ERROR ANALYSIS

A.J. Pintar and D. Caspary
Department of Chemical Engineering
Michigan Technological University
Houghton, MI 49931
September, 2012

OVERVIEW

Experimentation involves the measurement of "raw data" in the laboratory or field. It is assumed that there is always a difference between the measured value and true value for any measurement. Our estimate of the difference between the measured value and the true value is reported as *Uncertainty or Experimental Error*. Furthermore, experimental data will eventually be reported directly or they will be used in subsequent calculations. Therefore, when measured data are reported, or results calculated from measured data are reported, the reported values must include an estimation of the associated uncertainty. The uncertainty in the reported value is estimated using a tool called **Error Analysis**.

When a measured value is reported directly, the error analysis is complete when the error associated with that value is estimated and reported. If the value is to be combined mathematically with other measured values and the calculated result is to be reported, an additional step called **Propagation of Error** must be performed before reporting the result.

DATA CALCULATED FROM *EMPIRICAL CORRELATIONS*

Oftentimes you will be comparing measured values with values estimated from empirical relationships. These values will also have an error (or uncertainty) associated with them. Theoretical values for friction factors, heat transfer coefficients, mass transfer coefficients, etc. are usually obtained from correlating equations or diagrams and have an often overlooked error referred to as "engineering accuracy". Unless the specific reference states otherwise, engineering accuracy is assumed to be in the range of 10-20% error; therefore, using a ±15% uncertainty is recommended.

TWO TYPES OF ERROR IN DATA

1. Systematic Error:
   a. Has the same sign and magnitude for identical conditions; defined by instrument's *precision*. Systematic Error is predictable.
   b. Sources: Mis-calibration of Instruments
      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, adjusting zero and span
      Correction Factors or Calibration Curves
      Improved procedures
      Comparison to other methods.
   d. Must be corrected before data are reported or used in subsequent calculations.

2. Random Error:
   a. Can be positive or negative.
   b. Can not differentiate source of error between the instrument and the process itself.
   c. Sources: Random Process Fluctuations
      Random Instrument Fluctuations (Instrument *Accuracy*)
      Degree of Subdivision of Instrument's Scale

Equipment "goblins", "Phase of the Moon", Miscellaneous.

d. Must be dealt with using statistics.

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).

When reporting the results of replicated data for UO Lab you will typically report the mean value along with the estimated standard error. The following is the step-by-step procedure for estimating error:

1. Calculate the Mean Value (Estimation of the True Value)

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

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

2. Calculate the Variance

The variance is the sum of the squares of the difference between each measured value and the mean value, divided by the number of replicates minus one. Variance (σ^2) is defined statistically by:

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

3. Calculate Average Standard Deviation (Measure of the *Variability* of the data)

When a data set is small we use the average standard deviation to describe the magnitude of the spread in the data. Average Standard Deviation is simply called Standard Deviation (σ) and is defined as the square root of the variance, i.e. the square root of the expression labeled Eq. 2.

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

Called the *Standard Error of the Means* (e_s) is defined statistically by:

$$\boxed{\mathbf{e}_S = \frac{\sigma}{\sqrt{\mathbf{N}}}} . \qquad \text{(Eq. 3)}$$

It can be shown statistically that, for normally distributed data, the true value of x (the individual measurement) 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).

5. Determine Reading Error

Even though the data sampling shows no scatter (standard deviation of zero) there may still be a random error associated with the data due to the "reading error."

Sources of reading error (e_R) are:

Sensitivity of the instrument (the maximum change required for the instrument to respond)

Degree of subdivision of the scale of the instrument (one-half the smallest subdivision)

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. Students have a tendency to underestimate this quantity.

Some considerations for reading error in UO Lab:

How much does the rotameter or pressure gauge fluctuate between readings vs. the scale subdivisions?

How sensitive are the platform scales?

How precisely can the end point be determined in titrating an organic phase (+/- how many ml)?

What is the manufacturer's published accuracy for the instrument?

6. Adjust the Standard Error for Combined Random Error and Reading Error

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

If $e_R \ll \sigma$, then:

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

But, if $e_R \gg \sigma$, use:

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

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

If e_R and σ are of the same order of magnitude then use the average of the two errors:

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

ESTIMATION OF ERROR IN A CALCULATED RESULT

When measured values are used in calculations, the error associated with each measured value will affect the uncertainty in the final calculated result. The error in each term of the equation must be combined with the error in the other terms. This is called **Propagation of Error**. An estimation of the error in the calculated result must be calculated and reported along with the result.

Method:

If y is the desired quantity and all the individual u, v, w, ... are the raw data needed to calculate y, we can represent the general function as:

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

You would typically run a set of identical repeated experiments and find the individual values of u, v, w,... Next, calculate the mean value of each u, v, w, ... The mean value of y can be calculated by using the mean values of in the functional relationship:

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

Then, to estimate the error associated with y, use either of the two following methods:

A. Root Means Square Error (e_{RMS})

The Root Mean Square Error has a basis in statistics:

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

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

The RMS Error is tedious to calculate by hand and is best suited to spreadsheets.

B. Upper Estimate of the Propagated Error

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

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

where the mean values ($\bar{u}, \bar{v}, \bar{w}$,...) are used to evaluate the derivatives in the above expression. This method is easier to use for hand calculations.

Note that $e_{RMS} < e_{UL}$ always. Thus, using e_{UL} will give a more conservative estimate of the error.

SIGNIFICANT FIGURES

When reporting a value and its associated error use the appropriate number of significant figures (SF). For measured values, the number of SF is a function of the precision of the measuring device. When a calculated result combines more than one measured or estimated value the correct number of SF is the same as that of the least of all the measured values. The correct number of SF for estimated error is typically one less than the number of SF of the calculated result (sometimes two, but oftentimes only one.)

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 incorrect 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):

$$\dot{\mathbf{m}}_{\mathbf{j}} = \frac{(\mathbf{w}_{\mathbf{F,j}} - \mathbf{w}_{\mathbf{0,j}})}{\mathbf{t}_{\mathbf{j}}} \quad (\mathbf{j = 1,2,3,\ldots,N}).$$

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

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

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

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

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

$$(\mathbf{e}_{\mathbf{R,m}})_{\mathbf{j}} = \frac{[\mathbf{e}_{\mathbf{R,w_F}} + \mathbf{e}_{\mathbf{R,w_0}}]}{\mathbf{t}_{\mathbf{j}}} + \frac{[(\mathbf{w}_{\mathbf{F}} - \mathbf{w}_{\mathbf{0}})_{\mathbf{j}}\mathbf{e}_{\mathbf{R,t}}]}{\mathbf{t}_{\mathbf{j}}^2}.$$

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

$$\mathbf{e}_{\mathbf{R,m}} = \frac{\sum_{\mathbf{j=1}}^{\mathbf{N}} (\mathbf{e}_{\mathbf{R,m}})_{\mathbf{j}}}{\mathbf{N}}.$$

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 ($\mathbf{\overline{Q}}$):

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

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:

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

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:

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

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

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:

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

where

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

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

.

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

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

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

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).