

网络调查方法

Web Survey Methodology

马里奥·卡列加罗 卡佳·洛扎尔·曼弗雷达 瓦斯亚·维霍瓦尔 著

郭素然 译

网络调查方法

网络调查方法

马里奥·卡列加罗 卡佳·洛扎尔·曼弗雷达 瓦斯亚·维霍瓦尔 著

郭素然 译

马里奥·卡列加罗, 卡佳·洛扎尔·曼弗雷达, 瓦斯亚·维霍瓦尔
《网络调查方法》

Mario Callegaro, Katja Lozar Manfreda and Vasja Vehovar
Web Survey Methodology

出版方：斯洛文尼亚卢布尔雅那大学社会科学学院

出版人：萨莫·乌汉教授（院长）

卢布尔雅那，2026 年

Publisher: Faculty of Social Sciences, University of Ljubljana, Slovenia

For the publisher: prof. Samo Uhan, the dean

Ljubljana, 2026

© 马里奥·卡列加罗, 卡佳·洛扎尔·曼弗雷达, 瓦斯亚·维霍瓦尔

© Mario Callegaro, Katja Lozar Manfreda and Vasja Vehovar

中文翻译：郭素然

中文审校：刘丽丽

中文版式设计与封面设计：孙东亭

Chinese translation: Suran Guo

Chinese proofreading: Lili Liu

Chinese layout and cover design: Dongting Sun

英文原版：《网络调查方法论》，Sage 出版集团，2015 年

ISBN 978-085702-860-0，访问地址：<https://study.sagepub.com/web-survey-methodology>

English original edition: WEB SURVEY METHODOLOGY, Sage, 2015

ISBN 978-085702-860-0, accessible at <https://study.sagepub.com/web-survey-methodology>

中文版访问地址：<https://ebooks.uni-lj.si/ZalozbaUL> 与 <https://websm.org>

Chinese edition accessible at <https://ebooks.uni-lj.si/ZalozbaUL> and <https://websm.org>

<https://doi.org/10.51936/9789612951559>

斯洛文尼亚卢布尔雅那国家和大学图书馆编制的出版物编目记录（CIP）

Cataloging record of publication (CIP) prepared at the National and University Library in Ljubljana, Slovenia

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

[COBISS.SI](https://cobiss.si)-ID [279940611](https://cobiss.si/279940611)

ISBN 978-961-295-155-9 (PDF)

作者中文版序言

2015 年《网络调查方法》由 Sage 出版集团首次出版时，我们的目标是全面阐述网络调查的原理、实践与挑战。这一领域发展迅速，但仍缺乏系统的理论和方法论框架。此后，本书的反响与影响力让我们深感欣慰。

根据 Google Scholar 的数据，《网络调查方法》已成为网络调查领域被引用最多的专著，预计到 2026 年引用量将突破 1 000 次。该书被各学科的研究者、学生和专业人士广泛采用，自出版以来始终保持稳定的读者群体和销售量。据 Worldcat 的数据显示，全球已有 360 家图书馆收录了本书。

多年来，出版社、同行及读者均曾询问修订版的可能性。然而，由于三位作者都忙于其他事务，我们无法开展如此庞大的项目——尤其是初版的撰写已耗费了大量心血。此外，我们认为书中阐述的基本概念、术语、结构和流程至今仍具有重要意义。尽管该领域出现了一些显著发展——最明显的是在线样本库的扩展和移动设备的日益普及——但这些主要影响了调查设计和实施的实践层面，而没有根本改变其方法论原则。

真正新颖且具有变革潜力的，是近年来出现的生成式人工智能（AI）。生成式人工智能已开始影响本书所阐述的网络调查流程的十二个阶段，既带来了令人振奋的机遇，也引发了深刻的方法论、伦理和解释挑战，未来数年需要开展严谨的研究与创新探索。然而，现阶段全面探讨生成式人工智能与网络调查方法的融合可能仍为时尚早。

本书通过中文翻译版面向更广泛的读者群体，我们深感荣幸。我们衷心感谢郭素然博士的出色翻译，更重要的是感谢她率先推动中文版出版的远见卓识和积极行动。

同样感谢 Sage 出版集团将版权归还给作者。这一慷慨之举使我们能够将本书在线免费开放，进而为这部独立中文版的出版创造了条件。

我们还要感谢卢布尔雅那大学社会科学学院出版社，他们在斯洛文尼亚完成了这部英文原著的中文版出版工作，任务复杂且值得称赞。

我们希望本书的中文版能为中国网络调查研究与实践的持续发展做出有意义的贡献，同时促进这一方法论领域的国际对话与合作途径。

马里奥·卡列加罗 (Mario Callegaro)

卡佳·洛扎尔·曼弗雷达 (Katja Lozar Manfreda)

瓦斯亚·维霍瓦尔 (Vasja Vehovar)

译者序

网络调查作为社会科学研究、市场调研、公共政策分析等领域的重要途径，在数字时代的中国正迎来快速发展期。然而，相较于实践的蓬勃，国内相关方法论的系统梳理与规范指引仍有提升空间。原著三位作者基于长期的学术积累与实践经验，将网络调查的全流程拆解为十二个核心阶段，层层递进地阐释了其中的原理、实践与潜在挑战。这种结构化的呈现方式，既便于初学者搭建知识框架，也能为资深从业者提供进阶思考的切入点，对于中国当下的网络调查实践而言具有重要价值。

值得一提的是，本书原版出版后，网络调查领域涌现出诸多新变化——移动设备的全面普及、生成式人工智能技术的广泛应用，为这一方法学领域带来了新的机遇与挑战。原著作者在中文版序言中也提及，这些新发展并未动摇书中核心方法论的根基，反而为其提供了更广阔的实践场景。这也让我更加确信，将这部著作译为中文，不仅是对现有成熟方法学理论的引介，更是为中国学者参与相关领域的创新研究搭建了一座桥梁。由于译者水平有限，译文中难免存在疏漏之处，恳请各位读者批评指正。

郭素然

目 录

序 言	xvi
致 谢	xix
作者简介	xxi
译者简介	xxiii
缩略词列表	xxiv
第 1 章 调查研究与网络调查	1
1.1 定义与分类	1
1.1.1 调查和网络调查	2
1.1.2 网络调查的用户	4
1.1.3 网络调查的抽样	5
1.2 网络调查过程	7
1.3 网络调查的发展、应用和相关实践	10
1.3.1 网络调查与互联网和在线调查	11
1.3.2 网络调查的前驱和替代方案	12
1.3.3 现代网络调查的兴起	15
1.3.4 调查模式的背景	17
1.3.5 优势和局限性概述	19
1.3.5.1 成本	19

1.3.5.2	数据收集的速度	21
1.3.5.3	实施的便利性	23
1.3.5.4	问卷的计算机化	23
1.3.5.5	多媒体	24
1.3.5.6	时间和地理上的灵活性	25
1.3.5.7	自我管理	25
1.3.6	网络调查的应用	27
1.3.6.1	对高覆盖率的特定人群进行网络调查	27
1.3.6.2	客户满意度调查	28
1.3.6.3	普通人群调查	29
1.3.6.4	商业调查	31
1.3.6.5	在线样本库	32
1.3.6.6	在线社区调查	32
1.3.6.7	网站评估	33
1.3.6.8	娱乐性民意调查、问卷调查和小测验	34
1.3.7	与网络调查相关的实践	34
1.3.7.1	网络上的行政表格	35
1.3.7.2	互联网测试	35
1.3.7.3	心理学网络实验	36
第 2 章	预调查阶段	38
2.1	调查模式详述	38
2.1.1	选择网络调查模式	39
2.1.2	混合模式设置	40
2.1.2.1	混合模式调查、多模式联系和混合模式系统	40
2.1.2.2	混合模式调查的问题	41
2.1.2.3	混合模式下的网络调查	42
2.2	抽样	44
2.2.1	界定人群	46
2.2.2	基于概率的网络调查中的抽样设计原则	46
2.2.2.1	抽样框	47

2.2.2.2	抽样技术	51
2.2.2.3	样本量	52
2.2.3	非概率网络调查中的抽样设计	54
2.2.3.1	非基于名单的网络调查的招募	54
2.2.3.2	非概率情境下抽样设计的近似方法	58
2.2.4	概率和非概率调查中的统计推论	59
2.2.4.1	概率样本的统计推论	60
2.2.4.2	非概率样本的统计推论	61
2.2.4.3	决定概率—非概率抽样的困境	63
2.2.4.4	关于非概率网络调查的结论性评述	65
2.2.5	无覆盖问题	66
2.3	问卷准备	70
2.3.1	一般性问题	70
2.3.1.1	调查问题的类型	70
2.3.1.2	问题编制过程	72
2.3.1.3	问题和调查问卷设计原则	72
2.3.1.4	测量过程和测量误差	73
2.3.1.5	应答过程的认知方面	73
2.3.1.6	非实质性应答的具体内容	74
2.3.1.7	问题库	74
2.3.1.8	网络问卷的特殊性	75
2.3.2	问题类型	75
2.3.2.1	单一项目问题	78
2.3.2.2	多项目问题	89
2.3.2.3	其他问题类型	96
2.3.3	问卷结构、计算机化和布局	101
2.3.3.1	问卷结构	101
2.3.3.2	计算机化、互动性和动态性	106
2.3.3.3	视觉布局	111
2.3.4	确保调查的参与度	114
2.3.4.1	参与度、流程和用户体验	114
2.3.4.2	调查游戏化	116

2.3.4.3	问卷长度	117
2.3.4.4	调查满足	119
2.3.5	网络调查问卷测试	121
2.3.5.1	传统测试方法	122
2.3.5.2	网络问卷调查的具体测试方法	125
2.3.6	问卷编制、计算机化和测试的一体化	130
2.4	技术准备	133
2.4.1	采集和整理数据的准备工作	133
2.4.1.1	数据库的准备工作	134
2.4.1.2	受访者追踪和唯一标识符	135
2.4.1.3	电子邮件设置	137
2.4.1.4	样本管理	140
2.4.2	并行数据	141
2.4.2.1	联系信息并行数据	142
2.4.2.2	设备类并行数据	142
2.4.2.3	问卷导航并行数据	143
2.4.2.4	采集并行数据的软件	144
2.4.2.5	对并行数据的分析	145
2.4.2.6	并行数据的隐私和伦理规范	145
2.4.3	支持安全和隐私的准备	146
2.4.3.1	服务器安全	147
2.4.3.2	网络调查应答过程的安全	148
2.4.3.3	匿名性的技术问题	149
2.4.3.4	物理和电子访问数据文件的保证	150
2.4.3.5	访问审计	150
2.4.3.6	与法律和专业标准的关系	151
2.4.4	网络调查软件的技术预测试	151
2.5	无应答策略	152
2.5.1	网络调查无应答的一般性问题	153
2.5.1.1	无应答理论	153
2.5.1.2	网络调查中的无应答过程	154
2.5.1.3	影响网络调查无应答的因素	156

2.5.2	定义无应答、中断和完整性	159
2.5.2.1	单位无应答指标	159
2.5.2.2	中断	163
2.5.2.3	问题层面的项目无应答	164
2.5.2.4	单位层面的项目无应答	165
2.5.2.5	完整应答、部分应答与不可用应答单位	166
2.5.3	无应答率和无应答偏差之间的关系	168
2.5.4	网络调查的一般设计特征与无应答	172
2.5.4.1	调查发起方	172
2.5.4.2	主题突出性	174
2.5.4.3	问卷设计	175
2.5.5	网络调查中的激励措施	175
2.5.6	联系过程的作用：模式、时间和频率	178
2.5.6.1	联系方式	178
2.5.6.2	发出邀请的日期和时间	179
2.5.6.3	联系次数及其时间安排	180
2.5.6.4	调查实施的时间长度	181
2.5.7	邀请函的具体内容格式	182
2.5.7.1	邀请函的个性化	183
2.5.7.2	提供激励与请求语气	184
2.5.7.3	电子邮件邀请函的格式与视觉设计	185
2.5.8	无应答策略	187
2.6	综合管理	189
第 3 章	正式调查	194
3.1	招募	194
3.2	测量阶段	196
3.3	处理与监控	197
3.3.1	激活、软启动、全面启动和停用	197
3.3.2	提供帮助	199
3.3.3	实地调查工作的干预措施	200

3.3.4 实地调查阶段的监控	201
第 4 章 调查后阶段	204
4.1 数据准备	205
4.1.1 确定应答状态	205
4.1.1.1 样本单位的合格性、完整性状态和质量状况	205
4.1.1.2 缺失值和非实质性应答	207
4.1.1.3 应答状态报告	208
4.1.2 验证和编辑	208
4.1.3 插补	210
4.1.4 加权	212
4.1.5 编码和重新编码	214
4.2 初步结果	216
4.2.1 报告	216
4.2.2 发布	217
4.3 数据导出与文档化	218
4.3.1 数据导出	218
4.3.2 统计披露	219
4.3.3 文档化	219
4.3.4 数据归档	220
第 5 章 网络调查实施的选定主题	222
5.1 智能手机、平板电脑及其他设备	223
5.1.1 多设备介绍	223
5.1.2 处理移动设备的调查策略	226
5.1.3 设备效应	228
5.1.4 为智能手机设计网络问卷	232
5.1.5 基于浏览器的移动调查 vs 移动调查应用程序	234
5.1.6 多设备网络调查软件和移动调查应用程序	238
5.1.7 挑战和结论	239
5.2 在线样本库	241

5.2.1	术语与分类	241
5.2.2	招募、抽样和选择	242
5.2.2.1	非概率在线样本库的招募方法	243
5.2.2.2	概率在线样本库的招募方法	243
5.2.2.3	在线样本库成员的抽样与选择	244
5.2.3	维护在线样本库	245
5.2.4	在线样本库研究设计	246
5.2.5	在线样本库的数据质量	247
5.2.5.1	在线样本库数据质量和结果的比较	247
5.2.5.2	加权对在线样本库数据质量的影响	249
5.2.5.3	在线样本库研究中的其他数据质量问题	250
5.2.5.4	在线样本库数据质量的结论	251
5.2.6	购买在线样本库研究	252
5.2.7	在线样本库的未来	252
5.3	网络调查软件	253
5.3.1	网络调查过程的集成信息与通信技术 (ICT) 支持	255
5.3.2	网络调查软件市场	258
5.3.3	网络调查软件的特征	260
5.3.4	关于两个选定特征的讨论	263
5.3.4.1	界面速度	264
5.3.4.2	图形用户界面的关键特征	265
5.3.5	寻找合适的工具	265
5.3.6	网络调查软件的趋势	267
第 6 章	网络调查的广义方法论	270
6.1	更广泛的方法学背景	270
6.1.1	网络调查模式的修订	270
6.1.2	总调查误差和调查数据质量	272
6.1.3	对网络调查模式效应的评估	274
6.1.4	成本—误差优化	276
6.1.5	核心网络调查过程之外	277

6.1.5.1	初步研究活动的步骤	277
6.1.5.2	高级分析、处理与价值转化阶段	278
6.1.5.3	混合模式	279
6.1.5.4	混合方法	280
6.1.6	迈向交互式调查设计	281
6.2	项目管理框架内的网络调查	284
6.3	网络调查专业	286
6.3.1	相关学科	287
6.3.2	法律问题	288
6.3.3	一般伦理考量	289
6.3.4	正式认证、准则和标准	290
6.3.5	非正式指南、建议和其他资料	292
6.4	网络调查参考书目	293
第 7 章	网络调查的未来	297
7.1	一般技术发展	297
7.2	网络调查软件	299
7.3	方法学	300
7.4	更广泛的商业和社会问题	302
第 8 章	结 论	305
	参考书目	308
	作者索引	343
	主题索引	353

序 言

关于《网络调查方法》一书

20 世纪 90 年代初网络刚刚兴起后不久，网络调查便出现了，如今已成为数据采集的主要方式。借助网络，研究者发起一次网络调查，如同撰写并发送电子邮件一样简单。相应地，基本的计算机素养不仅是开展网络调查的必备技能，也是充分技能。

然而，开展网络调查会引发许多技术层面和概念层面的问题：网络调查是否适合我的研究问题？我如何从一般人群中精准招募受访者？如果参与调查的人来自在线社交媒体，我是否可以对获得的结果进行推广？可期待的样本规模究竟有多大？哪种方式可以维系合作意愿？一份网络问卷需要多长时间？是否应该强制受访者回答每一个问题？我该如何选择正确的网络调查软件？如何在智能手机上显示问卷以确保受访者阅读流畅？

诸如此类的问题有成千上百个，对这些问题的回答需要多方面的知识，既包括不断更新的技术细节，也涉及多重方法学知识，等等。很多时候，没有简单的答案。因为在大多数情境下，常识不足以给出可靠指引。一本“菜谱式”的方法学操作手册亦难以穷尽所有问题。真正需要的是，网络调查相关议题的系统化知识以及对研究情境的深层理解，以便在方法层面提供指引和依据，并创造性地生成契合特定网络调查情境的解决方案。本书的写作目标，恰在于提供这样一种知识理解框架。

本书是一部专论网络调查方法的系统性著作，内容覆盖全面且基于调查实践。该领域的其他出版物通常只涉及网络调查方法的某些部分，或将其置于更宏大方法论图景中作为“特例”一笔带过。与这些出版物相比，本书特点鲜明。

本书开篇先对网络调查的定义与实施流程进行介绍（第 1.1 节和第 1.2 节）。第 1.3 节还进一步阐释了计算机辅助调查信息收集（CASIC）的发展背景和历史演

变，以及网络调查实际实施的实际案例。接下来是三个核心章节，聚焦于网络调查过程的关键。第2章系统梳理了网络调查的前期准备工作，包括模式精细化设计、抽样方案和问卷准备、技术问题、无应答策略和综合管理。第3章围绕实际的数据收集展开讨论，包括被试招募、测量以及实地工作处理与监控支持等内容。第4章回顾了数据收集后的事宜：涉及数据整理准备、初步报告和数据导出。上述三章连同绪论部分共同构建起网络调查方法的基础概念框架，清晰呈现了该领域的重要结构、核心流程和专业术语。我们认为，这些章节作为本书的主要内容，具有较强的稳定性，不会因技术的快速迭代而迅速过时。与前文内容不同，第5章探讨的内容，在很大程度上依赖于技术的持续进步，属于网络调查方法论中更新较快的部分。在此背景下，第5章分析了网络调查方法中变化迅速却至关重要的三个方面，即设备、在线样本库调查和网络调查软件。第6章涉及更广泛的方法论、管理和专业框架，第7章则对未来趋势进行了展望。第8章为本书的结论。

本书配套的 WebSM 网站 (www.websm.org) 提供了丰富的补充材料。网站通过示例（例如样本量计算）和插图（例如对问题类型的交互式展示）进一步阐述了书中的概念，还包含额外的文献综述，术语表 (Glossary) 以及若干手册式的实用规则、建议和要点。

在本书撰写过程中，我们始终力求对每个讨论的问题做到三重呈现：既清晰阐释关键概念、梳理最新文献综述，又深入反思相关领域的最新研究进展，并针对性地提出实践建议。这一工作往往极具挑战性，尤其当实证证据薄弱、模糊或相互矛盾时。因此，本书整体上对网络调查方法的重要方面进行了系统性评估并提供指导，可以作为网络调查实践中方法论问题的参考手册。

我们努力在学术写作风格与应用实践者的风格之间寻求一种微妙但有时近乎难以实现的平衡，这得益于我们三位作者分别来自商业领域和学术界的多元背景。除了对网络调查过程进行概念化梳理，我们也涉及所有关键的实际问题，如从选择问题格式到发送电子邮件邀请的技术设置，等等。需要说明的是，编写一本手册式应用指南并非本书的目标，但相关的应用指导正逐步在 WebSM 网站上完善。

本书假定读者具备方法论或统计学的基本知识，同时也会对相关概念进行必要介绍。对于那些预设读者已经具备更多统计学和方法论知识或者熟悉网络调查实践的章节，WebSM 网站提供了相应的补充材料。因此，本书既适合在方法论或统计学方面缺乏正规教育但拥有丰富网络调查经验的读者，也适合接受过正规教育但缺乏相应实践经验的读者。在此背景下，我们将使用“新手”一词来指代这两类读者，并引导他们参考相应的补充材料。

本书可作为方法论教育和研究项目的教材供学生使用，同时也能为学术研究者了解特定主题的最新进展提供参考。除学术用途之外，实践者是本书最广泛的受众群体，包括市场研究和官方统计领域的专业人士，以及人力资源管理、健康研究、心理学、社会学等众多（社会科学）领域中使用网络调查的用户。对于那些缺乏充分方法论教育或经验，却需要开展严谨的网络调查的“自助式”（DIY）用户，本书同样适用。

对于涉及现代调查数据的任何研究，了解网络调查方法都愈发重要。我们预计，网络调查在未来将进一步扩展，延伸到新的领域和应用场景。我们还预计会有更多与技术进步相关的创新和更迭，特别是在新设备和整个调查研究过程的整合方面。我们相信，在不久的将来，几乎所有的调查都将至少包含网络调查方法框架的部分元素。上述因素突出了本书所探讨主题的重要性。

关于“网络调查方法”网站

截至 2015 年，网络调查研究领域已历经约 20 年的发展历程。与其他调查数据收集方式（如电话调查）相比，网络调查相关的研究成果分散在更广泛的资源载体、学科领域和出版物中。为整合网络调查方法的各类资源，瓦斯亚·维霍瓦尔（Vasja Vehovar）和卡佳·洛扎尔·曼弗雷达（Katja Lozar Manfreda）于 1998 年发起了网络调查方法（WebSM）网站，并持续维护至今。2003—2005 年，该项目获得欧盟框架项目（WebSM, HPSE CT 2002 50031）的支持。2009 年，二人凭借 WebSM 斩获美国民意研究协会（AAPOR）Warren J. Mitofsky 创新奖。自 2011 年以来，该网站的部分工作得到了 WebDataNet（Cost Action IS1004）的支持。

WebSM 的核心目标是提供关于网络调查方法以及新技术对调查数据收集影响的专业信息。该网站收集并发布相关的新闻动态、学术和专业活动信息、网络调查软件以及最重要的网络调查方法和议题的文献详情。截至 2014 年初，WebSM 数据库已收录 8 323 位作者的 6 890 条文献条目。

这个内容全面的网站是撰写本书的重要动力之一，如今，更成为本书的配套网站。网站根据相应章节对文献条目进行了结构化梳理，还包含本书各章节的各类补充材料。我们诚挚邀请读者关注网站的更新内容，访问地址 <http://www.websm.org>。

致 谢

本书初稿的撰写得到了多方支持。斯洛文尼亚研究机构为卢布尔雅那大学研究团队提供了研究基金（J5-5538、J5-4159、J5-4177 和 J5-6817）。此外，四位博士研究生围绕网络调查方法的各部分开展的研究工作，也为我们提供了大量的资金支持。我们还要感谢 WebDataNet Cost Action IS1004，为我们创造了在伦敦和卢布尔雅那之间往返交流的机会，助力本书撰写过程中的讨论和协调准备工作。

除了资金支持，我们还要感谢那些帮助我们撰写本书的人。Sage 出版集团的凯蒂·梅茨勒（Katie Metzler）全程为我们答疑解惑，协助解决写作中的各类难题，并始终鼓励我们推进这项工作。我们的首位编辑帕特里克·布林德尔（Patrick Brindle）是这个项目的启动者，若非他的坚持，本书或许难以问世。我们坚信 Sage 是一家具有独特风格的出版社。

马里奥感谢谷歌内部“调查邮件列表”上富有启发性的讨论。讨论中提出的挑战性问题促使我们完善了书中的部分内容，并深入探究了一些假设的理论依据。

卡佳和瓦斯亚感谢卢布尔雅那大学社会科学学院社会信息学与方法学系的支持，他们在该系教授统计学和方法论课程。该系的在读及毕业博士研究生提供了很大帮助，特别是格雷戈尔·切霍文（Gregor Čehovin）和内伊茨·贝尔泽拉克（Nejc Berzelak），其贡献几乎等同于共同作者。还有安德烈·彼得罗维奇（Andraž Petrovčič）、安泽·森德尔巴赫（Anže Sendelbah）、安娜·斯拉夫奇（Ana Slavec）和维斯娜·多尔尼查尔（Vesna Dolničar）也给予了重要支持。许多其他同事也提供了富有价值的建议，特别是瓦伦蒂娜·赫列贝克（Valentina Hlebec）在问卷方面、雅罗斯拉夫·贝策（Jaroslav Berce）在项目管理方面以及伊雷娜·维帕夫茨·布瓦尔（Irena Vipavc Brvar）在档案方面的建议。

最后，卡佳和瓦斯亚还要感谢阿努什卡·弗尔利戈伊（Anuška Ferligoj）教授，她鼓励他们进入这一方法论领域，并开启职业生涯。弗尔利戈伊教授不仅致力于追

求科学卓越，更以远见卓识在卢布尔雅那大学社会科学学院组建了方法论研究团队，该团队如今已获得全球学界的认可。

我们还收到了瓦利康的泽内尔·巴塔格尔（Zenel Batagelj）的宝贵反馈，并且听取了蒂姆·马瑟（Tim Macer）关于移动调查的建议。我们还需要感谢我们的家人。

马里奥感谢他的妻子安娜·维亚尔（Ana Villar），感谢她就网络调查方法和问题进行的多讨论，并分享了她在欧洲社会调查（European Social Survey, ESS）开展网络调查实验的经验。卡佳感谢她的丈夫马尔科（Marko）以及双方父母，在她专注工作和出差期间，他们不仅提供了帮助，有时甚至承担起照顾孩子的责任。瓦斯亚感谢妻子达林卡（Darinka）的温暖支持，以及来自阿廖沙（Aljoša）、米哈（Miha）和娜塔莎（Natasha）的宝贵反馈。

当然，尽管得到了如此多的支持，但书中难免存在不足，任何疏漏或不足之处都应归咎于作者本人。

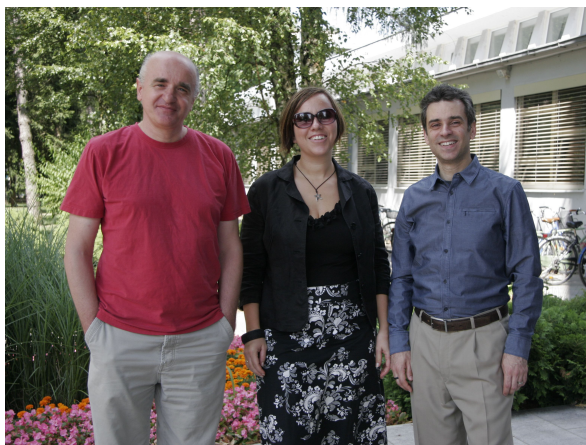
马里奥·卡列加罗（Mario Callegaro）

卡佳·洛扎尔·曼弗雷达（Katja Lozar Manfreda）

瓦斯亚·维霍瓦尔（Vasja Vehovar）

伦敦和卢布尔雅那，2015 年 1 月

作者简介



马里奥·卡列加罗（Mario Callegaro）是高级调查研究科学家，目前就职于谷歌伦敦分部，从事网络调查、客户满意度测量以及调查设计、问卷设计、抽样和报告方面的咨询工作。

此前，马里奥·卡列加罗在 Gfk（growth from knowledge 的缩写）——知识网络公司就任调查研究科学家，该公司负责管理美国一个基于概率的在线调查小组。马里奥拥有内布拉斯加大学林肯分校授予的调查研究与方法论博士学位。他是《调查研究方法》杂志的副主编，《国际市场研究》杂志的编辑委员会成员。马里奥在网络调查、在线调查小组、电话和手机调查、问题措辞、民意调查和出口民调、历史事件日志、纵向调查和调查质量等领域发表了大量文章。他最近（2014 年 5 月）出版了一本书，书名为《在线样本库研究：数据质量视角》（*Online Panel Research: A Data Quality Perspective*）。

瓦斯亚·维霍瓦尔（Vasja Vehovar）和卡佳·洛扎尔·曼弗雷达（Katja Lozar Manfreda）是斯洛文尼亚卢布尔雅那大学社会科学学院的统计学和方法学教授，他

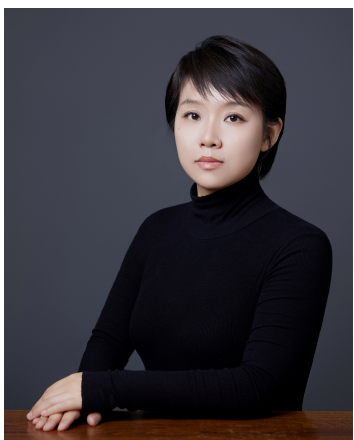
们也在该大学获得博士学位。瓦斯亚·维霍瓦尔和卡佳·洛扎尔·曼弗雷达是网络调查研究的先驱，于1996年开始进行网络调查实验。他们100多篇关于网络调查的学术论文、专著章节和其他成果，其中一些具有重要的国际影响力。有一项网络调查方法学领域的元分析文章，被引用次数最多，并成为领域内的经典（Lozar Manfreda et al., 2008）。

自1998年以来，他们一直负责管理最具权威性的网络调查方法参考数据库WebSM，并在2009年获得了美国民意研究协会（AAPOR）Warren J. Mitofsky 创新奖。他们的研究主题和相关出版物聚焦于网络调查的方法学问题：问卷设计、无应答、模式效应、以自我为中心的社会网络在线测量、测量误差和调查成本的优化、混合模式调查和数据质量。他们还参与了网络调查软件工具IKA的开发。

卡佳的博士论文是网络调查方法学领域的第一篇论文（2001年）（由Mick Couper 共同指导）。她曾任国际社会学协会逻辑与方法研究委员会RC33主席（2010—2014），也是成本行动WebDataNet核心小组成员和工作组负责人（2011—2015）。她是BMS（《社会学方法论通报》）科学委员会的成员和ESRA《调查研究方法》的副主编。她在GESIS暑期学校教授网络调查方法课程。

瓦斯亚在网络调查方法学领域指导了四名博士生。此外，他的研究领域还涵盖了广泛的互联网研究（如更好的互联网项目、斯洛文尼亚互联网研究项目、数字鸿沟研究），调查方法（例如抽样、调查无应答、调查替代、欧洲社会调查的权重计算）和特定统计方法（例如时距）等。

译者简介



郭素然，毕业于北京师范大学心理学院，获理学博士学位，国际关系学院副教授，硕士生导师。长期讲授“心理学研究方法”研究生课程，在多项研究中采用网络问卷调查和网络实验的方法。
邮箱：guosuran@163.com

缩略词列表

AAPOR	American Association for Public Opinion Research 美国民意研究协会
ABS	address-based sample 基于地址的抽样
ACASI	Audio Computer-Assisted Self-Interview/Interviewing 音频计算机辅助自填式面访
AoIR	Association of Internet Researchers 互联网研究协会
ARF	Advertising Research Foundation 广告研究基金会
BBS	Bulletin Board System 电子公告板系统
BCC	Blind Carbon Copy, Blind Copy Circulated 密送
CAI	Computer-Assisted Interviewing 计算机辅助调查
CAPI	Computer-Assisted Personal Interview/Interviewing 计算机辅助个人面访
CASI	Computer-Assisted Self-Interview/Interviewing 计算机辅助自填式面访
CASIC	Computer-Assisted Survey Information Collection 计算机辅助调查信息收集
CASM	Cognitive Aspects of Survey Methodology 调查方法学认知视角

CASRO	The Council of American Survey Research Organizations 美国调查研究机构协会
CATI	Computer-Assisted Telephone Interview/Interviewing 计算机辅助电话访问
CC	Carbon Copy, Copy Circulated 抄送
CESSDA	Consortium of European Social Science Data Archives 欧洲社会科学数据档案联盟
CRM	Customer Relationship Management 客户关系管理
CSAQ	Computerized Self-Administered Questionnaire 计算机辅助自填式问卷
CSS	Cascading Style Sheets 层叠样式表
CSV	Comma Separated Values 逗号分隔值
CTR	Click-Through Rates 点击率
DBM	Disk-By-Mail 邮寄软盘
DDI	Data Documentation Initiative 数据文档倡议
DIY	Do-It-Yourself 自助式
DOI	Digital Object Identifier 数字对象标识符
EFM	Enterprise Feedback Management 企业反馈管理
ELIPSS	Étude Longitudinal par Internet Pour les Sciences Sociales 社会科学纵向互联网研究（法国）
EMS	Electronic Mail Survey 电子邮件调查

ESOMAR	European Society for Opinion and Marketing Research 欧洲民意与市场研究协会
ESRA	European Survey Research Association 欧洲调查研究协会
F2F	Face-to-Face (Interview) 面对面（访谈）
FAQ	Frequently Asked Question 常见问题
FFRISP	Face-to-Face Recruited Internet Survey Platform 面对面 / 互联网调查平台
FOQ	Foundation Of Quality Initiative 质量基础倡议
GOR	General Online Research 通用在线研究
GPS	Global Positioning System 全球定位系统
GRS	Graphic Rating Scale 图形评定量表
GUI	Graphical User Interface 图形用户界面
HCI	Human-Computer Interaction 人机交互
HRM	Human Resource Management 人力资源管理
HTML	Hyper-Text Markup Language 超文本标记语言
IBR	Introduction Breakoff Rate 介绍性中断率
ICT	Information and Communication Technology 信息与通信技术
ILB	Item Level Breakoff 项目层面的中断

INR	Item Nonresponse Rate 项目无应答率
iOS	a mobile operating system developed and distributed by Apple Inc. 苹果移动操作系统
IP	Internet Protocol 互联网协议
IRB	Institutional Review Board 机构审查委员会
ISO	International Organization for Standardization 国际标准化组织
ITV	Interactive TV 交互式电视
IVR	Interactive Voice Recognition 交互式语音识别
KPI	Key Performance Indicator 关键绩效指标
LISS	Longitudinal Internet Studies for the Social Sciences 社会科学纵向互联网研究
MAR	Missing At Random 随机缺失
MCAR	Missing Completely At Random 完全随机缺失
MRIA	Marketing Research and Intelligence Association 市场营销研究与情报协会
MSE	Mean Squared Error 均方误差
NMAR	Not Missing At Random 非随机缺失
NOPVO	National Dutch Online Panel Comparison Study 荷兰国家在线样本库比较研究
OECD	Organization for Economic Cooperation and Development 经济合作与发展组织

P&P	Paper and Pencil questionnaire 纸笔问卷
PAPI	Paper and Pencil Interview/Interviewing 纸笔访谈
PSW	Propensity Score Weighting 倾向性评分加权
QBR	Questionnaire Breakoff Rate 问卷中断率
QR	Quick Response Code 二维码
QUAID	Question Understanding Aid 问卷理解辅助工具
RDD	Random Digit Dialing 随机数字拨号
RR	Response Rate 应答率
SaaS	Software as a Service 软件即服务
SAS	Statistical Analysis System 统计分析系统
SD	Standard Deviation 标准差
SE	Standard Error 标准误
SMS	Short Message Service 短信
SNS	Social Networking Site 社交网站
SPSS	Software Package used for Statistical Analysis 统计分析软件包
SQP	Survey Quality Prediction 调查质量预测

SRS	Simple Random Sample 简单随机抽样
SSL	Secure Sockets Layer 安全套接字层
TBR	Total Breakoff Rate 总中断率
TCP/IP	Transmission Control Protocol/Internet Protocol 传输控制协议 / 网际协议
TDE	Touchtone Data Entry 按键式数据输入
TQM	Total Quality Management 全面质量管理
TSE	Total Survey Error 总调查误差
UBR	Unit-level Breakoff Rate 单位层面的中断率
UCR	Unit-level Completeness Rate 单位层面的完整率
UIN	Unit-level Item Nonresponse 单位层面项目无应答率
UIT	Unproctored Internet Testing 无人监考的互联网测试
UMR	Unit-level Missing-data Rate 单位层面缺失数据率
URL	Uniform Resource Locator 统一资源定位符
UX	User Experience 用户体验
VAS	Visual Analogue Scale 视觉模拟量表
VCASI	Video Computer-Assisted Self-Interview/Interviewing 视频计算机辅助自填式面访

W3C	World Wide Web Consortium 万维网联盟
WAP	Wireless Application 无线应用
WCSC	Working Conditions in Slovenian Science 斯洛文尼亚科学家调查
WWW	World Wide Web 万维网
WYSIWYG	What You See Is What You Get 所见即所得
XML	Extensible Markup Language Protocol 可扩展标记语言

调查研究与网络调查

本章首先阐述网络调查的核心定义、基础概念及发展背景（第 1.1 节），为后续探讨奠定理论框架；接着解析网络调查的实施流程与结构体系（第 1.2 节），并以此为基础概述本书核心章节（第 2、3、4 章）的核心内容；第 1.3 节深入剖析网络调查的关键技术要点，追溯其历史演进脉络与实践应用场景。

1.1 定义与分类

网络调查隶属调查法范畴，虽具有一定的特殊性，但仍遵循一般调查方法论的原则。由此，我们时常参考并推荐读者参考 Groves、Fowler、Couper、Lepkowski、Singer 和 Tourangeau（2009）所著教科书《一般调查方法》，以系统掌握调查法的通用知识。本书的目的是构建独立的网络调查方法论体系，所以除了网络调查的特定内容，本书还包括一般调查方法的基本内容。

第 1.1.1 节阐述一般调查法的背景和调查模式的相关概念，第 1.1.2 节界定了网络调查的具体内涵及应用场景，第 1.1.3 节基于抽样特征，梳理了网络调查的主要类型。

1.1.1 调查和网络调查

调查 (*survey*) 是一种系统的数据收集方法, 通过标准化问卷向受访者提问, 目的是对一些目标人群 (*target population*), 即研究关注或感兴趣的人群进行定量分析 (Fowler, 2014; Groves, et al., 2009)。通常情况下, 调查数据仅从目标人群的样本中收集; 而从整个人群中收集数据, 被称为普查 (*census*) (例如, 特定组织的所有雇员)。调查可以是强制性的, 这种情况极为罕见 (例如某些国家的一些官方调查), 也可以是自愿性的, 本书只关注后者。

传统调查模式中, 访员通过电话或面对面 (F2F) 的访谈, 在纸上记录受访者的答案。自填式问卷则由受访者自行填写纸笔问卷 (P&P), 通常是通过邮寄的方式发放和回收问卷, 我们在下文中简称为邮寄调查 (*mail surveys*)。近几十年来, 信息和通信技术 (ICT) 的发展改变了这些传统调查模式 (*traditional survey modes*): 答案实现电子记录, 计算机技术的应用也带来了多方面改进, 同时催生了全新的调查方法。

随着互联网的兴起, 尤其是二十世纪九十年代中期现代网络浏览器的发展, 从每天填写的问卷数量以及开展的调查项目数量来看, 网络调查已经迅速发展成为主流的调查数据收集类型 (ESOMAR, 2013; Macer & Wilson, 2014)。

本书使用调查模式 (*survey model*) 这一术语来描述调查测量过程的具体特征。我们用测量 (*measurement*) 来表示数据收集 (*data collection*) 的核心阶段, 即受访者通过填写问卷提供真实信息。使用“测量”一词以避免与数据收集 (*data collection*) 混淆, 因为后者通常涉及人们对调查数据收集过程的宽泛理解。

网络调查模式 (*web survey model*) 是一种借助网络实施的自填式调查模式。相应地, 当使用网络调查模式时, 就指网络调查。在测量阶段, 当网络调查模式的应用产生特定结果时, 便是指网络调查模式效应 (*web survey mode effect*)。例如, 与访员主导的调查模式相比, 在网络调查模式下, 受访者更容易如实报告非社会赞许行为 (如赌博、吸烟)。调查模式效应 (*the survey mode effect*) 是一种特殊的测量误差 (*measurement error*), 体现为调查应答与真实值之间的差异。我们将在第 6.1.1 到 6.1.3 节进一步讨论网络调查模式效应的具体内容。

当然, 测量阶段只是调查实施的环节之一, 此外还包括抽样、招募被试和数据整理等其他流程。尽管这些环节都与调查模式密切相关, 对调查质量也非常重要, 但在一定程度上它们可视为相对独立的模块。例如, 当在测量阶段使用网络调查模式时, 可以使用多种抽样或招募方法。由于在网络调查模式中, 这些环节往往非常

特殊,本书将深入探讨,尤其关注以下问题和相关误差:未覆盖(*non-coverage*)(目标人群的某些部分未被纳入调查,如不使用互联网的人)、无应答(*nonresponse*)(已被纳入调查的对象没有回应,如拒绝回答)和自我选择(*self-selection*)(对招募被试的过程缺乏有效控制)。其他调查误差(*survey errors*)也可能出现,本书将用调查数据质量(*survey data quality*)泛指上述内容,第6.1.2节对此会详加论述。

本书将重点介绍基本网络调查模式(*basic web survey mode*)。我们将其定义为一种使用计算机化自填式问卷的调查模式,这些问卷存储于连接到互联网的特定计算机(即服务器)上,受访者通过网络浏览器访问这些问卷。受访者主要使用台式电脑或笔记本电脑,但近年来科技的发展导致智能手机、平板电脑和互联网电视的使用率上升。受访者阅读屏幕上直观显示的问题后,使用键盘、鼠标、触摸屏或其他一些手动电子设备(*manual electronic device*)(如指针、手写笔等)进行作答。虽然没有访员在场指导受访者,但这个过程仍具有互动性(*interactive*),如依据受访者的前期作答情况调整后续问题的呈现,或对未作答问题即时提示。受访者的答案通常逐页自动上传至研究者服务器的数据库。由于整个过程由服务器控制,因此被称之为基于服务器的(*server based*)调查模式。

需要说明的是,研究者(*researcher*)可以指个人,也可以是负责问卷设计和全程实施网络调查的组织。然而,在实践中,研究者的角色经常可以细分成以下几个子角色:(a)设定目标和设计调查(通常是由委托调查的客户完成);(b)资金筹措(通常由客户或者其他一些供资机构完成,但这些供资机构不干预调查设计);(c)赞助(除了资助机构,可以是任何其他正式支持调查的机构);(d)方法设计(原研究者可以进一步将调查分包给某研究机构,由该机构负责方法论设计并实施调查);(e)调查实施(可以委托专业网络调查供应商来执行)和(f)网络调查支持软件(多数网络调查依赖专业供应商提供的软件及网络托管服务,托管可单独作为一项服务)。尽管这些机构和组织对网络调查的实施非常重要,但由于我们重点关注方法层面,本书不展开讨论这些机构的具体职能。

随着语音识别和语音技术,以及视频、多媒体、动画(如虚拟采访者或主试)等各种技术的应用拓展,网络调查正变得日趋复杂。我们主要关注基本网络调查模式,这种模式目前是当代网络调查的主流实施方式。方便起见,除非特别说明,本书将基本网络调查模式简单表述为网络调查模式。在第5章、第6章和第7章的结论中将涉及这些拓展形式。

调查设计类型方面,仅进行一次、独立抽样调查的为横断面设计(*cross-sectional design*)。如果想观察时间维度上的变化,可以使用重复的横断面设计(*repeated*

cross-sectional design), 即在不同时间对新的独立样本实施相同的问卷调查。典型的例子是传统的民意调查, 它通常通过每个月的调查来追踪政治支持的趋势。此外, 纵向设计 (*longitudinal design*) 也可以用于观察变化, 即对同一批受访者进行连续、多次甚至定期 (如每月) 的调查; 这能够深入了解和精确观察与重要议题有关的变化, 如失业、收入动态或媒体消费等。纵向设计也适用于队列研究 (*cohort study*), 即追踪并反复调查 (例如每年一次, 每十年一次) 特定人群, 例如整个学校某一届的学生群体。

纵向设计与样本库设计 (*panel study*) 的关系密切, 有时甚至被当作同义词使用。但只在特定情况下使用“样本库”这一术语, 即当一组受访者作为重复抽样的框架, 用于开展独立调查时, 这些调查的主题可以相关或无关。后者也是样本库设计与纵向设计的主要区别, 在纵向设计中, 通常只关注一个具体问题。参与样本库调查的受访者需预先同意参与, 并提供社会人口学信息。从历史上看, 这种方法源于市场研究。在市场研究中, 有几千甚至几百万个家庭样本库为各种独立调查服务。如今, 在线样本库已经成为网络调查的核心工具, 相关内容在第 5.2 节中被讨论。

1.1.2 网络调查的用户

信息和通信技术的普及、互联网的广泛应用、相关成本的持续下降以及互联网速度的不断提升, 是推动网络调查广泛被应用的关键因素。然而, 不同的研究者群体的应用程度存在差异。

市场研究领域对网络调查的应用最为密集且技术先进, 这一趋势在商业领域具有代表性。在过去十年中, 77 个国家在定量研究中用于不同调查模式的资金 (ESOMAR, 2013) 一直在大幅增加, 其中网络调查的资金投入增长显著, 是增幅最大的调查模式类别。数据还显示, 面对面调查 (F2F) 大幅下降, 电话和邮寄调查略有下降。Greenbook¹、Macer 和 Wilson (2013) 也发现了类似的趋势, 即在市场研究中, 网络调查的资金占比已经超过其他所有调查模式资金的综合。见第 2 章的补充材料互联网统计 (*Internet statistics*) 中的详细趋势和数字, 网址: <http://websm.org/ch2>。

官方统计及政府部门对网络调查的应用相对滞后, 主要原因包括: 需获取具有代表性的、高质量普通人群样本, 对隐私保密问题的顾虑, 技术复杂性以及某些大型组织中长期固定的程序等。此外, 网络调查的方法论问题以及数据质量也存在待完善之处。在纵向调查中, 上述问题更有可能会出现, 如从传统模式到网络调

查的转变会影响纵向调查中与时间序列相关的数据。然而,许多国家的统计机构目前面临着严峻预算压力,这可能导致减少昂贵的面对面(F2F)调查和电话调查(Bethlehem, 2009b; Cobben & Bethlehem, 2013)。而网络调查可因成本低、快速联络到大量潜在受访者,成为颇具前景的替代方案。对于商业调查(*business surveys*)(即来自公司、组织和机构的调查,也称为机构调查, *establishment surveys*)来说尤其如此,在发达国家,企业普遍具有完整的互联网覆盖。

由于类似的原因,网络调查在学术研究中的应用相对较慢,一些重要的学术性社会调查(如美国社会综合调查、欧洲社会调查)虽然正在开展方法论研究,但是对引入网络调查模式仍持谨慎态度(Ainsaar, Lilleoja, Lumiste, & Roots, 2013; Villar, 2013)。

网络调查除了运用于专业的调查数据收集组织(商业、政府和学术部门),还对调查研究的普遍民主化作出了重要贡献。由于网络调查的成本较低,加之网络调查软件价格低廉且对用户友好,因此任何拥有互联网连接、计算机和基本 ICT 技能的人都可以创建并实施网络调查。网络调查不需要较多的调查研究方法知识,也不需要资金,这允许任何人以网络调查的形式提出问题和公布结果。我们称之为自助式研究(*do-it yourself*)。然而,自助式研究存在争议,由于这些研究人员通常没有或较少接受网络调查和一般调查的方法论教育/培训,缺乏调查的方法论知识,因此调查质量参差不齐。大量低质量的调查可能引发对人群的“过度调查”,对专业标准和网络调查行业造成恶劣影响。

自助式网络调查的占比(*share*)难以精确统计,可能占到受访者日常填写的各种网络调查问卷的大部分。即便排除许多无关紧要的民意调查、娱乐调查、小测试等,这一点也可能成立。自助式调查的实施者包括学生、博主、互联网社区的成员和许多其他有好奇心的人;此外,还有无数的小企业、组织、社区、协会和其他团体,他们经常专业调查机构体系之外自行进行网络调查。自助式网络调查的复杂程度不一,因此此类调查的发起者也是本书的主要受众之一。

商业、官方、学术和 DIY 等四大用户群体的需求存在显著差异,体现在网络调查类型、面临的方法论问题和所需的网络调查软件功能等方面。因此,网络调查的用户并非同质化群体,但本书将聚焦其共性特征展开论述。

1.1.3 网络调查的抽样

本部分在网络调查的类型和功能背景下,讨论其优点和局限性。对于某种目标

而言，一种特殊类型的网络调查可能提供高质量的结果，但对另一种目标而言，这种调查却没有科学价值。在这种情况下，除用户群体差异外，基于抽样特征的分类是理解网络调查背景的另一个重要方面。以下将详细介绍两个关键抽样特征。

首先，需要明确概率样本的定义，进而界定概率调查。概率调查中，目标人群的每个单元（即个体）被选入样本的概率已知且为正数（非零）。概率调查根据概率，使用随机程序（*randomized procedures*），本质上类似抽奖，选取样本单元。而非概率（*non-probability*）样本调查包括所有其他不遵循概率样本标准的调查。这种区分适用于所有调查模式，并非网络调查特有。

统计推论（*statistical inference*）是指能够将结果从样本推广到目标人群的过程，仅通过概率调查才能实现。因此概率调查有时也被称为科学抽样调查。尽管概率样本具有统计学推断的优势，但由于速度、成本和便利性等多种实际因素，非概率样本的网络调查也被广泛应用。

其次，需要区分基于名单（*list-based*）和非基于名单（*non-list based*）的网络调查，这一区分决定了许多其他的设计特征。如果事先掌握关于目标人群的完整名单，就可以认为是基于名单的调查，否则为非基于名单的调查。基于名单的网络调查可以是概率性的，也可以是非概率性的，表 1.1 提供了具体事例。

因此，基于名单的概率网络调查需满足两个条件：名单能充分覆盖目标人群，且使用概率方法进行抽样选取受访者或被试。例如，对互联网普及率高的特定人群进行网络调查，并且以其成员名单作为抽样框实施概率抽样（如学生、雇员、抄写员 / 客户、在线团体的注册会员等）。又如，使用传统模式（邮寄、面对面、电话）招募被试（单元）进入样本，开展基于名单的普通人群概率网络调查。此外，在线样本库调查也会依据概率进行样本选择。

当个体纳入样本的概率不详或已知为零时，则为基于名单的非概率网络调查，如目标人群的重要部分缺失。通过非概率的方式招募被试开展调查，如在网上公开邀请而获得的电子邮件地址（以下简称电子邮件）列表。

表 1.1 基于主要抽样特征的网络调查实例

	概率抽样	非概率抽样
基于名单的调查	对特定互联网人群的概率调查	目标人群名单不全的网络调查
	基于概率的普通人群网络调查	基于自填式名单的网络调查
	基于概率的在线样本库	非概率在线样本库
非基于名单的调查	概率网络拦截调查	无限制的网络调查（自选择）

当样本单位的名单事先不可知时,则为非基于名单的网络调查,通常这些也是非概率网络调查。在无限制网络调查(*unrestricted web surveys*)(也被称为自我选择或开放性网络调查, *self-selected or open access*)中存在各种典型案例,任何看到邀请链接(例如某些在线论坛上的链接)的人都可以参与,且不受研究者限制。当然,是否被纳入样本由受访者自主决定——通常是通过某种未知的机制,而研究者无法控制样本选择机制。常见的“每日一题”式在线调查多属此类,主要用于娱乐和简单观察。

非基于名单的概率式网络调查比较少见。如概率网络拦截调查(*web intercept survey*),即通过弹出窗口随机选择某个网站的访问者,邀请其参与网络调查。当然,这需要符合一些特定条件,方可视为基于概率的调查。

综上,值得重申的是,网络调查模式与抽样调查并无必然关联。如表 1.1 所示,网络调查可以与任何类型的抽样相结合。对这一独立性的概念性理解并非易事,特别是在网络调查早期。这不仅与非概率样本有关,也与较低的数据质量有关。因此,在公布网络调查的结果时,应该明确说明抽样类型,如概率网络调查、非概率在线样本库、概率网络拦截调查、非限制性网络调查,等等。网络调查的概念仅仅意味着使用了网络调查模式进行测量,但对相应的抽样方法没有要求。然而抽样方法却是决定数据质量的主要因素之一。

1.2 网络调查过程

网络调查程序是实施网络调查的流程,包含按顺序进行的几个步骤。其中与方法论和网络调查模式相关的具体活动被称为核心网络调查过程(*core web survey process*)。核心网络调查过程与其他一般性研究活动应当区分开来,例如研究问题的初步概念化、数据分析和结果发表,因为后者通常独立于调查模式。尽管这些活动仍然是实证研究的重要组成部分,但对其详细阐述并不在本书范围之内。尤其是网络调查并非独立的项目,而是大型研究工作中的一部分时,情况更是如此(例如,客户满意度调查可能只是复杂市场研究的环节之一)。

图 1.1 概述了网络调查的流程,重点呈现核心网络调查过程。这非常重要,因为它决定了第 2、3 和 4 章的内容框架,这些章节详细阐述了所有阶段(*stages*)的活动,已在图 1.1 中用方框表示。为便于入门者理解,本章先对 12 个阶段进行简要说明,同时建议入门者结合第 1.3 节中的插图和实例加深理解。

网络调查始于初步研究活动 (*preliminary research activities*), 这一阶段在开始准备网络调查的具体方法之前进行, 属于一般性研究活动 (详见第 6.1.5 节), 独立于任何调查模式, 因此不纳入核心网络调查过程。初步研究活动通常包括明确研究目标、概念界定、前期文献调研、定性调查、专家咨询等与实质性任务相关的活动, 还涉及项目管理等初期工作, 如可能的分包、聘请外部机构运行网络调查等。

与初步研究活动的实质性、管理性和组织性不同, 核心网络调查过程重点关注网络调查相关的方法学内容。

核心网络调查过程分为三个步骤。其中最关键或最基础的步骤是收集受访者的答案, 与调查实施 (*fieldwork*) 有关。这一概念源于面对面调查。在面对面调查中, 收集答案意味着访员必须到现场去获得应答。相应地, 我们把这一步骤称为正式调查 (*fielding*), 包括招募 (*recruitment*) (即联系和邀请潜在的受访者)、测量 (即完成问卷收集数据) 以及相关的处理与监控 (*processing and monitoring*) 等阶段。

当然, 在调查实施之前需要进行充分的准备工作。因此, 实际上是从预调查 (*pre-fielding*) 的步骤开始核心网络调查过程活动。其中包括以下阶段: 模式选择 (*mode elaboration*), 确定测量模式和招募模式; 抽样 (*sampling*), 界定目标人群、设计抽样框架、确定样本量及抽样方法; 准备问卷 (*questionnaire preparation*), 包括问卷设计、计算机化处理和预测试。此外, 前期准备工作还包括与数据库和案例管理有关的技术准备 (*technical preparations*), 如隐私、安全和电子邮件设置等, 以及无应答策略 (*nonresponse strategy*), 包含联系计划及实现预期合作的途径等, 而综合管理 (*general management*) 包括整体管理、组织、沟通和一些其他活动。

调查后阶段 (*post-fielding*) 从数据准备阶段 (*data preparation*) 开始, 包括编辑、编码、汇总和加权等具体活动。此外, 还生成初步结果 (*preliminary results*), 包括报告并发布 (例如自动生成汇总统计, 然后在网上发布)。同样, 初步的数据导出与文档化 (*data exporting and documentation*) 也可以整合到网络调查软件中, 从而大大地支持了存档功能。

仅在网络调查软件通信技术支持的范围内, 核心的网络调查过程才包括调查后阶段。这个边界可能是相当模糊的, 因为不同网络调查软件对各阶段活动的支持程度不同。进一步而言, 先进的网络调查软件在很大程度上可以支持所有三个阶段, 包括调查后阶段。本书后续在介绍信息与通信技术 (ICT) 支持的潜力时会提到。信息与通信技术 (ICT) 不仅支撑和整合整个核心网络调查过程, 也界定了边界, 因此其作用至关重要。

另一方面, 核心网络调查过程不包含任何一般性研究活动。调查后阶段进行的

实质性分析 (*substantive analysis*), 独立于网络调查的具体流程, 需借助网络调查软件之外的软件来完成。这一步骤通常涉及 (a) 高级分析 (*advanced analysis*), 在专门软件中运行, 与报告撰写和结果解释密切相关; 进一步 (b) 处理数据 (*processing of the data*), 如纳入客户关系管理 (CRM) 数据库, 或与外部数据合并和匹配, 以及 (c) 价值转化 (*valorization*), 这个术语涵盖成果传播 (*dissemination*) (即从传统出版到数字渠道的传播) 和应用 (*exploitation*) (即对成果的各类使用与整合)。

图 1.1 清晰呈现了核心网络调查过程的各个步骤和阶段。相应地, 本书后续几

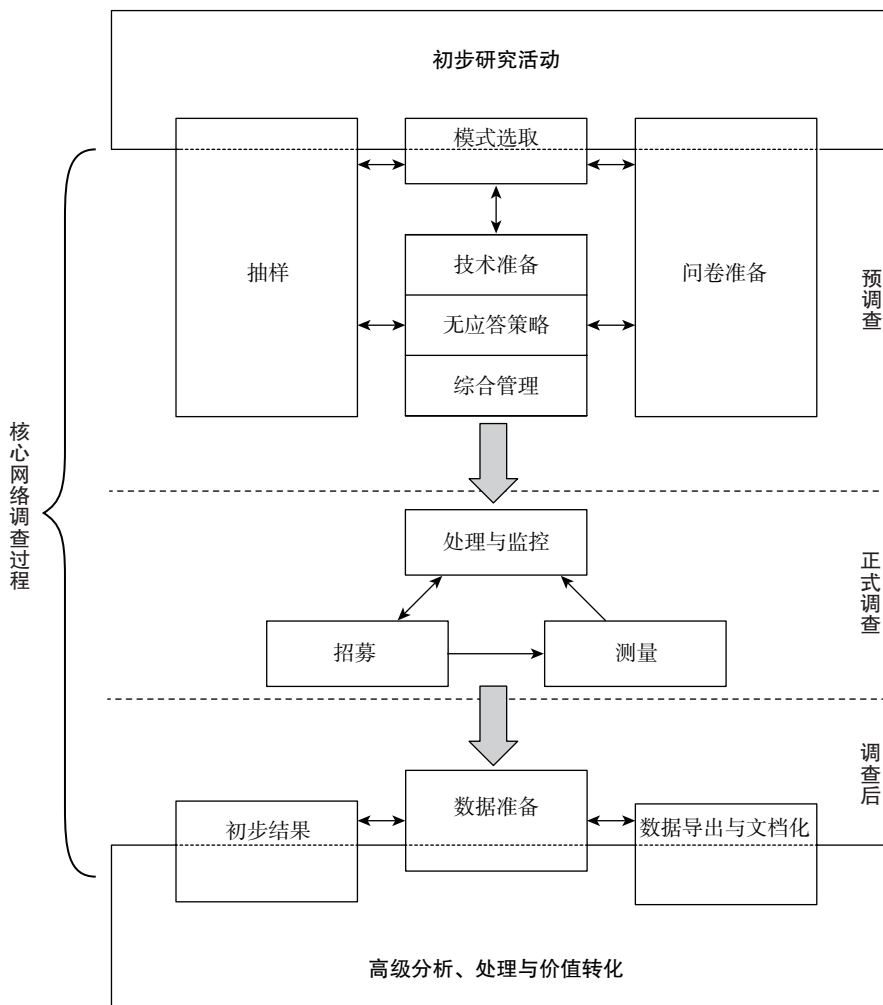


图 1.1 网络调查过程的步骤与阶段

个关键章节分别按顺序详述了这些步骤：预调查（第2章）、正式调查（第3章）和调查后阶段（第4章）。图1.1中的相关箭头标注了基本流程，以及各阶段之间存在的诸多相互作用。

请注意，这些步骤和相应的阶段并非总能截然分开。预调查阶段可能与初步研究活动重叠，因为在最初提出研究设想或在项目的早期构思时，可能已经做出专门针对网络调查的设计（如选择特定模式、提供现有样本或问卷初稿）。同样，预调查阶段的活动可能与正式调查交叠。例如，准备问卷时已经进行了实地测试，或者部分正式调查提前进行以便于进行修正。正式调查和调查后的工作也可能发生交叠。例如，在正式调查期间急需初步报告。有时也很难将调查后阶段的一些数据准备活动或文档管理工作与后续的数据分析、处理和价值转化区分开，尤其是在各种网络调查软件对相关工作的支持方式不同的情况下。

接下来的三章将介绍核心网络调查过程的简化流程，如图1.1所示，但不包括初步研究活动的步骤和高级统计分析、数据处理和价值转化等步骤。我们只介绍独立、横断面且独立开展的网络调查项目情境。这意味着假设只有一份问卷和一个测量（即完成问卷）环节，没有重复、递归、迭代、循环或其他在现实研究情境中经常出现的复杂情况。我们将在第6章一般方法论中进一步阐述扩展和复杂情况。

1.3 网络调查的发展、应用和相关实践

本节进一步阐述网络调查。首先区分与网络调查易混淆的概念，这些概念经常在类似的背景下使用，如互联网调查、电子邮件调查、在线调查、数字调查（第1.3.1节）。接下来，追溯网络调查的历史演进，从早期形式（第1.3.2节）到现代网络调查模式（第1.3.3节）。随后，从管理类型（访员管理与受访者自填式）和计算机化程度两个维度对调查模式进行分类（1.3.4节）；本节概述了网络调查的实用性，这对网络调查方法的入门者尤其有益。然后，概述网络调查的优势和局限（第1.3.5节），并举例说明不同研究领域网络调查的运用（第1.3.6节）。最后，本节总结了与网络调查相似的领域（第1.3.7节），如行政表格（*administrative forms*）、互联网测试和网络心理实验。这三个领域与网络调查存在诸多共同之处，但在网络调查的一般文献中却较少被纳入讨论。

1.3.1 网络调查与互联网和在线调查

网络 (*web*) 更确切地说是万维网 (*WWW*), 在经过十年的发展后于 1990 年正式推出。这比互联网 (*internet*) 的诞生晚了近三十年。互联网是遵循标准化通信协议 (*TCP/IP*) 的计算机电子网络。网络建构模块是超文本标记语言 (*HTML*), 承担着呈现与组织网络内容的功能。借助互联网, 用户能够通过特定软件 (即浏览器) 远程访问 *HTML* 格式网页中的相应内容。目前, 万维网联盟 (*W3C*) 推荐使用第五次修订版, 即 *TML5*。该版本不仅实现了跨操作系统的进一步标准化, 还增强了对互动性、图形及视频的支持。

某个网页 (*web page*) 对应的 *HTML* 内容拥有唯一的网络地址 (即 *URL*, 统一资源定位器), 并储存在一台连接到互联网的计算机 (即服务器) 上。单一网页的概念有时会与网站 (*website*) 产生混淆。通常来说, 网站有自己的域名 (如 *websm.org*) 或子域名 (如 *book.websm.org*), 且包含大量网页。

网络表单 (*web form*) 是网页的一种特殊子类型, 如今更倾向于将其视为一个组件。网络表单允许用户被动访问网页内容, 还支持用户在同一页面上输入一些数据并将内容传输到服务器上。在网络调查情境中, 网络问卷 (*web questionnaire*) 本质上是一种特殊的网络表单, 由用户 (即受访者) 填写, 输入内容 (即调查答案) 后传输到研究者的服务器。因此, 网络表单是网络调查数据收集的基础。在这种情况下, 标准 *HTML* 元素被定义为 *W3C* (制定网络标准的主要国际组织) 技术规范 2 中的 *HTML* 控件²。标准 *HTML* 元素在网络上广泛应用 (例如, 预订流程、行政表格、在线购物、用户注册等方面), 同时也构成了网络调查问卷的基本组件, 如单选按钮、复选框、下拉菜单和文本输入。

互联网调查 (*internet survey*) 的概念比网络调查更广泛。与之类似的术语包括通过互联网的调查 (*surveys via the Internet*) 或互联网媒介调查 (*Internet-mediated surveys*)。这些术语通常缺乏精确界定, 但核心都是指以互联网为媒介, 将受访者的回答传递 (传送) 给研究者的调查。值得一提的是, 依托互联网的信息传递并非必须依赖网络 (即万维网), 例如电子邮件作为互联网支持的另一项服务, 也可实现这一功能。

在线调查 (*online survey*) 一词的定义也通常不够清晰。尽管在线调查与互联网调查常被混为一谈, 但一般而言其涵盖范围比互联网调查更广泛。在此, 本节将通过在线调查的两个基本决定因素, 更精确地阐释二者之间的区别。第一个决定因素与计算机化的调查问卷有关, 需要有电子网络支持计算机之间的数据交换 (即受

访者的设备和研究者的服务器之间)。除了互联网,其他电信网络也可用于在线调查,如固定电话网络(如电话调查)、移动电话网络(如短信调查)和部分组织内部的其他各类计算机网络。第二个决定因素涉及计算机化问卷的自动传递:受访者不需要进行任何额外的操作(只需按一个按钮)就可以将回答发送给研究人员。电子邮件调查中,问卷是以文本文件的形式通过电子邮件接收的。受访者完成问卷后,还需保存问卷,并通过电子邮件上传附件回传给研究人员,因此超出了“不需额外操作的范畴”,不属于在线调查。

需要说明的是,数字调查(*digital surveys*)和电子调查(*electronic surveys*)这两个术语有时也会出现在网络调查的相关语境中,但由于其含义相当模糊,本书并未采用。数字调查或电子调查包括文本处理器中普通的计算机化问卷,这类问卷甚至可以与任何电子网络隔离。

如今,网络调查在互联网调查中占绝大比例(在在线调查中也占据大部分份额)。互联网调查中的“非网络”调查主要是电子邮件调查。下文将对电子邮件调查进行探讨,其背景是网络调查的早期雏形,以及电子邮件调查(EMS)和计算机辅助的样本库研究或电话样本库调查。

网络研究、互联网研究或在线研究,以及数字或电子研究,其涵盖范围比网络、互联网或在线调查宽泛得多。除了调查数据的收集,它们还包括对许多其他定量和定性研究技术的扩展应用。

1.3.2 网络调查的前驱和替代方案

本节回顾了网络调查的历史形式,特别关注了两种在当下有所复兴的替代方法:电子邮件调查以及作为应用程序在受访者计算机上运行的客户端问卷。

20世纪80年代中期出现的电子邮件调查(*electronic mail survey*, EMS)是首批借助电子网络开展的在线调查。在电子邮件调查中,计算机化的电子问卷被制作为软件文件(应用程序),直接粘贴到电子邮件通信中(使用TCP/IP或其他通信协议)。如果受访者来自同一组织机构(Ramos, Sedivi, & Sweet, 1998),且使用相同的电子邮件系统,调查实施会更为便捷。此外,电子问卷也可以通过调制解调器(借助电话线路),通常是经公告板系统(BBS)直接发送至受访者的电脑。在这两种情况下,受访者在运行应用程序并完成问卷填写后,都需要将其作为附件回传给研究人员。

20世纪80年代末和90年代初,随着互联网上标准化电子邮件服务的兴起,一

种统一的电子邮件形式应运而生并迅速发展,这推动了标准化电子邮件调查的使用。从形式上看,标准化电子邮件调查属于EMS的子集;然而,我们经常将EMS限制为标准化电子邮件调查出现之前的早期形态。

在电子邮件调查中,除了延续早期EMS将软件文件附于邮件的形式,还出现了两种新的选择方式:一是将调查表嵌入电子邮件正文,二是在电子邮件中附加电子文本文件(如MS Word、Excel或PDF文件;Porter, 2008)。其中,第一种方式逐渐成为电子邮件调查的主流。最初,这类调查多采用纯文本(ASCII)格式嵌入电子邮件正文。通常,这种基于文本的电子邮件调查,形式上与纸质问卷颇为相似(Schonlau, Fricker, & Elliott, 2002, p. 20)。文本会以简单方式进行格式化,以便在不同电子邮件客户端(系统)中的呈现保持一致。一般来说,受访者需通过回复邮件参与调查。回答的方式是在所选选项旁键入一个X,这些选项用方括号[]标识,或者键入与所选选项相对应的数字或字母(例子见Sheehan & Hoy, 1999)。这种方法非常原始,正如Grossnickle和Raskin(2001)在第一本网络市场研究手册中所写:“通过电子邮件进行调查的唯一理由是,所调查的目标人群可以使用电子邮件,但不能使用网络”(p. 140)。

上述方法可视为现代电子邮件客户端出现前的临时调查解决方案(Schaeffer & Dillman, 1998)。现代电子邮件客户端能够处理HTML(Porter, 2008),这在一定程度上推动了电子邮件调查的复兴。具体而言,在现代基于HTML的电子邮件调查中,问卷的外观、使用体验以及通过网络浏览器访问的问卷相同。但电子邮件调查既不依赖网络也不借助浏览器。HTML支持的各类数据输入形式在此均适用,比如单选按钮、复选框、下拉菜单和文本框等。嵌入的HTML表格还能实现自动记录答案到数据库中。尽管基于HTML的电子邮件调查在形式上不属于网络调查模式,但网络调查方法论的大部分原则在此仍能得到充分应用。

不过,这些基于HTML的电子邮件调查仍与文本型电子邮件调查存在相似局限,它们无法处理跳转逻辑(分支)或数据校验,互动性也远不及网络问卷。由于这些局限,它们仅适合短小且简单的问卷调查。另一个缺点是,若处理不当,问卷在小型设备上的显示可能出现問題。同时,不同电子邮件系统之间难以保持一致的呈现效果,这也给作答数据的提取带来了技术层面的困扰(Krasilovsky, 1996)。然而,相较于网络调查,基于HTML的电子邮件调查有显著优势:受访者不需要打开问卷就能直接看到问题,而网络调查则需要先打开包含调查URL的邮件邀请。基于HTML的电子邮件调查中,受访者打开邮件就能看到第一个问题可能会提升其配合度,尤其当首个问题设计得简单、有吸引力时。尽管存在明显缺陷,但在

涉及极简单、简短的问卷（例如关于消费者满意度的调查）时，在某种程度上可以预期，基于 HTML 的电子邮件调查的使用量仍有望增长。关于电子邮件调查的应用案例、优势、局限及最新研究，本书建议读者参考 Mesh（2012）或 Porter（2008）的文献。

将文本处理文件（如 MSWord）、电子表格（如 Excel）或其他文件格式（如 PDF）作为附件添加到电子邮件中，是另一种借助互联网发放问卷的方式。这些文件可以专门用于数据收集（如锁定说明文本和标记开放输入区）。这种方式的重要优势在于，它能实现更优的格式化效果——无论是在图形呈现和还是在格式化标准方面，均优于电子邮件（基于文本和 HTML）调查，甚至优于网络调查，网络调查中我们不能完全控制不同网络浏览器向受访者展示问卷的方式。此外，数据提取可以更加顺畅地实现自动化，且对于 PDF 文件，还可通过编程设置一些一致性检查（如范围验证）（Donald, 2010）。不过，其缺点是给受访者带来较重负担，他们必须保存文件、填写问卷并将其作为电子邮件的附件回复给发起者。然而，对于商业调查来说，附加一个可编辑的表格（即文本文件）是一个很好的解决方案。因为在商业调查中，同一个问卷是由多个受访者填写，这种很容易分发。尽管现代网络调查软件提供了更优越的格式，但对部分研究人员来说，准备相应的表格可能更简便、成本也更低。

当互动式问卷在受访者的电脑上（即客户端电脑）运行并被填写时，我们称之为基于客户端的（*client based*）调查。从问卷设计的角度来看，EMS 和电子邮件调查中，电子问卷以软件文本附件（即应用程序，*application*）的形式存在。这与通常基于服务器的网络调查（即计算机化问卷的互动性通过服务器来运行）相比，存在一些优势。例如，传统上基于客户端的调查可以更精细地控制问卷的标准化外观、跳转逻辑、可用性和随机化等方面（Saris, 1991, pp. 40-49），且运行速度更快，图形与多媒体呈现也更灵活。然而，与网络调查（通过浏览器在服务器上运行）不同，交互式调查需要在客户端（受访者的设备）上运行一个可执行文件。受访者有时会拒绝这类调查，因为他们不知道如何运行该文件，或者担心电脑会感染病毒（Porter, 2008）。此外，这类应用程序的正常运行往往需要额外软件，而客户端设备或平台可能并未预装，这就涉及操作系统对相关软硬件功能的加载。最常见的例子是 *Flash*（用于显示图形和动画的多媒体软件，可以作为插件用于大多数网络浏览器）和 *Java*（允许用户运行用 Java 编程语言编写的程序的软件）。另一个相关问题涉及附件的大小，若文件过大，可能会被部分电子邮件系统拦截。

现今，技术层面已攻克了上述基于客户端应用问卷的限制：整个问卷，包括通

过网络或电子邮件自动向研究人员传输应答, 都能顺畅运行。与基于服务器的网络调查相比, 现代客户端运行的问卷不会给受访者增加额外负担, 同时提供了更快的速度 (因在完成问卷时没有与服务器互动)、更高的多媒体灵活性 (突破浏览器的潜在限制), 并且不受特定网络浏览器特性的影响 (这一点在网络调查中研究者难以控制)。

虽然所有设备都支持标准的网络浏览器界面, 但基于客户端的应用程序问卷的运用仍面临阻碍, 主要因为相应的定制化应用程序成本较高, 往往需要在用户友好的高性能网络浏览器软件之外单独开发, 还需要为不同的操作系统单独编程。可见, 与网络调查相比, 基于客户端调查的应用占比微乎其微, 至少对于台式电脑或笔记本电脑用户来说是如此。基于客户端调查衰落的另一个关键性因素是, 需要开发功能强大的浏览器脚本, 并可以在客户端计算机上运行。这种脚本扩展了基于浏览器的网络调查功能, 使其具备了更多原本属于客户端调查应用程序的功能。需要补充说明的是, 从形式上看, 基于客户端的调查应用程序不属于网络调查, 因为填写问卷时不需要使用网络浏览器。其运行也不需要与服务器互动, 而“与服务器交互”正是 (基本) 网络调查模式的核心特征。

不过, 如今基于平板电脑和智能手机的客户端调查应用越来越流行, 使得客户端与服务器的权衡问题再次浮现, 这一点我们将在第 5.1 节进一步探讨。

计算机辅助调查, 或电话调查, 是网络调查更确切地说是在线调查的另一前驱形式。该形式可追溯至 20 世纪 80 年代, 荷兰开展了最早的尝试之一 (Saris, 1998), 其思路是通过概率抽样招募全国范围内的家庭样本, 为这些家庭提供台式电脑和电话调制解调器, 以便运行计算机化的自填问卷。当然, 当时尚无互联网、万维网和网络浏览器, 因此这并不属于网络调查。1999 年, 在首个电话调查 (*telepanel*) 出现 10 多年后, 美国知识网络公司 (Knowledge Networks) 启动了类似的基于概率的全国性在线调查。该公司通过随机数字拨号 (*Ran-dom digitdialling*, 简称 RDD) 招募家庭, 并为每个家庭提供网络电视 (*WebTV*) 作为问卷发放的技术支持 (Krotki & Dennis, 2001), 由此实现了现代意义上的网络调查。自首次出现以来, 在线样本的招募与维护方法以及向样本库成员提供设备的选择等方面, 均有了重要发展 (详见第 5.2 节)。

1.3.3 现代网络调查的兴起

20 世纪 80 年代网络调查的前身, 如电子邮件调查 (EMS)、标准化电子邮件

调查和电话调查等形式,虽未依托网络(或互联网),却已实现了在线计算机化调查的早期实践。直到现代网络浏览器出现,且其图形界面不断迭代升级后,网络作为调查媒介的功能才得以充分发挥。这始于1994年网景浏览器(*Netscape browser*)的推出。网景浏览器不仅推动了互联网的大众化应用,也促使区域性网络调查与全球网络调查(如GVU调查)³迅速拓展。在此阶段,电子邮件仍然是联系受访者的核心技术手段。值得一提的是,截至2015年,浏览器市场⁴由Chrome、Firefox、Internet Explorer、Safari和Opera占据。

20世纪90年代的首批网络调查呈现出较强的静态特征,其设计多依赖网络表格序列(即普通的HTML表格),问题以逐页呈现的方式排列在网页中。从本质上讲,这类调查仅是纸笔问卷(P&P问卷)的计算机(电子)化复刻。除借助互联网传递(即通过网络将答案自动传输到研究者的数据库)外,这些问卷几乎不具备任何计算机化问卷的先进功能,因此篇幅短小、设计简单,且缺乏分支逻辑。

1994年以后,随着具备灵活图形处理能力的主流网络浏览器持续发展,交互式网络调查开始逐步普及。这类网络调查主要依靠基于服务器的互动。每个网页(如一个单独的HTML表格)会被实时发送至服务器,服务器可根据作答内容即时评估,并动态生成下一页内容。这种交互性是网络调查的关键特征,也被纳入了基本网络调查模式的初始定义中。除非有明确限定(如在电子邮件调查中),一般来说,计算机化问卷通常默认具备交互属性。

近年来,尤其是从2000年开始,随着web2.0的引入,通过特定的浏览器脚本,网络验证器(*validations*,如范围验证)和各种交互行为(如基于作答内容的实时反馈)已可在客户端(即受访者设备端)实现。目前这主要通过JavaScript实现。JavaScript是一种面向对象的计算机编程语言,通常用于网络浏览器中创建互动效果。它可以在客户端设备上添加问题验证、逻辑分支和即时提示等功能,且无需与服务器进行交互。这增强了问卷调查界面的反馈效率和互动性。这不仅减少了服务器通信频次以提升响应速度,更在界面设计上赋予了更大灵活性,特别是交互功能。随着网络调查中基于客户端的互动功能不断增强(即JavaScript),完全基于客户端独立应用程序(如Java)运行的问卷的优势被进一步削弱,这也是基于客户端调查方法衰落的重要原因。如今,几乎所有的网络调查都是基于服务器的,通常需要受访者启用JavaScript才能确保正常运行。

早期的网络调查是由受访者使用其台式电脑、笔记本电脑进行访问和回答。在特殊情况下,如前文提到的知识网络调查(Knowledge Network panel),也会用到网络电视设备。然而,在过去的十年间,诸如智能手机和平板电脑等小型便携设备成

为访问互联网和参与网络调查的主流终端。这为网络调查问卷的设计提出新的挑战,比如如何在比台式机或笔记本小得多的屏幕上呈现调查问题。关于这一议题,我们将在 5.1 节展开详细讨论。

伴随网络调查的规模扩张,配套软件支持也迅速发展起来,使网络调查的实施更加便利。最初,网络调查软件仅支持 HTML 表格中最基本的浏览器兼容的元素:即单选按钮、下拉菜单、复选框、表格、文本框和图形图像等。这些元素至今仍然是网络调查问卷的核心组成部分。当然,现代网络调查软件的功能已经逐渐扩大,其完整概况将在 5.3 节进行介绍。

1.3.4 调查模式的背景

在探讨计算机辅助调查信息收集(CASIC)时(表 1.2),调查模式的差异主要体现在两个核心方面(Biemer & Lyberg, 2003),有无访员参与(访员管理式调查与自我管理式调查)和问卷是否实现计算机化(纸笔问卷与电子格式问卷)。

现代调查研究的发展已历经近百年(J. M. Converse, 2009)。其起点为纸笔问卷调查,早期形态包括访员管理的面对面(F2F)纸笔问卷调查和自填式邮寄调查,随后出现了访员管理的电话调查。直至 20 世纪 70 年代,调查数据收集的计算机化进程才正式开启,各类计算机辅助访员管理调查(即计算机辅助调查,CAI)相继涌现。计算机辅助电话访问(CATI)的出现得益于大型计算机在问卷处理与样本管理方面的技术突破(Couper & Nicholls II, 1998)。在 80 年代中期,便携式电脑的普及,使得计算机辅助个人面访(CAPI)逐渐可行(Couper & Nicholls II, 1998; Saris, 1991, p. 5),有时还可以结合计算机辅助自填式面访(CASI)完成部分问卷调查(de Leeuw & Nicholls II, 1996)。在 CASI 模式中,访员会向受访者介绍计算机化自填式问卷,问题除通过标准视觉(文本)形式呈现外,还可以以音频(ACASI)或视频格式(VCASI)传递给受访者。

20 世纪 80 年代中期,出现了计算机辅助自填式问卷(*computerized self-administered questionnaires*, CSAQs),即由受访者在没有访员参与的情况下独立完成问卷作答的整个过程(Ramos et al., 1998)。CSAQs 可以采取不同的形式。早期的 CSAQ 形式是邮寄软盘(DBM)。受访者通过邮件收到一张软盘。软盘可以在电脑上运行,并且包含用于处理问卷、记录答案的执行程序(应用程序)。问卷完成后,数据被保存在软盘上,然后受访者将软盘邮寄回数据收集机构(de Leeuw & Nicholls II, 1996; McNaughton, 1999)。

此外，计算机辅助自填式问卷还包括按键式数据输入（TDE）（Clayton & Werking, 1998）和后期出现的交互式语音识别（IVR）（Miller Steiger & Conroy, 2008）两种形式，依托电话来开展简短的自我管理调查。受访者收听录制的问卷，并使用电话按键（即 TDE）或自动语音识别系统（即 IVR）输入答案。所谓的机器人调查（*robo-surveys*）是这种方法的特殊形式，电话拨号由系统自动完成（即机器人拨号）。这代表了一种无访员（自我管理）和无纸质化媒介（计算机化）数据收集的典型案例。随着移动电话的出现，短信调查（*SMS surveys*）也成为 CSAQ 的新形式。

在 CSAQs 及各类调查数据收集形式的发展中，始于 20 世纪 90 年代的依托互联网的数据收集成为最显著的趋势。如前文所述，电子邮件调查和邮件调查率先出现，随后是网络调查逐步兴起。如今，在互联网覆盖率高且民众具备一定计算机操作能力的发达国家，网络调查已在调查行业中占据主导地位。凭借自我管理的便利性、较低的成本和较高的速度，网络调查逐步替代面对面调查和电话调查；与此同时，后两者的应答率也在持续下降（Brick & Williams, 2013; de Leeuw & de Heer, 2002）。

经历了初期对这种“替代技术”的热情追捧、发展波动与过高期待后（Black, 1998），网络调查如今正走向成熟。正如互联网未取代广播、电视、报纸等传统媒体，

表 1.2 调查模式的管理和计算机化

执行管理	计算机化程度	
	纸笔调查（P&P）	计算机辅助调查信息收集（CASIC）
访员管理的调查	P&P 访谈（PAPI）	计算机辅助调查（CAI）
	● 面对面 P&P 调查 ● 电话 P&P 调查	● 面对面计算机辅助个人调查（CAPI） ● 通过固定或移动电话进行的计算机辅助电话调查（CATI）
自我管理的调查	自我管理的 P&P 调查	计算机辅助自填式面访，由 F2F 访员介绍（CASI、ACASI、VCASI）
	● 邮寄调查 ● 自我管理的 P&P 调查，由访员交接 ● 传真调查	- 计算机辅助自填式问卷（CSAQs） ● 邮寄软盘（DBM） ● 电话自填式调查：电话按键数据输入（TED）、交互式语音识别（IVR）、机器人调查、短信调查 ● 电子邮件调查（EMSs） ● 邮寄调查 ● 各种设备上的网络调查 ● 移动设备调查：应用程序

而是对其进行了重新划分、改变和整合；同样，网络调查也不会取代其他调查模式，而是为整个数据收集过程带来深刻的变革。

受 Biemer 和 Lyberg（2003）和 Groves 等人（2009）研究的启发，表 1.2 总结了调查模式的基本分类。该分类呈现出从左上角到右下角的演变趋势，即从面对面访员管理的纸笔调查到 CSAQ。网络调查位于右下角，属于计算机辅助自填式问卷的范畴。在此需要说明的是，我们不采用“计算机辅助网络访谈”（CAWI）这一术语，因为“访谈”这一概念隐含着某种形式的访谈者参与/在场，而网络调查并非如此。关于调查模式的更多详细内容，将在第 6.1.1 节进一步阐述。

1.3.5 优势和局限性概述

网络调查具有显著优势，特别是早期应用时有些方面甚至会被不加批判地推崇，但这种调查模式也存在一定局限性。无论是优势还是劣势，都应基于客观事实进行评价，避免夸大其词。本章将对网络调查的主要特征进行初步的批判性梳理，为网络调查方法的核心问题提供全面视角，进而为后续章节的深入讨论奠定基础。在后续章节中，我们将对相关方法学问题展开系统且详细的分析。

1.3.5.1 成本

与邮件、电话和面对面调查相比，成本较低是网络调查最吸引人的优势之一。例如，Greenlaw 和 Brown-Welty（2009）对某协会成员的调查结果显示，通过电子邮件发送邀请的网络调查每份回复的成本为 0.6 美元，而邮件调查的单次成本为 4.8 美元，混合模式的成本为 3.6 美元。另一项以执业牙医为样本的研究对比了网络调查与纸笔问卷（P&P）的成本效益，结果表明使用电子邮件邀请的网络调查比纸质调查的成本收益高 2.7 倍（Hardigan, Succar, & Fleisher, 2012）。Strabac 和 Aalberg（2011）在美国与挪威开展的对比研究也证实：使用电子邮件邀请的网络调查，每个受访者的花费仅为电话调查的 1/3 至 1/4。值得注意的是，面对面调查的单位成本往往高达数百甚至数千欧元/美元，此处暂不展开详细对比。

与其他调查模式相比，网络调查的成本构成存在本质差异。数据收集的可变成本，取决于样本大小，通常极低。对于邮寄调查，单位可变成本涉及信封、问卷打印、邮票、回邮邮票和信封，以及数据录入的相关费用。对于电话调查，相应的可变成本取决于访问时长（即访员的成本和电话费）。而对于面对面调查，还需额外承担访员差旅费及样本家庭定位成本（如区域抽样、随机走访费用），成本会进一

步增加。对于使用电子邮件邀请的网络调查,所有这些可变成本都不存在。发送 10 份或 10,000 份电子邮件邀请的成本相同,获得 10 或 10,000 份完成问卷的成本也几乎相同。即使需要额外联络受访者(提醒受访者)或提供数字形式的材料(说明、项目背景、结果反馈等),费用也不会随着样本量的增加而上升。若涉及额外激励措施或邮寄实体邀请函/提醒通知,成本才会有所增加。

另一方面,在问卷开发和测试的固定成本方面,不同调查模式的差异较小,但网络调查的其他固定成本通常较低。以自助式网络调查为典型案例,其固定成本主要包括服务器费用与网络调查软件订阅费,如今这类软件的订阅成本已显著降低(见第 5.3 节);而其他调查模式通常有更高的固定成本,特别是访员协助模式中的培训和监测费用,或邮件调查中的管理流程成本。

以采用电子邮件邀请的简化网络调查为例(假设无无应答情况),其单位成本基本上表达为 F/n ,其中 F 代表固定成本, n 代表样本量,而其他模式的成本则为 $(F/n+c)$,其中 c 代表单位可变成本。由此可见,网络调查的成本优势并非细微差异,且随着样本量的增加,这种优势会愈发显著。不过,针对这一结论,我们需提出四点批判性思考:

- 当固定成本(F)包含维护普通人群在线样本库(尤其是概率样本库)的费用时,网络问卷每单位完成成本迅速增加,甚至可能与电话调查或邮件调查持平。在一些小型国家,在线样本库规模小,费用相对较高,网络调查的成本与面对面调查的费用相当。尤其是劳动力成本相对较低时,这一现象更为明显。这一案例有力反驳了对网络调查“固有成本优势”的过度简化认知。
- 网络调查相对较低的成本,不应该仅体现为给研究人员或客户带来财务收益,更应该转化为更高的调查数据质量。例如,网络调查节约的成本可以用于扩大样本量、增加受访者激励、强化宣传推广等,这些措施均有助于提升数据质量。
- 实施网络调查时需考虑受访者的潜在成本:在某些情况下,受访者可能需要为完成问卷支付额外的互联网费用(如通过手机作答时的流量费)。
- 与任何服务一样,成本讨论不应与质量割裂。成本固定的前提下,如果网络调查的数据质量与其他模式相当或更优,显然要优先选择网络调查。如果网络调查的数据质量较低,则需优先考量预算差异。如果网络调查的成本远远低于其他替代方案,却简单评估其数据质量较低,这种比较并不公平。事实上,数据质量和成本应同时考虑,此时情况会变得极为复杂(详见 6.1.4 小结)。

1.3.5.2 数据收集的速度

网络调查的另一大显著优势是数据收集的速度。这对时间敏感型研究而言尤为重要。这类研究需要迅速形成可操作的政策,例如选举活动期间的政治研究。同样,在市场研究中,简短的网络调查从启动到执行往往只需几天,甚至在几小时内就可以获得结果。若借助电子邮件、横幅广告等电子媒体发送调研邀请,能够在瞬间触达成千上万的受众,且几乎能即时收到反馈。不过,如果研究人员(或相关机构)此前从未开展过网络调研,也缺乏通过邮件、电话或面对面方式发送邀请的经验,那么上述优势可能会大打折扣。

网络调查在回复速度方面的优势非常明显,以至于最近发表的研究较少对此展开论述。然而,早期研究的结果不仅可以说明电子邮件(Shannon & Bradshaw, 2002; Truell, Bartlett, & Alexander, 2002)和传真邀请(Cobanoglu, Warde, & Moreo, 2001; Weible & Wallace, 1998)的情况,还能体现邮寄信件邀请的特点(Kwak & Radler, 2002a; Weible & Wallace, 1998)。Aoki 和 Elsmar(2000)的报告揭示了显著的差异:在对市场营销经理的调查中,邮寄问卷的平均回复时间为 50 天,而通过电子邮件邀请的网络调查回复时间仅为 3 天。同样 Nadler 和 Henning(1998)指出,在获取相同数量应答的情况下,网络调查需要 4 天时间,电话调查则需要 20 天。Heiervang 和 Goodman(2011)也提到,网络调查的正式调查时间仅为面对面调查所需时间的四分之一。

无论何种情况,与其他数据收集和邀请模式相比,采用电子邮件邀请的网络调查其应答数据的积累速度更快。图 1.2 呈现了 2011 年关于斯洛文尼亚科学家调查(WCSC)的典型回应模式。在该案例中,共发送了 4 551 封电子邮件邀请,随后又发出了两封提醒函,所使用的电子邮件地址均通过斯洛文尼亚科学家的官方登记处获得。从图中可以看到,大部分点击量出现在邀请发出当日,其次是提醒函发出之日,最终共回收 1 104 份有效回复。详细情况见第 1 章补充资料《斯洛文尼亚科学家调查的案例研究》(<http://websm.org/ch1>)。此外,与许多 F2F(Heerwegh & Loosveldt, 2008)和电话访谈(Lindhjem & Navrud, 2011),以及自填式纸笔 P&P 调查问卷(Hertel, Naumann, Konradt, & Batinic, 2002)相比,网络调查通常也为受访者节省时间,但也不总是如此。对于有经验的用户来说,使用键盘、鼠标和电脑屏幕回答网络问卷,比用手写回答或将答案告诉访员并由其记录,更为简便快捷。但需要注意的是,问卷完成时间的缩短可能意味着受访者对调查问题的重视程度降低,进而导致数据质量下降。

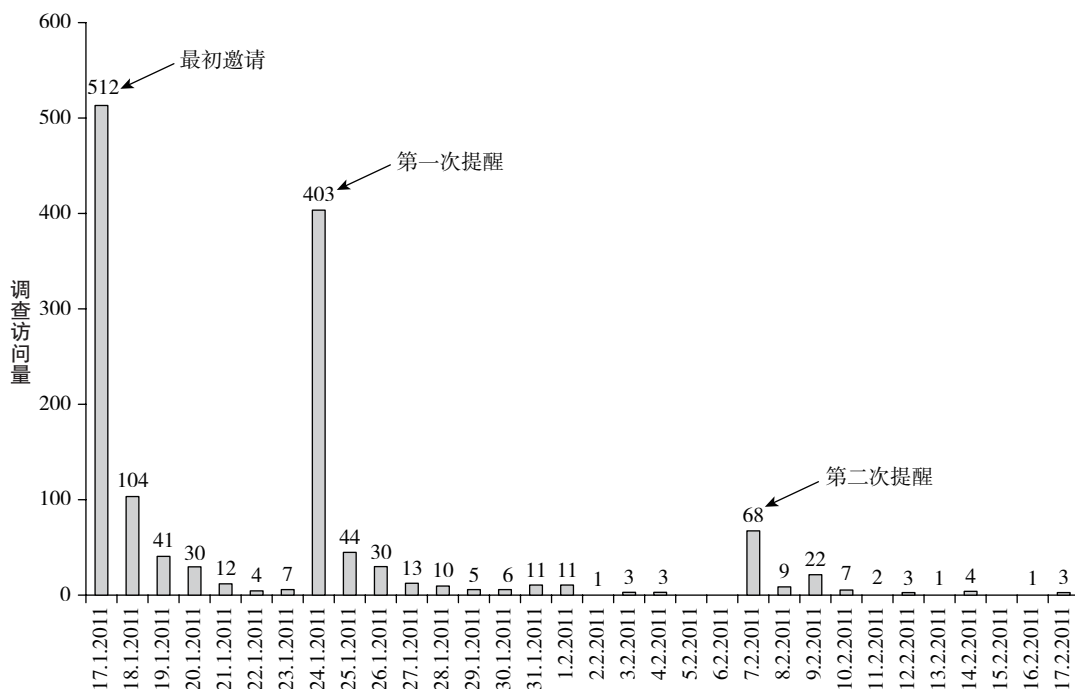


图 1.2 访问 WCSC 网络调查问卷，2011

总而言之，网络调查，特别是当使用电子邮件邀请时，通常比传统的调查模式更快，这对于时间敏感型研究（如投票意向的政治调查）而言是一个突出的优势。不过，我们还是提供了三个批判性见解。

- 与调查成本类似，速度仅是衡量数据质量的一方面。有时速度至关重要，但在另一些情况下，它可能无关紧要，甚至因被某些缺陷抵消而产生负面影响。
- 需注意的是，上述结果均来自观察性研究与准实验研究。如果在一项实验研究中，针对整个普通人群样本，其电子邮件和（移动/固定）电话的联系信息均可获得，那么与采用相应针对性呼叫策略的电话调查相比，使用电子邮件邀请的网络调查会更快地收集到答卷吗？也许会，但网络调查的回复率可能会大大降低（如 10% 对 40%），而这对研究来说是至关重要的。
- 前文只探讨了完成问卷阶段（实地调查环节）的速度优势，然而网络调查的许多更关键优势其实体现在问卷准备（实地调查前）和结果即时获取阶段（调查后阶段）。

1.3.5.3 实施的便利性

现今,网络调查软件越来越能实现用户友好型数据收集,操作也愈发便捷;并以类似的方式整合了网络调查过程中的其他各个环节。因此,即便是缺乏编程技能、甚至计算机知识有限的研究人员,也能够独立开展(网络)调查。加之,互联网简化了潜在应答者的访问,这使得实施网络调查变得容易。无论是对简单调查、自助式调查,还是对涉及高级问卷设计、复杂实施流程及严苛方法要求的复杂网络调查而言,其实施便捷性均优于传统调查模式及其他计算机辅助调查(CAI)模式。当然,整个网络调查过程仍然非常繁琐,需要由具备各类技能的人员(如程序员、抽样专家等)组成团队协作完成,但实施难度已远低于传统调查模式。

值得注意的是,网络调查系统在初期适应与启动阶段存在一些问题,尤其是专门调查机构将其用于集成数据收集服务时。综合来看,这仍然比为电话、面对面或邮寄调查引入复杂系统要容易得多。

实施便利性是网络调查较突出的特点之一,这几乎是一项绝对优势,至少对于熟悉计算机操作的研究人员而言。这种便利性在任何情况下都不会消失,也无法被其他方面的优势所替代。

1.3.5.4 问卷的计算机化

网络调查属于计算机辅助调查信息收集(CASIC)的范畴,其许多特点都有助于提升数据质量,并优于纸笔模式(P&P)(de Leeuw, Hox, & Snijkers, 1995)。具体体现为:可避免分支错误(如问题顺序错乱、遗漏或循环);可以即时检验数据(范围校验、一致性核查、无作答检测等);可对问题或选项进行随机化,以避免顺序效应的影响;支持自适应问卷(根据前期答案动态调整后续内容);便于实现问卷的随机分配和受访者问卷模块的分配;无需单独的数据录入环节,从而杜绝了数据录入错误。

此外,系统可记录每位受访者的活动时间戳(如访问问卷网页的时间)及其他调查过程数据(见第2.4节)。这些信息不仅有助于从终端系统故障中恢复数据,还有助于更高效地管理调查流程。

当然,计算机化的调查问卷需要正确编程。在实践中,因编程错误导致数据丢失或混淆的案例并不少见。同时,问卷还需要根据网络调查模式(以及一般的计算机辅助自填式问卷,CSAQ)进行调整,以满足方法学方面的要求。例如,从方法学的角度来看,研究者在设计强制性检验提示时应格外谨慎,如设定受访者在作答某

问题后才能继续回答后续问题，因为这种做法会在其他多个方面影响数据质量。又如，部分网络调查软件可能出现表格中各选项间距不一致的问题，这种设计是不恰当的，易引发测量误差。关于此类问题的更多细节，详见第 2.3.1 节。

网络调查（以及任何计算机辅助自填式问卷，CSAQ）获得高质量数据的另一个前提条件是受访者的接受程度。一般而言，受访者对计算机辅助自填式问卷的接受度较高（Beckenbach, 1995）。问卷的计算机化因涉及自动化的分支，通常会减轻受访者的负担。此外，在作答过程中，受访者若短暂离开，返回作答时计算机化问卷可准确定位至离开前的题目。另一方面，受访者的计算机素养可能会引发不可预知的错误（Olson, Smyth, Wang, & Pearson, 2011）。

综上，原则上计算机化的调查问卷可以提高数据质量及受访者满意度。但这一优势的实现需满足三个条件：问卷设计为对受访者友好、没有编程错误，并基于合理的方法论原则。本书将在第 2.3 节对此展开进一步讨论。

1.3.5.5 多媒体

上述内容多为计算机化问卷在应答过程中的普遍优势，而多媒体功能同样值得单独探讨。因为多媒体功能可以为网络调查问卷（以及一般意义的 CASIC）提供扩展和新颖的界面。通过不同的视觉和听觉刺激，调整背景与字体的颜色及类型、采用新颖的问题呈现方式，尤其是嵌入各类程序（如动画、配乐、视频片段、虚拟访员等）、运用个性化设计、展示产品三维图像及其他先进设计元素，可拓展调查问题的范围与性质。现今，网络调查已能远程实现多种问题类型的访问（例如对视频片段的评价），并且可以代替产品和广告测试中（例如品牌、标志、公司名称区分）使用的“真实”物理刺激（如面对面调查中的展示卡或实际产品）。此外，这些可以使调查对受访者更有吸引力，激励他们参与。然而，使用时需注意以下事项：

- 受访者设备的兼容性、图形分辨率以及所需下载时间。如果受访者需要等待太长时间来加载外部功能，因软件不支持这些功能而无法访问问卷或无法看到问卷的部分内容，其参与积极性会降低。出于这些原因，研究人员应该在网络调查中只使用经过核实无误的技术。Dillman、Tortora、Conradt 和 Bowker（1998）的一项研究比较了设计花哨与设计简约的网络问卷，发现后者的回复率更高、更易被完整作答，且受访者耗时更短。幸运的是，随着信息与通信技术（ICT）基础设施和软件的发展，特别是在使用独立的台式机或笔记本电脑，且连接到高速互联网（这正成为发达国家的标准）的情况下，这些技

术问题正日益减少。然而,目前当受访者使用智能手机或平板电脑等小型手持设备来回答网络问卷时,这仍需重点关注(Callegaro, 2010; 更多信息请参见第5.1节)。

- 另一个问题是多媒体功能是否影响及如何影响结果。图形图像可能会对受访者的回答产生不可预知的作用(见第2.3.3节)。因此,在选择多媒体功能来激励受访者或呈现问题时,需保持谨慎。一个基本原则是只有在必要时才使用这些功能。
- 极端情况下,调查工具可能会变得更趋近于游戏而非问卷,相关问题将在第2.3.4节进一步探讨。

1.3.5.6 时间和地理上的灵活性

时间和地理上的灵活性是网络调查的显著优势。网络调查(或网络调查软件服务)全天候(24/7)收集数据,不受时区限制、无地理限制、可即时访问世界上任何一个地方的网络问卷的特性,正是这种优势的直观体现。

通过网络调查进行国际调查和比较研究颇具优势。问卷可在中心节点以基准语言完成编程与测试,并整合计算机辅助自填式问卷(CSAQ)的所有分支逻辑、校验功能及视觉设计等要素;还可以单独添加翻译后的问题、错误信息提示页、导航按钮及其他系统信息的文本。这一模式能大幅节省时间与成本,同时减少误差。

网络调查还能以较低的成本轻松联系到特殊或低发人群。相较之下,其他调查模式需要对大规模人口和地理上广泛分散的人口进行筛选,成本较高。对于面对面或邮件调查来说,这一因素往往会阻碍数据收集。

另一方面,当对某一特定地区或国家的人口进行调查时,网络调查“无地理界限”的特性可能转化为劣势。因为在网络调查中,很难对其进行地理限制。此外,若需不同地区的参与者加入调查,还需额外提供语言选择功能,并对问卷及系统信息进行适当翻译。

1.3.5.7 自我管理

与所有自填式调查一样,网络调查对受访者来说很便捷。他们可以按照自己的节奏,自由选择参与调查的时间和地点,无需提前通过电话预约或安排面对面会面。在某些情况下(具体取决于调查问卷设计),受访者就像翻阅纸笔调查问卷那样可以浏览整个调查问卷,还可以随时向前和向后滚动以修改答案。若能提供足够的激励,受访者也会花时间查阅相关记录与文件,以确保问卷填写的准确性(Callegaro

& Wells, 2008)。这对商业调查尤其重要,因为受访者可能因工作中断需暂停作答,或需要在文件中查找答案或咨询知情人。经过编程设计的网络问卷,支持受访者在部分作答后暂停,稍后返回继续完成剩余内容。

网络调查无需访员参与,这一特点也构成其优势。因为访员可能会存在特定偏见且访员之间存在差异。而在访员管理的调查中,访员的情绪、偏见和态度会影响数据质量。网络调查因无访员介入,可避免这类潜在误差。相应地,网络调查更易获得有关敏感话题的高质量数据。网络调查受访者在回答问卷时可能会有更强的隐私感,给出的答案较少受社会赞许性影响。多项研究表明,与纸笔问卷(Booth-Kewley, Larson, & Miyoshi, 2007; Joinson, 1999)、电话调查(Chang & Krosnick, 2009; Kreuter, Presser, & Tourangeau, 2008; Lozar Manfreda, Couper, Vohar, Rivas, & Vehovar, 2002)、交互式语音应答(IVRs)(Kreuter et al., 2008)以及面对面调查相比(Heerwegh, 2009),网络调查更容易获得社会赞许性较低的应答。此外,网络调查还可以探讨线下难以施测的棘手主题,例如酒精和药物使用(McCabe, 2004; McCabe, Boyd, Couper, Crawford, & D'Arcy, 2002)、健康风险行为(Crutzen & Göritz, 2010)或性行为(Ross, Månsson, Daneback, Cooper, & Tikkanen, 2005)。值得注意的是,网络调查中社会赞许性偏差的程度也取决于受访者对匿名性的感知以及调查的个性化程度。例如Joinson(1999)发现,与匿名调查相比,非匿名调查中的社会赞许性答案更多。Heerwegh和Loosveldt(2007)发现,与非个性化的调查相比,个性化的网络调查中,社会赞许水平高的答案更多。

另一方面,网络调查缺乏人际互动,这也可能对数据质量产生负面影响。如可能会出现严重的自我选择偏差,受访者可能需要指导却无人协助,或参与动力不足,亦或难以在缺乏互动的情况下独立完成整份问卷。与其他调查模式相比,元分析结果表明,网络调查的应答率低于访员主导的调查(Lozar Manfreda, Bosnjak, Berzelak, Haas, & Vehovar, 2008)。此外,网络调查的受访者可能同时进行其他活动(如线上或线下多任务处理),导致回答网络问卷时注意力分散。此外,网络调查无法像面对面或电话调查情境一样进行追问,这对于有多重反应的问题(Smyth, Christian, & Dillman, 2008)、开放式问题和漏答来说(Heerwegh, 2009; Roster, Rogers, Albaum, & Klein, 2004; Stephenson & Crête, 2011),可能尤为棘手。由于数据收集过程不可观察,研究者无法确认受访者是否理解并遵循指导语;同时,也难以控制作答时的环境因素,如高温、噪声和受访者的情绪、疲劳程度或物质使用等。尽管其他调查模式也存在类似问题,但在某些情况下,例如在面对面访谈中环境因素至少可以测量。同样,网络调查无法捕捉诸如受访者的面部表情、身体语言、

语气、服装等背景信息，而这些信息可能有附加价值。

最后，许多调查机构会对超长问卷（如 CATI 或 CAPI 中时长 1 小时及以上的调查）采用混合模式收集数据，但现阶段没有足够的文献明确混合模式对无应答率和应答质量的影响。换言之，网络调查是否可以代替访员主导长达一小时的调查，并且仍然可以保证数据质量不受影响？对于网络调查，典型建议时长为 20 分钟，与电话调查时间相似（Macer & Wilson, 2013）。这一时长也适用于基于概率的在线样本库（如 Gesis 互联网样本库、LISS 样本库）中的网络调查，详见第 5.2 节的论述。

此外，缺乏经验或有特殊需求的互联网用户可能难以自主完成网络调查。例如有特殊阅读理解困难的失聪者（Fontaine, Jacquemain, & Italiano, 2013），或者患有唐氏综合征或威廉姆斯综合征的人（Marcell & Falls, 2001），在回答网络问卷时可能会遇到障碍。身体有缺陷的人（如盲人）、老年人或计算机经验不足的用户，在使用键盘、鼠标以及观看显示器时存在手眼协调问题，从而使用电脑完成问卷存在一定困难。因此，尽管计算机化的问卷比纸笔问卷更具客观性，但对部分特殊人群而言，计算机化的问卷可能造成困扰，进而降低其参与意愿（Couper, Tourangeau, & Smith, 1998, p. 435）。

1.3.6 网络调查的应用

本节介绍了现代网络调查的典型示例，并明确阐述了相关抽样类型。这有助于厘清调查模式和抽样特征之间的关系。此外，我们探讨了许多关于基本方法学困境的实际案例，这能帮助读者更好地理解后续章节（第 2-4 章）。

本节围绕网络调查在八个典型领域的应用展开论述。这八个领域的确定主要基于其普遍吸引力与认可度。因此，对领域的“分类”缺乏明确的基本维度和正式标准，而是融合了网络调查实施过程中呈现的若干重叠特征。除抽样类型外，这些特征还涉及受访者、研究者和委托方（发起方）的类型，以及调查主题和方法学细节等。

1.3.6.1 对高覆盖率的特定人群进行网络调查

当特定的目标人群完全（或大部分）接入互联网，且抽样框中包含该目标人群的完整联系信息（最好是电子邮件地址）时，网络调查便能发挥显著效果。在这种情况下，研究者很容易抽样或向全体对象发送电子邮件邀请（普查），且无论采用哪种方式，都有可能开展高质量的概率抽样网络调查。

这类调查的应用案例如下。其一为组织内部进行的调查。许多组织和协会的雇员 / 成员可以通过网络取得联系, 而组织调查可以充分利用这一优势。典型案例是对企业员工的网络调查 (如人力资源调查), 因为企业会给每个员工配备电子邮件地址。因此, 企业内部或外部合作方可以基于电子邮件地址抽取样本, 邀请员工参与调查。这类调查通常采用普查方式。对于这类调查, 最好的做法是加大宣传力度、鼓励全员参与, 以实现 100% 的应答率, 进而支持在极精细的层面 (如基层管理岗) 形成报告。

其二, 大学生网络调查 (即学生调查)。大学生群体是网络调查的理想对象, 因其互联网覆盖率与电子邮件拥有率近乎 100%, 但实践中也常遇问题。一是登记处的学生信息不准确。由于学生群体流动性强, 其邮寄地址、(多个) 电子邮件及电话号码频繁变动, 导致信息难以及时更新。解决这一问题的办法包括要求学生使用学校统一发放的电子邮件账户, 并定期检查信息有效性。二是保密性问题, 学校收集学生联系信息通常用于特定目的, 因此往往不允许研究人员 (或学术机构) 使用或传递这些信息。

其三, 特定协会 (如专业协会) 的成员调查。这类协会的成员通常可通过电子邮件联系, 但存在两类问题: 记录的电子邮件可能无效 (电子邮件错误或没有更新); 二是部分会员加入协会时没有提供电子邮件地址。在这种情况下, 如果目标人群是协会全体成员, 研究者需对无有效电子邮件的部分成员采用其他调查方式 (如邮寄调查)。

1.3.6.2 客户满意度调查

本部分将单独介绍客户满意度的网络调查, 这是针对高互联网覆盖率人群的另一类网络调查形式。此类调查的特点在于目标人群的特殊性 (即以客户名单作为抽样框架) 和调查内容的高度针对性 (即聚焦客户满意度)。企业常通过客户满意度调查收集客户反馈, 涵盖满意度、忠诚度、投入度等方面。网络调查作为电话、邮件和交互式语音应答 (IVR) 客户满意度调查的替代方案, 效率更高 (更快响应) 且成本更低。通常情况下, 这类调查高度标准化, 规模庞大, 并被整合至企业更广泛的信息系统中 (如客户关系管理、销售支持系统、企业资源规划)。有时也会在企业反馈管理 (EFM) 系统框架下开展客户满意度调查。客户满意度调查或许是大众最常接触的调查类型之一。关于此类调查的介绍, 读者可以参考 Hayes (2008) 和 Roman (2011) 的研究, 而针对测量时间和测量误差的论述, 可以参考 Dillman、Smyth 和 Christian (2009, p. 353)。

客户满意度调查主要分为两类：交易型和追踪型（transactional and tracking）。对于交易型满意度调查，最好是在客户与该公司完成交易（与购买或服务有关）后不久，通过发送电子邮件或短信邀请客户参加调查。此时网络调查的主要优点是，可即时获取用于改善交易质量的反馈。追踪型满意度调查，通常比交易型周期更长，是对客户群体的一种“把脉”。通常企业给整个客户群发送邀请进行调查，重点关注对企业、产品或服务的整体满意度。最近，这类调查的重心逐渐从满意度转向忠诚度。为了监测满意度或忠诚度水平，企业可以采用横断或纵向设计开展多次重复调查。在这两种情况下，客户都会收到带有网络调查链接的电子邮件邀请。近期，特别是在美国，基于 HTML 的电子邮件调查被越来越多用于实现这一目的。

在客户满意度研究中，网络调查的主要优势包括速度快、能即时得到客户反馈，且与线下方法相比可显著节约成本。然而，其最大优势在于抽样框的可获得性：客户名单（理想情况下包含电子邮件地址）可作为样本库。企业可以定期向其发送网络调查的电子邮件邀请函。在符合法律与伦理规范的前提下，还可将现有客户数据与个人应答数据进行匹配。

需要注意的是，当将调查结果推论至全体客户群体时，需密切监测未提供电子邮件地址的客户比例。如果这个比例较高，且有证据表明已覆盖的客户可能与未覆盖的客户存在异质性，企业或服务提供商应考虑转向混合模式调查，或开展额外的非覆盖误差研究。

企业经常建立所谓的专有样本库，将其自身客户纳入研究，并加入到在线社区样本库（见 5.2 节）。

1.3.6.3 普通人群调查

当对普通人群开展概率调查时，互联网普及率的提高使人们乐观地认为网络调查可以取代其他数据收集模式（Smyth & Pearson, 2011）。尽管可使用网络的人口比例持续增加（2013 年，欧洲和美国的这一比例均超过 75%，部分国家甚至达到 90%），但实际情况并非如此，网络调查无法完全替代其他调查模式。

另一个相关问题是，互联网用户和非用户在社会人口特征和生活方式方面并非随机分布。在许多国家，网络连接速度和宽带还是拨号的差异也是一个因素，这引发了关于“第二代数字鸿沟”的争论，即快速与慢速互联网连接之间的“采用差距”（Economic and Statistics Administration & National Telecommunications and Information Administration, 2010）。

除了可能出现严重的无覆盖误差，缺乏现成的电子邮件地址抽样框是这些调查

的另一个主要挑战。电子邮件抽样框是联系普通人群的最有效手段。但电子邮件地址不像电话号码那样具有结构性,无法像随机数字拨号(RDD)生成电话号码那样,随机生成具有全国代表性性的电子邮件地址列表。因此,仍需采用标准且成本较高的招募方式(例如邮件、电话或面对面邀请)。

第三个问题涉及伦理规范和国家的具体规定,即除非存在事先联系,否则不得使用电子邮件地址联系潜在的受访者(CASRO, 2011)。在未获事先同意或无任何关联的情况下,通过发送带有问卷的电子邮件与某人联系,就类似于访员未经许可而邮寄问卷。而对RDD电话样本进行“冷呼叫”,类似于访员在没有任何事先接触的情况直接致电家庭。有趣的是,这些未经许可的非互联网联系策略通常不受任何法规的限制,但电子邮件邀请却受到约束。这种差异可能在于,电子邮件邀请不需要任何成本(与邮件、电话或面对面有很大不同),并且网络调查邀请的数量规模较大,这与垃圾邮件密切相关。

上述问题并未完全阻碍通过网络对普通民众进行调查,此类调查可通过混合模式实现。最常见的做法是:向普通人群样本发出邮件邀请,先由愿意且能够完成网络问卷的受访者参与。对于未应答者,对其进行提醒并提供纸笔问卷。这种策略在美国的应用日益广泛。例如,在电话调查中,混合模式可以避免因仅覆盖使用手机的家庭而产生无覆盖误差(Smyth, Dillman, Christian, & O'Neill, 2010);在互联网普及率高的斯堪的纳维亚国家,该策略对于重要主题的调查效果显著(Gelder, Bretveld, & Roeleveld, 2010)。

针对普通人群的另一种调查策略是使用概率在线样本库,即通过邮件、电话或面对面等模式招募受访者加入概率性在线样本库。为无互联网的家庭提供设备(如台式机/笔记本电脑/笔记本、智能手机或平板电脑)和网络接入服务,以便其定期参与调查。例如,荷兰的社会科学纵向互联网研究(LISS)样本库(Scherpenzeel, 2011),法国的社会科学纵向互联网研究(ELIPSS)样本库(Lesnard, 2011),德国的互联网样本库(Blom, Gathmann, & Krieger, 2013)和美国的知识网络样本库(GfK)(Callegaro & DiSogra, 2008)。另一种选择是混合模式方法,即以另一种模式(如邮寄或电话)对非互联网家庭进行调查。如德国的Gesis样本库(Bosnjak, Haas, Galesic, Kaczmirek, Bandilla & Couper, 2013)和美国的盖洛普样本库(Rao, Kaminska, & McCutcheon, 2010)。

请注意,近期已出现多项研究协调行动,旨在强调网络调查的重要性并探索其在普通人群中的可行性(例如GenPopWeb项目⁶和WebDataNet成本行动IS1004⁷)。

1.3.6.4 商业调查

企业或机构调查 (Snijkers, Araldsen, Jones, & Willimack, 2013) 在性质上通常与针对家庭或组织中个人的调查非常不同。在商业调查中, 收集的信息极少涉及受访者本人, 更多聚焦其雇主或所在公司。换言之, 受访者充当代理人 (*proxy*) 的角色。此外, 商业调查很少涉及态度类问题, 而是侧重于有明确答案的事实性问题。调查的发起方也不同, 通常来自统计机构、联邦机构或金融机构, 但也有来自营销研究机构和其他企业。

商业网络调查往往具备较好的抽样框, 至少有电子邮件地址, 原则上可以实现概率抽样。电子邮件邀请虽然有其优势, 但也存在许多复杂问题 (地址准确性、需要公司内部转发等), 因此仍然经常使用邮寄邀请。

在设计测量工具时, 应该牢记网络商业调查的特殊性, 其一是从业人员习惯使用表格 (*form*) 而不是“问卷”这一术语。商业调查的回答过程 (无论线上或线下) 与个人调查的过程存在差异 (Snijkers & Morren, 2010), 导致测量误差也有所不同 (Bavdaž, 2010)。例如, 商业调查通常需经过正式审批流程, 包括由授权人员向数据收集机构提供调查所需信息 (S.W. Edwards & Cantor, 1991)。纸质表格可能作为草稿与网络表格配合使用, 再将数据录入在线系统。网络商业调查的独特之处在于经常需要多个受访者协作完成, 因为单人难以掌握所有答案, 也无法接触到作答所需的全部信息来源。出于这个原因, 应该为受访者提供简易登录机制, 方便不同人员都能够进入调查系统。此外, 在商业调查中可能涉及多次会话, 需具备随时保存已输入数据的功能, 以简化受访者的操作。基于上述原因, 可以考虑将调查问卷作为 PDF、MSWord 或 Excel 文件附在电子邮件中发送。由于这些问卷 (表格) 通常不是按顺序填写的, 因此网络问卷设计需支持各部分间的便捷导航 (*navigation*)。此外, 允许打印已经输入数据的问卷, 这会简化从多个来源检索信息的任务。商业调查通常需要大量的定量数据; 因此, 应该提供在线数据检验和提示, 如根据早期问题甚至早期调查中输入的数据进行恒定或一致性检验。这既可以为受访者提供帮助, 也有助于提升数据质量。最后, 对于许多企业来说, 安全性至关重要, 因此在设计调查时需要采取措施, 以确保数据的安全传输, 尤其是涉及机密和 / 或私人信息时 (Snijkers & Morren, 2010)。

Snijkers 和 Morren (2010) 以荷兰年度结构性商业调查 (Annual Dutch Structural Business Survey) 为例, 说明了典型商业调查的完成过程。平均而言, 完成该调查需要 1 小时 7 分钟, 而调查开放时间约为 8 天。总体而言, 43% 的受访者

在调查期间打印了调查表。在这个过程中，29%的人保存过信息，65%的人使用了工具的返回功能。平均而言，在填写问卷期间受访者进行了270次操作。这些数据表明，商业调查的设计和回答比其他类型的网络调查要复杂得多。事实上，商业调查也许是最复杂的网络调查。它不仅需要遵循所有的网络调查规则，也要考虑上述具体细节以及企业内部规则。更加复杂的是，标准的网络调查软件往往不具备或无法实现所有这些功能（见第5.3节）。因此，调查机构往往需要搭建内部系统以满足特殊需求。例如，美国人口普查局一直在使用名为Census Taker的内部系统，近期更名为Centurion，以收集商业调查数据（Morrison, Dillman, & Christian, 2010）。

1.3.6.5 在线样本库

市场研究中最常使用的网络调查类型是基于在线样本库的调查（Macer & Wilson, 2013）。在此仅对在线样本库进行简短介绍。由于高频率使用和高收益，以及在未来调查研究中的巨大潜力，在线样本库值得更多关注。因此，本书会有一个专门章节（第5.2节）对在线样本库的类型、构建和维护、基于样本库开展个人调查的方法以及数据质量等内容进行更详细的探讨。

在线样本库基本上可以分为两类：概率样本库和非概率样本库。概率样本是通过邮件、面对面或电话等方式，从具有代表性的家庭或个人样本中预先招募而来。为解决普通人群的非覆盖问题，会为无互联网家庭提供设备和互联网连接，以确保他们定期参与调查，或者以另一种模式对他们进行调查（如通过电话或邮件）。此类样本库是通过网络调查覆盖普通人群的重要工具（详见前文关于普通人群调查的内容），但由于建立和维护成本很高，目前相对较少。

非概率在线样本库（也称志愿者样本库、选择性加入样本库）比概率在线样本库更常见，规模更大，效果也不同。个人自愿加入样本库，通常会因参与调查而获得奖励。尽管部分概率在线样本库也使用奖励，但在非概率样本库中，奖励几乎是必备手段。这类样本库涵盖大量参与者，可以为市场研究和民意调查提供样本，甚至能触达极为特殊（稀有）的群体。在实际应用中，常通过配额抽样（Quota sampling）和加权技术来模拟样本的代表性。

1.3.6.6 在线社区调查

过去二十年间，网络社区吸引了社会科学家的广泛关注。从20世纪80年代末的电子公告板BBS，到20世纪90年代的讨论区和网络论坛，再到21世纪的虚拟环境（如第二人生，Second Life），乃至近年社会网络的融合，研究者已经在社会、

心理和文化方面对这些社区展开探索。虚拟民族志、“大数据”挖掘和社会网络分析等研究方法均被用于这一探索过程 (Preece & Maloney-Krichmar, 2005)。网络调查也是一种经常使用的方法。对网络社区开展网络调查通常较为便捷：目标人群即社区参与者，可在其访问社区网站时直接触达；若条件允许，也可通过注册用户名单联络。对此，研究者已经开发了多种招募和抽样策略来进行调查。最常见的策略是在不同的目标区域使用弹出式窗口或横幅广告以发出调查邀请。在网站上，也可以发布带调查邀请的帖子（如在论坛上），或者在目标社交网站墙上发布带调查邀请的广告。通常情况下，这些都是非概率的、非基于名单的网络调查案例。在虚拟世界（如第二人生）中，还可能涉及更多细节问题，如先以虚拟形象接近应答对象（即化身），再发送调查邀请 (Hill & Dean, 2009)。此外，在线社区的创建者也可以对注册用户发出电子邮件邀请，从而实现基于概率名单的网络调查。同样地，某些类型的弹出式邀请或横幅广告也可以用于概率样本选择（即拦截调查）。

根据 Preece (2001) 提出的概念框架，与网络社区相关的研究可分为两类。第一类涉及社会性维度，依据 Preece 的定义，它是指支持网上社会互动的软件、政策和实践 (p. 349)。此类研究试图了解人们在网络社区中的行为、自我表达方式、参与动机、自我管理模式，以及为什么有些人倾向于“潜水”（仅浏览但不贡献内容），等等 (Preece & Maloney-Krichmar, 2005)。第二类与在线社区可用性层面有关（用户体验或 UX），主要关注用户与技术本身的相互作用。社会互动涉及某网络社区的成员如何通过支持性技术（此处是互联网）相互作用 (Preece, 2001, p. 349)。因此，网络社区的可用性关注个人学习使用技术的直观性和简易性，也关注与支持社区的在线技术进行互动的直观性和简易性。当然，除了这两类核心问题，针对在线社区的网络调查也可用于探究其他一般问题。

在线社区研究中，被引用最多的网络调查当属 Andrews、Nonnecke 和 Preece (2003) 的研究，该研究考察了异步网络社区中的潜水者。近年来，社交网站相关研究日益增多，例如 Chou (2012) 以学生为被试，通过网络调查研究了 Facebook 作为班级网站的整合情况；以及 Woolley & Peterson (2012) 利用网络调查研究了与健康相关的 Facebook 粉丝页对健康寻求的行动、观念和行为的影响。

1.3.6.7 网站评估

网站评估的目的是测量网站所有者或广告商感兴趣的访客或会员的特征，如社会人口统计学背景、产品 / 服务使用和生活方式选择。此类评估有时与网站访问量测量或网络分析相结合，为网站覆盖面提供行业标准。访问量测量通常是通过在参

与测量的网站全域植入代码,用自动化技术手段完成。其结果可以与网络调查数据相匹配,从而深入了解访问者的情况。当然,这一过程可能会出现严重的法律和保密问题:如受访者是否允许将其网络调查的答复与所访问网站的数据相匹配?他们对此是否有充分的了解?另一方面,对广告商而言,这类匹配数据在决定购买广告位时具有重要价值。

近期网站评估调查的案例包括:Elling、Lentz、de Jong 和 van den Bergh (2012)使用标准化网站评估问卷对政府网站进行的评估;Witt、Rowland 和 Wilkinson (2012)通过网络调查来分析慈善网站用户的意见和需求。

网络评估的重点在于获取与网站使用相关的可用性和诊断性信息,例如用户对特定页面或功能的使用方式与评价、对网站主要功能的理解程度等。部分网站有时使用分屏的可用性测试:受访者在—个屏幕上填写网络问卷,而另一个屏幕同步显示被评估网站的相应网页(与另一个屏幕上的网络调查导航保持同步)。为此,研究者开发了专门的调查工具,如软件可用性量表(Software Usability Measurement Inventory, 简称 SUMI, Kirakowski & Corbett, 1993)、网络分析与测量量表(Web Analysis and MeasureMent Inventory, 简称 WAMMI, Kirakowski, Claridge, & Whitehand, 1998)及系统可用性量表(System Usability Scale, 简称 SUS, Brooke, 1996)。

1.3.6.8 娱乐性民意调查、问卷调查和小测验

此类调查寻求互动,提供反馈和潜在的趣味性或娱乐性的结果,但其方法论层面比较简单。有时,它们可能提供有用的见解;然而更多情况下并没有实质性的科学价值,其结果也无法推广到整个人群。新闻网站频繁使用这类调查,其普遍性甚至在一定程度上影响了普通民众对网络调查的认知。对斯洛文尼亚新闻网站的分析显示,这类调查会严重损害正规网络调查的声誉,尤其是当问题措辞粗糙、格式质量低下时(Ličen, Lozar Manfreda, & Hlebec, 2006)。

1.3.7 与网络调查相关的实践

行政表格、互联网测试和在线心理实验与网络调查存在诸多共通之处。网络调查从业人员可以从这些相关实践中汲取经验,反之,网络调查的若干原则也对这些实践具有参考价值。以下将对部分相关实践进行介绍。

1.3.7.1 网络上的行政表格

网络上的行政表格并非正式调查，因为它的目的不是描述被观察人群（参考我们在第 1.1.1 节中对调查的定义），而是收集用于特定行政流程的信息。行政表格的实例较多：就业申请表、活动注册表、订单或购买表单、贷款申请表、税务表格、预订表格等。在一些国家，网络表格还被用于电子投票（Herrnson, Niemi, Hanmer, Bederson, Conrad, & Traugott, 2008）。这类网络行政表格，在可用性及问卷设计排版方面，往往会面临与网络调查类似的问题。

有时，网络行政表格与网络调查的界限较为模糊，这在很大程度上源于术语使用的不明确。最初使用的网络表格是一种特殊类型的网页，可以让受访者输入信息（第 1.3.1 节），但在可用性专业领域（Jarrett, Gaffney, & Krug, 2009），网络表格包括简短（通常为一页）的网络调查。同样，在网络调查软件的背景下，“表格”通常指简短的单页调查（例如事件评估表）。如前文在商业调查部分所述，为了增强通用性，“表格”一词常被用来替代“问卷”。

1.3.7.2 互联网测试

教育和心理学领域的测试与评估工具，常采用由个体作答的封闭式问题，其流程与调查极为相似。传统上，这类测试通过纸笔问卷进行。直到十年前，互联网心理测试工作组（Naglieri, Drasgow, Schmit, Handler, Prifitera, Margolis, & Velasquez, 2004）才意识到纸笔测试与互联网测试的成本差异。即便扩大测试规模，互联网测试的额外成本也较低（p. 151），因此具备推广价值。互联网测试的另一个优势在于互联网提供分散化的测试体验。

互联网测试服务于特定行政目的（如工作招聘、学生注册、运营商咨询、团队建设等），而不是像网络调查那样进行人口推断。不过，在方法论层面，其与网络调查有很多相似之处，特别是在管理模式、问卷设计和保密性问题上。

互联网测试存在四种可能的管理模式（Bartram, 2006），这也与网络调查的实施特点有相似之处。

无人监考的两类测试

- 开放模式：任何人都可以通过互联网参与测试，不需要验证应试者的身份，且无人监督。
- 受控模式：需要向应试者提供一个登录名和密码，但没有人监督。

有人监考的两类测试

- 监督模式：可进行一定程度的监督（未必现场监督，如远程监督）。测试管理人或监考人负责登录考生，核实测试完成情况，同时确保测试材料安全。
- 管理模式：一般需要使用实体考场，严格核验考生身份，并在考试期间监考以防止作弊。考场的计算机运行专用软件，屏蔽所有可能协助作弊的功能，如打开新浏览器窗口以使用搜索引擎、保存或打印测试内容，与考场外人员通信等。

网络上最常见的测试是无人监考的控制模式，也被称为无人监考的互联网测试（Unproctored Internet Testing, 简称 UIT; Tippins, Beaty, Drasgow, Gibson, Pearlman, Segall, & Shepherd, 2006），多用于新员工招聘。有研究显示（Beaty, Dawson, Fallaw, & Kantrowitz, 2009），三分之二的雇主将 UIT 用于招聘新员工。这种测试的优势在于，当候选人数量较多且分布分散时，是具有成本效益的解决方案。与许多就业办公室的传统测验相比，UIT 在管理、评分和记录方面的一致性更高；UIT 还允许那些无法在工作时间内到达就业办公室的候选人，自主安排测试时间，从而大幅加快招聘流程。然而，这种测试存在作弊风险，如曾经参加测试的其他人在考试期间协助作弊、替考和篡改数据库等，不过在有人监考期间作弊行为也会发生（Drasgow, Nye, Guo, & Tay, 2009）。

目前已经研发出多种减少或防止 UIT 作弊行为的方法：远程摄像头监控、调用应聘者网络摄像头、按键分析、浏览器锁定、统计学检测和两步测试程序。通过两步程序，以监督的方式进行第一次测试，合格者进入第二次测试。这些方法也适用于有证据表明或担心受访者可能使用假身份的网络调查。

相关文献指出，在线测试中可能出现的另一问题（同样与网络调查相关）是分心。测试的最佳做法是营造有助于发挥真实水平的环境，因为分心和中断会严重影响成绩（Tippins, 2009）。相应地，网络调查中受访者的多任务操作也会产生干扰和中断，进而影响作答质量。

综上，国际测试委员会（International Test Commission, 2006）的指导方针和《工业与组织心理学杂志》（*Journal of Industrial and Organizational Psychology*）特刊（Sackett, 2009）是研究互联网测试计算机化相关问题的重要切入点，而网络调查从业者也会面临相似问题。

1.3.7.3 心理学网络实验

本质上，心理学网络实验是实验室计算机化心理学实验的自然延伸（Reips,

2002b)。参与者通过网络浏览器访问在线计算机实验平台(实则为网络服务器),并像在实验室中一样使用显示器、鼠标和键盘完成交互实验。网络调查目的在于推断目标人群的特征,而在线心理实验是为了对因果关系进行探究。

在过去的几年中,网络实验的受欢迎程度持续上升(Reips, 2007)。原因之一是其效率高于传统实验室实验(Reips & Krantz, 2010)。网络实验能够快速测试大量被试,不仅容易招募到异质性特征的大样本,在许多情况下还可触达稀有人群。在时间、空间和人力成本方面,网络实验具有成本效益(Reips & Birnbaum, 2011)。

目前有数百个网站可以让参与者报名参加在线心理实验,例如由汉诺威学院心理学系⁸维护的互联网心理研究(Psychological Research on the Net),以及由德乌斯托大学制作的网络实验表⁹(Web Experiment List, Reips & Lengler, 2005)。专门用于网络实验的软件较少(Reips & Birnbaum, 2011),且现有软件中可免费使用的更是寥寥无几(Reips, 2007)。不过,标准网络调查软件(见第 5.3 节)正逐步支持随机化、时间戳等功能,因此也可用于开展网络实验。

该领域的知名学者认为,与实验室实验相比,网络实验的优势大于劣势(Reips, 2002a; Reips & Krantz, 2010),且劣势也有相应解决办法(Reips, 2002a)。特别是,当网络实验与前文提及的互联网测试、网络调查等领域出现交叉融合时,各领域的知识体系都将得到强化。

注 释

1 <http://www.greenbookblog.org/grit/>

2 <http://www.w3.org/>

3 http://www.cc.gatech.edu/gvu/user_surveys/survey-04-1995/

4 http://www.w3schools.com/browsers/browsers_stats.asp

5 Some web survey software can restrict the geography by eliminating or flagging the IP address of devices connecting from a non-eligible region.

6 <http://www.natcenweb.co.uk/genpopweb/>

7 <http://webdatanet.cbs.dk/>

8 <http://psych.hanover.edu/research/exponnet.html>

9 <http://www.wexlist.net/>

预调查阶段

第 1.2 节所定义的核心网络调查过程（图 1.1）始于预调查，涉及网络调查的具体准备活动。本章首先介绍预调查阶段的调查模式选择（第 2.1 节）、抽样（第 2.2 节）和问卷准备（第 2.3 节）环节，还将对技术议题（第 2.4 节）、无应答策略（第 2.5 节）和综合管理（第 2.6 节）进行阐述。

2.1 调查模式详述

第 1.1.1 节已经介绍了网络调查模式，在该模式中网络问卷属于测量阶段的组成部分。网络调查模式既可以独立使用，也可以与其他 CASIC（见第 1.3.4 节）或传统调查模式结合使用。确定采用何种调查模式是开展调查首要考虑的因素，甚至早在初步研究阶段就应确定。然而，选择调查模式的决策过程非常复杂，因为这涉及预调查的其他环节，尤其是取样和问卷编制工作。因此，本书建议读者，特别是新手，在通读完全书后再回过头研读本节。因为对调查模式的探讨不可避免地涉及后续章节的许多概念（如无应答、无覆盖、测量、数据质量、数据准备等）。

第 2.1.1 节首先讨论与确定网络调查模式相关的问题，第 2.1.2 节则围绕相应的混合模式背景进行探讨。

2.1.1 选择网络调查模式

如果为某个封闭式群体（如某协会的成员）设计一份简单的问卷，并且已经掌握该群体的电子邮件地址，那么选择网络调查模式是非常简单的决策。在另一种情况下，要做出充分合理的决策需要丰富的知识经验，以及对具体调查情境的深刻理解，包括对研究问题性质的把握。

在考虑网络调查模式时，参考下述问题可能会有帮助。本书将其整理为一份简洁、非穷尽且未结构化的清单。

- 网络调查是否适合目标人群？能否精准定位并联系到目标单位？是否具备合适的单位名单？目标人群是否具备填写网络调查问卷的能力（如有足够的技能）？目标人群是否有足够的积极性来正确填写网络问卷？
- 网络调查模式是否适合所调查的主题和要收集的数据类型？能否通过这种自我管理的模式测量目标概念？是否存在隐私和保密方面的隐患？问卷是否足够简洁以适用自我管理模式？填写过程是否需要访员的协助，如提供激励、澄清疑问或进行物理干预？
- 调查质量需要达到何种水平？是否需要采用概率抽样？对无应答和无覆盖的容忍度是多少？能否预判可能出现的无应答或无覆盖情况？
- 在资源和支持方面有什么要求和限制（涉及基础设施、软件、人员、行政、财务和管理等）？能否承担邮件邀请成本，还是只能负担电子邮件邀请？是否有能力采取应对无应答的转换措施（如提供奖励、发送提醒或转换调查模式）？
- 时间安排上有多大的灵活性？
- 基于网络的数据收集有哪些潜在的法律限制？
- 现有的网络调查软件在技术层面有哪些选择和限制？
- 是否存在其他影响调查模式选择的因素？

对上述问题的回答涉及用非常复杂的方式筛选合适的调查模式。第1.3.6节回顾了网络调查的不同类型、应用场景及相应案例，这些内容为解决模式选择的困境提供了基本依据。但要选择有效且恰当的调查模式，还需借鉴过去的类似研究项目，并遵守更广泛的方法学原则（详见第6.1节）。

应该明确指出，在上述问题中，本书主要探讨的是与测量阶段（即问卷填答环节）的调查模式直接相关的内容。然而，这些内容与招募阶段的联系模式（如调查邀请）紧密交织。联系模式可以借助多种沟通渠道，从网上的帖子、横幅、含问

卷的 URL 弹出式链接到电子邮件、邮寄信件或短信邀请及传单等形式，甚至可以通过电话或面对面交流（F2F）或公共媒体（印刷品、电视、广播）进行联系。然而，当提及网络调查的个人联系时，比较典型的是使用电子邮件或邮寄邀请和提醒（Dillman & Messer, 2010; Grady, 2010; Kaplowitz, Hadlock, & Levine, 2004; Porter & Whitcomb, 2007）。为了提高覆盖率与合作度，在招募阶段采用多种模式结合，即多重沟通渠道联用，是一种常见策略。下文将这种情况称为多模式联系（*multimode contacting*）。选择正确的联系方式或模式组合取决于多种因素，例如抽样框中联系信息的可用性、时间、预算和其他具体因素，包括各类质量要求，其中无应答问题通常起着关键的作用。第 2.5 节和第 3.1 节将进一步探讨联系方式，而下文将聚焦测量模式。

2.1.2 混合模式设置

网络调查正越来越多地与其他调查模式相结合。本节首先澄清相关概念和术语，继而提示相关问题，最后概述网络调查中最常见的混合模式设置。

2.1.2.1 混合模式调查、多模式联系和混合模式系统

第 1.2 节仅探讨了网络调查模式，将其作为独立的横断面调查中使用的单一调查模式。这种情况属于独立的网络调查（*stand-alone web survey*）。然而，网络调查模式也可以与其他调查模式相结合。尽管本书主要关注独立网络调查，但鉴于混合模式的重要性，在此也会对涉及多种调查模式的替代方案展开讨论。

在测量阶段多重调查模式结合使用时，就称为混合模式（*mixed-mode*）测量，或者是混合模式数据收集。本书将其简称为混合模式，对应的调查和设计则分别称为混合模式调查和混合模式设计。

在此重申，本书中混合调查模式的概念仅适用于测量阶段（如问卷填写），而不涵盖调查过程的其他环节（如招募）。因此，采用多模式联系（*multimode contacting*）——使用电子邮件、邮寄、电话和其他通信渠道招募潜在的受访者参与独立网络调查，并不意味着是混合调查模式，而只是采用多模式联系的独立网络调查。

若需使用一个更具通用性的术语，以涵盖调查过程中招募与测量阶段的不同模式组合所涉及的各类活动、阶段、决策及结果，本书将采用 Biemer 和 Lyberg（2003, p. 208）的混合模式系统（*mixed-mode systems*）这一概念。该概念不仅包含测量阶段使用多重模式的混合模式情况，还包括独立（网络）调查招募阶段的多模式联系，

以及（单一）联系模式与相应（单一）数据收集模式不同的情况，如采用邮寄邀请参与网络调查。

因此，本书使用混合模式系统这一术语作为涵盖测量和招募阶段模式组合的通用概念，并将混合模式的概念限定于测量阶段的模式组合。

2.1.2.2 混合模式调查的问题

在测量阶段采用混合模式时，必须审慎考虑所涉及的每种模式、实地调查的同步性、管理方式和模式组合。当涉及的模式增多时，数据库管理、归档、文件编制和无应答计算也变得更加复杂。这也增加了行政管理、方法学和管理任务所需的资源。为了使复杂的和存在潜在风险的调查问题（见第2.3节）适配不同的调查模式，尤其需要扎实的方法学知识做支撑。不同的模式可能产生特定的模式效应。一个典型的例子是网络调查模式中的首呈效应（*primacy effect*），即受访者倾向于选择列表起始位置的选项。与之相反的是近呈效应（*recency effect*）常发生在电话调查模式中，受访者更倾向于选择列表末尾的选项。还需要仔细考虑“不知道”选项的更恰当且具有可比较性的处理方式，和对某些问题的无应答处理，以及对可能寻求快速简化回答方式的受访者的应对策略，本书将在第2.3.4节详细讨论。

上述所有问题都需要额外的资源来阐述和测试，否则调查问题在不同的模式下可能无法等效匹配。很多时候，当调查模式转换时，问题需要进行相应修改。例如，当把欧洲价值观调查的问卷从面对面模式调整到网络模式时，Bennink、Moors 和 Gelissen（2013）报告说，需要对256个问题中的156个问题的措辞进行修改。同样，欧洲社会调查从计算机辅助个人面访（CAPI）转换为网络问卷时，85%的问题需要调整（Ainsaar et al., 2013）。即便如此，某些特定模式的测量效果可能仍然存在差异。例如，对于敏感话题（如艾滋病、毒品、健康问题；见 Tourangeau & Yan, 2007），受访者在网络模式下可能比在面对面模式下报告更多的风险相关行为，因此，对这类数据进行合理组合以用于估计和分析就变得极为困难。

因此，混合模式会引起严重的数据整合性难题（de Leeuw & Hox, 2011）。幸运的是，很多证据表明，仅就测量误差而言，在不同的调查模式中，结果往往是稳健的。多项研究显示，与传统模式相比，网络调查在有效性和可靠性方面表现相当或更好（Börkan, 2010; Chang & Krosnick, 2010; Mangunkusumo, Duisterhout, de Graaff, Maarsingh, de Koning & Raat, 2006; Miller, Neal, Roberts, Baer, Cressler, Metrik, & Alan, 2002）。然而，模式差异可能出现在问题主题、格式和特征的特定组合中（Bennink et al., 2013），并且在与社会赞许性相关的问题中（如吸烟、赌博、喝

酒) 尤为常见。因此, 只有当不存在与某些特定模式有关的测量效应 (即模式效应) 时, 混合模式的数据才能被安全地组合和分析 (de Leeuw & Hox, 2011)。遵循统一 (*unified*) (或一致模式) 设计的原则 (Dillman et al., 2009, p. 326) 可能有助于设计出在不同模式中保持稳健的问卷。然而, 为了确保模式同质性 (*mode equivalence*), 即不同模式下的结果具有可比性, 通常需要针对特定模式进行调整和整合 (de Leeuw & Hox, 2011)。

基于上述所有问题, 在混合模式设计中结合网络调查模式必须具备充分的理由。这些理由通常与降低成本、快速收集数据以及提升数据质量的具体方面 (如减少无覆盖和无应答) 或提供奖励的特定需求有关。

2.1.2.3 混合模式下的网络调查

已有多位研究者探讨过一般意义上的混合模式组合 (Berzelak, 2014; Bethlehem & Biffignandi, 2012; de Leeuw, 2005; de Leeuw & Hox, 2011; Dillman et al., 2009; Groves et al., 2009)。下面我们概述了测量阶段涉及网络调查模式的最常见的混合模式调查情况。

- 网络作为问卷填写的平行模式: 在数据收集的同一阶段, 不同模式可用于不同的抽样部分。这主要是用来减少网络调查的覆盖率问题, 同时在一定程度上降低无应答。例如, 将目标人群中能够上网的部分人分配至网络调查, 另一部分则采用邮寄调查 (Rookey, Hanway, & Dillman, 2008)。国际调查也存在类似情况, 互联网普及率高的国家或地区使用网络模式, 而其他地方采用其他模式 (Blyth, 2008; De Beuckelaer & Lievens, 2009)。在上述例子中, 网络部分是完全独立的。然而, 这些部分及相应的抽样框也可能重叠, 如双框架 (*dual frames*) 案例。一方面, 基于某个包含电子邮件地址的抽样框联系互联网用户; 另一方面, 将整个目标人群 (包括互联网用户) 也纳入基于邮寄地址的一般抽样框进行邮寄调查。
- 网络作为受访者自选的问卷填写平行模式: 在数据收集的同一阶段, 可以提供不同的模式, 让受访者选择自己偏好的方式, 而不是由研究设计预先指定。例如, 寄给潜在受访者带有纸质问卷的信件, 受访者可以寄回所附的纸质问卷, 或者拨打免费电话接受电话采访或通过网络完成调查。关于网络和邮寄的组合, 见 Medway & Fulton (2012); 关于网络和电话的组合, 见 ORC Macro (2003)。然而, 大量实证研究表明, 提供更多的模式供受访者选择会降低回复率 (Medway

& Fulton, 2012), 因为这使决策过程复杂化。宁愿通过调查设计预先指定模式的另一个原因是: 自我选择往往会干扰特定受访者的特征, 同时带来测量、无应答和无覆盖误差。

- 针对网络调查无应答者的顺序混合模式: 可以按顺序将网络调查中的无应答者分配到另一种问卷填写模式。使用这种方法主要是为了提高应答率和降低成本, 最典型的做法是, 先开展成本最低的网络调查, 之后再采用更具侵入性且成本更高的模式, 如电话模式 (Dillman et al, 2009; Greene, Speizer, & Wiitala, 2008)、邮寄模式 (P. D. Converse, Wolfe, Huang, & Oswald, 2008; Kroth, McPherson, Leverence, Pace, Daniels, Rhyne ... Prime Net Consortium, 2009; Millar & Dillman, 2011; Vehovar, Berzelak, Lozar Manfreda, & Belak, 2009), 或 F2F 模式 (McMorris, Petrie, Catalano, Fleming, Haggerty, & Abbott, 2009)。然而, 由于实施的复杂性和数据质量问题的增加, 可能难以达到预期的成本节约效果。
- 针对传统调查中无应答者的网络模式: 在此网络是对初始阶段未回复者的一种跟进模式。该方法仅在特定情况下使用, 即最初使用各种传统模式, 例如电话 (Greene et al., 2008)、面对面 (McMorris et al., 2009) 或邮寄 (P. D. Converse et al., 2008), 然后邀请无应答者参加网络调查。
- 在特定数据收集阶段中采用网络模式: 当有两个或多个独立的数据收集阶段——例如纵向研究中的不同阶段——受访者可能在不同的时间点接受采用不同模式的调查。这样做的原因通常与某种模式的成本或特殊的便利有关。例如, Nigg、Motl、Wong、Yoda、McCurdy 等人 (2009) 的一项研究, 最初使用的是电话调查, 而后续受访者可以选择在网上完成问卷。
- (概率) 在线样本库招募: 一个非常特殊但被广泛使用的混合模式是招募在线小组, 特别是概率小组。这通常是通过面对面访谈完成的, 如荷兰的 LISS 小组 (Scherpenzeel, 2011)。此外, 可以向家庭邮寄邀请包, 受访者可选择通过邮寄、网络或拨打免费电话来回复邀请, 如美国的 GfK 知识网络小组 (DiSogra, Callegaro, & Hendarwan, 2010)。在成功招募后, 使用电子邮件邀请受访者参加网络调查, 以完成每个子阶段的数据收集。非互联网家庭可以获得互联网设备、通过其他传统模式参与调查, 或者直接从样本中排除。

虽然上述所有方式都涉及整个问卷填写环节的混合模式数据收集, 但网络也可仅用于以下某一环节的部分流程。

- 网络用于数据收集环节的一部分: 出于隐私保护或成本节约等多种原因, 数据

收集环节的某些部分可以通过其他模式进行。例如，在面对面访谈中，受访者可以选择在网上回答敏感问题。同样，一些消费者满意度调查的部分内容可以通过电话进行（例如，对于重要或复杂的问题），而其余部分则在网上进行。

- 网络作为受访者模式转换的选择：原则上，在一次调查中，我们可以为受访者提供机会，在同一份问卷填写过程中特意在某些数据采集模式之间切换（例如从网络到计算机辅助电话调查再到网络）。这种情况很少见，且在技术上要求很高，但对于一些复杂的调查来说，仍是一个切实可行的选择。

在此，本书不再进一步讨论混合模式环境下各种调查模式的比较及具体内容。读者可以参考 Groves 等人（2009）的研究，他们介绍了调查模式的关键维度（如访谈者参与、与受访者互动、隐私问题、通信渠道和所使用的技术）及其对测量、抽样、覆盖面和成本的影响。Biemer 和 Lyberg（2003）系统地概述了各种模式的特征，而 de Leeuw 和 Hox（2011）、Bethlehem 和 Biffignandi（2012）在网络调查背景下探讨了这些问题。Dillman、Smyth 和 Christian（2014）也对混合模式进行了较好地阐述。

最后，混合模式给调查过程带来了极大的变化。因此，在调查过程的各个阶段可能会出现许多直接和间接的不利因素，尤其是管理和行政方面的复杂性增加，以及由于模式效应造成的测量误差增大。所以，本书建议采取极为谨慎的态度，并进行仔细考量。同时建议可能需要专家参与，因为与调查模式相关的决策可能是整个网络调查过程中最复杂和困难的部分。

遗憾的是，这些决策也是网络调查过程中最先要做出的。相应地，本书在预调查步骤的一开始就讨论了模式问题，后续还会涉及与模式相关的各种问题，特别是与无应答（第 2.5 节）和正式调查（第 3 章）有关的内容。这些其他方面的考量同样重要，在决定调查模式时也应纳入考虑范围。

第 6.1.5 节将进一步讨论混合模式的广泛背景，第 6.1.1 节将修订网络调查模式的定义，第 6.1.3 节将回顾模式效应评估的问题。

2.2 抽样

样本（*sample*）涉及目标人群中的一个单位子集，调查通过样本来了解整个目标群体的情况。

此处还需要介绍变量 (*variable*) 这一术语, 它是一个统计学概念, 指与研究感兴趣的某些特征相关且涉及目标人群中所有单位的量化指标 (如性别、年龄、某种态度)。变量通常用一个字母符号来表示。当讨论样本中的单位时, 使用小写字母 (例如: y, x, p), 当讨论总体中的单位时, 则使用大写字母 (例如: Y, X, P)。

变量在每个单位上都有取值 (*value*)。对于调查变量和样本中的应答单位来说, 取值是受访者的应答答案。另外, 单位可以直接分配其他变量 (如关于单位所在区域的外部信息)。

变量具有一定的统计学特征 (如最小值、最大值、平均值、方差), 可以通过某些计算得出。当基于样本中的单位进行计算时, 所得结果被称为抽样统计 (*sampling statistics*) 或估计 (*estimates*), 因为它们是对目标人群特征的估计, 被称为总体参数 (*parameters*)。例如, 样本中女性的百分比为 $p = 50\%$, 也可以表述为比例系数 (*pro-part*), 即 0.50。然后将其作为相应总体参数 P 的估计值, 该参数表示女性在人口中的比例。理想情况下, 样本是目标人群的缩小版, 能够反映目标人群的每一个特征 (Lohr, 2010, p. 3)。在这一背景下, 本书经常使用代表性样本 (*representative sample*) 的概念, 指目标人群中研究感兴趣的特征 (如社会人口统计学特征) 可以通过样本已知信息以一定精度推测出来。

预调查 (*pre-fielding*) 的抽样阶段, 首先考虑目标人群 (第 2.2.1 节)。接下来, 讨论抽样设计: 抽样框、抽样方法、样本量和抽样单位的选择。基于第 1.1.3 节对抽样方法的初步分类 (表 1.1), 第 2.2.2 节和第 2.2.3 节将分别讨论概率和非概率网络调查的抽样设计。除了抽样设计——这只是抽样环节的一部分——第 2.2.4 节还将讨论抽样的估计部分 (即统计推论), 通过统计推论可以将样本结果推广到整个目标人群。在此背景下, 本书还阐述了是否以及何时使用非概率网络调查而非概率调查这一难题。在抽样阶段, 第 2.2.5 节将讨论最终的无覆盖问题, 即目标人群的某些组成单位没有被纳入调查 (如非互联网用户)。

上述诸多问题在很大程度上取决于研究目的, 且可能在初步研究活动中就已明确, 尤其是目标人群的定义和是否采用概率抽样的决定。

与其他预调查阶段类似, 本书的讨论仅限于独立的网络调查, 暂不涉及混合模式系统的复杂性; 主要关注网络调查的具体方面, 但也在一定程度上涵盖了基本的一般抽样原则, 以便对抽样问题进行独立阐述。

2.2.1 界定人群

预研究中，通常在调查的初始问题里列出单位、目标人群以及目标变量。同其他调查一样，网络调查的目标人群范围广泛，可以从学生到普通大众，以及各种特定人群，如外来动物的主人。第 1.3.6 节已经列举了很多案例。

与一般研究目标中较为宽泛的概念概述相比，为满足取样需求，需要更精准地描述目标人群。例如，在学生评价调查中，根据正式的入学标准或出勤率，学生存在多种状况，这就需要精确界定哪些单位属于目标范畴。这些被界定的单位是合格单位 (*eligible units*)，并与不合格单位 (*ineligible units*) 区分开来，后者不属于目标人群，但可能由于各种原因仍然出现在样本中（例如，某人已去世却未从名单中删除）。在学生评价的案例中，如果只关注国内学生，那么外国学生就不符合条件。这些多余的不符合条件的单位，有时被称为过度覆盖 (*overcoverage*)，只需要忽略即可。

出于实际考虑，我们常常会聚焦范围更狭窄的调查人群 (*survey population*)，该人群比理论上定义的目标人群更符合调查的实际情况。也就是说，将目标人群中一些规模可以忽略不计的极少部分纳入调查所需资源，可能与其带来的附加价值不匹配。例如，在学生评价调查中，可能仅调查在校学生，而排除缺席的学生。同样，在一般人口调查（如选举研究）中，我们通常会忽略某些机构中的人（如监狱服刑人员、护理院住户、军人等）。

当然，对待这些简化措施需要谨慎。如果由于某种原因，部分合格单位无法参与调查，比如在对普通人群的网络调查中，那些不使用互联网的人。第 2.2.5 节会探讨无覆盖的问题。

在网络调查中，数据收集成本较低这一特点可能促使研究者对整个人群进行调查（即普查）。特别是对于规模较小的目标人群，普查是一种便捷的解决方案，因为它省掉了所有与抽样设计和统计推论有关的程序。然而，关于目标人群和调查人群定义的讨论也至关重要。

2.2.2 基于概率的网络调查中的抽样设计原则

上述关于目标人群的讨论适用于所有网络调查，而以下将单独探讨抽样设计问题，包括概率抽样和非概率抽样。本节从概率抽样设置展开，主要聚焦基于名单的样本，因其比非名单样本的使用更为频繁。非名单样本较为罕见，第 1.1.3 节曾简

要介绍，同时也提及了概率网络拦截调查的案例。然而，就抽样设计原则而言，与基于名单的概率调查所确立的概念一致，因此此处不另加论述；拦截调查的具体实践可以参照 Poynter (2010)。

本节回顾了基于名单的概率网络调查的关键抽样设计原则：即抽样框、抽样技术和样本量。在很大程度上，这些问题并非专门针对网络调查，而适用于任何调查模式。但由于理解这些知识对网络调查至关重要，本节将进行较为广泛的讨论。

下面介绍的基于概率的样本抽样设计原则，也经常被应用于非概率情境，但这可能会带来各种风险和未知影响。因此，本节会偶尔对此问题做出具体评论，而第 2.2.3 节将专门详细介绍概率抽样原则在非概率环境中的实施。

2.2.2.1 抽样框

无论是调查整个人群还是只调查一个样本，基于名单的概率网络调查都需要抽样框。抽样框是一个基本的元素清单，使研究者能够确定和联系调查人群的合格单位。此处讨论普遍基于名单的情形，即抽样框是事先准备好的，可以从一些商业文件（如客户名单）或行政文件（如大学登记册）中获取，或通过初步程序（如网络招募）来准备。

理想状态下，抽样框提供相关的联系信息（电子邮件或邮寄地址、电话号码），并在合格单位和抽样框的元素之间实现一对一匹配，能够准确反映调查人群。例如，用选民的人口登记册来调查选民群体，或者用大学注册办公室的所有注册学生名单来调查学生，都非常接近于这种情况。

一旦获得或构建了作为抽样框的名单，就应当对名单质量进行评估。涉及的初步问题如下：

- 符合条件的单位中，缺失（即无覆盖）的比例是多少？这些单位的比例和特征可能涉及与抽样框有关的最基本和最关键的问题。这一内容将在第 2.2.5 节关于无覆盖的部分进一步阐述。
- 名单上有多少重复条目？如果是单一表单（如电子邮件地址），重复条目很容易被识别并被删除。然而，当同一单位在电子邮件名单中以不同条目出现时，识别就会变得困难。一些外部信息、专家评估、小规模初步调查或定性调查可能对此有所帮助。有时，只能调查后才能估计其重复比例，常用加权的方法来补偿，这是一种事后处理程序（见第 4.1.4 节）。加权是给不同单位分配不同

权重,以补偿纳入概率差异或提高样本的代表性。例如,在抽样框中,一个有多个条目(如两个电子邮件地址)的受访者应该得到与条目数成反比的权重(如权重为 $1/2$)。

- 名单中是否存在单位集群的条目?在这种情况下,应该将整个集群纳入。如果只纳入一个单位,需要用加权来补偿——相应的单位得到的权重与集群的大小成正比。如在网络购买行为的调查中,如果只有客户名单,而不是购买记录名单,一位有多次购买经历的客户便形成了一个集群。如果以相同概率选择顾客,那么通常会进行二次抽样,选取一个购买事件,并赋予该单元相应的权重(例如,有5次购买经历的顾客对应的权重与之成正比)。
- 名单的时效性如何?抽样框中过期的条目会增加无覆盖问题,并在数据收集中增加不确定性,导致数据收集效率较低。因此,应该清楚地记录最近一次刷新名单的时间(所谓的参考日期),以便准确估计其质量。
- 有多大比例的单位缺失联系信息或联系信息不完整?根据联系策略,可能需要多种类型的联系信息,如电子邮件地址、电话号码、姓名等。可以通过检查子样本的单位,来粗略检查和估计这些条目的相应质量。
- 不符合条件的单位,其百分比是多少?如果有可能提前识别,就应该从抽样框中剔除,但有时只有在实际接触时才能确定该单位是否符合条件。无论如何,估算不合格单元的比例很重要,这样就可以适当增加初始样本。对于那些不符合条件的单位,只需忽略不计。
- 可支配的额外信息有哪些?拥有其他信息能大幅提高抽样框的质量。姓名和头衔(先生、女士、博士)可用于个性化沟通,而实质性变量(如年龄、性别、教育)可用于控制并提高抽样设计(如分层)和调整(如加权)的效率。

此外,本书在下文提供了三种主要方法(电子邮件名单、传统抽样框和在线样本库),用于获取或建立基于概率的网络调查抽样框。原则上,这些准则也适用于非概率样本,所以在后面讨论非概率样本时不再重复。

2.2.2.1.1 以电子邮件名单作为抽样框

在网络调查中,特别是针对互联网覆盖率较高的人群(如学生、雇员、会员、客户、企业等)时,电子邮件地址名单(可能还包括其他信息)是能够符合预期和期望的抽样框。

电子邮件名单可能已经因为某些管理、商业或其他原因而早已存在,如果没有法律、伦理或技术限制,我们可以直接使用。然而,如今越来越多的时候需要获得明确许可,或者与建立某个名单(或登记册)有关的法律条文需要正式允许将该名单用于特定调查。

有时表单中只有名字,没有对应的电子邮件地址,所以我们不得不在网上借助各种搜索策略来获取电子邮件地址。当然,这样的搜索可能耗费大量时间,且有时可能无法确定匹配的电子邮件地址是否准确,甚至可能根本找不到。

还可以通过使用网络上的信息从头开始编制电子邮件名单。例如,如果想调查某所大学的学者,我们可以收集他们公开的电子邮件地址。但是,在一些国家拥有公开的电子邮件地址并不意味着研究者可以用于发送调查邀请。这与公开的电话号码和邮寄地址有很大不同,研究者合法使用后者进行传统的调查招募是被准许的。此外,还可以从网络上手动或通过自动程序提取公开的电子邮件地址,并用于网络调查。在网络调查早期,这些简单的“获取”方法是被接受的(Swoboda, Mühlberger, Weitkunat, & Schneeweiß, 1997)。但是,现今在大多数发达国家,由于存在各种法律限制以及调查行业的伦理准则,上述简单获取的方式不再被准许。这种“不请自来”的策略有时仍然用于获取一些商业电子邮件地址,原则上这些地址对任何营销、商业和研究目的都是开放的;但不能用于获取私人的电子邮件地址。同样,当受访者是某些非常特殊的目标人群时,如果这些目标人群的名单可以在网上找到(例如各种专业人员、业余爱好者、从业人员、特殊物品的所有者、植物或动物的所有者等),非商业性的调查邀请可以在符合条件的情况下使用公开的私人电子邮件,但在商业背景或受访者为普通人群时通常不适用。然而,在许多国家,这一点正受到越来越多的限制。

除非某项调查有法律规定,否则原则上需要征得潜在受访者的同意,才能使用其电子邮件地址发送调查邀请。现在,注册购买产品或服务越来越需要受访者本人同意(例如,明确勾选“同意接收后续客户满意度调查”选项),或者在申请加入某些协会、俱乐部或社区时,也需要本人同意。

一些国家(如斯洛文尼亚)的立法尤为严格。在这些国家,创建数据库并存储个人数据(如电子邮件地址)需要有正式的法律依据,或获取名单上个体的明确同意。如果未经同意而创建,相应的信息专员可以处以巨额罚款。

当然也有例外,使用关系密切的社会关系名单则不需要事先获得同意,如属于某个受访者圈子的群体。因此,自助式研究通常可以邀请研究者的家人、同事和朋友,以及社交媒体的“朋友”加入调查。

使用电子邮件名单进行邀请时，额外信息对于组合各种联系策略非常有用（例如，邮寄预通知、短信提醒）。其他一些信息（如性别、年龄、地区、受教育程度）对于优化样本选择也很重要。然而，电子邮件名单往往缺乏相关额外信息，因为其建立初衷并非用于调查。

鉴于限制越来越多，为网络调查获取电子邮件地址变得愈发困难。因此，各种专业公司出售电子邮件地址也就不足为奇了。虽然这些地址的价格和质量各不相同，但可以合法地用于调查，因为已经征得了这些地址所有者的同意。然而在大多数情况下，这些电子邮件名单并不是基于概率的样本。

有时，在获得、购买或建立大型的电子邮件地址数据集后，每次调查都会从该数据集中抽取一个单独的样本。在概率样本设置中，这种方法被置于两阶段抽样背景下讨论（Lohr, 2010）。

2.2.2.2.2 传统抽样框

当电子邮件地址的抽样框不存在或无法有效构建时，研究者需要依靠传统调查模式（面对面、电话、邮寄）的抽样框。这可能适用于各种特定人群（学生、客户、雇员、会员）和普通人群的情况。典型案例是，通过面对面招募到基于概率的在线样本库（Scherpenzeel, 2011），其中电子邮件地址是在面对面访谈中获得的。受访者连续（例如每月）参与网络调查可以补偿面对面访谈的高成本。使用电话号码的抽样框短期招募电话访谈，通过访谈获得电子邮件地址，然后通过电子邮件邀请受访者进行较长的网络问卷调查，这样可以节约成本（Bethlehem & Biffignandi, 2012, p. 241）。但遗憾的是，这种方式可能被证明成本较高，因为网络和电话阶段的累积回复率很低。

在网络调查招募中，最常见和最方便的抽样框选择是使用电子邮件地址。因此，相应的邮件邀请会包含网络问卷的 URL，受访者需要在浏览器中输入该 URL。采用正确的方法（如激励措施、多次接触、混合模式），可以获得可接受的成本和合理的回应率（Dillman et al., 2009, p. 235; Tijdens, 2014; Vehovar et al., 2009）。

上述传统抽样框除了成本较高，还可能存在其自身的问题，但在此不做讨论。读者可参考 Lohr（2010）和 Groves 等人（2009）的文章，了解调查研究中关于抽样框的一般讨论。

2.2.2.2.3 单位样本库

网络调查的另一个替代抽样框是在线样本库。这是一类规模庞大、信息丰富的

单位数据库, 每次可从中为不同的调查项目选择特定的样本。通常在招募阶段就收集相当多的额外信息, 可供选择时使用。由于可以反复访问相同的单位 (如每月), 所以称之为样本库。样本库中的单位可以来自特定的目标人群 (学生、消费者、雇员), 也可以来自一般的家庭或个人。前文案例中提到过基于概率的样本库。然而, 如今大多数在线样本库都是非概率型的, 第 5.2 节会进一步讨论相关内容。

2.2.2.2 抽样技术

调查研究通常研究样本 (*samples*), 即目标人群的一部分, 因为研究目标往往可以通过比目标人群小得多的一定数量的单位来实现。然而, 直到 20 世纪 30 年代相应的抽样理论得以应用, 现代调查的发展才出现重大突破。该理论的基础是随机化 (*randomization*), 指研究者通过随机化过程控制样本单位的选择, 在概念层面类似于抽签。当从包含 N 个单位的目标人群中抽取 n 个单位为样本时, 通常需要事先知道所有单位的被选概率 (即概率抽样)。这种方法可以节约巨大的成本, 因为通过恰当选择的几百个样本单位就可以准确地测量上亿目标人群的特征 (如特定态度)。但抽样也存在代价, 有可能选择到一个不理想的样本。因为存在许多潜在样本, 且其质量随机波动。通常采用抽样误差 (*sampling errors*) 来量化这种风险, 并用置信区间表示, 该区间与真实平均值和重复抽样可能获得的估计值之间的差异有关。

本书假定读者熟悉推断统计学的基础知识 (例如, 总体方差、样本方差、置信区间、标准误 (SE)、标准差 (SD)、显著性水平等), 因此在此将不再进一步阐述。但考虑到缺乏这些知识的研究人员也会进行网络调查, 我们在 WebSM 平台上概述了网络调查实践中的基本统计知识和抽样原则 (*Basic statistics and Sampling aspects in web survey practice*)、案例和实践经验法则, 作为第 2 章的补充 (<http://websm.org/ch2>)。

可以运用丰富的抽样技术来选择样本。对于网络调查和电子邮件招募, 简单随机抽样 (*simple random sample*, SRS) 通常占有优势, 它完全类似于抽签, 目标人群中的所有单位都有相同的概率被抽到。我们尤其推荐其近似方法系统抽样 (*systematic sampling*), 在系统抽样中首先随机确定一个起点, 然后按照固定的抽样间隔来选择单位。除了简单易行, 其优点是有可能进行隐性分层。为此, 首先将抽样框中的单位按重要控制变量 (如地区、年龄、性别) 进行分类; 当实施抽样间隔时, 所选单位会更均衡地分布在抽样框的所有部分 (阶层)。根据地区进行分层是较为常见的做法。

有时需要比简单随机抽样复杂的设计, 特别是涉及成本问题或混合模式时, 这

意味着要考虑使用显性分层、多阶段抽样、群组抽样等。在此我们不进一步讨论这些设计,因为它们并非专门针对概率网络调查(Bethlehem & Biffignandi, 2012),而建议读者参考一般调查研究(Groves et al., 2009)或抽样方法文献(Lohr, 2010)。

抽样是调查研究中一个高度专业化的组成部分。因此,面对复杂的抽样设计以及其他复杂的情况,要作出专业且可靠的判断需要丰富的培训和经验;在这种情况下,建议听取专家的意见。

2.2.2.3 样本量

样本量是实现研究目标(如对目标人群进行实质性分析),以及确保达到数据质量标准(如精确度)的基本要素。在传统调查模式中,成本的节约往往限制了样本量,但在网络调查中情况却并非如此。因此,研究人员有时会调查大量的网络样本,仅仅因为可以负担,或者因为他们错误地认为增加样本量可以弥补抽样框的质量不足和其他抽样缺陷,这在网络调查的早期经常发生(Lozar Manfreda, 2001)。本书强烈建议不要使用不必要的大样本,因为大量网络问卷会造成过度调查,长远来看可能会降低受访者参与网络调查的普遍意愿;而且不必要的大样本也违反了职业道德(Dillman et al., 2009, p. 358)。

确定样本大小是非常重要的,有时也是基于概率的网络调查以及非概率调查中面临的唯一取样难题。不幸的是,各种不适当的经验法则经常被使用,导致样本过小或过大。

在关于抽样的文献中,确定样本量的基本方法是基于相关置信区间(*confidence intervals*)所反映的所需绝对精度(*precision*)。举例来说,对于1-5级量表的变量,若采用简单随机抽样,样本量 $n = 100$,典型样本平均值 $y = 3.0$,相应的标准差 $SD(Y) = 1.0$,标准差 $SE(\bar{y}) = 0.1$,可以得到95%的置信区间为 $[3.0 \pm 0.2]$,这可能已经足以满足需要;但有时可能只需要一个非常粗略的方向,甚至 $n = 30$ 也可以,此时置信区间为 $[3.0 \pm 0.4]$ 。而另一个极端,如果 $n = 1\,000$,相应的区间为 $[3.0 \pm 0.06]$,则可能是不必要的,超出了实际需要的精度。因此,仔细评估研究需求是极其重要的。

当然,一些其他变量类型(如份额)可能需要更大的样本。这种方法的进一步细化在于使用相对精度和变异系数(*coefficient of variation*),我们将在第2章补充资料“网络调查实践中的抽样”(Sampling aspects in web survey practice)中进一步阐述(<http://websm.org/ch2>)。

上述方法只介绍了与整体样本估计有关的样本量。然而,通常还需要对领域(如地区)和子群体(如年龄—性别分类)进行估计,因此事先明确这些非常重要——

因为随后需要在每个子群体中达到一定的样本量。通常情况下,最小值被设定为20~30个单位,但这在很大程度上取决于所需的精度水平和变量性质。对于份额和重要性的估计,可能需要 $n' = 100$ 个单位或更多,而 $n' = 10$ 在实践中通常被视为绝对最小值。随着对更详细分析的要求(即精度的提高和子群体数量的扩大),样本量迅速增加。例如,如果把子群体的最小值设定为 $n' = 20$,包含2个性别类别、3个年龄组别和10个地区,就组成了60个单元, $n = 2 \times 3 \times 10 \times 20 = 1\,200$ 。

研究问题不同,所需的样本量也不同。在网络调查中,我们很容易对此进行优化。例如,在100个问题中,我们可以让所有单元都回答10个核心问题,并将剩余的90个问题拆分为两组,每组45个问题,随机分配给一半的样本(例如,根据单元的奇偶编号进行分配)。假设对于这90个问题, $n = 500$ 个单位的样本就足够了。这样一来,无需1000个单位回答100个问题,每人花费20分钟;而是1000个单位回答55个问题,每人花费12分钟。或者,如果让所有单位保持相同的20分钟作答时间,则可纳入更多问题,例如180个而不是100个,且不会延长作答时间。当然,这种拆分方式可能要复杂得多。尽管在本书看来,该方法是进行有效网络调查的关键建议之一;但在任何情况下,这种方法往往被遗忘并且用得较少。第2.3.3节问卷准备阶段将对此作进一步讨论。

无论确定哪种样本量,研究者应该根据预期的无应答率 and 不合格单位的比例加以考量。例如,如果估计有30%的单位会因电子邮件地址无效而不合格,且预期应答率(由于拒绝和无联系)为20%左右,那么若希望最终得到300个合格受访者,则需要邀请 $300 / (0.7 \times 0.2) = 2\,143$ 个单位。

一些网络调查软件提供了简单的样本量计算方法,以及在满足某些条件后结束调查的终止规则。这些规则通常与估计精确度有关,或者与样本结构有关(例如,需要50名30岁以下的女性等)。对于复杂样本、混合模式系统和序列样本选择,样本量的优化可能需要某些涉及成本和误差的模型(见第6.1.4节)。

最后的抽样工作与样本选择过程有关,即将上述讨论的抽样设计原则应用于抽样框,选出实际的样本单位用于正式调查。从技术角度看,样本选取通常是由一些外部统计软件完成,但网络调查软件也越来越多地提供相关的数据库管理功能(如排序、合并、连接),以及相应的样本选择程序(如SRS)。

在样本准备阶段,我们还可以准备额外的样本,也叫补充样本(*supplement sample*),以备在正式调查环节低应答率时的扩大样本之需。补充样本应该是独立的,并按照与主样本相同的原则进行选择。在这种情况下,事先准备好样本是非常重要的,而不是在正式调查过程中临时拼凑。

本书不再进一步阐述样本量的问题，因为网络调查在这方面并无太多特殊性，如需深入研究，可参考标准教科书（Lohr, 2010）。然而，如前所述，由于这个问题在实际调查中的重要性，我们在第2章的补充资料（<http://websm.org/ch2>）“网络调查实践中的抽样问题”中为新手提供了一些额外的指导方针和实际例子。

2.2.3 非概率网络调查中的抽样设计

在非概率调查情境中，样本单位的纳入概率是未知的。这无疑会带来诸多问题，因为前文所探讨的所有抽样原则（如随机化、抽样框、样本选择、样本量确定等）都是基于概率抽样背景制定的。概率网络调查有着悠久的历史，积累了丰富的理论和应用知识，正如上一节及其中提及的任何一本教科书所概述的那样。

本节首先回顾了非名单样本的招募程序，这类样本在非概率环境中比较普遍。相应的招募过程取代了基于名单的样本选择过程，而在基于名单的样本中，样本选择是通过从抽样框中选择来完成的。关于基于名单的非概率样本，在第2.2.2节最初讨论抽样框时已有所涉及，由于非概率背景没有增添新的内容，所以不做单独讨论。

基于概率的抽样原则，也可以用于非概率样本的抽样设计，因此本节仅探讨相应的实施问题。就规范角度而言，在不知道选择概率的情况下，抽样设计有许多隐含的假设，这些假设通常难以甚至无法检验，因此研究中对其进行检验（例如验证未知的选择机制是否随机）。尽管如此，当概率抽样设计原则用于非概率抽样时，仍然可以保留概率抽样的许多优点。下文将系统介绍相应的实践做法、近似方法和建议；第2.2.4节将阐述统计推论部分。

2.2.3.1 非基于名单的网络调查的招募

在非基于名单的网络调查中，样本选择并不是一个单独的阶段，而是与招募过程合并在一起。因此，受访者样本的构建与招募过程同步进行。下面简要介绍非基于名单、非概率式网络调查的主要招募方法。

调查邀请可以在网上发布，也可以通过线下发布，如印制传单或借助传统媒体（电视、广播、印刷品、广告牌等）投放各种邀请或广告。这些类型的网络调查属于无限制的网络调查（第1.1.3节），因为任何注意到邀请的人都可以参与。这种方法与方便取样（*convenience sampling*）的概念密切相关。在方便抽样中，被试因其易获取性和接近性而被选中，而不是通过适用于目标人群的随机化方法。因此，相

应的抽样选择机制不受研究者控制,而且通常是未知的。虽然研究者希望方便取样能够存在一定随机性,但实际情况很少如此。相反,它通常反映了主题的突出性和其他各种未知的驱动因素。

非名单型、非概率型网络调查的招募主要是通过网络进行,采用多种在线招募方法。最常见的方法列举如下。

- 最简单的方法是使用邀请文本,如“点击我”或“参加调查”,将其嵌入网页(如在线论坛),并附上网络问卷的 URL。
- 各种类型的横幅广告也特别常见,可视为一种展示性广告。研究人员在预期大量目标人群会浏览的特定网站上购买空间进行展示(例如,针对烹饪爱好者的调查将在与烹饪和食谱分享有关的网站上进行宣传)。
- 有时邀请是以弹出式“窗口”的形式出现,即调查邀请弹出在受访者的屏幕上。还可能包括图形动画和各种多媒体文件(如声音、视频)。这种方法是侵入性地拦截用户,所以通常被称为拦截调查(*intercept survey*),它与线下商场拦截类似,即我们在消费者购物时对其进行提问。弹出窗口通常只覆盖屏幕的一部分,但也可能是全屏。
- 弹出式方法的一种变式是浮动横幅广告,这类广告以不同程度的干扰形式出现在活动浏览器标签页的屏幕上,如果用户没有任何操作,就会在预定时间后消失或重新放置在网站的某个角落。
- 极端情况下,弹出式窗口会阻碍用户访问计划浏览网页的前期内容。用户首先被要求甚至被迫回应网络调查,之后才可以访问其他内容。这种方法也被称为中断浏览(*interrupting traffic*)(Henning, 2013, p. 30),是用访问者的时间(即填写网络调查)换取高端内容访问权的一种形式。类似地,在学生评价网络调查中,有时会要求学生先填写课程评价网络调查,否则不允许申请考试。由于受访者被迫参加调查,他们可能会认为这是不必要的负担,从而不会提供最佳答案。
- 搜索引擎广告也是付费广告的一种形式,但具有一定特殊性。其目的是利用个人搜索信息来吸引用户参加调查,方法是将邀请函放在点击键的顶部,或者用一个通常出现在屏幕右边的横幅广告。例如,在搜索引擎中输入“狗”这一搜索词的用户可能会看到付费广告,邀请他们参加有关狗的问题(食物、保护、立法等)的调查。搜索引擎广告对于调查难以接触到的人群特别有用。例如, Nunan 和 Knox (2011) 对特定的关键词进行投标,为英国私营医疗服务机构的消费者参与调查寻找受访者。

上述方法本质上是邀请潜在受访者点击链接,从而将其引导至网络问卷的第1页。然而,与收到此类邀请的访客数量相比,实际的点击量通常非常少。相应的比率,也称为点击率(*click-through rate*)(CTR),通常以百分比的小数形式衡量。当然,更具侵入性的方法(如弹出式广告)效果相对更好。涉及付费时,通常是按照每次点击邀请(如横幅广告、链接、弹出窗口)进行收费。每次点击的价格通常在几(几十)美分左右。然而,由于点击邀请并进入问卷第1页的用户不一定会继续参与问卷调查,因此通常一份完整的问卷收费高达数(10)欧元/美元的价格,但这仍然是合理的成本效益选择。

在线招募方法可用于一次性横断面调查、在线样本库招募,也可作为在线样本库的补充。在后一种情况下,我们假设并希望它们能提高代表性(R. P. Baker, Brick, Bates, Battaglia, Couper, Dever, ... Tourangeau, 2013)、增加多样性(Henning, 2013),并获得与现有小组成员相比“更新鲜”的受访者,即那些不太习惯接受在线调查的受访者(Poynter, 2010, p. 9)。

这些在线招募方法可以相互结合,也可以与其他方法结合,特别是与电子邮件邀请、线下推广和营销策略结合,形成各种组合和复杂的实施方式。我们在此介绍几个最常见的例子,即众包(*Crowdsourcing*)、在线社交媒体招募(*Online social media*)、网络抽样(*network sampling*)和游击队式招募(*guerrilla recruitment*)。

众包是一种从大量人群中征集服务的通用方式。该术语主要用于互联网社区,在那里用户愿意有偿执行任务(MTurk、Craigslist)。由于其可用性较高且成本低,众包也成为招募有偿参与调查或心理实验的流行方法(Buhrmester, Kwang, & Gosling, 2011)。

在线社交媒体为招募提供了新的、具体的潜力,因为用户越来越多地参与在线论坛、社区、博客、社交网站(SNS)、视频/照片分享服务、多人/团体交流、虚拟世界和其他合作平台。一个重要的优势是可以从用户资料中获得辅助信息,如兴趣小组的成员、帖子和他们之间的链接。当然,这些信息属于“公共社交媒体”的一部分(ESOMAR, 2011b)。研究者通过搜索个人资料、帖子和网络联系,可以用相对较少的资源有效找到范围较狭窄的目标人群(Wejnert & Heckathorn, 2008),并向其发送个人邀请。当然,明确且可用的抽样框通常不存在,而且所寻求的电子邮件地址通常不会公开发布。即使电子邮件地址是公开的,道德和法律法规也会限制将它们用于调查。因此,在线社交媒体中抽样和招募的主要方法是直接邀请,可以由研究人员(如发布消息)、网站管理部门(如管理员向所有注册会员发送邀请)或通过广告来完成。这些方式在根据用户资料(如年龄、性别、地点)进行

的精准目标定位方面具有较大优势。例如, Fenner、Garland、Moore、Jayasinghe、Fletcher、Tabrizi……Wark (2012) 成功在 Facebook 上使用这种有针对性的广告, 对澳大利亚某地区的年轻女性开展了健康研究调查。另一个成功的策略是在某些社交媒体网站内形成特殊的用户群, 邀请并让该用户群参与研究, 同时在社交媒体领域内 (即激励用户邀请他们的朋友) 和跨网站推广 (即发布加入该群的邀请)。Brickman Bhutta (2012) 建立一个 Facebook 小组来研究受洗天主教徒的状况, 结果显示参与者在基本人口统计学方面与一般人群不同, 但在相关分析和深入分析中得出了与概率样本非常相似的结果。

网络抽样方法利用网络个体之间的社会联系来定位和增加样本单位。这种方法尤其适用于具有罕见特征 (如罕见疾病和罕见行为) 的人群。网络抽样有多种方法, 例如标准的滚雪球抽样 (Baltar & Brunet, 2012), 即研究人员从受访者那里获得一份网络联系人名单, 然后在这个名单中进行招募。还有受访者驱动的抽样 (*respondent-driven sampling*, RDS) 则是一项重要改进 (Heckathorn, 1997)。WebRDS (Schonlau & Liebau, 2012; Wejnert & Heckathorn, 2008) 将这一方法扩展到网络情境中, 由受访者而不是研究者通过传递优惠券等方式来招募更多的受访者, 所有这些都可以在很大程度上缓解保密性担忧。如果仔细操作, 即在后续步骤中适当选择随机种子并进行控制, 这种方法也可以被视为一种概率抽样。

游击式招募指采用侵入性或有创意的方式进行在线招募。一方面, 这包括在各种新闻网站、一系列在线社交媒体网站和其他合适的虚拟场所广泛且系统地发布邀请, 通常需要雇用专人从事此工作。这对于某些研究目的来说可能是适合的, 例如在网络论坛中招募被试参与有关网络仇恨言论的调查 (Vehovar, Motl, Mihelič, Berčič, & Petrovčič, 2012)。但如果在烹饪论坛上发布有关野生动物的调查, 就可能会出现问題。另一方面, 病毒式营销方法 (*viral marketing approaches*) 也可用于此; 通常是基于有吸引力的多媒体内容, 其中包括调查邀请 (以及一些激励措施), 由用户自身通过电子邮件或社交媒体在互联网上复制传播。

关于在线招募, 有时也会出现实时抽样 (*real-time sampling*) 或河流抽样 (*river sampling*) 的概念。后者被定义为基于选择的网络抽样方法, 使用各种吸引注意力的技术, 如弹出框、超链接和横幅广告 (Baker et al., 2013, p. 18)。这一定义与网络招募的概念大致重合, 但有时并不确定它是否真正涵盖上述所有方法 (例如, 中断浏览的所有变式或所有类型的游击式招募)。

众多案例研究表明, 选择合适的招募方式在很大程度上取决于具体情况 (如文化、背景、研究问题、目标人群)。尽管如此, 近年来社交媒体招募往往表现出相

当的优势 (Antoun, Zhang, Conrad, & Schober, 2013; Wilkerson, Shenk, Grey, Rosser, & Noor, 2013)。然而, 这种优势可能会随着时间的推移而改变甚至消失, 因为网络服务的未来发展可能会为在线招募提供新的机会。

至此, 本节介绍了非名单型、非概率型样本招募, 接下来继续阐述在非概率情境下实施概率型抽样设计的一般原则。

2.2.3.2 非概率情境下抽样设计的近似方法

当采用上述招募方法时, 非概率样本既可以是非基于名单的, 也可以是基于名单的。当使用不控制和不提供纳入概率的程序 (如非概率在线样本库) 建立名单时, 基于名单的非概率抽样会发生 (如自我选择)。当抽样框存在严重缺陷时, 也会发生这种情况。这些缺陷 (如无覆盖、无应答) 会使某个抽样框从适用于概率抽样转变为只能用于非概率抽样, 然而其边界非常模糊, 取决于具体情况和相关知识。

在非概率抽样中, 本书尽可能地遵循概率抽样的设计原则。因此, 希望获得概率抽样所固有的特征, 特别是随机化和代表性。这些可以间接或直接实现, 本节将系统性地回顾这两种策略。

2.2.3.2.1 近似概率的样本设计

在非名单型、非概率型样本中, 基本抽样设计是尽可能广泛地发布样本招募信息。在实践中, 这意味着不仅使用一种招募渠道 (如只用一个网络社区或只在一个网站上做广告), 而是要将几种招募渠道结合起来 (例如在多个在线场所招募年轻人)。一个成功案例是在 30 多个国家实施的“工资指标”项目 (*Wage Indicator project*)。该项目采用精心设计的招募策略, 在广泛的网站上发布不受限制的网络调查邀请 (Tijdens, 2014)。同样成功的还有广泛发布 (即数百个网站上的广告横幅) 且不受限制的互联网使用网络调查 (例如 RIS 项目 1996—2006¹) 以及斯洛文尼亚的在线仇恨言论调查 (*on online hate speech*) (Vehovar, Motl, et al., 2012), 这些样本单位都是从大量在线论坛中招募的。

对于非名单型、非概率型样本的另一个间接策略是构建样本, 使其结构尽可能与调查人群的结构相同。这种想法通常与“代表性样本” (*representative samples*) 概念有关, 并且在传统上以各种配额规则 (*quota rules*) 为基础, 通常控制社会人口统计学特征 (如性别、年龄、受教育程度、地区), 从而使样本具备总体的特征。配额通常在招募阶段或筛选阶段应用。例如, 当招募的女性已经达到配额上限时, 为了实现性别平衡, 研究人员可能会拒绝主动报名的女性 (或者为她们提供一个简

短的娱乐性调查),而只对男性进行网络问卷调查;或者加强吸引更多男性的在线招募渠道(例如在体育网站上做广告)。

当抽样框附带单位的额外信息时,如基于名单的非概率样本(在线样本库),原则上可使用与上述非名单抽样类似的策略。然而,这种情况下研究人员可以从名单中抽取所有单位且通常能了解更多关于单位的信息,因此抽样操作更为有效。研究人员可以控制样本的分布并提前设定配额。通过根据目标人群的结构(如年龄、性别、受教育程度、地区等)预选单位,可以进一步优化甚至用精细分层完全取代配额设置。一个极端的例子是个体匹配。只能当存在大量的单位数据库可用时,个体匹配才完全有效,比如在大型在线样本库中。首先从一些理想的或高质量的控制源中设计一个“代表性”样本,然后在数据库(比如在线样本库)中寻找与所需单位最相似的单位(Rivers & Bailey, 2009)。在此,也常会用到观察数据的因果推断和倾向性评分模型的相关概念和解决方案(Terhanian & Bremer, 2012)。

2.2.3.2.2 直接纳入概率抽样原则

在相关的抽样设计或招募过程中,要遵循随机化和概率抽样原则。有时,抽样过程的早期步骤可以考虑概率抽样的内容。例如,在有配额的概率抽样中,我们可能首先选择一个学校的概率样本,然后根据一些配额要求(例如年龄、性别、学年),通过不受限制的横幅广告邀请来招募学生。同样,对于基于概率模型的抽样,非概率样本的元素(例如配额)可以作为附加限制加入其中(Berzofsky, Williams, & Biemer, 2009)。在线招募通过弹出式邀请进行随机化,使招募工作在几天或几个小时内完成,因此可以近似于概率网络拦截调查。

当然,上述方法都不能确保在应用这些抽样设计后,相应的统计推论与采用规范概率样本的情况等同。在工资指标的案例中,荷兰工资的最终测算结果(即使在加权后)得到了改善,但与荷兰统计局的官方数据相比,仍有偏差(Steinmetz, Bianchi, Tijdens, & Biffignandi, 2014)。然而,事实是这些方法通常能带来一定改善。实践者们还开发了许多其他方法,这些方法在不同的环境、研究问题和目标变量上的成功率有所不同。本书将在下一节讨论这些程序在统计推论方面的潜力。

2.2.4 概率和非概率调查中的统计推论

第2.2节开篇提到,抽样包括两个密切相关的部分:抽样设计,这是预调查阶段的主要关注点;以及估计,这与调查数据分析阶段的统计推论有关。由于这两部

分密切相关,本节介绍统计推论相关的内容,以呈现更为完整的调查全貌。这也可以为预调查阶段需做出的抽样决策提供依据。本节首先总结概率抽样的统计推论,接着阐述在非概率环境中应用这种方法存在的问题,最后针对概率抽样与非概率抽样的选择困境提出建议。

2.2.4.1 概率样本的统计推论

根据抽样调查的定义(第1.1.1节),其目标是推断目标人群的状况。因此,当使用样本数据时,研究人员需要在推论统计的背景下呈现结果,至少是关键结果。这意味着需要量化“仅研究样本而非整个人群”所带来的误差风险,具体通过测量因所选样本质量的随机波动而产生的错误风险来实现。例如,对于总体百分比 P ,我们不仅呈现点估计值 $p = 30\%$,还需给出相应的95%置信区间。除其他因素外,置信区间在很大程度上取决于样本量,所以取值可以在区间 $[30.00 \pm 0.05]$ 、 $[30.00 \pm 0.50]$ 甚至 $[30.00 \pm 5.00]$ 。在每种情况下,相应的置信区间都能很好地体现估计的精度,这是使用者需要了解的关键信息。例如, $[30.00 \pm 5.00]$ 的置信区间意味着,在这种类型和规模的样本中,平均20个中有19个样本的置信区间将包含总体真实值 P 。同样,我们进行统计检验或构建统计模型时,也会报告统计检验力、模型拟合和其他推断指标。

各种统计推论方法几乎都以概率抽样为前提。然而,本书通常会提到当前常用的频率论方法(*frequentist approach*)。该方法假定存在某些固定的总体真实值(参数),这些真值基于所有可能样本估计值的抽样分布。

估计和整个统计推论过程可以非常简单,比如简单随机抽样中的样本平均数,而对于复杂样本(如集群样本、分层样本)和复杂估计值(如比率、回归、校准)则需要专门的统计处理,特别是涉及加权时。在任何情况下,许多精细的推断程序已经被开发出来并用于概率抽样(J.F. Healey, 2011; Lohr, 2010),在此本书不再赘述。

标准误(SE)的计算是评价估计值精度的基础。对于精度较低的估计值,可放在括号中,并附上相应的方法学说明。而在精度极低的情况下,例如基于少于10个单位的统计推论,应避免公布估计值,使用一个点‘.’代替,意味着估计值不是零,但不可靠。否则,许多使用者可能会无视方法学警告,倾向于报告最不可靠的估计值。

让统计学新手感到困惑的另一个问题是采用普查(例如,调查目标人群中的所有学生、成员、客户、雇员等)。一旦进行普查,便不存在任何推论统计的风险,也不会用到推论统计;不需要置信区间和假设检验,因为已经测量到了真值。有时,特别是当检验某些模型时,我们仍可以将当前的研究总体视作某一无限放大的虚拟

超总体的一个样本。这一方法有一定合理性,但是需要澄清和论证。当然,普查也会存在其他的数据质量问题,如无应答。

第2章的补充资料“网络调查实践中的抽样问题”(Sampling aspects in web survey practice)会对这些问题进一步加以阐述,见<http://websm.org/ch2>。

值得注意的是,上述情况仅适用于概率调查的理想实施。第2.2.5节和第2.5.3节分别探讨无覆盖和无应答造成的各种偏差,而第6.1.2节介绍数据质量的其他内容。

2.2.4.2 非概率样本的统计推论

对于非概率样本,样本单位的纳入概率是未知的。因此,如果没有进一步的额外假设,原则上非概率样本无法进行任何标准的统计推论计算,而这些额外假设往往具有风险且难以验证。此外,非概率样本往往有一定的选择偏差(Bethlehem & Biffignandi, 2012)。例如精通互联网的受访者通常更频繁地自愿参与非概率样本,因此各种互联网服务的使用率普遍被高估。

尽管统计推论程序适用于概率抽样情境,但在实际应用中,研究人员经常将其用于非概率样本,尽管无法保证其有效性和有效程度。例如,可以使用标准程序来计算非概率样本的95%的置信区间,并得到 $[30.0 \pm 0.5]$ 。然而,这一程序产生的置信区间可能不包含总体真实值(过度重复抽样)的风险不再是5%,而是未知的。很有可能这种风险要比5%高得多,但无法计算。

然而,在处理非概率样本的数据时,研究人员常常采用与概率样本完全相同的方式进行统计推论。原则上不能、也不应该这样做,因为无法满足基本假设(即已知概率)。如果这样做的时候,没有考虑任何假设,也没有任何警告,这就是有问题的做法,有必要使用一个独立的术语将其与概率样本背景下的表述正式区分开。相应地,我们使用指标值(indications)一词来代替估计值(Baker et al., 2013)。

在实践中,非概率样本情境中使用标准的基于概率的统计推论程序时,有三种因素可能会对研究人员有利。

- 在几乎所有非概率样本中都存在一定程度的“自然”随机化,这种随机化在不同情况下和不同样本中有所差异。极端情况下可能与简单随机抽样的选择非常相似。
- 研究人员为改进非概率抽样设计而采取进一步措施可促进这一进程。第2.2.3节已介绍过这些措施(拓展、随机、配额、匹配)。
- 此外,在调查后的调整(归类、加权)和估计(复杂估计量、统计模型)环节已有多种强大的先进技术。第4.1节将对此进行介绍。

在实践中，这三种因素共同发挥作用时往往非常有效。关于网络仇恨言论的研究（Vehovar, Motl, et al., 2012）就是成功使用非概率网络调查的典型示例。将之与平行开展的概率电话样本调查的比较证实了一些原则，这些原则通常适用于“广泛分布的”非概率样本：

- 等级量表的估计平均值通常非常稳健，在该研究中，两个样本的平均值完全匹配。
- 相关性、等级和子群体分析也相互匹配，分析结果完全一致。
- 对占比（百分比）的估计通常更容易受到影响。在该研究中，非概率网络调查的占比估计值更高（例如，非概率网络调查的受访者参与度更高）。然而，类别之间的等级和关系仍完全一致。

这种模式具有典型性。与之相对应的是，在线样本库使用非概率样本取得了成功结果（Rivers, 2007），并且已经成为市场研究中网络调查的主流方法（Macer, 2014）。此外，我们愈发注意到，正如 Rivers（2010）所指出的，学术论文会计算非概率样本的置信区间，并且他还列出了三个支持非概率样本的基本论据。

- 第一个论据来自上述样本选择和估计程序。
- 第二个论据与许多概率样本的弱点有关，这些弱点由无覆盖和无应答引起。与一些高质量的非概率抽样相比，这会导致概率抽样质量更差。
- 第三个论据是某些邻近领域已经完全接受非概率抽样的推论，如实验和准实验研究、评估研究、观察研究和医学研究。

另一个间接论据涉及为非概率调查付费的客户。如果从长远来看，廉价的非概率样本的结果证明其价格的合理性，且符合调查目标，那么这就清楚地表明，在这些情况下，昂贵的概率样本的附加值过低。

另一方面，反对者坚持认为，在不知道纳入概率的情况下，没有理由基于非概率样本进行统计推论（Langer, 2013; Lohr, 2010）。同样，学者们也经常比较这两种方法（Yeager, et al., 2011），事实证明概率抽样设置总是更胜一筹。第 5.2 节提供了更多的依据，表明非概率在线样本库的估计质量存在某些问题和不稳定。当然，这种比较只限于统计方面，不包括效率、成本和其他各方面。

在任何情况下，只要使用、运行或讨论非概率样本，就应该明确区分它们，因为不同非概率样本之间存在巨大差异。一些样本认真地招募和拓展所覆盖的单位（如分层、配额、匹配），而另一些样本只是在某个狭窄的子区间内自主选择。调查实施后处理阶段采用的应用程序也是如此（计算、加权、校准、估计）。目标变量

和研究问题的具体内容也至关重要。前文已提及占比估计通常最容易受到影响，但相关分析也存在问题（Pasek & Krosnick, 2010）。另一方面，有些主题比较稳健，即使使用低质量的非概率样本也能成功处理，如市场研究和民意调查中的变量；而有些主题则需要严格的概率抽样，比如就业或健康状况。

总而言之，本书建议以更开放的态度接纳在非概率情境中使用标准统计推论方法。尽管如此，仍有两点需要注意。

- 首先，应该清晰描述、介绍并批判性地评估样本选择程序。我们支持美国民意研究协会的建议（Baker, 2013），即应该像概率样本那样，详细描述非概率样本的抽样、收集和调整方法。这也适用于数据质量指标的报告。
- 其次，详细说明所使用的假设和模型，并对概念分歧、风险和解释的局限性作出说明。当在任何非概率样本中使用标准的统计推论时，最基本的是要承认估计、置信区间、模型拟合和假设检验可能运用错误或根本不存在。当然，如果一些经验性证据支持在某些非概率样本中进行的统计推论多次有效，这应该作为一个额外依据。

严格地说，从形式上看，概率抽样和随机化并不是有效统计推论的必要前提，因为具体的建模方法也可以完成这项工作。然而，这些建模方法还没有发展成一般的推断程序，否则就可以直接用于实际操作，例如对样本中的所有变量进行描述性统计量的估计。本书使用标准统计推论（*standard statistical inference*）概念描述这些估计。相反，这些特定的建模方法主要关注目标变量、辅助信息、建模假设和各种外部输入，如过去的变量之间的关系和设定的统计模型。一个典型案例是对选举结果的描述估计，在这一场景中，来自非概率样本的数据只是预测模型的较小输入部分，而这些模型越来越可以成功预测（例如2013年的美国总统选举）²。另一个例子是对关系的关注，因果推断也可以在非随机情境下运行（Rosenbaum & Rubin, 1983）。然而，虽然这些方法有效，但由于其特殊性，本书不作介绍，只关注一般的统计推论方法。

2.2.4.3 决定概率—非概率抽样的困境

利用非概率样本进行统计推论缺乏理论基础，相应依据主要来自于实践经验的积累。如果某种方法在实践中反复被证明是成功的，这就成为支持其合理性的基本依据。当然，在某些情况下，某种非概率抽样方法屡屡取得好效果，但在其他情况下却未必。同样，有时一个缺陷较多的概率抽样（如应答率只有百分之几）反而优

于最好的非概率抽样，反之亦然。只有在特定情况下积累的反复实践证据才能证明非概率抽样的合理性。通常，经过多次成功的重复，加上相应的试错，从业人员也能预测出抽样设计和统计推论的变化所带来的不同效果。

因此，在决定采用概率抽样还是非概率抽样时，首先要考虑的是回顾之前的所有依据。如果可能，还应该检查变量的分布情况，并且在有外部对照的情况下将其与外部对照进行比较。这将有助于我们了解非概率样本在随机化和总体控制（即“代表性”）方面与概率样本的相似性。

另一层面的考量是系统评估更广泛的情境。即不应该只关注概率—非概率问题，而应该考虑与无覆盖、无应答、模式效应和其他测量误差等相关的问题，以及成本、时间、可比性等因素。特别要注意的是需将非概率抽样的真正问题（即选择偏差）与其他特征区分开。如果非概率抽样没有造成选择偏差，而实际问题是由模式效应等因素造成的，那么不选择非概率抽样就没有什么意义。

因此，很少孤立地决定采用概率还是非概率抽样，而是同时考虑一系列更广泛的因素。例如，在一项关于青年旅游的调查中，研究人员可能会在基于概率抽样的电话、邮寄或网络调查，与基于非概率抽样的在线样本库或通过在线社交媒体招募的样本之间做出选择。这需要考虑多个方面，其中所有的组成部分（如无应答、无覆盖、选择偏差、模式效应、成本、时间、可比性、准确性等）往往是不可分割的。例如，社交媒体选项可能不仅存在无覆盖和选择偏差问题，而且还具有低成本和方便性等特点。在这种情况下，由于使用非概率抽样而产生的选择偏差只是众多因素中的一种，不能被单独分开处理。

选择多种方法的组合有时会是较好的解决方案。由此，可以通过昂贵的概率调查获得关键变量的总体估计值，而通过非概率调查完成深入分析。例如，在一项关于网络仇恨言论的研究中（Vehovar, Motl, et al., 2012），研究者通过基于概率抽样的综合性电话调查获取关键问题的信息，而在广泛推送的自我选择网络调查中实施深度提问。

当没有必要的时间、预算和抽样名单时，概率抽样可能并非可行之选。然而，此时仍然应该考虑上述因素以优化程序，并详细说明非概率抽样是否值得进行。

使用非概率网络调查的一个很好的反面案例是，2014年斯洛文尼亚关于校服这一热门公共话题的调查。该网络调查是相应部委合作发起的，旨在了解关键群体的态度。调查方要求所有学校在其网站上发布网络调查的链接，对于一个200万人口的国家来说，应答数量是巨大的（20 731）。结果显示，教师、家长、学生这三个群体都反对校服³，反对和支持比例大致为60:40。这一结果随后被用于实际拒绝校服项目中，充分暴露了非概率样本的危险性。反对校服的群体可能只是“声音更大”且参与积极性更高，而采用概率样本则很可能得出相反结果。改进非概率抽样也可

以获得更好的结果,即抽样设计中引入更多的随机化元素。例如,首先,通过分层随机抽样选择大约30所学校,而不是向所有学校提供自我选择的机会。其次,在被选中的学校开展更有力度的宣传活动。还可以引入调查后的调整,基于一般特征(年龄、性别、社会阶层、地区)和具体特征(学校类型和年级),以及全国性调查中可能存在的服装品牌态度相关数据进行校准。

最后,应注意非概率样本很容易造成资源浪费,甚至可能产生不正确的估计而造成损失。必须重申的是,需要有令人信服的依据来证明实施程序在过去类似的情况下反复运行良好,因此对此次实施也是有效的。幸运的是,此类依据越来越多,因此需要努力去寻找、使用和参考它们。

2.2.4.4 关于非概率网络调查的结论性评述

第2.2.2节介绍了概率样本的抽样技术,并提到近一个世纪以来,概率抽样帮助调查实践取得了巨大进步。现在本书可以对非概率网络调查的优势作出同样论断:数以千计的网络调查项目得以开展,这在其他年代是不可能的。数以百万计的网络调查项目运行得更快、更简洁、更节省成本。尽管这种变化是逐渐发生的,几乎不引人注意,但如今我们可以说非概率网络调查彻底改变了调查领域。

本书在许多章节提到(或暗示),从形式上看,在非概率情境下,概率抽样原则和标准统计推论均不适用;除非将非概率的结果视为有待评估的指标和近似值。同样,它们也不能像在概率样本中那样被事前估计,而只能在事后估计,并且只能在一定程度上进行。然而,这些估计仍然非常有价值,因为我们可以改进相应的抽样和估计程序,或者至少认识到并有意识地将某些非概率的抽样和估计程序限制在特定类型的问题上。

根据定义,在概率抽样之外基本没有标准的统计推论,因为标准的统计推论与纳入概率密不可分。当纳入概率未知时,如在非概率抽样中,相应的推断程序隐含着对它们的假设。例如,通常的样本平均数(即简单平均数、算术平均数)计算,隐含地赋予样本中的所有单位 n/N 的纳入概率,其中 N 是总体规模, n 是样本规模。然后,需要相应的评估程序来判断这种“估计”的质量,或许更应该称之为“指标”或“近似值”。

因此,非概率抽样中的统计推论在很大程度上脱离了科学性,而进入了获得最佳实践程序的艺术和技艺范畴。此时,基于试错的经验以及研究人员的直觉至关重要。相应地,特别是对于复杂的问题和大型项目,专家或具有相关经验的专家小组的建议非常有价值。

所有这些听起来可能非常不科学,所以一些资深统计学家可能会对推断风险无

法事先量化的程序产生怀疑,甚至产生职业性反感。然而,这不正是应用统计学家的基本特征吗?这将他们与“脱离现实”的数学家区分开来,他们不会犹豫不决,而是接受那些能在各种实际应用中发挥作用的程序。

需要补充的是,这些问题并不罕见,因为从概率抽样诞生之初(近百年前),研究人员就一直在尝试各种替代方法来简化昂贵的概率抽样程序。只要这种方法足够有效,他们就会放弃严格的概率抽样,比如在网络调查出现的几十年前,大多数市场研究就经常使用大型的、自我选择的、非概率家庭小组。

还需要重申的是,网络调查模式与非概率取样之间并无内在关联。虽然大多数网络调查具有非概率性质,但这并不是网络调查的固有特点。根据第 1.1.1 节的定义,网络调查模式与抽样相互独立,非概率抽样同样出现在传统调查中(如街角的便利调查、商场拦截、电话调查、杂志邮寄调查),而许多网络调查是基于概率抽样的。因此,非概率网络调查的统计推论问题,本质上只是用成本较低的非概率样本代替昂贵而复杂的概率样本这一古老问题的表现形式。网络调查如此凸显这个问题,不过是因为它们成本较低、简单易行和使用率高。

这种概率—非概率的困境在实践中并不常见。相反,它被嵌入在“随机化程度高低不同的非概率样本”这一更常见的困境中。这通常也意味着非概率样本成本的多或少。正如此类问题的惯常处理方式,对该困境的解决通常取决于特定环境的实际情况(Galesic, 2009)。

AAPOR 关于非概率样本的报告中有关于这些问题的进一步讨论(Baker et al., 2013)。该报告严谨而详尽地分析了样本选择、加权、估计和质量评估措施,以及“适用于目标”(fit for purpose)和“适用于应用”(fitness to use)的概念。本书赞同这份报告的观点,即该领域尚有很多因素需要考虑。然而,在抽样设计程序或调查后的调整方面也许改进的空间相对较小。此外,我们期望在建模方面,特别是在标准(如报告)、分类(如非概率样本的类型)和诊断(如质量指标)方面有所改进。这些问题对于更广泛的当代调查研究议程也是至关重要的(Callegaro et al., 2014; Chang & Krosnick, 2009; Yeager et al., 2011)。我们将在第 4.2.1 节(初步报告)阐述推论原则的实施时重申这些问题。

2.2.5 无覆盖问题

无覆盖(non-coverage)是指目标人群中相当一部分合格单位被排除在样本之外的情况。在无覆盖情况下,部分单位的选择概率为零,这违反了概率抽样的基本

原则,导致标准统计推论无法正常开展。尽管这一问题通常在抽样阶段初期处理,但本书选择在抽样章节的最后部分进行阐述。原因是需要先建立抽样框、统计推论等核心概念,才能更深入地理解无覆盖问题的本质与影响。

在网络调查中,无覆盖往往是影响数据质量的关键缺陷。根据定义,非互联网用户已经被排除在任何网络调查之外。然而随着年轻一代的计算机知识越来越丰富,这种类型的无覆盖正在逐渐消失,至少在发达国家如此。只要抽样框没有覆盖整个目标人群,其他形式的无覆盖问题仍会持续存在。

本节的其余部分将重点讨论基于名单的概率抽样中的无覆盖问题,但相应原则同样适用于非基于名单的概率抽样(例如概率网络拦截调查)。在非概率抽样背景下,这些原则可作为有用的建议,发挥类似其他抽样原则的作用。

首先讨论由非互联网用户造成的无覆盖问题。在20世纪90年代的发达国家,这是一个特别突出的问题,当时互联网用户还只是非常特殊的少数群体。那时,研究者迫切希望找到解决无覆盖问题的方法,以充分利用网络调查低成本、高效率的优势。然而,尽管付出了诸多努力,仍未发现任何抽样或加权方法能够将从一个规模有限且特征特殊的群体(互联网用户)中获得的调查结果,可靠地推广到整个人群(Vehovar, Lozar Manfreda, & Batagelj, 1999)。

互联网普及率变化速度很快,更多相关内容可参见本书第2章补充资料“互联网统计”(Internet statistics),详见<http://websm.org/ch2>。即使在发达国家之间,互联网普及率也存在显著差异:例如,根据欧盟统计局数据,2013年欧洲家庭的平均普及率为79%,但土耳其仅为49%,冰岛则高达96%,各国家不等。许多其他发达国家的普及率,如美国、日本和澳大利亚,都有80%左右的高比例,但截至2014年,世界上大多数国家的互联网普及率仍在30%左右或以下。

即使在互联网普及率接近或超过80%的情况下,剩余非互联网用户通常也不是普通人群的随机子样本。在大多数发达的国家,20世纪90年代确定的非互联网用户的特征至今依然适用。根据美国⁴和欧盟⁵官方统计数据,老年人、非白人、某些少数族裔、受教育程度较低和低收入人群更有可能成为非互联网用户。这些差异有时被称为基本的数字鸿沟(digital divide)(Vehovar, Sicherl, Hüsing, & Dolničar, 2006),但其他形式的鸿沟也在显现,例如基于技术使用类型和结构的差异,这甚至引发了对“互联网用户”定义的争议(例如,交互式电视的用户是否属于互联网用户?)

网络调查中出现无覆盖的另一个原因与抽样框的缺陷有关。例如,缺乏某些学生、会员、客户等群体的电子邮件地址,就会产生影响无覆盖的一般性问题。几十

年前,研究者一般认为,当电话调查的覆盖率低于 80% 时 (Groves, Biemer, Lyberg, Massey, Nicholls & Waksberg, 2001),就需要采取应对措施,例如双框架设计,结合面对面模式来覆盖缺失的单位。荷兰的全国读者调查 (Petric & Appel, 2007)便是典型案例,尽管互联网渗透率超过 80%,仍然需要使用面对面调查对网络调查进行补充,因为非互联网用户的阅读习惯比较特殊。

通常而言,可接受的无覆盖水平取决于无覆盖偏差 (*non-coverage bias*)。无覆盖偏差指目标人群的真实值与无覆盖样本中的预期估计值之间的差异。对于变量 y ,其相应的平均值 $\bar{y}$ 的预期偏差称为 $\text{Bias}(\bar{y})$ 。在某些情况下即使无覆盖率很高,只要缺失的单位在关键变量上无特异性,部分变量的估计值也不会出现偏差。因此,这种情况与上述非概率样本中估计的稳健性非常相似。

由无覆盖引起的估计偏差通常与无覆盖率 (W_n) 成正比。无覆盖率被定义为目标人群中无法被纳入样本的单位数量与目标人群总数的比率。无覆盖概率还取决于调查估计值在覆盖人群 ($\bar{y}_c$) 和无覆盖人群 ($\bar{y}_n$) 之间的差异。无覆盖偏差可以表示为 (Biemer & Lyberg, 2003, p. 69):

$$\text{Bias}(\bar{y}) = W_n \times (\bar{y}_c - \bar{y}_n)$$

如果 $\bar{y}_c$ 和 $\bar{y}_n$ 之间的差异可以忽略不计,那么无论无覆盖率高,偏差都不会产生影响。同样,如果无覆盖率 (W_n) 极低,即使两部分人群存在较大差异,也无需过度关注。

但在多数情况下,研究者无法预先知晓 ($W_n, \bar{y}_c, \bar{y}_n$) 这些参数的具体数值。尽管如此,仍可通过上述公式估算和模拟各种方案,判断无覆盖风险是否在可接受范围内。例如,假设无覆盖率 $W_n = 40\%$,覆盖人群的估计值为 $\bar{y}_c = 0.20$,无覆盖人群的真实值 $\bar{y}_n = 0.30$,将得到无覆盖偏差为 $\text{Bias}(\bar{y}) = 0.4 \times (0.2 - 0.3) = -0.04$,此时,由于使用 $\bar{y}_c$ 而非真实值 $Y = 0.24$,便产生了偏差。

相对偏差比偏差 $\text{Bias}(\bar{y})$ 更具参考价值。相对偏差: $\text{Rbias}(\bar{y}) = \text{Bias}(\bar{y}) / \bar{Y}$,在这个例子中 $\text{Rbias}(\bar{y}) = 0.17 (0.04/0.24 \approx 0.17)$,不考虑偏差的正负号。这意味着估计值 ($\bar{y}_c = 0.20$) 与真实值 ($\bar{Y} = 0.24$) 之间存在 17% 的差距。这一差异已超出 5% 或 10% 的常见相对偏差基准线,因此对于多数研究目的而言,这种偏差是不可接受的;但实际的容忍度需根据调查目的和具体情境确定。

还需要将这种偏差与抽样变异性进行比较 (见第 2.2 节),用标准误 $\text{SE}(\bar{y})$ 表示。因此,可以得到标准化偏差: $\text{Sbias}(\bar{y}) = \text{Bias}(\bar{y})/\text{SE}(\bar{y})$ 。在简单随机抽样 (SRS) 假设下,标准化偏差可粗略反映 t 检验值。其中 $t = 1.96$ 或更高时意味着存在统计

学意义上的偏差。本案例中,如果 $SE(\bar{y}) = 0.03$,这将导致 $Sbias(\bar{y}) = 0.04/0.03 = 1.3$ 。由于该值很低,不能认为有统计上的显著偏差(尽管 $Rbias$ 很高)。更多示例可以在第2章的补充内容“互联网统计”中找到,见 <http://websm.org/ch2>。

基于上述计算、模拟和对参数的进一步分析,研究者可采取以下策略应对无覆盖问题。

- 可以在概念层面将目标人群限制在网络调查可以接触到的群体,即仅关注互联网用户,而不是整个人群。当然,如果这种限制与最初的研究目的冲突,问题就会再次出现。例如,调查公众舆论,非互联网用户同样拥有投票权,且投票行为可能与互联网用户不同,这种方法就可能行不通。
- 在抽样设计中可以采用同样的“无为”方法,即将重点放在整个目标人群上,仍然忽略无覆盖部分的单位。例如,在针对普通人群的网络调查中仅纳入互联网用户,然后通过加权、计算等调查后调整方法概括整个人群,以修正无覆盖率。这些调整通常能产生积极效果,但影响程度相对有限(见第4.1.4节)。它们被广泛用于一般非概率样本中,以弥补无覆盖、无应答和自我选择带来的偏差(Callegaro et al., 2014c)。
- 采用更积极的方法应对无覆盖涉及成本更高的招募方式,以覆盖缺失人群。这意味着采用混合模式系统,如用替代模式纳入缺失单位,正如第2.1.2节中所述。最典型的是网络调查与邮寄调查结合起来,也可能涉及双框架设计,即使用独立且存在重叠的样本(例如邮寄地址样本和电子邮件样本)。

无覆盖也是网络调查方法的重要挑战之一,在学术文献(Bethlehem & Biffignandi, 2012; Smyth et al., 2010)、国家和跨国一般人群调查(如欧洲社会调查, Ainsaar et al., 2013)、各种研究网络(如 NCRM 网络⁶或 Cost Action WebDataNet⁷),以及越来越多的国家概率在线样本库(见5.2节)中均有深入探讨。

综上,本书仅在概率抽样的背景下讨论了无覆盖问题。对于非概率抽样来说,这些原则仅能作为近似参考。无覆盖本身就是导致样本非概率性质的重要原因之一。此外,在调查普通人群时,网络调查中的无覆盖问题尤为关键。尽管互联网普及率不断提高,但在2014年发展中国家仍有20~30%的非互联网用户,很少有国家的互联网普及率超过90%。不过,未来几年内,发达国家(如经济合作与发展组织国家)的互联网无覆盖问题将逐渐消失,而世界其他地区的这一过程则会缓慢得多。

2.3 问卷准备

问卷准备通常是预调查阶段耗时最长、资源投入最大的环节。与其他阶段相同，我们重点聚焦网络调查特有的内容。但本节更多关注一般性问题，因为它们对网络问卷的质量至关重要。

本节将依次概述问卷编制的核心问题（第 2.3.1 节）、问题类型（第 2.3.2 节）、问卷结构、计算机化和版面布局（第 2.3.3 节）以及参与度（第 2.3.4 节），最后总结问卷测试（第 2.3.5 节）和问卷准备工作的整合（第 2.3.6 节）。

2.3.1 一般性问题

本节通过回顾调查问卷的定义、分类和概念，概述问卷编制的关键要点。

2.3.1.1 调查问题的类型

学术文献通常根据调查问题的内容（*content*）来架构问题（Bradburn, Sudman, & Wansink, 2004; Groves et al., 2009; Presser, Rothgeb, Couper, Lessler, Martin & Singer, 2004）。当问及受访者可观察的信息时，从简单的人口统计学特征到与行为有关的复杂描述，谈论的是事实性问题（*factual questions*）。另一方面，非事实性问题（*non-factual questions*）涉及态度、观点、意图、期望、信念、看法、自我分类、评估等内容。此外，还有心理学和知识性问题（*psychographic and knowledge questions*），但这些问题通常在调查范围之外，多见于心理学和教育测试中。调查内容还可以根据问题的敏感性或威胁性（*sensitivity or threatening nature*）进行区分。某些问题侵犯受访者隐私、保密性或私密性（例如收入、性、健康），或者与社会非赞许（*socially undesirable*）行为（例如撒谎、偷窃、欺骗等）有关，会让其感到不舒服。

另一种分类是基于应答格式（*response format*）（Alwin, 2007; Krosnick & Presser, 2010; Saris & Gallhofer, 2014）。在封闭式（*closed-ended*）问题中，受访者从预设选项（例如性别、地区）中选择答案；而开放式（*open-ended*）问题要求受访者输入文字（例如对某些服务的公开评论）或数值型答案（例如输入具体金额）。

测量水平（*measurement level*）是另一个关键区分维度。测量水平由提供给受

访者的应答选项的性质决定。由于其重要性,几乎所有的方法论和统计学教科书都会介绍相关内容。例如,幸福程度是与受访者内在状态相关的概念。受访者可以用特定的值来表达其幸福程度,而这些值可以用不同的测量水平表示。相应的选项范围可以从向受访者提供非结构化的开放式文本,到高度结构化的测量水平,即受访者选择一些数字,如在1~5的范围内选择数字代表自己的幸福程度。相应测量水平包含名义(*nominal*)、顺序(*ordinal*)、等距(*interval*)或比率(*ratio*)。

- 对于名义测量(*nominal measurement*)的问题,仅能区分类别。因测量水平本身没有内在顺序之分,所以不能进行排序。例如,关于性别、地区、宗教的封闭式问题格式,以及询问描述的开放式文本格式。
- 顺序测量(*ordinal measurement*)允许对应答类别进行排序(例如,可以说“每天使用”比“每周使用”更频繁,或者“高兴”这一选项比“不高兴”这一选项意味着更幸福),但不能够量化类别之间差异的大小。
- 等距测量(*interval measurement*)可以比较各类反应之间的距离(例如,可以说1994年和2004年之间的时间差异等于2004年和2014年之间的时间差异)。
- 比率测量(*ratio measurement*)具备等距特征且有绝对零点,由此可以计算数值间的比率(例如,40年的时长是20年的两倍)。

测量水平极为重要。不熟悉这些概念的读者可以在第2章补充资料“测量水平”(Measurement levels)(网址:<http://websm.org/ch2>)中找到进一步的解释和说明,也可以查阅有关基本调查方法的书籍(Groves et al., 2009)或统计学文献(Kirk, 2007)。

调查问题的测量水平与受访者提供答案的过程有关,通常反映在分析阶段变量的测量水平上,常被称为变量尺度(*scale of the variable*)。然而,问题的测量水平和变量尺度是两个独立的概念,不会自动匹配。例如,使用名义测量水平问题得到的回答可以作为顺序尺度变量进行编码和分析,或者使用顺序测量水平问题得到的回答在数据分析阶段可以作为等距尺度变量处理。

了解根据内容、应答格式和测量水平划分的问题类型,对于调查问卷的编制过程以及调查数据库(*datafile*)(第4.1节)的准备和组织至关重要,本书用数据库来指代包含所有应答的文件。数据库通常是矩形矩阵,受访者(单位)位于行,应答(变量)位于列。

2.3.1.2 问题编制过程

问题编制始于对理论构念的概念化和操作化,理论构念是研究者希望通过调查来测量的核心概念。虽然不同领域(社会、市场、健康研究等)的教科书对这一过程的阐述方式各异(Bradburn et al., 2004; Iacobucci & Churchill, 2015; Neuman, 2009; Saris & Gallhofer, 2014),但本书主要参考 Hox (1997) 的阐述。问卷开发的过程始于对某个理论建构或概念(*theoretical construct or concept*) 的阐述,以及对其子领域的定义,子领域被称为维度(*dimensions*)。这一过程为概念化。例如,在员工调查中,概念为“工作满意度”,可细分为维度(或子领域):“对工资的满意度”“对经理的满意度”和“对工作条件的满意度”。然后是操作化(*operationalization*),寻找每个概念或维度的经验指标(*empirical indicators*),并将之转化为实际的调查问题(*survey questions*)。

对于同一个维度,有多个经验指标时,情况会更复杂。例如,“对工资的满意度”这个维度可以使用不同的经验指标操作化,如“对实际收到的金额的满意度”“相对于组织的平均工资的满意度”,等等。在调查术语中,这些指标被称为项目(*items*)。当然,如果某维度只有一个指标,这样的维度就会与相应的项目相匹配。事实上,为了简化讨论,本书在下文中将假设维度只有一个指标。

当概念或维度涉及不同的主题(*subjects*) 时,情况也会复杂。例如,受访者要评价自己和伴侣的工作满意度。在这种情况下,应对每个主题的每个维度都设计单独的项目。

问卷中的问题可直观、单独呈现,也可组合为包含多个子问题的表格形式(如矩阵式问题)。项目包括独立问题和子问题,只要测量的是同一主题的某个概念的单一层面,均视为一个项目。在统计学术语中,项目对应抽样调查中的“变量”概念(第 2.2 节)。

多个问题或问题表格可以进一步整合为组块(*blocks*)。组块通常指一系列问题,它们形成了与内容有关的问题群(例如,社会人口相关的问题组块,工作满意度问题组块等)。在包括网络调查在内的自填式调查模式中,问卷通常横跨几个页面,每个页面(*page*) 可以包含一个或多个问题,这是非常重要的方法学决策。

2.3.1.3 问题和调查问卷设计原则

大量文献已经系统梳理了问题措辞和问卷设计的基础知识(Brace, 2008; Bradburn et al., 2004; Krosnick & Presser, 2010; Saris & Gallhofer, 2014)。本书在此总

结了 Krosnick 和 Presser (2010) 的建议。关于问题措辞, Krosnick 和 Presser (2010) 建议:使用简单的词语(而不是技术术语、行话或俚语)、简单的语法、具体的措辞、全面且互斥的应答选项。每个问题应该只测量一个维度,避免双管齐下;避免使用含义模糊的词语、诱导性问题、否定词,尤其是双重否定词。

更加具体的建议要结合问题的内容与问题的类型,这可以在相应的文献中找到。例如,对于行为类问题,需要注意记忆和回忆相关事项。对于敏感问题、替代回答(受访者替他人回答)、纵向研究和特殊群体(如儿童),也有许多具体建议。

另一套原则涉及整个调查问卷的设计,从调查问卷的问题结构、背景和问题的顺序到格式、填写指导语、视觉布局等。

2.3.1.4 测量过程和测量误差

在测量阶段施测问卷时,可能会出现随机或系统性的测量误差(*measurement errors*) (Biemer, Groves, Lyberg, Mathiowetz & Sudman, 2004; Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz & Trewin, 1997)。测量误差指观察值和真实值之间的差异。这两个值之间的相关通常用效度(*validity*)来表示。效度指是否真正测量了目标问题(Saris, 2012, p. 537)。反应偏差(*response bias*)与效度密切相关,反映了应答值和真实值之间的系统性差异(Groves et al., 2009, p. 279)。另一方面,信度(*reliability*)表达与重复测量中反应的稳定性有关的随机变化(例如,如果20分钟后或第二天再问,受访者对一般生活满意度的态度是否稳定?),这也与反应偏差密切相关(Groves et al., 2009, p. 282)。通常存在关于信度和效度数值的标准化信息,例如,ESS(调查质量预测工具)中问题的信息⁸,或者市场研究问题的信息(Bearden, Netemeyer, & Haws, 2011)。

2.3.1.5 应答过程的认知方面

在回答调查问题时,受访者需要完成一系列特定的认知步骤以形成答案。对这一过程的研究始于20世纪50年代(Biemer & Lyberg, 2003, p. 123)。此后,研究人员提出了多种模型来拆解应答过程(Tourangeau & Bradburn, 2010)。广为人知的是 Tourangeau、Rips 和 Rasinski (2000) 提出的模型,它包括四个认知成分:对问题的理解(*comprehension*)、从记忆中提取(*retrieval*)相关信息、对提取出的信息进行判断(*judgement*),最后是作出反应(*response*),选择和报告答案。这些组成部分也构成了研究由受访者导致的测量误差和推动问卷方法发展(如认知访谈,*cognitive interviewing*)的基础。

认知方法模型的一个重要分支是满足 (*satisficing*) 模型 (Krosnick, 1999), 该模型涉及在回答过程中可能出现的各种偏差。根据该模型, 受访者可以通过必要的努力执行所有的逻辑步骤来优化 (*optimize*) 应答行为, 从而给出合适的答案。另一方面, 当受访者粗略地执行一个或多个步骤时, 就会出现“满足”现象。强烈 (*strong*) 的满足意味着忽略信息提取或判断步骤, 受访者浅显地理解问题, 并给出在这种情况下“看似合理”的答案, 选择中性选项 (如没有意见)、默认选项、随机作答或没有区分地回答。较弱 (*Weak*) 的满足意味着受访者执行所有四个认知步骤, 但不够深入, 例如选择第一个看似可接受的选项或表现出默许偏差 (倾向于同意任何陈述)。

2.3.1.6 非实质性应答的具体内容

除了开放式问题或者封闭式问题中的实质性回答选项, 研究者有时会设置非实质性回答 (*non-substantive response*) 选项 (例如“不知道”“无意见”)。然而, 这类选项可能会增加满足性应答和其他形式的低质量应答行为 (Krosnick, 1991; Krosnick, Holbrook, Berent, Carson, Michael Hanemann, Kopp, ... Conaway, 2002; Thomas, Uldall, & Krosnick, 2002), 也没有证据表明会提升效度 (Alwin, 2007, p. 199)。因此, 此处遵循 Krosnick 和 Presser (2010) 的建议, 即原则上应避免非实质性的选项, 尤其是在非事实性问题上。除非这种选项是重要的、是在特定环境下的合理选择 (Manisera & Zuccolotto, 2014), 或者问题与受访者的知识范围有关 (Sturgis, Allum, & Smith, 2008)。当问题是必填项, 受访者必须选择封闭式问题的答案时, 可能需要非实质性选项。在视觉呈现上, 非实质性选项必须与实质性选项明显区分开 (Tourangeau, Couper, & Conrad, 2004), 否则受访者会误认为它是另一个实质性选项。网络调查中使用非实质性选项, 还需与应答提醒甚至强制应答机制相配合。

2.3.1.7 问题库

在开始编制问题前, 研究者应充分检索相关领域的文献或既有资源。数据库中存有大量经过验证来自官方及学术调查的问题。这一点对人口统计学特征的问题来说特别方便, 因为不需要做出任何改进且已经被证明更有效和可信。此外, 各类机构已构建专门的问题资源库 (例如 Q-Bank⁹ 或调查问题库¹⁰)。同样, 市场营销领域 (Bearden et al., 2011) 或心理测量领域 (Spies, Carlson, & Geisinger, 2010) 也存在丰富的包含问题清单的资料和手册, 不仅提供具体措辞, 还附带效度、信度数值。网络调查软件供应商也越来越多地开发主题丰富的多语言资料库。

2.3.1.8 网络问卷的特殊性

问卷编制的前期阶段主要是概念化和操作化,网络问卷几乎没有特殊性。然而,当涉及问题的措辞和调查问卷的版面布局时,网络调查问卷的特殊性愈发明显,特别是自填式作答和计算机化这两个方面。

由于采用自填式 (*self-administration*),网络问卷是研究者与受访者互动中的主要沟通工具。它的任务不仅是测量,而且包括传递调查的合法性和重要性,提供指导语和维持参与动机。自填式能增强受访者的隐私感知,这可能可以促进敏感信息的表露。网络问卷调查要求受访者具备基本的计算机知识,这对某些人群来说可能存在问题。此外,视觉沟通的有效性至关重要,需要精心设计指导语、格式和版面布局 (Dillman et al., 2009)。

网络问卷的另一个特点来自于计算机化 (*computerization*)。它带来了许多优势,包括交互性设计和多媒体元素整合。具体方面与阅读习惯有关:在网络上,人们倾向于快速浏览网页,抓取零散的单词和句子 (Nielsen, 1997),以 F 型轨迹阅读内容 (Nielsen, 2006)。眼动追踪研究指出,受访者在问卷页面的顶端 (Garland, Chen, Epstein, & Suh, 2013)、问题选项列表的开始 (Galesic, Tourangeau, Couper, & Conrad, 2008) 投入更多注意力,并且与纸笔问卷相比,应答方式更分散 (Fuchs, 2003)。

技术环境 (*technological environment*) 的动态变化与计算机化有间接关系,设备、界面和浏览器的多样化对用户体验有重要影响。网络调查软件需不断适配这些变化,这使得网络问卷更加具有特殊性。软件对网络问卷的编制特别重要,因为它可能提供精心设计的问题类型、高级功能、精细的格式化选项、问题库、整个过程的指导等。

对本节内容进行总结发现,调查问卷的编制是一个非常复杂的过程。这个过程受到简化操作、仓促设计及专业性不足的影响。因此,研究者应该充分考虑上述一般性原则以及网络调查的特殊性。后续章节也会系统地综述这些问题。

2.3.2 问题类型

本节将系统介绍网络问卷中的核心问题类型,包括各类问题的适用场景及版面布局格式建议。本节主要阐述方法论方面的内容,而非问题的具体内容——后者是各类主题文献中讨论问题类型时采用的另一种视角 (Bradburn et al., 2004)。同时,本节讨论范围将限定在第 1.1.1 节定义的基本网络调查模式。这意味着问题可通过电子屏幕呈现,可结合多媒体与动画元素,也意味着受访者使用键盘、鼠标、触摸

屏或其他手写电子设备（如指针、手写笔等）作答。我们进一步将受访者的手动电子输出限制在文字和数字上（即字母和数字），不涵盖以下内容：要求受访者绘画和其他替代性的调查测量方法（Groves et al., 2009, p. 150），如上传文件、拍照、附上扫描件、收据等；提供自动测量数据（如 GPS）和各种记录，如来自外部行政（如考试结果）或商业记录（如财务数据）的数据。本节会在末尾简要概述这些替代性调查，非因其不重要，而是因为超出了基础网络调查模式的核心范畴。

在这一背景下，我们把调查问题理解为一种特殊的测量工具，通过结构化语句形成刺激信息，引导受访者通过认知加工过程生成答案。

我们采用的方法是根据方法学标准对问题进行分类。因此，本书最初（第一层级）是按照方法学的复杂性来划分问题。

- 单一项目问题（*single item questions*）每次针对一个调查对象的某一概念的单一维度。在这种情况下，调查项目也与数据文件中的列和统计分析中的变量相匹配。一个例子是关于受访者年龄的问题，单位是岁。
- 多项目问题（*questions with multiple items*）可能有些复杂。在最简单的情况下，多个相同格式的子问题可测量同一调查对象的多个维度。例如，可以在三个维度上测量受访者的总体满意度，如“对工作的满意度”“对家庭生活的满意度”和“对健康的满意度”。这种情况要有三个调查项目，分别与数据文件中的三列和统计分析中的三个变量相匹配。当受访者对多个对象的同一维度进行评价时，也会出现类似的情况（例如，评价三个家庭成员的“健康满意度”）。同样会产生三个调查项目、数据文件中的三列和统计分析中的三个变量。还有更复杂的组合，例如，问卷表格可以由针对几个调查对象的某一概念的多个维度的子问题组成，如受访者和伴侣的工作满意度的三个维度。这就导致了 3×2 的调查项目与数据文件中的六列和统计分析中的六个变量相匹配。当然，更复杂的组合也是有可能的，有更多的维度和更多的调查对象。然而，此处仍然假设受访者按照顺序逐次处理每个项目。
- 此外，还存在其他问题类型（*other question types*）。其中一些问题要求受访者同时回答更多的项目。例如，一个问题可能涉及某一概念或几个调查对象（在某个质量维度上对更多的产品进行排名）。特定的问题也出现在涉及其他各种组合的情况下，并伴随图形的广泛使用。

本书根据测量水平（名义、顺序、等距和比率）将进一步对单项和多项问题进行分类（第二层级）。

在接下来的步骤（第三层级）中，根据问题的核心版面布局（以下简称为版面布局, *layout*）来构建每个测量层级。这取决于问题的应答方式、互动功能和图形外观。例如，对于名义测量层级的性别问题，受访者可以：(a) 输入开放式文本，描述其性别；(b) 选择两个单选按钮（男性 / 女性）中的一个；(c) 在有两个选项的下拉菜单中选择（男性 / 女性）；(d) 输入数字（如“1”代表女性，“2”代表男性）；以及 (e) 选择图形（如选择代表男性或女性的头像）。对于二分法（男性 / 女性）的名义测量使用 (f) 量表是不太恰当的。然而，一些研究会要求受访者在某种量表（如 1~100、1~10 或 1~5）上表达对自己性别的看法，介于自我感知的男性化和女性化两端之间。

网络调查的文献经常将标准的 HTML 元素作为对问题类型进行分类的起点 (Bethlehem & Biffignandi, 2012; Couper, 2008; Tourangeau, Conrad, & Couper, 2013)。这些元素被设定在某些问题的选项中：单选按钮、复选框、下拉菜单和文本输入。我们也承认这些 HTML 元素的重要性，但只在问题类型分类的最后（即第三层级）讨论版面布局时考虑。本书主要根据其测量水平来概念化单项问题。这么做有四个主要原因：

- 首先，它完全符合问题编制过程中的实质性活动，即根据研究目的和问题内容初步选择测量水平。通常情况下，研究目的和内容已经决定了测量水平。只有确定了后者，才能开始讨论版面布局，包括回答格式（开放式或封闭式问题）、HTML 元素的选择和其他图形要素。
- 其次，测量水平反映了受访者认知过程的必要细节，因此尤为重要。受访者首先了解需要什么样的测量精度（名义、顺序、等距或比率），每个水平都需要特定的认知过程，而在此版面布局的影响相对较小。
- 第三，测量水平在很大程度上决定了数据文件中相关变量的格式和结构，因此也决定了调查后阶段的基础。这些对于自动分析初步报告极其重要，例如，对于名义测量，就无法计算平均值。
- 第四，测量水平具有永久性概念特征，植根于人类认知过程的本质；而 HTML 元素的作用则随着可用性和技术的发展而改变。这对于使用小屏幕的设备来说尤其如此。在这些设备上，HTML 元素的作用、重要性和感知方式都会发生改变。

后续内容详细介绍关键问题的类型、分类、确定方法，并为它们在网络调查中的使用提供指导。第2章的补充资料“网络调查中的问题和版面布局” (*Questions and layouts in web surveys*) 中介绍了相关的交互式案例，便于对此深入理解（网址：<http://websm.org/ch2>），其中还包括与某些问题类型相关的默认的统计分析。

2.3.2.1 单一项目问题

最简单的调查问题类型是针对某一调查对象的某一概念的单个维度，形成一个项目。根据应答方式，这些问题可以是开放式，还可以是封闭式。对于开放式问题，受访者原则上可以任意以文字或数字（*text or numeric*）的形式写出任何类型的答案。而对于封闭式问题，受访者需要从某类别的列表中选择答案。这个列表应该是详尽且互斥的。选择通过点击（触摸）来完成，也可以是输入对应选项的编号（或字母）。

受访者从选项列表中选择唯一答案的封闭式问题，也被称为单选题（*single answer questions*）。它们与多选题（*multiple answer questions*）不同。多选题是指受访者可以从选项列表选择一个以上的答案，通常采用复选框的形式。然而，这本质上是一个多项目问题，因为每个应答选项都算作一个项目，需要受访者单独进行判断；而且每个应答选项在数据文件中对应单独的列（即一个变量），通常在选择时赋值为“1”，不选择时赋值为“0”。所以，此类问题将在多项目问题的相关论述中展开探讨。

2.3.2.1.1 名义测量的问题

名义测量水平下的应答，只能对其进行区分，而不能根据问题所固有的实质性维度对其进行排序。因此，当将这类应答转换为相应变量的测量水平后，除了统计频率分布，可以对其进行的单变量统计分析非常有限。在此，区分开放式文本条目和封闭式格式（有不同的版面布局）特别关键。

开放式文本条目。开放式文本条目问题可以收集文本格式的回答，既可以获得较长的叙述性答案（如评论、解释），也可以获得较短的叙述性答案（如姓名），甚至获得非叙述性答案（如 URL、电子邮件地址、验证码等）。

开放式文本条目是调查应答数据中结构化程度最低的测量水平。甚至可以说，从形式上看，它们几乎不属于名义测量水平。

因此，这种类型的问题也可以从对名义测量水平的介绍中排除（Saris & Gallhofer, 2014）。当受访者的作答时，如果没有与名义测量相关的类别来辅助完成测量，我们很难将其归为名义测量水平。同样，当处理原始格式的数据时（在清理和编码之前），研究者对开放式文本答案的区分仅能停留在答案本身的差异上。通常情况下，每个回答都是独特的，不同回答的数量与受访者的数量（即单位数量）完全一致。因此，研究人员无法直接获得可通过频率分布进行统计分析的名义水平

的测量值。只有在研究者将答案清理并编码为较少数量的类别之后，才有可能在调查后阶段中进行频率分析。此时，其测量水平通常保持名义测量水平，但也可以改变。例如，如果在编码过程中根据某些标准对开放式回答进行评级（如从不赞成到赞成），可能最终采用顺序测量水平。然而，这种情况只发生在调查后阶段中；在测量阶段，问题本身没有向受访者提供测量水平的提示。

开放式文本条目的编码过程指将相同的数字代码分配给相同特征的答案。例如，要求受访者评论新服务的问题中，所有赞扬新服务速度的回答都得到相同的编码。第4.1.5节将介绍编码问题。下面介绍的几种方法，可以在制定和结构化开放式文本输入问题时使用，并可以促进编码。

- 尽可能以结构化的方式提问开放式问题。例如，要求受访者将名字和姓氏作为单独的条目填写。同样，最好是分别提问优点和缺点。有时，要求受访者写下三个主要优点和三个主要缺点也是有用的。
- 对结构化的明确要求也可以采用单独输入字段的形式，特别是当寻求列表类信息时（例如，对于“说出三家顶级餐馆的名字”，可以向受访者提供三个独立的输入子字段）。但是这已经演变为一个多项目的问题。
- 开放式文本条目可以采用额外的规范实现某种结构。例如，要求受访者完成一个句子。
- 对于与特定非叙述性答案有关的文本条目，可以通过下述方式进一步结构化：符号和子字段（例如，电子邮件地址的两个字段之间有@）、各种输入字符限制、标签、指示和验证，包括外部验证（例如，用邮政编码或URL域名数据库进行检查）。由于受访者对问题的视觉布局很敏感，精心设计的视觉布局可以鼓励受访者在理想格式中回答问题。Couper、Kennedy、Conrad 和 Tourangeau（2011）以及 Dillman 等人（2014）为开放式文本输入字段的呈现提供了进一步的实践指南，例如对输入字段进行标记、将问题结构化为组成部分、提供模板等。
- 开放式文本输入也可以作为封闭式问题的补充。除非选项能够包含所有的类别（例如行政区域），否则增加“其他，请说明”的开放式选项是很有用的。例如，当问及受访者最喜欢的水果时，先确定15种最受欢迎的水果，并通过开放式选项补充其他水果。

开放式文本问题的输入字段可以大小不同。当预期答案只有一个或几个字时（例如姓名、最喜欢的水果），输入字段可以更短，一行就够了（例如30个字符），这被称为文本框（*text box*）。文本框也用于“其他，请说明”选项，在各种封闭式

回答格式中作为最后一个类别呈现。若预期应答为较长文本（如评论、解释），则需提供更宽（如 60 个字符）和多行（如 5 行）的输入区域。这类放大的文本框通常被称为文本区（*text area*）。网络调查软件通常支持自定义本输入区的行数、宽度和最大字符数。

文字输入栏的大小对受访者来说很重要。Dennis、de Rouvray 和 Couper（2000），Smyth、Dillman、Christian 和 McBride（2009）已经确定网络问卷中较大的输入字段会增加叙述性回答的长度。然而，当受访者面对较长的文本字段时，可能会感知到更多的负担，不作答的可能性也会增加（Zuell, Menold, & Körber, 2015）。受访者自动调整文本字段的大小，也称为可滚动框（当受访者的输入内容到达底部时，系统会自动出现额外的行），或者让受访者通过拖动其右下角来扩大文本字段的大小，均可以改善上述情况。滚动框的大小也可以根据之前的答案来定制（Emde & Fuchs, 2012b）。但是，不一定所有软件都支持该选项，同时该选项也没有经过广泛测试。

激发受访者强烈的参与动机（Smyth et al., 2009）或后续探究（Holland & Christian, 2009; Oudejans & Christian, 2010）可促使网络问卷获得更多样化反应。例如，受访者提供开放式的答案后，会收到一条确认回答的信息，同时会被追问一个后续问题，如是否有任何补充。研究已经证实与自我管理的纸笔问卷相比，网络问卷中文本输入能产生更长、更详细和更有启发性的回答（Barrios, Villarroja, Borrego, & Ollé, 2011; Kiernan, Kiernan, Oyler, & Gilles, 2005）。

开放式文本条目会增加受访者的认知负担，且需要更长的时间来回答。这可能会增加低质量回答（例如无效答案）、中断（初步终止调查）或不回答问题（遗漏对某些问题的回答）的数量。有证据表明，与纸笔问卷的自我管理模式相比，上述情况在网络调查中并不更常见（Barrios et al., 2011; Kiernan, 2005; Kwak & Radler, 2002）。但文本条目限制了定量分析的开展，并涉及编码，这不仅操作繁琐、耗时和成本高，同时具有一定主观性。

当可以指定详尽的反应选项列表，而且受访者知道所有的选项时，本书就建议使用封闭式格式。但在有些情况下，开放式文本条目是不可避免的。如回答选项太多（如姓名），复杂程度太高，无法结构化（如某些行动的复杂顺序），或者当选项无法预见时（如对某一背景下最重要问题的初步询问）。因此，决定是否采用这种问题类型需要研究者认真考虑实质性和方法性的情况，详见 Krosnick 和 Presser（2010）和 Saris 和 Gallhofer（2014）。问题类型可以影响不同的结果。例如，Reja、Lozar Manfreda、Hlebec 和 Vehovar（2003）发现，与封闭式类型相比，以开放式文本输入类型提问互联网相关的问题时，得到的结果存在相当大的差异。

单选按钮。单选按钮是封闭式问题的默认呈现形式。对于需要从名义测量的应答列表中选择单一答案的封闭式问题，单选按钮是最常用的选择，也是名义测量水平下的标准布局。单选按钮是纸笔问卷中单选框的视觉等价物，因此是几种调查模式中统一设计（*unified design*）的首选（第2.1节）。另一个优点是，单选按钮作为常见的HTML元素，广泛出现在表格、预订、付款等其他网络情境中，受访者可以立即识别其功能，选择一个选项。HTML单选按钮格式的问题在于其尺寸较小，无法调整。选择时需要鼠标进行较高精度的点击操作，这对于某些特殊人群可能存在障碍（Lumsden, 2007）。在布局设计上，单选按钮可以水平放置在问题文本的下方或侧方，应答选项标签多位于单选按钮右边（至少在西方文化中），但也可以放在单选按钮的上方或下方。单选按钮也可以垂直放置，这通常是小屏幕的唯一选择，也是保证不同设备显示一致性的默认推荐方案。研究表明，即使对于顺序测量问题，水平与垂直布局的差异对结果影响极小（Toepoel, Das, & van Soest, 2009）。所以对于名义测量，布局方向对数据质量的影响可以忽略。同一研究还表明，当选项数量较多时，对选项进行分组并无显著益处，但在类别过多的情况下，分组难以避免。此时可以采用答案树（*answer tree*），通过多步骤层级选择简化操作（例如，首先选择洲，再从该大洲中选择相应的国家）。

单选按钮，应支持“取消选择”的功能，允许在需要时取消选择。但并不是所有的软件都支持这个功能，所以一旦受访者选择了某个单选按钮，可能就无法撤销，而使问题得不到真实信息。

当问题仅有两个应答选项时（如“是/否”），这类二分法（*dichotomy*）问题可采用特定的复选框布局：勾选复选框表示“是”，不勾选则默认表示“否”。虽然独立复选框（*stand-alone checkbox*）在注册表单或行政表格中常见，但在网络调查中作为单一问题布局存在明显缺陷。不勾选的行为可以被解读为“否”，也可能是漏答或拒答，具有模糊性。另一方面，对于某些系列的二分法问题，复选框布局可作为特定场景的选择，这部分内容将在多项目问题的相关章节中详细探讨。

下拉菜单。下拉菜单，也被称为下拉列表或下拉框，是另一种常见格式，适用于在名义测量水平的封闭式问题，可呈现完整应答选项列表。它出现在许多CASIC模式中，但在纸笔模式中没有类似的选项；它也是日常网页浏览中常用的标准HTML元素，所以受访者对它很熟悉。下拉菜单特别适用于处理超长列表（如包含数百个国家的列表、出生年份列表）。当单选按钮布局因选项过多无法在一屏内完整显示时，下拉菜单成为更优选择。

下拉菜单在选项列表较长时可能增加应答耗时 (B. Healey, 2007)。自动补全功能 (*autocomplete function*), 也叫数据库查找 (*database lookup*) 可发挥作用, 当受访者输入前几个字母时, 系统会从下拉列表中预测匹配的词或短语, 帮助快速定位目标选项。例如, Couper、Zhang、Conrad 和 Tourangeau (2012) 在研究中成功应用该功能, 使受访者能从数千条药物治疗条目中高效筛选目标选项。

设计下拉菜单时, 需特别注意初始显示部分: 应仅呈现标有“点击选择”或“从列表中选择”的提示文字, 而非空白或用破折号“——”填充, 且列表中的首个实质性选项不应作为默认显示项。

如果应答选项数量极多, 下拉菜单可以使用系列相关问题, 后续问题根据前序答案动态显示。例如, 在选择汽车型号时, 受访者首先选择品牌, 然后选择所选品牌中的型号。在这里, “型号”下拉菜单会主动适配所选“品牌”菜单中选择的选项。这种设计能在一行内完成复杂选择, 替代单选按钮“答案树”的繁琐布局。这种布局有时被称为下钻菜单 (*drill-down menu*)。

与单选按钮布局相比, 下拉菜单存在明显局限。作为标准的 HTML 元素, 不能添加开放式选项“其他, 请说明”和内容搜索 (*content search*) 功能, 而这些功能在选项繁多且受访者明确目标答案 (如居住国) 时尤为重要。尽管通过额外的脚本和浏览器功能 (如可编辑下拉菜单的组合框, 允许添加未预设选项, Couper, 2008, p. 120) 可以弥补这些缺陷, 但有时受访者对这些拓展功能的识别度较低。

更关键的缺陷是, 菜单未展开时受访者无法预览全部选项, 这在受访者对选项不熟悉的场景中影响显著。

研究表明, 下拉菜单可能导致数据质量问题: 既可能引发更强的首呈效应, 即受访者倾向于点击列表顶部的应答选项 (Couper, Tourangeau, Conrad, & Crawford, 2004), 也会因需额外点击展开列表而增加应答时间 (B. Healey, 2007)。在一般情况下, 当有足够的空间和反应选项不多时, 下拉菜单与单选按钮相比没有任何优势 (Tourangeau, Conrad, & Couper, 2013, p. 93), 所以较少使用下拉菜单。

数字输入。有时可以使用带有数字输入的封闭式问题, 即受访者输入一个数字, 该数字对应名义测量水平上、一组固定选项的某个反应类别。例如输入某个众所周知的数字, 表示一个地区或邮政编码; 也可以要求受访者输入相应的数字 (例如“1”代表男性, “2”代表女性)。因为没有单选按钮方便, 选择数字输入应该有充足的理由。但当无法使用鼠标, 或者网络问卷用于使用外部编码列表记录行为时, 数字输入也有特殊优点。该方法通常用于纸笔问卷的数据输入, 以及 CAPI 和 CATI 调查中。

与通过点击（或触摸）来选择类别相比，输入数字的速度较慢，容易出错，而且不够直观。此外，应该禁止在反应选项范围之外输入，这一点本书将在等距和比率测量中详细讨论；在这些测量中，开放式数字输入是默认的布局格式。除数字外，也可以输入字母来表示所选择的类别（例如，用“A”“B”代表学校成绩；或者用“M”代表男性，“F”代表女性）。

高级图形呈现。采用名义测量的封闭式问题也可以通过各种图形化设计来实现。例如，单选按钮可以用特殊视觉设计的按钮（如单选按钮的大图像）或其他图像（如星星）代替。另外，文字描述的答案可以完全或部分地被某些图形或多媒体元素（如音频、视频）所替代，例如在选择最佳视频、最佳标识等时。

在极端案例中，单选按钮和文字回答都可以被选项的直接可视化所替代（例如，在选择最佳水果时，用不同类型的水果图片），通过直接点击特定的图片来完成选择（没有单选按钮）。特殊情况是带有图像区域选择的问题，即可点击的图像地图（*clickable image maps*），受访者在预先指定的图片区域中选择一个点，例如地图上的一个区域或人体图片中的某身体部位。这种形式在教育（如测验、电子学习）和市场研究中特别常见，有时被称为热点问题（*hotspot question*）。

高级的图形选项是使用拖放功能（*drag and drop*），这在市场研究的网络调查中很常见。在名义测量水平的情况下，受访者需要用鼠标选择某个类别（主题，通常以图片形式呈现），并将其拖放到某个预先选定的位置。与简单的一键选择相比，该案例所需的鼠标动作明显增加，但使用这种方法通常可以提高受访者的参与度。

单选按钮通常以标准化的方式在不同的技术（设备、浏览者和界面）中呈现，但多种技术不一定总支持图形呈现方式。然而，随着移动设备的普及，现今已经出现新的图形化问题。

本书将在后续内容介绍使用图形问题的案例，在此需要特别注意图形对受访者的回答可能产生非预期的影响。

2.3.2.1.2 顺序测量的问题

名义测量（如选择居住国）和顺序测量（如频率类选项“从不、有时、经常……”或态度类选项“非常不同意……非常同意”）的封闭式类别回答之间存在本质性和方法上的区别。与名义测量不同，顺序测量可以计算中位数、最小值、最大值和斯皮尔曼等级相关系数。在实践中，有时甚至会把顺序测量的数据当作是等距测量，这样也可以计算出平均数和方差。如果进一步假设反应是正态分布，就可以应用一

系列的多变量统计方法。

然而,这种重要的差异不一定反映在问卷调查问题的视觉布局上。例如,名义测量与顺序测量可能采用相同的单选按钮布局。下文将重点阐述,当相同布局用于顺序测量时所呈现的方法学特征。

对于顺序测量水平,事实性和非事实性的问题都可以适用。对于非事实性问题(如态度),经常以等级评定量表(*rating scales*)的形式呈现,受访者需在封闭的有序类别列表中对潜在概念进行评分(即赋值)。这些概念可分为单极概念(*unipolar concepts*)(例如强度,从低到强)和双极概念(*bipolar concepts*)(例如同意程度,从极端同意到极端不同意)。

等级评定量表常被泛称为“李克特量表”,特别是在表示5点等级评定的时候。然而,本书避免使用这种标签,因实际应用中其内涵与李克特的原始定义存在偏差(Neuman, 2009, p. 207)。

在实践中,包含5个及以上应答类别的评分量表在分析中常常被视为等距测量水平。前提是各个类别之间的视觉距离是相等的,这样才能类似于等距测量水平。此外,还应尽力使数据分布趋近正态分布。例如,如果所有受访者在被问及是否同意“食物非常好”这一陈述时都倾向于“非常同意”,可以考虑将问题改写为更有力的表述:“食物堪称绝佳”。

本书对顺序测量水平下问题布局的讨论,将主要集中在等级评定量表上;然而,它也将涵盖顺序测量中与事实性问题相关的其他应用(如频率、观察到的等级等)。

单选按钮。单选按钮是等级评定量表和一般顺序测量的首选布局格式。单选按钮具有在名义测量中介绍的所有特征。此外,等级评定量表也面临着许多源于传统调查模式的一般方法困境,而这些在网络问卷中又呈现出某些特殊性。由于这些困境对量表构建非常重要,下文将简要介绍其核心特征。

- 选项类别可以是数字(如1、2、3、4、5),也可以是文字(如同意、不同意……),还可以同时使用两种标记。在特殊情况下,数字也可以是负数(如-2、-1、0、+1、+2),但一般不建议使用负数,因为受访者可能倾向于避免负值(Toepoel et al., 2009)。除此之外,还存在其他的组合,如只标记两个极端(强烈同意和强烈反对),而其余点不做标记。这取决于具体情况,但一般而言,优先使用全文字标记(Dillman et al., 2014; Krosnick & Presser, 2010, p. 275)。
- 研究建议反应类别的垂直布局比水平布局要好(Krosnick, 2013),但这有可能超过所需的额外空间。
- 在垂直布局中,建议采用选项类别递减的顺序(例如从好到坏)。受访者通常

期望先看到积极的内容,且这种方式更常见,因此给受访者带来的认知负担更小(Holbrook, Krosnick, Carson, & Mitchell, 2000; Tourangeau, Couper, & Conrad, 2013)。这原则上意味着,水平布局也应该从积极类别开始。但这与数学中使用的坐标系统,以及日常实践测量长度、时间、重量等从低到高(例如从0到100)的做法相反。目前几乎没有证据表明,选项类别的方向对水平布局很重要。然而,Toepoel等人(2009)发现,虽然差异很小,但具有统计学显著性;同时有一些证据表明,垂直格式会产生首呈效应。

- 建议使用奇数点(如5或7),以便提供中间点的反应选项(Krosnick & Presser, 2010),特殊情况例外。
- 关于类别的数量,研究表明,超过7个类别数据质量的收益会变小,甚至出现质量下降(Malhotra, Krosnick, & Thomas, 2009)。Malhotra等(2009)建议单极概念采用5点量表(例如从低强度到高强度),双极概念采用7点量表(例如从极不满意到极满意)。然而在某些情况下,其他的计分方式可能是合理的,特别是2点、3点、4点、6点、10点、11点,甚至101点量表。一些重要的调查(如ESS)¹¹采用了11点计分的量表。目前研究还没有得出一致结论,因为为有研究支持类别数量少合理(Alwin, 2007),也有研究支持类别数量多是合理的(Saris & Gallhofer, 2014)。

关于文字标签,使用常用量表措辞(如同意—不同意)往往显得非常实用。但有研究(Krosnick & Presser, 2010)显示了内容和结构特定量表的某些优势(更少的默许,更高的信度和效度)。例如,不采用“是否同意食物美味”(不同意—同意)的方式,而是要求受访者对美味程度从极不美味到极美味进行评分。其他问题类型也是如此,所以应该使用自然指标法(*natural metrics*),而不是一般方法。例如,具体询问事件的频率(每天、每周、每月……),而不是使用常用的类别(非常经常、经常、不那么经常……)。

上述问题起源于纸笔模式的自我管理问卷,但对网络调查同样适用。当涉及问题和回答选项的视觉呈现时需要特别注意,因为受访者在认知层面解释类别之间的距离时会受到类别间距(即单选按钮)的影响。因此,如果计划在分析中使用带单选按钮的顺序测量作为等距测量,应该确保类别之间有相等的间距(Tourangeau, Conrad, & Couper, 2013, p. 79)。如果网络调查软件将列宽与应答标签的宽度相匹配,从而在整个量表中引入不一致的宽度,就会影响到这一点。

在网络调查中,由于与实时验证和提示的潜在交互,对于非实质性回答的问题(不知道、没有意见)还存在一些特殊情况,相关内容将在第2.3.3节讨论。

有各种评定量表的亚类型，如水平方向呈现两个选项，让受访者对二者进行比较（*compare one against another*）；语义差异（*semantic differential*）采用水平两极形式，仅对量表两端的选项进行标注，例如“丑陋－美丽”；或斯塔普量表（*Stapel scale*），属于单极 10 点垂直量表，以数字标注从－5 到＋5 的区间，且不存在中间点。

下拉菜单。在顺序量表中，更加不建议使用下拉菜单，理由与在名义测量的介绍相同。由于在等级评定量表和其他顺序测量水平的问题中，类别的数量通常相对较少，因此下拉菜单的优势——在有大量类别时使用，便不复存在。

数字输入。与名义测量水平一样，受访者可以在封闭式数字输入框中输入答案，以进行初步的测量。其限制与细节也同名义测量相同。然而在某些情况下，输入数字是非常简单和直观的，如学校成绩（如 1、2、3、4、5）。当然，超出反应选项范围的条目应该被禁用。

高级图形呈现。顺序测量问题的图形呈现与名义测量相同，此时图形和多媒体可以替代单选按钮、应答标签或两者。顺序测量中出现了更具体的替代物，如笑脸或拇指。例如，使用不同面部表情（从最不高兴到最高兴）表示对服务的满意度。有研究指出，在等级评定量表中使用图像的结果是积极的；特别是在测量快乐、唤醒和支配（SAM 量表）的心理学研究中，图像被认为是有益处的（Bradley & Lang, 1994）。但 Emde 和 Fuchs（2012a）并没有发现在网络调查中使用面部图像有任何优势，反而会造成作答延迟和应答分布异常问题。本书在后续内容中会进一步阐述图形、多媒体和图形呈现。

在顺序测量中，还可以选择使用图形线或色带呈现可能的反应，即在一条线上以图形形式设置若干离散点（如 5 个），供受访者选择对应数值。需要说明的是，这种呈现方式并不等同于完整的连续刻度。

2.3.2.1.3 等距和比率测量的问题

此节合并介绍等距和比率测量水平的内容。与顺序测量相比，等距测量水平下可以比较不同的值。比率测量与数字量（如年龄、收入、体重、身高、速度、时间等）有关，除了具有等距水平的特点，还有绝对零点，这使得反应值可以作为比率进行比较。原则上，比率测量是最受欢迎的测量水平，因为它可以使用的统计方法最广泛，而且除了一般的单变量和多变量统计，还可以使用相对量（*relative measures*）（例如变异系数，*coefficient of variation*）。

此处合并探讨等距和比率测量，是由于在测量阶段二者的差异甚小。在编制问题和布局实施阶段，这两个测量水平之间也几乎没有差异。

前文已提及，在分析阶段，有时可以将等级评定量表和其他顺序测量水平的问题视作等距测量数据进行处理。

开放式数字输入。开放式数字输入，即以数字形式输入回答，是等距和比率测量中最常见的问题布局选项，因为它提供了高精确度。原则上，我们期望对数字输入没有任何限制（如工资）。然而实际上，开放式和封闭式数字问题（从一组封闭的选项（如出生年份）中输入一个数字）之间的界限有时是模糊的。

输入字段的大小和格式也非常重要，类似于开放式文本输入。定制化的输入字段（例如询问子女数量时，采用两位数字的输入框而非十位数字的输入框）可以产生更精确的答案（Couper et al., 2011; Couper, Traugott, & Lamias, 2001）。因此，通过指定格式（如明确的货币符号 \$、€、£……）、小数点、长度和子字段（如年 - 月 - 日、小时 - 分钟 - 秒）来设计输入字段是很有益的。在此，实时验证（见第 2.3.3 节）至关重要，可有效避免数字输入超出合理范围（如年龄输入为 572 岁）。

在涉及等距和比率测量（例如工资）时，封闭式的回答格式，例如单选按钮和下拉菜单，原则上劣于开放式的数字输入；开放式输入更加精确和中立（Krosnick & Presser, 2010, p. 267）。当反应类别非常有限（如孩子的数量）时，单选按钮的布局会让作答更快。如果有更多的类别（如年龄），使用下拉菜单会更不容易出错。此外，有时我们可能会考虑在量表的预期范围内有聚合值的类别。例如，如果知道在数据分析中只使用三个年龄的编码组（如 <30, 30~49, 50<）（在顺序测量水平上，意味着失去了测量水平的精确性），最好使用带有三个年龄组的封闭式问题，而不是询问出生年份并在以后重新编码。从受访者的角度来看，封闭式的问题带来的负担较小。然而，事先分组都需要非常谨慎；因为有很多研究发现，在封闭式格式中对数值的不当分类会使结果产生偏差（Dillman et al., 2014, p. 161）。

连续量表。原则上，连续量表仅适用于反应值在连续体上具有无限数量可能、而非局限于有限封闭式选项的情况。具体而言，它要求受访者能够在精细的等距或比率测量水平上对答案进行观察与表达，但这种情况实际中并不常见。当涉及态度类问题时，其测量属性存在争议——研究者需明确相关概念应采用顺序测量还是等距测量。一旦使用连续量表，则默认所测项目属于等距或比率测量水平。

连续量表通常以简洁的图形形式呈现，由受访者在一条线上标记其选择的位置。这种方法起源于传统调查，并在传统调查中以“线生成法（line production）”

形式存在 (Saris & Gallhofer, 2014, p. 109), 目的是提升测量精度, 尤其是可避免数字四舍五入的问题 (例如以 10 或 5 的倍数作答)。四舍五入问题在采用开放式数字输入格式时经常发生。网络调查的优势在于, 图形化反应会自动转化为数字值 (如 1~100), 并可以选择实时向受访者显示该数值。

在网络问卷中, 连续量表最常见形式是视觉模拟量表 (*visual analogue scale*, VAS), 即呈现一条水平线和两端标注文字标签 (例如同意—不同意), 主要应用于健康或婚姻研究。受访者用鼠标点击选定相应位置 (Reips & Funke, 2008)。若在线上其他位置添加标签 (如“弱痛、中痛、强痛”) 或将线段划分为若干分段, 则称为图形评定量表 (*graphic rating scale*, GRS)。另外, 可以在选定的位置设置手柄, 以便受访者向左或向右拖动它, 这有时被称为滑动条 (*slider bar*)。如果在受访者选择答案之前, 手柄被放置在量表的任何位置 (例如量表中间), 那么滑动条的呈现方式就会引发问题, 导致无法区分受访者选择的是中间答案, 还是漏答了项目 (Funke, Reips, & Thomas, 2011, p. 223)。

连续量表有时优于单选按钮 (Funke & Reips, 2012; Reips & Funke, 2008; Torrance, Feeny, & Furlong, 2001), 因为它们更精确, 且受访者应答中的极端值和中值占比更少。Saris 和 Gallhofer (2014) 也主张使用连续量表, 但是建议在实施时要谨慎, 要注意受访者对问题的感知是否有很大差异。也有研究发现连续量表存在弊端。Couper、Tourangeau 和 Conrad (2006) 的研究表明, 使用连续量表比使用单选按钮耗时更长, 中断率和数据缺失率更高, 且在提供数字反馈时更易出现四舍五入问题。心理学研究 (Flynn, van Schaik, & van Wersch, 2004) 和 Funke 等人 (2011) 的研究中也发现了类似的问题, 特别是在受教育程度较低的受访者中。

由此看来, 连续量表更适合电脑操作熟练且能真正区分细微差别的受访者。否则, 连续量表的优点难以显示, 反而存在诸多缺点。因此, 研究者在使用连续量表时, 必须具备明确且充分的理由。

高级图形呈现。连续量表本身已涉及图形呈现, 但除上述视觉模拟量表 (VAS)、图形评定量表 (GRS) 和滑动条外, 还可包含其他图形呈现形式, 从替代文字描述的图片、量表上的数字标注到各类动画及多媒体元素。与其他问题类型相同, 本书建议使用图形时保持谨慎。

上文主要介绍了不同测量水平下单一项目问题的关键布局格式。表 2.1 总结了各种可能的组合。实心圆圈意味着默认的选项, 而空心圆圈指在某些情况下可能是合适的选项, 但使用它们需要明确的理由。其他组合则被认为是不可能或不合理 (例

表 2.1 不同测量水平下单一项目问题的关键布局

测量水平	关键反应格式					
	开放式文本输入	单选按钮	下拉菜单	数字输入	连续量表	高级图形
名义测量	○	●	○	○		○
顺序测量		●	○	○		○
等距或比率测量		○	○	●	○	○

如，名义测量水平的连续量表）。

部分空心圆圈对应的组合较为罕见，如名义测量中的数字输入；而其他组合则更为常见，如名义测量中的开放式文本输入。需要说明的是，截至目前本书仅介绍了单一项目问题，未涉及一系列开放式文本输入（如收集清单）的情况。此处指的是受访者提供答案阶段的测量水平，而不是统计分析阶段的测量水平（即变量的测量水平）。例如，在开放式文本输入问题中的回答可以被进一步编码，并作为顺序测量水平处理。同样，采用顺序测量法收集的数据，统计分析中有时可以被当作等距测量水平处理。

2.3.2.2 多项目问题

实践中，研究者常将问题与单个项目组合，以节省空间或加快受访者的填答速度。节省空间的理念源于纸笔模式，主要是为了降低印刷成本。网络调查中，空间本身不再是直接限制因素。但将相似问题归为一组，可以提高应答速度以及减轻受访者负担，这仍然是对问题进行分组的主要原因。此外，还可以将这些问题归入相同的背景。但研究者需要权衡潜在优势和数据质量下降的风险。与单一项目问题相比，复杂性的增加可能导致数据质量降低。

当多个单一项目问题被组合后，会形成包含子问题的复合问题，本节将重点讨论这类多项目问题。这样的问题组有时被称为矩阵（*matrix*）（Dillman et al., 2009, p. 179）或网格（*grid*），但网格通常表示具有相同应答选项的项目组合（Tourangeau, Conrad, & Couper, 2013, p. 72），主要用于搭配单选按钮布局的评定量表。对于一系列具有相同应答选项的项目，另一种表述是“问题矩阵”（*battery*）（Alwin, 2007; Saris & Gallhofer, 2014, p. 86）。本书倾向于用表格（*table*）这一概念来表示问题的分组或组合。

表格形式的问题存在诸多方法学挑战。此处首先讨论简单表格（*simple tables*）。简单表格由一系列关于某一同质问题的多个维度或多个主题组成，每个项目作为单独问题被分组呈现，且采用相同的布局设计。简单表格与矩阵和网格的实际用法大

致相近。因此，下文介绍的是具有同质性问题的简单表格。为方便起见，除非另有明确说明，将统一使用“表格”指代这种形式。至于更复杂的表格类型后续内容将进行讨论。

下面讨论的表格案例和插图在第2章的补充文件“网络调查中的问题和布局”中提出 (<http://websm.org/ch2>)。

2.3.2.2.1 名义测量问题的表格

首先介绍名义测量水平下相似单一项目问题的分组。布局设计尤为关键，因其直接决定表格的呈现形式。

开放式文本条目的表格。当带有开放式文本条目的类似项目被合并成一行（或一列）时，会形成单维表格，此时更适合称为“文本条目系列或矩阵”（*series or array*）。例如，可将姓名、家族姓氏和地址这三个文本条目放在同一行，分别对应三个测量项目，从而在数据文件中产生三列，在分析中产生三个变量。这种设计的优势在于节省空间，且无需重复问题的引导语，仅需在相应输入字段旁添加标签（如“姓名”“地址”）即可。若需针对5个家庭成员重复一系列文本条目，每个人都位于单独的行，则会形成二维文本条目表。在5个人的情况下，会有 $5 \times 3 = 15$ 个测量项目（数据文件中的15列和统计分析中的15个变量），通过简洁的界面整合呈现，既节省空间又便于用户理解。这种布局常见于家庭名册调查或社会网络测量，如改良后的布局（*alter-wise layout*）。针对每个网络成员（如朋友），依次收集其信息（姓名、家族名称、地址）并分配相应的答案（值）。另一种方式是项目式布局（*item-wise*），即为每个项目（如姓名、家族姓氏、地址）设定行，再为所有网络成员（如朋友）分配对应数值。然而，目前尚无明确证据表明，与针对每个家庭成员单独设计一系列问题相比，开放式文本条目的表格布局能显著提升或降低数据质量（Dillman et al., 2009, p. 180）。

单选按钮的表格。当受访者在水平单选按钮布局中从三个类别（如就业、未就业、退休）中选择就业状况时，这属于单一项目的单一问题测量。多名被试报告同一类别时，例如5个家庭成员，将家庭成员列在表格的不同行中呈现会更便捷。这样只需要一个指导语和一套应答标签。尽管表格中包含 $5 \times 3 = 15$ 个单选按钮，但其本质是通过一个维度（就业状况）测量5个对象（家庭成员），因此仅产生5个项目，对应数据文件中的5列和数据分析中的5个变量。与5个单独的问题相比，这种设计能加快应答速度并节省问卷空间，但需警惕复杂性增加导致的受访者注意力下降。

本节会在介绍顺序测量的表格时再来论述这个问题。若在表格中加入更多维度,得到双表或三表(*double tables or triple tables*)的单选按钮,例如在同一个表中加入5个家庭成员的性别单选按钮问题,则会形成真正的二维表,测量 $5 \times 2 = 10$ 个项目(5个家庭成员的2个社会人口统计学特征),数据质量风险也会随之升高。

下拉菜单的表格。在上述就业状况测量的例子中,使用下拉菜单布局几乎无优势,因为下拉菜单本身仅适用于单一项目测量。如果需要节省空间,可在一行中通过下拉菜单布局展示额外维度。例如,在一行中为每个受访者(如家庭成员)设置就业和性别的下拉菜单,形成一个 $5 \times 2 = 10$ 项的二维表。但这种布局复杂度高,易增加受访者负担并分散注意力,因此需有充分理由方可采用。

数字输入表格。上述就业状况测量也可通过封闭式数字输入框实现(如“1”表示就业,“2”表示失业,“3”表示退休)。但如前文对单一项目问题的分析,这种布局存在明显缺陷。虽然能在一定程度上节省空间,但其弊端与下拉菜单类似,需谨慎使用。

系列二分法问题。当存在一组二分法(是/否)问题时,组合呈现尤为便捷,例如询问受访者是否拥有某些商品、具备某些特征、有过某些经历或同意某些陈述。例如,可将“是否访问过四个国家”的问题设计为:仅显示一次问题文本,将四个国家作为子问题列出。此时有多种布局选择:系列“是/否”单选按钮、系列复选框或多选框。

第一种是系列是/否单选按钮(*series of YES/NO radio buttons*),每个国家单独成行,右侧放置两个单选按钮(“是”和“否”)。这种设计比为每个国家重复问题文本和应答选项更简洁。但是,对于“去年是否访问过28个欧盟国家或美国50个州”这类问题,系列“是/否”单选按钮并非理想解决方案,因为受访者需重复点击大量“否”选项。

在二分法的情况下,复选框(*checkbox*)是系列单选按钮的替代布局方案。类似于单选按钮和下拉式菜单,复选框也是一个标准的HTML元素。如前文所述,单一的复选框在网络问卷调查中很少使用,因为与单选按钮布局中选择“否”相比,复选框中选择“否”的含义是不明确的,即不勾选的选项可能是因为漏答或拒答。复选框在网络调查中可用于获得受访者的知情同意(例如“请检查您是否同意本研究的条件并愿意参与”)或允许他们注册并获取结果(例如“请检查您是否愿意收到本研究的结果”)。

另一方面,把系列复选框(*series of checkboxes*)放在一起,在网络问卷调查中较常见。这种布局常被简称为复选框问题(*checkbox question*)、复选所有适用或多

选问题 (*a check-all-that-apply or a multiple answer question*)。在上述“访问国家”的例子中,每个国家旁设置一个复选框,受访者仅勾选已访问过的国家,未勾选则默认表示未访问。尽管看似是单一项目问题,实则通过一个概念(访问)测量多个主题(国家),项目数量与主题数量一致,对应数据文件中的列数和分析中的变量数。

若限制复选框问题的选择数量(如最多选两个),则问题会更复杂,受访者需同时考虑所有选项,本质上属于排序问题,本节后续将进一步讨论。

尽管复选框是 HTML 元素,受访者通常知道使用复选框可以选择一个以上选项。但为了让受访者清晰了解,强烈建议明确说明可以选择几个答案。

复选框的潜在优势是,在某些情况下完成速度更快(Callegaro, Murakami, Tepman, & Henderson, 2015),而且比“是/否”单选按钮节省空间。在“是/否”单选按钮中,受访者明确面对所有的选项,需要为每个选项选择一个答案,所以有时也被称为强制选择题(*forced choice question*)。

然而,复选框也存在明显缺陷(Bradburn et al., 2004, p. 171)。在混合模式的调查中,引用复选框可能引发问题。因为在电话调查等听觉模式调查中,提出多个问题的标准方式为“是/否”的形式;而网络调查的复选框布局可能改变问题语境,引入模式间的方法学差异。

其次,系列复选框中的未勾选项目有多种解释。通常情况下,未被选中的项目被视为“否”。然而受访者可能漏答或拒答,实际上是没有回答。受访者还可能不确定,所以未勾选项目实际上意味着“不知道”。可通过在复选框列表末尾增加“以上都不是”或“其他,请说明”选项减少歧义,帮助识别未应答项目。

一项关于多个随机实验的元分析,比较了一系列复选框和“是/否”单选按钮的布局(Callegaro et al., 2015),发现“是/否”格式认可度更高。但在两种格式中,项目排序和相对顺序仍然是保持一致。

综上,本书更倾向于使用系列“是/否”的单选按钮,因为它能更清楚地识别项目遗漏情况,促使受访者认真思考每个选项,且在不同的调查模式下更具有可比性。使用复选框问题应有明确的理由。一种情况是,类别明确且易理解(例如种族起源问题)。另一种情况是研究主要关注项目之间的排序和关系,而不是精确的绝对比例(即精确的认可率或市场份额)。在这种情况下,减少受访者的时间压力使复选框更受青睐。当选项类别数量较多,而其中只有少数类别与每位受访者有关时,可以选用复选框。

多选框(*multiple selection box*)是系列是/否二分法问题的另一种变体,有时也被称为多选下拉菜单、多项选择甚至是组合框问题。其外观类似展开的下拉菜单,但是受访者可以通过按住 CTRL 键从列表中选择多个回答,而标准下拉菜单不支持

此功能。从视觉上看, 多选框与下拉菜单不同, 因为无需受访者采取任何操作就能立即显示很多选项。选项过多时会出现一个清晰可见的滚动条, 可以用来查看其他选项。在上述访问国家的案例中, 可以通过一个单一多选框来选择所有访问过的国家, 而非系列“是/否”单选按钮或一系列的复选框。这里强烈建议明确提示受访者使用 CTRL 键, 因并非所有用户都熟悉该操作。多选框的界面复杂, 用户熟悉程度降低, 使用这种布局需要明确的理由。与下拉菜单类似, 仅在选项数量多且空间限制严格时考虑使用。一个重要的改进设计是, 左侧显示所有潜在选项, 受访者通过拖放功能将选中项移至右侧框中。

复选框和多选框的表格。上面所有关于系列二分法问题的布局案例, 主要针对单一概念在多个维度上的测量。国家访问的案例中, 有4个国家对应4个测量项目(数据文件中的4列和分析中的4个变量)。更为复杂的情况是, 相同问题针对更多的受访者(如5个家庭成员)重复。系列复选框(复选框表, *table of check-boxes*), 会出现有 $5 \times 4 = 20$ 个项目的二维表。

如果为每个主题使用一个多选框(多选框表格), 则可进一步节省空间, 因为一个单列有5个多选框即可, 每个家庭成员一个。甚至可以再增加一列多选框, 例如, 每个家庭成员会说所列语言(英语、意大利语、斯洛文尼亚语)中的哪一种, 总共就会产生 $5 \times 4 + 5 \times 3 = 35$ 项。但若受访者对多选框的熟悉度低, 其缺陷会更加突出, 且这种情况下无法采用“拖放至另一框”的布局。

2.3.2.2.2 顺序测量问题的表格

与单一项目问题类似, 本书对顺序测量问题表格的介绍主要集中在等级评定量表表格上。这一研究重心的设定甚至更具合理性, 因为大量研究都是围绕等级评定量表的表格(也被称为矩阵、问题组或网格)进行的。

要强调的是, 与等级评定量表表格有关的许多方面(例如行和列的数量)都与一般的表格布局有关。

带单选按钮的等级评定量表表格。单选按钮布局在顺序测量表格中占主导地位。每个项目, 即被测概念的每个维度或每个主题, 单独成行显示。例如, 要求受访者对4个国家的总体旅游体验进行5点量表的评分。通常情况下, 问题的措辞描述放在左边, 重复的反应选项(即单选按钮)放在右边; 如果颠倒布局, 应答类别在行上, 维度在列中, 效果通常比较差(Galesic, Tourangeau, Couper, & Conrad, 2007)。应答选项的标签只在列顶部的标题中显示一次。正如在名义测量表格部分, 此处也主要

关注一系列单一等级评定量表。

与等级评定量表项目一样，应该注意表格中应答选项之间的列距，因为不均匀的空间会影响反应分布（Tourangeau, Conrad, & Couper, 2013, p. 79）。

当列和行的数量较多时，需要格外谨慎，这是所有类型表格的共性问题。即使需要减少刻度点的数量（例如从 7 减少到 5），也应该避免水平方向的滚动（左右方向）。垂直滚动（向上或向下）影响不大，但最好还是避免；因为向下滚动长表时，顶部的列标签会消失。为了在常规电脑屏幕上显示整个表格，每个表格的行（项）的维度数限制在 8~10 个。另一个解决方案是编程实现固定标题功能，即表格的标题是固定的，只滚动项目，或者在一定数量的项目之后重复标题标签。由于长的表格也会增加受访者的疲劳感，总体而言弊大于利。

建议用交替的阴影标记表格中的行，以更好地在行与行之间区分。例如，Crawford、McCabe 和 Pope（2005）建议在交替的行中使用浅灰色背景阴影来增强可读性。动态阴影，如行的字体颜色会改变，或者在选择回答后显示阴影，这会成功减少不作答的项目（Galesic et al., 2007）。但预先设定的阴影和鼠标悬停高亮效果可能会产生负面影响（Kaczmirek, 2011）。一般来说，设计表格格式时，应该避免任何冗余和视觉上的杂乱，以便表格的设计能够达到其基本目的：使受访者以用户友好的方式提供高质量答案。

正如 Lozar Manfreda 和 Vehovar（2008）的总结，带单选按钮的等级评定量表表格可以节省空间，使问卷看起来更短，方便制作；且因减少引导语和应答标签而降低受访者负担、减少鼠标移动、缩短应答时间。但是，正如多项研究所表明的，这种表格也存在诸多缺陷。

- 表格复杂性增加会加重受访者认知负担，导致其采取各种“满足策略”（如答案区分度较小），导致数据质量下降（Fricker, Galesic, Tourangeau, & Yan, 2005; C. Zhang, 2013）。
- 研究表明（Lozar Manfreda, Batagelj, & Vehovar, 2002; Toepoel et al., 2009）或担心（Couper, Tourangeau, Conrad, & Zhang, 2013, p. 113），与一连串的单一个问题相比，表格会增加项目无应答。
- 表格可能会改变问题的性质。项目被置于比较性语境中，易引发“语境效应”（*context effect*）。这些影响很弱，但在网络调查中仍然存在（Tourangeau et al., 2004）。
- 有报告指出表格是受访者放弃调查的最常见节点（Henning, 2011; McMahon & Stamp, 2009）。

目前尚无系统研究比较带单选按钮的等级评定量表表格的潜在负面影响（如数据质量降低）与一系列单选按钮问题的实际缺陷（如应答时间延长、问卷篇幅增加）。造成这一空白的部分原因是存在混淆因素。例如，表格布局对中断率的影响在问卷的开头和结尾会有所不同。同样，表格布局效应对在线样本库的“专业”受访者和“新手”受访者的满足策略影响也有差异。本书还建议谨慎对待关于应答时间的研究。因为应答时间常受到页面设计的影响，如每个项目作为一个独立的问题被放在一个单独的网页上。当表格与同一页面上的一连串单项问题进行比较时，应答时间差异微小（Bell, Mangione, & Khan, 2001），尽管后者导致问卷长度增加一倍。而将每个问题放在单独的一页上会导致应答时间增加 50% 以上（Callegaro, Yang, Bhola, Dillman, & Chin, 2009），这主要是由分页操作导致，与表格布局没有直接关系。

展开的表格—水平滚动矩阵。水平滚动矩阵（*horizontal scrolling matrix*, HSM）是表格布局的替代方案，在表格中用单选按钮来呈现子问题。子问题（即项目）依次在屏幕上显示，应答选项（单选按钮）以垂直或水平方式排列。受访者每次只看到一个子问题，回答完前一个自动显示下一个，无需点击“下一步”按钮。导航功能带有项目计数器和视觉进度条，便于受访者掌握应答数量和进度。这种形式在减少语境效应和降低复杂性方面显示出很好的效果（Klausch, de Leeuw, Hox, de Jongh, & Roberts, 2012），尤其适用于小屏幕设备（如智能手机）。某些网络调查软件在检测到小设备时，会自动将表格转化为这种独立问题形式（见第 5.1 节）。

其他顺序测量问题的表格。下拉菜单、开放式数字输入和高级图形呈现也可作为顺序测量问题表格的替代布局。这些形式不仅适用于带有等级评定量表的简单表格，也适用于其他带有顺序测量的问题。其局限性已在单一项目中讨论，同样，所有关于表格布局的一般问题也适用。

2.3.2.2.3 等距和比率测量问题的表格

在等距和比率测量水平中单一项目问题的潜在布局，如单选按钮、下拉菜单、数字输入和连续量表的具体特征，也适用于相应的表格设计。就布局本身而言，与表格有关的一般原则也适用。需补充的是，与更常见的单选按钮布局相比，连续量表的表格外观较繁琐，可能降低普通人群的参与意愿。

开放式数字输入格式（*open-ended numeric entry format*）是等距和比率测量水平的默认表格布局，与开放式文本输入的问题密切相关。这与在名义测量水平的表格中的问题类似，即一系列的开放式数字输入（如高度、宽度和长度）与一系列的

开放式文本输入（如姓名、家族名称、地址）的布局非常相似。

日期输入是经典案例，基本上是由三个数字问题组成的简单系列，常被用作讨论布局设计如何影响应答的范例。总而言之，一个由3个数字输入框组成的“年、月、日”系列，并在输入框旁标注“DD”“MM”和“YYY”符号，是较好的解决方案（Christian, Dillman, & Smyth, 2007; Couper et al., 2011b），下拉菜单也可以作为替代方案（Couper et al., 2011b）。

2.3.2.2.4 组合表格

上述案例中，同一测量水平、同一布局格式的问题被组合为表格，形成简单表格。然而，随着组合表格（*combined tables*）（有时也称为三维表格）结合更多的测量水平和布局格式，复杂性会增加。例如家庭名册调查中，对于每个人，一行内同时包含输入一个开放式文本条目（姓名）、一个开放式数字条目（年龄）和一个下拉菜单（性别）。此外，这些条目还会整合为一份完整的家庭成员信息表。一般而言，应对此类复杂性需保持谨慎，建议将组合表格拆解为几组较简单的任务集合。但在某些场景下，组合表格节省空间与时间的优势可能超过其潜在弊端。这对于商业调查中广泛使用的事实性问题尤其如此（例如，财务表中按年份划分的项目条目表）。另一种情况是，组合表格可充分匹配纸质问卷的平行表格形式，但这未必是网络调查问卷的最佳解决方案。Morrison 等人（2010）针对这种情况制定了通用的可用性原则，在网络问卷调查中应该仔细执行严格遵循。

网络问卷中是否使用表格并无统一答案，但仍可总结出若干结论。尽管这些结论主要是基于最常用的带有等级评定量表的简单表格，但它们也普遍适用于其他表格布局。大多数情况下，设计良好的表格可以节省空间，但是实施不当会严重影响数据质量。即使设计合理，也有研究表明表格会引发问题（如语境效应、项目无应答、满足应答）。另一方面，没有明确证据表明使用表格的显著优势，只有未经证实的观点认为表格可以直接节省应答时间（因为表格空间明显减少），或提高响应率和数据质量。因此，一般建议是尽可能避免使用表格，或至少尽量减少使用数量。由于表格会增加问卷难度，其对数据质量的损害可能超过略长的简单问题序列（Burdein, 2014）。尽管相关研究较多，但仍缺乏全面对比表格形式与评级标准表的研究，而这两者实际上密切相关。同样，所有关于表格布局的一般性问题也适用于此类场景。

2.3.2.3 其他问题类型

上面提到的问题类型不仅仅涉及单一项目问题或其组合（简单表格），还存在

其他各种问题类型。当受访者需要同时考虑更多项目（维度、主题）或涉及大量图形时，或当问题组合并在特定情境中实施时，它们就显得尤为重要。由于其复杂性，本书没有对其进行系统化的梳理，但下面列举一些常见的案例。

2.3.2.3.1 排名

排名法借鉴了面对面访谈中使用的卡片排序技术。受访者根据单一标准（概念维度）对一组卡片（对象）进行排序。例如，根据偏好对品牌进行排序。类似的方法是分组（*grouping*），受访者对某些对象进行分组，例如杂货店产品、服装产品等。可以要求受访者只对某些对象进行选择或排名，例如前三个对象。排名任务也可以拆解成几个步骤，如成对比较（*paired comparisons*），即每次不比较所有对象，而是仅对比或排序两个对象；或者在最大差异评分（*maximum difference scoring*, MaxDiff）中，仅需确定极端的对象，如最重要的和最不重要的。

排名通常基于单一维度进行；根据两个维度进行排名也是可能的，但非常罕见（Bradburn et al., 2004, p. 176）。

排名可以用短数字输入列表来实现，受访者为每个对象输入对应的排名数值。另一种可能是通过拖放功能，将对象从一个列表中移至另一个列表，或者在统一列表中重新排序对象。Blasius（2012）发现，在应答时间和项目无应答率方面，拖放布局优于所有替代方案（数字输入、用箭头移动、最差选择）。然而，由于技术限制，特别是在使用移动设备访问网络问卷的情况下，拖放技术对某些用户造成了操作障碍。

尽管排名看似是简单的顺序测量问题，但对受访者的认知要求较高。排名的一个重要特点是，两个对象不能从现有的顺序类别中获得相同的排名值。因此，受访者需要同时处理多个对象。这对于5个对象已颇具难度，若对象数量超过10个，则会变得异常复杂。

在传统的调查模式中，排名法与等级评定法的争议已久，而随着网络问卷中拖放功能的应用，这一争议再度升温（Neubarth, 2010）。然而，针对这一困境的研究仍然相对不足。与排名相比，等级评定可以确保每个主题的应答相互独立、具有更高的效度、更丰富的统计分析且认知负担更低。此外，基于评定结果，可随时对对象进行排序以获得排名。然而，评级差异很小，或者不允许更多对象有相同的等级时，排名法则更具优势。例如，受访者可能对4种产品给出最高评分（例如在1~5分的量表都是5分），但若需了解其购买其中一两种产品的决策依据，则只能通过排名法进一步排序。

2.3.2.3.2 常量

常量和 (*constant sum*) 问题, 有时也被称为实时累计计数 (*running tally*), 表示一系列开放式数字输入问题 (例如, 每天在各类活动上花费的时间), 其答案总和被限定为某个固定数值 (例如, 每天的总时长为 24)。这类问题对受访者的认知要求较高, 因为受访者需要同时考虑和处理更多的维度或主题 (项目)。在这种情况下, 实施实时自动求和, 并辅助相应提示, 可以明显提高数据质量 (Callegaro, DiSogra, & Wells, 2011; Conrad, Tourangeau, Couper, & Zhang, 2010)。

2.3.2.3.3 图形的广泛使用

图形是网络调查中一项极具优势的特征。尽管本书主要关注的是基本网络调查模式, 其中多媒体元素 (图像、音频、视频、动画) 使用相对有限, 但仍有必要在此介绍扩展图形在调查中应用的可能性, 并简要讨论它们与数据质量的关系。第 2.3.4 节将讨论网络问卷的整个视觉布局和游戏化设计。

用于说明调查问题的图形。与其他模式相比, 在网络调查中使用图形说明问题具有吸引力强、实施便捷且成本较低的特点。然而, 学术界 (Couper, 2008) 以及市场调研中 (Poynter, 2010, p. 55) 研究人员普遍担忧对图形效果的控制水平较弱。首先, 技术层面可能存在诸多问题: 图形下载速度慢、颜色与字体显示不一致, 以及在部分设备和浏览器上无法正常加载等。例如, Lozar Manfreda 等 (2002) 报告, 当在网络问卷中添加品牌标志时, 中断率增加。其次, 随着互联网连接速度的提高和技术的标准化, 技术问题正在减少, 但图形对应答的不可预测影响并未消除。

当图形是问题的必要元素时, 如关于标志偏好的调查, 几乎不会引发问题。但是当图形仅仅被作为附加的问卷元素 (例如试图使问卷更有吸引力时), 则可能会出现不可预知的副作用。当图形与语言环境不一致时, 问题更为突出, 因为语言信息通常优先于视觉效果 (Toepoel & Couper, 2011)。典型案例是, 在询问受访者健康状况的问题旁, 配上健康人士或病患的图片, 会导致显著的应答差异 (在 7 点量表上 3.4 与 2.5 的差异), 且效果还受图片大小和位置 (如置于前一页、标题旁、问题一侧) 的影响 (Tourangeau, Conrad, & Couper, 2013, p. 88)。Toepoel 和 Couper (2011) 的研究表明添加非必要的图片并不能提高数据质量, 也无法提高受访者对问卷的满意度, 而这正是使用图片背后的常见想法。Deutskens, de Ruyter, Wetzels, & Oosterveld (2004) 和 Ganassali (2008) 的研究也没有发现图片的优势。

一般来说, 应该非常谨慎地使用额外图形, 并考虑清楚潜在的负面作用。若存

在疑问,建议采取保守策略,尤其是在没有研究证据支持额外图形元素优势的情况下。这种态度看似保守甚至过时,毕竟图形和多媒体是网络调查模式的重要优势,但在缺乏优势依据和指导方针的情况下,保持谨慎更为稳妥。

用于模拟社会(社交)存在的图形。Póltorak 和 Kowalski (2013)的研究发现,网络问卷中使用访谈者的照片或视频来模拟社会存在,可以提高受访者积极性和数据质量。关于这个问题的实验已经重复验证了访谈者管理调查的效果,特别是性别(Fuchs, 2009; Tourangeau, Couper, & Conrad, 2003)和种族(Krysan & Couper, 2006)的影响。因此,应谨慎使用这些图像。

图形衍生的新问题类型。除了作为调查问题的附加元素,图形还衍生新问题类型和为现有问题提供替代布局。前文已提及部分案例(拖放、图形替代单选按钮的图像、热点问题和连续量表)。以下进一步列举充分利用图形优势的网络问卷问题案例。

- 日历布局(*calendar layout*)作为日期格式的图形界面是询问日期的替代性布局。通过弹出式日历选择日期,可完全替代数字输入或下拉菜单。在其他各种情况下的使用(如在线预订)使得这种格式在网络调查中越来越有优势。更复杂的在线日历版本会通过图形展示关键事件里程碑,简化复杂的回忆任务(Glasner & van der Vaart, 2013)。
- 在虚拟现实环境(*virtual reality environment*)中选择对象是排名任务的延伸。例如,在虚拟超市中,受访者可通过拖放功能拾取、查看商品,并按一定顺序放入购物篮(Brace, 2008, p. 169)。
- 热力图(*heat-map*)问题依赖于图形界面。受访者根据某个维度(如最重要、有吸引力或无吸引力的部分)点击图片中的特定点。与名义测量中预先定义区域的图形不同(例如选择区域),此处每个点均对应比率测量水平的二维坐标,可实现精细分析。热力图分析的结果是以彩色图表呈现的密度焦点,类似于报告眼动追踪或鼠标移动轨迹。这种布局可用于可用性评估研究。

读者可以在各种市场调研网站(例如 GMI)上获得更多图形衍生新问题类型的示例。

2.3.2.3.4 特定内容的实施

本节聚焦与特定研究内容相关的问题实施方式,暂不涉及方法论层面的通用问题类型(如工作满意度等),也不涵盖受版权保护的专用测量工具(例如盖洛

普咨询公司用于评估员工参与度的 Q12 量表 ; Harter, Schmidt, Killham, & Asplund, 2008)。同时, 与某些研究方法紧密关联的特定问题组合(如实验设计中的操纵检查题)也不在此列。尽管这些特定内容的问题未产生新的问题类型, 但它们形成了极具针对性的实施范式。以下结合网络调查的特点, 重点阐述几类典型案例。

- **以自我为中心的社会网络问题**是用姓名生成器构建社交网络成员 (alters) 列表。由受访者 (自我) 针对网络中的每个成员 (如朋友) 回答一系列问题 (如年龄、性别、关系亲密度)。姓名生成器 (*name generators*) 通常表现为一系列开放式文本框, 供受访者填写网络成员的姓名。同样的布局可用于生成产品、服务、目的地和其他对象的列表, 特别是在市场研究中。目前, 关于网络问卷中此类问题的标准化实施布局仍在探索中 (Lozar Manfreda, Vehovar, & Hlebec, 2004; Vehovar, Lozar Manfreda, Koren, & Hlebec, 2008)。研究面临的主要挑战集中在图形界面和收集名字的文本框的格式。例如, 采用“顺序呈现文本框”的设计 (即前一个文本框填写完成后才显示下一个) 被证实更优 (Hogan, Carrasco, & Wellman, 2007)。为收集网络成员的详细数据, 可针对每个成员设置系列问题, 可采用项目式或是成员式 (Coromina & Coenders, 2006; Vehovar, Lozar Manfreda et al., 2008)。此外, 如何用图形界面替代传统问题形式测量 (如可视化网络关系图) 仍是待攻克的难题 (Hogan et al., 2007; Koren & Hlebec, 2006; Lackaff, 2012)。
- **联合分析问题**将排名任务融入一系列模拟决策场景 (如产品购买选择)。通过精心设计的环境呈现不同特征水平 (排名) 的选项组合, 以便应用联合分析方法解析受访者偏好。例如, 让受访者在“快速、廉价的小型车”与“慢速、昂贵的大型车”等选项中做出选择, 并可结合等级评定量表。
- **360° 或多评估者反馈**是人力资源管理 (HRM) 中使用的社会网络调查问卷的一种变体。每个员工对来自上级 (经理)、下级 (下属) 和同级 (同事) 的人员名单进行评价。网络调查为这类任务提供了便利, 但也面临诸多方法论挑战 (LeDuff Collins, 2009)。

2.3.2.3.5 包括观察数据和外部数据的问题

网络调查中受访者的数据也可以通过其他方法收集 (Groves et al., 2009, p. 150), 例如借助录音、插件或文件上传。通过这类方式, 可以收集录音、绘图、签名扫描 (付款收据)、条形码或二维码、照片 (如自拍)、视频、生物标记 (如测量体重)、

媒体渠道识别（如广播、电视）或 GPS 定位等内容。这些数据收集方式体现了网络调查的显著优势，尤其在使用移动设备时更为突出。不过，由于本书聚焦基础网络调查模式，在此不展开详述。

本文概述了网络问卷调查中的问题类型（第 2.3.2 节）。其分类遵循方法论概念框架，首先依据复杂性划分，再进一步按测量水平与布局形式细分。

然而，在实际创建网络问卷时，所使用的网络调查软件会预设特定的问题类型结构，这种结构很大程度上取决于技术层面的子设置与子选项选择。在一级分类中，网络调查软件通常将问题划分为单一项目问题与表格问题、开放式问题与封闭式问题、多选问题等类别，这些分类与 HTML 元素存在密切关联。软件中问题选择流程的一级导航逻辑也与各类问题的使用频率相关。因此，最常见的类型（即单选按钮、复选框、开放式文本和数字条目、带单选按钮布局的简单等级评定量表）往往更易于获取。详见第 2 章的补充内容“网络调查中的问题出现频率”（*Frequency of appearance of question*），网址为 <http://websm.org/ch2>。

综上，选择最合适的问题类型是一项复杂的决策，需综合考量问题内容、表述清晰度、技术要求及受访者的任务难度等因素。尤其重要的是，不应仅因某种问题类型可实现或看似新颖有趣而选用，其核心评判标准应是该类型能否为研究目标提供高质量的数据支持。

2.3.3 问卷结构、计算机化和布局

调查问卷不仅仅是一系列的问题，特别是在自我管理模式（如网络调查）中，它被视为受访者和研究者之间的对话媒介（Schwarz, 1996）。问卷的作用是确保对话的顺畅推进，而这一功能需通过合理的问卷结构、与受访者的有效互动及优化的视觉布局共同实现。

2.3.3.1 问卷结构

决定问卷结构的因素涉及多个方面，概括而言包括：问题的排列顺序、问题在页面上的分布方式、非问题页面的设置、导航设计、问卷的模块与问题组块，以及为特定研究目的采用的特殊布局形式。

2.3.3.1.1 问题顺序和语境效应

众所周知，问卷中问题的顺序是很重要的，它会影响受访者对问题的理解和

应答。当一个问题影响到其他问题的处理和回答时,就称之为语境效应(*context effects*)。网络调查中同样存在此类效应,许多研究报告都提到了语境效应的存在(Couper et al., 2004; Ester & Vinken, 2010; Malhotra, 2008; Nielsen & Kjær, 2011; Siminski, 2008),包括在移动设备上的问卷(Mavletova, 2013; Peytchev & Hill, 2010)。因此,调查方法的通用原则可直接应用于网络问卷设计(Krosnick & Presser, 2010, p. 264):按主题分组呈现问题,从邀请函中提及的核心问题入手,再逐步过渡到次要问题;每个主题内遵循“从一般到具体”的逻辑;按格式和逻辑关系组织问题(例如,事件的时间顺序);开篇问题应简单且具有吸引力,人口统计学信息和敏感问题则适合放在问卷末尾。

语境效应可以通过以下方式避免:不使用表格呈现问题,增加分页数量以在视觉上分隔问题。此外,问题顺序效应(*order effects*)可以通过随机化来削弱,本节将进一步描述。关于语境效应的更多信息,可参考相关经典教材,比如Krosnick和Presser(2010, p. 291)、Dillman等(2009, pp. 157-165)以及Tourangeau等(2000, p. 197)的教材。

2.3.3.1.2 分页断点、分页与滚动

问题在问卷页面上的分布也采用多种方式。关于每页的问题数量,有两种极端的方法,一种是单页设计(*one-page design*)(也叫滚动, *scrolling*),即所有问题都集中在一个页面上呈现;另一种是单题单页设计(*one-question-per-pagedesign*)(也叫分页 *paging*)。这两种模式的对比是早期网络调查实验的重点内容(Lozar Manfreda, Batagelj & Vehovar, 2002; Vehovar & Batagelj, 1996)。实验结果显示,在时长约7分钟的一般主题性调查中,两者的中断率没有差异,但滚动模式下未作答的问题数量更多。同时,滚动模式的应答速度更快,总耗时更短。在很大程度上,两者在应答速度和时长上的差异是由于当时的互联网连接速度太慢。然而,即使在今天,这些基本的发现似乎仍然有效,只是两者的差距已大幅缩小。

分页设计的优势之一是模拟访谈式问卷的呈现思路,受访者每次只看到一个问题。另一个优势是便于在服务器端实现交互功能(如逻辑跳转),这类功能只有在受访者移到下一页后才会执行。然而,完成每个问题后都需要转到下一页,会增加受访者的操作负担,并且通常会略微延长总应答时间(Couper et al., 2001; Lozar Manfreda, Batagelj & Vehovar, 2002; Thorndike, Calbring, Smyth, Magee, Gonder-Frederick, Ost & Ritterbrand, 2009; Toepoel et al., 2009; Tourangeau et al., 2004)。

滚动设计的局限性主要体现在交互功能支持不足。例如,需要受访者跳过一些

问题（例如，如果性别是男性，则跳过关于生育的问题）的跳转是通过包含超链接以及附加的文字说明来实现的（Peytchev, Couper, McCabe, & Crawford, 2006），这对受访者来说操作繁琐且易引发困惑，同时会增加应答时间。另一种解决方案是采用客户端脚本实现动态跳转（*dynamic branching*），也叫混合式（*hybrid*）（Dillman et al., 2009, p. 202），即点击某个反应选项，相关附加问题会在当前页面即时显示。然而，Peytchev 等人（2006）发现，滚动设计中的跳转会导致受访者回避那些会导致显示大量附加问题的应答选项。滚动设计的另一个缺陷是，多数网络调查软件仅在进入下一页时保存答案。因此，滚动设计要求受访者在单次会话中完成问卷，若中途中断则可能丢失已填信息；同时，所有问题一次性可见也可能增加语境效应的风险。

另一方面，滚动设计无需频繁分页，减少了点击操作和应答时间，且能让受访者直观了解问卷全貌，更接近纸质问卷的自我管理模式。

现有研究普遍认为，分页与滚动设计之间没有实质性的效果差异（Thorndike et al., 2009; Toepoel et al., 2009）。在简单网络问卷中，滚动设计的部分缺陷会弱化，但这并不意味着其具备明显优势。Toepoel 等人（2009）研究发现，每屏 10 个题目的四个页面的应答时间与单屏 40 个问题的滚动设计的应答时间相比，两者差异微乎其微；且滚动设计提升的满意度非常小，在 10 点量表中仅从 7.06 提升到 7.37。滚动的唯一真正优势可能出现在需要上下文参考的情况下（例如，全部问题都应呈现在同一屏幕上，以便受访者看到所有的问题），或者在必须与纸质版本保持一致时。

在实践中，多数设计采用介于两种极端之间的折中方案。改良滚动设计（*a modified scrolling design*）是指只有在需要执行服务器端的功能（如分支）、减少语境效应或阶段性保存回答时，才会出现分页。另一种方法是改良分页设计（*a modified paging*），每个页面包含有限数量的问题，确保在常规屏幕尺寸内可完整显示，无需额外滚动。鉴于研究发现完全分页和完全滚动设计之间的差异很小，本书可以预期改良后的分页和改良后的滚动设计之间的差异会更小。

因此可以得出结论，改良后的分页策略几乎不会出错，而改良后的滚动方式可能仍然面临上述局限，特别是应答数据保存不及时、漏答风险较高。除了总体结论，还需要结合具体调查背景权衡两种模式的优劣。

值得注意的是，在过去的十年中，博客、社交网络和鼠标设备的普及改变了滚动操作的用户体验。这些都引起了网络可用性原则的重大变化（Nielsen, 2000a）。在互联网发展早期，滚动遭到排斥。近期，移动设备（智能手机、平板电脑）再次强化了滚动设计的适用性（Mavletova & Couper, 2014）。这些变化可能进一步弱化滚动设计的固有缺陷。

2.3.3.1.3 非问题页面和模块

除包含调查问题的问卷页面外,网络调查还需设置不含问题的页面或内容板块,用于承载介绍、附加说明和其他与受访者继续作答问卷之前有关的信息。下面列出常见类型的非问题页面,而其他内容通常放置在同一页面的其他问题旁边,仅与内容相对较短的特定问题有关。

- 介绍页 (*An introduction page, splash page, welcome screen*) 是出现在第一页的独立页面,用于向受访者介绍调查。只在简短调查中才会省略这部分(如评估表)。介绍页用于说服受访者参加调查。许多受访者访问介绍页后,没有继续参与调查,这使得该页面成为问卷调查的常见中断位置。因此,其内容和设计尤为重要:兼具吸引力和尊重性、专业性和礼貌性。它不能违背研究伦理,必须传递给参与者基本信息:即调查发起方、目的、内容、隐私问题、联系方式和预计完成时间。强调对受访者参与的额外鼓励也很重要,比如调查的重要性、参与者获益或激励措施等。第 2.5 节无应答部分将进一步介绍激励受访者参与调查的相关内容。在介绍页的末尾处,还有必要提供开始参与调查的说明,包括简单的指令“要继续,请单击‘下一步’按钮”或者输入调查访问代码的附加指令。一般来说,这一页应该简洁明了。某些研究表明 (Bauman, Jobity, Airey, & Hakan, 2000),简短精炼的“封面信”比冗长的文字密集的“求职信式”引言,更能增加受访者合作。当使用邮寄或电子邮件邀请时,介绍页的内容需要与邀请相一致,避免重复冗余(详见第 2.5.7 节)。
- 过渡页 (*transition page*) 将受访者引入新的主题。一些研究表明,过渡页在不提高中断率的情况下表现出积极作用 (Callegaro et al., 2009),用于调节问卷作答节奏,缓解从一个主题到另一个主题的过渡。
- 指示页 (*instruction page*) 针对更复杂的调查任务而提供操作(例如,如何执行排名任务)。当所需说明简短时,就没有必要呈现于单独页面,可以直接附在相关问题旁侧。
- 激励或抽奖页面 (*incentive or raffle page*) 提供了促使受访者完成问卷的激励措施等附加信息,例如可展示奖励图片、说明抽奖中奖概率、在线优惠券和其他电子奖励等。
- 文件上传页面 (*file upload page or section*) 用于收集受访者提供的图片、文本文件或其他文件等。例如,可以要求受访者上传问题中涉及的有缺陷产品的照片。这可以作为独立页面,或者嵌入相关问题旁,后者更常见。但很难说文件

上传页的内容属于调查问题。

- 感谢页 (*thank you page*) 是问卷调查的收尾页, 用于向受访者表达感谢。它还可能包含外部网站的链接 (如调查主办方的网站或与问卷主题相关的网站) 和结果可用性的信息等其他内容。有时, 感谢页会要求受访者提供联系信息, 以便通过电子邮件接收结果, 参与进一步的调查或加入在线样本库。可以将受访者填答完成的问卷附在感谢页面 (如 PDF 格式), 以便受访者留存。
- 除了非问题页面和模块, 还可以有非问题句子。这些句子是问卷页面的一部分 (例如介绍、解释、鼓励、感谢信)。

2.3.3.1.4 调查问卷导航

受访者通常通过问卷中的专用导航按钮在问卷中切换页面。其中, “下一步”按钮 (*next button*) 是必需的, 而 “上一步”按钮 (*previors button*) 为可选功能。后者让受访者获得一定的问卷控制权, 并允许他们修改答案 (如同纸笔问卷)。不过, 有时研究者认为这种行为不可取, 会选择移除 “上一步”选项, 但这可能增加中断率 (R. P. Baker & Couper, 2007)。本节后续关于视觉布局的内容将详细说明。

有时, 还有一种自动跳转 (*automatic forwarding*) 设计可替代 “下一步”按钮的手动操作。受访者选择任何选项后, 系统会自动进入问卷的下一页。但这种方法存在导航混乱的风险, 且仅适用于封闭式的单一答案问题。Hammen (2010) 也发现自动跳转会增加受访者的满足倾向, 即敷衍作答。因此仅在极特殊情况下 (如采用水平滚动矩阵呈现的项目表格, 详见第 2.3.2 节) 才建议使用, 且需有充分的合理性说明。

2.3.3.1.5 问题组块

问卷结构通常划分为若干内容模块, 在技术层面以 “组块” (blocks) 形式呈现。对于较长的问卷, 可通过副标题、章节标题、说明文本及激励语句等进行视觉分隔, 为受访者提供指向性和激励性信息。高阶组块可转化为导航标签, 贯穿问卷全程, 支持受访者在不同主题模块间灵活切换。这种设计在商业调查中尤为常见, 根据受访者特征分配不同组块内容。需注意的是, 导航逻辑不宜过于复杂, 否则可能导致受访者困惑 (Blanke, 2011)。

2.3.3.1.6 特殊布局

网络问卷默认在独立浏览器窗口或标签中运行, 受访者可根据需求调整窗口大

小。不过，针对特定场景，可采用以下替代布局：嵌入式问卷和分屏展示。嵌入式问卷（*embedded questionnaire*）被整合至现有网站中，作为某个网页的一部分出现。例如在在线商店网站的某个页面上，嵌入消费者满意度调查。这种方法的主要局限是网站可以分配给调查问卷的空间有限；并且由于依赖客户端技术（如 JavaScript）实现动态交互式功能，比较难适配复杂问卷。

分屏展示（*split-screen*）通常用于网站评估。具体而言，网络问卷通常呈现在浏览器窗口下方，而待评估的网站可以同时在其上方的窗口中使用和检查。

2.3.3.2 计算机化、互动性和动态性

在问卷编制过程中的某个阶段，需要将问卷初稿中固定、静态的内容转换为网络调查软件中的在线问卷版本，这些初稿通常是在文本处理器（如 MS Word、Google Docs）甚至纸质媒介上准备的。或者，可以直接在网络调查软件的动态格式中编制问卷，前提是该软件支持所需功能（Vehovar, Čehovin, & Močnik, 2014）。在上述两种情况下，研究者都面临着动态和互动性的基本挑战：逻辑跳转（分支）、元素随机化以及实时反馈机制等问题。

2.3.3.2.1 分支

分支（*branching*）也被称为路由（*routing*）、过滤（*filtering*）、跳转（*skipping*）、条件（*conditions*）、IF 句子（*IF sentences*），指根据受访者的前期回答动态显示后续问题（例如，只有受访者是女性时才问生育次数）。与纸质问卷中的书面说明相比，网络调查更有优势，可以自动处理非常复杂的条件。一种是循环分支（*loop branching*），对受访者提及的每个项目重复相同问题集。例如，对于每一个被勾选的“被访问国家”，都会出现一组进一步问题。正如上文讨论的分页和滚动设计一样，在基于服务器的网络调查中，分支过程在服务器上运行，有条件的问题只能出现在不同的页面上。另一种可以在不与服务器交互的情况下实现的替代方案是动态分支（*dynamic branching*），允许在同一页面内动态显示问题。它依赖于客户端技术（JavaScript）。正如滚动设计部分内容中所提及的，这使下一个问题呈现在同一页面上（例如，关于生育次数的问题与性别问题显示在同一页面上，在点击女性性别选项后，页面立刻出现“生育次数”）。这种方式在问题较少时可能有效，但一旦受访者意识到某些回答会触发更多的附加问题，易出现无应答或敷衍作答。

网络调查软件中支持分支功能尤为重要，但是在定义逻辑分支的用户界面的可用性方面，各个网络调查软件之间存在差异。

2.3.3.2.2 随机化

随机化是根据某种随机机制,在不同的受访者之间调整问题表述、反应选项、布局或模块等。例如,可以使用访问网络问卷的受访者的奇数和偶数记录号码来形成两个随机组。

第一种类型的随机化是对顺序进行操作,包括呈现给受访者的问题、项目或反应选项等。这可以通过对(a)名义测量水平封闭式问题中的应答选项;(b)表格中的子问题;(c)页面或一组问题集的问题;以及(d)问题模块本身进行随机排序来实现。其中,反应选项顺序的随机化最为常见,因为它可以处理反应顺序效应。在网络调查中,由于可能出现首呈效应,即受访者首先会更认真考虑列表中的某几个选项(Galesic et al., 2008),更倾向于选择列表开头的答案。还有一些证据表明,受访者对第一个选项反应更积极(Tourangeau, Couper, & Conrad, 2013)。随机化并没有改变受访者的问题行为,而只是将这种影响随机分散到所有的回答选项或子问题中。为了真正防止首呈效应,需要考虑其他策略,如通过适当指令和工具提示来提升受访者的积极性(Kunz & Fuchs, 2013)。

另一种随机化的方法是实验,与上述对受访者收到的所有选项的顺序进行随机化不同,每个受访者都被随机分配到事先设定的问卷元素之一(回答选项、问题或问题组块、实施布局等)。典型案例是包括两个随机小组的样本实验(split-sample)。这可以探索哪种问题措辞或布局能提供更好的数据质量。第1.3.6节对实验进行了讨论,而Reips(2007)则提供了进一步的建议。

类似的随机化原则也可以用来减轻受访者的负担。例如,将某些问题随机分配给一半受访者,将其余问题分配给另一半受访者(另见第2.2.2节)。这种方法有时被称为矩阵抽样(matrix sampling),可以在不增加受访者负担的情况下增加问卷中的问题数量。当然,拆分仅用于无需进行合并分析的问题。矩阵抽样可以扩展到模块调查(modular surveys)(Johnson, Siluk, & Tarraf, 2014)和分割问卷设计(split-questionnaire designs)(Raghunathan & Grizzle, 1995),然后使用数据整合技术(统计匹配)进行插补,以补充缺失部分。

随机化也可用于其他超出问卷编制范畴的情况,例如随机选择家庭中的成员。

2.3.3.2.3 实时验证和提示信息

网络问卷调查可依据受访者的答案和行为,通过实时验证和提示信息的形式提供反馈。这可以适用于各种调查目的(Peytchev & Crawford, 2005),包括(a)对未回答的项目进行提示,从而呈现项目未响应的状态;(b)根据问题类型、格式、长

度限制和数值范围进行控制,以确保答案符合规范;以及(c)一致性验证,即结合同一调查的其他一些数据(例如报告的受教育程度和年龄的逻辑匹配)、对同一受访者的历史调查记录或来自抽样框等其他来源的外部数据,检验应答的一致性。其他类型的验证和提示也是存在的,包括基于数据收集过程中行为数据(*paradata*)的反馈,如关于应答过快的提示。操作系统或网络调查软件发出的某些系统信息也可能相关,例如在长时间不活动的情况下,通知失去互联网连接或通知浏览器会话已过期。

基于验证的互动通过错误信息等形式向受访者提供反馈。技术层面涉及两种实现方式:一种可在服务器端运行,当受访者尝试继续进入下一页时提供反馈;另一种是使用客户端技术实现,即在同一问卷页面上的实时反馈。验证可以产生三种可能的结果:(a)无提示(*no prompt*),受访者可以继续回答,但数据文件中记录潜在无效答案的指标;(b)软提示(*soft prompt*),告知受访者错误,但允许其不按照要求进行修正而继续作答;(c)硬提示(*hard prompt*),强制要求受访者修正错误后才能继续作答。

由于验证信息会干扰受访者的作答流程,易被视为侵入性内容引发反感,因此需严格遵循专业标准及设计与可用性原则进行实施(Couper, 2008; Nielsen, 2000a)。最重要的是保持提示信息的礼貌性和尊重性;内容需要清楚说明问题及解决方案,且所有提示机制均需经过严谨测试。

下文将以项目无应答提示为例展开说明,这是验证中最典型、最重要和最频繁的案例。其他类型的验证信息,主要原则仍然是一样的。

2.3.3.2.4 项目无应答提示

当受访者未回答某个问题(即出现项目无应答)时,研究者可以使用软提示、硬提示或什么都不做。在实践中,硬提示在受访者能获得参与奖励的场景中应用较广。然而,研究者(Couper, 2008, p. 266; Dillman et al., 2009, p. 309)强烈建议避免使用硬提示,除非该问题对后续调查至关重要,如无应答问题较为关键,调查问卷其余部分的分支取决于这个问题。避免硬性提示的第一个原因是伦理道德层面考量,因为调查的参与通常是自愿的,不应该强迫受访者回答任何问题(AAPOR, 2010)。其次,方法学研究表明,硬提示可能导致调查中断率上升或受访者提供虚假答案。遗憾的是,目前尚无实证证据证实这一点。尽管已有研究显示,硬提示(Albaum et al., 2010)和软提示(DeRouvray & Couper, 2002)均能减少项目无应答;但与软提示相比,硬提示对中断的影响并不明显(Albaum et al., 2010; Couper, Baker, &

Mechling, 2011; Heerwegh, 2005)。此外,关于硬提示对答案准确性和数据质量的潜在负面影响,相关研究仍较为有限,但其风险不容忽视。

另一方面,硬提示的优势也十分明显:它能直接消除项目无应答,便于即时数据分析,节省后期处理缺失数据的资源。硬提示也可以在一开始就约束受访者,使他们更认真地对待调查任务。因此,在网络调查,尤其是市场和商业研究中,除了开放式文本输入问题,其他问题几乎都会使用硬提示。需要特别注意,若问题选项中不存在适合受访者的答案,硬提示可能迫使受访者做出不合理选择,这种情况下应避免使用。

项目无应答提示与非实质性反应选项(“不知道”“没有意见”等)的处理方法密切相关,本书在问卷编制部分已经介绍过(第2.3.1节)。我们遵循Krosnick和Presser(2010)的建议,除非有明确必要,否则不建议包括这些选项。另一种策略是,只在受访者不回答某一问题后显示非实质性的反应选项,并配合项目无应答提示使用。如果把处理非实质性选项的三种基本方法(提供、不提供、提示后提供)与处理无应答问题的三种提示方法(无、软、硬)结合起来,研究者在理论上可得到九种备选策略。此外,我们还可以通过“熟悉度前置问题”(familiarity pre-question)和“确定性后置问题”(certainty post-question)来获得对这一问题的重要见解,这进一步增加了组合的数量。熟悉度前置问题可以帮助研究人员只对足够熟悉该主题的受访者提出问题,而确定性后置问题则询问受访者对所提供的答案的确定程度。

目前,关于这些组合的实证研究仍较有限。总体而言,市场研究中主要使用硬提示处理项目无应答的问题。另一方面,无论是否包含实质性的问题,学术研究和政府研究都更倾向于使用软提示或无提示的方式来处理无应答问题。在访员主导的调查(面对面和电话模式)中,通常不主动提供非实质性选项,但受访者明确提出时,访员会记录相应答案。在网络调查中,相应的项目无回答软提示之后显示非实质性选项可能是适宜的策略(Ainsaar et al., 2013)。

实践的重要性与研究结论的不确定性,使得项目无应答提示成为未来研究的重点方向。理想情况下,未来的研究将考虑各种基本设置中项目无应答的提示和处理非实质性选项的所有组合:对所有问题或特定问题使用提示、比较不同的受访者群体(普通人群、特定人群或受过训练的在线样本库成员)、不同程度的主题突出性,不同的问题类型,以及问题的重要性和位置。同时,需综合评估数据质量的多个维度,即信度、效度、项目无应答、中断、满足、应答时间以及参与度和满意程度。此外,还需考虑这些方法与访员主导调查的标准流程的可比性,而非仅限于优化网络调查的互动优势。将非实质性的应答作为合理的应答类别,这也很重要;而不是将其

视为“懒惰的回答”或“被掩盖的不回答”。熟悉度前置问题和确定性后置问题应该包括在这样的调查中，因为它们涉及问题的基本实质性方面。

最后，软件支持对项目无应答提示的实施至关重要。例如，仅在项目无应答发生后，在同一页面为实质性选项添加“不知道”等非实质性选项，可能对多数调查软件构成技术挑战。

2.3.3.2.5 数据填充

数据填充（或回填）（*data-piping or fills*）指调用前期问题答案、历史调查记录或外部数据库（如抽样框）中的信息。它通常被用来构建更清晰的指导语，说明对受访者的要求（例如“你和你提到的马克是什么关系？”）。此外，它还可以提供个性化的介绍以提高受访者的参与度，如在激励语句中嵌入受访者姓名。使用这一功能时需谨慎：直接呈现敏感的个性化信息可能削弱受访者的隐私安全感。

2.3.3.2.6 进度指示器

进度指示器（*progress indicator*）可以为受访者提供调查问卷完成进度的信息，通常用于确保受访者的参与度，以防止中断。它的位置和格式有很大不同，可以直接显示，也可以是按需 / 点击呈现；可以在每一页上显示，也可以只在某些关键过渡点显示；可以是图形（如典型的进度条），也可以是文本格式（如已完成 60%/10 个问题中的 5 个 / 页面 / 部分）。几乎没有研究考量这些变化的性能。

在复杂问卷中，进度指示器会出现问题，如大量问题被跳过，这导致进度指示器的推进不一致。例如，如果由于分支而跳过几个页面，进度指示器可能会从 10% “跳” 到 35%，随后又缓慢推进，这些分支的影响在计算上很难校准（Kaczmirek, 2009, p. 146）。

当不使用进度指示器时，受访者会抱怨它的缺失（Lozar Manfreda, Batagelj & Vehovar, 2002）。研究表明，使用进度指示器一般不会降低中断率，甚至会稍微增加中断率。在一项针对中位时长为 18 分钟的中长篇网络调查的随机实验元分析中，这已得到证实（Villar, Callegaro, & Yang, 2013）。

由此，当使用大量分支时，进度指示器对于中等或长的网络问卷没有益处。然而，它适用于简单短小的问卷，在这些问卷中，它以一致和可检测的速度线性移动。纳入进度指示器也带来伦理困境：从受访者的角度来看，进度的基本定位是有利的，但从研究者的角度来看，它没有带来任何收益，反而会引发问题。

2.3.3.2.7 其他互动特征

本书已提及部分互动及动态功能案例（如表格中的动态阴影），后续关于视觉布局的章节还将补充其他形式。值得注意的是，网络问卷还可通过更多设计优化填写流程，例如适时插入鼓励性语句，这类细节并未在书中详述。总体而言，重要的是要考虑这些特征是否以及如何对受访者和数据质量有利，并将其与潜在的缺点进行权衡。在缺乏相关研究的情况下，依靠常识和保守做法是最好的选择。

2.3.3.3 视觉布局

在面对面调查中，我们希望访员衣着得体、举止专业，并在必要时为受访者提供参与指导、激励和支持。同样，在电话调查中，我们训练访员保持礼貌，并以非引导性的方式、中性的语气进行提问。在网络问卷调查中，这类与受访者友好互动的准则需要通过问卷的视觉布局来实现。

这种布局通常遵循一些基本设计原则。例如，在西方文化中，顶部和左侧的位置通常被视为“更受关注”或“更积极”，而视觉距离可能反映感知上的距离（Tourangeau, Couper, & Conrad, 2013）。从主导性的语言表述到数字和视觉表述，存在一定的层级关系。因此，语言描述应是首选。当添加数字和图形时（如在等级评定量表中），必须保持一致性。同时，还应该遵守基本的网页设计与可用性原则，避免使用大写字母（Nielsen, 2000a; Schriver, 1997）。此外，应该在视觉上突出重要内容，将其他内容设置为背景，或者只在必要时才显示。

大多数网络调查软件已经提供一些预先制作好的专业视觉设计，称为主题、皮肤或调查模板（*themes, skins or survey templates*）。若需对整个调查问卷的视觉布局进行精细定制，建议让具备足够网页设计经验的人员参与。如果有必要对整个调查问卷的视觉布局进行精细定制，建议让有丰富网页设计经验的人参与进来。视觉布局由一种特殊的视觉设计语言定义，称为CSS（层叠样式表），它可以确保整个网页的外观和格式一致。修改CSS通常需要设计知识以及一定的编程能力。不要求研究人员完全精通视觉设计，但可以进行一些小调整，如添加标识。

研究人员或程序员可以通过直接修改代码或使用软件提供的图形用户界面（GUI）来更改CSS文件。不同软件的界面功能不同，部分界面允许以用户友好的方式修改各种设计元素（如背景颜色、字体、页面结构等）。

除了直接调整基本视觉布局，研究者还可以修改结构元素，如布局设置（标识位置、进度指示器显示等）、页面结构（分页断点、模块、组块等）、指令、定义、

帮助功能以及其他非问题类句子或页面。编辑问题文本同样重要,但应非常谨慎,特别是当文本编辑器允许调整字符、数字和符号的字体和大小时。在这种情况下,图片的尺寸、结构和设置也很重要。

下文将介绍一些关键要素,而不对视觉原理做进一步详述。关于视觉微妙作用的更多讨论可以参考 Tourangeau、Conrad 和 Couper (2013, p. 88) 以及 Couper (2008, p. 84) 和 Dillman 等 (2009) 的研究。需要注意的是,在所有主流浏览器、设备和操作系统中,检查和测试网络问卷的外观和体验至关重要,尤其是在对标准视觉布局进行修改后。

2.3.3.3.1 基本布局

问卷的基本视觉布局要素包括网页的字体类型和大小、文本宽度、空间、颜色、结构以及其他常见的视觉元素。它们应该与其他网页上使用的传统风格保持基本一致。一般建议是使用白色或浅色背景,清晰易读的标准字体(如 Arial、Verdana 字体等),专业的外观和感受,并在整个问卷中保持设计风格统一。Couper (2008) 对所有这些问题有更详尽的阐述。尽管这些元素很重要,但关于网络调查基本布局的研究却相对匮乏。仅有初步证据表明简单的设计优于“花哨”的设计(Dillman et al., 1998)。需要牢记的是,应答质量是首要目标,美学设计应服务于这一目标,而非增加干扰因素。在这一前提下,Casey 和 Poropat (2014) 明确表明,经典的美学风格在网络问卷设计中表现更优,其与问卷的易用性以及受访者对研究方的信任度呈正相关。

2.3.3.3.2 标识

标识能够提升调查的公信力,提醒受访者调查的发起者、调查组织、研究项目或在线样本库成员身份。标识应该与整个视觉布局保持一致,大小合适,并放置在页面的某个角落,以防止出现杂乱的印象,并避免分散受访者的注意力。

2.3.3.3.3 其他行动按钮的位置和导航

其他行动按钮的位置和导航,使受访者能够从一个页面移动到另一个页面,在设计布局中也很重要。Couper、Baker 和 Mechling (2011) 的实验结果发现,“上一步”按钮的位置对中断率没有影响,但其置于“下一页”按钮的右边时,使用量增加。对此,使用量增加是由于错误布局造成的,因为这类设置与其他网页和设备中的惯常做法相反,通常做法是“后退”在“前进”的左边。有些观点认为将“上一页”

按钮放在“下一页”按钮的下面,或者以超链接的形式放在它的右边,可能是有益的。然而,安全起见,普遍的做法是:(a)同时保留两个按钮;(b)两个按钮相邻放置;(c)放在问卷右侧或页面中间;(d)两个按钮间距适中,并留有足够的空间;(e)‘下一步’在右侧。

有时,其他行动按钮有特殊用处,例如允许受访者在多个会话中回答网络问卷的“保存并稍后继续”按钮,或者特别用于商业调查中的“打印”和“保存”按钮。建议将其可见性和位置与导航按钮保持一致,但也要突出其重要性。

2.3.3.3.4 补充说明的位置和格式

针对特定调查问题,有时需补充说明性信息以指导受访者正确作答。这类附加说明需通过差异化格式与核心问题文本区分(例如,使用小号或不同颜色的字体)。附加说明的数量和详细程度需谨慎权衡。例如,需明确复选框允许多选时,应在问题旁添加简短提示,或者在进一步的说明中解释。在特定情况下,每个问题中的简短说明可能更合适。我们通常建议在真正需要的情况下使用说明,过度使用会导致受访者忽略关键信息。当然,这完全取决于具体情境和目标受访者。但无论如何,需要在重要性和视觉暴露间取得平衡。

2.3.3.3.5 定义的位置和格式

调查问题中术语定义的呈现方式,同样面临与补充说明类似的建议和困境。Peytchev、Conrad、Couper 和 Tourangeau (2010)的研究发现,增加定义的视觉可及性也会增加其使用,即当定义紧跟在问题文本之后且可通过鼠标悬停即时显示时,受访者对定义的使用频率最高;反之,若需通过额外点击操作才能查看定义,其使用率则降至最低。

2.3.3.3.6 帮助和其他调查相关信息的视觉布局

一般建议呈现提供帮助的电子邮件地址、免费咨询电话或帮助网页链接。有时也需要附加其他信息,如版权说明、免责条款、通用说明和常见问题。一般而言,这些元素的视觉布局应该体现其辅助属性,例如使用小号字体、不使用高亮颜色,以及置于顶部或底部的非中心位置,以避免干扰受访者,也可以在超链接的页面上提供细节。

2.3.3.3.7 对各种屏幕分辨率的适配性

因为网络问卷通常是在各种电脑和其他设备上显示,必须保证有一个强大的视觉外观。所有问卷元素(包括文本、表格和图片)的大小和位置都应该用相对规格来定义,并且要足够灵活以适应屏幕的分辨率。导航按钮的适当外观、防止水平滚动、图片和提示的适当位置等也很重要(Callegaro, 2010)。这类视觉适配工作网络调查软件承担了大部分,但研究者应该仔细核实。

需要注意的是,在结束对问卷视觉布局前,对视觉元素进行详尽评估至关重要,包括核实受访者对视觉布局的体验和评价,以及在常用浏览器、设备和操作系统上适当展示。第2.3.5节将进一步讨论测试问题。

2.3.4 确保调查的参与度

为了获得高质量的调查回复,需要确保受访者不仅开始参与调查,还能维持必要的参与度。本节首先概述调查参与度,进而探讨与问卷准备密切相关的问题(游戏化设计、问卷长度和满意度)。关于无应答的具体内容将在第2.5节详细阐述。

2.3.4.1 参与度、流程和用户体验

参与度的概念与“心流”(flow)密切相关(Csikszentmihalyi, 2009)。“心流”指深度沉浸的心理状态,个体在该状态下完全专注于某项活动并受其驱动。心流出现在各种人类活动中,并且有了详细测量工具,例如沉浸状态量表(Jackson & Eklund, 2002)。参与度(engagement)有时被认为是一种稍微被动的状态,即没有完全的“沉浸”或对外部世界失去感知(O'Brien & Toms, 2008)。

另一种理解参与度的视角来自用户体验(user experience, UX)领域。这一概念涵盖用户在使用产品或服务时的行为与心理状态,在软件应用场景中,参与度涉及认知、情感和行为层面,其核心是“使用户愿意持续投入”(O'Brien & Toms, 2010, p. 3)。“参与性使用”的构成要素包括集中注意力、感知可用性、美学体验、持续性、新颖性和参与等。O'Brien和Toms(2008)在计算机应用的一般背景下成功地测量了这些成分。

遗憾的是,上述概念尚未被系统应用于调查研究领域。只有一项研究将参与度作为满意理论中优化行为的隐性组成部分(Krosnick, 1999)。在实践中,参与度常被提及于提升调查合作度的研究中,特别是市场营销领域;然而,尚未形成清晰的

理论阐释, 最多被视为可用性与动机的交叉点 (McMahon & Stamp, 2009)。

在缺乏针对调查场景的参与度理论框架的情况下, 一致性检查常被用作参与度的替代衡量指标, 特别是使用各种陷阱问题 (*trap questions*), 即偶尔向受访者提出某些简单验证问题 (如在预设选项中指定某一答案)。Baker-Prewitt 和 Miller (2009) 建议这些问题应避免高要求、枯燥、怪异或具有冒犯性。研究显示, 通常有超过 10% 的受访者未能正确回答此类问题 (Thomas & Barlas, 2014)。

另一种间接了解参与度的方式是问卷末尾的评价问题, 通过询问受访者的满意度、感知复杂度等获取反馈。值得一提的是, Downes-Le Guin、Baker、Mechling 和 Ruyle (2012) 通过一致性检查来衡量参与度, 结果发现视觉吸引力强、互动性高的问卷能提升受访者满意度, 但对参与度没有影响。

技术测量也可以作为参与度的指标, 具体包括观察数据 (如应答时间、点击或鼠标移动轨迹)、眼动追踪和脑电波记录, 它们可用于检测回答网络问卷时的情绪投入 (Coombe, Jarrett, & Johnson, 2011)。部分研究指出, 问卷中含有认知要求高的问题, 例如表格、连续量表和开放式文本问题, 会对参与度产生负面影响 (Peytchev, 2009)。

由于参与度的概念尚未被充分阐释, 因此前文讨论问卷特征时未提前引入这一概念。尽管如此, 参与度仍是问题类型、布局设计、互动性等诸多方面的总体概括。在调查中断 (即无应答) 的语境中参与度尤为重要, 也是第 2.5 节讨论无应答过程的重要变量。在此引入这一概念, 是因为下文中讨论的其余问题 (游戏化设计、问卷长度和满足) 都与参与度的概念密切相关。

调查参与度影响测量误差。我们通常从效度和信度指标、非实质性回答的数量、开放式问题中的无效回答、不一致的回答以及各类“满足策略”的表现 (如无区分作答) 来观察这种影响。

参与度还会影响无应答误差的发生概率, 通常通过项目无应答率和中断率来估算。本书将采用应答质量 (*quality of responses*) 这个术语 (Ganassali, 2008, p. 25) 来表示测量误差和无应答误差的指标。这些问题可能由调查参与度低直接或间接 (通过问卷开发中的次优解决方案) 造成。当然, 应答质量只是数据质量的一个组成部分。第 1.1.1 节已初步介绍数据质量的概念, 第 6.1.2 节将进一步展开论述。

上文列举了与应答质量相关的多项指标, 但现有参与度方面的研究 (尤其是围绕问卷长度、游戏化和满足策略的研究) 往往仅聚焦其中某一维度, 难以形成全面评估。因此, 下文仅针对部分问题展开讨论, 无法得出决定性结论。

2.3.4.2 调查游戏化

游戏化 (*gamification*) 的理念是将游戏设计元素融入非游戏场景。这在许多领域似乎都是有效的 (Hamari, Koivisto, & Sarsa, 2014), 而且也越来越多地应用在调查中 (C. Roberts, 2013)。游戏思维和游戏机制有望增强参与度和趣味性, 进而提升调查应答质量 (Puleston, 2011, 2012; Sleep & Puleston, 2011)。

网络问卷与在线游戏在视觉呈现和互动设计上存在诸多共通点, 两者均通过特定元素维持用户参与。这些元素的复杂程度可分为多个层级。基本元素包括装饰性设计和简单交互功能, 比如滑块和下拉菜单、替代单选按钮 (Downes-Le Guin et al., 2012)。高级层面涉及与邀请流程 (如提醒、激励) 有关的各种图形和互动元素, 以及使用图片的问题布局, 以及问卷布局元素 (如鼓励和进度指示器)。不过, 游戏化中极端的机制, 如动画角色、头像、强烈的叙事背景、积分制、徽章奖励、进度水平、时间限制和过量图形, 在调查中较少应用。此外, 游戏化还可能结合定性技术 (如个性化叙事、投射法、角色扮演) 优化问题设计, 以增强吸引力。

尽管存在相似性, 调查与游戏之间存在着本质的概念差异。调查旨在测量现实, 而游戏以娱乐和逃避现实为导向。因此, 学术研究者通常不考虑调查的游戏化设计, 而游戏化倡导者则将网络问卷视为在线游戏的一个子集——“只是相当沉闷的游戏” (Puleston, 2013), 这并不令人惊讶。

类似概念调查娱乐化 (*surveytainment*) 涉及增强视觉布局和可及性驱动, 目的是通过使用图形等视觉效果和互动性, 为受访者提供视觉上更具吸引力的调查, 从而提高其参与度 (Póltorak & Kowalski, 2013)。显然, 这与游戏化的概念非常相似, 但其适用范围更狭窄, 它仅涉及网络调查, 且它在高级游戏机制方面与游戏化不同, 更强调娱乐属性。人类对认知类活动 (如问卷填答) 的注意持续时间通常短于娱乐活动, 后者长达数小时。

遗憾的是, 尽管许多研究涉及调查游戏化, 但是实证研究较少, 且结论也不一致。部分研究发现, 在调查游戏化标签下, 使用各种视觉和互动元素, 如化身、3D 环境、反馈机制、叙事背景、时间压力和奖励等, 通常会出现好的效果, 在某些方面提升应答质量 (Cechanowicz, Gutwin, Brownell, & Goodfellow, 2013)。这一点尤其体现在开放式回答更丰富、受访者满意度提高、跨文化适用性更强 (Puleston & Rintoul, 2012) 以及回应率更高等方面 (Puleston, 2012)。但是, 另一些研究 (Chien & Chang, 2012; Ewing, 2012) 发现, 动画问题产生了不可预测的应答差异, 其他研究 (Downes-Le Guin et al., 2012) 发现动画没有影响。此外, 后一项研究还报告动

画增加了中断和编程成本，而在陷阱问题所测量的参与度方面没有发现任何改善。

多数支持问卷游戏化的研究仅关注应答质量的几个孤立维度，未全面评估其整体影响。尤其关键的是信度与效度问题。例如，如果“游戏化”未经验证而改变研究者实际测量的内容，效度就会受到威胁。游戏化的效果往往会逐渐减弱，有时甚至几个问题之后就会减弱，信度则下降（Cechanowicz et al., 2013）。

总之，图形、互动设计与情境化本就是网络问卷的标准组成部分，无需刻意归入“游戏化”范畴。通过承诺一定程度的“趣味性”吸引受访者参与，也是长期存在的合理策略，这是调查参与的社会交换理论（*social exchange theory*）（Dillman et al., 2009）。因此，本书建议聚焦“视觉与互动设计对参与度的提升作用”，而非直接采用“调查游戏化”这一模糊概念。因为游戏化的概念相对不明确和比较含糊。例如，在某些研究者看来，连续量表被视为游戏化的元素，但其他研究者却不认为。尽管如此，仍可从所谓“调查游戏化”实践中借鉴视觉与互动设计的创新应用，尤其在这些设计的作用尚未被充分探索的调查场景中。

鉴于研究结论的不一致性，我们认为对游戏化应采取保守策略。除非研究明确地证明某类元素的优势，否则避免不必要或过度使用视觉效果、互动性和情境化设计。例外情况包括：针对儿童受访者的头像、用于场景对比的动画、知识类问题的限时答题（如5秒倒计时），或定性市场研究中的角色扮演等特定场景。

2.3.4.3 问卷长度

一般而言，较长的问卷会对调查参与的各个环节产生负面影响，容易导致受访者疲劳（Krosnick & Presser, 2010, p. 292），降低应答质量。值得注意的是，受访者感知的问卷长度与实际长度可能存在显著差异（Galesic & Bosnjak, 2009）。面对过长的问卷，受访者可能选择中途退出（中断），或采取敷衍策略（满足）。问卷的复杂性和认知负荷也至关重要，因为一份内容简单但篇幅较长的问卷，其应答质量往往优于内容复杂却篇幅较短的问卷（Burdein, 2014）。

影响调查参与度的因素同样会改变受访者对问卷长度的感知，因此关于“合理问卷长度”的建议存在较大差异。在市场研究中，网络问卷（尤其是通过移动设备访问的问卷）应尽可能简短，通常控制在几分钟内完成，尤其是使用移动设备时。另一方面，多项研究（Cape, 2010; Couper, 2008, p. 298; Macer & Wilson, 2014; McMahon & Stamp, 2009）共同指出，应答质量在问卷时长达到20分钟左右时开始下降，这与人类注意力持续时间的普遍规律一致。不过，也有案例表明长问卷在网络环境中可表现良好，例如欧洲社会调查（ESS）的网络问卷耗时近一小时，仍能

维持较高的数据质量 (Ainsaar et al., 2013; Villar, 2013)。

从具体指标来看, 问卷长度与中断率存在明显关联: 问卷长度增加一倍, 中断率通常上升约 50%。例如, Ganassali (2008) 发现中断率可从 21% 上升到 34%; Deutskens 等 (2004) 发现中断率从 17% 上升至 24%。Galesic (2006) 比较了 10 分钟、20 分钟和 30 分钟的问卷, 发现中断率分别为 32%、43% 和 53%。同样, Vehovar 和 Čehovin (2014) 发现平均不超过 10 个问题的短问卷中断率为 6%; 而如果有几百个问题, 平均中断率会增加到 18%。根据此项研究, 平均每增加一个项目, 中断率上升 0.06%¹², 这与 Survey-Monkey 的研究结果相似¹³。但这两项研究都是观察性的, 应该谨慎使用。

关于“提前告知问卷长度”的效果, 研究结论存在差异。Galesic 和 Bosnjak (2009) 发现, 宣布问卷长度对参与调查的受访者比例有一定影响: 当宣布的问卷长度耗时 10 分钟、20 分钟、30 分钟时, 应答率分别为 75%、65% 和 62%。Crawford、Couper 和 Lamias (2001) 也发现了类似的效果, 宣布耗时 10 分钟与 20 分钟的问卷产生了 63% 与 68% 的无应答率, 但随后被较短版本的较高中断率所抵消了。Kaplowitz、Lupi、Couper 和 Thorp (2012) 对比宣布耗时 10 分钟与 30 分钟的问卷的研究, 以及 Walston、Lisstiz 和 Rudner (2006) 对比宣布耗时 5 分钟与 15 分钟的问卷的研究, 均发现长度告知对参与率的影响较小。

与长问卷相比, 短于 5 分钟的问卷其优势证据有限; 但若能配合激励措施, 短问卷的优势可能更明显。除增加中断风险外, 冗长的问卷还存在其他弊端, 主要是应答质量方面。例如, 问卷中处于后半部分的问题应答质量往往较低 (Galesic & Bosnjak, 2009)。

此外, 语言和文化差异会影响问卷的实际完成时长。在国际市场营销研究中, 同一份问卷的完成时间相差 50%, 从日本的最短耗时到拉丁美洲的最长耗时不等¹⁴。在这种情况下, 准确计算时长也很重要。优质网络调查软件可以提供精确估算, 但也存在各种近似的经验法则, 如每个单词耗时 2.5 秒 (每分钟 24 个单词)。

综上, 网络问卷的可接受或最佳长度受到诸多因素影响, 特别是问卷的发起方、主题的突出性、激励措施和整体质量等。一个粗略的基准是: 最长不超过 20 分钟; 超过这一时长, 应答质量往往会快速下降。因此, 若需设计更长的问卷, 则需通过上述因素的优化组合, 单个会话持续一小时左右; 超过这个时长, 需要更多的会话。额外增加的问题对累积中断的影响越来越小。同样, “提前告知问卷长度”的效果也会随问卷延长而降低。

此外, 就网络调查中的问卷长度本书提出三条具体建议。首先, 应考虑随机

化（即分割问卷或矩阵抽样；见第2.3.2节）以缩短问卷长度。其次，应该使用实时在线诊断，尤其监测个别受访者的表现，以便采取适当的干预措施。例如，在中断发生之前若发现应答质量急剧下降（Galesic, 2006），可以进行提示、帮助、鼓励、选择分支甚至修改问卷；尤其适用于“愿完成长问卷”与“仅接受短问卷”两类受访者并存的场景（Garland et al., 2013）。第三，无论问卷长度如何，研究人员应该始终坚持伦理道德，向受访者实时反馈问卷填写进度，并宣布（现实的）调查问卷长度。

2.3.4.4 调查满足

调查满足是指受访者在填答过程中仅满足最低要求，而非尽力提供有价值的信息。第2.3.4节中已经对调查满足进行定义，这类行为常见于参与度较低的受访者，他们出于某种原因（如合同约束、激励吸引、避免尴尬、习惯或好奇心等）仍然参与调查。

与参与度类似，调查满足与应答质量密切相关，但关注范围更窄。例如，项目无应答可能是由受访者的“次优行为”（*suboptimal respondent behaviour*）引起的（Thomas, 2014）；这与低参与度相关，但通常不被视为满足的直接结果，因为受访者并未提供任何答案。调查满足通常表现为：应答顺序效应（如首呈效应）、过度选择非实质性选项（如“不知道”）、填答速度异常加快、应答不一致和特定回答模式（如随机选择、默认现状、无区分作答和极端值聚集），以及默许（倾向于同意）。关于网络调查中满足程度的研究通常只关注其中几个指标，创建各种满足程度指数来识别满足行为（Barge & Gehlbach, 2012; Garland et al., 2013; C. Zhang, 2013）。除参与程度（动机）和调查问卷的属性（任务难度）之外，受访者的认知能力也是影响满足行为的关键因素（Krosnick, 1991）。间接来看，调查经验（Toepoel, Das, & van Soest, 2008）、受教育水平（Malhotra, 2008）和文化差异也会产生影响。Fang、Wen 和 Prybutok（2013）进一步发现，相比集体主义文化（如中国），个体主义文化下（如美国）的满足行为发生更频繁。

与访员主导的调查模式相比，网络调查的满足行为更为普遍。这是因为缺乏访员的实时激励与澄清，受访者的隐私感知、匿名性及多任务处理机会（例如边填写边在线聊天、检查电子邮件、同时浏览等）降低其投入度。面对面调查中，访谈者提供的激励和澄清角色，很难用网络问卷的布局和互动功能完全复制。网络问卷对认知能力的要求更高，默认受访者具备一定阅读能力和电脑操作技能，这进一步增加了满足风险。因此，研究一致指出，与面对面调查相比，网络调查中出现的满足行为更为频繁（Heerwegh & Loosveldt, 2008; Lindhjem & Navrud, 2011），有时也多

于电话调查 (Fricker et al., 2005; Klein, Havens, & Thomas, 2009)。对于自我管理的纸笔问卷, 结果则存在分歧。部分研究发现网络模式的满足行为更多 (Fang et al., 2013), 另一些则得出相反结论 (Linchiat Chang & Krosnick, 2009)。

目前, 针对满足行为的全面研究仍较缺乏。多数研究仅关注部分指标, 未能覆盖应答质量的全部维度。此外, 网络问卷的视觉和互动功能在减少满足行为方面的潜力尚未被充分挖掘。下文将围绕核心网络调查流程的三个阶段, 讨论减少满足行为的一些方法。

- **预调查阶段。**迄今为止, 这是减少满足发生的最重要阶段。为了确保受访者的适度参与, 需要努力提升预调查活动的整体质量。简要建议包括: 保持问卷简洁、问题简单, 必要时要求受访者验证答案 (Krosnick & Presser, 2010, p. 281)。Krosnick (2000) 介绍了一套经典的措施, 最大限度地提高受访者的积极性、降低任务难度和减少由满足引起的反应效应。通过强调研究重要性、引导受访者承诺认真填答、使用随机探针监测质量、保持问卷简单, 并将重要问题置于问卷前部, 可以最大化填写动机。为了简化调查参与任务, 以下方法会有所帮助: 降低问题措辞复杂度、避免询问难以回忆的事件、将复杂问题拆解为简单问题、标记反应选项, 以及本章介绍的其他一般问题和问卷呈现方法。最后, 通过随机化反应选项或问题顺序、避免使用易引发默许偏差的同意式问题、减少表格使用、慎用非实质性应答选项, 可以最大限度地降低反应效应 (*response effects*)。此外, 上述关于参与度、游戏化和问卷长度的讨论也非常重要。
- **正式调查阶段。**可以在测量阶段 (问卷填写阶段) 减少满足行为。例如, 每当应答速度或一致性测量显示应答质量较低时, 可以通过提示受访者进行干预。正如 Conrad Schober 和 Coiner (2007) 和 Tourangeau、Couper 和 Steiger (2003) 所发现的, 类似人类对话的互动设计 (例如通过语音或视频) 也被证实有效。需注意的是, 不应过度依赖应答时间判断满足行为。尽管填答过快的受访者更易表现出顺序效应 (Malhotra, 2008; Smyth, Dillman, Christian, & Stern, 2006), 选择“不知道”选项 (Lindhjem & Navrud, 2011) 或无区分作答 (C. Zhang, 2013), 但满足行为的速度差异较大。Garland 等 (2013) 的研究发现, 应答速度与以不一致性衡量的满足程度无显著关联。然而, 一项对 18 个在线样本库的研究显示, 填答最快的前 5% 受访者应答质量较低, 部分源于更高的满足水平 (Thomas & Barlas, 2014)。
- **调查后阶段。**对满足行为的检验常推迟至调查后阶段。在此阶段, 满足行为的

证据可用于改善数据（第4.1节）。例如，删除满足程度过高的项目或个案，或者用插补法（*impute*）替换受满足行为影响的答案。

最后，需强调网络调查软件的作用。网络软件对于开发高质量的视觉和互动功能、提升参与度和减少满足行为是必不可少的。它还能支持监测质量，并提供最终的干预措施（第5.3节将进一步讨论软件功能）。

2.3.5 网络调查问卷测试

在问卷准备过程中，问卷编制与测试迭代进行。问卷测试（*questionnaire testing*）是一套通过预先规划应用于问卷准备阶段的正式评估方法。除本节讨论的正式测试外，还存在非正式反馈（*informal feedback*），涉及各阶段临时问卷版本的内部设计、起草和评估。这通常是由研究人员自行完成，在准备问卷的过程中对问卷草案进行测试和评估。也可以请其他人提供非正式的、临时帮助和建议（如研究小组、同事或社会网络成员）。这类反馈属于问卷编制过程的组成部分，并非本节聚焦的测试方法重点。

正式的问卷测试是问卷准备环节不可或缺的阶段。Bradburn等（2004, p. 314）在其经典著作中指出，没有人能够一次性地完成问卷编制而无需后续测试和修改。他们甚至强调——我们完全同意，若缺乏开展正式测试的资源，根本不应启动调查项目。

测试主要是试图识别潜在的测量误差与无应答误差来源。传统调查模式的测试程序已有详细阐述，相关综述可见Madans、Miller、Maitland和Willis（2011）以及Presser等（2004）的梳理。Campanelli（2008）则提供了实用指南，涵盖评分量表模板、问卷系统评审标准、编码方案等测试相关工具。Saris（2012）基于附加数据收集量及主观与实证评估标准，提出包含13种问卷评价方法的概念框架，其中多数方法可纳入下文所述的传统测试方法范畴。

网络调查与其他调查模式存在诸多共性。例如，对每个网络调查都需要检验其问题措辞、问题顺序和问卷长度等。然而，网络问卷的测试还包括网络模式所特有的维度，包括视觉布局（如屏幕布局、导航、问题呈现形式）、互动和多媒体功能以及技术问题（如分支逻辑的准确性和不同设备下的外观）。因此，网络问卷的总体测试应该包括：

- 与内容和效度相关的实质性问题。

- 选项的方法论与认知层面（理解难度、措辞准确性、格式合理性、设计适配性等，包括受访者的能力和社会赞许性）以及问卷的方法论与认知层面（布局、结构、互动性、长度等）。
- 技术问题（视觉呈现效果，以及跨设备、浏览器、操作系统、互联网速度的编程、分支、随机化和跳转功能等）；
- 问卷整体的可用性。

在实践中，所需的测试量取决于资源（预算、人员和时间）、网络问卷的具体情况（包括是否存在复杂问题、是否测量复杂概念）、问卷长度、先前在其他调查中的问题数量、对网络调查软件的熟悉程度、目标人群（如老人、儿童、企业）、调查的重要性以及其他因素。

本节只系统地讨论最重要的测试方法，而问卷编制、测试与实际执行的关联将在第 2.3.6 节进一步探讨。

2.3.5.1 传统测试方法

传统调查模式已发展出一系列适用于网络问卷的测试方法，这些方法通常用于问卷编制的早期阶段，可直接迁移至网络调查场景。下文重点阐述与网络问卷高度相关的核心方法。

传统预测试（*convetional pre-testing*）历来在面对面调查中占主导地位（Presser et al., 2004, p. 100），通常是以部分“排练”（*dress rehearsals*）的形式开展：通过方便抽样选取部分受访者，由经验丰富的访员进行面对面访谈，并在结束后进行正式反馈（Campanelli, 2008）。尽管普遍认为 12~25 次访谈足以揭示问卷的主要缺陷，但这一经验法缺乏实证依据。在网络问卷编制的早期阶段，传统预测试仍能有效识别实质性与认知层面的问题。

专家评估（*expert evaluation*）包括主题专家（*topic experts*）对实质性问题的评估，以及调查方法学家（*survey methodologists*）对问卷的认知逻辑与方法论适用性的评估（DeMaio & Landreth, 2004）。为此开发了各类专家评估表格、系统化评审工具及检查表（Campanelli, 2008; Groves et al., 2009, p. 261），但也可以提供完全非结构化的反馈。通常，参与评估的方法学家人数（至少一名）少于实质性主题专家人数（至少几名），因为方法学问题更加标准化。针对网络问卷，技术和人机交互专家（*technical and human-computer interaction experts*）也可以参与评估，这一点将在后续详述。网络调查的优势在于可通过在线形式整合多方反馈，提升专家评估程

序的效率。

认知访谈 (*cognitive interviews*) 是针对受访者回答调查问题时的认知过程, 典型的技术包括“出声思考”程序、追问、回顾性思考、释义、信心判断、情境模拟等 (K. Miller, Chepp, Willson, & Padilla, 2014)。认知访谈通常需要一个小时左右, 需由受过专业训练的访员主导 (Campanelli, 2008), 在实验室环境下进行。访谈技术可以通过各种不同的方式实施, 但会产生结果偏差 (DeMaio & Landreth, 2004)。样本量通常在 5 到 15 个受访者, 但有研究表明这一规模可能无法识别半数以上的问题 (J. Blair & Conrad, 2011)。认知访谈虽以面对面形式为主, 但也可通过远程方式开展, 如借助聊天工具、音视频会议甚至虚拟环境 (Dean, Head, & Swicegood, 2013)。在网络调查中, 认知访谈可完全通过自我管理的在线形式实施: 受访者回答问题后, 以文字或录音形式描述思考过程并提供反馈。这类定性数据可与应答时间、屏幕捕捉等参数结合分析 (Chaney, Barry, Chaney, Stellefson, & Webb, 2013)。然而, 虽然自我管理或网络介入的认知访谈帮助节省了大量资源, 但提供的信息量可能不及现场面对面实时互动, 后者可以确保更高的参与动机和反馈深度。到目前为止, 在线认知访谈方法的研究与应用仍较为有限 (Behr, Bandilla, Kaczmirek, & Braun, 2013)。

受访者汇报 (*respondents' debriefings*) 与简单的现场测试访谈有关。不同于认知访谈, 其实施场景总是在实验室外进行。该方法通常用于问卷编制的后期阶段, 以获得受访者对问题理解的额外见解。开放式或封闭式的后续评估问题是受访者汇报的重要工具。这些问题通常置于问卷末尾, 但也可紧跟单个特定问题之后。具体操作是采用情境问题 (*vignette question*), 即向受访者展示假设情境 (E. Martin, 2004, p. 154), 这已经在认知访谈部分提到过。例如, 当澄清“工作”的含义时, 询问受访者是否将在医院里进行义务工作的人视为工作人员。对于网络调查, 该技术可以很容易迁移到自我管理的在线环境 (Aviram, 2012)。

焦点小组 (*focus groups*) 通常是由 5~12 名潜在受访者参与半结构化的讨论。这些讨论需要专业人员主持, 通常持续一个小时左右。在问卷准备初期, 焦点小组对澄清核心概念尤为有效, 在后期也可用于评估问卷草稿。除受访者外, 专家或访员也可参与焦点小组。典型的焦点小组是在实验室中采用标准方法进行 (Stewart & Shamdasani, 2014)。然而, 随着视频会议系统和现代软件支持的发展, 在线焦点小组应用日益普遍 (Gaiser, 2008)。

行为编码 (*behaviour coding*) 是一种可靠且可重复的技术, 即观察 (或录像) 受访者 (和访员) 并对其行为进行编码, 以识别问题理解与回答中的障碍。基于的假设是, 受访者的言语反应和肢体语言可以反映误解和问题 (Ongena & Dijkstra,

2006)。行为编码通常在实验室进行,若使用网络调查,也可通过远程视频会议系统实施。

实验 (experiments) 通过将受访者随机分组,观察问卷设计差异对估计值、数据分布、应答质量指标以及可用性评估的影响,常用于测试问题措辞、视觉布局、邀请方式及激励措施的有效性。实验设计需应对诸多方法论与统计挑战,如分组数量、交互效应及因果关系确立等,因此需专业谨慎处理 (Tourangeau, 2004)。网络调查为实验提供了便利条件——技术实施相对简单,所需资源较少 (Reips & Krantz, 2010)。

上述方法 (从传统预测试到实验) 多应用于问卷编制早期,依赖受访者或专家的反馈。而下文描述的其余三种方法不同: 预研究使用问卷的最终版本进行测试,而在自我评估和统计建模中,无需受访者或专家提供额外数据。

预研究 (pilot study) 以目标人群的小样本为对象,在真实环境中评估最终问卷,相当于“彩排” (*dress rehearsal*)。通常在问卷末尾设置标准评估问题 (如“回答问卷的难易程度”) 以收集受访者反馈。对所获得的数据进行分析,特别是对回答的分布和过程数据进行分析,是预研究的重要组成部分。预研究不可或缺,因为它可以用来测试正式调查的其他环节 (如抽样、招募、应答率、长度) 和调查后阶段的其他环节 (如数据文件的记录、应答质量)。预研究的样本通常不超过目标人群的百分之几。最小样本量遵循经验法则,通常在 15~30 个左右,可以粗略地检验分布情况。另一方面,如果预研究的目的是检验不同子群体中问题的测量特征,则可能需要数百个样本。需要补充的是,当测试重点为研究设计的特征 (如应答率或激励措施),而非问卷本身时,研究者建议 (Bradburn et al., 2004, p. 319) 将其与预研究分开进行,并称之为现场试验 (*field trial*)。

统计模型 (statistical models) 基于预研究或既往研究的完整数据集,旨在理解各类误差,特别是信度和效度,为问卷优化提供依据。常用的统计建模方法包括潜变量模型 (如因子分析、潜在类别分析)、标准程序 (如项目反应理论)、准实验设计和模型以及多质多法模型 (Saris, 2012)。

问卷评估工具 (questionnaire assessment tools) 可以帮助研究者根据某些规则或程序来评估问卷。这可以通过审查问卷是否符合简单的检查表和专家方案来完成 (Campanelli, 2008; Groves et al., 2009, p. 261), 也可以基于更复杂的概念方法进行评估,比如 Saris 和 Gallhofer (2014) 的三步问题开发程序。部分工具可通过专用软件实现自我评估的自动化或半自动化诊断,例如调查质量预测 (Survey Quality Prediction, SQP) 系统和 QUAID (Graesser, Cai, Louwerse, & Daniel, 2006)。前者估

计调查问题的效度和信度 (Sarlis & Gallhofer, 2007), 后者分析问题的语言特征 (如是否存在生僻或专业词汇)。

2.3.5.2 网络问卷调查的具体测试方法

除了上述传统方法, 网络调查还可以采用特定的评价方法。这些方法部分源于其他计算机辅助调查模式 (CASIC), 部分则为网络调查所特有。

问卷技术测试 (*questionnaire technical testing*) 是问卷计算机化的重要组成部分。它对于确保问卷正确运行, 以及避免出现因技术问题而导致严重的无应答和测量误差来说至关重要。实施时需预先制定测试策略、标准、资源分配及计划, 以实现系统性评估。与网络问卷技术测试有关的任务通常包括:

- 技术正确性检查和调试 (*Checking for technical correctness and debugging*)。技术问题可能源于问卷设计规范的疏漏, 也可能是由于规范和实际编程之间的差异造成的。其中, 分支逻辑的准确性是检查重点, 条件语句定义中的微小错误会导致调查问卷分支出现严重偏差。对于简单问卷, 逐屏检查即可识别和解决这些问题。然而, 包含大量分支、随机化、计算、辅助数据、组合表和其他高级功能的复杂调查问卷需要精心设计的正式测试策略。在很大程度上, 这与测试软件应用的标准程序相通, 仅通过观察功能表现的“黑箱测试” (*black-box testing*) 往往不足以识别编程错误, 需结合“白箱测试” (*white-box testing*) 检查内部编程结构 (Baker, Crawford, & Swinehart, 2004a, p. 369)。Tarnai 和 Moore (2004) 系统梳理了计算机化问卷的常见错误、方法和相应管理策略。部分网络调查软件还具备自动生成测试数据和条目的功能, 可辅助技术验证。
- 跨浏览器、操作系统、设备的稳定性 (*stability across browsers, operating systems, devices*)。尽管网络浏览器提供相对标准化的体验, 但由于各种因素, 如屏幕尺寸、客户端脚本 (尤其是 JavaScript) 以及输入方式 (鼠标、触摸板或触摸屏) 的支持差异, 问卷的显示与交互效果可能存在差异。为了识别这些问题, 我们应该首先明确目标人群常用的技术环境, 例如通过分析类似或以往调查中的设备类型范例 (见第 2.4.2 节), 并检查不同设备、操作系统和浏览器的相应市场份额。测试重点包括检查视觉显示、交互功能的运行和网络问卷的高级性能 (例如多媒体、复杂问题类型、外部链接、特别编程功能)。为了验证用于网络调查软件的具体性能, 我们可以创建一个带有所有问题类型、标准布局 and 交互功能的主测试问卷, 并定期在各种平台上测试其性能。高质量的网

络调查软件能预先解决大部分兼容性问题。

- 遵守标准 (*compliance with standards*)。在设计网络问卷时, 遵循特定的技术标准往往是有益的。这些标准涵盖调查准备和实施的其他方面。我们在此阐述标准的所有方面, 既包括狭义的技术方面 (如 W3C 标准), 也关注问卷的布局和呈现 (Dillman et al., 2009)、可用性 (Nielsen, 2000a) 以及其他各种具体方面, 如网络可及性, 特别是针对视力受损的受访者的适配策略 (Kaczmirek & Wolff, 2007)。此外, 通过整合视觉设计方案, 如字体大小、背景颜色和标识, 可以建立一个标准的企业视觉形象。研究机构有时会制定特定的问卷实施标准, 例如使用硬性和软性方式提示无应答项目。语言规范性 (如语法、拼写、校对) 和整体质量保证政策也需纳入检查范围, 参见 Baker 等 (2004, p. 369) 了解市场研究机构中此类标准的案例。一旦建立, 在测试过程中就会按照此类标准检查每份网络问卷是否符合。
- 其他技术方面 (*other technical aspects*)。除上述内容外, 还需要检验数据记录是否正确、所需的互联网连接速度、实施调查的技术基础设施的各个方面 (服务器容量、备份和安全政策等) 以及其他具体问题。

电子观察 (*electronic observations*) 通过技术手段记录受访者行为, 这些行为指向问卷的潜在问题。录音可以作为独立的数据收集方法用于测试, 但更多时候是作为上述其他测试方法的支持而实施。这些方法包括以下内容:

- 自动收集的并行数据分析 (*analysis of automatically collected paradata*) 依托于受访者回答问卷时留下的计算机痕迹, 如时间戳、鼠标移动轨迹和导航路径。时间戳分析在网络问卷测试中应用广泛, 并且在历史上与反应延迟研究有关。反应延迟研究始于 20 世纪 90 年代, 当时 CASIC 模式实现了自动、详细的时间记录。反应延迟研究源于认知心理学, 刺激和反应之间的长时间延迟可能表明存在问题。学界已发表过关于网络问卷时间分析的研究 (Callegaro et al., 2009)。第 2.4.2 节将进一步讨论并行数据的收集和使用。
- 屏幕记录和受访者的视频会话 (*screen recording and respondents' video sessions*) 正变得越来越容易获得。可以单独分析屏幕序列的跟踪和录制受访者的行为, 也可以整合到其他测试方法中, 尤其是用户体验和各种定性技术。它也可以作为在线认知访谈的一部分, 上文已介绍。
- 眼动追踪 (*eye tracking*) 记录受访者看屏幕的位置 (眼睛注视问卷的部分)、注视点和浏览顺序 (Galesic & Yan, 2011)。结果通常以热图形式呈现, 并且可

以进行进一步分析 (Coombe et al., 2011), 以便能够优化某些元素的视觉呈现 (例如突出说明文本或回答选项)。该方法也常用于网络可用性测试 (Nielsen & Pernice, 2009)。

- 脑活动 (*brain activity*) 的测量可以用来深入了解受访者在回答问卷时的参与度、流畅度、兴奋度和注意力等各方面。可通过多种方法测量, 如脑电图 (EEG) 和功能性磁共振成像。尽管由于该类技术成本很高, 且对大脑过程和行为之间的关系了解尚不充分, 目前应用较少, 但其在调查测试中的应用潜力已得到初步验证 (Coombe et al., 2011)。

在线评论 (*online commenting*) 可用来收集对问题、页面或整个问卷的反馈, 同时在线回答或检查网络问卷。尽管评论本身在调查测试中并不罕见, 并且经常作为传统测试方法被使用, 但网络调查因在线实施给在线评论法带来了巨大便利。可通过在网络问卷的测试版本中添加简单的评论文本字段 (位于问题下方或旁边) 或链接选项收集一般评论来实现。评论系统需要以用户友好的方式融合至网络调查软件的测试模式中。评论可以来自于研究团队、专家或受访者。既可作为非正式反馈的一部分, 也可整合到正式测试或其他在线测试方法 (如受访者调查、认知访谈、专家访谈、焦点小组) 中。有效的评论系统需与网络调查软件的测试模式无缝融合, 支持反馈的提交、研究团队内部讨论及待办事项管理。遗憾的是, 多数网络调查软件缺乏此类功能, 仅有少数高级工具提供用户友好的评论模块 (Vehovar, Slavec, & Berzelak, 2011)。鉴于其对测试与问卷优化的重要性, 评论功能应作为选择网络调查软件时的关键考量因素。

可用性评估 (*usability evaluation*) 指可用性的程度, 即软件、硬件或其他任何元素的可用性和任务匹配性 (Kaczmirek, 2006)。依据此定义, 本书将网络问卷的可用性定义为: 问卷的简单程度以及与受访者任务的匹配程度。通常, 问卷的可用性可以通过本节中提到的问卷设计原则来实现, 包括适当地选择问题类型、问卷结构和布局、导航和提升参与度。然而, 当涉及可用性评估时, 这在一定程度上与上文已经阐述过的某些测试方法存在重叠。从狭义上讲, 它与人机交互 (HCI) 的原则密切相关。这些原则可用于一种特殊类型的交互系统的网络问卷调查。网络可用性的一般原则 (Nielsen, 2012; Roe, 2008) 也可以被考虑应用于网络问卷调查。在此背景下, Kaczmirek (2006) 梳理了网络问卷调查的可用性是如何与人机交互和 ISO 9241-10 标准的可用性原则相关联, 该标准是关于视觉显示终端的办公室工作的人机工效学要求。

可用性评估可在实验或自然环境中实施, 场景包括实验室或实地(如受访者家中)。常用技术是出声思考、其他观察和认知技术。

此外, 可通过分析预研究或正式调查中收集到的数据(即应答时间、中断率、错误率、导航行为), 或者通过人机交互专家来完成可用性评价。后者是专家评估, 也是启发式评估(*heuristic evaluation*), 专家们通常遵循某些结构化的启发式标准(如导航的简单性、字体的可读性或视觉呈现的整体清晰度)。

最常用的可用性评估方法是正式的可用性测试(*usability testing*)。通常是要求受访者按照预定方案完成特定任务, 然后观察、记录和分析他们成功完成情况的程序。网络问卷可用性测试多在实验室环境下进行, 由受访者完成各种可用性测试。总体而言, 一般的可用性测试方法基本上都适用(Nielsen, 2000a), 包括各种经验法则。例如, 研究者通常认为仅5个测试者的小样本量足以发现80%的可用性问题(Nielsen, 2000b)。在非常简化的情况下, 由几个受访者填写问卷并对过程进行出声思考评论, 可以提供足够的信息来改进问卷。在更复杂情境中, 通常需要10名左右的受访者, 有时会在问卷编制的各个阶段分别使用。Baker等(2004)实施了三个步骤: 首先观察受访者回答整个问卷的情况, 然后询问标准评估问题, 最后参与逐屏讨论和具体任务。Hansen和Couper(2004)提供了更多复杂案例研究范例。

可用性评估经常与上文讨论的电子观察法相结合。带有内置摄像头和屏幕记录软件的笔记本电脑, 为这种方法提供了便利。另一可行的替代方案是受访者在自然环境中进行测试。例如, Stettler(2013)研究使用视频会议软件, 在受访者和远程主持人之间进行屏幕共享和实时沟通。自我管理的远程会话使用在线评论系统或采用远程非主持式可用性测试软件(*remote unmoderated usability testing*)(McCloskey, 2014), 指定任务并对受访者的表现进行录制。Albert、Tedesco和Tullis(2009)提供了对在线可用性测试的指导。

尽管可用性评估研究广泛, 但关于网络调查中该方法具体实施的文献却较少, 且以案例为主。除了上述引用研究, 只存在少数几个案例(Da Costa, Schmoelz, Davies, di Pietro, Kupek & de Assis, 2013; Kaczmirek, 2011; Potaka, 2008)。虽然与一般可用性方法相比, 网络问卷的可用性评估相对较少, 但仍需要更多实践实施的指导。一些常见困境包括: 是在问卷编制结束时只进行一次可用性测试, 还是分阶段实施; 选择新手还是有经验的受访者; 是在实验室还是在远程环境下进行; 是使用调解会议还是非调解会议等; 还涉及最佳样本量、互动的结构化水平(情景与自然流露)、评估重点(措辞、视觉设计、交互功能、导航等)以及标准化性能指标的

开发。最后，还需要详细研究用于问卷调查可用性测试的范式、屏幕记录的方法及附加价值。

集成自动诊断 (*integrated automated diagnostics*) 建立在统计模型的基础上。但与上面提到的问卷评估工具 (如 SQP 或 QUAID) 不同, 它被嵌入到网络调查软件中, 并支持在线问卷编制的全流程。这些诊断工具能够协助检测各类与问卷相关的问题。在最简单的形式中, 它们能针对易识别的潜在问题发出提示, 例如, 表格中子问题数量过多。针对更复杂的问题, 自动诊断需依托先进技术 (如数据挖掘) 来自动估算问题的信度和效度, 借助语言技术识别表述模糊的词汇, 甚至利用基于语义的人工智能算法提供问题措辞优化建议。部分工具还会生成问卷质量总分 (Nallan, 2012)。但对这类分数的解读需审慎评估其计算方法, 以明确其局限性。在高度先进的形态下, 自动化系统能够引导研究者完成整个问卷编制流程。

前文提及的所有测试方法, 传统的和网络调查特定方法, 都应相互补充; 其中一些方法关联密切, 甚至有时可以合并。选择正确的测试方法组合是一个复杂问题, 下一节将展开讨论。

同样需要注意的是, 网络调查的测试方法需要特定的软件支持 (*software support*)。这类支持可能整合于问卷编制工具中, 也可能作为独立模块存在。前文已在不同测试方法中提及部分功能特点, 以下进行简要总结与拓展。

- 问卷预览功能 (*the questionnaire preview*) 是所有网络调查软件的标配, 并且在问卷编制过程中频繁使用。高级预览设置有助于提升测试效率, 允许我们忽略跳转逻辑与条件限制, 在问题和页面之间跳转, 在分页和滚动设计之间切换, 等等。此选项对于在数据发布和传播阶段问卷文档化也非常有用。因为即使调查终止, 仍能向数据使用者提供预览 URL。
- 测试问卷 (*test questionnaire*) 与问卷预览不同, 因为它支持输入问卷。测试问卷通常与主问卷的 URL 地址相连接, 测试答案通常存储在调查数据文件中。然而, 测试答案会存储于调查数据文件中, 但所有测试条目会被明确标记, 便于后续从数据文件中批量删除。
- 实时编程控制 (*real-time programming controls*) 可辅助识别分支逻辑 (条件语句)、问答循环、随机化等动态元素中的逻辑或技术错误。
- 应答模拟 (*response simulations*) 通过对每个问题生成随机答案自动填答问卷, 用于验证问卷逻辑的完整性。例如, 若某条件性问题未被模拟应答触发, 可能暗示存在分支错误导致该问题无法显示。

- 并行数据 (Paradata) 可收集有关受访者在完成预测试问卷时的行为数据, 可用于对问卷相关问题进行各种诊断。
- 在线评论系统 (online commenting system) 支持受访者和专家在线提供反馈, 而研究人员可以有效地处理和管理评论。
- 自动诊断法 (Automated diagnostics) 可评估问卷多方面特征的实施合理性, 包括表格中子问题数量、长度估算及问题措辞的语言规范性等。
- 支持捕捉和处理屏幕截图或视频集成 (Integration of the support for capturing and processing screenshots or videos), 如同网络应用的标准可用性测试流程, 这类功能能有力支持正式测试的开展, 请参考 Blaise Tools¹⁵ 对这些选项的讨论。

虽然上述方法的主要测试重点是调查问卷, 但调查流程的其他环节同样需要测试。第 2.4.4 节将进一步介绍技术准备阶段的一般技术预测试; 前文已提及的预研究与实地实验, 不仅用于评估与问卷相关的程序, 还可检验正式调查及调查后阶段的流程。第 3.3 节将继续讨论“软启动”(soft launch), 在某些情况下, 它可以作为正式调查开始时的测试工具。

2.3.6 问卷编制、计算机化和测试的一体化

第 2.3 节涵盖了网络问卷准备的几个环节, 而这些环节需要整合。问卷编制是整个过程的核心, 从最初的概念界定和操作化到最终的措辞和问题的布局。随后, 通过定义问卷结构、布局设计及计算机化处理, 将零散问题整合为完整问卷。为确保每个阶段的实施质量, 需运用上一节(第 2.3.5 节)提到的各种方法进行测试。所有这些阶段的实施本质上具有迭代性, 每一次迭代都会产生问卷初稿(draft versions)和/或其他一些产出物, 然后再进行进一步的评估和必要修改。

网络问卷的计算机化过程需整合外部软件工具。问卷的初稿通常使用标准的文本处理器(text processors)(如 MS Word、Google Docs)准备, 然后通过电子邮件或某些文件共享网络服务在研究者和客户之间进行流转。问卷编程说明, 如逻辑分支、随机化和验证, 也包含在这些初稿中。有时, 部分大型机构会专门采用外部问卷调查文档和开发工具完成此类工作(Lorenc, Biemer, Jansson, Eltinge, & Holmberg, 2013)。上述在网络调查软件之外完成的问卷准备工作, 统称为“外部问卷编制”(external questionnaire development)。

然而, 问卷最终需迁移至网络调查软件中实现数字化呈现, 在其中真正实现网

络问卷调查。相应地,所有在网络调查软件支持下进行的问卷编制和测试活动都被称为整合问卷编制(*integrated questionnaire development*)。随着软件技术发展,未来有望实现从初始概念、流程规范到最后测试,软件对问卷准备过程中的所有阶段提供更多的综合支持。Vehovar、Petrovčič 和 Slavec (2014)描述了该方向的实施步骤。尽管如此,外部编制方法仍将在一定范围内沿用,尤其是复杂或大型调查项目。

从方法学角度观察,问卷准备流程的复杂性可通过图 2.1 予以呈现。图 2.1 中的粗体框和大写字母表示问卷编制和测试的关键活动,这些活动通过迭代序列整合,每一轮迭代通常生成一版问卷初稿(图中切角方框所示),经测试后进入下一轮编制。

图 2.1 右上方的阴影区域是与外部问卷编制相关的活动。在图 2.1 中,其结构相对复杂,而综合问卷编制的其余活动则更为系统。外部编制并非本书重点,且涉及与在特定的外部环境中编制和测试问卷有关的各类特征。然而,在大多数情况下,外部环境主要指某些文本处理器。

如前文所述,调查问卷的准备过程往往会产生多版初稿。需注意的是,在测试某一版初稿前,还需完成若干辅助准备工作,这些工作已在图 2.1 中以方框明确标注,但其重要性常被低估。例如,为满足测试需求,需添加专门的引导语、测试说明及反馈收集项;有时,对于特定的测试,还需将问卷转换为印刷格式。在最终版问卷准备阶段,需完成最终校对、视觉布局微调及其他不涉及方法论层面的细节修订。

网络问卷草稿有时也可以导出为打印版、PDF 或文本处理格式,以便进一步优化(如在网络调查软件中修改)或用于外部开发(如通过外部工具进行新一轮编制)。

正式的测试通常按照顺序进行,可以使用上一节(第 2.3.5 节)中所阐述的各种方法。采用某类测试方法后,需先修订发现的问题,再进入下一阶段测试,以消除任何已发现的问题。

另一种做法是在同一测试阶段并行使用多种方法,但受资源限制,这种情况并不常见。此外,不同方法可能识别出相同问题,造成冗余;不同方法也可能出现冲突结果(Lenzner, 2012; Yan, Kreuter, & Tourangeau, 2012)。这种差异更多反映问卷编制的复杂性与多维性,而不是不同测试方法间本身的矛盾。如何将不一致、冲突的结果纳入问卷修订中会成为复杂难题。更多情况下,不同方法具有互补性(Hughes, 2004),因此在特定阶段同时采用多种测试方法可能更具价值。

在某次迭代后是否结束问卷测试往往由外部限制因素决定,比如时间和资源。换言之,预先设定终止规则十分必要,毕竟问卷优化理论上可无限进行。为了停止无休止地追求完美问卷,判断依据可是最后一次测试中发现的错误数量和严重程

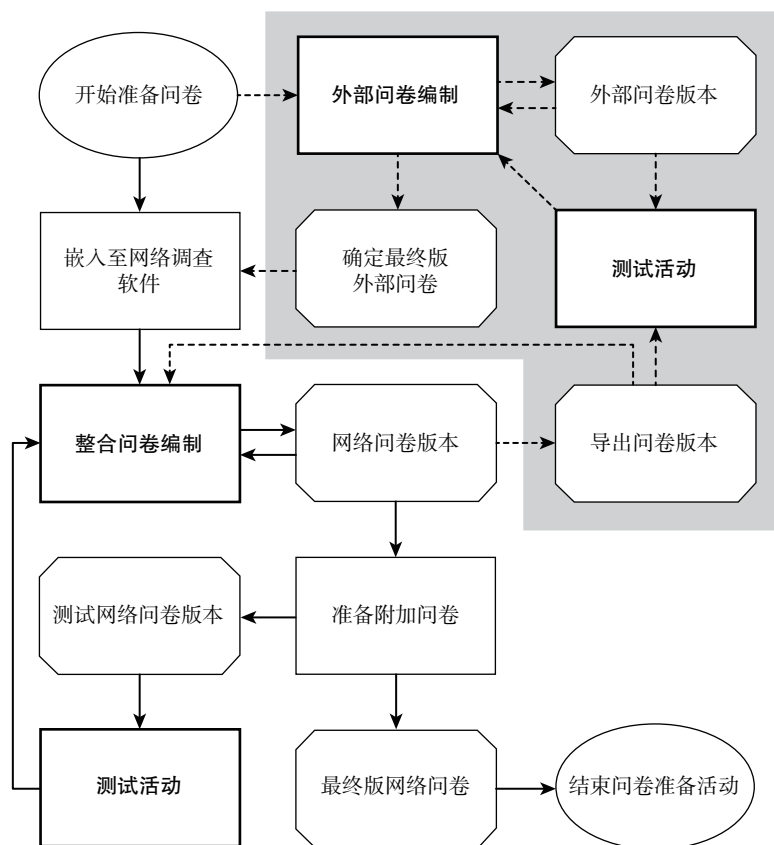


图 2.1 问卷准备过程的流程与结构

度。若仍存在大量问题或关键问题，则需继续测试，否则可进入最终版问卷阶段。

图 2.1 中所示流程可通过不同测试方法及序列组合实现，因此实际操作极为复杂。但这一概念框架有助于明确和结构化问卷编制各阶段及关键环节，尤其是测试轮次与评估方法的数量，进而合理规划测试所需的精力与资源。需强调的是，测试及相关准备工作所需资源往往远超预期，绝不可低估或忽视。

这一框架也为制定时间计划与管理方案提供了基础，当然，这些需在预调查的早期阶段完成（第 2.6 节）。此外，在更广泛的项目管理背景下（第 6.2 节），图 2.1 中间卷编制和测试的关键活动可以直接转化为相应工作包或其组成部分。对于复杂调查，建议咨询专家以确定最优流程与方法组合。

在此需再次强调，如果缺乏计划和资源来进行正式测试，就不应该开展网络调查。至少，测试应该包括以下内容：

- 初期需进行两项正式评估：专家评估（由至少一名调查方法学家或资深从业者完成）和受访者反馈或出声思考的认知访谈。为提高效率，两者均可通过在线评论系统远程实施。
- 基础可用性评估亦是必备环节，在简化场景下可与上述正式评估整合。
- 问卷技术测试是问卷准备的另一个必备环节。
- 还需要对真实受访者进行预研究。
- 如果有发起方或委托方，通常需要收集其反馈意见。

即便是上述最低限度的测试流程，也可能需要至少 5 个正式问卷版本；对于复杂调查（尤其是国际调查），版本数量会急剧增加。让我们以 Vehovar、Petrovčič 和 Slavec（2014）报告的案例结束本节。一份 30 分钟的英文主问卷需被翻译并在 7 国实施，前期需在文字处理软件中生成 6 版外部初稿，以及在带有集成在线评论系统的网络调查软件中准备 13 个计算机化在线版本。四个初步的在线版本也需被导出到离线文本处理器中，以方便外部评论。预研究后还需准备 7 版在线初稿。此后，因校对、翻译、回译、设计及格式化等需求，又产生更多版本。该案例清晰表明，整合问卷编制、测试与计算机化兼具重要性与必要性，因此每个网络调查项目都应为此分配充足的额外资源。

2.4 技术准备

除与调查模式、抽样方案及问卷设计相关的基础方法准备外，预调查阶段还包括技术准备、无应答策略及综合管理等其他活动。本书将从技术准备展开阐述，这部分主要涉及数据的采集与整理（第 2.4.1 节），并行数据的处理（第 2.4.2 节），安全与隐私保障的技术处理（第 2.4.3 节）和技术预测试（第 2.4.4 节）。

2.4.1 采集和整理数据的准备工作

清晰预判需记录的数据类型，以便在数据库准备、受访者追踪、电子邮件设置及样本管理等环节配备相应的技术支持。这些工作虽在很大程度上取决于所选的网络调查软件，但仍存在诸多可自主决策的空间与可选方案。此外，明晰这些问题对选择合适的软件具有重要指导意义。

2.4.1.1 数据库的准备工作

本书中数据库 (*database*) 是一系列结构化文件的集合, 涵盖调查前 (如联系信息)、调查中 (如问卷应答、并行数据) 及调查后 (如权重数据) 收集、获取或生成的数据。从形式上看, 数据库包含多个子数据库, 明确其组成部分、格式、结构、层次及水平, 对高效开展技术准备工作至关重要。以下区分四类原始数据。

- 问卷数据 (*questionnaire data*): 指测量阶段收集的每个应答单位的信息, 包括实质性应答 (*substantive responses*)、非实质性回答 (*non-substantive responses*) (如“不知道”) 和缺失数据的记录 (如项目无应答、中断)。问卷数据是数据库最基础的组成部分, 为与“数据库”这一通用概念区分, 前文已将其称为“数据文件” (见第 2.3.1 节)。
- 并行数据 (*paradata*) 指关于调查应答过程的记录数据 (Couper, 2000a), 以单位为层面收集。并行数据的主要来源是日志文件 (*log files*), 包括受访者每次操作后留下的标准化痕迹 (如时间戳、行动类型)。呼叫记录数据 (*all record data*) 也属于此类, 包括所有关于联系过程的相关信息; 还有一些其他的准数据类型 (如屏幕记录)。
- 辅助数据 (*auxiliary data*) 是指从抽样框或其他一些外部来源获得的单位特征 (如性别、地区、电子邮件地址)。
- 元数据 (*metadata*) 是指从外部获得的或由研究人员指定的数据。它们可以是调查项目层面的数据 (如项目描述、实地调查日期等), 也可以是问题层面的数据 (如敏感性、显著性、测量尺度)。

除上述四类原始数据外, 数据库中还存储着各类衍生数据 (*derived data*), 如权重数据、重新编码或计算得到的数值。

数据库中存储的数据量大、结构复杂且层次多样, 涵盖单位、页面、问题及变量等多个层面。其中, 变量是指统计学概念, 表示数据文件中的一列 (有时也是一行, 取决于数据方向); 也对应于一个调查项目。而问题是一个方法学术语, 表示问卷的一个组成部分, 它有自己的特征 (如测量水平、内容类型、复杂性、敏感性), 并且可以包括更多的项目 / 变量 (例如在表格的情况下)。

当然, 除了调查数据, 并非必须收集、存储和分析上述所有类型的数据, 这主要取决于研究目标和范围。一般来说, 我们不应该收集和储存超出需求的数据, 否则可能导致不必要的复杂性和成本增加。

此外,研究者对特定网络调查软件的数据库结构影响相对有限,一旦选定软件,其提供的功能也基本固定。不同网络调查软件在数据库功能上差异显著,具体取决于:

- 性能 (*capability*) (例如,可支持的问题或单位数量上限、可导出的数据类型及格式、是否能便捷生成并行数据、能否在调查阶段修改问卷内容、如何上传辅助数据、如何导入导出元数据、数据库结构如何等)
- 灵活性 (*flexibility*) (例如,可以改变并行数据收集的设置吗?在数据导出和数据导入方面有哪些选项?可以复制、合并和分类数据文件吗?)
- 运行速度 (*speed*) (例如,在创建问卷、填写答复和生成报告时,技术响应速度如何?以及处理大量变量或单位时的表现如何?)。

尽管上述大部分功能在选择网络调查软件时已确定,但研究者仍需负责部分工作,如定义变量的类型或设置并行数据(如打开 cookie 收集),以及预判并准备组织分析所需数据的程序。

2.4.1.2 受访者追踪和唯一标识符

本节聚焦基于名单与非基于名单的网络调查中,样本单位追踪的技术实现要点及唯一标识符的应用逻辑。追踪样本单位的原因很多,例如,若仅计划向未应答者发送提醒,首先需要追踪所有单位的应答状态。

基于名单的网络调查 (*list-based web surveys*) 邀请特定单位参与时,需通过唯一标识符 (ID) 实现精准追踪。电子邮件虽可直接关联受访者,但出于隐私保护原则,更建议生成独立的唯一 ID 代码,这样研究人员可以控制访问权限(例如,防止不需要的访问或避免重复提交)。对于有意推迟完成网络问卷的受访者,也提供“暂停并稍后继续”的选项;这样,受访者中断后再次访问时,可从上次终止处恢复作答。同样, ID 代码也有助于提醒中断单位继续填写。

ID 代码的登录方式分为自动登录与手动登录两类。自动登录 (*automated login*) 为每个受访者生成唯一的 URL,并将 ID 代码作为基本 URL 的扩展添加到网络问卷。然后, URL 会嵌入电子邮件邀请函中。受访者打开电子邮件邀请时,只需点击个性化 URL 就可以访问网络问卷。

在邮件或其他传统邀请的情况下,则需要手动登录,受访者首先使用邀请函中的网络调查问卷的初始 URL 地址,接下来输入在邀请函中收到的 ID 代码。这个过程被分成两个阶段,因为完整的个性化的 URL 地址会太长。

初始 URL 应该是简单而短小,最好有内容定制(例如 www.researcher.com/

projectname), 而不是用网络调查软件生成的默认技术链接。此外, 对于传统的邀请 (如邮件), 还应该确保初始 URL 地址出现在搜索引擎结果首位, 因为许多受访者错误地或有意地将 URL 地址输入搜索引擎, 而不是输入网络浏览器地址栏。

手动登录有可能输入错误的 URL 或 ID 代码。为了避免这种情况, URL 需要简单; 而密码需要简短, 同时足够复杂, 使人难以轻易猜测。因此, 密码避免使用连续数字, 而使用包括字母和数字 (例如 2X0A11) 的分散随机组合比较合适。成对的 ID 代码应该至少有两个字符不同, 以减少输入错误或恶意猜测。也可以结合半自动登录程序, 即通过邮件发送个性化 URL, 同时要求受访者手动输入另一组验证码, 既提升安全性, 又保留个性化体验。

有时会使用团体密码, 这类密码通常是基于一些容易记住的词 (如组织名称), 且通过离线传达 (如在组织内部或在教室里)。然而, 这仅适用于群体区分, 无法替代唯一 ID 的个体识别功能。此外, 部分调查软件可以在不使用团体密码的情况下, 支持为特定群体生成专属 URL。

自动登录可以伴随电子邮件邀请, 以减少受访者操作负担, 并有助于避免输入错误; 而手动登录则增加了个性化体验, 减少了感知匿名性。有时, 在手动登录的情况下, 由于个性化体验增加, 应答质量会更高 (中断更少, 开放式问题的答案更丰富) (Heerwegh & Loosveldt, 2002, 2003)。然而, 这些效果较弱, 且在其他一些研究中没有得到证实 (Crawford et al., 2001)。正如预期, 手动登录也会出现对敏感问题披露较少的情况 (Joinson, Woodley, & Reips, 2007); 而且与自动登录相比, 回复率也较低 (30% vs 35%)。一般来说, 本书推荐自动登录, 但有明确证据表明手动登录带来积极效果时, 可灵活调整。

无论采用何种登录方式, 我们都应该认真准备, 并检验对受访者的指导语。针对邮件邀请, 需提供更为清晰的指导。对互联网操作经验较少的群体, 则需要附加说明 (例如附上屏幕截图) (Dillman et al., 2009, p. 287)。说明应该简短, 并以无障碍和非要求性的语言撰写。针对邮件邀请, 需提供清晰的替代访问方案 (例如“输入 URL 地址”“点击链接”和“复制地址并粘贴到你的网络浏览器”)。

非基于名单的网络调查 (*non-list-based web surveys*) 无法实现为个体分配唯一标识, 因此受访者追踪面临更大挑战。在这种情况下, 使用 IP 地址 (*IP address*) 替代唯一标识符。IP 地址是分配给所有登录到互联网的设备的数字, 包括受访者的设备。然而, 也存在动态 IP 地址, 当同一台设备 (如电脑) 在不同时间登录互联网时, 可以分配不同的 IP。同样易出现问题的是组织 (团队) IP 地址, 组织中的所有设备都使用同一个 IP。如果每个 IP 地址只允许一个受访者访问, 那么只有一

个受访者可以参与调查。尽管如此,阻止来自同一 IP 地址的多个回应,如结合时间戳(在短时间内来自同一 IP 地址的大量访问),有助于防止简单欺诈(例如,虚假回应)或来自单一设备的服务器攻击。

另一个选择是小型文本文件(*cookies*),它是由托管网络问卷的服务器发送、存储在受访者设备上的小文本文件。服务器为 cookie 分配唯一代码,以及记录时间戳、最后回答的问题等信息(ESOMAR, 2011a, p. 10)。下次受访者访问相同的网络问卷时,网络调查软件就会识别这个单元,以防止重复访问,或者使受访者在暂停或中断后顺利继续。然而,cookies 与所使用的浏览器有关,所以受访者仍可通过在同一设备上使用不同浏览器、使用另一个设备或删除 cookies 来多次填写问卷。由于道德和法律限制,Cookies 的使用限制日益严格,且部分用户会主动禁用此类功能,因此 cookies 作为 ID 代码替代方案的作用较差。

Cookies 通常是临时文件,可以设置为与用户的浏览器会话一样长,但也可以设置为一个小时、一天、一个月甚至永久。某些软件需要 cookies 作为回答问卷的技术支持,并在会话终止时(完成、中断、暂停)失效。持续时间超过问卷调查时间的 Cookies 可能会受到法律限制,并且可能需要用户明确确认,这将使回答过程更加复杂。

一般来说,我们不建议使用 IP 和 cookies 来限制对调查的访问,因为它们会导致预期或意外的障碍。多数网络调查软件也默认关闭此类功能。然而,如上文所述,在某些特定情况下,它们可能是有用的,特别是当知道目标人群不会出现上述问题时。

由于这些限制,没有个人 IP 号码追踪且无 cookie 的网络调查正日益流行。

在非基于名单的调查中,很难防止同一人的重复多次参与,但可以在调查后采用某些措施(第 4.1 节)来检测并删除(例如如果两个单位的所有回答完全相同)。关于“暂停并稍后继续”,IP 地址或 cookies 可能更有帮助,特别是在假设大多数受访者使用它们时没有上述任何问题的情况下。否则,“暂停并稍后继续”选项要求受访者在暂停时提供电子邮件,以便稍后提醒其继续调查。

在非基于名单的网络调查中,也可以先招募单位,获得其电子邮件地址,然后发送定向邀请。在这种情况下,基于名单的调查的特征完全适用。

2.4.1.3 电子邮件设置

本节默认以基于名单的电子邮件邀请调查为讨论对象。在这种情况下,电子邮件邀请的技术设置是至关重要的。

对研究者而言,最简便的方式是使用默认电子邮件系统(*default email system*)。需将收件人地址填入邮件标题的“密送”(BCC)栏,而非“抄送”(CC)栏——若使用抄送,所有收件人将看到完整的收件人列表,这在道德和法律层面均不合适。

当然,这种方式局限性显著,最多仅能处理数百个地址;当地址数量更多时,电子邮件系统会限制批量发送数量。主流电子邮件服务(如雅虎、Gmail、MS Outlook等),默认限制通常设置为每天发送几百封电子邮件¹⁶,而机构内部的电子邮件系统可能允许发送数千封,甚至无明确限制。

更关键的限制出现在收件人端。电子邮件系统采用日益复杂的垃圾邮件检测过滤器,会拦截“抄送”或“密送”栏中包含大量地址的邮件。过滤器还对发件服务器的技术设置、邀请函的主题及内容敏感(关于电子邮件邀请函的格式,详见第2.5.7节)。

相比使用“抄送”或“密送”批量发送,通过改进脚本¹⁷可降低被识别为垃圾邮件的概率。这也使研究人员得以进行个性化的邀请(例如在电子邮件中包括姓名),增加发送的电子邮件数量,突破与抄送和密送有关的限制。

可以通过研究者的默认邮箱发送几百封邮件。如果采用一封邀请邮件,再加上一封对已回复者的“感谢”邮件或对未回复者的提醒邮件,会是很好的实践方案。这种方式在实际操作中被广泛提及,下文将其称为双邮件设置(*default email system*),即不单独追踪未应答者,而是向整个名单中的受访者发送两封邮件。

但对于高频使用或大规模的名单,通常需要专业工具。这类工具多为专门的外部大规模电子邮件服务,最初为各类警报、通讯及直接营销场景开发。此类工具能有效规避垃圾邮件拦截,并具有精心设计的功能,如个性化、检测电子邮件状态(如电子邮件发送、交付、打开)、管理电子邮件列表、邀请的图形设计、取消订阅(退出)等。若网络调查软件的邮件发送功能较弱或缺失,这类工具是理想选择。但需注意,其与网络调查软件的分离可能导致中断、暂停及部分受访者状态追踪困难。

最通用且理想的选择是利用网络调查软件的集成邮件支持(*integrated email support of the web survey software*)。此类功能通常能覆盖与专用电子邮件应用程序相同的功能,并与基于唯一标识符的追踪系统联动。具体可分为以下三种方案。

- 当网络调查软件供应商提供托管服务(SaaS)时,研究者通常需使用供应商提供的邮箱地址(例如:survey@supplier.com)作为发件人电子邮件。遗憾的是,多数供应商仅允许研究者自定义前缀中的“名称”(例如,name@supplier.com)或添加一个“研究人员”子域名(name@researcher.supplier.com)。然而,

网络调查软件很少提供改变域名（例如，使用“@researcher.com”而不是“@supplier.com”）需要的特殊程序。若缺少这些特殊程序，发件服务器的域名（又称域名密钥，如本例中的“supplier.com”）与电子邮件域名（如 @researcher.com）不匹配，会增加邀请邮件被标记为垃圾邮件的风险。因此，使用供应商域名（如 @supplier.com）是普遍做法。

- 当研究者将软件安装到自己的服务器上时，可完全自主设置发件人电子邮件地址，因为两个域（电子邮件地址域和服务器的域名）现在可以自动匹配（例如 researcher.com）。
- 通过网络调查软件电子邮件系统发送邀请这一过程，如果安装在专用服务器上（如 SaaS），则可以集成到研究者或研究机构的电子邮件系统中。这一方案优势显著，既能保留网络调查软件邮件系统的全部功能（如追踪、独立发送），又能借助研究者自有邮箱发送的优势（如发件人个性化、增强合法性）。当然，这种方式可能受限于研究者服务器的性能（如发送量、速度），且商业网络调查软件供应商极少提供此类选项。

第一种选择可能具备技术优势，因为供应商提供持续的专业支持，但在方法学层面可能存在不足，涉及发件人的电子邮件对受访者参与意愿的影响。要知道，参与调查的决策往往是在几秒钟内做出，因此发件人的电子邮件域名在这个过程中作用关键。这使得第三种综合方案非常有利，因为研究人员掌控域名。

通常情况下，研究者对上述方案的选择空间有限。但在有选择余地时，需审慎考量使用供应商、研究者（研究机构）、分包商或发起方的电子邮件域名（或子域名）的潜在影响。当然，具体情境决定最终选择：有时，知名网络调查软件供应商（如 info@supplier.com）或大型市场研究机构的邀请邮件，在收件人心中的认可度可能低于研究者或发起方的直接邮件；但在某些场景下，情况可能相反。

此外，邮件前缀的优化也很重要。通常而言，研究者的个性化前缀（如 ‘researchername@’）或项目专属前缀（如 ‘projectname@’）会比通用前缀，如 ‘info@’ 效果更佳。另外，第 2.5.7 节将讨论电子邮件邀请的格式。

还需注意包含“退订”选项，也称为“选择退出”或“从列表中删除”，因为这是法律和道德规范的明确要求，也能避免受访者因愤怒或反感等情绪选择将电子邮件邀请标记为垃圾邮件，进而永久地、全面地影响电子邮件发送地址的信誉。

技术层面需确保退信能被转发至研究者可访问的邮箱，而非无法查看的公共邮箱。这些退信应与对应的网络调查项目关联，并匹配至具体的受访者单位。通过基

于名单的样本发送电子邮件邀请时,需能技术识别以下邮件状态:(a)由于发件电子邮件服务器的技术问题而导致电子邮件发送失败;(b)电子邮件名称或域名未知或不存在;(c)收件箱已满;(d)自动回复(如缺席通知);(e)电子邮件未送达的其他原因(如收件人的服务器或设备的各种问题),以及追踪诊断;(f)电子邮件成功发送或(g)电子邮件被打开。此外,(h)在选择时也应记录明确的退订选项。这些技术状态将用于无应答指标的计算(见2.5节)。

最后,可以预见,随着技术复杂性的提升及反垃圾邮件限制趋严,网络调查软件的高质量邮件支持将成为实现有效电子邮件邀请的核心要素。

2.4.1.4 样本管理

样本管理整合了数据库准备、受访者追踪和基于名单的网络调查电子邮件设置等。当网络调查软件用于发送电子邮件邀请并追踪唯一标识符(ID)时,会出现默认情况。样本管理通常遵循以下流程:

1. 导入带有联系信息(电子邮件地址、姓名、职称等)的样本单位列表;
2. 标记需发送电子邮件邀请的样本单位;
3. 发送带有ID代码的电子邮件邀请函;
4. 通过ID追踪样本单位的应答状态;
5. 追踪被退回的电子邮件,并对其进行适当分类(地址不正确,某些不在办公室的回复,等等);
6. 标记未应答者或特殊状态单位(如自动回复),以便发出提醒;
7. 发出带有ID代码的提醒函;
8. 使用ID代码追踪提醒发出后的应答状态;
9. 处理被退回的电子邮件,并对其进行适当的分类(地址不正确,外出回复);
10. 重复步骤6-9,进行多轮提醒。

如果我们决定不追踪应答状态,仅采用单次邀请函或双邮件设置,则可简化上述流程。同样地,若使用外部邮件服务、邮寄邀请或多模式招募(如结合电子邮件与电话),流程也需相应调整。

在线样本库的样本管理需增加额外任务,包括样本库维护、将受访者分配至特定调查项目等。混合模式调查中,也会出现复杂的情况,如跨模式的单位追踪变得越来越复杂,这类情况通常在个案管理(*case management*)框架下处理。

高效的样本管理依赖于优质的网络调查软件，其需具备以下功能：可以追踪每个单位的电子邮件发送记录并邀请、提醒和联系，包括每个电子邮件邀请的档案（如谁、何时和什么）。这种管理产生的数据也是特殊类型的准数据，称为联络记录数据（见下文）。软件还应支持数据库操作，如合并、组合、创建和删除各种样本列表，以及识别和消除重复单位，这些都极其重要。

2.4.2 并行数据¹⁸

并行数据通常指关于调查实施过程的记录数据（Couper, 1998）。这类数据对于网络调查尤其重要，不仅丰富，还易于捕获，且大多并行数据与各种数字记录和正式调查步骤中留下的痕迹有关。

直接并行数据（*direct paradata*）是通过网络调查软件或外部脚本自动记录的信息，是正式调查过程的副产品。它们可以分为三大类：联系信息（*contact info*）并行数据、设备类（*device-type*）并行数据和问卷导航（*navigation*）并行数据。

另一方面，间接并行数据（*indirect paradata*）则需借助额外外部设备获取，如视频记录系统、脑电波监测设备、眼动追踪装置；或通过外部人员观察获取，如行为编码。这些间接并行数据主要用于测试环节，在第2.3.4节中已经论述过。

并行数据主要记录受访者的行为，但研究者在现场或实地调查步骤中的行为通常属于联系信息并行数据（例如发送电子邮件通知）；本书将其称为联络记录数据（*record data*）。

在这一背景下，并行数据还可以包括预调查阶段（例如，问卷的变化，谁改变了什么，什么时候改变的）和调查后阶段（例如，计算派生值或更改反应值：谁改变了什么，什么时候以及如何改变的）研究者行为（*researcher's actions*）的数字痕迹。这些数据对于记录、澄清和归档，以及最终优化流程都非常重要。高级网络调查软件应该支持捕捉这些数据，但这些轨迹已经超出了并行数据的概念范畴。

混合模式系统会出现一系列额外的并行数据，从多模式联系、访员特征和访员观察到额外的音频或视频记录（Kreuter, 2013）。

本节后续内容聚焦直接并行数据，先介绍典型案例，再探讨其与软件、分析及伦理相关的问题。

2.4.2.1 联系信息并行数据

联系信息并行数据包括关于尝试联系受访者的信息。它们与电话或面对面调查中的传统呼叫记录数据非常相似，以下是一些例子。

电子邮件邀请的结果可以被记录，并以上述方式进行结构化处理（例如，电子邮件已送达、电子邮件未送达、电子邮件发送错误，等等）。

访问问卷介绍页面是进一步的基本并行数据。它可以提醒尚未回复的受访者，以及那些只进入介绍页面后立即离开的受访者。这些基本并行数据是计算应答率的基础数据。

受访者中断前的最后作答问题，可用于判断应答完整性，以及该样本单位是否可以用于最终数据文件（详见 2.5 节）。它也可以用来发送更复杂的提醒，引导受访者从断点继续作答。

2.4.2.2 设备类并行数据

设备类并行数据包括被调查者在回答网络问卷时使用设备的技术信息，主要内容如下。

用户代理字符串是受访者访问网络问卷时，其浏览器向服务器发送的一串文本。该字符串所包含的信息可立即用于提供有针对性的内容和格式（例如，将问卷调整为智能手机版本），也可以用于方法学的目的（例如，评估不同浏览器、设备类型、操作系统、屏幕尺寸下的数据质量）。例如，Callegaro（2010）分析了美国的一项客户满意度调查，该调查针对移动设备、台式机/笔记本电脑进行了优化，结果发现使用移动设备的受访者中断率是桌面设备的 3 倍。如今，随着智能手机的兴起，统计设备类型（台式机/笔记本、平板电脑、智能手机）已成为常规诊断手段。

检测 JavaScript 和 Flash 信号，记录受访者的浏览器是否支持 JavaScript 和 / 或 Flash，第 1.3.2 节和第 1.3.3 节中介绍了这些信号。高级的网络问卷调查经常依赖于 JavaScript，所有的浏览器一般都默认启用 JavaScript，且许多基本的互联网服务也需要 JavaScript。尽管如此，仍有少数用户可能会禁用它，需监测其比例，然后研究者可以邀请他们启用这个选项，或者为其设计无需 JavaScript 的问卷。Flash 技术需要在受访者的设备上单独安装，而不仅仅是一个选项设置问题。因此，未支持的用户比例通常更高（达两位数）。

Cookie 和 IP 地址记录在网络调查中可能是有益的，但也可能是有问题的。尽管如此，当 IP 地址被单独记录且不直接关联单位信息时，可以据此获取有用的地理信息（如国家、地区）。

2.4.2.3 问卷导航并行数据

相较于联系信息与设备类并行数据,问卷导航并行数据更为复杂且丰富。下面将讨论几个常见的案例,也存在许多其他例子,例如检测在网络问卷窗口外点击的受访者或对回答过程的屏幕进行视频捕捉。

下文列举了相应的并行数据使用的经典案例。

浏览屏幕或反应延迟所花费的时间可能是此类并行数据最常见的类型,相关研究也出现在几篇文献中。Couper 和 Kreuter (2013)对反应延迟的并行数据进行了较好梳理,尤其是它在揭示应答过程方面的用途。在网络调查研究中,较长的反应延迟与深层次的认知加工有关,这可能是由于动机或参与度提高 (Callegaro et al., 2009)、中断或多任务处理 (Stieger & Reips, 2010)、问题复杂性增加 (Yan & Tourangeau, 2008),以及在态度类问题中的矛盾态度和在知识类问题中的知识缺乏 (Heerwegh, 2003)。Yan 和 Tourangeau (2008)的研究还强调了受访者特征的影响,年轻、受教育程度高或有更多调查经验的受访者通常应答更快。

关于反应延迟的研究更多关注的是快速 (*speeding*) 作答问题,即反应延迟过短,以至于研究者对受访者是否认真阅读了问题并在反应过程的各阶段都表现出最佳状态产生了合理怀疑 (Beckers, Siegers, & Kuntz, 2011; Kapelner & Chandler, 2010; Zhang, 2013)。因此,这种分析涉及检测极端欺诈、异常情况和虚假回应,但也可以作为对数据质量潜在威胁的指标。然而,受访者作答问卷过快,快到被宣布为超速者 (*speeder*) 的边界或基准,在技术上较为复杂。这主要由于时间戳多对应页面而非单个问题。反应延迟的非正态分布 (非对称性、偏态) 也会导致各种统计学上的复杂性。

为此,研究者开发了各种指标,从简单的方法,如计算偏离中位数或平均值的 2~3 个标准差 (West & Sinibaldi, 2013),到更复杂的解决方案,如“速度指数” (Steinbrecher, Roßmann, & Bergmann, 2013) 或基于平均阅读速度定义截止点 (C. Zhang, 2013)。通常使用每字 0.4 秒为截止值用来识别速度快的人,而使用每字 2.5 秒的高截止值界定速度过慢的人。现有研究中,超速者占比差异较大,从低于 2% (West & Sinibaldi, 2013) 到高达 16% 不等 (Beckers et al., 2011; Gutierrez, Wells, Rao, & Kurzynski, 2011)。

需注意,超速未必等同于低质量应答,因为受访者间的速度感差异较大 (Garland et al., 2013)。尽管如此,极端快速作答通常与低数据质量相关 (Thomas & Barlas, 2014; C. Zhang & Conrad, 2013)。同样,通过将单词按顺序显示在屏幕上 (例如每分钟 250 个) 迫使受访者缓慢阅读,可减少敷衍作答行为 (Kapelner & Chandler, 2010)。

按键和鼠标点击可通过特殊脚本捕获。例如, Stieger 和 Reips (2010) 在一项有 23 个问题的网络调查中捕捉了 1 046 名参与者的 336 262 个操作行为。这为测量每个页面的鼠标轨迹总长度(即受访者回答问题时的鼠标移动)提供了基础, 其中 11% 的受访者表现出过多的鼠标移动。分析结果还显示, 6% 的受访者过度点击, 4% 的受访者长时间无操作, 而大多数受访者至少有一次在没有足够时间阅读问题的情况下回答了问题。

改变答案指受访者修改了最初的答案, 这表明其对某一问题可能有困惑。例如, Stieger 和 Reips (2010) 的研究表明, 修改答案确实存在, 而且态度问题的答案改变(5.4%)比事实问题(1.5%)更为频繁。

每当受访者被提示和提醒的时候, 就会出现实时验证信息, 通常以弹出式窗口发出(例如, 项目无应答或无效输入)。相应的分析可以提供对受访者行为的额外了解。还可以使用 Haraldsen (2005) 提出的质量指数 (*quality index*), 通过激活的错误提示与潜在提示总数的比值衡量问卷, 优化问卷设计, 从而减少激活的提示信息数量。Callegaro (2014b) 也提供了类似的质量指数计算案例。

2.4.2.4 采集并行数据的软件

研究者可通过自己编写脚本、使用专门的并行数据软件或者使用网络调查软件的相应功能三种方式实现并行数据采集。

专门的并行数据软件类型丰富, 既有基本脚本(如客户端并行数据 CSP¹⁹ 和通用客户端并行数据 UCSP)(Kaczmirek, 2009)²⁰, 也有各种高级并行数据外部工具, 如 Stieger 和 Reips (2010) 使用的工具, 但网络调查软件的内置支持相对有限。Vehovar、Čehovin、Kavčič 和 Lenar (2012) 回顾了大量的网络调查软件供应商, 发现其中只有少数软件提供了对并行数据的简易收集和访问。30% 的软件可提供问卷填写总时长数据, 4% 的软件支持页面级时间记录, 仅 1% 的软件能够追踪受访者在问卷页面间的跳转轨迹。

从收集数据的角度来看, 可以区分基础并行数据 (*basic paradata*) 和高级并行数据 (*advanced paradata*)。基础并行数据是常规收集和报告的, 要么作为调查过程的一部分自动采集(例如, 发送电子邮件邀请的信息), 要么只需极少额外操作(例如, 时间戳、用户代理字符串、IP 地址), 因此通常包括在(高级)网络调查软件的默认功能中。高级并行数据 (*advanced paradata*) 需要激活网络调查软件中的特殊选项或使用外部脚本获得, 例如精确的鼠标点击次数、键盘输入、验证提示记录、鼠标滚动轨迹、屏幕捕捉、问卷窗口切换等。无论是哪种类型, 采集并行数据都可

能减慢调查进程并产生大量数据。因此,若不打算使用高级并行数据,应避免采集以减少不必要的负担。

2.4.2.5 对并行数据的分析

并行数据的分析可以在四个水平上进行(Kaczmarek, 2009, p. 83)。

- 第一水平并行数据记录受访者的个体行为,如时间戳或鼠标点击。这类数据通常按行为发生顺序记录,因单屏内的行为数量无法预先确定,第一水平数据产生了复杂的数据集(非矩阵),这使得数据分析从一开始就很困难。最终的邀请(时间戳、邀请或提醒的内容)和其他呼叫记录并行数据也属于第一水平。
- 第二水平并行数据是同一受访者的第一水平数据汇总。例如,在一个页面上点击鼠标的总次数或每个问题答案的更改次数。这些数据可以被结构化为一个矩形的矩阵,更容易后续分析。
- 第三水平并行数据是在不同受访者或不同变量间将第二水平数据进行汇总。在这个层面上,可以计算出每个受访者(对所有问题)或每个问题(对所有受访者)的平均答案变化数量。
- 第四水平是最高级别的汇总,发生在受访者和变量之间。它提供了每项调查的整体指标,如平均完成调查时间(秒)或项目无应答率。

第一水平数据是分析的基础,可以用多种方式组织加工。当与元数据、辅助数据和问卷数据结合时,可以组织成多个关系型子数据集,其中各特征对应分配到调查、单位、页面、问题、变量和个人回答条目的相应层次。对于后者,还可以区为实质性回答、非实质性回答或各种缺失值的状态。这些子数据集可以附带反应选项标签、变量标签和问题文本。分析时如何组织这些分类的子数据集,取决于研究者的目的。实际分析中,常将数据整理为矩阵形式,如一行包含某受访者在某页面的所有行为数据。当然,具体方式需依研究目标而定。

2.4.2.6 并行数据的隐私和伦理规范

当并行数据被视为个人信息时(例如来自用户代理字符串的IP地址),它们受法律保护,并需要特殊处理。例如,欧洲民意与市场研究协会(ESOMAR)的在线研究准则要求遵守相关地方法规,因为它与欧盟立法类似,将IP地址归类为个人数据(ESOMAR, 2011a, p. 6)。尽管美国未对此作出正式规定,但心理学领域的文献普遍认为:如果记录IP地址并整合到数据文件中,那么研究者不应该声称研究

是完全匿名的 (Barchard & Williams, 2008)。

并行数据的采集也引发了伦理争议与实践困境：如何向受访者告知数据采集行为？研究者可选择明确请求受访者同意，或在条款中以小字说明。但若完全不提及则可能被视为不道德。Couper 和 Singer (2013) 的系列实验表明，任何提及并行数据采集的行为都会降低受访者的参与意愿（例如从 65% 降至 55%），且在调查结束时才告知的情况下，受访者的同意率更低。

并行数据能为应答过程提供深入洞察，不仅对测试环节有价值，也有助于理解应答性质与质量。尽管本书开始声称并行数据含有丰富的信息且容易收集，但也同意 Couper (2005, p. 494) 和 Nicolaas (2011) 的观点，即分析并行数据实际上是非常有难度的。虽然基本数据（呼叫记录、设备字符串、时间戳）是默认采集的，而且其分析也形成了成熟的诊断流程，但非标准化的高级数据情况却非常不同，往往没有足够的结构化，需投入额外资源进行预处理。因此，研究者需审慎评估并行数据的附加价值，并合理规划资源投入。

2.4.3 支持安全和隐私的准备

网络调查常涉及敏感及机密信息的收集与交换，研究人员需要确保 (*secure*) 数据及相关流程的安全性。这些数据与流程需具备可靠性与安全性，能够抵御攻击并防止第三方非法访问。安全涉及数据传输 (*transmission*) 过程，意味着保护客户端（受访者的计算机）和服务器（托管调查问卷的服务器）之间的通信安全，以及将数据文件移交给客户端或用户的过程安全。安全存储数据 (*storage*) 是确保对网络调查软件运行服务器的访问，而安全存档数据 (*archiving*) 是指保护存储于其他计算机、服务器和设备上的数据。

上述问题需在相关合同中明确规定，以规范所有参与方（包括客户、研究人员、分包商、网络调查软件供应商等）的权责关系。这在市场营销研究中尤为关键，因为昂贵的网络调查需严格保密，防止被竞争对手获取。同样，在官方和学术调查中，关于数据预发布禁令及数据使用主导权的界定，也与数据及流程的保护密切相关。

安全问题与受访者的隐私担忧高度重叠，这一问题涉及诸多法律与技术层面，极为复杂。Cho 和 Larose (1999) 已揭示调查研究场景中隐私的关键维度：除上述适用于个人应答的安全问题外，还涉及与匿名性 (*anonymity*) 有关的特定问题。匿名性是指个人数据（标识符）(*identifiers*)、调查应答以及发布个人数据的条件控制间的联系。另一具体的隐私方面是个人应答的保密性 (*confidentiality of individual*

responses), 涉及个人信息的二次使用, 即禁止用于初始目的之外的场景。此外, 保密性还存在其他层面 (例如严格的商业保密性), 所以数据文件可以是保密的, 而不会对受访者回答的保密性造成任何威胁。

研究人员仅能部分保障安全与隐私, 因为这很大程度上取决于所使用的网络调查软件、服务器技术规格及托管环境。尽管如此, 研究人员及其他调查相关主体 (如机构审查委员会) 仍承担重要责任, 需具备一定的技术知识。为此, 下文简要介绍与隐私和安全相关的关键问题, 暂不涉及概念定义、理论阐释及深入技术细节, 重点聚焦对研究人员有实际意义的建议。关于这一主题的更全面的法律、专业及伦理层面的探讨, 可参阅 Corti、van den Eyden、Bishop 和 Woollard (2014) 和第 6.3 节。

2.4.3.1 服务器安全

网络调查的数据收集可依托企业或研究机构内部托管的服务器, 也可采用网络调查软件供应商普遍选择的 SaaS (软件即服务) 解决方案进行外部托管。外部托管还包括研究人员在非专用服务器上安装开源软件的场景, 此类服务器并非专门用于调查数据收集。尽管托管场景不同, 但所有情况下的安全准则是一致的。许多研究人员认为, 将数据托管在内部服务器更安全, 理由是机构了解并控制所有可以访问服务器的人员。此外, 从技术角度看, 一旦服务器接入互联网, 便面临相同的安全风险。事实上, SaaS 解决方案反而可能更安全: 安全是其核心业务, 相比小型机构, SaaS 供应商能投入更多资源和专业知识解决安全问题。这实际上形成了一个矛盾, 许多研究机构因安全顾虑无条件拒绝外部托管, 进而排斥不提供内部安装选项的网络调查软件。

除了服务器位置的困境, 还需关注其他关键安全方面。以下我们以提问形式对 Buchanan 和 Hvizdak (2009) 的研究加以改编和拓展, 作简要梳理:

- 托管服务器是否提供第三方安全认证? 如果是的话, 认证更新周期如何?
- 第三方能否访问数据, 相关访问流程是什么? 对可以访问数据的工作人员有哪些筛选、监测和控制措施?
- 渗透测试 (*pentest*) 的实施频率如何? 该测试通过模拟服务器攻击以排查安全漏洞。
- 服务器物理位置采取了哪些数据保护措施 (如电力保障、防火、防盗)?
- 对于 SaaS 解决方案, 企业在何处 (物理位置、国家) 存储数据及备份? 当地法规中是否有任何可能相关的具体规定?

- 服务器故障时（如一台服务器崩溃，另一台接管）的应对流程（包括通信机制和响应时间）是什么？数据传输至新服务器时是否采用加密形式？
- 研究人员完成工作后，数据将如何处理？数据何时以及如何被删除？数据将在服务器上保存多长时间，这是否与研究人员或机构政策所规定的时间框架冲突？
- 在正式调查中，如何提供数据镜像过程？如何频繁地进行备份和归档？
- 如果数据因安全漏洞而丢失或泄露，谁来负责，责任性质是什么？在数据丢失或被入侵的情况下，有什么机制可以提醒研究人员？
- 数据收集期间，调查平台抵御攻击的安全性如何？更具体地说，针对分布式拒绝服务（DDOS）攻击（即向服务器发送大量请求）有哪些预防和恢复程序？例如，在 2014 年，这些攻击使一些最大的 SaaS 网络调查软件供应商瘫痪数日。服务器正常运行时间（即服务器工作时间占比）是多少？99%、99.9% 还是 99.99%（后者正成为行业基准）？

上述问题需要在 ICT 人员的协助下进行评估，他们应澄清技术细节，确保答案符合当前安全标准。评估深度取决于调查的重要性：对于简单的自助式调查，对 SaaS 供应商的基本信任可能已足够；而更高的安全性则需要更多资源投入。具备 ISO 标准安全认证（如 ISO 27001 信息安全管理体系标准、ISO 27002 信息安全管理体系实践准则）或 SAS 70（服务组织控制报告）认证，无疑是朝着建立安全流程和文件管理迈出的重要一步。

2.4.3.2 网络调查应答过程的安全

问卷填写过程（即测量会话）涉及多项安全考量。首要问题是确保受访者设备与服务器之间的通信安全，需通过加密协议（如安全套接字层 SSL）实现数据传输加密，防止第三方截获或篡改信息。

另一风险源于网络调查软件可能通过注册程序遭遇恶意攻击，攻击者直接访问子页面或文件和测量会话，在开放式问题的回答输入中加入破坏性的脚本。

此外，程序漏洞和网络调查软件错误是潜在风险。有时，意外的大量应答（包括脚本和机器人提交）都会导致网络调查的应答过程变慢，甚至中断。这类问题可能仅影响特定调查，而服务器及其他调查仍可正常运行，但定位故障源头往往存在难度。

无论何种情况，均需明确问题应对流程。尤其是，需预先确定技术人员与研究

人员的故障通知机制及响应速度。理想状态下,自动监测程序应能实时发现异常并通过短信等方式立即提醒技术人员(例如通过短信)。这往往能使他们在几分钟内解决这个问题。如果技术人员无法即时解决问题,需及时通知(例如通过电子邮件)研究人员启动危机管理预案。如果是新的网络调查软件,或者服务器托管解决方案不是完全专业化的(例如,服务器不是专门用于网络调查的),建议研究者定期检查网络调查问卷的访问状态。

2.4.3.3 匿名性的技术问题

处理或收集个人数据时,若明确承诺为受访者保密并保障匿名性,研究者需要特别谨慎处理直接标识符。直接标识符可用于识别一个人,并与调查应答和其他个人数据相关联。识别资料不仅可以从抽样框中获得(如用户名、密码、姓名、联系信息),还可以从行政记录、既往调查和此次调查(如要求提供电子邮件地址)获得。如前所述,在一些国家(如欧盟国家)中,某些数据如IP地址,也被视为个人标识符。而许多标识符是正式调查过程中不可缺少的,如样本管理、激励、访问问卷、多次填写和追踪无应答者。

标识符应与调查应答数据分开存储(Barchard & Williams, 2008; Thiele & Kaczmirek, 2010),且需从用于分析的数据文件中完全排除标识符。这一点越来越成为法律要求和专业行为准则的建议,许多网络调查软件包会自动分离这两个文件。

此外,时间戳、记录编号等信息也不能用于间接识别个体,需谨慎处理。在线小组中,个人标识可将所有调查中的回答联系起来,所以在任何情况下都不应该对外使用。一旦在正式调查阶段和调查后阶段中不再需要将个人标识符用于行政管理目的,应永久删除。

我们还需警惕统计信息泄露风险(*statistical disclosure*)。交叉表或公开数据文件中,变量组合(如地区、性别、职业)可能间接识别受访者,这一问题需在调查后阶段处理(第4.3.2节)。

需明确的是,基于名单的调查中无法实现绝对匿名,因个案管理会将ID代码、个人标识、邀请及应答关联。受访者知道研究者实际上是在处理个人信息,原则上不存在匿名问题。在这种情况下,保密起着关键作用。受访者需要信任研究者采取所有措施来保证保密,不仅仅是提供匿名的程序,还需要一系列其他的安全措施。这对于敏感话题(如健康)和社会赞许性话题(如赌博)尤其重要,在人际关系调查(如学生评价调查、员工满意度调查)和商业调查(如财务数据)中也是如此。

2.4.3.4 物理和电子访问数据文件的保证

部分情况下,数据文件可能不涉及隐私或保密问题,可自由访问和发布,但通常并非如此。除服务器访问控制外,还需对其他数字存储介质建立严格的访问限制。许多研究机构的安全规则已明确相关流程,包括保护研究人员的计算机(台式机/笔记本)、外部通信(如网络浏览)和外部存储设备(如CD、USB 钥匙)。Thiele 和 Kaczmirek (2010) 建议对整个硬盘进行额外加密以确保数据安全,特别是增加对个人数据的保护。另一套安全措施是,确保相应的计算机在操作系统、网络浏览器以及反病毒和反间谍软件方面的安全补丁保持更新。

当交换数据时,文件应该额外加密(Barchard & Williams, 2008),并设置访问权限编码。同样地,当通过电子邮件发送数据时,Thiele 和 Kaczmirek (2010) 建议对整个电子邮件通信进行全程加密。另一个选择是,在同事、用户或客户之间共享文件时使用基于云的安全存储服务。当使用网络发布时,还应该包括用于传输文件的SSL 协议和加密。

除了技术保护,还需要特殊的安全政策来确保相关人员的责任行为。通常这一点会纳入至组织规则和协议中。而在非机构研究中(如自助式调查),建议由所有相关方签署保密声明(*statement of confidentiality*),明确数据使用目的、保护措施、期限及法律责任。当组织的机密数据被提供给研究人员时,也建议采取类似措施。

2.4.3.5 访问审计

往往会被忽视的环节是,对能够访问数据文件和其他机密数据的个人进行审计(Corti et al., 2014, p. 88)。这可以结构化为ICT 人员或SaaS 供应商的以下问题:

- 在网络调查软件中,哪些用户可直接访问特定调查?其权限级别具体为何(如仅查看、完整功能访问或有限访问)?
- 哪些用户拥有对调查问卷、数据文件及软件与服务器对应空间的全局访问权限(如系统工程师、组织高级管理员)?
- 这种访问的时间期限是什么,临时还是永久?访问控制列表多久被审查一次?
- 访问权限的授予流程是什么?由谁负责管理(如添加用户、修改密码、删除账户)?
- 谁有权访问服务器日志文件?日志文件的管理责任归属如何?
- 所用账户为个人账户还是团体账户?若为团体账户,需制定专项管理流程。

当用户层级复杂时,比如在企业反馈系统中(见第1.3.6节),这些问题尤其关键。然而,即使对于最简单的自助式网络调查,基础权限检查也有助于降低风险。

2.4.3.6 与法律和专业标准的关系

关于安全与隐私保护,及时掌握相关立法与专业标准至关重要(另见第6.3节)。在国际研究中,这一点要求更严格,因为各国在个人身份标识、存储、责任、元数据、调查未成年人规则等方面存在差异。欧洲民意与市场研究协会(ESOMAR)的《在线研究指南》(ESOMAR, 2011a, sec.6)进一步梳理了世界各地关于数据隐私和安全的关键立法问题,是重要的参考。审视任何SaaS网络调查软件供应商的安全和隐私政策以及服务条款,可直观了解该领域的复杂性。

学习和解读法律可能是一个漫长且复杂的过程。研究者可寻求当地统计、调查及市场研究协会的支持,这些机构通常为成员提供基础指导;更复杂的场景则可能需要专业法律顾问的介入。

目前有多项行业倡议有助于追踪最新动态,以下为与网络调查实施相关的重要方向。2010年计算机安全研究中的伦理问题研讨会(WECSR)举行了其第一次会议,汇集伦理学家、机构审查委员会成员、研究人员、律师和ICT社区。特别值得关注的是《伦理影响评估文件》(Kenneally, Bailey, & Maughan, 2010),它可以作为伦理、数据安全和数据的初步讨论指南,对任何在线收集的数据进行隐私保护。虽然它的范围比网络调查更广泛,但解决了关键问题,并且正在被广泛采用(Kenneally, Stavrou, McHugh, & Christin, 2011)。另一个重要资源是IEEE计算机学会安全与隐私研讨会²¹,尤其是Web 2.0安全与隐私(W2SP)。

2.4.4 网络调查软件的技术预测试

技术准备工作需要尽早进行,与抽样和问卷调查的准备活动同时进行。研究者面临着两方面的决策。

第一是确定或重新确认网络调查软件的选用,需尽早完成。如果软件存在严重的技术缺陷,应提前发现,核心问题涉及安全性及基本软件功能(如性能、灵活性、运行速度)。一旦研究者选择、接受或重新确认了某些网络调查软件(关于软件选择过程的更多信息详见第5.3节),需进一步决定数据采集与整理的相关设置及准备工作。

无论何种情况,均需规划并执行正式的软件技术预测试,以检验上述技术问题、

预判潜在风险并制定解决方案。其核心活动包括创建包含基础问题类型的测试问卷,模拟预调查前、正式调查及调查后阶段的所有技术环节。测试过程中需生成应答数据,以测试过程和核验数据文件。其中一些测试条目(反应)可能需要人工回答,而大规模数据则需借助自动程序生成,多数软件具备内置自动生成功能。此外,启动一个外部脚本(即机器人程序)来填写原型问卷(*prototype questionnaire*)有助于完全模拟真实测量环节。

对于简单调查,预测试的重点是软件响应速度、应答记录方式及标准化报告输出。问卷编制所需的关键特征也是预测试的重点,特别是问题布局和互动性。

当然,如果已经从过去经验或可靠推荐中充分了解了网络调查软件的技术特性,且调查无特殊要求,可能无需此类技术预测试。

软件的技术评估不能取代对问卷的技术测试(第 2.3.5 节)。尽管二者存在诸多相似之处,但问卷的技术测试在网络调查中实施时间更晚,且使用真实问卷和真实数据进行。同样,预研究也无法替代网络调查软件的技术评估与预测试,后者应在调查流程初期完成。

2.5 无应答策略

由于多种原因,在调查阶段可能无法从纳入调查的单位获取所有预期应答。例如,部分单位可能缺席或拒绝合作,这可能导致我们遗漏目标人群中的重要群体(如事务繁忙者、年龄较大者、经济状况较好的群体等),进而造成估计结果失真。我们称这一现象为无应答(*non response*),而相关的无应答策略(*nonresponse strategy*)涵盖了预防无应答发生所需的各项活动。

某些无应答的补救措施也可以在调查后阶段中使用(见第 4.1 节),但这些措施难以替代正式调查阶段实际受访者的真实应答。本节重点讨论预调查阶段需开展的无应答预防工作及相关准备事项。

研究人员需熟悉无应答的一般原则,以便在特定的网络调查中能够熟练掌握无应答问题,并准备适当的无应答策略。相应地,本节首先对无应答问题(第 2.5.1 节)、无应答指标(第 2.5.2 节)和无应答偏差(第 2.5.3 节)进行一般性介绍。接下来讨论预防无应答的方法:网络调查的设计特点(第 2.5.4 节)、激励措施(第 2.5.5 节)、联系过程(第 2.5.6 节)和邀请函格式(第 2.5.7 节)。最后,总结并概述无应答的策略(第 2.5.8 节)。

2.5.1 网络调查无应答的一般性问题

无应答指是未能从样本中合格单位获得回应。这里首先关注的是单位无应答 (*unit nonresponse*)，即未能从某个单位收集到任何回应（或至少没有可用的回应）。

无应答问题的原因通常分为三类（Groves et al., 2009）：无联系（未能成功接触单位）、明确拒绝（已接触但单位拒绝参与）及其他原因。其他原因占比通常较小且较为零散，涵盖第三人干预这一特殊原因，以及健康问题、语言障碍、行政流程限制及技术故障等。与访员主导的调查不同，网络调查中我们对这些原因的了解往往有限。

我们还关注了项目无应答 (*item nonresponse*)，即合格且配合的单位未回答特定项目（问题或子问题）。中断 (*breakoffs*) 则指符合条件的单位已开始作答，但在完成问卷前提前退出调查。

已经开始作答的合格单位，其完整性状态 (*completeness status*) 也是非常重要的。依据完整性状态，可以区分为完整作答 (*complete responses*) 或完整作答的单位 (*complete*)，指当一个单位符合某些预先指定的完成标准（例如对 80% 的问题作出回应）；以及具有部分完成的单位 (*partial responses*) 或不完整的应答 (*incomplete responses*)，简称为部分应答 (*partials*)，指当一个单位未能达到完整标准，但仍然符合最低限度应答的标准（例如对 50% 的问题作出应答）。剩余单位应答不充足，即没有达到纳入分析的最低标准（例如对问题的应答不到 50%）的单位，称为不可用的应答 (*unusable responses*) 或不可用的单位 (*unusables*)。不同调查的判定标准存在差异，这可能引发问题——因为完整性状态与无应答指标密切相关，而这些指标常需在不同调查间进行比较。

下文会进一步阐述这些状态，此处先总体概述无应答问题、网络调查中无应答的具体过程及影响因素。

2.5.1.1 无应答理论

调查无应答问题自 20 世纪上半叶现代抽样调查诞生以来便随之出现（Singer, 2006）。近几十年来，随着调查参与度 (*survey participation*) 的普遍下降，这一问题愈发凸显（Groves et al., 2009, p. 186）。无应答现象存在于所有调查模式中，但在自我管理式调查中尤为突出，因为这类调查缺乏访员的外部激励与社会压力。受访者拒绝与一台机器互动合作，比拒绝面对面访员的合作要更容易。网络调查的合作率通常低于其他模式（Lozar Manfreda et al., 2008），仅在特定场景（如组织内部

调查)中可能优于邮寄调查(Baruch & Holtom, 2008)。需补充的是,在比较此类研究时往往未纳入成本因素,因此对网络调查模式的结论并不完全公平。因为是将廉价的网络调查与昂贵的替代方案进行比较,网络调查的低成本本可用于优化无应答策略(如增加激励措施),却未在对比中体现。

关于无应答的研究已有丰富的文献(Albaum & Smith, 2012),特别是家庭访谈调查(Groves & Couper, 2008)、邮寄调查(Dillman, 1978; Goyder, 1988)、电话调查(Dillman, 1978)以及各种自我管理和混合模式调查(Dillman et al., 2009)。Fan和Yan(2010)对相关理论进行了系统梳理。网络调查领域应用最多的是社会交换理论(*social exchange theory*)(Ekeh, 1974),它是定制设计方法(Dillman et al., 2009)的核心;以及计划行为理论(*planned behaviour*),这是合理行动理论(*theory of reasoned action*)(Ajzen, 1991)的延伸,已被多位学者应用(Bosnjak, Tuten, & Wittmann, 2005; Fang, Shao, & Lan, 2009; Haunberger, 2011a; Heerwegh & Loosveldt, 2009)。其他各种理论有时也被使用:如经济交换理论(Biner & Kidd, 1994)、互惠规范(Gouldner, 1960)、社会心理学视角(Groves, Cialdini, & Couper, 1992)、交易分析(Comley, 2006)和杠杆优势理论(Groves, Singer, & Corning, 2000),以及认知失调、自我感知和承诺/参与等具体理论(Albaum & Smith, 2012; Heerwegh & Loosveldt, 2009)。

另一类极具针对性的理论视角是调查方法学认知视角(*cognitive aspects of survey methodology*)(CASM),其聚焦应答的心理与沟通过程(Schwarz, 2007),对解释项目无应答尤为有效。该理论形成于20世纪80年代,现已发展为跨学科领域,有时被称为CASM运动。

尽管学界普遍认为理解无应答的理论基础至关重要,但Fan和Yan(2010)在对网络调查无应答的300项实证研究的综述中指出,无应答研究缺乏理论整合,仅关注于某一个或两个无应答因素。Cook、Heath和Thompson(2000)的研究是少数探索因素间关系的案例。此外,关于网络调查中无应答的概念讨论通常是间接的,较多隐藏在网络调查的概述中(Couper, 2000b),而解决特定无应答维度的实证研究则占主导地位。在专门针对网络调查无应答过程的概念化研究中,Vehovar、Batagelj、Lozar Manfreda和Zaletel(2002)以及Fan和Yan(2010)的研究成果具有代表性。

2.5.1.2 网络调查中的无应答过程

结合Vehovar等人(2002)的研究,本书将在联系阶段(*contacting phase*)和合作阶段(*cooperatoin phase*)阐述网络调查中的无应答过程。以下以电子邮件邀

请的基于名单的调查为例展开说明,但这一过程也可适用于其他类型的网络调查。

联系阶段始于向样本单位发送电子邮件邀请。简单起见,最初假定所有的电子邮件均对应合格单位,且预期电子邮件信息会被网络(即互联网)成功传递。然而,这一过程可能因以下技术问题受阻:(a)从电子邮件地址的拼写错误、无效地址,到发件方邮件服务器、网络基础设施或邮件系统的故障等。通常这类邮件会被退回并标记为未送达,但有时可能不会收到任何通知。

邮件被网络接收后,需经过处理并送达目标单位的收件箱。在此过程中,可能遭遇另一类技术问题,阻止电子邮件的传递:(b)电子邮件的通信路径或收件人的电子邮件服务器出现问题。即使电子邮件成功抵达收件人的电子邮件服务器,(c)组织或个人的邮箱系统设置问题也可能导致无法正常接收,具体包括垃圾邮件过滤、自动转发、被误归至遗忘或错误的子文件夹、关联至其他邮箱账户,乃至收件设备的各类技术故障(如硬件问题、邮件软件故障、邮箱存储空间不足等)。

只有当邮件成功发送后,才能开始分析目标单位(潜在受访者)的行为。首先可能出现的问题是(d)调查期间目标单位可能处于缺席状态(如因病、休假等)。有时系统会自动发送邮件通知研究人员,但多数情况下,研究人员无法获得相关说明。此外,即便单位未缺席且正常可用,也可能因(e)其他原因没有查看电子邮件(如检查电子邮件的特定习惯,多个电子邮件帐户未同步),或者虽然查看了电子邮件但由于某种原因从未注意到邀请。有时(f),目标单位虽短暂注意到邮件通知,却可能因匆忙而忽略或遗漏了邀请信息。

当受访者明确注意到电子邮件邀请时,其处理该邀请的认知过程通常分为两个步骤。受访者首先需处理邮件列表中显示的简要提示信息。这通常对应收件箱列表中一两行的内容,包含部分邮件相关信息;具体信息内容取决于邮箱系统的个人设置,至少会显示发件人信息(姓名和/或邮箱地址)、发送时间及主题,有时还会显示邮件开头部分甚至完整内容。此外,鼠标悬停时可能会显示更多邮件信息。为简化说明,此处不展开这些细节,仅聚焦当前主流邮箱软件的通用场景。在这些软件中,用户首先接触的是收件箱信息列表中的单行信息。而在这此过程,电子邮件邀请很容易被误认为是垃圾邮件、营销广告或因其他原因(如没有吸引力的主题)而被视为不相关。因此,这类电子邮件邀请不会被打开,而是(g)直接被忽略、删除、放弃或以其他方式搁置。

受访者可能认为该行提示信息是有趣的和相关的,甚至会标记保存以便后续查看,但最终却未实际处理,导致该邮件(h)被无意中搁置。

只有当受访者点击邮件列表中的该行信息,并打开电子邮件的正文(文本)时,

才有可能阅读内容。但邮件也可能被打开后未被真正阅读,从而(i)被草率放弃,未得到充分考量。最后,受访者可能完整阅读并理解了邀请内容,但认为邮件内容与自己无关,(j)拒绝该邮件,即忽略、删除,甚至通过选择退出选项或发送拒绝的回复等方式明确拒绝。此外,受访者可能阅读了电子邮件邀请并决定参与,却推迟进入网络问卷填写流程,最终(k)被遗忘。

合作阶段始于受访者点击网络问卷 URL 地址。网络调查软件会记录这一行为,标志着该受访者正式进入应答流程。若未发生点击,研究人员通常难以准确判断前述联系过程的具体阶段与状态。诚然,某些类型的软件可以记录电子邮件的成功发送或被打开的状态,但这依赖于收件人的邮件系统支持,其反馈可能受到限制,因此实际作用有限。

受访者点击网络问卷后,无应答问题会以中断形式呈现。此处进一步区分介绍性中断(*introduction breakoffs*)和问卷中断(*questionnaire breakoffs*):(1)介绍性中断发生在网络问卷的介绍页面或首页,这类情况通常被归为单位无应答。如果中断发生在问卷后续页面,就会称为(m)问卷中断(*questionnaire breakoffs*),是否被归类为单位无应答取决于具体调查中使用的标准。

上文列举了 13 个可能发生无应答的节点(a-m),但还可以根据技术子类型(a, b, c)、无联系的子结构(d, e)、延迟子循环(h, k)和问卷中断的具体位置(l, m)进一步分类。鉴于无应答场景的多样性,网络调查应答率偏低具有一定必然性。不过,并非所有节点的风险程度均等:若调查具备合法性、主题突出、目标人群定期查看邮件、邀请设计专业、邮箱地址准确且技术架构完善,那么电子邮件邀请相关的无应答(g-j)多可归类为隐性拒绝或中断(l-m)。同样,在联系阶段结构较简单的情景(如通过邮件、短信或电话邀请的基于名单的网络调查,以及非基于名单的网络调查),上述无应答过程的逻辑依然适用。

2.5.1.3 影响网络调查无应答的因素

正目前针对网络调查无应答影响因素的系统性研究较少,本节在 Vehovar 等人(2002)的研究基础上,探讨社会技术环境、信息通信基础设施、受访者特征和调查设计有关的主题。不过,下文只对这些因素作简单概述。

社会技术环境是影响网络调查参与的重要因素。一方面,这种影响具有间接性,通过一般文化、经济发展、教育系统、社会机构的合法性、社会融合的程度、城市化、基础设施等产生影响。另一方面,一般的调查氛围、公众对调查的态度、对直接营销的看法、网络调查及其发起方的合法性、隐私立法、数据保护丑闻、进行的调查

数量等，都对参与度有更直接的影响。具体到网络调查，可以梳理为以下几点：

- **过度调查**，即网络调查行业的扩张，特别是在线样本库和自助式调查研究，导致受访者疲劳。虽然在线调查小组的应答者在很大程度上被归为各种类型的“专业受访者”，但他们实际上只占（互联网）总人口的一小部分（Fulgoni, 2005）。自助式调查研究在免费或廉价的 SaaS 平台上可以招募所有阶层的公民。最大的网络调查软件服务每天提供数百万份网络问卷。例如，在斯洛文尼亚这个拥有 200 万居民的国家，仅最大的网络调查软件服务²² 每年就提供超过 100 万份调查问卷（截至 2014 年），其中大部分是非商业和自助式调查，并且不属于专业调查行业。除了来自公共、学术、私人、自助式调查以及机构的正规网络调查邀请，公民也被各种娱乐性投票、测验、简短的营销调查和简短的个人询问所轰炸，特别是在综合性、休闲娱乐类网站和社交网站。此外，各种类型的调查都在与众多其他渠道竞争在线注意力，从直接营销到慈善、游戏、赌博及各种复杂的广告形式。遗憾的是，网民很难明确区分真正的研究要求和各种营销噪音。关于过度调查问题的进一步讨论和证据，详见 Weiner 和 Dalessio（2006）的组织研究以及 Adams 和 Umbach（2012）的高等教育研究。如果受访者易受过度调查影响，可以预期公众对调查邀请的倦怠会增加。例如，Peytchev（2011）的研究表明，在短时间内被邀请参加多项网络调查的学生更有可能不回答问题。
- **垃圾邮件和病毒**对终端用户来说似乎正在减少，但研究报告仍然指出，在 MS Outlook 电子邮件系统中，四分之三的电子邮件被过滤为垃圾邮件（Microsoft Corporation, 2013）。这对带有电子邮件邀请的网络调查来说尤其有害。因为它可能会被垃圾邮件拦截工具自动拦截，或被受访者主动认为是垃圾邮件。很难估计有多少网络调查的电子邮件邀请因垃圾邮件拦截工具而未被送达，这也导致应答率的计算出现问题（Dobrow, Orchard, Golden, Holowaty, Paszat, Brown, & Sullivan, 2008）。已有各种措施来对抗垃圾邮件的问题，从适当的信息格式到使用垃圾邮件分析工具（见 2.5.7 节）。受访者可能不想点击未知发件人发送的 URL 地址，因为担心这会进入“网络钓鱼”网站，即类似于合法网站的假网站，旨在获取用户名、密码、信用卡信息，还会植入恶意软件或病毒。
- **互联网隐私**与物理、信息、心理、法律、经济和互动等隐私方面密切相关（Cho & Larose, 1999），这可能导致互联网用户不愿意在网络上透露个人信息。因此，在考虑参与网络调查意愿（Couper, Singer, Conrad, & Groves, 2008）和自我暴露（Joinson, Paine, Buchanan, & Reips, 2008）时，受访者对隐私的一般态度和

对风险、伤害和利益的主观认知变得越来越重要。多项研究发现,隐私担忧会对网络调查中敏感信息的披露水平有负面影响(Heerwegh, Vanhove, Matthijs, & Loosveldt, 2005; Joinson et al., 2007)。此外,这些问题可能对目标人群的不同群体产生不同影响。敏感程度高的群体通常是受教育程度较高、年龄较大(Sheehan, 2002)、调查参与经验较少和来自隐私保护较严格的国家的用户(Bellman, Johnson, Kobrin, & Lohse, 2004)。

信息通信基础设施对网络调查的成功实施至关重要。正如在无覆盖率部分(第2.2.5节)所阐述的,不同国家的互联网普及率存在巨大差异。此外,互联网基础设施的质量(互联网连接速度)、互联网成本以及所使用的系统(电子邮件客户端、浏览器、设备)的差异也会影响应答情况,复杂问卷或高级交互功能对技术要求更高(例如利用位置数据的移动设备)。其他与ICT相关的方面也很重要,从免费WiFi点的性质和密度到平板电脑和智能手机的普及程度。

受访者的特征对无应答也有很大影响。常见影响因素为社会人口统计学信息(年龄、性别、教育或城市),但也包括调查经验、对调查主题的兴趣、个性特征和其他个人原因。对于网络调查,计算机知识、对计算机使用的态度和互联网使用模式也很重要。即使控制了社会人口统计学特征,计算机和互联网的使用强度仍被认为是网络调查合作行为的最重要预测因素(Batagelj & Vehovar, 1998; Kwak & Radler, 2002; Payne & Barnfather, 2012)。当计算机使用没有被控制时,它通过ICT用户特征来体现,这些特征为年轻、受过教育、男性、有较高社会地位(Batagelj & Vehovar, 1998; Diment & Garrett-Jones, 2007; Payne & Barnfather, 2012),但也更可能是学生(Shih & Fan, 2008),或在ICT或相关部门工作(Diment & Garrett-Jones, 2007)。有时种族因素也会产生影响(Couper, Kapteyn, Schonlau, & Winter, 2005)。一般来说,技术先进的用户更愿意或能够参与,例如,在一项使用Java的网络调查中,那些拥有技术知识的人较少出现中断现象(Stieger, Reips, & Voracek, 2007)。当然,目标人群的无应答特征也表现出群体性:一般公众通常无应答率较高,而各类管理人员或名人无应答率最高。不同理论所定义的因素中,还涉及其他具体预测因素:时间资源、对数据误用的恐惧、主题的突出性、对调查负担的感知(Haunberger, 2011b);道德义务、过往调查参与频率(Haunberger, 2011a);对参与网络调查的态度、内化的社会压力、感知的行为控制、参与调查的道德义务程度(Bosnjak et al., 2005);个人网络邀请和对赞助人的信任(Fang et al., 2009)。此外,某一调查的无应答者更可能拒绝其他调查(Peytchev, 2011)。已得出普遍性结论,社会行动和社

会参与水平较高与调查参与相关（Goyder, 1988）同样适用于网络调查。

网络调查设计是研究者唯一可以直接控制的因素。它广泛涉及调查发起方、主题突出性、问卷设计、激励措施、联系方式和邀请格式。本书将在第 2.5.4 至 2.5.8 节中加以详述。

2.5.2 定义无应答、中断和完整性

受访者（单位）无应答的情况与中断密切相关，也与其完整性状态有关。明确相关指标不仅对后续调查阶段及调查后阶段中不同调查间的对比分析至关重要，还能有效防止人为操纵应答率的潜在风险。因此需要明确确定相应指标的定义。

此处的讨论在概念上遵循了抽样部分（第 2.2 节），首先以基于概率列表的电子邮件邀请样本为研究对象展开阐述，随后将相关原则推广应用于其他类型的样本分析中。

2.5.2.1 单位无应答指标

结合前文对无应答过程的分析，单位无应答指标主要用于反映联系阶段的无应答情况，而中断、项目无应答及完整性状态相关指标则聚焦合作阶段的无应答问题。本节先围绕单位无应答展开讨论，再延伸至其他关联指标，各指标间存在密切逻辑关联。

作为核心参考依据，Smith（2002）与美国民意研究协会（AAPOR, 2011）对无应答指标的标准定义具有权威性，其中《针对特定人群的互联网调查》章节（第 32 页）明确了最终处置代码的规范。其他实用参考来源包括在线样本库相关研究（Callegaro & DiSogra, 2008; ISO, 2009）。

学界普遍认为，应答率（*response rate*, RR）是衡量单位无应答的关键指标，定义为应答单位与所有合格单位的比率。

$$RR = \frac{\text{应答单位}}{\text{所有合格单位}}$$

该公式的实际应用依赖于对单位状态的精准界定。为简化讨论并贴近实践，可将单位划分为四类核心状态：单位无应答与不合格单位（*ineligible units*）和应答单位（*responding units*），后者由完成者（*completes*）和部分完成者（*partials*）组成。

补充说明：应答率取值区间为 0~1，但是经常会用百分比形式来表达，对应区

间为 0~100%。本节讨论的其他一些比率也是如此。

在网络调查场景中,单位状态的判定存在诸多细节问题。第一类问题聚焦确定未启动问卷的单位(已被邀请但从未点击过问卷调查的 URL)。对于提供明确拒绝的单位,相对来说比较容易确定(例如,应答者通过电子邮件解释拒绝的原因,据此可以确定资格);对于电子邮件自动回复,有时也易判定(详见第 2.4.1 节)。根据自动回复的反馈信息,再确定合格(例如自动回复表示暂时缺席)或不合格的状态(例如自动回复表示死亡、搬家或永久缺席)。有时,即使收到电子邮件回复信息,仍然无法确定目标单位是否合格(例如,自动回复表示错误的电子邮件地址)。在该种状态,研究者面临的情况与其他未返回任何信息的情况类似。然而事实上,迄今为止,这往往是带有电子邮件邀请的网络调查中最常见的回复状态。实践中,经常将资格不明的单位简单视为合格单位,仅在确有证据表明其不合格时,才会使用估计值将其调整为不合格单位;因为根据定义,当一个合格单位没有回应时,则表明单位未应答。

第二类问题涉及已点击问卷 URL 的合格单位。根据定义,介绍性中断属于单位无应答。更复杂的是,需要将单位无应答与问卷中断、部分应答进行区分。因为上述涉及定义完整性状态的具体标准,而这些标准在不同调查中差异较大。下文将讨论这个问题,然后进一步阐述中断、项目无应答和完整性状态。

不过,在此我们可以引入一种简单的技术状态替代方案,即把调查问卷的所有分界点作为基准。据此可采用两种近似处理方式,将“无任何反馈”与“介绍性中断”一起作为单位无应答处理,而将问卷中断的单位视为部分完成。这种简化处理对于多数实际场景具有实用性,能够实现对应答率(RR)的简易估算。也就是说,通过确定前述 4 种状态,便能大致估算出符合美国民意研究协会标准的 RR2(应答率 2)。然而,由于美国民意研究协会关于基于概率列表的网络调查的正式标准(AAPOR, 2011)包含 22 种最终状态,而不是仅 4 种,因此将近似值表示为 RR2'。美国民意研究协会的 RR2 用于官方文件和报告,但 RR2' 的近似值在绝大多数情况下均适用,同样可用于概念讨论。两个具体示例见下文。

首先,在对斯洛文尼亚科学家的调查中(来自 1.3.5 节的 WCSC),发出了 4 551 封电子邮件,其中 3 185 封无任何反馈,149 封为介绍性中断和 113 封没有回答问题,由此合计有 3 447 个单位无应答。另一方面,通过默认技术方法识别出 103 封部分回答,以及 1 001 个完整应答,因此有 1 104 个应答单位。据此可计算 $RR2' = 24.3\%$ ($1\ 104/4\ 551$)。与按 AAPOR 指南全面处理所有处置代码的结果相比,差异甚微。

同样 Petrovčič、Lozar Manfreda 和 Petrič (2013) 在一项在线健康论坛注册会

员的调查中,采用 AAPOR 处置代码,使用调查的具体标准(至少有 50% 的项目被回答)来定义部分完成,得到 RR2 为 9.7%,与近似值 RR2' (9.9%) 只有细微差别。

对这两个例子的进一步阐述见第 2 章的补充内容,即“应答率计算实例”(Examples of response rate calculations) (<http://websm.org/ch2>)。

为了更全面地了解基于名单的网络调查的无应答特征,本书引用了 2011 年至 2012 年间 37 项基于名单的电子邮件邀请调查的研究结果(Lozar Manfreda, Berzelak, & Vehovar, 2012)。该研究显示,在所有被邀请者中,完全应答者(定义为回答 90% 以上项目)占 27%,部分应答者占 18%,无应答单位占 41%,介绍性中断占 14%。以合格单位为基数,无应答率为 55%,对应 RR2' 为 45%。这一结果相对偏高,主要归因于这些调查的特殊性——问卷简短且主题突出性强。

其他元分析研究(meta-content)也证实了网络调查应答率的大致范围(Anseel, Lievens, Schollaert, & Choragwicka, 2010; Baruch & Holtom, 2008; Mitra, Jain-Shukla, Robbins, Champion, & Durant, 2008; van Horn, Green, & Martinussen, 2009)。对于采用电子邮件邀请的基于名单的网络调查,若方法规范、发起方权威且主题突出,其应答率(RR2)通常在 30%~40%;但只有在方法健全、发起方良好和主题突出的调查中才会如此。若缺少上述任一条件,应答率可能降至个位数;仅在特殊场景(如极高的主题相关性、丰厚激励或密集提醒)下,应答率才可能达到或超过 60%(Murphy, Hamel, Harrison, & Hammer, 2011)。

前文假设所有状态不明的单位(如“无任何反馈”)均为合格单位,即计入单位无应答,这与 AAPOR 计算 RR2 时的设定一致。如果将他们都归类为不合格单位,并因此被排除在 RR 计算之外, AAPOR 将相应的计算结果称为 RR6。若他们被分成(基于一些特定的估计)无应答和不合格,则对应 RR4。如果从分子中排除部分应答单位,只保留完整应答单位,则分别得到进一步的变式(RR1、RR5 和 RR3)。

应答率及其互补指标(1-RR, 无应答率, nonresponse rate)是数据质量非常重要的指标;且在不同调查间具有可比性。此外,还有许多其他可比较且更具体的无应答指标。在传统调查中,合作率(cooperation rates)和拒绝率(refusal rates)非常重要,因为它们传达了无应答的原因(无联系、拒绝),但在网络调查中适用性有限,因为对于“无任何反馈”的单位,研究人员难以区分是未联系还是隐性拒绝。

联系率(contact rate)是电子邮件邀请中至少有应答的单位(如点击问卷、明确拒绝)占合格单位的比例。在这种情况下,公式中的分子包括完整应答或部分应答的受访者、中断者和明确拒绝者等。类似的还有点击率(click-through rate)(CTR),它表示在成功发送的电子邮件邀请中,执行目标操作(即访问网络问卷)的单位所

占比例。

若分母中的单位既包含合格单位，也包括其他所有邀请的单位，所得指标即为参与率，即应答单位占有所有邀请单位的比例（AAPOR, 2011）。当所有单位都符合条件时，参与率就等于RR。

为了评估电子邮件列表的质量，Vehovar 等人（2002）引入了吸收率（*absorption rate*），即在所有发送的电子邮件邀请中没有送达失败的单位比例。互补指标是失败率（*failure rate*），即未成功投递的邀请函份额。

上述指标主要适用于基于概率名单的横断面网络调查（采用电子邮件邀请），针对其他类型的网络调查，需做如下调整。

- **基于概率名单的邮件邀请**样本遵循类似原则，其联系状态可参照 AAPOR 邮件调查规则确定；电话招募场景亦适用此原则。
- **重复概率调查**，即特殊的纵向调查或概率在线样本库，同样适用上述原则，但需计算总体应答率。对于基于概率的在线样本库，总体应答率为招募率（*recruitment rate*）（指符合条件的邀请中的初始应答者）、简介率（*profile rate*）（初始应答者中完成介绍页面调查的应答者）以及特定调查的应答率的乘积（AAPOR, 2011, p. 12）。这里还需要考虑样本流失（*attrition*）（即从小组退出或不再合作的成员）。因此，在线调查小组的累积应答率（*cumulative response rates*）很少超过 20%，往往低于 10%（DiSogra & Callegaro, 待刊）；基于概率的非商业小组最高，接近 40%，如 FFRISP（Krosnick, Ackermann, Malka, Sakshaug, Tourangeau, de Bell & Turakhia, 2009）或 LISS（Scherpenzeel, 2011）。
- **基于非概率名单的网络调查**因样本不具目标人群代表性，所以在计算应答率时，合格单位做分母没有意义。因此，建议使用参与率来代替。这对于非概率性网络调查小组尤其重要，因为这些小组的参与率通常只有百分之几（Tourangeau, Conrad, & Couper, 2013, p. 41）。
- **非概率非名单网络调查**，如无限制的样本，因缺乏合适分母，难以计算传统应答率指标。对于横幅广告或带有超链接的网络调查，可采用点击率（CTR）——即点击量与总页面浏览量的比值，这一指标源于在线广告领域。实践中，这些比率通常仅为零点零几个百分点（如 0.03%）。
- **基于概率的非名单式样本**，如拦截式网络调查，可以将接触到邀请弹窗的网站访问者的数量作为应答率计算的分母。

2.5.2.2 中断

中断指合格单位已启动网络问卷调查,但在完成最后一个问题前退出的情况。本书统一使用“中断”这一术语,避免使用辍学(*dropout*)、放弃(*abandonments*)或流失等易产生歧义的表述。在传统的调查模式中,有时会使用中断来表示剔除部分完成和完整完成后的剩余合格单位(AAPOR, 2011),与本文定义不同。

本节还区分了介绍性中断(*introduction breakoffs*),这源于初始参与意愿不足,根源在于调查特征(如赞助商、突出性)、情境因素(如时间压力)和邀请细节(如频率、时间、介绍的内容和其他形式)等一系列广泛的原因。导致问卷中断(*questionnaire breakoffs*)的原因与上述不同,如受访者的情况(如疲劳、分心等)和问卷的特点(如设计、长度、内容、问题类型等)。相应地,可以计算出三种中断率(IBR、QBR、TBR)。

观察所有应答单位(启动问卷调查的单位)的问卷中断情况也很有趣,可以将其标记为毛问卷中断(*gross questionnaire breakoffs*),相应的比率表示为QBR',
 $QBR' = QBR \times (1 - IBR)$ 。

$$IBR(\text{介绍性中断率}) = \frac{\text{介绍性中断单位}}{\text{启动问卷调查的单位}}$$

$$QBR(\text{问卷中断率}) = \frac{\text{问卷中断单位}}{\text{启动问卷调查的单位} - \text{介绍性问卷中断单位}}$$

$$TBR(\text{总中断率}) = \frac{\text{介绍性中断} + \text{问卷中断}}{\text{启动问卷调查的单位}}$$

与中断率互补的是保留率(*retention rates*)(Göriz & Crutzen, 2012)。例如,总保留率(*total retention rates*)等于 $(1 - TBR)$,具有一个很好的特性: $(1 - TBR) = (1 - IBR) \times (1 - QBR)$ 。

多项元分析研究揭示了中断率的实际分布特征。在对5 752份网络调查(大部分由学生和自助式填写完成)的元分析研究中,Vehovar和Čehovin(2014)得出以下平均中断率:介绍性中断率为33%、问卷中断率为11%、总中断率为40%;对应的保留率为:通过介绍页的单位占67%、通过介绍页后完成问卷的单位占89%、所有启动者中完成问卷的单位占60%。当然,不同调查的中断率差异显著,部分调查总中断率仅为几个百分点,也有部分调查的保留率低至几个百分点。Vehovar和Čehovin(2014)的研究还发现,当前总中断率较十年前的元分析研究结果(16%~34%)有所上升(Lozar Manfreda & Vehovar, 2002; Musch & Reips, 2000)。

上述案例很好地说明了网络调查的现实情况，特别是在自助式调研研究领域。学术和政府调查的中断率比较低，通常在个位数范围内。例如，Ainsaar 等人（2013）在爱沙尼亚的一项 ESS 网络实验中，报告总中断率为 7%；基于概率的在线样本库 LISS 的总中断率为 10%~15%（Scherpenzeel, 2011）。营销研究调查的总中断率范围较广，常与自助式调查相当，甚至更高。许多非概率在线调查小组也是如此，其范围非常广泛。Vehovar 和 Čehovin（2014）的元分析研究也表明，大多数（83%）的中断为介绍性中断（即介绍性中断率 > 总问卷中断率）。相反的情况（总问卷中断率 > 介绍性中断率）多出现于高合法性、邀请设计良好的调查中，此类情况可能因问卷存在设计缺陷（如不友好、难度大或过长）而加剧。例如，Petrovčič 等人（2013）的研究中，介绍性中断仅占有所有中断的 24%。如第 2.3.4 节所述，问卷长度及设计、复杂性、主题等特征是影响中断的重要因素（Peytchev, 2009）。

除了上述调查层面的指标，还可以在项目（变量）层面计算中断率，即项目层面的中断（*item level breakoff*）（ILB），表示某个项目（变量）因中断而缺失的数值（单位）的份额。如果有一半的受访者在调查结束前退出问卷，那么就会出现某些项目的中断率为 0.5。同样，单位层面的中断率（*unit-level breakoff rate*）（UBR）表示单个受访者因中断而导致的缺失项占比。例如，仅完成 10% 问卷即退出的受访者，其 UBR 为 0.9。有关这些指标的更多变体与细节，可参见第 2 章补充内容“应答率计算实例”（Examples of response rate calculations）（<http://websm.org/ch2>）。

2.5.2.3 问题层面的项目无应答

当受访者接触到某一项目（问题或子问题）却未作答时，便产生项目无应答。衡量这一现象的指标为项目无应答率（*item nonresponse rate*）（INR），适用于特定调查问题或子问题，其计算公式为：

$$\text{INR} = \frac{\text{项目无应答单位数量}}{\text{接触该项目的所有合格单位}}$$

分母仅包含接触到该项目的合格单位，即完成者和部分完成者中仍在作答该项目的单位。由于分支、随机化、提前中断和其他可能的原因而没有接触到某一项目的单位应被排除。分子需要排除非实质性回答（“不知道”“其他”等），此类为合法应答，但需纳入“拒绝回答”或“不愿意回答”等明确的无应答选项。

在线样本库及市场研究调查中，除开放式文本条目外，多数问题会采用硬性提示（强制应答），因此项目无应答较少出现。但正如第 2.3.3 节所阐述的那样，亚

洲的社会科学家对硬性提示持保留态度。

项目无应答的重要原因包括问题敏感性,例如收入、性行为、社会污名行为、健康等方面。另外,此类问题可能引发受访者的应答偏差,如过度选择“不知道”、中间点和更符合社会赞许的答案。可使用现有技术处理敏感问题,从项目计数、随机回答和问题措辞技术到其他策略(Tourangeau & Yan, 2007),都可用来防止项目无应答。然而,研究表明网络调查的自我管理模式及计算机化特性,在一定程度上可减少敏感问题的项目无应答(见第1.3.5节)。

项目无应答也可能是缺乏参与度的间接表现(见第2.3.4节),特别是可能出现在复杂(如表格)、冗长、无聊、惊恐和/或设计不良的问卷中,因此提高参与度的策略会有帮助。

问题类型也会影响项目无应答的程度。开放式文本问题尤其易出现项目无应答现象。Vehovar 和 Čehovin (2014) 的研究显示,选答的开放式问题(如“您还有其他意见吗?”),其 INR 常超过 90%;而必填开放式问题(例如“描述您的工作职位”),其 INR 可能会低很多,甚至低于 10%。封闭式问题中附加的“其他,请说明”等开放式选项的 INR 通常高于 95%;而带数字条目的问题及预设应答类别的封闭式问题,其 INR 通常仅为几个百分点。

INR 虽然通常针对单个项目计算,也可扩展至问题组或整份问卷,通过计算范围、平均值、中位数等指标,实现不同调查间的比较(如问卷所有项目的平均 INR)。

2.5.2.4 单位层面的项目无应答

前文分析了所有单位针对某一项目的无应答情况,本节聚焦单位(应答者)层面的项目无应答——即单个单位在其接触的所有项目中,未应答项目所占的比例,称为单位层面项目无应答率(*unit-level item nonresponse*, UIN)。

$$\text{UIN} = \frac{\text{某一单位所有无应答项目}}{\text{该单位接触的所有项目}}$$

根据这个公式,例如,某一所有问题都未“回答”的单位,其单位层面项目无应答率 $\text{UIN} = 1$, 而一个对所有项目都回答的单位,其单位层面项目无应答率 $\text{UIN} = 0$ 。

但这一公式仅适用于完成问卷的受访者。对于中断的受访者,由于未接触后续项目而存在额外缺失数据,需结合单位层面中断率(UBR)综合考量。单个受访者的总缺失项目比例可表示为单位层面缺失数据率(*unit-level missing-data rate*)(UMR):

$$UMR = UMR + (1 - UMR) \times UIN$$

单位层面的完整率 (*unit-level completeness rate*) (UCR)

$$UCR = 1 - UMR$$

在 UBR 为 0.9 的情况下, 即仅接触 10% 的项目, 若受访者接触的项目中回答了一半 (即 $UIN = 0.5$), 则其 $UMR = 0.9 + (1 - 0.9) \times 0.5 = 0.95$, $UCR = 0.05$ 。UCR 是数据质量极其重要的指标, 因为它对完整应答、部分应答和不可用应答单位的定义具有决定性意义。

2.5.2.5 完整应答、部分应答与不可用应答单位

当引入完整性状态 (完整的、部分的、不可用的) 时, 需明确这些状态由调查特定规则和主观标准界定。由于这些规则直接影响关键的单位无应答指标, 本节将重点探讨这一问题。

传统调查模式中, 有时以受访者回答所有项目的 50% 作为基准 (AAPOR, 2011), 用于区分部分应答与不可用应答单位。衡量单位完整性状态的方法多样, 前文提及的单位层面的完整率 (UCR) 便是其一, 本节将基于此展开讨论。但总体而言, 我们更倾向于采用非形式化、更具实质性的标准定义部分应答。例如, 预先指定一组核心调查项目, 仅当这组核心项目未完成时, 单位才被归类为不可用应答; 否则, 视为部分应答。

我们重申不可用应答单位 (*unusable units*) 的核心特征是: 被排除在统计分析、归档及所有后续使用之外, 仅将可用应答单位 (*usable units*) 的补充集合纳入处理。这一概念借鉴了相应的 ISO 标准 (ISO, 2009, 2012)。可以补充说, 可用应答单位基本上与前文所说的应答单位或有应答的单位相匹配, 因此通常包括部分应答和完整应答单位。

不可用应答单位可定义为未达到最低完整性基准的单位 (如 $UCR < 0.50$)。此外, 它们还可能包括质量低下的应答 (如虚假应答、欺骗性应答、过度敷衍应答等)。完整性状态的判定仅适用于已开始作答的合格单位, 因此从定义上讲, 不可用应答单位属于单位无应答范畴, 多对应 AAPOR (2011) 中提到的“信息不足的部分应答或中断应答”类别。这在概念上比较清晰, 足以满足多数实践需求。然而, 不可用应答单位仍是一个复杂类别, 其交织了无应答问题与抽样 (如合格性)、测量 (如应答质量) 等多方面问题, 本节对此不再深入展开。

部分应答与完整应答的区分同样具有主观任意性。此处再次采用完整性水平判定标准（例如 UCR），例如以一定比例的高质量（即可用）应答定义完成度，如 80%（AAPOR, 2011）。然而，也可采用实质性方法，将完整应答单位界定为对预先指定的关键问题提供了高质量应答的单位。

上述所有完整性判定标准均依赖于随调查而异的主观规则，且在结构化问卷中也很难实现。在结构化问卷中，由于分支的关系，不同单位接触到的项目数量不同。为此，本书提出一种规避这些问题的替代方法。这个替代方法非常简单，即使用默认技术状态（*default technical statuses*）来定义完整性。

- a. 受访者仅访问介绍页面，然后退出。
- b. 受访者进入问卷初始页面并看到问题后退出。
- c. 受访者在初始页面后继续浏览，但未作答任何问题即退出（即空单位）。
- d. 受访者浏览至问卷末尾（即完成问卷流程），但未提供任何应答（即空单位）。
- e. 受访者对其所接触到的问题提供了部分或全部答复，但在问卷结束前退出。
- f. 受访者提供部分或全部应答，并结束调查问卷。

任何给定的调查软件都可以识别这些状态。它们也是 Bosnjak 和 Tuten（2001）对网络调查中（非）应答模式进行重要分类的基础，可用于界定不同层级的“潜伏者”和项目无应答。

基于上述状态，可轻松划分出（e）部分应答与（f）完整应答，其余类别（a-d）则归为不可用应答。当然，这种方法可能将仅应答一个项目的单位也纳入部分应答或完整应答，这是其明显缺陷。但优点在于易于计算，且在不同调查间具有可比性。采用这种技术方法，Vehovar 和 Čehovin（2014）在他们的元分析研究中观察到以下结构：

- 部分应答（e）：至少提供部分应答的问卷中断单位，占 7%。
- 完整应答（f）：提供部分应答且完成问卷的单位，占 57%。
- 不可用应答单位（（a）+（b）+（c）+（d））：无法进入分析的单位（36%），主要由介绍性中断构成（（a）+（b）= 33%），空单位占少数（（c）+（d）= 3%）。

在该技术方法及一般情况下，介绍性中断单位和空单位均属于不可用应答，而问卷中断单位的应答状态则可能是完整应答、部分应答或不可用应答（表 2.2）。因此，上述技术方法（在表 2.2 中用星号（*）标示）可能与其他采用 UCR 阈值的策略存在显著差异。Petrovčič 等人（2013）最初使用 $UCR = 0.5$ 的标准来定义部分应答，

表 2.2 基于单位应答项目数量与问卷完成程度的可能完成状态（默认技术状态以星号 * 标示）

已应答的项目数	问卷完成水平		
	介绍性中断	问卷中断	完成问卷
没有项目（UCR = 0）	不可用	不可用	不可用
一些项目（0 < UCR < 1）	—	不可用、部分应答 *、完整应答	不可用、部分应答、完整应答 *
所有项目（UCR = 1）	—	—	完整应答

以 $UCR = 0.9$ 来定义完整应答。然而,如果使用上述技术方法,基准的数量会增加六倍,不可用的单位数量会增加十分之一。因此,RR2 的估计值会从 9.7% 增加到 11.0%。

表 2.2 和上述案例说明,在网络调查中完整性状态的界定具有复杂性和高度敏感性。这一点极为重要,因其与无应答指标的计算密切相关。更多计算示例与说明可参见第 2 章补充内容“应答率计算实例”(Examples of response rate calculations) ([http:// websm.org/ch2](http://websm.org/ch2))。

2.5.3 无应答率和无应答偏差之间的关系

在追求高应答率时,研究者容易忽视无应答率与无应答偏差之间的关联,而这恰恰是制定无应答策略的核心考量。鉴于其重要性且在文献中经常被忽视,本书将对这一关键问题进行系统介绍。这一问题对网络调查尤为关键,因其往往受无应答影响更为显著。

本节讨论基于概率样本展开。对于非概率样本,可以采用类似的近似方法,正如第 2.2 节抽样部分所阐述的内容。

无应答偏差(nonresponse bias)是指调查中基于受访者样本得出的预期估计值与真实值之间的差异, $Bias(\bar{y}) = (\bar{Y}_r - \bar{Y})$, 其中 $\bar{Y}_r$ 是仅基于调查中受访者的估计值, $\bar{Y}$ 是目标人群中的“真实”(基准)值(Groves et al., 2009, p. 189)。需要注意的是,此处的无应答偏差与第 2.2.5 节讨论的无覆盖偏差不同。但由于分不同章节介绍这两种偏差,因此并没有采用不同统计符号来区分,如无应答偏差 $Bias_{NR}(\bar{y})$ 、无覆盖偏差 $Bias_{NC}(\bar{y})$ 。

无应答偏差可能与部分研究目标无关,但在进行重要的人口估计(如市场份额、失业率、疾病预后)或混合模式调查时,其影响可能至关重要。

需要注意的是,任何程度的无应答都会引起无应答偏差的存在,因为无应答单位与应答单位之间存在差异。例如,假设目标人群中拥有智能手机的单位占比 $\bar{Y}$

= 54% (基准值), 且智能手机用户也更有可能会参与关于 ICT 使用相关的网络调查。受智能手机用户的参与度较高影响, 他们在调查中的比例会被高估为 $\bar{Y}_r = 90\%$, 这是不正确的, 此时无应答偏差为 $\text{Bias}(\bar{y}) = (90\% - 54\%) = 36\%$ 。

简而言之, 若将样本划分为“始终应答者”与“始终无应答者”, 可直观证明: 无应答偏差与无应答者比例 (W_{nr}) 成正比, 且与应答者 $\bar{Y}_r$ 与无应答者 $\bar{Y}_{nr}$ 之间的差异成正比。当只依靠受访者 $\bar{Y}_r$ 估计总体时, 无应答偏差可以表示为:

$$\text{Bias}(\bar{y}) = (\bar{Y}_r - \bar{Y}) = W_{nr} \times (\bar{Y}_r - \bar{Y}_{nr})$$

如果无应答者比例 (W_{nr}) 极小, 或者应答者和无应答者之间的差异 ($\bar{Y}_r - \bar{Y}_{nr}$) 可以忽略, 则偏差无需过度关注。若 $W_{nr} = 0.60$, $\bar{Y}_r = 0.90$, 假设 $\bar{Y}_{nr} = 0.30$, 我们可以得到 $\text{Bias}(\bar{y}) = 0.6 \times (0.9 - 0.3) = 0.36$ 。

然而, 更为合理的假设是: 单位会以特定倾向成为无应答者, 因此无应答偏差本质上取决于无应答倾向与目标变量值的相关性。此外, 两者的方差及应答率的倒数也会放大偏差, 具体处理方法可参考 Bethlehem (2009a)。

相对偏差 (*relative response*) 是衡量无应答偏差的常用指标, 计算公式为 $\text{Rbias}(\bar{y}) = \text{Bias}(\bar{y})/\bar{Y}$ 。在上述案例中相对偏差 $\text{Rbias}(\bar{y}) = 0.36/0.54 \approx 0.66$, 数值是非常高的, 因为该偏差占目标人群值的 66%, 远超 5% 和 10% (相对偏差可接受的基准值, 第 2.2.5 节已经提到)。

还可以通过标准偏差 (*standardized bias*) 来进一步了解, 其公式为: $\text{Sbias}(\bar{y}) = \text{Bias}(\bar{y})/\text{SE}(\bar{y})$ 。第 2.2.5 节在讨论无覆盖偏差时引入了这一概念。在上述情况下, 如果 $n = 400$ 、 $\text{SE}(\bar{y}) = 2\%$, 则可以得到 $\text{Sbias}(\bar{y}) = 0.36/0.02 = 18$, 该值远高于 t 值 1.96, 所以无应答偏差具有统计学意义上显著性。

在现实中, 无无应答偏差通常未知, 因为总体真实值 $\bar{Y}$ 未知。总体真实值 $\bar{Y}$ 正是我们进行调查的目标。尽管如此, 我们还是可以从抽样框 (如性别)、外部辅助数据 (如选举结果), 或从调查的其他波段 (如初步筛选、调查后的追踪、纵向研究中的前一波段), 或从连续的调查努力 (例如拒绝转换、增加接触) 的比较中了解某些 $\bar{Y}$ 的情况。所有这些都可以提供分析无应答偏差模式的依据 (Groves & Peytcheva, 2008)。

若研究忽略无应答, 实则隐含“无偏差”假设。这通常意味着提高应答率会导致资源浪费, 除非提高应答率只是为了增加样本量。另一种极端是, 无论是否具备依据, 均认为提高应答率会减少无应答偏差。当实际情况并非如此时, 这也可能造成不必要的资源投入。

虽然很重要，但关于应答率与无应答偏差关系的研究却相对有限，多数仅观察到偏差与联系次数存在关联（Vehovar, 1999）。Fuchs、Bossert 和 Stukowski（2013）的研究综述表明，这类研究常因混杂因素过多，难以明确二者关系。因此，现有研究大多发现无应答偏差和无应答率之间几乎没有甚至完全没有相关性（Groves & Peytcheva, 2008）。这可能是因为，当预期无应答偏差较高时，研究者会刻意追求高应答率，从而会造成二者的虚假正相关。而在特定调查中，增加无应答干预（例如增加联系次数）通常会降低偏差，从而出现负相关，因为根据定义，高应答率会逐渐将无应答偏差降低到零。

图 2.2 展示了特定调查中某项目的无应答偏差与应答率之间的潜在关系。

- 线 A 代表了无应答率和无应答偏差之间的强相关，无应答偏差随应答率升高而显著降低。
- 线 B 和线 C 代表了二者的弱相关和 / 或仅存在中等偏差，应答率提升对偏差影响有限。
- 线 D 代表了一种特殊情况，应答率升高反而可能增加偏差。

所有线都以 0% 的无应答偏差和 100% 的应答率结束。这是理论层面，现实中可能永远难以实现。

Fuchs 等人（2013）在欧洲社会调查（ESS）对社会人口学变量的研究中用与图 2.2 非常相似的方式阐述了这一关系。在此研究中，他们运用联系信息并行数据，模拟了基于联系次数的无应答率及条件无应答偏差。他们的研究观察到线 B 和线 C

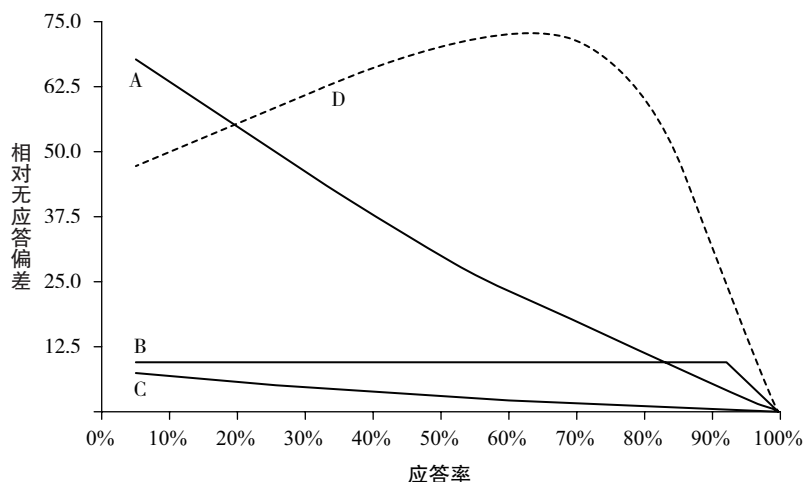


图 2.2 相对无应答偏差与应答率之间的关系

最为常见。我们还可以进一步推测,对于实质性目标变量(如对政府的态度),应答率对无应答偏差的影响会更小,尤其是如果实施了加权校正(这可能进一步减少偏差)。

当然,每个调查项目可能表现不同,需要在图 2.2 中单独绘制一条线。每条线均对应固定且特定调查情境中的某一项目。在该情境中,我们假设仅通过强化无应答干预措施(如连续增加联系轮次,增加激励措施,增加拒绝转换等)来调整。如果改变其他调查情境因素(如初始激励措施或初始数据收集模式),则会形成不同的调查情境及对应的线条。原则上,这些线条也可以在相应的实验设计中获得,如 Vehovar、Berzelak 和 Lozar Manfreda (2010)。

图 2.2 很好地说明了是否应投入资源以获取高应答率的困境。

- 在无应答率和无应答偏差之间强相关(线 A)的情况下,低应答率是不可接受的。上述智能手机用户比例估计的案例便是如此,其中 40% 的应答率产生的相关偏差为 $R_{\text{bias}}(\bar{y}) = 0.66$ 。通过增加联系、提供激励、实施拒绝转换或采用额外的混合模式等后续干预手段(作为特定调查情境的后续调整),可能将应答率提升至 80%,并将偏差降至 $R_{\text{bias}}(\bar{y}) = 0.10$ 。
- 在相关性较弱和/或只有中等偏差的情况下,提升应答率对降低偏差的作用有限(线 B 和线 C)。幸运的是,这种模式在现实中更为常见。例如,在斯洛文尼亚科学家调查(WCSC)中,若仅纳入初始邀请后、提醒前的受访者,最终样本量为 580,应答率 $RR = 13\%$,而不是最终的 $RR = 24\%$ 。然而调查的估计值不会发生很大变化。当 RR 从 13% 上升 24% 时,所有变量估计值的相对变化都低于 5%。当然,需注意的是,在这项研究中,即便应答率 $RR = 24\%$,仍有四分之三的合格单位未应答。然而,社会科学与市场研究领域的相关研究发现,即便对最后无应答者加以追踪(Stoop, 2005),最终结论均为:为达到 100% 的应答率而付出的昂贵资源不会改变估计值。同样,在 AAPOR 会员调查中,如果将后续波段的应答者包括在内,虽然应答率几乎翻倍,但估计值差异很小(Murphy et al., 2011)。还值得注意的是,在这次会员调查中,第一次提醒将会员的累积应答百分比从 20% 提升至 30%(这是常见效果)。而后续两次电子邮件提醒,并配合邮件和电话跟进,将应答率提高到了 61%。关键在于,这些额外投入是否真的在减少无应答偏差方面得到了回报。
- 线 D 特意说明应答率和应答偏差的关系几乎可以是任何形式,但它也反映了 Vehovar 等人所描述的真实但很罕见的情况。Vehovar 等人(2010)在混合模式调查中通过额外的连续无应答干预实验(即在网络调查的邮件邀请中提供 5 欧

元的预付现金奖励), 将应答率从 33% 提高到了 74%。然而, 这吸引了更多的互联网用户, 并因此增加了与互联网使用有关的目标变量的无应答偏差。例如, 无激励组的每日用户比例为 62%, 有激励组为 76%, 而官方数据报告普通人群的这一比例为 56%。

还需提及的是, 即便存在最轻微的非应答偏差迹象, 其背后通常也存在一定的情境性解释。在上述斯洛文尼亚科学家调查(WCSC)的例子中, 更成功的科学家较早应答, 且首轮应答者的科学成就估计值始终略高于后期应答者。尽管未达到统计上的显著水平但差异趋势稳定。例如, 采用 1~7 的评价量表, 在初始邀请后作出回应的科学家的平均科学成就水平是 5.0, 而仅在提醒后才作出回应的科学家是 4.8。如果在初始邮件后就停止调查, 会产生极小的非应答偏差。尽管如此, 检查估计值的变化方向是否与情境一致、符合逻辑且合理, 仍然是很重要的。

此处简要回顾第 2.2.3 节内容, 该节阐述了变量在概率抽样偏差下的稳健性。同样, 某些变量在非应答方面比较稳健(即不容易受到影响)。总体而言, 百分比(占比)比量表变量、相关性及子群体比较更易受非应答影响。同样, 对于趋势分析来说, 有偏差的估计有时并不那么重要, 只要趋势变化得到准确反映即可(例如满意度下降或市场份额下降)。

2.5.4 网络调查的一般设计特征与非应答

前文已阐述非应答的关键问题与核心概念, 本节将聚焦非应答的预防方法。首先从网络调查的一般特征切入——研究者对其中部分因素(如调查发起方、主题突出性)的控制力较弱, 对另一些因素(如问卷设计)则为间接控制。随后, 将依次探讨与激励措施(第 2.5.5 节)、联系过程(第 2.5.6 节)及邀请函格式(第 2.5.7 节)相关的直接非应答干预措施。

2.5.4.1 调查发起方

用调查发起方(*survey sponsor*)一词来指代任何可以出现在邀请函或调查介绍页中的权威机构或实体, 可以是研究者、分包商、合作伙伴、委托方和其他一些组织, 甚至是可以推荐调查的个人。文献一致指出, 在传统调查(Groves et al., 2009; Groves & Peytcheva, 2008)和网络调查(Dillman et al., 2009; Fang et al., 2009)中, 合法的权威身份会提升受访者信任, 使调查被认为更具善意、诚实性且有能力保护

隐私。

关于参与网络调查意愿 (*willingness to participate*) 的实证研究表明, 网络调查发起方的声誉和受访者对其的信任程度会影响参与意愿 (Bosnjak & Batinic, 2002; Fang et al., 2009; Fang & Wen, 2012)。对于实际参与 (*actual participation*) 行为也是如此。这点在科学家和学生发出调查邀请的例子中得到了证实, 两者应答率的差异非常大: 55% 对 7% (Guéguen & Jacob, 2002)。Kaplowitz 等 (2012) 的研究发现让大学副校长牵头参与教职员调查, 效果也是显著的, 应答率从 33% 上升至 43%。同样, 在一个非限制性的网站评估调查中, 在邀请横幅中显示政府发起将应答率从 11% 提高到了 15% (Walston et al., 2006)。当发起方的级别差异较小或不明确时, 效果就会消失, 比如由主任与行政助理发出的调查邀请 (Porter & Whitcomb, 2003a); 或在高 (发起方—政府部门) 与低 (研究者—大学研究中心) 显著性的赞助者中, 效果差异并不显著 (Boulianne, Klostad, & Basson, 2011)。

因此, 明确展示发起方身份至关重要, 例如通过标识或权威人物的签名。但另一方面, 将发起方标识放在问卷的每个页面上并没有带来任何附加价值, 至少在降低中断率方面作用不大 (Heerwegh & Loosveldt, 2006)。除了应答率, 发起方也会影响数据质量的其他方面。例如, 在一项性骚扰研究中, 当发起方替换为女性主义组织时, 数据质量会有不同 (Galesic & Tourangeau, 2007; Tourangeau, Groves, Kennedy, & Yan, 2009)。

发起方的效果因情况不同而有很大差异, 尤其与文化、组织环境及调查设计 (如电子邮件邀请或邮寄邀请、基于名单或非名单的网络调查) 相关。对发起方的认知及其实际角色的细微差异也很重要。事实上, Fang 和 Wen (2012) 发现, 调查发起方或委托方的声誉通常比执行调查的研究机构的声誉影响更大; 反之, 声誉较差的委托方可能会从研究者的高声誉中获益。

无论网络调查的类型如何, 宣传活动有助于提高目标人群对调查的整体认知。例如, 研究人员出现在传统或网络媒体上、媒体报道调查的重要性, 都有助于促成招募和合作。同样, 在招募过程之外, 来自管理层、上级或其他权威人士的一般性号召也可以激励参与, 特别是对会员、学生或雇员的调查 (Borg & Mastrangelo, 2008, p. 192)。当样本在目标人群中占比不可忽略时, 大型调查可以组织正式的宣传活动, 例如人口普查 (Datta, Yan, Evans, Pedlow, Spencer & Bautista, 2012)。遗憾的是, 关于付费推广 (如广告、公关) 的效果研究较少, 因为在这个领域进行实验难度较大, 潜在效果很难衡量, 很容易被高估和造成资源浪费。

总而言之, 在很多情况下, 能选择在目标人群中具有高声誉的发起方的机会非

常有限。商业或自助式网络调查很获得政府机构作为发起方的支持，但只要存在这种选择，研究人员就应该予以考虑。反之亦然，不应该选择那些可能损害调查声誉的主体作为发起方。

2.5.4.2 主题突出性

主题突出性 (*topic salience*) 指受访者感知到的特定主题的重要性和时效性 (C.Martin, 2004)。它被认为是调查参与的关键因素之一，不仅对应答率有显著作用，还会影响调查过程中的参与质量 (见 2.3.4 节)。

实证研究证实，主题突出性是预测调查参与度的重要积极因素，尤其在学生调查中表现显著 (Adams & Umbach, 2012; Porter & Whitcomb, 2005)，而在在线样本库中，这种影响通常较弱甚至不存在 (Tourangeau et al., 2009)。尽管 Keusch 的研究 (2012b, p. 68) 发现，对某些主题感兴趣的受访者更经常参与：高兴趣组参与率为 79%，比低兴趣组 (72%) 高出近 7%，且应答质量更优。然而，该研究还显示，电子邮件邀请中刻意突出主题对受访者的参与度并没有影响。

除对单位无应答率的影响外，当受访者对调查主题表现出高兴趣时，中断的情况也会更少，这一点由 Galesic (2006) 在一项横幅广告网络调查中发现。Holland 和 Christian (2009) 的研究也表明，那些对调查主题更感兴趣的受访者在开放式问题中会给出更为丰富的答案。因此，在与潜在受访者的交流中强调主题的重要性，有可能成为提高网络调查的应答率和答案质量的有效策略。

在无限制调查中，如客户拦截调查 (Comley, 2000)，主题突出性更加重要，因其在网络环境中更易传播。

对特定人群而言，突出性的作用尤为明显。例如，即使是事务繁忙的高层管理人员也可能因为主题的高突出性而参与其中。月度“商业趋势调查”便是典型案例，该调查经常获得高应答率。²³ 该调查由斯洛文尼亚的统计局实施，自 2014 年以来，他们在混合模式的调查中的累积应答率平均达 85%，其中超过一半的受访者完成了网络调查。²⁴

然而，强调突出性也可能对话题低兴趣的受访者产生负面影响 (Keusch, 2012b)，突出性会强化自我选择偏差，增加无应答偏差风险。Hansen 和 Smith (2012)、Chang 和 Krosnick (2009) 以及 Faas 和 Schoen (2006) 的研究发现，在受访者自我选择的非概率网络调查中，这种风险尤为突出。因为突出性易吸引对该主题有丰富知识的受访者，导致“专家偏向”。鉴于此，有时研究者可能会有意避免强调突出性，而宁愿将其隐藏在调查主题的一般标签后面。

研究人员在这方面的调整空间有限,因为调查主题通常固定。对于普遍关注低的话题,研究人员几乎无法提高其感知突出性。然而,我们在一定程度上可以使用一些直接的营销方法,在标题、邀请函、介绍页和问卷本身中强调调查最突出的方面;同时需平衡突出性的呈现,避免过度吸引特定群体而引发负面影响——这是该策略潜在的副作用。

2.5.4.3 问卷设计

问卷设计对受访者是否访问网络问卷页面的初始决策(单位不回答)影响较小,除非是部分无限制网络调查在首页即呈现问题。Walston 等人(2006)的研究显示,与采用单选按钮的评分量表相比,滑块的繁琐设计使应答率从 13% 降至 8%。问卷设计对调查中断、项目无应答和应答质量的其他方面影响要显著得多,尤其是结构、布局、图形、互动性、问题格式和长度。第 2.3.4 节已在调查参与度的背景下对这些内容展开详细讨论,Vicente 和 Reis(2010)也对影响无应答的一般网络问卷设计因素进行了补充回顾。

至此,本节概述了一般调查设计特点在无应答方面的作用。接下来的三节(2.5.5~2.5.7)将探讨更具体的无应答应对方法,即激励措施、联系策略和邀请函形式。

2.5.5 网络调查中的激励措施

传统调查模式的研究已表明,激励措施能够有效提高应答率(Singer, van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999)。Göritz(2006a)的元分析和其他研究(Gajic, Cameron, & Hurley, 2012; Parsons & Manierre, 2013)也在网络调查中验证了这一结论。本节首先讨论与网络调查中激励措施相关的一般问题,再具体分析抽奖等特殊形式的激励。

网络调查中的激励措施形式多样,对单位无应答、调查中断及应答质量的影响各有差异。以下几条一般的激励原则也适用于网络调查。

- 无条件的激励比有条件的激励更有效(Göritz, 2005)。
- 预付式激励措施一般比承诺式激励措施效果好,但在网络调查中,这种效果似乎比较弱;研究经常显示预付式激励措施没有优势(Bosnjak & Tute 2003)或者效果微弱(Downes-Le Guin, Janowitz, Stone, & Khorram, 2002)。
- 货币激励的效果通常优于非货币激励,如 Vehovar 等人(2009)的研究发现 5 欧元的预付现金激励将应答率从 33% 提高到 74%,而与邀请函一起发送的价

值相近的礼物（礼物为装硬币的钱包）几乎没有效果。

- 服务或权益类激励，如提供专属信息、会员资格、慈善捐赠等，很少有效果，而且通常无法与现金激励相比。例如，一项对 11 项邮寄问卷和电子问卷随机实验的元分析显示，承诺提供研究结果的组别与无激励组的应答率无显著差异（Edards, Roberts, Clarke, DiGuseppi, Wentz, Kwan, ... Pratap, 2009）。即便是基于概率的在线 LISS 小组，向参与者提供详细的多媒体结果也未提升应答率（Scherpenzeel & Toepoel, 2014）。这类激励仅在极少数高突出性主题中可能有效，如在健康调查中，为应答者提供定制化反馈（Balter, Fondell, & Balter, 2012）。
- 激励金额的设定需遵循社会交换理论，按照经验法则，象征性感谢的底线通常为最低面额纸币（如 5 欧元或 1 美元）。另一方面，若想体现对受访者付出时间的补偿，金额应大致匹配其所处环境的时薪水平，发达国家通常是每小时 5 欧元至 40 欧元。
- 很多时候，单增加货币激励的价值，影响非常小。例如 15 欧元对 25 欧元（Downes-Le Guin, Janowitz et al., 2002）、5 美元对 10 美元（Boulianne, 2013）或 15 英镑对 35 英镑（Villar, 2013）的比较均印证了这一点。后一项研究是在英国进行的网络版 ESS 实验，两种金额是以完成为条件的，但在邮件邀请中使用了 5 英镑的无条件预付金。这种两步走的方法对于长时间和高要求的调查来说是很好的模式，但这同样取决于环境；上述实验最终应答率为 20%。而在爱沙尼亚的一项类似研究（Ainsaar et al., 2013）仅使用象征性的非无条件预付式激励就达到了 40% 的应答率。另外，Alexander、Divine、Couper、McClure、Stopponi、Fortman 和 Johnson（2008）发现，在所研究的 24 种激励策略中，2 美元的无条件预付和 20 美元的完成条件支付的组合效果最佳。
- 对于网络环境下的特定激励形式（例如各种形式的在线支付、移动账户转账、网上书店优惠券）的实证研究较少，其潜力尚未被充分挖掘。

对在线样本库而言，通过积分累积实现的间接货币激励已成为表达感谢的标准方式。这种激励方式在小组招募（Scherpenzeel & Toepoel, 2012）和成员参与单个调查（Görizt & Luthe, 2013）中的作用已被广泛研究。

激励措施的效果也取决于网络调查背景。一项对网络调查的元分析研究（Lozar, Manfreda, & Vehovar, 2002）考察了影响应答率的几个因素，发现激励措施与其他手段结合时，主要作用是减少中断，对单位无应答的影响有限。需注意的是，元分析研究中可能存在虚假关联。因为冗长或乏味的调查更倾向于采用激励措施，而这类调查本身的预期应答率就较低（Cook et al., 2000）。

与强调研究主题的突出性相似,激励措施也可能不成比例地吸引特定单位,如收入较低的单位(Göritz, 2008),但这会增加无应答偏差(Parsons & Manierre, 2013)。此外,激励措施还可能使受访者更倾向于迎合调查者(如提供有利的回答),或更易采取满足策略,甚至在条件允许时重复填写问卷(Comley, 2000)。

一般而言,激励措施可以提高数据质量,特别是应答率、项目无应答、对开放式文本问题的丰富回复、回复一致性以及降低满足作答。同时,它们通常在完成时间或回复方式上很少出现问题(Göritz, Wolff, & Goldstein, 2008)。

在实施激励措施时,研究人员还需关注法律与税收环境。例如,部分国家(如斯洛文尼亚)规定,超过一定金额的激励需即时缴纳所得税,这对网络调查的操作会造成不便。

抽奖(奖券)在网络调查中具有特殊地位。通过承诺随机抽取获奖者并发放奖品,既能吸引受访者,也会降低成本。鉴于每次调查一般都有一个或多个奖品,与预付或后付的个人奖励相比,抽奖模式的成本低,因此对研究人员很有吸引力。部分研究显示,设置大量小额奖品比少数高额奖品的效果更好(Deutskens et al., 2004)。

抽奖同样受到严格的法律监管。研究者需遵守国家法规,尤其在跨国调查中需妥善处理相关文书工作。例如,美国调查研究机构协会(2012)提供了相关指导方针, Lord 和 Miller (2009)则进行了更详细的阐释;联系当地调查或市场研究协会是获取合规信息的有效途径。

Göritz (2006b)对研究网络调查中激励措施的实验进行了元分析。结果显示,相比不提供奖励,提供抽奖的效果很小,受访者参与调查的概率平均仅增加19%。对在线样本库的研究显示了类似的效果(Göritz & Luthe, 2013; Tuten, Galesic, & Bosnjak, 2008)。对于单次的网络调查,结果遵循类似的模式,但效果差异较大:从无影响(Cobanoglu & Cobanoglu, 2003)到中等影响(Heerwegh, 2006; Porter & Whitcomb, 2003b),最高可使应答率提升30%(Sauermann & Roach, 2013)。

最后需考虑激励措施的成本。首先,在评估替代方案时(如是否使用激励、选择何种类型与金额的激励),必须比较每个受访者的总成本。也就是说,即便无激励时可获得5%的应答率,其人均成本也可能比使用激励后获得30%应答率的方案人均成本更高(Downes-Le Guin et al., 2002)。当然,每个受访者的成本需结合对应的数据质量综合考量。其次,当使用激励措施时,应该寻求过往研究依据,证明它们已经对研究背景起作用,确认其在类似情境中有效。这可通过小型实验验证,若实验显示激励无效果(Žagar & Lozar Manfreda, 2012),则无需投入相关成本。

2.5.6 联系过程的作用：模式和频率

在讨论联系潜在受访者时，我们主要聚焦基于名单的网络调查，如第 2.2 节讨论抽样时所述；随后将这些原则推广应用于其他场景。

联系潜在受访者的邀请信息主要分为三类：预先通知、主要调查邀请和提醒。预先通知（*pre-notification*）是在主要邀请前几天发出，可采用与主要邀请不同的通信渠道（例如，通过邮寄或短信发送预先通知，再以电子邮件发送主要邀请）。其目的是告知受访者即将收到调查邀请，以激发其参与兴趣，并增强调查的合法性与重要性认知。尽管预先通知为可选环节，但主要邀请（*main invitation*）通常以电子邮件或邮寄形式发送，内容包含网络问卷的 URL 及其他必要信息。提醒（*reminders*）的主要目的是强化受访者对研究重要性的认知，通过重复接触调查请求来提高应答率。

2.5.6.1 联系方式

联系潜在受访者的方式需根据实际情况确定，但很多时候研究者的选择空间有限。例如，针对普通人群的调查，可能仅能获取邮政地址作为抽样框（Smyth et al., 2010; Vehovar et al., 2009）；有时则仅有电子邮件可用，或可同时使用电话号码。在后一种情况下，如何优化多模式联系方式成为复杂的决策难题。

与传统信件相比，电子邮件的优势是速度快、成本低；受访者只需要简单的点击就可以访问网络问卷，负担较小。此外，电子邮件还可以便捷地提供额外的材料（如传单、小册子），并直接与潜在受访者沟通；而邮件地址可能会干扰其他家庭成员。

然而，电子邮件也存在明显缺陷：地址变更频繁、易出现拼写错误、存在病毒威胁（例如，用户不敢点击一个未知发件人的 URL），或被误认为垃圾邮件。对于信件邀请，只有上述少数问题存在（如地址错误，收到未经请求的邮件），但影响程度较轻。此外，信件在传达调查请求的合法性方面更加有效，有助于建立信任和附带预付奖励。

实证研究表明，当潜在受访者与研究者之间已有联系时，电子邮件的效果可能与邮寄或明信片相当，甚至更优（Kaplowitz et al., 2012）。另一方面，当双方事先没有联系时，信件可能看起来更合法，因此更有效。正如 Vehovar、Lozar Manfreda 和 Batagelj（2001）的商业调查及 Bandilla、Couper 和 Kaczmirek（2012）的普通人群调查中所表明的。

当同时掌握电子邮件和邮寄地址时,两者的组合可能发挥协同作用。通常情况下,先邮寄信件发送预先通知,以增加调查的合法性,避免被视为不请自来的邀请;再使用电子邮件发送主要邀请,简化受访者访问问卷的流程;后续提醒则优先选择电子邮件,因其操作简便且成本低(Millar & Dillman, 2011)。当然,也可采用相反的策略:初期通过电子邮件联系,后续对未应答者改用信件或电话跟进,以提高应答转化率(Murphy et al., 2011)。

手机短信(即文本信息)也可作为联系受访者的替代方式,能有效避免垃圾邮件拦截问题,且更易吸引注意力(移动设备通常会通过声音和视觉提示新信息)。随着智能手机的普及,它也支持电子邮件通信,短信的优势可能会逐渐减弱。不过,短信邀请还是比邮寄信件和电话联系的成本更低。Bosnjak、Neubarth、Couper、Bandilla 和 Kaczmirek(2008)的研究发现,短信预先通知和电子邮件邀请的组合效果较好。Hansen 和 Pedersen(2012)则发现,在线样本库的短信招募效果不及电话招募,但与邮寄招募相当。

2.5.6.2 发出邀请的日期和时间

在访员管理的调查中,联系受访者的最佳时间通常是在工作日下午(面对面调查)或晚上(电话调查),周末则尤为适宜。然而,网络调查邀请的时间选择更为复杂,其效果会因目标人群(如雇员、学生)的不同而有很大差异。多项研究表明,一周中不同日期和一天中不同时段邀请效果存在较大差异(Basso & Rathod, 2004; Faught, Whitten, & Green, 2004; Granello & Weathon, 2003; Sauermann & Roach, 2013; Shinn, Baker, & Briers, 2007; Survey Monkey, 2011; Trouteaud, 2004)。避开受访者的忙碌时段,这是调查联系时间的一般原则。观察受访者使用电子邮件的习惯,以及当前调查项目中的回复时间,对于决定何时发送邀请至关重要。

Faught 等人(2004)发现,向美国制造业公司发送电子邮件邀请的最佳时期为星期三,但也有研究发现这一日期对调查结果无显著影响(Shinn et al., 2007)。Sauermann 和 Roach(2013)的研究中也是如此;不过,周末送达的邀请虽可能被及时查看,但潜在受访者当天回应的可能性比较小。Trouteaud(2004)针对美国大型企业的调查发现,清晨发送的电子邮件应答率高于午餐时间。Granello 和 Weathon(2003)的研究表明晚间6点发送的提醒,其第二天的完成率显著高于下午3点30分发送的提醒。关于由18~40岁的荷兰人组成的在线样本库调查发现,调查最好从周一开始(或至少包含周一),因为周五的邀请最有可能在周一产生回应;中午时段应纳入正式调查的时间规划(Basso & Rathod, 2004)。Survey Monkey

(2011)对超过10万份客户满意度调查和员工组织调查的观察研究中,周一也是应答率最高的一天。最后,2007年针对在线调查小组 Knowledge Networks (M. Callegaro & C. DiSogra, 个人沟通)的56.1万个案例的分析发现,周日和周一发送的邀请在次日获得的完成量最高;而在初始邀请三天后发送提醒,总体应答率会趋于稳定。

综合来看,当采用提醒策略时,不同日期或时段的初始应答率差异会被逐步抵消。因此,除非调查需快速结束或仅依赖单次电子邮件邀请,否则无需过度纠结特定时间点的选择。

2.5.6.3 联系次数及其时间安排

在传统调查模式中,增加联系次数通常能提高受访者的参与概率(P. J. Edwards et al., 2009)。这一规律同样适用于网络调查,联系次数是预测应答率的最重要因素之一(Cook et al., 2000),这也是自我管理调查的定制设计方法所倡导的(Dillman et al., 2009)。

预先通知可以提高电话和邮寄调查的应答率。例如,de Leeuw、Callegaro、Hox、Korendijk 和 Lensvelt-Mulders (2007)的研究显示,电话调查中预先通知可使应答率(RR1),平均从58%提高到66%。而Edwards等人(2009)的结果表明,预先通知对邮寄调查的效果更为显著,能使应答概率提升近50%。这一结论在网络调查中也得到了验证(Dykema, Stevenson, Day, Sellers, & Bonham, 2011; Keusch, 2012a; Porter & Whitcomb, 2007; Wiley, Han, Albaum, & Thirkell, 2009)。不过,也有部分研究未发现预先通知的显著效果(Felix, Burchett, & Edwards, 2011; Hart, Brennan, Sym, & Larson, 2009)。

同样,实验研究(Kaplowitz et al., 2004)发现,提醒在网络调查中同样有效,这一结论也被元分析所证实(Batanic & Moser, 2005; Göritz & Crutzen, 2012)。后者研究警告说,提醒可能会不成比例地吸引受访者,但他们经常中断调查。

关于联系次数(*number of contacts*),重复追踪的边际效益通常呈递减趋势,尤其是电子邮件提醒。过度频繁的提醒可能被视为垃圾邮件,引发受访者反感,甚至降低其未来参与调查的意愿。通常,两次电子邮件联系(即向所有受访者发出邀请函加感谢信提醒)通常已能覆盖大部分潜在应答者,而额外的提醒增益有限。但也存在例外情况,Kaplowitz等人(2012)的研究显示,第二次提醒(即第三次电子邮件联系)将教师的应答率从32%提高到40%,学生的应答率从15%提高到19%。同样,Rao和Pennington(2013)的研究也发现,第二次提醒能增加约10%的应答者,

但这些受访者的应答质量相对较低, 对整体估计值的影响较小。

除了联系次数, 联系间隔的时间安排同样重要。当发送网络调查的邮寄邀请时, 时间安排可能类似于邮寄调查, 通常建议在首次发送后一周内发送提醒, 并假设大多数人都对初始邀请作出回应。而电子邮件邀请的应答速度更快(很多时候, 大部分回应都是在第一天获得), 也随时间迅速下降(送达三天后, 与某联系人相关回复不到10%, 甚至不到5%), 因此对于电子邮件, 可适当压缩时间间隔。图1.2中的WCSC调查也是这种情况。根据上述研究结果, 建议在邀请发送后3-4天内发送首次提醒。Crawford等人(2001)发现, 高频次短间隔提醒(如小组在两天后收到第一次提醒, 又在两天后收到第二次提醒)比低频次长间隔(在五天后收到一次提醒的小组)更能快速提升应答率(36%对33%)。另一方面, Deutskens等人(2004)及Muñoz-Leiva、Sánchez-Fernández、Montoro-Ríos和Ibáñez-Zapata(2010), 还有Sánchez-Fernández、Muñoz-Leiva和Montoro-Ríos(2012)发现, 当联系之间的时间间隔延长至一周及以上时, 提醒次数多或少, 应答率没有明显差异, 这表明联系的时间间隔越短越有效果。

2.5.6.4 调查实施的时间长度

延长调查实施时间通常有助于获取更多应答, 但超过临界点后, 新增应答的数量会微乎其微, 对总体估计几乎无影响。同样, WCSC调查(图1.2)很好地说明了电子邮件邀请的网络调查中的普遍情况。

- 大多数应答是在邀请/提醒后立即得到的(一周后应答量几乎趋近于零)。
- 第一次提醒使应答率有了非常大的提升(从13%提高到23%), 而第二次提醒的增益微弱, 只使应答率从23%提高到24%。
- 从应答模式可推测, 额外提醒仅能带来不足1%的新增应答单位。
- 与第一次提醒之前获得的数据相比, 后续新增应答对估计值的影响可忽略不计。因此从数据质量角度, 此时可停止提醒(但更大样本量仍可能带来其他分析优势)。

对于在线样本库, 研究者可能会预期应答速度更快。例如, 在荷兰的一个在线样本库中(Basso & Rathod, 2004), 受访者在收到邀请后平均1.75天应答, 大约60%的受访者在收到邀请后的24小时内应答, 80%在2.5天内应答, 仅大约5%的受访者需要一周以上。

如果调查时间充裕, 适当延长调查周期几乎无负面影响。但是, 当时间不充

裕时,对于大多数电子邮件邀请的网络调查来说,最后一次提醒后的一周时间已经足够收集大多数潜在应答;而在时间压力极大的情况下,可缩短至三天甚至两天。

关键在于是否值得等待最后一批受访者。正如在 WCSC 调查中所显示的,这类受访者不会带来总体估计值的变化,在测量态度时尤其如此,因为其潜在贡献低于所有单位的 10% 甚至 5%。但在事实特征或行为测量中,情况则不同,最后一批受访者的贡献可能较大。例如, Kypri、Samaranayaka、Connor、Langley 和 MacLennan (2011) 及 Kypri、Stephenson 和 Langley (2004) 的研究发现,在最后五分之一的受访学生中,健康风险行为和不同饮酒特征的发生率更高。同样, Irani、Gregg 和 Telg (2004) 在最后一半的受访者中也观察到信息与通信技术技能的差异。无论何种情况,通过日常监测应答数据,研究者可有效判断截止时间,并评估额外提醒的潜在价值。需注意的是,目前尚无统一的截止标准,其判断需结合应答率与无应答偏差的关系(见 2.5.3 节)。此外,在商业调查等场景中,明确告知受访者截止日期也有助于提高应答效率。

2.5.7 邀请函的具体内容格式

网络调查邀请函的核心功能是说服受访者启动并完成问卷填答。经典调查研究文献指出,有效的邀请函需具备信息完整性、合法性、引导性和激励性四大特征,并明确包含以下要素(Dillman et al., 2009, p. 149):

- 说明对受访者的具体要求。
- 简要描述调查的核心内容。
- 解释受访者的选取方式及原因。
- 承诺数据保密措施。
- 提供额外联系方式(如回复电子邮件)以应对潜在问题。
- 明确访问网络问卷的操作指引。
- 若为电子邮件邀请,需附加电话号码、邮寄地址等辅助联系方式以增强合法性。
- 依据专业伦理规范,说明退出调查名单的途径(另见第 6.3.3 节关于伦理问题)。

这些关键原则达成了普遍的共识,但特定情况下可以酌情省略一些原则。此外,电子邮件营销领域(email marketing)的实践经验也可为邀请函设计提供参考,因为调查邀请类似于“打动”潜在客户以执行所需操作。相应的市场研究文献揭示了许多因素,这些因素与研究者需要采取的行动相重合,以使收件人打开电子邮件(取

决于主题行和电子邮件的发件人)、关注内容(电子邮件的特点和吸引力)并点击 URL 链接(激励措施的特征和吸引力)。市场研究文献强调,邀请函应传递紧迫感与重要性,让受访者感知到自身参与的独特价值;同时需重复行动呼吁、使用有效的设计并提供吸引人的、精心设计的主题行。更多案例见第2章补充内容“电子邮件营销方法概述”(Overview of email marketing approaches)(<http://websm.org/ch2>)。

研究者面临的核心问题是如何将上述要素整合为切实有效的实施方案。为此,下文将对网络调查邀请的个性化、语气和设计展开具体分析。我们主要关注的是电子邮件邀请,其原则亦适用于其他形式的邀请。

2.5.7.1 邀请函的个性化

根据 Dillman 等人(2009, p. 272)的研究,网络调查中邀请函的个性化能够在研究者与受访者之间建立连接,这是促成社会交流的必要前提,同时可提升受访者对自身参与重要性的感知。个性化的实现程度取决于抽样框中可用的联系信息,例如,如果知道受访者姓名,则可以在称呼中使用个人姓名(“亲爱的[个人姓名]”),否则只能使用“亲爱的女士”或“亲爱的先生”等泛化表述,如果只知道电子邮件地址,也是如此。正如在电子邮件设置部分(第2.4.1节)所讨论的,电子邮件应该单独发送,这意味着在收件人行(‘To:’)只出现收件人即受访者的电子邮件地址,而发件人(‘From:’)则需要仔细选择。

Cook 等人(2000)和 Edwards 等人(2009)的元分析发现,个性化对网络调查的应答率有积极影响。在电子邮件中,与使用的非个性化敬语相比,使用“亲爱的[个人姓名]”也能提高应答率(Heerwegh et al., 2005; Joinson et al., 2007; Muñoz-Leiva et al., 2010; Sánchez-Fernández et al., 2012; Sauermann & Roach, 2013),这一规律在邮寄邀请中同样适用(Sinclair, O’Toole, Malawaraarachchi, & Leder, 2012)。此外,对比单独发送与群体发送邀请的实验(Barron & Yechiam, 2002)也表明,个性化处理能提升参与度。不过,也有少数研究,如 Porter 和 Whitcomb(2003)以及 Mueller、Straatmann、Hattrup 和 Jochum(2014)的研究,未发现个性化对应答率的影响。

需注意的是,个性化可能产生潜在负面影响。在基于名单的样本中,受访者会意识到研究者设置了身份识别(唯一的 ID 代码),从而感知到匿名性的缺失。这可能导致数据质量下降,例如敏感问题的应答更趋社会赞许性、项目无应答率升高,或受访者回避披露敏感信息。尽管网络调查中这一关联的实证证据尚有限,但 Heerwegh 等人(2005)发现,收到个性化邀请的受访者更倾向于在敏感问题上给出符合社会期望的答案;Joinson 等人(2007)发现,个性化问候会增加受访者选

择“不愿回答”选项的比例。

因此,在利用个性化提升应答率的同时,需权衡其对敏感话题数据收集的潜在影响。如果调查涉及敏感内容,研究者可考虑放弃个性化处理。进一步的优化方案包括使用统一 URL 链接,因为受访者知道,电子邮件邀请函中使用独特的 URL 网络链接,自己的身份可以被识别,就像使用其名字一样。当然,非个性化的 URL 只能实现一种或两种电子邮件设置,因为我们无法追踪无应答者。此外,研究人员必须预先假设受访者不会多次填写问卷。卢布尔雅那大学的非官方学生课程评价与员工满意度调查中,采用统一的 URL 并在邀请中明确说明“链接非个人化,无法识别身份”,这不仅显著提高了应答率,还促使开放式问题获得了更丰富的反馈。该方式最初是在发送带有个性化 URL 的邮件之后,作为第二次提醒的配套策略使用,后续控制问题也证实,这种策略能增强受访者信任。此外,受访者的定性反馈也证实了这一策略的效果非常积极。另一方面,也没有虚假或多次应答的迹象。

然而,如何在保留唯一 ID 追踪功能的同时确保完全匿名,仍是实践中的难题。一种可行方案是,在封闭群体(如学校或组织)中,通过随机抽签分配含唯一 ID 的实体信件,但该方法的适用性有限。

2.5.7.2 提供激励与请求语气

邀请信息的主要作用之一是激发受访者的参与动机。如前文所述,调查参与动机的理论基础可追溯至社会交换理论(见第 2.5.1 节)。Dillman 等(2009)指出通过提升感知回报、降低感知成本及建立信任来促进(网络)调查中的合作。基于名单的网络调查中,邀请函的激励设计可归纳为以下维度。

- 提升感知回报(*benefits*)可以列举如下:(a)强调调查结果对受访者个人(b)及社会的价值,(c)重复表达研究需求以强化被重视感,(d)以积极语气表达感谢,(e)突出群体归属感,(f)让受访者知道其他潜在受访者已经作出回应,(g)强调参与机会的稀缺性。
- 通过提供信息和方法降低感知成本(*perceived costs*),(h)简化参与流程(例如在电子邮件邀请中提供网络调查的链接),(i)避免使用居高临下的措辞和(j)强调与受访者已经回应的其他任务之间的相似性,即关联受访者熟悉的情境(以此保持其态度、信念和行动是一致的)。
- 通过邀请函建立信任(*establishing trust*)的方法包括:(k)从合法机构获得赞助,(l)预先说明感谢机制,(m)强调调查的重要性,以及(n)承诺信息的保密性和安全性。

上面列出的网络调查邀请中的14项策略虽未穷尽所有可能,但为邀请函设计提供了具体方向。以下结合实证研究进一步说明关键策略的效果。

- 邀请函的恳求语气 (*Pleading tone of the invitation*)。两项研究 (Felix et al., 2011; Trouteaud, 2004) 均表明,与没有恳求语气的邀请相比,使用带有“恳求”语句的电子邮件邀请时,应答率有所增加。
- 稀缺性提示 (*scarcity*) (例如,告诉受访者他们是从小型选定群体中筛选出来的,或者明确调查的截止日期)。Porter 和 Whitcomb (2003) 以及 Henderson 和 Callegaro (2011) 的研究均表明,稀缺性会增加网络调查中的应答。但只有在有激励措施的调查中强调截止日期才有效,而在没有激励措施的调查中则无效。
- 与他人的相似性 (*similarity to others*): 对网络问卷应答率的实验研究进行系统回顾发现,如果邀请函包含“他人已经回复”的声明,那么应答率会上升 (P. J. Edwards et al., 2009)。

2.5.7.3 电子邮件邀请函的格式与视觉设计

在无访员参与的网络调查中,邀请函的格式和视觉设计是使受访者在阅读实际内容前就对调查感兴趣的重要因素。对于纸质信件邀请,可参考传统邮寄调查的设计原则 (Dillman et al., 2009); 而电子邮件邀请的设计需特别关注以下内容。

- 视觉外观 (*Visual appearance*) 应该是功能性的。带有不同颜色背景、使用不同字体类型和大小以及嵌入图形等花哨电子邮件,通常效果不佳。Whitcomb 和 Porter (2004) 在一项针对学生样本的实验中发现,从没有标题的纯文本电子邮件联系到有标题的文本电子邮件,再到有白色背景和简单标题的 HTML 电子邮件时,应答率逐步提升。关于图像的使用,Chittenden 和 Rettie (2003) 发现电子邮件调查活动中使用图像越多,应答率越高;但过度使用图像在移动设备上可能显示异常,且可能增加邮件体积导致加载缓慢,建议电子邮件中的平均图像数量为 1.7。此外,过于复杂的 HTML 背景和装饰性标题反而可能降低效果 (Whitcomb & Porter, 2004)。
- 邀请函的长度 (*length*) 通常应简洁精炼 (Dillman et al., 2009), 邮寄信函的长度为一页,电子邮件的长度更短 (例如不超过一屏显示内容)。所有的附加信息都可以在调查问卷的介绍页或研究项目的网站上发布。然而, Klofstad、Boulianne 和 Basson (2008) 的研究发现,在实践中电子邮件邀请的长度对应答率的影响并没有显著差异。而 Kaplowitz 等人 (2012) 甚至发现,在对

教职员工的网络调查中,较长的文字有积极的作用,但对学生群体则没有作用。这与较短的邀请函效果最好的预期相反——该预期的依据是,有观点指出(Nielsen, 2000a),互联网用户不是阅读而只是扫描屏幕上的文字。同时,这也与电子邮件营销的普遍经验相反。正如Kaplowitz等人(2012)所认为的,较长的文本可能传达“严肃性”或“重要性”,而这正是应答增加的原因。

- 垃圾邮件(spam)是电子邮件邀请函的一大挑战。避免垃圾邮件的一般方法是:确保发件人的IP地址无垃圾邮件记录、使用纯文本格式而不是HTML信息格式、单独发送邮件而不是群发邮件和避免使用营销敏感词(免费、现金、中奖、促销、奖品)以及通过垃圾邮件检测工具预筛内容,这些内容会触发垃圾邮件过滤器;等等。这类工具提供了垃圾邮件风险评分,并允许改变关键词和电子邮件的其他方面来降低该评分。即使在主题标题中使用“调查”一词,也会有风险,容易被垃圾邮件过滤器识别。如果受访者有来自同一机构的电子邮件地址,那么最好获得机构管理者的支持,以避免电子邮件邀请被大量拒绝。应特别注意发送邀请函和提醒函的电子邮件地址,要使用能被受访者电子邮件系统识别的域名。Hartford、Carey和Mendonca(2007)的研究发现,将发件人电子邮件地址从研究机构变更为大学域名,能够有效规避垃圾邮件过滤器的拦截。不过,正如第4.1.1节关于技术准备的内容所述,电子邮件软件(或综合网络调查软件)正逐渐承担起大部分反垃圾邮件的职责,为研究人员处理这一日益复杂的问题提供专业支持。
- 电子邮件邀请函的标题——“发件人”行(*Header of the email invitation – ‘From’ line*)需要说服受访者不删除未读邮件,而相信发件人的信誉和邮件的重要性。该行传达调查的合法性,所以要谨慎考量。正式全称通常优于昵称,机构域名邮箱(organization.com或university.org)比公共邮箱(如Yahoo、Gmail、Outlook)更易建立信任。然而,实际效果需结合具体情境判断,尤其是发件方与相关主体(发起方、委托方、研究者、软件供应商)的声誉、认可程度及相互关系。
- 电子邮件邀请函的标题——“主题”行(*Header of the email invitation – ‘Subject’ line*)需兼具信息性与专业性,以区别于垃圾邮件或营销信息,从而吸引受访者关注。实证研究显示主题行可采取多种方式增强吸引力。Porter和Whitcomb(2005)的研究表明,当样本群体对调查参与意愿极低时,主题行的文字差异对效果影响有限。关于在主题行中提及激励措施的结果也尚未在研究中形成定论。虽然Henderson和Callegaro(2011),Smith和Kiniorski(2003)和Zhang(2011)发现强调奖励是积极的,但Kent和Brandal(2003)却发现情况恰恰

相反,提及奖励会降低应答率。Kaplowitz 等人(2012)则发现包含权威人士请求的主题行比突出调查主题的更易获得回应。Henderson 和 Callegaro(2011)发现主题行中的礼貌性请求(“请提供您的反馈”)比疑问句式(“您愿意提供您的反馈吗?”)更有效。

综上,上述要素的优化需结合具体情境调整,如目标人群、主题及其突出性、与发件人的关系、调查目的等。一项关于电子邮件调查邀请的开创性研究(Tuten, 1997)表明,当主题行不能引起受访者的兴趣或受访者不认识发件人的名称时,电子邮件信息就会被删除。更多细节可参见 Dillman 等人(2014, chapter 9)以及第2章补充内容“电子邮件营销方法概述”(Overview of email marketing approaches)中对电子邮件营销文献的回顾(<http://websm.org/ch2>)。

2.5.8 无应答策略

基于名单的网络调查中,电子邮件地址很容易获得(例如,学生、雇员、会员、客户等),此时应对无应答现象相对简单。除发送电子邮件邀请及若干提醒外,通常无需额外措施。当有明确依据表明潜在的无应答偏差可以忽略不计时,也是如此,在此过度投入无应答策略反而会造成资源浪费。反之,当高应答率对研究至关重要时,则需整合前文所述各类方法,制定精细化的无应答策略,这一过程往往复杂、耗时且成本较高,在在线样本库调查中尤为常见。对于非基于名单的网络调查,其无应答策略与第2.2节中关于抽样方法的论述一致,均可参考基于名单的网络调查的原则进行近似设计。

研究者需对网络调查的应答率持有合理预期。如前文所述,仅在主题突出、发起方权威且调查流程专业的情况下,应答率才可能达到30%~40%;更高的应答率(如60%以上)仅出现于特殊场景,例如极具吸引力的主题(如活动后评估、购后反馈)、高度特定的目标人群(如凝聚力强的协会成员),或经过精心设计的调查方案(Murphy et al., 2011)。在许多营销调查及自助式网络调查中,应答率不仅偏低,其绝对数值的参考价值也有限。

需强调的是,在非名单式、非概率网络调查中,虽无法计算传统意义上的无应答率,但无应答及调查参与度仍是需关注的核心问题。研究者需在网络调查核心流程的所有三个阶段系统性应对无应答,其中预调查的策略制定最为关键。因为在预调查阶段研究者要确定无应答策略,具体可按以下步骤推进。

- 收集关于商业调查中应答率和无应答偏差的相关经验和专家证据。梳理同类调查中应答率与无应答偏差的案例。例如, 医疗从业人员的调查应遵循特殊规律 (Pit, Vo, & Pyakurel, 2014), 此时其他医疗调查的相应经验是至关重要的。
- 评估应答者和无应答者之间的潜在差异。如果对两者的差异缺乏认知, 制定无应答策略将存在较大风险 (详见第 2.5.3 节关于无应答偏差的讨论)。
- 考虑并模拟潜在的无应答偏差, 使用近似值 $\text{Bias}(\bar{y}) = W_{nr} \times (\bar{Y}_r - \bar{Y}_n)$, 在不同的调查设计、无应答策略和应答率假设下模拟偏差范围。
- 结合网络调查项目的目标、任务、时间、数据质量和成本等因素, 确定实现高应答的重要性及需达到什么水平的应答率。
- 从数据质量角度评估不同无应答策略的“单位应答成本”。小样本高质量的应答可能比大样本低质量的应答更具价值, 更能满足研究需求。
- 当网络调查措施无法有效解决无应答问题, 需考虑大幅度调整设计 (Groves et al., 2009), 采用不同的研究设计方法和无应答策略 (比如采用分层设计) 来分离和处理目标人群中易出现无应答现象的子群体。实施两阶段抽样, 即对无应答的受访者进行子抽样, 采用更密集的策略联系他们 (通常这种策略成本更高、更耗时)。实施混合模式设计, 甚至重新评估网络调查模式的适用性。
- 投入时间和资源规范无应答指标的计算, 对其进行监测, 并预先准备相应无应答的干预措施 (如应答率过低时追加提醒)。
- 对无应答策略的各个环节, 需要提前制定测试流程、时间和资源投入。例如, 事先规定电子邮件信息的测试主题、时间以及方式是非常重要的。
- 在正式调查阶段和调查后阶段, 预见到与无应答有关的活动。

无应答策略的制定建议从简单措施起步, 优先采用低成本且已证实有效的方法, 如预先通知、精心设计邀请函、改进问卷设计或发送提醒。当这些措施仍不满足需求时, 要考虑更复杂的无应答策略, 包括激励措施、多模式联系、混合模式设计或调整抽样方案 (如两阶段抽样)。当然, 由于此类策略涉及较高的复杂性与资源投入, 建议咨询专家或有经验的研究者。

无应答策略的基本组成部分可以被纳入招募实施计划。对于基于名单的网络调查, 需明确邀请工具格式 (信件、电子邮件、短信) 及联系流程 (顺序、频率、模式、时间)。对于非基于名单的网络调查, 需调整上述原则, 要包括明确招募渠道 (如横幅广告)。与这两种类型网络调查都相关的措施还包括: 激励措施、推广方案、介绍页设计, 及相应的技术、管理、行政支持措施等。最后, 需预设预期基准及应

对方案,包括已实施调查的监测方案及干预措施。虽然准备这些会产生形式化文档,但在调查后阶段的方法报告中仍需呈现。因此提前开展这些可以节省时间,且有助于提升准备工作的严谨性(也见第2.6节的管理)。

在正式调查阶段(*fielding step*),需严格执行预设计划并实时监测。当发现有严重偏离预期基准的情况时,立即启动预先指定的行动和干预措施,例如调整时间安排、扩大初始样本量、修改激励方式、联系和邀请格式等。第3章将进一步详细介绍这些方面。

某些对无应答的补救措施可以在调查后阶段(*post-fielding step*)进行,主要是对无应答的调整、准备无应答相关文档和评估无应答率和偏差(见第4章)。

此外,某些无应答活动可能超出核心网络调查过程。如果网络调查最终出现严重的无应答偏差现象,要对其进行高级统计调整,或者通过邮寄、电话或面对面形式进行后续调查。

总结第2.5节,网络调查的无应答策略与传统调查类似(Albaum & Smith, 2012; Lozar Manfreda & Vehovar, 2002),没有特定的无应答策略可以作为“万能钥匙”适用于所有网络调查环境;而且对无应答策略的阐述源于实践目的。在实际应用中,需注意复杂问题的过度简化,例如不同变量可能需要差异化无应答策略,并权衡无应答偏差、数据质量、时间与资源的相互作用,确保无应答策略与研究目标一致。

2.6 综合管理

管理作为一项组织功能,贯穿于网络调查全流程,承担着指导、组织、协调、优化、控制及筹备各阶段活动的核心职责。管理工作通常始于初步研究活动,因为在核心网络调查过程开始之前,项目管理计划通常已初步拟定。

管理中的组织结构和过程控制是非常重要的,因为它们衔接所有流程与决策的中枢。在某种程度上,管理通常是成功的网络调查过程的关键组成部分。正如本书在问卷准备部分(第2.3.6节)所述,良好的管理对于引导网络调查过程走向预设目标至关重要。对于大型或复杂调查,标准化管理方法与业务流程优化的价值更为凸显。然而,鉴于本书聚焦方法论,在此不进一步阐述这些方面。第6.2节将在更宽泛的网络调查背景下补充具体的项目管理要点。

管理任务非常重要的一部分是拟定综合管理计划,若已拟定,则需对其进行细化与修改。简单网络调查或许无需复杂形式,但预先构建基础流程图(*flowchart*)

仍然非常重要。这一基础流程包括时间安排、关键里程碑 (*milestones*) (表示过程中的关键事件) 和核心成果清单, 以及相关人员的工作时间与资源成本估算。

在网络调查过程的具体背景下, 可以列出以下基本管理活动。这些活动在项目中需要考虑到。

- 预调查阶段的管理支持 (第 2 章) 覆盖前期所有步骤, 尤其需协调可能存在的相互影响、重叠或矛盾情况。
- 正式调查阶段的管理支持 (第 3 章) 涉及招募、测量、处理和监测等任务, 包括研究者可能需要采取的具体干预措施 (例如, 决定追加电子邮件提醒或延长调查阶段周期)。
- 正式调查阶段后期的管理支持 (第 4 章) 包括相应阶段的活动, 特别是数据准备 (高级编辑与验证、缺失值估算、加权处理)、派生数据生成 (编码、重编码), 以及综合报告标准、结果分析、归档与输出规范。
- 整体管理还支持跨阶段流程衔接、全程监控与干预、内外部沟通 (含战略、计划、规则与活动)、行政保障、质量控制、合规性管理 (满足内外部标准)、文档归档, 以及确保所有活动与项目总体目标一致。

需要特别关注跨阶段活动的协调和优化。例如, 在技术准备阶段, 测试是作为技术性预测试出现的 (第 2.4 节), 但也作为问卷测试和现场试验出现在问卷准备阶段 (第 2.3 节), 以及调查阶段的软启动 (*soft launch*) (第 3 章)。

管理活动贯穿网络调查全过程, 并与其他阶段持续互动。因此, 与其他阶段不同, 其在核心过程示意图 (图 1.1, 第 1.2 节) 中的位置难以完全体现其概念性作用。尽管如此, 展开讨论这些问题的最好时机是概述所有的预调查步骤之后。

这一管理阶段是对所有预调查准备工作的正式确认与协调, 意味着预调查、正式调查及调查后各环节的计划与文件需在此阶段形成书面记录。在这种情况下, 几乎所有调查都需准备此类文件, 其中大部分内容取决于发起方或委托方的外部要求等。这些材料通常出现在方法学部分和 / 或报告的附录中。因此, 本书强烈建议将过程、方法与管理相关文件作为预调查准备工作的副产品, 至少达到最终报告所需的完整度。除了主要在预调查阶段准备相关计划和文件, 还需记录这些计划在正式调查中及调查后阶段的执行情况, 例如报告调查时间、无应答指标、权重计算程序、干预措施、计划调整等。这些文件的系统化和集中化制作、格式化处理与归档是在管理阶段统筹完成, 在复杂网络调查中尤为重要。

尽早形成书面文件很有帮助, 因为它可以明确流程 (如样本选择)、提升准备

质量（如编码或加权方法）、帮助解决执行问题（如调查干预）、厘清责任（例如在延误或安全问题的情况下）、加强研究者对委托方或发起方的承诺以及减少调查后报告撰写所需的时间，因为调查后阶段研究者通常更为繁忙。

为高效准备此类文件，建议在核心调查过程初期就制定相应计划，最好纳入综合管理计划。最低要求是提前规划最终报告中相应的方法学章节及相关附录的索引与内容。这能够更好地管理文件准备工作，同时也能提前说明哪些程序需以独立文件形式进行更正式的阐述，对招募计划、加权方法和安全政策来说尤其如此。此外，明确各项活动的成果、截止时间与责任人，也有助于提升管理规范性。

建立一套系统同样具有价值，最好是在线系统。这可以让研究人员或研究团队（以及潜在委托方或发起方）持续记录整个过程中的错误、经验教训、改革方式和创新建议，以改善未来网络调查项目的过程。

除上述标准管理及抽样、模式、问卷、技术准备外，预调查步骤还需考虑其他广泛问题：即专项公关活动、特殊技术要求（如特殊存档程序）、额外安全监测、对法律方面的持续关注（如处理法律投诉）、与行政有关的外部要求（如记录具体时间表、频繁和详细的管理或统计报告）以及与外部流程的整合（如客户满意度调查可能与销售和直接营销活动互动）等。从形式上看，本阶段需聚焦这些活动的计划、准备、记录、执行、协调与报告。

管理及其他准备活动的结构和形式因网络调查项目不同而有很大差异。尽管如此，本书仍可提炼预调查阶段需关注的关键问题，以便在该阶段应用。

- 网络问卷是否经过测试？能否在所有主流浏览器、设备和服务器上正常运行，且适配关键受访者群体？
- 是否已经分配资源和责任来定期监测现场？是否明确规定了相应的监测频率（每天、每周或每月）？
- 如果在调查阶段进行初步分析和自动报告，是否配备相应资源、规划及定期监测机制？
- 所有重要管理工作是否均有控制机制？例如，质量控制工作（如输出需经系统批准）、沟通工作（如明确谁负责回应有关调查的某些外部询问）、方法论相关工作（如测试和评估所有重要的方法学方案）和技术保障工作（如安全程序）是否均有预设流程与规则？
- 是否有书面的招募和无应答策略？
- 方法论、流程及调查后处理步骤（编辑、估算、加权、编码、报告、传播、输出、

归档)是否有明确规范?

- 针对收集方法学文件、准备和报告,以及写作、格式化和归档,是否已制定计划(明确责任人、工作内容、实施方法和完成时间)?
- 网络调查软件是否支持所需输出格式(如并行数据)?
- 数据收集的每一步是否均有归档与追踪(以便能够重现每份问卷的答案证明),以及是否每日定期备份?
- 是否能完全控制网络问卷的激活和停用,尤其是在无限制的网络调查中?
- 在网络调查软件托管服务(SaaS)的情况下,是否有紧急联系信息与支持机制?
- 是否有应急计划(风险管理)来应对潜在的技术问题(如服务器)或在调查中发现的方法错误(如一些不恰当的措辞),以及其他一些可能出现的问题,如法律问题或由某个受访者引发的公共丑闻(如一些隐私问题)?
- 是否有危机管理计划以应对团队或组织的意外变动?
- 如果委托分包商执行网络调查,或为特定委托方/发起方开展调查,责任划分是否明确、详细且有书面记录?
- 如果数据质量(如应答率)远低于预期,是否有应对方案?
- 若发现时间与资源严重不足,能否调整正式调查或调查后阶段的计划?是否有程序来评估研究人员是否仍能以现有资源实现目标,或者是否需要追加资源或调整目标?
- 在正式调查阶段严重拖延的情况下,能否在调查结束前开始进行实质性分析?
- 能否确认核心网络调查过程与项目的整体管理协同良好,特别是与调查后阶段的活动(包括高级统计分析、处理、传播和报告撰写)的协同推进?

如前所述,除了上述专门与网络调查有关的讨论,本书第6.2节将简要补充项目管理的相关内容。

注 释

1 <http://www.ris.org>

2 <http://fivethirtyeight.blogs.nytimes.com/2012/11/10/>

3 School uniform survey in Slovenia 2014, presentation of results: <http://www.websm.org/db/12/17584/>

4 Current Population Survey (<http://www.census.gov/hhes/computer/>).

- 5 Eurostat: http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/database
- 6 <http://www.natcen.ac.uk/genpopweb/>
- 7 <http://webdatanet.cbs.dk/>
- 8 <http://sqp.upf.edu/>
- 9 <http://wwwn.cdc.gov/qbank>
- 10 <http://survey.net.ac.uk/sqb>
- 11 <http://www.europeansocialsurvey.org/>
- 12 <http://www.websm.org/db/12/17443/>
- 13 https://www.surveymonkey.com/blog/en/blog/2010/12/08/survey_questions_and_completion_rates/
- 14 <http://question-science.blogspot.com/2012/07/how-to-calculate-length-of-survey.html>
- 15 <http://www.blaise.com/Tools>
- 16 For example, the limitations for Gmail at <https://support.google.com/a/answer/166852?hl=en>
- 17 For example, mailmerge script for Gmail: <http://www.labnol.org/software/mail-merge-with-gmail/13289/>
- 18 This section on paradata draws partially from Callegaro (2013b).
- 19 Written by Dirk Heerwegh, <https://perswww.kuleuven.be/~u0034437/public/csp.htm>
- 20 Available on Lars Kaczmirek's homepage at <http://www.kaczmirek.de/ucsp/ucsp.html>
- 21 <http://www.ieee-security.org/TC/SPW2014/>
- 22 1KA at the University of Ljubljana, <http://1ka.si>
- 23 <http://www.oecd.org/std/leading-indicators/31837055.pdf>.
- 24 Kozina, Lea, and, Čertanec, Barbara. 2014. Presentation of the Business Trends Survey. Ljubljana: Statistical Office of the Republic of Slovenia.

正式调查

若预调查阶段的准备充分，研究者在正式调查阶段的工作量会相对减少。这一点在测量阶段（第 3.2 节）尤为明显，因为网络调查是自我管理模式，即研究者或访员缺失，受访者自行填写网络问卷。不过，测量与招募（第 3.1 节）以及实地调查工作的处理与监控（第 3.3 节）密切相关，而在这些环节中，研究者需全程参与。

正如最初在第 1.2 节中所概述的，本书的论述方式聚焦研究者参与的阶段。因此，本章只探讨研究者在正式调查阶段中进行的实施类活动，相关准备工作已在预调查的各环节详细论述，研究者需在这些环节中提前完成探索。基于此，本章篇幅较短，其核心是整合预调查阶段准备工作的成果，同时重点呈现关键正式调查活动与实施细节。

3.1 招募

招募是网络调查过程中联系单位并邀请其参与调查的阶段。它是模式阐释（第 2.1 节）、样本选择（第 2.2 节）、技术准备（第 2.4 节）和无应答策略（第 2.5 节）的直接结果，同时招募工作还间接依赖于问卷编制（第 2.3 节）和综合管理（第 2.6 节）。因此，招募工作在很大程度上依赖于在本书已经详细讨论过的预调查阶段的问题，此处不再赘述，默认读者已对此类内容充分了解。

招募活动需遵循无应答策略及相应的招募实施计划（第 2.5.8 节和第 2.6 节），计划中已明确规定邀请、奖励、联系、宣传等所有相关要素。在正式调查中，研究者需执行这些预设程序，并在实施过程中灵活调整，例如：（a）启动预先准备的补充样本以扩大样本量，（b）强化或优化在线招募与推广的特定环节，（c）微调时间安排（如下次提醒的发送时机），（d）调整提醒次数（如是否追加一次提醒）和（e）完善邀请格式（如电子邮件中或横幅广告中的文字）。除此之外，没有太多要补充的内容，除非考虑先进的交互式调查设计（第 6.1.6 节）。因此，下文仅提供若干实施层面的实操建议。

在基于名单的网络调查中，招募工作通常依赖电子邮件邀请，这一点已在技术准备（第 2.4.1 节）和无应答策略（第 2.5.8 节）中介绍过。鉴于其实际操作的重要性，此处再次说明：使用研究者自身的电子邮件（组织邮箱或一些主流电子邮件服务，例如雅虎、Gmail、MS Outlook 等通用邮箱）还是网络调查软件系统的专用邮箱，是实践中常见的两难选择。在某些特定情况下我们允许使用前者，但通常优先选择集成于网络调查软件的专用电子邮件系统。因为专用电子邮件系统可以更好地处理垃圾邮件，也可以促进招募活动。例如，自动调度的高级功能可以控制发送电子邮件的时间（如根据时区控制）。

无论使用何种邮箱，建议在发送邀请前进行最终检查。检查重点是确认收件人信息，以及邀请内容在各类电子邮件系统和设备中的显示效果。同时，需重新核查邮件内容，避免因错别字或 URL 链接错误造成不良影响。

在非基于名单的网络调查中，招募实际上是在抽样阶段完成，因为样本的自我选择本身就构成了招募过程。此处需要强调网络调查问卷中介绍页的重要性。当涉及基于名单的网络调查时，介绍页被视为非问题页（第 2.3.3 节）。而在非基于名单的网络调查中（例如在横幅招募的情况下），介绍页面更为重要（另见抽样，第 2.2.3 节），因为它是向潜在受访者传达基本信息的主要场所，所以介绍页需参照邀请函的格式规范（第 2.5.7 节）进行设计。在这里，即便技术测试和实地测试均正常通过，研究者仍需重新检查邀请函、URL 链接及问卷在主流浏览器（如 Internet Explorer、Chrome、Firefox）中的实际展示效果。

此外，如第 2.5.4 节所述，调查的一般宣传与推广可在实地调查阶段开展，因此在条件允许且合理的情况下，应规划并实施此类推广活动。当然，在相关媒体上的临时（正面）曝光也同样有益。

3.2 测量阶段

一旦受访者决定参与调查,就会进入网络问卷的实际填写阶段,这正是测量阶段的主要环节。与其他阶段相比,受访者完全主导了这一自我管理的测量环节。研究者只是通过网络问卷设计间接地与受访者进行互动,而网络问卷的内容和格式已经固定。相应地,本节只关注与受访者直接相关的内容,而研究者在测量过程中的潜在干预将在后续处理与监控(*processing and monitoring*)等部分展开讨论。

测量的成功与否,取决于前期多个环节的质量,尤其是问卷编制(第2.3节)和招募(第3.1节)。如果前期准备充分,受访者则会顺利完成填写,并保持必要的参与度,提供高质量的应答,不会因为过度思考、疑惑、困惑、重读、不确定作答方向或反复寻找解释而中途停顿。

无应答、低动机和低参与度等在测量阶段会表现为作答中断、满足作答和其他数据质量问题,这些问题在关于问卷编制(第2.3节)和无应答(第2.5节)的讨论中都已得到深入研究。此外,测量阶段的具体关注点可能来自于受访者单方的各种中断。当这些中断与受访者同时或相继进行的其他一些活动有关时,我们称之为多任务处理(*multitasking*)。这可能与媒体类活动有关,要么是在同一设备的屏幕上(如电子邮件、网上冲浪等),要么是在另一设备上(也称为第二屏幕,如电视),以及各种非媒体类活动(如线下社交、饮食活动、个人事务干扰等)。

网络调查中的中断和多任务操作难以直接测量,因为这需要用摄像机记录受访者的实时状态。实际研究中,我们可以在问卷调查结束时询问受访者是否暂停或参与了其他活动。此外,并行数据可以用来核查受访者是否离开调查窗口(例如用户切换到另一个窗口或程序)或长时间未操作。需要说明的是,此处讨论的并非作答过程中故意长时间中断,而是同一问卷会话中出现的各类短暂延迟和中断。对于长时中断,研究者应该向受访者提供“暂停并稍后继续”的功能选项(第2.4.1节)。

这些中断与多任务行为的检测,在一定程度上与应答时间和延迟相关(第2.4.2节)。然而,关于多任务处理和较长应答时间两者关系的直接经验证据是有限的,因为关于应答时间的研究,重点通常是检测由矛盾态度、复杂认知加工、可用性问题、知识匮乏或参与度不足导致的超时或延迟(Olson & Parkhurst, 2013),而不是中断和多任务处理。尽管如此,较长的应答时间可能暗示存在中断或多任务行为,也可能反映受访者的异常或特定行为模式。研究者会采用多种标准作为基准

(Beckers et al., 2011; Callegaro et al., 2009a; Heerwegh, 2005; Stieger & Reips, 2010), 然而, 这些标准往往带有主观性, 且与具体调查场景相关。尽管如此, 受访者在填写问卷过程中的某些多任务处理行为, 在应答时间上几乎无体现(例如, 一些年轻的受访者能一边成功填写问卷, 一边同步聊天)。如前所述, 在某种程度上, 检测同一屏幕媒体多任务的另一种方法是, 观察受访者是否出现从调查窗口向其他窗口或浏览器标签的切换操作(Callegaro, 2013b)。

在网络调查中, 至少中断过一次测量的受访者, 其比例估计取决于具体调查类型和多任务测量方法。针对学生的调查显示, 这一比例从百分之几(Ainsaar et al., 2013; Stieger & Reips, 2010)到大约10%(Beckers et al., 2011; Stieger & Reips, 2010), 甚至高达50%(Heerwegh, 2005, p. 142; Vehovar, Slavec, & Sendelbah, 2014)。尽管50%这一数字看似较高, 但需指出的是, 电话调查中的多任务处理率与之相近(C. Kennedy, 2010; Lavrakas, Tompson, & Benford, 2010), 且均发现多任务处理与满足作答相关。事实上, 我们有理由相信, 网络调查中多任务处理更为普遍, 因为电子邮件、聊天、社交网络、电视、游戏等各类选项在同一屏幕上更易获取。

目前, 关于网络调查中中断与多任务行为的检测方法、成因及潜在影响的研究仍较为匮乏, 针对这些行为的预防、干预或纠正措施也有待探索。这一领域亟需进一步研究, 尤其是随着便携式调查设备的普及, 相关问题将更为突出。

3.3 处理与监控

通过上述内容可以看到, 招募与测量阶段主要是执行预调查阶段确定的招募计划与问卷设计。在正式调查阶段, 研究者所需开展的其他行动虽少, 但仍然非常重要。一方面, 它们与一些预先制定的策略、计划和时间表有关, 如激活/停用问卷(第3.3.1节)。另一方面, 不可预见的问题会引发具体帮助请求(第3.3.2节)和干预措施(第3.3.3节)。所有这些都需要对实际调查过程进行仔细监控(第3.3.4节)。

3.3.1 激活、软启动、全面启动和停用

实地调查工作是按照预先制定的计划进行的, 该计划已在预调查的综合管理阶段(第2.6节)完成制定和整合。通常, 调查始于激活问卷, 使受访者有权限访问网络问卷, 同时确保受访者的作答数据能被存储至正式数据文件中。激活问卷后立

刻启动招募（如发送邀请、发布访问链接），这通常需要研究者发送一些手动操作指令。值得注意的是，事先要从数据文件中剔除或明确标记测试项目，避免后续无法将其与真正的测验项目区分开。这对并行数据分析特别不利，因为为使受访者更快地完成，测试条目可能不使用分支，导致数据缺乏实质一致性。另一种解决方案是创建问卷副本，从全新的调查会话开始收集正式数据。

在全面激活和启动调查之前，建议先进行软启动（或试启动），即仅向小部分受访者开放参与渠道。在基于名单的样本中，这通常意味着只向部分样本发送邀请；而在非基于名单的样本中，这意味着限制推广范围（例如仅在少数网站上发布招募横幅）。软启动可以通过收集并分析初期应答数据，验证数据质量、排查问卷设计和招募流程等环节的潜在问题。此外，软启动可帮助估算问卷实际完成时长，为全面启动的节奏调控提供参考。显然，所有这些特征都类似于问卷测试中的预研究（第 2.3.5 节）；两者关键的区别在于，预研究的应答并不纳入真正的数据文件和统计分析，而软启动的数据则可作为正式数据的一部分。另一方面，“软启动”的本质是在没有足够时间进行预研究的情况下，替代或补充预研究；而这种情况经常发生在市场营销研究中。即便预研究顺利完成，仍建议在全面启动前通过软启动，再次核验流程与初步结果。

测量阶段结束后，问卷将按预设日期或停用规则（*stopping rules*）终止访问。停用规则类似于序列抽样（*sequential sampling*）的概念，即持续纳入样本直至满足特定统计标准（例如某部分中的最小单位数）。在网络调查中，这通常经过限制受访者的数量来实现，使用各种类型的配额设置来限制整体或某些部分的受访者数量。在符合停止规则后，调查自动停用。但如果研究者在自动停止调查前，已经知道了达到停止规则的条件，那就更好了。因为除了停用本身，关闭调查还涉及其他一些招募工作（例如从网上删除调查问卷的 URL，停止推广活动等）。

应该避免让愿意参与甚至迫切希望参与调查的受访者收到调查已经关闭的通知。尤其在自助式研究中，受访者经常遇到已经停用的网络调查邀请。这不仅会引发不满，更涉嫌违背研究伦理。如果无法删除问卷链接（例如在某个论坛发布了 URL），可以通过更改链接处的内容（即用其他网页替换问卷），或者设置自动重定向，使这个 URL 指向说明页、结果信息页、研究的后续行动页等。这些都要以网络调查软件的功能支持为前提。更糟糕的情况是，研究者未及时停用问卷或疏于管理，可能导致问卷实际已关闭但邀请链接仍有效，使得受访者的作答被无效存储。这种行为不仅不道德，还会浪费受访者的精力，因为研究者不会使用它们。由于这个原因，大多数网络调查软件都默认设置活动时长（例如一个或三个月），需研究

者手动操作才能延长，以此规避此类问题。

3.3.2 提供帮助

本书已经在第 2.3.3 节中讨论了关于问卷布局的介绍、定义和帮助内容。此外，在正式调查阶段，我们还应该确保提供实时帮助，以便对受访者的问题进行反馈。受访者可能会遇到各种技术问题，如网络连接速度慢、被禁止访问或问卷问题。在大多数情况下，研究者不仅要关注电子邮件邀请和调查问卷未经过充分预测试的问题，还要关注错误的 cookie 或 IP 设置等。然而，有时可能是受访者自身设备或操作导致的技术障碍。无论如何，上述问题都可能引发受访者的疑惑，需要研究者提供指导和解释。我们需要尽可能地避免受访者提出此类问题。例如，对于某些最低限度的技术要求（如 JavaScript、Flash 等），可自动向受访者推送说明与操作指引，从源头避免疑问产生。

至少需要在邀请函、介绍页或者网络问卷网页的页脚中，向受访者提供明确的联系信息，包括电子邮件地址或反馈表。有时，在问卷的每个页面设置一些带有常见问题（FAQ）和联系信息的帮助按钮也颇为实用。添加电话联系方式虽能增强调查的合法性与可信度，但需配备相应资源提供支持，需在资源投入与避免冗余沟通之间找到平衡。

受访者的反馈可能还包括各类顾虑与诉求：如希望获取填写完毕的问卷存档、要求知晓研究结果、申请从邀请名单中退出，或者希望解释他们被明确拒绝的原因等。受访者也会对调查提出批评，涉及技术层面（如字体大小）、方法层面（如没有进度指示器、没有返回按钮、指示不明确、问题质量低等）以及实质性问题（如“不充分”的问题、缺乏开放式问题来阐述自己的想法等）。对于在线样本库以外的网络调查中的突出主题，我们预计最多有 1% 的受访者会提供电子邮件回复（例如询问结果或评论主题）。受访者也会对调查提出抱怨和批评（例如，极为突出和有争议的主题会引发更多的反馈），特别是取决于是否明确提供和直观展示反馈选项。电话反馈通常更少，约 0.1% 的受访者可能致电。为减少过度反馈，我们应该提供充分的在线帮助、相关的 FAQ 和详尽的“致谢”页面，尤其是在调查末尾设置开放性评论题，征集受访者的任何意见。尽管如此，在大样本调查或主题敏感的情况下，研究者仍需做好处理更多反馈的准备。

当有外部网络调查软件服务或分包研究机构参与时，建议区分技术性问题与实质性问题，协调各方的帮助与反馈策略。

无论何种情况，一旦研究者在邀请函中承诺提供帮助与互动，就应建立机制确保专业、及时的回应。尤其在预期有电话反馈时，更需做好准备。

尽管为受访者提供一些互动式的帮助是必不可少的，但研究者的重点应该是编制好的问卷，并在邀请函和介绍页中充分沟通，而不是为设计不良的功能进行解释、修复、澄清和道歉。

3.3.3 实地调查工作的干预措施

本书所说的实地调查工作干预是指因在正式调查阶段发生意外事件，研究者所采取的应对行动。因此，此处不涉及实施过程中预先计划的调整（*implementation variations*），如预设的招募方案变动等，而是聚焦于研究者为提高数据质量（如提高应答率）或规避调查项目重大风险（如服务器遭攻击）所开展的额外活动。

常见干预措施多与补充招募相关（如延长实地调查工作、追加邀请、增加推广活动等），这些均不在初始招募计划的预期范围内。在更严峻的情况下，甚至可能需要调整整体研究设计（如转为混合调查模式）。

由于涉嫌违法行为（如未经同意使用电子邮件、侵犯隐私、垃圾邮件证据、非法内容等），也会被要求关闭调查或删除部分内容。此类诉求也可能基于实质性、方法论或伦理层面的考量，提出者可能来自受访者、目标人群、普通公众、方法学家、学科专家或利益相关者。

更频繁出现的问题与某些方法问题相关。例如，在实地调查过程中的严重分支错误或者问题措辞错误。对此，研究者需明确在正式调查阶段修改问卷的规范流程与技术操作。

网络调查软件或服务器的技术故障，可能导致网络问卷无法使用，这是需要紧急干预的典型情况。研究者还需针对受访者的技术问题干预和对问卷进行调整，例如浏览器兼容性不足（通过对浏览器提供支持）或通过与组织协商解决因安全政策导致的访问限制问题。

毋庸置疑，多数需干预的问题可通过预调查的质量管控措施预防，因此在专业筹备的网络调查中，上述问题极为罕见。尽管如此，研究者仍需意识到这些问题发生的可能性，并做好应对准备。

此处讨论仅限于意外事件引发的干预，这需要研究者采取一些额外的行动，而其他各种实地工作的干预可以事先预见并自动执行。

因此，本部分讨论不仅限于问卷应答流程中默认设置外的变动。这有可能意味

着,我们要对下列受访者进行具体干预,如应答过快(超速作答)、过慢(如存在认知障碍)或数据质量异常(如满足作答)等情况。原则上,在测量阶段,研究者可以在线检查多项应答质量指标(例如项目无应答、不一致、有效性、可靠性、满意度指标),进而为受访者提供帮助、给予激励,引导其跳转至简化版问卷,或推迟部分作答安排至更合适的时间。但此类在线干预因具有侵入性且可能引发“监控效应”(‘big brother’ effect),反而可能适得其反。

同样地,本部分不涉及与交互式实地设计相关的具体干预措施(也见第2.3.4节中关于参与度的讨论),相关内容将在更广泛的方法论框架下探讨(第6.1.5节)。

3.3.4 实地调查阶段的监控

“没有过程质量,就没有结果质量”(Lyberg, 2012)。相应地,为了实现网络调查项目的目标并产出高质量结果,我们必须保障调查过程的质量,而监控正是这一过程的核心环节。从形式上看,调查过程的监控与应答状态密切相关(第2.5.2节),并嵌入在综合管理中(见第2.6节)。

如果网络调查软件将实地调查过程中的重要发现以关键指标的形式提供给研究者,将极大提升效率。这些与过程表现相关的指标被称为关键绩效指标(key performance indicators, KPI),需要在预调查阶段尽早确定。在实地调查过程中,网络调查软件通常会在一个页面上汇总展示这些指标,该页面也被称为仪表盘页面(dashboard page)。相应的分析与图表虽在形式上类似网站流量分析工具,但具备诸多调查专属功能。网络调查软件自动显示的指标千差万别,研究者不仅可依托这些默认指标,还可进行额外计算。关于实用的仪表盘示例,可参见第3章补充资料“实地调查监控与诊断案例”(Examples of fieldwork monitoring and diagnostics, <http://websm.org/ch3>)。

基本关键绩效指标用于实时洞察应答过程,包括反映受访者技术应答状态的结构指标和非应答指标,如单位应答率、项目应答率、受访者应答完整率和中断率(见2.5.2节)。尤为重要的是,相关图表需能实时呈现应答动态(如第1.3.4节中的图1.2)。

此外,许多其他的关键绩效指标也很有实用价值。首先,可以根据应答状态、时间间隔(月、周、日、小时)对上述应答指标进行分类,或者根据语言、目标群体或内部分支(如随机组)对问卷最终版进行单独分类。对关键控制变量(如性别、年龄、地区)或并行数据相关变量(如设备类型、显示器尺寸、计算机操作系统、浏览器类型)进行细分也很有必要。其次,获得AAPOR标准应答率的计算结

果（AAPOR，2011；另见第 2.5.2 节）或至少其近似值，也是有帮助的。

以下各类额外洞察可能颇具价值：

- 受访者参与调查的 URL 来源结构以及基于 IP 地址的地理位置信息，对于非基于名单的网络调查特别有价值。
- 在基于名单的网络调查中，将电子邮件邀请集成到网络调查软件系统中。我们希望对发送的电子邮件进行详细统计，包括发送内容、时间及对象。除整体应答状态分析外，还需支持对单个受访者的追踪（如记录联系历史数据）。
- 监控问卷长度及每页作答时间时，计算平均数需要进行一定的删减以去除极端值（如 5% 的极端情况）。将实际时长与预期对比，可及时发现偏差与潜在问题。
- 监控各类提示信息的出现频率（第 2.3.3 节）。

理想情况下，研究者希望能实时掌握受访者的应答质量细节，如无效应答、欺诈行为、极端超速作答、满足作答和其他有问题的数据质量等。但这类信息的生成依赖复杂的统计程序，通常这些程序仅在数据准备阶段运行（见第 4.1 节）。尽管如此，任何质量诊断方法（如直线作答比例等）都具有参考价值。未来，网络调查软件将集成越来越多的此类诊断功能，其可用性将成为选择软件的重要标准。

除上述活动外，监控的另一重要内容是在数据收集过程中对问卷所有变量进行基础汇总（如频率分布、描述性统计），以核查数据分布与潜在偏差。

应该在特定的背景下观察上述指标。理想流程是先确定需监控的活动（如受访者参与度），再选择对应指标（如应答率）设定评价标准（如 40% 的基准值），并明确纠正措施（如再次发送提醒）。

下面举例说明 KPI 在网络调查中的应用以及相应的实地工作干预。

- 当两次提醒后的应答率低于预期基准时（例如 20% 而不是 40%），可以考虑另一种提醒策略。
- 当问卷的中断率高于预期基准时（如实际 40% 而非预期的 20%），要重新检查问卷，尤其关注中断高发页面。
- 当某类提示信息出现频率过高时（如 40% 的受访者输入了错误数字），要考虑改进问题说明。
- 当一定时间段内（如一天、几个小时）没有应答时，需检查服务器性能及技术故障。
- 当受访者的社会人口结构偏离预期时（例如，女性太多，占比 80% 而不是

50%)，调整招募方式（例如，优化推广渠道、额外联系、差异化激励）。当转发（例如，我们预计 50% 的单位来自 Facebook，但实际只有 20%）或者 IP（预计 30% 来自某些城市，但只有 3%）偏离了预期时，也是如此。

需注意区分正式调查阶段的监控与调查后阶段的其他诊断工作（第 4.1 节）。在正式调查阶段，需逐一核查每个受访者，确定其状态（完整、部分、不可用）和可能的补救措施（如插补）。详见第 3 章补充资料“实地调查监控与诊断案例”（*Examples of fieldwork monitoring and diagnostics*, <http://websm.org/ch3>）。

最后，需要补充说明的是，在混合模式调查、在线样本库、纵向调查和复杂调查情况下，或当交互式实地策略替代固定招募计划时，监控工作会更为复杂。由于这类情况超出了基础调查模式范畴，本书仅在第 6.1 节简要提及。

调查后阶段

正式调查阶段一旦结束，便进入调查后的工作环节。除非存在某些特殊情况，否则一般不在调查后阶段收集任何额外的应答数据。

根据核心网络调查过程的结构（图 1.1，第 1.2 节），调查后阶段的章节将涵盖以下阶段内容：数据准备（第 4.1 节），初步结果（第 4.2 节）以及数据导出与文档化（第 4.3 节）。

将调查后阶段与高级统计分析、处理和价值转化（不属于核心网络调查过程）区分出来的标准是：一是关注方法论问题而不是实质性问题，二是在网络调查软件中实现 ICT 支持的整合。因此，部分与调查后有关的活动超出了核心网络调查过程的范畴，因为它们与实质性内容有关（如报告撰写），或者因为涉及外部软件支持（如统计分析）。

尽管本章重点围绕网络调查相关方面展开（为保证论述的完整性），但也会介绍一些重要的正式调查后活动。这些活动具有一般性，与具体的调查模式无关。因此，我们经常参考标准的调查方法学教科书，如 Groves 等（2009）、Biemer 和 Lyberg（2003）和 Iarossi（2006）。

与正式调查类似，调查后阶段的活动需要在预调查阶段进行规划和准备（第 2.6 节）。然而，在预调查阶段，为正式调查相关任务所做的准备需要投入大量资源，而这些任务在正式调查阶段仅需简单执行即可（例如无应答策略）。调查后的情况则恰好相反，因为与调查后有关的任务主要是在正式调查结束后开展。与实施调查相比，为调查后阶段所做的准备工作所需资源相对较少。

此外,一般而言调查后阶段的概念与调查后的调整、数据收集后处理(Groves et al., 2009)和数据处理阶段(Biemer & Lyberg, 2003)等概念的内涵大致重叠。

4.1 数据准备

数据准备涉及说明应答状态(第4.1.1节)、编辑和验证(第4.1.2节)、插补(第4.1.3节)、加权(第4.1.4节),以及编码和重新编码(第4.1.5节)。原始输入是一个既包含网络调查回复,也包含其他主要数据(并行数据、元数据、辅助数据)的数据文件,而输出则是最终数据库(*final database*),用于统计分析、初步结果、数据导出、存档和进一步利用。

数据准备的质量取决于预调查和正式调查阶段中的各项活动。计划疏漏、环节缺失、问卷设计失误、推进仓促、资源匮乏、管理不当、实地监控不力等,均会导致严重后果。它们会在多个方面造成数据质量低下。在数据准备阶段,我们只能一定程度上解决其中的部分问题。

其中尤为关键的是技术性数据库准备(第2.4.1节)和问卷编制(第2.3节),具体包括:问题措辞的精准性、应答选项的合理选择、各类问题的恰当布局(如为开放式问题预留充足空间)、分支设置的准确性、缺失值与非实质性应答的妥善处理、实时验证的合理性,以及定义、解释、帮助信息的清晰阐述。

在整个调查后阶段,需要在整体管理范畴考虑数据准备活动(2.6节)。其中,规划与安排至关重要,因为除了问卷准备,数据准备通常是核心网络调查过程中耗时最长的阶段。

4.1.1 确定应答状态

第一项活动是确立并核查单位及单个条目的状态(即受访者对具体项目的应答情况)。这是后续所有数据准备工作的基础,包括计算应答率和诊断数据质量。从广义上讲,这类工作属于数据编辑范畴;但由于此处的重点是确认数据的准确状态,而非开展检查、评估或编辑,因此我们将这部分内容单独处理。

4.1.1.1 样本单位的合格性、完整性状态和质量状况

这里主要关注以下问题:不合格样本单位、重复提交、单位无应答、应答完整

性状态和数据质量诊断。

抽样相关章节已对不合格单位 (*ineligible units*) 进行了定义 (第 2.2 节), 理论上这类单位应在抽样阶段即被排除, 但在实际操作中, 部分单位的合格性需在正式调查阶段通过特定问题核实 (例如投票研究中需确认是否为外国公民), 或确认部分单位搬迁或消亡等情况。因此, 应该重新检查单位是否合格。如果有些单位不属于目标人群, 我们应该将其从最终的数据库中排除或做明确标记。

重复提交 (*Multiple submissions*) 带来更多的挑战, 指同一单位提交多次应答或多次启动应答流程的情况。本书曾在数据库准备中讨论过这个问题 (第 2.4.1 节)。如果掌握了受访者的唯一标识 (ID 代码), 处理会相对简单。当没有控制操作防止受访者重复时, 最终可能会出现具有相同 ID 代码的同一受访者的多项记录。此时, 即便存在重复提交, 只需识别同一 ID 对应的多条记录, 核查后保留质量最高的应答即可。当缺乏 ID 代码时, 问题就更加复杂, 比如在非基于名单的网络调查中。这种情况下, 最多可以借助的是 cookies 和 IP 号码, 但是这两种方法都具有局限性, 且已在第 2.4.1 节中讨论过。另外, 我们可以使用统计匹配算法来发现重复作答的受访者。这种算法基于一定概率识别重复的单位。通常, 通过社会人口统计学信息 (例如相同的年龄、性别等)、关键的实质性回答和并行数据 (如设备、浏览器等) 来识别重复单位。需说明的是, 此处主要针对受访者因操作误解 (如未确认应答是否保存) 导致的重复提交。如果是受访者恶意伪造, 则属于不同性质的问题, 这将在后续内容展开探讨。

第 2.5.2 节已经概述了单位无应答 (*Unit nonresponse*) 识别, 其中详细说明了基于名单的、有电子邮件邀请的样本无应答处理流程。第 2.4.1 节也提及与电子邮件设置相关的技术准备要点。实践中需制定明确规则, 将没有点击邮件中问卷 URL 的受访者 (例如没有返回、邮件发送错误、自动回复等) 转化为不合格、无应答和未知状态等。最好能更精确地将它们转化为 AAPOR 标准处置代码 (AAPOR, 2011)。

完整性状态 (*Completeness statuses*) 的界定存在一定复杂性。第 2.5.2 节重点讨论了这个问题, 包括如何识别完整应答、部分应答及无效单位。并行数据可作为重要补充依据, 例如每页作答时间、鼠标移动轨迹、长时间无操作记录、应答内容修改痕迹、过度点击行为及项目缺失情况 (Stieger & Reips, 2010)。这些并行数据可以帮助评估状态。举例而言, 若受访者完成问卷的时间远低于中位数水平, 可判定为无效单位。

应答质量诊断 (*Quality response diagnostics*) 需结合应答数据与并行数据, 从

而能够计算单位的各种质量指标,如项目无应答的百分比、作答模式一致性(如直线作答)、数据一致性水平、多任务处理痕迹等。这类诊断功能可集成于网络调查软件,也可通过专门的外部服务实现(尤其适用于在线样本库调查)。

理想情况是,我们使用所有可用的数据,涵盖合格、完整性状态及数据质量等方面,建立一套精确且统一的方法,为数据文件中的每个单位设定准确状态或最佳近似状态。部分状态可在正式调查阶段自动生成,其余则需在调查后阶段通过专用程序确定。无论何种情况,每个单位至少需设定通用状态,再进一步细分为合格状态、应答状态、中断状态和完整性状态。在先进的网络调查中,每个单位也应该至少有一个质量状态,再进一步细分为与上述各种数据质量维度相关的子项。

4.1.1.2 缺失值和非实质性应答

除明确单位状态外,还需识别每个具体应答的状态。除预期的实质性应答外,数据中还存在各类缺失值和非实质性应答,有时也被称为无效或非有效应答(*invalid or non-valid responses*)。对这些应答的部分编码是由网络调查软件在正式调查阶段自动完成的,而其余部分则需要在调查后完成。无论如何,我们都必须认真检查和核对这些状态的设置准确性。

(a) 项目无应答状态与“合理的非实质性回答”二者的编码应不同,比如(b)“不知道”,(c)“倾向于不回答或拒绝回答”,(d)“不相关”(或任何其他非实质性回答)。当没有呈现给受访者某个问题时,需要根据具体原因进行不同的编码:(e)由于问卷设计(如分支)而出现的“不适用”状态,与项目无应答存在本质差异;(f)因前期作答中断未触及该问题;(g)在某个单位开始回答问卷后的新增问题;或由于(h)问卷完全空白。

以吸烟状况调查为例:若先筛查吸烟状态,仅向吸烟者询问昨天抽了多少支烟,那么这个问题的答案可以是任何数量的香烟(1、2、3、4……),包括“0”,这应该与“没有答案”明确区分开,后者意味着项目无应答(a)。我们也可以提供选项“拒绝回答”(c)或者“不知道”(b)。“不适用”状态(e)应自动分配到第一个问题所筛选出的非吸烟者。

此外,我们还应保留另一组状态来标记因质量较低而被删除的值(如不一致、欺诈、满足作答)。理想情况下,数据文件中的每个单位、每个项目、每个输入都应该关联一个相关的元变量(meta-variable),用于标记描述和处理特征,如变量的类型和来源、单位的状态以及最终对单位(尤其是单个项目)的编辑、修改或插补记录。这一设置也符合创建特殊数据填充(插补)标志的要求(Heeringa, West,

& Berglund, 2010), 便于用户区分原始值与插补值。但在实际操作中, 最终数据库通常为每个项目仅保留一维变量。除正值之外, 负值通常用于表示缺失值(如 -1, -2, ……) 和非实质性回答(如 -99, -98, ……)。但有时实质性的反应也是负值(如温度), 则需针对此类特殊情况单独处理。

4.1.1.3 应答状态报告

只有当样本单位层面和个人项目层面的所有应答状态均设置准确后, 才能确定各类应答指标, 具体包括:(a) 报告标准应答率(AAPOR, 2011), 至少是 RR2 或 RR1, 最好涵盖所有最终处置代码;(b) 报告中断率, 最好分为介绍页中断、问卷中断和总中断, 并分析中断高发的问题环节;(c) 问卷的完整性状态(完整、部分和无效的应答, 最好根据技术层面和实质层面进行分类);(d) 项目无应答的分布情况(需结合不同变量和不同单位进行分析);在高级应用场景中, 还需包含(e) 各类其他数据质量诊断指标(例如作答速度、满足作答指标、多任务处理痕迹)。

尽管某些诊断指标可能需要人工计算, 但多数情况下, 网络调查软件的调查模块(*survey dashboards*)会自动生成这些报告;此外, 还可按正式调查时段、地区/国家(基于 IP 地址的地理位置)、转发渠道及设备类型等维度, 对关键应答状态进行更细致的分析。详见相应的补充资料, “实地调查监控与诊断案例”(*Examples of fieldwork monitoring and diagnostics*, <http://websm.org/ch3>) 及 “应答率计算实例”(*Examples of response rate calculations*, <http://websm.org/ch2>)。

4.1.2 验证和编辑

验证与编辑是指对收集到的数据进行检查, 必要时进行修正, 以最大程度保留原始测量设计的特性(Groves et al., 2009, p. 345)。换言之, 就是通过检查(验证)识别错误与不一致的条目, 并进行纠正(编辑)(Waal, Pannekoek, & Scholtus, 2011)。

在网络调查中, 部分验证可在问卷填写过程中实时执行。实时验证会产生提示信息, 引导甚至强制受访者检查并更正答案。然而, 通常很难事先预测所有可能的验证规则。无法实时执行所有验证的另一个原因是, 此举会导致许多提示信息, 从而使受访者感到厌烦, 增加中断率。下面我们从不同的角度和层面来讨论验证和编辑问题。

基本验证(*Basic validations*)是最基础的数据验证, 通常在正式调查阶段完成。

它们主要与开放式输入（文本和数字）有关，如防止错误的类型和格式（在数字框中输入文本）或范围检查（如年龄为200岁），同时涉及关键的基础一致性校验（例如12岁的受访者不应报告有投票行为），以及单一变量的简单计算验证（例如常量和问题）。这类验证必不可少，若受访者未实时纠正，需在数据准备阶段加以补充验证。

项目无应答验证（*Item nonresponse validation*）是在实地测量中最常实施的验证类型。如第2.3.3节所述，当受访者未提供答案却继续作答时，系统会自动弹出提示信息。然而，这类验证具有可选性，因为项目无应答可能不是错误、遗漏或不一致，而是自愿调查中的合理行为，因此强制受访者回答可能会引起争议。尽管如此，研究者通常倾向于通过各类提示减少项目无应答。如果数据文件中仍然存在无应答项目，将在后续通过插补法处理（详见下一节）。

高级验证（*Advanced validations*）是验证与编辑活动的核心，主要在数据准备阶段进行，通常涉及多个关联变量。这类验证需系统全面，对研究者认为重要的所有方面进行深入核查，具体标准因情况而异。例如，在商业调查中，不同项目的回答需要保持一致（例如雇员人数或营业额，在各组织单位加总时，不应超过机构总体水平）。这可以通过网络调查软件、通用统计软件（如SAS、Stata、IBM-SPSS或R）手动运行各类条件校验，或者使用专门的数据编辑软件来完成核查（Groves et al., 2009, p. 346）。

当验证发现有严重错误或不一致时，需要进行替代（编辑），用特定类型的缺失值替代（例如，如果受访者回答了不适用自身的问题），或者用一些合理的替代值替换（例如，如果报告的支出超过了报告的收入，那么收入就被替换为支出的总和）。

需注意的是，不应在验证和编辑上过度投入精力，而是要接受受访者可能存在轻微不连贯的情况。在简单的态度调查中，验证的数量通常较少，但在行为和事实类问题（如事件、支出、收入、旅行等）中，验证的数量会急剧增加。此时，过多验证会变成对资源的低效利用。在部分商业调查中，编辑活动的成本甚至可能占据调查总预算的最大份额。此外，过度验证，也被称为过度编辑，未必能显著提升数据质量（Grandquist & Kovar, 1997）。

实质性验证（*Substantive validations*）包括对所有变量的单变量统计进行细致和系统的观察：对名义变量和顺序变量进行频率统计，对等距变量和比率变量进行描述性统计（平均值、中位数、方差、最小值、最大值），通常与图形分布相结合（如直方图）。同时可核查关键细分维度（如不同性别或年龄组的平均值）。这类验证需

经验丰富的研究者结合方法学知识和实质领域知识开展，与前述的纯粹技术化执行的验证不同。实质性验证是非常重要的，它为调查数据的后续处理和使用提供了必要的合理性依据。类似的核查在测试和正式调查阶段的工作中也有用，可及时发现并采用适当的干预。

进一步的数据核查可以根据用户的需求进行调整。例如，用于数据质量保证和欺诈检测的高级筛选程序（Blasius & Thiessen, 2012），或各种数据清理程序（Osborne, 2012），如核查正态分布或异常。然而，这些通常超出了核心网络调查过程中调查后阶段的范畴，因为其结果通常不纳入最终数据库。在此阶段，我们会尽量保留原始报告数据，仅在充分理由时才进行修改。

可见，验证细节和整体编辑方式是结构化极强的议题，因此需要在预调查的数据管理阶段（第 2.6 节）精心组织、具体化和规划。

4.1.3 插补

前文已提及一种与编辑相关的特殊插补类型，其源于数据验证过程。通常而言，插补被定义为针对缺失数据（主要是项目无应答）人工赋值的常规做法。第 2.5.2 节已详细探讨了无应答的发生机制、成因、后果、偏差和预防，而本节仅聚焦于如何通过插补进行补充处理。除项目无应答之外，插补也用于其他类型的缺失值，甚至用于非实质性应答（例如“不知道”），但此类应用需额外的假设与论证支持。例如，需要证明“不知道”的应答实际可视为项目无应答的替代形式，类似于斯洛文尼亚公民投票调查（Rubin, Stern, & Vehovar, 1995）。

在考虑对无应答项目进行插补时，需先完成应答状态的界定工作（见第 4.1.1 节）。因此，无应答项目应被清晰识别，并与其他缺失值、非实质性应答及实质性应答（如零值）明确区分。同时，还需区分受访者层面的无应答与项目层面的无应答（第 2.5.2 节）。

相较于数据验证，网络调查中关于插补法的专项研究较少。虽有部分案例研究涉及网络调查数据的插补问题（Chang, Frost, Chao, & Ree, 2010），但据我们所知，目前尚未有研究深入探讨其潜在的方法学细节。本节将提供总体概述，重点讨论针对合格单位预期应答中缺失值的插补处理。

确定缺失数据及相应指标后，首先需观察缺失数据的模式。这可通过标准统计软件包实现，软件包越来越多地包含了这样的程序（例如 Stata 中的“mvpatterns”命令可显示缺失比例与模式）；部分先进的网络调查软件或专门的缺失数据处理工

具也具备此功能。

关于缺失多少数据可无需插补的问题,目前还没有达成共识。有学者提出 1% 至 2% (Heeringa et al., 2010, p. 108),但这取决于数据缺失的原因和方式,尤其是与偏差有关的主题(见第 2.5.3 节)。

默认的做法是通过采用“整列删除”来忽略缺失的数据,这意味着只有完整的案例才被用于后续分析。然而,这可能会导致严重的数据损失。例如,如果我们有 10 个变量,每个变量有 10% 的缺失值随机出现,最终仅能保留 35% 的完整单位。尤其是如果需要进行多变量分析(如回归),这种损失往往不可接受。因此,本书强烈建议避免这种简化处理,尤其是针对关键变量。

恰当的插补方法取决于缺失数据的类型(Little & Rubin, 2002),包括完全随机缺失(*missing completely at random*, MCAR),即缺失值的概率与其他任何变量无关;随机缺失(*missing at random*, MAR),即缺失值的概率可以通过其他已观测变量来解释;或者非随机缺失(*not missing at random*, NMAR),即缺失值的概率与一个(几个)未观察到的变量值或者未知的缺失数据机制有关。因此,调查中纳入丰富的相关变量至关重要,这些变量可以潜在地解释缺失数据的模式(Peytchev, 2012),并创造随机缺失条件。

上述“整列删除”程序仅在完全随机缺失的假设下才有效。然而,现实中符合该模式的缺失数据极少。因此,通过观察目标变量的缺失值在其他变量(如年龄、性别等)上的分布差异,对随机缺失进行基础检验颇具意义。如果发现显著变异,则提示可能存在随机缺失,可采用标准插补方法处理。

插补方法多样,从较简单的均值插补(直接采用应答单位的均值),到适用于随机缺失的标准方法[冷卡插补、热卡插补、最近邻法、从模型中得出预测值、期望最大化(EM)算法]。

最完善且要求最高的是多重插补法(Little & Rubin, 2002)。多重插补法中创造更多的插补值,从而获得更多的数据集。其核心优势是可以对任何变量的抽样方差进行合理估计。单一插补策略的主要缺陷是往往低估变异。使用者可能误将插补数据视为真实数据,实际上有效的样本量远小于通过插补建立的人工样本量。因此,置信区间过窄。

本章不再对各种插补方法做进一步的描述,但会建议读者参考适当的教科书(Enders, 2010; Little & Rubin, 2002; Särndal & Lundström, 2005; Waal et al., 2011)。需补充说明的是,部分网络调查软件已集成简单插补程序,而完整的插补功能可在主流统计软件包及专门工具中获取。

遗憾的是,应用统计学领域还没有开发出“研究人员按下按钮就能将所有缺失值适当地估算出来”这样的一键式插补程序。尽管如此,相关技术正逐步完善。目前,插补工作仍需投入时间、资源并具备专业能力,因此插补策略往往成为区分高级与基础网络调查的标志。自助式调查通常不进行插补。然而,现实情况是,除非调查规模极小、项目无应答可忽略不计或明确符合 MCAR 模式的调查,否则均需实施插补,并建议在专业统计指导下进行。插补错误可能会给数据质量带来严重损害。这是网络调查最关键的方面,并驳斥了网络调查可以“轻松”开展的片面认知。

减少插补资源投入的最佳方式是预防项目无应答 (de Leeuw & Hox, 2008),第 2.5.2 节已探讨相关措施。鉴于插补工作的复杂性,人们便不难理解为何许多网络调查会对所有问题设置强制提示以预防无应答,尽管这种操作可能引发受访者干扰、中断作答、满意度下降及伦理争议。

插补法也可以用于单位无应答,且常与加权程序配合使用(下文会讨论)。一个非常有趣的方法是问卷分割技术(在抽样和问卷编制阶段已经提到,见第 2.2.2 节和第 2.3.3 节)。将问卷模块随机分配给子样本,再通过统计匹配(即数据融合)算法填补调查中未接收特定模块的单位缺失数据 (Johnson et al., 2014)。

4.1.4 加权

与插补类似,加权在网络调查中的具体应用相对有限,但已有大量文献涉及其实施方式与方法学细节。早期网络调查虽能以较低成本收集大量样本,但样本往往存在偏差且缺乏代表性,这推动了加权程序的应用,以解决该问题 (Vehovar et al., 1999)。事实上,通过加权修正偏态样本并非新问题,也不仅限于网络调查。因此,本节仅做概述,并重点强调网络调查中与加权相关的实际问题。

一般而言,加权的确定包含三个主要阶段。首先是建立基本权重,它源于抽样选择设计,与抽样选择概率的倒数成正比。例如,在欧洲晴雨表调查 (*Eurobarometer surveys*, 也是欧盟民意调查) 中,为获得欧盟总体估计,每个国家大约有 1 000 个样本单位,需通过加权调整各国人口规模的巨大差异。简单来说,我们可以说 1 000 个样本中的每个单位在一个 200 万人口的国家中代表 2 000 个公民,在一个 2 000 万人口的国家中代表 20 000 个公民。

其次,计算无应答权重,以弥补单位无应答的影响。例如,若某一年龄—性别组(如 50 岁以上的女性)的应答率为 40%,我们可以应用加权调整。然后,该组所有应答单位的权重按 $1/0.4 = 2.5$ 的比例设定,这样一来,每个应答单位就代表了

原样本中的 2.5 个合格单位。

最后，将前两步得到的权重相乘，进一步调整为与已知人口参数匹配的人口权重（*population weighting*）。这样一来，样本与目标总体（如地区等）完全匹配。这为数据的实际使用者提供了便利。通常情况下，它对估计值的影响相对较小，因此有时也被称为表面加权（*cosmetic weighting*）。此处可采用多种标准方法（事后分层、多变量反复加权），以及更复杂的方法（如高级校准）。

加权默认假定理想的概率抽样，此时其作用通常较小。由于无应答、无覆盖和选择性偏差等，实际样本与理想概率样本差异增大时，加权会对估计值产生更大的影响。这在非概率样本中尤为明显；在这种情况下，我们可以借鉴抽样调查中，在非概率环境下实施概率抽样的思想（第 2.2 节）。

多数情况下，加权对估计值的影响相对有限，但仍可感知且值得实施。我们需关注加权后估计值的变化方向，因为真实值往往沿同一方向进一步偏移。多项研究证实加权最多可消除一半偏差，如 Vehovar 等（1999）的研究，以及由 Tourangeau 等人（Tourangeau, Conrad, & Couper, 2013, p. 33）进行的关于网络调查中加权的元分析研究。

网络调查中经常被提及的具体方法是倾向值加权（*propensity score weighting*, Steinmetz et al., 2014）。这个想法源于 Rosenbaum 和 Rubin（1983）的观察研究，并被广泛用于非概率抽样的网络调查。早期网络调查中，该方法已用于营销研究，解决非概率抽样、覆盖不足及无应答问题（Taylor, Bremer, Overmeyer, Siegel, & Terhanian, 2001）。后续研究虽发现其在特定场景下略优于其他备选方法（R. P. Baker et al., 2013）。但是，这种差异仍然非常小，无法彻底解决非概率样本的推断问题。

下面我们进一步介绍在网络调查实践中常见的加权情况。

- 在网络调查中，研究者所使用的样本往往无需特定基础权重，尤其是普查或简单随机抽样（SRS）。
- 非概率样本中，没有基础权重，也没有无应答权重，而是使用人口权重同时补偿选择、无应答和无覆盖。
- 加权会增大抽样方差，可通过标准化权重的方差（即标准差的平方）估算。平均权重通常为 1.0，方差增幅若超过 40%，意味着置信区间将扩大近 20%。为控制这种增幅，常在标准化后，对权重进行截断——将超过 4 或 2 的权重标准化后削减。
- 如果使用的辅助信息及细节一致，不同加权方法的结果差异通常极小。
- 纳入正确的控制变量加权（也纳入调查）是非常重要的，可以考虑从关于同一

主题的其他可靠调查中获得控制变量。

- 建议使用标准化的社会人口学问题（例如受教育程度），而非编制新问题。

Groves 等（2009）对加权进行了一般性介绍，Valliant, Daver 和 Kreuter（2013）对加权进行了专门论述，而 Bethlehem 和 Biffignandi（2012）对网络调查进行了具体概述。与缺失值插补类似，部分高级网络调查软件已集成部分加权方法，主流统计软件包及专门的统计软件中也有详细程序。最简单的实用方法可能是多重反复加权（*raking*）（Battaglia, Hoaglin, & Frankel, 2009），即通过指定人口边界使样本与之匹配。

与插补相同，加权需投入时间、资源并具备专业能力。当面临复杂的问题时，需要专家的指导。

4.1.5 编码和重新编码

在数据准备阶段，研究者可以生成各种衍生数据（第 2.4.1 节），然后将其导入最终数据库，从而供给所有用户使用。

编码是通过为开放式文本条目（如单词、短语、评论）赋值而生成衍生数据的过程，这一方法也适用于多媒体条目（如声音、图片、视频）。为文本或多媒体赋予数字代码，目的是将其纳入定量分析，通过计算频率、相关性、双变量分析、细分等方式深化对数据的理解。本书主要关注开放式文本项目（即单词、短语、观点）的编码，但其原则也适用于其他类型。

进行编码时，需要一个目标值（类别）的编码列表，也称为编码方案或编码结构。这类列表有时是预先存在的（如官方的职业代码），有时则需研究者为特定变量专门制定（如关于新服务的公开评论）。编码列表中的类别具有以下特征（Groves et al., 2009, p. 332）：（a）每个分类都有唯一的编号；（b）通过标签描述该类别涵盖的所有答案；（c）类别具有穷尽性，确保所有答案均可归入某一类别；（d）类别之间互斥；以及（e）类别必须符合研究目的。

对于大样本的编码，建议先从随机子样本或分层子样本（例如 100~300）入手，提炼反复出现的主题（类别与代码）。如果原有类别无法涵盖大量的答案编码，就需要增加新的类别。对于出现频率最高的类别，可以进入另一层编码。有时每个条目还可包含多个维度，例如对新服务的公开评论，除主题（类别）外，还可增加强度（程度）维度。

初步的编码方案设定后，可以进一步应用并检验它对答案进行编码的效果。人

工编码时,最好由另一位编码员同时对小部分样本进行独立编码,再通过编码者信度指标评估编码员与编码方案的质量 (Biemer & Lyberg, 2003)。

编码方法在很大程度上取决于问题的性质。数百条评论与数百万条评论需要采用不同的方法 (例如全球企业服务的客户反馈)。在官方统计中,工作人员通常经专门培训以应对大规模编码工作;而商业领域中,编码外包十分普遍,逐字编码 (*verbatim coding*) 的标签经常出现在这些服务中。

用户友好的界面对于支持人工编码是极其重要的,对于半人工编码也是如此。半人工编码中,词根被自动识别为同一代码 (例如苹果们和苹果),但仍需要人工核查。值得注意的是,词语编码 (如品牌、名称、水果) 与语境编码 (*contextual coding*) 有相当大的区别。语境编码是指从段落或长文中提取概念、主题和维度。依托人工智能、数据/文本挖掘及机器学习发展起来的自然语言处理技术,正推动完全自动化编码的应用 (Esuli & Sebastiani, 2010), 无论是否存在预设编码方案。针对某些特定问题已开发出各类具体方法,如情感分析编码,并与可视化和各种类型的大数据分析建立了强有力的联系。

目前有许多专门用于编码的软件包。此外,标准统计软件包 (如 IBM-SPSS、SAS、Stata、R) 以及一些网络调查软件也具备编码功能。后者是非常实用和节省时间的,其综合功能与专业软件相差无几。

在调查后阶段,编码存在多种替代方法。正如第 2.3.2 节中所述,开放式文本问题的常见误区是在可采用封闭式问题时仍使用开放式设计。例如,当有多达 20 个选项,而且所有受访者都很熟悉这些选项时,采用封闭式问题可节省大量编码资源。

在事先知道编码方案的情况下,另一种选择是使用客户端脚本的交互潜力,激励受访者进行实时编码。第 2.3.2 节中已详细介绍此类方法,包括答案树、下拉菜单及自动完成功能等。另一种替代方法是文本字符串匹配 (*text string matching*),用于职业编码 (Tijdens, 2011)。受访者输入与其职业相关的词汇或其首字母后,系统从预设数据库中返回最佳匹配选项,供其最终选择。

除对开放式文本条目编码外,还可通过为其他问题类型赋予额外值生成衍生数据。这通常被称为重新编码 (*recoding*),因为原有类别 (即原始编码选项) 已经存在,但由于某种原因需要对其进行调整和修改。然而,数据准备阶段只适合最基础的重新编码。例如,受访者报告的出生年份,通过重新编码计算出具体年龄,如 10 岁。

有时,更复杂的衍生数据也可以在数据预处理阶段生成和赋值,例如在知识或阅读表现调查中,利用多个变量计算综合得分。

当然,每个数据分析员均可根据特定分析目的,生成更多专属的衍生数据。

4.2 初步结果

关于初步结果,本节将讨论报告(*reporting*)(第4.2.1节)和发布(*dissemination*)(第4.2.2节)功能在网络调查软件中的整合程度。网络调查软件的功能特性差异显著,用户需求也各不相同。有时,这些初步结果已能满足研究人员、客户或用户的需求,但更多时候仍需开展更深入的分析。此处仅探讨基于综合网络调查软件的智能自动统计分析所生成的报告与发布功能;而在专门统计软件中进行的高级分析,以及各类人工发布形式,均不属于核心网络调查过程的范畴。

4.2.1 报告

数据准备阶段完成后,初步报告不仅能帮助理解实质性结果,还可支持数据验证与编辑核查(第4.1.2节),辅助正式调查过程的监控(第3.3.4节),并为后续分析提供基础输入。对于简单的网络调查而言,这类初步报告有时已能满足所有需求。

初步报告已逐渐成为核心网络调查过程中不可或缺的部分。93%的网络调查软件已经支持自动生成单变量汇总统计(频率、描述性统计)和默认图表(92%),其中一半还提供交叉分析(Vehovar, Čehovin et al., 2012; 另见第5章补充材料“网络调查软件分析”, *Web survey software analysis*, <http://websm.org/ch5>)。

测量水平和问题类型对于选择默认的统计分析至关重要(J.F. Healey, 2011)。单变量的名义、顺序、等距或等比测量,以及表格和高级题型,均需匹配相应的分析方法。关于各类问题类型的默认报告示例,读者可以参考补充资料《网络调查中的问题和布局及测量水平》(*Questions and layouts in web surveys and Measurement levels*)(<http://websm.org/ch2>)。

部分高级网络调查软件,已集成了类似标准统计软件包的强大统计功能。但需明确的是,无论是在网络调查软件内完成,还是借助外部标准或专门统计软件包开展的高级统计分析,均不属于调查后阶段的范畴,因为它主要是实质性研究工作。

本书关于统计推论部分(第2.2.4节)已指出,概率抽样可确保估计值的合理性。任何样本结果均需在统计推论框架下报告,包括标准误、置信区间等指标。所有标准统计软件包默认假定为简单随机抽样(SRS)。简单抽样方差的估计公式也被用来计算方差、标准差、标准误和置信区间。但这种简化假设在复杂样本(如阶段抽样、分层抽样、多期抽样、重复抽样、样本库抽样,或涉及加权、插补)或复

杂的估计（如比率或回归参数）中并不适用，这些需要特殊的估计程序（Heeringa et al., 2010）。因此，即便假设为简单随机样本，若实际样本更复杂（含加权、插补）、估计方式更复杂，或为非概率样本，也需明确这是近似估计。此时，首先应做出明确警示，其次第 2.2.4 节的相关内容完全适用。这也提示可能存在不可靠的估计值，以及区分分析总体（普查）与样本的情况。

此外需注意，初步报告可能受限于数据准备工作的完成程度。如果编辑、插补等数据准备步骤尚未执行或未完成，报告结果会存在一定局限性。

4.2.2 发布

调查结果的发布过程复杂且耗时，需要在初期研究活动中进行细致规划（第 2.6 节）。

网络调查软件具备多项强大功能，能帮助研究者、客户及其他用户高效获取丰富的初步报告。网络调查软件不仅可自动发布特定初步报告，还能在用户查询时在线生成实时报告。后者对于调查来说特别有价值，因为在正式调查阶段需要持续掌握实时状态。此外，系统可根据预设标准自动向目标用户推送提醒，例如周期提醒（如每周报告）或条件提醒（如达到临界样本量）。同时，公开原始数据的实时状态也颇具价值，例如便于审查开放式文本问题或追踪具有特定特征的样本单位。

网络调查软件集成了丰富的制表功能，同时也已经存在丰富的专门工具支持。这里我们指的是通过用户友好的在线界面，使用户能够执行交互式统计分析（交叉表、数据分割、筛选、图表甚至多变量统计）。与使用标准统计软件包的替代性离线分析相比，集成于网络调查软件的功能具有诸多优势：可通过任何设备访问、提供实时数据，无需数据准备或转换、省去数据文件传输给用户的环节以及不需要安装软件。此外，还能对系统访问权限进行多层次精细控制，增强数据安全性。上述这些功能甚至可替代纸质概要，纸质报告往往是呈现结果的基础形式。

高级网络调查软件的重要功能之一是支持分级访问，可为不同层级用户分配特定功能权限（分析、数据查看、导出、编辑等）。在大型组织中，用于收集各种利益相关者（客户、员工）反馈的工具，其功能设计尤为详尽。也就是说，在这类结构庞大、结构化和等级化的组织中，信息获取渠道的设定与控制需要极为精确。第 1.3.6 节以企业反馈管理（EFM）为例已经对此进行说明。

4.3 数据导出与文档化

除了依托于网络调查软件内生成的统计结果的初步报告和发布,网络调查的其他产出还涉及最终数据库的导出(第 4.3.1 节)、统计披露问题(第 4.3.2 节)、文档化(第 4.3.3 节)和数据归档(第 4.3.4 节)。这些环节有助于外部软件(如统计分析工具)的高效使用,同时保障数据存档的广泛可及性。

4.3.1 数据导出

网络调查数据有时无需导出,因网络调查软件的综合功能已能满足汇总统计、制表、绘图及其他分析需求。但如需开展深入统计分析、外部数据处理或归档,仍需进行数据导出。

无论是否导出,作为数据准备阶段最终产物的最终数据库需具备初始描述、版本控制、编制说明、时间戳及变更追踪等基本特征,这一点至关重要。

数据导出的准备工作始于预调查阶段,需处理技术层面的问题,尤其是数据库的定义(第 2.4.1 节)和技术检测(第 2.4.4 节),以及预测试和实地试验(第 2.3.5 节)。通过这些环节,可检验实际导出流程的运行情况、标准导出格式(如 Excel)的有效性,以及复杂数据结构(如分支、模块、嵌套数据等)的导出效果。导出变量名称和标签、变量的测量水平和类型属性时,需要特别注意,因为这些信息在导出过程中可能无法完整保留。此外,在变量的数量、案例的数量、变量名称(长度,是否允许字符串)和复杂数据结构的处理等方面也存在约束。还需要特别注意标识符变量(*identifier variables*),必须将其与调查数据严格分开。根据项目需求,还应核查特殊并行数据和元数据的导出情况:明确需导出的并行数据范围(仅基础默认数据如时间戳、用户设备字符串,或更高级数据)及导出形式(结构化可直接分析的格式,或仍为字符串形式)。

网络调查软件通常支持某些格式的输出,至少是 Excel 和 CSV(逗号分隔值)格式,这意味着表格数据(数字和文本)以纯文本形式存储。但对于后者,一些数据可能会在导出过程中丢失(例如标签)。然而,这种格式应用普遍,可作为其他导出方式出现问题时的折中选择。部分变量类型(如联合表、社会网络数据)在直接导出至标准统计包(如 IBM-SPSS、SAS、Stata、R)时可能因标签和数据结构无法被完整保留而难以识别。

4.3.2 统计披露

网络调查通常会向受访者作出隐私保证,包括声明其应答将被保密,且仅以汇总形式呈现结果,不披露任何个人信息。正如第2.4.3节技术准备阶段中已经提到的,首先需将调查应答与个人标识信息分离。标识信息常包括电子邮件地址、姓名、物理地址、受访者或客户识别号码以及详细的地理信息等。一些设备类型相关的并行数据,如IP地址,也属于个人信息范畴。

当公开汇总表、交叉列表或调查数据集(微观数据)时,还存在另一种风险。间接识别受访者的机会增加,这在商业调查和官方统计调查中尤为突出。依据专业准则与伦理标准,研究者有责任防范统计信息泄露,这同时也是一个法律问题。因此,我们需要限制控制变量的细化程度,如年龄、性别、受教育程度、地理信息等。当这些变量纳入交叉表或最终数据库时,若单元格样本量过小,可能导致受访者被识别。例如,当涉及地理信息时,我们可以剔出邮政编码和城市名称等,只保留地区、国家或州;再结合年龄、性别等控制变量进行交叉分析,以确保每个单元格的样本量足够(如不少于10个),从而避免个体识别。为防范交叉分析及微观数据中的信息泄露,研究者已开发出多种方法。关于这一问题的深入探讨,可参考Hundepool等(2012)的研究,而Groves等(2009, p. 308)和Vehovar, Zaletel, & Seljak(2008)的概述也颇具参考价值。

4.3.3 文档化

数据文档化是核心网络调查过程中极其重要却经常被忽视的环节。文档记录不完善可能引发潜在错误与误解,还会浪费其他用户及分析人员使用和解读数据文件的时间。此外,即便是撰写调查报告的研究者,也可能在一年后就忘记诸多细节。

为减少此类问题,需结合特殊的元数据并以标准格式描述调查。以下为需要记录的关键方法学信息(ICPSR, 2012):(a)抽样设计和数据收集方法;(b)使用的激励措施;(c)关于正式调查的实施场景、时间和数据收集组织的信息;(d)关于合格性判定、最终筛选流程、应答过程指标及其他数据质量指标的详细信息;(e)问卷最好导出为稳定的标准化格式(例如XML),有时还附带相应的翻译;(f)编码手册与编码方案、重新编码程序和最终的衍生变量;(g)数据插补程序和加权程序的说明;(h)基于数据集生成的内容摘要和其他报告;以及(i)用于预览问卷的URL,以便了解交互功能。此外,为满足正式归档需求,(j)还需包含其他管理性

数据，如作者联系方式、研究目标、组织信息、主要研究者、使用条件及各类归档专用数据。更规范的做法是，这类描述应遵循国际公认的结构与标准，例如数据文档倡议（DDI）¹。

理想情况下，用户通过阅读调查文档，无需向数据收集团队进一步咨询，就能重现研究结果与发现。

编码手册（*codebook*）（f）尤为重要，它应该包含关于调查问题的标准信息，从问题措辞到回答类别和相关代码（Mohler & Uher, 2003）。在网络调查中，第 4.1.1 节所述与应答状态相关的代码尤其重要。编码手册通常可由统计软件包及多数网络调查软件自动生成，近年来这些软件已开始生成先进的、交互式、用户友好且带超链接的编码手册。

关于问卷（e），除文本文件（如 PDF、RTF）和电子表格格式（如每行一个问题）外，网络调查软件通常可以将问卷及其结构（如分支）一同导出。最终的编程代码应单独导出。此外，问卷预览的 URL 也很重要，便于查看问题及流程的实际呈现形式与交互效果。部分网络调查软件还可生成 PDF 格式的整个调查截图系列。

正如第 2.6 节管理阶段所述，强烈建议在预调查前及正式调查阶段就开始收集文档资料。这不仅能节省时间，还有助于澄清细节、辅助管理，同时起到记录细节的提醒作用。

4.3.4 数据归档

仅完成数据导出与记录并不足以支持调查数据的长期使用，还需确立稳定的归档格式与存储位置。

首先需区分两类归档：研究人员或研究机构的内部归档，以及专业数据归档机构的标准化外部归档（如全球 CESSDA 归档网络中列出的）。后者基于精细的程序和复杂的归档标准，往往需要投入大量资源，且这类数据档案主要是利用公共资源进行的调查，如官方和学术调查。各国在归档实践上存在差异，具体可参考各国数据档案机构的相关规范以进一步了解情况。

商业机构通常不会对调查数据进行公开归档，但也存在例外情况。即便如此，为实现高质量归档，仍需遵循专业的归档原则。目前已有多种标准与认证为专业归档方法提供了参照（如 ISO, DSA²）。

许多网络调查可能根本不需要专门的归档程序，因为网络调查软件服务已经自动保存了基本档案：数据文件、问卷版本、输出、结果、含基础实地调查活动与诊

断信息的仪表盘、问卷预览链接等。这些内容几乎能满足前述所有的文档要求 [(a)-(j)]。仅少数一般性描述无法自动生成,需手动添加。当然,这些档案均以特定网络调查软件的格式存储,长期来看可能面临无法访问的风险。

除此之外,现代网络调查软件通常还支持问卷和问题的进一步复制和重复使用。因此,每个用户都可以建立专属问卷和问题档案库,同时部分软件还能生成公共资源库。

表面上看,网络调查软件似乎已较好地解决了归档问题。但从长远来看,考虑到未来5年或10年,我们无法确保开展特定网络调查时使用的软件仍能存在,并支持旧版本调查数据的访问。这对个人开展的自助式网络调查来说关系非常大,对组织而言则更为关键。毫无疑问,保留历史调查数据具有价值,毕竟收集这些数据已耗费大量精力,但如何平衡归档资源投入仍是需要考量的问题。

数据归档领域已形成保障长期归档的方法、标准、格式与流程(ICPSR, 2012)。部分网络调查软件(如Blaise)已符合相关归档标准,并能按标准格式导出数据。此外,还可为特定调查数据分配唯一的数字对象标识符DOI(Digital Object Identifier)。这不仅能确保数据的永久识别,也为建立与科学引文索引并行的数据引用体系奠定了基础。

如需进一步了解相关内容,如前所述,数据归档机构提供了大量关于其活动的信息(ICPSR, 2012);也有专门的著作(Corti et al., 2014)以及简明在线手册(例如参见英国数据服务机构的手册³),总结了重要的基本准则。而在开展外部正式归档时,强烈建议在预调查阶段咨询专业人士。与其他调查后活动相同,归档需在预调查阶段就纳入规划(第2.6节),因其贯穿整个网络调查项目的全流程(例如会影响变量名称的选择)。

注 释

1 <http://www.ddalliance.org>

2 <http://www.datasealofapproval.org/>

3 <http://ukdataservice.ac.uk/media/440320/depositsurvey.pdf>

网络调查实施的选定主题

本章将聚焦网络调查实践中的三个特定主题：适用于网络问卷作答的新型设备、在线样本库和网络调查软件。尽管这三类主题在本书的核心章节（第 2 章、第 3 章和第 4 章）中已有所提及，但鉴于其内容的丰富性以及调查从业者对它们的高度关注，值得本书进行更深入的探讨。我们特意设置独立章节进行阐述，以避免干扰核心网络调查过程。在本章中，我们将对相关实施步骤展开一般性讨论，这些步骤相对比较稳定，不会像受访者使用设备的多样性、在线样本库的潜在应用以及市场上调查软件的更新迭代那样迅猛。

在第 5.1 节中，我们将关注一个快速增长的受访者群体。该群体或因自身选择，或应调查机构要求，使用智能手机和平板电脑参与调查。我们将梳理当前关于这一主题的认知，因为在当下设计网络调查，已不能再默认所有受访者都会通过台式机或笔记本电脑访问并完成问卷。

在线样本库是通过网络调查从样本库成员那里收集信息。这些在线样本库在网络调查中占据着相当大的比例。对研究人员而言，它们在数据收集的便利性、效率和成本等方面更具优势，但同时数据质量和样本代表性方面带来了挑战。第 5.2 节将对此加以详细探讨。

最后，网络调查的开展离不开网络调查软件，第 5.3 节专门加以介绍。本节不仅会对软件进行分类，还将回顾研究人员在选择和使用软件时应重点关注的关键特征。

5.1 智能手机、平板电脑及其他设备

近年来,网络调查领域中最显著且变化最为剧烈的趋势,或许就是越来越多的人开始使用多种设备(尤其是智能手机)来回答网络调查问卷。本节首先介绍新设备的一般术语、类型划分和相关趋势(第5.1.1节);接下来,我们将阐述研究人员可以采取的策略(第5.1.2节)和这些设备的相关特征(第5.1.3节);我们还会概述在移动设备上设计网络调查的基本建议(第5.1.4节),并讨论移动调查应用程序引发的相关问题(第5.1.5节);最后,我们介绍软件的特征(第5.1.6节)和未来面临的挑战(第5.1.7节)。

5.1.1 多设备介绍

在网络调查发展初期,主要的技术焦点是确保网络问卷——此处指的是基本的网络调查——在不同的浏览器和不同的操作系统上可以正确显示。主要假设是受访者使用台式机或笔记本电脑来填写网络调查问卷。本章将讨论近年出现且发展迅速的现象,即针对多种设备的网络调查(*web surveys for multiple devices*)。对于此概念,本书指的是受访者在不同设备上参与的网络调查,这些设备可以接入互联网,但不能归类为传统的台式电脑和笔记本电脑。首先,这些设备包括移动设备(*mobile devices*)——小型的手持(便携式)计算设备,通常有触摸屏输入或一个小键盘。此部分主要指智能手机(*smartphones*)和平板电脑(*tablets*),前者是具有先进计算能力和互联网连接功能的移动电话,后者是主要通过触摸屏幕来操作的小型移动计算机。

截至2014年1月,美国成年群体中,58%的人拥有智能手机,42%的人拥有平板电脑¹。在同期,欧盟人群中,52%拥有智能手机的人订阅了移动互联网服务(European Commission, 2014)。新近的数据可以在“互联网统计”(Internet statistics)中找到,见第2章补充资料,网址为<http://websm.org/ch2>。

就社会人口统计学特征而言,移动设备的使用情况与互联网的普及模式类似。与使用台式机/笔记本电脑参与调查的受访者相比,智能手机用户尤其是早期用户,更年轻、受教育程度更高(Peterson, 2012; Stapleton, 2011)。

然而,我们应该意识到,其他移动设备也能够接入互联网并用于开展网络调查。在这种情况下,我们应该提到功能手机(*feature phones*)和平板手机(*phablets*)。

功能手机较智能手机的功能更有限,但通常仍可以浏览网页。功能的限制通常与 Wi-Fi 网络接入、触摸屏功能和下载应用程序的能力有关。然而,目前没有明确和标准化的方法来区分功能手机和智能手机,所以此处不做详细阐述。平板手机是集智能手机和平板电脑功能于一体的移动设备,尺寸介于两者之间。其他移动设备包括电子阅读器 (*e-readers*) (或是电子书阅读器,主要是为阅读电子书和期刊而设计的移动电子设备) 和 MP3 播放器 (*MP3 players*, 能够存储和播放数字媒体的便携式数字消费电子设备)。除了移动设备,视频游戏机 (*video game consoles*) 和交互式电视 (*interactive TV, ITV*) 也可以接入互联网。在传统电视技术的基础上,它们增加了各种数据服务,如按需提供内容、网上购物和网络银行服务等。

欧盟统计局 2013 年对斯洛文尼亚的调查数据显示,有 8% 的互联网用户 (16~74 岁) 除了使用台式机 / 笔记本电脑、平板电脑和手机,还通过其他设备接入互联网。这一结果说明了这些设备的普及程度。就各种通信技术普及率而言,这与欧盟的平均水平相近²。尽管这些设备占比相当大,但因其都是为单一和特定目的而设计 (例如玩游戏、听音乐、阅读、观看视频),不太可能用于与网络调查相关的活动。目前,较少实证研究发现,这些设备用于访问网络调查,其比例可以忽略不计。然而,这种情况在未来可能会改变,尤其是随着交互式电视使用的增加。尽管交互式电视在市场上已经出现了十多年,但由于各种技术和商业原因,仍处于早期应用阶段。这一领域出现了各种相互竞争且相互重叠的概念、解决方案和产品,如交互式网络电视、互联网电视、网络电视、智能电视、苹果电视、数字电视棒,上述都指向电视和互联网的融合。

随着可接入互联网的设备种类如雨后春笋般涌现,在不同设备上设计或填写问卷的可能性也成倍增加。基于 Peterson (2012) 提出的特定设备的有意使用和无意使用的概念并加以扩展,本书定义了定向设备和非定向设备网络调查 (*intentional and unintentional device-specific web surveys*),其中的意图 (*intent*) 是指研究人员 (即调查设计者) 的意图,而不是受访者的意图。在一项定向设备网络调查中,问卷的设计设备与受访者的设备之间需要一对一匹配。例如,研究者设计一份在台式机 / 笔记本电脑上进行的调查问卷,而实践中受访者也在台式机 / 笔记本电脑上完成网络调查问卷,便符合这一定义。在非定向设备的网络调查中,问卷的设计设备和受访者的设备之间不需要匹配。

最需关注的问题是,研究者设计用于在台式机 / 笔记本电脑上完成的问卷,受访者却在智能手机上完成作答。越来越多的受访者使用智能手机阅读和回复电子邮件,这也是非定向设备网络调查中最常见的情况。例如,2013 年底的数据显示,移

动设备占电子邮件阅读量的近50%³。因为电子邮件邀请是网络调查的主要邀请方式,所以很多受访者也仅仅通过点击智能手机屏幕上的链接来完成网络调查。

市场调研公司 Kinesis (2013b) 通过观察受访者访问网络调查时使用的设备发现,2013 年第三季度,在美国,由其平台发布的网络调查中,有 51% 受访者通过智能手机或平板电脑发起填写。而在欧洲, Kinesis 报告称这一比例要低得多,只有 10%。Peterson (2012) 发现,美国通过移动设备参与调查的人数比例因目标人群和调查主题而异,从最高 30% 到最低 1% 不等。最后,在非概率在线小样本库中,高达 20% 的样本库成员使用移动设备 (Fisher & Bernet, 2014),而在概率在线样本库中,这一比例约为 10% (Toepoel & Lugtig, 2014)。

这些数据清晰显示了以下趋势:在发达国家,未来几年随着智能手机的进一步普及和功能完善,其在网络调查中的使用比例将持续上升。因此,几乎所有网络调查都需将智能手机视为重要的访问渠道。

由于受访者在适应新技术方面已经领先于调查行业,非定向设备网络调查是存在问题的。Meaning Ltd 2011 年对 36 个国家 230 家市场调研公司开展技术调查,发现只有 15% 的公司针对智能手机设备修改了问卷,而许多公司甚至无法提供使用智能手机的受访者的百分比 (Macer, 2012)。这表明这些公司没有针对特定设备修改问卷。同样,在 2012 年和 2013 年,至少有 30% 的公司不知道(或无法提供估计)受访者是通过台式机/笔记本电脑、智能手机还是平板电脑访问网络调查 (Macer & Wilson, 2014)。即使是 WebSM 网站上关于软件的最新概述 (WebSM, 2014) 也指出,当前软件对简易网络问卷的移动端显示支持有限,且常忽略必要的适配调整。

在展开更详细的讨论之前,我们先明确基本概念。首先,我们将使用术语“移动调查”(mobile surveys)来表示在测量阶段涉及移动设备(通常是移动电话)的任何调查。这些可以是访员执行的,如移动计算机辅助访谈(也称为 mCATI,一种计算机辅助电话访谈,受访者使用移动电话回答)或移动计算机辅助个人面访(也称为 mCAPI,即访员使用移动设备来进行的计算机辅助个人调查),也可以是由受访者自行完成的。在后一种情况下,我们主要讨论与网络相关的调查。的确,移动设备也可以用于短信调查,不需要接入互联网,而且历史上曾有人使用特定的无线应用协议(WAP)方法进行调查。然而,由于这两种协议在网络调查实践中已经过时,故不纳入讨论。

我们将使用“移动网络调查”(mobile web surveys)一词来表示在智能手机或任何其他可访问互联网的移动设备上进行的自我管理式调查,其中网络技术直接服

务于实地数据收集。

移动网络调查进一步区分为基于浏览器的移动网络调查 (*browser-based mobile web surveys*) 和移动调查应用程序 (*mobile survey applications*)。前者与常见的网络调查类似,即受访者使用移动设备浏览问卷。后者实际上超出了基本的网络调查模式,即仅限于基于浏览器和基于服务器的交互。也就是说,移动调查应用程序可以在没有浏览器的情况下运行问卷,支持离线交互。在使用此类应用程序时,网络仅用于接收问卷,并在问卷完成后将其发送到研究人员的服务器。

多设备调查 (*multiple device survey*) 这一概念主要强调调查问卷可使用不同的设备来回答,并且研究人员会针对性优化问卷来支持这种情况。现今,几乎所有的网络调查都在无意中实现了多设备访问。

还应该注意,移动调查是更广泛的移动研究 (*mobile research*) 概念的子集。从最广泛的意义上讲,移动研究涵盖了任何涉及移动设备的研究方法:从移动网络调查(包括移动日志调查),到移动计算机辅助电话访谈和移动计算机辅助个人面访,移动定性研究方法(如移动人种学研究),尤其是各种市场研究方法(如移动购物者研究、移动广告测试等)和/或其他非调查方法(如基于位置的研究、快速响应二维码数据收集)。

5.1.2 处理移动设备的调查策略

Callegaro (2010) 曾提出若干应对智能手机、平板电脑等移动设备参与网络调查的可行方案。Couper (2013) 在此基础上进行了补充,并对相关研究进行了全面综述。下文将拓展上述研究者对移动网络调查的初步论述,归纳出以下策略。

1. **不做任何改动 (*Doing nothing*)** 指不根据访问设备类型修改网络问卷,仅对设备进行识别和分类。这种方法实施起来最容易,但可能对于使用智能手机的受访者而言,其数据质量方面存在隐患。研究者可根据此类用户的数量或其他标准,在数据分析阶段排除部分样本,或依据明确规则谨慎处理。例如,在开展问题措辞实验时,在数据分析中使用有关设备类型的信息以控制不同的影响因素是非常重要的。
2. 可以在多个环节 **明确劝阻使用移动设备参与调查 (*Explicitly discouraging the use of a mobile device for surveying*)**:
 - 2(a) 在调查邀请中 (*In the survey invitation*): 在调查邀请中劝阻受访者使用

移动设备。这是另一种无需研究者费力的方法,但遗憾的是,如下所述,它似乎并不奏效。

- 2 (b) 在介绍页面 (*On the introduction page*): 在网络问卷的初始界面就劝阻受访者,但这种策略也可能是无效的。例如,某市场策略公司针对约 71 000 名美国客户开展的简短满意度调查 (6~7 分钟) 中,将受访者随机分为三组:(i) 在电子邮件邀请中劝阻使用智能手机参与调查;(ii) 在介绍界面不鼓励使用智能手机参与调查;(iii) 发送标准电子邮件邀请的对照组 (Peterson, 2012)。结果发现,实验条件对使用智能手机参与网络调查受访者的百分比没有实质性影响。这一结果表明,无论传达何种信息,主动切换到另一设备的受访者的比例没有变化。在大学生群体中获得的结果略好一些,将智能手机受访者分为提示组和无提示组,提示一半受访者切换到台式机/笔记本电脑,而另一半则没有相关提示。在提示组中,14% 的大学生切换了设备;而在对照组中,只有 7% 的人自发地切换了设备 (Guidry, 2012)。
- 2 (c) 阻止移动设备访问 (*Blocking mobile device access*): 这可以通过识别受访者所使用浏览器的用户代理字符串来完成。标记被屏蔽的受访者,并发送定制的电子邮件提醒,邀请他们从其他设备完成调查。由于这种解决方案可能会导致较高的无应答率,因此需要非常谨慎使用和仔细检测。换言之,该策略在这些用户的占比非常小且对问卷进行移动端适配带来的收获微乎其微情况下适用。
3. 针对移动浏览器优化网络问卷 (*Optimizing web questionnaire for mobile browsers*):
 - 3 (a) 针对最常见的设备和相应的操作系统优化网络问卷 (*Optimize the web questionnaire for the most common devices and corresponding operating systems*): 重点调整问卷在移动浏览器中的显示效果。通常,当服务器识别出移动设备后,会自动适配布局,但受访者也可手动切换至普通网页版。优化内容包括外观调整 (如将水平刻度改为垂直刻度) 和技术适配 (如增大单选按钮尺寸)。首先,我们需要确定针对哪些设备 (即哪些系列的智能手机或平板电脑) 来优化问卷外观。这通常取决于特定设备在调查目标人群 (或他们居住的国家) 中的市场覆盖率,也取决于可用的资源。如果研究人员从头开始开发和编程一个独立的网络调查——尽管可能性很小——那么需要优化很多方面。更多时候,优化效果取决于调查软件是否支持移动适配功能。

- 3 (b) 针对所有设备优化网络问卷 (*Optimize the web questionnaire for all devices*): 这个解决方案要求很高, 几乎难以实现, 因为它要求网络调查软件可以处理所有类型的移动设备和所有类型的相关平台 (操作系统), 以及支持所有版本的浏览器。与台式 / 笔记本电脑的标准浏览器相比, 移动浏览器的标准化程度要低得多。然而, 针对常见的屏幕尺寸和关键浏览器版本, 处理两个最常见的平台——安卓 (Android) 和苹果 (iOS) ——目前 (2013 年) 就能覆盖 95% 以上的市场。⁴
4. 将网络问卷作为移动应用程序提供, 为特定调查进行开发 (*定制移动调查应用程序*) (*Providing a questionnaire as a mobile application, developed for a particular survey (custom mobile survey application)*) 是指为特定调查开发移动应用程序, 使部分或全部问卷在移动设备端 (客户端) 运行。这可以利用移动设备的许多特定功能 (全球定位系统、闹钟、照片和视频), 但需要大量资源投入, 尤其是因为它必须针对不同的智能手机平台单独开发。此外, 应用程序在上线应用商店 (谷歌应用商店、苹果 iTunes、微软 Windows Phone 应用商店) 之前可能需要经历耗时的流程。最后, 受访者需要额外的精力来下载和安装特定应用程序 (Ferneer & Scherpenzeel, 2013)。
5. 通过移动设备上的通用调查应用程序 (*通用移动调查应用程序*) 提供问卷 (*Providing a questionnaire through a general application for surveying on mobile devices, general mobile survey application*), 受访者只需下载一次相应的通用调查应用程序, 然后以优化的方式自动接收后续问卷。这种专门用于移动调查的通用应用程序可以是独立的, 也可以是现有网络调查软件的补充应用程序。而现有网络调查软件已经支持在台式机 / 笔记本电脑上的基本网络调查模式。对于调查者和受访者之间有长期关系的网络调查来说, 例如在线样本库, 通用应用程序比较适用。

5.1.3 设备效应

本节的剩余部分主要关注使用智能手机回答网络问卷的移动网络调查, 但结尾部分将简要介绍平板电脑和交互式电视的相关情况。下面讨论的许多问题适用于各种类型的移动设备。

之所以重点关注智能手机, 是因为它是网络调查中最常用的移动设备。然而, 我们应该注意到, 直到 2013 年, 智能手机才首次占到手机总出货量的一半⁵, 所以

功能手机仍然占主导地位。

显然,随着越来越多的受访者在移动设备上参与调查,每份网络问卷都应该针对智能手机进行优化(上述策略3、4和5)。然而,即便问卷经过设备适配优化,如果同一份网络问卷是在台式机/笔记本电脑或智能手机上回答的(无论是否针对设备进行了优化),结果可能会有所不同,这便涉及具体的设备效应(*device effects*)。需要明确的是,在此处设备效应仅指在调查过程的测量阶段,用于描述调查的特定设备(例如台式机/笔记本电脑或智能手机)对受访者答案的影响。因此,必须将设备效应与模式效应加以区分,后者是由于使用某种调查模式(例如基本的网络调查模式)来完成问卷而产生的。研究者通过比较智能手机和台式机/笔记本电脑的网络调查,探究潜在的设备效应,发现存在以下问题。

智能手机用户的响应更快,但完成问卷的时间更长(Faster reaction, but longer questionnaire completion times on smartphones): Peterson (2012)指出,平均而言,从电子邮件邀请发出的那一刻开始计算,智能手机受访者平均比台式机/笔记本电脑受访者更早地开始填写网络问卷。与此同时,尽管进行了优化,但受访者在智能手机上参与调查可能比在台式机/笔记本电脑上参与调查要多花25%到50%的时间。Bruijne和Wijnant(2013)、Grenville(2012)、Schmidt和Wenzel(2013)以及Wells、Bailey和Link(2014)也都报告了类似的结果。使用智能手机的受访者完成问卷的时间较长,可能与移动网络速度(通常低于宽带连接)和网络可用性问题有关。一些研究表明,滚动设计比智能手机的分页有一定优势,可见问卷设计也可能是原因之一(Mavletova, 2013)。

智能手机端的参与率较低(Lower participation rates on smart-phones): 研究发现,将受访者随机分配到不同设备,智能手机调查对象的参与率较低。例如,在2011年底进行的两项研究中,智能手机网络调查的参与率约为台式机/笔记本电脑网络调查的一半(Mavletova, 2013; Mavletova & Couper, 2013)。同样,在荷兰的一项实验中,智能手机受访者的完成率为47%,而台式机/笔记本受访者的完成率高于60%(Bruijne & Wijnant, 2013)。这可能是因为,在智能手机上完成问卷所需的时间更长,且设备屏幕尺寸较小使任务复杂化,受访者认知负荷增加。此外,针对智能手机的网络调查仍处于起步阶段,许多优化方案尚未完善,也可能影响参与率。

智能手机端的中断率更高(Higher breakoff rates on smartphones): 在三项针对专业人士(即亚洲、欧洲和北美的在线广告商)的研究中,网络问卷没有针对智能手机进行优化,Callegaro(2010)发现,在智能手机上参与调查的中断率,比在台式机/笔记本电脑上参与调查高出2~3倍。在一项客户满意度研究中,Stapleton

(2011)发现,使用台式机/笔记本电脑的网络受访者的中断率为3.3%,而使用智能手机的受访者的中断率为16.9%。在大学生群体中也发现了类似的结果(McClain, Crawford, & Dugan, 2012),两个德国在线样本库的研究(Bosnjak, Poggio, Becker, Funke, Wachenfeld & Fisher, 2012; Schmidt & Wenzel, 2013)亦是如此。特别是,当遇到表格、开放式问题和Flash技术时,中断率更高。例如,Jue和Luck(2014)发现,在10个问题的网络调查中遇到12列10行表格时,受访者会出现许多中断。对于智能手机用户来说,该页面的中断率是台式机/笔记本电脑用户的两倍。这也可能与智能手机缺乏问题优化有关。正如Callegaro(2010)所指出的,表格尤其有风险,因为应答选项是水平布局的,加之智能手机的屏幕较小,因此受访者很难在智能手机上完成。换言之,如果没有水平滚动,水平响应选项就无法完全显示。在一项随机实验中,Stapleton(2011)在将列表从水平格式切换到连续垂直格式时,智能手机的中断率降低了30%。智能手机上更难操作带有开放式问题的网络问卷。Jue和Luck(2014)在一份带有开放式问题的网络调查问卷中发现,智能手机用户的中断率比台式机/笔记本电脑用户高出50%。此外,Flash技术不适用于某些智能手机(苹果手机),Bosnjak等人(2012)也注意到,当采用Flash技术时,使用智能手机的受访者的中断率增加了一倍。

智能手机端开放式文本问题的答案长度较短(Shorter length of answers to open-ended text questions on smartphones):大多数研究(实验或观察)显示,智能手机端受访者针对开放式问题给出的答案较短(Mavletova, 2013; Peterson, 2012; Wells et al., 2014)。Peterson(2012)还观察到,在对开放式文本问题进行编码后,尽管智能手机端受访者写的字符数量较少,但结果和其他设备受访者相似。

关于社会赞许性的调查结果喜忧参半(Mixed findings on social desirability):在一项实验中,俄罗斯非概率在线样本库的成员被随机分配至台式机/笔记本电脑网络调查(电子邮件邀请)或智能手机网络调查(短信邀请),Mavletova和Couper(2013)发现智能手机群体存在社会赞许性偏差,智能手机端受访者报告的饮酒量比台式机/笔记本电脑端受访者少,但在其他敏感问题上没有显示社会赞许性偏差。然而,在另一个相似的实验设计(例如随机分配在线样本库成员)中,研究者没有发现社会赞许性效应(Mavletova, 2013)。智能手机网络调查中的潜在社会赞许性偏差较高,部分是因为受访者使用智能手机时减少了隐私感,例如受访者可能在公共场所回答问卷。此外,由于手机是个人设备,手机用户对隐私的担忧可能会更大,因为人们认为手机可能会被识别。

关于智能手机端满足行为(敷衍作答行为)更多的证据尚不明确(Unclear

evidence on differential satisficing behavior on smartphones): 目前尚无明确证据表明智能手机端受访者出现满足行为的比率高于台式机/笔记本电脑端的受访者。许多研究测量了满意即可行为的不同维度, 并没有发现设备之间的差异。值得注意的是, Buskirk 和 Andrus (2013)、Mavletova 和 Couper (2013) 以及 Wells 等人 (2014) 在首呈效应方面没有发现任何差异; Mavletova (2013) 发现非实质性应答的水平没有差异; Wells 等人 (2014) 发现在使用“其他”类别方面没有差异; Buskirk 和 Andrus (2013) 发现项目无应答率上无差异。与上述不同的是, 项目区分度似乎受到问卷填写设备的影响。在一项针对大学本科生群体的网络调查中, Guidry (2012) 发现与台式机/笔记本电脑端受访者 (9%) 和平板电脑 (如 iPad) 端受访者 (8%) 相比, 智能手机端受访者的项目无区分度作答的比例翻了一番 (18%)。

然而, 研究者需谨慎考虑上述所有结果。需要注意的是, 针对智能手机的网络调查还处于初始阶段, 因此并非所有可能的优化方案都已经得到探索和整合。随着技术和方法的进步, 现有研究发现的设备效应可能会逐渐减弱。智能手机正在不断改进, 为用户提供更好的网络浏览体验。基于 Wi-Fi 接入点覆盖范围的扩大和对新一代移动宽带网络支持力度的提升, 智能手机的互联网接入速度也提高了。基于浏览器的智能手机调查的方法和设计优化也在迅速改进。

与此同时, 用户每天在智能手机上执行的任务数量和范围也在不断扩大, 因此通过移动设备参与调查很可能成为一种日常做法。

事实上, Toepoel 和 Lugtig (2014) 的研究结果可以提供强有力的支持。他们使用了一份基于浏览器的问卷, 需花费大概 10 分钟完成, 并完全针对智能手机进行了优化, 结果发现智能手机和台式机/笔记本电脑版本的问卷在应答率、中断断、项目无应答、填写完成时间、开放式问题的答案以及在多项选择题的回答数量方面, 都不存在差异。

对于屏幕尺寸约为 10 英寸 (25 厘米) 的大型平板电脑 (例如 iPad), 研究表明, 与台式机/笔记本电脑用户的中断率相似 (Callegaro, 2010)。其他研究发现, 在无应答行为方面, 平板电脑用户的表现与台式机或笔记本电脑用户更为接近, 而与智能手机用户存在明显差异 (Guidry, 2012)。综合上述因素来看, 平板电脑无疑是一项极具发展潜力的技术。事实上, 一些在线样本库已经开始探索为那些缺乏互联网接入条件的成员提供平板电脑, 以便他们能够参与调查 (例如法国的 ELIPSS 小组⁶)。然而, 随着小型平板电脑 (如 iPad mini、Nexus 7 等) 的大量涌现, 还需要更多的研究来考察潜在的设备效应。

另一在网络问卷作答中具有潜力的设备是交互式电视（ITV）。消费者可通过交互式电视进行投票、评论、购物、玩游戏或参与竞拍（例如拍卖、赛马、赌博等），甚至可以进行一般的网络浏览，因此理论上也可用于填写网络问卷。事实上，在 20 世纪初网络调查发展早期，知识边缘网络样本库的成员已通过公司提供的网络电视，借助特殊的遥控器和无线键盘完成调查（Huggins & Eyerman, 2001）。尽管在解决关键商业、技术和可用性问题后，交互式电视未来可能在调查设备中占据一定比例，但目前尚未达到这种程度。相应的方法学挑战尚不明晰，但我们可以预期，这些挑战与隐私问题和用于应答的设备扩展遥控器（例如指针，鼠标）的特征有关。

综上，截至 2014 年，现有研究证据仍较为有限。对于何种设备将在调查中占据主导地位、多设备非定向网络调查的影响等问题，尚无法得出确凿结论。阻止受访者使用移动设备来完成问卷（策略 2）可以解决这一问题，但可能会造成更多的受访者流失。此外，在网络调查中忽视或不适应新设备，则会限制受访者的自然行为环境。因此，建议根据不断变化的调查环境进行调整。下一节将为智能手机问卷设计提供具体建议。

5.1.4 为智能手机设计网络问卷

正如 Macer（2011）明确指出的，即便在今天，移动网络调查领域仍缺乏广泛共识，相关研究及最佳实践的实证依据也较为有限。然而，我们基于自身经验并总结 Callegaro（2013a）的研究结果，提出了一些设计层面的考虑因素。这些因素既涉及问卷设计，也涉及调查软件。一方面，问卷设计的决策包括问题措辞、问题类型的选择、邮件邀请的措辞等，研究者需要在特定网络调查软件中做出这些决策。另一方面，由调查软件驱动的考量因素同样重要，尤其是那些与软件特定功能相关的因素，它们会影响问卷在移动设备上的显示效果（Mavletova, 2013）。

我们提出以下调查设计的考量因素（*survey design considerations*），但这些因素仍应遵循一般问卷准备的原则（第 2.3 节），并且与有意和无意（*intentional and unintentional*）的使用行为相关：

- 在电子邮件邀请中，保持主题、内容和调查链接的简洁——在无意设计的情况下，这也会影响非智能手机用户。冗长的主题和内容会导致显示多行文本，需要受访者滚动查看。此外，尽快使调查网址链接可见也很重要。网址应尽可能简短，如果可能的话，避免跨多行显示。尽管可使用网址缩短服务，但需注意

其可能隐藏域名合法性的问题。

- 在问题页面中，删除或减少所有非必要的、与问题无关的内容。例如，标志、免责声明和帮助链接，此类信息可以仅限在介绍页显示，以便为问题本身提供更多空间。由于进度条占用大量空间，却不能真正减少中长篇调查的中途退出情况，可以考虑避免使用（Villar et al., 2013）。
- 避免使用表格形式的问题（例如单选按钮表格）。表格不适用较小的屏幕，需要受访者缩放和水平滚动才能阅读文本。考虑为每个项目设置成一个单独的问题。如今，大多数调查软件在智能手机上显示表格时，已经自动将其展开为一系列独立问题（*unfold tables*）（Pferdekaemper & Batanic, 2009）。另一种解决方案是将表格自动转换为一系列下拉菜单问题，每个问题都放置在单独的一行中。
- 将复杂问题分解（*Decompose*）并使用具有标准应答格式和布局的单一项目问题类型（例如单选按钮）。许多高级问题类型（例如排序）可能无法在小屏幕上很好运作。
- 对于拖放问题（例如排序）要谨慎使用，并进行充分测试。因为在触摸屏和其他小屏幕上，要流畅地完成这类问题极其困难。
- 考虑将双极量表拆分为（*splitting*）两个问题。首先是两个主要选项（例如满意或不满意），然后是每个选项中的程度（例如极其、非常、一些和有点）。Malhotra 等（2009）举例说明了有 7 个或更多回答选项的问题。在智能手机上，由于问题文本的长度，不能显示超过 5 分的评分。
- 避免让受访者进行水平滚动。除了上述修改（例如分解、拆分、展开），还可以采用将水平量表转换为垂直量表、限制文本区域大小和其他一些方法。
- 遵循西方语言文化的“向上意味着好”的逻辑安排垂直评分，使顶部的应答选项具有积极含义，而底部的应答选项具有消极含义（Tourangeau, Couper, & Conrad, 2013）。
- 考虑不要在问题文本中提及应答选项（仅应答文本中提及），以节省宝贵的页面空间。
- 尽可能避免使用视频。因为视频可能出现兼容性问题，并且加载时间较长。对于（大尺寸）图片应谨慎使用。

虽然上述修改大多是基于研究者深思熟虑的决定，不受网络调查软件的限制，但以下与调查软件相关的考量因素（*survey software considerations*）通常无法由研

究者自行完成，而是需要修改网络调查软件：

- 优化字体大小和导航栏以适用于智能手机。这还包括可能放弃使用标准的超文本标记语言（HTML）元素，转而使用图形，例如用图像代替单选按钮或用图片取代应答选项的语言描述。
- 确保不使用 Flash 技术，因为运行 iOS 操作系统的智能手机和平板电脑（例如 iPhone、iPad）将无法支持它。
- 定制提示信息和错误信息，使其适合智能手机的字体大小和长度显示。
- 针对小屏幕优化图片大小、调整文本区域大小。提供将网格自动展开为水平滚动矩阵的功能。
- 当询问开放式问题时，考虑为智能手机用户提供语音记录答案的选项。在市场调研实践中已经有一些成功的应用案例。

这些只是一般性的考虑和建议。重要的是要在多种设备和相关操作系统上全面测试问卷，以充分发现可能出现的问题。我们还建议向网络调查软件的开发人员询问他们是如何处理智能手机和平板电脑操作的，并获取他们为这些设备提供的自动调整列表。

上述许多考量因素与 Nielsen 和 Budiu（2013）提出的移动网站可用性指南不谋而合。例如，在为多种设备设计网络问卷时，区分鼠标和手指这两种输入设备十分重要。两者主要区别在于鼠标的精度远高于手指触摸；鼠标不会遮挡屏幕的视野，而手指触摸则会；最后，鼠标通常至少有两个按钮和一个滚轮，而一根手指甚至两根手指对网页的控制要有限得多。因此，在移动网络调查中，问卷交互元素（例如单选按钮或复选框）的大小和位置应该是不同的。此外，关于屏幕尺寸的限制，Nielsen 和 Budiu（2013）建议设计无需水平滚动的网站，保持下拉菜单简短，确保文本不被截断，正确处理错误信息和提示信息，并对移动网站进行恰当的格式设置。如上所述，这些建议中的每一条都能成功应用于智能手机上的网络调查。尽管“移动优先”的设计理念已被提出，遗憾的是，目前可用性设计与移动网页调查设计之间的跨领域借鉴仍较为有限，这一现状有待改善。

5.1.5 基于浏览器的移动调查 vs 移动调查应用程序

到目前为止，我们主要讨论了基本的网络调查问卷，这类问卷可以通过网络浏览器在多种设备上完成。我们将这种调查方式称为基于浏览器的移动调查（策略 3a

和 3b)。然而,如上所述,网络调查也可以通过使用专门的移动调查应用程序(策略 4 和 5)在移动设备进行回答(例如智能手机、平板电脑)。移动设备上的软件应用程序通常简称为应用程序(*apps*),因此我们在此之后将其称为移动调查应用程序(*mobile survey apps*)。

随着技术的进步,基于浏览器的移动调查变得越来越强大,这一趋势可结合当前广泛发展的自适应设计(*responsive design*)和自适应式网站(*responsive websites*)背景来理解。这些网站会根据用户使用的具体设备自动调整界面布局。这极大地增强了基于浏览器的移动网络调查的固有优势:

- 受访者无需下载任何东西,就可以从电子邮件邀请中直接回答问卷。当然,对于在线样本库而言,这一优势就不那么突出了,因为在线样本库通过专门的应用程序发送邀请会更加便捷。
- 使用标准超文本标记语言(HTML)元素的问卷能够实现更高的覆盖率,因为它们不依赖于不同设备、平台和浏览器的操作系统。
- 仅针对台式机/笔记本电脑或智能手机/平板电脑管理特定问卷,比为各类设备、平台和浏览器开发和协调平行调查问卷要容易得多。

然而,目前基于浏览器的移动调查仍然面临某些限制,这些限制往往使得采用移动调查应用程序更合适:

- 互联网连接是回答问卷的必要条件。
- 一些用于支持高级功能的技术可能无法在移动设备正常工作,甚至根本无法运行(如 JavaScript 和 HTML5,更不用说 Java 和 Flash)。
- 为了充分利用设备的硬件功能,例如获取位置信息或拍照,受访者需要接受来自智能手机浏览器的提示,同意相应权限请求;而使用应用程序时,这可能只需要在应用程序安装期间授权一次。
- 可能无法完全支持移动功能的集成,例如设备运动检测、二维码扫描、GPS 定位或摄像头使用。
- 可能缺乏某些调查所需的灵活性(例如时间使用日记中定期的铃声/提示信息)。
- 需要相应支持的与基于位置的服务相关的调查。这类调查中,问卷的呈现是由受访者的位置触发的(例如,在离开商店后或接近某个地点时)。这有时也在地理围栏(*geofencing*,基于 GPS 或移动运营商位置识别技术)或信标(*beaconing*,基于蓝牙技术)的背景下进行讨论。
- 需要支持在不同地点收集连续报告和照片的调查,如在民族志研究中。

另一方面,受访者需要通过网络(例如苹果应用商店、谷歌应用商店)将移动调查应用程序下载到移动设备上。这可以针对特定的调查问卷单独完成(策略4),或者可能有通用的软件应用程序,只需要完成一次安装,后续就可以通过移动设备上的该程序接收和完成更多的调查问卷(策略5)。

移动调查应用程序非常重要的区别在于两项基本功能:问卷显示和数据管理。在基于浏览器的移动调查中,这两项功能都是由服务器和网络浏览器运行的。当然,用台式机或笔记本电脑的基本网络调查模式也是如此。然而,移动调查应用程序的情况有所不同,因为应用程序可以控制这两项功能:

- 问卷显示 (*Displaying the questionnaire*): 当移动调查应用程序控制调查问卷的布局时,它可以依托于设备上本地安装的互联网浏览器,或者依托于模拟浏览器功能的应用程序。后者是非常简单的功能,称为超文本标记语言容器 (*HTML container*),可以为移动设备屏幕重新调整 HTML 元素的显示方式。例如,它可以作为一个独立的功能设备(而不是浏览器的扩展)将水平布局切换为垂直布局,或者调整单选按钮的大小。在另一种极端情况下,该功能可以扩展为完全不受浏览器限制和 HTML 元素的约束。它可以像在高级移动应用程序中一样,完全控制问卷的呈现方式,例如脸书 (Facebook)、谷歌邮箱 (Gmail) 或优兔 (YouTube) 等应用程序,它们极大地改善了移动图形用户界面(尽管数据交换仍然依赖于服务器)。移动操作系统为移动应用程序开发人员制定的指南可用于此类目的 (Owen, 2013)。
- 数据管理 (*Data management*): 即便在高级移动调查应用程序中,当应用程序不支持数据管理时,我们仍然需要基于服务器且全程保持互联网连接的网络调查。只有当数据管理成为应用程序的集成功能时,移动调查应用程序才会一定程度上摆脱与网络服务器的持续交互。在极端情况下,这样的应用程序可以作为客户端独立于服务器运行计算机辅助自填式问卷 (CSAQ),仅在问卷调查结束时或当研究者触发相应程序时,才将应答提交给研究者的服务器。

支持在移动设备上显示问卷是移动调查应用程序的关键特征,我们将仅具备这一特征的应用程序称为被动式移动调查应用程序 (*passive mobile survey app*); 对于同时支持数据管理的情况,我们称之为主动式移动调查应用程序 (*active mobile survey app*)。两种类型之间的主要区别在于,前者的数据连接是同步的(始终在线),而后的数据连接是异步的(在线或离线均可)。严格意义上讲,主动式移动调查应用程序已经超出基本的网络调查模式范畴,因为问卷填写无需网络浏览器,也无

需与服务器交互（见 1.1 节中基本网络调查模式的定义）。在这种情况下，网络仅用于交换空白问卷和已完成的问卷，所以主动式移动调查应用程序属于网络关联型调查模式（*web-related survey modes*）的范畴（第 6.1.1 节）。

依据 Macer（2011, pp. 9-10）的观点，移动调查应用程序具有以下优势（主要涉及主动式移动调查应用程序，不过随着基于浏览器的移动调查的进步，被动式移动调查应用程序正在日益扩展其功能）：

- 受访者可以在没有互联网连接的情况下填写调查问卷，只需要下载问卷并提交时连网，且提交可以在回答环节结束后立即完成，也可以稍后完成。
- 应用程序可以整合自身的交互功能，如提示、发送消息、触发警报（例如振动或发出蜂鸣声），并且即使没有互联网连接也可以保持“活跃”状态。
- 应用程序可以充分利用设备的功能，如记录 GPS 定位数据、拍摄照片（例如某一调查问题需要受访者拍照）、录制视频、语音录音以及扫描条形码或二维码。

无论何种类型，移动调查应用程序总体上更适用于各类在线样本库，尤其是旅行研究、位置服务研究、多媒体数据收集（语音、图片、视频）以及像日记记录、生物标志物采集和生态瞬时评估（EMA）等持续报告。关于主动式移动调查应用程序，需要补充的是，它们对于计算机辅助个人面访（CAPI）以及在各种业务流程中收集数据也非常便捷。

另一方面，移动调查应用程序也有缺点：

- 在回答调查之前，受访者必须将应用程序下载并安装到设备上。
- 每当更换设备时，受访者需要重新安装应用程序，而这种情况经常发生。这增加了受访者的负担，并可能引发新的安全顾虑。
- 应用程序需要针对每个特定的操作系统进行编程和设计（至少针对苹果系统和安卓系统，但也可能包括 Windows phone 系统）。这大大增加了成本和开发时间，至少对于定制移动调查应用程序如此（策略 4）。对于通用应用程序（策略 5），这一问题相对轻微，但仍需额外投入，因为通常还需同步开发基于浏览器的网页调查。
- 同时实现通过应用程序和网络浏览器管理问卷可能更复杂，要求更高。

考虑到当前的技术现状，我们可以得出结论，当基于浏览器的移动调查确实面临上述局限性时，移动调查应用程序会是合适的选择。然而，在未来，当对高级脚

本和 HTML5 的支持更加广泛时，基于浏览器的移动调查能够处理比现在更高级的功能。事实上，随着基于浏览器的移动网络调查不断模拟移动调查应用程序的丰富功能，两者之间的界限将日益模糊。Buskirk 和 Andrus（2012）将这一趋势中的一种实现方式称为类似应用程序（*app-like*）的基于浏览器的移动调查，其运行依赖 JavaScript 和 HTML5。

5.1.6 多设备网络调查软件和移动调查应用程序

如前所述，在撰写本书时，标准的网络调查软件尚未完全支持多设备调查。WebSM 调查软件研究（WebSM, 2014）表明，2014 年初，许多主流网络调查软件供应商，对基于浏览器的移动调查支持仍然有限。例如，不少软件即便在基础功能上也存在问题，如无法自动展开表格或自动将水平量表转换为垂直量表。然而，正如上文所述，这些问题将在几年内逐渐消失。在不久的将来，通过浏览器激活的相关客户端脚本可能会将基于浏览器的移动调查提升到更高水平，届时对移动调查应用程序的需求可能会大幅减少。

在通用移动调查应用程序（即用于创建这些应用程序的通用开发工具）的背景下，对灵活问卷开发的支持也在快速提升，但相较于已有的网络调查软件功能薄弱。后者主要用于为台式机 / 笔记本电脑创建基于浏览器的标准网络调查问卷，功能很丰富。因此，复杂的问卷可能仍需作为定制化移动应用程序，针对不同移动平台分别进行编程开发。

尽管如此，移动调查应用程序的数量正在迅速增加。2012 年 7 月，研究期刊 *Quirk's* 列出了 32 家专门从事移动网络调查的供应商（Townsend, 2012）。近期对苹果和安卓应用商店中可用移动应用程序的梳理显示，与网络调查相关的应用程序超过 1 000 个（WebSM, 2014）。虽然其中大多数针对在线样本库中的受访者，但也有许多主动式和被动式的移动调查应用程序与成熟的网络调查软件配套使用。此外，还有一些用于在移动设备上创建简单网络调查（例如用于投票）的应用程序，具体将在下文论述。

除了供受访者使用的各类移动调查应用程序，我们还必须明确区分那些为研究者提供支持的应用程序。具体而言，许多网络调查软件供应商已经开发了专门的移动调查编写应用程序（*mobile survey authoring apps*），支持在移动设备上创建基于浏览器的网络问卷。通常（但并非总是如此），受访者仍然可以通过网络浏览器访问这类问卷。为研究者提供专门的移动应用程序主要原因在于：网络调查软件的

用户界面非常复杂,几乎无法在移动设备上使用。因此,如果研究者希望在移动设备上快速创建简单调查(例如反馈表),这类应用程序可提供经过适配的特殊界面。在该界面中,导航和选项相应地进行了优化,这非常有帮助。当然,这类应用作为独立工具运行,与供受访者使用的移动调查应用程序并无直接关联。

鉴于与调查相关的移动应用程序种类繁多,在讨论不同类型的应用程序时,需格外谨慎。因此,我们建议使用明确的附加标签来区分是在讨论主动式还是被动式的移动调查应用程序,定制型还是通用型移动调查应用程序,以及该应用程序面向受访者还是研究者使用。

5.1.7 挑战和结论

尽管自首次尝试短信(Cooke, Nielsen, & Strong, 2003)和无线应用协议(WAP)调查(Fuchs, 2008)以来,基于智能手机和其他设备的调查已经取得了长足的进步,但许多挑战仍然摆在面前。在发达国家,我们预期智能手机将在未来几年内实现几乎全面覆盖(90%或更高)。此外,在大多数国家中,互联网普及率的增长主要是依托智能手机来实现。在非洲和南美洲的许多国家,用户正在跳过通过台式机/笔记本电脑使用的互联网阶段,直接通过智能手机接入互联网。⁷

随着智能手机和平板电脑的普及,受访者会“未经我们许可”就使用手头的任意设备来回答问卷,这一趋势愈发明显(Peterson, 2012)。如今,很明显,试图阻止或引导受访者使用其他设备的做法是行不通的。未来唯一可行的解决方案是谨慎规划多设备网络调查。这种模式的转变既包括调查平台驱动的设计决策,也包括涵盖问卷设计、内容和电子邮件邀请等的调查设计决策。

然而,阻碍移动网络调查更广泛应用的障碍仍然存在。有时,部分用户需支付数据传输费用,因此研究者需始终确认目标人群是否能真正免费通过移动设备参与调查。此外,移动浏览器的功能仍然落后于台式机浏览器,而使用专门的移动调查应用程序对受访者和研究者来说均不够便捷。不过,我们可以预期,随着互联网连接可用性的提升、成本的降低,稳定且几乎随时可用的网络连接将成为常态;同时,随着类应用程序的移动浏览器方案的逐步成熟,多数技术问题将得到解决。

台式机/笔记本电脑端基本网络调查模式(第1.3.3节)的发展经验表明,只有在关键的技术问题得到解决之后,方法学层面的问题才会凸显重要性。与台式机/笔记本电脑端网络调查的发展过程类比,网络调查发展的第一个十年(1994~2004)的动荡通常与互联网速度的提高——互联网速度长期受到拨号上网限制——以及客

户端脚本（例如 JavaScript）的改进有关。在随后的十年（2005~2014）中，技术变得更加稳定，技术变革更少，因此方法论效应与技术细节的关联也随之减弱。尽管移动网络调查可能会遵循类似的路径，但相关方法学问题的重要性已得到关注和研究。这也涉及是使用基于浏览器的调查还是使用移动调查应用程序的两难困境。随着技术进步，这一困境可能逐渐消解，因为对于受访者来说，这两种选择之间的差异会越来越小。（本书将在 5.3 节关于软件的总结中进一步讨论与移动调查相关的软件支持及其未来趋势。）

在对各类研究的回顾中（第 5.1.3 节），移动设备调查面临若干方法论难题，尤其是体现在无应答率、中断、问题类型、问卷长度和所需时间方面。然而，值得注意的是，几年前用老式设备进行网络调查，由于技术限制造成的研究方法缺陷，在今天已得到解决。此外，需要非常谨慎地解释使用不同设备进行调查的研究结果，因为即使在实验中，也很难将被试随机分配到不同的设备组别。智能手机的覆盖率可能远低于台式机 / 笔记本电脑，如果受访者自行选择设备（这一点很难切实避免），可能会出现显著的自我选择偏差。在这种情况下，设备特征与用户特征相互干扰，导致难以确定效应的来源。此外，设备效应也可能与其他方面（例如无应答）相互混淆。因此，分离并探究移动设备的独立影响极具挑战性。

另一个需要谨慎处理的问题，与使用移动设备进行调查时可能面临的环境多样性有关。周围有其他人在场、所处位置不方便以及同时处理多项任务等情况，都可能对应答质量产生影响。此外，移动网络调查，特别是通过调查应用程序时，能够从受访者设备收集其他类型的数据（例如位置或互联网流量行为）。因此，为解决隐私和安全担忧，要与受访者进行清晰、公开的沟通，这至关重要（见第 6.3 节）。

综上所述，越来越多的受访者将通过移动设备访问调查问卷是不可避免的趋势。这就需要大量新的方法学研究和资源投入，以保障高质量网络调查的开展。尽管该领域研究尚处于起步阶段，但本节梳理了若干关键原则，并从现有研究发现中提炼出一些初步见解。

最后，需强调的是，移动调查只是移动研究的一部分（Poynter, Williams, & York, 2014）。在市场调研（以及相关的网络调查软件）领域，新的移动调查解决方案不断涌现，例如在短信或电子邮件调查中的创新形式。

5.2 在线样本库

在线样本库⁸主要为了拥有一组预先招募的受访者而建立的。这些受访者可以随时准备接受网络调查的邀约并及时响应网络调查。他们由一组注册的受访者组成,在注册时已同意定期参与网络调查。样本库提供商投入大量资源来招募和维护他们的样本库,以便使整个数据收集过程更加高效,在几周内、甚至几天或几小时内获得特定调查的数据。在某种程度上,在线样本库是消费者(或家庭)调查小组的自然演变,过去这些调查是通过邮件、电话和面对面的方式进行(Wansink & Sudman, 2002)。

此处使用广义上的“在线”(online)一词,类似于本书在第1.3.1节中定义在线调查、互联网调查和网络调查之间差异的方式。此外,术语样本库(panel)使用非常具体,与前面第1.1.1节中概述的含义一致,即可以反复被包括在各种调查中的潜在受访者抽样框。另一方面,对“样本库”一词更常见的理解与纵向调查有关。纵向调查是对同一组实际受访者就同一主题进行反复调查(例如劳动力调查),但本节所指的在线样本库通常并非如此。

我们首先概述与在线样本库的术语(第5.2.1节)、招募(第5.2.2节)和维护(第5.2.3节)相关的一般问题。接下来,我们将讨论在线样本库设计(第5.2.4节)和数据质量(第5.2.5节)的具体内容。最后,我们讨论购买在线样本库研究(第5.2.6节)和在线样本库的未来(第5.2.7节)。

5.2.1 术语与分类

“在线样本库”是一个广义术语,涵盖多种不同类型。不同样本库在成员组成、数据收集类型以及数据收集方式等方面存在差异。根据成员构成,最常见的三种在线样本库是普通人群样本库、专业样本库和专属样本库(Baker, et al., 2010, p. 8)。普通人群样本库(*General population panels*)是最常见的小组类型,会尽量纳入所有可能类型的受访者,包括难以接触到的人群。该样本库被用作一个抽样框,研究者从中抽取符合客户需求的特定样本。专业样本库(*Specialty panels*)是通过招募特定类型的人群而建立。这些人因为年龄、其他人口统计数据(例如拉丁美洲人小组)或因其职业(例如医疗专业人员小组)而被招募,或者针对特定的人群招募(例如选举样本库,符合条件的选民在选举前后的几个月里被招募并接受调查,以研究

其态度形成过程)。专属样本库 (*Proprietary panels*) 是专业样本库的一个子类, 其成员参与特定公司的研究。例如, 研究者可以使用“某一特定类型产品所有者的电子邮件地址”来构建专属样本库。近年来, 上述样本库也与在线社区样本库 (*online community panels*) 有相当大的重叠, 后者的样本库成员共享共同的特定虚拟空间。

在线样本库可以通过两种方式收集数据: 一是网络调查问卷 (即在线调查样本库), 二是被动收集受访者的在线行为。后者被称为互联网评级样本库 (*Internet ratings panels*, Napoli, Lavrakas, & Callegaro, 2014)。它们主要是根据市场占比、访问者人口统计特征和在线广告等方面来测量互联网受众。与在线调查样本库的不同之处在于, 互联网评级样本库还涉及另一个维度: 在线行为的被动测量。为了测量这种行为, 开展互联网评级的公司要求样本库成员在其设备上安装一个在线软件检测器, 随后, 以被动的方式追踪受访者计算机上的所有互联网活动。

本节的其余部分只关注在线调查样本库 (简称在线样本库), 即使用网络调查问卷收集数据的样本库。

5.2.2 招募、抽样和选择

在线样本库供应商之间的主要区别在于招募方式。根据本书对网络调查类型的初步分类 (参见第 1.1.3 节), 在线样本库可以分为两种类型: 非概率在线样本库和概率在线样本库。

非概率在线样本库 (*Non-probability online panels*) 也被称为接入式、自主选择参与的志愿者样本库, 其成员招募涉及想要加入样本库的个体自我选择的过程。几乎任何人都可以选择成为这类样本库的成员。样本库成员也可以邀请朋友和家人加入, 或推荐其他人。从抽样的角度来看, 由于招募的自愿性质, 样本库成员被选中的概率未知、无法确定。需要说明的是, 此类样本库不包含非互联网用户群体。

概率在线样本库 (*Probability online panels*) 采用概率抽样机制招募样本库成员, 为目标群体中的每个成员提供已知且非零的入选概率。通常使用线下方式 (电话随机拨号、邮件和面对面招募) 进行初始招募。如果被调查者没有通过上述任何一种方法被选中和招募, 则他们不被允许自愿加入样本库。样本库成员不能推荐朋友、家人或同事加入。为了能够代表整个人群, 概率在线样本库为那些无法连接互联网的招募成员提供设备和 / 或互联网接入 (DiSogra & Callegaro, 2015)。

5.2.2.1 非概率在线样本库的招募方法

建立非概率在线样本库的研究者和公司通常不会透露他们的招募方法,因为他们视之为竞争优势(Baker et al., 2010)。出于这个原因,关于招募方法的公开信息很少。以下内容基于少量资料来源(Baker et al., 2010; Comley, 2007; Postoaca, 2006)以及我们与供应商打交道和浏览在线样本库网站的个人经验。

线下招募方式(*Offline recruitment methods*)包括线下样本库招募、线下调查结束时招募、从现有邮寄或电子邮件地址营销数据库中招募、通过电话、邮寄或面对面直接招募,以及借助线下媒体曝光、散发传单等方式招募。

线上招募方式(*Online recruitment methods*)比线下招募成本更低(见第2.2.3节),且更具多样性。最简单的招募方法是让在线样本库的志愿者直接在其网站或样本库门户网站注册;也用横幅广告来招募潜在的志愿者。这些广告将潜在的受访者引导至可以自行注册的网站。横幅广告可以针对特定的网站,从而根据用户的兴趣招募特定群体。在新闻组或邮件列表中邀请受访者加入样本库也很常见,且成本较低。搜索引擎广告(Nunan & Knox, 2011)被越来越多地用于招募在线样本库的潜在成员。在线社交网络(如脸书、推特等)是另一种招募潜在成员的方式。联合注册协议是当用户注册某项在线服务时,同时提议加入在线样本库。门户网站、电子商务网站、新闻网站和特殊兴趣网站都是共同注册协议的典型案例。联盟中心,也被称为在线样本库整合平台,通常以承诺金钱作为奖励,同时将个人注册到一个或多个在线样本库的网站。“成员推荐成员”机制通过提供激励措施,让现有的样本库成员招募他们的朋友和家人。在网络调查结束时招募成员是建立样本库的另一种方式。最后,向购买的邮件列表发送未经请求的电子邮件是另一种方法,而信誉良好的公司(至少在美国和欧洲)较少使用这种招募方式(Baker et al., 2010, p. 721)。

5.2.2.2 概率在线样本库的招募方法

概率在线样本库使用传统的线下方法来招募潜在成员(Callegaro & DiSogra, 2008; DiSogra & Callegaro, 2015)。此类样本库的概率样本可以是基于地址的抽样(ABS)、人口登记抽样、区域概率抽样或随机拨号抽样。潜在的样本库成员可以通过提前邮寄或直接由调查员接触的方式联系。例如,荷兰的LISS样本库向潜在家庭预先邮寄信件和样本库宣传手册。对于可以匹配到电话号码的地址,调查员通过电话联系该家庭。对于无法匹配到电话号码的地址,调查员则亲自入户联系潜在的样本库成员(Scherpenzeel & Das, 2010)。在美国,GfK(原 Knowledge Networks)

使用基于地址的抽样框 (DiSogra et al., 2010) 向抽样家庭邮寄招募方案资料包, 其中包括如何通过邮件、电话或网络加入样本库的说明 (Knowledge Networks, 2011)。EKOS¹ (加拿大) 和盖洛普样本库 (美国) (Tortora, 2009) 采用随机拨号抽样 (RDD) 招募样本库成员。最后, 美国的面对面 / 互联网调查平台 (FFRISP) 则通过基于地址的抽样 (ABS) 进行面对面招募 (Krosnick et al., 2009)。

为了使样本库能够代表全体人口, 需要使用不同的方法来联系或调查没有互联网接入的家庭。GfK 和 LISS 样本库为这些非互联网家庭免费提供计算机并支付互联网连接费用。其他公司, 如 EKOS 和盖洛普样本库 (DiSogra & Callegaro, 2015) 或德国的 Gesis 样本库 (Bosnjak et al., 2013), 通过电话和 / 或邮寄方式调查线下人群。

第三种方式是所有样本库成员 (而不仅仅是那些没有互联网接入的人) 提供相同的设备和互联网连接。例如, 美国 FFRISP 为所有成员提供笔记本电脑和宽带连接 (Krosnick et al., 2009); 法国 ELIPSS 则为其所有成员提供带有数据连接的平板电脑 (Legleye & Lesnard, 2013)。

最后, 样本库公司可能会不招募没有互联网接入的家庭。例如, 如果特定国家的互联网普及率非常高, 可以通过适当的加权调整来弥补覆盖误差 (如冰岛社会科学研究所的在线样本库)。

5.2.2.3 在线样本库成员的抽样与选择

鉴于在线样本库提供了由预先招募的受访者组成的抽样框, 公司究竟如何选择受访者使其成为特定研究的一部分? 概率样本库一般采用传统的概率抽样方法, 如简单随机抽样或分层抽样。

非概率在线样本库通常使用配额抽样方法 (Rivers, 2007)。这需要在关键子群体 (通常依据人口统计学特征, 但有时也依据行为特征) 中设置配额或最大数量的受访者。在完成特定问卷的过程中实时执行配额, 一旦某个配额已满, 而符合该组别要求的新受访者将被筛选排除。当然, 大型样本库, 尤其是拥有数百万成员的样本库, 处于更有利地位。

部分样本库机构还采用更复杂的抽样方法以提升代表性。所有这些新方法往往都是专有的。其中最著名的是 YouGov 公司开发的样本匹配 (*sample matching*) 方法 (Rivers, 2007)。该方法首先使用商业数据库 (在美国, 如 Acxiom、Experian 和 InfoUSA, 或者涉及选举投票时使用选民登记名单) 中的抽样框对目标人群进行枚

¹ <http://www.probit.ca/>

举。然后,从抽样框中抽取随机样本,并将其与具有相同选定特征的样本库成员进行匹配。为增加获得应答的可能性,为抽样框中的每一行选择多个样本库成员。该方法适用于样本库机构开展的所有同规格开放研究。如果某一样本库成员进入某项研究,而另一个与之最为匹配的人已经完成该研究,他或她会被实时重新定向到次优匹配的研究,不会被拒之门外。

其他使用的方法有 Toluna 的倾向得分选择 (Propensity Score Select) 或智能选择法 (SmartSelect) (Terhanian & Bremer, 2012) 以及全球市场洞察公司 (Global Market Insight) 的 Pinnacle 法 (*Pinnacle methodology*) (GMI, 2012)。Callegaro 等人 (2014) 对此进行了详细的解释。不过,这些替代性的抽样和选择方法尚未经过广泛测试,关于它们的调查研究文献也寥寥无几。

5.2.3 维护在线样本库

在线样本库一旦建立,概率与非概率样本库的维护方式存在共通性 (Callegaro & DiSogra, 2008)。由于样本库具有动态变化特征,其维护需要持续投入大量工作。正如 Callegaro 和 DiSogra (2008) 所阐述的,样本库始终处于动态调整中。大多数样本库会持续招募新成员,以应对各类成员流失问题,如自愿流失、被动流失、达到任期退出和因死亡导致的流失等。

自愿流失 (*voluntary attrition*) 是成员主动联系机构并申请退出的行为。这种情况并不常见。更为常见的是,成员在不通知机构的情况下停止参与调查或更改电子邮件地址。在这些情况下,我们称之为被动流失 (*passive attrition*)。一些样本库机构还设定成员的最长任期 (例如两年),当成员达到任期时,他们就会“退休” (*retire*)。最后,当成员离世或因身心原因不再具备应答能力时,会出现死亡流失 (*attrition by mortality*)。

明确特定时间点的活跃成员数量,是衡量在线样本库质量的关键指标。但不同供应商对“活跃样本库成员”的定义及统计周期差异很大。国际标准化组织 (ISO) 的 26362 标准 (ISO, 2009) 将活跃样本库成员定义为“在过去 12 个月内至少参加过一次被邀请的调查,更新过他/她的个人资料数据或注册加入可访问样本库⁹ (access panel)” (p. 2)。

各机构管理在线样本库被动流失的方式存在很大差异。更积极的样本库管理策略,如从数据库中清除不太活跃的成员,从而提高每次调查的参与率 (Callegaro & DiSogra, 2008)。这种做法确实是有代价的,会减少活跃样本库的规模,并且存在风

险,即更活跃成员的应答行为可能与普通成员存在差异。纵向研究设计需要对同一批样本库成员进行长期追踪调查,人员流失尤为突出。在这种情况下,有效样本量减少可能限制特定分析的开展,而潜在的流失偏差还会影响数据质量。为此,样本库提供商可以通过特定举措来解决流失问题,并尝试减少流失(Kruse et al., 2010),例如重新招募离开在线样本库的成员,制定激励计划以使成员停留更长时间。

激励策略对减少潜在流失至关重要。通常,机构会通过消费者积分计划为完成问卷的成员提供奖励。一般来说,每一分钟的回答时间对应的激励金额为几美分或几十美分。虽然金钱激励最有效,但也可以使用其他方式,如彩票、随机抽奖、慈善捐赠或象征性礼物(Göriz, 2010)。对于在线社区而言,各种活动都可以用作激励措施,包括在线或线下活动。

受访者的负担也会影响流失。很难确切知道样本库机构平均发送多少份调查邀请。在在线样本库门户网站上进行注册,并以样本库成员的身份监控调研活动,可以获得相关信息,但前提是该公司准许监控。Grey Matter Research公司采用过该方法。公司员工和其他志愿者注册了11个不同的美国在线样本库,并监测了三个月的流量(Grey Matter Research, 2012)。结果显示,这些调查的时长不同,平均时长为17.3分钟,而平均每月收到20.4份邀请,范围从6.5份到最多51份不等。

5.2.4 在线样本库研究设计

在线样本库可用于横断面研究和纵向研究(见第1.1节中的定义)。在横断面研究中,成员参与各种不同主题的调查。这些调查可以是一次性的(横断面),也可以是重复横断面设计,即同一项调查可以多次进行,也称为追踪研究(*tracking studies*)。重复横断面研究是针对同一个主题进行的,但每次使用的受访者样本不同。经典案例是品牌知名度研究,即一家公司长期监测其品牌知名度。在线样本库也可以用于纵向(*longitudinal*)研究,即在不同时间点对同一主题的相同受访者进行调查,以研究随时间发生的变化(Göriz, 2007)。这种类型的设计最接近传统意义上的家庭样本库概念。在此类设计中,每次新的调查被称为一波/轮(*wave*)。除追踪随时间的变化外,在不同的时间点重复提问相同的问题,还可以用于研究项目的效度和信度(测试—重测, *test-retest*),从而提高整体数据质量(Sikkel & Hoogendoorn, 2008)。此外,纵向设计也可整合横断面研究的成分,例如部分基础问题仅询问一次,其他问题则在不同时间点重复测量。

通常,在线样本库会通过普查式调查收集全体成员的特征信息,这一策略被称

为样本库特征分析 (*panel profiling*) (Callegaro & DiSogra, 2008)。此类调查主要聚焦于相对稳定且不会频繁变动的特征、态度与行为,例如人口统计信息(在开始或加入小组时收集)、健康状况、一般政治倾向和主要购物者身份。这种特征 (*profiling*) 分析有几个优点:(a)避免在每次客户委托的研究中重复提问,从而缩短问卷长度;(b)先前收集的数据可以帮助对特定研究进行问卷分支设计,提高效率;(c)客户可以直接购买概况调查的数据,无需另行开展调查;(e)当客户希望对特定目标人群(例如患有某种疾病的受访者)进行研究时,供应商可以根据特征分析数据抽取受访者样本,这称为目标抽样 (*targeted sampling*) (Sikkel & Hoogendoorn, 2008) 或预筛选 (*pre-screening*, Callegaro & DiSogra, 2008)。有效的样本库管理需为每位受访者维护多维度数据集,通过降低调查负担提升整体效率。

在线样本库的另一个优势(与特征分析的概念密切相关)是能够接触到小众人群和稀有人群,而这类人群的数据若通过其他方式收集往往需付出极高成本(Boyle, et al., 2013; G. Wright & Peugh, 2012)。

5.2.5 在线样本库的数据质量

在衡量概率与非概率在线样本库的数据质量时,本节重点关注两个核心指标:将点估计值与基准值(黄金标准)对比,以及变量间关系的稳健性¹⁰。在比较点估计值时,讨论选举民意调查估计值这一特殊情况非常有趣。此外,本节还将简要涉及其他质量问题,并专门讨论在线样本库中备受关注的权重调整问题。

5.2.5.1 在线样本库数据质量和结果的比较

自在线样本库研究开展以来,典型的数据质量研究是将在线样本库的结果与已知基准数据进行比较。这些基准来自官方统计的人口普查数据,或者是其他一些高质量的官方统计调查基准数据。鉴于这方面的研究数量众多,我们重点关注规模较大的比较研究,请参阅 Callegaro 等人(2014c)的著作章节以获取更多细节。

这类研究的常见设计是在多个在线样本库中同步开展相同问卷调查,再将结果与已知基准进行对比。首个针对不同非概率样本库进行商业化大规模比较的是,荷兰国家在线样本库比较研究(NOPVO)(Vonk, van Ossenbruggen, & Willems, 2006)。该研究对来自荷兰的19个不同在线样本库的1000名受访者进行相同的问卷调查,并比较了调查结果。作为研究设计方面的里程碑,NOPVO的设计后来被其他国家的商业和学术机构多次采用。例如,广告研究基金会(ARF)启动了质量基础1

(FOQ1) 倡议 (Walker, Pettit, & Robinson, 2009), 该倡议比较了 17 家美国非概率样本库供应商以及一个随机拨号电话样本库和一个邮寄样本库。2012 年, 美国广告研究基金会 (ARF) 的 FOQ2 项目 (Advertising Research Foundation, 2014) 以及加拿大市场营销研究与情报协会 (MRIA) (Chan & Ambrose, 2011) 在本国 14 个样本库中重复了类似的研究。上述三项研究 (NOPVO、ARF 和 MRIA) 得出一致的主要结论: 在线调查样本库之间不可以互换, 在不同在线样本库上进行的相同调查会产生不同的结果。与已知基准进行比较时, 各样本库之间的差异非常大; 即便针对同一样本库开展同一项调查, 其结果与基准数据的偏差范围也较大。

在学术研究领域, Yeager 等 (2011) 将通过随机拨号 (RDD) 进行的相同问卷调查结果与六个非概率样本库、一个概率在线样本库和一个基于河流抽样 (river sampling) 收集数据的样本 (见第 2.2.3 节) 的结果进行了比较, 总共涉及八个数据集。结果表明, 随机拨号和在线概率样本库数据比任何其他非概率样本库数据都更接近基准。后续采用类似方法的研究, 也得出相近结论, 即概率样本的数据质量更优 (Krosnick, MacInnis, Suh, & Yeager, 2013)。

当研究变量之间的关系时, 情况并不像上述关注点估计值的研究那样清晰。这方面的研究数量很少, 研究问题在于当客户或研究者基于不同样本库观察变量之间的关系时, 是否会做出不同的决策或得出不同的结论。Callegaro 等人 (2014c) 回顾了此类研究, 指出大多数研究集中在政治学领域。尽管人们普遍认为变量之间的关系和相关性受非概率抽样选择的影响不大 (即它们具有稳健性), 但相关系数的符号方向并不一致且相关强度因样本而异。换言之, 正如 Pasek 和 Krosnick (2010) 所强调的, 根据所使用的样本库类型, 可能会做出不同的决策。我们希望未来能有更多的研究关注这一主题。

选举民意调查的特殊案例 (The special case of election polls)。选举民意调查的优势在于, 所有样本库的核心统计指标 (即最终选举结果) 均可被直接验证。总体而言, 研究显示, 非概率在线样本库在准确性方面与概率在线样本库相当, 甚至在某些情况下预测得更准确。例如, 在美国, 类似现象可以追溯到 2000 年的选举 (Taylor et al., 2001), 在英国可以追溯到 2001 年 (You-Gov, 2011)。最近, 在 2012 年美国总统大选中, 非概率在线样本库的表现与传统概率民意调查一样好, 有时甚至更胜一筹 (Silver, 2012)。因此, 来自选举民意调查的结论与前文所述研究结论大相径庭。前文所述研究显示概率在线样本库的数据质量高于非概率在线样本库。对于这种差异, 有几种可能的解释。首先, 需要注意的是, 选举民意调查主要估计一个变量 (选举结果), 而这个变量大多数时候 (取决于国家) 是二元变量。其次,

选举研究通常是在特殊环境中进行,即在选举前几周许多其他研究(通常是电话民意调查)的结果已公开发布。事实上,与大多数调查不同,在选举民意调查中,网站会实时汇总调查结果,并在选举结束前持续提供指导,直至选举结束。这种指导使得抽样不断被完善、能够识别潜在的选民、优化问题措辞、处理犹豫不决的人和无应答者,以及完善加权机制。因此,准确性的差异更多地反映了这些调查环节的组合效果,而非非概率样本库整体的准确性。这些数据高度模型化,整合了过往选举的大量经验及选举民意调查的建模知识。正如最近美国民意研究协会(AAPOR)关于非概率样本的报告所述:“尽管非概率样本在选举民意调查中通常表现良好,但在其他领域及测量不同现象的更复杂调查中,其准确性的证据尚未明确”(Baker et al., 2013a, p. 108)。尽管如此,选举民意调查的案例仍极具参考价值,从中吸取的经验教训也可以推广到其他领域。

5.2.5.2 加权对在线样本库数据质量的影响

NOPVO、ARF FOQ1 和 ARF FOQ2 以及 MR1A 的研究中使用的加权方法,主要是基于人口统计学变量,差异较小。然而,这种加权技术(基于人口统计学变量的后分层加权)对使非概率样本库的估计值彼此接近,几乎没有效果。这表明在非概率在线样本库中存在其他影响数据质量的自我选择机制。在 FOQ2 研究中,事后分析显示,将诸如互联网使用频率之类的行为变量与人口统计学变量一起纳入考量,有助于使估计值更接近基准。Yeager 等人(2011)在其比较研究中使用了后分层加权,并指出它提高了两个概率样本估计值的准确性,但非概率样本库的估计值准确性有时增加、有时减少,且这种变化是非系统的。最后,Tourangeau、Conrad 和 Couper(2013)针对加权对非概率在线样本库的影响开展了元分析,以减少覆盖偏差和选择偏差。分析不同的研究后,得出结论:此类调整最多能消除五分之三的偏差,且不同变量之间仍存在较大差异,而不同加权方法之间的差异并不显著。

另一种在在线样本库中越来越受欢迎的加权方法是倾向性评分加权(*propensity score weighting*, PSW),也称为参考调查(*using a reference survey*)加权法。倾向性评分加权法基于的假设是,仅通过基于人口统计学的加权无法弥补非概率在线样本库和“真实”目标人群之间的差异。正如前文所述,仅通过人口统计特征加权并不能解决样本库成员与目标人群的本质差异。为了纠正这些偏差,需要在模型中加入额外的协变量,被称为网络特征变量(*webographics*)(Schonlau, et al., 2004)。应用倾向性评分加权法,需要一项参考调查为基准,通常来自高质量的概率调查,这类调查通常是通过电话(随机拨号抽样, RDD)或面对面的方式收集数据。参考调查

用于校准或平衡在线样本库。倾向性评分是受访者的条件概率,确切而言为在线样本库受访者而不是参考调查受访者的条件概率 (Schonlau et al., 2004)。倾向性评分加权已经在不同的领域得到应用,且效果值得关注,但对于在模型中应纳入哪些网络特征变量仍然不明确 (Steinmetz et al., 2014)。详见第 4.1.4 节的讨论。

5.2.5.3 在线样本库研究中的其他数据质量问题

过去 15 年间,研究者探讨了诸多在线样本库研究中的其他数据质量问题。其中部分是全新议题,另一些则是其他数据收集方法中已被熟知、如今在在线样本库研究文献中再度受到关注的问题。鉴于篇幅有限,本节仅聚焦若干核心概念。其中一类问题与在线样本库成员频繁填写大量调查问卷有关,这引发了关于数据质量、职业受访者 (*professional respondents*)、快速答题者 (*speeders*) 以及样本库条件效应 (*panel conditioning effects*) 影响的讨论。

鉴于样本库成员在其注册期内会参加许多调查,那么反复填写调查问卷究竟会增加还是减少问卷测量误差?从一个角度来看,填答问卷的行为可以通过练习变得更好。换句话说,正如 Chang 和 Krosnick (2009) 所表明的,与电话调查受访者相比,在线样本库成员的报告准确性更高。从另一个角度来看,在非概率在线样本库中,任何人都可以注册,这引发了职业受访者 (*professional respondents*) 及其对数据质量的威胁问题。职业受访者被定义为经验丰富的调查参与者,可能是为了获得相关的奖励,例如现金或积分,而填写大量问卷 (Baker et al., 2010, p. 756)。关于职业受访者与数据质量关系的证据较复杂。正如 Hillygus、Jackson 和 Young (2014) 在其对职业受访者的批判性回顾中所阐述的,有时职业受访者提供高质量的数据,有时则不然。

与职业受访者相关的是快速答题者 (*speeders*)。这类受访者完成问卷的时间远低于中位数,因此可能提供的答案质量较低。Greski、Meyer 和 Schoen (2014) 的研究发现,在非概率样本库中,快速答题者的比例高于基于概率的样本库。然而,他们对数据质量的影响并不大。

最后,与多次调查有关,样本库条件效应 (*panel conditioning*) 的老问题再次出现在在线样本库的文献中。样本库条件效应 (Warren & Halpern-Manners, 2012) 与纵向调查有关,即对同一受访者多次提出相同或相似的问题,这使得受访者进行调整,从而使应答质量下降。有证据表明,样本库条件效应在知识性问题中比其他类型的问题中更突出 (Kruse et al., 2010),但正如 Warren 和 Halpern-Manners (2012) 所提醒的,研究样本库条件作用是非常复杂的,它经常与流失效应相互作用。

研究者在覆盖误差、无应答偏差和流失等领域研究了其他数据质量问题。如果样本库的目标是得出代表整个人群的结果,那么覆盖误差(*coverage error*)与国家互联网普及率密切相关(Callegaro & Krosnick, 2014)。正如第5.2.5节所述,基于概率的在线样本库有不同的处理覆盖误差的方法(例如,给受访者提供设备来完成调查)。对于非概率在线样本库而言,要得出能代表全体人口的结果,唯一的选择是进行加权,因为根据定义,非互联网家庭或受访者不会参与非概率在线样本库。

“无应答”的含义因样本库类型而异。对概率样本库而言,从招募到单次调查的完成,无应答影响每个阶段(Lee, 2006)。对于非概率在线样本库而言,由于缺乏抽样框(第2.5.2节),讨论无应答的意义有限。此外,对于单次调查而言,讨论无应答的意义也不大,因此问卷的完成率更多地取决于样本库机构如何管理和邀请样本库成员,而不是其他因素(Callegaro & DiSogra, 2008)。受访者流失问题可以从不同角度来分析。从成本角度来看,流失会减少样本库的规模,迫使样本库机构不断招募成员,以保持其规模的稳定性。从无应答角度来看,受访者的流失几乎从来都不是随机的(即不同类型的受访者以不同的速率流失)(Lugtig, Das, & Scherpenzeel, 2014)。最后,对于纵向研究设计而言,受访者流失减少了有效样本量,并引入了无应答测量误差。

5.2.5.4 在线样本库数据质量的结论

现有证据已引发对在线样本库数据质量及非概率抽样方法的质疑。从大规模比较研究来看,使用不同类型的在线样本库可能导致截然不同的研究决策。与此同时,当前状况并未比互联网普及率较低的几年前有所改善。在线样本库的数据质量问题仍未得到有效解决。

Baker等人(2013)指出,现有证据表明“最具有前景的非概率抽样方法是那些基于模型的方法,这类模型试图处理在抽样和估计阶段(*in both the sampling and estimation stages*)应对推断的挑战”(p. 101)。这一观点颇具启发性。因此,我们认为,比较不同样本来源的研究议程,以及对抽样与加权策略的进一步探索,均应持续推进。两份AAPOR工作组报告(在线样本库和非概率样本)(Baker et al., 2010, 2013)、ARF第二季度发布的研究结果和WebDataNet倡议中比较欧洲概率在线样本库的研究(WebDataNet工作组04)¹¹均为这一研究方向提供了指引。

尽管调查文献已经重点关注了其他数据质量问题,如职业受访者、快速答题者、样本库条件效应以及在线样本库的覆盖误差、无应答偏差和受访者流失的影响等,但从量化这些影响到制定针对性应对措施,仍有大量工作亟待完成。

5.2.6 购买在线样本库研究

截至目前,回顾的各项研究引发了关于在线样本库数据质量的讨论。在购买在线样本库研究时,专业协会有时并不清楚应该给其成员什么样的建议。尽管如此,专业协会仍在这方面上采取了一些措施。例如,欧洲民意与市场研究协会提供了一套包含 28 个问题的清单(向供应商提出),旨在帮助研究者(ESOMAR, 2012)。美国调查研究机构协会(CASRO)要求其成员遵守《调查研究标准与伦理准则》(CASRO, 2011)。该准则包含了针对在线样本库研究各个方面的八项披露规则,例如样本库招募实践、样本库成员活动和样本库质量实践。其他地方协会(如德国市场和社会研究经济协会 Arbeitskreis Deutscher Markt, 2007)在其成员与在线样本库供应商打交道时,提供了指导方针。尽管这些指导方针无疑有助于让客户更好地了解他们购买的样本类型,但大多数在线样本库供应商会对他们的招募和样本库管理实践加以保护,不会透露相关信息。在实践中,供应商通常会以所谓的“样本库手册”(panel books)的形式考虑到上述建议的基本内容,这些手册描述、记录并推广其服务。

适用性原则在这一背景下最为适用(Baker et al., 2013)。在这方面,作为在线样本库的购买者,我们应该问自己以下问题:在同等价格下,除在线样本库外,还有哪些其他的数据收集方法可供选择?目标是估计全国性的还是区域性的统计数据?代表性真的有必要吗?在线样本库的实验是研究的主要目标吗?研究结果只适用于互联网人群还是适用于整个人群?关注点是否在相关性上,这可能比点估计更稳健?

如果代表性和可推广性是必要的,一种策略是在研究中添加一些问题,而这些问题要有与主要研究主题相关的已知基准。通过这种方式,可以了解到所获得的结果与基准数据的差距有多大。随着越来越多的公共数据集易于获取,对于大多数研究而言,这种策略似乎是可行的。

5.2.7 在线样本库的未来

正如第 1 章所述,超过一半的市场调研相关收入已经来自网络调查,而第三方(非概率)在线样本库已经占据了所有样本的近 35%(ESOMAR, 2013)。在线样本库已然存在,面对面调查和电话调查的减少进一步推动了这一转变。与此同时,在线调查数据质量问题真正成为需要关注的焦点。虽然非概率在线样本库正逐渐成为一种常见选择,但概率在线样本库的情况并非如此。因此,我们常常需要退一步,

思考需要哪种类型的在线样本库，或者是否有其他比网络调查更合适的方法，甚至是否需要原始数据收集。

移动在线调查样本库是一个最近迅速发展的趋势，随着移动应用程序的发展，其规模正迅速扩大。用户通过下载相应的智能手机或平板电脑应用程序成为样本库成员；这些应用程序随后会推送调查问卷（见第 5.1 节）。

我们已经提到了针对非概率库在线样本库的各种行业标准和准则（Arbeitskreis Deutscher market, 2007; CASRO, 2011; ESOMAR, 2012）。对于概率在线样本库，基于实证的实践方法也已经发布且易于获取（Scherpenzeel & Das, 2010）。然而，遵循这些实践方法会使建立在线样本库（前期初始成本）和维护成本高昂。可能正是因为如此，概率在线样本库在整个在线样本库业务中只占少数。另一方面，随着互联网普及率的提高以及用于浏览网页的设备价格降低，建立和维护概率在线样本库的总体成本应该会下降。例如在欧洲，越来越多的概率样本库正在建立（Das, 2012）、推广¹²或处于规划之中（Nicolaas, 2013）。国家统计局也在考虑建立概率在线样本库（Amarov & Rendtel, 2013; Cobben & Bethlehem, 2013），尽管质量和代表性是关键问题，引发了诸多争议（Svenssoon, 2013）。

如前所述，关于非概率样本库的抽样和加权，我们仍然有许多需要学习的地方。我们鼓励读者针对在线样本库的数据质量设计并开展更多研究。遗憾的是，迄今为止的证据表明，在线样本库数据收集在成本、速度以及对特定群体覆盖方面的优势，并不总是与数据质量相关。

最后，非常重要的一点是数据质量与成本密切相关。由于非概率在线样本库完成一次调查的成本可能低至 2 美元（Terhanian & Bremer, 2012）；而概率样本库或电话访谈的成本要比非概率在线样本库高出 10 倍，面对面访谈的成本甚至可能高 100 倍，这该如何衡量？如此大的价格差异是否真正反映了数据质量的差异？我们将在第 6.1.4 节中对上述问题进行简要思考，并期待未来对在线样本库数据质量进行更多的研究，同时希望本节内容能对这一讨论有所贡献。

5.3 网络调查软件

开展网络调查离不开软件支持。20 世纪 90 年代初，这类软件多为满足特定问卷需求而定制开发的计算机程序。至今，这种做法在某些情况下仍然存在。此外，许多内容管理系统、在线学习平台、社交网站和其他网络服务中，都有一些功能有

限的网络调查模块和插件，而我们关注的重点，是专门用于网络调查的软件。

学术文献很少涉及网络调查软件，少量研究有所关注，如 Baker (2013)、Crawford (2002)、Kaczmirek (2008)、Wright (2005) 和 Macer (2014)。一些会议报告、在线评估、杂志评论、博客和论坛也涉及这些问题（参见 WebSM 对博客的概述）¹³。来自调查软件行业的作者出版物和帖子也可以提供帮助，但往往是促销性的（Bhaskaran & LeClaire, 2010），而独立专家的分析则非常罕见¹⁴。

尽管网络调查软件的重要性不言而喻，但其复杂性及快速变化的动态特征，阻碍了学术界对其展开深入研究。目前市场上有数百种网络调查软件，涵盖数千项功能，并且可以根据各种评估标准进行分类。这些标准涵盖了多个维度，从可用性、交互性、统计和方法学到技术、法律、管理以及成本等方面。此外，网络调查软件的更新迭代速度非常快，许多供应商每周都会发布改进版本，新的市场参与者也不断涌现。更广泛的环境变化进一步加剧了不确定性，例如技术发展（如 SaaS、HTML5）、调查设计原则演变（如自适应设计）、用户体验理念更新（如 web 2.0）、新服务的出现（如社交网站）、新设备的推出（如智能手机和平板电脑），以及与方法论（如无应答编码）、法规（如网络 cookie/ 垃圾邮件法）或安全标准（如 ISO 标准）相关的专业标准的变化。

相应地，网络调查软件中“必备功能”“实用功能”与“锦上添花功能”的界定（Kaczmirek, 2008, p. 238）也在迅速变化。今天的最优选择可能转瞬即逝，而这种“过时”可能仅需一年，因为约半数的市场调研公司计划在未来两年内更换其用于收集和分析数据的平台（Macer & Wilson, 2014）。

总体而言，理想的软件能够节省资源、提高数据质量，并确保受访者、研究者以及委托方 / 发起方的满意度。选择合适的软件，其重要性与价值往往超过了众多方法论层面的决策，这些决策在相关文献及本书前几章中已详尽阐述。

当然，不同类型的用户对网络调查软件的需求存在差异。对于专业用户而言（*professional users*），好的软件尤其重要，比如专业的数据收集机构（统计局、市场研究机构）和其他调查密集型机构（研究 / 学术 / 非营利机构），以及开展高级或复杂网络调查项目的研究者，如在线样本库、非常复杂的问卷调查（如研究实验）和大型（如人口普查）或重复性（如客户满意度）的调查。专业用户能够理解高级功能的细微差异，并且通常高度重视长期合作关系。对他们来说，更换网络调查软件会引发严重的技术、方法论和组织层面的难题，还会造成低效性。

普通用户（*ordinary users*）的需求则相对简单。他们经常在特定项目中实施简单或中等复杂程度的网络调查。这类工作通常仅为辅助性或支持性活动，可能定期

开展。但这并非其业务或专业活动的核心，在方法论上也无需过于先进。在这个群体的低端，存在许多DIY用户，他们实际创建了当今多数网络调查，并且通常只需要一组有限的功能。普通用户通常可以相对容易地更换软件。

下文将首先梳理网络调查软件的宏观概念框架（第5.3.1节），然后回顾网络调查软件市场（第5.3.2节）、网络调查软件的特点（第5.3.3节），并讨论两个选定的易用性方面（第5.3.4节）。最后，我们分析软件选择过程（第5.3.5节）和未来趋势（第5.3.6节）。

尽管现实情况中，网络调查日益被视为更广泛的社会科学研究环境的一部分，我们仍然主要关注对核心网络调查过程的支持（第1.2节图1.1）。这一范畴有时被归入电子社会科学（e-social science）范畴（Vehovar, Petrovčič, & Slavec, 2014）。理想情况下，整个研究过程应得到完整的信息与通信技术（ICT）解决方案支持，类似于电子学习平台。另一方面，网络调查流程也正在成为各种应用领域的一个元素，如市场调研、人力资源管理、客户满意度、客户关系管理（CRM）、商业智能、直接营销、销售等。在复杂组织中，与客户（和其他利益相关者）调查相关的企业反馈管理（EFM）概念经常出现，其中存在多个层级的用户（例如问卷设计者、数据管理者、报告撰写者、翻译人员和仪表盘高管等）。

5.3.1 网络调查过程的集成信息与通信技术（ICT）支持

考虑到核心网络调查过程——预调查阶段（pro-fielding）、正式调查阶段（fielding）和调查后阶段（post-fielding）——我们通常期望ICT能够整合完整流程。然而，从表5.1的总结性概述中可以看出，情况并非如此。在表5.1中，当今网络调查软件中ICT支持尚未达到全流程整合水平，而是根据技术支持等级对应匹配到了网络调查流程的各个阶段。这种水平的划分主要基于实践经验，并根据对网络调查软件市场的全面分析（Vehovar, Čehovin et al., 2012）。

强支持（*Strong support*）指所有网络调查软件都支持和整合了网络调查过程的某个阶段或子阶段。其中问卷的计算机化（2.3b）是所有工具都支持的功能。高级软件可以为这一过程提供额外的用户友好性、指导和诊断功能，如在线帮助。ICT支持完全整合的另一个阶段是测量（3.2），其中完全支持填写问卷和保存数据。这也是网络调查软件最基本的功能，因此所有软件都完全支持。此处，一些高级工具也可能提供额外的功能，如实时质量控制（Peytchev & Crawford, 2005）、暂停和继续填写等。

中等支持（*Moderate support*）指大多数工具已经集成了这些功能，包括招募（3.1）、详细技术支持（2.4）和初步报告（4.2）。本书的相应章节详细讨论了需要 ICT 支持的各种功能。招募和初步报告尤为重要，它们通常与两个强支持阶段（2.3b 和 3.2）共同出现在网络调查软件的导航菜单中，直观呈现基本流程：

创建 → 招募 → 测量 → 分析

在一些网络调查软件中，“招募”和“测量”阶段合并为“收集”（Collect）阶段。当然，各种不同的标签可以用来代替“招募”，如“分发”（Distribute）、“发布”（Publish）或“邀请”（Invite）。

表 5.1 网络调查过程的各阶段及集成 ICT 支持水平 *

调查研究过程中的步骤	集成 ICT 支持
1 初步研究活动	无
2 预调查	
2.1 模式设计	弱
2.2 抽样	弱
2.3 问卷准备	
(a) 问卷编制	几乎没有
(b) 问卷计算机化	强
(c) 测试	弱
2.4 无应答策略	弱
2.5 技术准备	中等
2.6 管理	几乎没有
3 正式调查	
3.1 招募	中等
3.2 测量	强
3.3 处理和监控	弱
4 调查后	
4.1 数据准备	弱
4.2 初步结果	中等
4.3 数据导出与文档化	弱
5 高级分析、处理、价值转化	无

* 表中的前两个层次遵循核心章节第 2、3、4 章的结构。

弱支持 (*Weak support*) 指特定阶段通常依赖外部 ICT 解决方案, 仅有少数先进网络调查软件提供部分支持。这适用于对混合模式调查 (2.1) 的支持。混合模式调查很少与网络调查软件集成 (例如支持网络和电话调查的软件)。此外, 这也适用于样本选择 (2.2), 通常是通过外部数据管理工具完成的, 以及测试 (2.3c)。测试阶段通常需要各种外部工具进行高级问卷诊断 (例如调查质量预测器)¹⁵, 用于捕获和分析可用性文档 (屏幕截图、元数据) 和各种问卷评估方法 (Madans et al., 2011) 等。只有最先进的工具 (Stark & Gatward, 2010) 部分地集成了这些方面。对无应答策略 (2.5) 的支持通常也非常有限, 甚至应答率通常需要手动计算, 因为很难“即时”确定单位应答状态。

弱支持也涉及调查阶段工作处理与监控 (3.3)。网络调查软件通常仅能为网络调查过程提供有限的信息, 需要外部计算来支持某些决策。同样, 在调查阶段也不会提供关于数据质量的详细信息 (例如满足作答、不可用单位等)。

这种情况同样适用于数据准备 (4.1), 在这一阶段, ICT 支持的集成程度非常有限, 通常需要使用各种外部工具进行编辑、编码、重新编码、加权、插补缺失值以及数据管理 (例如文件排序、合并、更新) 等。这些活动都很复杂, 尤其在需要详细记录时, 如审计追踪 (元数据)、记录变更 (谁、何时、变更了什么) 或检查数据文件中更改的一致性。数据导出与文档化 (4.3) 也是如此, 除少数罕见的工具 (如 Blaise)¹⁶ 外, 复杂的元数据导出、文档化以及导出到标准图书馆格式等通常由外部工具完成或根据特殊要求额外完成。

几乎没有支持 (*Almost no support*) 意味着通常需要外部 ICT 服务来支持特定阶段的某些功能, 即使最先进的工具也只是提供极少的 ICT 集成。问卷编制 (2.3a) 是最典型的例子, 正如第 2.3.6 节所述, 它通常涉及文本处理器 (例如 MS Word、Google Docs) 或一些专门的问卷规范工具, 以及用于起草调查问卷的异步通信工具 (例如电子邮件, 网络数据交换平台)。然而, Vehovar、Čehovin 等人 (2014) 的研究表明, 将这个阶段的 ICT 与计算机化结合起来 (2.3b), 只需要相对较少的扩展功能 (例如在线编辑、基于评论系统进行问卷反馈等)。这种方法还能够实现问卷编制过程的全面记录 (即追踪变更), 并利用语言技术 (问题措辞的语言质量)、数据挖掘技术 (从题库中寻找潜在问题) 和人工智能技术 (从理论文献中形式化和实现问卷开发规则), 增强该过程中的交互式诊断功能。这无疑是未来网络调查软件发展的关键挑战之一。

几乎没有集成 ICT 支持也是综合管理 (2.6) 的特征, 这通常需要外部软件支持 (例如 BaseCamp、微软项目管理软件 MS Project)。团队内部和外部 (发起方/委托方) 的沟通也是如此。理想情况下, 这些沟通功能应该与管理功能以及网络调查软件本

身整合在一起。然而，现今的沟通通常通过外部渠道进行，例如电子邮件、论坛、聊天或某些专门的在线协作工具（如维基）。现在，也有少数的高级工具逐步整合了沟通和管理功能，特别是在跨用户层级、多调查项目、复杂调查或在线样本库的场景中。

没有支持（*no support*）意味着没有任何网络调查软件整合了相应的 ICT 支持。这可能是因为相应的 ICT 服务独立于网络调查软件之外，也可能是因为根本没有这样的支持。这与初步研究活动（1）有关（表 5.1）。在初步研究活动中使用了一套单独的 ICT 服务，从电子邮件和视频会议到专门用于头脑风暴、思维导图、结构化想法、计算机化定性研究的特殊软件和问题库系统（例如 Survey Question Bank）¹⁷。高级分析（5）的情况也与之相同，例如统计分析是通过电子表格或专业的统计工具（例如 IBM-SPSS, R, Stata, SAS）在外部进行，高级处理（例如集成到某些客户关系管理系统中，Customer Relationship Management, 简称 CRM）和价值转化呈现（例如为印刷做准备）亦是如此。

综上所述，当前信息与通信技术的支持呈现显著碎片化特征。部分原因在于缺乏综合概念框架，从而导致调查过程被分割成孤立的组成部分，每个部分都由单独的 ICT 解决方案支持。这种分离既有历史上各种技术限制（例如 web 2.0 仅在十年前才出现）的原因，目前这些限制正在逐渐消除，也因为软件开发人员长期以工程视角为主（Vehovar, Čehovin, et al., 2014），而方法学家并未参与核心架构设计。

理想的软件应该将信息与通信技术支持集成到核心网络调查过程的所有步骤中。实际上我们确实观察到了这一总体趋势。当然，预调查和调查后阶段的活动尤其需要正式化和整合，这在 Blaise 用户会议上展示的各种研究论文中体现的最为明显¹⁸。还应该补充说明的是，用户普遍倾向于使用单一工具完成调查设计、数据收集、分析与呈现，这种整合需求已得到验证（Macer & Wilson, 2014）。

当然，对于专门的网络调查软件来说，在 ICT 支持的整合方面（而不是依赖于专门的外部 ICT 解决方案）存在一定的限制。这种整合应该考虑到各个阶段之间的差异，每个阶段都有特定的限制、动态和挑战。例如，开发高度专业化的缺失值插补功能作为网络调查软件的集成组件，既复杂又昂贵，因此合理的边界是必要的。

5.3.2 网络调查软件市场

自 20 世纪 70 年代起，随着计算设备从大型机转向台式机、操作系统从微软 DOS 迭代至 Windows，再到后来向互联网的转型，计算机辅助调查信息收集（CASIC）市场始终保持活跃。尽管如此，到 20 世纪 90 年代末，网络调查软件

行业已经很成熟,包括当今多家领先供应商,既有复杂的商业工具(例如 Blaise、GlobalPark/Questback 和 Confirm)或开源工具(例如 PhPSurveyor/ limesurvey),也包括具备网络图形用户界面的低成本托管平台。所有这些工具都能使用户无需任何编程脚本就可以轻松创建网络调查(例如 SurveyMonkey)。

如今,市场动态主要体现在初创企业涌现、风险投资活跃以及频繁的并购活动上¹⁹。对利润的追求进一步推动网络调查软件供应商建立自己的在线样本库,除软件之外,也为研究者提供受访者样本。

随着越来越多的软件功能逐渐成为“大众化商品”,消费者也因此从中受益。与此同时,与其他全球 ICT 行业一样,网络调查软件市场也呈现集中化趋势。大型企业在行业中逐渐挤压低端市场,许多低端供应商只能满足普通用户需求。市场集中化还体现在高端市场,即面向专业用户的高级工具,因为高昂的开发和营销成本需要足够庞大的客户群体分担。

我们观察到,那些在市场上存在了十年或更久的供应商,由于其初始软件架构的限制,灵活性日益不足。转向 SaaS(软件即服务)、实施 web 2.0(例如拖放功能)和支持移动设备是许多供应商失去竞争力的典型案例。与此同时,进入网络调查软件市场的壁垒正在增加。2013 年和 2014 年,全球市场上仅新增少数软件包(WebSM, 2014)。专门为某些组织创建的内部工具也面临类似困境,由于难以应对持续的修改与集成需求,这类工具在市场调研公司中的占比已降至 10% 以下(Macer & Wilson, 2014)。

目前市场上有数百种专门的网络调查软件工具。由于行业动态性强,全面的市场概述愈发难以维持。在此,我们参考 WebSM 的网络调查软件列表(WebSM, 2014),该列表虽以英语工具为主,但也涵盖了西班牙语、德语和俄语的核心工具。除亚洲外,我们相信全球性的工具都在这个列表内。2013 年,WebSM 名单上有 350 种活跃工具,其中大多数供应商位于美国(42%)和欧洲(35%)。这一分布格局在 2014 年保持稳定,活跃工具数量基本持平(352 种工具)。

根据 2012 年 WebSM 的研究(Vehovar, Čehovin et al., 2012),大多数网络调查软件(51%)是面向普通用户的商业工具,通常提供固定价格套餐。此外,普通用户还会获得各种免费(3%)或开源工具(7%)。四分之一(25%)的工具面向专业用户,而其余部分(15%)的工具仅部分支持网络调查(例如赛事支持、测验、表单、投票、人力资源管理),且常作为副产品提供,其功能往往有限。

开源工具的短板及其功能局限性,部分源于需要用户自行安装服务器。这对普通用户而言往往难以实现,对于缺乏处理开源解决方案的技术专业人员来说也是挑

战。此外,对于这种特定的网络服务来说,及时和专业的用户支持极其重要,这正是商业解决方案相对于开源工具的另一优势。近期对开源网络调查软件的研究证实了这一点,同时发现了一种趋势,即开源工具正开始针对部分服务(例如安装、托管)收费(参见 WebSM, 2014 年的开源研究)。

5.3.3 网络调查软件的特征

网络调查软件之间最明显的区别就是价格(*price*)。低价软件通常面向普通用户,包括以下软件包(Vehovar, Čehovin et al., 2012):

- 免费版本(有各种限制,通常每份问卷最多有 100 份应答或 10 个问题,功能有限)。
- 基本版本,每月费用几十欧元/美元(但也有低至每月 3 欧元的情况),功能有限(通常不支持条件逻辑和高级问题类型)。
- 一级、二级甚至三级等分级版(例如高级版、专业版和企业版),价格和功能逐渐增加,但通常每月的费用保持在几百欧元/美元以内。

当前趋势是供应商提供的套餐数量逐渐减少,有时只提供一种固定费率(按月或按年计费)。同样,接受访者数量收费模式正逐渐消失,2012 年只有 12% 的工具采用这种模式。相应的价格取决于每个月的受访者总数,范围从每位受访者 0.002 欧元到 0.938 欧元不等(平均值为 0.194 欧元,中位数为 0.145 欧元)。2014 年的更新数据显示了非常类似的情况(参见 2014 年 WebSM 的定价研究)。

低价工具在 Vehovar、Čehovin 等人(2012)的研究中得到了详细阐述,而按不同功能包层级划分的相应功能概述可以在 Čehovin 和 Vehovar (2012a)的研究中找到。通过与过去研究(Vehovar, et al., 2005)的比较,我们可以清楚看到诸如高级问题类型、条件支持、实时验证、随机化、数据管道等曾经的区分性特征,如今正日益成为行业标准。

另一方面,高价的网络调查软件面向专业用户,且没有按照套餐进行结构化。相反,它通常需要用户提供报价,随后供应商提供定制化的规格,继而双方进行谈判。高端工具在高级方法论功能(*methodological features*)(Vehovar, Čehovin, et al., 2012)方面存在巨大差异,而面向普通用户的低端工具通常不支持这些功能。一般而言,这些功能涉及上述与表 2.1 相关的所有高级方面。我们在下面提供了一个更具体的功能概述,并首先揭示了六个基本方法特征。这些特征最清楚地定义了高级网络调查软件:

1. 集成的电子邮件系统 (*email system*), 具备数据库管理 (例如合成电子邮件列表)、邀请设计 (例如精心设计的电子邮件邀请格式)、电子邮件诊断 (例如: 查看电子邮件是否被发送 / 已打开)、监控功能 (例如功能强大的仪表盘, 用于详细追踪每个单位的应答历史)、设置发件人的电子邮件 (例如 `anyname@anyorganisation.com`) 以及非常强大的垃圾邮件预防 (见第 2.4.1 节) 等方面的高级功能。
2. 混合模式调查 (*mixed-mode surveys*) 支持网络、电话、面对面、邮寄等多种调查模式的招募、个案管理和数据库同步。
3. 对多种设备 (*multiple devices*) (台式机 / 笔记本、智能手机、平板电脑、智能电视) 的集成支持, 包括对基于浏览器的移动调查的全面优化, 以及创建移动调查应用程序和基于浏览器的网络调查编写应用程序的集成支持 (第 5.1.6 节)。
4. 语言和翻译 (*Language and translations*): 支持多语言回答、在单一数据文件中存储多语言回答、允许受访者选择语言、为不同语言提供专属 URL、功能强大的翻译界面 (与问卷更改严格分离)、灵活地更改验证型消息和导航按钮的默认文本、多语言问题库和多语言管理用户界面。
5. 第 2.3.2 节中概述的高级问卷功能 (*Advanced questionnaire features*), 具有高度灵活的问题修改功能、高级问题类型 (例如 360° 评估、社交网络分析、联合分析、生物标志物、多媒体等)、对图形和多媒体的强大支持、与基于位置的服务 (地图、地理信息系统、GPS) 的集成, 支持综合问卷的编制和测试 (如详细的评论系统)。
6. 支持纵向调查 (*Support for longitudinal surveys*), 包括支持纵向调查样本和案例管理的一般功能, 纵向调查中样本单位被调查多次 (从样本选择到单位追踪), 以及对在线样本库的具体支持: 从招募到小组成员画像、各种流程优化、激励管理、欺诈检测和调查后调整等。详细内容请参考 Macer (2014)。

除上述方法论问题外, 高端和低端网络调查软件也在技术、管理和行政 (*technical, managerial and administrative*) 等相关功能方面同样存在显著差异, 这些功能在本书的其他章节中从实质性的角度进行了讨论。

7. 数据收集可以在客户端 (*client*) 进行, 也可以在供应商服务器 (*supplier server*) (SaaS) 上完成, 每种选择都有与安全要求相关的优点和缺点 (见第 2.4.3 节)。
8. 在台式机 / 笔记本安装 (*installation*) 中, 可以提供启用离线问卷创建的选项。

同样，也可以在移动设备端用户友好地创建网络调查（见第 5.1.5 节）。

9. 一个重要的选项是将编程脚本的自定义部分添加到现有代码中，以及基于独立编程语言（*programming language*）开发网络问卷。这可以提升长 / 复杂问卷（例如数千个问题、具有复杂实时验证）的编程速度和灵活性。
10. 高级技术功能涉及容量（*capacity*）（例如对用户数量、调查项目数量、应答数量都没有限制）、隐私（例如高级 IP 和 Cookie 设置）、安全（*security*）（例如镜像、复制、备份、追踪、云）、技术标准（*technical standards*）（如 W3C）以及为残障人士提供无障碍网络标准（如第 508²⁰ 节）。
11. 支持多用户层级（*multiuser hierarchy*）并分配给不同层级用户（例如企业反馈管理，EFM）相关的详细选择功能和权限。
12. 可以提供对执行、管理、监控和报告（*administration, management, monitoring and reporting*）的高级支持，从计算自动应答率（例如 AAPOR 比率）和详细的缺失数据文档到各种数据处理和项目管理报告、见解和干预机制等。
13. 与业务流程的集成（*Integration with business processes*）极其重要，从直接插件（例如用于销售的插件）到与其他软件链接的一般灵活性（例如用于营销、分析、商业智能等的软件），尤其强调与组织内部数据的集成及互通的便捷性和能力。
14. 支持工具的更换 / 切换（*changing/switching*），尤其是导入和导出功能，避免用户受限于单一工具（Kinesis, 2013）。
15. 用户支持（*user support*）质量，涵盖手册、文档、教程及互动（例如聊天、电话、操作时间、响应时间），与本地化水平（*level of localization*）密切相关：（a）以研究者的母语提供高质量的用户界面（不仅仅是英语界面）；（b）以本地语言提供问题和问卷库以及（c）有效的本地支持（电话和实时帮助、培训、干预和额外的软件开发）。
16. 清晰且详尽阐述法律问题（*legal issues*）和财务条款（*financial terms*）非常重要，并且可能有很大差异。

高端解决方案通常以数千欧元 / 美元的年费起步，对于专业用户在专门的数据收集机构中，费用很容易达到六位数。因此，高端工具在价格水平上已经非常明确地区分开来，并且没有套餐选项。

高端工具和低端工具之间的一个简单但重要且非常具体的区分特征——也适用于高端工具自身内部的结构化——是是否存在自己的编程语言（见上述第 9 点）。同样具有区分性的是符合正式无障碍标准（见上述第 10 点）。

尽管存在这些差异，但低端和高端工具之间的界限正变得越来越模糊：前者越来越多地支持专业用户所需的功能，而后者的一些工具则扩大其供应和调整价格以接近普通用户。在一定程度上，这主要是由于网络调查软件的特性所导致。这些软件解决了所有工具的问题，并与用户友好性、适应新趋势和可用性有关：

17. 通用可用性 (*General usability*) 包含许多方面 (Nielsen, 2012)，从设计、一般用户友好性、执行操作所需的点击次数、速度和整体复杂性（例如该工具可能过于技术化）到架构和导航。
18. 功能强大的图形用户界面 (*GUI*)，为处理复杂操作提供友好的支持（例如条件逻辑），从而在很大程度上取代编写脚本（即编写程序代码，见第9点）。
19. 设计灵活性 (*Design flexibility*) 与默认调查模板的质量、友好的设计构建器界面（例如更改标志、页眉 / 页脚、字体、颜色）、集成的 CSS 编辑器，强大的外部导入功能以及自动适应各种设备布局的响应式网页设计功能有关。

这里可以观察到计算机程序员在网络调查过程中的参与程度 (*involvement of computer programmers*) 呈现有趣的演变趋势。具体来说，计算机程序员最初是作为 CASIC（计算机辅助调查信息收集）革命的一个无法避免的附属组成部分被纳入调查流程。为了创建一份计算机化的问卷，需要专业的编程技能来编写描述和运行调查问卷的程序脚本。当然这使得该过程变得复杂化，因为程序员需要详细规范、反馈、沟通和各种程度的控制。所有这些都给流程增加了一个额外的阶段，并且也带来成本、复杂性、冲突和延期等问题。直到 20 世纪 90 年代末，网络图形用户界面 (*GUI*) 才使得非程序员也能创建（简单的）网络调查。在 21 世纪的 web 2.0 革命和其他技术进步之后，图形用户界面的发展进一步优化了这一流程。然而，对于非常复杂的问卷，离线编程脚本——当然需要一定的计算机编程技能——有时似乎仍然更为高效。但新型工具正不断突破边界，逐步回归 CASIC 时代之前的状态——调查研究者可独立创建、把控问卷，无需计算机程序员参与。

5.3.4 关于两个选定特征的讨论

我们将进一步阐述界面速度和图形用户界面这两个关键特征——二者已包含在上述 19 个方面中（第 17 和 18 点）。之所以专门讨论，既是为了揭示这些特征的复杂性，也因为它们在软件选择中常被忽视。

5.3.4.1 界面速度

在网络调查软件的通用可用性方面,界面速度(*speed*)是创建和编辑网络问卷时的页面加载时间,在创建问卷的过程中会给研究者带来严重困扰。如果问卷在带有脚本的台式机/笔记本电脑上离线运行,那么相应的响应时间(几乎)为零,这是一个重要优势。离线编辑的速度优势——尤其是当编辑有数百或数千个问题的问卷时——是一些高级网络调查软件甚至完全舍弃网络图形用户界面(GUI),转而依赖编程脚本的关键原因之一。另一方面,虽然图形用户界面是当今的主流解决方案,但问卷在网页上运行的速度可能会成为瓶颈。多数网络延迟并非源于互联网连接,即便是跨区域距离(如欧洲与美国),简单操作(如添加一个问题)的延迟通常也仅0.5秒左右,而是由服务器端运行的网络调查软件本身性能所致。

一项比较了16种SaaS工具的研究发现,速度差异相当大(Čehovin & Vehovar, 2012b)。具体而言,这些工具的速度差异体现在关键问卷编制命令方面:添加、编辑和删除问题;创建、预览、保存和删除问卷;点击内部和外部导航按钮。问卷保存命令的平均延迟时间最长(2.8秒),最长延迟甚至可达8.7秒。另一个平均延迟较大的命令是点击应用程序的外部功能(例如帮助、关于等),平均延迟时间为3.0秒,最慢的工具则需要8.1秒。其他命令的平均延迟时间大多在1秒到2秒之间,但存在相当大的差异。

此外,研究涉及的16个SaaS托管调查软件,其网站的谷歌网页速度观察平均得分(Google PageSpeed Insights score,分数越高越好,满分100)为77分(最低57分,最高99分)。这些网站的“主页”或“第一页”的平均响应时间为2.3秒(最低0.6秒,最高6.6秒)。

将这些数据与标准基准对比可见,网络调查软件实际上属于响应较慢的网络服务。一般来说,我们应该认识到人类对0.1秒以内的延迟无明显感知。超过这个时间后,延迟的影响会增加,直到1.0秒左右,延迟感变得明显可以觉察到——但此时通常不成问题,仍然可以忍受(Nielsen, 1993)。超过这个延迟时间点后,用户的厌烦程度会增加,并可能在10秒后达到临界水平,用户变得越来越不耐烦、紧张甚至考虑放弃操作。在使用图形用户界面的网络问卷开发时,我们很容易在处理大型问卷(例如数百或数千个问题)中触及这些延迟阈值。但是不断发展的技术正在持续提升图形用户界面的性能表现。

所有这些基准都是基于前互联网时代,由30~40年前关于人机交互的研究建立的。然而,这些基准至今仍然有效,同样适用于解释不同任务和软件类型的容忍度差异,以及延迟对效率的影响。在我们的案例中,这与编制问卷调查有关

(Shneiderman, 1984)。

无论如何,速度是易于测量的指标,且可与不同容忍度水平对比。因为使用图形用户界面创建问卷的 SaaS 类工具的速度之间,确实存在相当大的差异,所以在选择网络调查软件时,研究人员可以检查这一特性。当使用编程脚本来描述问卷时,延迟问题可完全消除(延迟为零)。

5.3.4.2 图形用户界面的关键特征

尽管存在上述速度问题,图形用户界面正变得越来越流行,特别是在“所见即所得”(WYSIWYG)的环境中。以文字处理软件(例如 MS Word、Google Docs、Open/ Libre Office)数十年的发展为例,其图形用户界面已实现高度趋同与标准化(例如所有工具都使用具有标准化位置和功能的按钮)。网络调查软件实现了一定程度的标准化,因此竞争工具的登录页面和内部界面都非常相似。然而,有若干问题仍悬而未决,有待进一步完善。

目前来看,研究者在选择时往往根据主观的个人偏好。但我们认为,某些方案基本上已经过时,效率低下。其中大多数涉及以下方面:(a) 问卷的布局,(b) 编辑问题的基本流程和步骤,(c) 条件逻辑(分支)的创建界面,(d) 信息架构和导航。关于这些问题的更多见解,可参见第5章补充材料“网络调查软件的特征”(*Characteristics of web survey software*) (<http://websm.org/ch5>)。

检验图形用户界面效率的一个客观指标是完成某个任务所需的点击次数(*number of clicks needed to accomplish a certain task*)。理论上,功能更丰富的高级工具因导航更复杂,所需点击次数更多;但即便功能相似的工具,差异依然显著。一项对27种同质工具的研究表明,创建标准化问卷所需的点击次数范围很广(从149到524次点击)(Berzelak & Lozar Manfreda, 2009)。近期对主流网络调查软件的评估也表明,创建标准化问卷时,单项任务的平均点击次数在3.1到6.3次之间(Čehovin & Vehovar, 2012b)。任何用户都可以重复这个任务,并检查相应软件的点击次数。

综上,研究者可通过标准化方式测量界面速度和点击次数,并与基准对比。此外,建议在真实场景中由经验丰富的用户或可用性专家进行定性评估,因为这些问题在实际使用体验发生之前,难以充分预判。

5.3.5 寻找合适的工具

对于任何服务或产品,用户都面临着将营销噪音与现实区分开来的问题;但另

一方面,如果没有营销推广,潜在的好产品很难被发现。我们在选择网络调查软件时也面临同样的问题。因此,本节将寻找网络调查软件的过程分为四个阶段,这些阶段对于专业用户和普通用户来说均可适用。

初步规格清单 (Preliminary list of specifications)。首先,必须明确研究者的真实需求和现实条件。对于普通用户来说,第一步是查看标准套餐功能的初步清单 (Čehovin & Vehovar, 2012a)。专业用户则应该进一步细化上面提到的高端工具的 19 个功能,以明确自身需求和立场。实验研究者的需求最为独特、非标准化且非常具体 (如需支持鼠标移动追踪等特定元数据导出功能)。

本书系统梳理了网络调查全流程及相关方法学特征,所有这些都需要相应的信息通信技术支持。我们也明确指出了网络调查软件所需的功能。因此,通过梳理前文对调查各阶段的讨论,可形成最贴合需求的初步功能规范。

此处,需注意的是,不同网络调查软件对同一功能的描述术语可能存在差异。因此有时很难理解是否支持某种特定功能,这就需要额外的努力。

筛选规格和标准 (Shortlisting the specifications and criteria)。我们需要在初步清单中,确定必不可少的功能。这与在购买其他产品或服务时的情况相同。例如,如果我们购买汽车,会首先使用一些简单且易于识别的基本标准。这些标准可能与预算和其他一些关键特征 (尺寸、每年行驶的公里数、维护成本、环保问题) 有关。这就决定了我们的核心需求层 (例如家庭轿车),而从功能的实质性角度来看,其他细节则相对不那么重要 (例如颜色、品牌、附加功能)。对于网络调查软件而言,基本标准也与预算范围有关,同时也与规模有关。这里的规模涉及调查的数量、每次调查的问题数量、受访者数量、网络调查软件用户数量和层级数量。其他标准更具体地满足我们的需求,应该从标准功能 (Čehovin & Vehovar, 2012a) 或上述 19 个高级功能中仔细识别。对于普通用户来说,可用性或易用性方面具有一定重要性,而专业用户则更关注软件即服务 (SaaS) 的特定功能和编程脚本。

用户还会有各种额外的偏好。例如,他们可能需要在正式调查过程 (即实时) 轻松且用户友好地进行问卷干预,或者需要详细地在线查看结果 / 统计数据。因此所选工具需要支持这些功能。同样,如果需要某些高级问题类型,如非常具体的组合表 (第 2.3.2),就得明确地检查该功能是否得到支持。

工具的候选列表 (Shortlisting of the tools)。明确需求后,就可以检查相应的软件供应情况。应该与同行、同事、合作伙伴等进行广泛的,最好是系统、正式的咨询。这是最强大的信息来源,但只有在得当且充分的情况下才是如此;否则我们可能会得到误导性的结果 (例如信息来自背景不同、不可比较或有问题的用户)。

我们也可在互联网上系统开展调查。遗憾的是,没有任何一个列表(包括 WebSM 和 Quirk 的列表)²¹是完全详尽的、结构友好的并且频繁得以维护。对于 SaaS 解决方案,根据大致的页面访问量对工具进行排名(Vehovar, Čehovin et al., 2014)可能会提供一些见解。然而,排名也存在许多不足之处。

选择过程 (Selection process)。最后一个阶段是多标准的练习,其中在前几个阶段建立的一组基本标准由关键用户(或评估者或专家)进行评估,他们对每个维度进行评分(通常在 1 到 5 的量表上),而平均值的总和决定“胜出者”。对于更复杂的案例,可以使用群体决策领域中与软件选择过程相关的某些专业策略(Lai, Wong, & Cheung, 2002)。首先衡量某些标准的重要性以及根据这些标准进行的评估,然后将其结合。

当决策涉及每年数万或数十万欧元/美元的投入,并且对组织流程和相关员工的满意度有重要影响时,选择过程应该更详尽复杂。聘请专家或专家小组可能非常明智。对任何机构来说,更换一些复杂的软件都是潜在的噩梦。没有任何简单的方法可以将调查从一个软件包转换并整合到另一个软件包。

5.3.6 网络调查软件的趋势

我们预计,在预调查阶段(pre-fielding)、正式调查阶段(fielding)和调查后阶段(post-fielding),ICT(信息与通信技术)的整合将进一步得到加强。因此网络调查过程中更多阶段将得到更高级别的支持(见表 5.1)。用户将以更低的成本获得更好的服务。特别是,数据挖掘、语言技术和人工智能在各个阶段的参与度将增加(Vehovar, Petrovčič, & Slavec, 2014)。

软件向托管解决方案(SaaS)转型的趋势将持续推进,这一模式已受到多数普通用户青睐。然而,出于安全性与隐私保护需求,大型机构是否会完全放弃本地部署、全面采用托管模式,仍有待观察。

网络图形用户界面(GUI)取代编程脚本的趋势将持续强化。然而,正如技术专家、出版商和科技学科研究者未因文本处理软件的图形用户界面而放弃 LaTeX 编程工具(他们更倾向于使用 LaTeX 编程工具来描述文本),专业用户的细分市场可能会继续使用编程脚本来处理复杂问卷。

用于创建问卷的网络图形用户界面,可能会出现类似文本处理器类软件的趋同现象,朝一个标准化的界面解决方案发展;还可能会遵循在线文本处理软件界面,同时集成评论工具,并提供各种视图/模式选项。但需注意,与相应的离线替代方

案相比,在线文本处理软件目前在速度和功能方面仍然存在一些不足,因此这个过程可能会比较缓慢。

市场的集中化与全球化趋势将进一步凸显。在许多方面,只有大型供应商才能在整个设备和操作系统范围内提供有效的支持,以及在网络调查流程的各个阶段实现 ICT 集成。同样具有挑战的是,将高级算法(数据挖掘、语言资源和技术、人工智能)持续集成到问卷开发过程中。另一方面,技术进步确实促进并简化了跨平台、跨设备实施新解决方案(如 HTML5)。然而,许多功能确实需要大量资源,如涉及本地化各个方面的全面多语言支持,依然只有大型企业才能负担得起。尽管如此,新技术仍然为新手和细分市场参与者提供了机会,如近年来许多新工具主要专注于移动网络调查(Vehovar, Čehovin & Močnik, 2014)。

设备适配仍是核心挑战。用户不仅通过台式机和笔记本电脑进行网络互动,还日益借助一系列其他具有特定界面和屏幕尺寸的设备进行互动。现在网络调查也用到这些设备。正如第 5.1.2 节所讨论的,这造成了许多复杂性,因为这些设备在许多方面远不如台式机和笔记本电脑标准化。显然,网络调查软件应该识别设备并相应地优化问卷(例如省略拖放问题和鼠标悬停文本)。这意味着,软件需为小屏幕自动优化呈现方式(例如,将表格拆分成单独的条目,将长水平选项垂直排列等)。然而,正如 5.1.6 节所讨论的,适应这些特定需求是一个相对缓慢的过程。2014 年大多数网络调查软件,即使是在基础功能上(例如消除缩放或水平滚动的需求),仍未能实现基于浏览器的移动调查自动适配(参见移动功能研究,WebSM, 2014)。

关于第 5.1.5 节讨论的移动调查应用程序,我们预计,随着 HTML5 标准的普及与浏览器高级脚本功能的完善,其发展在趋于稳定前仍将经历数年波动。20 世纪 90 年代末,面向台式机的客户端问卷应用程序开始兴起(见第 1.3.2 节)。由于在设计灵活性和速度上优于基于浏览器的网络调查,客户端网络调查应用程度获得了发展动力。然而,随后互联网速度提升,浏览器得以改进和客户端脚本得以开发,基于客户端的台式机/笔记本电脑应用程序便失去了优势,最终几乎消失。我们相信,当代的移动调查应用程序本质上也是一种基于客户端的应用程序,也将遵循类似的趋势。不过,在需要基于位置服务的场景中,移动应用程序仍可能保持独特价值。

注 释

1 <http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/>

2 Uporaba informacijsko-komunikacijske tehnologije. Statistični urad Republike

- Slovenije. Retrieved September 1, 2014, from http://www.stat.si/tema_ekonomsko_infidruzba_informacijsko.asp
- 3 The ultimate mobile email statistics overview. EmailMonday. Retrieved September 1, 2014, from <http://www.emailmonday.com/mobile-email-usage-statistics>
 - 4 Apple cedes market share in smartphone operating system market as Android surges and Windows phone gains, according to IDC. IDC Analyze the Future. Retrieved September 1, 2014, from <http://www.idc.com>
 - 5 IDC Analyze the Future. Retrieved September 1, 2014, from <http://www.idc.com>
 - 6 *ELIPSS*. Retrieved September 1, 2014, from <http://www.elipss.fr>
 - 7 Seven ways mobile phones have changed lives in Africa. Retrieved September 1, 2014, from <http://www.cnn.com>
 - 8 This section draws on some material from Callegaro et al. (2014a).
 - 9 Note that for the ISO, access panels are ‘a sample database of potential respondents who declare that they will cooperate for future data collection if selected’ (p. 1).
 - 10 For space reasons we omit here a third quality metric: test–retest reliability (running the same study at two close points in time on the same panel but with a different set of respondents)
 - 11 <http://webdatanet.cbs.dk/>
 - 12 GESIS Panel Roadshow 2013/2014. Retrieved September 1, 2014, from <http://www.gesis.org>
 - 13 <http://www.websm.org/blogs/>
 - 14 For example <http://www.meaning.uk.com/blogs/tim-macer/>
 - 15 <http://sqp.upf.edu/>
 - 16 <http://www.blaise.com/>
 - 17 <http://survey.net.ac.uk/sqb/>
 - 18 <http://www.websm.org/blaise>
 - 19 Mergers in web survey software & service industry. Retrieved September 1, 2014, from <http://www.websm.org/merges>
 - 20 <http://www.websm.org/508>
 - 21 <http://websm.org/sw> and <http://www.quirks.com/directory/software/>

网络调查的广义方法论

在前述章节中，我们已聚焦于核心网络调查过程的基础方法学方面（第 2 章、第 3 章、第 4 章）以及相关的实施情况（第 5 章）。在本章中，我们将讨论网络调查过程中更为广泛的方法学内容（第 6.1 节）、管理内容（第 6.2 节）以及专业内容（第 6.3 节）。这些对于理解和管理网络调查过程都非常重要。我们还将对网络调查的相关文献进行概述（第 6.4 节）。

6.1 更广泛的方法学背景

本节将首先探讨与网络调查方法学间接关联的维度，深入剖析网络调查模式的定义（第 6.1.1 节）、数据质量（第 6.1.2 节）、网络调查模式效应（第 6.1.3 节）和成本（第 6.1.4 节）。

此外，我们将回顾超出核心网络调查过程之外的相关活动（第 6.1.5 节）：初步研究活动的步骤、高级分析、数据处理环节与价值转化，以及混合模式和混合方法。最后，我们讨论了自适应、响应式和交互式设计（第 6.1.6 节）。

6.1.1 网络调查模式的修订

我们在第 1.1.1 节对基本网络调查模式的初步定义只是概述性介绍，而现在将

对其进行更精确的界定。首先,需要再次强调,在概念层面,我们将调查模式与抽样和招募过程完全剥离开。在对其定义时,我们仅考虑了对测量阶段有直接影响的因素,即受访者填写问卷的阶段。

研究文献通常采用某些属性(维度、因素、特征)来定义调查模式,这些属性与测量阶段的特点及流程有关。它们共同构成了受访者与问卷间信息传递的基础机制。尽管不同学者(Bierner & Lyberg, 2003; Couper, 2011; Groves et al., 2009; Tourangeau & Bradburn, 2010)对调查模式的维度划分存在差异,但近年 Berzelak (2014)提出的“六大固有模式特征”具有较强代表性。这些特征完全在研究者的控制之下,在实际调查过程中保持稳定,并且会对应答过程产生影响。基于在第1章中的讨论,现将这些特征及其选项概述如下,其中重点标注与基础网络调查模式相关的内容:

- a. 访员参与程度 (*Interviewer involvement*): 由访员进行的面访、无访员在场的自填式、访员被动在场的自填式 (CASI, 见第 1.3.4 节)。
- b. 计算机技术的使用 (*Usage of computer technology*): 交互式计算机化问卷、非交互式计算机化调查问卷、纸笔记录 (见第 1.1 节和第 1.3.2 节)。
- c. 信息传输媒介 (*Information transmission medium*, 即受访者将填写好的问卷提交给研究者的渠道): 面对面、电话、邮寄、电子邮件附件、电子邮件嵌入、自动在线传输 (例如通过互联网或其他电子网络, 见第 1.3.1 节)。
- d. 问题呈现输入通道 (*Question presentation input channel*): 听觉、视觉、双通道 (见第 1.3.4 节和第 1.1 节)。
- e. 应答输出渠道 (*Response output channel*): 口头报告、手写、手动电子化 (鼠标、键盘、指针、触控笔、触摸屏) (见第 1.3.4 节和第 1.1 节)。
- f. 访员互动的接近性 (*Closeness of the interviewer interaction*): 面对面、远程音频或视频、预先录制的音频或视频、虚拟音频或视频 (如虚拟的访员)、无互动。

此外,还有其他与调查方式相关的特征,这些特征与各种实施和情境的具体细节 (*implementation and contextual specifics*) 有关,通常超出研究者的控制范围,并且因实施方式、访员和受访者而异。因此,这些特征并不构成调查模式的定义性要素,仅用于描述特定调查的具体实施形式。例如,访谈的节奏、认知要求、技术 (如设备类型、屏幕分辨率、音质等)、环境干扰、受访者的多任务处理、调查的非个性化特征,以及访员和受访者的特征等。实施还可能在其他技术方面存在差异,

但通常对调查过程影响甚微,例如服务器的位置(例如 SaaS)、网络调查软件,以及客户端(例如作为移动应用程序运行的调查)与基于服务器的调查之间的区别。相应地,使用智能手机和平板电脑来回答网络问卷并不会引入一种新的调查方式。当受访者使用台式机/笔记本电脑、平板电脑或智能手机来回答网络问卷时,我们仍然将其称之为网络调查模式。这并非否定移动设备可能产生的特定设备效应(参见第 5.1.3 节),或者意味着也不需要考虑有意针对特定设备的网络调查实施情况(例如,使用移动调查应用程序的在线样本库)。这六个维度定义了数百种不同的调查模式。上面(a)至(f)中的下划线选项界定了一类具有基本网络调查模式特征的调查形式,但实际上它涉及我们在第 1.3.2 节中介绍过的、更广泛的基本在线 CSAQ 调查模式,其中输入(d)仅限于视觉选项,输出(e)仅限于手动电子化输出。在此情况下,基本网络调查模式只是基础在线 CSAQ 的一个特定子类型,具有以下三个子特征:在线传播媒介(c)是互联网;采用基于服务器的通信方式进行交互;而界面则是基于网络浏览器。因此,基本网络调查模式排除了在移动应用程序中运行的调查,这些调查独立于浏览器和与服务器的交互(参见第 5.1.5 节)。

若去除输入(d)和输出(e)渠道的限制,剩下的四个固有模式特征——自我管理(a)、交互式计算机化(b)、在线传输回答(c)和缺乏访员互动(f)——定义了一个更广泛的在线 CSAQ 调查模式系列,其中也涉及语音技术。当网络被专门用作自动在线传递应答的传输媒介时,被称为与网络关联型调查模式(*web-related survey modes*)。这是部分移动调查应用程序的情况,其中网络仅用于接收问卷并将其发送回服务器。

上述澄清表明了调查方式的多样性,并表明基本网络调查模式实际上只是一个非常具体的案例。尽管如此,仍需重申的是,当前绝大多数与网络相关的调查都是用这种方式进行的。更详细的阐述、示例以及调查方式的图形展示可以在第 6 章的补充材料“调查模式”(Survey modes)(<http://websm.org/ch6>)中找到。

6.1.2 总调查误差和调查数据质量

总调查误差(*total survey error*, TSE)是一个涵盖所有类型调查误差的总称。正如 Groves 和 Lyberg (2010)所讨论的,不同作者在处理总调查误差时存在一定分类和术语上的差异,而且仍在不断修订。例如,Smith (2011)提供了一个非常详细的总调查误差组成部分的分类体系。尽管如此,以下调查误差的类型是最常被提

及的：抽样误差、覆盖（抽样框）误差、无应答误差、定义误差、测量误差和处理误差（Biemer & Lyberg, 2003）。每种类型的误差都包含系统性和随机性成分，分别表示为偏差和方差。研究者仅在概率抽样的背景下对总调查误差进行定义和概念阐述；然而，在实践中，正如第2.2节关于抽样部分所述，研究者也会参照近似方法。

模式效应（*mode effect*）是测量误差的一个特定亚类型，也是总调查误差的一个子成分。它是由调查模式对估计值的影响而产生的。相应地，不同调查方式（例如面对面与网络）的回答分布差异是由特定方式的固有特征引起的。Biemer 和 Lyberg（2003）使用了更精确的术语“纯模式效应”（*pure mode effect*），以区别于更广泛的概念“模式系统效应”（*mode system effect*），后者还包括特定调查方式的其他相关误差来源（例如抽样、无应答、招募、激励、研究设计背景）。例如，这些模式系统效应是指当特定调查方式并不是随机分配给单位，而是由受访者通过某种未知的选择机制选择时产生的误差，这类误差往往难以与纯模式效应区分。

总调查误差的每个组成部分都可能损害用于推断总体参数的估计值的统计质量。这些偏差最常见的指标与抽样精度（用抽样方差表示的随机波动）、偏差（总体真实值与估计值的期望值之间的差异）、信度（重复测量中回答的一致性）以及效度（实际测量到真正想测量的概念的程度）有关。

准确性通过均方误差（*mean squared error*, MSE）来衡量，用于估计总调查误差的各种组成部分（Dietrich, 2008）。均方误差是总调查误差中包含的相关组成部分中随机方差和偏差平方的总和。然而，处理总调查误差所有组成部分的随机误差和系统误差并非易事。在实践中，相关讨论通常简化为抽样方差和由无应答或覆盖不全引起的偏差。最小化均方误差的目标可为使用网络调查的合理性提供依据：网络调查能以较低成本获取大规模样本，其较小的抽样方差可以抵消潜在偏差，而潜在偏差在网络调查中较为常见，主要源于较高的无应答率和/或覆盖不全问题。

虽然整合到总调查误差中的调查误差与估计值的统计特征（例如准确性）有关，但调查质量的概念要广泛得多（*survey quality is much broader*, Lyberg, 2012）。它包括调查结果（统计结果）的其他方面，如相关性、可比性、一致性、及时性、可获取性和清晰度（*relevance, comparability, coherence, timeliness, accessibility and clarity*）。更广泛的数据质量框架还涉及信任和用户的感知（*perception of the users*），以及相关生成过程的特征（*production process*）（Ehling & Körner, 2007）。类似的处理组件也包含在 ISO（2012）关于调查过程的标准中。各种通用的质量管理方法也可以应用于此，如六西格玛统计过程控制和全面质量管理（TQM）。专门的调查数据组织（例如统计局）通常会开发自己的（一般是非常复杂的）系统，通过标准化

的质量报告来监控数据质量 (Eurostat, 2009)。然而, 特定模式效应作为这些一般系统的组成部分, 目前文献对其的讨论相对较少且不够深入, 但我们相信随着混合模式调查的使用增加, 这种情况将会改变。

特别是对于大型和复杂的调查以及决策支持调查而言, 这些更广泛的调查质量维度非常重要。例如, 无论调查的准确性如何以及在其他方面质量有多高, 但是未能在特定时间 (例如在决策点) 提供结果, 意味着整个调查毫无用处。研究人员在专注于网络数据收集过程的方法学细节时, 有时可能会忽略这些额外的但非常重要的数据质量组成部分, 从而忽略整体视角。因此, 建议在网络调查的概念化、操作化和规划中, 以及在对该过程的常规管理监控中, 应明确纳入数据质量的各个方面。

6.1.3 对网络调查模式效应的评估

调查模式的复杂性及其与其他类型调查误差的相互作用, 使得对模式效应的实证评估要求非常高。因此, 在许多研究中, 我们无法真正分离出纯模式效应。模式效应还可能受到设计不佳的问卷 (例如错误的量表格式、过多的颜色或图片等)、受访者的问题 (例如计算机素养较低)、技术问题 (例如设备故障) 以及网络调查实施中其他错误等多重干扰。最重要的是, 纯模式效应通常与无应答、覆盖不足和样本选择机制相互作用, 这些因素之后难以有效区分。

在比较调查模式时, 研究者常默认 (或隐含地假设) 某一模式是准确的 (即黄金标准), 通常是面对面调查 (F2F)。我们需要非常谨慎地对待这种情况, 除非已知应答的真实值 (例如通过辅助数据), 否则我们宁愿谈论模式间效应 (*between-mode effect*), 或者确切而言是模式间差异 (*between-mode difference*)。模式间差异通常被观察为不同实施方式 (如实验组) 之间均值估计的简单差异。

此外, 模式效应研究还可以关注相关性或测量等效性 (Revilla, 2013)、可靠性 (Buchanan & Smith, 1999; Chang & Krosnick, 2009; Hertel et al., 2002; Mangunkusumo et al., 2006; Miller et al., 2002)、潜在结构 (Buchanan & Smith, 1999; Deutskens et al., 2004; Roster et al., 2004) 和共同效度 (Linchiati, Chang & Krosnick, 2009; Hertel et al., 2002)。

Spijkerman 等 (2009) 的研究比较了荷兰普通人群的非概率在线样本库调查 ($n = 57\ 125$) 和概率面对面调查 (计算机辅助个人面访 $n = 7\ 204$) 的结果, 这是比较调查模式而没有分离纯模式效应的典型研究。他们发现在线样本库调查中, 药物使用率明显更高 (例如, 15 至 24 岁年龄组中, 曾经使用大麻的比例为 39%, 而面

对面调查中则是 28%)。然而,目前尚不清楚这些差异是源自调查模式(与面对面调查相比,受访者在网络调查中承认使用更多药物),还是源自无应答、自我选择、抽样、无覆盖或涉及不同人群等方面的差异。在许多研究中,模式效应常常与抽样框问题(Burr, Levin, & Becher, 2001; Tomsic, Hendel, & Matross, 2000)、样本差异(Braunsberger, Wybenga, & Gates, 2007; Chang & Krosnick, 2009; Roster et al., 2004)、电子邮件地址的可用性(Deutskens, de Ruyter, & Wetzels, 2006; Griffis, Goldsby, & Cooper, 2003),以及用于减少无应答的顺序混合模式设计(Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003)等因素混淆。这些不受控制的差异可能被错误地归因于纯模式效应。

剥离纯模式效应的标准方法是开展实验研究,随机分配样本单位至采用不同模式的子样本中。然而,无应答和相当大的随机误差仍然无法与模式效应分离(Vannieuwenhuyze & Loosveldt, 2013)。人口加权可能会有所帮助,但其作用范围通常是未知的,且仅提供有限的帮助(见第 4.1.4 节)。一些特定的方法(Vannieuwenhuyze, Loosveldt, & Molenberghs, 2010, 2012),例如在混合模式的背景中使用单一模式组,可能会部分减少这一问题。

更优的研究设计是在实验室中进行随机化,正如 Chang 和 Krosnick (2010)所述,参与者在到达实验室后才被随机分配到实验组。同样地,Jäckle、Roberts 和 Lynn (2006)报告了一项实验,研究者在街头招募参与者,他们同意合作后,将之随机分配到四种模式中的一种。通过这种方法可潜在减少无应答差异,但也存在外部效度问题,即在人工(实验室)实验条件下得出的发现是否也适用于现实生活条件。

另一种衡量纯模式效应的方法是重测法(*re-surveying*)(测试-重测或重复测量),这种方法可以有效控制组间样本、覆盖范围以及无应答的差异。因为是在两次调查中测量相同的单位,这种方法还显著降低了抽样方差,从而极大地提高了发现潜在差异的能力(Alwin, 2007)。然而,该方法实际实施起来较为复杂且成本高昂,这一点从采用它的研究数量有限便可见一斑(Mangunkusumo et al., 2006; Miller et al., 2002; Rivara, et al., 2011)。此外,受访者可能会记住第一次调查中的相同问题,而且两次测量之间的一些外部效应或事件可能会改变目标变量的值。为了控制这些因素,还需要另一个实验组(即未经第一次调查处理的“事后”对照组)。调查模式应用顺序也可能产生影响,为了正确控制这些影响,还需要再设立另一个实验组,其处理顺序与原组相反。

多项研究表明,网络调查模式通常优于传统的面访(面对面)和电话调查模式,

也优于纸笔自填模式，或至少与它们相当（Börkan, 2010; Braunsberger et al., 2007; Buchanan & Smith, 1999; Chang & Krosnick, 2009; Hertel et al., 2002; Mangunkusumo et al., 2006; Roster et al., 2004）。具体而言，Tourangeau、Conrad 和 Couper（2013）的一项元分析研究表明，对于通常在传统调查模式中被低估的敏感行为，受访者在没有访员在场的网络调查中更倾向于给出更真实的回答。第 1.3.5 节已指出，自填方式在其他方面也有优势（例如受访者可以自行选择填写问卷的时间和节奏）。

另一方面，网络调查模式的特定效应有时可能会产生负面结果，这主要是由于各种形式的“走捷径”行为，例如满足作答、过多无应答（或“不知道”）选项、快速作答、回答不一致等。这种情况大多是由于缺乏访员参与，在某些情况下访员是必不可少的，例如需要激励（例如长问卷）和澄清问题（例如复杂问题）时。但通过精心结合激励措施与交互式问卷设计，可在很大程度上弥补访员缺席的不足。

6.1.4 成本—误差优化

在讨论不同调查模式以及调查数据质量的差异时，成本因素往往被忽略。于是问题随之而来，在没有控制成本的情况下，将网络调查模式与其他替代方案进行比较是否准确和公平？也就是说，只有在固定预算下进行比较——而不是通常采用的固定样本量的做法（例如，每种模式的样本量 $n = 1\,000$ ）——才可能表明网络调查能够让我们负担得起更大的样本量。或者，网络调查节省的成本可用于激励措施、混合模式设计等改进措施，从而显著提高数据质量，并改变比较的结果。尤其令人惊讶的是，尽管成本往往是我们在选择调查模式时最重要的因素，但文献中缺乏对成本的考虑。

如第 1.3.5 节所述，网络调查中某些固定成本通常相对独立于样本量，而单位可变成本则相对较低，并且取决于样本量。计算受访者每分钟的总成本（*total costs per minute*）是一个有效的工具，可用于比较网络调查与传统替代方式。例如，在发达国家，像一般社会调查这样的基于概率抽样的面对面（F2F）调查，其每分钟的成本通常是几欧元 / 美元，有时甚至超过 5 欧元或 10 美元，而基于概率的在线样本库调查的每分钟总成本通常在 1 到 2 欧元 / 美元之间。来自非概率样本的网络调查，其每分钟的成本可能要低得多，但这种成本差异很大，主要取决于应答质量。成本需要与数据质量相平衡，这是一个非常复杂的问题，当涉及混合模式时可能会更加复杂。在概念层面，这些问题相对较早得到了研究关注（Groves, 1989），

但在实践中我们仍然缺乏有效的操作标准和程序。Vehovar 等人（2010）提出了同时优化成本与准确性的方法，即最小化成本与均方误差（MSE）的乘积。Roberts、Vendenplas 和 Stähli（2014）以及 Vannieuwenhuyze（2014）则提出了具体操作方法，而 Schouten、Calinescu 和 Luiten（2013）在交互式设计的背景下讨论了成本优化，该背景下成本—误差优化直接影响正式调查的实施顺序。

6.1.5 核心网络调查过程之外

到目前为止，我们观察到的核心网络调查过程似乎是完全独立的，并且处于一个简化的预调查、调查实施和调查后处理的流程中。然而，实际情况很少是这样——网络调查通常更为复杂，并且嵌入到更大的研究背景中。因此，下文将回顾其几个最重要的方面。

6.1.5.1 初步研究活动的步骤

第 1.2 节（图 1.1）将前期研究活动定义为核心网络调查过程之前的准备工作。与预调查阶段的区别在于，这些活动并不直接涉及网络调查的具体内容；相反，关注的是一般的准备工作，而这些准备工作与任何调查模式无关。然而，它们对于网络调查的整个背景来说是非常重要的。这些前期的研究活动通常——但现在总是——包括以下组成部分，我们采用一个简单、非结构化且非详尽的列表来呈现它们：

- 对一般问题的澄清，如最初的研究思路、更广泛的（研究、商业、行政）框架、基本的范式和概念设置、问题的表述、目标确定和研究问题定义；同时还包括研究问题（例如客户满意度测量）与最终决策问题（例如如何提高客户满意度）之间关系的详细阐述；
- 进一步阐述研究设计方面，例如对关键概念进行更具体的概述、构念的操作化、研究假设的提出、所使用研究方法的说明、目标人群的定义，以及有时还包括对关键问题的选择或阐述；
- 伦理方面的考虑，这也可能涉及潜在的正式审批程序；
- 关键的行政、管理和过程质量问题，包括时间、成本和其他资源等重要方面；
- 在复杂的研究设计中，此处会开展某些形式的初步研究活动，尤其是二手资料研究、定性研究和专家咨询；

- 当研究者角色（前文将其视为单一实体）分散于发起方、委托方、研究提供者、软件供应商等多方时（见第 1.2 节），会产生额外活动，这将大幅增加初步研究的复杂性，需要额外的行动、互动与流程支持。

尽管并非上述所有活动都会出现在每次调查中，但其中大多数都会出现，至少以某种非结构化和非正式的方式出现。即使在简单的自助式研究中，也需要一定的决策和活动，才能对上述大多数问题找到相应的解决方案。在更复杂的研究项目中，方法学方面通常在正式的研究设计（*research design*）中加以阐述，该设计会明确所使用的研究方法，以及相关研究活动的阶段、顺序和其他具体细节。一般而言，根据研究的探索性、描述性或因果性特点，可以涉及各种定性和定量研究方法。网络调查的作用需要在总体研究设计中确定。当然，上述一些前期研究活动可以在一定程度上与预调查阶段重叠，特别是早期对方法的详细阐述（例如调查问卷已经由委托方预先确定）。

另一类规范化工作与研究计划（*research proposal*）有关，其概述拟议研究的所有重要方面，而非仅局限于方法论。在 Iacobucci 和 Churchill（2015）的书中可找到相关实践案例。通常，研究计划需经过管理层、委托方、发起方或其他资助机构的审批。

当研究角色分散于多个实体且需明确彼此关系时，研究计划尤为重要。在这种情况下，各种形式的招标或研究基金申请通常会频繁涉及其中。这需要额外且细致的行政工作（申请表格、流程指南、法律问题、选择流程及标准）。在涉及网络调查时，除价格外，调查过程的质量指标（如应答率）应成为决策的关键依据。

由于初步研究活动通常复杂、结构化且与调查紧密关联，文献中很少将其作为整体讨论。相反，各种组成部分或方面通常是分开讨论的（例如研究范式、研究方法、采购程序、法律问题）。为了更全面地了解前期研究活动的方法学方面，读者可以参考一般性的调查方法文献（Blair, Czaja & Blair, 2014; Groves et al., 2009）、社会科学方法文献（Babbie, 2013）或特定学科的文獻（例如市场研究；Iacobucci & Churchill, 2015）。其他方面很少专门针对调查背景进行阐述。例外的是，Iacobucci 和 Churchill（2015）、Iarossi（2006）以及 Sue 和 Ritter（2012）的著作对招标流程细节有所涉及。

6.1.5.2 高级分析、处理与价值转化阶段

如第 1.2 节（图 1.1）所定义，本阶段属于广义调查活动的范畴，与网络调查的

具体内容无直接关联。因此,这些活动是核心网络调查过程的后续环节,且如第4章所述,与调查后期步骤存在明确区分。

相应的活动主要涉及高级分析(*analyses*),包括报告撰写、统计分析程序、方差估计、结果的实质性解释以及调查报告和演示文稿的准备。

这一步骤还包括各种高级形式的调查后数据处理(*data processing*),例如将调查数据整合到更广泛的行政或商业流程中的管理与决策程序中。调查数据可能经过进一步处理,涉及数据合并、记录关联、统计匹配(即数据融合)及其他多元应用场景,这些操作正日益成为高级研究及行政或商业流程的重要组成部分。

价值转化部分包括对调查结果的深度利用(在其他研究或决策过程中使用和再使用)以及通过公共存档数据文件和发布、推广、发表(在简报通讯、研究摘要、文件、完整报告以及学术论文和书籍中)研究结论来传播结果。

6.1.5.3 混合模式

我们在第2.1节中介绍了混合模式的概念,将其定义为在测量阶段结合多种调查模式的组合形式;并且使用“混合模式系统”这个术语来涵盖招募阶段的多模式联系,以及在实施阶段使用多种沟通渠道所涉及的其他方面和后果。除此之外,我们还讨论了某些更广泛的实施方面。

首先,需要注意的是,在商业调查中,网络调查模式和纸笔调查或电话调查模式通常会常规地结合使用,通常是以标准的顺序方式,从成本较低的(网络)模式开始,逐步过渡到更昂贵的(邮寄、电话)模式。

其次,我们应该指出,尽管存在各种关于混合模式(理解为在测量阶段结合多种调查模式)使用不断增加的说法和指标,但总体而言,情况并非如此,至少在市场研究和自助式领域并非如此。在市场研究中(Macer & Wilson, 2013, p. 40),混合模式调查在市场调研行业的营收占比相对较低(混合模式为6%,而网络调查为52%)。此外,在过去的七年里,这一占比已经非常稳定,且预计未来不会增加。这并不令人意外,因为高质量的混合模式调查通常非常复杂且成本高昂,相比之下,网络、邮寄或电话调查这些单一模式目前在市场研究领域占据主导地位。

在昂贵的基于概率的面对面入户调查中,情况则不同。由于此类调查存在抽样、招募、未覆盖和无应答等问题,网络调查不能简单地取代传统调查模式,但是可以通过混合模式进入到调查过程中。然而,正如第2.1节所讨论的,这在方法学上比较复杂。尽管如此,在最初的实施(Beukenhorst & Wetzels, 2009)之后,涉及网络调查模式的混合模式似乎正在慢慢成为这类调查的主流(Buelens & van Den brake,

2014)。

在官方统计中, 一个重要的混合模式应用挑战与劳动力调查 (*Labour Force Survey*) 有关, 该领域在引入网络调查作为数据收集模式方面可以观察到持续的进展, 相关案例可见 Körner 和 van Der Valk (2011) 的研究或 DCSS 项目 (Blanke & Luiten, 2014)。

在学术调查中, 我们也可以观察到在考虑混合模式数据收集方面的进展, 但要高度谨慎, 如爱沙尼亚 (Ainsaar et al., 2013) 和英国 (Villar, 2013) 的欧洲社会调查 (ESS) 混合模式实验所示。

让我们举一个典型的案例来说明混合模式存在的问题。斯洛文尼亚健康研究所最近进行的一项烟草、酒精和药物使用情况调查 (平均时长 40 分钟)¹ 采用了标准的顺序混合模式调查方法 (网络→电话→面对面)。在完成网络调查后, 累积应答率为 20%, 在完成电话调查后, 累积应答率为 45%, 在完成面对面调查后, 累积应答率达到 60%。与传统的面对面调查方式相比, 应答率和成本节约取得了巨大成功, 同时也纳入了在传统调查模式下不会合作的群体。然而, 模式效应的影响仍然不清楚。例如, 在对每个模式群体的社会人口统计特征进行加权后, 大麻使用率 (曾经使用过大麻) 在网络、电话和面对面模式群体中分别为 25%、10% 和 9%。问题在于, 现在不可能将模式效应与模式的选择 (阶段)、差异性无应答偏差、调查过程阶段和调查人口的差异区分开来。因此, 通过网络调查收集的应答比例不断上升对未来趋势的影响尚不清楚。如果大麻使用率的估计值增加, 我们可能无法确定这是因为人们真的消费了更多, 还是仅仅是因为越来越多的人在采用网络作答, 而他们在网络上更愿意承认较高的购买量。

解决这一问题的可行方案是开展平行实验, 即对部分单位进行面对面调查, 以便可以估计纯模式效应和模式间效应。或者, 可以使用不同的调查模式对单位进行重新访谈。这些方法在 Klausch、Hox 和 Schouten (2014) 的实验中得以检验, 各种模式效应随后被分离出来。另一个可能的解决方案是采用先进的加权方法 (Buelens & van den Brakel, 2014)。似乎我们需要同时使用这两种方法: 一方面基于模式效应的外部估计 (例如通过重新访谈获得) 进行一些模型假设; 另一方面采用先进的加权技术。

6.1.5.4 混合方法

混合方法是指将数据收集的定量方法 (如调查) 与定性方法 (如深度访谈、焦点小组、专家解释、叙事研究、行动研究、民族志研究和头脑风暴) 相结合的研究。

究设计。与调查类似,定性研究技术也越来越多地转向网络(例如通过聊天进行的在线深度访谈或在线焦点小组)。

学界已在范式、概念及操作层面针对混合方法展开广泛讨论,有时甚至将其视为独立于定性与定量研究之外的第三种研究范式(Johnston, 2007)。近年来,混合方法在文献与实践中的应用均呈扩展趋势(Morgan, 2013)。

需要补充的是,在网络调查过程中,研究者出于多种方法论目的的考虑,会在设计问题、编制问卷或评估调查题目时借鉴定性研究的见解。对于实质性方面,研究者通常在解释中涉及定性研究,使用专家评估或来自目标人群的定性反馈。

互联网非常适合多种方法的结合。在传统的调查模式中,超出两个步骤的方法组合(例如焦点小组→调查)往往需要大量的时间和资源。相比之下,互联网使得定性方法和定量方法的结合变得简单、低成本且高效。Lobe 和 Vehovar (2009)表明,定性(在线访谈)和定量(网络调查)方法的序列可以通过迭代方式深化和提高研究的深度和质量,并可以提供相当的优势:在给定的资源范围内,四步法的研究设计(定性→定量→定性→定量)可能优于单步法的研究设计(定性或定量),也可能优于更传统的两步法(定性→定量)或三步法(定性→定量→定性)的研究设计。

混合方法尤其适用于所谓的众包(crowdsourcing)以及在线社区研究中的各种其他方法的应用场景。在这些方法中,企业与市场研究会建立并维护目标人群的在线社区,用于开展定性与定量数据的协同收集。

6.1.6 迈向交互式调查设计

根据最初讨论的范围(第1.2节),我们在前面的章节中主要阐述了横断面网络调查,依赖于单次测量交互。在这种情况下,实地调查过程相对简单,尤其是开展基于名单的网络调查时,仅需通过向所有调查对象发送电子邮件邀请并推送提醒即可完成。

另一方面,调查设计的复杂性可能源于更复杂的方案。调查数据的收集方式多样:既可以对独立样本重复实施一系列相同调查,也可以对独立样本开展一系列互补调查,还可以通过各种类型的纵向设计来收集。此外,不同的外部数据可以迭代地纳入网络调查过程,如行政数据(例如使用公民的特定表单或问询记录)、商业数据(例如消费者的商业交易记录)或技术数据(例如用户在网络上执行的某些操作的日志文件信息)。

当涉及混合模式系统时,复杂性会进一步增加,尤其是在招募阶段(第3.1节)中相对固定的招募计划及有限的实施变体,被各类交互式实地调查策略取代时。

自适应设计 (*Adaptive design*, Schouten et al., 2013) 基于下述事实, 即调查实地工作通常面临许多不确定性, 而这些不确定性可以通过使用实地工作本身产生的数据来降低。也就是说, 通常情况下, 样本中的所有单位都接受相同的处理 (*treatment*), 这意味着他 (她) 们会收到相同数量的提醒, 相同的邀请模式和相同的调查测量模式。某些辅助数据 (如年龄、地区、成本结构、对应答率的估计) 被用于构建和优化抽样或研究设计 (如分层抽样、多阶段抽样、双重抽样、重复调查、混合模式)。然而, 这些工作是在开始调查之前完成的, 不包括来自实地调查过程的反馈。

自适应设计通常在响应式设计 (*responsive design*, Groves & Heeringa, 2006) 这一较为狭窄的背景下讨论。响应式设计要求调查分阶段实施, 前一阶段的数据会被用于优化下一阶段 (Wagner et al., 2012)。如果某些子群体可能有较低的应答率, 因此响应式设计在下一阶段会分配额外的资源 (例如增加联系次数)。

值得注意的是, 自适应设计和响应式设计主要与面对面调查或电话调查相关 (Schouten et al., 2013)。尽管丰富的辅助数据 (*paradata*)、灵活性、信息通信技术 (ICT) 支持和复杂的监测实际上提供了一个更便利的环境 (Bianchi & Biffignandi, 2014), 但它们在网络调查中的应用却相当罕见。

在此, 我们将自适应设计的概念扩展至更宽泛的交互式调查设计 (*interactive fieldwork design*), 这意味着在 (a) 实地调查工作期间生成的 (b) 单位层面的任何数据——即调查数据、辅助数据以及外部数据 (技术、行政、商业)——被用于对与招募和测量阶段相关的 (c) 实地调查工作程序进行实时优化。

上述设计方式与第 3.1 节中讨论的实施变异性适应和交互式调查设计之间的界限有时非常模糊。例如, 我们预先在招募计划中指定了确切标准, 在此情况下, 当应答率较低时, 就需要激活补充样本、发送额外提醒或启动额外的推广活动。二者的主要区别在于, 交互式设计还包括在单位层面的干预, 而不仅仅是针对部分群体层面 (*segments*)。例如, 并非简单地向已经收到三次电子邮件邀请、但未回应的所有人发送邮件提醒, 而是利用单位层面的附加信息 (如年龄、性别) 分别处理每个对象, 并在该层面决定具体行动。

图 6.1 (更新自 Vehovar & Batagelj, 1996) 在混合模式背景下对交互式调查设计中潜在的流程进行了概念化呈现。该流程涉及单位层面的信息和数据通信过程, 这与图 1.1 (第 1.2 节) 中对一般网络调查过程的概念化不同。集中处理和监控交互式地连接了两项基本的调查活动: 招募和测量 (即问卷填写环节)。不同的调查模式可以用于招募受访者 (即多模式联系), 并且可以重复尝试联系, 直到确定无法获得应答 (例如我们得知该单位不符合条件)、实地调查工作时间已经结束, 或者尝

试进一步联系某个单位已不再合理。关于后者的决定可以基于简单原则（例如最多五次联系）、无应答的背景（例如某个部分的最低应答率或最低应答数量）、统计参数（例如某些估计所需的精度）或某些复杂的成本—误差标准，它们通过成本（例如成本和均方误差的乘积）来优化数据质量。

可以针对不同的单位制定差异化决策，以优化时间、成本和数据质量。例如，如果某单位在两次电子邮件邀请后仍未回应，智能的集中处理——具备该单位的所有辅助信息（例如年龄、性别）以及所有实地工作数据、辅助数据和外部数据——可能会根据某种优化算法自动决定下一次联系尝试的最佳时机、方式和格式。例如，可能最适合在一周或两周后发送下一封电子邮件提醒，或者改为邮寄提醒，甚至最适合的调查方式可能是纸笔问卷（P&P）而不是网络问卷。

依托集中化信息通信技术（ICT）支持的处理系统，可显著提升整合性、交互性与优化程度，这尤其得益于其能借鉴过往调查经验，并引入统计与决策模型。我们还可以进一步优化单位层面的测量环节结构，该环节甚至可以被拆分成多个阶段。

在上述讨论和图 6.1 中，我们讨论的是基于名单的网络调查，但在非基于名单的网络调查中，情况会略有不同：研究者不直接向个人发送邀请，而是通过改变推广策略（例如横幅广告）与单位进行互动，因此互动的重点将转向修改在线招募工作上（第 2.2.3 节）。例如，研究者可以决定减少直接横幅广告，同时增加在在线社交媒体上的推广。

交互式框架设计需要在预调查阶段加以考虑，可以补充或替代招募计划。在招

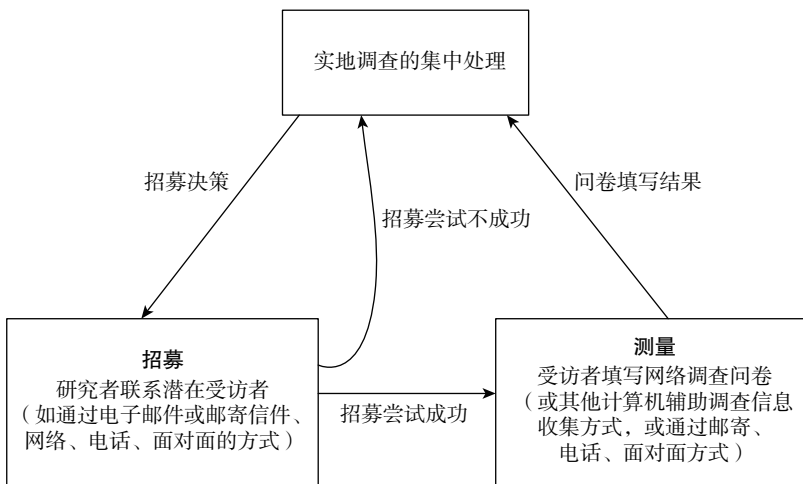


图 6.1 基于名单样本的交互式混合模式调查设计中的单位层面流程

募计划中, 我们需进一步扩展与时间、成本和数据质量相关的设计特征组, 以及监控指标, 这些指标将提示需要采取行动或进行调整的情况。如前所述, 实时计算功能可以通过智能信息与通信技术 (ICT) 的支持进行整合和扩展。

文献中也有所提及混合模式设置中整体优化的某些内容 (Schouten et al., 2013; Vannieuwenhuyze, 2014; Vehovar et al., 2010)。当然, 通过结合混合方法及与外部行政、商业和 / 或技术流程的交互, 迭代式调查设计可进一步扩展, 不再局限于混合模式。

6.2 项目管理框架内的网络调查

在第 2.6 节中, 我们讨论了与核心网络调查过程相关的综合管理问题, 而在本节中, 我们将讨论项目管理的更广泛背景。

网络调查可以被视为与任何其他项目一样, 通用的项目管理原则在此同样适用: 研究者需对各项活动进行组织、规划与协调, 以在既定时间与资源约束下达成目标。针对这类任务——尤其是一次性、首次开展或内容重复但形式多变的工作——专业的项目管理标准已逐步发展成熟。

项目管理有助于规范化工作描述 (*description of work*, DoW) 的制定, 将任务和活动结构化为工作包 (*work packages*, WP), 并实现资源的最优分配等。它还涉及计划和时间安排, 包括里程碑 (即关键日期) 和可交付成果 (*work packages*) 的确定。在项目管理的语境中, “可交付成果” 指的是报告、结果和其他可衡量的项目输出。在这里, 提供控制以及监督、协调和干预也非常重要。

开展网络调查项目时, 研究者需协调大量活动。从图 1.1 (第 1.2 节) 中展示的网络调查过程的结构可以看出, 12 个阶段的活动相互关联。这些阶段可直接转化为项目管理计划中的工作包结构。仅在问卷准备过程中已经出现了相当大的复杂性 (见第 2.3.6 节中的图 2.1), 因为测试和问卷开发与复杂的管理过程密切相关。此外, 初步研究活动以及与核心网络调查过程外部环境的其他互动也带来了额外的复杂性。

因此, 网络调查项目需要周密的计划与协调, 且潜在风险点较多。此时, 管理技能的重要性不亚于成功实施调查所需的方法论技能, 甚至更为重要。我们可以用几个与复杂程度相关的案例来说明项目管理任务:

- 在事后评估调查中——我们需要收集参与者的反馈，同时也处理他们的电子邮件——项目管理可以是一个较为简单的流程：准备问卷、设置电子邮件、发送邀请和提醒，然后生成初步报告，并在线发布。对于经验丰富的研究者而言，如果有合适的软件和标准化的调查问卷，所有这些操作可能只需要几分钟。
- 在上述案例中，如果我们增加新问题的开发与预测环节，则会涉及更多活动，因此整个项目可能需要研究者花费数小时。无论如何，这个例子充分展示了网络调查的力量，即其高效性与便捷性，尤其是如果受访者通过移动设备立即回复时。其时间和成本效率是任何传统调查模式的替代方案无法比拟的。
- 当为委托方开展调查或需要组建项目团队时，情况会变得更加复杂。此时需处理额外的互动、沟通及各类复杂问题。一个典型的网络调查项目，包括某些实质性的前期调查、网络调查的实施和最终报告，可能需要小型团队投入数周甚至数月，总计数百小时。所有这些可能还需要至少在最低限度上进行形式化的项目管理任务。
- 随着复杂性的增加，项目管理所需的资源通常会快速增加，而与此同时，核心网络调查过程本身所占的比例则逐渐降低。运行预算接近或超过六位数（例如10万美元）的网络调查已属于严谨的系统性工程，需要采用正式的项目管理机制。
- 在线样本库、纵向研究、跨国研究以及来自不同组织和部门的合作伙伴参与的研究中，项目管理的重要性进一步增加。在国际研究中，仅翻译和回译本身就已经成为一个严峻的项目管理挑战。

专业的数据收集机构，无论其所属领域（商业领域还是公共领域），通常都有自己详细制定的项目管理规则。而另一方面，不专门从事调查数据收集的组织，特别是学术和非营利部门的机构，也参与开展网络调查。有时，其研究者并没有准备好管理复杂的网络调查项目，因为不能仅靠一般知识和常识来处理管理问题。下面我们将概述一些典型的项目管理问题，如果这些问题缺乏足够资源与关注，可能引发风险：

- 项目团队会议的适当形式（即结构、时长与频率），以及会议纪要的严格记录、监控和报告。
- 团队成员和委托方之间清晰的（如书面）责任划分，包括沟通策略（例如委托方是否可以直接与网络程序员沟通）
- 质量保证标准和相应的评估程序，这些需要提前在独立文件中正式确定，并且

引入外部专家监督或指导委员定期监督项目也会很有用,尤其适用于新的、长期的或复杂的项目。

- 提前细化基本管理程序,包括初步计划、应急计划、风险管理、危机管理和冲突解决方案。
- 对核心网络调查过程进行正式监控,特别是正式调查(例如,应答率、数据质量)。
- 定期控制资源,包括花费的时间,并通过关键时间点、截止日期、关键路径和甘特图等监控时间安排,这是项目管理中的标准监控方法。
- 提供用于记录、备份、归档和沟通的平台。该平台最好超越电子邮件附件,并允许在线访问最新的材料(例如问卷草稿)。
- 明确规定(如合同)关于数据所有权、知识产权以及在数据收集延迟或错误情况下的处罚条款。

上述内容适用于任何调查研究项目,而对于网络调查而言,通过使用集成的网络调查软件和其他在线服务,有许多额外的机会来缩短、简化和优化管理任务。

基本的专业资源与具体的管理方法有关,这些方法中也详细阐述了标准和认证,例如 PMIbook²、IPMA³ 和 PRINCE2⁴。此外,也有许多项目管理的教科书和指南,但令人惊讶的是,直接涉及调查的(Stouthamer-Loeber & van Kammen, 1995)或间接涉及调查的(Lyberg et al., 1997)专门讨论却寥寥无几。同样,关于调查方法的教科书很少有专门的章节涉及项目管理,只是在调查实施的章节中部分涉及这些问题。Iarossi(2006)以及 Kennedy、Tarnai 和 Wolf(2010)的著作涉及调查项目管理问题的更结构化的讨论,但并不是专门针对网络调查,而是针对一般调查。

6.3 网络调查专业

进行网络调查不仅需要熟悉调查方法,还需要熟悉信息通信技术(ICT)、统计学、管理学、法律法规以及更广泛的社会背景,包括伦理问题。相关的能力体现在多种专业的交叉领域。此外,是否可以将网络调查方法视为一个独立的专业领域也值得探讨。

一般来说,当以下要素存在时,我们才会谈到一个“专业”:正规的教育项目、明确的职业描述、专业期刊、专门的国际协会和活动,以及某些通用规范和标准。对于网络调查方法而言,这些属性大多缺失。然而,它们确实存在于调查方法这一

更广泛的领域,网络调查方法主要属于这一领域。

另一方面,确实存在许多与网络调查方法相关的特定专业活动,包括定期举办的活动、专门的课程(特别是在各种暑期学校中)、众多研究项目、专业网络、专门的标准和教科书。值得一提的还有专门的会议议程、数千篇学术论文、演讲和其他材料,包括众多在线资源,以及数百种网络调查软件包,这些软件包通常包括相关的方法指南。WebSM (<http://www.websm.org>) 对这些方面进行了详尽的概述。

从这个角度出发,我们可以讨论特定的专业知识以及与之相关的专业要素,而本书就是这一领域的见证。在本节的其余部分,我们首先回顾应用网络调查以及开展方法论研究的科学学科(第6.3.1节)。我们还将讨论相关的法律问题(第6.3.2节)和伦理问题(第6.3.3节)、正式的准则和标准(第6.3.4节),以及非正式的指南和其他资料(第6.3.5节)。

6.3.1 相关学科

网络调查在形式上隶属于调查研究框架,且处于社会科学方法论的更广泛背景中。从最狭义来讲,社会科学方法论主要与社会学相关,在较小程度上也涉及政治学和传播学;而这些领域中网络调查的实施,均与网络调查方法高度关联。

在更广泛的意义上(e.g. OECD, 2002),社会科学以及网络调查的应用还包括其他三大分支:心理学、经济学和教育学;地理学、城市规划、人口统计学、管理学、公共行政学、法学、管理学、组织科学;以及各种具体领域,如评价研究、社会人类学和语言学。尤其是嵌套于更广泛的经济学和商学领域内的市场研究,对网络调查尤为重要,因为大多数专业网络调查都是在这些领域开展的。网络调查实施的另一个非常重要的领域,是在线研究和互联网研究的特定跨学科背景。

除了社会科学,网络调查还出现在其他科学领域,尤其是人文科学和医学(健康研究)中,同时也出现在各种自然科学和技术科学中。统计学、计算机科学和信息学(如人机交互, HCI)领域与网络调查方法相关密切。

为理解网络调查在不同场景下的作用,需认识到诸多学科已形成自身独特的方法论体系。实施网络调查时,应结合相应领域的专业知识,可以参考相关的教科书:市场研究领域可参考Iacobucci和Churchill(2015)的著作,健康研究领域可参考Eysenbach和Wyatt(2002)的著作,组织研究领域可参考Bocarnea、Reynolds和Baker(2012)的著作。

网络调查的一个非常自然的方法论背景是社会科学方法论,它为网络调查提供

了更广泛的方法论框架 (Babbie, 2013)。具体而言, 这一框架涉及问题表述、研究步骤、决策与研究问题的关联、概念化与操作化过程, 以及对定性和定量研究方法实施的整体认知。其中, 调查研究背景在这里尤为重要。在本书中, 我们多次引用了该领域的标准教科书, 即 Groves 等人 (2009) 以及 Biemer 和 Lyberg (2003) 的教科书。

另一方面, 对于各种具体内容, 需要参考其他学科。例如, 调查数据分析需要统计学知识 (Heeringa et al., 2010), 而网络可用性涉及人机交互 (HCI) 的专业知识 (Roe, 2008)。

根据以上对相关学科的概述, 网络调查方法论的各个方面分布广泛, 这也是调查方法论的普遍特征。因此, 相关内容可以在各类专业协会中找到, 这些协会主要涉及调查研究 (AAPOR, CASRO, ESRA, ESOMAR), 但也与互联网研究 (AoIR, AAAS, APA)、社会科学方法论 (ISARC33)、统计学 (ISI, IASS, AMSTAT, RSS) 和市场研究 (ESOMAR, AMA, MRS) 有关, 还有些与计算机科学和信息学相关 (ASC, ACM, IFIP-HCC)。大型组织 (如欧盟统计局) 也可能开展与数据收集新技术相关的特定活动, 如“统计学新方法和新技术” (NTTS) 会议。上述专业组织都不是专门从事网络调查的, 其开展的相应活动和事件也是如此。尽管如此, 确实存在专门针对网络调查方法论的持续性活动: 即每两年举办一次的互联网调查方法研讨会 (Internet Survey Methodology Workshop)⁵ 和每年举办一次的互联网调查与调查方法国际研讨会 (International Workshop on Internet Survey and Survey Methodology)⁶。

6.3.2 法律问题

网络调查与法律法规密切相关, 尤其是在开展国际研究时, 这可能会变得非常复杂。总体而言, 鉴于该问题的复杂性, 在简要概述中, 我们只指出并说明最关键的内容。更具体的信息和指导可以从当地的市场研究、统计或调查研究协会获得。

尽管存在全球化的趋势, 但各国的法律环境差异仍然很大, 尤其是在刑事规范、隐私保护、知识产权、电信规则及消费者权益保护等方面。以下为网络调查涉及的核心法律问题:

- 未经请求的电子邮件邀请: 是否可以向在互联网上公开的私人或商务电子邮件地址发送未经请求的电子邮件邀请?
- Cookie: 法律是如何规定 Cookie 的使用的?

- 垃圾邮件：具体的反垃圾邮件法规是什么？
- 激励措施：允许何种类型和级别的激励措施，以及相应的行政程序是什么？
- 抽奖或有奖竞赛：从立法的角度来看，抽奖或有奖竞赛是如何安排的？
- 隐私、保密性、匿名性：这些问题在法律上是如何规定的（技术讨论见第2.4.3节）？如果研究者将收集到的应答挪作他用（例如用于营销），该如何处理？
- 辅助数据：谁有权获取、处理、分析、发布和存档这些信息？
- 未成年人：儿童如何参与网络调查，他们参与的最低年龄是多少？
- 知识产权：使用某个特定问题是否需要获得许可？如何使用？某些问题的版权是如何确立和保护的呢？
- 数据存档：如果数据丢失，谁来负责？数据存档的程序是什么？
- 软件：如果软件漏洞造成损害，谁应该承担责任？
- 非法内容：如果研究者或受访者将非法内容上传到网络问卷中，谁该承担责任？
- 数据泄露：如果能够识别个人信息的表格或数据被发布，谁负责？
- 数据安全：如果未经授权的人在（非）加密回答过程中，通过盗用帐户或直接侵入服务器获取数据，谁应承担责任？

当网络调查以托管服务（SaaS）的形式运行时，取消政策（*cancellation policy*）和税收（*taxes*）问题也可能浮现。此外，受访者、研究者和网络调查服务供应商之间的权责关系也需要明确。如果托管服务与软件供应商分离，情况会变得更加复杂，更不用说所有这些主题和服务都在不同的国家，具有不同的立法。

网络调查软件的使用条款及法律说明日益详尽，通常包含隐私、反垃圾邮件、Cookie、安全、奖励机制、取消政策等独立章节，这一趋势并不意外。这种复杂性的一个很好的例证是流行的网络调查软件服务，我们可以在其中找到关于法律问题（例如使用条款、隐私问题）的最新详细说明。然而，这种详细的处理方式仍然不是标准做法。WebSM的一项研究（Vehovar, Čehovin et al., 2012）发现，只有60%的供应商在其官方网站上包含隐私声明，而只有44%的供应商列出了使用条件。

6.3.3 一般伦理考量

伦理困境存在于网络调查过程的所有步骤和阶段。一般而言，这些问题涉及网络调查过程相关的数据、程序、行动和不作为是否可能会给他人或群体造成潜在伤害。我们所说的“伤害”是指某些负面的身体、心理或社会体验以及相关的感受。

是否提前告知网络问卷的预计完成时间是一个典型的问题, 尽管相对轻微。当预计完成时间超过 30 分钟时, 合作率有下降的风险 (Crawford et al., 2001; Galesic, 2006)。同样, 前文已经提到, 寻求明确同意收集辅助数据会显著降低合作意愿 (Couper & Singer, 2013), 这对整个知情同意 (*informed consent*) 方法 (Losch, 2008) 构成了挑战。该方法要求受访者在参与调查之前明确同意, 他们知晓研究目的、自身权利和潜在危害。

一个更极端的例子是受访者在网络调查中报告各种敏感信息, 从非法活动 (如偷窃、毒品交易、非法持枪) 和可能的严重心理状态 (如抑郁、自杀企图), 再到严重的犯罪问题 (如谋杀未遂、恋童癖行为)。在这种情况下, 遵守匿名性和保密承诺可能会给研究者带来非常严重的伦理困境。我们在第 6 章补充资料 “网络调查中伦理困境案例” 中提供了网络调查中的典型伦理困境案例 (<http://websm.org/ch6>)。

与法律问题的差异相比, 不同国家在伦理方面的处理差异较小。然而, 在不同的机构和协会中, 伦理准则方面的差异则更为显著。与检查和司法系统不同, 各种伦理委员会和理事会负责监督规则, 并可能施加特定职业的制裁 (例如正式警告, 甚至从专业协会中开除)。有时 (例如在健康研究中) 在调查开始前需要从伦理机构获得明确的预先批准。例如, 在美国, 大多数机构 (学术机构、公共机构, 有时还包括私人机构) 的研究都受到机构审查委员会 (IRBs) 的监管, 调查会接受审查以保护受访者的安全、权利和福祉。美国民意研究协会 (AAPOR) 在其网站上专门设立了一个版块, 致力于介绍 IRBs 和调查研究实践。

我们建议在开始网络调查项目之前熟悉相关的伦理准则, 以便识别并解决潜在的问题。然而, 关于应该参考哪一种伦理准则的建议并不明确。Groves 等人 (2009) 的研究为一般调查场景下的伦理考量提供了有益起点。

6.3.4 正式认证、准则和标准

前面的章节提到了在开展网络调查时需要考虑的一般法律和伦理问题。更具体的阐述通常可以在进行研究的各种准则和标准中找到。这些准则和标准是由专业研究协会在正式认证与标准体系框架下制定的。由于方法论、法律、伦理和专业等方面的内容通常在这些规范中相互交织, 我们在本节中将它们放在一起讨论。

在项目管理领域, 我们已提到一些具体标准 (例如 PMI-book、IPMA 和 PRINCE2), 以及其他质量管理方法和认证 (例如六西格玛全面质量管理, TQM)。

综合管理、行政和处理标准（如 ISO 9001）可从专业认证机构（如 ISO⁷ 或 BSI⁸）获得。正如在可用性、安全性或存档的相关讨论中提到的那样，一些更具体的规范可能与网络调查密切相关。同样，本书参考了两个针对调查的 ISO 标准：

- ISO 20252：2012《市场、民意和社会研究—词汇和服务要求》（ISO, 2012）。
- ISO 26362：2009《市场、民意和社会研究的访问样本库—词汇和服务要求》（ISO, 2009）。

获得并维持这些认证不仅能证明组织符合特定标准，还有助于研究机构获得大客户的认可。但相关认证通常需投入较高成本（例如 ISO 或 BSI）。

除上述商业化认证外，还有其他各种与网络调查直接或间接相关的规范、标准和指南，并且通常由专业协会和组织制定，可以免费获取。这些可能是专门处理网络调查的独立文件，或者是某些更广泛文件中的明确章节。与组织/协会的使命和特点（例如调查方法、市场营销或统计）相关的内容会被突出为重点。也可能有监督和/或申诉机构，以审查和澄清相关问题。

在与网络调查相关的具体国际规范方面，ESOMAR 是重要的资源来源。它提供了几个指南，包括“在线研究指南”“社交媒体研究指南”“移动市场研究指南”以及“帮助在线样本购买者的提问清单”。类似地，市场研究与情报协会也提供了“向您的在线调查供应商提出的十个问题”等指南（MRIA, 2013）。

在互联网调查研究的伦理规范方面，互联网研究协会（AoIR, 2012）对伦理决策与互联网研究的阐述较为充分。全球研究商业网络（GRBN）也关注国际层面的专业伦理，其研究对比了澳大利亚、加拿大、德国、英国及美国的专业行为准则，并设有专门章节探讨互联网研究⁹。

关于网络调查的具体规范，美国调查研究机构协会的《调查研究标准与伦理准则》（CASRO, 2011）以及德国市场与社会研究协会理事会（Rat der Deutschen Market - und Sozialforschung eV, 2007）的《在线调查指南》中均有专门讨论。

此外，美国民意研究协会（AAPOR）在标准与伦理领域也提供了丰富资源，包括制定处置准则（AAPOR, 2011）的详细阐述，以及每年对民意调查、调查方法和调查统计领域新书（仅限英语）（Callegaro, 2014a）的年度审查。

在通用统计领域，欧盟统计局的专业准则具有重要参考价值（Eurostat, 2011）。此外，各国协会及大型组织也制定了各类专业准则与标准，相关概述可在 WebSM 网站查询。

值得一提的是，万维网联盟（W3C）制定的网络通用技术标准（例如 HTML

或 CSS 代码) 尤为重要, 特别是要符合网络无障碍标准(例如美国《康复法案》第 508 条¹⁰)。

6.3.5 非正式指南、建议和其他资料

尽管上述提到了认证、准则和标准, 但在实践中, 关于如何运行网络调查的实用指南仍然相当缺乏。这一空白部分由各种非正式指南填补, 这类资料多以网络资源形式呈现。

网络调查软件供应商是这类资料的主要来源。他们深知, 销售软件及相关服务仅是业务的一部分; 从更广泛的方法论层面提供指南、提示、培训及最佳实践材料, 不仅是对特定软件功能的支持, 更是其使命的重要组成部分。这些资料形式多样, 包括简单的操作清单、系统的教程、专门的用户会议活动以及深入的方法论资料, 如白皮书甚至书籍 (Bhaskaran & LeClaire, 2010)。

根据 WebSM 的一项研究 (Vehovar, Čehovin et al., 2012), 在 377 个软件供应商中, 67% 提供了关于如何开展调查的基本方法文档, 30% 提供了具体的白皮书, 17% 包括广泛的文档, 9% 以正式的 PDF 文件形式提供了方法论文档。相关供应商的建议概览可见第 6 章的补充材料“网络调查软件方法建议清单” (*List of web survey software methodological recommendations*, <http://websm.org/ch6>)。

博客和其他定期在网上发布的帖子是另一个填补网络调查方法论最佳实践普及空白的渠道。这些文章中可能会有一些宣传性的内容, 但也能从中获得关于网络调查方法论现状的许多有价值的见解。WebSM 将相关博客分为以下四大类:

- 专家博客, 由调查方法专业人士和市场营销专业人士维护。
- 协会和媒体博客, 力求客观地讨论当代调查方法, 并且不直接参与销售调查软件或提供方法建议。
- 网络调查软件供应商的博客, 这是供应商的市场营销和销售策略的一部分, 但也提供相关信息, 因为它们必须吸引更广泛的受众才能取得成功。
- 来自私人、公共和非营利部门的研究组织的博客, 广泛涉及网络调查方法。

2014 年, WebSM 数据库包含了大约 100 个博客和其他在线资源, 这些资源或多或少定期解决与网络调查方法相关的问题¹¹。

6.4 网络调查参考书目

序言部分介绍了 WebSM 网站 <http://www.websm.org>, 并在各章节中都提到过该网站, 特别是关于补充资料、网络调查软件列表、博客等信息。然而, WebSM 的核心任务是其建立于 1998 年的文献数据库。文献的选择和收录过程是常规化的, 依赖于对大约 100 个来源的定期回顾, 也通过明确策略处理搜索引擎捕捉到的新条目。

在 WebSM 中, 文献条目根据三个标准收录: (a) 文献来源可公开获取; (b) 至少有摘要; (c) 条目基于与网络调查方法相关的研究。由于方法论研究重点对于建立这种相关性至关重要, 因此各种描述性研究项目报告或市场研究会议论文不在数据库的范围内。另一方面, 部分间接关联网络调查方法的一般性文献被收录, 涵盖了一般社会科学方法论、一般调查方法论和计算机辅助调查信息收集 (CASIC)、可用性研究、人机交互 (HCI)、在线定性方法、互联网研究、在线社区研究、电子学习, 以及不同实质性领域使用网络调查来收集数据的重要案例研究。这种间接相关性的标准更具主观性, 但我们相信这种方法多年来一直保持一致。

文献目录是根据学术性条目 (期刊论文、书籍、编著、书籍章节) 和其他类型 (杂志论文、通讯、会议论文集、会议资料、学位论文、报告、商业资料等) 来构建的。每条文献条目都用标准代码描述, 包括出版类型、标题、作者、来源、年份、摘要和全文源数据库, 同时还用特定代码标注了学科领域、主题、国家、目标人群和相关主题。

我们在这里提供一些关于 WebSM 文献数据库结构的见解, 涵盖截至 2013 年的出版物。在 6 890 个条目中, 学术条目占 2 888 个, 其余的主要属于各种会议材料、会议记录和其他文件。

在 WebSM 中数据库中, 每本期刊收录条目超过 10 条的 10 种期刊按条目数量降序排列如下: *Quirk's Marketing Research Review*、*Public Opinion Quarterly*、*Social Science Computer Review*、*Journal of Official Statistics*、*International Journal of Market Research*、*Survey Practice*、*Behavior Research Methods*、*Computers in Human Behavior*、*Journal of Medical Internet Research* 和 *Marketing News*。这些期刊基本上反映了第 6.3.1 节讨论的相关科学领域的结构。大部分条目来自社会科学领域, 其次是市场营销、健康、互联网研究、心理学和教育。这些领域与一般社会科学方法论、计算机科学和统计学有很强的重叠。当然, 这种结构与实际网络调查项目的结构, 以及与日常完成的网络问卷的结构有很大的不同。

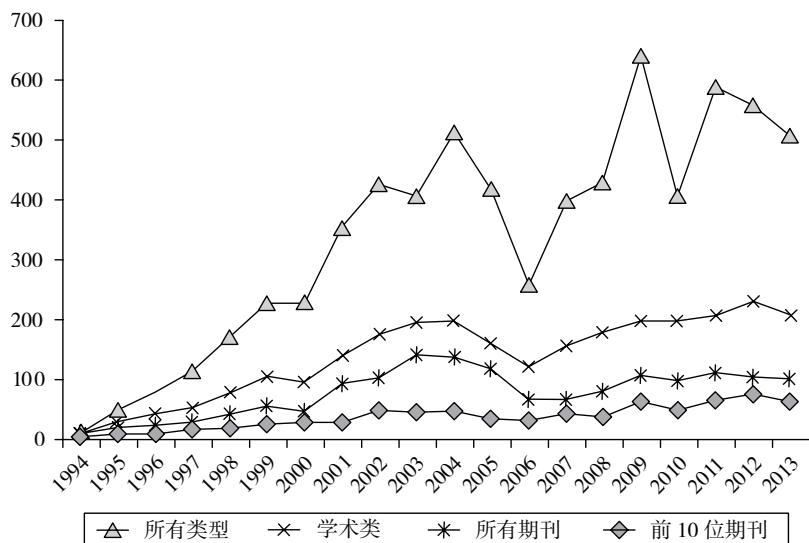


图 6.2 按出版年份（1994—2013）和出版类型分类的 WebSM 数据库中的文献条目

从地区分布看，与美国相关的研究条目占比最高（53%）；就会议而言，美国民意研究协会（AAPOR）的条目数量最多，其次是欧洲调查研究协会（ESRA）和通用在线研究（GOR）。

如图 6.2 所示，自 1994 年首篇直接聚焦网络调查方法论的学术文献（Pitkow & Recker, 1994）发表以来，近二十年间新条目数量稳步增长。2004 年之后，年度新增条目数量的增长速度放缓至每年约 500 条，其中 150~200 条为学术条目。2006 年出现了不明原因的下降趋势。值得注意的是，针对中国网络调查研究相关出版物的元研究也发现了类似趋势（Zhang, Shao, & Fang, 2008）。

截至 2013 年底，数据库共收录 8 323 位作者，其中 4 741 位贡献了学术性内容。在这两类作者中，大多数作者（分别为 76% 和 84%）只贡献了一个条目。在这一领域发表文章的门槛相对较低，但这更可能是由于作者的主要研究重点在网络调查方法之外。

在图 6.3 中，我们使用 Pajek 网络分析工具（de Nooy, Mrvar, & Batagelj, 2011）来展示网络调查方法学领域的作者合作网络。为此，我们在 2014 年 11 月对 WebSM 文献数据库进行了额外的分析，其中只包括与网络调查方法直接相关的学术出版物。在图 6.3 中，只有 72 位作者发表了 5 篇或更多的学术文献条目，他们总共产生了 528 条被纳入此次分析的条目。每位作者用一个圆圈表示，其大小反

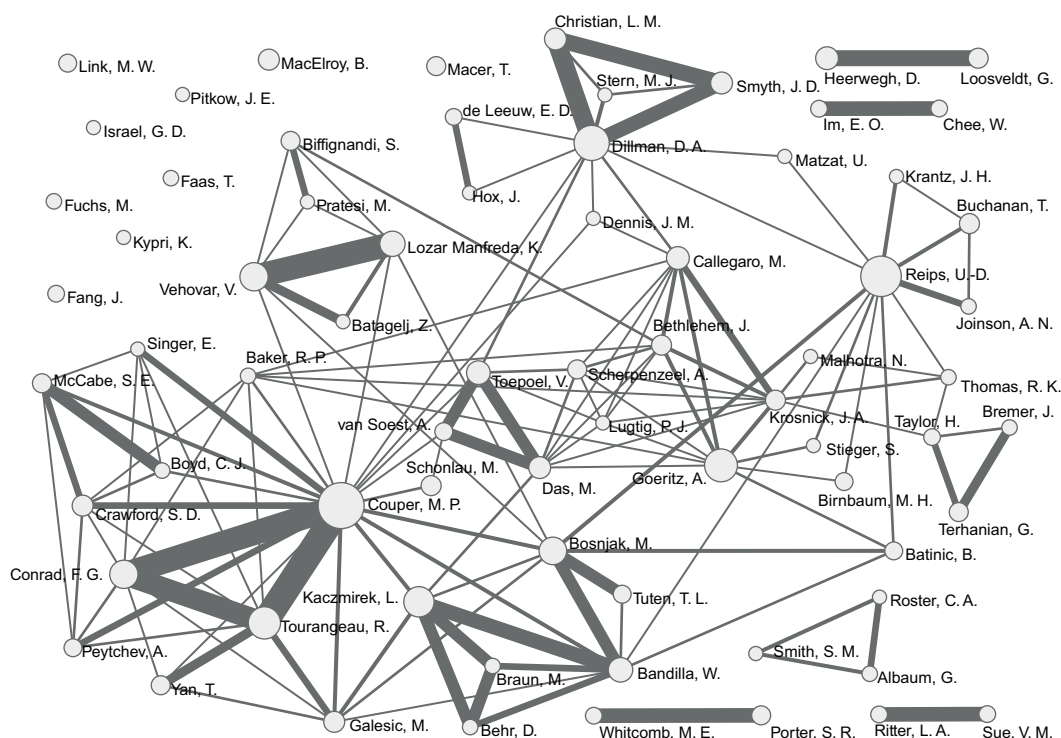


图 6.3 2014 年 11 月，WebSM 中拥有五篇或更多与网络调查方法直接相关学术条目的作者之间的合作图谱

映了他们在 WebSM 数据库中的条目数量，例如 Couper 有 53 条，Reips 有 42 条，Dillman 有 31 条。总共有 25 位作者拥有超过 10 条条目。连接作者的线条的粗细反映了他们合作撰写的学术条目数量，例如 Couper 和 Tourangeau 合作撰写了 14 条。

需说明的是，图中仅包含与网络调查方法直接相关的文献条目，未纳入作者在其他学科领域的成果。

有趣的是，72 位作者中，有 52 位构成了网络中最大的连通分量（即最大的作者合作子网络）。由此可见，这一合作网络规模相对较小。作为本书作者的我们几乎认识所有这些研究人员，且观察到网络调查方法仅是少数人的主要研究方向，多数人同时涉足社会科学方法论、一般调查方法论及各实体领域（如社会学、心理学）。

上图中包含的所有文献条目列表，以及关于文献条目的进一步统计数据（研究主题、语言、来源等），作为补充资料，可以在 WebSM 网站上找到（*Statistics of WebSM bibliography database*, <http://websm.org/ch6>）

综上，与网络调查方法论相关的专业活动已在一定程度上趋于稳定。在未来，网络调查有望将更加深入地融入一般社会科学方法论的背景中，同时也将更好地整合至一般的管理框架。网络调查方法论与一般调查方法论之间的界限将变得越来越模糊，我们将在第 7 章关于未来趋势的部分中进一步讨论这一问题。

注 释

- 1 See pages 20–21 in Institute of Public Health presentation on data aspects, available at <http://www.websm.org/nijz13>
- 2 <http://www.pmi.org/PMBOK-Guide-and-Standards.aspx>
- 3 <http://ipma.ch/resources/ipma-publications/ipma-competence-baseline/>
- 4 <http://www.prince2.com/prince2-methodology>
- 5 <http://workshop.websm.org>
- 6 <http://kostat.go.k/wis/inde.html>
- 7 ISO standards: <http://www.is.orgliso/home/standards.htm>
- 8 BSI certification: <http://www.bsigroup.com>
- 9 <http://www.websm.org/grbn12>
- 10 <http://www.websm.org508>
- 11 <http://www.websm.orgblogs>

网络调查的未来

毫无疑问，网络调查将持续扩展，而且其方法体系中的某些方面也将日益适用于各类调查场景。除了这一普遍观察结论，下文将进一步探讨更具体的发展论断。部分论断已在前面章节提及，此处仅作归纳。同时结合了对行业、专业领域及科学层面的观察与推测，并融入了我们在网络调查实践、相关活动及专业网络中的经验思考。

未来趋势可以大致分为四个方向，且各个方向之间可能会部分重叠：技术发展（第 7.1 节）、网络调查软件（第 7.2 节）、方法论（第 7.3 节）以及与专业领域、科学研究、商业和社会相关的更广泛的问题（第 7.4 节）。

7.1 一般技术发展

信息通信技术（ICT）的整体发展趋势将持续催生成本更低廉、使用更便捷、性能更优越的设备与服务，例如从云计算到三维图像（全息图）。此外，我们预计未来的网络调查将特别受益于以下几类技术突破。

语音输入。尽管过去数十年语音识别技术已取得快速发展，但我们仍在期待其重大突破，实现低成本的标准化产品，至少能覆盖主要官方语言。在众多应用场景中，技术的进一步发展将让受访者可通过语音完成网络调查，以更贴合用户习惯的交互方式与设备对话。一旦该技术全面落地，部分网络调查可能会转向这种替代性

作答方式，且这一趋势在智能手机和平板电脑调查中尤为适用。

智能数据分析。这是我们期待了数十年却屡次推迟实现的技术突破。将人工智能应用于统计分析中——涵盖从处理异常到创建适配图表的全流程——对于网络调查来说尤其具有吸引力，因为它可以顺利地整合到网络调查过程中，以支持实时数据质量诊断、数据编辑和插补，调查干预、初步报告撰写等。这有望促进信息与通信技术（ICT）支持的整合，同时也能满足当前对用于数据质量诊断、迭代调查设计、数据清理、初步数据分析以及自动准备高级传播材料的独立软件的部分需求。精心设计的仪表盘也能借助这些技术进展，使用户可对收集的数据进行实时深度分析，并与团队其他成员共享结果。

在线协作工具。我们预计在开发功能强大且用户友好型平台方面将取得进展。这些平台将被集成到网络调查软件中，支持研究全流程的协作与沟通环节。这将为当前依赖电子邮件和文本处理器的做法提供更有效的替代方案，尤其是复杂调查项目，因为这类项目需要协调众多任务和相关人员。

多种设备。不难预测，智能手机、平板电脑、网络电视等已经替代台式机/笔记本电脑，且这一趋势将进一步延伸，设备种类会更丰富，功能也将持续更新。在完成网络调查问卷的设备中，标准计算机（台式电脑和笔记本电脑）可能很快就会只占很小的份额。硬件发展和软件标准化（如平台、浏览器、HTML5）将对网络调查产生重要影响，并将决定网络调查软件和方法学方面的众多变革与调整。

身体语言。传统调查中，受访者主要通过口头表述或手动输入提供答案。这方面可以朝着游戏主机（例如 Xbox Kinect、PlayStation Move 或 Wii MotionPlus）已经探索的方向进行拓展，其中反馈也可以通过特定的身体动作来传递。这不仅涉及手的动作（例如表示“是”或“否”的手势），还包括身体任何部位的任何动作，甚至包括各种眼神交流。

电子小额支付的发展。建立便于小额款项轻松支付的通用标准，尤其是使其在其他场景中得到更广泛的应用，可能会进一步推动网络调查中更高效激励策略的发展。这可能让受访者只需一键操作即可获得奖励。

发展中国家直接向智能手机网络调查转变。许多发展中国家跳过了固定电话技术，直接进入移动电话时代。我们可以预期在互联网方面也会出现类似的趋势：基于家庭互联网连接的台式机/笔记本电脑可能会被跳过，移动设备将更受青睐。这一转变的实现取决于现有技术基础，在智能手机市场占有率较高的国家，移动设备有望成为主要的调查工具。

7.2 网络调查软件

调查软件对网络调查至关重要，其范畴直接界定了核心网络调查过程（第1.2节图1.1）。虽然许多信息与通信技术（ICT）服务（如电子邮件、固定电话服务等）已基本成熟，但网络调查软件的潜力尚未完全释放，仍有广阔的发展空间。语言技术、文本/数据挖掘、人工智能以及其他进步将使我们能够充分利用现有的知识和数据库，以及调查数据、辅助数据和元数据。这些数据将共同支持网络调查过程中每一个步骤的整合与优化。具体而言，在核心网络调查过程的三个步骤中，预计网络调查软件会有以下进展：

1. **预调查阶段**：技术将整合、自动化并促进整个问卷设计过程。智能交互式帮助将提供建议，并引导用户完成从概念化和操作化到问题措辞（借助丰富的题库提供支持）、计算机化和测试的整个过程。测试结果将通过各种评估方法（例如测试用户的鼠标移动）自动处理和提取。此外，文本处理器将集成到网络调查软件中，并提供对评论管理的高级支持。这一功能将使网络调查问卷的开发在其天然场景——网络环境中得到充分支撑，摆脱对普通文本处理器的依赖。在普通的文本处理器中，仅能通过追踪修订（tracking changes）、用不同颜色编辑文字以及在文本中添加评论来管理反馈信息，而所有这些操作之后还需要通过电子邮件或其他互联网服务进行信息交换。
2. **正式调查阶段**：自动且实时地检测低质量的回答以及其他潜在问题，同时为受访者提供定制化的交互式帮助，这些举措将带来诸多益处，并且在很大程度上能够取代访员所承担的解释说明和激励受访者的角色。
3. **调查后阶段**：目前通常在系统外完成的数据准备工作（编辑、编码、插补和加权）将逐步实现自动化并整合至系统中，报告撰写与数据导出环节亦是如此。这种整合将进一步拓展核心网络调查过程的范畴，使其延伸至高级分析和数据应用领域。

除了网络调查软件技术的进步，我们还预计供应商之间将通过并购实现越来越多的横向整合。网络调查软件的发展及其整合涉及大量开发者与用户，其所需资源的最小规模与最优规模将持续扩大。目前尚不确定各类地方性、全国性及小规模调查软件的存续周期，以服务特定的用户群体，但集中化绝对是一个趋势。具有竞争力的价格进一步推动了这一趋势，因为要确保用于营销、技术支持和开发的资源，就需要越来越多的客户。

随着在线样本库和各种业务流程（例如调研客户的心声）的整合，网络调查插件与主流内容管理系统（如 WordPress）及社交平台（如 Facebook）的融合，以及微软、谷歌、苹果等大型信息通信技术企业办公服务支持范围的拓展，纵向集中化的重要性愈发凸显。

7.3 方法学

除了一般性和特定的技术发展趋势之外，我们还预计在调查方法学的发展方面会有重大改进。以下讨论将纳入那些对方法学有直接影响的技术进步。

混合模式数据收集的兴衰演变。我们认为，在市场与商业领域，调查模式的混合应用预计不会增多。因为对于他们的特定需求而言，混合模式既成本高昂又操作复杂。然而，当网络调查被引入官方统计和学术概率调查时，混合模式在过渡阶段会有暂时的增加。最终，几乎所有的电话调查和面对面调查至少会部分迁移到网络上。在第一步中，这通常意味着转变为混合模式调查。由于这种转变的复杂性，这将是网络调查历史上最有趣和最具挑战性的方法学实践之一。

在这一初步转变后，约十年至二十年间，混合模式调查的总体数量预计将下降。当“数字原住民”（即出生于数字技术普及的环境中的人）数量增加、大多数国家的互联网普及率达到 95% 以上以及计算机素养水平提高时，技术和方法学的进步、软件的整合以及进一步的社会变革将推动这一趋势。精心设计且量身定制的网络调查将不需要访员的介入，仅在极少数极为特殊的调查中，访员才是不可或缺的。

并行数据使用的增加。鉴于网络调查中无访员参与的特点，并行数据也将越来越多地被用作一种工具，以便更好地理解问题回答过程，并阐明网络调查数据质量问题。

加强与定性研究的整合。与混合模式不同，网络调查与各种定性研究方法的整合有望稳步增长。所谓的混合研究方法，即将调查数据收集与定性数据收集相结合的方法，目前在概念、方法及技术方面（即集成软件支持）来看，其探索仍然相对薄弱。一旦这两种方法在网络上得到充分整合并获得支持，它们之间的互补性和协同效应将带来新的可能性。

技术增强的补充数据收集。我们已列举过智能手机的相关案例，这类设备可以半自动（例如条形码扫描）或自动收集特定数据（例如位置、运动、距离等）。同样，受访者使用一些可穿戴技术后，他们接触广播和电视信号以及户外广告牌广

告的数据也可以被自动记录,因此在调查中不需要询问受访者关于广告的信息。在生物标志物数据收集以及通过可穿戴技术收集的其他数据方面,预计也会取得显著进展。

技术增强的调查问卷。从问题的多媒体呈现、问卷视觉布局(界面设计),到游戏化元素与高级动画(如智能虚拟面试官)的引入,调查问卷的呈现形式与交互体验将愈发丰富。此外,“物联网”的兴起将带来实质性变革:物体(设备、电器、消费品等)将被赋予IP地址并接入互联网。例如,冰箱可能会定期(例如一年一次)用自身语音或触摸屏功能邀请用户参与调查。或者,调查邀请可以转移到同一消费者名下其他更便捷的调查设备上(例如电视、智能手机等)。

调查数据与大数据的融合拓展。调查数据正日益实现相互融合(如统计匹配),并与行政数据、商业数据,以及与交易和在线行为相关的众多电子痕迹数据建立关联。现代科技进一步使获取、处理和分析海量数据(即大数据)成为可能。这些大数据可能在一定程度上取代某些类型的调查(例如关于受访者的媒体消费、零售购买、活动和旅行的调查)。但除少数例外,这并不会削弱对调查的整体需求,特别是在许多情况下,直接向受访者提问往往比从大量非结构化、非针对性收集的数据中提取信息更简便、高效。另一方面,这无疑将在对各种数据源的综合利用方面提高协同效应。

客户样本库和在线社区的兴起。除普通大众在线样本库的规模不断扩大之外,越来越多的机构正基于自己的目标群体(客户、选民、会员、顾客等)建立专属在线样本库。此外,他们还建立各种类型的在线社区(例如通过社交媒体粉丝群建立的社区)。在这些社区中,除了开展网络调查,他们还开展定性研究、众包和数据挖掘等活动。

社交媒体招募、实时招募和动态样本。一些公司已经将所有互联网用户视为潜在的调查受访者群体,并综合运用不同的招募技术,以便实时创建在线样本并进行调查。由于传统调查以及在线样本库都存在应答率方面的问题,采用创新方法与激励策略即时招募受访者的需求将愈发迫切。

基于非概率样本进行推断。非概率样本已经在网络调查中占据主导地位,因此该领域有望取得重要进展:将建立更详细的非概率调查规则和标准,并形成对其潜力和局限性的共识,尤其是明确这种方法适用的条件。相应地,当前替代科学统计推论的相关技能,也有望得到规范化并受到科学关注。因此,我们预计将出现涵盖概念与实践的相关教科书。

7.4 更广泛的商业和社会问题

在很大程度上，更广泛的商业和社会背景也决定了网络调查的作用及其应用情况。

与行政和商业流程的进一步整合。各类行政与商业流程将更深度地融入网络调查的数据收集环节，涵盖从基于移动应用程序的情报报告（例如，从条形码扫描到照片和视频收集）、客户关系管理（CRM）数据库、对销售部门的支持，到人力资源管理（HRM）系统及其他各种评估和管理应用程序等全部流程。

自助式（DIY）研究的拓展。随着网络调查软件为调查全过程提供越来越智能的支持，具备基础计算机操作能力与常识的人将被进一步赋能，不仅可以开展简单的调查，还可以处理许多更高级的方法学问题。随着这一发展，专业网络调查和自助式网络调查之间的界限将变得更加模糊。自助式研究的进一步扩展也将是网络调查增长的主要因素，因为越来越多的新用户将有能力实施更复杂的调查。

研究的民主化。自助式研究的扩展已经推动研究走向民主化。过去，调查数据收集需要大量的资源（时间、人力和资金方面），只有专业的调查机构才能负担得起。今天，任何人——即使没有受访者资源、资金或专业知识——都可以进行网络调查，这使得研究领域越来越向小型机构、非专业人士和整个公民社会开放，他们可能会提出官方、学术或市场研究机构永远不会提上议程的研究课题。我们认为这一趋势不会停止：越来越多的主体（个人、组织）将能够开展越来越复杂的网络调查。

调查参与度下降。我们在第2.5节中提到了导致调查合作度持续下降的各种原因。我们预计这些趋势在未来会在一定程度上加剧，直至过度调查导致网络调查达到严重饱和的状态。在这一背景下，由于隐私和保密问题，以及未经请求的通信越来越被受访者反感，尤其需要指出网络调查的社会限制将日益增加。因此，受访者知情同意这一点将愈发重要。方法学研究者将需要主要通过精心设计的激励措施、优化的迭代调查设计，甚至更友好的网络问卷来应对这些问题，而这些问卷往往包含调查娱乐化和游戏化的元素。

网络调查专业的兴起与衰落。当前，网络调查方法学领域仍有诸多问题待解。对于一些非常基本的问题，我们尚未形成最终结论。比如单页还是多页设计、硬提示还是软提示、处理“不知道”答案的方式，甚至是日历日期输入问题的最佳格式，等等。未来，新设备的出现及传统调查向互联网迁移过程中的技术动荡，将会带来更多的方法学挑战。因此，我们可以预期，对专业支持的需求将会增加，相应的方

法学研究也会增多。

然而，当几乎所有调查都完全纳入网络调查范畴后，网络调查方法学将逐渐成为调查方法学不可分割的一部分。同样，我们不再单独且长篇大论地讨论信息社会的概念，因为信息通信技术已经渗透到现代社会的各个角落，将不再存在信息社会之外的现代社会。

除了在形式上和术语上的淡化，从长远来看（例如几十年以后），网络调查方法的实质性地位也可能下降。其中一个原因是，在过去二十年中出现的许多方法动荡往往是暂时的，因为网络调查技术（如设备、网络调查软件、互联网速度）尚未变得稳定和成熟，处于持续的技术改进之中。一旦技术平台及其相关方法论基本稳定下来，就像固定电话访谈调查一样，所有这些动荡都将慢慢消失。因此，我们也可以预期，调查方法学专业领域将出现一定程度的衰退，因为大多数调查过程将实现标准化与智能化自动操作。当然，这与过去的趋势完全相反。在过去，对方法学家处理问卷设计、无应答、非概率样本选择和调查后调整的需求持续增加。

调查行业会衰落吗？我们认为未来调查的总量不会减少。在此背景下，回顾 20 世纪 90 年代中期的猜测颇有意义。当时，随着计算机辅助调查信息收集（CASIC）的初步增长，有人猜测，由于替代方法（如自动化数据、模型）的发展，调查行业在 2020 年后将会衰退。尽管自动化数据收集和建模实际上正在迅速扩展，但我们今天看到，调查数量并没有随之而下降。相反，调查数量正快速增加，这一趋势可以从供应商、研究者以及受访者那里得到证实。我们完全可以预期，某些自动数据收集和处理将取代对调查的需求，因此相关服务（例如商业智能）实际上将与调查行业所提供的服务，尤其是市场研究领域提供的服务，展开日益激烈的竞争。尽管如此，由于自助式调查的扩展及信息需求的增长，这种替代无法抵消网络调查的增长势头。尤其是在商业领域，以数据驱动的方式来做出决策的概念正在迅速普及。正因如此，我们看到越来越多的决策者委托进行调查，将其作为支持商业决策的另一个数据来源。

每日完成的网络调查问卷数量可能稳步增加。但如前所述，网络调查专业领域的发展可能不会紧随这一趋势。它将发生诸多变化，主要朝着与管理学、统计学、信息学和特定领域信息加强整合的方向发展，这可能会进一步减少对专业（网络）调查方法学家的需求。

最后，以一个极具前瞻性的推测作结：随着神经科学领域出现某些根本性的突破，如能直接检测调查问题的答案，我们如今所熟知的调查行业或许会真得走向衰

落。这实际上可能会促使实现调查问卷自动填写，并催生一种全新的数据收集方式（例如，直接通过脑电波扫描获取应答，而无需受访者进行任何主动配合）。但需再次强调，这或许不会缩减传统调查行业的规模，反而可能进一步增加问卷完成量。

结 论

网络调查作为一种相对新兴的数据收集方法，其方法论研究已取得诸多进展，本书尝试对这些成果进行批判性梳理与呈现。

以下简要总结网络调查为调查方法学领域带来的突破。首先，它们重新引发了传统邮寄调查文献中讨论的诸多议题。也就是说，我们此处讨论的是一种自我管理的调查模式，它既有优点也有缺点。网络调查另一个重要且独特的方面在于，受访者需要与设备及互联网进行交互，才能完成问卷作答。

网络调查也凸显了视觉呈现的重要性。实践表明，即便对调查问题的布局稍作调整，也可能严重改变回答的分布情况。与此相关的是，网络调查具有诸多与日常网络体验相似的共性。换言之，诸如单选按钮、复选框、输入框等元素，不仅出现在网络调查中，还出现在许多其他情境中。

此外，网络调查也凸显了抽样和受访者招募方面的挑战。互联网为快速招募低成本的受访者群体提供了可能，却也导致了向非概率抽样的转向。尽管其他数据收集方法也采用过非概率抽样，但在线招募开辟了许多新的可能性，同时也带来了挑战。

自我管理、视觉设计和抽样是网络调查方法学的三个核心要素，并贯穿于本书的各个章节。本书的整体架构则是根据网络调查过程（预调查阶段、正式调查阶段、调查后阶段）进行组织。在论述过程中，我们提供了超过 700 条关键参考文献，并从相关的网络调查方法学（WebSM）书目数据库中精心挑选了其中最优秀的文献。

截至 2014 年初, 该数据库包含 6 890 个条目。

经过多年对网络调查的研究, 我们对不同的抽样和在线招募机制、视觉设计的效果、网络问卷中的交互性和图像使用、激励措施、多次联系、其他提高应答率的方法, 以及并行数据等方面有了深入理解。在本书讨论中, 我们多次提到“需要开展更多研究”这一普遍观点。在这个快速发展的领域, 技术和方法日新月异, 潜在研究思路也层出不穷。然而, 我们还是要再次强调需进一步深入研究的关键特定主题:

- 深入研究针对无效应答或无应答的硬提示, 它们与非实质性应答选项(例如“不知道”)使用之间的相互关系, 以及它们对总体数据质量的影响;
- 明确使用表格而非系列单个问题的适用场景与标准;
- 探究交互式与视觉元素的潜在作用;
- 厘清多设备开展网络调查的具体实施路径;
- 建立非概率网络调查的标准体系(涵盖指标、报告、文档记录等);
- 完善混合模式系统。

尽管网络调查的研究最初是基于调查方法学的一般知识, 但事实证明, 它比与传统调查模式相关的研究更具跨学科性。仅仅通过浏览网络调查方法学(WebSM), 研究者便能清晰感受到这种多样性。网络调查的实施涵盖了社会科学、市场营销、健康、教育和互联网研究等领域, 相关研究成果涉及社会科学方法论、统计学、计算机科学、可用性研究和网页设计等多个方面。

网络调查的未来发展方向如何? 结合第 7 章对潜在趋势的分析, 我们确信网络调查的使用将持续稳步增长。随着许多国家互联网普及率的提高, 将会有更多的调查在网络上进行, 尤其是近年来全球向移动设备转变的趋势下。另一方面, 对于网络调查使用量的增加, 也存在一些替代方式。自动数据收集、统计建模和大数据技术, 特别是在市场研究领域, 为商业智能行业的竞争者创造了发展空间。在这种背景下, 网络调查方法学的核心价值在于助力产出具有竞争力的成果, 即以低成本收集高质量的调查数据。

关于最具前景的未来发展领域, 我们认为应重点关注信息与通信技术(ICT)对整个网络调查过程在概念、方法和技术方面的进一步整合。同样, 我们预计知识技术将越来越多地应用于问卷设计过程, 以及应答质量的诊断过程中。这些进展将显著提升网络调查的整体质量。

最后需要说明的是，本书的内容并非终点。网络调查方法学（WebSM）网站（<http://www.websm.org>）将持续更新并补充新内容。读者可在该网站获取各类补充材料与相关信息。读者反馈可发送至 book@websm.org，以便我们优化补充材料，助力未来推出更优质的版本。

参考书目

- AAPOR. (2010). *AAPOR code for professional ethics and practices*. Retrieved from <http://www.aapor.org/>.
- AAPOR. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. Retrieved from <http://www.aapor.org/>.
- Adams, M. J. D., & Umbach, P. D. (2012). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education*, 53(5), 576–591.
- Advertising Research Foundation. (2014). *Foundations of quality 2.0: Launching what we've learned*. Retrieved from Presented at ARF's Symposium on FOQ2 Initiative. Retrieved from <http://www.thearf.org/>.
- Ainsaar, M., Lilleoja, L., Lumiste, K., & Roots, A. (2013). *ESS mixed mode experiment results in Estonia (CAWI and CAPI mode sequential design)*. Report. Retrieved from http://www.yti.ut.ee/sites/default/files/ssi/ess_dace_mixed_mode_ee_report.pdf
- Ajzen, I. (1991). A theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Albaum, G., Roster, C. A., Wiley, J. B., Rossiter, J., & Smith, S. M. (2010). Designing web surveys in marketing research: Does use of forced answering affect completion rates? *Journal of Marketing Theory & Practice*, 18(3), 285–294.
- Albaum, G., & Smith, S. M. (2012). Why people agree to participate in surveys. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 179–193). New York: Springer.
- Albert, W., Tedesco, D., & Tullis, T. (2009). *Beyond the usability lab: Conducting large-scale online user experience studies*. Amsterdam: Morgan Kaufmann.
- Alexander, G. L., Divine, G. W., Couper, M. P., McClure, J. B., Stopponi, M. A., Fortman, K. K., ... Johnson, C. C. (2008). Effect of incentives and mailing features on online health program enrollment. *American Journal of Preventive Medicine*, 34(5), 382–388.
- Alwin, D. F. (2007). *Margins of error: a study of reliability in survey measurement*. Hoboken, New Jersey: Wiley.
- Amarov, B., & Rendtel, U. (2013). The recruitment of the access panel of German Official Statistics from a large survey in 2006: Empirical results and methodological aspects. *Survey Research Methods*, 7(2), 103–114.
- Andrews, D., Nonnecke, B., & Preece, J. (2003). Conducting research on the Internet: Online survey design, development and implementation guidelines. *International Journal of Human-Computer Interaction*, 16(2), 185–210.
- Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology*, 25(3), 335–349.
- Antoun, C., Zhang, C., Conrad, F. G., & Schober, M. F. (2013). *Comparisons of online recruitment strategies: Craigslist, Facebook, Google Ads and Amazon's Mechanical Turk*. Presented at the The American Association for Public Opinion Research (AAPOR) 68th Annual Conference, Boston, Massachusetts.

- AoIR. (2012). *Ethical decision-making and Internet research*. Retrieved from <http://aoir.org/>.
- Aoki, K., & Elsmar, M. (2000). Opportunities and challenges of conducting web surveys. Results from a field experiment. *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 930-935. (pp. 930-935). Alexandria, Virginia: AMSTAT.
- Arbeitskreis Deutscher Markt. (2007). *Guideline for online surveys*. Retrieved from http://www.adm-ev.de/fileadmin/user_upload/PDFS/R08_E_07_08.pdf
- Aviram, H. (2012). What would you do? Conducting web-based factorial vignette surveys. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences*. pp. 463-473. Berlin: Springer.
- Babbie, E. R. (2013). *The practice of social research* (13th ed.). Bedford, CA: Wadsworth.
- Baker, J. D. (2013). Online survey software. In *Online instruments, data collection, and electronic measurements: organizational advancements*. (pp. 328-334). Hershey, PA: IGI Global.
- Baker, R. P., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., ... Zahs, D. (2010). Research synthesis. AAPOR report on online panels. *Public Opinion Quarterly*, 74(4), 711-781.
- Baker, R. P., Brick, M. J., Bates, N., Battaglia, M. P., Couper, M. P., Dever, J. A., ... Tourangeau, R. (2013). Report of the AAPOR task-force on non-probability sampling. Retrieved from <http://www.aapor.org/>.
- Baker, R. P., & Couper, M. P. (2007). The impact of screen size, background color, and navigation button placement on response in web surveys. Presented at the General Online Research Conference (GOR), March, University of Leipzig, Germany.
- Baker, R. P., Crawford, S. D., & Swinehart, J. (2004). Development and testing of web questionnaires. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, J. Martin, & E. Singer (Eds), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 361-384). Hoboken, New Jersey: Wiley.
- Baker-Prewitt, J., & Miller, J. (2009). *Beyond 'trapping' the undesirable panelist: The use of red herrings to reduce satisficing*. Presented at 2009 CASRO Panel Quality Conference, New Orleans, Louisiana. Retrieved from <http://www.casro.org/>. http://www.burke.com/Library/Conference/Beyond%20Trapping%20the%20Undesirable%20Panelist_FINAL.pdf
- Baltar, F., & Brunet, I. (2012). Social research 2.0: Virtual snowball sampling method using Facebook. *Internet Research*, 22(1), 57-74.
- Balter, O., Fondell, E., & Balter, K. (2012). Feedback in web-based questionnaires as incentive to increase compliance in studies on lifestyle factors. *Public Health Nutrition*, 15(6), 982-988.
- Bandilla, W., Couper, M. P., & Kaczmirek, L. (2012). The mode of invitation for web surveys. *Survey Practice*, 5(3).
- Barchard, K. A., & Williams, J. (2008). Practical advice for conducting ethical online experiments and questionnaires for United States psychologists. *Behavior Research Methods*, 40(4), 1111-1128.
- Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, 53(2), 182-200.
- Barrios, M., Villarroya, A., Borrego, Á., & Ollé, C. (2011). Response rates and data quality in web and mail surveys administered to PhD holders. *Social Science Computer Review*, 29(2), 208-220.
- Barron, G., & Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, 18(5), 507-520.
- Bartram, D. (2006). Testing on the Internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram & R. K. Hambleton (Eds), *Computer-Based Testing and the Internet: Issues and Advances* (pp. 13-38). Chichester: Wiley.
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139-1160.
- Basso, L., & Rathod, S. (2004). Response and field period effects: *The effect of time in online market*

- research and consequences for future online survey strategies. Survey Sampling International White Paper. Retrieved from <https://www.surveysampling.com/>.
- Batagelj, Z., & Vehovar, V. (1998). WWW surveys. *Advances in Methodology and Statistics (Metodološki Zvezki)*, 14, 209–224.
- Batanić, B., & Moser, K. (2005). Determinanten der rücklaufquote in online panels. *Zeitschrift Für Medienpsychologie*, 17(2), 64–74.
- Battaglia, M. P., Hoaglin, D. C., & Frankel, M. R. (2009). Practical considerations in raking survey data. *Survey Practice*, 5(2).
- Bauman, S., Jobity, N., Airey, J., & Hakan, A. (2000). *Invites, intros and incentives: Lessons from a web survey*. Presented at the The American Association for Public Opinion Research (AAPOR) 55th Annual Conference, Portland, Oregon.
- Bavdaž, M. (2010). Sources of measurement errors in business surveys. *Journal of Official Statistics*, 26(1), 25–42.
- Bearden, W. O., Netemeyer, R. G., & Haws, K. L. (2011). *Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research* (3rd ed.). Los Angeles, CA: Sage.
- Beaty, J. C., Dawson, C. R., Fallaw, S. S., & Kantrowitz, T. M. (2009). Recovering the scientist-practitioner model: How IOs should respond to unproctored internet testing. *Industrial and Organizational Psychology*, 2(1), 58–63.
- Beckenbach, A. (1995). Computer-assisted questioning: The new survey methods in the perception of the respondents. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 48(1), 82–100.
- Beckers, T., Siegers, P., & Kuntz, A. (2011). *Speeders in online value research. Presented at the General Online Research Conference (GOR)*, Düsseldorf, Germany.
- Behr, D., Bandilla, W., Kaczmirek, L., & Braun, M. (2013). Cognitive probes in web surveys: On the effect of different text box size and probing exposure on response quality. *Social Science Computer Review*.
- Bell, D. S., Mangione, C. M., & Khan, C. E. (2001). Randomized testing of alternative survey formats using anonymous volunteers on the world wide web. *Journal of the American Medical Informatics Association*, 8(6), 616–20.
- Bellman, S., Johnson, E. J., Kobrin, S. J., & Lohse, G. L. (2004). International differences in information privacy concerns: A global survey of consumers. *The Information Society*, 20(5), 313–324.
- Bennink, M., Moors, G., & Gelissen, J. (2013). Exploring response differences between face-to-face and web surveys: A qualitative comparative analysis of the Dutch European Values Survey 2008. *Field Methods*, 25(4), 319–338.
- Berzelak, N. (2014). *Mode effects in web surveys*. Ljubljana: PhD thesis. Retrieved from <http://www.fdv.uni-lj.si/>.
- Berzelak, N., & Lozar Manfreda, K. (2009). *Relations between usability and functionality of web survey software tool: An empirical evaluation*. Presented at the 7th Internet Survey Methodology Workshop, Bergamo, Italy.
- Berzofsky, M., Williams, R., & Biemer, P. (2009). Combining probability and non-probability sampling methods: Model-aided sampling and the O*NET Data Collection Program. *Survey Practice*, 2(6).
- Bethlehem, J. (2009a). *Applied survey methods: A statistical perspective*. Hoboken, New Jersey: Wiley.
- Bethlehem, J. (2009b). Can we make official statistics with self-selection web surveys? *Proceedings of Statistics Canada Symposium 2008*. Canada: Statistics Canada.
- Bethlehem, J., & Biffignandi, S. (2012). *Handbook of web surveys*. Hoboken, New Jersey: Wiley.
- Beukenhorst, D., & Wetzels, W. (2009). *A comparison of two mixed mode designs: cati-capi and web-caticapi*. Presented at the European Survey Research Association Conference. Presented at the European Survey Research Association (ESRA), Warsaw, Poland.

- Bhaskaran, V., & LeClaire, J. (2010). *Online surveys for dummies*. Indianapolis, IN: Wiley.
- Bianchi, A., & Biffignandi, S. (2014). Responsive design for economic data in mixed mode panels. In F. Mecatti, P. L. Conti, & M. G. Ranalli (Eds), *Contribution to sampling statistics*. (pp. 85–101) Berlin: Springer.
- Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds). (2004). *Measurement Errors in Surveys*. New York: Wiley.
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, New Jersey: Wiley.
- Biner, P. M., & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology and Marketing*, 11(5), 483–492.
- Black, G. S. (1998). *Internet surveys: A replacement technology*. Presented at The American Association for Public Opinion Research (AAPOR) 53rd Annual Conference, St. Louis, Missouri.
- Blair, J., & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly*, 75(4), 636–658.
- Blair, J., Czaja, R. F., & Blair, E. A. (2014). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Thousand Oaks, CA: SAGE.
- Blanke, K. (2011). *Scrolling or paging – it depends*. Presented at the 5th Internet Survey Methodology Workshop, Hague, Netherlands.
- Blanke, K. & Luiten, A. (2014). *Query on Data Collection for Social Surveys*. Deliverable for ESSnet Data Collection for Social Surveys using Multiple Modes. Luxemburg: Eurostat. Retrieved from http://www.cros-portal.eu/sites/default/files/Query_report_DCSS.pdf.
- Blasius, J. (2012). Comparing ranking techniques in web surveys. *Field Methods*, 24(4), 382–398.
- Blasius, J., & Thiessen, V. (2012). *Assessing the quality of survey data*. London: SAGE.
- Blom, A. G., Gathmann, C., & Krieger, U. (2013, July). The German Internet Panel: Method and results. Mannheim: Universität Mannheim. Retrieved from http://reforms.uni-mannheim.de/internet_panel/home/gip_concept/gip_overview_for_sfb_website.pdf.
- Blyth, B. (2008). Mixed-mode: The only ‘fitness’ regime? *International Journal of Market Research*, 50(2), 241–266.
- Bocarnea, M. C., Reynolds, R. A., & Baker, J. D. (Eds). (2012). *Online instruments, data collection, and electronic measurements: Organizational advancements*. Hershey, PA: IGI Global.
- Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior*, 23(4), 463–77.
- Borg, I., & Mastrangelo, P. M. (2008). *Employee surveys in management: Theories, tools, and practical applications*. Cambridge, MA: Hogrefe.
- Börkan, B. (2010). The mode effect in mixed-mode surveys: Mail and web surveys. *Social Science Computer Review*, 28(3), 371–380.
- Bosnjak, M., & Batinic, B. (2002). Understanding the willingness to participate in online surveys: the case of e-mail questionnaires. In B. Batanic, U.-D. Reips, & M. Bosnjak (Eds), *Online social sciences* (pp. 115–116). Seattle: Hogrefe.
- Bosnjak, M., Haas, I., Galesic, M., Kaczmirek, L., Bandilla, W., & Couper, M. P. (2013). Sample composition discrepancies in different stages of a probability-based online panel. *Field Methods*, 25(4), 339–360.
- Bosnjak, M., Neubarth, W., Couper, M. P., Bandilla, W., & Kaczmirek, L. (2008). Prenotification in Web based access panel surveys: The influence of mobile text messaging versus e-mail on response rates and sample composition. *Social Science Computer Review*, 26(2), 213–223.
- Bosnjak, M., Poggio, T., Becker, K. R., Funke, F., Wachenfeld, A., & Fisher, B. (2012). *Mobile survey participation rates in commercial market research: A meta-analysis*. Presented at the 6th Internet Survey Methodology Workshop, Ljubljana, Slovenia.
- Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in Web-based surveys. *Journal of Computer-Mediated Communication*, 6(3).

- Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in web surveys: An experiment. *Social Science Computer Review*, 21(2), 208–217.
- Bosnjak, M., Tuten, T. L., & Wittmann, W. W. (2005). Unit (non)response in Web-based access panel surveys: An extended planned-behavior approach. *Psychology and Marketing*, 22(6), 489–505.
- Boulianne, S. (2013). Examining the gender effects of different incentive amounts in a web survey. *Field Methods*, 25(1), 91–104.
- Boulianne, S., Klofstad, C. A., & Basson, D. (2011). Sponsor prominence and responses patterns to an online survey. *International Journal of Public Opinion Research*, 23(1), 79–87.
- Boyle, J. M., Ball, S., Ding, H., Euler, G., Starcie, G., & Lewis, F. (2013). Understanding bias in probability and non-probability samples of a rare population. Presented at the The American Association for Public Opinion Research (AAPOR) 68th Annual Conference, Boston, Massachusetts.
- Brace, I. (2008). *Questionnaire design: How to plan, structure and write survey material for effective market research* (2nd ed.). London: Kogan Page.
- Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design for market research, political polls, and social and health questionnaires* (Revised.). San Francisco: Jossey-Bass.
- Bradley, M., & Lang, P. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59.
- Braunsberger, K., Wybenga, H., & Gates, R. (2007). A comparison of reliability between telephone and webbased surveys. *Journal of Business Research*, 60(7), 758–64.
- Brick, J. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 36–59.
- Brickman Bhutta, C. B. (2012). Not by the book: Facebook as a sampling frame. *Sociological Methods & Research*, 41(1), 57–88.
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds), *Usability evaluation in industry* (pp. 189–194). London: Taylor and Francis.
- Bruijne, M. de, & Wijnant, A. (2013). Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computerassisted web survey. *Social Science Computer Review*, 31(4), 482–504.
- Buchanan, E. A., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of human research ethics committees. *Journal of Empirical Research on Human Research Ethics: An International Journal*, 4(2), 37–48.
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90(1), 125–144.
- Buelens, B., & van den Brakel, J. A. (2014). Measurement error calibration in mixed-mode sample surveys. *Sociological Methods & Research*. Published online before print.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high quality, data? *Perspectives on Psychological Science*, 6(1), 3–5.
- Burdein, I. (2014). Shorter isn't always better. *CASRO Journal*, 2014, 5–42.
- Burr, M. A., Levin, K., & Becher, A. (2001). Examining Web vs paper mode effects in a federal government customer satisfaction study. Presented at The American Association for Public Opinion Research (AAPOR) 56th Annual Conference, Montreal, Quebec, Canada.
- Buskirk, T. D., & Andrus, C. (2012). Smart surveys for smart phones: exploring various approaches for conducting online mobile surveys via smartphones. *Survey Practice* 5(1).
- Buskirk, T. D., & Andrus, C. (2013). Online surveys aren't just for computers anymore! Exploring potential mode effects between smartphone and computer-based online surveys. *Proceedings of the Joint Statistical Meeting, American Association for Public Opinion Research Conference* (pp. 5678–5691). Washi ngton, DC: AMSTAT.

- Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice* 3(6).
- Callegaro, M. (2013a). From mixed-mode to multiple devices. Web surveys, smartphone surveys and apps: Has the respondent gone ahead of us in answering surveys? *International Journal of Market Research*, 55(2), 317–320.
- Callegaro, M. (2013b). Paradata in web surveys. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic use of process information* (pp. 261–279). Hoboken, New Jersey: Wiley.
- Callegaro, M. (2014a). Recent books and journals in public opinion, survey methods, and survey statistics. *Survey Practice*, 7(2).
- Callegaro, M. (2014b). Using Paradata to Improve the Quality of Web Surveys: Some Examples and Applications. Presented at the Conference on Paradata – From survey research to practice, Royal Statistical Society, London. Retrieved from http://www.southampton.ac.uk/s3ri/news/events/2014/06/26_paradata_from_survey_research_to_practice.page.
- Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (Eds). (2014a). *Online Panel Research: A Data Quality Perspective*. Chichester, United Kingdom: Wiley.
- Callegaro, M., Baker, R. P., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014b). Online panel research: History, concepts, applications and a look at the future. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 1–22). Chichester: Wiley.
- Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, 72(5), 1008–1032.
- Callegaro, M., DiSogra, C., & Wells, T. (2011). How the order of response options in a running tally can affect online survey estimates. Presented at The American Association for Public Opinion Research (AAPOR) 66th Annual Conference, Phoenix, Arizona.
- Callegaro, M., & Krosnick, J. A. (2014). Introduction to part I. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 56–60). Chichester: Wiley.
- Callegaro, M., Murakami, M., Tepman, Z., & Henderson, V. (2015). Yes-No answers versus check-all in self-administered modes. A systematic review and analyses. *International Journal of Market Research*. 57(2).
- Callegaro, M., Villar, A., Yeager, D. S., & Krosnick, J. A. (2014). A critical review of studies investigating the quality of data obtained with online panels. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 23–53). Chichester: Wiley.
- Callegaro, M., & Wells, T. (2008). Do online respondents go the extra mile and take on inconvenient tasks? Retrieved from http://www.knowledgenetworks.com/ganp/docs/Do-Online-Respondents-Go-the-Extra-Mile-and-Take-on-Inconvenient-Tasks_6-25-08.pdf.
- Callegaro, M., Yang, Y., Bhola, D. S., Dillman, D. A., & Chin, T.-Y. (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 103(1), 5–25.
- Campanelli, P. (2008). Testing survey questions. In E. de Leeuw, J. J. Hox, & D. A. Dillman (Eds), *International handbook of survey methodology* (pp. 176–200). New York: Lawrence Erlbaum.
- Cape, P. (2010). Questionnaire length, fatigue effects and response quality revisited. Retrieved from http://www.surveysampling.com/ssi-media/Corporate/white_papers/SSI_QuestionLength_WP.image.
- Carini, R., Hayek, J., Kuh, G., Kennedy, J. M. & Ouimet, J. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education*, 44(1), 11–19.
- Casey, T. W., & Poropat, A. (2014). Beauty is more than screen deep: Improving the web survey respondent experience through socially-present and aesthetically-pleasing user interfaces. *Computers in Human Behavior*, 30(2014), 153–163.

- CASRO. (2011). Code of standards and ethics for survey research. Council of American Survey Research Organizations. Port Jefferson, NY: Council of American Survey Research Organizations.
- Cechanowicz, J., Gutwin, C., Brownell, B., & Goodfellow, L. (2013). Effects of gamification on participation and data quality in a real-world market research domain. In *Gamification 2013* (pp. 58–65). Stratford, Ontario.
- Chan, P., & Ambrose, D. (2011). Canadian online panels: Similar or different? *Vue* (January/February), 16–20.
- Chaney, B. H., Barry, A. E., Chaney, J. D., Stollefson, M. L., & Webb, M. C. (2013). Using screen video capture software to aide and inform cognitive interviewing. *Quality & Quantity*, 47(5), 2529–2537.
- Chang, L., Frost, L. Z., Chao, S., & Ree, M. J. (2010). Instrument development with Web surveys and multiple imputations. *Military Psychology*, 22(1), 7–23.
- Chang, L., & Krosnick, J. A. (2009). National surveys via Rdd telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73(4), 641–678.
- Chang, L., & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered computerized questionnaires: An experiment. *Public Opinion Quarterly*, 74(1), 154–167.
- Chien, Y.-T., & Chang, C.-Y. (2012). Exploring the impact of animation-based questionnaire on conducting a web-based educational survey and its association with vividness of respondents' visual images. *British Journal of Educational Technology*, 43(3), E81–E85.
- Chittenden, L., & Rettie, R. (2003). An evaluation of e-mail marketing and factors affecting response. *Journal of Targeting, Measurement and Analysis for Marketing*, 11(3), 203–217.
- Cho, H., & Larose, R. (1999). Privacy issues in Internet surveys. *Social Science Computer Review*, 17(4), 421–434.
- Chou, P.-N. (2012). The integration of Facebook into class management: An exploratory study. *Educational Research*, 3(7), 572–575.
- Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping respondents get it right the first time: The influence of words, symbols, and graphics in Web surveys. *Public Opinion Quarterly*, 71(1), 113–25.
- Clayton, R. L., & Werking, G. S. (1998). Business surveys of the future: The world wide web as a data collection methodology. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds), *Computer-assisted survey information collection* (pp. 543–562). New York: Wiley.
- Cobanoglu, C., & Cobanoglu, N. (2003). The effect of incentives in web surveys: Application and ethical considerations. *International Journal of Market Research*, 45(4), 475–488.
- Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax, and web-based survey methods. *International Journal of Market Research*, 43(4), 441–452.
- Cobben, F., & Bethlehem, J. (2013). *Web panels for official statistics*. Discussion Paper 201307. The Hague, The Netherlands: Statistics Netherlands.
- Comley, P. (2000). Pop-up surveys: What works, what doesn't work and what will work in the future. In ESOMAR (Ed.), *Proceedings of the ESOMAR Worldwide Internet Conference: Net Effects 3*. Amsterdam: ESOMAR.
- Comley, P. (2006). The games we play: A psychoanalysis of the relationship between panel owners and panel participants. Presented at the *Panel Research Conference*. Amsterdam: ESOMAR.
- Comley, P. (2007). Online market research. In M. van Hamersveld & C. de Bont (Eds), *Market research handbook* (5th ed., pp. 401–419). Chichester: Wiley.
- Conrad, F. G., Schober, M. F., & Coiner, T. (2007). Bringing features of human dialogue to Web surveys. *Applied Cognitive Psychology*, 21(2), 165–87.
- Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2010). *Professional web respondents and data quality*. Presented at The American Association for Public Opinion Research (AAPOR)

- 65th Annual Conference, Chicago, Illinois.
- Converse, J. M. (2009). *Survey research in the United States: Roots and emergence 1890–1960*. Piscataway, NJ: Transaction Publishers.
- Converse, P. D., Wolfe, E. W., Huang, X., & Oswald, F. L. (2008). Response rates for mixed-mode surveys using mail and e-mail/web. *American Journal of Evaluation*, 29(1), 99–107.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in Web or internet-based surveys. *Educational and Psychological Measurement*, 60(6), 821–836.
- Cooke, M., Nielsen, A., & Strong, C. (2003). The use of SMS as a research tool. In R. Banks, J. Currall, J. Francis, L. Gerrard, R. Khan, T. Macer, ... A. Westlake (Eds), *The impact of technology on the survey process. Proceedings of the Fourth International Conference on Survey and Statistical Computing* (pp. 267–276). Chesham, Bucks, UK: Association for Survey Computing. Retrieved from <http://www.asc.org.uk/>.
- Coombe, R., Jarrett, C., & Johnson, A. (2011). Usability testing of market research questionnaires. Presented at the European Survey Research Association Conference. Presented at the European Survey Research Association (ESRA), Lausanne, Switzerland.
- Coromina, L., & Coenders, G. (2006). Reliability and validity of egocentered network data collected via web: A meta-analysis of multilevel multitrait multimethod studies. *Social Networks*, 28(3), 209–231.
- Corti, L., van den Eyden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. London: SAGE.
- Council of American Survey Research Organizations. (2012). *Rules for governing sweepstakes*. Port Jefferson, NY: Council of American Survey Research Organizations. Retrieved from http://c.ymcdn.com/sites/www.casro.org/resource/resmgr/docs/rules_for_governing_sweepsta.pdf.
- Couper, M. P. (1998). Measuring survey quality in a CASIC environment. In American Statistical Association (Ed.), *Proceedings of the Joint Statistical Meeting, Survey Research Methods Section* (pp. 41–49). Washington, DC: AMSTAT.
- Couper, M. P. (2000a). Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review*, 18(4), 384–396.
- Couper, M. P. (2000b). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464–494.
- Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review*, 23(4), 486–501.
- Couper, M. P. (2008). *Designing effective web surveys*. New York: Cambridge University Press.
- Couper, M. P. (2011). The future of modes of data collection. *Public Opinion Quarterly*, 75(5), 889–908.
- Couper, M. P. (2013). *Surveys on mobile devices: Opportunities and challenges*. Presented at the NCRM Conference: Web surveys for the general population: How, why and when?, London, United Kingdom. Retrieved from <http://www.natcenweb.co.uk/genpopweb/documents/Mick-Couper.pdf>.
- Couper, M. P., Baker, R. P., & Mechling, J. (2011). Placement and design of navigation buttons in web surveys. *Survey Practice*, 4(1).
- Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2005). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, 36(1), 131–148.
- Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing input fields for non-narrative open-ended responses in web surveys. *Journal of Official Statistics*, 27(1), 65–85.
- Couper, M. P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society, Journal of the Royal Statistical Society, Series A*, 176(1), 271–286.
- Couper, M. P., & Nicholls II, W. L. (1998). The history and development of computer assisted survey information collection. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds), *Computer assisted survey information collection* (pp. 1–21).

- New York: Wiley.
- Couper, M. P., & Singer, E. (2013). Informed consent for web paradata use. *Survey Research Methods*, 7(1), 57–67.
- Couper, M. P., Singer, E., Conrad, F. G., & Groves, R. M. (2008). Risk of disclosure, perceptions of risk, and concerns about privacy and confidentiality as factors in survey participation. *Journal of Official Statistics*, 24(2), 255–275.
- Couper, M. P., Tourangeau, R., & Conrad, F. G. (2006). Evaluating the effectiveness of visual analog scales: A Web experiment. *Social Science Computer Review*, 24(2), 227–245.
- Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What they see is what we get: Response options for web surveys. *Social Science Computer Review*, 22(1), 111–127.
- Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review*, 31(3), 322–345.
- Couper, M. P., Tourangeau, R., & Smith, T. W. (1998). Collecting sensitive information with different modes of data collection. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds), *Computer assisted survey information collection* (pp. 431–454). New York: Wiley.
- Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65(2), 230–253.
- Couper, M. P., Zhang, C., Conrad, F. G., & Tourangeau, R. (2012). *Database lookup in web surveys*. Presented at the 6th Internet survey methodology workshop, Ljubljana, Slovenia.
- Crawford, S. (2002). Evaluation of web survey data collection systems. *Field Methods*, 14(3), 307–321.
- Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perception of burden. *Social Science Computer Review*, 19(2), 146–162.
- Crawford, S. D., McCabe, S. E., & Pope, D. (2005). Applying Web-based survey design standards. *Journal of Prevention and Intervention in the Community*, 29(1–2), 43–66.
- Crutzen, R., & Göritz, A. S. (2010). Social desirability and self-reported health risk behaviors in web-based research: Three longitudinal studies. *BMC Public Health*, 10(1), 720.
- Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience*. New York: HarperCollins.
- Čehovin, G., & Vehovar, V. (2012a). *WebSM study: Overview of features of software packages*. Retrieved from <http://www.websm.org/db/12/16586/>.
- Čehovin, G., & Vehovar, V. (2012b). *WebSM Study: Speed and efficiency of online survey tools*. Retrieved from <http://www.websm.org/db/12/16583/>.
- da Costa, F. F., Schmoelz, C. P., Davies, V. F., di Pietro, P. F., Kupek, E., & de Assis, M. A. A. (2013). Assessment of diet and physical activity of Brazilian schoolchildren: Usability testing of a web-based questionnaire. *JMIR Research Protocols*, 2(2), 1–15.
- Das, M. (2012). Innovation in online data collection for scientific research: The Dutch MESS project. *Methodological Innovations Online*, 7(1), 7–24.
- Datta, R. A., Yan, T., Evans, D., Pedlow, S., Spencer, B., & Bautista, R. (2012, March 13). 2010 Census Integrated Communications Program Evaluation (CICPE). Retrieved from https://www.census.gov/2010census/pdf/2010_Census_ICP_Evaluation.pdf.
- De Beuckelaer, A., & Lievens, F. (2009). Measurement equivalence of paper-and-pencil and internet organisational surveys: A large scale examination in 16 Countries. *Applied Psychology*, 58(2), 336–361.
- de Leeuw, E. (2005). To mix or not to mix? Data collection modes in surveys. *Journal of Official Statistics*, 21(2), 1–23.
- de Leeuw, E., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys. *Public Opinion Quarterly*, 71(3), 413–443.
- de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds),

- Survey Nonresponse* (pp. 41–54). Hoboken, New Jersey: Wiley.
- de Leeuw, E., & Hox, J. (2011). Internet surveys as a part of a mixed-mode design. In M. Das, P. Ester, & L. Kaczmirek (Eds), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies* (pp. 45–76). New York: Routledge.
- de Leeuw, E., & Hox, J. J. (2008). Missing data. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*, pp. 468–472. Thousand Oaks, CA: SAGE.
- de Leeuw, E., Hox, J. J., & Snijders, G. (1995). The effect of computer-assisted interviewing on data quality: A review. *Journal of the Market Research Society*, 37(4), 325–344.
- de Leeuw, E., & Nicholls II, W. L. (1996). Technological innovations in data collection: Acceptance, data quality and costs. *Sociological Research Online*, 1(4).
- Dean, E., Head, B., & Swicegood, J. (2013). Virtual cognitive interviewing using Skype and Second Life. In C. A. Hill, E. Dean, & J. Murphy (Eds), *Social Media, Sociality, and Survey Research* (pp. 107–132). Hoboken, New Jersey: Wiley.
- DeMaio, T., & Landreth, A. (2004). Examining expert reviews as a pretest method. In P. Prüfer, M. Rexroth, & F. J. Fowler (Eds), *ZUMA-Nachrichten Spezial Band 9, QUEST 2003. Proceedings of the 4th Conference on Questionnaire Evaluation Standards* (pp. 60–72). Mannheim, Germany: Gesis.
- Dennis, J. M., de Rouvray, C., & Couper, M. P. (2000). Questionnaire design for probability-based Web surveys. Presented at The American Association for Public Opinion Research (AAPOR) 55th Annual Conference, Portland, Oregon.
- de Nooy, W., Mrvar, A., and Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (2nd Edition). New York: Cambridge University Press.
- DeRouvray, C., & Couper, M. P. (2002). Designing a strategy for reducing ‘No opinion’ responses in Webbased surveys. *Social Science Computer Review*, 20(1), 3–9.
- Deutsdens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of Internet-based surveys: An experimental study. *Marketing Letters*, 15(1), 21–36.
- Deutsdens, E., Ruyter, K. de, & Wetzels, M. (2006). An assessment of equivalence between online and mail surveys in service research. *Journal of Service Research*, 8(4), 346–355.
- Dietrich, B. J. (2008). Mean square error (MSE). In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*, (pp. 456–458). Thousand Oaks, California: SAGE.
- Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.
- Dillman, D. A., & Messer, B. L. (2010). Mixed-mode surveys. In P. V. Marsden & J. D. Wright (Eds), *Handbook of survey research* (2nd ed., pp. 551–574). Howard House, United Kingdom: Emerald Group Publishing.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, New Jersey: Wiley.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, New Jersey: Wiley.
- Dillman, D. A., Tortora, R., Conradt, J., & Bowker, D. K. (1998). Influence of plain vs fancy design on response rates for web surveys. *Proceedings of the Joint Statistical Meeting, Survey Research Methods Section*. Dallas, Texas: AMSTAT.
- Diment, K., & Garrett-Jones, S. (2007). How demographic characteristics affect mode preference in a postal/web mixed-mode survey of Australian researchers. *Social Science Computer Review*, 25(3), 410–417.
- DiSogra, C., & Callegaro, M. (2015). Metrics and design tool for building and evaluating probability-based online panels. *Social Science Computer Review*.
- DiSogra, C., Callegaro, M., & Hendarwan, E. (2010). Recruiting probability-based web panel members using an address-based sample frame: Results from a pilot study conducted by Knowledge Networks. In *Proceedings of the Annual Meeting of the American Statistical Association* (pp. 5270–5283). Hollywood, FL: AMSTAT.
- Dobrow, M. J., Orchard, M. C., Golden, B., Holowaty, E., Paszat, L., Brown, A. D., & Sullivan,

- T. (2008). Response audit of an Internet survey of health care providers and administrators: implications for determination of response rates. *Journal of Medical Internet Research*, 10(4), e30.
- Donald, E. (2010). Using a fillable PDF together with SAS ® for questionnaire data. Presented at the 23rd NorthEast SAS Users Group, Baltimore, MD. Retrieved from <http://www.nesug.org/Proceedings/nesug10/cc/cc10.pdf>.
- Downes-Le Guin, T., Baker, R. P., Mechling, J., & Ruyle, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research*, 54(5), 613–633.
- Downes-Le Guin, T., Janowitz, P., Stone, R., & Khorram, S. (2002). Use of pre-incentives in an internet survey. *Journal of Online Research*.
- Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial and Organizational Psychology*, 2(1), 46–48.
- Dykema, J., Stevenson, J., Day, B., Sellers, S. L., & Bonham, V. L. (2011). Effects of incentives and prenotification on response rates and costs in a national web survey of physicians. *Evaluation & the Health Professions*, 34(4), 434–447.
- Economic and Statistics Administration, & National Telecommunications and Information Administration. (2010). *Exploring the digital nation: Home broadband internet adoption in the United States*. Washington, DC: US Department of Commerce.
- Edwards, P. J., Roberts, I., Clarke, M. J., DiGuseppi, C., Wentz, R., Kwan, I., ... Prata, S. (2009). Methods to increase response to postal and electronic questionnaires. *Cochrane Database of Systematic Reviews* 8(3). Chichester: Wiley.
- Edwards, S. W., & Cantor, D. (1991). Towards a response model in establishment surveys. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds), *Measurement errors in surveys* (pp. 211–236). New York: Wiley.
- Ehling, M., & Körner, T. (Eds). (2007). *Handbook on data quality assessment methods and tools*. Wiesbaden: Eurostat.
- Ekeh, P. P. (1974). *Social exchange theory: The two traditions*. Cambridge, MA: Harvard University Press.
- Elling, S., Lentz, L., de Jong, M., & van den Bergh, H. (2012). Measuring the quality of governmental websites in a controlled versus an online setting with the ‘Website Evaluation Questionnaire’. *Government Information Quarterly*, 29(3), 383–393.
- Emde, M., & Fuchs, M. (2012a). Exploring animated faces scales in web surveys: Drawbacks and prospects. *Survey Practice*, 5(1).
- Emde, M., & Fuchs, M. (2012b). Using adaptive questionnaire design in open-ended questions: A field experiment. *Proceedings of the Survey Research Methods Section, American Statistical Association*. San Diego, California.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.
- ESOMAR. (2011a). *ESOMAR guideline for online research*. Amsterdam: ESOMAR.
- ESOMAR. (2011b). *Esomar guideline on social media research*. Amsterdam: ESOMAR. Retrieved from <http://www.websm.org/db/35/17108/>
- ESOMAR. (2012). *28 Questions to help research buyers of online samples*. Amsterdam: ESOMAR. Retrieved from <http://www.websm.org/db/35/17126/>
- ESOMAR. (2013). *Global market research 2013*. Amsterdam: ESOMAR.
- Ester, P., & Vinken, H. (2010). Measuring attitudes toward controversial issues in internet surveys: Order effects of open and closed questioning. In M. Das, P. Ester, & L. Kaczmarek (Eds), *Social and behavioral research and the internet. Advances in applied methods and research strategies* (pp. 245–268). New York: Routledge.
- Esuli, A., & Sebastiani, F. (2010). Machines that learn how to code open ended survey data. *International Journal of Market Research*, 52(6), 775–800.
- European Commission. (2014). *Eurobarometer special surveys: Special Eurobarometer 414*.

- Brussels, European Commission. Retrieved from http://ec.europa.eu/public_opinion/archives/ebs/ebs_414_en.pdf.
- Eurostat. (2009). *ESS standard for quality reports*. Luxemburg: Eurostat.
- Eurostat. (2011). *European Statistics Code of Practice*. Luxemburg: Eurostat.
- Ewing, T. (2012). Not just playing around. *Quirk's Marketing Research Review*, 26(3), 30–34.
- Eysenbach, G., & Wyatt, J. (2002). Facilitating research. In B. C. McKenzie (Ed.), *Medicine and the internet* (3rd ed., pp. 211–226). Oxford: Oxford University Press.
- Faas, T., & Schoen, H. (2006). Putting a questionnaire on the Web is not enough: A comparison of online and offline surveys conducted in the context of the German Federal Election 2002. *Journal of Official Statistics*, 22(2), 177–190.
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26(2), 132–139.
- Fang, J., Shao, P., & Lan, G. (2009). Effects of innovativeness and trust on web survey participation. *Computers in Human Behavior*, 25(1), 144–152.
- Fang, J., & Wen, C. (2012). Predicting potential respondents' decision to participate in web surveys. *International Journal of Services Technology and Management*, 18(1/2), 16.
- Fang, J., Wen, C., & Prybutok, V. (2013). The equivalence of Internet versus paper-based surveys in IT/IS adoption research in collectivistic cultures: The impact of satisficing. *Behaviour & Information Technology*, 32(5), 480–490.
- Faught, K. S., Whitten, D., & Green Jr, K. W. (2004). Doing survey research on the internet: Yes, timing does matter. *Journal of Computer Information Systems*, 44(3), 26–34.
- Felix, L. M., Burchett, H. E., & Edwards, P. J. (2011). Factorial trial found mixed evidence of effects of prenotification and pleading on response to web-based survey. *Journal of Clinical Epidemiology*, 64(5), 531–536.
- Fenner, Y., Garland, S. M., Moore, E. E., Jayasinghe, Y., Fletcher, A., Tabrizi, S. N., ... Wark, J. D. (2012). Web-based recruiting for health research using a social networking site: An exploratory study. *Journal of Medical Internet Research*, 14(1), e20.
- Ferneer, H., & Scherpenzeel, A. (2013). The smartphone in survey research: Experiments for time use data. *Survey Statistician*, 67(January), 19–25.
- Fisher, B., & Bernet, F. (2014). Device effects: How different screen sizes affect answer quality in online questionnaires. Presented at the General Online Research Conference (GOR), Cologne, Germany.
- Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item Likert and Visual Analogue Scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*, 20(1), 49–58.
- Fontaine, S., Jacquemain, M., & Italiano, P. (2013). Specific mixed-mode methodologies to include sensory disabled people in quantitative surveys. Presented at the European Survey Research Association (ESRA), Ljubljana, Slovenia.
- Fowler, F. J. (2014). *Survey research methods* (5th ed.). Thousand Oaks, CA: SAGE.
- Fricker, S. S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of Web and telephone surveys. *Public Opinion Quarterly*, 69(3), 370–392.
- Fuchs, M. (2003). Kognitive prozesse und antwortverhalten in einer Internet-befragung (Cognitive processes and behavioral response in an Internet survey). *Österreichische Zeitschrift für Soziologie*, 28(4), 19–45.
- Fuchs, M. (2008). Mobile web surveys: A preliminary discussion of methodological implications. In F. G. Conrad & M. F. Schober (Eds), *Envisioning the survey interview of the future* (pp. 77–94). Hoboken, NJ: Wiley.
- Fuchs, M. (2009). Gender-of-interviewer effects in a video-enhanced web survey: Results from a randomized field experiment. *Social Psychology*, 40(1), 37–42.
- Fuchs, M., Bossert, D., & Stukowski, S. (2013). Response rate and nonresponse bias: Impact of the number of contact attempts on data quality in the European Social Survey. *Bulletin of*

- Sociological Methodology/Bulletin de Méthodologie Sociologique*, 117(1), 26–45.
- Fulgoni, G. (2005). *The professional respondent problem in online panel surveys today*. Presented at the Market Research Association annual conference, Chicago, Illinois.
- Funke, F., & Reips, U.-D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*, 24(3), 310–327.
- Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*, 29(2), 221–231.
- Gaiser, T. J. (2008). Online focus groups. In N. G. Fielding, R. M. Lee & G. Blank (Eds), *The SAGE handbook of online research methods* (1st ed., pp. 290–306). London: SAGE.
- Gajic, A., Cameron, D., & Hurley, J. (2012). The cost-effectiveness of cash versus lottery incentives for a webbased, stated-preference community survey. *The European Journal of Health Economics*, 13(6), 789–799.
- Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, 22(2), 313–328.
- Galesic, M. (2009). *Comparing the results of web surveys on volunteer versus probabilistically selected panels of participants in Germany and the US*. Presented at the 4th Internet Survey Methodology Workshop, Bergamo, Italy.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360.
- Galesic, M., & Tourangeau, R. (2007). What is sexual harassment? It depends on who asks! Framing effects on survey responses. *Applied Cognitive Psychology*, 21(2), 189–202.
- Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). Using change to improve navigation in grid questions. Presented at the General Online Research Conference (GOR), Leipzig, Germany.
- Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly*, 72(5), 892–913.
- Galesic, M., & Yan, T. (2011). Using eye-tracking for studying survey response process. In M. Das, P. Ester, & L. Kaczmarek (Eds), *Social and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 349–370). New York: Routledge.
- Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, 2(1), 21–32.
- Garland, P., Chen, K., Epstein, L., & Suh, A. (2013). Speed (necessarily) doesn't kill: A new way to detect survey satisficing. *CASRO Journal*, 2012–2013, 21–23.
- Gelder, M. M. H. J. van, Bretveld, R. W., & Roeleveld, N. (2010). Web-based questionnaires: The future in epidemiology? *American Journal of Epidemiology*, 172(11), 1292–1298.
- Glasner, T., & van der Vaart, W. (2013). Life history calendars: A viable method for web-based data collection? Presented at the NCRM Conference: Web surveys for the general population – How, why and when?, London, United Kingdom.
- GMI. (2012). GMI Pinnacle.
- Göriz, A. S. (2005). Contingent versus unconditional incentives in WWW studies. *Advances in Methodology and Statistics (Metodološki Zvezki)*, 2(1), 1–14.
- Göriz, A. S. (2006a). Cash lotteries as incentives in online panels. *Social Science Computer Review*, 24(4), 445–459.
- Göriz, A. S. (2006b). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science*, 1(1), 58–70.
- Göriz, A. S. (2007). Using online panels in psychological research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds), *The Oxford handbook of Internet psychology* (pp. 473–485). Oxford: Oxford University Press.
- Göriz, A. S. (2008). The long-term effect of material incentives on participation in online panels. *Field Methods*, 20(3), 211–225.

- Görizt, A. S. (2010). Using lotteries, loyalty points, and other incentives to increase participant response and completion. In S. D. Gosling & J. A. Johnson (Eds), *Advanced methods for conducting online behavioural research* (pp. 219–233). Washington, DC: American Psychological Association.
- Görizt, A. S., & Crutzen, R. (2012). Reminders in web-based data collection: Increasing response at the price of retention? *American Journal of Evaluation*, 33(2), 240–250.
- Görizt, A. S., & Luthe, S. C. (2013). Effects of lotteries on response behavior in online panels. *Field Methods*, 25(3), 219–237.
- Görizt, A. S., Wolff, H.-G., & Goldstein, D. G. (2008). Individual payments as a longer-term incentive in online panels. *Behavior Research Methods*, 40(4), 1144–1149.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25(2), 161.
- Goyder, J. (1988). *The silent minority: Nonrespondents on sample surveys*. Boulder, Colorado: Westview Press.
- Grady, J. (2010, February 23). *Libraries nationwide receiving ALA-APA Library Salary Survey invitation*. Retrieved from <http://www.ala.org>.
- Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question understanding aid (QUAID): A web facility that tests question comprehensibility. *Public Opinion Quarterly*, 70(1), 3–22.
- Grandquist, L., & Kovar, J. G. (1997). Editing of survey data: How much is enough? In L. E. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds), *Survey measurement and process quality* (pp. 415–436). New York: Wiley.
- Granello, D. H., & Weathon, J. E. (2003). Using web-based surveys to conduct counseling research. In J. W. Bloom & G. R. Walz (Eds), *Cybercounseling and cyberlearning: An encore* (pp. 287–305). Alexandria, Virginia: American Counseling Association.
- Greene, J., Speizer, H., & Wiitala, W. (2008). Telephone and web: Mixed-mode challenge. *Health Services Research*, 43(1 Pt 1), 230–248.
- Greenlaw, C., & Brown-Welty, S. (2009). A comparison of web-based and paper-based survey methods: Testing assumptions of survey mode and response cost. *Evaluation Review*, 33(5), 464–480.
- Grenville, A. (2012). *Debunking mobile research*. Retrieved from http://www.visioncritical.com/system/files/WHITE_PAPER_Debunking_Mobile_Research-Its_just_the_same_or_is_it.pdf.
- Greski, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Görizt, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 238–262). Chichester: Wiley.
- Grey Matter Research. (2012). *More dirty little secrets of online panel research*. Retrieved from http://www.greymatterresearch.com/index_files/Online_Panels_2012.htm.
- Griffis, S. E., Goldsby, T. J., & Cooper, M. (2003). Web-based and mail surveys: A comparison of response, data and cost. *Journal of Business Logistics*, 24(2), 237–258.
- Grossnickle, J., & Raskin, O. (2001). *The handbook of online marketing research*. New York: McGraw-Hill.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.
- Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L., & Waksberg, J. (2001). *Telephone survey methodology*. New York: Wiley.
- Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4), 475–495.
- Groves, R. M., & Couper, M. P. (2008). *Nonresponse in household interview surveys*. Hoboken, New Jersey: Wiley.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: Wiley.
- Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys. *Journal of the*

- Royal Statistical Society. Series A*, 169(3), 439–457.
- Groves, R. M., & Lyberg, L. E. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74(5), 849–879.
- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72(2), 167–189.
- Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, 64(3), 299–308.
- Guéguen, N., & Jacob, C. (2002). Solicitation by e-mail and solicitor's status: A field study of social influence on the web. *CyberPsychology & Behavior*, 5(4), 377–383.
- Guidry, K. R. (2012). *Response quality and demographic characteristics of respondents using a mobile device on a web based survey*. Presented at the The American Association for Public Opinion Research (AAPOR) 68th Annual Conference, Orlando, Florida.
- Gutierrez, C., Wells, T., Rao, K., & Kurzynski, D. (2011). *Catch them when you can: Speeders and their role in data quality*. Presented at the 36th Midwest Association for Public Opinion Research (MAPOR) Annual Conference, Chicago, Illinois.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 3025–3034), Hawaii, USA.
- Hammen, K. (2010). The impact of visual and functional design elements in online survey research. Presented at the General Online Research Conference (GOR), Pforzheim, Germany.
- Hansen, J. M., & Smith, S. M. (2012). The impact of two-stage highly interesting questions on completion rates and data quality in online marketing research. *International Journal of Market Research*, 54(2), 241–260.
- Hansen, K. M., & Pedersen, R. T. (2012). Efficiency of different recruitment strategies for web panels. *International Journal of Public Opinion Research*, 24(2), 238–249.
- Hansen, S. E., & Couper, M. P. (2004). Usability testing to evaluate computer-assisted instruments. In *Methods for testing and evaluating survey questionnaires* (pp. 337–360). Hoboken, NJ: Wiley.
- Haraldsen, G. (2005). *Using client side paradata as process quality indicators in web surveys*. Presented at the Exploratory workshop on Internet survey methodology: Towards concerted European research efforts, Dubrovnik, Croatia.
- Hardigan, P. C., Succar, C. T., & Fleisher, J. M. (2012). An analysis of response rate and economic costs between mail and web-based surveys among practicing dentists: A randomized trial. *Journal of Community Health*, 37(2), 383–394.
- Hart, A. M., Brennan, C. W., Sym, D., & Larson, E. (2009). The impact of personalized prenotification on response rates to an electronic survey. *Western Journal of Nursing Research*, 31(1), 17–23.
- Harter, J. K., Schmidt, F. L., Killham, E. A., & Asplund, J. W. (2008). *Q12 meta-analysis*. Retrieved from http://strengths.gallup.com/private/resources/q12meta-analysis_flyer_gen_08%2008_bp.pdf.
- Hartford, K., Carey, R., & Mendonca, J. (2007). Sampling bias in an international internet survey of diversion programs in the criminal justice system. *Evaluation & the Health Professions*, 30(1), 35–46.
- Haunberger, S. (2011a). Explaining unit nonresponse in online panel surveys: An application of the extended theory of planned behavior. *Journal of Applied Social Psychology*, 41(12), 2999–3025.
- Haunberger, S. (2011b). To participate or not to participate: Decision processes related to survey non response. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 109(1), 39–55.
- Hayes, B. E. (2008). *Measuring customer satisfaction and loyalty: Survey design, use, and statistical analysis methods* (3rd ed.). Milwaukee, WI: ASQ Quality Press.
- Healey, B. (2007). Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in web surveys. *Social Science Computer Review*, 25(1),

- 111–128.
- Healey, J. F. (2011). *Statistics: A tool for social research* (9th ed.). Belmont, California: Cengage Learning.
- Heckathorn, D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 4(2), 174–199.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied survey data analysis*. Boca Raton, Florida: Chapman and Hall/CRC.
- Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a Web survey. *Social Science Computer Review*, 21(3), 360–73.
- Heerwegh, D. (2005). *Web surveys: Explaining and reducing unit nonresponse, item nonresponse, and partial nonresponse* (Unpublished doctoral dissertation). Katholieke Universiteit Leuven, Leuven, Belgium.
- Heerwegh, D. (2006). An investigation of the effect of lotteries on web survey response rates. *Field Methods*, 18(2), 205–220.
- Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1), 111–121.
- Heerwegh, D., & Loosveldt, G. (2002). Web surveys: The effect of controlling survey access using PIN numbers. *Social Science Computer Review*, 20(1), 10–21.
- Heerwegh, D., & Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in web surveys. *Journal of Official Statistics*, 22(2), 191–210.
- Heerwegh, D., & Loosveldt, G. (2007). Personalizing e-mail contacts: Its influence on web survey response rate and social desirability response bias. *International Journal of Public Opinion Research*, 19(2), 258–268.
- Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-Internet-coverage population differences in response quality. *Public Opinion Quarterly*, 72(5), 836–846.
- Heerwegh, D., & Loosveldt, G. (2009). Explaining the intention to participate in a web survey: A test of the theory of planned behaviour. *International Journal of Social Research Methodology*, 12(3), 181–195.
- Heerwegh, D., Vanhove, T., Matthijs, K., & Loosveldt, G. (2005). The effect of personalization on response rates and data quality in web surveys. *International Journal of Social Research Methodology*, 8(2), 85–99.
- Heiervang, E., & Goodman, R. (2011). Advantages and limitations of web-based surveys: Evidence from a child mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, 46(1), 69–76.
- Henderson, V., & Callegaro, M. (2011). *Increasing (or decreasing) response rate by changing the subject of email invitations*. Presented at the The American Association for Public Opinion Research (AAPOR) 6th Annual Conference, Phoenix, Arizona.
- Henning, J. (2011). *Causes of survey incompletes: Why panelists say they abandon surveys*. Retrieved from <http://blog.verint.com/>.
- Henning, J. (2013). Improving the representativeness of online surveys. *Alert!*, 4, September.
- Hernson, P. S., Niemi, R. G., Hanmer, M. J., Bederson, B. B., Conrad, F. C., & Traugott, M. W. (2008). *Voting technology: The not-so-simple act of casting a ballot*. Washington, DC: Brookings Institution Press.
- Hertel, G., Naumann, S., Konradt, U., & Batinic, B. (2002). Personality assessment via Internet: Comparing online and paper-and-pencil questionnaires. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds), *Online Social Sciences* (pp. 115–133). Seattle: Hogrefe.
- Hill, C. A., & Dean, E. (2009). *Survey research in virtual worlds: Second life R as a research platform*. Presented at the European Survey Research Association (ESRA), Warsaw, Poland.
- Hillygus, S. D., Jackson, N., & Young, M. (2014). Professional respondents in nonprobability online

- panels. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 219–237). Chichester: Wiley.
- Hogan, B., Carrasco, J. A., & Wellman, B. (2007). Visualizing personal networks: Working with participant-aided sociograms. *Field Methods*, 19(2), 116–144.
- Holbrook, A. L., Krosnick, J. A., Carson, R. T., & Mitchell, R. C. (2000). Violating conversational conventions disrupts cognitive processing of attitude questions. *Journal of Experimental Social Psychology*, 36(5), 465–494.
- Holland, J. L., & Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review*, 27(2), 196–212.
- Hox, J. J. (1997). From theoretical concepts to survey question. In L. E. Lyberg, P. P. Biemer, M. Collins, E. de Leeuw, C. Dippo, & D. Trewin (Eds), *Survey measurement and process quality* (pp. 47–69). New York: Wiley.
- Huggins, V., & Eyerman, J. (2001). Probability based internet surveys: A synopsis of early methods and survey research results. Paper presented at Research Conference of the Federal Committee on Statistical Methodology, Arlington, Virginia.
- Hughes, K. A. (2004). Comparing pretesting methods: Cognitive interviews, respondent debriefing, and behavior coding. Statistical Research Division Study Series, Survey Methodology 2004(02). U.S. Census Bureau. Retrieved from <http://www.census.gov/srd/papers/pdf/rsm2004-02.pdf>.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., & de Wolf, P.-P. (2012). *Statistical disclosure control*. Hoboken, New Jersey: Wiley.
- Iacobucci, D., & Churchill, G. A., Jr. (2015). *Marketing research: Methodological foundations* (11th ed.). Nashville, TN: Earlie Lite Books.
- Iarossi, G. (2006). *The power of survey design: A user's guide for managing surveys, interpreting results, and influencing respondents*. Washington, DC: World Bank Publications.
- ICPSR. (2012). *Guide to social science data preparation: Best practice throughout the data life cycle* (5th ed.). Ann Arbor, Michigan: ICPSR Institute for Social Research University of Michigan.
- Irani, T. A., Gregg, J. A., & Telg, R. (2004). Choosing to use the web: Comparing early and late respondents to an online web-based survey designed to assess IT computer skills perceptions of county extension agents. *Journal of Southern Agricultural Education Research*, 54(1), 168–179.
- ISO. (2009). *ISO 26362: Access panels in market, opinion, and social research – Vocabulary and service requirements*. Geneva: International Organization for Standardization.
- ISO. (2012). *ISO 20252: Market, opinion and social research – Vocabulary and service requirements* (2nd ed.). Geneva: International Organization for Standardization.
- Jäckle, A., Roberts, C., & Lynn, P. (2006). Telephone versus face-to-face interviewing: Mode effects on data quality and likely causes. Report on phase II of the ESS-Gallup mixed mode methodology project. *ISER Working Paper* 2006-41. Colchester: University of Essex.
- Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The flow state scale-2 and dispositional flow scale-2. *Journal of Sport & Exercise Psychology*, 24(2), 133–150.
- Jarrett, C., Gaffney, G., & Krug, S. (2009). *Forms that work: Designing web forms for usability*. Amsterdam: Morgan Kaufmann.
- Johnson, E. P., Siluk, L., & Tarraf, S. (2014). Cyborgs vs monsters: Assembling modular surveys to create complete datasets. *CASRO Journal*, 2014, 49–54.
- Johnston, M. (2007). Spoken and multimodal dialog systems for survey research. In *The Challenges of a Changing World*. Southampton, United Kingdom: Association for Survey Computing.
- Joinson, A. N. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society*, 31(3), 433–438.
- Joinson, A. N., Paine, C., Buchanan, T., & Reips, U.-D. (2008). Measuring self-disclosure online: Blurring and nonresponse to sensitive items in web-based surveys. *Computers in Human*

- Behavior*, 24(5), 2158–2171.
- Joinson, A. N., Woodley, A., & Reips, U.-D. (2007). Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior*, 21(1), 275–285.
- Jue, A., & Luck, K. (2014). Update: Participation of mobile users in online surveys. Retrieved from <http://ww2.focusvision.com>.
- Kaczmirek, L. (2006). A short introduction to usability in online surveys. Presented at the IFD&TC, Montreal, Canada.
- Kaczmirek, L. (2008). Internet survey software tools. In N. G. Fielding, R. M. Lee, & G. Blank. (Eds), *The SAGE handbook of online research methods* (1st ed., pp. 236–54). London: SAGE.
- Kaczmirek, L. (2009). *Human survey-interaction: Usability and nonresponse in online surveys*. Cologne: Herbert Von Halem Verlag.
- Kaczmirek, L. (2011). Attention and usability in internet surveys: Effects of visual feedback in grid questions. In M. Das, P. Ester, & L. Kaczmirek (Eds), *Social and behavioral research and the Internet. Advances in applied methods and research strategies* (pp. 191–214). New York: Routledge.
- Kaczmirek, L., & Wolff, K. G. (2007). Survey design for visually impaired and blind people. In *Proceedings of the 4th International Conference on Universal Access in Human Computer Interaction: Coping with Diversity* (pp. 374–381). Berlin, Heidelberg: Springer-Verlag.
- Kapelner, A., & Chandler, D. (2010). Preventing satisficing in online surveys: A ‘kapcha’ to ensure higher quality data. Presented at the CrowdConf 2010, San Francisco, California.
- Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, 68(1), 94–101.
- Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The effect of invitation design on web survey response rates. *Social Science Computer Review*, 30(3), 339–349.
- Kenneally, E., Bailey, M., & Maughan, D. (2010). A framework for understanding and applying ethical principles in network and security research. In *Proceedings of the 14th International Conference on Financial Cryptography and Data Security: Workshop on ethics in computer security research (WECSR)* (pp. 240–246). Tenerife, Canary Islands, Spain: Springer-Verlag.
- Kenneally, E., Stavrou, A., McHugh, J., & Christin, N. (2011). *Moving forward, building an ethics community (Panel statement)*. Presented at the Workshop on ethics in computer security research (WECSR), Saint Lucia.
- Kennedy, C. (2010). Nonresponse and measurement error in mobile phone surveys. University of Michigan.
- Kennedy, J. M., Tarnai, J., & Wolf, J. G. (2010). Managing survey research projects. In P. Marsden & J. D. Wright (Eds), *Handbook of survey research* (2nd ed., pp. 575–592). Howard House, UK: Emerald Group Publishing.
- Kent, R., & Brandal, H. (2003). Improving email response in a permission marketing context. *International Journal of Market Research*, 45(4), 489–503.
- Keusch, F. (2012a). How to increase response rates in list-based web survey samples. *Social Science Computer Review*, 30(3), 380–388.
- Keusch, F. (2012b). The role of topic interest and topic salience in online panel web surveys. *International Journal of Market Research*, 55(1), 59–80.
- Kiernan, N. E., Kiernan, M., Oyler, M. A., & Gilles, C. (2005). Is a web survey as effective as a mail survey? A field experiment among computer users. *American Journal of Evaluation*, 26(2), 245–252.
- Kinesis. (2013a). Online survey platform conversion: Processes, issues and projected costs. Austin, Texas: Kinesis Survey Technologies. Retrieved from http://www.kinesisurvey.com/wp-content/uploads/2013/09/OnlineSurveyPlatformConversion_KinesisWhitepaper.pdf.
- Kinesis. (2013b). Online survey statistics for the mobile future. Austin, Texas: Kinesis Survey Technologies. Retrieved from <http://www.kinesisurvey.com/2013/10/09/latest-mobile-survey-traffic-stats-q3-2013/>.

- Kirakowski, J., Claridge, N., & Whitehand, R. (1998). Human centered measures of success in web site design. In *Proceedings of the 4th Conference on Human Factors & the Web*. Basking Ridge, NJ.
- Kirakowski, J., & Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3), 210–212.
- Kirk, R. E. (2007). *Statistics: An introduction* (5th ed.). Belmont, California: Cengage Learning.
- Klausch, L. T., de Leeuw, E. D., Hox, J. J., de Jongh, A., & Roberts, A. (2012). *Matrix vs single question formats in web surveys: Results from a large scale experiment*. Presented at the General Online Research Conference (GOR), Mannheim, Germany.
- Klausch, L. T., Hox, J., & Schouten, B. (2014). Evaluating mixed-mode redesign strategies against benchmark surveys: The case of the Crime Victimization Survey. Presented at the General Online Research Conference (GOR), Cologne, Germany.
- Klein, J. D., Havens, C. G., & Thomas, R. K. (2009). Comparing adolescent response bias between internet and telephone surveys. *Journal of Adolescent Health*, 44(2), S36–S36.
- Klofstad, C. A., Boulianne, S., & Basson, D. (2008). Matching the message to the medium: Results from an experiment on internet survey email contacts. *Social Science Computer Review*, 26(4), 498–509.
- Knowledge Networks. (2011). Knowledgepanel design summary. Retrieved from <http://www.knowledgenetworks.com/knpanel/KNPanel-Design-Summary.html>.
- Koren, G., & Hlebec, V. (2006). Web application of Antonucci's hierarchical approach for measuring social networks. Presented at the SMABS-EAM conference, Budapest, Hungary.
- Körner, T., & van der Valk, J. (2011). Mixing modes in the LFS: Computer-assisted, cost effective and respondent friendly. Presented at the 97th DGINS conference, Wiesbaden, Germany.
- Krasilovsky, P. (1996). Surveys in cyberspace. *American Demographics: Tools Supplement*, 11(Nov–Dec), 18–22.
- Kreuter, F. (2013). Improving surveys with paradata: Introduction. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic use of process information* (pp. 1–11). Hoboken, New Jersey: Wiley.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567.
- Krosnick, J. A. (2000). The threat of satisficing in surveys: The shortcuts respondents take in answering questions. *Survey Methods Newsletter*, 20(1).
- Krosnick, J. A. (2013). *Lecture at Web Surveys Day 2013, Faculty of Social Sciences*. Ljubljana, Slovenia.
- Krosnick, J. A., Ackermann, A., Malka, A., Sakshaug, J., Tourangeau, R., de Bell, M., & Turakhia, C. (2009). *Creating the face-to-face recruited internet survey platform (FFRISP)*. Presented at the Third Annual Workshop on Measurement and Experimentation with Internet Panels, Santpoort, The Netherlands.
- Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Michael Hanemann, W., Kopp, R. J., ... Conaway, M. (2002). The impact of 'no opinion' response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66(3), 371–403.
- Krosnick, J. A., MacInnis, B., Suh, A., & Yeager, D. S. (2013). The accuracy of survey data collected by various American on-line survey firms: A 2012 comparison. Presented at the MESS Workshop 2013, Scheveningen. The Netherlands, Amsterdam: CenterData.
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–313). Howard House, UK: Emerald Group Publishing.
- Kroth, P. J., McPherson, L., Leverence, R., Pace, W., Daniels, E., Rhyne, R. L., ... Prime Net

- Consortium. (2009). Combining web-based and mail surveys improves response rates: A PBRN study from PRIME Net. *Annals of Family Medicine*, 7(3), 245–248.
- Krotki, K., & Dennis, J. M. (2001). *Probability-based survey research on the Internet*. Presented at the 53rd Conference of the International Statistical Institute, Seoul, Korea.
- Kruse, Y., Callegaro, M., Dennis, J. M., DiSogra, C., Subias, T., Lawrence, M., & Tompson, T. (2010). Panel conditioning and attrition in the AP-Yahoo news election panel study. *Proceedings of the Joint Statistical Meeting, American Association for Public Opinion Research Conference* (pp. 5742–5756). Washington, DC: AMSTAT.
- Krysan, M., & Couper, M. P. (2006). Race of interviewer effects: What happens on the Web? *International Journal of Internet Science*, 1(1), 17–28.
- Kunz, T., & Fuchs, M. (2013). Reducing response order effects in check-all-that-apply questions by use of dynamic tooltip instructions. Presented at the General Online Research Conference (GOR), Mannheim, Germany.
- Kwak, N., & Radler, B. (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics*, 18(2), 257–273.
- Kypri, K., Samaranayaka, A., Connor, J., Langley, J. D., & MacLennan, B. (2011). Nonresponse bias in a web-based health behaviour survey of New Zealand tertiary students. *Preventive Medicine*, 53(4–5), 274–277.
- Kypri, K., Stephenson, S., & Langley, J. (2004). Assessment of nonresponse bias in an internet survey of alcohol use. *Alcoholism, Clinical and Experimental Research*, 28(4), 630–634.
- Lackaff, D. (2012). New opportunities in personal network data collection. In M. Zacarias & J. V. de Oliveira (Eds), *Human-computer interaction: The agency perspective* (pp. 389–407). Berlin: Springer.
- Lai, V. S., Wong, B. K., & Cheung, W. (2002). Group decision making in a multiple criteria environment: A case using the AHP in software selection. *European Journal of Operational Research*, 137(1), 134–144.
- Langer, G. (2013). Comment on summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1(2), 130–136.
- Lavrakas, P. J., Tompson, T. N., & Benford, R. (2010). Investigating data quality in cell phone surveying. Presented at The American Association for Public Opinion Research (AAPOR) 65th Annual Conference, Chicago, Illinois.
- LeDuff Collins, M. (2009). *The thin book of 360 feedback: A manager's guide*. Plano, Texas: Thin Book Publishing.
- Lee, S. (2006). An evaluation of nonresponse and coverage errors in a prerecruited probability web panel survey. *Social Science Computer Review*, 24(4), 460–475.
- Legleye, S., & Lesnard, L. (2013, May 19). ELIPSS: Étude Longitudinale par Internet Pour les Sciences Sociales. Retrieved from <http://www.cinis.fr>.
- Lenzner, T. (2012). An evaluation of two non-reactive web questionnaire pretesting methods. Presented at the General Online Research Conference (GOR), Mannheim, Germany.
- Lesnard, L. (2011). ELIPSS: A new mobile web panel for social scientists. Presented at the MESS Workshop, Amsterdam, Netherlands.
- Ličen, S., Lozar Manfreda, K., & Hlebec, V. (2006). The quality of survey questions published on Slovenian journalistic web sites. *Advances in methodology and statistics (Metodološki Zvezki)* 3(2), 355–368.
- Lindhjem, H., & Navrud, S. (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics*, 70(9), 1628–1637.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.
- Lobe, B., & Vehovar, V. (2009). Towards a flexible online mixed method design with a feedback loop. *Quality & Quantity*, 43(4), 585–597.
- Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Boston, Massachusetts: Brooks/Cole.

- Lord, T. H., & Miller, L. C. (2009). Playing the game by the rules: A practical guide to sweepstakes and contest promotions. *Francise Law Journal*, 29(1).
- Lorenc, B., Biemer, P. P., Jansson, I., Eltinge, J. L., & Holmberg, A. (2013). Prelude to the special issue on systems and architectures for high-quality statistics production. *Journal of Official Statistics*, 29(1), 1–4.
- Losch, M. E. (2008). Informed consent. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*, pp. 335–337. Thousand Oaks, California: SAGE.
- Lozar Manfreda, K. (2001). *Web survey errors*. PhD thesis. Ljubljana: University of Ljubljana, Faculty of Social Sciences.
- Lozar Manfreda, K., Batagelj, Z., & Vehovar, V. (2002). Design of web survey questionnaires: Three basic experiments. *Journal of Computer Mediated Communication*, 7(3).
- Lozar Manfreda, K., Berzelak, N., & Vehovar, V. (2012). *Paradata insight into survey response behaviour: An analysis of a set of hosted web surveys*. Presented at the General Online Research Conference (GOR), Mannheim, Germany.
- Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50(1), 79–104.
- Lozar Manfreda, K., Couper, M. P., Vohar, M., Rivas, S., & Vehovar, V. (2002). Virtual selves and web surveys. *Advances in Methodology and Statistics*, 18, 187–213.
- Lozar Manfreda, K., & Vehovar, V. (2002). *Survey design features influencing response rates in web surveys*. Presented at the International Conference on Improving Surveys, Copenhagen, Denmark.
- Lozar Manfreda, K., & Vehovar, V. (2008). Internet surveys. In E. De Leeuw, J. J. Hox, & D. A. Dillman (Eds), *International handbook of survey methodology* (pp. 264–284). New York: Lawrence Erlbaum.
- Lozar Manfreda, K., Vehovar, V., & Hlebec, V. (2004). Collecting ego-centred network data via the web. *Advances in Methodology and Statistics (Metodološki Zvezki)* 1(2), 295–321.
- Lugtig, P. J., Das, M., & Scherpenzeel, A. (2014). Nonresponse and attrition in a probability-based online panel of the general population. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 135–153). Chichester: Wiley.
- Lumsden, J. (2007). Online-questionnaire design guidelines. In R. A. Reynolds, R. Woods, & J. D. Baker (Eds), *Handbook of Research on Electronic Surveys and Measurements*, (pp. 44–64). Hershey: Idea Group Reference.
- Lyberg, L. E. (2012). Survey quality. *Survey Methodology*, 38(2), 107–130.
- Lyberg, L. E., Biemer, P. P., Collins, M., Leeuw, E. de, Dippo, C., Schwarz, N., & Trewin, D. (Eds). (1997). *Survey measurement and process quality*. New York: Wiley.
- Macer, T. (2011). Making it fit: How survey technology providers are responding to the challenges of handling web surveys on mobile devices. In D. F. Birks, R. Banks, L. Gerrad, A. Johnson, R. Khan, T. Macer, ... P. Willis (Eds), *Shifting the boundaries of research* (pp. 1–23). London: Association for Survey Computing.
- Macer, T. (2012). Developments and the impact of smart technology. *International Journal of Market Research*, 54(4), 567–570.
- Macer, T. (2014). Online panel software. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, P. J. Lavrakas, & J. A. Krosnick (Eds), *Online panel research: A data quality perspective* (pp. 413–440). Chichester, United Kingdom: Wiley.
- Macer, T., & Wilson, S. (2013). *The Confinity annual market research software survey, 2012*. United Kingdom: Meaning Ltd.
- Macer, T., & Wilson, S. (2014). *The Confinity annual market research software survey, 2013*. United Kingdom: Meaning Ltd.
- Madans, J., Miller, K., Maitland, A., & Willis, G. B. (2011). *Question evaluation methods*:

- Contributing to the science of data quality*. Hoboken, New Jersey: Wiley.
- Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5), 914–934.
- Malhotra, N., Krosnick, J. A., & Thomas, R. K. (2009). Optimal design of branching questions to measure bipolar constructs. *Public Opinion Quarterly*, 73(2), 304–324.
- Mangunkusumo, R. T., Duisterhout, J. S., de Graaff, N., Maarsingh, E. J., de Koning, H. J., & Raat, H. (2006). Internet versus paper mode of health and health behavior questionnaires in elementary schools: Asthma and fruit as examples. *Journal of School Health*, 76(2), 80–86.
- Manisera, M., & Zuccolotto, P. (2014). Modelling ‘don’t know’ responses in rating scales. *Pattern Recognition Letters*, 45(August), 226–234.
- Marcell, M., & Falls, A. (2001). Online data collection with special populations over the World Wide Web. *Down Syndrome Research and Practice*, 7(3), 106–123.
- Martin, C. (2004). The impact of topic interest on mail survey response behaviour. *International Journal of Market Research*, 36(4), 327–338.
- Martin, E. (2004). Vignettes and respondent debriefing for questionnaire design and evaluation. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds), *Methods for testing and evaluating survey questionnaires* (pp. 149–172). Hoboken, New Jersey: Wiley.
- Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, 31(6), 725–743.
- Mavletova, A., & Couper, M. P. (2013). Sensitive topics in PC web and mobile web surveys: Is there a difference? *Survey Research Methods*, 7(3), 191–205.
- Mavletova, A., & Couper, M. P. (2014). *Mobile web survey design: Scrolling versus paging, SMS versus e-mail*. Presented at the General Online Research Conference (GOR), Cologne, Germany.
- McCabe, S. E. (2004). Comparison of web and mail surveys in collecting illicit drug use data: A randomized experiment. *Journal of Drug Education*, 34(1), 61–72.
- McCabe, S. E., Boyd, C. J., Couper, M. P., Crawford, S., & D’Arcy, H. (2002). Mode effects for collecting alcohol and other drug use data: Web and US mail. *Journal of Studies on Alcohol*, 63(6), 755–761.
- McClain, C. M., Crawford, S. D., & Dugan, J. P. (2012). *Use of mobile devices to access computer-optimized web surveys: Implications for respondent behavior and data quality*. Presented at the The American Association for Public Opinion Research (AAPOR) 67th Annual Conference, Orlando, Florida.
- McCloskey, M. (2014). Selecting an online tool for unmoderated remote user testing. Retrieved from http://www.nngroup.com/articles/unmoderated-user-testing-tools/?utm_source=Alertbox&utm_campaign=cd82fc86c0-Alertbox_email_06_02_2014&utm_medium=email&utm_term=0_7f29a2b335-cd-82fc86c0-24345537.
- McMahon, L., & Stamp, R. (2009). *Questionnaire intelligence: New rules of engagement for online survey design*. Presented at the Worldwide Readership Research Symposia, Valencia, Spain.
- McMorris, B. J., Petrie, R. S., Catalano, R. F., Fleming, C. B., Haggerty, K. P., & Abbott, R. D. (2009). Use of web and in-person survey modes to gather data from young adults on sex and drug use: An evaluation of cost, time, and survey error based on a randomized mixed-mode design. *Evaluation Review*, 33(2), 138–158.
- McNaughton, R. B. (1999). Disk by mail for industrial survey research: A review and example. *Industrial Marketing Management*, 28(3), 293–304.
- Medway, R. L., & Fulton, J. (2012). When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly*, 76(4), 733–746.
- Mesh, G. (2012). Email surveys. In L. Gideon (Ed.), *Handbook of survey methodology for the social sciences* (pp. 313–325). New York: Springer.
- Microsoft Corporation. (2013). Microsoft security intelligence report, Volume 15. Key findings and

- summary. Redmond, Washington.
- Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, 75(2), 249–269.
- Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., & Alan, G. (2002). Test-retest reliability of alcohol measures: Is there a difference between Internet-based assessment and traditional methods? *Psychology of Addictive Behaviors*, 16(1), 56–63.
- Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (2014). *Cognitive interviewing methodology*. New York: Wiley.
- Miller Steiger, D., & Conroy, B. (2008). IVR: Interactive voice technology. In E. de Leeuw, J. Hox, & D. A. Dillman (Eds), *International handbook of survey methodology* (pp. 285–298). New York: Lawrence Erlbaum.
- Mitra, A., Jain-Shukla, P., Robbins, A., Champion, H., & Durant, R. (2008). Differences in rate of response to web-based surveys among college students. *International Journal on E-Learning*, 7(2), 265–281.
- Mohler, P. P., & Uher, R. (2003). Documenting comparative surveys for secondary analysis. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds), *Cross-cultural survey methods* (pp. 311–327). Hoboken, New Jersey: Wiley.
- Morgan, D. L. (2013). *Integrating qualitative and quantitative methods: A pragmatic approach*. London: SAGE.
- Morrison, R. L., Dillman, D. A., & Christian, L. M. (2010). Questionnaire design guidelines for establishment surveys. *Journal of Official Statistics*, 26(1), 43–85.
- MRIA. (2013). *Improving the quality of your online research: Ten questions to ask your online survey provider*. Toronto, Ontario.
- Mueller, K., Straatmann, T., Hattrup, K., & Jochum, M. (2012). Effects of personalized versus generic implementation of an intra-organizational online survey on psychological anonymity and response behavior: A field experiment. *Journal of Business and Psychology*, 1–13.
- Muñoz-Leiva, F., Sánchez-Fernández, J., Montoro-Ríos, F., & Ibáñez-Zapata, J. Á. (2010). Improving the response rate and quality in web-based surveys through the personalization and frequency of reminder mailings. *Quality & Quantity*, 44(5), 1037–1052.
- Murphy, J., Hamel, L., Harrison, C., & Hammer, H. (2011). AAPOR 2011 membership survey results and methodology. Deerfield, Illinois.
- Musch, J., & Reips, U.-D. (2000). A brief history of web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet*, (pp. 61–87). San Diego: Academic Press.
- Nadler, R., & Henning, J. (1998). *Web surveys: For knowledge, lead management, and increased traffic*. New York: Perseus.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59(3), 150–162.
- Nallan, S. (2012, April 5). Benchmarking for better surveys. Retrieved from <http://blog.surveymonkey.com/blog/2012/04/05/benchmarking-for-better-surveys/>.
- Napoli, P. M., Lavrakas, P. J., & Callegaro, M. (2014). Internet and mobile ratings panels. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 387–407). Chichester: Wiley.
- Neubarth, W. (2010). Drag & drop: A flexible method for moving objects, implementing rankings, and a wide range of other applications. In S. D. Gosling & J. A. Johnson (Eds), *Advanced methods for conducting online behavioral research* (pp. 63–74). Washington, DC: American Psychological Association.
- Neuman, W. L. (2009). *Social research methods: qualitative and quantitative approaches* (7th ed.). Boston: Pearson.
- Nicolaas, G. (2011). Survey paradata: A review. National Centre for Research Methods: Southampton. Retrieved from <http://eprints.ncrm.ac.uk/1719/>.

- Nicolaas, G. (2013). A probability-based web panel for the UK: What would it look like? Presented at the NCRM Conference: Web surveys for the general population – How, why and when?, London, United Kingdom. Retrieved from <http://www.natcenweb.co.uk/genpopweb/documents/UK-prob-based-web-panel-Gerry-Nicolaas.ppt>.
- Nielsen, J. (1993). *Usability engineering*. Burlington, MA: Morgan Kaufmann.
- Nielsen, J. (1997, October 1). How users read on the web. Retrieved from <http://www.nngroup.com/articles/how-users-read-on-the-web/>.
- Nielsen, J. (2000a). *Designing web usability: The practice of simplicity*. Indianapolis: New Riders Press.
- Nielsen, J. (2000b). Why you only need to test with 5 users. Retrieved from <http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>.
- Nielsen, J. (2006, April 17). F-shaped pattern for reading web content. Retrieved from <http://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/>.
- Nielsen, J. (2012). Usability 101: Introduction to usability. Retrieved from <http://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
- Nielsen, J., & Budiu, R. (2013). *Mobile usability*. Berkeley, CA: New Riders Press.
- Nielsen, J., & Pernice, K. (2009). *Eyetracking web usability*. Berkeley, CA: New Riders Press.
- Nielsen, J. S., & Kjær, T. (2011). Does question order influence sensitivity to scope? Empirical findings from a web-based contingent valuation study. *Journal of Environmental Planning and Management*, 54(3), 369–381.
- Nigg, C. R., Motl, R. W., Wong, K. T., Yoda, L. U., McCurdy, D. K., Paxton, R., ... Dishman, R. K. (2009). Impact of mixed survey modes on physical activity and fruit/vegetable consumption: A longitudinal study. *Survey Research Methods*, 3(2), 81–90.
- Nunan, D., & Knox, S. (2011). Can search engine advertising help access rare samples? *International Journal of Market Research*, 53(4), 523–540.
- O'Brien, H. L., & Toms, E. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1), 50–69.
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938–955.
- OECD. (2002). *Frascati manual: Proposed standard practice for surveys on research and experimental development* (6th ed.). Paris: OECD. Retrieved from <http://www.oecd.org/science/inno/frascatimanualproposedstandardpracticeforsurveysonresearchandexperimentaldevelopment6thedition.htm>.
- Olson, K., & Parkhurst, B. (2013). Collecting paradata for measurement error evaluation. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic use of process information* (pp. 43–72). Hoboken, New Jersey: Wiley.
- Olson, K., Smyth, J. D., Wang, Y., & Pearson, J. E. (2011). The self-assessed literacy index: Reliability and validity. *Social Science Research*, 40(5), 1465–1476.
- Ongena, Y. P., & Dijkstra, W. (2006). Methods of behavior coding of survey interviews. *Journal of Official Statistics*, 22(3), 419–451.
- ORC Macro. (2003). Report of the results of the asthma awareness survey. Retrieved from http://www.nasn.org/Portals/0/resources/asthma_survey.pdf.
- Osborne, J. W. (2012). *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Thousand Oaks, California: SAGE.
- Oudejans, M., & Christian, L. M. (2010). Using interactive features to motivate and probe responses to openended questions. In M. Das, P. Ester & L. Kaczmarek (Eds), (pp. 215–244). *Social and behavioral research and the internet*. New York: Routledge.
- Owen, R. (2013). Survey optimisation considerations for Android, Apple and Windows 8 mobile devices. Presented at the Market Research Society (MRS) mobile research conference, London.

- Parsons, N. L., & Manierre, M. J. (2013). Investigating the relationship among prepaid token incentives, response rates, and nonresponse bias in a web survey. *Field Methods*, 26(2), 191–204.
- Pasek, J., & Krosnick, J. A. (2010). Measuring intent to participate and participation in the 2010 census and their correlates and trends: Comparisons of RDD telephone and non-probability sample internet survey data. Statistical Research Division Study Series, Survey Methodology #2010(15). US Census Bureau. Retrieved from <http://www.census.gov/srd/papers/pdf/ssm2010-15.pdf>.
- Payne, J., & Barnfather, N. (2012). Online data collection in developing nations: An investigation into sample bias in a sample of South African university students. *Social Science Computer Review*, 30(3), 389–397. doi:10.1177/0894439311407419
- Peterson, G. (2012). What we can learn from unintentional mobile respondents. *CASRO Journal*, 2012–2013, 32–35.
- Petric, I., & Appel, M. (2007). *The readership currency: Dutch design. How a new methodology for AIR measurement opens up new perspectives for the print advertisers and publishers*. Presented at the Worldwide Readership Research Symposium, Vienna, Austria. Retrieved from <http://www.nommedia.nl/upload/docu%20menten/2007-wrrs-petric-appel.pdf>.
- Petrovčič, A., Lozar Manfreda, K., & Petrič, G. (2013). The effect of email invitation characteristics and response reluctance on nonresponse in web forum surveys. Presented at the European Survey Research Association (ESRA), Ljubljana, Slovenia.
- Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, 73(1), 74–97.
- Peytchev, A. (2011). Breakoff and unit nonresponse across web surveys. *Journal of Official Statistics*, 27(1), 33–47.
- Peytchev, A. (2012). Multiple imputation for unit nonresponse and measurement error. *Public Opinion Quarterly*, 76(2), 214–237.
- Peytchev, A., Conrad, F. G., Couper, M. P., & Tourangeau, R. (2010). Increasing respondents' use of definitions in web surveys. *Journal of Official Statistics*, 26(4), 633–650.
- Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web survey design: Paging versus scrolling. *Public Opinion Quarterly*, 70(4), 596–607.
- Peytchev, A., & Crawford, S. D. (2005). A typology of real-time validations in web-based surveys. *Social Science Computer Review*, 23(2), 235–249.
- Peytchev, A., & Hill, C. A. (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. *Social Science Computer Review*, 28(3), 319–335.
- Pferdekaemper, T., & Batanic, B. (2009). Mobile surveys from a technological perspective. In E. Maxl, N. Döring, & A. Wallisch (Eds), *Mobile market research* (pp. 116–133). Cologne: Herbert von Halem.
- Pit, S. W., Vo, T., & Pyakurel, S. (2014). The effectiveness of recruitment strategies on general practitioner's survey response rates: A systematic review. *BMC Medical Research Methodology*, 14(1), 76.
- Pitkow, J., & Recker, M. (1994). Results from the first World Wide Web user survey. *Computer Networks and ISDN Systems*, 27(2), 243–254.
- Półtorak, M., & Kowalski, J. (2013). *Pros and cons of virtual interviewers: Vote in the discussion about surveytainment*, Mannheim, Germany.
- Porter, S. R. (2008). Email survey. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*, pp. 231–33. Thousand Oaks, CA: SAGE.
- Porter, S. R., & Whitcomb, M. E. (2003a). The impact of contact type on web survey response rates. *Public Opinion Quarterly*, 67, 579–588.
- Porter, S. R., & Whitcomb, M. E. (2003b). The impact of lottery incentives on student survey response rates. *Research in Higher Education*, 44(4), 389–407.
- Porter, S. R., & Whitcomb, M. E. (2005). E-mail subject lines and their effect on web survey viewing and response. *Social Science Computer Review*, 23(3), 380–387.

- Porter, S. R., & Whitcomb, M. E. (2007). Mixed-mode contacts in web surveys: Paper is not necessarily better. *Public Opinion Quarterly*, 71(4), 635–648.
- Postoaca, A. (2006). *The anonymous elect: Market research through online access panels*. Berlin: Springer.
- Potaka, L. (2008). Comparability and usability: Key issues in the design of internet forms for New Zealand's 2006 Census of populations and dwellings. *Survey Research Methods*, 2(1), 1–10.
- Poynter, R. (2010). *The handbook of online and social media research: Tools and techniques for market researchers*. Chichester: Wiley.
- Poynter, R., Williams, N., & York, S. (2014). *The handbook of mobile market research: Tools and techniques for market researchers*. Chichester: Wiley.
- Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behaviour and Information Technology*, 20(5), 347–356.
- Preece, J., & Maloney-Krichmar, D. (2005). Online communities: Design, theory, and practice. *Journal of Computer-Mediated Communication*, 10(4).
- Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, J., & Singer, E. (Eds). (2004). *Methods for testing and evaluating survey questionnaires*. Hoboken, New Jersey: Wiley.
- Puleston, J. (2011). Online research – game on!: A look at how gaming techniques can transform your online research. In D. F. Birks, R. Banks, L. Gerrad, A. Johnson, R. Khan, T. Macer, ... P. Willis (Eds), *Shifting the boundaries of research* (pp. 20–50). London: Association for Survey Computing.
- Puleston, J. (2012). Gamification 101: From theory to practice part I. *Quirk's E-Newsletter*, January.
- Puleston, J. (2013). *Online research, game on!* Presented at the NCRM Conference: Web surveys for the general population – How, why and when?, London, United Kingdom. Retrieved from <http://www.natcenweb.co.uk/genpopweb/documents/Jon-Puleston.pdf>.
- Puleston, J., & Rintoul, D. (2012). *Can survey gaming techniques cross continents?* Examining cross cultural reactions to creative questioning techniques. Presented at the Asia Pacific 2012, Shanghai: Esomar.
- Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, 90(429), 54–63.
- Ramos, M., Sedivi, B. M., & Sweet, E. M. (1998). Computerized self-administered questionnaires. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds), *Computer assisted survey information collection* (pp. 389–408). New York: Wiley.
- Rao, K., Kaminska, O., & McCutcheon, A. L. (2010). Recruiting probability samples for a multi-mode research panel with Internet and mail components. *Public Opinion Quarterly*, 74(1), 68–84.
- Rao, K., & Pennington, J. (2013). Should the third reminder be sent? The role of survey response timing on web survey results. *International Journal of Market Research*, 55(5).
- Rat der Deutschen Markt- und Sozialforschung e.V. (2007). *Richtlinie für Online-Befragungen (Directive for Online Surveys)*.
- Reips, U.-D. (2002a). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4), 243–256.
- Reips, U.-D. (2002b). Theory and techniques of web experiments. In B. Batanic, U.-D. Reips, & M. Bosnjak (Eds), *Online social sciences* (pp. 229–250). Seattle: Hogrefe & Huber.
- Reips, U.-D. (2007). The methodology of internet-based experiments. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds), *The Oxford handbook of internet psychology* (pp. 373–90). Oxford: Oxford University Press.
- Reips, U.-D., & Birnbaum, M. H. (2011). Behavioral research and data collection via the Internet. In K.-P. L. Vu & R. W. Proctor (Eds), *Handbook of human factors in web designs* (2nd ed., pp. 563–585). Boca Raton, FL: Taylor and Francis.
- Reips, U.-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in

- Internet-based research: VAS Generator. *Behavior Research Methods*, 40(3), 699–704.
- Reips, U.-D., & Krantz, J. H. (2010). Conducting true experiments on the Web. In S. D. Gosling & J. A. Johnson (Eds), *Advanced methods for conducting online behavioral research* (pp. 193–216). Washington, DC: American Psychological Association.
- Reips, U.-D., & Lengler, R. (2005). The web experiment list: A web service for the recruitment of participants and archiving of Internet-based experiments. *Behavior Research Methods*, 37(2), 287–292.
- Reja, U., Lozar Manfreda, K., Hlebec, V., & Vehovar, V. (2003). Open-ended vs close-ended questions in web questionnaires. *Advances in Methodology and Statistics (Metodološki Zvezki)*, 19, 159–177.
- Revilla, M. (2013). Measurement invariance and quality of composite scores in a face-to-face and a web survey. *Survey Research Methods*, 7(1), 17–28.
- Rivara, F. P., Koepsell, T. D., Wang, J., Durbin, D., Jaffe, K. M., Vavilala, M., ... Temkin, N. (2011). Comparison of telephone with world wide web-based responses by parents and teens to a follow-up survey after injury. *Health Services Research*, 46(3), 964–981.
- Rivers, D. (2007). Sampling for web surveys. *Proceedings of the Joint Statistical Meeting, Survey Research Methods Section*. Salt Lake City: AMSTAT.
- Rivers, D. (2010). *AAPOR report on online panels*. Presented at the The American Association for Public Opinion Research (AAPOR) 68th Annual Conference, Chicago, Illinois and at the PAPOR 2010 Annual Conference, San Francisco.
- Rivers, D., & Bailey, D. (2009). *Inference from matched samples in the 2008 US national elections*. Presented at the The American Association for Public Opinion Research (AAPOR) 64th Annual Conference, Hollywood, Florida.
- Roberts, C. (2013). *Participation and engagement in web surveys of the general population: An overview of challenges and opportunities*. Presented at the NCRM Conference: Web surveys for the general population – How, why and when?, London, United Kingdom.
- Roberts, C., Vandenplas, C., & Stähli, M. E. (2014). Evaluating the impact of response enhancement methods on the risk of nonresponse bias and survey costs. *Survey Research Methods*, 8(2), 67–80.
- Roe, D. J. (2008). Usability testing. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*, pp. 933–934. Thousand Oaks, CA: SAGE.
- Roman, E. (2011). *Voice-of-the-customer marketing: A revolutionary 5-step process to create customers who care, spend, and stay*. New York: McGraw-Hill.
- Rookey, B. D., Hanway, S., & Dillman, D. A. (2008). Does a probability-based household panel benefit from assignment to postal response as an alternative to Internet-only? *Public Opinion Quarterly*, 72(5), 962–984.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Ross, M. W., Månsson, S.-A., Daneback, K., Cooper, A., & Tikkanen, R. (2005). Biases in internet sexual health samples: Comparison of an internet sexuality survey and a national sexual health survey in Sweden. *Social Science & Medicine*, 61(1), 245–252.
- Roster, C. A., Rogers, R. D., Albaum, G., & Klein, D. (2004). A comparison of response characteristics from web and telephone surveys. *International Journal of Market Research*, 46(3), 359–373.
- Rubin, D. B., Stern, H. S., & Vehovar, V. (1995). Handling ‘don’t know’ survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association*, 90(431), 822–828.
- Sackett, P. R. (2009). From the editor. *Industrial and Organizational Psychology*, 2(1), 1.
- Sánchez-Fernández, J., Muñoz-Leiva, F., & Montoro-Ríos, F. J. (2012). Improving retention rate and response quality in web-based surveys. *Computers in Human Behavior*, 28(2), 507–514.
- Saris, W. E. (1991). *Computer-assisted interviewing*. Newbury Park, CA: SAGE.
- Saris, W. E. (1998). Ten years of interviewing without interviewers: The telepanel. In M. P. Couper,

- R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II, & J. M. O'Reilly (Eds), *Computer assisted survey information collection* (pp. 409–429). New York: Wiley.
- Saris, W. E. (2012). Discussion evaluation procedures for survey questions. *Journal of Official Statistics*, 28(4), 537–551.
- Saris, W. E., & Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, 1(1), 29–43.
- Saris, W. E., & Gallhofer, I. (2014). *Design, evaluation, and analysis of questionnaires for survey research* (2nd ed.). Hoboken, New Jersey: Wiley.
- Särndal, C.-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. Chichester: Wiley.
- Sauermann, H., & Roach, M. (2013). Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Research Policy*, 42(1), 273–286.
- Schaeffer, D. R., & Dillman, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly*, 62(3), 378–397.
- Scherpenzeel, A. (2011). Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 109(1), 56–61.
- Scherpenzeel, A., & Das, M. (2010). True longitudinal and probability-based Internet panels: Evidence from the Netherlands. In *Social and behavioral research and the internet: Advances in applied methods and research strategies* (pp. 77–104). New York: Routledge.
- Scherpenzeel, A., & Toepoel, V. (2012). Recruiting a probability sample for an online panel: Effects of contact mode, incentives, and information. *Public Opinion Quarterly*, 76(3), 470–490.
- Scherpenzeel, A., & Toepoel, V. (2014). Informing panel members about study results. In M. Das, P. Ester, & L. Kaczmirek (Eds), *Social and behavioral research and the Internet. Advances in applied methods and research strategies* (pp. 77–104). New York: Routledge.
- Schmidt, S., & Wenzel, O. (2013). *Mobile research performance: How mobile respondents differ from PC users concerning interview quality, drop-out rates and sample structure*. Presented at the General Online Research conference (GOR), Mannheim, Germany.
- Schonlau, M., Fricker, R. D., & Elliott, M. N. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica, California: RAND.
- Schonlau, M., & Liebau, E. (2012). Respondent driven sampling. *The Stata Journal*, 12(1), 72–93.
- Schonlau, M., Zapert, K., Payne Simon, L., Hayness Sanstand, K., Marcus, S. M., Adams, J., ... Berry, S. H. (2004). A comparison between responses from a propensity-weighted web survey and an identical RDD survey. *Social Science Computer Review*, 22(1), 128–138.
- Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey Methodology*, 39(1), 29–58.
- Schrivver, K. A. (1997). *Dynamics in document design: Creating text for readers*. New York, USA: Wiley.
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21(2), 277–287.
- Shannon, D. M., & Bradshaw, C. C. (2002). A comparison of response rate, response time, and costs of mail and electronic surveys. *Journal of Experimental Education*, 70(2), 179–192.
- Sheehan, K. B. (2002). Toward a typology of Internet users and online privacy concerns. *Information Society*, 18(1), 21–32.
- Sheehan, K. B., & Hoy, M. G. (1999). Using e-mail to survey Internet users in the United States: Methodology and assessment. *Journal of Computer Mediated Communication*, 4(3).
- Shih, T.-H., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods*, 20(3), 249–271.
- Shinn, G., Baker, M., & Briers, G. (2007). Response patterns: Effect of day of receipt of an e-mailed

- survey instrument on response rate, response time, and response quality. *Journal of Extension*, 45(2).
- Shneiderman, B. (1984). Response time and display rate in human performance with computers. *Computing Surveys*, 16(3), 265–285.
- Sikkel, D., & Hoogendoorn, A. (2008). Panel surveys. In E. de Leeuw, J. Hox, & D. A. Dillman (Eds), *International handbook of survey methodology* (pp. 479–499). New York: Lawrence Erlbaum Associates.
- Silver, N. (2012, November 10). Which polls fared best (and worst) in the 2012 presidential race? Retrieved from <http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/>.
- Siminski, P. (2008). Order effects in batteries of questions. *Quality & Quantity*, 42(4), 477–490.
- Sinclair, M., O'Toole, J., Malawaraarachchi, M., & Leder, K. (2012). Comparison of response rates and cost-effectiveness for a community-based survey: Postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC Medical Research Methodology*, 12(1), 1–8.
- Singer, E. (2006). Introduction nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 637–645.
- Singer, E., van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, 15(2), 217–230.
- Sleep, D., & Puleston, J. (2011). *The game experiments: Researching how gaming techniques can be used to improve the quality of feedback from online research*. Presented at Impact: Research Reloaded, Amsterdam: Esomar.
- Smith, R. M., & Kniorski, K. (2003). *Participation in online surveys: Results from a series of experiments*. Presented at the The American Association for Public Opinion Research (AAPOR) 58th Annual Conference, Nashville, Tennessee.
- Smith, T. W. (2002). Developing nonresponse standards. In R. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds), *Survey nonresponse* (pp. 27–40). New York: Wiley.
- Smith, T. W. (2011). Refining the total survey error perspective. *International Journal of Public Opinion Research*, 23(4), 464–484.
- Smyth, J. D., Christian, L. M., & Dillman, D. A. (2008). Does 'yes or no' on the telephone mean the same as 'check-all-that-apply' on the Web? *Public Opinion Quarterly*, 72(1), 103–113.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys: Can increasing the size of answer-boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, 73(2), 325–337.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & O'Neill, A. C. (2010). Using the Internet to survey small towns and communities: Limitations and possibilities in the early 21st century. *American Behavioral Scientist*, 53(9), 1423–1448.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, 70(1), 66–77.
- Smyth, J. D., & Pearson, J. E. (2011). Internet survey methods: A review of strengths, weaknesses, and innovations. In M. Das, P. Ester, & L. Kaczmarek (Eds), *Social and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 11–44). New York: Routledge.
- Snijders, G., Araldsen, G., Jones, J., & Willimack, D. K. (2013). *Designing and conducting business surveys*. Hoboken, New Jersey: Wiley.
- Snijders, G., & Morren, M. (2010). *Improving the web and electronic questionnaires: The case of audit trails*. Presented at the European Conference on Quality in Official Statistics, Helsinki: Statistics Finland. Retrieved from <https://q2010.stat.fi>.
- Spies, R. A., Carlson, J. F., & Geisinger, K. F. (2010). *The eighteenth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

- Spijkerman, R., Knibbe, R., Knoop, K., Mheen, D. van de, & Eijnden, R. van den. (2009). The utility of online panel surveys versus computer-assisted interviews in obtaining substance-use prevalence estimates in the Netherlands. *Addiction*, 104(10), 1641–1645.
- Stapleton, C. (2011). *The smart(phone) way to collect survey data*. Presented at the The American Association for Public Opinion Research (AAPOR) 66th Annual Conference, Phoenix, Arizona.
- Stark, R., & Gatward, R. (2010). *Computer Assisted Interview Testing Tool (CTT): A review of new features and how the tool has improved the testing process*. Ann Arbor, Michigan: University of Michigan. Retrieved from <http://www.blaiseusers.org/2010/papers/2a.pdf>.
- Steinbrecher, M., Roßmann, J., & Bergmann, M. (2013). *The short-term campaign panel of the German Longitudinal Election Study 2009: Design, implementation, data preparation, and archiving*. Retrieved from http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2013/TechnicalReport_2013-20.pdf.
- Steinmetz, S., Bianchi, A., Tijdens, K. G., & Biffignandi, S. (2014). Improving web survey quality: Potential constraints of propensity score adjustments. In M. Callegaro, R. P. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds), *Online panel research: A data quality perspective* (pp. 273–298). Chichester: Wiley.
- Stephenson, L. B., & Crête, J. (2011). Studying political behavior: A comparison of internet and telephone surveys. *International Journal of Public Opinion Research*, 23(1), 24–55.
- Stettler, K. (2013). Using Web Ex to conduct usability testing of an on-line survey instrument. Presented at the The American Association for Public Opinion Research (AAPOR) 68th Annual Conference, Boston, Massachusetts and at the Quest Workshop 2013, Washington D.C.
- Stewart, D. W., & Shamdasani, P. N. (2014). *Focus group: Theory and practice* (3rd ed.). Thousand Oaks, CA: SAGE.
- Stieger, S., & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, 26(6), 1488–1495.
- Stieger, S., Reips, U.-D., & Voracek, M. (2007). Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science and Technology*, 58(11), 1653–1660.
- Stoop, I. (2005). *Nonresponse in sample surveys: The hunt for the last respondent*. The Hague, Netherlands: Social and Cultural Planning Office.
- Stouthamer-Loeber, M., & van Kammen, W. B. (1995). *Data collection and management: A practical guide*. Newbury Park, CA: SAGE.
- Strabac, Z., & Aalberg, T. (2011). Measuring political knowledge in telephone and web surveys: A crossnational comparison. *Social Science Computer Review*, 29(2), 175–192.
- Sturgis, P., Allum, N., & Smith, P. (2008). An experiment on the measurement of political knowledge in surveys. *Public Opinion Quarterly*, 72(1), 90–102.
- Sue, V. M., & Ritter, L. A. (2012). *Conducting online surveys* (2nd ed.). Thousand Oaks, California: SAGE.
- Survey Monkey. (2011, August 16). *What day of the week should you send your survey?* Retrieved from <https://www.surveymonkey.com/blog/en/blog/2011/08/16/day-of-the-week/>.
- Svensson, J. (2013). *Web panel surveys: Can they be designed and used in a scientifically sound way?* Presented at the 59th ISI World Statistics Congress. Hong Kong, China.
- Swoboda, W. J., Mühlberger, N., Weitkunat, R., & Schneeweiß, S. (1997). Internet surveys by direct mailing: An innovative way of collecting data. *Social Science Computer Review*, 15(3), 242–255.
- Tarnai, J., & Moore, D. L. (2004). Methods for testing and evaluating computer-assisted questionnaires. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds), *Methods for testing and evaluating survey questionnaires* (pp. 319–336). Hoboken, NJ: Wiley.
- Taylor, H., Bremer, J., Overmeyer, C., Siegel, J. W., & Terhanian, G. (2001). The record of internet-based opinion polls in predicting the results of 72 races in the November 2000 US elections.

- International Journal of Market Research*, 43(2), 127–136.
- Terhanian, G., & Bremer, J. (2012). A smarter way to select respondents for surveys? *International Journal of Market Research*, 54(6), 751–780.
- The International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143–171.
- Thiele, O., & Kaczmarek, L. (2010). Security and data protection: Collection, storage, and feedback in internet research. In S. D. Gosling & J. A. Johnson (Eds), *Advanced methods for conducting online behavioral research* (pp. 235–253). Washington, DC: American Psychological Association (APA).
- Thomas, R. K. (2014). Research quality: Fast and furious ... or much ado about nothing? Sub-optimal respondent behavior and data quality. *Journal of Advertising Research*, 54(1), 17–31.
- Thomas, R. K., & Barlas, F. M. (2014). *Respondents playing fast and loose? Antecedents and consequences of respondent speed of completion*. Presented at the The American Association for Public Opinion Research (AAPOR) 69th Annual Conference, Anaheim, California.
- Thomas, R. K., Uldall, B. R., & Krosnick, J. A. (2002). *Reliability and validity of web-based surveys: Effects of response modality, item format and number of categories*. Presented at the The American Association for Public Opinion Research (AAPOR) 57th Annual Conference, Florida, USA.
- Thorndike, F. P., Calbring, P., Smyth, F. L., Magee, J. C., Gonder-Frederick, L., Ost, L.-G., & Ritterbrand, L. M. (2009). Web-based measurement: Effect of completing single or multiple items per webpage. *Computers in Human Behavior*, 25(2), 393–401.
- Tijdens, K. G. (2011). Text string matching to measure occupations in web-surveys. Presented at the Data collection for social surveys using multiple modes workshop. Luxembourg.
- Tijdens, K. G. (2014). Dropout rates and response times of an occupation search tree in a web survey. *Journal of Official Statistics*, 30(1), 23–43.
- Tippins, N. T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology*, 2(1), 2–10.
- Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59(1), 189–225.
- Toepoel, V., & Couper, M. P. (2011). Can verbal instructions counteract visual context effects in web surveys? *Public Opinion Quarterly*, 75(1), 1–18.
- Toepoel, V., Das, M., & van Soest, A. (2008). Effects of design in web surveys: Comparing trained and fresh respondents. *Public Opinion Quarterly*, 72(5), 985–1007.
- Toepoel, V., Das, M., & van Soest, A. (2009). Design of web questionnaires: A test for number of items per screen. *Field Methods*, 21(2), 200–213.
- Toepoel, V., & Lugtig, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of internet users. *Social Science Computer Review*, 32(4), 544–560.
- Tomsic, M. L., Hendel, D. D., & Matross, R. P. (2000). A World Wide Web response to student satisfaction surveys: Comparisons using paper and Internet formats. Presented at the 40th Annual Meeting of the Association for Institutional Research.
- Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual Analog Scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making*, 21, 329–34.
- Tortora, R. (2009). Recruitment and retention for a consumer panel. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 235–249). Hoboken, New Jersey: Wiley.
- Tourangeau, R. (2004). Experimental design considerations for testing and evaluating questionnaires. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds), *Methods for testing and evaluating survey questionnaires* (pp. 209–224). Hoboken, NJ: Wiley.
- Tourangeau, R., & Bradburn, N. M. (2010). The psychology of survey response. In P. V. Marsden & J. C. Wright (Eds), *Handbook of survey research* (2nd ed., pp. 315–346). Howard House, United

- Kingdom: Emerald Group Publishing.
- Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The science of web surveys*. Oxford: Oxford University Press.
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2003). The impact of the visible: Images, spacing, and other visual cues in web surveys. *Statistical Policy Working Paper 36* (pp. 23–44). Washington, DC: Federal Committee on Statistical Methodology.
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, 68(3), 368–393.
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2013). ‘Up means good’: The effect of screen position on evaluative ratings in web surveys. *Public Opinion Quarterly*, 77(S1), 69–88.
- Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: Experiments on social presence in web and IVR surveys. *Computers in Human Behavior*, 19(1), 1–24.
- Tourangeau, R., Groves, R. M., Kennedy, C., & Yan, T. (2009). The presentation of a web survey, nonresponse and measurement error among members of web panel. *Journal of Official Statistics*, 25(3), 299–321.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883.
- Townsend, L. (2012). Flowing with the mainstream: Is mobile market research finally living up to the hype? *Quirk’s Marketing Research Review*, July, 36–41.
- Trouteaud, A. R. (2004). How you ask counts: A test of internet-related components of response rates to a webbased survey. *Social Science Computer Review*, 22(3), 385–92.
- Truell, A. D., Bartlett, J. E., & Alexander, M. W. (2002). Response rate, speed, and completeness: A comparison of Internet-based and mail surveys. *Behavior Research Methods, Instruments, & Computers*, 34(1), 46–49.
- Tuten, T. L. (1997). Electronic methods of collecting survey data: A review of e-research. *ZUMA-Arbeitsbericht*, 9, 1–26.
- Tuten, T. L., Galesic, M., & Bosnjak, M. (2008). Optimizing response rates and data quality in web surveys: The immediacy effect and prize values. In L. O. Petrieff & R. V. Miller (Eds), *Public opinion research focus* (pp. 149–157). Hauppauge, New York: Nova Publisher.
- Valliant, R., Daver, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.
- van Horn, P., Green, K. E., & Martinussen, M. (2009). Survey response rates and survey administration in counseling and clinical psychology: A meta-analysis. *Educational and Psychological Measurement*, 69(3), 389–403.
- Vannieuwenhuyze, J. T. A. (2014). On the relative advantage of mixed-mode versus single-mode surveys. *Survey Research Methods*, 9(1), 31–42.
- Vannieuwenhuyze, J. T. A., & Loosveldt, G. (2013). Evaluating relative mode effects in mixed-mode surveys: Three methods to disentangle selection and measurement effects. *Sociological Methods & Research*, 42(1), 82–104.
- Vannieuwenhuyze, J. T. A., Loosveldt, G., & Molenberghs, G. (2010). A method for evaluating mode effects in mixed-mode surveys. *Public Opinion Quarterly*, 74(5), 1027–1045.
- Vannieuwenhuyze, J. T. A., Loosveldt, G., & Molenberghs, G. (2012). A method to evaluate mode effects on the mean and variance of a continuous variable in mixed-mode surveys. *International Statistical Review*, 80(2), 306–322.
- Vehovar, V. (1999). Field substitution and unit nonresponse. *Journal of Official Statistics*, 15(2), 335–350.
- Vehovar, V., & Batagelj, Z. (1996). *The methodological issues in WWW surveys*. Presented at the International Conference on Computer-Assisted Survey Information Collection, San Antonio,

- Texas.
- Vehovar, V., Batagelj, Z., Lozar Manfreda, K., & Zaletel, M. (2002). Nonresponse in web surveys. In R. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds), *Survey nonresponse* (pp. 229–242). New York: Wiley.
- Vehovar, V., Berzelak, N., & Lozar Manfreda, K. (2010). Mobile phones in an environment of competing survey modes: Applying metric for evaluation of costs and errors. *Social Science Computer Review*, 28(3), 303–318.
- Vehovar, V., Berzelak, N., Lozar Manfreda, K., & Belak, E. (2009). *Optimising survey costs in mixed mode environment*. Presented at the New Techniques and Technologies for Statistics (NTTS) conference Eurostat February.
- Vehovar, V., & Čehovin, G. (2014). *Questionnaire length and breakoffs in web surveys: A meta study*. Presented at the 7th Internet Survey Methodology Workshop. Bolzano, Italy.
- Vehovar, V., Čehovin, G., Kavčič, L., & Lenar, J. (2012). *WebSM study: Survey software features overview*. Ljubljana: University of Ljubljana, Faculty of Social Sciences. Retrieved from <http://www.websm.org/db/12/15753/>.
- Vehovar, V., Čehovin, G., & Močnik, A. (2014). *WebSM study: Survey software in 2014*. Retrieved from <http://www.websm.org/db/12/17467/>.
- Vehovar, V., Koren, G., Lozar Manfreda, K., & Berzelak, J. (2005). *What is important when choosing web survey software?* Presented at the Internet Survey Methodology Workshop. Retrieved from <http://workshop.websm.org/db/14/118/>.
- Vehovar, V., Lozar Manfreda, K., & Batagelj, Z. (1999). Web surveys: Can the weighting solve the problem? *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 962–967. Alexandria, Virginia: AMSTAT.
- Vehovar, V., Lozar Manfreda, K., & Batagelj, Z. (2001). Sensitivity of e-commerce measurement to survey instrument. *International Journal of Electronic Commerce*, 6(1), 31–52.
- Vehovar, V., Lozar Manfreda, K., Koren, G., & Hlebec, V. (2008). Measuring ego-centered social networks on the web: Questionnaire design issues. *Social Networks*, 30(3), 213–222.
- Vehovar, V., Motl, A., Mihelič, L., Berčič, B., & Petrovčič, A. (2012). Zaznava sovražnega govora na slovenskem spletu. *Teorija in Praksa*, 49(1), 171–189.
- Vehovar, V., Petrovčič, A., & Slavec, A. (2014). E-social science perspective on survey process: Towards an integrated web questionnaire development platform. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel & P. Sturgis (Eds) *Improving Survey Methods* (pp. 170–183). London: Taylor and Francis.
- Vehovar, V., Sicherl, P., Hüsing, T., & Dolničar, V. (2006). Methodological challenges of digital divide measurements. *The Information Society: An International Journal*, 22(5), 279–290.
- Vehovar, V., Slavec, A., & Berzelak, J. (2011). *Web survey software*. Presented at the 5th Internet Survey Methodology Workshop. Hague, Netherlands.
- Vehovar, V., Slavec, A., & Sendelbah, A. (2014). *Investigating respondent multitasking in web surveys: Comparison of self-reports and paradata indicators*. Presented at the 7th Internet Survey Methodology Workshop, Bolzano, Italy.
- Vehovar, V., Zaletel, M., & Seljak, R. (2008). Probabilistic methods in surveys and official statistics. In T. Rudas (Ed.), *Handbook of probability: Theory and applications* (pp. 225–239). Thousand Oaks, CA: SAGE.
- Vicente, P., & Reis, E. (2010). Using questionnaire design to fight nonresponse bias in web surveys. *Social Science Computer Review*, 28(2), 251–267.
- Villar, A. (2013). *Feasibility of using web to survey at a sample of addresses: A UK ESS experiment*. Presented at Genpoweb, London. Retrieved from <http://www.natcenweb.co.uk/genpoweb/documents/ESS-experiment-Ana-Villar.ppt>.
- Villar, A., Callegaro, M., & Yang, Y. (2013). Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys. *Social Science Computer Review*, 31(6), 744–762.
- Vonk, T., van Ossenbruggen, R., & Willems, P. (2006). The effects of panel recruitment and

- management on research results: A study across 19 online panels. Presented at the ESOMAR Panel Research Conference, Barcelona, November.
- Waal, T. de, Pannekoek, J., & Scholtus, S. (2011). *Handbook of statistical data editing and imputation*. Hoboken, New Jersey: Wiley.
- Wagner, J., West, B. T., Kirgis, N., Lepkowski, J. M., Axinn, W., & Kruger-Ndiaye, S. (2012). Use of paradata in a responsive design framework to manage a field data collection. *Journal of Official Statistics*, 28(4), 477–499.
- Walker, R. W., Pettit, R., & Robinson, J. (2009). A special report from the Advertising Research Foundation – The foundations of quality initiative: A five-part immersion into the quality of online research. *Journal of Advertising Research*, 49(4), 464–485.
- Walston, J. T., Lissitz, R. W., & Rudner, L. M. (2006). The influence of web-based questionnaire presentation. variations on survey cooperation and perceptions of survey quality. *Journal of Official Statistics*, 22(2), 271–291.
- Wansink, B., & Sudman, S. (2002). Predicting the future of consumer panels. *Journal of Database Marketing*, 9(4), 301–311.
- Warren, J. R., & Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research*, 41(4), 491–534.
- WebSM. (2014). Overview of web survey software market. Retrieved from <http://www.websm.org/sw>.
- Weible, R., & Wallace, J. (1998). Cyber research: The impact of the Internet on data collection. *Marketing Research*, 10(3), 18–31.
- Weiner, S. P., & Dalessio, A. T. (2006). Oversurveying: Causes, consequences and cures. In A. I. Kraut (Ed.), *Getting action from organizational surveys. New concepts, technology, and applications* (pp. 294–311). San Francisco: Jossey-Bass.
- Wejnert, C., & Heckathorn, D. D. (2008). Web-based network sampling efficiency and efficacy of respondent-driven sampling for online research. *Sociological Methods & Research*, 37(1), 105–134.
- Wells, T., Bailey, J. T., & Link, M. W. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review*, 32(2), 238–255.
- West, B. T., & Sinibaldi, J. (2013). The quality of paradata: A literature review. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic uses of process information*. Wiley.
- Whitcomb, M. E., & Porter, S. R. (2004). E-Mail contacts. A test of complex graphical designs in survey research. *Social Science Computer Review*, 22(3), 370–376.
- Wiley, J. B., Han, V., Albaum, G., & Thirkell, P. (2009). Selecting techniques for use in an internet survey. *Asia Pacific Journal of Marketing and Logistics*, 21(4), 455–474.
- Wilkerson, J. M., Shenk, J. E., Grey, J. A., Rosser, B. R. S., & Noor, S. W. (2013). Recruitment strategies of methamphetamine-using men who have sex with men into an online survey. *Journal of Substance Use*, 1–5.
- Witt, J., Rowland, K., & Wilkinson, C. (2012). Redeveloping the research section of Meningitis UK's website — A case study report. *International Journal of Nonprofit and Voluntary Sector Marketing*, 17(3), 209–218.
- Woolley, P., & Peterson, M. (2012). Efficacy of a health-related Facebook social network site on health-seeking behaviors. *Social Marketing Quarterly*, 18(1), 29–39.
- Wright, G., & Peugh, J. (2012). Surveying rare populations using a probability-based online panel. *Survey Practice*, 5(3).
- Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, 10(3). Retrieved from <http://jcmc.indiana.edu/vol10/issue3/wright.html>
- Yan, T., Kreuter, F., & Tourangeau, R. (2012). Evaluating survey questions: A comparison of methods. *Journal of Official Statistics*, 28(4), 503–529.

- Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22(1), 51–68.
- Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, A. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, 75(4), 709–747.
- Zhang, C. (2011). *Impact on data quality of making incentives salient in web survey invitations*. Presented at the American Association for Public Opinion Research (AAPOR) 66th Annual Conference, Phoenix, Arizona.
- Zhang, C. (2013). *Satisficing in web surveys: Implications for data quality and strategies for reduction* (A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Survey Methodology) in the University of Michigan 2013.). Ann Arbor, Michigan.
- Zhang, C., & Conrad, F. G. (2013). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127–135.
- Zhang, Q., Shao, P., & Fang, J. (2008). Bibliometric Analysis of Current Web Survey Research in China. *Tsinghua Science & Technology*, 13(3), 420–424.
- Zuell, C., Menold, N., & Körber, S. (2015). The influence of the answer box size on item nonresponse to openended questions in a web survey. *Social Science Computer Review*, 33(1), 115–122.
- Žagar, S., & Lozar Manfreda, K. (2012). *Effect of different stimulus on data quality in online panels*. Presented at the 6th Internet Survey Methodology Workshop, Ljubljana, Slovenia.

作者索引

- Aalberg T., 19
Adams M. J. D., 157, 174
Ainsaar M., 41, 69, 109, 118, 164
Airey J., 104
Ajzen I., 154
Albaum G., 26, 108, 154, 180, 189, 295
Albert W., 128
Alexander G. L., 176
Alexander M. W., 21
Allum N., 74
Alwin D. F., 70, 74, 85, 89, 275
Amarov B., 253
Ambrose D., 248
Andrews D., 33
Andrus C., 231
Anseel F., 161
Antoun C., 58
Aoki K., 21
Appel M., 68
Araldsen G., 31
Asplund J. W., 100
Aviram H., 123
- Babbie E. R., 278, 288
Bailey D., 59
Bailey J. T., 229
Bailey M., 151
Baker J. D., 287
Baker R. P., 56, 57, 61, 63, 66, 105, 108, 112, 115, 125, 126, 128, 179, 213, 241, 243, 249-253, 254
Baker-Prewitt J., 115
Baltar F., 57
Balter K., 176
Balter O., 176
Bandilla W., 123, 178, 179
Barchard K. A., 146, 149, 150
Barge S., 119
Barlas F. M., 115, 120, 143
Barnfather N., 158
Barrios M., 80
Barron G., 183
Barry A. E., 123
Bartlett J. E., 21
Bartram D., 35
Baruch Y., 154, 161
Basso L., 179, 181
Basson D., 173, 185
Batagelj Z., 67, 102, 110, 154, 158, 178, 282, 294
Batinic B., 21, 173
Battaglia M. P., 56, 214
Bauman S., 104
Bavdaž M., 31
Bearden W. O., 73, 74
Beaty J. C., 36
Becher A., 275
Beckenbach A., 24
Beckers T., 143, 197
Behr D., 123
Belak E., 43
Bell D. S., 95
Bellman S., 158
Benford R., 197
Bennink M., 41
Berčič B., 57
Berglund P. A., 208
Bergmann M., 143
Bernet F., 225
Berzelak J., 127, 265
Berzelak N., 26, 42, 161, 171, 271
Berzofsky M., 59
Bethlehem J., 5, 42, 44, 50, 52, 61, 69, 77, 169, 214, 253
Beukenhorst D., 279
Bhaskaran V., 254, 292
Bhola D. S., 95
Bianchi A., 59, 282
Biemer P. P., 17, 19, 40, 44, 59, 68, 73, 130, 204, 205, 215, 271, 273, 288
Biffignandi S., 42, 44, 50, 52, 59, 61, 69, 77, 214, 282
Biner P. M., 154
Birnbaum M. H., 37

- Bishop L., 147
 Black G. S., 18
 Blair E. A., 278
 Blair J., 123
 Blair J. E., 278
 Blanke K., 105, 280
 Blasius J., 97, 210
 Blom A. G., 30
 Bocarnea M. C., 287
 Bonham V. L., 180
 Booth-Kewele S., 26
 Borg L., 173
 Börkan B., 41, 276
 Borrego Á., 80
 Bosnjak M., 26, 30, 117, 118, 154, 158, 167, 173, 175, 177, 179, 230, 244
 Bossert D., 170
 Boulianne S., 173, 176, 185
 Bowker D. K., 24
 Boyd C. J., 26
 Boyle J. M., 247
 Brace I., 72, 99
 Bradburn N. M., 70, 72, 73, 75, 92, 97, 121, 124
 Bradley M., 86
 Bradshaw C. C., 21
 Brandal H., 186
 Braun M., 123
 Braunsberger K., 275, 276
 Bremer J., 59, 213, 245, 253
 Brennan C. W., 180
 Bretveld R. W., 30
 Brick J. M., 18, 56
 Brickman Bhutta C. B., 57
 Briers G., 179
 Brooke J., 34
 Brownell B., 116
 Brown-Welty S., 19
 Bruijne M. de, 229
 Brunet I., 57
 Buchanan E. A., 147
 Buchanan T., 157, 274, 276
 Budiu R., 234
 Buelens B., 279, 280
 Buhrmester M., 56
 Burchett H. E., 180
 Burdein I., 96, 117
 Burr M. A., 275
 Buskirk T. D., 231
 Cai Z., 124
 Calinescu M., 277
 Callegaro M., 25, 30, 43, 66, 69, 92, 95, 98, 104, 110, 114, 126, 142, 143, 144, 159, 180, 185, 187, 197, 226, 229, 230, 231, 232, 242, 243, 244, 245, 247, 248, 251, 291
 Cameron D., 175
 Campanelli P., 121, 122, 123, 124
 Cantor D., 31
 Cape P., 117
 Carrasco J. A., 100
 Carey R., 186
 Carini R., 275
 Carlson J. F., 74
 Carson R. T., 85
 Casey T.W., 112
 Cechanowicz J., 116
 Champion H., 161
 Chan P., 248
 Chandler D., 143
 Chaney B. H., 123
 Chaney J. D., 123
 Chang C.-Y., 116
 Chang L., 26, 41, 66, 120, 174, 210, 250, 274, 275, 276
 Chao S., 210
 Chen K., 75
 Chepp V., 123
 Chien Y.-T., 116
 Chin T.-Y., 95
 Chittenden L., 185
 Cho H., 146, 157
 Choragwicka B., 161
 Chou P.-N., 33
 Christian L. M., 26, 28, 30, 32, 44, 80, 96, 120, 174
 Christin N., 151
 Churchill G. A. Jr., 72, 278, 287
 Cialdini R. B., 154
 Claridge N., 34
 Clayton R. L., 18
 Cobanoglu N., 177
 Cobben F.S., 253
 Coenders G., 100
 Coiner T., 120
 Comley P., 154, 174, 177, 243
 Connor J., 182
 Conrad F. G., 35, 58, 74, 75, 77, 79, 82, 85, 88, 89, 93, 95, 98, 99, 107, 111, 112, 113, 120, 123, 143, 157, 162, 213, 233, 249, 276
 Conradt J., 24
 Conroy B., 18
 Converse J. M., 17

- Converse P. D., 43
 Cook C., 154, 176, 180, 183
 Cooke M., 239
 Coombe R., 115, 127
 Cooper A., 26
 Cooper M., 275
 Corbett M., 34
 Corning A., 154
 Coromina L., 100
 Corti L., 147, 150, 221
 Couper M. P., 1, 17, 26, 27, 30, 56, 70, 74, 75, 77, 79, 80, 82, 85, 87, 88, 89, 93, 94, 96, 98, 99, 102, 103, 105, 107, 108, 111, 112, 113, 117, 118, 120, 128, 134, 141, 143, 146, 154, 157, 158, 162, 176, 178, 179, 213, 226, 229, 230, 231, 233, 249, 271, 276, 290, 294, 295
 Crawford S. D., 26, 82, 94, 103, 107, 118, 125, 136, 181, 230, 254, 255, 290
 Crête J., 26
 Crutzen R., 26, 163, 180
 Csikszentmihalyi M., 114
 Czaja R. F., 278
 Čehovin G., 106, 118, 144, 163, 164, 165, 167, 216, 255, 257, 258, 259, 260, 264, 265, 266, 267, 268, 289, 292

 da Costa F. F., 128
 Dalessio A. T., 157
 Daneback K., 26
 Daniel F., 124
 D'Arcy H., 26
 Das M., 81, 119, 243, 251, 253
 Datta R. A., 173
 Daver J. A., 214
 Dawson C. R., 36
 Day B., 180
 De Beuckelaer A., 42
 de Heer W., 18
 de Jong M., 34
 de Jongh A., 95
 de Leeuw E., 17, 18, 23, 41, 42, 44, 73, 95, 180
 de Nooy, 294
 de Rouvray C., 80
 de Ruyter K., 98, 275
 Dean E., 33, 123
 DeMaio T., 122, 123
 Dennis J. M., 80
 Dennis M. J., 15
 DeRouvray C., 108
 Deutskens E., 98, 118, 177, 181, 274, 275
 Dietrich B. J., 273
 Dijkstra W., 123

 Dillman D. A., 13, 24, 26, 28, 30, 32, 40, 42, 43, 44, 50, 52, 75, 79, 80, 84, 87, 89, 90, 95, 96, 102, 103, 108, 112, 117, 120, 126, 136, 154, 172, 179, 180, 182, 183, 184, 185, 187
 Diment K., 158
 DiSogra C., 30, 43, 98, 159, 180, 242, 243, 244, 245, 247, 251
 Dobrow M. J., 157
 Dolničar V., 67
 Donald E., 14
 Downes-Le Guin T., 115, 116, 175, 176, 177
 Drasgow F., 35, 36
 Dugan J. P., 230
 Durant R., 161
 Dykema J., 180

 Edwards P. J., 180, 183, 185
 Edwards S. W., 31
 Ehling M., 273
 Ekeh P. P., 154
 Eklund R. C., 114
 Elling S., 34
 Elliott M. N., 13
 Elsmar M., 21
 Eltinge J. L., 130
 Emde M., 80, 86
 Enders C. K., 211
 Epstein L., 75
 Ester P., 102
 Esuli A., 215
 Ewing T., 116
 Eyerman J., 232
 Eysenbach G., 287

 Faas T., 174
 Fallaw S. S., 36
 Falls A., 27
 Fan W., 154
 Fan X., 158
 Fang, J., 119, 120, 154, 158, 172, 173, 294
 Faught K. S., 179
 Feeny D., 88
 Felix L. M., 180, 185
 Fenner Y., 57
 Fernée H., 228
 Fisher B., 225
 Fleisher J. M., 19
 Flynn D., 88
 Fondell E., 176
 Fontaine S., 27
 Fowler F. J., 2

- Frankel M. R., 214
 Fricker R. D., 13
 Fricker S. S., 94, 120
 Frost L. Z., 210
 Fuchs M., 75, 80, 86, 99, 107, 170, 239
 Fulton J., 42, 43
 Funke F., 88, 230
 Furlong W., 88
- Gaffney G., 35
 Gaiser T. J., 123
 Gajic A., 175
 Galesic M., 30, 66, 75, 93, 94, 107, 117, 118, 119, 126, 173, 174, 177, 290
 Gallhofer L., 70, 72, 78, 80, 85, 87, 88, 89, 124, 125
 Ganassali S., 98, 118
 Garland P., 57, 75, 119, 120, 143
 Garrett-Jones S., 158
 Gates R., 275
 Gathmann C., 30
 Gatward R., 257
 Gebler N., 175
 Gehlbach H., 119
 Geisinger K. F., 74
 Gelder M. M. H. J. van, 30
 Gelissen J., 41
 Gilles C., 80
 Glasner T., 99
 Goldsby T. J., 275
 Goldstein D. G., 177
 Goodfellow L., 116
 Goodman R., 21
 Göritz A. S., 26, 163, 175, 176, 177, 180, 246
 Gosling S. D., 56
 Gouldner A. W., 154
 Goyder J., 154, 159
 Grady J., 40
 Graesser A. C., 124
 Grandquist L., 209
 Granello D. H., 179
 Green K. E., 161
 Green Jr, K. W., 179
 Greene J., 43
 Greenlaw C., 19
 Gregg J. A., 182
 Grenville A., 229
 Greski R., 250
 Grey J. A., 58
 Griffiths S. E., 275
 Grizzle J. E., 107
- Grossnickle J., 13
 Groves R. M., 2, 19, 42, 44, 50, 52, 68, 70, 71, 73, 76, 100, 122, 124, 153, 154, 157, 168, 169, 170, 172, 173, 188, 204, 205, 208, 209, 214, 219, 271, 272, 276, 278, 282, 288, 290
 Guéguen N., 173
 Guidry K. R., 227, 231
 Guo J., 36
 Gutierrez C., 143
 Gutwin C., 116
- Haas L., 26, 30
 Hadlock T. D., 40
 Hakan A., 104
 Halpern-Manners A., 250
 Hamari J., 116
 Hamel L., 161
 Hammen K., 105
 Hammer H., 161
 Han V., 180
 Hansen J. M., 174
 Hansen K. M., 179
 Hansen S. E., 128
 Hanway S., 42
 Haraldsen G., 144
 Hardigan P. C., 19
 Harrison, C., 161
 Hart A. M., 180
 Harter J. K., 100
 Hartford K., 186
 Hattrup K., 183
 Haunberger S., 154, 158
 Havens C. G., 120
 Haws K. L., 73
 Hayek J., 275
 Hayes B. E., 28
 Head B., 123
 Healey B., 82
 Healey J. F., 60, 216
 Heath F., 154
 Heckathorn D. D., 56, 57
 Heeringa S. G., 207, 211, 217, 282, 288
 Heerwegh D., 21, 26, 109, 119, 136, 143, 154, 158, 173, 177, 183, 197
 Heiervang E., 21
 Hendarwan E., 43
 Hendel D. D., 275
 Henderson V., 92, 185, 186, 187
 Henning J., 21, 55, 56, 94
 Hertel, 21, 274, 276

- Hill C. A., 33, 102
 Hillygus S. D., 250
 Hlebec V., 34, 80, 100
 Hoaglin D. C., 214
 Hogan B., 100
 Holbrook A. L., 74, 85
 Holland J. L., 80, 174
 Holmberg A., 130
 Holtom B. C., 154, 161
 Hoogendoorn A., 246, 247
 Hox J. J., 23, 41, 42, 44, 72, 95, 180, 212, 280
 Hoy M. G., 13
 Huang X., 43
 Huggins V., 232
 Hughes K. A., 131
 Hundepool A., 219
 Hurley J., 175
 Hüsing T., 67
 Hvizdak E. E., 147

 Iacobucci D., 72, 278, 287
 Iarossi G., 204, 278, 286
 Ibáñez-Zapata J. Á., 181
 Irani T. A., 182
 Italiano P., 27

 Jäckle A., 275
 Jackson N., 250
 Jackson S. A., 114
 Jacob C., 173
 Jacquemain M., 27
 Jain-Shukla P., 161
 Janowitz P., 175, 176
 Jansson I., 130
 Jarrett C., 35, 115
 Jobity N., 104
 Jochum M., 183
 Johnson A., 115, 158
 Johnson E. J., 176
 Johnson E. P., 107, 212
 Johnston M., 281
 Joinson A. N., 26, 136, 157, 158, 183
 Jones J., 31
 Jue A., 230

 Kaczmirek L., 30, 94, 110, 123, 126, 127, 128, 144, 145, 149, 150, 178, 179, 254
 Kaminska O., 30
 Kantrowitz T. M., 36
 Kapelner A., 143
 Kaplowitz M. D., 40, 118, 173, 178, 180, 185, 186, 187
 Kapteyn A., 158
 Kavčič L., 144
 Kenneally E., 151
 Kennedy C., 79, 173
 Kennedy C. K., 197
 Kennedy J. M., 275, 286
 Kent R., 186
 Keusch F., 174, 180
 Khan C. E., 95
 Khorram S., 175
 Kidd H. J., 154
 Kiernan M., 80
 Kiernan N. E., 80
 Killham E. A., 100
 Kiniorski K., 186
 Kirakowski J., 34
 Kirk R. E., 71
 Kjær T., 102
 Klausch L. T., 95, 280
 Klein D., 26
 Klein J. D., 120
 Klostad C. A., 173, 185
 Knox S., 55, 243
 Kobrin S. J., 158
 Koivisto J., 116
 Konradt U., 21
 Körber S., 80
 Koren G., 100
 Korendijk E., 180
 Körner T., 273, 280
 Kovar J. G., 209
 Kowalski J., 99, 116
 Krantz J. H., 37, 124
 Krasilovsky P., 13
 Kreuter F., 26, 131, 141, 143, 214
 Krieger U., 30
 Krosnick J. A., 26, 41, 63, 66, 70, 72, 73, 74, 80, 84, 85, 87, 102, 109, 114, 117, 119, 120, 162, 174, 244, 248, 250, 251, 274, 275, 276
 Kroth P. J., 43
 Krotki K., 15
 Krug S., 35
 Kruse Y., 246, 250
 Krysan M., 99
 Kuh G., 275
 Kuntz A., 143
 Kunz T., 107
 Kurzynski D., 143
 Kwak N., 21, 80, 158

- Kwang T., 56
 Kypri K., 182
- Lackaff D., 100
 Lai V. S., 267
 Lamias M. J., 87, 118
 Lan G., 154
 Landreth A., 122, 123
 Lang P., 86
 Langer G., 62
 Langley J. D., 182
 Larose R., 146, 157
 Larson E., 180
 Larson G. E., 26
 Lavrakas P. J., 197, 242
 LeClaire J., 254, 292
 Leder K., 183
 LeDuff Collins M., 100
 Lee S., 251
 Legleye S., 244
 Lenar J., 144
 Lengler R., 37
 Lensvelt-Mulders G., 180
 Lentz L., 34
 Lenzner T., 131
 Lesnard L., 30, 244
 Levin K., 275
 Levine R., 40
 Ličen S., 34
 Liebau E., 57
 Lievens F., 42, 161
 Lilleoja L., 5
 Lindhjem H., 21, 119, 120
 Link M. W., 229
 Lissitz R. W., 118
 Little R. J. A., 211
 Lobe B., 281
 Lohr S. L., 45, 50, 52, 54, 60, 62
 Lohse G. L., 158
 Loosveldt G., 21, 26, 119, 136, 154, 158, 173, 275
 Lord T. H., 177
 Lorenc B., 130
 Lugtig P., 225, 231, 251
 Luiten A., 277, 280
 Lumiste K., 5
 Lumsden J., 81
 Lundström S., 211
 Lupi F., 118
 Luthe S. C., 176, 177
 Lyberg L. E., 17, 19, 40, 44, 68, 73, 201, 204, 205, 215, 271, 272, 273, 286, 288
- Lynn P., 88, 275
- MacInnis B., 248
 MacLennan B., 182
 Macer T., 2, 4, 27, 32, 62, 117, 225, 232, 237, 254, 258, 259, 261, 279
 Madans J., 121, 257
 Maitland A., 121
 Malawaraarachchi M., 183
 Malhotra N., 85, 102, 119, 120, 233
 Maloney-Krichmar D., 33
 Mangione C. M., 95
 Mangunkusumo R. T., 41, 274, 275, 276
 Manierre M. J., 175, 177
 Manisera M., 74
 Månsson S.-A., 26
 Marcell M., 27
 Martin C., 174
 Martin E., 123
 Martinussen M., 161
 Mastrangelo P. M., 173
 Matross R. P., 275
 Matthijs K., 158
 Maughan D., 151
 Mavletova A., 102, 103, 229, 230, 231, 232
 McBride M., 80
 McCabe S. E., 26, 94, 103
 McClain C. M., 230
 McCutcheon A. L., 30
 McGonagle K., 175
 McHugh J., 151
 McMahon L., 94, 115, 117
 McMorris B. J., 43
 McNaughton R. B., 17
 Mechling J., 109, 112, 115
 Medway R. L., 42, 43
 Mendonca J., 186
 Menold N., 80
 Mesh, G., 14
 Messer B. L., 40
 Meyer M., 250
 Mihelič L., 57
 Millar, M. M., 43, 179
 Miller E. T., 41, 274, 275
 Miller J., 115
 Miller K., 121, 123
 Miller L. C., 177
 Miller Steiger D., 18
 Mitchell R. C., 85

- Mitra A., 161
 Miyoshi D. K., 26
 Močnik A., 106
 Mohler P. P., 220
 Molenberghs G., 275
 Montoro-Ríos F., 181
 Moore D. L., 57, 125
 Moors G., 41
 Moreo P. J., 21
 Morgan D. L., 281
 Morren M., 31
 Morrison R. L., 32, 96
 Moser K., 180
 Motl A., 57, 58
 Mrvar, 294
 Mueller K., 183
 Mühlberger N., 49
 Muñoz-Leiva F., 181, 183
 Murakami M., 92
 Murphy J., 161, 171, 179, 187
 Musch J., 163
- Nadler R., 21
 Naglieri J. A., 35
 Nallan S., 129
 Napoli P. M., 242
 Naumann S., 21
 Navrud S., 21, 119, 120,
 Netemeyer R. G., 73
 Neubarth W., 97, 179
 Neuman W. L., 72, 84
 Nicholls II W. L., 17
 Nicolaas G., 146, 253
 Nielsen A., 239
 Nielsen J., 75, 103, 108, 111, 126, 127, 128, 186, 234, 263,
 264
 Nielsen J.S., 102
 Nigg C. R., 43
 Nonnecke B., 33
 Noor S. W., 58
 Nunan D., 55, 243
 Nye C. D., 36
- O'Brien H. L., 114
 O'Neill A. C., 30
 O'Toole J., 183
 Ollé C., 80
 Olson K., 24, 196
 Ongena Y. P., 123
 Oosterveld P., 98
- Osborne J. W., 210
 Oswald F. L., 43
 Oudejans M., 80
 Ouimet J., 275
 Overmeyer C., 213
 Owen R., 236
 Oyler M. A., 80
- Padilla J. L., 123
 Paine C., 157
 Pannekoek J., 208
 Parkhurst B., 196
 Parsons N. L., 175, 177
 Pasek J., 63, 248
 Payne J., 158
 Pearson J. E., 24, 29
 Pedersen R. T., 179
 Pennington J., 180
 Pernice K., 127
 Peterson G., 223, 224, 225, 227, 229, 230, 239
 Peterson M., 33
 Petrič G., 160
 Petric L., 68
 Petrovčič A., 57, 131, 133, 160, 164, 167
 Pettit R., 248
 Peugh J., 247
 Peytchev A., 102, 103, 107, 113, 115, 157, 158, 164, 211, 255
 Peytcheva E., 169, 170, 172
 Pferdekaemper T., 233
 Pit S. W., 188
 Pitkow J., 294
 Pope D., 94
 Poropat A., 112
 Porter S. R., 13, 14, 40, 173, 174, 177, 180, 183, 185, 186
 Postoaca A., 243
 Potaka L., 128
 Poynter R., 47, 56, 98, 240
 Preece J., 33
 Presser S., 26, 70, 72, 73, 74, 80, 84, 85, 87, 102, 109, 117,
 120, 121, 122
 Prybutok V., 119
 Puleston J., 116
 Pyakurel S., 188
- Radler B., 21, 80, 158
 Raghunathan T., 107, 175
 Ramos M., 12, 17
 Rao K., 30, 143, 180
 Rasinski K. A., 73
 Raskin O., 13

- Rathod S., 179, 181
 Recker M., 294
 Ree M. J., 210
 Reips U.-D., 37, 88, 107, 124, 136, 143, 144, 157, 158, 163, 197, 206, 294
 Reis E., 175
 Reja U., 80
 Rendtel U., 253
 Rettie R., 185
 Revilla M., 274
 Reynolds R. A., 287
 Rintoul D., 116
 Rips L. J., 73
 Ritter L. A., 278
 Rivara F. P., 275
 Rivas S., 26
 Rivers D., 59, 62, 244
 Roach M., 177, 179, 183
 Robbins A., 161
 Roberts A., 95
 Roberts C., 116, 275, 277
 Roe D. J., 127, 288
 Roeleveld N., 30
 Rogers R. D., 26
 Roman E., 28
 Rookey B. D., 42
 Roots A., 5
 Rosenbaum P. R., 63, 213
 Ross M. W., 26
 Rosser B. R. S., 58
 Roßmann J., 143
 Roster C. A., 26, 274, 275, 276
 Rowland K., 34
 Rubin D. B., 63, 210, 211, 213
 Rubinson J., 248
 Rudner L. M., 118
 Ruyle E., 115
 Sackett P. R., 36
 Samaranyaka A., 182
 Sánchez-Fernández J., 181, 183
 Saris W. E., 14, 15, 17, 70, 72, 73, 78, 80, 85, 87, 88, 89, 121, 124, 125
 Särndal C.-E., 211
 Sarsa H., 116
 Sauermann H., 177, 179, 183
 Schaeffer D. R., 13
 Scherpenzeel A., 30, 43, 50, 162, 164, 176, 228, 243, 251, 253
 Schmidt F. L., 100
 Schmidt S., 229
 Schneeweiß S., 49
 Schober M. F., 58, 120
 Schoen H., 174, 250
 Schollaert E., 161
 Scholtus S., 208
 Schonlau M., 13, 57, 158, 249, 250
 Schouten B., 277, 280, 282, 284
 Schriver K. A., 111
 Schwarz N., 73, 101, 154
 Sebastiani F., 215
 Sedivi B. M., 12
 Seljak R., 219
 Sellers S. L., 180
 Sendelbah A., 197
 Shamdasani P. N., 123
 Shannon D. M., 21
 Shao P., 154, 294
 Sheehan K. B., 13, 158
 Shenk J. E., 58
 Shih T.-H., 158
 Shinn G., 179
 Shneiderman B., 264
 Sicherl P., 67
 Siegal J. W., 213
 Siegers P., 143
 Sikkel D., 246, 247
 Siluk L., 107
 Silver N., 248
 Siminski P., 102
 Sinclair M., 183
 Singer E., 146, 153, 154, 157, 175, 290
 Sinibaldi J., 143
 Slavec A., 127, 131, 133, 197, 255, 267
 Sleep D., 116
 Smith J. L., 274, 276
 Smith P., 74
 Smith R. M., 186
 Smith S. M., 154, 174, 189
 Smith T. W., 27, 159, 272
 Smyth F. L., 102
 Smyth J. D., 24, 26, 28, 29, 30, 44, 69, 80, 96, 120, 178
 Snijkers G., 23, 31
 Speizer H., 43
 Spies R. A., 74
 Spijkerman R., 274
 Stähli M. E., 277
 Stamp R., 94, 115, 117
 Stapleton C., 223, 229, 230
 Stark R., 257
 Stavrou A., 151

- Steiger D. M., 120
 Steinbrecher M., 143
 Steinmetz S., 59, 213, 250
 Stellefson M. L., 123
 Stephenson L. B., 26
 Stephenson S., 182
 Stern H. S., 210
 Stern M. J., 120
 Stettler K., 128
 Stevenson J., 180
 Stewart D. W., 123
 Stieger S., 143, 144, 158, 197, 206
 Stone R., 175
 Stoop I., 171
 Stouthamer-Loeber M., 286
 Straatmann T., 183
 Strabac Z., 19
 Strong C., 239
 Stukowski S., 170
 Sturgis P., 74
 Succar C. T., 19
 Sudman S., 70, 73, 241
 Sue V. M., 278
 Suh A., 75, 248
 Svensoon J., 253
 Sweet E. M., 12
 Swicegood J., 123
 Swinehart J., 125
 Swoboda W. J., 49
 Sym D., 180
- Tarnai J., 125, 286
 Tarraf S., 107
 Tay L., 36
 Taylor H., 213, 248
 Tedesco D., 128
 Telg R., 182
 Tepman Z., 92
 Terhanian G., 59, 213, 245, 253
 Thiele O., 149, 150
 Thiessen V., 210
 Thirkell P., 180
 Thomas R. K., 74, 85, 88, 115, 119, 120, 143
 Thompson R. L., 154
 Thorndike F. P., 102, 103
 Thorp L., 118
 Tijdens K. G., 50, 58, 59, 215
 Tikkanen R., 26
 Toepoel V., 81, 84, 85, 94, 98, 102, 103, 119, 176, 225, 231
 Tompson T. N., 197
- Toms E. G., 114
 Tomsic M. L., 275
 Torrance G. W., 88
 Tortora R., 24
 Tourangeau R., 26, 27, 41, 56, 73, 74, 75, 77, 79, 82, 85, 88, 89, 93, 94, 98, 99, 102, 107, 111, 112, 113, 120, 124, 131, 143, 162, 165, 173, 174, 213, 233, 249, 271, 276, 295
 Townsend R., 238
 Traugott M. W., 35, 87
 Trouteaudo A. R., 179, 185
 Truell A. D., 21
 Tullis T., 128
 Tuten T. L., 154, 167, 177, 187
- Uher R., 220
 Uldall B. R., 74
 Umbach P. D., 157, 174
- Valliant R., 214
 van den Bergh H., 34
 van den Brakel J. A., 280
 van den Eyden V., 147
 van der Vaart W., 99
 van der Valk J., 280
 van Hoewyk J., 175
 van Horn P., 161
 van Kammen W. B., 286
 van Ossenbruggen R., 247
 van Schaik P., 88
 van Soest A., 81, 119
 van Wersch A., 88
 Vanhove T., 158
 Vannieuwenhuyze J. T. A., 275, 277, 284
 Vehovar V., 26, 43, 50, 57, 58, 62, 64, 67, 80, 94, 100, 102, 106, 110, 118, 127, 131, 133, 144, 154, 156, 158, 161, 162, 163, 164, 165, 167, 170, 171, 175, 176, 178, 189, 197, 210, 212, 213, 216, 219, 255, 257, 258, 259, 260, 264, 265, 266, 267, 268, 277, 281, 282, 284, 289, 292
 Vendenplas C., 277
 Vicente P., 175
 Villar A., 5, 110, 118, 176, 233, 280
 Villarroya A., 80
 Vinken H., 102
 Vo T., 188
 Vohar M., 26
 Vonk T., 247
 Voracek M., 158
- Waal T. de, 208, 211
 Wagner J., 282

- Walker R. W., 248
Wallace J., 21
Walston J. T., 118, 173, 175
Wang Y., 24
Wansink B., 70, 241
Warde B., 21
Warren J. R., 250
Weathon J. E., 179
Webb M. C., 123
Weible R., 21
Weiner S. P., 157
Weitkunat R., 49
Wejnert C., 56, 57
Wellman B., 100
Wells T., 26, 98, 143, 229, 230, 231
Wen C., 119, 173
Wenzel O., 229, 230
Werking G. S., 18
West B. T., 143, 207
Wetzels M., 98, 275
Wetzels W., 279
Whitcomb M. E., 40, 173, 174, 177, 180, 183, 185, 186
Whitehand R., 34
Whitten D., 179
Wiitala W., 43
Wijnant A., 229
Wiley J. B., 180
Wilkerson J. M., 58
Willems P., 247
Williams D., 18
Williams J., 146, 149, 150
Williams N., 240
Williams R., 59
Willimack D. K., 31
Willis G. B., 121
Willson S., 123
Wilson S., 2, 4, 27, 32, 117, 225, 254, 258, 259, 279
Winter J., 158
Witt J., 34
Wittmann W. W., 154
Wolf J. G., 286
Wolfe E. W., 43
Wolff H.-G., 177
Wolff K. G., 126
Wong B. K., 267
Woodley A., 136
Woollard M., 147
Woolley P., 33
Wright G., 247
Wright K. B., 254
Wyatt J., 287
Wybenga H., 275

Yan T., 94, 95, 110, 126, 131, 143, 165, 173
Yan Z., 154
Yang Y., 95, 110
Yeager D. S., 62, 66, 248, 249
Yechiam E., 183
York S., 240
Young M., 250

Žagar, S., 177
Zaletel M., 154, 219
Zhang C., 58, 82, 94, 98, 119, 120, 143, 186, 267
Zhang Q., 294
Zuccolotto P., 74
Zuell C., 80

主题索引

- 美国民意研究协会 (AAPOR), 63, 160–162, 166, 171, 202, 206, 249, 290, 291, 294
 - 最终处置代码, 159, 208
 - 关于非概率样本的报告, 66, 249, 另见非概率样本《标准定义报告》(2011), 159, 另见最终处置代码
- 访问审计, 150
- 活跃样本库成员 (ISO 2662 标准), 245
- 基于地址的抽样 (ABS)
- 行政表格, 10, 34, 35, 81
- 优势与局限性概述, 19
- 广告研究基金会 (ARF), 248
- 匿名性, 26, 119, 136, 146, 149, 183, 184, 289, 290
- 应用程序, 234–239, 另见移动调查应用程序
- 网络调查的应用 (及 / 用于), 27–34
 - 商业, 31, 32
 - 客户满意度, 28, 29
 - 娱乐性民意调查、问卷调查和小测验, 34
 - 普通人群, 29, 30
 - 高覆盖率的特定人群, 27, 28
 - 在线社区, 32, 33
 - 在线样本库, 32
 - 学生, 27, 28
 - 网站评估, 33, 34
- 流失
 - 自愿流失, 245
- 网络调查的更广泛背景, 见网络调查
- 各章注释
 - 网络调查的更广泛背景, 296
 - 调查后阶段, 221
 - 预调查阶段, 192, 193
 - 网络调查实施的选定主题, 268, 269
 - 调查研究与网络调查, 37
 - 点击率 (CTR), 56, 161
 - 调查方法学认知视角 (CASM), 154
 - 计算机辅助个人面访 (CAPI), 17, 27, 41, 82, 225–237, 271
 - 计算机辅助调查信息收集 (CASIC), 17, 23, 38, 81, 125, 126, 258, 263, 293, 303
 - 计算机辅助电话访问 (CATI), 17, 27, 82, 225
 - 计算机辅助自填式问卷 (CSAQ), 17–19, 23–25, 236, 272
 - 美国调查研究机构协会 (CASRO), 30, 252, 288
 - 《调查研究标准与伦理准则》(2011), 252, 291
 - 覆盖率, 另见无覆盖率与合作, 40
 - 误差, 244, 251
 - 互联网 6, 18, 48
 - 众包, 56, 281, 301
 - 客户满意度测量, 277
 - 研究 / 调查, 7, 28, 29, 49, 142, 180, 191, 254
 - 数据库, 3, 8, 9, 13, 16, 36, 41, 49, 53, 59, 71, 74, 79, 82, 110, 133–135, 140, 141, 205, 206, 208, 210, 214, 215, 218, 219, 243, 244, 245, 261, 292, 293, 294, 299
 - 数据文件, 76–78, 90, 92, 93, 108, 124, 129, 134, 135, 142, 145–147, 149, 150, 152, 197, 198, 205, 207, 209, 217, 219, 220, 257, 261, 279
 - 数据集, 50, 211, 219, 247, 248, 252, 262
 - 子数据集, 145
 - 数据文档倡议 (DDI), 220
 - 数据导出与文档化, 218–220
 - 归档, 220
 - 导出, 218
 - 文档化, 219, 220
 - 统计披露, 219

数据准备, 205–215

编码和重新编码, 214, 215

确定应答状态, 205–208

样本单位的合格性、完整性状态和质量状况,
205–207

缺失值和非实质性应答, 207, 208

应答状态报告, 208

插补, 210–212

验证和编辑, 208–210

加权, 212–214

数据质量, 4, 7, 20–24, 26, 27, 31, 32, 38, 42, 43,
52, 61, 63, 67, 81, 82, 85, 89–91, 94–96, 98,
99, 107, 109, 111, 115, 118, 142, 143, 161,
166, 173, 177, 181, 183, 188, 189, 192, 196,
198, 200–202, 205–210, 212, 219, 222, 226,
241, 246–254, 257, 270–274, 276, 283, 284,
286, 298, 300, 306

定义与分类, 1–7

网络调查的抽样, 5–7

调查和网络调查, 2–4

网络调查的用户, 4, 5

活跃样本库成员定义 (ISO 2662 标准), 245

自愿流失, 245

网络调查模式, 44

《在线调查指南》(德国市场与社会研究协会理事会,
2007), 291

自助式 (DIY) 研究, 5, 49, 198, 255, 279, 302

选举民意调查, 247–249

电子邮件, 6, 8, 11–16, 39, 48–50, 56, 69, 104,
119, 130, 133, 135, 137–140, 149, 155, 157,
160, 178, 182, 185, 195–197, 200, 219, 242,
257, 275, 298, 299

退回, 139

联系方式, 40, 178, 179

设置, 8, 137–140

调查邀请, 19–22, 29–31, 33, 39, 43, 50, 56, 104,
135–142, 144, 155–157, 159–162, 173, 174,
179–187, 195, 199, 202, 206, 225, 227, 229,
230, 232, 239, 261, 281, 282, 288

调查, 10–18, 29, 185

调查参与度, 114–121, 153, 174, 175

企业反馈管理 (EFM), 28

错误信息, 25, 108, 234

机构调查, 5, 31, 另见调查

伦理准则, 290

考量, 200, 212, 289, 290

标准, 219

伦理, 29, 30, 104, 108, 110, 141, 147, 277, 287,
291

以及《调查研究标准与伦理准则》(美国调查研究机构
协会), 252

社会科学纵向互联网研究 (ELIPSS) (法国), 30

欧洲社会调查 (ESS): 调查质量预测工具, 73

欧洲民意与市场研究协会 (ESOMAR), 2, 4, 288

欧洲调查研究协会 (ESRA), 288

网络调查的发展、应用和相关实践, 10–27, 另见网络调
查

优势与局限性概述, 19–27

问卷的计算机化, 23, 24

调查模式的背景, 17–19

成本, 19, 20

实施的便利性, 23

多媒体, 24, 25

自我管理, 25–27

数据收集的速度, 21, 22

时间和地理上的灵活性, 25

网络调查的前驱和替代方案, 12–15

现代网络调查的兴起, 15–17

网络调查与互联网和在线调查, 11, 12

Facebook, 33, 57, 203, 236, 300

正式调查阶段 (及), 194–203, 299

关键绩效指标 (KPIs), 201–203, 另见主题条目测量,
196, 197

处理和监控 (及), 197–203

启动、软启动、全面启动和停用, 197, 198

实地调查干预, 200, 201

实地调查过程监控, 201–203

提供帮助, 199, 200

招募, 194, 195

实地调查, 8, 41, 53, 124, 130, 141, 149, 179, 181,

- 187, 189–191, 194, 195, 197–208, 210, 215–221, 257, 266, 277, 281–283, 286, 298, 302
- 最终处置代码, 159, 208
- Flash 技术, 14, 142, 199, 230, 234, 235
- 网络调查的未来 (及), 297–304
 - 更广泛的商业和社会问题, 302–304
 - 一般技术发展, 297, 298
 - 方法学, 300, 301
 - 网络调查软件, 299, 300
- 游戏化, 98, 114–116, 301
- 通用在线研究 (GOR), 294
- 德国的 Gesis 互联网样本库, 27, 30, 244
- 全球研究商业网络 (GRBN), 291
- 谷歌, 228, 264, 300
 - 谷歌网页速度观察, 265
- HTML/HTML5 (超文本标记语言 / 第五代超文本标记语言), 11, 13, 14, 16, 17, 29, 77, 81, 82, 91, 92, 101, 185, 186, 234, 235, 238, 254, 268, 291, 298
- 插补, 210–212
- 互联网, 非互联网用户, 4, 5, 29, 45, 67, 68, 69
- 互联网测试, 10, 35–37, 另见测试
 - 指导方针, 36
 - 国际测试委员会关于互联网测试计算机化的指导方针, 36
 - 相关文献, 36
 - 无人监考的互联网测试 (UIT), 36
- Java/Javascript (编程语言), 14, 16, 106, 125, 142, 158, 199, 235, 238, 240
- 关键绩效指标 (KPIs), 201–203
 - 及其在网络调查中的应用, 202, 203
- 知识网络 (公司名, Knowledge Networks), 15, 180, 243, 244
- 荷兰的社会科学纵向互联网研究 (LISS), 27, 30, 162, 164, 176, 243, 244, 另见荷兰
 - 样本库与面对面访谈, 43
- 测量水平 70, 71, 77
 - 等距测量, 83–87
 - 名义测量, 78, 86
 - 顺序测量, 83, 86
 - 过程, 73, 77, 144
 - 比率测量, 71, 83, 86, 87, 95
- 调查模式详述 (及), 38–44
 - 混合模式设置, 40–44
 - 混合模式调查、多模式联系和混合模式系统, 40, 41
 - 混合模式调查的问题, 41, 42
 - 混合模式下的网络调查, 42–44
 - 选择网络调查模式, 39, 40
- 模型
 - 认知方法——满足 (Krosnick), 74
 - 决策模型, 283
 - 预测模型, 63
 - 按受访者数量收费模型, 260
 - 应答过程 (Tourangeau, Rips and Rasinski), 73
 - 统计模型, 124, 129
- 荷兰, 15, 30, 31, 43, 59, 179, 181, 229, 243, 247, 274
 - 全国读者调查, 68
- 名义测量, 71, 78
- 非概率样本, 6, 7, 32, 48, 54, 58, 61–66, 68, 69, 168, 213, 217, 245, 247–250, 276, 301, 303
- 无应答 (及), 152–189
 - 定义无应答、中断和完整性, 159–168
 - 中断, 163, 164
 - 完整应答、部分应答与不可用应答单位, 166–168
 - 单位无应答指标, 159–162
 - 问题层面的项目无应答, 164, 165
 - 单位层面的项目无应答, 165, 166
 - 一般性问题, 153–159
 - 影响网络调查无应答的因素, 156–159
 - 激励措施, 175–177
 - 网络调查中的无应答过程, 154–156
 - 相关理论, 153, 154
- 网络调查的一般设计特征与无应答, 172–175
 - 问卷设计, 175
 - 调查发起方, 172–174
 - 主题突出性, 174, 175
- 网络调查中的激励措施, 175–177
- 无应答率和无应答偏差之间的关系, 168–172
- 联系过程的作用: 模式、时间和频率, 178–182

- 发送邀请的日期和时间, 179, 180
- 实地调查的时间长度, 180, 181
- 联系方式, 178, 179
- 联系次数及其安排, 181, 182
- 邀请函的具体内容格式, 182, 183
 - 格式与视觉设计, 185-187
 - 个性化, 183, 184
 - 提供激励与请求语气, 184, 185
 - 无应答策略, 187-198
- 在线样本库(及), 241-253
 - 购买在线样本库研究, 252
 - 数据质量, 247-251
 - 质量和结果的比较, 247, 248
 - 相关结论, 251
 - 加权对在线样本库数据质量的影响, 249, 250
 - 在线样本库研究中的问题, 250, 251
 - 未来展望, 252, 253
 - 维护, 245, 246
 - 概率在线样本库, 32
 - 招募、抽样和选择, 242-245
 - 非概率在线样本库的招募方法, 243
 - 概率在线样本库的招募方法, 243, 244
 - 在线样本库成员的抽样与选择, 244, 245
 - 研究设计, 246, 247
 - 相关术语与分类, 241, 242
- 调查后阶段, 204-221, 299
 - 数据导出与文档化, 218-221, 另见主题条目
 - 数据准备, 205-215, 另见主题条目
 - 初步结果, 216, 217
 - 发布, 217
 - 报告, 216, 217
- 预调查(及), 38-192, 299, 另见各主题条目
 - 综合管理, 189-192
 - 调查模式详述, 38-44
 - 无应答策略, 152-189
 - 问卷准备, 70-133
 - 抽样, 44-69
 - 技术准备, 133-152
- 问题类型(及), 75-101
 - 问题库, 74, 258
 - 常量和, 98
 - 特定内容的实施, 99, 100
 - 编制过程, 72, 77
 - 图形的广泛使用, 98, 99
 - 多项目, 89-96
 - 组合表格, 96
 - 包含等距与比率测量问题的表格, 95, 96
 - 包含名义测量问题的表格, 90-93
 - 包含顺序测量问题的表格, 93-95
 - 其他, 96-101
 - 包含观察数据和外部数据的问题, 100, 101
 - 排名, 97
 - 单一项目, 78-89
 - 等距和比率测量, 86-89
 - 名义测量, 78-83
 - 顺序测量, 83-86
 - 等距和比率测量问题
 - 名义测量问题, 78-83
 - 顺序测量问题, 83-86, 86-89
 - 高级图形呈现, 86
 - 下拉菜单, 86
 - 开放式数据输入, 86
 - 开放式文本输入, 78-80
 - 单选按钮, 84, 85
- 问卷准备(及), 70-133
 - (通过以下方式)确保调查的参与度, 114-121
 - 参与度、流程和用户体验, 114, 115
 - 问卷长度, 117-119
 - 调查游戏化, 116, 117
 - 调查满足, 119-121
 - 一般性问题, 见问卷准备: 一般性问题
 - 问卷编制、计算机化和测试的一体化, 130-133
 - 问题类型, 见主题条目
 - 结构, 见问卷结构
 - 网络调查问卷测试, 121-130
 - 网络问卷调查的具体测试方法, 125-130
 - 传统测试方法, 122-125
- 问卷准备: 一般性问题, 70-75
 - 计算机化、互动性和动态性, 106-111
 - 测量过程和测量误差, 73

- 问题和调查问卷设计原则, 72, 73
- 问题库, 74
- 问题编制过程, 72
- 非实质性应答的具体内容, 74
- 网络问卷调查的特殊性, 75
- 结构, 101–114
- 调查问题的类型, 70, 71
- 视觉布局, 111–114
- 排名, 97
- 比率测量, 71, 86–88
- 拒答, 152, 162
- 关于计算机辅助调查 / 电话调查研究, 15
 - 自助式 (DIY), 5, 49, 198, 255, 279, 302
 - 在线样本库与非概率样本 (美国民意研究协会工作组), 249
- 抽样 (及), 44–69, 另见定义与分类
 - 界定人群, 46
 - 非概率网络调查中的设计问题, 54–59
 - 招募, 54–58, 另见调查招募方法
 - 非概率情境下抽样设计的近似, 58, 59
 - 概率网络调查中的设计原则, 46–54
 - 样本量, 52–54
 - 抽样框, 47–51, 另见主题条目
 - 抽样技术, 51, 52
 - 无覆盖问题, 66–69
 - 在线样本库, 241
- 抽样框, 47–51
 - 电子邮件列表作为抽样框, 48–50
 - 以及单位样本库, 50, 51
 - 传统抽样框, 50
- 满足, 见问卷准备
- 网络调查实施的选定主题, 222–269, 另见各主题条目
 - 在线样本库, 241–253
 - 智能手机、平板电脑及其他设备, 223–240
 - 网络调查软件, 253–268
- 敏感问题、敏感主题, 41, 44, 102, 136, 149, 183
- 斯洛文尼亚, 34, 49, 58, 64, 93, 157, 160, 171, 174, 177, 210, 224, 280
- 智能手机、平板电脑及其他设备 (及), 223–240
 - 基于浏览器的移动调查与移动调查应用程序, 234–238
 - 挑战和结论, 239, 240
 - 为智能手机设计网络问卷, 232–234
 - 设备效应, 228–232
 - 多设备, 223–226, 另见调查
 - 多设备网络调查软件和移动调查应用程序, 238, 239
 - 处理移动设备的调查策略, 226–228
- 社会交换理论, 117, 154, 176, 184
- 软启动, 197
- 概率和非概率调查中的统计推论, 59–66
 - 关于非概率网络调查的结论性评述, 65, 66
 - 关于概率的决策, 63–65
 - 针对非概率样本, 61–63
 - 针对概率样本, 60, 61
- 研究 (关于)
 - 网络调查与纸笔问卷的应答, 26
 - 客户满意度调查和员工组织调查 (Survey Monkey, 2011), 179
 - 电子邮件调查邀请 (Tuten, 1997), 187
 - 《网络调查实施指南》(WebSM), 292
 - 荷兰国家在线样本库比较研究 (NOPVO), 247
 - 使用网络调查的在线社区 (2003), 32
 - 问卷设计 (Dillman, Tortora, Conradt and Bowker, 1998), 24
 - 社交网站, 33
- 调查招募方法, 55–58
 - 众包, 56
 - 游击式招募, 56–58
 - 网络抽样, 56
 - 在线 / 在线社交媒体, 55, 56
- 调查研究与网络调查, 1–37, 另见网络调查
 - 定义与分类, 1–7, 另见主题条目
 - 网络调查的发展、应用和相关实践, 见主题条目
 - 相关实践, 见网络调查的相关实践, 34–37
 - 网络调查过程, 7–10, 另见主题条目
- 调查满足, 见问卷准备
- 调查软件, 151, 253, 258, 267, 299
- 调查 (关于), 另见网络调查
 - 学术社交, 5
 - 商业, 5
 - 客户满意度调查, 142

- 机构, 5
- 欧洲社会调查, 5, 41
- 面对面, 4, 23, 26
- 劳动力, 280
- 电子邮件无应答 (斯洛文尼亚), 160
- 移动设备与多设备, 224–226
- 移动调查应用程序, 234–238
- 基于概率的面对面家庭 (调查), 279
- 烟草、酒精和药物 (斯洛文尼亚健康研究所), 280
- 移动设备的使用, 斯洛文尼亚 (Eurostat, 2013), 224
- WCSC 调查, 21, 181, 182
- 平板电脑, 见智能手机、平板电脑及其他设备
- 技术准备, 133–152
 - 用于数据采集与整理, 133–141
 - 数据库准备, 134, 135
 - 电子邮件设置, 137–140
 - 受访者追踪与唯一标识符, 135–137
 - 样本管理, 140–141
 - 并行数据收集, 141–146
 - 并行数据分析, 145
 - 联系信息并行数据, 142
 - 设备类并行数据, 142
 - 并行数据的隐私与伦理, 145, 146
 - 问卷导航并行数据, 143, 144
 - 用于并行数据的软件, 144, 145
- 保障安全与隐私的准备工作, 146–151
 - 访问审计, 150, 151
 - 与法规和行业标准的关系, 151
 - 确保对数据文件的物理和电子访问安全, 150
 - 保障服务器安全, 147, 148
 - 保障网络调查应答过程安全, 148, 149
 - 匿名性的技术问题, 149
- 网络调查软件的技术预测试, 151, 152
- 电话调查, 12, 15
- 测试, 另见互联网测试
 - 网络问卷调查的具体测试方法, 125
 - 传统测试方法, 122–125
 - 主题突出性, 174
- 总调查误差 (TSE), 272
- 唯一标识符, 135
- 美国 (US), 29, 145, 243, 244, 248, 259
 - 美国调查研究机构协会:《调查研究标准与伦理准则》(CASRO, 2011), 252
 - 美国调查研究机构协会 (2012): 抽奖指导方针, 177
 - 客户满意度调查, 142
 - 盖洛普样本库, 30
 - 社会综合调查, 5
 - GfK (知识网络), 30, 43
 - 美国机构审查委员会 (IRBs), 290
 - 互联网普及率, 67
 - 纸笔问卷, 30
 - 《康复法案》, 291
 - 满足, 119
 - 网络调查, 224, 225, 238
- 虚拟环境, 32, 123
- (问卷的) 视觉布局, 79, 84, 98, 101, 111–114, 121
- 网络实验, 37, 227
- 网络调查过程, 7–10
- 网络调查专业, 286–292
 - 正式认证、准则和标准, 290, 291
 - 一般性伦理考量, 289, 290
 - 非正式指南、建议和其他材料, 292
 - 法律问题, 288, 289
 - 相关学科, 287, 288
- 网络调查软件 (及), 253–268
 - 特征, 260–263
 - 讨论, 263–265
 - 图形用户界面 (GUI) 的关键特征, 265
 - 界面速度, 264, 265
 - 预期进步, 299, 300
 - 寻找合适的工具, 265–267
 - 网络调查过程的集成信息与通信技术 (ICT) 支持, 255–258
 - 市场, 258–260
 - 趋势, 267, 268
- 网络调查 (及), 10–37, 270–296, 另见网络调查的未来;
 - 网络调查的发展与网络调查专业
 - 更广泛的方法学背景, 见网络调查: 更广泛的方法学背景
 - 未来, 见网络调查的未来

- 网络调查中的激励措施, 175-177
- 非正式指南、建议和其他材料, 292, 另见 WebSM
- 网络调查缺乏人际互动 / 访员, 26
- 非基于名单的, 136, 137
- 网络调查的专属招募方法, 56, 另见众包
- 网络商业调查的特殊性, 31
- 项目管理框架, 284-286
- 由智能手机 / 平板电脑支持的, 225, 226
- 软件, 见网络调查软件
- 网络调查的参考文献, 293-296
- 网络调查中的加权, 212-214
- 网络调查 : 更广泛的方法学背景, 270-284
 - 核心网络调查过程之外 (的步骤), 277-284
 - 高级分析、处理与价值转化, 278, 279
 - 混合方法, 280, 281
 - 混合模式, 279, 280
 - 初步研究活动, 277, 278
 - 成本—误差优化, 276-277
 - 网络调查模式效应评估, 274-276
 - 总调查误差和调查数据质量, 272-274
 - 迈向交互式实地调查设计, 281-284
 - 网络调查模式的修订, 270-272
- 网络调查的相关实践, 34
 - 网络上的行政表格, 35
 - 互联网测试, 35, 36
 - 有监考与无监考测试, 35, 36
 - 心理学网络实验, 36, 37
- 网络调查工具, 见网络调查软件
- 加权, 212-214
- WebSM, 292
 - 博客, 292
 - 调查实施文档, 292
- 万维网联盟 (W3C), 11