間に成り立つ関係について述べる．さらに，サンプルスコアの平均と分散を特定の値に固定するためにおこなうカテゴリースコアの一次変換（寸法調整）による，各スコア間の関係の保存性についても言及する．

表 1　カテゴリカルデータの反応表

サンプル＼カテゴリー	1	⋯	j	⋯	M	反応数
1	1	⋯	$\delta_1(j)$	⋯	0	m_1
⋮	⋮		⋮		⋮	⋮
i	0	⋯	$\delta_i(j)$	⋯	1	m_i
⋮	⋮		⋮		⋮	⋮
N	0	⋯	$\delta_N(j)$	⋯	1	m_N
反応したサンプル数	n_1	⋯	n_j	⋯	n_M	T

2　データと表記法

（2－1）対象（サンプル）について，1つの特性（アイテム）に1つのカテゴリー（小分類項目）を対応させて得られるデータをカテゴリカルデータといい，各サンプルがM個の特性で記述されるときには，M個のカテゴリーが設けられる．

また，サンプルについてアイテムが複数個あり，さらに各アイテムが複数個のカテゴリーに分かれているとき，このよう分類はアイテム・カテゴリー分類といわれる．アイテム・カテゴリー分類によるデータで，制約条件"各サンプルは，各アイテムにおいて唯一つのカテゴリーにのみ反応する"を満たすものを，本稿ではアイテム・カテゴリーデータという．

（2－2）N個のサンプルがM個のカテゴリーに反応しているカテゴリカルデータにたいして，カテゴリー変量を次のように定義する：

$$\delta_i(j)=\begin{cases}1\cdots\cdot\text{サンプル}\ i\ \text{がカテゴリー}\ j\ \text{に反応したとき}\\[2ex] 0\cdots\cdot\text{そうでないとき}\end{cases}$$

カテゴリー j に反応したサンプルの数を n_j，サンプル i が反応したカテゴリーの数（反応数）を m_i とすると，$n_j=\sum_i \delta_i(j)$ （$j=1,\ 2,\ \cdots,\ M$），　$m_i=\sum_j \delta_i(j)$ （$i=1,\ 2,\ \cdots,\ N$） であって，反応総数 T は，$T=\sum_i m_i=\sum_j n_j=\sum_i\sum_j \delta_i(j)$ となる（表1）．

アイテム・カテゴリーデータでは，カテゴリー変量は次のように定義される：

$$\delta_i(\alpha j) = \begin{cases} 1 \cdots\cdots サンプル i が \alpha アイテムの j カテゴリーに反応したとき \\ 0 \cdots\cdots そうでないとき \end{cases}$$

アイテムの数をK, αアイテムにおけるカテゴリーの数をk_α, αアイテムにおけるjカテゴリーに反応したサンプルの数を $n_{\alpha j}$ とすると，$n_{\alpha j}=\sum_i \delta_i(\alpha j)$ ($\alpha=1$，2，…，K; $j=1$，2，…，k_α) であり，各アイテムに反応したサンプルの数はすべて等しく $N=\sum_j n_{\alpha j}$ ($\alpha=1$，2，…，K) となる．さらに，制約条件は $\sum_j \delta_i(\alpha j)=1$ ($i=1$，2，…，N; $\alpha=1$，2，…，K) と表わされ，従って各サンプルの反応数はすべて等しく $K=\sum_\alpha\sum_j \delta_i(\alpha j)$ ($i=1$，2，…，N) となる．

以下においては，アイテムの区別は本質的な役割を果さいのでその区別はしない．そして，制約条件に留意しつつカテゴリカルデータの形で表現し，その表記法を用いて簡略に記述することにする．

3　カテゴリースコア，サンプルスコアおよびグルーピングスコア

（3－1） 数量化法を適用した結果，各カテゴリーと各サンプルに与えられる数値を，それぞれカテゴリースコアとサンプルスコアという．

カテゴリースコア　x_j ($j=1$，2，…，M) とサンプルスコア　y_i ($i=1$，2，…，N) は，xとyとの相関係数　$\rho=S(x,y)/S_xS_y$　($S(x,y)$ は共分散；S_x^2, S_y^2 は分散) を最大にするように与えられる．すなわち，x_jは条件　$\bar{x}=\sum_j n_jx_j/T=0$　のもとで基本方程式

$$\sum_j \left[\sum_i \delta_i(j)\ \delta_i(k)/m_i\right] x_j=\rho^2 n_k x_k \quad (k=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots (1)$$

を解くことによって求められ，また y_i は正規方程式

$$\sum_j \left[\delta_i(j)/T-n_jm_lT^2\right] x_j=\rho(S_x/S_y)(m_l/T)\left[y_l-\sum_i m_iy_i/T\right] \quad (l=1,\ 2,\ \cdots,\ N)$$

と　$\bar{x}=0$　とから，　$\sum_i m_iy_i/T=0$，　　$S_y/(\rho S_x)=1$ とおいて

$$y_i=\sum_j \delta_i(j)\ x_j/m_i \quad (i=1,\ 2,\ \cdots,\ N) \qquad \cdots\cdots (2)$$

と与えられることは周知のとおりである（参考文献(1), (3), (4), (5)）．

アイテム・カテゴリーデータに対しては，（１）式（２）式において m_i がK（反応数＝アイテムの数）に変るだけである．

（3－2） サンプルの属性を表わす変数をグルーピング変数といい，グリーピング変数の各カテゴリーごとにそのカテゴリーに反応するサンプルのサンプルスコアの平均をとり，それをそのカテゴリーに与えられるグルーピングスコアとよぶことにする．

グルーピングスコア z_k は，

$$z_k=\sum_i \delta_i(k)\ y_i/n_k \qquad (k=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots (3)$$

で求められる.

アイテム・カテゴリーデータの場合も全く同様である.

（3－3）軸の算出に用いたアイテムも，グルーピング変数として取り扱うことができる．このとき，同一のカテゴリーにカテゴリースコア x とグルーピングスコア z が与えられるが，これら2つのスコア間の関係を導く．

$$z_k=\sum_i \delta_i(k)\ y_i/n_k=\frac{1}{n_k}\ \sum_j \left[\sum_i \frac{\delta_i(j)\ \delta_i(k)}{m_i}\right]\ x_j$$

これを基本方程式（1）に代入して

$$z_k=\rho^2 x_k \qquad (k=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots(4)$$

すなわち，グルーピングスコアは，固有値 ρ^2 を比例定数としてカテゴリースコアに正比例する．

4 一次変換と各スコア

（4－1）一般に，各スコアを算出するためのコンピュータプログラムでは，出力されるスコアに寸法調整がなされる．

いま，"サンプルスコア y_i の平均を0，分散を a^2（a は定数）とする"ような寸法調整を考える．

基本方程式の解を x'_j とし，これにより算出されるサンプルスコア

$$y'_i=\sum_j\ \delta_i(j)\ x'_j/m_i \qquad (i=1,\ 2,\ \cdots,\ N) \qquad \cdots\cdots\cdots(5)$$

の平均を $\bar{y}'$，分散を $S^2_{y'}$ とし，一次変換

$$y_i=a\,(y'_i-\bar{y}')/S_{y'} \quad (a\text{は定数};\ i=1,\ 2,\ \cdots,\ N) \qquad \cdots\cdots\cdots(6)$$

をおこなう．（5）を（6）に代入して

$$y_i=\frac{a}{S_{y'}}\ \sum_j \delta_i(j)\ (x'_j-\bar{y}')/m_i$$

ここで

$$x_j=a(x'_j-\bar{y}')/S_{y'} \qquad (j=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots(7)$$

とおくと

$$y_i=\sum_j \delta_i(j)\ x_j/m_i \qquad (i=1,\ 2,\ \cdots,\ N) \qquad \cdots\cdots\cdots(8)$$

すなわち，（7）によって，出力されるカテゴリースコアが計算され，同時に（8）によって，寸法調整されたサンプルスコアが算出される．

（4－2）アイテム・カテゴリーデータの場合は，事情は少し異なる．

それは，アイテム・カテゴリーデータであることの制約条件から $m_i=K$，さらに基本方程式の解からサンプルスコアを求めるときの条件 $\sum_i m_i y'_i/T=0$ から $\bar{y}'=\sum_i Ky'_i/T=0$ となる特

殊な性質があることによる．

このことからアイテム・カテゴリーデータでは

$$x_j = ax'_j/S_{y'} \quad \text{または} \quad x_j = ax'_j/(KS_{y'}) \qquad (j=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots (9)$$

とおいて，それぞれ

$$y_i = \sum_j \delta_i(j)\ x_j/K,\ \ y_i = \sum_j \delta_i(j) x_j \qquad (i=1,\ 2,\ \cdots,\ N) \qquad \cdots\cdots\cdots (10)$$

となる．

5　正比例関係の保存性

（５－１）一次変換（７）によるグルーピングスコアとカテゴリースコアとの間の正比例関係の保存性について調べる．

カテゴリカルデータに対しては

$$z_k = \sum_i \delta_i(k)\ y_i/n_k = (a/S_{y'})\left[\sum_j\left\{\sum_i \frac{\delta_i(j)\delta_i(k)}{m_i n_k}\right\}x'_j - \bar{y}'\right]$$

これがx'_jに関する基本方程式（１）により

$$z_k = (a/S_{y'})\ (\rho^2 x'_k - \bar{y}')$$

さらに，一次変換（７）によって

$$z_k = \rho^2 x_k + \{a(\rho^2 - 1)/S_{y'}\}\ \bar{y}' \qquad (k=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots (11)$$

となる．

すなわち，寸法調整のための一次変換を施すと，カテゴリカルデータでは，出力されるグルーピングスコアとカテゴリースコアとの間には正比例関係が必ずしも成立しなくなる．それが保存されるのは，　$\overline{y'}=0$　または$\rho^2=1$　のカテゴリカルデータの場合で，そのとき，
$z_k=\rho^2 x_k$　または　$z_k=x_k$　となる．

（５－２）アイテム・カテゴリーデータは，　$\bar{y}'=0$　を満たすカテゴリカルデータであるから正比例関係は保存される．しかし，その比例定数は，一次変換（９）のいずれの式を用いるかによってK倍だけ異なる．

実際，（５－１）と全く同様にして

$$z_k = \rho^2 x_k \quad \text{または} \quad z_k = K\rho^2 x_k \qquad (k=1,\ 2,\ \cdots,\ M) \qquad \cdots\cdots\cdots (12)$$

となる．

6　むすび

以上によって次のことがわかる：

1）固有ベクトルの要素としてのカテゴリースコアと，それに対応するグルーピングスコアとの間には，固有値を比例定数とする正比例関係がある．

2）寸法調整のための一次変換によってこの関係は，

（ｉ）　カテゴリカルデータでは一般には保存されない．

（ii）　アイテム・カテゴリーデータでは保存される．

なお，本稿を草するにあたって，北海道大学大型計算機センターを利用したことを付記する．

参考文献

(1)　斉藤堯幸，小川定暉，野嶋栄一郎（1972）データ解析（１）—— 数量化理論に関する総合報告，総研紀要，２（１），23－140ページ，（ユニバック総合研究所）．

(2)　坪内昭夫（1979）数量化理論第３類におけるカテゴリースコアとグルーピングスコア，北教大紀要（第２部Ａ），29（２），11－16．

(3)　林知己夫，樋口伊佐夫，駒沢勉（1970）情報処理と統計数理．産業図書．東京．400頁．

(4)　三宅一郎，中野嘉弘，水野欽司，山本嘉一郎（1977）ＳＰＳＳ統計パッケージII解析編．東洋経済新報社．東京．318頁．

(5)　安田三郎，海野道郎（1977）（改訂２版）社会統計学．丸善．東京．340頁．