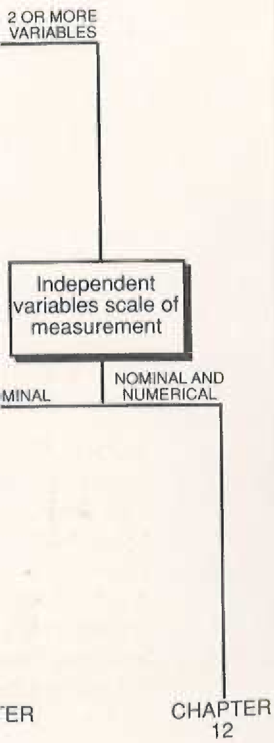
2

second edition

Basic & Clinical Biostatistics

Beth Dawson-Saunders
Robert G. Trapp

a LANGE medical book



a LANGE medical book

Basic & Clinical
Biostatistics

second edition

Beth Dawson-Saunders, PhD
Professor and Chief; Informatics, Biostatistics & Research
Department of Internal Medicine
Southern Illinois University
School of Medicine

Robert G. Trapp, MD
Medical Director
The Arthritis Center
Springfield, IL
Formerly
Assistant Professor and Chief of Rheumatology
Department of Internal Medicine
Southern Illinois University
School of Medicine

APPLETON & LANGE
Norwalk, Connecticut

97 98 / 10 9 8 7 6

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*
Prentice Hall, *Englewood Cliffs, New Jersey*

ISBN 0–8385–0542–2
ISSN 1045–5523

Acquisitions Editor: John Dolan
Production Editor: Christine Langan
Designer: Kathy Hornyak

PRINTED IN THE UNITED STATES OF AMERICA

ISBN 0-8385-0542-2
90000
9 780838 505427

Table 4–3. Calculations for standard deviation of typical change in vessel diameter (X) of 18 patients with no lesion growth.

Patient	X	$X-\overline{X}$	$(X-\overline{X})^2$
1	0.13	0.01	0.0001
2	0	−0.12	0.0144
3	−0.18	−0.30	0.0900
4	−0.15	−0.27	0.0729
5	0.11	−0.01	0.0001
6	0.43	0.31	0.0961
7	0.41	0.29	0.0841
8	−0.12	−0.24	0.0576
9	0.06	−0.06	0.0036
10	0.06	−0.06	0.0036
11	−0.19	−0.31	0.0961
12	0.39	0.27	0.0729
13	0.30	0.18	0.0324
14	0.18	0.06	0.0036
15	0.11	−0.01	0.0001
16	0.94	0.82	0.6724
17	−0.07	−0.19	0.0361
18	−0.23	−0.35	0.1225
Sums	$\Sigma X = 2.18$		$\Sigma(X-\overline{X})^2 = 1.4586$

the variation in the data, because the sum of the squared deviations is 1.4586. The standard deviation of the remaining 17 patients (after eliminating Patient 16) is substantially smaller, 0.22, demonstrating the overwhelming effect that even one outlying observation can have on the value of the standard deviation.

The standard deviation, like the mean, requires numerical data. Also like the mean, it is very important in statistics. First, it is an essential part of many **statistical tests** (which are discussed in detail in later chapters). Second, the standard deviation is very useful in describing the spread of the observations about the mean value. Two rules of thumb for using the standard deviation follow.

1. Regardless of how the observations are distributed, at least 75% of the values *always* lie between these two numbers: $\overline{X} - 2s$ and $\overline{X} + 2s$. In the vessel diameter example, the mean change $\overline{X}$ is 0.12 mm and the standard deviation s is 0.29 mm; therefore, at least 75% of the 18 observations are guaranteed to be between 0.12 − 2(0.29) and 0.12 + 2(0.29), or between −0.46 and +0.70 mm. In this example, 17 of the 18 observations (94%) actually are between these limits.

2. If the distribution of observations is a **bell-shaped distribution,** then even more can be said about the percentage of observations that lie between the mean and −2 standard deviations. For a bell-shaped distribution, approximately:

67% of the observations lie between $\overline{X} - 1s$ and $\overline{X} + 1s$.

95% of the observations lie between $\overline{X} - 2s$ and $\overline{X} + 2s$.

99.7% of the observations lie between $\overline{X} - 3s$ and $\overline{X} + 3s$.

For further discussion on the use of the mean and standard deviation with a bell-shaped distribution, see Chapter 5.

4.3.1.c The Coefficient of Variation: The coefficient of variation is a useful measure of *relative* spread in data and is used frequently in the biologic sciences. For example, suppose the authors of the study on diet and lipoproteins want to compare the variability in the ratio of total/HDL cholesterol with the variability in vessel diameter change for the 18 patients who had no lesion growth. The mean and the standard deviation of total/HDL cholesterol (in millimoles per liter) are 5.81 and 1.20, respectively; for the vessel diameter change (in millimeters), they are 0.12 and 0.29, respectively. A comparison of 1.20 and 0.29 makes no sense because cholesterol and vessel diameter are measured on different scales. The coefficient of variation adjusts the scales so that a sensible comparison can be made.

The coefficient of variation is defined as the standard deviation divided by the mean times 100%. It produces a measure of relative variation—variation that is relative to the size of the mean. The formula for the **coefficient of variation** is

$$CV = \left(\frac{s}{\overline{X}}\right)(100\%)$$

From this formula, the CV for total/HDL cholesterol is (1.20/5.81) (100%) = 20.7%, and the CV for vessel diameter change is (0.29/0.12) (100%) = 241.7%. Therefore, we can conclude that the relative variation in vessel diameter change is much greater than (more than 10 times as great as) that in cholesterol ratio.

A frequent application of the coefficient of variation is in laboratory testing and quality control procedures. For example, screening for neural tube defects is accomplished by measuring maternal serum alpha-fetoprotein. DiMaio et al (1987) evaluated the use of this test in a prospective study of 34,000 women. The reproducibility of the test procedure was determined by repeating the assay ten times in each of four pools of serum. They calculated the mean and the standard deviation of the ten assays in each pool of serum and then used them to find the coefficient of variation for each pool. The coefficients of variation for the four pools were 7.4%, 5.8%, 2.7%, and 2.4%. These values indicate relatively good reproducibility of the assay because the variation, as measured by the standard deviation, is small relative to the mean. Therefore, readers of their article can be confident that the assay results were consistent.

4.3.1.d Percentiles: A **percentile** is a number that indicates the percentage of a distribution that is equal to or below that number. For example, consider the standard physical growth chart for girls from birth to 36 months of age given in Fig 4–2 (Hamill et al, 1979). For girls 21 months of age, the 95th percentile of weight is 13.4 kg, as noted by the arrow in the