心理学探新 2014,Vol. 34,No. 1,43 - 47
PSYCHOLOGICAL EXPLORATION

# 两种估计多维测验合成信度置信区间方法比较*

杨 强[1],叶宝娟[2],温忠麟[3]
(1. 江西师范大学教育学院,南昌 330022;2. 江西师范大学心理学院,江西省心理与认知科学重点实验室,南昌 330022;
3. 华南师范大学心理应用研究中心,广州 510631)

**摘 要:** 有两种方法可以估计多维测验合成信度的置信区间: Bootstrap 法和 Delta 法。本文用模拟研究比较这两种方法,结果发现,Delta 法与 Bootstrap 法得到结果的差异很小。因为 Bootstrap 法得到的是实证结果,通常被认为是真值的反映,而 Delta 法比 Bootstrap 法简单得多,所以可以用 Delta 法估计合成信度的置信区间。举例演示如何计算多维测验的合成信度以及用 Delta 法计算其置信区间。

**关键词:** 多维测验; 合成信度; 置信区间; Bootstrap 法; Delta 法
**中图分类号:** B841. 2 **文献标识码:** A **文章编号:** 1003 - 5184(2014)01 - 0043 - 05

心理与教育测验中,信度(reliability)是衡量测验质量的一个重要指标。没有信度的测验毫无用处(Aiken,2003)。即使一个完美的研究设计也无法弥补不可靠和不精确测量所带来的缺陷,因此,估计测验的信度就成了数据分析的必须前提和关键性步骤(Biemer,Christ,& Wiesen,2009; Vangeneugden et al,2010)。

最常用的 $\alpha$ 系数在测验不是基本 $\tau$ 等价或误差相关的情况下可能会低估(例如,刘红云,2008; Sijtsma,2009a)或高估(例如,Green & Yang,2009; Revelle & Zinbarg,2009; Sijtsma,2009b)测验信度。利用验证性因子模型,测验信度可以用合成信度(composite reliability)比较准确地估计(Bentler,2009; 邱皓政,林碧芳,2009; 温忠麟,叶宝娟,2011)。在报告合成信度时,最好是同时报告其置信区间,以了解信度估计的误差范围。

叶宝娟和温忠麟(2011)研究了单维测验合成信度的区间估计问题,比较了三种估计方法: Bootstrap 法、Delta 法和直接用 LISREL 软件输出的标准误进行计算。Bootstrap 法得到的结果是实证结果,比较可信,但需要数据模拟技术,非常麻烦。Delta 法是一种近似计算,可以根据计算公式在 SEM 软件中编程,添加额外参数,在估计合成信度的同时,结果文件会给出计算标准误所需要的参数值,进行简单计算即可得到标准误,比 Bootstrap 法简单。LISREL 软件添加额外参数估计合成信度时,结果文件中会直接给出其标准误,比 Bootstrap 法和 Delta 法都要简单。他们的研究发现,Delta 法与 Bootstrap 法得到的置信区间非常接近,而用 LISREL 输出的标准误计算的置信区间与 Bootstrap 法得到的置信区间相差很大,推荐使用 Delta 法计算合成信度的置信区间。

实证研究中还经常用到多维测验(multidimensional test),本文的目的是在多维情况下比较 Delta 法与 Bootstrap 法得到的置信区间,以评估 Delta 法的精度。

## 1 多维测验合成信度

若一个多维测验由 $p$ 个题目 $x_1,x_2,\cdots,x_p$ 测量了 $n$ 个因子 $\xi_1,\xi_2,\cdots,\xi_n$,$\delta_1,\delta_2,\cdots,\delta_p$ 为 $x_1,x_2,\cdots,x_p$ 的测量误差,则有:

$$x_i=\sum_{j=1}^{n}\lambda_{ij}\xi_j+\delta_i,i=1,2,\cdots,p \tag{1}$$

其中,$\lambda_{ij}$ 表示题目 $x_i$ 在因子 $\xi_j$ 上的负荷,每个题目只在一个因子上的负荷不是零。如果整份测验的分数相加有意义,测验分数 $X=x_1+x_2+\cdots+x_p$ 的合成信度为(Raykov & Shrout,2002; Yang & Green,2010)

$$\rho_m = \mathrm{var}(\sum_{i=1}^{p}\sum_{j=1}^{n}\lambda_{ij}\xi_j)/[\mathrm{var}(\sum_{i=1}^{p}\sum_{j=1}^{n}\lambda_{ij}\xi_j)+\sum_{i=1}^{p}\mathrm{var}(\delta_i)] \tag{2}$$

* 基金项目: 国家自然科学基金(31271116,31360237),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),江西省博士后科研择优资助项目(2013KY08),江西省博士后日常经费资助项目(2013RC18),江西省教育规划课题(13ZD3L011),江西师范大学青年英才培育资助计划,江西师范大学 2012 年度第二批青年成长基金项目。

通讯作者: 温忠麟,E - mail: wenzl@ scnu. edu. cn。

公式(2)得到的只是多维测验合成信度的点估计。对于要评价的总体合成信度，点估计提供的信息量有限并且不能给出估计的偏差(Raykov，Dimitrov，& Asparouhov，2010)。在评价一个测验质量时，最好用信度的区间估计来补充信度点估计得到的信息(Raykov & Shrout，2002；Zinbarg，Yovel，Revelle，& McDonald，2006)。

## 2 估计多维测验合成信度的置信区间的方法

使用 Bootstrap 法对一个固定的样本(当作总体)进行重复取样，以估计合成信度的标准误，进而计算置信区间。这种方法比较麻烦，但得到的是一种实证结果，通常作为真值的反映，可以用来比较其他方法计算结果是否合理(叶宝娟，温忠麟，2011)。本文所用的 Bootstrap 法是从一个给定的样本中有放回地重复取样以产生出许多样本，即将原始的样本当作总体，从这个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)。

Delta 法先用验证性因子分析得到模型参数及其方差和协方差，进而得到参数的光滑函数(smooth functions)的近似标准误(Raykov & Marcoulides，2004)。这种方法简单易行，但只是一种近似算法，其精度需要权衡。

用 Delta 法估计多维测验合成信度的标准误的公式(叶宝娟，温忠麟，2012)如下

$$SE = \sqrt{D_1^2 \mathrm{var}(s) + D_2^2 \mathrm{var}(t) + 2D_1 D_2 \mathrm{cov}(s,t)} \quad (3)$$

其中，SE 是标准误，$u$ 是标准化因子负荷的和，$v$ 是误差方差的和：

$$s = \mathrm{var}\left(\sum_{i=1}^{p}\sum_{j=1}^{n}\lambda_{ij}\xi_j\right),\ t = \sum_{i=1}^{p}\mathrm{var}(\delta_i) \quad (4)$$

$D_1$ 和 $D_2$ 由下面公式计算得到：

$$D_1 = \frac{\hat{t}}{(\hat{s}+\hat{t})^2},\ D_2 = -\frac{\hat{s}}{(\hat{s}+\hat{t})^2} \quad (5)$$

## 3 研究方法

### 3.1 模拟研究设计

考虑 5 个自变量：因子个数，每个因子题数，因子负荷，样本容量，估计方法。

(1)因子个数，设置的因子个数为 2，4。(被试间因素)

(2)每个因子上题数，设置的题数为 3，6，10。(被试间因素)

(3)因子负荷，设置的因子负荷在 0.3 ~0.9 之间，分为 3 种情况：高负荷(0.7 ~0.9)，中负荷(0.5 ~0.7)，低负荷(0.3 ~0.5)。(被试间因素)

(4)样本容量，设置的样本容量分别为 300，500，1000。(被试间因素)

(5)估计方法，Bootstrap 法和 Delta 法。(被试内因素)

本模拟实验是一个 2 ×3 ×3 ×3 ×2 的设计，前四个因素是被试间因素，最后一个因素是被试内因素。对于被试间因素，不同水平对应的模拟数据不同，共有 2 ×3 ×3 ×3 =54 种水平组合。因子个数、每个因子题数、因子负荷和样本容量是数据的条件，估计方法为关注的对象——合成信度区间估计方法，据此可以在不同数据条件下进行方法比较。

在每种处理(即水平组合)中，模拟一批正态数据，然后随机取一个拟合好的样本，即 RMSEA 和 SRMR 小于 0.08，NNFI 和 CFI 大于 0.9(温忠麟，侯杰泰，Marsh，2004)，用 Delta 法和 Bootstrap 法计算合成信度的标准误。比较两个标准误的差异，也就是比较了合成信度的置信区间的差异。并且还可以了解当测验题目、因子负荷和样本容量变化时，合成信度的置信区间及其不同方法的差异是如何变化的。

### 3.2 用 Bootstrap 法估计合成信度的置信区间

用 Bootstrap 法计算合成信度的置信区间有 3 个步骤：第一步，从原始样本中重复取样 1000 次，得到 1000 个 Bootstrap 样本，容量与原始样本的容量相同。这一步使用 LISREL8.72 实现；第二步，计算 1000 个 Bootstrap 样本的合成信度(点估计)；第三步，计算第二步得到的 1000 个 Bootstrap 样本合成信度的标准差，这个标准差就是用 Bootstrap 法计算得到的合成信度的标准误，进而计算合成信度的置信区间，区间中点是原始样本的合成信度，区间半径是 Bootstrap 法得到的标准误的 1.96 倍。

### 3.3 用 Delta 法估计合成信度的置信区间

在用 Delta 法计算合成信度的置信区间时，先用 LISREL 程序计算合成信度点估计值和区间估计所需要的参数，再根据公式(5)和(3)计算合成信度的标准误，进而计算区间估计，用这种方法比用 Bootstrap 法估计合成信度的置信区间要方便得多。

## 4 研究结果

计算用 Bootstrap 法得到的标准误时只使用收敛到恰当解且模型拟合较好的样本的结果。在 54 个处理条件中，有 49 个处理条件对应的 Bootstrap 样本有恰当解且结构模型拟合较好的样本比例为

100%，其余 5 个处理条件对应的 Bootstrap 样本有恰当解且结构模型拟合较好的样本比例都在 93% 以上。因为 Bootstrap 法赖以计算的样本不少于 930 个，所以计算结果有效。

因为 Bootstrap 法得到的标准误是一种实证结果，可以看作是真值，将 Delta 法计算的标准误与其比较来计算偏差。比较的标准与单维测验比较的标准相同，也就是说，如果标准误相差 0.01，信度的下限就会相差 0.02。从应用角度看，如果信度下限相差 0.01（相应的标准误相差 0.005），只是微小差别；如果信度下限相差 0.02（相应的标准误相差 0.01），是有点差别；如果信度下限相差 0.05（相应的标准误相差 0.025），是有实质差别。

### 4.1　Delta 法的标准误

比较 Delta 法和 Bootstrap 法的结果（见表 1 的 Bias_D 列）。Delta 法的标准误偏差（绝对值）有正有负，但都很小，均小于 0.005，在题目较多、负荷为高时，标准误偏差均小于或等于 0.001，尤其是在负荷为高时，标准误偏差为 0。因此，可以认为 Delta 法估计的标准误是近似无偏的。

### 4.2　合成信度的标准误变化情况

Bootstrap 法的结果是一种实证结果，因此检视 Bootstrap 法标准误的变化情况可以了解不同条件下合成信度估计的精度情况。保持其他条件不变，随着因子个数的增加（或者随着题数的增加、或者随着负荷的升高或者随着样本容量的增加），用 Bootstrap 法估计的标准误减小。

总之，因子越多、题目越多、负荷越高或者样本容量越大，合成信度的标准误越小，从而估计的合成信度越精确。用 Delta 法得到的标准误与 Bootstrap 法得到的标准误变化情况一致。

**表 1　正态分布下多维测验合成信度的两种标准误比较**

| 因子 | 题数 | 负荷 | $N$ | 信度 | SE_B | SE_D | Bias_D |
|---|---|---|---|---|---|---|---|
| 2 | 3 | 低 | 300 | 0.525 | 0.043 | 0.042 | -0.001 |
| | | | 500 | 0.531 | 0.032 | 0.033 | 0.001 |
| | | | 1000 | 0.433 | 0.028 | 0.028 | 0.000 |
| | | 中 | 300 | 0.665 | 0.034 | 0.031 | -0.003 |
| | | | 500 | 0.686 | 0.024 | 0.022 | -0.002 |
| | | | 1000 | 0.700 | 0.014 | 0.015 | 0.001 |
| | | 高 | 300 | 0.876 | 0.011 | 0.011 | 0.000 |
| | | | 500 | 0.872 | 0.009 | 0.009 | 0.000 |
| | | | 1000 | 0.881 | 0.006 | 0.006 | 0.000 |
| | 6 | 低 | 300 | 0.617 | 0.035 | 0.033 | -0.002 |
| | | | 500 | 0.608 | 0.029 | 0.026 | -0.003 |
| | | | 1000 | 0.612 | 0.019 | 0.018 | -0.001 |
| | | 中 | 300 | 0.835 | 0.014 | 0.014 | 0.000 |
| | | | 500 | 0.814 | 0.012 | 0.012 | 0.000 |
| | | | 1000 | 0.806 | 0.009 | 0.009 | 0.000 |
| | | 高 | 300 | 0.942 | 0.005 | 0.005 | 0.000 |
| | | | 500 | 0.931 | 0.005 | 0.005 | 0.000 |
| | | | 1000 | 0.935 | 0.003 | 0.003 | 0.000 |
| | 10 | 低 | 300 | 0.731 | 0.020 | 0.023 | 0.003 |
| | | | 500 | 0.716 | 0.018 | 0.018 | 0.000 |
| | | | 1000 | 0.701 | 0.014 | 0.014 | 0.000 |
| | | 中 | 300 | 0.889 | 0.010 | 0.009 | -0.001 |
| | | | 500 | 0.889 | 0.007 | 0.007 | 0.000 |
| | | | 1000 | 0.881 | 0.006 | 0.005 | -0.001 |
| | | 高 | 300 | 0.948 | 0.004 | 0.004 | 0.000 |
| | | | 500 | 0.957 | 0.003 | 0.003 | 0.000 |
| | | | 1000 | 0.958 | 0.002 | 0.002 | 0.000 |
| 4 | 3 | 低 | 300 | 0.608 | 0.031 | 0.034 | 0.003 |

续表 1

| 因子 | 题数 | 负荷 | $N$ | 信度 | SE_B | SE_D | Bias_D |
|---|---|---|---|---|---|---|---|
| | | | 500 | 0.543 | 0.031 | 0.031 | 0.000 |
| | | | 1000 | 0.532 | 0.022 | 0.022 | 0.000 |
| | | 中 | 300 | 0.762 | 0.022 | 0.021 | -0.001 |
| | | | 500 | 0.775 | 0.014 | 0.015 | 0.001 |
| | | | 1000 | 0.773 | 0.011 | 0.011 | 0.000 |
| | | 高 | 300 | 0.911 | 0.008 | 0.008 | 0.000 |
| | | | 500 | 0.909 | 0.006 | 0.006 | 0.000 |
| | | | 1000 | 0.913 | 0.004 | 0.004 | 0.000 |
| | 6 | 低 | 300 | 0.710 | 0.026 | 0.024 | -0.002 |
| | | | 500 | 0.695 | 0.020 | 0.020 | 0.000 |
| | | | 1000 | 0.691 | 0.015 | 0.014 | -0.001 |
| | | 中 | 300 | 0.876 | 0.011 | 0.010 | -0.001 |
| | | | 500 | 0.868 | 0.009 | 0.009 | 0.000 |
| | | | 1000 | 0.868 | 0.006 | 0.006 | 0.000 |
| | | 高 | 300 | 0.955 | 0.004 | 0.004 | 0.000 |
| | | | 500 | 0.963 | 0.002 | 0.002 | 0.000 |
| | | | 1000 | 0.955 | 0.002 | 0.002 | 0.000 |
| | 10 | 低 | 300 | 0.807 | 0.016 | 0.016 | 0.000 |
| | | | 500 | 0.777 | 0.016 | 0.014 | -0.002 |
| | | | 1000 | 0.780 | 0.010 | 0.010 | 0.000 |
| | | 中 | 300 | 0.915 | 0.007 | 0.007 | 0.000 |
| | | | 500 | 0.918 | 0.005 | 0.005 | 0.000 |
| | | | 1000 | 0.913 | 0.004 | 0.004 | 0.000 |
| | | 高 | 300 | 0.971 | 0.002 | 0.002 | 0.000 |
| | | | 500 | 0.973 | 0.002 | 0.002 | 0.000 |
| | | | 1000 | 0.971 | 0.001 | 0.001 | 0.000 |

注：SE_B 和 SE_D 分别表示 Bootstrap 法和 Delta 法得到的标准误。将 Bootstrap 法估计的标准误看作真值，Bias_D 表示 Delta 法的标准误的偏差。

## 5 用 Delta 法估计合成信度的置信区间示例

接下来用一个例子说明如何用 Delta 法来计算多维测验合成信度的置信区间。基本心理需要测验(basic psychological needs test)是经常用到的一个测验，由 Deci 和 Ryan(2000)编制、喻承甫等人(2012)修订的基本心理需要测验，评定的是个体生活中基本心理需要获得满足的程度，共 21 个项目，分别测量了青少年关系需要(8 个项目)、能力需要(6 个项目)和自主需要(7 个项目)三个方面。将基本心理需要测验施测于 600 个被试，模型的拟合指数为：$\chi^2(186) = 190.704$，RMSEA = 0.001，NNFI = 0.998，CFI = 0.999，SRMR = 0.031，模型拟合很好(温忠麟等，2004)。

LISREL 的输出结果可以给出公式(2)中的合成信度的点估计值，以及公式(3)中的所有参数(但 $D_1$ 和 $D_2$ 需要计算)。将 LISREL 的输出结果代入公式(3)容易求得合成信度的标准误，进而计算其置信区间。本例合成信度的点估计值为 0.796，用 Delta 法求得的合成信度的标准误为 0.012，合成信度 95% 的置信区间为(0.772，0.820)。如果在 LISREL 中用 Bootstrap 法抽样 1000 次，本测验的 1000 个 Bootstrap 样本都有恰当解并且结构模型拟合好，1000 个 Bootstrap 样本合成信度的标准误是 0.012，95% 的置信区间为(0.772，0.820)，与用 Delta 法得到的结果相同。

本例用 Delta 法估计的合成信度的置信区间的长度为 0.048，信度的误差范围相对较大。因此，很有必要估计和报告合成信度的置信区间，以便对基本心理需要测验的质量进行比较客观的评价。

## 6 讨论和结论

在评价测验质量时，不仅需要信度的点估计，区间估计也很重要，通过区间估计可以衡量信度估计的精确性。如果预先设定的可接受的信度小于测验信度的置信区间的下限，那么此测验的信度可以接

受,可以接受此测验。如果可接受的信度大于测验信度的置信区间的上限,那么此测验的信度不高,不应当使用此测验。如果测验信度的置信区间包含了可接受的信度,那么此测验的信度值得怀疑,在做出接受或者拒绝此测验的决定时应谨慎。

本文比较了两种估计多维测验合成信度标准误的方法: Bootstrap 法和 Delta 法。结果发现 Delta 法的标准误与 Bootstrap 法的标准误差异很小。Bootstrap 法的结果是一种实证结果,可以作为真值看待,但 Bootstrap 法计算过程相当麻烦。Delta 法通过近似计算得到标准误,与 Bootstrap 法的结果差别很小,而且比 Bootstrap 法简单得多,所以推荐使用 Delta 法。

随着因子增加、题数增加、负荷的升高或样本容量的增多,合成信度的标准误倾向于减小。因此,在编制测验时,应适当增加测验题目,选用因子负荷高的题目,尽可能多施测一些被试。

需要说明的是,Mplus5.0 以上的版本也可以在程序中添加额外变量,计算合成信度,并且其是用 Delta 法计算参数的标准误(Muthén & Muthén, 2010),可以直接给出置信区间。因此 Mplus 软件结果文件输出合成信度的标准误与本文用 Delta 法求得的合成信度标准误是相同的(计算误差除外)。

同单维测验相同(叶宝娟,温忠麟,2011),求多维测验合成信度的置信区间,可以只考虑使用单侧概率。例如,本文示例计算得到的合成信度的 95% 的置信区间为(0.772,0.820),则合成信度低于 0.772 的可能性为 2.5%。实际应用中,不需要有 95% 的把握说多维测验合成信度在置信区间中,只需要有 95% 的把握说,多维测验合成信度不低于置信区间下限。

本文在数据正态,误差不相关的情形之下,用模拟研究比较了两种估计多维测验合成信度的置信区间的方法,并由此得出了结论。本文的结论是否适用于误差相关、数据非正态的情况呢? 有待进一步研究。

## 参考文献


刘红云.(2008).α 系数与测验的同质性.*心理科学*,*31*(1),185－188.

邱皓政,林碧芳.(2009).*结构方程模型的原理与应用*.北京: 中国轻工业出版社.

温忠麟,叶宝娟.(2011).测验信度估计: 从 α 系数到内部一致性信度.*心理学报*,*43*(7),821－829.

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*(4),453－461.

叶宝娟,温忠麟.(2012).用 Delta 法估计多维测验合成信度的置信区间.*心理科学*,*35*(6),1213－1217.

喻承甫,张卫,曾毅茵,叶婷,胡谏萍,李丹黎.(2012).青少年感恩、基本心理需要与病理性网络使用的关系.*心理发展与教育*,*27*(1),575－582.

Aiken,R. L.(2003). *Psychological testing and assessment*(11th ed.). Needham Heights,MA: Allyn and Bacon.

Bentler,P. M.(2009). Alpha,dimension－free,and model－based internal consistency reliability. *Psychometrika*,*74*,137－143.

Biemer,P. P.,Christ,S. L.,& Wiesen,C. A.(2009). A general approach for estimating scale score reliability for panel survey data. *Psychological Methods*,*14*,400－412.

Deci,E. L.,& Ryan,R. M.(2000). The "what" and "why" of goal pursuits: Human needs and the self－determination of behavior. *Psychological Inquiry*,*11*(4),227－268.

Green,S. B.,& Yang,Y.(2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*,*74*,121－135.

Muthén,L. K.,& Muthén,B. O.(2010). *Mplus user's guide*(6th ed.). Los Angeles: Muthén & Muthén.

Raykov,T.,Dimitrov,D. M.,& Asparouhov,T.(2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*,*17*,265－289.

Raykov,T.,& Marcoulides,G. A.(2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*,*11*,621－637.

Raykov,T.,& Shrout,P. E.(2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling*,*9*,195－212.

Revelle,W.,& Zinbarg,R. E.(2009). Coefficients alpha,beta,omega,and the glb: Comments on Sijtsma. *Psychometrika*,*74*,145－154.

Sijtsma,K.(2009a). On the use,the misuse,and the very limited usefulness of cronbach's alpha. *Psychometrika*,*74*,107－120.

Sijtsma,K.(2009b). Reliability beyond theory and into practice. *Psychometrika*,*74*,169－173.

Vangeneugden,T.,Molenberghs,G.,Laenen,A.,Geys,H.,Beunckens,C.,& Sotto,C.(2010). Marginal correlation in longitudinal binary data based on generalized linear mixed models. *Communications in Statistics－Theory and Methods*,*39*,3540－3557.

Wen,Z.,Marsh,H. W.,& Hau,K. T.(2010). Structural equation model of latent interactions: An appropriate standardized solution and its scale－free properties. *Structural Equation Model*,*17*,1－22.


(下转第 52 页)

续附表 3

```
Wf1 = Uf1/v1;! 计算三个组因子的 ωh。
Wf2 = Uf2/v1; Wf3 = Uf3/v1;
EVCg = U2/(U2 + U3 + U4 + U5);! 计算因子 G 的 EVC;
EVC1 = U3/(U2 + U3 + U4 + U5);! 计算因子 F1 的 EVC;
EVC2 = U4/(U2 + U3 + U4 + U5);! 计算因子 F2 的 EVC;
EVC3 = U5/(U2 + U3 + U4 + U5);! 计算因子 F3 的 EVC;
OUTPUT: STANDARDIZED; CINTERVAL;
```

# Estimating Test Reliabilities Using Mplus

Wang Mengcheng[1], Ye Baojuan[2]
(1. Center for Psychology and Brain Science, Guangzhou University, Guangzhou 510006;
2. School of Psychology, Jiangxi Normal University, Nanchang 330022)

**Abstract:** Cronbach's alpha as a reliability estimate resides in many shortages. To overcome those drawbacks, researchers have introduced a plentiful of alternative estimators of reliability. At present, however, those alternative reliabilities are not available in popular statistical packages(e. g., SPSS), leading to less adopted by applied researchers. In order to narrow the gap between theory and practice, in this paper, several popular estimators of reliability(composite reliability; single item reliability and reliability $\omega_h$) were demonstrated though analyzing an example using Mplus.

**Key words:** reliability; Cronbach's alpha; composite reliability; reliability $\omega_h$; Mplus

(上接第 47 页)

Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, *17*, 66 - 81.

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for $\omega_h$. *Applied Psychological Measurement*, *30*, 121 - 144.

# A Comparison of Two Methods Estimating the Confidence Interval of Composite Reliability of Multidimensional Test

Yang Qiang[1], Ye Baojuan[2], Wen Zhonglin[3]
(1. School of Education, Jiangxi Normal University, Nanchang 330022;
2. Lab of Psychology and Cognition Science of Jiangxi, School of Psychology, Jiangxi Normal University, Nanchang 330022;
3. Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631)

**Abstract:** There are two methods to estimate the confidence interval of composite reliability of a multidimensional test: Bootstrap method and Delta method. A simulation study was conducted to compare these two methods. The simulation results indicated that the difference between the standard errors obtained by Delta method and Bootstrap method was ignorable. Because the result from Bootstrap method can be treated as the true value, whereas Delta method is much simpler than Bootstrap method, we recommended that Delta method could be adopted to estimate the confidence interval of composite reliability of a multidimensional test. An example of a multidimensional test was used to illustrate how to calculate composite reliability and its confidence interval by using Delta method.

**Key words:** multidimensional test; composite reliability; confidence interval; Bootstrap method; Delta method