

HZ BOOKS
华章科技

ELSEVIER
爱思唯尔

Amazon五星级畅销书，资深用户体验专家、统计分析师、心理学专家10余年工作经验结晶，着眼于用户体验设计人员工作中所遇到的疑难问题，推荐最佳解决方案

从科学的量化视角给从业者以思考，注重实战，每个知识点都辅之以设计精巧的案例，包含大量操作技巧和最佳实践

UI/UE系列丛书

用户体验度量

量化用户体验的统计学方法

Quantifying the User Experience

Practical Statistics for User Research



(美) Jeff Sauro James R. Lewis 著
殷文婧 徐沙 杨晨燕 隋馨缘 陶伯仲 译



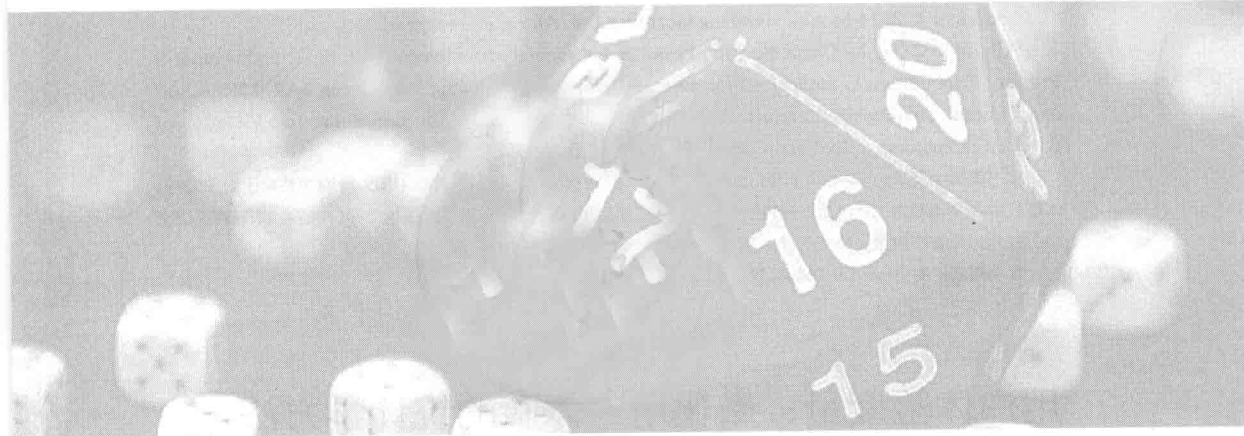
机械工业出版社
China Machine Press

用户体验度量

量化用户体验的统计学方法

Quantifying the User Experience

Practical Statistics for User Research



(美) Jeff Sauro James R. Lewis 著
殷文婧 徐沙 杨晨燕 隋馨缘 陶伯仲 译



机械工业出版社
China Machine Press

01408482

图书在版编目 (CIP) 数据

用户体验度量: 量化用户体验的统计学方法 / (美) 邵罗 (Sauro, J.) 等著; 殷文婧等译. —北京: 机械工业出版社, 2014.3

(UI/UE 系列丛书)

书名原文: Quantifying the User Experience: Practical Statistics for User Research

ISBN 978-7-111-45904-0

I. 用… II. ①邵… ②殷… III. 软件设计 IV. TP311.5

中国版本图书馆 CIP 数据核字 (2014) 第 031056 号

本书版权登记号: 图字: 01-2012-8344

Quantifying the User Experience: Practical Statistics for User Research

Jeff Sauro and James R. Lewis

ISBN 978-0-12-384968-7

Copyright © 2012 by Jeff Sauro and James R. Lewis. Published by Elsevier Inc. All rights reserved.

Authorized Simplified Chinese translation edition published by the Proprietor.

Copyright © 2014 by Elsevier (Singapore) Pte Ltd. All rights reserved.

Printed in China by China Machine Press under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macau SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to civil and criminal penalties.

本书简体中文版由 Elsevier (Singapore) Pte Ltd. 授权机械工业出版社在中国大陆境内独家出版和发行。本版仅限在中国境内 (不包括香港特别行政区、澳门特别行政区及台湾地区) 出版及标价销售。未经许可之出口, 视为违反著作权法, 将受法律之制裁。

本书封底贴有 Elsevier 防伪标签, 无标签者不得销售。



用户体验度量: 量化用户体验的统计学方法

(美) Jeff Sauro 等著

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 秦 健

印 刷: 北京瑞德印刷有限公司

版 次: 2014 年 4 月第 1 版第 1 次印刷

开 本: 170mm × 242mm 1/16

印 张: 17.75

书 号: ISBN 978-7-111-45904-0

定 价: 69.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有 · 侵权必究

封底无防伪标均为盗版

本书法律顾问 北京大成律师事务所 韩光 / 邹晓东

译者序

首先感谢机械工业出版社编辑和我们亲爱的同事周薇将此书介绍给我们，同时也带给了我们一次书籍翻译的经历。

时下用户体验的概念甚为时兴，相关书籍俯拾即是，但作为一名从业者，想找到一本真正对实践有指导意义的资料却十分困难。看到此书时我们甚为惊喜：作者为业界资深专家，有丰富的从业经验；书中的讨论重点着眼于用户体验设计人员工作中所遇到的疑难问题。由此，我们欣然接下了本书的翻译任务，也希望将其介绍给国内更多关注用户体验的朋友。

译者小团队有5名成员，皆为腾讯公司的用户体验从业者。团队成员来自上海、深圳、成都三地，部分成员虽是同事但之前少有机会合作，都因此书结缘而走到一起。译者团队有用户研究人员，也有设计师，从业经验最长6年，最短两年。其中两名成员还是腾讯公司用户体验与设计部“CDC 翻客”团队的主力成员。

写书难，译书亦不简单。互联网行业的从业者工作之余的时间少得可怜，加之本书涉及众多统计概念和推导分析，极大地增加了翻译工作的难度。各位译者迎难而上，牺牲了周末、年假、工作中的休息时间，持续4个月进行翻译、交叉校对、讨论、统稿，最终完成本书翻译工作。其中辛苦只有自知。凭借团队成员的知识背景、工作经验和辛勤付出，相信我们为读者提供的是一份有诚意的高质量译本。

在此我们也向团队成员互表感激，感谢大家在这4个月中的付出和精诚合作。同时感谢鼓励我们的家人、团队领导和同事们。

由于时间有限，工作难免有纰漏之处，期待读者们提出反馈，帮助我们修正。我们的邮箱地址：Quantify.team@QQ.com。

以下是团队成员的翻译感言，排序不分先后。

殷文婧：翻译本书，感触最深的是作者面对问题的坚持精神和实证方法。我认为这正是自己对这本书和作者的敬佩之处，也是业界对用户研究从业者的“专业”给予认可和尊重之源。

徐沙：从事用户体验相关工作以来，我不断对自己原有学术的“科学方法体系”进行调整。类似可用性测试这类方法，如何科学衡量结果，一直困扰着我们。“科学的一个本质特征

是发展新的创意以及将新老想法进行实证测试”，在实际工作中我们采用可用性测试进行评估，但往往忽视了对结果的评估，导致可用性测试变成了信息搜集而不能直接帮助决策。当然评估的标准不限于本书，本书的价值在于从科学的量化视角给从业者以思考，本书推荐的很多问卷也可以在实践中尝试使用。

杨晨燕：用户研究从业者在研究过程中常会遇到这样的问题和挑战：“测试多少样本足以揭示问题？”“小样本用户研究结果能说明总体情况吗？”本书针对用户研究中的这些实践性问题，全面深入地梳理了作者及国际上其他最前沿的研究结果，并结合自己的工作经验，推荐了最佳解决办法，特别是在可用性测试和测量方面。对用户研究从业者或关注用户体验的人来说，相信本书能够带给大家帮助和启发。

隋馨缘：在平日的工作中，常常会遇到资源紧缺的问题：时间资源、样本资源、财务预算。你甚至会怀疑在多大程度上对自己的结论有信心。统计学能够帮助你评估，在你能承受的误差范围之内需要多少资源。也能让你审视自己的资源，更有效地利用资源来帮助你做出决策。

陶伯仲：作为设计师，我们一直倡导“以用户需求出发”来做设计，但在实际工作中往往由于时间、资源上的限制只能在没有用户测试数据的情况下自己拍板。由于对“需要多少样本？”“结果可不可信？”等问题心里没底，普通设计师想要自己搜集用户信息时还是举步维艰。本书作者根据自身多年的工作经验，深入浅出地介绍了不同情况下可用性测试中的数据搜集方法和衡量方式，对于想用心做用户喜欢的产品的设计师来说，是一本不可错过的指南和工作手册。最后，还想对翻译过程中一起辛苦熬夜的同事和家人说一声：“谢谢！辛苦了！”

致 谢

非常感谢 Elisa Miller、Lynda Finn、Michael Rawlins、Barbara Millet、Peter Kennedy、John Romadka 和 Arun Martin，感谢他们对本书各章节初稿进行的细致复核。对于他们付出的时间和给予的有价值的建议，我们表示深深的感激。

Jeff 的致谢

这本书代表了我 10 年来研究、重新取样和阅读许多学科期刊文章的结果，这些工作都是为了帮助回答一个激动人心的领域中的问题。在这个过程中，我不仅对找到的答案感到满意，而且学到了很多，遇到了很多人，尤其是我的合著者——Jim Lewis。感谢我的家人在著书过程中给予我的耐心和鼓励。

Jim 的致谢

写一本书要拿走一大块你的生活。非常幸运，我的家人容忍了我的这种痴迷。感谢我的妻子 Cahty，感谢她的耐心和爱的支持。我的孩子们，Mickael 和 Patrick——你们又可以来我的办公室啦。

作者介绍

Jeff Sauro 是一名接受过 6 sigma 培训的统计分析师，也是 Measuring Usability 公司的主要创始人。15 年来，他一直为诸如 PayPal、Walmart、Autodesk 和 Kelley Blue Book 这类公司进行可用性项目和统计分析，也曾在 Oracle、Intuit 和 General Electric 等公司工作。

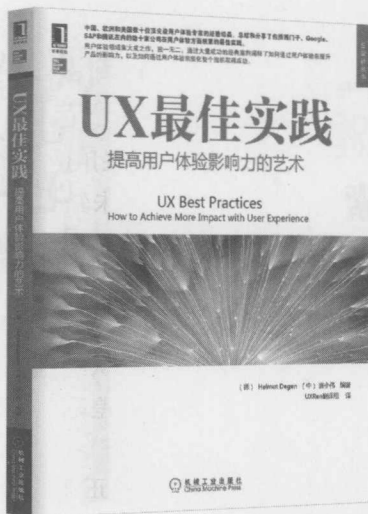
Jeff 发表过超过 15 篇同行评审的研究文章，他也是《Journal of Usability Studies》的编辑委员。另外，Jeff 也是人机交互（Computer Human Interaction, CHI）和可用性专家协会大会（Usability Professionals Associations, UPA）的常任演讲者和讲师。

Jeff 在斯坦福大学获得了设计和技术专业的硕士学位，期间主要研习统计概念。在斯坦福大学之前，他在锡拉丘兹大学获得了信息管理与技术专业以及电视广播电影专业的双学士学位。如今 Jeff 和他的妻子以及 3 个孩子生活在丹佛。

James R. (Jim) Lewis 博士是一名资深的人因工程师（1981 年起在 IBM 工作），现今主要从事语音应用的设计和评估，他也是《Practical Speech User Interface Design》一书的作者。他是认证的人因专家，获得了实验心理学（语言心理学）的博士学位和工程心理学硕士学位，另外也是音乐理论和作曲专业的音乐硕士。Jim 是国际公认的可用性测试和测量方面的专家。他（受邀）撰写了《Handbook of Human Factors and Ergonomics》第 3 版和第 4 版的可用性测试章节，并在各种专家会议上演讲关于可用性测试和度量指标的内容。

Jim 是 IBM 大师级的发明家，拥有 77 项美国专利局颁发的专利。他最近在《International Journal of Human - Computer Interaction》和《Journal of Usability Studies》的编辑委员会工作，同时也在老龄化与技术增强研究与交互中心（Center for Research and Education on Aging and Technology Enhancement, CREATE）的科学咨询委员会工作。他也是可用性专家协会（Usability Professionals Association, UPA）、人因和工效学会（Human Factors and Ergonomics Society, HFES）、心理科学协会（Association for Psychological Science, APS）和美国心理协会（American Psychological Association, APA）的会员。另外，Jim 还是跆拳道黑带 5 级和美国跆拳道协会（American Taekwondo Association, ATA）的认证导师。

推荐阅读



UX最佳实践：提高用户体验影响力的艺术

作者：（德）Helmut Degen ISBN: 978-7-111-41108-6 定价：59.00元

中国、欧洲和美国数十位顶尖级用户体验专家的经验结晶，总结和分享了包括西门子、Google、SAP和腾讯在内的数十家公司在用户体验方面积累的最佳实践。

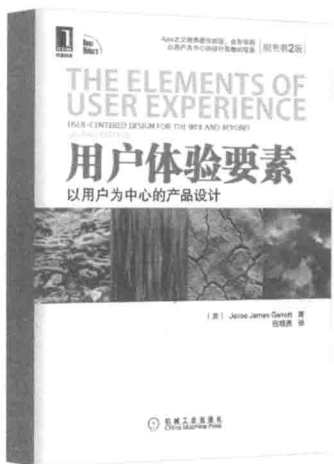
用户体验领域集大成之作，独一无二，通过大量成功的经典案例阐释了如何通过用户体验来提升产品的影响力，以及如何通过用户体验来催化整个组织取得成功。

本书探讨了在世界不同的地区中，用户体验活动是如何整合到多种产品的开发中，并最终减少产品复杂度、提高产品质量和增加收益的。本书涵盖多种用户体验方法，例如分析用户需求和期望、进行概念设计、创建原型、使用敏捷开发、进行可用性测试、创建界面设计原则、定义界面设计模式，以及设定用户体验指标等。

本书内容：

- 在设计大纲中，传达设计目标和用户需求。
- 建立端到端的以用户为中心的开发流程。
- 促进管理层、设计师及工程师之间的协作。
- 将用户体验指标融入商业目标框架及产品开发流程中。
- 采用敏捷开发流程和设计思维的方法。
- 收集、测量并分析可用性指标。
- 采用用户体验评估系统来识别问题。
- 通过原型快速传达和评判设计思想。
- 通过界面设计模式和模式库达成跨产品的一致性。

推荐阅读



用户体验要素：以用户为中心的产品设计（原书第2版）

作者：Jesse James Garrett ISBN: 978-7-111-34866-5 定价：39.00元



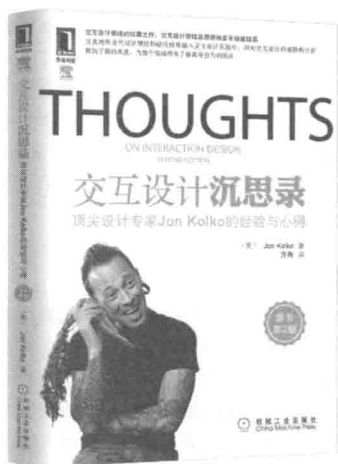
认知设计

作者：Julie Dirksen ISBN: 978-7-111-38832-6 定价：69.00元



优秀网站设计：打造有吸引力的网站（原书第3版）

作者：Patrick J. Lynch等 ISBN: 978-7-111-39959-9 定价：69.00元



交互设计沉思录：顶尖设计专家Jon Kolko的经验与心得（原书第2版）

作者：Jon Kolko ISBN: 978-7-111-39678-9 定价：49.00元

目 录

译者序	1
致谢	1
作者介绍	1
第 1 章 导论	1
1.1 简介	1
1.2 本书的组织结构	1
1.3 如何使用本书	2
1.3.1 应该使用何种检验	2
1.3.2 我需要多大的样本量	5
1.3.3 你不必手动计算	6
1.4 本章要点	7
1.5 本章思考题	7
1.6 参考资料	8
第 2 章 量化用户研究	9
2.1 什么是用户研究	9
2.2 用户研究的数据	9
2.3 可用性测试	10
2.3.1 样本量	10
2.3.2 代表性和随机性	10
2.3.3 数据收集	12
2.3.4 任务完成率	12
2.3.5 可用性问题	13

2.3.6	任务时间	14
2.3.7	出错数	14
2.3.8	满意度评分	14
2.3.9	复合分数	14
2.4	A/B 测试	15
2.5	调查数据	15
2.5.1	等级量表	15
2.5.2	净推荐值	16
2.5.3	评论和开放性数据	16
2.6	需求收集	16
2.7	本章要点	17
2.8	参考资料	17
第3章 我们的估算到底有多准确		19
3.1	简介	19
3.1.1	置信区间 = 误差幅度的两倍	19
3.1.2	置信区间提供了精确度和位置	20
3.1.3	置信区间的三个组成部分	20
3.2	完成率的置信区间	20
3.2.1	置信区间的历史	21
3.2.2	Wald 区间：对于小样本来说就太不靠谱了	21
3.2.3	精确置信区间	22
3.2.4	Wald 校正区间：增加两个成功与两个失败	22
3.2.5	完成率的最佳点估计	24
3.2.6	遭遇可用性问题的置信区间	26
3.3	等级量表和其他连续性数据的置信区间	26
3.3.1	任务时长数据的置信区间	29
3.3.2	任务时长均值还是任务时长中位数	30
3.3.3	几何均值	30
3.3.4	大样本任务时长的置信区间	32
3.3.5	围绕中位数的置信区间	33
3.4	本章要点	34
3.5	本章思考题	35

3.6 参考资料	36
第4章 我们达到或超过目标了吗	38
4.1 简介	38
4.2 单侧检验和双侧检验	40
4.3 完成率与基准的比对	41
4.3.1 小样本检验	42
4.3.2 大样本检验	44
4.4 满意度分数与基准的比对	46
4.5 任务时间和基准的比对	50
4.6 本章要点	54
4.7 本章思考题	54
4.8 参考资料	57
第5章 不同设计之间有统计学差异吗	59
5.1 简介	59
5.2 比较两个平均值（等级量表和反应时）	59
5.2.1 被试内设计比较——配对 t 检验	60
5.2.2 比较任务时长	62
5.2.3 组间比较（双样本 t 检验）	64
5.2.4 t 检验的假设	68
5.3 比较完成率、转换率以及 A/B 测试	69
5.3.1 组间比较	69
5.3.2 组内比较	78
5.4 本章要点	86
5.5 本章思考题	88
5.6 参考资料	94
第6章 我们需要多大的样本量 第一部分：总结性研究	95
6.1 简介	95
6.1.1 我们为何要关注	95
6.1.2 可用性研究的类型至关重要	96
6.1.3 总结性可用性测试样本量预估的基本原则	96

6.2	预估数值	97
6.3	比较数值	104
6.4	如何控制变异性	109
6.5	二项置信区间样本量的估计	110
6.5.1	大样本的二项样本量估计	110
6.5.2	小样本的二项样本量估计	112
6.5.3	与基准比例相比较的样本量	115
6.6	卡方检验的样本量预估 (独立比例)	117
6.7	MCNEMAR 精确检验的样本预估 (配对比例)	120
6.8	本章要点	123
6.9	本章思考题	124
6.10	参考资料	130
第7章 我们需要多大的样本量 第二部分: 形成性研究		131
7.1	简介	131
7.2	使用发现问题的概率模型来估计形成性用户研究的样本量	131
7.2.1	著名方程: $P(x \geq 1) = 1 - (1 - p)^n$	131
7.2.2	从 $1 - (1 - p)^n$ 中推导出样本量估计方程	133
7.2.3	使用表格计划形成性用户研究样本量	135
7.3	二项概率模型的假设	136
7.4	模型的附加应用	137
7.4.1	估计多重问题或其他事件的复合 p 值	137
7.4.2	校正小样本 p 的复合估计值	138
7.4.3	估计可发现的问题数和未被发现的问题数	143
7.5	影响 p 值的是什么	144
7.6	什么是合理的目标问题发现率	145
7.7	调解“神奇的数字5”和“8还不够”	147
7.7.1	一段历史: 20世纪80年代	147
7.7.2	又一段历史: 20世纪90年代	148
7.7.3	“神奇的数字5”的起源	149
7.7.4	“8还不够”: 一个调解方法	151
7.8	更多关于二项概率公式和其小样本校正	155
7.8.1	二项概率公式的起源	155

7.8.2	紧缩校正是如何起作用的	156
7.9	针对问题发现的其他统计模型	159
7.9.1	对问题发现使用二项式模型的批评	159
7.9.2	扩展的二项式模型	160
7.9.3	Capture-recapture 模型	161
7.9.4	在计划形成性用户研究时为什么不用其他模型	161
7.10	本章要点	165
7.11	本章思考题	166
7.12	参考资料	168
第8章	标准化的可用性问卷	171
8.1	简介	171
8.1.1	什么是标准化的问卷	171
8.1.2	标准化可用性问卷的优点	171
8.1.3	什么样的标准化可用性问卷是有用的	172
8.1.4	标准化问卷的质量评估: 信度、效度和灵敏度	173
8.1.5	问卷的步距数	174
8.2	整体评估问卷	174
8.2.1	QUIS (用户交互满意度问卷)	175
8.2.2	SUMI (软件可用性测试问卷)	176
8.2.3	PSSUQ	178
8.2.4	SUS (软件可用性问卷)	182
8.2.5	可用性整体评估问卷的实验比较	194
8.3	任务评估问卷	197
8.3.1	场景后问卷	197
8.3.2	单项难易度问卷	198
8.3.3	主观脑力负荷问题	198
8.3.4	期望评级	199
8.3.5	可用性等级评估	201
8.3.6	任务评估问卷的实验比较	203
8.4	网站感知可用性的评估问卷	205
8.4.1	网站分析和测量问卷	206
8.4.2	标准通用的百分等级问卷	207

8.4.3	其他评估网站的问卷	209
8.5	其他有趣的问卷	210
8.5.1	计算机系统可用性问卷	210
8.5.2	有用性、满意度、易用性	211
8.5.3	用户经验的可用性度量	211
8.5.4	享受性质量	212
8.5.5	美国消费者满意度指标	213
8.5.6	净推荐值	213
8.5.7	福雷斯特客户体验指数	215
8.5.8	技术接受模型	215
8.6	本章要点	216
8.7	本章思考题	217
8.8	参考资料	220
第9章 测量和统计的六大持久论战		224
9.1	介绍	224
9.2	对多点量表数据进行平均合理吗	225
9.2.1	一方观点	225
9.2.2	另一方观点	226
9.2.3	我们的推荐	228
9.3	需要测试至少30名用户吗	229
9.3.1	一方观点	229
9.3.2	另一方观点	230
9.3.3	我们的推荐	231
9.4	所有的实验都要进行双侧检验吗	231
9.4.1	一方观点	231
9.4.2	另一方观点	232
9.4.3	我们的推荐	233
9.5	当 $p > 0.05$ 时,我们能拒绝原假设吗	233
9.5.1	一方观点	233
9.5.2	另一方观点	234
9.5.3	我们的推荐	235
9.6	能将各种可用性度量指标合并到一个分数中吗	236

9.6.1	一方观点	236
9.6.2	另一方观点	238
9.6.3	我们的推荐	239
9.7	假使你需要进行多次检验该怎么办	239
9.7.1	一方观点	239
9.7.2	另一方观点	241
9.7.3	我们的推荐	241
9.8	本章要点	245
9.9	本章思考题	246
9.10	参考资料	249
第 10 章	总结	251
10.1	简介	251
10.2	更多信息	251
10.3	好运	254
10.4	本章要点	254
10.5	参考资料	254
附录 A	基础统计概念速成	255

第1章 导 论

1.1 简介

统计是很多用户体验行业设计师和研究员最不愿考虑的事情。实际上，之所以很多从业人员认为这个行业有吸引力，正是因为他们认为可以很大程度避免接触没有人情味的数字。他们认为，如果可用性和设计是定性的活动，就可以安全跳过公式和数字。

虽然设计和一些可用性活动确实是定性的，但是好坏设计的效果可以容易地通过转换量、完成率、完成时间、感知的满意度、推荐值和销售量来量化。越来越多的可用性从业者和用户研究人员期望量化这项工作的收益。如果他们不这么做，其他人将去做——遗憾的是，其他人可能会错误地使用度量标准或方法。

1.2 本书的组织结构

本书为那些测量人与界面交互时的行为与态度的读者设计。本书不涉及可能被发现有部分用处的抽象数学理论。相反，本书关乎用户研究项目实施、分析、汇报时所遇最常见疑虑和问题的分析。总体上，这些活动分成三个领域：

- 1) 汇总数据及计算误差幅度（第3章）。
- 2) 与基准比较（第4章）或组间比较（第5章）时判定是否存在统计显著的差异。
- 3) 找到适合研究的样本量（第6章和第7章）。

我们还提供：

- 综述量化用户研究常用方法的背景章节（第2章）及很多基础统计概念的快速介绍/回顾（附录）。
- 对标准化可用性问卷的广泛讨论（第8章）。
- 对持久的统计争论的讨论，用户研究者应该了解这些争论并能够为他们的