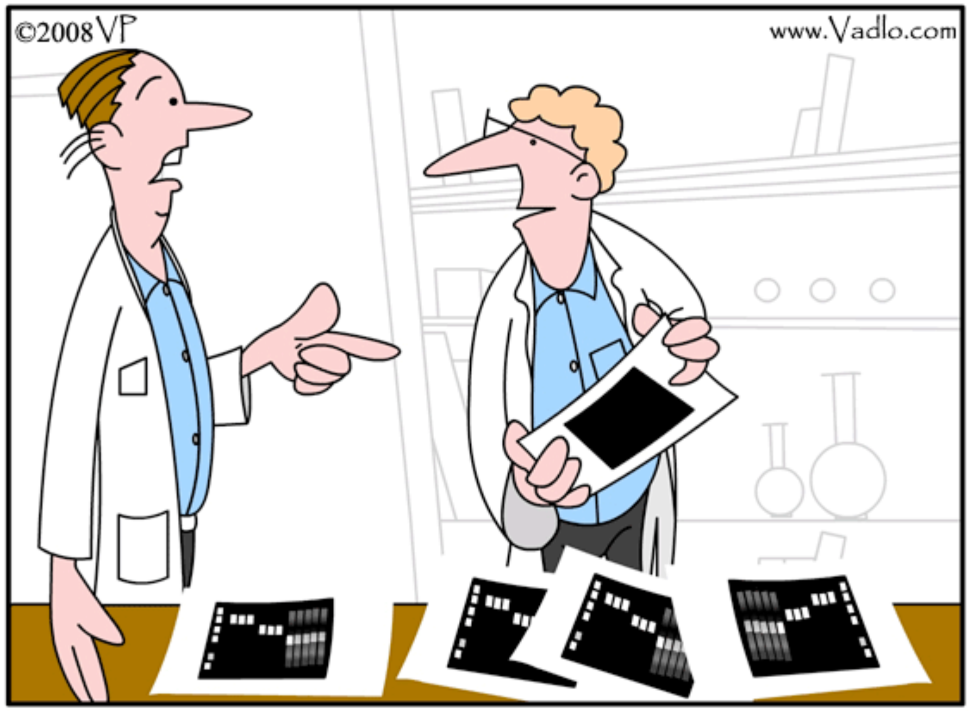
Errors in Chemical Analysis



*Data don't make any sense, we will have to resort to **statistics**.*

"Lies, damn lies, and statistics..."

Chem M3LC Lectures Week 1

Accuracy versus Precision

In casual conversation we use the words "accurate" and "precise" interchangeably - not so in Analytical Chemistry.

Accuracy versus Precision

In casual conversation we use the words "accurate" and "precise" interchangeably - not so in Analytical Chemistry.

Accuracy	*The closeness of a measurement, or the mean of multiple measurements, to its true or accepted value.*

Accuracy versus Precision

In casual conversation we use the words "accurate" and "precise" interchangeably - not so in Analytical Chemistry.

Accuracy	*The closeness of a measurement, or the mean of multiple measurements, to its true or accepted value.*
Precision	*The agreement between multiple measurements made in the same way.*

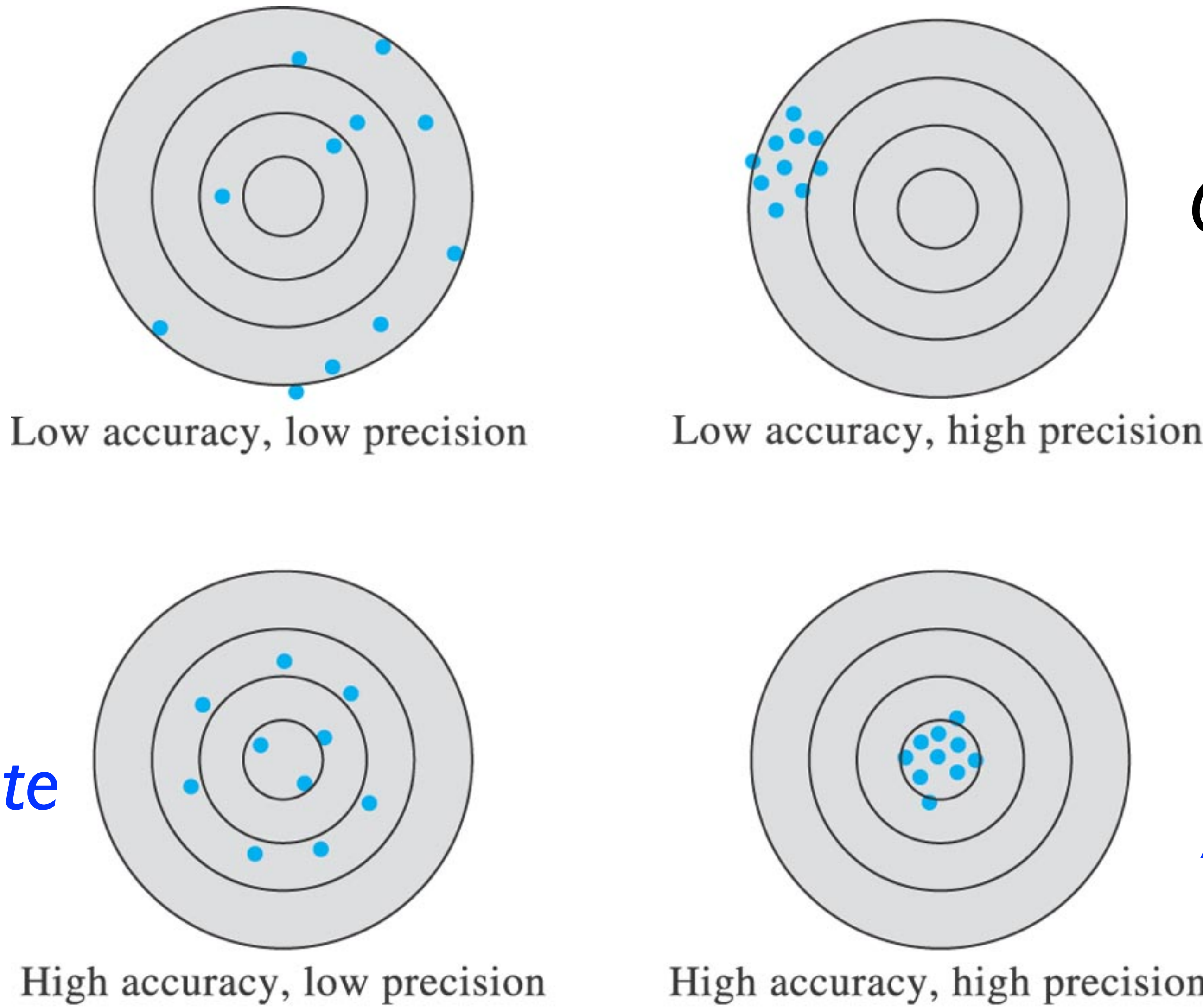
Accuracy versus Precision

Neither Accurate nor Precise

Only Precise

Only Accurate

Precise and Accurate



Determinate versus Indeterminate Error

In any measurement, there are two types of errors: determinate and indeterminate.

Determinate versus Indeterminate Error

In any measurement, there are two types of errors: determinate and indeterminate.

Determinate Error	*Errors that cause the measured mean value (x bar) for any series of measurements to be displaced, in one particular direction by one particular amount, from the true mean value (μ).*

Determinate versus Indeterminate Error

In any measurement, there are two types of errors: determinate and indeterminate.

Determinate Error

Errors that cause the measured mean value (x bar) for any series of measurements to be displaced, in one particular direction by one particular amount, from the true mean value (μ).

Indeterminate Error

Errors that cause the measured value for each measurement to be scattered randomly about μ.

In other words:

Accurate measurements have low Determinate Error.

Precise measurements have low Indeterminate Error.

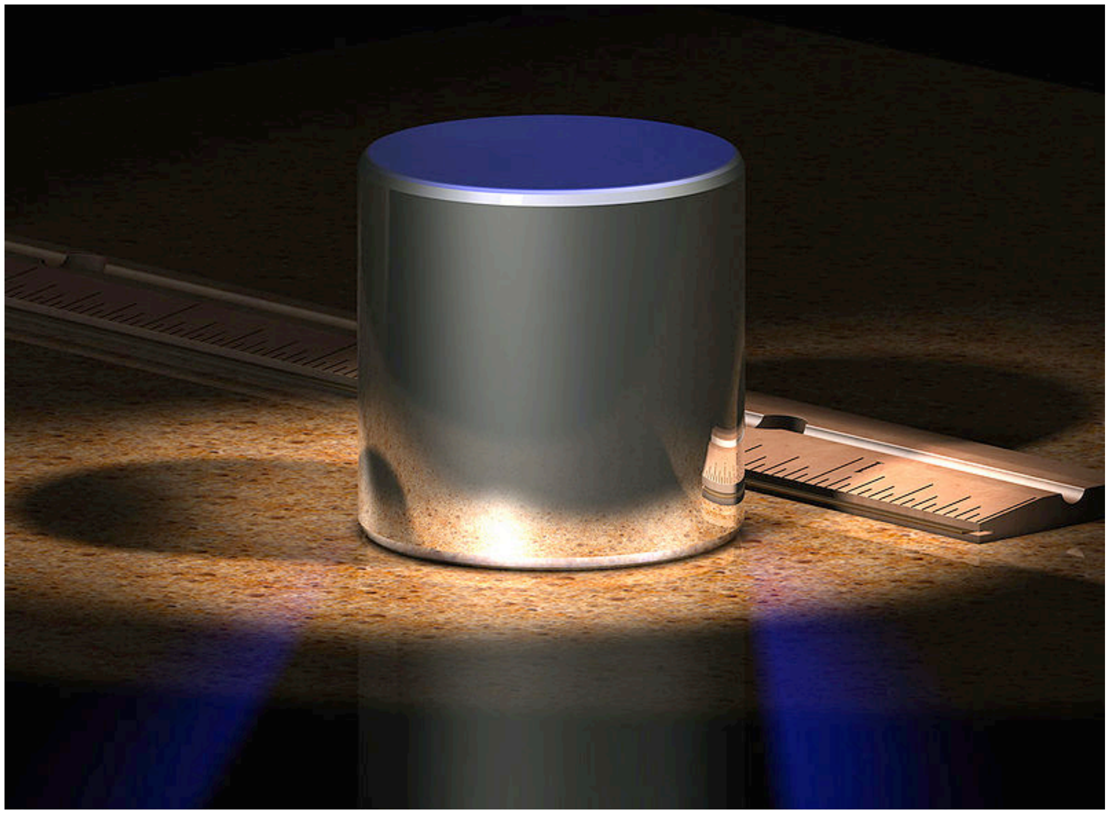
Determinate error is exposed by calibration against a sample with a known value, aka a STANDARD.

For example: standard weights

If determinate error is present at an unacceptable level, you must track down its source and eliminate it.

Determinate error is exposed by calibration against a sample with a known value, aka a STANDARD.

For example: standard weights



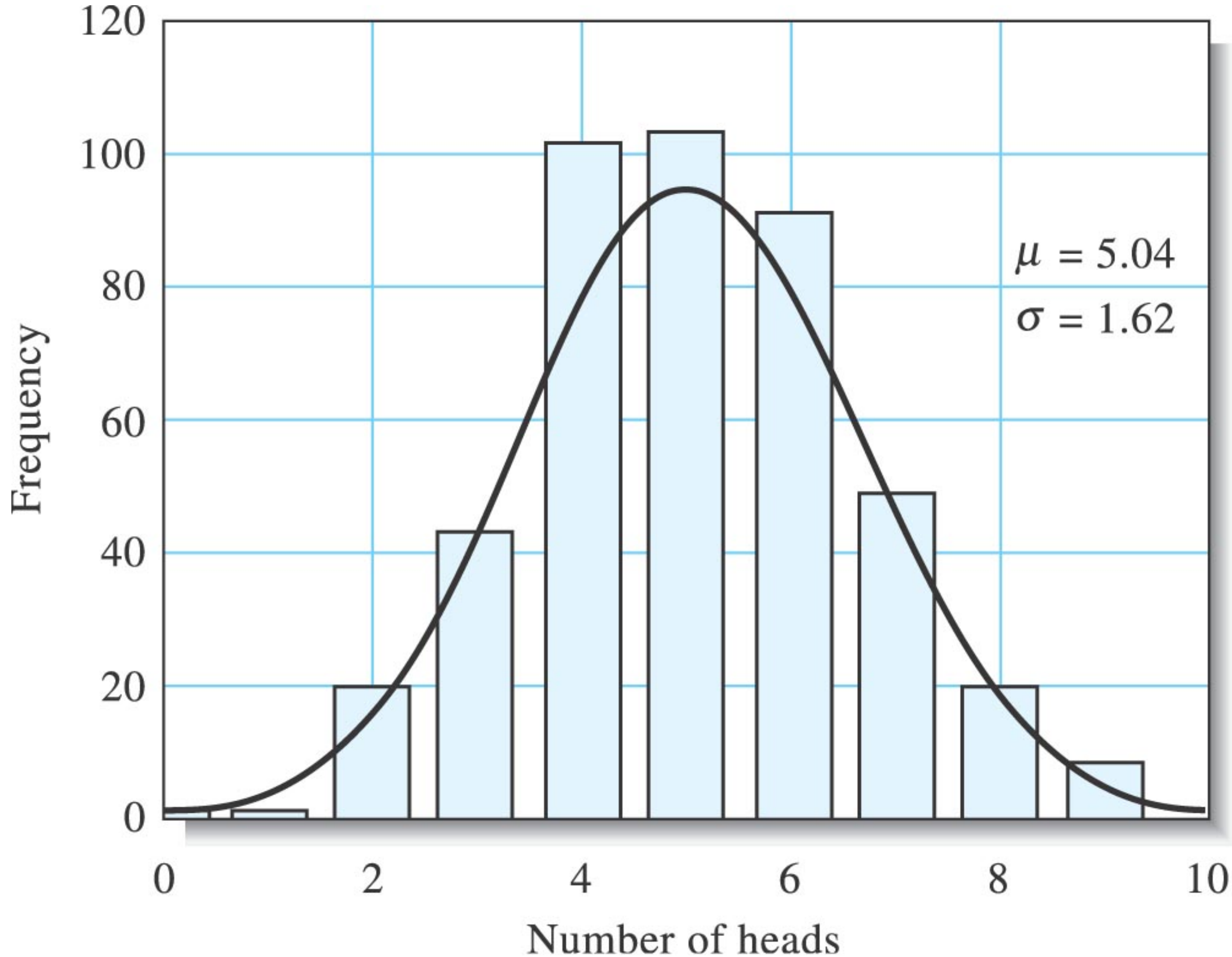
one kilogram
- exactly!

If determinate error is present at an unacceptable level, you must track down its source and eliminate it.

Understanding the Nature of Indeterminate Error

Random errors behave under the laws
of large numbers called
"Gaussian Statistics"

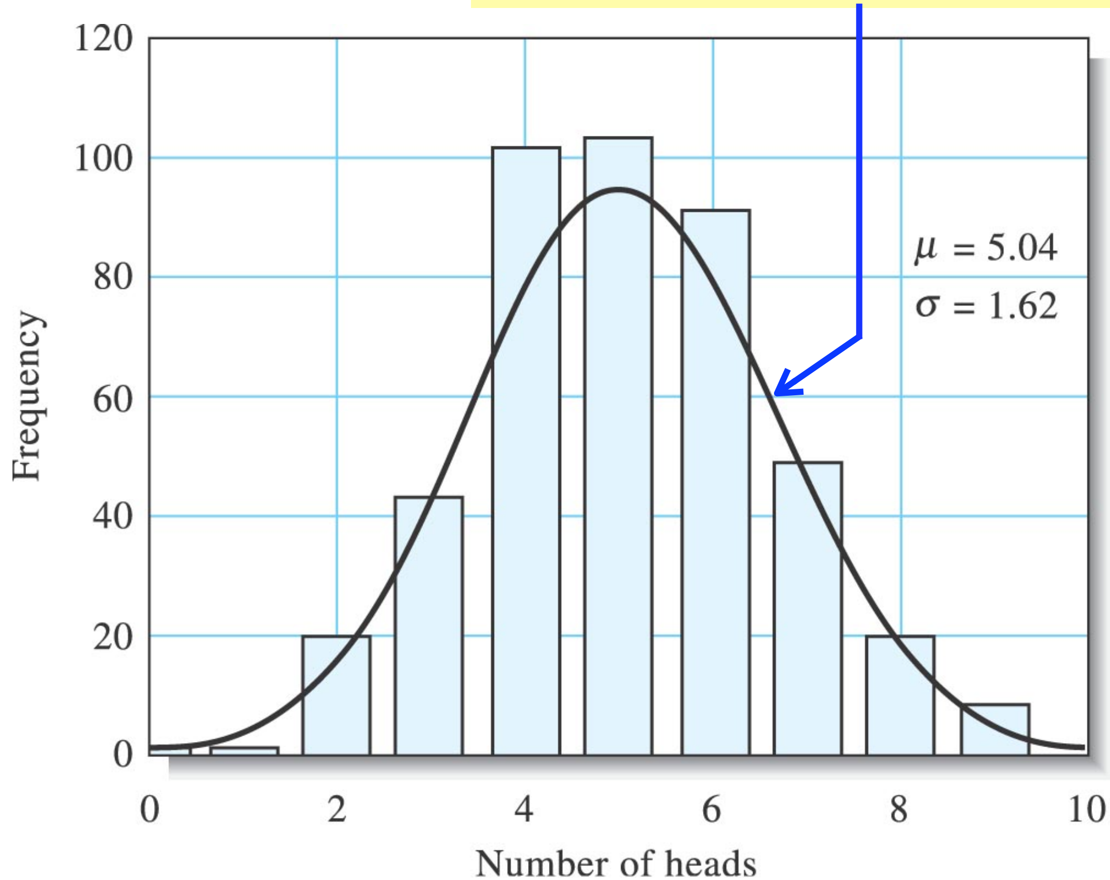
Say you flip a coin ten times, you tally the results, and you do this 395 times (3950 coin flips!)...



this distribution is Gaussian...

Gaussian distribution function

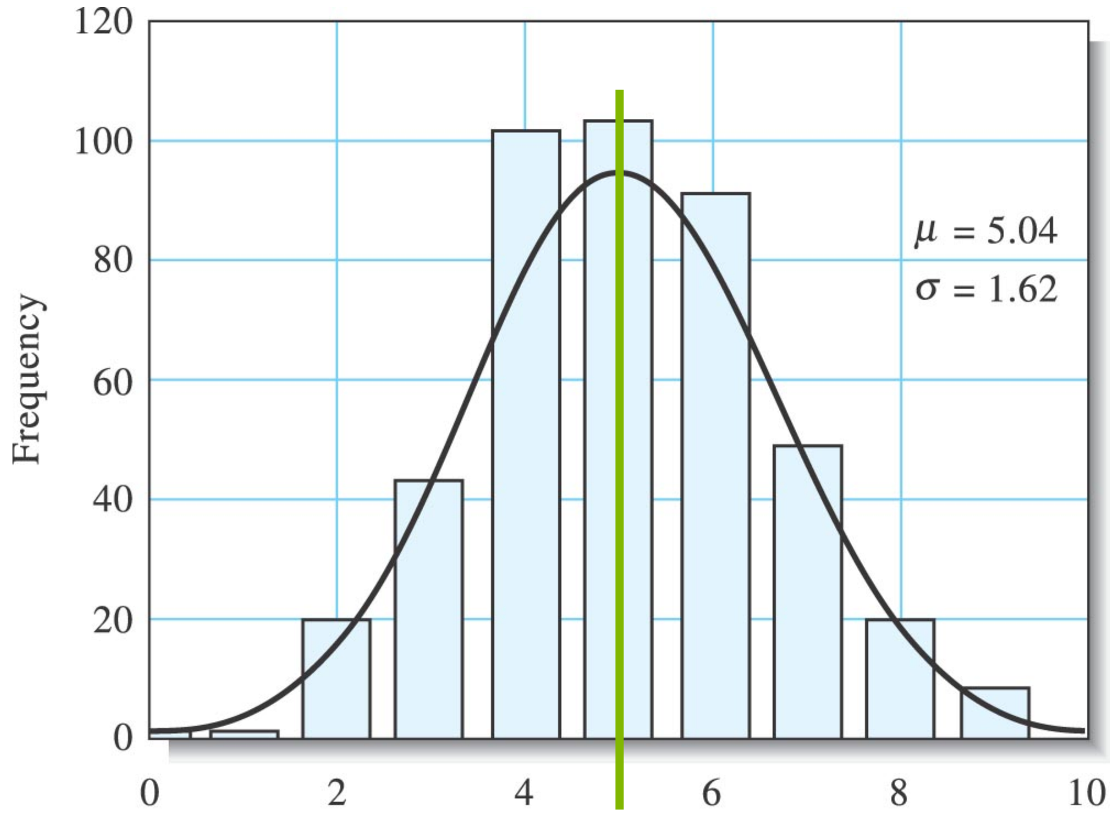
$$y = \frac{\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$



this distribution is Gaussian...

$$y = \frac{\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$

μ is called the mean

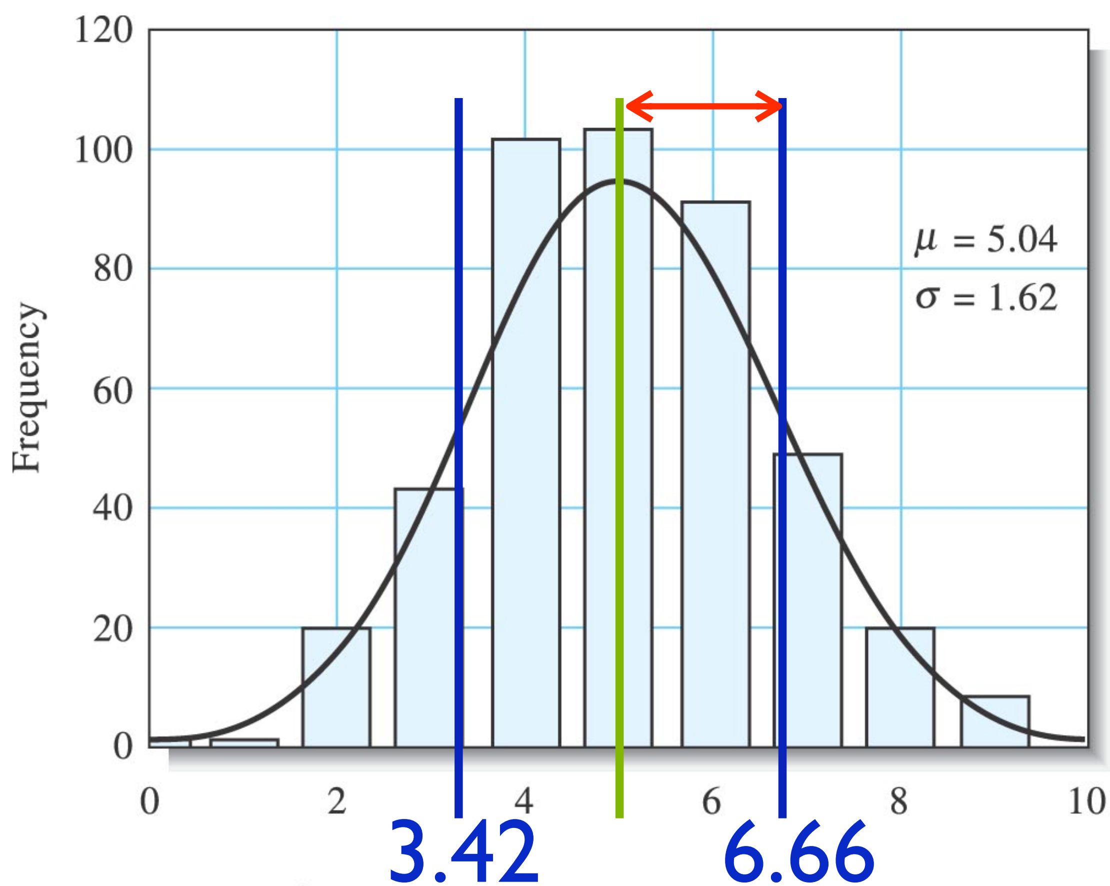


$\mu = 5.04$

this distribution is Gaussian...

$$y = \frac{\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$

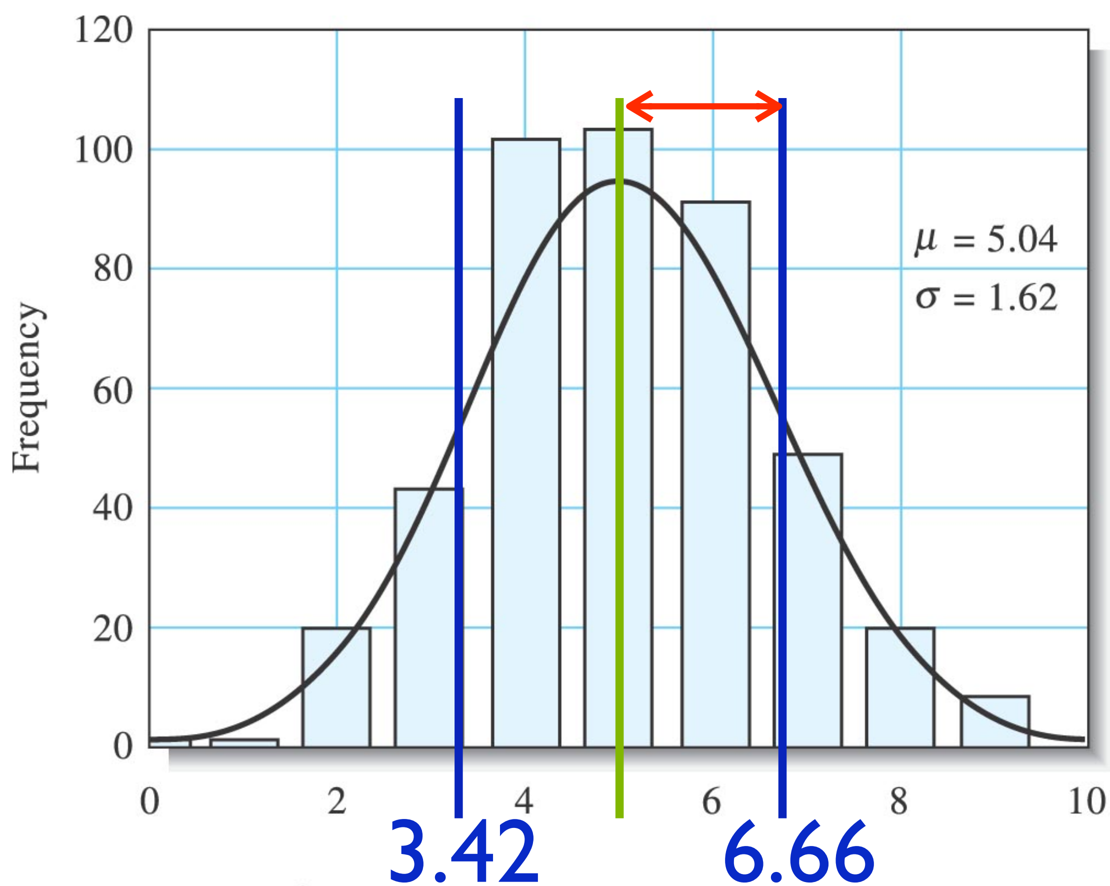
σ is called the standard deviation



this distribution is Gaussian...

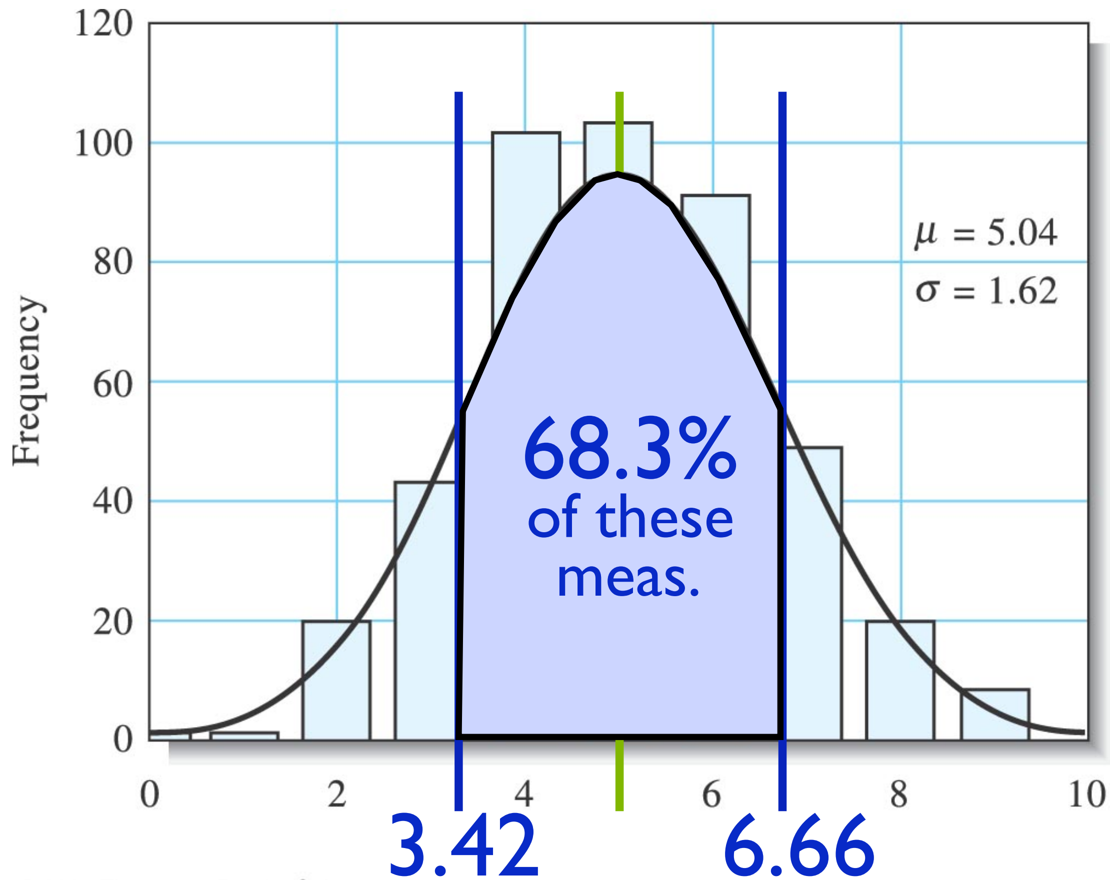
$$y = \frac{\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$

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

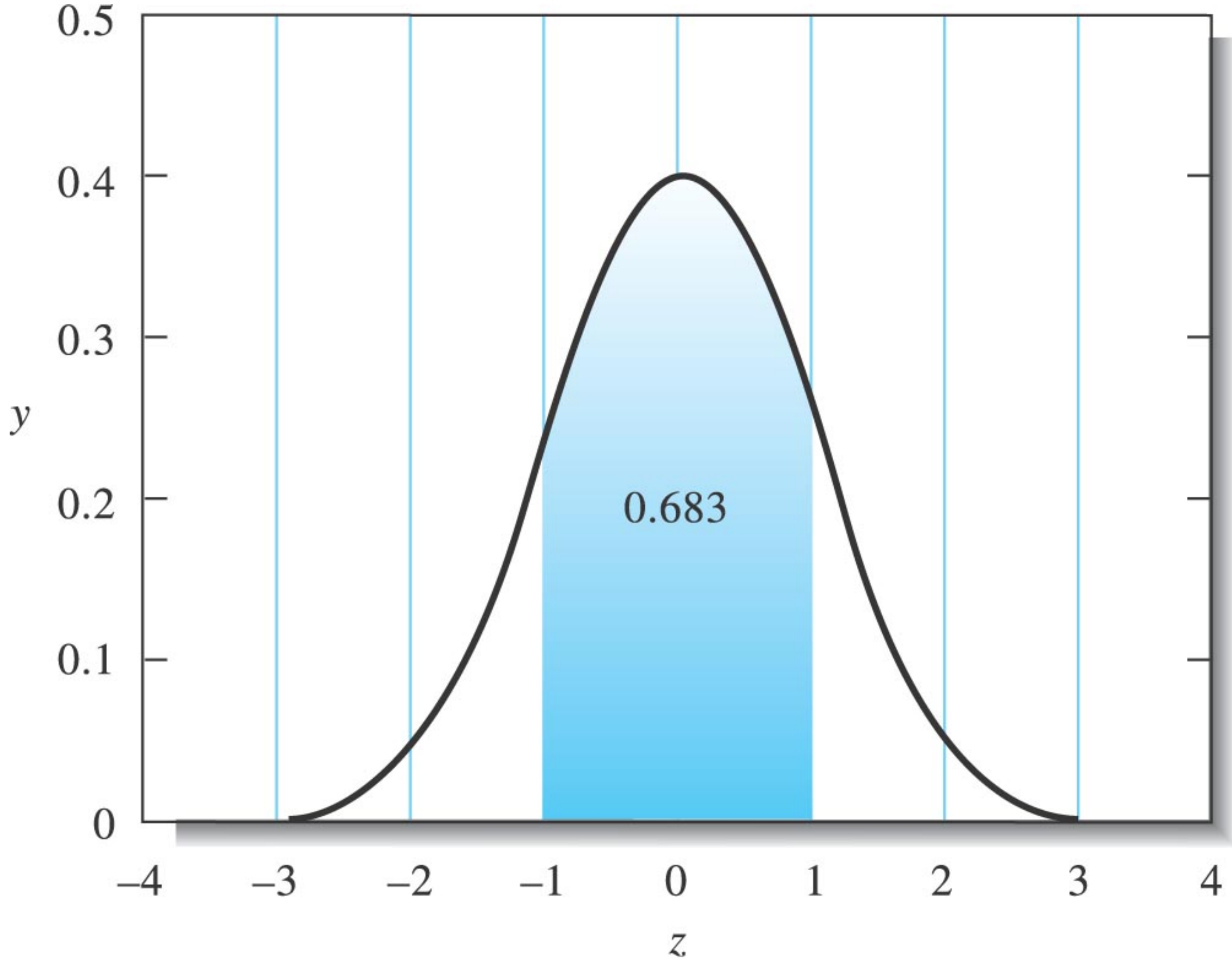


this distribution is Gaussian...

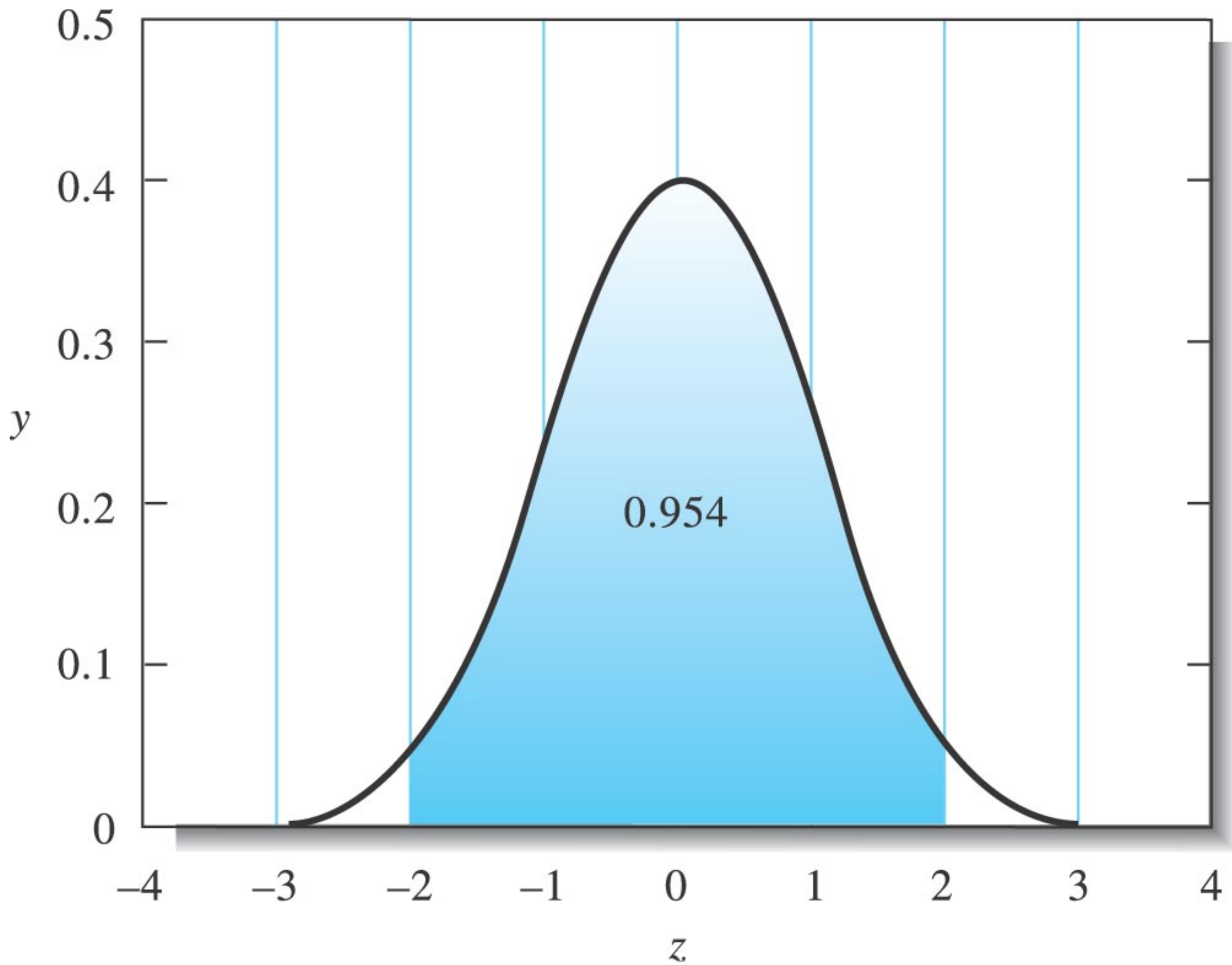
$$y = \frac{\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]}{\sigma\sqrt{2\pi}}$$



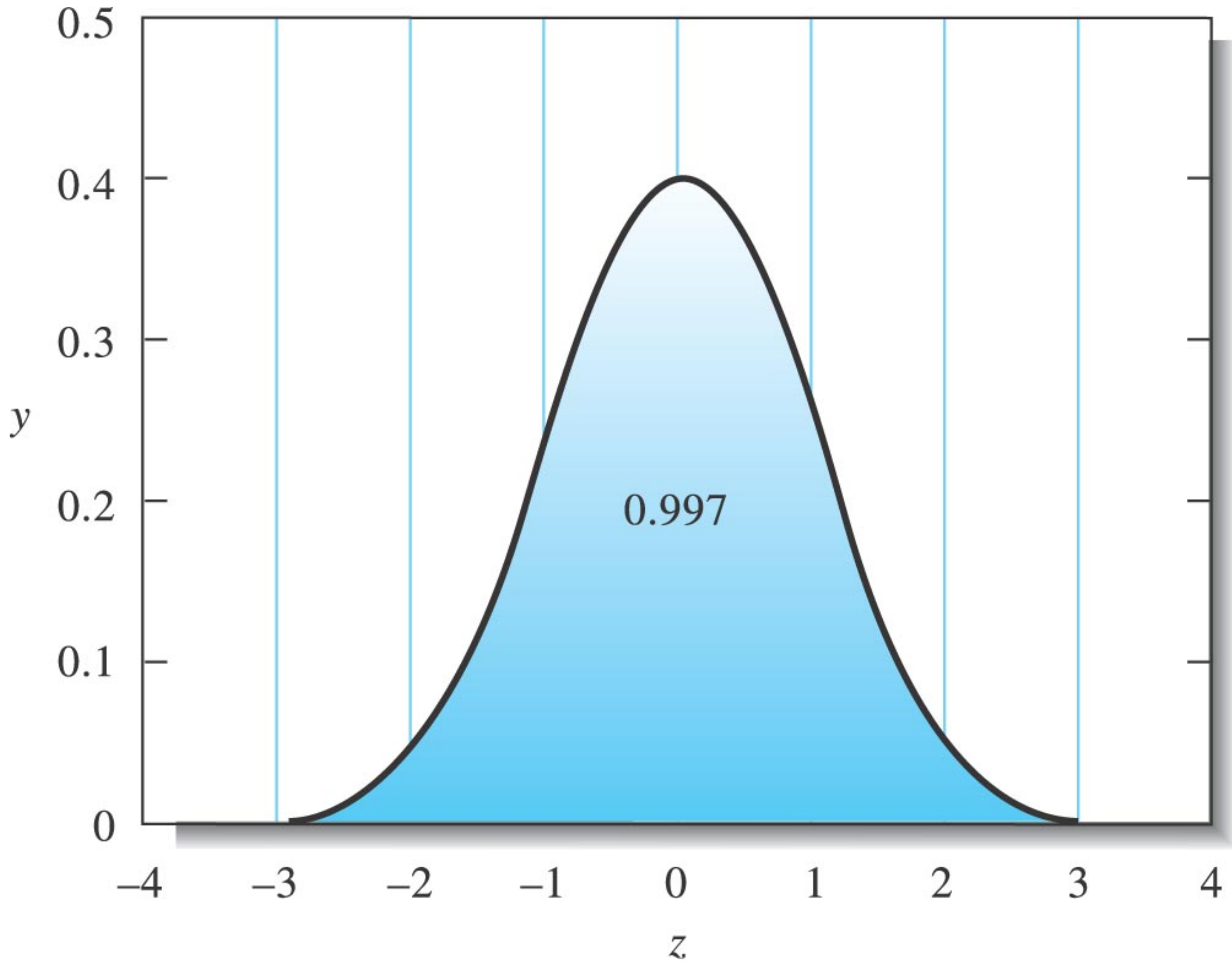
±1σ encompasses 68.3% of the measurements...



±2σ encompasses 95.4% of the measurements...



±3σ encompasses 99.7% of the measurements.



±6σ = 99.99966%

We distinguish between two flavors of means:

μ - the *true or population mean*.

You know what the mean should be (e. g. for coin flips, 50%).

We distinguish between two flavors of means:

$\boldsymbol{\mu}$ - the *true or population mean*.

You know what the mean should be.

$\overline{x}$ - the *sample mean*. ("x-bar")

You have no idea what the value of the mean should be, and must calculate x-bar.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

N = # of measurements

and this means we've got two flavors of standard deviations too:

when you know the
true or population mean:

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

and this means we've got two flavors of *standard deviations* too:

when you know the *true or population mean:*

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

...and when you don't:

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

We will use "standard deviation" to mean s, and "true standard deviation" to mean σ.

and this means we've got two flavors of *standard deviations* too:

when you know the *true or population mean:*

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

...and when you don't:

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

read more about the "degrees of freedom" in any statistics book...

N-1 is called the "degrees of freedom" for this set of measurements!

The Confidence Interval.

What is it? Well:

The Confidence Interval or Level

A textbook definition:

"The confidence interval for the mean is the range of values within which the population mean is expected to lie with a certain probability."

Huh?

Confidence intervals (CIs).

Since the mean is involved, there are two flavors of these.

i) Confidence intervals (CIs) when you know μ.

If you know μ, you can calculate σ:

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

And you report the answer as:

$$\mu = \bar{x} \pm \frac{2\sigma}{\sqrt{N}}$$

This is the 95% confidence interval for the next N measurements when μ is known.

i) Confidence intervals (CIs) when you know μ.

$$\mu = \bar{x} \pm \frac{2\sigma}{\sqrt{N}}$$

This is the "95% confidence interval"; i.e. there is a 95.4% chance that the average of the next N measurements will fall in this range.

This 95.4% is the area under the Gaussian curve for $\pm 2\sigma$.

Note that this interval decreases as N increases.

This a table of Confidence Levels from the book.

TABLE 7-1

Confidence Levels for Various Values of z

Confidence Level, %	z
50	0.67
68	1.00
80	1.28
90	1.64
95	1.96
95.4	2.00
99	2.58
99.7	3.00
99.9	3.29

In general we us z:

$$\mu \pm \frac{z\sigma}{\sqrt{N}}$$

For 95% we have just used 2 instead of 1.96 for z to make is simple.

We will say that z=2 is the "95% confidence level"

ii) Confidence intervals (CIs) when you DO NOT know μ.

If you DO NOT know μ,
you must use x-bar ($\bar{x}$)
and calculate s:

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

what is t?!

You then report the answer as:

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

This is the 95% confidence interval.

ii) Confidence intervals (CIs) when you DO NOT know μ.

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

t is called the "Student t factor"

The value of t depends on the "degrees of freedom" (dof) for your experiment. You will need to first determine the degrees of freedom, and then look up t in a table.

For a set of N measurements, the dof is N-1.

Example: if N=5, then the dof=4. We use t= 2.78 in our 95% confidence level.

TABLE 7-3

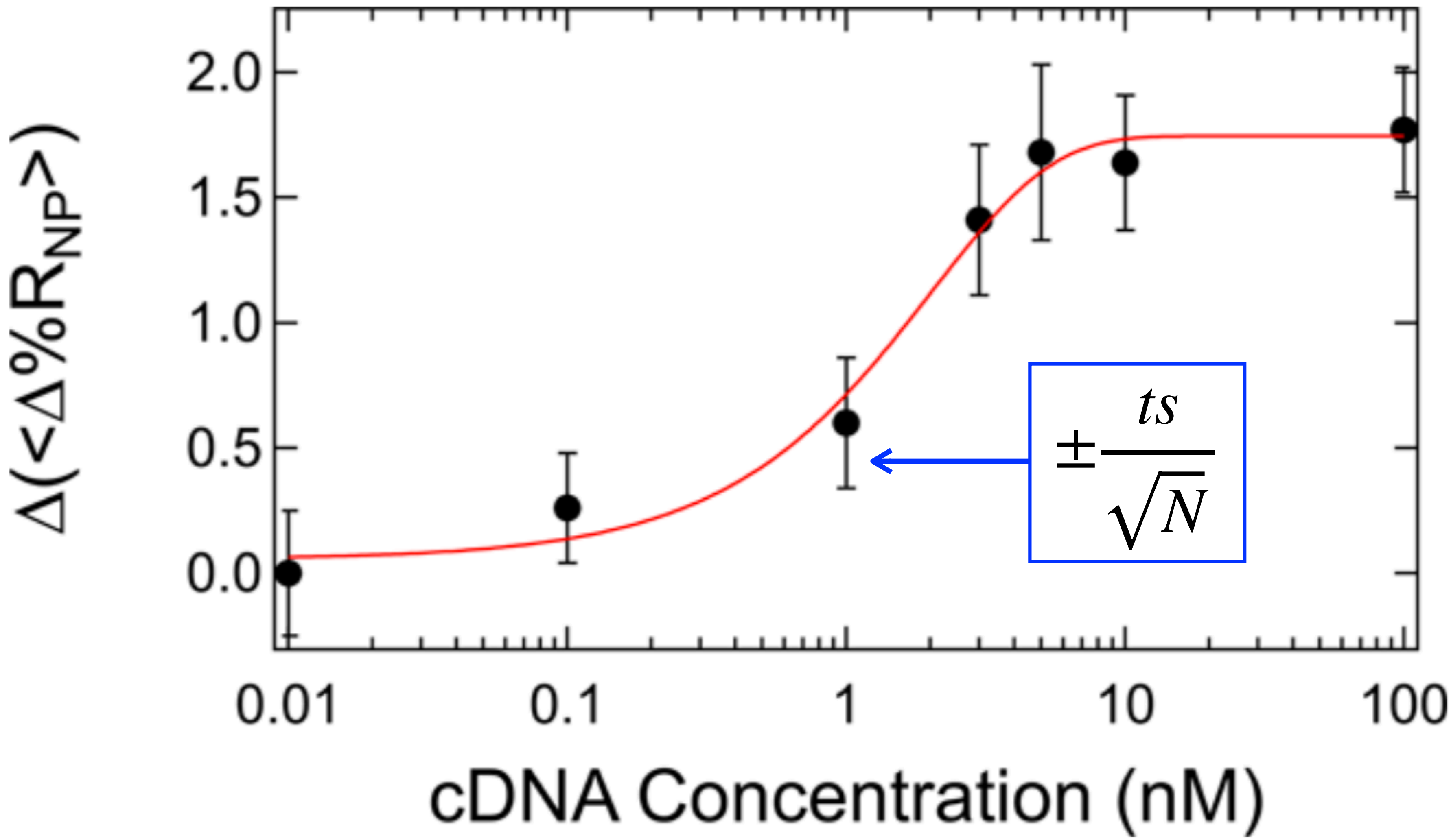
Values of *t* for Various Levels of Probability

Degrees of Freedom	80%	90%	95%	99%	99.9%
1	3.08	6.31	12.7	63.7	637
2	1.89	2.92	4.30	9.92	31.6
3	1.64	2.35	3.18	5.84	12.9
4	1.53	2.13	2.78	4.60	8.61
5	1.48	2.02	2.57	4.03	6.87
6	1.44	1.94	2.45	3.71	5.96
7	1.42	1.90	2.36	3.50	5.41
8	1.40	1.86	2.31	3.36	5.04
9	1.38	1.83	2.26	3.25	4.78
10	1.37	1.81	2.23	3.17	4.59
15	1.34	1.75	2.13	2.95	4.07
20	1.32	1.73	2.09	2.84	3.85
40	1.30	1.68	2.02	2.70	3.55
60	1.30	1.67	2.00	2.62	3.46
∞	1.28	1.64	1.96	2.58	3.29

For N>20, just set t=2.

Table from Skoog (8th ed).

When you plot data, please use error bars that are the size of the 95% confidence intervals.



Graph taken from: B. M. Matthews, Adam M. Maley, Kellen M. Kartub, and Robert M. Corn, "Characterizing the Incorporation of DNA into Single NIPAm Hydrogel Nanoparticles with Surface Plasmon Resonance Imaging Measurements," *J. Phys. Chem. C*, **123** 6090-6096 (2019).

Let's run through this again!

Calculating 95% Confidence intervals (CIs).

If you DO NOT know μ, you must use x-bar ($\bar{x}$) and calculate s and t and report this number:

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

I make 5 measurements (N =5).

Let's run through this again!

Calculating 95% Confidence intervals (CIs).

If you DO NOT know μ,
you must use x-bar ($\bar{x}$)
and calculate s and t
and report this number:

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

I make 5 measurements (N =5).

→ *i) Calculate the mean (x-bar).*

$$\bar{x} = \frac{\Sigma x_i}{N}$$

Let's run through this again!

Calculating 95% Confidence intervals (CIs).

If you DO NOT know μ,
you must use x-bar ($\bar{x}$)
and calculate s and t
and report this number:

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

I make 5 measurements (N =5).

i) Calculate the mean (x-bar).

→ *ii) Calculate the standard deviation (s).*

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

Let's run through this again!

Calculating 95% Confidence intervals (CIs).

If you DO NOT know μ,
you must use x-bar ($\bar{x}$)
and calculate s and t
and report this number:

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

I make 5 measurements (N =5).

i) Calculate the mean (x-bar).

ii) Calculate the standard deviation (s).

iii) Look up t in the table for a dof=4 (N-1).

t= 2.78

N=5, so the dof=4. We use t= 2.78 for our 95% confidence level.

Values of *t* for Various Levels of Probability

Degrees of Freedom	80%	90%	95%	99%	99.9%
1	3.08	6.31	12.7	63.7	637
2	1.89	2.92	4.30	9.92	31.6
3	1.64	2.35	3.18	5.84	12.9
4	1.53	2.13	2.78	4.60	8.61
5	1.48	2.02	2.57	4.03	6.87
6	1.44	1.94	2.45	3.71	5.96
7	1.42	1.90	2.36	3.50	5.41
8	1.40	1.86	2.31	3.36	5.04
9	1.38	1.83	2.26	3.25	4.78
10	1.37	1.81	2.23	3.17	4.59
15	1.34	1.75	2.13	2.95	4.07
20	1.32	1.73	2.09	2.84	3.85
40	1.30	1.68	2.02	2.70	3.55
60	1.30	1.67	2.00	2.62	3.46
∞	1.28	1.64	1.96	2.58	3.29

Table from Skoog (8th ed).

Confidence intervals (CIs) determine the number of significant figures (sig figs) in reporting a result.

An example:

You calculate a mean value of 26.2345 mg/mL for a series of measurements, and a confidence interval of 0.0245 mg/mL.

What do you report?

Confidence intervals (CIs) determine the number of significant figures (sig figs) in reporting a result.

An example:

You calculate a mean value of 26.2345 mg/mL for a series of measurements, and a confidence interval of 0.0245 mg/mL.

What do you report?

Correct Answer: 26.23 ± 0.02 mg/mL.

Only ONE non-zero digit here. Always. Round up or down as necessary.

Confidence intervals (CIs) determine the number of significant figures (sig figs) in reporting a result.

An example:

You calculate a mean value of 26.2345 mg/ml for a series of measurements, and a confidence interval of 0.0245 mg/ml.

What do you report?

Correct Answer: 26.23 ± 0.02 mg/mL.

The last digit here matches the place of the single digit in the CI.

Confidence intervals (CIs) determine the number of significant figures (sig figs) in reporting a result.

An example:

You calculate a mean value of 26.2345 mg/ml for a series of measurements, and a confidence interval of 0.0245 mg/ml.

What do you report?

Correct Answer: 26.23 ± 0.02 mg/mL.

NOTE: If you are performing additional calculations with a result, keep the trailing digits and use them to generate a final answer. Only round off the CI at the end of the calculation.

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

incorrect *answers:* 26.2345 ± 0.0245 mg/ml

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

***incorrect** answers:* 26.2345 ± 0.0245 mg/ml

one sig fig only.

must match single sig fig of CI

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

incorrect *answers:* 26.2345 ± 0.0245 mg/ml

26.2 ± 0.02 mg/ml

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

***incorrect** answers:* 26.2345 ± 0.0245 mg/ml

26.2 ± 0.02 mg/ml

decimal place of last digit of mean does not match CI.

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

***incorrect** answers:* 26.2345 ± 0.0245 mg/ml

26.2 ± 0.02 mg/ml

26.234 ± 0.024 mg/ml

Correct Answer: 26.23 ± 0.02 mg/ml

Also ok: 26.23 (± 0.02) mg/ml

incorrect *answers:* 26.2345 ± 0.0245 mg/ml

26.2 ± 0.02 mg/ml

26.234 ± 0.024 mg/ml

↑

one sig fig only.

Let's do an example Confidence Interval Calculation:

Suppose you have a new analytical method for measuring %Ni in a metal sample. You make four measurements of the nickel concentration and get the following results:

sample	%Ni
1	0.0329
2	0.0322
3	0.033
4	0.0323
$\bar{x}$	-
s	-

What do you report as the 95% Confidence Interval?

Let's do an example Confidence Interval Calculation:

Suppose you have a new analytical method for measuring %Ni in a metal sample. You make four measurements of the nickel concentration and get the following results:

sample	%Ni
1	0.0329
2	0.0322
3	0.033
4	0.0323
$\bar{x}$	0.03260
s	-

calculate x-bar

What do you report as the 95% Confidence Interval?

Let's do an example Confidence Interval Calculation

Suppose you have a new analytical method for measuring %Ni in a metal sample. You make four measurements of the nickel concentration and get the following results:

sample	%Ni
1	0.0329
2	0.0322
3	0.033
4	0.0323
$\bar{x}$	0.03260
s	0.00041

calculate s

What do you report as the 95% Confidence Interval?

sample	%Ni
1	0.0329
2	0.0322
3	0.033
4	0.0323
$\bar{x}$	0.03260
s	0.00041

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

what's the value of *t*?
N=4, and thus the d.o.f. = 3.

TABLE 7-3

Values of *t* for Various Levels of Probability

Degrees of Freedom	80%	90%	95%	99%	99.9%
1	3.08	6.31	12.7	63.7	637
2	1.89	2.92	4.30	9.92	31.6
3	1.64	2.35	3.18	5.84	12.9
4	1.53	2.13	2.78	4.60	8.61
5	1.48	2.02	2.57	4.03	6.87
6	1.44	1.94	2.45	3.71	5.96
7	1.42	1.90	2.36	3.50	5.41
8	1.40	1.86	2.31	3.36	5.04
9	1.38	1.83	2.26	3.25	4.78
10	1.37	1.81	2.23	3.17	4.59
15	1.34	1.75	2.13	2.95	4.07
20	1.32	1.73	2.09	2.84	3.85
40	1.30	1.68	2.02	2.70	3.55
60	1.30	1.67	2.00	2.62	3.46
∞	1.28	1.64	1.96	2.58	3.29

$t = 3.18$

sample	%Ni
1	0.0329
2	0.0322
3	0.033
4	0.0323
$\bar{x}$	0.03260
s	0.00041

solution:

95% confidence interval:

$$\mu = 0.03260 \pm \frac{(3.18)(0.00041)}{\sqrt{4}} = 0.03260 \pm 0.00065$$

You will report this value as:

0.0326 ± 0.0007 %Ni

William Sealy Gosset, Guinness Chief Brewer, 1876-1937

Photo from Wikipedia.

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

William Sealy Gosset

Biometrika, Volume 6, Issue 1, 1 March 1908, Pages 1–25.

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

Student t factor.

As an employee of Guinness, Gosset applied his statistical knowledge – both in the brewery and on the farm – to the selection of the best yielding varieties of barley.

Another researcher at Guinness had previously published a paper containing trade secrets of the Guinness brewery. To prevent further disclosure of confidential information, Guinness prohibited its employees from publishing any papers regardless of the contained information. However, after pleading with the brewery and explaining that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, he was allowed to publish them, but under a pseudonym ("Student"), to avoid difficulties with the rest of the staff.

Guinness St. James's Gate Brewery, Dublin. Estd. 1759.

Text from Wikipedia.

If you have multiple sources of error, you calculate the Standard Deviation using the method of "Propagation of Errors."

Propagation of Errors.
R. Corn - Chem M3LC.

Addition and Subtraction: sum of the squares of the absolute standard deviations:

$$y = a + b - c$$

$$s_y^2 = s_a^2 + s_b^2 + s_c^2$$

Multiplication and Division: sum of the squares of the relative standard deviations:

$$y = a\, b/c$$

$$\left(\frac{s_y}{y}\right)^2 = \left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2$$

Please see my handout on "Propagation of Errors for more information.

Comparison of Two Experimental Means

Used to determine whether two experimentally measured values are statistically different.

Experiment A:
Number of data points: N_A
Mean: x_A
Std. Dev : s_A

Experiment B:
Number of data points: N_B
Mean: x_B
Std. Dev : s_B

Are x_A and x_B statistically different?

i) Calculate the pooled standard deviation (s_P):

$$s_P = \sqrt{\frac{\sum_{i=1}^{N_A}(x_i - x_A)^2 + \sum_{j=1}^{N_B}\left(x_j - x_B\right)^2}{N_A + N_B - 2}}$$

Note that the total DOF is $N_A + N_B - 2$

Comparison of Two Experimental Means

Used to determine whether two experimentally measured values are statistically different.

ii) Calculate a t-value (t_{calc}) using the equation:

$$t_{calc} = \frac{|x_A - x_B|}{s_P}\sqrt{\frac{N_A N_b}{N_A + N_b}}$$

iii) Compare t_{calc} with the t-value in the 95% table for the total DOF (t_{table}):

If $t_{calc} > t_{table}$, then the two numbers are statistically different.

(95% Confidence Level)

Q-Test for the Rejection of Data Points

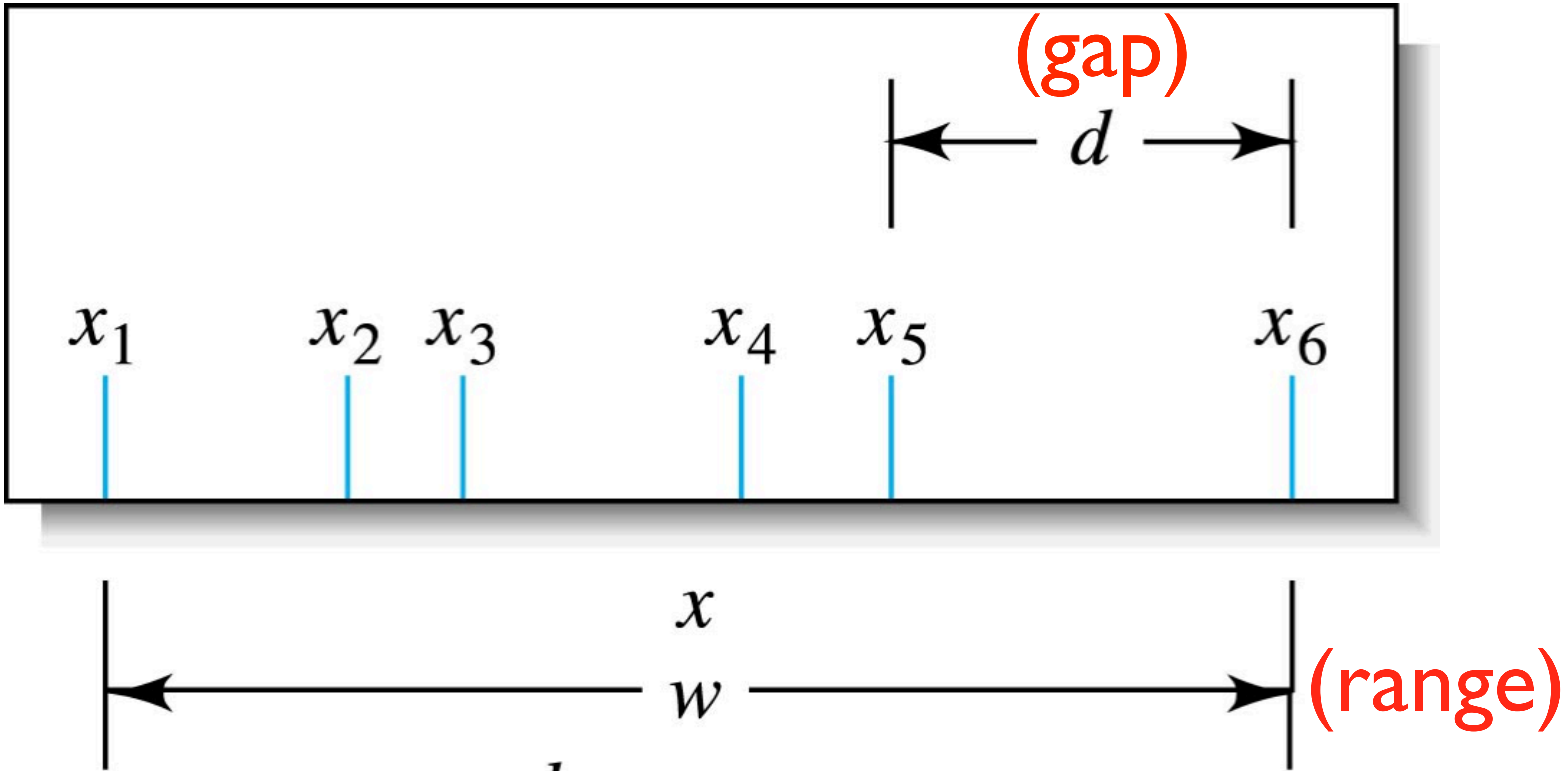
Used to determine whether a data point can be rejected on the basis of determinate error.

$$Q = \frac{gap}{range}$$

Compare to the tabulated value of Q_{crit}

reject if $Q > Q_{crit}$

Example of a Q-test:



$$d = x_6 - x_5$$

$$w = x_6 - x_1$$

$$Q = d/w$$

If $Q > Q_{crit}$, reject x_6

TABLE 7-5

Critical Values for the Rejection Quotient, Q*

Number of Observations	Q_{crit} (Reject if $Q > Q_{crit}$) 90% Confidence	95% Confidence	99% Confidence
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

*Reprinted with permission from D. B. Rorabacher, *Anal. Chem.*, **1991,** *63*, 139. Copyright 1991 American Chemical Society.

At a 95% Confidence Level, Q must be greater than 0.625 to reject the data point.