Power Analysis for Experimental Research

A Practical Guide for the Biological, Medical and Social Sciences

Power analysis is an essential tool for determining whether a statistically significant result can be expected in a scientific experiment prior to the experiment being performed. Many funding agencies and Institutional Review Boards now require power analyses to be carried out before they will approve experiments, particularly where they involve the use of human subjects. This comprehensive, yet accessible, book provides practicing researchers with step-by-step instructions for conducting power/sample size analyses, assuming only basic prior knowledge of summary statistics and the normal distribution. It contains a unified approach to statistical power analysis, with numerous easy-to-use tables to guide the reader without the need for further calculations or statistical expertise. This will be an indispensable text for researchers and graduates in the medical and biological sciences needing to apply power analysis in the design of their experiments.

R. BARKER BAUSELL is Professor at the University of Maryland School of Medicine and Director of Research of the Complementary Medicine Program. He has served as the editor-in-chief of the peer-reviewed methodology journal *Evaluation and the Health Professions* for the past 25 years and has conducted statistical research in a number of disciplines related to the medical sciences.

YU-FANG LI is a Data Analyst at the Department of Veteran's Affairs, Puget Sound Health Care System in Seattle, Washington. She previously worked as a statistical consultant with a particular interest in meta-analysis.

Power Analysis for Experimental Research

A Practical Guide for the Biological, Medical and Social Sciences

R. BARKER BAUSELL
YU-FANG LI

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Univers and Bembo 11/13pt. *System* QuarkXPress® [SE]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

ISBN 0 521 80916 9 hardback

This book is dedicated to Jesse Turner Bausell

Contents

Introduction

The primary purpose of this book is to provide an *easy-to-use*, unified approach to statistical power analysis in order to enable investigators to avail themselves of the advantages of this powerful tool in the design of their experiments. It is our firm conviction that no other process possesses more potential for increasing the scientific and societal yields accruing from our experiments. It is also our firm belief that the a priori consideration of power is so integral to the entire design process that its consideration should not be delegated to individuals not integrally involved in the conduct of an investigation, hence the present volume has been written to be completely accessible to practicing researchers. For this reason we have studiously avoided the use of technical terms and formulas until the appendix to make it as accessible (and hopefully interesting) to individuals without advanced statistical training as possible.

This is not to say that statistical collaboration in the conduct of most experiments is not desirable. It is, in fact, often absolutely essential and we have written this work to make it as helpful as possible to statisticians charged with the task of performing a power or sample size analysis. It has been our experience, however, that while principal investigators are well versed in formulating research hypotheses, they often conceptualize the determination of power (or the sample size necessary to achieve a desired value thereof) as a technical exercise better delegated to someone with the appropriate expertise. Our purpose in writing this book is to simplify the power analytic process to the point that it can become the integral component of experimental design in *practice* that it occupies in *theory*. The *true* value of the concept of statistical power, in fact, lies in the fact that its consideration forces investigators to think in terms of the strength of the effects their experiments are likely to produce, which is absolutely crucial to the design process itself. It is for this reason that it is not wise or fair to delegate the power analytic process to a statistician, no matter how skilled, who is not immersed in the subject matter and previous research surrounding the experiment being designed. With this principle in mind, we hope that this work will facilitate the statistician's role both with respect to computing the power analysis for a wide range of experimental designs and to involving his/her non-statistical collaborators in the process.

What makes this collaboration so important is the fact that the types of hypotheses that most scientists have been trained to write are not nearly as scientifically (or clinically) relevant as they could be. In designing most experiments, for example, it is almost trivial to posit something to the effect that "Patients exposed to intervention X will experience significantly fewer occurrences of symptom Y at the end of Z weeks than patients receiving standard care." In designing such an experiment it is almost a foregone conclusion that the investigator (and his/her potential funding agency) *believes* that the proposed intervention will be better than a control. From an epistemological point of view it is tautological that the intervention will not produce exactly the same effects as the control. What the investigator's true job description involves, therefore, is the design of experiments that are capable of *demonstrating* the intervention's effectiveness or, at the very least, of designing experiments that provide an adequate chance of demonstrating the intervention's effectiveness.

Many researchers consider this the province of **statistical significance**, and in one sense they are correct. Statistical significance, however, is only one of two pillars upon which the process of accepting or rejecting scientific hypotheses rests. The other pillar is statistical power, or the probability that statistical significance will be obtained and that probability is determined primarily by the **size of the effect** that an experiment is most likely to produce. Statistical significance, the supreme arbiter of an intervention's effectiveness, is also determined primarily by the size of the effect that is actually obtained after the experiment is conducted and the data are collected. If the experiment is not designed with sufficient power to detect the intervention's true **effect size**, then statistical significance will not be obtained once the data are collected and the intervention will be declared non-effective, even if a clinically relevant difference occurs between it and its control and even if the intervention "truly" is effective and is capable of saving thousands of lives (or of improving their quality) if implemented. It is therefore absolutely incumbent upon investigators to design their experiments in such a way that societally important effect sizes will be statistically significant. This, then, is the essence and true purpose of a power analysis. It also represents the true value of the power analytic process in the sense that it forces the investigator to consider what size of effect must be obtained in order to provide a reasonable chance of obtaining statistical significance. Said another way, statistical power involves forcing the investigator to perform a hypothetical statistical analysis prior to collecting data. This is accomplished by simply substituting the minimum effect the intervention is expected to have upon the outcome within the experimental context being designed. This in turn unequivocally yields a discrete, numerical probability of how likely this result will be to occur in the actual data analysis performed at the end of the experiment.

About the book

This book constitutes the most comprehensive power analytic tool available and this statement is not made lightly because it was written standing upon the shoulders of giants – notably Jacob Cohen's seminal *Statistical power analysis for the behavioral sciences* and Mark Lipsey's delightful *Design sensitivity: statistical power for experimental research*. (We would also like to acknowledge Professor Karen L. Soeken for her contributions made to this project from its inception.) What this book basically does is extend the work of these and numerous others to additional designs via tables (and detailed templates to facilitate their use) that permit a one-step approach to power analysis, while making a few advances of its own including the computation of power for multiple comparison procedures and mixed interactions.

The book itself is organized around those parametric statistical procedures most commonly employed in the analysis of experiments involving continuous outcome data. In a sense the volume need not be read from cover to cover, although we do recommend a perusal of the first three introductory chapters since they lay the conceptual foundation for the use of the tables and templates that follow.

Supplementary software

While we believe the treatment of power for *experimental* research is as comprehensive as is possible within the confines of a single volume, there are relatively rare occasions when an investigator or a statistician does need to compute power for, say, an alpha level other than 0.05 or for a different parameter than our tables permit. We have, therefore, in collaboration with Mikolaj Franaszczuk (who at the time of this writing is a brilliant undergraduate computer science student at Cornell University), prepared a computer program entitled *Power analysis for experimental research* that exactly mirrors each of the procedures covered in this volume, but which permits different (and more exact) parameters to be input. The program may be obtained free of charge to readers of this book by email from the first author (bbausell@compmed.umm.edu).