249

BIOSTATISTICAL ANALYSIS

FOURTH EDITION

JERROLD H. ZAR

Department of Biological Sciences
Northern Illinois University

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging in Publication Data

Zar, Jerrold H.
Biostatistical analysis / Jerrold H. Zar. — 4th ed.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0-13-081542-X (alk. paper)
1. Biometry. I. Title
QH323.5.Z37 1999
570'.1'5195—dc21 98-34062
CIP

Editorial/production supervision: *Interactive Composition Corporation*
Cover director: *Jayne Conte*
Cover designer: *Bruce Kenselaar*
Manufacturing manager: *Trudy Pisciotti*
Editor: *Teresa Ryu*
Senior editor: *Sheri L. Snavely*
Editorial assistants: *Nancy Bauer* and *Lisa Tarabokjia*

Printed in the United States of America

10 9

ISBN 0-13-081542-X

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S. A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

CONTENTS

27 CIRCULAR DISTRIBUTIONS: HYPOTHESIS TESTING (continued)

APPENDIX A ANALYSIS OF VARIANCE HYPOTHESIS TESTING App1

APPENDIX B STATISTICAL TABLES AND GRAPHS App11

24731

APPENDIX B *STATISTICAL TABLES AND GRAPHS (continued)*

would be described by the *binomial distribution* (sometimes referred to as the "Bernoulli distribution"*). Let us now examine binomial probabilities.

24.1 Binomial Probabilities

Consider a population consisting of two categories, where p is the proportion of individuals in one of the categories and $q = 1 - p$ is the proportion in the other. Then the probability of selecting at random from this population a member of the first category is p, and the probability of selecting a member of the second category is q.†

For example, let us say we have a population of female and male animals, in proportions of $p = 0.4$ and $q = 0.6$, respectively, and we take a random sample of two individuals from the population. The probability of the first being a female is p (i.e., 0.4) and the probability of the second being a female is also p. As the probability of two mutually exclusive events both occurring is the product of the probabilities of the two separate events (Section 5.7), the probability of having two females in a sample of two is $(p)(p) = p^2 = 0.16$; the probability of the sample of two consisting of two males is $(q)(q) = q^2 = 0.36$.

What is the probability of the sample of two consisting of one male and one female? This could occur by the first individual being a female and the second a male (with a probability of pq) or by the first being a male and the second a female (which would occur with a probability of qp). The probability of either of two mutually exclusive outcomes is the sum of the probabilities of each outcome (Section 5.6), so the probability of one female and one male in the sample is $pq + qp = 2pq = 2(0.4)(0.6) = 0.48$. Note that $0.16 + 0.36 + 0.48 = 1.00$.

Now consider another sample from this population, one where $n = 3$. The probability of all three individuals being female is $ppp = p^3 = (0.4)^3 = 0.064$. The probability of two females and one male is ppq (for a sequence of ♀♀♂) + pqp (for ♀♂♀) + qpp (for ♂♀♀), or $3p^2q = 3(0.4)^2(0.6) = 0.288$. The probability of one female and two males is pqq (for ♀♂♂) + qpq (for ♂♀♂) + qqp (for ♂♂♀), or $3pq^2 = 3(0.4)(0.6)^2 = 0.432$. And, finally, the probability of all three being males is $qqq = q^3 = (0.6)^3 = 0.216$. Note that $p^3 + 3p^2q + 3pq^2 + q^3 = 0.064 + 0.288 + 0.432 + 0.216 = 1.000$ (meaning that there is a 100% probability—that is, it is certain—that the three animals will be in one of these three combinations of sexes).

*The binomial formula discussed in the following section was first described in 1676 by Sir Isaac Newton (1642–1727), the great English scientist and mathematician; and its first proof, for positive integer exponents, was given by the Swiss mathematician, Jacques (also known as Jakob or James) Bernoulli (1654–1705) in a 1713 publication (Cajori, 1954). Each observed event from a binomial distribution is sometimes called a "Bernoulli trial." David (1995) ascribes the first use of the term "binomial distribution" to G. U. Yule, in 1911.

†This assumes "sampling with replacement." That is, each individual in the sample is taken at random from the population and then is returned to the population before the next member of the sample is selected. Sampling without replacement is discussed in Section 24.2. If the population is very large compared to the size of the sample, then sampling with and without replacement are indistinguishable in practice.

24733

If we performed the same exercise with $n = 4$, we would find that the probability of four females is $p^4 = (0.4)^4 = 0.0256$, the probability of three females (and one male) is $4p^3q = 4(0.4)^3(0.6) = 0.1536$, the probability of two females is $6p^2q^2 = 0.3456$, the probability of one female is $4pq^3 = 0.3456$, and the probability of no females (i.e., all four are male) is $q^4 = 0.1296$. (The sum of these five terms is 1.0000, a good arithmetic check.)

If a random sample of size n is taken from a binomial population, then the probability of X individuals being in one category (and, therefore, $n - X$ individuals in the second category) is

$$P(X) = \binom{n}{X} p^X q^{n-X}. \tag{24.1}$$

In this equation, $p^X q^{n-X}$ refers to the probability of sample consisting of X items, each having a probability of p, and $n - X$ items, each with probability q. The *binomial coefficient*,

$$\binom{n}{X} = \frac{n!}{X!(n - X)!}, \tag{24.2}$$

is the number of ways X items of one kind can be arranged with $n - X$ items of a second kind, or, in other words, the number of possible *combinations* of n items divided into one group of X items and a second group of $n - X$ items. (See Section 5.3 for a discussion of combinations; Equation 5.3 explained the factorial notation, "!".) Therefore, Equation 24.1 can be written as

$$P(X) = \frac{n!}{X!(n - X)!} p^X q^{n-X}. \tag{24.3}$$

Thus, $\binom{n}{X} p^X q^{n-X}$ is the Xth term in the expansion of $(p + q)^n$, and Table 24.1 shows this expansion for powers up through 6. Note that for any power, n, the sum of the two exponents in any term is n. Furthermore, the first term will always be p^n, the second will always contain $p^{n-1}q$, the third will always contain $p^{n-2}q^2$, etc., with the last term always being q^n. The sum of all the terms in a binomial expansion will always be 1.0, for $p + q = 1$, and $(p + q)^n = 1^n = 1$.

As for the coefficients of these terms in the binomial expansion, the Xth term of the nth power expansion can be calculated by Equation 24.3. Furthermore, the examination of these coefficients as shown in Table 24.2 has been deemed interesting for centuries.

TABLE 24.1 Expansion of the Binomial, $(p + q)^n$

n	$(p + q)^n$
1	$p + q$
2	$p^2 + 2pq + q^2$
3	$p^3 + 3p^2q + 3pq^2 + q^3$
4	$p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$
5	$p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$
6	$p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 - q^6$

TABLE 24.2 Binomial Coefficient, ${}_nC_X$

n	$X = 0$	1	2	3	4	5	6	7	8	9	10	Sum of coefficients
1	1	1										$2 = 2^1$
2	1	2	1									$4 = 2^2$
3	1	3	3	1								$8 = 2^3$
4	1	4	6	4	1							$16 = 2^4$
5	1	5	10	10	5	1						$32 = 2^5$
6	1	6	15	20	15	6	1					$64 = 2^6$
7	1	7	21	35	35	21	7	1				$128 = 2^7$
8	1	8	28	56	70	56	28	8	1			$256 = 2^8$
9	1	9	36	84	126	126	84	36	9	1		$512 = 2^9$
10	1	10	45	120	210	252	210	120	45	10	1	$1024 = 2^{10}$

This arrangement is known as *Pascal's triangle*.* We can see from this triangular array that any binomial coefficient is the sum of two coefficients on the line above it, namely,

$$\binom{n}{X} = \binom{n-1}{X-1} + \binom{n-1}{X}. \tag{24.4}$$

This can be more readily observed if we display the triangular array as follows:

$$\begin{array}{ccccccccccc} & & & & & 1 & & & & & \\ & & & & 1 & & 1 & & & & \\ & & & 1 & & 2 & & 1 & & & \\ & & 1 & & 3 & & 3 & & 1 & & \\ & 1 & & 4 & & 6 & & 4 & & 1 & \\ 1 & & 5 & & 10 & & 10 & & 5 & & 1 \end{array}$$

Also note that the sum of all coefficients for the nth power binomial expansion is 2^n. Appendix Table B.26a presents binomial coefficients for much larger n's and X's, and they will be found useful later in this chapter.

Thus, we can calculate probabilities of category frequencies occurring in random samples from binomial population. If, for example, a sample of five (i.e., $n = 5$) is taken from a population composed of 50% males and 50% females (i.e., $p = 0.5$ and $q = 0.5$) then Example 24.1 shows how Equation 24.3 is used to determine the probability of the sample containing 0 males, 1 male, 2 males, 3 males, 4 males, and 5 males. These

*Blaise Pascal (1623–1662), French mathematician and physicist and one of the founders of probability theory (in 1654, immediately before abandoning mathematics to become a religious recluse). He had his triangular binomial coefficient derivation published in 1665, although knowledge of the triangular properties appears in Chinese writings as early as 1303 (Cajori, 1954; David, 1962; Struik, 1967: 79). Pascal also invented (at age 19) a mechanical adding and subtracting machine which, though patented in 1649, proved too expensive to be practical to construct (Asimov, 1982: 130–131). His significant contributions to the study of fluid pressures have been honored by naming the international unit of pressure the pascal, which is a pressure of one newton per square meter (where a newton—named for Sir Isaac Newton—is the unit of force representing a one-kilogram mass accelerating at the rate of one meter per second per second). Pascal is also the name given to a modern computer language. The relationship of Pascal's triangle to ${}_nC_X$ was first published in 1685 by the English mathematician, John Wallis (1616–1703) (David, 1962: 123–124).

probabilities are found to be 0.03125, 0.15625, 0.31250, 0.31250, 0.15625, and 0.03125, respectively. This enables us to state that if we took 100 random samples of five animals each from the population, about three of the sample [i.e., (0.03125)(100) = 3.125 of them] would be expected to contain all females, about sixteen [i.e., (0.15625)(100) = 15.625] to contain one male and four females, thirty-one [i.e., (0.31250)(100)] to consist of two males and three females, etc. If we took 1400 random samples of five, then (0.03125)(1400) = 43.75 [i.e., about 44] of them would be expected to contain all females, etc. Figure 24.1a shows graphically the binomial distribution for $p = q = 0.5$, for $n = 5$. Note, from Fig 24.1a and Example 24.1, that when $p = q = 0.5$ the distribution is symmetrical [i.e., $P(0) = P(n)$, $P(1) = P(n-1)$, etc.], and Equation 24.3 becomes

$$P(X) = \frac{n!}{X!(n-X)!}0.5^n. \tag{24.5}$$

Appendix Table B.26b gives binomial probabilities for $n = 2$ to $n = 20$, for $p = 0.5$.

Example 24.2 presents the calculation of binomial probabilities for the case where $n = 5$, $p = 0.3$, and $q = 1 - 0.3 = 0.7$. Thus, if one were sampling a population consisting of 30% males and 70% females, 0.16807 (i.e., 16.807%) of the samples would be expected to contain no males, 0.36015 to contain one male and four females, etc. Fig. 24.1b presents this binomial distribution graphically, whereas Fig. 24.1c shows the distribution where $p = 0.1$ and $q = 0.9$.

For calculating binomial probabilities for large n, it is often convenient to employ logarithms. For this reason, Appendix Table B.40, a table of logarithms of factorials, is provided. Alternatively, it is useful to note that the denominator of Equation 24.3 cancels out much of the numerator, so that it is possible to simplify the computation of $P(X)$,

EXAMPLE 24.1 Computing binomial probabilities, $P(X)$, where $n = 5$, $p = 0.5$, and $q = 0.5$ (following Equation 24.3).

X	$P(X)$
0	$\frac{5!}{0!5!}(0.5^0)(0.5^5) = (1)(1.0)(0.03125) = 0.03125$
1	$\frac{5!}{1!4!}(0.5^1)(0.5^4) = (5)(0.5)(0.0625) = 0.15625$
2	$\frac{5!}{2!3!}(0.5^2)(0.5^3) = (10)(0.25)(0.125) = 0.31250$
3	$\frac{5!}{3!2!}(0.5^3)(0.5^2) = (10)(0.125)(0.25) = 0.31250$
4	$\frac{5!}{4!1!}(0.5^4)(0.5^1) = (5)(0.0625)(0.5) = 0.15625$
5	$\frac{5!}{5!0!}(0.5^5)(0.5^0) = (1)(0.03125)(1.0) = 0.03125$

24736

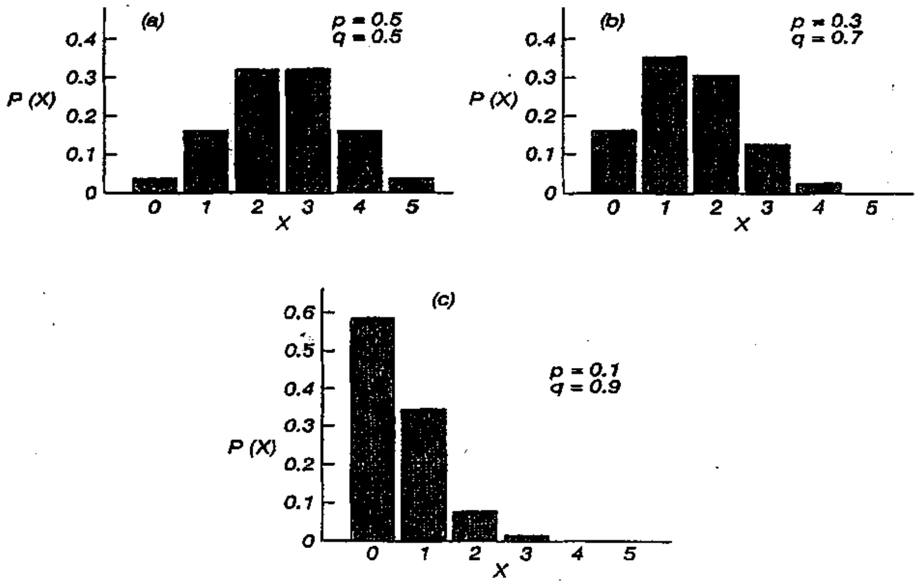

Figure 24.1 The binomial distribution, for $n = 5$. (a) $p = q = 0.5$. (b) $p = 0.3$, $q = 0.7$. (c) $p = 0.1$, $q = 0.9$. These graphs were drawn utilizing the proportions given by Equation 24.1.

especially in the tails of the distribution (i.e., for low X and for high X), as shown in Example 24.3. If p is very small, then the use of the Poisson distribution (Section 25.1), should be considered.*

The mean of a binomial distribution of counts X, is

$$\mu_X = np, \tag{24.6}$$

the variance† is

$$\sigma_X^2 = npq, \tag{24.8}$$

and the standard deviation of X is

$$\sigma_X = \sqrt{npq}. \tag{24.9}$$

---

*Raff (1956) and Molenaar (1969a, 1969b) discuss several approximations to the binomial distribution, including the normal and Poisson distributions.

†A measure of symmetry (see Section 6.1) for a binomial distribution is

$$\gamma_1 = \frac{q - p}{\sqrt{npq}}, \tag{24.7}$$

so it can be seen that $\gamma_1 = 0$ only when $p = q = 0.05$, $\gamma_1 > 0$ implies a distribution skewed to the right (as in Figs. 24.1b and 24.1c) and $\gamma_1 < 0$ indicates a distribution skewed to the left.

EXAMPLE 24.2 Computing binomial probabilities, $P(X)$, where $n = 5, p = 0.4, q = 0.7$ (following Equation 24.3).

X	$P(X)$
0	$\frac{5!}{0!5!}(0.3^0)(0.7^5) = (1)(1.0)(0.16807) = 0.16807$
1	$\frac{5!}{1!4!}(0.3^1)(0.7^4) = (5)(0.3)(0.2401) = 0.36015$
2	$\frac{5!}{2!3!}(0.3^2)(0.7^3) = (10)(0.09)(0.343) = 0.30870$
3	$\frac{5!}{3!2!}(0.3^3)(0.7^2) = (10)(0.027)(0.49) = 0.13230$
4	$\frac{5!}{4!1!}(0.3^4)(0.7^1) = (5)(0.0081)(0.7) = 0.02835$
5	$\frac{5!}{5!0!}(0.3^5)(0.7^0) = (1)(0.00243)(1.0) = 0.00243$

EXAMPLE 24.3 Computing binomial probabilities, $P(X)$, with $n = 400, p = 0.02$, and $q = 0.98$. (Many calculators can operate with large powers of numbers; otherwise, logarithms may be used.)

X	$P(X)$
0	$\frac{n!}{0!(n-0)!}p^0q^{n-0} = q^n = 0.98^{400} = 0.00031$
1	$\frac{n!}{1!(n-1)!}p^1q^{n-1} = npq^{n-1} = (400)(0.02)(0.98^{399}) = 0.00253$
2	$\frac{n!}{2!(n-2)!}p^2q^{n-2} = \frac{n(n-1)}{2!}p^2q^{n-2} = \frac{(400)(399)}{2}(0.02^2)(0.98^{398}) = 0.01028$
3	$\frac{n!}{3!(n-3)!}p^3q^{n-3} = \frac{n(n-1)(n-2)}{3!}p^3q^{n-3} = \frac{(400)(399)(398)}{(3)(2)}(0.02^3)(0.98^{397}) = 0.02784$

and so on.

Thus, if we have a binomially distributed population where p (e.g., the proportion of males) $= 0.5$ and q (e.g., the proportion of females) $= 0.5$ and we take ten samples from that population, the mean of the ten X's (i.e., the mean number of males per sample) would be expected to be $np = (10)(0.05) = 5$ and the standard deviation of the ten X's would be expected to be $\sqrt{npq} = \sqrt{(10)(0.5)(0.5)} = 1.58$. Our concern typically is with the distribution of the expected probabilities rather than the expected X's, as will be explained in Section 24.3.