

# 因子分析 (Factor Analysis)

- 変数の背後にある要因の探索 -

立教大学社会学部 村瀬洋一

## 1. 因子分析とは何か

### 1.1. 分析目的と具体例

因子分析の目的は、複数の変数の背後にある、かくれた要因を明らかにすること。または、似ている変数をまとめ分類すること（変数間の構造の解明）。因子分析は、実在の変数については、とくにXとYを定めない。すべての変数が説明変数Xである。量的変数（連続変数）のみを用いることができる。

具体例としては、例えば、ふつう国語の成績がよい人は社会や英語も成績が良い。これは、表面的なテストの点数の背後に、文化系能力や、理科系能力のような、かくれた要因（総合的な能力）が存在する、と考えることができるだろう。このことを図で表すと、以下のようなになる。

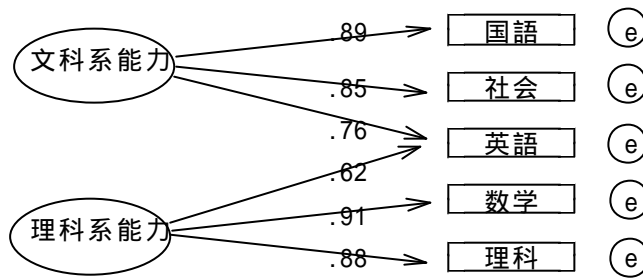


図1. 5科目試験成績の構造に関する因子分析結果（架空例）

つまり観測可能な変数の背後に、文科系能力や理科系能力など潜在的な要因（因子）が存在すると考えることができる。それならば、文科系総合テストがあれば1科目だけで済むはずだが、現実には、総合的な文科系能力を直接測定することはできないので3科目やっていると言ってよい。直接測定できない要因のことを、因子または潜在変数という。

調査データでの具体例としては、例えば、価値観に関する4段階回答の質問が10問あった場合に、それらを分析して2因子となることがある。その場合、それらの質問は、2つの隠れた要因（例えば、伝統的価値観と、政治的無力感など）に規定されていることになる。4段階回答等は厳密な量ではないが、連続変数と考えて良い。

### 1.2. 因子分析の考え方と基本モデル

因果関係図では通常、モデルとして自分で設定した因子（潜在変数）Fを楕円、観測変数（実在する変数）を四角でかく。上記を記号で表現したものが以下の図である。観測変数Xは、因子Fと、それ以外の要因eによって規定されている、と考える。

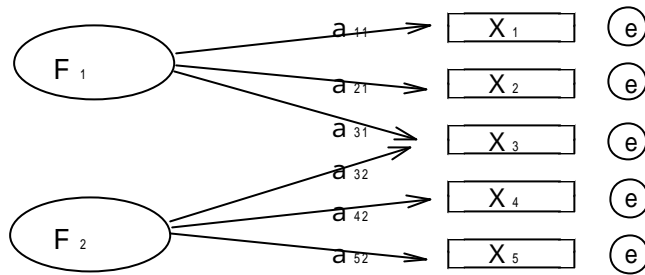


図2.記号による表現

$X_1$ について数式で表現すると以下の(1)式ようになる。 $a$ のことを因子負荷(factor loading)あるいは因子パターンと呼ぶ。上図では、 $a_{12}$ は0なので省略している。

$$X_1 = a_{11}F_1 + a_{12}F_2 + d_1e_1 \quad \dots\dots(1)$$

この式は、 $X$ が因子 $F$ と誤差項 $e$ によって規定されている、ということを表している。 $F$ によって説明される部分を共通性 $h^2$ (重回帰分析の決定係数 $R^2$ と同じもの)、誤差項部分を独自性という言葉で呼ぶこともある。例えば、 $h^2$ が0.30ならば、因子 $F$ により、ある観測変数は、分散の30%が説明されていることになる。

実際に分析する時は、因子 $F$ は分析後に出てくるので、自分の頭でどのような性質の因子かを解釈して、因子に名前をつけることが多い。これは、自由に解釈してつければよい。

回転後の負荷量の平方和(全ての $a$ を二乗した合計)は、回転後の因子寄与と一致する。因子寄与とは、回転後の因子の説明力の大きさである。例えば元の質問が5個あり、第1因子が2.2、第2因子が1.8の場合、元の質問4個分の分散を2因子で説明したことになる。

### 1.3. 因子分析の特徴

因子分析での観測変数は、すべて $X$ であり、特定の被説明変数 $Y$ はない。 $X$ と $Y$ をとくに設定しない点が、重回帰分析等と異なる。因子と観測変数との関係(偏回帰係数)が因子負荷である。これは直交解の場合、相関係数と同じ値である。

因子分析は、数学的には、もともとの相関行列を要約しているだけである。相関行列さえあれば、もとのデータが無くても因子分析を行うことができる。

なお、 $X$ 、 $F$ 、 $e$ はすべて連続変数(量的変数)である。 $X$ が離散変数(カテゴリー)の時は、数量化3類(対応分析、コレスポンデンス分析と同じもの)や、カテゴリカル因子分析等を用いる。

### 1.4. 因子分析の種類

探索的因子分析 - 従来型の因子分析。SASやSPSSなどで可能。

確証的因子分析 - この資料では触れない。モデルを探索するのではなく、事前にまずモデルを作り、構造方程式モデル(SEM; Structural Equation Mode

1, 共分散構造分析ともいう) のソフトによって分析する方法。  
AMOSやEQS、LISREL、S A Sのproc calisなどの分析ソフトを用いる。

### 1.5. 探索的因子分析の推定法 ( 因子抽出のための計算方法 )

主因子法(Principal Factor Method)

- 第 1 因子の因子寄与をもっとも大きくするように解を求める。

主成分法(Principal Component Method)

- 各因子の因子寄与が大きくなるように解を求める。

主成分法で回転をしない結果は、主成分分析の分析結果と同じ。

最尤法(Likelihood Method)

- 確率密度により解を推定する。共分散構造分析でよく使われる。

分布の歪んでいるデータでも正確な推定ができるとされる。

### 1.6. 因子分析と主成分分析

因子分析を主成分法で行い回転をしない場合、主成分分析と同じ結果になる。この 2 つは、数学的にはまったく同じものである。

主成分分析は、変数の要約が目的であり回転はしない。複数の観測変数を要約し、総合得点(主成分得点)を作るのが、主成分分析の目的である。主成分分析の結果を回転すると、因子分析を主成分法で行ったものと同じ結果になる。

### 1.7. 探索的因子分析の回転法

単純構造(各観測変数が 1 因子のみと対応する構造)になるように回転する。

バリマックス法 - 各因子間の相関は 0 となる(直交解)。説明できる分散を最大にするという基準で、回転後の分析結果を計算する計算法。

プロマックス法 - 各因子間の相関がある(斜交解)。自然な解釈が可能になることもあるが、データのあてはまりが悪いと分析結果が出ないこともありやや難しい。最近では、主因子法でプロマックス回転が良いとされることも多い。

## 2 . S P S S による分析

### 2.1. S P S S の操作

シンタックスの例は以下。太字の所に、自分の使いたい変数を入れる。因子数は何種類かを試して、適切に解釈できる結果を選ぶのがもっとも良く、最適な因子数についてとくに基準はない。ただし、因子数を増やしすぎて固有値の小さい因子が出ると良くない。自分の頭で解釈を考え、因子に名前を付けると良いだろう。抽出法は様々なものがあるが、通常は、主成分法(あるいは主因子法)を選べばよい。

## シンタックス例 1

### FACTOR

```
/VARIABLES Q3 Q4 Q6a to Q6n      分析したい変数名を書く
/CRITERIA FACTORS(3)           因子数を変える場合は(3)のところを書き換える
/EXTRACTION PC                 因子計算法の指定   ここでは主成分法
/FORMAT SORT                   因子負荷の大きさ順に変数を並べかえて出力表示
/ROTATION VARIMAX              回転法の指定
```

### 解説

```
/CRITERIA 因子数を書く   MINEIGEN(1)と書くと固有値 1 以上の因子数となる
/EXTRACTION 計算法を指定 PCと書くと主成分法
/ROTATION 回転法を指定している   VARIMAX回転とすることが多い
```

## シンタックス例 2

因子数を設定せず、MINEIGEN(1)コマンドを書き、固有値 1 以上の因子数とした例。

### FACTOR

FAC だけでも動く

```
/VARIABLES Q2 Q3 Q8a to Q8g      分析したい変数名を書く
/EXTRACTION PC                   主成分法PCを指定している。主因子法にする場合はPAF
/MINEIGEN(1)                    因子数は固有値 1 以上となる数にする（省略可）
/FORMAT SORT                     因子負荷の大きさ順に変数を並べかえて出力表示
/ROTATION VARIMAX               回転法の指定
/SAVE REG(ALL)                  この行を書くと、因子得点が新変数として作られる
/PLOT ROTATION                  プロットの図を出す命令文。なくてもよい
```

シンタックスを書かずに、SPSSで画面上の「分析」をクリックし、データの分解因子分析を選んでよい。回転法は、普通まずは「バリマックス」を選べばよい。オプションボタンを押して、因子負荷プロットの図を出すと分かりやすい。

## 2.2. 分析時のポイント

男女別（かつ年齢別）に分析した方が、明確な結果になることが多い。SPSSのデータウィンドウで、画面上の「データ」をクリックし、ファイルの分割を選び、性別の変数により分割する。

「得点」ボタンを押すと、因子得点を新変数として保存することができる。これは、各因子における総合得点のようなものである。分析後に、SPSSのデータウィンドウで、変数リストの最後を見ると、Fac1、Fac2などの名前がついた新変数ができる。

因子数をいくつにするかは、特に決定的な方法はない。いくつかの結果を出して、自分で解釈しやすい因子数とすればよい。また、因子に名前をつけることが多いが、これも自分で自由に解釈して、名前をつければよい。SPSSの初期設定では、固有値が1以上となる因子数が、まず出てくる。固有値が1とは、元の質問で言えば1個分の分散を説明する因子、という意味である。

### 2.3. 結果のまとめ方

下記のように、因子負荷構造を表にする。回転後の結果のみを表にすればよい。

SPSS出力の、

- ・ 共通性（因子抽出後） - 上記の因子分析モデルで説明したもの
- ・ 回転後の負荷量平方和 - 回転後の因子寄与。因子の説明力の大きさ
- ・ 回転後の成分行列 - 因子負荷の数字が並んでいるもの

を見て、適切に表を作る。

分析結果を視覚的に提示したい場合は、因子負荷構造を散布図にすると良い。

表1. 政治と権威に関する質問項目の変数間の構造 1995年××調査男性  
主成分法 バリマックス回転後の因子負荷構造

	第1因子 権威主義	第2因子 有効感	共通性 (h <sup>2</sup> )
権威のある人には敬意を払うべき	0.71	0.00	0.50
以前からなされてきたやり方を守る	0.77	-0.15	0.62
伝統や習慣にしたがうべき	0.67	-0.15	0.48
指導者や専門家に頼るべき	0.69	-0.12	0.49
政治のことは理解できる	-0.24	0.72	0.58
政治のことはやりたい人にまかせない	-0.28	0.59	0.43
国民の意見は国の政治に反映されている	0.14	0.71	0.52
因子寄与	2.47	1.14	3.61
因子寄与率 (%)	35.3	16.3	51.6

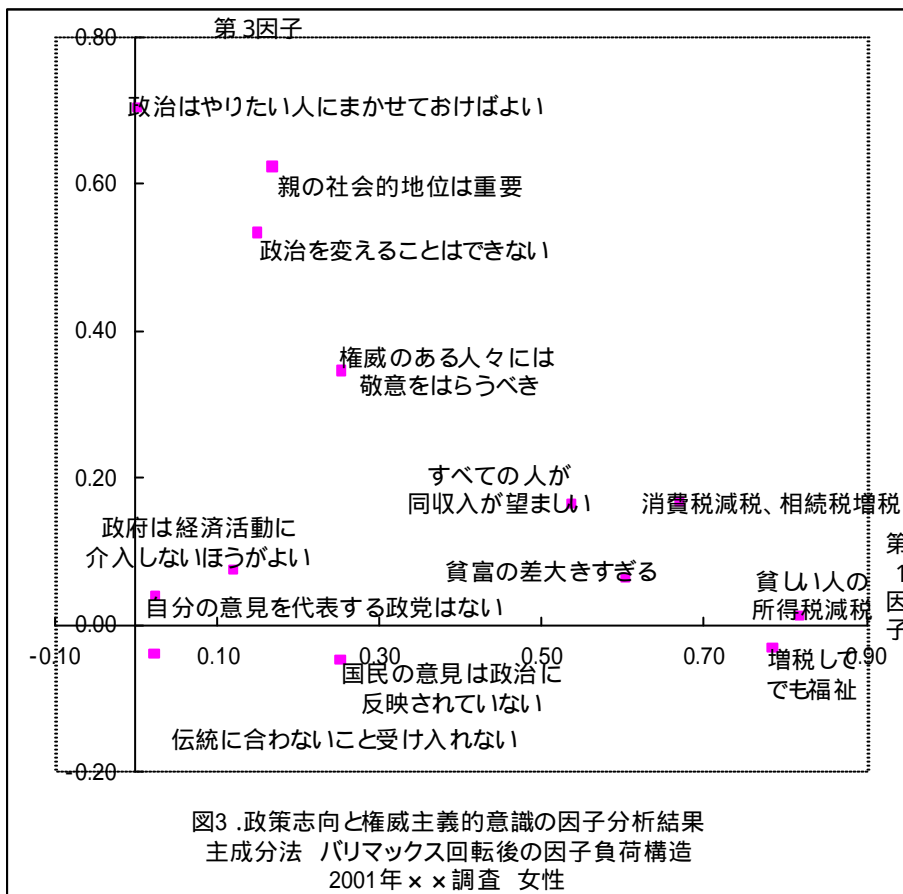
注 各項目は1-4の値をとり「そう思う」と答えるほど高い値である。

因子数は固有値1.00以上とした。因子負荷が0.40以上を太字とした。

#### 図表作成時の注意点

- ・ 図表だけを見て、第3者に内容が分かるようにするのが図表作成の基本。表の注やタイトルを詳しく書くのが重要なこつである。抽出法や回転法を書くこと。
- ・ 図表には必ず番号をつける。表タイトルは表上に、図のタイトルは図下に書く。
- ・ 上記の表で、各因子負荷を縦の二乗和は因子寄与、横の二乗和は共通性になる。因子寄与が1とは、因子の持つ情報量が、もとの質問項目1つと同じ大きさという意味。上記の例では、もとの7つの質問のうち、3.61個分をモデルで説明したという意味になる。寄与率は、もとの変数7個を100%として、自分で計算する。
- ・ 因子には自分で名前を付ける（自由に解釈を加える）。

このような表を散布図で表した場合は、以下のようなになる。成分行列をエクセルなどでグラフにすればよい。エクセルの場合、点の説明は自分で文字を打たなくてはならない。点をクリックして文字を打ち込み、エンターキーを押せばよい。



確証的因子分析の結果は、以下のような図にまとめるとよい（数字は架空の例）

観測変数（実在変数）は四角で、潜在変数は丸で書くのが慣例

確証的因子分析の場合、以下のように、図タイトル中にモデル全体の各種の適合度係数（モデルのフィッティングの良さを表す数字）などを詳しく書くと良い。

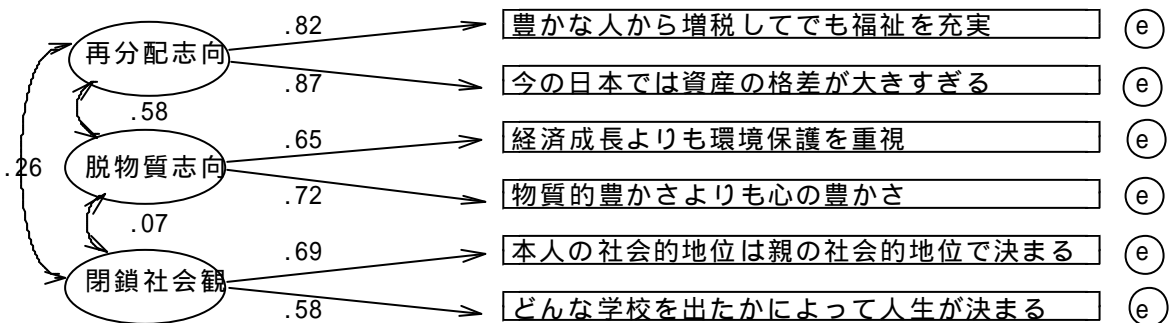


図4. 政策志向と政治の有効性感覚の確証因子分析 1996年 × × × 調査 男性データ  
n=764 欠損値は個人単位で削除 再尤法で推定 df=34 p=.12  
Chi-square=154.97 GFI=.98 AGFI=.95 CFI=.91 AIC=218.97 RMR=.09 RMSEA=.04

### 3. 分析時の注意点

#### 3.1. 分析の前に必ず欠損値処理をすること

多くの場合、欠損値は9か99。SPSSの場合、missing valuesコマンドを用いる。

回答が2桁の場合、欠損値99である。まず単純集計をとって確認するとよい。分析は、まず単純集計を見て分布を確認することが重要である。

#### 3.2. 分析の前に変数の向きを必要に応じて逆転し、わかりやすく設定する

分析を行う前に、原則として、すべての変数を、数字が大きいほど肯定になるように直すこと。数字が小さいほど肯定となる変数が混ざっていると、とても分かりにくい。変数逆転後の新変数を使って分析すればよい。また、すべての説明変数が量的変数である。カテゴリ変数は用いてはいけない。

#### 3.3. モデルの考え方とデータ人数

似たような変数が多数あれば因子分析の結果は安定するが、数学的には、単に変数を要約しただけであり、因子が存在するとは言えない。心理学では、大学生のみを対象としたデータを用い、因子得点を被説明変数として分散分析を行うことが多い。社会学では、対象者は均一ではないし、質問数は多くないので、因子分析の結果は不安定になりがちで、あまり信用されないことも多い。因子を作るのに適切な質問項目があるならば、重回帰分析等を行う前の参考として、因子分析を行うと良いだろう。

つまり、因子分析は、あくまでも現実の変数群を要約したにすぎないので、因子が実在するかどうかは不明確なのである。因子数や推定法は何種類かあり、それにより、さまざまな因子が出てきてしまうことに注意すること。また、データ人数は、最低400人程度はないと、因子分析で信用できる結果を得ることはできない。人数が少ないと結果は不安定である。

ただし、因子分析により、変数間の関係（各質問項目のどれとどれが似ているか）の全体像を理解し、その上で重回帰分析など詳しい分析をすると、分析を適切に進めやすいという点で意味がある。また、似たような質問項目が多いときは、それらを要約して1因子にすると分析しやすい。もっとも社会調査データの場合は、要約して因子得点を作るよりも、単なる合計得点の方が分かりやすいし、その方が分析しやすいことも多い。

### 4. 課題

質問群を見て、いくつかの因子が背後にあることを考え、自分の好きな質問項目（変数）を使って分析する。5変数以上、できれば10変数以上を使うと良い。3変数を3因子に要約しても意味がないので、なるべく、5変数を2因子に要約等、もとの観測変数の数は多い方が良い。分析時には、上記の注意点を気をつけること。

結果を表にまとめ、因子に名前をつける。数行の解釈をつけて提出する。

## 5. 参考 的確な構成の論文を作るコツとは何か - 論文の流れを明確に

論文や報告書を書くときは、文章の流れを明確にするとよい。つまり、まず分析目的や仮説を明確に書く。次に、主な被説明変数となる質問項目の、単純集計や男女別クロス集計結果を（表かグラフで）載せる。基本的な分布に関する図表があると、具体的にどのような質問項目を使ったのか、読み手にとって分かりやすくなるので良い。

次に、変数間の相関（または因子分析結果）の表を載せる。これにより、用いた変数間の関連の構造が、どのようなものかが分かる。その後、重回帰分析や分散分析（またはパス解析やSEMなど重回帰分析系の分析結果）を載せる。

最後に、主な変数について、さらにクロス集計結果を載せてもよい。それにより、主要な説明変数と被説明変数の間に、事実としてどのような関連があるのかが分かる。

このような流れで、計量分析の結果を載せた論文は、読み手にとって分かりやすく説得力がある。

調査データの分析の場合は、分析目的にそって、まず、ひとつの被説明変数Yを決めることが、論文のコツである。そして、その原因となるXとは何かについて、分析結果を載せ、解明していくと良いのである。まず、論文冒頭で、分析目的を明確に書くこと。当然だが、目的と仮説、分析結果、最後の結論が、一貫していれば分かりやすい論文である。これは論文の基本だが、現実には一貫していない論文も多いのである。

なお仮説とは、原因と結果を含む文である。原因については、説明変数が何かを書けばよい。説明変数は、基礎的な質問項目や、心理学的項目（態度や意識）、社会学的項目（収入や財産、学歴、職業、社会的地位）に関する項目、社会構造に関する項目（居住地の都市度や産業化に関する項目）など、複数の種類の項目が、あるとよいだろう。一部しか説明変数として使っていない場合は、視野が狭い論文ということになる。ただし、複雑すぎる分析はよくないし、目的をしぼるために、あえて説明変数の種類を少なくすることも、時には必要だろう。

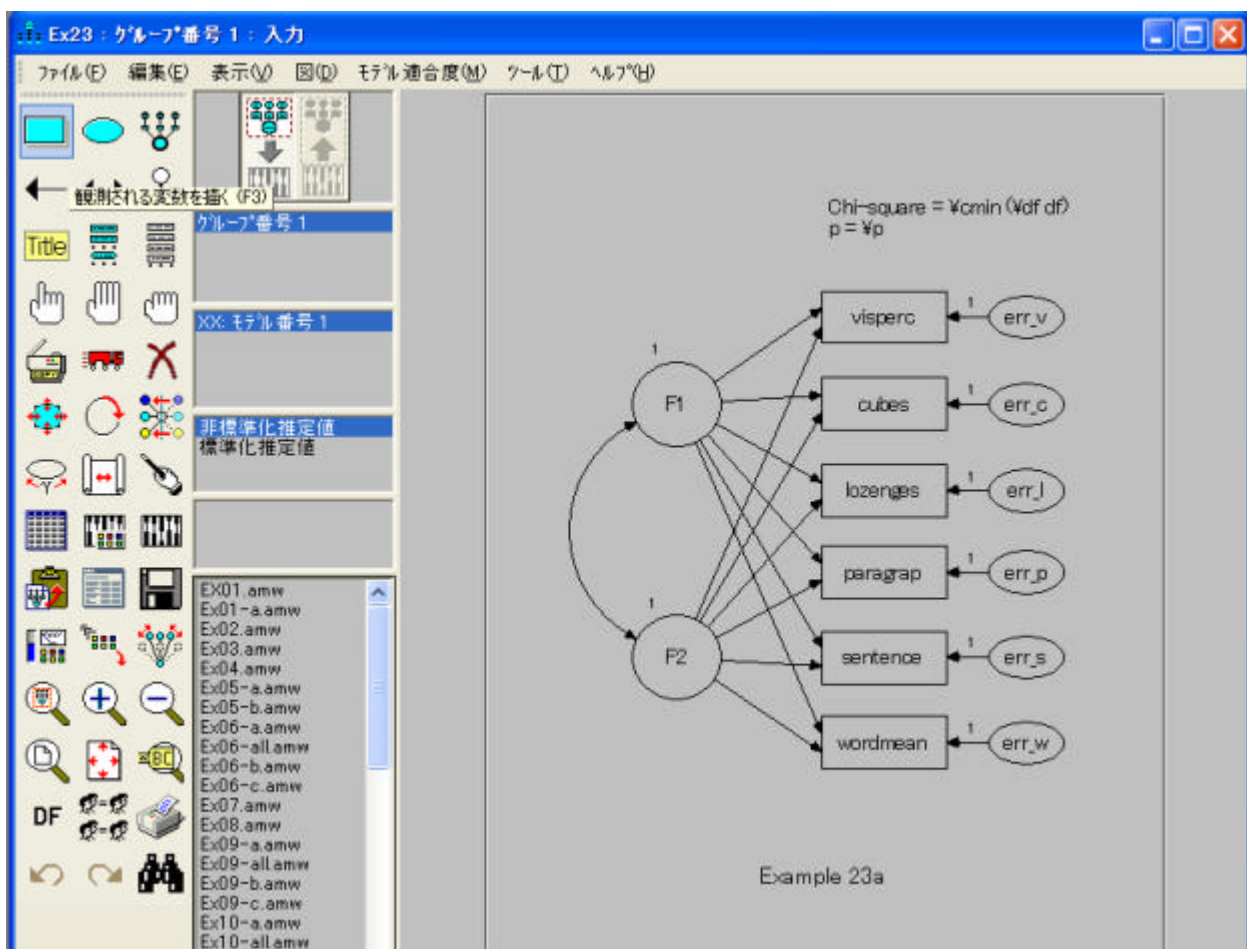


## 6. 発展版 - AMOSの使い方(一部のPC教室とノートPCのみ入っている)

エイモスはSPSS社が売っているオプションのソフトだが、SPSS本体とはまったく別個に動かすことができる。確証的因子分析や、因子を使った重回帰分析やパス解析等、構造方程式モデル(セム、共分散構造分析とも言う)と呼ばれる高度な分析を行うことができる。

1) シンタックスとテキスト形式データファイルを用いて、SPSS形式データファイルをあらかじめ作っておく。この際に、欠損値はすべて除いたデータファイルを作るとよい。

2) AMOSの画面で、自分の好きなモデルの図を書く



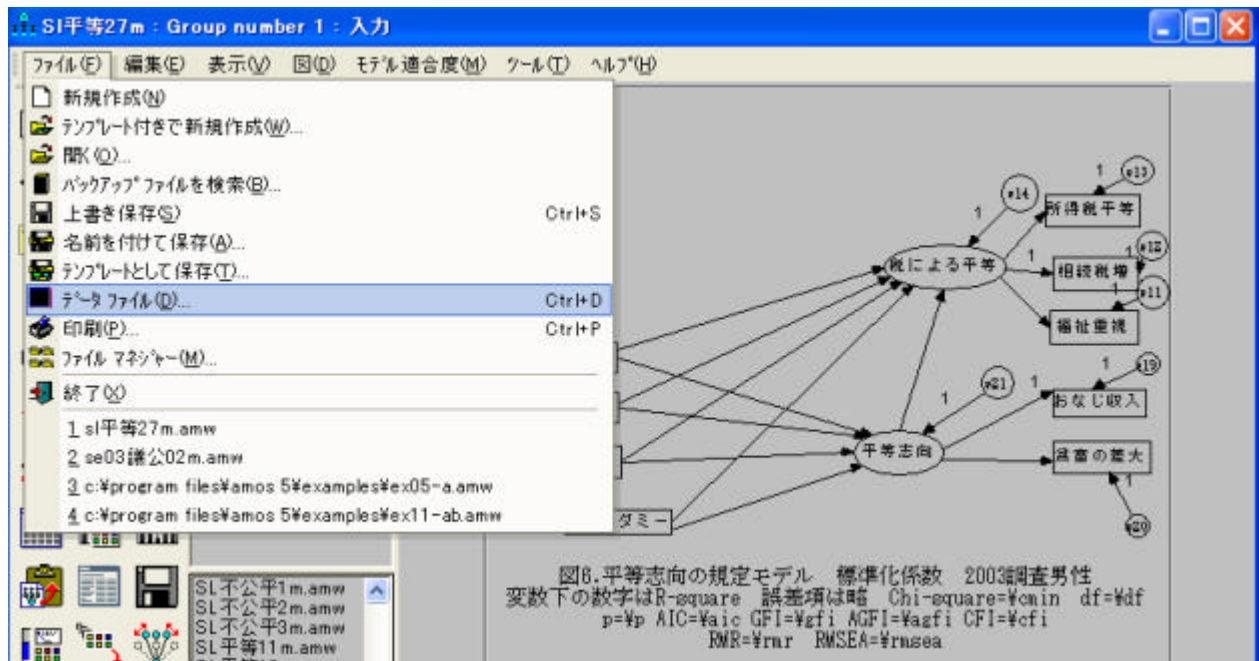
AMOSを起動しモデルの図を自由に書けばよい。四角の変数(観測変数)は、データファイルの中にある変数名を、正確に書く。四角を書いたら右クリックして、変数名などを書けばよい。矢印の刺さっている変数には、必ず誤差項からの矢印も刺さることになるので注意。この矢印を忘れるとエラーが起きる。誤差項をつけるボタン(四角の上に が付いているボタン)を使って、四角い変数に誤差項をつけると良い。

なお画面上の「ツール」をクリックしてマクロを選び、潜在変数を名付けるを選ぶと、誤差項に自動的に名前を付けてくれる。また、ペンのボタンを押してから各変数をクリックすると、図を整えてくれる。

3) 画面上の「表示」をクリックして「分析のプロパティ」を選び、分析の詳しい内容を

決める。ふつう、分析法は「最尤法」、標準化推定値と重相関係数の二乗を選べば、あとはとくに変更しなくてよい。ただし、データファイルの中に欠損値がある場合は、平均値等の推定を選ばないと、分析ができない。

4)画面上「ファイル」をクリックして「データファイル」を選び、「ファイル名」ボタンを押して、分析に用いるSPSSデータファイルを指定する。



5)自分の書いたモデルは、画面上「ファイル」をクリックして名前を付けて保存する。

6)モデルの図を書き終わったら分析する。画面上「モデル適合度」をクリックして「推定値を計算」を選ぶか、または分析ボタン（鍵盤のようなボタン）を押せばよい。上記の点に注意し、的確にモデルを書いているならば、OKと出る。

7)結果表示のボタン（画面左上付近の、上向き矢印の大きなボタン）を押すと、数字が図の中に書き込まれる。このボタンの下にある「標準化推定値」をクリックすると、標準化された係数が出る。四角い変数の上には $R^2$ が表示される。この状態で図全体をコピーをすれば、図をワード等の画面に貼ることもできる。数字の位置を動かしたいときは、矢印や四角などを右クリックして「パラメータ移動」を選ぶ。

8)画面上「表示」をクリックしてテキスト出力で、より詳しい分析結果も表示される。パラメータ推定値で「確率」と出ているのが有意水準（有意確率）である。

9)有意でないパス（矢印）を削除し、分析を繰り返す。

図を書く際に、因子から観測変数へのパスのうち、どれか一つは必ずパラメータを1とする（このパラメータを固定母数という）。そうでないと、どれを基準に因子を作れば良いのか、ソフトが判断できない。パスを右クリックしてパラメータを入れればよい。

エラーが出て分析結果が出ない場合、その原因の多くは、自分で書いた観測変数名が間

違っている、上記のパラメーター1を忘れた、あるいは誤差項をつけ忘れたことである。また多重共線性が強い場合や、もともと観測変数間の関連が小さいので因子を作ることができない場合等、無理なモデルを設定した場合はエラーが出て、分析結果が表示されないことがある。いくつかの因子をどのように作るべきか等を、よく考えてモデルを作れば良い。

元のデータファイルに欠損値がある場合もエラーが出る。画面上「表示」をクリックし、「分析のプロパティ」を選び、「平均値と切片を推定」にチェックを入れれば、欠損値を含むデータも分析できるのだが、この場合、各種の適合度係数は出なくなるのが欠点である。事前に、SELECT文を使って欠損値を除いたSPSSデータファイルを作った方がよい。

重回帰分析の図を書いて分析しても良い。この場合、普通の重回帰と違い、最少二乗法でなく最尤推定法で計算した結果が出る。因子を使った重回帰分析やパス解析が簡単にできることが利点である。

## 7. 参考文献

市川伸一・大橋靖雄・岸本淳司・浜田知久馬．1993．『SASによるデータ解析入門 第2版』東京大学出版会．

海保博之編著．1985．『心理・教育データの解析法10講 基礎編』福村出版．

狩野裕・三浦麻子．2002．『AMOS, EQS, CALISによるグラフィカル多変量解析（増補版） - 目で見る共分散構造分析』現代数学社．

丘本正．1986．『因子分析の基礎』日科技連出版社．

芝祐順．1979．『因子分析法 第2版』東京大学出版会．

豊田秀樹．1992．『SASによる共分散構造分析』東京大学出版会．

豊田秀樹．1998．『共分散構造分析 - 構造方程式モデリング 入門編』朝倉書店．

豊田秀樹編．1998．『共分散構造分析 - 構造方程式モデリング 事例編』北大路書房．

豊田秀樹編．2007．『共分散構造分析 - 構造方程式モデリング Amos編』東京図書．

豊田秀樹・前田忠彦・柳井晴夫．1993．『原因を探る統計学 - 共分散構造分析入門』講談社ブルーバックス．

渡部洋編著．1988．『心理・教育のための多変量解析法入門（基礎編）』福村出版．

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．

群馬大学 青木繁伸氏 <http://aoki2.si.gunma-u.ac.jp/lecture/mva.html>

大阪大学 狩野 裕氏 <http://koko15.hus.osaka-u.ac.jp/~kano/lecture/>

香川大学 堀 啓造氏 <http://www.ec.kagawa-u.ac.jp/~hori/spss/factorlink.html>

8 . おまけ 因子分析のS A S プログラム - 市川他 ( 1993 : 209 , 222 など ) を参照

因子数を指定しないとき、nfact=は省略してよい。固有値1以上の因子数となる。

```
proc factor rotate=varimax;
```

```
var q6a--q6g;
```

因子数を 4、回転法をバリマックス回転と指定した例。

```
proc factor nfact=4 plot rotate=varimax;
```

```
var q6a--q6g;
```

主因子解にするときはmethodを指定 ( pp.215- 参照 )。何も指定しないと主成分解。

```
proc factor method=prinlit nfact=4 plot rotate=varimax;
```

```
var q6a--q6g;
```

新データセットD2を作る例。因子得点が factor1、factor2などの名前で新変数となる。

```
proc factor rotate=varimax out=d2;
```

```
var s15a--s15f s16a--s16e;
```

```
proc sort data=d2; by f26;
```

```
proc means data=d2;
```

```
var factor1 factor2 factor3; by f26;
```