

Uncertainty and Bias

UIUC, 403 Advanced Physics Laboratory, Spring 2013

Matthias Grosse Perdekamp

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



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 $1/(20/6) = 0.3$ 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 0.27 so it goes.

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

The probability that 1 or more measurements are outside 2 sigma for 6 measurements is $1-0.95^6=0.27$

Our case study: A very simple first step

Data set (picked off the graph by hand)

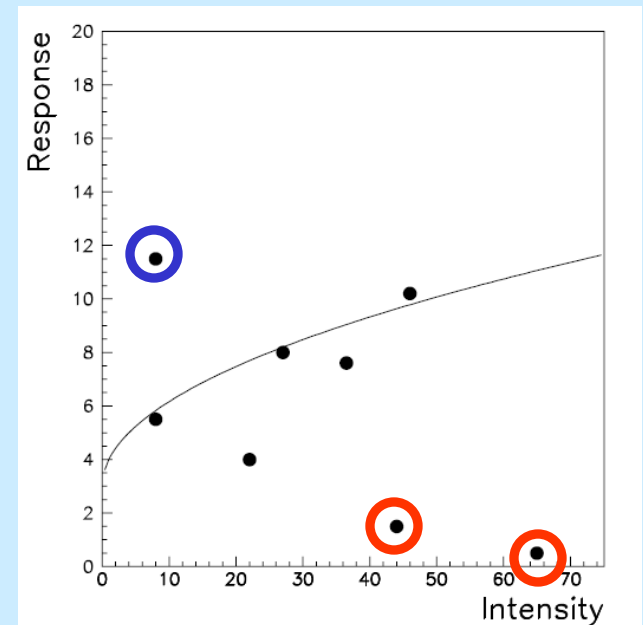
11.5, 5.5, 4.0, 8.0, 7.6, 1.5, 10.2, 0.5 (note, at same beam intensity!)

Mean: = 6.1

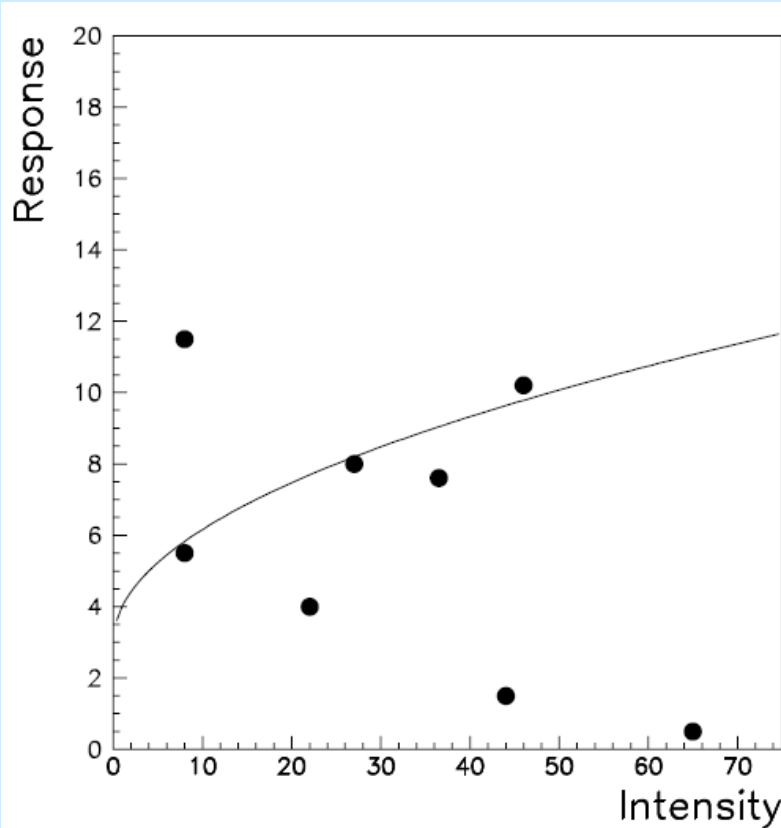
Standard deviation = 4.0

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)

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

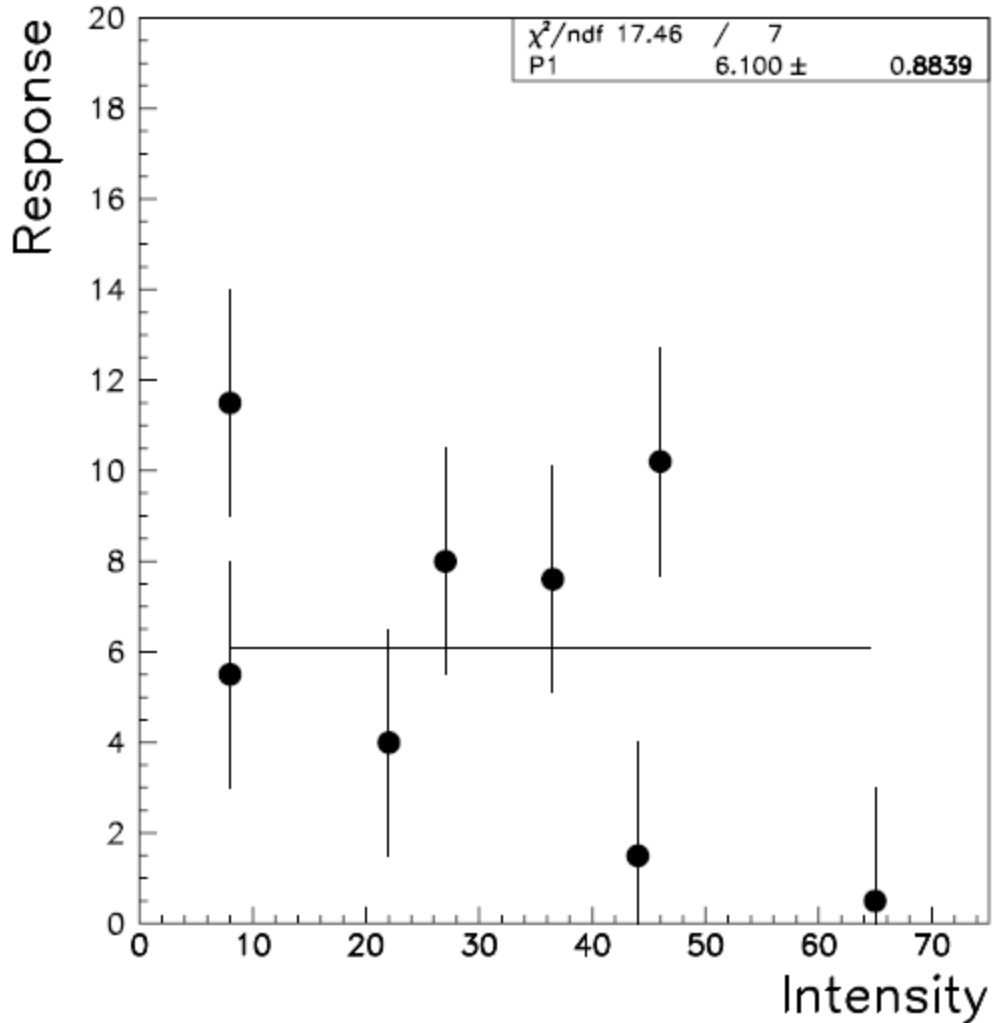


What are the uncertainties ?

Can we attach the power fluctuations to particular data points?

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

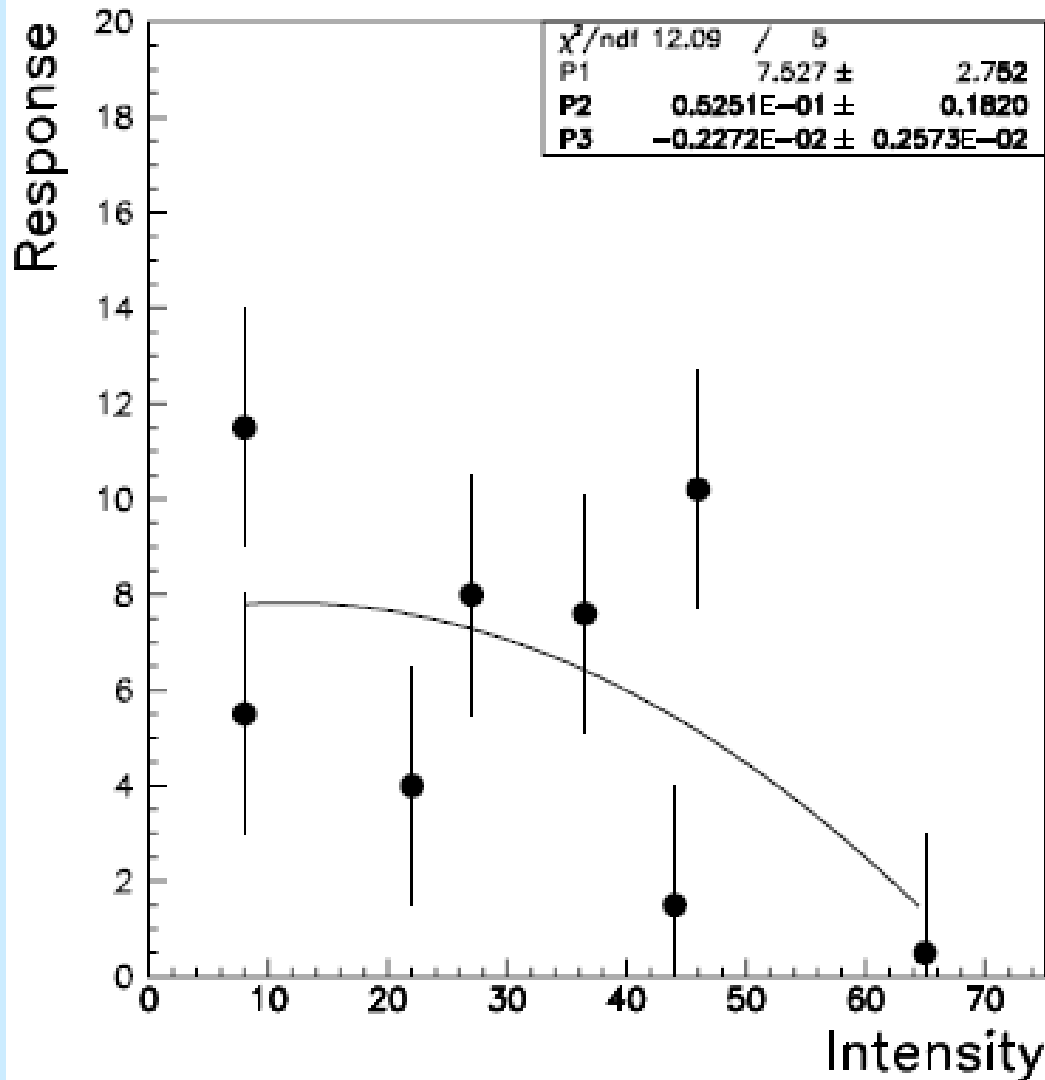
Let's look at our data



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



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

Are the data compatible with a polynomial .
Not sure chi2/ndf is 2.4

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

Repeat: Is all data good data?

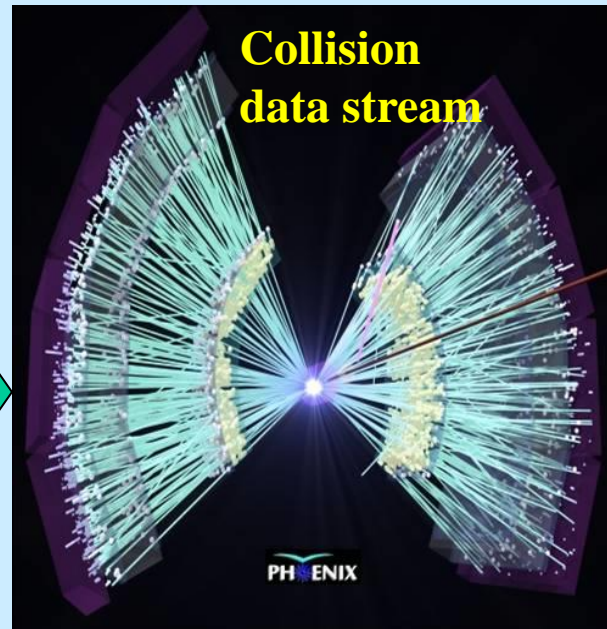
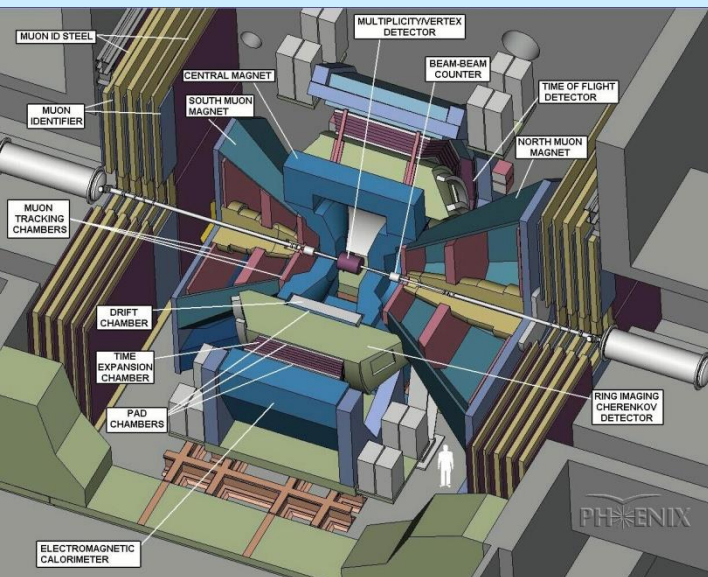
NO!

- 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

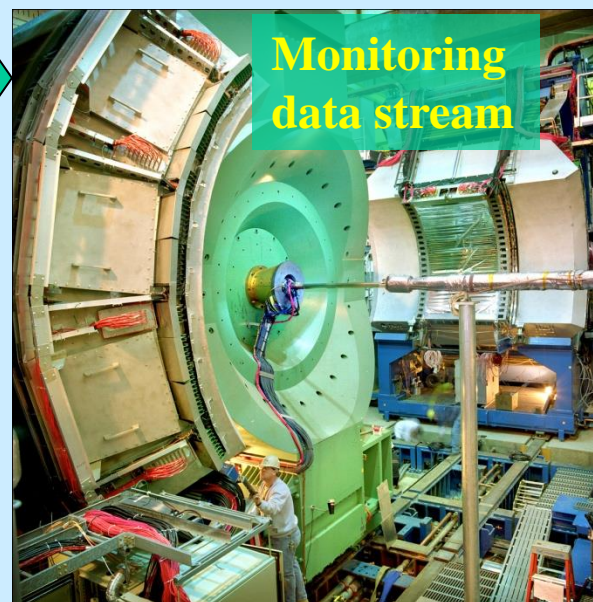
Some additional points

- **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

Example PHENIX

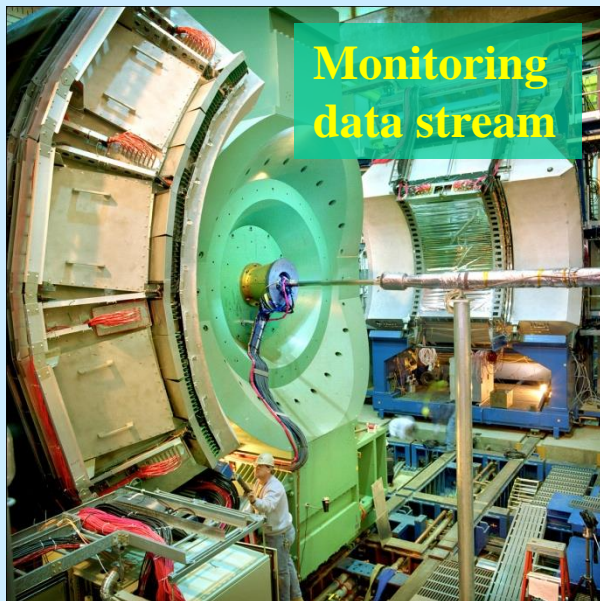


- (I) collision data stream
- o *collision id*
 - o *table of final state particles id, px, py, pz*



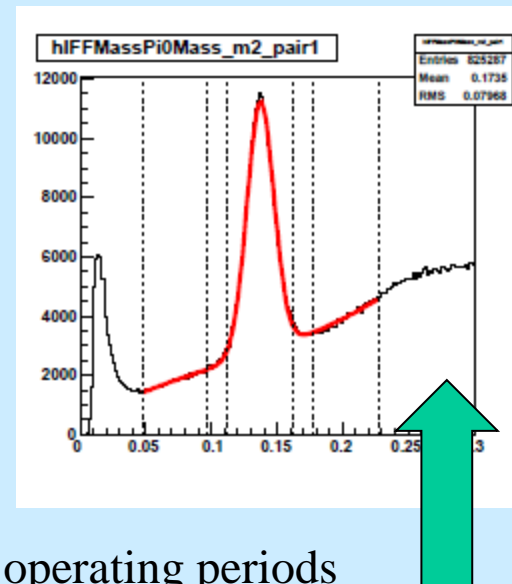
- (II) monitoring data stream
- o *time*
 - o *beam current*
 - o *beam backgrounds*
 - o *environment T, P*
 - o *detector LV, HV, electronics T*
 - o *gas composition*
 - o *DAQ health*

Example PHENIX



(I) carry out analysis of stable operating condition based on monitoring data

- o *time*
- o *beam current*
- o *beam backgrounds*
- o *environment T, P*
- o *detector LV, HV, electronics T*
- o *gas composition*
- o *DAQ health*



(II) Confirm choice of stable operating periods with standard candles from collision data stream

(III) Carry out physics analysis based on collision data stream

- o *collision id*
- o table of final state particles
id, px, py, pz

