

大野研サブゼミ第2回事前資料

高田真也

2021.11.17

実施概要

2021年11月17日(水) 10:00-12:00

- (1) 共分散構造分析の特徴とモデル構築方法についてレクチャー(30分程度)
- (2) 応用例の比較を通じたよいモデル構築のためのディスカッション(40分程度)
- (3) B4学生に対するアドバイス(15分程度)

1. はじめに

狩野によれば、共分散構造分析とは、直接観測できない潜在変数を導入し、潜在変数と観測変数との間の因果関係を同定することにより社会現象や自然現象を理解するための統計的アプローチとされ、因子分析とパス回帰の拡張である(狩野,2002)。

活用方法としては、①調査項目間の因果関係を調べる、②項目をまとめて単純化する、③項目をまとめて単純化してから因果関係を調べるといったことが挙げられる(狩野,2002)。回帰分析では従属変数に関する誤差のみが設定されるが、共分散構造分析では、因果関係上の誤差、原因系の潜在変数を測る指標の誤差、結果系の潜在変数を測る誤差など原則、すべての誤差を検討することができる(狩野・三浦,1997)。またモデル構築を通じて直接効果、間接効果、総合効果への分解が可能となる。

理論的には、構造方程式において、未知係数(パス係数の値や、残差、外生変数の挙動を示すパラメーター(分散・共分散)を求めるために、方程式を標本共分散行列という方程式に変換し、これを解くことを構造方程式モデリングと呼ぶ。

2. 潜在変数と構造方程式

直接観察できない変数を潜在変数と呼び、観察変数と区別する。観測変数 X は、潜在変数 F と誤差で構成される。すなわち、

$$X = F + e$$

潜在する共通因子のために観測変数間に相関が生じる場合は、以下の様に示す。

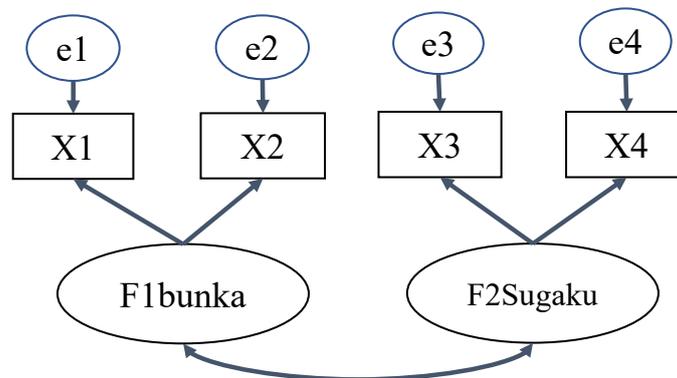
$$\begin{aligned}v_1 &= \alpha_{11}f_1 + e_1 \\v_2 &= \alpha_{21}f_1 + e_2\end{aligned}$$

ここで、観測変数 v_1 および、 v_2 は共通して f_1 という因子を含んでいる。 α_{11} は因子負荷と呼ばれ、影響力の強さを示す。なおモデルの中で積極的な意味を持つ潜在変数を構造変数と呼ぶことがある。

3. 検証的分析と探索的分析

分析には大きく分けて検証的分析と探索的分析の二種類がある。例えば検証的分析では2因子を想定したモデルの構造は次のようになるが、ここでは因子 F_1 は X_3 、 X

4には影響を及ぼさず、因子 F2 は X1, X2 には影響を及ぼさないという仮説に基づいている。一方、探索的因子分析では、すべての因子がすべての観測変数に影響を与えると解釈して因子数を選定し影響の大きい因子を特定する(狩野・三浦,2020)。



4. パス図の引き方

実際のパス図を引く際に、次のような決まりがある。

- ・観測変数は四角形で囲む
- ・構造的な潜在変数（構成概念）は円または楕円で囲む
- ・誤差変数は囲まない。
- ・影響を与える変数から与えられる変数に単方向の矢印を書き、矢印に影響力を示す数値を付与する。
- ・共変動を示す2つの外生変数に因果関係を仮定しないときは双方向の矢印を引き、矢印に共分散（または相関）を示す数値を付与する。

また潜在変数の設定の仕方については、以下の様な点を考慮する事が重要とされる。

- ・共変する似た項目は潜在変数でまとめ、違った側面を持つ項目は潜在変数化しない方がよい。
- ・誤差変数と観測変数の間に共分散を設定しない
- ・誤差変数の間では、共分散は設定しない
- ・観測変数・構造変数（重要な潜在変数）の間には共分散を設定する
- ・最初から潜在変数間にパスを設定し、多重指標モデルを作るのではなく、適合度の高いモデルを作ったうえで、因子間の相関関係から説明するモデルを考え、多重指標モデルに進化させるのが良い（豊田、1998:69）

5. モデルの評価

適合度として代表的なものを示すと、

χ^2 検定・・・モデルとデータのずれを評価する。標本サイズが多くなると検出力が高くなるため、標本サイズが数百までなら実用的とされている。

GFI・・・母共分散推定値行列が標本共分散行列を説明している割合を表して

いる。自由母数を追加すると増加するために、AGFIを用いることが多い。

$$GFI = 1 - \frac{\text{tr}\{\Sigma^{-1}(S - \Sigma)\}^2}{\text{tr}\{(\Sigma^{-1}S)^2\}}$$

$$AGFI = 1 - \frac{n(n+1)(1-GFI)}{2df}$$

GFI、AGFIは1に近いほど適合度が高く、0.9あるいは0.95より大きい場合を適合度が高いモデルと判断される事が多い。

RMSE・・・ χ^2 が期待値 df より大きくなった分を自由度1つあたり、観測値1つあたりの値として評価する。0.05以下であれば良好、0.1以下であれば適合度が認められない。

$$RMSE = \sqrt{\text{Max}\left[\frac{\chi^2 - df}{df(N-1)}, 0\right]}$$

(小島,2003)

パスを入れ、パラメーターを増やすと、カイ2乗値の値は低下する。しかしパラメーターのより多いモデルが良いモデルではない。情報量基準は引いたパスの個数ペナルティをかけ、カイ2乗値とパラメーター数のバランスを取る意味を持つ。

$$AIC = T_{ML} - 2d$$

(但し、 T_{ML} はカイ二乗値、 d は自由度)

4. 分析のすすめかた

構造化モデリングの応用例は広い。豊田(2003)をもとに代表的なものを示す。

(1) 1要因の経時データをもとにモデルを組む場合

→ LCA (Latent Curve Analysis:潜在曲線分析) という手法が用いられる。

(2) 欠損値の処理

→ データを欠損した変数ごとに分類し、母集団間で母数が等しいという制約をかけた多母集団の同時解析を行うのが望ましいとされている。

(3) 2母集団の同時解析

→ 目的としては、母集団間のモデルの同質性・異質性を考慮することにある。

(4) 順序尺度・名義尺度

→順序尺度については、2値データについて、最尤推定量に基づく多分相関係数(ポリコリック相関係数)や多分系列相関係数(ポリシリアル相関係数)を用いるのが望ましく、3段階尺度・4段階尺度はグレイゾーン、5段階以上だと連続変数として処理して問題ないとしている(狩野・三浦,2020)。萩生田-繁梲(1996)も5段階尺度・7段階尺度を勧めている。

男女などの名義尺度については、ダミー変数の形で独立変数として用いることには問

題がないが、従属変数に用いる場合には注意が必要である。

(5) サンプルサイズ

→自由母数の数の5倍から10倍の標本数を薦めるものもあるが、200が目安になると考えられている(Boomsma,1985)。

(6) 誤差間の相関

→観測変数の誤差間に相関を設定することの意味は、変数間にモデルで導入された構造変数以外の共通の変動要因が存在することを意味する。多くの場合、標本を収集したときの状況(同じ対象に反復的に測定した場合や、大人数から同時に収集した観測変数と1人1人別個に取った観測変数が混在する場合)に起因するが、多変量LM検定を用いて誤差項に相関を引くかどうかを検討することができる。AMOSのデフォルトでは誤差項の設定すべきかどうかは検討されていないため、Build LMTTestダイアログを用いて設定する必要がある。詳しくは狩野・三浦(2020:170)参照のこと。

5. 分析がうまくいかない時

分析がうまくいかない時、具体的にはGFIやCFIの値が0.7や0.8にしかならない時、その原因として(1)異常値が混入したり、質問項目が不適切であるなどデータに問題がある、(2)モデルが適切ではない(3)母集団分布にあう適切な推測方法が指定されていないといったことが考えられる(狩野・三浦,2020)。このうち、モデルを修正する場合、具体的には既に引いてあるパスを外してシンプルなモデルにするためには、ワールド検定に基づいて有意でないパラメーターを外すのが有効な手段である(狩野・三浦,2020)。

参考文献

- ・小島隆也(2017)、「EXCELで学ぶ共分散構造分析とグラフィカルモデリング」、オーム社
- ・狩野裕(2002)、「共分散構造分析の基礎と実際－基礎編－」SSJデータアーカイブ第6回公開セミナー資料
- ・狩野裕・三浦麻子(2020)、「AMOS,EQS,CALISによるグラフィカル多変量解析」新装版、現代数学社
- ・豊田秀樹(1998)、「共分散構造分析[入門編]」、朝倉出版
- ・豊田秀樹(2003)、「共分散構造分析[疑問編]」、朝倉出版