Statistics for Nuclear Medicine

Introduction, Descriptive Statistics, and Graphic Displays

Peter C. O'Brien, Marc A. Shampo, and James S. Robertson

Mayo Clinic and Mayo Foundation, Rochester, Minnesota

J Nucl Med 24: 83–88, 1983

Because the field of statistics has become so important in medicine, it is well worth the time of a physician in nuclear medicine to become acquainted with the language of statistics, the elementary concepts, and a few of the more commonly used procedures. For this purpose we are presenting these essentials in a series of six short, nontechnical papers. Obviously, what can be accomplished in the brief space allotted each of them is limited. However, the reader can expect to gain an understanding of what statistics is, an ability to understand the statistical aspects of much of the medical literature, a feel for when it will be necessary to consult with a statistician, and—for those occasions—an ability to communicate effectively with him.

Unfortunately, such an elementary acquaintance as we offer may lead a reader to overestimate his statistical capabilities and fail to consult a statistician in undertaking a research effort. We do not believe that this series of papers, or any review of statistics at the introductory level, will enable anyone to proceed without professional assistance in medical research requiring statistical expertise.

Medical research studies may be classified into two broad categories. Descriptive studies are intended to describe the characteristics of only the study group, using observations obtained from every member of the group. Inferential studies, on the other hand, are designed to enable the investigator to use observations from selected individuals (a sample) to form conclusions about the larger group (population) from which they were drawn.

First, we will deal with descriptive studies, focusing on summary statistics (such as the mean and median) and graphic techniques (such as histograms and scatter diagrams). We then describe how one may estimate characteristics of the population from characteristics of a small number of its members randomly selected. These principles are then applied to the problem of testing hypotheses about the population by the use of some of the more common testing procedures.

The last three papers discuss topics in medical research common to both types of research studies. Included are some problems that arise in evaluating the association between two characteristics, analyzing survival data (where one must be careful if not all persons in the study were observed until death, or if some may have been followed up longer than others), determining normal values, evaluating a new medical procedure, and applying sequential statistical methods (which enable the investigator to test hypotheses while the study is in progress, with a view toward terminating the study early).

As we have indicated, our purpose is to offer an acquaintance with these topics for only a small investment of the reader's time. Throughout, the discussion will be kept at an elementary level, omitting all mathematical derivations and, as much as possible, mathematical formulas. It is our hope that, upon completion of this series, the reader will be encouraged to go on to a further study of statistics. Many excellent elementary textbooks are available.

DESCRIPTIVE STATISTICS

Statistics is the mathematical technique or process of gathering, describing, organizing, analyzing, and interpreting numerical data. In describing a set of numerical data, we are especially interested in two of their characteristics: typical values and variability.

Typical values. It is often desirable to characterize a set of numbers by a single value that is considered to be typical. Among several kinds of such values, the ones used most commonly are the mean and median.

Mean. This is computed by summing the individual data points, then dividing this sum by the number of

For reprints contact: Dr. O'Brien, Section of Medical Research Statistics, Mayo Clinic and Mayo Foundation, 200 First St. SW, Rochester, MN 55905.

observations (N) in the data set. We illustrate with the following hypothetical data:

$$-2, 0, 2, 4, 6 \ (N = 5).$$

$$\text{The mean is } \frac{-2+0+2+4+6}{5} = \frac{10}{5} = 2.$$

Median. If N is odd, the median is defined as the middle value: half the other observations are equal to it or smaller, and half are equal to it or larger. For the data set (−2, 0, 2, 4, 6), the median is 2. If N is even, one takes the midpoint between the two inner values: the median of (1, 5, 6, 7) is 5.5; and the median of (4, 10, 18, 36) is 14.

Others. Another kind of central value is the mode, defined as the most frequent or most common value. Other types of averages are sometimes encountered in addition to the arithmetic mean described above, such as a geometric mean or a harmonic mean, but these will not be discussed further in our series on elementary statistics.

Variability. Regardless of which summary term (mean or median) has been used to characterize the center of the data, the question of variability arises. Specifically, one is interested in the range of values that occur most commonly and how closely individual values tend to cluster around the center.

A useful method is to determine the 25th percentile (P_{25}) and the 75th percentile (P_{75}). Of all the values under consideration, 25% lie below P_{25} and 75% lie below P_{75}. The interquartile range (also called the semiquartile range) extends from P_{25} to P_{75} and this range includes 50% of the data points. In some instances, an investigator may find other percentiles more appropriate.

In very small data sets, an informative statement regarding variability is given by the range—the smallest value and the largest. However, a disadvantage of the range is that it depends heavily on the size of N: as more observations are included (as N becomes larger), the range usually gets larger (though it may remain unchanged). The range also may be greatly influenced by outliers, as will be illustrated below.

Another statistic that is commonly used to describe the variability in a set of data is the standard deviation. This usage of the standard deviation appears to derive largely from the mistaken belief that 95% of the observations can be expected to lie within two standard deviations from the mean. The falsity of this proposition is easily demonstrated, for it is true only under special, infrequently occurring conditions (such as the error observed in a repeated series of measurements). Thus the appropriateness of the standard deviation for descriptive purposes is somewhat limited. However, it is useful in other contexts (relating to the sample mean) that will be discussed in later papers. The computations required for calculating the standard deviation are illustrated below.

Step 1. Square the deviation of each individual value from the mean.

Step 2. Sum the squared deviations.

Step 3. Divide the sum by N − 1. The result is called the variance (s^2).

Step 4. Obtain the standard deviation (that is, s) by taking the square root of the variance ($\sqrt{s^2}$).

Example

Step 1.

Original data	Deviation from mean of +2	Deviation squared
−2	4	16
0	2	4
+2	0	0
+4	2	4
+6	4	16
Step 2. Sum of squared deviations =		40

Step 3. (N = 5)

$$s^2 = \frac{\text{sum of squared deviations}}{N-1} = \frac{40}{4} = 10$$

Step 4. $s = \sqrt{10} = 3.16$.

Outliers and skewness. Although the mean and standard deviation are the most commonly used statistics for describing typical values and variability exhibited by a set of data, they are not appropriate when outliers or skewness is present. For example, suppose measurements of bone mineral density of the lumbar spinal column in seven 70-yr-old women produced the values 0.71, 0.73, 0.77, 0.77, 0.78, 0.80, and 1.50 g/cm^2. The value of 1.50, clearly dissimilar to the other six observations, is termed an "outlier." When it is included in the series, the mean is 0.866, which is larger than six of the seven data points. The standard deviation, 0.281, is more than twice the range of the remaining six points when the outlier is omitted. Clearly, in this instance, the mean and standard deviation do not provide an accurate description of the set of data. In this case, the data would be described more accurately by a statement that the median value is 0.77, six values range from 0.71 to 0.80, and one value is 1.50.

As an example of skewness, suppose the seven values had been 0.71, 0.73, 0.77, 0.83, 0.94, 1.20, and 1.50 g/cm^2. The mean and standard deviation are 0.954 and 0.294, respectively. Note that the span from the smallest value to the median is only 0.12 unit (0.71 to 0.83), whereas the span from the median to the largest value is 0.67 unit (0.83 to 1.50). (When the values are arranged in order of increasing size and those greater than the median are more spread out than those smaller than the median, we say the data are skewed to the right. This is a common occurrence, particularly with data that cannot

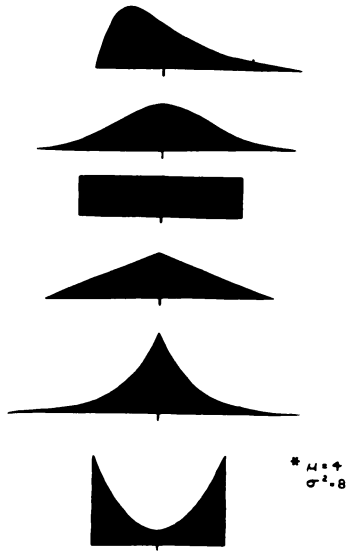

FIG. 1. Six data sets with same mean ($\bar{x}$ = 4) and same standard deviation (s = 2.83). (From Elveback LR: A discussion of some estimation problems encountered in establishing normal values. In *Clinically Oriented Documentation of Laboratory Data*. Edited by ER Gabrieli. New York, Academic Press, 1972, pp 117–138. By permission.)

be negative, such as the usual laboratory measurements. Less frequently one encounters data that are skewed to the left.) Again the mean and standard deviation fail to represent accurately the typical values and dispersion. The median (0.83) and range (0.71 to 1.50) would convey this information better.

Generally, when data are highly skewed or when outliers are present, the center is represented more meaningfully by the median. Variability usually is best described by quoting appropriate percentiles or the range (or both), especially when outliers or skewness is present. Ultimately, of course, the descriptive statistics discussed above remain a summary. Considerably more information may be conveyed by a graphic display.

Limitations of the mean and the standard deviation are illustrated by Fig. 1, which shows the manner in which individual values of six hypothetical data sets are distributed about the mean. (For example, notice that in the distribution at the top of Fig. 1, most of the values are less than the mean, with the data skewed to the right.) Although the data sets depicted are very different, all six have the same mean ($\bar{x}$ = 4) and the same standard deviation (s = 2.83).

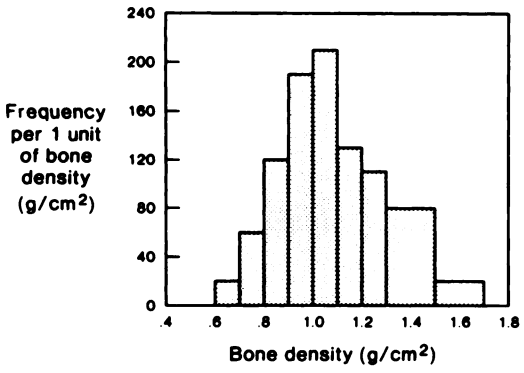

FIG. 2. Histogram of bone density values from 105 normal women corresponding to data in Table 1. Abscissa (x-axis) has unequal intervals corresponding to column B in Table 1. Ordinate (y-axis) has values corresponding to column D in Table 1.

GRAPHIC DISPLAYS

Graphic displays can be quite useful in conveying a quick visual impression of large data sets.

Histograms. A very useful graph for this purpose is the histogram, in which frequency is represented by area. For example, Fig. 2 shows the distribution of bone density values from 105 normal women (*1*). It can be seen that there are more values in the interval from 1.0 to 1.1 g/cm^2 than in any other interval of comparable width, and that most of the values are less than 1.3 g/cm^2 (the area to the left of 1.3 g/cm^2 is most of the total area). To provide an understanding of histograms, we will work through the steps that produced Fig. 2.

The first step is to list the observations in order of size, indicating the frequency with which each observation occurs (Table 1). One then forms class intervals, grouping the data according to intervals of interest, or in such a way as to ensure that each interval contains at least some minimal number of observations. On occasion, one may wish to use unequal class intervals. For example, in describing the age at which death has occurred in a group of subjects, the first year of life may be of special interest; if so, class intervals 0–1, 2–9, 10–19, 20–29, and so forth may be desirable. To illustrate the technique for this expedient in the example involving the bone density values, unequal class intervals (columns A and B of Table 1) have been chosen, which will cause the columns in the histogram to be of unequal width.

In column C of Table 1 are the frequencies, or the number of observations that fall within each interval (in the example, the numbers of subjects whose bone density values fall within each interval). If all of the intervals

TABLE 1. DISTRIBUTION OF BONE DENSITY IN 105 NORMAL WOMEN

Bone density, g/cm^2 (A)	Width of interval (B)	Frequency (C)	Frequency ÷ width (D)	Cumulative % of subjects (E)
0.60–0.69	0.10	2	20	1.9
0.70–0.79	0.10	6	60	7.6
0.80–0.89	0.10	12	120	19.0
0.90–0.99	0.10	19	190	37.1
1.00–1.09	0.10	21	210	57.1
1.10–1.19	0.10	13	130	69.5
1.20–1.29	0.10	11	110	80.0
1.30–1.49	0.20	16	80	95.2
1.50–1.69	0.20	5	25	100.0

were of equal size, these frequencies would suffice to determine the relative heights of the bars to be plotted in the histogram. Since the widths of the intervals are unequal and the frequency is to be represented by area (width × height), one must solve: frequency = width × height. Thus,

$$\text{height} = \frac{\text{frequency}}{\text{width}}$$

= frequency per unit measurement (frequency per 1 g/cm² of bone density).

With that information, Fig. 2 can be constructed.

Except in the case of *very* large data sets, one must consider the problem of choosing interval widths, keeping in mind the twin objectives of accurate detail and reliable overall description of the distribution. These considerations are illustrated in Fig. 3. Apparently, many of the peaks that are seen with use of 0.1 as the interval width are artifacts—notice that they disappear when an interval width of 0.3 is used. Conversely, with intervals of 1.0 virtually all detail is lost. However, no recommendation will be made for choosing between the two histograms in the middle (class intervals of 0.3 or 0.5) other than to point out that—as will often be the case—the informed judgment of the investigator will probably serve better than any rule of thumb.

Whatever class intervals are chosen, whether of equal or unequal width, the horizontal axis should be marked at regular intervals (like a ruler), as in Fig. 2. The vertical axis should start at 0 and also be marked at regular intervals, and should not be broken.

Frequency polygons. Frequency polygons provide a useful method for comparing two data sets on the same graph. (If the sets are not of the same size, their distributions first are made proportional, usually by conversion to a percentage basis.) To draw a frequency polygon, one simply connects the midpoints of the tops of successive bars of the histogram (made with percentage frequencies), as shown in Fig. 4A. A frequency-polygon comparison of bone density values from the 105 normal women in our previous example with the corresponding values from 82 normal men is shown in Fig. 4B.

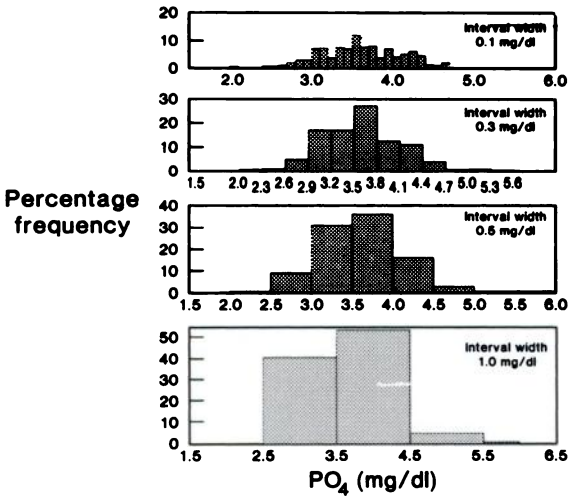

FIG. 3. Histograms of PO_4 levels in 329 females, plotted with interval widths of 0.1, 0.3, 0.5, and 1.0 mg/dl. (From O'Brien PC, Shampo MA. Statistics for Clinicians. 2. Graphic Displays—Histograms, Frequency Polygons, and Cumulative Distribution Polygons. *Mayo Clin Proc* 56:126–128, 1981.)

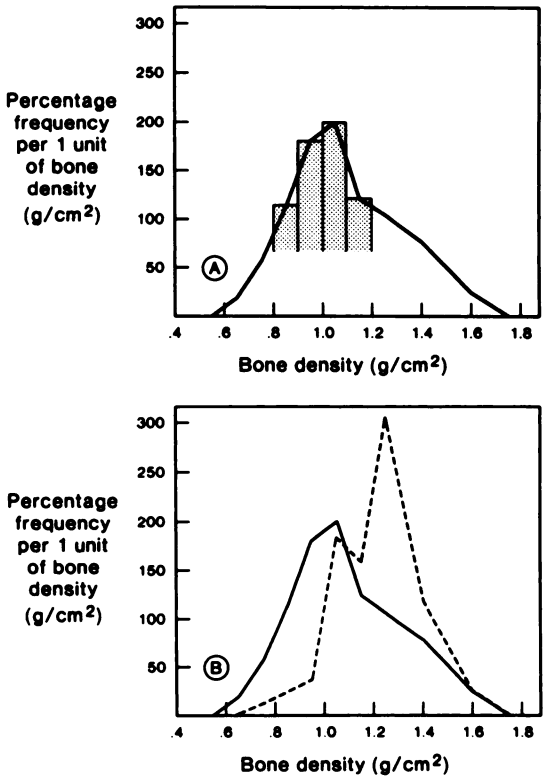

FIG. 4. Frequency polygons representing bone density values. Frequencies are expressed as percentages of the respective totals. A: data from 105 normal women (From Fig. 2 and Table 1). B: data from 105 normal women (—) and 82 normal men (- - -).

Cumulative distribution polygons. Another very useful method for displaying the distribution of a data set is provided by the cumulative distribution polygon (Fig. 5), which shows the percentage of observations less than

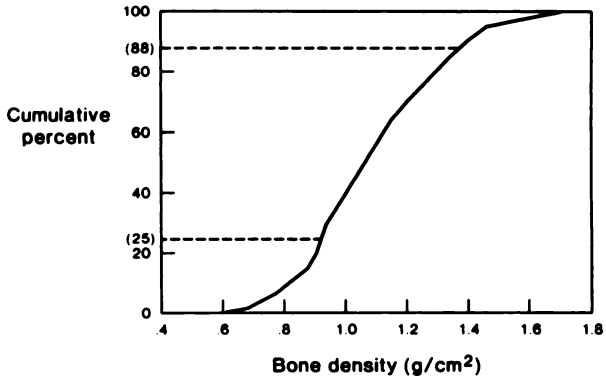

FIG. 5. Cumulative percentage polygon of bone density values from 105 normal women (from column E in Table 1).

any given value. Any desired percentile can be obtained from it as well. For example, Fig. 5 indicates that among the set of 105 bone density values in our familiar example, 1.38 corresponds to the 88th percentile (88% of the observations were less than 1.38).

The graph is constructed by connecting consecutive points from the cumulative distribution (column E of Table 1) with straight-line segments. Cumulative frequency polygons can be plotted together, just as frequency polygons can; and this provides another way to compare sets of data.

Scatter Diagrams. Some graphic displays can be used for presenting data on a single continuous variable, such as bone density. However, we now consider graphing the relationship between two continuous variables—for example, between age and serum IgE concentration. tween age and serum IgE concentration.

The appropriate graph is a scatter diagram (Fig. 6). Each point in the scatter diagram is determined by two values. In our example, each patient will be represented by a single point whose location is determined by his age (on the horizontal scale) and his IgE value (on the vertical scale).

The first step in preparing a scatter diagram is to determine the range for each variable, so that the axes may be properly labeled. The graph should be approximately square, with no values plotted on the axes themselves. For a scatter diagram—unlike the graphs in the previous section—it is not necessary to start either axis at 0.

A scatter diagram should be one of the first steps in data analysis. Data features that otherwise might go undetected may become obvious on the scatter diagram.

For example, in Fig. 6 it is apparent that one subject (→) is considerably older than the others in the group. Also, there are more patients with IgE values below the mean value (286 ng/ml) than above it. It can be said that the age of 70 yr is an outlier and the data on IgE are skewed. These are important elements to consider the selecting appropriate descriptive statistics. For the present data, medians and ranges would be preferable to means and standard deviations.

This example illustrates a general rule that should always be kept in mind when displaying data graphically: *The purpose of a graph is to convey a quick visual impression.* Figure 6 accomplishes this by exposing the presence of outliers and skewness. However, it would be inappropriate to expect the reader to determine individual IgE measurements from such a graph, as that information could be obtained more conveniently from a table.

Previously we showed how two large data sets may be compared by use of frequency polygons. With smaller data sets, individual points may be plotted in a scatter diagram. For example, IgE values in males and females are compared in Fig. 7A.

When data are strongly skewed, as the data on Fig. 7A are, the display sometimes can be made more convenient by a suitable transformation, such as taking logarithms of the original measurements (Fig. 7B). The same transformation may be accomplished simply by plotting the original values on semilogarithmic paper.

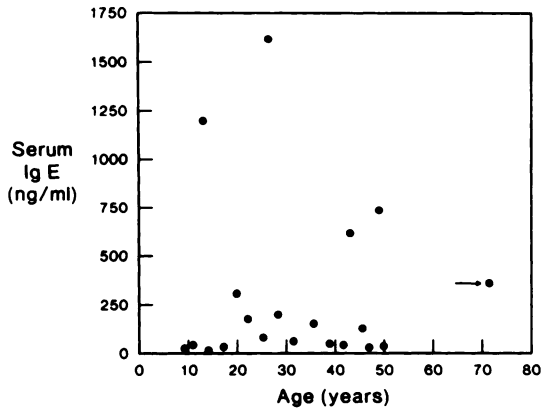

FIG. 6. Scatter diagram showing data (fictitious) relating IgE value with age. (From O'Brien PC, Shampo MA. Statistics for Clinicians. 3. Graphic Displays—Scatter Diagrams. *Mayo Clin Proc* 56: 196–197, 1981.)

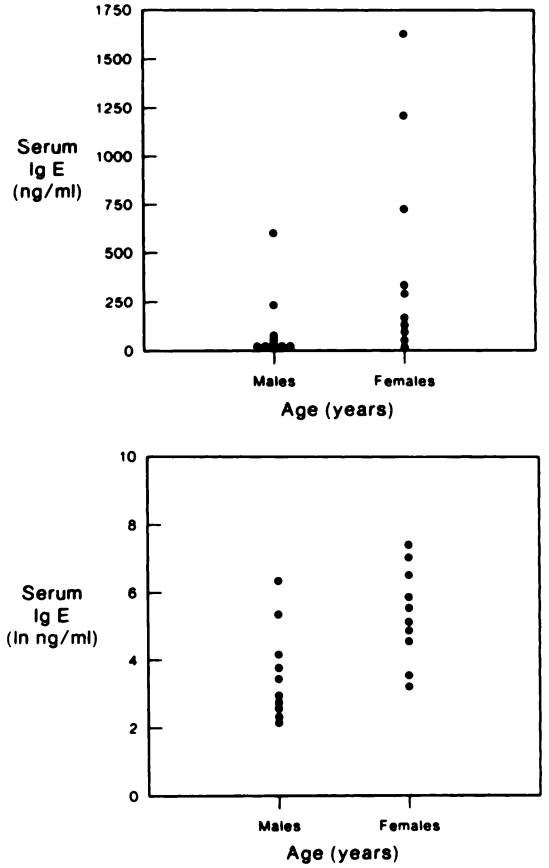

FIG. 7. IgE values by sex. A, original measurements. B, logarithms of measurements. (From O'Brien PC, Shampo MA. Statistics for clinicians. 3. Graphic displays—Scatter diagrams. *Mayo Clin Proc* 56:196–197, 1981.)

Although the logarithmic transformation probably is the kind most commonly used, it is by no means the only one to be considered. Another transformation that is useful (especially when logarithms overcorrect, producing skewness in the opposite direction) is taking the square root of the variable.

If the transformation is successful in eliminating skewness, conceivably one could compute descriptive statistics (means and standard deviations) from the transformed data. Although this may be useful in some applications, it usually produces less satisfying results than would be obtained by choosing a more appropriate descriptive statistic that preserves the original unit of measurement. Generally, transformations are more useful in inferential than in descriptive statistics.

ACKNOWLEDGMENTS

Although we assume all responsibility for the content of this series of papers, much of the organizational framework is patterned after a course taught at the Mayo Clinic for many years by Dr. Lila Elveback. We gratefully acknowledge her contribution, including many helpful conversations relating to these papers. Also we wish to thank Dr. Guy Whitehead, who contributed much to the writing and editing of the series. This series is based in large part on the series "Statistics for Clinicians," *Mayo Clinic Proceedings* 56:Jan-Dec, 1981.

REFERENCE

1. RIGGS BL, WAHNER WH, DUNN WL, et al: Differential changes in bone mineral density of the appendicular and axial skeleton with aging: Relationship to spinal osteoporosis. *J Clin Invest* 67:328-335, 1981