最近のデータ解析法

大蔵省印刷局研究所 山崎 秀彦

Recent Methods For Data Analysis

Hidehiko Yamazaki

Research Institute of Printing Bureau, Ministry of Finance

Recently, computer-oriented statistical approaches for data analysis as are called Multivariate Data Analysis and Quantification Theory are paid attention to as methods for prediction of unknown data and for abstraction of complicated phenomena in various kinds of fields like psychology, economics, chemistry, medicine, etc.

The methods can be classified into two groups as follows.

- (1) Methods to describe a given property using some other properties, which can be employed to predict unknown data.
- (2) Methods to summarize data.

In this article the application of these methods to studies on paper is described. Included are quantification theory type 1, 3 and 4, multiple regression analysis, time series analysis, principal component analysis, factor analysis, multidimensional scaling.

Keywords: Multivariate data analysis, Quantification theory, Prediction, Abstraction, Multiple regression analysis, Time series analysis, Principal component analysis, Factor analysis, Multidimensional scaling.

1. はじめに

紙パルプ分野の研究で用いられるデータ解析法には、基本統計量としての平均、標準偏差をはじめ、相関係数、一変数の回帰式、要因分析のための実験計画法などがあり、更に複雑でコンピュータを必要とするものでは重回帰分析の利用がしばしば見受けられる。後者は、複数の変数を同時に取り扱いそれらの関連性を分析する多変量解析と呼ばれるデータ解析法の一つである。多変量解析は近年のコンピュータの発達と普及により多くの分野で応用され、新たな研究分野の発達をみている。例えば、計量心理学を初めとして計量経済学、計量法学、計量化学、計量診断学などが挙げられる。使用する目的により解析手法を大きく分けると、

- (1) 複数の特性の組合せで特定の一つの特性を説明しようとするものであり、しばしば予測に用いられるものと、(2) データの要約に用いられるものに分けられる。

ここでは、これらの手法の紙の研究への応用について述べる。ただし、ここでは数量化理論も多変量解析に含めた。また、本稿では個々の手法の計算の詳細については触れない。それらについては章末の文献を参照して頂きたい。

2. 要因分析と予測

2.1 数量化理論第1類¹⁾⁻⁵⁾⁸⁾

これは統計数理研究所の林が開発した一連の数量化理論の一つである。ある量的な特性を複数の質的な特

本総説は去る昭和61年11月19日、繊維学会第21回紙パルプシンポジウムにおいて講演した内容に加筆したものである。

性（以下、要因と呼ぶ*）の組合せで説明するもので、要因のウェイトを計算し、未知データの予測を行うことができる。

*) 数量化理論では通常、これはアイテムと呼ばれる。

予測式は一般的な形で次のように表現する。

$$\begin{aligned} \text{予測量} &= Y = X_1 + X_2 + X_3 + \cdots + X_m \\ &= \sum_{j=1}^m X_j \end{aligned} \quad (1)$$

ここで、 $X_j (j=1, 2, \cdots, m)$ は要因 j の数量を意味する。

更に、個々の要因が複数のカテゴリから構成されているものとし、個々の個体（サンプル）は各要因において必ず一つ該当するカテゴリで示される特徴を有しているものとする。

要因 j の数量 X_j はカテゴリ数量 $x_{jk} (k=1, 2, \cdots, l_j)$ の一次結合により次式で表す。ただし、 l_j は要因 j に含まれるカテゴリ数である。

$$X_j = \sum_{k=1}^{l_j} \delta_{jk} x_{jk} \quad (2)$$

ここで、 δ_{jk} は個体が要因 j におけるカテゴリ k に該当すれば 1、該当しなければ 0 の値をとる。 x_{jk} は実測値と予測値の差 $Y - \hat{Y}$ の 2 乗平均を最小にするような値として計算される。

結局、次式を解く。

$$(A'A)X = A'Y \quad (3)$$

$$A = \{\delta_{i(jk)}\} \quad (i=1, 2, \cdots, n; j=1, 2, \cdots, m; k=1, 2, \cdots, l_j)$$

$$Y = \{y_i\} \quad (i=1, 2, \cdots, n)$$

ただし、 $\delta_{i(jk)}$ は個体 i が要因 j のカテゴリ k に該当するとき 1、そうでないとき 0 の値をとる。 y_i は個体 i の特性値である。実際の計算においては行列 $(A'A)$ にランク落ちが生ずるので、式(3)から要因 $j=2, \cdots, m$ の第 1 カテゴリを除外した式を解く。

また、回帰式の精度は実測値と計算値の相関係数に相当する重相関係数で表す。各要因の特性への寄与の程度は偏相関係数や要因中のカテゴリ数量の最大値と最小値の差に相当するレンジで表す。

偏相関係数の意味は次のごとくである。今、仮に図 1 に示すようにある説明したい特性 Y 、その要因 X_1 及び X_2 があるとする（いずれもベクトルで表されている）。一般に Y と X_1 の間の関係を知りたい場合、よく Y と X_1 の相関係数の値の大小で判定されること

図 1 相関係数と偏相関係数の関係

が多い。しかし、この場合は第 2 の要因 X_2 の影響が加わっているため、真に Y と X_1 の相関を知りたいならば X_2 の影響を除外した Y' と X_1' の間の相関係数を調べるのが本当であるといえる。これを偏相関係数と呼ぶ。ここで、 Y' と X_1' はそれぞれ Y と X_1 を X_2 に垂直な平面 S 上に投影したベクトルである。以上のことは要因の数が更に多くても同じ論法が当てはまる。

なお、カテゴリ数量 x_{jk} の値は実用的には特性平均からのずれを示すように基準化したものを用いるのが普通である（次の応用例を参照）。

1) 応用例 1

表 1 は *Pinus radiata* の BSP について漂白、こう解及びアンモニア処理を行った Allison⁶⁾ の実験からデータのみ引用したものである。この場合、物性を支配する要因は漂白、こう解、アンモニア処理条件の 3 つであり、漂白には ZEP（オゾン→カセイソーダ抽出→過酸化処理）、CEH（塩素→カセイソーダ抽出→次亜塩素酸ソーダ処理）及び UT（未漂白）の 3 カテゴリが、こう解には PFI ミル 2000、4000 及び 8000 回の 3 カテゴリが、アンモニア処理には AT1（紙料濃度 30%、760 kPa のアンモニアガス処理 45 分を 3 サイクル）、AT2（紙料濃度 15%、760 kPa のアンモニアガス処理 45 分を 1 サイクル）及び UT（未処理）の 3 カテゴリが含まれている。これらの実験条件を分かり易くカテゴリ化したのが表 2 である。

今、No. 1~17 のデータをもとに数量化分析を行い、得られたカテゴリ数量を用いて予測式を構成し、次いでそれをもとにして No. 18 に示された処理条件で作製される手すきシートの物性を予測してみる。

入力データとしては、比引裂強さを例にとれば、 δ_{jk} には表 2 の実験条件パターンを、 Y には比引裂強さの測定値をそれぞれ対応させる。これをもとに数量化分析を行い、得られたカテゴリ数量を用いて、次のような予測式を構成することができる。

表 1 漂白、ころ解、アンモニア処理後の BSP 手すきシートの物理的性質

No.	漂白処理	PFI ころ解 回転数	アンモニ ア処理	フリーネス (CSF)	密度 kg/m ³	度	比引裂強さ mN.m ² /g	比破裂強さ kPa.m ² /g	比引張強さ N.m/g	伸び率 %	ヤング率 MN/m ²	比散乱係数 cm ² /g	透気度 sec/100ml	白色度 (Eitrepheo)
1	未漂白	2000	UT	720	633		12.7	5.6	74	2.5	4600	180	3	55.4
2		2000	AT1	750	533		24.4	2.2	27	1.8	2100	242	1	49.0
3		4000	UT	650	657		11.1	6.4	84	2.8	5200	158	4	52.7
4		4000	AT2	690	629		15.7	5.1	61	3.0	3900	183	4	41.8
5		8000	UT	400	695		10.0	7.3	91	3.2	5300	141	19	52.1
6	ZEP	8000	AT2	540	653		13.7	5.9	70	3.3	4300	163	16	40.4
7		2000	UT	690	666		10.9	5.8	67	2.8	4900	170	4	77.8
8		2000	AT1	760	543		19.1	1.5	23	2.1	1900	254	1	76.1
9		4000	UT	530	687		9.0	6.9	85	3.0	5400	149	18	77.4
10		4000	AT2	620	666		11.1	5.3	65	3.1	4400	170	6	69.5
11	CEH	8000	UT	260	732		8.3	7.4	93	3.3	6000	130	500	76.2
12		8000	AT2	190	751		8.6	7.7	94	3.1	6200	126	1000	69.1
13		2000	UT	700	676		11.1	6.1	73	3.1	4700	166	3	81.0
14		2000	AT1	760	548		22.4	1.8	24	2.3	1900	262	1	76.9
15		4000	UT	630	698		9.9	7.0	81	3.2	5000	148	8	80.2
16		4000	AT2	680	668		13.2	5.5	61	3.4	3800	173	4	72.1
17		8000	UT	420	725		8.9	7.6	88	3.3	5600	132	100	79.4
18		8000	AT2	500	692		11.3	6.4	75	3.6	4700	151	29	71.4

備考 UT = 未処理

AT1 = 紙料濃度 30%, 760kPa のアンモニアガス処理 45分を 3 サイクル

AT2 = 紙料濃度 15%, 760kPa のアンモニアガス処理 45分を 1 サイクル

ZEP = オゾノン→カセイソーダ抽出→過酸化水素処理

CEH = 塩素→カセイソーダ抽出→次亜塩素酸ソーダ処理

表 2 カテゴリー化した実験条件

No.	漂白処理			PFIミル(回)			アンモニア処理		
	UT	ZEP	CEH	2000	4000	8000	UT	AT ₁	AT ₂
1	1			1			1		
2	1			1				1	
3	1				1		1		
4	1				1				1
5	1					1	1		
6	1					1			1
7		1		1			1		
8		1		1				1	
9		1			1		1		
10		1			1				1
11		1				1	1		
12		1				1			1
13			1	1			1		
14			1	1				1	
15			1		1		1		
16			1		1				1
17			1			1	1		
18			1			1			1

$$\begin{aligned}
\text{Tear} = & 1.73\delta_{\text{BL}}(\text{UT}) - 1.70\delta_{\text{BL}}(\text{ZEP}) \\
& - 0.03\delta_{\text{BL}}(\text{CEH}) + 1.27\delta_{\text{PFI}}(2000) \\
& - 0.01\delta_{\text{PFI}}(4000) - 1.51\delta_{\text{PFI}}(8000) \\
& - 2.65\delta_{\text{AM}}(\text{UT}) + 7.75\delta_{\text{AM}}(\text{AT}_1) \\
& + 0.12\delta_{\text{AM}}(\text{AT}_2) + 12.95 \text{ mN} \cdot \text{m}^2/\text{g} \quad (4)
\end{aligned}$$

ただし、 δ_{jk} はここでは実験条件を反映させ以下のように表示した。

- $\delta_{\text{BL}}(\quad)$ = 該当する漂白条件に対して、1
 該当しない漂白条件に対して、0
 $\delta_{\text{PFI}}(\quad)$ = 該当するこう解回転数に対して、1
 該当しないこう解回転数に対して、0
 $\delta_{\text{AM}}(\quad)$ = 該当するアンモニア処理条件に対して、1
 該当しないアンモニア処理条件に対して、0

とする。

重相関係数は0.987であり、データによくフィットしていることを示している。偏相関係数は漂白処理条件が0.890、こう解処理条件が0.805、アンモニア処理条件が0.997であり、アンモニア処理条件の寄与がとりわけ大であることが分かる(表3)。更に、個々の要因の中での個々のカテゴリの寄与もカテゴリ数量を用いて明瞭に示すことができる。図2はその例であり漂白、こう解及びアンモニア処理の比引裂強さへの影

図 2 漂白、こう解及びアンモニア処理の比引裂強さへの影響

響が平均値との差で示されている。漂白処理については、比引裂強さは未漂白に比べて ZEP では $3.4 \text{ mN} \cdot \text{m}^2/\text{g}$ 、CEH では $1.8 \text{ mN} \cdot \text{m}^2/\text{g}$ 低下することが分かる。こう解処理については、比引裂強さは PFI こう解回転数の増加とともに低下し、2000 回に比べて 4000 回では $1.3 \text{ mN} \cdot \text{m}^2/\text{g}$ 、8000 回では $2.8 \text{ mN} \cdot \text{m}^2/\text{g}$ 低下することが分かる。アンモニア処理については、比引裂強さは未処理に比べて AT1 では $10.4 \text{ mN} \cdot \text{m}^2/\text{g}$ 、AT2 では $2.8 \text{ mN} \cdot \text{m}^2/\text{g}$ 増加することが分かる。

さて、式(4)を用いて No. 18 のシートの比引裂強さを予測するには No. 18 の実験条件、すなわち漂白 CEH、PFIミル 8000 回、アンモニア処理 AT2 に該当する δ_{jk} に 1、その他に 0 を代入することにより次のように計算できる。

$$\begin{aligned}
\text{Tear} = & 1.73 \times 0 - 1.70 \times 0 - 0.03 \times 1 + 1.27 \times 0 \\
& - 0.01 \times 0 - 1.51 \times 1 - 2.65 \times 0 + 7.75 \times 0 \\
& + 0.12 \times 1 + 12.95 = 11.5 \text{ mN} \cdot \text{m}^2/\text{g}
\end{aligned}$$

他の物性についての要因分析の結果を表3に示す。比散乱係数と透気度を除いて、重相関係数は0.95以上と高く、良好な回帰式が得られていることが分かる。また、偏相関係数の寄与の目安を0.85とすれば、フリーネスにはこう解条件が、密度、比破裂強さ、比引張強さ及び白色度にはこう解及びアンモニア処理条件が、比引裂強さには漂白及びアンモニア処理条件が、伸び率及びヤング率にはアンモニア処理条件がそれぞれ大きく影響していることが分かる。表4は No. 18 の他の物性値を予測したものである。フリーネス、比散乱係数、透気度以外は予測誤差5%以内の良好な予

表 3 物理的性質に影響を及ぼす要因の分析結果

フリーネス (CSF)	密度 kg/m ³	比引裂強さ mN・m ² /g	比破裂強さ kPa・m ² /g	比引強強さ N・m/g	伸び率 %	ヤング率 MN/m ²	比散乱係数 cm ² /g	透気度 sec/100ml	白 色 度 (Elrepho)
重相関係数	0.952	0.987	0.981	0.977	0.977	0.970	0.732	0.744	0.997
相 関 係 数									
漂白処理	0.334	0.308	0.067	0.099	0.259	0.211	0.373	0.442	0.099
こう解処理	0.894	0.590	0.689	0.687	0.740	0.622	0.077	0.569	0.687
アンモニア処理	0.275	0.913	0.918	0.904	0.871	0.894	0.693	0.273	0.904
偏相関係数									
漂白処理	0.705	0.890	0.420	0.339	0.842	0.632	0.465	0.532	0.339
こう解処理	0.941	0.805	0.872	0.855	0.828	0.786	0.179	0.628	0.855
アンモニア処理	0.402	0.977	0.963	0.955	0.929	0.948	0.672	0.316	0.955

表 4 物理的性質の予測結果

	フリーネス (CSF)	密 度 kg/m ²	比引裂強さ mN・m ² /g	比破裂強さ kPa・m ² /g	裂 断 長 N・m/g	伸び率 %	ヤング率 MN/m ²	比散乱係数 cm ² /g	透 気 度 sec/100ml	白 色 度 (Elrepho)
予 測 値	413	714	11.5	6.7	76	3.5	4796	203	36	70.1
実 測 値	500	692	11.3	6.4	75	3.6	4700	151	29	71.4
誤 差 率(%)	-17.3	3.2	2.0	4.5	1.0	1.3	2.0	34.6	23.9	- 1.9

測結果を得ている。予測誤差の大きな物性については適切な数値変換、例えば対数変換などをした後、数量化分析するのも一策である。

図3は比引裂強さについて No. 1~18 のデータを一つずつ順に除外し、残る17のデータより除外したサンプルの比引裂強さを予測した結果である。前に示した No. 18 以外のサンプルについても良好な予測が行われていることが分かる。実測値と予測値の相関係数は0.965である。このように試料を順に一つずつ除外し予測式のチェックをする方法を Leave-one-out 法とか Jack-knife 法といい、後出の重回帰分析によく用いられる。

図3 除外したサンプルの比引裂強さの予測

2) 応用例2

コート紙を試作する目的で3種類のラテックスを選び、バインダー比率を15、20、25%の3水準、並びにカゼイン／ラテックス比を1/1、1/2、1/3の3水準としてテストコーターでシングルコートし、スーパーカレンダー仕上げをして27種類の用紙を得た。これら用紙について紙質特性及びウェットピックス強度を評価した。ウェットピックス強度の評価には IPI インキを用いて FOGRA 印刷適性試験機に

よる紙むけ試験を印刷速度1m/s、印刷圧力20kg/cmで行い、4段階の官能評価をした。ただし、湿し水にはイソプロピルアルコール20%水溶液を用い、印刷ディスク上のインキ皮膜厚さ9μm、加湿量20mm³/40cm²、加湿ロールの円周速度0.5m/s、加湿ロール上での水の分布時間10s、加湿から印刷開始までの時間間隔1sとした。

表5は数量化理論を用いて要因分析をした結果である。重相関係数の値はいずれも0.9以上と高く精度の良い回帰式が得られていることが分かる。白色度にはバインダー比率が、75°光沢にはラテックスの種類、バインダー比率、カゼイン／ラテックス比すべてが、吸油度にはバインダー比率が、ウェットピックス (IPI No. 4 インキ) にはバインダー比率が大きく寄与していることが分かる。

予測式の一例として75°光沢を求める式を示す。

$$\begin{aligned} \text{Gloss} = & 0.47\delta\text{Latex}(A) + 3.04\delta\text{Latex}(B) \\ & - 3.51\delta\text{Latex}(C) + 3.37\delta\text{Binder}(15) \\ & - 0.65\delta\text{Binder}(20) - 2.72\delta\text{Binder}(25) \\ & - 2.94\delta C/L(1/1) + 1.06\delta C/L(1/2) \\ & + 1.88\delta C/L(1/3) + 23.36\% \end{aligned} \quad (5)$$

ただし、 $\delta\text{Latex}()$ = 該当するラテックスの種類に対して、1

該当しないラテックスの種類に対して、0

$\delta\text{Binder}()$ = 該当するバインダー比率に対して、1

該当しないバインダー比率に対して、0

$\delta C/L()$ = 該当するカゼイン／ラテックス比に対して、1

該当しないカゼイン／ラテックス比に対して、0

こうして、式(5)を用いることにより設定された実験条件の任意の組合せから作製されるコート紙の75°光沢が容易に算出できる。

表5 物理的性質に影響を及ぼす要因の分析結果

	白色度	75°光沢	吸油度	耐ウェットピックス
重相関係数	0.962	0.976	0.909	0.923
偏相関係数				
ラテックスの種類	0.277	0.942	0.504	0.755
バインダー比率	0.947	0.935	0.888	0.958
カゼイン／ラテックス比	0.833	0.910	0.632	0.186

2.2 重回帰分析⁷⁾⁸⁾

要因（説明変量）が数量データで与えられ且つ予測したい特性（目的変量）もまた数量で与えられる場合の要因分析と予測には重回帰分析が用いられる。

説明変量が一つの場合の回帰式は、一般によく知られているとおりで、目的変量と説明変量との間で n 個のサンプルについて図4におけるようなデータの散らばりがある場合に、データから目的変量の軸と並行に線分を引き、その平方和 $e_1^2 + e_2^2 + \dots + e_n^2$ を最小にするような一次式のことである。

図4 y の x に対する回帰直線

説明変量が複数（ m 個）ある場合、すなわち多変量の場合も基本的には同様な考え方をすればよく、そのとき回帰式は、

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (6)$$

のように説明変量の1次結合の形式で与えられる。個々の変量に掛る係数（偏回帰係数） b_j ($j=0, 1, \dots, m$) は予測値と実測値の差 $\hat{Y} - Y$ の2乗平均を最小にする条件から決定する。

実際には次式を解く。

$$\mathbf{x}'\mathbf{x}\mathbf{b} = \mathbf{x}'\mathbf{y} \quad (7)$$

ただし、 $\mathbf{x}$ = 第1列の要素がすべて1、第2～ $m+1$ 列に n 個のサンプル、 m 個の説明変量に対するデータ行列を有する $n \times (m+1)$ の行列。

$$\mathbf{b} = \{b_j\} \quad (j=0, 1, \dots, m)$$

$$\mathbf{y} = \{y_i\} \quad (i=1, 2, \dots, n)$$

また、偏回帰係数 b_j ($j=0, 1, \dots, m$) の有意性は t 値の大きさと自由度 $n-m-1$ 、危険率 α の t 分布表による $t(n-m-1, \alpha)$ の値との比較により次のように判定する。

$t \geq t(n-m-1, \alpha)$ ならば、その係数は有意。

$t < t(n-m-1, \alpha)$ ならば、その係数は有意とはい

えない。

ここで、 n = サンプル数

m = 変量数

α = 危険率。1 または 5 %

$t = b_j / (b_j \text{ の標準偏差})$

更に、数量化理論第1類のときと同様、回帰式の精度を意味する重相関係数及び目的変量に対する個々の説明変量の寄与を意味する偏相関係数も求める。

1) 応用例1

Clark¹⁰⁾¹¹⁾ は紙の引張強さ T （又は裂断長）を五つの基本物性、すなわちゼロスパン裂断長 Z 、粘着性 H 、長さで重み付けした繊維長 L 、湿潤圧縮率 V 、粗度 G のべき乗の積で表わすことを提唱した。その式の一般形は次のごとくである。

$$T = kZ^a H^b L^c G^d V^e \quad (8)$$

Clark は実データをもとにして k, a, b, c, d, e を決定するのに、裂断長の対数と基本物性の一つの対数をプロットし、得た直線の勾配から算出する方法を用いた。しかし、この方法には基本物性が相互に独立であるという前提がある。実際には、むしろ基本物性相互間に多少なりとも相関があると考えられるため、個々の定数を個別にではなく、データに最も良く適合するように同時決定するのが合理的と考えられる。

この目的には重回帰分析の利用が便利である。式(8)の対数をとると次の式が得られる。

$$\begin{aligned} \log T = & \log k + a \log Z + b \log H + c \log L \\ & + d \log G + e \log V \end{aligned} \quad (9)$$

これは次のように書くことができる。

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 \quad (10)$$

ここで、 $Y = \log T$

$x_i = i$ 番目の基本物性の対数

($i=1, 2, 3, 4, 5$)

$b_0 = \log k$

$b_i = x_i$ の係数 ($i=1, 2, 3, 4, 5$)。それぞれ a, b, c, d, e に相当。

Y, x_1, x_2, x_3, x_4 及び x_5 を実測より求め、重回帰分析により、 b_0, b_1, b_2, b_3, b_4 及び b_5 を決定することにする。

表6¹¹⁾ は針葉樹の亜硫酸パルプとクラフトパルプをさまざまに処理したときの四つの基本物性とシート強度を示す。粗度はこの場合一定と仮定されているため、 G^d は定数 k に含められ省略できる。裂断長を説明するため、これらデータに Clark 式を当てはめ重回帰分析を行った。

表6 紙質特性

処	理	回/g	粘着性 kg/37.5mm ²	長さ平均 繊維長 mm	かさ cm ³ /g	ゼロスパン 裂断長 km	裂断長 km	比破裂強さ	比引強さ	
亜硫酸パルプA										
なし		0	0.94	2.67	1.79	(16)	1.74	9	126	
		8	4.5	1.90	1.38	16	5.70	22	115	
		16	(5)	1.69	(1.32)	(16)	5.82	31	95	
		32	6.5	1.39	1.25	16	6.36	32	75	
64		8.2	1.04	1.17	16	6.26	28	59		
こう解後, 23℃で2N HClに½時間浸せき		0	1.2	(2.67)	(1.80)	(15)	1.60	8	130	
32		6.5	(1.39)	1.29	15	6.26	31	85		
64		7.4	(1.04)	1.21	15	6.20	30	70		
こう解後, 50℃で2N HClに7½時間浸せき		0	1.1	(2.67)	1.33	(10)	1.65	6	90	
8		4.6	(1.90)	1.41	10	4.65	24	65		
16		(5)	(1.69)	1.34	(10)	5.20	24	60		
こう解後, 50℃で2N HClに15時間浸せき		0	1.3	(2.67)	1.77	(8.8)	1.65	7	60	
8		4.1	(1.90)	1.40	8.8	3.79	15	(60)		
16		4.4	(1.69)	1.32	(8.8)	4.78	21	60		
32		7.0	(1.39)	1.27	8.8	5.10	20	40		
64		8.0	(1.04)	1.19	8.8	5.24	18	35		
こう解後, 50℃で2N HClに40時間浸せき		0	1.1	(2.67)	1.69	(5.8)	1.35	5	35	
32		(6.0)	(1.39)	1.24	5.8	3.90	10	(21)		
64		7.2	(1.04)	1.12	5.8	4.31	14	23		
こう解後, 50℃で2N HClに2時間浸せき		8	4.2	(1.0)	1.41	6.0	2.1	6	...	
16	4.5	(0.7)	1.21	(6.0)	2.1	6	...			
8	5.3	(1.9)	1.41	16	6.0	32	105			
8	4.3	(2.0)	1.50	(16)	5.6	...	...			
なし なし-200メッシュ通過分除去										
なし なし-200メッシュ通過分除去										
なし なし-200メッシュ通過分除去										
なし-1.13kg/cm ² ウエットプレス										
なし-100メッシュ 通過分除去										
なし-46.5kg/cm ² ウエットプレス										
なし-100メッシュ 通過分除去										
標準 (4.2kg/cm ²)										
標準, 200メッシュ通過分除去										

Dis は少し離解

備考 () 内は外挿, 内挿又は他の実験による

図5は実測値と式による計算値を比較した結果である。このとき得た式は

$T = 1.5Z^{0.40} H^{0.81} L^{0.38} V^{0.028}$ である。また、図6は Clark が同じデータから求めた関係 $T = kZ^{0.5} H^{0.5} L^{0.5} V^{-1}$ を用いた場合の同様な関係*2)を示したものである。このとき得た式は

$T = 0.184 + 0.616Z^{0.5} H^{0.5} L^{0.5} V^{-1}$ である。それぞれの場合の式の定数の有意性を検討した結果を表7及び8に示す。定数の同時決定で得られた式の精度 ($r = 0.944$) が、個別に定数を決定して得られた式の精度 ($r = 0.909$) より良いことが分かる。後者の定数項 a

($=0.184$) は、危険率5%でも有意ではなく0と開けることが分かる。

*2) 文献 11) には Clark による図が掲載されているが、ここでは筆者が

$$T = a + kZ^{0.5} H^{0.5} L^{0.5} V^{-1} \quad (11)$$

の関係式に最小2乗法を適用し a 及び k を求めた。

表7 重回帰分析による裂断長の Clark 式 $T = kZ^a H^b L^c V^d$ ($\log T = \log k + a \log Z + b \log H + c \log L + d \log V$) の決定

変量	係数	t 値	有意性	偏相関係数
$\log k$	0.397	5.07	*	
$\log Z$	0.400	4.58	*	0.661
$\log H$	0.810	11.53	*	0.912
$\log L$	0.380	3.75	*	0.585
$\log V$	0.028	0.13	—	0.026

重相関係数 0.944

備考 * 危険率1%有意
— 危険率5%で有意性なし

表8 重回帰分析による裂断長の Clark 式 $T = a + kH^{0.5}Z^{0.5}L^{0.5}V^{-1}$ の決定

変 量	係数	t 値	有意性	偏相関係数
a	0.184	0.48	—	
$H^{0.5}Z^{0.5}L^{0.5}V^{-1}$	0.616	11.96	*	0.909

重相関係数 0.909

備考 * 危険率1%有意
— 危険率5%で有意性なし

2) 応用例2

Clark¹¹⁾ は引張強さについての Page 理論に修正を加え、次の式(修正 Page 式)を導いた。

$$1/T = K/Z + K'/HLD \quad (12)$$

ここで、 T = 裂断長

Z = ゼロスパン裂断長

G = 粗度

H = 粘着性

L = 長さで重み付けした平均繊維長

D = シート密度

ここでも、定数 K 及び K' の決定に重回帰分析を応

図5 Clark 式 (式(8)~(10)) による裂断長の計算値と実測値

図6 Clark 式 (式(11)) による裂断長の計算値と実測値の関係

用することができる。修正 Page 式では裂断長の逆数 (Y) がゼロスパン裂断長の逆数項, すなわち繊維強度に関する項 (x_1) と粗度, 粘着性, 長さで重み付けした平均繊維長及びシート密度からなる項, すなわち繊維間結合に関する項 (x_2) の一次結合の形で表現される。簡略化すると次の形になり, 重回帰分析の利用が容易となる。

$$Y = b_0 + b_1 x_1 + b_2 x_2 \quad (13)$$

ここで, b_0 =定数。修正 Page 式が理想的にデータにフィットする場合は 0 となる。

b_1 =繊維強度に関する項の係数, K

b_2 =繊維間結合に関する項の係数, K'

表 6 のデータに修正 Page 式を当てはめ, 重回帰分析を行った。ただし, 粗度 G が一定と仮定されているため, 式(12)中で G は K' に含められる。また, $D=1/V$ とした。

図 7 修正 Page 式による裂断長の計算値と実測値の関係

表 9 重回帰分析による裂断長の修正 Page 式 $1/T = a + K/Z + K'V/HL$ の決定

変量	係数	t 値	有意性	偏相関係数
a	-0.014	0.36	—	
1/Z	1.458	4.04	*	0.601
V/HL	0.657	9.82	*	0.877
重相関係数	0.882			

備考 * 危険率 1% 有意
— 危険率 5% で有意性なし

図 7 は実測値と式による計算値を比較した結果である。式の精度 ($r=0.882$) は Clark 式の場合よりやや低い。式の定数の有意性を検討した結果を表 9 に示す。得られた修正 Page 式は,

$1/T = -0.014 + 1.458/Z + 0.657V/HL$ である。定数項 $b_0 (= -0.014)$ は危険率 5% でも有意ではなく, 0 とおけることが分かる。また, $K (= 1.458)$ は Page が理論的に誘導した値 9/8 よりやや大きい。

3) 応用例 3

説明変量が数量化理論第 1 類で扱った場合のように, 数量でなく性質の違いを表すカテゴリで与えられている場合にも, 重回帰分析はしばしば用いられる。

表 10 及び 11 は Grandmaison らの実験¹²⁾ によるもので, Sybron-Brinkman ZM-1 をミルに用いて針葉樹及び広葉樹パルプを処理したときの条件並びに処理後の平均繊維長を示したものである。表 11 のデータを重回帰分析に適用するためには説明変量が数量で表される必要がある。ここでは個々の変量は“高”か“低”かの 2 カテゴリのうち必ず一つが当てはまるため, これに便宜上 +1 と -1 の数量をそれぞれ割当て計算する。この仮りの数量は一般にダミー変数と呼ばれる。

表 12 は Grandmaison らが得た偏回帰係数^{*3)} である。針葉樹パルプではスクリーンサイズとパルプの前処置湿度に, 広葉樹パルプではスクリーンサイズに有意性が認められる。また, ここでは変量の交互作用も含めて検討されているが, それらに有意性は認められなかった。

ダミー変数の利用は他の分野, 例えば化学や薬学において近年構造活性相関として知られる分野において, 分子構造の質的特徴を重回帰式に盛り込む目的で利用されている¹³⁾。

*3) 筆者の計算では若干異なる値を得ている。

表 10 7 つの変量の低及び高値

	低	高
1. ローターのタイプ, 歯の数	6	12
2. スクリーンサイズ, mm	0.5	5.0
3. パルプ供給速度, g/min	5	10
4. ローター速度, rpm	10,000	20,000
5. パルプ前処置湿度, %RH	0	70
6. パルプ試料の大きさ, mm	6.4	12.7
7. 空気流速, g/min	90	190

表 11 実験計画及び繊維長の測定値

実験番号	変 量							平均繊維長, mm	
	1	2	3	4	5	6	7	針 葉 樹	広 葉 樹
1	—	—	—	—	+	+	+	0.472	0.197
2, 17	+	—	—	+	—	—	+	0.220, 0.193	0.199, 0.262
3	—	+	—	+	—	+	—	0.589	0.257
4, 18	+	+	—	—	+	—	—	0.772, 0.599	0.456, 0.354
5	—	—	+	+	+	—	—	0.483	0.270
6, 19	+	—	+	—	—	+	—	0.355, 0.207	0.194, 0.231
7	—	+	+	—	—	—	+	0.583	0.457
8, 20	+	+	+	+	+	+	+	0.490, 0.433	0.348, 0.278
9	+	+	+	+	—	—	—	0.325	0.446
10, 21	—	+	+	—	+	+	—	0.750, 0.726	0.484, 0.463
11	+	—	+	—	+	—	+	0.442	0.271
12, 22	—	—	+	+	—	+	+	0.302, 0.146	0.188, 0.164
13	+	+	—	—	—	+	+	0.446	0.292
14, 23	—	+	—	+	+	—	+	0.619, 0.515	0.331, 0.337
15	+	—	—	+	+	+	—	0.188	0.268
16, 24	—	—	—	—	—	—	—	0.288, 0.101	0.235, 0.238

実験番号 1~8 は 2⁷-1 実験計画, 実験番号 9~16 は逆の条件によるもの, 実験番号 17~24 は繰り返し実験である。正負の符号は高及び低のレベルを示す。

表 12 偏回帰係数の推定値

変 量	偏 回 帰 係 数	
	針葉樹	広葉樹
定 数	0.430	0.302
1. ローターのタイプ	-0.051	0.002
2. スクリーンサイズ	0.119 ^a	0.070 ^a
3. 供給速度	0.012	0.025
4. ローター速度	-0.050	-0.016
5. パルプ前処置湿度	0.074 ^a	0.014
6. パルプ試料の大きさ	-0.005	-0.029
7. 空気流速	-0.005	-0.019

^a 95%有意

4) 応用例 4

紙には関係ないが一つの問題に異なる側面から重回帰分析と数量化理論を適用した例を以下に示す。

図 8 L-Aspartylaminoethyl ester 類

岩村¹³⁾は甘味物質である L-Aspartylaminoethyl ester 類 (図 8) の甘味度を次式のような置換基の立体パラメータと電子的パラメータの組合せで説明した。

$$\log SP = 0.67\sigma + 3.36L_2 - 0.29L_2^2 + 4.18W_{u1} - 0.85W_{u1}^2 - 0.53L_1 - 11.33 \quad (14)$$

ここで, SP=甘味度, Sweet potency の略
 L_1, L_2, W_{u1} = 置換基の立体パラメータ
 σ = 電子的パラメータ

この場合, 説明変量に 2 乗項が含まれているが, それを個別の変量とみなせば重回帰分析は適用できるわけである。計算に用いられた分子数は 51 で, 重相関係数は 0.88 であった。

これに対して, 筆者¹⁴⁾が R_1 及び R_2 の置換基の種類をカテゴリ化し数量化理論第 1 類を用いて得た予測モデルが次式である。

$$\begin{aligned} \log SP = & -0.51\delta_{R_1(H)} + 0.57\delta_{R_1(CH_3)} \\ & + 0.10\delta_{R_1(CH_2OH)} - 0.09\delta_{R_1(C_2H_5)} \\ & - 0.19\delta_{R_1(i-Pr)} + 0.12\delta_{R_1(COOCH_3)} \\ & - 0.25\delta_{R_2(CH_3)} + 0.15\delta_{R_2(C_2H_5)} \\ & - 0.47\delta_{R_2(n-Pr)} + 0.10\delta_{R_2(i-Pr)} \\ & + 0.22\delta_{R_2(c-Pr)} + 0.15\delta_{R_2(c-Bu)} \\ & + 0.17\delta_{R_2(i-Bu)} + 1.40 \end{aligned} \quad (15)$$

ここで, $R_1 = \text{H}, \text{CH}_3, \text{CH}_2\text{OH}, \text{C}_2\text{H}_5, i\text{-Pr}, \text{COOCH}_3$ のうちいずれかの置換基

$R_2 = \text{CH}_3, \text{C}_2\text{H}_5, n\text{-Pr}, i\text{-Pr}, c\text{-Pr}, c\text{-Bu}, i\text{-Bu}$ のうちいずれかの置換基

$\delta_{Ri(i)}$ = 該当する置換基 R_i の種類に対して,
1
該当しない置換基 R_i の種類に対して,
0
($i=1, 2$)

図 9 2つの手法による甘味度 (log SP) の実測値と推定値の比較

注 * 岩村が用いた51分子のうち数量化モデルには36分子を使用した

表 13 Aspartylaminoethyl esters における置換基のタイプによる甘味度 (log SP) の差

(a) 置換基 R_1 の場合

基準カテゴリー	比較すべきカテゴリ				
	CH_3	CH_2OH	C_2H_5	$i\text{-Pr}$	COOCH_3
H	1.07	0.61	0.42	0.31	0.62
CH_3		-0.47	-0.66	-0.76	-0.45
CH_2OH			-0.19	-0.29	0.01
C_2H_5				-0.10	0.20
$i\text{-Pr}$					0.31

(b) 置換基 R_2 の場合

基準カテゴリー	比較すべきカテゴリ					
	C_2H_5	$n\text{-Pr}$	$i\text{-Pr}$	$c\text{-Pr}$	$c\text{-Bu}$	$i\text{-Br}$
CH_3	0.40	-0.22	0.35	0.47	0.39	0.41
C_2H_5		-0.62	-0.05	0.07	-0.01	0.01
$n\text{-Pr}$			0.57	0.69	0.61	0.63
$i\text{-Pr}$				0.12	0.04	0.06
$c\text{-Pr}$					-0.07	-0.06
$c\text{-Bu}$						0.02

ただし, ここでは岩村が用いた51分子のうちカテゴリ化の容易な36分子を使用した。また, 岩村の結果と筆者の結果を比較したのが図9である。重相関係数は前者で0.84, 後者で0.87であり, 両手法とも同程度の精度が得られていることが分かる。

表13は式(4)のカテゴリ数量をもとに置換基間の甘味度の差を示したものである。表の見方は, 例えば(a)において R_1 の位置に水素原子がある場合, それをエチル基に換えると甘味度が log SP 値で0.42増加することを意味している。

こうして, 甘味度を説明するのに重回帰分析モデルでは置換基の立体パラメータや電子パラメータを用い, 数量化モデルでは置換基の種類のような質的情報のみを用いて異なる側面からの解析を行えることが分かる。

2.3 時系列解析⁹⁾¹⁹⁾

時系列解析のモデルは, 過去のデータの時間的な並びをもとに次の時点に得られるデータを予測するもので, 定常な時系列については次式で表される。

$$\varepsilon(s) = x(s) - \sum_{m=1}^M a(m)x(s-m) \quad (16)$$

ここで, $x(s)$ = 時点 s における特性値

$a(m)$ = 時点 $(s-m)$ における特性値 $x(s-m)$ の偏回帰係数

M = 過去の時点の数

$\varepsilon(s)$ = 残差

これは, 自己回帰モデルと呼ばれ, 前述の重回帰分析の説明変量のところに目的変量の過去の時系列を置き換えると理解しやすい。やはり最小2乗法により $\varepsilon(s)$ の2乗平均を最小にするような $a(m)$ ($m=1, 2, \dots, M$) を求める。

ここで, 最適な M の決定, すなわち過去のどの時点までデータをとるかが重要な問題となる。詳細は文献¹⁹⁾に譲るが, 最終予測誤差 FPE や情報量規準 AIC によりデータから決定する。

また, パワースペクトル密度が, 得られた残差分散と偏回帰係数をもとに次式により計算できる。

$$p_{xx}(f; \Delta t) = \Delta t \sigma^2 / \left| 1 - \sum_{m=1}^M a(m) \exp(-i2\pi f \Delta t m) \right|^2 \quad (17)$$

$$(-1/2\Delta t \leq f \leq 1/2\Delta t)$$

ここで, f = 周波数

Δt = データ読み取りの時間間隔

σ^2 = 残差分散

1) 応用例

自己回帰モデルを紙の地合評価に用いた例²⁰⁾を次に示す。

紙のベータ線透過像(図10)をフィルムに現像すると、微小部分の坪量の変動が濃度変化として再現される。マイクロフォトメータを使用してフィルムを直線的に走査し、フィルムの濃度を連続的に測定すると、図11のようなチャートが得られる。測定間隔を決めてデータを採り、この濃度変化を定常時系列なデータと考え、これに自己回帰モデルを当てはめて、パワースペクトル密度の推定を行った。

図12は地合の異なる2種類の試験紙の地合のパワースペクトルである。ここで、マイクロフォトメータの測定面積は $250\mu\text{m}^2$ 、データの測定間隔は 0.625mm で、紙の流れ方向に走査した。Aの試験紙は全周波数領域のスペクトル密度が小さい値で一定していることから、坪量の変動が小さく、大きいフロックの少ない均一な地合であることが分かる。一方、Bの試験紙は低周波数領域のスペクトル密度が急に高くなっており、データの分散も大きいことから坪量の変動が大きく、フロックの大きい悪い地合であることが分かる。

図13は上のデータの3点ずつを一まとめに平均化し、見かけ上、データの測定間隔を3倍の 1.875mm とした場合のパワースペクトルである。Bの試験紙は周波数 0.1cycle/mm 付近のスペクトル密度が高く、フロックの大きさは波長の $1/2$ と考えると 5mm 程度のフロックが多いことが分かる。

図10 ベータ線透過像

図12 地合のパワースペクトル(1)

図13 地合のパワースペクトル(2)

3. 要約及びデータ構造の解明

3.1 多次元尺度構成法 MDS¹⁵⁾

Torgerson は、予め複数の個体について相互の距離が数量的に与えられている場合にそれら個体をもとの距離関係を保ちながらユークリッド空間に配置できる方法を開発した。これは多次元尺度構成法 MDS と呼ばれ、心理学の分野で発達した。Kruskal や Shepard はこれに対して一対ずつの個体のランクオーダーのみのデータから多次元空間に個体の配置を行う方法を開発した。Torgerson のメトリックな MDS に対して後者はノンメトリック MDS と呼ばれる。

メトリック MDS においては、基本的にはピタゴラ

図11 ミクロフォトメータで測定した地合の濃度変化

スの定理を多次元空間に適用した2点間の距離と座標との次の関係から出発する。すなわち、

$$d_{ij} = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2} \quad (i, j = 1, 2, \dots, n) \quad (18)$$

d_{ij} : 個体 i と j の間で認められた差

X_{ik}, X_{jk} : 第 k 次元 (軸) 上の個体 i, j それぞれの座標

m : 次元数

n : 個体数

であり、 d_{ij} を測定により得て X_{ik}, X_{jk} を逆推定するわけである。主観的な反応の場合は物理量のべき関数で表した次式が用いられる。

$$d_{ij} = V \left[\sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2} \right]^p \quad (19)$$

V, p は前式の単調性を変えことなく各判定者の反応形式の差を調整するための実数値である。

1) 応用例

Lyne⁽⁶⁾ は印刷物の品質評価に MDS を導入した。それによると、判定者に図 14 のような八つの試験印刷物を一対ずつ見せ、2 者の差及びし好の判定 (図 15) を求めた。図 16 は判定者グループの差のデータに MDS を適用した例である。第 1 因子 (次元) は、その配置モデルにおける試料間の距離の最大分散を説明する軸である。第 2 及びその他の因子は第 1 因子に直交している。得られた試料の序列を参考に各因子に対して主観的な意味付けを行い、モットリングとコントラストを得た。また、し好データをもとに個々の印刷物についての次のようなユーティリティという尺度を設けた。

図 14 パブリカテストプリント

$$U_i = \sum P_{ij}/n$$

(20)

U_i : 第 i 番目のユーティリティ

P_{ij} : モデルから推測した第 i 番目と第 j 番目の印刷物のし好値

n : 印刷物の数

これを目的変数として差の座標を用いて重回帰式を

試料 A 及び B		
差の判断		
↑ きわめてよ く似ている		↑ 全く異な っている
し (嗜) 好の判断		
↑ A のほうが 好きである	↑ どちらで もない	↑ B のほうが 好きである

図 15 印刷物品質評価に用いる質問用紙

図 16 差の判断に基づく試料の布置と判定者のし (嗜) 好ベクトル

図 17 判定者グループ別の各因子の重み

作成し、製紙者、印刷者、広告者及び読者それぞれの判定者グループごとに各因子の嗜好への重みを算出し比較した(図17)。印刷者は彼らの制御可能な因子、例えばコントラストに大きなウェートを置く傾向があり、他方、製紙者や広告者は紙の色に対して敏感な傾向を示した。モットリングには4グループとも顕著な関心を示した。

3.2 主成分分析及び因子分析⁷⁾⁸⁾

主成分分析の概略は次のとおりである。今、 n 個の試料について p 種類の特性を測定した場合を考える。 p 種類の特性のうちのどの2者を選んでも互いに無関係であるような組合せはほとんど無いと考えるのが一般的である。そのため、特性間の係わり合いを知るには考えられるすべての組合せについて相関を吟味しなければならないことになる。しかし、このようなやり方は特性の数が少ないときには問題ないが、数十もの特性を扱わなければならないときには実用的とは言えない。

そこで、主成分分析法では p 個の特性の組合せからなる新しい $m(m < p)$ 個の特性を考える。その新しい特性は個々のもとの特性に重みをもたせることによって次式のように表される。

$$\begin{aligned} Z_1 &= l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ Z_2 &= l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ &\vdots \\ Z_m &= l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \end{aligned} \quad (21)$$

ここで、 $l_{k1}^2 + l_{k2}^2 + \cdots + l_{kp}^2 = 1 (k=1, 2, \cdots, m)$ とする。 $Z_1, Z_2, \cdots, Z_m$ はそれぞれ第1主成分、第2主成分、 $\cdots$ 、第 m 主成分と呼び、これら主成分は互いに直交している。また、これら主成分は最大分散のものから順に $Z_1, Z_2, \cdots, Z_m$ とする。結果として、主成分を求めることはもとの特性 $x_1, x_2, \cdots, x_p$ のもつ情報の損失を最小にとどめながらデータを要約することに相当する。

これら主成分の意味は図18の2変数の場合を考えたと理解しやすい。図中で Z_1 はデータが最も広く散

図18 主成分の意味

らばっている方向を通る軸であり、 Z_2 はそれに垂直な方向の軸である。

最終的には主成分は次式を解くことにより得られる。

$$\Sigma \mathbf{l} = \lambda \mathbf{l} \quad (22)$$

ここで、 Σ = データの分散共分散行列

$\mathbf{l}$ = 固有ベクトル。もとの特性 $x_i (i=1, 2, \cdots, p)$ に掛る係数

λ = 固有値

ただし、 $x_i (i=1, 2, \cdots, p)$ の測定単位が異なっている場合は x_i の代わりに、平均0、分散1に標準化した変量、

$$x_i^* = (x_i - \text{平均}) / \text{標準偏差}$$

を用い、式(22)は分散共分散行列 Σ の代わりに相関係数行列 R を用いて解く。

最大固有値に対応する固有ベクトルが第1主成分 Z_1 の特性 $x_i (i=1, 2, \cdots, p)$ の係数である。以下、固有値の大きさの順に第2主成分、第3主成分、 $\cdots$ が求まる。また、個々の主成分の意味は主成分ともとの特性との相関係数である因子負荷量から解釈する。

これに対して、因子分析モデルは雑な言い方をすれば、主成分分析モデルを一見逆さまにしたようである。上式の Z_k の位置にもとの特性を置き、 x_{ij} の位置に共通因子と呼ばれる新しい特性(主成分分析の場合の主成分に対応)を置いたモデルを考えると分かり易い。このほか、特殊因子と呼ばれる共通因子とは独立な因子による影響が加わる。主成分分析は因子分析の手法の一つにみなされることもある。

1) 応用例1

筆者ら¹⁷⁾は市中での流通を終えた古い紙幣(現行券ではなく旧券)80枚について取扱いによる位置汚れを調べるために、周辺の余白の部分について表裏28箇所の反射率の測定を行った。得た反射率から色差を計算し、個々の測定位置における色差を変量として主成分分析を行った。通常、各主成分の意味はもとの特性との相関係数に相当する因子負荷量の値(表14)から解釈するが、ここでは因子負荷量の符号が測定位置によってグループとしての挙動を示したことに着目し図19を得た。各主成分は次のように解釈できる。

Z_1 : 紙幣の表裏のどの位置においても、因子負荷量の絶対値及び符号が等しいことから、表裏一様な汚れを表す因子であると考えられる。

Z_2 : 因子負荷量の符号は紙幣の表ですべて正、裏ですべて負になっており、表裏の汚れ方の違いを表す因子であると考えられる。この違いは二つ折りのような

表 14 銀行券の汚れの主成分分析 (因子負荷量)

測定位置	Z ₁	Z ₂	Z ₃
表 1	0.6504	0.2927	0.3852
2	0.6155	0.4832	0.0031
3	0.6788	-0.1188	0.1263
4	0.5366	0.3978	-0.0564
5	0.6036	0.2581	-0.3559
6	0.6679	0.1032	-0.1989
7	0.5538	0.2411	-0.3909
8	0.6370	0.1800	-0.0306
9	0.6332	0.4443	-0.1839
10	0.6105	0.0597	-0.0490
11	0.5047	0.3281	0.0535
12	0.7559	0.2421	0.0681
13	0.6367	0.3462	0.1612
14	0.5835	0.4781	0.2782

裏 15	0.6954	-0.0049	0.0136
16	0.6438	-0.0722	0.4395
17	0.6398	-0.5267	-0.0103
18	0.6731	-0.2800	0.0059
19	0.6439	-0.2169	-0.3017
20	0.7469	-0.1781	-0.2498
21	0.6723	-0.2892	-0.3067
22	0.6618	-0.2395	-0.3988
23	0.7128	-0.1572	-0.1254
24	0.6088	-0.4740	-0.0886
25	0.6052	-0.2086	0.4342
26	0.6391	-0.2719	0.3022
27	0.6103	-0.2684	0.3150
28	0.5974	-0.3247	0.2258
寄与率 %	40.8	8.9	6.1
(累積寄与率)		(49.7)	(55.8)

第 1 主成分 (Z₁) 第 2 主成分 (Z₂) 第 3 主成分 (Z₃)

図 19 汚れを支配する主な因子

行為に基づくものと思われる。

Z₃: 因子負荷量の符号は紙幣の右半分がすべて負、左半分がすべて正になっており、左右の汚れ方の違いを表す因子であると考えられる。紙幣の左半分と右半分は扱われ方に違いのあることを示している。

以上の 3 因子が紙幣の汚れの主な形態を示すものと

考えらる。

2) 応用例 2

高橋ら¹⁸⁾は上質紙、コート紙、ポンド紙、木綿紙を含む 29 種の用紙を用いて、継次範ちゅう法により 21 の風合関連感覚項目について尺度化を行い、相関分析の結果、被験者の判断の一致性の良い感覚尺度を導き

表 15 触系官能 8 尺度についての因子分析 (主因子法)

(a) 固有値と寄与率

No.	固有値	%
1	5.2311	65.3885
2	1.6987	86.6219
3	0.6952	95.3121
4	0.0736	96.2316

(b) 初期抽出因子負荷行列

感覚尺度	F 1	F 2	F 3	F 4
1 厚 い	0.9101	0.1704	-0.2839	-0.0843
2 堅 い	0.8888	0.2982	0.3443	-0.0650
3 滑 ら か	-0.6134	0.7911	0.0203	0.0771
4 ごわ ごわ	0.8883	-0.4004	0.2135	-0.0297
5 ばり ばり	0.8160	0.2441	0.4835	0.0921
6 丈 夫 な	0.7706	0.4309	-0.2715	0.1407
7 軽 い	-0.8166	-0.4104	0.3036	0.0913
8 つかみやすい	0.7219	-0.6172	-0.2247	0.1374

(c) 共通性

感覚尺度	抽出した因子数		
	2	3	4
1 厚 い	0.85732	0.93792	0.94502
2 堅 い	0.87889	0.99743	1.00166
3 滑 ら か	1.00210	1.00251	1.00845
4 ごわ ごわ	0.94940	0.99498	0.99586
5 ばり ばり	0.72544	0.95921	0.96769
6 丈 夫 な	0.77950	0.85321	0.87301
7 軽 い	0.83526	0.92744	0.93577
8 つかみやすい	0.90208	0.95257	0.97144

(d) バリマックス回転後の因子負荷行列

感覚尺度	2 因子解		3 因子解			4 因子解			
	1	2	1	2	3	1	2	3	4
1 厚 い	0.8571	-0.3503	0.8883	-0.3404	0.1818	0.8888	-0.3407	0.1787	-0.0843
2 堅 い	0.9085	-0.2314	0.6193	-0.2209	0.7517	0.6218	-0.2212	0.7496	-0.0650
3 滑 ら か	-0.0865	0.9973	-0.0951	0.9962	-0.0311	-0.0948	0.9963	-0.0306	0.0771
4 ごわ ごわ	0.5293	-0.8181	0.3612	-0.8119	0.4531	0.3625	-0.8121	0.4517	-0.0297
5 ばり ばり	0.8180	-0.2373	0.4717	-0.2279	0.8275	0.4744	-0.2282	0.8259	0.0921
6 丈 夫 な	0.8811	-0.0558	0.9000	-0.0456	0.2028	0.9007	-0.0459	0.1998	0.1407
7 軽 い	-0.9087	0.0979	-0.9403	0.0875	-0.1889	-0.9409	0.0878	-0.1857	0.0913
8 つかみやすい	0.2719	-0.9100	0.3567	-0.9068	-0.0546	0.3562	-0.9069	-0.0559	0.1374

出した。それらは、厚い、堅い、滑らか、ばりばり、ごわごわ、丈夫な、軽い、つかみやすいの8尺度であった。これらをもとに因子分析を行い、表15に示すような主要な3因子の因子負荷行列を得た。表には4因子解も求められているが寄与は極めて小さく無視できる。ここでの因子負荷量は抽出された因子に軸の回転(バリマックス回転)を施し、各尺度への抽出因子の寄与を明確化したものである。これらの因子はそれぞれ厚み感因子、平滑感因子、堅さ因子と判定された。

3.3 数量化理論第3類¹⁾⁻⁵⁾⁹⁾

個体が n 個あり、それぞれが m 個の特性のうち幾つかに反応したとき、個体と特性を順次入れ替え、似たもの同士が集るようにすることを考える。

もとのデータ行列を次のように表す。

$$D = \{\delta_{ij}\} = \begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_{nm} \end{pmatrix} \begin{matrix} \text{特 性} \rightarrow \\ \\ \\ \downarrow \\ \text{個 体} \end{matrix}$$

ここで、 δ_{ij} は個体 i が特性 j をもっているか否かに依存して1か0の値をとる。

この行と列の入れ替えから、似たもの同士が集るようにするためには、数学的には個体と特性にそれぞれ $x_i (i=1, 2, \dots, n)$, $y_j (j=1, 2, \dots, m)$ のような数量を与え、個体数量と特性数量間の相関係数が最大となるような $\{x_i\}$ 及び $\{y_j\}$ を求めることになる。

最終的には次の固有方程式を解く問題に帰着する。

$$C Y = \lambda Y \quad (23)$$

ここで、 $C = (C_{jl})$

$$C_{jl} = 1/\delta_{.j} \sum_{i=1}^n \delta_{ij} \delta_{il} / \delta_{.i} - \delta_{.l} / n$$

$$(j, l = 1, 2, \dots, m)$$

$$x_i = 1/\delta_{.i} \sum_{j=1}^m \delta_{ij} y_j$$

$$(i = 1, 2, \dots, n)$$

ただし、 $n \geq m$, $\delta_{.j} = \sum_{i=1}^n \delta_{ij}$, $\delta_{.i} = \sum_{j=1}^m \delta_{ij}$, $\lambda =$ 固有値

目的の数量 $\{x_i\}$, $\{y_j\}$ は固有値最大のものから順に対応する2, 3の固有ベクトルを選び、得られた数量 $\{x_i\}$, $\{y_j\}$ を基に個体と特性の関連を多次的に分析する。ここで、固有値の平方根は個体数量と特性数量の相関係数に等しく、最大のものから第1相関係数、第2相関係数、……と呼ぶ。

1) 応用例1

官能試験に応用した例¹⁷⁾を次に示す。

被験者85名に各国の銀行券30種類を提示し、銀行券として好ましく感じる券種の指摘を求めた。各被験者の回答を次のように集計し、被験者を個体変量に、券種を特性変量にして数量化を行った。

被験者	券種				
	A	B	C	D	E
1	○		○		
2	○	○			
3		○	○		
.....

図20 好ましいとされた銀行券の布置

ここでは○印の箇所の δ_{ij} が 1, 無印の箇所の δ_{ij} が 0 の値をとる。

図 20 は好ましいとされた銀行券の布置 ($\{y_j\}$ に対応) を第 1 及び第 2 相関軸からなる平面上に示したものである。類似の判断をされた券種間の距離は近く, 異なった判断をされた券種間の距離は遠くなるように数量化されている。日本銀行券(現行券ではなく旧券)の近辺にはカナダ, イギリス, スペイン, オーストリアなどの銀行券が位置し, 他方, オランダ, フランス, スイスなどの銀行券は日本銀行券とはかなり異なったイメージで捕らえられていることが分かる。また, これらの券種を支持した被験者の布置 ($\{x_i\}$ に対応) を第 1 及び第 2 相関軸からなる平面上に年齢別に示したのが図 21 である。これを図 20 と重ね合わせてみると, 若年者層の側では好みはかなりバラエティに富んでいるが, 高年令者層の側では日本銀行券に近いイメージを有した券種に好みが集まる傾向にあることが分かる。

図 21 好ましいと判断した被験者の布置

この例におけるように, 数量化理論第 3 類は特性に対する個体の反応数に制限を要しないため, データに欠測値があってもよいし, 特に人間の心理の表現の場合には自由な発想をデータにそのまま生かすことができる点で優れているといえよう。

3.4 数量化理論第 4 類²⁾

これは前述の MDS と同じような目的で使用できる。取り扱うデータとしては 2 者比較に基づく親近性データである。

今, 仮に n 個の個体があって, その i 番目と j 番目の間で相互の親近性を表す何等かの数量 e_{ij} ($i, j=1, 2, \dots, n$) が与えられているとき, 親近性の強い個体間では距離が小さく, 親近性の弱い個体間では距離が大きくなるように, 各個体に多次元的な数値を付与しようというものである。親近性の数量 e_{ij} を用いて, 類似する個体間の距離を最小にするような個体の数量 x_i は, x_i の分散を一定とする条件のもとに次式を最大にすることにより求められる。

$$Q = - \sum_{i=1}^n \sum_{j=1}^n e_{ij} (x_i - x_j)^2 \quad (24)$$

結局, 次の固有方程式を解く問題に帰着する。

$$H X = \lambda X \quad (25)$$

ここで, $X = n$ 個の個体の数値を表すベクトル

$H =$ 要素 h_{ij} からなる行列

$$h_{ij} = \begin{cases} e_{ij} + e_{ji} & (i \neq j) \\ - \sum_{k=1}^n (e_{ik} + e_{jk}) & (i = j) \end{cases}$$

$\lambda =$ 固有値

固有値最大のものから順に対応する 2, 3 の固有ベクトルを選べば, それが求める個体の数量 x_i であり多次元の値が付与されることになる。

1) 応用例

表 16 かたさ感における用紙相互の親近性

	書籍用紙	N・BKP紙	コート紙	堅紙	証券用紙A	みつまた紙A	L・BKP紙	みつまた紙B	木綿紙	証券用紙B
書籍用紙		3	1	0	0	1	1	3	1	0
N・BKP紙	3		15	13	6	22	20	9	19	7
コート紙	1	15		20	6	7	16	21	7	2
堅紙	0	13	20		4	13	21	9	14	4
証券用紙A	0	6	6	4		10	7	6	15	20
みつまた紙A	1	22	7	13	10		11	4	7	10
L・BKP紙	1	20	16	21	7	11		20	10	3
みつまた紙B	3	9	21	9	6	4	20		3	1
木綿紙	1	19	7	14	15	7	10	3		10
証券用紙B	0	7	2	4	20	10	3	1	10	

表 17 疲労処理紙の感覚と物性の相関

	坪 量	厚 さ	密 度	か さ	平滑度	曲げ剛性	透気度	引張強さ
か た さ 感	0.146	-0.441	0.418	-0.428	-0.049	0.948	0.789	0.782

上質紙、コート紙、じん皮繊維紙を含む10種類の用紙を紙幣の流通を仮想した一定の処理条件のもとに疲労させ、それらを2枚1組として被験者(31名)に提示し、かたさ感についてサンプル間の類似性の有無の判定を求めた¹⁷⁾。

表16はかたさ感についてサンプル間に差を感じないとしたデータを集計し、度数を表示したものである。これを親近性のデータ e_{ij} とし数量化した結果を図22に示す。

図 22 数量化理論第4類によるかたさ感の分析

書籍用紙を除き、サンプルは第II軸に沿って一列に並ぶことが分かる。このことは、書籍用紙だけが他とは異なる基準で判断され、しかも弁別も容易に行われていることを意味する。次に、判断の基準を異にする書籍用紙を除き、図22の第II軸と物性との相関を調べたのが表17である。第2軸で表されるかたさ感は曲げ剛性との相関が高く ($r=0.948$)、このことから疲労紙のかたさ感を曲げ剛性で評価できることが分かる。

4. おわりに

データ解析法には予測の手法と要約の手法がある。また、データのタイプは質的データと数量データとに分けられ、前者には数量化理論が、後者には狭義での多変量解析法が適用されるのが一般的である。

しかし、実際には重回帰分析におけるダミー変数の利用のように、質的データに多変量解析が適用される例はしばしばある。逆に、数量データ、例えば実験の結果得られた測定値でも実験誤差を考慮すれば細かい

端数の値は意味がなく、せいぜい上、中、下程度で数値を捕らえるほうが実際的な場合もあり、そのときは数量化理論の適用が可能である。

解析の進め方も、まず要約の手法でデータ構造に探りを入れてポイントを絞り、次いで予測の手法を適用するのが効果的であるように思われる。例えば、ある物性を他の物性の組合せて説明する場合、何を説明変量に選ぶかが重要であり、それは予測式の構成の前段で考察されねばならないことである。

ここでは触れなかったが他の関連する有益な手法として、数量化理論第2類、判別分析、情報量規準AICがあることを付記しておく。手元にこれらの手法を適用するための紙に関係したデータがないだけの理由で割愛した。

最後に、以上の計算はすべて手計算では実行不可能であるため、コンピュータの使用が必要である。16ビットパソコンの場合はBasicプログラムで十分である。市販のソフトウェアを使うか、以下の参考文献を頼りにソフトウェアを自作するのもよいと思われる。計算に様々なオプションを加える可能性を考えると後者の方が多様な使用の仕方ができると筆者は考える。

文 献

- 1) C. Hayashi, Annals of The Institute of Statistical Mathematics, 3, 69~98 (1952)
- 2) 特集 数量化理論, 数理科学, [204] (1980)
- 3) 駒沢, 多元的データ分析の基礎, 朝倉書店 (1978)
- 4) 駒沢, 数量化理論とデータ処理, 朝倉書店 (1982)
- 5) 林, 樋口, 駒沢; 情報処理と統計数理, 産業図書 (1978)
- 6) R.W. Allison, Appita, 34, [6], 465~470 (1981)
- 7) 奥野ほか, 多変量解析法, 日科技連 (1978)
- 8) 河口, 多変量解析入門Ⅰ, 森北出版 (1978)
- 9) 河口, 多変量解析入門Ⅱ, 森北出版 (1978)
- 10) J.d'A. Clark, Tappi, 45, [8], 628~634 (1962)
- 11) J.d'A. Clark, Tappi, 56, [7], 122~125 (1973)
- 12) E. W. Grandmason, A. Gupta; Tappi, 69, [8], 110~113 (1986)
- 13) 構造活性相関懇話会編, 薬物の構造活性相関Ⅱ, 化学の領域 増刊136号, 南江堂 (1982)
- 14) H. Yamazaki, D.G. Howerly; J. Chemometrics, submitted for publication
- 15) 林, 飽戸, 多次元尺度解析法, サイエンス社 (1976)
- 16) M. B. Lyne, Tappi, 62, [11], 103~107 (1979)
- 17) 山崎, 池田, 馬場, 長谷川, 森本; 印刷局研究所報告 No. 52, 1~62 (1983)
- 18) 高橋, 長谷川, 柳川; 印刷局研究所報告 No. 50, 1~100 (1981)
- 19) 赤池, 中川; ダイナミックシステムの統計的解析と制御, サイエンス社 (1972)
- 20) 高津, 山崎; 未発表