共分散構造分析の基礎知識

心理学データ解析演習 2006/5/10
発表者：志波　泰子

1

発表のながれ

1. 共分散構造分析とは何?
   その一般的なSEMの手順は?
2. SEMの基礎知識とパス図のルール
3. なぜ共分散構造分析というのか?
4. SEMの構造方程式と測定方程式
5. SEMの利用法ーおもなモデルの紹介
6. SEMの難問：推定と適合度の問題

2

1 ❏ 共分散構造分析とは何?

▮ LISRELの開発者Jöreskogら(1968) の提案
➔因子分析の弱点を補うもの、
従来の探索的因子分析(exploratory factor analysis)（観測変数に潜在する因子の数と意味だけの探索）でなく、仮説を検証する検証(確認)的因子分析法(confirmatory factor analysis)へ
➔パス解析を用いた構成概念間の因果関係の分析へ発展

3

■ 構成概念と因果関係を同時に扱うもの

- 検証的分析法：仮説を立ててその妥当性を検討し構成概念間の因果関係などを分析
  ☞因子分析は構成概念と観測変数間を明らかにするが因果関係は扱わないー潜在変数の想定し、因子パターンの推定
  ☞回帰分析とパス解析は観測変数間の関係を明らかにするだけで構成概念は扱わないー因果関係を仮定し、偏回帰係数を推定
- 仮説について自由にモデルを表現できる
- 既に得られている知見をモデルに反映できる

4

❏ SEMの一般的進め方

手順：
1. 仮説の設定：
   検証すべきものは何か、分析者の自由な仮説をモデル化
2. データの収集：
   どのような観測値が適当かを考えてデータを収集
3. モデルの構成：
   仮説を表すモデルをパス図に表現する
4. 分析の実行：
   SPSS上のデータをAmosなどを用いて分析

5

5. 結果の判定
   結果からモデル全体の適否をまず判断、χ^2検定や適合度指標などを利用して、全体の適合状況、因果関係の有意性の評価、部分的適合性評価
6. モデルの修正・改良：必要ならより適切なモデルへしかし合理的な範囲で修正すること

☛SEMの目的は仮説の検証であって、得られたデータに適合するモデルの探索が目的ではない
☛モデルの適合度が高くても因果関係が支持されたのではなく、方程式のモデルがデータと整合的であったということでしかない

6

2❏パス図の基礎知識とルール

- 構成概念：購買意欲、食物意識や数学的能力、文科的能力のような直接には計測できないもの
- 潜在変数：直接観測できない「構成概念」を測る
- 観測変数：直接観測可能な変数
- 内生変数（従属変数）：どこからかの影響を受けている、片方矢印を一つでも受けている変数

相関関係は設定できないが 因果関係は設定できる

- 外生変数（独立変数）：どこからも影響を受けていないパスを出すだけ（従属変数以外の）変数

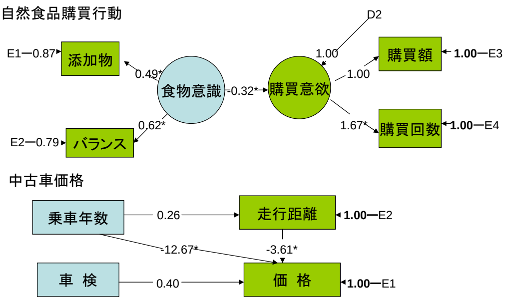
相関関係を表す双方矢印↔は独立変数間のみに設定する

- 誤差変数:内生変数への外部からの誤差

どこからも矢印を受けないので独立変数であり、直接観測されないので潜在変数でもある

観測変数には誤差変数（E）、潜在変数には攪乱変数（D）がある

■多重指標モデルとパス解析モデル（狩野,2003）



■パス図のルール

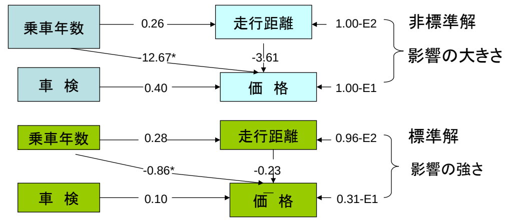
- 観測変数は四角形で囲む
- 潜在変数は円または楕円で囲む
- 誤差変数は記号のみ（あるいは楕円）
- 片方矢印→は因果を表す、その上にパス係数の推定値を付す
- 片方矢印→を受けた従属変数には、必ず誤差変数が付属
- 双方矢印↔は単なる相関関係を表す

その上に相関係数または共分散の推定値を付す

■標準解と非標準解はどう違うか

▮標準解：すべての変数の分散を1に標準化

☞従属潜在変数を含む分析は分散をそろえるため
通常、標準解を解釈、多母集団の同時分析では
非標準解



（狩野,2003)

■偏回帰係数と標準偏回帰係数

▮非標準解は偏回帰係数：単位に依存する

価格＝-3.61走行距離-12.67乗車年数+0.40車検
+152.68（定数）

（推定された回帰式）

価格＝-3.61走行距離-12.67乗車年数+0.40車検

➔乗車年数が1年多くなると12.67万円下がる

➔走行距離が1万キロ長くなると3.6万円下がる

➔車検の残りが1ヶ月多いと0.4万円上がる

・他の独立変数が一定のときに、当該変数が1単位動いたとき従属変数への影響の大きさである

■標準解には標準偏回帰係数：要因に関わる

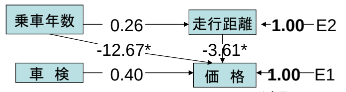
・単位の不ぞろい、ばらつきの不ぞろいを調整
・要因の影響の強さを比較する
・標準解（変数の平均を0に、分散を1に変換）
推定された回帰式は
価格＝-0.23走行距離-0.86乗車年数+0.10車検
➔乗車年数が一番強い因果構造をあらわす

13

■直接効果、間接効果、総合効果

- 直接効果：変数が別の変数へ直接的に影響
- 間接効果：他の変数を経由して影響を及ぼす
- 直接効果+間接効果=総合効果（媒介効果）

ex. 乗車年数から価格への直接効果＝-12.67
間接効果＝ 0.26×(-3.61)=-0.94
総合効果＝-12.67+(-0.94)=-13.61



（狩野,2003）

14

3 ❏ なぜ共分散構造分析というのか

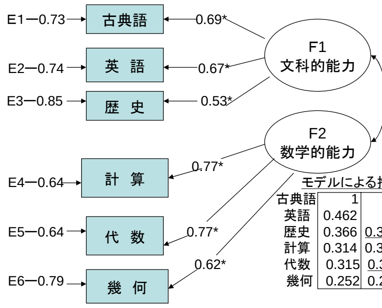
・相関(共分散)構造：パス図から理論的に計算された相関行列
・ パス図から理論的に相関係数を算出
パス係数などの推定値はモデル(Σ^) がデータ(S)に最も近くなるように計算される

・ SとΣ^の違いの程度で適合度を測る

・ モデルの適合度は変数間の共分散(相関係数)についてデータからの計算値とモデルに基づいた計算値の近さによって評価される

15

■パス図から理論的に計算した相関行列(狩野,2003)



ex.（ 英語と歴史）
=0.67x0.53x1=0.355

ex.（英語と代数）
=0.67x0.77x0.6=0.309

モデルによる推定された相関行列 Σ^

古典語	1					
英語	0.462	1				
歴史	0.366	0.358	1			
計算	0.314	0.307	0.244	1		
代数	0.315	0.308	0.244	0.589	1	
幾何	0.252	0.247	0.196	0.472	0.473	1

16

■SとΣ^ の差で適合度を測る（狩野,2003）

相関行列 S

古典語	1					
英語	0.439	1				
歴史	0.410	0.351	1			
計算	0.288	0.354	0.164	1		
代数	0.329	0.320	0.190	0.595	1	
幾何	0.248	0.329	0.181	0.470	0.464	1

モデルによる(推定された)行列 Σ^

古典語	1					
英語	0.462	1				
歴史	0.366	0.358	1			
計算	0.314	0.307	0.244	1		
代数	0.315	0.308	0.244	0.589	1	
幾何	0.252	0.247	0.169	0.472	0.473	1

残差行列 S− Σ^

古典語	0					
英語	−0.023	0				
歴史	0.044	−0.007	0			
計算	−0.026	0.047	−0.080	0		
代数	0.014	0.012	−0.054	0.060	0	
幾何	−0.004	0.082	−0.015	−0.002	−0.009	0

ex.
（ 古典語と英語） **0.462=0.69x0.67**

（ 古典語と幾何） **0.257=0.69x0.62x0.6**

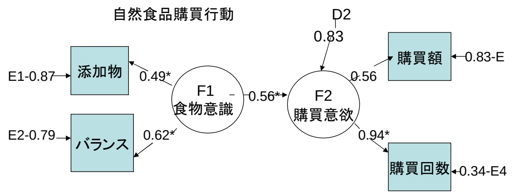
1

17

・ モデルとデータができるだけ適合するようなパス係数（因果係数）や分散を求めること
・ それらの解を求める基準として、
最小2乗法(least squares estimation)や最尤法(maximum likelihood estimation)が用いられる
最小2乗法：（S－Σ^）の差の2乗について基準関数値を なるべく最小にするパラメータを求める
＊一般化最小2乗推定法
F_{GLS}=1/2tr $[\{(S-\Sigma^{\wedge}) S^{-1}\}^2]$

18

４❑多重指標モデルの構造方程式と測定方程式



19

■SEMの名称はStructural Equation Modeling（構造方程式モデル）

- 構造方程式は真の因果関係を表すもの、平均の構造も分析するので共分散構造分析より適切（狩野,2003）

- SEMは構造方程式と測定方程式で構成

20

■構造方程式：

- 直接観察されなかった潜在変数（構成概念）を生成
- 因果関係を表現するための方程式

（自然食品）購買行動モデルでは

食物意識＝0.49添加物+0.62バランス+誤差

購買意欲＝0.56食物意識+誤差

21

■測定方程式：

- 実際に観察された変数を生成
- 潜在変数が観測変数によってどのように測定されているかを記述する方程式

（自然食品）購買行動モデルでは

添加物＝0.49食物意識 ＋誤差

バランス＝0.62食物意識＋誤差

購買額＝0.56購買意欲＋誤差

購買回数＝0.94購買意欲＋誤差

22

５❑SEMの主な3つの利用法

1. 項目間の因果関係を調べる

 ➔重（単）回帰分析（パス解析）モデル

 パス解析は本来は潜在変数を扱わない、変数間の因果関係を相関係数をもとに調べるもの、SEMの部分的モデルとなる

2. 項目をまとめて単純化（潜在変数化）

 ➔検証的因子分析：観測変数間の関係を導入し、その仮説がデータと矛盾しないかを検討する、影響しないパスは積極的に0に固定して推定

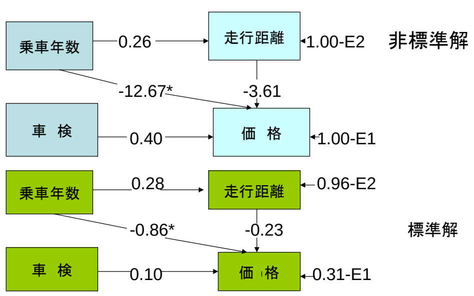
3. 単純化（潜在変数化）してから因果関係を調べる

 ➔多重指標モデル（典型的な共分散構造モデル）

 共分散の推定と検定：観測変数の分散と共変関係を調べる

23

（1）パス解析モデル：中古車価格（狩野,2003）



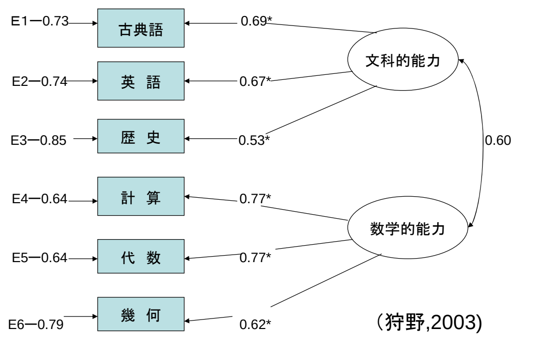
24

■SEMによるパス解析と従来のパス解析

- 従来のパス解析のように内生観測変数の数だけパス回帰分析を繰り返すよりAmosなどを用いて正確な解を求めるべき
  ➔正確な最尤解や最小2乗解が報告されない
- 因果モデルの改良が容易
- 誤差相関を入れたモデリングが可能
- 双方向因果モデルが可能

25

（2） 文科的能力と数学的能力：検証的因子分析モデル



（狩野,2003）

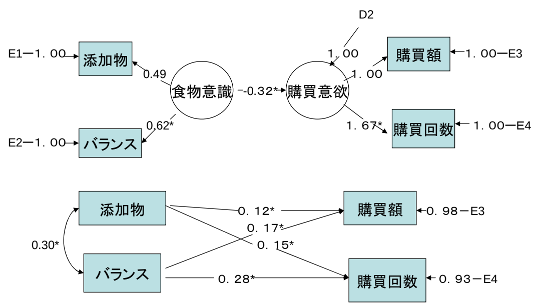
26

■探索的因子分析（EFA）と検証的因子分析（CFA）

- どちらの分析も可能なとき、CFAは検定によってパスの有意性が確認できる
- 因子に関する様々な仮説が検証できる
  EFAには事前の知識がほとんど反映されない、独自因子の相関を仮定しない
- EFAは因子回転法に依存するのみで細かい統計的推測は無理
  直交解も斜交解も予測値は同じで適合性を比較できない

27

（3）代表的多重指標モデル：購買行動（狩野,2003）



28

■パス解析と多重指標分析

- 「バランス」が一定の下での「添加物」の影響は何かがよくわからない
- 共変する似た項目は潜在変数でまとめるほうがよい
- しかし、違った側面を持つ項目は潜在変数化しないほうがよい（中古車価格の例）

29

6❏SEMの難問： 推定と適合度の問題

A. 推定の問題:

- パス係数などの推定値はモデルがデータに最も近くなるように計算される
  観測変数についてはデータがあり、その共分散は分かるが、潜在変数の分散、誤差変数、攪乱変数の分散が分っていない
  ➔識別性の確保について
- データとモデルにおいて解（パス係数の推定値）が一意に求められない
- 求める母数の数が得られる方程式の数より多い

30

☛調査されたケースが未知数より多いことが必要条件：

ex. 未知数の値を固定するかまたは、方程式をもうひとつ得るか?

➔モデルに制約を入れて母数の一部を既知とする

☛分析の目的やデータの性質を損なわない範囲で拘束条件を前もって設定する必要

拘束条件とされているもの：

- 独立変数である潜在変数の分散を1に固定してその尺度を定める
- 従属変数である潜在変数から観測変数へのパス係数のひとつを1に固定することによってその尺度を定める
- 誤差変動は互いに独立としてその間の共変動を1に定める
- 誤差変数からのパス係数を1に固定する

■ さらに求める解が得られるためには

☛自由度がプラスであることが必要

- 自由度0では唯一解となり、SEMには不適
- SEMではどの方程式も満たすいくつかの解を得たい

➔識別過剰：方程式の数が未知の数より多い方（不能）がSEMには適当、妥協の産物

➔その妥協が妥当かどうか

モデルの検定につながる

＊df=0のモデルを飽和モデル（Amosの出力モデル）という

飽和モデルは母数に拘束を課さないためデータはモデルと完全に適合するが、他のデータへの一般化可能性がない

■ 自由度(degrees of freedom)

- χ^2値の自由度である
- 推定する母数は通常はパス係数と独立変数の分散・共分散
- 観測変数の数をpとしたとき

既知の観測変数の分散・共分散の数はp(p+1)/2

- 自由度(df)は

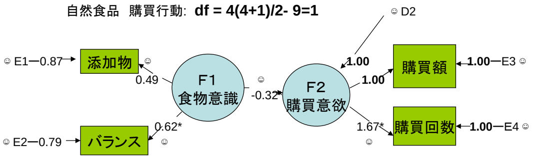
= p(p+1)/2－q

= 観測変数の分散・共分散の数

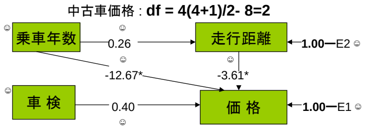
－推定する母数の数（q）

■推定するべき母数(☺)と自由度（狩野,2003）

自然食品　購買行動: **df = 4(4+1)/2- 9=1**



中古車価格：**df = 4(4+1)/2- 8=2**



（狩野,2003）

▮それでも、解が求まる必要十分条件ではない

今のところ前もって識別問題を解決する方法はない

☛ 識別条件の現実的解決はAmosを走らせて 結果を見ること

B. 適合度の問題：モデルの評価

■SEMは、自由な仮説をモデル化して分析

そのため仮説の妥当性をデータから検証し、吟味することが不可欠

・モデル全体を評価し、モデルがどの程度受容できるかを判定、モデルの部分評価では個々の因果関係を判定する必要がある

37

（1）χ^2検定：

・モデルが正しいことを帰無仮説で検定

5%の有意差ではモデルと観測データが等しいという仮説は棄却できない➔ 帰無仮説は正しい、モデルは観測データに適合

☛注意：本来、χ^2検定は棄却されない（「正しい」）というより棄却される（「正しくない」）と判断する危険度の確率であること

ケース数にも左右されやすく、ケースの増加はモデルを棄却しやすい（n=300~400以上のとき）

・ χ^2 だけでモデルの採否はできない

38

（2）適合度指標：

GFI, AGFI, CFI, AIC,RMSEAなどは

連続量的判断指標で、データとモデルの距離を表す

▮ **GFI**（Goodness of Fit Index）：

・モデルがデータのもつ分散共分散をどの程度説明するかの割合の指標、モデルとデータが完全に適合すれば、1になる

・通常0~1までの値を取り、1に近いほど説明力のあるモデルといえる

・推定母数に対する拘束の数に影響を受ける

39

▮ **AGFI**（Adjusted GFI）：

・自由度調整済み指標といえる（GFI ≧AGFI）

・GFIの欠点を修正、パラメータが多く複雑なモデルにペナルティを加える

・**GFI、AGFI**は0.9以上であることが目安

▮ **CIF**（comparative fit index）

・独立モデルと比較してモデルの適合度がどれほど改善されたかでモデルを評価

・0~1までの値をとり、1に近いほど適合がよい

＊独立モデル：観測変数間に相関がないことを仮定

40

▮**RMSEA**（root mean square error of approximation）：

・モデルの分布と真の分布との乖離を1自由度あたりの量として表現した指標

・一般に、0.05以下であれば当てはまりがよいとされ、0.1以上であれば、そのモデルは採択されるべきではないとされている

▮**AIC**(Akaike's Information Criterion)：赤池情報量基準

・相対的なモデルの良さを表す指標

・モデル間の比較に適している

・AICの値が小さいほど真のモデルに近い良いモデル

・χ^2検定やGFIによって複数のモデルを選択した後、AICが最小のモデルを選択する

41

■モデルの部分評価（因果関係の評価）

・個々の因果関係はパス係数や相関係数として求められる

・モデルが全体としてデータに適合しても個々の因果関係が有意でなければならない

・モデルの部分評価はｔ検定などでおこなう

・パス係数が有意であること、0に近ければ2つの変数間の関係はない

42

(1)Wald検定：t＝パス係数（推定値）÷標準誤差

・t ＞1.96 であれば5%有意水準でパス係数0
　を棄却、因果関係があるといえる
　（因果係数が0でない）

・ 帰無仮説（因果係数=0）を棄却できない場合
　その観測変数を除く方がよい

・1.96未満であれば不要なパスとみなす

（AmosではC.R.として 出力）

43

(2)決定係数（r^2）：

=1ー（誤差変動÷従属変数の変動）

・ パス係数に対し独立変数が従属変数の変動を
　どの程度説明できるか

☛独立変数が従属変数をすべて説明できれば
　1-0=1である
　従って、r^2が1に近いほどそのパス係数は有意

（AmosではSMCとして出力）

44

C. モデルの修正・改良

モデルがデータに対して不適切な場合の修正

▮ 修正指数としてLM検定（Lagrange Multiplier）：

・ パスを増やしたいときにそこにパスを入れてどの程
　度χ^2値が減少するか➔3.84以上の減少が必要

パスを減らしたいときはWald検定で

・ C.R.(Critical Ratio)が1.96未満なら不要なパスとする

☛パスは最も有意でないものから1つずつ削除、分析
　するべき

（Amosでは複数モデルの同時分析機能が利用できる）

45

必要に応じて適切なモデルへ修正・改良できる
しかし**合理的な範囲**で修正すること

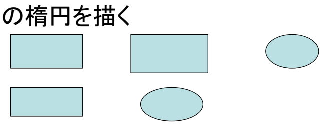
☛SEMの目的は仮説の検証であって、得られた
データに適合するモデルの探索が目的ではな
い

☛モデルの適合度が高くても因果関係が支持さ
れたのではなく、方程式のモデルがデータと整
合的であったということでしかない

46

■Amosを使って見る

- SPSSの「分析（A)」メニューから「Amos　5」選択、「Amos　Graphics」でAmosを起動、保存したSPSSのデータファイルを読み込む
- 1.変数を描く：
  ツールバーのアイコン（ ）をクリック
  ・3つの長方形を描いてみる
  ・次にツールバーのアイコン（ ）をクリック
  ・2つの楕円を描く



47

- 2.変数の命名：
  - 長方形をダブルクリック、 「オブジェクトのプロパティ(O)」のウィンドウが表示される
  - 「文字」のタブを選択、「変数名（N）」に入力
  - 同様に長方形、楕円をクリックして「変数名（N）」に入力
  - 「オブジェクトのプロパティ(O)」を閉じる

48

- ３．パスの矢印を描く：ツールバーのアイコン（←）をクリック
  - 仮説にあうように矢印を描く
- ４．パラメータの制約：
  - 誤差の変数を定義：
  - 誤差と変数間の矢印をダブルクリック
  - 「オブジェクトのプロパティ(O)」のウィンドウが表示される、「パラメータ」のタブをクリック
  - 係数を1に固定することにして「係数（R）」に1を入力
  - 「オブジェクトのプロパティ(O)」を閉じる

- ５．分析の設定：
  - 「分析のプロパティ」アイコン(　)をクリック（あるいは、「表示（V）」メニューから「分析のプロパティ(A)」を選択）
  - 「出力」のタブをクリック
  - 「最小化履歴（H）」「標準化推定値（T）」「重相関係数の平方（Q）」にチェックをいれる
  - 「分析のプロパティ（A）」を閉じる

- 6.分析の実行：
  - ツールバーの〔推定値計算〕アイコン（　）をクリック、（あるいは、〔モデル適合度（M）〕メニューから〔推定値を計算（C）〕を選択
  - ファイルの保存の指示➔適当な場所に保存
  - 保存するとAmosが分析を始める
  - 「最小値に達しました」「出力の書き込み」が表示される

引用文献:

狩野裕 (2003). 心理・教育測定論　講義（講義資料）

豊田秀樹 (1998). 共分散構造分析ー構造方程式モデリング—[入門編]　朝倉書店

豊田秀樹 (1998). 共分散構造分析ー構造方程式モデリング—[事例編]　朝倉書店

山本嘉一郎・小野寺孝義編著 (1999). Amosによる共分散構造分析と解析事例　ナカニシヤ出版

小塩信司 (2005). SPSSとAmosによる心理・調査データ解析　東京図書

服部環・海保博之 (1996). 心理データ解析　福村出版