ERROR PROPAGATION

And an introduction to the Method of Least Squares

One case study of data analysis for a single variable

AtMol. Durham Atomic & Molecular Physics

493

nature10104

tron

Yb ta
Solen
valv

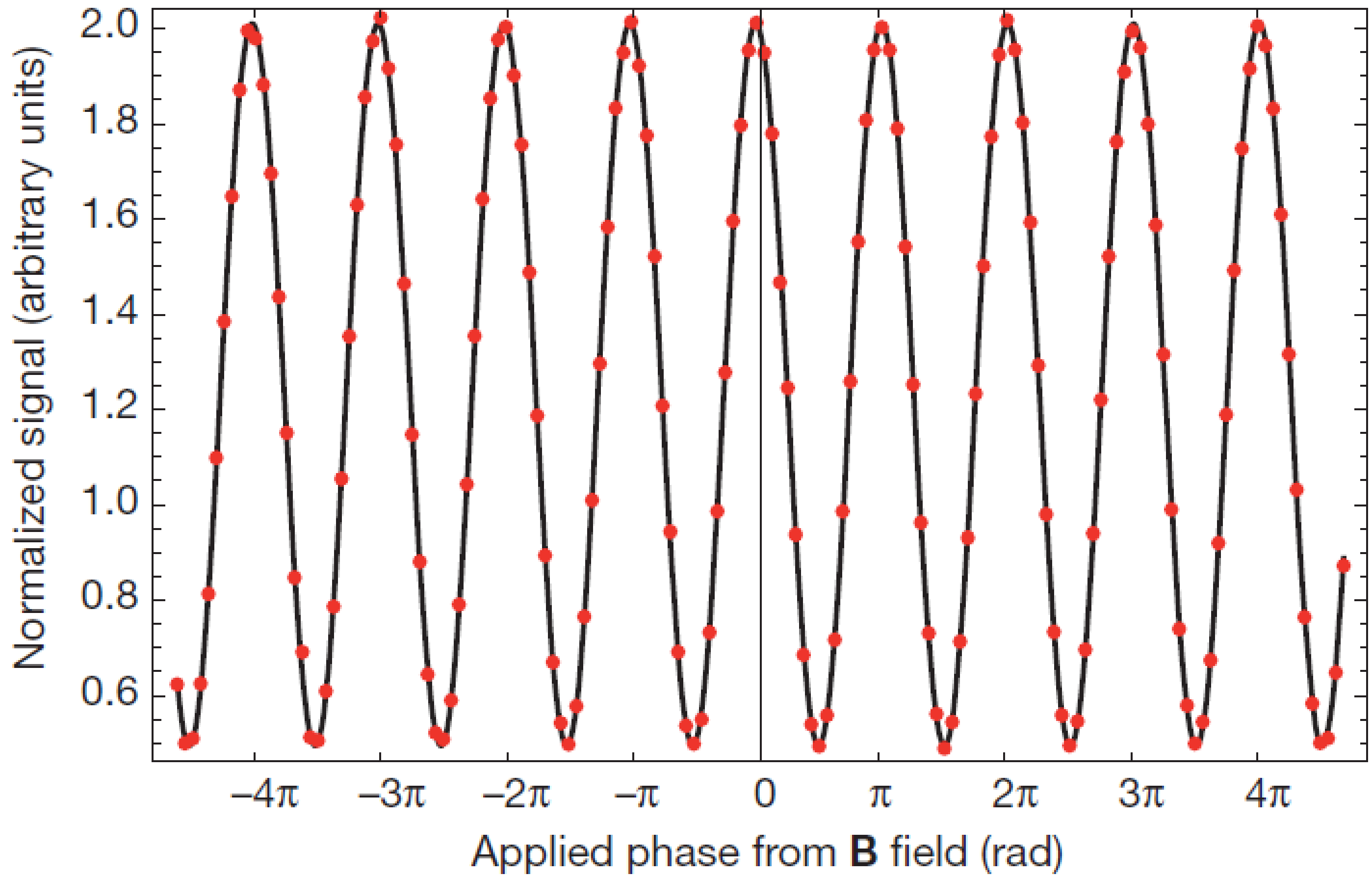


Figure 3 | Interferometer fringes produced by magnetic field scan. Dots indicate the probe fluorescence normalized to the pump fluorescence. The line is the fit to the cosine-squared model.

Figur
particl
matter
particl
electro
provid
d_e = (
the ele
with 9
indica
precisi
a mol
scale[2].

f YbF mol-
hose mole-
l state. The
the pump
t the $F = 1$
s, between
ere $\hat{\mathbf{z}}$ is the
agnetically
r molecules
where m_F is
z-axis. The
e $m_F = \pm 1$
$d_e E_{eff}) T/h$,
eman shift
e effective
lse is then
l to $\cos^2\phi$,
asures. For
the pump
l is shown

Our measurement is derived from 6,194 blocks of data taken in 2010, comprising 25 million molecular beam pulses, together with many subsidiary measurements used to search for systematic errors. To analyse the data, we select the central 130 μs of each probe pulse

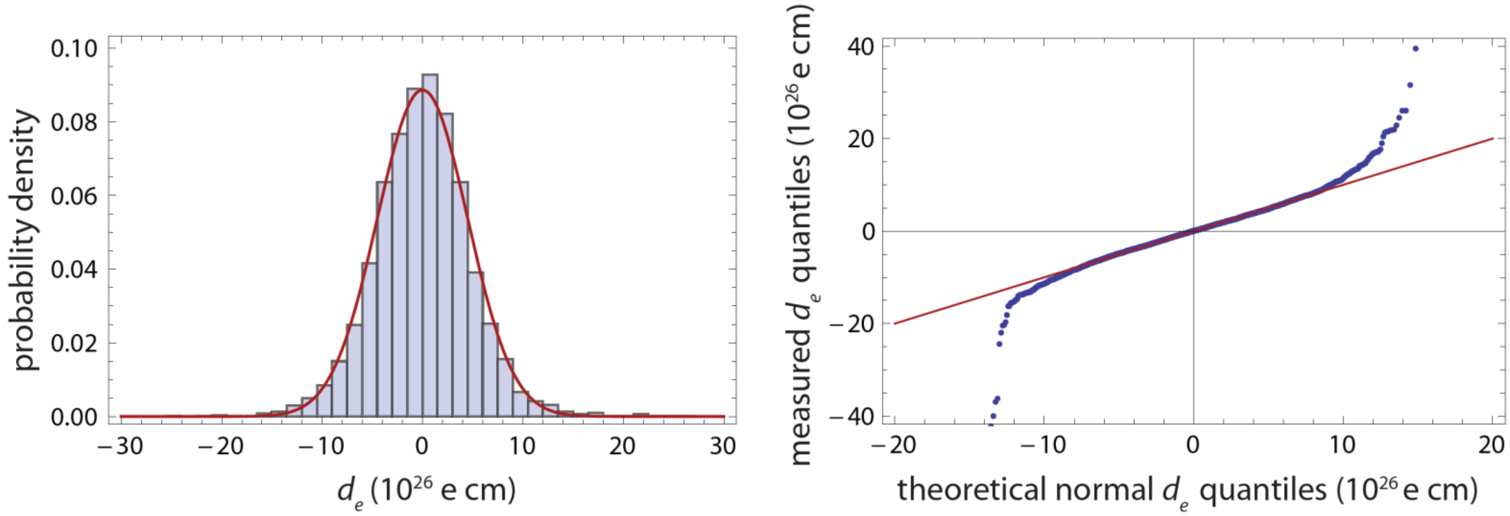
The EDM values obtained from the set of blocks are almost normally distributed but there tend to be a few more points in the wings of the distribution than in a normal distribution. The same is true of other quantities of interest that we extract from the data. For all these quantities, we calculate the 5% trimmed mean[19], a simple robust statistic that drops the largest and smallest 5% of the data. We use the bootstrap method[20] to determine the associated statistical uncertainty. For non-normal distributions, these methods give more reliable measures than the mean and standard error.

unknown offset, which was only removed once the data collection and analysis were complete.

Robust statistics and **bootstrap sampling methods** used for non-normal distributions

Electron EDM measurement – Smallman thesis

AtMol. Durham Atomic & Molecular Physics



(a) (b)

FIGURE 3.6: The distribution of 6194 d_e values that make up the 2011 dataset. For the most part, the distribution is fit well by a Gaussian distribution (red line) with a standard deviation given by the 68.3% statistical uncertainty of the bootstrapped trimmed mean multiplied by $\sqrt{6194}$. The distribution does, however, deviate from the Gaussian fit out in the wings, as demonstrated by the quantile-quantile plot on the right.

Brute-force numerical techniques are becoming more prevalent in data analysis

In total, we collected 1024 blocks (360.3 hours) of

Selected for a Viewpoint in *Physics*
PRL **119**, 153001 (2017) PHYSICAL REVIEW LETTERS week ending 13 OCTOBER 2017

Precision Measurement of the Electron's Electric Dipole Moment Using Trapped Molecular Ions

William B. Cairncross,[*] Daniel N. Gresh, Matt Grau,[†] Kevin C. Cossel,[‡] Tanya S. Roussy, Yiqi Ni,[§] Yan Zhou, Jun Ye, and Eric A. Cornell
JILA, NIST and University of Colorado, Boulder, Colorado 80309-0440, USA
and Department of Physics, University of Colorado, Boulder, Colorado 80309-0440, USA

(Received 21 April 2017; published 9 October 2017)

We describe the first precision measurement of the electron's electric dipole moment (d_e) using trapped molecular ions, demonstrating the application of spin interrogation times over 700 ms to achieve high sensitivity and stringent rejection of systematic errors. Through electron spin resonance spectroscopy on $^{180}Hf^{19}F^{+}$ in its metastable $^{3}\Delta_1$ electronic state, we obtain $d_e = (0.9 \pm 7.7_{\mathrm{stat}} \pm 1.7_{\mathrm{syst}}) \times 10^{-29}$ e cm, resulting in an upper bound of $|d_e| < 1.3 \times 10^{-28}$ e cm (90% confidence). Our result provides independent confirmation of the current upper bound of $|d_e| < 9.4 \times 10^{-29}$ e cm [J. Baron *et al.*, New J. Phys. **19**, 073029 (2017)], and offers the potential to improve on this limit in the near future.

DOI: 10.1103/PhysRevLett.119.153001

The reduced chi-squared statistic for fitting a weighted mean to the eEDM data set is $\chi_r^2 = 1.22(5)$. This overscatter is present in all frequency channels, and is attributable to

$$d_e = (4.3 \pm 3.1_{\text{stat}} \pm 2.6_{\text{syst}}) \times 10^{-30} e \text{ cm} \quad (4)$$

where the combined statistical and systematic uncertainty, $\sigma_{d_e} = 4.0 \times 10^{-30} e$ cm, is a factor of 12 smaller than the previous best result, from ACME I[1,9].

An upper limit on $|d_e|$ is computed by applying the Feldman–Cousins prescription[9,33] to a folded normal distribution, which yields

$$|d_e| < 1.1 \times 10^{-29} e \text{ cm} \quad (5)$$

at 90% confidence level. This is 8.6 times smaller than the best previous limit, from ACME I[1,9]. Because paramagnetic molecules are sensitive

Code availability. The computer codes used for the analysis of the data are available from the corresponding authors on reasonable request.

Data availability

The data that support the conclusions of this article are available from the corresponding authors on reasonable request.

Standard deviations from mean

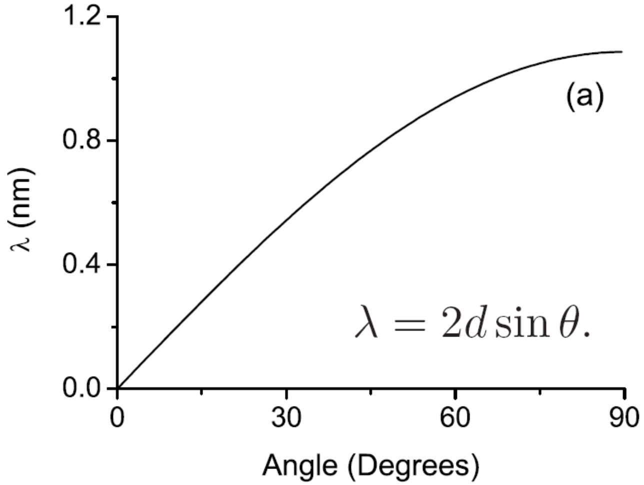
University

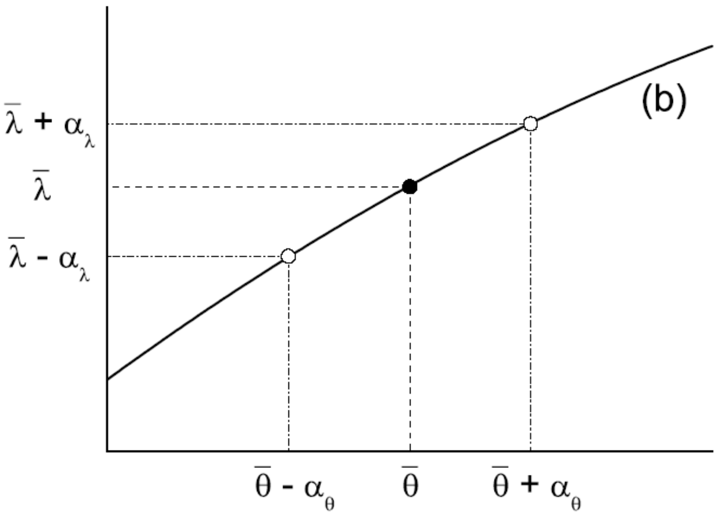
In most experiments we have to combine measurements of several variables into a single quantity

Understanding how to propagate the error is a vital part of **data analysis** and **reduction**

Understanding which factors contribute to the **limiting error** is a vital part of **experimental design**

Error propagation in a single-variable function – functional approach





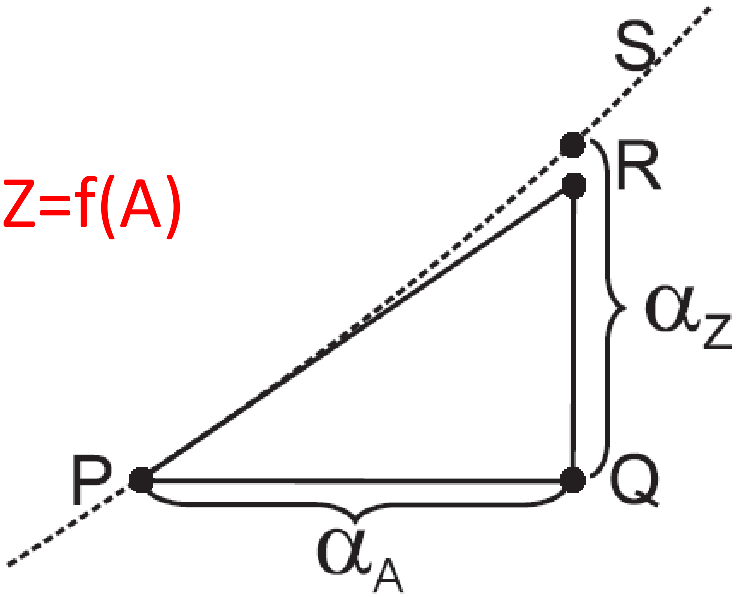
Z=f(A)

$$\begin{aligned}\bar{Z} \pm \alpha_Z &= f\left(\bar{A} + \alpha_A\right), \\ \bar{Z} &= f\left(\bar{A}\right), \\ \bar{Z} \mp \alpha_Z &= f\left(\bar{A} - \alpha_A\right).\end{aligned}$$

$$\alpha_Z = \left| f\left(\bar{A} + \alpha_A\right) - f\left(\bar{A}\right) \right| .$$

Measurements and their Uncertainties, p 37

Z=f(A)



$R \cong S$

$$P = \left(\bar{A}, f\left(\bar{A}\right)\right)$$

$$Q = \left(\bar{A} + \alpha_A, f\left(\bar{A}\right)\right)$$

$$R = \left(\bar{A} + \alpha_A, f\left(\bar{A}\right) + \frac{\mathrm{d}f}{\mathrm{d}A} \times \alpha_A\right)$$

$$S = \left(\bar{A} + \alpha_A, f\left(\bar{A} + \alpha_A\right)\right)$$

$$f\left(\bar{A}\right) + \frac{\mathrm{d}f\left(A\right)}{\mathrm{d}A}\alpha_A = f\left(\bar{A} + \alpha_A\right),$$

$$\alpha_Z = \left|\frac{\mathrm{d}Z}{\mathrm{d}A}\right| \alpha_A.$$

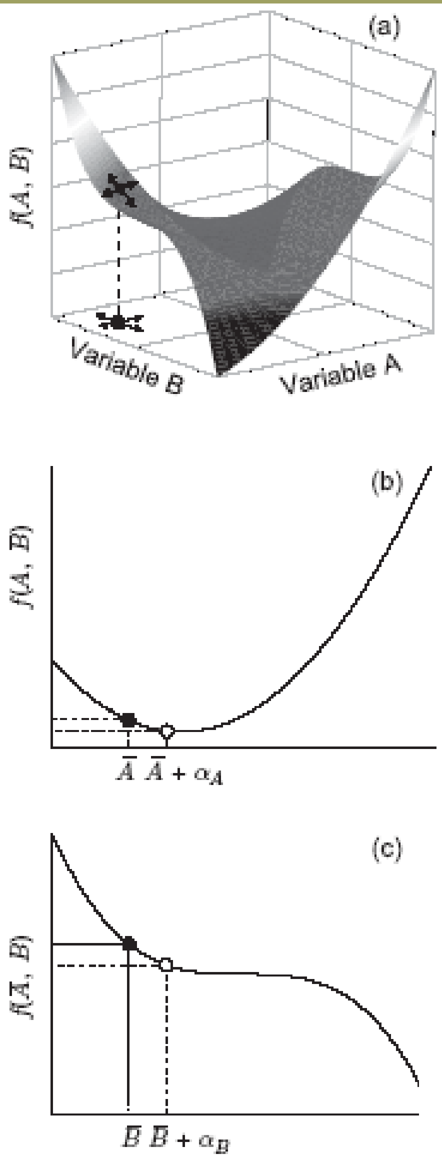
AtMol. Durham Atomic & Molecular Physics

- Calculus approach

Function, $Z(A)$	$\frac{dZ}{dA}$	Uncertainty
$\frac{1}{A}$	$-\frac{1}{A^2}$	$\alpha_z = \frac{\alpha_A}{A^2} = Z^2\alpha_A$ OR $\left\lvert\frac{\alpha_Z}{Z}\right\rvert = \left\lvert\frac{\alpha_A}{A}\right\rvert$
$\exp A$	$\exp A$	$\alpha_z = \exp A\ \alpha_A = Z\ \alpha_A$
$\ln A$	$\frac{1}{A}$	$\alpha_z = \frac{\alpha_A}{A}$
$\log A$	$\frac{1}{\ln(10)\ A}$	$\alpha_z = \frac{\alpha_A}{\ln(10)\ A}$
A^n	$n\ A^{n-1}$	$\alpha_z = \left\lvert nA^{n-1}\right\rvert \alpha_A$ OR $\left\lvert\frac{\alpha_Z}{Z}\right\rvert = \left\lvert n\frac{\alpha_A}{A}\right\rvert$
10^A	$10^A \ln(10)$	$\alpha_z = 10^A\ \ln(10)\,\alpha_A$
$\sin A$	$\cos A$	$\alpha_z = \lvert\cos A\rvert\ \alpha_A$
$\cos A$	$-\sin A$	$\alpha_z = \lvert\sin A\rvert\ \alpha_A$
$\tan A$	$1 + \tan^2 A$	$\alpha_z = \left(1 + Z^2\right)\ \alpha_A$

Measurements and their Uncertainties, inside front cover

AtMol. Durham Atomic & Molecular Physics



$$\alpha_Z^A = f\left(\bar{A}+\alpha_A, \bar{B}\right) - f\left(\bar{A}, \bar{B}\right). \quad \alpha_Z^B = f\left(\bar{A}, \bar{B}+\alpha_B\right) - f\left(\bar{A}, \bar{B}\right)$$

$$(\alpha_Z)^2 = \left(\alpha_Z^A\right)^2 + \left(\alpha_Z^B\right)^2 + \left(\alpha_Z^C\right)^2 + \cdots$$

$$\begin{aligned}(\alpha_Z)^2 = &\left[f\left(\bar{A}+\alpha_A, \bar{B}, \bar{C}, \ldots\right) - f\left(\bar{A}, \bar{B}, \bar{C}, \ldots\right)\right]^2 \\ &+\left[f\left(\bar{A}, \bar{B}+\alpha_B, \bar{C}, \ldots\right) - f\left(\bar{A}, \bar{B}, \bar{C}, \ldots\right)\right]^2 \\ &+\left[f\left(\bar{A}, \bar{B}, \bar{C}+\alpha_C, \ldots\right) - f\left(\bar{A}, \bar{B}, \bar{C}, \ldots\right)\right]^2 \\ &+\cdots\end{aligned}$$

A calculus approximation to this result is

$$(\alpha_Z)^2 = \left(\frac{\partial Z}{\partial A}\right)^2 (\alpha_A)^2 + \left(\frac{\partial Z}{\partial B}\right)^2 (\alpha_B)^2 + \left(\frac{\partial Z}{\partial C}\right)^2 (\alpha_C)^2 + \cdots$$

Percentage errors – dominant error

It is frequently possible to bypass the need to use the rigorous expression if one performs a quick back of the envelope calculation.
For example, if $Z = A \times B \times C \times D$, and A is known to 5%, and B, C and D to 1%, what is the percentage error in Z?

The experienced practitioner will not have to use the appropriate formula from Table 4.2 as the addition of the percentage errors in quadrature will yield, to one significant figure, 5%.

RULE OF THUMB: Perform a quick calculation to identify any dominant errors. Consider whether a more rigorous calculation is useful.

Correlated errors – a health warning

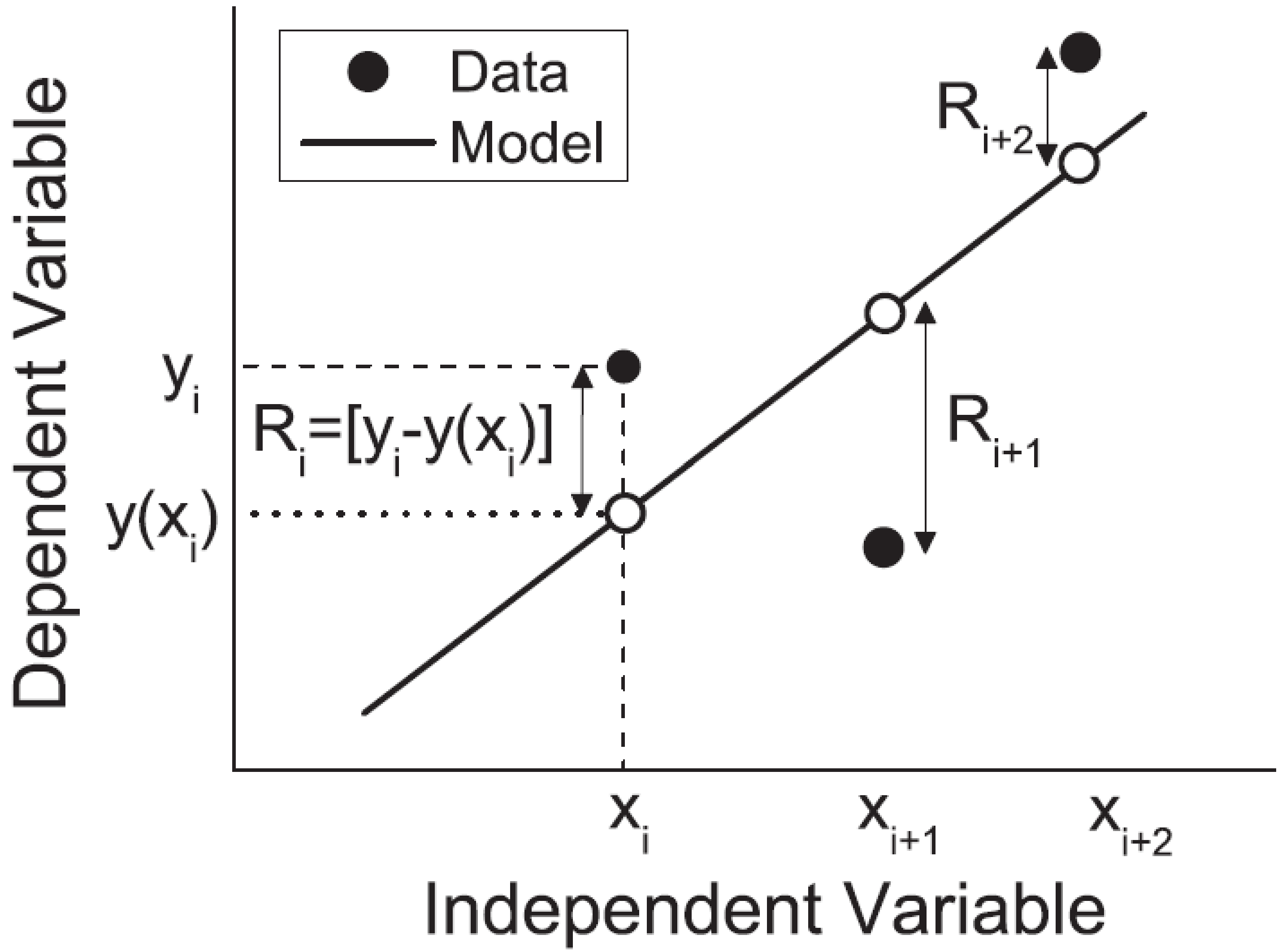
$$Z = \frac{(A - B)}{(A + B)}$$

One might be tempted to use the look-up tables and make the substitutions
X (A, B) = (A − B) and *Y (A, B) = (A + B)* such that *Z = X/Y*

This would be WRONG as the errors in *X* and *Y* are correlated

Only substitute single-variable functions when using the look-up tables. Recall that the look-up tables are only valid for independent variables or single-variable functions.

The importance of residuals

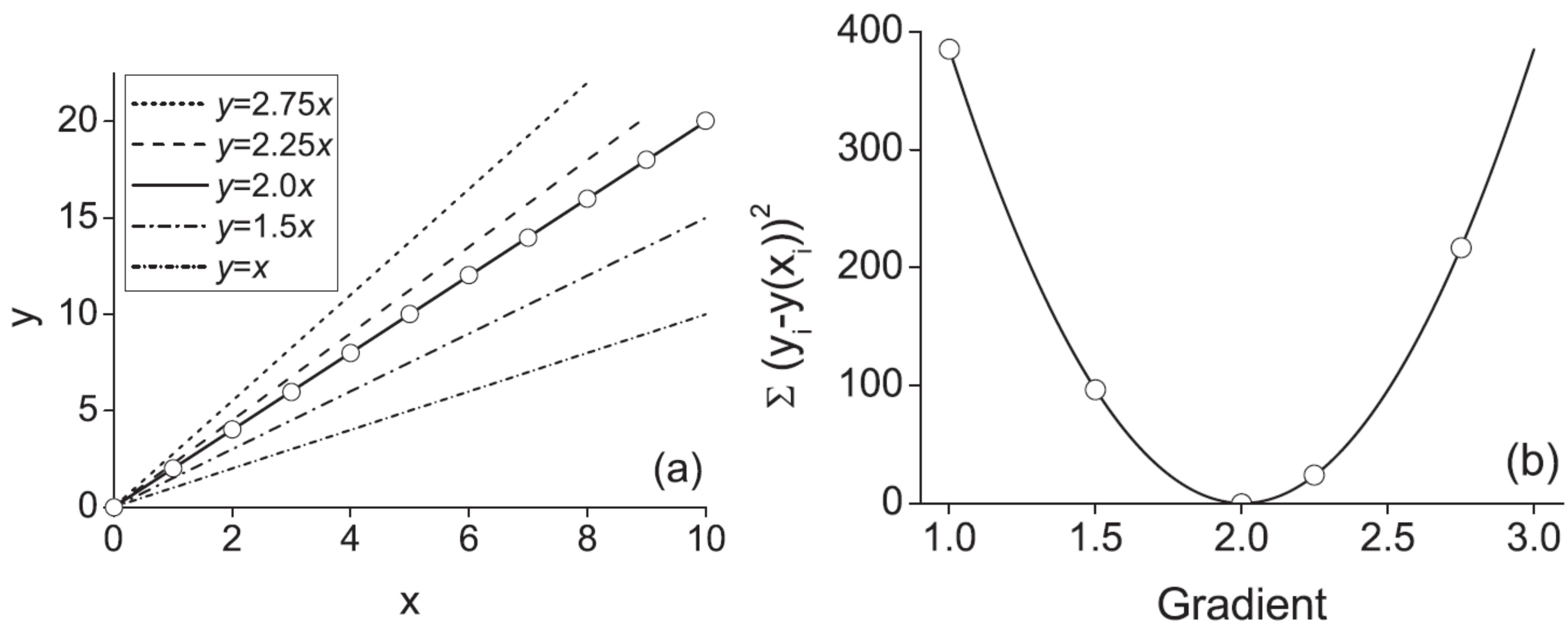


Measurements and their Uncertainties, Pg. 60

The goodness-of-fit parameter

$$\chi^2 = \sum_i \frac{\left(y_i - y\left(x_i\right)\right)^2}{\alpha_i^2}$$

Note that χ^2 is dimensionless



Measurements and their Uncertainties, Pg. 61

χ^2 for data with Poisson errors

AtMol.
Durham Atomic & Molecular Physics

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

i is the number of counts (integer)

O_i is the observed number of occurrences (integer)

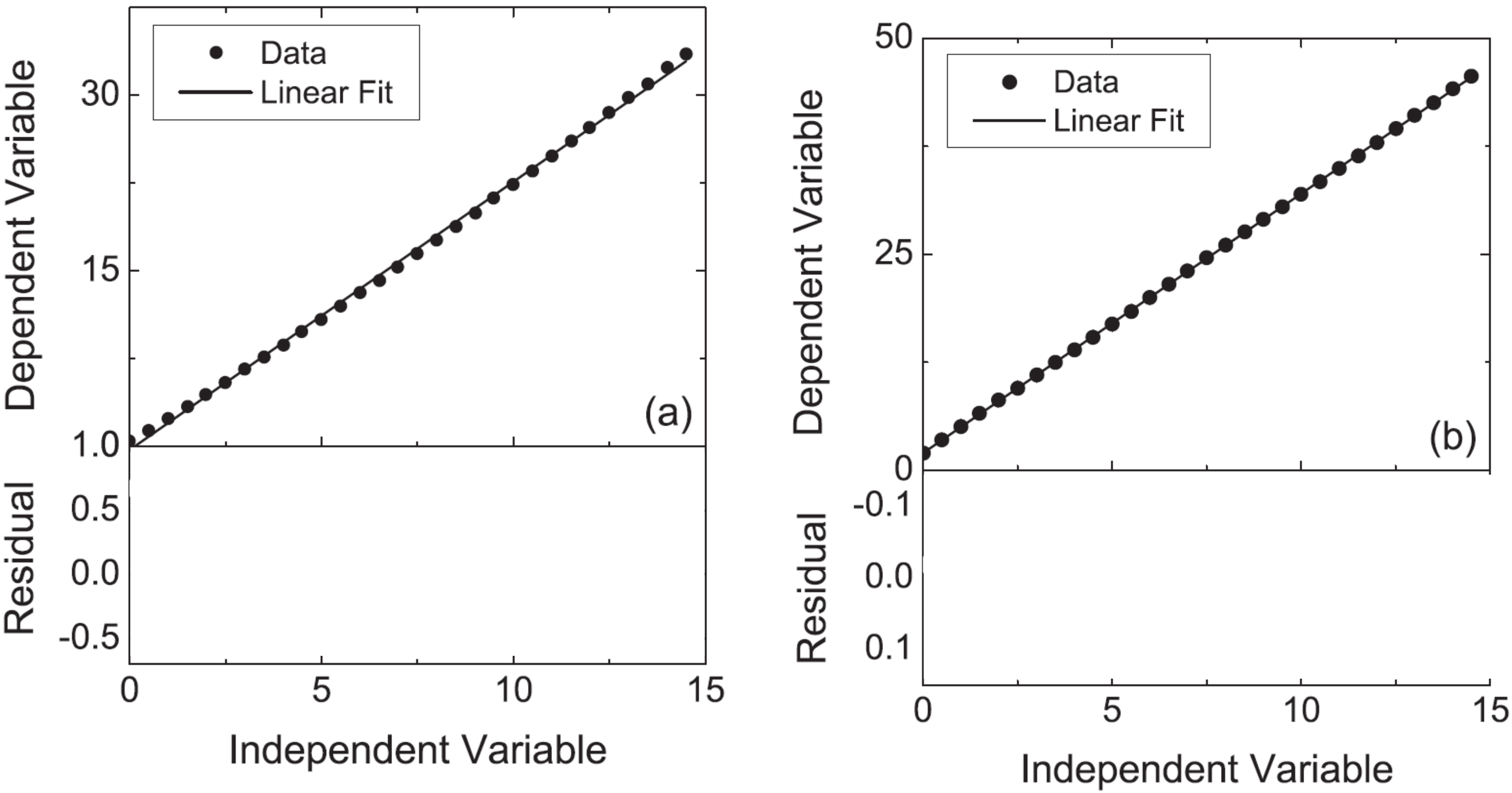
E_i is the expected number of occurrences

O_i is the observed number of occurrences (integer)

Note that χ^2 is (again) dimensionless

Measurements and their Uncertainties, Pg. 68

After fitting a straight-line graph – plot residuals

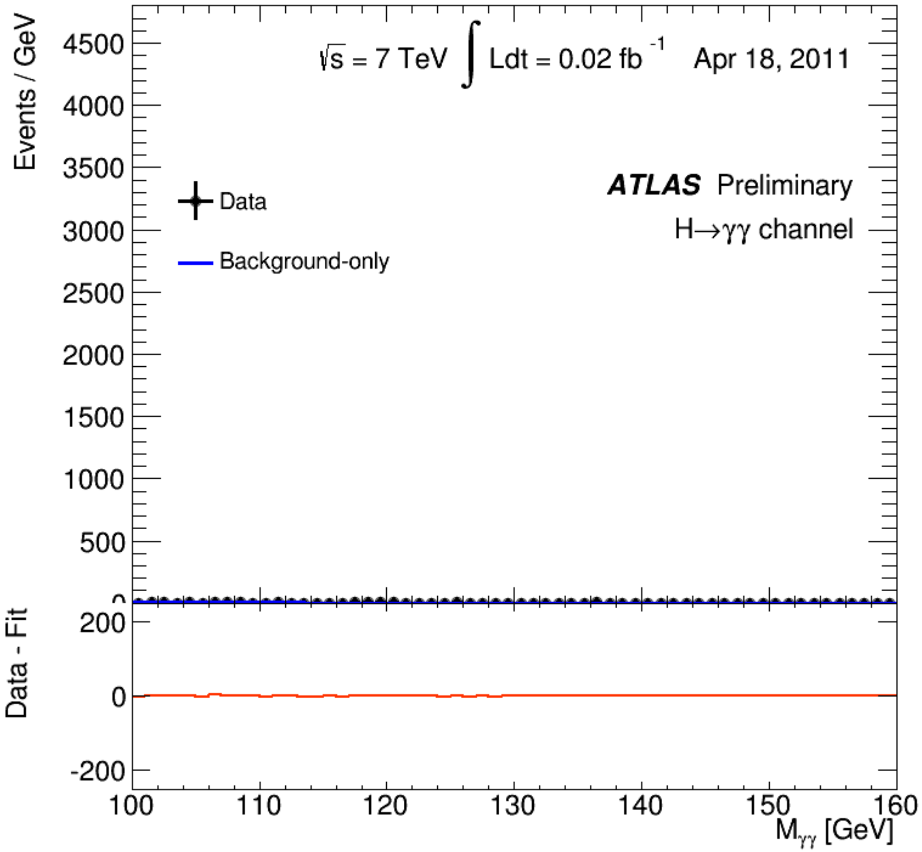


Measurements and their Uncertainties, Pg. 63

After fitting a nonlinear function – plot residuals

AtMol.
Durham Atomic & Molecular Physics

Plotting the residuals makes certain features (the existence of a Higgs boson) more apparent



Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC
Phys.Lett. B716 1-29 (2012)

Minimisation: The Method of Least Squares

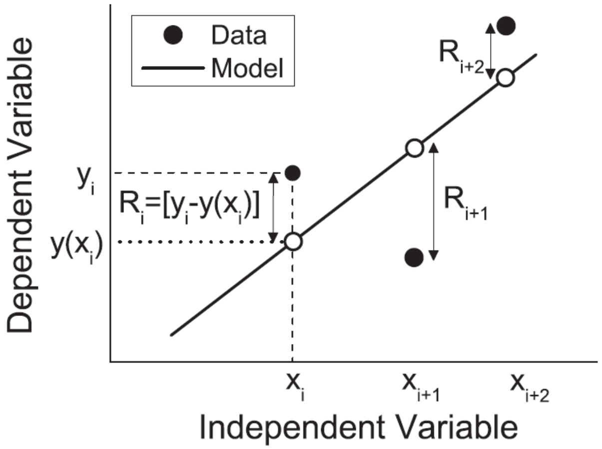
$$S=\sum_i\left(y_i-y\left(x_i\right)\right)^2=\sum_i \mathrm{R}_i^2$$

$$S=\sum_i\left(y_i-mx_i-c\right)^2$$

$$S \text{ is a minimum when } \frac{\partial S}{\partial m}=\frac{\partial S}{\partial c}=0$$

$$\frac{\partial S}{\partial m}=-2\sum_i\left(x_i\left[y_i-mx_i-c\right]\right)=0$$

$$\frac{\partial S}{\partial c}=-2\sum_i\left(y_i-mx_i-c\right)=0.$$



Measurements and their Uncertainties, p 58

Therefore the required values of m and c are obtained from the two simultaneous equations

$$m\sum_i x_i^2 + c\sum_i x_i = \sum_i x_i y_i$$
$$m\sum_i x_i + cN = \sum_i y_i$$

Note that the last equations demonstrates that the best-fit straight line must go through the point $(\bar{x}, \bar{y})$, where $\bar{x} = \frac{1}{N}\sum_i x_i$, $\bar{y} = \frac{1}{N}\sum_i y_i$.

Measurements and their Uncertainties, p 58

In this case by ignoring the *y*-error bars we obtain a simple ANALYTIC SOLUTIONS for the GRADIENT, INTERCEPT and their UNCERTAINTIES

$$\Delta = N\sum x_i^2 - \left(\sum x_i\right)^2$$

$$c = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{\Delta},$$

$$\alpha_{\mathrm{CU}} = \sqrt{\frac{1}{N-2}\sum_i (y_i - mx_i - c)^2}$$

$$m = \frac{N\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\Delta},$$

$$\alpha_c = \alpha_{\mathrm{CU}}\sqrt{\frac{\sum_i x_i^2}{\Delta}} \qquad \alpha_m = \alpha_{\mathrm{CU}}\sqrt{\frac{N}{\Delta}}$$

Measurements and their Uncertainties, p 58

Best-fit straight line – non-uniform error bars

We use a weighted fit using the **Normalised residual** R_i

$$R_i = \frac{y_i - y(x_i)}{\alpha_i}$$

The sum of the squares of the normalised residuals is called χ^2

This is now a **weighted fit** with the weight w_i being the inverse of the square of the error bar – points with small errors are more important!

Measurements and their Uncertainties, Pg. 67

Best-fit straight line – non-uniform error bars

AtMol. Durham Atomic & Molecular Physics

$$\chi^2 = \sum_i \frac{\left(y_i - y\left(x_i\right)\right)^2}{\alpha_i^2} = \sum_i R_i^2$$

$$\chi^2 = \sum_i w_i \left(y_i - m x_i - c\right)^2$$

$$\chi^2 \text{ is a minimum when } \frac{\partial \chi^2}{\partial m} = \frac{\partial \chi^2}{\partial c} = 0$$

Measurements and their Uncertainties, Pg. 67

Best-fit straight line – non-uniform error bars

Therefore the required values of m and c are obtained from the two simultaneous equations

$$m\sum_i w_i x_i^2 + c\sum_i w_i x_i = \sum_i w_i x_i y_i$$
$$m\sum_i w_i x_i + c\sum_i w_i = \sum_i w_i y_i$$

Note that the last equations demonstrates that the best-fit straight line must go through the point $(\tilde{x}, \tilde{y})$, where $\tilde{x} = \frac{\sum_i w_i x_i}{\sum_i w_i}$, $\tilde{y} = \frac{\sum_i w_i y_i}{\sum_i w_i}$.

Best-fit straight line – non-uniform error bars

$$c = \frac{\sum_i w_i x_i^2 \sum_i w_i y_i - \sum_i w_i x_i \sum_i w_i x_i y_i}{\Delta'}, \quad \alpha_c = \sqrt{\frac{\sum_i w_i x_i^2}{\Delta'}}$$

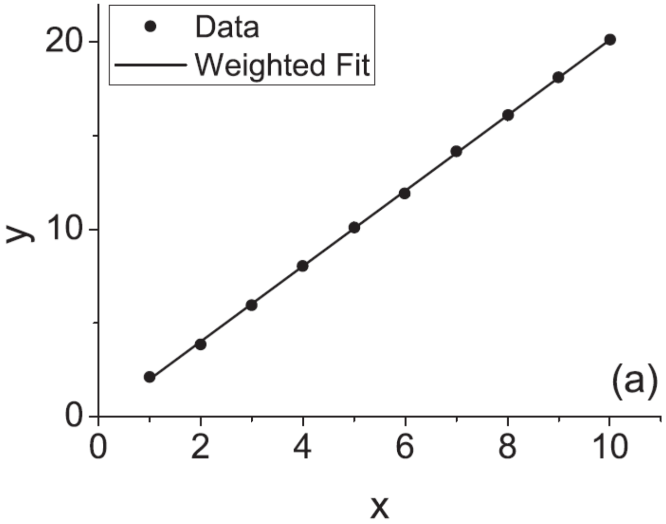
$$m = \frac{\sum_i w_i \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i}{\Delta'}, \quad \alpha_m = \sqrt{\frac{\sum_i w_i}{\Delta'}}$$

$$\Delta' = \sum_i w_i \sum_i w_i x_i^2 - \left(\sum_i w_i x_i\right)^2 \qquad w_i = \alpha_i^{-2}$$

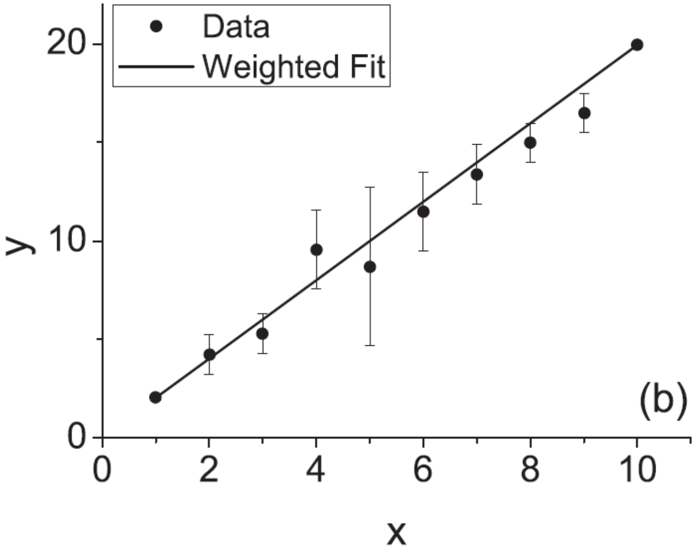
Measurements and their Uncertainties, Pg. 69

Strategies for a straight-line fit

AtMol.
Durham Atomic & Molecular Physics

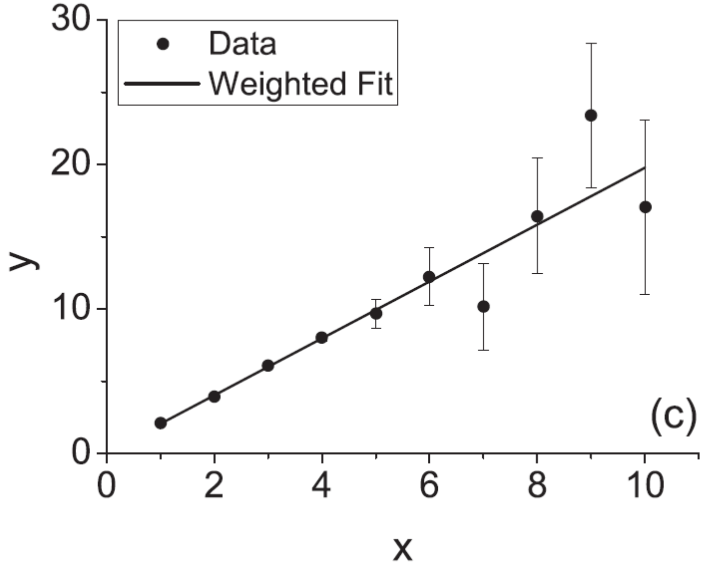


$$m = 2.01 \pm 0.01$$
$$c = -0.02 \pm 0.07$$

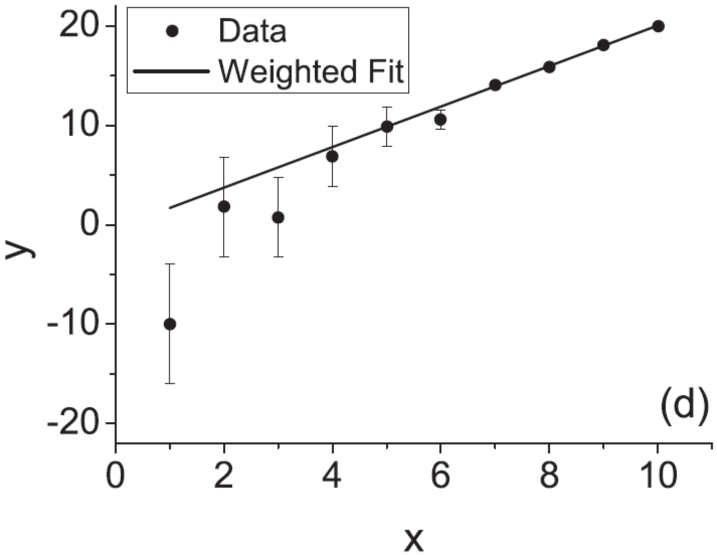


$$m = 1.99 \pm 0.01$$
$$c = 0.04 \pm 0.08$$

Measurements and their Uncertainties, Pg. 71

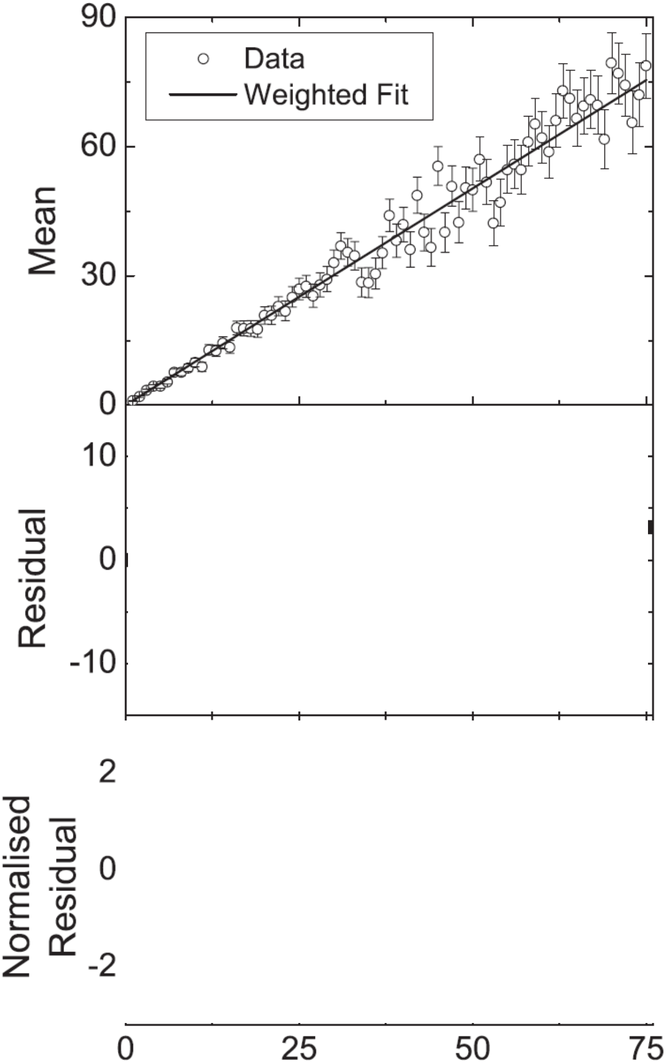


$$m = 1.97 \pm 0.03$$
$$c = 0.09 \pm 0.06$$



$$m = 2.04 \pm 0.04$$
$$c = -0.3 \pm 0.4$$

Measurements and their Uncertainties, Pg. 71



normalised residual,

$$R_i = \frac{y_i - y(x_i)}{\alpha_i},$$

65% are scattered within ± 1

96% are within ± 2.

Measurements and their Uncertainties, Pg. 72

"…error analysis is a participation, rather than a spectator, sport."

Please attempt homework before 5pm next Monday

(21st October).

For MiSCaDA students via https://notebooks.dmaitre.phyip3.dur.ac.uk/miscada-da/hub/login

For PhD students either the notebook server OR email me a document