# Chapters 5 - 7

# Errors in Chemical Analysis

1

## Experimental Errors

All measurements have errors

(1) Precision-Uncertainty

- Reproducibility (Closeness of each test)

(2) Accuracy

- Nearness to the "truth"

(3) Our goals are to

- minimize errors and to calculate the size of the errors.

2

## Some Important Terms

For a set of measurements:

| No. ($i$) | $1$ | $2$ | $3$ | … | $i$ | … | $N$ |
|---|---|---|---|---|---|---|---|
| Data ($x$) | $x_1$ | $x_2$ | $x_3$ | … | $x_i$ | | $x_N$ |

$N$: number of measurements
$N-1$: degrees of freedom

***Mean (Average)***

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Sample mean (average): $N < 20$

Population mean (average): $N \geq 20$

$\bar{x} = \mu$ (true value): $N \rightarrow \infty$ (infinite)

3

***Median***

Middle result when the data are arranged by size

a. If the data is an odd numbered set, the median is the middle value.

b. If the data is an even numbered set, the median is the average of the middle two values. For example,

An odd-numbered set:

2.9
2.6
2.4
2.3
2.2

Sum = 12.4

$\bar{x}$ = 12.4/5 = 2.5

Median = 2.4

An even-numbered set:

0.1000
0.0902
0.0886
0.0884

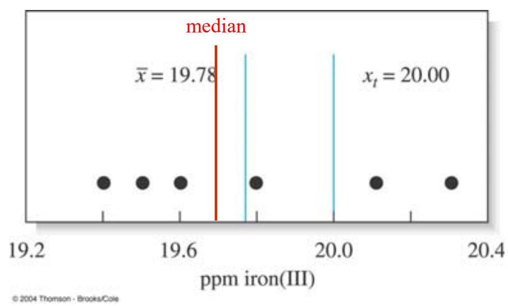
Sum = 0.3672

$\bar{x}$ = 0.3672/4 = 0.0918

Median = (0.0902+0.0886)/2 = 0.0894

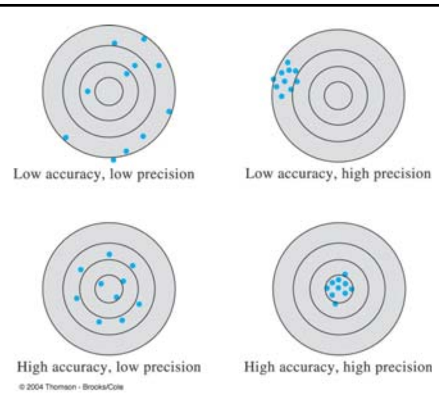
4

## Mean, true, and median value




Results from 6 replicate determination of Fe(III) in aq. samples of a standard soln containing 20.0 ppm Fe(III)

5




Accuracy vs precision using the pattern of darts on a dartboard.

6

## *Precision*

--the closeness of results to others obtained in exactly the same way.

**Sample standard deviation ($s$)**

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

($N < 20$)

**Population standard deviation ($\sigma$)**

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N\to\infty}(x_i - \mu)^2}{N}}$$

($N \geq 20$)  $\mu$-true value

## *"Accuracy"*

***Deviation from the mean*** $d_i$

$$d_i = |x_i - \bar{x}|$$

7

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

$$s = \sqrt{\frac{N\Sigma x_i^2 - (\Sigma x_i)^2}{N(N-1)}} = \sqrt{\frac{\Sigma x_i^2 - (\Sigma x_i)^2 / N}{(N-1)}}$$

EXAMPLE 6-1

The following results were obtained in the replicate determination of the lead content of a blood sample: 0.752, 0.756, 0.752, 0.751, and 0.760 ppm Pb. Calculate the mean and the standard deviation of this set of data.

To apply Equation 6-5, we calculate $\Sigma x_i^2$ and $(\Sigma x_i)^2/N$.

| Sample | $x_i$ | $x_i^2$ |
|---|---|---|
| 1 | 0.752 | 0.565504 |
| 2 | 0.756 | 0.571536 |
| 3 | 0.752 | 0.565504 |
| 4 | 0.751 | 0.564001 |
| 5 | 0.760 | 0.577600 |
| | $\Sigma x_i = 3.771$ | $\Sigma x_i^2 = 2.844145$ |

$$\bar{x} = \frac{\Sigma x_i}{N} = \frac{3.771}{5} = 0.7542 \approx 0.754 \text{ ppm Pb}$$

$$\frac{(\Sigma x_i)^2}{N} = \frac{(3.771)^2}{5} = \frac{14.220441}{5} = 2.8440882$$

Substituting into Equation 6-5 leads

$$s = \sqrt{\frac{2.844145 - 2.8440882}{5-1}} = ,$$

to

$$\sqrt{\frac{0.0000568}{4}} = 0.00377 \approx 0.004 \text{ ppm Pb}$$

See Excel example

8

For example, given the following data set:

| *Data* | *Deviation* |
|---|---|
| 2.3 | \|2.3 - 2.5\| = \| -0.2\| |
| 2.6 | \|2.6 - 2.5\| = \|+0.1\| |
| 2.2 | \|2.2 - 2.5\| = \| -0.3\| |
| 2.4 | \|2.4 - 2.5\| = \| -0.1\| |
| 2.9 | \|2.9 - 2.5\| = \|+0.4\| |

$$\bar{x} = \frac{2.3+2.6+2.2+2.4+2.9}{5} = 2.5$$

$$s = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N-1}}$$

$$s = \sqrt{\frac{(-0.2)^2 + (0.1)^2 + (-0.3)^2 + (-0.1)^2 + (0.4)^2}{5-1}}$$

s = 0.3

9

***Relative standard deviation (RSD)***

$$\text{RSD} = s_r = \frac{s}{\bar{x}}$$

***Coefficient of variation (CV)-RSD expressed as %***

$$\text{CV} = \frac{s}{\bar{x}} \text{ x100\%}$$

RSD can be also expressed in parts per thousand ("ppt", ‰)

$$\text{RSD in ppt} = \frac{s}{\bar{x}} \text{ x1000 ‰}$$

- RSD and CV usually give a clear picture of data quality
- Large RSD or CV implies poor quality/precision

10

***Standard deviation of the mean ($s_m$)***

$$s_m = s/\sqrt{N}$$

***Variance ($s^2$)***

$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$

***Spread or range (w)***

Another way to describe the precision of a set of replicate results.

$$w = x_{max} - x_{min}$$

11

EXAMPLE 6-3

For the set of data in Example 6-1, calculate (a) the variance, (b) the relative standard deviation in parts per thousand, (c) the coefficient of variation, and (d) the spread.

In Example 6-1, we found

$\bar{x} = 0.754$ ppm Pb and $s = 0.0038$ ppm Pb

(a) $s^2 = (0.0038)^2 = 1.4 \times 10^{-5}$

(b) $\text{RSD} = \frac{0.0038}{0.754} \times 1000 \text{ ppt} = 5.0 \text{ ppt}$

(c) $\text{CV} = \frac{0.0038}{0.754} \times 100\% = 0.50\%$

(d) $w = 0.760 - 0.751 = 0.009$ ppm Pb

| | |
|---|---|
| 1 | 0.752 |
| 2 | 0.756 |
| 3 | 0.752 |
| 4 | 0.751 |
| 5 | 0.76 |

Excel 6-3

12

## *Accuracy versus precision*

1. Accuracy is the closeness of a measurement to the true (or accepted) value ($\mu$ or $x_t$).
2. Accuracy is expressed by the *absolute error* or the *relative error*:

Absolute Error $E = x_i - x_t$

where $x_t$ is the true or acepted value of the quantity

$$\text{Relative Error: } E_r = \frac{x_i - x_t}{x_t} \times 100\%$$

$$\text{Relative Error: } E_r = \frac{x_i - x_t}{x_t} \times 1000‰$$

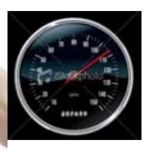
Parts per thousand

13

## Types of Errors in Experimental Data

1. **Systematic Error** (Determinate)
   --changes in signal from true value:
   - predictable
   - correctable (using a ref.)
   - in the same direction ( + or -)



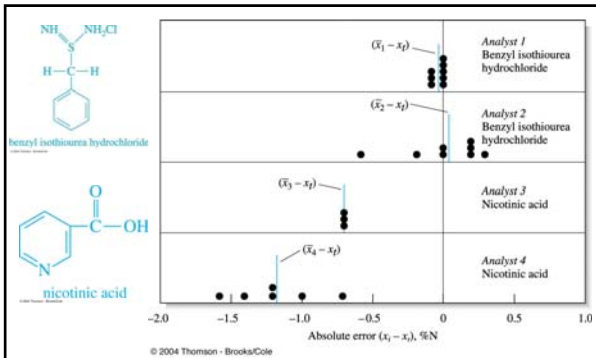



14

**2. Random Error--**changes in signal for replicate measurements:

- always present
- unpredictable
- non-correctable (equal probability of being + or -)
- can be reduced by averaging multiple measurements
- can be treated mathematically (with statistical methods)

15




"Method Error"

Absolute error in the micro-Kjeldahl ("K"-"yell" -"doll") determination of nitrogen (Random and Systematic Errors)

http://en.wikipedia.org/wiki/Kjeldahl_method

16

## Kjeldahl method (N% determination)

1. Degradation:
Sample + $H_2SO_4 \rightarrow (NH_4)_2SO_4(aq) + CO_2(g) + SO_2(g) + H_2O(g)$

2. Liberation of ammonia:
$(NH_4)_2SO_4(aq) + 2NaOH \rightarrow Na_2SO_4(aq) + 2H_2O(l) + 2NH_3(g)$

3. Capture of ammonia:
$B(OH)_3 + H_2O + NH_3 \rightarrow NH_4^+ + B(OH)_4^-$

4. Back-titration:
$B(OH)_3 + H_2O + Na_2CO_3 \rightarrow NaHCO_3(aq) + NaB(OH)_4(aq) + CO_2(g) + H_2O$

17

**3. Gross Error**—(Human) silly mistakes:

- occur only occasionally
- often large (+ or -)
- undetected mistakes during the experiment
- can be verify by "Q-test"

Examples:

0.1000 recorded as 0.0100
1.00 g as 1.00 mg
Wrong connection of electrode wires

18

## Comparison of Random and Systematic Errors

Random Error (affect measurement precision)

- Usually small in values, and not avoidable;
- Equal distributed (+/-) around the mean value;
- Can be treated easily by statistics, normally can be removed by average; (may be quantified by statistical parameters)
- Related to the precision of measurement.

Systematic Error (affect the accuracy of results)

- Due to poor technique or false calibration, sometimes large in values;
- Always in same direction (+ or -), can be detected with calibration or comparison with standard reference samples.
- Difficult to deal with statistically;
- Related to accuracy of measurement.

19

## Sources of Systematic (Determinate) Errors

- ***Instrumental errors***--caused by nonideal instrument behavior, by fault calibration, or by use under inappropriate conditions.
- ***Method errors***—arise from nonideal chemical or physical behavior of analytical systems.
- ***Personal errors***—result from e.g., personal limitations of the experimenter.

20

## Instrumental Errors

- "Drift" in electronic circuits (e.g., inproper zero)—lamp warm-up
- Temperature controls—PMT sensitivity
- Poor power supply—High voltage supply for PMT
- Instruments calibrations—pipets, burets and volumetric flasks volumes, pH meter with standard pH buffers, Reference electrode potentials

21

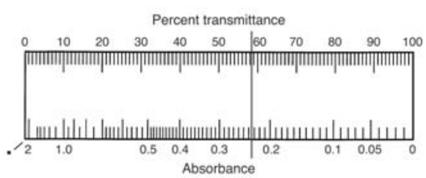
## Method Errors

- Instability of the reagent
- Slowness of some reactions
- Loss of solution by evaporation
- Interferences (pH measurements at high/low pHs)
- Contaminants

22

## Personal Errors

- Estimating the position of a pointer between two scale divisions
- The color of a solution at the end position in a titration—color blindness




**Analog Spectrometer**

23

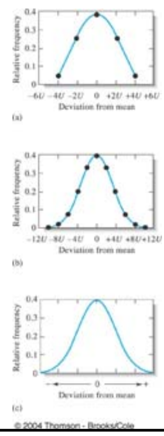
## Detection of Systematic Instrument and Personal Errors

- analysis of standard samples
- independent analysis
- blank determinations
- variation in sample size

24

## Random (Indeterminate) Errors

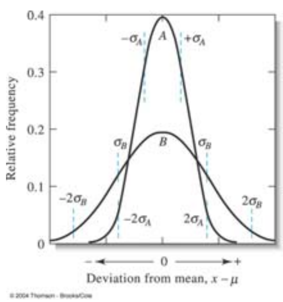
- Affect precision but not accuracy
- Follows a Gaussian or normal distribution
- Most values fall close to the mean, with values farther away becoming less likely. The width of the distribution tells us something about the precision of our measurement.




25

## Gaussian Distribution

Two parameters define a Gaussian distribution for a population, the population mean, μ, and the population standard deviation, σ.



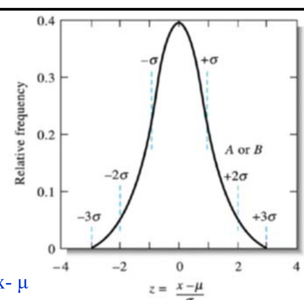

$$\sigma = \sqrt{\frac{\sum_{1}^{N} (x_i - \mu)^2}{N}}$$

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

26

Another way to think about this curve is with the variable z, which is the deviation relative to the standard deviation.

$$z = \frac{x - \mu}{\sigma}$$




So when x - μ = σ, z = 1; when x- μ = 2σ, z = 2.

$$y = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \rightarrow y = \frac{e^{-z^2/2}}{\sigma\sqrt{2\pi}}$$

*ALL normal error curve plots will be the same shape in these z units*

27

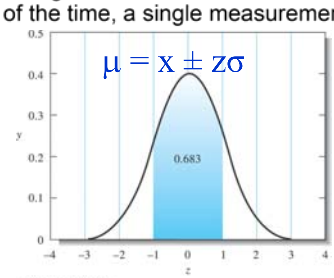
## General properties of a normal error curve (a normalized Gaussian distribution of errors)

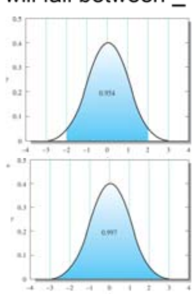
a. The mean (or average) is the central point of maximum frequency (i.e., the top of the bell curve).
b. The curve is symmetric on both sides of the mean (i.e., 50% per side).
c. There is an exponential decrease in resulting frequency as you move away from the mean.
d. If time and expense permit, you need to perform more than 20 replicates when possible to be sure that the sample mean and standard deviation are sufficiently close to the population mean and standard deviation.

28

## Probability and the Area Under the Curve

The area under a Gaussian curve between two values describes the likelihood of any single measurement falling between those two values. For example, 68.3% of the time, a single measurement will fall between $\pm$ 1$\sigma$



$\sigma$ --population standard deviation

29

## Review: Sample vs Population Standard Deviations

The sample standard deviation $s$ is given by

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

where $\bar{x}$ is the sample mean.

The population standard deviation $\sigma$ is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

where $\mu$ is the population mean.

30

| | N ≥ 20 | N < 20 |
|---|---|---|
| statistic | population | sample |
| mean | $\mu$ | $\bar{x}$ |
| st. dev. | $\sigma$ | $s$ |
| variance | $\sigma^2$ | $s^2$ |

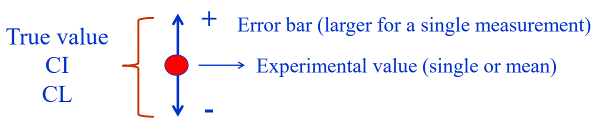
31

## Confidence Intervals (CI) or Confidence Limits (CL)

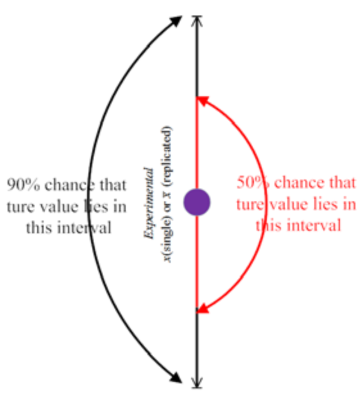
Questions:

How can we predict the true value of the sample with a limited measurements?

Or
How confident would be to locate the true value on the basis of a single or several replicate measurements?




32




33

**The error bar (uncertainty) depends on:**

- Number of measurements (N) (smaller at larger N)
- Confidence % (smaller at less confidence)
- Standard deviations (σ, s) (smaller at smaller σ, s)

**Two cases:**

- When σ is known → Z table
- When σ is unknown → *s* value to replace σ → *t* table

**Single vs replicate measurements (N ≥ 1)**

34

## Calculation of CI ($\mu$) when $\sigma$ is known

| Number of measurements (N) | Confidence Intervals [CI, CL, or μ] |
|---|---|
| Single | $\mu = x \pm \frac{z\sigma}{\sqrt{1}}$ |
| Replicate | $\mu = \overline{x} \pm \frac{z\sigma}{\sqrt{N}}$ |

35

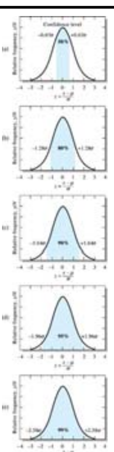


**TABLE 7-1**
**Confidence Levels for Various Values of *z***

| Confidence Level, % | $z$ |
|---|---|
| 50 | 0.67 |
| 68 | 1.00 |
| 80 | 1.28 |
| 90 | 1.64 |
| 95 | 1.96 |
| 95.4 | 2.00 |
| 99 | 2.58 |
| 99.7 | 3.00 |
| 99.9 | 3.29 |

36

EXAMPLE 7-1

N = 7

Determine the 80% and 95% confidence intervals for (a) the first entry (1108 mg/L glucose) in Example 6-2 (page 124) and (b) the mean value (1100.3 mg/L) for month 1 in the example. Assume that in each part, $s = 19$ is a good estimate of $\sigma$.

(a) From Table 7-1, we see that $z = 1.28$ and 1.96 for the 80% and 95% confidence levels. Substituting into Equation 7-1,

$$80\%\ \text{CI} = 1108 \pm 1.28 \times 19 = 1108 \pm 24.3\ \text{mg/L}$$

$$95\%\ \text{CI} = 1108 \pm 1.96 \times 19 = 1108 \pm 37.2\ \text{mg/L}$$

From these calculations, we conclude that it is 80% probable that $\mu$, the population mean (and, *in the absence of determinate error,* the true value), lies in the interval 1083.7 to 1132.3 mg/L glucose. Furthermore, the probability is 95% that $\mu$ lies in the interval between 1070.8 and 1145.2 mg/L.

37

(b) For the seven measurements,

$$80\%\ \text{CL} = 1100.3 \pm \frac{1.28 \times 19}{\sqrt{7}} = 1100.3 \pm 9.2\ \text{mg/L}$$

$$95\%\ \text{CL} = 1100.3 \pm \frac{1.96 \times 19}{\sqrt{7}} = 1100.3 \pm 14.1\ \text{mg/L}$$

Thus, there is an 80% chance that $\mu$ is located in the interval between 1091.1 and 1109.5 mg/L glucose and a 95% chance that it lies between 1086.2 and 1114.4 mg/L glucose.

$$\text{CI for } \mu = \bar{x} \pm \frac{z\sigma}{\sqrt{N}}$$

38

EXAMPLE 7-2

How many replicate measurements in month 1 in Example 6-2 are needed to decrease the 95% confidence interval to 1100.3 ± 10.0 mg/L of glucose?

Here, we want the term $\pm\frac{z\sigma}{\sqrt{N}}$ to equal ±10.0 mg/L of glucose.

$$\frac{z\sigma}{\sqrt{N}} = \frac{1.96 \times 19}{\sqrt{N}} = 10.0$$

$$\sqrt{N} = \frac{1.96 \times 19}{10.0} = 3.724$$

$$N = (3.724)^2 = 13.9$$

$$\text{CI for } \mu = \bar{x} \pm \frac{z\sigma}{\sqrt{N}}$$

We thus conclude that 14 measurements are needed to provide a slightly better than 95% chance that the population mean will lie within ±14 mg/L of the experimental mean.

10 mg/L

39

## Calculation of CI ($\mu$) when $\sigma$ is unknown

| **Number of measurements ($N$)** | **Confidence Intervals [CI, CL, or μ]** |
|---|---|
| Single | $\mu = x \pm \dfrac{ts}{\sqrt{1}}$ |
| Replicate | $\mu = \overline{x} \pm \dfrac{ts}{\sqrt{N}}$ |

40

**TABLE 7-3**

**Values of *t* for Various Levels of Probability**

| Degrees of Freedom (N-1) | 80% | 90% | 95% | 99% | 99.9% |
|---|---|---|---|---|---|
| 1 | 3.08 | 6.31 | 12.7 | 63.7 | 637 |
| 2 | 1.89 | 2.92 | 4.30 | 9.92 | 31.6 |
| 3 | 1.64 | 2.35 | 3.18 | 5.84 | 12.9 |
| 4 | 1.53 | 2.13 | 2.78 | 4.60 | 8.61 |
| 5 | 1.48 | 2.02 | 2.57 | 4.03 | 6.87 |
| 6 | 1.44 | 1.94 | 2.45 | 3.71 | 5.96 |
| 7 | 1.42 | 1.90 | 2.36 | 3.50 | 5.41 |
| 8 | 1.40 | 1.86 | 2.31 | 3.36 | 5.04 |
| 9 | 1.38 | 1.83 | 2.26 | 3.25 | 4.78 |
| 10 | 1.37 | 1.81 | 2.23 | 3.17 | 4.59 |
| 15 | 1.34 | 1.75 | 2.13 | 2.95 | 4.07 |
| 20 | 1.32 | 1.73 | 2.09 | 2.84 | 3.85 |
| 40 | 1.30 | 1.68 | 2.02 | 2.70 | 3.55 |
| 60 | 1.30 | 1.67 | 2.00 | 2.62 | 3.46 |
| ∞ | 1.28 | 1.64 | 1.96 | 2.58 | 3.29 |

© 2004 Thomson - Brooks/Cole

41

**EXAMPLE 7-3**

A chemist obtained the following data for the alcohol content of a sample of blood: % $C_2H_5OH$: 0.084, 0.089, and 0.079. Calculate the 95% confidence interval for the mean assuming (a) the three results obtained are the only indication of the precision of the method and (b) from previous experience on hundreds of samples, we know that the standard deviation of the method $s$ = 0.005% $C_2H_5OH$ and is a good estimate of $\sigma$.

(a) $\Sigma x_i = 0.084 + 0.089 + 0.079 = 0.252$

$\Sigma x_i^2 = 0.007056 + 0.007921 + 0.006241 = 0.021218$

3 $s = \sqrt{\dfrac{0.021218 - (0.252)^2/3}{3 - 1}} = 0.0050\%\ C_2H_5OH$

Here, $\bar{x} = 0.252/4 = 0.084$. Table 7-3 indicates that $t$ = 4.30 for two degrees of freedom and the 95% confidence level. Thus,

$$95\%\ \text{CI} = \bar{x} \pm \frac{ts}{\sqrt{N}} = 0.084 \pm \frac{4.30 \times 0.0050}{\sqrt{3}}$$

$$= 0.084 \pm 0.012\%\ C_2H_5OH$$

42

(b) Because $s = 0.0050\%$ is a good estimate of $\sigma$,

$$95\%\ \text{CI} = \bar{x} \pm \frac{z\sigma}{\sqrt{N}} = 0.084 \pm \frac{1.96 \times 0.0050}{\sqrt{3}}$$

$$= 0.084 \pm 0.006\%\ C_2H_5OH$$

Note that a sure knowledge of $\sigma$ decreases the confidence interval by a significant amount. See Feature 7-1 for a description of alcohol analyzers.

43

## Error Propagation

$$y = (a \pm s_a) + (b \pm s_b) + (c \pm s_c)$$

$$y = (a + b + c) \pm s_y$$

$$s_y = \sqrt{s_a^2 + s_b^2 + s_c^2}$$

44

## Standard Deviation of Calculated Results

- Standard Deviation of a Sum or Difference

| | |
|---|---|
| +0.50 | (±0.02) |
| +4.10 | (±0.03) |
| -1.97 | (±0.05) |
| 2.63 | (?) |

$s_{max} = +0.02 + 0.03 + 0.05 = +0.10$

$s_{min} = -0.02 - 0.03 - 0.05 = -0.10$

Possibly

$s_{min} = -0.02 - 0.03 + 0.05 = 0$

The variance ($s^2$) of a sum or difference is equal to the sum of the individual variances

$$s_y^2 = s_a^2 + s_b^2 + s_c^2$$

$$s_y = \sqrt{s_a^2 + s_b^2 + s_c^2}$$

$$s_y = \sqrt{(0.02)^2 + (0.03)^2 + (0.05)^2} = 0.06$$

45

## Error Propagation

$$y = a \times b / c$$

$$\text{or } y = (a \pm s_a) \times (b \pm s_b) / (c \pm s_c)$$

$$y = a \times b / c \;\; \pm s_y$$

$$\frac{s_y}{y} = \sqrt{(\frac{s_a}{a})^2 + (\frac{s_b}{b})^2 + (\frac{s_c}{c})^2}$$

46

*e.g.*,

$$a = 10.05 \pm 0.050$$

$$b = 1005.0 \pm 5.000$$

$$y = a / b = ? = \frac{10.05 \pm 0.050}{1005.0 \pm 5.000} = \frac{10.05}{1005.0} \pm s_y = 0.01000 \pm s_y$$

$$\frac{s_y}{y} = \sqrt{(\frac{s_a}{a})^2 + (\frac{s_b}{b})^2} = \sqrt{(\frac{0.05}{10.05})^2 + (\frac{5.000}{1005.0})^2} = 7.03 \times 10^{-3}$$

$$s_y = y \times 7.03 \times 10^{-3} = 0.01000 \times 7.03 \times 10^{-3} = 7.03 \times 10^{-5}$$

$$y = 0.01000 \pm 7.03 \times 10^{-5} = 0.01000 \pm 0.00007$$

47

**EXAMPLE 6-4**

Calculate the standard deviation of the result of

$$\frac{[14.3(\pm 0.2) - 11.6(\pm 0.2)] \times 0.050(\pm 0.001)}{[820(\pm 10) + 1030(\pm 5)] \times 42.3(\pm 0.4)} = 1.725(\pm ?) \times 10^{-6}$$

First, we must calculate the standard deviation of the sum and the difference. For the difference in the numerator,

$$s_a = \sqrt{(0.2)^2 + (0.2)^2} = 0.283$$

and for the sum in the denominator,

$$s_b = \sqrt{(10)^2 + (5)^2} = 11.2$$

48

We may then rewrite the equation as

$$\frac{2.7(\pm 0.283) \times 0.050(\pm 0.001)}{1850(\pm 11.2) \times 42.3(\pm 0.4)} = 1.725 \times 10^{-6}$$

The equation now contains only products and quotients, and Equation 6-12 applies. Thus,

$$\frac{s_y}{y} = \sqrt{\left(\frac{0.283}{2.7}\right)^2 + \left(\frac{0.001}{0.050}\right)^2 + \left(\frac{11.2}{1850}\right)^2 + \left(\frac{0.4}{42.3}\right)^2} = 0.107$$

To obtain the absolute standard deviation, we write

$$s_y = y \times 0.107 = 1.725 \times 10^{-6} \times 0.107 = 0.185 \times 10^{-6}$$

and round the answer to $1.7(\pm 0.2) \times 10^{-6}$.

49

**TABLE 6-4**

**Error Propagation in Arithmetic Calculations**

| Type of Calculation | Example* | Standard Deviation of $y$† | |
|---|---|---|---|
| Addition or subtraction | $y = a + b - c$ | $s_y = \sqrt{s_a^2 + s_b^2 + s_c^2}$ | (1) |
| Multiplication or division | $y = a \times b/c$ | $\frac{s_y}{y} = \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2}$ | (2) |
| Exponentiation | $y = a^x$ | $\frac{s_y}{y} = x\left(\frac{s_a}{a}\right)$ | (3) |
| Logarithm | $y = \log_{10} a$ | $s_y = 0.434 \frac{s_a}{a}$ | (4) |
| Antilogarithm | $y = \text{antilog}_{10} a$ | $\frac{s_y}{y} = 2.303\, s_a$ | (5) |

*$a$, $b$, and $c$ are experimental variables with standard deviations of $s_a$, $s_b$, and $s_c$, respectively

†These relationships are derived in Appendix 9. The values for $s_y/y$ are absolute values if $y$ is a negative number.

© 2004 Thomson - Brooks/Cole

50

## Detection of Gross Errors—*Q* test

- On occasion, a set of data may contain a result that appears to be an ***outlier*** (*i.e.* outside of the range of that accounted for by random error).
- Inappropriate or unethical to discard data without a reason.
- The criterion used to decide whether or not to remove the potential ***outlier*** from the data set is the *Q* Test.

51

- The quantity $Q$ (the rejection quotient) is calculated as:

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

where $x_q$ is the questionable result, $x_n$, is the nearest neighbour to the questionable result and $w$ is the spread of the entire set.

- $Q$ is compared to critical $Q$ values, $Q_{crit}$, looked up from the $Q$-Table at a given confidence level.

- If $Q$ is greater than $Q_{crit}$, then the questionable result may be rejected at the indicated confidence level.

52

**TABLE 7-5**

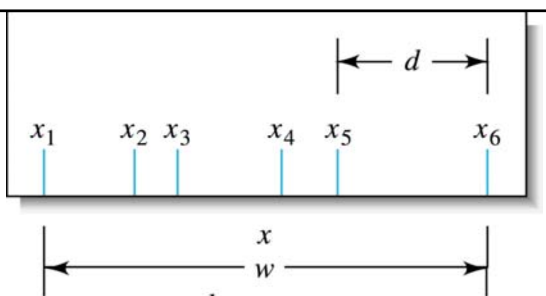
**Critical Values for the Rejection Quotient, $Q$***

| Number of Observations | $Q_{crit}$ (Reject if $Q > Q_{crit}$) 90% Confidence | 95% Confidence | 99% Confidence |
|---|---|---|---|
| 3 | 0.941 | 0.970 | 0.994 |
| 4 | 0.765 | 0.829 | 0.926 |
| 5 | 0.642 | 0.710 | 0.821 |
| 6 | 0.560 | 0.625 | 0.740 |
| 7 | 0.507 | 0.568 | 0.680 |
| 8 | 0.468 | 0.526 | 0.634 |
| 9 | 0.437 | 0.493 | 0.598 |
| 10 | 0.412 | 0.466 | 0.568 |

*Reprinted with permission from D. B. Rorabacher, *Anal. Chem.*, **1991,** *63,* 139. Copyright 1991 American Chemical Society.

© 2004 Thomson - Brooks/Cole

53




$$d = x_6 - x_5$$
$$w = x_6 - x_1$$
$$Q = d/w$$

If $Q > Q_{crit}$, reject $x_6$

© 2004 Thomson - Brooks/Cole

54

## Procedures for Q-test

- Re-arrange the set of data from small to large or large to small
- Identify the smallest and largest questionable data values
- Calculate the Q values for both of the above isolated values
- Compare calculated Q values with the Q values obtained from the Q table at certain confidence levels
- Discard the experimental value if the calculated Q > table Q and keep the value if the calculated Q < table Q

55

**Example**

An Analyst is given a solution containing an unknown concentration of strong base. Titration with 0.1000 M strong acid was used to determine the following results:

| Concentration of base (M) | Volume of Titrant (mL) |
|---|---|
| 0.1012 | 25.30 |
| 0.1014 | 25.35 |
| 0.1015 | 25.37 |
| 0.1035 | 25.88 |

Can the 0.1035 M result be rejected with 90% confidence?

$$Q_{\text{Exp}} = \frac{0.1035 - 0.1015}{0.1035 - 0.1012} = 0.87$$

$$Q_{(90\%\,\text{Conf},\,N=4)} = 0.76$$

Data Point May Be Discarded Since $Q_{\text{Exp}} > Q_{\text{Theoretical}}$

56

**EXAMPLE 7-11**

The analysis of a calcite sample yielded CaO percentages of 55.95, 56.00, 56.04, 56.08, and 56.23. The last value appears anomalous; should it be retained or rejected at the 95% confidence level?

The difference between 56.23 and 56.08 is 0.15%. The spread (56.23 − 55.95) is 0.28%. Thus,

$$Q = \frac{0.15}{0.28} = 0.54$$

For five measurements, $Q_{crit}$ at the 95% confidence level is 0.71. Because 0.54 < 0.71, we must retain the outlier at the 95% confidence level.

57

## Grubbs Test (*G*-Test)

The recommended way of identifying outliers is to use the Grubb's Test. A Grubb's test is similar to a *Q*-test however $G_{exp}$ is based upon the mean and standard deviation of the distribution instead of the next-nearest neighbor and range.

$$G_{exp} = \frac{|x_q - \bar{x}|}{s}$$

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

If $G_{exp}$ is greater than the critical *G*-value ($G_{crit}$) found in the Table then you are statistically justified in removing your suspected outlier .

**Table: Critical Rejection Values for Identifying an Outlier: *G*-test**

| N | 90% C.L. | 95% C.L. | 99% C.L. |
|---|---|---|---|
| | | $G_{crit}$ | |
| 3 | 1.1.53 | 1.154 | 1.155 |
| 4 | 1.463 | 1.481 | 1.496 |
| 5 | 1.671 | 1.715 | 1.764 |
| 6 | 1.822 | 1.887 | 1.973 |
| 7 | 1.938 | 2.020 | 2.139 |
| 8 | 2.032 | 2.127 | 2.274 |
| 9 | 2.110 | 2.215 | 2.387 |
| 10 | 2.176 | 2.290 | 2.482 |

58

## Comparison of Two Experimental Means --the *t* test for differences in means

**Ex:** The homogeneity of the chloride level in a water sample from a lake was tested by analyzing portions drawn from the top and from near the bottom of the lake, with the following results in ppm Cl:

| Top | Bottom |
|---|---|
| 26.30 | 26.22 |
| 26.43 | 26.32 |
| 26.28 | 26.20 |
| 26.19 | 26.11 |
| 26.48 | 26.42 |

Question:
Apply the *t*-test at the 95% confidence level to determine if the means are different?

59

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled}\sqrt{\frac{N_1 + N_2}{N_1 N_2}}}$$ Eqs (7-7) (p155) and (6-7) (p124)

$$s_{pooled} = \sqrt{\frac{\sum_{i=1}^{N_1}(x_i - \bar{x}_1)^2 + \sum_{j=1}^{N_2}(x_j - \bar{x}_2)^2 + \sum_{k=1}^{N_3}(x_k - \bar{x}_3)^2 + \ldots}{N_1 + N_2 + N_3 + \ldots - N_t}}$$

$$= \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}}$$ (when comparing 2 sets of data)

$\bar{x}_1$, $\bar{x}_2$ – mean of the 1st and 2nd set data

$N_1$, $N_2$ – number of the 1st and 2nd set tests

$s_{pooled}$ – pooled standard deviation

$N_t$ – total number of data sets that are pooled

60

## Calculated $t$ vs. critical (theoretical) $t$

(from Table 7-3, where degrees of freedom: $N_1+N_2-2$)
(Page 147)

If $t_{calculated} < t_{critical}$,
NO significant difference between two sets of data

If $t_{calculated} > t_{critical}$,
Significant difference between the means

61

## One Sample $t$-test

| **Number of measurements ($N$)** | $t$ |
|---|---|
| Single | $t = \frac{x - \mu}{s}$ |
| Replicate | $t = \frac{\bar{x} - \mu}{s / \sqrt{N}}$ |

$t$ depends on the desired confidence level and is used to determine if the difference between the experimental mean and the accepted value is due to random error or a systematic error.

62

## Calculated $t$ vs. critical (theoretical) $t$

(from Table 7-3, where degrees of freedom: $N-1$)
(Page 147)

If $t_{calculated} < t_{critical}$,
Measured average agrees with the “true value”

If $t_{calculated} > t_{critical}$,
Significant difference between the measured average and the “true value”; systematic error exists.

63

## Solution to the Ex.

For the Top data set: $\bar{x} = 26.338$

For the Bottom data set: $\bar{x} = 26.254$

$s_{pooled} = 0.1199$

degrees of freedom $= 5 + 5 - 2 = 8$

For 8 degrees of freedom at 95% confidence $t = 2.31$ (Table 7-3)

$$t = \frac{26.338 - 26.254}{0.1199\sqrt{\frac{5+5}{5 \times 5}}} = 1.11$$

Since 1.11 < 2.31, no significant difference exists at 95% confidence

64

## Comparison of Precision

### --the *F* test for differences in standard deviations

An *F*-test can provide insights into two areas:

1) Is method A more precise than method B?
2) Is there a difference in the precision of the two methods?

$$F = \frac{\text{Larger Variance}}{\text{Smaller Variance}} = \frac{s_1^2}{s_2^2} = \frac{\text{numerator}}{\text{denominator}} > 1$$

$s$ : standard deviation

If $F_{calculated} < F_{critical}$ (Table 7-4)
NO significant improvement in precision

65

**TABLE 7-4**

**Critical Values of *F* at the 5% Probability Level (95% confidence level)**

| Degrees of Freedom (Denominator) | Degrees of Freedom (Numerator) 2 | 3 | 4 | 5 | 6 | 10 | 12 | 20 | ∞ |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.40 | 19.41 | 19.45 | 19.50 |
| 3 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.79 | 8.74 | 8.66 | 8.53 |
| 4 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 5.96 | 5.91 | 5.80 | 5.63 |
| 5 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.74 | 4.68 | 4.56 | 4.36 |
| 6 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.06 | 4.00 | 3.87 | 3.67 |
| 10 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 2.98 | 2.91 | 2.77 | 2.54 |
| 12 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.75 | 2.69 | 2.54 | 2.30 |
| 20 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.35 | 2.28 | 2.12 | 1.84 |
| ∞ | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 1.83 | 1.75 | 1.57 | 1.00 |

© 2004 Thomson - Brooks/Cole

66

**EXAMPLE 7-8**

A standard method for the determination of the carbon monoxide (CO) level in gaseous mixtures is known from many hundreds of measurements to have a standard deviation of 0.21 ppm CO. A modification of the method yields a value for $s$ of 0.15 ppm CO for a pooled data set with 12 degrees of freedom. A second modification, also based on 12 degrees of freedom, has a standard deviation of 0.12 ppm CO. Is either modification significantly more precise than the original?

$$F_1 = \frac{s_{\text{Std}}^2}{s_1^2} = \frac{(0.21)^2}{(0.15)^2} = 1.96$$

and for the second,

$$F_2 = \frac{(0.21)^2}{(0.12)^2} = 3.06$$

67

For the standard procedure, $s_{\text{std}}$ is a good estimate of $\sigma$, and the number of degrees of freedom from the numerator can be taken as infinite. From Table 7-4, the critical value of $F$ at the 95% confidence level is $F_{\text{crit}} = 2.30$.

Since $F_1$ is less than 2.30, we cannot reject the null hypothesis for the first modification. We conclude that there is no improvement in precision. For the second modification, however, $F_2 > 2.30$. Here, we reject the null hypothesis and conclude that the second modification does appear to give better precision at the 95% confidence level. *(continued)*

It is interesting to note that if we ask whether the precision of the second modification is significantly better than that of the first, the $F$ test dictates that we must accept the null hypothesis. That is,

$$F = \frac{s_1^2}{s_2^2} = \frac{(0.15)^2}{(0.12)^2} = 1.56$$

In this case, $F_{\text{crit}} = 2.69$. Since $F < 2.69$, we must accept $H_0$ and conclude that the two methods give equivalent precision.

68

## *The t*-test versus the *F*-test

- *t* test is valid for comparison of different sets of data obtained with the same experimental methodology
- *F* test is used to compare precisions obtained with different analytical techniques, e.g., spectroscopic vs electrochemical method

69

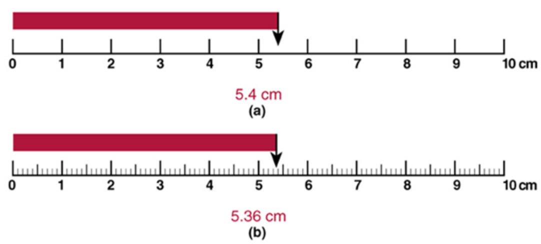
**Significant Figures/Significant Digits**

**(sig figs/sig digs)**

**"digits that carry meaning contributing to a number's precision"**

70

**Significant Figures**




**Significant figures -** all digits in a number representing data or results that are known with certainty plus one uncertain digit.

71

## RECOGNITION OF SIGNIFICANT FIGURES

- All nonzero digits are significant.
  - 3.51 has 3 sig figs
- The number of significant digits is independent of the position of the decimal point
- Zeros located between nonzero digits are significant
  - 4055 has 4 sig figs

72

- Zeros at the end of a number (trailing zeros) are significant *if the number contains a decimal point.*
  - 5.7000 has 5 sig figs
- Trailing zeros are ambiguous if the number does not contain a decimal point
  - 2000. versus 2000
- Leading zeros are not significant.
  - 0.00045 (note: 4.5 x $10^{-4}$)

73

## Examples:

How many significant figures are in the following?

4 7.500

4 2009

3 600.

4 0.003050

6 80.0330

74

6 2.30900

2 0.00040

4 30.07

1,2,or 3 300

2 0.033

75

**SCIENTIFIC NOTATION & Sig Figs**

- Often used to clarify the number of significant figures in a number.
- Example:

  4,300 = 4.3 x 1,000 = 4.3 x $10^3$

  0.070 = 7.0 x 0.01 = 7.0 x $10^{-2}$

76

**SIGNIFICANT FIGURES IN CALCULATION OF RESULTS**

I. Rules for Addition and Subtraction

- the result should have as many decimal places as the **measured** number with the smallest number of ***decimal*** places
- example: 54.4 cm +2.02 cm

| | |
|---|---|
| 54.4 cm | 15.02 |
| 2.02 cm | 9,986.0 |
| 56.42 cm | 3.518 |
| | 10004.538 |

correct answer 56.4 cm

77

Class Practice:

2.0118 + 0.009567 = 2.021367?

$\mu$ = 2.0123 ± 0.008167 = ?

78

## II. Rules for Multiplication and Division

- the result should have as many significant figures as the **measured** number with the smallest number of significant figures.

$$\frac{(4.2\times10^{3})(15.94)}{2.255\times10^{-4}} = 2.9688692\times10^{-8} \text{ (on calculator)}$$

Which number has the fewest sig figs?

The answer is therefore, $3.0 \times 10^{-8}$

79

- For example, if you measured the length, width, and height of a block you could calculate the volume of a block:

  Length: 0.11 cm

  Width: 3.47 cm

  Height: 22.70 cm

Volume = 0.11cm x 3.47cm x 22.70cm

= 8.66459 $cm^3$

Where do you round off?

= 8.66? = 8.7? 8.66459?

80

## III. Rules for Logarithms and Antilogarithms

**1.** In a logarithm of a number, keep as many digits to the right of the decimal point as there are significant figures in the original number.

**2.** In an antilogarithm of a number, keep as many digits as there are digits to the right of the decimal point in the original number.

**EXAMPLE 6-7**

Round the following answers so that only significant digits are retained: (a) $\log 4.000 \times 10^{-5} = -4.3979400$, and (b) antilog $12.5 = 3.162277 \times 10^{12}$

**Solution**

(a) Following rule 1, we retain 4 digits to the right of the decimal point

$$\log 4.000 \times 10^{-5} = -4.3979$$

(b) Following rule 2, we may retain only 1 digit

$$\text{antilog } 12.5 = 3 \times 10^{12}$$

81

## Rules for Rounding Off Numbers

- When the number to be dropped is less than 5 the preceding number is not changed.
- When the number to be dropped is 5 or larger, the preceding number is increased by one unit.
- Round the following number to 3 sig figs: 3.34966 x $10^4$

  =3.35 x $10^4$

82

## Practice:

(51.954-51.874)/0.02000 = ?

(a) 4
(b) 4.0
(c) 4.00
(d) 4.000

Correct answer: (b)

? sig figs

5 sig figs

(1.235 - 1.02) x 15.239 = 2.923438 = ?
1.12

3 sig figs

1.235
-1.02
0.215 = 0.22

83