

# 论样本代表性的评估

王晓晖 风笑天 田维绪

(贵州民族大学民族学与社会学学院,贵州 贵阳 550025;南京大学社会学院,江苏 南京 210093)

[摘要] 关于如何评估样本代表性,存在“概率样本说”和“结构相似说”两类观点和方法。前者认为评估样本代表性的关键是判断该样本是不是概率样本,后者认为应依据变量在样本中的分布与在总体中的分布是否相似来评估样本代表性。两类观点和方法各有优缺点。明智的做法是,取长补短,将两类观点和方法结合起来,共同确保和提高样本的代表性。

[关键词] 样本代表性; 概率样本说; 结构相似说

[中图分类号] C915 [文献标识码] A [文章编号] 1003-4145 [2015] 03-0088-05

DOI:10.14112/j.cnki.37-1053/c.2015.03.013

抽样调查始终面临如何评估样本代表性的问题。关于如何评估样本代表性,存在两种主要观点:第一种观点认为,评估样本代表性的关键是判断它是不是概率样本,本文称该观点为“概率样本说”;第二种观点认为,评估样本代表性的基本方法是考察部分变量在样本中的分布与在总体中的分布是否相近,本文称该观点为“结构相似说”。由于第一种观点的关注点是抽样设计和抽样方案实施环节,也就是样本被抽出之前的环节,而第二种观点的关注点是样本被实际抽取之后的问题,故宋子轩、冷燮、陈瑶瑶等人将前者称为“事前保证方法”,将后者称为“事后评估方法”。<sup>①</sup>本文的观点是,无论是“概率样本说”还是“结构相似说”,皆有其缺陷,应同时结合两类方法,才能更好地评估和确保样本的代表性。为论证该观点,本文将分别考察两类方法的特征和缺陷,并探讨如何将两类方法结合起来。在展开论证之前需指出的是,样本代表性是一个程度问题,不是“有”或“无”代表性的问题。本文赞同这种观点,即样本的代表性是指样本反映总体特性的能力及基于样本构造的统计量的优良程度。<sup>②</sup>样本代表性有多高,最终体现在抽样误差的大小上。抽样误差越小,样本代表性越高;抽样误差越大,样本代表性则越低。

## 一、概率样本说及其缺陷

支持“概率样本说”的学者们的基本观点,可以概括为三个:第一个观点是,概率抽样的样本即“概率样本”就是对总体有很好代表性的样本。其理由是概率抽样能完全避免人为干扰,消除人为误差。因此,评估样本代表性的关键,是依据对概率抽样方案和具体抽样程序的考察,判断该样本是否是一个概率样本;<sup>③</sup>第二个观点是,样本代表性的比较,是关于不同抽样设计的抽样效果的比较,比较采用同一种抽样设计抽取的不同样本是无意义的。<sup>④</sup>其理由是,样本代表性的评估归根结底要看统计量(也即是总体参数的估计值)的性质(如无偏性、有效性、一致性<sup>⑤</sup>),而同一抽样设计下的统计量的性质是确定的,因此需要比较的是不同抽样设计下的统计量的性质;第三个观点是,不能依据实际抽取的样本的特征来判断其代表性,具体说,不能通过考察部分变量在样本中的分布与在总体中的分布是否相近来评估某个具体样本的代表性。也即是说,他们

收稿日期:2015-01-12

作者简介:王晓晖(1973—),男,博士,贵州民族大学民族学与社会学学院副教授,主要研究方向为经济社会学、社会研究方法。

风笑天(1954—),男,博士,南京大学社会学院教授,主要研究方向为家庭社会学、社会学研究方法。

田维绪(1973—),男,贵州民族大学民族学与社会学学院副教授,主要研究方向为消费社会学、社会研究方法。

①宋子轩、冷燮、陈瑶瑶《概率抽样条件下样本代表性事后评估方法探讨》,《统计研究》2012年第7期。

②俞纯权、王曰人《论样本的代表性》,《统计与信息论坛》2003年第2期。

③参见俞纯权《关于样本代表性的理解》,《江苏统计》2000年第10期;李文华《社会调查研究中样本的代表性问题探讨》,《决策与统计》2006年第17期;游正林《应该如何评估样本的代表性》,《华中师范大学学报》(人文社会科学版)2009年第3期。

④参见俞纯权《关于样本代表性的理解》,《江苏统计》2000年第10期;李文华《社会调查研究中样本的代表性问题探讨》,《决策与统计》2006年第17期。

⑤关于无偏性、有效性、一致性的详细论述,参见卢淑华编著《社会统计学》,北京大学出版社2009年版。

反对“结构相似说”。他们或者认为依据“结构相似说”来抽取样本,研究者可能因怀疑样本代表性不佳而人为地更换构成样本的元素<sup>①</sup>或者认为“结构相似说”不可行、不科学。<sup>②</sup>上述三个观点相互联系,构成“概率样本说”的基本内容,其中第一个观点是核心。为便于讨论,本文首先以简单随机抽样为例子来回应前两个观点,同时将在文章第二部分回应第三个观点,并通过第三个观点的批判来论证“结构相似说”。

概率样本的突出优点是能够确定抽样分布和抽样误差,从而能对总体特征进行统计推断。抽样分布是样本统计量(如样本均值 $\bar{X}$ 、样本成数 $\hat{P}$ 、样本方差 $S^2$ 等)的分布,是指从规模为 $N$ 的总体中抽出的规模为 $n$ 的“所有样本”的某个统计量的分布。比如就样本均值而言,从规模为 $N$ 的总体中抽取规模为 $n$ 的样本,若采取可放回的简单随机抽样方法,可以抽出 $N^n$ 个样本,而样本均值 $\bar{X}$ 的分布,就是指依据 $N^n$ 个样本计算出的 $N^n$ 个样本均值 $\bar{X}$ 的分布。而依据中心极限定理,在样本规模 $n$ 很大时,样本均值的分布接近于以总体均值 $\mu$ 为数学期望、 $\delta^2/n$ (其中 $\delta^2$ 是总体方差)为方差的正态分布, $\delta^2/n$ 开平方根即为抽样均值的标准误差 $SE = \delta/\sqrt{n}$ 。同时,统计学已经论证过,根据简单随机样本所构造的样本均值 $\bar{X} = \sum x/n$ 、样本方差 $S^2 = \sum (x - \bar{X})^2 / (n - 1)$ 、样本成数 $\hat{P} = \sum x/n$ (此时 $x$ 为虚拟变量,取值为0或1),分别是总体均值 $\mu$ 、总体方差 $\delta^2$ 、总体成数 $P$ 的无偏估计值,且满足有效性、一致性的要求。

但是,我们必须注意的事实是,统计学关于抽样分布、估计值的无偏性、有效性和一致性等的讨论,都是基于“从总体中抽出 $N^n$ 个样本”这样一个理论前提来展开的;而在实际的研究中,我们并没有把 $N^n$ 个样本都抽出来,只是从中抽出一个特定的样本,并依据这个特定样本的信息来推断总体。由于样本具有随机性,在实际抽样之前,我们根本不知道我们将抽中 $N^n$ 个样本中的哪一个样本。而且,这 $N^n$ 个样本的抽样误差是有所不同的,有的样本抽样误差很大,有的抽样误差很小甚至为零。如果我们抽中的那个特定样本的抽样误差较大,则该样本的样本均值 $\bar{X}$ 、样本成数 $\hat{P}$ 、样本方差 $S^2$ 与总体均值 $\mu$ 、总体方差 $\delta^2$ 、总体成数 $P$ 的差值相应的也会比较大。例如某个总体包括6户人,6户人的家庭人口规模分别是1、2、3、4、5、6人,则6户人的户均人口数 $\mu = 3.5$ 人。现从6户人中抽2户人构成的样本,若采用重复抽样,可以抽出36个样本,从而可以计算出36个样本均值。若其中一个样本是由家庭人口数为1和6的这两户构成的,则样本均值 $\bar{X} = 3.5 = \mu$ ,我们可用该样本推断总体均值。然而,如果我们抽中一个由家庭人口数为1和2的这两户构成的样本, $\bar{X} = 1.5$ ,与 $\mu = 3.5$ 相差较大,此时用该样本推断总体就会出现偏差。这个例子说明,从规模为 $N$ 的总体中抽出的规模为 $n$ 的 $N^n$ 个样本中,有的样本的抽样误差比较大,有的样本的抽样误差比较小。就连赞同“概率样本说”的俞纯权教授自己也曾指出“在概率抽样下,由于样本的抽取带有随机性,因此依据不同的具体样本得到的目标量估计值是不同的,从而实际误差依不同的具体样本而异”。<sup>③</sup>总之,样本所固有的随机性,可能会使我们在实际调查中抽出的那个样本是严重偏离总体的样本,<sup>④</sup>或者说是抽样误差很大的样本,该样本对总体的代表性很低,用它去推断总体,有可能会得到错误的结论。

关于样本随机性可能会导致我们抽中严重偏离总体的样本,我们还可以用假设检验逻辑和假设检验中的“弃真错误”来说明。假设检验的基本逻辑是直接检验原假设( $H_0$ ),间接检验研究假设( $H_1$ ),目的是排除抽样误差的可能性。假设已有的数据显示,某地青年的平均初婚年龄是23岁(即总体均值 $\mu = 23$ ),现统计局从当地青年中采用简单随机抽样抽取100人,发现他们的平均初婚年龄为26岁(即样本均值 $\bar{X} = 26$ )。那么,我们能否可以就此下结论,认为该地青年的平均初婚年龄已经推迟呢?显然不能。这是因为出现样本均值 $\bar{X} = 26$ 这样一个结果,有两种可能的原因:第一种原因是,平均初婚年龄并没有推迟,而是因为此次调查刚好抽中一个有偏差的样本(即代表性低的样本),该样本的抽样误差使得 $\bar{X} \neq \mu$ ;第二种原因是平均初婚年龄已经推迟,所以抽样调查的结果 $\bar{X} = 26 \neq 23$ 。那么,究竟是什么原因导致了 $\bar{X} = 26$ 这个结果呢?我们可成立两个相互矛盾的假设来解决这个问题:原假设( $H_0$ )  $\mu = 23$ ,研究假设( $H_1$ )  $\mu \neq 23$ 。原假设的意思是,青年

①俞纯权《关于样本代表性的理解》,《江苏统计》2000年第10期。

②参见俞纯权、王曰人《论样本的代表性》,《统计与信息论坛》2003年第2期;游正林《应该如何评估样本的代表性》,《华中师范大学学报》(人文社会科学版)2009年第3期。

③俞纯权、王曰人《论样本的代表性》,《统计与信息论坛》2003年第2期。

④Firestone, William A. "Alternative Arguments for Generalizing from Data as Applied to Qualitative Research." Educational Research, 1993, No.

平均初婚年龄仍为 23 岁,100 人抽样调查的平均初婚年龄之所以是 26 岁,是因为我们抽中了误差较大的样本;研究假设的意思是,青年的平均初婚年龄已经推迟,因此样本均值为 26 岁。显然,这两个假设是相互矛盾的,否定其中之一,就需承认另外一个。统计学的基本做法是,直接检验原假设( $H_0$ ),如果原假设被否定了,也就是抽样误差导致样本均值不等于总体均值的可能性被排除了,就可以接受研究假设( $H_1$ )。

检验原假设的方法是“概率反证法”,而其基本原理是“小概率事件实际不可能发生的假设”(又称“小概率原理”)。“小概率事件实际不可能发生的假设”有两个含义:第一,若某事件是小概率事件,则在一次观察中它不应该发生;第二,若在某次观察中我们假设的小概率事件发生了,则我们应该否定认为它是小概率事件的这个假设。在实际的运用中,我们首先假设原假设( $H_0$ )是真的,然后根据这个前提推导出这样一个结果,即“在  $H_0$  为真的情况下,我们抽中一个有很大抽样误差的样本,从而使得  $\bar{X} = 26 \neq 23$ ”这一随机事件是一个小概率事件,并根据“小概率事件实际不可能发生的假设”认为它不会发生。小概率的大小是人为确定的,通常选择 1%、5%、10%。选择小概率值之后,就可以确定小概率事件发生的标准,即确定接受域与拒绝域。比如选择了 1% 这个概率,则可将样本均值分布范围划分为三段:  $[\mu - 2.58SE, \mu + 2.58SE]$ 、 $(-\infty, \mu - 2.58SE)$ 、 $(\mu + 2.58SE, +\infty)$  (其中 SE 是样本均值的标准误差)。平均而言,100 个样本均值中,有 99 个将落入  $[\mu - 2.58SE, \mu + 2.58SE]$  这个区间,只有一个会落入  $(-\infty, \mu - 2.58SE)$  或  $(\mu + 2.58SE, +\infty)$  这两个区间。假设检验中的小概率事件,就是指在原假设为真的前提下,抽样误差使得样本均值落入  $(-\infty, \mu - 2.58SE)$  或  $(\mu + 2.58SE, +\infty)$  这两个区间。如果我们抽中的特定样本的样本均值落入了该区域,则应否定“原假设为真”这个前提。然而,依据小概率事件的发生而否定原假设是有可能犯错误的,此错误就是所谓的“弃真错误”,也就是当原假设本来是真实的而我们否定它的时候所犯的误差。之所以会犯“弃真错误”,就是因为我们在抽样时,恰好抽中了一个有较大抽样误差的样本。而导致我们犯“弃真错误”的样本,对总体是缺乏较好代表性的。

总之,从上面的论述我们可以得出一个结论,那就是并非所有概率样本都具有代表性,依据同一抽样设计所抽出的不同样本也由于其抽样误差不同而具有不同的代表性。其原因就在于样本具有随机性,该随机性可能导致我们在特定一次抽样中恰好抽中一个有很大偏差的样本。这也说明,单纯依据对实际抽样方法和抽样程序的考察来判断特定样本是不是概率样本的方法,不足以确保我们在实际研究中抽中的那个特定样本能够具有较好的代表性。

## 二、结构相似说及其缺陷

支持“结构相似说”的学者们的基本观点是,通过考察部分变量在样本中的分布与在总体中的分布是否相近可以评判特定样本的代表性。很多知名的研究方法专家都赞同“结构相似说”。艾尔·巴比认为,一个样本的质量,是指某些特征在样本中的分布与在总体中的分布相同,如果样本的综合特征与总体的同类综合特征相近,则该样本具有代表性。比如,如果总体中女性的比例为 50%,有代表性的样本中女性所占的比例应接近 50%。<sup>①</sup>袁芳和王汉生等认为,在依据事先确定的样本规模和抽样方法抽出样本后,应先对抽出的样本进行评估,其目的是初步检查样本对于总体的代表性,以剔除那些偏差太大的样本,重新抽取代表性高的样本。评估的方法是依据若干容易得到的资料例如年龄、性别、文化程度等,比较样本与总体在这些特征上的相似程度。<sup>②</sup>风笑天指出,将可得到的、反映总体某些重要特征及其分布的资料与样本中同类指标的资料进行比较,若二者之间差别很小,我们对样本的代表性就有较大的信心;若二者之间差别十分明显,则样本的质量和代表性就不一定会很高。用来比较的指标越多越好,而各种指标对比结果越近越好。<sup>③</sup>王宁也认为,问卷调查中所谓样本的代表性,指的是样本的特征与总体的特征的一致性问题,<sup>④</sup>而他所说的特征一致性,就是样本分布与总体分布相似。

依据“结构相似说”对样本代表性进行“事后评估”的具体操作方法目前有两类:第一类是逐项评估方法,也就是依据单个变量来比较样本与总体,计算“平均数代表性检验系数”或“结构代表性检查差异率”。

①[美]艾尔·巴比《社会研究方法》,邱泽奇译,清华大学出版社 2007 年版,第 189 页。

②袁芳、王汉生《社会研究方法教程》,北京大学出版社 1997 年版,第 208 页。

③风笑天《现代社会调查方法》,华中科技大学出版社 2009 年版,第 66-67 页。

④王宁《个案研究的代表性问题与抽样逻辑》,《甘肃社会科学》2007 年第 5 期。

前者的测算公式是:平均数代表性检验系数 =  $[(\mu - \bar{X}) / \mu] \times 100\%$ , 样本平均数与总体平均数差值越小越好; 后者的测算公式是:结构代表性检查差异率 =  $[(P - \hat{P}) / P] \times 100\%$ , 样本成数与总体成数差值越小越好。有多少个变量, 就可以计算出多少个“平均数代表性检验系数”或“结构代表性检查差异率”, 并依据每个计算结果分别评估样本代表性。一般规定, 平均数的代表性检验要求检验系数控制在  $\pm 2\%$  或  $\pm 3\%$  以内, 结构的代表性检查要求差异率控制在  $\pm 5\%$  以内。<sup>①</sup> 第二类是综合评估方法, 即同时依据多个变量来评估样本的代表性, 并且最终计算出一个样本与总体的“整体差异率”。具体做法是, 将样本与总体的各个属性变量之间的差异率加权汇总成一个确切的指数。计算综合值时, 每个属性变量被赋予不同的权重。权重的设计从属性变量相对于抽样调查的目标变量(即研究变量)之间的相关关系入手, 相关程度越高则表明该属性相对于调查目的的重要性就越高, 就赋予较大的权重。<sup>②</sup> 两类方法的差别, 主要在于前者用多个单变量逐项评估样本代表性, 后者则把样本与总体在每个变量上的差异综合成一个指数。

如前所述, 赞同“概率样本说”的学者们反对“结构相似说”。他们有两个理由: 一是认为依据“结构相似说”来抽取样本, 研究者可能因怀疑样本代表性不佳而人为地更换构成样本的元素, 二是认为“结构相似说”不可行、不科学。然而, 这两个理由都有问题。首先, 以研究者可能更换样本元素为由而拒斥“结构相似说”是完全站不住脚的。人们在批判“结构相似说”时, 完全忽视“结构相似说”的前提即概率抽样。“结构相似说”的真正观点是, 在采用概率抽样方法抽出一个特定样本之后, 还要考察部分变量在样本和总体中的分布是否相近。如果相近, 则可认为该样本具有较好的代表性; 如果不相近, 则认为该样本存在较大的偏差, 并应依照先前的抽样方法重新抽出一个样本, 再评估新样本的代表性。如果有研究者因特定样本代表性不佳而人为地更换构成样本的元素, 事实上他已经不再遵循“结构相似说”的要求了; 若他仍然佯称自己的抽样是概率抽样并按“结构相似说”来评估样本, 此为研究者的学术不端。我们只能追究该研究者的学术道德, 而不能追究“结构相似说”本身。事实上, 在实际的概率抽样中, 我们也常看见某些抽样员偷工减料未严格执行概率抽样方案的现象, 难道我们也因此而否定“概率相似说”? 其次, 认为“结构相似说”不可行、不科学的观点同样站不住脚。反对“结构相似说”的学者指出, 由于我们在实际调查之前并不知晓总体特征, 因而无法对某些变量在样本中的分布与其在总体中的分布予以比较。即使能比较某些变量在样本中的分布和在总体中的分布, 当两者的分布差别很小时, 我们亦无从判断差别“很小”所代表的真正意义。<sup>③</sup> 确实, 样本代表性评估面临一个先天的矛盾: 我们是因为不了解总体特征才去开展抽样调查、推断总体特征, 然而在未实际调查之前我们又需要依据总体特征来评估样本代表性, 也就是说, 我们处于要依据我们正欲去推断的未知特征来判断样本代表性的尴尬境地。但是, 在信息已然发达、调查项目和数据库众多的今天, 我们对很多总体并非完全不了解, 我们或多或少掌握了一些基本数据, 这些基本数据通常被称为辅助信息,<sup>④</sup> 笔者称它们为“辅助变量”。我们可以利用这些辅助信息来比较样本特征和总体特征, 比如现在很多社会学研究者常依据人口普查数据中的性别、年龄、教育程度、婚姻等变量来评估样本质量。<sup>⑤</sup> 另外, 变量在样本中的分布与其在总体中的分布的差别“很小”的意义很明确。如果是实际调查之前就考察变量在样本和总体中的分布, 分布差别很小就是指样本中的经验分布与总体中的概率分布相近, 而分布相近意味着抽样误差较小, 则我们更有把握用该样本数据推断总体特征。如果是在依据某样本展开调查之后再用调查所收集的资料来比较变量在样本和总体中的分布, 两种分布之间的差异除了包含抽样误差外, 还包含非抽样误差。若此时经验分布与概率分布之间的差别仍然很小的话, 同样说明样本对总体有较好的代表性。总之, 部分学者对“结构相似说”的批判是站不住脚的。

本文认为, “结构相似说”在理论上不存在问题, 但在实践中确实存在缺陷。在实际的研究项目中, 所需考察的变量少则数十个, 多则上千个。比如中国人民大学开展的“中国综合社会调查”(CGSS)、中国社会科学院开展的“中国社会状况综合调查”(CSS)、中山大学开展的“中国劳动力状况动态调查”(CLDS) 等大型

①李文华《社会调查研究中样本的代表性问题探讨》,《决策与统计》2006年第17期。

②宋子轩、冷雯、陈瑶瑶《概率抽样条件下样本代表性事后评估方法探讨》,《统计研究》2012年第7期。

③游正林《应该如何评估样本的代表性》,《华中师范大学学报》(人文社会科学版)2009年第3期。

④参见俞纯权《辅助信息在抽样调查中的应用》,《浙江统计》2000年第4期;陈培培、金勇进《辅助信息及其在抽样设计中的应用》,《中国统计》2014年第4期。

⑤郝大海《应答率的意义及其他——对中国“高”调查回收率的另一种解读》,《社会学研究》2007年第6期。

社会调查,其问卷问题通常比较多,包含近千个变量。为评估样本代表性,我们必须拥有关于“研究变量”的信息,知晓它们在总体中的分布。<sup>①</sup>然而,我们依据已有信息评估样本质量时,通常能获得的是性别、年龄、文化程度、籍贯等人口统计信息,难于获得关键的研究变量的信息。即使某样本在人口统计信息方面与总体的分布相似,也仅说明该样本在我们已掌握的这些信息上对总体有代表性。至于在关键的研究变量上是否有代表性,我们是没有把握的。但若由于只能依据辅助变量、不能依据研究变量来评估样本代表性而拒斥“结构相似说”,是“将婴儿与洗澡水一起倒掉”的不智之举。毕竟,能够依据某些辅助变量达到对样本代表性一定程度的了解,总比完全不了解更令研究者宽心。因此,在依据“结构相似说”评估样本时,要尽量多收集更多辅助信息,掌握更多的辅助变量,尤其是那些与关键研究变量有较高统计相关性的变量,然后依据这些变量来评估样本代表性。我们所掌握的辅助变量越多,样本与总体在这些辅助变量上的分布越相近,我们对样本代表性的信心就越大。

### 三、概率样本说和结构相似说的结合

鉴于“概率样本说”与“结构相似说”各有其优缺点,我们不应该像赞同“概率样本说”的学者那样,固执己见而片面地否定别的观点和方法,而应吸收两类方法的优点,将两类方法结合起来,取长补短,以便更好地评估和确保样本的代表性。在抽样设计阶段,应尽量多地收集现有的有关研究总体的各类辅助信息,根据研究总体的实际情况,制定一套可行的、能尽量提高抽样精度的抽样方案。在实际抽样阶段,要尽可能地严格执行既定的抽样方案。在实际抽出样本之后,既要实际抽取样本的具体方法和程序进行分析和检查,也要依据已经掌握的信息,考察变量在样本中的分布和在总体中的分布,对样本的代表性作初步的检查和评估。

上述论述所勾画的是两类方法结合的理想状态。现实的抽样调查实践非常复杂,存在众多损害样本代表性的因素。抽样框是否完备或某些元素是否在抽样框中重复出现、非住宅户、无法找到地址、无法接触、两次或两次以上访问无人在家或无人应答、无人符合访问条件、家人或调查对象完全拒访或中途拒访、调查对象外出、生病、聋哑、语言不通等造成的无法接受访问、回收率低时的样本轮换、能够被找到且愿意接受访问的调查对象与不能找到或不愿接受访问的调查对象之间可能存在的显著差异等等这些原因,都将对整个抽样程序和抽样结果造成影响。尤其是降低调查回收率的因素、能够接触到的且愿意接受访谈的调查对象与不能接触到或不愿接受访谈的调查对象之间的差异,可能会增大我们抽中有系统性偏差样本的可能性。<sup>②</sup>更为糟糕的是,在抽样程序复杂的社会研究中,各种造成概率抽样沦落为非概率抽样和降低样本代表性因素的影响的性质和大小往往难于评估。此时我们对样本代表性的信心,恐怕只能来自样本与总体的结构相似了。而为了增加对评估结果的信心,我们除了依据性别、年龄、学历、婚姻状况、职业等基本人口统计信息来评估样本代表性外,更要依据研究变量或与研究变量的相关性较强的样本来评估样本。

(责任编辑:陆影)

<sup>①</sup>Firestone, William A. . "Alternative Arguments for Generalizing from Data as Applied to Qualitative Research." Educational Research, 1993, No. 22.

<sup>②</sup>郝大海《应答率的意义及其他——对中国“高”调查回收率的另一种解读》,《社会学研究》2007年第6期。