Physics 259

Uncertainties on DMM measurements

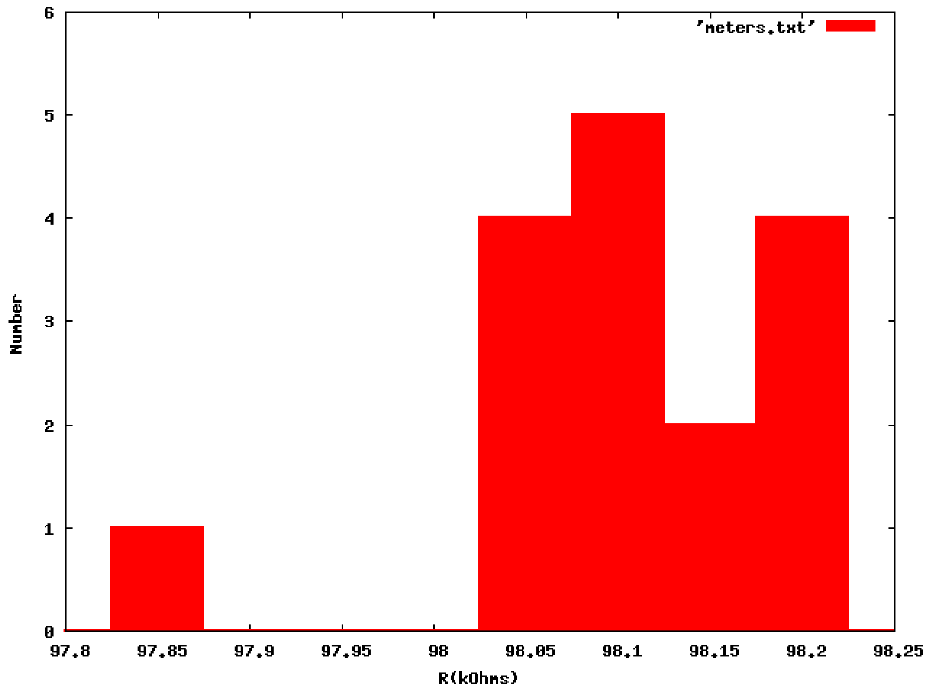
- During one day, we measured the *same* “100k” resistor with most of the Keithley DMMs
- Let’s recall what each of these measurements means
  - We have a TRUE value of the resistor R_0
  - Each measurement is an *estimator* of the true value
  - Each measurement is a *sample* of size 1 that should reflect some probability density function $P(R|R_0)$
    - This pdf is something the manufacturer knows – given a large group of their meters, if we measured the same value with each, what would the resulting distribution look like, and how much of a spread would there be between values.
- YOU only have 1 DMM
  - Your best *estimator* of the true value of the resistor is what the meter shows
  - Your *estimator* of the uncertainty on your estimate is based on the manufacturers data sheet, which encodes their pdf, in principle assuming a Gaussian distribution

DMM uncertainties

- Here is the data

R(kOhms)

98.01
98.06
98.10
98.19
98.02
98.16
97.82
98.08
98.01
98.02
98.07
98.18
98.16
98.13
98.11
98.09



DMM uncertainties

- The resistor was labeled 100k, and had 10% precision
- The *mean value* or *average* of our data is

$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R_i = 98.08k\Omega$$

- This, by the way, is our best estimator for the true value of the resistor IF ALL THE MEASUREMENTS HAVE EQUAL UNCERTAINTY
- Before we talk about the uncertainty on this number, let’s compare our data spread to the manufacturers claim

DMM Uncertainties

- Recall we quantify the "width" of a distribution with its variance:

$$
\begin{aligned}
\sigma^2 &= E((x-\bar{x})^2) \\
&= E(x^2 - 2\bar{x}x + \bar{x}^2) \\
&= E(x^2) - 2\bar{x}E(x) + E(\bar{x}^2) \\
&= E(x^2) - 2\bar{x}^2 + \bar{x}^2 \\
&= E(x^2) - \bar{x}^2 \\
&= \overline{x^2} - \bar{x}^2
\end{aligned}
$$

- The Standard Deviation is just σ

DMM uncertanties

- For a finite sample of data, the variance (stdev squared) is, as defined above:

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

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

- Now, a real subtlety to see if you have things straight:
- This is the variance of the given sample
  - We want something else – we want an estimator of the variance of the underlying probability distribution from which the sample was drawn
- The variance of the sample is NOT the best estimator of the true variance!

Uncertainties of DMM

- For the *value* of the measurement, the mean of all measurements is the best estimator of the true value
- For the variance, or standard deviation, the right estimator is actually

$$\sigma_{est}^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N}(x_i - \bar{x})^2\right)$$

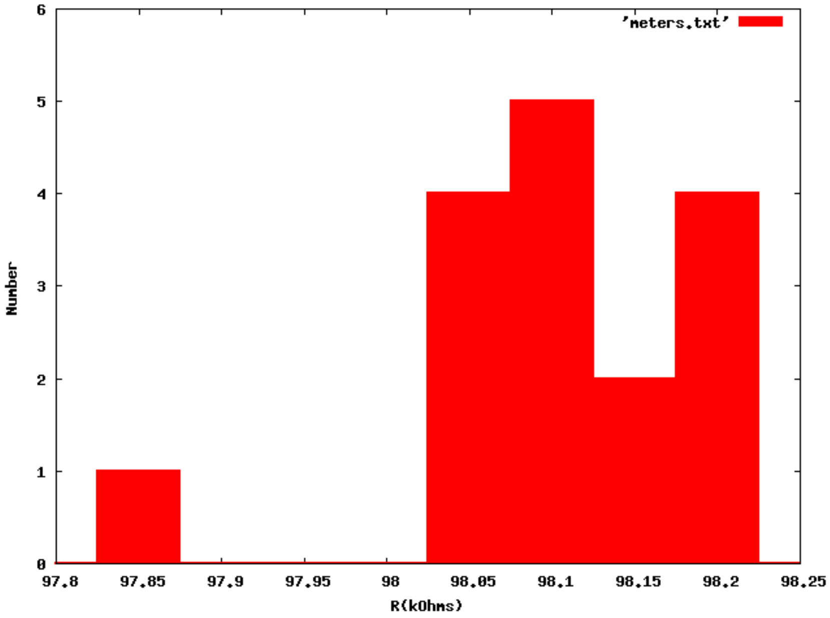
- Note the N-1, vs N
- If you use N, you have what is called a *biased estimator* – the average value of the estimator over many samples is not the true value

- Here is the data again, with the correct estimator of the mean and σ

	R(kOhms)
	98.01
	98.06
	98.10
	98.19
	98.02
	98.16
	97.82
	98.08
	98.01
	98.02
	98.07
	98.18
	98.16
	98.13
	98.11
	98.09
Mean	**98.08**
Stdev	**0.09**

- Our estimate of the true resistance: 98.08 kOhm
- Now let's estimate the uncertainty

DMM uncertainty



R-Mean(kOhm)	(R-Mean)**2
-0.07	0.0043
-0.02	0.0002
0.02	0.0006
0.11	0.0131
-0.06	0.0031
0.08	0.0071
-0.26	0.0653
0.00	0.0000
-0.07	0.0043
-0.06	0.0031
-0.01	0.0000
0.10	0.0109
0.08	0.0071
0.05	0.0030
0.03	0.0012
0.01	0.0002

Kiethley: 0.04%+1d=0.05kOhm

Our estimate dominated by one outlier …

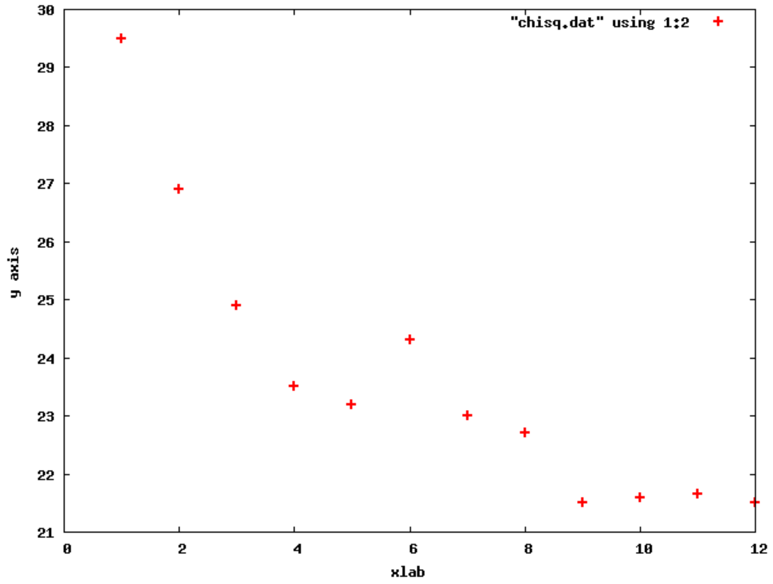
0.1236 SUM
0.0082 SUM/15
0.09 SQRT(kohm)

Fitting with gnuplot

- Latest gnuplot is 4.2
- You should follow a tutorial and make sure you know how to plot a function, plot a data file, etc
- See the web page, help file, net …
- You should feel comfortable plotting a function, plotting data points from a file and fitting the data points

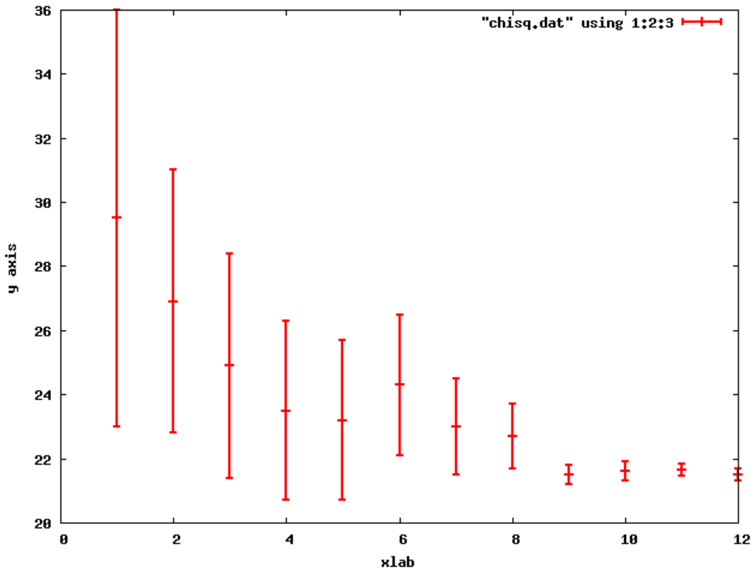
Fitting with gnuplot

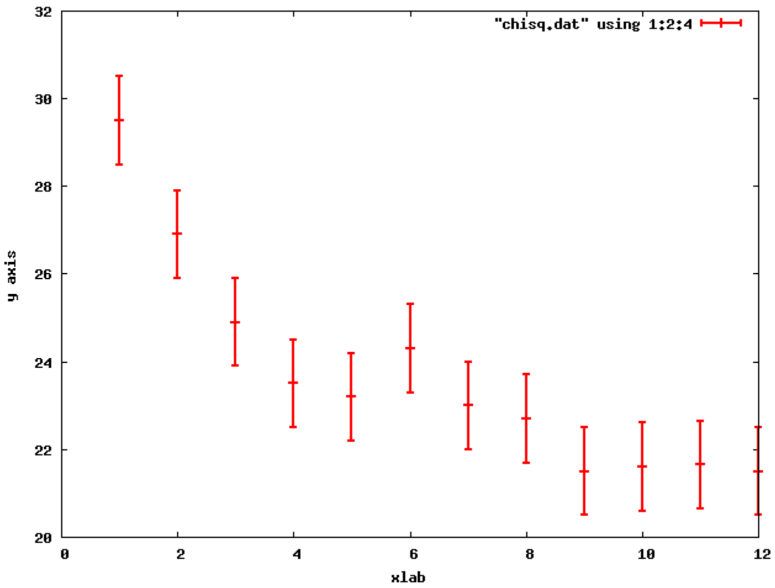
- You will be fitting using the Least-Squares method (aka chi-squared fit)
- Suppose you have a set of data points, and each point's uncertainty $\{x_i, y_i, \sigma_i\}$ in some file "chisq.dat"
- You can plot the data using gnuplot
  - plot "chisq.dat" using 1:2
  - plot "chisq.dat" using 1:2:3



Fitting

- The data points alone don't tell the story
  - File "chisq.dat" has x, y, dy1, dy2, where dy1 uncertainties vary point to point, dy2 are uniform:





- Your best estimate for the true value of the quantity measured will be different in these two cases
  - Quality of measurements must be taken into account!!

Fitting

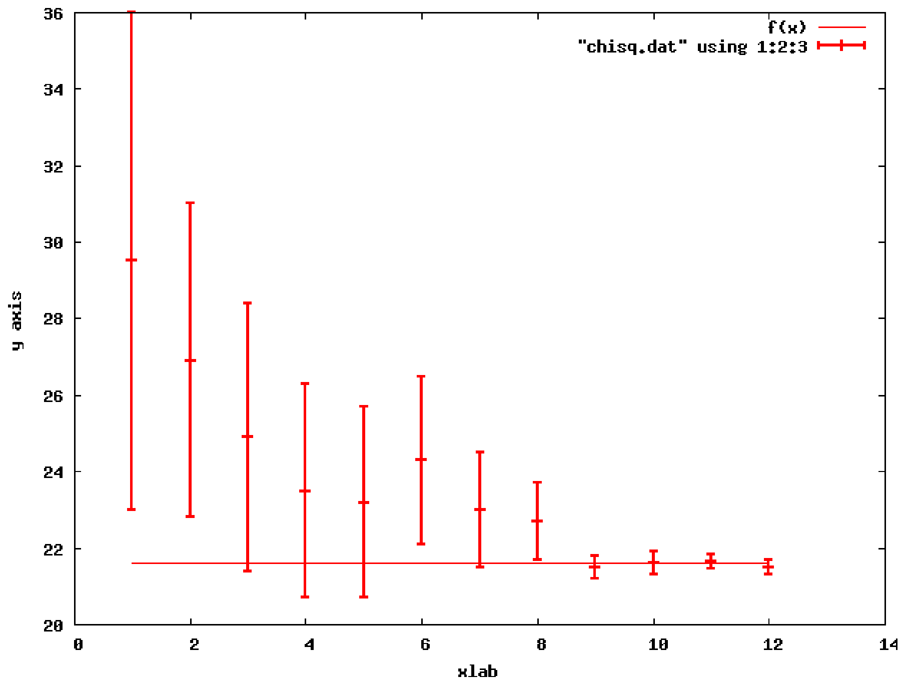
- To fit the data to a model, we postulate a functional form that the data points should follow
  - Eg might be an exponential + constant for the RC lab
- Form the χ^2 (chi squared) as follows:

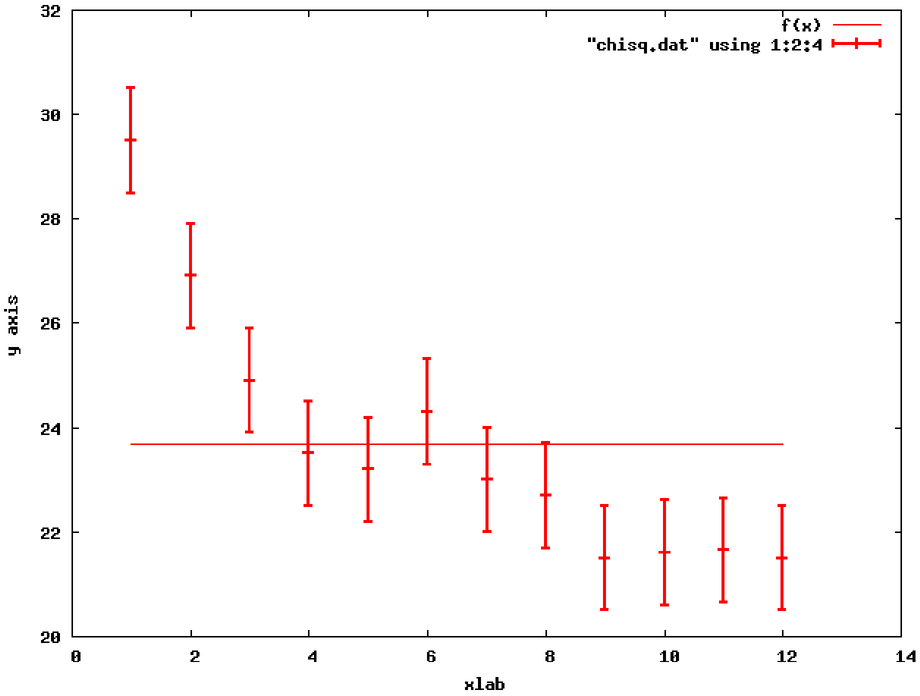
$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i \mid \vec{\theta}))^2}{\sigma_i^2}$$

  - Where f is the value of your function at each point x, and θ is the set of parameters of the function
    - Eg if f is a straight line, f(x)=m*x+b, the parameters are m and b
- That is, take your *measured value – predicted value* at each point. Square these. Divide by the uncertainty on the point's measurements. Add all these up.
- The Least squares principle states that the best estimator for the function's parameters are those that minimize the χ^2

FItting

- Here are fits to a constant, using both uniform errors and non-uniform errors. Note these are different!





Fitting a constant

- Let's do one fit by hand – fit data to a constant
  - This is like finding the mean of the data (weighted by errors).

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{y_i - a}{\sigma_i}\right)^2$$

$f(x) = a$ is our fit function

Find the minimum – set the derivative =0

$$0 = \frac{d\chi^2}{da} = -2\sum_{i=1}^{N}\left(\frac{y_i - a}{\sigma_i^2}\right)$$

$$a\sum_{i=1}^{N}\frac{1}{\sigma_i^2} = \sum_{i=1}^{N}\frac{y_i}{\sigma_i^2}$$

$$a = \frac{\sum_{i=1}^{N}\frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}}$$

Weighted mean formula

Fitting a constant

- Note that if all the uncertainties are the same we get:

$$a = \frac{\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{N} y_i}{\frac{1}{\sigma^2}\sum_{i=1}^{N} 1} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

- i.e. the usual MEAN value. The full formula properly weights the better measurements over the poorer ones.

Degrees of freedom

- What do we expect the “typical” value of chisquared to be for a good fit?
- If the model fits the data, then each term in

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i \mid \vec{\theta}))^2}{\sigma_i^2}$$

Should be about 1 (by definition of the uncertainty \sigma)

- So we expect the sum to be about N
- However, you have to subtract off the number of parameters in the fit function
    - Eg. Straight line: has 2 free parameters, slope+intercept
        - A fit to 2 data points has $\chi^2 = 0$ as a line can always be drawn through 2 points. It has 0 degrees of freedom (dof)

Degrees of freedom

- A fit of a straight line to 3 data points has 1 dof (3 points – 2 fit parameters). There is real information here.
- In general, #dof = N points – M fit parameters
- A “good” fit will have $\frac{\chi^2}{ndof} \approx 1$
- A bad fit will have a large value
- A very small value indicates your errors are overestimated!
  – This is as bad as underestimating errors!