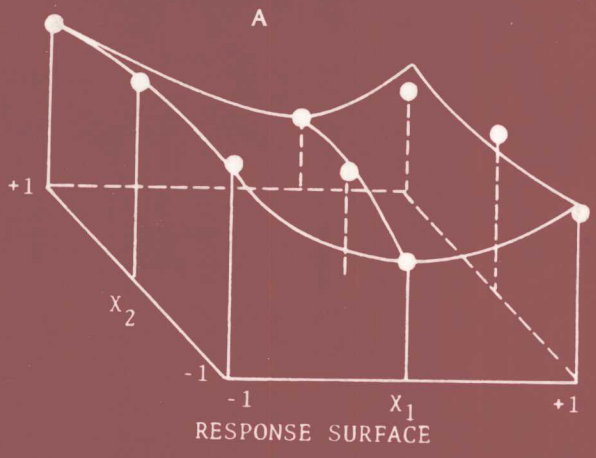
DRUGS AND THE PHARMACEUTICAL SCIENCES VOLUME 80

Pharmaceutical Statistics

Practical and Clinical Applications

Third Edition



Sanford Bolton

Pharmaceutical Statistics

Practical and Clinical Applications

Third Edition

Sanford Bolton
Visiting Professor
University of Arizona
Tucson, Arizona

MARCEL DEKKER, INC. NEW YORK • BASEL • HONG KONG

Library of Congress Cataloging-in-Publication Data

Bolton, Sanford
Pharmaceutical statistics: practical and clinical applications / Sanford Bolton. — 3rd ed.
p. cm. — (Drugs and the pharmaceutical sciences ; 80)
Includes bibliographical references and index.
ISBN 0-8247-9812-0 (alk. paper)
1. Pharmacy—Statistical methods. I. Title. II. Series: Drugs and the pharmaceutical sciences ; v. 80.
[DNLM: 1. Technology, Pharmaceutical—methods. 2. Statistics. W1 DR893B v.180 1997 / QV 25 B694p 1997]
RS57.B65 1997
615'.1'072—dc21
DNLM/DLC
for Library of Congress

96-49907
CIP

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the address below.

This book is printed on acid-free paper.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

The first edition of this book was published in 1984. Its modest aim—to help publicize the importance of statistics and expand its use among pharmaceutical scientists—has been met, as attested by its wide use and popularity. I can think of few more satisfying and heartwarming feelings than those imparted by my pharmaceutical colleagues when I am told how they have read and used the book to advantage. Most of these stories have a common theme that confirms the fulfillment of my original objective. Use of a simple, practical approach to current statistical problems that are commonly encountered in pharmaceutical research is an effective way of communicating statistics to the practicing pharmaceutical scientist. One statistician, the director of a statistical group at a major pharmaceutical company, told me that he gives a copy of the book to new statisticians so that they can get a general overview of the types of problems encountered in the pharmaceutical industry. Another scientist told me that he keeps three copies, one for himself and two to lend to friends and colleagues. This news is encouraging and gratifying, considering my aims and hopes for this book.

Therefore, the theme of the book remains the same: to present an uncomplicated explanation of the approaches and solutions to commonly encountered statistical problems, with examples that are relevant to scientists involved in pharmaceutical and related research. Because real examples often have different twists, one should be cautioned to take each example in context, and always seek expert advice, until experience gives sufficient confidence.

This third edition follows the approach of the previous editions. It includes more examples and techniques that reflect current practice and problems, and the continued emphasis and reliance on statistical techniques in drug and pharmaceutical research. Statistical tables and concepts have been expanded. In particular, analysis of bioequivalence studies, clinical studies, quality control, and concepts in GMP validation topics have been expanded. The ''Barr decision,'' resulting from a trial in which Barr Laboratories challenged an FDA injunction, has had a profound effect on the entire pharmaceutical industry. Some of the decisions resulting from the trial have been controversial, and my evaluation is presented in this new edition.

Among the new and expanded topics in this edition are recent developments in stability data analysis, concepts of statistical outliers, more detailed discussion of bioequivalence studies including some nonparametric applications, problems in sampling and devising limits for product release, covariance analysis, tolerance intervals, multiple endpoints, more nonparametric tests, and additional topics in clinical data analysis.

Although details of the statistical computations are presented in the book, current-day statistical analyses cannot be performed economically without computers. Otherwise, the structure remains similar to the previous editions. The initial chapters are general and introductory, whereas the later and expanded chapters are more specific and applied. No previous statistical education or higher mathematics is needed to understand the concepts of this book.

The exercises are an integral part of the book and can be used to test the reader's comprehension of the text. More difficult or advanced problems are designated by double asterisks.

There are many people who deserve credit and thanks for this endeavor. In particular, I want to thank the graduate students with whom I have had the privilege to work. They are part of my family and have always inspired me. A very special expression of gratitude is extended to Jenny Chen, who volunteered to redo the figures using computer software. Also, many thanks are due to the PMA and generic pharmaceutical companies, both scientists and management, who have given me the opportunity of working with real-life stimulating problems. Once again, I want to acknowledge the encouragement and inspiration instilled in me by the late Dr. John Fertig of Columbia University and the late Dr. Takeru Higuchi, my mentors.

Finally, my sincere wish is that this book will continue to be useful to my fellow scientists and statisticians.

Sanford Bolton

CONTENTS

**A more advanced topic.

**A more advanced topic.

1

BASIC DEFINITIONS AND CONCEPTS

Statistics has its own vocabulary. Many of the terms that comprise statistical nomenclature are familiar: some commonly used in everyday language, with perhaps, somewhat different connotations. Precise definitions are given in this chapter so that no ambiguity will exist when the words are used in subsequent chapters. Specifically, such terms as *discrete* and *continuous variables*, *frequency distribution*, *population*, *sample*, *mean*, *median*, *standard deviation*, *variance*, *coefficient of variation* (*CV*), *range*, *accuracy*, and *precision* are introduced and defined. The methods of calculation of different kinds of means, the median, standard deviation, and range are also presented. When studying any discipline, the initial efforts are most important. The first chapters of this book are important in this regard. Although most of the early concepts are relatively simple, a firm grasp of this material is essential for understanding the more difficult material to follow.

1.1 VARIABLES AND VARIATION

Variables are the measurements, the values, which are characteristic of the data collected in experiments. These are the data that will usually be displayed, analyzed, and interpreted in a research report or publication. In statistical terms, these observations are more correctly known as *random variables*. Random variables take on values, or numbers, according to some corresponding probability function. Although we will wait until Chapter 3 to discuss the concept of probability, for the present we can think of a

random variable as the typical experimental observation that we, as scientists, deal with on a daily basis. Because these measurements may take on different values, repeat measurements observed under apparently identical conditions do not, in general, give the identical results (i.e., they are usually not exactly reproducible). Duplicate determinations of serum concentration of a drug 1 hr after an injection will not be identical no matter if the duplicates come from (a) the same blood sample or (b) from separate samples from two different persons or (c) from the same person on two different occasions. Variation is an inherent characteristic of experimental observations. To isolate and to identify particular causes of variability requires special experimental designs and analysis. Variation in observations is due to a number of causes. For example, an assay will vary depending on:

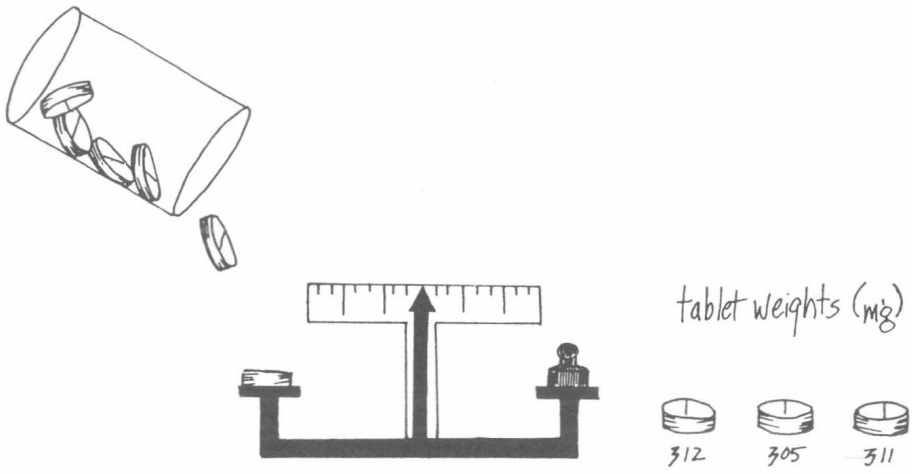
1. The instrument used for the analysis
2. The analyst performing the assay
3. The particular sample chosen
4. Unidentified, uncontrollable background error, commonly known as "noise"

This inherent variability in observation and measurement is a principal reason for the need of statistical methodology in experimental design and data analysis. In the absence of variability, scientific experiments would be short and simple: interpretation of experimental results from well-designed experiments would be unambiguous. In fact, without variability, single observations would often be sufficient to define the properties of an object or a system. Since few, if any, processes can be considered absolutely invariant, statistical treatment is often essential for summarizing and defining the nature of data, and for making decisions or inferences based on these variable experimental observations.

1.1.1 Continuous Variables

Experimental data come in many forms*. Probably the most commonly encountered variables are known as *continuous variables*. A continuous variable is one that can take on *any* value within some range or interval (i.e, within a specified lower and upper limit). The limiting factor for the total number of possible observations or results is the sensitivity of the measuring instrument. When weighing tablets or making blood pressure measurements,

*For a further discussion of different kinds of variables, see Sec. 15.1 in Chapter 15, Nonparametric Methods.



Tablet weights: an example of a variable measurement (a random variable).

there are an infinite number of possible values that can be observed if the measurement could be made to an unlimited number of decimal places. However, if the balance, for example, is sensitive only to the nearest milligram, the data will appear as discrete values. For tablets targeted at 1 g and weighed to the nearest milligram, the tablet weights might range from 900 to 1100 mg, a total of 201 possible integral values (900, 901, 902, 903, . . . , 1098, 1099, 1100). For the same tablet weighed on a more sensitive balance, to the nearest 0.1 mg, values from 899.5 to 1100.4 might be possible, a total of 2010 possible values, and so on.

Often, continuous variables cannot be easily measured but can be ranked in order of magnitude. In the assessment of pain in a clinical study of analgesics, a patient can have a continuum of pain. To measure pain on a continuous numerical scale would be difficult. On the other hand, a patient may be able to differentiate slight pain from moderate pain, moderate pain from severe pain, and so on. In analgesic studies, scores are commonly assigned to pain severity, such as no pain = 0, slight pain = 1, moderate pain = 2, and severe pain = 3. Although the scores cannot be thought of as an exact characterization of pain, the value 3 does represent more intense pain than the values 0, 1, or 2. The scoring system above is a representation of a continuous variable by discrete ''scores'' which can be rationally ordered or ranked from low to high. This is commonly known as a rating scale, and the ranked data are on an ordinal scale. The rating scale is an effort to quantify a continuous, but subjective, variable.

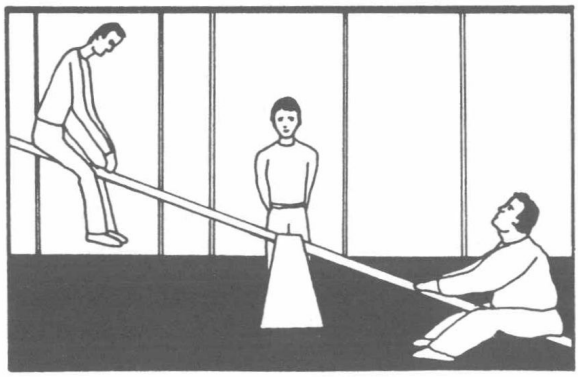
1.1.2 Discrete Variables

In contrast to continuous variables, *discrete variables* can take on a countable number of values. These kinds of variables are commonly observed in biological and pharmaceutical experiments and are exemplified by measurements such as the number of anginal episodes in 1 week or the number of side effects of different kinds after drug treatment. Although not continuous, discrete data often have values associated with them which can be numerically ordered according to their magnitude, as in the examples given earlier of a rating scale for pain and the number of anginal episodes per week.

Discrete data that can be named (nominal), categorized into two or more classes, and counted are called categorical variables, or *attributes*; for example, the attributes may be different side effects resulting from different drug treatments or the presence or absence of a defect in a finished product. These kinds of data are frequently observed in clinical and pharmaceutical experiments and processes. A finished tablet classified in quality control as ''defective'' or ''not defective'' is an example of a categorical or attribute type of variable. In clinical studies, the categorization of a patient by sex (male or female) or race is a classification according to attributes. When calculating ED_{50} or LD_{50}, animals are categorized as ''responders'' or ''nonresponders'' to various levels of a therapeutic agent, a categorical response. These examples describe variables that cannot be ordered. A male is not associated with a higher or lower numerical value than a female.

Continuous variables can always be classified into discrete classes where the classes are ordered. For example, patients can be categorized as ''un-

Classification by attributes: patients categorized by weight.



Underweight Normal weight Overweight