Statistics for the Clinical Laboratory

Irwin M. Weisbrot, M.D.

Statistics for the Clinical Laboratory

Irwin M. Weisbrot, M.D.

Senior Attending Pathologist
Department of Pathology
Norwalk Hospital
Norwalk, Connecticut

Associate Clinical Professor of Pathology
Yale University School of Medicine
New Haven, Connecticut

Lecturer
Department of Pathology
Mt. Sinai School of Medicine
New York, New York

1792

J. B. LIPPINCOTT COMPANY
Philadelphia
LONDON MEXICO CITY NEW YORK
ST. LOUIS SÃO PAULO SYDNEY

Acquisitions Editor: Lisa A. Biello
Sponsoring Editor: Sanford J. Robinson
Manuscript Editor: Dorothy Hoffman
Indexer: Pamela Fried
Art Director: Maria S. Karkucinski
Production Supervisor: J. Corey Gray
Production Assistant: Rosanne Hallowell
Compositor: Centennial Graphics, Inc.
Printer/Binder: R. R. Donnelley & Sons Company
Cover Printer: Philips Offset Co., Inc.

1 3 5 6 4 2

Library of Congress Cataloging in Publication Data

Weisbrot, Irwin M.
Statistics for the clinical laboratory.

Bibliography: p.
Includes index.
1. Diagnosis, Laboratory—Statistical methods.
2. Chemistry, Clinical—Statistical methods. I. Title.
[DNLM: 1. Statistics. QA 276.A2 W426s]
RB38.3.W45 1985 616.07'5'015195 83-26767
ISBN 0-397-50620-1

Preface

Like most laboratorians, you probably already have one or perhaps two books on statistics at home lying on a bookshelf. Like most, you have probably passed it by many times, selecting a spy thriller instead; feeling guilty, of course, for such a transgression; despoiling again the secret visions you have of mastering the damned thing and confronting your colleagues with quick statistical wit, wielding a piece of chalk like a rapier, reducing their simple numerical inferences into reversed, inside out, and exactly opposite truths.

While such motives may be less than honorable, you nevertheless have my sympathy, for they reflect the frustration many of us endure when confronted with the need to follow basic arguments concerning quantitative data. Let me tell you immediately that this little book will not convert a mathematical milquetoast into a computerized paragon of statistical virtue.

Why, then, ought you to buy or borrow *Statistics for the Clinical Laboratory*? In truth, there is really no earthly reason. There are many good books around, and I refer to them frequently. The best I have found for my purposes are listed at the end of Chapter 1, and to master any one of them would provide a very ample base upon which to expand. I would at first eschew the average college statistical text, as most are too concerned with mathematical proof and justification. Frequently they state the equation, and presume the way the statistic works flows naturally from that. (It does, but not for most of us.)

Undoubtedly, most of you have already had a statistics workshop or two at regional or national meetings; some of you may even have taken and passed a required course in statistics, and yet, somehow, something remains elusive, discomfiting.

Mostly, I think, the difficulty exists because you have not come away from those episodes with any way to relate the *t*-test and the F-test to reality, to your own experiences. In fact, they are only convenient extensions of reality, and of our intuitions, merely set to very specific and limited ground rules. Indeed, most of you probably perform the essences of *t*-tests and F-tests 10 or 15 times a day every time you decide that the run of tests must be valid if the standards and controls look good, or the result must be correct if three different techs got the same value. It's just that the statistical test gives us some idea of the *probability* that our data support the conclusion when we announce this acceptance or that rejection of our hypothesis—but only if we and the data have played by the rules of the statistic and its assumptions.

For most of us there is no holy grail of statistics which will enlighten effortlessly. Select any book you will, but work with it; think about it. Try to relate the statistics to what they are trying to say. Lean back, close your eyes, try to see what the diagrams and explanations are—5, 10, 15 times a day if necessary, at odd, wasted moments.

Modesty prohibits me from telling of the circumstances during which the standard error of the mean became clear to me, and I dare not describe in any detail whatsoever my circumstances when the central limit theorem made itself known to me. For each it will be different, but worth it. For once you have internalized and understood a statistical concept, however simple, you have become less vulnerable to the random, less hostage to the exotic.

Why, then, buy or borrow this book? First, because I have tried to introduce simple and basic statistics using explanations that worked for me when I groped with them. Second, I have borrowed generously from some good friends, and from many I wish I knew as friends, to show how statistics help us to run better laboratories and to produce more rational results from the clinical laboratory.

Irwin M. Weisbrot, M.D.

Contents

1

Basic Statistics

This chapter introduces basic statistical concepts and describes those basic statistical tests used daily in the clinical laboratory. Its aim is to assist the reader in gaining an intuitive feel for the mechanisms of a few fundamental tests, their applications, their limitations, and, most of all, their pitfalls if misunderstood or misapplied. It is not meant to be a complete statistical text or compendium of all possible tests available. My professional statistical colleagues agree that amateur statisticians will seldom err if they use a few statistical tests they thoroughly understand, however inelegant these tests may be.

Should laboratory workers have access to expert statistical consultation, they ought to design their experiments or processes using the statistics they know prior to such consultation. Then, if revision is suggested, they should request clarification contrasting the opposing statistical approaches. Frequently, they will find that there is no fundamental difference. Often they will have learned something new and useful. Sometimes they will have to teach the statistician about the vagaries and fickleness of biologic systems.

Remember, we do not use statistics because we are smart; rather we use them because our minds cannot grasp large numerical arrays or attributes and produce abstract generalizations from them. We must chip away, order them, and rank them before we can describe their central tendencies, dispersions, and distributions, followed sometimes by a great leap toward inferences about them. Nevertheless, with practice, rough guesses about the means and distributions of the data can be made, a skill worth having to avoid accepting ludicrous statistics.

Finally, statistical nomenclature and notation often lack uniformity from one author or publisher to the next. Learning to recognize the process will minimize this difficulty.

Significant Figures and Rounding Off

In general, the original determination should contain only numbers that are analytically certain. For instance, ordinary serum urea and glucose levels are reported to a whole number without a decimal. The means and standard deviations of the series of such determinations theoretically have an infinite number of significant figures but ordinarily are reported to one or at most two figures beyond analytic capability. In fact, with modern calculators and computers there is no difficulty in carrying 10 or 15 places, and it is wise to do the computations carrying as many significant places as the computer can handle and rounding off only in the final statement.

The traditional way of rounding off is based on the integer following the last significant number. If the integer is 4 or less, it is discarded, and no change is made in the preceding number. If the integer is 6 or more, the preceding digit is rounded up by 1. If the number is 5 and the last significant digit is odd, the number is increased by 1, but if the digit to the left is even, it remains unchanged. Zero is an even number. This produces small errors that are not generally important.[6]

In the examples that follow it has sometimes been necessary to round off values for ease of tabulation, but usually at least three decimal places have been carried. Expect small rounding errors to yield small differences in your calculations.

Some Basic Definitions

Descriptive statistics are concerned with the mean, range, variability, and distribution of a data set.

Inferential statistics are concerned with the relationships among different sets or samples of data. Is the mean of one set greater than the other? Is the dispersion of one sample really different from that of the second sample? Yes, we find a difference between the means of the two samples, but is it real or significant?

Certain definitions are worth stating here and repeating as necessary in the discussion to follow.

Population refers to the universe of values or attributes such as all the fasting plasma glucose levels of all hospitalized patients in all hospitals in the northeastern United States.

Sample refers to a portion or subset of a population such as the fasting plasma glucose levels on ward CP5 of Norwalk Hospital in Connecticut. Even this could be considered a population in its own right, further stratified by age, gender, or other factors.

Parameter describes a quantitative attribute of a population.

Statistic describes a quantitative attribute of a sample.

PARAMETER		STATISTIC
μ (mu)	Arithmetic mean or average	$\bar{x}$ ($\bar{x}$-bar)
σ (sigma)	Standard deviation	s
$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n}$	Variance	$(s^2) = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}$
$Z = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$	Distribution of means	$t = \frac{\bar{x}_1 - \bar{x}_0}{s/\sqrt{n}}$
$\mu \pm 1.96\sigma$	95% confidence interval	$\bar{x} \pm t_{1-\alpha}s$

Don't be concerned if these symbols are not very clear now. Observe that parameters usually have Greek letters symbolizing them, whereas statistics must usually be content with less elegant Roman or italic characters. But as you make progress in your statistical education you will realize that a slightly different test format, with substantial difference in your resulting inference, may depend upon whether your data are to be treated as population parameters or sample statistics. Return to this list after reading the first few chapters and see if it doesn't make more sense then.

Bias is commonly used with the following slightly different meanings in the clinical laboratory:

1. The difference between two means, the mean difference.
2. The presence of nonrandom events, which make the sampled population nonrepresentative of the target population. Example: estimation of the average value of plasma glucose in a mixed hospital population just as all the specimens from diabetic clinic arrive. Example: severe electrical power fluctuations during the morning's run for plasma glucose while data on repeatability is being collected. Example: preference of even over odd in reading a burette.
3. Lack of accuracy.

Random means unpredictable: no algorithm or formula can predict the next event. Having a rectangular distribution, all events appear equally probable.

Measures of Central Tendency

The *arithmetic average* or *mean* (signified as $\bar{x}$, pronounced "x-bar") is the most useful and common method of describing data (*e.g.*, plasma glucose, body weight, intelligence). It may be generalized as

$$\frac{1}{n} \sum_{n}^{i=1} x_i$$

which tells you to sum the n number of individual values of x and divide by n. The data are not grouped; the interval equals 1.

Consider the numbers in Column 1 of Table 1-1. The arithmetic mean, $\bar{x}$, is 9.61 mg/dl, and for Column 2 it is 9.54 mg/dl. The numbers 9.61 and 9.54 do not appear as actual values; they are idealized descriptors.

Table 1-1. Serum Calcium Values (mg/dl)

	DETERMINATION 1	DETERMINATION 2	DETERMINATION 3
1	9.2	9.0	8.7
2	9.3	9.2	8.9
3	9.5	9.3	9.3
4	9.6	9.4	9.6
5	9.6	9.6	9.6
6	9.6	9.6	9.7
7	9.7	9.6	9.9
8	9.8	9.8	9.9
9	9.9	9.9	9.9
10	9.9	10.0	10.0
Σ	96.1	95.4	95.5
$\bar{x}$	9.61	9.54	9.55
s	0.2331	0.3169	0.4478
Median	9.6	9.6	9.6
Mode	9.6	9.6	9.9
Range	9.2–9.9	9.0–10.0	8.7–10
Central 80%	9.3–9.9	9.2–9.9	8.9–9.9

Each column represents replicate determinations on one patient serum.

(Weisbrot IM: Basic statistics, quality control, normal values, and comparison of methods. In Race GJ [ed]: Laboratory Medicine, Vol 3, Chap 32, p 3. Philadelphia, J B Lippincott, 1983)

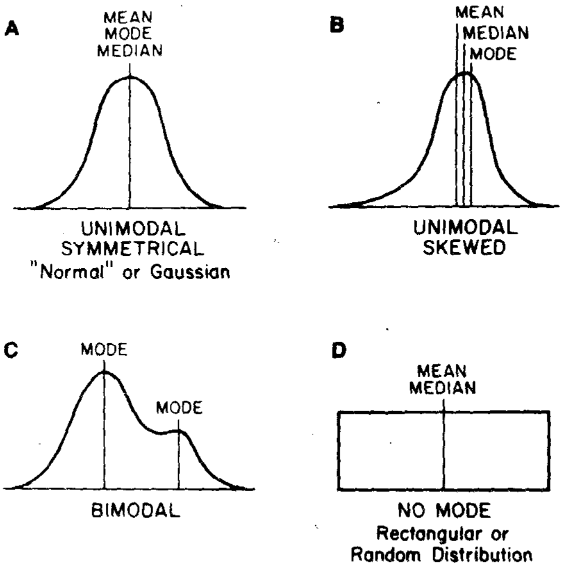

Fig. 1–1. Typical distributions of data; many more are possible.

The *median* is the middle value or the 50th percentile value when the data are rank ordered by magnitude. Half the data points are above the half below the median. Calculation is not always as straightforward as the arithmetic mean, nor can the median be easily manipulated.

The *mode* is the data point that occurs most frequently. In both Columns 1 and 2 of Table 1-1, 9.6 is the mode. Data distributions with one mode are unimodal (Fig. 1-1). Data may be bimodal or even polymodal or have no distinct mode at all. Although the mode, like the median, is not easily manipulated, it is graphically demonstrable. It may permit recognition of two or more distinct populations, sometimes permitting the identification of a normal and abnormal subset within the large population.

When the mean, median, and mode coincide, the data follow a symmetric distribution of fundamental importance to statistical concepts.

When data are very skewed, neither the mean nor the mode may estimate central tendency well, although the median retains its distinction. One approach is to use the geometric or the logarithmic mean:

$$\bar{x}_{geom} = \sqrt[n]{(x_1)(x_2)(x_3)\ldots(x_n)}$$

(take the *n*th root of the product)

$$\bar{x}_{log} = \text{antilog}\,\frac{1}{n}\,\Sigma \log x_i$$

(sum the log of each value, average them, take the antilog)

The log mean is simpler to calculate, especially with the small calculators available today. Data requiring such manipulation are not common in clinical chemistry laboratory work. They occur mostly in determining lethal dose levels in pharmacology and in dilution titers in clinical immunology.

Consider a series of titers obtained on diluted serum—1:10, 1:20, 1:20, 1:40, 1:160—from which the mean is desired.

Converting to decimals, we have

0.1, 0.05, 0.05, 0.025, and 0.00625

The geometric mean for the series would be

$$\sqrt[5]{.1 \times 0.05 \times 0.05 \times 0.025 \times 0.00625} = \sqrt[5]{3.9063 \times 10^{-8}}$$

Because most pocket computers will raise to a power more readily than they will extract unusual roots, and remembering that the n^{th} root is the same as the $1/n^{th}$ power, we have

$$(3.9063 \times 10^{-8})^{0.2} = 0.032988 = 1{:}30.3$$

as the mean titer for the series.

Similarly, using the log formula, and remembering that multiplying a logarithm raises to an exponent and dividing a logarithm yields a root, the result would be equal to the antilog of

$$0.2\,\Sigma(-1 - 1.30103 - 1.30103 - 1.60206 - 2.20412) = 0.032988 = 1{:}30.3$$

By comparison, the arithmetic average would be 1:21.6.

Measures of Dispersion

RANGE

One measure of how the data are distributed around or in relationship to the mean is range. For instance, in Column 1 of Table

1-1 the calcium range is 9.2 to 9.9 in ten replicates and 9.0 to 10.0 in Column 2. Experience tells us that the whole range of data frequently includes values that are borderline or marginal and not entirely representative. We could assume that the first and last values are somehow suspect and we could determine a *truncated range*. We could select the central 80% interval, discarding the lower and upper 10% of values. This yields a range of 9.3 to 9.9 and 9.2 to 9.9, respectively, for Columns 1 and 2.

Although useful, range is limited in describing the relationships between the data and the mean, and, although it tells us what the spread of the data was, it provides no formal way of predicting what the dispersion is likely to be in the future. There is no easily derived or easily used algorithm or parameter that enables us to anticipate future data (provided, of course, no system changes have occurred). Nor can we easily determine the probability that the dispersion has changed during the next trial, should the next range not be identical to the first.

MEAN DEVIATION

Mean deviation is the average amount each data point differs or deviates from the mean. It is determined as the following:

$$\Sigma \frac{|x_i - \bar{x}|}{n}$$

or the sum of the absolute differences of each value from the mean, disregarding the ± sign, divided by n, or

$$\Sigma \frac{|d|}{n}$$

The calcium values from Column 1 in Table 1-1 are again listed in Table 1-2 to illustrate the following calculations.

Summing the absolute $|x - \bar{x}|$ values, we get 1.72. Dividing by 10, the mean or average deviation becomes 0.17 mg of calcium per deciliter. This statement seems to make sense upon inspection of the original data because roughly half the calcium values are within 0.17 mg of the mean. However, if the signs are retained and the sum of $x - \bar{x}$ is taken, we find that the positive deviations cancel the minus deviations and there is no deviation. Such mis-

Table 1-2. Calculation of Mean Deviation and Standard Deviation Data from Column 1, Table 1

	x	$x - \bar{x}$	$\|x - \bar{x}\|$	$(x - \bar{x})^2$
	9.2	−0.41	0.41	0.1681
	9.3	−0.31	0.31	0.0961
	9.5	−0.11	0.11	0.0121
	9.6	−0.01	0.01	0.0001
	9.6	−0.01	0.01	0.0001
	9.6	−0.01	0.01	0.0001
	9.7	0.09	0.09	0.0081
	9.8	0.19	0.19	0.0361
	9.9	0.29	0.29	0.0841
	9.9	0.29	0.29	0.0841
Σ	96.1	0.00	1.72	$0.4890 = \Sigma(x - \bar{x})^2$
$\bar{x}$	9.61	0.00	0.172	
s^2				$\frac{0.4890}{9} = 0.05433 =$ variance
s				$\sqrt{0.05433} = 0.2331$

(Weisbrot IM: Basic statistics, quality control, normal values, and comparison of methods. In Race GJ [ed]: Laboratory Medicine, Vol 3, Chap 32, p 4. Philadelphia, J B Lippincott, 1983)

constructions have appeared in respectable medical literature and they remind us that when all else fails, one should look at the data.

Mean deviation has an efficiency of 0.88, meaning that an n of 100 is required to yield as good an estimate of dispersion as can be attained with an n of 88 used to calculate standard deviation. Mean deviation was more widely used before the days of calculators and computers. It has many of the defects described for range as a measure of dispersion, but it does give a single numerical value.

To estimate standard deviation, multiply mean deviation by 1.253 for large samples.[3]

STANDARD DEVIATION

Standard deviation, s, has become the principle estimator of dispersion because (1) the problem of negative differences is obviated