Chemistry 120

Basic Statistics in Chemistry

Beyond Sig. Figs.

- Significant Figures are first step in describing Quality of a Measurement
  - Critical skill required for all sciences
- No Measurement is complete without
  - Correct significant figures
  - Units
  - Statement of the measurement's uncertainty
- You are responsible for Understanding Concepts for In-Class Exams
  - Actual calculations in lab and on take-home exams

Uncertainty

- Anything that makes the Measurement deviate from the true Value or decreases Confidence in the Result
- Often referred to as *Error*
  - Not necessarily a mistake
  - May result from no fault of experimentalist
- No Measurement is Perfect
  - Measurement tools are limited
  - Seek to minimize uncertainty
  - Imprecision is inherent to the Universe

Types of Uncertainties

- *Gross Error*
  - – Simply a mistake
  - – Correct mistake or discard result
- *Systematic Error*
  - – Decreases accuracy
  - – Detect and correct with *standards*
- *Random Error*
  - – Decreases precision
  - – Treat with statistics

Expressing Accuracy

- If true Value is known, use *Percent Error*
  - – Either '+' if measured value is too high, or '-' if measured value is too low

$$\text{Percent Error} = \frac{\text{measured value} - \text{true value}}{\text{true value}} \text{x}100$$

- – Averaging (finding average, $\overline{x}$) can increase accuracy in the <u>absence</u> of systematic error
- – Use standards

Percent Error Example

- You've determined ρ_{Cu} to be 9.56 g/cm^3. The accepted value is 8.96 g/cm^3. What is the percent error?

$$\text{Percent Error} = \frac{\text{measured value} - \text{true value}}{\text{true value}} \text{ x } 100$$

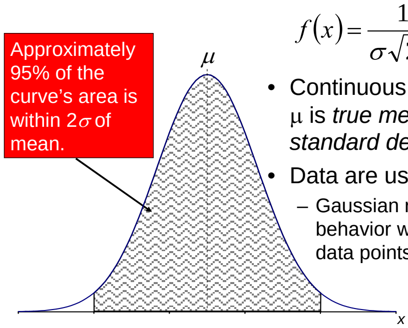
$$\text{Percent Error} = \frac{9.56 \text{ g}\cdot\text{cm}^{-3} - 8.96 \text{ g}\cdot\text{cm}^{-3}}{8.96 \text{ g}\cdot\text{cm}^{-3}} \text{ x } 100 =$$

Dealing with Random Errors

- Range = Highest Value – Lowest Value
  - Sometimes reported as $\bar{x} \pm \frac{\text{Range}}{2}$
  - Small range = high precision
  - Limited use
- Better Treatment assumes that Random Errors have a Gaussian Distribution about the Mean
  - Can also treat data with other distributions
  - See statistics books for more details

Gaussian Distribution

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



- Continuous function, where μ is *true mean* and σ is *standard deviation*
- Data are usually discrete
  - Gaussian represents limiting behavior when number of data points, *N*, is large

Dealing with Random Errors

- Standard Deviation, σ (for N>30)

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

- Estimated Standard Deviation, *S* (N<30)

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

  - Smaller σ or *S* means higher precision
  - Precision increases as *N* increases

Dealing with Random Errors

- *Confidence Limit*
  - Related to *confidence interval*
  - Expresses the confidence that the true value is within a specified range *in the absence of systematic error*
  - Reported as the average ± an uncertainty (Δ) at a specific confidence level
  - Example: 8.99 ± 0.01 g/cm^3 at the 95% confidence limit

Dealing with Random Errors

- Confidence Limit calculated from (Estimated) Standard Deviation

$$\Delta = t \frac{S}{\sqrt{N}}$$

  - *t* (Student's *t*) found in a table for a specific confidence level (usually 95%)
  - Higher confidence level usually more precise
  - Tables list *t* by degrees of freedom (*N*-1)

Reporting Uncertainties

- You have three measurement of ρ_{Cu} (9.54, 9.55 and 9.56 g/cm^3). How do you report this result?

Step 1: Calculate the average.

$\overline{X}$ = 9.55 g/cm^3

Step 2: Calculate the estimated standard deviation.

S = 0.01 g/cm^3

Reporting Uncertainties

Step 3: Determine uncertainty at desired confidence limit.

For N = 3, degrees of freedom = (N-1) = 2.

Look in table of t values at 95% confidence level to find t.

$$\Delta = t\frac{S}{\sqrt{N}} = 4.30\frac{0.01\ \text{g}\cdot\text{cm}^{-3}}{\sqrt{3}} =$$

Reporting Uncertainties

- Report the Result in any of three Ways
  - Confidence interval
    - 9.55 ± 0.02 g/cm^3 at the 95% limit
  - With (estimated) standard deviation stated
    - An average density of 9.55 g/cm^3 with an estimated standard deviation of 0.01 g/cm^3.
  - With (estimated) standard deviation shown in parentheses after last digit
    - 9.55(1) g/cm^3

Reporting Uncertainties

- Sometimes individual Data are more precise than the Data Set as a Whole
  - Example: Data for ρ_{Cu} are 9.541, 9.559 and 9.568 g/cm^3
  - $\overline{x}$ = 9.556 g/cm^3, S = 0.014 g/cm^3 and Δ = 0.034 g/cm^3
  - First uncertain digit (last significant figure) is in 1/10$^{\text{ths}}$ place

9.56(1) g/cm^3 — Round at last significant figure

or

9.556(14) g/cm^3 — Reflect precision of each point

Precision and Accuracy

- Remember Precision $\neq$ Accuracy
- But *in Absence of Systematic Error*, High Precision gives confidence that the Average is true Value
- **Key Point**: when high Accuracy is required high Precision and Elimination of Systematic Error are essential
  - Use most precise measurement tools
  - Run standards

Q-Test

- Used to identify a suspect Datum
- Procedure
  - Place the data in ascending order
  - Look at extremes
  - Calculate Q for suspect data point
  - Compare calculated Q with Q_c in table
  - If $Q > Q_c$, data point may be excluded

$$Q = \frac{|\text{suspect value} - \text{closest value}|}{|\text{highest value} - \text{lowest value}|}$$

Q-Test Example

- Density of Cu is measured to be 9.43, 8.97, 8.96, 8.95 and 8.93 g/cm^3. Can any of these points be eliminated?

Step 1: Identify the suspect point.

9.43 g/cm^3 is suspect (all others are clustered).

Step 2: Calculate Q for this point.

$$Q = \frac{|\text{suspect value} - \text{closest value}|}{|\text{highest value} - \text{lowest value}|} =$$

Q-Test Example

Step 3: Look up Q_c, compare to calculated Q.

Look in table under correct N (= 5) to find Q_c.

Q is greater than Q_c for 5 points (0.92 > 0.64).

The point 9.43 g/cm^3 may be eliminated from the data set on the basis of *Q*-test.

- Important Restrictions on the *Q*-test
  - Only one point can be eliminated
  - Can't test and eliminate duplicate points
  - Once removed a point should not be included in further calculations

Propagation of Error

- Errors NEVER go away
  - Shouldn't random errors cancel?
  - Every calculated value has an uncertainty that results from each measured value used
  - Significant figure rules
- Propagated Error vs. Statistics for a Group
  - Propagated error represents minimum uncertainty based on errors in measurement
  - Statistics for group includes this and other uncertainties

Propagation of Error

- General Formula

$$\Delta f = \pm\sqrt{\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 (\Delta x_i)^2}$$

- Propagated uncertainty depends on
  - Absolute uncertainty (Δx_i)
  - How strongly function depends on quantity $\partial f / \partial x_i$
- What does this Equation tell us about how to make a Measurement?

Example of Error Propagation

- What is the propagated uncertainty in the volume of the small copper block? In the block's density?

Length (cm)	Width (cm)	Height (cm)	Mass (g)
0.49	0.49	0.49	1.105

$$\Delta V = \pm V\sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2 + \left(\frac{\Delta z}{z}\right)^2}$$

$$\Delta d = \pm d\sqrt{\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta m}{m}\right)^2}$$

Summary

- No Measurement is Perfect
- Accuracy and Precision
  - Definitions and how they are connected
- Three Types of Uncertainties (Errors)
  - Gross and systematic errors affect accuracy
  - Random error affects precision
- Error Treatment
  - Gross and systematic: use standards, *Q*-test
  - Random: use statistics, increase *N*
- Errors Propagate