Chapter 3

EXPERIMENTAL ERRORS

1

Other Common **Analytical Terms**

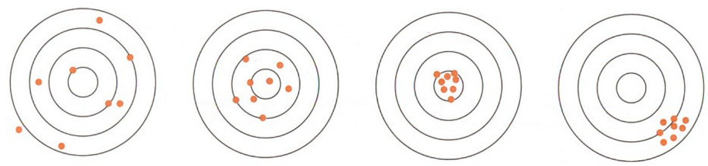
1. **Precision** refers to reproducibility of repetitive measurements of equivalent analyte solutions (e.g. aliquots)
   - Often expressed as (absolute) **standard deviation**, **s**, or **relative standard deviation**, **RSD**

vs. ***Accuracy***, which indicates *how close* the measured analyte concentration is to the true analyte concentration in the sample.

- Often expressed as **bias** (by comparison with standard reference materials)

2

Accuracy vs. Precision



Neither	A, but not P	A & P	P, but not A
		➢ Results close together and close to true value	➢ Results close together, but far from true value

- ➢ Good precision does not assure accuracy
- ➢ However, the ***higher the precision***, the *greater is the probability of obtaining the true value.*
    - ❖ It helps to get highly reproducible (highly precise) replicate measurements

Analytical Terms – *Cont.*

2. **Sensitivity** ***(of an instrument)*** – The smallest amount of a substance that can be detected by an instrument

Illustration of instrument sensitivity:

What is the sensitivity of each balance?

Top loading electronic balance with ***0.01 g sensitivity*** and 1810 g capacity.

Digital analytical balance with ***0.0001 g sensitivity (or 0.1 mg)*** and 100 g capacity.

- More sensitive

Images available at http://www.acmescales.com/

The results are back from the lab: John Smith is pregnant.

[Courtesy 3M Company, St. Paul, MN.]

Some laboratory errors are more obvious than others, but there is error associated with every measurement. There is no way to measure the "true value" of anything. The best we can do in a chemical analysis is to carefully apply a technique that experience tells us is reliable. Repetition of one method of measurement several times tells us the *precision* (reproducibility) of the measurement. If the results of measuring the same quantity by different methods agree with each other, then we become confident that the results are *accurate*, which means they are near the "true" value.

This picture shows that errors, including ***experimental errors***, are real.

The results are back from the lab. *Now what do we do?*

Data handling and error analysis:

Accuracy = closeness of result to the *accepted or* true value; measured with:

- **bias**

Precision refers to the reproducibility of results; measured with:

- **standard deviation (s)**, or
- **relative standard deviation**, **RSD**

7

Bias: A measure of accuracy

Bias is a quantitative term describing the difference between the measured quantity (or its mean) and its true or known value.

Bias = measured value – known value

8

Ways of expressing precision

1. (**Absolute**) **standard deviation, s**

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

where x = individual measurements; $\bar{X}$ = mean of all measurements and n = total number of measurements

2. **Relative standard deviation, RSD**

$$RSD = \frac{s}{\bar{X}}$$

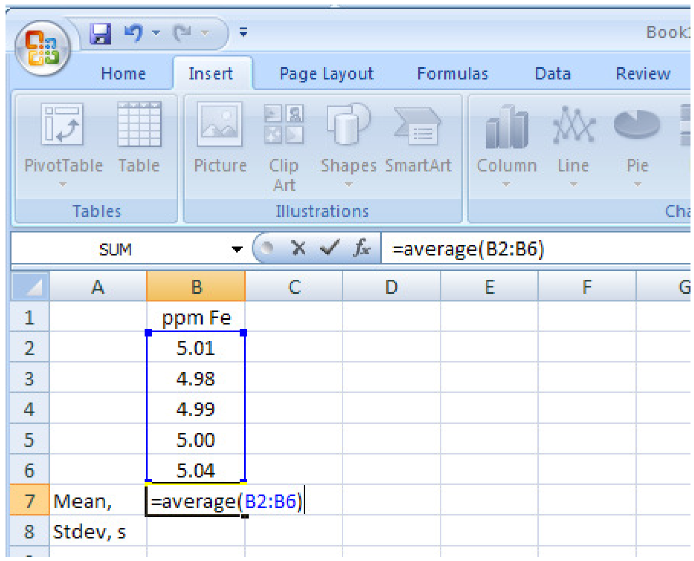
Example: Determining the precision of a set of measurements

The following data were obtained for the determination of iron in a vitamin tablet using flame atomic absorption spectrometry. Calculate the mean ($\bar{x}$) and the relative standard deviation (RSD).

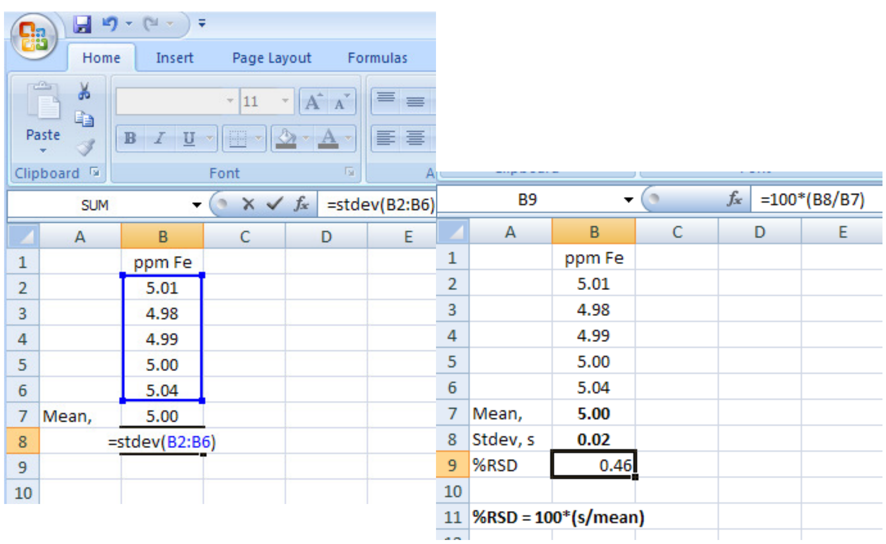
sample #	**ppm Fe**
1	5.01
2	4.98
3	4.99
4	5.00
5	5.04

$\bar{x} = 5.00_4$

Do it in Excel!



11



Reported ppm Fe: **5.00 ± 0.02 ppm**
or %RSD = 0.46 % error. Pretty reproducible, isn't it?
A very precise measurement!

12

Accuracy is more difficult to measure than precision. *WHY?*

- The *true* or *accepted value* is not always available.
  - However, remember that the ***higher the precision,*** the ***greater is the probability of obtaining an accurate result***

13

Chapter 4

STATISTICS

14

Scenario:

The presence of ***dissolved copper*** in drinking water is typically due to corrosion of household plumbing systems. Its levels in drinking water is regulated by the EPA because of its health effects, such as gastrointestinal distress and liver or kidney damage

The **maximum contaminant level** for **Cu** is **1.3 ppm** (or 1300 ppb)

(http://www.epa.gov/safewater/contaminants/index.html#mcls).

15

The maximum contaminant level for Cu is **1.3 ppm**. The following data were obtained from the analysis of water samples collected over a 5-day period.

Sample ID	Amount of Cu in mg/L
Day 1	1.28
Day 2	1.25
Day 3	1.34
Day 4	1.29
Day 5	1.55

Mean ppm Cu = 1.34

Based on the *5-day mean copper l*evels, is this water safe for drinking?

16

Statistics in Chemistry

Measurements in chemistry

- Often *variable*
  - Difficult to make a conclusion with certainty

  Example: See problem in previous slide

Questions:

(1) Is water from the source safe to drink?

(2) Is the last reading (day 5) off?

- Hard to say
- Need a tool to decide whether the mean ppb Cu is significantly higher than the limit, or to decide whether or not to *discard* the last result

17

Importance of Statistics in Chemistry

- Gives us a *tool to accept or reject conclusions*
  - **High probability** of being correct = **ACCEPT**
  - **Low probability** of being correct = **REJECT**

18

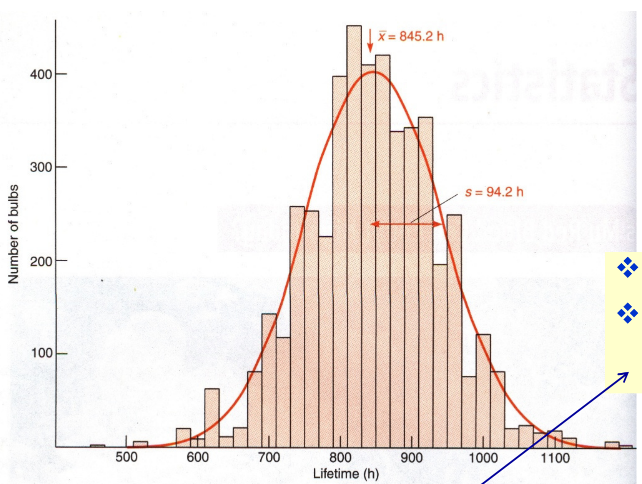
Distribution of Experimental Results

Scenario: Let's say the analysis of the same sample of water for Cu was repeated 100 times by atomic absorption spectroscopy. Let's also limit errors to random error.

Question: How would the results look like?

- Results will tend to cluster around the mean value for Cu levels
- Gaussian distribution (Figure 4-1, p. 69)
  - Increase repetitions, smoother curve

Gaussian Distribution



- Bell-shaped
- Normal distribution of variation in expt'l data

Figure 4-1, p. 69. Bar graph and Gaussian curve for lifetimes of hypothetical light bulbs

Gaussian Distribution (Cont.)

In real life, we repeat experiments **3-4 times**, not 100 times!

Small set of results

⇓

Estimate statistical parameters

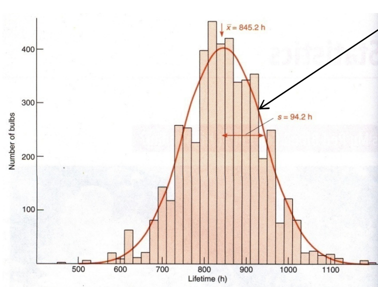
⇓ **Describes**

Large set of results

Allows estimation of statistical behavior from a small number of repetitions (analysis)

21

Characteristics of the Gaussian Curve



(Smooth orange line)

1. **Mean, $\bar{x}$** (also called *average*)
   - ➢ Sum of measured values, x_i, divided by the number of measurements, n

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

22

Characteristics of the Gaussian Curve – *Cont.*

2. Standard deviation, s

- Measures how closely the data are clustered about the mean

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

- Measures the <u>width</u> of the Gaussian curve
  - ***Smaller s, narrower curve*** – Fig. 4-2, p. 54
    - *Smaller variability*
    - *Individual data are close together*

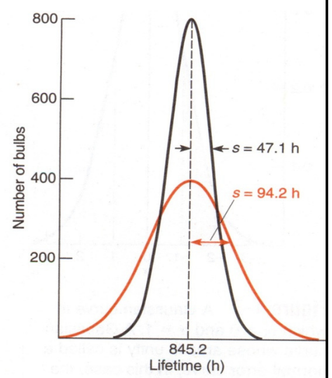


Figure 4-2 Gaussian curves for two sets of light bulbs, one having a standard deviation half as great as the other. The number of bulbs described by each curve is the same.

23

Characteristics of the Gaussian Curve – *Cont.*

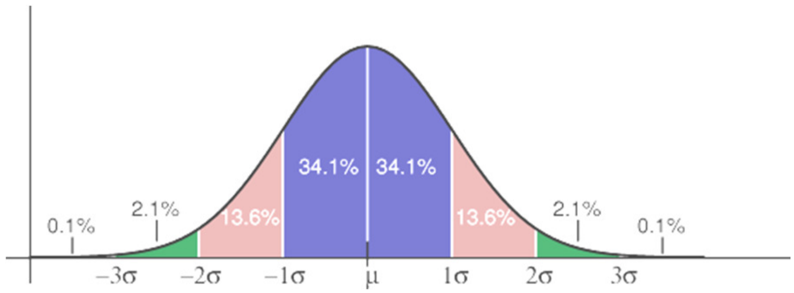
Gaussian curve and standard deviation – See table on p. 57

Note that sigma, σ, is used instead of s for standard deviation of larger sets of data

Range	% of Measurements	Meaning
$\mu \pm 1\sigma$	68.3	**More than 2/3 (68 %) of measurements are within 1 std. dev. from the mean**
$\mu \pm 2\sigma$	95.5	**Over 95 % of measurements are within 2 std. dev. from the mean**
$\mu \pm 3\sigma$	99.7	**Almost 100 % of measurements are within 3 std. dev. from the mean**

24

Gaussian Distribution



Gaussian distribution. Image available at
http://introcs.cs.princeton.edu/java/11gaussian/

Gaussian curve and standard deviation – Cont.

Practical significance

- Comparing two methods (and two $\bar{x}$) to measure % Fe in ore

	Method A	Method B
$\bar{x}$	32.4	31.8
s	0.8	1.1

Interpretation of results

- About 68 % of measurements from Method A will fall between 31.6 - 33.2 (vs. 30.7 - 32.9 from Method B)
- 95.5 % of measurements from Method A will fall between 30.8 - 34.0 (vs. 29.6 - 34.0 from Method B)

Parameters for Finite vs. Infinite Set of Data

Finite set	Infinite set
Sample mean, $\bar{x}$	Population mean, μ
Standard deviation, s	Population std. dev., σ

NOTE: μ and σ cannot be measured, but …

- … *as the number of measurements increases values of $\bar{x}$ and s approach μ and σ*

27

Other Statistical Parameters

Degrees of freedom, n – 1

- Used in the calc. of s

Variance, s^2 (or σ^2)

- Squared std. dev.

Relative standard deviation, RSD (or the *coefficient of variation*)

- Std. dev. expressed as percentage of the mean

Confidence interval

- An expression which states that the true mean, μ, is likely to lie within a certain distance from the measured mean, $\bar{x}$

28

Confidence Intervals

- An expression which states that *the true mean, μ, is likely to lie within a certain distance from the measured mean,* $\bar{x}$

Confidence interval:

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{n}}$$

Where t is the *Student's* t, taken from Table 4-2, p. 73

- Allows us to determine μ from the measured mean, $\bar{x}$, and std. deviation, s

29

Confidence interval:

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{n}}$$

Exercise: (Prob. 4-11, p. 75) The % of an additive in gasoline was measured six times with the following results: 0.13, 0.12, 0.16, 0.17, 0.20 and 0.11 %. Find the 90 % and 99 % confidence intervals for the percentage of the additive.

30

Answers: **90 %: 0.15 ± 0.03; 99 %: 0.15 ± 0.05**

Table 4-2 Values of Student's t

n – 1 =

Degrees of freedom	Confidence level (%) 50	90	95	98	99	99.5	99.9
1	1.000	6.314	12.706	31.821	63.657	127.32	636.619
2	0.816	2.920	4.303	6.965	9.925	14.089	31.598
3	0.765	2.353	3.182	4.541	5.841	7.453	12.924
4	0.741	2.132	2.776	3.747	4.604	5.598	8.610
5	0.727	2.015	2.571	3.365	4.032	4.773	6.869
6	0.718	1.943	2.447	3.143	3.707	4.317	5.959
7	0.711	1.895	2.365	2.998	3.500	4.029	5.408
8	0.706	1.860	2.306	2.896	3.355	3.832	5.041
9	0.703	1.833	2.262	2.821	3.250	3.690	4.781
10	0.700	1.812	2.228	2.764	3.169	3.581	4.587
15	0.691	1.753	2.131	2.602	2.947	3.252	4.073
20	0.687	1.725	2.086	2.528	2.845	3.153	3.850
25	0.684	1.708	2.060	2.485	2.787	3.078	3.725
30	0.683	1.697	2.042	2.457	2.750	3.030	3.646
40	0.681	1.684	2.021	2.423	2.704	2.971	3.551
60	0.679	1.671	2.000	2.390	2.660	2.915	3.460
120	0.677	1.658	1.980	2.358	2.617	2.860	3.373
∞	0.674	1.645	1.960	2.326	2.576	2.807	3.291

31

WORK:

	0.12		
	0.16		
	0.17		
	0.20		
	0.11		
mean	**0.15**	(0.148)	
stdev	**0.03**	(0.028)	
deg. freedom	5.00	sqrt (n) =	2.4495
t at 90 % CI	2.015		
t at 99 % CI	4.032		
CI (90%) =	0.148 +/- (2.015*0.028)/sqrt(6		
=	0.148 +/-	0.023	
=	**0.15 +/- 0.02**		
CI (99%) =	0.148 +/- (4.032*0.028)/2.449		
=	0.148 +/-	0.046	
=	**0.15 +/- 0.05**		

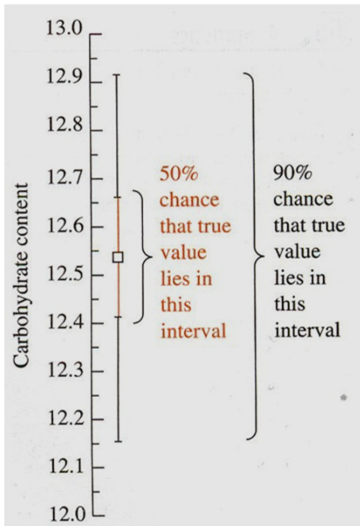
95 % chance that the true mean lies within the range 0.12 % to 0.18 % (of additive)

99 % chance that the true mean lies within the range 0.10 % to 0.20 % (of additive)

32

Confidence Intervals -Cont.

- A closer look at the meaning of **confidence interval**



- 50 % C.I: *There is a 50 % chance that true mean, μ, lies between 12.4 to 12.7 % carbs*
- 90 % C.I: *There is a 90 % chance that true mean, μ, lies between 12.1 to 12.9 % carbs*

33

Comparing Means with *Student's t*

- Allows for comparison of *two sets of measurements* (e.g. using 2 different methods) to decide whether or not they are the same
- Recall the following results for measuring % Fe in ore using two different methods

	Method A	**Method B**
$\bar{x}$	**32.4**	**31.8**
s	**0.8**	**1.1**

- Are the results from the 2 methods different? (i.e. *Are the two means different?*)
    - Perform a t-test

34

The *Student's t*

Three cases:

1. Comparing a measured result with a "known" value

Example: You purchased a standard solution of Fe as 1000 ppm Fe (± 5 ppm). You are testing a new analytical method for Fe analysis to see whether it can reproduce the known value. A t-test can be used to compare the mean of your measurements with the known value of 1000 ± 5 ppm Fe.

- Read p. 76 for more information on case 1

2. Comparing two replicate measurements

- Use a t-test to decide whether 2 sets of replicate measurements give "the same" or "different" results, within a stated confidence level

35

Case 2: Comparing two replicate measurements– Cont.

- Perform a t-test to determine if the result from Method B is significantly lower that that from Method A. Assume that each method consisted of 5 replicate measurements.

	Method A	Method B
$\bar{x}$	32.4	31.8
s	0.8	1.1

- Performing a t-test for 2 sets of data consisting of n_1 and n_2 measurements:

Step 1: Calculate a value of *t* with the formula:

Calc. of t:

$$t_{calc} = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

36

where

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}}$$

NOTE: s_{pooled} is a *pooled* standard deviation from both sets of data

Step 2: Compare t_{calc} with tabulated t (Table 4-2) for $n_1 + n_2 - 2$ degrees of freedom.

- If *t_{calc} is greater than t_{table} at the 95 % confidence level, the two results are considered to be different*

37

Case 2 – Cont.

Exercise: Determine if the two means from the measurements below are significantly different. Assume that each method consisted of 5 replicate measurements.

	Method A	**Method B**
$\bar{x}$	**32.4**	**31.8**
s	**0.8**	**1.1**

- Can be performed with Excel (Case 2 only) – read Section 4-5, p. 82, "***t* tests with a spreadsheet**"

38

Cases: The Student's t – *Cont.*

3. Comparing individual differences (Paired t Test)

- Two different methods are used to make *single measurements on several different samples* (NOTE: There are no replicate measurements)
- Answers the question: *Is method A systematically different from method B?*
- Not as commonly used in chem labs as Case 2
- For more information on Case 3, read p. **78**

39

Grubb's Test for an Outlier

Outlier = a data point that is far from other points

Given twelve results for determining the mass % Zn in galvanized nail:

	Mass % Zn		Mass % Zn
	10.2		10.0
	10.8		9.2
	11.6		11.3
	9.9		9.5
	9.4		10.6
Outlier? →	7.8		11.6

Q. Should the reading 7.8 be discarded?

- We can't simply discard bad data
- But we can perform a ***Grubbs test*** to determine if an outlier can be discarded

40

Grubbs test for outliers

- Limited to at least **4** repetitions/replicates
- Calculated from the mean ($\bar{x}$) and standard deviation (s) of the *complete data set* using the equation:

$$G_{\text{calculated}} = \frac{|\text{questionable value} - \bar{x}|}{s}$$

- ***If G_{calc} > $G_{critical}$ discard the outlier.*** *Otherwise, retain it.*

Table 4-5, p. 83

Calculation:

$\bar{X}$ = 10.16; s = 1.11; outlier = 7.9

$$G_{calc} = \frac{|7.8 - 10.16|}{1.11} = 2.13$$

41

TABLE 4-5 Critical values of *G* for rejection of outlier

Number of observations	*G* (95% confidence)
4	1.463
5	1.672
6	1.822
7	1.938
8	2.032
9	2.110
10	2.176
11	2.234
12	2.285
15	2.409
20	2.557

$G_{calculated}$ = |questionable value − mean|/s. If $G_{calculated}$ > G_{table}, the value in question can be rejected with 95% confidence. Values in this table are for a one-tailed test, as recommended by ASTM.

SOURCE: *ASTM E 178-02 Standard Practice for Dealing with Outlying Observations, http://webstore.ansi.org; F. E. Grubbs and G. Beck, Technometrics* 1972, *14*, 847.

- Since G_{calc} (2.137) < $G_{critical}$ (2.285 when n=12), we cannot discard the outlier 7.8.
- There is > 5% chance that the value 7.8 is a member of the same population as the other measurements
- HOWEVER, if the value 7.8 is due to a faulty procedure (e.g. a spill during the expt.), then it must be discarded.

42

Practice problems

Chapter 4 Problems (pp. 93-95)
Gaussian distribution: 1 and 3
Confidence intervals, t-test and Grubbs test: 8, 9, 12, 13, 16, 19, 21, 22 and 23.

43