ERROR ANALYSIS

A.J. Pintar
Department of Chemical Engineering
Michigan Technological University
Houghton, MI 49931
September, 2001

OBJECTIVE

Generally, laboratory experimentation involves the measurement of “raw data” in the laboratory and then the use of these “raw data” to calculate other quantities afterwards. The objective of an error analysis is to determine the effect of errors in the raw data on the calculated results:

$$\textbf{RAW DATA}\left\{\begin{array}{c}\underline{\textbf{Instruments}}\\ \textbf{Gauges}\\ \textbf{Scales}\\ \textbf{Timers}\end{array}\right\} \Rightarrow \textbf{CALCULATED RESULTS}$$

ERRORS IN THE RAW DATA ⇒ ERRORS IN THE CALCULATED RESULTS

TYPES OF ERROR

1. Systematic Error:
   a. Has the same sign and magnitude for identical conditions.
   b. Sources: Miscalibration of Instrument
      Natural Phenomena, i.e. heat transfer in a thermowell or in a thermometer stem.
      Consistent Operator Error, i.e. parallax.
   c. Often can be removed or compensation made:
      Recalibration
      Correction Factors or Calibration Curves
      Improved procedures
      Comparison to other methods.

2. Random Error:
   a. May be positive or negative.
   b. Sources: Random Process Fluctuations and Upsets
      Instrument Sensitivity
      Degree of Subdivision of Instrument Scale
      Operator Error in Reading Instrument Scale
      Equipment “goblins”
      “Phase of the Moon”??
      Miscellaneous.
   c. Can be dealt with using statistics.

The discussion that follows deals only with Random Error.

RAW DATA - STATISTICAL ANALYSIS OF REPLICATED DATA

Suppose that there are N measurements of the quantity x, i.e. x_1, x_2 , x_3, x_4, ..., x_N

1. Mean Value (Measure of the True Value)

The mean value ($\bar{x}$) is defined statistically by:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} . \qquad \text{(Eq. 1)}$$

2. Variance

The variance (σ^2) is defined statistically by:

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{(N-1)} = \frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{(N-1)} . \qquad \text{(Eq. 2)}$$

3. Standard Deviation (Measure of the "Precision" or Scatter in Data)

The standard deviation (σ) is defined as the square root of the variance, i.e. the square root of the expression above (Eq. 2).

4. Standard Error (Measure of the deviation of x from the true value)

The standard error (e_S) is defined statistically by:

$$e_S = \frac{\sigma}{\sqrt{N}} . \qquad \text{(Eq. 3)}$$

It can be shown statistically that the true value of x lies somewhere between:

$\bar{x} - e_S$ and $\bar{x} + e_S$ (with 68.3% confidence)

$\bar{x} - 2e_S$ and $\bar{x} + 2e_S$ (with 95.0% confidence)

$\bar{x} - 3e_S$ and $\bar{x} + 3e_S$ (with 99.7% confidence).

RAW DATA - STATISTICAL TREATMENT OF THE READING ERROR

Even though the data sampling shows no scatter (standard deviation of zero) there can still be a random error associated with the data due to the "reading error." The reading error (e_R) takes into account the sensitivity of the instrument (the maximum change required for the instrument to respond); the degree of subdivision of the scale of the instrument (one-half the smallest subdivision); and random fluctuations in the instrument reading in between sampling times (one-half the difference between the maximum and minimum values). The value used for the reading error (e_R) is the largest of the possible values. Generally, some judgment and familiarity with the instrument are needed to come up with a good estimate of the reading error. There is a tendency to underestimate this quantity.

Some considerations for reading error:

(a) How much does the rotameter fluctuate between readings vs. the rotameter scale subdivisions?
(b) How much does the pressure gauge (vacuum gauge) fluctuate vs. the scale subdivisions?
(c) How sensitive are the large weighing scales?
(d) How precisely can the end point be determined in titrating an organic phase (+/- how many ml)?

Once a value is determined for the reading error (e_R) it is combined with the standard deviation (σ) from (Eq. 2) to obtain the standard error as follows:

1. If $e_R \ll \sigma$, then

$$\mathbf{e_S = \frac{\sigma}{\sqrt{N}}} \quad \textbf{(as before)} \; . \qquad \text{(Eq. 4)}$$

2. If $e_R \gg \sigma$, then

$$\mathbf{e_S = \frac{e_R}{\sqrt{3}}} . \qquad \text{(Eq. 5)}$$

The origin of the $\sqrt{\mathbf{3}}$ in (Eq. 5) is the Poisson Distribution.

3. If e_R and σ are of the same order of magnitude then

$$\mathbf{e_S = \frac{1}{2}\left(\frac{\sigma}{\sqrt{N}} + \frac{e_R}{\sqrt{3}}\right)} . \qquad \text{(Eq. 6)}$$

Calculated Data - Empirical Correlations

Theoretical values for friction factors, heat transfer coefficients, mass transfer coefficients, etc. are usually obtained from correlating equations or diagrams that have an often overlooked associated with them. These values have “engineering accuracy”, which usually means 10-20% error; an error of 15% is recommended.

PROPAGATION OF ERROR FROM RAW DATA TO CALCULATED DATA

As mentioned at the beginning, any errors in the raw data will be propagated and will create errors in the calculated data.

If y is the desired quantity and u, v, w, ... are the raw data, in general,

$$y = f(u, v, w, ...).$$

The mean value of y can be obtained by using the mean values of u, v, w, ... in the functional relationship:

$$\bar{\mathbf{y}} = \mathbf{f}(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\mathbf{w}}, \ldots).$$

There are two methods that can be used to estimate the error in y.

1. ROOT MEAN SQUARE ERROR (RMS)

The Root Mean Square Error has a basis in statistics:

$$\mathbf{e}_{\mathbf{RMS,y}} = \sqrt{\left\{\left[\left(\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right)_{\mathbf{v,w}} \mathbf{e}_{\mathbf{S,u}}\right]^2 + \left[\left(\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\right)_{\mathbf{u,w}} \mathbf{e}_{\mathbf{S,v}}\right]^2 + \left[\left(\frac{\partial \mathbf{f}}{\partial \mathbf{w}}\right)_{\mathbf{u,v}} \mathbf{e}_{\mathbf{S,w}}\right]^2 + \ \ldots\right\}_{\bar{\mathbf{u}},\bar{\mathbf{v}},\bar{\mathbf{w}}}} \quad . \qquad \text{(Eq. 7)}$$

where the mean values ($\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\mathbf{w}}$,...) are used to evaluate the derivatives in the above expression.

The RMS Error is tedious to calculate by hand and is probably best suited to computer calculations.

2. UPPER ESTIMATE OF THE PROPAGATED ERROR

An upper limit to the error can be estimated as follows:

$$\mathbf{e}_{\mathbf{UL,y}} = \left|\left(\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right)_{\mathbf{v,w}}\right| \mathbf{e}_{\mathbf{S,u}} + \left|\left(\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\right)_{\mathbf{u,w}}\right| \mathbf{e}_{\mathbf{S,v}} + \left|\left(\frac{\partial \mathbf{f}}{\partial \mathbf{w}}\right)_{\mathbf{u,v}}\right| \mathbf{e}_{\mathbf{S,w}} + \ \ldots \qquad \text{(Eq. 8)}$$

where the mean values ($\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\mathbf{w}}$,...) are used to evaluate the derivatives in the above expression.
This method is easier to use for hand calculations and is recommended even for computer usage.
Note that $e_{RMS} < e_{UL}$ always. Thus, using e_{UL} will give a more conservative, realistic estimate of the error.

ERROR ANALYSIS OF FLOW RATE BY REPLICATED "PAIL AND SCALE" MEASUREMENTS

One common method of measuring flow rate is to measure the mass of liquid collected in a barrel or pail (w_F-w_0) over a time interval (t). If replicated measurements (N) have been made of the final and initial mass ($w_{F,j}$ and $w_{0,j}$) and the time interval (t_j), it would be <u>incorrect</u> to determine the mean, variance, etc. of (w_F, w_0, and t) and then calculate the mass flow rate (m) and its error. The correct procedure would be as follows:

1. Calculate the mass flow rate for each measurement (m_j):

$$\dot{m}_j = \frac{(w_{F,j} - w_{0,j})}{t_j} \quad (j = 1,2,3,\ldots,N).$$

2. Calculate the mean value of the flow rate ($\overline{\dot{m}}$):

$$\overline{\dot{m}} = \frac{\sum_{j=1}^{N} \dot{m}_j}{N}.$$

3. Calculate the standard deviation of $\dot{m}$:

$$\sigma_m = \sqrt{\frac{\sum_{j=1}^{N} (\dot{m}_j - \overline{\dot{m}})^2}{(N-1)}}.$$

4. Determine the reading error associated with each mass flow rate ($\dot{m}_j$) due to propagation of the reading errors in w_F, w_0, and t:

$$(e_{R,m})_j = \frac{[e_{R,w_F} + e_{R,w_0}]}{t_j} + \frac{[(w_F - w_0)_j e_{R,t}]}{t_j^2}.$$

5. Determine the average reading error associated with the mass flow rate:

$$e_{R,m} = \frac{\sum_{j=1}^{N} (e_{R,m})_j}{N}.$$

6. Combine the reading error and the standard deviation as before:

If $e_{R,m} << \sigma_m$, then

$$\mathbf{e_{S,m} = \frac{\sigma_m}{\sqrt{N}}}\,.$$

If $e_{R,m} >> \sigma_m$, then

$$\mathbf{e_{S,m} = \frac{e_{R,m}}{\sqrt{3}}}\,.$$

If $e_{R,m}$ and $\boldsymbol{\sigma_m}$ are of the same order of magnitude then

$$\mathbf{e_{S,m} = \frac{1}{2}\left(\frac{\sigma_m}{\sqrt{N}} + \frac{e_{R,m}}{\sqrt{3}}\right)}\,.$$

<u>ERROR ANALYSIS OF FLOW RATE BY REPLICATED MEASUREMENTS OF CHANGE IN LIQUID LEVEL IN A TANK</u>

One common method of measuring volumetric flow rate (Q) is to measure the change in liquid level in a tank (h_F-h_0) over a time interval (t). If replicated measurements (N) have been made of the final and initial liquid levels ($h_{F,j}$ and $h_{0,j}$) and the time interval (t_j), an error analysis can be performed in the same way as for the “pail and scale” method:

1. Calculate the volumetric flow rate for each measurement (Q_j):

$$\mathbf{Q_j = \frac{\frac{\pi D^2}{4}(h_F - h_0)_j}{t_j} \qquad (j = 1,2,3,\ldots,N)}\,.$$

where D is the inside diameter of the tank (assumed to have no error associated with it).

2. Calculate the mean value of the flow rate ($\overline{\mathbf{Q}}$):

$$\mathbf{\overline{Q} = \frac{\sum_{j=1}^{N} Q_j}{N}}\,.$$

3. Calculate the standard deviation of Q:

$$\sigma_Q = \sqrt{\frac{\sum_{j=1}^{N}(Q_j - \overline{Q})^2}{(N-1)}} .$$

4. Determine the reading error associated with each flow rate (Q_j) due to propagation of the reading errors in h_F, h_0, and t:

$$(e_{R,Q})_j = \frac{\pi D^2}{4}\left[\frac{e_{R,h_F} + e_{R,h_0}}{t_j} + \frac{(h_F - h_0)_j}{t_j^2}\right].$$

5. Determine the average reading error associated with the flow rate:

$$e_{R,Q} = \frac{\sum_{j=1}^{N}(e_{R,Q})_j}{N} .$$

6. Combine the reading error and the standard deviation as before:

If $e_{R,Q} << \sigma_Q$, then

$$e_{S,Q} = \frac{\sigma_Q}{\sqrt{N}} .$$

If $e_{R,Q} >> \sigma_Q$, then

$$e_{S,Q} = \frac{e_{R,Q}}{\sqrt{3}} .$$

If $e_{R,Q}$ and σ_Q are of the same order of magnitude then

$$e_{S,Q} = \frac{1}{2}\left(\frac{\sigma_Q}{\sqrt{N}} + \frac{e_{R,Q}}{\sqrt{3}}\right) .$$

EXAMPLE -- ERROR IN CALCULATED VALUE OF THE OVERALL HEAT TRANSFER COEFFICIENT

The overall heat transfer coefficient (U) is obtained from:

$$\overline{U} = \frac{\overline{Q}}{A\overline{(T_h - T_c)}_{LM}}$$

where $\overline{(T_h - T_c)}_{LM} = LMTD = \dfrac{[\overline{(T_h - T_c)_2} - \overline{(T_h - T_c)_1}]}{\ln[\dfrac{\overline{(T_h - T_c)_2}}{\overline{(T_h - T_c)_1}}]}$.

The error in the calculated value of U due to errors in Q, A, and the temperatures (T_{h2}, T_{h1}, T_{c2}, T_{c1}) is given by:

$$e_{S,U} = \frac{e_{S,Q}}{A(LMTD)} + \frac{Qe_{S,LMTD}}{A(LMTD)^2} + \frac{Qe_{S,A}}{A^2(LMTD)}$$

where

$$e_{S,LMTD} = \{\left\{\frac{[\overline{(T_h - T_c)_2} - \overline{(T_h - T_c)_1}]}{(T_h - T_c)_2} - \ln\left[\frac{(T_h - T_c)_2}{(T_h - T_c)_1}\right]\right\}[e_{T_{h2}} + e_{T_{c2}}]$$

$$+ \left\{\frac{[\overline{(T_h - T_c)_2} - \overline{(T_h - T_c)_1}]}{(T_h - T_c)_1} - \ln\left[\frac{(T_h - T_c)_2}{(T_h - T_c)_1}\right]\right\}[e_{T_{h1}} + e_{T_{c1}}]\} / \left\{\ln\left[\frac{(T_h - T_c)_2}{(T_h - T_c)_1}\right]\right\}^2.$$

If $(T_h\text{-}T_c)_2$ and $(T_h\text{-}T_c)_1$ are approximately equal then:

$$LMTD \approx \frac{1}{2}[\overline{(T_h - T_c)_2} + \overline{(T_h - T_c)_1}]$$

$$e_{S,LMTD} \approx \frac{1}{2}[e_{T_{h2}} + e_{T_{c2}} + e_{T_{h1}} + e_{T_{c1}}].$$

TABLE OF NOMENCLATURE

A	Heat Transfer Area
D	Inside Diameter of Tank
e_S	Standard Error
e_R	Reading Error
h_F, h_0	Final and Initial Liquid Levels, respectively, in Volumetric Flow Rate Measurement
i, j	Refer to a Particular Sample or Data Point
LMTD	Log-Mean Temperature Difference
m	Mass Flow Rate
N	Number of Data (Sample) Points
Q	Volumetric Flow Rate; Heat Transfer Rate
σ	Standard Deviation
σ^2	Variance
T_h, T_c	Temperature of Hot and Cold Fluids, respectively
t	Time Interval for Flow Rate Measurement
U	Overall Heat Transfer Coefficient
u, v, w	Independent Variables Used in a Calculation
w_F, w_0	Final and Initial Mass, respectively, in "Pail and Scale" Method
$\bar{\mathbf{x}}$	Mean Value of x
x_i	Sampled Value of x
y	Dependent Variable Determined in a Calculation

REFERENCES

Bragg, G.B., *Principles of Experimentation and Measurement*, Prentice-Hall, Englewood Cliffs, NJ, (1974).

Barry, A.B., *Errors in Practical Measurement in Science, Engineering, and Technology*, Wiley, NY, (1978).

Lyon, A.J., *Dealing with Data*, Pergamon Press, NY, (1970).