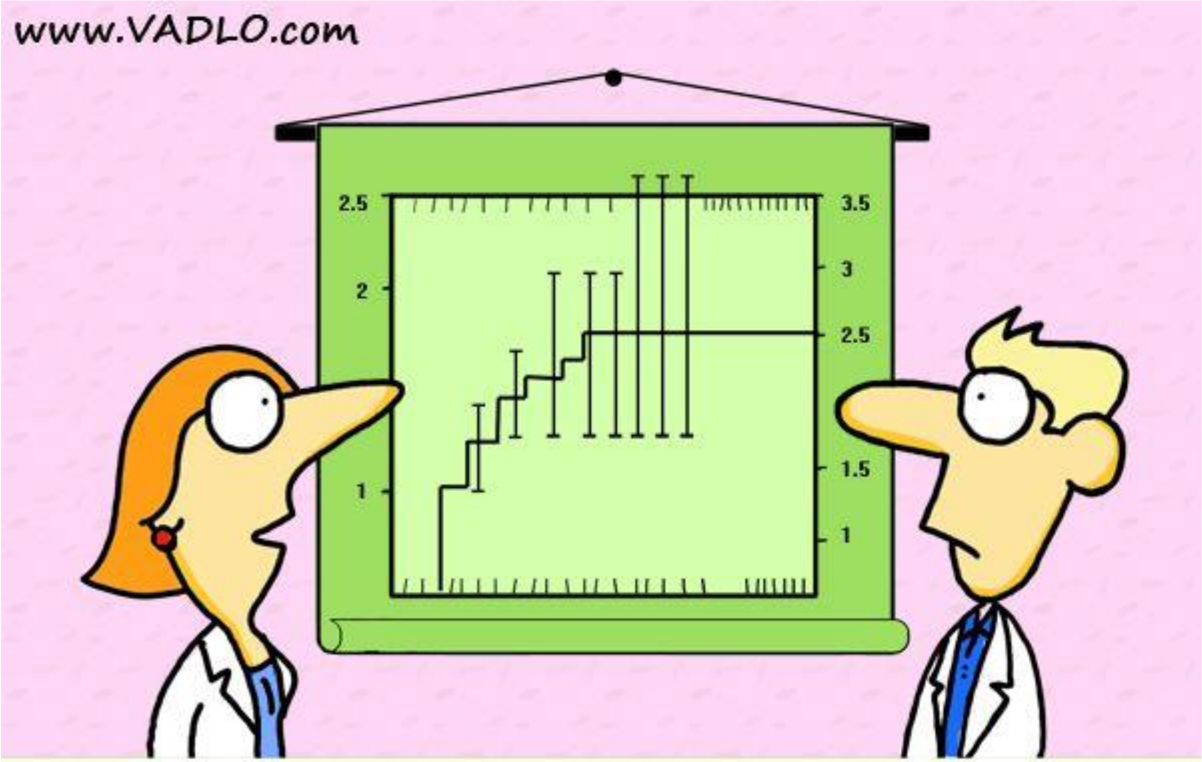
Error Analysis



"Did you really have to show the error bars?"

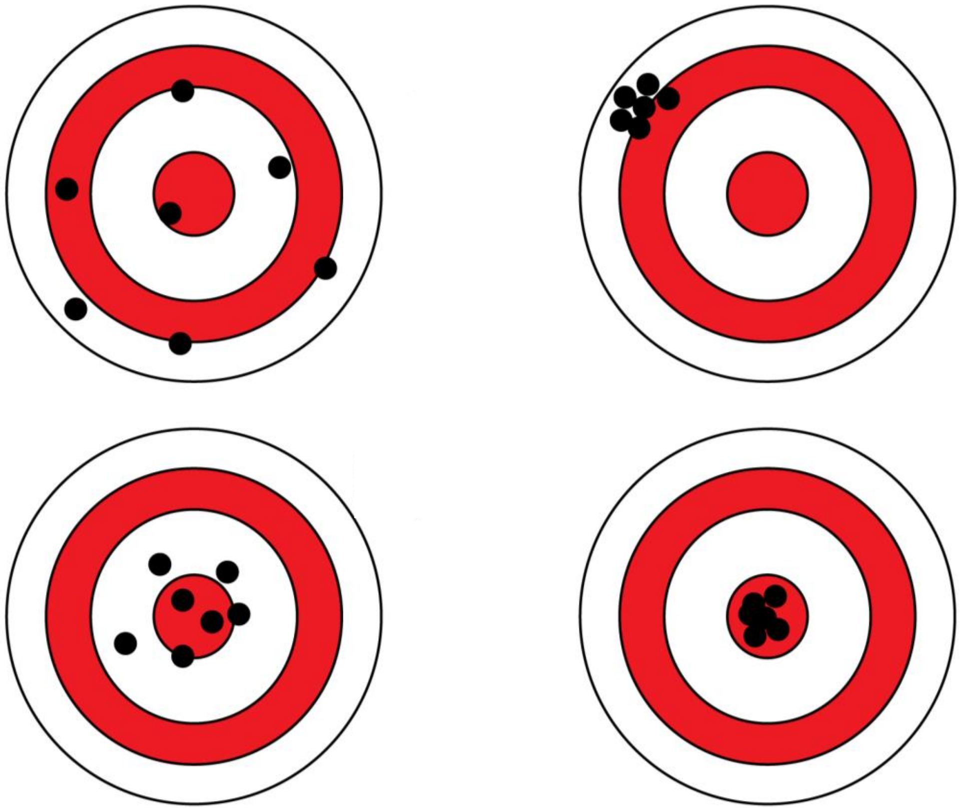
V. Lorenz, E. Colla, L. Yang, M. Perdekamp, D. Hertzog, R. Clegg

PHYS403

Fall 2018

Accuracy versus Precision

- **Accuracy**: how close the result of the experiment comes to the true value
  - A measure of the "correctness" of the result
- **Precision**: how exactly the result is determined (without reference to what the result means)
  - Absolute precision – same units as value
  - Relative precision – fractional units of value



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

Types of Uncertainties

- Systematic
  - **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.
- Statistical
  - Uncertainties due to **stochastic fluctuations**
  - Generally there is **no correlation** between successive measurements.
  - **Multiple measurements** can be used to reduce the uncertainty.

Condensed Matter Experiments (Ferroelectrics, Second Sound, Superconductivity, NMR, AFM)

- Sources of Uncertainty
  - Sample preparation
  - Sample/detector degradation
  - Temperature calibration
  - Voltage calibration
  - Pressure calibration/regulation

Atomic, Molecular, Optical Experiments (Optical Pumping, Quantum Eraser, Spectroscopy)

- Sources of Uncertainty
  - Optical alignment
  - Atomic density calibration
  - Magnetic field calibration
  - Optical waveplate / polarizer calibration
  - Cleanliness of samples, optics
  - Single-photon counting statistics

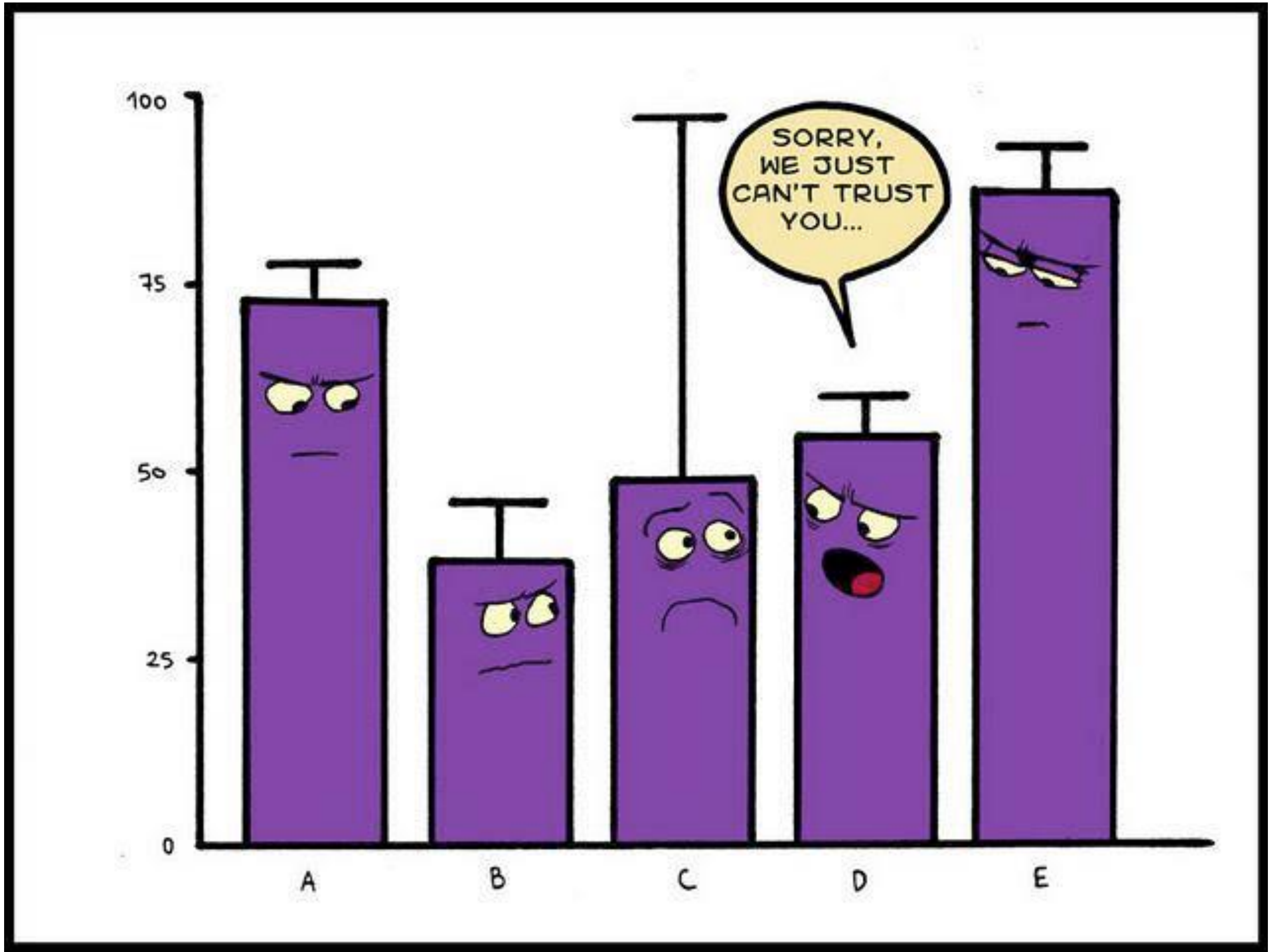
Nuclear Experiments (Alpha Range, Gamma-Gamma, Muon, Moessbauer)

- Sources of Uncertainty
  - Spatial positioning of equipment
  - Voltage drift
  - Heating
  - Energy calibration
  - Pressure calibration/regulation
  - Particle counting statistics

Ways to check for error/uncertainty

- Reproducibility: would the same results be produced for
  - a 2^{nd}, 3^{rd}, experiment?
  - an "identical" sample? If not, check preparation/degradation
  - slightly different parameters, such as speed of change in T, direction of change? If not, check for time lags, response of equipment.
  - different quality of experimental components such as connectors, paint, wires?
- Statistical analysis, error propagation
- Fitting to theory

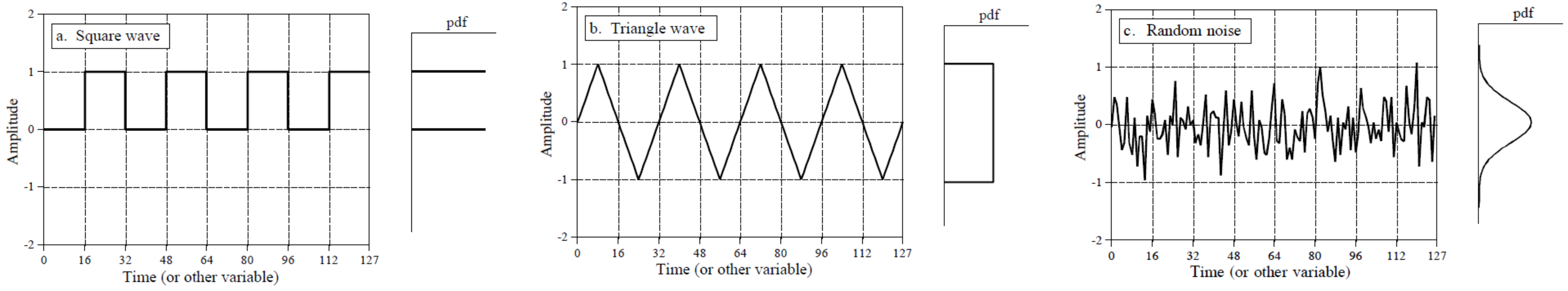
Statistical Analysis & Fitting, or, “How to get error bars”



Statistical Analysis

- Probability density functions: what is the probability of getting a particular value?

- Examples:

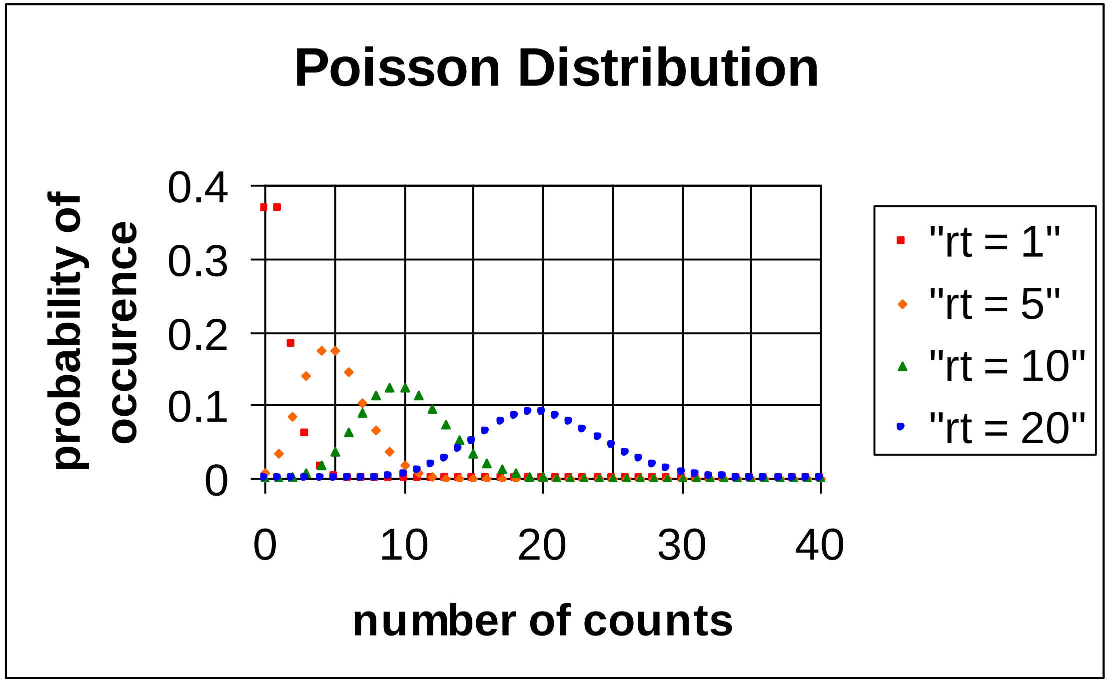


Counting Experiments: Poisson Distribution

- Probability to have *n* decays in time interval *t*:

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

r: decay rate [counts/s]



Is nuclear decay a random process?
Yes, follows Poisson Distribution!
(Rutherford and Geiger, 1910)

A statistical process is described through a Poisson Distribution if:

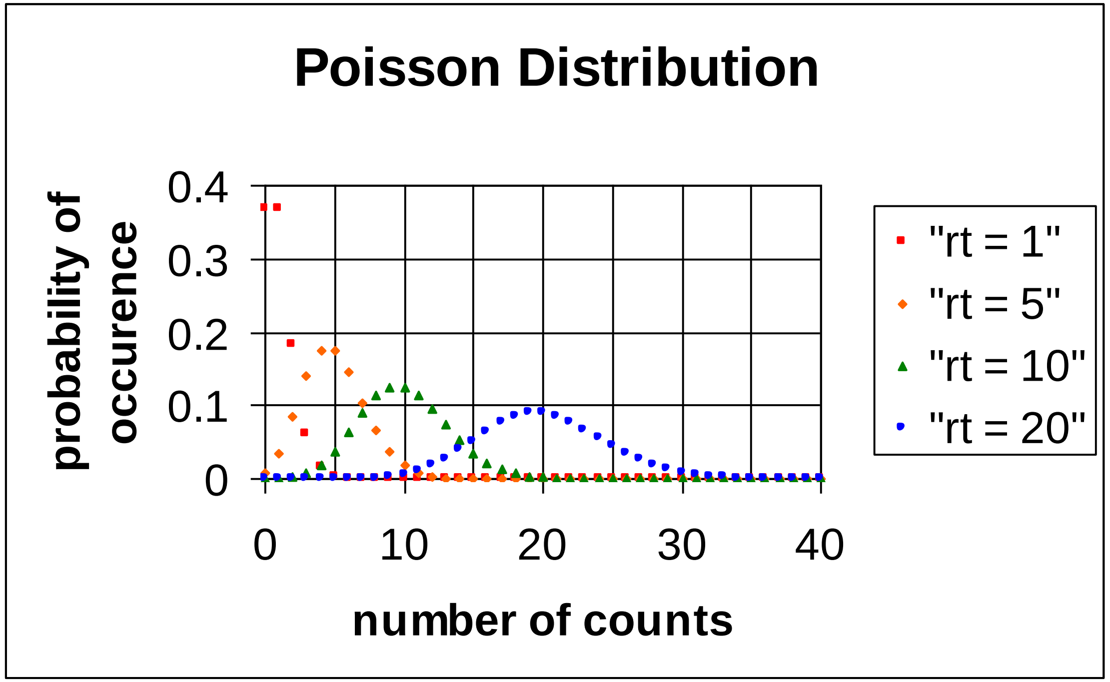
- **random process** → for a given nucleus probability for a decay to occur is the same in each time interval
- **universal probability** → the probability to decay in a given time interval is same for all nuclei
- **no correlation between two instances** → the decay of one nucleus does not change the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

Counting Experiments: Poisson Distribution

- Probability to have *n* decays in time interval *t*:

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

r: decay rate [counts/s]



Some properties:

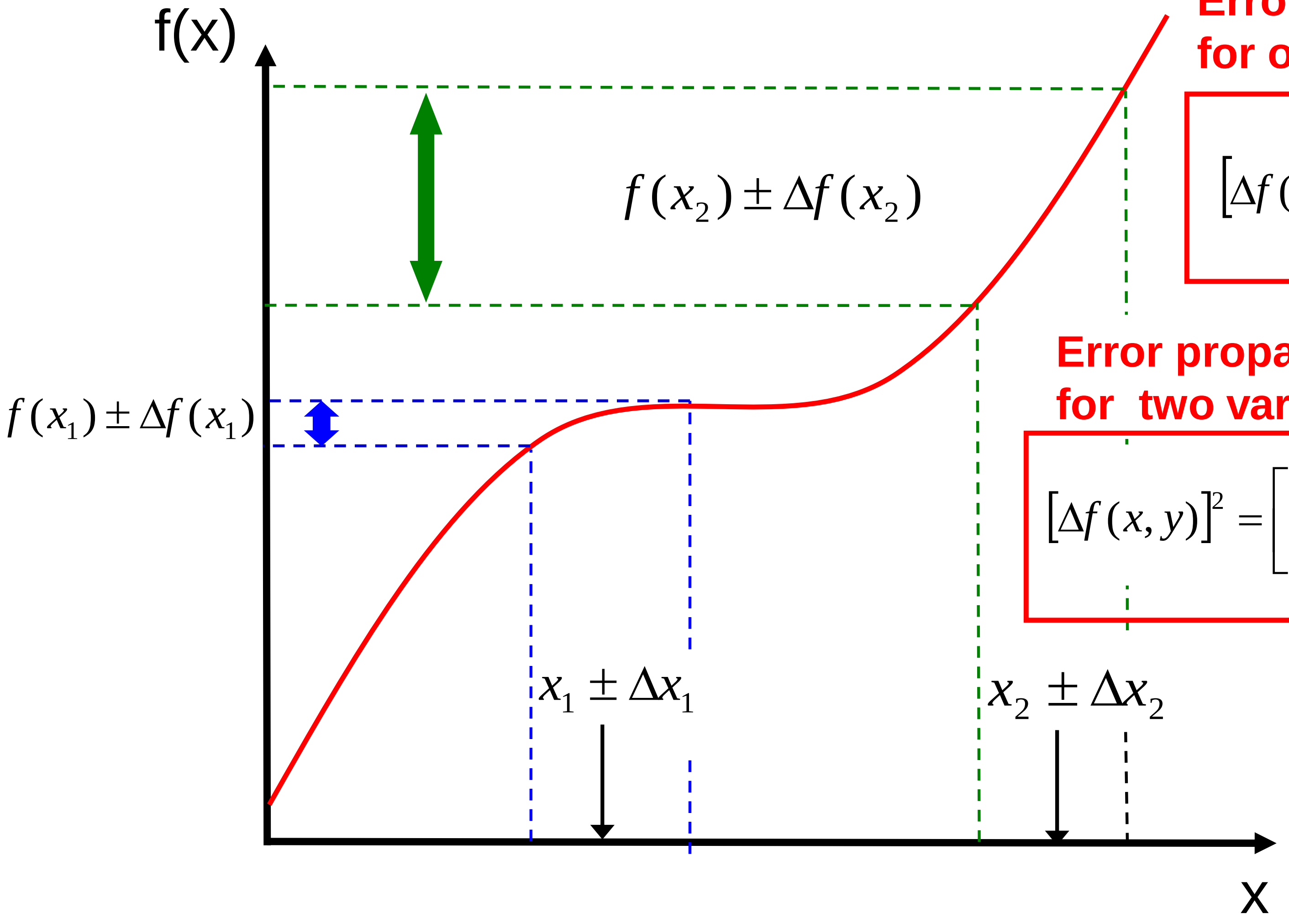
$$\sum_{n=0}^{\infty} P_n(rt) = 1 \ , \ \text{probabilities sum to 1}$$

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

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

standard deviation

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:

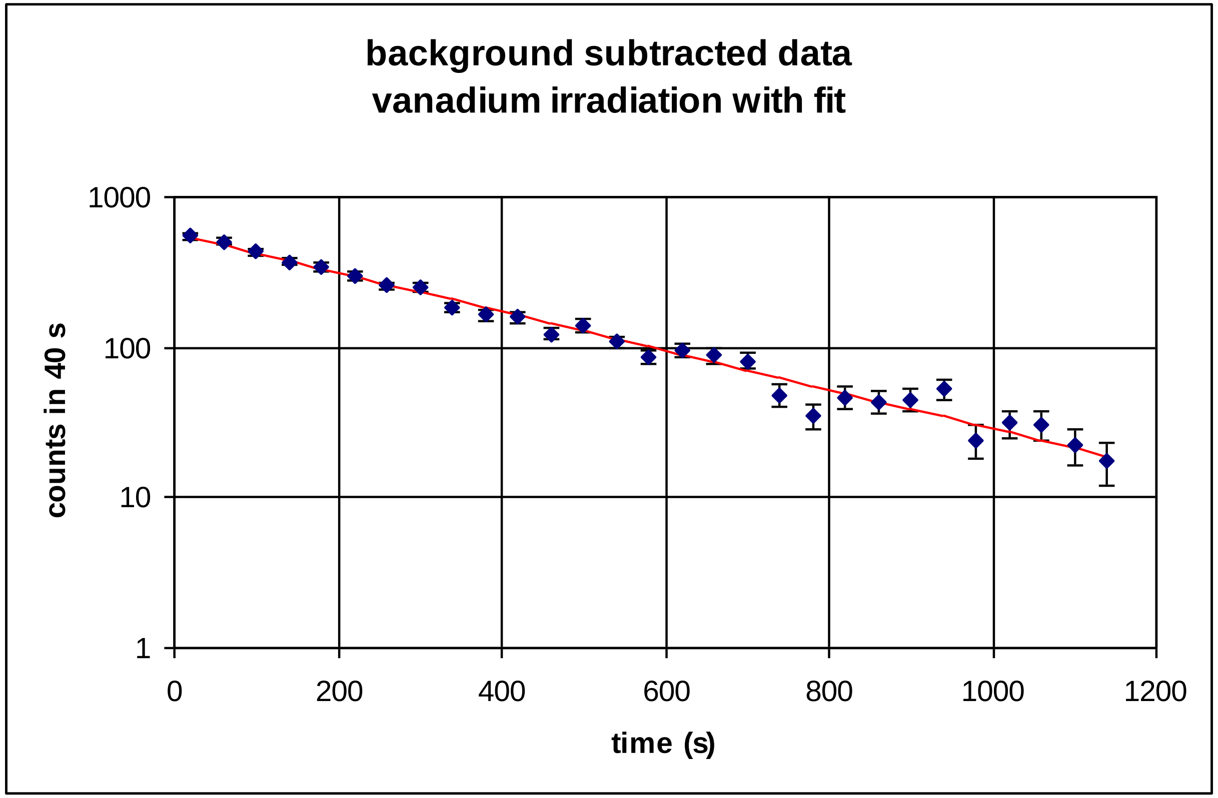
$$[\Delta f(x,y)]^2 = \left[\frac{\partial f}{\partial x}\right]^2 \cdot \Delta x^2 + \left[\frac{\partial f}{\partial y}\right]^2 \cdot \Delta y^2$$

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, ΔS’ = √S’
- Measured background rate: B, ΔB = √B
- Signal: S = S’ – B
- Error:

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

$$\frac{\partial S}{\partial B} = -1, \quad \frac{\partial S}{\partial S'} = 1, \quad \Delta \mathrm{B} = \sqrt{B}, \ \Delta \mathrm{S}' = \sqrt{S'}$$

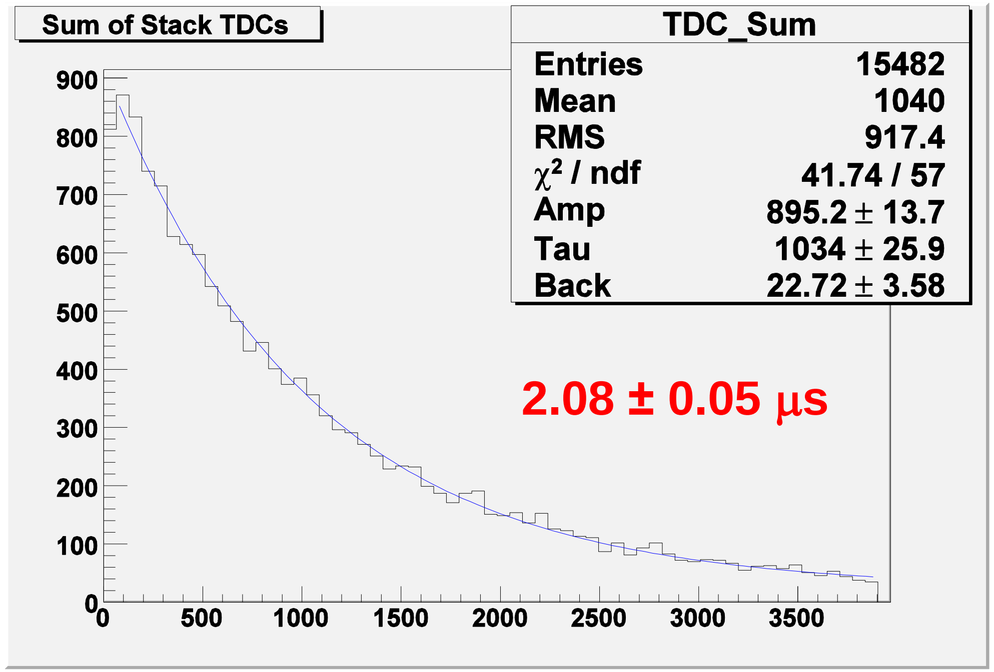
$$\Rightarrow (\Delta \mathrm{S})^2 = \left(\frac{\partial S}{\partial B}\right)^2 \Delta \mathrm{B}^2 + \left(\frac{\partial S}{\partial S'}\right)^2 \Delta \mathrm{S}'^2$$

$$= \quad 1 \quad \cdot \ B \quad + \ 1 \quad \cdot \ S'$$

$$\Rightarrow \ \Delta \mathrm{S} \quad = \sqrt{B + S'}$$

Fitting

Chi-square: $\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$, where x = measured value, μ = mean, σ^2 = variance



Goodness-of-fit test for **Poisson** distribution:

$$\chi^2/\mathrm{ndf} = 1 \pm \sqrt{\frac{2}{\mathrm{ndf}}}$$

(# points - # constraints)

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

Too low means errors are underestimated

Too high means fit is bad

Note: Muon lifetime will usually be lower compared to real one of 2.2 μs due to negative muon capture.

WARNING: DO NOT try to determine goodness of fit from Origin’s Chi-square: not normalized!

Instead, use the r-square value: the closer to 1, the better the fit

Reporting measurement results

- Include uncertainty **with correct number of significant figures**

Too many digits for given error

For errors starting with 1, can keep 1 additional digit

Fit or measurement result	**Uncertainty**	**Wrong**	**Right**
V = 0.122349 m/s	σ = 0.01298 m/s	0.122389 +/- 0.01298 m/s	0.123 +/- 0.013 m/s
T = 3.745 x 10^{-3} s	σ = 0.0798 ms	3.745 x 10^{-3} s +/- 0.0798 ms	(3.75 ± 0.08) x 10^{-3} s

Too many digits for given error

Units don't match

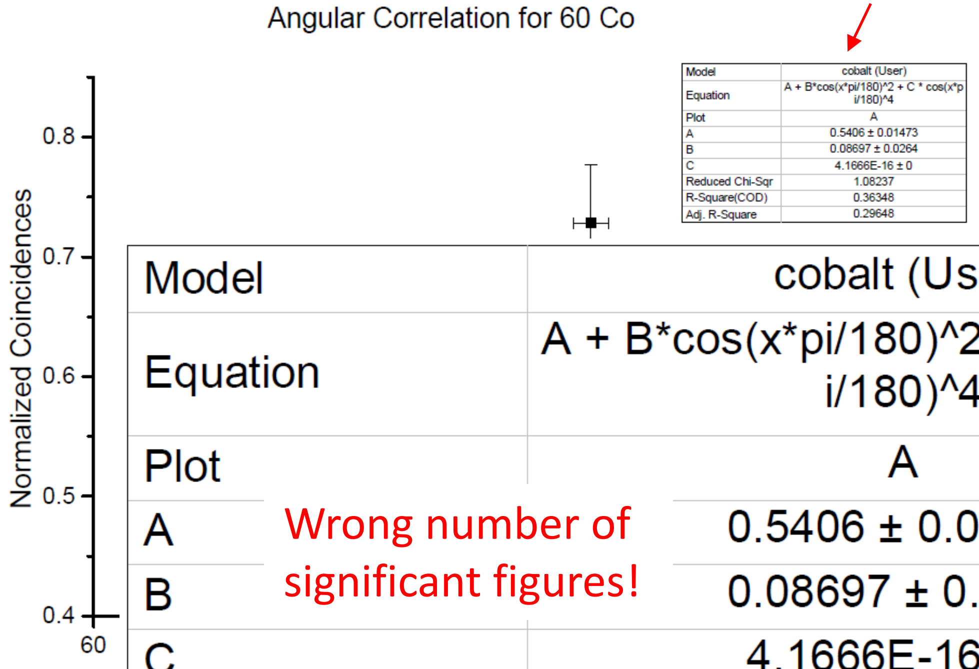
Use scientific notation

Examples:	5.24 x 10^{-2} +/- 0.01 T	1.2687 +/- 0.0019 g	56.8889 +/- 2.5487 s
	0.05 +/- 0.01 T	1.2687 +/- 0.0019 g	57 +/- 3 s

Reporting measurement results

Bad

Too small to read! Quote in text, not in plot.

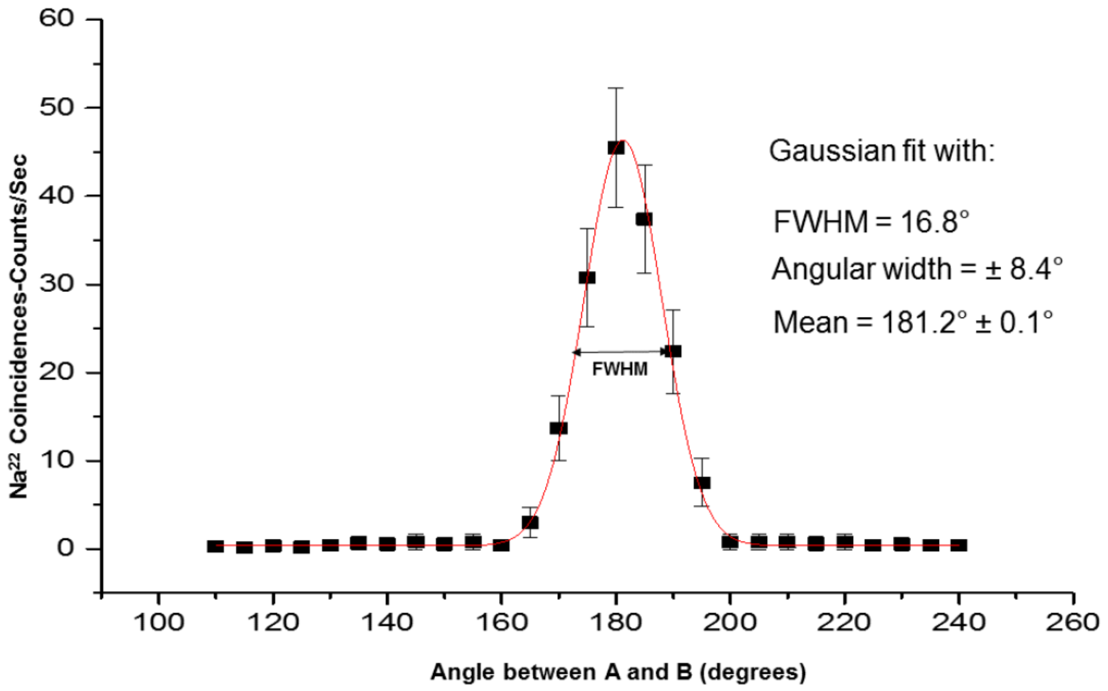


Model	cobalt (User)
Equation	A + B*cos(x*pi/180)^2 + C * cos(x*pi/180)^4
Plot	A
A	0.5406 ± 0.01473
B	0.08697 ± 0.0264
C	4.1666E-16 ± 0
Reduced Chi-Sqr	1.08237
R-Square(COD)	0.36348
Adj. R-Square	0.29648

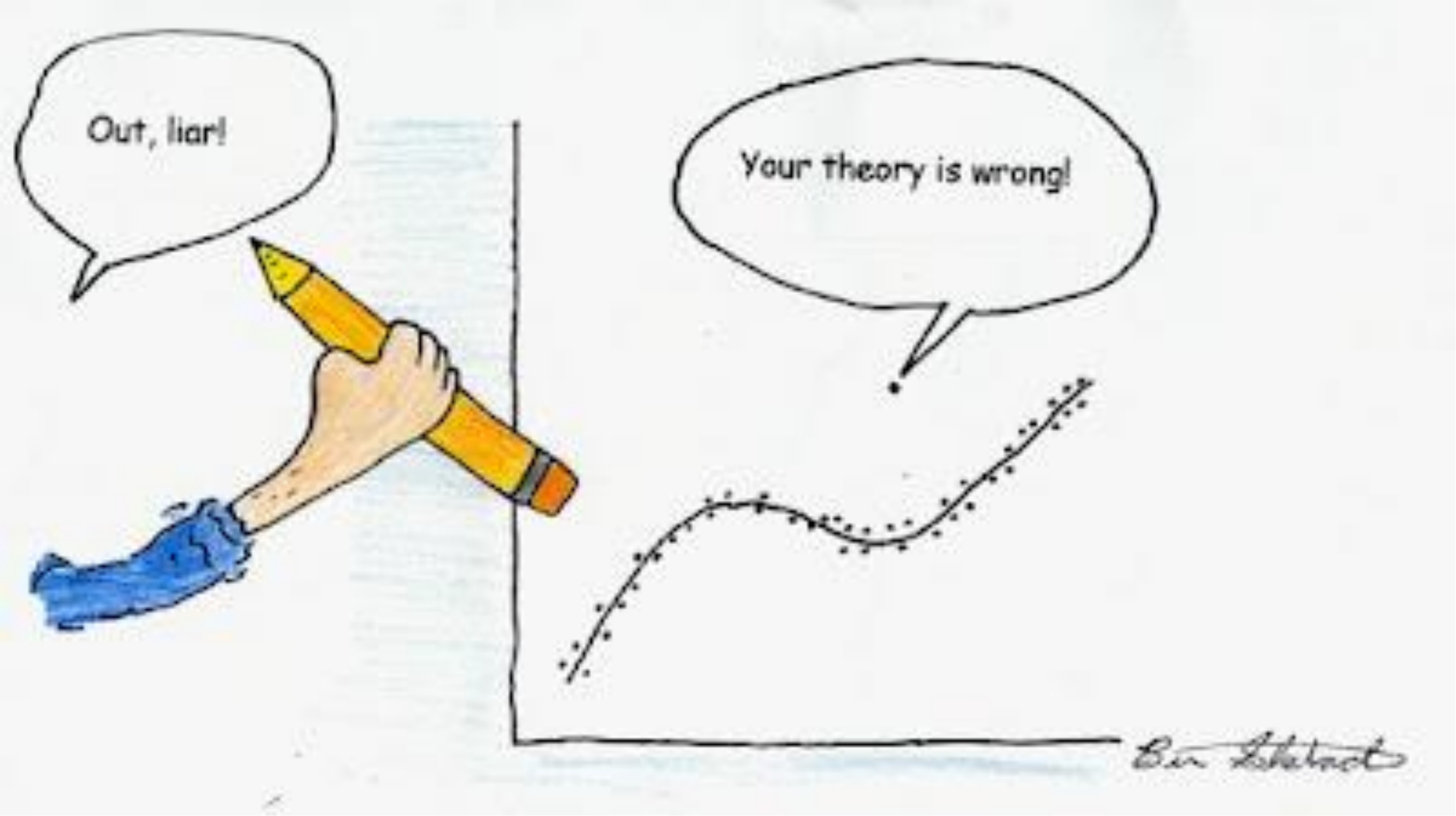
Wrong number of significant figures!

OK

Better to quote in text, not in plot.
Has right sig figs.



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?
- 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?

Possible answers:

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 set up a judgement that is unbiased?

First, compute some simple statistics:

Mean of measurements: $\bar{x} = 3.8$

Standard deviation: $\sigma_x = \sqrt{\sum (x_i - \bar{x})^2 / N} = 0.8$

Is the measurement anomalous? It differs by 2σ (1.6) from the mean.

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 6*0.05= 0.3 measurements should fall outside the 2σ bound.

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

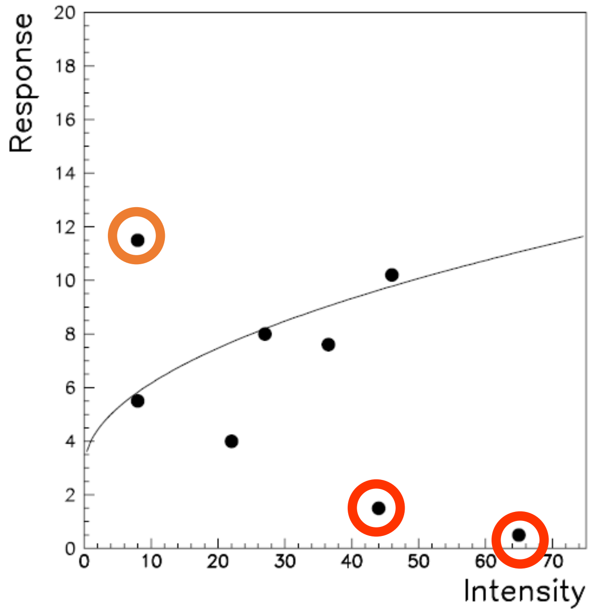
Chauvenet's criterion is the following: If the suspect measurement has a lower probability than 1/2, you should reject it. Our measurement has 0.3 so it goes.

A case study

Data set:
11.5, 5.5, 4.0, 8.0, 7.6, 1.5, 10.2, 0.5

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



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

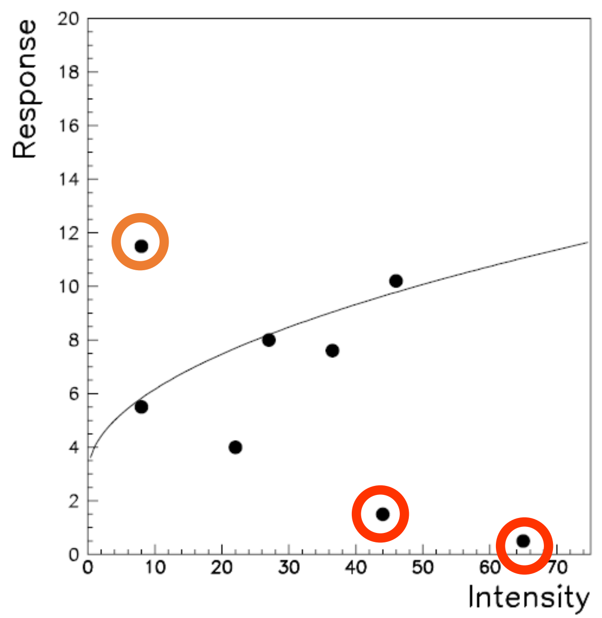
A case study

Data set:
11.5, 5.5, 4.0, 8.0, 7.6, 1.5, 10.2, 0.5

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

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



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 …

A case study

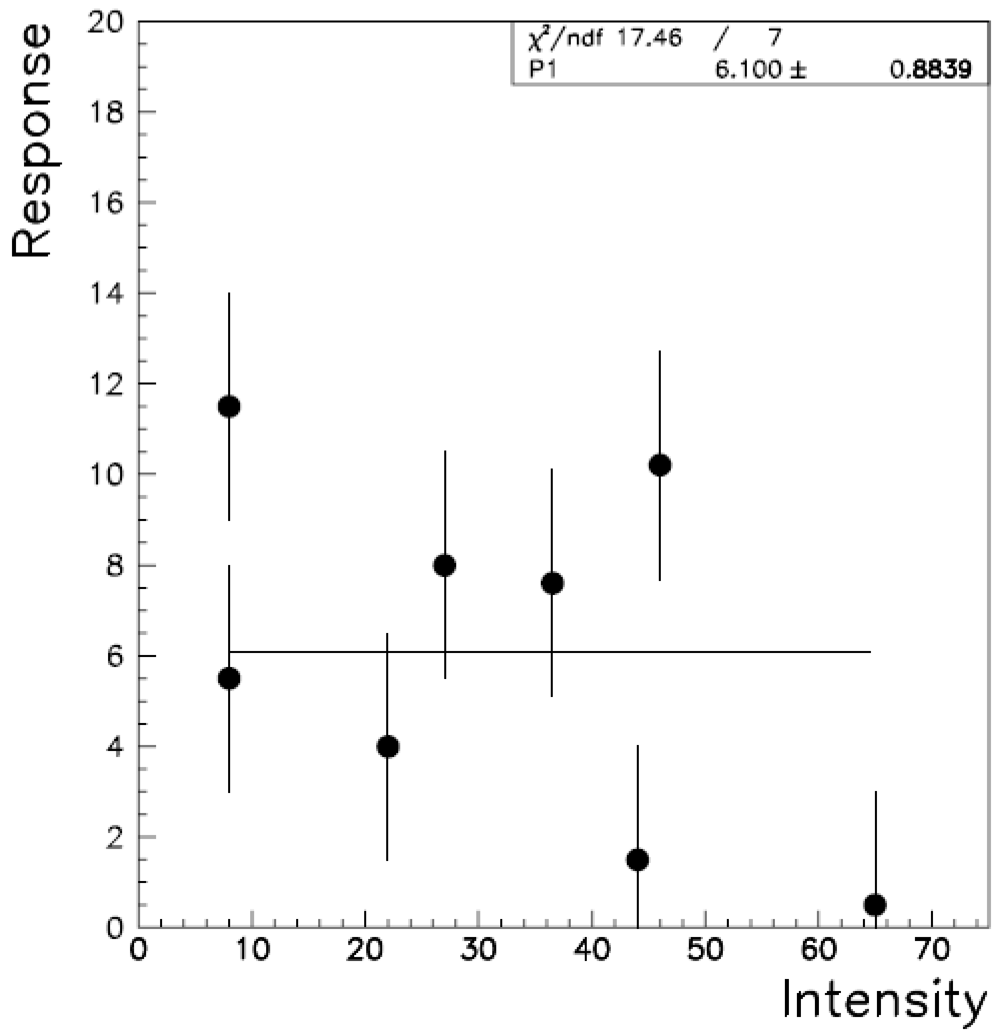
$$\chi^2/\text{ndf} = 1 \pm \sqrt{\frac{2}{\text{ndf}}}$$

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



A case study

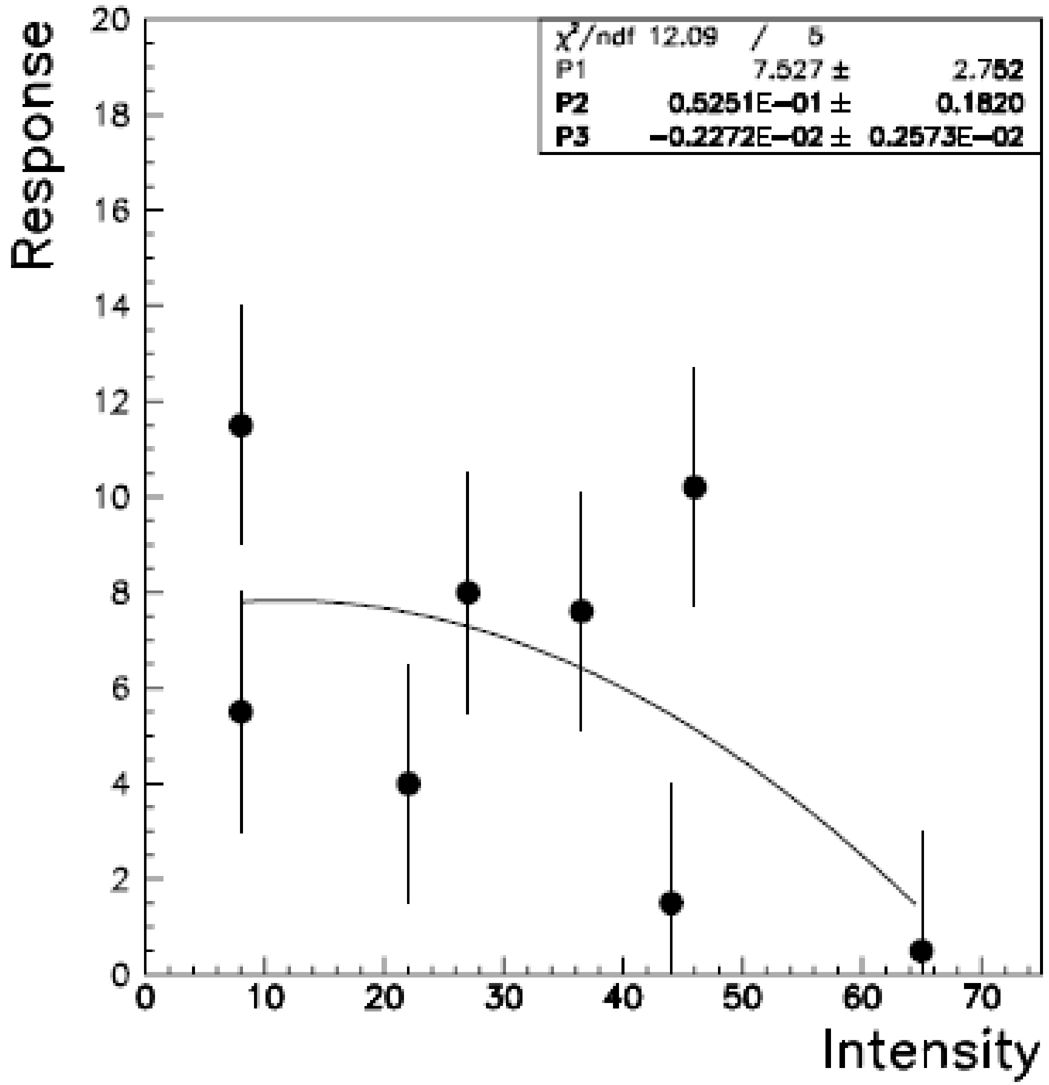
$$\chi^2/\mathrm{ndf} = 1 \pm \sqrt{\frac{2}{\mathrm{ndf}}}$$

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 & experi-mental apparatus, data cannot be rejected !



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.