Basic Statistics and Pharmaceutical Statistical Applications

James E. De Muth

University of Wisconsin–Madison
Madison, Wisconsin

Marcel Dekker, Inc. New York • Basel

Library of Congress Cataloging-in-Publication Data

De Muth, James E.
Basic statistics and pharmaceutical statistical applications / James E. De Muth.
p. cm.-(Biostatistics; 2)
Includes bibliographical references and index.
ISBN 0-8247-1967-0 (alk. paper)
1. Pharmacy-Statistical methods. 2. Statistics. I. Title II. Series: Biostatistics (New York, N. Y.); 2.
RS57.D46 1999
615.1'072-dc21 99-30733
CIP

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, Inc.
270 Madison Avenue, New York, NY 10016
tel: 212-696-9000; fax: 212-685-4540

Eastern Hemisphere Distribution
Marcel Dekker AG
Hutgasse 4, Postfach 812, CH-4001 Basel, Switzerland
tel: 41-61-261-8482; fax: 41-61-261-8896

World Wide Web
http://www.dekker.com

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

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

PRINTED IN THE UNITED STATES OF AMERICA

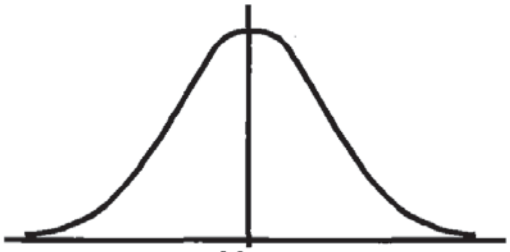
6

The Normal Distribution and Confidence Intervals

Described as a "bell shaped" curve, the normal distribution is a symmetrical distribution which is one of the most commonly occurring outcomes in nature and its presence is assumed in several of the most commonly used statistical tests. Properties of the normal distribution have a very important role in the statistical theory of drawing inferences about population parameters (estimating confidence intervals) based on samples drawn from that population.

The Normal Distribution

The normal distribution is the most important distribution in statistics. This curve is a special frequency distribution that describes the population distribution of many continuously distributed biological traits. The normal distribution is often referred to as the **Gaussian distribution**, after the mathematician Carl Friedreich Gauss, even though it was first discovered by the French mathematician Abraham DeMoivre (Porter, 1986).



It is critical at this point to realize that we are focusing our initial discussion on the *total population, not a sample*. As mentioned in the previous chapter, in the population, the mean is expressed as μ and standard deviation as σ.

The characteristics of a normal distribution are as follows. First, the normal distribution is continuous and the curve is symmetrical about the mean. Second, the mode, median, and mean are equal and represent the middle of the distribution. Third, since the mean and median are the same, the 50th percentile is at the mean with an equal amount of area under the curve, above and below the mean. Fourth, the probability of all possible outcomes is equal to 1.0, therefore, the total area under the curve is equal to 1.0. Since the mean is the 50th percentile, the area to left or right of the mean equals 0.5. Fifth, by definition, the area under the curve between one standard deviation above and one standard deviation below the mean contains an area equal to approximately 68% of the total area under the curve. At two standard deviations this area is approximately 95%. Sixth, as distance from the mean (in the positive or negative direction) approaches infinity, the frequency of occurrences approaches zero. This last point illustrates the fact that most observations cluster around the center of the distribution and very few occur at the extremes of the distribution. Also, if the curve is infinite in its bounds we cannot set absolute external limits on the distribution.

The frequency distribution (curve) for a normal distribution is defined as follows:

$$f_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2} \qquad \text{Eq. 6.1}$$

where: π (pi) = 3.14159 and e = 2.71828 (the base of natural logarithms).

In a normal distribution, the area under the curve between the mean and one standard deviation is approximately 34%. Because of the symmetry of the distribution, 68% of the curve would be divided equally above and below the mean. Why 34%? Why not a nice round number like 35%, 30%, or even better 25%? The standard deviation is that point of inflection on the normal curve where the frequency distribution stops its descent to the baseline and begins to pull parallel with the x-axis. Areas or proportions of the normal distribution associated with various standard deviations are seen in Figure 6.1.

The term "*the* bell-shaped curve" is a misnomer since there are many bell-shaped curves, ranging from those which are extremely peaked with very small ranges to those which are much flatter with wide distributions (Figure 6.2). A normal distribution is completely dependent on its parameters of μ and σ. A

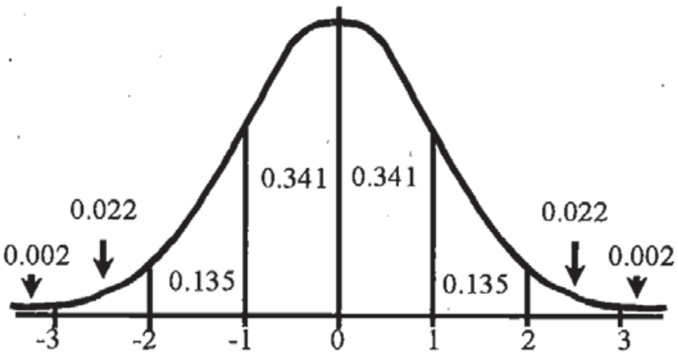


Figure 6.1 Proportions between various standard deviations under a normal distribution.

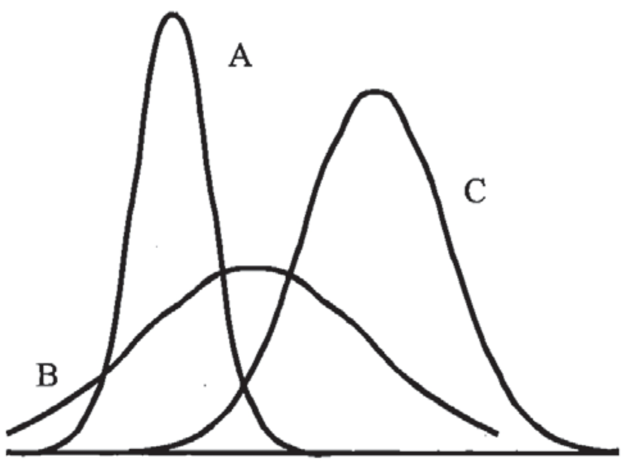


Figure 6.2 Example of three normal distributions with different means and different standard deviations.

standardized normal distribution has been created to compare and compute variations in such a distribution regardless of center or spread from the center. In this standard normal distribution the mean equals 0 (Figure 6.3). The spread of the distribution is also standardized by setting one standard deviation equal to +1 or -1, and two standard deviations equal to +2 or –2 (Figure 6.4).

As seen previously, the area between +2 and -2 is approximately 95%. Additionally, fractions of a standard deviation are calculated and their equivalent areas presented. If such a distribution can be created (with a mean equal to zero and standard deviation equal to one) then the equation for the frequency distribution can be simplified to:

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.