

PHY-4103

Dr. Eram Rizvi & Dr. Alston Misquitta

Lecture 7 - Hypothesis Testing







How well do we know our errors? Why do we not quote results as 8.214 ± 0.132 ? We know σ - what is the error on this quantity? - the error of the error

$$\frac{\sigma_{\sigma}}{\sigma} = \frac{1}{\sqrt{2n}}$$

Depends only on number of measurements n

For n=10 $\frac{\sigma_{\sigma}}{\sigma} \approx 0.22 = 22\%$ For n=50 $\frac{\sigma_{\sigma}}{\sigma} \approx 0.1 = 10\%$

It is difficult to know error better than 10% ! Never quote error to more than 2 sig.figs!



<u>Rule of Thumb:</u> Look at first 3 sig.figs of the error quote 2 sig.figs if < 355 quote 1 sig.fig if > 355

Examples: 0.713 ± 0.059 should be given as 0.71 ± 0.06 12.1749 ± 0.0112 should be given as 12.175 ± 0.011 0.07341 ± 0.0462 should be given as 0.73 ± 0.05

> Keep the precision the same for error and central value i.e. same number of decimal places e.g. <u>never</u> 123.57 ± 2 !!!

Dr Eram Rizvi

Scientific Measurement - Lecture 7



Fitting Data

Scientists confront data with theory If data and theory disagree? Theory is wrong! If data and theory agree? Theory is right

No!

Scientific data can only falsify theories. Theory may be correct, but parameters of theory may be wrong

Strongest statement you can make is that data and theory are consistent

Theory predicts y=f(x)Theory may have a free parameter

$$F = G \frac{m_1 m_2}{r^2}$$

G is not predicted!

Slope of plot F vs $1/r^2$ determines G

F

→ $1/r^2$



What if we measure data like the following?



We fit the data to the theory and determine the best value for G i.e. we extract the best value for a and b in a straight line fit y=ax+b

Dr Eram Rizvi

Scientific Measurement - Lecture 7



Consider n data points (x_i, y_i) with error σ_i Assume error on x $\sigma_i=0$

We calculate residuals of data to theory:

$$r_i = [y_i - f(x_i)]^2$$
$$R = \sum_{i=1}^n r_i = \sum_{i=1}^n [y_i - (ax_i + b)]^2$$

example: theory predicts linear relationship between f and x with two parameters a and b

$$f(x) = ax + b$$

require:

$$\frac{\partial R}{\partial a} = 0$$
 and $\frac{\partial R}{\partial b} = 0$

after some algebra....

$$a = \frac{n \sum_{i} x_{i} y_{i} - \sum_{i} x_{i} \sum_{i} y_{i}}{n \sum_{i} x_{i}^{2} - \left(\sum_{i} x_{i}\right)^{2}}$$

$$b = \frac{n \sum_{i} y_{i} - a \sum_{i} x_{i}}{n}$$

no need to remember computers do it for you!



$\begin{array}{cccccccccccccccccccccccccccccccccccc$	x _i	y _i (±1.5)	r _i (residual)	r _i /σ _i	Errors are ± 1.5 on the y _i measurements
I8.1 $+1.06$ $+0.71$ 29.1 -0.27 -0.18 312.5 $+0.80$ $+0.53$ 414.7 $+0.67$ $+0.45$ 518.9 $+2.54$ $+1.69$ 617.6 -1.09 -0.73 718.9 -2.12 -1.41 824.5 $+1.15$ $+0.77$ 924.9 -0.28 -0.19	0	2.6	-2.11	-1.41	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	I	8.1	+1.06	+0.71	
312.5 $+0.80$ $+0.53$ 414.7 $+0.67$ $+0.45$ calculating: $a=2.327$ 518.9 $+2.54$ $+1.69$ $b=4.708$ 617.6 -1.09 -0.73 thus $y = 4.71 + 2.33 \times$ 718.9 -2.12 -1.41 thus $y = 4.71 + 2.33 \times$ 824.5 $+1.15$ $+0.77$ 924.9 -0.28 -0.19	2	9.1	-0.27	-0.18	
414.7 $+0.67$ $+0.45$ calculating: $a=2.327$ 518.9 $+2.54$ $+1.69$ $b=4.708$ 617.6 -1.09 -0.73 thus $y = 4.71 + 2.33 \times$ 718.9 -2.12 -1.41 thus $y = 4.71 + 2.33 \times$ 824.5 $+1.15$ $+0.77$ 924.9 -0.28 -0.19	3	12.5	+0.80	+0.53	
5 18.9 $+2.54$ $+1.69$ $b=4.708$ 6 17.6 -1.09 -0.73 7 18.9 -2.12 -1.41 8 24.5 $+1.15$ $+0.77$ 9 24.9 -0.28 -0.19	4	14.7	+0.67	+0.45	calculating: a=2.327
6 17.6 -1.09 -0.73 thus y = 4.71 + 2.33 x 7 18.9 -2.12 -1.41 8 24.5 +1.15 +0.77 9 24.9 -0.28 -0.19	5	18.9	+2.54	+1.69	b=4.708
7 18.9 -2.12 -1.41 8 24.5 +1.15 +0.77 9 24.9 -0.28 -0.19	6	17.6	-1.09	-0.73	thus $y = 4.71 \pm 0.33 y$
8 24.5 +1.15 +0.77 9 24.9 -0.28 -0.19	7	18.9	-2.12	-1.41	$u_{105} y = 4.71 + 2.55 x$
9 24.9 -0.28 -0.19	8	24.5	+1.15	+0.77	
	9	24.9	-0.28	-0.19	

residuals: $r_i = y_i - f(x_i)$

This is the deviation of any one measurement from the <u>calculated</u> slope

Often the quality of the fit is stated in terms of the χ^2 test

Think of it as summed residual in units of the error

$\gamma^2 - \sum_{i=1}^{n} \left(\frac{\mathbf{r}_i}{\mathbf{r}_i}\right)^2$	$-\sum_{i=1}^{n} \left(\underline{y}_{i} \right)$	$y_i - f(x_i)$	
$\lambda = \sum_{i=1}^{\infty} \left(\overline{\sigma_i} \right)$	$-\sum_{i=1}$	σ_{i}	Ţ

For a good fit, expect average deviation to be about 1σ , i.e. $\chi^2\approx n~$ i.e. $\chi^2/n\approx 1$

Dr Eram Rizvi

Scientific Measurement - Lecture 7

30.0



This is the data plotted with PhysPlot

Could fit using a polynomial e.g. quadratic with 3 parameters, or cubic with 4 params So, we "penalise" the χ^2 , use no. degrees of freedom (n.d.f.) = n - #params It tells us the χ^2 = 9.24 for 8 degrees of freedom (10 data - 2 params = 8 d.f.) Thus χ^2 /n.d.f. = 1.16 not bad!





In this example the data scatter more: similar straight line fit look at χ^2 /n.d.f. = 35.8 / 8 = 4.48

This means that either the model is wrong (i.e. straight line) or errors too small

Dr Eram Rizvi

Scientific Measurement - Lecture 7



In this example the data scatter much less: similar straight line fit look at χ^2 /n.d.f. = 0.078 / 8 = 0.0098

when χ^2 /n.d.f. is too small, probably means errors are underestimated