

# 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

目的 — 複数の変数の背後にある、隠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）。

量的変数（連続変数）のみを用いることができる。

具体例 — 通常、国語の成績が良い人は社会や英語も成績が良い。これは、表面的な点数の背後に、文科系能力や理科系能力のような、隠れた要因（総合的能力）が存在する、と考えることができる。このことを図で表すと以下のようなになる。

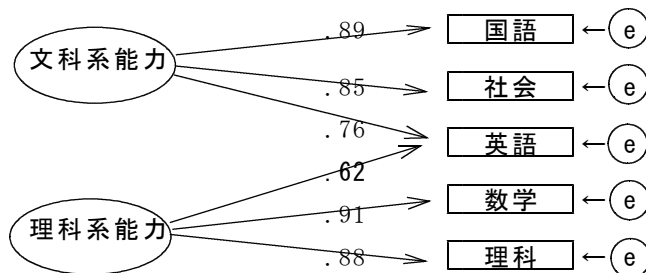


図1. 5科目試験成績の構造に関する因子分析結果（架空例）

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、総合的能力を直接測定することはできない。そのため3科目やっていると見える。直接測定できない要因のことを、因子または潜在変数 (latent variable) という。

調査データでの具体例 — 価値観や社会意識に関する4段階回答の質問が10問あり、それらを分析すると10問が2グループに分かれることがある。その場合、それらの質問は、実は2つの隠れた要因（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。上記の図の各科目を、各質問項目に置き換えて、具体例を考えてみると良い。

### 1.2. 因子分析の考え方と基本モデル

因果連関図では通常、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）Xを長方形でかく。以下は上記の図を記号で表現したもの。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、と考えるのがポイント。 矢印の向きに注意。考え方としては、あくまでもFが原因でXが結果である。

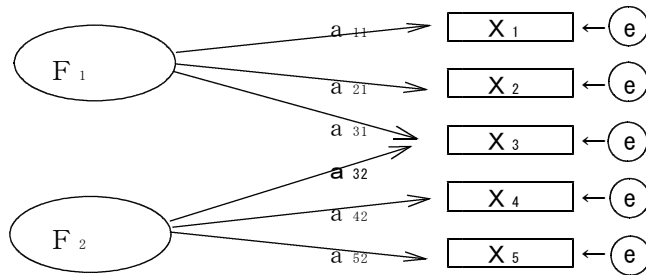


図2. 記号による表現

$X_1$ について数式で表現すると以下の(1)式のようになる。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。

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

数式で書くと難しくみえるが、これは上記の図と同じもの。

Xが因子Fと誤差項eによって規定されていることを表している。

Fによって説明される部分 — **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 — **独自性**

例えば、 $h^2$ が0.40ならば、因子Fにより、ある観測変数Xは、分散の40%が説明されていることになる。共通性と独自性を足すと1となる (昔は誤差を特殊因子と呼んだこともあるが、そのような名前だったというだけであまり意味はない)。

実際の分析においては、因子Fは分析後に出てくるので、どのような性質の因子かを自分で解釈し、因子に名前をつける。これは、自由に解釈してつければ良い。普通、まず回転しない因子を計算し、その後に回転を行い、回転後の結果のみを用いる。

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1.6の場合、元の質問3.8個分の分散を2因子で説明したことになる。回転前の値を固有値というが、回転後は、固有値ではない (回転については後述)。

### 1.3. 因子分析の特徴と種類

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。実在する変数については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 (因子パターン) である。 これは直交解 (各因子が無相関とした分析結果) の場合、相関係数と同じ値である。

因子分析は、数学的には以下のような複数の観測変数間の相関を要約しているだけである。 このような相関行列のみあれば、元のデータが無くても分析はできる。実際の因子分

析とは、隠れている因子を発見するというよりは、手元にあるデータを要約するために便利な尺度をいくつか計算して作り、それを因子と呼んでいるだけである。例えば最初の図で、文科系に関する試験ばかりがたくさんあった場合は、因子寄与の大きな文科系因子が作られるが、だからと言って文科系因子が確実に存在するとか、理科系因子がない、ということではない。因子分析の結果を見て、因子の存在を発見したということとはできない。

表0. 変数間の相関行列の例

	問 1	問 2	問 3	問 4	問 5
問 1		.87	.82	.28	.34
問 2			.74	.34	.28
問 3				.32	.16
問 4					.85

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。

因子分析には大きく次の 2 種類がある。

- 1) 探索的因子分析 — 従来型の因子分析。SASやSPSSなどで可能。
- 2) 確証的因子分析 — 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 Model, 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、SASのproc calisなどの分析ソフトを用いる。

#### 1.4. 探索的因子分析の推定法 (因子抽出のための計算方法 詳しくは村瀬他 (2007) 参照)

因子は架空のものであり、確実に存在する因子を得るための方法はない。そのため、各種の計算法が存在するが、どれも恣意的なものであり、因子がいくつか出たからといって、因子が必ず存在するとはいえない。つまり推定法に決定版はない。主な推定法は以下。

主因子法 (Principal Factor Method) : 第一因子最大化

- 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なるべく均等に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。

ただしデータ人数が十分に大きければ主因子法の結果と変わらない。

## 因子数の決定

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良い。村瀬他(2007:267)のようにスクリープロットを見て因子寄与の大きなもののみの因子数としても良い。ただし、因子寄与が1未満などの小さな因子を作っても、あまり意味がない。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、名前をつければ良い。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(元の変数1個くらいの情報はある)因子、という意味である。なお、元々5変数を、5因子に要約しても、要約した意味がないから、ある程度少なめの因子数にすると良い。

### 1.5. 探索的因子分析の回転法 村瀬他(2007, p. 257)の図など参照

因子の意味は、因子と関連する観測変数をもとに解釈する。そこで、解釈しやすいように、因子負荷  $a$  の散布図における分析結果を、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。例えば図2では、 $F_1$ から $X_4$ や $X_5$ への矢印はない( $X_3$ は2つの因子に規定されているので厳密な単純構造ではない)。普通は、回転後の結果のみを用いる。回転前の結果を論文で使うことはない。主な回転法は以下。

- 直交回転 ー各因子間の相関は0となる。普通、バリマックス回転という方法が用いられる。全因子によって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。
- 斜行回転 ー各因子間の相関がある。普通、プロマックス回転という方法が用いられる。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。

初心者は、主成分法、バリマックス回転、因子数は固有値1以上として、まず分析すれば良い。この場合、因子間の相関は0なので、結果を解釈しやすい。固有値の意味については、参考文献を読むこと。ただし、因子同士にまったく相関がないというのは、現実にはあまりないことなので、モデルの前提として厳しすぎる(不自然すぎる)ともいえる。この後に、主因子法のプロマックス回転(最近はこれが用いられることが多い)などを試してみると良いだろう。

ただ、因子同士の関連がかなり強いならば、複数の因子を作らず、1因子で良いともいえる。したがって、プロマックス回転を用いたモデルだと、どのモデルが良いのか判断するのは難しい。しかしプロマックス回転の方が、データへのあてはまりが良い(自然なモデルになる)場合もある。もともと因子間の相関が小さいと考えられる場合は、バリマックス回転で良いだろう。なお直交解での因子負荷は、因子と観測変数との相関係数である。斜行回転の場合の因子負荷とは、観測変数を $Y$ とした場合の、各因子の回帰係数である。

### 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、実在する変数を要約することが目的であり、背後にある要因の抽出は目的ではないため、回転はしない。複数の観測変数をすべて要約し、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。以下の散布図でいうと、多くの点（観測変数）の真ん中に引いた線が第一主成分、それと直角に交わるものが第二主成分である。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。

## 2. SPSSによる分析

### 2.1. SPSS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）を選べば良い。

#### シンタックス例 1

```
FACTOR                                ←FAC のみでも動く
/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 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
```

#### 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 この行を省略するとPCになる
/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。なくても良い
```

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→因子分析を選んで良い。回転法は、初心者は「バリマックス」を選べば良い。オプション

ンボタンを押して、因子負荷プロットの図を出すことができる（村瀬，2007：262の図）。

なお分析前に、用いる変数は欠損値処理済みか、変数の方向は適切か、各変数の度数分布表を出してよく確認すること。

## 2.2. 分析時のポイント

社会調査の意識や価値観に関する項目では、男女別（かつ年齢別）に分析した方が、経験的には、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ている。

因子数をいくつにするかは、とくに正解はないので、色々と結果を出し、自分で解釈しやすい因子数とすれば良い。

## 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良い。 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。因子寄与率は、もとの変数が7個ならば、7を100%として、因子寄与は何%となるか、自分で計算する。因子寄与合計と共通性合計は同じ値になる。以下の表を見て確認すること。

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	<b>0.71</b>	0.00	0.50
以前からなされてきたやり方を守る	<b>0.77</b>	-0.15	0.62
伝統や習慣にしたがうべき	<b>0.67</b>	-0.15	0.48
指導者や専門家に頼るべき	<b>0.69</b>	-0.12	0.49
政治のことは理解できる	-0.24	<b>0.72</b>	0.58
政治のことはやりたい人にまかせない	-0.28	<b>0.59</b>	0.43
国民の意見は国の政治に反映されている	0.14	<b>0.71</b>	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。

因子分析の値に、有意確率の\*を付けることはない。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。

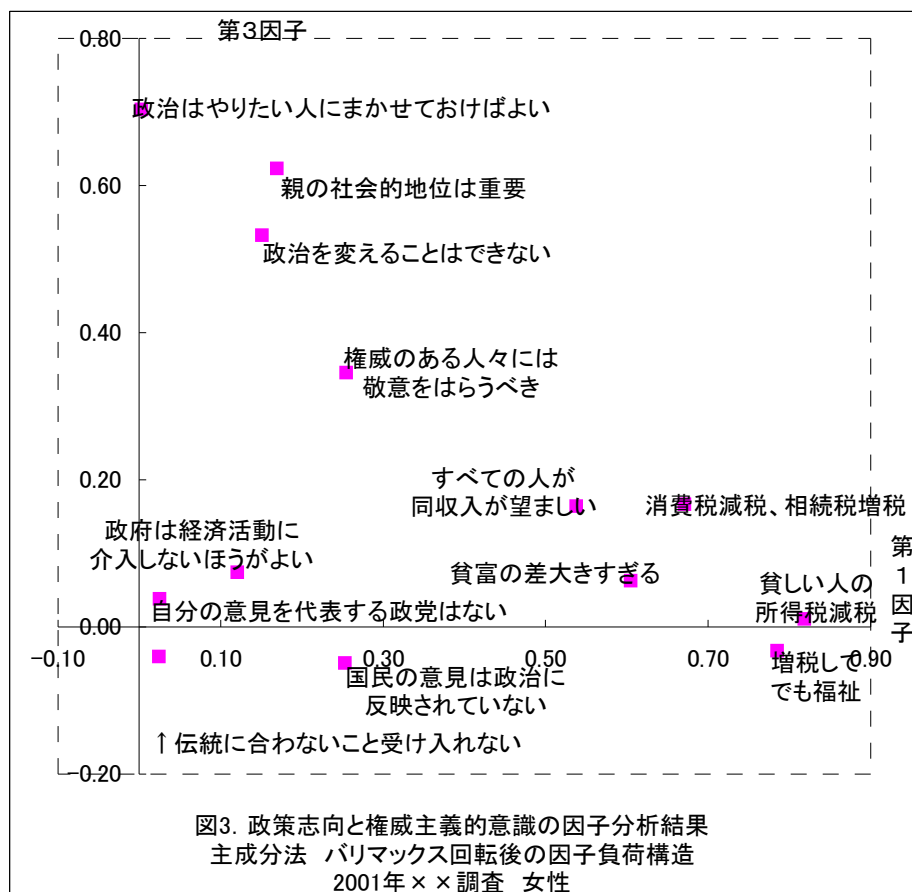
回転後の共通性が出ない場合、自分で因子負荷の二乗和を計算する。上記で、2つの因子負荷を二乗して足すと共通性になる。

エクセルで二乗和を出す関数 `( )`内はコロン `=SUMSQ(B5:C5)`

★図表作成時の注意点

- ・図表だけを見て、第三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。
- ・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2因子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）。
- ・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を書く。

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良い。エクセルの場合、バージョンによっては、点の説明をつけたい時はオプションでデータラベルを付けるか、あるいは、テキストボックスを挿入して文字を自分で打ち込む。



エクセルで散布図を作る時は、数字部分のみをマウスで囲んでからグラフ作成ボタン（挿入→グラフ）を押すこと。一太郎附属の表計算ソフトでグラフを作るなどしても良い。

この図は第一因子と第三因子を2次元グラフにした例である。3次元グラフは分かりにくいので、個別に2次元の散布図にすると良い。

確証的因子分析の結果は、以下のような図にまとめると良い（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。

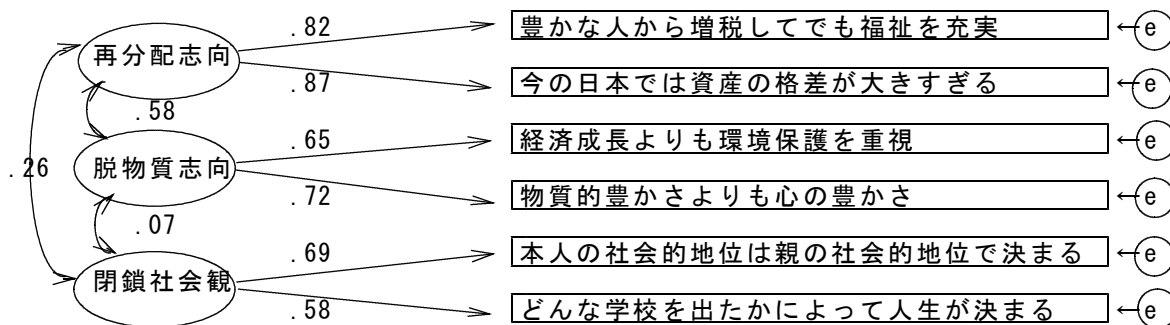


図4. 政策志向と政治的有効性感覚の確証因子分析 1996年×××調査 男性データ  
 n=764 欠損値は個人単位で削除 再尤法で推定 df=34 p=.12  
 Chi-square=154.97 GFI=.98 AGFI=.95 CFI=.91 AIC=218.97 RMR=.09 RMSEA=.04

### 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

社会調査データは多くの場合、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。まず度数分布を出して確認すると良い。データ分析は、まず分布を確認し、偏りや欠損値をみ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、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る。また、すべての説明変数が量的変数である。

#### 3.3. モデルの考え方とデータ人数

似たような変数X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Yとして分散分析を行うことが多い。しかし大学外で行う社会調査では対象者は均一ではないし、同じような質問をあまり多く設けることはない。そのため、因子分析の結果は不安定になりがちで、あまり信用できない。結果は砂上の楼閣といっても良いだろう。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(表0のような相関行列の要約)にすぎないので、因子が実在するかどうかは不明である。厳密には、因子分析により因



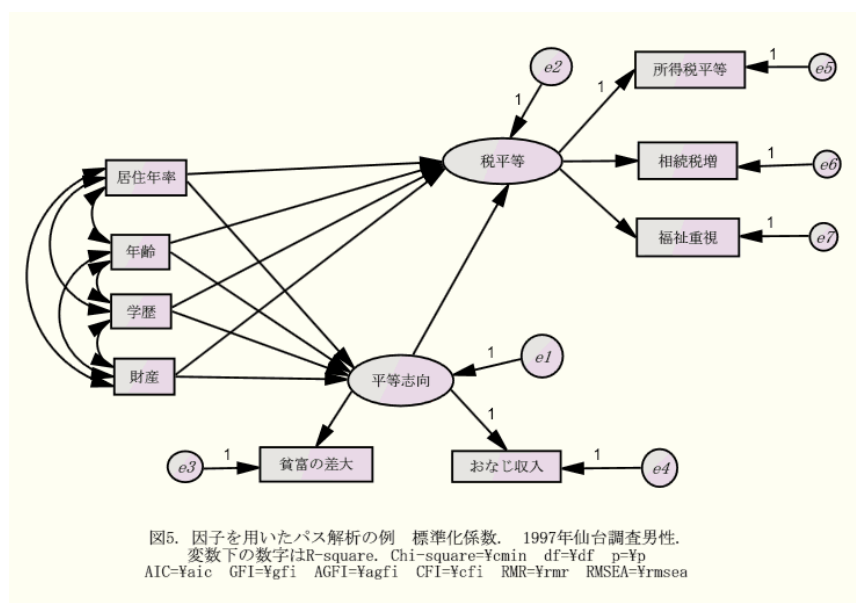
子を発見したという言い方は不適切である。因子数や推定法は何種類もあり、それにより、さまざまな因子が出てきてしまうことに注意すること。また、データ人数は数百人以上ないと、結果は不安定であり、信用できる分析結果にはならない。

ただし社会調査データでも、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。重回帰の前か後に、回転後の因子分析結果を載せても良いだろう。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、老人を含め一般人が対象の社会調査の場合、回答者の疲労を考えると、似ている質問をたくさん設けることはできない。また回答者も多様であるから、適切な因子を作れないことも多い。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く、その方が分析しやすいことも多い。

### 3.4. 構造方程式における因子の考え方

AMOSなどの構造方程式モデルのソフトを用いて、確証的因子分析をする場合、因子から観測変数へのパスはどれか一つに係数1が付く。また、矢印が刺さっている変数には、必ず誤差項が付く。例えば以下の図で「税平等」という因子（楕円形で描いた物）は、3つの観測変数（実変数）からなるが、3つのパスのうち1つだけ推定前に係数1を付ける（3つのうちのどれでも良い）。また、誤差項から出ているパスにも1をつける。矢印の向きは、楕円から四角い観測変数へ、となる。つまり考え方としては、因子が観測変数を規定していることになる。

AMOSを用いる場合、画面上の「表示」をクリックして、分析のプロパティを選ぶことができる。普通、推定法として最尤法を指定し、「出力」タブで「標準化係数と重相関係数の平方」を指定すれば、あとはとくに指定しなくて良い。ただし、データファイル中に欠損値があると、AMOSは分析結果を出せない。その場合、「平均値と切片を推定」にチェックを入れると、AMOSが欠損値を推定した上で分析結果を出す。ただこの場合、分析結果が必ずしも適切ではないし、適合度係数が出ない。社会調査データの場合、対象者があまり均質でないし、欠損値の推定は、うまくいく保証はない。まずSPSS側で、欠損値を削除したデータファイルを作成



してから、それ使って分析した方が良い。

### 3.5. 構造方程式モデルにおける 因子分析とパス解析

ところで、基本的な確証型因子分析は、以下のようなモデルである。つまり、図5のように、因子に対して刺さっているパスはない（したがって因子に誤差項は付かない）。しかし、以下のような因子分析のモデルにおいては、因子には、他の変数からの影響がない（規定されていない）ということになり、因子間の相関（両方向矢印）が存在する。モデルは、分析目的

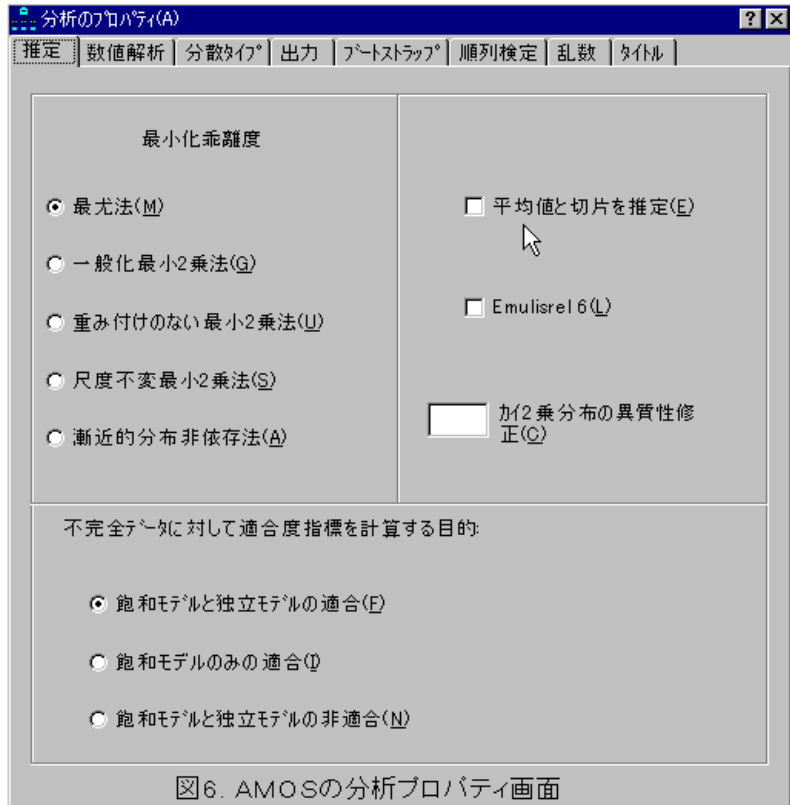
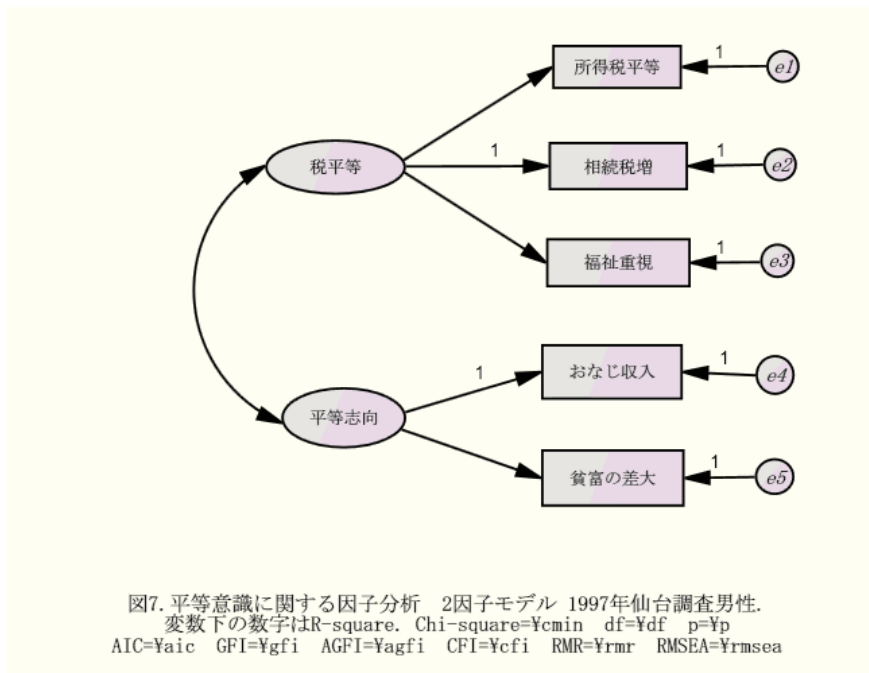


図6. AMOSの分析プロパティ画面

に応じて分析者が自由に決めればよく、どちらのモデルでも良い。心理学的な因子分析のモデルは、もともと図7のようなモデル（基本的な因子分析モデル）であった。心理学における実験データの多くは、学生のみから取ったデータなので、年齢や学歴は均質であり、図5のような、基本属性や社会的地位を考慮した分析はできない。しかし大学外から取った社会調査データの場合、因子が回答者年齢やその他の変数の影響を、まったく受けていない、ということは、現実にはあまりない。したがって、図5のように、因子が何らかの社会学的な変数に影響されているというモデルの方が、モデル全体の当てはまりがよく、適合度係数はより良くなりがちであろう。少なくとも、回答者年齢や教育年数など、何らかの基本属性を考慮したモデルの方が、現実的な場合が多い。ただこれは、因子の性質や、データの内容にもよるので、一概には言えない。



#### 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。7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、7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、上記の注意点に気をつけること。

回転後の結果を表にまとめ、因子に名前をつける。結果に関して数行の解釈の文章も書いて提出する。上記の結果のまとめ方を参考に、適切な表を作ること。

初めに、自分の名前やテーマ、用いたデータも必ず書く。因子抽出法や回転法も書く。

AMOSでは、サンプルモデルをかき換えると良い。因子から観測変数へのパス1箇所固定母数1をつける必要がある。また、矢印の向きや、誤差項を忘れるとエラーになる。

#### 5. 課題 発展版

AMOSを使って確証的因子分析を行う。因子を使ったパス解析のモデルを作って分析する。例えば、上記の図5のように、左から右へ、因果関係の流れが分かりやすいように図をかき、何らかのデータを読み込んで、分析を行う。その際、因子を2つ以上作ること。1つの因子は、2つ以上の観測変数（実在する変数）からなる。また調査データの場合、左端に、年齢や教育年数などの基本属性を、いくつか用いる方が、モデル全体の当てはまりがよくなることが多い。

なおパス解析では、モデル途中にある因子同士の因子間相関は、誤差項間にかく。

#### 6. 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良い。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。その後に、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良い。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良い。説明変数は、基礎的な質問項目や、心理学的要因（態度や意識）、社会学的要因（収入や財産、学歴、職業、社会的地位）に関する項目、社会構造に関する要因（居住地の都市度や産業化に関する項目）など、複数の種類の要因があると良い。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、目的を絞るために、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。

## 参考文献

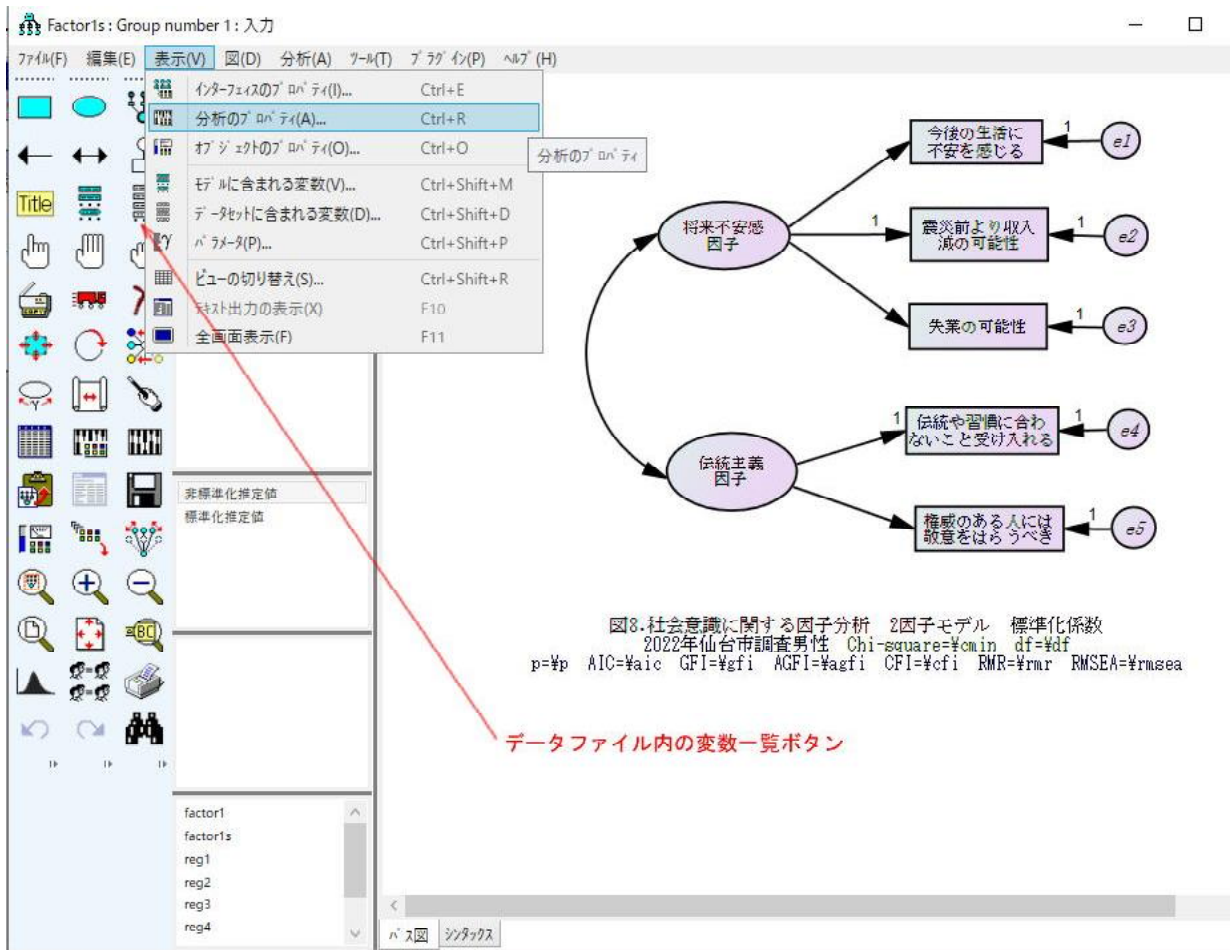
- 足立浩平. 2006. 『多変量データ解析法 ー心理・教育・社会系のための入門』ナカニシヤ出版.
- 市川雅教. 2010. 『因子分析』朝倉書店.
- 市川伸一・大橋靖雄・岸本淳司・浜田知久馬. 1993. 『S A Sによるデータ解析入門 第2版』東京大学出版会.
- 海保博之編著. 1985. 『心理・教育データの解析法10講 基礎編』福村出版.
- 狩野裕・三浦麻子. 2002. 『AMOS, EQS, CALISによるグラフィカル多変量解析（増補版） 一目で見る共分散構造分析』現代数学社.
- 丘本正. 1986. 『因子分析の基礎』日科技連出版社.
- 豊田秀樹. 1992. 『S A S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 ー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 ー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 ー構造方程式モデリング Amos編』東京図書.
- 豊田秀樹. 2012. 『因子分析入門』東京図書.
- 豊田秀樹. 2014. 『共分散構造分析[R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 ー共分散構造分析入門』講談社ブルーバックス.
- 村瀬洋一他編. 2007. 『SPSSによる多変量解析』オーム社.
- 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.
- 柳井晴夫・繁梶算男他. 1990. 『因子分析 ーその理論と応用』朝倉書店.
- 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://www.sigmath.es.osaka-u.ac.jp/~kano/>
- 香川大学 堀 啓造氏 <http://www.ec.kagawa-u.ac.jp/~hori/spss/factorlink.html>

## 7. 発展版 ーAMOSの使い方

エイモスはSPSSのオプションとして売っているソフトだが、SPSS本体とはまったく別個に動くものである。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（SEM、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 (ファイル名にsavとつくもの) 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良い。SPSSの欠損値処理はAMOSでは意味がないので、SELECT文を使い、欠損値がある人を削除して人数を減らしたデータファイルを作る(後述)。

2) AMOSの画面で、自分の好きなモデルの図をかき、矢印の向きに注意する。また、矢印が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。以下の図で、どの矢印に1が付くか注意する。

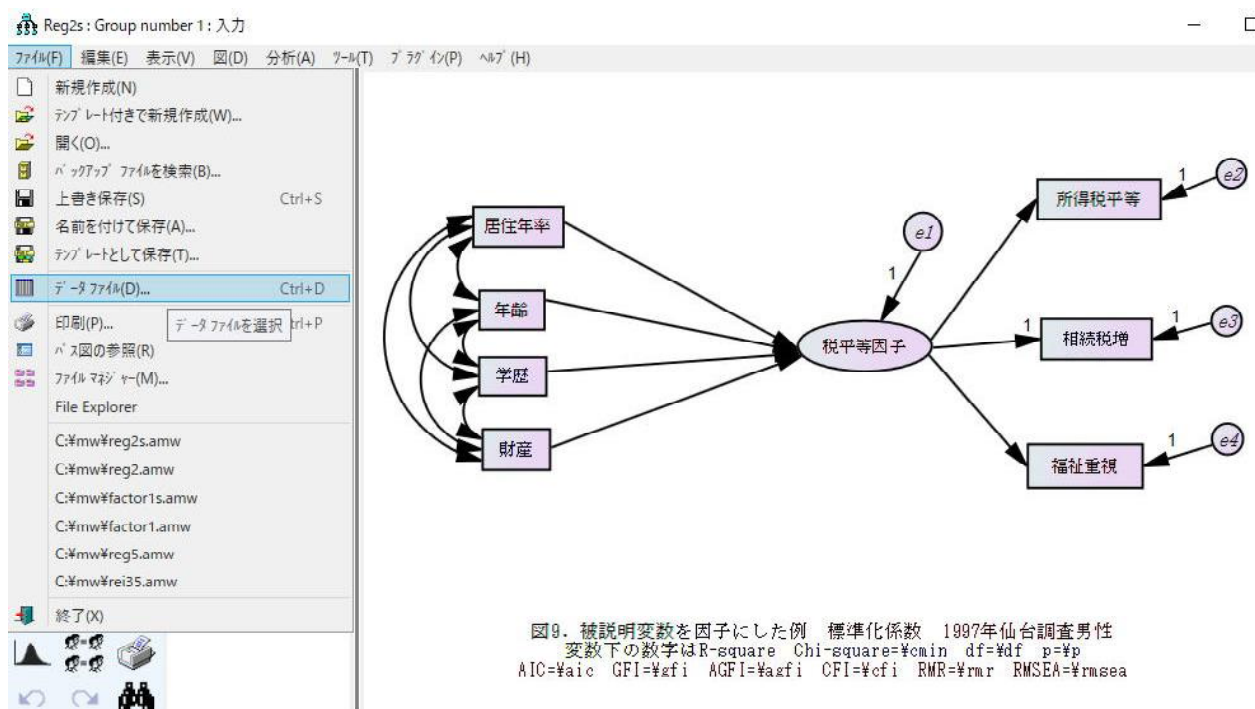


AMOSを起動しモデルの図を自由にかく。上記の因子分析の図の場合、矢印の向きがどうなるか理解すること。因子から四角の変数(観測変数)への方向となる。観測変数は、データファイルの中にある変数名を、正確に書く。四角を書いたら右クリックして、変数名などを書けば良い。また、ある因子から観測変数への矢印のうち1つは、必ずパラメーターを1にする。矢印を右クリックしてパラメーターのボックスに1を入れれば良い。誤差項から観測変数へのパスにも必ず1を付ける。

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良い。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形式データファイルを指定する。

5) 自分の書いたモデルの図は、「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）を押せば良い。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窓の「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が表示される。パラメーター推定値で「確率」と出ているのが有意水準（有意確率）である。どのパスが有意なのかを確認する。

9) 有意でないパス（矢印）を削除し、分析を繰り返す。

## 8. AMOSでエラーが出た場合の対処

モデルの図を書く際に、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ーを1とする（このパラメーターを固定母数という）。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ー1を入

れば良い。

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か、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。これは、あまりやらない方がよい。事前に、SELECT文を使って欠損値を除いたSPSSデータファイル(sav)を作った方がよい。モデルに10の観測変数がある場合、SPSS上でSELECT文を10行書いて実行してからデータファイルを保存すればよい。

まずSPSSのシンタックスウィンドウにて、変数逆転と、SELECT文を書いて実行する。その後、各変数の度数分布をみて、欠損値がないことを確認する。その上で、人数を減らしたデータファイル(sav)を、名前を付けて保存してから、AMOSで利用するとよい。

## 9. 因子を用いた重回帰分析モデルの作り方

重回帰分析の図を書いて分析してもよい。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。上記の図9は、被説明変数Yのみを因子（3つの観測変数を用いて作った潜在変数）にした例である。まずはこの例を参考にモデルを作ってみるとよい。

観測変数2つでも因子を作ることができるが、3つ以上あった方が安定した結果になることが多い。ただし、それは元々の観測変数の性質による。あまり似ていない観測変数を3つ用いて因子を作ろうとしても、うまくいかない。まず、各観測変数の内容（質問文の意味や変数間の相関係数など）をよく確認した上で、モデルを作ることが重要。

図9のように、タイトル中に、AGFI=YAGFIなどを書くと、モデル全体の適合度係数が表示される。この場合のp値は、モデルとデータの距離についてのものなので、大きい方がよい。5%以上の値になればよいが、この数字は普通、あまり大きくならない。これが0%でも、それほど気にしなくてもよい。AGFIが0.85以上で、RMSEAが0.10以下くらいであれば、良いモデルと言ってよい。RMSEAは、モデルに無駄に複雑な部分がないかどうかの数字で、小さい方がよい。SEMの主な適合度係数は以下。係数の定義は統計学の文献を読むこと。

### カイ二乗値

モデル全体がデータに当てはまっているかについての検定。モデルと現実の距離を表す。値が一定の値より小さければ、モデルが現実<sub>に</sub>当てはまっている（モデルとデータの距離が小さい）。値が大きいとモデルは不採択となる。

GFI(Goodness of Fit Index)

範囲は0~1。1に近いほど説明力のあるモデルといえる。しかし、多くの場合に0.90以上となってしまい、あまり目安にならない。

AGFI(Adjusted Goodness of Fit Index ; 修正適合度指標)

範囲は0~1となる。自由度調整をした指標。

RMR(Root Mean square Residual; 残差平方平均平方根)

値が0に近いほど、モデルが当てはまっているとする。モデルによって推定された分散と、実際のデータの分散の間の距離を表している。

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

モデルと現実の距離をあらわす。複数のモデルを比較する際には、値が小さい方がよい。

RMSEA(Root Mean Square Error of Approximation)

モデルに無駄に複雑な部分がないかについての値。小さい方がよい。

一般的には0.05以下であればかなりよいモデルであり、0.10以上であれば当てはまりが悪いとする。

## 10. SPSSシンタックスの見本

AMOSを使う前に欠損値があるケース（個人）について、以下のシンタックスでデータ人数を減らすとよい。人数が減ったSPSS形式データファイル(sav)は、名前をつけて保存する。自分のモデルに4変数あるなら、以下のようにシンタックスを4行書けばよい。多くの場合、欠損値は9か99だが、データによって異なるので注意する。

```
/****** SELECT文により欠損値9がある人を削除しデータ人数を減らす *****/  
SELECT IF Q6C <9.  
SELECT IF Q30A <9.  
SELECT IF Q30B <9.  
SELECT IF Q30C <9.
```

```
/****** 逆転した新変数作成 変数の方向をそろえる *****/  
COMPUTE GIIN =5-Q6C.  
COMPUTE axlife =5-Q30A.  
COMPUTE axincom=5-Q30B.  
COMPUTE axjobls=5-Q30C.
```

```
/****** 度数分布表を出力 *****/  
FRE VAR = AGE EDU GIIN axlife axincom axjobls.
```

```
/****** 参考までに相関係数を出力 *****/  
COR /VARIABLES= AGE EDU GIIN axlife axincom axjobls.
```

シンタックスで、逆転した新変数を作り変数の方向をそろえ、その後に分析すること。自分が作ったシンタックス(sps)は名前を付けて保存すること。



11. おまけ 因子分析のSAS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=は省略して良い。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=prinit 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;
```