Principle of Maximum Likelihood
Weighted Averages
Linear Least Squares Fitting

Lecture # 5

Physics 2BL

Summer 2010

Outline

- Principle of maximum likelihood
- Weighted averages
- Least Squares Fitting
- example

Principle of Maximum Likelihood

- Best estimates of X and σ from N measurements (x_1 - x_N) are those for which $Prob_{X,\sigma}(x_i)$ is a maximum

The Principle of Maximum Likelihood

Recall the probability density for measurements of some quantity x (distributed as a Gaussian with mean X and standard deviation σ)

$$P_{X,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-X)^2}{2\sigma^2}}$$

Normal distribution is one example of $P(x)$.

Now, lets make repeated measurements of x to help reduce our errors.

$$x_1,\ x_2,\ x_3,\ \ldots,\ x_n$$

We define the Likelihood as the product of the probabilities. The larger L, the more likely a set of measurements is.

$$L = P(x_1)P(x_2)P(x_3)\ldots P(x_n)$$

Is L a Probability?

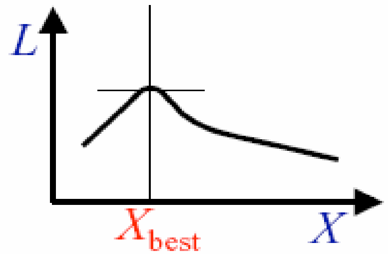
Why does max L give the best estimate?

The best estimate for the parameters of $P(x)$ are those that maximize L.

Yagil

Using the Principle of Maximum Likelihood:

Prove the mean is best estimate of X



Assume X is a parameter of $P(x)$.
When L is maximum, we must have: $\frac{\partial L}{\partial X} = 0$

Lets assume a Normal error distribution and find the formula for the best value for X.

$$L = P(x_1)P(x_2)\ldots P(x_n) = \prod_{i=1}^{n} P(x_i)$$

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - X)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\sum_{i=1}^{n}\frac{(x_i - X)^2}{2\sigma^2}}$$

$$L = Ce^{-\chi^2/2}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - X)^2}{\sigma^2}$$

Defininition

$$\frac{\partial L}{\partial X} = 0 = Ce^{-\chi^2/2} \frac{-1}{2}\frac{\partial \chi^2}{\partial X}$$

$$\Rightarrow \quad \frac{\partial \chi^2}{\partial X} = 0$$

$$\frac{\partial \chi^2}{\partial X} = \frac{1}{\sigma^2}\sum_{i=1}^{n} 2(x_i - X) = 0$$

$$\sum_{i=1}^{n}(x_i - X) = 0$$

$$\sum_{i=1}^{n} x_i - nX = 0$$

$$X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Q.E.D.

the mean

What is the Error on the Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Formula for mean of measurements. (We just proved that this is the best estimate of the true x.)

Now, use propagation of errors to get the error on the mean.

$$\sigma_{\bar{x}} = \frac{\partial \bar{x}}{\partial x_1}\sigma_{x_1} \oplus \frac{\partial \bar{x}}{\partial x_2}\sigma_{x_2} \oplus \ldots \oplus \frac{\partial \bar{x}}{\partial x_n}\sigma_{x_n}$$

$$\frac{\partial \bar{x}}{\partial x_i} = \frac{1}{n}$$

$$\sigma_{\bar{x}} = \sqrt{\sum_{i=1}^{n}\left(\frac{\sigma_{x_i}}{n}\right)^2} = \sqrt{n\left(\frac{\sigma}{n}\right)^2} = \frac{\sigma}{\sqrt{n}}$$

What would you do if the x_i had different errors?

We got the error on the mean (SDOM) by propagating errors.

Weighted averages (Chapter 7)

We can use maximum Likelihood (χ^2) to average measurements with different errors.

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - X}{\sigma_i} \right)^2$$

We derived the result that:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Using error propagation, we can determine the error on the weighted mean:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{\sum_{i=1}^{n} w_i}}$$

What does this give in the limit where all errors are equal?

$$\frac{\partial \chi^2}{\partial X} = 0 = -2\sum_{i=1}^{n} \frac{x_i - X}{\sigma_i^2}$$

$$\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} - X \sum_{i=1}^{n} \frac{1}{\sigma_i^2} = 0$$

$$w_i \equiv \frac{1}{\sigma_i^2}$$

$$\sum_{i=1}^{n} w_i x_i = X \sum_{i=1}^{n} w_i$$

$$X = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Example: Weighted Average

Suppose 2 students measure the radius of Neptune.

- Student A gets r=80 Mm with an error of 10 Mm and
- Student B gets r=60 Mm with an error of 3 Mm

What is the best estimate of the true radius?

$$\bar{r} = \frac{w_A r_A + w_B r_B}{w_A + w_B} = \frac{\frac{1}{100}80 + \frac{1}{9}60}{\frac{1}{100} + \frac{1}{9}} = 61.65 \text{ Mm}$$

What does this tell us about the importance of error estimates?

<u>Example</u>:

Compatibility of measurements
Best estimate, Weighted Average

Two measurements of the speed of sound give the answers:

u_A=332 ± 1 and u_B=339 ± 3

(Both in m/s.)

- a) Are these measurements consistent at the 5% level? At the 1% level?
- b) What is your best estimate for the speed of sound and its uncertainty?

a) To check if the two measurements are consistent, we compute:

$$q = u_A - u_B = 339-332 = 7 \text{ m/s}$$

and: $$\sigma_q = \sqrt{\sigma^2_{uA} + \sigma^2_{uB}} = 3.16 \text{ m/s}$$

so that: $$t = \frac{q}{\sigma_q} = \frac{339-332}{3.16} = 2.21$$

From Table A we get that 2.21 sigma corresponds to: 97.21%

Therefore the probability to get a worse result is 1-97% ~3%.

The results are not consistent at the 5% and are consistent at the 1% confidence level.

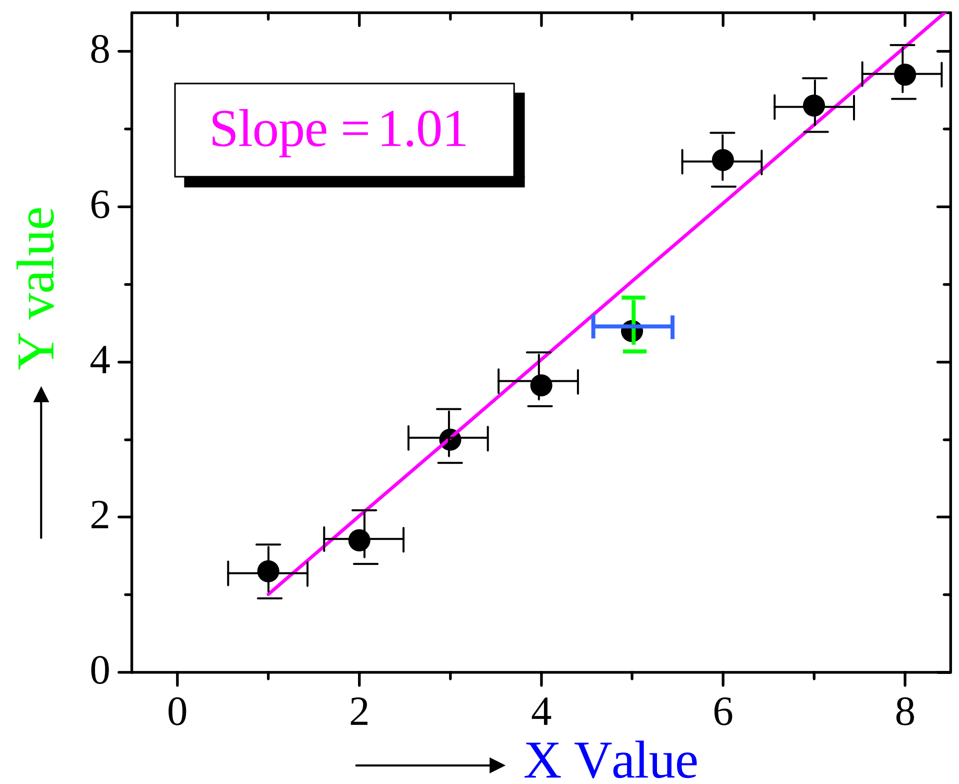
b) Best estimate is the weighted mean:

$$\bar{u} = \frac{w_A u_A + w_B u_B}{w_A + w_B} = \frac{\frac{1}{1}332 + \frac{1}{9}339}{\frac{1}{1} + \frac{1}{9}} = 332.7 \text{ m/s}$$

$$\sigma_{\bar{u}} = \frac{1}{\sqrt{1/w_A + 1/w_B}} = \frac{1}{\sqrt{1/1 + 1/9}} = 0.9 m/s$$

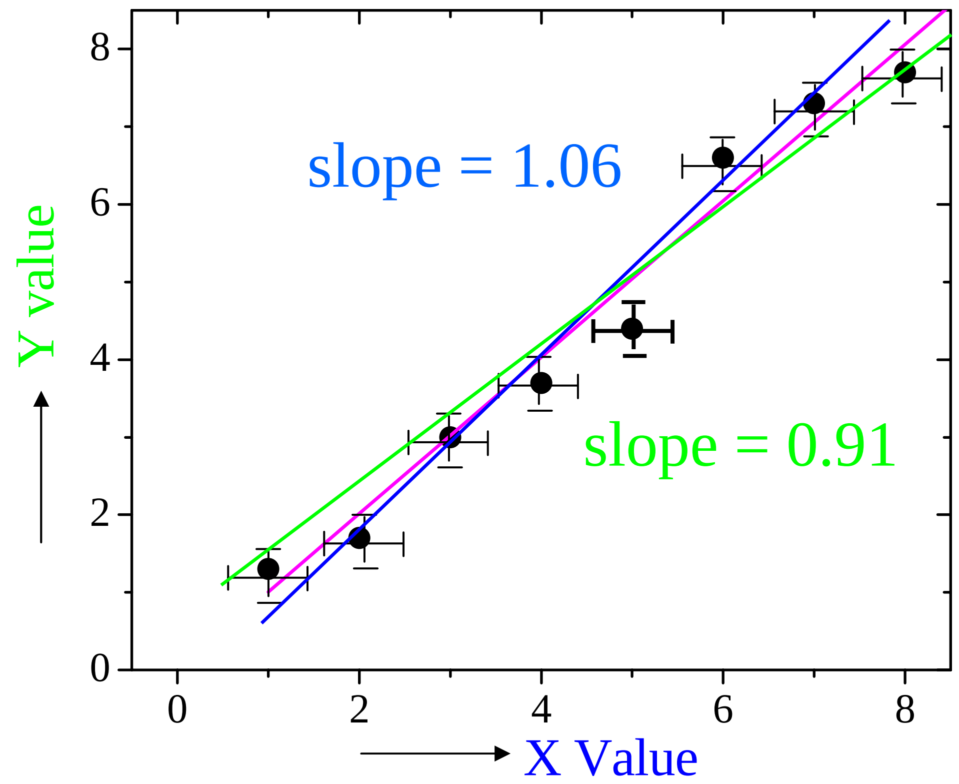
Linear Relationships: $y = A + Bx$ (Chapter 8)

- Data would lie on a straight line, except for errors
- What is 'best' line through the points?
- What is uncertainty in constants?
- How well does the relationship describe the data?



A Rough Cut

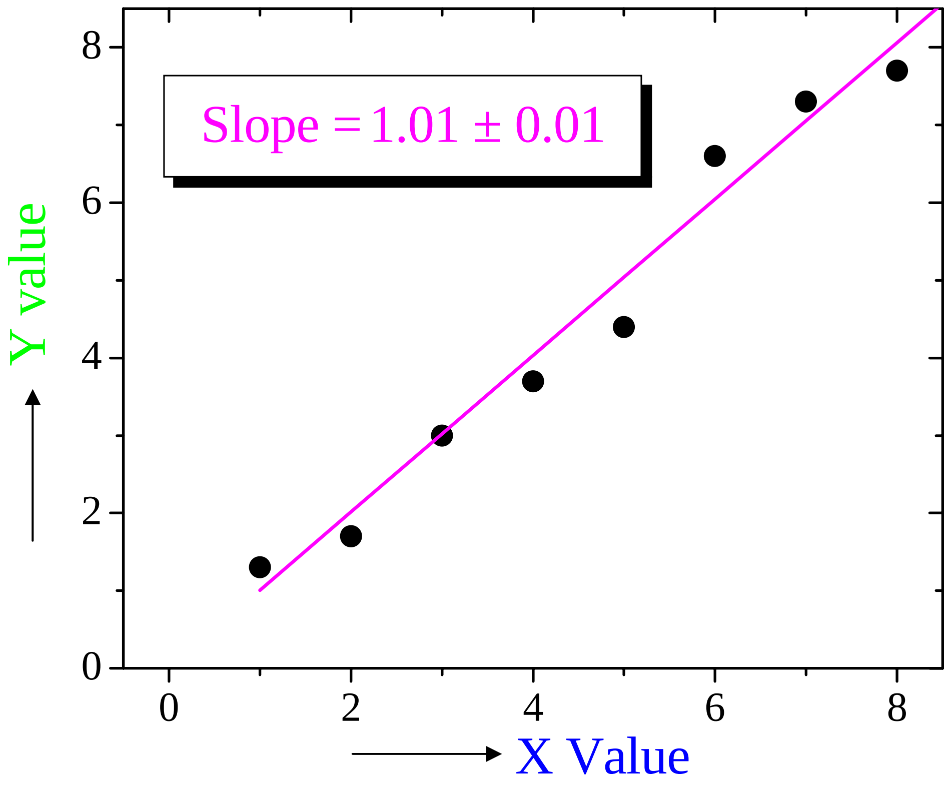
- Best means ‘line close to all points’
- Draw various lines that pass through data points
- Estimate error in constants from range of values
- Good fit if points within error bars of line



slope = 1.01 ± 0.07

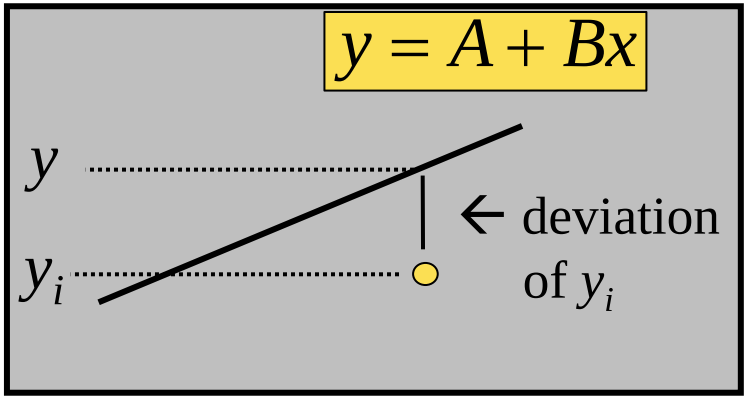
More Analytical

- Best means ‘minimize the square of the deviations between line and points’
- Can use error analysis to find constants, error



The Details of How to Do This (Chapter 8)

- Want to find A, B that minimize difference between data and line
- Since line above some data, below other, minimize sum of squares of deviations
- Find A, B that minimize this sum



$$y_i - y = y_i - A - Bx_i$$

$$\sum_{i=1}^{N} (y_i - A - Bx_i)^2$$

$$\frac{\partial}{\partial A} = \sum y_i - AN - B\sum x_i = 0$$

$$\frac{\partial}{\partial B} = \sum x_i y_i - A\sum x_i + B\sum x_i^{\,2} = 0$$

Finding *A* and *B*

- After minimization, solve equations for *A* and *B*
- Looks nasty, not so bad…
- See Taylor, example 8.1

$$\frac{\partial}{\partial A} = \sum y_i - AN - B\sum x_i = 0$$

$$\frac{\partial}{\partial B} = \sum x_i y_i - A\sum x_i + B\sum x_i^2 = 0$$

$$A = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{\Delta}$$

$$B = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\Delta}$$

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

Uncertainty in Measurements of *y*

- Before, measure several times and take standard deviation as error in *y*
- Can't now, since y_i's are different quantities
- Instead, find standard deviation of deviations

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

$$\sigma_y = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - A - Bx_i)^2}$$

Uncertainty in *A* and *B*

- *A*, *B* are calculated from x_i, y_i
- Know error in x_i, y_i ; use error propagation to find error in *A*, *B*
- A distant extrapolation will be subject to large uncertainty

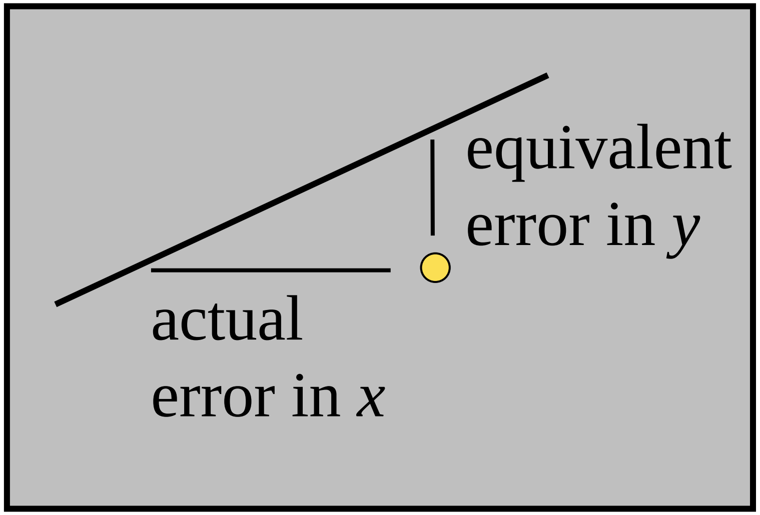
$$\sigma_A = \sigma_y \sqrt{\frac{\sum x_i^2}{\Delta}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

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

Uncertainty in x

- So far, assumed negligible uncertainty in x
- If uncertainty in x, not y, just switch them
- If uncertainty in both, convert error in x to error in y, then add errors



$$\Delta y = B \Delta x$$

$$\sigma_y(equiv) = B\sigma_x$$

$$\sigma_y(equiv) = \sqrt{\sigma_y{}^2 + \left(B\sigma_x\right)^2}$$

Other Functions

- Convert to linear
- Can now use least squares fitting to get ln *A* and *B*

$$y = Ae^{Bx}$$

$$\ln y = \ln A + Bx$$

How to minimize uncertainty

- Can always estimate uncertainty, but equally important is to minimize it
- How can you reduce uncertainty in measurements
  - Use better components
  - Have both partners read value – demand consistency
  - Measure multiple times

Weighted averages

- $X = x = \dfrac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$ where $w_i = \dfrac{1}{\sigma_i^2}$

- $\sigma_{wav} = \dfrac{1}{\sqrt{\sum_{i=1}^{n} w_i}}$

Example problem

Measure wavelength λ four times:

503 ± 10 nm

491 ± 8 nm

525 ± 20 nm

570 ± 40 nm

What is the weighted average?

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

$$\lambda_{wav} = \frac{[(1/100)*503 + (1/64)*491 + (1/400)*525 + (1/1600)*570]}{[(1/100) + (1/64) + (1/400) + (1/1600)]} \text{ nm}$$

$$\lambda_{wav} = 499.8 \text{ nm}$$

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{\sum_{i=1}^{n} w_i}}$$

$$\sigma_\lambda = \frac{1}{\sqrt{[(1/100) + (1/64) + (1/400) + (1/1600)]}} \text{ nm} = 6 \text{ nm}$$

$$\lambda_{wav} = (500 \pm 6) \text{ nm}$$

Example problem

Measure wavelength λ four times:

503 ± 10 nm

491 ± 8 nm

525 ± 20 nm

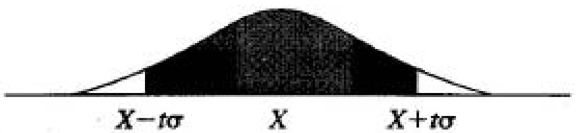
570 ± 40 nm

Should we reject the last data point?

$$t_{sus} = \Delta\lambda = \frac{|570 - 500|\ \text{nm}}{\sqrt{6^2 + 40^2}\ \text{nm}} = 1.73\ \sigma$$

Prob of λ outside $\Delta\lambda$ =

Table A. The percentage probability, $Prob(\text{within } t\sigma) = \int_{X-t\sigma}^{X+t\sigma} G_{X,\sigma}(x)\,dx$, as a function of t.



t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00	0.80	1.60	2.39	3.19	3.99	4.78	5.58	6.38	7.17
0.1	7.97	8.76	9.55	10.34	11.13	11.92	12.71	13.50	14.28	15.07
0.2	15.85	16.63	17.41	18.19	18.97	19.74	20.51	21.28	22.05	22.82
0.3	23.58	24.34	25.10	25.86	26.61	27.37	28.12	28.86	29.61	30.35
0.4	31.08	31.82	32.55	33.28	34.01	34.73	35.45	36.16	36.88	37.59
0.5	38.29	38.99	39.69	40.39	41.08	41.77	42.45	43.13	43.81	44.48
0.6	45.15	45.81	46.47	47.13	47.78	48.43	49.07	49.71	50.35	50.98
0.7	51.61	52.23	52.85	53.46	54.07	54.67	55.27	55.87	56.46	57.05
0.8	57.63	58.21	58.78	59.35	59.91	60.47	61.02	61.57	62.11	62.65
0.9	63.19	63.72	64.24	64.76	65.28	65.79	66.29	66.80	67.29	67.78
1.0	68.27	68.75	69.23	69.70	70.17	70.63	71.09	71.54	71.99	72.43
1.1	72.87	73.30	73.73	74.15	74.57	74.99	75.40	75.80	76.20	76.60
1.2	76.99	77.37	77.75	78.13	78.50	78.87	79.23	79.59	79.95	80.29
1.3	80.64	80.98	81.32	81.65	81.98	82.30	82.62	82.93	83.24	83.55
1.4	83.85	84.15	84.44	84.73	85.01	85.29	85.57	85.84	86.11	86.38
1.5	86.64	86.90	87.15	87.40	87.64	87.89	88.12	88.36	88.59	88.82
1.6	89.04	89.26	89.48	89.69	89.90	90.11	90.31	90.51	90.70	90.90
1.7	91.09	91.27	91.46	91.64	91.81	91.99	92.16	92.33	92.49	92.65
1.8	92.81	92.97	93.12	93.28	93.42	93.57	93.71	93.85	93.99	94.12
1.9	94.26	94.39	94.51	94.64	94.76	94.88	95.00	95.12	95.23	95.34
2.0	95.45	95.56	95.66	95.76	95.86	95.96	96.06	96.15	96.25	96.34
2.1	96.43	96.51	96.60	96.68	96.76	96.84	96.92	97.00	97.07	97.15
2.2	97.22	97.29	97.36	97.43	97.49	97.56	97.62	97.68	97.74	97.80
2.3	97.86	97.91	97.97	98.02	98.07	98.12	98.17	98.22	98.27	98.32

Example problem

Measure wavelength λ four times:

503 ± 10 nm

491 ± 8 nm

525 ± 20 nm

570 ± 40 nm

Should we reject the last data point?

$$t_{sus} = \Delta\lambda = \frac{|570 - 500| \text{ nm}}{\sqrt{6^2 + 40^2} \text{ nm}} = 1.73\,\sigma$$

Prob of λ outside $\Delta\lambda$ = 100 % - 91.6 % = 8.4 %

Total Prob = N x Prob = 4 * 8.4 % = 33.6 %

Is Total Prob < 50 % ? Yes, therefore can reject data point

Remember

- Finish Lab Writeup for Exp. 2
- Read next week’s lab description, prepare
- Homework Taylor #6.4, 7.2, 8.6, 8.10
- Read Taylor through Chapter 9