LightStone

Stata で簡単に試せる SEM

第3回　2ファクタモデルの推定

SEM の第三回目です. 今回の目的はこれまでの知識を利用して 2 ファクタモデルを推定することです. そして, 第 2 回の最後に少し触れた"識別"について, もう少し詳しく解説します.

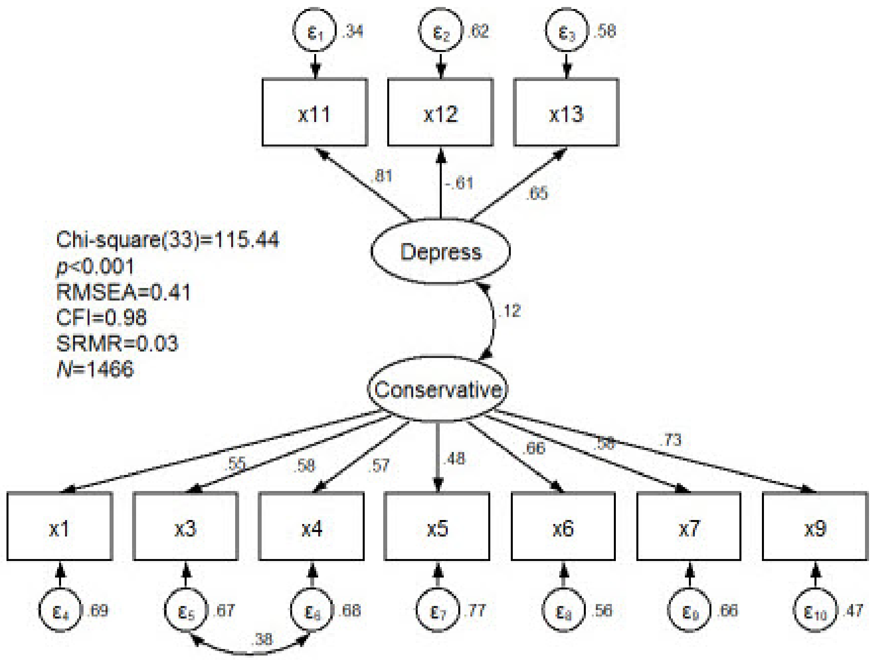
利用する書籍は Alan C. Acock, 2013. *Discovering Structural Equation Modeling Using Stata, Revised Edition*, Stata Press の第一章の後半です.

第3章　2ファクタモデル

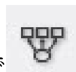
ここでは潜在変数である Conservative と Depress の統計的な関係を考察する.

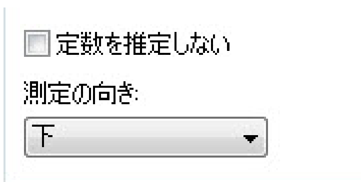
3.1　パス図の作成

最初にここで作成する 2 ファクタモデルを以下に示す.

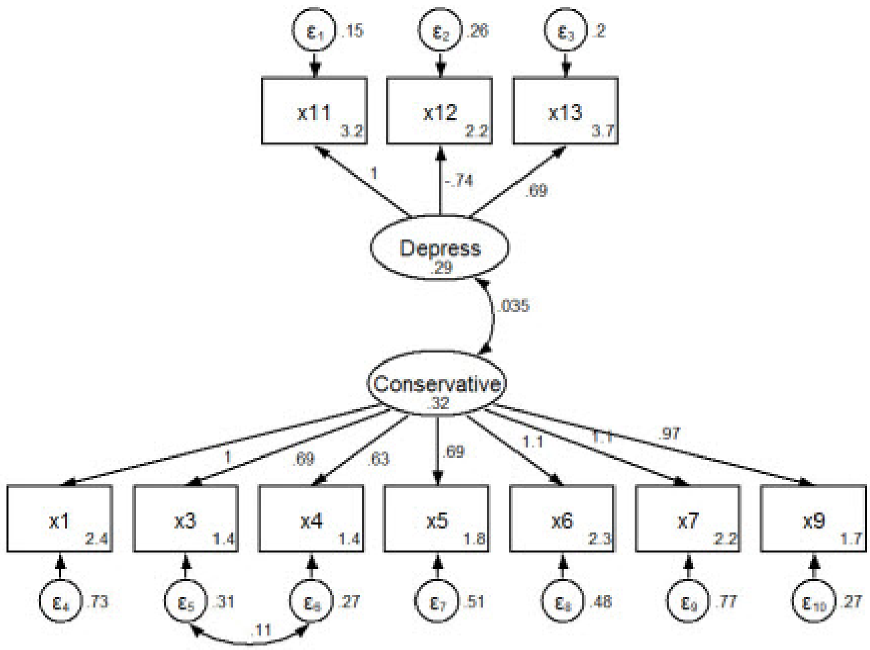


- SEM ビルダーを使って次のような手順に従って作成する.

1. 今回のように計測可能な変数 x が多い場合は SEM ビルダの画面で  のアイコンをクリックする. そして上部の Depress を作図する. 測定変数の項目には x11 x12 x13, または, x11-x13 と入力する.
2. 次にもう一度同じアイコンをクリックし, Conservative の部分を作成する. この時, ダイアログの下部にある次に示す項目で計測の向きを調整する. 同じく測定変数の項目に x1 x3-x7 x9 と入力する.

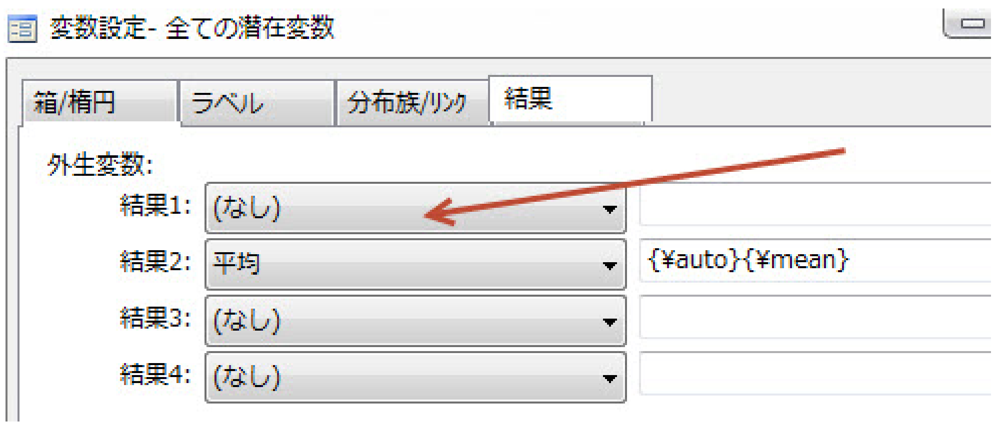


3. 共分散を設定する箇所に  のアイコンを利用して両矢印の曲線を引く.
4. 作図が完了したら, SEM ビルダーで推定/推定と操作する.



この図をもう少しスッキリさせることを考える.

- 潜在変数 (楕円)Depress と Conservative の分散を非表示にする
  SEM ビルダで設定/変数/すべての潜在変数と操作する. 次に示す結果のタブで, 分散の項目を「なし」に変更する.



- 観測可能な変数 (矩形) の定数項を非表示にする
  設定/変数/線形の内生観測変数と操作する. 結果のタブで切片とあるところを「なし」に変更する.

- 標準化回帰係数を表示する

  表示/標準化回帰係数を表示と操作する. 推定値が更新されたら, Stata のコマンドウィンドウを表示し, 適合度検定のコマンドを実行する.

```
. estat gof,stats (all)
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(33)	115.438	model vs. saturated
p > chi2	0.000	
chi2_bs(45)	3630.536	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.041	Root mean squared error of approximation
90% CI, lower bound	0.033	
upper bound	0.050	
pclose	0.958	Probability RMSEA <= 0.05
Information criteria		
AIC	30276.919	Akaike´s information criterion
BIC	30446.209	Bayesian information criterion
Baseline comparison		
CFI	0.977	Comparative fit index
TLI	0.969	Tucker-Lewis index
Size of residuals		
SRMR	0.033	Standardized root mean squared residual
CD	0.952	Coefficient of determination

- 仕上げとしてパス図に適合度検定の一部の情報を追加する.
- テキストアイコン  をクリックしてモデルの適合度に関する情報を次のように入力する.

```
Chi-square(33)=115.44
{it:p}<0.01
RMSEA=0.41
CFI=0.98
SRMA=0.03
{it:N}=1466
```

- {it:p} は p をイタリックで表示する書式である.

分析チェック

SEM による分析の結果, 2 つの潜在変数 Depress と Conservative の間の相関は弱いことが分かった.

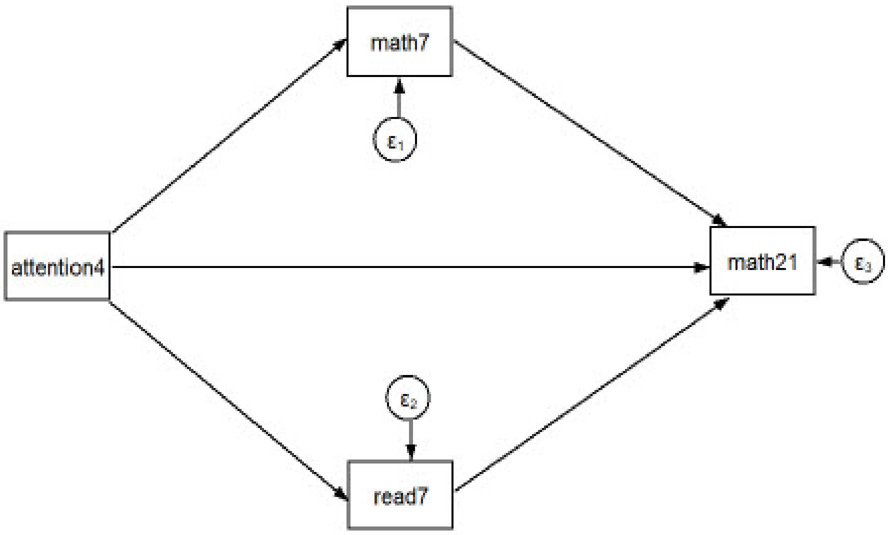
3.2 パスモデル

ここでは SEM を利用してパスモデルを推定する. [1]

- このセクションで紹介するパスモデルの特徴は観測可能な変数だけを利用する事.
- 一般的な連立方程式を考える訳であるが, ポイントは誤差項間の相関の設定にある.

幼児期の集中力と学力の関係

- McClelland et al. (2013) のデータ path.dta を利用して 4 才の時の集中力の持続性と学力の関係を仮説検定する.
- 中間変数として 7 才の時の読解力と計算力を利用する.
- データを開いたら, 次に示すパス図を SEM ビルダーで作成する.

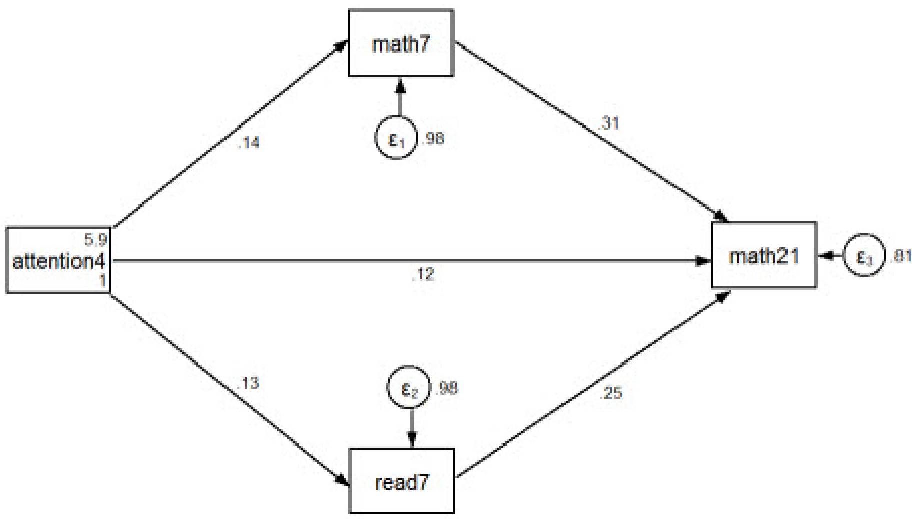


- SEM ビルダーのウィンドウにある推定メニューを利用してモデルを推定する.
- 欠損値はランダムで, 変数が正規分布に従うことを仮定して推定手法は `mlmv`(欠損値のある) を利用する.
- 推定ダイアログのレポートタブで, 「標準化係数と値を表示する」を選択する.

[1]Alan C. Acock, 2013. *Discovering Structural Equation Modeling Using Stata, Revised Edition*, Stata Press の第 2 章にある A substantive example of a path model の要約になります.

```
. sem (attention4 -> math7, ) (attention4 -> read7, ) (attention4 -> math21, ) (math7 -> math21, ) (read
> 7 -> math21, ), method(mlmv) standardized
Endogenous variables
Observed:  math7 read7 math21
Exogenous variables
Observed:  attention4
Fitting saturated model:
(省略)
Structural equation model                      Number of obs     =        430
Estimation method  = mlmv
Log likelihood     =  -4246.557
------------------------------------------------------------------------------
             |                 OIM
Standardized |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Structural   |
  math7 <-   |
  attention4 |    .141458   .0486307     2.91   0.004     .0461437    .2367723
       _cons |    3.04888   .3344304     9.12   0.000     2.393408    3.704351
  -----------+----------------------------------------------------------------
  read7 <-   |
  attention4 |   .1289838   .0491968     2.62   0.009     .0325598    .2254077
       _cons |   3.163475   .3383925     9.35   0.000     2.500238    3.826712
  -----------+----------------------------------------------------------------
  math21 <-  |
       math7 |   .3075685   .0481426     6.39   0.000     .2132108    .4019262
       read7 |   .2520422   .0489132     5.15   0.000      .156174    .3479104
  attention4 |   .1171187   .0467622     2.50   0.012     .0254664     .208771
       _cons |   1.380531    .361878     3.81   0.000     .6712636    2.089799
-------------+----------------------------------------------------------------
 var(e.math7)|   .9799896   .0137584                      .9533913     1.00733
 var(e.read7)|   .9833632   .0126912                      .9588009    1.008555
var(e.math21)|   .8075246   .0341705                      .7432537    .8773531
------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(1)   =     27.56, Prob > chi2 = 0.0000
```

パス図は次のようになる.



分析チェック

- math7 に対して attention4 は有意である (符号は正).
- read7 に対しても有意である (符号は正).
- math21 に対して math7 と read7 はどちらも有意であり, math7 の影響が大きい
- math21 に対する attention4 の効果は有意であるが, math7 と read7 よりも小さい.

内生変数の分散に対する考察

内生変数の分散に対するモデルの説明力を考察する.

```
. estat eqgof
```

Equation-level goodness of fit

depvars	fitted	Variance predicted	residual	R-squared	mc	mc2
observed						
math7	7.621122	.1525014	7.46862	.0200104	.141458	.0200104
read7	64.70388	1.076467	63.62742	.0166368	.1289838	.0166368
math21	6.920939	1.33211	5.588828	.1924754	.4387202	.1924754
overall				.0515245		

```
mc  = correlation between depvar and its prediction
mc2 = mc^2 is the Bentler-Raykov squared multiple correlation coefficient
```

- 観測した変数 math7 と read7 について, モデルで説明可能な分散の割合は 2% 程度である.
- 一方, math21 に対する説明力は 19% 程度ある.

$$R^2 = \frac{\text{内生変数の分散の予測値}}{\text{内生変数の分散}}$$

- mc2 (Bentler-Raykov R^2) は非再帰形モデルの場合に参照する

モデルの適合度に対する考察

- 適合度の考察を行う.

```
.  estat gof,stats(all)
```

```
----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Likelihood ratio     |
          chi2_ms(1) |     27.561   model vs. saturated
            p > chi2 |      0.000
          chi2_bs(6) |    130.877   baseline vs. saturated
            p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
               RMSEA |      0.249   Root mean squared error of approximation
 90% CI, lower bound |      0.174
         upper bound |      0.332
              pclose |      0.000   Probability RMSEA <= 0.05
---------------------+------------------------------------------------------
Information criteria |
                 AIC |   8515.114   Akaike's information criterion
                 BIC |   8559.816   Bayesian information criterion
---------------------+------------------------------------------------------
Baseline comparison  |
                 CFI |      0.787   Comparative fit index
                 TLI |     -0.276   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
                  CD |      0.052   Coefficient of determination
----------------------------------------------------------------------------
Note: SRMR is not reported because of missing values.
```

- 尤度比検定の項目で $\chi^2(1) = 27.56, p < 0.001$ とありますから, 共分散構造をモデルで再現できていないことが分かる.
- 誤差に関する項目から RMSEA が約 0.25 で, 0.05 以下という規準を大きく上回っており, 誤差が大きすぎることが分かる.
- CFI の目安である 0.9 をクリアしていない

変数間の相関

変数間に相関を設定することによってモデルがどの程度, 改良できるか次のコマンドによって調べる.

```
. estat mindices
```

```
Modification indices
```

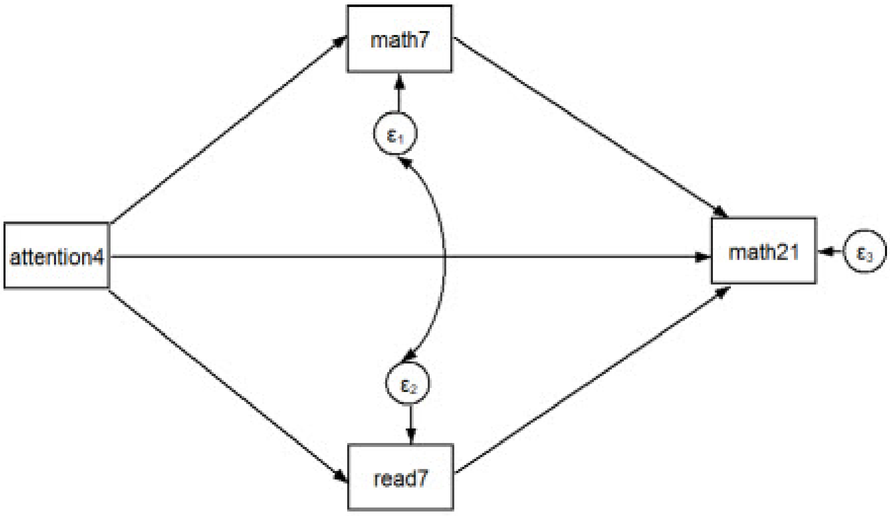
	MI	df	P>MI	EPC	Standard EPC
Structural					
math7					
read7	26.885	1	0.00	.0899778	.2621748
math21	26.885	1	0.00	1.091552	1.040202
read7					
math7	26.885	1	0.00	.7665476	.2630773
math21	26.885	1	0.00	2.615316	.8553455
cov(e.math7,e.read7)	26.885	1	0.00	5.725053	.2626257

```
EPC = expected parameter change
```

- カイ二乗検定統計量はどこに相関を設定しても同じ値だけ減少し, モデルを改良することが分かる
- 時間の流れを考えて, math21 から read7 や math7 に与える影響は存在しないので, 実質的な意味はない.
- 計算に際し, 問題文を読む力が必要とされるケースでは read7 から math7 への影響は考えられる.
- しかし, その逆は考えずらい.
- 計算力や読解力に子供の家庭の社会経済的地位が影響したり, 性差が影響すると考えることができるなら誤差項の間に相関を考えることは合理的である.

3.3 誤差項の相関

- math7 と read7 の誤差項間に相関を設定する.
- 先のモデルの推定結果から自由度が 1 であることが分かる.
- ここで新たに共分散を設定すると perfect fit になり, 自由度がゼロになるが, 設定に合理性があれば, 自由度ゼロを気にする必要はない.



このパス図を使って, 標準化係数を mlmv によって推定した結果を次に示す.

```
 Structural equation model                        Number of obs     =        430

Estimation method  = mlmv
Log likelihood     = -4232.7763
```

Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
math7 <-						
attention4	.1424678	.0485568	2.93	0.003	.0472983	.2376373
_cons	3.043008	.3341487	9.11	0.000	2.388089	3.697928
read7 <-						
attention4	.1296611	.0491074	2.64	0.008	.0334123	.22591
_cons	3.162223	.3378934	9.36	0.000	2.499964	3.824482
math21 <-						
math7	.3008525	.0467369	6.44	0.000	.20925	.3924551
read7	.2462258	.0473568	5.20	0.000	.1534081	.3390434
attention4	.1147871	.0460627	2.49	0.013	.024506	.2050683
_cons	1.365485	.3571808	3.82	0.000	.6654235	2.065547
var(e.math7)	.9797029	.0138355			.9529576	1.007199
var(e.read7)	.983188	.0127347			.9585427	1.008467
var(e.math21)	.7779759	.0384575			.7061369	.8571234
cov(e.math7,e.read7)	.2599788	.0469642	5.54	0.000	.1679306	.352027

```
LR test of model vs. saturated: chi2(0)    =      0.00, Prob > chi2 =      .
```

分散の説明力を確認します.

```
. estat eqgof
```

```
 Equation-level goodness of fit
```

depvars	fitted	Variance predicted	residual	R-squared	mc	mc2
observed						
math7	7.619575	.154655	7.46492	.0202971	.1424678	.0202971
read7	64.62929	1.086548	63.54274	.016812	.1296611	.016812
math21	7.178428	1.593784	5.584644	.2220241	.4711943	.2220241
overall				.0448891		

```
mc  = correlation between depvar and its prediction
mc2 = mc^2 is the Bentler-Raykov squared multiple correlation coefficient
```

- math7 と read7 の分散の説明力はほぼ変化しない
- math21 に関しては 0.19 から 0.22 に向上している
- 今, 自由度はゼロなので estat gof,stats(all) や estats mindices コマンドに意味はない

3.4 各種効果の推定

パス図を見ると math21 に対して 3 つの矢印が引かれている

- attention4 からの効果を「直接効果」と呼ぶ
- math7 および read7 を経由している効果は「間接効果」と呼ぶ
- 2 つの効果は次のコマンドを用いて一覧表示できる

```
. estat teffects,standardize
```

Direct effects

	Coef.	OIM Std. Err.	z	P>|z|	Std. Coef.
Structural					
math7 <-					
attention4	.1290411	.0446933	2.89	0.004	.1424678
read7 <-					
attention4	.342035	.1313072	2.60	0.009	.1296611
math21 <-					
math7	.2920135	.0470959	6.20	0.000	.3008525
read7	.0820604	.0161567	5.08	0.000	.2462258
attention4	.1009146	.0408745	2.47	0.014	.1147871

Indirect effects

	Coef.	OIM Std. Err.	z	P>|z|	Std. Coef.
Structural					
math7 <-					
attention4	0	(no path)			0
read7 <-					
attention4	0	(no path)			0
math21 <-					
math7	0	(no path)			0
read7	0	(no path)			0
attention4	.0657493	.0202659	3.24	0.001	.0747877

Total effects

	Coef.	OIM Std. Err.	z	P>|z|	Std. Coef.
Structural					
math7 <-					
attention4	.1290411	.0446933	2.89	0.004	.1424678
read7 <-					
attention4	.342035	.1313072	2.60	0.009	.1296611
math21 <-					
math7	.2920135	.0470959	6.20	0.000	.3008525
read7	.0820604	.0161567	5.08	0.000	.2462258
attention4	.1666639	.0440972	3.78	0.000	.1895748

- 大切な情報は右端の Std. Coef. にある標準化係数である
- 1 番上のテーブルは直接効果, つまり, 間に変数が介在していない部分の効果を示している
- これは普通の標準化係数の出力と同じ
- 2 番目のテーブルは間接効果を示している
- つまり, attention4 から math7 と read7 を経由している効果で 0.074.

$$0.074 = 0.142 \times 0.300 + 0.129 \times 0.246$$

- 最後のテーブルは直接効果と間接効果を合計したもの
- 標準化した右端の列をみると分かり易い

それぞれの効果をまとめてみると, 次のようになる.

アウトカム	直接	間接	合計
Math7			
attention4→math7	0.14**	-	0.14**
Read7			
attention4→read7	0.13**	-	0.13**
Math21			
attention4→math21	0.11*	0.07**	0.19***
math7→math21	0.30***	-	0.30***
read7→math21	0.25***	-	0.25***

- 有意水準は非標準化の情報を示している
- $^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

簡単なまとめ

- 潜在変数を利用した 2 ファクタモデルを推定し, 観測できない潜在変数間の関係をモデル化した
- 全て観測可能な変数だけを用いた連立方程式モデル (パスモデル) を式を記述する代わりにパス図を用いて構築し, 係数を推定した