

· 主编特邀(Editor-In-Chief Invited) ·

编者按:

中介效应分析在许多领域都有广泛应用,因为它可以分析变量之间影响的过程和机制,相对于回归分析,可以得到比较深入的结果。虽然中介分析不能肯定地说“证实”了什么,但可以帮助我们支持某种理论而排除其竞争的理论。温忠麟教授及其合作者有关中介分析方法的研究,引领和推动了中介分析在国内的应用,他们2004年发表在《心理学报》上的论文,在Google学术搜索上可以查到被上千篇论文引用。最近10年,中介效应分析方法和模型都有了许多发展,本文作者做了相当全面的综述,并尝试澄清一些有争议的问题。这些看似不大的争议,在统计上还是很重要的,会影响检验流程。作者还就如何分析因果关系,提出了理论分析的若干思路。本文对中介效应分析的逻辑和统计方法有相当透彻的理解,在中介分析的立论、建模、检验和解释方面有很高的参考价值。本刊希望通过温忠麟和叶宝娟的这篇文章,提高应用工作者涉及中介分析的论文质量。

(本文责任编辑:侯杰泰)

## 中介效应分析:方法和模型发展\*

温忠麟<sup>1,2</sup> 叶宝娟<sup>3</sup>

(<sup>1</sup> 华南师范大学心理应用研究中心/心理学院, 广州 510631)  
(<sup>2</sup> 香港考试及评核局, 香港) (<sup>3</sup> 江西师范大学心理学院, 南昌 330022)

**摘要** 在心理学和其他社科研究领域,大量实证文章建立中介效应模型,以分析自变量对因变量的影响过程和作用机制。检验中介效应最流行的方法是Baron和Kenny的逐步法,但近年来不断受到批评和质疑,有人甚至呼吁停止使用其中的依次检验,改用目前普遍认为比较好的Bootstrap法直接检验系数乘积。本文对相关的议题做了辨析,并讨论了中介分析中建立因果关系的方法。综合新近的研究成果,总结出一个中介效应分析流程,并分别给出显变量和潜变量Mplus程序。最后介绍了中介效应模型的发展。

**关键词** 中介效应; 间接效应; 逐步法; Bootstrap法; 因果

**分类号** B841

近年来,在心理学和其他社科研究领域,大量实证文章建立中介效应(mediation effect)模型进行分析。例如, Rucker, Preacher, Tormala 和 Petty (2011)统计发现,2005至2009年发表在Journal of

Personality and Social Psychology (JPSP) 和 Personality and Social Psychology Bulletin (PSPB) 上的文章,分别有59%和65%使用了中介检验。国内心理学期刊上有关中介效应文章所占的比例,也可以用引人注目来形容。中介效应模型可以分析自变量对因变量影响的过程和作用机制,相比单纯分析自变量对因变量影响同类研究,中介分析不仅方法上有进步,而且往往能得到更多更深入的结果,这可以解释为什么中介分析受到重视。

检验中介效应最流行的方法是Baron和

收稿日期: 2014-02-06

\* 国家自然科学基金(31271116)、教育部人文社会科学重点研究基地项目(11JJD190005)、中国博士后科学基金项目(2013M540535)、教育部人文社会科学研究青年基金项目(13YJC190029)和江西省博士后科研择优资助项目(2013KY08)资助。

通讯作者: 温忠麟, E-mail: wenzl@scnu.edu.cn

Kenny (1986)的逐步法(causal steps approach)。但是,近年来逐步法受到几乎是一边倒的批评和质疑(例如, Edwards & Lambert, 2007; Hayes, 2009; Spencer, Zanna, & Fong, 2005; Zhao, Lynch, & Chen, 2010)。有人甚至呼吁停止使用依次(piecemeal)检验(逐步法中的一个步骤),改用目前普遍认为比较好的 Bootstrap 法直接检验系数乘积的显著性(如 Zhao et al., 2010)。本文在介绍各种中介分析方法后,对这些批评逐一做了辨析,特别谈到中介分析如何从理论上建立因果关系;综合新近的中介效应方法研究成果,总结出一个中介效应分析流程;还介绍了中介效应模型的新近发展。

## 1 中介效应模型及检验方法介绍

考虑自变量  $X$  对因变量  $Y$  的影响,如果  $X$  通过影响变量  $M$  而对  $Y$  产生影响,则称  $M$  为中介变量。例如,“家庭经济地位”影响“家庭功能”,进而影响“青少年疏离感”。为了行文简便,避免在回归方程中出现与方法讨论无关的截距项,假设所有变量都已经中心化(即将数据减去样本均值,中心化数据的均值为 0)或者标准化(均值为 0, 标准差为 1),可用下列回归方程来描述变量之间的关系(图 1 是相应的路径图):

$$Y = cX + e_1 \quad (1)$$

$$M = aX + e_2 \quad (2)$$

$$Y = c'X + bM + e_3 \quad (3)$$

其中方程(1)的系数  $c$  为自变量  $X$  对因变量  $Y$  的总效应;方程(2)的系数  $a$  为自变量  $X$  对中介变量  $M$  的效应;方程(3)的系数  $b$  是在控制了自变量  $X$  的影响后,中介变量  $M$  对因变量  $Y$  的效应;系数  $c'$

是在控制了中介变量  $M$  的影响后,自变量  $X$  对因变量  $Y$  的直接效应; $e_1 \sim e_3$  是回归残差。对于这样的简单中介模型,中介效应等于间接效应(indirect effect),即等于系数乘积  $ab$ ,它与总效应和直接效应有下面关系(MacKinnon, Warsi, & Dwyer, 1995):

$$c = c' + ab \quad (4)$$

检验中介效应最常用的方法是逐步检验回归系数(Baron & Kenny, 1986; Judd & Kenny, 1981; 温忠麟, 张雷, 侯杰泰, 刘红云, 2004),即通常说的逐步法:(i)检验方程(1)的系数  $c$  (即检验  $H_0: c = 0$ );(ii)依次检验方程(2)的系数  $a$  (即检验  $H_0: a = 0$ )和方程(3)的系数  $b$  (即检验  $H_0: b = 0$ ),有文献称之为联合显著性检验(test of joint significance, Hayes, 2009)。如果(i)系数  $c$  显著, (ii)系数  $a$  和  $b$  都显著,则中介效应显著。完全中介过程还要加上: (iii)方程(3)的系数  $c'$  不显著。

上述 Baron 和 Kenny (1986)的逐步法,第一步检验的是  $X$  对  $Y$  的总效应;第二步实际上是检验系数乘积的显著性(即检验  $H_0: ab = 0$ ),通过依次检验系数  $a$  和  $b$  来间接进行;第三步检验用来区分完全中介还是部分中介。这三步其实是可以分开进行的。区分每一步的目的对理解和讨论逐步法很重要。特别地,我们将检验系数乘积(即检验  $H_0: ab = 0$ )的依次检验与逐步法区分开来,前者是后者的一个步骤。文献提到逐步法时,有的是指全部步骤,有的却是指依次检验,不小心的读者容易引起混淆。

系数乘积的检验(即检验  $H_0: ab = 0$ )是中介效应检验的核心,下面先集中讨论其检验方法。依

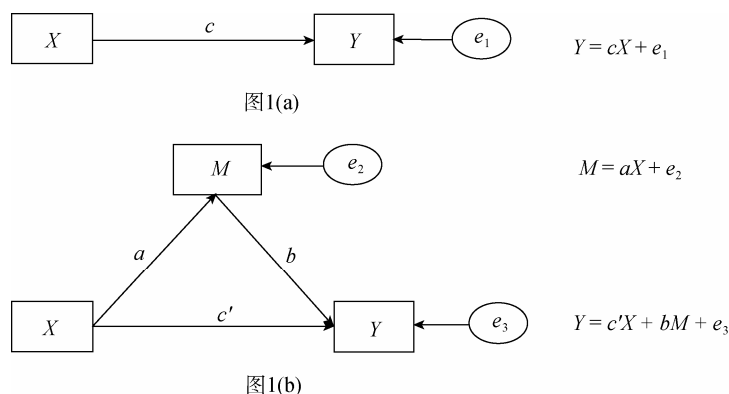


图1 中介模型示意图

次检验是对系数乘积的间接检验,想法很直观,如果检验结果是 $a \neq 0$ 且 $b \neq 0$ ,就可以推出 $ab \neq 0$ 。这个推理在代数上没有问题,但在统计检验上如何呢?模拟研究发现,用依次检验来检验 $H_0: ab = 0$ ,第一类错误率较低,低于设定的显著性水平(如0.05)(MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; 温忠麟等, 2004)。这就是说,如果依次检验结果 $a$ 和 $b$ 都显著,已经足够支持所要的结果,即 $ab$ 显著。但依次检验的检验力(power)也较低,即系数乘积实际上显著而依次检验比较容易得出不显著的结论(Fritz & MacKinnon, 2007; MacKinnon et al., 2002)。

检验系数乘积更多的是直接针对假设 $H_0: ab = 0$ 提出的检验方法。Sobel (1982)法就是比较有名的一种。检验统计量为 $z = \hat{a}\hat{b} / s_{ab}$ ,其中 $\hat{a}$ 和 $\hat{b}$ 分别是 $a$ 和 $b$ 的估计, $s_{ab} = \sqrt{\hat{a}^2 s_b^2 + \hat{b}^2 s_a^2}$ 是 $\hat{a}\hat{b}$ 的标准误, $s_a$ 和 $s_b$ 分别是 $\hat{a}$ 和 $\hat{b}$ 的标准误。模拟研究发现,Sobel法的检验力高于依次检验(MacKinnon et al., 2002; 温忠麟等, 2004)。但这个检验统计量的推导需要假设 $\hat{a}\hat{b}$ 服从正态分布,就算其中每一个系数都是正态分布,其乘积通常也不是正态的,因而上面标准误 $s_{ab}$ 的计算只是近似的,可能很不准确。这样,Sobel检验的局限性是很明显的(方杰, 张敏强, 2012; Hayes, 2009; MacKinnon, 2008; MacKinnon, Lockwood, & Williams, 2004)。

试图用来替代Sobel法直接检验 $H_0: ab = 0$ 的方法至少有三类(方杰, 张敏强, 2012),包括乘积分布法、Bootstrap法和马尔科夫链蒙特卡罗(MCMC)法。乘积分布法默认 $\hat{a}\hat{b}$ 分布是两个正态变量的乘积分布,根据乘积分布构建临界值进行检验和区间估计(Fritz & MacKinnon, 2007; MacKinnon et al., 2004; Tofighi & MacKinnon, 2011)。

Bootstrap法是一种从样本中重复取样的方法,前提条件是样本能够代表总体(当然这也是通常取样进行统计推论的要求)。Bootstrap法有多种取样方案,其中一种简单的方案是从给定的样本中有放回地重复取样以产生出许多样本,即将原始样本当作Bootstrap总体,从这个Bootstrap总体中重复取样以得到类似于原始样本的Bootstrap样本(Wen, Marsh, & Hau, 2010)。例如,将一个容量为500的样本当作Bootstrap总体,从中有放回地

重复取样,可以得到一个Bootstrap样本(容量还是500)。类似的可以得到很多Bootstrap样本(比如1000个),对这1000个Bootstrap样本,可以得到1000个系数乘积的估计值,其全体记为 $\{\hat{a}\hat{b}\}$ 。将它们按数值从小到大排序,其中第2.5百分位点和第97.5百分位点就构成 $ab$ 的一个置信度为95%的置信区间,据此就可以进行检验了:如果置信区间不包含0,则系数乘积显著(方杰, 张敏强, 2012; Preacher & Hayes, 2008; Preacher, Rucker, & Hayes, 2007; 温忠麟, 刘红云, 侯杰泰, 2012)。这样的检验方法称为非参数百分位Bootstrap法,检验力高于Sobel检验(Fritz & MacKinnon, 2007; MacKinnon et al., 2004)。检验力更高的是使用偏差校正后的置信区间,即所谓的偏差校正的非参数百分位Bootstrap法(Edwards & Lambert, 2007; 方杰, 张敏强, 2012; Fritz & MacKinnon, 2007; MacKinnon, 2008; Preacher & Hayes, 2008; Taylor, MacKinnon, & Tein, 2008; 温忠麟等, 2012)。在Bootstrap法前面冠以“非参数”,是因为所论的Bootstrap法不涉及总体分布及其参数(因而不要求正态假设),利用样本所推导的经验分布代替总体分布,属于非参数方法。

在中介效应分析中,马尔科夫链蒙特卡罗(MCMC)法是一种贝叶斯统计方法(Ntzoufras, 2009; Yuan & MacKinnon, 2009)。具体一点说,MCMC法是在贝叶斯理论框架下,将马尔科夫链过程引入到蒙特卡罗模拟中,实现抽样分布随模拟的进行而改变的动态模拟(方杰, 张敏强, 2012)。MCMC法需要较多的统计知识和复杂的算法,还会涉及引起争论的先验分布(prior distribution)问题,这里不拟多说。

研究发现,上述三类方法中,用偏差校正的非参数百分位Bootstrap法或者先验信息的MCMC法计算系数乘积的置信区间比Sobel法得到的置信区间更精确,有更高的检验力(方杰, 张敏强, 2012; Hayes & Scharkow, 2013; MacKinnon, 2008; Preacher & Hayes, 2004; Yuan & MacKinnon, 2009)。常用的统计软件Mplus (Muthén & Muthén, 2012)已经有简单的指令,调用Bootstrap法和MCMC法计算系数乘积的置信区间,实现系数乘积的Bootstrap法检验和MCMC法检验。不熟悉Bootstrap法和MCMC法的读者,可以将其和最小二乘法、极大似然法那样看待,利用统计软件提

供的功能实现计算。但多数方法学文章都只推荐 Bootstrap 法(例如 Biesanz, Falk, & Savalei, 2010; Cheung & Lau, 2008; Fritz, Taylor, & MacKinnon, 2012; Hayes & Scharkow, 2013; MacKinnon et al., 2004; Pituch & Stapleton, 2008; Pituch, Stapleton, & Kang, 2006; Taylor et al., 2008), 而且 MCMC 法的先验分布通常也无法得到, 所以到目前为止, Bootstrap 法是公认的可以取代 Sobel 法而直接检验系数乘积的方法。不过, 偏差校正的非参数百分位 Bootstrap 法在某些条件下的第一类错误率会超过设定的显著性水平(如 0.05) (方杰, 张敏强, 2012; Fritz & MacKinnon, 2007; MacKinnon et al., 2004), 而非参数百分位 Bootstrap 法没有这个问题(Fritz et al., 2012)。

因为  $ab = c - c'$ , 所以检验间接效应也可以通过检验  $H_0: c - c' = 0$  来进行(Clogg, Petkova, & Shihadeh, 1992; Freedman & Schatzkin, 1992), 称为系数差异检验法, 以区别上面讨论的系数乘积检验法。但因为系数差异检验法的第一类错误率明显高于系数乘积检验法(可能远高于 0.05), 所以它们很早就输给了系数乘积检验法(MacKinnon et al., 2002; 温忠麟等, 2004), 后面不提。

总结一下, 检验间接效应可以分成两类, 一类是检验  $H_0: ab = 0$ , 另一类是检验  $H_0: c - c' = 0$ 。检验  $H_0: ab = 0$  又可以分成间接检验和直接检验两类。依次检验是间接检验  $H_0: ab = 0$ , 而 Sobel 检验、Bootstrap 法、MCMC 法等是直接检验  $H_0: ab = 0$ 。直接检验  $H_0: ab = 0$  比较好的方法是偏差校正的非参数百分位 Bootstrap 法, 应当取代 Sobel 检验。研究者如果在乎检验的第一类错误率, 使用非参数百分位 Bootstrap 法比较妥当(Fritz et al., 2012; Hayes & Scharkow, 2013)。但很多文献想当然就认为依次检验也应当让位给 Bootstrap 法, 至少从检验的角度来说是没有道理的, 下一节会详细讨论。

## 2 对 Baron 和 Kenny 逐步法的质疑和辨析

Baron 和 Kenny (1986)的逐步法的每一步, 都有人提出批评和质疑, 最彻底否定的要数 Zhao 等人(2010)的文章。下面看看这些批评要点, 逐一辨析。

### 2.1 依次检验还有用吗?

就间接效应的检验而言, 依次检验方程(2)

的系数  $a$  和方程(3)的系数  $b$ , 是最多人使用的方法。尽管早有方法文章已经建议使用 Bootstrap 法直接检验系数乘积, 但很多应用工作者还是照用依次检验。依次检验受欢迎的原因是方法简单, 容易理解和解释。方法学者不推荐也可以理解, 因为依次检验的检验力在各种方法中是最低的(Fritz & MacKinnon, 2007; Hay, 2009; MacKinnon et al., 2002)。就是说, 依次检验比较不容易检验到中介效应显著。但如果研究者用依次检验已经得到显著的结果, 检验力低的问题对其而言就不是问题! 此时, 依次检验的结果甚至好过 Bootstrap 法的结果, 奇怪的是似乎很少人理解到的这一点, 下面给出理由。

设想甲乙两人用同一组数据检验系数乘积(即检验  $H_0: ab = 0$ ), 甲做依次检验结果显著, 乙用 Bootstrap 法检验结果也显著, 甲的结果更好: (1)看着甲的结果, 我们几乎可以肯定乙的检验结果也是显著, 因为 Bootstrap 法的检验力高于 Sobel 检验(Fritz & MacKinnon, 2007; MacKinnon et al., 2004), 后者又高于依次检验(MacKinnon et al., 2002); 而看着乙的结果却不能判断甲的检验结果是否显著(因为依次检验比较不容易得到显著的结果)。(2)看着甲的结果, 我们知道  $X$  显著影响  $M$ , 而且  $M$  显著影响  $Y$ , 推论是间接效应显著; 看着乙的结果, 我们只知道间接效应是显著的, 但不知道  $X$  是否显著影响  $M$ , 也不知道  $M$  是否显著影响  $Y$ 。(3)当检验结果是显著的时候, 可能的错误只是第一类的。检验力比较高的方法, 通常也有比较高的第一类错误率。前面说过, 当设定显著性水平 0.05 时, 依次检验的第一类错误率低于 0.05(MacKinnon et al., 2002; 温忠麟等, 2004), 而 Bootstrap 法的第一类错误率可能会超过 0.05 (Fritz & MacKinnon, 2007; MacKinnon et al., 2004)。一般而言, 甲的错误率不会比乙的错误率高。综合上面三点可以说, 如果检验结果都显著, 依次检验的结果强于 Bootstrap 法检验结果。

但要注意, 上面说的是已经得到“显著”结果的情况(就像大多数投稿中的结果那样), 才是依次检验的结果强于 Bootstrap 法检验结果。但并不意味着要推荐依次检验、不推荐 Bootstrap 法, 因为两者的检验力高低不同。我们推荐的检验方法和步骤详见第3节。



## 2.2 要先检验总效应吗?

逐步法中第一步是检验方程(1)的系数  $c$ , 有些人认为没有必要(例如, MacKinnon, Krull, & Lockwood, 2000; Zhao et al., 2010)。他们的论据是, 间接效应( $ab$ )的符号可能和直接效应( $c'$ )的符号相反, 使得总效应( $c$ )不显著, 但中介效应还是存在; 也可能存在两条中介路径, 其间接效应大小相近但符号相反, 使得总效应不显著。就是说, 即使总效应不显著, 间接效应还是可能存在。

这里其实涉及两个问题, 一是要不要检验系数  $c$ ? 二是中介效应要不要以系数  $c$  显著为前提条件? 第一个问题的答案是肯定的, 因为研究者肯定会关心  $X$  是否显著影响  $Y$ 。对于特定的两个变量  $X$  和  $Y$ , 如果根据理论、经验或者与他们关系密切的第三个变量  $M$ , 都无法设想  $X$  和  $Y$  之间有关系的话, 还会去研究  $X$  如何影响  $Y$  吗? 文章将如何立论? 所以说, 研究者肯定会关心  $X$  和  $Y$  之间关系。

对于第二个问题, 则涉及到“中介效应”概念的定义问题。以系数  $c$  显著为前提条件是一种定义, 不用这个前提条件是另一种定义, 从外延来看, 后者包含了前者。按概念的外延与内涵的反变关系, 后者的内涵缩小了。这是要引起重视的, 而不是仅仅支持或者反对这个系数  $c$  显著为前提就完事。如果不加区分的使用两种不同定义的概念, 就会造成混乱。这方面应当向数学家学习, 他们将“数”的概念不断扩张的同时, 也用不同的名称进行区分, 如整数、有理数、实数和复数。

按 Baron 和 Kenny (1986)定义, 中介效应是以系数  $c$  显著为前提, 即  $X$  显著影响  $Y$  为前提。在这个定义下, 分析中介效应可以解释“ $X$  如何影响  $Y$ ”, 中介过程提供了“ $X$  对  $Y$  的作用机制”(MacKinnon & Fairchild, 2009; 温忠麟, 侯杰泰, 张雷, 2005)。涉及中介的应用文章往往会声称要研究“ $X$  对  $Y$  的作用(或影响)机制”。如果系数  $c$  不显著, 就说明  $X$  对  $Y$  的影响不显著, 如果还问“ $X$  如何影响  $Y$ ”或者“ $X$  对  $Y$  的作用机制是什么”, 不合常理。此时, 合理的问题应当是“ $X$  为何不影响  $Y$ ”, 建模的逻辑已经与前面说的中介模型的逻辑不同了。所以比较好的做法是将这种情形与通常中介效应区分开来, 不少文献称之为“遮掩效应”(suppressing effects) (Kenny, 2003; MacKinnon, 2008; MacKinnon et al., 2000, 2002; Shrout &

Bolger, 2002)。如果间接效应和直接效应符号相反, 总效应就出现了被遮掩的情况, 其绝对值比预料的要低。

温忠麟等人(2012)的书上, 将系数  $c$  不显著的情形归入“广义中介分析”, 既说明这种情形与通常的中介分析有区别, 也可以看出与中介分析有联系。如果根据前后文可以自明, 也可以删去“广义”两字。说到底, 如果是按传统的目的研究中介效应, 是要以系数  $c$  显著为前提, 否则就是另一个故事了。比方说, 如果一个人买了房子, 你可以问“他是通过中介买的, 还是自己直接买的?”但如果一个人没有房子, 此时的问题应当是“他为啥没有房子?”, 可能根本就没有买过, 也可能买了又卖掉了(类似于符号相反的抵消)。从问题的提出, 到结果的解释, 两种情形可能是很不同的故事。

这样说来, 我们不用去争论中介效应要不要以系数  $c$  显著为前提, 而是应当根据实际情况进行立论, 合理地提出相应的问题, 建立模型进行分析, 并作出相应的解释。虽然系数  $c$  不显著还是可以继续分析, 但应当明白, 系数  $c$  显著与否, 是不同的事情, 用不同的名称区分开来是明智的做法。与此相关的两个同义词, 中介效应和间接效应, 也是有区分的。中介效应一定是间接效应, 因而有的场合两者是一回事; 但间接效应不一定是中介效应(Mathieu & Taylor, 2006; 温忠麟等, 2004)。

## 2.3 区分完全中介和部分中介是否合适?

逐步法中最后一步, 通过检验方程(3)的系数  $c'$  来区分完全中介还是部分中介。如果系数  $c'$  不显著, 属于完全中介(James & Brett, 1984)。Baron 和 Kenny (1986)认为完全中介是中介效应存在的最强有力的证明。区分完全中介和部分中介, 是对中介效应模型的效应量的一种文字描述(Preacher & Kelley, 2011), 可以帮助解释结果。

但完全中介和部分中介概念是有问题的。第一, 在总效应小(但显著)的时候, 间接效应可能不到总效应的七成, 直接效应已经不显著了, 结果是完全中介, 与常理相悖。一般地说, 当总效应小且样本也小的时候, 容易得到完全中介的结果(Preacher & Hayes, 2008), 但其实完全中介的情况是很少的(Baron & Kenny, 1986; Iacobucci, 2008)。第二, 当说  $M$  是  $X$  和  $Y$  关系的完全中介时,

排除了将来探索其他中介的可能性(Pituch, Whittaker, & Stapleton, 2005)。Preacher 和 Hayes (2008)呼吁放弃完全中介的概念,将所有中介都看作是部分中介。Zhao 等人(2010)建议直接报告间接效应和直接效应的显著性,是可取的做法。

#### 2.4 逐步法能验证因果关系吗?

逐步法的英文“causal steps approach”按字面翻译是“因果逐步法”,使得部分研究者误以为逐步法可以证明因果关系,也有部分研究者质疑逐步法其实得到的是相关关系,不是因果关系(Stone-Romero & Rosopa, 2008)。

没错,逐步法建立模型的过程中,假设了变量之间的影响关系,即 $X$ 影响 $M$ , $M$ 影响 $Y$ ,而且,逐步法确实也不能验证因果关系。但这是所有统计方法都存在的问题,即统计无法验证因果关系。其实,连简单的回归分析,也有同样的问题。

变量之间因果关系的提出和假设,应当是模型建立之前要做的事情。中介模型中的每一个箭头表示的因果关系,包括“ $X \rightarrow Y$ ”,“ $X \rightarrow M$ ”,“ $M \rightarrow Y$ ”,都要有有理据,或者有某种学科理论支持,或者有文献做铺垫,或者有经验常识作为佐证。总之,因果链中的每一个关系,都要在提出假设和建模之前得到支持,否则假设的模型就没有根基。如果只有“ $X \rightarrow M$ ”和“ $M \rightarrow Y$ ”,也可以推论“ $X \rightarrow Y$ ”,做出假设。如果两个变量 $X$ 和 $Y$ 的因果关系比较明确,或者人们对因果关系比较有信心,可以说“ $X$ 影响 $Y$ ”,否则可以说得委婉一点:“ $X$ 对 $Y$ 有预测作用”。

那么应当如何分析变量之间的因果关系呢?这里不拟涉及哲学上的因果论,先看看社科中比较经典的推论因果关系的3个准则(Cook & Campbell, 1979): (1)因和果共同变化; (2)因在果之前发生; (3)排除因果联系外的其他解释。根据这些准则,社科研究可以采用下面方法分析因果关系:一是理论分析进行因果推理,二是实验设计验证因果关系,三是追踪研究观察因果关系。

先说理论分析,这是通过问卷测量收集数据进行中介分析必须要做的。对于两个相关的变量 $X$ 和 $Y$ (相关变量满足了因果关系的第一个准则),通过下面理论分析可以增加对“ $X \rightarrow Y$ ”的信心:

(1)看 $X$ 和 $Y$ 的变量属性, $X$ 是比 $Y$ 更加本质的(或者是长久的、稳定的)属性。一般来说,本质属性影响状态属性,长期属性影响临时属性,稳

定属性影响不稳定属性。例如,学生的智力和学业成绩,前者更加本质,应当是智力影响成绩。又如,成年人的身高比体重稳定,因而身高影响体重。至于性别,从生物学知道是由染色体决定的,相对于社科中的其他变量,性别更加本质和长久稳定,所以社科中凡是与性别相关的变量,都可以说受到性别的影响,即性别是这些变量的“因”。

(2)颠倒 $X$ 和 $Y$ 的因果顺序,难以解释。就是说“ $Y \rightarrow X$ ”还不如“ $X \rightarrow Y$ ”那样好解释。例如,性犯罪数量与气温正相关,哪个是因?如果说性犯罪数量影响气温,解释不通;而“气温升高使得性犯罪数量增加”比较容易解释,一种解释是:气温高 $\rightarrow$ 衣着暴露 $\rightarrow$ 诱发性犯罪;另一种解释是:气温高 $\rightarrow$ 荷尔蒙分泌旺盛 $\rightarrow$ 引发性犯罪。又如,有人调查发现,参加过补习的学生,平均成绩还不如从未参加过补习的学生,得到结论是补习不能提高成绩。这里显然忽视了一个事实:成绩不好是参加补习的一个原因,而不是结果。

(3)尽量排除共同原因引起的虚假效应(有关虚假效应,参见侯杰泰,温忠麟,成子娟,2004)。例如,婚龄与年医药费支出正相关,谁是原因呢?都不是,年龄(代表时间)才是这两个变量的共同原因。重要期刊要求作者检验方法效应,也是一样道理,就是排除方法对各变量的共同影响。

上面前两个分析都在支持因果关系的第二个准则,即因在果之前发生。最后一个分析是支持第三个准则,即排除其他解释。

再说实验设计,通常认为是验证因果关系最有效的一种做法。上面说到的补习能否提高成绩的例子,如果进行实验设计,将同年级成绩相当的学生分成实验组(参加补习,当然教师水平还不能太差)和控制组(不参加补习),并且两组除了是否补习不同外,其他会影响成绩的因素都控制到基本相同,得到的结果能说明补习是否提高成绩。实验设计的关键在于控制无关因素,即控制与实验目的无关但可能会影响因变量的因素,然后看看自变量的变化是否会引起因变量的变化,读者可以参考研究方法中有关实验设计的论著(如莫雷,温忠麟,陈彩琦,2007)。

既然实验可以验证因果关系,那么设计“ $X \rightarrow M$ ”、“ $M \rightarrow Y$ ”和“ $X \rightarrow Y$ ”三个实验(是否需要第三个实验,参考2.2节的讨论),就应当可以验证

中介效应的因果关系。如果第一个实验能验证  $X$  是导致  $M$  的原因,第二个实验能验证  $M$  是导致  $Y$  的原因,那么间接路径的因果链就得到了验证 (Spencer et al., 2005)。当然,相同的变量,在每个实验中都应当有相同的定义和测量。

如果要验证完全中介,则要通过所谓的过程调节设计(moderation-of-process design),先设计一个实验验证“ $X \rightarrow Y$ ”;然后设计另一个实验,如果控制了  $M$ ,无论  $X$  如何变化, $Y$  都不会变化了 (Spencer et al., 2005)。前面说过,完全中介的情形罕见,不拟赘述。

最后说说追踪研究。实验设计通常是在时间很短的一次实验中观测涉及的变量,因而需要假设自变量的变化会对因变量产生即时的影响。但在心理和其他社科领域,有些变量之间虽然有因果关系,也可能需要一段时间才能观测到变化,即所谓的滞后效应。例如,引进有效的教学手段,难以立即改变学生的成绩。又如,学业成绩的改变可能即时影响自信,但自信的改变却要一段时间后才能影响学业成绩。这时,基于横断的实验数据不能够对因果作出推断,应当采用追踪研究获取纵向数据(Cole & Maxwell, 2003)。

追踪研究是间隔一段时间对研究对象重复测量(两次或多次)自变量、中介变量和因变量,获得历时性的数据,用追踪数据分析技术(刘红云,张雷,2005),检验自变量是否会影响后续观测的中介变量,中介变量是否会影响后续观测的因变量。Mackinnon, Fairchild 和 Fritz (2007)总结了至少 3 种可以检验追踪数据的中介效应的方法,包括自回归模型(autoregressive modeling),潜增长模型(latent growth modeling),和潜差异分数模型(latent difference score modeling)。

上述三类方法,都有一定的局限性。理论分析只是一种因果推理,推理只要符合逻辑就是有效推理,但不能保证推理一定正确。前面已经提到过横断的实验数据的不足,追踪研究也有局限,包括重复测量会引起的练习效应、疲劳效应等(温忠麟,2009)。

总之,中介效应分析是否仅仅是相关分析,关键要看内容。相关是不能证明因果的,回归或者中介分析中的因果关系依靠统计以外的理论来支撑(包括理论分析、实验、追踪调查等)。不能一看到中介分析就认定仅仅是相关分析。如果有理

论分析,加上统计验证,我们就提高了对因果的信心。所有方法都只能在一定程度上证实因果关系,其中实验设计是最为可靠的方法。但所有方法都不能说最终证实了因果关系,只能证伪,因果关系的确立最终还是要经过实践的检验。最后说明一下,这里所说的因果,包括直接因果和间接因果。还有,我们是在统计意义上谈论因果关系,只是集体规律,对个体可能无效。

### 3 新的中介效应检验流程

对于系数乘积的检验,温忠麟等人(2004)早就意识到,如果检验结果都显著,依次检验结果强于 Sobel 检验结果,所以在他们提出的检验流程中,先进行依次检验,不显著才需要做 Sobel 检验。现在, Sobel 法由 Bootstrap 法取代,根据前面的讨论,对中介效应的检验流程进行相应的修改(见图 2),步骤如下。

第一步,检验方程(1)的系数  $c$ ,如果显著,按中介效应立论,否则按遮掩效应立论。但无论是否显著,都进行后续检验。

第二步,依次检验方程(2)的系数  $a$  和方程(3)的系数  $b$ ,如果两个都显著,则间接效应显著,转到第四步;如果至少有一个不显著,进行第三步。

第三步,用 Bootstrap 法直接检验  $H_0: ab = 0$ 。如果显著,则间接效应显著,进行第四步;否则间接效应不显著,停止分析。

第四步,检验方程(3)的系数  $c'$ ,如果不显著,即直接效应不显著,说明只有中介效应。如果显著,即直接效应显著,进行第五步。

第五步,比较  $ab$  和  $c'$  的符号,如果同号,属于部分中介效应,报告中中介效应占总效应的比例  $ab/c$ 。如果异号,属于遮掩效应,报告间接效应与直接效应的比例的绝对值  $|ab/c'|$ 。

对这个流程,有几点说明:

(1)当间接效应显著时,如果第一步检验后按遮掩效应立论,最后结果按遮掩效应解释。如果第一步检验后按中介效应立论,要根据  $ab$  和  $c'$  的符号进行解释,如果符号相反,按遮掩效应解释。就是说,开始按中介效应立论,不排除最后要按遮掩效应解释,但这样的情况少见。

(2)关于中介效应的效应量,起码应当报告  $ab/c$  或者  $|ab/c'|$ ,并酌情报告其他效应量。Preacher 和 Kelley (2011)给出了上 10 种效应量计算方法,

其中 MacKinnon (2008, 也见 Fairchild, MacKinnon, Taborga, & Taylor, 2009)定义的类似于回归中的  $R^2$  那样有方差解释率意义的效应量值得注意, 但这些效应量没有单调性(即中介效应  $ab$  上升时, 效应量可能反而下降), 不好理解。至于 Preacher 和 Kelley (2011)提出并推荐的中介效应量指标  $\kappa^2$ , 不仅缺乏统计意义, 而且没有单调性(Wen, under review), 不用为好。

(3)这个流程主要是从参数检验的角度考虑的。从参数估计角度看, 一般认为, 单单给出点估计是不够的, 应当给出区间估计。系数乘积  $ab$  的置信区间计算应当用 Bootstrap 法代替 Sobel 法。这样, 为了做区间估计, Bootstrap 法成为一个必须的方法, 而且依次检验也可以通过 Bootstrap 法进行, 即用 Bootstrap 法求出系数  $a$  和  $b$  的置信区间进行检验(但单个的系数不像系数乘积那样肯定会违背正态分布, 所以使用 Bootstrap 法依次检

验  $a$  和  $b$  与通常的依次检验结果基本上不会有出入)。有经验的研究者还可以通过区间估计看出方程(3)是否存在多重共线性问题。如果 Bootstrap 法的置信区间过大, 说明参数估计摇摆不定, 可能存在多重共线性问题。尽管如此, 还是应当先做依次检验, 因为如果显著的话, 结果强于直接检验系数乘积。用附录 1 那样的简单 Mplus 程序, 可以一次性得到所要的全部结果, 包括通常的依次检验结果和 Bootstrap 法置信区间。

(4)如果直接效应显著, 不排除存在其他中介变量的可能, Zhao 等人(2010)建议在讨论部分说明这种可能性。

#### 4 基于结构方程的中介分析

如果测验的信度很高(如 0.9 以上), 使用显变量和使用潜变量分析变量之间关系得到的结果会很接近。但如果测验信度不是很高(如只是通常

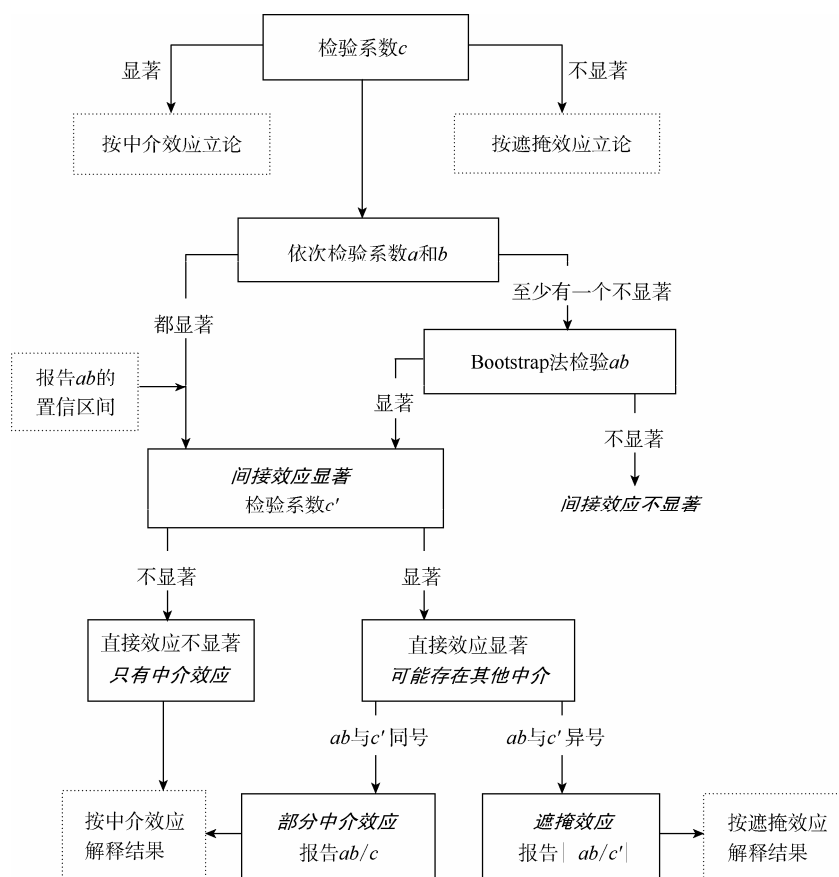


图2 中介效应检验流程



可接受的 0.7 左右),使用显变量分析变量之间关系往往会低估效应(侯杰泰等,2004)。中介效应因为涉及两个路径系数的乘积,受到的影响可能更大。Ledgerwood 和 Shrout (2011)的模拟研究发现,如果使用指标的均值作为显变量,中介效应的估计值低估实际中介效应的比例与合成信度(叶宝娟,温忠麟,2011)的乘积有关。例如,如果信度都是 0.9,则中介效应的估计值是实际中介效应的 80%( $\approx 0.9 \times 0.9$ )左右。但使用潜变量的弱点是中介效应估计的标准误较大,降低了检验力。就是说,使用潜变量的检验力通常低于使用显变量的检验力。为此,Ledgerwood 和 Shrout 建议,如果测验信度不够高,使用两步分析策略:第一步用显变量建模检验中介效应,第二步用潜变量建模估计中介效应。不过,如果用潜变量检验中介效应已经显著,就没有必要报告显变量分析结果了。

为了方便,我们前面讨论的模型所涉及的变量都是显变量,但有关的概念、模型路径图、分析思路、检验方法和流程同样适用于潜变量。对于潜变量的中介分析,只需将路径图中的长方形框(表示显变量)换做椭圆形图框(表示潜变量)(温忠麟等,2012)。使用结构方程,不仅可以同时处理显变量和潜变量,还可以同时分析多个自变量、多个因变量和多个中介变量的关系,也可以分析追踪数据(侯杰泰等,2004)。

对于显变量模型的分析,模型的拟合指数一般都很好,因此,可以把重点放在效应分析上。但是,对于潜变量,模型会变得更复杂,模型中既包括结构方程,也包括测量方程,此时,模型的拟合检验变得很重要(温忠麟等,2012)。对于题目指标多的潜变量,可能还涉及到题目打包问题(吴艳,温忠麟,2011)。

具体到用结构方程分析图 1 所示的中介模型,不失一般性,假设  $X$  有 4 个指标  $x_1, x_2, x_3, x_4$ ,  $M$  有 3 个指标  $m_1, m_2, m_3$ ,  $Y$  也有 3 个指标  $y_1, y_2, y_3$ 。首先,建立一个图 1(a)相应的结构方程模型,检验  $c$  的显著性。这一步,懂得结构方程的读者都会做。然后建立一个图 1(b)相应的结构方程模型,按附录 2 的 Mplus 程序,就可以得到需要的结果,包括依次检验的结果和 Bootstrap 法置信区间。程序中注明了如何解读结果,已有结构方程知识的读者容易看明白,并修改程序应用在自己的研究中。

## 5 中介效应模型的发展

随着统计理论的发展和分析软件的进步,中介效应分析模型也有了长足的发展,包括类别变量的中介模型、多重中介模型、多水平中介模型、有中介的调节模型与有调节的中介模型等。这些模型的分析方法,一般都基于简单中介模型分析方法或者思路,并受到简单中介模型分析方法发展的影响。

### 5.1 类别变量的中介效应模型

通常的中介效应模型,假设自变量、中介变量和因变量均为连续变量。对于自变量  $X$  为分类或者等级变量的情景,可以通过定义虚拟变量(dummy variable)的方法来处理,中介效应的分析与连续变量的步骤完全相同。对于因变量为分类或等级变量的情景,研究相对较少(Mackinnon, Lockwood, Brown, Wang, & Hoffman, 2007)。

如所知,如果因变量是分类或等级变量,自变量是连续变量,应当用 Logistic 回归取代通常的线性回归(Pregibon, 1981),回归系数的尺度转换为 Logit 量尺。对于因变量  $Y$  是分类或者等级变量、中介变量( $M$ )和自变量( $X$ )是连续变量的中介效应模型, $M$  对  $X$  的回归系数(连续变量的量尺)与  $Y$  对  $M$  的回归系数(Logit 量尺)和  $Y$  对  $X$  的回归系数(Logit 量尺)均不在相同的尺度上,因此不能简单采用处理连续变量中介效应的方式,直接将回归系数  $a$  和  $b$  相乘得到中介效应大小。因而,这样的模型需要通过标准化转换实现回归系数的等量尺化(MacKinnon, 2008; MacKinnon, Lockwood et al., 2007)。Iacobucci (2012)对因变量为分类数据的中介效应进行了理论探讨,强调在实际应用中研究者应该关注因变量的性质,选择合适的中介效应分析方法。

### 5.2 多重中介效应模型

对于情景比较复杂的研究,经常需要多个中介变量才能清晰地解释自变量对因变量的作用(MacKinnon, 2008),这就涉及多重中介(multiple mediation)模型。根据多个中介变量之间是否存在相互影响,多重中介模型可以分为单步多重中介模型和多步多重中介模型(Hayes, 2009)。单步多重中介模型,也称为并行多重中介模型,是指中介变量之间不存在相互影响。多步多重中介模型,也称为链式多重中介模型,是指中介变量之间存

在影响关系,中介变量表现出顺序性特征,形成中介链(柳士顺,凌文铨,2009;温忠麟等,2012)。

多重中介效应分析可以从3个角度进行,(1)总的中介效应(total mediation effect),即估计和检验所有间接效应的总和;(2)特定路径的中介效应(specific mediation effect),即估计和检验某个感兴趣的特定路径的间接效应;(3)对比中介效应,即估计和检验某两个路径的间接效应的差异(Hayes, 2009; MacKinnon, 2008; Preacher & Hayes, 2008)。多重中介模型因为涉及的变量较多、路径比较复杂,即使只涉及显变量,一般也要使用结构方程模型进行分析(Cheung, 2007; Lau & Cheung, 2012; Macho & Ledermann, 2011; Preacher & Hays, 2008)。检验多重中介效应比较好的方法是Bootstrap法(Cheung, 2007; 方杰,温忠麟,张敏强,孙配贞,印刷中; Lau & Cheung, 2012)。

### 5.3 多层中介效应模型

当数据具有层级结构时,变量分属不同的层级(也称为水平),如学生变量是第一层,学校变量是第二层。多层中介效应模型是多层线性模型和中介效应模型的结合,可以研究组织层面自变量对个体层面因变量的影响,增强中介效应的解释力(方杰,张敏强,邱皓政,2010)。考虑最简单的情形,模型中含有一个预测变量 $X$ ,一个中介变量 $M$ 和一个结果变量 $Y$ ,以两水平中介模型为例,根据 $X$ 、 $Y$ 和 $M$ 所在的层级不同,可以组合出多种类型的中介模型,但研究中常见的真正涉及多层中介效应的模型只有3类,包括2-2-1模型(这3个数字依次代表自变量、中介变量和因变量的层级)、2-1-1模型和1-1-1模型3种(温忠麟等,2012; Zhang, Zyphur, & Preacher, 2009)。

Bauer, Preacher 和 Gil (2006)根据中介变量 $M$ 是在第一层测量还是在第二层测量,区分低层中介模型和高层中介模型,2-2-1模型是高层中介模型,2-1-1模型和1-1-1模型是低层中介模型(温忠麟等,2012)。对于高层中介模型,模型中的路径系数都是固定的,此时中介效应的估计和检验相对简单,但是对于低层中介模型,模型中的路径系数可能是随机的,此时中介效应的估计及检验可能比较麻烦(Kenny, Korchmaros, & Bolger, 2003)。不同的模型和路径系数类型分析方法可能不同,其中Bootstrap法起着重要作用(方杰,温忠麟,张

敏强,任皓,2014; Li & Beretvas, 2013; Preacher, Zhang, & Zyphur, 2011; Preacher, Zyphur, & Zhang, 2010; 温忠麟等,2012)。

### 5.4 有中介的调节模型与有调节的中介模型

如果自变量 $X$ 与因变量 $Y$ 的关系受到第三个变量 $U$ 的作用,此时 $U$ 是调节变量,影响 $X$ 和 $Y$ 之间关系的方向(正或负)和强弱,比如“感觉寻求”与“青少年烟酒使用”的关系,受到“生活事件”的影响。调节效应分析的目的是探究 $X$ 何时影响 $Y$ 或何时影响较大(Baron & Kenny, 1986; Muller, Judd, & Yzerbyt, 2005)。中介变量和调节变量有时候会被混淆或者换用,温忠麟等(2005)从不同角度对这两种变量做了比较和区分。

如果一个模型包含不止3个变量,可能同时包含调节变量和中介变量,这些变量在模型中的位置和作用不同会产生不同的模型,有中介的调节模型(mediated moderation model)和有调节的中介模型(moderated mediation model)都是同时包含调节变量和中介变量的模型。

有中介的调节模型意味着自变量对因变量的效应受到调节变量的影响,而调节效应(至少部分地)通过中介变量而起作用(Baron & Kenny, 1986; Edwards & Lambert, 2007; Muller et al., 2005; 温忠麟,张雷,侯杰泰,2006)。叶宝娟和温忠麟(2013)对文献上已有的检验有中介的调节模型的方法或步骤进行梳理,总结出一个检验流程。

有调节的中介模型意味着自变量通过中介变量对因变量产生影响,而中介过程受到调节变量的调节(Baron & Kenny, 1986; Edwards & Lambert, 2007; Muller et al., 2005; 温忠麟等,2006)。温忠麟和叶宝娟(2014)分析了文献上出现的检验有调节的中介模型的3种常见方法:依次检验、系数乘积的区间检验和中介效应差异检验,从显著性结果的强弱关系、包含的信息量和可解释性等方面进行比较,认为它们之间是替补关系:优先使用依次检验,将系数乘积的区间检验(Bootstrap法)作为其替补,而中介效应的差异检验(Bootstrap法)又是系数乘积的区间检验的替补,当前一种检验方法不显著的时候,使用后一种检验方法。

在某些情况下,有调节的中介模型与有中介的调节模型会有一模一样的路径图,区别在于立

论和解释不同。对于有中介的调节模型,重心在于考虑自变量与因变量之间关系的方向(正或负)和强弱受到的影响,即调节效应;其次考虑调节变量是如何起作用的,即是否通过中介变量而起作用。对于有调节的中介模型,重心在于考虑自变量对因变量的作用机制,即中介效应;其次考虑中介过程是否受到调节,即中介作用何时较强、何时较弱。重要的是,两种模型分析方法和步骤不同,特别是分析步骤需要与模型解释相配合(温忠麟,叶宝娟,2014)。

## 6 小结

中介效应分析的发展主要是两个方面,一个是简单中介模型如何更加精确地检验间接效应(即检验  $H_0: ab = 0$ ),另一个是针对不同的变量和数据类型,发展相应的中介模型,扩大应用范围。两个方面的发展也互有影响。

对于间接效应的检验,针对传统的 Sobel 检验的缺陷,发展了多种替代的检验方法,目前得到公认的是 Bootstrap 法直接检验系数乘积  $ab$ ,在各种复杂模型中都有用。但由于受到非此即彼的思维影响,不少文献认为依次检验法也应当让位于 Bootstrap 法。我们提出的检验流程,综合了依次检验法和 Bootstrap 法的优点,推荐先尝试简单的依次检验  $a$  和  $b$ ,如果不显著则用 Bootstrap 法直接检验系数乘积  $ab$  以提高检验力。用这个流程,无论是考虑第一类错误率、检验力还是结果的解释性,与单纯的 Bootstrap 法检验系数乘积相比,只会更好不会更差。

模型发展方面,一是应对类别变量、多个中介变量、层级数据以及追踪数据,发展新的模型,二是将中介变量和调节变量组合起来建模,这样就将中介分析的方法和应用扩大到很广的范围。相信这是未来的主要发展方向,即新的应用领域的模型建构和分析方法研究。包括 Bootstrap 法在内的现代统计方法将是兴趣所在。但即使是复杂模型,也还是可以看看简单的依次检验是否能派上用场,简单而有效是一个方法能够流行的关键。

还有一个发展方向是扩展中介效应的概念,考虑非线性效应、甚至脱离具体统计模型分析中介效应(Imai, Keele, & Tingley, 2010; Valeri & VanderWeele, 2013)。但根据内涵与外延的反变原

理,扩展中介效应概念后的研究将比较困难。

## 参考文献

- 方杰, 温忠麟, 张敏强, 任皓. (2014). 基于结构方程模型的多层中介效应分析. *心理科学进展*, 22, 530-539.
- 方杰, 温忠麟, 张敏强, 孙配贞. (印刷中). 基于结构方程模型的多重中介效应分析. *心理科学*.
- 方杰, 张敏强. (2012). 中介效应的点估计和区间估计: 乘积分布法、非参数 Bootstrap 和 MCMC 法. *心理学报*, 44, 1408-1420.
- 方杰, 张敏强, 邱皓政. (2010). 基于阶层线性理论的多层级中介效应. *心理科学进展*, 18, 1329-1338.
- 侯杰泰, 温忠麟, 成子娟. (2004). *结构方程模型及其应用*. 北京: 教育科学出版社.
- 刘红云, 张雷. (2005). *追踪数据分析方法及其应用*. 北京: 教育科学出版社.
- 柳士顺, 凌文铨. (2009). 多重中介模型及其应用. *心理科学*, 32, 433-435.
- 莫雷, 温忠麟, 陈彩琦. (2007). *心理学研究方法*. 广州: 广东高等教育出版社.
- 温忠麟. (2009). *教育研究方法基础* (第2版). 北京: 高等教育出版社.
- 温忠麟, 侯杰泰, 张雷. (2005). 调节效应与中介效应的比较和应用. *心理学报*, 37, 268-274.
- 温忠麟, 刘红云, 侯杰泰. (2012). *调节效应和中介效应分析*. 北京: 教育科学出版社.
- 温忠麟, 叶宝娟. (2014). 有调节的中介模型检验方法: 竞争还是替补? *心理学报*, 46, 714-726.
- 温忠麟, 张雷, 侯杰泰. (2006). 有中介的调节变量和有调节的中介变量. *心理学报*, 38, 448-452.
- 温忠麟, 张雷, 侯杰泰, 刘红云. (2004). 中介效应检验程序及其应用. *心理学报*, 36, 614-620.
- 吴艳, 温忠麟. (2011). 结构方程建模中的题目打包策略. *心理科学进展*, 19, 1859-1867.
- 叶宝娟, 温忠麟. (2011). 单维测验合成信度三种区间估计的比较. *心理学报*, 43, 453-461.
- 叶宝娟, 温忠麟. (2013). 有中介的调节模型检验方法: 甄别和整合. *心理学报*, 45, 1050-1060.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11, 142-163.
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for

- indirect effects. *Multivariate Behavioral Research*, 45(4), 661–701.
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11, 296–325.
- Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling*, 14, 227–246.
- Clogg, C. C., Petkova, E., & Shihadeh, E. S. (1992). Statistical methods for analyzing collapsibility in regression models. *Journal of Educational Statistics*, 17(1), 51–74.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112, 558–577.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. MA: Houghton Mifflin.
- Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, 12, 1–22.
- Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009).  $R^2$  effect-size measures for mediation analysis. *Behavior Research Methods*, 41, 486–498.
- Freedman, L. S., & Schatzkin, A. (1992). Sample size for studying intermediate endpoints within intervention trials or observational studies. *American Journal of Epidemiology*, 136, 1148–1159.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18, 233–239.
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47, 61–87.
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76, 408–420.
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24, 1918–1927.
- Iacobucci, D. (2008). *Mediation analysis*. Thousand Oaks, CA: Sage.
- Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, 22, 582–594.
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15, 309–334.
- James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, 69, 307–321.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602–619.
- Kenny, D. A. (2003). “Mediation,” <http://www.users.rcn.com/dakenny/mediate.htm/>
- Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8, 115–128.
- Lau, R. S., & Cheung, G. W. (2012). Estimating and comparing specific mediation effects in complex latent variable models. *Organizational Research Methods*, 15, 3–16.
- Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, 101, 1174–1188.
- Li, X., & Beretvas, S. N. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling*, 20, 241–264.
- Macho, S., & Ledermann, T. (2011). Estimating, testing, and comparing specific effects in structural equation models: The phantom model approach. *Psychological Methods*, 16, 34–43.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- MacKinnon, D. P., & Fairchild, A. J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science*, 18, 16–20.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614.
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science*, 1, 173–181.
- MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, 4, 499–513.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83–104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004).



- Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39, 99–128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41–62.
- Mathieu, J. E., & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in Organizational Behavior. *Journal of Organizational Behavior*, 27, 1031–1056.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89(6), 852–863.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken, NJ: Wiley.
- Pituch, K. A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research*, 43(2), 237–267.
- Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*, 41(3), 367–400.
- Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, 40, 1–23.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36, 717–731.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891.
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, 16, 93–115.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42, 185–227.
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, 18, 161–182.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15, 209–233.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 9, 705–724.
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5, 359–371.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422–445.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89, 845–851.
- Stone-Romero, E. F., & Rosopa, P. J. (2008). The relative validity of inferences about mediation as a function of research design characteristics. *Organizational Research Methods*, 11, 326–352.
- Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11, 241–269.
- Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, 43, 692–700.
- Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, 18, 137–150.
- Wen, Z. (Under review). Monotonicity of effect sizes: Comment on kappa-squared as mediation effect size measure.
- Wen, Z., Marsh, H. W., & Hau, K. T. (2010). Structural equation models of latent interactions: An appropriate standardized solution and its scale-free properties. *Structural Equation Modeling*, 17, 1–22.
- Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, 14, 301–322.
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, 12, 695–719.
- Zhao, X., Lynch Jr., J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37, 197–206.

### 附录1 用偏差校正的非参数百分位Bootstrap法检验显变量中介效应Mplus程序

DATA: FILE IS p.dat; ! p.dat是原始数据文件, 按X M Y顺序排列

VARIABLE: NAMES ARE X M Y; !变量名称

Analysis: bootstrap=1000; ! Bootstrap 法抽样1000 次

MODEL:

Y on X; !做Y对X的回归

M on X (a); !做M对X的回归, X的回归系数命名为a

Y on X

M (b); !做Y对X和M的回归, M的回归系数命名为b, 需要单独一行

MODEL CONSTRAINT:

new (H); !定义辅助变量

H=a\*b; !系数乘积ab的估计

OUTPUT:

cinterval (bcbootstrap);

!输出各个系数及系数乘积ab的偏差校正的非参数百分位Bootstrap 法置信区间。若要得到(不校正的)非参数百分位Bootstrap 法置信区间, 只需将OUTPUT中的 cinterval (bcbootstrap)改为cinterval (bootstrap)即可。

**注释:** 各个系数及系数乘积 ab 的估计值, 对应于 Mplus 输出结果中的“MODEL RESULTS”部分中的各个系数及“New/Additional Parameters”中的 H 的参数估计值, 各个系数及 ab 的 95%(或 99%)置信区间的下限和上限, 对应于 Mplus 输出结果中的“CONFIDENCE INTERVALS OF MODEL RESULTS”部分中的各个系数及“New/Additional Parameters”中的 H 的“Lower 2.5%”(或“.5%”)和“Upper 2.5%”(或“.5%”)的值。

### 附录2 用偏差校正的非参数百分位Bootstrap法检验潜变量中介效应Mplus程序

DATA: FILE IS p.dat; ! p.dat是原始数据文件, 按x1-x4 m1-m3 y1-y3顺序排列

VARIABLE: NAMES ARE x1-x4 m1-m3 y1-y3; !变量名称

Analysis: bootstrap=1000; ! Bootstrap 法抽样1000 次

MODEL:

Y by y1-y3; ! y1-y3是潜变量Y的指标

M by m1-m3; ! m1-m3是潜变量M的指标

X by x1-x4; ! x1-x4是潜变量X的指标

Y on X; !做Y对X的回归

M on X (a); !做M对X的回归, X的回归系数命名为a

Y on X

M (b); !做Y对X和M的回归, M的回归系数命名为b, 需要单独一行

MODEL CONSTRAINT:

new (H); !定义辅助变量

H=a\*b; ! 系数乘积ab的估计

OUTPUT:

cinterval (bcbootstrap);

!输出各个系数及系数乘积 ab 的偏差校正的非参数百分位 Bootstrap 法置信区间。若要得到(不校正的)非参数百分位Bootstrap 法置信区间, 只需将 OUTPUT 中的 cinterval (bcbootstrap)改为 cinterval (bootstrap)即可。

## Analyses of Mediating Effects: The Development of Methods and Models

WEN Zhonglin<sup>1,2</sup>; YE Baojuan<sup>3</sup>

(<sup>1</sup> Center for Studies of Psychological Application / School of Psychology,  
South China Normal University, Guangzhou 510631, China)

(<sup>2</sup> Hong Kong Examinations and Assessment Authority, Hong Kong, China)

(<sup>3</sup> School of Psychology, Jiangxi Normal University, Nanchang 330022, China)

**Abstract:** Mediation models are frequently used in the research of psychology and other social science disciplines. Mediation indicates that the effect of an independent variable on a dependent variable is transmitted through a third variable, which is called mediator. In most applied research, Baron and Kenny's (1986) causal steps approach has been used to test mediating effect. In recent years, however, many methodological researchers questioned the rationality of the causal steps approach, and some of them even attempted to stop its use. Firstly, we clarify the queries on the causal steps approach one by one. Secondly, we propose a new procedure to analyze mediating effects. The new procedure is better than any single method that constitutes the procedure in terms of Type I error rate and power. The proposed procedure can be conducted by using observed variables and/or latent variables. Mplus programs are supplied for the procedure with observed variables and/or latent variables. Finally, this article introduces the development of mediation models, such as mediation model of ordinal variables, multilevel mediation, multiple mediation, moderated mediation, and mediated moderation.

**Key words:** mediating effect; indirect effect; causal steps approach; Bootstrap method; causality