Table of Contents

Statistical Methods of Data Analysis

Chemistry in Context

You have been contracted by an environmental law firm to measure the concentration of oil that may have formed “plumes” deep under water in the Gulf of Mexico. How do you perform the analysis so that your data have meaning?

The law firm also suspects that many plumes exist and wants to know if the hydrocarbon concentration is the same in every plume. How can you answer this question with statistical meaning?

Finally, the law firm has requested that two different methods for measuring the hydrocarbon concentration in plumes be used. How can you determine if the two methods provide the same answer?

Sample Exam Questions

(1) The carbohydrate content of a glycoprotein is determined to be 12.6, 11.9, 13.0, 12.7, and 12.5 g of carbohydrate per 100 g of protein in replicate analyses. Find the 50% and 90% confidence intervals (μ) for the carbohydrate content. Write out in words what these calculations mean about where the true mean lies.

(2) Lord Rayleigh received the 1904 Nobel Prize for discovering argon. This discovery occurred when he noticed a small discrepancy between two sets of measurements of density of nitrogen gas.

In Rayleigh's time, it was thought that dry air was composed of about one-fifth oxygen and four-fifths nitrogen. Rayleigh removed all O_2 from air by passing the air through red-hot copper to make copper oxide. He then measured the density of the remaining gas by collecting it in a fixed volume at constant temperature and pressure. He also prepared the same volume of pure nitrogen by chemical decomposition of N_2O, NO, or ammonium nitrite.

The average mass of gas from air was 2.31011 g with an s of 0.000143 for 7 measurements. The mass of gas from the chemical sources was 2.29947 g with an s of 0.00138 for 8 measurements.

Was Lord Rayleigh's gas from air different from the N_2 produced chemically?

(3) Consider the cholesterol content of six sets of human blood plasma measured by two different techniques. Each sample is recorded as having a different cholesterol content. But are the two techniques yielding different answers at the 95% CL?

Sample	Method A	Method B
1	1.46	1.42
2	2.22	2.38
3	2.84	2.67
4	1.97	1.80
5	1.13	1.09
6	2.35	2.25

(4) Calculate the average, stdev, and μ at the 95% confidence interval of the following numbers: 32.76, 30.25, 34.40, 25.39, 40.15.

(5) Can any of the numbers in problem (4) be discarded from the data set?

(6) Find the answer, absolute and percent relative uncertainty for each of the following and express each answer with the correct number of significant figures.

a. $15.8(\pm 0.3) + 207.3(\pm 0.4) =$

b. $6.35(\pm 0.04) \times 10^{-3} \div 3.256(\pm 0.002) \times 10^{6} =$

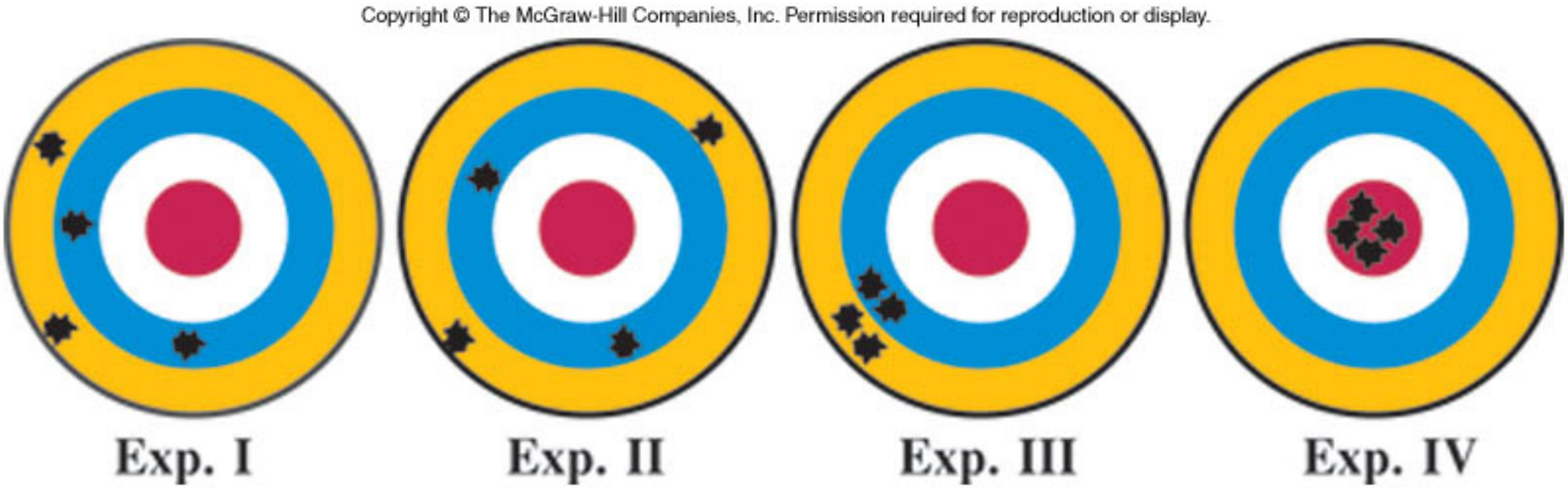
c. $4.2817(\pm 0.0003) \times 10^{-3} \div \{12.62(\pm 0.02) - 5.312(\pm 0.003)\} =$

Measurements and Calculations

- Analytical Precision: How close a group of measurements are to each other. (reproducibility)
- Analytical Accuracy: How close a measurement is to the "true" value
- Analytical Sensitivity: Slope of the calibration curve; ability to measure a difference between two concentrations
- Analytical Specificity: How well an assay measures the intended target vs. closely related targets

Concept Te$t

If the "bulls eye" represents the true value, which of the following is precise but not accurate?



A.	B.	C.	D.
Neither	Accurate? Consider the average.	Precise	Accurate & Precise

Diagnostic Sensitivity, Specificity, and Accuracy

Diagnostic Sensitivity: The true positive rate or recall rate. Percentage of sick people who are correctly identified as sick.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{Total with Disease}} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

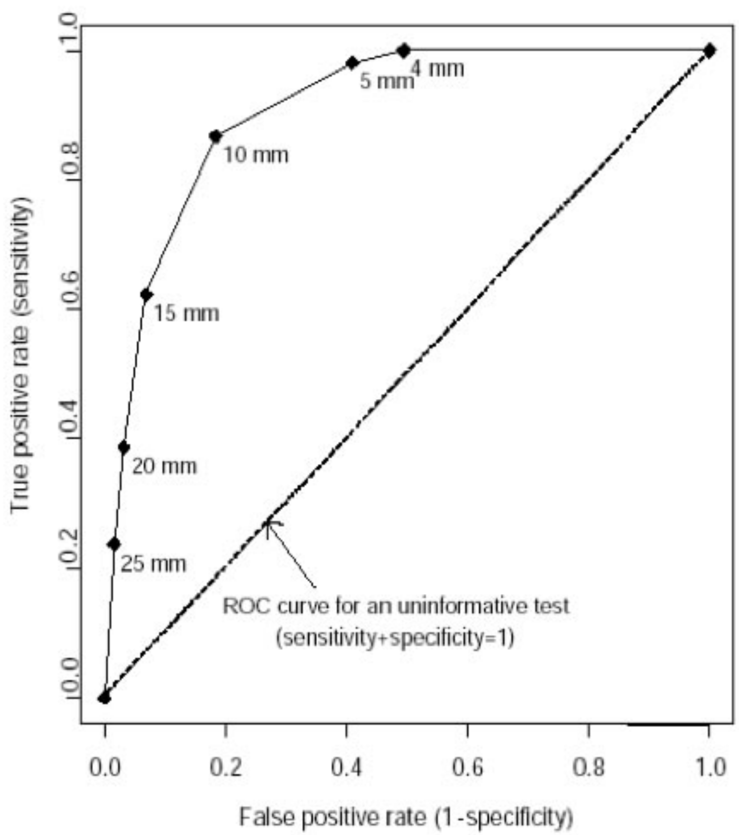
Diagnostic Specificity: The true negative rate. Percentage of healthy people who are correctly identified as healthy.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{Total without Disease}} = \frac{\text{True Negatives}}{\text{True Negatives + False Positives}}$$

Diagnostic Accuracy: Ability to identify diseased cases from normal cases. Typically defined by the area under the ROC curve (Receiver Operator Characteristic)

ROC Curves in Diagnostics

ROC Curve: Plot of Sensitivity (True+ Rate) vs. 1-Specificity (False+ Rate) at different cut-off points (for instance, concentrations or tumor sizes).



Most accurate test has an area under the curve = 1.

Data for a uterine cancer test that measures the thickness of the uterine wall.

If we say the thickness must be 20 mm to conclude cancer, we will only correctly identify people with cancer 40% of the time because patients with much smaller thicknesses often have cancer. However, we will hardly ever tell someone who doesn't have cancer that they do (few false positives).

If we say that any thickness above 4 mm indicates cancer, we will correctly identify 99% of cancers, but we will also increase the number of false positives.

Significant Figures

For multiplication and division:
Your answer should be reported to the least number of **significant figures** used in the calculation.

Ex. (2.345 x 4120)/198.2 = 48.74 = 48.7 (3 sf)

Without more information, we do not know if the 0 in 4120 is significant.

For addition and subtraction:
Your answer should be reported to the least number of **decimal places** used in the calculation.

Ex. 127.5 + 150. + 133 = 410.5 = 411 (3 sf)

Note that the “point” is used to indicate that the zero to the left is significant.

Significant Figures

Problem: A sheet of gold (area = 5200 in^2) weighs 1705 mg. What is the thickness in cm? (density = 19.32 g/cm^3)

Visualizing the problem:

If we know the area and we can calculate the volume, then we should be able to solve for the thickness.

Plan: mass(mg) → mass(g) → volume(cm^3) → thickness(cm)

Solution: $$1705\ \text{mg} \times \frac{1\ \text{g}}{1000\ \text{mg}} \times \frac{1\ \text{cm}^3}{19.32\ \text{g}} \times \frac{1}{5200\ \text{in}^2} \times \left(\frac{1\ \text{in}}{2.54\ \text{cm}}\right)^2$$

(4 sf) (∞ sf, exact) (4 sf) (2 sf) (∞ sf, exact)

$= 2.6 \times 10^{-6}$ cm (2 sf)

Concept Te$t

Express the following product to the correct number of significant figures.

$$(0.0340\ \text{yd}) \times (27.08\ \text{yd}) \times (3\ \text{ft/yd})^2$$

(The calculator gives 8.28648)

A. 8

B. 8.3

C. 8.29

D. 8.286

E. 8.2865

0.0340 yd = 3 sf

27.08 yd = 4 sf

3 ft/yd = ∞ sf

$\therefore$ answer = 3 sf

Concept Te$t

Express the following sum to the correct number of significant figures.

997.4 + 6.0321
(The calculator gives 1003.4321)

A. 1003
B. 1003.4
C. 1003.43
D. 1003.432
E. 1003.4321

997.4 = 1 decimal place
6.0321 = 4 decimal places

$\therefore$ answer = 1 decimal place

Mixed Multiplication and Addition Problems

Use all the numbers given to calculate an answer, then go back and determine the precision of each manipulation.

$$(0.108 - 0.100) \times \frac{54.2}{33.1} = 0.0130997$$

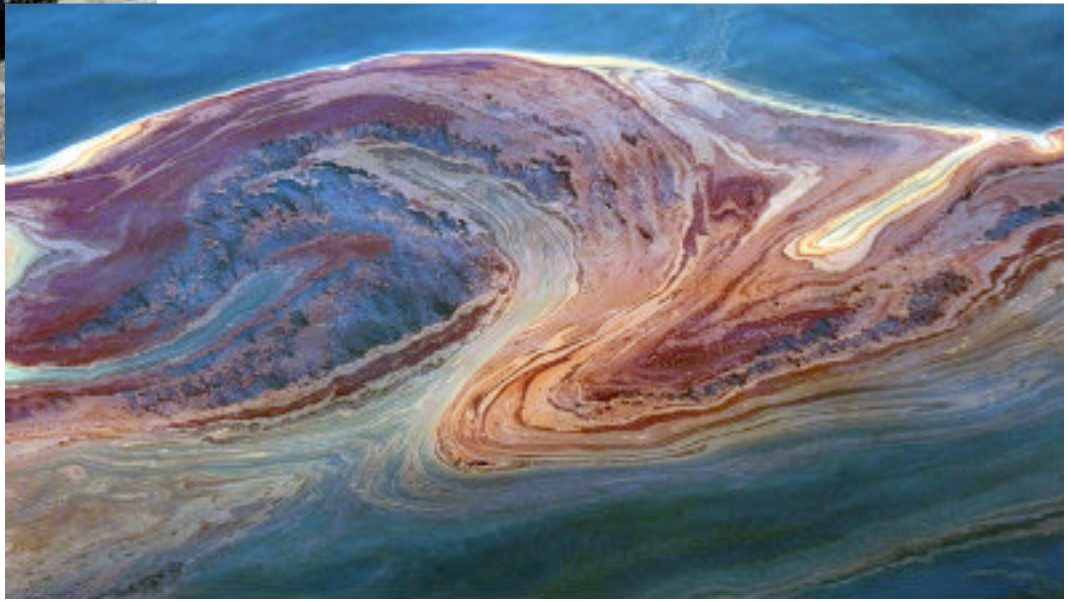
$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$(0.008) \qquad\qquad = 0.01$$

1sf

That's nice Dr. D, but how do we really think about collecting and reporting experimental data?

How would you test for the oil spilled in the gulf?



Statistical Methods of Data Analysis

No matter what measurement you are making, at a minimum you should think about making multiple measurements and reporting an average and standard deviation. (This may seem obvious to you, but this is typically not done in the research literature!) To test a diagnostic assay, you would also do multiple measurements on different days, using different reagent lot numbers and different technicians.

Why multiple measurements?

Because all measurements contain experimental error. It is never possible to be completely certain of a result. The data we report should thus be accompanied by some indication of the statistical significance of the data.

Types of Error

* Key Concepts

Random Error: Errors that arise from the effects of uncontrolled variables in the measurement. Random error has an equal chance of being positive or negative. It cannot be eliminated but it can be reduced by better experiments.

Example: How a volumetric flask or syringe is filled with solution each time. It might be filled a little high or low each time vs. the desired amount.

Systematic Error: Arises from a flaw in equipment or design of an experiment. If the same experiment is conducted twice in exactly the same manner, the error is reproducible and will always be either negative or positive. Cannot be treated statistically; must determine its source and try to minimize.

Example: pH meters are typically calibrated with a standard solution that has a known pH of 7.00. If the standard solution really had a pH of 7.09, then all of the measurements using that meter will be low by 0.09 pH units (when you think a solution is pH 7 it will really be 7.09).

Accounting for Random Error: The Gaussian Distribution

If the errors in a measurement are purely random and the measurement is repeated many times, then the results tend to cluster symmetrically about the average value.

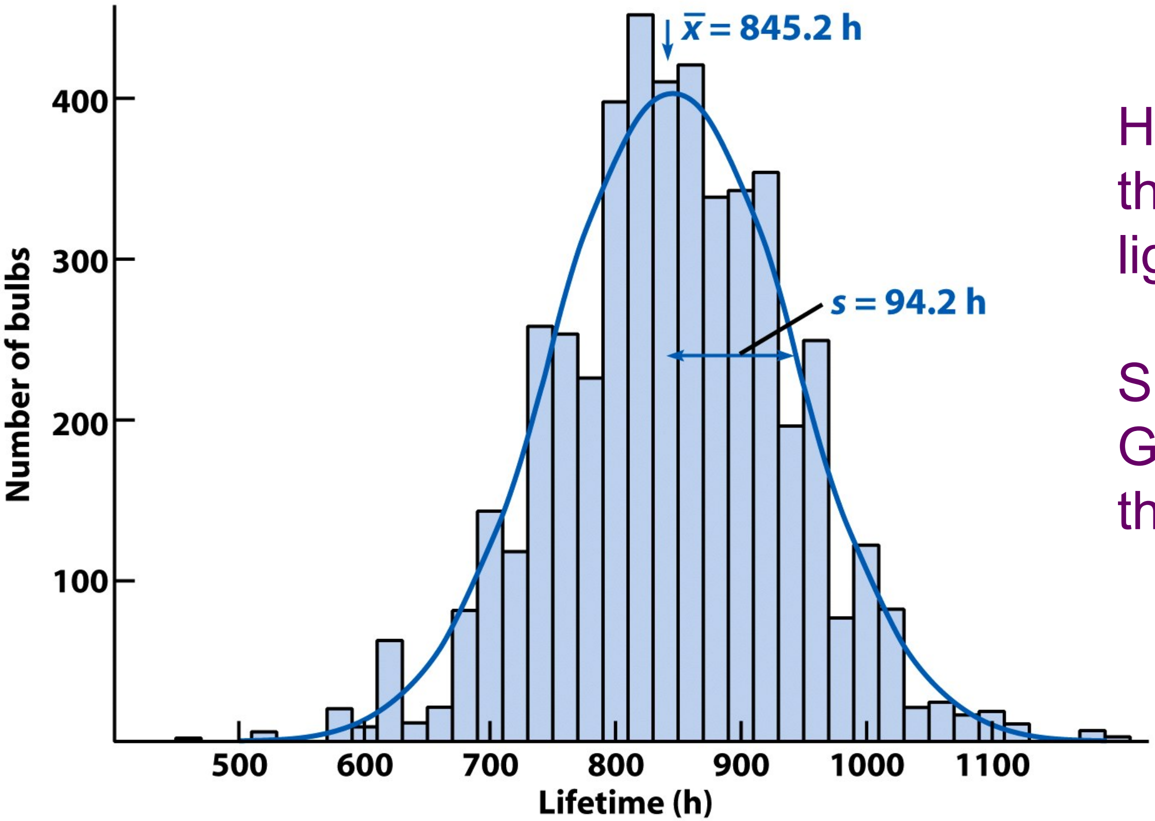


Figure 4-1
Quantitative Chemical Analysis, Seventh Edition

Hypothetical data for the lifetimes of 4,768 light bulbs.

Smooth curve is the Gaussian distribution that best fits the data.

The Gaussian Distribution

The Gaussian distribution is characterized by two parameters. The arithmetic mean, $\bar{x}$ (also called the average), is the sum of the measured values divided by n, the number of measurements:

$$\bar{x} = \frac{\sum x_i}{n}$$

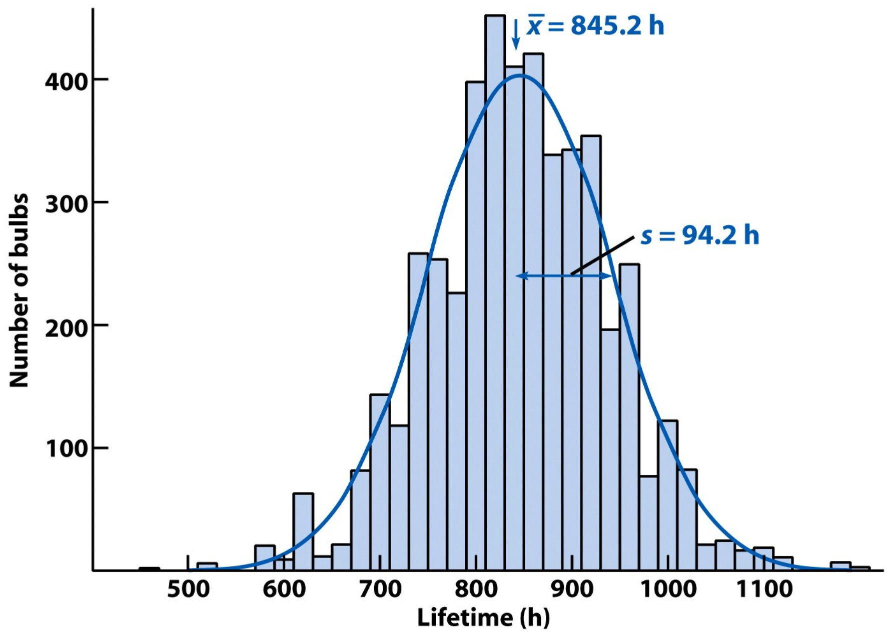


Figure 4-1
Quantitative Chemical Analysis, Seventh Edition

The Gaussian Distribution

The standard deviation, s, measures how closely the data are clustered about the mean. The smaller the standard deviation, the more closely the data are clustered about the mean and the more precise the measurement.

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

n-1 is called the "degrees of freedom" (Sometimes you will see just n in the denominator, which is for very large n)

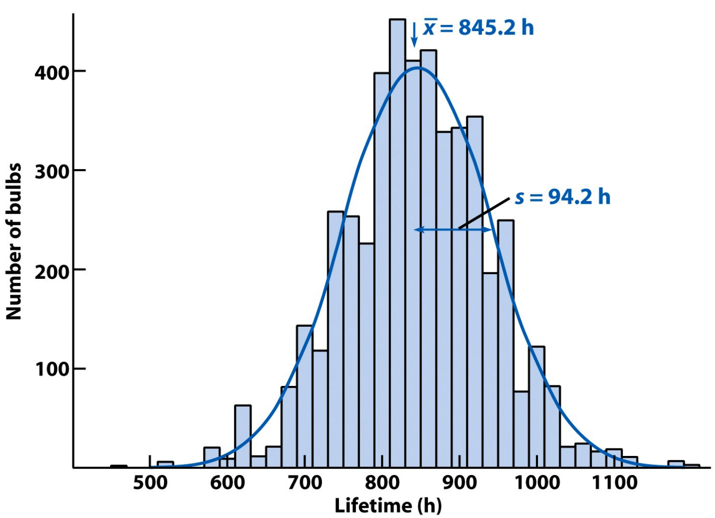


Figure 4-1
Quantitative Chemical Analysis, Seventh Edition

It is also useful at times to express s as a % of the mean value. This is the relative standard deviation or coefficient of variance.

$$\boxed{CV = 100 \times (s/\bar{x})}$$

Huge in the real world!

Degrees of Freedom

DOF for a set of data points is the number of values that must be specified to determine all of those data points.

Example:
If all you are given is the set of points 5, 7, 3, 9, 1 then the only way to specify all of the points is to know all of those values. The DOF is thus 5.

However, if you are given the set of points 5, X, 3, 9, 1 AND the average = 5, then you can determine X. Thus the DOF on the data set is N-1 = 4. If we know the average, we don't have to specify all of the points, we can calculate it. The DOF thus decreases.

Chemical Analysis in Context

You have been contracted by an environmental law firm to measure the concentration of hydrocarbons that may have formed “plumes” deep under water in the Gulf of Mexico. How do you perform the analysis so that your data have meaning?

Assume you can collect samples from the plume and determine the hydrocarbon concentration using gas chromatography.

How would you do the analysis and report the data?

Oil Plume Data Analysis

Trial	Concentration (ppm)
1	5.04
2	4.14
3	6.02
4	3.85
5	5.39

From Excel: Average = 4.888 ppm ± 0.89351553 ppm

What do you do with all of these numbers?

Oil Plume Data Analysis

From Excel: Average = 4.888 ppm ± 0.89351553 ppm

Key Concept: It is common to state the deviation to 1 significant figure and then round the average off to the same decimal place (this is not just arbitrary, it has a mathematical derivation).

4.9 ppm ± 0.9 ppm

Notice that the deviation (error) sets the precision of the average).

Oil Plume Data Analysis

A measurement with less error has more precision (more sig figs may be reported).

Trial	Concentration (ppm)
1	5.04
2	4.95
3	5.02
4	4.92
5	5.03

From Excel: Average = 4.992 ppm ± 0.05357238 ppm

4.99 ppm ± 0.05 ppm

Oil Plume Data Analysis

You may also wish to sample different depths of the plume and report an average and standard deviation for the entire plume.

Depth (ft)	Concentration (ppm)
500	4.99 ± 0.06
1000	8.5 ± 0.3
2000	22 ± 5

What is the overall average and standard deviation?

Propagation of Errors

Depth (ft)	Concentration (ppm)
500	4.99 ± 0.06
1000	8.5 ± 0.3
2000	22 ± 5

- The sum of the measurements is 35.49.
- To calculate the uncertainty <u>of the sum or difference of two or more numbers</u>, take the square root of the sum of the squares of each uncertainty: $(e_1^2 + e_2^2)^{1/2}$
- For the uncertainties above this is 5.01.
- The reported average plume concentration is

$$\text{Avg over all depths} = \frac{35 \text{ ppm} \pm 5 \text{ ppm}}{3}$$

The average is 11.6666666, but the new s must be calculated by propagating the error using a new rule for division.

Propagation of Errors

To calculate the uncertainty when performing multiplication or division, we use the fractional uncertainty.

$$\text{Fractional uncertainty} = \frac{\text{Standard deviation}}{\text{Measured Value}}$$

For an average and stdev of 4.99 ± 0.06

$$\text{Fractional uncertainty} = \frac{0.06}{4.99} = 0.012$$

The propagated error in a product or quotient is

$$(\%e_1^2 + \%e_2^2)^{1/2}$$

Propagation of Errors

Avg over all depths = $\dfrac{35 \text{ ppm} \pm 5 \text{ ppm}}{3}$

% uncertainty = 14.285%

$$(\%e_1^{\,2} + \%e_2^{\,2})^{1/2}$$

$$(14.285^2 + 0)^{1/2} = 14.285\%$$

14.285% of 11.666666666 = 1.66666 = 2

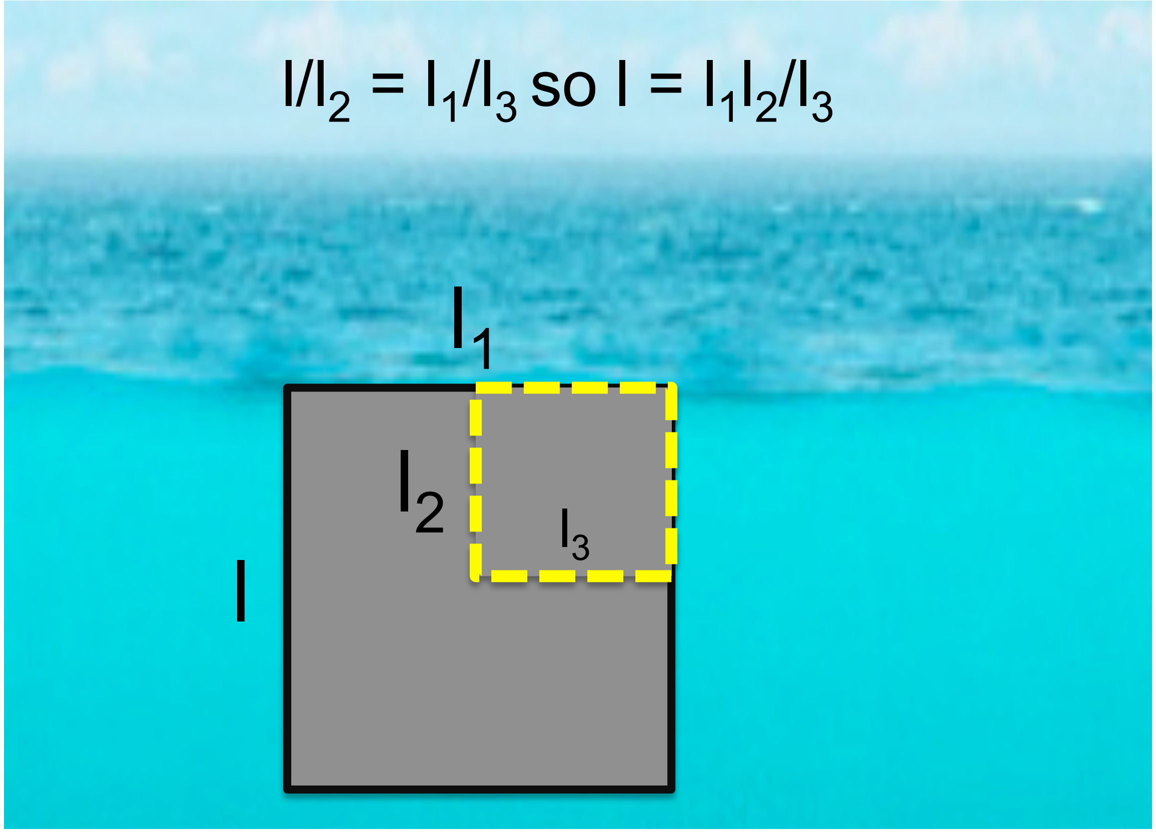
12 ppm ± 2 ppm

(Or notice in this case that we could have simply divided 5 ppm by 3 because there is no uncertainty in the number 3.)

Concept Te$t

Suppose we try to determine the length l of an oil plume by measuring three other lengths, l_1, l_2, and l_3 for which $l = l_1 l_2 / l_3$.

(Perhaps we can measure l_1 and l_2, but we can't go deep enough to measure l directly.)



Concept Test

Suppose we try to determine the length l of an oil plume by measuring three other lengths, l_1, l_2, and l_3 for which $l = l_1 l_2 / l_3$.

Suppose $l_1 = 200 \pm 2$ ft., $l_2 = 5.5 \pm 0.1$ ft., and $l_3 = 10.0 \pm 0.4$ ft.

The uncertainty in l is:

(A) 1.5 ft
(B) 5 ft
(C) 18 ft

Hint: Error propagation is the same for multiplication and division:

$$(\%e_1^{\,2} + \%e_2^{\,2} + \ldots\ldots)^{1/2}$$

Concept Test

Suppose we try to determine the length l of an oil plume by measuring three other lengths, l_1, l_2, and l_3 for which $l = l_1 l_2 / l_3$.

Suppose $l_1 = 200 \pm 2$ ft., $l_2 = 5.5 \pm 0.1$ ft., and $l_3 = 10.0 \pm 0.4$ ft.

The % uncertainties are 1%, 1.8181%, and 4%, respectively.

Uncertainty in $l = (1^2 + 1.8181^2 + 4^2)^{1/2} = 4.50616\%$

This tells us that the error is 4.50616% of l. If we want the absolute error, s, in ft., we need to determine l and multiply it by 0.0450616.

$l = 110.$ ft

4.50616% of 110. = 5 ft

Confidence Intervals

"Student's t" is a statistical tool used most frequently to express confidence intervals and to compare results from different experiments.

A "confidence interval" is another way of reporting the uncertainty in a measurement.

It is an expression stating that the true mean, μ, is likely to lie within a certain distance from the measured mean, $\bar{x}$. (The true mean and measured mean are different because we can never do enough trials to find the true mean).

The confidence interval is: $$\mu = \bar{x} \pm \frac{ts}{n^{1/2}}$$

Confidence Intervals

$$\mu = \bar{x} \pm \frac{ts}{n^{1/2}}$$

s is the stdev, n is the number of measurements, and t is Student's t, taken from a table (I'll show you in 6 slides).

The equation states that the true mean is expected to equal the measured mean ± a term that depends upon the measured standard deviation and the number of measurements.

Student was the pseudonym of W. S. Gosset, who worked at the Guinness Brewery. Guinness normally would not let its employees publish scientific papers, but felt Gosset's work was so important that they let him publish under the name Student (Biometrika, 1908, 6, 1).

Confidence Intervals

(The true mean and measured mean are different because we can never do enough trials to find the true mean.)

Concept Test

I ask 200 randomly selected CU students how much money they donated to charity over the past week. The sample mean for the 200 students is \$42.35. Therefore, I can make the following claim:

CU students donated an average of \$42.35 to charities last week.

(A) True (B) False

Confidence Intervals

(B) False

It *could* be the case that the 200 students you selected just happened to be bigger (or smaller) donors than the other CU students.

In fact, the average for all students (the *population or true mean*) could be very different from the *sample mean* of $42.35.

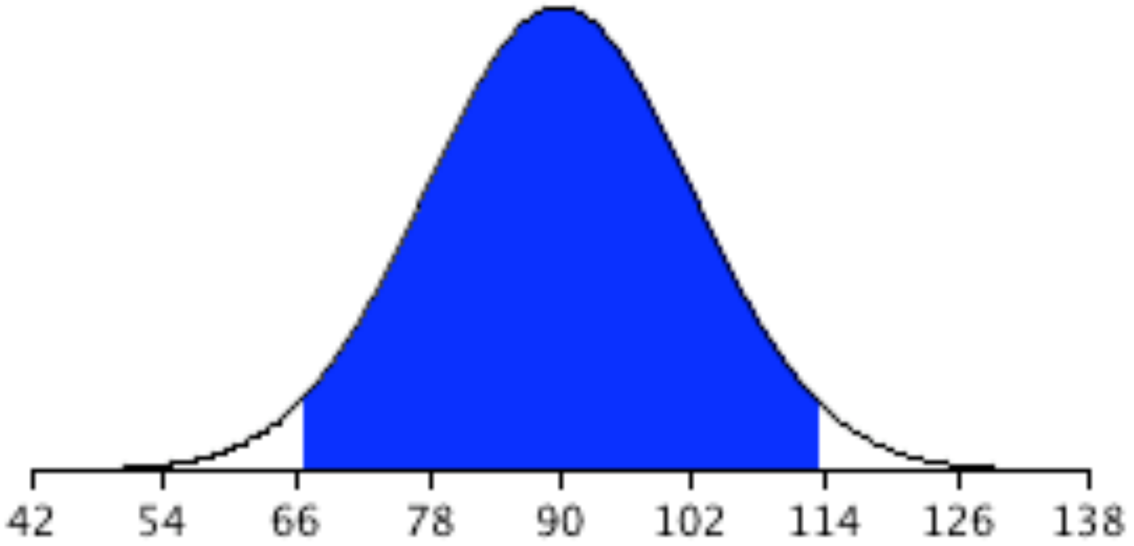
For instance, what if one student not polled happened to donate $10 million last week? The effect of including that student might be to raise the mean figure to over $1,000.

Confidence Intervals are designed to address this.

Confidence Intervals

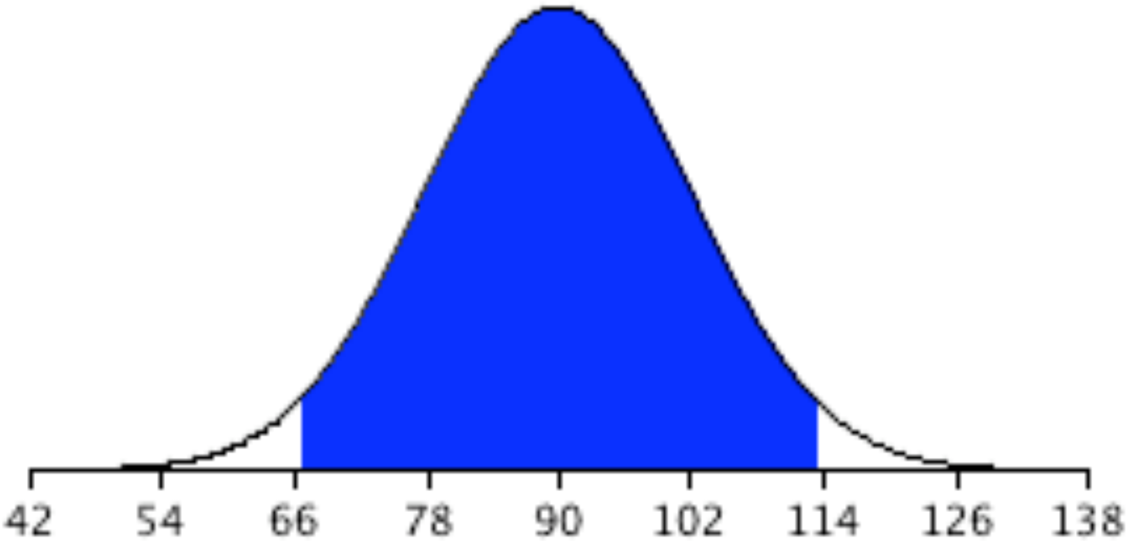
What does the CI tell us?

Assume that 9 measurements of an oil plume were recorded and the mean was 90 ppm and stdev 12 ppm.



The shaded region corresponds to the middle 95% of the distribution.

Confidence Intervals



This means that if we were to take 9 more measurements of the oil plume, we could be 95% certain that the mean of those 9 measurements would lie within this blue region. In fact, if you repeatedly take 9-trial measurements, the average of those 9 will be in the blue region 95% of the time.

A CI is an interval likely to include a parameter.

Using Confidence Intervals

Case 1: Comparing a measured result with a known (accepted) value.

Case 2: Comparing replicate measurements. Do two sets of measurements give the same result?

Case 3: Comparing two methods for analyzing a sample. Do two methods give the same result?

Using Confidence Intervals

Case 1: Comparing a measured result with a known (accepted) value.

You are developing methods to remove arsenic from water in Guatemala. To determine how well your purification method is working you must measure the amount of arsenic before and after purification. The "known" value of arsenic before purification is 3.19 mM (determined presumably by making a large # of precise measurements). Your measured values are 3.29 mM, 3.22 mM, 3.30 mM, and 3.23 mM, giving a mean of 3.26 mM and stdev of 0.04 mM. Does your method agree with the known value at the 95% confidence interval?

$$\mu = \bar{x} \pm \frac{ts}{n^{1/2}} = 3.26 \pm \frac{(t)(0.04)}{4^{1/2}}$$

t comes from a table

http://books.google.com/books?id=x5eEW76lizEC&pg=PA74&lpg=PA74&dq=student%27s+t+table+daniel+harris&source=bl&ots=9Cgdfw39hJ&sig=-7w8o7Ut2Gl4Rp_pAdlvziCysDY&hl=en&ei=LJ8bTI_hO5GCNJrIvI0N&sa=X&oi=book_result&ct=result&resnum=1&ved=0CBIQ6AEwAA#v=onepage&q&f=false

df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Using Confidence Intervals

Case 1: Comparing a measured result with a known (accepted) value.

$$\mu = \bar{x} \pm \frac{ts}{n^{1/2}} = 3.26 \pm \frac{(3.182)(0.04)}{4^{1/2}}$$

$$= 3.26 \text{ mM} \pm 0.06 \text{ mM}$$

The measured answer should fall within 3.195 and 3.325.

The known value, 3.19, was outside of the measured value.

We thus conclude that there is less than a 5% chance that our method is giving the expected answer (95% of the time we should get 3.26 mM ± 0.06 mM and we didn't this time).

Using Confidence Intervals

Case 1: Comparing a measured result with a known (accepted) value.

An alternative method:

Determine what is known as $t_{calculated}$

$$t_{calc} = \frac{|\text{Known value} - \bar{x}|}{s} n^{1/2}$$

$$t_{calc} = 3.5$$

Compare t_{calc} to t_{table} at the desired confidence limit and DOF.

If $t_{calc} > t_{table}$ then the two results are different.

With 3 DOF, 3.5 > 3.182 so the result differs from the known value.

Concept Te$t

The quality control department at Pfizer tests Lipitor tablets to ensure that they each contain the correct amount of the drug. The standard value is 12.0 mg/tablet. When twelve tablets were tested, a mean value of 11.5 mg/tablet was found. The standard deviation on these measurements was 0.6. Do these results differ from the standard value at the 95% confidence level?

(A) Result agrees with the standard value

(B) Result is different from standard value

Concept Te$t

$$t_{calc} = \frac{|\text{Known value} - \bar{x}|}{s} n^{1/2}$$

$t_{calc} = [|12.0 - 11.5|/0.6](12)^{1/2} = 2.886_7$

$t_{calc} = 2.887 > t_{table} = 2.201$

So data are statistically different!

Concept Te$t

At which CL are the data the same?

(A) 50%

(B) 90%

(C) 99%

Using Confidence Intervals

What about a different confidence level?

$$t_{\text{table, 99\%}} = 5.841$$

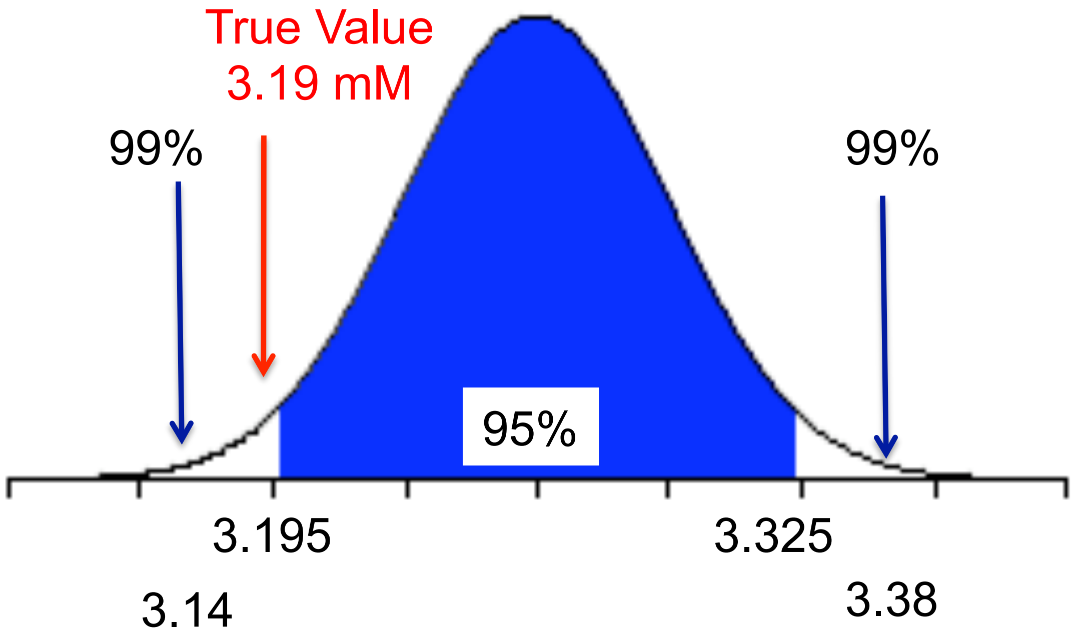
$$\mu = \bar{x} \pm \frac{ts}{n^{1/2}} = 3.26 \pm \frac{(5.841)(0.04)}{4^{1/2}}$$

$$= 3.2_6\ \text{mM} \pm 0.1_2\ \text{mM}$$

Now we expect our result to fall between 3.14 and 3.38.

This does fall within the known value.

Seems weird, but the higher the % confidence, the lower the stringency (agreeing at the 50% CL is better than the 99.9% CL).



Now we expect our result to fall between 3.14 and 3.38.

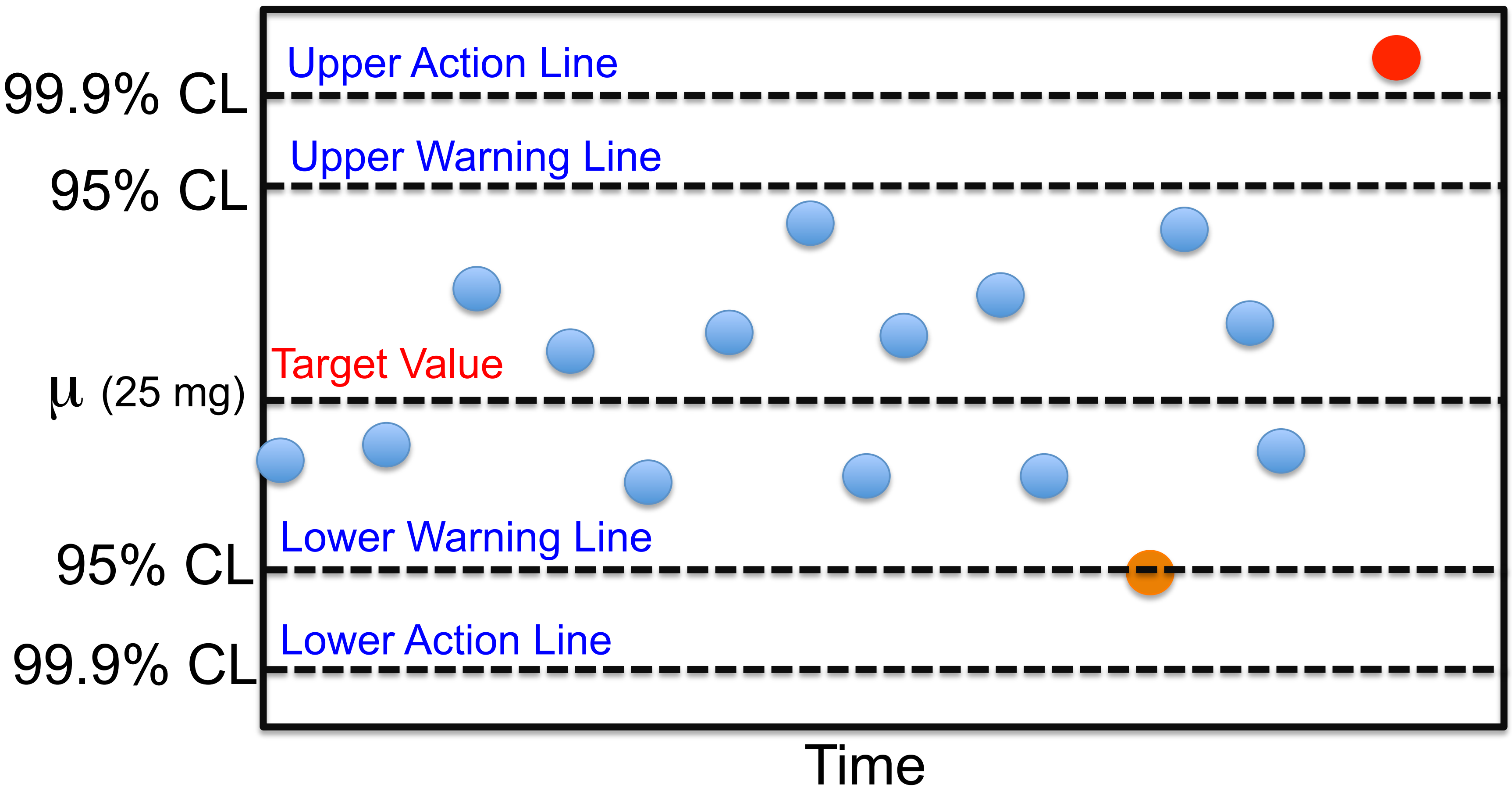
What this means is that we have actually lowered our expectations by widening the range of values that we expect our answer to fall within.

The cynic would say, “of course you will match the known value 99% of the time, you gave yourself a huge range to hit!”

By widening the target we are accepting a larger standard deviation.

Control Charts

You are QC officer for Bayer and you must monitor the production line to ensure that the target amount of ibuprofen is contained within each Advil tablet (25 mg).
A large # of trials tells you where the 95% and 99.9% CL lie.
You then test 25 tablets every hour and create a Control Chart.
You expect that 1 out of 20 trials (5%) will fall at the 95% line, so if this happens it serves as a warning that something might be wrong. Only 1 out of 1000 trials should fall on the 99.9% line, so this triggers an action (shutting down production).



Using Confidence Intervals

Case 2: Comparing replicate measurements. Do two sets of measurements give the same result?

The law firm also suspects that many plumes exist and wants to know if the hydrocarbon concentration is the same in every plume. How can you answer this question with statistical meaning?

Plume 1 (ppm)	Plume 2 (ppm)
2.30986	2.30143
2.31010	2.29890
2.31001	2.29816
2.31024	2.30182
2.31010	2.29869
2.31028	2.29940
	2.29849
2.31011 ± 0.00014	2.29947 ± 0.00138

Using Confidence Intervals

Plume 1 (ppm)	Plume 2 (ppm)
2.30986	2.30143
2.31010	2.29890
2.31001	2.29816
2.31024	2.30182
2.31010	2.29869
2.31028	2.29940
	2.29849
$2.310_1 \pm 0.0001_4$	$2.299_5 \pm 0.001_4$

While it is tempting to conclude that these are "close enough" to each other, it is possible to define "close enough" with statistical meaning.

This is accomplished with a Student's t test.

Using Confidence Intervals

For two sets of data consisting of n_1 and n_2 measurements with their corresponding averages, we calculate a value of t with the formula:

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

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

s_{pooled} is a “pooled” stdev that makes use of both sets of data.

Using Confidence Intervals

To determine if the two sets of data give the "same" number within experimental error, t_{calc} is compared to t_{table} with the appropriate degrees of freedom and desired confidence interval.

In this case the DOF represents both data sets:

$$DOF = n_1 + n_2 - 2$$

If $t_{calc} > t_{table}$, the two results are considered to be statistically different.

Using Confidence Intervals

Concept Test

The two oil plumes have the same hydrocarbon concentration at the 95% CL.

(A) True

(B) False

Using Confidence Intervals

$s_{pooled} = 0.00103$

$t_{calc} = 18.35$

DOF = 11

$t_{table} = 2.210$. $t_{calc} >> t_{table}$ so the difference between the oil concentration in the plumes is significant.

Using Confidence Intervals

Case 3: Comparing two methods for analyzing a sample. Do two methods give the same result?

Finally, the law firm has requested that two different methods for measuring the hydrocarbon concentration in plumes be used. How can you determine if the two methods provide the same answer?

Consider 11 plume samples, each analyzed once by two methods.

Sample #	Method 1	Method 2	Difference, d
1	17.2	14.2	-3.0
2	23.1	27.9	4.8
3	28.5	21.2	-7.3
4	15.3	15.9	0.6
5	23.1	32.1	9.0
6	32.5	22.0	-10.5
7	39.5	37.0	-2.5
8	38.7	41.5	2.8
9	52.5	42.6	-9.9
10	42.6	42.8	0.2
11	52.7	41.1	-11.6

Avg d = -2.491 ± 6.748

Using Confidence Intervals

The paired t test

$$t_{calc} = \frac{|d_{avg}|}{s_d}\sqrt{n}$$

$$s_d = \sqrt{\frac{(d_1 - d_{avg})^2 + (d_2 - d_{avg})^2 \ldots\ldots}{n - 1}}$$

Using Confidence Intervals

<u>The paired t test</u>

For the two methods:

$$t_{calc} = 1.2 < t_{table} \text{ for 95\% CL}$$

The two techniques are not significantly different.

Using Confidence Intervals

Make sure you keep Cases 2 and 3 straight:

Case 2

Two sets of replicate measurements of different samples by one technique

Masses of gas isolated by Lord Rayleigh

From Air	From Chemical Decomp.
2.31017	2.30143
2.30986	2.29890
2.31010	2.29816
2.31001	2.30182

Case 3

Two methods of analysis. Single measurement on each sample is made by the two methods.

Cholesterol content (g/L)

Sample	Method A	Method B
1	1.46	1.42
2	2.22	2.38
3	2.84	2.67
4	1.97	1.80
5	1.13	1.09
6	2.35	2.25

The Q test for Bad Data

How do you know when you can throw out data from a particular trial?

Do we keep or throw out data when it fits our model or expectations?

Plume 1 (ppm)
2.30
2.31
2.31
4.31
2.33
2.29

The Q test for Bad Data

How do you know when you can throw out data from a particular trial?

Do we keep or throw out data when it fits our model or expectations?

No! We can perform a Q test:

$$Q_{calc} = \text{gap/range}$$

Gap: difference between the questionable point and the nearest value.

Range: total spread in the data.

The Q test for Bad Data

If $Q_{calc} > Q_{table}$ then the point may be rejected with XX% confidence.

# of values, not DOF

N	Q_{crit} (CL:90%)	Q_{crit} (CL:95%)	Q_{crit} (CL:99%)
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

The Q test for Bad Data

Caution: When possible the sample that yielded the rejected data point should also be re-analyzed several more times to verify that an error was made initially.

Concept Test

The much larger data point may be discarded at the 90% CL.

(A) True

(B) False

Q_{calc} = (4.31-2.33)/(4.31-2.29) = 0.98

Q_{table}, 90% = 0.560