

# 经典测量理论中的信度理论与信度系数

## ——与张敏强教授商榷

席仲恩,汪顺玉

(重庆邮电大学 外国语学院,重庆 400065)

**摘要:**文章指出了《教育测量学》一书关于信度理论和信度系数值得商榷之处:系数的定义表述不严谨,未区分信度系数的定义式和估计式,对信度系数定义的评注不符合科学现状,定理的证明含有错误,证明的条件不全,符号使用混乱;建议在入门教材中删除有关信度系数的证明过程。

**关键词:**经典测量理论;信度;信度理论;信度系数

中图分类号:G642

文献标志码:A

文章编号:1008-5831(2007)06-0116-05

教育测量学中的经典测量理论是一套关于明显学业表现的测量学理论,它主要由信度理论、项目分析理论和分数等值理论组成。经典理论中的信度理论指的是20世纪初由斯皮尔曼和布朗发展出的一套关于信度系数估计的理论。由于学术传统习惯的原因,这套理论至今仍有很强的生命力,因此,也是心理测验和教育测量学教材的必讲内容,更是入门教材的核心内容之一。由张敏强教授编著、人民教育出版社出版的“教育科学分支学科丛书”之二——《教育测量学》<sup>[1]</sup>——当然也分专章介绍了这一理论。在阅读过程中,我们发现该教材关于经典信度理论的介绍有多处疑点,兹提出来,和张教授商榷,也向广大的读者请教。

### 一、关于信度系数的定义及其评注

《教育测量学》第102页把信度定义为“真分数方差与观察分数方差之比例”,语言欠规范,用词欠准确。确切些说,这里的“信度”应该是“信度系数”。用规范的现代汉语表达,信度系数的定义应该是:真分数方差与观察分数方差之比;或者,真分数方差占观察分数方差的比例。

我们之所以强调“信度”和“信度系数”之间区别的原因是,在现代教育测量学中,“信度”这个术语已经变成一个表示类属概念的词,其中包括信度系数、测量标准误、信噪比、分类一致性等<sup>[2]</sup>,即使在经典理论中,“信度”就已经包含了信度系数和测量标准误两层意思<sup>[3]</sup>。

收稿日期:2007-09-22

作者简介:席仲恩(1958-),男,陕西人,重庆邮电大学外国语学院副教授,博士,主要从事语言测试中的计量学理论研究;汪顺玉(1965-),男,重庆邮电大学外国语学院副教授,博士,主要从事语言测试理论与实践研究。

教材第 102 页把信度系数形式化地定义为

$$r_{XX} = \frac{S_T^2}{S_X^2} \quad (1)$$

也有些欠妥。首先,做为信度系数的定义,式中的参数最好用希腊字母表示,而不是用拉丁字母,当然,为了方便,也可以全部用拉丁字母,但最好加以说明,而且要统一。该教材的主要问题出在符号约定的不统一上。一会儿用希腊字母表示参数,用拉丁字母表示估计(第 51 页,第 107 页等),一会儿又不加区分。其次,信度系数的下脚号也最好不用同一个拉丁字母,要么用两个截然不同的拉丁字母,要么用同一个字母及该字母加上其他标号。如果没有证明过程,用同一个字母做脚号也可以。如果有证明过程,用同一个字母做脚号容易引起混淆(下节讨论的错误可能就是由于这个原因造成的)。因此,就本教材的符号体系而言,建议把(1)式用作信度系数的估计式,把下式用作信度系数的定义式

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2} \quad (2)$$

式中  $\rho_{XX'}$  表示信度系数,  $\sigma_T^2$  表示真分数的方差,  $\sigma_X^2$  表示观察分数的方差。

该教材对于信度系数定义的评注是:“经典理论定义的信度是一个理论上构想的、不可操作的概念,因此大部分实际中的信度资料都以相关系数来表示。”这个评注在 20 世纪 60 年代之前是成立的,此后,情景就有所不同。评注科学内容应该评它的现状,而不是它的历史,如果我们的重心不是历史的话。事实上,60 年代之后,克伦巴赫及其同仁就改写了这个历史,用概化理论既解放了经典信度理论的条件限制,也拓宽了经典信度理论的适用范围<sup>[4]</sup>。在满足内部一致性的条件下,信度系数的定义自身就是一个相关系数的函数,不过是组内相关系数,而不是组间相关系数。发现了这层关系,就没有必要假定平行测验条件得到满足(但实际上并没有满足)的前提下,用平均组间相关系数间接估计,而是直接根据组内相关系数的方差分量式定义信度系数的模型,这就是概化理论的全部内容。不幸的是,该教材并没有介绍概化理论这一称做经典理论现代版本的测量理论的任何内容。

概化理论充分挖掘方差分量分解技术的潜力,不仅估计出了真分数方差,也估计出了各种误差源的方差,而真分数方差与误差方差之和便是观察分

数的方差。最近,笔者在统一信度理论时,还发现了一种远比概化理论更简单、更明了的信度系数估计方法,完成了经典理论、概化理论、强真分数理论以及项目反应理论在信度理论上的统一,由于与本文关系不大,在此不便赘述。但无论怎么说,从概化理论的角度看,教材中的信度系数定义是完全可以操作的,而不是不可以操作的。

## 二、关于组内—组间相关系数等价的证明

在经典理论发展的早期,由于研究者并不知道信度系数用组内相关系数(关于组内相关系数的原始定义以及组内相关系数可以表示成两个方差之比最少可以追溯到 R. A. Fisher 1925 年出版的统计学著作<sup>[5]</sup>)的函数定义比用组间相关系数、组间相关系数的函数或平均组间相关系数的函数定义更方便,条件也更容易满足,于是才提出严格的平行测验条件,以便通过组间相关系数间接地估计组内相关系数。不难证明,当一组受考在两次测验上的分数的平均数和方差都相等时,组间相关系数与组内相关系数相等。可是,教材第 102 页下方关于这个关系的证明显然是错误的。请看教材的证明过程:

对一测验 g:

$$\begin{aligned} r_{X_g X_g} &= \frac{S_{X_g X_g}}{S_{X_g} S_{X_g}} = \frac{S_{(T_g + E_g)(T_g + E_g)}}{S_{X_g}^2} \\ &= \frac{S_{T_g}^2 + S_{T_g E_g} S_{E_g T_g} + S_{E_g}^2}{S_{X_g}^2} \\ &= \frac{S_{T_g}^2 + 0 + 0 + 0}{S_{X_g}^2} \\ &= \frac{S_{T_g}^2}{S_{X_g}^2} = r_{XX} \end{aligned}$$

显然,以上的证明过程是错误的,因为其中的  $S_{E_g}^2 \neq 0$ 。因此,以上的证明(为了便于比较,此处也用和教材一致的符号)结果应该是

$$\begin{aligned} r_{X_g X_g} &= \frac{S_{X_g X_g}}{S_{X_g} S_{X_g}} = \frac{S_{(T_g + E_g)(T_g + E_g)}}{S_{X_g}^2} \\ &= \frac{S_{T_g}^2 + S_{T_g E_g} S_{E_g T_g} + S_{E_g}^2}{S_{X_g}^2} \\ &= \frac{S_{T_g}^2 + 0 + 0 + S_{E_g}^2}{S_{X_g}^2} \\ &= \frac{S_{T_g}^2 + S_{E_g}^2}{S_{X_g}^2} \\ &= \frac{S_{X_g}^2}{S_{X_g}^2} = 1 \end{aligned}$$

这样的结果是显然的,因为一个随机变量跟它自身

的相关系数必然是1。事实上,严格的平行测验条件就是教材第一章第三节中的“真分数模型”,但其中的定义(2) $\bar{T} = \bar{X}$ 是错误的,正确的表示应该是 $T = EX$ 。但是,这个证明并不是组间相关系数和组内相关系数在平行测验条件满足前提下的等价性的证明。教材中出现错误的原因可能是理解上的,也可能是符号使用不当造成的。以下证明过程才是关于等价性的证明。

设 $X$ 和 $X'$ 为同一测验在同一组受考上的两次测量结果,假定前一次测量不影响后一次测量,则

$$\begin{aligned}\rho_{XX'} &= \frac{\sigma(X, X')}{\sigma(X)\sigma(X')} \\ &= \frac{\sigma(T + E, T + E')}{\sigma_X \sigma_{X'}} \\ &= \frac{\sigma_T^2 + \sigma_{TE'} + \sigma_{ET} + \sigma_{EE'}}{\sigma_X \sigma_{X'}} \\ &= \frac{\sigma_T^2 + 0 + 0 + 0}{\sigma_X^2} \\ &= \frac{\sigma_T^2}{\sigma_X^2}\end{aligned}$$

等价性得证。对于平行测验,其证明过程也一样。

### 三、组内—组间相关系数等价的条件

严格说来,以上证明并不完整,而只是按照教材的“范例”,把证明过程叙述了一遍。要使证明过程完整,还必须补上等价条件。这些条件称做平行测验条件,即教材第50-51页上的“真分数模型”, $\rho_{TE} = 0$ (等价于 $s_{TE'} = 0, s_{ET} = 0$ ),(等价于 $\sigma_{E_1E_2} = 0, \sigma_{EE'} = 0$ ),外加 $X = T + E, X' = T + E'$ ,以及 $s_X = s_{X'}$ 。显然,只有附加上这些条件,关于等价性的证明才能完成。可见,组间相关系数与组内相关系数等价是有条件的,而不是无条件的。如果实践中这些条件不能在足够高的程度上得到满足时,仅是通过组间相关系数估计信度系数,其结果便是无效的。

不难看出,教材中的另一个显著问题是没有提供等价性证明所必须的条件。缺少了这些条件,不仅第102页的再测相关系数与信度系数的等价性无法得以证明,对于第103页的几个推论的证明也无法完成。

### 四、符号使用不一致

符号使用不一致是本教材又一个值得商榷的方面。可以说,整个教材中的符号使用都是混乱的,本节我们仅举出与信度问题讨论有关的符号使用不一致之处。

关于信度问题的讨论离不开“真分数”模型和平行条件。在教材第50-51页关于真分数模型和平行条件的介绍中,相关系数、方差和协方差分别用希腊字母 $\rho$ 和 $\sigma$ 加上相应的脚号表示,但是到了第102-104页,却用了拉丁字母 $r$ 和 $S$ 脚号表示;第50页的数学期望算子用大写花体拉丁字母 $E$ 表示,第51页却改用小写斜体希腊字母 $\varepsilon$ 表示;在第102-103页的证明中, $S$ 已经扮演了数学期望算子的角色。这种做法的直接结果是很难让读者适从。可见,作者在写这部教材时可能参考了不同作者的文献,这些作者可能用了不同的符号体系,但该教材的作者并没有把不同文献作者的不同符号系统一起来。

### 五、信度概念混乱

教材中的信度概念相当混乱,有些证明也证而不明。在第106-107页证明斯皮尔曼—布朗分半矫正公式时,无故引入了信度的另一个条件等价定义:

$$\rho_{XX'} = \rho_{XT} = \frac{\sigma_T^2}{\sigma_X^2} \quad (3)$$

原本这个等价定义的目的对真分数进行点估计(关于这一点目前学界尚存在争议),此外,别无其他用途,与斯皮尔曼—布朗矫正公式的证明没有任何关系。而且,教材中对于矫正公式的证明过程几乎全部错误,证而不明。此外,把第51页的公式1-17搬到了第106页,就把原来的 $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$ 变成了 $\sigma_X^2 = \sigma_T^2 + \sigma_E$ ,最后一项掉了一个平方“2”。

实际上,矫正公式的证明非常容易。设测验 $X$ 被分成两个严格的平行子测验 $X_1$ 和 $X_2$ ,就观察分数而言,则

$$X = X_1 + X_2 \quad (1a)$$

即总测验的分数等于两个子测验分数之和。根据方差的性质,给1)式两边取方差得

$$\sigma_X^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + 2\rho_{X_1X_2}\sigma_{X_1}\sigma_{X_2} \quad (2a)$$

由于两个子测验平行, $\sigma_{X_1} = \sigma_{X_2}$ ,于是2a)式可以缩减为

$$\sigma_X^2 = 2\sigma_{X_1}^2(1 + \rho_{X_1X_2}) \quad (3a)$$

同样,对于真分数而言,

$$T = T_{X_1} + T_{X_2} \quad (4a)$$

对4a)式两边取方差得

$$\sigma_T^2 = \sigma_{T_{X_1}}^2 + \sigma_{T_{X_2}}^2 + 2\rho_{T_{X_1}T_{X_2}}\sigma_{T_{X_1}}\sigma_{T_{X_2}} \quad (5a)$$

由于两个子测验平行, $\sigma_{T_{X_1}} = \sigma_{T_{X_2}}$ ,且 $\rho_{T_{X_1}T_{X_2}} = 1$ ,于是5a)式可以缩减为

$$\sigma_T^2 = 4\sigma_{T_{X_1}}^2 \quad (6a)$$

6a) ÷ 3a) 得

$$\frac{\sigma_T^2}{\sigma_X^2} = \frac{4\sigma_{T_{X1}}^2}{2\sigma_{X1}^2(1 + \rho_{X1X2})} \quad 7a)$$

根据信度系数的定义,  $\rho_{X1X2} = \sigma_{T_{X1}}^2 / \sigma_{X1}^2$ , 代入上式并化简得

$$\frac{\sigma_T^2}{\sigma_X^2} = \frac{4\rho_{X1X2}^2}{2(1 + \rho_{X1X2})} = \frac{2\rho_{X1X2}}{1 + \rho_{X1X2}} \quad 8a)$$

斯皮尔曼—布朗分半信度矫正公式得证。即

$$\rho_{XX'} = \frac{2\rho_{X1X2}}{1 + \rho_{X1X2}} \quad (4)$$

如果按照斯皮尔曼 1910 年文章的原意, 以上的证明过程应该如下:

$$X = \frac{1}{2}(X_1 + X_2) \quad 1b)$$

即总测验的分数等于两个子测验分数之和。根据方差的性质, 给 1b) 式两边取方差得

$$\sigma_X^2 = \frac{1}{4}(\sigma_{X1}^2 + \sigma_{X2}^2 + 2\rho_{X1X2}\sigma_{X1}\sigma_{X2}) \quad 2b)$$

由于两个子测验平行,  $\sigma_{X1}^2 = \sigma_{X2}^2$ , 于是 2b) 式可以缩减为

$$\sigma_X^2 = \frac{1}{2}\sigma_{X1}^2(1 + \rho_{X1X2}) \quad 3b)$$

同样, 对于真分数而言,

$$T = \frac{1}{2}(T_{X1} + T_{X2}) \quad 4b)$$

对 4b) 式两边取方差得

$$\sigma_T^2 = \frac{1}{4}(\sigma_{T_{X1}}^2 + \sigma_{T_{X2}}^2 + 2\rho_{X1X2}\sigma_{X1}\sigma_{X2}) \quad 5b)$$

同样, 由于两个子测验平行,  $\sigma_{X1}^2 = \sigma_{X2}^2$ , 且  $\rho_{T_{X1}T_{X2}} = 1$ , 于是 5b) 式可以缩减为

$$\sigma_T^2 = \sigma^2 T_{X1} \quad 6b)$$

6b) ÷ 3b) 得

$$\frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma^2 T_{X1}}{\frac{1}{2}\sigma_{X1}^2(1 + \rho_{X1X2})} \quad 7b)$$

根据信度系数的定义,  $\rho_{X1X2} = \sigma_{T_{X1}}^2 / \sigma_{X1}^2$  代入上式并化简得

$$\frac{\sigma_T^2}{\sigma_X^2} = \frac{\rho_{X1X2}}{\frac{1}{2}(1 + \rho_{X1X2})} = \frac{2\rho_{X1X2}}{1 + \rho_{X1X2}} \quad 8b)$$

实际上, 教材中并没有关于分半信度矫正公式的证明。那么, 教材的第 106 - 107 页证明的是什么呢? 它证明了两个定理。

定理 1: 如果满足独立性条件(即误差分数与真

分数无关)时, 真分数方差与观察分数方差之比(信度系数)等于真分数与观察分数之间的相关系数的平方。关于该定理的用处本节开始已经论述过, 此处无需赘述。

定理 2: 如果满足平行测验条件时, 真分数方差与观察分数方差之比(信度系数)等于两个平行测验的观察分数之间的相关系数。

定理 2 是用两个平行测验上的观察分数来估计信度系数的理论依据。对于经典信度理论的批评, 主要是针对定理 2 的, 因为它要求的平行条件在实践中很难满足。概化理论也正是在这一点上解放了经典理论中的信度理论。

定理 1、定理 2 再加上信度系数的定义, 就是我们常说的经典理论中关于信度系数的三个等价定义。就学科发展的水平而言, 定理 2 今天已经没有任何必要, 定理 1 的用处也非常有限, 斯皮尔曼—布朗矫正公式也没有多大用场。关于这几点, 国内学界还没有认识到。我们相信, 随着心理和教育测量学知识的普及, 很快国内教育和心理测量界的同仁就会认识到这几点。

## 六、关于信度系数模型的证明

证明对于透彻地理解模型含义非常重要。但是, 对教材而言, 还应该考虑学生和授课老师能否看懂证明过程, 是否有足够的基础读懂证明过程。关于信度系数模型的证明所用数学知识不算很多, 主要是概率论中的方差和协方差算子。但是, 据笔者所知, 国内教育和心理相关专业的学生通常并不学这部分内容, 即使到了硕士和博士阶段也不学, 他们主要学一些简单的一元函数微积分学和应用统计知识。所以, 这样的证明过程, 教育和心理学科的学生很难读懂, 很多授课老师, 如果是纯正的心理学或教育学专业出身的, 也很难读懂。鉴于这种情况, 我们以为, 在像《教育测量学》这样的入门教材中, 对于信度系数模型的证明似乎没有多大的必要, 即使证明过程没有任何错误。

## 七、结语

教材是人类间接经验的重要源泉之一, 也是传播人类文明知识的高效途径。教材的主要读者是学生, 是对教材所讲内容的正确性缺乏足够判断力的学生。他们往往把教材当做“圣经”, 几乎百分之百地相信。因此, 教材的内容应该绝准确。一部教材个别地方出现疏忽在所难免, 但是, 如果在实质内容方面出了错,

其波及面就大了。如果再加上连锁反应,后果不堪设想。我们真心地希望,就文中讨论的问题而言,我们自己错了,但终因我们无法说服自己,只好公开和张教授商榷,也向广大读者求教。如果真的是我们错了,我们定会衷心感谢任何一位指出者;如果万一我们的所言没有错误,那么,还请大家携起手来,共同消除由于这本教材中的错误对于全社会已经造成的非我们所期望的影响。

#### 参考文献:

[1]张敏强.教育测量学[M].北京:人民教育出版社,1998.

[2]HAERTEL E H. Reliability [M]// R. L. BRENNAN. Educational Measurement (4th ed.). USA: American Council on Education /Praeger, 2006:65 - 110.

[3]STANLEY, J C RELIABILITY [M]// R. L. BRENNAN. Educational Measurement (2th ed.). Washington D. C. : American Council on Education, 1971:365 - 442.

[4]CRONBACH L J, RAJARATNAM N, GLEESER G C. Theory of Generalizability: A liberalization of Generalizability Theory [J]. British Journal of Statistical Psychology, 1963;16,137 - 163.

[5] FISHER R A. Statistical Methods for Research Workers [M]. London: Oliver & Boyd, 1925:177 - 211.

## Reliability Theory and Reliability Coefficient in Classical Test Theory: A Discussion with Prof. Zhang Minqiang

XI Zhong-en, WANG Shun-yu

(College of Foreign Languages, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** This paper elaborates on the doubtful points with regard to reliability theory and reliability coefficient in the textbook *Educational Measurement*. These points include loose definition of reliability, unjust remarks on the definition, erroneous proof of theorems and inadequacy of necessary conditions for the proof, inconsistency in notation and the use of symbols. It suggests that in textbooks at the introductory level, mathematical proofs of theorems concerning the estimation of reliability coefficient be cut out.

**Key words:** classical test theory; reliability; reliability theory; reliability coefficient

(责任编辑 胡志平)