Chapter 15

The Bootstrap Discovery Behaviour Model: Why Five Users are Not Enough to Test User Experience

Simone Borsci  
Brunel University, UK

Stefano Federici  
University of Perugia, Italy

Maria Laura Mele  
Sapienza University of Rome, Italy

Domenico Polimeno  
University of Perugia, Italy

Alessandro Londei  
Sapienza University of Rome, Italy

ABSTRACT

The chapter focuses on the Bootstrap statistical technique for assigning measures of accuracy to sample estimates, here adopted for the first time to obtain an effective and efficient interaction evaluation. After introducing and discussing the classic debate on p value (i.e., the discovery detection rate) about estimation problems, the authors present the most used model for the estimation of the number of participants needed for an evaluation test, namely the Return On Investment model (ROI). Since the ROI model endorses a monodimensional and economical perspective in which an evaluation process, composed of only an expert technique, is sufficient to identify all the interaction problems—without distinguishing real problems (i.e., identified both experts and users) and false problems (i.e., identified only by experts)—they propose the new Bootstrap Discovery Behaviour (BDB) estimation model. Findings highlight the BDB as a functional technique favouring practitioners to optimize the number of participants needed for an interaction evaluation. Finally, three experiments show the application of the BDB model to create experimental sample sizes to test user experience of people with and without disabilities.

DOI: 10.4018/978-1-4666-1628-8.ch015
The Bootstrap Discovery Behaviour Model

INTRODUCTION

The ROI model, which was proposed in 1993 by Nielsen and Landauer, shows that, generally, the least number of users required for a usability test ranges from three to five. This model is an asymptotic test which allows practitioners to estimate the number of users needed through the following formula:

\[
\text{Found}(i) = N \left[ 1 - (1 - p)^i \right] \quad (1)
\]

In (1), the \( N \) value corresponds to the total number of problems in the interface, the \( p \) value is defined by Nielsen and Landauer (1993) as “the probability of finding the average usability problem when running a single average subject test” (i.e., discovery detection rate), and the \( i \) value corresponds to the number of users. For instance, by applying formula (1), practitioners can estimate whether five users are sufficient for obtaining a reliable assessment and, if not, how many users (\( N \)) are needed in order to increase the percentage of usability problems. Nielsen, starting from the results obtained by many applications of the ROI model, suggests that the practitioners, in order to test different categories of users, have to divide users into multiple groups composed as follows (Nielsen, 2000):

- 5 subjects of a category if testing 1 group of users;
- 3-4 subjects from each category if testing 2 groups of users;
- 3 users from each category if testing three or more groups of users.

The value “\( p \)” (see formula 1) may be considered an index for assessing the effectiveness and efficiency of an Evaluation Method (EM). As some international studies (Lewis, 1994; Nielsen, 2000; Nielsen & Mack, 1994; Virzi, 1990, 1992; Wright & Monk, 1991) have shown, a sample size of five participants is sufficient to find approximately 80% of the usability problems in a system when the individual detection rate (\( p \)) is at least .30. The value of 30% was derived through Monte Carlo (MC) resampling of multiple evaluators, and could also be estimated using the full matrix of problems as discovered by independent evaluators (Lewis, 2001).

However, as Nielsen and Landauer (1993, p. 209) underline when discussing their model, the discoverability rate (\( p \)) for any given usability test depends on at least seven main factors:

- The properties of the system and its interface;
- The stage of the usability lifecycle;
- The type and quality of the methodology used to conduct the test;
- The specific tasks selected;
- The match between the test and the context of real world usage;
- The representativeness of the test participants;
- The skill of the evaluator.

As Borsci, Londei, and Federici (2011) claim, many studies underline that these factors have an effect on the evaluation of the interaction between system and user that the ROI model is not able to estimate (Caulton, 2001; Hertzum & Jacobsen, 2003; Lewis, 1994, 2006; Schmettow, 2008; Spool & Schroeder, 2001). In this sense, the ROI model cannot guarantee the reliability of the evaluation results obtained by the first five participants.

One of the most relevant problems that Borsci et al. (2011) underline is that the ROI model starts with a “one evaluator” condition and not at zero condition; this means that the characteristics of the system are considered only as the differences between problems found by the first evaluators. Nielsen and Mack (1994) pointed out that the first evaluator (a user or an expert) generally finds 30% of the problems, because these problems are generally the most evident. The subsequent evaluators usually find a smaller percentage of
Related Content

Symmetry vs. Duality in Logic: An Interpretation of Bi-Logic to Model Cognitive Processes Beyond Inference
[www.igi-global.com/article/symmetry-vs-duality-in-logic/133297?camid=4v1a](www.igi-global.com/article/symmetry-vs-duality-in-logic/133297?camid=4v1a)

Visualization of Brain Activation During Attention-Demanding Tasks Using Cognitive Signal Processing

Human Centricity and Perception-Based Perspective and Their Centrality to the Agenda of Granular Computing
[www.igi-global.com/article/human-centricity-perception-based-perspective/63621?camid=4v1a](www.igi-global.com/article/human-centricity-perception-based-perspective/63621?camid=4v1a)

Single SNN Architecture for Classical and Operant Conditioning using Reinforcement Learning
[www.igi-global.com/article/single-snn-architecture-for-classical-and-operant-conditioning-using-reinforcement-learning/179931?camid=4v1a](www.igi-global.com/article/single-snn-architecture-for-classical-and-operant-conditioning-using-reinforcement-learning/179931?camid=4v1a)