

抽样调查的工程化：从面访到问卷星

■ 黄向阳

先请读者们想一想：每年会接到几个电话调查，又会帮孩子的同学们填写几次问卷星？当然，第二个问题主要适合家有中小学生的家长。

在 2022 年，抽样调查（或简称调查）已经成为日常词汇的一部分，因为其自带数据光环，还成了媒体热搜中的高频词，每每被引用来作为重要论据。此情此景，似乎应该让统计人感到自豪吧，然而大众对抽样调查、问卷和量表的解读其实相当随意，不说小学生和中学生的问卷调查了，即使在大学里，情况也没有好到哪里去。有时面对这些所谓问卷调查分析报告，真的很想批个不及格。那么问题出在哪里呢？为什么很多学过统计课程的大学生，对问卷调查的理解、掌握和执行却如此拉胯？按照某位名人说过的思考方式，文科生苛责于人，理科生反求诸己。就是说，文科生在分析社会现象的时候，喜欢要求大家加强学习，提高素质；理科生（尤其是工程师）则相反，更倾向于在自己身上找原因，是不是产品设计不合理，或者哪些地方讲得不好。

这里留下了一个小问题：纯数学家属于哪一类？纯数学家属于第一类，因为你学不好数学的确是你自己的问题。有一位统计学家曾经敏锐地指出了统计和数学的区别：“数学家只管教数学就行了，统计学家还要教大家怎么‘用统计’。”这就是说，有些人指责统计会撒谎，并不是完全没有道理，尽管真正的原因是没有学会怎么“用统计”。怎么用好统计的确是个巨大挑战。到 20 世纪 70 年代后期，计算机的普及已经成为可以预见的发展趋势了，这个新的技术手段带来了更加物美价廉的统计软件和计

算机辅助电话调查系统（Computer Assisted Telephone Interviewing System, CATI）。然而，怎么用好统计的难题是技术手段能够解决的吗？也许恰好相反，貌似简单的、即插即用的黑箱式统计工具无疑是普及统计方法的利器，但副作用也不能忽视，它会造成更多的“伪调查”和“伪统计”，让统计方法沦为措辞数学的应用（措辞数学，就是将数学公式和数据分析当作貌似科学的论据）。在 20 世纪 90 年代初，有一位统计学家对统计软件的滥用表示过担忧：“如今随便哪个大学生都会用统计软件做回归分析，但是把统计软件交给没有受过严格训练的人，就像把链锯交给一个小孩，不是伤着自己就是伤着别人。”

因为“用统计”本质上是个工程化问题，工程化需要设计能力和执行能力，而这些内容恰恰是大学统计课程的弱项，本文将对抽样调查方法的工程化做一个简单的回顾。

在课程中不受重视的抽样方法

都说读史使人明智，但我的个人经验却是读史让我吃惊。比如，在 1998 年翻译出版的《统计的未来——国际统计发展展望》中，对调查方法在大学统计课程中的地位就有令人惊讶的说法。这本书第 44 页的论文标题是“政府统计机构中统计方法创新的挑战”，作者是密执安大学与美国调查方法联合研究中心的罗伯特·格罗夫斯（Robert Groves），论文主题是普查和抽样调查中的统计方法创新不足现象及其改进对策。文中指出：

但是抽样技术始终处在一个相对弱势的地位，



看看统计系专业课程设置就可以知道，没有独立的量表设计课程，在其他课程中也很少展开论述量表设计和调查项目实施过程或质量控制。

接着在第 63 页回顾历史：

统计系在某种意义上说自然是调查统计学家知识的家园……（但是）许多大学统计系中都不曾有过调查统计的黄金时代。科克伦、戴明、汉森、哈特利、拉扎斯菲尔德、马哈拉诺比斯、斯蒂芬、耶茨等人都教过许多人，但从总体上看，他们在学院外部进行的教学多于学院内部。

在同一页上提到了过去 20 年中（即从 20 世纪 70 年代中期到 90 年代初），大规模调查统计取得了许多重大进展。包括，心理学观点大幅改善了量表设计，贝叶斯方法用来补救无回答，计算机技术迅速改变了调查实施和统计分析。不过书中有个很消极的评价：“大量的革新和进展发生在各个相关学科的边缘地带，结果是，调查统计学家需要了解的知识很少能在大学中学到。”不过，创新活动边缘化在诸多学科中都不是什么新问题，更不是在统计领域中格外恶劣。本书在第 52 页曾引用罗杰斯 (Rogers) 在《创新扩散》中的观点：“在所有情况下，新知识的发现到应用它的时间似乎太长，超过了必要的程度。”

所以对于创新扩散迟缓的问题，统计学家不用过于纠结，这毕竟需要一个过程，重点在于“必要”的程度是多少年。须知，理论知识从学术认可到落地实施的工程化，再到说服用户和社会大众，需要一个时间，那么抽样调查经历的时间到底有多长呢？答案是一个世纪。

抽样调查的艰难历史

随机抽样从实践到理论差不多经历了一个世纪，陈希孺在《数理统计学简史》中对这段历史有精彩记录和评述，这里仅仅做个简单说明。早在 1802 年，拉普拉斯用“比例法”，通过抽样实施过对法国人口的估计。1861 年，英国的法尔也实施过人口抽样调查。除了零星个案之外，抽样调查的实践和理论在 19 世纪都没有太大进展。到 19 世纪最后 20 年中，挪威的凯尔 (A. N. Kiaer) 才在挪威的人口和农业普查工作中，逐渐发展出“代表性抽样”的思想。凯尔在 1895 年的国际统计学会大会上倡导抽样调查，但引发了激烈争论。不过到了 1903 年召开的国际统计学会大会时，凯尔的抽样调查主张就得到了多数人的赞同。凯尔的代表性抽样有两个要点，一个是样本必须有代表性，另一个是如果样本有很好的代表性，那么就不需要特别大的样本量。显然，第二个要点是违反直觉的。实际上，即使到了 21 世纪，仍然有不少人对于民意调查只需要几千个样本仍然感到惊讶和困惑。

遗憾的是，凯尔没有提出现代的随机抽样理论，所以 1903 年的胜利仍然缺乏根基。到 1906 年，英国的鲍莱 (A. L. Bowley) 将概率引进到抽样调查中，从而开启了现代的随机抽样理论和实践。一直到 1926 年的国际统计学会大会召开，基于概率的随机抽样方法终于得到了普遍承认。当然，理论上的论证还需要数理统计学家的的工作，其里程碑是奈曼在 1934 年发表的论文《关于代表性方法的两个不同方

面：分层抽样和目的性抽样》，从此以后，投入实施的抽样调查一般都要做分层抽样设计。在组织建设和推广普及方面的启动一直到 1947 年，联合国统计司才设立了一个抽样分委员会，从而推动了抽样方法在全球各国的普及。

了解战后社会经济统计发展历程的人还会注意到另一个问题，在 20 世纪 50 年代，苏联和中国的官方统计都曾经反对用基于概率的方法来处理社会统计问题。当然，这些反对意见后来都逐渐归于沉寂了，因为实践证明随机抽样方法是对普查方法的极好补充。这个波折其实有很深刻的科学方法论根源。我们现在都习惯于数学模型在各个应用领域之间的迁移和借鉴，但是这种做法在比较谨慎的科学家看来是不够妥当的。比如在 19 世纪围绕正态分布的争论中，很多人觉得正态误差论源自物理学和工程应用，不宜移植到社会统计问题的分析中。陈希孺在《数理统计学简史》中有多处分析过造成这种现象的原因，下面是一个不完整的列举。

在第 142 页指出，在 19 世纪很多社会学家认为社会数据不同于科学实验数据，很难保证数据的同质性，对于这些异质数据进行分析就比较困难。同理，对异质总体进行抽样的时候就必须用分层抽样。在第 143 页指出了 19 世纪的观念障碍，当时人们普遍认为，“适用于误差的规律未必一定适用于其他的数据”。类似的分析在 170 页再次出现，不过重点是对这个观念的理论突破，即埃奇沃斯在 1885 年的两篇论文《统计学方法》和《观测数据与统计数据》，指出概率模型可以用来分析社会经济统计问题。现在

学习概率论和统计学的人，都会快速接受基本操作：如果数据来自某个总体，就可以适用分析该总体的各种统计方法。也就是说，和数据的具体领域来源无关，而各种分析方法可以进行大胆的跨领域移植。

但是，从批判性思维的角度来看，19 世纪的学者们的质疑和犹豫并非吹毛求疵或者烦琐哲学，毕竟两种不同来源的数字能否用同一种方法来处理是值得认真思考的：一类数据是对单个对象的重复观测值（在自然科学领域很常见），另一类是对一批个体的各自观测值。

抽样调查的工程化和 CATI 系统

早期的抽样调查实践往往用来充当普查的补充和替代手段，从实践角度来看，普查意味着调查对象分散在很大的空间中，在依赖访员面访的时代里，无论普查还是抽样调查都需要足够的人力和资源，一言以蔽之，就是实施成本高昂，需要花很多钱。所以敢于实施抽样调查的机构都会高度重视有关的组织工作，所有这些工作都可以称为理论方法的工程化。

抽样调查的工程化包含两个基本方向。一个是问卷和量表的设计，另一个是访员的遴选、培训和督查。问卷不同于量表，汉语字面就不同，英文分别是 questionnaire 和 scale。问卷的内容相对来说比较自由，还可以包括开放式问题，你可以把问卷理解为一种访谈提纲。量表则不然，英文 scale 的本义是天平，就是测量工具，它的学科源头是心理学和教育学。到现在，重视量表设计的仍然是这两个学科，外加社会学，但是显然不包括统计学（这是本文第一节中引述的

格罗夫斯的评价，也符合我们的个人感受）。至于访员的培训和督查，如果没有严重社交恐惧症的话，受过中等教育的人就足以胜任访员的工作了。由此可见，除了调查方案设计之外，抽样调查的工程化在面访时代并不需要更多的数学工具，相反需要更多的心理学和语言学常识。据我所知，很多项目和学生论文经常忽视的环节居然是，问卷和量表设计阶段的可读性测试和预调查修正。最重要的原因可能是预调查和反馈调整太花时间了。

有了计算机之后，自然就会想到计算机化，中文有个形象的简称“电算化”。这个电算化几乎特指会计系统的电算化，想想也是，会计活动的第一要素就是合规，除了少数职业判断以外，绝大部分会计分录工作都可以机械化，而随后的数据处理更需要的是数据库管理，不是微积分和线性代数。这个案例告诉我们，适合电算化的系统本身要足够成熟，足够机械。而抽样调查的实施和问卷设计到20世纪70年代，已经比较成熟了。此外，当时在美国，计算机的软硬件成本正在持续下降中，固定电话也已经高度普及，这样就基本凑齐了电算化所需要的天时地利人和。电话普及率是CATI生效的前提（强调“固定”电话是因为当时移动电话基本上还在开发中），作为一个对照例子，我国到新千年之初还很少使用CATI系统，主要原因就是电话普及率不足。

最早编写和实施CATI系统的是加州洛杉矶分校的基于计算机的行为研究中心(the Center for Computer-based Behavioral Studies)和美国人口普查局(Census Bureau)，有兴趣的读者可以查阅项目组在

1978年发表的论文。该文指出“设计良好”的CATI系统有如下优点：提升数据质量；节约问卷准备和数据处理的时间；能够对问卷进行更加彻底的预调查测试；能够按照受访者的特点定制问卷中的问题；可以提高实施方法论研究的能力。

前面两条优点都是可以想象的，第三条的要点是CATI系统能够降低预调查的成本和时间，第四五两条才是CATI带来的新能力。这是新技术催生观念和行为的典型例子，其中第五条需要多做一点说明。有了CATI系统，就可以在调查过程中根据受访者的特点选用若干问题，如果是手工时代，只有很长的问卷才能容纳各种备选问题，在计算机辅助下则可预先准备大量备选问题，在调查过程中根据受访者的特点做适当调整。此外，还可以对问卷进行小的修改以测试不同问卷设计的实用效果，这个做法有些像网络时代的A/B测试。

结语：从问卷到问卷星

随着网络普及，尤其是智能手机的普及，新的技术手段又催生了一种新的调查平台，甚至不用依赖通过电话的即时交流。家有中小学生的恐怕都尝试过高度一体化的问卷调查，比如问卷星，你甚至可以从平台上下载数据文件，获得基本的可视化图表。这种工程化成就到底是好是坏呢？我想它取决于教师们传授正确统计实践的能力，真正的挑战仍然是，统计学家必须教会大家怎么用好统计。❏

作者单位：中国人民大学统计学院