Manual of Petroleum Measurement Standards Chapter 13—Statistical Aspects of Measuring and Sampling

Section 1—Statistical Concepts and Procedures in Measurement

FIRST EDITION, JUNE 1985
REAFFIRMED, MARCH 1990
Reaffirmed 3/2002

American Petroleum Institute
1220 L Street, Northwest
Washington, D.C. 20005

Manual of Petroleum Measurement Standards
Chapter 13—Statistical Aspects of Measuring and Sampling

Section 1—Statistical Concepts and Procedures in Measurement

Measurement Coordination Department

FIRST EDITION, JUNE 1985

American
Petroleum
Institute

Nothing contained in any API publication is to be construed as granting any right, by implication or otherwise, for the manufacture, sale, or use in connection with any method, apparatus, or product covered by letters patent nor as insuring anyone against liability for infringement of letters patent.

API publications may be used by anyone desiring to do so. Every effort has been made by the Institute to assure the accuracy and reliability of the data contained in them; however, the Institute makes no representation, warranty, or guarantee in connection with API publications. The Institute hereby expressly disclaims any liability or responsibility for loss or damage resulting from their use; for the violation of any federal, state, or municipal regulation with which an API publication may conflict; or for the infringement of any patent resulting from the use of an API publication.

FOREWORD

This publication covers statistical concepts and procedures used in bulk oil measurement.

Suggested revisions are invited and should be submitted to the director of the Measurement Coordination Department, American Petroleum Institute, 1220 L Street, N.W., Washington, D.C. 20005.

CONTENTS

SECTION 1–STATISTICAL CONCEPTS AND PROCEDURES IN MEASUREMENT

PAGE

13.1.0 Introduction . . . 1
13.1.1 Scope . . . 1
13.1.2 Definitions . . . 1
13.1.3 Nomenclature . . . 2
13.1.4 Statistical Control . . . 2
13.1.5 Measurements . . . 2
13.1.5.1 True Value . . . 3
13.1.5.2 Uncertainty of Measurement . . . 3
13.1.5.3 Confidence Level . . . 3
13.1.5.4 Reporting Results . . . 3
13.1.6 Types of Errors . . . 3
13.1.6.1 Spurious Errors . . . 3
13.1.6.2 Systematic Errors . . . 3
13.1.6.3 Random Errors . . . 4
13.1.7 Accuracy and Precision . . . 4
13.1.7.1 Repeatability . . . 4
13.1.7.2 Reproducibility . . . 4
13.1.7.3 Application of Precision to a Single Measurement . . . 5
13.1.8 Statistical Procedures . . . 5
13.1.8.1 Statistical Procedure for a Single Set of Data . . . 5
13.1.8.2 Statistical Procedure for Two or More Sets of Data . . . 9
13.1.8.3 Rounding Statistical Estimates . . . 10
13.1.8.4 Example . . . 11

APPENDIX A–NORMAL (GAUSSIAN) DISTRIBUTION . . . 15
APPENDIX B–DIXON'S TEST FOR OUTLIERS . . . 17

Tables
1–Range Conversion Factor . . . 7
2–Distribution Values for 95 Percent Probability (Double-Sided) . . . 8
3–Derived Statistics for Example . . . 13
4–Symbols for Example . . . 13
5–Volume Measurement Statistics for Example . . . 13
B-1–Dixon's Test for Outliers . . . 17

Figures
A-1–Frequency Histogram . . . 15
A-2–Bell-Shaped Curve . . . 16

Chapter 13—Statistical Aspects of Measuring and Sampling

SECTION 1—STATISTICAL CONCEPTS AND PROCEDURES IN MEASUREMENT

13.1.0 Introduction

The nature of physical measurements makes it impossible to measure a physical variable without error. Absolute accuracy is only achievable when it is possible to count the objects or events; even then, when large numbers are involved, it may be necessary to approximate. With the best equipment and directions, the potential for errors in fluid volume measurements involving large amounts of material is large.

Minimizing errors, estimating the remaining errors, and keeping all parties informed of errors is increasingly important to the petroleum industry. Equally important is an understanding of the size and significance of errors. Providing estimates of errors and statements concerning errors in a standard form can help avoid disputes and dispel delusions of accuracy in statements of quantity.

Chapter 13 of the *Manual of Petroleum Measurement Standards* is designed to help those who make measurements of bulk oil quantities improve the value of their result statement by making proper estimates of the uncertainty or probable error involved in measurements. During the development of Chapter 13.1, reference was made to Part XIV, Section 1 (Tentative) of the *Petroleum Measurement Manual* published by the Institute of Petroleum, London, England.

13.1.1 Scope

This chapter covers the basic concepts involved in estimating errors by statistical techniques and ensuring that results are quoted in the most meaningful way. The statistical procedures that should be followed in estimating a true quantity from one or more measurements and in deriving the range of uncertainty of the results are discussed. Sources of error are examined and examples are provided showing how a statement of the overall uncertainty in completed measurements is derived.

The subsequent sections (in preparation at the time this section was published) of Chapter 13 will deal with the application of the concepts discussed in Section 1 to various methods for bulk oil measurement widely used in the petroleum industry. Chapter 13.1 is a reference document explaining theory and the application of statistical procedures whereas subsequent sections will provide statistical equations and typical examples for various types of measurement.

13.1.2 Definitions

The following terms are used throughout Chapter 13.

Accuracy is the ability to indicate values closely approximating the true value of the measured variable.

Bias is any influence on a result that produces an incorrect approximation of the true value of the variable being measured. Bias is the result of a predictable systematic error.

Confidence interval or range of uncertainty, C, is the range or interval within which the true value is expected to lie with a stated degree of confidence.

Confidence level is the degree of confidence that may be placed on an estimated range of uncertainty.

Degrees of freedom is the number of independent results used in estimating the standard deviation.

Direct measurement is a measurement that produces a final result directly from the scale on an instrument.

Error is the difference between true and observed values.

Indirect measurement is a measurement that produces a final result by calculation using results from one or more direct measurements.

Mean, $\bar{x}$, is the average of two or more observed values.

Measurement is a procedure for determining a value for a physical variable.

Normal (Gaussian) distribution (see Appendix A).

The *observed value* is the result obtained from a measurement.

An *outlier* is a result that differs considerably from the main body of results in a set.

Parameters are the values that characterize and summarize the essential features of measurements.

Precision is the degree to which data within a set cluster together.

A *random error* is an error that varies in an unpredictable manner when a large number of measurements of the same variable are made under effectively identical conditions.

Range, w, is the region between the limits within which a quantity is measured.

Repeatability, r, is a measure of the agreement between the results of successive measurements of the same variable carried out by the same method, with the same instrument, at the same location, and within a short period of time.

Reproducibility is a measure of the agreement between the results of measurements of the same variable where individual measurements are carried out by the same methods, with the same type of instruments, but by different observers, at different locations, and after a long period of time.

A *result* is the observed value of a variable determined by a single measurement.

A *spurious error* is a gross error in procedure (for example, human errors or machine malfunctions).

Standard deviation, s, is the root mean square deviation of the observed value from the average.

Standard normal deviate (see Appendix A).

Student's t is a statistical function that varies in magnitude with degrees of freedom.

A *systematic error, e,* is one that, in the course of a number of measurements made under the same conditions, on material having the same true value of a variable, either remains constant in absolute value and sign or varies in a predictable manner. Systematic errors result in a bias.

True value, X, is the correct value of a variable.

Variance, V or v, is the measure of the dispersion or scatter of the values of the random variable about the mean μ.

13.1.3 Nomenclature

The following algebraic symbols are used throughout Chapter 13.

- A True limit of range of uncertainty for random errors.
- a Estimate of A.
- B True limit of range of uncertainty for systematic errors.
- b Estimate of B.
- C True total limit of range of uncertainty.
- c Estimate of C.
- D Conversion factor (used to derive s from w).
- e Estimate of systematic error.
- n Number of repeated measurements.
- m Number of quantities incorporated in a final indirect quantity measurement.
- p Number of independent sources of systematic error.
- P, Q Constants.
- r Estimate of repeatability.
- S True value of standard deviation.
- s Estimate of standard deviation.
- t Value of Student's t distribution.
- V True variance, S^2.
- v Estimate of variance, s^2.
- w Range of a set of data.
- X True value of a variable.
- $\bar{x}$ Observed mean value of a set of data.
- x Observed value of a variable.
- $\bar{y}$ Observed mean value corrected for bias.
- y Observed value of a variable corrected for bias.
- μ Mean of Gaussian normal distribution.
- σ Standard deviation of a Gaussian normal distribution.
- ϕ Degrees of freedom.

13.1.4 Statistical Control

Proper use of statistical techniques requires that the measurement process be in a state of statistical control. Unless this is achieved, any statement concerning the estimate of the true value of the quantity being measured, and the statistical uncertainty associated with it, is not strictly valid and may even be meaningless. A measurement process that is under statistical control will, if measurements on the same quantity are repeated by the same method and under essentially the same conditions, show stability of the mean value and regular scatter of individual results (see also 13.1.7).

Repeatability and reproducibility, when properly established, can be used to monitor statistical control on a routine basis (see 13.1.7.1 and 13.1.7.2).

Strict statistical control is usually very difficult to ensure. An important step in establishing any measurement procedure is to decide which variables should be used to monitor statistical control and to establish target values required to maintain an appropriate degree of consistency. Some essential elements in statistical control are listed here.

1. The entire measurement procedure and instructions must be clearly defined and closely followed.
2. Independent procedures for checking and maintaining equipment must be available.
3. Means for detecting and eliminating equipment malfunctions and human mistakes (leading to spurious errors) should be incorporated (see 13.1.6.1).

These features of the measurement procedure must be adhered to at all times. Furthermore, control charts and other records of equipment performance, maintenance, and calibration checks must be used as an integral part of statistical control procedures.

13.1.5 Measurements

13.1.5.1 TRUE VALUE

One primary assumption is made; that is, an exact or true value exists for any variable, valid for the conditions that exist at the moment when the result is deter-

mined. Generally, the true value X cannot be determined, but a valid estimate $\bar{x}$ can be obtained by rigorous application of the appropriate method of measurement using the specified instruments. By statistical analysis of the various errors involved, it is possible to use observed values to obtain an estimate of the true value and to quantify the reliability of that estimate. In any set of measurements, the best estimate of X will be the mean $\bar{x}$ after rejecting outliers and correcting for systematic errors.

13.1.5.2 UNCERTAINTY OF MEASUREMENT

The usefulness of a result is greatly increased when it is accompanied by a statement of its reliability. The statistical calculations provided in this chapter give a range or interval within which the true value of the variable can be expected to lie with a stated degree of confidence. The statistical term for such an interval is the *confidence interval* (also referred to as the *range of uncertainty* of the measurement). The limits of a confidence interval about an estimate $\bar{x}$ are expressed as $\bar{x} \pm C(\bar{x})$; the magnitude of $\pm C(\bar{x})$ depends on the random variability of the measurements, unknown systematic errors, and the confidence level. As an example, consider the following statement: $10° \pm 1°C$. In this statement, the estimate $\bar{x}$ is 10° and the confidence interval is $\pm 1°$.

13.1.5.3 CONFIDENCE LEVEL

Setting absolute limits to a range of uncertainty is rarely possible. It is more practical to give an indication of the degree of confidence that may be placed on an estimated range of uncertainty. This degree of confidence, or confidence level, indicates the probability that the range quoted will include the true value of the quantity being measured. The most common statistical practice is to use the 95 percent confidence level. This level implies that there is a 95 percent probability (19 chances in 20) that the true value will lie within the stated range. The 95 percent level is recommended for all commercial applications in petroleum measurement and will be used throughout this chapter. In certain limited circumstances, a different degree of confidence may be required.

NOTE: Strictly, a confidence level or confidence interval can only be used to account for Gaussian random errors or errors that may be so treated. Systematic errors must be accounted for before the confidence level and interval are applied, and substantial contributions to the total uncertainty should be separately recorded.

13.1.5.4 REPORTING RESULTS

All results should be reported so that the estimate of the true value, and the limits within which the true value is expected to lie with a given level of confidence, can be seen at a glance. Results should be written as follows: $\bar{y} \pm C(\bar{y})$ 95, n (95 percent confidence level, n measurements) from which the following relevant information can be obtained:

1. $\bar{y}$ is a mean value of n measurements, is corrected for all known systematic errors, and is the estimate of the true value.
2. There is 95 percent probability that the true value lies between $\bar{y} - C(\bar{y})$ and $\bar{y} + C(\bar{y})$.
3. There is 95 percent probability that any further single measurement will lie within $\bar{y} \pm = C(\bar{y})/n^{0.5}$.

Expanding on the example, assume that the following temperature measurements were taken for a delivery batch of crude oil: 10, 8, 11, 9, and 12°C; and the confidence interval was determined to be $\pm$ 2°C. Then $\bar{y} = 10$, $n = 5$ and the result statement would be: $10° \pm 2°C$ (95 percent confidence level, 5 measurements).

13.1.6 Types of Errors

The difference between the observed value of a variable and its true value includes all errors associated with the person actually taking and recording the results, instrument errors, procedure errors, and errors resulting from sampling procedures or changes in conditions during the period of measurement. There are three basic types of errors that must be considered: spurious errors, systematic errors, and random errors.

13.1.6.1 SPURIOUS ERRORS

Spurious errors are gross errors, such as misapplication of method, incorrect reading or recording, and instrument malfunction. These errors cannot be incorporated into any statistical analysis and the results must be discarded.

There are statistical methods of testing for outliers (see Appendix B), but these methods should only be applied if there is good reason to believe that spurious errors exist. Data should not be discarded lightly, and the observer should record what information has been discarded and state the reasons.

13.1.6.2 SYSTEMATIC ERRORS

A systematic error is one that, in the course of a number of measurements made under the same conditions on the same variable, remains constant or varies predictably. Thus, systematic errors cause bias in the results. The bias can be positive or negative, leading to over- or underestimation of the true value of the variable being measured. In many liquid measurement applications, systematic errors may make a larger contribution

to the overall uncertainty of a result than random errors (see 13.1.6.3). Systematic errors must be identified and either eliminated or compensated for before interpreting overall results statistically (see note in 13.1.5.3).

Ideally, bias could be considered as constant for all measurements made by the same operator and equipment. Unfortunately, assessment of bias is complicated by the fact that some contributions to bias do vary with time. For example, knowledge and control of test conditions may be inadequate or an instrument may wear progressively. Such factors will probably not change significantly during the course of one set of measurements but could change both in magnitude and sign over a longer period.

Assessment of systematic errors by experimental means is difficult, especially when variation with time is involved. Errors introduced by the observer or by changes in operating conditions are probably easiest to identify, but any experimental evaluation of systematic error may involve a complete change of equipment, which is often not feasible. The alternative to experimentation is to make a subjective assessment on the basis of experience and knowledge of the instruments involved.

In any event, if the conditions of measurement are unchanged, increasing the number of measurements will not reduce the effects of systematic error.

All conceivable sources of error must be identified, examined methodically, and assessed quantitatively to establish whether they make a significant contribution to bias.

13.1.6.3 RANDOM ERRORS

Random errors are caused by small independent influences that prevent a repeated measurement from giving an identical result, although the true value of the variable involved remains the same. Results that contain only random errors are amenable to statistical analysis. Random errors are assumed to follow a normal (Gaussian) distribution, described in Appendix A. Provided that all systematic errors can be accounted for, the corrected mean value $\bar{y}$ and the range of uncertainty $\pm C(\bar{y})$ can be calculated. Increasing the number of measurements reduces the value of $C(\bar{y})$ and hence improves the reliability of the final estimate $\bar{y}$.

13.1.7 Accuracy and Precision

A set of measurements subject to the smallest systematic errors will be expected to have their mean closest to the true value, and this is said to give the most accurate set. It is also evident that the set of measurements subject to the smallest random errors will be expected to be clustered closest together, and so form the most precise set. Within these rather narrow statistical definitions, precise measurements are not necessarily accurate, since they could cluster about a point that is not the true value. Conversely, it is possible to have a set of measurements that are accurate taken as a group although widely scattered and of doubtful reliability when taken singly.

This distinction is important since any statement concerning reliability must account for both systematic errors and random errors, as statistically defined. In practice, accuracy in measurement cannot exist without precision, so every effort must be made to satisfy both criteria of reliability.

The precision of a method of measurement can be determined quantitatively and is conventionally expressed as repeatability and reproducibility.

13.1.7.1 REPEATABILITY

The repeatability of a method of bulk measurement is a quantitative measure of the random error associated with a single operator at a given location, obtaining successive measurements on the same body of material over a short time interval, with the same measuring devices, and under constant operating conditions. Repeatability is defined as the difference between two such measurements that would be exceeded in the long run in 1 case in 20 in the normal and correct operation of the method of measurement. Repeatability is the range of uncertainty (95 percent confidence level) for the difference between two measurements obtained under the same conditions.

The short time interval between measurements is essential to ensure that external conditions are kept as nearly constant as practicable. The time interval should be of the same order of magnitude as the duration of a single measurement. For example, if a measurement takes 5 minutes to carry out, the interval before a second measurement should not exceed 10 minutes.

13.1.7.2 REPRODUCIBILITY

The reproducibility of a bulk measurement method is a quantitative expression of the random error associated with different operators working in different locations with different instruments, with each operator obtaining single measurements on the same body of material by using the same method and the same types of measuring devices. It is defined as the difference between two such single and independent measurements that would be exceeded in the long run in only 1 case in 20 in the normal and correct operation of the method of measurement. Reproducibility is the range of uncertainty (95 percent

confidence level) for the difference between two measurements obtained under the same conditions.

Good reproducibility indicates that random errors are acceptably small (good repeatability) and that systematic errors other than those inherent in the method are probably also very limited in size and number. Reproducibility conditions as defined, can rarely be met in quantitative bulk oil measurement because the identity of a body of oil is almost invariably lost during its movement from one place to another (the only possible exception being measurement by gage or weighing of a vehicle or ship). However, a close approximation to reproducibility conditions could be achieved for measurements such as gaging a storage tank if two operators each set up their own apparatus for the prescribed method of measurement at the same location.

13.1.7.3 APPLICATION OF PRECISION TO A SINGLE MEASUREMENT

As stated previously, reproducibility of a measurement method, strictly defined, is not a concept that can often be utilized. The application of repeatability is also limited in normal commercial transfer measurement, since the second measurement required to establish a difference between two results is not a practical proposition in everyday work using meters.

In practice it is necessary to conduct a special exercise and obtain repeated determinations of the result with the apparatus that is to be used at a given site and to use these determinations to estimate the random error of a single measurement. This random error is expressed as the range of uncertainty about a single measurement rather than the range for the difference between two measurements as would be the case for repeatability and reproducibility. This estimate of the range of uncertainty is then used for all routine measurements until such time as a complete re-check of the apparatus and method is undertaken (usually at prescribed regular intervals).

The estimated range of uncertainty so obtained would cover errors both from the instruments used and from the calibration system employed. It should also be noted that, without repeated measurements, it is impossible to use the range of uncertainty as a means of monitoring statistical control on a short-term basis, as was the case with repeatability (see 13.1.7.1).

13.1.8 Statistical Procedures

True value and range of uncertainty are two important characteristics describing the measurement of any physical variable. There are characteristics that describe other features of the results, such as random error, standard deviation, and bias, and these are called the parameters of the population of observed variables. Parameters are all assumed to have true values. The procedures described in this section are used to derive estimates of the parameters, known as statistics, from the set of measurements obtained. Parameters will be represented algebraically either by Greek letters (for example, μ and σ) or true values by capital Roman letters, and observed values will be represented by lowercase Roman letters.

In general, the result in question will be a function of one or more intermediate results, each of which could contribute to both the final result estimate and its range of uncertainty. The statistics for each intermediate result should be established first. Intermediate results will be combined to give the statistics that relate to the final result.

13.1.8.1 STATISTICAL PROCEDURE FOR A SINGLE SET OF DATA

Statistics are derived from a single set of n repeated measurements x_i, for $i = 1$ to n. Each measurement will be an estimate of X, the true value of the variable but will be subject to both systematic and random errors (see 13.1.6). All known sources of systematic error should be accounted for before the true value and range of uncertainty are estimated. In the interest of clarity, it is good practice to record the source and magnitude of each error separately.

Random errors are assumed to follow the normal (Gaussian) distribution (see 13.1.6.3 and Appendix A), which is fully determined if its parameters μ (mean) and σ (standard deviation) are determined. These two parameters are estimated from the measurements obtained. The possible sources and magnitudes of the systematic errors to be found in measuring systems are given in detail in 13.1.8.1.1 through 13.1.8.1.7.

13.1.8.1.1 Number of Repeated Measurements Required

There is no fixed value for the optimum number of measurements required to establish a true value and a range of uncertainty. On the one hand, n, the number of repeated measurements, has no bearing on the determination of systematic errors that are present to the same extent in all measurements made under the same operating conditions (see 13.1.7.2). On the other hand, the statistics relating to random errors (for example, mean and standard deviation) are not independent of n, since the larger n becomes, the closer estimates will approach their true values and the smaller will be the range of uncertainty (see 13.1.7.3).

Very often it is only practical to obtain from five to ten measurements in the field. This is perfectly accepta-

ble for the day-to-day estimate of a mean value, but greater reliability is required for a statistic that is to be used as a standard measure. This is the case for repeatability (see 13.1.7.1), which should be estimated from at least 20 and preferably 30 or more repeated measurements. A similar argument applies when estimating the range of uncertainty for single measurements (see 13.1.7.3).

13.1.8.1.2 Outlying Results

Results that are subject to spurious errors (see 13.1.6.1) may differ considerably from the remaining results in the set. These are called outlying results. If a result is suspected to be an outlier but is not easily identifiable, then the set of results should be tested for outliers according to the procedures given in Appendix B. The suspect result should be discarded if the test proves significant. It should be stressed, however, that a good reason is required before a result is rejected, and that reason should be clearly stated.

When the repeatability of the method of measurement has been established, it is possible to make a preliminary check for outliers by the range test illustrated in 13.1.8.1.7. Note that constant systematic errors cannot be detected in an outlier test because they are present to the same extent in all results of the quantity in question (see 13.1.6.2).

13.1.8.1.3 Correcting for Bias

If a constant systematic error e is known to exist, for example, a depth gage is known to give a consistent bias 1 millimeter above the true reading due to faulty calibration, then each of the measurements x_i should be adjusted accordingly. The adjusted results, y_i, will then be the most accurate available (see 13.1.7) and are given by the expression:

$$y_i = x_i - e \quad (1)$$

Note that the systematic error could be level dependent, that is, a constant function (for example, percentage) of the measurement itself:

$$e = f(x) \quad (2)$$

An example of this would be a direct reading meter known from experience to give a consistent bias 1 percent above the true value. In that case, Equation 1 becomes:

$$y_i = x_i - f(x_i) \quad (3)$$

There are times, however, when the systematic error is unknown in magnitude and/or sign, usually due to variations in operating conditions over a long interval (see 13.1.6.2). In that case, the average systematic error $\bar{e}$ should be estimated, taking into account the conditions that affect the measurements at the time. Very often the only way to estimate the average is to calculate the mean range in which the errors could lie. If the errors were estimated to range from e_1 to e_2, the average systematic error would be:

$$\bar{e} = (e_1 + e_2)/2 \quad (4)$$

Individual measurements should then be adjusted by the mean value $\bar{e}$ as in Equation 5:

$$y_i = x_i - \bar{e} \quad (5)$$

Note that when the systematic error takes positive or negative values up to the same maximum ($e_1 = -e_2$) no correction will be made. Note also that unknown systematic errors will contribute to the range of uncertainty for the true value estimate (see 13.1.8.1.6.3).

13.1.8.1.4 Estimating True Value

The results y_i (x_i corrected for bias) are now subject to random errors and unknown systematic errors. As previously stated, the measurements y_i are assumed to follow the normal distribution with mean μ and standard deviation σ. The estimate of the mean that is most likely to be correct, or the "maximum likelihood" estimate of μ, is the average $\bar{y}$ of the set of corrected measurements:

$$\bar{y} = \frac{1}{n}(y_1 + y_2 + \ldots + y_n) = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (6)$$

If only one result is available, the result is the estimate of the true value.

13.1.8.1.5 Estimating Standard Deviation

The standard deviation $\sigma(y)$ describes the random error of a single measurement.

The maximum likelihood estimate $s(y)$ of the standard deviation is calculated from the set of corrected results (y_i) as follows:

$$s(y) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

or

$$s(y) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{n\bar{y}^2}{n-1}} \quad (7)$$

A less complicated but more approximate estimate is:

$$s(y) = \frac{w}{D(n)} \quad (8)$$

Where:

w = the range of the set of measurements (for $n < 12$).
$D(n)$ = a conversion factor (see Table 1).

A further approximation can be made by replacing $D(n)$ by $(n)^{0.5}$. It should be stressed, however, that Equation 8 is approximate since it should theoretically apply to the average range $\overline{w}$ of a number of sets of n measurements. A more reliable estimate would be obtained from the average range of six pairs of results than from the range of a single set of twelve repeated results. For this reason, the equation should only be used as a quick check to monitor statistical control and not for data interpretation (see 13.1.4).

The standard deviation of the average of n repeated results can be calculated as:

$$\sigma(\bar{y}) = \frac{\sigma(y)}{n^{0.5}} \quad (9)$$

In terms of estimates, the standard deviation, or as it is more commonly called, the standard error, of the average becomes:

$$s(\bar{y}) = \frac{s(y)}{n^{0.5}} \quad (10)$$

As the number of measurements increases, the standard error of the average will decrease. Therefore, an average based on a large number of measurements would in this sense be more reliable than one based on a small number of measurements (see 13.1.8.1.1). Furthermore, since the distribution of any average tends toward the normal as n becomes larger, Equation 10 would still hold true if the distribution of individual results deviated from the normal distribution.

13.1.8.1.6 Estimating Range of Uncertainty

For a measurement function, here called G, the limit $C(G)$ of the range of uncertainty (see 13.1.5.2) consists of two parts, the limit $A(G)$ due to random errors and the limit $B(G)$ due to unknown systematic errors (see 13.1.8.1.3). The estimation of A, B, and C depends to a large extent on the nature of G, be it a single measurement or an average, and on the nature of the errors present. (In this section, the expression "limit of the range of uncertainty" will often be referred to in shortened form as "limit of uncertainty" or "uncertainty limit.")

Table 1—Range Conversion Factor

Number of Measurements, n	Conversion Factor, $D(n)$	Number of Measurements, n	Conversion Factor, $D(n)$
2	1.128	8	2.847
3	1.693	9	2.970
4	2.059	10	3.078
5	2.326	11	3.173
6	2.534	12	3.258
7	2.704		

SOURCE: Davies, O. L., *Statistical Methods in Research and Production*, 2nd Edition, Longman, 1984.

13.1.8.1.6.1 Uncertainty Due to Random Errors

The limit $A(y)$ of the range of uncertainty due to random errors about a single measurement y is simply the product of the standard deviation $\sigma(y)$ and the standard normal deviate (see Appendix A). For 95 percent probability, the standard normal deviate has a value of 1.96, that is,

$$A(y) = 1.96\ \sigma(y) \quad (11)$$

In general, the standard deviation will be estimated from Equation 7 as $s(y)$. To take this into account, the limit of random uncertainty calculated from $s(y)$ should be based not on the standard normal deviate, but on a value known as Student's t, which varies in magnitude with the degree of freedom. For the purpose of this document, degrees of freedom may be regarded as the number of independent measurements used in estimating the standard deviation, which for n measurements will be $n - 1$ (1 degree of freedom having been used in calculating the average). The limit of the range of uncertainty for single measurements (see 13.1.7.3), will, in this case, be estimated as:

$$a(y) = (t_{95,\ n-1})\ s(y) \quad (12)$$

Where:

$t_{95,\ n-1}$ = the value of the t-distribution for $(n-1)$ degrees of freedom and for a two-sided probability of 95 percent (two-sided since the range of uncertainty covers both sides of the true quantity estimate).

Values of the t-function are given in Table 2. Once again, by using Equation 10, the limit for an average will become:

$$a(\bar{y}) = (t_{95,\ n-1}) \times s(\bar{y}) = (t_{95,\ n-1}) \times \frac{s(y)}{n^{0.5}} \quad (13)$$

Table 2—*t*-Distribution Values for 95 Percent Probability (Double-Sided)

Degrees of Freedom ϕ	$t_{95,\phi}$	Degrees of Freedom ϕ	$t_{95,\phi}$
1	12.706	18	2.101
2	4.303	19	2.093
3	3.182	20	2.086
4	2.776	21	2.080
5	2.571	22	2.074
6	2.447	23	2.069
7	2.365	24	2.064
8	2.306	25	2.060
9	2.262	26	2.056
10	2.228	27	2.052
11	2.201	28	2.048
12	2.179	29	2.045
13	2.160	30	2.042
14	2.145	40	2.021
15	2.131	60	2.000
16	2.120	120	1.980
17	2.110	--	1.960

SOURCE: Fisher and Yates, *Statistical Tables for Biological, Agricultural, and Medical Research.*

Thus, combining Equations 12 and 13:

$$a(\bar{y}) = \frac{a(y)}{n^{0.5}} \qquad (14)$$

It is worth noting that as n becomes very large, so the t-value approaches the standard normal deviate, and the standard deviation estimate $s(y)$ approaches the true value $\sigma(y)$.

13.1.8.1.6.2 Uncertainty Due to Systematic Errors

Systematic errors can affect results by creating a bias, an uncertainty, or both (see 13.1.6.2). When the errors are known, be they level dependent or not, the bias can be removed according to 13.1.8.1.3 and no uncertainty will exist. On the other hand, when the errors are unknown in sign and/or magnitude and even though bias can be allowed for according to Equation 4, there will still be an uncertainty as to the true value of the variable. Nor is it theoretically possible, due to the very nature of systematic errors, to express the true limit B in terms of the measurements obtained. It is necessary, therefore, to estimate the limit for each source of systematic error by calculating the absolute value by which the corrected results could deviate from their true value with 95 percent confidence. Assuming that systematic errors are uniformly distributed and using the symbols defined in 13.1.8.1.3, this means that:

$$b(y) = 0.95 \left| \frac{(e_1 - e_2)}{2} \right| \qquad (15)$$

Note that $b(y)$ is the limit of a range of uncertainty, and should not be confused with the maximum value (e_1 or e_2) that a systematic error could take. If the error takes positive or negative values up to the same maximum ($e_1 = e_2$) then:

$$b(y) = 0.95 \mid e_1 \mid = 0.95 \mid e_2 \mid \qquad (16)$$

Note also that since systematic errors are present to the same extent for all measurements made under the same conditions (see 13.1.6.2), the limit of the range of uncertainty about an average result $\bar{y}$ will be identical, that is:

$$b(\bar{y}) = b(y) \qquad (17)$$

Although it is difficult to handle systematic errors with theoretical justification, this should not detract from their importance in measuring systems. In many cases, systematic errors create greater uncertainty than random errors, and for this reason, great care should be taken in their identification and estimation.

13.1.8.1.6.3 Combining Random and Systematic Uncertainties

In attempting to allow for systematic uncertainties, difficulties will arise because systematic errors are often variable with time and cannot be identified from a single set of measurements obtained under constant operating conditions (see 13.1.6.2). This is not to say that systematic errors cannot be estimated at all since good estimates can be derived from calibration exercises or from experience with and knowledge of the measuring system involved. A combination of uncertainties is required because it is of great value to state the range in which a true value is expected to lie.

There are two schools of thought on how uncertainty limits should be combined: (1) by simple addition or (2) by a method called root sum square. The latter method is theoretically correct if only random uncertainty limits are to be combined (see 13.1.8.2.2), but gives a narrower range and, therefore, a more optimistic view than the former method. In choosing between the two methods, consideration must be given not only to theoretical implications, but also to the manner in which errors are found to behave in practice.

Consider first a set of measurements of the same quantity subject to p independent sources of systematic error, all of which are unknown but have been estimated according to 13.1.8.1.3. Since the errors affect the same measurement, they tend to cancel each other out, at least to a certain extent. With this in mind, it would be sensible to take the more optimistic view and combine the systematic uncertainty limits by the root sum square method. Assuming that systematic errors follow a uni-

form distribution—that there is an equal probability that the error lies anywhere throughout its full range—there would be a theoretical justification for this choice. As a general rule, the total limit of the range of uncertainty due to systematic errors should be calculated as:

$$b(y) = \sqrt{b_1^2(y) + b_2^2(y) + \ldots + b_p^2(y)} \quad (18)$$

In this equation, the systematic uncertainties have been combined in exactly the same manner as random uncertainties (see 13.1.8.2.2). On a theoretical basis, systematic and random uncertainties should be combined in the same fashion. There is further justification for this approach in practical terms, since systematic and random errors would be expected to average each other out to a certain extent. This leads to the root sum square method for combining systematic and random uncertainties, which, in terms of average measurements, is:

$$c(\bar{y}) = \sqrt{a^2(\bar{y}) + b^2(\bar{y})} \quad (19)$$

Note that $a(\bar{y})$ becomes smaller as n becomes larger (see 13.1.8.1.1), whereas $b(\bar{y})$ will remain unchanged. The total limit $c(\bar{y})$ will approach the limit $b(\bar{y})$ as n increases. This shows the need for care in estimating systematic errors, regardless of the number of repeated measurements obtained (see 13.1.6.2). Note also that by using Equations 14 and 17, the limit of the range of uncertainty for any further single measurement (see 13.1.5.4) becomes:

$$c(y) = \sqrt{na^2(y) + b^2(y)} \quad (20)$$

13.1.8.1.7 Estimating Repeatability

Since repeatability is defined as the range of uncertainty due to random errors for the difference between two measurements (see 13.1.7.1), it can be estimated directly from Equation 12. In this case, the standard deviation relates to the absolute difference between two repeated measurements, y_1 and y_2, and for a normal distribution of errors this is:

$$\sigma(|y_1 - y_2|) = \sqrt{2}\,\sigma(y_1) = \sqrt{2}\,\sigma(y_2) = \sqrt{2}\sigma(y) \quad (21)$$

In terms of estimates this becomes:

$$s(y_1 - y_2) = \sqrt{2}s\,(y) \quad (22)$$

Substituting Equation 22 into Equation 12, the estimate r of repeatability will be given by:

$$r = (t_{95,\ n-1})\,[\sqrt{2}\,s(y)] \quad (23)$$

$t_{95,\ n-1}$ is described in 13.1.8.1.6. This estimate can be compared with a predetermined repeatability value for control purposes. If r were excessively great, it would imply that measurements were subject to unusually large errors. Note that a repeatability estimate that is to be used as a standard measure should be based on as many results as possible, at least 20 and preferably 30 or more (see 13.1.8.1.1) and would normally be calculated at the end of a carefully controlled study.

Repeatability is most commonly used as a range test of the difference between two repeated measurements (see 13.1.4 and 13.1.7.1). It can also be used to construct a test on the range of three or more measurements. By combining Equations 23 and 8, the range can be represented by:

$$w = \frac{d(n)r}{\sqrt{2}t_{95,\ n-1}} \quad (24)$$

By substituting a previously determined repeatability value into this expression, a critical value can be calculated for the range of a set of n measurements. However, it is advisable not to use this as a formal outlier test. Because the range represents only a part of the information on the variability within a set of measurements (that is, the smallest and largest values), the test will only be approximate. Nevertheless, it can be used to monitor statistical control within a set of measurements (see 13.1.4 and 13.1.6.2) and flag the need for rigorous analysis.

13.1.8.2 STATISTICAL PROCEDURE FOR TWO OR MORE SETS OF DATA

In some cases, the quantity in question is obtained indirectly from m intermediate and independent results, each of which will have been estimated from a separate set of data according to the procedures in 13.1.8.1. In this section, procedures are given in which the estimates for the intermediate quantities are combined to give those relating to the final quantity.

13.1.8.2.1 Estimating True Value

The value X of the final result is assumed to be a function F of the m intermediate quantities $X_1, X_2 \ldots X_m$.

Algebraically, this can be represented by:

$$X = F(X_1, X_2, \ldots X_m) \quad (25)$$

The maximum estimate of X is obtained simply by substituting into Equation 25 the appropriate estimates for $X_1, X_2, \ldots X_m$. In terms of measurements corrected for bias (see 13.1.8.1.3), the estimate $\bar{y}$ of the true variable will become:

$$\bar{y} = F(\bar{y}_1, \bar{y}_2, \ldots \bar{y}_m) \quad (26)$$

As an example, consider a relationship between quantities of the form:

$$X = PX_1X_2 + QX_3 \quad (27)$$

P and Q are known constants. The estimate of the final quantity, according to Equation 26, is then:

$$\bar{y} = P\bar{y}_1 \bar{y}_2 + Q\bar{y}_3 \quad (28)$$

Note that the calculation resulting from such an equation could give rise to a further source of systematic error (see 13.1.8.2.3). This would be the case, for example, in estimating the volume of a tank from tables of liquid depth. The intermediate results would include assumed values for tank dimensions, with further assumptions on environmental conditions, and these could all contribute to unknown systematic errors.

13.1.8.2.2 Combining Random Uncertainties

Random error is represented statistically by the standard deviation (sometimes called standard error) associated with a particular measurement function (see 13.1.8.1.5). It is useful when combining random errors to consider another parameter called variance. Standard deviation σ is simply the square root of the variance V:

$$V(X) = \sigma^2(X) \quad (29)$$

In terms of estimates corrected for bias:

$$v(y) = s^2(y) \quad (30)$$

Any of the expressions dealing with standard deviation may be converted to the corresponding expressions for variance by substituting Equation 29 or 30.

Now consider the random errors associated with the m intermediate quantities. The variance of the indirect measurements of the final quantity is given approximately by the expression:

$$V(X) = \left(\frac{\sigma F}{\sigma X_1}\right)^2 V(X_1) + \left(\frac{\sigma F}{\sigma X_2}\right)^2 V(X_2) + \ldots + \left(\frac{\sigma F}{\sigma X_m}\right)^2 V(X_m) \quad (31)$$

$\sigma F/\sigma X_1$ represents the partial differential coefficient of F with respect to X_1, and F is Equation 25. $\sigma F/\sigma X_1$ may be regarded as the change in F brought about by unit change in X_1. Equation 31 only holds true, however, if the quantities $X_1, X_2, \ldots X_m$ are independent of each other. Furthermore, the equation leads to the root sum square method of combining random uncertainty limits (see 13.1.8.1.6.3), for by substituting into it Equations 11 and 29, it becomes:

$$A(X) = \sqrt{\left[\frac{\sigma F}{\sigma X_1} A(X_1)\right]^2 + \left[\frac{\sigma F}{\sigma X_2} A(X_2)\right]^2 + \ldots + \left[\frac{\sigma F}{\sigma X_m} A(X_m)\right]^2} \quad (32)$$

The corresponding equation in terms of estimates will be:

$$a(\bar{y}) = \sqrt{\left[\frac{\sigma F}{\sigma X_1} a(\bar{y}_1)\right]^2 + \left[\frac{\sigma F}{\sigma X_2} a(\bar{y}_2)\right]^2 + \ldots + \left[\frac{\sigma F}{\sigma \bar{y}_m} a(\bar{y}_m)\right]^2} \quad (33)$$

13.1.8.2.3 Combining Systematic Uncertainties

As previously stated, there are theoretical difficulties when attempting to combine systematic uncertainties (see 13.1.8.1.6). The choice is between the arithmetic and root sum square methods of combining and should take into account the manner in which the errors behave in practice. This is sometimes difficult to judge, particularly for the multiplicative terms in a relationship (see Equation 27). Assuming a uniform distribution of systematic errors, however, it is theoretically correct to combine the systematic errors in a multiplicative function by the root sum square method. This, coupled with the fact that systematic errors combined in an additive fashion are expected to cancel each other out to a certain extent, leads to the choice of the root sum square method as applicable in the general case. The appropriate formula is identical in form to Equation 33 but with random uncertainty limits replaced by the corresponding systematic limits:

$$b(\bar{y}) = \sqrt{\left[\frac{\sigma f}{\sigma X_1} b(\bar{y}_1)\right]^2 + \left[\frac{\sigma f}{\sigma X_2} b(\bar{y}_2)\right]^2 + \ldots + \left[\frac{\sigma f}{\sigma X_m} b(\bar{y}_m)\right]^2} \quad (34)$$

The point to remember when combining systematic uncertainties is that the relationship between quantities (see Equation 25) may only be approximate (see 13.1.8.2.1). In that case, a further unknown systematic error could be present, and the corresponding uncertainty limit should be estimated according to Equation 15. This should be included as another squared term in the uncertainty expression (Equation 34).

13.1.8.2.4 Estimating Total Uncertainty

For the reasons already explained in 13.1.8.1.6.3, the random and systematic components of the total uncertainty should be combined by quadrature according to Equation 19. In this case, however, $a(\bar{y})$ will be estimated from Equation 33 and $b(\bar{y})$ from Equation 34.

13.1.8.3 ROUNDING STATISTICAL ESTIMATES

When applying the procedures of 13.1.8.1 and 13.1.8.2, it is important to consider the effect of rounding on the statistical estimates derived. Rounding that is too coarse will become a significant source of error. Any particular result will be reported to the smallest unit of measure of the instrument involved, and the statistics

that relate to that result should reflect this level of accuracy and should be rounded accordingly. For example, a gage reading would be reported to the nearest millimeter if that was the scale unit of the tape measure. Estimates of the mean gage, standard deviation, and the limit of the range of uncertainty should also be rounded to the nearest millimeter, and the calculations leading up to those estimates should include a sufficient number of digits to achieve this.

Particular care should be taken when considering more complicated functions, such as would be found in the indirect estimation of a parameter from a number of intermediate calculations. It is useful to relate the calculations to the units in which the final estimate is to be reported. In a root sum square estimate, for example, which is to be reported to two decimal places, the squared terms should be calculated to at least four decimal places to achieve the required level of accuracy. From the opposite viewpoint, in terms of a root sum square estimate, if one or more of the squared terms were calculated to only two decimal places, it would be incorrect to report the final estimate to any greater accuracy than one decimal place.

All estimates, except repeatability, should be rounded up or down to the smallest unit of measure (rounding unit). As a result of its definition (see 13.1.7.1), a repeatability estimate should always be rounded up to the nearest rounding unit.

13.1.8.4 EXAMPLE

Consider the indirect measurement of the volume at standard temperature of liquid in a tank. This is to be estimated from a set of repeated gage readings, a calibration table, a set of repeated temperature measurements, and a temperature correction formula. Each set of direct measurements (gage readings and temperatures) will be considered separately according to 13.1.8.1, and the appropriate statistics will be derived. These will then be combined to give estimates in terms of the liquid volume corrected for expansion in the tank resulting from nonstandard temperature.

For the purpose of this example, the procedures of 13.1.8.1 will only be described in detail with respect to the set of gage readings. Statistics for the set of temperature data will be given. Note that statistics that are to be used at a later stage in the calculations will be stated to a greater level of accuracy (one or two more decimal places) than that achieved in the corresponding measurements. This is to ensure that the final estimate of volume includes no rounding errors. Note also that the figures used in the example were chosen strictly for illustrative purposes, and are not necessarily typical of those to be found in practice.

13.1.8.4.1 Direct Measurements

In this section, the procedures of 13.1.8.1 will be applied to the single set of tank gage measurements. This can be considered as separate steps as follows:

Step 1—Information available.

Six gage measurements x_i, for i = 1 to 6 (see 13.1.8.1) were recorded to the nearest millimeter: 6534, 6544, 6542, 6540, 6543, and 6544. Unknown systematic errors were expected as a result of sludge at the bottom of the tank and inaccuracy in the tank gage tape. These errors (see 13.1.8.1.3) recorded in millimeters as:

Source of Systematic Error	Maximum Range of Error e_1	e_2
Sludge	−4	0
Tape	−1	+1

It is also known from an independent study that the repeatability for tank gaging was 7 millimeters.

Step 2—Outlying results.

The first gage reading differs from the others by what appears to be an appreciable amount. As a quick check on its validity, the critical range for the set of measurements, rounded to the nearest millimeter, is calculated from Equation 24 as:

$$w = \frac{D(n)r}{\sqrt{2} \times (t_{95,\ n-1})}$$

Where:

$n = 6$.
$D(6) = 2.543$ (see Table 1).
$r = 7$.
$t_{95,5} = 2.571$ (see Table 2).

Therefore:

$$w = \frac{D(6)r}{\sqrt{2}\ t_{95,\ 5}}$$
$$= \frac{2.534 \times 7}{2 \times 2.571}$$
$$= 5 \text{ millimeters}$$

The observed range of 10 exceeds this value, so Dixon's outlier test was applied (see Appendix B). The appropriate Dixon ratio for six measurements and for testing a low value is:

$$R_{10} = \frac{6540 - 6534}{6544 - 6534}$$
$$= 0.6$$

which exceeds the critical ratio at the 95 percent

probability level. The first measurement was rejected as a faulty reading (outlier), and all following calculations disregard it.

Step 3—Correcting for bias.

According to Equation 4, the average systematic error due to tape inaccuracy is zero, but that for sludge is given by:

$$\bar{e} = \frac{(e_1 + e_2)}{2} = \frac{-4 + 0}{2} = -2 \text{ millimeters}$$

The results x_i must be adjusted according to Equation 1 to give the corrected measurements Y_i, for i = 1 to 5: 6546, 6544, 6542, 6545, and 6546.

Step 4—Estimating true gage reading.

This will be the average $\bar{y}$ of the results corrected for bias (Equation 5), that is:

$$\bar{y} = \frac{6546 + 6544 + \ldots + 6546}{5} = 6544.6 \text{ millimeters}$$

Step 5—Estimating standard deviation.

The standard deviation of corrected measurements can be estimated both from Equation 7 as:

$$s(y) = \tfrac{1}{4}\sqrt{(1.4^2 + 0.6^2 + 2.6^2 + 0.4^2 + 1.4^2)} = 1.67$$

and from Equation 8 as:

$$s(y) = \frac{w}{D(5)} = \frac{4}{2.326} = 1.72$$

Statistics derived from the second and more approximate estimate will be given in parentheses for comparative purposes.

Step 6—Estimating range of uncertainty.

By substituting the standard deviation estimates into Equation 13, the limit of the range of uncertainty due to random errors becomes:

$$a(\bar{y}) = \frac{[t_{95,4}]\,[s(y)]}{\sqrt{5}} = \frac{2.776 \times 1.67}{\sqrt{5}} = 2.07\ (2.14) \text{ millimeters}$$

Since there are two unknown sources of systematic error, the corresponding limits of uncertainty will be estimated for each according to Equations 15 and 16, respectively, as:

$$\text{limit due to sludge } b_1(y) = 0.95 \times \left|\frac{(-4 - 0)}{2}\right| = 1.9 \text{ millimeters}$$

$$\text{limit due to tape } b_2(y) = 0.95 \times |-1| = 0.95 \text{ millimeter}$$

Combining the systematic limits by the root sum square method (Equation 18) gives the total limit for systematic errors:

$$b(y) = \sqrt{1.9^2 + 0.95^2} = 2.12 \text{ millimeters}$$

The limits for systematic and random uncertainties should be combined in a similar manner (Equation 19) to give:

$$c(\bar{y}) = \sqrt{a^2(\bar{y}) + b^2(\bar{y})} = \sqrt{2.07^2 = 2.12^2} = 2.96\ (3.01) \text{ millimeters}$$

Step 7—Estimating repeatability.

To compare the variability within the set of measurements to that expected in practice, the repeatability can be estimated from Equation 23 as:

$$r = (t_{95,4}) \times \sqrt{2} \times s(y) = 2.776 \times \sqrt{2} \times 1.67 = 6.6\ (6.7) \text{ millimeters}$$

Rounding up to the nearest unit of measure of 1 millimeter (see 13.1.8.3), the repeatability estimates would become 7 millimeters. This is identical to the value derived from the independent study.

Step 8—Stating the result.

The estimate of the true gage reading should be stated after rounding to the nearest unit of measure (see 13.1.5.4):

$$C(\bar{y}) = 2.96 \sim 3$$

The result statement thus becomes:

True gage reading = 6545 ± 3 millimeters (95 percent confidence level, 5 measurements)

NOTE: One further result is rejected as a faulty reading (outlier).

Table 3—Derived Statistics for Example

Value	Gage Reading, millimeters	Thermometer Reading, °C
n	5	9
$\bar{y}$	6544.6	23.38
$a(\bar{y})$	2.07	1.362
$b(\bar{y})$	2.12	0.500

Table 4—Symbols for Example

Measurement	True Value	Estimate Corrected for Bias
Depth	X_1	y_1
Absolute volume	X_2	y_2
Temperature	X_3	y_3
Corrected volume	X_4	y_4

13.1.8.4.2 Measuring Volume

Next, the statistics derived from the two sets of direct measurements are combined according to the procedures of 13.1.8.2 as follows:

Step 1—Information available.

The information corresponding to the direct measurements can be summarized in the form of derived statistics as in Table 3. Let us also assume that the symbols allocated to each quantity are as listed in Table 4.

The calibration table, by which a gage reading in millimeters can be converted to a volume in liters, was obtained from an unknown function of tank dimensions. No random error is created in the use of such a table, but an unknown systematic error is expected resulting from the approximate nature of the function. This was assumed to be level dependent, and the corresponding limit of uncertainty is estimated to be:

$$b(X_2) = 0.2\%\ X_2$$

Finally, the function (see Equation 25) used to correct the volume for temperature and expansion in the tank is:

$$X_4 = F(X_2, X_3)$$
$$= f(X_3)\ X_2\ [1 + 0.000022\ (X_3 - 15)]$$

Table 5—Volume Measurement Statistics for Example

Value	Gage Reading, milliliters	Volume Measurement, liters
n	5	5
$\bar{y}$	6544.6	17016
$a(\bar{y})$	2.07	5
$b_1(\bar{y})$	2.12	6

Where:

$f(X_3)$ = a factor (read from tables) corresponding to a temperature X_3.

Step 2—Estimating absolute volume.

The calibration table may be regarded as a means of converting a liquid depth measurement (gage reading) in millimeters to a liquid volume measurement in liters and may be represented by the function:

$$X_2 = PX_1$$

P is nearly constant in this example. The statistics that relate to volume results should, therefore, be read directly from the table. In this case, they are assumed to be those in Table 5.

The systematic error brought about by inaccuracies in the table should be considered at this point. In terms of results corrected for bias, the corresponding limit of uncertainty will be estimated as:

$$b_2(\bar{y}_2) = 0.2\%\ \bar{y}_2$$
$$= 0.002 \times 17016$$
$$= 34 \text{ liters}$$

The two limits for systematic errors that affect volume results are then combined by the root sum square method (Equation 18) to give:

$$b(y_2) = \sqrt{b_1^2(y_2) + b_2^2(y_2)}$$
$$= \sqrt{6^2 + 34^2}$$
$$= 35 \text{ liters}$$

Step 3—Estimating corrected volume.

According to Equation 26, the estimate of correct volume will be obtained by substituting estimates directly into the appropriate equation. If we assume that the temperature factor $f(x_3)$ is read from tables as:

$$f(\bar{y}_3) = f(23.38)$$
$$= 0.98 \text{ (given)}$$

Then the true corrected volume is estimated to be:

$$\bar{y}_4 = 0.98 \times 17016\ [1 + 0.000022(23.38 - 15)]$$
$$= 16678.9 \text{ liters}$$

Step 4—Estimating random uncertainty limit.

The random errors for volume and temperature measurements are combined according to Equation 33. In our case, the derivatives of the function are:

$$\frac{\alpha F}{\alpha X_2} = 0.98\ [1 + 0.000022\ (X_3 - 15)]$$

and

$$\frac{\alpha F}{\alpha X_3} = 0.98 \times X_2 \times 0.000022$$

Substituting the estimated values,

$\bar{y}_3 = 23.38$ (Table 3)
$\bar{y}_2 = 17016$ (Table 5)

For X_3 and X_2, respectively, gives:

$$\frac{\alpha F}{\alpha X_2} = 0.98018$$

$$\frac{\alpha F}{\alpha X_3} = 0.36638$$

The total limit of random uncertainty will then be given by:

$$a(y^4) = \sqrt{(0.98018 \times 5)^2 + (0.36686 \times 1.362)^2}$$
$$= 4.9 \text{ liters}$$

Step 5—Estimating systematic uncertainty limit.

Systematic uncertainty limits should be combined in a fashion similar to random uncertainty limits according to Equation 34. The total limit of systematic uncertainty will be:

$$b(\bar{y}_4) = \sqrt{(0.98018 \times 35)^2 + (0.36686 \times 0.5)^2}$$
$$= \sqrt{34.3036^2 + 0.18343^2}$$
$$= 34.3 \text{ liters}$$

Note that the systematic error in temperature measurements makes only a small contribution compared with that created by inaccuracies in the calibration table.

Step 6—Stating the result.

Combining the random and systematic components of uncertainty by quadrature (Equation 19), the total uncertainty limit becomes:

$$c(\bar{y}_4) = \sqrt{a^2(\bar{y}_4) + b^2(\bar{y}_4)}$$
$$= \sqrt{4.9^2 + 34.3^2}$$
$$= 34.6 \text{ liters}$$

Rounding to the nearest unit of measurement, which from the calibration table was 1 liter, the final statement will be:

True corrected volume = 16,679 ± 35 liters (95 percent confidence level, 5 gage measurements, 9 temperature measurements).

NOTE: One further gage reading is rejected as a faulty reading.

APPENDIX A
NORMAL (GAUSSIAN) DISTRIBUTION

Consider a set of n repeated measurements x_i, lying in the range a to b so that $a \leq x_i \leq b$. If the total range is split into p equal subranges of length $dx = (b - a)/p$, a frequency histogram can be drawn. The histogram (see Figure A-1) consists of a series of p contiguous rectangles, with base equal to the subrange dx and height proportional to the number of measurements falling in that range.

The height of each rectangle could just as easily represent the proportion of the total number falling in the subrange or the relative frequency. The total area of the histogram would then be 1, and the area in each rectangle would become the probability of a measurement falling in the subrange.

Now consider the number of measurements n becoming very large, and the length dx of each subrange becoming very small. A continuous line drawn through the midpoint of the tops of each rectangle, which represents the relative frequency of measurements, would give a bell-shaped curve similar to Figure A-2.

For the normal distribution, the curve is symmetrical about the mean and has the formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(x-\mu)^2}{\sqrt{2\alpha}}\right)$$

Where:

σ = standard deviation.

The area under the curve once again represents probability. Each of the shaded regions shown has an area of:

$$P = \int_{-\infty}^{\mu-c} f(x)\,dx = \int_{\mu+c}^{+\infty} f(x)\,dx$$

When $c = 1.96\sigma$, the probability P (one shaded area) will be 0.025, or 2½ percent of the total area under the curve.

Now if measurements x_i follow the normal distribution with mean μ and standard deviation σ, then values μ_i will follow a normal distribution with zero mean and unit standard deviation where:

$$\mu_i = \frac{(x_i - \mu)}{\sigma}$$

The value μ_i is termed the *standard normal deviate*, and has been tabulated for different probabilities P. For a

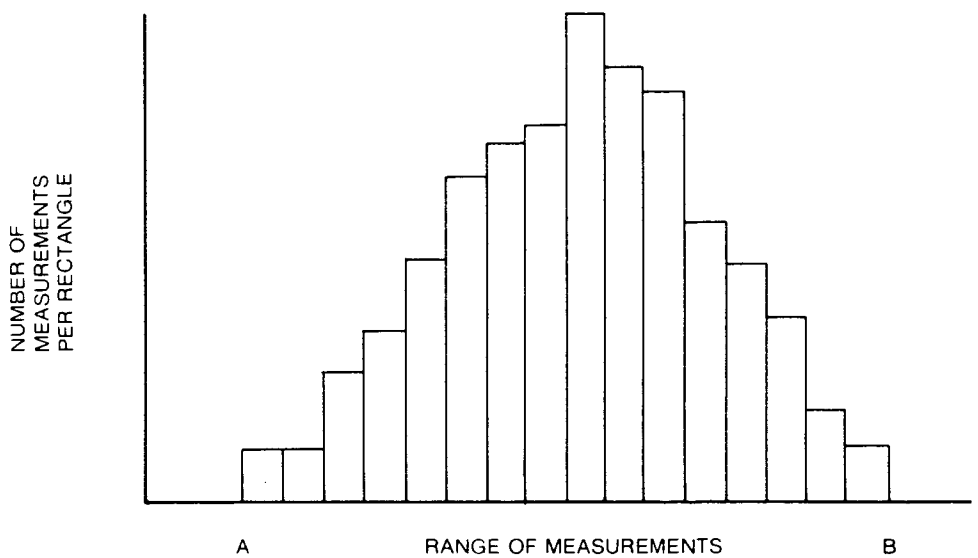

Figure A-1—Frequency Histogram

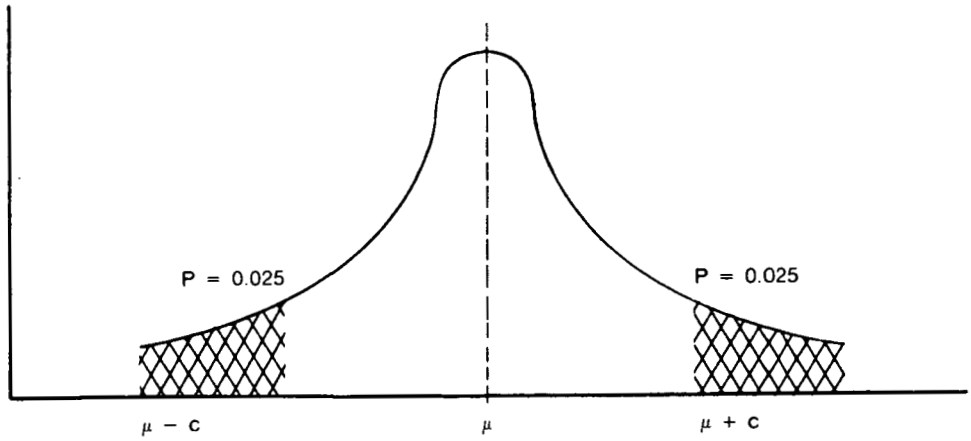

Figure A-2—Bell-Shaped Curve

probability P = 0.05, however, the standard normal deviate has a value 1.96. This probability is represented by both shaded areas in the distribution (Figure A-2) and includes all values of x which differ from the mean μ by more than 1.96σ.

APPENDIX B
DIXON'S TEST FOR OUTLIERS

The following steps should be followed (see Table B-1) to use Dixon's test for outliers.

1. Arrange the set of measurements x_i in ascending order of magnitude $x_1, x_2, \ldots x_n$.
2. Choose the appropriate test criterion, depending on the value of n and whether the measurement in question is low or high.
3. Calculate the Dixon R ratio. If this exceeds the critical ratio at the 5 percent probability level (P = 0.95), then the measurement in question is highly suspect and could possibly be rejected.

If the critical ratio at the 1 percent probability level (P = 0.99) is exceeded, then the measurement in question should be discarded.

When a measurement is rejected, the outlier test should be repeated.

NOTE: The two suffixes in the Dixon ratio refer to the differences in the numerator and denominator respectively.

Table B-1—Dixon's Test for Outliers

Number of Values, n	Critical Values P = 0.95	Critical Values P = 0.99	Test Criterion Low Values	Test Criterion High Values
3	0.941	0.988		
4	0.765	0.889		
5	0.642	0.780	$R_{10} = \frac{x_2 - x_1}{x_n - x_1}$	or $\frac{x_n - x_{n-1}}{x_n - x_1}$
6	0.560	0.698		
7	0.507	0.637		
8	0.554	0.683		
9	0.512	0.635	$R_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	or $\frac{x_n - x_{n-1}}{x_n - x_2}$
10	0.477	0.597		
11	0.576	0.679		
12	0.546	0.642	$R_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	or $\frac{x_n - x_{n-2}}{x_n - x_2}$
13	0.521	0.615		
14	0.546	0.641		
15	0.525	0.616		
16	0.507	0.595		
17	0.490	0.577		
18	0.475	0.561		
19	0.462	0.547	$R_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	or $\frac{x_n - x_{n-2}}{x_n - x_3}$
20	0.450	0.535		
21	0.440	0.524		
22	0.430	0.514		
23	0.421	0.505		
24	0.413	0.497		
25	0.406	0.489		

SOURCE: *Biometrics*, Vol. 9, p. 89, 1953.

Order No. 852-30321

1-1700— 6/85—1.5M
1-1701 10/89— 3C (2A)
1-1700— 3/91—2.5C (5D) C

American Petroleum Institute
1220 L Street, Northwest
Washington, D.C. 20005