Data Analysis I

CU- Boulder

CHEM-4181

Instrumental Analysis Laboratory

Prof. Jose-Luis Jimenez

Spring 2007

Presentation will be posted on course web page – based on lab manual, Skoog, web links

Objective of Data Analysis Section

- *Treat data in your lab reports and student choice exp. in a professional way*
  - Very easy to generate lots of numbers with modern instruments, but can you quantify their quality?
    - "Recent years have seen the introduction of many [instruments] that are capable of generating data in truly prodigious quantities." (recent paper)
  - "Data of unknown reliability are essentially worthless"
- What you need to know
  - Data analysis section of manual (p. 11-18)
  - Appendix 1 of Skoog, Holler, and Nieman
  - How to use Excel for plotting & linear regression
    - Access to Excel?
  - Useful tutorial linked on web page
- Will go quickly since you've probably seen most of this before
- Data Evaluation Homework Set
  - Due Wed. Jan. 31st at start of class

Review of Significant Figures I

- CQ: A = 1 ; B = 2 ; C = 3 ; D = 4 ; E = 5
- How many significant figures in?
  - 4308
  - 47,000
  - 4.00
  - 35.01 + 7986.0 + 3.152 = 8024.162 ?
  - (56.0 x 0.003460 x 43.42)/1.684 = 4.99587 ?

Review of Significant Figures II

- Any number you report should have correct number of sigfigs
  - Convey to reader how well number is known
    - All certain digits plus 1st uncertain digit (e.g. 2.351)
- Rules
  - All non-zero numbers are significant
  - Leading zeros are always **in**significant
  - Captive zeros are always significant
    - 4308
    - 40.05
  - Trailing zeros are significant only if number contains decimal point
    - 47,000
    - 4.00

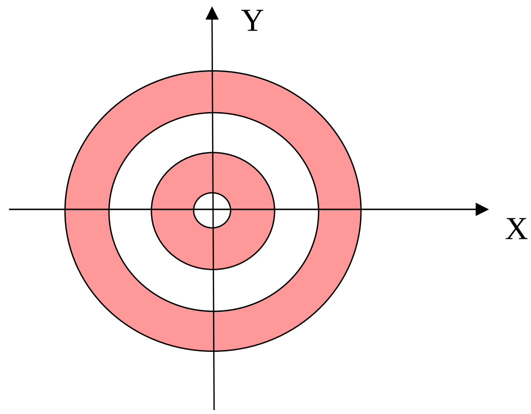
Review of Significant Figures III

- When values are added or subtracted
  - the answer cannot have more sigfigs to right of decimal than the input with the *least* sigfigs
  - 35.01 + 7986.0 + 3.152 = 8024.162 ?
- When values are multiplied or divided
  - the answer has the same sigfigs as the input with the *least* sigfigs
  - (56.0 x 0.003460 x 43.42)/1.684 = 4.99587 ?

Review of Concentration Units

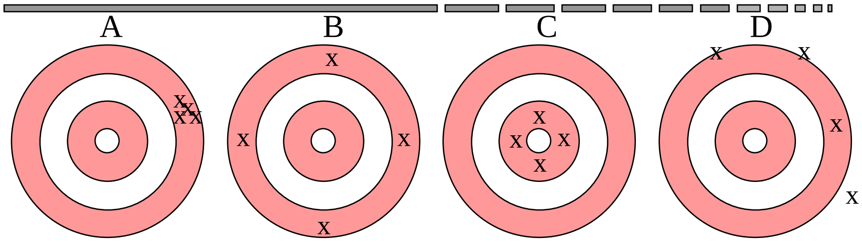
- Mass-to-mass ratios
  - percent, parts-per-hundred
  - ppth, parts-per-thousand
  - ppm, parts-per-million
  - ppb, parts-per-billion (1 part in 10^9)
  - ppt, parts-per-trillion (1 part in 10^{12})
- Volume-to-volume ratios
  - For gases
  - ppmv, ppbv, pptv, etc.
- Q: how many ppt are in 0.031 ppth?

Precision vs. Accuracy I



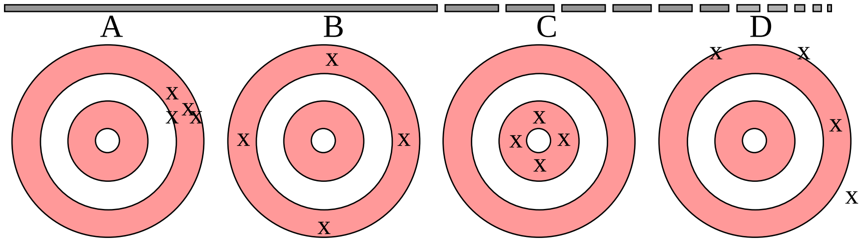
- Measure two variables
  - E.g. concentration of Na^+ and Cl^- in seawater
- Accepted value at origin

Precision vs. Accuracy II



- CQ: Which is the most precise?
  - A
  - B
  - C
  - D
  - I don't know

Precision vs. Accuracy III

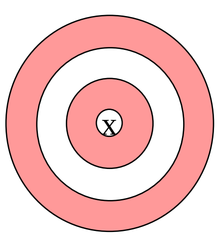


- CQ: Which is the most <u>accurate</u>?
  - A
  - B
  - C
  - D
  - I don't know
- Which is better, A or B?

Precision vs. Accuracy IV

- Precision
  - Agreement between two or more measurements made in an identical fashion
- Accuracy
  - Accuracy is the nearness of a measurement to the accepted value

---



CQ: This measurement is

A. Accurate
B. Precise
C. Precise and Accurate
D. Neither
E. I don't know

Error Notation

- x_i : individual measurement
- x_t : true value
- $\bar{x}$: average value
- E_a : absolute error
- $RE(\%)$: relative error

$$E_a = \bar{x} - x_t$$

$$RE(\%) = \frac{\bar{x} - x_t}{x_t} \cdot 100$$

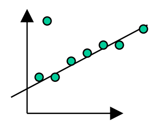
- Q1: what are the units of E_a and RE?
- Q2: you count 2570 cattle on a herd, but the actual value is 2630 cattle
  - E_a? RE?

Types of Experimental Errors

- Random of indeterminate errors
  - Related to precision
  - Treat them with statistics
- Systematic or determinant errors
  - Related to accuracy
  - Get rid of them

$$E_a = E_s + E_r$$

- “Gross” errors
  - Examples
    - Using the wrong scale on a meter
    - Mistake in writing down instrument readout
  - Give rise to “outliers”
    - We’ll deal with these later



Clicker Q

CQ: Which of the following procedures would lead to systematic errors?

A. Using a 1-quart milk carton to measure 1-liter samples of milk.

B. Using a balance that is sensitive to +/-0.1 gram to obtain 250 milligrams of vitamin C.

C. Using a 100-milliliter graduated cylinder to measure 2.5 milliliters of solution.

D. None of the above

E. B & C

http://chemed.chem.purdue.edu/genchem/topicreview/bp/ch1/errors.html

Random Error in Replicate Measurements

- 50 replicate absorbance measurements
  - Spectrophotometer, measuring Fe(III) after treating with excess thiocyanate (*Similar to you Exp. #3*)

	A	B	C	D	E	F	G	H
1	Replicate Absorbance Measurements*							
2	Trial	Absorbance		Trial	Absorbance		Trial	Absorbance
3	1	0.488		18	0.475		35	0.476
4	2	0.480		19	0.480		36	0.490
5	3	0.486		20	0.494		37	0.488
6	4	0.473		21	0.492		38	0.471
7	5	0.475		22	0.484		39	0.486
8	6	0.482		23	0.481		40	0.478
9	7	0.486		24	0.487		41	0.486
10	8	0.482		25	0.478		42	0.482
11	9	0.481		26	0.483		43	0.477
12	10	0.490		27	0.482		44	0.477
13	11	0.480		28	0.491		45	0.486
14	12	0.489		29	0.481		46	0.478
15	13	0.478		30	0.469		47	0.483
16	14	0.471		31	0.485		48	0.480
17	15	0.482		32	0.477		49	0.483
18	16	0.483		33	0.476		50	0.479
19	17	0.488		34	0.483			
20	*Data listed in the order obtained							
21	Mean	0.482		Maximum	0.494			
22	Median	0.482		Minimum	0.469			
23	Std. Dev.	0.0056		Spread	0.025			

From Skoog

Histograms

- Data in table form are unwieldy
- Organize: count number in equally-sized bins

TABLE a1-2 Frequency Distribution of Data from Table a1-1

Absorbance Range, *A*	Number in Range, *y*	Relative Frequency, *y/N*[a]
0.469–0.471	3	0.06
0.472–0.474	1	0.02
0.475–0.477	7	0.14
0.478–0.480	9	0.18
0.481–0.483	13	0.26
0.484–0.486	7	0.14
0.487–0.489	5	0.10
0.490–0.492	4	0.08
0.493–0.495	1	0.02

[a]*N* = total number of measurements = 50.

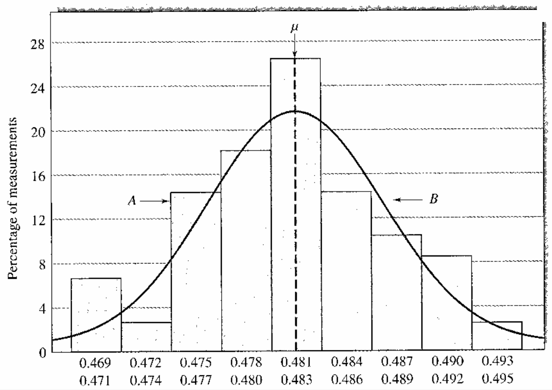


FIGURE a1-1 *A*, Histogram showing distribution of the 50 results in Table a1-1. *B*, Gaussian curve for data with the same mean and standard deviation as those in *A*.

From Skoog

Properties of Normal Distribution

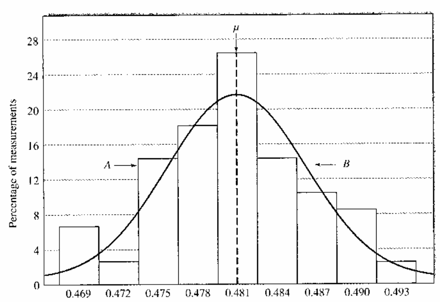


FIGURE a1-1 *A*, Histogram showing distribution of the 50 results in Table a1-1. *B*, Gaussian curve for data with the same mean and standard deviation as those in *A*.

- Most frequently observed value (“median”) is also the mean (μ)
- Results cluster symmetrically around mean
- Small deviations from mean are more common than large ones
- In the absence of systematic errors, the mean approaches the true value
  - Gets better as you add more measurements

Systematic Errors

- Have a definite value
- Same magnitude for replicate measurements
- It has a sign
- Does not “average away”
- 3 Types
  - Instrumental: non-ideal instrument behavior, faulty calibrations
    - E.g.: drift in electronics, leaks into vacuum systems, temperature effects on detectors, pickup from 110 V wall power, drained batteries
    - Detectable and correctable by calibration with standards

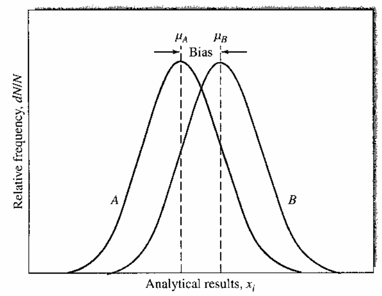


FIGURE a1-2 Illustration of systematic error in analytical results. Curve A is the frequency distribution for the accepted value by Method A, which has no bias. Curve B illustrates the distribution of results by Method B, which has a significant bias = $\mu_B - \mu_A$.

Systematic Errors II

- Types
  - Personal: from judgment of experimentalist
    - E.g.: estimating when color changes in titration, reading a buret, reading a needle on a scale
    - Prejudice: we want results to fall closer to what we think is correct result
    - Minimizing by care and personal discipline. Double-check!
  - Method: non-ideal chemical and physical behavior or reagents and reactions
    - E.g.: slow or incomplete reactions, losses by evaporation, adsorption onto solid surfaces, instability of reagents, contaminants, and chemical interferences
    - Harder to detect. Need *validation of the method* by analyzing materials that ressemble the samples in physical state, composition, concentration, matrix (NIST SRMs)