NBS Special Publication 747

Statistical Concepts in Metrology – With a Postscript on Statistical Graphics

Harry H. Ku

Statistical Engineering Division
Center for Computing and Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Gaithersburg, MD 20899

August 1988

U.S. Department of Commerce
C. William Verity, Secretary

National Bureau of Standards
Ernest Ambler, Director

Library of Congress
Catalog Card Number: 88-600569
National Bureau of Standards
Special Publication 747
Natl. Bur. Stand. (U.S.),
Spec. Publ. 747,
48 pages (Aug. 1988)
CODEN: XNBSAV

U.S. Government Printing Office
Washington: 1988

For sale by the Superintendent
of Documents,
U.S. Government Printing Office,
Washington, DC 20402

Contents

List of Figures

List of Tables

Statistical Concepts in Metrology—With a Postscript on Statistical Graphics

Harry H. Ku

Statistical Engineering Division, National Bureau of Standards, Gaithersburg, MD 20899

"Statistical Concepts in Metrology" was originally written as Chapter 2 for the Handbook of Industrial Metrology published by the American Society of Tool and Manufacturing Engineers, 1967. It was reprinted as one of 40 papers in NBS Special Publication 300, Volume I, Precision Measurement and Calibration; Statistical Concepts and Procedures, 1969. Since then this chapter has been used as basic text in statistics in Bureau-sponsored courses and seminars, including those for Electricity, Electronics, and Analytical Chemistry.

While concepts and techniques introduced in the original chapter remain valid and appropriate, some additions on recent development of graphical methods for the treatment of data would be useful. Graphical methods can be used effectively to "explore" information in data sets prior to the application of classical statistical procedures. For this reason additional sections on statistical graphics are added as a postscript.

Key words: graphics; measurement; metrology; plots; statistics; uncertainty.

STATISTICAL CONCEPTS OF A MEASUREMENT PROCESS

Arithmetic Numbers and Measurement Numbers

In metrological work, digital numbers are used for different purposes and consequently these numbers have different interpretations. It is therefore important to differentiate the two types of numbers which will be encountered.

Arithmetic numbers are exact numbers. 3, $\sqrt{2}$, $\frac{1}{3}$, e, or π are all exact numbers by definition, although in expressing some of these numbers in digital form, approximation may have to be used. Thus, π may be written as 3.14 or 3.1416, depending on our judgment of which is the proper one to use from the combined point of view of accuracy and convenience. By the usual rules of rounding, the approximations do not differ from the exact values by more than ± 0.5 units of the last recorded digit. The accuracy of the result can always be extended if necessary.

Measurement numbers, on the other hand, are not approximations to exact numbers, but numbers obtained by operation under approximately the same conditions. For example, three measurements on the diameter of a steel shaft with a micrometer may yield the following results:

No.	*Diameter in cm*	*General notation*
1	0.396	x_1
2	0.392	x_2
3	0.401	x_3
	Sum 1.189	$\sum_{i=1}^{n} x_i$
	Average 0.3963	$\bar{x} = \frac{1}{n} \sum_{1}^{n} x_i$
	Range 0.009	$R = x_{max} - x_{min}$

There is no rounding off here. The last digit in the measured value depends on the instrument used and our ability to read it. If we had used a coarser instrument, we might have obtained 0.4, 0.4, and 0.4; if a finer instrument, we might have been able to record to the fifth digit after the decimal point. In all cases, however, the last digit given certainly does not imply that the measured value differs from the diameter D by less than ± 0.5 unit of the last digit.

Thus we see that measurement numbers differ by their very nature from arithmetic numbers. In fact, the phrase "significant figures" has little meaning in the manipulation of numbers resulting from measurements. Reflection on the simple example above will help to convince one of this fact.

Computation and Reporting of Results. By experience, the metrologist can usually select an instrument to give him results adequate for his needs, as illustrated in the example above. Unfortunately, in the process of computation, both arithmetic numbers and measurement numbers are present, and frequently confusion reigns over the number of digits to be kept in successive arithmetic operations.

No general rule can be given for all types of arithmetic operations. If the instrument is well-chosen, severe rounding would result in loss of information. One suggestion, therefore, is to treat all measurement numbers as exact numbers in the operations and to round off the final result only. Another recommended procedure is to carry two or three extra figures throughout the computation, and then to round off the final reported value to an appropriate number of digits.

The "appropriate" number of digits to be retained in the final result depends on the "uncertainties" attached to this reported value. The term "uncertainty" will be treated later under "Precision and Accuracy"; our only concern here is the number of digits in the expression for uncertainty.

A recommended rule is that the uncertainty should be stated to no more than two significant figures, and the reported value itself should be stated to the last place affected by the qualification given by the uncertainty statement. An example is:

> "The apparent mass correction for the nominal 10 g weight is +0.0420 mg with an overall uncertainty of ±0.0087 mg using three standard deviations as a limit to the effect of random errors of measurement, the magnitude of systematic errors from known sources being negligible."

The sentence form is preferred since then the burden is on the reporter to specify exactly the meaning of the term uncertainty, and to spell out its components. Abbreviated forms such as $a \pm b$, where a is the reported value and b a measure of uncertainty in some vague sense, should always be avoided.

Properties of Measurement Numbers

The study of the properties of measurement numbers, or the Theory of Errors, formally began with Thomas Simpson more than two hundred years ago, and attained its full development in the hands of Laplace and Gauss. In the next subsections some of the important properties of measurement numbers will be discussed and summarized, thus providing a basis for the statistical treatment and analysis of these numbers in the following major section.

The Limiting Mean. As shown in the micrometer example above, the results of *repeated measurements of a single physical quantity under essentially the same conditions* yield a set of measurement numbers. Each member of this set is an estimate of the quantity being measured, and has equal claims on its value. By convention, the numerical values of these n measurements are denoted by $x_1, x_2, \ldots, x_n$, the arithmetic mean by $\bar{x}$, and the range by R, i.e., the difference between the largest value and the smallest value obtained in the n measurements.

If the results of measurements are to make any sense for the purpose at hand, we must require these numbers, though different, to behave as a group in a certain predictable manner. Experience has shown that this is indeed the case under the conditions stated in italics above. In fact, let us adopt as the postulate of measurement a statement due to N. Ernest Dorsey (reference 2)*:

> "The mean of a family of measurements—of a number of measurements for a given quantity carried out by the same apparatus, procedure, and observer—approaches a definite value as the number of measurements is indefinitely increased. Otherwise, they could not properly be called measurements of a given quantity. In the theory of errors, this limiting mean is frequently called the 'true' value, although it bears no necessary relation to the true quaesitum, to the actual value of the quantity that the observer desires to measure. This has often confused the unwary. Let us call it the limiting mean."

Thus, according to this postulate, there exists a limiting mean m to which $\bar{x}$ approaches as the number of measurements increases indefinitely, or, in symbols $\bar{x} \longrightarrow m$ as $n \longrightarrow \infty$. Furthermore, if the true value is τ, there is usually a difference between m and τ, or $\Delta = m - \tau$, where Δ is defined as the bias or systematic error of the measurements.

*References are listed at the end of this chapter.

In practice, however, we will run into difficulties. The value of m cannot be obtained since one cannot make an infinite number of measurements. Even for a large number of measurements, the conditions will not remain constant, since changes occur from hour to hour, and from day to day. The value of τ is unknown and usually unknowable, hence also the bias. Nevertheless, this seemingly simple postulate does provide a sound foundation to build on toward a mathematical model, from which estimates can be made and inference drawn, as will be seen later on.

Range, Variance, and Standard Deviation. The range of n measurements, on the other hand, does not enjoy this desirable property of the arithmetic mean. With one more measurement, the range may increase but cannot decrease. Since only the largest and the smallest numbers enter into its calculation, obviously the additional information provided by the measurements in between is lost. It will be desirablê to look for another measure of the dispersion (spread, or scattering) of our measurements which will utilize each measurement made with equal weight, and which will approach a definite number as the number of measurements is indefinitely increased.

A number of such measures can be constructed; the most frequently used are the variance and the standard deviation. The choice of the variance as the measure of dispersion is based upon its mathematical convenience and maneuverability Variance is defined as the value approached by the average of the sum of squares of the deviations of individual measurements from the limiting mean as the number of measurements is indefinitely increased, or in symbols:

$$\frac{1}{n} \Sigma (x_i - m)^2 \longrightarrow \sigma^2 = \text{variance, as } n \longrightarrow \infty$$

The positive square root of the variance, σ, is called the standard deviation (of a single measurement); the standard deviation is of the same dimensionality as the limiting mean.

There are other measures of dispersion, such as average deviation and probable error. The relationships between these measures and the standard deviation can be found in reference 1.

Population and the Frequency Curve. We shall call the limiting mean m the location parameter and the standard deviation σ the scale parameter of the population of measurement numbers generated by a particular measurement process. By population is meant the conceptually infinite number of measurements that can be generated. The two numbers m and σ describe this population of measurements to a large extent, and specify it completely in one important special case.

Our model of a measurement process consists then of a defined population of measurement numbers with a limiting mean m and a standard deviation σ. The result of a single measurement X* can take randomly any of the values belonging to this population. The probability that a particular measurement yields a value of X which is less than or equal to x' is the proportion of the population that is less than or equal to x', in symbols

$$P\{X \leq x'\} = \text{proportion of population less than or equal to } x'$$

*Convention is followed in using the capital X to represent the value that might be produced by employing the measurement process to obtain a measurement (i.e., a random variable), and the lower case x to represent a particular value of X observed.

Similar statements can be made for the probability that X will be greater than or equal to x'', or for X between x' and x'' as follows: $P\{X \geq x''\}$, or $P\{x' \leq X \leq x''\}$.

For a measurement process that yields numbers on a continuous scale, the distribution of values of X for the population can be represented by a smooth curve, for example, curve C in Fig. 2-1. C is called a frequency curve. The area between C and the abscissa bounded by any two values (x_1 and x_2) is the proportion of the population that takes values between the two values, or the probability that X will assume values between x_1 and x_2. For example, the probability that $X \leq x'$, can be represented by the shaded area to the left of x'; the total area between the frequency curve and the abscissa being one by definition.

Note that the shape of C is not determined by m and σ alone. Any curve C' enclosing an area of unity with the abscissa defines the distribution of a particular population. Two examples, the uniform distribution and the log-normal distribution are given in Figs. 2-2A and 2-2B. These and other distributions are useful in describing certain populations.

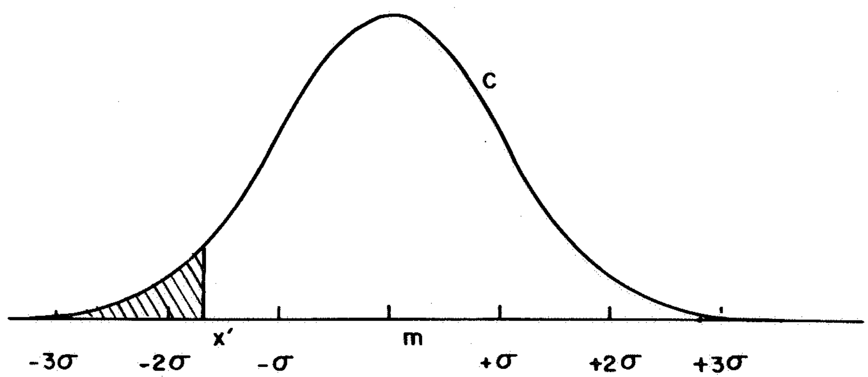

Fig. 2-1. A symmetrical distribution.

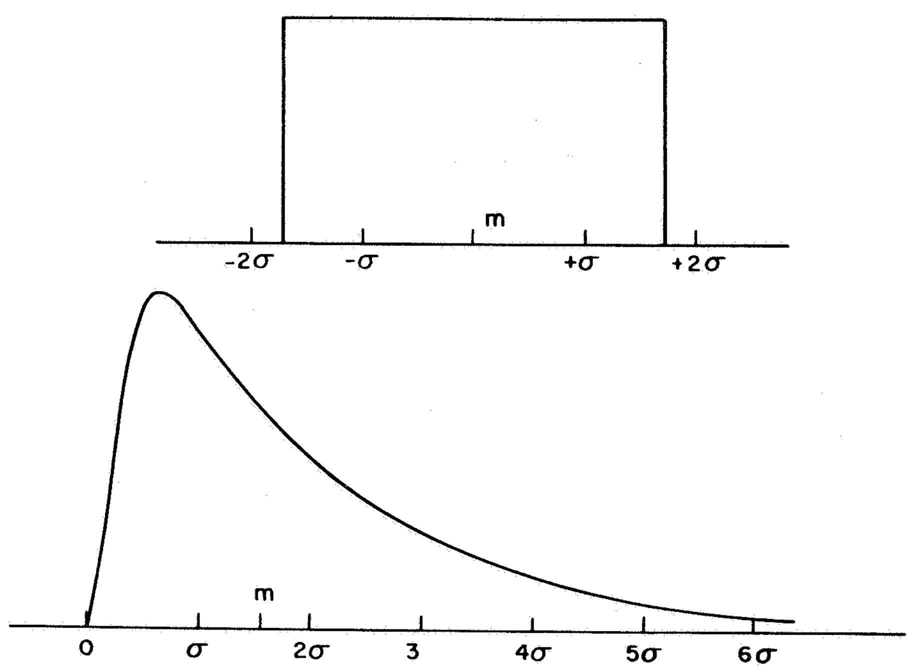

Fig. 2-2. (A) The uniform distribution (B) The log-normal distribution.