

# 合理进行多元分析——复杂结构方程模型分析

胡纯严<sup>1</sup>, 胡良平<sup>1,2\*</sup>

(1. 军事科学院研究生院, 北京 100850;

2. 世界中医药学会联合会临床科研统计学专业委员会, 北京 100029

\*通信作者: 胡良平, E-mail: lphu927@163.com)

**【摘要】** 本文目的是介绍与复杂结构方程模型分析有关的基本概念、计算方法、两个实例以及使用 SAS 实现计算的方法。基本概念包括自由参数与固定参数、直接效应与间接效应、模型的基本构成和模型修正; 计算方法涉及模型识别、模型估计和模型评价; 两个实例的资料分别是“小学教师工作满意度”和“932 名受试者疏离感与社会经济地位的关系”; 借助 SAS 软件, 对两个实例中的数据分别进行了结构方程模型分析, 对 SAS 输出结果做出了解释。

**【关键词】** 自由参数; 固定参数; 模型识别; 模型修正; 模型估计; 模型评价

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20230925004

## Reasonably carry out multivariate analysis: complex structural equation model analysis

Hu Chunyan<sup>1</sup>, Hu Liangping<sup>1,2\*</sup>

(1. Graduate School, Academy of Military Sciences PLA China, Beijing 100850, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

\*Corresponding author: Hu Liangping, E-mail: lphu927@163.com)

**【Abstract】** The purpose of this article was to introduce the basic concepts, calculation methods, two examples related to the complex structural equation model analysis and how to implement calculations using SAS. Basic concepts included free parameters and fixed parameters, direct effects and indirect effects, basic composition of the model and model modification. The computational methods involved the model identification, the model estimation and the model evaluation. The data in the two examples were "primary school teachers' job satisfaction data" and "relationship of sense of alienation and socioeconomic status among 932 subjects". With the help of SAS software, the structural equation model analysis was conducted on the data in the two examples, and an explanation was given to the SAS output results.

**【Keywords】** Free parameters; Fixed parameters; Model identification; Model modification; Model estimation; Model evaluation

在某些研究中,常涉及诸如智力、性格、社会地位、个人成就等变量,这些变量虽客观存在但却不能直接进行准确测量。为解决此类问题,统计学家们提出了结构方程模型分析。其理论依据是:虽然隐变量无法被直接观察,但可以通过一个或多个观测变量(称为显变量)来表征,即可以通过这些观测变量来研究隐变量之间、隐变量与显变量之间的关系。本文将介绍与结构方程模型分析有关的基本概念和计算方法,并采用 SAS 软件对实例进行分析。

## 1 基本概念

### 1.1 自由参数与固定参数

自由参数是指结构方程模型分析过程中需要估计的参数。固定参数是指结构方程模型分析过程中无需估计的参数,即其取值被设置为某已知常

数。固定参数有以下两种常见的使用场景。第一种是两个变量(显变量或隐变量)之间没有关系,将代表该关系的矩阵元素固定为 0;第二种是设定隐变量的度量单位或尺度,因为隐变量本身没有单位,若不设定单位,则无法进行计算,设定方法有两种,一是将所有隐变量的方差设置为 1(或其他常数),简称固定方差法,二是将隐变量对用来表征它的某个观测变量的因子载荷设置为 1(或其他常数),简称固定负荷法。无论是固定方差法还是固定负荷法,无论方差或负荷固定为多少,模型的标准参数估计结果以及所有的拟合效果统计量都是唯一确定的,区别只在于模型的非标准化参数估计结果。在固定负荷法中,选择将隐变量对应的观测变量的负荷固定为 1,不影响模型的标准参数估计结果。

## 1.2 直接效应与间接效应

直接影响是指一个变量对另一个变量的影响无需通过其他中间变量来传递或连接,影响的数量大小被称为直接效应;间接影响是指一个变量对另一个变量的影响需要通过其他中间变量来传递或连接,影响的数量大小被称为间接效应。

## 1.3 模型的基本构成

结构方程模型由两部分组成:一部分是测量模型,用来描述显变量(即观测变量)与隐变量之间的关联关系,其方程称为测量方程(组);另一部分是结构模型,用来描述隐变量之间的结构关系,其方程称为结构方程(组)<sup>[1-2]</sup>。测量方程见式(1)和式(2)。结构方程见式(3)。

$$X = \Lambda_x \xi + \delta \quad (1)$$

$$Y = \Lambda_y \eta + \varepsilon \quad (2)$$

$$\eta = B\eta + \Gamma\xi + \zeta \quad (3)$$

从形式上看,测量方程式(1)和式(2)实际上就是证实性因子分析模型。其中式(1)是外源变量的测量方程, $X$ 是由 $p$ 个外源观测变量组成的 $p \times 1$ 向量, $\xi$ 是由 $k$ 个外源隐变量组成的 $k \times 1$ 向量, $\Lambda_x$ 是 $X$ 对 $\xi$ 的因子载荷矩阵, $\delta$ 是由 $p$ 个测量误差组成的 $p \times 1$ 向量。式(2)是内生变量的测量方程, $Y$ 是由 $q$ 个内生观测变量组成的 $q \times 1$ 向量, $\eta$ 是由 $m$ 个内生隐变量组成的 $m \times 1$ 向量, $\Lambda_y$ 是 $Y$ 对 $\eta$ 的因子载荷矩阵, $\varepsilon$ 是 $q$ 个测量误差组成的 $q \times 1$ 向量。

式(3)中, $B$ 是 $\eta$ 对 $\eta$ 的结构系数矩阵,反映内生隐变量之间的关系, $\Gamma$ 是 $\xi$ 对 $\eta$ 的结构系数矩阵,反映外源隐变量对内生隐变量的影响, $\zeta$ 是结构残差向量。

对于结构方程模型,有如下假定:①误差项 $\delta$ 、 $\varepsilon$ 、 $\zeta$ 的均值为0;②误差项 $\delta$ 、 $\varepsilon$ 与隐变量 $\xi$ 、 $\eta$ 之间均不相关,且误差项 $\delta$ 、 $\varepsilon$ 之间也不相关(但 $\delta$ 和 $\varepsilon$ 的内部可以相关);③误差项 $\zeta$ 与 $\delta$ 、 $\varepsilon$ 、 $\xi$ 之间均不相关。此外,通常还假定显变量和隐变量都是中心化的,但此条件不是必须的。

## 1.4 模型修正

对初始的设定模型进行评价后,若对其拟合效果不满意,则应对模型进行适当的修正,使其能更好地拟合数据,反映客观现实。模型修正是建立在模型评价基础之上的,其内容包括以下6个方面:查看拟合效果指标、查看拟合残差结果、查看参数估计

结果、查看参数检验结果、查看修正指数结果、查看决定系数大小,具体内容参阅文献[3-4]。

## 2 计算方法

### 2.1 模型识别

模型设定以后,尚不能直接进行数据分析。在此之前,需判断模型在统计上是否具有可操作性,能否得到稳定且合理的结果,即进行模型识别。根据方程的个数和模型中待估计的参数个数,可将结构方程模型分为不可识别的、恰好识别的、过度识别的。其中,不可识别是指待估计的参数个数多于样本中能得出的方程个数,此时包含的信息不充足,参数估计可得到很多解。恰好识别是指待估计的参数个数等于样本中能得出的方程个数,此时可求出参数的估计值,但由于自由度和卡方值均为0,无法检验模型的合理性。过度识别是指待估计的参数个数少于样本中能得出的方程个数,实际上是对待估参数附加了不同的条件,此时可求出参数的估计值并进行检验。所以,结构方程模型所能拟合的模型通常都是过度识别的模型。

由于结构方程模型的复杂性,模型识别问题目前尚无标准的判别方法<sup>[5-6]</sup>。已有的判断模型识别程度的法则包括 $t$ 法则、两步法则、MIMIC法则、递归法则。下面对最常用的 $t$ 法则进行介绍。

在结构方程模型(1)、(2)、(3)中,共有 $(p+q)$ 个显变量,可产生 $(p+q)(p+q+1)/2$ 个不同的方差和协方差。如果理论模型成立,则可得到 $(p+q)(p+q+1)/2$ 个不同的方程。设 $t$ 为模型中未知参数的个数,则模型可识别的一个必要但非充分条件见式(4)。

$$t \leq (p+q)(p+q+1)/2 \quad (4)$$

方程个数与待估计参数个数之差是模型的自由度,记为 $DF$ ,其计算见式(5)。

$$DF = (p+q)(p+q+1)/2 - t \quad (5)$$

在具体使用 $t$ 法则时,有以下标准:①当 $DF < 0$ 时,模型不可识别;②当 $DF = 0$ 时,模型恰好识别;③当 $DF > 0$ 时,模型过度识别。由于模型识别的复杂性和现有识别法则的局限性,在具体操作时,对模型识别问题一般不作深入考量,多交由统计软件来完成,当软件报告模型无法拟合时,才进一步探讨模型无法识别的原因。

### 2.2 参数估计

结构方程模型分析是基于协方差矩阵进行的,

这里的协方差矩阵是指样本协方差矩阵(记为  $S$ ),它是作为总体协方差矩阵(记为  $\Sigma$ )的估计而存在。如果设定的理论模型为真,则基于此模型计算而得的理论协方差矩阵[也称再生协方差矩阵,记为  $\Sigma(\theta)$ ]应等于总体协方差矩阵<sup>[7-8]</sup>。以下介绍  $\Sigma(\theta)$  的求解过程。

完整的结构方程模型分析求解,需要估计 8 个参数矩阵: $\Lambda_X, \Lambda_Y, B, \Gamma, \Phi, \Psi, \Theta_\delta$  和  $\Theta_\varepsilon$ 。其中,  $\Phi$  为  $\xi$  的协方差矩阵,  $\Psi$  为  $\zeta$  的协方差矩阵,  $\Theta_\delta$  和  $\Theta_\varepsilon$  分别是  $\delta$  和  $\varepsilon$  的协方差矩阵。

欲求得全部显变量组成的  $(p+q) \times 1$  个向量  $(X', Y')$  的协方差矩阵,可以先求出  $X$  和  $Y$  的协方差矩阵及它们之间的协方差矩阵。由前文“1.3 模型的基本构成”部分中结构方程模型分析的三个假定可知,应有式(6)和式(7)。

$$E(\xi\delta') = E(\delta\xi') = 0, E(\eta\varepsilon') = E(\varepsilon\eta') = 0 \quad (6)$$

$$E(\delta\varepsilon') = 0, E(\xi\varepsilon') = 0, E(\delta\eta') = 0, E(\xi\zeta') = 0 \quad (7)$$

由式(1)可得式(8)。

$$\Sigma(\theta) = \begin{pmatrix} \sum_{XX}(\theta) & \sum_{XY}(\theta) \\ \sum_{YX}(\theta) & \sum_{YY}(\theta) \end{pmatrix} = \begin{pmatrix} \Lambda_X\Phi\Lambda_X' + \Theta_\delta & \Lambda_X\Phi\Gamma'\tilde{B}'\Lambda_Y' \\ \Lambda_Y\tilde{B}\Gamma\Phi\Lambda_X' & \Lambda_Y\tilde{B}(\Gamma\Phi\Gamma' + \Psi)\tilde{B}'\Lambda_Y' + \Theta_\varepsilon \end{pmatrix} \quad (14)$$

结构方程模型参数估计的基础是方差差异最小化的思想,如果理论模型为真,则  $\Sigma(\theta) = \Sigma$ 。但总体协方差矩阵  $\Sigma$  是未知的,通常用样本协方差矩阵  $S$  代替。根据模型求得的参数,需要使  $S$  和  $\Sigma(\theta)$  之间的差异尽可能小。两者之间的差异可用一个拟合函数来定义,记为  $F(S, \Sigma(\theta))$ 。  $F(S, \Sigma(\theta))$  的定义不同,所对应的参数估计方法也不同,常用的参数估计方法包括非加权最小二乘法、广义最小二乘法、加权最小二乘法以及最大似然函数法。

### 2.3 模型评价

完成设定模型的参数估计后,就得到了假定的初始模型。下一步要做的是对这个初始模型进行整体评价,即利用各种统计量,检验模型是否很好地拟合了样本数据。如果检验结果达到了可接受的程度,说明研究者设定的模型与实际资料相吻合,所建立的理论体系可以解释所研究的现象。如果结果未能达到可接受的程度,则需要对设定模型进行修改和调整。

如何对初始模型进行整体评价,一般而言,好的结构方程模型对应的拟合函数  $F(S, \Sigma(\theta))$  应足够小。在模型可以识别的情况下,结构方程模型整体

$$\sum_{XX}(\theta) = \text{cov}(X) = E(\Lambda_X\xi + \delta)(\Lambda_X\xi + \delta)' \quad (8)$$

$$= \Lambda_X\Phi\Lambda_X' + \Theta_\delta$$

同理,由式(2)可得式(9)。

$$\sum_{YY}(\theta) = \text{cov}(Y) = E(\Lambda_Y\eta + \varepsilon)(\Lambda_Y\eta + \varepsilon)' \quad (9)$$

$$= \Lambda_Y E(\eta\eta') \Lambda_Y' + \Theta_\varepsilon$$

由式(3)可得式(10)。

$$\eta = (I - B)^{-1}(\Gamma\xi + \zeta) = \tilde{B}(\Gamma\xi + \zeta) \quad (10)$$

记  $\tilde{B} = (I - B)^{-1}$ ,其中  $I$  为单位矩阵。这里隐含了一个关于模型的假定,即  $I - B$  是可逆的。由式(10)可得式(11)。

$$E(\eta\eta') = \tilde{B}(\Gamma\Phi\Gamma' + \Psi)\tilde{B}' \quad (11)$$

将式(11)代入式(9),可得式(12)。

$$\sum_{YY}(\theta) = \Lambda_Y\tilde{B}(\Gamma\Phi\Gamma' + \Psi)\tilde{B}'\Lambda_Y' + \Theta_\varepsilon \quad (12)$$

计算  $X, Y$  之间的协方差矩阵,得式(13)。

$$\sum_{YX}(\theta) = E(YX') = E(\Lambda_Y\eta + \varepsilon)(\Lambda_X\xi + \delta)' \quad (13)$$

$$= \Lambda_Y E(\eta\xi') \Lambda_X' = \Lambda_Y\tilde{B}\Gamma\Phi\Lambda_X'$$

据此,可计算出向量  $(X', Y')$  的协方差矩阵,见式(14)。

拟合优度检验的统计假设如下。  $H_0: S = \Sigma(\theta)$ ,即数据被完全拟合;  $H_1: S \neq \Sigma(\theta)$ ,即数据未被完全拟合。评价这个统计假设的统计量有很多,例如绝对拟合统计量(包括 GFI、AGFI、RMR、SRMR、 $\chi^2$ 、RMSEA 和 PCLOSE),相对拟合统计量(包括 NFI、RFI、NNFI、IFI、CFI)以及简约拟合统计量(包括 PGFI、PNFI、AIC、CAIC、SBC、ECVI、CN)。因篇幅所限,前面提及的三类统计量的名称和计算可参阅文献[9]。在前述诸多统计量中,只有少数属于检验统计量(例如  $\chi^2$ ),绝大多数属于非检验统计量,即一般统计量。

## 3 实例与 SAS 实现

### 3.1 问题与数据结构

#### 3.1.1 两个实际问题及数据

【例 1】某研究者根据影响小学教师工作满意度的文献与理论分析,提出了一个关于小学教师成就动机、自尊、专业承诺、工作绩效及工作满意度这五个隐变量的因果模型。随机抽取 100 名小学教师为样本,并测量 11 个观测变量,各观测变量的标准差及相关矩阵见表 1<sup>[10]</sup>。教师工作满意度的因果模型见图 1。

表 1 11 个观测变量之间的相关矩阵

Table 1 Correlation matrix between 11 observational variables

项 目	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
$Y_1$	1.000										
$Y_2$	0.938	1.000									
$Y_3$	0.810	0.805	1.000								
$Y_4$	0.788	0.786	0.970	1.000							
$X_1$	0.157	0.196	-0.139	-0.160	1.000						
$X_2$	0.223	0.261	-0.102	-0.095	0.778	1.000					
$X_3$	0.288	0.322	-0.114	-0.121	0.765	0.834	1.000				
$X_4$	0.223	0.240	-0.169	-0.183	0.575	0.635	0.854	1.000			
$X_5$	0.213	0.252	-0.125	-0.128	0.544	0.579	0.812	0.768	1.000		
$X_6$	0.492	0.497	0.539	0.572	0.340	0.416	0.282	0.051	0.066	1.000	
$X_7$	0.430	0.434	0.533	0.561	0.359	0.395	0.287	0.071	0.075	0.864	1.000

注： $Y_1$ 的标准差为 1.790,  $Y_2$ 的标准差为 1.621,  $Y_3$ 的标准差为 2.203,  $Y_4$ 的标准差为 2.513,  $X_1$ 的标准差为 1.167,  $X_2$ 的标准差为 1.400,  $X_3$ 的标准差为 1.950,  $X_4$ 的标准差为 1.173,  $X_5$ 的标准差为 1.319,  $X_6$ 的标准差为 1.192,  $X_7$ 的标准差为 1.638

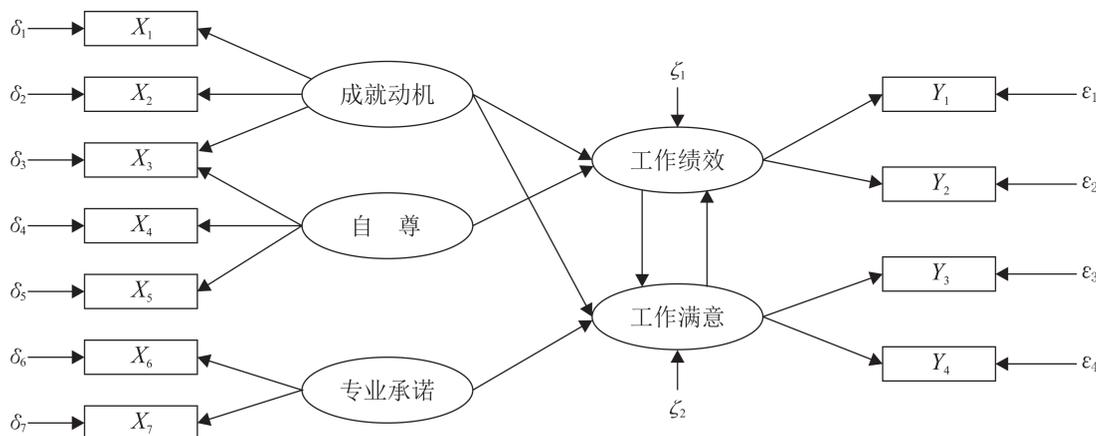


图 1 小学教师工作满意度结构方程模型路径图

Figure 1 Path diagram of the structural equation model for job satisfaction of primary school teachers

【例 2】某研究收集了 932 名受试者在 1967 年和 1971 年的生命无价值感和无力感相关数据, 作为他们疏离感的指标。同时还获得了受试者的教育水平(在学年数)和 Duncan 社会经济指数, 以这两个变量作为受试者的社会经济地位指标。各观测变量之间的协方差矩阵见表 2<sup>[10]</sup>。各变量路径图见图 2。

表 2 6 个观测变量之间的协方差矩阵

Table 2 Covariance matrix between six observational variables

项 目	$X_{1967}$	$Y_{1967}$	$X_{1971}$	$Y_{1971}$	edu	eco
$X_{1967}$	11.834					
$Y_{1967}$	6.947	9.364				
$X_{1971}$	6.819	5.091	12.532			
$Y_{1971}$	4.783	5.028	7.495	9.986		
edu	-3.839	-3.889	-3.841	-3.625	9.610	
eco	-2.190	-1.883	-2.175	-1.878	3.552	4.503

注： $X_{1967}$ 和  $Y_{1967}$  分别代表 1967 年对应的无价值感和无力感； $X_{1971}$ 和  $Y_{1971}$  分别代表 1971 年对应的无价值感和无力感；edu 和 eco 分别代表教育水平和社会经济指标

### 3.1.2 对数据结构的分析

例 1 和例 2 的资料都涉及多个自变量、多个因变量和多个隐变量, 故它们都属于单组设计多自变量、多中介变量(隐变量)和多因变量的定量资料。说明: 在测量方程模型中, 隐变量只能作为自变量, 出现在模型等号右边; 而在结构方程模型中, 有些隐变量可以作为因变量, 出现在模型等号左边, 有些隐变量可以作为自变量, 出现在模型等号右边。

### 3.2 用 SAS 实现统计分析

#### 3.2.1 分析例 1 的资料

设所需要的 SAS 程序如下<sup>[9]</sup>：  

```
data a1(type=corr);
_type_='corr';
input _name_ $ Y1-Y4 X1-X7;
if _n_=1 then _type_='std';
else _type_='corr';
cards;
```

```

std 1. 78997 1. 62142 2. 20341 2. 512966
1. 16748 1. 40000 1. 95013 1. 17303 1. 31947
1. 19248 1. 63829
Y1 1. 00000 . . . . .
Y2 0. 93788 1. 00000 . . . . .
Y3 0. 81084 0. 80473 1. 00000 . . . . .
Y4 0. 78810 0. 78585 0. 97037
1. 00000 . . . . .
X1 0. 15744 0. 19599 -0. 13878 -0. 16054
1. 00000 . . . . .
X2 0. 22307 0. 26079 -0. 10244 -0. 09522
0. 77762 1. 00000 . . . . .
X3 0. 28820 0. 32227 -0. 11380 -0. 12060
0. 76513 0. 83364 1. 00000 . . . . .
X4 0. 22289 0. 23975 -0. 16946 -0. 18285
0. 57540 0. 63511 0. 85375 1. 00000 . . . . .
X5 0. 21255 0. 25194 -0. 12486 -0. 12817
0. 54400 0. 57924 0. 81224 0. 76820 1. 00000 . . . . .
X6 0. 49192 0. 49651 0. 53891 0. 57197
0. 34047 0. 41570 0. 28166 0. 05076 0. 06610
1. 00000 . . . . .
X7 0. 42967 0. 43443 0. 53271 0. 56085

```

```

0. 35866 0. 39545 0. 28702 0. 07077 0. 07494
0. 86404 1. 00000
;
run;
proc calis cov nobs=100 res;
lineqs X1=f1+e1,
X2=a2 f1+e2,
X3=a3 f1+a4 f2+e3,
X4= f2+e4,
X5=a6 f2+e5,
X6= f3+e6,
X7=a8 f3+e7,
Y1= f4+e8,
Y2= a10 f4+e9,
Y3= f5+e10,
Y4= a12 f5+e11,
f4=b1 f1+b2 f2+b3 f5+d1,
f5=b4 f1+b5 f3+b6 f4+d2;
std e1-e11 d1 d2 f1 f2 f3=16*var;
cov f1 f2=cf1f2, f1 f3=cf1f3, f2 f3=cf2f3, d1 d2=
cd1d2;
run;

```

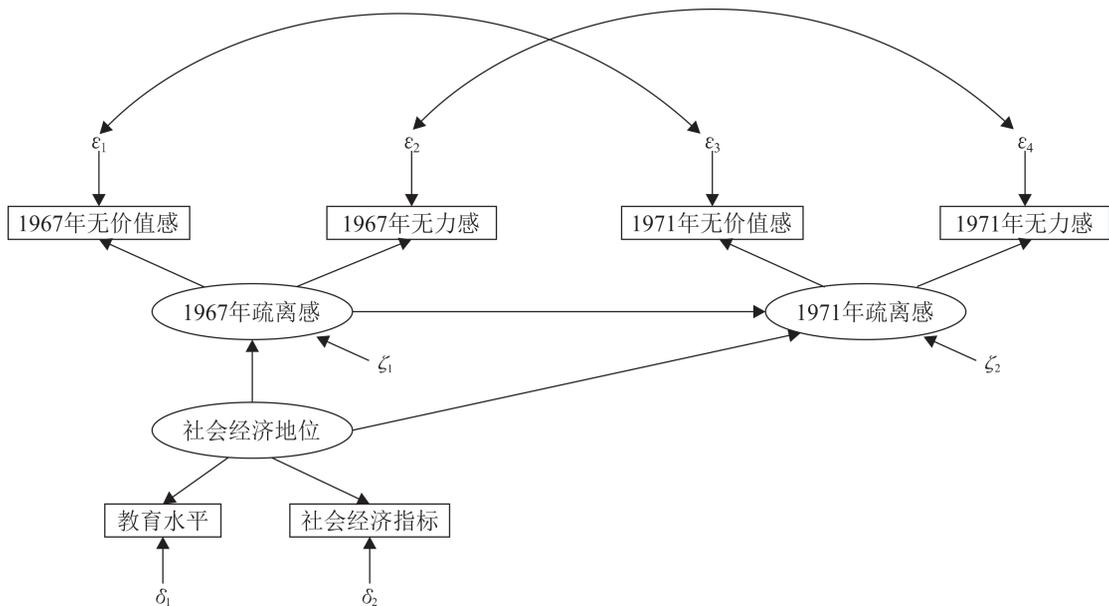


图2 疏离感结构方程模型路径图  
Figure 2 Path diagram of the structural equation model for sense of alienation

【SAS输出结果及解释】线性方程中,标准回归系数的估计结果见表3。结构方程模型中,12个参数与0之间差异均有统计学意义,只有系数 $b_1$ 与0之间的差异无统计学意义(可以在SAS程序中将此系数设置为0,重新运行SAS程序),说明通过分析得到的结构方程模型基本成立。

外生变量方差的标准化的估计值见表4。外生变量中协方差的标准化的估计值见表5。由表5可知,两个隐变量 $f_2$ 与 $f_3$ 之间的协方差的估计值为0.058,此值与0之间差异无统计学意义,即可以认为这两个隐变量是互相独立的。

基于表3、表4、表5的计算结果,可以写出包含

13 个方程的结构方程组, 还可将计算出来的系数填入图 1 相应位置上, 呈现带系数的结构方程模型路径图。因篇幅所限, 此部分内容从略。

表 3 线性方程中的标准化效应  
Table 3 Standardization effects in linear equation

变 量	预测变量	参 数	估 计	标准误	t	Pr> t
X <sub>1</sub>	f <sub>1</sub>		0.707	-	-	-
X <sub>2</sub>	f <sub>1</sub>	a <sub>2</sub>	0.847	0.022	39.196	<0.001
X <sub>3</sub>	f <sub>1</sub>	a <sub>3</sub>	0.452	0.090	5.028	<0.001
X <sub>3</sub>	f <sub>2</sub>	a <sub>4</sub>	0.552	0.084	6.599	<0.001
X <sub>4</sub>	f <sub>2</sub>		0.707	-	-	-
X <sub>5</sub>	f <sub>2</sub>	a <sub>6</sub>	0.827	0.025	33.094	<0.001
X <sub>6</sub>	f <sub>3</sub>		0.707	-	-	-
X <sub>7</sub>	f <sub>3</sub>	a <sub>8</sub>	0.891	0.015	58.677	<0.001
Y <sub>1</sub>	f <sub>4</sub>		0.932	0.010	90.218	<0.001
Y <sub>2</sub>	f <sub>4</sub>	a <sub>10</sub>	0.920	0.012	74.872	<0.001
Y <sub>3</sub>	f <sub>5</sub>		0.955	0.007	141.700	<0.001
Y <sub>4</sub>	f <sub>5</sub>	a <sub>12</sub>	0.965	0.005	183.700	<0.001
f <sub>4</sub>	f <sub>1</sub>	b <sub>1</sub>	0.131	0.089	1.472	0.141
f <sub>4</sub>	f <sub>2</sub>	b <sub>2</sub>	0.219	0.088	2.485	0.013
f <sub>4</sub>	f <sub>5</sub>	b <sub>3</sub>	0.660	0.051	12.987	<0.001
f <sub>5</sub>	f <sub>1</sub>	b <sub>4</sub>	-0.450	0.064	-7.004	<0.001
f <sub>5</sub>	f <sub>3</sub>	b <sub>5</sub>	0.554	0.074	7.489	<0.001
f <sub>5</sub>	f <sub>4</sub>	b <sub>6</sub>	0.442	0.062	7.093	<0.001

表 4 外生变量方差的标准化结果  
Table 4 Normalized results of the variance of exogenous variables

变量类型	变 量	参 数	估 计	标准误	t	Pr> t
误差	e <sub>1</sub>	var	0.500	-	-	-
	e <sub>2</sub>	var	0.282	0.037	7.711	<0.001
	e <sub>3</sub>	var	0.159	0.020	8.067	<0.001
	e <sub>4</sub>	var	0.500	-	-	-
	e <sub>5</sub>	var	0.316	0.041	7.652	<0.001
	e <sub>6</sub>	var	0.500	-	-	-
	e <sub>7</sub>	var	0.205	0.027	7.584	<0.001
	e <sub>8</sub>	var	0.131	0.019	6.807	<0.001
	e <sub>9</sub>	var	0.153	0.023	6.787	<0.001
	e <sub>10</sub>	var	0.088	0.013	6.868	<0.001
	e <sub>11</sub>	var	0.069	0.010	6.848	<0.001
扰动	d <sub>1</sub>	var	0.151	0.026	5.915	<0.001
	d <sub>2</sub>	var	0.097	0.015	6.261	<0.001
隐藏	f <sub>1</sub>	var	1.000	-	-	-
	f <sub>2</sub>	var	1.000	-	-	-
	f <sub>3</sub>	var	1.000	-	-	-

表 5 外生变量中协方差的标准化结果  
Table 5 Normalized results of covariance in exogenous variables

Var1	Var2	参 数	估 计	标准误	t	Pr> t
f <sub>1</sub>	f <sub>2</sub>	cf <sub>12</sub>	0.666	0.085	7.859	<0.001
f <sub>1</sub>	f <sub>3</sub>	cf <sub>13</sub>	0.404	0.105	3.851	0.001
f <sub>2</sub>	f <sub>3</sub>	cf <sub>23</sub>	0.058	0.120	0.483	0.629
d <sub>1</sub>	d <sub>2</sub>	cd <sub>12</sub>	0.142	0.037	3.891	<0.001

3.2.2 分析例 2 的资料

设所需要的 SAS 程序如下<sup>[9]</sup>:

```

data a2(type=cov);
_type_='cov';
input _name_ $ x1967 y1967 x1971 y1971 edu eco;
datalines;
X1967 11.834 . . . . .
Y1967 6.947 9.364 . . . .
X1971 6.819 5.091 12.532 . . .
Y1971 4.783 5.028 7.495 9.986 . .
edu -3.839 -3.889 -3.841 -3.625 9.610
eco -2.190 -1.883 -2.175 -1.878 3.552 4.503
;
run;
proc calis cov nobs=932 res toteff;
lineqs X1967=f1+e1,
Y1967=a2 f1+e2,
X1971=f2+e3,
Y1971=a4 f2+e4,
edu=f3+e5,
eco=a6 f3+e6,
f1=b1 f3+d1,
f2=b2 f1+b3 f3+d2;
std e1-e6 d1 d2 f3=9*var;
cov e1 e3=ce1e3,e2 e4=ce2e4;
run;

```

【SAS 输出结果及解释】线性方程中, 标准回归系数的估计结果见表 6。

表 6 线性方程中的标准化效应  
Table 6 Standardization effects in linear equation

变 量	预测变量	参 数	估 计	标准误	t	Pr> t
X <sub>1967</sub>	f <sub>1</sub>	-	0.799	0.008	94.986	<0.001
Y <sub>1967</sub>	f <sub>1</sub>	a <sub>2</sub>	0.762	0.012	61.246	<0.001
X <sub>1971</sub>	f <sub>2</sub>	-	0.819	0.007	110.700	<0.001
Y <sub>1971</sub>	f <sub>2</sub>	a <sub>4</sub>	0.776	0.012	66.414	<0.001
edu	f <sub>3</sub>	-	0.707	-	-	-
eco	f <sub>3</sub>	a <sub>6</sub>	0.513	0.027	19.164	<0.001
f <sub>1</sub>	f <sub>3</sub>	b <sub>1</sub>	-0.659	0.025	-26.239	<0.001
f <sub>2</sub>	f <sub>1</sub>	b <sub>2</sub>	0.522	0.055	9.437	<0.001
f <sub>2</sub>	f <sub>3</sub>	b <sub>3</sub>	-0.252	0.062	-4.042	<0.001

由表 6 可知, 结构方程模型中 8 个参数与 0 之间差异均有统计学意义, 说明通过分析得到的结构方程模型基本成立。外生变量方差的标准化结果的估计值见表 7。外生变量中协方差的标准化结果的估计值见表 8。由表 8 可知, 两对隐变量各自的协方差的估计值均有统计学意义, 可以认为这两对隐变量各自之间是互相独立的。基于表 6、表 7、表 8 计算结果, 可以写出含 8 个方程的结构方程组, 还可将计

算出的系数填入图 2 相应位置上,呈现带系数的结构方程模型路径图。因篇幅所限,内容从略。

表 7 外生变量方差的标准化结果

Table 7 Normalized results of the variance of exogenous variables

变量类型	变 量	参 数	估 计	标准误	t	Pr> t
误差	$e_1$	var	0.362	0.013	26.934	<0.001
	$e_2$	var	0.419	0.019	22.122	<0.001
	$e_3$	var	0.330	0.012	27.208	<0.001
	$e_4$	var	0.397	0.018	21.901	<0.001
	$e_5$	var	0.500	-	-	-
	$e_6$	var	0.737	0.027	26.818	<0.001
扰动	$d_1$	var	0.567	0.033	17.187	<0.001
	$d_2$	var	0.492	0.027	18.241	<0.001
隐藏	$f_3$	var	1.000	-	-	-

表 8 外生变量中协方差的标准化结果

Table 8 Normalized results of covariance in exogenous variables

Var1	Var2	参 数	估 计	标准误	t	Pr> t
$e_1$	$e_3$	$ce_1e_3$	0.097	0.017	5.608	<0.001
$e_2$	$e_4$	$ce_2e_4$	0.098	0.018	5.476	<0.001

自变量  $f_1$ 、 $f_2$ 、 $f_3$  对所有因变量  $X_{1967}$ 、 $Y_{1967}$ 、 $X_{1971}$ 、 $Y_{1971}$ 、edu、eco、 $f_1$ 、 $f_2$  的直接效应、间接效应和总效应的计算结果很多,因篇幅所限,此部分内容从略。

## 4 讨论与小结

### 4.1 讨论

一个好的结构方程模型通常应满足以下条件:①模型整体上较好地拟合了样本数据,主要拟合指标基本满足要求;②因子载荷和结构系数的估计值与 0 的差异有统计学意义,且符号、大小均与实际情况相符;③固定参数的修正指标值都不大,一般要求小于 3.84;④标准化拟合残差比较小,一般要求小于 1.96;⑤测量方程与结构方程的决定系数  $R^2$  一般要求大于 0.50。在实际应用中,有时并非以上条件都能满足,需根据数据和实际情况选择拟合相应模型。

根据待比较的两模型中所含参数的关系,可将模型之间的关系分为嵌套模型和非嵌套模型。嵌套模型:假设有 A 和 B 两个模型,其中模型 A 包含参数较少且其所包含的参数是模型 B 所包含的参数的子集,此情形称为模型 A 嵌套于模型 B。嵌套模型拟合效果的比较,最常用的方法是  $\chi^2$  检验。根据两嵌套模型拟合  $\chi^2$  值的差值  $\Delta\chi^2$  及相应自由度的差值  $\Delta df$  进行假设检验。因  $\Delta\chi^2$  受样本量的影响较大,故不能以  $\Delta\chi^2$  为唯一标准。非嵌套模型是两个模型所包含的参数无法完全包含,比较非嵌套模型拟合效果,可通过信息准则来实现,同时也应比较两模型其他拟

合效果评价统计量的差异。

### 4.2 小结

本文介绍了与结构方程模型分析有关的基本概念、计算方法、两个实例以及使用 SAS 实现计算的方法。基本概念包括自由参数与固定参数、直接效应与间接效应、模型的基本构成和模型修正;计算方法涉及模型识别、模型估计和模型评价;借助 SAS 软件,对两个实例中的数据分别进行了结构方程模型分析,并对 SAS 输出结果做出了解释。

## 参考文献

- [1] 李卫东. 应用多元统计分析[M]. 北京: 北京大学出版社, 2008: 351-370.  
Li WD. Applied multivariate statistical analysis [M]. Beijing: Peking University Press, 2008: 351-370.
- [2] 何晓群. 多元统计分析[M]. 2 版. 北京: 中国人民大学出版社, 2008: 329-349.  
He XQ. Multivariate statistical analysis [M]. 2<sup>nd</sup> edition. Beijing: China Renmin University Press, 2008: 329-349.
- [3] 万崇华, 罗家洪. 高级医学统计学[M]. 北京: 科学出版社, 2014: 218-234.  
Wan CH, Luo JH. Advanced medical statistics [M]. Beijing: Science Press, 2014: 218-234.
- [4] 王建华. 实用医学科研方法[M]. 北京: 人民卫生出版社, 2003: 486-512.  
Wang JH. Practical medical research methods [M]. Beijing: People's Medical Publishing House, 2003: 486-512.
- [5] 张岩波. 潜变量分析[M]. 北京: 高等教育出版社, 2009: 83-112.  
Zhang YB. Latent variables analysis [M]. Beijing: Higher Education Press, 2009: 83-112.
- [6] Armitage P, Colton T. Encyclopedia of biostatistics [M]. 2<sup>nd</sup> edition. New York: John Wiley & Sons, 2005: 5704-5712.
- [7] 易丹辉. 结构方程模型 方法与应用[M]. 北京: 中国人民大学出版社, 2008: 42-164.  
Yi DH. Structural equation modeling methods and applications [M]. Beijing: China Renmin University Press, 2008: 42-164.
- [8] Lattin JM, Carroll JD, Green PE. 多元统计分析[M]. 北京: 机械工业出版社, 2003: 352-385.  
Lattin JM, Carroll JD, Green PE. Analyzing multivariate data [M]. Beijing: China Machine Press, 2003: 352-385.
- [9] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 1383-3062.
- [10] 胡良平. 面向问题的统计学: (3) 试验设计与多元统计分析 [M]. 北京: 人民卫生出版社, 2012: 165-213.  
Hu LP. Problem-oriented statistics: (3) experimental design and multivariate statistical analysis [M]. Beijing: People's Medical Publishing House, 2012: 165-213.

(收稿日期: 2023-09-25)

(本文编辑: 吴俊林)