Measurements to Statistics

Collection and Interpretation of Data

Part I

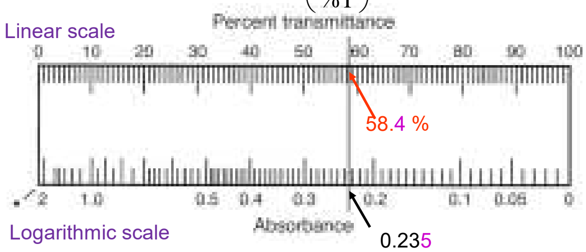
Single measurements:

Physical quantity ~ measurement: *number* *unit*
(measure)

The accuracy of the *measurement* is limited by the capability of the measuring instrument.

The last digit of the *number* of a measurement is a *considered judgment*, *estimate*. (a source of uncertainty).

$$A = \log\left(\frac{100}{\%T}\right)$$



58.4 means actual value is between 58.0 and 59.0.
0.235 means actual value is between 0.230 and 0.240.

The last digit of the *number* of a measurement is a *considered judgment*, *estimate*. (a source of uncertainty).

Better instruments will allow more precise measurements - better estimates - uncertainty can be minimized but never eliminated.

Significant figures:

measurement - 104.036 m
↑
uncertain, but *must be included* in the number

The uncertain position in the number limits the number of digits in a measurement, hence the need for the definition of significant figures.

Significant figures:

measurement - 104.036 m
↑

Expressed in scientific notation:

$$104.036 = 1.04036 \times 10^2 \quad (6)$$
$$= 0.104036 \times 10^3$$
$$= 0.0104036 \times 10^4$$

↑ pre-exponent ↑ exponent

significant figures = non-place-holding digits in a reported measurement = # of digits in the pre-exponent.

ZEROs of a number are significant in a number if (i) in the middle of a number (ii) at the end on the rhs of decimal point.

1) All non-zero digits are significant
   1.5 has 2 sig. figs.
2) Interior zeros are significant
   1.05 has 3 sig. figs.
3) Leading zeros are **NOT** significant
   0.001050 has 4 sig. figs.
   1.050 x 10^{-3} has 4 sig. figs.
4) Trailing zeros may or may not be significant
   i. Trailing zeros after a decimal point are significant
      1.050 has 4 sig. figs.
   ii. Zeros at the end of a number without a written decimal point are *ambiguous* and should be avoided by using scientific notation
      if 150 has 2 sig. figs. then 1.5 x 10^2
      but if 150 has 3 sig. figs. then 1.50 x 10^2

ALL DIGITS OF A MEASUREMENT INCLUDING THE DIGIT AT THE UNCERTAIN POSITION are called SIGNIFICANT FIGURES.

Or

Significant figures is the *proper number of digits* in the number.

Exact numbers have an *unlimited number* of significant figures (meaning there are no uncertainties, perfect – do not worry about it's sig. figs.; uncertainties creep in via numbers with decimal points).

A number (e.g. integers) whose value is known with complete certainty **(exactly)** are

a. integral powers of 10

b. numbers from counting individual objects (integers)

c. numbers from definitions and defined constants
   1 cm = 0.01 m; c = 299792458 m s^{-1} (vacuum)
   http://physics.nist.gov/cuu/Constants/

and d. integer values (in equations)

$$\text{radius of a circle} = \frac{\text{diameter of a circle}}{2}$$

Arithmetic operations:

The precision of a calculated result is determined by the number with the lowest precision.

a. addition and subtraction:

Result has decimal places same as the # with the *least* decimals.

b. multiplication and division:

Result has the same # sig. fig. as the # with the least # sig. fig.

Disregard the uncertainty of **integers and powers of 10**, because they are **exact**.

a. addition and subtraction:
i. Add/subtract numbers
ii. round off at the proper decimal place

Result has decimal places same as the # with the *least* decimals

$$\begin{array}{r} 12.0 \\ 3.0045 \\ 61.830452 \\ \hline 76.834952 \end{array} \text{ (calculator)}$$

76.834952 (roundoff); 76.8

$$\begin{array}{l} 1.632\times10^5 \\ 4.107\times10^3 \\ 0.934\times10^6 \end{array} \rightarrow \begin{array}{r} 1.632\ \ \times10^5 \\ 0.04107\times10^5 \\ 9.34\ \ \ \ \times10^5 \\ \hline 11.51307\times10^5 \end{array}$$

Express all numbers with *same* exponent

11.51×10^5

b. multiplication and division:

Result has the same # sig. fig. as the # with the least # sig. fig.

Integers and powers of 10 has no uncertainty

$$x = \frac{41.3600\times0.02328\times\frac{122.123}{1000}}{3.4842}\times100$$

x = 3.374876 = 3.375 (note *rounding off)*

i. perform the calculation
ii. round off at the proper decimal place

Rounding Off:

Look at the left most digit to be dropped

<5, no change of retained digit
>5, increase retained digit by 1
=5, increase *retained digit* by 1, if *it* is odd ←

$$69.5\text{in}\times\frac{1\text{yd}}{36\text{in}}\times\frac{1\text{m}}{1.0936\text{yd}}=1.765321466\text{m}=1.76\text{m}$$

c. logarithms: log 25.158 = 1.40070 (6)

↑ characteristic ↑ mantissa

mantissa has the same # sig.fig. as that of the number.
round off at the correct position.

d. antilogs:
will carry sig. figs. equal to the number of digits in the mantissa.

antilog (1.4007)=25.16 (4)
↑ round off such that it contains same # sig. fig. as mantissa

e. y=√x y has same # sig. figs. as x

A single run of a protocol to determine an 'unknown' is meaningless.

The protocol must be *replicated* many times, *minimum - triplicate*.

The need for replication; (a). To alleviate any doubt about the homogeneity of the material (lab sample) under 'study'.
(b) To express the *level of confidence* sought from the 'study' requires *statistical validation*, which requires a data set not a single data point.

Replications (GLP) increase the credibility of a measurement.

This course presents the *statistical procedures* used in analytical chemistry and not necessarily the statistical theories.

Replicate measurements:

Replicate a protocol (say n times) for validation.

The replicates will produce different (a spread of) values;
$x_1, x_2, \ldots, x_n$ (data set) n replications

True value = population average = μ; μ is elusive, often

For each observation x_i: absolute uncertainty $= |x_i - \mu|$

relative uncertainty $= \frac{|x_i - \mu|}{\mu}$

x_i values from the replicates is a spread of values even though all measurements are done nominally under same conditions.

Frequency of getting a specific value for a replicated measurement (data value, x_i) is not the same.

All measurements are associated with an uncertainty.

Regardless of the care taken in generating replicates, the replicates are not perfect.

The uncertainty of the measurement can be *quantified*.

Uncertainty would yield *a range of values* and the true value is expected to be within the range.

REPLICATE DATA

Mass of '10.00mL' water (g)

1	9.988 ←	18	9.975	35	9.976
2	9.973	19	9.980	36	9.990
3	9.986 ←	20	9.994 #	37	9.988 ←
4	9.980	21	9.992	38	9.971 ←
5	9.975	22	9.984	39	9.986 ←
6	9.982	23	9.981	40	9.978
7	9.986 ←	24	9.987	41	9.986 ←
8	9.982	25	9.978	42	9.982
9	9.981	26	9.983	43	9.977
10	9.990	27	9.982	44	9.977
11	9.980	28	9.991	45	9.986 ←
12	9.989	29	9.981	46	9.978
13	9.978	30	9.969 *	47	9.983
14	9.971	31	9.985	48	9.980
15	9.982	32	9.977	49	9.983
16	9.983	33	9.976	50	9.979
17	9.988 ←	34	9.983		
Mean	9.982				
Median	9.982				
Spread	0.025				
s	0.0056				

n=50

Sample mean

Uncertainty has one sig. fig. (sometimes two sig. fig.).

The origin of the spread of the values of a determined quantity, x, is attributed to random error. Random error (statistical fluctuations) arises due to factors beyond the control of the experimentalist.

Examples of random error, small changes in the ambient temperature, electrical power level variations, subjective reading of scales, electronic noise of the instrument, etc.

Instruments are associated with two (a. fundamental noise and b. technical noise) types of random error.

Instrument noise.

a. Limitations imposed by physical phenomena.

Johnson noise (in low voltage measurements due to thermal agitation of charge carriers).

Shot noise (in low intensity light measurements due to the finite nature of energy in photons).

b. Technical noise (minimized with better design of experiments).

Experimental measurements always contain some variability.

Variability leads to the inevitable ***spread*** of the observed values within a range of values, thus the quantity estimated is associated with a degree of **uncertainty** (also referred to as *error*).

Quantification and reporting the *uncertainty* (error) is essential, it also *gives the experimenter* a means to identify sources of error and thereby improve the experimental protocols.

Statistics gives us tools to accept conclusions that have a high probability of being correct, and therefore accept results; or to reject conclusions that do not satisfy statistical significance.

REPLICATE DATA

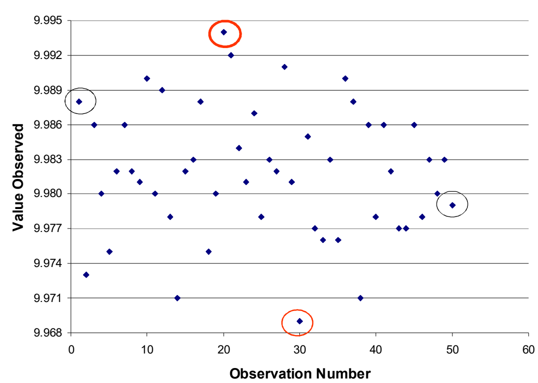
Mass of '10.00mL' water (g)

1	9.988 ←	18	9.975	35	9.976
2	9.973	19	9.980	36	9.990
3	9.986 ←	20	9.994 ##	37	9.988 ←
4	9.980	21	9.992	38	9.971 ←
5	9.975	22	9.984	39	9.986 ←
6	9.982	23	9.981	40	9.978
7	9.986 ←	24	9.987	41	9.986 ←
8	9.982	25	9.978	42	9.982
9	9.981	26	9.983	43	9.977
10	9.990	27	9.982	44	9.977
11	9.980	28	9.991	45	9.986 ←
12	9.989	29	9.981	46	9.978
13	9.978	30	9.969 **	47	9.983
14	9.971	31	9.985	48	9.980
15	9.982	32	9.977	49	9.983
16	9.983	33	9.976	50	9.979
17	9.988 ←	34	9.983		
Mean	9.982				
Median	9.982				
Spread	0.025				
s	0.0056				

n=50

Sample mean

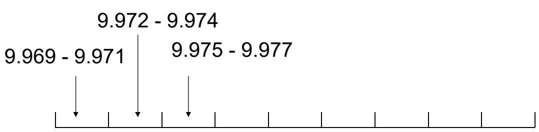
Quantitative Aspects of Random Error and Visualization of Data sets:



Quantitative Aspects of Random Error and Visualization of Data sets:

Divide the data range values into equal data intervals (bins), bin size 0.002.

Count the number of data values in the intervals.



Quantitative Aspects of Random Error and Visualization of Data sets:

Results *tend* to cluster symmetrically about an 'average value', sample mean.

FREQUENCY DISTRIBUTION OF DATA

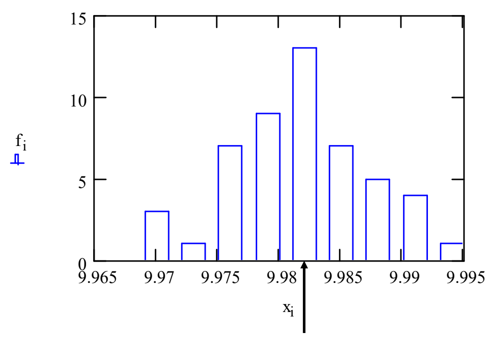
9 bins
bin size 0.002

frequency

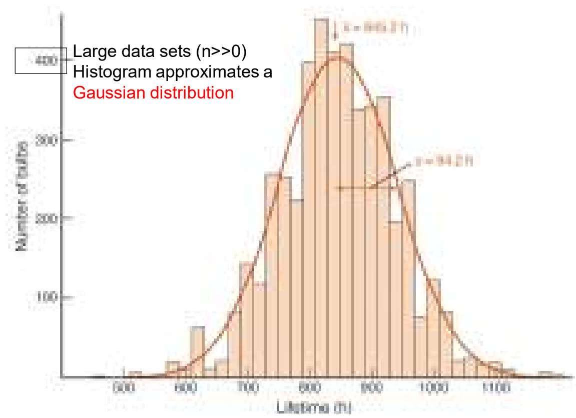
Range, mL	Count in Range	% Counts in Range
9.969 to 9.971	3	6
9.972 to 9.974	1	2
9.975 to 9.977	7	14
9.978 to 9.980	9	18
9.981 to 9.983	13	26
9.984 to 9.986	7	14
9.987 to 9.989	5	10
9.990 to 9.992	4	8
9.993 to 9.995	1	2

Step 1: Histogram
(frequency distribution)

Sample mean = 9.982

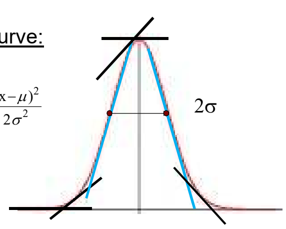


Lifetime of bulbs study



Gaussian Curve:

$$y = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The distribution of values determined from a *very large number of replications* with only random error as the source of error, conforms to a smooth curve; Gaussian shape (distribution).

Most measurements with natural error/uncertainty are distributed are Gaussians.

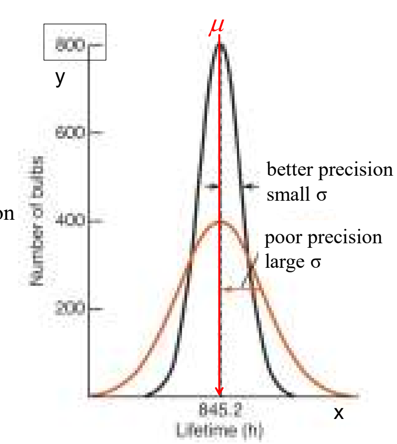
Gaussian Function

$n \to \infty$

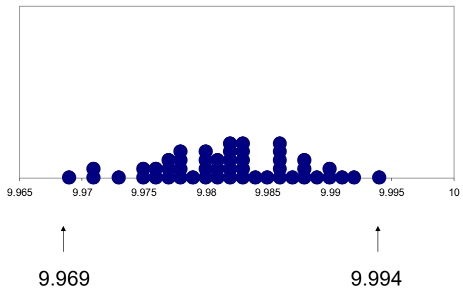
μ = true value

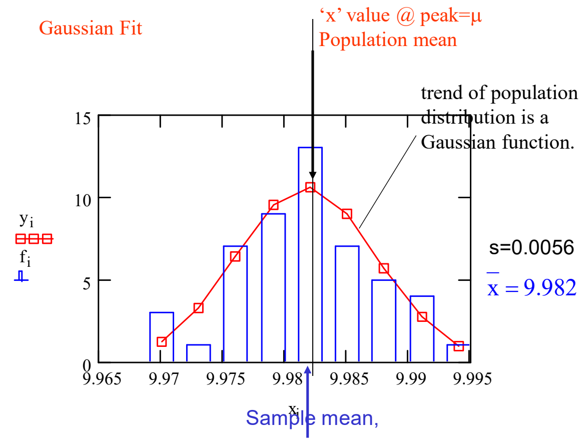
σ = population standard deviation

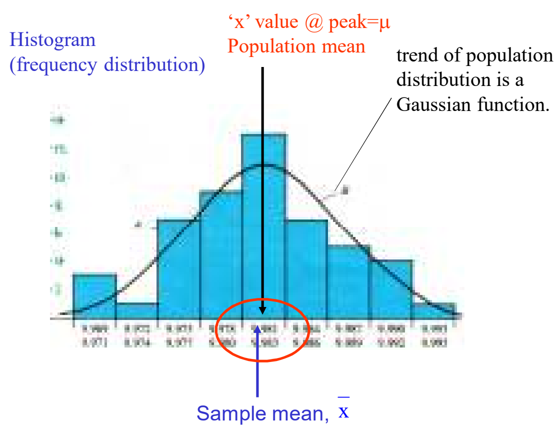
$$y = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

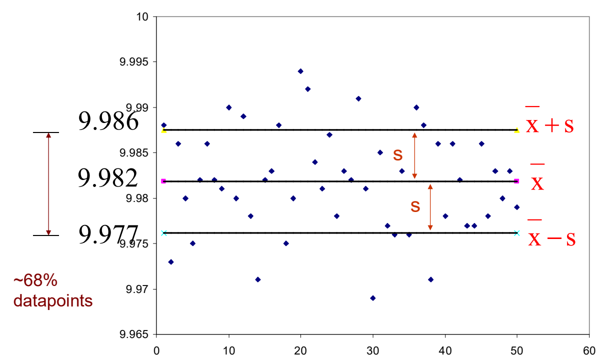


Buret Readings - Data points arranged along the x axis (smaller bins) - bar graph.

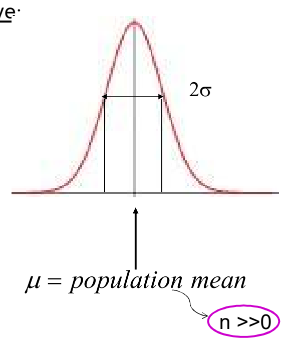








Gaussian Curve



The **μ** and **σ** (std. dev.) statistics characterize the Gaussian distribution.

Symmetric function (bell shape) about a central peak value (population (true) mean = μ); is characteristic of a data set with *random errors* (and vice versa).

μ is unique to the 'test' material (e.g. concentration of an analyte).

Gaussian distribution has two points of inflections. The distance between *them* is defined as the width of the distribution =2σ.

σ = standard deviation and variance= σ^2

σ (precision of measurements) is a measure of the extent of the spread (i.e. the variability) of a data set (closeness of the replicated measurements), measures the agreement among *repeated* measurements of the same quantity. Smaller σ implies a small variability (high precision) due to unavoidable random errors.

Population average = arithmetic mean

$$= \text{population mean} = \mu = \frac{\sum_{1}^{n} x_i}{n}$$

$n \rightarrow \infty$

Replications (GLP) increase the credibility of a measurement.

Precision, σ : measures the closeness of the replicate values, i. e. reproducibility, of a population

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

σ is a measure of precision as n →infinity

(σ ~ spread of replicates for large n; smaller spread means higher/superior precision).

σ is unique to the 'protocol of analysis',
for well established methods, σ is known.

If the area under the Gaussian curve is 100% *then* the area between inflection points is 68.3%

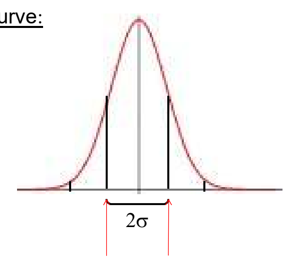
i.e. the probability of getting a value, for a single measurement between (population mean ± σ) is 68.3%.

Uncertainty should be interpreted as a statement of probability.

result = population mean ± σ

68.3% of measurements lie between (μ-σ) and (μ+σ)

Gaussian Curve:



The area within the 2σ limit is 68.3% of the total area.

mean $\pm \sigma$	68.3%;
mean $\pm 2\sigma$	95.4%
mean $\pm 3\sigma$	99.7%

Some observations - Gaussian Curves:

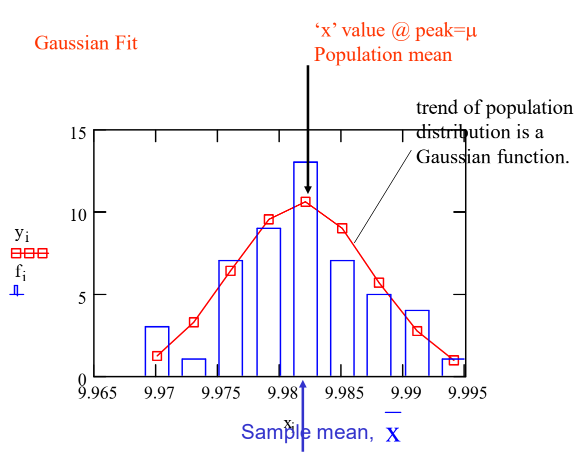
Each set of data has a unique population mean and a standard deviation.

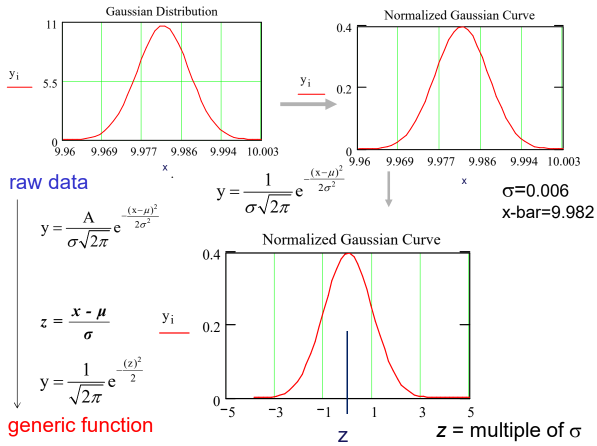
All data distributions with random errors conform to the Gaussian shape.

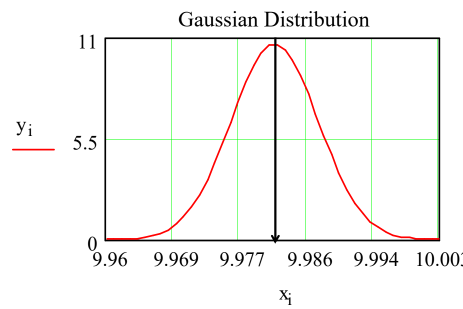
a. positive and negative errors/uncertainties are equally likely.
b. smaller errors/uncertainties are more probable than larger errors.

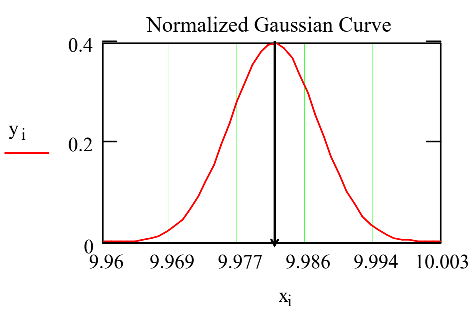
Area of the frequency distribution curve equals the sum of all the observations.

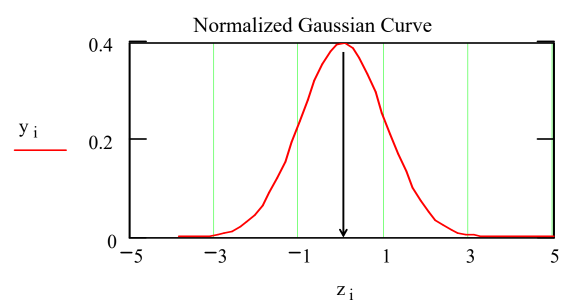
All distributions can expressed in one generic function - normalized Gaussian function; probability function.









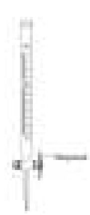


generic function

All determinations/measurements are associated with errors/ uncertainties.

Uncertainties (errors) arise from all measuring instruments.

Table 2-2 Tolerances of Class A burets



Buret volume (mL)	Smallest graduation (mL)	Tolerance (mL)
5	0.01	±0.01
10	0.05 *or* 0.02	±0.02
25	0.1	±0.03
50	0.1	±0.05
100	0.2	±0.10

Table 2-3 Tolerances of Class A volumetric flasks

Flask capacity (mL)	Tolerance (mL)
1	±0.02
2	±0.02
5	±0.02
10	±0.02
25	±0.03
50	±0.05
100	±0.08
200	±0.10
250	±0.12
500	±0.20
1 000	±0.30
2 000	±0.50

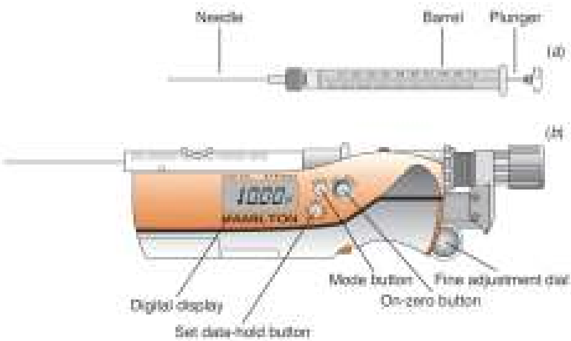
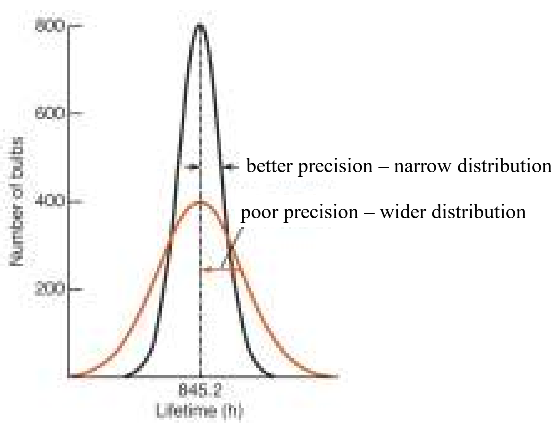


Table 2-5 Manufacturer's tolerances for micropipets

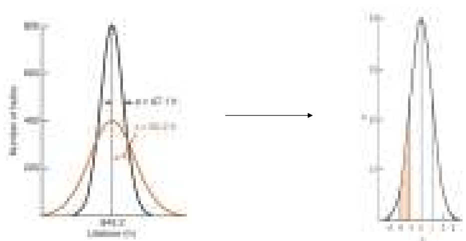
Pipet volume (μL)	At 10% of pipet volume: Accuracy (%)	At 10% of pipet volume: Precision (%)	At 100% of pipet volume: Accuracy (%)	At 100% of pipet volume: Precision (%)
Adjustable pipets				
0.2–2	±8	±4	±1.2	±0.6
1–10	±2.5	±1.2	±0.8	±0.4
2.5–25	±4.5	±1.5	±0.8	±0.2
10–100	±1.8	±0.7	±0.6	±0.15
30–300	±1.2	±0.4	±0.4	±0.15
100–1 000	±1.6	±0.5	±0.3	±0.12
Fixed pipets				
10			±0.8	±0.4
25			±0.8	±0.3
100			±0.5	±0.2
500			±0.4	±0.18
1 000			±0.3	±0.12

SOURCE: Data from Hamilton Co., Reno, NV.



SKIP

$$y = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

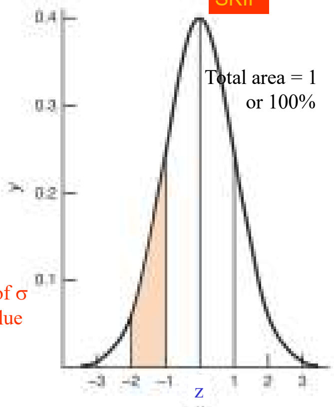


All Gaussians can be converted to generic function;
Normalized standard Gaussian Distribution.

Normalized standard Gaussian Distribution.

SKIP

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$$

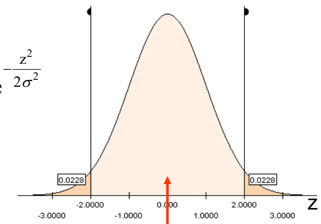


x axis (labeled z) is in units of σ measures how many σ's a value is deviated from mean.

The probability of measuring a value in a certain range of values equal to the area of that range in the Gaussian curve.

SKIP

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$$



$$(0.5 - 0.0228) \times 2$$
$$= 0.4772 \times 2 = 0.9544$$
$$or \;\; 95.44\%$$

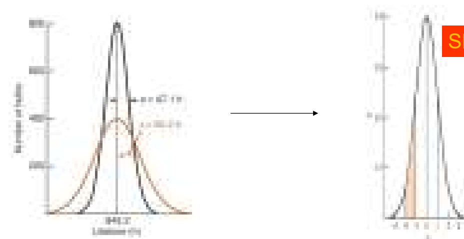
SKIP

Table 4-1 Ordinate and area for the normal (Gaussian) error curve

$\lvert z \rvert$	y	Area	$\lvert z \rvert$	y	Area	$\lvert z \rvert$	y	Area
0.0	0.398 9	0.000 0	1.4	0.149 7	0.419 2	2.8	0.007 9	0.497 4
0.1	0.397 0	0.039 8	1.5	0.129 5	0.433 2	2.9	0.006 0	0.498 1
0.2	0.391 0	0.079 3	1.6	0.110 9	0.445 2	3.0	0.004 4	0.498 650
0.3	0.381 4	0.117 9	1.7	0.094 1	0.455 4	3.1	0.003 3	0.499 032
0.4	0.368 3	0.155 4	1.8	0.079 0	0.464 1	3.2	0.002 4	0.499 313
0.5	0.352 1	0.191 5	1.9	0.065 6	0.471 3	3.3	0.001 7	0.499 517
0.6	0.333 2	0.225 8	2.0	0.054 0	0.477 3	3.4	0.001 2	0.499 663
0.7	0.312 3	0.258 0	2.1	0.044 0	0.482 1	3.5	0.000 9	0.499 767
0.8	0.289 7	0.288 1	2.2	0.035 5	0.486 1	3.6	0.000 6	0.499 841
0.9	0.266 1	0.315 9	2.3	0.028 3	0.489 3	3.7	0.000 4	0.499 904
1.0	0.242 0	0.341 3	2.4	0.022 4	0.491 8	3.8	0.000 3	0.499 928
1.1	0.217 9	0.364 3	2.5	0.017 5	0.493 8	3.9	0.000 2	0.499 952
1.2	0.194 2	0.384 9	2.6	0.013 6	0.495 3	4.0	0.000 1	0.499 968
1.3	0.171 4	0.403 2	2.7	0.010 4	0.496 5			

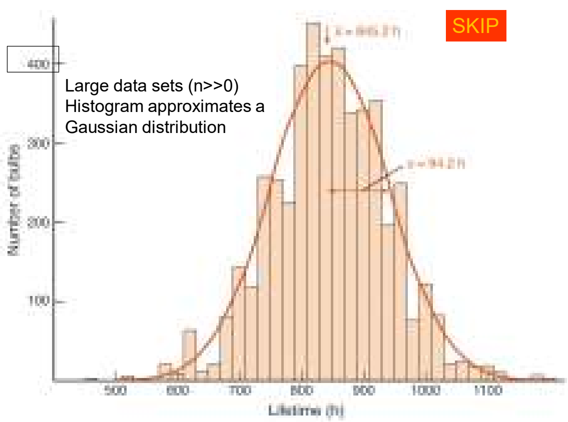
[illegible]

SKIP



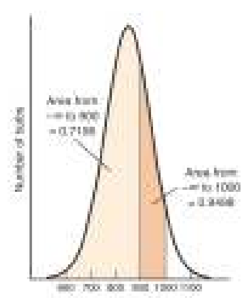
Use the NORMDIST function of Microsoft Excel to calculate the area under a Normalized Gaussian curve.

SKIP



Fraction of bulbs with lifetime between 900h to 1000h?

SKIP



Mean	Std dev
845.2	94.2

Area from -∞ to 900 =	0.719629
Area from -∞ to 1000 =	0.949841
Area from 900 to 1000 =	0.230211

"NORMDIST(X,Mean,StdDev,TRUE)"

Area under the curve from –∞ to a specified point of value X.

SKIP

Mean	Std dev
845.2	94.2

Area from -∞ to 900 =	0.719629
Area from -∞ to 1000 =	0.949841
Area from 900 to 1000 =	0.230211

"NORMDIST(X,Mean,StdDev,TRUE)"

=NORMDIST(900,A2,C2,TRUE)

Cells containing mean and stdev values

Classification of experimental errors:

a. ***Random/indeterminate errors*** - makes each x_i different from another.

* never eliminated but can be minimized; arises due to random variations beyond experimenter's control.

frequency of x_i vs x_i $\Rightarrow$ distribution function ~ Gaussian function ($n \to$ v.large)

b. **Systematic/determinate error** – shifts μ & x_i's by a constant value.

Can be detected and corrected by calibration and standardization.

Avoid ***zero error*** (non-zero reading on an instrument when it is measuring something that should read zero), ***calibration error*** and ***insertion error***. Check protocol and calibration of the analysis using *standard reference materials* (SRM/CRM).

SRM (NIST definition) : Material or substance, one or more of whose property values are sufficiently homogeneous, stable, and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials.

c. **Gross errors** – mistakes; produces outliers, usually rejected.

Most experiments are replicated for a finite number, i.e. a low number of data points, n (a data set = sample), therefore both μ and σ are incalculable.

Best estimate of μ and σ are;

sample mean $\bar{x}$

sample standard deviation s

The number of replications are very limited; $n << \infty$.

The limited number of data points (in the sample) are assumed to be drawn from the members of a data set (population) of an 'infinite' number of data points.

Sample Mean:

best estimate of 'unknown' = arithmetic mean of data set

$$= \text{sample mean} = \bar{x} = \frac{\sum_{1}^{n} x_i}{n}$$

n = # replications = # data points in a data set.

$$\bar{x} \to \mu \quad \text{as } n \to \infty$$

Sample Standard Deviation:

For a limited number of replications, estimate of uncertainty sample std. deviation, s

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

Degrees of freedom ($n-1$)

$$s \to \sigma \quad \text{as } n \to \infty$$

(s ~ spread of replicates for limited n; smaller spread means higher/superior precision).

Degrees of freedom, number of observations *less* the number of quantities measured.

Quick and Approximate Calculation of *s*:

From a limited number of measurements (n≤10), the s can be estimated from the following (whichever is higher).

$$s = \frac{2}{3}(x_{max} - \bar{x}) \text{ and } s = \frac{2}{3}(\bar{x} - x_{min})$$

Any outlier however would have an inordinate influence on calculated s from above.

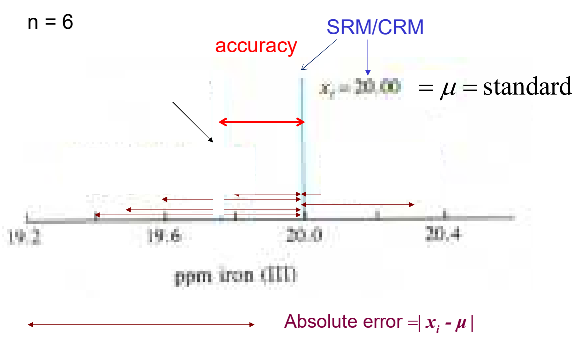
Practical measure of precision of a data set is *s* not σ.

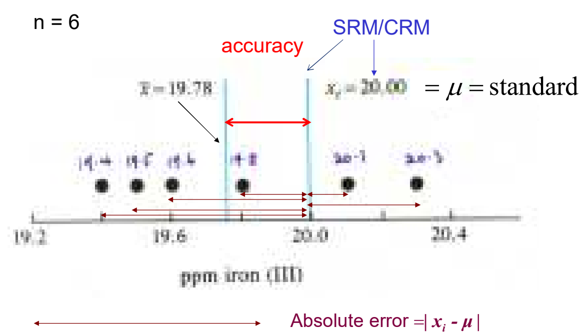
Practical estimate of a value of a measurement is $\bar{x}$ not μ.

$s \rightarrow \sigma$ and $\bar{x} \rightarrow \mu$; as n→ ∞

The higher the number of replications the closer will be the sample mean to the population mean. *The estimated mean value depends on the number of data points, n, used for it's calculation.* Thus, the 'estimated mean' is associated with an uncertainty.

The *uncertainty of the mean value, decreases*, i.e. the confidence in the mean value increases, *as n increases*.





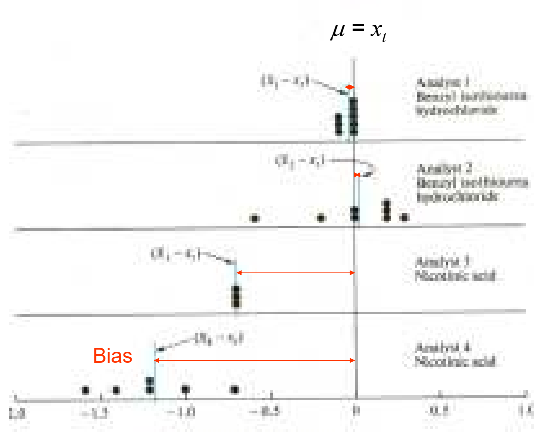
Accuracy: refers to the closeness of the sample mean, $\bar{x}$ to the true value, μ; |$\bar{x}$- μ|;

Bias: defined as |$\bar{x}$ - μ|; smaller bias imply a higher accuracy.

STRIVE FOR ACCURATE & HIGH PRECISION RESULTS.

s → σ & $\bar{x}$ → μ ;as n → infinity (≥20, practically)

Replications (GLP) increase the accuracy of a measurement (assuming only random error present) and therefore credibility of the measurement.



REPLICATE DATA

1	9.988	18	9.975	35	9.976
2	9.973	19	9.980	36	9.990
3	9.986	20	9.994 #	37	9.988
4	9.980	21	9.992	38	9.971
5	9.975	22	9.984	39	9.986
6	9.982	23	9.981	40	9.978
7	9.986	24	9.987	41	9.986
8	9.982	25	9.978	42	9.982
9	9.981	26	9.983	43	9.977
10	9.990	27	9.982	44	9.977
11	9.980	28	9.991	45	9.986
12	9.989	29	9.981	46	9.978
13	9.978	30	9.969 *	47	9.983
14	9.971	31	9.985	48	9.980
15	9.982	32	9.977	49	9.983
16	9.983	33	9.976	50	9.979
17	9.988	34	9.983		
Mean	9.982				
Median	9.982				
Spread	0.025				
s	0.0056				

The uncertainty of the mean value, *decreases*, i.e. the confidence in the mean value increases, as n increases.

The means calculated from a limited number of replications would have its own distribution of values. The variation of such a distribution depends on the number of replications n per data set. The *standard deviation such a set of 'means'* is referred to as ***standard error* or *standard deviatiation of the mean (SDOM)***. It is related to standard deviation from the data points;

$$std.\,error,\ s_n = \frac{s}{\sqrt{n}}$$

$$mean = \bar{x} \pm \frac{s}{\sqrt{n}}$$

Larger the number of replications, n, of a data set the lower is the standard deviation of the mean calculated from such a set replicates (data points).

Measures of precision; RSD,CV and s_m

$$RSD = \frac{s}{\bar{x}} \qquad CV = \%RSD = \frac{s}{\bar{x}} \times 100$$

CV = coefficient of variation

$$std.\,error,\ s_n = \frac{s}{\sqrt{n}}$$

** For n>20, $s \cong \sigma$

The number of sig. figs of the uncertainty is *one*, for the limited number of observations we usually make, e.g. ±0.1, ± 0.002, ± 0.00006, etc.

Note: *Do not exceed one sig. fig. for uncertainty under normal circumstances.*

Pooled data yields better estimation of s:

Data accumulated over a long period of time or from different data sets, *using the same method* can be pooled to obtain a better estimate of std. deviation, σ. Homogeneous variances (assumption: *sources of error are the same* for all data points).

$$s_{pooled} = \sqrt{\frac{\sum_i^{n_1}(x_i - \bar{x}_1)^2 + \sum_j^{n_2}(x_j - \bar{x}_2)^2 + ..}{(n_1 + n_2 + ...) - k}}$$

$$s_{pooled} \to \sigma$$

n_k = number of data points from set k
k = n_{sets} = number of data sets = number of 'quantities)

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

k = number of data sets

Propagation of Errors

When calculations are performed, using measurements that are associated with errors (uncertainties) themselves, the errors get *propagated*.

The final result must carry an error (uncertainty) as well.

The error associated with the final result must also be stated and must be expressed *to the correct significant figure*.

Example:

$$c_{HCl} = \frac{w_{Na_2CO_3}}{(MW)_{Na_2CO_3}} \times \frac{2M_{HCl}}{V_{HCl}} + c_0$$

Most terms in an expression has an error associated with it!

Propagation of Errors

Additions and subtractions

$$y = (k_1 a_1 + k_2 a_2 - k_3 a_3)$$

k_i = constants (assumed 'perfect')

$$y \pm s = k_1(a_1 \pm s_1) + k_2(a_2 \pm s_2) - k_3(a_3 \pm s_3)$$

Multiplications and divisions

$$y = k\frac{a_1 \times a_2}{a_3}$$

$$y \pm s = k\frac{(a_1 \pm s_1) \times (a_2 \pm s_2)}{(a_3 \pm s_3)}$$

Std. deviation of computed results, propagation of errors:

a. addition & subtraction

$$y = (k_1 a_1 + k_2 a_2 - k_3 a_3) \ \ or$$
$$y = (k_1 a_1 + k_2 a_2 + k_3 a_3) \ \ or....$$

Calculate y in the usual manner.
Do not round or adjust for sig. figs.
until s is calculated.

$$s = \sqrt{(k_1 s_1)^2 + (k_2 s_2)^2 + (k_3 s_3)^2}$$

(same approach for σ, use σ in place of s)

b. multiplication & division

$$y = k\frac{a_1 \times a_2}{a_3} \ \ or \ \ y = k\frac{a_1}{a_2 \times a_3} \ \ \text{or ...}$$

first calculate y and the relative errors,

Calculate y in the usual manner. Do not round or adjust for sig. figs. until s is calculated.

$$\frac{s}{y} = \sqrt{(\frac{s_1}{a_1})^2 + (\frac{s_2}{a_2})^2 + (\frac{s_3}{a_3})^2}$$

$$s = \sqrt{(\frac{s_1}{a_1})^2 + (\frac{s_2}{a_2})^2 + (\frac{s_3}{a_3})^2} \times y$$

(same approach for σ), the significant figure of the final result must be consistent with uncertainty.

Significant figures of a calculated result is determined by the ***first** non zero digit* of the uncertainty (σ or s) associated with it. Calculate the uncertainty first before deciding on the significant figures of the final computed value. Both the computed result and the uncertainty must be consistent in terms of the number of figures in them.

4 $\quad \dfrac{0.002364}{0.02500} = 0.0946 \pm 0.0002 \quad$ 3

4 $\quad \dfrac{0.002664 \pm 0.000003}{0.02500 \pm 0.00005} = 0.1066 \pm 0.0002 \quad$ 4

3 $\quad \dfrac{0.821}{0.803} = 1.022 \pm 0.004 \quad$ 4

Some other uncertainty calculations:

The error in y for an error of s in x (a and k are constants)

$y = kx^a$	$ya\,(s/x)$
$y = \log x$	$0.43429(s/x)$
$y = \ln x$	s/x
$y = 10^x$	$2.3026\,ys$
$y = e^x$	ys

In general, use differential calculus to calculate the error.

$$e_i = s_i$$

TABLE 3-1 Summary of rules for propagation of uncertainty

Function	Uncertainty	Function[a]	Uncertainty[b]
$y = x_1 + x_2$	$e_y = \sqrt{e_{x_1}^2 + e_{x_2}^2}$	$y = x^a$	$\%e_y = a\%e_x$
$y = x_1 - x_2$	$e_y = \sqrt{e_{x_1}^2 + e_{x_2}^2}$	$y = \log x$	$e_y = \frac{1}{\ln 10}\frac{e_x}{x} = 0.434\,29\frac{e_x}{x}$
$y = x_1 \cdot x_2$	$\%e_y = \sqrt{\%e_{x_1}^2 + \%e_{x_2}^2}$	$y = \ln x$	$e_y = \frac{e_x}{x}$
$y = \frac{x_1}{x_2}$	$\%e_y = \sqrt{\%e_{x_1}^2 + \%e_{x_2}^2}$	$y = 10^x$	$\frac{e_y}{y} = (\ln 10)e_x = 2.302\,6\,e_x$
		$y = e^x$	$\frac{e_y}{y} = e_x$

a. [illegible]
b. [illegible]

Harris, Quantitative Chemical Analysis, 8e

Note: The higher the number of replications the closer will be the *sample mean* to the *population mean*.

Also a higher number of replications makes the *uncertainty of the mean* (also known as standard error or SDOM) decrease as $1/\sqrt{n}$; and therefore confidence increases)

Municipal water lab - Concentration calculations:

$$M^{+n} + EDTA^{4-} \rightleftharpoons MEDTA^{-(4-n)} \qquad M = Ca \text{ /or } Mg$$

Stoichiometric ratio:

$$\frac{\text{moles } M^{+n}}{\text{MolesEDTA}} = 1 \Rightarrow C_M V_{M-\text{solution}} = C_{EDTA} V_{EDTA-\text{solution}}$$

Define $C_1 = C_{Ca} + C_{Mg}$ and $C_2 = C_{Ca}$

Setup the equations, Solve for C_i's and their standard deviations s_1 and s_2.

Municipal water lab - uncertainty calculations:

From 1st set of replicates, the total concentration of ions C_1 and the uncertainty s_1 can be calculated:

Part 1 $\Rightarrow C_1 \pm s_1 = (C_{Ca} + C_{Mg}) \pm s_1$

From 2nd set of replicates, concentration of Ca ions C_2 (= C_{Ca}) and the uncertainty s_2 (= s_{Ca}) would be:

Part 2 $\Rightarrow C_2 \pm s_2 = C_{Ca} \pm s_{Ca}$

$C_{Mg} = C_1 - C_2$

and $s_{Mg}^2 = s_1^2 + s_2^2 \Rightarrow s_{Mg} = [s_1^2 + s_2^2]^{1/2}$ (Uncertainty of C_{Mg})

decimal places must be set accordingly