When an industrial hygiene engineer is called upon to make many similar environmental surveys, it is often advisable to furnish him with an appropriate specific form to fill out at the time of sampling. One such form for air contaminants, reproduced in Figure 7, is punched for safe carrying and easy removal back at the laboratory.

In all cases the importance of drawing up correct special forms to be used by the investigator in the field can hardly be overemphasized. Upon the form itself may depend whether or not all needed information will be obtained, or whether a return trip may be required in important instances. Figure 8 shows a good form for recording sound-level values—determinations which are destined to become an increasingly important part of industrial hygiene.

Detailed and very inclusive plant survey forms employed by the United States Public Health Service, along with recommendations for their use, were given by Bloomfield[4] in 1945.

IV. Statistical Analysis of Data

A. FIRST INSPECTION

To the industrial hygienist who wishes to draw all the conclusions justified by the data he has collected—and none that are not justified—a certain working knowledge of the fundamentals of statistics is essential.[5]

The first and simplest calculation he will make is a total of the number of his items. Second, he will determine, by inspection, the item with the lowest value and the highest one. This gives him the *range* of his values. The logical question next occurring in regard to these data is: "Are they spread out more or less evenly throughout the range, or is there some tendency to group around a central value?" If the latter is true, "What is that central value?"

B. AVERAGES

There are a number of statistics that *measure the central tendency* of a *distribution,* which is the name given to an arrangement of numerical data. It is desirable to have some familiarity with five of them that are most commonly used and one or more of which will be the most valuable typical value in analyzing any given related data.

The *arithmetic mean,* which is commonly called the *average,* or simply *the mean,* is the best known and most often used. It should always be determined, for two reasons. First, nearly everyone is familiar with it and will understand its

[4] J. J. Bloomfield, *Am. J. Pub. Health,* **35,** 559 (1945).

[5] This discussion aims to present fundamental definitions and techniques, and to point out pitfalls of statistics, in a simple, usable way, particularly for persons not mathematically trained, who nevertheless find themselves with data to be used. For more complete coverage of statistics, standard texts should be consulted.

meaning. Second, it lends itself to mathematical calculations to determine other *parameters,* variables which characterize the series of data. The mean is determined by adding together all the values and dividing by the number of values.

The second measure of central tendency in order of usage is the *median.* This value may give a better picture of the center of the data than the mean does and should always be considered. While it cannot be mathematically manipulated to determine other aspects of the series of data, it is deservedly appearing more often in the literature. The *median* is the middle value in an odd number of items arranged in order of magnitude, or the average of the two central values of an even number of items. Obviously, it is not affected by the range in the values nor by a few unusually high or low values. For this latter reason, particularly, it is very often more characteristic of the entire group than is the mean, which can be pulled either up or down by a few abnormally high or low values.

The third measure of central tendency that should be considered is the *mode.* The mode is less well known but simple to determine and often quite useful in an evaluation of data. The mode is the numerical value of most frequent occurrence. It may be identical with either the mean or the median and it may not. If it is not, then it should be reported as giving a useful picture of the arrangement of the data. There may be enough items on the opposite side of the median or mean to balance the modal value, but they may be spread out so that no one of them is as characteristic a value (occurs as often) as the mode. In the conclusions to be drawn, the action to follow, evaluation of any particular set of data, in such a case, knowledge of the existence and magnitude of the mode can prove valuable.

The fourth value of the central tendency of a group of data with which the industrial hygienist should have at least some familiarity is the *geometric mean.* He may not wish to calculate this value for his own data but he is sure to meet it in reading technical literature. The geometric mean is the nth root of the product of n items. This means that the logarithm of the geometric mean is equal to the average of the logarithms of the items. The geometric mean should be used as the measure of central tendency when the constant *rate* of change is desired, that is, if the quantities are constant percentages (which will plot as a straight line on logarithmic graph paper). In Figure 9, page 203, the geometric mean is the 50 per cent size as plotted on logarithmic probability paper.

A fifth type of average value is the *quadratic mean.* While the quadratic mean is not so often used as a measure of the central tendency of the original quantities themselves, it is the basis for the determination of a whole series of exceedingly valuable statistics known as *standard deviations.* The quadratic mean is the square root of the arithmetic mean of the squares of the items. The steps for computing it are (*1*) each item is squared; (*2*) all the squares are added; (*3*) the sum is divided by the number of items; and (*4*) the square root of the above mean of the squares is determined (cf. "The R.M.S. Sound Pressure," page 627).