

Chapter 4 – Statistics

Since every measurement has uncertainty associated with it, how do we interpret quantitative results?

<u>RBC count on “normal” days (cells/μL x 10^6)</u>	<u>Today’s RBC count (cells/μL x 10^6)</u>
5.1	5.6
5.3	
4.8	
5.4	
5.2	

Is the red blood cell count high? Or is the variability from random error enough to account for today’s higher measurement, assuming no systematic error.

- The question becomes how much random error is there? What is the precision of the measurement?
- This is much of what Chapter 4 addresses, we’ll come back to the above problem later when equipped to provide the answer.

Hopefully we can all agree that the lifetime of a single light bulb from a batch of the same type of light bulbs is a random process.

An experiment that produces a small standard deviation is more *precise* than one that produces a large standard deviation. Greater precision does not necessarily imply greater *accuracy*, which means nearness to the "truth."

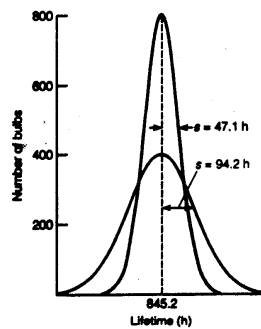


Figure 4-2 Gaussian curves for two sets of light bulbs, one having a standard deviation half as great as the other. The number of bulbs described by each curve is the same.

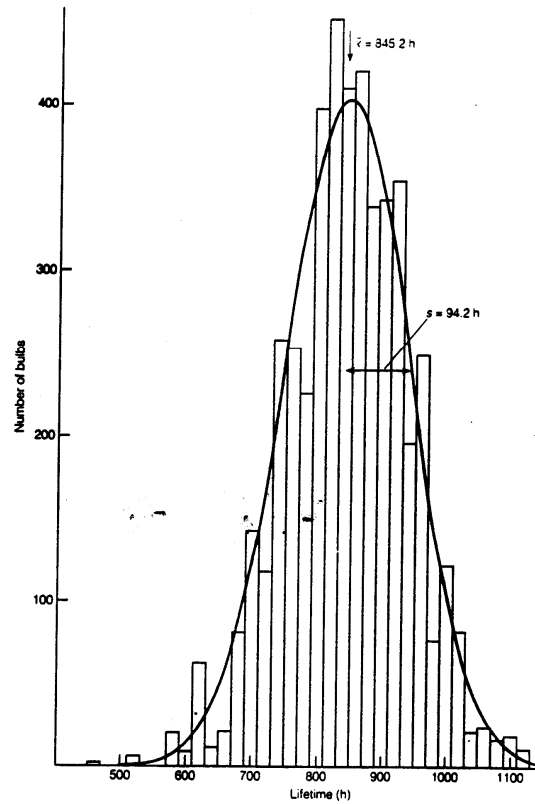


Figure 4-1 Bar graph and Gaussian curve describing the lifetime of a hypothetical set of electric light bulbs. The smooth curve has the same mean, standard deviation, and area as the bar graph. Any finite set of data, however, will differ from the bell-shaped curve.

On the right is a plot of the number of light bulbs with lifetimes in 20 h intervals. 4768 light bulbs total tested. If an infinite number of light bulbs are tested, get the smooth curve shown.

From this data 2 important parameters are obtained.

1. The arithmetic mean or average is 845.2 h.

2. The standard deviation (s) measures how closely data is clustered around the mean.

Take a look at Figure 4-2 above. The smaller the s value the more closely data is clustered around the mean. Less random error, higher precision.

It is worth figuring out how to calculate s on your calculator, which surely has a function to do this. The equation that is used is below.

Make sure you calculate the **sample** standard deviation (s) rather than the **true or population** standard deviation (σ). You can tell the difference in that the true standard deviation has n in the denominator, not $n - 1$.

It may be worth doing 1 example by hand.

Ex. A solution concentration is determined 4 times by titration with the following data:

0.1018 M, 0.1021 M, 0.1015 M, 0.1020 M

What does this mean? (No pun intended.)
 s , if well defined (4768 measurements does it, 4 measurements do not) measures the probable deviation of a result the from the mean.

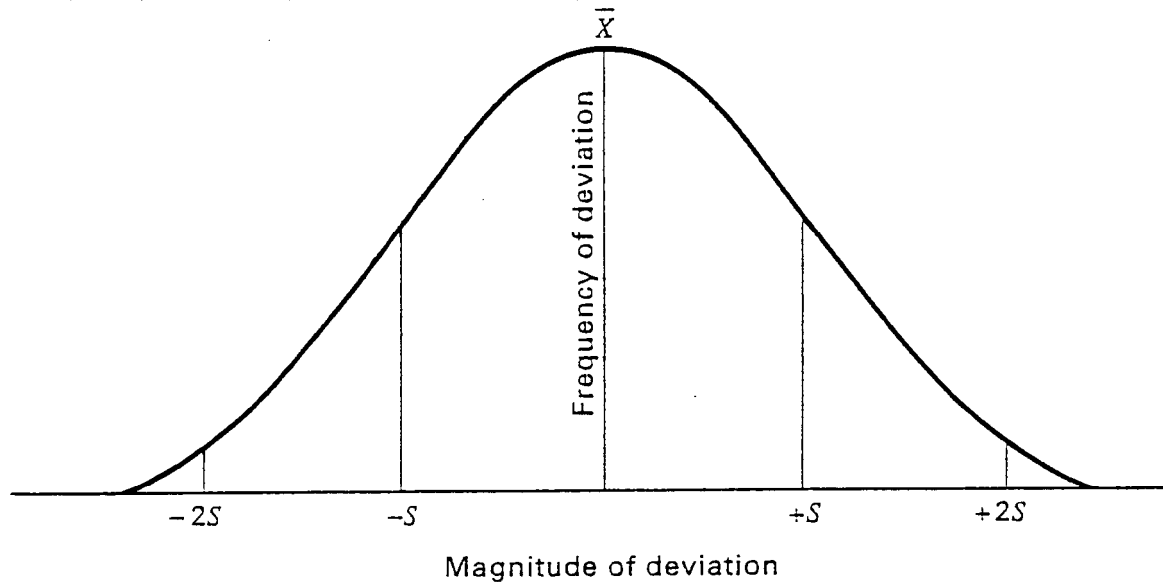


FIGURE 2.2 Normal error curve.

68 % of measurements are within $\pm 1s$ of the mean
 96 % of measurements are within $\pm 2s$ of the mean
 99.7 % of measurements are within $\pm 3s$ of the mean
 Etc. Can also do things like take slices from the Gaussian to determine the probability of a result between any 2 values, as Harris illustrates but is not covered here.

For the light bulb example, where s is well defined:
 68 % of the manufactured light bulbs will have a lifetime:
 $845.2 \text{ h} \pm 94.2 \text{ h}$

96 % of the manufactured light bulbs will have a lifetime:
 $656.8 \text{ h} \pm 188.4 \text{ h}$

Flipping this logic on its head, you can say that for any individual light bulb:

There is a 68 % chance that it will have a lifetime in the range

There is a 96 % chance that it will have a lifetime in the range

With a lot of data, it is OK to report it as the

because the Gaussian distribution, or the error curve, is well defined. Thus the standard deviation can be directly interpreted as a measure of the random error, and the probable error of another measurement.

The molarity example is different, and it is the situation we commonly find ourselves in. There are 4 measurements which are nowhere near enough to define the Gaussian. Thus the standard deviation for the molarity example does not carry as much information with it as with the light bulb example. So how do we estimate the probable error in the molarity case?

A few concepts and definitions first.

In the absence of systematic error

- As $n \rightarrow \infty$, the mean $\rightarrow \mu$, where μ is the “true” value

- As $n \rightarrow \infty$, $s \rightarrow \sigma$

Where σ is the true or population standard deviation

So the more times you measure something in the absence of systematic error;

- The closer your average () is likely to be to the true value (μ)
- and
- The closer your sample standard deviation (s) is likely to be the population standard deviation (σ).

A very important question: as n decreases, how far away is the average (), the result you get, likely to be from the “true” value (μ)? What is the probable error of another measurement?

Section 4.2 – Confidence Interval

Due to the additive nature of random error, the likely difference between the mean and μ (i.e. absolute error) is inversely proportional to the square root of the number of measurements.

Eq. 1:

Since s , standard deviation, describes the precision of a set of measurements with only random error, or the Gaussian curve width, a larger s is directly proportional to the random error.

Eq. 2:

Putting the 2 proportionalities together Eq. 3 is found:

Eq. 3:

If s is well defined by numerous measurements, then great!

Eq. 4:

With few measurements, s is an estimate of the true or population standard deviation (σ). Eq. 3 remains valid, but not Eq. 4.

Use Student's t as an additional factor to take into account the uncertainty of s when calculated with just a few measurements.

Eq. 5:

Taking away the absolute value and rearranging

Eq. 6

Table 4-2

Values of Student's t

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	2.068	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

Note: In calculating confidence intervals, σ may be substituted for s in Equation 4-6 if you have a great deal of experience with a particular method and have therefore determined its "true" population standard deviation. If σ is used instead of s , the value of t to use in Equation 4-6 comes from the bottom row of Table 4-2.

Ex. % CI in a salt measured gravimetrically, 4 measurements: 28.63 %, 28.59 %, 28.65 %, 28.68 %

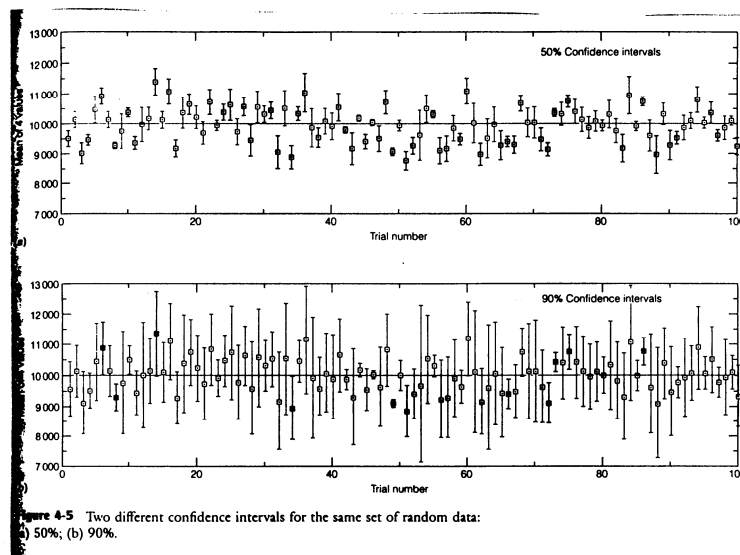
- Find the 90 % confidence interval
- Find the 95 % confidence interval

The mean is 28.64 %, $s = 0.04$ %.

What if we obtained the same mean and standard deviation from 9 measurements rather than 4?

We could have predicted this trend since uncertainty (absolute error) is reduced by making more measurements.

The meaning of a confidence interval is illustrated by using a spreadsheet to create a random error curve or Gaussian distribution with parameters that you can input. Data below was defined with $\mu = 10,000$ and $\sigma = 1000$.



4 numbers were chosen at random from the set of 100 data points described above with $\mu = 10,000$ and $\sigma = 1,000$.

These 4 numbers were used to calculate a mean, standard deviation, and confidence intervals at 50 % ($t = 0.765$) and 90 % ($t = 2.353$). This was repeated 100x.

At the 50 % confidence interval error bars encompass μ , the true value, 45 times. (Theoretically should be 50 times or 50%).

At the 90 % confidence interval error bars encompass μ , the true value, 89 times. (Theoretically should be 90 times or 90%).

Now back to the original question beginning the chapter.

Is the red blood cell count high today? Or is the variability from random error enough to account for today's higher measurement, assuming no systematic error?

<u>RBC count on "normal"</u> <u>days (cells/μL x 10^6)</u>	<u>Today's RBC count</u> <u>(cells/μLx 10^6)</u>
5.1	5.6
5.3	
4.8	
5.4	
5.2	

Calculate a confidence interval and see if it encompasses the RBC count today.

Mean = $5.1_6 \times 10^6$ cells/ μ L (subscripted # = extra sig. fig.)

$s = 0.2_3 \times 10^6$ cells/ μ L

95 % confidence interval:

99 % confidence interval:

Section 4.3 - 2 other ways to use Student's t

1. Compare quantitative data from a method and compare to an accepted answer.

a) Can be done just like we did earlier.

Ex. A new procedure you have developed measures Cu in biological samples. You buy a NIST standard which has 11.7 ppm. Measure biological sample 5x.

Mean = 10.8 ppm, $s = 0.8$ ppm

Is there systematic error in the procedure?

In general do tests at the 95 % confidence interval – but 5% chance you will be wrong!

If the confidence interval includes the accepted value then there is no systematic error. If the confidence interval does not include the accepted value then there is systematic error.

However the accepted value is not included in the 90% confidence interval. So there is a 5 – 10% chance that the method has a systematic error.

b) This same problem can be done a different way which will be useful soon. You can calculate a t value using the data, and compare the calculated t value (t_{calc}) with the tabulated t value (t_{table}).

First the confidence interval equation must be rearranged and solved for t:

2. Comparison of means. The same sample is measured with two different methods. The means are almost always different, but are they equivalent given the vagaries of random error? In other words, do the 2 methods give “the same” result, or are they different? Must use a paired t-test.

Ex. 2 methods are used to determine glucose content in blood. Do the 2 methods give equivalent results?

Method 1 (mg/dL)	Method 2 (mg/dL)
127	130
125	128
123	131
130	129
131	127
126	125
129	

$$n_1 = 7, \bar{X}_1 = 127.3, s_1 = 2.9$$

$$n_2 = 6, \bar{X}_2 = 128.3, s_2 = 2.2$$

1. Calculate s for the entire data set, called s_{pooled}
2. Calculate an experimental t value (t_{calc}) from the confidence interval equation, substituting the difference between the 2 means rather than comparing a mean to a true value (μ), taking into account n for each data set.
3. Compare the calculated t value with the tabulated t value with $n_1 + n_2 - 2$ degrees of freedom, and interpret as before.

Section 4.4 – comparison of precision, F-test

Recall that the standard deviation, s , is a measure of experimental precision. s^2 is the variance, and variance ratios are used to compare experimental precision.

Important: the largest variance always goes in the numerator! No numbers in F-test less than 1 allowed.

From the glucose content example, a paired t-test was used to compare if the results were equivalent, essentially a comparison of means. To compare the precision of these results use the F-test.

As before, if $F_{\text{calc}} < F_{\text{table}}$, the random error in the 2 methods are equivalent at the confidence level tested.

Table 4-5 Critical values of $F = s_1^2/s_2^2$ at 95% confidence level

Degrees of freedom for s_2	Degrees of freedom for s_1													
	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
2	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5
3	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.81	8.79	8.74	8.70	8.66	8.62	8.53
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.63
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.36
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.67
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.58	3.51	3.44	3.38	3.23
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	2.93
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.71
10	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.70	2.54
11	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.57	2.40
12	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.30
13	3.81	3.41	3.18	3.02	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.38	2.21
14	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.31	2.13
15	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.07
16	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.19	2.01
17	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.15	1.96
18	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.11	1.92
19	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.07	1.88
20	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.84
30	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.62
∞	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.00

Method 1, by definition that method with a larger variance, has 6 degrees of freedom in the glucose example; method 2 has 5.

Section 4.6 – discarding data and Q-test.

If you perform replicate analyses, and one analysis looks off (for no apparent reason, not because of a systematic error that you are aware of, in which case throw it out) do the following:

Generate a quotient by dividing the difference between a questionable result and its nearest numerical neighbor (gap) by the difference between the smallest and largest values (range)

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

As always, if $Q_{\text{calc}} > Q_{\text{table}}$ then the data can be discarded at the confidence level tested.

Ex. 4 samples are analyzed for % Cl^- with following results
33.03%, 33.09%, 33.05%, 32.89%

Can run 4 be rejected?

TABLE 2.3
Rejection Quotient, Q , at Different Confidence Limits^a

No. of Observations	Confidence level		
	Q_{90}	Q_{95}	Q_{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
15	0.338	0.384	0.475
20	0.300	0.342	0.425
25	0.277	0.317	0.393
30	0.260	0.298	0.372

^aAdapted from D. B. Rorabacher, *Anal. Chem.*, 63 (1981) 139.

Use this table, it is more complete than our text's Q table.

Chapter 4 Exercises A, E, F
Problems 1, 8, 9, 11, 12-14, 18-21

Section 4.7 will be covered in Labs 2 and 3, method of least squares.

Section 4.8 will also be covered in Labs 2 and 3 in the context of spectrophotometry.

Chapter 18, spectrophotometry, is next.