INTERNATIONAL

Designation: E1970 – 23

Standard Practice for Statistical Treatment of Thermoanalytical Data[1]

This standard is issued under the fixed designation E1970; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope*

1.1 This practice details the statistical data treatment used in some thermal analysis methods.

1.2 The method describes the commonly encountered statistical tools of the mean, standard derivation, relative standard deviation, pooled standard deviation, pooled relative standard deviation, the best fit to a (linear regression of a) straight line (or plane), and propagation of uncertainties for all calculations encountered in thermal analysis methods (see Practice E2586).

1.3 Some thermal analysis methods derive the analytical value from the slope or intercept of a linear regression straight line (or plane) assigned to three or more sets of data pairs. Such methods may require an estimation of the precision in the determined slope or intercept. The determination of this precision is not a common statistical tool. This practice details the process for obtaining such information about precision.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.5 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E456 Terminology Relating to Quality and Statistics

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

E2161 Terminology Relating to Performance Validation in Thermal Analysis and Rheology

E2586 Practice for Calculating and Using Basic Statistics

F1469 Guide for Conducting a Repeatability and Reproducibility Study on Test Equipment for Nondestructive Testing (Withdrawn 2018)[3]

3. Terminology

3.1 *Definitions*—The technical terms used in this practice are defined in Practice E177 and Terminologies E456 and E2161 including *precision, relative standard deviation, repeatability, reproducibility, slope, standard deviation, thermoanalytical*, and *variance*.

3.2 *Symbols* (*1*):[4]

a, c, m = slope
b, d = intercept
n = number of data sets (that is, x_i, y_i)
r = correlation coefficient
R = gage reproducibility and repeatability (see Guide F1469) an estimation of the combined variation of repeatability and reproducibility (**2**)
R^2 = coefficient of determination
RSD = relative standard deviation
s = standard deviation
s_b = standard deviation of the line intercept
s_i = standard deviation of the "*i*th" measurement
s_m = standard deviation of the slope of a line
s_{pooled} = pooled standard deviation
s_r = within laboratory repeatability standard deviation (see Practice E691)
s_R = between laboratory repeatability standard deviation (see Practice E691)
s_y = standard deviation of Y values
X = mean x value
x_i = an individual independent variable observation
Y = mean y value
y_i = an individual dependent variable observation

[1] This practice is under the jurisdiction of ASTM Committee E37 on Thermal Measurements and is the direct responsibility of Subcommittee E37.10 on Fundamental, Statistical and Mechanical Properties.

Current edition approved Jan. 1, 2023. Published February 2023. Originally approved in 1998. Last previous edition approved in 2021 as E1970 – 16(2021). DOI: 10.1520/E1970-23.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] The last approved version of this historical standard is referenced on www.astm.org.

[4] The boldface numbers in parentheses refer to a list of references at the end of this standard.

***A Summary of Changes section appears at the end of this standard**

Z = mean z value
z_i = an individual independent variable observation
Σ = mathematical operation which means "the sum of all" for the term(s) following the operator
δy_i = variance in y parameter

4. Summary of Practice

4.1 The result of a series of replicate measurements of a value are typically reported as the mean value plus some estimation of the precision in the mean value. The standard deviation is the most commonly encountered tool for estimating precision, but other tools, such as relative standard deviation or pooled standard deviation, also may be encountered in specific thermoanalytical test methods. This practice describes the mathematical process of achieving mean value, standard deviation, relative standard deviation and pooled standard deviation and other terms relating to the statistical treatment of thermoanalytical data.

4.2 In some thermal analysis experiments, a linear or a straight line, response is assumed and desired values are obtained from the slope or intercept of the straight line through the experimental data. In any_i practical experiment, however, there will be some uncertainty in the data so that results are scattered about such a straight line. The linear regression (also known as "least squares") method is an objective tool for determining the "best fit" straight line drawn through a set of experimental results and for obtaining information concerning the precision of determined values.

4.2.1 For the purposes of this practice, it is assumed that the physical behavior, which the experimental results approximate, are linear with respect to the controlled value, and may be represented by the algebraic function in Eq 1:

$$y = mx+b \tag{1}$$

4.2.2 Should the physical behavior be linear with respect to two control variables, then the relationship is a plane and may be represented by the algebraic function in Eq 16.

4.2.3 Experimental results are gathered in pairs (or sets), that is, for every corresponding x_i (and z_i) (controlled) value(s), there is a corresponding y_i (response) value.

4.2.4 The best fit (linear regression) approach assumes that all x_i (and z_i) values are exact and the y_i values (only) are subject to uncertainty.

NOTE 1—In experimental practice, both x and y values are subject to uncertainty. If the uncertainty in x_i and y_i are of the same relative order of magnitude, other more elaborate fitting methods should be considered. For many sets of data, however, the results obtained by use of the assumption of exact values for the x_i data constitute such a close approximation to those obtained by the more elaborate methods that the extra work and additional complexity of the latter is hardly justified **(2 and 3)**.

4.2.5 The best fit approach seeks a straight line, which minimizes the uncertainty in the y_i value.

4.3 The law of propagation of uncertainties is a tool for estimating the precision in a determined value from the sum of the variance of the respective measurements from which that value is derived weighted by the square of their respective sensitivity coefficients.

4.3.1 Variance is the square of the standard deviation(s). Conversely the standard deviation is the positive square root of the variance.

4.3.2 The sensitivity coefficient is the partial derivative of the function with respect to the individual variable.

5. Significance and Use

5.1 The standard deviation, or one of its derivatives, such as relative standard deviation or pooled standard deviation, derived from this practice, provides an estimate of precision in a measured value. Such results are ordinarily expressed as the mean value ± the standard deviation, that is, $X \pm s$.

5.2 If the measured values are, in the statistical sense, "normally" distributed about their mean, then the meaning of the standard deviation is that there is a 67 % chance, that is 2 in 3, that a given value will lie within the range of ± one standard deviation of the mean value. Similarly, there is a 95 % chance, that is 19 in 20, that a given value will lie within the range of ± two standard deviations of the mean. The two standard deviation range is sometimes used as a test for outlying measurements.

5.3 The calculation of precision in the slope and intercept of a line, derived from experimental data, commonly is required in the determination of kinetic parameters, vapor pressure or enthalpy of vaporization. This practice describes how to obtain these and other statistically derived values associated with measurements by thermal analysis.

6. Calculation

6.1 Commonly encountered statistical results in thermal analysis are obtained in the following manner.

NOTE 2—In the calculation of intermediate or final results, all available figures shall be retained with any rounding to take place only at the expression of the final results according to specific instructions or to be consistent with the precision and bias statement.

6.1.1 The mean value (X) is given by:

$$X = \frac{x_1+x_2+x_3+....+x_i}{n} = \frac{\Sigma x_i}{n} \tag{2}$$

A similar relation exists for Y and Z.

6.1.2 The standard deviation (s) is given by:

$$s = \left[\frac{\Sigma(x_i - X)^2}{(n-1)}\right]^{1/2} \tag{3}$$

6.1.3 The relative standard deviation (RSD) is given by:

$$RSD = (s \cdot 100\,\%)/X \tag{4}$$

6.1.4 The pooled standard deviation (s_p) is given by:

$$s_p = \left[\frac{\Sigma(\{n_i - 1\} \cdot s_i)}{\Sigma(n_i - 1)}\right]^{1/2} \tag{5}$$

NOTE 3—For the calculation of pooled relative standard deviation, the values of s_i are replaced by RSD_i.

6.1.5 The gage repeatability and reproducibility (R) is given by:

$$R = [s_R{}^2 + s_r{}^2]^{1/2} \tag{6}$$

NOTE 4—For the calculation of relative gage repeatability and reproducibility, the values of s_r and s_R are replaced with RSD_r and RSD_R.

6.2 *Linear Regression (Best) Fit Straight Line:*

6.2.1 The slope (m) is given by:

$$m = \frac{n\Sigma(x_iy_i) - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \quad (7)$$

6.2.2 The intercept (b) is given by:

$$b = \frac{(\Sigma x_i^2)(\Sigma y_i) - (\Sigma x_i)(\Sigma x_iy_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \quad (8)$$

6.2.3 The individual dependent parameter variance (δy_i) of the dependent variable (y_i) is given by:

$$\delta y_i = y_i - (mx_i + b) \quad (9)$$

6.2.4 The standard deviation s_y of the set of y values is given by:

$$s_y = \left[\frac{\Sigma(\delta y_i)^2}{n-2}\right]^{1/2} \quad (10)$$

6.2.5 The standard deviation (s_m) of the slope is given by:

$$s_m = s_y\left[\frac{n}{n\Sigma x_i^2 - (\Sigma x_i)^2}\right]^{1/2} \quad (11)$$

6.2.6 The standard deviation (s_b) of the intercept (b) is given by:

$$s_b = s_y\left[\frac{\Sigma x_i^2}{n\Sigma x_i^2 - (\Sigma x_i)^2}\right]^{1/2} \quad (12)$$

6.2.7 The denominators in Eq 7, Eq 8, Eq 11, and Eq 12 are the same. It is convenient to obtain the denominator (D) as a separate function for use in manual calculation of each of these equations.

$$D = n\Sigma x_i^2 - (\Sigma x_i)^2 \quad (13)$$

6.2.8 The linear correlation coefficient (r), a measure of the mutual dependence between paired x and y values, is given by:

$$r = \frac{n\Sigma xy - (\Sigma x_i)(\Sigma y_i)}{[n\Sigma x_i^2 - (\Sigma x_i)^2]^{1/2}[n(\Sigma y_i^2) - (\Sigma y_i)^2]^{1/2}} \quad (14)$$

NOTE 5—r may vary from –1 to +1, where values of +1 or –1 indicate perfect (100 %) correlation and 0 indicates no (0 %) correlation, that is, random scatter. A positive (+) value indicates a positive slope and a negative (–) indicates a negative slope.

6.2.9 The coefficient of determination (R^2) is given by:

$$R^2 = r^2 \quad (15)$$

6.3 *Multiple Linear Regression to a Plane:*

6.3.1 The equation for the best fit to a plane is given by:

$$y = a\,x + c\,z + d \quad (16)$$

6.3.2 The slope (a) for the variable x is given by:

$$a = (\alpha\;\beta - \gamma\;\delta)/(\varepsilon\;\alpha - \gamma^2) \quad (17)$$

where:

$$\alpha = \sum z^2 - \left(\sum z\right)^2/n \quad (18)$$

$$\beta = \sum xy - \left(\sum x \sum y\right)/n \quad (19)$$

$$\gamma = \sum xz - \left(\sum x \sum z\right)/n \quad (20)$$

$$\delta = \sum zy - \left(\sum z \sum y\right)/n \quad (21)$$

$$\varepsilon = \sum x^2 - \left(\sum x\right)^2/n \quad (22)$$

6.3.3 The slope (c) for the variable z is given by:

$$c = (\varepsilon\;\delta - \gamma\;\beta)/(\varepsilon\;\alpha - \gamma^2) \quad (23)$$

6.3.4 The intercept (d) is given by:

$$d = Y - a\,X - c\,Z \quad (24)$$

6.3.5 Calculation of the planar correlation coefficient is sufficiently complex that it is seldom manually performed and is not included in this standard. However, most multiple linear regression software packages do provide the planar correlation coefficient should it be needed.

6.4 *Propagation of Uncertainties:*

6.4.1 The law of propagation of uncertainties, neglecting the cross terms, is given by:

$$s_z^2 = \Sigma[(\partial z/\partial i)s_i]^2 \quad (25)$$

or

$$s_z = \{\Sigma\,[(\partial z/\partial i)\;s_i]^2\}^{1/2} \quad (26)$$

6.4.2 For example, given the function $z = a\,d\,/c$, then the sensitivity coefficient for a is $\partial z/\partial a = d/c$, for d is $\partial z/\partial d = a/c$, and for c is $\partial z/\partial c = -ad/c_2$.

6.4.3 Eq 26 becomes:

$$s_z = \{[(\partial z/\partial a)s_a]^2 + [(\partial z/\partial d)s_d]^2 + [(\partial z/\partial c)s_c]^2\}^{1/2} \quad (27)$$

or

$$s_z = \{(d\;\;s_a/c)^2 + (a\;\;s_d/d)^2 + (a\;\;d\;\;s_c/c^2)^2\}^{1/2}$$

6.4.4 Dividing both sides of the equation by $z = a\,d/c$, yields:

$$s_z/z = \{(s_a/a)^2 + (\;s_d/d)^2 + (\;s_c/c)^2\}^{1/2} \quad (28)$$

6.4.5 The form of Eq 27 has been determined for a number of functions and is presented in Table 1.

6.5 *Example Calculations:*

6.5.1 Table 2 provides an example set of data and intermediate calculations which may be used to examine the manual calculation of slope (m) and its standard deviation (s_m) and of the intercept (b) and its standard deviation (s_b).

6.5.1.1 The values in Columns A and B are experimental parameters with x_i being the independent parameter and y_i the dependent parameter.

6.5.1.2 From the individual values of x_i and y_i in Columns A and B in Table 2, the values for x_i^2 and x_iy_i are calculated and placed in Columns C and D.

6.5.1.3 The values in columns A, B, C, and D are summed (added) to obtain $\Sigma x_i = 76.0$, $\Sigma_{yi} = 86.7$, $\Sigma x_i^2 = 1540.0$, and $\Sigma x_iy_i = 1753.9$, respectively.

6.5.1.4 The denominator (D) is calculated using Eq 13 and the values $\Sigma x_i^2 = 1540.0$ and $\Sigma x_i = 76.0$ from 6.5.1.3.

$$D = (6 \cdot 1540.0) - (76.0 \cdot 76.0) = 3464.0 \quad (36)$$

6.5.1.5 The value for m is calculated using the values $n = 6$ $\Sigma x_i \cdot y_i = 1753.9$, $\Sigma x_i = 76.0$, $\Sigma y_i = 86.7$, and $D = 3640.0$, from 6.5.1.3 and 6.5.1.4 and Eq 7:

$$m = \frac{n\Sigma(x_i\,y_i) - \Sigma x_i\,\Sigma y_i}{D} \quad (37)$$

$$m = \frac{(6 \cdot 1753.9) - (76.0 \cdot 86.7)}{3464.0} = \frac{10523.4 - 6589.2}{3464.0} \quad (38)$$

$$= 1.1357$$