Lecture 2- Experimental Errors & Statistical Analysis

- no analysis is free of error or "uncertainty"

Goal: limit errors and define tolerable level or limits

Analytical Procedure

2-3 replicates are performed and carried out through the entire experiment

-results vary, must calculate "central" or best value for data set.

Mean – "arithmetic mean", average ($\bar{x}$)

$$Mean = \frac{\sum x}{n}$$

Median **–** arrange results in increasing or decreasing order,

Rules:

For odd # values, median is middle value

For even # values, median is the average of the two middle values

Ex: 19.4, 19.5, 19.6, 19.8, 20.1, 20.3

$$Mean = \frac{\Sigma x}{n} = \frac{19.4 + 19.5 + 19.6 + 19.8 + 20.1 + 20.3}{6}$$

$$= 19.78$$

$$\text{Median} = \frac{19.6 + 19.8}{2} = 19.7$$

How do we determine error?

Accuracy – closeness of measurement to its true or accepted value

Precision – agreement between 2 or more measurements of the sample made in exactly the same way

Absolute error (E) – difference between true and measured value

$E = x_i – x_t$

where x_i = experimental value

x_t = true value

*Ex: See Fig. Iron data

$\bar{x}$ = 19.78 ppm Fe

x_t = 20.00 ppm Fe

E = 19.78 – 20.00 ppm = -0.22 ppm Fe

(-) value too low, (+) value too high

Relative error – expressed as % or in ppt

$$E_r = \frac{x_i - x_t}{x_t} \times 100 \quad \text{(as \%)}$$

$$E_r = \frac{x_i - x_t}{x_t} \times 1000 \quad \text{(as ppt)}$$

Several ways to describe precision

Standard Deviation (S)

$$S = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}}$$

Variance (S^2) = square of the standard deviation

n-1 = degrees of freedom

Another name for precision…..

Relative Standard Deviation (RSD)

RSD (ppt) = $\left(\frac{S}{\overline{\overline{X}}}\right)$ x 1000

When express as a percent, RSD termed the coefficient of variation (Cv).

Cv (%) = $\left(\frac{S}{\overline{\overline{X}}}\right)$ x 100

Ex: Four measurements: 51.3, 55.6, 49.9 and 52.0 – Calculate the %RSD of the experiment.

1.

2.

3.

Types of Errors in Chemical Analysis

1. Indeterminate (random) errors
- no identifiable cause
- can't be eliminated
- random, yields both high and low values

2. Determinate (systematic) errors
- have a definite source, identifiable and corrected

a. Instrumental – imperfections in measuring device
b. Method – non ideal chemical or physical behavior
c. Personal – carelessness, personal limitations

Reporting Analytical Data

1. RSD or Cv, with # of data used, i.e. n=4

2. Significant figures

3. Report 90% or 95% confidence intervals

Significant figures – all the certain digits and the 1st uncertain, min # of digits needed to write a value in scientific notation.

Rules for determining sig. fig’s…

1. Disregard all initial zeros

0.03026
not significant – only mark with decimal place

2. Zeros are significant (1) in the middle of a # or (2) at the end of a # on the right hand side of decimal.

104 0.0106   0.106   0.1060

Use of sig figs in computations
Answer can only have the same # of decimal places as that carried by the # with the fewest.

3.4 + 0.020 + 7.310 = 10.730 = 10.7

*uncertainty lies in the tenth place of 3.4

Addition and subtraction – sci notation
1. Express all #’s with the same exponent
2. Align all #’s w/respect to the decimal point
3. Round off #’s according to the # of decimal places in the # w/the fewest decimal places.
Ex:
1.62×10^5
$+ 4.107 \times 10^3$
$+ 0.984 \times 10^6$

Multiplication and division

* normally limited to # of digits in the # w/ fewest sig figs
* the power of 10 has no influence on the # of figs retained

3.26×10^{-5}	4.3179×10^{12}	34.60
x 1.78	x 3.6 x 10-19	÷ 2.462 87
----------------	------------------	-------------
5.80×10^{-5}	1.6×10^{-6}	14.05

Logarithms and antilogarithms

logarithm – composed of a characteristic and a mantissa. Characteristic is the integer part, mantissa the decimal.

1. In a logarithm of a #, keep as many digits to the right of the decimal (mantissa) as there are sig figs in original #

2. In an antilogarithm, keep as many digits as there are digits to the right of the decimal in the original #.

log 339 = 2.530   How to write in sci. notation = 3.39×10^{2}

a. Rule 1
log 4.000×10^{-5} = -4.3979400 = -4.3979

b. Rule 2
antilog 12.5 = 3.16227×10^{12} = 3×10^{12}

Rules for Rounding final data

1. If last digit of a # is 0-4, round down

2. If last digit of a # is 6-9, round up

3. If last digit of a # is 5, round to nearest even #

Don't round any # until calculation is complete…round at end.

Ex: mean of data set is 61.555 and std = 0.069
For the mean, do we round to 61.55 or 61.56?
Look at rule 3, we see round to nearest even # = 61.56 and std = 0.07

Distribution of repeated measurements (large # of data)

Standard deviation gives the measure of the spread of a set of results, but not shape of distribution.

X and S are reasonable estimates of mean and standard deviation, for small # of samples.

population = large # of data when determining distribution and comparing two results.

$X = \mu$ and $S = \sigma$

Where μ is the population mean, σ = population std *the exact value of μ for a population of data can never be determined; it requires an infinite # of measurements to be made.

However, we can set limits (confidence limits) = an interval around x that probably contains μ

The magnitude of the confidence limits is termed the confidence interval.

<u>Comparison of Two Standard Deviations (precision) using F-test</u>

$$F_{calculated} = \frac{s_1^2}{s_2^2}$$

Rules: If $F_{calculated} > F_{table}$ (95%) = significant

Table 4-3 Masses of gas isolated by Lord Rayleigh

From air (g)	From chemical decomposition (g)
2.310 17	2.301 43
2.309 86	2.298 90
2.310 10	2.298 16
2.310 01	2.301 82
2.310 24	2.298 69
2.310 10	2.299 40
2.310 28	2.298 49
—	2.298 89
Average	
2.310 11	2.299 47
Standard deviation	
0.000 14_3	0.001 38

SOURCE: R. D. Larsen, *J. Chem. Ed.* **1990**, *67*, 925; see also C. J. Giunta, *J. Chem. Ed.* **1998**, *75*, 1322.

Are the two standard deviations sig. different?

$F_{calculated} = 0.00138^2/0.00014^2$
$= 93.1$

Significant? Yes according to table

Treatment of Outliers

1. Reexamine for gross errors.

2. Estimate precision to be expected.

3. Repeat analysis if time and sufficient sample is available.

4. If analysis can't be repeated, perform Q-test.

5. If Q-test indicates retention of value, consider reporting median.