



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。

ELSEVIER

UI/UE系列丛书

# Quantifying the User Experience

Practical Statistics for User Research , 2nd Edition

# 用户体验度量

## 量化用户体验的统计学方法

(原书第2版)

[美] 杰夫·绍罗 (Jeff Sauro) 詹姆斯 R. 路易斯 (James R. Lewis) 著 顾盼 译



海外借



机械工业出版社  
China Machine Press

Quantifying the User Experience

Practical Statistics for User Research, 2nd Edition

# 用户体验度量

量化用户体验的统计学方法

(原书第2版)

[美] 杰夫·绍罗  
(Jeff Sauro)

詹姆斯·R·路  
(James R. Lewis)



著

顾盼 译



机械工业出版社  
China Machine Press

## 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 (原书第 2 版) / (美) 杰夫·绍罗 (Jeff Sauro), (美) 詹姆斯 R. 路易斯 (James R. Lewis) 著; 顾盼译. —北京: 机械工业出版社, 2018.1  
(UI/UE 系列丛书)

书名原文: Quantifying the User Experience: Practical Statistics for User Research, 2nd Edition

ISBN 978-7-111-58965-5

I. 用… II. ①杰… ②詹… ③顾… III. 统计方法—应用—人—机系统—系统设计 IV. TP11

中国版本图书馆 CIP 数据核字 (2018) 第 014508 号

本书版权登记号: 图字 01-2016-7274

Elsevier (Singapore) Pte Ltd.  
3 Killiney Road, #08-01 Winsland House I, Singapore 239519  
Tel: (65) 6349-0200; Fax: (65) 6733-1817

Quantifying the User Experience: Practical Statistics for User Research, 2nd edition  
Jeff Sauro, James R Lewis  
© 2016 Jeff Sauro and James R Lewis. Published by Elsevier Inc. All rights reserved.  
ISBN-13: 978-0128023082

This translation of Quantifying the User Experience: Practical Statistics for User Research, 2nd edition by Jeff Sauro, James R Lewis was undertaken by China Machine Press and is published by arrangement with Elsevier (Singapore) Pte Ltd.

Quantifying the User Experience: Practical Statistics for User Research, 2nd edition by Jeff Sauro, James R Lewis 由机械工业出版社进行翻译, 并根据机械工业出版社与爱思唯尔 (新加坡) 私人有限公司的协议约定出版。

《用户体验度量: 量化用户体验的统计学方法》(原书第 2 版)(顾盼译)

ISBN: 978-7-111-58965-5

Copyright © 2018 by Elsevier (Singapore) Pte Ltd.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from Elsevier (Singapore) Pte Ltd. Details on how to seek permission, further information about the Elsevier's permissions policies and arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: [www.elsevier.com/permissions](http://www.elsevier.com/permissions).

This book and the individual contributions contained in it are protected under copyright by Elsevier (Singapore) Pte Ltd. and China Machine Press (other than as may be noted herein).

### 注意

本书涉及领域的知识和实践标准在不断变化。新的研究和经验拓展我们的理解, 因此须对研究方法、专业实践或医疗方法作出调整。从业者和研究人员必须始终依靠自身经验和知识来评估和使用本书中提到的所有信息、方法、化合物或本书中描述的实验。在使用这些信息或方法时, 他们应注意自身和他人的安全, 包括注意他们负有专业责任的当事人的安全。

在法律允许的最大范围内, 爱思唯尔、译文的原文作者、原文编辑及原文内容提供者均不对因产品责任、疏忽或其他人身或财产伤害及/或损失承担责任, 亦不对由于使用或操作文中提到的方法、产品、说明或思想而导致的人身或财产伤害及/或损失承担责任。

This edition is printed in China by China Machine Press under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong SAR, Macau SAR and Taiwan. Unauthorized export of this edition is a violation of the contract.

本书简体中文版由 Elsevier (Singapore) Pte Ltd. 授权机械工业出版社在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 出版与发行。未经许可之出口, 视为违反著作权法, 将受民事和刑事法律之制裁。

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 (原书第 2 版)

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: 张锡鹏

责任校对: 李秋荣

印刷: 北京市兆成印刷有限责任公司

版次: 2018 年 3 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 20.25

书号: ISBN 978-7-111-58965-5

定价: 79.00 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) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: [hzit@hzbook.com](mailto:hzit@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

HZBOOKS | 华章IT | Information Technology



## 译者序

这是一次奇妙的翻译之旅，每翻开新的一页，都会为作者详尽的解析而赞叹。感谢机械工业出版社，使我拥有了这宝贵的经历。

我们身边时时刻刻都发生着交互，人与人、人与物、人与环境、物与环境等。这些交互给我们留下了一段段体验，形成了我们对事物的认知。体验很重要，也无处不在，然而，自从用户体验提出以来，它就不是一个容易解读的概念。它看不见、摸不着，没有实在的产物。如何评估用户体验，更是一个难题。

2012年出版的本书第1版是一本系统阐述如何量化和评估用户体验的书籍。它利用统计学的知识，使用户体验从主观感受变成了一个可以量化的参数，在无形和有形之间搭建起了桥梁。这也使本书成为用户体验领域的必读书籍之一，有着特殊的地位。感谢第1版的译者殷文婧、徐沙、杨晨燕、隋馨缘和陶伯仲——五位来自腾讯公司的资深用户研究人员和设计师。他们专业的知识背景、丰富的工作经验和辛勤的付出让本书第一时间拥有了高质量的中文版。在我翻译第2版的过程中，他们的翻译也给我带来了许多指导性的建议。本书作者没有停止对统计学和用户体验的科学研究，用户体验实践最为活跃的互联网领域也发展得很快。此次再版，作者增加了近几年用户体验度量的新方法，并结合互联网的发展，尤其是在移动互联网领域的实践，为我们带来了新的解读。

数据科学在我的工作经历中也十分重要。通过数据，我们观察每一个功能的表现，分析用户的使用流程，评估产品的用户体验。科学系统的用户体验度量不仅可以使我们在做产品设计的时候有理有据，还可以促进问题的发现和解决。希望本书可以帮助更多的从业者了解统计学在用户体验度量中的应用，丰富个人和团队的技术，提高产品的用户体验。

刘司南、李祯和金煜曦也参与了此次翻译工作。本书中文版的问世，少不了她们的努力和支持。以下是她们在翻译过程中的一些感想。

在实际工作中经常会由于项目紧张而无法进行充分的用户研究和分析。如果有机会进行用户调研，定性分析也往往是大多数。如今数据驱动产品成为趋势，于是设计师从数据的角度发现和分析问题的能力变得非常重要。也就是说，用户研究中的定量分析也越来越值得重视。本书作者围绕量化用户体验的主题，结合在理论和实践中的不断积累，内容从第1版到第2版得到了极大地丰富，

推荐了许多新的分析和解决方法。相信本书的再版能给从业人员带来更多的启发。

——刘司南

本书运用统计学的方法，详细介绍了用户研究中各种定量测量方法以及如何评估通过这些方法得到的参数。用户体验相关从业者能从本书中获得非常多的理论知识，借此来评估我们所收集的数据，并得出提高产品体验的方法。甚至企业的决策者也可以从中得到启发，做出更符合用户需求的明智决定。

——李楨

通过参与本书的翻译工作，我对用户体验以及测量方法有了更深入的了解。以测试用户认知和满意度为目标的标准化问卷有许多种类型，量化问卷结果的统计方式种类也是极尽丰富的。不同类型的网站，要选用恰当的问卷和统计方法。另一方面，问卷的选项和用语也是一门学问，需要结合心理学等学科的理论，根据不同类型的网站和用户来设计。

——金煜曦

qux\_cn@163.com 是我专门为本书设立的电子邮箱，如果您有任何关于本书的建议和批评都欢迎发给我。我渴望得到您的反馈，并会跟进本书的勘误和改进。

另一方面，如果有机会通过此次翻译认识更多从事用户体验工作或对本书感兴趣的朋友，也将是我极大的收获。我十分期待和大家交流，任何读书心得和想法，也非常欢迎给我发邮件。

顾盼

2017年10月

## 作者介绍

Jeff Sauro 博士是一位经过六西格玛 (six-sigma) 训练的统计分析师，也是 MeasuringU 公司的首席创始人。15 年来，他一直在 Google、eBay、Walmart、Autodesk、Lenovo 和 Dropbox 等公司中推行可用性项目并进行统计分析，也曾在 Oracle、Intuit 和 General Electric 等工作。

Jeff 发表过超过 20 篇文章，出版过 5 本专业书籍，其中包括大家所熟知的《Customer Analytics for Dummies》。此外，他每周在 measuringu.com 网站上发表一篇与用户体验和测量相关的文章。

Jeff 在丹佛大学获得研究方法和统计学的博士学位；在斯坦福大学获得设计和技术的硕士学位；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。目前，Jeff 和他的妻子以及 3 个孩子生活在丹佛。

James R. (Jim) Lewis 博士是一位资深人因工程师 (1981 年起在 IBM 工作)，目前专注于用户体验的测量和评估。他是认证的人因专家，取得实验心理学 (语言心理学) 的博士学位和工程心理学的硕士学位，以及音乐理论和作曲专业的音乐硕士学位。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，在各类专家会议上演讲关于可用性测试和度量指标的内容，并担任 2014 年 HCII 大会的主讲人。他是 Simon——如今被认为是历史上的第一款智能手机的首席交互设计师。此外，Jim 也是书籍《Practical Speech User Interface Design》的作者。

Jim 是 IBM 大师级的发明家，拥有 88 项美国专利局颁发的专利。目前他担任《International Journal of Human-Computer Interaction》的编委以及《Journal of Usability Studies》的联合主编，并任职于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。同时，他是多个相关专业领域国际权威协会的会员，包括可用性专家协会 (Usability Professionals Association, UPA)、人因和工效学会 (Human Factors and Ergonomics Society, HFES) 和人机交互 ACM 特别兴趣小组 (ACM Special Interest Group in Computer-Human Interaction, SIGCHI)。他曾担任语音交互设计协会 (Association for Voice Interaction Design, AVIXD) 的主席。另外，Jim 是跆拳道黑带 5 段和美国跆拳道协会 (American Taekwondo Association, ATA) 的认证导师。



## 推 荐 序

当2012年本书第1版出版后，我立刻订购了一本。我和本书的两位作者很熟，他们热衷于用户体验研究，经常发表颇有见解的专业文章。当我在《Journal of Usability Studies》担任主编的时候，他们都是该杂志的审稿编辑。我常常发送这些专业文章的初稿给他们审核和修订，在工作中，他们总是使用大量的统计学方法。

在等待这本书送达的时候，我开始思考：在用户体验行业是否也需要我们自己的基础统计学书籍？最近有数十本为心理学家、社会科学家、工程师、市场和医学研究人员等撰写的统计学方法书籍，我们为什么不直接从中选择一本最好的呢？我们真的需要那么多种不同的方式来解释该怎么做 $t$ 检验吗？

然后我意识到我可能问了一个错误的问题。首先我们应该思考为什么会有这么多关于统计学的书籍。我坚信的答案是，统计学的概念和公式在不考虑背景因素的情况下只是抽象概念。在什么时候需要进行一个 $t$ 检验，又用哪种方式来检验都要基于这个背景因素。这个在今天更是如此，因为实在有太多的软件和工具可以帮我们完成基础的公式化计算了。

很多学生害怕学习统计学的基础概念，我也像个本科生一样有同样的恐惧。而且如果由另一个学科的老师来为该学科编写书籍，例如由一位统计学老师来为用户体验编写统计学书籍，那么学习这些概念将会变得更难。所以，我相信如果有一本书可以基于用户体验来学习相关的统计学概念，将会使学习变得容易。

那么本书在基于用户体验这个背景来表达统计学概念的方面做得有多好呢？当我使用了第1版并且阅读了第2版后，我的答案是非常好。几乎在每个段落中，作者都使用统计学来阐明用户体验的问题。例如，在关于 $t$ 检验的章节中，比较两种评级量表的方式或可用性测试任务时间的时候，就很好地利用了统计学来揭示：如何通过形成性和总结性的研究，以及在产品开发的阶段确定样本大小。纵览全书，没有任何一处脱离背景因素的抽象讨论。

作者也很善于系统化地组织全书的内容，以方便学生和老师使用。例如，作者在第1章中提供了一系列的决策导图，以图形的方式为后续章节铺垫了背景。而第1章中的表格则将这些决策与本书中的相关部分进行了结合。希望我在授课的时候，这样的内容组织方式也是可行的。另外，每章的最后有本章要点、本章思考题及其答案。本书的作者花费了大量的时间和精力来组织全书的结构，以帮助读者更好地理解。这很好地体现了本书英文书名副书名中的“实用”（practical）一词。



我喜欢本书的写作风格。作者以一个用户体验从业者的姿态向另一位用户体验从业者娓娓道来，而不是像一位专家在为新手写书。每当作者谈到有关他个人轶事的问题时，都会使用“根据文档记载来看……”这样很友好的表达。这也增强了内容的丰富性和个性，使阅读变得非常有趣，感觉像正在和作者聊天。

值得一提的是，作者在第2版中引用了一些新的素材。如我们所知，在写作第1版的时候，用户体验文献中还鲜有关于任务测试和实验后评估量表的讨论，所以当时书中包含了大量关于主观测量的章节。在第2版中，作者不再需要如同我们过去做的那样，自己去创建评估量表，而是将那些在第1版发行以后出现的评估量表使用在了书中。例如，1996年公布的10项软件可用性量表(SUS)激起了大家对实验后快速评估用户满意度的兴趣。它出现在可用性测试已经成为最被认可的依托经验主义的可用性评估方法的时候，于是SUS在当时成为实验后最受欢迎的评估量表，人们希望利用它来缩短开发时间。而如今又有了几个可靠的快速评估量表可以用在任务测试后和实验后，本书不仅介绍了所有这些量表，还对它们进行了比较。

第10章是作者在第2版中新增的一章，其中介绍了相关分析、回归分析和方差分析这三种统计学的检验方法。有些讽刺意味的是，正如作者指出的，统计学往往就蕴藏在最古老的测量方法中。该章最让我喜爱的是作者整合这三种检验方法的方式，非常好地展现了三者之间的相关性。在我以往学习和教授统计学课程的过程中，我不曾把它们联系在一起，特别是方差分析向来被看作是与相关分析和回归分析无关的。

关于这本书，我还有一个问题是想问自己的：在10年甚至15年前，就有人可以为我们的专业——用户体验撰写这样的一本统计学书籍。可它为什么偏偏诞生在现在呢？首先，我相信这样的等待是值得的。我们使用的这些统计学方法正在日趋成熟，而对这些方法研究的积累也使其与本书更加相关。其次，自2000年以来，信息行业 and 用户体验领域已经有了长足的发展，这帮助我们更多地了解到统计学如何受到背景因素的影响。显而易见，互联网改变了一切。网站和网页本身已经成为一门学科，使一些新的方式成为可能，例如远程和无人机测试。同时也诞生了一些新的检验方法，例如点击路径和转换率的研究，以及对A/B测试新的探索。所有这些内容都可以在本书中读到。

互联网的另一个重大影响是丰富了我们对人机交互的思考方式。以前我们所关注的可用性、生产力和问题定位法，目前已经应用到了用户体验领域，这也激起了我们对于探索新检验方式的兴趣。本书也包含了这些新的检验方式，我们在第2版中进行了更新。例如，你将会在书中看见关于客户忠诚度净推荐值的讨论，愉悦程度的享受性度量，以及全流程用户体验的可用性等级评估。

毫无疑问，当我需要一些关于用户体验度量的新灵感时，我会参考本书第2版。在我之后的统计学教学过程中，本书也会是我的第一选择。如果这是你关于这方面的第一本书，相信我，你会从中受益良多。

Joseph Dumas

用户体验设计顾问

# 前 言

欢迎阅读本书第2版！在这一版中，我们自始至终致力于提供关于如何检验和分析用户研究数据的最新且实用的信息。第2版的主要变化如下：

- 我们重新审阅了2012年以来的相关文献，并根据需要更新了内容和参考文献列表。
- 在本书涉及的话题中，自2012年以来在标准化的可用性问卷方面，有了最新的研究成果发表，所以我们全面更新了第8章的内容。
- 关于第1版，反馈最多的是希望加入相关性分析的内容。所以在第2版中我们增加全新的一章来介绍相关性分析、回归分析和方差分析。
- 由于全新的第10章，我们调整了第1章中的决策导图，以引导读者更好地理解书中的主题。

撰写本书的一个最主要目的是为尽可能广泛的用户体验设计师和研究人员提供切实可行的用户体验评估方法。我们从熟悉用户的研究人员最常遇到的数据着手，紧接着介绍适合分析这类数据的方法——显著性检验、置信区间和样本大小评估等。这其中包含一些较为常见的检验方法，例如  $t$  检验。然而，有一些检验方法是我们希望被纳入统计学的教学中，但目前还没有的，例如“调整的Wald二项分布比例的置信区间”（adjusted-Wald binomial confidence interval）。

我想我们可以做到的是，在基于背景因素的情况下，提供有理有据的方法建议。当遇到难以给出建议的案例时，我们也希望可以提供有效的信息，以帮助人们更好地做出决策。所有的这一切都是希望能够为创造更好的产品和服务出一份力，为使用它们的人们带来益处，为提供它们的企业带来价值。

# 致 谢

非常感谢 Morgan Kaufmann 的团队邀请我们进行第 2 版的撰写。非常感谢 Lynda Finn、Linda Lior 和 Phil Kortum 对本书各章节初稿深思熟虑地审校。对于他们付出的宝贵时间和给予的重要建议，我们表示深深的感激。

## Jeff 的致谢

本书代表了近 10 年来重新采样和研究的成果。我们阅读并参考了来自各个学科的数十篇期刊论文，以帮助我们更好地解答用户体验领域的问题。在这个过程中，我不仅找到了满意的答案，还学到了很多知识，结识了许多朋友。这其中便包括我的合著者——James R. Lewis。感谢在著书的整个过程中我的家人对我的包容和鼓励。

## James 的致谢

感谢每一位使我对可用性研究和用户体验设计如此感兴趣的人，是他们让我对我这份事业乐此不疲——从我的合著者 Jeff Sauro 到 IBM 的用户体验团队，以及州立农业保险公司，他们是最宝贵的学习财富，他们鼓舞了我对教育事业的不断追求。当然，还要感谢我的家人一直以来对我的支持和鼓励。

# 目 录

译者序	
作者介绍	
推荐序	
前言	
致谢	
<b>第1章 导论</b> ..... 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 ..... 6	
1.5 本章思考题 ..... 6	
1.6 本章思考题答案 ..... 6	
1.7 参考资料 ..... 7	
<b>第2章 量化用户研究</b> ..... 8	
2.1 什么是用户研究 ..... 8	
2.2 用户研究的数据 ..... 8	
2.3 可用性测试 ..... 9	
2.3.1 样本量 ..... 9	
2.3.2 代表性和随机性 ..... 9	
2.3.3 用户研究的三种研究 类型 ..... 11	
2.3.4 数据收集 ..... 11	
2.3.5 任务完成率 ..... 12	
2.3.6 可用性问题 ..... 12	
2.3.7 任务时间 ..... 13	
2.3.8 出错数 ..... 13	
2.3.9 满意度评分 ..... 14	
2.3.10 复合分数 ..... 14	
2.4 A/B 测试 ..... 14	
2.5 调查数据 ..... 15	
2.5.1 等级量表 ..... 15	
2.5.2 净推荐值 ..... 15	
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 ..... 16	
2.8 参考资料 ..... 17	
<b>第3章 我们的估算到底有多准确</b> ..... 18	
3.1 简介 ..... 18	
3.1.1 置信区间 = 误差幅度 的两倍 ..... 18	
3.1.2 置信区间提供了精确度 和位置 ..... 19	

3.1.3	置信区间的三个组成要素 .....	19	4.3.2	大样本检验 .....	42
3.2	完成率的置信区间 .....	19	4.4	满意度评分与基准的比对 .....	43
3.2.1	置信区间的历史 .....	20	4.5	任务时间和基准的比对 .....	47
3.2.2	Wald 区间: 对于小样本来说太不靠谱了 .....	20	4.6	本章要点 .....	51
3.2.3	精确置信区间 .....	21	4.7	本章思考题 .....	51
3.2.4	Wald 校正区间: 增加两次成功与两次失败 .....	21	4.8	本章思考题答案 .....	52
3.2.5	完成率的最佳点估计 .....	23	4.9	参考资料 .....	55
3.2.6	遭遇可用性问题的置信区间 .....	24	第5章	不同设计之间有统计学差异吗 .....	56
3.3	等级量表和其他连续性数据的置信区间 .....	25	5.1	简介 .....	56
3.3.1	任务时长数据的置信区间 .....	27	5.2	比较两个平均值(等级量表和反应时) .....	56
3.3.2	任务时长均值还是任务时长中位数 .....	28	5.2.1	组内比较——配对 $t$ 检验 .....	57
3.3.3	几何平均数 .....	28	5.2.2	比较任务时长 .....	59
3.3.4	大样本任务时长的置信区间 .....	30	5.2.3	组间比较(双样本 $t$ 检验) .....	61
3.3.5	围绕中位数的置信区间 .....	30	5.2.4	$t$ 检验的假设 .....	65
3.4	本章要点 .....	32	5.3	比较完成率、转换率以及 A/B 测试 .....	66
3.5	本章思考题 .....	33	5.3.1	组间比较 .....	66
3.6	本章思考题答案 .....	33	5.3.2	组内比较 .....	74
3.7	参考资料 .....	34	5.4	本章要点 .....	82
第4章	我们达到或超过目标了吗 .....	35	5.5	本章思考题 .....	84
4.1	简介 .....	35	5.6	本章思考题答案 .....	86
4.2	单侧检验和双侧检验 .....	38	5.7	参考资料 .....	90
4.3	完成率与基准的比对 .....	39	第6章	关于样本量大小的总结性研究 .....	91
4.3.1	小样本检验 .....	39	6.1	简介 .....	91
			6.1.1	我们为何要关注 .....	91
			6.1.2	可用性研究的类型至关重要 .....	92

6.1.3	总结性可用性测试 样本量预估的基本 原则 .....	92	7.2.3	使用表格计划形成性 用户研究样本量 .....	129	
6.2	预估数值 .....	93	7.3	二项概率模型的假设 .....	130	
6.3	比较数值 .....	99	7.4	模型的附加应用 .....	131	
6.4	如何控制变异性 .....	104	7.4.1	估计多重问题或其他 事件的复合 $p$ 值 .....	131	
6.5	二项置信区间样本量的 估计 .....	105	7.4.2	校正小样本 $p$ 的复合 评估值 .....	132	
6.5.1	大样本的二项样本量 估计 .....	105	7.4.3	估计可发现的问题数和 未被发现的问题数 .....	136	
6.5.2	小样本的二项样本量 估计 .....	106	7.5	影响 $p$ 值的是什​​么 .....	138	
6.5.3	与基准比例相比较的 样本量 .....	109	7.6	什么是合理的目标问题 发现率 .....	138	
6.6	卡方检验的样本量预估 (独立比例) .....	111	7.7	调解“神奇的数字5”和 “8还不够” .....	140	
6.7	McNemar 精确检验的样本 预估(配对比例) .....	114	7.7.1	一段历史——20世纪 80年代 .....	141	
6.8	本章要点 .....	117	7.7.2	又一段历史——20世纪 90年代 .....	141	
6.9	本章思考题 .....	119	7.7.3	“神奇的数字5”的 起源 .....	142	
6.10	本章思考题答案 .....	120	7.7.4	“8还不够”—— 一个调解方法 .....	144	
6.11	参考资料 .....	123	7.8	更多关于二项概率公式及其 小样本校正 .....	147	
<b>第7章 关于样本量大小的形成性 研究 .....</b>			125	7.8.1	二项概率公式的 起源 .....	147
7.1	简介 .....	125	7.8.2	紧缩校正是如何 起作用的 .....	149	
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 .....	125	7.9	针对问题发现的其他 统计模型 .....	151	
7.2.1	著名方程: $P(x \geq 1) =$ $1 - (1 - p)^n$ .....	125	7.9.1	对问题发现使用二项 分布模型的批评 .....	151	
7.2.2	从 $1 - (1 - p)^n$ 中推导 出样本量估计方程 .....	127				

7.9.2	扩展的二项分布模型 .....	152	8.2.5	UMUX (用户体验的可用性指标) .....	186
7.9.3	捕获-再捕获模型 .....	153	8.2.6	UMUX-LITE (用户体验的可用性指标-简化版) .....	188
7.9.4	在计划形成性用户研究时为什么不用其他模型 .....	153	8.2.7	整体评估可用性问卷的实验比较 .....	189
7.10	本章要点 .....	156	8.3	任务评估问卷 .....	192
7.11	本章思考题 .....	158	8.3.1	ASQ (场景后问卷) .....	192
7.12	本章思考题答案 .....	159	8.3.2	SEQ (单项难易度问卷) .....	193
7.13	参考资料 .....	159	8.3.3	SMEQ (主观脑力负荷问题) .....	194
<b>第8章</b>	<b>标准化的可用性问卷 .....</b>	<b>163</b>	8.3.4	ER (期望评分) .....	194
8.1	简介 .....	163	8.3.5	UME (可用性等级评估) .....	196
8.1.1	什么是标准化的问卷 .....	163	8.3.6	任务评估问卷的实验比较 .....	198
8.1.2	标准化可用性问卷的优点 .....	163	8.4	网站感知可用性的评估问卷 .....	199
8.1.3	什么样的标准化可用性问卷是有用的 .....	164	8.4.1	WAMMI (网站分析和测量问卷) .....	200
8.1.4	标准化问卷的质量评估: 信度、效度和灵敏度 .....	165	8.4.2	SUPR-Q (标准化的用户体验百分等级问卷) .....	201
8.1.5	其他项目的特征 .....	166	8.4.3	其他评估网站的问卷 .....	203
8.2	后续研究的调查问卷 .....	167	8.5	其他有趣的问卷 .....	204
8.2.1	QUIS (用户交互满意度问卷) .....	168	8.5.1	CSUQ (电脑系统可用性问卷) .....	204
8.2.2	SUMI (软件可用性测试问卷) .....	169	8.5.2	USE (有用性、满意度、易用性) .....	206
8.2.3	PSSUQ (整体评估可用性问卷) .....	170			
8.2.4	SUS (系统可用性量表) .....	175			



8.5.3	HQ (享受性质量) …	206	9.5	当 $p > 0.05$ 时, 我们能拒绝原假设吗 ……………	230	
8.5.4	EMO (情绪指标结果) ……………	207	9.5.1	一方观点 ……………	230	
8.5.5	ACSI (美国顾客满意度指数) ……………	208	9.5.2	另一方观点 ……………	230	
8.5.6	NPS (净推荐值) …	208	9.5.3	我们的推荐 ……………	232	
8.5.7	CxPi (福雷斯特客户体验指数) ……………	209	9.6	能将各种可用性度量指标合并到一个分数中吗 ……………	234	
8.5.8	TAM (技术接受模型) ……………	210	9.6.1	一方观点 ……………	234	
8.6	本章要点 ……………	211	9.6.2	另一方观点 ……………	235	
8.7	本章思考题 ……………	212	9.6.3	我们的推荐 ……………	236	
8.8	本章思考题答案 ……………	213	9.7	假使你需要进行多次检验该怎么办 ……………	236	
8.9	参考资料 ……………	215	9.7.1	一方观点 ……………	237	
<b>第9章 测量和统计的六大持久论战</b> ……………			221	9.7.2	另一方观点 ……………	238
9.1	介绍 ……………	221	9.7.3	我们的推荐 ……………	239	
9.2	对多重评分量表数据进行平均合理吗 ……………	222	9.8	本章要点 ……………	242	
9.2.1	一方观点 ……………	222	9.9	本章思考题 ……………	243	
9.2.2	另一方观点 ……………	223	9.10	本章思考题答案 ……………	244	
9.2.3	我们的推荐 ……………	225	9.11	参考资料 ……………	246	
9.3	需要测试至少 30 名用户吗 ……………	226	<b>第10章 相关性分析、回归分析和方差分析</b> ……………			248
9.3.1	一方观点 ……………	226	10.1	简介 ……………	248	
9.3.2	另一方观点 ……………	226	10.2	相关性分析 ……………	249	
9.3.3	我们的推荐 ……………	228	10.2.1	如何计算相关性 …	250	
9.4	所有的实验都要进行双侧检验吗 ……………	228	10.2.2	$r$ 的统计意义 ……	252	
9.4.1	一方观点 ……………	228	10.2.3	$r$ 的置信区间 ……	253	
9.4.2	另一方观点 ……………	228	10.2.4	解释 $r$ 的大小 ……	254	
9.4.3	我们的推荐 ……………	230	10.2.5	$r$ 的样本量评估 …	254	
			10.3	决定系数 ( $R^2$ ) ……	255	
			10.3.1	与二进制数据的 相关性 ……………	255	
			10.3.2	计算 phi 相关性 …	256	

10.4	回归分析 .....	258	10.7	本章思考题 .....	278
10.4.1	评估斜率和干扰性 .....	258	10.8	本章思考题答案 .....	279
10.4.2	斜率和预测值的 置信区间 .....	259	10.9	参考资料 .....	284
10.4.3	线性回归的样本量 评估 .....	262	10.10	附录：回归分析中样本量 公式的推导 .....	285
10.5	方差分析 .....	264	第11章	总结 .....	287
10.5.1	比较两个以上变量 的方法 .....	264	11.1	简介 .....	287
10.5.2	评估交互 .....	272	11.2	更多信息 .....	287
10.5.3	方差分析的置信区间 和样本量评估 .....	275	11.3	好运 .....	290
10.6	本章要点 .....	275	11.4	本章要点 .....	290
			11.5	参考资料 .....	290
			附录A	基础统计概念速成 .....	291