

表 6.11: 初期値の設定 (EQS)

```

/EQUATIONS
V1 = 30*V999 + 1F1 + E1;
V2 = 25*V999 + .54*F1 + E2;
V3 = 16*V999 + 1.46*F1 + E3;
V4 = 10*V999 + 1F2 + E4;
V5 = 19*V999 + 1.29*F2 + E5;
V6 = 17*V999 + 2.22*F2 + E6;

```

CALIS では、個々のパラメータに初期値を設定することはできません。¹²

¹²CALIS では初期値の推定にかなりの工夫がある。ユーザーが初期値を与える必要はないということであろう。

第7章 潜在曲線モデル

本章では、共分散構造モデルの中でも最近注目を浴びている潜在曲線モデル (latent curve model) について紹介します。潜在曲線モデルは、例えば、子供の成長 (身長伸びや体重増加) の時系列変化といったように、各個体から経時的に反復測定したデータ (縦断的データ) の解析に用いることができます。

適用事例が多いのは1次元の縦断的データです。ここで紹介する解析事例もアルコールの消費回数という1次元データです。潜在曲線モデルに現れる潜在変数は、切片や傾きというような、直線や曲線を記述する量です。¹ このように、前章までに紹介した、いわゆる因子分析の発展形としての共分散構造モデルとは、少し様相が異なります。潜在曲線モデルや第8章で紹介する二段抽出モデルなどは、1990年代以降に発展した「第二世代の共分散構造モデル」と言えるものです。²

しかし、潜在曲線モデルは数学的には平均構造のある共分散構造モデルですので、標準的な共分散構造分析のソフトウェアで実行することができます。AMOS や EQS では、潜在曲線モデルを実行しやすくするオプショ

¹もちろん多変量の潜在曲線モデルもあるし、従来の潜在変数に潜在曲線モデルを適合させることもできる。

²McArdle(1988) や Meredith-Tisak(1990) がその端緒を開いた。潜在曲線モデルを扱った成書として Duncan et. al.(1999) がある。邦書では豊田 (2000, 12章) に解説がある。

ンが備えられています。このモデルは、発達・教育心理学分野を含めたより広範な領域への応用が期待されています。³

7.1 潜在曲線モデルの考え方と分析例

潜在曲線モデルで分析されるのは縦断的データであり、オブザベーション(個体)は「個人」で変数は「時間」であることが多いです。ただし、必ずしも測定値は「成長」をあらわすもの(例えば「身長」や「体重」)である必要はありません。変数が順序付けられている必要はありますが、測定値は「経年」に伴う「アルコール摂取量」「コレステロール値」や、「試行回数」に伴う「反応時間」「反応回数」等であってもかまいません。また、オブザベーションについても個人に限定する必要はなく、例えば「日本」「大阪府」「大阪大学」といったものであってもかまいません。その場合は「国民総生産」「財政赤字額」「学位授与者数」等が測定値になりえます。

成長や変化を分析するとき、個体差をどう記述するかという大問題があります。一口に成長・変化と言っても色々なパターンがあります。最初がんなばるが途中で息切れする人や大器晩成型の人がいるかと思えば、着実に積み重ねていくタイプもいます。また、まったく変化しない個体や途中で低下をきたす個体もあるでしょう。従って、母集団全体で総括してどのような変化をするかということはもちろん大事なのですが、それに加えて、どのように個体差を記述するかということが問題になるわけです。

個体ごとにその推移を追っていけばすべての個体の変化のパターンが分かるのですが、サンプルの大きさが数百以上になると何らかの括りがないと理解不能になります。その一つの解決方法は、個体の属性によって変化のパターンを記述することです。例えば、男女で経時変化のパターンは違うのか、年齢による違いはあるのか、国籍はどうか、環境要因はどのように効くのか、といったものが考えられます。このように、経時変化の個体差を何らかの説明変数で記述できるならば、現象の理解は深まり、有効なアクションがとれるかもしれません。そのような方法論を提供するのが、潜在曲線モデルなのです。

ここでは、アルコールの消費回数という実際の縦断的データを用いて、

³潜在曲線モデルを用いた邦文の応用研究としては清水(1999)などがある。

潜在曲線モデルとその適用例を説明していきます。

7.1.1 アルコールデータ

表 7.1 に、10 歳から 15 歳までの青少年を対象にした、アルコール消費回数に関するデータを示しています。⁴ この調査は、同一回答者に対して、同一内容の調査を、1 年間隔の 3 時点 (T_1, T_2, T_3) で縦断的に行ったもので、標本サイズは $n = 363$ です。各観測時点 (T_1, T_2, T_3) におけるアルコール消費回数の変数 (X_1, X_2, X_3) には、ビール・ワインやウィスキーを飲んだ回数、泥酔した回数などが下位尺度として含まれています。この変数の平均を見ると、1.36, 2.12, 3.18 と、時が進む(経年)につれて線形(直線)に近い形で増加していることが分かります。また、残りの 3 つの変数は回答者の属性を表しています。年齢 (X_4)、性別 (X_5 ; 男性=1, 女性=0 の二値変数) の他に、回答者の飲酒経験に深く関わるであろう環境変数として「親のアルコール依存 (X_6)」を測定しています。これは二値変数で、両親またはいずれかの親がアルコール依存症と診断されている回答者の場合は 1、そうでなければ 0 の値をとります。

ここで、アルコール消費回数と時間経過について、二つの視点からデー

表 7.1: 青少年のアルコール消費回数のデータ (n=363)

	X1	X2	X3	X4	X5	X6
相関行列						
T1 におけるアルコール消費回数	1.00					
T2 におけるアルコール消費回数	.68	1.00				
T3 におけるアルコール消費回数	.50	.68	1.00			
年齢	.31	.31	.23	1.00		
性別 (男性 1, 女性 0)	.01	.06	.10	.14	1.00	
親のアルコール依存	.12	.17	.20	.12	-.15	1.00
平均	1.36	2.12	3.18	12.91	0.52	0.56
標準偏差	2.81	3.98	4.79	1.41	0.50	0.50

⁴出典は Curran(2000)。

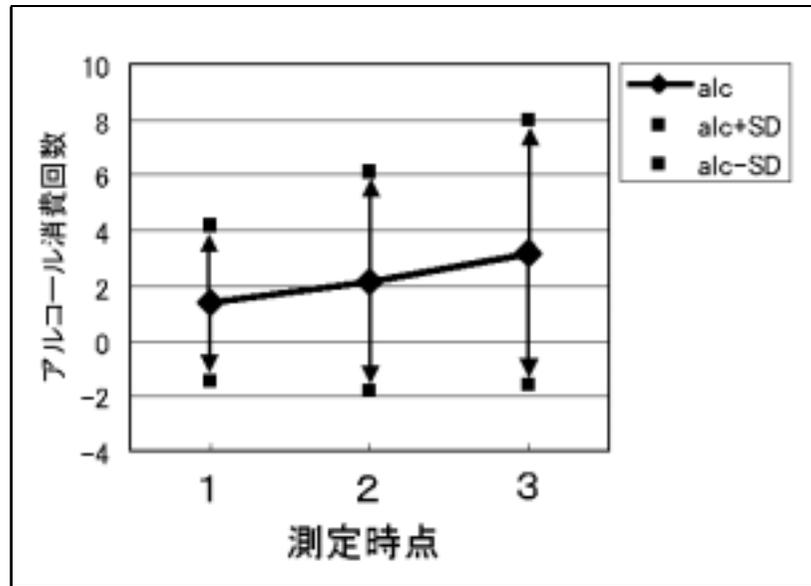


図 7.1: アルコール消費回数の経時変化 (回答者全体)

タを眺めてみましょう。まず図 7.1 は、回答者全体の平均値と標準偏差をプロットしたものです。先ほど述べたとおり、時間経過につれて平均値は直線的に増加する傾向にあり、それと同様に標準偏差も大きくなっていることがわかります。つまり、アルコール消費回数は年を追うごとにばらつきが大きくなっており、個人差が顕著になってきているというわけです。次に、図 7.2 に、回答者 3 名の個人プロフィールを示しています。各観測時点において各回答者のアルコール消費回数をプロットし、近似直線を当てはめています。先に示唆されたとおり、各観測時点でのアルコール消費回数に見られる回答者間のばらつきはかなり顕著であるものの、回答者ごとの近似直線の当てはまりは比較的よさそうです。

分析の最終目的は、回答者の時系列的な「アルコール消費回数」から、その初期値や増加率に、「 X_4 :年齢」「 X_5 :性別」「 X_6 :親のアルコール依存」がどのように関係しているかを調べることです。「アルコール消費回数」の初期値と増加率については、当てはめられた近似直線の切片と傾きで記述

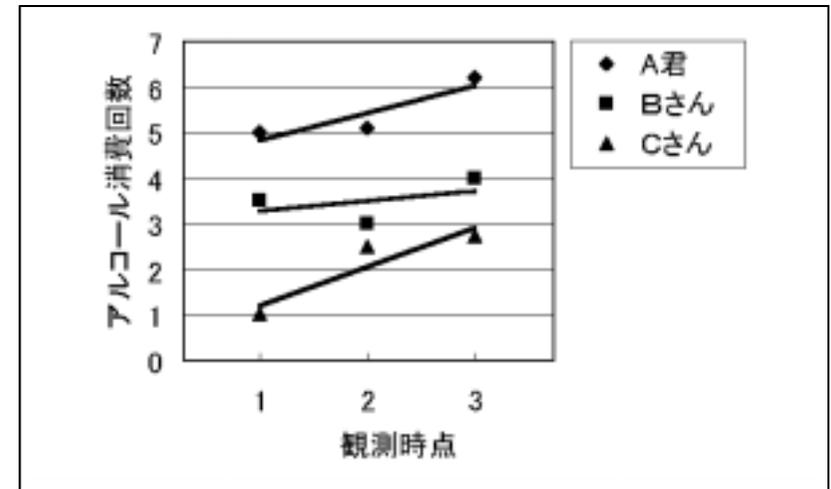


図 7.2: アルコール消費回数の経時変化 (個人プロフィール)

することができそうです。そして、各回答者ごとに計算された切片と傾きが、「 X_4 :年齢」「 X_5 :性別」「 X_6 :親のアルコール依存」とどのような関係にあるか、すなわち、切片または傾きを回帰分析の基準変数とし、(回答者の属性変数である) X_4 , X_5 , X_6 を説明変数とした回帰分析を実行することで、分析の目的が達せられるでしょう。

このように、従来の分析を適用するとすれば、回帰分析を 2 回行うこととなります。このモデルを階層的線形モデル (HLM, Hierarchical Linear Model)、もしくは、MLM (MultiLevel Model) といい、これらのモデルを実行するソフトウェアもあります。⁵

共分散構造分析における潜在曲線モデルは、階層的線形モデルを一つの共分散構造モデルで実現し、一回の“run”で実行します。モデルの当てはまりや各回帰式の当てはまりなどを吟味することができます。

次節では、まず時間にアルコール消費回数を回帰する回帰分析を行うモデルを例にとって、潜在曲線モデルの基本的な考え方を説明します。

⁵例えば、Raudenbush et. al. (2000) による HLM5 がある。Bryk-Raudenbush(1992) も参照のこと。

7.1.2 1次のモデル：変化に直線を当てはめる

ここでは、観測時点 t におけるアルコール消費回数を x_t で表すことにします。先年のアルコールデータの例では、 $t = 1, 2, 3$ となります。⁶ 観測時点に対して x_t をプロットし、直線を当てはめてみましょう。切片と傾きを β_0, β_1 とすると、

$$x_t = \beta_0 + (t-1)\beta_1 + e_t \quad (t = 1, 2, 3) \quad (7.1)$$

となります。⁷ ここでの e_t は、直線からのずれを表す誤差項です。回答者を表す添え字を付加すれば、

$$x_t^{(i)} = \beta_0^{(i)} + (t-1)\beta_1^{(i)} + e_t^{(i)} \quad (t = 1, 2, 3; i = 1, \dots, 363)$$

となります。このように書くと、切片や傾きが回答者ごとに変化するということがよく分かります。しかし、この表現はやや複雑ですから、以後は式 (7.1) で議論を進めていきましょう。

もう一度式 (7.1) を見てみましょう。この式は、 β_0, β_1 を共通因子とする因子分析モデルの形をしています。3 時点でのモデルを具体的に表現すると

$$x_1 = \beta_0 + 0 \times \beta_1 + e_1$$

$$x_2 = \beta_0 + 1 \times \beta_1 + e_2$$

$$x_3 = \beta_0 + 2 \times \beta_1 + e_3$$

となり、また、行列表記にすると、

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \quad (7.2)$$

となります。つまり、因子分析モデルに当てはめれば、 $[\beta_0, \beta_1]'$ が共通因子、その前にある $0, 1, 2$ からなる定数行列が因子負荷量にあたるわけです。

⁶ 観測時点が等間隔でない場合は、 $t = 1, 4, 5$ などとする

⁷ $x_t = \beta_0 + t\beta_1 + e_t$ ($t = 0, 1, 2$) と書く方が直感的には理解しやすいかもしれない。しかし、共分散構造分析では、変数は X_1, X_2, \dots と (0 からではなく) 1 から始めるのが普通なので本文のようにした。

因子分析モデルでは、共通因子は確率変数とするのが一般的です。共分散構造モデルにおける潜在変数も確率変数です。すなわち、ここでの β_0, β_1 も確率変数です。もしかすると、切片や傾きを確率変数と考えることに少々の抵抗を感じる読者があるかもしれません。しかし、こう考えてみるとよいでしょう。図 7.2 に示したように、アルコール消費回数は個人によって異なる傾向にあります。それゆえに、切片・傾きは回答者個人に「貼り付いている」量と考えることができます。この意味で、切片や傾きはこれまでの章で取り扱った個人の知能や学力といった確率変数と同じであると考えることができます。よって、切片・傾きを確率変数と考えるのに抵抗感を持つ必要はないわけです。

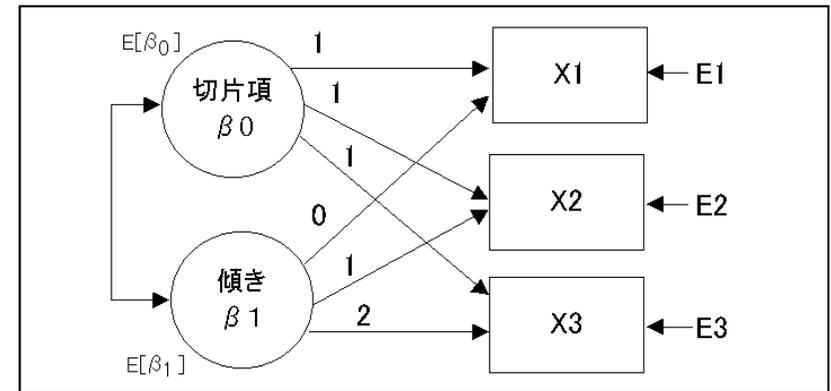


図 7.3: 1次 (直線) のモデル

さて、式 (7.2) のモデルをパス図で表すと図 7.3 のようになります。このパス図で、 β_0 と β_1 の間に相関 (共分散) を設定していることに注目してください。これは、4.2 節で既に述べた共分散構造分析の基本ルール「(誤差変数を除く) 独立変数間には通常、相関 (共分散) を設定する」に則ったものです。しかし、実際的な意味として「観測開始時点 (T_1) でアルコール消費回数が多い回答者は、その後の消費回数の増加率も高い」とか、逆に「観測開始時点 (T_1) でアルコール消費回数が多い回答者は、既にその時点で消費の臨界点に達しているから、以降あまり消費回数は伸びない」といった可能性が考えられるため、切片と傾きには何らかの相関が認めら

れても不思議ではありません。そして、もし、このような知見が得られたならば、それは重要な発見ではないでしょうか。

潜在変数モデルと通常の因子分析モデルとの相違点は、確率変数 β_0, β_1 の平均が0ではないことです。⁸ 平均 $E(\beta_0)$ は、観測開始時点 (T_1) での回答者の平均アルコール消費回数を表しており、平均 $E(\beta_1)$ は、平均伸び率 (1年あたりの消費回数の伸び率) を表しています。パス図にはその意味で $E(\beta_0)$ と $E(\beta_1)$ を入れてあります。図 7.3 で分析を行うためには、平均構造のある共分散構造モデル (6.2 節参照) を用いることになります。ただし、観測変数に切片が含まれていないことに注意しましょう。なお、推定すべきパラメータは、因子の平均と分散・共分散、そして誤差分散です。

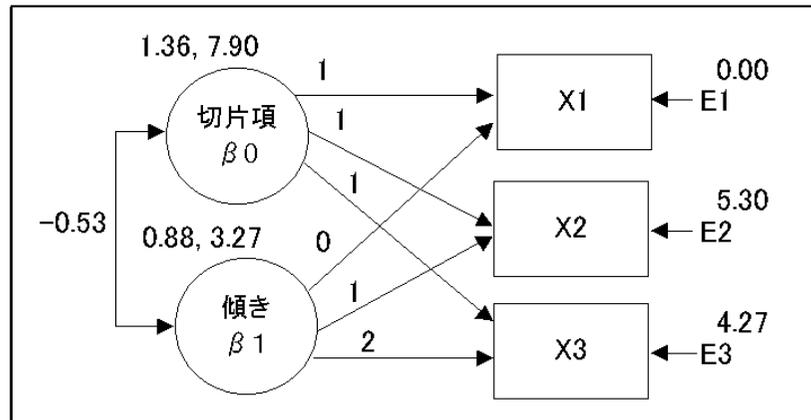


図 7.4: 1 次のモデルの推定結果

1 次のモデルの分析結果は図 7.4 のようになります。因子に付与されている数値は左側が平均、右側は分散の推定値です。モデルの適合度は

$$\text{カイ 2 乗値} = 1.897, \quad \text{自由度} = 1, \quad \text{p 値} = .168 \quad (7.3)$$

となって、十分良い当てはまりであることが分かります。

EQS と CALIS では、平均の効果を表す「定数変数 (CONSTANT)」を

⁸もう一つ付け加えるならば、因子負荷量がすべて既知であり推定しないこと。

導入するので、パス図での表示は図 7.4 と少し異なります。EQS での分析結果は図 7.5 のようになります。(因子である) 切片と傾きの誤差が 7.90、

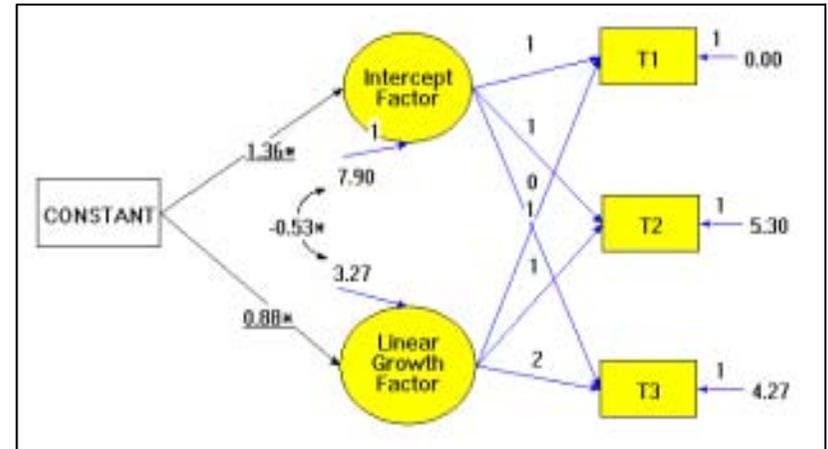


図 7.5: 1 次のモデルの推定結果

3.27 と推定されており、これらはそれぞれ切片「Intercept Factor(β_0)」の分散と傾き「Linear Growth Factor(β_1)」の分散を表します。定数変数「CONSTANT」からのパス係数は切片と傾きの平均を表し、それぞれ 1.36、0.88 と推定され、有意になっています。

切片の平均は調査時点の全回答者のアルコール消費回数の平均値ですから、表 7.1 にある $T_1(X_1)$ の平均値 1.36 と一致します。また、回帰式から、 T_2 と T_3 における消費回数は、 $1.36 + 0.88 \times 1 = 2.24$ 、 $1.36 + 0.88 \times 2 = 3.12$ と予測することができ、これらの値は実際のデータ 2.12、3.18 と近い値であることが分かります。このように、潜在変数 (共通因子) である切片と傾きの平均が、全データを一括りにしたものの回帰分析結果を表しています。

一方、個人差 – 個人による直線の違い – はこれらの因子の因子得点として表されます。生データがありませんので各回答者の因子得点は出力できませんが、そのバラツキの程度は因子の分散で評価できます。切片の分散は 7.90 と推定されており、これに誤差分散 $\text{Var}(e_1)$ の推定値を加えたものが X_1 の分散、すなわち、表 7.1 にある $T_1(X_1)$ の標準偏差 2.81 の 2

乗に一致します。⁹ 傾きのバラツキは分散で 3.27 と評価されています。これは、各回答者ごとの傾きが、 $0.88 \pm 2 \times \sqrt{3.27} = 0.88 \pm 3.62$ ぐらいばらついていることを教えてくれます。両因子の共分散は -0.53 と推定されており、相関係数に直すと

$$\frac{-0.53}{\sqrt{7.90 \times 3.27}} = -0.10$$

となりあまり大きな値でなく有意でもありません。もちろん、この値は標準解を表示させることで見ることができます。 T_1 時点で消費回数が多い回答者は増加率が低い、つまりこれらの間に負の相関があるということは、統計的には確認できなかったということになります。

観測変数に付随する誤差 $e_1 \sim e_3$ の分散は、各時点における回帰直線の当てはまりの悪さを表しています。 e_1 の分散が 0 と推定されていることに注意してください。これを不適解といい大きな問題をはらんでいる可能性があります。採用したモデルが不適切であるために不適解が生じているかもしれないからです。実は、0 と推定されたこの値は「分散は非負」という制約を課して推定したもので、もしこの制約を課さなければマイナスのとても大きな値になる可能性もあります。¹⁰ そうした場合はモデル規定に大きな誤りがあることとなります。このことを確認するため、その制約をはずして再分析します。そのためには、EQS モデルファイルに、不等式制約

```
/1 INEQUALITY
-99999<(E1, E1);
```

を記述します。¹¹ すると e_1 の分散の推定値、標準誤差、z-値はそれぞれ、-.586, .746, -.785 と出力されます。 e_1 の分散は負に異常に大きな値ではなく z-値も有意になっていません。この事実と適合度が良いということを考え合わせると、 e_1 の分散は、標本変動によって不運にも偶然負の値に推

⁹ここでは、 $\widehat{\text{Var}}(e_1) = 0$ と推定されているので、 $7.90 = 2.81^2$ となっている。

¹⁰分散の推定値に非負制約をおくことがデフォルトであるかどうかはソフトウェアによる。AMOS, CALIS は非負制約をおかない。不適解についてのこのような吟味は、分析の初期段階で行う。

¹¹モデルファイル (*.eqx) で、プルダウンメニューから Inequality を選択する。(E1, E1) を選び Lower Range に -99999 を入力、 をクリックする。

定されてしまったという解釈が可能です。不適解の原因が標本変動であると判断できるときは、不適解を重大な問題と考えなくてもよいのです。なお、論文やレポートでは、負の値に推定された分散を報告してはいけません。報告する最終解は、非負の制約をおいた解、すなわち、図 7.4 です。

7.1.3 説明変数のあるモデル

前節で解説した潜在曲線モデルは、経年によって青年のアルコール消費回数が、2年間3時点で直線的に変化すると考えてよいかどうかを検討する単純なものでした。個人による直線の違いは、因子で表された「切片と傾き」の違いによって表すことができました。本節ではもう一步分析を進めて、回答者ごとに異なる切片や傾きが、どのような要因によって左右されているのかを調べてみましょう。男女差があるのでしょうか、年齢による影響が強いのでしょうか、それとも環境(両親のアルコール依存)が深く関わっているのでしょうか。これらの変数 ($X_4 \sim X_6$) をモデルの中に取り込んで説明変数とし、切片と傾きを基準変数としたパス解析を実行すれば、要因間の関係について有益な情報が得られる可能性があります。

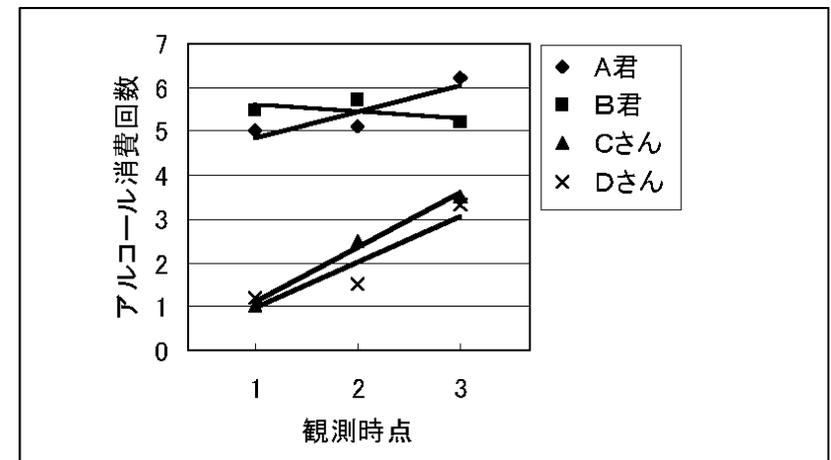


図 7.6: アルコール消費回数と時間経過 (男女別プロフィール)

アルコール消費回数と時間経過に関する個人プロフィールが、(仮に) 図 7.6 のようになったとしましょう。図 7.2 とはやや異なり、こちらでは男性 2 名と女性 2 名のデータです。このグラフを見ると、女性の場合は時間経過にしたがってアルコール消費回数が線形に(かなり急な傾きで)増加しているのに対して、男性は初期値 (T_1 時点) での消費回数が高いが増加傾向は小さいことが見て取れます。¹² そこで、先ほどの分析モデルに性別の違いを取り込んだモデルを考えてみましょう。パス図で表すと図 7.7 のようになります。

このモデルは、性別という属性によって、 β_0 や β_1 が系統的に影響を受けることを仮定したモデルです。解析は、先ほどの 1 次のモデルと同様に、平均構造モデルを用いて行います。図中の α_0, α_1 は切片¹³を、 $E[\text{性別}]$ は性別の平均を示しています。性別変数の影響を方程式で表すと

$$\begin{cases} \beta_0 = \alpha_0 + \gamma_0 X_4 + d_1 \\ \beta_1 = \alpha_1 + \gamma_1 X_4 + d_2 \end{cases}$$

となります。これを用いると、性別変数 X_4 で男性を 1, 女性を 0 とコーディングしてあれば、切片と傾きの期待値は、男女別に表 7.2 のように計

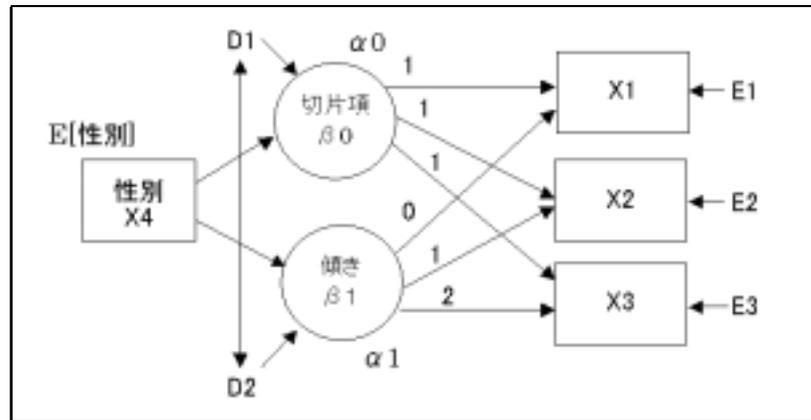


図 7.7: 説明変数のあるモデル

¹²例であり実際の分析結果とは異なる。

¹³従属変数である因子 β_0 と β_1 の切片である。切片 β_0 と間違わないようにすること。

算できます：

表 7.2: 切片と傾きの期待値

	$X_4 = 0$ (女性)	$X_4 = 1$ (男性)
$E(\beta_0)$	α_0	$\alpha_0 + \gamma_0$
$E(\beta_1)$	α_1	$\alpha_1 + \gamma_1$

このように、切片 α_0, α_1 は女性の平均を、 γ_0, γ_1 は男女差を表すパラメータであることが分かります。 $\gamma_0 \neq 0$ や $\gamma_1 \neq 0$ が判断できるときは男女差があることとなります。もし

$$\alpha_0 = 1, \alpha_1 = 1; \gamma_0 = 4, \gamma_1 = -1$$

であったならば、図 7.6 のような特性になります。

また、誤差変数 d_1 と d_2 に相関が設定されていることに注意してください。これは、 β_0 と β_1 の関係が、性別だけでは完全に説明できない可能性を考慮したものです。

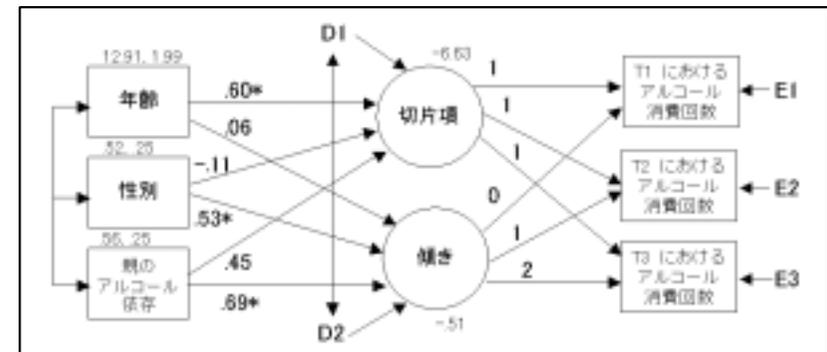


図 7.8: 3 つの説明変数を導入した潜在曲線モデル (推定結果)

図 7.6 をもう一度見てください。これらの直線は平行ではなく、特に、性別によって大きく傾きが変わっています。このことは、アルコール消費

回数に影響を及ぼす要因として時間と性別を考えたとき、交互作用「時間×性別」があることを示しています。この交互作用は、性別から傾きへの影響 γ_1 が0でないことから生み出されたものです。¹⁴ 従って、 $\gamma_1 = 0$ の検定は、交互作用の検定に他なりません。

説明変数を一つではなく、複数にした場合も手続きは同様です。次に、「性別」「年齢」「親のアルコール依存」という3つの説明変数を導入した潜在曲線モデルの分析結果を見てみましょう(図7.8)。適合度は表7.3に示すとおりで、良い適合を示していると言えます。

表 7.3: 分析モデルの適合度

カイ2乗値	自由度	p値	GFI	CFI	RMSEA
4.982	4	0.289	0.997	0.999	0.026

パス係数の推定結果によると、年齢や性別という属性、あるいは親のアルコール依存という環境によって、 T_1 時点でのアルコールの消費回数(切片)と増加の割合(傾き)は変化するようです。具体的には、以下のような関係にあることが考察できます。

1. T_1 時点で年齢が一つ上だと消費回数は0.60多い。年齢から消費回数の増加量(傾き)への影響は認められない。
2. 性別は T_1 時点での消費回数には有意な影響をもたないが、経年による増加量には影響し、傾きで0.53だけ男性が高い。
3. 親がアルコール依存であるかどうかは、経年による増加量に影響(0.69)をもつ。 T_1 時点での消費回数へのパス係数は0.45と推定されているが、統計的有意性は認められなかった。

この節を終わるにあたって、2種類の誤差 e と d の役割を確認しておきましょう。誤差 e は、既に述べたように、各回答者の散布図へ入れた直線とデータのずれを表しています。 e の分散が大きいと直線の当てはまりが悪いこととなります。一方、誤差 d は、切片や傾きのばらつきで説明変数によって説明できない程度を表しています。すなわち、この例に則し

¹⁴ $\gamma_1 = 0$ ならば平均的な傾き $E(\beta_1)$ は性別に依存せず、男女ごとの直線は平行になる。

たとえば、年齢、性別、親のアルコール依存を決めても変動する、切片や傾きの程度を表しています。

まとめると、潜在曲線モデルは、個体差を考慮して個体ごとに異なった直線を入れたとき、その直線はどの程度当てはまっているか ($\text{Var}(e_i)$)、個体ごとに異なる直線は説明変数でどの程度説明できるか ($\text{Var}(d_i)$) を区別して、「一つのモデル」で評価してくれるのです。

7.1.4 2次のモデル

これまで紹介してきた潜在曲線モデルは、1次式(直線)の当てはめを行うものでしたが、2次曲線の当てはめも行うことができます。¹⁵ モデル式は、

$$x_t = \beta_0 + (t-1)\beta_1 + (t-1)^2\beta_2 + e_t \quad (t = 1, 2, 3)$$

となります。この式も、 $\beta_0, \beta_1, \beta_2$ を共通因子とする因子分析モデルの形をしています。3時点でのモデルを具体的に表現すると、

$$x_1 = \beta_0 + 0 \times \beta_1 + 0 \times \beta_2 + e_1$$

$$x_2 = \beta_0 + 1 \times \beta_1 + 1 \times \beta_2 + e_2$$

$$x_3 = \beta_0 + 2 \times \beta_1 + 4 \times \beta_2 + e_3$$

となり、行列表記にすると、

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

であり、パス図で表現すると図7.9のようになります。

7.2 AMOS, EQS, CALIS による実行例

以下の節では、潜在曲線モデルの解析方法について、1次(直線)のモデルを例にとり、その手順をソフトウェアごとに解説していきます。

¹⁵非線形の潜在曲線を扱うことも可能だが、本書では紙幅の関係上ふれることができない。豊田(2000)などを参考にされたい。

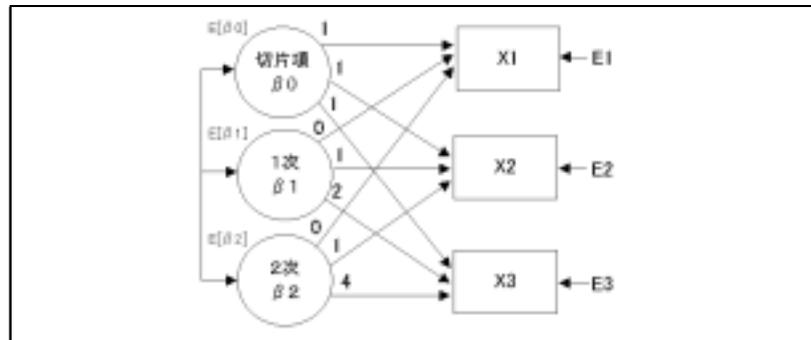


図 7.9: 2次曲線モデル

7.2.1 AMOS

まずデータセットの用意をします。ここでは MS-Excel で図 7.10 のように作成しました。

	A	B	C	D	E	F	G	H
1	rowtype_	varname_	T1	T2	T3	年齢	性別(M=1)	観アルコール
2	n		363	363	363	363	363	363
3	corr	T1	1	0.68	0.5	0.31	0.01	0.12
4	corr	T2	0.68	1	0.68	0.31	0.06	0.17
5	corr	T3	0.5	0.68	1	0.23	0.1	0.2
6	corr	年齢	0.31	0.31	0.23	1	0.14	0.12
7	corr	性別(M=1)	0.01	0.06	0.1	0.14	1	-0.15
8	corr	観アルコール	0.12	0.17	0.2	0.12	-0.15	1
9	stddev		2.81	3.98	4.79	1.41	0.5	0.5
10	mean		1.36	2.12	3.18	12.91	0.52	0.56

図 7.10: データ

AMOS では、潜在曲線モデルのパス図を簡単に描くためのマクロが用意されています。ツールバーから **ツール** | **マクロ** | **Growth Curve Model** を選択し、測定時点の数 (Number of time points, ここでは 3) を入力すると図 7.11 が得られます。ただし、1,1,1, 0,1,2 と固定するパス係数はオブジェクトのプロパティでユーザーが入力しなければなりません。データ

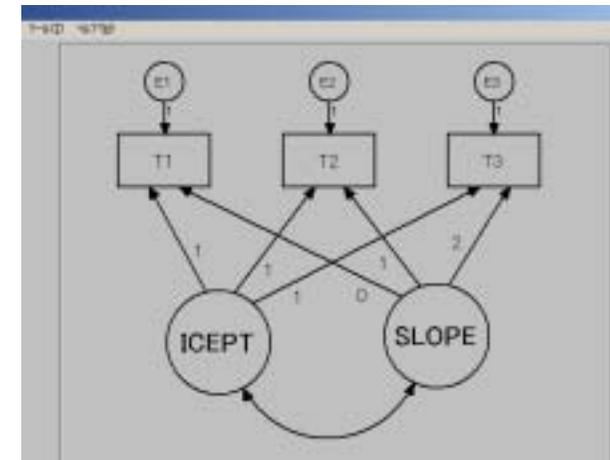


図 7.11: 潜在曲線モデルのマクロ

ファイルを指定して、観測変数名を X_1, X_2, X_3 から T_1, T_2, T_3 へ変更します。モデルに含まれる変数を表示させて変数名を観測変数の長方形へドラッグすると間違いがありません。平均構造の分析を行うために、分析のプロパティから推定のタブを選択して **平均値と切片を推定** にチェックを入れます。観測変数の切片を 0 とおくために、観測変数のプロパティを開いて切片に 0 を入れます。切片 (ICEPT) と傾き (SLOPE) の因子平均を設定するには、因子のプロパティを開いて平均にある 0 を消去します。以上でモデル規定は終了です。AMOS を走らせると図 7.12 の出力が得られます。推定値が微妙に異なるときは、分析のプロパティにある分散タイプのタブで、二つの共分散を共に「不偏推定値共分散」としてください。

$\widehat{\text{Var}}(e_1) = -.59$ となっていることから分かるように、AMOS のデフォルトでは分散の推定値が負の値になることがあります。表出力の表示を見ると表 7.4 のようになっており、負の値と推定された $\text{Var}(e_1)$ ですが有意ではありません。そこで、 $\text{Var}(e_1) \geq 0$ として推定することになります。ところが、AMOS には不等式制約をおく機能がないので、 $\text{Var}(e_1) = 0$ とおいて再推定します。 e_1 のオブジェクトのプロパティを開き、パラメー

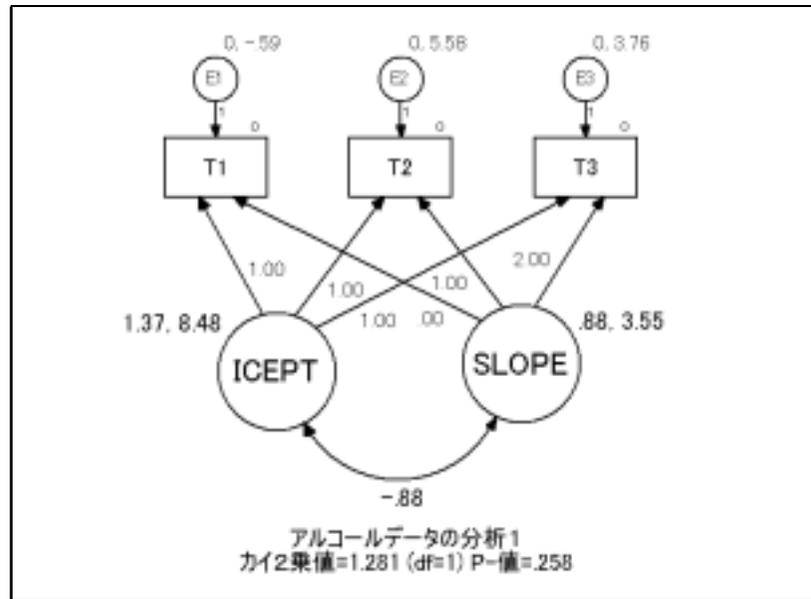


図 7.12: 1 次のモデルの推定結果：説明変数なし

タのタブで分散に 0 を設定します。¹⁶ 変更したモデルで推定すると、(7.3) と図 7.4 と同等の結果が得られます。しかし、モデルの自由度と p 値が異なり、AMOS の方は、df=2, p 値=.387 となります。これは、 $Var(e_1)$ を自由パラメータとみなすかどうかによる違いを反映したものです。不適解の場合、モデルの自由度をどのように定義すべきかは、現在のところ未解決の難しい問題です。個人的には、自由度を大きくとるとモデルを受容

表 7.4: 不適解の吟味

	推定値	標準誤差	検定統計量	確率
e_1	-0.586	0.746	-0.785	0.432

¹⁶推定値が表示されているモードでは変更はできない。入力モードで設定されたい。

しやすくなり第二種の過誤が大きくなるので、小さい自由度 (今の場合は df=1) として判断すべきではないかと考えています。

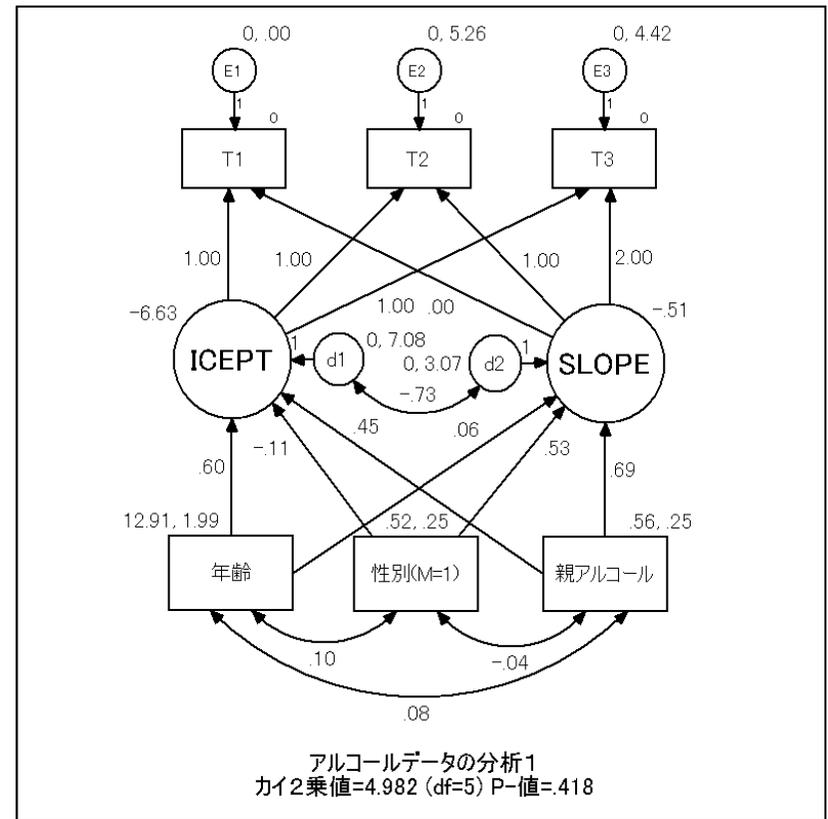


図 7.13: 1 次のモデルの推定結果：説明変数あり

続いて、説明変数のあるモデルでの分析に移りましょう。観測変数「年齢」「性別 (M=1)」「親アルコール」を描き相関を設定します。因子「ICEPT」と「SLOPE」間の相関を外してから、これらの観測変数の各々から二つの因子へパスを引きます。誤差変数 (攪乱変数) d_1 と d_2 を描き相関を設定します。¹⁷ 図 7.13 の推定値が得られます。説明変数のない分析結果と同

¹⁷多くの独立変数間に相関を設定するには、ツールバーから ツール マクロ

様、自由度の違いに注意しておく必要があります (df=4 とする)。

7.2.2 EQS

EQS では、潜在曲線モデルのパス図を簡単に描くことのできるメニューがパスモデルヘルパーに用意されています。このヘルパーメニューを使えば、基準変数となる変数と観測時点数を指定することによって、説明変数を含まない1次(直線)のモデルと2次曲線モデルを描くことができます。

まず、EQS のエディタか外部エディタで表 7.1 のような相関行列のデータを作成し、alcohol.ess などの名前で保存します。

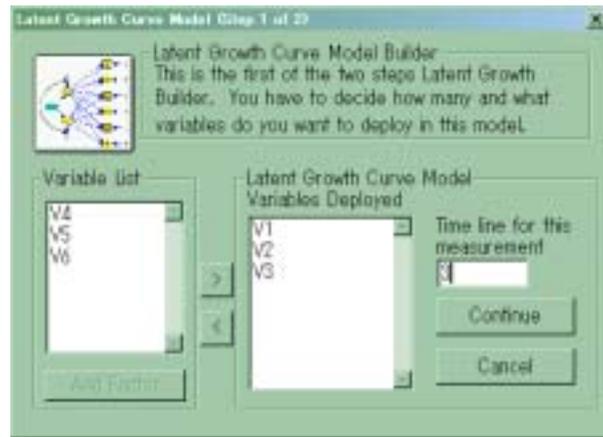


図 7.14: Latent Growth Curve Model (Step 1 of 2)

次に、パスモデルヘルパーを立ち上げ、New Model Helper ダイアログで、上から3番目にある Latent Growth Curve Model を選択します。すると、Latent Growth Curve Model (Step 1 of 2) が開きます(図 7.14)。ここでは、時系列で測定された変数とその回数を設定します。Variable List から、各観測時点のアルコール消費回数を示す変数 $V_1 \sim V_3$ を選択し、 をクリックして左の Variable Deployed に移動させます。そして、Time

Line for this measurement ボックスに、観測回数 3 を入力して、 ボタンをクリックします。すると、画面は Latent Growth Curve Model (Step 2 of 2) に切り替わります(図 7.15)。

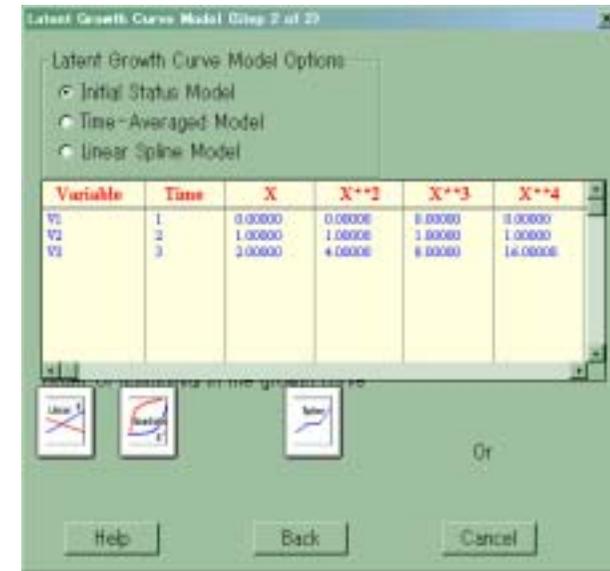


図 7.15: Latent Growth Curve Model (Step 2 of 2)

ここでは分析モデルに関するオプションを設定します。上のボックスでは潜在曲線モデルのオプションを指定します。ここでは Initial Status Model を選択してください。次に下に3つ並んでいるボックスから、どのようなモデルに当てはめるかを選択します。Linear が直線、Quadratic が2次曲線、そして Spline は非線形の潜在曲線の当てはめを行うオプションです。図 7.3 のような1次のモデルの場合は、直線の当てはめを行うので、Linear と書かれたボタンをクリックします。また、図 7.9 のような2次のモデルの場合は、2次曲線の当てはめを行うので Quadratic と書かれたボタンをクリックします。いずれにせよ、この2ステップで作図は完了し、自動的に gwl i near. eds という名前で図 7.16 のようなパス図が作

成されます.¹⁸

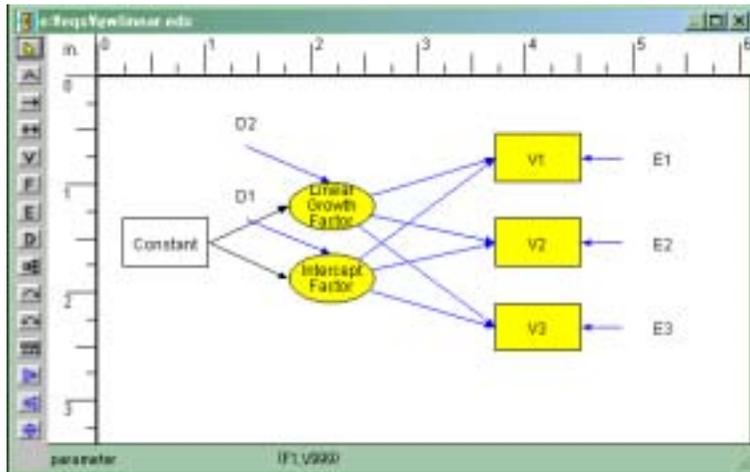


図 7.16: パスモデルヘルパーを用いて描いたパス図

パス図は全体がグループ化されていますので、各要素の位置を変更したり、パスの追加・削除などを行いたい場合は、**Layout** - **Break Group** でグループを解除してから行ってください。

ここでは、図 7.8 にあるような、アルコール消費回数に影響を及ぼすことが予想される 3 つの説明変数を導入したモデルを検査することにしましょう。そのためには、ヘルパーによって描かれたパス図に新たに変数を追加する必要があります。パス図ウインドウの左端に並んでいるパス図作成メニューから **V** を選択し、ウインドウ上に V_4, V_5, V_6 を表示したら、後は図 7.8 のモデルにしたがってパスを引いていきます。図 7.17¹⁹ のようなパス図が完成したら、やはり **File** - **Save As** で保存しておきましょう。

パス図からモデルファイルを作成するためには、EQS のメニューから **Build.EQS** - **Title/Specifications** を選択します。EQS Model Specification ダイアログでは、タイトルを付ける以外は OK ボタンをクリックするだ

¹⁸ **View** - **Labels** をチェックしているため、変数名が表示されている。

¹⁹ **View** - **Labels** をチェックしていないため、変数名が表示されていない。

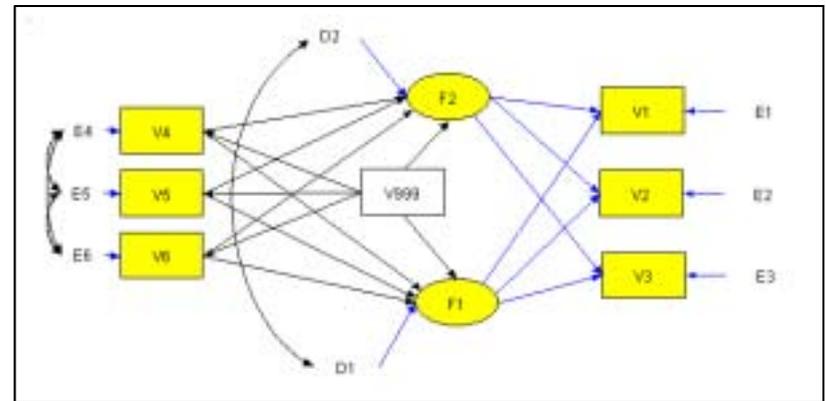


図 7.17: 説明変数のある 1 次直線モデルのパス図

けです。Type of Analysis メニューの Structural Mean Analysis には、既にチェックマークが入っているはずですが、`gwl i near. eqx` という名前の EQS モデルファイルが自動的に作成されたら、**Build.EQS** - **Run EQS** を選択して、`alcohol. eqx` などのファイル名でモデルファイルを保存すると、別の窓が開いて推定値を計算します。計算が終わった後、出力ファイル `gwl i near. out` が開きます。**Window** - `gwl i near. eds` を選択してパス図ウインドウに移り、推定値を表示させることができます。

表 7.5 に説明変数のあるモデルのモデルファイル例を紹介しています。²⁰

7.2.3 CALIS

CALIS では多母集団の同時分析はできないのですが平均構造の分析は行うことができますから、潜在曲線モデルが扱えます。

1 次のモデル (図 7.3) で分析する SAS プログラムは表 7.6 のようになります。データステップの解説は省略します。平均構造 $\mu(\theta)$ があるモデルで分析するとき、CALIS は

$$E[XX'] = \mu(\theta)\mu(\theta)' + \Sigma(\theta)$$

²⁰ EQS がパス図から自動作成するモデルファイルを、見やすくするため若干変更してある。

表 7.5: 1 次モデルの推定結果：説明変数があるモデル

```

/TITLE
  Latent Curve Model with Independent Variables
  Created by EQS 6 for Windows
/SPECIFICATIONS
  DATA='alcohol.ess';
  VARIABLES=6; CASES=363; METHODS=ML;
  MATRIX=CORRELATION; ANALYSIS=MOMENT;
/LABELS
  V1=T1; V2=T2; V3=T3; V4=AGE; V5=SEX; V6=ALC;
/EQUATIONS
  V1 = 1F1 + 0F2 + 1E1;
  V2 = 1F1 + 1F2 + 1E2;
  V3 = 1F1 + 2F2 + 1E3;
  V4 = *V999 + 1E4;
  V5 = *V999 + 1E5;
  V6 = *V999 + 1E6;
  F1 = *V4 + *V5 + *V6 + *V999 + 1D1;
  F2 = *V4 + *V5 + *V6 + *V999 + 1D2;
/VARIANCES
  V999 = 1.00; E1 TO E6 = *; D1 TO D2 = *;
/COVARIANCES
  E4 TO E6 = *; D2, D1 = *;
/PRINT
  FIT=ALL;
  TABLE=EQUATION;
/OUTPUT
  PARAMETER ESTIMATES; STANDARD ERRORS;
  RSQUARE; LISTING;
/END

```

の関係式を利用します。この方法は McDonald(1980) によって開発されたもので、平均構造を意識しないソフトウェアでも、この方法を用いれば平均構造モデルで分析できるというスグレモノです。普通の分析は、平均の周りの 2 次モーメントである分散共分散行列を利用しますが、McDonald の方法は、その代わりに原点まわりの 2 次モーメントを使うのです。これを平均に関して調整していない分散共分散行列 (covariance matrix for uncorrected for the mean) といいます。このことを宣言するのが、UCOV

表 7.6: 1 次モデルの SAS プログラム

```

DATA alcohol (TYPE=corr);
  INPUT _TYPE_ $ _NAME_ $ t1 t2 t3 age sex alc;
CARDS;
N      .      363   363   363   363   363   363
CORR t1   1.00   .      .      .      .      .
CORR t2   .68   1.00   .      .      .      .
CORR t3   .50   .68   1.00   .      .      .
CORR age  .31   .31   .23   1.00   .      .
CORR sex  .01   .06   .10   .14   1.00   .
CORR alc  .12   .17   .20   .12   -.15   1.00
MEAN    .   1.36  2.12  3.18 12.91 0.52 0.56
STD     .   2.81  3.98  4.79 1.41 0.50 0.50
;

PROC CALIS UCOV AUG DATA=alcohol;
  LINEQS
    t1 = 1 f_icept + 0 f_slope + e1,
    t2 = 1 f_icept + 1 f_slope + e2,
    t3 = 1 f_icept + 2 f_slope + e3,
    f_icept = alpha_1 INTERCEP + d1,
    f_slope = alpha_2 INTERCEP + d2;
  STD
    e1-e3 = del 1-del 3,
    d1-d2 = psi 1-psi 2;
  COV
    d1-d2 = psi 21;
  RUN;

```

というオプションです。また、平均構造を表すには 1 という固定変数を使います。LINEQS にある INTERCEP のことです。この変数を認識させるために AUG というオプションを付けることになります。²¹

ここでは、切片と傾きを表す潜在変数を f_{icept} と f_{slope} で表しています。d1 と d2 は、 f_{icept} と f_{slope} からそれらの平均効果を除いたもの、すなわち中心化した量を表しており、実質的に f_{icept} と

²¹Augmented covariance matrix の意味。観測変数ベクトルに固定変数 1 を付け加えたものの 2 次モーメントを考えると平均と 2 次モーメントの両者を含んだ情報を得ることができ、McDonald の方法はこの行列に対するモデリングを行っている。

f_slope と同等です。従って、d1 と d2 に分散・共分散を設定することは、f_cept と f_slope に分散・共分散を設定することになっています。

表 7.6 でのプログラムを実行すると、表 7.7 のようなウォーニングが出され、誤差分散 e_1 の分散が負の値に推定されていることが問題とされます。この不適解の t Value が有意でないことを確認 (絶対値 < 1.96) した

表 7.7: ウォーニングと $\widehat{\text{Var}}(e_1)$

WARNING: The central parameter matrix _PHI_ has probably 1 negative eigenvalue(s).

Vari ances of Exogenous Vari ables

Variable	Parameter	Estimate	Standard Error	t Value
E1	DEL1	-0.586345	0.746109	-0.786

後、 $\text{Var}(e_1) \geq 0$ の制約をおいて再分析します。そのために、RUN; の手前の行に、以下のステートメントを挿入します：

```
BOUNDS
del1 >=0;
```

すると、表 7.8 の最終の出力が得られます。

CALIS では、不等式制約をおいて推定し推定値がその境界にきたとき、そのパラメータを自由なものとして扱いません。例えば、 $\text{Var}(e_1) \geq 0$ の制約の下で $\widehat{\text{Var}}(e_1) = 0$ となると、 $\text{Var}(e_1)$ は自由パラメータではなく、その値が $\text{Var}(e_1) = 0$ と既知であったものとみなされます。その結果、自由度が 1 増加します。²²

説明変数を入れて分析するための SAS プログラム (PROC ステップのみ) を表 7.9 に示しています。

²²237 ページで述べたように、このような場合でも $\text{Var}(e_1)$ を自由パラメータとして扱った方がよいというのが著者の考えである。

表 7.8: 1 次のモデル : CALIS による推定値

Manifest Variable Equations

T1 = 1.0000 F_ICEPT + 1.0000 E1
T2 = 1.0000 F_ICEPT + 1.0000 F_SLOPE + 1.0000 E2
T3 = 1.0000 F_ICEPT + 2.0000 F_SLOPE + 1.0000 E3

Latent Variable Equations

F_ICEPT = 1.3600*INTERCEP + 1.0000 D1
Std Err 0.1475 ALPHA_1
t Value 9.2212

F_SLOPE = 0.8849*INTERCEP + 1.0000 D2
Std Err 0.1071 ALPHA_2
t Value 8.2655

Vari ances of Exogenous Vari ables

Variable	Parameter	Estimate	Standard Error	t Value
INTERCEP		1.002762	0	0.000
E1	DEL1	0	0	.
E2	DEL2	5.302662	0.477406	11.107
E3	DEL3	4.266742	1.123232	3.799
D1	PSI 1	7.896100	0.586912	13.454
D2	PSI 2	3.272405	0.363435	9.004

7.3 潜在曲線モデルのまとめ

成長・変化を分析する際、読者の皆さんがまず考えつくのは次の二つの方法でしょう。すなわち、母集団全体の変化を一つの曲線²³で記述する方法と各個体の変化に一つの曲線 - 曲線は個体の数だけある - を入れて観察する方法です。もちろん両方ともやってみる必要がありますが、前者は、個体差が大きい分野では大雑把すぎて、ぼんやりとした全体像を眺めるだけになってしまいます。また、後者は、個体数が多いと「木を見て森を見ず」という感じになり、そのままでは有益な情報は得られません。こ

²³ここでいう曲線とは直線を含んだ概念である。

表 7.9: 1 次モデルの推定結果：説明変数があるモデル

```

PROC CALIS UCOV AUG DATA=al cohort;
LINEQS
  t1 = 1 f_i cept + 0 f_s lope + e1,
  t2 = 1 f_i cept + 1 f_s lope + e2,
  t3 = 1 f_i cept + 2 f_s lope + e3,
  f_i cept = alpha_1 INTERCEP + gamma_11 age + gamma_12 sex
            + gamma_13 ALC + d1,
  f_s lope = alpha_2 INTERCEP + gamma_21 age + gamma_22 sex
            + gamma_23 alc + d2,

  age = m4 INTERCEP + e4,
  sex = m5 INTERCEP + e5,
  alc = m6 INTERCEP + e6;
STD
  e1-e6 = del 1-del 6,
  d1-d2 = psi 1-psi 2;
COV
  d1-d2 = psi 21,
  e4-e6 = del 54
            del 64 del 65;
BOUNDS
  del 1 >=0;
RUN;

```

の「帯に短し たすきに長し」という状況を打破するのが潜在曲線モデルだと言えるでしょう。すなわち、成長・変化の個体差を、説明変数の導入で有効に分類しようとするのです。説明変数が名義尺度ならば、完全に分類に対応します。その分類に価値があるのか、分類ごとに共通な曲線はどのようなものか、より具体的には、男女によって曲線が異なるのか、親がアルコール依存であるかどうかで飲酒回数は影響を受けるのかなど、これらの疑問に明快に答えてくれます。また、説明変数が連続量の場合は、通常の回帰分析と同様に、説明変数の1単位の増加が、特性の成長や変化にどのように影響するかを教えてください。

潜在曲線モデルは、観測変数に当てはめる直線や曲線の係数を潜在変数に見立てるところから、このようによばれています。「潜在成長モデル (latent growth model)」、「潜在成長曲線モデル (latent growth curve model)」

「成長曲線モデル (growth curve model)」といった用語が用いられることもあります。

通常の共分散構造モデルは、パラメータの値を推定することによって母集団の性質を明らかにすることを目的としています。ここでいうパラメータとは、個々の個体 (測定対象) を直接記述するものではなく、母集団の特徴を記述するものでした。しかし、われわれは、因子分析における因子得点 (e.g., 個人の能力) のように、個体差を記述する道具も持ち合わせています。潜在曲線モデルは、個体差を確率変数とみなすことで、母集団の特徴を記述するパラメータのみを推定するが、個体差も許すというフレームワークを、成長・変化の記述に応用したものであると言えるでしょう。因子分析での個体差が因子得点で表されたのと同様に、潜在曲線モデルでも、切片や傾きは潜在変数ですが、それらの得点で個体差を検討することができます。しかし、因子分析や潜在曲線モデルの一つの特徴は、個体差を記述する因子や切片、傾きを具体的に推定することなく、必要なパラメータ (因子負荷量, 説明変数からのパス係数) が見積もれる点にあります。

共分散構造分析としての潜在曲線モデルには一つ欠点があります。それは、測定時点が各個体ごとにそろっていなければならないという制約です。測定時点が個体ごとに異なっても、各個体の推移に直線や曲線を入れて、それらを記述する量 (切片・傾き) を計算して説明変数へ回帰するという方法は可能です。HLM などのソフトウェアではこのようなデータの分析が可能なのですが、現在のところ、残念ながら共分散構造分析のフレームワークでは分析できないのです。²⁴

²⁴フリーの共分散構造分析ソフトウェアに Neale の Mx がある。Mx は各個体に対して平均と共分散構造を指定できるので、このようなデータでも分析できると思われる (著者未確認)。興味のある読者は以下の URL を訪問されたい: <http://views.vcu.edu/mx/>