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 desirable 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:

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

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.

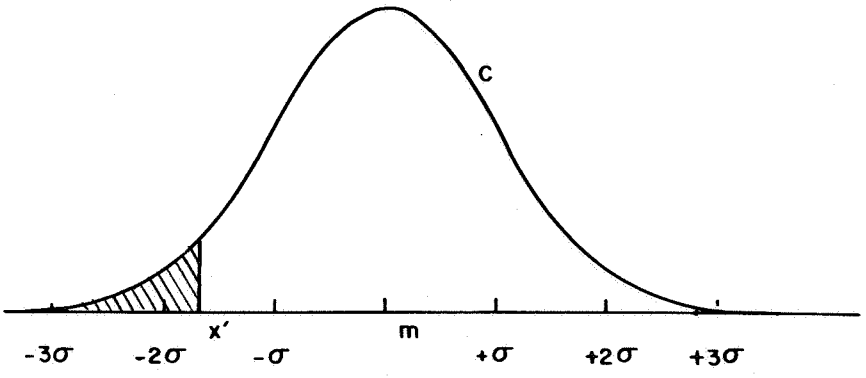

Fig. 2-1. A symmetrical distribution.

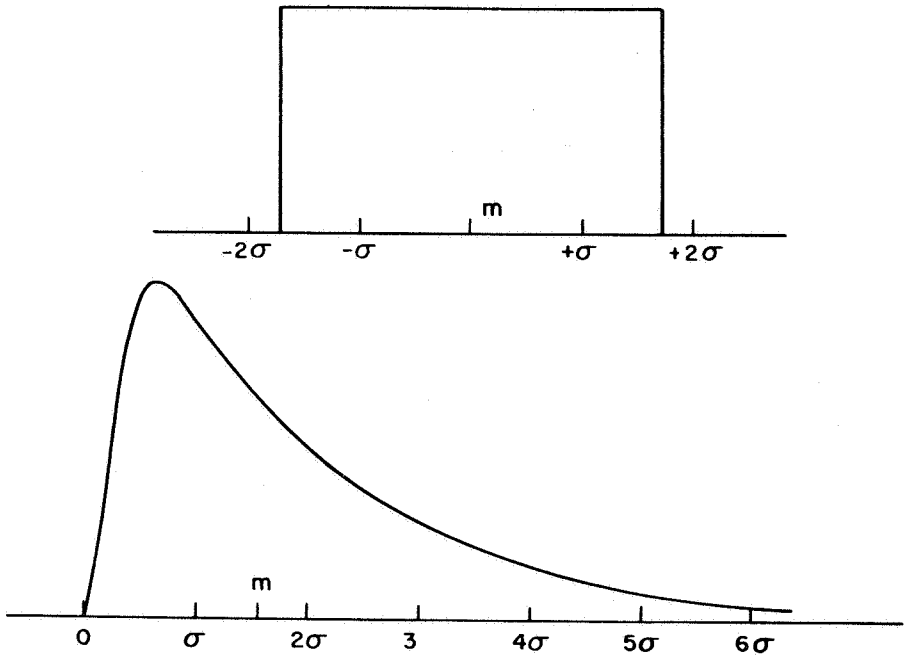

Fig. 2-2. (A) The uniform distribution (B) The log-normal distribution.

The Normal Distribution. For data generated by a measurement process, the following properties are usually observed:

1. The results spread roughly symmetrically about a central value.
2. Small deviations from this central value are more frequently found than large deviations.

A measurement process having these two properties would generate a frequency curve similar to that shown in Fig. 2-1 which is symmetrical and bunched together about m. The study of a particular theoretical representation of a frequency curve of this type leads to the celebrated bell-shaped normal curve (Gauss error curve.). Measurements having such a normal frequency curve are said to be normally distributed, or distributed in accordance with the normal law of error.

The normal curve can be represented exactly by the mathematical expression

$$y = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-1/2[(x-m)^2/\sigma^2]} \tag{2-0}$$

where y is the ordinate and x the abscissa and $e \doteq 2.71828$ is the base of natural logarithms.

Some of the important features of the normal curve are:

1. It is symmetrical about m.
2. The area under the curve is one, as required.
3. If σ is used as unit on the abscissa, then the area under the curve between constant multiples of σ can be computed from tabulated values of the normal distribution. In particular, areas under the curve for some useful intervals between $m - k\sigma$ and $m + k\sigma$ are given in Table 2-1. Thus about two-thirds of the area lies within one σ of m, more than 95 percent within 2σ of m, and less than 0.3 percent beyond 3σ from m.

Table 2-1. Area under normal curve between $m - k\sigma$ and $m + k\sigma$

k:	0.6745	1.00	1.96	2.00	2.58	3.00
Percent area under curve (approx.):	50.0	68.3	95.0	95.5	99.0	99.7

4. From Eq. (2-0), it is evident that the frequency curve is completely determined by the two parameters m and σ.

The normal distribution has been studied intensively during the past century. Consequently, if the measurements follow a normal distribution, we can say a great deal about the measurement process. The question remains: How do we know that this is so from the limited number of repeated measurements on hand?

The answer is that we don't! However, in most instances the metrologist may be willing

1. to assume that the measurement process generates numbers that follow a normal distribution approximately, and act as if this were so,
2. to rely on the so-called Central Limit Theorem, one version of which is the following*: "If a population has a finite variance σ^2 and mean m, then the distribution of the sample mean (of n independent

*From Chapter 7, *Introduction to the Theory of Statistics*, by A. M. Mood, McGraw-Hill Book Company, New York, 1950.

measurements) approaches the normal distribution with variance σ^2/n and mean m as the sample size n increases." This remarkable and powerful theorem is indeed tailored for measurement processes. First, every measurement process must by definition have a finite mean and variance. Second, the sample mean $\bar{x}$ is the quantity of interest which, according to the theorem, will be approximately normally distributed for large sample sizes. Third, the measure of dispersion, i.e., the standard deviation of the sample mean, is reduced by a factor of $1/\sqrt{n}$! This last statement is true in general for all measurement processes in which the measurements are "independent" and for all n. It is therefore not a consequence of the Central Limit Theorem. The theorem guarantees, however, that the distribution of sample means of *independent* measurements will be *approximately* normal with the specified limiting mean and standard deviation $\sigma/\sqrt{n}$ *for large n.*

In fact, for a measurement process with a frequency curve that is symmetrical about the mean, and with small deviations from the mean as compared to the magnitude of the quantity measured, the normal approximation to the distribution of $\bar{x}$ becomes very good even for n as small as 3 or 4. Figure 2-3 shows the uniform and normal distribution having the same mean and standard deviation. The peaked curve is actually two curves, representing the distribution of arithmetic means of four independent measurements from the respective distributions. These curves are indistinguishable to this scale.

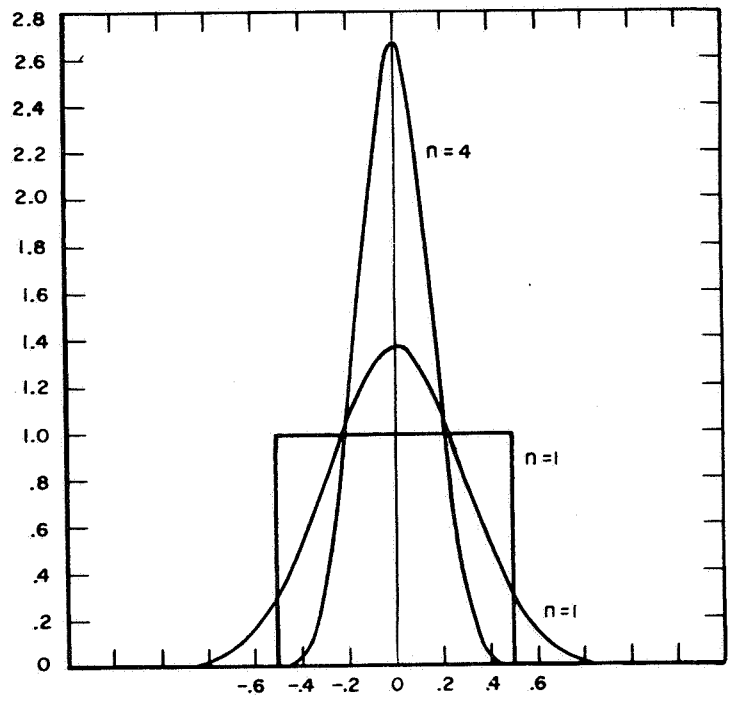

Fig. 2-3. Uniform and normal distribution of individual measurements having the same mean and standard deviation, and the corresponding distribution(s) of arithmetic means of four independent measurements.

A formal definition of the concept of "independence" is out of the scope here. Intuitively, we may say that n normally distributed measurements are independent if these measurements are not correlated or associated in any

way. Thus, a sequence of measurements showing a trend or pattern are not independent measurements.

There are many ways by which dependence or correlation creeps into a set of measurement data; several of the common causes are the following:

1. Measurements are correlated through a factor that has not been considered, or has been considered to be of no appreciable effect on the results.
2. A standard correction constant has been used for a factor, e.g., temperature, but the constant may overcorrect or undercorrect for particular samples.
3. Measurements are correlated through time of the day, between days, weeks, or seasons.
4. Measurements are correlated through rejection of valid data, when the rejection is based on the size of the number in relation to others of the group.

The traditional way of plotting the data in the sequence they are taken, or in some rational grouping, is perhaps still the most effective way of detecting trends or correlation.

Estimates of Population Characteristics. In the above section it is shown that the limiting mean m and the variance σ^2 completely specify a measurement process that follows the normal distribution. In practice, m and σ^2 are not known and cannot be computed from a finite number of measurements. This leads to the use of the sample mean $\bar{x}$ as an estimate of the limiting mean m and s^2, the square of the computed standard deviation of the sample, as an estimate of the variance. The standard deviation of the average of n measurements, $\sigma/\sqrt{n}$, is sometimes referred to as the standard error of the mean, and is estimated by $s/\sqrt{n}$.

We note that the making of n independent measurements is equivalent to drawing a sample of size n at random from the population of measurements. Two concepts are of importance here:

1. The measurement process is established and under control, meaning that the limiting mean and the standard deviation do possess definite values which will not change over a reasonable period of time.
2. The measurements are randomly drawn from this population, implying that the values are of equal weights, and there is no prejudice in the method of selection. Suppose out of three measurements the one which is far apart from the other two is rejected, then the result will not be a random sample.

For a random sample we can say that $\bar{x}$ is an unbiased estimate of m, and s^2 is an unbiased estimate of σ^2, i.e., the limiting mean of $\bar{x}$ is equal to m and of s^2 to σ^2, where

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

and

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right]$$

In addition, we define

$$s = \sqrt{s^2} = \text{computed standard deviation}$$

Examples of numerical calculations of $\bar{x}$ and s^2 and s are shown in Tables 2-5 and 2-6.

Interpretation and Computation of Confidence Interval and Limits

By making k sets of n measurements each, we can compute and arrange k, $\bar{x}$'s, and s's in a tabular form as follows:

Set	*Sample mean*	*Sample standard deviation*
1	$\bar{x}_1$	s_1
2	$\bar{x}_2$	s_2
.	.	.
.	.	.
.	.	.
j	$\bar{x}_j$	s_j
.	.	.
.	.	.
.	.	.
k	$\bar{x}_k$	s_k

In the array of $\bar{x}$'s, no two will be likely to have exactly the same value. From the Central Limit Theorem it can be deduced that the $\bar{x}$'s will be approximately normally distributed with standard deviation $\sigma/\sqrt{n}$. The frequency curve of $\bar{x}$ will be centered about the limiting mean m and will have the scale factor $\sigma/\sqrt{n}$. In other words, $\bar{x} - m$ will be centered about zero, and the quantity

$$z = \frac{\bar{x} - m}{\sigma/\sqrt{n}}$$

has the properties of a single observation from the "standardized" normal distribution which has a mean of zero and a standard deviation of one.

From tabulated values of the standardized normal distribution it is known that 95 percent of z values will be bounded between -1.96 and $+1.96$. Hence the statement

$$-1.96 < \frac{\bar{x} - m}{\sigma/\sqrt{n}} < +1.96$$

or its equivalent,

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < m < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

will be correct 95 percent of the time in the long run. The interval $\bar{x} - 1.96(\sigma/\sqrt{n})$ to $\bar{x} + 1.96(\sigma/\sqrt{n})$ is called a *confidence interval* for m. The probability that the confidence interval will cover the limiting mean, 0.95 in this case, is called the confidence level or confidence coefficient. The values of the end points of a confidence interval are called confidence limits. It is to be borne in mind that $\bar{x}$ will fluctuate from set to set, and the interval calculated for a particular $\bar{x}_j$ may or may not cover m.

In the above discussion we have selected a two-sided interval symmetrical about $\bar{x}$. For such intervals the confidence coefficient is usually denoted by $1 - \alpha$, where $\alpha/2$ is the percent of the area under the frequency curve of z that is cut off from each tail.

In most cases, σ is not known and an estimate of σ is computed from the same set of measurements we use to calculate $\bar{x}$. Nevertheless, let us form a quantity similar to z, which is

$$t = \frac{\bar{x} - m}{s/\sqrt{n}}$$

and if we know the distribution of t, we could make the same type of statement as before. In fact the distribution of t is known for the case of normally distributed measurements.

The distribution of t was obtained mathematically by William S. Gosset under the pen name of "Student," hence the distribution of t is called the Student's distribution. In the expression for t, both $\bar{x}$ and s fluctuate from set to set of measurements. Intuitively we will expect the value of t to be larger than that of z for a statement with the same probability of being correct. This is indeed the case. The values of t are listed in Table 2-2.

Table 2-2. A brief table of values of t

Degrees of freedom	Confidence Level: $1 - \alpha$			
ν	0.500	0.900	0.950	0.990
1	1.000	6.314	12.706	63.657
2	.816	2.920	4.303	9.925
3	.765	2.353	3.182	5.841
4	.741	2.132	2.776	4.604
5	.727	2.015	2.571	4.032
6	.718	1.943	2.447	3.707
7	.711	1.895	2.365	3.499
10	.700	1.812	2.228	3.169
15	.691	1.753	2.131	2.947
20	.687	1.725	2.086	2.845
30	.683	1.697	2.042	2.750
60	.679	1.671	2.000	2.660
∞	.674	1.645	1.960	2.576

*Adapted from *Biometrika Tables for Statisticians*, Vol. I, edited by E. S. Pearson and H. O. Hartley, The University Press, Cambridge, 1958.

To find a value for t, we need to know the "degrees of freedom" (ν) associated with the computed standard deviation s. Since $\bar{x}$ is calculated from the same n numbers and has a fixed value, the nth value of x_i is completely determined by $\bar{x}$ and the other $(n - 1)x$ values. Hence the degrees of freedom here are $n - 1$.

Having the table for the distribution of t, and using the same reasoning as before, we can make the statement that

$$\bar{x} - t\frac{s}{\sqrt{n}} < m < \bar{x} + t\frac{s}{\sqrt{n}}$$

and our statement will be correct $100\,(1 - \alpha)$ percent of the time in the long run. The value of t depends on the degrees of freedom ν and the probability level. From the table, we get for a confidence level of 0.95, the following lower and upper confidence limits:

ν	$L_l = \bar{x} - t(s/\sqrt{n})$	$L_u = \bar{x} + t(s/\sqrt{n})$
1	$\bar{x} - 12.706(s/\sqrt{n})$	$\bar{x} + 12.706(s/\sqrt{n})$
2	$\bar{x} - 4.303(s/\sqrt{n})$	$\bar{x} + 4.303(s/\sqrt{n})$
3	$\bar{x} - 3.182(s/\sqrt{n})$	$\bar{x} + 3.182(s/\sqrt{n})$

The value of t for $\nu = \infty$ is 1.96, the same as for the case of known σ. Notice that very little can be said about m with two measurements. However, for n larger than 2, the interval predicted to contain m narrows down steadily, due to both the smaller value of t and the divisor $\sqrt{n}$.

It is probably worthwhile to emphasize again that each particular confidence interval computed as a result of n measurements will either include m or fail to include m. The probability statement refers to the fact that if we make a long series of sets of n measurements, and if we compute a confidence interval for m from each set by the prescribed method, we would expect 95 percent of such intervals to include m.

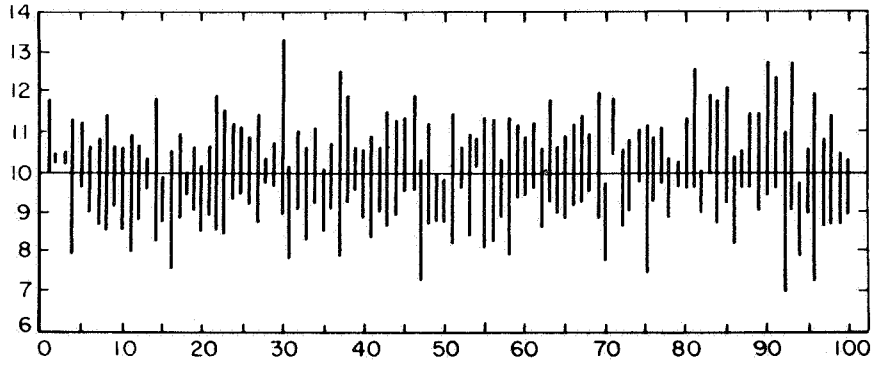

Fig. 2-4. Computed 90% confidence intervals for 100 samples of size 4 drawn at random from a normal population with $m = 10$, $\sigma = 1$.

Figure 2-4 shows the 90 percent confidence intervals ($P = 0.90$) computed from 100 samples of $n = 4$ from a normal population with $m = 10$, and $\sigma = 1$. Three interesting features are to be noted:

1. The number of intervals that include m actually turns out to be 90, the expected number.
2. The surprising variation of the sizes of these intervals.
3. The closeness of the mid-points of these intervals to the line for the mean does not seem to be related to the spread. In samples No. 2 and No. 3, the four values must have been very close together, but both of these intervals failed to include the line for the mean.

From the widths of computed confidence intervals, one may get an intuitive feeling whether the number of measurements n is reasonable and sufficient for the purpose on hand. It is true that, even for small n, the confidence intervals will cover the limiting mean with the specified probability, yet the limits may be so far apart as to be of no practical significance. For detecting a specified magnitude of interest, e.g., the difference between two means, the approximate number of measurements required can be solved by equating the half-width of the confidence interval to this difference and solving for n, using σ when known, or using s by trial and error if σ is not known. Tables of sample sizes required for certain prescribed conditions are given in reference 4.

Precision and Accuracy

Index of Precision. Since σ is a measure of the spread of the frequency curve about the limiting mean, σ may be defined as an index of precision. Thus a measurement process with a standard deviation σ_1 is said to be more precise than another with a standard deviation σ_2 if σ_1 is smaller than σ_2. (In fact, σ is really a measure of imprecision since the imprecision is directly proportional to σ.)

Consider the means of sets of n independent measurements as a new derived measurement process. The standard deviation of the new process is $\sigma/\sqrt{n}$. It is therefore possible to derive from a less precise measurement process a new process which has a standard deviation equal to that of a more precise process. This is accomplished by making more measurements.

Suppose $m_1 = m_2$, but $\sigma_1 = 2\sigma_2$. Then for a derived process to have $\sigma_1' = \sigma_2$, we need

$$\sigma_1' = \frac{\sigma_1}{\sqrt{n}} = \frac{2\sigma_2}{\sqrt{4}}$$

or we need to use the average of four measurements as a single measurement. Thus for a required degree of precision, the number of measurements, n_1 and n_2, needed for measurement processes I and II is proportional to the squares of their respective standard deviations (variances), or in symbols

$$\frac{n_1}{n_2} = \frac{\sigma_1^2}{\sigma_2^2}$$

If σ is not known, and the best estimate we have of σ is a computed standard deviation s based on n measurements, then s could be used as an estimate of the index of precision. The value of s, however, may vary considerably from sample to sample in the case of a small number of measurements as was shown in Fig. 2-4, where the lengths of the intervals are constant multiples of s computed from the samples. The number n or the degrees of freedom ν must be considered along with s in indicating how reliable an estimate s is of σ. In what follows, whenever the terms standard deviation about the limiting mean (σ), or standard error of the mean ($\sigma_{\bar{x}}$), are used, the respective estimates s and $s/\sqrt{n}$ may be substituted, by taking into consideration the above reservation.

In metrology or calibration work, the precision of the reported value is an integral part of the result. In fact, precision is the main criterion by which the quality of the work is judged. Hence, the laboratory reporting the value must be prepared to give evidence of the precision claimed. Obviously an estimate of the standard deviation of the measurement process based only on a small number of measurements cannot be considered as convincing evidence. By the use of the control chart method for standard deviation and by the calibration of one's own standard at frequent intervals, as subsequently described, the laboratory may eventually claim that the standard deviation is in fact known and the measurement process is stable, with readily available evidence to support these claims.

Interpretation of Precision. Since a measurement process generates numbers as the results of repeated measurements of a single physical quantity under essentially the same conditions, the method and procedure in obtaining these numbers must be specified in detail. However, no amount of detail would cover all the contingencies that may arise, or cover all the factors that may affect the results of measurement. Thus a single operator in a single day with a single instrument may generate a process with a precision index measured by σ. Many operators measuring the same quantity over a period of time with a number of instruments will yield a precision index measured by σ'. Logically σ' must be larger than σ, and in practice it is usually considerably larger. Consequently, modifiers of the words "precision" are recommended by ASTM* to qualify in an unambiguous manner what

*"Use of the Terms Precision and Accuracy as Applied to the Measurement of a Property of a Material," ASTM Designation, E177-61T, 1961.