8. STATISTICAL TECHNIQUES

Statistics are used in metrology to summarize experimental data, to provide the basis for assessing its quality, and to provide a basis for making probabilistic decisions in its use. The essential basic statistical information for describing a simple data set is:

The mean of the set,	$\bar{x}$
The standard deviation of the set,	s
The number of individuals in the set,	n

If the set is a random sample of the population from which it was derived, if the measurement process is in statistical control, and if all of the observations are independent of one another, then s is an estimate of the population standard deviation, σ, and $\bar{x}$ is an unbiased estimate of the mean, μ.

The population consists of all possible measurements that could have been made under the test conditions for a stable test sample. In this regard, the metrologist must be aware that any changes in the measurement system (known or unknown) could possibly result in significant changes in its operational characteristics, and, hence the values of the mean and standard deviation. Whenever there is doubt, statistical tests should be made to determine the significance of any apparent differences before statistics are combined.

The following discussion reviews some useful statistical techniques for interpreting measurement data. In presenting this information, it is assumed that the reader is already familiar with basic statistical concepts. For a detailed discussion of the following techniques and others not presented here, it is recommended that the reader consult NBS Handbook 91 - Experimental Statistics, by Mary G. Natrella [19]. That handbook also contains comprehensive statistical tables from which the tables contained in Chapter 9 of this publication were taken.

8.1 Estimation of Standard Deviation from a Series of Measurements on a Given Object

Given n measurements x_1, x_2, x_3, ……,x_n

Mean, $\bar{x} = \dfrac{(x_1 + x_2 + x_3 + ... + x_n)}{n}$

Standard deviation estimate, $s = \sqrt{\dfrac{\sum(x_n - \bar{x})^2}{n-1}}$

The estimate, s, is based on n-1 degrees of freedom

8.2 Estimation of Standard Deviation from the Differences of k Sets of Duplicate Measurements

Given k differences of duplicate measurements, d_1, d_2, d_3, …, d_k, a useful formula for estimating the standard deviation is:

$s_d = \frac{\sum d_i^2}{2k}$ where s_d is based on k degrees of freedom.

Note that $d_1 = \bar{x}_i' - \bar{x}_i''$, for example.

The values d_1, d_2 etc., may be differences of duplicate measurements of the same sample (or object) at various times, or they may be the differences of duplicate measurements of several similar samples (or objects).

8.3 Estimation of Standard Deviation from the Average Range of Several Sets of Measurements

The range, R, is defined as the difference between the largest and smallest values in a set of measurements.

Given R_1, R_2, R_3, …, R_k

Mean, $\bar{R} = \frac{(R_1 + R_2 + R_3 + ... + R_k)}{k}$

Standard deviation can be estimated by the formula, $s_R = \frac{\bar{R}}{d_2^*}$

The value of d_2^* will depend on the number of sets of measurements used to calculate s_R, and on the number of measurements in each set, i.e., 2 for duplicates, 3 for triplicates, etc. Consult a table such as Table 9.1 for the appropriate value of d_2^* to use. The effective number of degrees of freedom for s_R is in the table.

8.4 Pooling Estimates of Standard Deviations

Estimates of the standard deviation obtained at several times may be combined (pooled) to obtain a better estimate based upon more degrees of freedom. The following equation may be used for this purpose:

$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 + ... + (n_k-1)s_k^2}{(n_1-1)+(n_2-1)+(n_3-1)+...+(n_k-1)}}$ where

s_p will be based on $(n_1-1)+(n_2-1)+(n_3-1)+...+(n_k-1)$ degrees of freedom.

8.5 "Within" and "Between" Standard Deviation

Estimation of the within-series, s_w, and between-series, s_b standard deviation, (also referred to as short-term and long-term standard deviations in the applications described here) is an important way to characterize a measurement process. The former provides guidance as to how many repetitions of a measurement are required to obtain a result on a single occasion with a given precision, while the latter is a better estimate of the precision of replication (reproducibility) of a result on various occasions and is a more realistic evaluation of measurement variability.

To estimate these standard deviations, sets of measurements may be made on several occasions. To simplify the calculations, each set should consist of the same number of measurements. For most measurements, it is recommended that duplicate measurements be made on at least 12 separate occasions when estimating s_w and s_b.

Given k sets of duplicate measurements made on k occasions the following table and calculation can be made.

Table 8.1

Occasion	Measured Values		Range	Mean
1	x_1'	x_1''	R_1	$\frac{(x_1' + x_1'')}{2} = \bar{x}_1$
2	x_2'	x_2''	R_2	$\frac{(x_2' + x_2'')}{2} = \bar{x}_2$
3	x_3'	x_3''	R_3	$\frac{(x_3' + x_3'')}{2} = \bar{x}_3$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
k	x_k'	x_k''	R_k	$\frac{(x_k' + x_k'')}{2} = \bar{x}_k$

1. Calculate $\bar{R} = \frac{(R_1 + R_2 + R_3 + ... + R_k)}{k}$

2. Calculate $s_w = s_R - \frac{\bar{R}}{d_2^*}$

Note: One may use the procedure of 8.2 to calculate s_w if preferred.

3. Calculate s_x as follows:

$$\bar{\bar{x}} = \frac{(\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + ... + \bar{x}_k)}{k}$$

$$s_{\bar{x}} = \sqrt{\frac{\sum(\bar{x}_k - \bar{\bar{x}})^2}{k-1}}$$

4. Calculate s_b (for the case of duplicate measurements)

$$s_b = \sqrt{s_x^2 - \frac{s_w^2}{2}}$$

Note that s_b is an estimate of the long term component of the standard deviation of a single measurement. The long term standard deviation of the mean of *n* measurements taken at a single occasion is estimated by:

$$s_b(\bar{x}_n) = \sqrt{s_b^2 - \frac{s_w^2}{n}}$$

Important note: Do not use this approach for handling within-series and between-time standard deviations with the Mass Code. Separate formulas are available for that application.

8.6 Confidence Interval for the Mean

The estimation of the confidence interval for the mean of *n* measurements is one of the most frequently used statistical calculations. The formula used will depend on whether the population standard deviation, σ, is known or whether it is estimated on the basis of measurements of a sample(s) of the population.

Using Population Standard Deviation, σ

Strictly speaking, σ, is never known for a measurement process. However, the formula for use in such a case is:

$$\bar{x} \pm \frac{z\sigma}{\sqrt{n}}$$

Variable	Description
$\bar{x}$	sample mean
σ	known standard deviation
n	number of measurements of sample
z	standard normal variate, depending on the confidence level desired

For 95 % confidence *z* = 1.960; for 99.7 % confidence *z* = 3.0.
For other confidence levels, see Table 9.2

Using Estimate of Standard Deviation, s

In the usual situation, s is known, based on υ degrees of freedom and the formula for use is:

$$\bar{x} \pm \frac{ts}{\sqrt{n}}$$

Variable	Description
$\bar{x}$	sample mean
s	estimate of standard deviation
n	number of measurements on which the mean is based
t	Student's t value, based on the confidence level desired and the υ degrees of freedom associated with s (see Table 9.3).

Note that $t \rightarrow z$ as $n \rightarrow \infty$. For many practical purposes, the standard deviation may be considered as known when estimated by at least 30 degrees of freedom.

8.7 Confidence Interval for σ

The standard deviation, σ, is ordinarily not known but is, instead, an estimated value based on a limited number of measurements, using procedures such as have been described above. Such estimates may be pooled, as appropriate, to obtain better estimates. In any case, the uncertainty of the estimated value of the standard deviation may be of interest and can be expressed in the form of a confidence interval, computed as indicated below.

The interval is asymmetrical because the standard deviation is ordinarily underestimated when small numbers of measurements are involved due to the fact that large deviations occur infrequently in a limited measurement process. Indeed, it is the general experience of metrologists that a few measurements appear to be more precise than they really are.

The basic information required to compute the interval is an estimate of the standard deviation, s, and the number of degrees of freedom on which the estimate is based. The relationships to use are:

Lower limit $B_L s$

Upper limit $B_U s$

Interval $B_L s$ to $B_U s$

The values for B_L and B_U depend upon the confidence level and degrees of freedom associated with s. Values for use in calculating the confidence level are given in Table 9.7. A more extensive table (Table A-20) is available in NBS Handbook 91 [19].

8.8 Statistical Tolerance Intervals

Statistical tolerance intervals define the bounds within which a percentage of the population is expected to lie with a given level of confidence. For example, one may wish to define the limits within which 95 % of measurements would be expected to lie with a 95 % confidence of being correct. The interval is symmetrical and is computed using the expression

$$\bar{x} \pm ks$$

where k depends on three things

Variable	Description
p	the proportion or percentage of the individual measurements to be included
γ	the confidence coefficient to be associated with the interval
n	the number of measurements on which the estimate, s, is based

Table 9.6 may be used to obtain values for k for frequently desired values of γ and p. A more extensive table is Table A-6 found in NBS Handbook 91 [19].

8.9 Comparing Estimates of a Standard Deviation (F Test)

The F-test may be used to decide whether there is sufficient reason to believe that two estimates of a standard deviation differ significantly. The ratio of the variances (square of the standard deviation) is calculated and compared with tabulated values. Unless the computed ratio is larger than the tabulated value, there is no reason to believe that the observed standard deviations are significantly different.

The F ratio is calculated using the equation

$$F = \frac{s_L^2}{s_S^2}$$

where s_L is the numerically larger value and s_S is the smaller value of the two estimates under consideration.

The critical value of F depends on the significance level chosen for the decision (test) and the number of degrees of freedom associated with s_L and s_S, respectively.

Table 9.4 contains critical values for F at the 95 % level of confidence. The tabulated values of F are not expected to be exceeded with 95 % confidence on the basis of chance alone. As an example, if both the numerator and the denominator values for s were each based on 9 degrees of freedom, an F value no larger than 4.03 is expected with 95 % confidence, due to the uncertainties of the s values, themselves. Table A-5 of NBS Handbook 91 [19] contains values for F for other confidence levels.

The F-test is useful for comparing the precision of methods, equipment, laboratories, or metrologists, for example. An inspection of Table 9.4 shows that when either of the values of s is based on a small number of degrees of freedom, the F value is large. Consequently, the significance of decisions based on small changes in precision can be supported statistically only by a relatively large number of measurements. If such changes are suspected, but the data requirement is difficult to meet, the decision may need to be made on the basis of information about the measurement process itself.

The F-test is also useful for deciding whether estimates of the standard deviation made at various times differ significantly. Such questions need to be answered when deciding on whether to revise control limits of a control chart, for example.

8.10 Comparing a Set of Measurements with a Given Value

The question may arise as to whether a measured value agrees or significantly disagrees with a stated value for the measured object. The evaluation can be based on whether or not the confidence interval for the measured value encompasses the stated value. The confidence interval is calculated using the expression

$$\bar{x} \pm \frac{ts}{\sqrt{n}}$$

as previously described in Section 8.6. In using this expression, n represents the number of measurements used to calculate the mean, $\bar{x}$, and t depends on the degrees of freedom, υ, associated with s and the confidence level needed when making the decision. Note that one may use historical data for estimating s, such as a control chart for example, in which case υ will represent the degrees of freedom associated with establishment of the control limits and may be considerably larger than $n-1$.

8.11 Comparing Two Sets of Measurements with Regard to Their Means

This discussion is concerned with deciding whether the means of two measured values, $_A$ and $_B$, are in agreement. The data sets used for this purpose may consist of the following:

$$\begin{array}{ll} \bar{x}_A & \bar{x}_B \\ s_A & s_B \\ n_A & n_B \end{array}$$

The first question to be resolved is whether s_A and s_B can be considered to be different estimates of the same standard deviation or whether they do, indeed, differ. An F test may be used for this purpose. However, it will be recalled that this is not sensitive to small real differences, so the decision may need to be based on physical considerations, such as the known stability of the measurement process, for example.

Case I

Confirming (or assuming) that s_A and s_B are not significantly different, they are pooled, as already described (but repeated here for convenience) and used to calculate a confidence interval for the difference of the means. If this is larger than the observed difference, there is no reason to believe that the means differ. The steps to follow when making the calculation described above are:

Step 1. Choose α , the level of significance for the test

Step 2. Calculate the pooled estimate of the standard deviation, s_p

$$s_p = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A - 1) + (n_B - 1)}}$$

s_p will be estimated with $n_A + n_B - 2$ degrees of freedom

Step 3. Calculate the respective variances of the means

$$v_A = \frac{s_A^2}{n_A} \text{ and } v_B = \frac{s_B^2}{n_B}$$

Step 4. Calculate the uncertainty of $\left|X_A - X_B\right| = \Delta$

$$U_\Delta = t\sqrt{(V_A + V_B)}$$

using a value for t based on $\frac{\alpha}{2}$ and $\upsilon = n_A + n_B - 2$.

Step 5. Compare U_Δ with Δ

If $U_\Delta \geq \Delta$, there is no reason to believe that Δ is significant at the level of confidence chosen.

Case II

Confirming (or assuming) that s_A and s_B are significantly different, their individual values are used to calculate U_Δ as outlined below.

Step 1. Choose α, the level of significance for the test.

Step 2. Calculate the respective variances of the means.

$$v_A = \frac{s_A^2}{n_A} \text{ and } v_B = \frac{s_B^2}{n_B}$$

Step 3. Calculate the uncertainty of $\left|X_A - X_B\right| = \Delta$

$$U_\Delta = t^* \sqrt{(V_A + V_B)}$$

using a value for t^* based on $\frac{\alpha}{2}$ and f, the effective number of degrees of freedom calculated as described in Step 4.

Step 4. Calculate f, the effective number of degrees of freedom as follows:

$$f = \left(\frac{(V_A + V_B)^2}{\frac{V_A^2}{n_A + 1} + \frac{V_B^2}{n_B + 1}} \right) - 2.$$

Step 5. Compare U_Δ with Δ.
If $U_\Delta \geq \Delta$, there is no reason to believe that Δ is significant at the level of confidence chosen.

8.12 Use of Random Numbers

Conducting operations in random sequences can avert problems of bias that might stem from a particular order of operations. For example, in the measurement of a series of items, it might be difficult to determine whether systematic trends in the measured values were due to differences in the items or to measurement system drift unless the items were measured in random order.

Use of tables of random numbers is a convenient means for randomizing measurement operations. The operations, test objects, and other matters requiring randomization may be assigned serial numbers. The order of selection is then determined by use of a random number table, as described below. When the number of operations or test items is less than 100, a table such as Table 9.11, reproduced from NBS Handbook 91 [19], may be used conveniently. One may start from any arbitrarily selected position in the table and proceed from it in any pre-determined arbitrary manner. If the first number encountered is not that of one of any item, ignore it and proceed until a valid match is encountered. This becomes the first item in the sequence. Continuing in the same manner, items are selected in the sequence in which their serial numbers are encountered ignoring the repetition of previously identified items. The procedure is continued until all items have been randomly selected.

As an example, select 10 specimens (numbered 01 to 10) in random order. Start from a randomly selected place, say column 2, row 5 of Table 9.11. Proceed from this point along the table as one would read a book. The starting number is 14, which is not usable. The first useful number encountered is 08, the next 03, and so on. Using the procedure described above, the following random order was found:

Specimen No.									
08	03	09	05	06	02	07	10	04	01
Order									
1	2	3	4	5	6	7	8	9	10