Types of "Error”

Illegitimate Errors

- These are outright mistakes or blunders in computation, measurement or recording of a result.
- Dealt with by performing previously erroneous operation correctly.
- This is a source of errors one typically does not discuss in reporting scientific results:

Your colleagues (and professors!) expect that if you are reporting a result to them you are competent enough to do the corresponding computations and measurements correctly and that you have checked your work before reporting them to others!

42

Systematic Errors

- Systematic errors can arise from:
    1. Faulty calibration of equipment.
    2. Faulty use of equipment.
    3. Bias of the observer (e.g., how one observer's eyes perceives a star against the reticle in an eyepiece compared to the perception by another observer, or e.g., how one reads the needle on a dial compared to another person).
    4. Poor assumptions of how to do experiment properly or how to analyze/interpret the results.

43

Systematic Errors

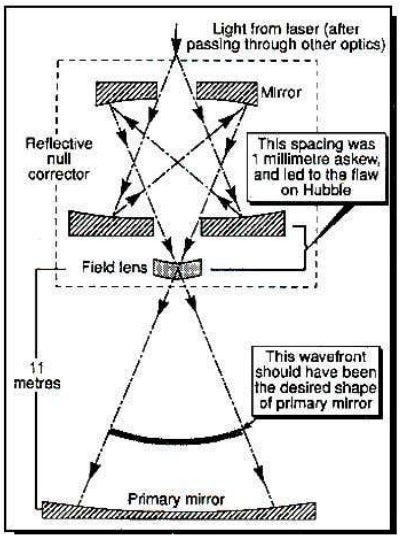
- ***Systematic errors are usually not easy to detect!***
  - Have to sense there is a problem (e.g., from theory or by seeing differences between different experimental apparati).
- Have to assess experimental conditions or techniques for possible errant influences.

44

- **A key to systematic errors is that they are *reproducible*:**
  When the experiment is repeated in the same way, you always get an answer that is offset systematically from truth in the same way.
- We refer to this systematic offset as a ***bias*** in the experiment.
- Examples of systematic errors:
  1. Hubble Space Telescope spherical aberration: Test rig at wrong distance, mirror ground incorrectly.
  2. Using a metal measuring stick ruled at 25° C at a temperature of 0°.
  3. The history of measuring Hubble's constant.
  4. The history of the Millikan Oil Drop Experiment

45

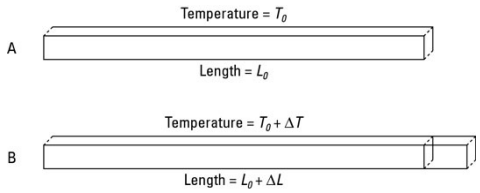
Hubble Space Telescope spherical aberration.



Most perfect/imperfect mirror ever made!

46

Measuring Stick thermal Expansion



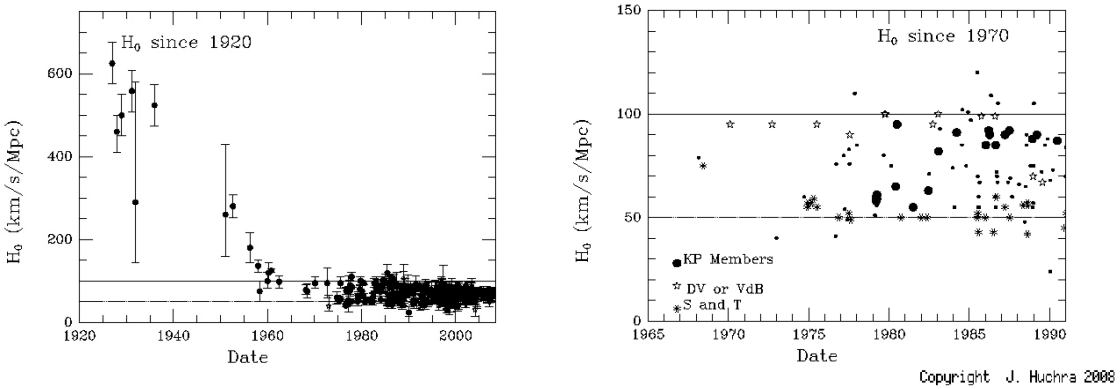
Metal expands with increasing temperature

273 K

350 K

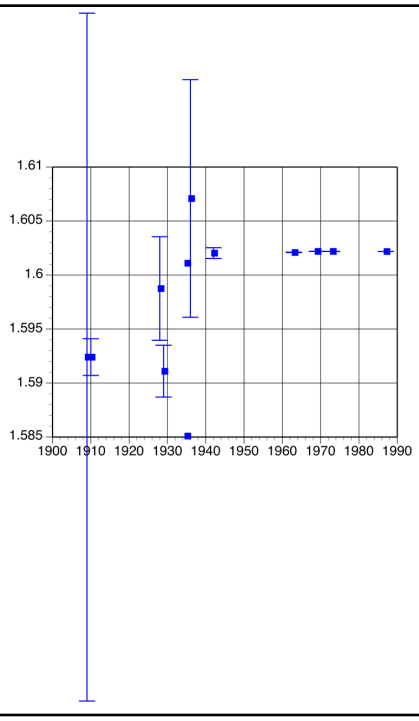
47

Hubble Law



48

Millikan Oil Drop – Charge of the Election



Millikan (notebooks)
Millikan (published)
Erik Backlin, Nature 1929
[Birge, 1929]
Backlin and Flemberg, Nature 1936
Backlin and Flemberg, cited in HR Robinson RPP 1937
Gunnar Kellström PR 1936
[Birge, 1942]
[Dummond and Cohen, 1963]
[Taylor et al, 1969]
[Cohen and Taylor, 1973]
[Cohen and Taylor, 1987]

49

Ideally systematic errors should be absent, but they should always be checked for.

- A key way to do this is to test your apparatus on something where you know the correct answer.

  For example, if you suspect your ohmmeter is measuring resistances incorrectly, one could measure some resistors with known values.

  Or to check your wristwatch, you could compare its reported time against that from the atomic clock at the U.S. Naval Observatory.

50

***Calibration* of your experimental apparati is a key part of experimentation that is intimately related to whether it will bias your results.**

For example, if you find your wristwatch is running 10 minutes fast, then to use your watch to obtain accurate time you could either:

- calibrate (i.e. correct) your watch against an accurate timepiece (e.g., the atomic clock), or
- you could remember that your watch is *biased* to be 10 minutes fast and always account for (i.e., subtract) that bias every time you get the time from your watch.

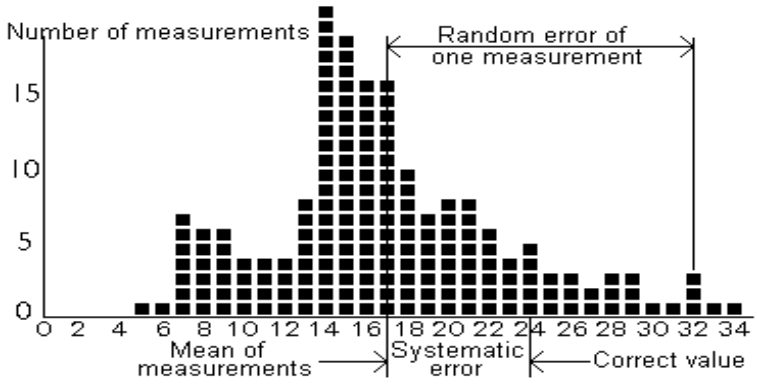
51

Random Errors (Statistical Errors)

- These are fluctuations in observations that lead to **different** results each time same experiment performed.
- These are **irreproducible** - *You don't get exactly the same answer each time the experiment is performed in the same way.*
- Random errors are an indefiniteness coming from either:
  1. An inability of your measuring device to give infinitely accurate answers.
  2. Nature: Fluctuations that occur in observations of a small sample drawn from a large population.
- The effect of random errors is to produce a spread of answers around some mean value (which is hopefully near the correct value).

52

For example, note the spread out distribution of results of repeated experiments shown below; **random errors create the spread**, and result in the displacement of, e.g., the highlighted measurement from the mean value (From http://www.uiah.fi/projects/metodi/evirheet.gif):



Note that in the case shown, the mean value of the distribution is actually offset from the correct value that should have been gotten, so **that there is *also* apparently a systematic error** affecting the results of the experiment.

53

As we will see below, we can get some idea of the size of random errors affecting our experiment by simply analyzing the spread in the distribution of answers we get when we repeat the experiment.

Examples of sources of random errors:

- Making ***mm*** measurements with a meterstick ruled only in ***cm*** increments.
- Counting # of photons received from a weak source in a short time span (e.g., 1 sec).

While systematic errors are often hard to overcome, we often have an easier time with random errors, which we can do something about:

1. Improve precision of equipment.
2. Count more events (*e.g., expand the time interval to count photons arriving to 100 sec*).
3. Repeat experiment and average results (*the more the better!!*).

54

Examples: Systematic vs. Random Error

- Thought experiment: Imagine a wristwatch with only hour and minute hands but no second hand and running 2 minutes late.
  - In this case, what is the character of the expected systematic and random errors?

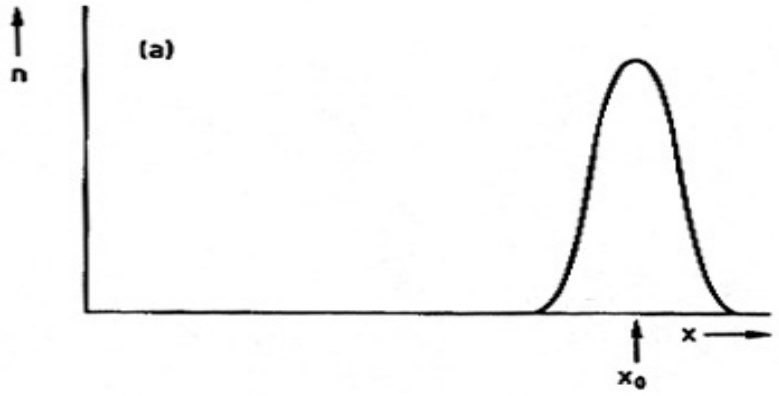
Accuracy vs. Precision - THEY ARE NOT THE SAME!!

- **Accuracy** - How close experiment comes to true value. Has primarily to do with ***Systematic Errors***.
- **Precision** - How exactly is result determined without reference to what result means or how it compares to "truth". Has primarily to do with ***Random Errors***.

55

Example distributions of measurements in an experiment (where x_0 denotes the "truth value sought"):

(a) This distribution shows *random errors* (i.e. the *mean value* of many trials of the experiment gives a result that is **accurate** but the individual measures themselves are **not very precise**).

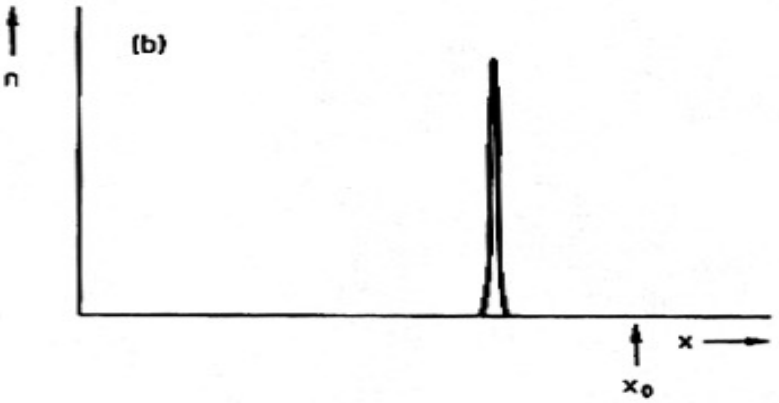


56

Example distributions of measurements in an experiment (where x_0 denotes the "truth value sought"):

(b) This distribution shows primarily *systematic errors* (i.e. the experiment is able to give **very precise** results but they are **not very accurate**).

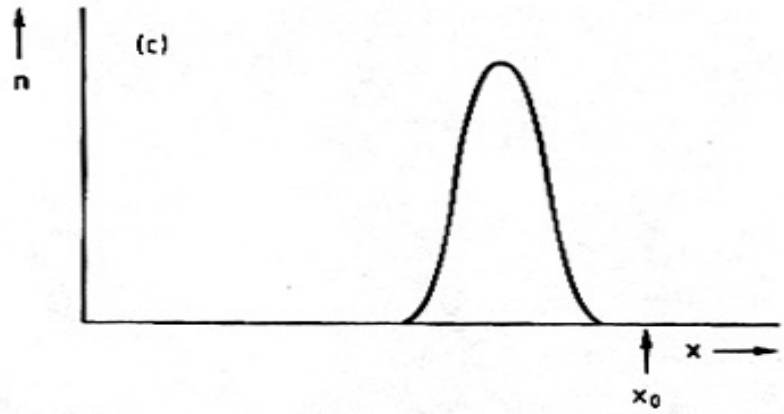
Example: The Hubble Space Telescope mirror was ground very precisely but to an inaccurate shape. This resembles distribution (b).



57

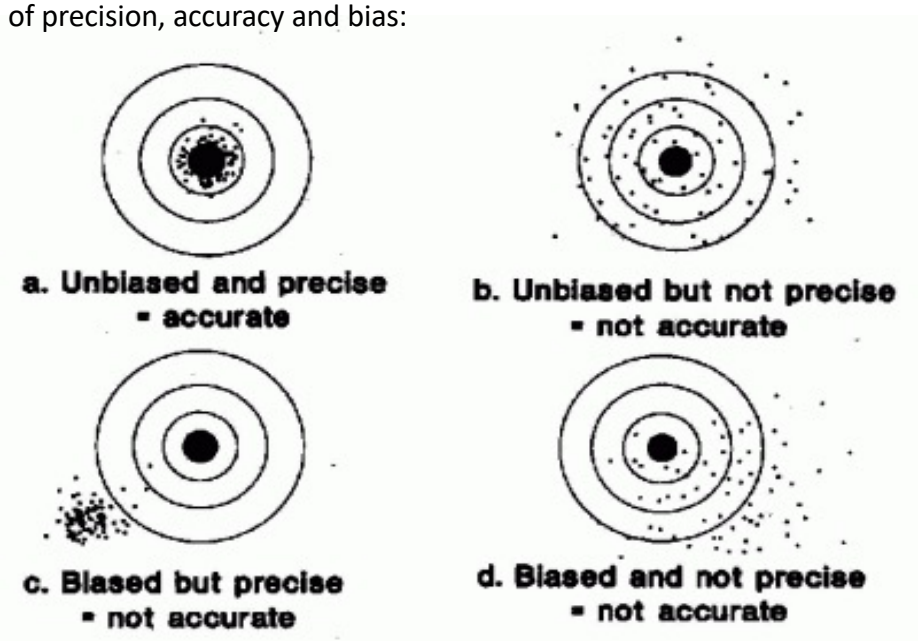
Example distributions of measurements in an experiment (where x_0 denotes the "truth value sought"):

(c) This distribution shows *both* random & systematic errors (the experiment is giving results that are **neither precise nor accurate**).



58

Here is another example -- shots fired at a target – that vividly demonstrates the concepts of precision, accuracy and bias:



59

We often want to describe the precision of an experiment or a piece of equipment. This can be done in several ways:

Absolute Precision

Magnitude of uncertainty given in same units as result.
e.g. **30 km/s +/- 3 km/s**

Relative Precision

This is a ***fractional*** uncertainty, often given as a percentage.
e.g. (3 km/s) / (30 km/s) **= 10 % error**

From Lyons: ***A good experimentalist is one who minimizes and realistically estimates the random errors of his apparatus, while reducing the effect of systematic errors [as much as possible].***

60

Estimating the Truth - Simple Cases

- In doing a measurement or experiment we hope to asymptote to truth value. (we'll call that $\boldsymbol{\mu}$) by beating down random errors (and hoping systematics are not present).
- Normally we repeat experiments and get a distribution of results $\mathbf{x_i}$ meant to get at the value of $\boldsymbol{\mu}$ (like the distributions we have shown above).
- In error analysis, we use the $\mathbf{x_i}$ to estimate truth value and to assess the reliability of the result.
  - An estimator of the truth value is also called the "expectation" of the true value.
  - There are often a variety of estimators possible to give an expectation of the truth.

61

Estimating the Truth - Simple Cases

- A simple estimator you know is the **mean** value:

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

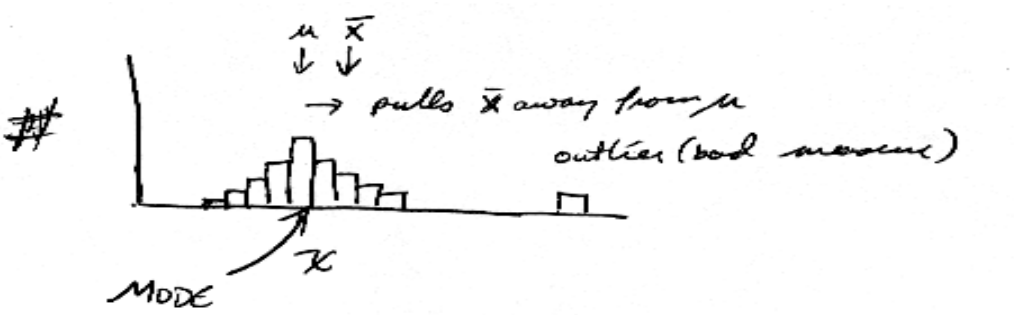
- The hope is that we can use averaging to approach the truth:

$$\mu = \lim_{N \to \infty} \left(\frac{1}{N} \Sigma x_i \right)$$

62

But the mean value of the many trials of an experiment may not be a good estimator if the distribution $\mathbf{x}_i$ is not symmetric.

For example, look at the mean value of this distribution, which is affected by the presence of one experiment trial that gave a very high ("outlier") result:



(NOTE: T
identifyi

63

A more reliable assessment of the truth might be the "**mode**" (the most common $\mathbf{x}_i$ value) --- indicated below by $\mathbf{x}_{max}$.

- BUT - The precision of this estimator depends on the "bin size" of how you "histogram" the data, because this sets the "resolution" spacing of $\mathbf{x}_i$.

  That is, the precision in the modal value can be no finer than the selected bin size.

- Note that the size of bins does have some influence in determining the modal value, e.g.:
  - One might want to make the bins narrow to improve resolution, but making the bins so fine that all bins have either 1 or 0 counts in them is of no use.
  - Using the mode as an estimator requires both a large number of measures and an appropriate "bin-size" to overcome random fluctuations that could make any particular bin the "highest" by chance.

64

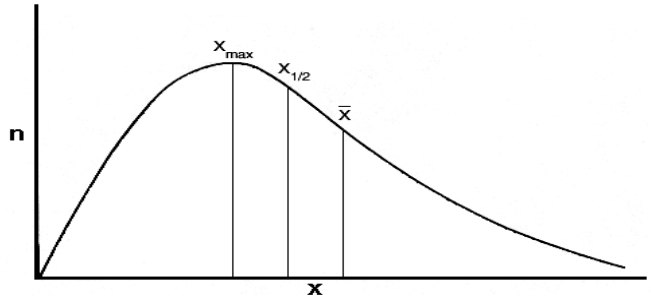
The "**median**" is an estimator that is more robust to outliers and is useful when you don't have a very large number of x_i.

The median is defined as a value ($\mathbf{x_{1/2}}$) for which half of observations/measures lie above and half below:

$$P(x_i \leq x_{1/2}) = P(x_i \geq x_{1/2}) = 50\%$$

65

Example of mean, median and mode for a particular distribution **x**:



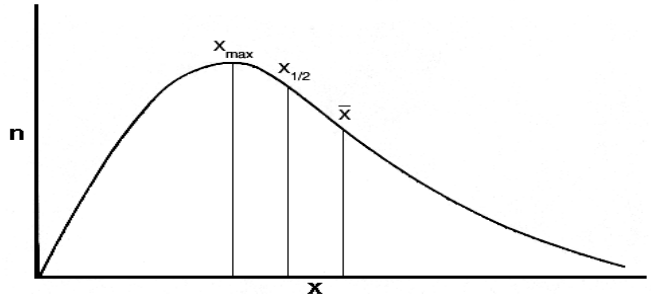
x_{max}, $x_{1/2}$, and $\bar{x}$ can each be used as an estimator of μ, but each is best used under different circumstances:

$\bar{x}$: Useful when you have small N (few measures).

But these rules are not firm.

66

Example of mean, median and mode for a particular distribution **x**:



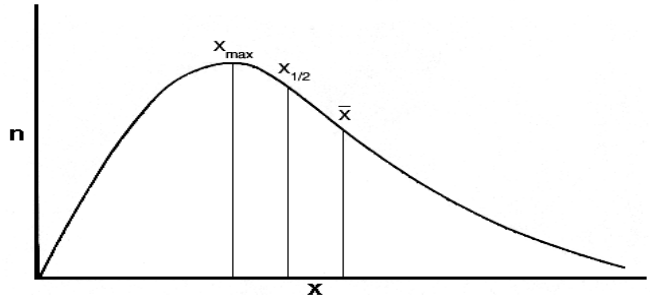
x_{max}, $x_{1/2}$, and $\bar{x}$ can each be used as an estimator of μ, but each is best used under different circumstances:

$x_{max} : N \to \infty$, useful when you have very many measures.

But these rules are not firm.

67

Example of mean, median and mode for a particular distribution **x**:



x_{max}, $x_{1/2}$, and $\bar{x}$ can each be used as an estimator of μ, but each is best used under different circumstances:

$x_{1/2}$: The median is useful when you have a moderate number of measures (say, N >~ 10). Also when it is apparent that there are "outliers" in the distribution.

But these rules are not firm.

68

Characterizing the Uncertainty in the Truth
Note, the word "error" is sometimes used for the word "uncertainty".

- Some scientists abhor this usage, but MANY others (like me!) were raised routinely using the word "error" to mean "uncertainty".
- As John Taylor says in his *An Introduction to Error Analysis*, the words *error* and *uncertainty* can be treated as interchangeable in the current context.

Characterizing the "width" of the distribution leads to an estimation of the "uncertainty" in **μ**.

There are various ways of doing this.

Assume that our preferred estimator of **μ** is the mean value of our multiple experiments:

69

Deviation (for one measure):

$$x_i - \bar{x}$$

Mean Deviation:

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

Variance (Easier to calculate):

$$\sigma^2 = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)$$

Standard Deviation = "Root Mean Square"

$$\sigma = \sqrt{\sigma^2} = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{1/2}$$

70

NOTE - when ***N*** is not large, should use **Sample Standard Deviation**:

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

This is because the *true* standard deviation *should* have been measured against the truth value μ, whereas we do not know μ and are measuring a dispersion against an *estimator* of μ, namely the mean value of *x*. In this case, the *number of degrees of freedom (NDOF)* in the situation has been lowered by 1, leaving only *N* - 1 as our NDOF.
(Think about it this way, if *N* = 1 you would set μ = x_i and you would find an unrealistic σ = 0....
... at least with the *s* definition above, the standard deviation is indeterminate.)

Note that as *N* approaches infinity, the sample standard deviation approaches the standard deviation, since 1 / (*N* - 1) approaches 1 / *N*.

71

It is most customary to quote the random errors of an experiment as gauged by the standard deviation or the sample standard deviation.

- When you see or give values given in a form such as *X +/- Y* , generally the *Y* value (the error or uncertainty) is the standard deviation or sample standard deviation. Unless stated explicitly otherwise, it is typical to assume this is what is meant.
- Under these circumstances, often the *Y* value is referred to as the **"1-sigma" (1σ) errors.**
- In plots of the values *X*, it is customary to plot *error bars* that are the size of "1-σ"; that is, the error bar extends 1-σ from *X* up to *X+Y*, and 1-σ down to *X-Y*.
- On occasion you may find people giving other-sized error bars (e.g., "2-σ"), but unless explicitly stated, generally 1-σ are assumed.

72

Combining Results from Different Experiments

When several experiments measure the same physical quantity and give a set of answers, a_i , and with different uncertainties, σ_i , you must combine differently.

- When you have measurements as *X +/- Y* , with different *Y* values (the error or uncertainty), you need to weight by the uncertantities.
- The best combined estimate is:

$$a = \sum \left(\frac{a_i}{\sigma_i^2} \right) \Big/ \sum \left(\frac{1}{\sigma_i^2} \right)$$

- And uncertainty of:

$$1/\sigma^2 = \sum \left(\frac{1}{\sigma_i^2} \right)$$

73

A Final Comment on Characterizing the Spread in a Distribution

In the above discussions we have assumed that the spreads in the distributions of results from multiple trials of the same experiment arose completely from random errors.

- In this case, the spreads we have calculated are giving us an estimation of the uncertainty in our estimator of the truth value.
  Then we can say that the values of σ or *s* are the uncertainty in the mean, for example.

74

But sometimes we are measuring something that has an *intrinsic* spread in values, e.g., the height of students taking PHYS 10273.

- In this case, because not everyone in the class has the same height, we will measure a spread of values if we measure the height of everyone in the class, even if those measurements are infinitely precise.

- The *measured* spread will have contributions not only from the *intrinsic* spread of people in the class (which is relatively large), but also due to errors in our ability to measure the height of everyone in the class accurately (which is presumably much smaller).

75

But sometimes we are measuring something that has an *intrinsic* spread in values, e.g., the height of students taking PHYS 10273.

- It is important to remember that in the case where there is a true *intrinsic spread* in what we are measuring, then the values of σ or *s* are *not* a measure of the uncertainty in the mean, but are measures (mostly) of the intrinsic spread of heights.

  The *error in the mean* will actually be much smaller than σ or *s*, and this is a more useful descriptor of the uncertainty in the result (if the result is to determine the mean value of the experiment and not its intrinsic spread).

- One can prove that this is the case by thinking about what would happen to our value of σ or *s* if we substantially increased the number of students in the class (say by a factor of 10 or 100).

  The spread of heights (σ or *s*) will be relatively unchanged, even though we know we should be getting a better estimate of the mean height of someone of college age as *N* substantially increases.

76

CENTRAL LIMIT THEREOM AND COMBINATION OF ERRORS
REFERENCE: Lyons Chapter 1.5-1.12.

We have discussed earlier the issue of experimental errors, including the concepts of random and systematic errors and their connection to precision and accuracy.

In this lecture I want to give some important practical aspects of dealing with errors, including the concepts of *Gaussian distributions* and the important *Central Limit Theorem*, and proper ways to both *combine* and *propagate* errors.

Much of this lecture is taken directly out of the book:
"*A Practical Guide for Physical Science Students*" by Louis Lyons

77

The Gaussian distribution is central to any discussion of the treatment of errors.

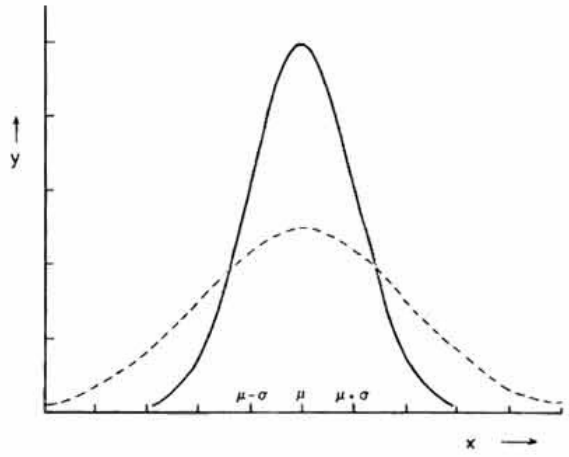
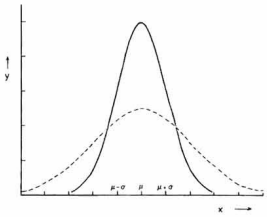


Fig. 1.6. The solid curve is the Gaussian distribution of eqn (1.10). The distribution peaks at the mean μ, and its width is characterised by the parameter σ. The dashed curve is another Gaussian distribution with the same value of μ, but with σ twice as large as the solid curve. Because the normalisation condition (1.11) ensures that the areas under the curves are the same, the height of the dashed curve is only half that of the solid curve at the maxima. The scale on the x axis refers to the solid curve.

78

The general form of the Gaussian distribution in one variable x is:



$$y = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-(x-\mu)^2/2\sigma^2\right\}$$

- The curve of y as a function of x is symmetric about the value of x = μ, at which point y has its maximum value.
- The parameter σ characterizes the width of the distribution.
- The factor $(\sqrt{2\pi}\sigma)^{-1}$ ensures that the distribution is normalized to have unit area underneath the whole curve, i.e.,

$$\int_{-\infty}^{\infty} y dx = 1$$

79

The parameter μ is the mean of the distribution, while σ has the following properties:

- The mean square deviation of the distribution from μ is σ^2, called the *variance*. (The reason that the curious factor of 2 appears within the exponent in the equation for *y* above is to make sure that σ is the RMS deviation. Otherwise the root mean square deviation from the mean would have been $\sigma/(2^{1/2})$, which is unaesthetic.)
- σ is known as the *standard deviation*.
- The height of the curve at x= μ ± σ is $e^{-1/2}$ of the maximum value. Since

$$1/\sqrt{e} = 0.607$$

σ is very *roughly* the half width at half height of the distribution.

80

- In fact, one finds that the Full-Width-Half-Maximum is equal to

FWHM = 2.354 σ

So that HWHM = 1.177 σ.

(Prove this to yourself!).

- The fractional area underneath the curve in the range (i.e., within ± σ of the mean μ) is 0.68.

This is a very important thing to keep in mind: $\mu - \sigma \leq x \leq \mu + \sigma$

When you repeat an experiment with random errors, about 2/3 of the results will be within 1σ (one standard deviation) of μ and 1/3 of your results will be beyond 1σ.

81

This is a very important thing to keep in mind: When you repeat an experiment with random errors, about 2/3 of the results will be within 1σ (one standard deviation) of μ and 1/3 of your results will be beyond 1σ.

Note that 95% of the results will be within 2σ, and 99.7% are within 3σ.

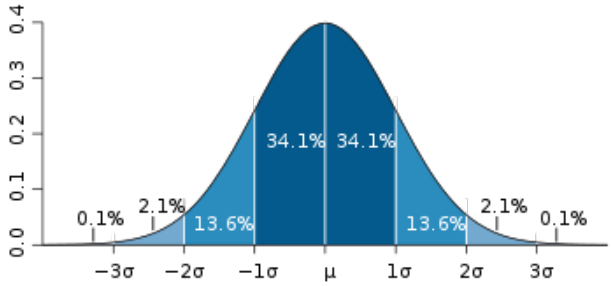


Image and caption from Wikipedia: Dark blue is less than one standard deviation from the mean. For the normal distribution, this accounts for 68.27 % of the set; while two standard deviations from the mean (medium and dark blue) account for 95.45 %; and three standard deviations (light, medium, and dark blue) account for 99.73 %.

The height of the distribution at its maximum is $(\sqrt{2\pi}\sigma)^{-1}$. As σ decreases the distribution becomes narrower, and hence, to maintain the normalization condition, also higher at the peak.

82

By a suitable change of variable to

$$x' = (x - \mu)/\sigma$$

any normal distribution can be transformed into a standardized form

$$y = \frac{1}{\sqrt{2\pi}} \exp\left(-x'^2/2\right)$$

with mean zero and unit variance

83

The Meaning of σ

- It is customary to quote σ, the standard deviation, as the accuracy of a measurement.
- Since σ is not the maximum possible error, we should not get too upset if our measurement is more than 1 σ away from the expected value. Indeed, we should expect this to happen with about 1/3 of our experimental results.
  Since, however, the fractional areas beyond ± 2σ and beyond ± 3σ are only 4.6% and 0.3% respectively, we should expect such deviations to occur much less frequently.

84

Central Limit Theorem

A feature that helps to make the Gaussian distribution of such widespread relevance is the Central Limit Theorem. One statement of this is as follows.

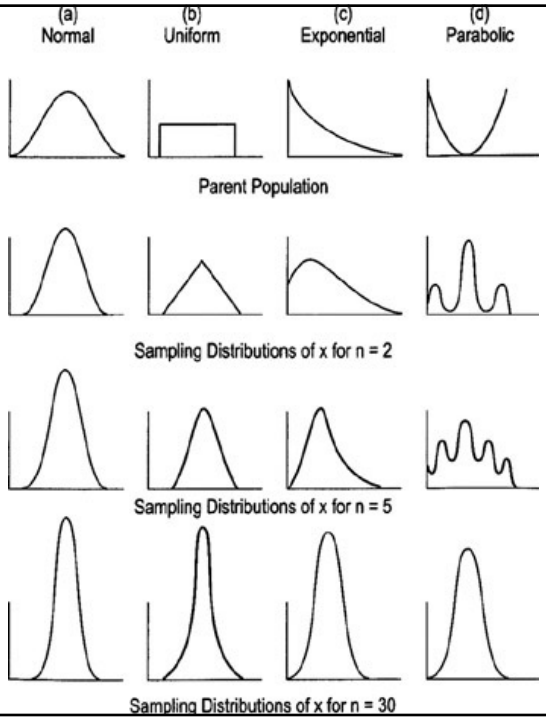
- Consider a set of n independent variables x_i, taken at random from a population with mean μ and variance σ^2, and then calculate the mean ($\bar{x}$) of these n values.
- If we repeat this procedure many times, since the individual x_i are random, then the calculated means $\bar{x}$ will have some distribution.
- *The surprising fact is that, for large n, the distribution of $\bar{x}$ tends to a Gaussian (of mean μ and variance σ^2 / n). The actual SHAPE of the distribution of the x_i is actually irrelevant.*

85

- The only important feature is that the variance σ^2 should be finite.
- Thus: If the x_i are already Gaussian distributed, then the distribution of $\bar{x}$ is also Gaussian for all values of n from 1 upwards.
  But even if the x_i have some other distribution --- say, for example, a *uniform* distribution over a finite range --- then the distribution of the sum or average of a few x_i will already look Gaussian.
- *Thus, whatever the original distribution, a linear combination of a few representatives from the distribution variables almost always approximates to a Gaussian distribution.*

86

Regardless of the shape of the parent population, the distribution of the means calculated from samples quickly approaches the normal distribution as shown below for four very different parent populations (left to right) and by doing averages of an increasing number of independent "draws" from the parent population.



87

Some practical rules of thumb:

- If the parent population is normal, $\bar{x}$ will always be normal for any sample size.
- If the population is at least symmetric, sample sizes of *n* = 5 to 20 should be OK (Gaussian-like).
- Worst-case scenario: Sample sizes of 30 should be sufficient to make $\bar{x}$ approximately normal (i.e., Gaussian) no matter how far the parent population is from being normal.
- If making a distribution of $\bar{x}$, use a standard subgroup size (e.g., all subgroups with averaging 5 observations, or 30 observations).

88

An important aspect of the Central Limit Thereom (**what makes it important for empirical science**) is that if we adopt the Gaussian distribution of the $\bar{x}$ represent the error in our knowledge of some mean measurement, , a $\bar{x}$ we adopt the width of ***that*** distribution as the 1-σ error of the derived mean value , ther $\bar{x}$ e find that

$$\sigma^2_{\bar{x}} = \sigma^2 / n$$

$$\sigma_{\bar{x}} = \sigma / n^{1/2}$$

We call $\sigma_{\bar{x}}$ the *standard deviation of the mean* or the *error in the mean*.

The above implies the the *error in the mean*, $\sigma_{\bar{x}}$, is smaller than the error *in each individual measure*, σ, by the square root of *n*.

The reason why the error in the mean ($\sigma_{\bar{x}}$) is smaller than the error in individual measurements (σ) for n > 1 is that when we have more than one measure we have more than one estimate of the actual value μ. With a *distribution of measures*, you always have a better idea of where the mean might lie than with only one measure, and intuitively you know that the more measures you have (i.e., as *n* gets large) the better is your estimate of the mean value.

89

Be clear about the difference between σ and $\sigma_{\bar{x}}$:

σ is the ***spread*** in the distribution of x values, and represents the standard deviation of a single measure. It is always the same no matter how many times you repeat the experiment with the same measurement apparatus.

$\sigma_{\bar{x}}$ is the standard distribution of your estimates of $\bar{x}$, and both $\bar{x}$ and σ can always be improved by increasing the number of measures *n*, with the latter improving as $\sigma / n^{1/2}$.

90

Propagation/Combination of Errors

We are frequently confronted with a situation where the result of an experiment is given in terms of two (or more) measurements (of either the same or different types).

We want to know what is the error on the final answer in terms of the errors on the individual measurements.

We first consider in detail the case where the answer is a linear combination of the measurements. Then we go on to consider products and quotients and the general case of combining errors.

91

As a very simple illustration, consider:

$$a = b - c$$

Provided that the errors on *b* and *c* are *uncorrelated*, the rule is that we add the contributions of error in *b* and *c* in quadrature (because we have seen above that in dealing with errors it is most sensible to consider RMS deviations):

$$\sigma_a^2 = \sigma_b^2 + \sigma_c^2$$

The errors in two measurements are *uncorrelated* when the measurement of one variable has no bearing on the measurement of another -- they are independent.

92

Combining Results of Experiments Having Differing Errors

When several experiments measure the same physical quantity and give a set of answers a_i with different errors σ_i, then the best estimates of *a* and its accuracy σ are given by:

$$a = \Sigma(a_i/\sigma_i^2) / \Sigma(1/\sigma_i^2)$$

and

$$1/\sigma^2 = \Sigma(1/\sigma_i^2)$$

Thus each experiment is to be weighted by a factor of $1/\sigma_i^2$ In some sense, $1/\sigma_i^2$ gives a measure of the information content of that particular experiment.

The weighting to get the mean *a* above makes intuitive sense: We want contributions from experiments with worse errors to be less than the contributions from experiments with better errors.

The simplest case is when all the errors σ_i are equal. Then the best combined value a from the above equation becomes the ordinary average of the individual measurements of a_i, and the error σ on a is $\sigma_i/\sqrt{N}$, where N is the number of measurements.

93