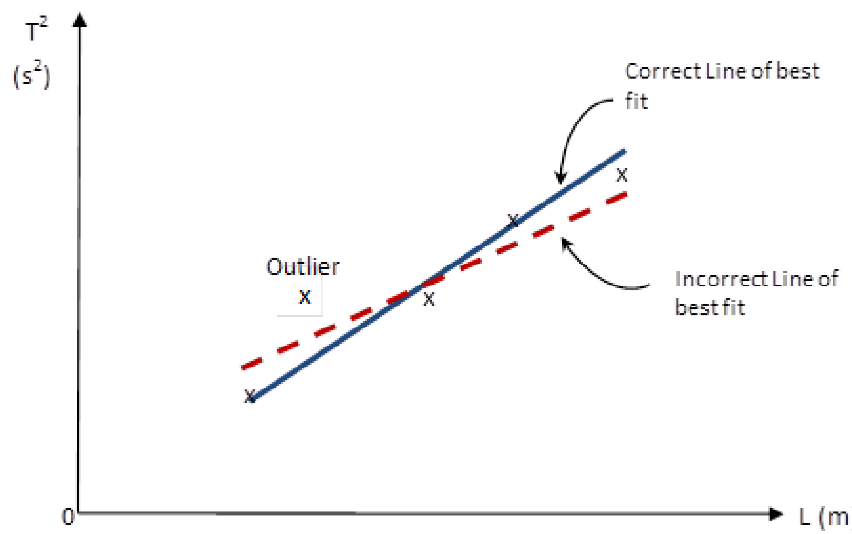
Making Science Graphs and Interpreting Data

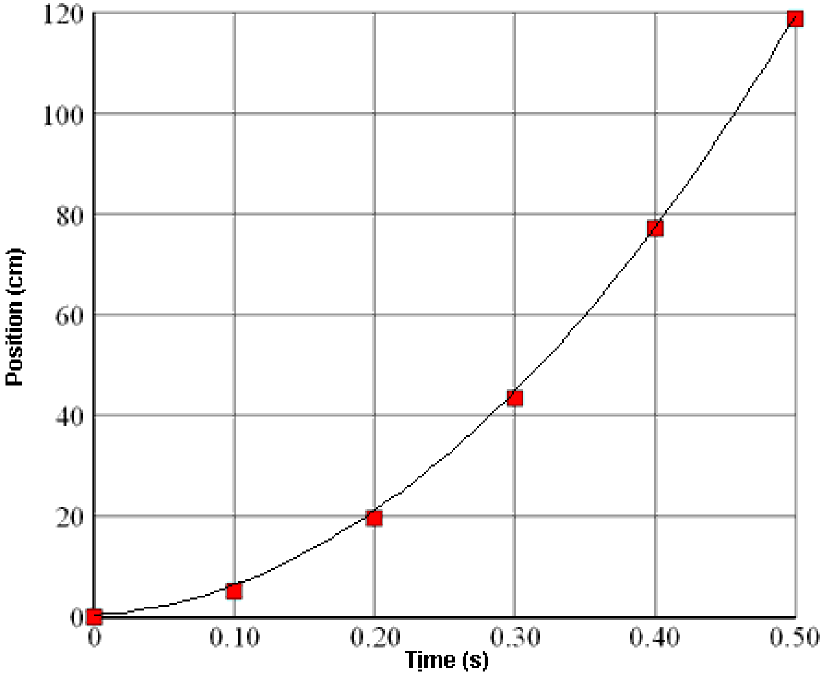
Scientific graphs

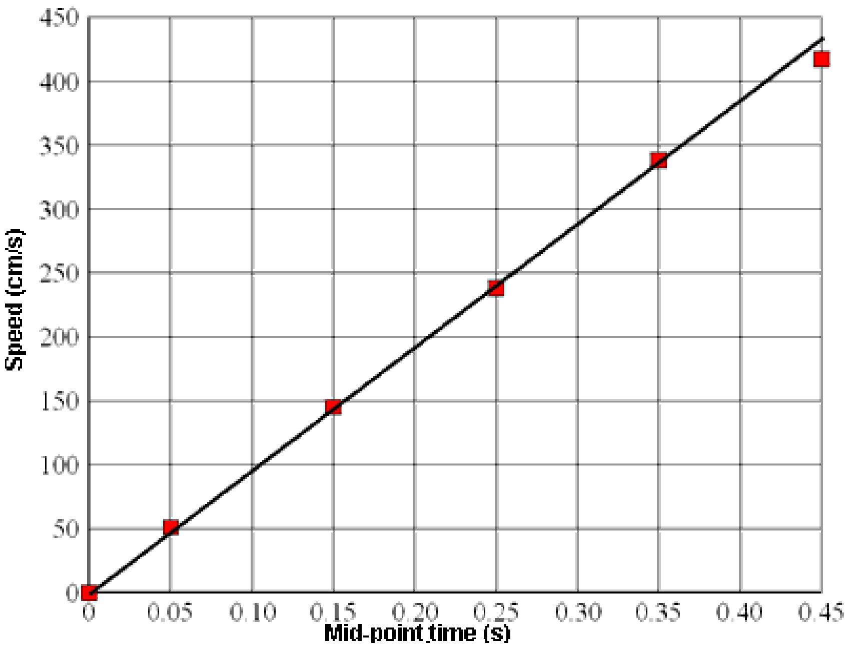
- A graph is a visual representation of a relationship between two variables, x (independent variable) and y (dependent variable).
- The graphs make it easy to identify trends in data that we have collected.

A graph showing a relationship between T^2 and L for a simple pendulum experiment



Scientific Graphs





- Most scientific graphs are made as **line** graphs. There may be times when other types would be appropriate, but they are rare.

- The lines on scientific graphs are usually drawn either **straight** or **curved**. These "smoothed" lines do not have to touch all the data points, but they should at least get close to most of them. They are called **best-fit lines**.

- In general, scientific graphs are not drawn in connect-the-dot fashion.

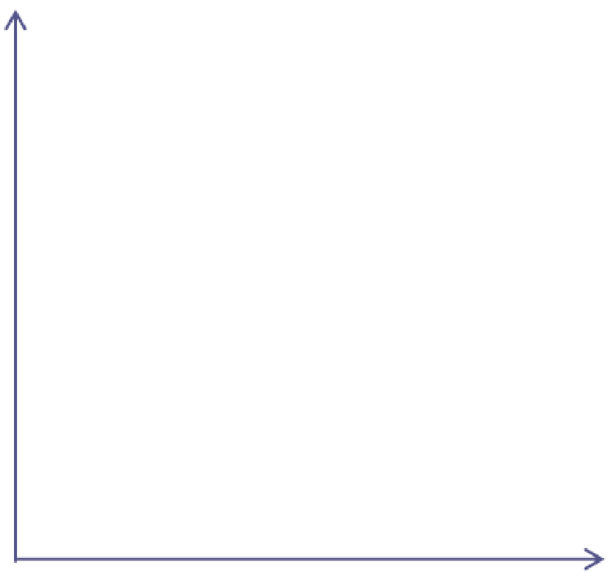
Graphical Analysis

- **Drawing of best fit straight line graph:**
- To draw the best fit straight line graph through a set of scattered experimental data points we will follow a standard statistical method, known as *least squares fit* method.

Convention for labelling tables and graphs

t/s	v/ms^{-1}
0	2.5
1.0	4.0
2.0	5.5

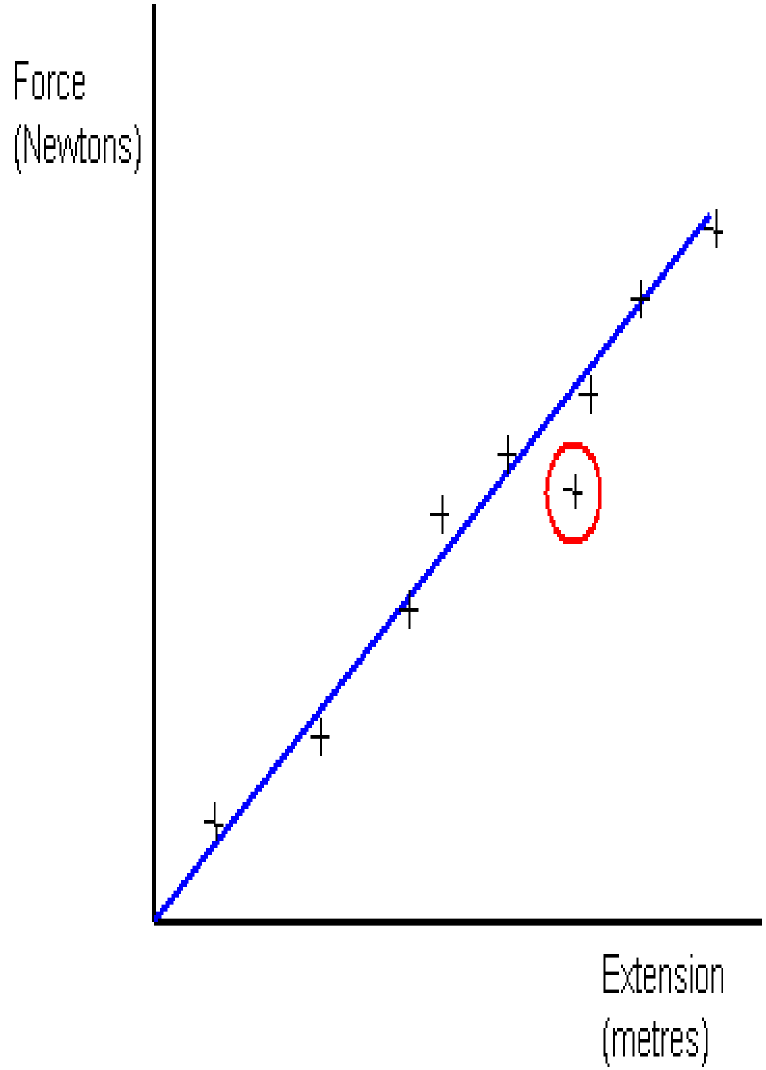
- The symbol / unit is indicated at the italics as indicated in the data column left.
- Then fill in the data with pure numbers.
- Then plot the graph after labelling x axis and y axis



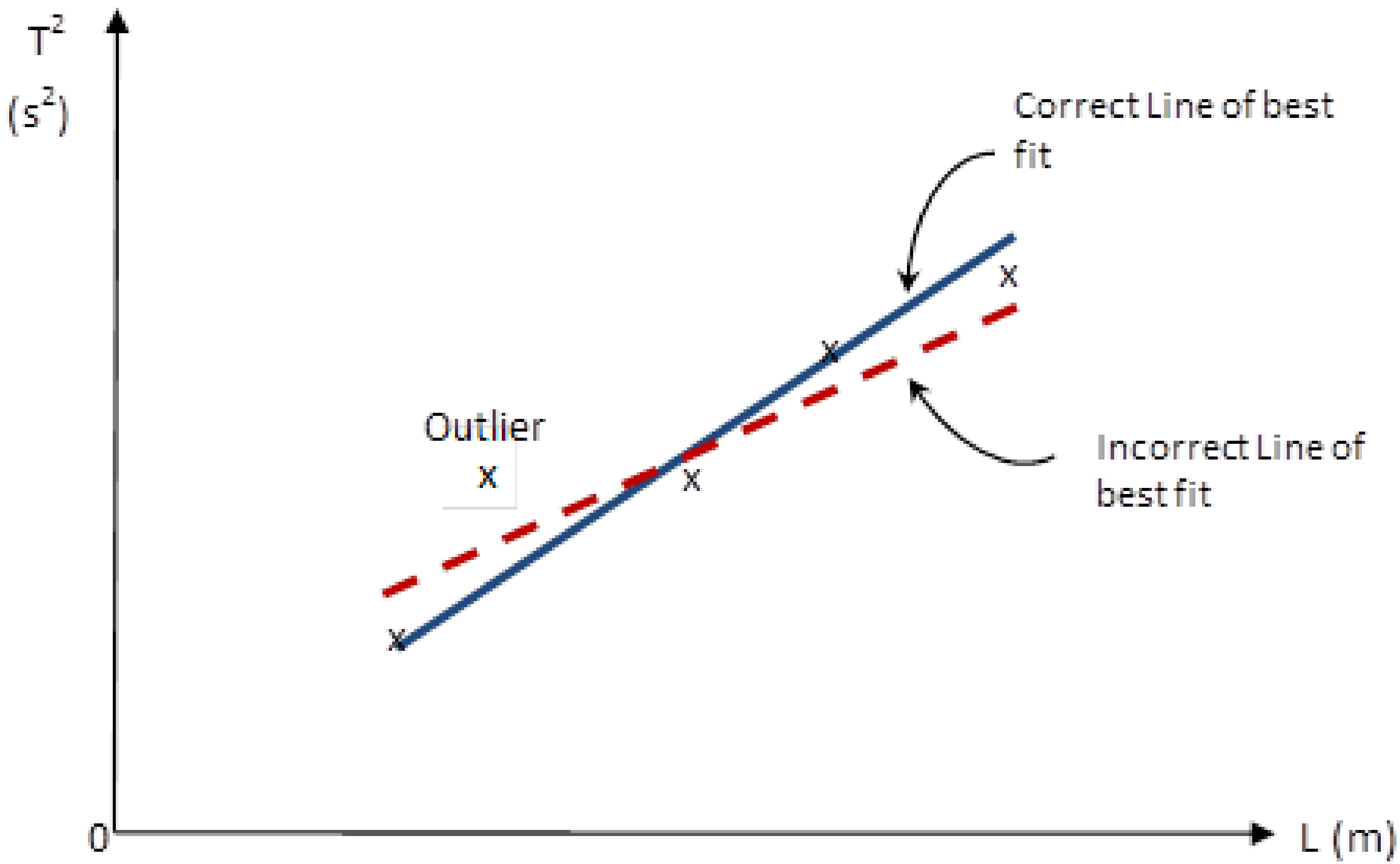
[Illustration with sample graph on left]

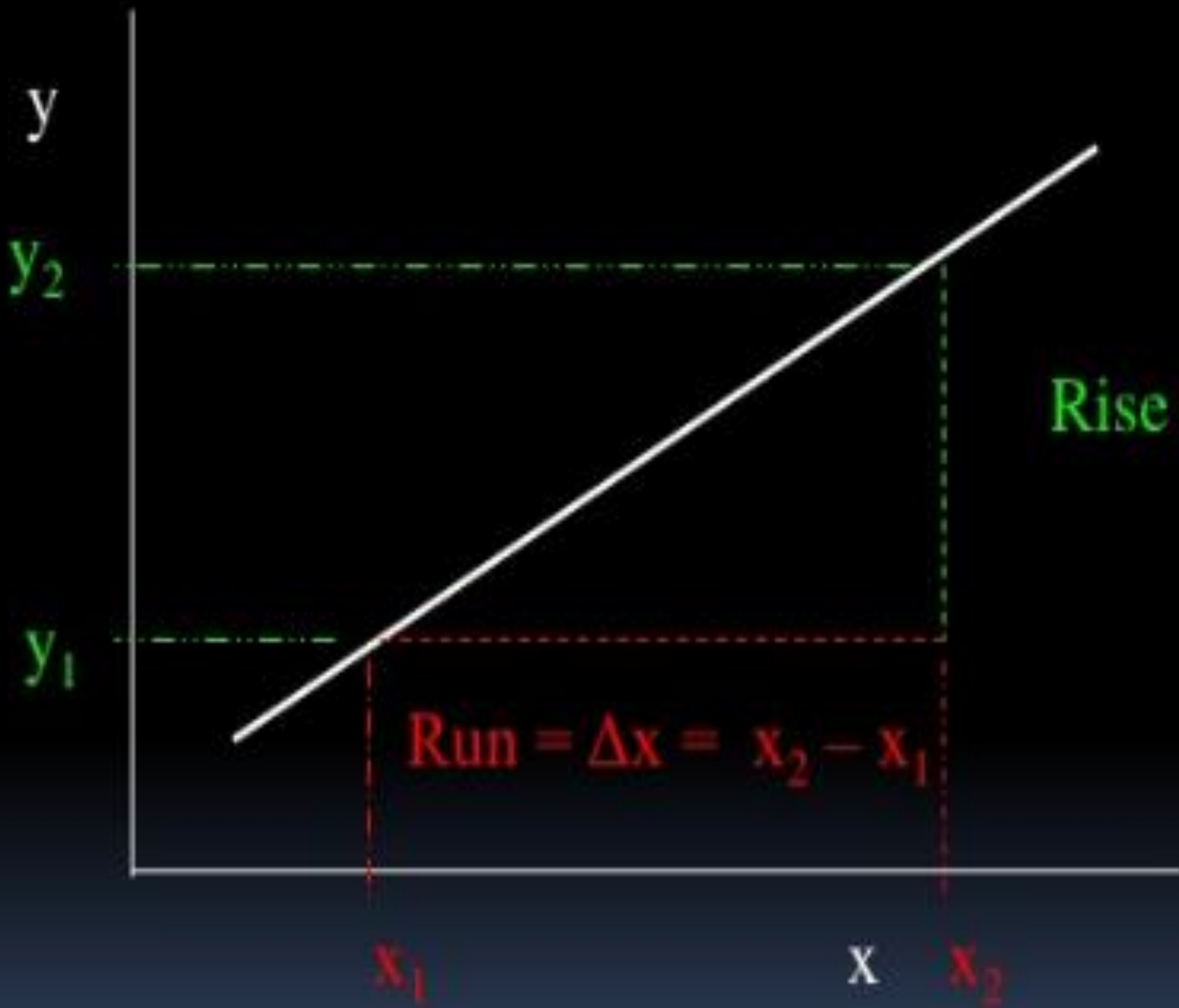
Drawing graphs

1. Give your graph a title
2. Chose a sensible scale for both axes
3. Try and draw a graph that will fill the graph paper
4. Label the axes
5. Put units on both axes
6. Plot the points correctly
7. Mark the points clearly + or a circle with a dot in it
8. Draw a smooth best fit line through the points. This is often called a trend line. It shows the general trend but you do not have to join up all the points – in fact you should not do this unless they actually fit the trend line.
9. Recognise any anomalous points (ones that are way off the general trend).
10. Draw the line clearly and finely



A graph showing a relationship between T^2 and L for a simple pendulum experiment





Rise = $\Delta y = y_2 - y_1$

$$\text{Slope} = \frac{\text{Rise}}{\text{Run}}$$

$$\text{Slope} = \frac{y_2 - y_1}{x_2 - x_1}$$

Once you know the slope then the equation of a line is very easily determined.

Slope Intercept form for any line:

$$y = mx + b$$

slope

y-intercept

(the value of y when x =0)

Of course in Physics we don't use "x" & "y". We could use F and m, or d and t, or F and x etc.)

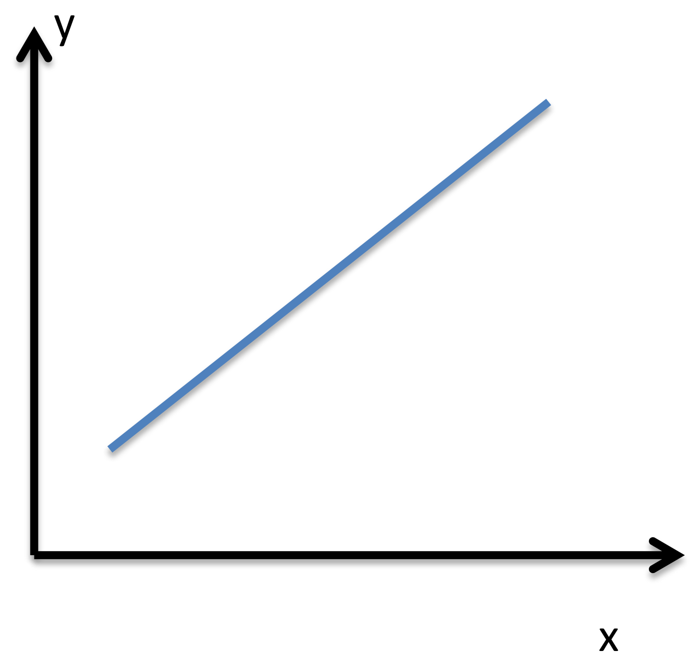
Linear Graphics

- Let us consider a set of N experimental data points (x_1,y_1), (x_2,y_2), ...(x_N,y_N).
- It is well known that a straight-line graph is described by the equation

$y = mx + C$

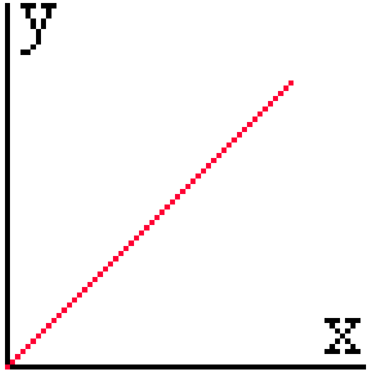
m: slope

x	y



Directly Proportional and Inversely Proportional Graphs

Directly Proportional

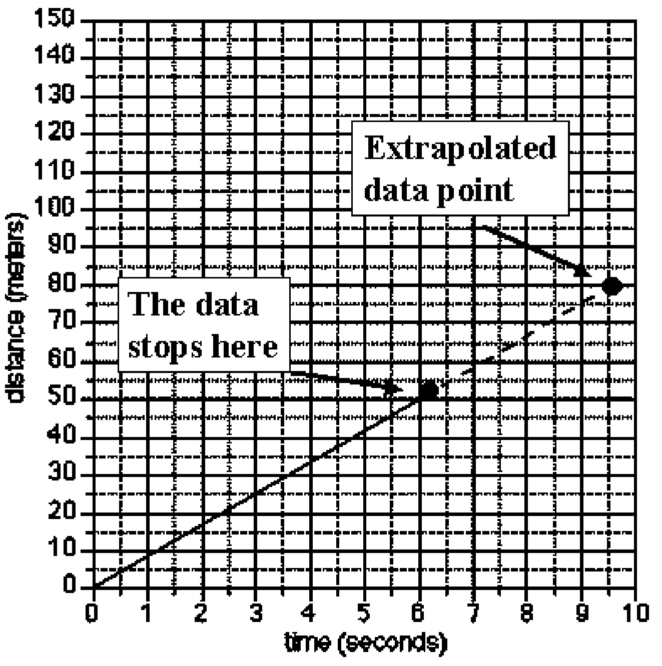


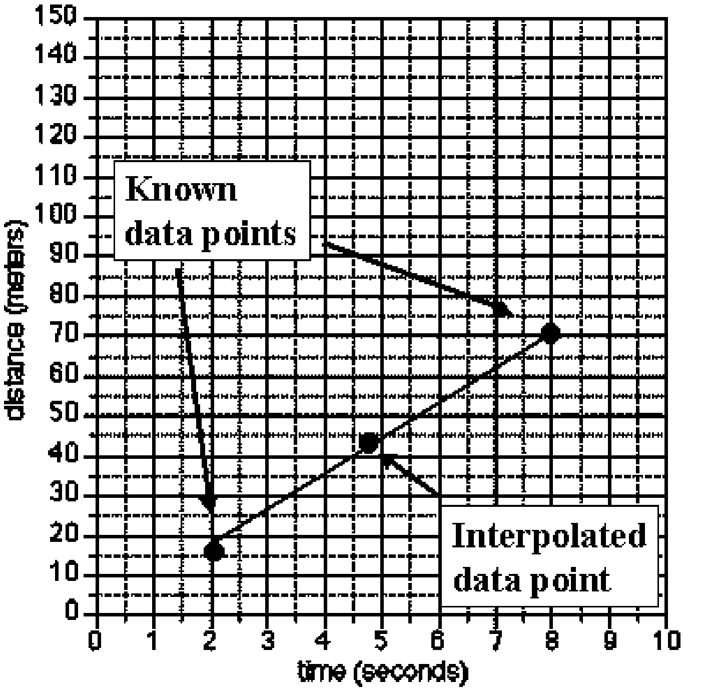
As the independent variable increases, the dependent variable increases as well.

Inversely Proportional

As the independent variable increases, the dependent variable decreases.

Predicting Data on a Graph





- Graphs are a useful tool in science. The visual characteristics of a graph make trends in data easy to see.
- One of the most valuable uses for graphs is to "predict" data that is not measured on the graph.
  - **Extrapolate:** extending the graph, along the same slope, above or below measured data.
  - **Interpolate:** predicting data between two measured points on the graph.

How to Construct a Line Graph

1. **Identify the variables**
   a. Independent variable
      -Goes on the X – axis (horizontal)
      -Should be on the left side of a data table

   b. Dependent variable
      -Goes on the Y – axis (vertical)
      -Should be on the right side of a data table
2. **Determine the scale of the Graph**
   a. Determine a scale (numerical value for each square) that best fits the range of each variable
   b. Spread the graph to use MOST of the available space

How to Construct a Line Graph

3. Number and Label Each Axis

a. This tells what the lines on your graph represent. Label each axis with appropriate units.

4. Plot the Data Points

a. plot each data value on the graph with a dot.

5. Draw the Graph

a. draw a curve or line that best fits the data points.

b. Most graphs of experimental data are not drawn as "connect the dots".

6. Title the Graph

a. Your title should clearly tell what the graph is about.

b. If your graph has more than one set of data, provide a key to identify the different lines.

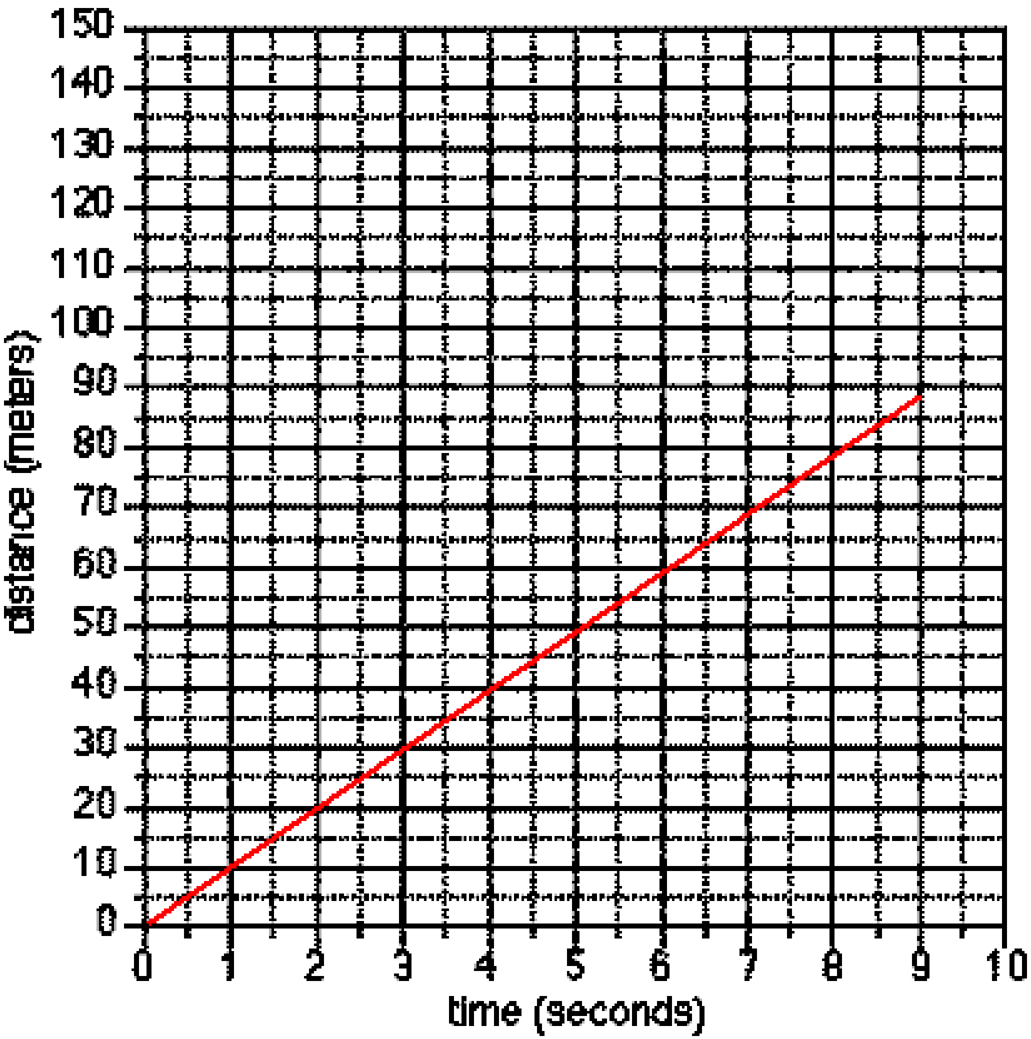
Graphing Practice Problem #1a

Time (seconds)	Distance (meters)
0	0
1	2
2	8
3	18
4	32
5	50
6	72
7	98
8	128
9	162
10	200

A. **Graph the data.**

B. **What does the graph represent?**

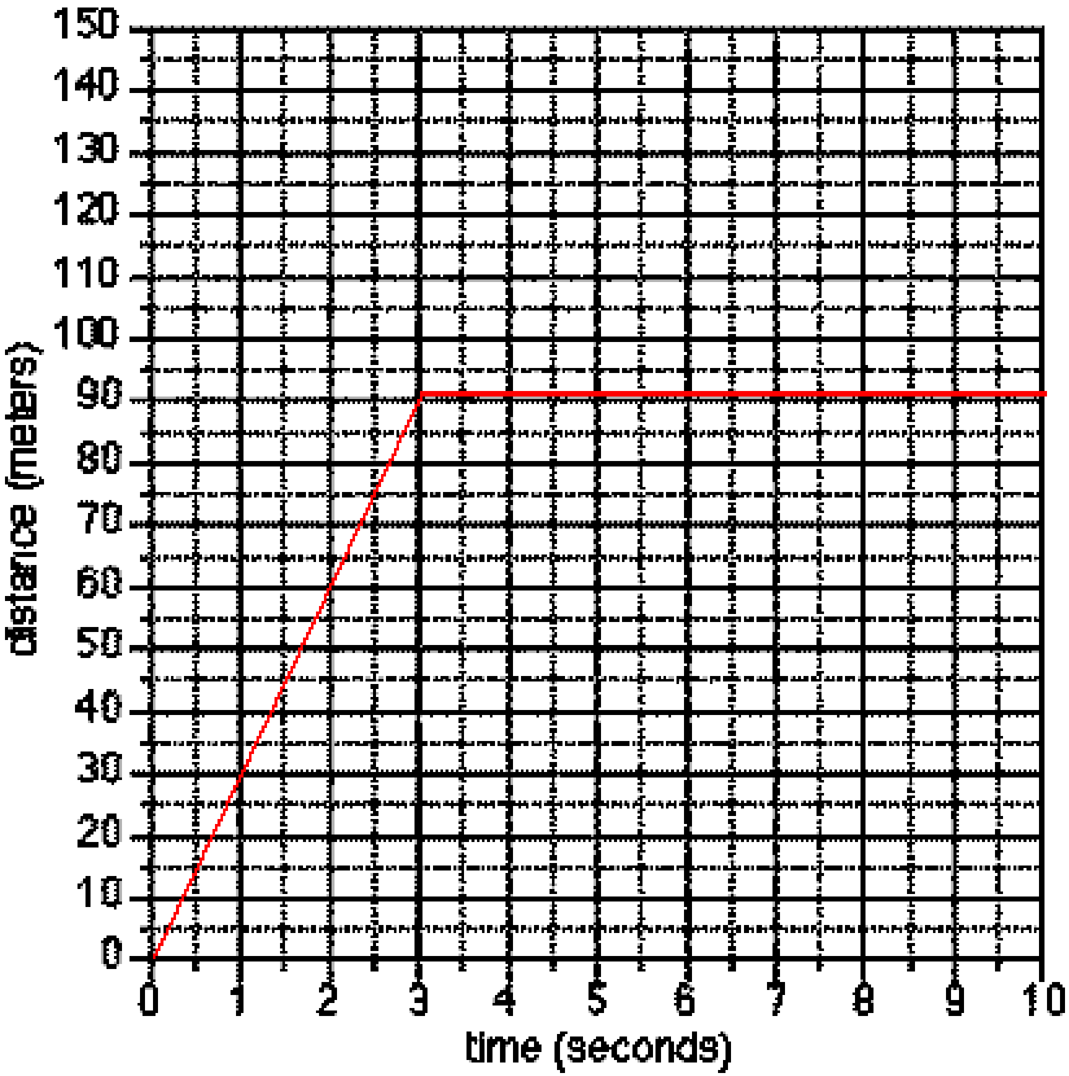
Graphing Practice Problem #1b



A. What type of motion does this graph represent?

B. Put the data from this graph into a table.

Graphing Practice Problem #1c



A. Describe what happens during the time represented by this graph.

B. Put the data from this graph into a table.

Common Graph Forms in Physics

Working with graphs – interpreting, creating, and employing – is an essential skill in the sciences, and especially in physics where relationships need to be derived.

As an introductory physics student you should be familiar with the typical forms of graphs that appear in physics.

Below are a number of typical physical relationships exhibited graphically using standard X-Y coordinates (e.g., no logarithmic, power, trigonometric, or inverse plots, etc.).

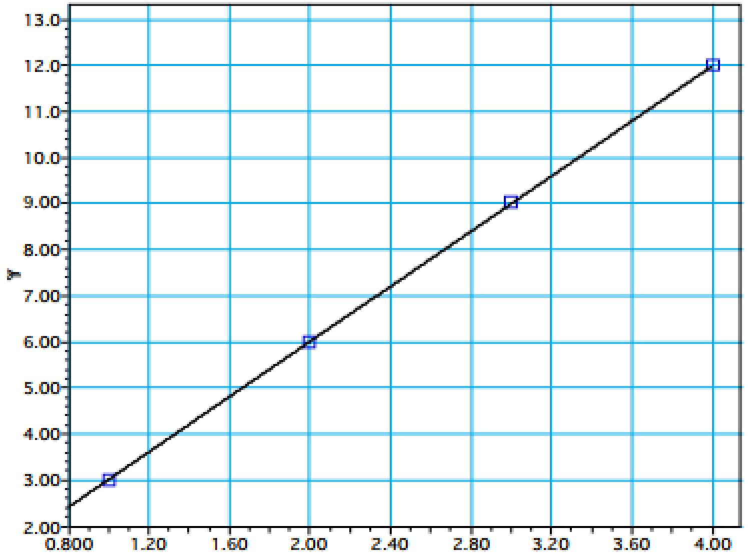
Common Graph Forms in Physics

Study the forms of the graphs carefully, and be prepared to use the program Graphical Analysis to formulate relationships between variables by using appropriate curve-fitting strategies.

Note that all non-linear forms of graphs can be made to appear linear by "linearizing" the data.

Linearization consists of such things as plotting X versus Y2 or X versus 1/Y or Y versus log (X), etc.

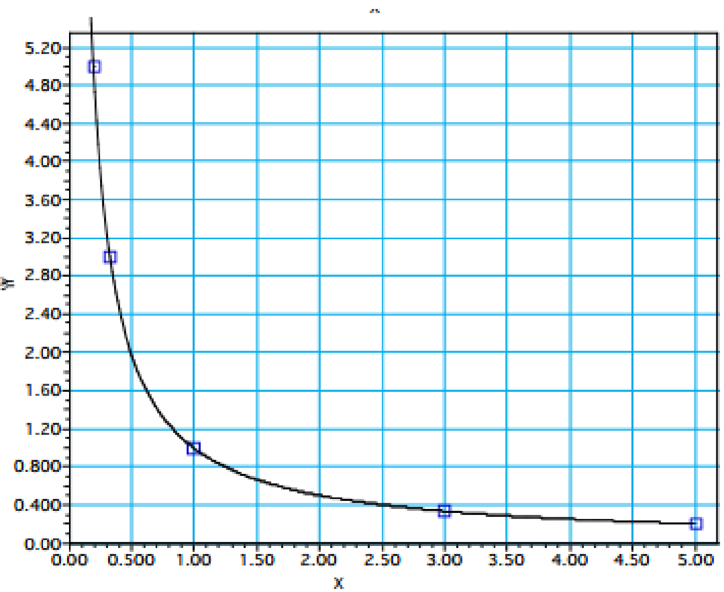
Note: While a 5th order polynomial might give you a better fit to the data, it might not represent the simplest model.



LINEAR RELATIONSHIP: What happens if you get a graph of data that looks like this? How does one relate the X variable to the Y variable? It's simple, Y = A + BX where B is the slope of the line and A is the Y-intercept. This is characteristic of Newton's second law of motion and of Charles' law:

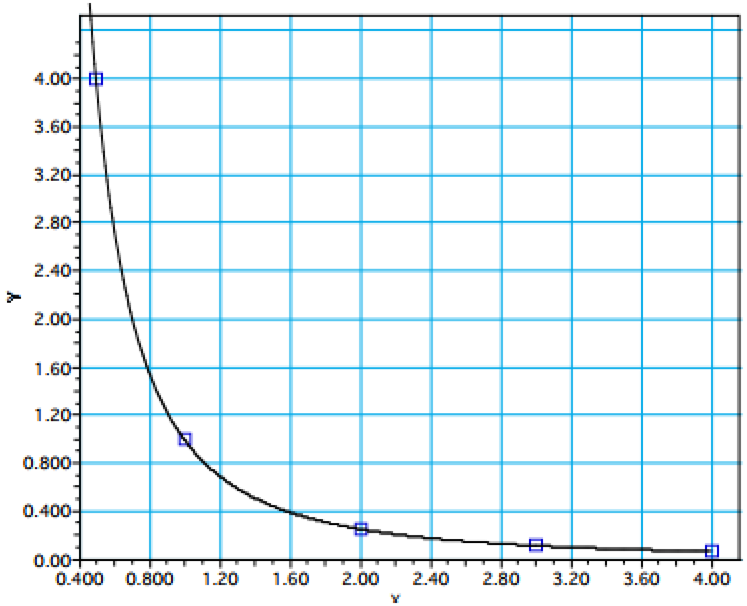
$$F = ma$$

$$\frac{P}{T} = const.$$



INVERSE RELATIONSHIP: This might be a graph of the pressure and temperature for a changing volume constant temperature gas. How would you find this relationship short of using a computer package? The answer is to simplify the plot by manipulating the data. Plot the Y variable versus the inverse of the X variable. The graph becomes a straight line. The resulting formula will be Y = A/X or XY = A. This is typical of Boyle's law:

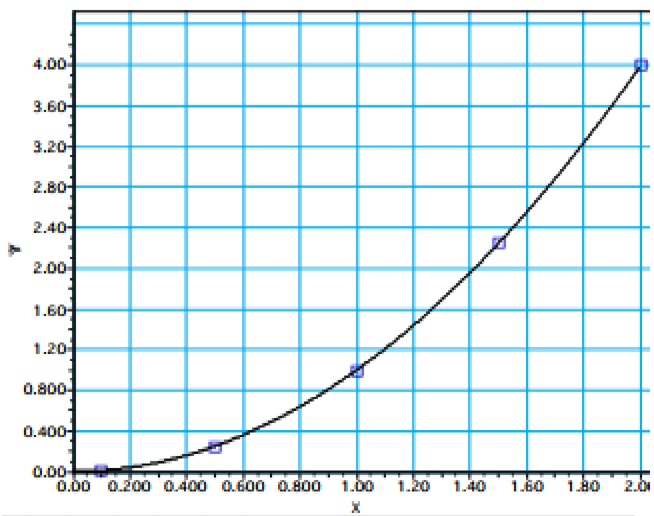
$$PV = const.$$



INVERSE-SQUARE RELATIONSHIP: Of the form Y = A/X^2. Characteristic of Newton's law of universal gravitation, and the electrostatic force law:

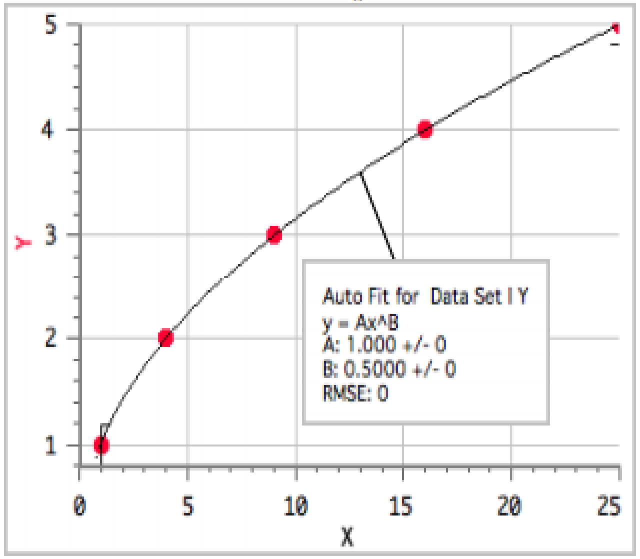
$$F = \frac{Gm_1m_2}{r^2}$$

$$F = \frac{kq_1q_2}{r^2}$$



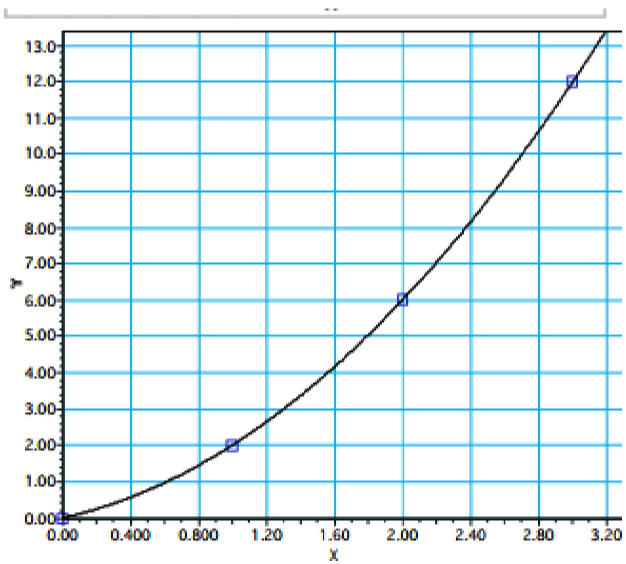
POWER RELATIONSHIP: *Top opening parabola.* Of the form $Y = AX^2$. Typical of the distance-time relationship:

$$d = \frac{1}{2}at^2$$



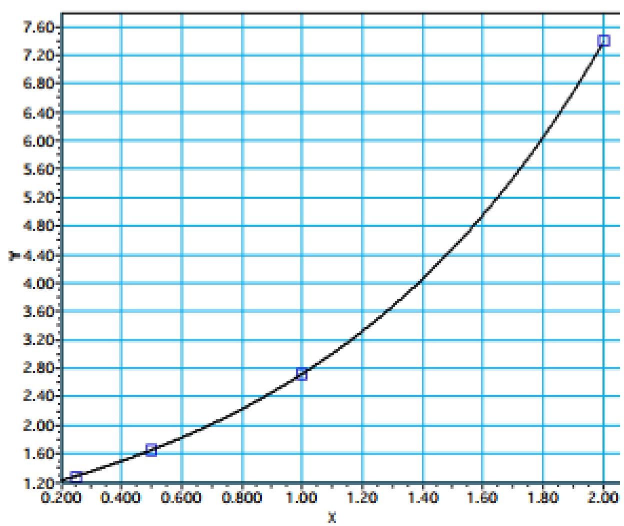
POWER RELATIONSHIP: *Side opening parabola.* Of the form $Y^2=A_1X$ or $Y=A_2X^{1/2}$. Typical of the simple pendulum relationship:

$$P^2 = k_1 l \quad \text{or} \quad P = k_2\sqrt{l}$$



POLYNOMIAL OF THE SECOND DEGREE: Of the form $Y = AX + BX^2$. Typical of the kinematics equation:

$$d = v_o t + \frac{1}{2} a t^2$$



EXPONENTIAL RELATIONSHIP: Of the form Y = A * exp(BX). Characteristic of exponential growth or decay. Graph to left is exponential growth. The graph of exponential decay would look not unlike that of the inverse relationship. Characteristic of radioactive decay.

$$N = N_o e^{-\lambda t}$$

In the latter two examples above there are only subtle differences in form.
Many common graph forms in physics appear quite similar.

Only be looking at the “RMSE” (root mean square error provided in Graphical Analysis) can one conclude whether one fit is better than another. The better fit is the one with the smaller RMSE.

Figure 1 displays the data points along with the best fit model.

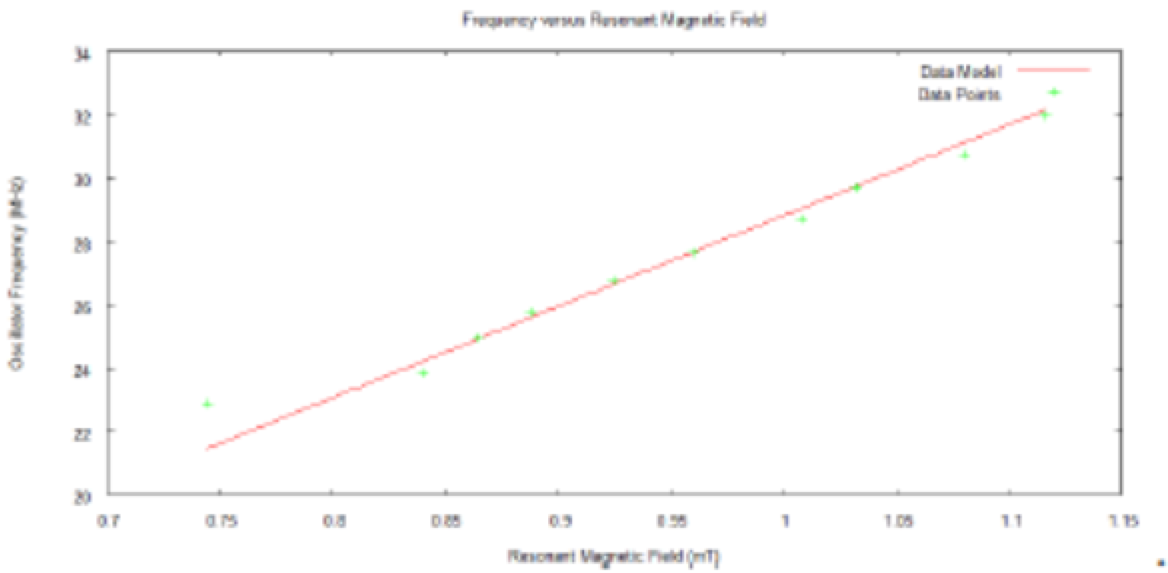


FIG. 1. Setting equation 8 equal to $h\upsilon$ and solving for υ gave rise to an equation suitable for finding μ_s using the least sum of squares method for a linear equation in GNUPlot.

<u>Good</u>

- Caption
- Fig. 1 is mentioned by name in text above.

<u>Bad</u>

- All fonts too small

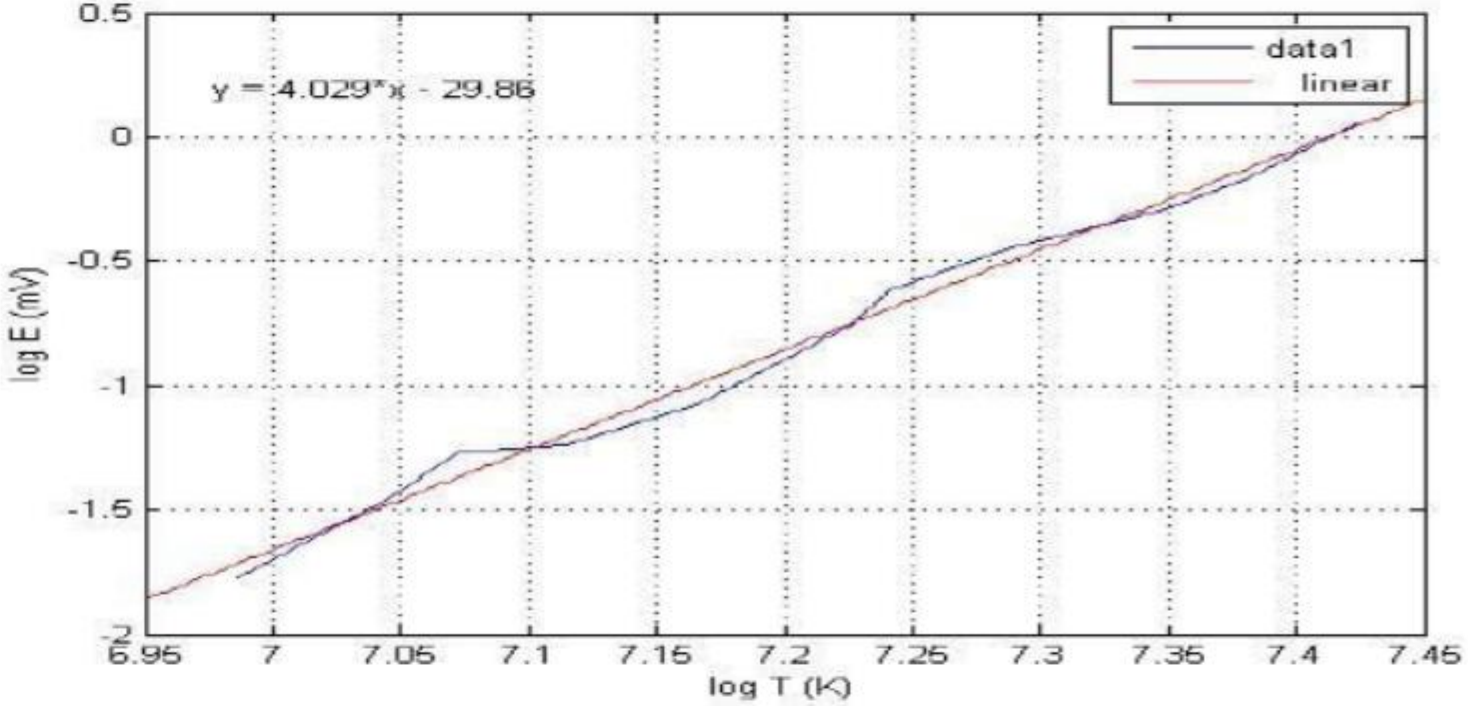


Figure 2: Displays the slope of log T vs. log E. A linear fit is placed on our data and we take a linear fit to get the slope equivalent to approximately four for the Stefan-Boltzmann experiment.

Good

- Caption
- Axes are labeled and units are shown
- Legend

Bad

- No data symbols shown
- Instead, data points connected with lines

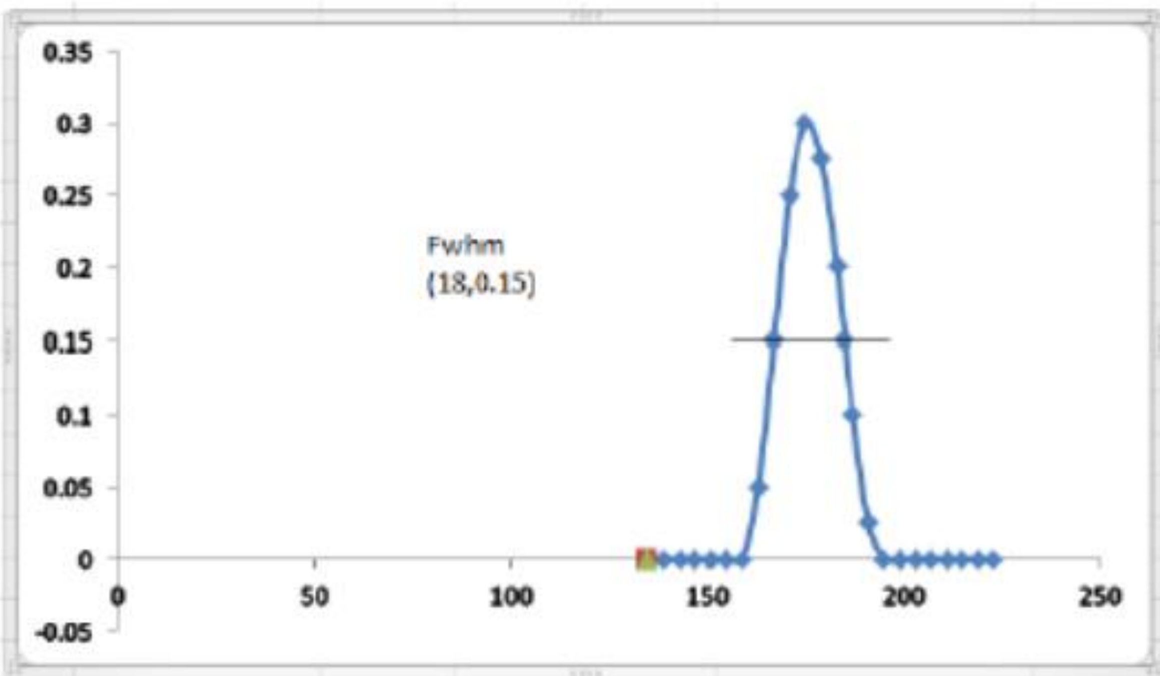


Figure 1: FWHM (18.0,0.15), a point was made to on point (185,0.15) to repersent the fwhm while the rest are data points.

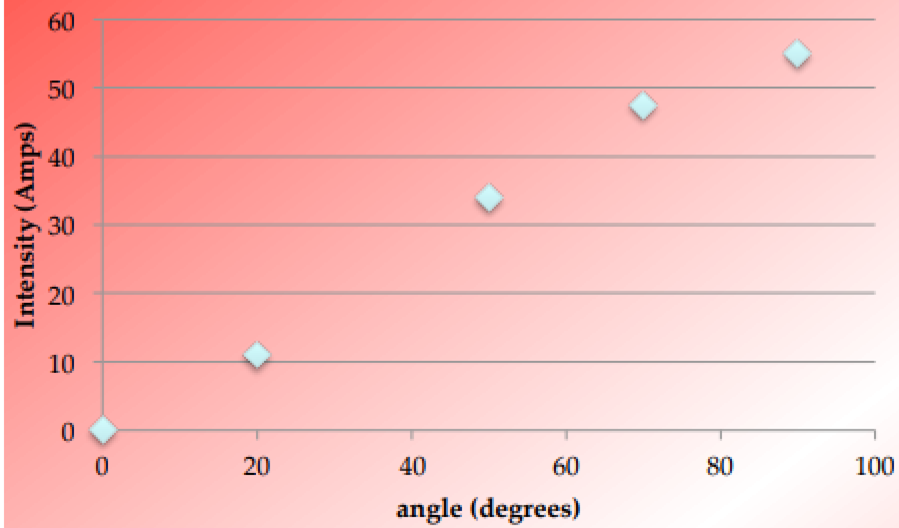
<u>Good</u>

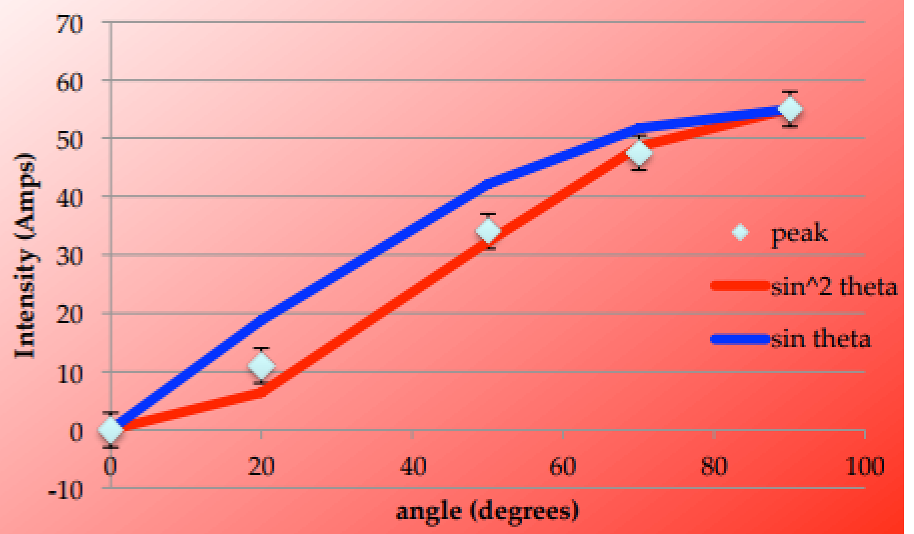
- Caption
- Data points

<u>Bad</u>

- Caption doesn't make sense
- No axis labels
- X-axis should focus on data of interest

Trend Analysis





Fitting with Computer Software

- Most common approach is Least Squares Fitting
- Excel
  - Chart: Add Trendline
  - Limited function choices
  - Goodness of fit: R-squared
- Mathematica
  - Fit[data,funs, vars]
  - Goodness of fit: "χ^2" = $\Sigma_i |F_i - f_i|^2$, sum of residuals
- Origin
  - Several Choices
- Gnuplot
- SciDavis

Types of Error

There are three types of limitations to measurements:

1) **Instrumental limitations**
   Any measuring device can only be used to measure to with a certain degree of fineness. Our measurements are no better than the instruments we use to make them.

2) **Systematic errors and blunders**
   These are caused by a mistake **which does not change** during the measurement. For example, if the platform balance you used to weigh something was not correctly set to zero with no weight on the pan, all your subsequent measurements of mass would be too large. Systematic errors do not enter into the uncertainty. They are either identified and eliminated or lurk in the background producing a shift from the true value.

3) **Random errors**
   These arise from unnoticed variations in measurement technique, tiny changes in the experimental environment, etc. Random variations affect precision. Truly random effects average out if the results of a large number of trials are combined.

Systematic errors are not random and therefore can never cancel out.

They affect the <u>accuracy</u> but not the <u>precision</u> of a measurement.

Statistical Analysis of Small Data Sets

Repeated measurements allow you to not only obtain a better idea of the actual value, but also enable you to characterize the uncertainty of your measurement. Below are a number of quantities that are very useful in data analysis. The value obtained from a particular measurement is x. The measurement is repeated N times. Oftentimes in lab N is small, usually no more than 5 to 10. In this case we use the formulae below:

Mean (x_{avg})	The average of all values of x (the "best" value of x)	$x_{\text{avg}} = \frac{x_1 + x_2 + \cdots + x_N}{N}$
Range (R)	The "spread" of the data set. This is the difference between the maximum and minimum values of x	$R = x_{\text{max}} - x_{\text{min}}$
Uncertainty in a measurement (Δx)	Uncertainty in a single measurement of x. You determine this uncertainty by making multiple measurements. You know from your data that x lies somewhere between x_{max} and x_{min}.	$\Delta x = \frac{R}{2} = \frac{x_{\text{max}} - x_{\text{min}}}{2}$
Uncertainty in the Mean (Δx_{avg})	Uncertainty in the mean value of x. The actual value of x will be somewhere in a neighborhood around x_{avg}. This neighborhood of values is the uncertainty in the mean.	$\Delta x_{\text{avg}} = \frac{\Delta x}{\sqrt{N}} = \frac{R}{2\sqrt{N}}$
Measured Value (x_{m})	The final reported value of a measurement of x contains both the average value and the uncertainty in the mean.	$x_{\text{m}} = x_{\text{avg}} \pm \Delta x_{\text{avg}}$

Example

You measure the length of an object five times.
You perform these measurements twice and obtain the two data sets below.

Measurement	Data Set 1 (cm	Data Set 2 (cm)
x_1	72	80
x_2	77	81
x_3	8	81
x_4	85	81
x_5	88	82

Quantity	Data Set 1 (cm)	Data Set 2 (cm)
x_{avg}	**81**	**81**
R	**16**	**2**
Δx	**8**	**1**
Δx_{avg}	**4**	**0.4**

For Data Set 1, to find the best value, you calculate the mean (i.e. average value):

$$x_{\text{avg}} = \frac{72\text{cm} + 77\text{cm} + 82\text{cm} + 86\text{cm} + 88\text{cm}}{5} = 81\text{cm}$$

The range, uncertainty and uncertainty in the mean for Data Set 1 are then:

$$R = 88\text{cm} - 72\text{cm}{=}16\text{cm}$$

$$\Delta x = \frac{R}{2} = 8\text{cm}$$

$$\Delta x_{\text{avg}} = \frac{R}{2\sqrt{5}} \approx 4\text{cm}$$

Data Set 2 yields the same average but has a much smaller range.

We report the measured lengths x_{m} as:

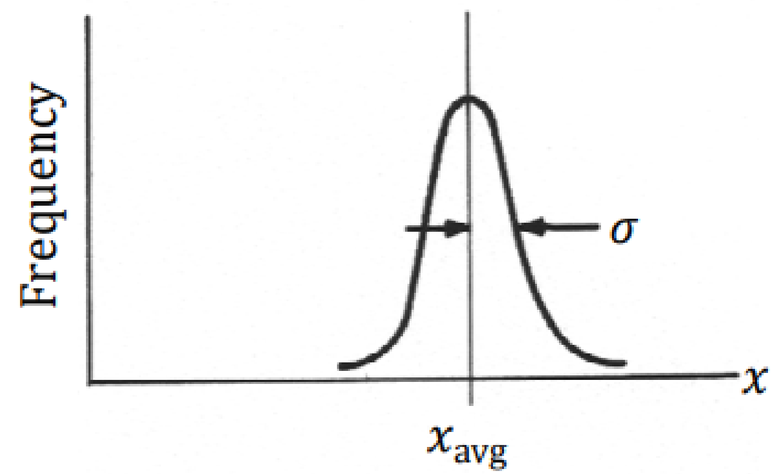
Data Set 1: $\boldsymbol{x_{\text{m}} = 81 \pm 4\text{cm}}$

Data Set 2: $\boldsymbol{x_{\text{m}} = 81.0 \pm 0.4\text{cm}}$

Notice that for Data Set 2, Δx_{avg} is so small we had to add another significant figure to x_{m}.

Statistical Analysis of Large Data Sets

If only random errors affect a measurement, it can be shown mathematically that in the limit of an **infinite** number of measurements ($N \rightarrow \infty$), the distribution of values follows a **normal distribution** (i.e. the bell curve on the right). This distribution has a peak at the mean value x_{avg} and a width given by the standard deviation σ.

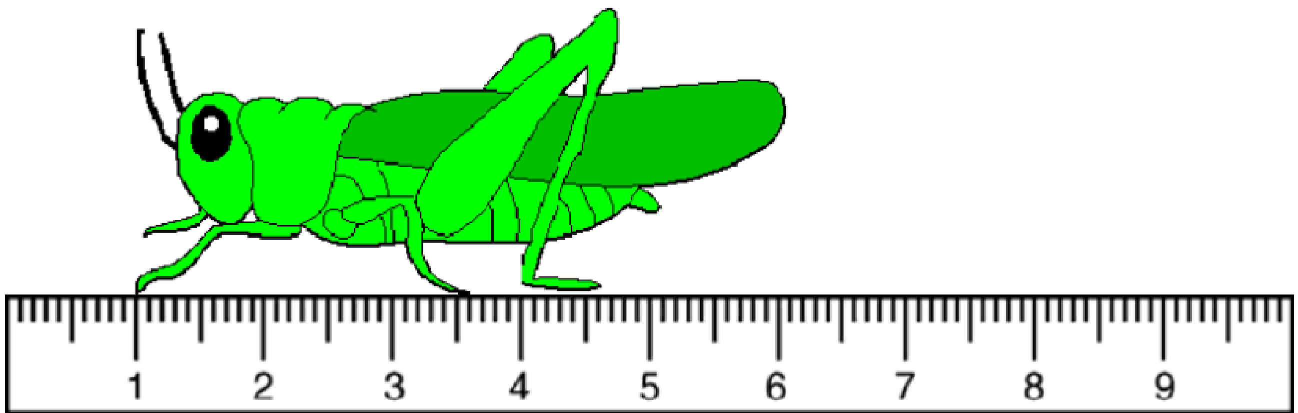


Obviously, we never take an infinite number of measurements. However, for a large number of measurements, say, $N \sim 10 - 10^2$ or more, measurements may be approximately normally distributed. In that event we use the formulae below:

Mean (x_{avg})	The average of all values of x (the "best" value of x). This is the same as for small data sets.	$x_{\text{avg}} = \frac{\sum_{i=1}^{N} x_i}{N}$
Uncertainty in a measurement (Δx)	Uncertainty in a single measurement of x. The vast majority of your data lies in the range $x_{\text{avg}} \pm \sigma$	$\Delta x = \sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - x_{\text{avg}})^2}{N}}$
Uncertainty in the Mean (Δx_{avg})	Uncertainty in the mean value of x. The actual value of x will be somewhere in a neighborhood around x_{avg}. This neighborhood of values is the uncertainty in the mean.	$\Delta x_{\text{avg}} = \frac{\sigma}{\sqrt{N}}$
Measured Value (x_{m})	The final reported value of a measurement of x contains both the average value and the uncertainty in the mean.	$x_{\text{m}} = x_{\text{avg}} \pm \Delta x_{\text{avg}}$

Propagation of Uncertainties

Oftentimes we combine multiple values, each of which has an uncertainty, into a single equation. In fact, we do this every time we measure something with a ruler. Take, for example, measuring the distance from a grasshopper's front legs to his hind legs. For rulers, we will assume that the uncertainty in all measurements is one-half of the smallest spacing.



$x = 1.0\text{cm} \pm 0.05\text{cm}$ $\quad$ $y = 4.63\text{cm} \pm 0.05\text{cm}$

The measured distance is $d_\text{m} = d \pm \Delta d$ where $d = 4.63\text{cm} - 1.0\text{cm} = 3.63\text{cm}$. What is the uncertainty in d_m? You might think that it is the sum of the uncertainties in x and y (i.e. $\Delta d = \Delta x + \Delta y = 0.1\text{cm}$). However, statistics tells us that if the uncertainties are independent of one another, the uncertainty in a sum or difference of two numbers is obtained by quadrature: $\Delta d = \sqrt{(\Delta x)^2 + (\Delta y)^2} = 0.07\text{cm}$. The way these uncertainties combine depends on how the measured quantity is related to each value. Rules for how uncertainties propagate are given below.

Addition/Subtraction	$z = x \pm y$	$\Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2}$
Multiplication	$z = xy$	$\Delta z = \lvert xy \rvert \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}$
Division	$z = \frac{x}{y}$	$\Delta z = \left\lvert \frac{x}{y} \right\rvert \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}$
Power	$z = x^n$	$\Delta z = \lvert n \rvert x^{n-1} \Delta x$
Multiplication by a Constant	$z = cx$	$\Delta z = \lvert c \rvert \Delta x$
Function	$z = f(x, y)$	$\Delta z = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 (\Delta x)^2 + \left(\frac{\partial f}{\partial y}\right)^2 (\Delta y)^2}$

Examples

Addition

The sides of a fence are measured with a tape measure to be 124.2cm, 222.5cm, 151.1cm and 164.2cm. Each measurement has an uncertainty of 0.07cm. Calculate the measured perimeter P_{m} including its uncertainty.

$$P = 124.2\mathrm{cm} + 222.5\mathrm{cm} + 151.1\mathrm{cm} + 164.2\mathrm{cm} = 662.0\mathrm{cm}$$

$$\Delta P = \sqrt{(0.07\mathrm{cm})^2 + (0.07\mathrm{cm})^2 + (0.07\mathrm{cm})^2 + (0.07\mathrm{cm})^2} = 0.14\mathrm{cm}$$

$$\boxed{P_{\mathrm{m}} = 662.0 \pm 0.1\mathrm{cm}}$$

Multiplication

The sides of a rectangle are measured to be 15.3cm and 9.6cm. Each length has an uncertainty of 0.07cm. Calculate the measured area of the rectangle A_{m} including its uncertainty.

$$A = 15.3\mathrm{cm} \times 9.6\mathrm{cm} = 146.88cm^2$$

$$\Delta A = 15.3\mathrm{cm} \times 9.6\mathrm{cm}\sqrt{\left(\frac{0.07}{15.3}\right)^2 + \left(\frac{0.07}{9.6}\right)^2} = 1.3\mathrm{cm}^2$$

$$\boxed{A_{\mathrm{m}} = 147 \pm 1\mathrm{cm}^2}$$