

Mplusによる構造方程式モデリング 入門

資料ダウンロード: <http://bit.ly/1gjldfV>

清水裕士
広島大学

simizu706@hiroshima-u.ac.jp

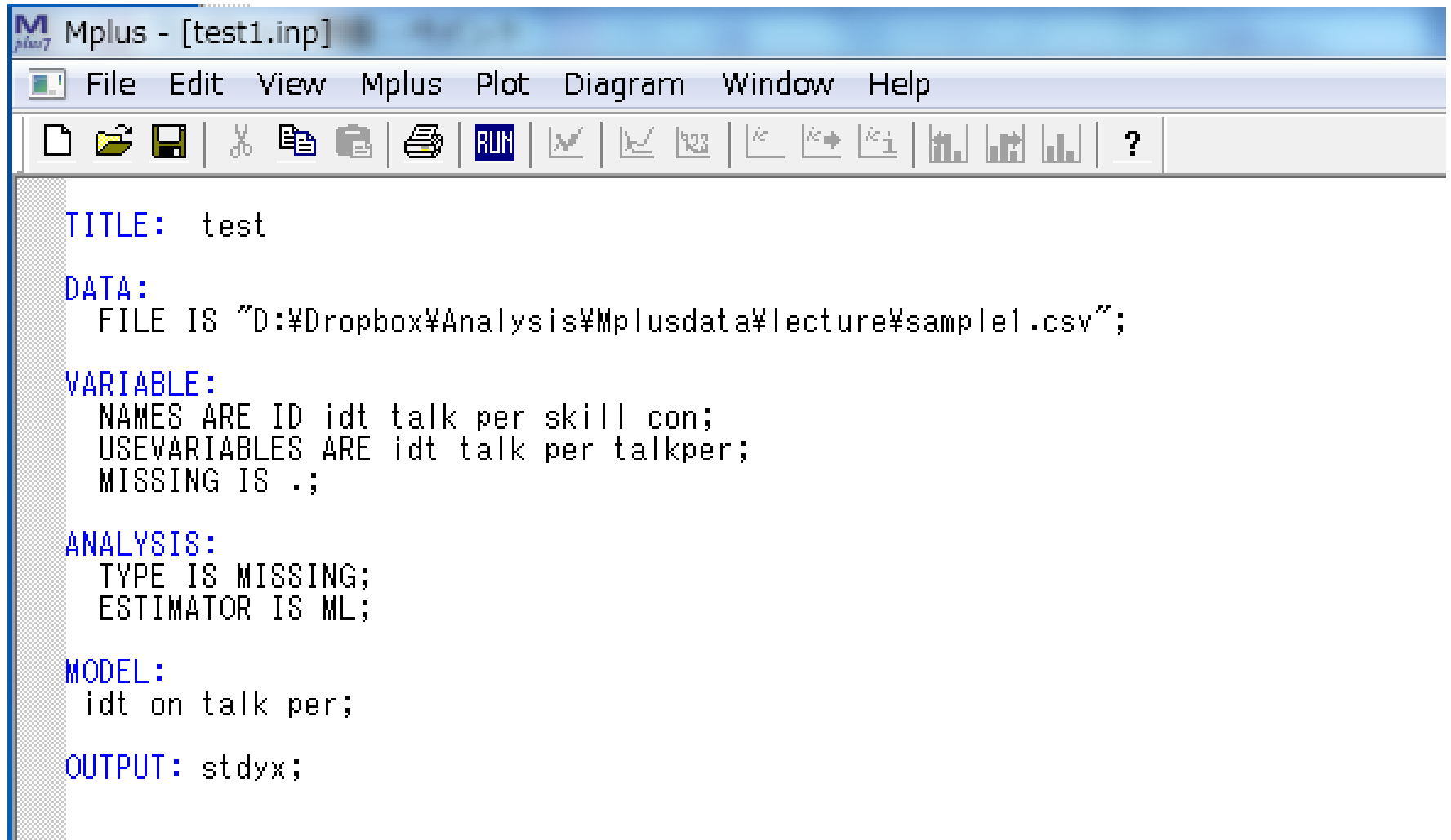
<http://norimune.net>

Mplus導入編

Mplusとは

- Muthén & Muthénが作った統計ソフトウェア
 - 主に構造方程式モデリングのためのソフト
 - <http://www.statmodel.com/>
- 非常に高度な分析が可能
 - Amosができることは全部可能
 - カテゴリカルデータ分析、マルチレベル分析、潜在クラス分析、それらの組み合わせ、など
- CUI (Character User Interface)を採用
 - 最初はやや慣れが必要
 - 最新版(Mplus7)は、一部GUIに対応

Mplusの編集画面



The screenshot shows the Mplus software interface. The title bar reads "Mplus - [test1.inp]". The menu bar includes "File", "Edit", "View", "Mplus", "Plot", "Diagram", "Window", and "Help". The toolbar contains icons for file operations (New, Open, Save, Cut, Copy, Paste), a "RUN" button, and various plot and diagram tools. The main text area contains the following syntax file content:

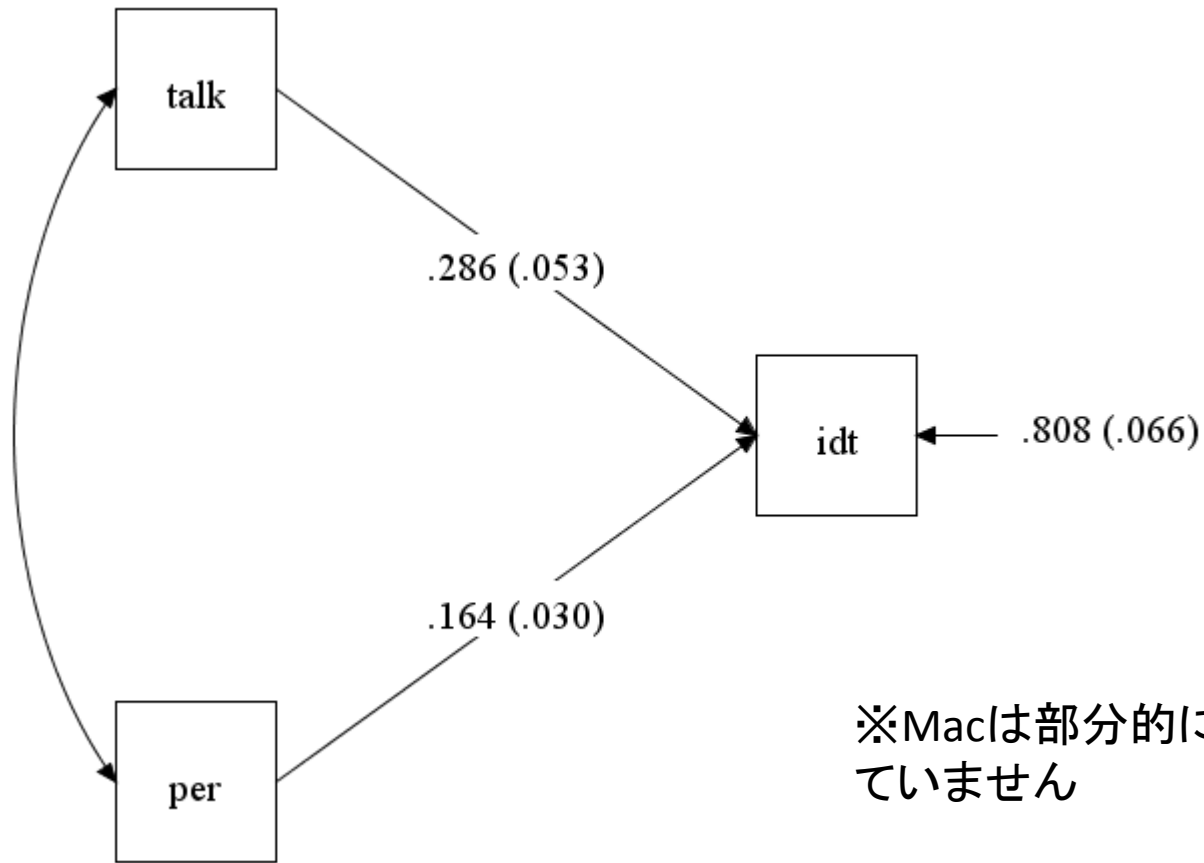
```
TITLE: test  
DATA:  
  FILE IS "D:\Dropbox\Analysis\Mplusdata\lecture\sample1.csv";  
VARIABLE:  
  NAMES ARE ID idt talk per skill con;  
  USEVARIABLES ARE idt talk per talkper;  
  MISSING IS .;  
ANALYSIS:  
  TYPE IS MISSING;  
  ESTIMATOR IS ML;  
MODEL:  
  idt on talk per;  
OUTPUT: stdyx;
```

Mplusの出力

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
IDT ON				
TALK	0.286	0.053	5.426	0.000
PER	0.164	0.030	5.462	0.000
Intercepts				
IDT	1.801	0.211	8.518	0.000
Residual Variances				
IDT	0.808	0.066	12.166	0.000

Mplusの出力 ダイアグラム



どんな分析ができるのか？

- 構造方程式モデル
 - 内生変数の分布も様々なものから指定可能
 - 順序・名義カテゴリーカル → 項目反応理論も可能
 - ポアソン, 負の二項分布
 - 打ち切りデータ
 - 様々な推定法に対応
 - 最尤法, 一般化最小二乗法, 重みつき最小二乗法, ベイズ推定法,
- 潜在混合分布モデル (Mixtureアドインが必要)
 - 潜在的なグループにサブジェクトを分ける
- マルチレベルモデル (Multilevelアドインが必要)
 - ネストされたデータを分析する手法

でも、お高いんでしょう？

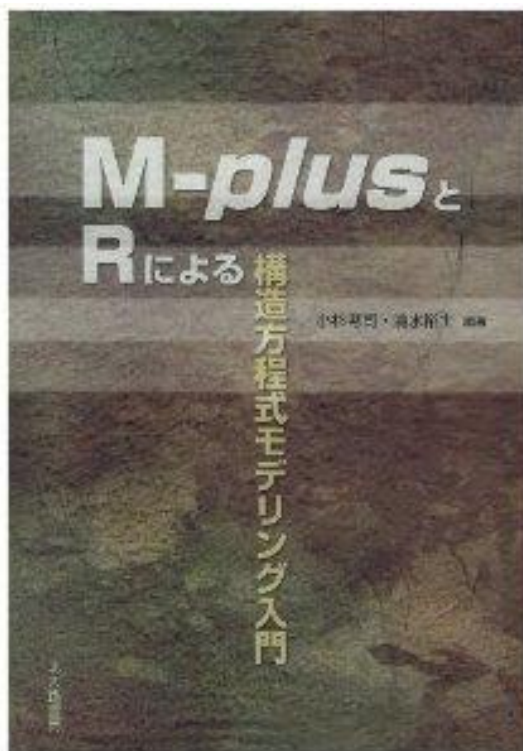
- 機能の割にかなり安い
 - 学生の場合 (Student Pricing)
 - 基本モデル 2万円
 - 全部入り 3万5千円
 - 教員の場合 (University Pricing)
 - 基本モデル 6万円
 - 全部入り 9万円
 - しかも、同じバージョンなら永久に使える
 - アップデートする場合は、1年間は無料。それ以降は有料

Amosとの違い

- GUI (Amos) か CUI (Mplus) か
 - これには一長一短がある
 - 簡単なパスモデルはAmosのほうが簡単
 - Mplusだと、“F1 by v1-v40;” というコードだけで40項目の因子分析ができるが、Amosだと大変。
- 分析の機能
 - 完全にMplus > Amos
 - Amosでできることはすべてできる
- ライセンスの扱い
 - AmosはIBMなので高価 & 年単位で更新が必要
 - Mplusは一度購入すれば、同じバージョンならずっと使える
- 日本語への対応
 - Amosは日本語に完全対応 Mplusは完全不対応
 - 今後もMplusは日本語に対応することはない・・・が、

【宣伝】Mplusの日本語解説本，登場【販促】

- 小杉考司・清水裕士 編著 (2014)
MplusとRによる構造方程式モデリング 北大路書房



Mplus初級編

資料のダウンロード

- 以下のURLからダウンロードできます

<http://bit.ly/1gjldfV>

サンプルデータ紹介

- 仮想的なデータを利用
 - 3人集団が集団討議を行う実験(100集団300人)
 - ※実際に実験は行っていません!
 - 何が集団アイデンティティを高める要因となるのか?
- 測定変数
 - 集団アイデンティティ → idt
 - 発言量 → talk
 - 集団パフォーマンス → per

 - コミュニケーションスキル → skill
 - 実験条件(0が同じ情報条件、1が違う情報条件) → con

初級編メニュー

- データの読み込み
 - データセットの準備
 - 基本的なコードの書き方
- とりあえず回帰分析
 - 重回帰分析
 - パス解析
- とりあえず因子分析
 - 確認的因子分析
 - 探索的因子分析
- 便利な機能や知っておくと便利な事柄

Mplusの構成

- データファイル
 - Mplusのデフォルトでは.datの拡張子。でも.txtとかでもよい。
- 入力ファイル
 - inpファイルと呼ばれる。これにすべてのコードを書く。
 - 拡張子は.inp
- 出力ファイル
 - outファイルと呼ばれる。これに分析結果が出力される。
 - 拡張子は.out

データファイルの読み込み

- 読み込みファイル
 - datファイル(.dat)
 - テキストファイル(.txt)
 - csvファイル(.csv)
- データ区切り
 - スペース
 - タブ
 - カンマ
- 僕個人のオススメは、タブ区切りのテキストファイル

例: エクセルのデータからMplusへ 0

- Mplus用のフォルダを作る
 - Mplusは日本語のパスを読めない!
 - ユーザー名が日本語の人は、マイドキュメントにおいても認識されない
 - 清水からのオススメ
 - C:¥mplusfile¥~など、Cドライブに直接フォルダを作る
 - 僕はDドライブにDropboxを置いているので、そこに。

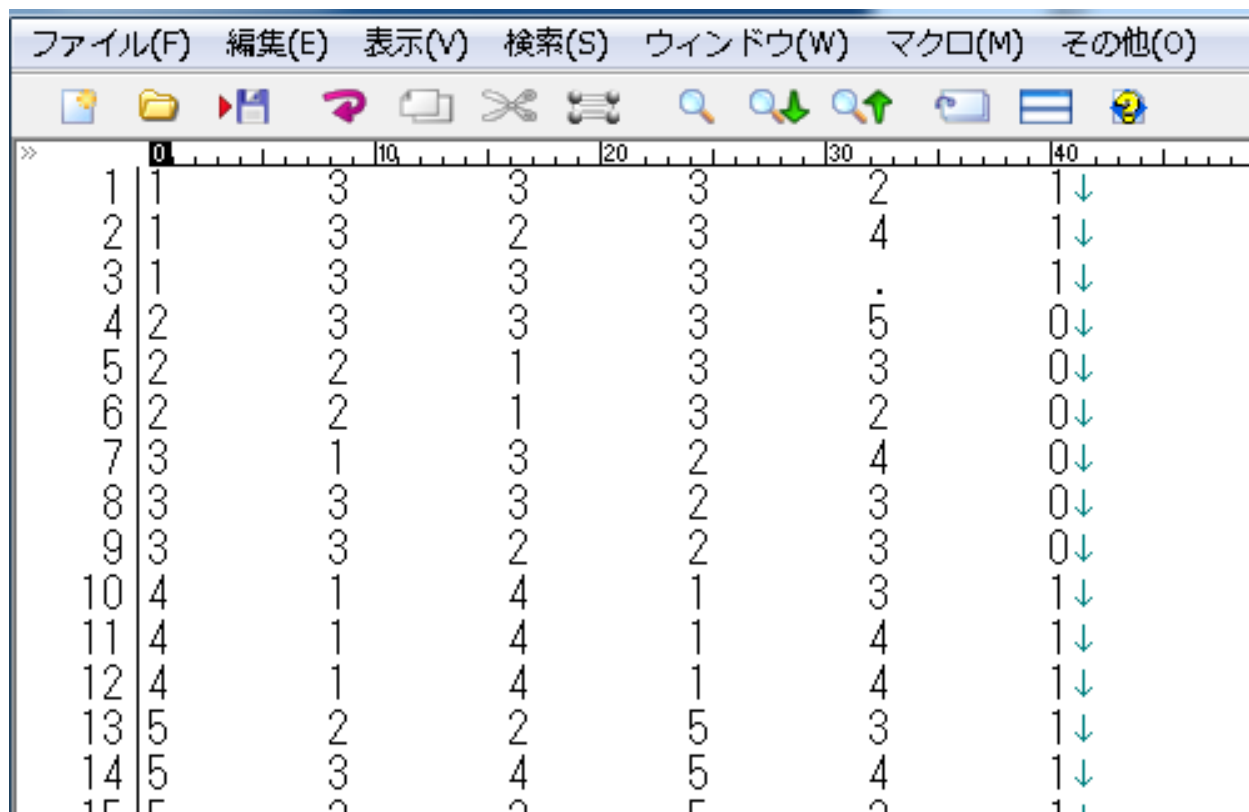
例: エクセルのデータからMplusへ 1

- エクセルのデータを選択してコピー
 - 変数名は含めない・・・変数名はあとで入力

	A	B	C	D	E	F	G
1	group	idt	talk	per	skill	con	
2	1	3	3	3	2	1	
3	1	3	2	3	4	1	
4	1	3	3	3		1	
5	2	3	3	3	5	0	
6	2	2	1	3	3	0	
7	2	2	1	3	2	0	
8	3	1	3	2	4	0	
9	3	3	3	2	3	0	
10	3	3	2	2	3	0	
11	4	1	4	1	3	1	
12	4	1	4	1	4	1	
13	4	1	4	1	4	1	
14	5	2	2	5	3	1	
15	5	3	4	5	4	1	
16	5	2	3	5	2	1	

例: エクセルのデータからMplusへ 2

- メモ帳にそのまま張り付ける



The screenshot shows a Notepad window with a menu bar (ファイル(F), 編集(E), 表示(V), 検索(S), ウィンドウ(W), マクロ(M), その他(O)) and a toolbar. The main area contains a table of data with 17 rows and 7 columns. The data is as follows:

1	1	3	3	3	2	1↓
2	1	3	2	3	4	1↓
3	1	3	3	3	.	1↓
4	2	3	3	3	5	0↓
5	2	2	1	3	3	0↓
6	2	2	1	3	2	0↓
7	3	1	3	2	4	0↓
8	3	3	3	2	3	0↓
9	3	3	2	2	3	0↓
10	4	1	4	1	3	1↓
11	4	1	4	1	4	1↓
12	4	1	4	1	4	1↓
13	5	2	2	5	3	1↓
14	5	3	4	5	4	1↓
15	5	3	4	5	4	1↓

例: エクセルのデータからMplusへ 3

- ファイルの保存場所
 - Mplusの入力ファイル(.inp)と同じフォルダに入れることをオススメする。詳細は後述。
 - さっき作った、Mplus用のフォルダの中に入れる
- 保存方法
 - 保存するとき、拡張子は.datのほうが便利かもしれないけど、そこは好みで。
 - .datか、.txt、.csvあたりで。
 - 日本語は不対応なので、半角英数字で。

Mplus言語のルール

- 文章はすべて半角英数字
 - 日本語不可
 - 大文字・小文字の区別はない
- 文末にはセミコロン”;"を付ける
 - 改行だけでは、同じ文章だと判断される
- “IS”や”ARE”は”=”で代用できる
 - マニュアルには”IS”とか”ARE”が使われているが、”=”で大丈夫。
- その他
 - コメントアウトは”!”

こんな感じ

```
TITLE: test
```

```
DATA:  
FILE IS "D:\Dropbox\Analysis\Mplusdata\lecture\sample1.txt";
```

```
VARIABLE:  
NAMES ARE idt talk per skill con;  
USEVARIABLES ARE idt talk per;  
MISSING IS .;
```

```
ANALYSIS:  
TYPE IS GENERAL;  
ESTIMATOR IS ML;  
ITERATIONS = 1000;  
CONVERGENCE = 0.00005;
```

```
OUTPUT: STDYX;
```

Mplusの基本コード

- DATA:
 - データファイルのある場所を指定
- VARIABLE:
 - 分析に使う変数の指定
- ANALYSIS:
 - 分析方法の指定
- MODEL:
 - モデルの指定
- OUTPUT:
 - 出力の設定
- その他
 - PLOT、SAVEDATA、DEFINEなどなど

DATA: コマンド

- 使うデータファイルを指定する
 - ファイルのパスを指定する
 - FILE = “D:¥Dropbox¥・・・¥sample1.dat”;
 - パスの最後にセミコロンを忘れずに！
 - inpファイルと同じフォルダに入れる場合・・・
 - FILE = “sample1.dat”;
 - ファイル名だけでよい。こちらが簡単なのでオススメ
- リストワイズ削除の時には以下を書く
 - LISTWISE = ON;
 - ただ、LISTWISEは基本は使わなくてよい

VARIABLE: コマンド

- 読み込んだ変数を指定する
 - データファイルにある変数名をすべて書く
 - NAMES = v1 v2 v3;
 - 上の場合、v1-v3; と書くこともできる
- 分析に使用する変数を指定する
 - USEVARIABLES = v1 v2;
 - 書き方は上と同じ
- 欠損値を指定する
 - MISSING = .;
 - ピリオドの場合は上のよう書く
 - ダブルコーテーションはつけない

ANALYSIS: コマンド

- 分析方法を指定する

- TYPE = GENERAL;

- 基本的なSEMをする場合はこれだけでOK。
 - 他に、TWOLEVEL(マルチレベルモデル)、MIXTURE(潜在クラス分析)、EFA(探索的因子分析)などがある。

- ESTIMATOR = MLR;

- デフォルトはMLR。ロバスト最尤法のこと。
 - 他に、ML(最尤法)、WLSMV(重みつき最小二乗法)、BAYES(ベイズ法)、GLS(最小二乗法)などがある。
 - Amosと結果を一致させたいなら、MLを使用するといい。

MODEL: コマンド

- モデルを指定する
 - モデルの書き方の基本は以下の5つ。
 - 回帰分析(構造方程式)は ON
 - 因子分析(測定方程式)は BY
 - 共分散は WITH
 - 変数名だけを書くと、分散を推定する指定
 - [変数名]と書くと、平均値を推定する指定
- パラメータの制約をする
 - MODEL CONSTRAINT:の後に書く
 - これについては、後述

OUTPUT: コマンド

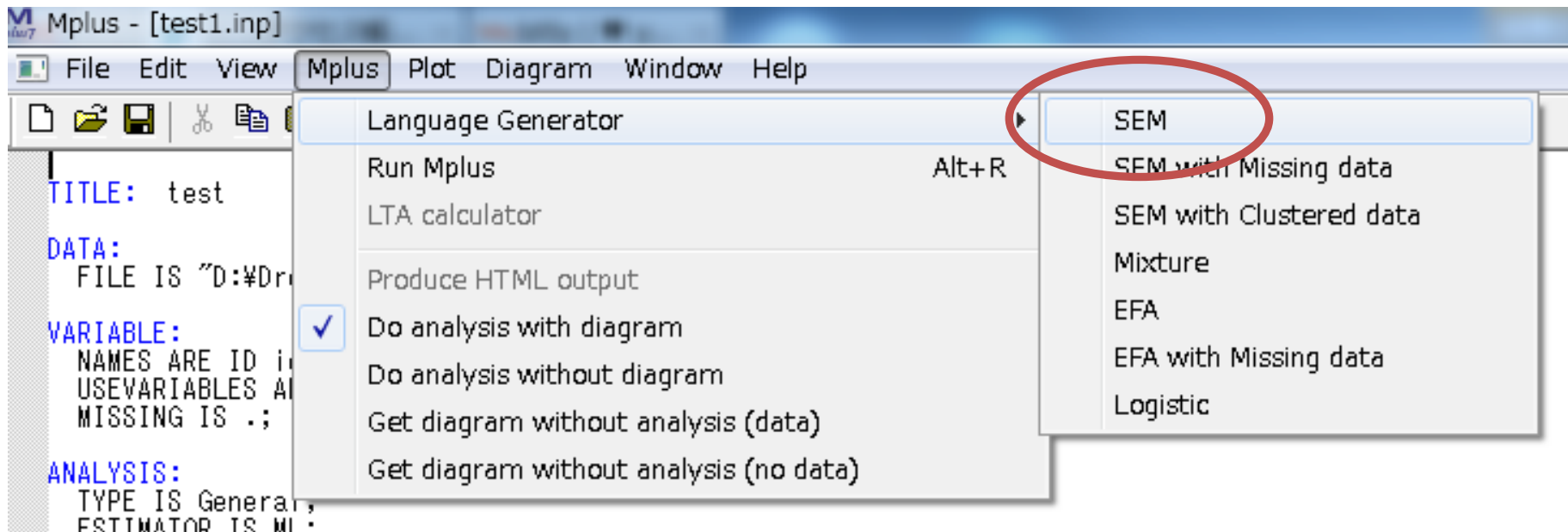
- 出力の設定をする
 - 記述統計量を出力する
 - SAMPSTAT
 - 標準化係数を出力する
 - STANDARIZED
 - しかし、STDYXと書いた方が余計なのが出なくて便利。
 - 信頼区間を出力する
 - CINTERVAL
 - ブートストラップ信頼区間を出力する場合は、CINTERVAL(BOOTSTRAP)と書く。

その他のコマンド

- **SAVEDATA: コマンド**
 - 因子得点などを保存する
 - `SAVE = FSOCRES;` で因子得点を保存
 - `File = ###.txt;` でファイル名を指定。日本語不可。
- **PLOT: コマンド**
 - グラフなどを出力する
 - IRTやベイズ推定の際に使う
- **DEFINE: コマンド**
 - 新しい変数を作ったり、変換したりするのに使う

Mplusの基本コードの書き方

- 最初は、Language Generatorで。



- 基本的なモデリングはこれで解決する。

これだけは覚えておこう

DATA:

FILE IS "sample1.dat"; !ここにファイル名を書く。日本語不可。

VARIABLE:

NAMES = ID idt talk per skill con; !ここにデータに入っている変数名すべて書く

USEVARIABLES = idt talk per; !ここに分析に使う変数を書く。

MISSING = .; !欠損値記号を書く。

ANALYSIS:

TYPE = GENERAL; !普通のSEMをするなら、GENERALでよい。

ESTIMATOR = ML; !MplusのデフォルトはMLR。MLにすればAMOSと同じ結果になる。

MODEL:

!ここにモデルを書く。

OUTPUT:

SAMPSTAT STDYX MODINDICES(ALL); !この3つは常に書いておいてよい。

Mplusを使う上での注意点

- ファイルの場所に注意！
 - パスの中に日本語が入っていると読み込めない
 - ユーザー名が日本語の場合, Cドライブに直接フォルダを置くなどの工夫が必要。
- 一行に80文字までしか読みこまない
 - 横に長すぎると, エラーが出る
 - こまめに改行をする必要がある
- Usevariablesには, 使う変数のみを指定すること
 - モデルに含まれていない変数が指定されていると, 適合度を下げる要因にしかない。

MODEL: コマンドの書き方

- 重回帰分析をやってみよう・・・ONを使う。
 - 目的変数: idt
 - 説明変数: talk per

idt on talk per;

- これで終わり！

重回帰分析のコード

DATA:

```
FILE IS "sample1.dat";
```

VARIABLE:

```
NAMES = ID idt talk per skill con;
```

```
USEVARIABLES = idt talk per;
```

```
MISSING = .;
```

ANALYSIS:

```
TYPE = GENERAL;
```

```
ESTIMATOR = ML;
```

MODEL:

```
idt on talk per;
```

OUTPUT:

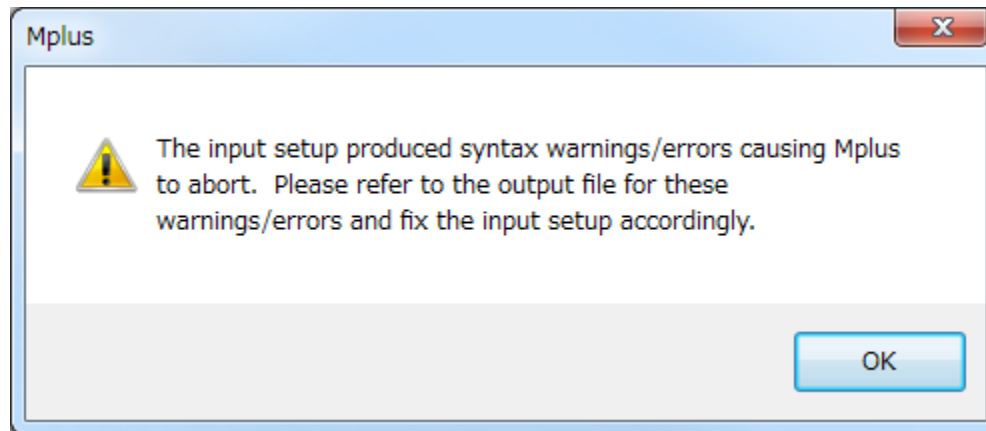
```
SAMPSTAT STDYX MODINDICES(ALL);
```

入力ファイルの保存と実行

- 分析を実行する前に、inpファイルを保存
 - 場所は、さっきつくったフォルダの中に
 - データファイルがある場所が便利
 - フルパスを書かなくてよいから
 - 拡張子は.inp
- inpファイルが保存できたら、RUNボタンを押す
 - 最初は、何かしらエラーが出る。
 - こころが折れそうになることもあるが、がんばる。

プログラムを実行！

- 最初は，ほぼ間違いなくエラーが出る



– この画面と仲良くなるのが最初の一步

警告とエラーの違い

- WARNING
 - 分析は走るけど、一応ゆっとくね、的なもの。
 - どの変数とも相関してない変数があるよー
 - モデルの記述が重複してるよー
 - その推定法はこのデータでは使えないからデフォルトに変えたよー
 - シレッと書いてあるので、注意すること
- ERROR
 - コードに問題あり、分析が走らない
 - ERRORに書いてあることをよく読んで、inpファイルを訂正する

よくあるコードのミス

- ~という変数が見つかりません
 - “Unknown variables in a ~~~~~:”
 - USEVARIABLESに指定していない
 - たまにあるのが、USEVARIABLESの行が長すぎて、80文字を超えてしまっていると、このエラーが出る
 - セミコロンが抜けている場合にも出ることがある
- 書いているオプションがわかりません
 - “Unrecognized setting for ~~~~ option:”
 - スペルミスか、セミコロンが抜けている

出力の見方

- 結果を見るとき的心得
 - できるだけ詳細をちゃんと見る
 - 分析がうまくいってない場合、走っても警告が出ることがある。
 - したかった分析がちゃんとできているか、チェック。
 - 手違いで分析法などが間違えているかもしれない
 - 適合度は必ず見る
 - 適合度がかなり悪いなら、データ入力のミスの可能性
 - 慣れないCUIでのモデリングは、ミスも多い

分析の要約

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	296
Number of dependent variables	1
Number of independent variables	2
Number of continuous latent variables	0
Observed dependent variables	
Continuous	
IDT	
Observed independent variables	
TALK	
PER	
Estimator	ML
Information matrix	OBSERVED
Maximum number of iterations	1000
Convergence criterion	0.500D-04
Maximum number of steepest descent iterations	20
Maximum number of iterations for H1	2000
Convergence criterion for H1	0.100D-03

適合度指標

MODEL FIT INFORMATION		
Number of Free Parameters		4
Loglikelihood		
H0 Value	-388.450	
H1 Value	-388.450	
Information Criteria		
Akaike (AIC)	784.900	
Bayesian (BIC)	799.661	
Sample-Size Adjusted BIC	786.976	
(n* = (n + 2) / 24)		
Chi-Square Test of Model Fit		
Value	0.000	
Degrees of Freedom	0	
P-Value	0.0000	
RMSEA (Root Mean Square Error Of Approximation)		
Estimate	0.000	
90 Percent C.I.	0.000	0.000
Probability RMSEA <= .05	0.000	
CFI/TLI		
CFI	1.000	
TLI	1.000	
Chi-Square Test of Model Fit for the Baseline Model		
Value	57.712	
Degrees of Freedom	2	
P-Value	0.0000	
SRMR (Standardized Root Mean Square Residual)		
Value	0.000	

- 情報量基準
 - モデル比較に使う
- X2乗値
 - モデルがデータと等しいという帰無仮説の確率
- RMSEA
 - 0.05 (0.10) 以下がよい
 - 儉約性を考慮した指標
- CFI
 - 0.95以上がよい
 - 儉約性をやや考慮した指標
- SRMR
 - 0.05 (0.10) 以下がよい
 - モデルとデータの距離

推定結果

- Estimate: 推定値
- S.E.: 標準誤差
- EST./S.E. z値 1.96以上で5%有意

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
IDT ON				
TALK	0.286	0.053	5.426	0.000
PER	0.164	0.030	5.462	0.000
Intercepts				
IDT	1.801	0.211	8.518	0.000
Residual Variances				
IDT	0.808	0.066	12.166	0.000

推定結果2 標準化係数

STANDARDIZED MODEL RESULTS

STDYX Standardization

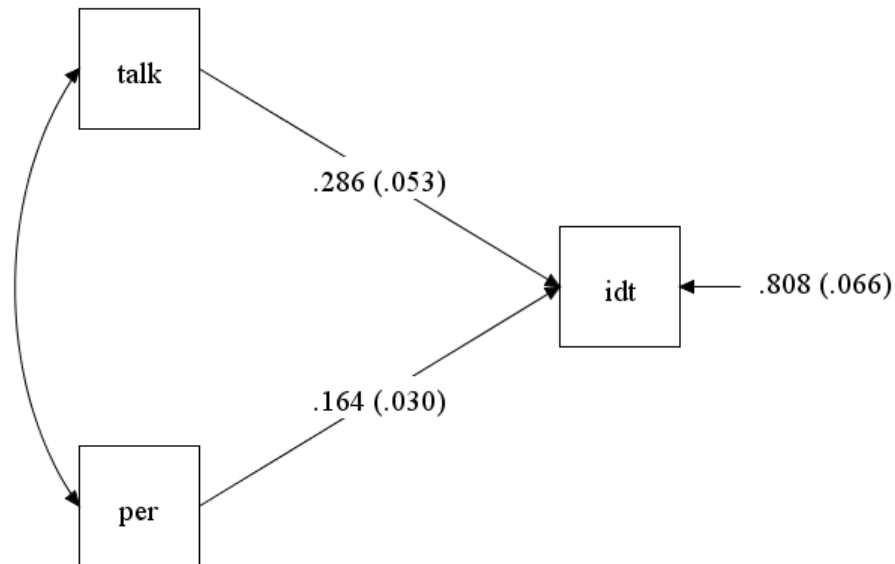
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
IDT ON				
TALK	0.287	0.051	5.637	0.000
PER	0.289	0.051	5.678	0.000
Intercepts				
IDT	1.818	0.260	6.983	0.000
Residual Variances				
IDT	0.823	0.040	20.439	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
IDT	0.177	0.040	4.400	0.000

ダイアグラムの出力

- メニューバーのDiagramを選択
 - View diagramをクリックすると、パス図が見られる
 - ()内は標準誤差 出力設定で変えられる



分散・共分散の推定

- 分散と共分散を推定してみよう・・・WITHを使う。
 - さっきと同じモデルで
 - ただし、独立変数の分散も同時に推定する

idt on talk per;

talk per;

talk with per;

- Mplusでは、デフォルトでは内生変数の分散しか推定しない。
- 外生変数の分散を指定すると、それらの変数間の共分散も自動的に推定する。
 - すべての欠損値を推定するためには、独立変数も分散を推定する必要がある。ただ、データ規模が大きくなるため、適合度が悪くなることもある

結果は・・・

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
IDT	ON				
	TALK	0.285	0.053	5.425	0.000
	PER	0.164	0.030	5.498	0.000
PER	WITH				
	TALK	0.131	0.101	1.297	0.195
Means					
	TALK	3.021	0.057	52.698	0.000
	PER	4.690	0.101	46.333	0.000
Intercepts					
	IDT	1.802	0.211	8.548	0.000
Variances					
	TALK	0.983	0.080	12.230	0.000
	PER	3.074	0.251	12.247	0.000
Residual Variances					
	IDT	0.806	0.066	12.186	0.000

共分散を推定したくない場合

- 固定母数の制約をする・・・@を使う。
 - 共分散を0に固定する

`talk with per@0;`

- これでtalkとperの共分散が0に固定される。

`idt on per@0;`

- とすれば、パス係数を0に固定することもできる。

共分散の指定あれこれ

- 一度に全部の変数の共分散を指定
idt talk per WITH idt talk per;
– これで3変数間の共分散を推定できる
- ペアごとの共分散を指定
v1 v2 v3 PWITH v4 v5 v6;
– V1とv4、v2とv5、v3とv6の共分散を推定
- 共分散の自動推定をオフにする
– ANALYSISコマンドにMODEL = NOCOVARIANCES と書く。
 - これで、因子間や分散を推定した外生変数間の共分散が自動で推定されなくなる

平均値の指定

- 平均構造の推定
 - TYPE = GENERALにしておけば、自動的に平均構造も推定される
 - 自分で指定しておきたい場合は、[]を使って
[idt];
 - と書く。
 - 平均値に固定制約を与えたい場合は、
[idt@0];
 - と書く。
 - 平均値を推定したくない場合は、ANALYSISコマンドのところに、MODEL = NOMEANSTRUCTURE;と書く。

パス解析

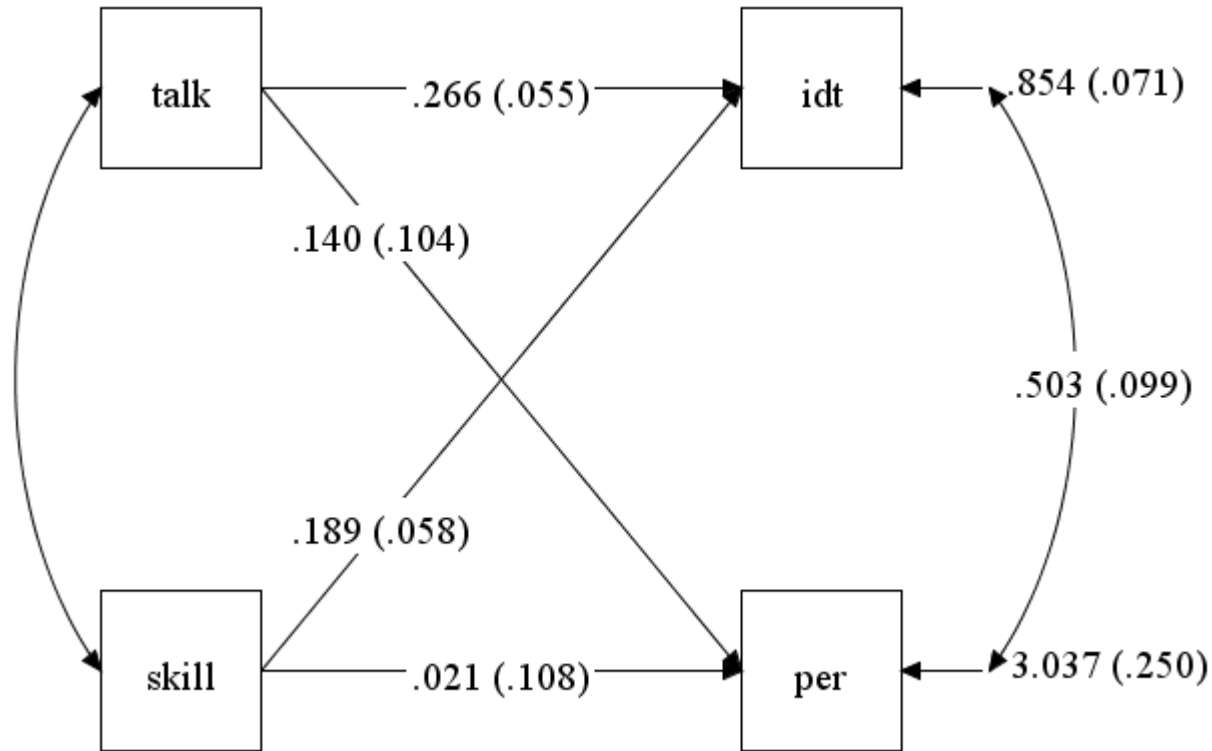
- ONとWITHでパス解析が可能
 - idt と per を talk と skill で予測するモデル
 - skillをusevariablesに追加するのを忘れずに！

idt per on talk skill;

idt with per;

- 複数の目的変数に同じ説明変数が予測する場合は、一行で書ける。

パス解析のダイアグラム



媒介分析

- 間接効果を検定する・・・INDを使う。
 - MODEL INDIRECTオプションを利用
 - Idt と conの関係をtalkが仲介するかを検定

idt on talk con;

talk on con;

MODEL INDIRECT:

idt IND talk con;

- con → talk → idtという媒介効果を検定

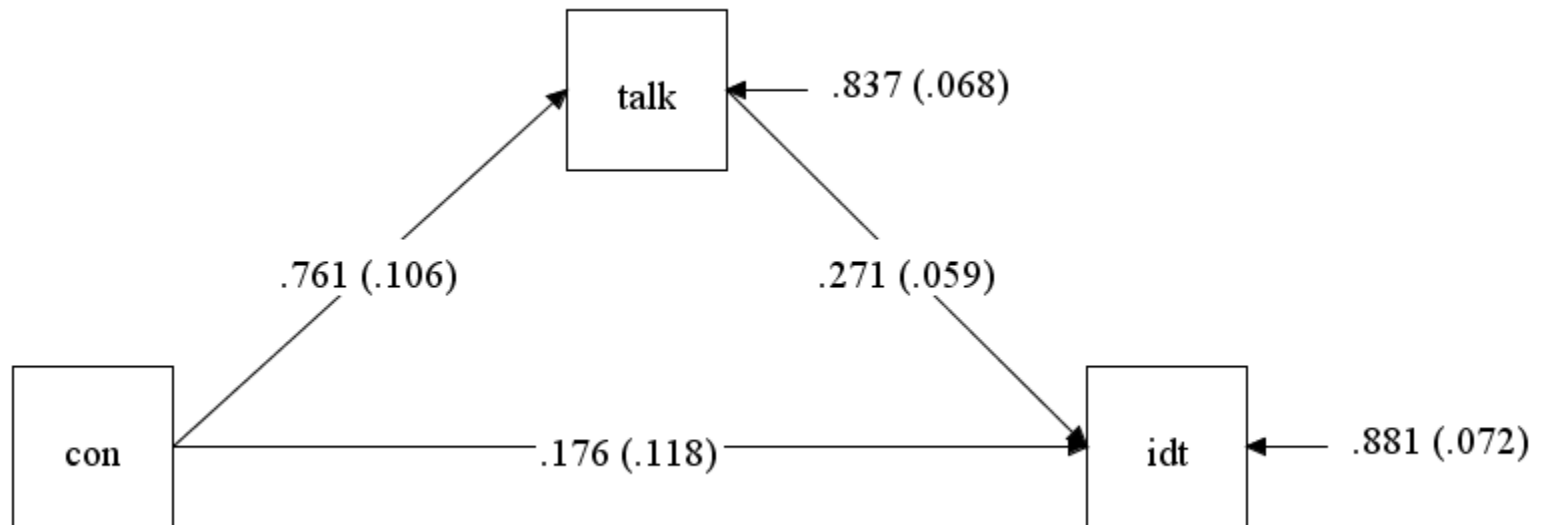
間接効果の検定結果

- Specific indirectのところを見る。
 - いわゆるSobel testの結果。

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from CON to IDT				
Sum of indirect	0.207	0.054	3.860	0.000
Specific indirect				
IDT				
TALK				
CON	0.207	0.054	3.860	0.000

媒介分析のダイアグラム



ブートストラップ信頼区間

- 間接効果のブートストラップ信頼区間
 - BOOTSTRAPオプションを使う
 - ANALYSISコマンドに以下の文を書く。
BOOT = 1000;
 - 回数は1000～10000回の範囲で。
 - CINTERVALオプションで信頼区間を出力する
 - OUTPUTコマンドに以下の文を書く。
CINTERVAL(BOOT)

コードは以下の通り

ANALYSIS:

```
TYPE = GENERAL;  
ESTIMATOR = ML;  
BOOT = 1000;
```

※リサンプリング回数を指定

MODEL:

```
idt on talk con;  
talk on con;
```

MODEL indirect:

```
idt IND talk con;
```

OUTPUT:

```
SAMPSTAT STDYX MODINDICES(ALL) CINTERVAL(BCBOOT);
```

※BCBOOTはバイアス修正信頼区間

ブートストラップ信頼区間の結果

CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
Effects from CON to IDT							
Sum of indirect	0.043	0.096	0.117	0.207	0.285	0.297	0.308
Specific indirect							
IDT							
TALK							
CON	0.043	0.096	0.117	0.207	0.285	0.297	0.308

- 信頼区間に0が含まれてなければ有意

確認的因子分析

- sample2.datを使う
 - IDと6項目の変数
 - v1-v3が同じ因子、v4-v6は別の因子
- 因子分析をする・・・BYを使う。
 - F1とF2という潜在変数がv1-v10で測定されている

F1 by v1-v3;

F2 by v4-v6;

- F1とF2の因子間相関は自動的に推定される
 - もちろん、F1 with F2@0; と書けば、直交回転となる。

確認的因子分析

- いくつかの注意点

- 普通にF1 by v1-v3;とかくと、v1への因子負荷量は、自動的に1に固定される

- これは、モデルの識別のために必要な制約

- ただ、分散を1に固定しても因子分析は可能

- ただし、因子が外生変数の場合に限る

F1@1;

- と書けば、F1の分散を1に固定することができる。

- また、v1を自由推定するように指定することができる。

F1 by v1* v2-v3;

- あるいは F1 by v1-v5*; *は自由推定の指定。

以下のコードを書く

DATA:

```
FILE IS "sample2.dat";
```

VARIABLE:

```
NAMES = ID v1-v6 sex;  
USEVARIABLES = v1-v6;  
MISSING = .;
```

ANALYSIS:

```
TYPE = GENERAL;  
ESTIMATOR = ML;
```

MODEL:

```
F1 by v1-v3*;  
F2 by v4-v6*;  
F1-F2@1;
```

OUTPUT:

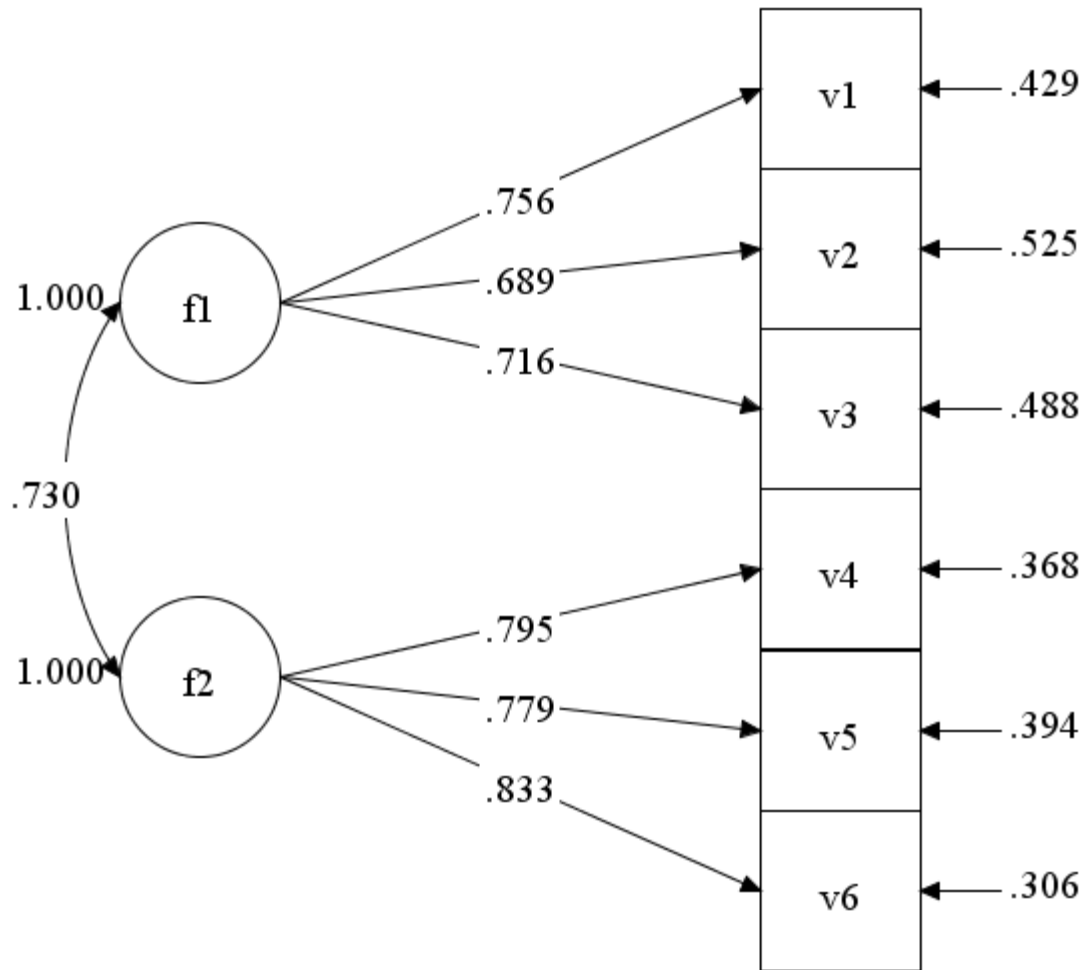
```
SAMPSTAT STDYX MODINDICES(ALL);
```

確認的因子分析の結果

STDYX Standardization

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F1	BY				
	V1	0.756	0.045	16.929	0.000
	V2	0.689	0.049	14.060	0.000
	V3	0.716	0.047	15.185	0.000
F2	BY				
	V4	0.795	0.035	22.711	0.000
	V5	0.779	0.037	21.233	0.000
	V6	0.833	0.033	25.567	0.000
F2	WITH				
	F1	0.730	0.053	13.875	0.000

- 因子分析の場合は標準化のほうが解釈しやすい



確認的因子分析の適合度

- 因子分析は適合度が低くなりがち
 - SRMRはかなり悪くなるので、あまり参照しない
 - むしろ、CFIやRMSEAに注目する
 - CFIは.85ぐらいでもそこそこ報告例がある
 - RMSEAは0.05以下が望ましいが、0.10以下でもあり
- 因子数をどう決定するか
 - BICなどの情報量基準を用いる
 - BICが一番小さくなる因子数を採用
 - それだけではなくRMSEAなども適宜参照

確認的因子分析の誤差相関

- 誤差相関とは
 - 測定項目の誤差項の間に共分散を仮定する

F1 by v1-v3;
v1 with v2;

 - 例えば、上のような感じ。
- どういう時に誤差相関を仮定するか
 - 測定方法論上、測定誤差に相関が仮定される場合
 - 質問項目が似ている(同じ言葉が含まれている、など)
 - 同じことについて、違う側面から質問している
 - 適合度を上げるためだけにむやみに仮定するべきではない
 - しかし、相関を仮定することが妥当なら、仮定する方がよい。

モデルの識別

- 識別問題とは？
 - 特に潜在変数などを利用すると、推定するべきパラメータが多くなり、モデルが識別できない
 - Mplusの場合、標準誤差の推定ができなくなる
 - 固定母数を利用して、モデルを識別させる
- 典型的な識別問題への対処
 - 因子負荷量を1に固定
 - 因子の分散を1に固定（因子が外生変数のとき）
 - 因子負荷量を等値にする（2項目の場合など）

因子負荷量の等値制約

- 各項目の因子負荷量が等しいという仮定
 - モデルの儉約性が向上する可能性がある
 - パラメータ名を設定し、等値に指定する

F1 by v1-v3*(p1);

F2 by v4-v6*(p2);

- v1-v3は、p1という同じパラメータであることを指定
- v4-v6も同様に、p2という同じパラメータ

左が普通 右が等値制約

MODEL FIT INFORMATION			
Number of Free Parameters		19	
Loglikelihood			
H0 Value		-1232.666	
H1 Value		-1228.609	
Information Criteria			
Akaike (AIC)		2503.332	
Bayesian (BIC)		2566.000	
Sample-Size Adjusted BIC		2505.806	
		($n^* = (n + 2) / 24$)	
Chi-Square Test of Model Fit			
Value		8.115	
Degrees of Freedom		8	
P-Value		0.4224	
RMSEA (Root Mean Square Error Of Approximation)			
Estimate		0.008	
90 Percent C.I.		0.000	0.084
Probability RMSEA <= .05		0.731	
CFI/TLI			
CFI		1.000	
TLI		1.000	
Chi-Square Test of Model Fit for the Baseline Model			
Value		491.834	
Degrees of Freedom		15	
P-Value		0.0000	
SRMR (Standardized Root Mean Square Residual)			
Value		0.019	

MODEL FIT INFORMATION			
Number of Free Parameters		15	
Loglikelihood			
H0 Value		-1233.953	
H1 Value		-1228.609	
Information Criteria			
Akaike (AIC)		2497.906	
Bayesian (BIC)		2547.381	
Sample-Size Adjusted BIC		2499.859	
		($n^* = (n + 2) / 24$)	
Chi-Square Test of Model Fit			
Value		10.688	
Degrees of Freedom		12	
P-Value		0.5558	
RMSEA (Root Mean Square Error Of Approximation)			
Estimate		0.000	
90 Percent C.I.		0.000	0.066
Probability RMSEA <= .05		0.865	
CFI/TLI			
CFI		1.000	
TLI		1.003	
Chi-Square Test of Model Fit for the Baseline Model			
Value		491.834	
Degrees of Freedom		15	
P-Value		0.0000	
SRMR (Standardized Root Mean Square Residual)			
Value		0.043	

潜在変数を含むパス解析

- 因子間に因果関係を仮定する
 - F1がF2を回帰するような場合。

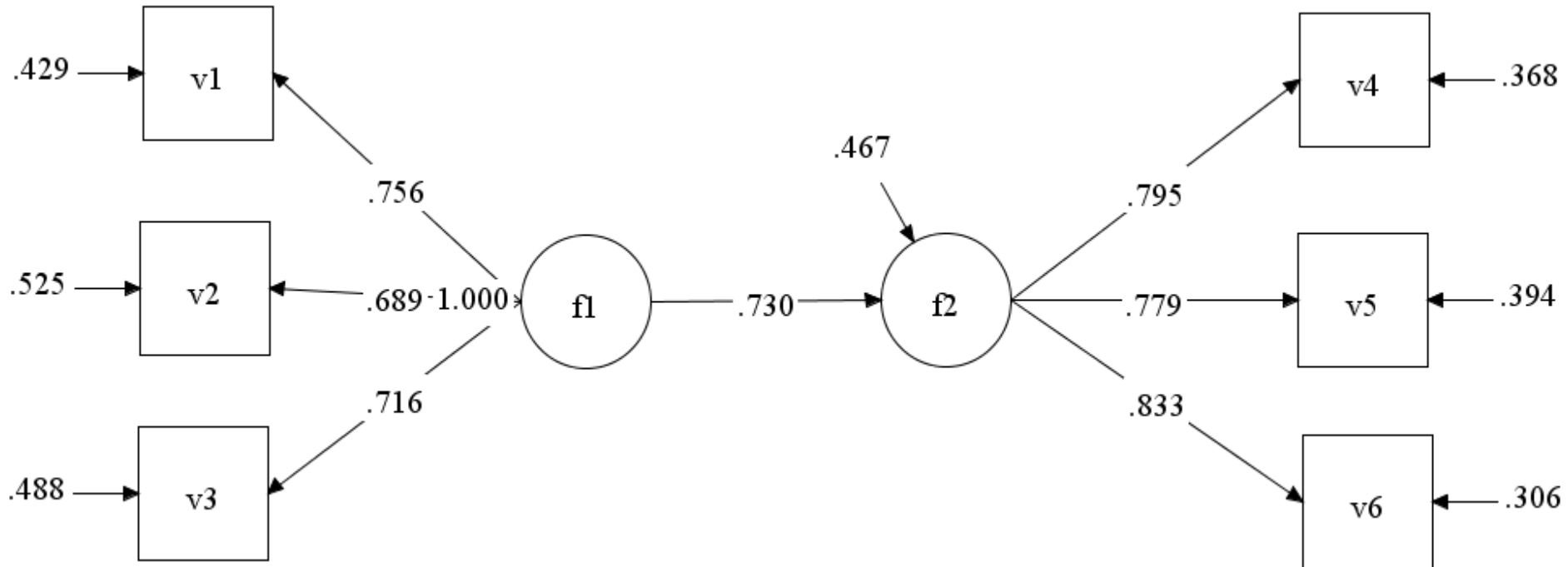
F1 by v1- v3;

F2 by v4-v6;

F2 on F1;

- 今回は、因子負荷量を1に固定して推定

分析結果(標準化係數)



MIMICモデル

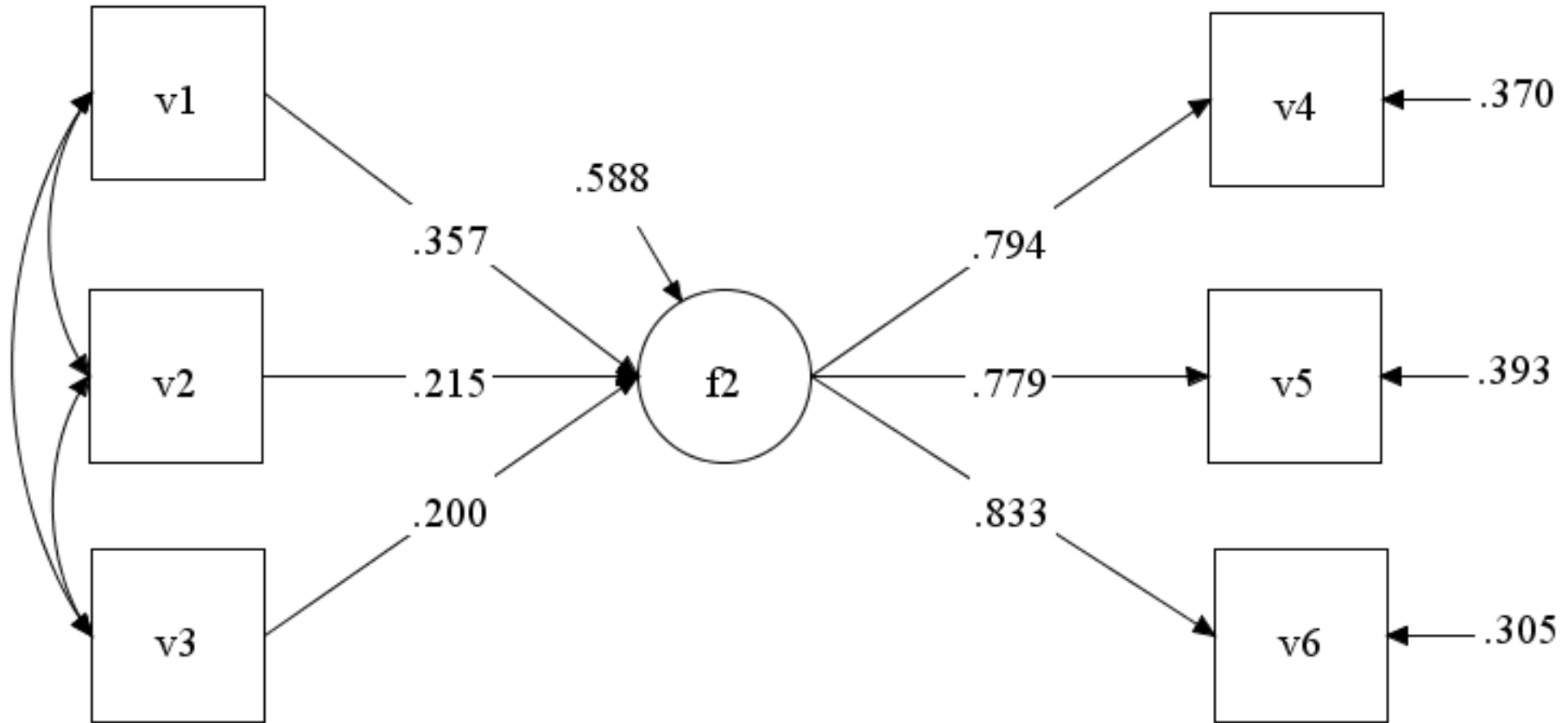
- Multiple indicator Multiple Cause Model
 - 複数の説明変数が、潜在変数を経て複数の目的変数に影響するモデル。

F2 by v4-v6;

F2 on v1-v3;

- 特に難しい制約は必要ない。

MIMICモデル ダイアグラム



探索的因子分析

- 事前にどの項目がどういう因子を構成するかわかってない場合に使う方法
 - SPSSなどに入っている因子分析と同じ。
- 2種類の方法がある
 - EFA(探索的因子分析)を使う
 - プロマックス回転など使える回転が多い
 - ESEM(探索的構造方程式モデル)を使う
 - 因子得点をそのままSEMに使える
 - 回転は、プロマックスは使えないが、たいていは使える

EFAでのやり方

- ANALYSISコマンドをTYPE = EFAとする。
 - EFAのあと、何因子を指定するかを最小値と最大値を指定。
 - 下の例の場合、1因子と2因子の2つを走らせる。
 - ROTATIONオプションに、回転法を指定。
 - VARIMAX, PROMAX, OBIIMIN, GEOMIN
 - デフォルトはGEOMIN回転
- MODELコマンドはいらない。

ANALYSIS:

TYPE = EFA 1 2;

ESTIMATOR = ML;

ROTATION = oblimin;

EFAの結果

OBLIMIN ROTATED LOADINGS (* significant at 5% level)

	1	2
V1	0.674*	0.089
V2	0.697*	-0.001
V3	0.760*	-0.039
V4	-0.071	0.865*
V5	0.029	0.759*
V6	0.105	0.743*

OBLIMIN FACTOR CORRELATIONS (* significant at 5% level)

	1	2
1	1.000	
2	0.697*	1.000

ESEMでのやり方

- MODELコマンドを書く必要がある
 - モデルの最後に(*1)と書く。
 - 回転法も指定してやる

ANALYSIS:

TYPE = GENERAL;

ESTIMATOR = ML;

ROTATION = oblimin;

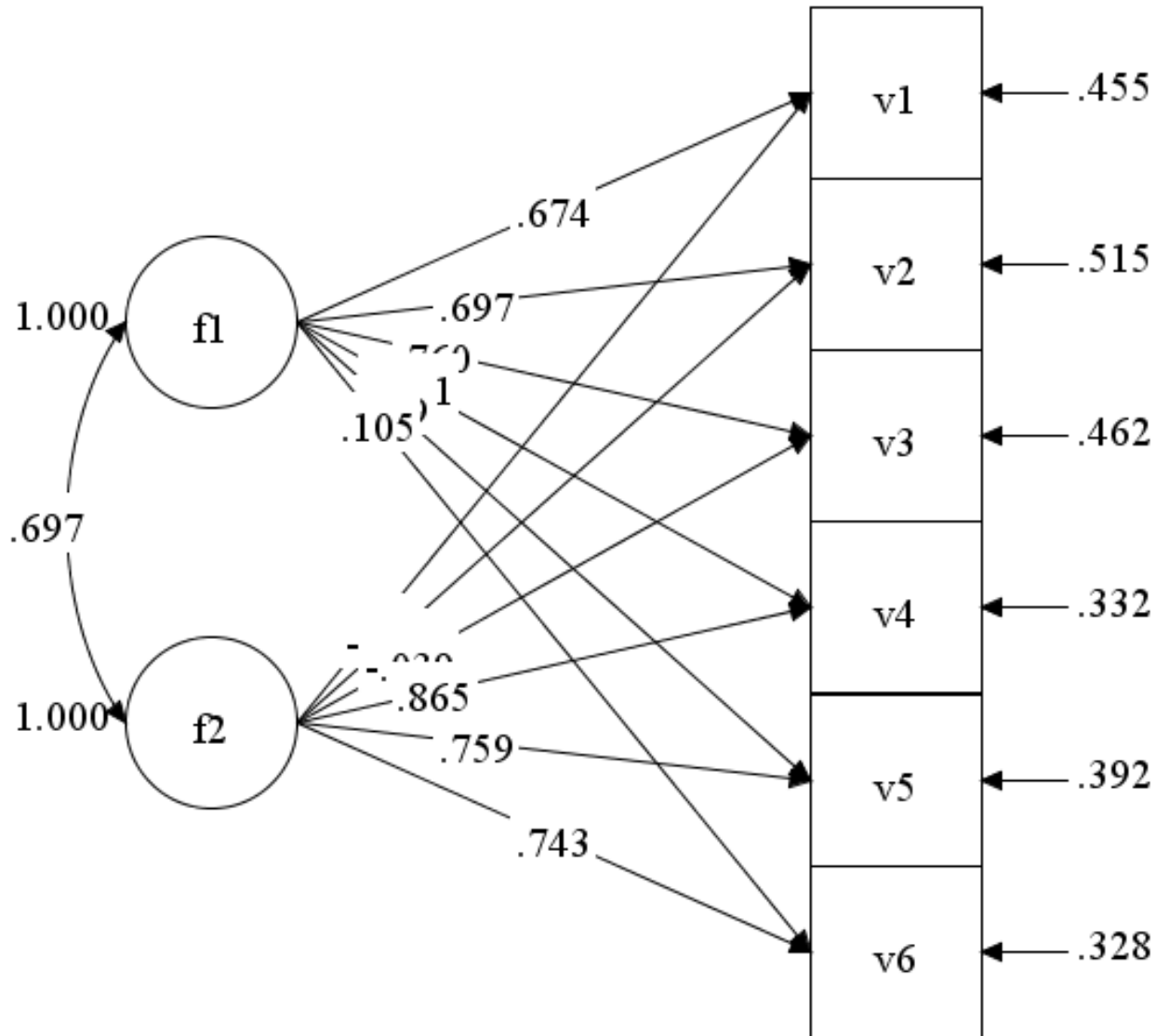
MODEL:

F1-F2 by v1-v6(*1);

ESEMの結果 EFAと同じ

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F1	BY				
	V1	0.674	0.111	6.082	0.000
	V2	0.697	0.101	6.893	0.000
	V3	0.760	0.084	9.001	0.000
	V4	-0.071	0.045	-1.576	0.115
	V5	0.029	0.089	0.329	0.742
	V6	0.105	0.091	1.155	0.248
F2	BY				
	V1	0.089	0.107	0.824	0.410
	V2	-0.001	0.092	-0.008	0.994
	V3	-0.039	0.064	-0.607	0.544
	V4	0.865	0.058	14.950	0.000
	V5	0.759	0.084	8.999	0.000
	V6	0.743	0.084	8.822	0.000
F2	WITH				
	F1	0.697	0.061	11.482	0.000

ESEMのダイアグラム



2つ以上の探索的因子分析

- モデルの中に二つ以上の探索的因子分析が含まれている場合
 - (*1)とは別に(*2)とする。

ANALYSIS:

TYPE = GENERAL;

ESTIMATOR = ML;

ROTATION = oblimin;

MODEL:

F1 by v1-v3(*1);

F2 by v4-v6(*2);

変数の作成や便利なオプション

DEFINEコマンドの利用 P574

- 新しい変数を作成
 - 読み込んだ変数を利用して、新しい変数を作る
 - 数値変換
 - 線形・非線形変換、合成を行う
 - 条件式を利用した変数の作成
 - IF文を使って、条件に当てはまる場合の得点を与える
 - 関数を利用した変数の作成
 - Mplusに入っている関数を使って変数を合成・変換

数値変換

- 既存の変数で新しい変数を作る
 - $X1 = v1 + v2;$
 - $X2 = v1 * v2 + v3 / v4 ;$
 - $X3 = \text{LOG}(v1);$
 - 自然対数。常用対数は $\text{LOG10}()$
 - $X4 = \text{ABS}(v2);$
 - 絶対値
 - $X5 = \text{SQRT}(v3);$
 - 平方根

条件式を使う変換

- IF文を使う

- IF(v1 == 1 AND v2 == 1) THEN group = 1;

- IF(v1 == 2 OR v2 == 1) THEN group = 2;

- groupは新しく作った変数名

- “==”は等値を意味しており、“=”は代入を意味している
 - 混同しないように注意！

- IF(v1 == 4) THEN group = _MISSING;

- 欠損値に指定

Mplusの関数を使う

- MEAN: 平均値を算出する
 - $F1 = \text{MEAN}(v1-v3);$
- SUM: 合計値を算出する
 - $F2 = \text{SUM}(v4-v6);$
- CUT: カテゴリに分割する
 - $\text{CUT } v1-v3(3);$
 - $v1-v3$ を、3以下を0、3より大を1にコードする
 - $\text{CUT } v1 v3 v6(2 4);$
 - $v1$ と $v3$ と $v6$ を、2以下を0、2より大4以下を1、4より大を2にコードする

DO文を使うと便利

- 複数の変数を一度に作成

```
x1 = SQRT(v1);
```

```
x2 = SQRT(v2);
```

```
x3 = SQRT(v3);
```

```
x4 = SQRT(v4);
```

```
x5 = SQRT(v5);
```

– こんなのも、下の1行で書ける

```
DO(1, 5) x# = SQRT(v#);
```

DEFINEで定義した変数の利用

- USEVARIABLESオプションで指定する。
 - ただし、既存の変数の後に指定。
 - 例えば、次のように書く

VARIABLE:

```
NAMES = ID v1-v6;
```

```
USEVARIABLES = f1 f2;
```

```
MISSING = .;
```

DEFINE:

```
f1 = mean(v1-v3);
```

```
f2 = mean(v4-v6);
```

DEFINE機能の便利な使い方

- いろんな従属変数で分析したい場合
 - USEVARIABLESとMODELの両方を変更しないといけない。
 - 次のように書いてみる
 - xの変数を変えるだけで、モデルを変更することなくいろんなモデルを試せる

VARIABLE:

```
NAMES = ID v1-v6;
```

```
USEVARIABLES =v2-v5 x;
```

```
MISSING = .;
```

DEFINE:

```
x = v1;
```

MODEL:

```
x on v2-v5;
```

DATA: コマンドのTYPEオプション

- デフォルトはINDIVIDUAL
 - 個人と変数の行列形式
 - 指定しなくてもこのタイプになる
- 要約データでの入力
 - 共分散行列形式
 - FULLCOV
 - 下三角行列の場合はCOV
 - 相関行列形式
 - FULLCORR
 - 下三角行列の場合はCORR

要約データでの入力

- 要約データを用意する
 - sample3.datを利用する
 - 平均値が入っているので、MEANSも書く
 - 人数も別で指定する。NOBSERVATIONS=人数。

DATA:

```
FILE IS "sample3.dat";
```

```
TYPE = MEANS FULLCOV;
```

```
NOBSERVATIONS = 200;
```

VARIABLE:コマンドのUSEOBSERVATIONS

- 分析で使用するサブジェクトを指定する
 - 例えば、conが0のサブジェクトだけ使う場合

```
USEOBSERVATIONS = CON == 0;
```

- “==”は等しい、という意味。“=”は代入なので違いに注意。
- “<“や”>”、“<=”も使える。ANDやORなども使用可。

入門編 おさらい

- Mplusのコードの基本
 - 回帰はON
 - 相関はWITH
 - 因子分析はBY
 - 固定パラメータは@
 - 自由パラメータは*
 - パスの等値制約は()
 - 変数の作成は, DEFINE:で

知っておくと便利な知識

推定法について

- 我々がよく使う推定法
 - 最小二乗法(Un-weighted Least Square Method)
 - 分散分析、回帰分析などはすべてこの方法
- 構造方程式モデルの推定法
 - 最尤法(Maximum Likelihood Method)
 - 手元のデータが最も得られやすいようなモデルを推定
 - ESTIMATOR = ML;あるいはMLR;
 - 一般化最小二乗法(Generalized Least Square Method)
 - 残差の共分散行列の逆行列で重みづける最小二乗法
 - ESTIMATOR = GLS;
 - 重みつき最小二乗法(Weighted Least Square Method)
 - 残差を共分散行列の(共)分散で重みづける最小二乗法
 - ESTIMATOR = WLS; あるいはWLSMV;
 - MCMCによるベイズ推定法(Markov Chain Monte Carlo Method)
 - ベイズの定理に基づいた、マルコフ連鎖モンテカルロ法を使った推定法
 - ESTIMATOR = BAYES;

標準誤差の推定法

- 普通の推定方法
 - MLやGLS
 - モデルの仮定に基づいて標準誤差を推定
 - 特に最尤法は, 多変量正規性の仮定
- ロバスト標準誤差の推定
 - MLRやWLSMV
 - データの分布に合わせて標準誤差を補正
 - 多少多変量正規性から逸脱しても、頑健な結果
 - 推定値そのものは同じ
 - Mplusのデフォルトはこちらを使う。
 - ただ、ロバスト標準誤差の推定では、 χ^2 乗検定がやや複雑になる。そのため、今回はMLやWLSで書いている。

欠損値の推定法

- 従来の方法

- リストワイズ削除による分析

- 分散分析や回帰分析、因子分析はすべてこれを使う
 - 推定結果は多くの場合、バイアスを受ける

- 構造方程式モデルの欠損値推定

- 完全情報最尤法(Full information ML)

- 特に指定はいらない。基本はこれを使ってくれる。
 - サブジェクト全体を消すのではなく、欠損していない部分をすべて活用して推定
 - EMアルゴリズムで推定する

欠損値推定における注意点

- 完全情報最尤法の適用範囲
 - 欠損パターンがMARの場合に有効
 - 分散を推定している変数のみに適用
 - 内生変数はすべて分散が推定される
 - 外生変数は共分散を指定するか、分散を推定するように指定する必要がある
 - 分散が推定されている変数すべてが欠損の場合は、その欠損は推定できない
- 多重代入法も有効
 - DATA IMPUTATIONコマンドを使う
- 欠損値推定はしないよりも、したほうがいい！