多変量解析の“からくり”をどう捉えるか
その１　主成分分析・因子分析

木下　英明　　加來　秀俊

Mechanisms of multivariate statistical analysis
Part 1: Principal component analysis and factor analysis

Hideaki KINOSHITA　　Hidetoshi KAKU

While students are often capable of performing statistical analysis with aid of a computer, they also often do not understand the mathematical principles behind the operation. This understanding is essential for an effective and thorough investigation. The primary aim of this article was to assist students in understanding the numerical value displayed on the computer screen during multivariate statistical analysis. Mechanisms of principal component analysis and factor analysis were explained using as few mathematical term as possible.

１．緒言

世の中の現象は、いくつもの要因が重なってできており、それを解明するために多変量解析法が考案された。解析に必要な計算は大変複雑で、極めて時間がかかるので理論だけが存在した時代もあったが、近年はパーソナルコンピューター（パソコン）の急速な進歩のおかげで、身近なものになった。目的とする計算値は楽に得られるようになった反面、パソコンの導いた数値が何を意味するかがよく解らないままに解析を終えることがある。何をやったかを考えなおしてみるとばく然とした気分になることがあるが、これは多変量解析の“からくり”を理解していないためにおきることかと思う。ここでは多変量解析をより身近なものとするために、そのからくりについて説明したい。

幾つかの組織や幾人かについてｐ種類の測定値（x_1、x_2…… x_pと表示され、各々は変数と呼ばれる、例えば英語、社会、理科 …… の成績はx_1、x_2、x_3…… と表記されることを知ると分かりやすい）が得られたとき、これらの変数に対する重みをa_1、a_2…… a_pとすれば、重みづけの合計点Ｙは、

$Y=a_1x_1+a_2x_2+\cdots\cdots+a_px_p$　　と表され、重みを変化させることで、いくつもの合計点（合成得点）を作ることができる。入学試験では、幾つかの教科の合計得点を判断の基準に

しているが、すべての教科得点に対する重みを１とした場合の合成得点が多く利用されている。このように重みの係数を１にすることが一般には多いが、より詳しい判断をするためには重みを変えて考察することが肝要である。

教科の試験成績あるいは食品嗜好の調査測定値は、幾つかの要因からなっていることが多く、その一つである x_1 は

$x_1 = a_1f_1 + a_2f_2 + \cdots\cdots + a_pf_p$ のような合計点として表すことができる。

ここでも a_i は重み付けの係数であり、f_i は変数間の相関に関わる要因を示すもので、測定値から推測される。ここに示したような一次式を基に、多くの変数間の相関関係を考慮して、各変数ないし変数から推測されたものに、目的に応じた最適な重みを与え、その結果として得られる合成得点を多次元空間に位置づけて（ベクトルの概念）分析する法が多変量解析と呼ばれるものである。

参考：数学一般では変数間の関係を方程式で表現するとき、一方を従属変数、他方を独立変数とに区別する。統計学では目的変量（基準変量、従属変量）と説明変量（予測変量、独立変量）とに区別される。前述した合計点 **Y** は目的変量で、教科の得点 x_1、x_2、……は説明変量である。元の変量に重みをつけたものを繋いで、新しい変量を合成することを線形結合といい、この変換が多変量解析で大きな役割を果たしている。

多種多様の特性からなる事象に対応して測定された変数の相互関連を分析して、少数の総合的特性値にまとめる主成分分析、そして変数の背後にあって、変数間の相関関係を規定している潜在因子を探る因子分析は多数の変数の特徴を分析するものである。性格検査や入学試験での合成得点がある値（外的基準－重みづけの基準となる変数）より大きいかどうかでセールスマンとしての適性や入学後の学業成績を判別・予測するものに重回帰分析がある。多変量解析はこれ以外にもいくつかの分析法があるが、ここでは外的基準がなく変数の内的構造を分析する主成分分析、および因子分析についてまず説明したい。そこでは独特の演算形式が定義されるベクトルや行列の理解が必要な場合がある。これについては、必要最低限について説明する。高等学校で使用される教科書レベルのベクトル・行列について一通り目を通して、できれば幾つかの練習問題を解いてみると、多変量解析を理解しやすくなる。それが叶わないなら、ベクトル・行列の演算は私たちが具現できない事象を、数学の言葉で構築し、一定の法則の下で処理していることを知ればよいかと思う。

２．主成分分析（principal component analysis）

ある問題に対して、いくつかの要因が考えられるとき、要因を独立に扱うのではなく総合的に扱おうとするのが主成分分析である。

幾つかの説明変量、x_1～x_pの総合的特性を、

$a_1x_1+a_2x_2+\cdots\cdots+a_px_p$のような一次式で表現するのが主成分分析で、多くの変量の値をできるだけ情報の損失がないように、互いに独立な少数個の総合的指標で表そうとするものである。“幾つかの説明変量の分散の和”は、“情報損失量の二乗和”と“主成分の分散の和”を合わせたものに等しいと定義されている。したがって、情報の損失を少なくすることは主成分の分散を最大にすることと同じ意味を持つ。p個の変数があれば、理論的には以下に示すようにp個の主成分があり、

$$z_1=a_{11}x_1+a_{12}x_2+\cdots\cdots+a_{1p}x_p$$
$$z_2=a_{21}x_1+a_{22}x_2+\quad\quad+a_{2p}x_p$$
$$\cdot\qquad\qquad\qquad\cdot$$
$$\cdot\qquad\qquad\qquad\cdot$$
$$z_p=a_{p1}x_1+a_{p2}x_2+\quad\quad+a_{pp}x_p$$

のように示され、それぞれを第一主成分、第二主成分……第p成分と呼ぶ。主成分がp種類あれば、一つの変数に対してp種類の係数（重みをつけるもので、総合的特性を示す）があり、a_{ik}のように示される。ある主成分についてのみ考えるときはaに付ける添え字は、はじめを省略して一文字にしている。

売上高、資本金、事業損益、負債を説明変量として主成分分析を行うと

“a_1×売上高＋a_2×資本金＋a_3×事業損益＋a_4×負債”

が主成分の一つで、パソコンが算出した何種類かの説明変量の係数a_1～a_4を眺めながら、この主成分は“企業の優良さを示す総合的特性値である”というふうに判断していくのである（この場合は変数が4種類なので最大4種類の主成分あり、それぞれの係数も最大4種類ある）。投手のスピード、コントロール、試合度胸などを説明変量として、あるいは何種類かの科目の試験成績を説明変量として、主成分分析をすると、第一主成分としては投手としての総合力や総合学力と判断されるものが得られるかと思う。

p個の変数があるときはp個の主成分があることについては既に述べたが、普通は第一主成分（場合によっては第三主成分まで）だけを求めることが多い。

２．１　主成分の基本的理解と重み付けをする係数の算出法

主成分分析の基本的な理解のために２変量を例にして考える。表２－１に７つの県（A－G）の人口10万に対する博物館や資料館などの文化施設（x_1）と教育施設（x_2）数を示した。この表は今後、具体的な問題を説明するときに度々使用する。

表２－１　10万人当たりの文化施設数(x_1)と教育施設数(x_2)

県	x_1	x_2	
A	22.9	13.7	
B	24.9	16.2	
C	19.3	11.3	
D	22.0	10.4	
E	28.6	24.9	
F	42.6	26.5	
G	41.3	20.3	
平均	28.8	17.6	
分散	88.9	41.4	
相関係数	0.80	共分散	48.7

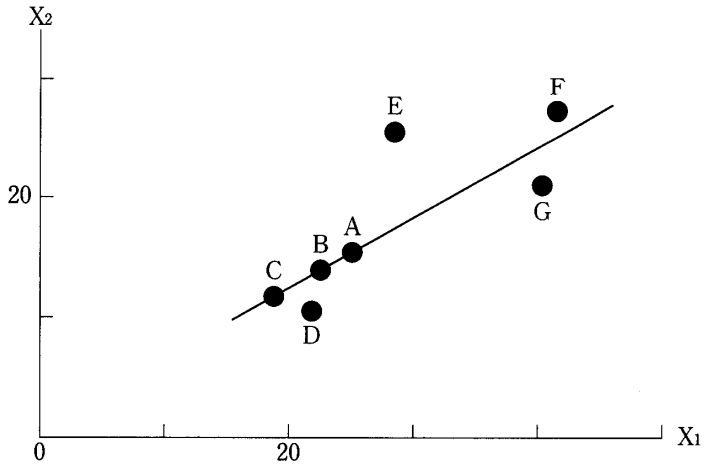

図２－１　文化施設数と教育施設数の散布図

この二つの説明変量の総合的特性は$a_1x_1+a_2x_2$という一次式で表されるのだが、問題はa_1、a_2をどう決定すればよいかである。

x_1を横軸に、x_2を縦軸にとって７個の点を描き、７点から引いた垂線の長さ（情報損失量と考えられる）の和が最も小さくなるように引かれた直線（すべての測定値が直線上にないことは、情報量の全てが主成分の中に反映されないことを意味する）が主成分なのである（図２－１）。このときの直線の傾きはa_1：a_2で、$a_1^2+a_2^2=1$という条件のもとでa_1とa_2を求めれば第一主成分が得られる。

情報損失量をa_1とa_2を含む数式で表現し、これが最小値をとるときのa_1とa_2が求めるものである。この算出には垂線の長さを示す式（ヘッセの標準形）や偏微分（いくつかの変数からなる式を、ある変数についてだけ微分することを偏微分するといい、極大、極小になるときは偏微分した式の値がゼロになる事を利用する）と呼ばれる方法を用いるのだが、詳細や具体的な演算はできなくてもかまわない。

一方、主成分の分散を最大にするときも、分散をa_1とa_2を含む式で表現し、その最大値をとるときのa_1とa_2を算出すればいいのである。これも偏微分して得られたいくつかの連立方程式を解くことで得られる。

この二つのいずれの方法を用いても、第一主成分のa_1は0.85、a_2は0.53と算出され、第一主成分は、$z_1=0.85x_1+0.53x_2$で示される。

２．２　説明変量のベクトル表示

ベクトルとは大きさと方向を持った量で、n本のベクトルはn次元空間で定義される。私たちは３次元までは、互いに直交する３本の座標軸を設定して、その方向・位置を頭に描けるが、４次元以上については通常の思考では考えることができない。

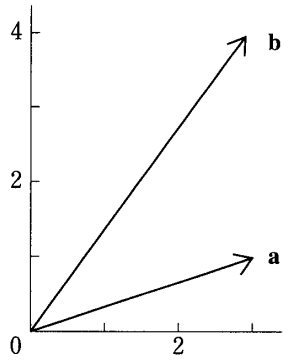

図２－２　２次元ベクトルの一例

２．２．１　ベクトルの内積とノルム

図２－２に示すように、２次元座標（平面）で（０、０）から（３、１）および（３、４）に向かう２個のベクトル a **(a)**およびベクトル b **(b)**を例にして考える。（ベクトル a は a を太字にして示すことが多い）。これらは要素が２個なので２次元ベクトルと呼ばれる。

図２－２に示した **a** と **b** の“内積”は（a、b)=(３－０)×(３－０)+(１－０)×(４－０)＝13と定義される。

同一ベクトル同士の内積の平方根は

$\|a\|=\sqrt{a, a}=\sqrt{9+1}=\sqrt{10}$、$\|b\|=\sqrt{b, b}=\sqrt{9+16}=\sqrt{25}$で示され、これらはベクトルの長さに相応しており、“ノルム”と呼ばれる。

これらのことを、多変量解析において使用できるように一般化してみる。

n 人のテスト得点 x から平均値 $\bar{x}$ を引いたものを縦一列に並べてみる。テスト y についても同様にする。これらは数学的に n 次元ベクトルと呼ばれ以下のように示され、列ベクトルとも呼ばれる（横に並べる n 次元ベクトルもあり、これは行ベクトルと呼ばれる。行列の演算において、列ベクトルと行ベクトルは異なる機能を持つので、両者をきちんと区別しなければならない）。

$$x=\begin{bmatrix} x_1-\bar{x} \\ x_2-\bar{x} \\ \\ \\ x_n-\bar{x} \end{bmatrix} \qquad y=\begin{bmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \\ \\ y_n-\bar{y} \end{bmatrix}$$

これらの内積は $\Sigma(x_i-\bar{x})(y_i-\bar{y})=(x_1-\bar{x})(y_1-\bar{y})+\cdots\cdots+(x_n-\bar{x})(y_n-\bar{y})$、
二つのノルムは $\Sigma(x_i-\bar{x})^2$ の平方根、および $\Sigma(y_i-\bar{y})^2$ の平方根で表される。

内積を n で割ったものは共分散、ノルムを$\sqrt{n}$で割ったものは標準偏差と呼ばれるものに同じである。

二つのベクトル、**x** と **y** の交わる角度の余弦、$\cos\theta$ は

$\cos\theta$ ＝(x と y との共分散)／(x の標準偏差)(y の標準偏差)

で表され、これはピアソンの積率相関係数と同じである。完全に相関する、つまり相関係数が１どうしの二つのベクトルは重なり（$\cos 0^\circ=1$）、まったく相関しない、つまり相関係数がゼロどうしの場合は直交する（$\cos 90^\circ=0$）。相関係数がゼロどうしの概念を“直交する”と表現するのはこのことによる。算術的に導かれる相関係数や標準偏差が方向を持ったベクトルでも表現できることを知ると、多変量解析の本質に迫りやすいといえるであろう。

2．3　行列

ベクトルと密接な関係があり、多変量解析を理解するのに有用な行列という数学の方法・定義がある。行列（matrix）は、m（行）×n（列）個の数を

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

の形に並べ、その算法を定義したものである。

$a_{i,k}=(i=1,2,,\ \cdots\cdots m,\ \ k=1,2,\cdots\cdots n)$ を行列の成分と呼ぶ。

同じ行（よこ）あるいは列（たて）の組をひとつのベクトルの成分と考えて、行ベクトル、あるいは列ベクトルと呼ぶことは既に示した。2．2項で示したn次元ベクトルはiかkが1の場合の行列ともいえる。行列どうしの演算は日常の算術のそれとは異なることが多く、頭の中にイメージを描けないことが多い。たとえば、和と積はある形どうしの行列にしか定義されない、行列Aと行列Bの積が0の場合に、行列A＝0または行列B＝0とは限らないことなどである。

多変量解析でよく利用される分散共分散行列および相関行列の表示法を2行2列の場合について、一般的な場合と表2－1に基づくものとに分けて、以下に示した。変量がn個であれば、n×nの数を含む行列になる（行と列の成分の数が同じものを正方行列という）。相関行列で右下がりの対角線上の数値は1と示されているが、本来はx_1とx_1およびx_2とx_2の相関係数（いずれも1）と書くべきものである。

分散共分散行列

$$\begin{bmatrix} x_1\text{の分散} & x_1\text{と}\ x_2\text{の共分散} \\ x_1\text{と}\ x_2\text{の共分散} & x_2\text{の分散} \end{bmatrix} \qquad \begin{bmatrix} 88.9 & 48.7 \\ 48.7 & 41.4 \end{bmatrix}$$

（表2－1に基づく具体例）

相関行列

$$\begin{bmatrix} 1 & x_1\text{と}\ x_2\text{の相関係数} \\ x_1\text{と}\ x_2\text{の相関係数} & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0.80 \\ 0.80 & 1 \end{bmatrix}$$

（表2－1に基づく具体例）

2．4　固有値と固有ベクトル

行列Sに対して、実数λとベクトルzが等式　$Sz=\lambda z$　をみたすとき、λをSの固有値（eigenvalue）、zを固有値λに属する固有ベクトル（eigenvector）という。固有ベクトルの成分は主成分の係数、a_1、a_2……と同じである。（ベクトルzを行列Sによって変換したものは、元のベクトルをλ倍したものになるという関係を示しており、この理解にはベク

トル・行列の演算に馴染む必要がある。数学の世界で決められたことと割きっても進んでもよい）

行列Sを$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$とすると、固有値λは、

行列式$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda)-bc=0$　　を成立させる値である。

行列Sに対して、行列式Sはad－bcのように定義され、determinantと呼ばれる。（行列と行列式は定義が異なる。3×3以上の正方行列式について、それぞれ定義されているが、ここでは基本を理解するために2×2の正方行列式についてだけ示す）

この定義に従って、表2－1に基づく分散共分散行列の固有値と固有ベクトルを求めてみる。$(88.9-\lambda)(41.4-\lambda)-48.7\times48.7=0$から$\lambda_1=119.3$および$\lambda_2=10.9$が得られる。

これは2次方程式を解いた結果で簡単にλを計算できる。2変量ではλは2つある。

最大固有値119.3（λ_1）の固有ベクトルは$\begin{bmatrix} 88.9 & 48.7 \\ 48.7 & 41.4 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=119.3\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$

から求められる。行列の算法から$88.9a_1+48.7a_2=119.3a_1$、あるいは$48.7a_1+41.4a_2=119.3a_2$が導かれる。それと重みの値に一定の条件を付ける式、$a_1^2+a_2^2=1$（重みの種類がいくつあっても、それらの二乗の和を1とする）から連立方程式を解いて

$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=\begin{bmatrix} 0.85 \\ 0.53 \end{bmatrix}$で示される固有ベクトルが得られる。

$\lambda_2=10.9$からは固有ベクトル$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=\begin{bmatrix} -0.53 \\ 0.85 \end{bmatrix}$が得られる。

これらは、第一主成分および第二主成分の係数で、既に2．1の項で示したベクトル・行列を用いないで、連立方程式を解いて得られた値と同じである。固有値119.3のとき第一主成分の係数が得られ、固有値10.9のとき第二主成分の係数が得られる。各々の係数は変数と主成分の相関の強さを示すものである　固有値の和は表2－1に示した変数x_1とx_2の分散の和に同じである。各固有値λは、その合計が変数の分散の和に等しいという条件下で、目的に叶うように再配分されたものといえる。一般的には最も大きい固有値は第一主成分に、次に大きい固有値は第一主成分と直交する第二主成分に、その次に大きい固有値は第一と第二主成分と直交する第三主成分に、というように対応している。つまり大きな分散を持つ主成分から抽出されていくのである。ここで求めた第一主成分と第二主成分が直交することは，二つのベクトルの内積（$0.85\times-0.53+0.53\times0.85$）がゼロになることから裏付けられる。このように簡単に計算できるのは変数が2個だけだからである。

一般にp個のベクトルがある場合に、それらはp次元の座標に描かれるが、ベクトル間の角度によってはpより小さい次元の空間に収められる場合がある。後に例示するが、三本の

ベクトルが2次元平面に収まることは理解できるものである。(このときこれらのベクトルは一次従属、あるいは一次独立でないといわれる。)

各変数間の分散共分散行列、あるいは相関行列の固有値と固有ベクトルを求めたとき、正の値を持つ固有値の数が合成得点の種類数を示し、それぞれの固有値に対応する固有ベクトルが、合成得点を得るために各変数に重みを与えるものとなる。固有値の数は、いくつかの変量をベクトルで表して相関を考えるとき、どの次元まで考察すればよいかを説明するものである。3個のベクトルは普通は3次元座標に描かれるが、つまり3つの固有値を持つが、そのうちの1つが0のときは2次元平面に描ける。つまり2次元で考察できるということである。(このときは三本のベクトルが特殊な相関関係にある。一般的には、ゼロないしゼロに近い固有値が算出されたときは、本来の次元数からそれに相応する固有値の数を引いた次元で考察すればよい)

2変量の場合にはこれまでに示したように，具体的な計算がパソコンに頼らなくてもできるが、3変量以上になると高次の連立方程式を解く必要があり、パソコンに頼らざるを得ない。2変量で示したことと本質的には同じ作業をパソコンはおこない、固有値や重みの係数などを算出していることを知ればよい。

固有値の理解が曖昧な人への参考：固有値λは主成分の分散（そのベクトルの大きさ）に相応する数であることに気づかれたと思う。会合でいくつかの意見があったとき、総括して一番に言いたいことが第一主成分である。それはみんなの意見をなるべく尊重して作られるべきであるが、反映されない意見もあるのが世の常である。だから第一主成分は反映されない部分（の分散）が極力小さくなるようにして得られる。反映されない部分が小さいとき(情報の損失が小さいと表現される)は大きな固有値が算出される(多くの人の意見が反映されている)。このようなときは第一主成分以外の主成分（意見）を見出すことは難しい、あるいは意味がないといえる。第一主成分の固有値がさほど大きくないときは（それ以外の意見がまだあることを意味する)、それと直交する（相関しない）成分の分散、つまり第二主成分を求める意義がある。その固有値から第二主成分の重みが判断できる。変数が多いときは、変数間の相関要因もそれだけ多くなり、変数の特質を知るためには様々な視点で相関を捉える必要があると考えられる。固有値の数は、その視点の数と考えたらどうであろう。広い景色を見るとき、その中には色々見たいものが（互に異なる範疇）あるとすれば、それらを見るのにふさわしい場所がそれぞれあるはずである。一番大きな値の固有値は、一番見たいものを見る場所に相応し、固有値の大きさは見たさの程度を示すものであろう。

2．5　説明変量がp個の場合の考え方

p個の説明変量それぞれをn人あるいはn個について調べると、p×n　個のデータが得られる。

変量／個体	x_1	x_2	…	x_p
1	x_{11}	x_{21}	…	x_{p1}
2	x_{12}	x_{22}	…	x_{p2}
⋮	⋮	⋮		⋮
n	x_{1n}	x_{2n}	…	x_{pn}

このデータから以下に示す分散共分散行列S（相関行列の場合もある）が得られる。

$$S=\begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{12} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_p^2 \end{bmatrix}$$

次に示す行列式から、Sの固有値λを求める。この固有値は正の実数をもち、$\lambda_1>\lambda_2>\cdots\cdots$である。

$$\begin{vmatrix} s_1^2-\lambda & s_{12} & \cdots & s_{1p} \\ s_{12} & s_2^2-\lambda & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_p^2-\lambda \end{vmatrix}=0$$

最後に、固有値λ_iの固有ベクトルを求める。

$$\begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{12} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_p^2 \end{bmatrix}\begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}=\lambda_i\begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}$$

これらの過程は、はじめの素データを入力すると、適当なソフトを使用すると、パソコンが次々に導いてくれる。

ただし、$a_{i1}^2+a_{i2}^2+\cdots\cdots+a_{ip}^2=1$という条件が設定されている。

こうして、最大固有値λ_1の固有ベクトルに対して、

第1主成分　$z_1=a_{11}x_1+a_{12}x_2+\cdots\cdots+a_{1p}x_p$を得る。以下同様にして

第p主成分　$z_p=a_{p1}x_1+a_{p2}x_2+\cdots\cdots+a_{pp}x_p$までが導かれる。

2．6　寄与率と累積寄与率

第i成分の固有値を固有値全部の合計で割ったものを第i成分の寄与率（proportion）、第1から第i成分までの固有値の合計を固有値全部の合計で割ったものを累積寄与率（cumula-

tive proportion）という。これらについてもパソコンの画面に表示される。なるべく少ない主成分数でデータの情報を反映できることが望ましい。累積寄与率が0.8を目安として主成分をとらえればよいといわれている。つまり、この値はどの成分まで解析したらよいかを示している。たとえば第３主成分までの累積寄与率が0.8を越したら、以後の主成分を求めることはあまり意味がないと判断するのである。

２．７　説明変量の単位を変えて入力すると分析値が異なる場合がある。

表２－１のデータは人口10万人当たりの施設数を示しているが、そのどちらかを人口１万人当たりの数で入力した場合や、身長をメートルで入力していたのをセンチメートルで入力した場合は、分散、共分散の値は異なるので、主成分を示す式も異なる。類似した内容および同じスケールどうしの変量を扱うときに限って分散共分散行列が利用できる。多くの場合は相関行列による主成分分析がなされる。相関係数は測定量のスケールを変えて入力しても同じ値になるからである。主成分分析プログラムを用いると、分散共分散行列と相関行列の両方を用いることが可能で、どちらかを選択できる。

表２－１から得られた分散共分散より第一および第二主成分が得られたが、これを相関行列より求めると、λ_1は1.80、λ_2は0.20と算出され、第一主成分は、$0.71x_1+0.71x_2$、第二主成分は、$-0.71x_1+0.71x_2$のように表現される。重み付けの係数値および寄与率は分散共分散行列より導かれたものと多少異なる。これは数学的にやむをえないことで、２変量のときはことに異なる（分散共分散行列を用いる方が詳細な情報が得られる）。後に例示する臨床検査では、項目により測定の諸内容がそれぞれ異なっており、得られた数値どうしは同じ土俵では比較できない。そのような場合はすべてが標準化された値（後述）を使うのが望ましいので相関行列を使用する。表２－１のデータはともに人口10万人あたりの施設数で、それらの数が類似していたので分散共分散行列を使えたのである。

２．８　主成分の解釈

表２－１のデータに基づく分散共分散行列から得られた第一主成分は$0.85x_1+0.53x_2$であった。この値はＦ県とＧ県（施設数が多い）で大きく、Ｃ県とＤ県（施設数が少ない）で小さい。このことから第一主成分は施設の充実度を意味していると考えられる。このときの第一主成分の寄与率は0.92＝119.3／(119.3＋10.9) で、充分１に近いので、第二主成分がなにを意味するかを考えることは、あまり意味がないといえる。

変量x_1の方が、x_2より大きな係数をもつことは、この主成分には変量x_1の方の影響が大きい（主成分との相関が大きい）ことを示している。

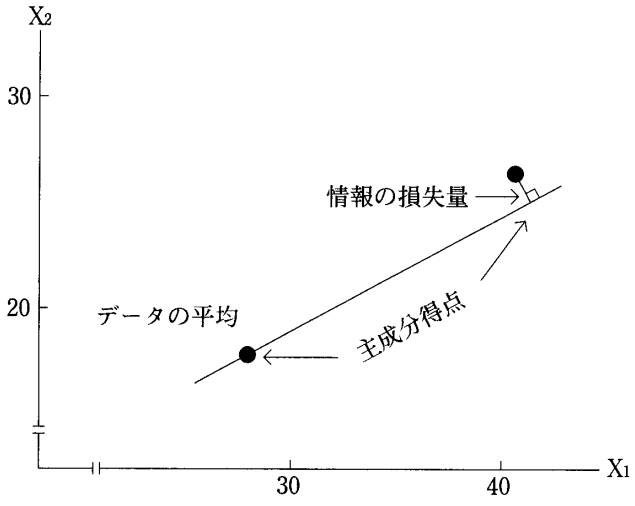

図2－3　F県の主成分得点

2．9　主成分得点

表2－1の説明変量であるx_1およびx_2の平均値の座標（28.8、17.6）は主成分を表す直線の上にあり、この点を原点とする。各県の座標点（x_1、x_2）から直線に下ろした垂線の交点と原点の距離を主成分得点（垂線の交点が原点より小さい座標にあるときは負の値）と定義する。図2－3にF県の場合について例示した。F県はこの中では主成分得点が最も大きく、主成分に示される内容を多く持つと解釈する。

ある主成分得点zは一般的には

$z=a_1x_1+a_2x_2\cdots\cdots a_px_p-(a_1\bar{x}_1+a_2\bar{x}_2\cdots\cdots a_p\bar{x}_p)$ のように示される。

肝臓病の指針といわれている何種類かの臨床検査を何人かにしてもらい、それらのデータをパソコンに入力すると、相関行列が導かれ、幾つかの主成分の係数が得られる。その中で肝臓病の重症度を意味すると考えられる主成分も得られるかと思う。その主成分得点を各個人について算出すると、各人の肝臓病態が推測できる。またその式中で、それぞれの変数にかかる係数の大きさから、その検査の肝臓病診断に対する重要性も判断できる。

3　因子分析（factor analysis）

主成分分析では、たとえば心理統計の成績をx_1、心理測定法の成績をx_2として、それらの成績にかける重みをa_1およびa_2としたときに得られる合成変数z（主成分）は、$z=a_1x_1+a_2x_2$で表され、この分散を大きくするようなa_1およびa_2を求めることで、第一主成分が得られる。この場合は第一主成分として“数理的な思考力”あるいは“人間の行動に対する理解力”が捉えられるであろう。

因子分析では変数を幾つかの共通因子f_i（後述する）と独自因子U（後述する）に分離して、それぞれに重みをつけたものの和として表す。x_1が心理統計を好む程度あるいは成績を示す変数、心理測定法に対するそれがx_2のとき、f_1を数理的な思考に対する興味・理解力を示す因子、f_2を人間の行動に対するそれとすれば、変数x_1およびx_2は、共通してf_1とf_2を含む式で表現できるはずである（一般には、多くの因子が含まれる場合を扱うが、分かりやすく説明するためにここでは2変数2因子にした）。このようなことからf_iを共通因子と呼ぶ。心理統計を教えた教員が自分にとって嫌だった場合などは、心理統計と聞くだけでいやになるもので、それが成績に影響する場合が考えられる（逆もある）。このような場合は、変数

のなかに特有な事情を反映するものとして独自因子による項を導入する必要がある。独自因子には誤差による影響も含め、重みをdと記す。以上を整理すると変数x_1が二つの因子を含むときは

$$x_1 = a_{11}f_1 + a_{12}f_2 + d_1U_1$$　　と表現される。

ここで重み付けの係数a（因子負荷、後述する）の最初の添え字は変数の番号を、次の添え字は因子の番号を示す。変数全体をまとめて考えるときは最初の添え字は省略されることが多い。

主成分を示す式zではx_1、x_2……という測定値を入力することで、分散共分散行列あるいは相関行列を求めることから、重み付けの係数aを算出できた。因子分析でも重み付けの係数を算出することが大きな課題である。独自因子の項を何らかの方法で消去し、f_iを推定できれば、主成分分析の場合と類似した方法で、これらの係数が求まることが推定できる。（主成分を示す式でのx_iと変数を因子に分解した式でのf_iが同じ関係にある）

前述した例で、“数理的な思考力”が主成分分析での主成分として、また因子分析での因子として捉えられた。主成分分析では、幾種類かの変数の測定値それぞれに適当な重みを与え、重みと測定値の積の和が、たとえば“数理的思考力の尺度を意味する”とするのである。一方、因子分析では、幾種類かの変数の測定値がどのような要素（因子）からなっているかを調べるものであるが、類似した変数を扱うときに、その因子の一つに数理的な思考力が導入されることは必然である。このように主成分分析と因子分析では同じ概念が推定・判断されることが多いので、それらの違いがよく分からないという人が多いが、それらを定義する数式を見て、その意図することを読み取れば違いが分かるはずである。数式は文章のようなもので、それらの意味を読み取ることが大事である。

3．1　因子分析の基本モデルと因子負荷

何人かの学生に、生化学、有機化学および物理化学のテストをする、あるいはそれらに対する興味の程度を質問して、それらの点数をベクトルa、ベクトルb、ベクトルcで表現し図3－1に示した。三つから二つとったベクトル間のピアソンの積率相関係数（r、相関係数）を求めたところ、r_{ab}とr_{bc}はともに$\sqrt{3}/2 = 0.87$、r_{ac}は$1/2 = 0.50$であった。これらの値はベクトルの交わる角度の余弦であるので、**a**と**b**および**b**と**c**は30度、**a**と**c**は60度で交わり、図3－1に示すように3つのベクトルが平面上に描ける。（三つのベクトルが同じ平面に描けるのは特別な場合で、以後の説明を分かりやすくするためにこうした）。このとき各ベクトルの大きさはどれも1になるように標準化されている。すでに述べたように個々の測定値の集合をベクトルで表示したとき、ベクトルの大きさは標準偏差の大きさに対応する。同じ測定をしても変数の単位のとり方が異なると、ベクトルの大きさは異なる。そのようなことに対処するには、“測定値から平均値を引いた値を標準偏差で割る”という“標準

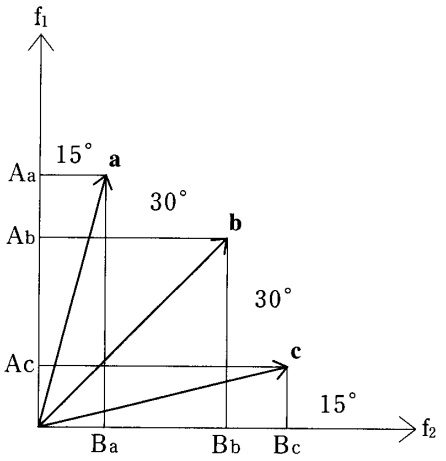

図3－1　平面上にある三つのベクトル

化”がなされた数値を用いるとよい。相関係数はデータの単位の採り方には無関係に算出され、ベクトル間の角度のみを規定するもので、座標軸の位置についての情報は与えない。そこで一例としてcと15度、aと15度で交わり、互いに直交する二つのベクトルf_1とf_2を引いてみる。（これらのベクトルの引き方は無数にある。このことは後に述べるバリマックス法を理解するための基本である。）

f_1もf_2も**a**、**b**、**c**、と同様になんらかのテストの成績、あるいは何らかに対する興味を示すものと考えられるのでそれについて考えてみる。

f_1はa、すなわち生化学の成績あるいは興味と高い相関を持つ。一方f_2は物理化学のそれと高い相関をもつ。そこでf_1は生物的なことに、f_2は物理的なことに対する理解度あるいは興味を示すものと解釈できる。（“化学”といってもその内容は大変広く、その理解や研究には多様な興味・能力が必要で、これらはどれも因子である。このことは“心理学”とだけ言う場合も同じである）。

図3－1に示したように**a**、**b**、**c**からf_1およびf_2に下ろした垂線の座標を、Aa、Ab、AcおよびBa、Bb、Bcとすれば、**a**、**b**、**c**は以下のように分解される。

$\mathbf{a}=Aaf_1+Baf_2=0.97f_1+0.26f_2$

$\mathbf{b}=Abf_1+Bbf_2=0.71f_1+0.71f_2$

$\mathbf{c}=Acf_1+Bcf_2=0.26f_1+0.97f_2$（$\cos15^\circ=0.97$、$\cos75^\circ=0.26$、$\cos45^\circ=0.71$）

Aa、とBaは、aとその二つの因子f_1とf_2に対する関連の強さを示すもので“因子負荷”とよばれる。AbとBbおよびAcとBcについても同様である。既に述べたように、因子負荷のうち第一因子に重みをつけるものはa_1、以下第n因子に対しての負荷はa_nと示すのが普通である。f_1とf_2はそれぞれの変数すべてに共通に寄与するので、“共通因子”と呼ばれる。

実際の測定値は独自因子（誤差も含む）を含んでおり、このような共通因子だけに関わる座標でのベクトル表示はできない。また二つ以上の因子が見出されることが多いので、このような2次元座標を用いた単純な方法で因子負荷を求めることはほとんどない。ここに示したものは因子負荷の意味を理解するための例である。

表3－1　5教科間の成績の相関行列

	x_1	x_2	x_3	x_4	x_5
国語：x_1	1	.60	.50	.44	.24
英語：x_2	.60	1	.44	.40	.23
社会：x_3	.50	.44	1	.30	.14
理科：x_4	.44	.40	.30	1	.52
数学：x_5	.24	.23	.14	.52	1

3．2　相関行列と因子の推定

ここで5種類の試験（国語、英語、社会、理

科、数学）を何人かにしてもらい、表３－１に示すような相関行列を得たとする。

この表に示した５個の変数間の相関係数はどれも0.14以上で、程度の差はあるが互に相関していることが分かる。その中でも国語（x_1）、英語（x_2）、社会（x_3）の成績は互いによく相関し、理科（x_4）の成績とも適度に相関しているが数学（x_5）の成績とはあまり相関していない。一方理科と数学の成績はよく相関している。このとから二つの因子が想定できる。一つの因子に対する因子負荷はx_1、x_2、x_3で大きく、もう一つの因子に対してはx_5で大きいこと、x_4では両因子がほぼ同程度の因子負荷をもつことが読み取れる。この二つ以外にもう一つの因子を想定するのは難しい。

相関行列の固有値を求めて（主成分分析と同じ方法）、１より大きい固有値の数だけを因子数にすればよいことが知られている。この行列の固有値を算出すると、３番目以降の固有値はゼロに近く、二つの因子を考えればよいことになる。このことは相関行列を見ることで推定したことと同じである。

３．３　主因子法による因子負荷の抽出

３．２の項末で述べた、表３－１に示した相関行列から固有ベクトルを求めることで、因子負荷とみなせるものが算出できるが、独自因子を考慮していないなどの条件下で得られたもので理想的な解ではない。

因子負荷を求める方法のうち、最も多く用いられるのは主因子法（principal factor method）と呼ばれているものである。主因子法においては、第一因子の寄与が可能な限り大きくなるように因子の抽出が行われる。第一因子によって最大限の分散が説明できるように（より大きな固有値が求まるように）、つまり第一因子が全体として各変数と最大限の相関をもつようにするのである。第二因子以下も、順次、それ以前の因子によって説明されていない分散を最大限説明できるように選ばれる。この方法は、主成分分析における重みをつける係数の計算とよく似ている。主因子法における各因子の負荷の算出は、相関行列の対角成分が“共通性”の推定値であることが主成分分析と異なる（主成分分析では対角成分がすべて１、つまり同じ変数間の相関係数を１とみなして計算される）。

３．３．１　共通性とその推定

変数が独自因子を含まないか、含んでも非常に小さい場合は、変数間の相関係数は、共通因子のみに基づくといえる。独自因子（誤差も含む）を含む変数間の相関係数は“見掛けの相関係数”とでもいうべきものであろう。そのような場合、変数の値のうちどれだけが変数間本来の相関に寄与するかを示しておかないと、正しい解析ができなくなる。このために導入されたのが共通性（communality）で、h^2で示される。h^2は標準化された各変数の、共通因子による部分の分散で、具体的には各因子負荷の二乗和である。d^2を独自因子の重み係数の二乗とし、共通因子と独自因子が無相関であれば、

$h^2+d^2=1$にしたがって共通性が定義されている。繰り返しになるが、変数の分散のうち、共通因子によって説明される部分の分散が共通性である。共通性が低い変数は独自性が高いと普通は解釈されるが、その変数の測定誤差の分散が大きい場合も考えられるので、解析にあたっては測定法を検討することも大事である。

共通性の推定値として“重相関係数の二乗”（square of multiple correlation coefficient, SMC）がよく使われる。重相関係数とは、幾つかの変数を合成して得られる変数と、ある一つの変数との間の相関係数で、しかもそれが最も大きくなるように合成した場合の相関係数である。SMCの値は、すべての変数間での相関行列の逆行列が求まれば計算できる。これについて詳細を知るためには重回帰分析を学ぶ必要がある。この方法を用いると、表３－１のデータから、各変数についての共通性は0.68、0.53、0.37、0.52、および0.65と算出される（表３－２参照）。

３．３．２　因子負荷の抽出と因子解のプロット

表３－１に示した相関行列の対角線上の１に替えて、３．３．１の項に示した共通性の値を入れて、主因子法を適用すると、表３－２に示すような因子負荷が得られる。（共通性を考慮に入れないで、対角線上に１を入れて計算した場合に比べて、固有値は少し小さくなり、それから導かれる固有ベクトルの値も異なる。）

因子負荷を示した行列の行の二乗和が各変数の共通性である。列の二乗和は、各因子の寄与（contribution）と呼ばれ、それぞれの因子が、変数の分散を説明する上でどの程度関わっているか、全体として変数とどの程度の相関をもっているかを示す指標である。また各因子の寄与は固有値に対応している。各因子についての寄与の和は各変数の共通性の和と同じである。

表３－２　主因子法で得られた各教科の因子負荷

変量	a_1	a_2	共通性
x_1	.78	−.28	.68
x_2	.69	−.22	.53
x_3	.55	−.25	.37
x_4	.67	.28	.52
x_5	.53	.61	.65
寄与	2.11	0.64	(2.75)

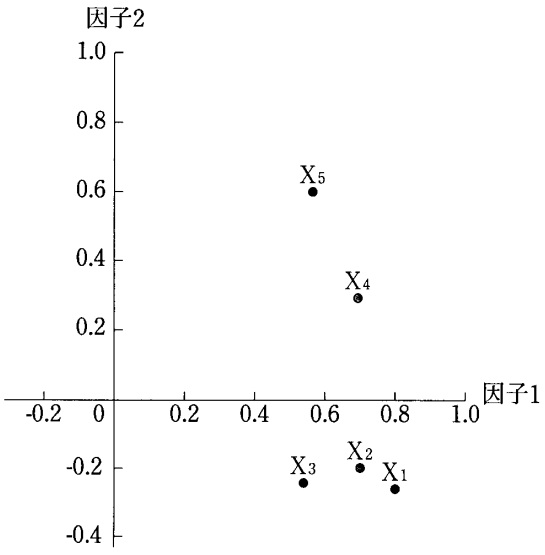

図３－２　主因子解による因子負荷の散布図

表３－２に見られるように、因子１の負荷が一様に大きく算出されるのは主因子法の特性

である。

二つの因子が直交する座標軸に因子負荷の値を各変数についてプロットした散布図を利用すると因子の解釈がしやすい。(因子が３以上の場合は２因子ずつ、すべての組み合わせの図を作って検討する)。図３－２に主因子法で得られた因子負荷の散布図を示した。

因子１に偏った分布がみられる。x_1は国語、x_2は英語、x_3は社会、x_4は理科、x_5は数学の成績である。因子１は一般的な学力を表すものと解釈されるが因子２はその解釈が難しい。一般に主因子法では第一因子以外の因子がどのような性質かを解釈しにくい。

それぞれの因子が一部の変数のみに寄与し、また、それぞれの変数も少数(できれば一つ)の因子にのみに高い負荷をもつ場合は、因子の解釈がしやすくなる。このような因子解は“単純構造”をもつといわれる。

３．３．３　バリマックス法（座標軸の回転）

３．１の項で示したように相関係数はベクトル間の交角のみを規定するもので、座標軸については自由である。図３－２のグラフを34度回転させた場合の因子負荷と、因子負荷の散布図を、表３－３および図３－３に示した。

表３－３　主因子解を34度回転させたときの因子負荷

変量	a_1	a_2	共通性
x_1	.80	.20	.68
x_2	.70	.20	.53
x_3	.60	.10	.37
x_4	.40	.60	.52
x_5	.10	.80	.65
寄与	1.66	1.09	(2.75)

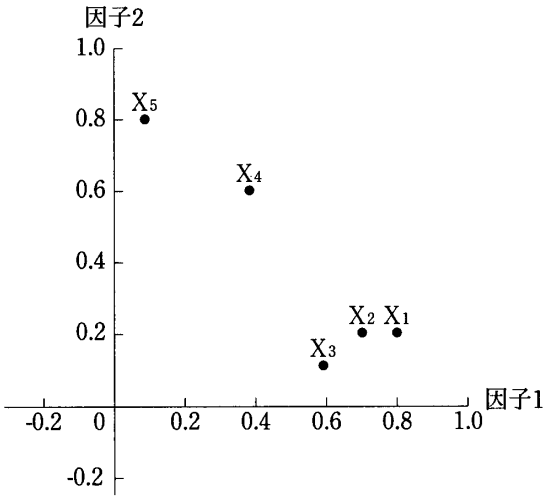

図３－３　表３－３に基づく因子負荷の散布図

座標軸を設定したベクトルを回転させると、座標軸をはじめ因子負荷の方向が変化するのでそれぞれの意味が変化する。回転後の第一因子に対する因子負荷は国語、英語、社会で大きく、第二因子に対しては数学で大きい。これから第一因子は文章理解に対する能力、第二因子は数理的な能力に加えて計算力をも示すものと推定できる。理科の成績は文章理解と数理的な能力を示す因子負荷が共に大きく、理科の理解には二つの因子が関わる事を示しており納得できるものである。座標軸を回転させることで各変数の因子負荷が分離されると（単純構造になると)、回転前に比べて、情報として好ましい因子の抽出がしやすくなる。

回転の前後で各変数について共通性の値は同じである。社会の共通性が低いのは科目の特性かもしれないが、試験の内容が適切でなかったことも要因として考えられるので、再度の

試験をして考察することが望ましい。

各因子の寄与は、主因子法による場合と回転後では異なる。主因子法では因子１と因子２の寄与の比率は3.3：1.0であるが、回転後は1.5：1.0に変化している。

単純構造への回転方法として一般的に用いられているものはバリマックス法と呼ばれるものである。最も単純な構造の指標（バリマックス基準）は、各因子における負荷の二乗の分散をすべての因子にわたって合計したものを大きくすることである。ある因子負荷の二乗の分散が、大きくなることは、その因子について、大きな負荷を持つ変数と０に近い負荷しかもたない変数とに分離されることである。したがって、そのような分散をすべての因子について合計した値を最大化すれば、因子解全体として単純構造が実現できる。

回転後に得られた各因子の負荷の二乗は、主因子法で得られたものに比べて広い範囲にある、つまり分散が大きいことが表３－２と表３－３を見比べると確認できる。このような結果を得るために座標軸の回転がなされるのである。

（AとBが正の整数で、その和が一定のとき、A^2+B^2はAとBが近い値のとき小さく、離れているときに大きい。つまり因子負荷の二乗の値がばらついているほどその分散は大きい）。ここに示した解はバリマックス法によるものとほぼ同じである。

参考：主成分分析と因子分析の相違

大まかに言えば、主成分も因子もほぼ同じもので、変数の特徴を示す。多くの変数をまとめたものが主成分、変数を分解したものが因子である。主成分分析は現象の要約的記述を目的とし、因子分析は現象の背後にある真実を探ることを目的としたものである。主成分はいくつかの変数ベクトルの一次結合ベクトルとして表現でき、数学的には行列の固有値・固有ベクトルを求めることに帰着されるが、因子分析では独自因子が導入されるので共通因子は変数ベクトルの一次結合ベクトルとして表現できない。因子分析モデルに独自因子が導入されたのは、抽出される共通因子の数を、変数の数に比べてできる限り少なくしたいためである。

4　食品の嗜好調査結果に対する主成分分析および因子分析

少し古くなるが、1970年に日本人の食生活で重要と考えられる100種類の食品に対する大規模な嗜好調査が行われた。この調査データに基づいて、主成分分析および因子分析の具体例を示す。

食品としては、主食、副食、飲み物、菓子、果物の中から代表的なものが100選ばれた。このような調査対象の選択は実験結果をより良いものにするために大変重要である。

調査対象は15歳以下（男１、女６番）、16－20歳（男２、女７番）21－30歳（男３、女８番）31－40歳（男４、女９番）41歳以上（男５、女10番）の10グループに分けられた。それぞれのグループは100人から成る。現在（2003年）のような高齢化社会では、これに加えて

老齢者のグループを、最低１つは設定する必要があるであろう。

嗜好を測定する尺度として、

“最も好きな食品である”　　を９点、
“いつも食べたい”　　を８点、
“機会があればいつも食べたい”　を７点、
“好きだから時々食べたい”　　を６点、
“時には好きだと思うことがある”を５点、
“たまたま手に入れば食べてみる”を４点、
“ほかに何もないときに食べる”　を３点、
“もし強制されれば食べる”　　を２点、
“おそらく食べる気にならない”　を１点とした。

このような点数化、つまり嗜好の程度を数値で示すことをしないと統計的な取り扱いができない。点数化の手順は、よく吟味して計画しないと、良い解析結果が導かれないので統計処理以上に重要である。

調査結果の集計は、行ごとに食品名を100行、列ごとに性・年齢別のグループを10列にして、そこに平均嗜好度を示すのが妥当である。一例として、ご飯、煮魚、バナナについての結果を以下に示す。

食品名＼グループ	1	2	3	4	5	6	7	8	9	10
1. ご　飯	7.69	7.31	7.47	7.76	7.87	7.51	7.24	7.70	7.91	7.95
：　：										
40. 煮　魚	3.84	3.84	4.47	4.29	5.37	3.97	3.73	3.88	5.08	5.50
：　：										
100. バナナ	8.29	7.45	7.00	6.76	6.69	8.14	7.09	6.84	6.83	7.13

10グループの平均嗜好度を10個の変数として、各変数の平均および分散を表４－１に、変数間の100の食品に対する相関行列（右半分は対称なので省略）を表４－２に示した。

一般的に予想されることであるが、表４－１からは、年齢の若い層では分散・平均が大きく、同じ年齢層では女性グループのほうで平均が小さく分散が高い傾向が見られる。表４－２からは男女ひっくるめて年齢層が近いどうしでは相関が高い傾向が見られる。

表4－1
性・年齢別に分けたグループの食品嗜好度の平均と分散

グループ	平均	分散
1	6.038	1.536
2	5.785	1.069
3	5.947	0.682
4	5.670	0.838
5	5.641	0.782
6	5.781	1.675
7	5.564	1.399
8	5.378	1.259
9	5.517	1.033
10	5.542	1.279

表4－2　グループ間の食品嗜好に対する相関行列

	1	2	3	4	5	6	7	8	9	10
1	1.000									
2	.871	1.000								
3	.516	.759	1.000							
4	.370	.604	.852	1.000						
5	.172	.402	.726	.874	1.000					
6	.938	.821	.517	.358	.208	1.000				
7	.811	.838	.658	.488	.354	.889	1.000			
8	.615	.709	.698	.620	.523	.746	.894	1.000		
9	.500	.647	.701	.721	.710	.621	.768	.852	1.000	
10	.330	.457	.558	.632	.748	.493	.642	.778	.911	1.000

4．1　食品嗜好調査の主成分分析

表4－2の相関行列の固有値と固有ベクトル（主成分の係数）を表4－3に示す。

表4－3　表4－2の相関行列より算出された主成分の係数と固有値

グループ ＼ 主成分の係数	a_1	a_2	a_3
1	0.286	0.446	0.194
2	0.331	0.240	0.336
3	0.323	−0.166	0.442
4	0.299	−0.359	0.375
5	0.261	−0.507	0.128
6	0.309	0.408	−0.084
7	0.344	0.253	−0.171
8	0.348	0.032	−0.290
9	0.346	−0.164	−0.322
10	0.303	−0.267	−0.522
固有値	6.83	1.76	0.75
寄与率	0.683	0.176	0.075
累積寄与率	0.683	0.859	0.934

この表をよく見ると、第一主成分は性・年齢に無関係な全般的な嗜好度を表し、第二主成分は年齢による嗜好の違いを表すものと解釈される。第三主成分は男女の嗜好の違いを表すものと解釈される。第三主成分までで累積寄与率は93.4％に達しているのでこれ以上の主成

分を取り上げていない。

4．1．1　第一主成分と第二主成分が直交する座標での各食品の得点

ゼロを交点として、横軸に第一主成分の意味する“一般的な食品の好み”をとり、縦軸に第二主成分の意味する“年齢による嗜好の違い”をとる。そのとき一般に好まれる方と若い人に好まれる方を正にとることにする（図４－２）。

ご飯、煮魚、バナナについては、10グループの平均嗜好度を既に示した。また各グループにおける、すべて100種類の食品に対する嗜好の平均値は表４－１に示した。

ここでは“ご飯”を例にして、その座標について考えてみる。各グループにおいて、ご飯に対する得点（調査した結果の平均）から表４－１に示した全食品に対する嗜好の平均値を引く（いずれのグループでも正数となり、ご飯が好まれていることが分かる）。次にそれら10の値に、表４－３に示した各グループについて算出された主成分の係数をそれぞれかけて、

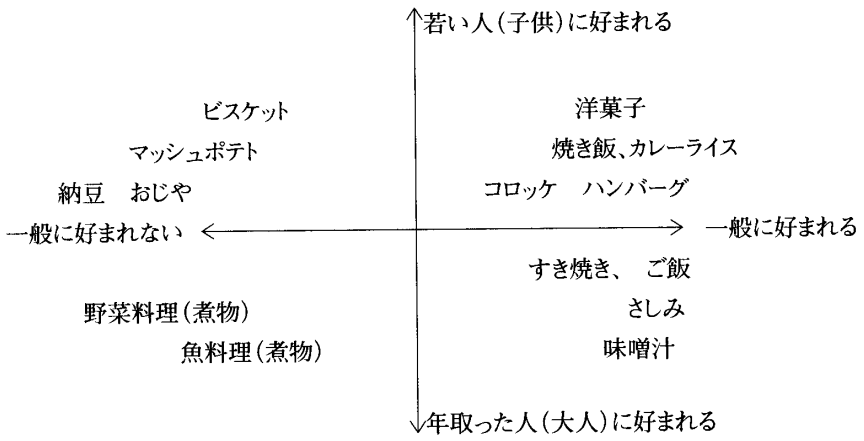

図４－２　第一主成分－第二主成分平面における食品の位置

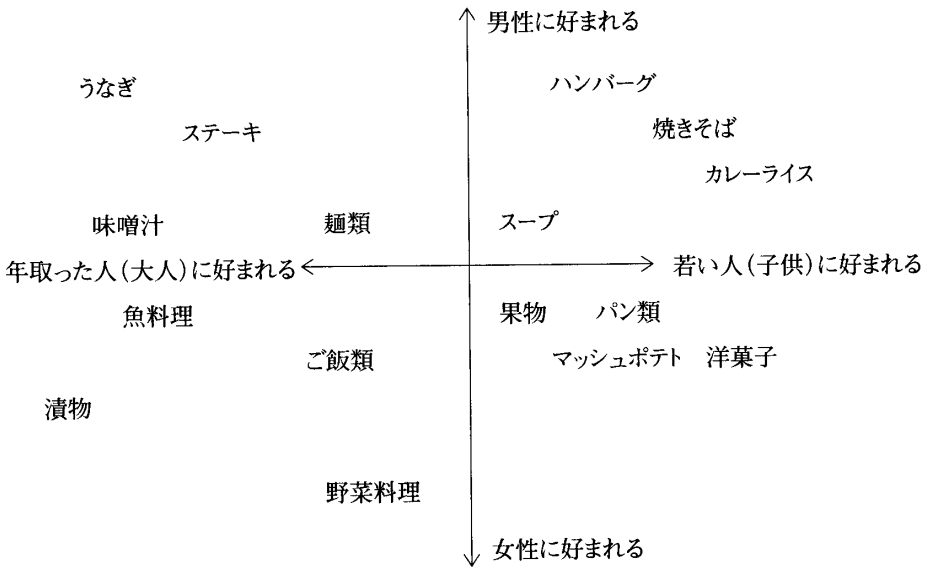

図４－３　第二主成分－第三主成分平面における食品の位置

それら10個を合計する。第一主成分についてはほぼ6、第二主成分については−0.7となる。これをプロットした点は座標軸で4つに分けられた右側におおきく、やや下に入る（第四象限)。このことは“ご飯”が一般に好まれ、どちらかというと年齢の大きい人に好まれることを示すものである。ちなみに第一象限（右上）にはコロッケ、やきめし、洋菓子など、第二象限（左上）にはビスケット、納豆、マッシュポテトなど、第三象限（左下）には幾つかの野菜料理、魚料理など、第四象限には前述のご飯、味噌汁、刺身などが入っていた。

これと同じことを第二主成分と第三主成分についてもやってみたものを図4－3に示した。これらを3次元の空間における位置にまとめると、色々な食品が“一般に好まれるか”、“大人に好まれるか子供に好まれるか”、“男性に好まれるか女性に好まれるか”ということが一目で分かる。

4．2　食品嗜好調査の因子分析

因子数を決めるとき、相関行列から求めた固有値のうち1以上のものが幾つあるかがその目安となることを既に示した。λ_1からλ_3までは、表4－3に示したように、6．83、1．76、0．75であった。さらにλ_4は0.26、λ_5は0.12である。λ_3は1に近くλ_4との間に大きな落差が見られるので因子として捨てがたい。

次に行列の対角要素に共通性の値を入れた場合の固有値は、λが1から7までは、それぞれ6.74、1.69、0.66、0.18、0.035、0.0024、−0.013のようになる。この場合は固有値が正というのが因子にするかどうかの基準であるが、ここでもλ_4以下は0に近い。そこで因子数を3にして、因子負荷量を、主因子法で得られたものと、それをバリマックス回転したものについて表4－4に示した。

表4－4　食品嗜好調査のデータより主因子法およびバリマックス法で得られた因子負荷

グループ	主因子法			バリマックス法			共通性
	a_1	a_2	a_3	a_1	a_2	a_3	
1	0.74	−0.57	0.15	0.94	0.14	0.09	0.90
2	0.86	−0.31	0.27	0.84	0.43	0.13	0.91
3	0.83	0.21	0.34	0.45	0.77	0.22	0.85
4	0.78	0.47	0.33	0.23	0.90	0.28	0.94
5	0.67	0.65	0.11	0.00	0.82	0.45	0.88
6	0.81	−0.53	−0.07	0.92	0.07	0.31	0.94
7	0.90	−0.33	−0.14	0.82	0.20	0.48	0.94
8	0.90	−0.04	−0.21	0.60	0.34	0.62	0.86
9	0.90	0.22	−0.26	0.40	0.47	0.74	0.93
10	0.80	0.36	−0.45	0.20	0.40	0.87	0.97
寄与	6.74	1.69	0.66	3.92	2.81	2.36	

主因子法による第一因子の負荷量はいずれのグループでも正で0.8前後の値であるので、性・年齢に無関係に好まれるかどうかを示す因子と解釈される。第二因子の負荷量は男女を問わず年齢の若い層で負、高年齢で正であるから、嗜好の年齢差を表す因子と解釈される。第三因子の負荷量は、男性で正、女性では負であるから、嗜好の男女差を示す因子と解釈される。これらの因子の意味するものは主成分分析による主成分の意味と同じである。

バリマックス回転を適用すると、第一因子は性を問わず、年齢20歳以下の負荷量が大きいので、若年層の嗜好の程度を示す因子と考えられる。第二因子は男性で21歳以上の負荷量が大きいので、大人男性の嗜好の程度を示す因子と解釈される。第三因子は女性で21歳以上の層の負荷量が大きいので、大人女性の嗜好の程度を表す因子と解釈される。

主因子法では、主成分分析の主成分の意味と類似した因子が抽出されたことを前述したが、それは係数の算出法が類似しているからである。一方バリマックス展開後の因子の解釈は主因子法のそれとは異なる。バリマックス展開後の解釈の方が有用であることが多いが、主因子法による解釈も１つの参考になると思う。

主成分分析の項で示した方法と同じようにして、因子の得点を因子を示す座標にプロットすることで各食品の好まれ方が分かる。各食品の因子得点も、主成分の場合と同様に求められる。第三因子まで取り上げればよい場合は、主成分分析の場合で示したように平面座標を３種類作って検討すればよい。また三次元空間を作って、各食品がどこに位置するかを見れば、その好まれ方が分かる。

文　　献

1）田中豊、脇本和昌：多変量統計解析法、現代数学社（京都）(1983)
2）石村貞夫、有馬哲：多変量解析のはなし、東京図書（東京）(1987)
3）渡部洋：多変量解析入門、福村出版（東京）(1988)
4）柳井晴夫、岩坪秀一：複雑さに挑む科学、講談社（東京）(1976)

(2004年１月31日受理)