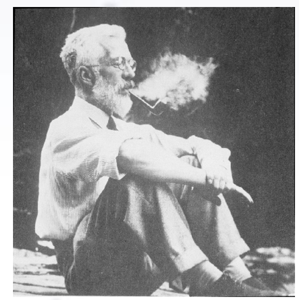
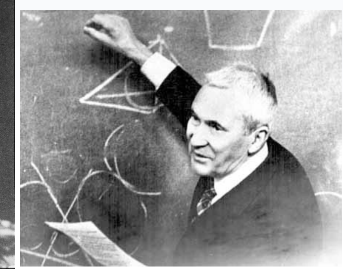
Applied Statistics

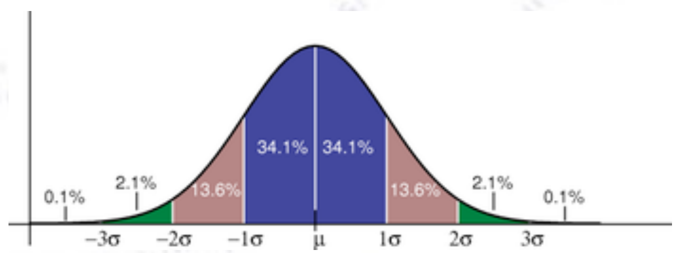
Good Experimental and Statistical practices





Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Optimal data analysis & cross checks

As an observational astronomer, you have survived a long airline flight, a terrifying taxi ride to the summit of a volcano, days of dodgy weather, hours coaxing flakey equipment back into an orderly lifestyle, exhaustion. At last, you attain that exalted state of resonance with machine and sky. Your equipment is working in miraculous defiance of Murphy's Law.

Everything that could go wrong did, but now you have emerged to savour a long clear night plucking data from the sky. Thus you succeed in acquiring an astronomical dataset. **After such an ordeal, giving birth to the data, it seems shameful, even criminal, to analyse your data with anything less than optimal methods.**

[Keith Horne (Univ. of St. Andrews), "The Ways of Our Errors" (book in preparation)]

Error propagation

The simplest experiments and subsequent extraction of a measurement are usually the best! Propagating errors to the final result (as below) is key.

When you count something, then there is (mostly) not doubt about the result:

$35000 \pm (\sqrt{35000} = 187)$

175 ± 8

$$\sigma^{fid} = \frac{N_{data} - N_{bkg}}{C_{fid} L}$$

0.552 ± 0.006

$(81 \pm 2)\ \mathrm{pb}^{-1}$

Error propagation

The simplest experiments and subsequent extraction of a measurement are usually the best! Propagating errors to the final result (as below) is key.

When you count something, then there is (mostly) not doubt about the result:

$35000 \pm (\sqrt{35000} = 187)$

175 ± 8

$$\sigma^{fid} = \frac{N_{data} - N_{bkg}}{C_{fid} L}$$

0.552 ± 0.006

$(81 \pm 2)\ \mathrm{pb}^{-1}$

Quantifying your uncertainties is hard work, that requires forethought!
Did you carefully think about your experimental setup and data taking **before** you actually did the experiment?

$$\boxed{\sigma^{fid} = 0.781 \pm 0.004\ (\mathrm{stat}) \pm 0.008\ (\mathrm{syst}) \pm 0.016\ (\mathrm{lumi})}$$

Defining the Chi-Square

<u>Problem Statement:</u> Given N data points (x,y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_i^N \frac{(y_i - f(x_i,\theta))^2}{\sigma_i^2}$$

The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Chi-Square probability interpretation

The Chi-Square probability can roughly be interpreted as follows:

- If χ^2 / Ndof ≃ 1 or more precisely if $0.01 < p(\chi^2, Ndof) < 0.99$, then all is good.
- If χ^2 / Ndof ≫ 1 or more precisely if $p(\chi^2, Ndof) < 0.01$, then your fit is bad, and your hypothesis is probably not correct.
- If χ^2 / Ndof ≪ 1 or more precisely if $0.99 < p(\chi^2, Ndof)$, then your fit is TOO good and you probably overestimated the errors.

Note:
If the statistics behind the plot is VERY high (great than 10^6), then you might have a hard time finding a model, which truly describes all the features in the plot (as now tiny effects become visible), and one hardly ever gets a good Chi-Square probability.
The "missing uncertainty" that would make your Chi2 reasonable is often a good measure of the systematic uncertainties (effects that you don't understand or know).

A little bit on experiments…

Measurement situation

There are four possible situations in experimental measurements of a quantity:

One measurement, no error:

X = 3.14

Situation: You are f***ed!

You have no clue about uncertainty, and you can not obtain it!

Several measurements, no errors:

X1 = 3.14
X2 = 3.21
X3 = …

Situation: You are OK

You can combine the measurements, and from RMS get error on mean.

One measurement, with error:

X = 3.14 ± 0.13

Situation: You are OK

You have a number with error, which you can continue with.

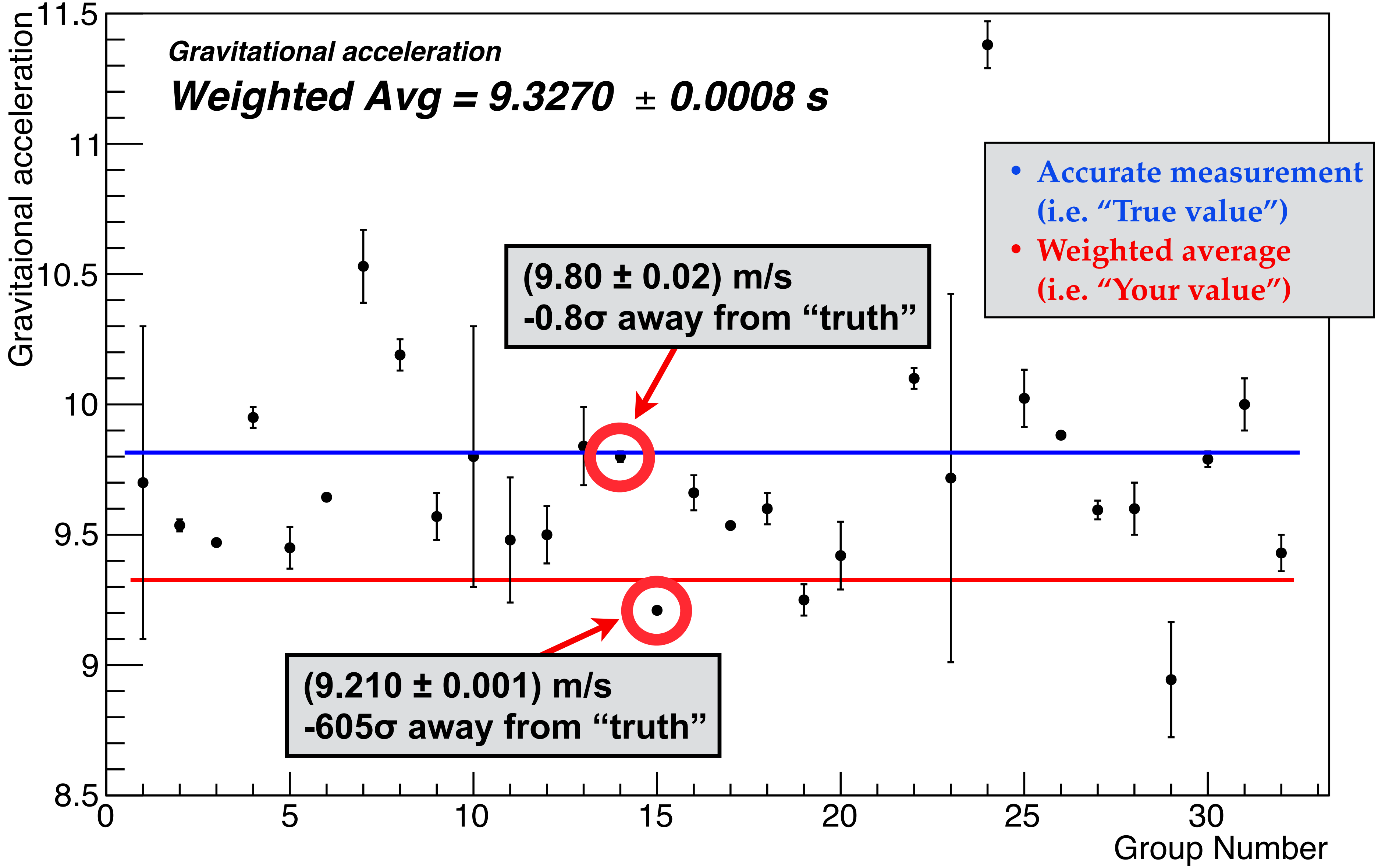
Several measurements, with errors:

X1 = 3.14 ± 0.13
X2 = 3.21 ± 0.09
X3 = …

Situation: You are on top of things!

You can both combine to a weighted average and check with a chi-square.

Example results (Ball on Incline)



The good experiment

It is not simple to say exactly what a **good experimental setup** comprises of. Experience helps a lot, which is why I'm asking you to do experiments in the first place. That may help you towards becoming a **good experimenter**.

Below I've tried to list the things that I think characterises a good experiment, but the list is by no means unbiased nor exhaustive.

<u>Characteristics of good experimental design:</u>

- Provides unbiased estimates and associated uncertainties.
- Enables the experimenter to perform cross checks.
- Includes a plan for analysis and reporting of results.
- Gives results that are easy to interpret.
- Permits conclusions that have wide validity.
- Shows the direction of better results.
- Is as simple as possible.

The good experimenter?

The good experimenter

<u>The good experimenter will always:</u>

- Plan the experiments carefully.
- Attempt individual measurements.
- Inspect data visually.
- Test assumptions.
- Keep an accurate record.
- **Perform cross checks.**
- Do a ChiSquare test when possible.
- Try to “blind” results until final.
- Check if numbers are “reasonable”.

<u>The good experimenter will never:</u>

- Not look at the raw data.
- Rely on untested assumptions.
- “Just let someones program do it”.
- Make changes in data.
- Look only for effects in one direction.