Treatment of Errors

Ron Robertson
Physical Chemistry, APSU

a f:\files\courses\361-2lab\08labs\treatment of errors.docx

Types of Errors

- **Mistakes – data can be thrown out**
- **Scale – the error from reading the measuring scale**
- **Replicate – the error from the scatter of multiple analyses**
- **Propagated – the error generated in a calculated result**

Replicate Analysis Error

- **Systematic – inherent in procedure and can be controlled**
- **Random – cannot be controlled**

1. Data takes shape of Gaussian error curve

2. Best estimate of true value is the average $\bar{x}$

3. The standard deviation is a measure of the scatter or precision of the data

For a large number of data points in a replicate analysis where the error is controlled by randomness, 68% of the measurements would lie within one standard deviation (s) of the average ($\bar{x}$).

4. **The standard deviation is a measure of scatter among single measurements. We actually need to know how much the average ($\bar{x}$) might be expected to vary if we repeated our experiments and obtained new data. This is called the standard deviation of the mean and is given the symbol s_m. It is equal to the standard deviation divided by the square root of the number of data pts.**

5. **The standard deviation of the mean is also called the “standard error”. It can be obtained from the <u>Analysis Toolpak</u> in Excel. The <u>Toolpak</u> must be installed as an Add-in under “Excel options”. It then resides in the Excel 2007 ribbon under the “Data” tab. Select “Descriptive Statistics” in the <u>Analysis Toolpak</u>, select the cells containing your data, and generate the standard error.**

6. S_m is smaller than the standard deviation itself because it is the predicted standard deviation of the $\bar{x}$ values taken from an infinite number of data sets.

7. The best estimate of the "true value" of a set of replicate analyses is the average ± s_m *unless* the scale error or propagated error is greater.

8. Report s_m to 1 sf and round off the average to the same place value as the place value of the last sf in s_m.

Scale Errors

If the scale error of each measurement is greater than s_m, then report the best value of a replicate series of analyses as $\bar{x}$ ± scale error. Round off the average to the same place value as the scale error sf.

Rejection of Data

For data points that appear to be obvious mistakes you can reject the data point if the following criteria holds

<u>Q Test</u>

$$Q = \frac{|x_q - x_n|}{\omega}$$

$x_q = questionable\ point$

$x_n = nearest\ neighbor$

$\omega = spread\ of\ data$

If $Q_{exp} > Q_{tab}$ then reject the point

Q Values for 90% Confidence Level

You can use any reasonable confidence level as long as you specify the confidence level. Having a $Q_{exp} > Q_{tab}$ at the 90% confidence level means that it is 90% likely that the examined point is not a member of the data set.

Number of points	Q
4	0.76
5	0.64
6	0.56
7	0.51
8	0.47
9	0.44
10	0.41

From Skoog and West, "Analytical Chemistry", 5th edition, p. 56

Sample problem

A student obtains the following results for the Molarity of an NaOH solution.

0.504

0.510

0.514

0.530

Can 0.530 be rejected?

Calculate the best reported value

Propagation of Errors

If more than 1 physical measurement is used in a computation to obtain a result, we say that error is <u>propagated</u> by the computation.

How do you calculate propagated error?

Suppose F is a function of a, b, c each of which has an error Δa, Δb, Δc. (This error could be scale or replicate error).

The theory of errors states that

$$\Delta F^2 = \Delta F_a^2 + \Delta F_b^2 + \Delta F_c^2$$

ΔF_a is the amount that the function F changes when variable "a" changes by Δa. The same is true for ΔF_b and ΔF_c.

To calculate these terms one needs to use the partial derivative of F with respect to a, b, and c. Since the partial derivative is the slope of F with respect to that variable (holding all other variables constant), then multiplying the partial derivative by the error in that variable tells us how much F changes.

In calculus notation then

$$\Delta F_a = \left(\frac{\partial F}{\partial a}\right)\Delta a$$

So that to get ΔF^2 we find that

$$\Delta F^2 = \left[\left(\frac{\partial F}{\partial a}\right)\Delta a\right]^2 + \left[\left(\frac{\partial F}{\partial b}\right)\Delta b\right]^2 + \left[\left(\frac{\partial F}{\partial c}\right)\Delta c\right]^2$$

What is the best reported value for "F"?

Substitute into the function F for variables a, b, and c. Then solve for ΔF^2 using the values for a, b, and c as well as the errors in a, b, and c. Finally take the square root to get ΔF itself.

The best reported value is

The value of the function F ± ΔF

Round ΔF to 1 sf and report F to the same place value as the sf in ΔF.

Sample problem

If we measure a rectangular computer chip and find that the length is 1.0145 cm and the width is 0.5170 cm, what is the best reported value possible for the area of the chip? Use the error in length and width as 0.0001 cm.

Curve fitting and Modeling

Often we have data and we wish to find the best mathematical function that fits the data. This is called “curve fitting” or “modeling” the data. There are at least 3 options for this:

- **We want to find an equation that best fits our data within the range of the data points. (Ex. 5a on the problem set)**
- **We want to find an equation that can be logically extended outside our data points. (Ex. 5c on the problem set)**

- **We have an equation that the data should fit and we are looking for coefficients of that curve fit that are related to a physical property. (Ex – Ht of Vaporization lab)**

Options for this curve fitting would be to use *Excel* or a dedicated graphing program like *Graphical Analysis*. Parameters like the correlation coefficient (R^2), the Root Mean Square Error (RMSE), or the general shape of the graph can be used to decide on the best fit equation.

Simple example

Try modeling the following data. Try different functions.

Time (s)	Temp (◦C)
0.0	0.0
1.0	1.0
2.0	4.0
3.0	8.5
4.0	16.2