

再谈样本规模和调查回收率

——对《应答率的意义及其他》一文的回应

风笑天

提要: 本文以事实为依据,对《应答率的意义及其他》一文的质疑进行了回应。作者指出,首先,“应文”质疑的出发点是不恰当的;其次,笔者关于调查回收率的定义与福勒的定义完全一致;第三,笔者对样本规模的定义、以及对 CGSS2003 调查项目中的样本规模——5900 的定义和理解也是清楚的和正确的,笔者并没有“误读”这一样本规模的含义。第四,真正错误理解和定义“样本规模”含义的恰恰是“应文”的作者。由于“应文”作者不恰当地将“样本规模”的概念变换成“实际发生的访谈数量”,不恰当地在有关“样本规模”的讨论中引入“实地调查过程”的内容,因而形成了对样本规模含义的不正确理解,做出了对样本规模含义和计算的不正确划分,同时也导致了 CGSS2003 的样本规模计算以及其他调查项目回收率统计中的混乱现象。

关键词: 调查研究 样本规模 回收率

笔者的一点学习心得,在很短的时间内就得到了同行的关注,深感欣慰。编辑部转来的《应答率的意义及其他》一文(以下简称“应文”)对拙文提出了一些不同意见。本着学术探讨的目的,笔者给出以下回应。

一、笔者前文质疑的究竟是什么?

“应文”一开始就写道,笔者的前文“对中国目前社会调查中的高回收率提出了质疑,指出由于严格执行随机抽样和实地调查所必然面临的许多客观障碍和所必然造成的回收率下降,所以,过高的回收率实际上意味着调查过程的不严格性,其结果则是调查资料的低质量”。在提要中,“应文”作者也写道,笔者的前文“对中国社会调查质量的质疑也缺乏足够的事实依据”(郝大海,2007)。

笔者前文的内容真的如上所述吗?这不符合事实。事实是,笔者前文的内容既没有对“中国目前社会调查中的高回收率(现状)”提出质疑,也没有“对中国社会调查质量”提出质疑,更没有“指出”上文中所指

出的那一段结论。

首先,笔者将前文的提要完整照录,读者可看看笔者前文论述的究竟是什么:

论文由美国学者的质疑引入对调查回收率所具有的方法论意义的探讨,并结合国内外的研究实例,对调查回收率的衡量标准及其依据提出了自己的看法。作者指出,一方面,调查回收率所代表的实际样本的大小,是衡量一项调查结果所具有的代表性的最终指标,因而回收率越高越好。但另一方面,由于实际调查中存在着许多制约和影响回收率的因素,如果不切实际地单纯追求高回收率可能会影响到调查资料的质量,因此回收率并非越高越好。研究认为,样本的完整与资料的质量是我们在调查中应该同等关注的两个方面。既要尽可能地提高回收率,也要尽可能地防止和避免低质量的高回收率。(风笑天,2007)

实际上,笔者前文主要是针对笔者以往认识中(当然也可能包括其他研究者认识中)存在的“回收率越高越好”的片面看法提出的质疑。论文集中探讨的是“美国学者为什么会对高回收率提问”、“高回收率是否比低回收率更好”、“有哪些因素影响调查回收率的高低”、“应该如何正确看待调查回收率”等问题,研究的结论所回答的也是“回答率是否越高越好”的问题。关于这一点,读者无论是从笔者前文的大小标题、上述的论文提要、还是从整篇论文的立论和论述中,都不难看清。笔者对“回收率越高越好”的看法进行质疑,与“应文”所说的对“中国目前社会调查中的高回收率(现状)”的质疑、或者对“中国社会调查质量”的质疑显然并不是一回事。

其次,在笔者前文第六小节的最后一段中,明明白白地写着这样一段文字:

正是由于十分了解严格执行随机抽样和实地调查所必然面临的许多客观障碍和所必然造成的回收率下降的现象,所以,当听到高回收率时,有经验的研究者就会很自然地调查实际实施的严格性提出疑问。这就是那位美国教授为什么会以一种对高回收率表示怀疑和将高回收率看作缺点而提问的原因。在他眼里,过高

的回收率实际上意味着调查过程的不严格性,其结果则是调查资料的低质量。(风笑天,2007)

我们如果加上这一段叙述中的主语(括号中部分),读者对这段话的意思可能会看得更加清楚:

正是由于(那位美国教授)十分了解严格执行随机抽样和实地调查所必然面临的许多客观障碍和所必然造成的回收率下降的现象,所以,当(他)听到高回收率时,(他)就会很自然地_{对调查实际实施的严格性}提出疑问。这就是(他)为什么会以_{一种对(笔者的)高回收率表示怀疑和将(笔者的)高回收率看作缺点而提问}的原因。在他眼里,过高的回收率实际上意味着调查过程的不严格性,其结果则是调查资料的低质量。

其实,即使不像这样增加叙述中的主语,读者也会很清楚地从笔者前文的上下文中看明白这段话的含义。看清“是谁”“说了什么”。但“应文”作者却很不恰当地从上述文字中抽出两句话安在笔者的头上,从而构造出一种并不存在的“笔者’指出’的看法”,并由此作为其质疑的出发点。

二、笔者对调查回收率的定义与福勒的定义是否不同?

“应文”作者在列举了笔者前文中关于调查回收率的定义以及福勒关于应答率的定义后指出:

比较以上有关调查回收率的两 个定义,发现对于调查回收率的分子,存在一些差异。前者是“计划调查的样本数”,后者是“从总体中抽取的所有人,包括那些虽被抽中但因拒绝、语言问题、疾病或者缺乏易得性而没有应答的人”。由于前 一个定义没有对“计划调查的样本数”给出进一步的说明,不知道计划中是否也包括了“没有应答的人”。(郝大海,2007)

笔者前文真的没有“给出进一步的说明”吗？“应文”作者真的“不知道计划中是否也包括了‘没有应答的人’”吗？事实是，就在笔者给出调查回收率的定义及计算方法的下面，紧接着对调查回收率的含义和计算方法进行了举例说明：

比如，一项调查从总体中抽取了400名调查对象作为样本，研究者采用自填问卷的方法收集资料，发出问卷400份，实际收回问卷380份。那么，根据上述公式计算，该项调查的回收率为：

$$(380/400) \times 100\% = 95\%$$

同样，如果研究者采用的是结构式访问的方法来收集资料，而在实际访问400名调查对象的过程中，有10位调查对象由于地址错误、外出打工、生病住院等客观原因而无法接触到，还有10位调查对象拒绝接受访问。结果，实际完成访问的对象为380人，回收率亦是95%。（风笑天，2007）

如果说，“应文”作者仅凭笔者定义中文字的表面意思还有可能产生不同的理解的话，那么，当读到上面这一段专门例举的具体详细的例子，还能说“不知道计划中是否也包括了‘没有应答的人’”吗？

笔者上面所说的“由于地址错误、外出打工、生病住院等客观原因而无法接触到”的人，以及“拒绝接受访问”的人，与福勒定义中的“那些虽被抽中但因拒绝、语言问题、疾病或者缺乏易得性而没有应答的人”不一样吗？

任何一个实事求是的读者都不难看出，笔者前文对调查回收率的理解和说明与福勒的定义完全一致。笔者定义中的“计划调查的样本数”，实际上包括了“由于地址错误、外出打工、生病住院等客观原因而无法接触到”的对象，也包括了“拒绝接受访问的对象”。“应文”作者面对笔者前文中如此清楚明白的说明，却依然说笔者没有“给出进一步的说明”。

因此，可以确认的第一点是：笔者前文中对调查回收率的定义和理解是清楚和确切的。并且，如果“应文”作者同意福勒关于应答率的定义的话，他也不应该对笔者的定义和解释产生疑义。

三、笔者是否“误读”了样本规模的含义？

在“应文”的提要、第三节开头以及结语中，其作者三度明确指出：笔者对“国内调查回收率过高的看法，只是对样本规模含义误读的结果”（郝大海，2007）。这是“应文”作者质疑笔者前文的焦点所在。问题是：笔者真的“误读”了样本规模的含义吗？

首先看看笔者对样本规模含义的理解与其他学者的理解之间有无不同。

在笔者撰写的两本方法教材中，笔者对样本规模给出的定义都是：“样本中所含个案的多少”（风笑天，2005a：78，2005b：147）。

中国大百科全书社会学卷中的定义是：“样本内所含单位的数量”（中国大百科全书总编辑委员会《社会学》编辑委员会，1991：20）。

北京大学社会学系袁方教授主编的《社会研究方法教程》中的定义是：“样本内所含个体数量的多少”（袁方主编，1997：224）。

“应文”作者自己的教材中给出的定义是：“样本中所包含元素的数量”（郝大海，2005：38）。其在“应文”中所给出的定义是：“调查样本中所包含的被调查者数量”（郝大海，2007）。

仅从定义来看，以上定义的含义基本一致，不能说笔者的定义对样本规模的含义有什么“误读”。实际上，“应文”作者也没有指出笔者在定义上的“误读”。其指出笔者“误读”的内容和依据，主要体现在其所列举的一项调查实例上。^①

“应文”作者指出：“不难看出，‘高文’作者在计算调查回收率时，对5900这个数字的确切含义发生了误读……可以认定使用CGSS2003数据的研究论文给出的5900，应该是没有计入无应答数量的样本规模，因此，CGSS2003的回收率不应该据此被计算成99.9%”（郝大海，2007）。

的确，笔者前文中所理解的5900，就是CGSS2003调查的样本规模，就是该调查“计划调查的总的个案数”，或者说，就是抽样设计所抽取的

^① 需要说明的是由于“应文”许多地方将这种讨论直接具体化到CGSS2003调查，所以，尽管笔者主观上并不愿意如此进行讨论，但作为回应，笔者也只能被动地陷入到对这一具体调查的讨论中。不过，在讨论有关这一调查的具体问题之前，笔者首先声明：本文对该调查中任何问题的讨论，只针对该具体问题，而不涉及对该项调查总体质量的评价。

总个案数,也即是福勒定义中的“从总体中抽取的所有人,包括那些虽被抽中但因拒绝、语言问题、疾病或者缺乏易得性而没有应答的人”。

先来看看笔者是依据什么做出这种理解的。正如“应文”作者所说,从采用这一调查数据发表的5篇论文的样本介绍中,读者(包括笔者)所能看到的都只有“按抽样方案抽取了5900位城镇居民,实际完成样本5894个”这样的陈述。笔者的确是按照前述样本规模的定义,并根据“抽取了5900位城镇居民”这样的陈述而将5900作为CGSS2003调查的样本规模的。

笔者对该调查中5900这个数字的这种理解错了吗?请看下面由中国人民大学统计学系统计科学研究中心为该项调查撰写的“抽样设计、抽样框和样本抽取过程的说明”中对5900这一数字的来源、含义的具体说明:

由于调查的结果主要是估计各种比例数据以及比例数据之间的比较,所以在调查样本量的确定上是以估计简单随机抽样的总体比例 P 时的样本量为基础。在95%的置信度下按抽样绝对误差不得超过3%的要求进行计算,需要抽取样本量:

$$n_0 = \frac{u_{\alpha}^2 p(1-p)}{d^2} = 1067$$

这里 d 为抽样绝对误差取0.03, u_{α} 在置信度为0.95时为1.96, $p(1-p)$ 最大取0.25。由于采用多阶段的复杂抽样,设计效应 $deff$ 一般会在2和2.5之间,我们把 $deff$ 定为2,这样需要的样本量就为2000个。

综合考虑精确度、费用以及调查实施的可行性等因素,以及以往若干全国社会调查的经验;再加上考虑到在调查实施中通常会存在一部分户内找不到、或没有合格调查对象、以及各种原因造成的无回答等情况,根据对回答率的估计,需要将上述样本量适当扩大为2380个。

这个样本量是对全国精度而言的,由于我们的抽样还要满足沿海、中、西部和省会城市直辖市不同区域的对比,对不同区域总体估计精度的要求,所以样本量要求在此基础上相应扩大。考虑到总体分为沿海、中、西部和省会城市直辖市这四个层,我们把样本量初步确定在 $2380 \times 4 = 9520$ 个。

除此之外,考虑到北京、天津和上海三大直辖市的特殊地位,我们给每个城市的市区分别增加160个住户样本。本期调查的总样本量为 $9520+160\times 3=10000$ 。这10000个样本不仅能满足对总体的估计,而且也能满足对抽样框各自总体的估计,所以是比较合适的样本量……

考虑本次调查的具体内容,在10000个样本中,欲实现城市样本数与农村样本数的对比约为6000:4000……

最终抽样单元中城市样本与农村样本的对比为5980:4020,基本满足总样本量在城乡分配上的要求。(详见中国综合社会调查,CGSS2003中抽样方案,83.rar)

在该中心提供的另一份“抽样框和抽样过程说明”文件中,最后一段的表述是:

最终抽样结果是,共计抽出125个区县(初级抽样单元),二级抽样单元中街道数与乡镇数的对比为295:205,对应的三级抽样单元中居委会与村委会的对比为590:410,对应最终抽样单元中城市样本与农村样本的对比为5900:4100,基本满足总样本量10000在城乡分配上的要求。(详见中国社会调查开放数据库网站,CGSS2003相关文件中5.rar)

上述抽样说明中的“再加上考虑到在调查实施中通常会存在一部分户内找不到、或没有合格调查对象,以及各种原因造成的无回答等情况,根据对回答率的估计,以及总体划分为五个抽样框,我们把样本量确定为10000个”。“对应最终抽样单元中城市样本与农村样本的对比为5900:4100”等内容,不是清清楚楚地告诉我们:5900是“考虑了各种原因造成的无回答情况”后所确定的城市调查的样本规模吗?事实证明,笔者的理解与抽样设计中的定义和说明完全一致。

因此,可以确认的第二点是:不仅笔者对样本规模的定义与其他学者的定义是一致的;笔者对“应文”所列举的CGSS2003调查项目中的样本规模——5900的定义和理解也是清楚的和确切的。笔者并没有“误读”这一样本规模的含义。

四、为什么不能按“应文”的方式来计算样本规模？

其实，“应文”作者之所以认为笔者“误读”了5900，依据的并不是前述的样本规模定义，而是依据其在该文中提出的对样本规模的新定义和计算方式（即第一种方式）：

在制定抽样方案时，首先要确定样本规模，即确定调查样本中所包含的被调查者数量。由于在调查执行中无法百分之百地调查到所有预先选定的被调查者，因此，根据是否计入调查中的无应答数量，样本规模可以有两种取值方式。如果不计入无应答数量，那么样本规模只是依据研究需要和经费预算确定的一个理论值。而在调查执行中，对于那些无应答的被调查者，需要使用备选的替换样本来补足，这样实际调查的被调查者数量就要大于样本规模的理论值。如果制定抽样方案时，事先计入了无应答数量，那么就要设定一个大于研究需要的被调查者数量的样本规模。在调查执行中，那些由于无应答而减少的被调查者数量，就会使成功完成调查的人数接近研究需要的被调查者数量。（郝大海，2007）

笔者认为，“应文”作者所依据的这种样本规模的定义和计算方式是不正确的。不能按这种方式来定义和计算样本规模。理由有三。

第一，按“应文”这种方式定义和计算的并不是样本规模。

让我们来分析上述关于样本规模及其计算方式的论述中所存在的错误（括号中是笔者的提示，粗体为作者所加）：

在制定抽样方案时，首先要确定样本规模，即确定调查样本中所包含的被调查者数量。（为防止产生误解，请读者记住，这里的“被调查者数量”是福勒定义中所说的“所有的”被调查者，也即是“应文”作者下面所说的“所有预先选定的被调查者”。）由于在调查执行中无法百分之百地调查到所有预先选定的被调查者，因此，根据是否计入调查中的无应答数量，样本规模可以有两种取值方式。（读者可以思考一下，无论是否计入调查中的无回答数量，这种“所有预先选定的被调查者”的数量会发生改变吗？）如果不计入无应

答数量,那么样本规模只是依据研究需要和经费预算确定的一个理论值。(注意,“应文”作者在这里变换了概念,这句话中的“样本规模”已不再是前面定义中的那个“样本规模”,不再是前面定义中的“样本中所包含的被调查者数量”,也不再是“应文”作者上面所说的“所有预先选定的被调查者”了。他将前述定义中的样本规模变成了一个规模更小的“理论值”。这句话也就成为他否定5900作为样本规模的依据。)而在调查执行中,对于那些无应答的被调查者,需要使用备选的替换样本来补足,(注意,“备选的替换样本”是并不包括在前述样本中的另一个样本!这句话就为他得出比“理论值”更大的“样本规模”提供了可能。)这样实际调查的被调查者数量(注意,“应文”作者在这里又一次偷换了概念,他将前述的“样本规模”用“实际调查的被调查者数量”来做了替换。正是因为有了这样的替换,他在后面就可以名正言顺地去按后者来计算新的样本规模和回收率了。)就要大于样本规模的理论值。(两个正数之和当然要大于其中之一!)如果制定抽样方案时,事先计入了无应答数量,那么就要设定一个大于研究需要的被调查者数量的样本规模。在调查执行中,那些由于无应答而减少的被调查者数量,就会使成功完成调查的人数接近研究需要的被调查者数量。

上述分析表明,“应文”作者在上述论述中所说的“样本规模的两种取值方式”,实际上定义了两个不同的“样本规模”:其中的第二种方式(即所谓计入无应答数量的方式)定义的是前述几个定义中所说的样本规模。而其第一种方式(即所谓不计入无应答数量的方式)所定义的并不是前述几个定义中所说的样本规模。它实际上是将原有样本规模的概念做出改变,并增加了另一个样本后,在这两个样本中的“实际发生的访谈数量”!这样一种“实际发生的访谈数量”与上述这段论述的开头所定义的样本规模,在含义上已有本质的不同。它再也不是“所有预先选定的被调查者”了,再也不是福勒定义中的“从总体中抽取的所有的人”了。

第二,实际调查中样本元素有无回答的问题与样本规模无关。

本质上,样本规模是属于“抽样”范畴中的概念,而非属于“实地调查过程”中的概念。如果要介绍或讨论有关样本规模的内容,各种研究方法著作的作者们都只会介绍和讨论抽样问题的地方去介绍和讨论

它。并且,所有的作者在介绍如何计算样本规模大小时,涉及的也往往只是诸如总体的异质性程度、可接受的置信水平、允许的误差大小等因素。比如,“应文”作者就是在自己著作的“基本抽样方法”一章中对样本规模进行介绍的。他也明确指出过:“如果先不考虑经费问题,那么样本规模主要取决于抽样精度和总体标准差。一般来说,抽样希望达到的精确度越高,总体的异质性程度越显著,所需的样本规模就越大”(郝大海,2005:39)。

至于样本中的个体在实际调查过程中是否可以接触到、他们是否愿意接受调查、他们最终是否完成了调查等等,均与抽样设计所确定的样本规模无关!换句话说,样本规模是研究者在实地调查开始之前,为解决如何选取有代表性的调查对象这一问题时,所做出的计划和决定之一。只有在抽样设计中才会涉及到样本规模的问题。至于样本中的这些被抽中的个体最终完成调查的情况(比如问卷收回或未收回、访问成功或不成功、电话有应答或无应答等等),并不是样本规模所涉及和所反映的问题。

第三,按“应文”的第一种方式定义的样本规模是抽样设计时所无法确定的。

正如“应文”作者所说,所有调查的样本规模都是研究者在调查开始之前决定的,“从调查执行的发生流程看,无论是邮寄问卷,还是当面或电话访谈调查,都先要制定抽样方案。而在制定抽样方案时,首先要确定样本规模,即确定调查样本中所包含的被调查者数量”(郝大海,2007)。所以,样本规模是在实地调查开始前就确定好的和已知的。并且,样本规模一旦确定后,它“所包含的被调查者数量”,无论是800,还是1000,也都是确定不变的。不管实际调查过程中无法接触、无应答的人数是200,还是300,都不会使原有的样本规模数字发生任何改变。但是,按照“应文”第一种方式定义的样本规模,即“实际发生的访谈数量”,却是一个在实地调查前无论如何都无法确定、无论如何无法知道的数字,是一个必须要到实地调查全部结束后、通过统计才能知道的数字。因此,按这种方法来定义和计算的“样本规模”(即实际发生的访谈数量)是任何一种抽样设计都无法设计出来的!

正是由于“应文”作者不恰当地将“样本规模”的概念变换成“实际发生的访谈数量”,不恰当地在有关“样本规模”的讨论中引入“实地调查过程”的内容,同时又将调查回收率的定义和计算掺杂在样本规模的

讨论之中,因而形成了对样本规模含义的不正确理解,做出了对样本规模含义和计算的不正确划分。

五、CGSS2003 的样本规模究竟是多少？

笔者注意到,虽然“应文”作者认为笔者对 CGSS2003 项目的样本规模理解有误,可他在“应文”的通篇论述中,却始终没有明确告诉我们 CGSS2003 的样本规模究竟是多少!换句话说,“应文”作者始终没有明确地告诉读者,对于 CGSS2003 项目来说,这个福勒定义中的“从总体中抽取的所有人”究竟是多少!照理说这本该是个很简单的问题。因为对于任何一项调查研究来说,这都是研究者事先必须确定的。但现在由于“应文”作者对样本规模、对调查回答率的不正确理解和不正确的计算方法,使得这一问题的答案变得有点复杂了。

先来看看我们已有的两个数字:5900 和 8430。下面是它们的来源和含义。

1. 5900。这本来是该项调查的样本设计单位根据研究方案的要求,考虑了大约 20% 的无回答比例以及众多其他因素后确定的计划调查的总样本规模(详见前述的 CGSS2003 调查抽样方案)。但是,由于实际调查中研究者针对无应答现象补充了大量的替换样本,因而 5900 最终演变成了 CGSS2003 调查中“通过使用替换样本后成功完成访问的总数量”,即由计划中的“分母”,变成了实际中的“分子”。所以,它在客观上已不再是原有的样本规模。

2. 8430。这是“应文”作者根据 30% 的无应答比例和 5900 个成功访问的数量推算出来的“实际发生的访问数量”,也即是他所说的“计入了无应答数量后的样本规模”。根据前面笔者的分析,无论从定义上,还是从其实际内涵上,这个所谓的“计入了无应答数量后的样本规模”并不是真正意义上的“样本规模”。它只是 CGSS2003 调查中“实际发生的访问数量”,而不是“从总体中抽取的所有对象”。

事实上,8430 这一数字甚至连“实际发生的访问数量”都不是。因为作者计算出这一数字时所依据的“30% 无应答比例”本身就不确定、不可靠。在“应文”中,作者写道:“相关资料显示,在 CGSS2003 中,因为各种原因访谈未成功的比例大约是 30% (边燕杰等,2006: 108)”。光看

这一句话,读者可能会觉得这个 30%是客观的、清楚的、明确的,但实际上,这一“相关资料”究竟是如何“显示”的呢?下面是这一资料的原文照录:

这次上海的调查,在第一批入户名单 400 户中,因为各种原因访问未成功的占 47%(全国的情况,综合各地督导的经验大概是 30%),其中拒访和访问中途拒访的正好是 100 户,占样本数的 25%(见表 5-1)。

表 5-1 访问接触记录表

访问失败原因	频数	百分比(%)
成功访问	212	53.0
访问失败	188 ^①	47.0
拒访或中途拒访	100	25.0
地址不清或无法找到	20	5.0
空户,无人居住	13	3.3
三次接触后找不到被访者	14	3.5
被访者在访问期间无法访问到	10	2.5
非居民户	10	2.5
未注明失败原因	21	5.3
合计	400	100

资料来源:边燕杰、李路路、蔡禾,2006:108-109。

上表中上海调查的各方面情况、人数、比例都清清楚楚。其访问未成功的比例 47%,是根据各种未成功访问的 188 人计算得到的。但这里的 30%则与 47%大不相同,30%仅仅是来自于“综合各地督导的经验”,并且只是一种“大概”的比例,而不是来自客观的、清楚明确的调查记录和访问统计!那么,这种通过“综合”“经验”后得到的“大概”比例,能给我们一个确切的、客观的数字吗?我们又有什么理由来相信依据这样的数字所推算出来的结果呢?!

笔者在这里根据相关信息,再给出与 CGSS2003 样本规模相关的另

① 原书中误印为 198。

外两个不同的数字: 8850 和 7662。下面是这两个数字的含义及来源。

3. 8850。这是笔者根据香港科技大学社会科学部社会调查中心网页上(香港科技大学调查研究中心, 2004a)公布的“在抽取户样本时, 应至少多抽取 50% 的备用样本”的抽样说明, 按多抽取 50% 的备用样本计算得到的(计算方法: $5900 \times (100\% + 50\%) = 8850$)。笔者认为, 如果实际多抽取的备用样本的确是 50%, 同时, 如果该调查将备用样本全部调查完, 那么, 这个数字就是该调查实际的“计划调查的总样本规模”。但遗憾的是, 在实际调查中, 研究者是按照“当 5900 的预定目标达到时, 调查就结束了”的方式进行的, 因此, 当达到 5900 个成功访问的个案时, 多抽的 50% 的备用样本是否全部调查完我们并不知道。所以, 如果 8430 是“实际发生的访谈数量”, 那么, 8850 就没有用完, 因而它也不能作为真正意义上的样本规模。

4. 7655。这是笔者根据香港科技大学社会科学部社会调查中心网页上公布的“访问结果统计, 共在 559 个居委会中成功完成问卷 5894 份, 访问成功率为 77%”的信息(香港科技大学调查研究中心, 2004b)计算得到的(计算方法: $5894 / 77\% = 7655$)。根据前述的调查方式, 这也是“应文”作者所定义的那种“计入了无应答数量后的样本规模”, 也即“应文”所说的“实际发生的访谈数量”, 这个与 8430 含义相同、但取值却不相同的数字, 虽然同样也不是我们所要寻找的样本规模, 但它却从另一个方面对“应文”作者推算的 8430 的可靠性提出了质疑。

上述分析表明, 尽管我们总共找到了四个相关的数字, 但它们中一个也不是我们所希望找到的 CGSS2003 的样本规模! CGSS2003 的样本规模究竟是多少呢? 这成了一个在该调查结束几年以后仍无答案的问题! 为什么对于这样一个看起来本不应该成为问题的问题, 我们却会陷于这样一种很尴尬的境地?!

下面, 让我们来设想另一种情况: 如果 CGSS2003 的研究者在整个调查中, 始终把依据严格的抽样设计所确定的 5900 户作为样本规模, 不再另外抽取备用的替换样本, 在实际调查访问过程中也不进行样本替换。这样, 假定调查中真的出现了 30% 的无应答现象, 即访问失败了 1770 户, 最终得到的成功访问的数量为 4130 户。那么, 无论是回答样本规模的问题, 还是回答调查回收率的问题, 一切都非常简单, 一切都非常清楚:

样本规模为 5900;

调查回收率(按笔者的定义,等于访问成功的数量/计划访问的数量;按福勒的定义,等于应答者数量/抽取的所有的人)= $4130/5900=70\%$ 。

虽然有人可能会说,4130 这个有效样本的数量比 5900 少了一些,但事实上这样的有效样本数量似乎也够了。更关键的是,这样做的结果不仅能使调查的总工作量大大减少,或许还能使得调查的总偏差变得更小。

六、在研究报告中应该怎样报告样本规模和调查回收率?

“应文”作者为了说明笔者不仅对 CGSS 的样本规模发生了“误读”,而且对其他调查的样本规模也发生了“误读”,专门打电话询问了一些调查的负责人,得出了“应文”中的统计表 1。对表 1 中 10 项随机抽样调查的样本规模和回收率进行了再次分析后,“应文”作者认为,由于缺乏“与实际发生的访谈数量有关的数据”,总共 10 项大型调查中,只有 4 项调查“可以计算出调查回收率”。“其他几位调查主持人均报告,各项调查在执行过程均发生过样本替换,但没有保留样本替换的相关资料,因此无法提供准确实际发生的访谈数量”,因而无法计算回收率。这也即是说,按“应文”作者的理解,这几项调查的研究者虽然已经用调查数据发表了论文,但是,他们实际上连自己所做调查的样本规模是多少都不知道!他们在论文中对抽取的样本数目、回收的问卷数目等等所作的描述,也都不能用来说明他们调查的回收率。这能说得过去吗? (“应文”作者认真询问当事人的做法,是一种实事求是的态度。但笔者的疑问是:为什么只询问他们“有没有替换样本的情况”,而没有询问他们:“你们的样本规模是多少”?)

让我们实际看看这些作者是如何表述自己的样本规模和回收率的(请注意说明中的粗体字部分):

1. 中国社会结构变迁调查。“我们采用多阶段复合抽样的方法,从全国 3300 个市县级抽样单位抽取了 73 个市/县/区的 408 个居民委员会/村民委员会的 6240 名 16—70 周岁的居民进行调查。”“最终获取的有效样本为 6193”(陆学艺, 2004: 363; 李春玲, 2005: 18)。

2. 北京市社会网络与健康调查。“按照简单随机原则获得 1677 个

住宅地址作为调查样本……最终成功访问了 1004 位 18 岁以上的在职或退休的城市居民”(张文宏, 2005)。

3. 长春市居民入户问卷调查。“具体的抽样方案是: 从吉林省长春市的 5 个区中, 每个区抽取 4 个街道(总共 20 个街道), 每个街道抽取 5 个居民委员会(总共 100 个居民委员会), 每个居民委员会抽取 5 个家庭(总共 500 个家庭), 每个家庭访问 1 人, 总共组成 500 个样本……调查过程中有 2 个家庭无法访问, 资料的整理过程中又发现有 2 份问卷大部分没有回答, 实际统计样本数为 496 个”(张海东, 2004)。

4. 转型时期中国大陆城市居民的阶层意识调查。“共发放调查问卷 830 份, 回收有效问卷 812 份, 有效回收率为 90%^①”(余红、刘欣, 2004)。

5. 性社会学全国抽样调查。“结果有效应答率是 76.3%(3824 个样本)”(潘绥铭, 2004)。

6. 全国综合社会调查(城镇部分)(GSS2003)。“本次调查利用 2000 年进行的第五次全国人口普查数据, 采用 PPS 抽样从省级单位抽到社区, 再在每个社区以系统抽样抽取 10 个调查户, 用 Kish 表每户抽取 1 人作为调查对象。调查对象为 18—69 岁的成年人; 共抽取 5900 人。调查方式为入户面访, 实际完成问卷 5894 份。经过数据清理, 最后样本量为 5894”(冯仕政, 2006)。“采用 PPS 与分层设计的抽样方法共抽取了 5900 位城镇居民, 实际完成样本 5894 个”(刘精明、李路路, 2005)。

7. 中国单位调查。“原定的调查计划是在 10 个城市中抽取 10 个单位, 再在每个单位中抽取 40 个人……这次调查共获取 3293 个有效问卷”(王天夫、王丰, 2005)。

8. 当代中国社会变迁、社会意识与社会流动调查。“共发放问卷 3780 份, 回收有效问卷 3183 份”(许欣欣, 2005)。

9. 福建二市县居民入户问卷调查。“再从各个乡镇中按随机方法各抽取 4 个行政村, 每个行政村再按随机方法抽取 25 个 18 岁以上村民, 共 1000 人……本次调查共成功访问村民 913 人。其间, 每个村大约成功访问 20 至 25 人”(胡荣, 2006)。

10. 湖北三城市居民入户问卷调查。“本次共发放问卷 1000 份, 回收有效问卷 973 份”(慈勤英、王卓祺, 2006)。

① 该百分比计算有误, 原文如此。

笔者多少有点替这些作者感到冤枉, 他们的说明中明明有样本规模, 明明可以计算回收率(有的甚至已经说出了回收率的多少), 可是被“应文”作者用不正确的标准一衡量, 他们的样本规模就变得不再是样本规模了, 他们的回收率也不再是回收率了。问题是: 这些作者都会承认这一点吗?

谈到研究报告中对样本规模的报告问题, 我们还可以用“应文”作者所举 CGSS2003 为例来说明。细心的读者可以发现一个非常奇怪的现象: 从 CGSS2003 调查结束直到今天发表的“应文”出现之前的几年时间内, 无论是从已经发表的论文中, 还是从主持该调查的单位网站上所公布的调查基本情况中, 全都看不到这次调查的样本规模是多少! 所有这些地方能看到的, 都只有“抽取了 5900 个样本”这样的说法! 我想, 如果不是“应文”作者通过质疑的文章告诉我们, 恐怕不仅仅是笔者, 所有读者可能都不会知道, 在“预定样本规模”5900 的背后, 还有着 8430 这样一个“实际样本规模”的数字!

因此, 笔者在此不禁要问一句: 造成这一现象的原因, 究竟是由于那些撰写论文的作者、调查的主持者不愿意向读者告知这一描述调查基本情况的数据? 还是由于他们自己本身就不知道有 8430 这样一个“样本规模”的数字存在? 甚至是“应文”作者自己, 在此之前也不知道有 8430 这样一个数字存在。不是吗? 看看“应文”作者在其著作中介绍样本规模计算时所举例子的原话吧: “在 2003 年全国综合社会调查 (General Social Survey, GSS) 中, 样本规模为 10000, 其中城市样本规模为 5894” (郝大海, 2005: 41)。而上述现象所反映的, 究竟是这些研究者不重视报告样本规模的问题, 还是反映了“应文”作者对样本规模的认识上存在着偏差?

此外, “应文”中对样本规模的表述与其表 1 中的界定之间也存在着明显的自相矛盾。比如, “应文”作者前面曾反复指出笔者“误读”了 CGSS2003 的样本规模, 指出 5900 并不是这一调查的样本规模。可是, 在表 1 中, 他却又清清楚楚、明明白白地在 CGSS2003 的“样本规模”一栏里写着 5900, 而不是写着 8430 或者别的什么数字! 这种自相矛盾的状况, “应文”作者又该如何解释呢? 我们能不能也提出同样的问题: 应文作者“误读”了 CGSS2003 的样本规模吗? 又比如, 按他前面的说法, “可以认定使用 GSS2003 数据的研究论文给出的 5900, 应该是没有计入无应答数量的样本规模”, “在访谈调查中, 调查员会按被抽中的被调查

者地址对其进行访谈,而访谈数量则取决于样本规模,如果成功访谈数量达到了样本规模,调查就结束了,否则就需要继续访谈”。因此,5900是CGSS2003调查中按上述方式所得到的“成功访谈数量”。那么,它与表1中的“完成问卷数量”5894之间又是一种什么样的关系呢?CGSS2003调查究竟成功完成了多少份问卷?是5900还是5894?

还有,“应文”表1中“严格回收率”的提法,实际上也是不充分的:因为他自己也知道,表1所指出的四项可以计算“严格回收率”的调查中,其他三项调查的回收率都不是按他所提出的方法来计算的。读者可以看到:在有“严格回收率”记录的四项调查中,另外三项的“访谈被调查者数量”实际上全部都是笔者所说的“计划访谈数量”,即样本规模。所以,表1中这三项调查的“样本规模”一栏中的数字与“访谈被调查者数量”一栏中的数字完全一样(北京市社会网络与健康调查中的1004为笔误,应为1677),惟独CGSS2003一项调查的这两个数字不一样!如果“应文”作者真的对自己的算法有信心,他就应该在表1中的“访谈被调查者数量”一栏中写上“样本规模”的字样,而在表1中现在的“样本规模”一栏中写上“预先确定的理论值”。这样就更加符合他所提出的样本规模的含义及其计算方式。他之所以不能这样写,主要原因就是其他几项调查的“样本规模”事实上都是真正的样本规模,都不可能标为“预先确定的理论值”!

在结束本文前,笔者再次主张:不管我们的实际调查做得是好是差,严格或不严格、科学或不科学,质量高或不高,研究者都应如实地、清楚地、详尽地在报告中告诉读者自己是如何做的。这是科学共同体中的一条基本原则。

参考文献:

- 边燕杰、李路路、蔡禾主编,2006,《社会调查方法与技术:中国实践》,北京:社会科学文献出版社。
- 慈勤英、王卓祺,2006《失业者的再就业选择——最低生活保障制度的微观分析》,《社会学研究》第3期。
- 冯仕政,2006,《单位分割与集体抗争》,《社会学研究》第3期。
- 风笑天,2005a,《现代社会调查方法》(第三版),武汉:华中科技大学出版社。
- ,2005b,《社会学研究方法》(第二版),北京:中国人民大学出版社。
- ,2007,《高回收率更好吗?——对调查回收率的另一种认识》,《社会学研究》第3期。
- 郝大海,2005,《社会调查研究方法》,北京:中国人民大学出版社。

- , 2007,《应答率的意义及其他》,《社会学研究》第6期。
- 胡荣, 2006,《社会资本与中国农村居民的地域性自主参与——影响村民在村级选举中参与的因素分析》,《社会学研究》第2期。
- 刘精明、李路路, 2005,《阶层化: 居住空间、生活方式、社会交往与阶层认同——我国城镇社会阶层化问题的实证研究》,《社会学研究》第3期。
- 陆学艺, 2004,《当代中国社会流动》,北京: 社会科学文献出版社。
- 李春玲, 2005,《断裂与碎片》,北京: 社会科学文献出版社。
- 潘绥铭, 2004,《性社会学基本命题的实证》,《社会学研究》第6期。
- 王天夫、王丰, 2005,《中国城市收入分配中的集团因素: 1986-1995》,《社会学研究》第3期。
- 香港科技大学调查研究中心, 2004a,《2003年度中国综合社会调查》, http://www.ust.hk/%7Ewebssc/survey/HHsampling03_C.pdf
- , 2004b,《2003年度中国综合社会调查》, <http://www.ust.hk/%7Ewebssc/survey/GSS2003c4.html>
- 许欣欣, 2005,《社会、市场、价值观: 整体变迁的征兆》,《社会学研究》第4期。
- 杨伟民, 2005,《当前中国的社会保险在社会分层中的作用》,《社会学研究》第5期。
- 余红、刘欣, 2004,《单位与代际地位流动: 单位制在衰落吗》,《社会学研究》第6期。
- 袁方主编, 1997,《社会研究方法教程》,北京: 北京大学出版社。
- 张海东, 2004,《城市居民对社会不平等现象的态度研究——以长春市调查为例》,《社会学研究》第6期。
- 张文宏, 2005,《城市居民社会网络资本的阶层差异》,《社会学研究》第4期。
- 中国大百科全书总编辑委员会《社会学》编辑委员会, 1991,《中国大百科全书社会学》,北京: 中国大百科全书出版社。
- 中国社会调查开放数据库网站, http://www.cssod.org/show_survey.php?SurveyId=3
- 中国综合社会调查网站, <http://www.chinagss.org>

作者单位: 南京大学社会学系
责任编辑: 杨可

reflections about the social workers' experiences. In their journey with the villagers, the researchers began to realize that cultural illiteracy is an essential obstacle of social work intervention in rural China. The authors argue that if we are insensitive to the local culture and context, professional knowledge can be useless to the local people, or even worse, if not handled carefully, can become cultural invasion to the local community.

DEBATES

The Significance of Response Rates and Others: Another explanation to “high” response rates for Chinese social survey *Hao Dahai* 180

Abstract: This is a response to the critique of the paper “Is It Really Better to Have Higher Response Rates?” published earlier in this journal. The author demonstrates in the present article that the previous critique was based on misunderstanding of the key concept of survey research: the sample size. This fundamental misunderstanding invalidates the central arguments made in that earlier article, including the critique to the Chinese General Social Survey (CGSS). This paper presents evidence showing that several recent large-scale social surveys yield similar response rates. I argue that the issue of low response rate has not attracted adequate attentions from survey researchers. Survey researchers in China are facing very different social reality from 10 years ago, when response rate was a trivial issue. As it is getting increasingly difficult to get adequate cooperation from respondents especially urban respondents, it is the time for us to give some serious thoughts on the issue of low response rate and try to work collectively and creatively on possible solutions.

On the Sample Size and the Response Rate *Feng Xiaotian* 195

Abstract: The paper responds to the critique by the paper “The Significance of Response Rates and Others”. The author points out firstly that the critique’s starting point is improper. The author argues that the definition of the response rate given by himself is the same as that given by Foler, and the author’s understandings to the concept of sample size, and to the sample size of CGSS2003 are clear and correct, in other words, the author does not misunderstand the meaning of the sample size. The paper points out that it’s the author of the article “The Significance of Response Rates and Others” who misunderstands it by changing the conception of the “sample size” to “the number of people interviewed”, and has improperly introduced the “course of survey” to the discussion of the “sample size”. At the same time, it brings confusion about the sample size of CGSS2003.

BOOK REVIEW