

構造方程式モデリング (SEM)

- 構造方程式モデリングを利用することによって、線形回帰モデルから同時方程式まで幅広いモデルの推定が可能になります。
- 確証的因子分析 (CFA)、誤差項の相関したモデル、潜在成長モデル、MIMIC モデル、IRT モデルは、専用のコマンドだけでなく **sem** コマンドを使って推定することもできます。
- SEM とは特別なモデルに限定された推定手法ではありません。実際、Stata の **regress** や **probit**、**stcox** などのコマンドによる推定と同じことが SEM で実行可能です。SEM は計量モデルに対する一つの考え方であると理解してください。

この例題集でできること

- 構造方程式モデリングを行い、コマンド操作でモデルの係数を推定します。
- SEM ビルダー (グラフィック操作) でパス図を作成します。パス図から推定を行うこともできます。
- 作成したモデルに対して、適合度検定を行います。数種類の検定が実行されるので、用途に合った検定を使います。

例 1 Single-factor measurement モデル

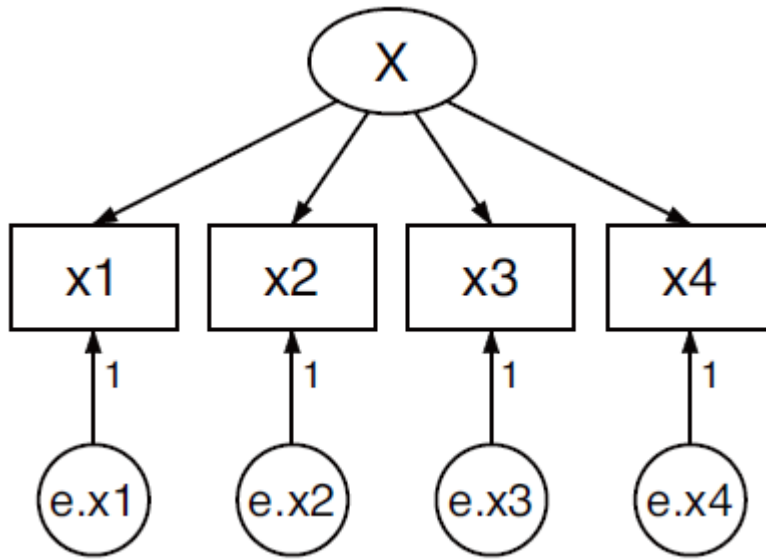
- サンプルデータをダウンロードし、データを確認します。

```
. use https://www.stata-press.com/data/r16/sem_1fmm
. summarize
. notes
```

Variable	Obs	Mean	Std. Dev.	Min	Max
x1	500	99.518	14.35402	60	137
x2	500	99.954	14.1939	52	140
x3	500	99.052	14.26395	59	150
x4	500	94.474	70.11603	-113	295

```
_dta:
1. fictional data
2. Variables x1, x2, x3, and x4 each contain a test score designed to measure X.
   The test is scored to have mean 100.
```

- 下図のモデルにフィットさせます。



```
. sem(x1 x2 x3 x4 <- X)
```

Endogenous variables

Measurement: x1 x2 x3 x4

Exogenous variables

Latent: X

Fitting target model:

Iteration 0: log likelihood = -8487.5905
 Iteration 1: log likelihood = -8487.2358
 Iteration 2: log likelihood = -8487.2337
 Iteration 3: log likelihood = -8487.2337

Structural equation model Number of obs = 500
 Estimation method = ml
 Log likelihood = -8487.2337

(1) [x1]X = 1

		OIM				
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Measurement	x1					
	X	1 (constrained)				
	_cons	99.518	.6412888	155.18	0.000	98.2611 100.7749
x2	X	1.033249	.0723898	14.27	0.000	.8913676 1.17513
	_cons	99.954	.6341354	157.62	0.000	98.71112 101.1969
x3	X	1.063876	.0729725	14.58	0.000	.9208526 1.2069
	_cons	99.052	.6372649	155.43	0.000	97.80298 100.301
x4	X	7.276754	.4277638	17.01	0.000	6.438353 8.115156
	_cons	94.474	3.132547	30.16	0.000	88.33432 100.6137
	var(e.x1)	115.6865	7.790423			101.3823 132.0089
	var(e.x2)	105.0445	7.38755			91.51873 120.5692
	var(e.x3)	101.2572	7.17635			88.12499 116.3463
	var(e.x4)	144.0406	145.2887			19.94838 1040.069
	var(X)	89.93921	11.07933			70.64676 114.5001

LR test of model vs. saturated: chi2(2) = 1.46, Prob > chi2 = 0.4827

- このモデルの方程式は、下記の通りです。

$$x_1 = \alpha_1 + X\beta_1 + e.x_1$$

$$x_2 = \alpha_2 + X\beta_2 + e.x_2$$

$$x_3 = \alpha_3 + X\beta_3 + e.x_3$$

$$x_4 = \alpha_4 + X\beta_4 + e.x_4$$

- 変数 X は、外生的な潜在変数であるため、標準化の制約が必要です。変数 X は最初に観測された変数 x1 に固定されるため、パス係数に制限がかかります。
- X->x1, X->2, X->3 のパス係数は、それぞれ 1 (制約付き) , 1.30, 1.06 と推定されます。一方、X->4 のパス係数は 7.28 ですが、異常値ではありません。このデータは Stata 社で作成したものです。モデルの真の係数は、1, 1, 1, 7 です。作成したモデルと、観測可能な変数の共分散も含むモデルの検定結果が表の下部に表示されます。 $\chi^2(2)$ 統計量は 1.14、有意水準は 0.4827 なので、「モデル化していない共分散は不要である」という帰無仮説は、棄却できません。
- これはフィットの悪いモデルにおける適合度検定です。したがって、モデルのフィットは良くないという結果が出力されます。より数学的に正確な帰無仮説は「近似共分散行列と観測された変数の平均ベクトルが、母集団で観測された行列とベクトルに等しい」です。

同じモデルを gsem にフィットさせる

- sem と gsem は、standard linear SEM では同じ結果になります。

```
. gsem(x1 x2 x3 x4 <- X)
```

```

Fitting fixed-effects model:

Iteration 0:  log likelihood = -8948.2394
Iteration 1:  log likelihood = -8948.2394

Refining starting values:

Grid node 0:  log likelihood = -8487.5916

Fitting full model:

Iteration 0:  log likelihood = -8487.5916
Iteration 1:  log likelihood = -8487.5051
Iteration 2:  log likelihood = -8487.3836
Iteration 3:  log likelihood = -8487.2697
Iteration 4:  log likelihood = -8487.2337
Iteration 5:  log likelihood = -8487.2337

Generalized structural equation model          Number of obs   =       500

Response      :  x1
Family        :  Gaussian
Link          :  identity

Response      :  x2
Family        :  Gaussian
Link          :  identity

Response      :  x3
Family        :  Gaussian
Link          :  identity

Response      :  x4
Family        :  Gaussian
Link          :  identity

Log likelihood = -8487.2337

( 1) [x1]X = 1

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1						
X	1 (constrained)					
_cons	99.518	.6412888	155.18	0.000	98.2611	100.7749
x2						
X	1.033249	.0723898	14.27	0.000	.8913676	1.17513
_cons	99.954	.6341354	157.62	0.000	98.71112	101.1969
x3						
X	1.063876	.0729725	14.58	0.000	.9208526	1.206899
_cons	99.052	.6372649	155.43	0.000	97.80298	100.301
x4						
X	7.276753	.4277636	17.01	0.000	6.438352	8.115154
_cons	94.474	3.132547	30.16	0.000	88.33432	100.6137
var(X)	89.93923	11.07933			70.64678	114.5001
var(e.x1)	115.6865	7.790422			101.3822	132.0089
var(e.x2)	105.0444	7.387549			91.51872	120.5692
var(e.x3)	101.2572	7.176349			88.12498	116.3463
var(e.x4)	144.0408	145.2886			19.94848	1040.067

- 結果は **sem** コマンドの時とほぼ同じです。係数、分散、標準誤差は最後の桁が異なります。**gsem** はより膨大な数値処理を行う別のアプローチを行っているため、精度が低下します。
- 対数尤度値は同じです。これは珍しいことで、一般的に **gsem** の対数尤度は **sem** のメトリックスは異なります。ただし、モデルに観測される変数が含まれない場合は、**gsem** と **sem** で同じメトリックスが使用されます。
- **gsem** と **sem** の両方で同じモデルをフィットさせると、**sem** の方が若干高速かつ正確で、モデル構築後により多くの推定機能を使用できます。

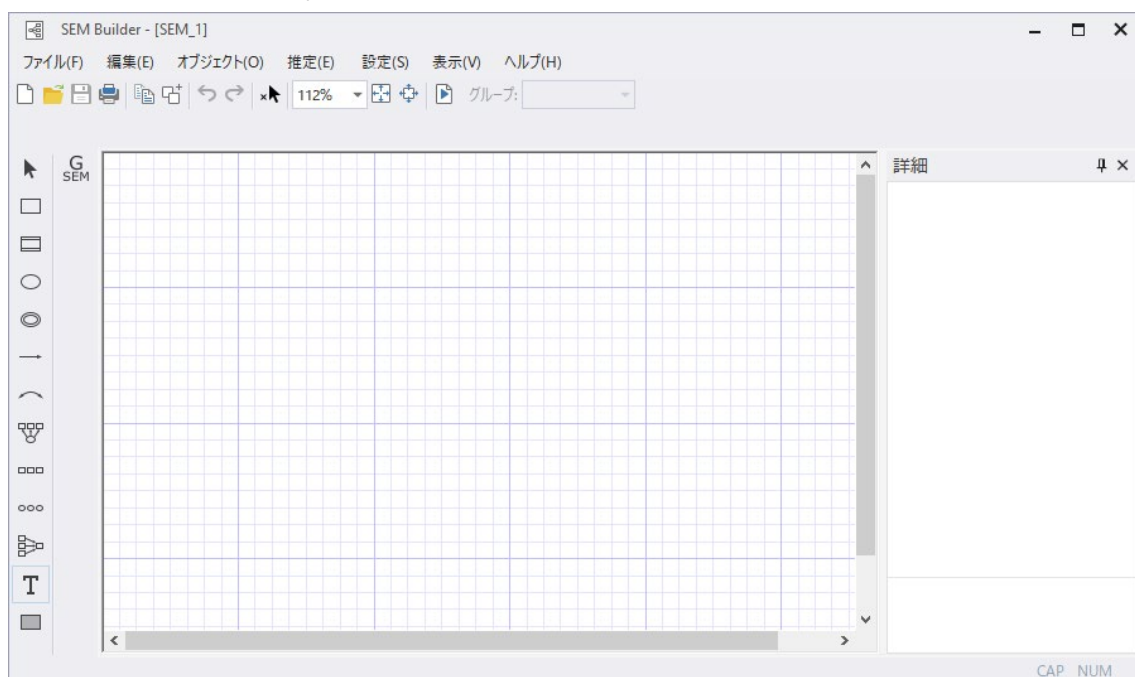
Builder を使ってフィットさせる

- コマンドではなく、Builder を使ってモデルをフィットさせる方法もあります。

1. コマンドウィンドウで下記を入力し、サンプルデータをダウンロードします。

```
.use https://www.stata-press.com/data/r16/sem_1fmm
```

2. メニューの統計 > SEM (構造方程式モデリング) > モデル構築 / 推定から、新規の SEM Builder を開きます。



3. X のコンポートを作成します。

 計測要素を追加アイコンを選択し、SEM Builder の描画領域をクリックします。

ダイアログボックスを下記のように設定します。

- グループ変数名を「X」にします。
- 測定変数に x1, x2, x3, x4 を指定します。
- 測定の向きを下にします。
- OK をクリックします。

測定成分

メイン 距離 接続線

潜在変数

観測レベルの潜在変数(標準)

マルチレベル潜在変数

グループ変数名:

X

測定変数

変数を選択する

変数の数を指定する

測定変数:

x1 x2 x3 x4

測定変数を一般化する

分布族/リンク: Gaussian, Identity

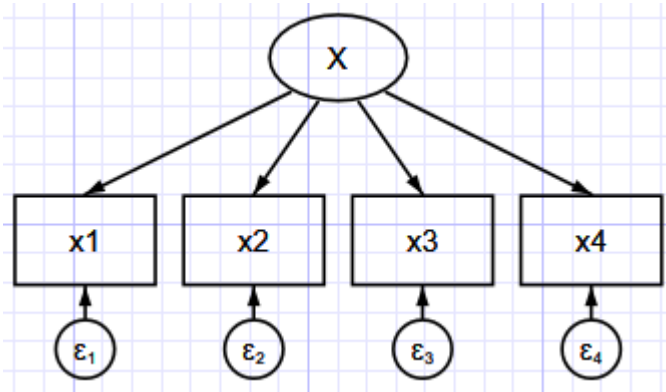
定数を推定しない

測定の向き:

下


OK キャンセル

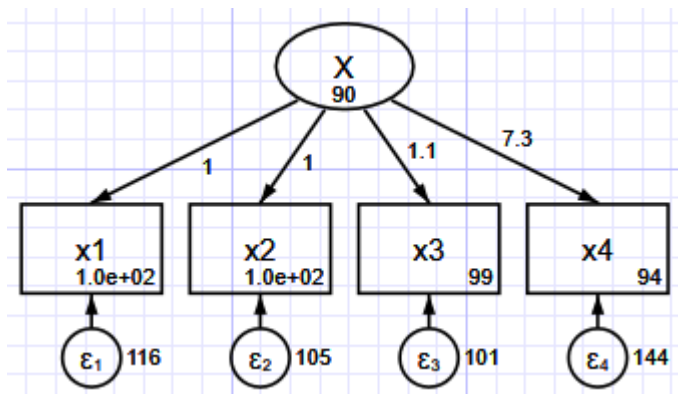
下記の図が作成されます。誤差項からの各 X へのパスには1の制限がかかっているため、図には追加する必要がありません。



☞ 選択をクリックし、変数などを選択してコンポーネントの位置を調整できます。

4. 推定を行います。

ツールバーの  推定アイコンをクリックして、SEM 推定オプションウィンドウで OK をクリックします。



下記のコマンドで Builder に構築されたダイアグラムを呼び出すことができます。

```
.webgetsem sem_1fmm
```

例 2 published covariances からデータセットを作成する

- Williams, Eaves, Cox(2002)では、データから得られた共変量が公開されています。次の例で、SEM をフィットさせるためにこの共変量を使用します。そのために、例 2 ではまず summary statistics dataset (SSD) の作り方を説明します。
- Williams, Eaves, Cox(2002)が発表した共変量マトリックスは、下記の表の通りです。

	Affective		Miniscale		Cognitive		Miniscale	
	1	2	...	5	1	2	...	
Affective								
1	2038.035	1631.766	...	1721.830	659.795	779.519	...	
2		1932.163	...	1688.292	702.969	790.448	...	
.								
.								
5				2061.875	775.118	871.211	...	
Cognitive								
1					630.518	500.128	...	
.								
.								

SSD の作成

- do ファイルで下記のコマンドを実行し、SSD を作成します。

```
.clear all
. ssd init a1 a2 a3 a4 a5 c1 c2 c3 c4 c5
. ssd set obs 216
. #delimit ;
. ssd set cov 2038.035 /
> 1631.766 1932.163 /
> 1393.567 1336.871 1313.809 /
> 1657.299 1647.164 1273.261 2034.216 /
> 1721.830 1688.292 1498.401 1677.767 2061.875 /
> 659.795 702.969 585.019 656.527 775.118 630.518 /
> 779.519 790.448 653.734 764.755 871.211 500.128 741.767 /
> 899.912 879.179 750.037 897.441 1008.287 648.935 679.970
> 1087.409 /
> 775.235 739.157 659.867 751.860 895.616 550.476 603.950
> 677.865 855.272 /
> 821.170 785.419 669.951 802.825 877.681 491.042 574.775
> 686.391 622.830 728.674 ;
.#delimit cr
```

- #delimit コマンドは、do ファイルでのみ使用できます。#delimit から #delimit cr

までの間で改行が無視され、;が入力された時のみ改行します。1行にすべてのコマンドを入力する場合には、使用する必要はありません。

データの保存

- 下記のコマンドでデータを保存し、解析を中断することができます。

```
. save sem2fmm
```

- 下記のコマンドで、他に設定しなければならない項目があるかどうかを調べることができます。

```
. ssd status
```

```
Status:
           observations:  set
           means:        unset
           variances or sd:  set
           covariances or correlations:  set
```

- 平均が設定されていません（上記の論文には平均の記載がありませんでした）。

SSD のラベリング

- 下記のコマンドでデータを確認します。

```
. ssd describe
```

```
Summary statistics data from sem2fmm.dta
  obs:          216
  vars:         10          5 Nov 2020 09:48
```

variable name	variable label
a1	
a2	
a3	
a4	
a5	
c1	
c2	
c3	
c4	
c5	

- 下記のコマンドで、データセットにラベルとノートを書き加えます。

```
. label data "Affective and cognitive arousal"  
. label var a1 "affective arousal 1"  
. label var a2 "affective arousal 2"  
. label var a3 "affective arousal 3"  
. label var a4 "affective arousal 4"  
. label var a5 "affective arousal 5"  
. label var c1 "cognitive arousal 1"  
. label var c2 "cognitive arousal 2"  
. label var c3 "cognitive arousal 3"  
. label var c4 "cognitive arousal 4"  
. label var c5 "cognitive arousal 5"  
. #delimit ;  
. notes: Summary statistics data containing published covariances  
> from Thomas O. Williams, Ronald C. Eaves, and Cynthia Cox,  
> 2 Apr 2002, "Confirmatory factor analysis of an instrument  
> designed to measure affective and cognitive arousal",  
> _Educational and Psychological Measurement_,  
> vol. 62 no. 2, 264-283. ;  
. notes: a1-a5 report scores from 5 miniscales designed to measure  
> affective arousal. ;  
. notes: c1-c5 report scores from 5 miniscales designed to measure  
> cognitive arousal. ;  
. notes: The series of tests, known as the VST II  
> (Visual Similes Test II) were administered to 216 children  
> ages 10 to 12. The miniscales are sums of scores of  
> 5 to 6 items in VST II. ;  
. #delimit cr  
. ssd describe  
. save sem2fmm, replace
```

- SSD は下記のように設定されます。

```
Summary statistics data from sem2fmm.dta
obs:          216          Affective and cognitive arousal
vars:         10          5 Nov 2020 09:48
                   (_dta has notes)
```

variable name	variable label
a1	affective arousal 1
a2	affective arousal 2
a3	affective arousal 3
a4	affective arousal 4
a5	affective arousal 5
c1	cognitive arousal 1
c2	cognitive arousal 2
c3	cognitive arousal 3
c4	cognitive arousal 4
c5	cognitive arousal 5

- 以上のように、変数やデータにラベルを付けることができます。データセットにノートを付けることもできます。
- SSD はデータセットと同様に、保存していつでも使用することができます。

SSD のリスト

- 下記のコマンドで、SSD のリストを作成することができます。

```
. ssd list
```

```
Observations = 216

Means undefined; assumed to be 0

Variances implicitly defined; they are the diagonal of the covariance matrix.

Covariances:
      a1      a2      a3      a4      a5      c1      c2      c3
2038.035
1631.766 1932.163
1393.567 1336.871 1313.809
1657.299 1647.164 1273.261 2034.216
1721.83 1688.292 1498.401 1677.767 2061.875
659.795 702.969 585.019 656.527 775.118 630.518
779.519 790.448 653.734 764.755 871.211 500.128 741.767
899.912 879.179 750.037 897.441 1008.287 648.935 679.97 1087.409
775.235 739.157 659.867 751.86 895.616 550.476 603.95 677.865
821.17 785.419 669.951 802.825 877.681 491.042 574.775 686.391

      c4      c5
855.272
622.83 728.674
```

例3 2因子の measurement model

- 複数因子の measurement model について、例2で作成した SSD を使用して解説します。
例2の SSD のデータは、下記から入手することもできます。

```
. use https://www.stata-press.com/data/r16/sem_2fmm
```

- データを入手したら、内容を確認します。

```
. ssd describe
.notes
```

Summary statistics data from sem2fmm.dta

```
obs:          216          Affective and cognitive arousal
vars:          10          5 Nov 2020 10:12
                          (_dta has notes)
```

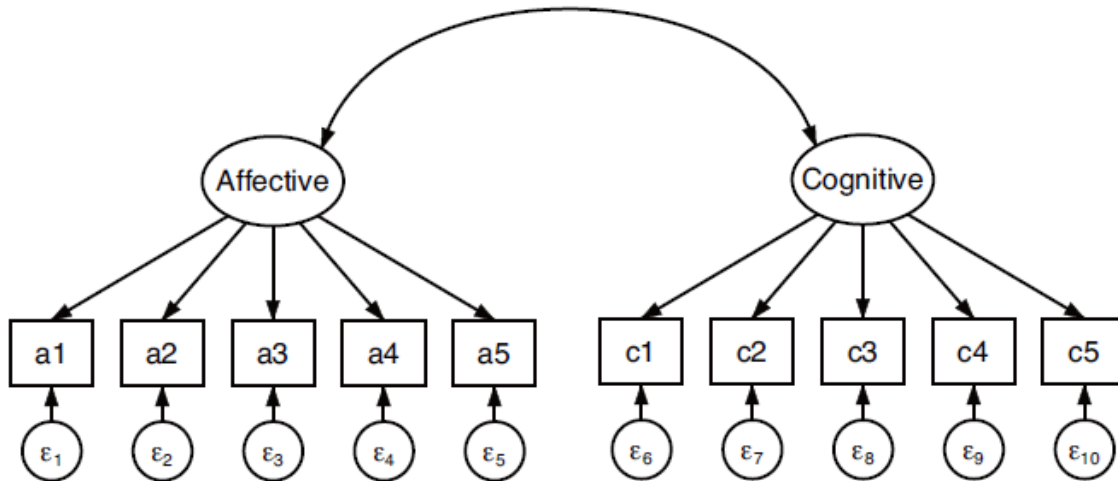
variable name	variable label
a1	affective arousal 1
a2	affective arousal 2
a3	affective arousal 3
a4	affective arousal 4
a5	affective arousal 5
c1	cognitive arousal 1
c2	cognitive arousal 2
c3	cognitive arousal 3
c4	cognitive arousal 4
c5	cognitive arousal 5

_dta:

1. Summary statistics data containing published covariances from Thomas OWilliams, Ronald CEaves, and Cynthia Cox, 2 Apr 2002, "Confirmatory factor analysis of an instrument designed to measure affective and cognitive arousal", *_Educational and Psychological Measurement_*, vol62 no2, 264-283
2. a1-a5 report scores from 5 miniscales designed to measure affective arousal
3. c1-c5 report scores from 5 miniscales designed to measure cognitive arousal
4. The series of tests, known as the VST II (Visual Similes Test II) were administered to 216 children ages 10 to 12The miniscales are

複数因子の measurement model をフィットさせる

- 下図の Kline (2005, 70-74, 184) のモデルにフィットさせます。



```
. sem(Affective -> a1 a2 a3 a4 a5)(Cognitive -> c1 c2 c3 c4 c5)
```

Endogenous variables

Measurement: a1 a2 a3 a4 a5 c1 c2 c3 c4 c5

Exogenous variables

Latent: Affective Cognitive

Fitting target model:

```
Iteration 0: log likelihood = -9542.8803
Iteration 1: log likelihood = -9539.5505
Iteration 2: log likelihood = -9539.3856
Iteration 3: log likelihood = -9539.3851
```

Structural equation model

Number of obs = 216

Estimation method = ml

Log likelihood = -9539.3851

(1) [a1]Affective = 1

(2) [c1]Cognitive = 1

		OIM						
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
Measurement								
a1	Affective	1 (constrained)						
a2	Affective	.9758098	.0460752	21.18	0.000	.885504	1.066116	
a3	Affective	.8372599	.0355086	23.58	0.000	.7676643	.9068556	
a4	Affective	.9640461	.0499203	19.31	0.000	.866204	1.061888	
a5	Affective	1.063701	.0435751	24.41	0.000	.9782951	1.149107	
c1	Cognitive	1 (constrained)						
c2	Cognitive	1.114702	.0655687	17.00	0.000	.9861901	1.243215	
c3	Cognitive	1.329882	.0791968	16.79	0.000	1.174659	1.485105	
c4	Cognitive	1.172792	.0711692	16.48	0.000	1.033303	1.312281	
c5	Cognitive	1.126356	.0644475	17.48	0.000	1.000041	1.252671	
	var(e.a1)	384.1359	43.79119			307.2194	480.3095	
	var(e.a2)	357.3524	41.00499			285.3805	447.4755	
	var(e.a3)	154.9507	20.09026			120.1795	199.7822	
	var(e.a4)	496.4594	54.16323			400.8838	614.8214	
	var(e.a5)	191.6857	28.07212			143.8574	255.4154	
	var(e.c1)	171.6638	19.82327			136.894	215.2649	
	var(e.c2)	171.8055	20.53479			135.9247	217.1579	
	var(e.c3)	276.0144	32.33535			219.3879	347.2569	
	var(e.c4)	224.1994	25.93412			178.7197	281.2527	
	var(e.c5)	146.8655	18.5756			114.6198	188.1829	
	var(Affective)	1644.463	193.1032			1306.383	2070.034	
	var(Cognitive)	455.9349	59.11245			353.6255	587.8439	
cov(Affective,Cognitive)		702.0736	85.72272	8.19	0.000	534.0601	870.087	
LR test of model vs. saturated: chi2(34) = 88.88, Prob > chi2 = 0.0000								

- 例1では、**sem** を **raw** データで実行しました。この例3では、**SSD** で実行しています。そのため、特別なオプションを指定する必要はありません。
- 推定された係数は、**unstandardized** な係数または因子負荷量です。
- 表の下の **ver(e.)** で推定されている係数は、誤差分散です。
- 結果は(Kline 2005, 184)とは正確には一致しません。**sem** オプションに **nm1** を指定した場合、結果は3桁または4桁目で一致する可能性があります。**nm1** オプションを指定すると、分散と共分散を計算する際に、**N** ではなく **N-1** で割り算します。

標準化された結果を表示する

- パス係数ではなく、標準化されたパスの値を表示させると、結果が見やすくなります。標準化するには、**standardized** オプションを使用します。
- 標準化された値とは、標準偏差単位で表示されます。ある変数と、その変数に従って変化した変数の両方の変化を標準偏差単位で表しています。

```
. sem, standardized
```



```

Structural equation model                Number of obs   =       216
Estimation method   = ml
Log likelihood      = -9539.3851

( 1) [a1]Affective = 1
( 2) [c1]Cognitive = 1

```

	Standardized	Coef.	OIM Std. Err.	z	P> z	[95% Conf. Interval]	
Measurement							
a1	Affective	.9003553	.0143988	62.53	0.000	.8721342	.9285765
a2	Affective	.9023249	.0141867	63.60	0.000	.8745195	.9301304
a3	Affective	.9388883	.0097501	96.29	0.000	.9197784	.9579983
a4	Affective	.8687982	.0181922	47.76	0.000	.8331421	.9044543
a5	Affective	.9521559	.0083489	114.05	0.000	.9357923	.9685195
c1	Cognitive	.8523351	.0212439	40.12	0.000	.8106978	.8939725
c2	Cognitive	.8759601	.0184216	47.55	0.000	.8398544	.9120658
c3	Cognitive	.863129	.0199624	43.24	0.000	.8240033	.9022547
c4	Cognitive	.8582786	.0204477	41.97	0.000	.8182018	.8983554
c5	Cognitive	.8930346	.0166261	53.71	0.000	.8604479	.9256212
	var(e.a1)	.1893602	.0259281			.1447899	.2476506
	var(e.a2)	.1858097	.0256021			.1418353	.2434179
	var(e.a3)	.1184887	.0183086			.0875289	.1603993
	var(e.a4)	.2451896	.0316107			.1904417	.3156764
	var(e.a5)	.0933991	.015899			.0669031	.1303885
	var(e.c1)	.2735248	.0362139			.2110086	.354563
	var(e.c2)	.2326939	.0322732			.1773081	.3053806
	var(e.c3)	.2550083	.0344603			.1956717	.3323385
	var(e.c4)	.2633578	.0350997			.2028151	.3419733
	var(e.c5)	.2024893	.0296954			.1519049	.2699183
	var(Affective)	1	.			.	.
	var(Cognitive)	1	.			.	.
	cov(Affective,Cognitive)	.8108102	.0268853	30.16	0.000	.758116	.8635045

```

LR test of model vs. saturated: chi2(34) = 88.88, Prob > chi2 = 0.0000

```

- 標準化された係数に加えて、誤差分散が推定されます。
- `sem` コマンドを使うと、潜在変数の分散も表示されます。最初の出力では、潜在変数の Affective は 1644.46 で標準誤差 193 でした。標準化した出力では、正規化されているため、分散が 1 で標準誤差はありません。
- 正規化された共分散は相関係数となるため、Affective と Cognitive の相関は 0.81 となります。
- このモデルの標準化された係数は、各 indicator がひとつの要素を測定しているため、indicator と潜在変数の相関係数を表します。例えば、標準化されたパス係数 $a1 \leftarrow$ Affective は 0.90 なので、 $a1$ と Affective の相関は 0.90 となります。

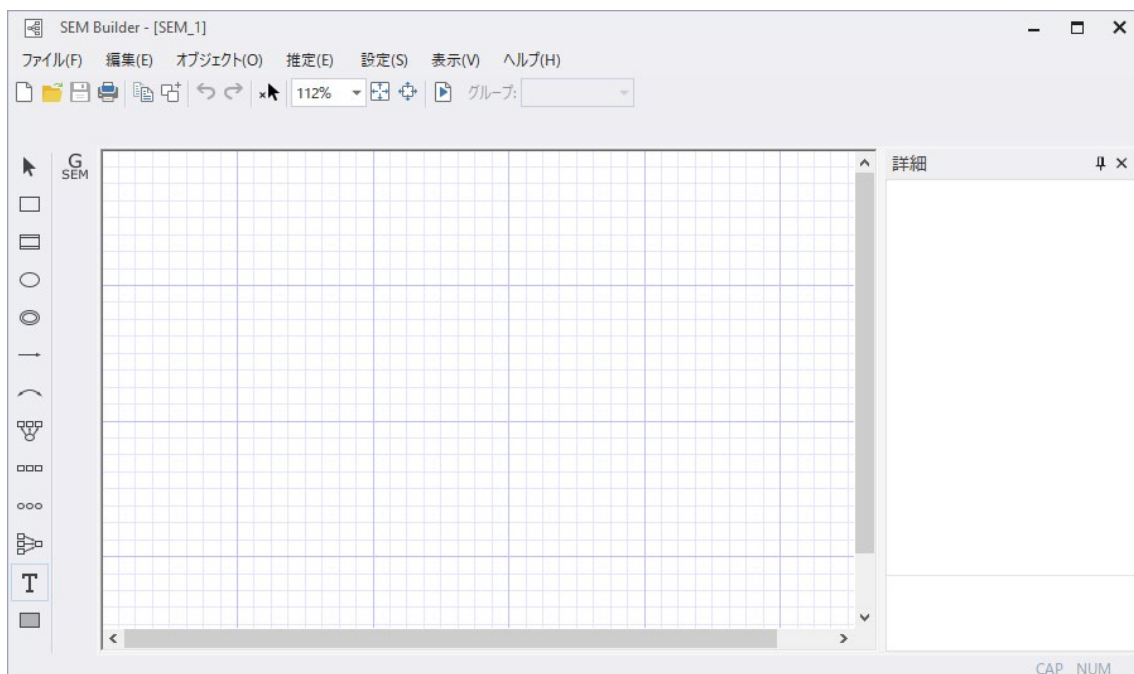
Builder を使ってフィットさせる

- コマンドではなく、Builder を使ってモデルをフィットさせる方法もあります。

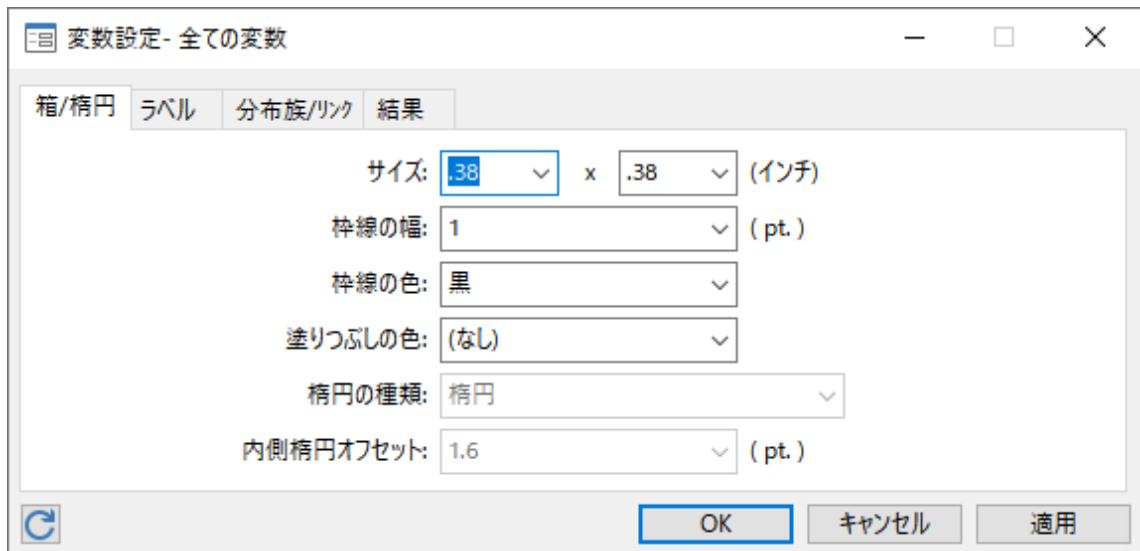
1. コマンドウィンドウで下記を入力し、サンプルデータをダウンロードします。

```
.use https://www.stata-press.com/data/r16/sem_2fmm
```


2. メニューの統計 > SEM (構造方程式モデリング) > モデル構築/推定から、新規の SEM Builder を開きます。



3. 変数ボックスの大きさを設定します。SEM Builder のメニューから、設定 > 変数 > すべて...を選択します。



ダイアログボックスのサイズを.38に変更してOKをクリックします。

4. affective arousal の measurement コンポーネントを作成します。  計測要素を追加ア

アイコンを選択し、SEM Builder の描画領域をクリックします。

ダイアログボックスを下記のように設定します。

- a. グループ変数名を「Affective」にします。
- b. 測定変数に a1, a2, a3, a4, a5 を指定します。
- c. 測定の向きを下にします。
- d. OK をクリックします。

測定成分

メイン 距離 接続線

潜在変数

観測レベルの潜在変数(標準)

マルチレベル潜在変数

グループ変数名:

Affective;

測定変数

変数を選択する

変数の数を指定する

測定変数:

a1 a2 a3 a4 a5

測定変数を一般化する

分布族/リンク: Gaussian, Identity

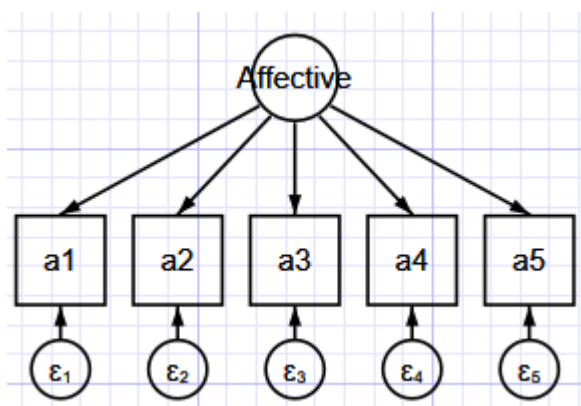
定数を推定しない

測定の向き:

下

OK キャンセル

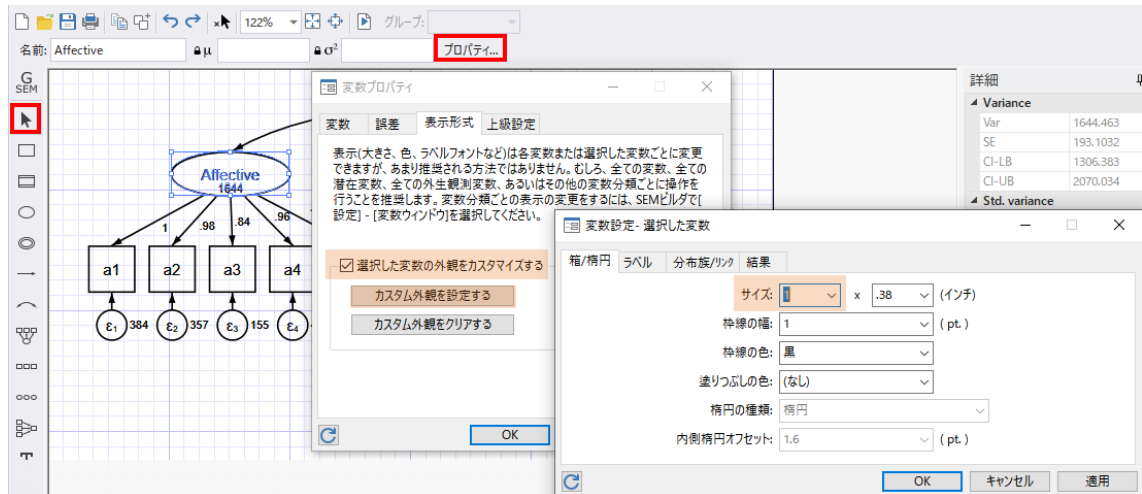
下記の図が作成されます。



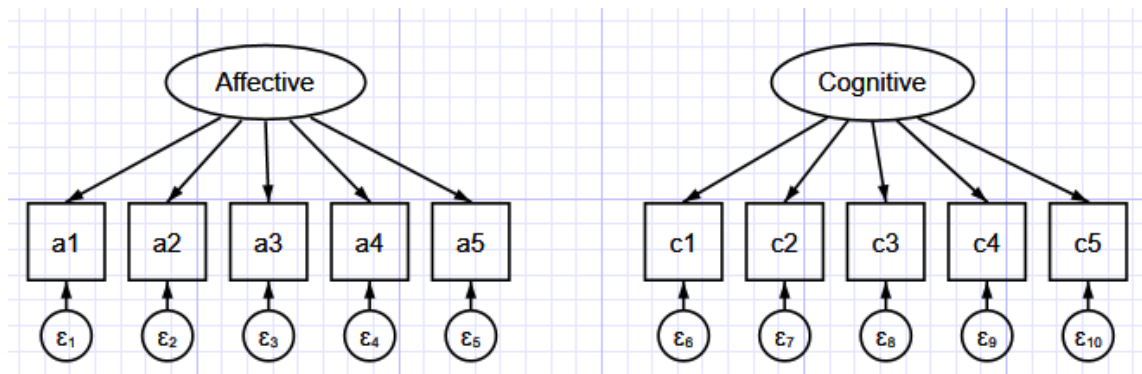
選択をクリックし、変数などを選択してコンポーネントの位置を調整できます。画

面の右下に配置します。

文字を収めるために円を楕円に変更するには、円を選択してプロパティをクリックします。表示形式タブの「選択した変数の外観をカスタマイズする」にチェックを入れ、「カスタム外観を設定する」をクリックし、「サイズ」を修正して適用します。

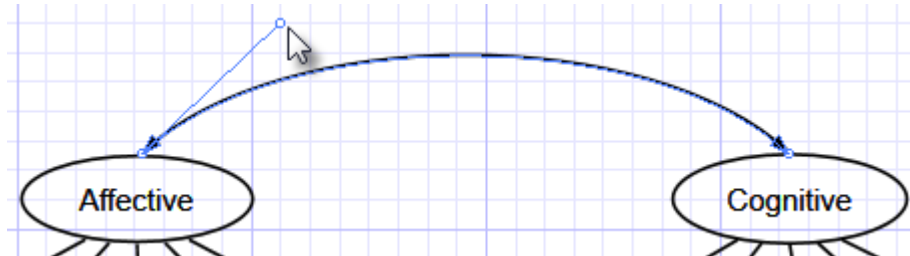


- 手順4を繰り返して、画面の右下に cognitive arousal の measurement コンポーネントを作成します。グループ変数名に「Cognitive」、測定変数に c1, c2, c3, c4, c5 を指定します。




- 相関を作成します。
 - 共分散を追加アイコンをクリックします。
 - 右上の「Affective」の楕円をクリックしてハイライトし、左上の「Cognitive」の楕円にドラッグします。
- 図を整えます。

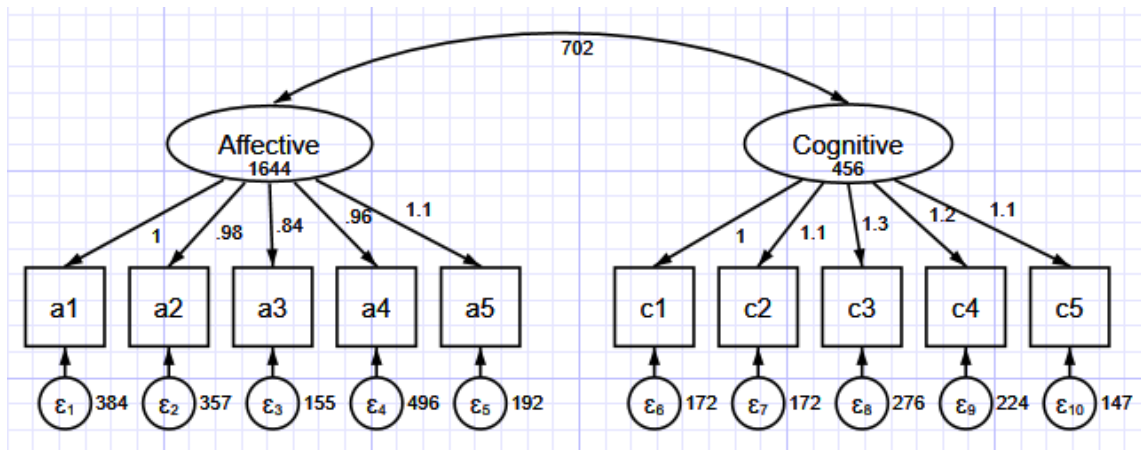
選択ツールでアイコンをクリックしてドラッグし、共変量や変数を移動させることができます。また、共分散の円弧も、端点をクリックして接点の位置や角度を調整することができます。

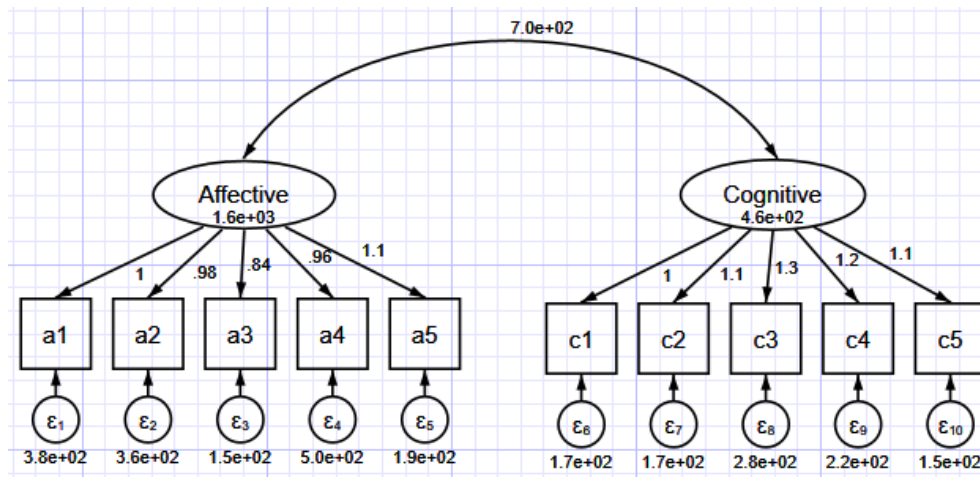


8. 推定を行います。

ツールバーの  推定アイコンをクリックして、SEM 推定オプションウィンドウで OK をクリックします。

9. SEM Builder のメニューから表示 > 推定値を表示にチェックを入れると、推定値が表示されます。





- 下記のコマンドで、例3のモデルを呼び出すことができます。

```
. webgetsem sem_2fmm
```

推定式のフィットの良さを評価する

- a1 と Affective の相関は 0.90 なので、Affective によって説明される a1 の分散は $0.90^2 = 0.81$ で、説明されない分散は $1 - 0.81 = 0.19$ となります。これらを自動的に計算するには、`estat eqgof` コマンドを使います。

```
. estat eqgof
```

Equation-level goodness of fit

depvars	fitted	Variance predicted	residual	R-squared	mc	mc2
observed						
a1	2028.598	1644.463	384.1359	.8106398	.9003553	.8106398
a2	1923.217	1565.865	357.3524	.8141903	.9023249	.8141903
a3	1307.726	1152.775	154.9507	.8815113	.9388883	.8815113
a4	2024.798	1528.339	496.4594	.7548104	.8687982	.7548104
a5	2052.328	1860.643	191.6857	.9066009	.9521559	.9066009
c1	627.5987	455.9349	171.6638	.7264752	.8523351	.7264752
c2	738.3325	566.527	171.8055	.7673061	.8759601	.7673061
c3	1082.374	806.3598	276.0144	.7449917	.863129	.7449917
c4	851.311	627.1116	224.1994	.7366422	.8582786	.7366422
c5	725.3002	578.4346	146.8655	.7975107	.8930346	.7975107
overall				.9949997		

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

- 観測された・されないにかかわらず、内生変数のフィットした分散が表示されます。この例では、観測された内生変数が表示されています。
- **predicted** 欄には、内生変数の分散の予測値が表示されます。
- **residual** 欄には、残差の分散が表示されます。
- **R-squared** 欄には、各 indicator によって説明される分散の割合が表示されます。例えば、1 行目には上で計算した **Affective** によって説明される **a1** の分散 $0.90^2 = 0.81$ が表示されています。**R2** は決定係数とも呼ばれます。
- **mc** は多重相関 (multiple correlation)、**mc2** は多重相関の二乗を表します。
- **R-squared**、**mc**、**mc2** は従属変数の関係性を表します。回帰的なモデルでは、この 3 つの統計量は実質的に同じ数値になります。**mc** は **R-squared** の 2 乗根に等しく、**mc2** は **R-squared** に等しくなります。**mc** と **R-squared** が負の値になることがありますが、これはモデルが負の予測をしている、モデルが妥当ではないという意味ではありません。**mc2** (mc^2) は、非再帰的でないモデルで **R-squared** の代わりに使用することが推奨されています (Bentler and Raykov, 2000)。

例 4 適合度 (goodness-of-fit) の 統計

- **estat gof** コマンドの例を示します。例 3 と同じデータを使用します。

```
. use https://www.stata-press.com/data/r16/sem_2fmm
. sem(Affective -> a1 a2 a3 a4 a5)(Cognitive -> c1 c2 c3 c4 c5)
```

- このモデルを例 3 でフィットさせた時に、計算結果の下の方に次のような結果が出力されました。

```
LR test of model vs. saturated: chi2(34) = 88.88, Prob > chi2 = 0.0000
```

- この飽和モデルを、「model χ^2 test」と呼びます。
- 結果は、フィットが良くないことを示しています。**sem** でのフィットの結果が良くても、フィットは、標本サイズ、相関、モデルに無関係の分散、multivariate nonnormality に影響を受ける可能性があります (Kline, 2016)。
- **sem** での適合度とは、観測される値をどの程度フィットさせることができるかということです。この例では、**a1** から **a5**、**c1** から **c5** の間の共分散です。測定されるモデルでは、根本的な原因 (この例では潜在変数の **Affective** と **Cognitive**) は観測されません。
- 観測される **a1** から **a5**、**c1** から **c5** を **Affective** と **Cognitive** でフィルタリングできると仮定することは手軽かもしれませんが、元の変数に含まれる情報が多く失われている場合のみ有効です。

- 観測されない値がある場合はモデルの適合度は重要ですが、すべての変数が観測される場合は重要ではなくなります。
- 次のコマンドで、適合度を調べます。

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

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(34)	88.879	model vs. saturated
p > chi2	0.000	
chi2_bs(45)	2467.161	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.087	Root mean squared error of approximation
90% CI, lower bound	0.065	
upper bound	0.109	
pclose	0.004	Probability RMSEA <= 0.05
Information criteria		
AIC	19120.770	Akaike's information criterion
BIC	19191.651	Bayesian information criterion
Baseline comparison		
CFI	0.977	Comparative fit index
TLI	0.970	Tucker-Lewis index
Size of residuals		
SRMR	0.022	Standardized root mean squared residual
CD	0.995	Coefficient of determination

- 望ましい値は、検定ごとに異なります。
- 例では、すべての適合度の検定が行われています。オプションを使用して、特定の検定のみ行うこともできます。
- 尤度比検定については、2つの検定が行われます。
- ひとつめは `sem` コマンドで表の下部に出力される `model 2 test` と同じです。この `saturated model` は、共分散を完全にフィットさせるモデルです。
- ふたつめは、baseline 対 saturated comparison です。baseline モデルは、平均と、観測されるすべての変数の分散 + すべての外生変数の共分散のみを含みます。
- どちらも 5% レベル（または他のレベル）で、モデルの saturated モデルに対する適合性を棄却することができます。このレベルの設定は、研究者によって異なります。

- RMSEA 値は、90%信頼区間の上限と下限が表示されます。多くの場合、90%信頼区間の上限か下限のどちらかで検定し、95%信頼区間を持たせます。下限が 0.05 未満または上限が 0.10 を超える場合は、適合度が低いという仮説を棄却しません。
- このモデルは上限が 0.10 を超えているので、適合度が低いという仮説を棄却できません。
- **pclose** は、RMSEA 値が 0.05 未満の確率です。これは、予測された積率が母集団の積率にどのくらい近いかを表しています。このモデルは、近いとはいえません。
- **AIC** と **BIC** は、モデルを比較するために使用されます。値が小さいほど良いとみなされます。
- **baseline comparison** の **CFI** と **TLI** は、1 に近い程良いとみなされます。**TLI** は、**NNFI** (non-normed fit index) とも呼ばれます。
- 最後に、standardized root mean squared residual (**SRMS**) と決定係数 (**CD**) が表示されます。**SRMS** は 0 で完全に適合、0.08 までの値であれば良く適合しているとされます。このモデルは、よく適合しています。
- 決定係数は、モデル全体の **R2** のような値です。1 に近いと良いとされます。
- **estat gof** コマンドは、分野や研究者によって必要な統計が異なるため、複数の適合度の検定を行います。必要な検定の結果のみをご利用ください。