〈研究ノート〉

探索的多変量解析によるデータ解析*

——多変量対応分析——

中　山　慶一郎**

1. はじめに

今回は多重対応分析、MCA（Multiple Correspondence Analysis）をとりあげ、その理論と、R を用いた分析について述べる。MCA は、意識調査で一般的に行はれている調査に対応する分析方法であり、主成分分析（PCA）、対応分析（CA）に類似した手法である。幾つかの変数のカテゴリーについて選択したデータを、低次元のパターンに表示する手法の一つである。MCA では、カテゴリー変数（categorical variables）を処理するが、量的変数（continuous variables）もその分析の中に取り入れて処理することもできる。

本稿では、MCA の考え方を簡単な例を提示して例証し、その実行過程を、R を用いて説明する。更に、すでに分析した日本とドイツの調査データを用いて、2 国間の意識の相違いを探ることにし、幾つかの分析方法を提示し、MCA の性質を調べることを目標にする。

2. Multiple Correspondence Analysis について

a. 社会調査に於いては、n 個の個体が Q 個の質問に答えるのが一般的な形式である。Q 個の質問は、それぞれ幾つかの選択肢から 1 つを選んで回答する形式を有する。

簡単な例として、ここで取り上げるデータ（taste）は、食品の味覚について、酸味と苦みの程度の報告の結果である。9 人の回答者に酸味 source と、苦み bitter の 2 つの変数から、その程度を表すカテゴリー 1 つを選択し、それを番号で示したものである。（第 1 表）

9 個の個体 indiviual の category の選択を 0, 1 の数値で示した表が第 2 表である。

この表は individual × category $(n \times \Sigma q)$ の dummy variable table であり、その値は n_{ij} は、0, 1 の 2 値で表される。行 row と列 col の周辺度数 n_{i+}, n_{+j} と、その分布 $n_{i+}/n = r_i$ 及び $n_{+j}/n = c_j$ を表示している。周辺度数の分布は、MCA の理論に重要な役割を果たしている。第 2 表のデータを比率にしたものが、第 3 表である。

周辺度数分布は、individual（行）と変数 variables（列、（Q＝2））の変動を示し、profile ともいい、行または列の変動の状態を表す。これらを比率で示したものを average profile といい、masses ともいう。MCA のデータでは、individual の masses は CA に比べて、1/n となるのが特色である。masses は、行と列の weights で、それぞれの重要性を表している。これらの表を見ると、individual

第 1 表　データ（taste）

	source	bitter
1	1	1
2	1	1
3	1	1
4	2	1
5	2	2
6	2	2
7	3	2
8	3	2
9	3	2

*キーワード：多重対応分析、MCA、R
**関西学院大学名誉教授

第 2 表　Indicator matrix Data Table n_{ij}

	s 1	s 2	s 3	b 1	b 2	Total n_{i+}	average profile
1	1	0	0	1	0	2	0.111111
2	1	0	0	1	0	2	0.111111
3	1	0	0	1	0	2	0.111111
4	0	1	0	1	0	2	0.111111
5	0	1	0	0	1	2	0.111111
6	0	1	0	0	1	2	0.111111
7	0	0	1	0	1	2	0.111111
8	0	0	1	0	1	2	0.111111
9	0	0	1	0	1	2	0.111111

第 3 表　比率、original profile $p_{ij}=n_{ij}/n$

	s 1	s 2	s 3	b 1	b 2	r_j
1	0.055556	0	0	0.055556	0	0.111111
2	0.055556	0	0	0.055556	0	0.111111
3	0.055556	0	0	0.055556	0	0.111111
4	0	0.055556	0	0.055556	0	0.111111
5	0	0.055556	0	0	0.055556	0.111111
6	0	0.055556	0	0	0.055556	0.111111
7	0	0	0.055556	0	0.055556	0.111111
8	0	0	0.055556	0	0.055556	0.111111
9	0	0	0.055556	0	0.055556	0.111111
$'c_i$	0.166667	0.166667	0.166667	0.222222	0.277778	1

第 4 表　列プロファイル　a_{ij}

	s 1	s 2	s 3	b 1	b 2	Σ
1	0.5	0.5	0	0	0	1
2	0.5	0.5	0	0	0	1
3	0.5	0.5	0	0	0	1
4	0	0.5	0.5	0	0	1
5	0	0	0.5	0.5	0	1
6	0	0	0.5	0.5	0	1
7	0	0	0	0.5	0.5	1
8	0	0	0	0.5	0.5	1
9	0	0	0	0.5	0.5	1

第 5 表　行プロファイル　b_{ij}

	s 1	s 2	s 3	b 1	b 2
1	0.333333	0.25	0	0	0
2	0.333333	0.25	0	0	0
3	0.333333	0.25	0	0	0
4	0	0.25	0.333333	0	0
5	0	0	0.333333	0.2	0
6	0	0	0.333333	0.2	0
7	0	0	0	0.2	0.333333
8	0	0	0	0.2	0.333333
9	0	0	0	0.2	0.333333
Σ	1	1	1	1	1

について、(1, 2, 3) と (5, 6,) と (7, 8, 9) が類似の反応を示しているのが見られる。

第 3 表は元のデータを比率に変換した（$p_{ij}=n_{ij}/n$）もので、これを profile という。これを行と列についての profile に変換すると、observed row profile（$a_{ij}=n_{ij}/n_{i+}$）、第 4 表、および observed column profile（$b_{ij}=n_{ij}/n_{+j}$）、第 5 表を求める。

b. 距離（distance）について

MCA では、データは、individual から見ると、9 個のデータは 5 つの category が示す 5 次元の空間の点とみなし、variable の各データは 5 つの category が示す 9 次元の空間の点と解することが出来る。これら多次元の空間の点の集まりを低次元の空間の集まりに縮約し可視化してパターン化するように工夫するものである。ここで問題となるのが各点間の距離を如何にとらえるかである。MCA では χ^2 距離で測定する。χ^2 統計量は、統計学ではクロス表の独立性の検定に用いられてきたものである。

カイ 2 乗統計量は、

$$\chi^2=\Sigma\frac{(観測値-期待値)^2}{期待値}$$

a_{ij} 及び b_{ij} から χ^2distance を求めるために、列

第 6 表

	s 1	s 2	s 3	b 1	b 2	ChiDist
1	0.666667	0.166667	0.1666667	0.347222	0.2777778	1.274755
2	0.666667	0.166667	0.1666667	0.347222	0.2777778	1.274755
3	0.666667	0.166667	0.1666667	0.347222	0.2777778	1.274755
4	0.166667	0.666667	0.1666667	0.347222	0.2777778	1.274755
5	0.166667	0.666667	0.1666667	0.222222	0.1777778	1.183216
6	0.166667	0.666667	0.1666667	0.222222	0.1777778	1.183216
7	0.166667	0.166667	0.6666667	0.222222	0.1777778	1.183216
8	0.166667	0.166667	0.6666667	0.222222	0.1777778	1.183216
9	0.166667	0.166667	0.6666667	0.222222	0.1777778	1.183216

第 7 表

	s 1	s 2	s 3	b 1	b 2
1	0.44444444	0.11111111	0.11111111	0.173611111	0.11111111
2	0.44444444	0.11111111	0.11111111	0.173611111	0.11111111
3	0.44444444	0.11111111	0.11111111	0.173611111	0.11111111
4	0.11111111	0.44444444	0.11111111	0.173611111	0.11111111
5	0.11111111	0.44444444	0.11111111	0.111111111	0.07111111
6	0.11111111	0.44444444	0.11111111	0.111111111	0.07111111
7	0.11111111	0.11111111	0.44444444	0.111111111	0.07111111
8	0.11111111	0.11111111	0.44444444	0.111111111	0.07111111
9	0.11111111	0.11111111	0.44444444	0.111111111	0.07111111
ChiDist	1.41421356	1.41421356	1.41421356	1.118033989	0.89442719

について、i 番目の row profile a_{ij} と、その average profile c_i を用いて、$(a_{ij}-c_j)^2/c_j$ を計算する（第 6 表）。同様に、行についても、j 番目の column profile b_{ij} と、その average profile, centroid, masses r_j を用いると、$(b_{ij}-r_j)^2/r_j$ を計算する。それらの結果が第 6、7 表である。

ここで、ChiDist は χ^2-distance を示している。

$$\chi^2-\text{distance}$$
$$\|a_i-c\|_c=\sqrt{\Sigma_j(a_{ij}-c_j)^2/c_j}$$
$$\|b_j-r\|_r=\sqrt{\Sigma_i(b_{ij}-r_i)^2/r_i}$$

次に、χ^2 統計量と Inertia（分散）Φ^2 との関係を示す。

$$\Theta^2=\frac{\chi^2}{n}=\sum_i r_i\,\|\,a_i-c\,\|_c^2$$
$$=\sum_i r\sum_j(\frac{p_{ij}}{r_i}-c_j)^2/c_j=\sum_j c_j\,\|\,b_i-r\,\|_r^2$$
$$=\sum_j c_j\sum_i(\frac{p_{ij}}{c_j}-r_i)^2/r_i$$
$$=(n_{ij}-n_{i+*}n_{+j})/n_{i+*}n_{+j}$$
$$=(p_{ij}-c_{i*}r_j)^2/(c_{i*}r_j)$$

上式を用いて、inertia を求めた結果が第 8 表である。この表の値は、MCA の分析のもとになる。元の多次元の情報をまとめたデータ行列を表現したものである。$\Phi^2=1.5$ で $\chi^2=1.5\times18(=n)=27$ となる。

第 8 表において、行と列および各要素の inertia を表示したものが、第 9 表である。この数値はグラフ表示した時の中心からの距離を表している。

次に、第 9 表のデータ行列 **S** を特異値分解 Singular Value Decomposition しよう。まず、データ行列 S を定義する。**S** を求めるには、列プロファイル a_{ij}、行プロファイル b_{ij}、比率 p_{ij} から計算できる。

第 8 表

	s 1	s 2	s 3	b 1	b 2	Inertia
1	0.074074	0.018519	0.0185185	0.03858	0.0308642	0.180556
2	0.074074	0.018519	0.0185185	0.03858	0.0308642	0.180556
3	0.074074	0.018519	0.0185185	0.03858	0.0308642	0.180556
4	0.018519	0.074074	0.0185185	0.03858	0.0308642	0.180556
5	0.018519	0.074074	0.0185185	0.024691	0.0197531	0.155556
6	0.018519	0.074074	0.0185185	0.024691	0.0197531	0.155556
7	0.018519	0.018519	0.0740741	0.024691	0.0197531	0.155556
8	0.018519	0.018519	0.0740741	0.024691	0.0197531	0.155556
9	0.018519	0.018519	0.0740741	0.024691	0.0197531	0.155556
Inertia	0.333333	0.333333	0.3333333	0.277778	0.2222222	1.5

第 9 表　Inertia

Inertia	s 1	s 2	s 3	b 1	b 2	row inertia
1	49	12	12	26	21	120
2	49	12	12	26	21	120
3	49	12	12	26	21	120
4	12	49	12	26	21	120
5	12	49	12	16	13	104
6	12	49	12	16	13	104
7	12	12	49	16	13	104
8	12	12	49	16	13	104
9	12	12	49	16	13	104
col inertia	222	222	222	185	148	1000

第 10 表　データ行列 S

	s 1	s 2	s 3	b 1	b 2
1	0.272166	0.170103	–0.13608	–0.15713	–0.17568
2	0.272166	0.170103	–0.13608	–0.15713	–0.17568
3	0.272166	0.170103	–0.13608	–0.15713	–0.17568
4	–0.13608	0.170103	0.272166	–0.15713	–0.17568
5	–0.13608	–0.13608	0.272166	0.125708	–0.17568
6	–0.13608	–0.13608	0.272166	0.125708	–0.17568
7	–0.13608	–0.13608	–0.13608	0.125708	0.351364
8	–0.13608	–0.13608	–0.13608	0.125708	0.351364
9	–0.13608	–0.13608	–0.13608	0.125708	0.351364

1）$s_{ij} = \sqrt{r_i}(a_{ij} - c_j)/\sqrt{c_j}$,

2）$s_{ij} = \sqrt{c_j}(b_{ij} - r_i)\sqrt{r_i}$

3）$s_{ij} = (p_{ij} - c_j r_i)/\sqrt{c_j r_i}$

すなわち、1)、2)、3）で定義するが、いずれも、同じ **S** となる。

データ行列 **S** と第 8 表の Inertia とは同じ構造を持っているので、分析結果は同じである。**S** は行列表示すれば、$\mathbf{S} = \mathbf{D_r}^{-1/2}(\mathbf{P} - \mathbf{rc}^{\mathrm{T}})\mathbf{D_c}^{-1/2}$ となる。$\mathbf{D_r}$ は **r** の対角行列で、$\mathbf{D_c}$ は **c** の対角行列である。

c. 特異値分解

データ行列 **S** を低次元に縮約するには、特異値分解 Singular value decomposition を用いる。いま、R でデータ行列 **S** の SVD を実行すると、

この表から、特異値 **d**（singular values）、と左特異行列 **U**（left singular matrix）、右特異行列 **V**（right singular matrix）が得られる。**S** は 9×5 の

第 11 表　SVD の計算結果

＞	svd(S)		
$d			
[1]	9.58 E−01	7.07 E−01	2.86 E−01
$u	[,1]dim 1	[,2]dim 2	[,3]dim 3
[1,]	−0.42686	−0.1543033	0.127296
[2,]	−0.42686	−0.1543033	0.127296
[3,]	−0.42686	−0.1543033	0.127296
[4,]	−0.14797	0.46291	−0.8079
[5,]	0.202041	0.46291	0.365763
[6,]	0.202041	0.46291	0.365763
[7,]	0.341487	−0.3086067	−0.10184
[8,]	0.341487	−0.3086067	−0.10184
[9,]	0.341487	−0.3086067	−0.10184
$v	[,1]dim 1	[,2]dim 2	[,3]dim 3
[1,]s 1	−0.54554	−2.67 E−01	0.545545
[2,]s 2	0.109109	8.02 E−01	−0.10911
[3,]s 3	0.436436	−5.35 E−01	−0.43644
[4,]b 1	−0.52705	−1.22 E−16	−0.52705
[5,]b 2	0.471405	1.18 E−16	0.471405

行列であり、SVD 分解するのに、その rank は 3 である。そこで結果の表示は 3 列までにした。行列表示をすると

$$\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{D}\mathbf{c}^{-1/2} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^T$$

$$\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$$

D_α は **d** の対角行列である。

なる関係が得られる。この式を展開すると、

$$\mathbf{s}_{ij} = \Sigma_k \alpha\, \mathbf{u}_{ik}\mathbf{v}_{jk}$$

ここで、$\mathbf{u}_{si}/\sqrt{r_i}$を行列表示した、$\mathbf{D}_r^{-1/2}\mathbf{u} = \mathbf{a}_s$ を行の標準座標 standard coordinates といい、$\mathbf{v}_{sj}/\sqrt{c_j}$、即ち、$\mathbf{D}_c^{-1/2}\mathbf{v} = \mathbf{b}_s$ は列の standard coordinates という。standard coordinate $\mathbf{a}_s$, $\mathbf{b}_s$ は **r**, **c** を weight とする値で、平均 0、分散 1 となる。また、

$$\mathbf{u}^T\mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{D}_c^{-1/2}\mathbf{v} = \alpha$$

となることから、

$$\mathbf{a}_s = \mathbf{D}_r^{-1/2}\mathbf{u} \qquad \mathbf{b}_s = \mathbf{D}_c^{-1/2}\mathbf{v}$$

であるので、$\mathbf{a}_s(\mathbf{P} - \mathbf{r}\mathbf{c}^T)\mathbf{b}_s = \alpha_s$ であることは、この式が共分散を示している。

ここで、Inertia に関する式、inertia $= \Sigma\alpha_s = \Sigma\lambda_s$、即ち、固有値の和が inertia であるので、標準座標を各次元の固有値をもちいて尺度変換して、**F**

第 12 表　座標

a）列座標

col standard coordinate				col principle coordinate			
	Dim 1	Dim 2	Dim 3		Dim 1	Dim 2	Dim 3
s 1	−1.336	−6.55 E−01	1.336306	s 1	−1.28058	−4.63 E−01	0.381889
s 2	0.2673	1.96 E＋00	−0.26726	s 2	0.256115	1.39 E＋00	−0.07638
s 3	1.069	−1.31 E＋00	−1.06905	s 3	1.024461	−9.26 E−01	−0.30551
b 1	−1.118	−2.59 E−16	−1.11803	b 1	−1.07141	−1.83 E−16	−0.31951
b 2	0.8944	2.24 E−16	0.894427	b 2	0.857125	1.58 E−16	0.255609

b）行座標

row standard coordinate				row principle coordinate			
	Dim 1	Dim 2	Dim 3		Dim 1	Dim 2	Dim 3
1	−1.281	−0.46291	0.381889	1	−1.22717	−0.32733	0.109136
2	−1.281	−0.46291	0.381889	2	−1.22717	−0.32733	0.109136
3	−1.281	−0.46291	0.381889	3	−1.22717	−0.32733	0.109136
4	−0.444	1.3887301	−2.42371	4	−0.42539	0.981981	−0.69265
5	0.6061	1.3887301	1.09729	5	0.580844	0.981981	0.313583
6	0.6061	1.3887301	1.09729	6	0.580844	0.981981	0.313583
7	1.0245	−0.92582	−0.30551	7	0.981736	−0.65465	−0.08731
8	1.0245	−0.92582	−0.30551	8	0.981736	−0.65465	−0.08731
9	1.0245	−0.92582	−0.30551	9	0.981736	−0.65465	−0.08731

$=\mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\alpha$ 及び、$\mathbf{G}=\mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\alpha$ をそれぞれの主成分座標 principal coordinates という。

d. グラフ表示

標準座標も主成分座標も同じ座標軸上で、スケールが異なる。principle coordinate による座標が一般に用いられている。直交する方向への射影された分散を最大にする座標軸個数の選択は、分散比、すなわち、inertia の比によって決まるが、通常は 2 次元のグラフで表現されるのが、普通である。

Principal inertia（eigenvalue）

	1	2	3	4	total（inertia）
Value	0.91833	0.5	0.08167	0	1.5
Percentage	61.22%	33.33%	5.44%	0%	100.00%

固有値＝特異値^2 の関係がある。$0.91833 = 0.958^2$, また、固有値の合計と inertia の合計も等しい。共に、1, 5 である。第 12 表の principal coordinates について、変数と個体について、1 次元と 2 次元の座標のグラフ表示したものが、第 1 図である。

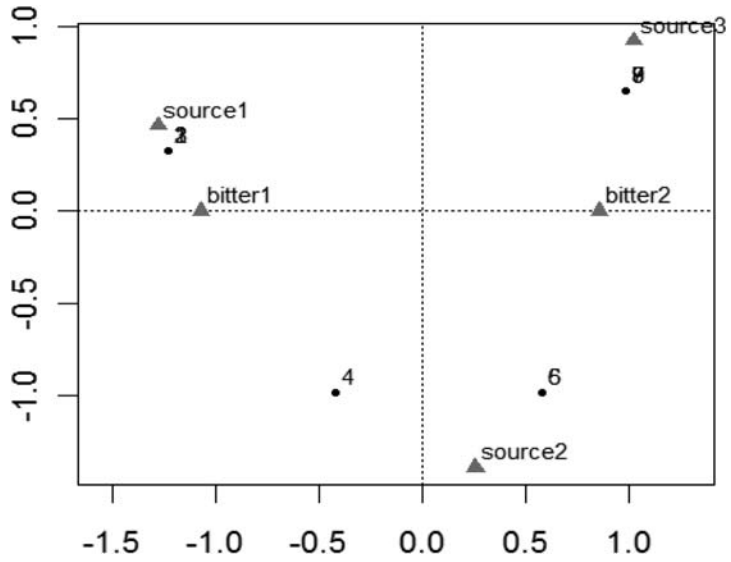
グラフで、行（個体）、列（変数）の分布の状態から、1 から 9 まで individual が変数 s、b を選択した結果がその配置に表れている。このグラフはデータを 2 次元に縮約した結果を現わしている。全体として U 字型を示し、Guttman effect などと呼ばれているが、これは、第 2 表のデータを次表のように組み替えると、表の対角線を中心とした線形の配置がグラフになっていることが理解できる。

	s 1	b 1	s 2	b 2	s 3
1	1	1	0	0	0
2	1	1	0	0	0
3	1	1	0	0	0
4	0	1	1	0	0
5	0	0	1	1	0
6	0	0	1	1	0
7	0	0	0	1	1
8	0	0	0	1	1
9	0	0	0	1	1

e. 若干の理論的考察

ここで、多次元データの縮約を骨子とする理論について、すこし考察を加えてみる。主成分、対応分析を主体とするデータは、n×p の行列から展開する。変数は、量的と質的とに分類できるが、処理から見るとこの区別は便宜的であり、カテゴリーに分類されるデータの性質によって、分析方法は主成分分析（数値）になり、対応分析（頻度）になる。

第 1 図　Principal coordinate のグラフ



特異値分解 Singular value decomposition について、データ行列（n×p）の SVD は、

$$\mathbf{S}=\mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}} \qquad \mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I}$$

は、

$$\mathbf{S}^{\mathrm{T}}\mathbf{S}=(\mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}})^{\mathrm{T}}\mathbf{U}\mathbf{D}_\alpha\mathbf{V}=\mathbf{V}\mathbf{D}_\alpha\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}}=\mathbf{V}\mathbf{D}_\alpha^2\mathbf{V}^{\mathrm{T}}$$

$$\mathbf{S}\mathbf{S}^{\mathrm{T}}=\mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}}(\mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}})^{\mathrm{T}}=\mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}}\mathbf{V}\mathbf{D}_\alpha\mathbf{U}^{\mathrm{T}}$$
$$=\mathbf{U}\mathbf{D}_\alpha^2\mathbf{U}^{\mathrm{T}}$$

上の 2 式は、固有値分解を示しているので、その固有根 Λ は、$D_\alpha^2=\Lambda$ なる関係がある。ここで、$\mathbf{D}_\alpha$ の対角要素は特異値 singular value で非負の値である。その大きさ **r** は、min（n, p）で、**S** のランクという。

行列 **S** の SVD を展開すると、

$$\mathbf{S}=\Sigma\alpha_i\mathbf{u}_i\mathbf{v}_i^{\mathrm{T}} \qquad i=1, 2, \ldots, r$$

更に、この展開式を、k＜r に留めると、

$$\mathbf{S}^{\divideontimes}=\Sigma\alpha_i\mathbf{u}_i\mathbf{v}_i^{\mathrm{T}} \qquad i=1, 2, \ldots, k$$

となる。$\mathbf{S}^{\divideontimes}$は **S** の rank k＜r での最小 2 乗に意味での最良の近似と理解できる、この点から理論を展開したものが、Gifi system と呼ばれている。$\mathbf{S}=\mathbf{S}^{\divideontimes}+\varepsilon$ となるので誤差項を最小にすることは、**S** を最大にすることになる。これは **S** の分散を最大にするようなパラメータを定めることを意味する。これを別の観点から云えば、最適な尺度を決めることであり、幾何学的にいえば、互いに直交する主軸の上に推定値を得ることになる。

3. MCA の計算方法

ここでは、R を用いて、MCA の計算を taste のデータを使って説明する。各表の導出には Excel を用いたが、ここでは、Excel での計算とともに、R の Program を提示することにする。

1．まず、第 1 表のデータを Excel 上につくり、R に読み込む。

```
taste<− read.table(“clipboard”, header＝TRUE)
```

2．このデータ taste を、indicator variable を利用して、第 2 表をつくる。

```
taste<− make.dummy(taste)　　make.dummy() の program は別途参照。
```

3．第 3 表は、第 2 表を比率に変換したものである。$p_{ij}=n_{ij}/n$

これを profile table という。

```
Z<− taste　#taste を Z とする。
P<− Z/sum(Z)　#P は比率の表である。
cm<− apply(P, 2, sum)
　　#cm は列の周辺度数 masses を表す。
rm<− apply(P, 1, sum)
　　#rm は行の masses を表す。
```

4．第 4, 5 表は行 indiviuals と、列 variables についての profile a_i, b_j を求めたものである。

その要素は、$a_{ij}=n_{ij}/n_i$, 及び $b_{ij}=n_{ij}/n_{+j}$ である。

```
cs<− apply(Z, 2, sum)　#cs は列の和を表す。
rs<− apply(Z, 1, sum)　#rs は行の和を表す。
A<− sweep(Z, 1, rs,”/”)
　　#A は第 4 表を表す。
B<− sweep(Z, 2, cs,”/”)
　　#B は第 5 表を表す。
```

5．第 4 表から、第 6 表、第 5 表から第 7 表を計算し、各表から、行と列の χ^2distance を求める。

```
AP<− sweep(A, 2, cm,”− “)^2
APP<− sweep(AP, 2, cm,”/”)
　　#APP は第 6 表の計算結果を表す。
ChiDistA<− apply(APP, 1, sum)^(1/2)
　　#変数の χ²distance
cbind(APP, ChiDistA)　　#第 6 表を表す。
BP<− sweep(B, 1. rm,”− “)^2
BPP<− sweep(BP, 1, rm,”/”)
　　#BPP は第 7 表の計算結果を表す。
ChiDistB<− apply(BPP, 2, sum)^(1/2)
　　#個体の χ²distance
rbind(BPP, ChiDistB)　　#第 7 表を表す。
```

6．次に、Inertia（分散）を求める。Inertia を得るには、第 6, 7 表から求めても、直接求めても同じ結果を得る。式で表現すると、

```
InertiaP<− diag(rm)　%*%　APP
InertiaP<− BPP　%*%　diag(cm)
rc<− rm%*%t(cm)
InertiaP<−(P−rc)^2/rc
```

第 8 表の周辺和は、行および列の inertia を得る。

```
InP<− addmargins(InertiaP)
round(InP/sum(InertiaP)*1000,0)
　　#第 9 表の出力。
```

7．SVD を計算するデータ行列を S とすると、

S を計算するには、行 row profile、列 column profile から計算する方法と、行と列を同時に計算する方法があるが、いずれも同じ S が得られる。

```
S<−(P−rc)/sqrt(rc) #S の算出、第 9 表。
dec<−svd(S)　　#svd の出力、第 10 表。
```

8．標準座標 standard coordinate と、主成分座標 principal coordinate を求める。

```
lam<−dec$d[1 : 3] ^ 2　　#固有値を求める。
b.s 1<−dec$v[,1]/sqrt(cm)
#col standard coordinate を求める。
                                    (dimension 1)
b.s 2<−dec$v[,2]/sqrt(cm)
b.s 3<−dec$v[,3]/sqrt(cm)
g.s 1<−b.s 1*sqrt(lam[1])
#col principal coordinate を求める。
                                    (dimension 1)
g.s 2<−b.s 2*sqrt(lam[2])
g.s 3<−b.s 3*sqrt(lam[3]),
f.s 1<−dec$u[,1]/sqrt(rm)
#row standard coordinate を求める。
                                    (dimension 1)
f.s 2<−dec$u[,2]/sqrt(rm)
f.s 3<−dec$u[,3]/sqrt(rm)
a.s 1<−f.s 1*sqrt(lam[1])
#row principal coordinate を求める。
                                    (dimenson 1)
a.s 2<−f.s 2*sqrt(lam[2])
a.s 3<−f.s 3*sqrt(lam[3])
```

座標の値は、第 11 表にある。

9．グラフ。グラフは主座標 principal coordinate を表示したものであるが、ここでは、R の package ca を利用している。

```
ca を使うには、
library(ca)
data(taste)
res<−mjca(taste, lambda = "indicator")
plot(res)
#グラフを描く。(第 1 図)
```

註

R の package による計算

MCA を計算するに利用できる R の package について詳しい説明は、R のサイトである http : //cran.r−project.org/web/views/Multivariate.html などを参照されたい。ここで、主として利用した package としては、ca と FactoMineR である。以下では、この 2 つの package を用いて計算することにする。

ca は、CA, MCA に特化した package で使い易い、FactoMineR は探索的多変量解析を目標としているが、graphic 機能が豊富である。それぞれ、manual, package の解説を参照されたい。

なお、調査データ分析、文献は次号にて。

Exploratory Multivariate Statistical Data Analysis

—— multivariate correspondence analysis ——

ABSTRACT

Multiple variable correspondence analysis has a wide application area as exploratory multivariate statistical analysis, which can be applied to tables with individuals in the rows and categorical variables in the columns. Here I explain the outline of multiple correspondence analysis with a simple table as an example, and calculate thes procedure using Excel and R.

I then analyze the cultural data from Japan and Germany, and present several methods analyses.

Key Words: Multiple correspondence analysis（MCA）, R, FactoMineR