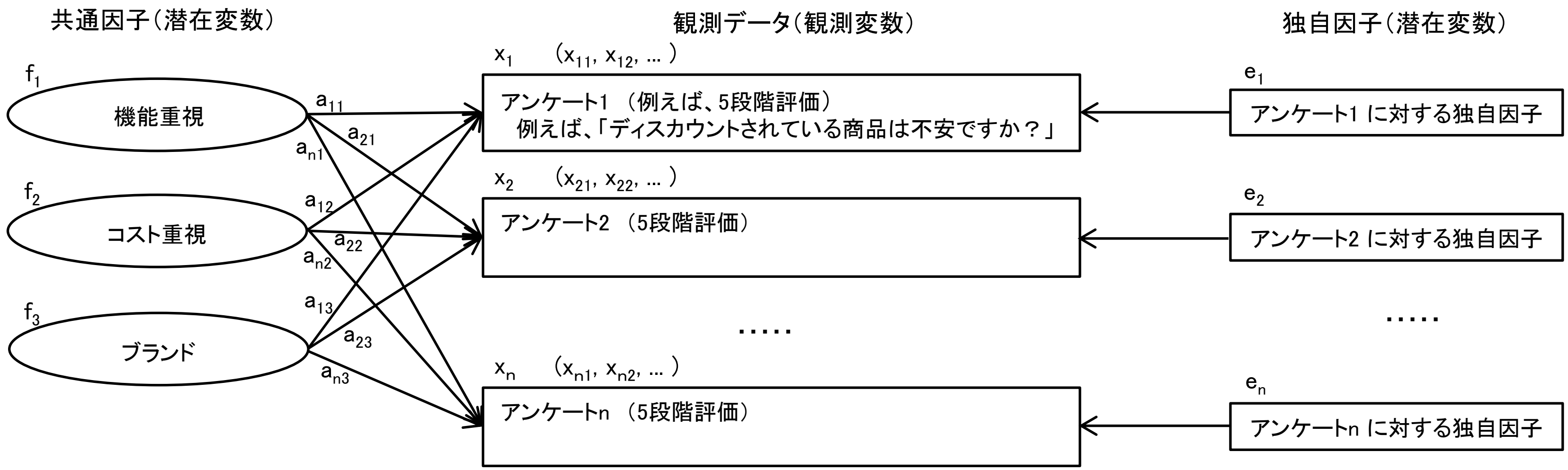
因子分析の手順

2021年 ver. 1.1

倉谷 隆博

11－1. 因子分析とは

因子分析（Factor Analysis）は、沢山の観測データの背後には観測データ（結果）を引き起こす何らかの複数の因子（原因）が潜んでいると捉え、これらの潜在因子をあぶりだそうとする分析方法である。潜在因子は観測変数間の相関関係の強弱に影響与えていると考え、相関関係を分析して潜在因子を見つけ出す。そして、見つけ出した潜在因子と観測変数の間の関係を明らかにする。つまり、因子分析では、潜在因子を使って観測変数の挙動を説明する。
因子分析は、例えば化粧品の購入意識アンケートの結果をもとに、何を（どんな潜在因子を）重視して化粧品を選んでいるのか、機能なのか、コストなのか、それともブランドなのか、などの潜在因子を明らかにする目的で使用される。潜在因子の名前は、分析者が因子分析の結果を見て付ける。潜在因子によって消費者を分類し、その結果に基づいて店舗ごとの販売方法を変える、というようなことが可能になる（マーケティング分野での活用事例）。



（図１）因子分析の概要

（図１）に因子分析の概要を示す。観察データ $x_1, x_2, \cdots, x_n$ は潜在因子（共通因子） $f_1, f_2, \cdots$ からの影響を受けており、また、それぞれ独自の因子 $e_1, e_2, \cdots, e_n$ からの影響も受けている、というモデルを考える。この関係は（式１）で表すことができる。観測データ = 共通性 + 独自性であると捉える。因子分析では（式１）の係数 $a_{11}, a_{21}, \cdots$ を観測データ間の相関関係から求める。その上で、共通因子 f1, f2, ・・・ が何を表すのかを考察する。

$$\left.\begin{aligned} x_1 &= a_{11}f_1 + a_{12}f_2 + a_{13}f_3 + e_1 \\ x_2 &= a_{21}f_1 + a_{22}f_2 + a_{23}f_3 + e_2 \\ &\cdots \\ x_n &= a_{n1}f_1 + a_{n2}f_2 + a_{n3}f_3 + e_n \end{aligned}\right\} \cdots \text{（式１）}$$

なお、因子分析を一般化したものは、構造方程式モデリング（SEM : Structural Equation Modeling）と呼ばれる。この手法は、統計モデルに多くのパラメータを持たせ、パラメータの最適値を最尤法（後述）で求めていく。

11－2. 因子分析の進め方

手順1）（式1）の係数a_{11}, a_{12}, ・・・ を、観測データ x_1, x_2, ・・・ , x_n から求める。
係数a_{11}, a_{12}, ・・・ を因子負荷量と呼ぶ。因子負荷量を算出（抽出）する代表的な方法を要約して挙げる。

（1）最尤（さいゆう）推定法

統計モデル（分布の形）を仮定したとき、その統計モデルから得られるデータの生起確率は統計モデルのパラメータ（平均などの母数）の式で表すことができる。この式を尤度（ゆうど）関数という。
最尤推定法は、データが生起される確率が最大になるときの統計モデルの母数を推定する方法である。
すなわち、尤度関数が最大になるときが、その統計モデルの母数の最適値であるとする推定法である。
母数の最適値は尤度関数の微分値が 0 になるときの値として求める。これを因子分析に当てはめれば、
因子負荷量を最適化すべきパラメータとして尤度関数に組み込んで、求めるということになる。

（2）主因子法

下記の前提条件のもとで（式2）が成立する。また、観測変数の相関行列 R は固有値と固有ベクトルを使って（式3）のように、ベクトルの積の和で表すことができる（相関行列のスペクトル分解という）。
（式2）＝（式3）より、相関行列の固有値と固有ベクトルから因子負荷行列 A を求めることができる。

$\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathrm{T}} + \mathbf{E}$　・・・　（式2）

$\mathbf{R} = \Sigma\, \lambda_i \boldsymbol{\nu}_i \boldsymbol{\nu}_i^{\mathrm{T}}$　・・・　（式3）

R ：　標準化した観測データの相関行列（分散共分散行列）
A ：　因子負荷行列　（$^{\mathrm{T}}$は転置を表す）、　E ：　独自因子の分散行列
λ：　R の固有値、　$\boldsymbol{\nu}$：　R の固有ベクトル（$^{\mathrm{T}}$は転置を表す）

前提条件：　共通因子は標準化されている（f_1, f_2, ・・・　の平均 ＝ 0, 分散 ＝ 1）、　独自因子の平均は 0
異なる2つの因子は独立している、つまり f_1 と f_2, f_1 と e_1, e_1 と e_2 などの共分散 ＝ 0

手順2）観測変数ごとの因子負荷量を散布図にプロットして共通因子の持っている意味を解釈する。この解釈の助けになるよう、散布図の軸まわりに因子負荷量のプロットが集まるように散布図の軸を回転させる。
共通因子の間に相関がないと仮定して軸の直交を保ったまま軸を回転させるバリマックス法、共通因子の間に相関があると仮定して複数の軸をいろいろな角度で回転さるプロマックス法などが使用される。

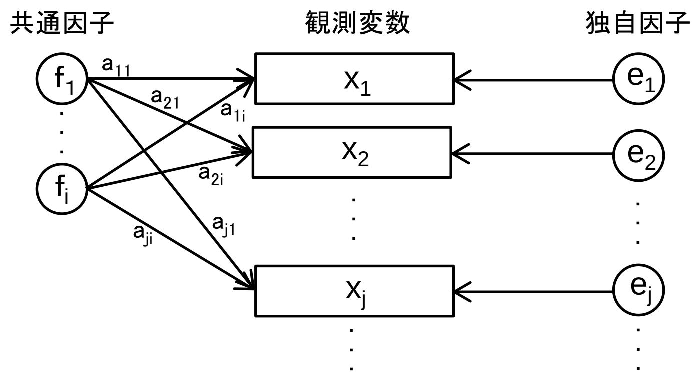
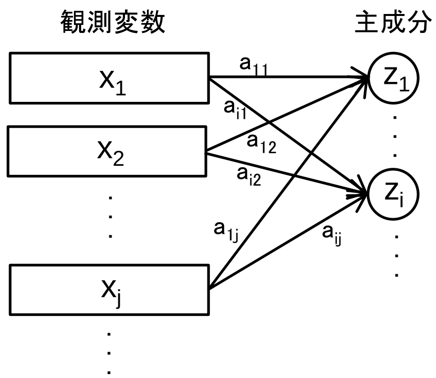
手順3）因子負荷量のプロットをもとに、共通因子（潜在変数）の持つ意味を考察し名前を付ける。

11－3. 因子分析と主成分分析との比較

因子分析の結果と主成分分析の結果はよく似ていると言われる。主成分分析で得られた主成分をもとに観測変数を説明するという取り組みを行えば、つまり主成分分析の方向を主成分から観測変数を見るようにすれば、それは、まさしく因子によって観測変数を説明するという構造に似てくる。また、いずれの分析においても観測変数間の相関をもとに共通因子と観測変数、主成分と観測変数との間の関係を明らかにする分析手法である。

しかし、分析の「思想」という視点に立てば大きな違いがある。つまり、因子分析は、原因を探る、構造を知るための分析であり、一方、主成分分析は情報を圧縮することで変数の数を少なくして分析しやすくするための方法である。

（表１）因子分析と主成分分析の違い

<table>
<tr><th></th><th>因子分析</th><th>主成分分析</th></tr>
<tr><td>模式図</td><td>
</td><td>
</td></tr>
<tr><td>構造式</td><td>$x_j = a_{j1}f_1 + a_{j2}f_2 + \cdots + a_{ji}f_i + e_j$
x： 観測変数
a： 因子負荷量（構造式の係数）
f： 共通因子
e： 独自因子（誤差は独自因子に含む）</td><td>$z_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ij}x_j$
x： 観測変数
a： 主成分負荷量（構造式の係数）
z： 主成分（誤差を含む）
制約条件 ： $a_{i1}^2 + \cdots + a_{ij}^2 = 1$</td></tr>
<tr><td>思想</td><td>観測データという結果の背後にある共通の原因を探し出すいわば、「結果」から「原因」にさかのぼる分析手法</td><td>観測データという「結果」をもとに、観測変数をいくつかの変数にまとめる（主成分に合成し要約する）分析手法</td></tr>
<tr><td>構造式の係数の求め方</td><td>観測変数の相関係数をもとに、最尤（さいゆう）法、主因子法などを使って因子負荷量を求める</td><td>観測変数の相関係数をもとに、固有ベクトルを求めて座標変換を行うための主成分負荷量を求める</td></tr>
</table>

付録. 因子分析の手順　1/3

ここでは、主因子法と直交回転法を使用した因子分析の手順を、（付表１）の観測データを使用して示す。
なお、統計解析ソフト SPSS などを使った因子分析では、適用範囲が広い最尤法と斜交回転法がよく使用される。

手順１）（付表１）の観測データを標準化して（付表２）を得る。　標準化データ = （観測データ　−平均値）/ 標準偏差

（付表１）準備した観測データ

	A	B	C	D	E	F	G	H	I	J
x_1	57.0	22.0	47.0	89.0	46.0	63.0	43.0	77.0	61.0	53.0
x_2	80.0	47.0	87.0	77.0	53.0	79.0	51.0	64.0	80.0	65.0
X_3	45.0	65.0	68.0	46.0	62.0	55.0	46.0	57.0	80.0	73.0

（付表２）標準化した観測データ

	A	B	C	D	E	F	G	H	I	J
x_1	0.1	-1.8	-0.5	1.8	-0.5	0.4	-0.7	1.1	0.3	-0.2
x_2	0.8	-1.5	1.3	0.6	-1.1	0.8	-1.2	-0.3	0.8	-0.2
x_3	-1.2	0.4	0.7	-1.1	0.2	-0.4	-1.1	-0.2	1.7	1.1

手順２）（付表2）の分散共分散行列（標準化されているので相関行列になる）を求める。

S_{x1x1} = (Σ(x_{1i} − x_{1i} の平均)²) / (10 −1) = 1.0　同様に、S_{x2x2} = 1.0,　S_{x3x3} = 1.0

S_{x1x2} = S_{x2x1} = (Σ(x_{1i} − x_{1i} の平均)(x_{2i} − x_{2i} の平均) / 10 = 0.4842,　同様に、S_{x2x3} = S_{x3x2} = 0.0627,　S_{x3x1} = S_{x1x3} = −0.2645

よって、相関行列 R は対称行列で（付式１）のようになる。

$$\mathbf{R} = \begin{bmatrix} S_{x1x1} & S_{x1x2} & S_{x1x3} \\ S_{x2x1} & S_{x2x2} & S_{x2x3} \\ S_{x3x1} & S_{x3x2} & S_{x3x3} \end{bmatrix} = \begin{bmatrix} 1 & 0.4842 & -0.2645 \\ 0.4842 & 1 & 0.0627 \\ -0.2645 & 0.0627 & 1 \end{bmatrix} \quad \cdots \text{（付式１）}$$

手順３）固有方程式（付式２）を（付式３）に変形し、（付式３）から導かれる行列式（付式４）を解いて、固有値 λ を求める。

$$\mathbf{R}\boldsymbol{\nu} = \lambda \boldsymbol{\nu} \quad \cdots \text{（付式２）}$$

$$(\mathbf{R} - \lambda)\boldsymbol{\nu} = 0 \quad \cdots \text{（付式３）}$$

$$\begin{vmatrix} S_{x1x1}-\lambda & S_{x1x2} & S_{x1x3} \\ S_{x2x1} & S_{x2x2}-\lambda & S_{x2x3} \\ S_{x3x1} & S_{x3x2} & S_{x3x3}-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & 0.4842 & -0.2645 \\ 0.4842 & 1-\lambda & 0.0627 \\ -0.2645 & 0.0627 & 1-\lambda \end{vmatrix} = \lambda^3 - 3\lambda^2 + 2.6916\lambda - 0.6756 \quad \cdots \text{（付式４）}$$

（付式４）の３次方程式は、右記のサイトを使用して解いた。　三次方程式の解 - 高精度計算サイト (casio.jp)

また、（付式１）から、翔泳社の Excel のアドインソフトであるビジネス統計の数学付録、固有値を使用して求めることもできる。
求めた固有値は、値の大きな順に以下の通りである。

$$\left.\begin{array}{l} \lambda_1 = 1.5272 \\ \lambda_2 = 1.0526 \\ \lambda_3 = 0.4203 \end{array}\right\} \quad \cdots \text{（付式５）}$$

付録. 因子分析の手順　2/3

手順4）（付式2）を使って固有ベクトルを求める。

・固有値 λ_1 = 1.5272 のとき

$$\begin{bmatrix} 1 & 0.4842 & -0.2645 \\ 0.4842 & 1 & 0.0627 \\ -0.2645 & 0.0627 & 1 \end{bmatrix} \begin{bmatrix} \nu_{11} \\ \nu_{12} \\ \nu_{13} \end{bmatrix} = 1.5272 \begin{bmatrix} \nu_{11} \\ \nu_{12} \\ \nu_{13} \end{bmatrix} \quad \cdots \text{（付式6）}$$

（付式6）を解いて（付式7）が求まり、ベクトルの大きさが 1 になるように再計算すると固有ベクトル（付式8）が求まる。

$$\left.\begin{aligned} \nu_{11} &= -2.5110 \\ \nu_{12} &= -2.1875 \\ \nu_{13} &= 1.0 \end{aligned}\right\} \cdots \text{（付式7）} \qquad \left.\begin{aligned} \nu_{11} &= -0.7222 \\ \nu_{12} &= -0.6291 \\ \nu_{13} &= 0.2876 \end{aligned}\right\} \cdots \text{（付式8）}$$

・固有値 λ_2 = 1.0526 のとき

$$\left.\begin{aligned} \nu_{21} &= -0.0624 \\ \nu_{22} &= 0.4732 \\ \nu_{23} &= 0.8787 \end{aligned}\right\} \cdots \text{（付式9）}$$

・固有値 λ_3 = 0.4203 のとき

$$\left.\begin{aligned} \nu_{31} &= 0.6889 \\ \nu_{32} &= -0.6167 \\ \nu_{33} &= 0.3810 \end{aligned}\right\} \cdots \text{（付式10）}$$

手順5）固有値と固有ベクトルを使って因子負荷量を求める。

そのために、本文の（式2）と（式3）の式の構成の類似性に着目した近似式である（付式11）を使って求める。
ただし、固有値は大きな値をとる λ_1 と λ_2 の2つに絞る。つまり因子数を2つにして因子負荷量を求める。

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} \fallingdotseq \lambda_1^{0.5} \begin{bmatrix} \nu_{11} \\ \nu_{12} \\ \nu_{13} \end{bmatrix}, \qquad \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \fallingdotseq \lambda_2^{0.5} \begin{bmatrix} \nu_{21} \\ \nu_{22} \\ \nu_{23} \end{bmatrix} \quad \cdots \text{（付式11）}$$

$$\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} = (1.5272)^{0.5} \begin{bmatrix} -0.7222 \\ -0.6291 \\ 0.2876 \end{bmatrix} = \begin{bmatrix} -0.8924 \\ -0.7775 \\ 0.3554 \end{bmatrix}, \qquad \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} = (1.0526)^{0.5} \begin{bmatrix} -0.0624 \\ 0.4732 \\ 0.8787 \end{bmatrix} = \begin{bmatrix} -0.0640 \\ 0.4855 \\ 0.9015 \end{bmatrix} \quad \cdots \text{（付式12）}$$

よって、因子負荷量行列 A は（付式13）で表される。

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} -0.8924 & -0.0640 \\ -0.7775 & 0.4855 \\ 0.3554 & 0.9015 \end{bmatrix} \quad \cdots \text{（付式13）}$$

付録. 因子分析の手順　3/3

手順6）本文（式１）に（付式１３）の因子負荷量を代入して、モデル式（付式１４）を作成することができる。

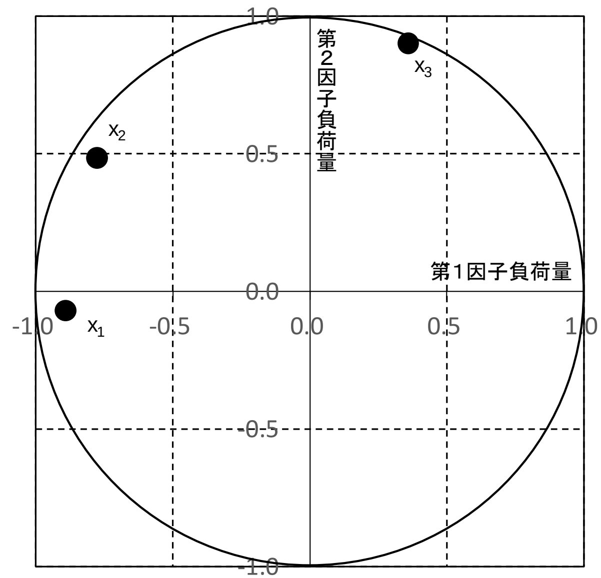
$$\left.\begin{aligned} x_1 &= -0.8924\, f_1 - 0.0640\, f_2 + e_1 \\ x_2 &= -0.7775\, f_1 + 0.4855\, f_2 + e_2 \\ x_3 &= 0.3554\, f_1 + 0.9015\, f_2 + e_2 \end{aligned}\right\} \quad \cdots \quad \text{（付式１４）}$$

手順7）モデル式（付式１４）の観測変数ごとの因子負荷量をレーダーチャート上にプロットする（２因子の場合は容易に見える化できる）。

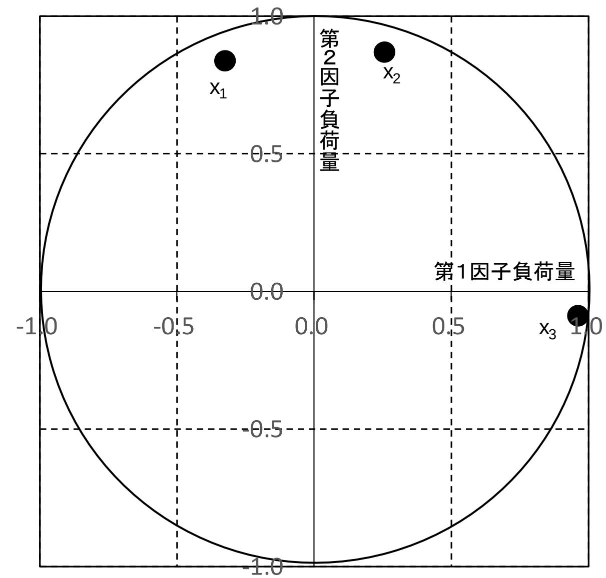
ついで、この結果をもとに、第１因子負荷量の軸と第２因子負荷量の軸が何を意味しているのか、観測変数のプロット位置をもとに解釈する。第１因子負荷量は第１共通因子に、第２因子負荷量は第２共通因子に、それぞれ働く係数である。

しかし、（付図１）の x_2 のように軸から離れた位置にプロットされている観測変数がある場合は、軸の意味の解釈が難しい。そこで、プロットされる観測変数ができるだけ軸の周りに集まるように原点を固定して回転する。その結果が（付図２）である。この図から、第１因子負荷量は観測変数 x_3 のプラス方向への影響が大きな意味を持つ変数である、また、第２因子負荷量は観測変数 x_1 と x_2 のプラス方向への影響が大きな意味を持つ変量であると解釈できる。

これらの因子負荷量の解釈をもとに、それぞれの共通因子の意味を考えて、潜在因子としての共通因子に名前を付ける。



（付図１）因子負荷量軸を使ったレーダーチャート



（付図２）回転させたレーダーチャート

参考にした書籍

河本薫：「会社を変える分析の力」、 講談社現代新書
小山昇：「数字は人格」、 ダイヤモンド社
永野裕之：「ビジネス×数学＝最強」、 すばる舎
竹内薫：「数学×思考＝ざっくりといかにして問題をとくか」、 丸善出版
中西達夫：「統計データをすぐに分析できる本」、 アニモ出版
中西達夫：「すぐれた判断が統計データ分析から生まれる」、 実務教育出版
豊田裕貴：「マンガでわかる　ビジネスを成功に導くデータ分析」、 ナツメ社
向後千春、冨永敦子：「統計学がわかる」、 技術評論社
石井俊全：「意味がわかる統計学」、 ベレ出版
涌井良幸、涌井貞美：「中学数学でわかる統計の授業」、 日本実業出版
涌井良幸、涌井貞美：「統計学の図鑑」、 技術評論社
西内啓：「統計学が最強の学問である」、 ダイヤモンド社
西内啓：「統計学が最強の学問である（実践編）」、 ダイヤモンド社
西内啓：「統計学が最強の学問である（ビジネス編）」、 ダイヤモンド社
森岡毅、今西聖貴：「確率思考の戦略論」、 角川書店
デビッド・マクアダムス：「世界一流企業はゲーム理論で決めている」、 ダイヤモンド社
河村真一ほか：「本物のデータ分析力が身に付く本」、 日経BPムック
末吉正成、末吉美貴：「Excel　ビジネス統計分析　この分析できますか？」、 翔泳社
谷岡一郎：「社会調査のウソ　リサーチ・リテラシーのすすめ」、 文藝春秋
林知己夫：「調査の科学」、 ちくま学芸文庫
八谷大岳：「データ解析」シリーズ　全１５回、（例）データ解析 第1回 ベクトルの復習 – YouTube