

# 老年健康数据质量评估问题探讨<sup>\*</sup>

顾大男

(北京大学 人口研究所, 北京 100871)

**摘要:** 评价老年健康数据的标准可分为效度、信度和准确度。效度就是用来测度老年健康的标识或变量是否正确恰当地度量了想要研究的内容。信度是指度量方法的一致性程度或排除随机误差的程度。准确度是效度和信度的统一和综合,是观察值与真值间的差距。在老年健康数据质量的评价中,我们可以应用诸如单变量描述统计、交叉表分析、均数分析、方差分析、敏感性分析、相关分析、多元回归分析、因子分析、内部一致性分析等众多统计分析方法和诸如条目应答理论等特殊模型对构成老年健康内容的各个变量和指标进行各种效度和信度评估。

**关键词:** 老年健康; 数据质量评估

中图分类号: C913.6 文献标识码: A 文章编号: 1000-4149(2001)02-0038-06

## A Discussion about Assessment of the Quality of Elder's Health Data

GU Da-nan

(Population Institute, Beijing University, Beijing, 100871)

**Abstract:** The criterion of the quality of elder's health data includes the validity, the reliability and the accuracy. The validity is used to measure whether the health variables meet the standards what the researcher wants to study. The reliability is the consistency of the measurements or the level of removing the random errors. The accuracy is a disparity between the observing numerical value and the reality's. It is the combination of validity and reliability. The descriptive statistic analysis, crosstab mean analysis, variation analysis, sensitive analysis, correlation, multiple regression, factor analysis, internal consistency analysis and some special model such as IRT (Item Response Theory) can be applied in the assessment of the validity and reliability of elder's health data.

**Key words:** the elder's health; the assessment of the quality of data

一般地,我们是通过调查或某些医院诊所等的有关记录来获得老年健康资料的。问题是多数研究老年健康的数据在使用前,没有经过质量评估,而是直接使用从中得出有关老年发病率、护理、照料、身

收稿日期: 1999-11-23

作者简介: 顾大男(1965-),男,江苏苏州人,北京大学人口研究所博士生,南京人口管理干部学院副教授。主要研究方向: 人口学方法及其应用,老年健康与长寿。

\*作者感谢导师曾毅教授和副所长柳玉芝副教授在最后成文时提出的宝贵意见。

体功能、心理、社会参与等一系列健康方面的指标数值并进行分析得出结论。显然,这些结论免不了存在准确性问题,从而提出的某些提高老年健康的对策必然存在偏差。因此,为了增强我们老年健康数据质量意识,为了对老年健康作出正确判断,作者认为对老年健康数据质量进行探讨和分析既有理论意义更具实际价值。

## 一、老年健康数据质量评价标准

### (一)健康数据的效度

#### 1. 效度的概念

最简单地讲,老年健康数据的效度就是用来度量老年健康的变量或指标是否恰当,是否正确地度量了想研究的内容<sup>[1]</sup>。效度可细分为内容效度、标准效度、建构效度、增量效度、演绎效度等。

#### 2 效度的主要内容

(1)内容效度是整个健康数据效度的中心和主要方面。它是指某一健康指标中的条目(Item)或某一健康概念中的一组指标变量是否代表了一个给定的老年人口健康内容中的主要方面<sup>[2-4]</sup>。通俗地讲就是“所有与健康相关的概念能否在度量方法中得以体现?”以及“某一变量(比如抑郁)的各个方面是否都可由相应的一个变量来与之对应?”

(2)标准效度表明某一度量健康的方法是否与“金标准”方法高度相关。比如“一套新的度量抑郁的方法是否与‘金标准’高度相关?”“一个条目数较少的度量身体功能方法是否与一个条目数较多但已被证明有效的度量方法高度相关?”“用一套度量健康感知的方法得到的结果能否很好地预测以后年份健康服务的利用情况?”等都是这方面的内容。

(3)建构效度(construction validity)就是用来度量健康的方法是否与假设中的其他方法相关。建构效度中用得最多的是趋同效度(Convergent validity),即度量同一类健康概念的不同变量所得到的结果应该是高度相关的。比如,身体功能变量、户外活动能力变量、身体能力的满意度变量间应该存在至少中等相关的关系(相关系数一般在0.4以上),因为它们都是用来评价身体功能的变量。建构效度中还有识别效度。它指与不相关的变量间的相关性是否要比与相关变量间的相关性为低。比如,身体功能类变量与抑郁或孤独类变量间不应该具有较高的相关性;再比如,身体功能类变量与流动类变量间的相关性应大于它与心理类变量间的相关性。由于健康是一个很广泛而又复杂的体系,因此,需要从很多方面

和角度来加以考察。这时仅仅用上述两种效度是不够的,需要用一种较为复杂的方法。坎贝尔(Campbell)和菲斯克(Fiske)把趋同效度和识别效度相结合提出了一种多特征——多方法途径(Multitrait-multimethod approach)<sup>[5]</sup>。它实际上是一种矩阵法。这矩阵中包含了由两个或更多方法来度量的两组或以上的健康概念(特征)间的相关。比如,抑郁自评变量与调查者测评变量间的相关性是否要比抑郁自评变量与焦虑自评变量间的相关性高?由于几乎所有度量老年健康方法的目的之一不外乎考察疾病、治疗对老年人身体功能和状况的影响程度,因此,我们要检验这些方法能否把一个患有某一特殊疾病的已知老年群体从一般老年群体中区别开来。这种效度我们称之为已知群效度。比如,老年病人的生理功能应该比一般老年人群来得差。患有严重疾病的老人的健康状况应该比病情较轻的老人的健康状况差。基于这些已知群的认识,我们就可以来评价所用指标的正确性。建构效度还包括因子效度,即考察反映同一健康内容的一组变量在因子分析时是否负载在同一因子上。

(4)增量效度主要包括以下两个基本方面:一个条目较多的方法是否比条目较少的方法所揭示的信息多得多?在经费许可下,追加的一组新的健康变量所提给的追加信息是否比原先变量所提供的信息多得多?比如,增加一个流动变量是否增加了很多原有身体功能变量没有揭示的信息?增量效度的高低主要基于其预测能力的大小。一般用回归分析中因变量被解释的比例变化来确定。

(5)演绎效度是指同一方法对某一特殊老年群体(伤残老年)所体现出来的效度是否对其他老年群体也有效。对不同老年群体而言,其平均健康水平肯定是不一样的,但不同老年群体中,变量间相关模式应该是类似的、稳定的。

### (二)信度

信度或可靠性是指度量方法的一致性程度或排除随机误差的程度。信度的高低由信度系数或称可靠性系数来表示。这种系数等于排除随机误差变异后的变异占全部变异的百分比<sup>[6]</sup>。在不同老年群体间进行比较时,信度要求在0.5以上;在老年个体间比较或对同一个老年人在不同时间上比较时,要求在0.9以上<sup>[7]</sup>。纽纳利(Nunnally)提出在研究起步阶段,0.5已足够<sup>[8]</sup>;但在后期研究中,至少要在0.8以上。以日常生活自理能力(Activities of daily living

ADLs) 为例, 美国医疗效果研究项目 (Medical Outcome Study) 中的 ADL 信度为 0.92<sup>[9]</sup>; 加拿大马尼托巴省 65 岁老年人调查的数值为 0.89<sup>[10]</sup>。

信度包括两大类: 内部一致性信度和重复度量信度。前者是用来反映度量同一类健康变量的很多条目间的内部一致程度。它把同一度量方法所有条目中奇异的方差视为误差, 而所有条目共有方差视为总方差中排除了误差以后的方差。用公式表示就是:  $r_{ii} = kr_{ii} / (1 + (k-1)r_{ii})$ 。其中,  $r_{ii}$  称为某一度量方法的内部一致性信度系数;  $r_{ii}$  单个条目的信度系数可用条目间的平均相关系数表示;  $k$  为该度量方法中的条目数。内部一致性信度检验对老年人健康的横断面调查特别有用。重复度量信度是指在相同条件不同时间下, 用同一方法重复对同一老年群体进行健康度量所得结果间的一致性程度。它反映健康数据的稳定性, 因此, 有时也称稳定性(度)。这里, 两次度量的时间间隔不宜太长, 以免老年人的健康水平发生变化, 从而低估信度; 但也不宜太短, 以免老年人记住原先的回答, 从而高估信度。

单个条目的变量也可计算信度系数。在重复度量中, 它可用两个时点上的变量计算一般相关系数、Cohen Kappa 一致性系数<sup>①</sup> 等来获得。在横断面调查中, 根据“两个度量方法间的相关系数永远小于它们各自信度系数乘积的平方根”<sup>[11]</sup> 这一准则, 来推算某单一条目的信度系数。比如, 已知某一方法的信度系数为 0.8, 某单一条目的变量与它的相关系数也为 0.8, 那么, 这一单项目变量的信度系数肯定不小于 0.8。若设计得好, 也可用 Cohen Kappa 一致性系数来获得。

### (三) 准确度

准确度就是观察值与真值之间的差距。差距愈小, 准确度愈高。准确度是效度和信度的综合。效度高, 并不等于准确度高。同样, 信度高, 也不等于准确度高。只有效度和信度都高时, 准确度才高。准确度高, 才说明健康数据的质量高。

## 二、影响老年健康数据质量的主要因素

准确度的反面说法就是误差, 即准确度高, 误差就小, 数据质量就高; 准确低, 误差就大, 数据质量就差。经典统计理论将误差分为三大类, 即系统误差、随机误差(偶然误差)和过失误差。

### (一) 系统误差

又称偏差, 是指在多次观察中按某一确定性规律变化的误差, 它有明显的倾向性。形成系统性误

差的原因很多, 主要有:

#### 1. 建构性误差

它又可分为:

(1) 问卷设计误差。主要是在问卷设计中, 由于缺乏对健康概念的全面理解造成调查内容的片面性和不完整性而引起的。根据世界卫生组织的定义, 健康是一个人在“身体、精神、社会方面所处的良好完整状态, 而不仅仅没有疾病或不虚弱”。身体健康主要由与躯体有关的解剖、体质和生理等内容组成。精神健康主要包括情感、心智以及自我意识、完整的人格、自我实现等方面<sup>[12-13]</sup>。社会健康主要是指人们具有遵守社区规范从事社交活动以及发挥作用的能力<sup>[14-15]</sup>。不仅健康的内容是多维的, 度量这些内容的指标也是多种类型的。比如, 有功能类的指标, 有疾病症状类的指标, 有情感类的指标, 还有多种诊断类方面指标, 等等。因此, 健康是一个庞大的系统。如果对它缺乏全面的理解, 将健康的定义、范围、计算方法等规定得不恰当, 就很难提高数据的内容效度和标准效度。当然, 问卷中的健康内容的设计取决于研究的目的。

(2) 不应答误差。它是指由于部分老年人拒绝接受采访或参与调查而引起的误差。国外的经验显示, 老年人特别是高龄老人不喜欢接受调查<sup>[16-18]</sup>。老年人不应答所产生的偏差大小取决于应答者与不应答者间是否存在系统偏差<sup>[19]</sup>②。如果不应答者是具有某种特定特征的老年人, 那么, 就会产生偏差。西方的经验表明, 不应答的老年人主要是那些年龄较大、健康较差、智力有较大障碍、社会经济地位较低、社会交往少的老人<sup>[20-21]</sup>。

(3) 代答误差。在老年健康调查中, 代答主要是为了减少(2)中提到的不应答偏差而经常被加以应用的一种方法。这种方法主要是出于老年人因健康状况包括诸如听觉、视觉等的丧失和记忆方面原因而让其他知情人来代替老年人本人来回答有关事实性问题。对一个代答的老年人来说, 若代答人的回答中没有体现出比没有代答老人的情况差, 就会产生回答偏差; 另一方面, 若该代答老人的情况确实比

① Kappa 系数用来测量在同一次调查中, 同一问题不同形式所得到的结果间的一致程度; 或不同时间上同一问题在同一老年群体中所得到的结果间的一致性。它用观察到的一致性比例与由于偶然性而预期的一致性比例之差除以非偶然性而预期的一致比例。它一般对定距以下变量而言。

② 即使没有偏差, 在抽样调查中, 若不应答比例过高也会影响样本的代表性。

没有代答者的老年人为好,调查员又会对代答人回答的准确性感到怀疑。

(4)数据缺失。是指老年人在调查中某些条目上没有给出信息。没有回答某一问题的老年人与没有回答该问题的老年人有系统差异,那么,没有回答某一问题会带来统计偏差。比如,在健康长寿研究中,老年人少报了一个活到100多岁但现已去世的同胞兄弟,这时就可能丧失了一个证实生物学因素影响的佐证。这种信息的缺失有时对分析和研究是致命的。另一种是某一信息的部分缺失,即在某一组信息中遗漏了某几项或对某些问题没有作回答。这种缺失造成的危害要比前一种来得小。然而,国内外几乎所有老年健康调查,仅仅基于后一种缺失来计算整个调查的缺失情况,这实际上是一种误区,或者至少说这是不完整的。

(5)偏好。年龄偏好在各个国家特别是发展中国家普遍存在<sup>[22,23]</sup>。老年人年龄的高报<sup>①</sup>,会歪曲老年群体的健康模式。

(6)社会需要应答。这种应答是老年人为了迎合某种社会需要而作出的回答。它没有真实反映老年人自身的健康状况。这种应答多见于意向性的态度和观点问题中。社会需要应答导致社会不需要的低报,社会需要的高报<sup>[24]</sup>。

(7)其他因素。由调查人员对概念理解的不正确或某种习惯,以及工作人员在编码、录入时由于理解不正确等原因也能造成建构性误差。其他因素还包括测量健康特别是身体类健康设备的精确性、调查环境等。

2 内容误差。主要是指无意的误报如年龄、就诊情况的漏报、某些敏感性问题的瞒报等。

### (二)随机误差

又称抽样误差,是指由于偶然因素(随机因素,即抽取样本的随机性)而产生的误差。它在多次观察中误差没有规律,或大或小。当样本规模很大时,随机误差可降低到任意低的水平。

### (三)过失误差

过失误差是指超出预想的、因某种突发性的异常因素所造成的误差。比如,由于工作人员的一时疏忽在记录、编码或录入时发生的误差。

## 三、评价老年健康数据质量的主要方法

高尖端计算机和高功能统计分析软件的应用、学科的不断发展为检验和评价老年健康数据提供了众多方法,主要有:

(一)单变量描述统计,包括频数分布、方差、标准差大小等,主要用以检验单条目的不同效度和信度。

(二)列联表交叉分析主要用以考察度量方法已知群效度。比如,健康老人群体中生活完全自理的比例要比带病老人群体为高。同时,还可以考察缺失值在不同特征老年群体中的分布,以便发现这种缺失是否存在偏差。

(三)均数分析也主要用来考察已知群效度。

(四)方差分析一般可用来检验演绎效度。

(五)多元回归可用来分析增量效度。

(六)敏感性分析用来检验两次调查中的变量间的一致性。若问卷设计得好,在一次性调查中也可用这一方法。

(七)相关分析和关联度分析主要考察变量的趋同效度、识别效度、信度。可用列联相关系数、Cohen Kappa 一致性系数、皮尔逊相关系数、复相关指数等。

(八)因子分析可用来检验同类变量中多条目间的信度。实际操作时,还可以将数据按某一特征分成若干组,单独进行因子分析,以便发现某些信度低(或高)的群体。因子分析还可以对不同老年群体分别作变量的增量效度分析。

(九)内部一致性信度分析也是考察同类变量中多条目间的信度。一般的大型统计应用分析软件中都有这一功能。

(十)条目应答理论 IRT(Item Response Theory)以往主要应用于教育和心理测试<sup>[25]</sup>。只是到最近,IRT才被广泛地运用于健康<sup>[26,27]</sup>。这些研究在把ADLs、IADLs<sup>②</sup>、功能缺损、功能受限等不同类条目合并的基础上用IRT来评估老年人的失能状况。最近两年,把IRT用之与ADLs和IADLs的标准条目集合中来分析老年人的失能状况。由于它在分析中给出每个条目的信度系数和条目偏差,因此,它被视为具有广阔应用前景的检验健康数据质量的方法之一。

(十一)与外部其他同类可获数据进行比较。

另外,我们还可以基于某些变量间的逻辑关系,对它们进行内部逻辑检查并对之进行修正。

## 四、讨论

健康是一个庞大的概念体系,内容复杂、指标繁

<sup>①</sup> 老年人的年龄高报,还会产生抽样框的不准确性从而产生随机误差,见本文(二)随机误差。

<sup>②</sup> IADLs, Instrumental Activities of Daily Living,日常生活照料能力。

多。特定一次调查要面面俱到几乎是不可能的,因此,我们关心的是调查内容中如何在兼顾全面的基础上突出研究重点。同时,由于各个国家的文化差异,对健康特别是心理健康和社会健康的操作性定义及其理解是不同的,因此,在沿用国外的健康内容框架时,应谨慎为好。特别是在跨国比较时更应如此。另外,即使在同一文化下,每一个人特别是老年人对心理健康和社会健康的理解也是不同的。因此,老年人所回答的信息有时会出现偏差,尤其在健康自评中。

为能全面衡量老年群体的总体健康水平,在调查设计中,应将长期住院老人包括在内以减少偏差或随机误差。同时,由于老年群体在调查中的代答和条目缺失比例较大,我们在问卷设计中最好能反映出各条目的代答情况。而且,“不知道”和“拒绝回答”两者性质是不同的,在设计时,应考虑这一问题。国外的经验告诉我们,若老年人在发觉条目答案中有“不知道”或“不清楚”这一选项,就会倾向于选择这样的选项,从而使我们丧失很多信息。较好的办法是在条目答案选项中不给出“不知道”或“不清楚”,而由调查员在调查时用特殊码加以标明。问卷设计中还应尽量避免可能的社会需要应答式的问题存在。另外,由于年龄较大、健康状况差的老年人代答或数据缺失的比例最高,我们在实际调查中要特别给予关注。

实际上,效度还包括一个方面即如何来解释变量得分。它可分为两个问题:一是不同尺度水平的含义是什么?比如,2或5各意味着什么?同时,由于度量老年人健康状态的主要目的之一(特别纵向调查)就是来确定治疗方法或健康服务的效果,所以,了解某一时间内得分的变化是很有意义的,这就是解释变量得分的第二个问题:得分变化的含义是什么?比如,5分或10分的变化意味着什么?举一个简单例子,若一个变量按“非常好、好、一般、差、很差”次序排列,则一般赋值是从1到5或相反。这样,最简单明了。但当变量是由很多条目构成时,变量总体得分的取值范围将很大,而且得分没有原来意义上的含义了,这时就需要将它们转化。一般是将它们转为0~100的得分范围。所有条目的理论最大值的总和为100,所有条目理论最小值的总和为0,其余不同的实际组合则根据相应比例求得。近几年来,曼顿(Manton)和他的同事们研究出了一种隶属等级(Grade of Membership, GoM)方法。每一个人

可以同时隶属于不同的几个组,极大地提高了健康的度量效度<sup>[29]</sup>。

在因子分析中,若变量为二分变量,建议用四次项相关,而不用皮尔逊相关。因为四次项相关假定所观察的二分变量代表未观察到的连续变量呈二元正态分布,它的范围并不受观察变量离中和偏态的边际分布所限制。当然,当变量高度偏态时,应尽量避免用因子分析,但当全部变量呈现一致性偏态时仍可用<sup>[29,30]</sup>。

在与外部数据进行比较时,应首先注意外部数据的质量问题。只有当外部数据质量高时,方能进行。

对于纵向数据,当我们在根据这些数据比较老年健康水平的变化时,也应先对这些数据进行质量检验。国外某些研究表明,在相邻两次间隔二三年的纵向调查中,老年人健康水平的变化与老年人的年龄以及其他特征无关,而与调查者的特征有关。这说明老年人的健康水平没有真正发生变化。这些研究告诫我们,在纵向资料比较时,必须对纵向数据进行质量检验,只有在数据可接受的情况下,我们才可对他们进行对比。

国内对老年健康数据质量的研究尚处于起步阶段,需要我们大家给予更多关注和共同探讨。本文的写作意图之一就是力求抛砖引玉,以促进深层次的研究,并不断完善这一领域的方法论。

#### 参考文献:

- [1] Anastasi, A. Psychological testing, 4<sup>th</sup> ed. New York: Macmillan, 1976.
- [2] Nunnally, J. C. Psychological Theory, 2<sup>nd</sup> ed. New York: McGraw-Hill, 1978.
- [3] Ware, E. J., Manning, W. G., Duan, N., et al. Health Status and the Use of Outpatient Mental Health Services. *American Psychologist* 1984, 39: 1090—1100.
- [4] Ware, E. J. Standards for Validating Health Measures: Definition and Content. *Journal of Chronic Disease*, 1987, 40: 473—80.
- [5] Campbell D. T., and Fiske, D. W. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 1959, 56: 81—105.
- [6] Kerling, F. N., and E. J. Pedhazour. *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston, 1973.
- [7] Helmstader, G. C. *Principles of Psychological Measurement*. New York: Appleton-Century-Crofts, 1964.

- [ 8 ] 同[ 2 ]
- [ 9 ] Stewart, A. L., Ron D. Hays, and John E. Ware, Jr. Methods of Constructing Health Measure. In Stewart A. L. and John E. Ware Jr., Editors. *Measuring Function and Well-being—The Medical Outcome Study Approach*. Duke University Press, 1992.
- [ 10 ] Penning, M. J., Laurel A. Strain. Gender Differences in Disability, Assistance and Subjective Well-being in Later Life. *Journal of Gerontology: Social Science*, 1994, 49, s204.
- [ 11 ] 同[ 2 ]
- [ 12 ] Bradburn, N. M. *The Structure of Psychological Well-being*. Chicago: Aldine Publishing, 1969.
- [ 13 ] Dupuy, H. J. The Psychological General Well-being (PGWB) Index. In *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*, edited by N. K. Wenger, M. E. Mattson, C. D. Furberg, and J. Elinson. New York: Le Jacq, 1984.
- [ 14 ] Ware, J. E., David-Avery, A., And Donald, C. Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Volumes 5: General Health Perceptions. 1978. Publication no. Santa Monica, Calif.: Rand Corporation, R-1987/5-HEW.
- [ 15 ] Spitzer, W. O. 1987. State of Science 1986: Quality of Life and Functional Status as Target Variable for Research. *Journal of Chronic Disease* 40: 465-71.
- [ 16 ] Lowe, F. E., and T. C. McConick. Some Sampling Biases. *Public Opinion Quarterly*, 19, 303—15.
- [ 17 ] DeMaio, T. J. Refusals: Who, Where and Why. *Public Opinion Quarterly*, 1980, 44: 223—33.
- [ 18 ] Herzog, A. R., and W. L. Rodgers. Survey of Older Americans. Some Methodological Investigations. Final Report to National Institute on Aging. Ann Arbor, Mi: Institute for Social Research, 1982.
- [ 19 ] Moser C. A., and G. Kalton. *Survey Methods in Social Investigations*. 2<sup>nd</sup> ed. New York: Basic Books, 1972.
- [ 20 ] Goudy, W. J. Sample Attrition in Multivariate Analysis in Retirement History Study. *Journal of Gerontology*, 1985, 40: 358—67.
- [ 21 ] Norris, F. H. Characteristics of Older Nonrespondents over Five Waves of a Panel Study. *Journal of Gerontology*, 1985, 40: 627—36.
- [ 22 ] Coale, A. J., and E. E. Kisker. Mortality Crossovers; Reality or Bad Data. *Population Studies*, 1986, 40: 389—40.
- [ 23 ] Coale, A. J., Li, Shaomin. The Effect of Age Misreporting in China on the Calculation of Mortality Rates at very High Ages. *Demography*, 1991. Vol. 28, No. 2, p293—301.
- [ 24 ] 同[ 2 ]
- [ 25 ] Lord, F. M. *Applications of Item response Theory to Practical testing Problems*. Hillsdale, NJ: Lawrence Erlbaum, 1980.
- [ 26 ] Haley, S. M., McHomey, C. A., and J. E. Ware 1994. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): I. Unidimensionality and Reproducibility of Hawkins. D. F. Estimation of Nonresponse Bias. *Sociological Methods and Research*, 1975, 3: 461—88.
- [ 27 ] Granger, C. V., Hamilton, B. B., Linacre, J. M., Heinemann, A. W., & Wright, B. D. Performance Profiles of the Functional Independence Measure. *American Journal of Physical Medicine and Rehabilitation*, 1993, 72, 84—89.
- [ 28 ] Manton, K. G., M. A. Woodbury, and D. Tolley. *Statistical Applications Using Fuzzy Sets*. A Wiley-Interscience Publication, John Wiley & Sons Inc, 1994.
- [ 29 ] Rummel, R. J. *Applied Factor Analysis*. Evanston: Northwestern University Press, 1970.
- [ 30 ] Muthen, B. O., & Hofacker, C. Testing the Assumptions underlying Tetrachoric Correlations. *Psychometrika*, 1988, 53, 563—578.

(责任编辑 周皓)