

第9章 因子分析

§9-1 因子分析とは

- ・ 因子分析(factor analysis)

実験や観測によって得られた「観測変数」の背後に存在する「因子」を推定する統計的分析手段。

- ・ 観測変数(observed variable)

実験や観測を通して得られたデータ(観測値)。

- ・ 因子(factor)

得られた観測変数に対し影響を及ぼしている、一見すると表には出て来ていない潜在的な要因のこと。潜在変数(latent variable)とも言われる。

- ・ 共通因子(common factor)

複数の観測変数の背後にあり、影響を与える因子のこと。共通因子によって説明される観測変数の分散のことを、共通分散(common variance)といい、その値を共通性(communality)という。

- ・ 独自因子(unique factor)

1つの観測変数にのみ影響を与えている因子のこと。独自因子によって説明される観測変数の分散のことを、独自分散(unique variance)という。個別に各変数に影響を及ぼす要因としては、この独自因子の他に測定誤差も存在している。

- ・ 因子負荷量(factor loading)

因子と観測変数間の関係性の強弱のこと。推定された因子の解釈に使用される。

因子分析では、複数の観測変数に共通した要因である共通因子を推定したい。しかし、実際に実験や観測によって得られた観測変数は、共通因子の他にも、その変数独自の因子や測定誤差を含んでいる。また、各観測変数に対し、どの因子が強く影響を及ぼして、どの因子があまり影響を与えていないのかという、関係性の強弱を知ることも必要であり、因子負荷量の値が大事になる。つまり、観測変数 y は因子負荷量 a と共通因子 x によって、以下の式のように分解することが出来る。

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \dots + \text{誤差}$$

因子分析には大きく分けて2つの手法が存在する。

- (1) 探索的因子分析(exploratory factor analysis: EFA)

因子と観測変数間の関係に付いて、先行する仮説や制約を分析時には考えず

に、あくまで観測によって自分が得られた観測変数(データ)のみから相関係数を計算し、観測変数間に相関関係をもたらす因子は何かを推定する手法のこと。

しかし、全く仮説をおかずに実験や観測を行う訳ではない。観測変数の間にどのような関係性があるか想定して関連がありそうな観測変数をそろえなければ因子を推定できなくなってしまうからである。「どんな関係性があるかは分からないが、何かしら関係はしていそう」な観測変数を選ぶということである。

この探索的因子分析では、とりあえず分析内の全ての環境変数と因子が関連するというモデルを想定して分析する。そして得られた推定値を考えて、本当に関連している因子は何かを解釈する。

(2) 検証的因子分析(confirmatory factor analysis: CFA)

確率的因子分析とも呼ばれる。先行する理論を元に、因子の数や意味、因子と環境変数の関係をあらかじめ、規定しておいた仮説を作っておき、その仮説が本当に正しいか検証する為に、得られた観測変数を分析する手法のこと。「観測変数間の相関関係は仮定した因子構造と一致するか否か」という問いに答える為に分析を行い、データと仮説の当てはまり具合を検証する。探索的因子分析とは異なり、想定される因子は全ての観測変数に関連するとは見なさずに因子と関連が強い観測変数にのみ焦点を当てている。

探索的因子分析は以下の目的で使われる。

① 構成概念を探る

観測変数間の相関関係をまとめ、そのパターンを統計的にまとめることで、背後にある因子構造や概念を理解・定義したりする場合に用いられる。

検証的因子分析では、あくまで元からある仮説が正しいか検証するだけなので、予期していなかった因子構造や概念は見つかりにくい。しかし、探索的因子分析では、必要最低限の仮定しかおかない為、考えていなかった関係性が現れることもあるかもしれない。

② 妥当性の高い質問用紙を作成する

測定が難しいと思われるような概念や要因を引き出す為に、とりあえず考えられる質問項目を大量に用意しておき、試作の質問用紙を作る。そして、その回答結果を分析し、因子に分ける。その際に、どの因子にも属さない(妥当でない)質問項目を削除し選別を行う。これを何回か繰り返すことで、より妥当性の高い質問用紙を作成することができる。

③ 変数をまとめる

質問項目のように大量に観測変数がある場合、出来るだけ数少ない変数にまとめることで一般化を行い、因子間の関係性を検証しやすくしたり、他の目的に利用しやすくすることが可能となる。

ただし、ただ単に多くの観測変数をまとめて単純に数値化することのみが目的であるのならば、主成分分析(principal component analysis)を用いる方が適切である。より一般化され多くの観測変数に共通して影響を及ぼす因子を見つける為に変数をまとめるのである。

以上3つのような用途を挙げたが、行った探索的因子分析の結果が妥当かどうかを確認する為に確認としてさらに、異なるサンプルを用いて検証的因子分析を行うこともある。

§ 9-2 因子分析の基本事項

(1) サンプルサイズ

サンプルサイズに絶対的な基準はなく、データの質や環境変数や因子の数、相関関係の強さや共通性等によっても、適切なサイズは変わる。

一般的には、信頼できる相関係数の算出の為に、全体として 300、1 変数辺り 10～15 は必要である。どのくらいが妥当であるかの理論的根拠としては、KMO の標本妥当性の測度(Kaiser-Meyer-Olkin measure of sampling adequacy)で検証すると良い。また、十分なサンプルサイズが確保できないような場合でも、測定する観測変数の数を絞り込み、少なくとも観測変数の 10 倍程度は確保する必要がある。

(2) データの種類

観測変数として、間隔尺度・比率尺度の連続データを扱う。

(3) 観測変数の数

3～4つの観測変数が1つの因子に対して高い因子負荷量を示すことが1つの目安である。観測変数を増やせば、多面的に因子を分析することが可能になり信頼性は増すが、そのぶん分析は大変になる。

(4) 観測変数間の相関

因子の分析に用いられる相関係数が.30以上を示す観測変数が見られない場合、共通因子を十分に抽出することが出来ず、分析の意味がない。反対に.90以上の場合多重共線性や単一性の恐れがある。

因子負荷・因子寄与・共通性

(1) 因子負荷量

因子と観測変数がどのくらい関連しているかを-1.00~1.00の間の値で表す。この値が大きいかほど強い関連性があり、一般に.60以上で高く、.30未満で低いとされる。

(2) 因子寄与・因子寄与率・累積因子寄与率

① 因子寄与

各因子が観測変数の変動をどの程度説明しているのかを表す数値。各因子に置ける観測変数の因子負荷量の2乗の和。各因子が観測変数に対して全体的にどの程度貢献しているのかを示す。

② 因子寄与率

各因子の因子寄与を最大値で割った数値。各因子が全観測変数に対してどの程度の割合で貢献しているのかを示す。観測寄与の最大値は因子の理論的最大数で観測変数の総数に相当する。

③ 累積因子寄与率

因子寄与率を第1因子から順に足して行った数値。

(3) 共通性

観測変数を説明している各因子負荷量の2乗和で算出され、抽出された因子全体が、個々の観測変数をどの程度説明しているかを表す。全因子との共通分散の合計が観測変数の共通性となり、0~1の値をとる。よって1からそれぞれの観測変数の共通性の値を引くと独自性が分かる。

因子の推定方法

複数の観測変数から因子を推定し妥当な解釈を行うためにはまず最初に因子の推定方法を選択する必要がある。

仮説検証の行う為の前提として、母集団まで一般化しないのであれば、主成分分析、主因子法、イメージ因子法が適しており、母集団にまで結果を拡張して一般化したいのであれば、最尤法(さいゆうほう)もしくはアルファ因子法が適している。

因子分析では、因子抽出の段階で、共通性が1以上の値を示す不適解という問題が起こることがある。その原因としては、以下の2点がある。

- ① 推定する因子負荷量の数に対してサンプルサイズが十分でない場合
- ② データに適合する因子のモデルが定まらない場合

主因子法は、因子モデルとデータの適合度を計算できる最小2乗法や最尤法と比べて不

適解が出にくい。しかしこれは同時に不適解というかたちで現れるデータの問題を見落としてしまう可能性もある。よって現在では最尤法が良く使用される。

因子の回転方法

因子をより解釈可能なものとする為に、因子を回転させる必要がある。通常、因子分析を行うと因子が抽出されるが、その段階では因子と観測変数はうまく合致していない。そこで因子軸を引き、それを回転させた上で、傾向の似たいくつかの観測変数の固まりを解釈できるようにグルーピングするプロセスが必要となる。つまり、「因子軸をそれぞれ観測変数群に重なるように動かす」ことが因子の回転である。

回転には、直行回転と斜交回転の2通りがあり、そのそれぞれにいくつかの回転方法がある。

直行回転は、「因子間には相互作用がなく互いに独立している」という制約を置いた回転方法であり、代表的な方法としてはバリマックス回転が挙げられる。

一方、斜交回転は「因子は互いに相関がある」と想定した方法で、プロマックス回転が使われる。

回転させる前の時点では、観測変数がどちらの因子と関係があるのか判断しにくいいため、2本の軸を直行回転、あるいは斜交回転させて、更に観測回数を重ねて行くことで、因子の解釈を容易にするのである。この2種類の回転方法には以下のような相違点がある。

① 回転軸の動かし方

直交回転では2本の因子軸を直角に固定したまま動かす。角度を保って回転させるため、軸と観測変数の重なりに制約が生まれる。

斜交回転では2本の因子軸を個別に動かす。そのため、観測変数の散らばり具合に対して柔軟に軸と観測変数を重ねることができ、因子の解釈が行いやすい。

② 因子負荷量と相関係数

直交回転では観測変数同士の相関を想定せず、因子負荷量と相関係数は同じものとして扱われる。

斜交回転では因子負荷量と相関係数の2つは異なるため、別々の票に出力される。

③ 因子寄与・因子寄与率・累積因子寄与率の算出

直交回転では、回転させても一方の因子への因子負荷量が高まれば、他方への因子負荷量が低下するという関係性が維持されるので、因子寄与の最大値が観測変数の数になり、因子寄与の計算が可能となる。そのため、初期・因子抽出後・回転後の3か所においてこれら3つの値が表示される。

一方、斜方回転ではどの因子に対しても高い因子負荷量を示すことが可能であるため最大値が定まらず因子寄与の計算を行うことができない。そのため、回転後の因子寄与の値が出力されず、全体的にどの程度の貢献度なのかについては相对比较に留まる。

このように、直交回転・斜交回転ともに、メリットデメリットがある。因子を回転させることの最大の目的は、単純構造を達成し、因子の解釈を容易にすることである。

単純構造とは、観測変数が強い関係性を持つ因子に対してはより強い因子負荷量を、関係性の低い因子にはより低い因子負荷量を示し、各因子がより明確に構成されている状態のことである。

この、単純構造を達成するという点においては、回転軸の角度を自由に動かせる斜交回転の方が優れている。さらに、因子間の相関を推定しないのは不自然であることが多く、これらの理由から斜交回転の方が広く使われている。

因子分析では、データに分析をかけると推定地の算出・因子の回転が客観的に行われている。その後の最終的な因子数の決定や因子の解釈は、分析者自らの判断で行う。

それはすなわち、観測変数に対して最も解釈が妥当である因子を探ることが探索因子分析において、大きな作業となる。そのため、因子の推定方法や回転方法の選択は、非常に重要なプロセスである。

§ 9-4 主成分分析

相関関係にある観測変数を限りなく少ない成分へと分解し、データを要約することで新たな合成変数を作ることを目的とする統計的分析手段である。ここでは、主成分分析と因子分析では多くの類似点が見られるが、以下のような違いがある。

- ① 因子分析では、観測変数の分散が、共通分散と独自分散に分けられ、共通分散のみを因子の推定に利用した。それに対し、主成分分析では、全ての分散がグルーピングに用いられる。そのため、主成分分析では、共通性が 1 に固定される。誤差は各変数に分離されずに含まれたままで、最後に誤差が足されるということがない。また、不適解などの共通性に関する問題も見られない。
- ② 因子分析では、観測変数間の相関係数の原因となる潜在的な因子を推定し、その因子から観測変数への影響を推定する。一方、主成分分析は観測変数を要約するという主要観点から、観測変数から合成変数への影響関係を想定する。各観測変数がどの程度合成変数に影響を与えるのかを示す量が主成分負荷量である。因子分析では各観測変数を探る式を立てるが、主成分分析では、共通概念である主成

分を求める式を立てることになる。

- ③ 因子分析では、複数の因子を仮定することにより、観測変数全体を説明しようと試みる。一方、主成分分析では第 1 主成分が最も分散を最大限に説明するように計算が行われる。そのため、最も大事なものは第 1 主成分である。
- ④ データの要約に主眼を置く主成分分析では、相関が高いほど、情報の集約性が高まることから、多重共線性や単一性が問題にならない。これを利用して重回帰分析において、多重共線性や単一性が問題になる場合に、主成分分析で得られた主成分得点を独立変数として重回帰分析を行う、主成分回帰という手法が存在する。

以上の説明から明らかなように、主成分分析は複数の観測変数を集約し、主成分得点により、変数化や数値化をすることが主な目的となる。因子分析でも因子の変数化や数値化は可能ではあるが、因子分析の本来の目的は観測変数の背後に共通して存在する因子の推定である。このように、主成分分析と因子分析は目的や特徴が異なるので、適切に使い分ける必要がある。