

1. Record Nr.	UNINA9910299296903321
Autore	Sakai Tetsuya
Titolo	Laboratory Experiments in Information Retrieval : Sample Sizes, Effect Sizes, and Statistical Power // by Tetsuya Sakai
Pubbl/distr/stampa	Singapore : , : Springer Singapore : , : Imprint : Springer, , 2018
ISBN	981-13-1199-4
Edizione	[1st ed. 2018.]
Descrizione fisica	1 online resource (IX, 150 p. 53 illus., 43 illus. in color.)
Collana	The Information Retrieval Series, , 1871-7500 ; ; 40
Disciplina	025.04
Soggetti	Information storage and retrieval Statistics Information Storage and Retrieval Statistics for Engineering, Physics, Computer Science, Chemistry and Earth Sciences
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	1 Preliminaries -- 2 t-tests -- 3 Analysis of Variance -- 4 Multiple Comparison Procedures -- 5 The Correct Ways to Use Significance Tests -- 6 Topic Set Size Design Using Excel -- 7 Power Analysis Using R -- 8 Conclusions.
Sommario/riassunto	Covering aspects from principles and limitations of statistical significance tests to topic set size design and power analysis, this book guides readers to statistically well-designed experiments. Although classical statistical significance tests are to some extent useful in information retrieval (IR) evaluation, they can harm research unless they are used appropriately with the right sample sizes and statistical power and unless the test results are reported properly. The first half of the book is mainly targeted at undergraduate students, and the second half is suitable for graduate students and researchers who regularly conduct laboratory experiments in IR, natural language processing, recommendations, and related fields. Chapters 1–5 review parametric significance tests for comparing system means, namely, t-tests and ANOVAs, and show how easily they can be conducted using Microsoft Excel or R. These chapters also discuss a few multiple comparison procedures for researchers who are interested in comparing every system pair, including a randomised version of Tukey's Honestly

Significant Difference test. The chapters then deal with known limitations of classical significance testing and provide practical guidelines for reporting research results regarding comparison of means. Chapters 6 and 7 discuss statistical power. Chapter 6 introduces topic set size design to enable test collection builders to determine an appropriate number of topics to create. Readers can easily use the author's Excel tools for topic set size design based on the paired and two-sample t-tests, one-way ANOVA, and confidence intervals. Chapter 7 describes power-analysis-based methods for determining an appropriate sample size for a new experiment based on a similar experiment done in the past, detailing how to utilize the author's R tools for power analysis and how to interpret the results. Case studies from IR for both Excel-based topic set size design and R-based power analysis are also provided.

---