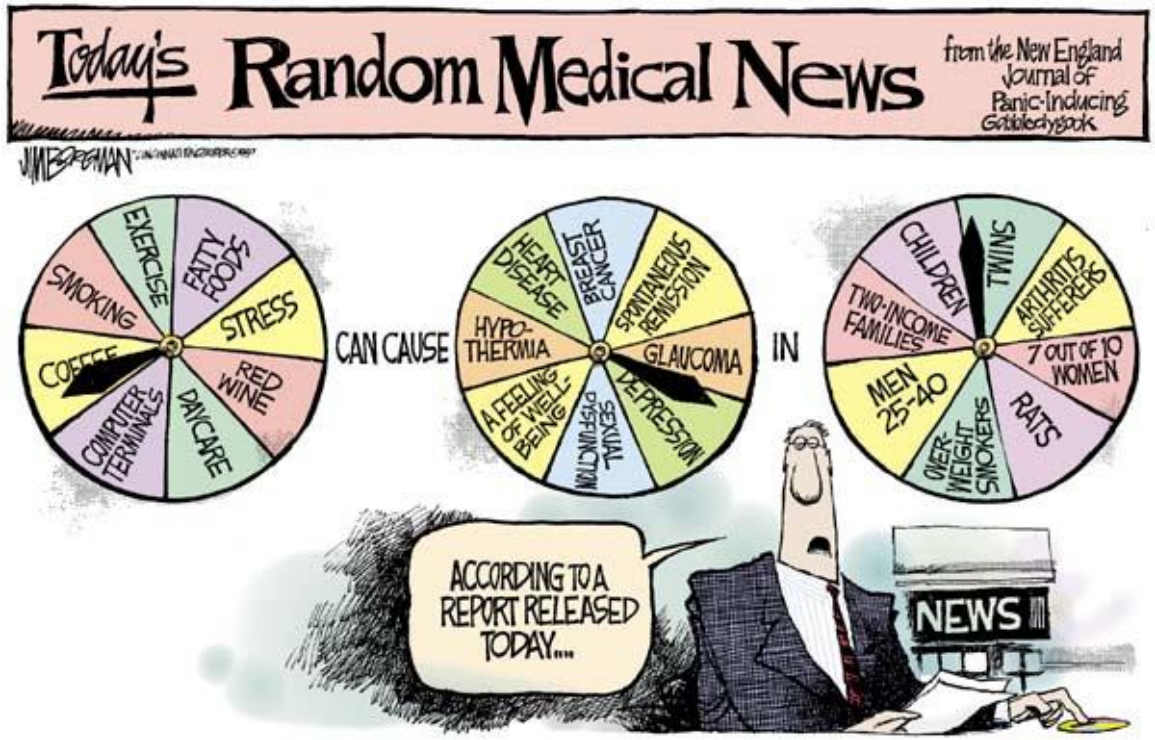
Error Analysis



V. Lorenz
L. Yang, M. Grosse Perdekamp, D. Hertzog, R. Clegg
PHYS403
Spring 2016

Reporting measurement results

- Always include uncertainty estimates in your results
- Have the correct number of significant digits
- Examples:
  - Origin fit result:
    V = 0.122349 m/s, σ = 0.01298 m/s
    You should report V = 0.122 ± 0.013 m/s
  - Measurement result:
    T = 3.745 x 10^{-3} s, σ = 0.0798 ms
    You should report T = (3.75 ± 0.08) x 10^{-3} s

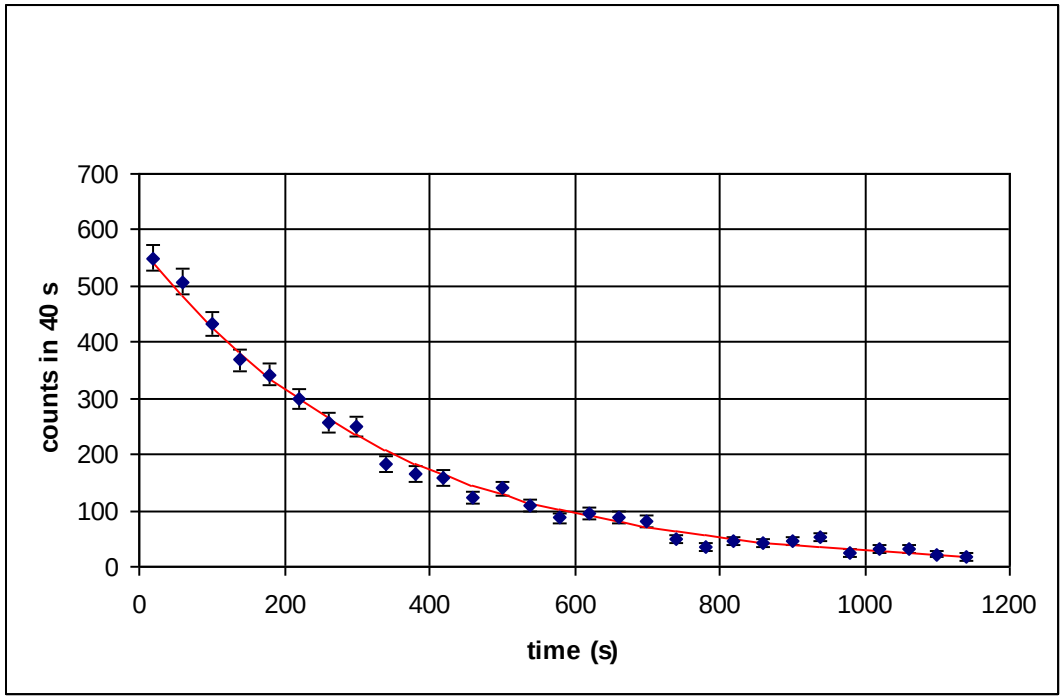
Types of Uncertainties

- **Statistical**
  - Uncertainties due to stochastic fluctuations
  - Generally there is no correlation between successive measurements.
  - Multiple measurements can be used to reduce the uncertainty.
- **Systematic**
  - Uncertainties associated with imperfect knowledge of measurement apparatus, other physical quantities needed for the measurement, or the physical model used to interpret the data.
  - Generally correlated between measurements. Cannot be reduced by multiple measurements.
  - Better calibration, or measurement of other variable can reduce the uncertainty.

Example: Statistical, Error Propagation

Lifetime Measurement

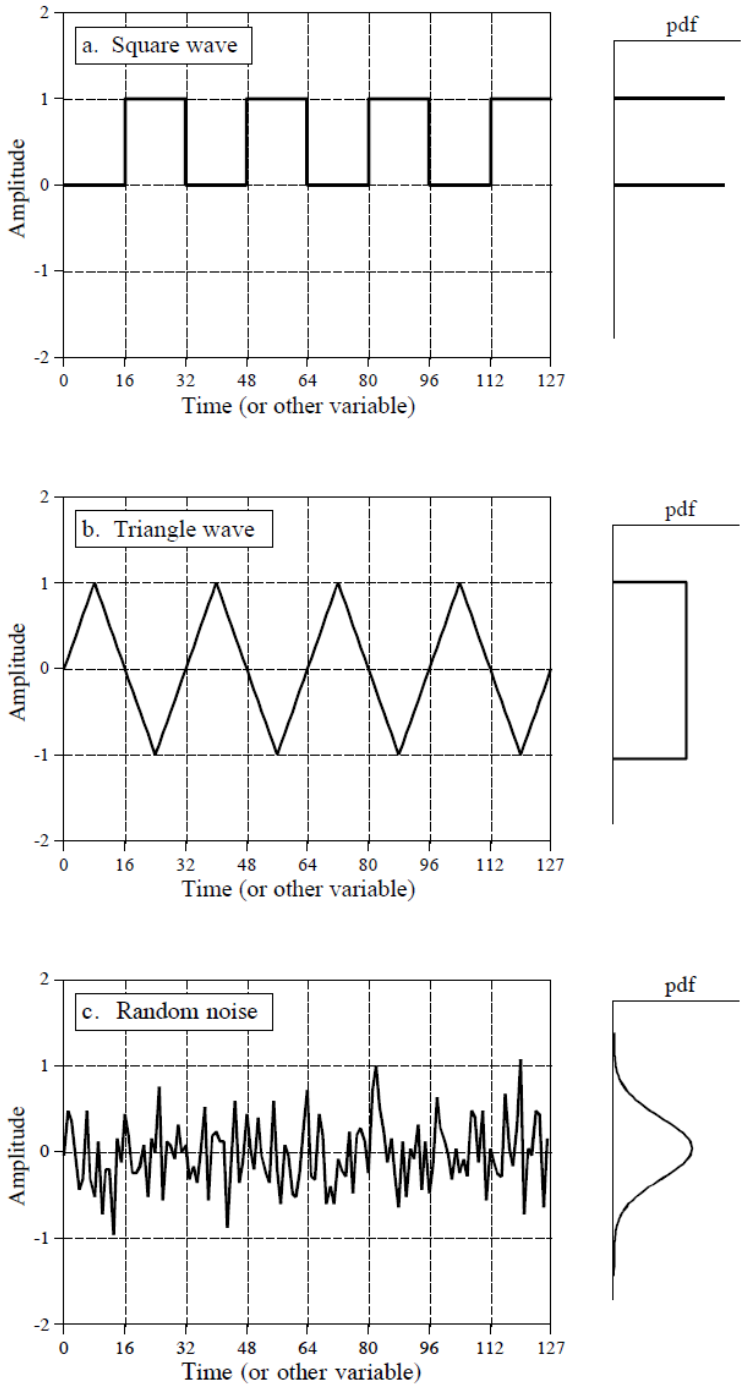
$$N(t) = N_0 e^{-\lambda t} \quad , \quad t_{1/2} = \frac{\ln 2}{\lambda}$$



Measurement:

- observe radioactive decay
- measure counts/Δt vs t
- exponential fit to determine decay constant (or lifetime)

Examples of probability density functions



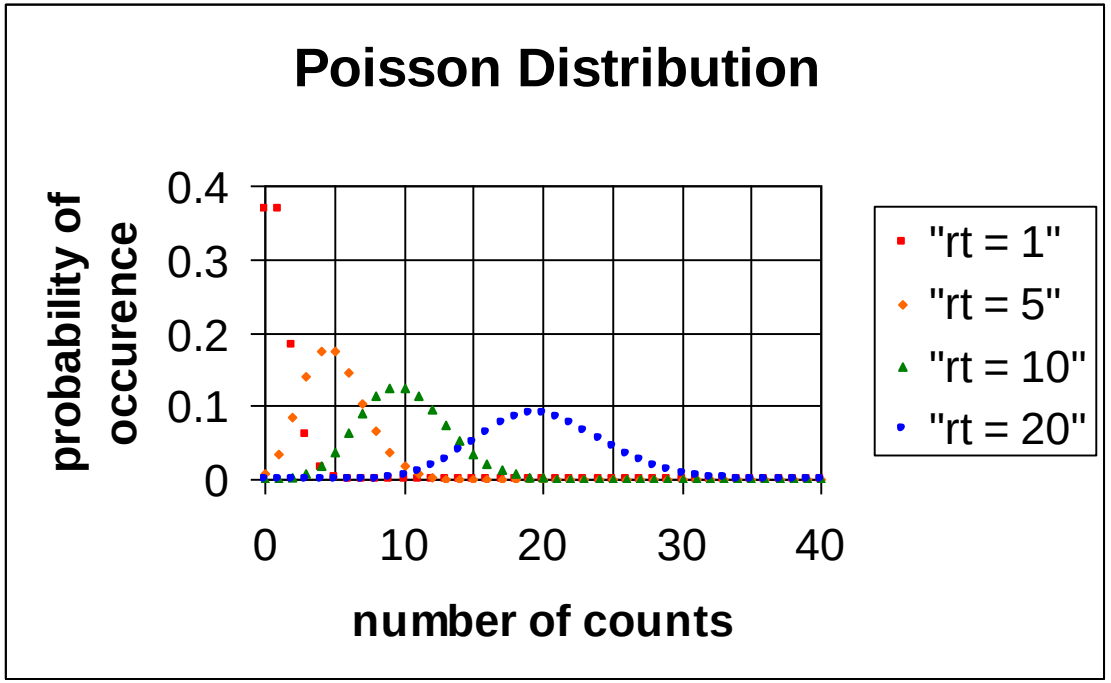
The Poisson Distribution

$$P_n(rt) = \frac{(rt)^n}{n!} e^{-rt}, \quad n = 0, 1, 2, \ldots$$

r: decay rate [counts/s]

t: time interval [s]

$P_n(rt)$: Probability to have *n* decays in time interval *t*!



A statistical process is described through a Poisson Distribution if:

- **random process** → for a given nucleus probability for a decay to occur is the same in each time interval
- **universal probability** → the probability to decay in a given time interval is same for all nuclei
- **no correlation between two instances** → the decay of one nucleus does not change the probability for a second nucleus to decay

The Poisson Distribution

$$P_n(rt) = \frac{(rt)^n}{n!} e^{-rt}, \quad n = 0, 1, 2, \ldots$$

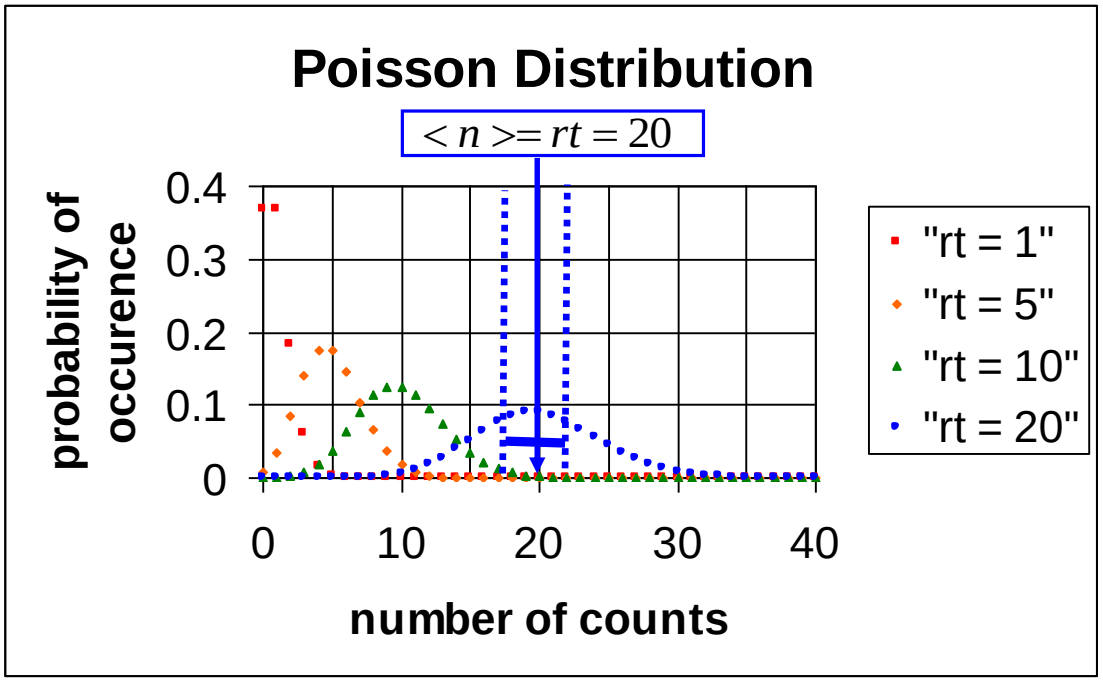
r: decay rate [counts/s]

t: time interval [s]

$P_n(rt)$: Probability to have *n* decays in time interval *t*!

Is nuclear decay a random process?
Yes, follows Poisson Distribution!

(Rutherford and Geiger, 1910)



$$\sum_{n=0}^{\infty} P_n(rt) = 1 \ , \ \text{probabilities sum to 1}$$

$$< n >= \sum_{n=0}^{\infty} n \cdot P_n(rt) = rt \ , \quad \text{the mean}$$

$$\sigma = \sqrt{\sum_{n=0}^{\infty} (n - < n >)^2 P_n(rt)} = \sqrt{rt} \ ,$$

standard deviation

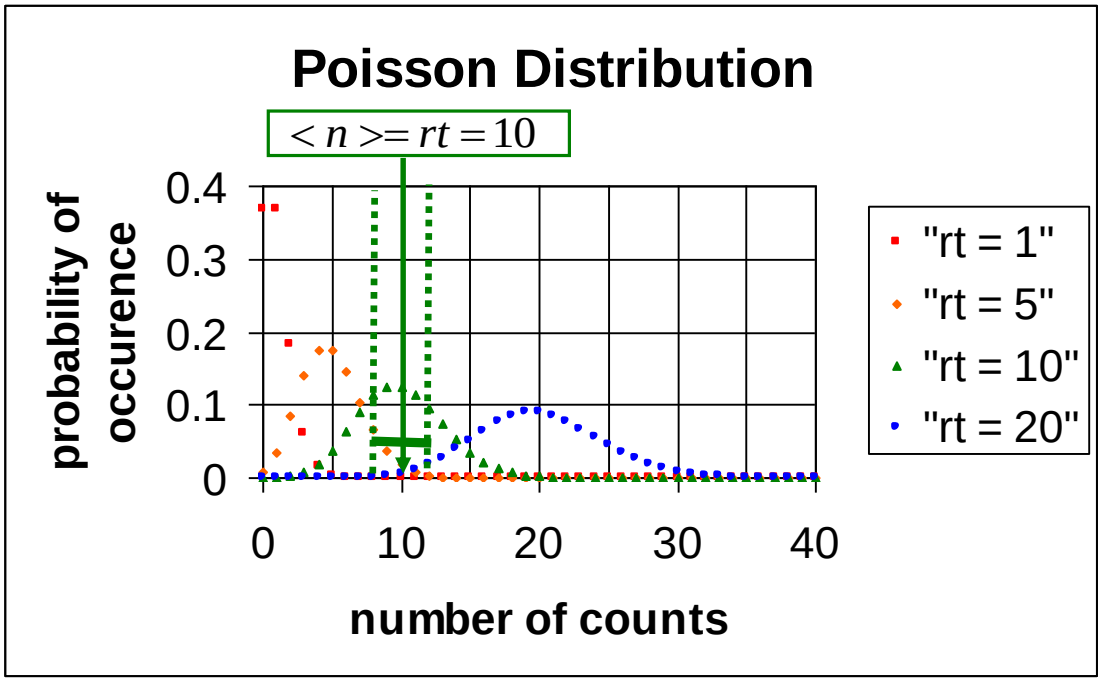
The Poisson Distribution

$$P_n(rt) = \frac{(rt)^n}{n!} e^{-rt}, \quad n = 0, 1, 2, \ldots$$

r: decay rate [counts/s]

t: time interval [s]

$P_n(rt)$: Probability to have *n* decays in time interval *t*!

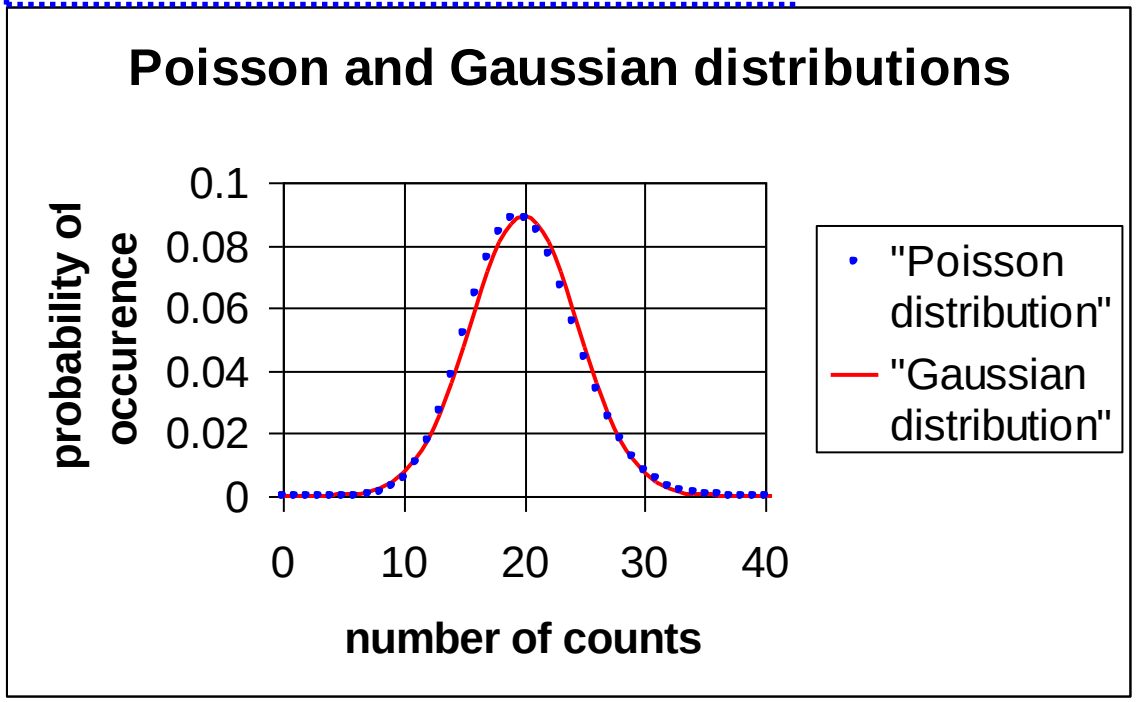


n	$P_n(rt = 10)$
0	4.5x10-5
5	0.038
10	0.125
15	0.035
20	0.002

The Poisson Distribution at large *rt*

$$P_n(\langle n\rangle = rt) = \frac{(\langle n\rangle)^n}{n!} e^{-\langle n\rangle}$$

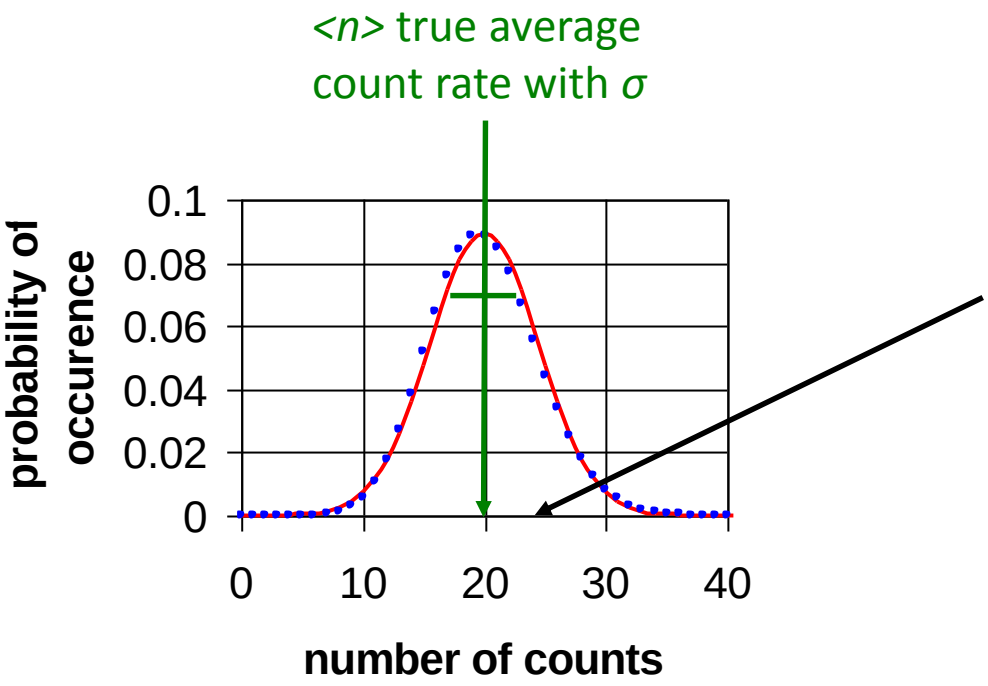
Poisson distribution: discrete



$$P_n(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(x-\langle n\rangle)^2}{2\sigma^2}}$$

Gaussian distribution: continuous

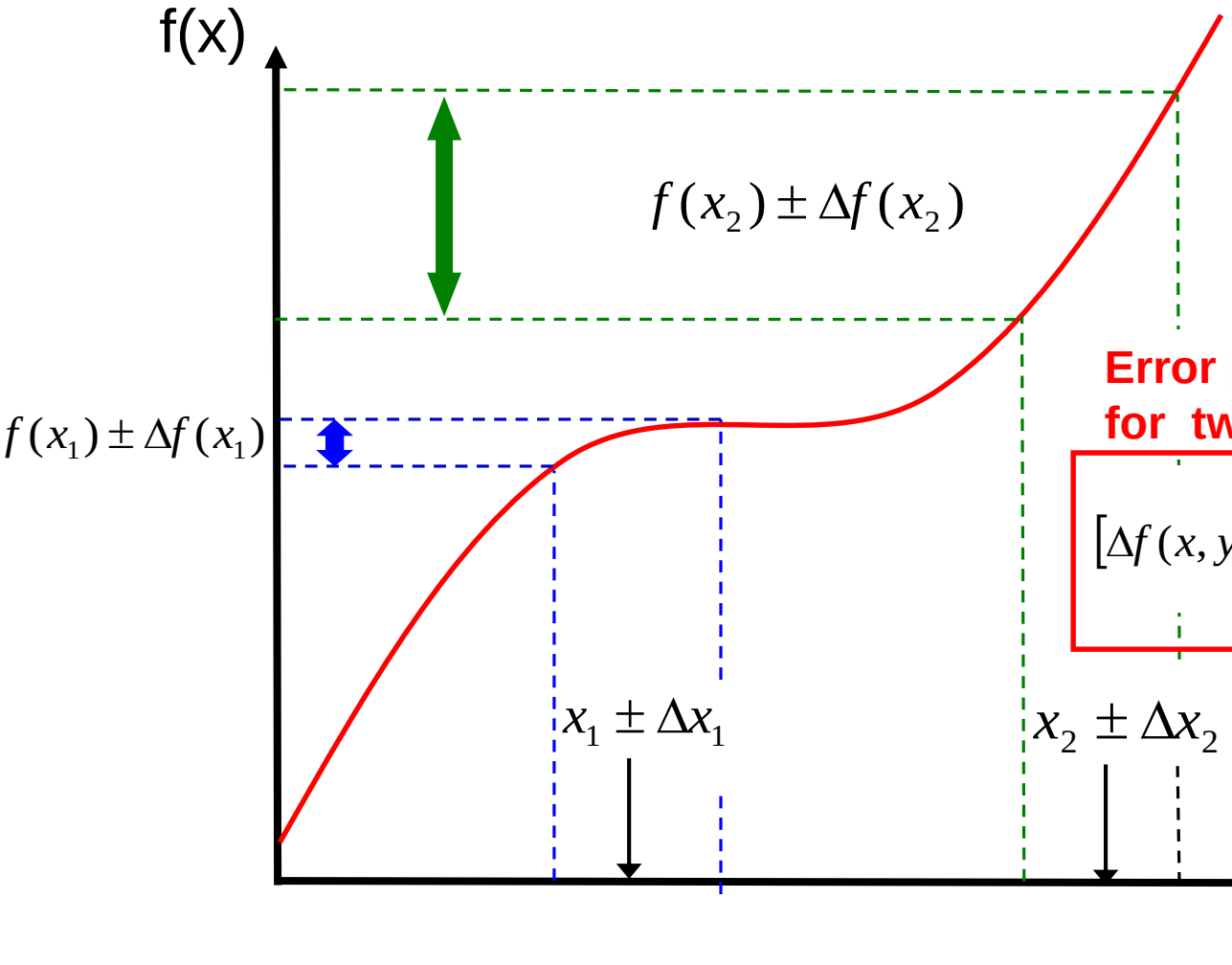
Measured count rate and errors



Single measurement $N_m = r_m t = 24$
And $\sigma_m = \sqrt{N_m} = \sqrt{24} = 4.5$

Experimental result: $\langle n \rangle = 24 +/- 4.5$

Propagation of errors



Error propagation for one variable:

$$[\Delta f(x)]^2 = \left[\frac{\partial f}{\partial x}\right]^2 \cdot \Delta x^2$$

Error propagation for two variables:

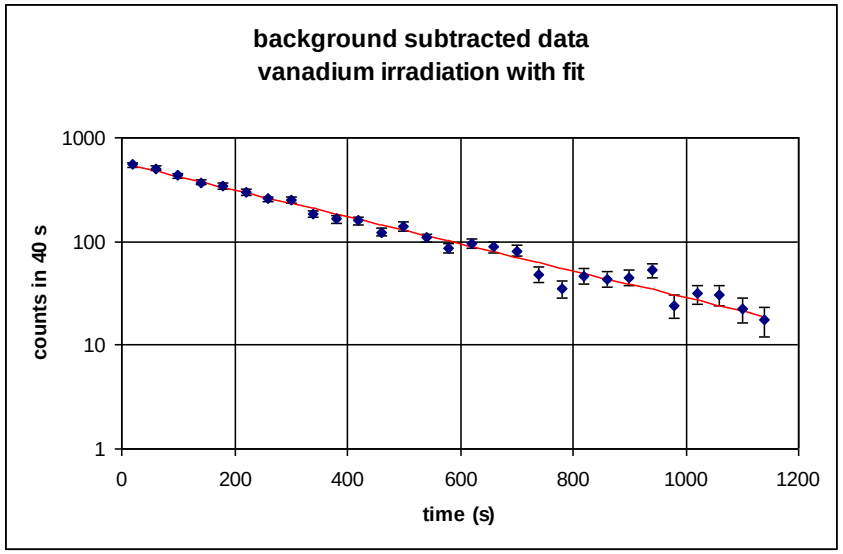
$$[\Delta f(x,y)]^2 = \left[\frac{\partial f}{\partial x}\right]^2 \cdot \Delta x^2 + \left[\frac{\partial f}{\partial y}\right]^2 \cdot \Delta y^2$$

Example I, Error on Half-Life

Propagate error in decay constant λ into half life:

$$t_{1/2} = \frac{\ln 2}{\lambda}, \lambda = 2.944 \pm 0.092 \cdot 10^{-3} s^{-1}$$

$$\Delta t_{1/2} = \sqrt{\left(\frac{\ln 2}{\lambda^2}\right)^2 \cdot \Delta\lambda^2} = 7.4s$$



Example II, Rates for $\gamma\gamma$ Correlations

Measured coincidence rate: $S' = S + B$, $\Delta S' = \sqrt{S'}$

Measured background rate: B, $\Delta B = \sqrt{B}$

Signal: $S = S' - B$

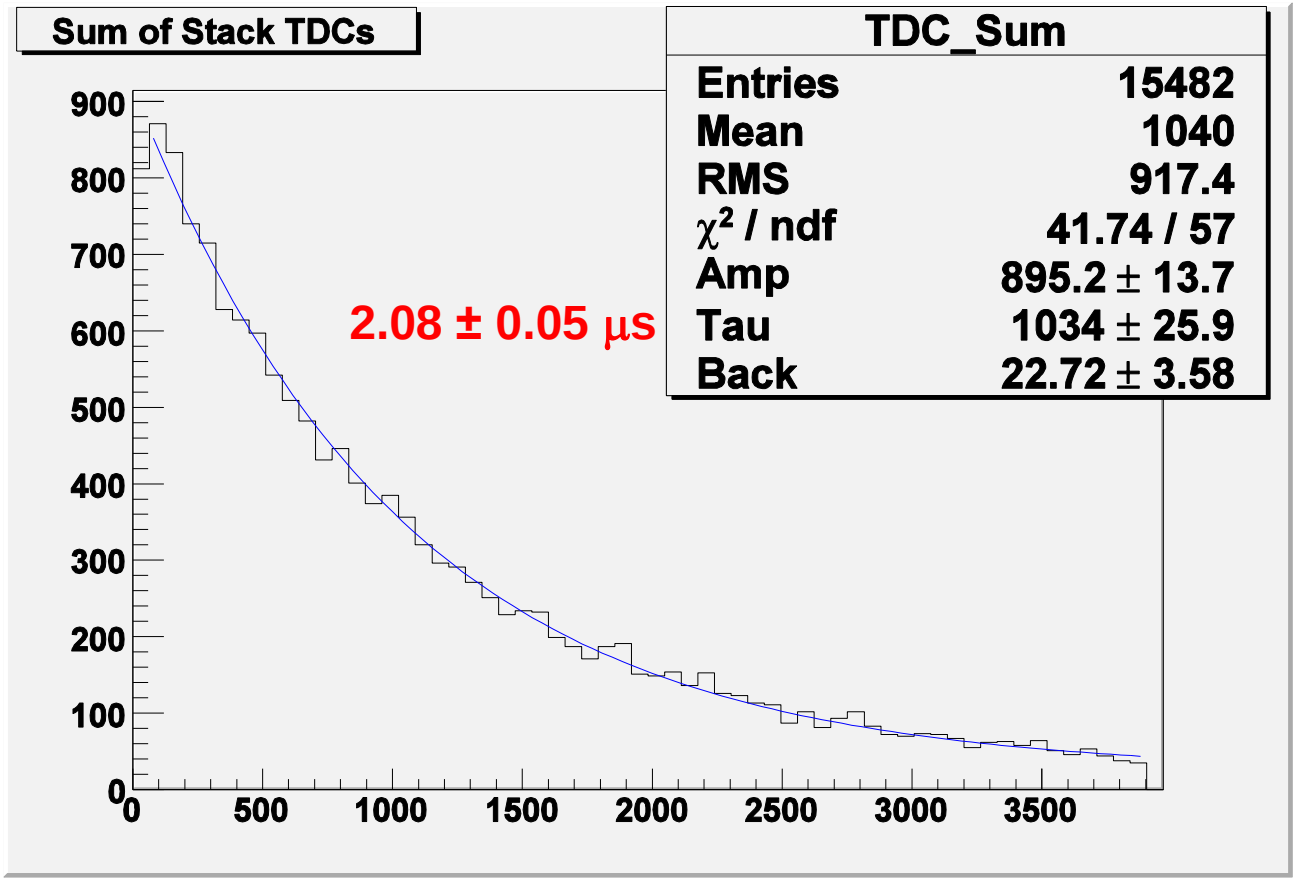
Error:

$$\left[\Delta f(x_i)\right]^2 = \left[\frac{\partial f}{\partial x_i}\right]^2 \cdot \Delta x_i^2$$

$$\begin{cases} \dfrac{\partial S}{\partial B} = -1, \quad \dfrac{\partial S}{\partial S'} = 1, \quad \Delta B = \sqrt{B}, \ \Delta S' = \sqrt{S'} \\ \Rightarrow (\Delta S)^2 = \left(\dfrac{\partial S}{\partial B}\right)^2 \Delta B^2 + \left(\dfrac{\partial S}{\partial S'}\right)^2 \Delta S'^2 \\ \qquad\qquad = \quad 1 \quad \cdot B \quad + \ 1 \quad \cdot \ S' \\ \Rightarrow \ \Delta S \quad = \sqrt{B + S'} \end{cases}$$

Interpreting fitting results

χ^2: the extent to which the data match the assumed distribution



$$\chi^2 / dof = 1 \pm \sqrt{\frac{2}{\#dof}}$$

Here, 0.73 ± 0.18; a bit "low" but okay.

Too low means errors are underestimated

Too high means fit is bad

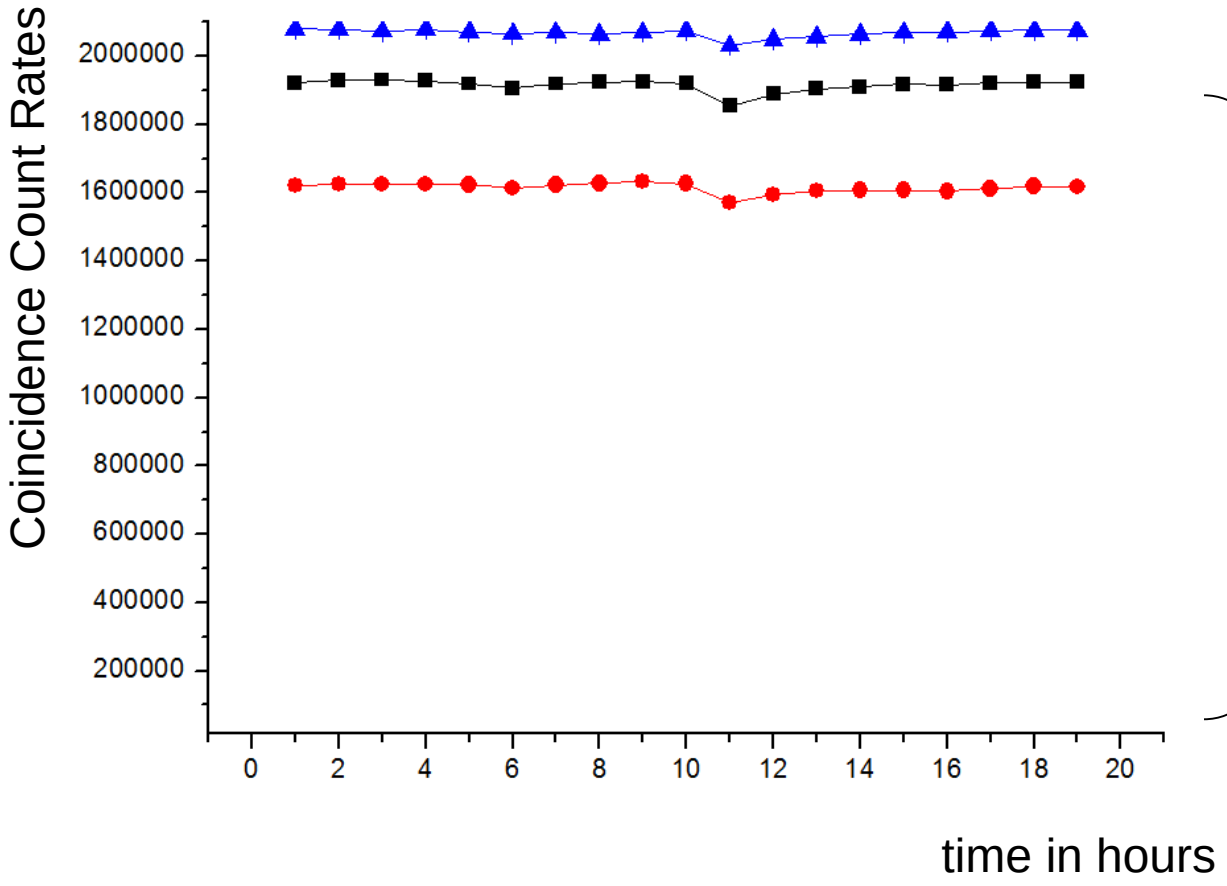
Some of you will be comparing a slightly low lifetime with the real one of 2.2 μs. Yours will usually be lower due to negative muon capture.

Let's consider systematic uncertainties

A table is helpful; consider how you might develop and fill in a table for the Quantum Eraser (or other optics) experiment

Error	Value	Method to determine
Synchronization of the reconvergence of the split beams		
Instability of interference pattern due to air fluctuations	__% in reading high/low	Stability of reading vs time
Beam splitter imperfections	__% on visibility	

From the $\gamma\gamma$ experiment:



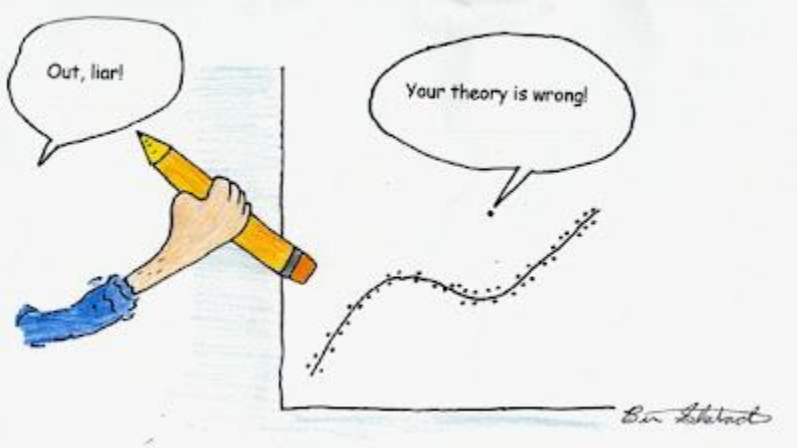
Systematic test of count rate stability ➔ stability of detectors + electronics

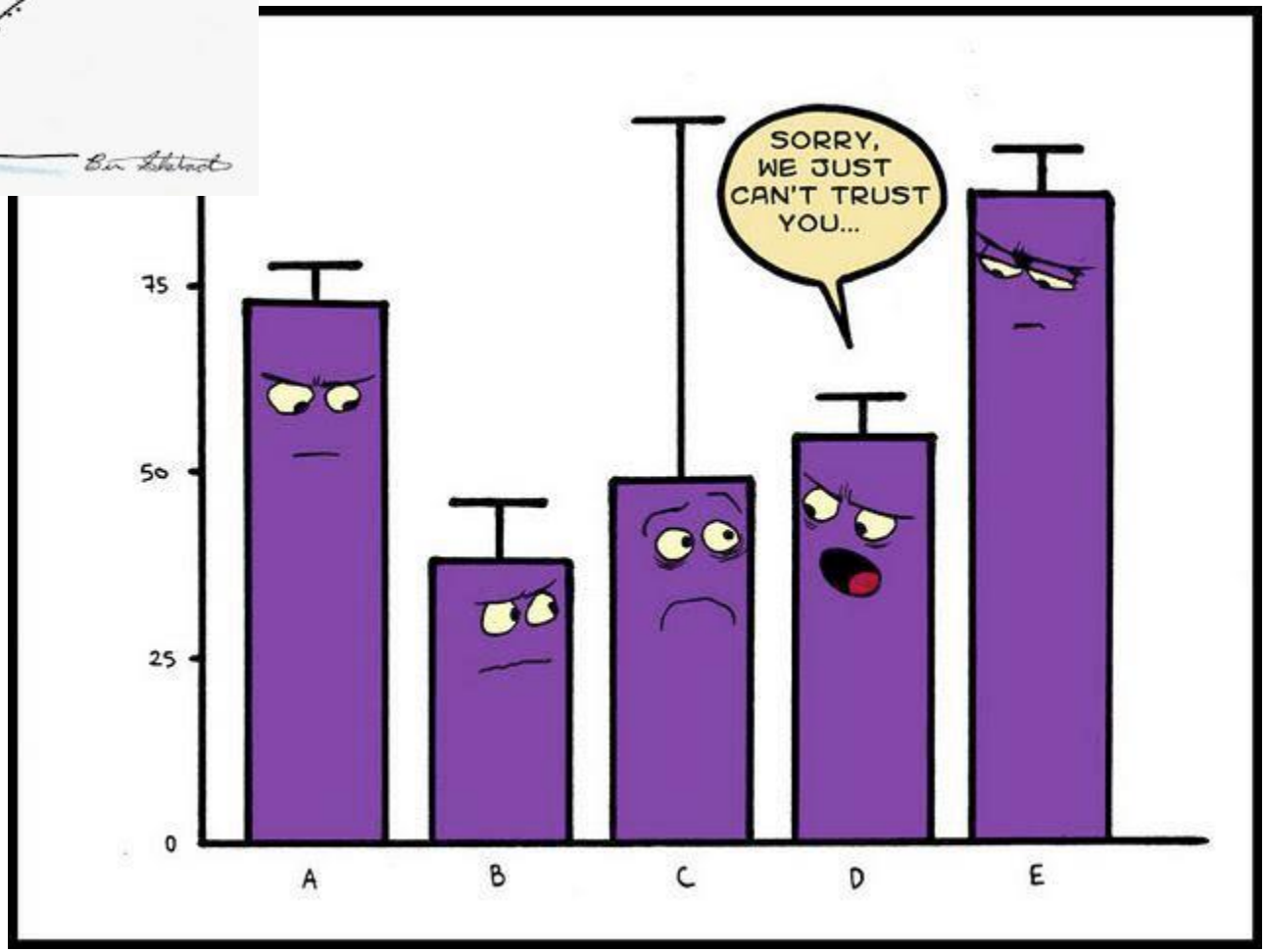
➔Large drifts , temperature effects from power dissipation in discriminators

In Ferroelectric analysis:

- Calibration of temperature as sample is being evaluated. How accurate is it? How well are phase transitions identified?
- Quality of particular sample. Would 2nd , 3rd "identical" sample produce same results?
- General reproducibility of traces for multiple paths.
  - Do results depend on speed of change in T?
  - How about the direction of the T change (heating vs cooling?)
    - How do you tell the difference between physics that depends on heating versus cooling compared to just instrumentation effects, like time lag ?
    - Temp lag between thermometer and sample…leads to ?
  - How do external factors like quality of the lead connections affect results? Are connections being evaluated too?

Data rejection





Data rejection

What if an experiment doesn't give the result you expected?

What if it gives a result that you just **know** is wrong in some way?

Do you keep trying until you get the "right" result?

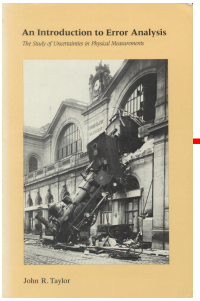
Note: especially relevant here in our modern physics lab course where the "right" result was, in general, published a long time ago

This happens. Data rejection is common. But be careful.

Realities of complex experiments
- Stuff goes wrong
- Equipment malfunctions
- People make mistakes

Burden on the physicist
- Record everything

Responsibility of physicist
- Develop a "result-unbiased" algorithm for data rejection
  - Make decisions before you look at the results
  - Keep answer in a "blind" or unbiased space
  - You can rarely use the result to determine inclusion



Rejection of Data

from J. Taylor, Ch. 6 of An Introduction to Error Analysis

Consider 6 measurements of a pendulum period : **3.8, 3.5, 3.9, 3.9, 3.4, 1.8**

Should the last measurement be rejected?

Yes: If some aspect of the experiment was changed ... new "slow" stopwatch, etc.

No: Never! You must always keep all data !! (diehards; beware)

Maybe? The usual case. You don't know why, but something may have made this measurement "bad." How do you judge in an unbiased manner ?

Chauvenet's Criterion

The probability (assuming a Gaussian distribution) is 0.05 for this to be an acceptable measurement. What's wrong with that? We would even *expect* that 1 out of 20 measurements would fall outside of the 2σ bound.

But, we only made 6 measurements.

So, we expect that only 1/(20/6) = 0.3 measurements should fall outside the 2σ bound.

Now it is a bit about personal taste. Is this unreasonable?

Chauvenet's criterion is the following: If the suspect measurement has a lower probability than 1/2, you should reject it. Our measurement has 0.27 so it goes.

Is all data good data? NO!

- Write down everything
  - in the logbook; take your time; use sentences; record numbers (values)
  - glitch in the power? note the time
  - temperature "cold" or "hot"? comment about it
  - somebody "reset" the system? note it please and when
- Record (electronically if possible) everything reasonable
  - as parallel information to the main data set
  - temperatures; voltages; generally called "slow controls"
- You *WILL* (almost certainly) have to go back and hunt for this documentation when something possibly anomalous arises … and it will

Some additional points

- Data rejection does exist and is necessary.
  - If you can document a problem, then it is easy to discard
  - There still may be some data you would like to throw out.
    - this is tricky and takes some carefully prepared, bias-free statistical tests to justify
- Theory curves can be misleading and should generally (always?) be avoided when dealing with issues of data rejection
- You must also think in reverse. How self-consistent is your data set?
  - There are then many sophisticated tests of the data set itself
  - You will be expected to demonstrate this in many cases

Summary (for your report)

- Always include uncertainty estimates for all your measurements if applicable (use correct number of significant digits)
- Compare your results with published values if applicable
  - Do your measurements agree within uncertainty?
  - If not, is your estimate of systematic or statistical uncertainty correct? Are there other factors that can influence your result that you forgot to consider?
- If you need to reject certain sets or points of data, you should describe the reason that data should not be included. The reason should be based on changes in environment, setup, etc., and not solely result driven.

Error Propagation

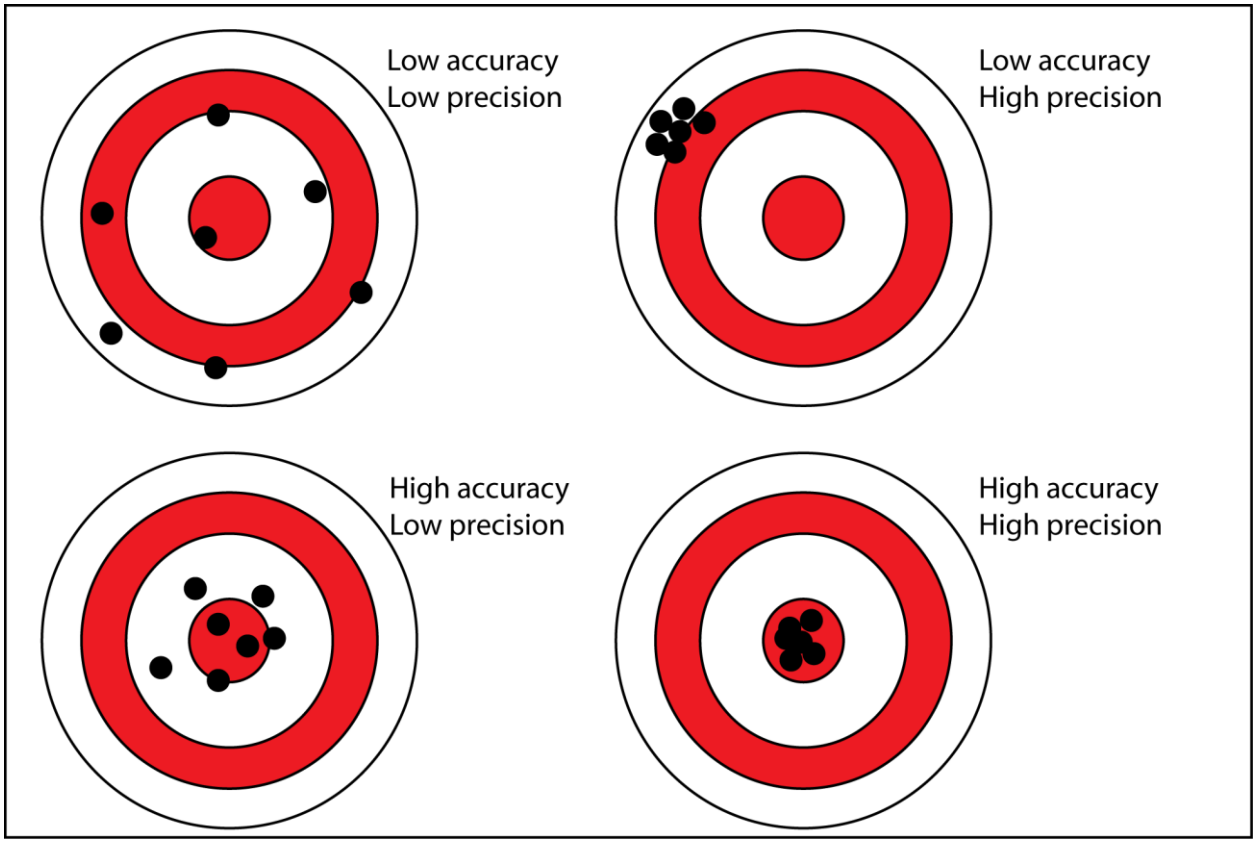
Function	Variance	Standard Deviation
$f = aA$	$\sigma_f^2 = a^2\sigma_A^2$	$\sigma_f = \lvert a\rvert\sigma_A$
$f = aA + bB$	$\sigma_f^2 = a^2\sigma_A^2 + b^2\sigma_B^2 + 2ab\,\sigma_{AB}$	$\sigma_f = \sqrt{a^2\sigma_A^2 + b^2\sigma_B^2 + 2ab\,\sigma_{AB}}$
$f = aA - bB$	$\sigma_f^2 = a^2\sigma_A^2 + b^2\sigma_B^2 - 2ab\,\sigma_{AB}$	$\sigma_f = \sqrt{a^2\sigma_A^2 + b^2\sigma_B^2 - 2ab\,\sigma_{AB}}$
$f = AB$	$\sigma_f^2 \approx B^2\sigma_A^2 + A^2\sigma_B^2 + 2AB\,\sigma_{AB}$	$\sigma_f \approx \sqrt{B^2\sigma_A^2 + A^2\sigma_B^2 + 2AB\,\sigma_{AB}}$
$f = \frac{A}{B}$	$\sigma_f^2 \approx f^2\left[\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 - 2\frac{\sigma_{AB}}{AB}\right]$ [11]	$\sigma_f \approx \lvert f\rvert\sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 - 2\frac{\sigma_{AB}}{AB}}$
$f = aA^b$	$\sigma_f^2 \approx \left(abA^{b-1}\sigma_A\right)^2 = \left(\frac{fb\sigma_A}{A}\right)^2$	$\sigma_f \approx \left\lvert abA^{b-1}\sigma_A\right\rvert = \left\lvert\frac{fb\sigma_A}{A}\right\rvert$
$f = a\ln(bA)$	$\sigma_f^2 \approx \left(a\frac{\sigma_A}{A}\right)^2$ [12]	$\sigma_f \approx \left\lvert a\frac{\sigma_A}{A}\right\rvert$
$f = a\log_{10}(A)$	$\sigma_f^2 \approx \left(a\frac{\sigma_A}{A\ln(10)}\right)^2$ [12]	$\sigma_f \approx \left\lvert a\frac{\sigma_A}{A\ln(10)}\right\rvert$
$f = ae^{bA}$	$\sigma_f^2 \approx f^2\left(b\sigma_A\right)^2$ [13]	$\sigma_f \approx \lvert f\left(b\sigma_A\right)\rvert$
$f = a^{bA}$	$\sigma_f^2 \approx f^2\left(b\ln(a)\sigma_A\right)^2$	$\sigma_f \approx \lvert f\left(b\ln(a)\sigma_A\right)\rvert$
$f = A^B$	$\sigma_f^2 \approx f^2\left[\left(\frac{B}{A}\sigma_A\right)^2 + \left(\ln(A)\sigma_B\right)^2 + 2\frac{B\ln(A)}{A}\sigma_{AB}\right]$	$\sigma_f \approx \lvert f\rvert\sqrt{\left(\frac{B}{A}\sigma_A\right)^2 + \left(\ln(A)\sigma_B\right)^2 + 2\frac{B\ln(A)}{A}\sigma_{AB}}$

Example formulae from Wikipedia

In most cases for this class, variables are uncorrelated, therefore the correlation term can be safely ignored. Before using the formula, you should check if your assumption about variable correlation is warranted.

Accuracy vs. Precision

- Accuracy: a measure of how close the result of the experiment comes to the true value
- Precision: a measure of how exactly the result is determined (without reference to what the result means)



Source: P. Bevington, Data Reduction and Error Analysis for the Physical Sciences

https://sites.google.com/a/apaches.k12.in.us/mr-evans-science-website/accuracy-vs-precision

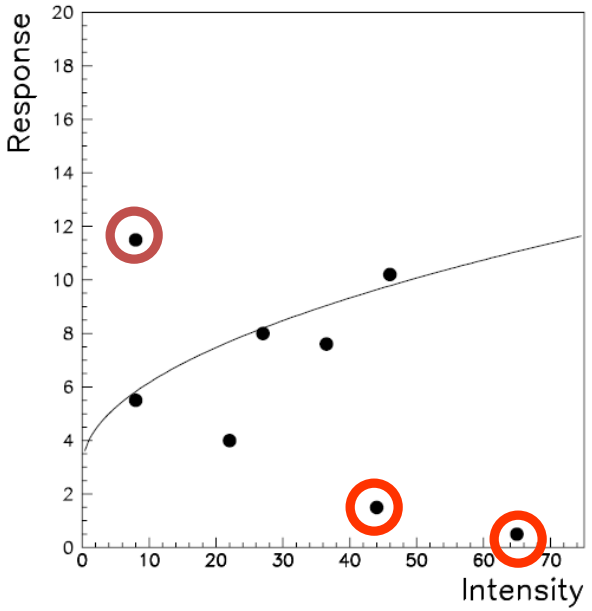
A case study

Data set (picked off the graph by hand)
11.5, 5.5, 4.0, 8.0, 7.6, 1.5, 10.2, 0.5 (note, at same beam intensity!)

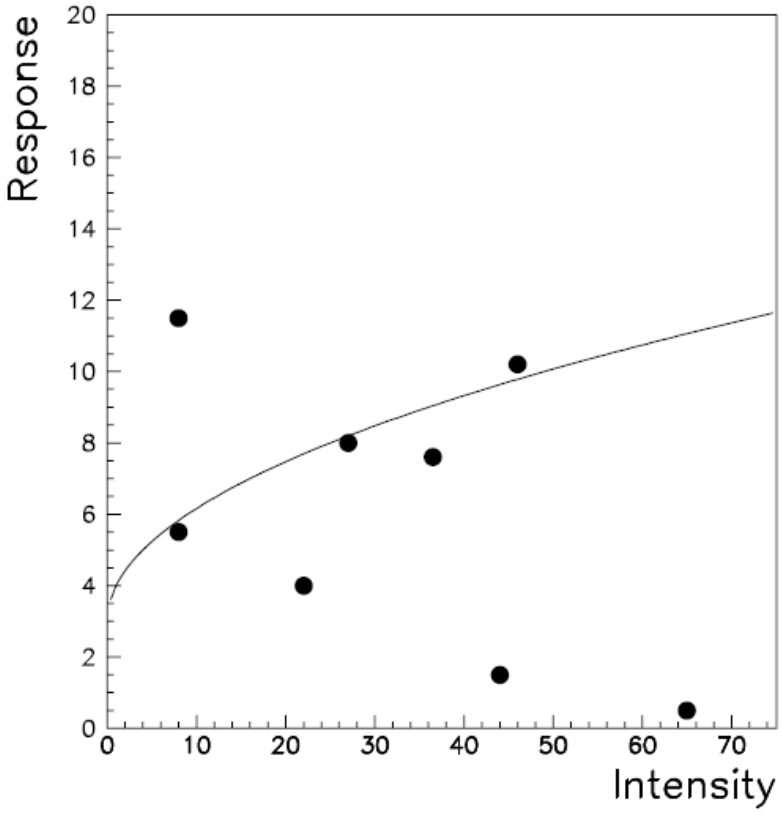
Mean: = 6.1
Standard deviation = 4.0

List of "deviations" in sigma: 1.35, -0.15, -0.53, 0.48, 0.38, **-1.15**, 1.03, **-1.40**
(these are the "bad" guys)

Data Points	prob in 1	prob in 8
(8,11.5)	0.09	0.53
(44,1.5)	0.07	0.44
(65,0.5)	0.15	0.72



Let’s look at our data

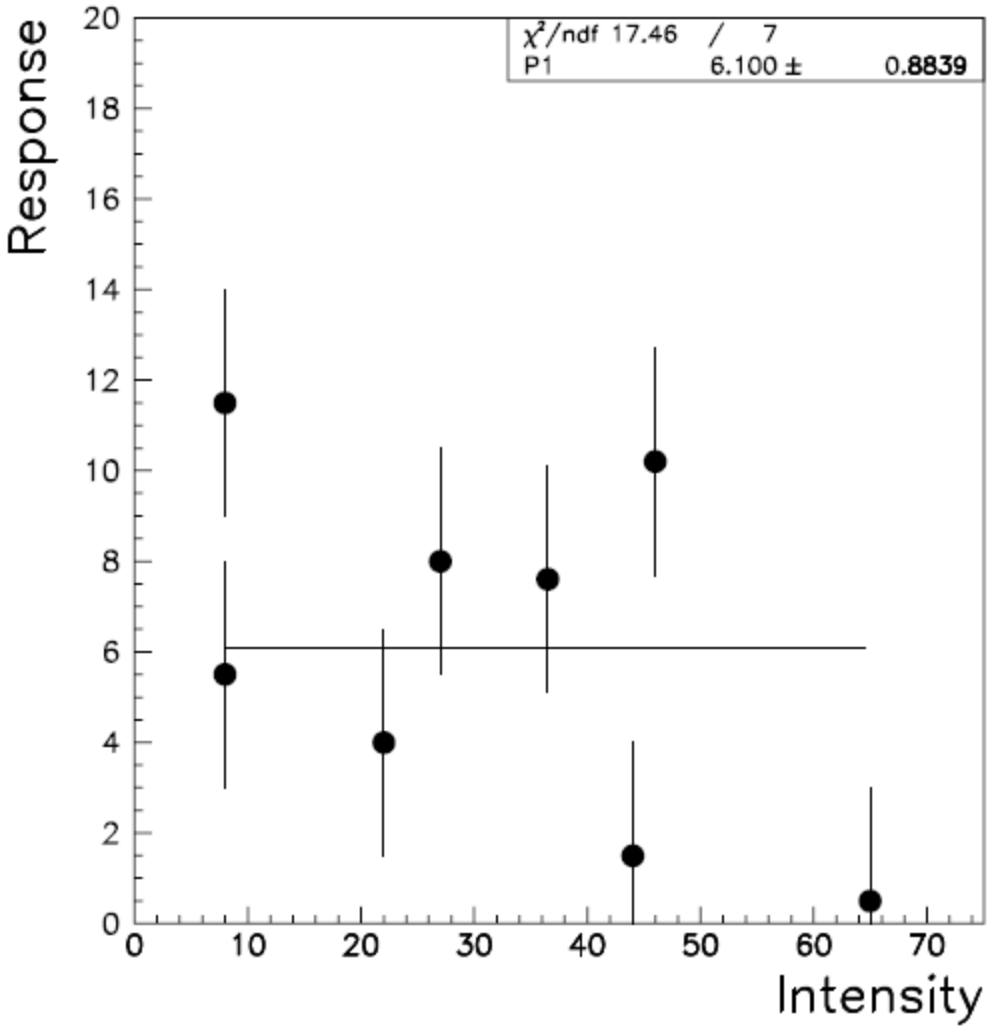


What are the uncertainties?

Can we relate power fluctuations to particular data points?

Why should we trust the theory prediction? It could be simply wrong …

Let's look at our data



$$\chi^2 / dof = 1 \pm \sqrt{\frac{2}{\#dof}}$$

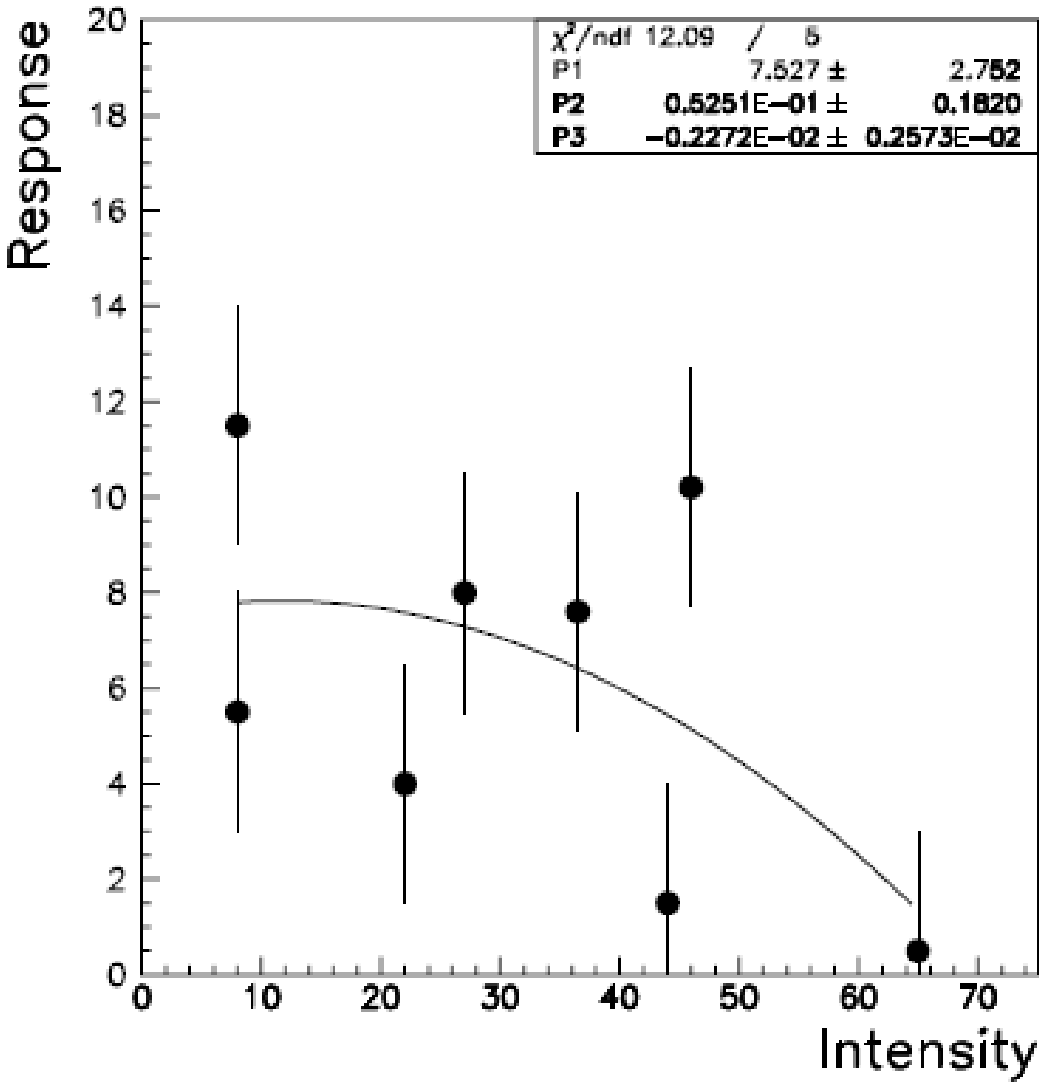
Too low means errors are underestimated

Too high means fit is bad

Assume we find the errors to be +/- 2.5 independent of beam intensity

Are the data compatible with a constant behavior?
Not sure: χ^2/ndf is 2.5

Let's look at our data



$$\chi^2 / dof = 1 \pm \sqrt{\frac{2}{\#dof}}$$

Too low means errors are underestimated

Too high means fit is bad

Are the data compatible with a polynomial?
Not sure: χ^2/ndf is 2.4

In absence of slow control data for beam & experimental apparatus, data cannot be rejected !

Uncertainty and Bias

I can live with doubt and uncertainty and not knowing. I think it is much more interesting to live not knowing than to have answers that might be wrong.

- Richard Feynman

But what if an experiment doesn't give the result you expected?

What if it gives a result that you just know is wrong in some way?

Don't you keep trying until you get the "right" result?

Note: especially relevant here in our modern physics lab course where the “right” result was, in general, published a long time ago

- Henry H. Bauer, Professor of Chemistry & Science Studies, **VPI**

How common is data “rejection”?

Answer: Common

- Realities of complex experiments
  - Stuff goes wrong
  - Equipment malfunctions
  - People make mistakes
- Burden on the physicist
  - Record everything
- Responsibility of physicist
  - Develop a “result-unbiased” algorithm for data rejection
    - Make decisions before you look at the results
    - Keep answer in a “blind” or unbiased space
    - You can rarely use the result to determine inclusion

Rejection of Data

from J. Taylor, Ch. 6 of An Introduction to Error Analysis

Consider 6 measurements of a pendulum period : **3.8, 3.5, 3.9, 3.9, 3.4, 1.8**

Should the last measurement be rejected?

Yes: If some aspect of the experiment was changed ... new "slow" stopwatch, etc.

No: Never! You must always keep all data !! (diehards; beware)

Maybe? The usual case. You don't know why, but something may have made this measurement "bad." How do you judge in an unbiased manner ?

Chauvenet's Criterion

The probability (assuming a Gaussian distribution) is 0.05 for this to be an acceptable measurement. What's wrong with that? We would even *expect* that 1 out of 20 measurements would fall outside of the 2σ bound.

But, we only made 6 measurements. So, we expect that only 0.05 x 6 = 0.3 measurements should fall outside the 2s bound.

Now, it is a bit of personal taste. Is this unreasonable?

Chauvenet's criterion is the following: If the suspect measurement has a lower probability than 1/2, you should reject it. Our measurement has 1/3 so it goes.

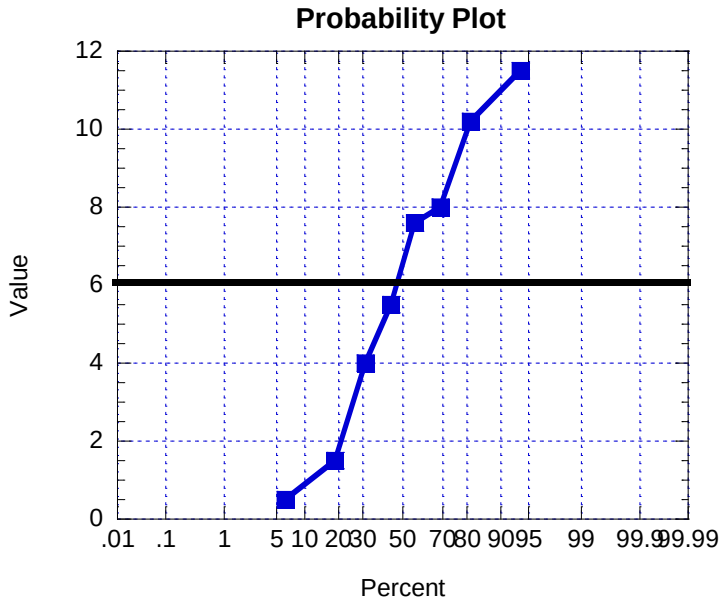
New results: Mean = **3.7**
Standard deviation = **0.2** ! much smaller !

Our case study:
A very simple first step

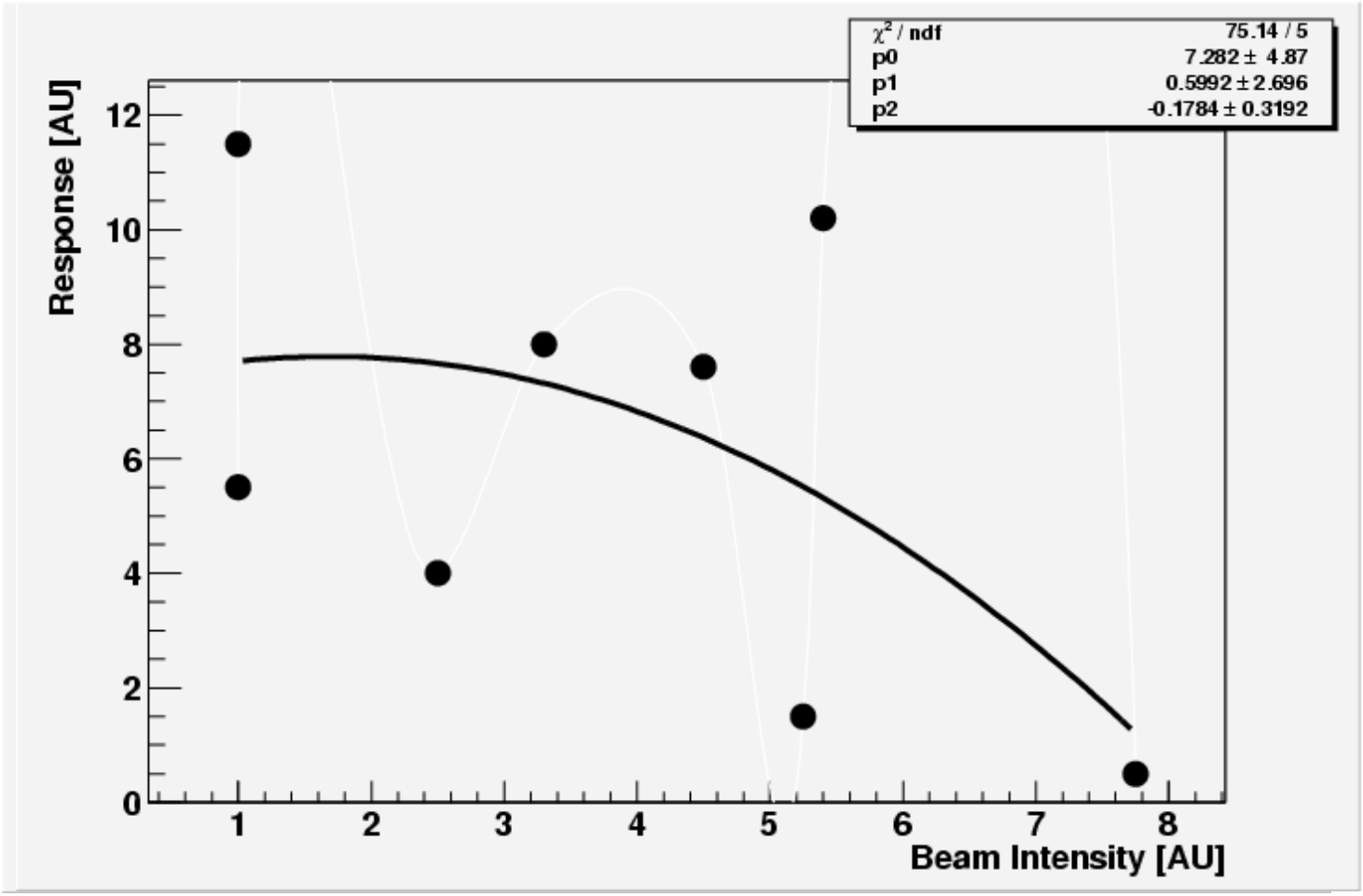
Data set (picked off the graph by hand)
11.5, 5.5, 4.0, 8.0, 7.6, 1.5, 10.2, 0.5 (note, at same beam intensity!)

Mean: = 6.1
Standard deviation = 4.0

List of "deviations" in sigma: 1.35, -0.15, -0.53, 0.48, 0.38, **-1.15**, 1.03, **-1.40**
(these are the "bad" guys)



My plot of our data



Where are the "bad" points ?

What was the theory prediction ?

Repeat: Is all data good data?

NO!

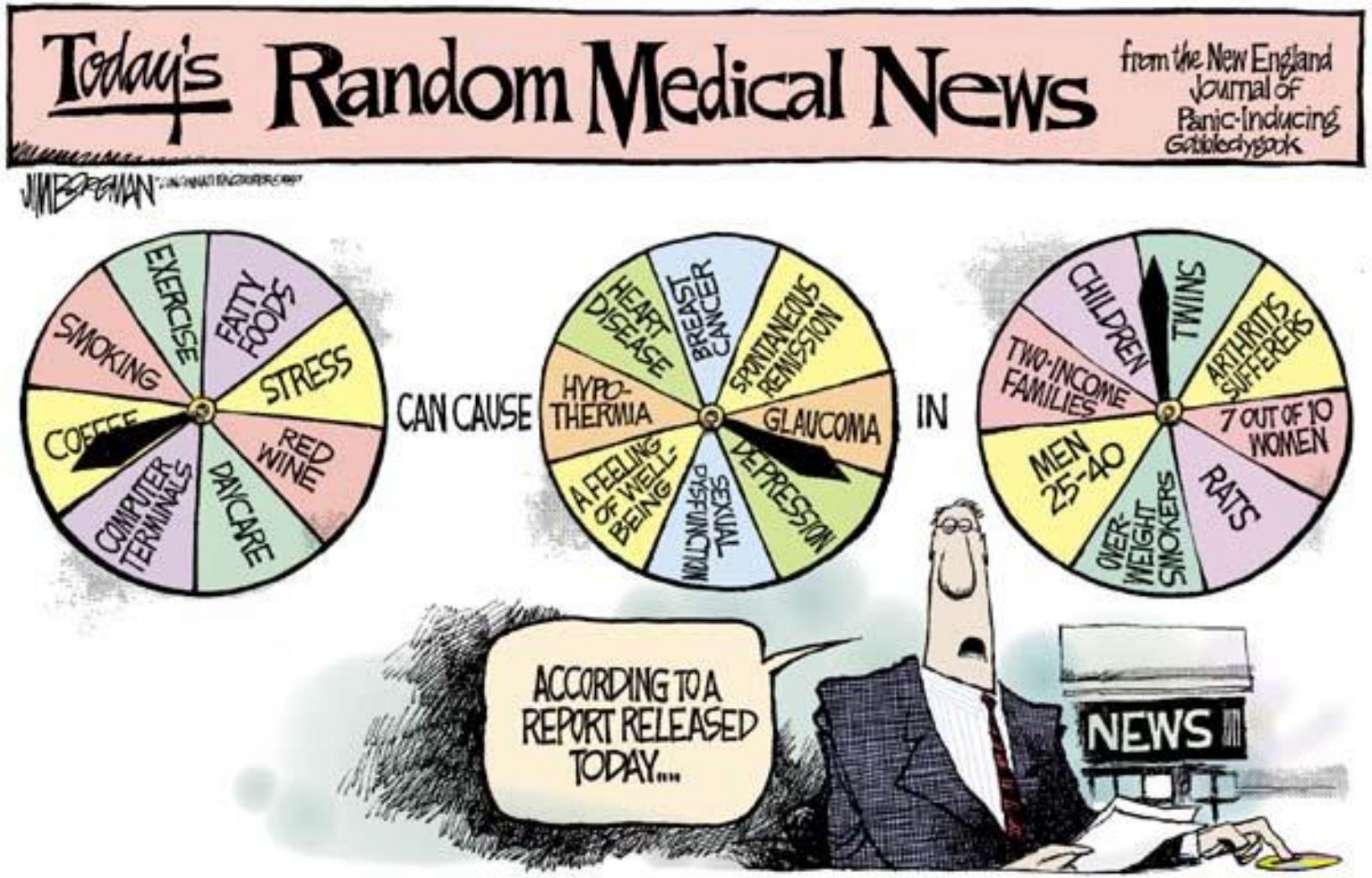
- **Write down everything**
  - **in the logbook; take your time; use sentences; record numbers (values);**
  - **glitch in the power? note the time**
  - **temperature "cold" or "hot"? comment about it**
  - **somebody "reset" the system? note it please and when**
- **Record (electronically if possible) everything reasonable**
  - **as parallel information to the main data set**
  - **temperatures; voltages; generally called "slow controls"**
- **You *WILL* (almost certainly) have to go back and hunt for this documentation when something possibly anomalous arises … and it will**

Some additional points

- **Data rejection does exist and is necessary.**
  - **If you can document a problem, then it is easy to discard**
  - **There still may be some data you would like to throw out.**
    - **this is tricky and takes some carefully prepared, bias-free statistical tests to justify**
- **Theory curves can be misleading* and should generally (always?) be avoided when dealing with issues of data rejection**
- **You must also think in reverse. How self-consistent is your data set?**
  - **There are then many sophisticated tests of the data set itself**
  - **You will be expected to demonstrate this in many cases**

*** there are trivial exceptions: counts vs intensity is linear, etc.**

The importance of statistics
And error analysis



An Introduction to Error Analysis

The Study of Uncertainties in Physical Measurements

John R. Taylor

Errors and Data Analysis

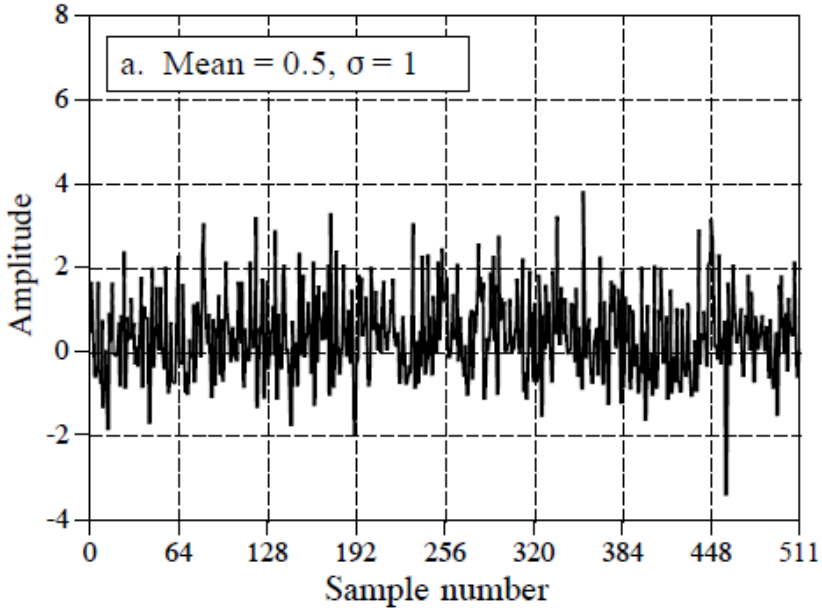
Types of errors:

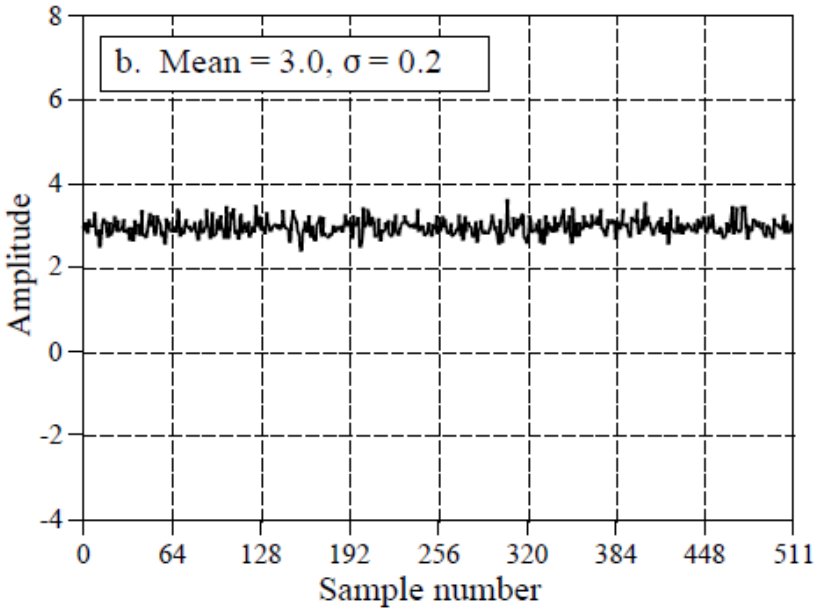
1) **Precision errors** – these are random errors. These could also be called repeatability errors. They are caused by fluctuations in some part (or parts) of the data acquisition. These errors can be treated by statistical analysis.

2) **Bias errors** – These are systematic errors. Zero offset, scale errors (nonlinear output *vs* input) , hysteresis, calibration errors, etc. If these are hidden, they are essentially impossible to correct. These are often negligible in instruments used for calibration for a long time. But new instruments and devices can easily have bias errors. For instance, when reducing scales from meters and millimeters to a scale of nanometers bias errors can creep in due to unforeseen new effects.

3) **Analysis errors** – wrong theory or wrong analysis applied to data, which are used to ”fit” the data. This is uauslly not considered as a error in the data acquisition, but nevertheless can waste a lot of time.

Examples of a constant signal and random noise from time acquired data





Where does the "randomness" come from?

- Counting statistics – small numbers (radioactive decay and photon counting
- Electronic noise from an electronic circuit
- Small number fluctuations in number of molecules or nano-sized objects

Some helpful “rules” when dealing with errors of an experimental set-up

1: As soon as an error from a particular source is seen to be significantly smaller than other errors present, it is given no further consideration.

2: The major concern of most error analyses is the quantitative estimate of bias errors, and correction of data accordingly when possible.

3: Whenever feasible, precision errors should be estimated from repeated tests or from observed scatter in graphed results.

4: In planning an experiment where it appears that significant bias errors will be present, an effort should be made to ensure that precision errors are much smaller.

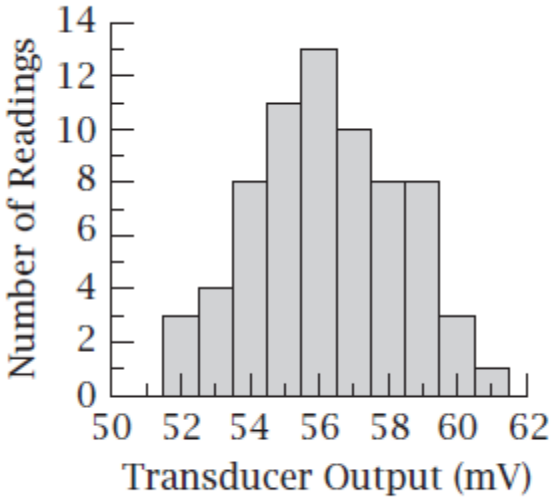
How to handle data samples of multiple measurements taken of the same configuration.

The mean value of the sample values is: $$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The usual measure of the scatter is **the standard deviation**, which is the **square root of the variance**:

$$S_x = \left[\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2\right]^{1/2}$$

Example:



Histogram of a large data sample.

Notice that the shape of the histogram is similar to the familiar normal (Gaussian) probability distribution. Indeed, most precision errors have the characteristic that, as the sample size becomes large, the shape of the histogram tends to that of the normal distribution. This characteristic allows many powerful methods of statistical analysis to be applied to the analysis of precision errors.

Running Statistics

Calculation trick using the two definitions for μ and σ:

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i \qquad \sigma^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \mu)^2$$

You can show the following, which is a faster way to keep a running calculation of the variance, and has less digital round-off

$$\sigma^2 = \frac{1}{N-1}\left[\sum_{i=0}^{N-1} x_i^2 - \frac{1}{N}\left(\sum_{i=0}^{N-1} x_i\right)^2\right]$$

or using a simpler notation,

$$\sigma^2 = \frac{1}{N-1}\left[sum\ of\ squares - \frac{sum^2}{N}\right]$$

While moving through the signal, a running tally is kept of three parameters: (1) the number of samples already processed, (2) the sum of these samples, and (3) the sum of the squares of the samples (that is, square the value of each sample and add the result to the accumulated value). After any number of samples have been processed, the mean and standard deviation can be efficiently calculated using only the current value of the three parameters.

The standard deviation of the mean is:

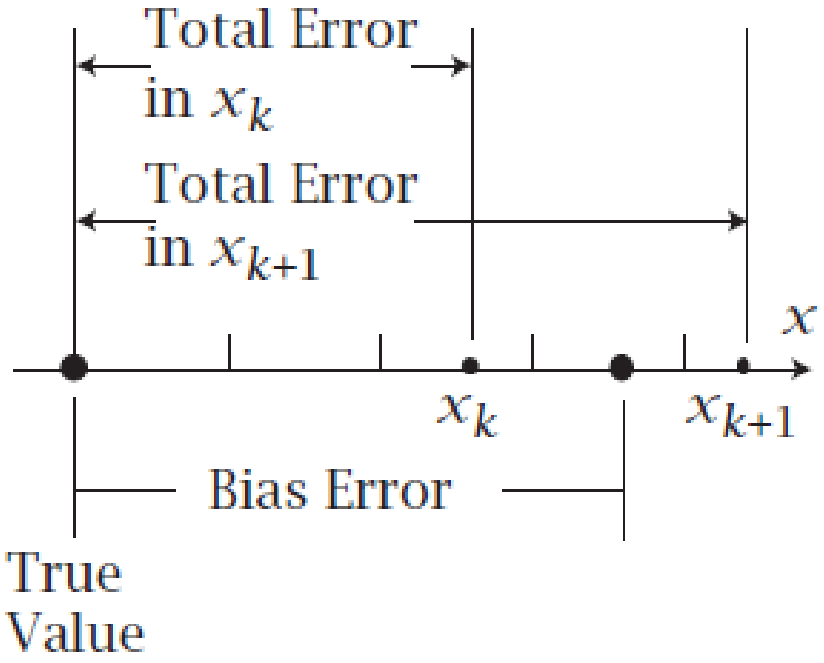
$$S_{\bar{X}} = \frac{S_X}{N^{1/2}}$$

This is NOT the standard deviation of one measurement from the mean of one set of experiments! If the experiment is carried out in many times data sets, and in each set of data many measurements are taken, the standard deviation of the mean values of the sets of data have a much lower standard deviation than the standard deviation of the values of the individual sets. That is, there is always less precision error in a sample mean than in the individual measurements, and if the sample size is large enough the error can be negligible.

Remember this is only for the statistical precision error – NOT the bias error.

A statistical analysis of a sample tells a lot about precision errors, having a sample tells us nothing about *bias errors.*

The total error in a measurement is the difference between the measured value and the true value. BUT we do not know what the true value is! If we take a large enough sample we could say that a good estimate of the bias error is $x - x_{true}$. But the catch is that we do not know x_{true} *a priori:* x_{true} is the unknown we want to determine. Thus, determination of bias errors has nothing to do with samples of data and statistical analysis. To find the bias errors you have to compare with data from similar instruments, or with standard measurements, or patiently find the bias in your instrument.



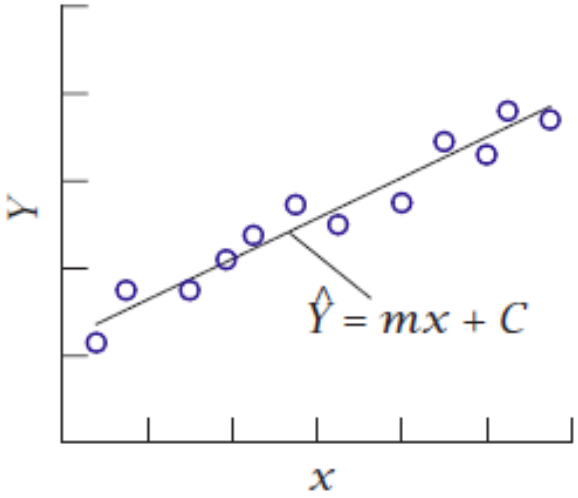
Total and bias errors in a measurement.

How about **least square curve fits** – that is, one parameter depends on another.

Take the example of a **straight line dependence**.

$$y = Mx + C \qquad (x_i, y_i);\ i = 1, 2, \ldots, N$$

assume that y has significant precision error,
but the x precision error is negligible



$$\sum D_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - mx_i - C)^2$$

Sum of squared of differences

How to determine the slope and intercept

$$\sum D_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - mx_i - C)^2$$

$$\frac{\partial \sum D_i^2}{\partial m} = \sum x_i Y_i - C \sum x_i - m \sum x_i^2 = 0$$

$$\frac{\partial \sum D_i^2}{\partial C} = \sum Y_i - NC - m \sum x_i = 0.$$

$$m = \frac{\sum x_i Y_i - N\overline{x}\,\overline{Y}}{\sum x_i^2 - (\sum x_i)^2 / N};$$

$$C = \frac{\overline{Y} \sum x_i^2 - \overline{x} \sum x_i Y_i}{\sum x_i^2 - (\sum x_i)^2 / N} = \overline{Y} - m\overline{x}.$$

$$\overline{x} = (1/N) \sum x_i$$

$$\overline{Y} = (1/N) \sum Y_i$$

Standard error for the curve fit is defined as:

$$S_Y = \left[\frac{1}{N-2} \sum D_i^2 \right]^{1/2}$$

Comments:

- It was assumed that all the variance was in "y". If "x" also has significant variance, the expressions are more complex.

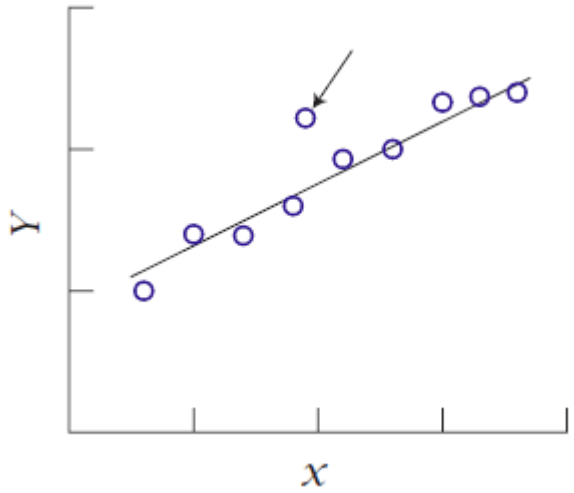
- If the plot is seen to be nonlinear, maybe we can linearize the data: for instance

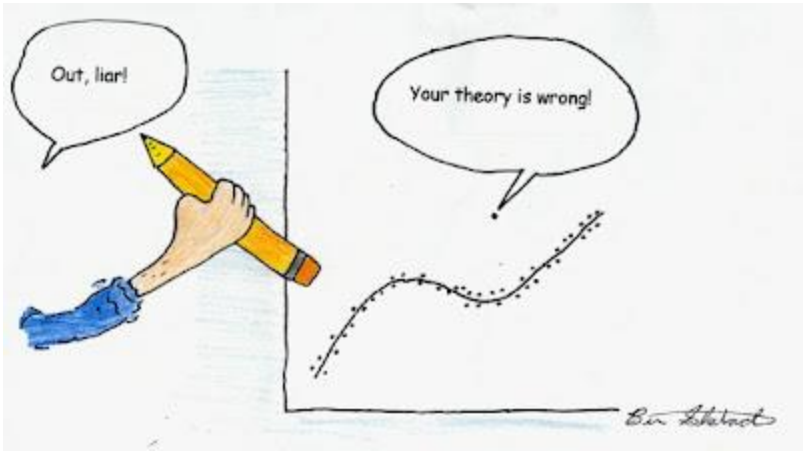
$$\text{If } y = ae^{-kx}, \text{ then } \ln(y) = \ln a - kx; \text{ plot } \ln(y) \text{ vs } x; \text{ slope} = -k, \text{ and intercept} = \ln a.$$

$$\text{If } y = ax^n; \text{ then } \ln y = \ln a + n \ln x \text{ ; plot } \ln y \text{ vs } \ln x$$

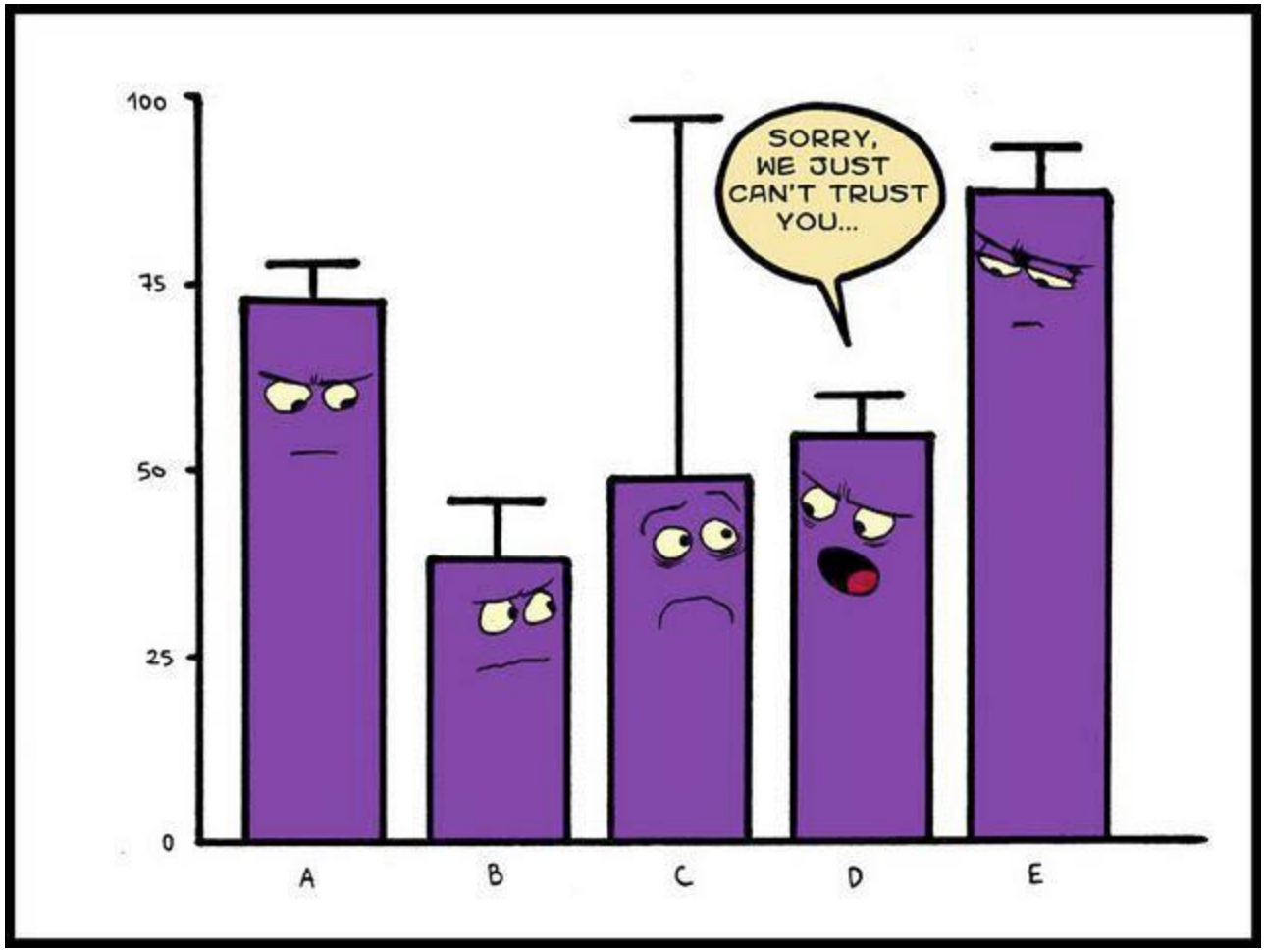
- Often the data points can be fit to **several models**. If you are testing a theory you know the model; or maybe you are searching for a hint for a theory.

- How do you handle **outliers** (see figure below and later)?





Another type of “outlier”



Uncertainty

We do not know the actual value of the parameter(s) we are measuring – we only know an estimate of this value. So we have to deal with estimated – or probable - errors. If we say we are C% confident that the true value X_{true} *of a measurement* X_i *lies within the* interval $X_i \pm P_X$ *: then* P_X *is called the precision uncertainty* at a *confidence level of C%. This means that if we specify a* 95% confidence level estimate of P_X, *we would expect* X_{true} to be in the interval $X_i \pm P_X$ *about 95 times out of a 100.*

We **usually assume a normal distribution** if N>10; then P_X is approximately 2x the standard deviation for 95% confidence:

$$P_X \cong 2S_X \left(C = 95\%, N > 10\right)$$

This is the uncertainty at 95% confidence for individual samples drawn from a normal population and the total sample is large

For small samples this must be amended – so always try to keep N>10.

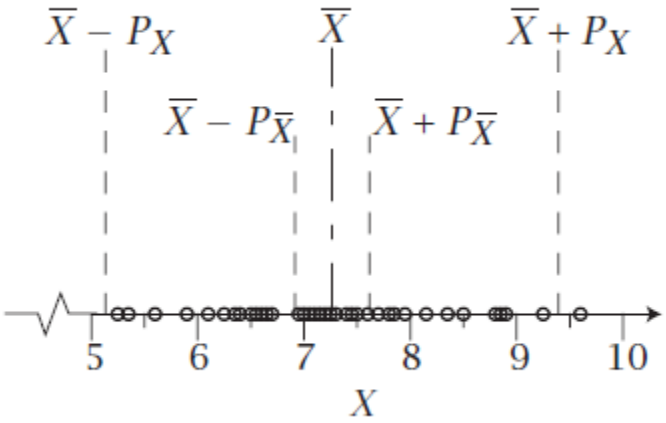
Now what about **the precision in the uncertainty of the value of the mean of repeated sets of measurements**, each set consisting of a certain number of individual measurements?

Remember: $$S_{\bar{x}} \cong \frac{S_X}{N^{1/2}}$$

Then the corresponding **precision uncertainty in the sample mean** is: $$P_{\bar{X}} \cong 2S_{\bar{X}} \left(C = 95\%, N > 10\right)$$

So, The probable error in a sample mean is much less than in the individual measurements. Why is this important?

- We usually average individual measurements over a time interval before recording the averaged values.
- When precision error is important, we usually are interested in the sample mean, not in individual measurements in any particular set of measurements.



Can we know estimates of our error when we only take a single measurement?

Yes, if we have independent data for the variance of the measurement from previous measurements, or from an examination of the instrument from the factory or from control measurements. But in general it is best to take several measurements.

How about the precision error for a curve fit? Then one can show:

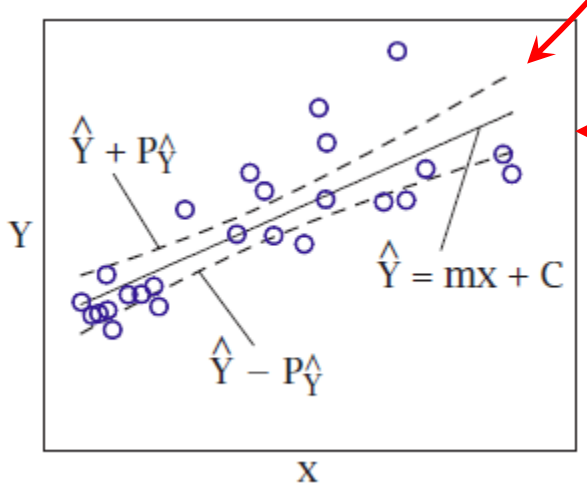
$$S_Y = \left[\frac{1}{N-2}\sum D_i^2\right]^{1/2} \qquad S_{xx} = \sum x_i^2 - \left(\frac{1}{N}\right)\left(\sum x_i\right)^2$$

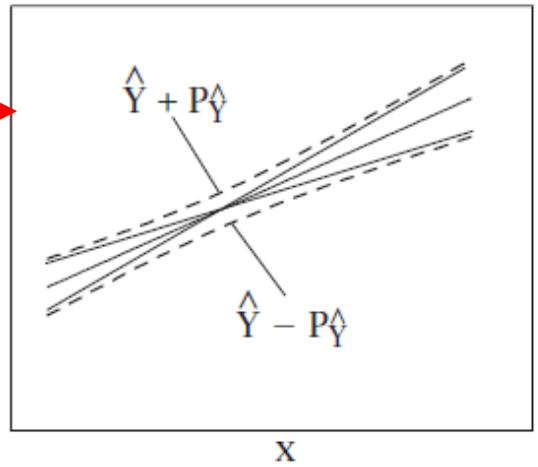
$\hat{Y}$ *for a* curve-fit is like a "mean" value analogous to $\overline{X}$ *for a sample* of values of a single variable.

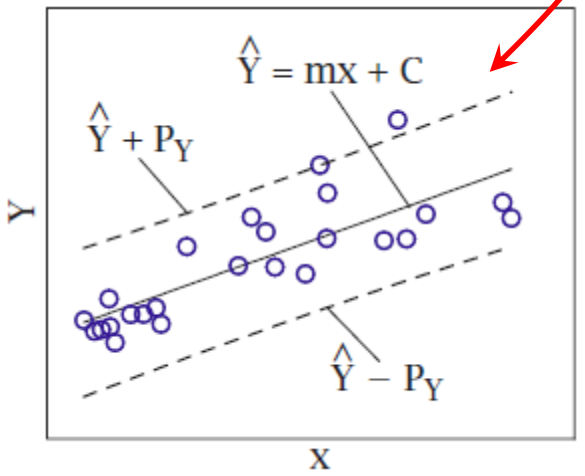
$$P_{\hat{Y}} = 2\left\{S_{\hat{Y}}^2\left[\frac{1}{N} + \frac{(x-\overline{x})^2}{S_{xx}}\right]\right\}^{1/2} \qquad (C = 95\%, N > 10)$$

$$P_Y = 2\left\{S_{\hat{Y}}^2\left[1 + \frac{1}{N} + \frac{(x-\overline{x})^2}{S_{xx}}\right]\right\}^{1/2} \qquad (C = 95\%, N > 10)$$

P_Y is always larger than $P_{\hat{Y}}$, just like P_x is larger than $P_{\overline{x}}$







$P_{\hat{Y}}$ *depends on how far x is away from* $\overline{X}$*: it is a minimum at* $X = \overline{X}$

The range where the curve fits will fall 95% of the time <u>for repeated **sets**</u> of measurements

The range in which we are 95% confident a **single data point** will fall

Bias uncertainty differs from precision uncertainty:

- We are usually concerned with the precision uncertainty of a sample mean or a curve-fit.

- Precision uncertainties can be reduced by increasing the number of data points used.

- Bias uncertainty is independent of sample size: it is the same for one data point as for a sample of 100 data points.

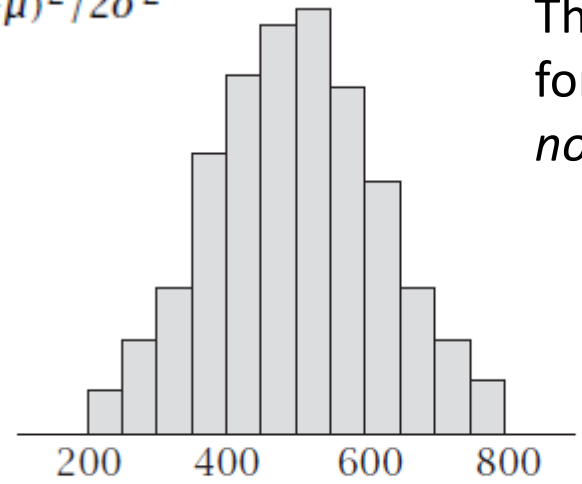
The Normal Probability Distribution

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

$$\int_{-\infty}^{\infty} f(X)\,dX = 1.0$$

$$\mu = \int_{-\infty}^{\infty} X f(X)\,dX$$

$$\sigma^2 = \int_{-\infty}^{\infty} (X-\mu)^2 f(X)\,dX$$



A histogram of a sample from a normal population.

The probability density function for a random variable *X having a normal distribution*

A single measurement that is assumed to be from a normal parent population.

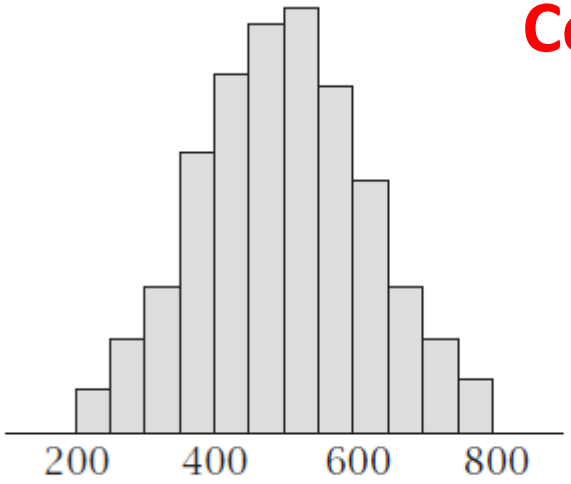
$$P(\mu - \Delta X \le X \le \mu + \Delta X) = \int_{\mu-\Delta X}^{\mu+\Delta X} \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

$$\rightarrow z = \frac{X-\mu}{\sigma} \rightarrow P(-z_1 \le z \le z_1) = \frac{1}{\sqrt{2\pi}} \int_{-z_1}^{z_1} e^{-z^2/2}\,dz$$

$$\frac{1}{\sqrt{2\pi}} \int_{-z_1}^{z_1} e^{-z^2/2}\,dz = \operatorname{erf}\left(\frac{z_1}{\sqrt{2}}\right)$$

$$\operatorname{erf} z_1 = \frac{2}{\sqrt{\pi}} \int_0^{z_1} e^{-z^2}\,dz$$

Confidence levels



A histogram of a sample from a normal population.

E.g. probability that a measurement will fall within 1 standard deviation.

$$P(-1 \le z \le 1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-z^2/2} dz$$

The probability that a measurement will fall within a certain fraction of standard deviations (*σ's) of the mean:*

Probability	Range about mean value
50	$\pm 0.675\sigma$
68.3	$\pm 1.0\sigma$
95	$\pm 1.96\sigma$
99.7	$\pm 3\sigma$
99.99	$\pm 4\sigma$

t-statistics – small number of samples

Remember

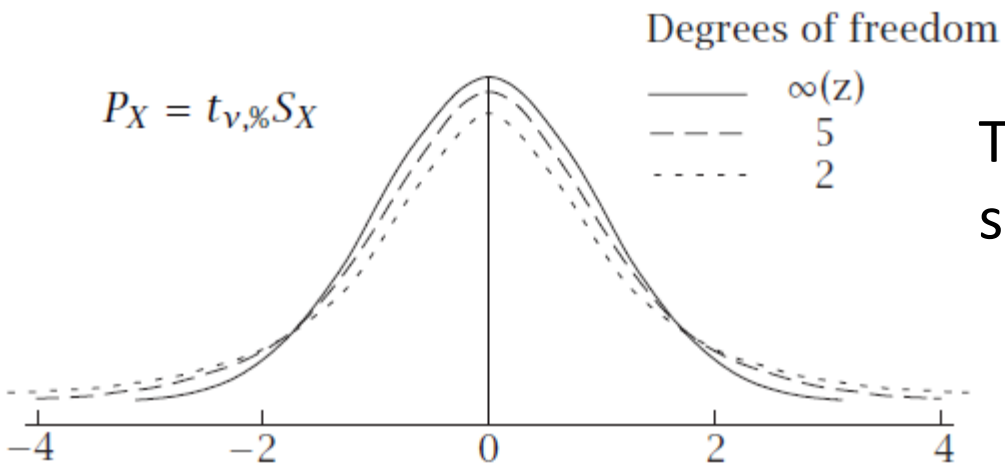
$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad S_x = \left[\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2\right]^{1/2} \qquad \sigma_X = \frac{\sigma}{N^{1/2}}$$

$$N - 1 = \nu = \text{degrees of freedom}$$

The precision uncertainty P_X *of an individual measurement* at a confidence level C% is defined such that we are C% sure that the population mean μ *lies in the interval* $X_i \pm P_X$. *BUT we do not know the population standard* deviation σ.

ν	$t_{95\%}$	ν	$t_{95\%}$
4	2.770	15	2.131
5	2.571	20	2.086
7	2.365	30	2.042
10	2.228	60	2.000

$$\overline{X} = \mu \pm t_{\nu,\%} S_{\overline{X}}(C\%)$$



The smaller the # of samples, the larger is “t”

How well does the sample mean $\bar{X}$ *estimate the population mean* μ*?*

Because the sample means are normally distributed, the *t-distribution can be used:*

$$\bar{X} = \mu \pm t_{\nu,\%} S_{\bar{X}}(C\%) \qquad S_{\bar{X}} = \frac{S_X}{N^{1/2}}$$

That is, one can say with *C% confidence that the population mean* μ *is within* $\pm t_{\nu,\%} S_{\bar{X}}$ of $\bar{X}$.

NOTE: Sample means are normally distributed even when the parent population is not Gaussian.

What do you do with outliers? Find the problem, or if there is no reason found use:

Chauvenet's criterion *is recommended: It states that* points should be discarded if the probability (calculated from the normal distribution) of obtaining their deviation from the mean is less than 1*/2N.*

Ratio of the maximum acceptable deviation to the standard deviation is given as a function of *N.*

N	$\frac{(X_{\max} - \bar{X})}{S_X}$
5	1.65
7	1.80
10	1.96
15	2.13
20	2.24
50	2.57
100	2.81

Standard error of a fit to a straight line

$$\hat{Y} = mx + C$$

$$S_Y = \left[\frac{1}{N-2} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \right]^{1/2}$$

Y_i is a random variable and can be taken to have a normal distribution for each value of x_i.

$$P_{\hat{Y}} = t_{\nu,\%} \left\{ S_{\hat{Y}}^2 \left[\frac{1}{N} + \frac{(x_i - \overline{x})^2}{S_{xx}} \right] \right\}^{1/2}$$

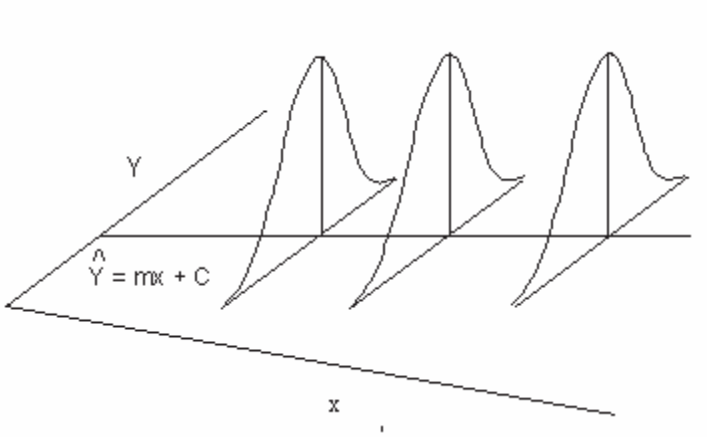
$$S_{xx} = \sum x_i^2 - (1/N) (\sum x_i)^2$$

$$\nu = N - 2$$

For *N large and a 95% confidence level, we set* $t_{\nu,\%} = 2$

$$P_{\hat{Y}} = 2 \left\{ S_{\hat{Y}}^2 \left[\frac{1}{N} + \frac{(x - \overline{x})^2}{S_{xx}} \right] \right\}^{1/2}$$

$$(C = 95\%, N > 10)$$



Standard error of a fit to a straight line

Standard deviation for the **slope**

$$S_m = \left(\frac{S_Y^2}{S_{xx}} \right)^{1/2}$$

Standard deviation for the intercept

$$S_C = \left[S_Y^2 \left(\frac{1}{N} + \frac{\overline{x}^2}{S_{xx}} \right) \right]^{1/2}$$

Precision uncertainty for the **slope**

$$P_m = t_{\nu,\%} S_m$$

Precision uncertainty for the **intercept**

$$P_C = t_{\nu,\%} S_C$$

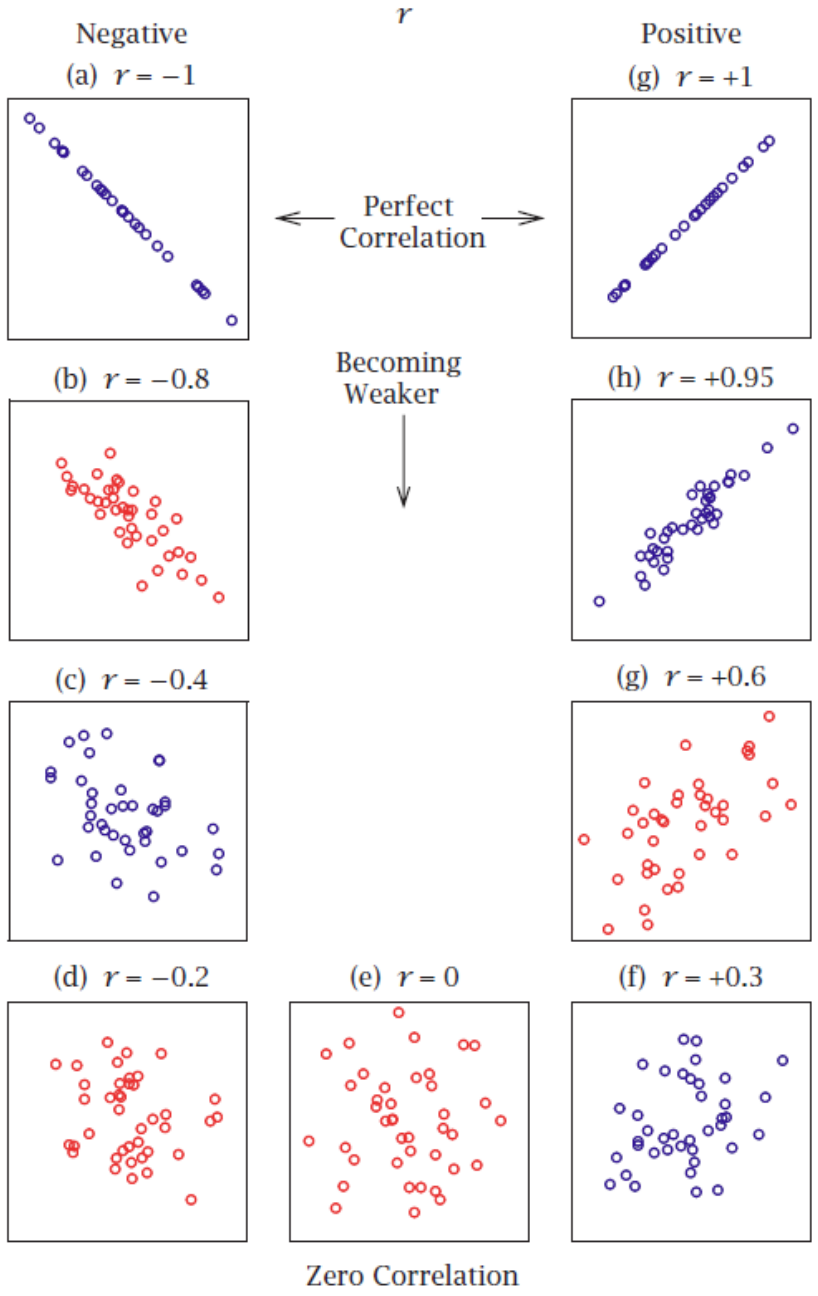
And for *N large and a 95% confidence level,*
we set $t_{\nu,\%} = 2$

The Correlation Coefficient

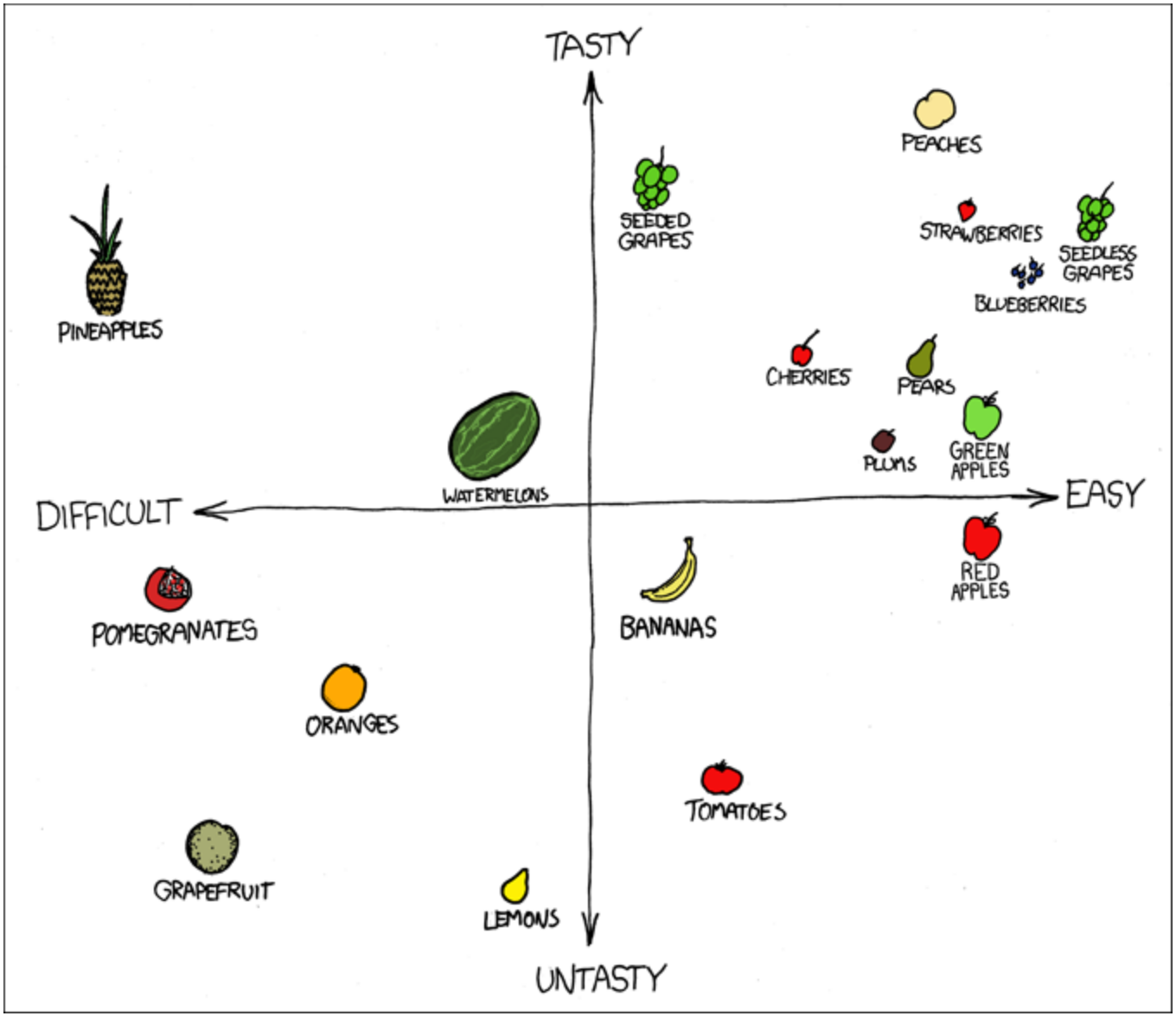
$$r = \frac{1}{N-1}\sum \frac{\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{S_X S_Y}$$

In statistics practice a straight-line curve-fit is considered reliable for *±0.9 ≤ r ≤ ±1 (the sign indicates that Y increases or decreases* with *X).*

The correlation coefficient is useful when precision errors are large, such as in experiments in the life sciences and medicine. Then the central question is whether there is any correlation whatsoever. In physics and engineering experiments the precision errors are usually much smaller and the precision uncertainties of $\hat{Y}$, m, and C are more useful.



But be careful! You can correlate anything, even if ill or subjectively defined.



Autocorrelation shows how similar data is over certain distances

correlation between observations separated by *k time steps*

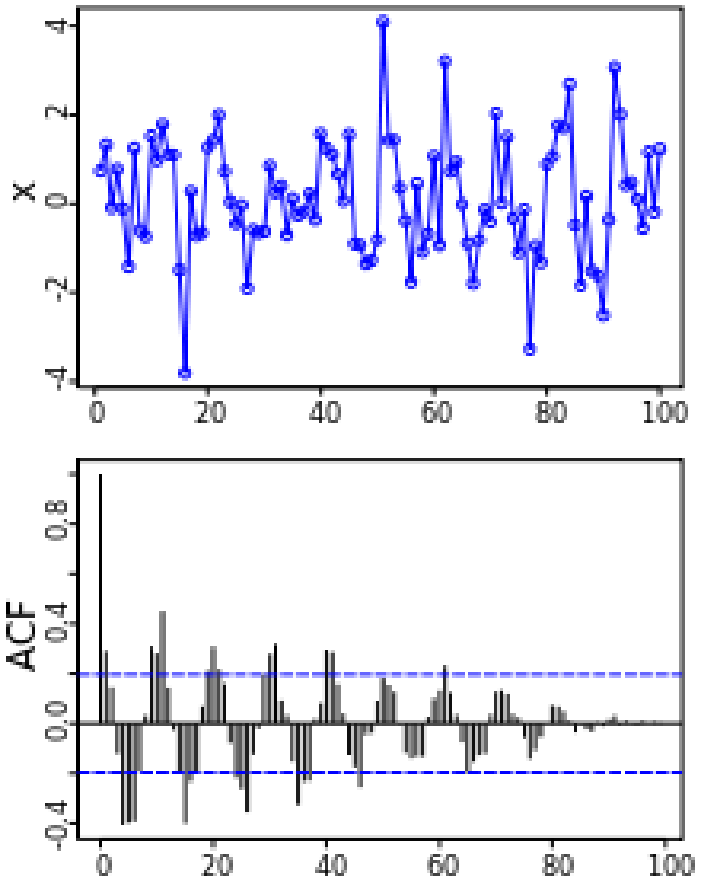
$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Autocovariance

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

$$r_k = c_k / c_0$$

c_0 is the variance



A plot showing 100 random numbers with a "hidden" sine function, and an autocorrelation (correlogram) of the series on the bottom.

http://en.wikipedia.org/wiki/Autocorrelation

Propagation of Precision Uncertainties

Say Y is a function of *N independent measurements* X_i. *If the uncertainties* P_i are small enough we can use a first order Taylor expansion of *Y to write*

$$Y(X_1 + P_1, X_2 + P_2, \ldots, X_N + P_N)$$

$$\cong Y(X_1, X_2, \ldots, X_N) + \frac{\partial Y}{\partial X_1}P_1 + \frac{\partial Y}{\partial X_2}P_2 + \cdots + \frac{\partial Y}{\partial X_N}P_N$$

Since *Y is a linear function of the independent variables,* a theorem of mathematical statistics says:

$$P_Y = \left[\sum_{i=1}^{n}\left(\frac{\partial Y}{\partial X_i}P_i\right)^2\right]^{1/2} \quad \text{or} \quad P_Y = \left[\sum_{i=1}^{N}\Delta Y_i^2\right]^{1/2}$$

All the uncertainties in the X_i *must be at* the same confidence level.

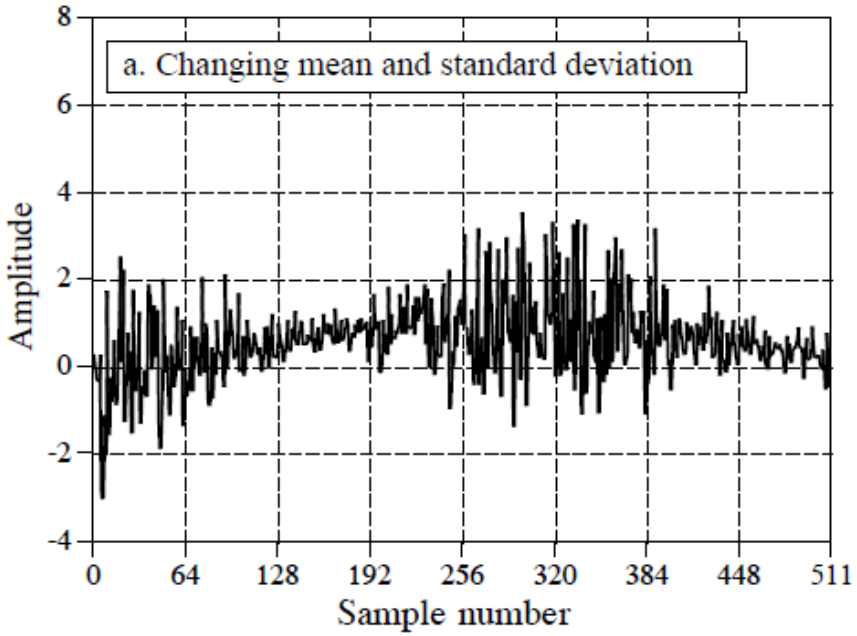
If *Y depends only on a product of the independent measurements* X_i

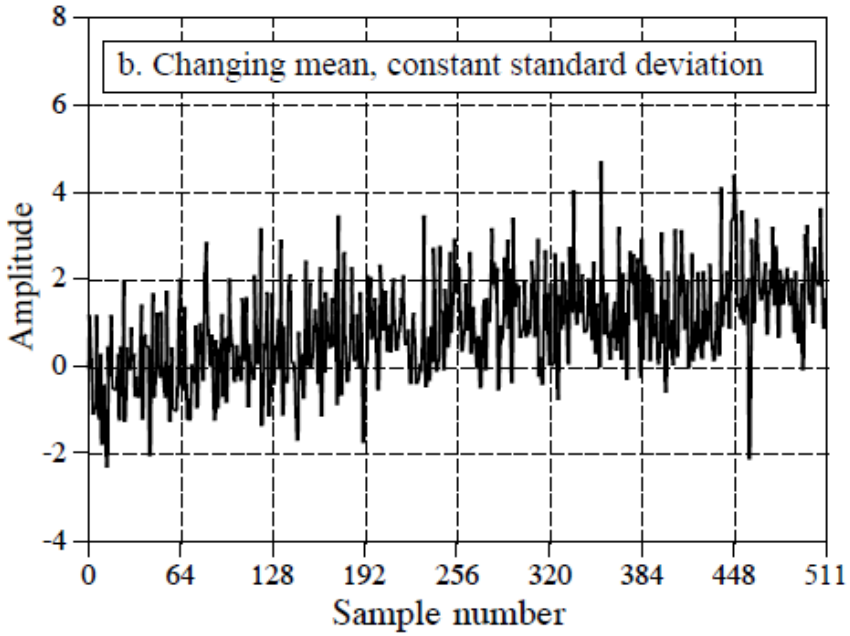
$$Y = CX_1^{m_1}X_2^{m_2}\ldots \quad \text{then} \quad \frac{P_Y}{Y} = \left[\sum_i\left(m_i\frac{P_i}{X_i}\right)^2\right]^{1/2}$$

What about:

- **weighting,**
- **Precision and accuracy**
- **Histograms**
- **Poisson statistics**
- **Non-linear fitting**
- **Chi-square analysis**

Weighting





Examples of signals generated from non-stationary processes. In (a), both the mean and standard deviation change. In (b), the standard deviation remains a constant value of one, while the mean changes from a value of zero to two. It is a common analysis technique to break these signals into short segments, and calculate the statistics of each segment individually.

Least Square fitting of a straight line

minimize

$$S = \sum_i w_i \left(Y_i - y_i\right)^2 = \sum_i \frac{1}{\sigma_i^2}\left(Y_i - y_i\right)^2$$

If the variance varies, you want to minimize chi-square

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - y_{fit})^2}{\sigma_i^2}$$

Goodness of fit parameter that should be unity for a "fit within error"

$$\chi^2_{reduced} = \frac{1}{\nu} \sum_{i=1}^{n} \frac{(y_i - y_{fit})^2}{\sigma_i^2}$$

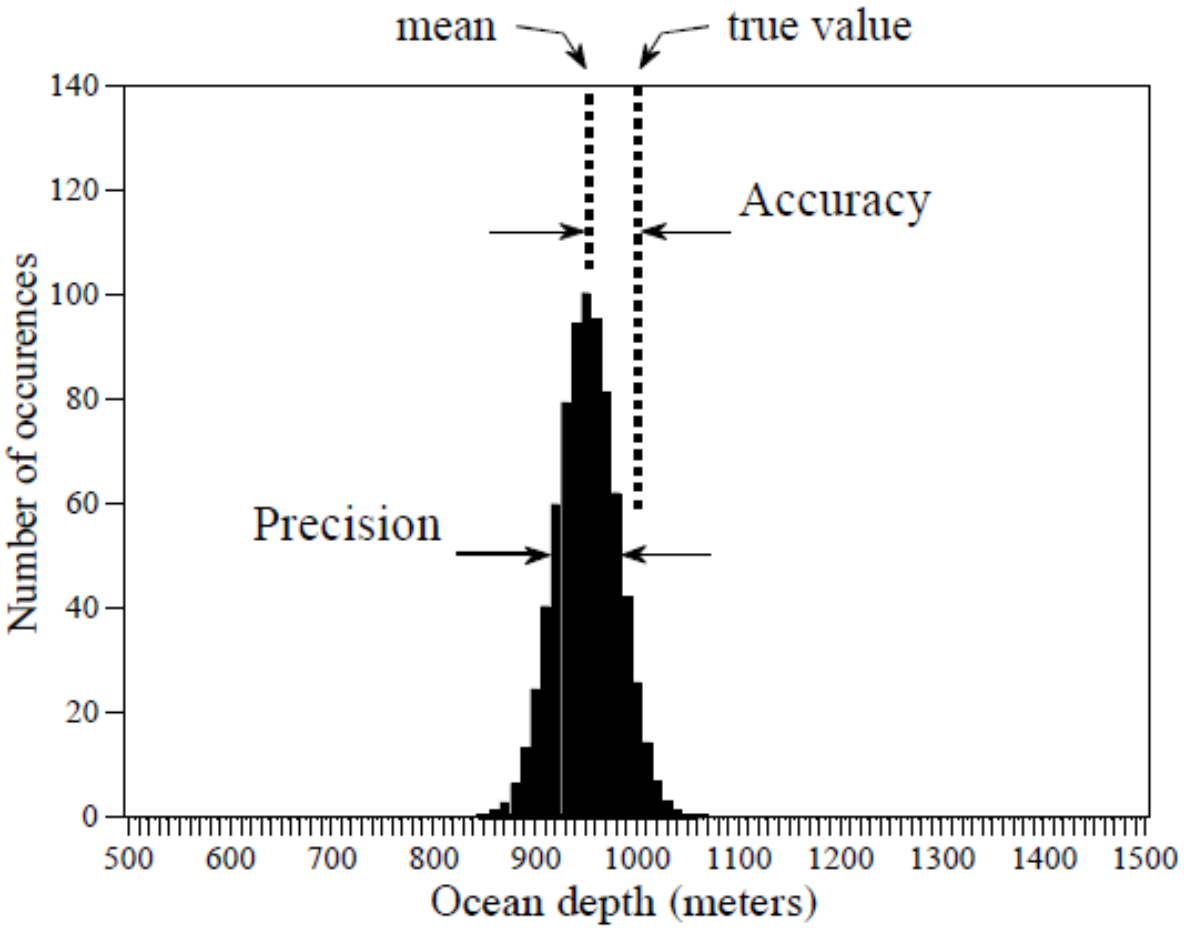
ν is the # of degrees of freedom

$\nu \cong$ n-# of parameters fitted

χ^2 caveats

- Chi-square lower than unity is meaningless…if you trust your s^2 estimates in the first place.
- Fitting too many parameters will lower c^2 but this may be just doing a better and better job of fitting the noise!
- A fit should go smoothly THROUGH the noise, not follow it!
- There is such a thing as enforcing a "parsimonious" fit by minimizing a quantity a bit more complicated than c^2. This is done when you have *a-priori* information that the fitted line must be "smooth".

Graphical description of precision and accuracy



Poor **accuracy** results from **systematic errors.**

Precision is a measure of random noise. Averaging several measurements will always improve the precision.

Poisson distribution:

λ= mean value,

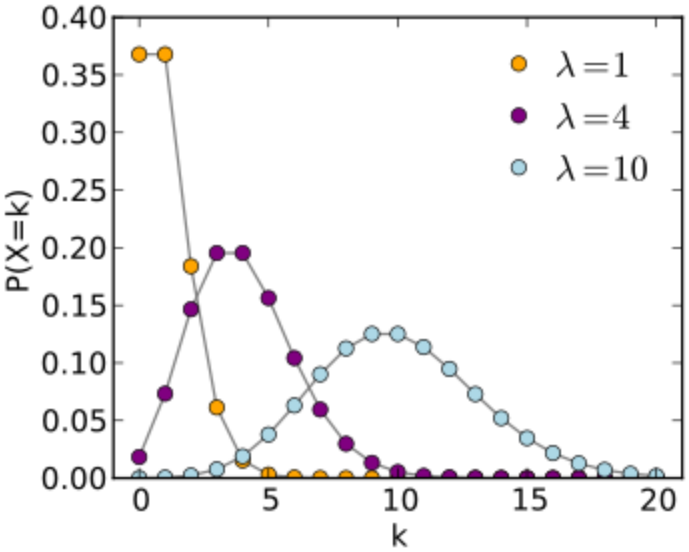
k = # of times observed

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!},$$

Probability of observing k occurrences in tine t, where λ is the average rate per time

$$\Pr(N_t = k) = f(k;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

The variance is equal to the mean



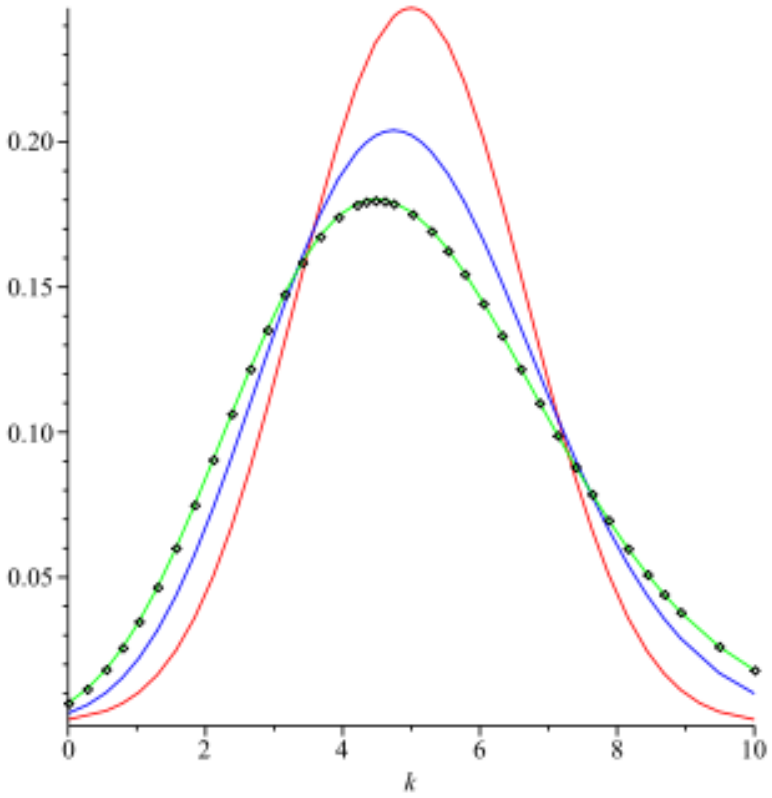
The classic Poisson example is the data set of von Bortkiewicz (1898), for the chance of a Prussian cavalryman being killed by the kick of a horse.

See:

http://www.umass.edu/wsp/statistics/lessons/poisson/index.html

http://mathworld.wolfram.com/PoissonDistribution.html

Comparison of the Poisson distribution (black dots) and the binomial distribution with n=10 (red line), n=20 (blue line), n=1000 (green line). All distributions have a mean of 5. The horizontal axis shows the number of events *k*.



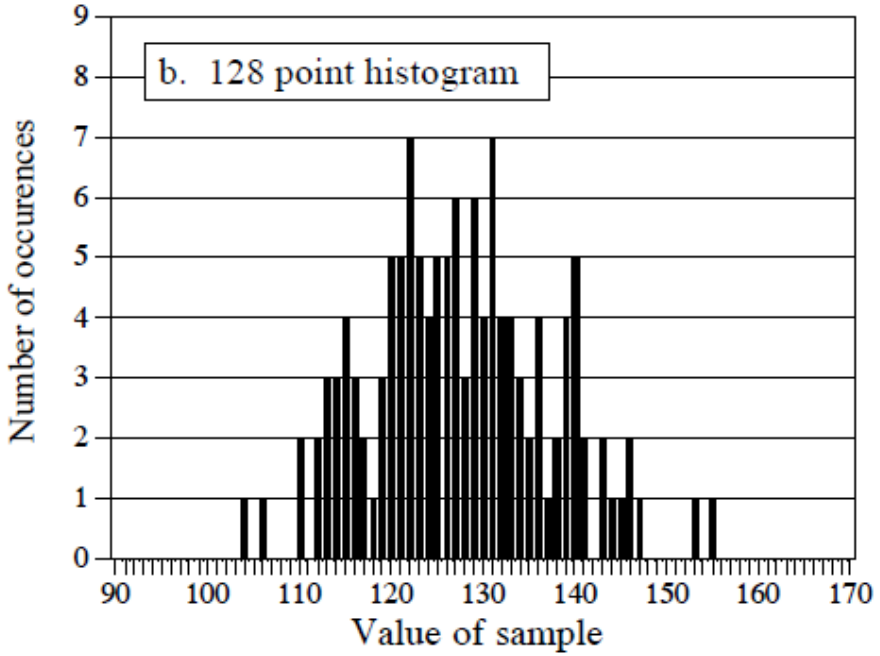
http://en.wikipedia.org/wiki/Poisson_distribution

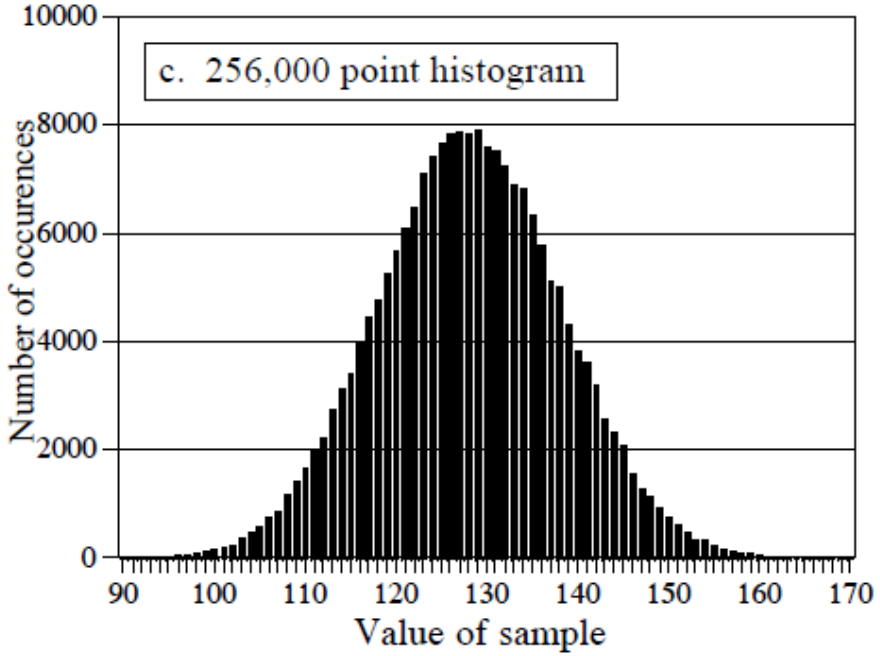
Histograms

$$N = \sum_{i=0}^{M-1} H_i$$

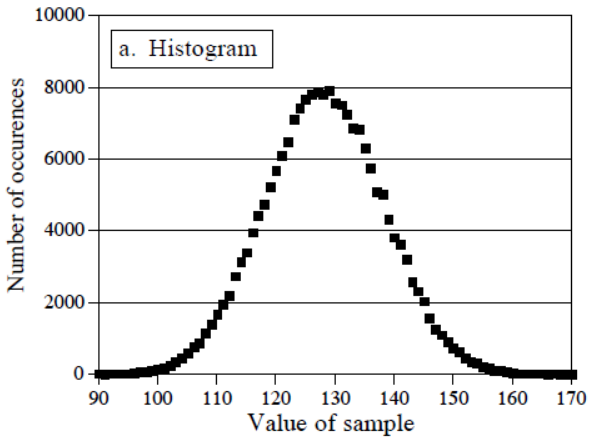
$$\mu = \frac{1}{N} \sum_{i=0}^{M-1} i\, H_i$$

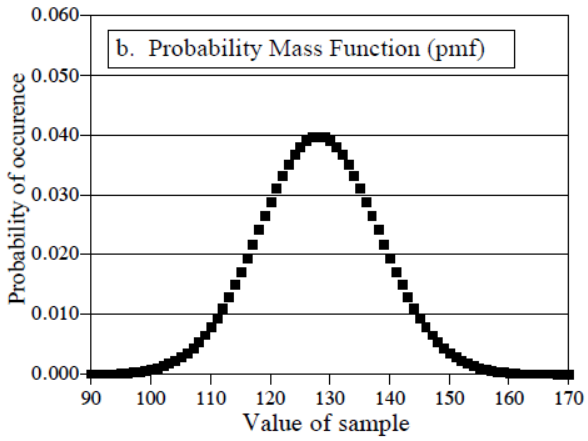
$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{M-1} (i - \mu)^2 H_i$$

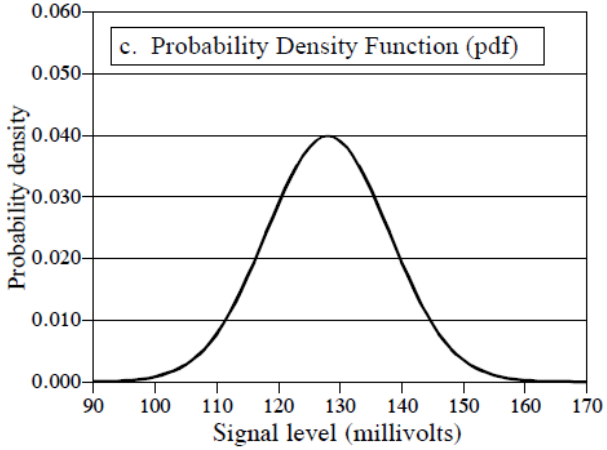




(a) the histogram, (b) the probability mass function (pmf) and (c) the probability density function (pdf)







The amplitude of these three curves is determined by: (a) the sum of the values in the histogram being equal to the number of samples in the signal; (b) the sum of the values in the pmf being equal to one, and (c) the area under the pdf curve being equal to one.

Examples of
probability density functions.