Appendix C: Graphs

Vern Lindberg

1 Making Graphs

"A picture is worth a thousand words."

Graphical presentation of data is a vital tool in the sciences and engineering. Good graphs convey a great deal of information and can be used to extract new conclusions. Bad graphs are simply confusing.

1.1 Basic layout of a graph

Certain conventions are used when plotting graphs. Figure 1 shows some basic elements of a graph. Other examples come later in Figures 5, 6, and 7.

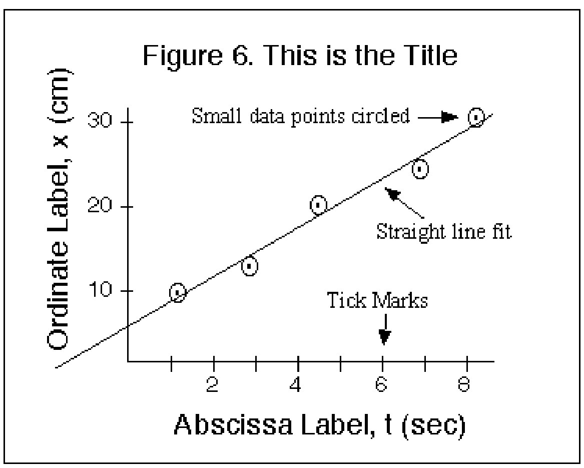

Figure 1: Some basic elements of a graph

(a) The horizontal axis is called the *abscissa* and the vertical axis is called the *ordinate*. You can use the terms horizontal and vertical just as well.

(b) The graph must have a title which clearly states the purpose of the graph. This should be located on a clear space near the top of the graph. A possible title for a graph would be "Figure 6. Variation of Displacement With Elapsed Time for a Moving Car"

The title should uniquely identify the graph—you should not have three graphs with the same title. You may wish to elaborate on the title with a brief caption. Do not just repeat the labels for the axes! Having figure numbers allows you to easily refer to the graph.

Examples of poor titles

y vs t	The title should be in words and should not just repeat the symbols on the axes!
Displacement versus time	This title is in words, but just repeats the names on the axes. The title should add information.
Data from Table 1	Again, this adds minimal information. It may be useful to include this information, but tell what the graph is and what it means.

(c) Normally you plot the independent variable (the one over which you have control) on the abscissa (horizontal), and the dependent variable (the one you read) on the ordinate (vertical). If for example you measure the position of a falling ball at each of several chosen times, you plot the position on the ordinate (vertical) and the time on the abscissa (horizontal.)

In speaking of a graph you say "I plotted vertical versus horizontal or ordinate versus abscissa." The graph we have just discussed is a plot of position versus time. If you are told to plot current versus voltage, voltage goes on the abscissa (horizontal).

(d) The scale should be chosen so that it is easy to read, and so that it makes the data occupy more than half of the paper. Good choices of units to place next to major divisions on the paper are multiples of 1, 2, and 5. This makes reading subdivisions easy. Avoid other numbers, especially 3, 6, 7, 9, since you will likely make errors in plotting and in reading values from the graph. The zero of a scale does not need to appear on the graph.

Computer plotting packages should allow you control over the minimum and maximum values on the axis, as well as the size of major and minor divisions. The packages should allow you to include a grid on the plot to make it look more like real graph paper.

(e) Tick marks should be made next to the lines for major divisions and subdivisions. Look at Figures 5 and 7 to see examples. Logarithmic scales are pre-printed with tick marks.

(f) Axis label. The axes should be labeled with words and with units clearly indicated. The words describe what is plotted, and perhaps its symbol. The units are generally in parentheses. An example would be

Displacement, y, of ball (cm)

On the horizontal axis (abscissa) the label is oriented normally, as are the numbers for the major divisions. The numbers for the major divisions on the vertical axis are also oriented normally. The vertical axis label is rotated so that it reads normally when the graph paper is rotated 1/4 turn clockwise. See Figures 5, 6, and 7 for examples.

Avoid saying "Diameter in meters ($\times 10^{-4}$)" since this confuses the reader. (Do I multiply the value by 10^{-4} or was the true value multiplied by this before plotting?) Instead state "Diameter ($\times 10^{-4}$ meters)" or use standard prefixes like kilo or micro so that the exponent is not needed: "Diameter (mm)."

(g) Data should be plotted as precisely as possible, with a sharp pencil and a small dot. In order to see the dot after it has been plotted, put a circle or box around the dot. If you plot more than one set of data on the same axes use a circle for one, a box for the second, etc., as shown in Figure 2.

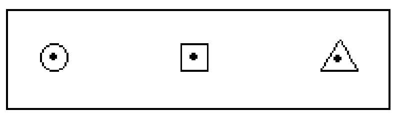

Figure 2: For hand drawn data, use a small dot surrounded by a circle, square, or triangle.

1.2 Uncertainties and Graphs: Error Bars

Data that you plot on a graph will have experimental uncertainties. These are shown on a graph with error bars, and used to find uncertainties in the slope and intercept.

Consider a point with coordinates $X \pm \Delta X$ and $Y \pm \Delta Y$. (a) Plot a point, circled, at the point (X, Y).

(b) Draw lines from the circle to $X + \Delta X, X - \Delta X, Y + \Delta Y,$ and $Y - \Delta Y$ and put bars on the lines, as shown in Figure 3. These are called error bars.

(c) The true value of the point is likely to lie somewhere in the oval whose dimension is 2σ, i.e. twice the size of the error bars. as shown in Figure 3(c). It is not usually drawn on graphs. Often the error bars may be visible only for the ordinate (vertical), as Figure 3(b). Draw the best error bars that you can! If they cannot be seen, make a note to that effect on the graph.

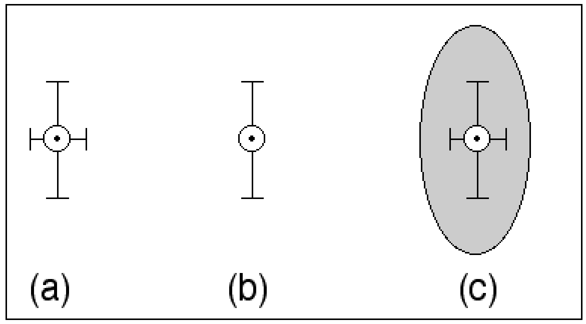

Figure 3: A data point with error bars. (a) If uncertainties are known for both vertical and horizontal coordinates, extend bar in each direction by 1σ (b) Often an error will only be significant for one coordinate, here the vertical. (c) The shaded region is the 2σ area in which we are likely to have the true value occur.

2 Straight line graphs on linear graph paper.

Suppose that we have plotted a graph with Y on the ordinate and X on the abscissa and the resulting plot suggests a straight line. We know that the general equation for a straight line is

$$Y = MX + B \tag{1}$$

where M is the *slope* and B is the *intercept* on the Y-axis i.e. the Y-intercept.

The capital forms of Y and X are chosen to represent any arbitrary variables we choose to plot. For example we may choose to plot position, x, on the Y-axis versus mass, m, on the X-axis. Figure 4 illustrates the terms.

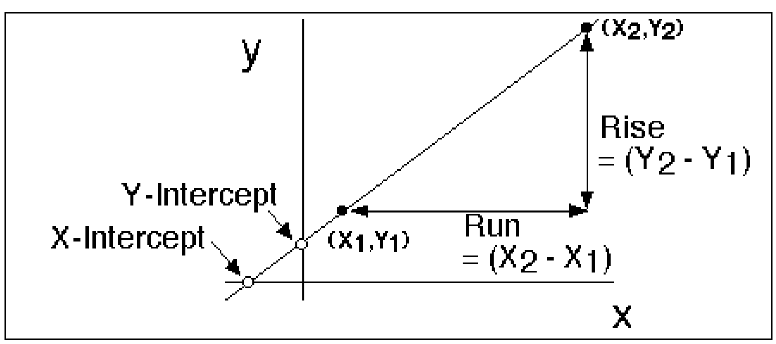

Figure 4: Rise, run, and intercepts defined.

While there are statistical methods of drawing the “best fit straight line”, your common sense will serve just fine. Draw a line (use a ruler) so that half the data are above the line and half below the line.

To determine the slope, choose two points, (X_1, Y_1) and (X_2, Y_2), from the straight line that are not data points and that lie near opposite ends of the line so that a precise slope can be calculated. $(Y_2 - Y_1)$ is called the *rise* of the line, while $(X_2 - X_1)$ is the *run*. The slope is

$$M = \frac{Y_2 - Y_1}{X_2 - X_1} \tag{2}$$

Slope has units and these must be included in your answer!

The point where the line crosses the vertical axis is called the intercept (or the Y-intercept) with coordinates (0, Y-intercept). The intercept has the same units as Y.

The line can be extended to cross the horizontal axis as well. The value of X where this happens is called the X-intercept, with the same units as variable X, and coordinates $(X\text{-intercept}, 0)$.

If the line goes directly through the origin, with intercepts of zero, we say that Y is directly proportional to X. The word proportional implies that not only is there a linear (straight line) relation between Y and X, but also that the intercept is zero, $Y = MX$.

2.1 Uncertainties in Slope and Intercept Using Error Bars

In this discussion we will describe simple means for finding uncertainties in slope and intercept; a full statistical discussion would begin with "Least Squares Fitting" and then proceed to "analysis of variance."

Once the graph is drawn and a linear fit is made and the slope and intercept are determined, we wish to find uncertainties in the slope and intercept. Figure 5 shows position versus time for a snail. Data are plotted on this graph with error bars shown. The uncertainty in time is so small that no horizontal error bars are visible. Look closely to see that this meets the criteria for a good graph.

A solid line is shown which best fits the data. To get the slope, pick two points on the line located far apart on the graph. The points are not data points since they may not actually be on the line. The slope is (points chosen and calculation shown on the graph)

$$\text{Slope} = \frac{13.3 - 0.3}{6.5 - 0.5} = 2.09 \text{ cm/s}.$$

The intercept is directly read from the intersection of the line with the y-axis to be -0.68 cm.

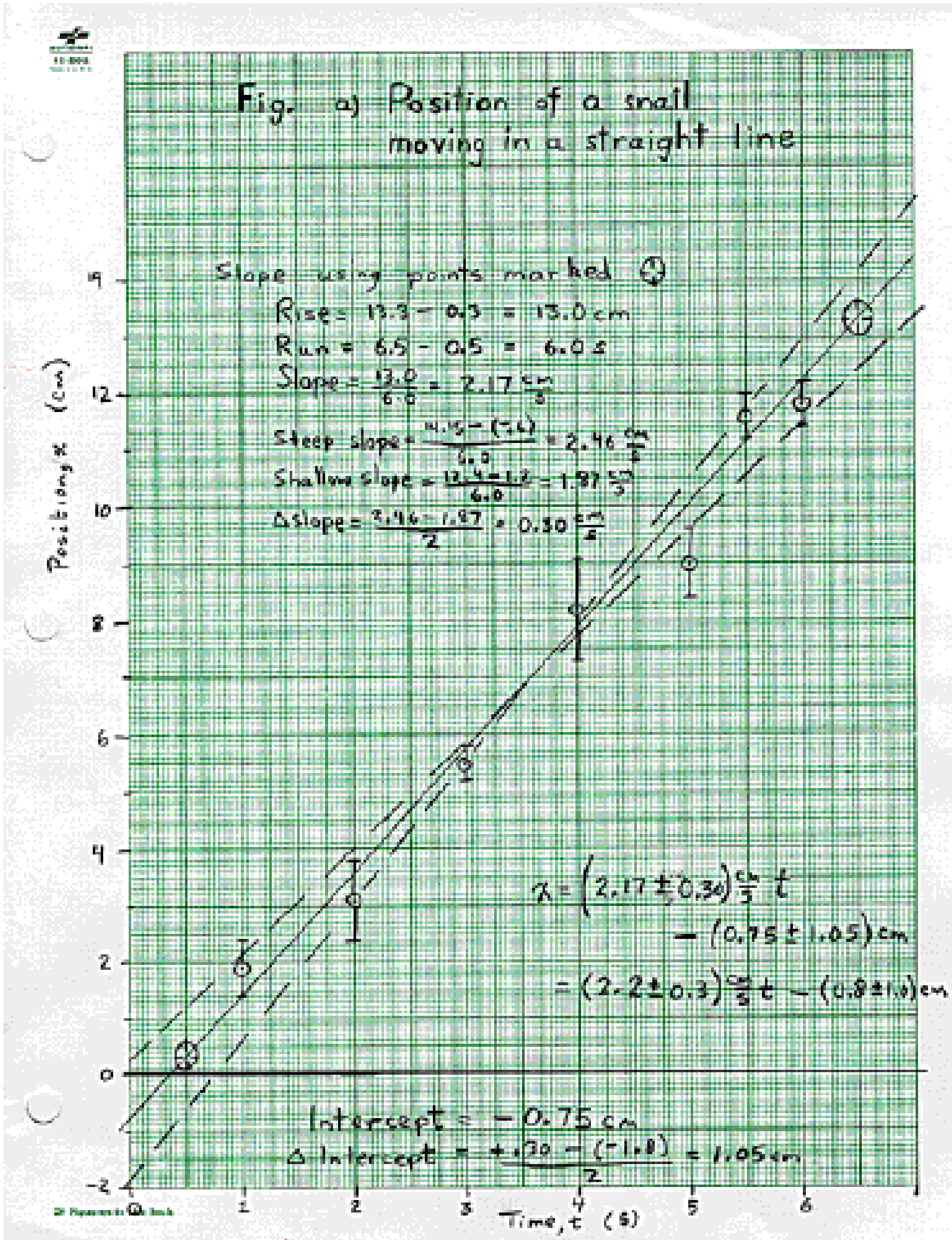

Figure 5: A hand drawn graph of data on the motion of a snail. Notice the features of a well-drawn graph including title, axis labels, tick marks, circled data points, and error bars.

Using the error bars as a guide we have drawn dashed lines which conceivably fit the data, although they are too steep or too shallow to be considered best fits. This is a judgment call on your part.

The slopes of the dashed lines (calculations on the graph) are 2.32 cm/s and 1.79 cm/s. Half the difference of these is 0.27 cm/s which we take as the uncertainty in the slope of the best line. We round off the uncertainty to the proper number of significant figures, and round the slope to match, resulting in

slope = (2.1 ± 0.3) cm/s.

The differences between the best slope and either of the extreme slopes should equal the uncertainty in the slope. Here the differences are (2.09 - 1.79) = 0.30 cm/s and (2.32 - 2.09) = 0.23 cm/s, which are basically the same as the 0.3 cm/s above.

We try to make the three lines cross in the middle of our data. If we draw them this way we can determine the uncertainty in the intercept. The dashed lines have intercepts of -1.52 cm and $+0.20$ cm and half of the difference between these is -0.86 cm which we use as the uncertainty in the intercept.

Intercept = (-0.7 ± 0.9) cm.

It is more difficult to do this on the computer graph, but we can try as is suggested on Figure 6. While Excel will compute the slope and intercept for the best fit line (they call it a trend line), the slopes of the steep and shallow lines must be computed by hand. The results are the same as those from the hand-drawn graph.

2.2 Uncertainties in Slope and Intercept When There Are No Error Bars

Even if we lack error bars we use the same approach to find the errors in slope and intercept. Using this method it is possible to get good estimates of uncertainty in the slope and intercept. Generally you will have less confidence in the intercept uncertainty.

2.3 What is being done in statistical terms

The process described in Sections 2.1 and 2.2 estimates the statistical procedure of finding standard errors in the slope and intercept. Statistics programs will allow this to be done automatically (in Excel see the LINEST function). The values of uncertainties you get by visual estimation will be similar to the values obtained by a full regression analysis.

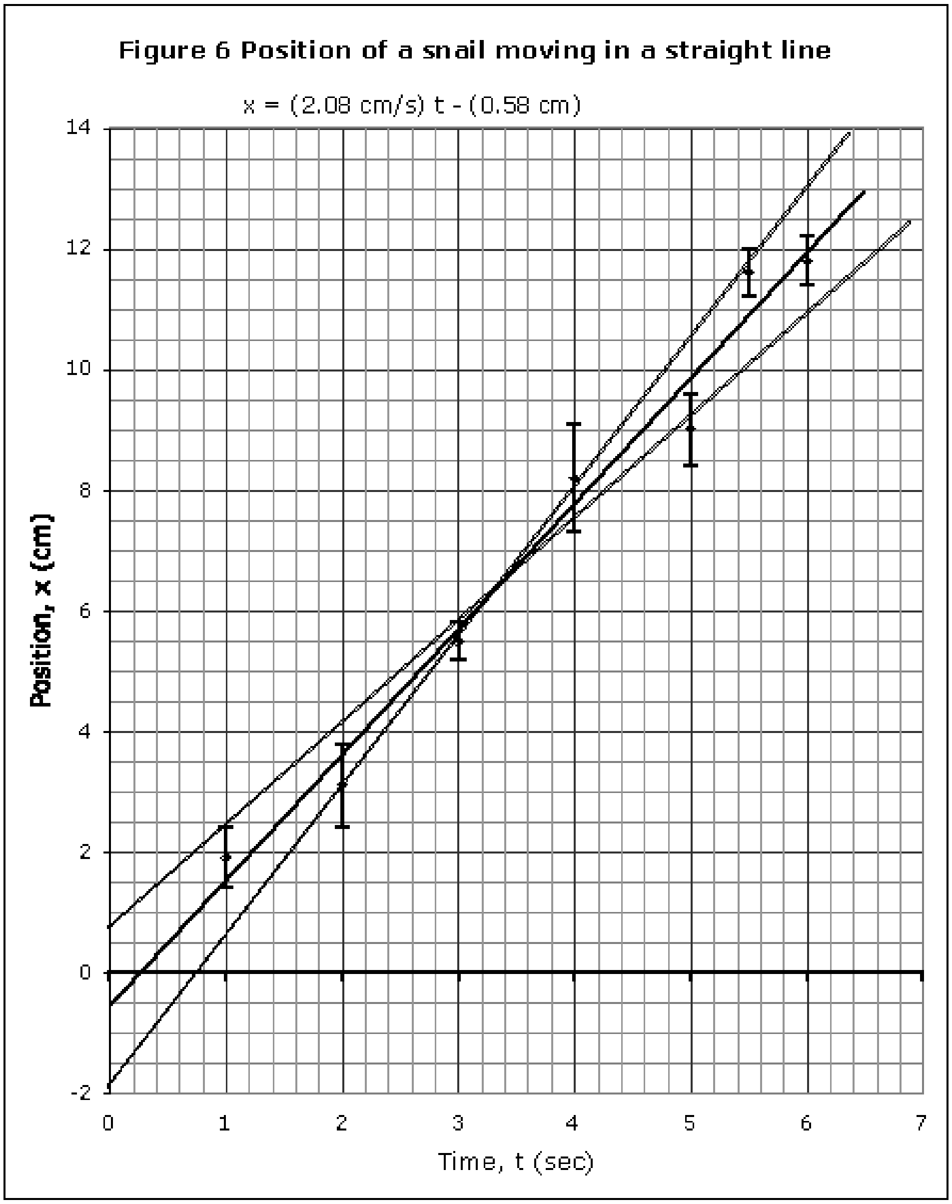

Figure 6: A computer generated graph of the same data as are used in Figure 5. The best-fit line is generated within Excel (Insert trend line) with its equation displayed. The steep and shallow lines are added using the drawing palette in Excel, good titles and axis labels are used.

2.4 Another example of estimating uncertainty of a slope on a graph—M. Richmond

If one has more than a few points on a graph, one should calculate the uncertainty in the slope as follows. In Figure 7, a graph of data collected in an electrical experiment are plotted. The data points are shown together with error bars to indicate the uncertainty in each measurement. It appears that current is measured to ± 5 milliamps, and voltage to about ± 0.2 volts.

A best fit line is drawn, and its slope determined using the coordinates of the points marked with hollow triangles. Note that these are not data points, but they are picked near the ends of the line.

$$\text{Best slope} = \frac{147\ \text{mA} - 107\ \text{mA}}{10.0\ \text{V} - 4.5\ \text{V}} = 7.27\ \text{mA/V}$$

Next a line is drawn with as small a slope as you decide is reasonable that still fits the data, considering the error bars. The hollow triangles on this line are used to find the slope.

$$\text{Shallow slope} = \frac{145\ \text{mA} - 115\ \text{mA}}{10.5\ \text{V} - 5.0\ \text{V}} = 5.45\ \text{mA/V}$$

Finally a line is drawn with as steep a slope as you decide is reasonable that still fits the data, considering the error bars. The hollow triangles on this line are used to find the slope.

$$\text{Steep slope} = \frac{152\ \text{mA} - 106\ \text{mA}}{10.0\ \text{V} - 5.0\ \text{V}} = 9.20\ \text{mA/V}$$

The uncertainty in the slope is one-half of the difference between max and min slopes,

$$\frac{9.20 - 5.45}{2} = 1.88\ \text{mA/V}$$

So keeping in mid that we want to round the uncertainty to one significant figure, the slope of this line is (7 ± 2) mA/V. Your instructor may prefer 2 sig fig in the uncertainty in which case the slope is (7.3 ± 1.9) mA/V.

3 Curve Fitting

We are free to make many plots from a given set of data. For instance if we have position x as a function of time t we can make plots of x versus t, x versus t^2, $\log x$ versus t, or any number of any choices.

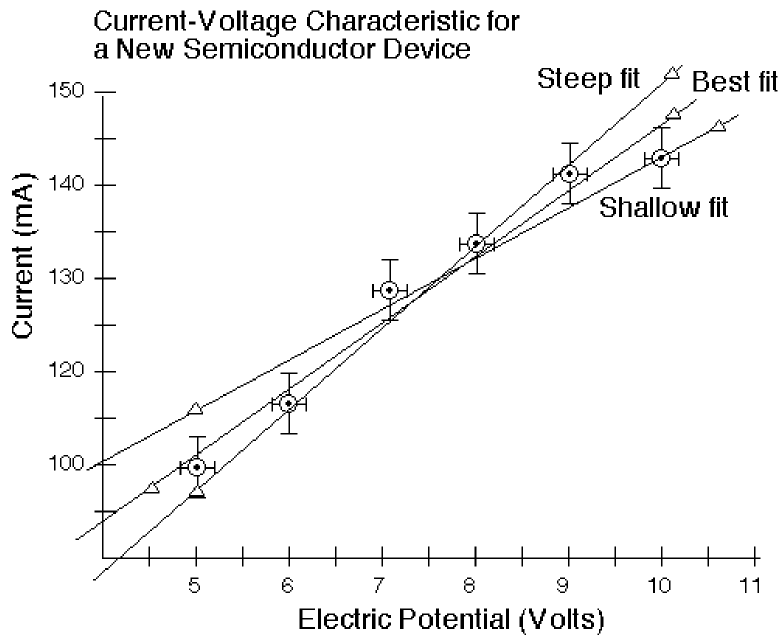

Figure 7: A graph with extra lines to allow estimating uncertainty in the slope.

If possible, we choose our plot so that it will produce a straight line. A straight line is easy to draw, we can quickly determine slope and intercept of a straight line, and we can quickly detect deviations from the straight line.

Table 1: Different graphs for different functions. This summarizes some of the most common mathematical relations and the graphing techniques needed to find slopes and intercepts. Values marked with * require special techniques that will be discussed when such graphs are required.

Form	Plot to yield straight line	Slope	Intercept
$y = ax + b$	y vs x linear paper	a	b
$y^2 = cx + d$	y^2 vs x linear paper	c	d
$y = ax^m$	$\log y$ vs $\log x$ linear paper	m	$\log a$
$y = ax^m$	y vs x log-log paper	m^*	a at $x = 1$
$xy = K$	y vs $1/x$ linear paper	K	0
$y = ae^{bx}$	$\ln y$ vs x linear paper	b	$\ln a$
$y = ae^{bx}$	y vs x log-linear paper	b^*	a

If we have the guidance of a theory we can choose our plot variables accordingly. If we are using data for which we have no theory we can empirically try different plots until we arrive at a straight line. Some common functions and suggested plots are in Table 1.

4 More information

The website at `http://www.rit.edu/cos/uphysics/graphing/graphingpart1.html` has more detailed information, including examples of bad graphs. It also has a link to a sample Excel sheet explaining how to produce graphs in Excel that meet scientific criteria.

Another good reference is the interactive simulation at `http://phet.colorado.edu/simulations/sims.php?sim=Curve_Fitting` that can run in your browser ("Run Now") or downloaded ("Run Offline".)

Check the "Adjustable Fit" box and place the following points on the graph. The linear fit makes a line $a + bx$, and you have two sliders for the values of a and b.

Get the best fit by picking values that minimize the χ_r^2 bar on the left of the animation. Record the values of a and b.

Now change the slope to a steeper value so that the minimum χ_r^2 is 10. Record these values and subtract the best fit values to estimate the uncertainty in slope and intercept.

x	y	Δy
-14	-4.7	0.8
-10	0	0.8
-6	0.7	0.8
1	7	0.8
11	12.5	0.8