

第1回 確証的因子分析入門

5 回にわたって SEM の考え方と Stata での操作方法を紹介します。ここで利用するテキストは Alan C. Acock, 2013. *Discovering Structural Equation Modeling Using Stata, Revised Edition*, Stata Press です。今回は第一章“Introduction to confirmatory factor analysis”のセクション 9 までの内容を簡単にまとめてご説明します。操作説明には Stata14 日本語版を利用します。

第1章 確証的因子分析入門

- SEM について学ぶ前に因子分析と主成分分析の基礎知識とコマンドを確認する.¹

1.1 因子分析

- 因子分析は潜在的な因子を使ってアウトカムをモデル化する.
- 因子の意味は分析者が与える

因子分析

p 個の質問や問題があるとき, その回答をなるべく少ない因子で表現する.

$$y_{ij} = z_{i1}b_{1j} + z_{i2}b_{2j} + \cdots + z_{iq}b_{qj} + e_{ij}$$

- y_{ij} は j 番目の質問に対する i さんの答え
- z_{ik} は k 番目の共通因子 (common factor) における i さんの係数
- b_{kj} は因子負荷量 (factor loadings)
- e_{ij} は独自因子 (unique factor)
- y_{ij} 以外はすべて推定値である
- p 個の質問をより少ない因子 q でモデル化する

医師のコスト意識に関する調査

- Tarlov et al. (1989) によるデータとその相関行列から再現した bg2.dta を利用する
- 質問は 6 問, 回答は 5 段階.
- 1 が賛成で, 5 は不賛成とする

```
. use bg2,clear
. des
```

¹本講習会のテキストは Alan C. Acock, 2013. *Discovering Structural Equation Modeling Using Stata, Revised Edition*, Stata Press の基本的解説部分と Stata のマニュアルから抜粋したものです.

```

Contains data from http://www.stata-press.com/data/r15/bg2.dta
obs:          568          Physician-cost data
vars:         7           11 Feb 2016 21:54
size:        14,768      (_dta has notes)

```

variable name	storage type	display format	value label	variable label
clinid	int	%9.0g		Physician identifier
bg2cost1	float	%9.0g		Best health care is expensive
bg2cost2	float	%9.0g		Cost is a major consideration
bg2cost3	float	%9.0g		Determine cost of tests first
bg2cost4	float	%9.0g		Monitor likely complications only
bg2cost5	float	%9.0g		Use all means regardless of cost
bg2cost6	float	%9.0g		Prefer unnecessary tests to missing tests

Sorted by: clinid

質問	内容
bg2cost1	最高の医療は高額なものである
bg2cost2*	費用を重視する
bg2cost3*	最初に検査費用を決定する
bg2cost4*	可能性のある合併症だけを監視する
bg2cost5	費用に関係なく、すべての手段を利用する
bg2cost6	抜けがないように多くの検査を行う

因子分析の実行

```
. factor bg2cost1-bg2cost6
```

(obs=568)

```

Factor analysis/correlation          Number of obs   =      568
Method: principal factors           Retained factors =       3
Rotation: (unrotated)              Number of params =     15

```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	0.85389	0.31282	1.0310	1.0310
Factor2	0.54107	0.51786	0.6533	1.6844
Factor3	0.02321	0.17288	0.0280	1.7124
Factor4	-0.14967	0.03951	-0.1807	1.5317
Factor5	-0.18918	0.06197	-0.2284	1.3033
Factor6	-0.25115	.	-0.3033	1.0000

LR test: independent vs. saturated: $\chi^2(15) = 269.07$ Prob> $\chi^2 = 0.0000$

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Uniqueness
bg2cost1	0.2470	0.3670	-0.0446	0.8023
bg2cost2	-0.3374	0.3321	-0.0772	0.7699
bg2cost3	-0.3764	0.3756	0.0204	0.7169
bg2cost4	-0.3221	0.1942	0.1034	0.8479
bg2cost5	0.4550	0.2479	0.0641	0.7274
bg2cost6	0.4760	0.2364	-0.0068	0.7175

分析チェック

- 固有値 (Eigenvalue) が正のものを候補とする
- 固有値が 1 以上のある因子を用いて解釈を与える
- 因子 1:費用に対する医師の平均的な態度
- 因子 2:すべての質問で正. 全ての提案に賛成するという姿勢

その他の用語

$$y_{ij} = z_{i1}b_{1j} + z_{i2}b_{2j} + z_{i3}b_{3j} + e_{ij}$$

- 共通性 z_{ik} (因子), b_{kj} (因子負荷量)
- 独自性 (Uniqueness) とは共通性 z_{ik} では説明できない, それ以外の分散の比. 1-独自性=共通性.
- 尤度比検定は全ての変数の分散共分散をモデル化した saturated と分散だけの情報による independent モデルを比較したもの

1.2 主成分分析

- 主成分分析 (principal component factor analysis) は数多くのデータをなるべく少ない情報 (主成分) に集約することを目的とする分析手法
- 主成分の意味は分析者が与える
- 例えば, 2 つのアウトカムがある場合. その合成変量を次のように定義する.

$$z = a_1x_1 + a_2x_2$$

- この合成変量の分散を最大化したり, アウトカム x との重相関係数の二乗和を最大化する
- この最大化問題は解析的に解くことができる
- その計算過程で固有値と固有ベクトルが登場する

データの内容

- サンプルデータ:nlsy97cfa.dta
- “政府の役割”について 20 代の人を対象にした調査.
- より良い生活を実現するために政府はより積極的に行動すべきか?

質問内容

- x1: より多くの雇用を提供すべき
- x2: 物価をコントロールすべき
- x3: 医療制度をより充実させるべき
- x4: 高齢者の自立を支援すべき
- x5: 産業界への支援を充実させるべき
- x6: 失業者への自立を支援すべき
- x7: 所得格差を解消すべき
- x8: 大学への補助金を充実すべき
- x9: 住宅への補助金を充実すべき
- x10: 環境問題に取り組むべき

- 回答は1から4
- 1は当該の役割は政府の為すべき事柄である
- 4は政府の為すべき事柄ではない (Conservative)
- 主成分分析の実行

```
. use "nlsy97cfa.dta", clear
. codebook x1-x10, compact
```

(結果は省略)

- 合計の数字が大きいほど Conservative である
- Conservative に関する質問項目は x1-x10 まで

操作

```
. factor x1-x10, pcf
```

(obs=1,617)

```
Factor analysis/correlation          Number of obs =      1,617
Method: principal-component factors   Retained factors =      2
Rotation: (unrotated)                 Number of params =     19
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	3.91523	2.90094	0.3915	0.3915
Factor2	1.01429	0.13285	0.1014	0.4930
Factor3	0.88144	0.11496	0.0881	0.5811
Factor4	0.76648	0.02404	0.0766	0.6577
Factor5	0.74243	0.04889	0.0742	0.7320
Factor6	0.69354	0.08649	0.0694	0.8013
Factor7	0.60705	0.06820	0.0607	0.8620
Factor8	0.53886	0.09140	0.0539	0.9159
Factor9	0.44746	0.05424	0.0447	0.9607
Factor10	0.39322	.	0.0393	1.0000

LR test: independent vs. saturated: $\chi^2(45) = 4083.46$ Prob> $\chi^2 = 0.0000$

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
x1	0.6064	-0.3789	0.4888
x2	0.5810	0.0438	0.6605
x3	0.7221	0.2140	0.4328
x4	0.7174	0.3200	0.3830
x5	0.5780	-0.0261	0.6653
x6	0.6091	-0.4536	0.4233
x7	0.6050	-0.3327	0.5233
x8	0.5994	0.3252	0.5350
x9	0.7330	-0.1621	0.4365
x10	0.4543	0.5211	0.5221

主成分の選択

- 固有値 (Eigenvalue) が 1 以上なのは第 2 主成分まで
- 固有値と因子負荷量 (loadings) の関係

$$3.91 = \sum_{i=1}^{10} \text{factor}_i^2$$

- 因子負荷量は 0.4 以上 (0.3 という主張もある)
- 主成分分析は変数を標準化し, 合成変量 (尺度) を相関行列または分散共分散行列で作成する.
- その合成変量を最大化するように固有値や因子負荷量を計算する

操作

- 質問項目 10 は環境問題に関する質問なので, 試しに省いてみる

```
. factor x1-x9, pcf
```

```
(obs=1,625)
```

```
Factor analysis/correlation      Number of obs   =      1,625
Method: principal-component factors  Retained factors =      1
Rotation: (unrotated)             Number of params =      9
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	3.76124	2.80650	0.4179	0.4179
Factor2	0.95473	0.10627	0.1061	0.5240
Factor3	0.84847	0.10176	0.0943	0.6183
Factor4	0.74671	0.05561	0.0830	0.7012
Factor5	0.69110	0.07429	0.0768	0.7780
Factor6	0.61681	0.07780	0.0685	0.8466
Factor7	0.53900	0.09177	0.0599	0.9065
Factor8	0.44723	0.05252	0.0497	0.9561
Factor9	0.39471	.	0.0439	1.0000

LR test: independent vs. saturated: $\chi^2(36) = 3863.18$ Prob> $\chi^2 = 0.0000$
 Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Uniqueness
x1	0.6243	0.6103
x2	0.5883	0.6539
x3	0.7222	0.4785
x4	0.7131	0.4915
x5	0.5818	0.6615
x6	0.6197	0.6160
x7	0.6085	0.6297
x8	0.5968	0.6439
x9	0.7392	0.4535

- 固有値が 1 以上なのは第 1 主成分だけになった
- 9 個の項目を一つの主成分に集約した
- Uniqueness とは, その質問 (項目) 独自の変動
- 例えば 61% とあるのは, 項目 x1 の分散のうち, 61% は主成分で説明できないことを示す
- 主成分分析は Uniqueness をゼロと仮定しているのだから, 小さいほど良い

信頼度

- 9 つの項目 (質問) は Conservative という概念を計測するための合理的な質問になっているのか?
- クロンバックの α を用いて項目の適切さを調べる

$$\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$$

```
. alpha x1-x9,item label
```

```
Test scale = mean(unstandardized items)
```

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem cov.	alpha	Label
x1	1833	+	0.6638	0.5055	.1985677	0.7892	GOVT RESPONSIBILITY - PROVIDE JOBS 2006
x2	1859	+	0.5894	0.4545	.2184795	0.7933	GOVT RESPNSBLTY - KEEP PRICES UND CTRL 2006
x3	1874	+	0.6689	0.5728	.2157707	0.7814	GOVT RESPNSBLTY - HLTH CARE FOR SICK 2006
x4	1872	+	0.6584	0.5680	.2195358	0.7834	GOVT RESPNSBLTY -PROV ELD LIV STAND 2006
x5	1815	+	0.5818	0.4410	.2186451	0.7947	GOVT RESPNSBLTY -PROV IND HELP 2006
x6	1811	+	0.6502	0.5031	.2045592	0.7879	GOVT RESPNSBLTY -PROV UNEMP LIV STAND 2006
x7	1775	+	0.6559	0.4870	.1984449	0.7934	GOVT RESPNSBLTY -REDUCE INC DIFF 2006
x8	1875	+	0.5396	0.4409	.234795	0.7966	GOVT RESPNSBLTY -PROV COLL FIN AID 2006
x9	1847	+	0.7171	0.6216	.2050897	0.7739	GOVT RESPNSBLTY -PROV DECENT HOUSING 2006
Test scale					.2126229	0.8072	mean(unstandardized items)

- Test scale $\alpha = 0.81$ (目安は 0.70 以上)
- 例えば, 項目 x1 を除外すると α は 0.789 に減少する
- ここではどれも除外しない時に信頼度が最も高い
- 項目 (質問) の信頼度は確保されている
- ただし, 例えば, $\bar{r} (= 0.17)$ でも項目数が 40 になると α は 0.8 という高い値になってしまう

回答の平均

各人の保守的傾向を回答の平均で調べる

```
. egen conserve = rowmean(x1-x9)
```

- 各人 (横方向) に回答の合計を求める
- conserve の欠損値が 7,097 行 (全体 8,985 行)
- conserve の記述統計量

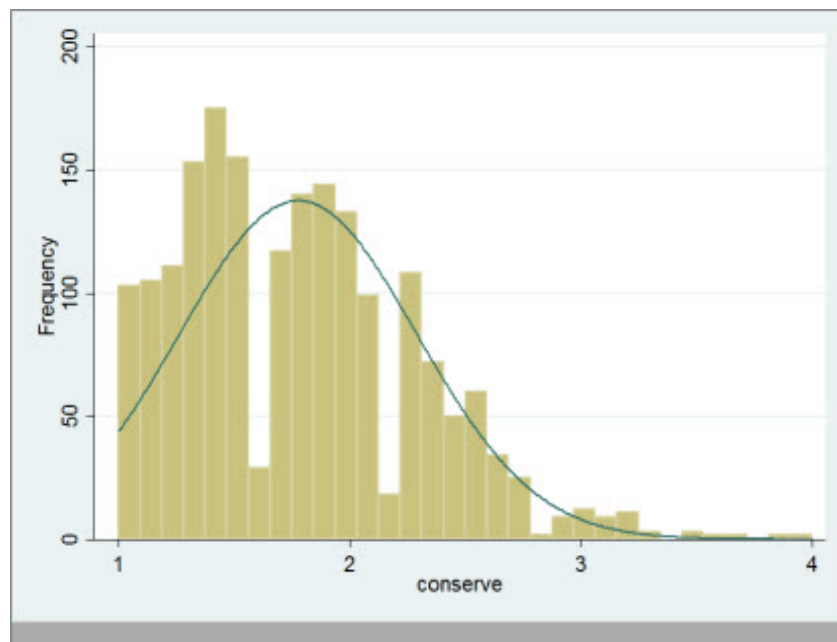

```
. summarize conserve, detail
```

conserve					
	Percentiles	Smallest			
1%	1	1			
5%	1	1			
10%	1.111111	1	Obs	1,888	
25%	1.354167	1	Sum of Wgt.	1,888	
50%	1.690476		Mean	1.775299	
		Largest	Std. Dev.	.5132186	
75%	2.111111	3.888889	Variance	.2633934	
90%	2.444444	3.888889	Skewness	.7200074	
95%	2.666667	4	Kurtosis	3.537959	
99%	3.222222	4			

分析チェック

- 平均 1.78, 標準偏差 0.51, 1 問以上質問に回答した人数 1,888 人
- 回答の平均値のヒストグラム

```
. histogram conserve, norm freq
```



- 数字 (回答番号の平均値) が小さいほど Conservative である
- 曲線は正規分布
- 分布は右裾が長い。つまり, 積極的な政府の役割を期待している人が多い

主成分得点

- 主成分分析の第一主成分の因子負荷量を利用して個人ごとの主成分得点 `conservf1` を計算する
- これは一般的に平均や合計などの統計量と強い相関を持つ

```
. factor x1-x9, pcf
```

(結果は省略)

```
. predict conservf1
```

```
(regression scoring assumed)
```

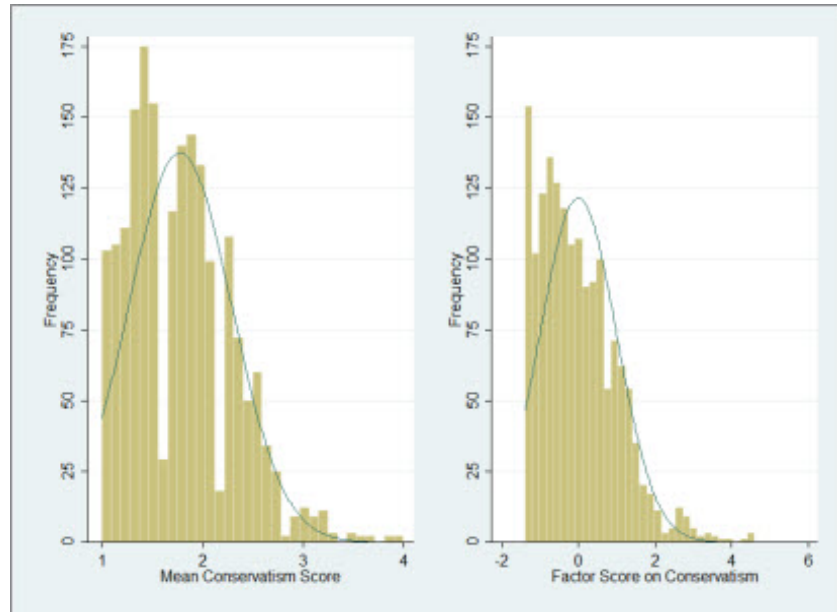
```
Scoring coefficients (method = regression)
```

Variable	Factor1
x1	0.16598
x2	0.15641
x3	0.19200
x4	0.18958
x5	0.15468
x6	0.16475
x7	0.16179
x8	0.15866
x9	0.19654

- この因子負荷量は標準化したもの
- 各個人の回答と因子負荷量の積和を主成分得点とする
- 第9番目の係数は約 2.0, 一方, 2番目の係数は約 0.16. 主成分に与える影響は9番目の項目が大きい
- 平均値 `conserve` と主成分得点 `conservf1` の相関はかなり高い
- 次に示す `do` ファイル `hist01.do` を使ってヒストグラムを作成する

(`hist01.do` の内容)

```
histogram conserve, norm freq name(A, replace) ///
xtitle(Mean Conservatism Score) ylabel(0(25)175)
histogram conservf1, norm freq name(B, replace) ///
xtitle(Factor Score on Conservatism) ylabel(0(25)175)
graph combine A B
```



分析チェック

- 主成分分析は各項目（質問）の相関行列, または, 分散共分散行列を利用して主成分を計算する
- モデルは作成しないので, すべての分析者の答えは同じになる. 主成分の解釈のみ, 分析者が考えて与える
- 各項目（質問）の合理性は信頼度で確認できる
- 主成分得点は各項目の平均と強い相関を持つ

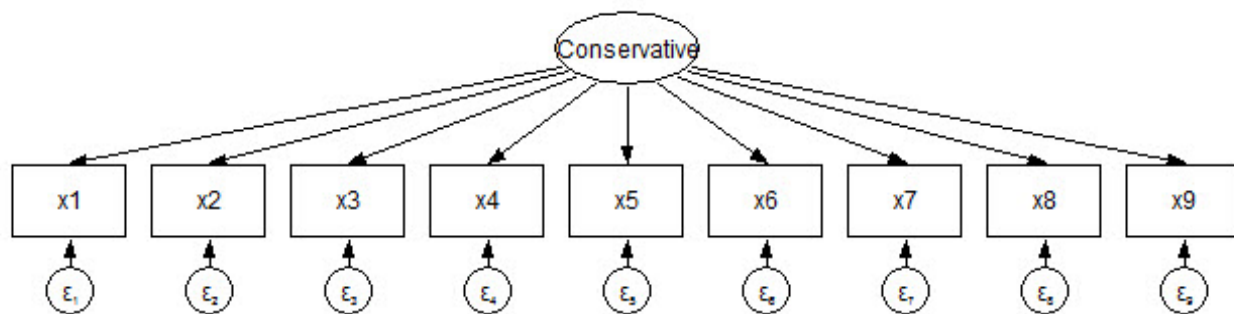
1.3 確証的因子分析

- 確証的因子分析 (CFA: Conirmatory Factor Analysis)
- 主成分分析は変数を標準化する (分散 1) が, CFA と一般の因子分析は分散に制約なし
- 潜在変数として Conservative の存在を仮定する
- Conservative は 9 つの質問項目 ($x_1 - x_9$) で共有される
- 仮定: 誤差項 $\epsilon_1 - \epsilon_9$ は正規分布に従い, 互いに独立である
- 潜在変数は全ての質問 (項目) に対する応答に共通する

パス図の作成

- メニューを利用する場合は統計/SEM/モデル構築, 推定と操作して SEM ビルダの画面を表示する
- コマンドで SEM ビルダを起動する場合は `sembuilder` と入力する

- オブジェクトを追加するアイコンをクリックしたら、グループ変数名に“Conservative”，測定変数に“x1-x9”と入力する
- 必要に応じて微調整し、次のようなパス図を作成する



モデル推定

- SEM ビルダーで推定/推定と操作してダイアログを表示する
- 推定手法のデフォルトは最尤法 (ml)
- mlmv: 多変量正規分布を仮定して、リストワイズにデータを削除しない
- adf: 正規分布の仮定を利用しない加重最小二乗法

```
. sem (Conservative->x1-x9)
```

```
(7360 observations with missing values excluded)
```

```
Endogenous variables
```

```
Measurement: x1 x2 x3 x4 x5 x6 x7 x8 x9
```

```
Exogenous variables
```

```
Latent: Conservative
```

```
Fitting target model:
```

```
(省略)
```

```
Structural equation model
```

```
Number of obs = 1,625
```

```
Estimation method = ml
```

```
Log likelihood = -15593.729
```

```
( 1) [x1]Conservative = 1
```

	OIM				[95% Conf. Interval]	
	Coef.	Std. Err.	z	P> z		
Measurement						
x1 <-	1 (constrained)					
Conservative	2.329846	.0253521	91.90	0.000	2.280157	2.379535
_cons						
x2 <-	.7377011	.0451423	16.34	0.000	.6492237	.8261784
Conservative	1.617231	.0198829	81.34	0.000	1.578261	1.656201
_cons						
x3 <-	.8267157	.0432635	19.11	0.000	.7419209	.9115105
Conservative	1.414154	.0167434	84.46	0.000	1.381337	1.44697
_cons						
x4 <-	.7555335	.0403806	18.71	0.000	.676389	.834678
Conservative	1.362462	.0155865	87.41	0.000	1.331913	1.39301
_cons						
x5 <-	.7380149	.0462134	15.97	0.000	.6474383	.8285914
Conservative	1.769846	.0202603	87.36	0.000	1.730137	1.809556
_cons						
x6 <-	.9146378	.053406	17.13	0.000	.8099639	1.019312
Conservative	2.259692	.0229301	98.55	0.000	2.21475	2.304634
_cons						
x7 <-	1.028027	.0614681	16.72	0.000	.9075522	1.148503
Conservative	2.219692	.0266439	83.31	0.000	2.167471	2.271913
_cons						
x8 <-	.5486913	.033463	16.40	0.000	.483105	.6142775
Conservative	1.307077	.0141374	92.46	0.000	1.279368	1.334786
_cons						
x9 <-	.9278118	.0479147	19.36	0.000	.8339008	1.021723
Conservative	1.705231	.0187041	91.17	0.000	1.668571	1.74189
_cons						
var(e.x1)	.7287257	.0280851			.6757076	.7859038
var(e.x2)	.4706031	.0178489			.4368885	.5069195
var(e.x3)	.2397812	.0104761			.2201029	.2612188
var(e.x4)	.2145611	.009255			.1971672	.2334895
var(e.x5)	.4950753	.0186802			.4597838	.5330757
var(e.x6)	.590299	.0229507			.5469876	.6370399
var(e.x7)	.8199315	.0314634			.7605262	.8839769
var(e.x8)	.2297334	.0087974			.213122	.2476396
var(e.x9)	.2967257	.0129788			.2723476	.3232858
var(Conservative)	.3157048	.0287081			.264167	.3772973

LR test of model vs. saturated: chi2(27) = 419.01, Prob > chi2 = 0.0000

ここでは次のような回帰分析を行っている.

$$x1 = \alpha_1 + \beta_1 \text{Conservative} + \epsilon_1$$

- この単純回帰モデルを $x1 - x9$ まで 9 本連立させる
- 表の下の方には誤差項 ϵ の分散を表示

分析チェック

- 全ての質問に共通するのは Conservative
- 因子分析とは異なり, Conservative に第一因子、第二因子などはない

推定結果の解釈

- デフォルトの最尤法では欠損値に対してリストワイズな削除を行うので, 7,360 行のデータを推定から除外
- 被説明変数の $x1, \dots, x9$ の事は内生変数と呼ぶ
- $x1$ に制約を付けて推定した係数のことを非標準化 β と呼ぶ
- 因子負荷量を 1 に固定した $x1$ のことをリファレンスインジケータと呼ぶ. 計算上, どこかに制約が必要
- 非標準化 β を計算するにはリファレンスインジケータが必要
- 一番大きな負荷量の変数をリファレンスインジケータとすると考えもある
- 例えば, 7 番目の変数をリファレンスインジケータとする場合は次のようにする

参考 (ここでは操作しない)

```
. sem (Conservative -> x7 x1-x6 x8 x9 )
```

- 標準化 β (観測できる変数と潜在変数の分散をともに 1 とする)

```
. sem, standardized
```

```
Structural equation model           Number of obs   =   1,625

Estimation method   = ml
Log likelihood      = -15593.729
( 1) [xi]Conservative = 1
```

Standardized	OIM		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
Measurement						
x1 <-						
Conservative	.549795	.0200518	27.42	0.000	.5104942	.5890958
_cons	2.279751	.0470588	48.44	0.000	2.187518	2.371985
x2 <-						
Conservative	.5171478	.0208436	24.81	0.000	.4762951	.5580005
_cons	2.017742	.0432214	46.68	0.000	1.93303	2.102454
x3 <-						
Conservative	.6882205	.016421	41.91	0.000	.656036	.720405
_cons	2.09521	.044341	47.25	0.000	2.008303	2.182116
x4 <-						
Conservative	.6756463	.0168425	40.12	0.000	.6426355	.7086571
_cons	2.16845	.0454115	47.75	0.000	2.079445	2.257455
x5 <-						
Conservative	.5077306	.0210256	24.15	0.000	.4665212	.54894
_cons	2.167021	.0453905	47.74	0.000	2.078057	2.255985
x6 <-						
Conservative	.555978	.0200541	27.72	0.000	.5166727	.5952834
_cons	2.444653	.0495404	49.35	0.000	2.347556	2.541751
x7 <-						
Conservative	.5378005	.0204071	26.35	0.000	.4978034	.5777977
_cons	2.066659	.0439268	47.05	0.000	1.980564	2.152753
x8 <-						
Conservative	.5409708	.0202147	26.76	0.000	.5013507	.580591
_cons	2.29354	.0472646	48.53	0.000	2.200903	2.386177
x9 <-						
Conservative	.6914122	.0162749	42.48	0.000	.6595139	.7233104
_cons	2.26162	.046789	48.34	0.000	2.169916	2.353325
var(e.x1)	.6977254	.0220488			.6558217	.7423066
var(e.x2)	.7325582	.0215584			.6914999	.7760543
var(e.x3)	.5263525	.0226025			.4838655	.5725702
var(e.x4)	.5435021	.0227592			.5006764	.589991
var(e.x5)	.7422096	.0213507			.7015209	.7852583
var(e.x6)	.6908884	.0222993			.6485363	.7360063
var(e.x7)	.7107706	.0219499			.6690257	.7551202
var(e.x8)	.7073505	.0218711			.6657569	.7515427
var(e.x9)	.5219492	.0225054			.4796519	.5679763
var(Conservative)	1	.			.	.

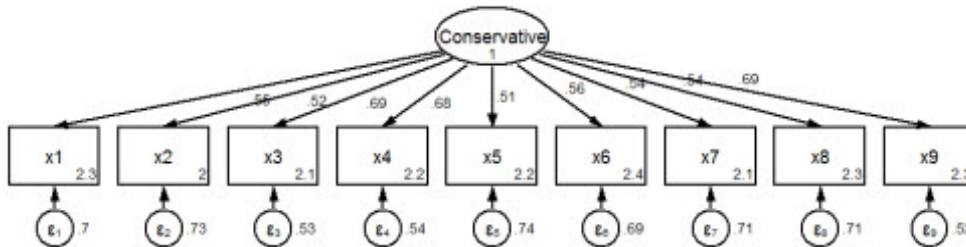
```
LR test of model vs. saturated: chi2(27) = 419.01, Prob > chi2 = 0.0000
```

- 尤度比検定の saturated model とは観測可能な変数間の共分散を内包するモデル
- 帰無仮説は「推定したモデルは全ての共分散構造の情報を有している」

- ここでは帰無仮説が棄却されているので、モデルの見直しが必要

パス図のダイアログで同じ推定を実行する

- 推定ダイアログの表示タブで標準化係数と値を表示するを選択する



- 因子負荷量はすべて有意
- 例えば、Conservative が 1 単位大きくなった時、質問 x1 に対して 0.55 だけ大きく応答する
- 同様に、質問 2 に対しては 0.52 だけ高く応答する

適合度

推定した SEM モデルの良さを評価する

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

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(27)	419.007	model vs. saturated
p > chi2	0.000	
chi2_bs(36)	3872.316	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.095	Root mean squared error of approximation
90% CI, lower bound	0.087	
upper bound	0.103	
pclose	0.000	Probability RMSEA <= 0.05
Information criteria		
AIC	31241.457	Akaike's information criterion
BIC	31387.075	Bayesian information criterion
Baseline comparison		
CFI	0.898	Comparative fit index
TLI	0.864	Tucker-Lewis index
Size of residuals		
SRMR	0.049	Standardized root mean squared residual
CD	0.835	Coefficient of determination

尤度比検定

chi2_ms(27) 構築した SEM モデルと saturated model(飽和モデル) の尤度比検定

帰無仮説: CFA のモデルで観測できる変数の共分散構造を完全に表現できている。

- これは SEM の推定結果の一番下に表示されている情報である
- saturated model とは制約のないモデルを使ってフィットを実行し, 変数の共分散構造を再現するモデルの事. 詳細は第二章で解説する
- この検定は, 自分の作製したモデルの良さを調べるためのもの

chi2_bs(36) baseline model と saturated model の尤度比検定

帰無仮説: baseline model は saturated model(飽和モデル) と同様の説明力 (共分散構造) を持つ

- baseline model とは殆どの共分散はゼロという制約を掛けたモデルの事. 詳細は第二章で解説する
- この検定は共分散の有無によって, 尤度がどれくらい異なるのかを調べるためのもの

RMSEA

- Root Mean Squared Error of Approximation
- モデルに不要な複雑さが存在すると数値が大きくなる
- 一般的に 0.05 が良く, 0.08 でほどほどのフィットと評価する

$$\text{RMSEA} = \sqrt{\frac{T(N-1)}{df}}$$

- $T = \max(\text{model chi-squared} - df, 0)$, N は標本サイズ.
- よって, この例題の結果は良いものではない

情報量規準

- 同じ質問項目を利用した異なるモデルとの比較の際に利用する
- 単独のモデルの評価には利用しない

ベースラインとの比較

- 尤度比検定から CFA モデルは完全なものではない事が分かった
- それではどの程度, 良いものなのか?
- CFI(comparative fit index) を利用する
- CFI:89.8%. 全ての項目 (質問) が独立であるとするヌルモデルに比べると, 89.8% 優れている
- 一般的に 0.90 や 0.95 よりも大きければ良いとされている. ここでは 0.898

残差のサイズ

- Standardized Root Mean Squared Residual(SRMR)
- SRMA = 0.045. 一般的に 0.05 以下が良いとされる
- 観測可能な 9 つの内生変数の相関をモデルで表現するための指標とされる
- 相関とは異なるので, 注意すること
- CD(Coefficient of Determination) はモデルフィットの良さを示す指標. 最大値は 1.

共分散構造を確認する

- SEM において潜在変数 Conservative を想定せずに, 変数の平均/分散/共分散をすべて求める

```
. sem x1-x9
```

(出力は省略)

```
.estat framework,fitted
```

Fitted covariances of observed variables

	Sigma	observed x1	x2	x3	x4	x5
observed						
x1		1.044432				
x2		.3016396	.6424108			
x3		.2498545	.2046792	.4555535		
x4		.1992127	.1793545	.2498852	.3947755	
x5		.2482231	.2165196	.1611653	.1806527	.6670292
x6		.3475723	.1387868	.2081397	.19141	.2302307
x7		.37492	.2434761	.2320905	.2028315	.2111784
x8		.1393272	.118155	.1491306	.1465426	.1309826
x9		.2843054	.1899406	.2217721	.1948425	.2170808

	Sigma	observed x6	x7	x8	x9
observed					
x6		.8544061			
x7		.3817168	1.153581		
x8		.1122545	.1651529	.3247807	
x9		.3362416	.3459893	.1828245	.5684957

Fitted means of observed variables

	mu	observed x1	x2	x3	x4	x5
mu						
mu		2.329846	1.617231	1.414154	1.362462	1.769846

	mu	observed x6	x7	x8	x9
mu					
mu		2.259692	2.219692	1.307077	1.705231

- 質問項目は9個なので、共分散行列の要素は45個(共分散35+分散9)
- 推定する負荷量8個(最初の項目には1という制約がある)+誤差分散9個+Conservativeの分散=18個のパラメータを推定
- よって、カイ二乗検定の自由度は $45 - 18 = 27$
- フィットした共分散や、フィットした平均の詳細は第二章で解説する

モデルの改良

- 推定したモデルの改良を試みる
- つまり, saturated model との尤度比検定の検定統計量(カイ二乗値)を小さくできるか, ということを考える
- そのためにここではパラメータを追加する
- 項目(質問)の誤差項間に相関を想定する
- どの質問間に相関を想定するか?

```
. sem (Conservative -> x1-x9)
```

(結果は省略)

```
. estat mindices
```

Modification indices

	MI	df	P>MI	EPC	Standard EPC
cov(e.x1,e.x2)	27.046	1	0.00	.0830313	.1417856
cov(e.x1,e.x4)	22.437	1	0.00	-.0549145	-.1388768
cov(e.x1,e.x6)	16.165	1	0.00	.0727568	.1109317
cov(e.x1,e.x7)	8.434	1	0.00	.0615727	.0796558
cov(e.x1,e.x8)	13.662	1	0.00	-.041522	-.1014807
cov(e.x2,e.x5)	16.394	1	0.00	.0526548	.1090874
cov(e.x2,e.x6)	39.084	1	0.00	-.090013	-.1707818
cov(e.x2,e.x9)	11.090	1	0.00	-.0364539	-.0975527
cov(e.x3,e.x4)	147.976	1	0.00	.0889115	.3919898
cov(e.x3,e.x5)	18.664	1	0.00	-.0433453	-.1258053
cov(e.x3,e.x6)	15.379	1	0.00	-.0438055	-.1164353
cov(e.x3,e.x7)	15.291	1	0.00	-.0510741	-.1151874
cov(e.x3,e.x9)	16.559	1	0.00	-.035565	-.133333
cov(e.x4,e.x6)	12.898	1	0.00	-.0375737	-.1055778
cov(e.x4,e.x7)	22.944	1	0.00	-.0586131	-.1397431
cov(e.x4,e.x8)	11.218	1	0.00	.0217217	.0978375
cov(e.x4,e.x9)	30.371	1	0.00	-.0449624	-.1781953
cov(e.x6,e.x7)	29.683	1	0.00	.1041754	.1497409
cov(e.x6,e.x8)	31.435	1	0.00	-.0568032	-.1542497
cov(e.x6,e.x9)	62.385	1	0.00	.0984096	.2351387
cov(e.x7,e.x9)	19.055	1	0.00	.0635887	.128918
cov(e.x8,e.x9)	16.553	1	0.00	.0314136	.1203174

EPC = expected parameter change

- MI は当該の相関を想定した時に減少する検定統計量 (カイ二乗値)
- $P > MI$ の列から減少分は有意であることが分かる
- MI が一番大きなところに着目する (x3 と x4 の誤差項の共分散)
- 項目 2 は政府の物価への介入なので, Conservartive とは関係ないものとしてパス図から削除する
- 項目 3 は「健康」を損ねた人への支援. 項目 4 は「高齢者」に対する政府の支援を問うもの.

```
. codebook x3 x4,compact
```

```
. codebook x3 x4,compact
```

```
Variable   Obs Unique      Mean  Min  Max  Label
```

```
x3          1874      4  1.416222    1    4  GOVT RESPNSBLTY - HLTH CARE FOR SICK 2006
x4          1872      4  1.365385    1    4  GOVT RESPNSBLTY -PROV ELD LIV STAND 2006
```

- パス図で誤差項に相関を設定する
- さらに項目 8 は大学に対する助成で, 保守性との関連も考えられるが, 社会よりも個別の家庭に対する利益と考えられるので, これも除外する
- Conservartive の分散を 1 にするという制約を掛けて非標準化係数を推定する
- コマンドの場合は次の通り

```
. sem (Conservative -> x1 x3-x7 x9),cov(e.x3*e.x4) variance(Conservative@1)
```

(7354 observations with missing values excluded)

Endogenous variables

Measurement: x1 x3 x4 x5 x6 x7 x9

Exogenous variables

Latent: Conservative

Fitting target model:

(省略)

Structural equation model Number of obs = 1,631

Estimation method = ml

Log likelihood = -12634.282

(1) [var(Conservative)]_cons = 1

	OIM		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
Measurement						
x1 <-						
Conservative	.5731977	.0262634	21.82	0.000	.5217223	.624673
_cons	2.332311	.0253234	92.10	0.000	2.282679	2.381944
x3 <-						
Conservative	.3928852	.0173651	22.62	0.000	.3588501	.4269203
_cons	1.41447	.0167666	84.36	0.000	1.381608	1.447332
x4 <-						
Conservative	.3541811	.0162626	21.78	0.000	.322307	.3860552
_cons	1.364194	.0155794	87.56	0.000	1.333659	1.394729
x5 <-						
Conservative	.4042766	.0213058	18.97	0.000	.3625179	.4460353
_cons	1.769467	.0201988	87.60	0.000	1.729878	1.809055
x6 <-						
Conservative	.5942818	.0230962	25.73	0.000	.5490141	.6395494
_cons	2.261189	.0229305	98.61	0.000	2.216247	2.306132
x7 <-						
Conservative	.6178507	.0273552	22.59	0.000	.5642355	.6714658
_cons	2.219497	.0265943	83.46	0.000	2.167373	2.271621
x9 <-						
Conservative	.5474424	.0182758	29.95	0.000	.5116225	.5832622
_cons	1.705089	.0186488	91.43	0.000	1.668538	1.74164
var(e.x1)	.7173593	.0287599			.6631487	.7760015
var(e.x3)	.3041442	.0124605			.2806769	.3295736
var(e.x4)	.2704286	.0109494			.2497976	.2927635
var(e.x5)	.5019923	.0192822			.4655875	.5412437
var(e.x6)	.5044216	.0218112			.4634339	.5490344
var(e.x7)	.7717982	.0311739			.7130544	.8353815
var(e.x9)	.2675341	.0134682			.2423975	.2952774
var(Conservative)	1	(constrained)				
cov(e.x3,e.x4)	.110268	.0091451	12.06	0.000	.092344	.128192

LR test of model vs. saturated: chi2(13) = 56.02, Prob > chi2 = 0.0000

分析チェック

- model vs.saturated の検定統計量は小さくなったか?

419 から 56 まで小さくなった

- estat gof コマンドで次の項目を確認してみましょう

- CFI は 95% 程度になったか?

0.898 から 0.985 になった

- RMSEA は 0.05 程度になったか?

0.095 から 0.045 になった

- SRMR は 0.08 以下を保っているか?

0.049 から 0.023 で 0.08 以下を保っている

- 情報量規準は小さくなったか?

AIC は 31241 から 25312 と小さくなって