

Contents

Acknowledgments.....	xiii
About the Authors.....	xv
CHAPTER 1 Introduction and How to Use This Book.....	1
Introduction.....	1
The Organization of This Book.....	1
How to Use This Book.....	2
What Test Should I Use?.....	2
What Sample Size Do I Need?.....	6
You Don't Have to Do the Computations by Hand.....	7
Key Points from the Chapter.....	7
Reference.....	8
CHAPTER 2 Quantifying User Research.....	9
What is User Research?.....	9
Data from User Research.....	9
Usability Testing.....	9
Sample Sizes.....	10
Representativeness and Randomness.....	10
Data Collection.....	12
Completion Rates.....	12
Usability Problems.....	13
Task Time.....	14
Errors.....	14
Satisfaction Ratings.....	14
Combined Scores.....	14
A/B Testing.....	15
Clicks, Page Views, and Conversion Rates.....	15
Survey Data.....	15
Rating Scales.....	15
Net Promoter Scores.....	16
Comments and Open-ended Data.....	16
Requirements Gathering.....	16
Key Points from the Chapter.....	17
References.....	17

CHAPTER 3	How Precise Are Our Estimates? Confidence Intervals.....	19
	Introduction.....	19
	Confidence Interval = Twice the Margin of Error.....	19
	Confidence Intervals Provide Precision and Location.....	19
	Three Components of a Confidence Interval.....	20
	Confidence Interval for a Completion Rate.....	20
	Confidence Interval History.....	21
	Wald Interval: Terribly Inaccurate for Small Samples.....	21
	Exact Confidence Interval.....	22
	Adjusted-Wald Interval: Add Two Successes and Two Failures.....	22
	Best Point Estimates for a Completion Rate.....	24
	Confidence Interval for a Problem Occurrence.....	26
	Confidence Interval for Rating Scales and Other Continuous Data.....	26
	Confidence Interval for Task-time Data.....	29
	Mean or Median Task Time?.....	30
	Geometric Mean.....	31
	Confidence Interval for Large Sample Task Times.....	33
	Confidence Interval Around a Median.....	33
	Key Points from the Chapter.....	36
	References.....	38
CHAPTER 4	Did We Meet or Exceed Our Goal?.....	41
	Introduction.....	41
	One-Tailed and Two-Tailed Tests.....	44
	Comparing a Completion Rate to a Benchmark.....	45
	Small-Sample Test.....	45
	Large-Sample Test.....	49
	Comparing a Satisfaction Score to a Benchmark.....	50
	Do at Least 75% Agree? Converting Continuous Ratings to Discrete.....	52
	Comparing a Task Time to a Benchmark.....	54
	Key Points from the Chapter.....	58
	References.....	62
CHAPTER 5	Is There a Statistical Difference between Designs?.....	63
	Introduction.....	63
	Comparing Two Means (Rating Scales and Task Times).....	63
	Within-subjects Comparison (Paired <i>t</i> -test).....	63
	Comparing Task Times.....	66
	Between-subjects Comparison (Two-sample <i>t</i> -test).....	68
	Assumptions of the <i>t</i> -tests.....	73

Comparing Completion Rates, Conversion Rates, and A/B Testing..... 74
 Between-subjects..... 75
 Within-subjects..... 84
 Key Points from the Chapter..... 93
 References..... 102

CHAPTER 6 What Sample Sizes Do We Need? Part 1: Summative Studies..... 105

Introduction..... 105
 Why Do We Care?..... 105
 The Type of Usability Study Matters..... 105
 Basic Principles of Summative Sample Size Estimation..... 106
 Estimating Values..... 108
 Comparing Values..... 114
 What can I Do to Control Variability?..... 120
 Sample Size Estimation for Binomial Confidence Intervals..... 121
 Binomial Sample Size Estimation for Large Samples..... 121
 Binomial Sample Size Estimation for Small Samples..... 123
 Sample Size for Comparison with a Benchmark Proportion..... 125
 Sample Size Estimation for Chi-Square Tests
 (Independent Proportions)..... 128
 Sample Size Estimation for McNemar Exact Tests
 (Matched Proportions)..... 131
 Key Points from the Chapter..... 135
 References..... 141

CHAPTER 7 What Sample Sizes Do We Need? Part 2: Formative Studies..... 143

Introduction..... 143
 Using a Probabilistic Model of Problem Discovery to Estimate
 Sample Sizes for Formative User Research..... 143
 The Famous Equation: $P(x \geq 1) = 1 - (1 - p)^n$ 143
 Deriving a Sample Size Estimation Equation from $1 - (1 - p)^n$ 145
 Using the Tables to Plan Sample Sizes for Formative User Research..... 146
 Assumptions of the Binomial Probability Model..... 148
 Additional Applications of the Model..... 149
 Estimating the Composite Value of p for Multiple Problems
 or Other Events..... 149
 Adjusting Small Sample Composite Estimates of p 149
 Estimating the Number of Problems Available for Discovery and the
 Number of Undiscovered Problems..... 155
 What affects the Value of p ?..... 157

What is a Reasonable Problem Discovery Goal?..... 157

Reconciling the “Magic Number 5” with “Eight is not Enough”..... 160

 Some History: The 1980s..... 160

 Some More History: The 1990s..... 161

 The Derivation of the “Magic Number 5”..... 162

 Eight Is Not Enough: A Reconciliation..... 164

More About the Binomial Probability Formula and its Small
Sample Adjustment..... 167

 Origin of the Binomial Probability Formula..... 167

 How does the Deflation Adjustment Work?..... 169

Other Statistical Models for Problem Discovery..... 172

 Criticisms of the Binomial Model for Problem Discovery..... 172

 Expanded Binomial Models..... 173

 Capture–recapture Models..... 174

 Why Not Use One of These Other Models When Planning Formative
 User Research?..... 174

Key Points from the Chapter..... 178

References..... 181

CHAPTER 8 Standardized Usability Questionnaires..... 185

Introduction..... 185

 What is a Standardized Questionnaire?..... 185

 Advantages of Standardized Usability Questionnaires..... 185

 What Standardized Usability Questionnaires Are Available?..... 186

 Assessing the Quality of Standardized Questionnaires:
 Reliability, Validity, and Sensitivity..... 187

 Number of Scale Steps..... 187

Poststudy Questionnaires..... 188

 QUIS (Questionnaire for User Interaction Satisfaction)..... 188

 SUMI (Software Usability Measurement Inventory)..... 190

 PSSUQ (Post-study System Usability Questionnaire)..... 192

 SUS (Software Usability Scale)..... 198

 Experimental Comparison of Poststudy Usability Questionnaires..... 210

Post-Task Questionnaires..... 212

 ASQ (After-scenario Questionnaire)..... 213

 SEQ (Single Ease Question)..... 214

 SMEQ (Subjective Mental Effort Question)..... 214

 ER (Expectation Ratings)..... 215

 UME (Usability Magnitude Estimation)..... 217

 Experimental Comparisons of Post-task Questionnaires..... 219

Questionnaires for Assessing Perceived Usability of Websites..... 221
 WAMMI (Website Analysis and Measurement Inventory)..... 222
 SUPR-Q (Standardized Universal Percentile Rank Questionnaire)..... 223
 Other Questionnaires for Assessing Websites..... 224
 Other Questionnaires of Interest..... 225
 CSUQ (Computer System Usability Questionnaire)..... 225
 USE (Usefulness, Satisfaction, and Ease of Use)..... 227
 UMUX (Usability Metric for User Experience)..... 227
 HQ (Hedonic Quality)..... 228
 ACSI (American Customer Satisfaction Index)..... 229
 NPS (Net Promoter Score)..... 229
 CxPi (Forrester Customer Experience Index)..... 230
 TAM (Technology Acceptance Model)..... 231
 Key Points from the Chapter..... 232
 References..... 236

CHAPTER 9 Six Enduring Controversies in Measurement and Statistics..... 241

Introduction..... 241
 Is it Okay to Average Data from Multipoint Scales?..... 242
 On One Hand..... 242
 On the Other Hand..... 243
 Our Recommendation..... 245
 Do you Need to Test at Least 30 Users?..... 246
 On One Hand..... 246
 On the Other Hand..... 247
 Our Recommendation..... 248
 Should you Always Conduct a Two-Tailed Test?..... 248
 On One Hand..... 248
 On the Other Hand..... 250
 Our Recommendation..... 250
 Can you Reject the Null Hypothesis when $p > 0.05$?..... 251
 On One Hand..... 251
 On the Other Hand..... 251
 Our Recommendation..... 253
 Can you Combine Usability Metrics into Single Scores?..... 254
 On One Hand..... 254
 On the Other Hand..... 255
 Our Recommendation..... 256
 What if you Need to Run more than One Test?..... 256
 On One Hand..... 256

On the Other Hand..... 258
Our Recommendation..... 258
Key Points from the Chapter..... 262
References..... 266

CHAPTER 10 Wrapping Up..... 269

Introduction..... 269
Getting More Information..... 269
Good Luck!..... 272
Key Points from the Chapter..... 272
References..... 272

Appendix: A Crash Course in Fundamental Statistical Concepts..... 273

Introduction..... 273
Types of Data..... 273
Populations and Samples..... 274
 Sampling..... 274
Measuring Central Tendency..... 274
 Mean..... 274
 Median..... 275
 Geometric Mean..... 275
Standard Deviation and Variance..... 276
The Normal Distribution..... 276
 z-scores..... 278
Area Under the Normal Curve..... 278
Applying the Normal Curve to User Research Data..... 280
Central Limit Theorem..... 280
Standard Error of the Mean..... 282
Margin of Error..... 283
t-Distribution..... 283
Significance Testing and p-Values..... 284
 How much do Sample Means Fluctuate?..... 285
The Logic of Hypothesis Testing..... 287
Errors in Statistics..... 288
Key Points from the Appendix..... 289

Index..... 291