## **Project Booking Sheet**

**Hypothesize**

What is the hypothesis, research question(s), or goal(s) of the study?

**Operationalize**

1. What type of method are you using?

(e.g. task-based study, simple survey, tree test, card sort)

* 1. Are there tasks? If so, what are they.

**Usually 2‒4 tasks, total duration of <15 minutes**

* 1. How will this be tested?

(e.g., live site, product, prototype, wireframes)

* 1. Desktop or mobile (Android or iOS)?

**Desktop only for UX Bootcamp**

1. What are your metrics? (How will success be measured?)
2. Is this a standalone or comparative study?
	1. If this is a comparative study, is it between- (different users in each condition/group) or within-subjects (same users in each condition/group)?

**Between-subjects only for UX Bootcamp**

1. Who are the participants?

(e.g., Geo, Membership, Age, Income, Tenure, etc.)

**General consumer profile for UX Bootcamp (or with approved attributes)**

* 1. Are there distinct subgroups (e.g. Account Holders vs. Prospects?)
1. What sample size will you use and why? Use the tables on the next pages to help with sample size planning.

**Max 100 participants for UX Bootcamp**

##

## **Sample Size for Standalone Studies**

Most studies sample only part of the population of interest, so there will be sampling error and some uncertainty in the metrics and therefore conclusions. When discussing sample sizes for surveys it’s best to think in terms of the precision of the estimate. Precision refers to the amount of margin of error around each metric. There is an inverse square root relationship between the sample size (total number of responses) and the margin of error.

The table below breaks down the margin of error for different sample sizes and can be a helpful guide for subgroup analysis (e.g., differences by region or brand). For example, at a sample size of 53, the expected margin of error will be ~13% for 95% confidence. That means if 70% of respondents agree to a statement, we can be 95% confident between 57% and 83% of all customers would agree to the statement.

|  |  |  |
| --- | --- | --- |
| **Margin of Error****Margin of Error** | **90% Confidence (+/-)****Sample Size** | **95% Confidence (+/-)** |
| 24% | 10 | 13 |
| 20% | 15 | 21 |
| 15% | 28 | 39 |
| 14% | 32 | 46 |
| 13% | 38 | 53 |
| 12% | 45 | 63 |
| 11% | 54 | 76 |
| 10% | 65 | 93 |
| 9% | 81 | 115 |
| 8% | 103 | 147 |
| 7% | 136 | 193 |
| 6% | 186 | 263 |
| 5% | 268 | 381 |
| 4% | 421 | 597 |
| 3% | 749 | 1,064 |
| 2% | 1,689 | 2,398 |

To achieve a margin of error of +/- 10%, you should plan on a sample size of 93 for 95% confidence and 65 for 90% confidence.

## **Sample Size for Comparisons**

The table below shows the approximate difference we can detect in metrics such as the completion rate or any other binary measure (at a 50% completion rate). For example, at a sample size of 426 (~200 in each group, we can detect a 12% difference for a between-subjects design. So, if 50% agree to a statement on one website and 62% on a competitive site, the difference would be statistically significant. These estimates are the most conservative but are recommended when planning a study without prior data. For continuous measures like perceived difficulty, branding and overall perceptions, we will be able to detect smaller differences at the same sample size. For example, many 7-point scales have standard deviations of 1 point, meaning a difference of .2 points would be statistically significant (e.g., 5 vs. 5.2).

|  |  |  |
| --- | --- | --- |
| **Difference to Detect90% Confidence & 80% Power** | **Sample SizeWithin Subjects** | **Sample Size Between Subjects** |
| **50%** | 17 | 22 |
| **40%** | 20 | 34 |
| **30%** | 29 | 64 |
| **20%** | 50 | 150 |
| **12%** | 93 | 426 |
| **10%** | 115 | 614 |
| **9%** | 130 | 760 |
| **8%** | 148 | 962 |
| **7%** | 171 | 1258 |
| **6%** | 202 | 1714 |
| **5%** | 246 | 2468 |
| **4%** | 312 | 3860 |
| **3%** | 421 | 6866 |
| **2%** | 640 | 15452 |
| **1%** | 1297 | 61822 |

See Chapter 6 in *Quantifying the User Experience* for the formulas on computing sample sizes for within- and between-subjects proportions.

## **Sample Size for Discovering Problems**

Start with the likelihood of detecting a problem at least once, usually .85 or higher. Next identify how common the problem is (if it exists) in the column “Problem Detection Probability.” For example, to have an 85% chance of seeing a problem that affects 25% of all users, plan on testing with 7 users. To have a 90% chance of seeing a problem that affects 10% of all users, plan on testing with 22 users.

|  |  |
| --- | --- |
| Problem Detection Probability | Likelihood of Detecting the Problem at Least Once |
| **.50** | **.75** | **.85** | **.90** | **.95** | **.99** |
| **.01** | 69 | 138 | 189 | 230 | 299 | 459 |
| **.05** | 14 | 28 | 37 | 45 | 59 | 90 |
| **.10** | 7 | 14 | 19 | 22 | 29 | 44 |
| **.15** | 5 | 9 | 12 | 15 | 19 | 29 |
| **.25** | 3 | 5 | 7 | 9 | 11 | 17 |
| **.50** | 1 | 2 | 3 | 4 | 5 | 7 |
| **.90** | 1 | 1 | 1 | 1 | 2 | 2 |