社会
2014·1
CJS
第34卷

问卷调查质量研究:应答代表性评估

任莉颖　邱泽奇　丁　华　严　洁

摘　要:随着问卷调查中无应答现象的增多,应答样本代表性问题成为问卷调查研究者的关注焦点。由于自身局限性,应答率无法提供有效的调查质量信息,建构新的应答代表性指标因此就成为研究重点。在综合比较相关研究成果的基础上,本文认为,R指标对于评估并提升问卷调查质量有较好的应用前景,并探讨了R指标的概念界定、计算方法、指标构成和注意事项,同时运用这一指标对"中国家庭动态跟踪调查"2010年初访数据的应答代表性进行了评估。

关键词:社会调查　调查质量　应答代表性　R指标

Research on Survey Quality: Evaluation of the Representativeness of Survey Responses

REN Liying　QIU Zeqi　DING Hua　YAN Jie

Abstract: As non-responses in surveys have increasingly become more common in recent years, the representativeness of survey responses has called attention from survey researchers. Response rate is commonly used as an indicator of survey quality. However, theoretically and empirically there is not necessarily a direct link between response rates and nonresponse biases. So how to get alternative indicators of response representativeness has become a focus in

* **作者1:任莉颖　北京大学中国社会科学调查中心**(Author 1: REN Liying, Institute of Social Science Survey, Peking University) E-mail: isssrenly@pku.edu.cn;**作者2:邱泽奇　北京大学社会学系**(Author 2: QIU Zeqi, Department of Sociology, Peking University);**作者3:丁　华　北京大学中国社会科学调查中心**(Author 3: DING Hua, Institute of Social Science Survey, Peking University);**作者4:严　洁　北京大学政府管理学院**(Author 4: YAN Jie, School of Government, Peking University)

** **本研究受国家自然科学基金资助项目"并行数据和调查数据质量管理"(71171004)资助。**[This research was supported by the project "Paradata and the Quality Management of Survey Data"(71171004), which was sponsored by National Natural Science Foundation of China.]

research.

After reviewing multiple measures of assessing response representativeness, we consider that R-indicator is the most promising compared with others. Regarding its construction, R-indicator is guided by sound theories, based on rich sample-frame data and paradata, and can be obtained by relatively simple algorithm. Regarding its application, R-indicator can be used for comparing different surveys with the same target population, different waves of a panel survey, or times of measurement in different stages of the same survey.

This article introduces the definition of the concept of R-indicator, its computation, composition and limitations. R-indicator was applied to the evaluation of the representativeness of the survey responses in the China Family Panel Studies (CFPS), 2010. We divided the whole fieldwork process into three stages and computed the R-indicator, Maximal Absolute Bias, and partial R-indicators for each stage. The analysis of these indicators led to a discovery that the samples were over- or under-represented in the areas with variations in community attributes, economic development, population density, non-agricultural population ratio, and support from community commissions. Moreover, the seriousness of these problems changed as the survey went on.

Keywords: social survey, survey quality, response representativeness, R-indicator

---

一项严谨的问卷调查,不仅要在设计阶段尽量避免各种抽样误差,还要关注在调查执行阶段因无应答造成的样本代表性问题的影响。随着调查对象流动性和社会对个体隐私保护意识的增强,无应答样本在各国的问卷调查中愈来愈多(de Leeuw and de Heer,2002),社会调查行业使用应答率(response rate)测量调查质量的做法就越来越受到质疑。一般认为,应答率越高,应答样本的代表性就越强,问卷调查的质量也就越高,在一些关于问卷调查的教科书中甚至还设定了"足够好"、"好"、"可接受"等不同级别的应答率区间(Babbie,2007;Singleton and Straits,2005)。但也有研究证明,应答率和代表性之间没有必然的关系(Groves, 2006; Groves and Peytcheva, 2008; Heerwegh, *et al.*, 2007)。在这种情况下,如何评估一项调查的应答代表性便成为问卷调查研究共同关注的问题。

以往研究都从多个维度建构应答代表性指标,本文重点探讨其中

的R指标(R-indicator)。R是"代表性"(representativeness)英文单词的首字母,这个指标由欧洲的研究者在2007年开始研究构建,并致力于被社会调查行业广泛接受。[1]本文首先讨论应答率作为代表性指标的局限性,归纳目前问卷调查研究在建构应答代表性指标方面的进展;然后详细讨论R指标的建构原理和计算方法,并基于"中国家庭动态跟踪调查"(CFPS)2010年的初访数据对其应用性进行评估;最后评述R指标在国内问卷调查中的应用前景和对其发展的促进作用。

一、应答率的局限

问卷调查的科学性来自代表性抽样(representative sampling),而这个概念的用法却是五花八门。克鲁斯卡和莫斯泰勒(Kruskal and Mosteller,1979a,1979b,1979c)曾就此连续发表三篇文章,汇集了他们在非科学文献、科学文献和统计文献中的发现。他们将代表性抽样的用法概括为9类:

1. 代表性抽样可以使结论显得更为科学和可信;
2. 代表性抽样指没有受到选择性压力(absence of selective force);
3. 代表性样本是总体的微缩或镜像,具有和总体相同的构成;
4. 代表性样本指典型或理想的个案;
5. 代表性样本须囊括总体的多样性;
6. 代表性抽样是一个含混的概念,有待澄清;
7. 代表性抽样是一种特定的抽样方法;
8. 代表性抽样可以得出好的估计值;
9. 代表性抽样取决于特别的研究目的。

这些用法从松散到严谨,反映了研究者不同的背景。从统计角度理解,在问卷抽样调查中,获取样本代表性的目的就是希望将样本的统计值合理推论到总体。因此,可以认为代表性样本就是总体的微缩或镜像,或者囊括了总体的多样性。为保证这一点,在调查过程中需要关注的一个重要因素就是,样本是否受到其他选择的压力,如样本自选择。

理论上获取代表性抽样并不等于在实践中会得到代表性应答,评估

1. 关于R指标的构建与应用的论文已相继在国际一些知名的社会调查以及统计的学术会议及学术期刊上发表。想了解R指标的研究进展和成果,可访问http://www.risq-project.eu。

一项问卷调查质量的好坏需要通过特定的指标反映应答样本在多大程度上具有代表性。这个指标至少需要关注两个标准:与总体特征的估计偏差有关;能侦测出调查执行过程中样本应答的非随机选择机制。

应答率是问卷调查中最常用的质量指标之一。在美国民意调查研究协会(AAPOR)给出的标准定义中,应答率是指"完成访问的单元个数与样本中符合资格的单元个数的比值"(AAPOR,2011:5)。应答率简单易懂,便于计算,但社会调查研究已经越来越认识到,应答率并不能用来评估应答代表性。

首先,从理论上看,应答率与无应答造成的偏差没有直接联系。计算应答偏差的公式由两部分组成:一个是无应答样本的比例,另一个是应答者与无应答者在总体均值上的差异,即:

$$B(\bar{Y}_r) = \left(\frac{M}{N}\right)(\bar{Y}_r - \bar{Y}_m)$$

其中 $\bar{Y}_r$ 是应答者的总体均值,$\bar{Y}_m$ 是无应答者的总体均值,N 是总体样本规模,M 是无应答样本的个数,$B(\bar{Y}_r)$ 是应答偏差。只有当应答者和无应答者在总体均值上的差异保持不变时,应答率越高(即无应答率越低),应答偏差才会越小。在前者不确定的情况下,应答率和应答偏差之间不存在直接对应关系。

其次,应答率不能反映无应答样本缺失的状态。数据缺失机制可以归纳为三类:完全随机缺失(Missing-Completely-at-Random,MCAR)、随机缺失(Missing-at-Random,MAR)和非随机缺失(Not-Missing-at-Random,NMAR)。在无应答样本处于完全随机缺失的状态时,样本缺失不会影响估计值的偏差。应答率高,应答样本的规模就大,如此可以降低估计值的方差,追求高应答率就会有很好的回报。对于随机缺失和非随机缺失两种情形,如果样本的缺失会直接影响估计值的偏差,应答率基本上就与应答样本的代表性无关。

第三,研究证明,应答率与无应答偏差(nonresponse bias)没有必然关系(Groves,2006;Groves and Peytcheva,2008;Heerwegh, *et al.*,2007)。斯克顿等(Schouten, Cobben and Bethlehem,2009)设计了一个简单的例子说明这个问题。他们选取了 1998 年荷兰的一个调查项目(POLS),这项调查历时两个月,采访在进行了一个月和两个月时,分别有不同的应答率。研究者从注册数据中选取了两个变量:荷兰人口中

接受社会保障津贴的人数比例和父母至少有一方在荷兰境外出生的人数比例，并分别将两个变量的数据与调查进行一个月和两个月后的应答样本匹配。最后估算出在不同应答率情况下两个变量的估计值（见表1）。数据显示，POLS在进行一个月后的应答率为47.2%，两个月后的应答率为59.7%，增长了12.5%。而在这两个变量的应答均值上，高应答率并没有带来更接近真值的结果，估计误差反而增大了。

表1：POLS进行一个月和两个月后的变量估计值比较（%）

变量	一个月后	两个月后	样本
社会保障津贴	10.5	10.4	12.1
非国内出生	12.9	12.5	15.0
应答率	47.2	59.7	100

数据来源：Schouten，Cobben and Bethlehem（2009：102）。

此外，应答率的计数方法是“一人一票”，适用于等概率入选的抽样设计。对于不等概率的复杂抽样设计，则采用简单的应答率计算方式，其结果将只能了解采访执行过程的进展情况，而无法了解应答样本的代表性。

因此，把应答率作为应答代表性指标存在多个误区。简单的应答率计算方法对复杂的抽样设计不适用，应答率也不能反映无应答样本的缺失机制。最关键的是，应答率与无应答的偏差没有必然联系，因此，在问卷调查中不能把应答率作为评价调查质量好坏的指标。

二、应答代表性指标的研究进展

问卷调查研究一直致力于寻求替代应答率的应答代表性指标。格鲁夫斯等（Groves，Kirgis，*et al.*，2008）提出，可以在两个维度上建构应答代表性指标：一个是在调查维度上构建一个单一指标，另一个是在估计维度上可以包含多个个体指标。调查维度上的单一指标相对较直观和便于使用，也是研究者需要攻克的重点。瓦格纳（Wagner，2012）根据建构指标的数据类型，将在权威学术期刊上发表的文献里的调查维度指标分为三大类：仅根据是否应答的样本信息来建构的指标；除了是否应答的样本信息，还需要基于样本框数据或并行数据（paradata）建构的指标；除了以上三种数据信息外，还需要用到调查数据建构的指标。

在第一类指标中，最有代表性的就是应答率。前面提到，应答率作

为代表性指标隐含了一个很强的假定,即应答总体的均值与无应答总体的均值有恒定的差异。所以,应答率与应答偏差成反比,也即,应答率越高,应答偏差越小。如果无应答是完全随机的,上述命题成立,应答率就可以作为有效的代表性指标使用。

应答率在应用上有很多便利,既可以在不同调查之间进行比较,也可以在同一调查的不同时间点进行比较。追求高应答率常常会影响数据采集的策略及人力物力的投入。

第二类指标主要包括分组应答率(Groves,Brick,*et al.*,2008)和R指标(Schouten,Cobben and Bethlehem,2009)。这一类指标不仅要用到样本是否应答的信息,同时还要借助样本框数据或并行数据进行分组或建构模型。样本框数据是关于样本总体的数据,如人口特征、人口密度、地区经济等统计结果,并行数据则来自数据采集过程,如联系记录、访员观察和访问痕迹等。无论是对应答者还是无应答者,这些数据都可以采集到,因而,可以用来分析应答者和无应答者在这些变量上的区别。

分组应答率就是根据这些变量类别将样本分为若干子样本,计算每个子样本的应答率,然后得出这些子样本应答率的变异系数(coefficient of variation)作为评估应答代表性的指标。系数越低,说明应答样本越趋向于总样本在这些变量上的一个无偏的子样本。

R指标则要基于样本框数据和并行数据建构应答倾向模型,并根据这个模型估计出每个样本的应答概率。应答概率的方差越小,R指标的数值就越大,也意味着应答样本的代表性越强。

第二类指标有三个优势。第一,其指标建构包含了更多的信息,而且这些信息对于所有样本都是完整和无缺失的;第二,这类指标可以在不同调查之间进行比较,其条件是这些调查具有相同的样本框,在指标计算上要选取相同的样本框数据和并行数据;第三,这类指标也可以用于同一调查不同时间点上的比较,条件也是计算时要选用相同的样本框数据和并行数据。在数据采集过程中,组织者可以根据这些指标提供的信息来调整工作策略,以求在各个子样本上取得平衡的应答效果。

这类指标的弱点也很明显。第一,分组或模型的建构依赖自变量的选取,不同研究者采用不同的分组标准和模型建构会得出不同的结果;第二,指标的质量还取决于数据质量,特别是在不同调查之间进行

比较时，很难保证并行数据具有相同的质量；第三，这类指标只基于样本框数据和并行数据，也隐含了一个很强的假定，即选用的样本框数据和并行数据与调查主题的所有估计值密切相关。实际上，大多数调查都不只包含一个主题，因而这个假定很难得到证实。

第三类指标在一定程度上针对的是第二类指标的第三个弱点。这些指标从处理调查数据的缺失值入手，通过统计手段，利用样本框数据和并行数据，或对辅助数据与调查数据的关系进行分析，或对事后权重进行调整，或对缺失值进行插补，然后在这些分析结果的基础上评价应答样本的代表性。

这类指标的例子很多，如辅助数据与调查变量的相关性(Kreuter，*et al.*，2010)、事后权重与调查变量的相关性(Olson，2006)、缺失信息分值(Fraction of Missing Information，FMI)(Wagner，2010)等。这一类指标在计算上要相对复杂，而且会遇到第二类指标同样的问题，即模型建构中变量的选取和数据质量问题。这些评价方法更多是建立在估计维度基础上，会受到调查变量的影响，在不同调查间的比较和调查过程的指导上具有局限性。

通过对上述应答代表性指标或评估方法的了解和比较，本文认为R指标具有明显的优势。首先，相对于第一类指标中的应答率，R指标不仅有更强的理论依据，也借助样本框数据和并行数据，提升了指标的信息量；其次，与同一类指标中的分组应答率相比，R指标可以避免变量在多类别多情况下的分组复杂性；第三，在计算难度和应用上，R指标也明显优于第三类指标，并能不受调查内容的影响，可以实现不同调查项目的横向比较。同时，R指标也不受变量值变动的影响，可以实现同一项目在不同执行阶段的纵向比较。因此，在问卷调查中，R指标有很好的应用前景。

接下来将对R指标的建构和应用进行详细介绍。

三、R指标的建构

为进一步讨论R指标对应答样本的代表性，首先要对概念进行界定。

(一)“应答代表性”的定义

“代表性”有“强”和“弱”两个定义(Schouten，Cobben and Bethlehem，2009)：如果目标总体中所有单元的应答概率完全相同并相

互独立,那么样本的应答就具有代表性(强);如果对于分类变量 X,样本在不同类别上的平均应答倾向(response propensity)是相同的常数,那么样本的应答相对于 X 来说就具有代表性(弱)。

实际上,这两个定义都认为,当无应答样本的缺失机制属于完全随机缺失(MCAR)时,其应答样本是有代表性的。"强"定义是理论上的代表性,因为每个采访对象的应答概率无从得知;"弱"定义则是可操作的定义,可以选择某几个分类变量,然后利用统计方法估计样本在这些类别上的应答倾向,比较不同类型的平均应答倾向值是否相同。因此,要满足"弱"定义的代表性,需考虑以下三方面的问题:

第一,分类变量的选择。用作建构代表性指标的分类变量首先要满足一个重要条件,即该变量必须在所有样本上有值,无论是应答样本还是无应答样本,该变量值都不能缺失;其次,该分类变量要与抽样设计密切相关,这样在比较应答样本和目标样本的相似程度时才会更有参考价值;第三,这些分类变量最好是研究变量估计值的有效预测因子,这样就可以更好地捕捉到无应答对于估计值偏差的影响。问卷调查中能够满足这些条件的数据有两种,一种是辅助性变量(auxiliary variable),如目标总体的普查数据、样本框的各类数据等;另一种是并行数据,也就是关于调查执行过程的数据,如联系记录、访员观察、访员调配记录等。

第二,样本应答倾向的计算。在这方面,最常用的模型是逻辑斯蒂回归,模型的因变量是代表目标样本是否应答的二分变量。也可以根据研究的需要选用其他模型,如研究不同联系尝试次数情况下的应答代表性问题,可以选用离散时间的风险模型(discrete-time hazard model)。[2]

第三,不同类别平均应答倾向的比较。当根据应答倾向模型计算出不同类别上应答倾向的估计值,用来比较这些估计值与平均应答倾向值差距最简单的方法就是计算方差。方差小,证明应答倾向估计值接近相同,应答样本就有很好的代表性;方差大,则应答倾向的差异就大,应答代表性就弱。

(二) R 指标的测量

经过以上理论界定和操作上的考虑,研究者将代表性的"强"定义

2. Geert Loosveldt and Koen Beullens. RISQ-Fieldwork Monitoring. Work Package 6, Deliverable 5, version 2, The RISQ Project, 7th Framework Programme (FP7) of the European Union, 2009: www.r-indicator.eu。

中的 R 指标定义为：

$$R(\rho) = 1 - 2S(\rho)$$

其中，ρ 代表目标样本的应答概率，$S(\rho)$是应答概率的标准差，即：

$$S(\rho) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\rho_i - \bar{\rho})^2}$$

公式中 N 是目标总体的个数，ρ_i 是“强”定义中目标总体的一个单元基于其被选择为目标样本做出应答的概率，也就是：

$$\rho_i = P[r_i = 1 \mid s_i = 1]$$

$S(\rho)$的取值范围在 0 和 0.5 之间不难证明。我们通过一个简单的数学转换，将代表性指标 $R(\rho)$的取值范围限定在 0 和 1 之间，这样更方便解释和使用。当 R 指标越接近 1，应答概率的方差越小时，应答的代表性就越强；当 R 指标越接近 0，应答的代表性就越弱。

在实际操作时，目标样本的应答概率往往无从得知，这就不得不借助可操作的“弱”定义，将 R 指标的计算公式加以调整。先把应答概率 ρ 用基于模型预测的应答倾向 $\hat{\rho}$ 来代替，然后对应答倾向的平均值采用加权算法，即设定 s_i 代表单元 i 是否为目标样本，π_i 代表单元 i 的入选概率，那么，

$$\hat{\bar{\rho}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i \frac{s_i}{\pi_i}$$

这样，应答代表性指标的估计值计算公式就转换为：

$$\hat{R}(\rho) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\frac{s_i}{\pi_i}(\hat{\rho}_i - \hat{\bar{\rho}})^2}$$

简而言之，R 指标计算过程中包括两个步骤。第一步是选择合适的分类变量，建构应答倾向模型，第二步是在应答倾向模型估计值的基础上计算应答代表性 R 指标。

（三）其他辅助性指标

1. 最大绝对偏差(maximal absolute bias)

如果 R 指标测量的是无应答样本的缺失机制是否为完全随机缺失，那么能否根据 R 指标的大小得知无应答带来的偏差大小？答案是可以的，但会有一定局限。研究者在 R 指标的基础上建构了相对于变量 X 的最大绝对偏差的估算方法：

$$B_m(\rho_x) = \frac{1 - R(\rho_x)}{\bar{\rho}_x}$$

在实践中,以上公式中的参数用估计值来代替,于是公式转换为:

$$\hat{B}_m(\hat{\rho}_x) = \frac{1 - \hat{R}(\hat{\rho}_x)}{\hat{\bar{\rho}}_x}$$

需要注意的是,最大绝对偏差相对应的变量 X 应是估计 R 指标时采用的相同的分类变量,所以,该指标反映的偏差不是针对具体调查项目中的某个具体研究变量的无应答偏差,而是对同一或不同调查项目所共有的辅助性变量或并行数据在应答样本中的估计偏差,不能用这个指标来评估某个具体研究变量的估计偏差。

2. 偏 R 指标(partial R-indicator)

R 指标可以用来评估应答样本的代表性程度,却不能确定哪个组别的人群被过低或过高代表。研究者于是在 R 指标的基础上又发展出偏 R 指标的计算方法,[3] 用来评估某一个具体的分类辅助变量对应答代表性的影响。

偏 R 指标有两种:一种是无条件的偏 R 指标(unconditional partial R-indicator),意义上类似关联分析中的相关系数,测量的是某一个变量对于应答代表性的直接影响;另一种是有条件的偏 R 指标(unconditional partial R-indicator),测量的是某一变量基于其他变量对于应答代表性的影响,意义上类似偏相关系数。这两个偏 R 指标可以帮助我们进一步观测到应答样本在代表性问题上的细节,并具有实践上的指导作用。

(四)使用 R 指标的注意事项

R 指标可以用来比较共同目标总体下不同问卷调查的应答样本代表性,也可以比较追踪调查中不同次调查的应答样本代表性,还可以比较同一调查中不同数据采集时间点上应答样本代表性。但在应用时还有一些注意事项,斯克顿等(Schouten,Cobben and Bethlehem,2009:27

3. 因计算方法相对复杂,本文不再赘述。感兴趣者可参看:Natalie Shlomo,Chris Skinner,Barry Schouten, Thaya Carolina, and Mattijn Morren. RISQ-Partial Indicators for Representative Response. Work Package 5,Deliverable 4,version 2, The RISQ Project,7th Framework Programme (FP7) of the European Union,2009:www. r-indicator. eu。

—28)对此作了总结,[4]概述如下:

1. 报告 R 指标和最大绝对偏差时,必须同时报告用来预测应答倾向的辅助变量;

2. 在比较不同的问卷调查项目时,要求使用具有相同类别的完全相同的辅助变量,以及相同的辅助变量的交互项;

3. 应答倾向模型中变量的多少会影响到 R 指标的估计误,因此建议只选择理论上或文献中已经证明的和应答行为相关的辅助性变量作为模型的预测变量;

4. R 指标和最大绝对偏差的估计会受到样本规模的影响,因而更适用于大规模的社会调查;

5. R 指标测量的是应答样本的总体代表性,不反映无应答对于调查变量估计偏差的影响。

四、R 指标的应用:以 CFPS 为例

"中国家庭动态跟踪调查"(CFPS)是一项全国性大规模、多学科的社会跟踪调查项目,目标样本覆盖全国 25 个省、市、自治区(不含西藏、青海、新疆、宁夏、内蒙古、海南、香港、澳门、台湾),样本类型分别包括村居、村居样本下的家户,以及家户内的所有家庭成员。调查内容涉及社会、经济、教育、人口和健康等诸多主题,并自 2010 年初访调查后对这些主题进行跟踪调查。

在调查技术上,CFPS 结合了调查问卷的计算机辅助面访(Computer-Assisted Personal Interview, CAPI)和访问活动的网络管理,实现了对大型综合性调查的高效管理,采集到了丰富的调查数据和并行数据,在很大程度上也提高了数据质量。

(一) R 指标计算变量的选取

分类变量的选取是计算 R 指标的重要步骤。根据 CFPS 的特点,我们主要从抽样设计、调查经验和数据质量三个方面筛选进入模型的分类变量。

---

4. Barry Schouten, Mattijn Morren, and Jelke Bethlehem. RISQ-How to use R-indicator? Work Package 4, Deliverable 3, The RISQ Project, 7th Framework Programme (FP7) of the European Union, 2009: www.r-indicator.eu。

首先,我们希望选取的变量能够和调查主题的重要估计值密切相关。CFPS是一个综合的社会科学数据采集平台,主题多样,涵盖面广。一般这样的调查在抽样设计上都会尽可能获取样本在地理分布上的代表性,所以CFPS采用了内隐分层(implicit stratification)的多阶段概率抽样方法,主要排序变量为区县的人均GDP、非农人口比例和人口密度。[5]也就是说,从理论上看,设计者认为样本在这些变量上的代表性会影响调查主题变量的估计值,因此,在选取计算R指标的分类变量时,要首先考虑使用这些辅助性变量,这些数据可以从国家发布的统计资料直接获取。

其次,根据以往的调查实践和研究经验,样本是否能够访问成功与村/居所处的地域类型(如城市、城镇、农村或郊区)、村/居委会对调查的配合程度,以及访员的性别密切相关。一般来说,农村居民比较容易接受采访,城市居民则相对较难;村/居委会对于调查的支持程度会直接影响居民对访问的态度;女性访员相对男性访员的入户难度要低一些。在CFPS 2010年的调查中,这些信息通过访员的观察记录、对访员的调查问卷,以及调查支持系统存储的访员调配记录被有效采集。

第三,选取变量时,我们尽量避免使用理论上认为有作用,但数据质量没有保证的变量。如访员的努力程度(表现为访员联系受访者的次数、时间或方式等)会直接影响访问是否能够顺利进行。CFPS通过访员的联系记录采集了这些数据,但访员反映,由于开关电脑比较麻烦,如果住户无人响应,一般不会马上在电脑里插入联系记录,而是或者当天工作完成后几个家庭一起插入记录,或者索性就漏掉不记了,这样会导致联系次数记录不足,电脑自动记录的联系时间(即插入联系记录的时间)也不准确。对于这些数据有明显问题的变量,我们没有用来计算R指标。

因此,CFPS建构R指标的分类变量主要有两大类:来自抽样设计的辅助性变量数据和来自调查实施过程的并行数据,具体如表2所示:

---

5. 谢宇、邱泽奇、吕萍. 2012. CFPS－1:中国家庭动态跟踪调查抽样设计. 参见网址:http://www.isss.edu.cn/index.php?catid=201&action=index。

表 2:建构 R 指标的分类变量

变量	类别	变量	类别
辅助性变量数据		并行数据	
区县人均 GDP(元)	＜5 000	村居类型	城市
	5 000－9 999		城镇
	10 000－19 999		农村
	20 000－39 999		郊区
	＞＝40 000		
区县非农人口比例	＜0.1	村/居委会配合程度	很少
	0.1－0.2		较少
	0.2－0.4		较多
	0.4－0.8		很多
	0.8－1		
区县人口密度(人/km^2)	＜200	访员性别	女
	200－399		男
	400－599		
	600－1 599		
	＞＝1 600		

（二）不同调查阶段 R 指标的比较

前面提到，R 指标的一个重要应用就是比较同一调查在不同数据采集时间点上应答样本的代表性。CFPS 2010 年的初访调查大致分三个阶段：第一阶段从 4 月实地入户开始至暑假前（以 7 月 31 日为截点），这一阶段根据抽样进度，对不同地区的访员进行多批次培训，一批访员培训结束，便开始一些地区的入户访问；第二阶段（8 月 1－31 日）是访问全面展开阶段，这一时期抽样工作已结束，部分样本区县的大学生访员暑假回到家乡入户调查，极大地推动了调查进展；第三阶段是调查收尾阶段，包括对因世博会或自然灾害等未能访问成功的部分村居的补访，以及对春节期间对外出回乡家庭的补访等，这一阶段一直延续到初访调查结束。

表 3 列出了这三个阶段的主要测量结果。首先，随着调查的推进，访到率明显上升，从第一阶段的 55.9％升高到第二阶段的 79.7％。第三阶段是攻坚阶段，访到率上升减缓，只增加了大约 3％。然而，与访到率相比，R 指标的变化不大，只是从第一阶段的 0.731 升高到第三阶段的 0.791，但这并不意味着在第一阶段就已访问到有代表性的样本，以后的访问徒劳无功。从样本代表性的另一个指标最大绝对偏差看，

表 3:CFPS2010 不同调查阶段访到的家户样本代表性测量

	第一阶段	第二阶段	第三阶段
访到率(%)	55.9	79.7	82.8
R 指标	0.731	0.783	0.791
最大绝对偏差	0.240	0.136	0.126
无条件偏 R 指标			
区县人均 GDP	0.064	0.071	0.069
区县非农人口比例	0.096	0.069	0.066
区县人口密度	0.068	0.073	0.068
村/居类型	0.105	0.088	0.084
村/居委会配合程度	0.067	0.050	0.051
访员性别	0.013	0.004	0.004
条件偏 R 指标			
区县人均 GDP	0.017	0.016	0.016
区县非农人口比例	0.033	0.010	0.008
区县人口密度	0.021	0.020	0.018
村/居类型	0.041	0.039	0.037
村/居委会配合程度	0.037	0.021	0.023
访员性别	0.029	0.020	0.020

这一数值在第一阶段高达 0.240,而在第三阶段降低到0.126,意味着这些辅助性变量或并行数据在应答样本中的估计偏差明显降低。

我们可以从偏 R 指标观察影响样本代表性的变量情况,主要有三个发现:第一,这些变量的条件偏 R 指标和无条件偏 R 指标差异明显,证明这些变量之间并不是彼此独立的,而是彼此关联的,这与常识相符;第二,村/居类型在这些变量中是最有影响的,在所有三个阶段的影响力中都是最强的,说明访问到的样本在不同类型的村/居内分布不均衡;第三,随着调查进入后两个阶段,所有变量数据都有不同程度的降低,其中区县非农人口比例表现最为显著,其条件偏 R 指标从第一阶段的0.033降至第三阶段的 0.008,这表明后期的访问明显提高了应答样本在这些变量上的代表性,尤其是有效地弥补了前期访到样本在不同非农人口比例地区的失衡。

我们还可以通过变量类别的偏 R 指标考察应答样本在不同类别上的代表性情况,其结果按照第三阶段的无条件偏 R 指标从小到大排序可参见表 4。无条件偏 R 指标如果呈现负值,表示此类别上的应答样本代表性不足,需要增加该类别上应答样本比例。负值越小,表明需

要增加的应答样本比例越高。如果呈现正值，则意味着该类别被过度代表，正值越大，过度代表的程度越大。因此，那些拥有较低的负向无条件偏 R 指标和较高的条件偏 R 指标最应引起调查执行人员和研究人员的注意。

在变量的偏 R 指标上，我们已经发现，村/居类型对应答样本代表性的影响较大，表 4 更是清楚地显示，这一变量中处于城市的村/居的应答样本呈现代表性不足的状况，其影响在所有调查阶段和所有变量类别中居于首位。与其对应的是处于农村的村/居样本被过度代表，其影响也比其他变量类别大。

除了城市的村/居，区县 GDP 大于或等于 40 000 元，区县人口密度大于等于 1 600 人/km^2，区县非农人口比例在 0.8 和 1 之间，村/居委会配合程度很少的子样本的应答代表性也相对不足，这些都反映了目前中国抽样入户调查的困难主要集中在具有这些特征的地区。

此外，变量类别对于应答代表性的影响力在不同的调查阶段也不相同，这表现在表 4 中三个阶段变量类别的不同排序上。如区县 GDP 大于或等于 40 000 元的变量类别在第一阶段的影响力位居第 5，而到访问全部结束时，其影响力上升至第 2。这表明，后期调查访问的样本较多集中在区县 GDP 小于 40 000 元的地区，加重了前者的代表性不足问题。同时我们也发现，区县非农人口比例在 0.8 和 1 之间的变量类别的影响力从第一阶段的第 2 降到采访结束时的第 4，说明后期的调查在这类地区访到了较多的样本，在一定程度上增加了应答样本的代表性。

（三）R 指标的作用

在 CFPS 2010 年的初访调查中，我们是在调查结束后开始计算 R 指标的。事后计算的好处在于，变量选取时更为谨慎，一方面可以从所有获取数据中有效筛选相关分类变量，另一方面对所选变量的数据质量也有全面认识。通过对 R 指标及其辅助指标结果的分析，我们可以对调查数据的代表性有更深入的了解，并为数据使用者提供重要的参考信息。同时，因为 CFPS 是跟踪调查，我们也可以从初访调查中吸取经验和教训，有针对性地制定管理方案，提高一些子样本的代表性，应用到下一轮的跟踪调查中。

表 4:变量类别对应答样本代表性的影响

	第一阶段			第二阶段			第三阶段		
	Pu	Pc	排序	Pu	Pc	排序	Pu	Pc	排序
村/居类型:城市	-0.079	0.024	1	-0.070	0.024	1	-0.068	0.023	1
区县人均 GDP(元):>=40 000	-0.037	0.005	5	-0.053	0.011	4	-0.051	0.010	2
区县人口密度(人/km²):>=1 600	-0.050	0.008	4	-0.053	0.007	3	-0.051	0.007	3
区县非农人口比例:0.8-1	-0.069	0.010	2	-0.054	0.002	2	-0.050	0.004	4
村/居委会配合程度:很少	-0.055	0.027	3	-0.041	0.013	5	-0.040	0.014	5
区县人口密度(人/km²):600-1 599	-0.017	0.014	8	-0.019	0.011	6	-0.017	0.008	6
区县人均 GDP(元):20 000-39 999	-0.028	0.004	6	-0.013	0.007	7	-0.015	0.007	7
区县非农人口比例:0.4-0.8	-0.016	0.009	9	-0.007	0.002	9	-0.012	0.004	8
村/居委会配合程度:较少	-0.012	0.011	10	-0.006	0.006	10	-0.011	0.008	9
村/居类型:城镇	-0.017	0.019	7	-0.009	0.018	8	-0.006	0.016	10
村/居类型:郊区	-0.010	0.011	11	-0.001	0.014	12	-0.004	0.013	11
访员性别:男	-0.008	0.019	12	-0.002	0.014	11	-0.003	0.014	12
访员性别:女	0.010	0.021	15	0.003	0.014	13	0.003	0.015	13
区县非农人口比例:0.2-0.4	0.010	0.013	14	0.013	0.004	17	0.010	0.004	14
区县人口密度(人/km²):400-599	0.029	0.013	20	0.010	0.011	15	0.012	0.011	15
区县人均 GDP(元):10 000-19 999	0.015	0.008	16	0.012	0.009	16	0.014	0.008	16
村/居委会配合程度:较多	0.027	0.015	19	0.010	0.006	14	0.015	0.009	17
区县人口密度(人/km²):200-399	0.003	0.004	13	0.023	0.008	18	0.021	0.007	18
区县人均 GDP(元):<5 000	0.017	0.012	17	0.030	0.004	21	0.024	0.002	19
村/居委会配合程度:很多	0.025	0.017	18	0.027	0.014	20	0.025	0.014	20
区县非农人口比例:<0.1	0.054	0.024	24	0.033	0.007	22	0.027	0.004	21
区县非农人口比例:0.1-0.2	0.034	0.014	22	0.023	0.005	19	0.030	0.002	22
区县人口密度(人/km²):<200	0.034	0.004	21	0.039	0.006	24	0.034	0.004	23
区县人均 GDP(元):5 000-9 999	0.038	0.008	23	0.033	0.004	23	0.035	0.004	24
村/居类型:农村	0.066	0.024	25	0.052	0.021	25	0.050	0.020	25

注:Pu 代表无条件偏 R 指标,Pc 代表条件偏 R 指标。

R 指标除了用于事后评估和经验教训的总结外，也可以用于调查过程中的实时监测。如格鲁夫斯和海瑞纳(Groves and Heeringa，2006)提出的回应式社会调查设计(Responsive Survey Design)就很需要在调查过程中及时计算各种指标，并以此为指导改变调查设计，达到有效提高成本效益和测量精度的目的。对于 CFPS 和大多数问卷调查来说，抽样框数据在调查前就已获取，在调查过程中可以通过 R 指标的计算实时掌控样本在这些变量上的分布，及时采取应对措施来平衡应答样本的代表性。

五、结语

在科学抽样调查兴起的 20 世纪中期，问卷调查的应答率一般高于 90%。而最近几年，应答率不断下滑，研究者开始担心科学概率抽样获取的样本代表性已经被低比例的应答破坏。但应答率自身不能回答这样的问题，研究者开始致力于开发新的综合指标来揭示应答样本的代表性，R 指标就是这种趋势下的产物。

本文讨论了应答样本代表性的含义，指出应答率作为代表性指标的局限，并比较了目前在应答代表性评估方面的研究成果，然后介绍 R 指标的界定及计算方法，并借助 CFPS 数据对 R 指标的应用做了尝试性分析。R 指标的优势在于，可以在变量甚至变量类别层面上观察应答样本的代表性情况，并在同一个调查的不同时间点追踪调查中不同次的访问，以及在共同目标总体下对不同社会调查的应答样本代表性进行比较。

R 指标在国内有较好的应用前景。经历了从国外引进到本土发展，国内问卷调查在方法上越来越成熟，应用也越来越广泛。近年来，各种社会科学调查项目收集了大量数据，然而，大多数调查并非由专业的调查机构来执行，调查过程不透明，调查数据也不共享，使研究者对数据质量虽多有担忧，却无从评估。

R 指标通过报告应答样本的总体代表性和在一些重要变量上的具体的代表性偏差，能帮助研究者对数据质量有更深刻的认识，并在使用数据做研究时能更好地阐释分析结果，使结论更为合理可信。对从事问卷调查的专业人员来说，他们可以通过 R 指标了解调查进程中存在的代表性问题，有针对性地调整管理策略，提高调查数据质量。

同时,R 指标的应用对于国内问卷调查的发展具有重要的促进作用。首先,R 指标主要依靠的数据来源是样本框数据和并行数据。样本框数据可以通过权威的统计资料获得,并行数据则需要主动采集。因此,R 指标的应用会促使问卷调查从业者注重调查过程及并行数据的采集,从而使调查过程可以通过数据分析得以观测,这些过程信息也有助于对问卷缺失数据的科学补值。

其次,它还可以促使国内问卷调查技术由传统向现代转变。传统的纸笔调查模式下,并行数据的采集会有一定的难度,数据质量也很难保证,这都会影响 R 指标的计算。现代信息技术支持下的计算机辅助访问可以有效采集到大量高质量的并行数据,因此,国内调查界应尽快采用先进的调查技术,提高问卷调查的科学性。

第三,R 指标的应用还会促进问卷调查机构的专业化发展。R 指标计算的一个重要环节就是建构应答倾向模型,选取的变量是否可以有力预测应答倾向将直接影响到该模型的效能。这就要求问卷调查机构的研究人员要有较高的研究素质和实践经验,选取合理的分类变量和分析模型计算 R 指标。此外,如何根据 R 指标提供的信息提高调查质量是一个很现实的问题。调查策略能否适时调整,调整的效果能否得到及时评估,都要依靠一个团队的紧密合作和信息流转的通畅,这也将促使调查机构在管理模式和技术手段上实现新的突破。

总之,在未来的问卷调查过程中,管理者和研究者在追求高应答率的同时,也应该关注应答样本在哪些子群体中出现代表性不足或过度代表的情形,并根据这些信息在调查过程中适时调整访问策略,或制定新策略以应用到下一次调查中,提高问卷调查的质量。R 指标在这方面是一个重要的工具,不仅在国内有较好的应用前景,还会促进国内问卷调查进一步发展。

参考文献(References)

American Association for Public Opinion Research. 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*(7th Edtion), http://www.aapor.org/Standard_Definitions2.htm#. UsojAmzxvIV.

Babbie, Earl. 2007. *The Practice of Social Research*(11th Edtion). Belmont, CA: Wadsworth.

de Leeuw, Edith and Wim de Heer. 2002. "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison." In *Survey Nonresponse*, edtied by

Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little. New York: Wiley: 41－54.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70(5): 646－675.

Groves, Robert M. and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72(2): 167－189.

Groves, Robert M., Michael Brick, Mick Couper, William D. Kalsbeek, Brian Harris-Kojetin, Frauke Kreuter, Beth-Ellen Pennell, Trivellore E. Raghunathan, Barry Schouten, Tom W. Smith, Roger Tourangeau, Ashley Bowers, Matthew Jans, Courtney Kennedy, Rachel Levenstein, Kristen Olson, Emilia Peytcheva, Sonja Ziniel, and James Wagner. 2008. "Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools." *Survey Practice* 1(3): 14－22.

Groves, Robert M., Nicole Kirgis, Emilia Peytcheva, James Wagner, William G. Axinn, and William D. Mosher. 2008. "Responsive Design for Household Surveys: Illustration of Management Interventions Based on Survey Paradata." NCRM Research Methods Festival, St Catherine's College, Oxford, UK.

Groves, Robert M. and Steve Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169 (3): 439－457.

Heerwegh, Dirk, Koen Abts, and Geert Loosveldt. 2007. "Minimizing Survey Refusal and Noncontact Rates: Do Our Efforts Pay Off?" *Survey Research Methods* 1(1): 3－10.

Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64(2): 125－148.

Kreuter, Frauke, Kristen Olson, James Wagner, Ting Yan, Trena M. Ezzati-Rice, Carolina Casas-Cordero, Michael Lemay, Andy Peytchev, Robert M. Groves, and Trivellore E. Raghunathan. 2010. "Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples from Multiple Surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173(2): 389－407.

Kruskal, William and Frederick Mosteller. 1979a. "Representative Sampling I: Non-Scientific Literature." *International Statistical Review* 47(1): 13－24.

Kruskal, William and Frederick Mosteller. 1979b. "Representative Sampling II: Scientific Literature Excluding Statistics." *International Statistical Review* 47(2): 111－123.

Kruskal, William and Frederick Mosteller. 1979c. "Representative Sampling III: Current Statistical Literature." *International Statistical Review* 47(3): 245－265.

Olson, Kristen. 2006. "Survey Participation, Nonresponse Bias, Measurement Error Bias and Total Bias." *Public Opinion Quarterly* 70(5): 737－58.

Schouten, Barry, Fannie Cobben, and Jelke G. Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35(1): 101－113.

Singleton, Royce and Bruce Straits. 2005. *Approaches to Social Research* (4th Edtion). New York: Oxford University Press.

Wagner, James. 2010. "The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data." *Public Opinion Quarterly* 74(2): 223－243.

Wagner, James. 2012. "A Comparison of Alternative Indicators for the Risk of Nonresponse Bias." *Public Opinion Quarterly* 76(3): 555－575.

责任编辑：张　军