Physics 1140 Fall 2013
Introduction to Experimental Physics

Joanna Atkin

Lecture 5:

Recap of Error Propagation and Gaussian Statistics

Graphs and linear fitting

Experimental analysis

- Typically make repeat measurements of a quantity to estimate the uncertainty

1. Take many measurements of quantity.
2. Calculate the sample average. $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$
3. Calculate the sample standard deviation. $\sigma_x = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}$
4. Calculate the standard error on the mean. $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$
5. Report measurement in standard format.
   - One significant figure for uncertainty (or two, if first digit is a 1) and value rounded to same number of decimal places as uncertainty.

$$\bar{x} \pm \sigma_{\bar{x}} \text{ [units]}$$

Experimental analysis

- Derived quantity based on measured quantities, e.g. $T \approx 2\pi\sqrt{\frac{L}{g}}$
- Measure L and T and estimated associated uncertainty.
- Calculate uncertainty in derived quantity (e.g. g) using master equation of error propagation:

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 \sigma_z^2 + \ldots}$$

- Works for any function!
  - But all uncertainties must be independent/uncorrelated

Experimental analysis

- Discussion / Conclusion
  - Your own words!
  - what was measured
  - did it meet up to expectations – always compare with theory, accepted value, or a value measured in another way, and report the discrepancy and the significance of the discrepancy
  - what were the sources of error – random, systematic
  - how could experiment be improved.

- Look at the rubric posted on the course websites, and sample lab reports for more feedback about how to structure the lab report.

Clicker Question 1

Which of the following will get "better" or decrease with an increase of the number of measurements?

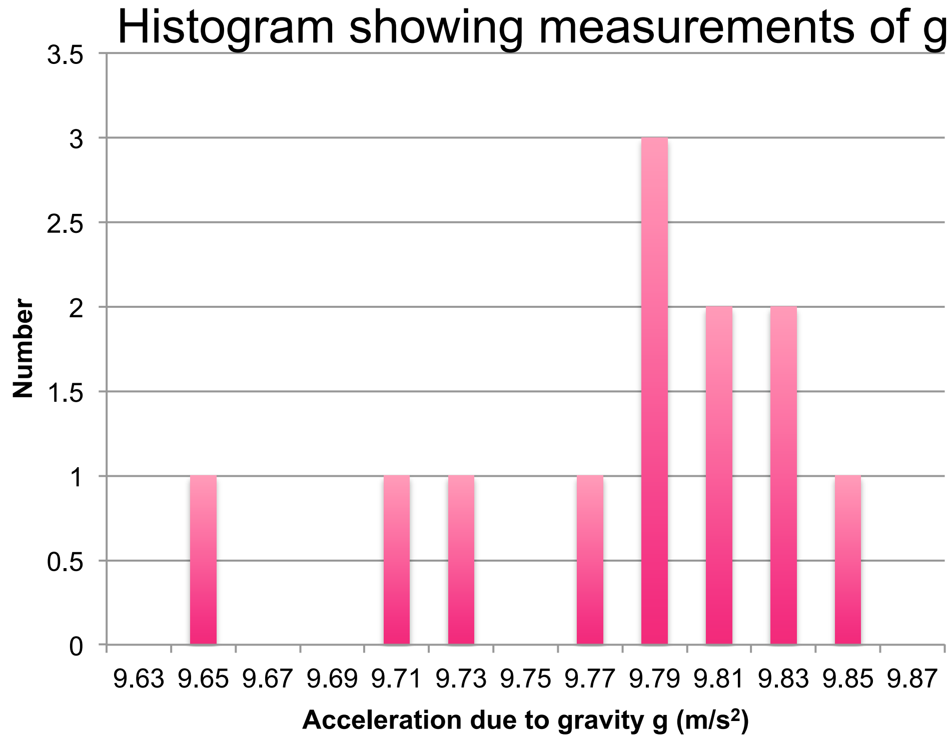
A. The Standard Deviation

→ B. The Standard Error on the Mean

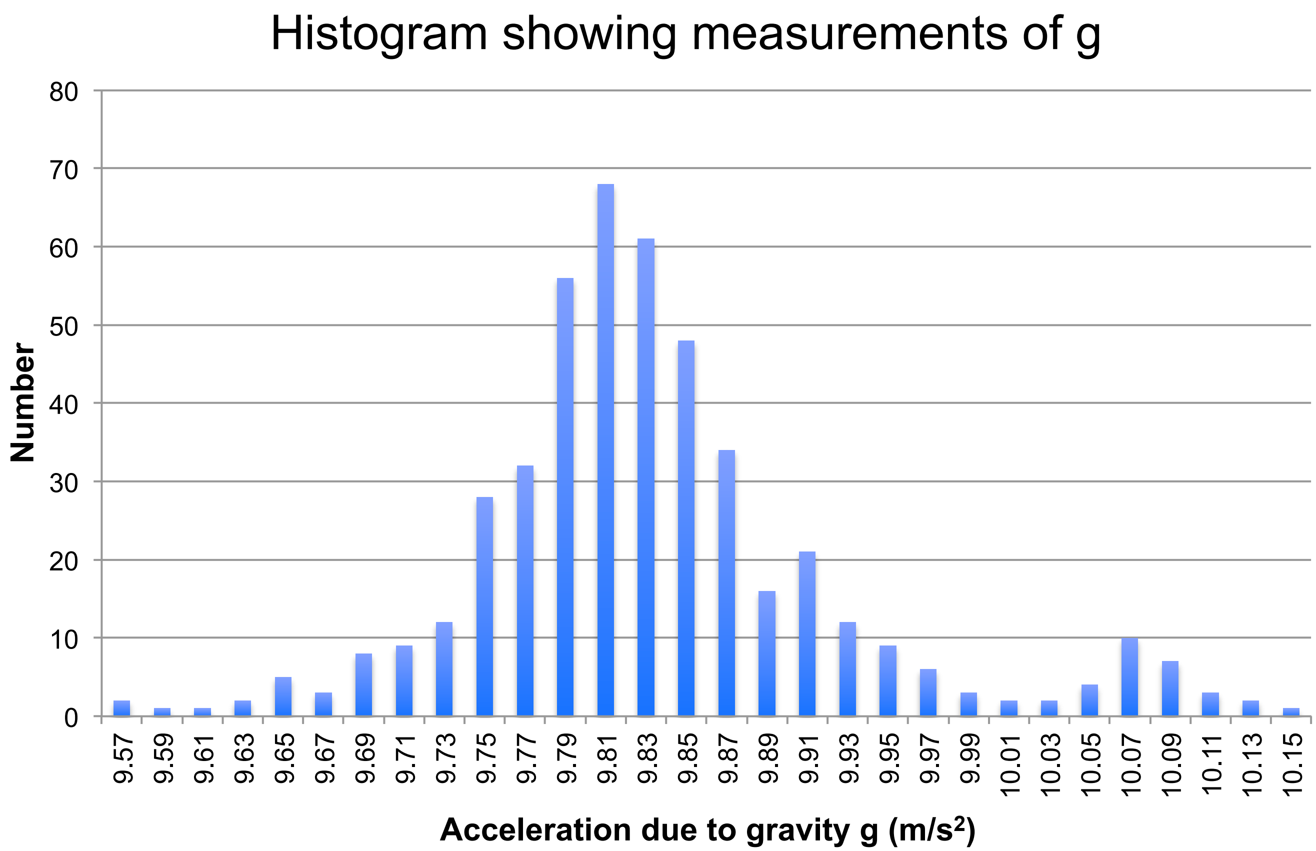
Gaussian distribution

- If we make measurements of the same quantity using the same method
- If all measurements are uncorrelated/random

- The measurements distribute themselves according to the Normal or Gaussian distribution

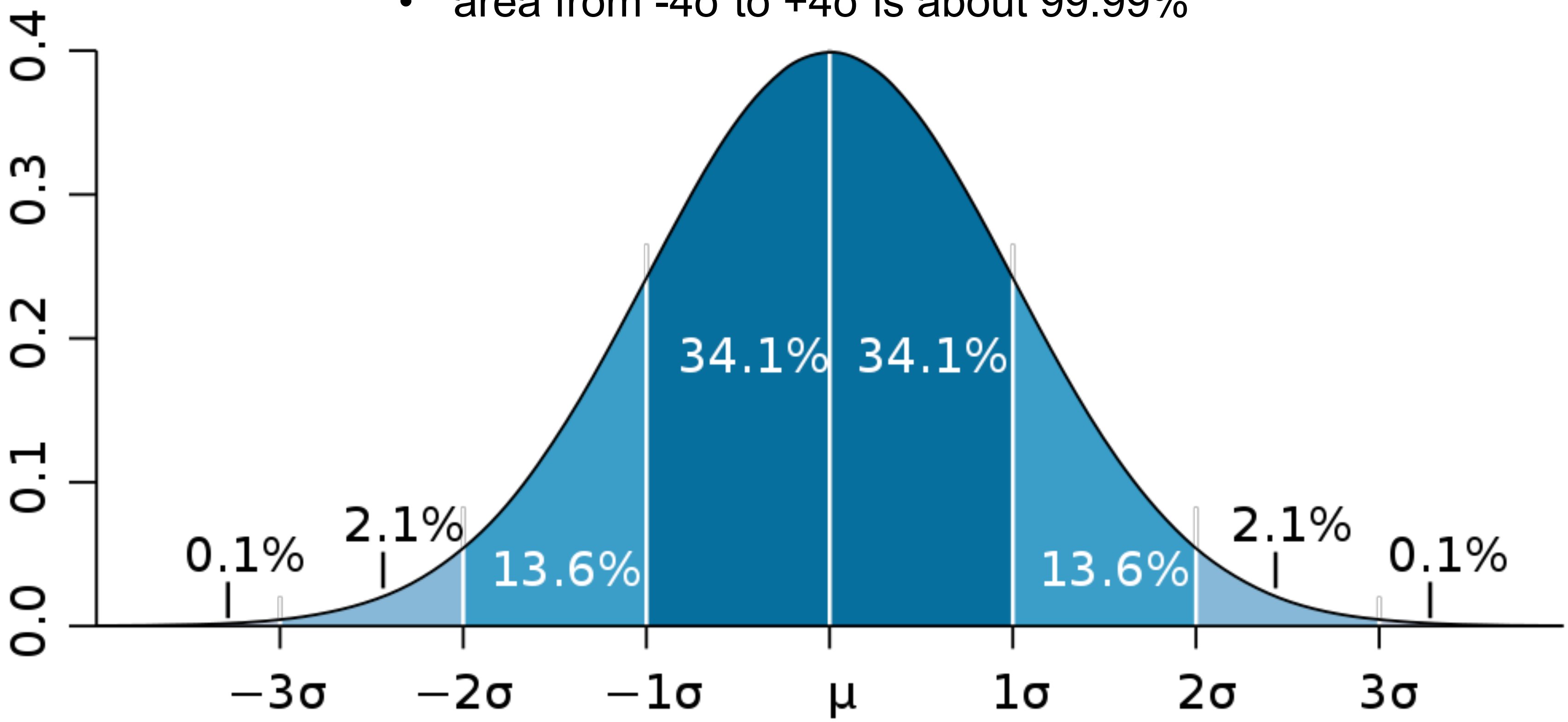


Large number of measurements



Significance and the Gaussian distribution

- area from -∞ to +∞ is 1
- area from $-\sigma$ to $+\sigma$ is about 68%
- area from -2σ to $+2\sigma$ is about 95%
- area from -3σ to $+3\sigma$ is about 99.7%
- area from -4σ to $+4\sigma$ is about 99.99%

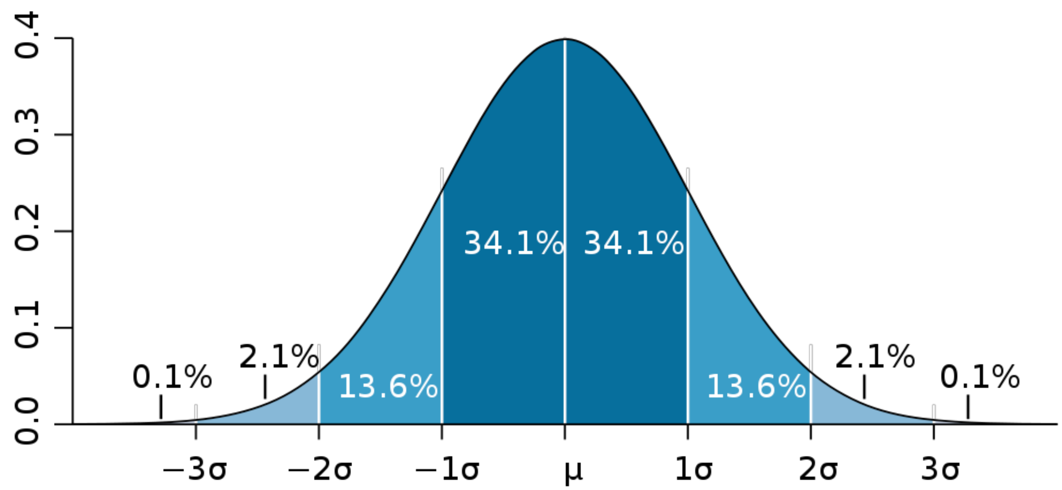


- Keep these values in mind as you write the conclusions of your reports.
- A discrepancy is significant if the probability to deviate that much by chance is small.

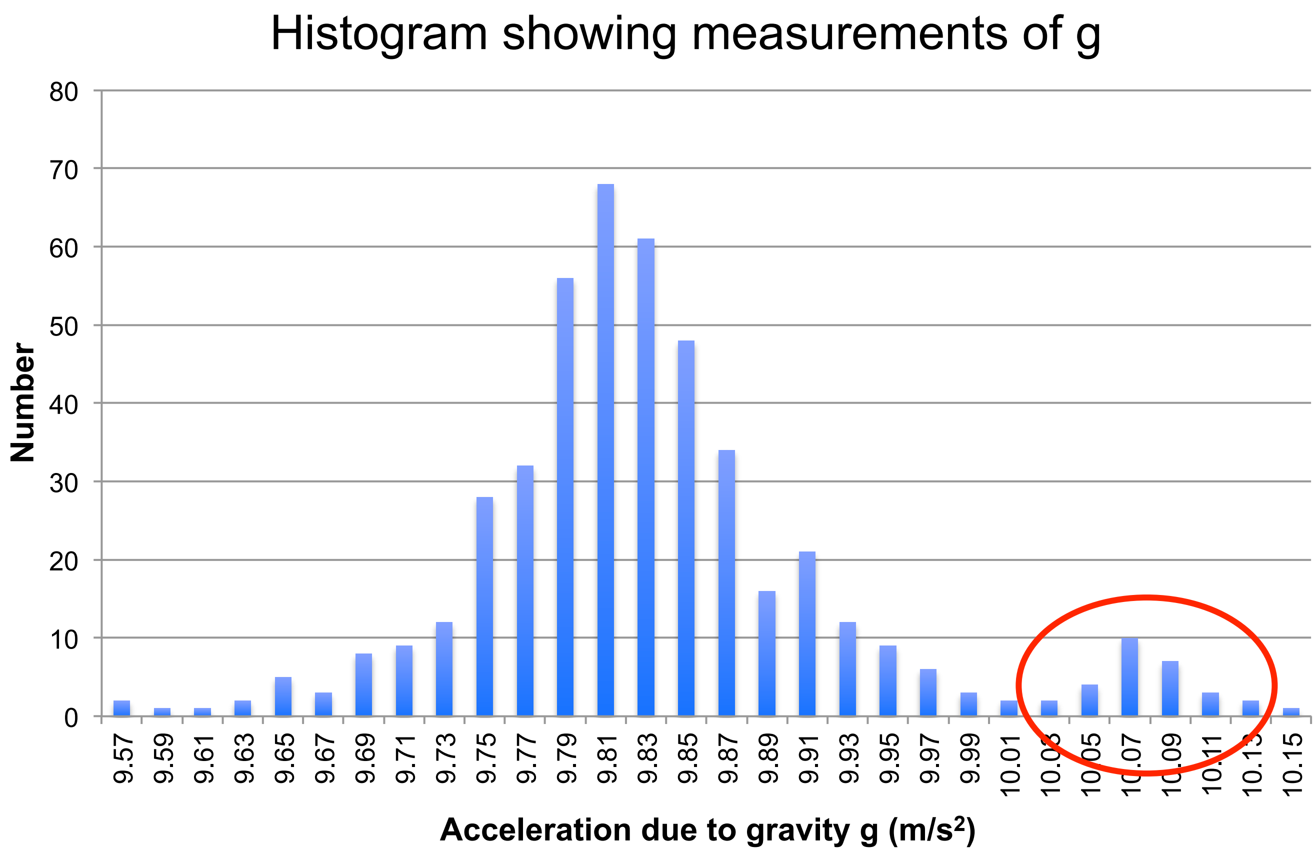
Clicker question 2

In a distant foreign country, the distribution of heights of adult males is found to be Gaussian with a mean of 5'10'' and a standard deviation 3''. What fraction of males are taller than 6'4''?

A) less than 1%
B) about 5%
C) about 16%
D) about 2.5%
E) about 20%

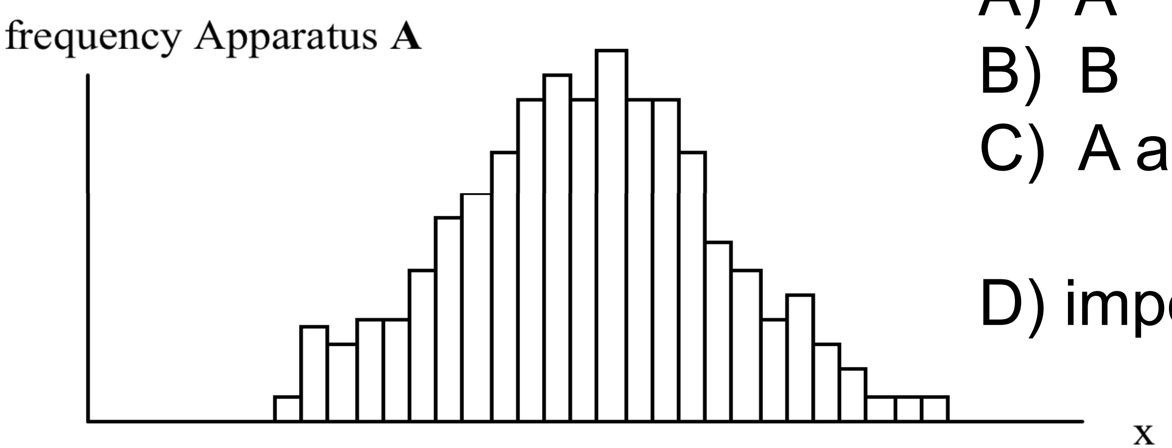


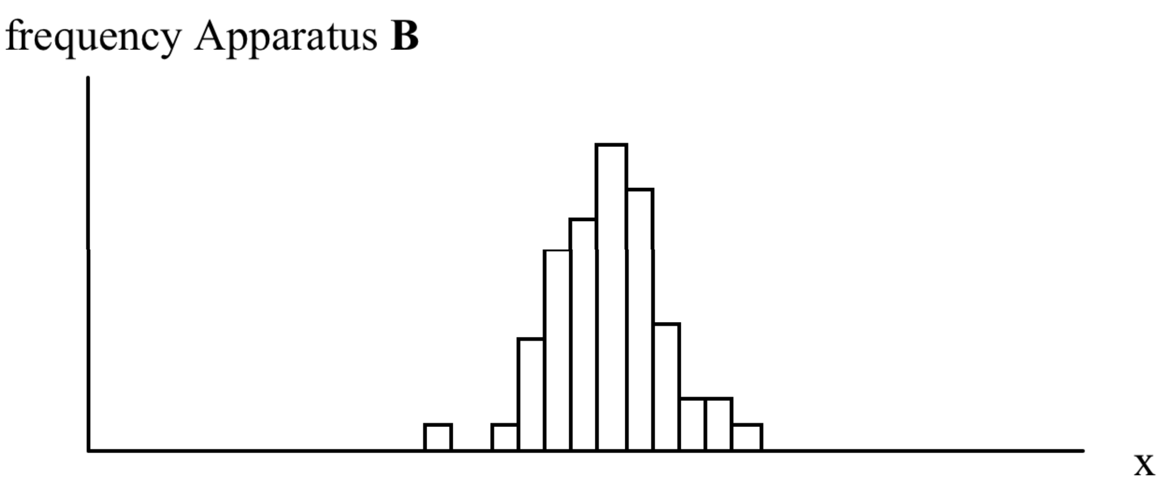
Random and systematic uncertainty



Clicker question 3

Which series of measurements has the smaller standard deviation σ?





A) A
B) B
C) A and B have the same standard deviation.
D) impossible to tell from the information given.

Graphical methods

Often you will take a series of measurements where
the input parameters are varied.

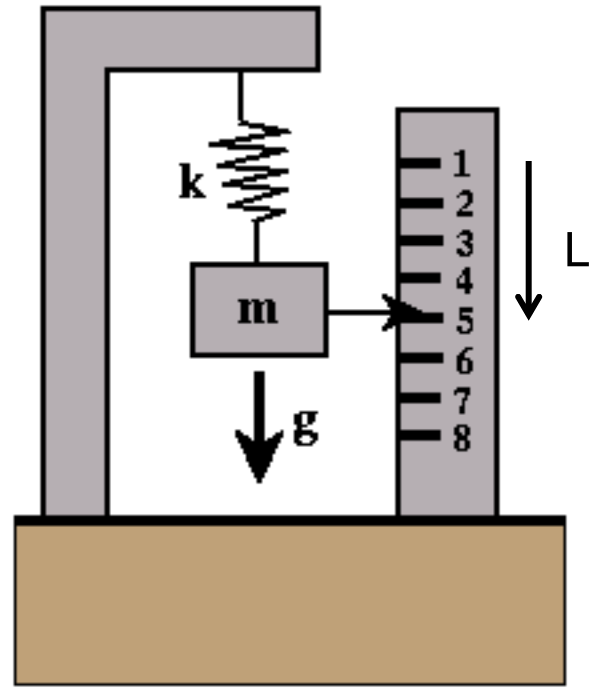
- Example: measure a spring constant:

$$F = mg = kL$$

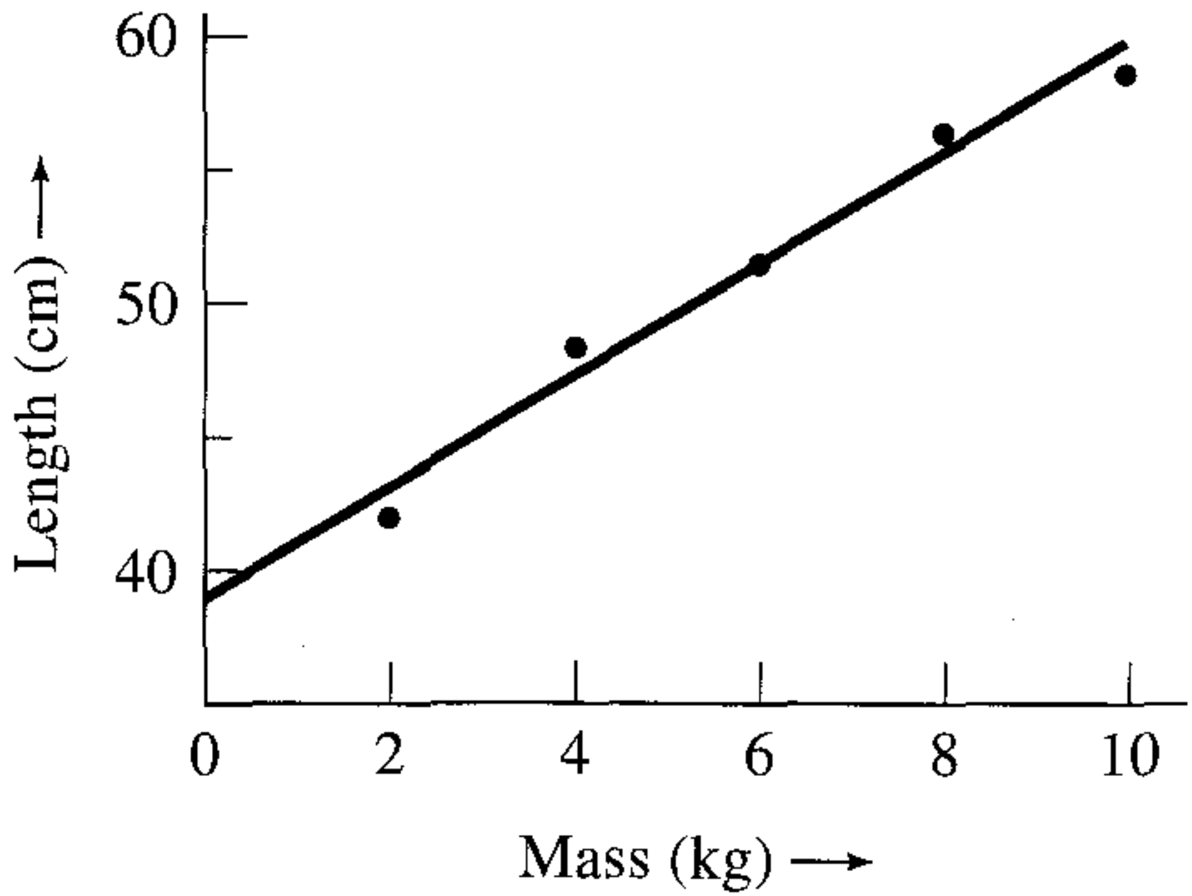
– measure many values of m and x

$$L = \frac{g}{k} m \text{ or } L = A + B \cdot m$$

– Where $B=g/k$ and A is the constant "zero" displacement



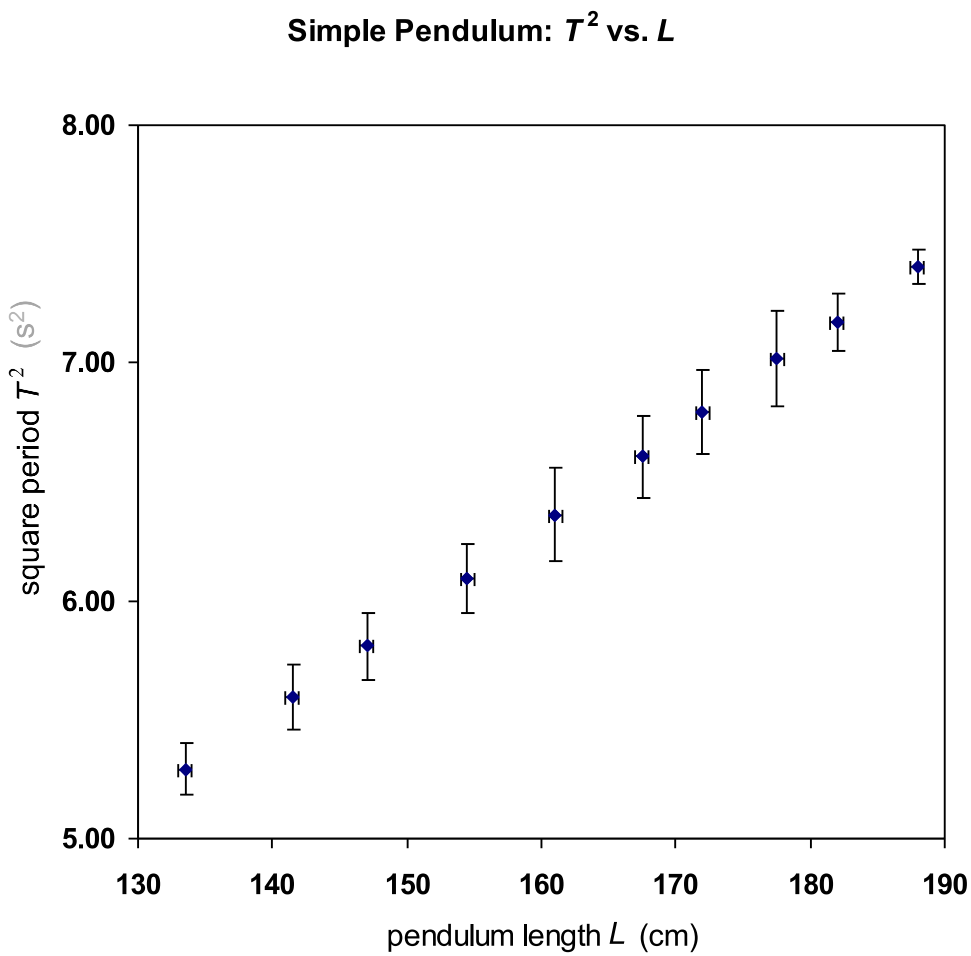
Linear relationship



Legend, axis labels, units, captions...

Graphical methods

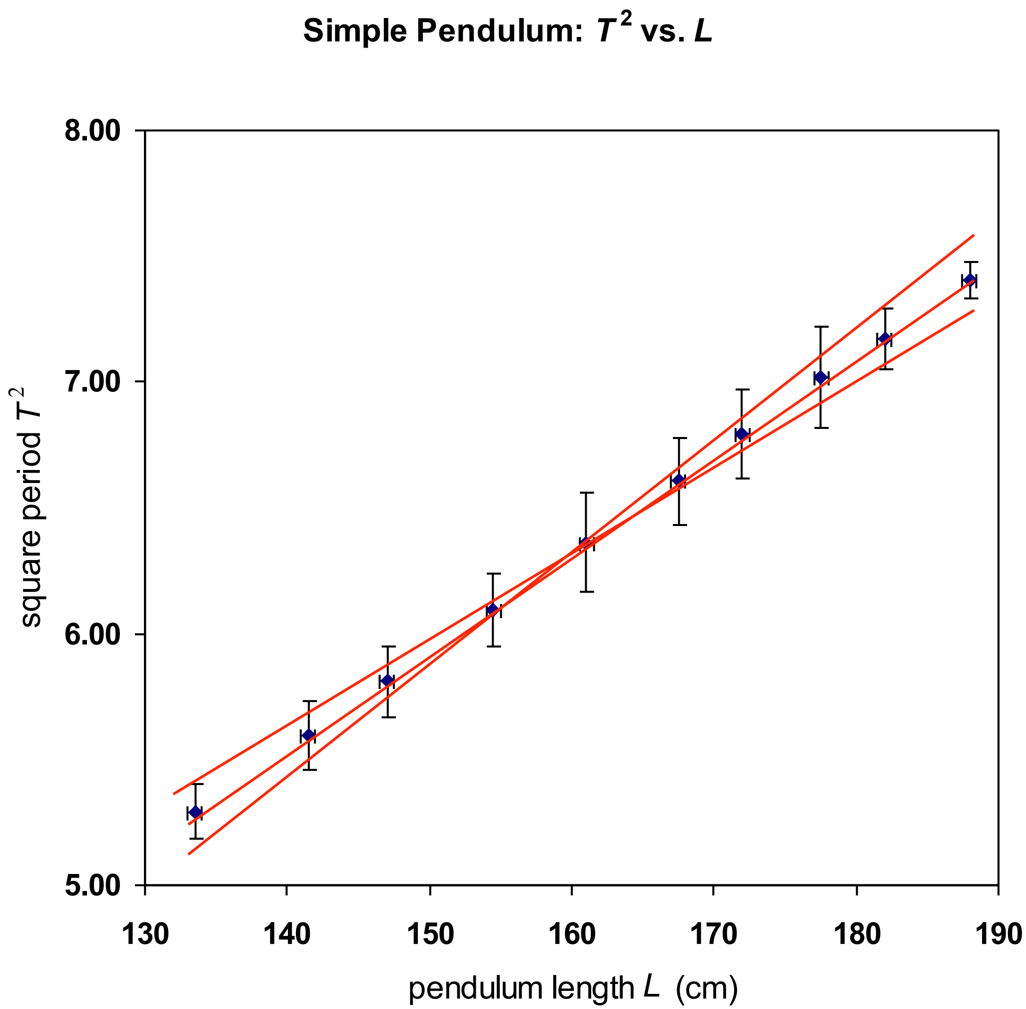
- When plotting data, you want to somehow show that there are uncertainties on your data points.
- Convention: show the 1σ errors on each data point using "error bars."



Fitting

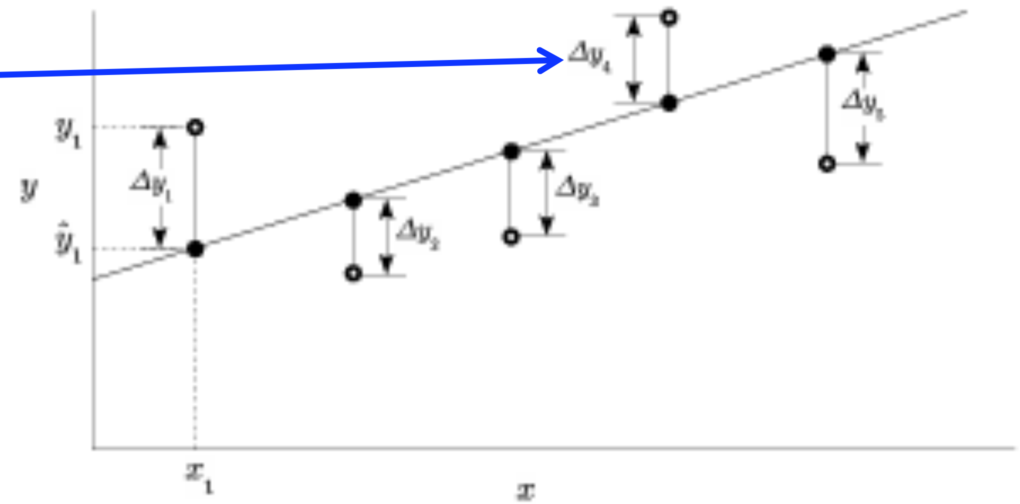
$$y = A + B \cdot x$$

- Fitting: find A and B that create the "best" line through the data.
- But, how do we define a line as the "best" fit?



Best fit line: least squares minimization

This difference, summed in quadrature for each point



- Least squares method: minimize the sum of the distance between each point and the line of best fit.
- The distances Δy are called the residuals of the fit.

$$\Delta y_i = y_i - \hat{y}_i \text{, where}$$

$$\hat{y}_i = A + Bx_i$$

Least squares fitting

$y = A + B \cdot x$ - want to find A and B that best predict y for a given x.

$$\text{total error} = \sqrt{\sum_i (y_i - (A + Bx_i))^2}$$

- Want to minimize total error
  - set derivative = 0!

$$\frac{\partial}{\partial A}\left(\sum_i (y_i - (A + Bx_i))^2\right) = 0$$

$$\frac{\partial}{\partial B}\left(\sum_i (y_i - (A + Bx_i))^2\right) = 0$$

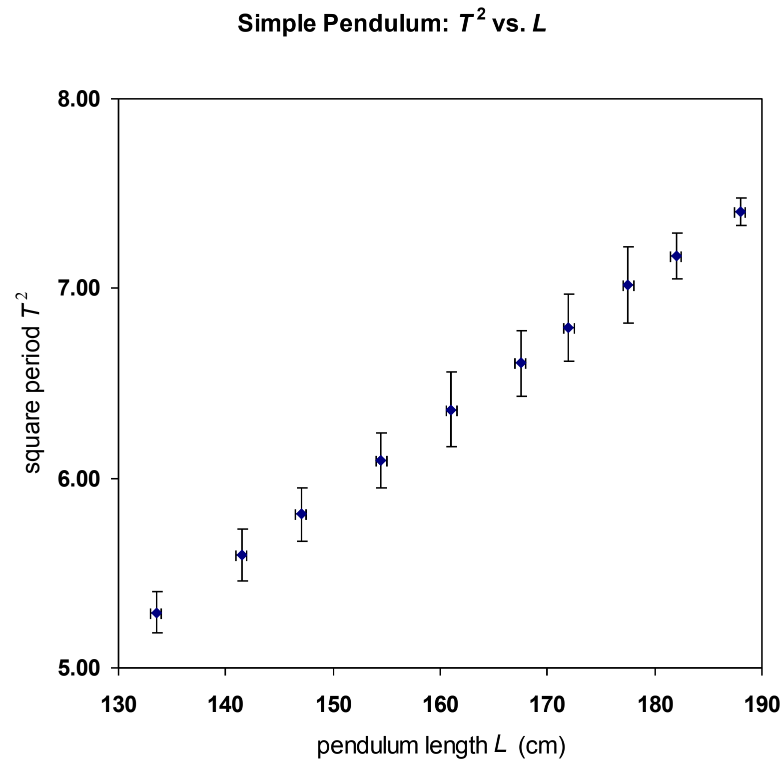
Result: get an estimate for A, B, σ_A, σ_B.

The formulae for A, B and their uncertainties are rather complicated, but very regular and easy for computers

Fitting

- Mathematica code for doing linear fits provided: Linfit.nb
  - calculates A, B, σ_A, σ_B.

- Can have extensions to least squares fitting to take into account points with different errors: weighted least squares fit.

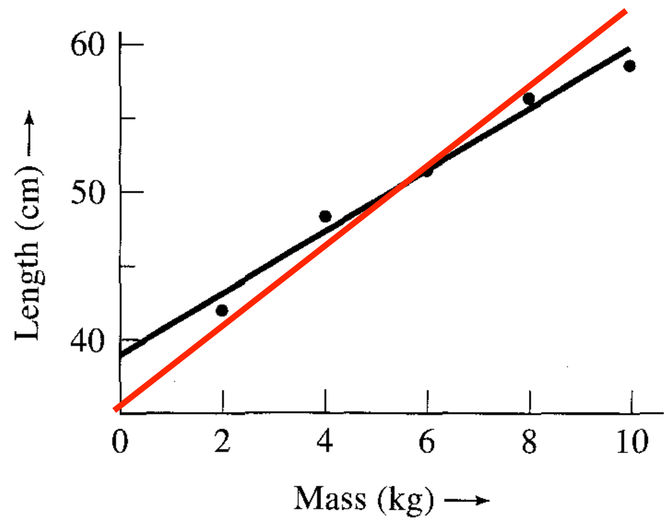


Rules for linear fitting

- Sometimes, you expect from physical arguments that a linear fit should have intercept = 0 (e.g. $x = v_0 t$, $F = ma$)

- Some programs allow you to force the intercept A = 0 when performing a linear fit.

- Suggested procedure:
  - Start by fitting your data to $y = A + Bx$. Do not force A = 0!
  - Using σ_A, see if the difference between A and zero (the expected value) is statistically significant
  - If not, you can try to perform the fit again forcing A = 0.
  - If so, keep the intercept. You may have found evidence for a systematic error in the location of your zero.



Goodness of fit

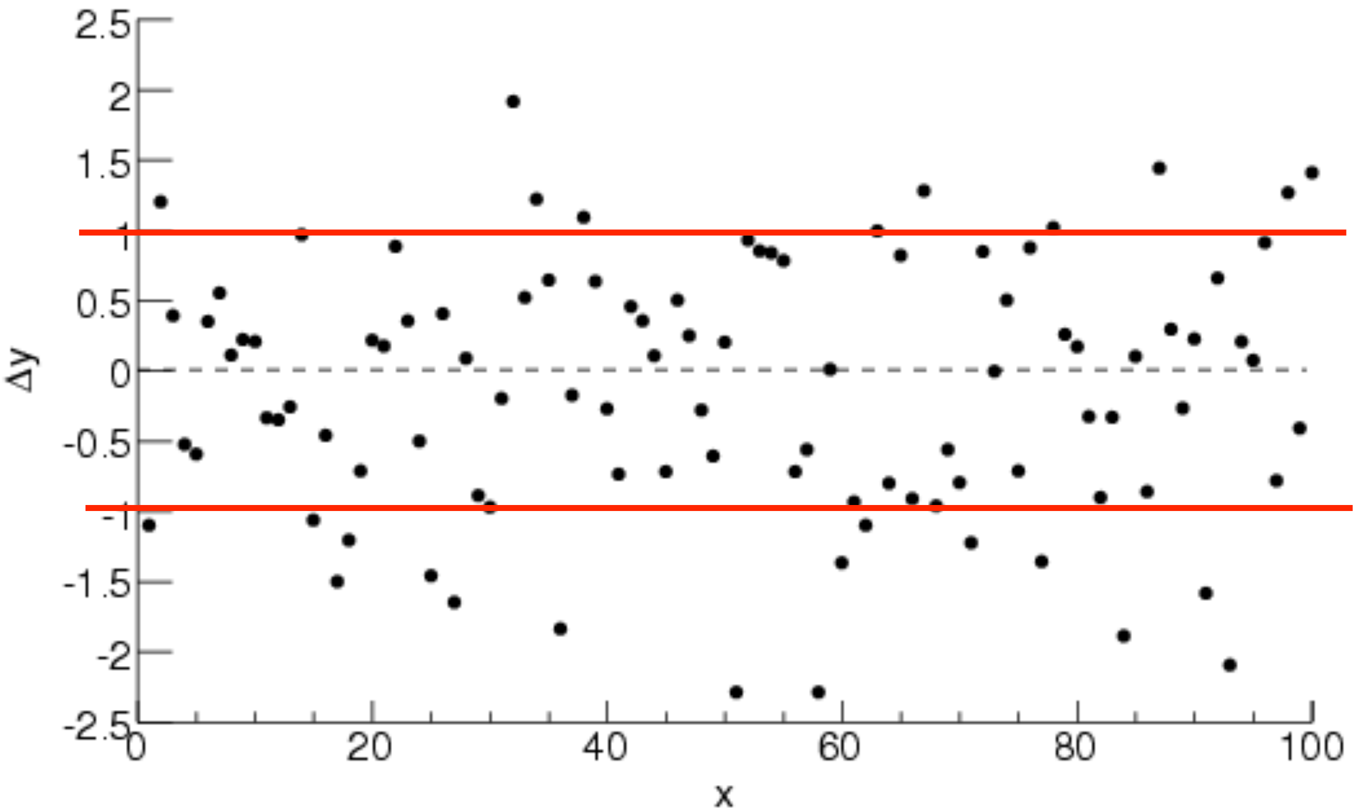
- You can perform a fit on any data set, but the function you choose may or may not be appropriate.

- One way to tell if your fit is good is to plot the residuals Δy as a function of x.

- You should expect that 68% of the residuals (~$^2/_3$) should fall within 1σ of zero.
  - If less than 68% of the residuals are within 1σ of zero, you have probably of underestimated the random errors in your data, or there is a systematic error.
  - If more than 68% of the residuals are within 1σ of zero, the fit is "too good." You have probably overestimated your errors.
  - Remember, this is not exact. But *about* 2/3 should fall within the error bars.

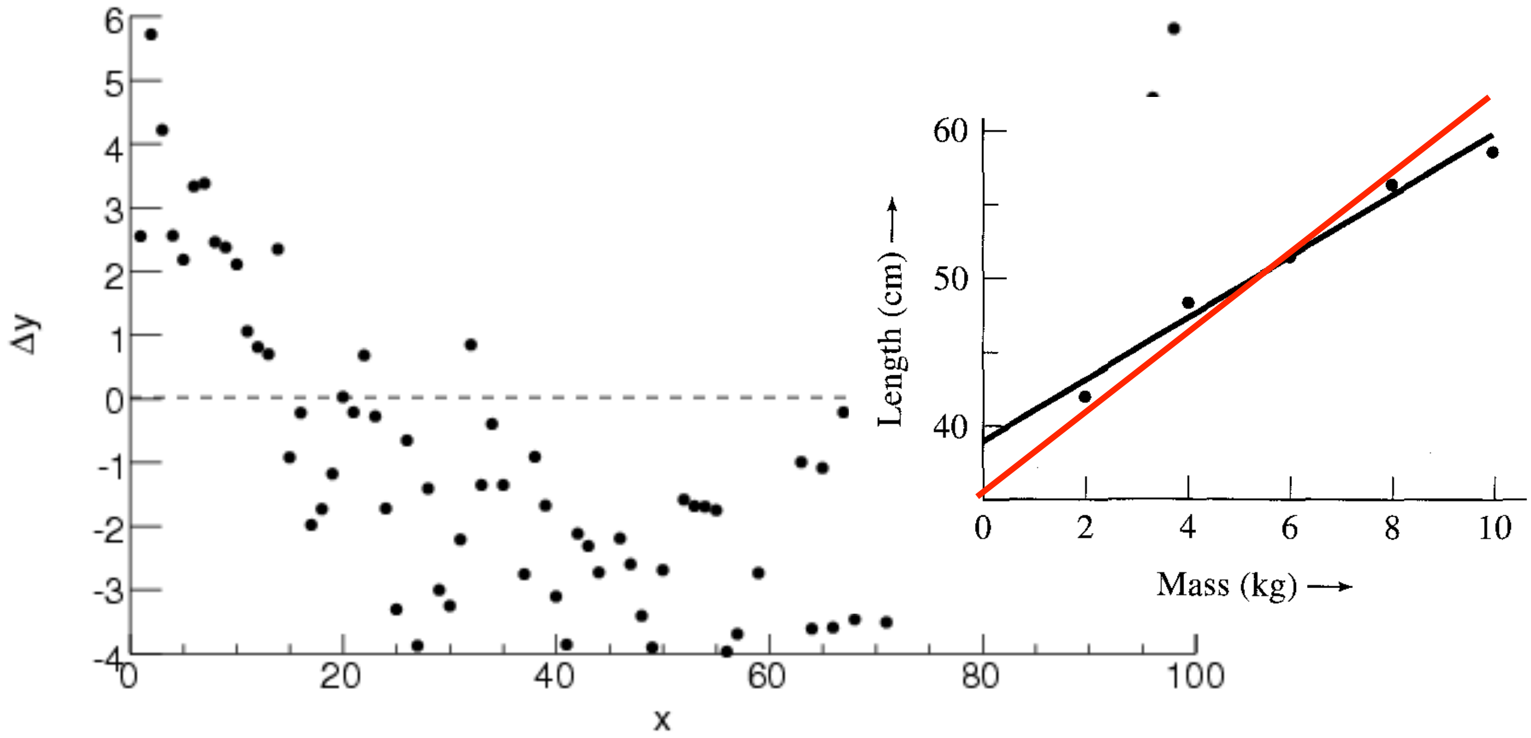
Appropriate residuals

- The ideal distribution of residuals has no observable pattern.



Incorrect fit or bias in the data

- If the wrong equation is fitted to the data, or there is some bias, your residuals may look like this:



Nonlinear data

Example: exponential decay

$$N(t) = N_0 e^{-t/\tau}$$

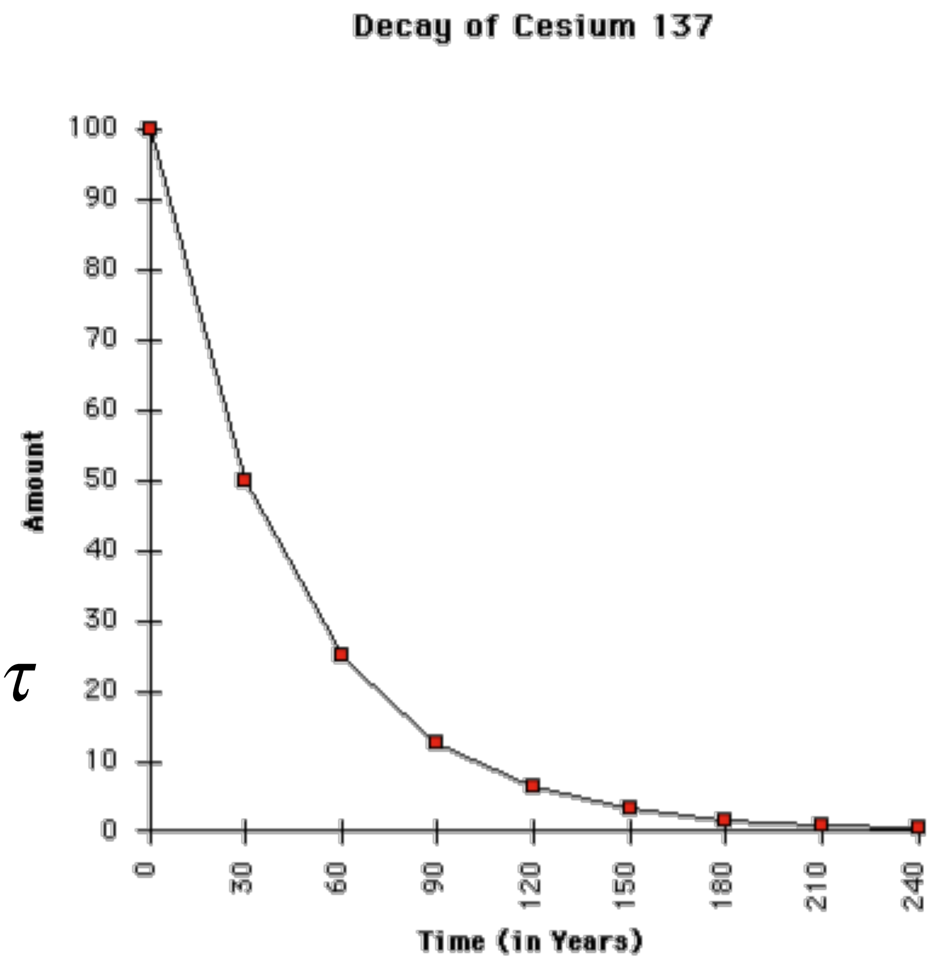
- Doesn't fit to a line

- BUT

$$\ln N(t) = \ln(N_0 e^{-t/\tau}) = \ln N_0 - t/\tau$$

$y = A + B \cdot x$ where $A = \ln N_0$

and $B = -1/\tau$



Admin

- Last lecture!
- Last homework due next Wednesday, 10/9
- Midterm through D2L on 10/14-10/16
  - available for 48 hours
  - 10% of grade
  - similar material to homeworks
  - I'll send out an email with more details