APPENDIXES

APPENDIX A
PRECISION INDEX FOR UNIFORM DISTRIBUTION OF ERROR

The precision index for a uniform distribution of error is easily calculated by considering the definition of the variance:

$$\sigma^2 = \int_a^b (x - \mu)^2 \, p(x) \, dx \qquad \text{(the general formula for the variance).}$$

For a uniform distribution (Fig. A-1) between the limits of a and b, the formula is

$$\sigma^2 = \int_a^b \left(x - \frac{a + b}{2}\right)^2 \frac{1}{(b - a)} \, dx$$

where (a + b)/2 is the mean (μ) of the uniform distribution.

$$\sigma^2 = \int_a^b \left(x^2 - (a + b)\, x + \frac{a^2 + 2ab + b^2}{4}\right) \frac{1}{(b - a)} \, dx$$

$$\sigma^2 = \frac{1}{(b - a)}\left(\int_a^b x^2 \, dx - \frac{(a + b)}{1}\int_a^b x \, dx + \frac{a^2 + 2ab + b^2}{4}\int_a^b dx\right)$$

$$\sigma^2 = \frac{1}{(b - a)}\left[\frac{b^3 - a^3}{3}\right] - \frac{(a + b)}{(b - a)}\left[\frac{b^2 - a^2}{2}\right] + \frac{a^2 + 2ab + b^2}{4(b - a)}\left[b - a\right]$$

$$\sigma^2 = \frac{(b - a)^2}{12} \quad \text{or} \quad \sigma = \sqrt{\frac{(b - a)^2}{12}}$$

When a = -1 and b = +1, $\sigma = \sqrt{\frac{[1.0 - (-1.0)]^2}{12}} = \sqrt{\frac{1}{3}} \approx 0.577$

when a = -1/2 and b = +1/2, $\sigma = \sqrt{\frac{(\frac{1}{2} - (-\frac{1}{2}))^2}{12}} = \sqrt{1/12} \approx 0.3$

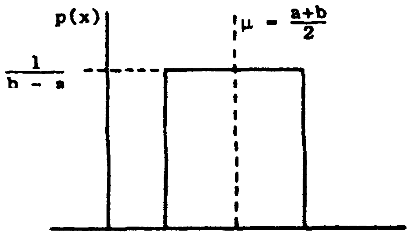


Fig. A-1 Uniform Distribution of Error

APPENDIX B
PROPAGATION OF ERRORS BY TAYLOR'S SERIES

GENERAL

The proofs in this section are shown for two- and three-variable functions. These proofs can be easily extended to functions with more variables, although, because of its length, the general case is not shown here.

TWO INDEPENDENT VARIABLES

If it is assumed that response Z is defined as a function of measured variables (x and y), the two restrictions that must be considered are

1. Z is continuous in the neighborhood of the point (μ_x, μ_y). Both x and y will have error distributions about this point and the notation (μ_x and μ_y) indicates the mean values of these distributions.
2. Z has continuous partial derivatives in a neighborhood of the point (μ_x, μ_y).

These conditions are satisfied if the functions to be considered are restricted to smooth curves in a neighborhood of the point with no discontinuities (jumps or breaks in the curve). The Taylor's series expansion for Z is

$$Z = Z_{\mu_x, \mu_y} + \frac{\partial Z}{\partial x}(x - \mu_x) + \frac{\partial Z}{\partial y}(y - \mu_y) + R_2 \qquad \text{(B-1)}$$

where $\frac{\partial Z}{\partial x}$ and $\frac{\partial Z}{\partial y}$ are evaluated at the point (μ_x, μ_y).

$$R_2 \leq \frac{1}{2}\left[\frac{\partial^2 Z}{\partial x^2}(x - \mu_x)^2 + \frac{\partial^2 Z}{\partial y^2}(y - \mu_y)^2\right] \qquad \text{(B-2)}$$

where $\frac{\partial^2 Z}{\partial x^2}$ and $\frac{\partial^2 Z}{\partial y^2}$ are evaluated at (θ_1, θ_2) with θ_1 between x and μ_x, and θ_2 between y and μ_y.

The quantity R_2, the remainder after two terms, is not significant if either:

1. $(x - \mu_x)$ and $(y - \mu_y)$ are small
2. The second partials $\frac{\partial^2 Z}{\partial x^2}$ and $\frac{\partial^2 Z}{\partial y^2}$ are small or zero. These partials are zero for linear functions.

By assuming R_2 to be small or zero, Eq. (B-1) becomes

$$Z \doteq \mu_Z + \frac{\partial Z}{\partial x}(x - \mu_x) + \frac{\partial Z}{\partial y}(y - \mu_y) \tag{B-3}$$

or

$$Z - \mu_Z \doteq \frac{\partial Z}{\partial x}(x - \mu_x) + \frac{\partial Z}{\partial y}(y - \mu_y) \tag{B-4}$$

By defining μ_Z as the average value of the distribution of Z, the difference $(Z - \mu_Z)$ is the difference of Z about its average value. This difference may be approximated by (Eq. (B-4))

$$Z - \mu_Z \approx \frac{\partial Z}{\partial x}(x - \mu_x) + \frac{\partial Z}{\partial y}(y - \mu_y) \tag{B-5}$$

where the partials are evaluated at the point (μ_x, μ_y).

The variation in Z is defined by

$$\sigma_Z^2 \equiv \int (Z - \mu_Z)^2 p_Z dZ$$

where p_Z is the probability density function of Z. Therefore,

$$\sigma_Z^2 \approx \iint \left[\frac{\partial Z}{\partial x}(x - \mu_x) + \frac{\partial Z}{\partial y}(y - \mu_y)\right]^2 p_{xy}\, dxdy \tag{B-6}$$

$$= \iint \left[\frac{\partial Z}{\partial x}(x - \mu_x)\right]^2 p_{xy}\, dydx + \iint \left[\frac{\partial Z}{\partial y}(y - \mu_y)\right]^2 p_{xy}\, dxdy$$
$$+ 2\iint \left[\frac{\partial Z}{\partial x}(x - \mu_x)\right]\left[\frac{\partial Z}{\partial y}(y - \mu_y)\right] p_{xy}\, dxdy \tag{B-7}$$

where P_{xy} is the joint distribution function of x and y. Integrating the first term of Eq. (B-7) with respect to y and the second term of Eq (B-7) with respect to x gives

$$\sigma_Z^2 \approx \int \frac{\partial Z}{\partial x}^2 (x - \mu_x)^2 p_x dx + \int \frac{\partial Z}{\partial y}^2 (y - \mu_y)^2 p_y dy$$
$$+ 2\iint \frac{\partial Z}{\partial x}(x - \mu_x)\frac{\partial Z}{\partial y}(y - \mu_y) p_{xy} dxdy \tag{B-8}$$

If μ_x and μ_y are the means of the distributions of x and y, then define the following:

$$\sigma_x^2 = \int (x - \mu_x)^2 p_x dx \tag{B-9}$$

$$\sigma_y^2 = \int (y - \mu_y)^2 p_y dy \tag{B-10}$$

$$\rho_{xy}\sigma_x\sigma_y = \iint (x - \mu_x)(y - \mu_y) p_{xy}\, dxdy \tag{B-11}$$

where ρ_{xy} is the coefficient of correlation between x and y. Combining the definitions and Eq. (B-8) gives

$$\sigma_Z^2 \approx \left[\frac{\partial z^2}{\partial x}\sigma_x^2 + \frac{\partial z^2}{\partial y}\sigma_y^2 + 2\frac{\partial z \partial z}{\partial y \partial x}\rho\sigma_x\sigma_y\right] \tag{B-12}$$

If x and y are independent variables, then $\rho = 0$ and

$$\sigma_Z^2 \approx \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 \tag{B-13}$$

THREE INDEPENDENT VARIABLES

If it is assumed that Z is a function of variables x, y, and w, two restrictions must be considered:

1. Z is continuous in a neighborhood of the point (μ_x, μ_y, μ_w)
2. Z has continuous partial derivatives in a neighborhood of (μ_x, μ_y, μ_w)

If these restrictions are satisfied, then the Taylor's series expansion for Z in the vicinity of (μ_x, μ_y, μ_w) is

$$Z = \mu_Z + \frac{\partial z}{\partial x}(x - \mu_x) + \frac{\partial z}{\partial y}(y - \mu_y) + \frac{\partial z}{\partial w}(w - \mu_w) + R_2 \tag{B-14}$$

where $\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}$, and $\frac{\partial z}{\partial w}$ are evaluated at (μ_x, μ_y, μ_w),

$$R_2 \leq \frac{1}{2}\left[\frac{\partial^2 z}{\partial x^2}(x - \mu_x)^2 + \frac{\partial^2 z}{\partial y^2}(y - \mu_y)^2 + \frac{\partial^2 z}{\partial w^2}(w - \mu_w)^2\right] \tag{B-15}$$

These second partials are evaluated at a point θ_1, θ_2, θ_3 defined so that θ_1 is between μ_x and x, θ_2 is between μ_y and y, and θ_3 is between μ_w and w. The same restrictions apply to R_2 as defined for two-variable functions.

By assuming R_2 to be small or zero, Eq. (B-14) becomes

$$Z - \mu_Z \doteq \frac{\partial z}{\partial x}(x - \mu_x) + \frac{\partial z}{\partial y}(y - \mu_y) + \frac{\partial z}{\partial w}(w - \mu_w) \tag{B-16}$$

where the partials are evaluated at the point (μ_x, μ_y, μ_w).

The variation in Z is defined by

$$\sigma_Z^2 \equiv \int (Z - \mu_Z)^2 \; p_Z dZ \tag{B-17}$$

where p_Z is the probability density function of Z. Therefore,

$$\sigma_Z^2 \approx \iiint \left[\frac{\partial z}{\partial x}(x - \mu_x) + \frac{\partial z}{\partial y}(y - \mu_y) + \frac{\partial z}{\partial w}(w - \mu_w)\right]^2 p_{x,y,w} \; dxdydw \tag{B-18}$$

$$= \iiint \left[\frac{\partial z}{\partial x}(x - \mu_x)\right]^2 p_{x,y,w} \; dwdydx + \ldots$$
$$+ 2 \iiint \frac{\partial z}{\partial x}\frac{\partial z}{\partial y}(x - \mu_x)(y - \mu_y) p_{x,y,w} \; dwdxdy + \ldots \tag{B-19}$$

where $p_{x,y,w}$ is the joint distribution function of x, y, and w. Integrating in the proper order produces these results:

$$\sigma_Z^2 \approx \int \left(\frac{\partial z}{\partial x}\right)^2 (x - \mu_x)^2 p_x dx + \ldots + 2\iint \frac{\partial z}{\partial x}\frac{\partial z}{\partial y}(x - \mu_x)(y - \mu_y) \; p_{xy} dxdy \tag{B-20}$$

Therefore,

$$\sigma_Z^2 \approx \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y{}^2 + \left(\frac{\partial z}{\partial w}\right)^2 \sigma_w^2 + 2\,\frac{\partial z}{\partial x}\frac{\partial z}{\partial y}\rho_{xy}\sigma_x\sigma_y$$
$$+ 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial w}\rho_{xw}\sigma_x\sigma_w + 2\frac{\partial z}{\partial y}\frac{\partial z}{\partial w}\rho_{yw}\sigma_y\sigma_w \tag{B-21}$$

If x, y, and w are independent variables, then $\rho_{xy} = \rho_{xw} = \rho_{yw} = 0$ and

$$\sigma_Z^2 \approx \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial z}{\partial w}\right)^2 \sigma_w^2 \tag{B-22}$$

MONTE CARLO SIMULATION

To determine the restrictions that must be placed on applications of the method of partial derivatives, a Monte Carlo Simulator was designed to provide simulation checks for the computation of various functions. Comparative results are listed in Tables B-1 and B-2.

Table B-1 contrasts the results of the Monte Carlo simulation of the functions tabulated, column (7), with the estimates using Partial Derivatives, column (6). One thousand functional values were obtained in each simulation. Column (1) identifies the function simulated and column (2) gives the number of the simulation run. Column (3), Theoretical Input, includes the parameters of the populations from which the random

Table B-1 Results of Monte Carlo Simulation for Theoretical Input ($\sigma_x{}^2$, μ_x, $\sigma_y{}^2$, μ_y)

(1) Function	(2) Simulation Run Number	(3) Theoretical Input				(4) Method of Partials Estimated Variance (Theoretical)	(5) Method of Partials Estimated Variance (Actual Input)	(6) Input Variance Corrected for Nonindependence (Method of Partials)	(7) Observed Variance (Simulator Results)
		σ_x^2	μ_x	σ_y^2	μ_y				
x + y	1	1.0	10	4.0	20	5.0	4.9477	4.8496	4.8567
	2	1.0	10	4.0	20	5.0	4.9186	4.8435	4.8506
	3	1.0	10	4.0	20	5.0	5.0786	4.9493	4.9564
	4	1.0	10	4.0	20	5.0	5.1639	5.2444	5.2515
x - y	1	1.0	10	4.0	20	5.0	4.9477	5.0358	5.0410
	2	1.0	10	4.0	20	5.0	4.9186	4.9937	4.9885
	3	1.0	10	4.0	20	5.0	5.0786	5.2079	5.2028
	4	1.0	10	4.0	20	5.0	5.1639	5.0834	5.0782
(x)(y)	1	1.0	10	4.0	20	800.0	792.81	773.27	768.63
	2	1.0	10	4.0	20	800.0	794.33	779.29	797.48
	3	1.0	10	4.0	20	800.0	802.28	776.41	775.78
	4	1.0	10	4.0	20	800.0	867.67	883.85	883.38
x/y	1	1.0	10	4.0	20	0.005	0.0050	0.0051	0.0054
	2	1.0	10	4.0	20	0.005	0.0050	0.0051	0.0054
	3	1.0	10	4.0	20	0.005	0.0050	0.0052	0.0055
	4	1.0	10	4.0	20	0.005	0.0054	0.0053	0.0057

numbers were drawn. Column (4) lists the method of partials estimates of variance for the function based on the theoretical input (column 3). Column (5) lists the estimates of variance for the function calculated using the method of partial derivatives from the observed variation of the variables x and y. Column (6) gives column (5) corrected for the observed correlation between the pairs of x, y input values. The correction factor is:

$$2\rho\sigma_x^2\sigma_y^2 \frac{\partial Z}{\partial x}\frac{\partial Z}{\partial y}$$

where ρ is the observed correlation between paired values of x and y, $\sigma_x{}^2$ and $\sigma_y{}^2$ are the observed variances of x and y, and $\frac{\partial Z}{\partial x}$ and $\frac{\partial Z}{\partial y}$ are the partial derivatives of the function Z. Column (7) lists the simulator results for the function (column 1) for 1000 data points.

Table B-2 Results of Monte Carlo Simulation for Theoretical Input μ_{x_i}, $\sigma_{x_i}^2$

(1) Function Z	(2) Number of Simulations	(3) Theoretical Input		(4) Estimated Parameters (Method of Partials)		(5) Simulation Results	
		μ_{x_i}	$\sigma_{x_i}^2$	μ_z	σ_z^2	μ_z	σ_z^2
$(x_1x_2)/x_3$	2	20	1.0	20	3.00	20.2 20.6	2.56 3.24
$(x_1x_2)/(x_3x_4x_5)$	1	20	1.0	0.05	3.12×10^{-5}	0.0505	3.6×10^{-5}
$(x_1x_2x_3x_4)/(x_5x_6x_7)$	2	20	1.0	20	7.00	20.04 20.25	8.41 8.41
$(x_1x_2x_3)\Big/\left(\prod_{i=4}^{9} x_i\right)$	1	20	1.0	1.25×10^{-4}	3.52×10^{-10}	1.29×10^{-4}	4.0×10^{-10}
$\left(\prod_{i=1}^{6} x_i\right)\Big/(x_7x_8x_9)$	2	20	1.0	8000	1.44×10^{6}	8150 8300	1.69×10^{6} 1.82×10^{6}

Columns (1) through (3) of Table B-2 present the input to the Monte Carlo Simulator. The theoretical imput column (3) shows the parameters of the population of random numbers that were used to produce the functional values. Column (5) summarizes the results of the simulation. These results may be compared with the estimates from the method of partials, column (4).

Simulation results have shown that the method of partial derivatives is most accurate for functions involving sums and differences of the observed variables. For these functions, if the variables are mutually independent, the Taylor's series is exact for any magnitude of error in the measured parameters. If the variables are not mutually independent, a correction factor can be computed that will ensure exactitude of the method. (The correction factor $(2\rho_{xy}\sigma_x\sigma_y\frac{\partial Z}{\partial x}\cdot\frac{\partial Z}{\partial y})$ is the third term in Eq. (B-12). If ρ_{xy} is not zero, this term should be included in estimating $\sigma_Z{}^2$. From data, ρ_{xy} may be estimated with

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\Sigma(x_i-\bar{x})\ (y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\ \Sigma(y_i-\bar{y})^2}}$$

where n pairs of observations are available and $\bar{x}$ and $\bar{y}$ are the average of the x_i and y_i values, respectively.)

Close approximations can be made for errors that exist in functions involving products and quotients of independently varying observed values if the ratio of measured errors to their respective nominal values is small (less than 0.1). The approximation improves as measured errors decrease in relation to their nominals. For all of the functions examined involving two or more independent variables, the approximation is within 10 percent of the true error. The simulation results are summarized in Tables B-1 and B-2.

Table B-3 shows the Taylor's formula for several functions. In addition, the Taylor's formula for the coefficient of variation is also listed. The coefficient of variation is easily converted to a percentage variation by multiplying by 100.

Table B-3 Error Propagation Formulas

Function	Taylor's Formula	Coefficient of Variation Formula
$w = f(x,y)$	$S^2_w \approx \left(\frac{\partial w}{\partial x} S_x\right)^2 + \left(\frac{\partial w}{\partial y} S_y\right)^2$	- - - - - - - - - - -
$w = Ax + By$	$S^2_w \approx A^2 S^2_x + B^2 S^2_y$	$V^2_w \approx \frac{A^2 x^2 V^2_x + B^2 y^2 V^2_y}{(Ax + By)^2}$
$w = \frac{1}{y}$	$S^2_w \approx \frac{S^2_y}{y^4}$	$V^2_w \approx V^2_y$
$w = \frac{x}{x + y}$	$S^2_w \approx \left(\frac{y\,S_x}{(x+y)^2}\right)^2 + \left(\frac{x\,S_y}{(x+y)^2}\right)^2$	$V^2_w \approx y^2 (V^2_x + V^2_y) / (x+y)^2$
$w = \frac{x}{1 + x}$	$S^2_w \approx \frac{S_x^2}{(1 + x)^4}$	$V^2_w \approx \frac{V^2_x}{(1+x)^2}$
$w = xy$	$S^2_w \approx (yS_x)^2 + (xS_y)^2$	$V^2_w \approx V^2_x + V^2_y$
$w = x^2$	$S^2_w \approx 4x^2 S_x^2$	$V^2_w \approx 4V^2_x$
$w = x^{1/2}$	$S^2_w \approx \frac{S_x^2}{4x}$	$V^2_w \approx \frac{V_x^2}{4}$
$w = \ln x$	$S^2_w \approx \frac{S_x^2}{x^2}$	$V^2_w \approx \left(\frac{V_x}{\ln x}\right)^2$
$w = kx^a y^b$	$S^2_w \approx \left(aky^b x^{a-1} S_x\right)^2 + \left(bkx^a y^{b-1} S_y\right)^2$	$V_w^2 \approx \left(aV_x\right)^2 + \left(bV_y\right)^2$

where:

$$V_x = \frac{S_x}{\bar{x}}$$

$$V_y = \frac{S_y}{\bar{y}}$$

$$V_w = \frac{S_w}{\bar{w}} \; ; \quad \bar{w} = f(\bar{x},\bar{y})$$

APPENDIX C
ESTIMATES OF THE PRECISION INDEX FROM MULTIPLE MEASUREMENTS

INTRODUCTION

The derivatives, proofs, and examples of this section are presented for measurement schemes using two instruments measuring the same parameter. These instruments are read simultaneously and the difference between the readings is analyzed statistically to estimate instrumentation precision index. The proofs and derivations can easily be extended to instrumentation setups using more than two instruments by considering all independent combinations of instrument pairs. Instrumentation pairs which have a fixed constant locational bias between them can also be analyzed to provide estimates of precision index.

DISCUSSION

In any measurement system, if the parameter which is being measured is a constant, the precision error of the measurement instruments is easy to identify. The variance (S^2) of the measurements (X_i) about the average measurement ($\bar{X}$) provides an unbiased estimate of the variance of the instrument:

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{N - 1} \quad \text{(for N measurements)} \tag{C-1}$$

In the usual case, the measured parameter is not constant. Then the variation of the readings is increased by the variation of the parameter, and any directly computed statistic is subjected to large errors. However, multiple instruments can be analyzed to estimate the precision error in the measurement system.

For example, if two instruments are measuring a parameter X, then an appropriate mathematical model for each reading would be $X_i + \epsilon_{ij}$ where X_i is a typical parameter value and ϵ_{ij} is the corresponding precision error for the jth instrument. (For simplicity, bias error is ignored as it has no effect on these methods.) A series of such readings is illustrated in Table C-1.

Table C-1 Multiple Measurements of a Parameter

True Value	Instrument One	Instrument Two
X_1	$X_1 + \epsilon_{11}$	$X_1 + \epsilon_{12}$
X_2	$X_2 + \epsilon_{21}$	$X_2 + \epsilon_{22}$
X_3	$X_3 + \epsilon_{31}$	$X_3 + \epsilon_{32}$
⋮	⋮	⋮
X_n	$X_n + \epsilon_{n1}$	$X_n + \epsilon_{n2}$

This list is provided by making simultaneous readings of instruments one and two. By subtracting the reading of instrument one from instrument two, a column of differences can be produced which is independent of the variation of the parameter (Table C-2).

Table C-2 Multiple Measurement Difference

Instrument One	Instrument Two	Average	Difference
$X_1 + \epsilon_{11}$	$X_1 + \epsilon_{12}$	$X_1 + \frac{\epsilon_{11} + \epsilon_{12}}{2}$	$\epsilon_{11} - \epsilon_{12} = \Delta_1$
$X_2 + \epsilon_{21}$	$X_2 + \epsilon_{22}$	$X_2 + \frac{\epsilon_{21} + \epsilon_{22}}{2}$	$\epsilon_{21} - \epsilon_{22} = \Delta_2$
$X_3 + \epsilon_{31}$	$X_3 + \epsilon_{32}$	$X_3 + \frac{\epsilon_{31} + \epsilon_{32}}{2}$	$\epsilon_{31} - \epsilon_{32} = \Delta_3$
.	.	.	.
.	.	.	.
.	.	.	.
$X_n + \epsilon_{n1}$	$X_n + \epsilon_{n2}$	$X_n + \frac{\epsilon_{n1} + \epsilon_{n2}}{2}$	$\epsilon_{n1} - \epsilon_{n2} = \Delta_n$

The variation of these differences (S_Δ) provides an unbiased estimate of the precision error of the average reading of the two instruments. That is, S_Δ can be used to estimate S_{reading} based on Eq. (C-1).

$$S_\Delta = \sqrt{\frac{\sum_{i=1}^{n} (\Delta_i - \bar{\Delta})^2}{n - 1}} \qquad \text{(C-2)}$$

where $\bar{\Delta}$ is the average difference between the meters and n is the number of differences. It can be proved that the estimate of the precision index of the average reading is then

$$S_\Delta/2 = \sqrt{\frac{\sum_{i=1}^{n} (\Delta_i - \bar{\Delta})^2}{4(n - 1)}} \qquad \text{(C-3)}$$

This is based on the assumption that the two meters have independent precision errors. The formulation is derived from Taylor's Series expansion in Appendix B, for precision errors in calculated values. For example, if a single pair of multiple readings is made, the error is $(\epsilon_{i1} + \epsilon_{i2})/2$ if the average of the two readings is recorded. The precision index of this average value may be estimated using the Taylor's series expansion:

$$s_{\left(\frac{\epsilon_{i1} + \epsilon_{i2}}{2}\right)} \approx \sqrt{\left(\frac{\partial\left(\frac{\epsilon_{i1} + \epsilon_{i2}}{2}\right)}{\partial\epsilon_{i1}} s_{\epsilon_{i1}}\right)^2 + \left(\frac{\partial\left(\frac{\epsilon_{i1} + \epsilon_{i2}}{2}\right)}{\partial\epsilon_{i2}} s_{\epsilon_{i2}}\right)^2} \tag{C-4}$$

Thus

$$s_{\left(\frac{\epsilon_{i1} + \epsilon_{i2}}{2}\right)} \approx \sqrt{\left(\frac{1}{2} s_{\epsilon_{i1}}\right)^2 + \left(\frac{1}{2} s_{\epsilon_{i2}}\right)^2} \tag{C-5}$$

and

$$s_{\left(\frac{\epsilon_{i1} + \epsilon_{i2}}{2}\right)} \approx \sqrt{\frac{1}{4}\left(s^2_{\epsilon_{i1}} + s^2_{\epsilon_{i2}}\right)} = \frac{1}{2}\sqrt{s_{\epsilon_{i1}}{}^2 + s_{\epsilon_{i2}}{}^2} \tag{C-6}$$

In the same manner, the precision index of $\epsilon_{i1} - \epsilon_{i2}$ (Eq. (C-2)) can be shown to be:

$$s_{\left(\epsilon_{i1} - \epsilon_{i2}\right)} \approx \sqrt{\left(\frac{\partial\left(\epsilon_{i1} - \epsilon_{i2}\right)}{\partial\epsilon_{i1}} s_{\epsilon_{i1}}\right)^2 + \left(\frac{\partial\left(\epsilon_{i1} - \epsilon_{i2}\right)}{\partial\epsilon_{i2}} s_{\epsilon_{i2}}\right)^2} \tag{C-7}$$

or

$$s_{\left(\epsilon_{i1} - \epsilon_{i2}\right)} \approx \sqrt{(1)^2\, s^2_{\epsilon_{i1}} + (-1)^2\, s^2_{\epsilon_{i2}}} \tag{C-8}$$

By combining Eqs. (C-2), (C-6), and (C-8), the precision error estimate for the average reading of Table C-2 is

$$\begin{aligned} s_{\left(\frac{\epsilon_{i1} + \epsilon_{i2}}{2}\right)} &= \frac{1}{2}\sqrt{s^2_{\epsilon_{i1}} + s^2_{\epsilon_{i2}}} \approx \frac{1}{2}\, s_{\left(\epsilon_{i1} - \epsilon_{i2}\right)} = \frac{1}{2} s_\Delta \\ &= \sqrt{\frac{\sum\limits_{i=1}^{n} (\Delta_i - \bar{\Delta})^2}{4(n-1)}} \end{aligned} \tag{C-9}$$

The method described in the preceding discussion does not include long-term drifts, calibration errors, and other precision errors that do not vary from data point-to-data point, from stand-to-stand, and from calibration-to-calibration. A new mathematical model must be developed based on an engineering analysis including these effects.

To illustrate the method, the mathematical model derived for a turbojet engine will be used. Again, X is the parameter being measured; then

$$\text{Reading}_{\ell} = X_{ij\ell} + \alpha_i + \beta_{ij} + \epsilon_{ij\ell}$$

is the mathematical model of reading ℓ of a typical instrument during run j and since calibration i. The model assumes a within test (short-term) component of precision error ($\epsilon_{ij\ell}$), a test-to-test (long-term) component or precision error (β_{ij}) which is constant within a test and a calibration-to-calibration error (α_i) which is constant within each calibration and within test within each calibration.

Table C-3 illustrates a series of readings of a multiple instrumented parameter, assuming this model and the differences of the simultaneous readings of instruments one and two. Notice that, within each run, the run-to-run and calibration-to-calibration components are constant. Between the averages of each run, the within run components are reduced and the calibration-to-calibration components are constant allowing estimates of run-to-run components. Finally, the between two calibrations, the within run, and run-to-run components are reduced allowing an estimate of calibration-to-calibration error. The method of analysis is that of a one-way nested analysis of variance.

The analysis can be completed using the formulas of Table C-4. The last column of this table provides the estimates of the precision index for calibration-to-calibration error, run-to-run error, and within run error. The third column, mean square, is calculated by dividing the value of the sum of squares (column 1) by the proper degrees of freedom (column 2).

The estimate of the total precision index, for a multiple instrumented parameter (average of two measurements) on a particular run and calibration, would be the root-sum-square of the three estimates: within test precision index (S_{wt}), test-to-test precision index (S_{tt}), and calibration-to-calibration precision index (S_{cc}); that is,

$$S_T^2 = S_{wt}^2 + S_{tt}^2 + S_{cc}^2$$

The degrees of freedom for this estimate can be calculated from the degrees of freedom for each estimate using the Welch-Satterthwaite formula.

Table C-3 Differences in Readings of Multiple Instruments

		1ST INSTRUMENT	2ND INSTRUMENT	DIFFERENCES	DELTAS
Cal 1	1st Run	$X_{111} + \alpha_{11} + \beta_{111} + \epsilon_{1111}$	$X_{111} + \alpha_{12} + \beta_{112} + \epsilon_{1112}$	$\alpha_{11} + \beta_{111} + \epsilon_{1111} - \left[\alpha_{12} + \beta_{112} + \epsilon_{1112}\right]$	$= \Delta_{111}$
		$X_{112} + \alpha_{11} + \beta_{111} + \epsilon_{1121}$	$X_{112} + \alpha_{12} + \beta_{112} + \epsilon_{1122}.$	$\alpha_{11} + \beta_{111} + \epsilon_{1121} - \left[\alpha_{12} + \beta_{112} + \epsilon_{1122}\right]$	$= \Delta_{112}$
		. . . .	. . . .	. . . .	. . . .
		$X_{11n} + \alpha_{11} + \beta_{111} + \epsilon_{11n1}$	$X_{11n} + \alpha_{12} + \beta_{112} + \epsilon_{11n2}$	$\alpha_{11} + \beta_{111} + \epsilon_{11n1} - \left[\alpha_{12} + \beta_{112} + \epsilon_{11n2}\right]$	$= \Delta_{11n}$
	2nd Run	$X_{121} + \alpha_{11} + \beta_{121} + \epsilon_{2111}$	$X_{121} + \alpha_{12} + \beta_{122} + \epsilon_{2112}$	$\alpha_{11} + \beta_{121} + \epsilon_{2111} - \left[\alpha_{12} + \beta_{122} + \epsilon_{2112}\right]$	$= \Delta_{211}$
		$X_{122} + \alpha_{11} + \beta_{121} + \epsilon_{2121}$	$X_{122} + \alpha_{12} + \beta_{122} + \epsilon_{2122}$	$\alpha_{11} + \beta_{121} + \epsilon_{2121} - \left[\alpha_{12} + \beta_{122} + \epsilon_{2122}\right]$	$= \Delta_{212}$
		. . . .	. . . .	. . . .	. . . .
		$X_{12n} + \alpha_{11} + \beta_{121} + \epsilon_{21n1}$	$X_{12n} + \alpha_{12} + \beta_{122} + \epsilon_{21n2}$	$\alpha_{11} + \beta_{121} + \epsilon_{21n1} - \left[\alpha_{12} + \beta_{122} + \epsilon_{21n2}\right]$	$= \Delta_{21n}$
	Kth Run	$X_{1K1} + \alpha_{11} + \beta_{1K1} + \epsilon_{K111}$	$X_{1K1} + \alpha_{12} + \beta_{1K2} + \epsilon_{K112}$	$\alpha_{11} + \beta_{1K1} + \epsilon_{K111} - \left[\alpha_{12} + \beta_{1K2} + \epsilon_{K112}\right]$	$= \Delta_{K11}$
		$X_{1K2} + \alpha_{11} + \beta_{1K1} + \epsilon_{K121}$	$X_{1K2} + \alpha_{12} + \beta_{1K2} + \epsilon_{K122}$	$\alpha_{11} + \beta_{1K1} + \epsilon_{K121} - \left[\alpha_{12} + \beta_{1K2} + \epsilon_{K122}\right]$	$= \Delta_{K12}$
		. . . .	. . . .	. . . .	. . . .
		$X_{1Kn} + \alpha_{11} + \beta_{1K1} + \epsilon_{K1n1}$	$X_{1Kn} + \alpha_{12} + \beta_{1K2} + \epsilon_{K1n2}$	$\alpha_{11} + \beta_{1K1} + \epsilon_{K1n1} - \left[\alpha_{12} + \beta_{1K2} + \epsilon_{K1n2}\right]$	$= \Delta_{K1n}$
Cal 2	1st Run	$X_{211} + \alpha_{21} + \beta_{211} + \epsilon_{1211}$	$X_{211} + \alpha_{22} + \beta_{212} + \epsilon_{1212}$	$\alpha_{21} + \beta_{211} + \epsilon_{1211} - \left[\alpha_{22} + \beta_{212} + \epsilon_{1212}\right]$	$= \Delta_{121}$
		$X_{212} + \alpha_{21} + \beta_{211} + \epsilon_{1221}$	$X_{212} + \alpha_{22} + \beta_{212} + \epsilon_{1222}$	$\alpha_{21} + \beta_{211} + \epsilon_{1221} - \left[\alpha_{22} + \beta_{212} + \epsilon_{1222}\right]$	$= \Delta_{122}$
		. . . .	. . . .	. . . .	. . . .
		$X_{21n} + \alpha_{21} + \beta_{211} + \epsilon_{12n1}$	$X_{21n} + \alpha_{22} + \beta_{212} + \epsilon_{12n2}$	$\alpha_{21} + \beta_{211} + \epsilon_{12n1} - \left[\alpha_{22} + \beta_{212} + \epsilon_{12n2}\right]$	$= \Delta_{12n}$
	2nd Run	$X_{221} + \alpha_{21} + \beta_{221} + \epsilon_{2211}$	$X_{221} + \alpha_{22} + \beta_{222} + \epsilon_{2212}$	$\alpha_{21} + \beta_{221} + \epsilon_{2211} - \left[\alpha_{22} + \beta_{222} + \epsilon_{2212}\right]$	$= \Delta_{221}$
		$X_{222} + \alpha_{21} + \beta_{221} + \epsilon_{2221}$	$X_{222} + \alpha_{22} + \beta_{222} + \epsilon_{2222}$	$\alpha_{21} + \beta_{221} + \epsilon_{2221} - \left[\alpha_{22} + \beta_{222} + \epsilon_{2222}\right]$	$= \Delta_{222}$
		. . . .	. . . .	. . . .	. . . .
		$X_{22n} + \alpha_{21} + \beta_{221} + \epsilon_{22n1}$	$X_{22n} + \alpha_{22} + \beta_{222} + \epsilon_{22n2}$	$\alpha_{21} + \beta_{221} + \epsilon_{22n1} - \left[\alpha_{22} + \beta_{222} + \epsilon_{22n2}\right]$	$= \Delta_{22n}$
	Kth Run	$X_{2K1} + \alpha_{21} + \beta_{2K1} + \epsilon_{K211}$	$X_{2K1} + \alpha_{22} + \beta_{2K2} + \epsilon_{K212}$	$\alpha_{21} + \beta_{2K1} + \epsilon_{K211} - \left[\alpha_{22} + \beta_{2K2} + \epsilon_{K212}\right]$	$= \Delta_{K21}$
		$X_{2K2} + \alpha_{21} + \beta_{2K1} + \epsilon_{K221}$	$X_{2K2} + \alpha_{22} + \beta_{2K2} + \epsilon_{K222}$	$\alpha_{21} + \beta_{2K1} + \epsilon_{K221} - \left[\alpha_{22} + \beta_{2K2} + \epsilon_{K222}\right]$	$= \Delta_{K22}$
		. . . .	. . . .	. . . .	. . . .
		$X_{2Kn} + \alpha_{21} + \beta_{2K1} + \epsilon_{K2n1}$	$X_{2Kn} + \alpha_{22} + \beta_{2K2} + \epsilon_{K2n2}$	$\alpha_{21} + \beta_{2K1} + \epsilon_{K2n1} - \left[\alpha_{22} + \beta_{2K2} + \epsilon_{K2n2}\right]$	$= \Delta_{K2n}$

Table C-4 Analysis of Precision Error from Multiple Instrument Differences

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Variance Estimates
Between Cals	$\sum_{i=1}^{M} (T_i^2/N_i) - \frac{T^2}{N}$	M-1	$M_2 = \sigma_0^2 + \bar{N}_2\sigma_1^2 + \bar{N}_3\sigma_2^2$	$\sigma^2_{\substack{Cal\\Cal}} = \sigma_2^2 = \frac{M_2 - \sigma_0^2 - N_2\sigma_1^2}{\bar{N}_3}$
Between Runs	$\sum_{i=1}^{M} \sum_{j=1}^{K_i} (T_{ij}^2/N_{ij}) - \sum_{i=1}^{M} (T_i^2/N_i)$	R-M	$M_1 = \sigma_0^2 + \bar{N}_1\sigma_1^2$	$\sigma^2_{\substack{Run\\Run}} = \sigma_1^2 = \frac{M_1 - \sigma_0^2}{\bar{N}_1}$
Within Runs	$\sum_{i=1}^{M} \sum_{j=1}^{K_i} \sum_{l=1}^{N_{ij}} \Delta_{ijl}^2 - \sum_{i=1}^{M} \sum_{j=1}^{K_i} (T_{ij}^2/N_{ij})$	N-R	$M_0 = \sigma_0^2$	$\sigma^2_{Within} = \sigma_0^2$
Total	$\sum_{i=1}^{M} \sum_{j=1}^{K_i} \sum_{l=1}^{N_{ij}} \Delta_{ijl}^2 - \frac{T^2}{N}$	N-1		

$$T = \sum_{i=1}^{M} T_i = \sum_{i=1}^{M} \sum_{j=1}^{K_i} T_{ij} = \sum_{i=1}^{M} \sum_{j=1}^{K_i} \sum_{l=1}^{N_{ij}} \Delta_{ijl}$$

$$N = \sum_{i=1}^{M} N_i = \sum_{i=1}^{M} \sum_{j=1}^{K_i} N_{ij} \qquad R = \sum_{i=1}^{M} K_i$$

$$\bar{N}_1 = \left[N - \sum_{i=1}^{M} \cdot \left(\sum_{j=1}^{K_i} N_{ij}^2/N_i\right)\right] / (R-M)$$

$$\bar{N}_2 = \left[\sum_{i=1}^{M} \left(\sum_{j=1}^{N_i} N_{ij}^2/N_i\right) - \sum_{i=1}^{M} \sum_{j=1}^{K_i} N_{ij}^2/N\right] / (M-1)$$

$$\bar{N}_3 = \left(N^2 - \sum_{i=1}^{M} N_i^2\right) \Big/ \left((N-1)N\right)$$

APPENDIX D
OUTLIER DETECTION

GENERAL

All measurement systems may produce wild data points. These points may be caused by temporary or intermittent malfunctions of the measurement system, or they may represent actual variations in the measurement. Errors of this type cannot be estimated as part of the uncertainty of the measurement. The points are out of control points for the system and are meaningless as steady-state test data. They should be discarded. Figure D-1 shows two spurious data points (sometimes called outliers).

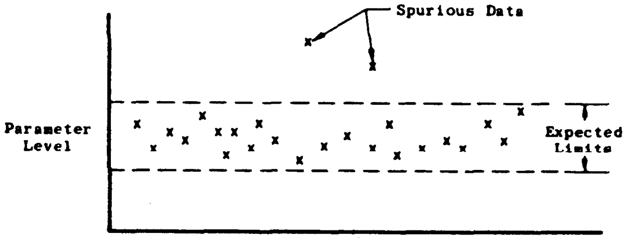


Fig. D-1 Outliers Outside the Range of Acceptable Data

All data should be inspected for wild data points as a continuing quality control check on the measurement process. Identification criteria should be based on engineering analysis of instrumentation, thermodynamics, flow profiles, and past history with similar data. To ease the burden of scanning large masses of data, a computerized routine is available to scan steady-state data and flag suspected outliers. The flagged points should then be subjected to a comprehensive engineering analysis.

This routine is intended to be used in scanning small samples of data from a large number of parameters at many time slices. The work of paging through volumes of data can be reduced to a manageable job with this approach. The computer will scan the data and flag suspect points. The engineer, relieved of the burden of scanning the data, can closely examine each suspected wild point.

ARNOLD ENGINEERING DEVELOPMENT CENTER OUTLIER METHOD

Several general purpose outlier techniques were reviewed and discussed in the ICRPG Handbook CPIA 180. The U.S. Air Force Arnold Engineering Development Center (AEDC) has developed a new technique to flag outliers in small or moderately sized samples of data. This technique was compared with the Thompson's Tau Technique used in ICRPG CPIA 180. The AEDC method as compared with the Thompson's Tau method detects a larger proportion of the outliers in the data, and when no outliers are present, it flags fewer good points. The AEDC method is useful for computer routines since it is fast and requires little core storage. The method discriminates between good data and outliers by examining how far each point lies from the average value.

The first step is to calculate an average value ($\overline{X}$) and a standard deviation (s) from the data.

$$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N} \qquad s = \pm\sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}}$$

Then, from the sample size (N), a test value (C) is calculated from

$$C = \frac{-1.6819236 + 1.6386898N - 0.00721312N^2}{1.0 + 0.59286772N - 0.00355709N^2}$$

when $N < 65$. When $N \geqslant 65$, $C = 3$ is used. Each data point is tested to determine if it falls in the interval, average value plus or minus the standard deviation times C, i.e.,

$$\bar{X} \pm Cs$$

If a data point falls outside the interval, it is flagged as an outlier.

The formula for calculating C was determined at AEDC and was based on engineering judgment to determine the expected intervals for good data in the range of 10 to 30 samples. Then the curve (C) was fit to the data (Fig. D-2). For sample sizes of 65 or greater, C = 3 should be used.

AN EXAMPLE OF THE AEDC METHOD

The AEDC Fortran Subroutine is used to test the data in Table D-1 for outliers.

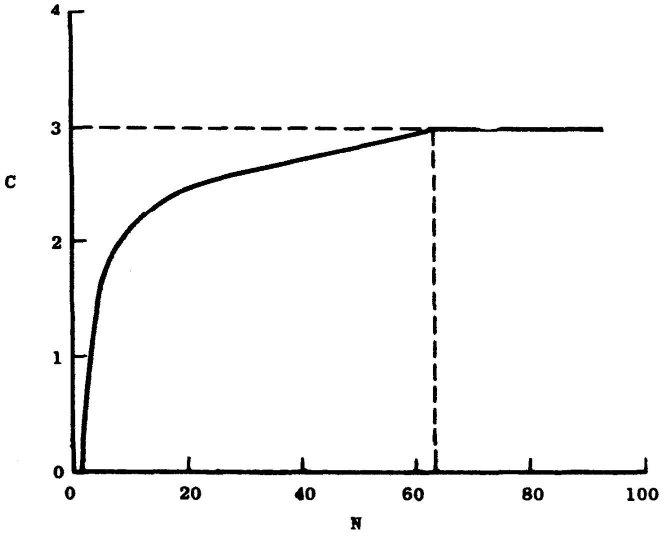


Fig. D-2 Parameter C versus Sample Size

Table D-1
Pressure Data
for Outlier
Test

Point No.	Pressure, psia
1	12.96
2	13.15
3	13.01
4	13.11
5	13.30
6	13.68
7	13.26
8	13.10
9	12.84
10	13.19
11	13.25
12	13.39
13	13.11
14	13.03
15	12.96

The first step is to calculate the mean ($\bar{X}$) and standard deviation (s) for the data:

$$\bar{X} = \frac{\sum_{i=1}^{15} X_i}{15} = 13.156$$

$$s = \pm\sqrt{\frac{\sum_{i=1}^{15} (X_i - \bar{X})^2}{14}} = \pm 0.2057$$

The test value (C) is calculated based on the sample size (N).

$$C = \frac{-1.6819236 + 1.6386898N - 0.007213l2N^2}{1.0 + 0.59286772N - 0.00355709N^2} = 2.3398$$

Each data point is checked to determine if it falls in the interval $\bar{X} \pm Cs$, which in this case is 12.6747 to 13.6373. A convenient method for doing this is to subtract $\bar{X}$ from each data point to determine the deviation $(X_i - \bar{X})$. Then, this is checked against Cs = 2.3398 times 0.2057 = 0.4813.

Table D-2 lists the deviations $(X_i - \bar{X})$ for the data. In this case, point number six has a deviation which is greater than 0.4813. It is flagged as an outlier and printed out by the subroutine.

Table D-2

Point No.	Deviation $(X_i - \bar{X})$
1	-0.196
2	-0.006
3	-0.146
4	-0.046
5	0.144
6	0.524
7	0.104
8	-0.056
9	-0.316
10	0.034
11	0.094
12	0.234
13	-0.046
14	-0.126
15	-0.196

THE FORTRAN SUBROUTINE

The following is a listing of a subroutine to implement the AEDC outlier rejection method in a computer program. The subroutine is written in American National Standard Fortran IV. The subroutine was compiled and run on an IBM 360, model 75, with a G-level compiler operating under IBM release 20.1.

```
      SUBROUTINE CAVG ( XBAR, SIG , X , N , IELIM )
C
C                 AEDC OUTLIER REJECTION SUBROUTINE
C
C
C     XBAR        MEAN CALCULATED AFTER OUTLIERS HAVE BEEN REMOVED FROM
C                 DATA
C
C     X           INPUT ARRAY OF SAMPLE DATA
C
C     SIG         STANDARD DEVIATION CALCULATED AFTER OUTLIERS HAVE BEEN
C                 REMOVED FROM THE DATA
C
C     N           NUMBER OF DATA POINTS IN DATA SAMPLE, X
C
C     IELIM       NUMBER OF OUTLIERS REJECTED IN SAMPLE OF DATA
C
C
```

```
C      IF A SAMPLE DATA POINT IS FOUND TO BE AN OUTLIER, ITS VALUE IS
C      SET TO 0.0 .
C
C      SAMPLE DATA POINTS EQUAL TO ZERO INPUT TO THE SUBROUTINE ARE
C      DISCARDED FROM THE SAMPLE.
C
C
       DIMENSION X(1) , DEL(100)
       IND = 0
       IELIM = 0
       WRITE(6,102)
   102 FORMAT(/' BEFORE REMOVING OUTLIERS'/)
     1 SUM = 0.
       NN = 0
       DO 3 I=1,N
       IF ( X(I) .EQ. 0. ) GO TO 3
       NN = NN + 1
       SUM = SUM + X(I)
     3 CONTINUE
       IF ( NN .LT. 2 ) RETURN
       XBAR = SUM / NN
       SUM2 = 0.
       DO 4 I = 1,N
       IF ( X(I) .EQ. 0. ) GO TO 4
       DEL(I) = ABS( X(I) - XBAR )
       SUM2 = SUM2 + DEL(I)**2
     4 CONTINUE
       C = 3.
       SIG = SQRT ( SUM2 / ( NN - 1 ))
       WRITE(6,101)XBAR,SIG,NN
   101 FORMAT(/' MEAN=',E20.7,5X,'STD. DEV.=',E20.7,5X,'N=',I5/)
       IF ( IND .EQ. 1 ) RETURN
       IF ( NN .LE. 65 ) C = (-1.6819236D0 + 1.6386898D0 * NN
      1                      -.00721312D0 * NN * NN ) / ( 1.D0
      2                      +.59286772D0 * NN - .00355709D0 * NN * NN )
       CS = C * SIG
       K = NN
       DO 5 I = 1,N
       IF ( X(I) .EQ. 0. .OR. DEL(I) .LE. CS ) GO TO 5
       NN = NN - 1
       WRITE(6,100)I,X(I)
   100 FORMAT(/5X,'POINT NO.',I5,' VALUE=',E20.7,' HAS BEEN REJECTED'/)
       X(I) = 0.
     5 CONTINUE
       IELIM = K - NN
       IND = 1
       IF ( K .EQ. NN) RETURN
       WRITE(6,103)
   103 FORMAT(/' AFTER REMOVING OUTLIERS'/)
       GO TO 1
   999 CONTINUE
       END
```

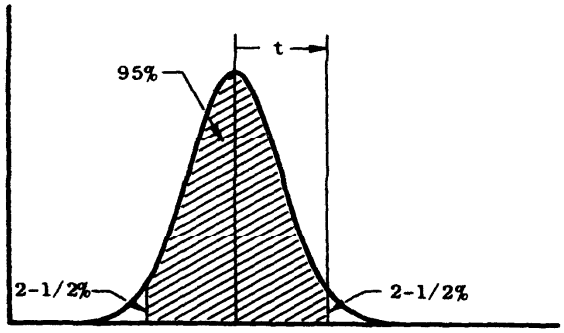
APPENDIX E
TABLES

This section of the Appendix presents the tables of the Student's "t" distribution, the F table for comparison of precision indices, and the Thompson's Tau table for the outlier test.

STUDENT'S "t" TABLE

The table of Student's "t" distribution (Table E-1) presents the two-tailed 95-percent "t" values for the degrees of freedom from one to 30. Above 30, round the value to 2.0.

Table E-1 Two-Tailed Student's "t" Table



Degrees of Freedom	"t"	Degrees of Freedom	"t"
1	12.706	17	2.110
2	4.303	18	2.101
3	3.182	19	2.093
4	2.776	20	2.086
5	2.571	21	2.080
6	2.447	22	2.074
7	2.365	23	2.069
8	2.306	24	2.064
9	2.262	25	2.060
10	2.228	26	2.056
11	2.201	27	2.052
12	2.179	28	2.048
13	2.160	29	2.045
14	2.145	30	2.042
15	2.131		
16	2.120	31 or more use 2.0	

The table is used to provide an interval estimate of the true value about an observed value. The interval is the measurement plus and minus the standard deviation of the observed value times the "t" value (for the degrees of freedom of that standard deviation):

$$\text{interval} = \text{measurement} \pm t_{95}S$$

The 95-percent Student's "t" value for a standard deviation of 50 lb with 17 deg of freedom is 2.110. The interval is

$$\text{measurement} \pm 2.11 \times 50 = \text{measurement} \pm 105.50 \text{ lb}$$

F TABLE

The table of 95-percent F values (Table E-2) is presented for tests for a significant increase in precision index. The test is performed by dividing the square of the new precision index by the square of the old index:

$$F_{\text{calculated}} = \frac{S_{\text{new}}^2}{S_{\text{old}}^2}$$

This calculated value is compared with the F table value for f_1 and f_2 degrees of freedom; f_1 is the degrees of freedom for $S_{new}{}^2$, and f_2 is the degrees of freedom for $S_{old}{}^2$.

For example, suppose that the pooled precision index for a force measuring device is 0.05 percent based on four sets of 5 tests each (total of 16 degrees of freedom). A new estimate of precision index is 0.10 percent based on a sample size of 5. The F calculated value is:

$$F_{\text{calculated}} = \frac{S_{\text{new}}^2}{S_{\text{old}}^2} = \frac{(0.1)^2}{(0.05)^2} = \frac{0.0100}{0.0025} = 4.00$$

$f_1 = 4$ and $f_2 = 16$, the table value of F is 3.01. Because 4.00 is greater than 3.01, this indicates the new index is significantly larger than the old index.

THOMPSON'S TAU TABLE

Thompson's Tau Table (Table E-3) is presented to aid in detecting outliers or bad data in a sample. The table is given for several levels of significance (p). If $p = 0.05$ level is used, this sets the probability of a good point at 5 percent. The procedure for the test is this:

1. From the data (X_j; $i = 1,n$) containing the suspected outlier X', two statistics are calculated:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Table E-2 5-percent Fractiles of the F Distribution

$$F(f_1, f_2) = \frac{S^2_{new}}{S^2_{old}}$$

Degrees of Freedom for the Numerator (f_1)

Degrees of Freedom for the Denominator (f_2)	1	2	3	4	5	6	7	8	9	10	12	15	17
1	161	200	216	225	230	234	237	239	241	242	244	246	247
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.68
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.83
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.59
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.91
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.48
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.19
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.97
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.81
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.69
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.58
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.50
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.43
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.37
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.32
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.27
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.23
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.20
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.17
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.14
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.11
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.09
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.07
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.05
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	2.03
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	2.00
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.98
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.05	1.97	1.93
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.02	1.94	1.90
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.89
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	1.98	1.90	1.86
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.96	1.88	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.83
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.93	1.85	1.81
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.80
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.90	1.82	1.78
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.79	1.75
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.77	1.73
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.73	1.69
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.72	1.67
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.68	1.63
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.62

Table E-2 (Concluded)

$$F(f_1, f_2) = \frac{S^2_{new}}{S^2_{old}}$$

Degrees of Freedom for the Numerator (f_1)

19	20	22	24	26	30	35	40	50	100	500	∞	Degrees of Freedom for Denominator (f_2)
248	248	249	249	249	250	251	251	252	253	254	254	1
19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	2
8.67	8.66	8.65	8.64	8.63	8.62	8.60	8.59	8.58	8.55	8.53	8.53	3
5.81	5.80	5.79	5.77	5.76	5.75	5.73	5.72	5.70	5.66	5.64	5.63	4
4.57	4.56	4.54	4.53	4.52	4.50	4.48	4.46	4.44	4.41	4.37	4.37	5
3.88	3.87	3.86	3.84	3.83	3.81	3.79	3.77	3.75	3.71	3.68	3.67	6
3.46	3.44	3.43	3.41	3.40	3.38	3.36	3.34	3.32	3.27	3.24	3.23	7
3.16	3.15	3.13	3.12	3.10	3.08	3.06	3.04	3.02	2.97	2.94	2.93	8
2.95	2.94	2.92	2.90	2.89	2.86	2.84	2.83	2.80	2.76	2.72	2.71	9
2.78	2.77	2.75	2.74	2.72	2.70	2.68	2.66	2.64	2.59	2.55	2.54	10
2.66	2.65	2.63	2.61	2.59	2.57	2.55	2.53	2.51	2.46	2.42	2.40	11
2.56	2.54	2.52	2.51	2.49	2.47	2.44	2.43	2.40	2.35	2.31	2.30	12
2.47	2.46	2.44	2.42	2.41	2.38	2.36	2.34	2.31	2.26	2.22	2.21	13
2.40	2.39	2.37	2.35	2.33	2.31	2.28	2.27	2.24	2.19	2.14	2.13	14
2.34	2.33	2.31	2.29	2.27	2.25	2.22	2.20	2.18	2.12	2.08	2.07	15
2.29	2.28	2.25	2.24	2.22	2.19	2.17	2.15	2.12	2.07	2.02	2.01	16
2.24	2.23	2.21	2.19	2.17	2.15	2.12	2.10	2.08	2.02	1.97	1.96	17
2.20	2.19	2.17	2.15	2.13	2.11	2.08	2.06	2.04	1.98	1.93	1.92	18
2.17	2.16	2.13	2.11	2.10	2.07	2.05	2.03	2.00	1.94	1.89	1.88	19
2.14	2.12	2.10	2.08	2.07	2.04	2.01	1.99	1.97	1.91	1.86	1.84	20
2.11	2.10	2.07	2.05	2.04	2.01	1.98	1.96	1.94	1.88	1.82	1.81	21
2.08	2.07	2.05	2.03	2.01	1.98	1.96	1.94	1.91	1.85	1.80	1.78	22
2.06	2.05	2.02	2.00	1.99	1.96	1.93	1.91	1.88	1.82	1.77	1.76	23
2.04	2.03	2.00	1.98	1.97	1.94	1.91	1.89	1.86	1.80	1.75	1.73	24
2.02	2.01	1.98	1.96	1.95	1.92	1.89	1.87	1.84	1.78	1.73	1.71	25
2.00	1.99	1.97	1.95	1.93	1.90	1.87	1.85	1.82	1.76	1.71	1.69	26
1.97	1.96	1.93	1.91	1.90	1.87	1.84	1.82	1.79	1.73	1.67	1.65	28
1.95	1.93	1.91	1.89	1.87	1.84	1.81	1.79	1.76	1.70	1.64	1.62	30
1.90	1.89	1.86	1.84	1.82	1.80	1.77	1.75	1.71	1.65	1.59	1.57	34
1.87	1.85	1.83	1.81	1.79	1.76	1.73	1.71	1.68	1.61	1.54	1.53	38
1.85	1.84	1.81	1.79	1.77	1.74	1.72	1.69	1.66	1.59	1.53	1.51	40
1.83	1.81	1.79	1.77	1.75	1.72	1.69	1.67	1.63	1.56	1.49	1.48	44
1.81	1.79	1.77	1.75	1.73	1.70	1.67	1.64	1.61	1.54	1.47	1.45	48
1.80	1.78	1.76	1.74	1.72	1.69	1.66	1.63	1.60	1.52	1.46	1.44	50
1.78	1.76	1.74	1.72	1.70	1.67	1.64	1.61	1.58	1.50	1.43	1.41	55
1.76	1.75	1.72	1.70	1.68	1.65	1.62	1.59	1.56	1.48	1.41	1.39	60
1.75	1.73	1.71	1.69	1.67	1.63	1.60	1.58	1.54	1.46	1.39	1.37	65
1.72	1.70	1.68	1.65	1.63	1.60	1.57	1.54	1.51	1.43	1.35	1.32	80
1.69	1.68	1.65	1.63	1.61	1.57	1.54	1.52	1.48	1.39	1.31	1.28	100
1.66	1.64	1.61	1.59	1.57	1.53	1.50	1.48	1.44	1.34	1.25	1.22	150
1.64	1.62	1.60	1.57	1.55	1.52	1.48	1.46	1.41	1.32	1.22	1.19	200
1.60	1.58	1.55	1.53	1.51	1.47	1.44	1.41	1.36	1.26	1.13	1.08	1000
1.59	1.57	1.54	1.52	1.50	1.46	1.42	1.39	1.35	1.24	1.11	1.00	∞

Table E-3 Thompson's Tau

Sample Size	Level of Significance			
N	P= .1	.05	.02	.01
3	1.3968	1.4099	1.41352	1.414039
4	1.559	1.6080	1.6974	1.7147
5	1.611	1.757	1.869	1.9175
6	1.631	1.814	1.973	2.0509
7	1.640	1.848	2.040	2.142
8	1.644	1.870	2.087	2.207
9	1.647	1.885	2.121	2.256
10	1.648	1.895	2.146	2.294
11	1.648	1.904	2.166	2.324
12	1.649	1.910	2.183	2.348
13	1.649	1.915	2.196	2.368
14	1.649	1.919	2.207	2.385
15	1.649	1.923	2.216	2.399
16	1.649	1.926	2.224	2.411
17	1.649	1.928	2.231	2.422
18	1.649	1.931	2.237	2.432
19	1.649	1.932	2.242	2.440
20	1.649	1.934	2.247	2.447
21	1.649	1.936	2.251	2.454
22	1.649	1.937	2.255	2.460
23	1.649	1.938	2.259	2.465
24	1.649	1.940	2.262	2.470
25	1.649	1.941	2.264	2.475
26	1.648	1.942	2.267	2.479
27	1.648	1.942	2.269	2.483
28	1.648	1.943	2.272	2.487
29	1.648	1.944	2.274	2.490
30	1.648	1.944	2.275	2.493
31	1.648	1.945	2.277	2.495
32	1.648	1.945	2.279	2.498
∞	1.64485	1.95996	2.32634	2.57582

and

$$SD = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}}$$

These are the average value and standard deviation of the sample.

2. Calculate the difference in (absolute value) between the average value $\bar{X}$ and the outlier X′:

$$\delta = |\bar{X} - X'|$$

3. Entering the table for N and P, (take P = 0.05) read a value of Tau.

4. The comparison is made between δ and the product of SD and Tau. If δ is larger or equal to that product, the data point is declared an outlier. A new mean and standard deviation must be calculated. If δ is smaller, the point is not rejected.

Example:

In the following sample of 15 data points, $\bar{X}$ and S were calculated to be 9.949 and 0.997:

9.558	10.478	9.609	9.582	9.583
11.447	11.485	11.067	9.173	10.303
10.472	10.310	7.416	9.488	9.257
		↑——Suspected outlier		

The test is to compare δ (the absolute difference between the suspected outlier and the average value) with the product Tau value times calculated standard deviation.

$$\delta = |\bar{X} - X'| = |9.949 - 7.416| = 2.533$$

$$\text{Tau} \times SD = 1.923 \times 0.997 = 1.917$$

where Tau is the Table E-3 value for P = 0.05 and N =15. Since δ is greater, the point 7.416 is discarded.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1980 311-046/125