Appendix VII. Graphs

Revised April 21, 2006

A. Introduction

Although most graphs are now produced using computer software, you must still understand the basic principles involved in making a good graph. Graphical representations should illuminate relations between variables and present results to the viewer in a readily understandable form. A good graph makes information much easier to understand than does a long block of text.

B. Some Suggestions

a. Make large graphs.

Hand-drawn graphs of laboratory data should be no smaller than half a notebook page. If you will be performing data analysis directly from the graph (*e.g*., *determining slopes and intercepts*), the graph should be full-page size. Computer-generated graphs can be smaller but should be large enough to display all important details.

b. Choose and label the axes carefully.

The independent variable is usually plotted along the *x*-axis (the *abscissa*) with the dependent variable along the *y*-axis (the *ordinate*). Both axes should be labeled and include units, *e.g.* "LENGTH (cm)".

c. Choose convenient scales and intervals.

One often must read numbers from the graph, so choose starting points and intervals which are easy to interpret. For example, if you are plotting a distance measurement, you might choose the axis scale so that 1 cm = 1, 2, 5 or 10 units, but not 3.33 units. Try to scale your graph so that variations in the data are clear and error bars are big enough to be easily drawn and seen.

d. Plot data points with error bars.

In most experiments, the major uncertainty is in the *y*-axis variable and only the *y*-axis error bars are shown. However, if the *x*-axis values also have important uncertainties, then these too are shown on the graph.

e. Put a title or caption on your graph.

A graph meant for stand-alone presentation (as part of an oral presentation or poster paper) should have a meaningful title. (*x-axis variable vs. y-axis variable is not a meaningful title!*) Graphs in written papers must have a caption describing in a sentence or phrase what is being presented. It may also be appropriate to include a meaningful title.

C. Slope of a Straight Line

Suppose you expect your data to fall on a straight line, *i.e*., to follow the function

$$f(x; A, B) = A + Bx \qquad (1)$$

as illustrated in Figure 1. Note that, although the data are expected to obey a linear relationship, the points are scattered about a straight line because of the measurement uncertainties. Your object is to estimate the best straight line from these data and to obtain the intercept *A* and slope *B* of that line and their uncertainties. The best way to solve this problem would be to use a least-squares

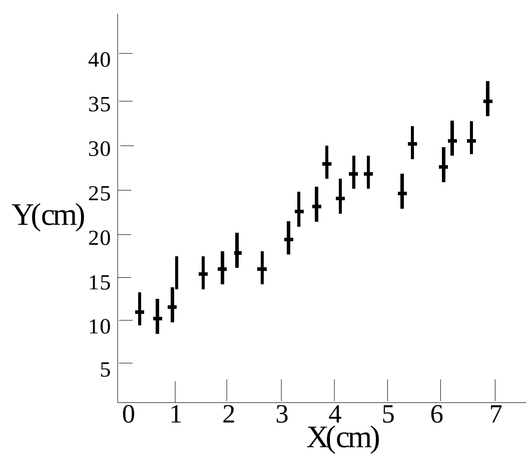

Figure 1: Linear Data

fitting routine (*see Appendix VIII*). However, it is useful to understand how to estimate a slope and the uncertainty in its value without resorting to using a computer.

For a manual estimate of the slope and its uncertainty, attempt to bracket the correct line by two lines that represent your estimates of one standard deviation steeper than and one standard deviation flatter than the true slope. Note that these are not the "maximum" and "minimum" possible slopes (which would be a vertical and a horizontal line, respectively). Rather, they are our attempt to estimate lines which, allowing for the uncertainties, would bracket about two-thirds of the data points. Figure 2 indicates a possible choice of lines, *a* and *b,* with slopes one standard deviation smaller and one standard deviation larger, respectively, than a best-fit line.

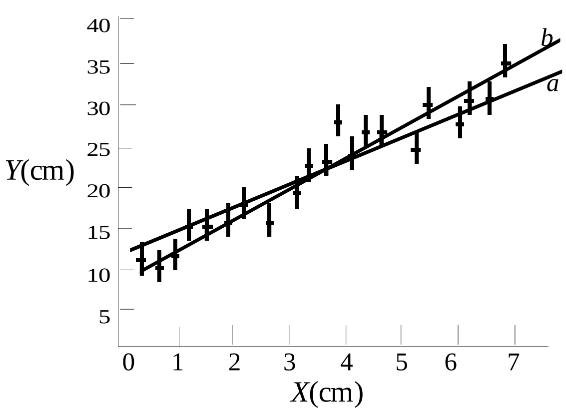

Figure 2: Linear Data with Slope Bracketed

The next step is to find the slopes B_a and B_b. The best estimate of the slope is then the mean of the two

$$B=(B_a + B_b)/2$$

and the estimate of the uncertainty in B is

$$\delta_B=(B_a - B_b)/2.$$

A similar procedure can be used to obtain the intercept.

D.Histograms

A *histogram* or bar graph shows the frequency with which particular values of measurements are obtained. For example, suppose that we have made repeated measurements of the time it takes a ball to fall from a height of 3 meters, and have obtained the data listed in Table 1. We want to convert those data to a histogram. Computer routines are available to calculate and produce histograms, however, you should be familiar with the histogramming process and in particular should know how to select reasonable ranges and bin sizes for your data.

To make a histogram, follow this procedure.

a. Select the ranges and binning interval.

The interval should be such that your data are distributed over several bins, never just 1, and preferably 4 or more, but there should not be so many bins that there are only 1 or 2 measurements in each bin. We illustrate histograms of Table 1 data with two different bin widths, Δt=0.2 s in Figure 3 and Δt=0.1 s in Figure 4.

Table 1 Times (s)
0.99493
0.79048
1.04097
0.60052
1.08496
0.70595
0.98048
0.74406
0.78633
0.8182
0.92053
0.61908
0.65506
0.73261
0.25645
0.9899
0.62636
0.85021
0.68984
0.76326
0.70103
1.01237
0.56961
0.58217
0.62454

Table 1: Times for ball to drop 3 m

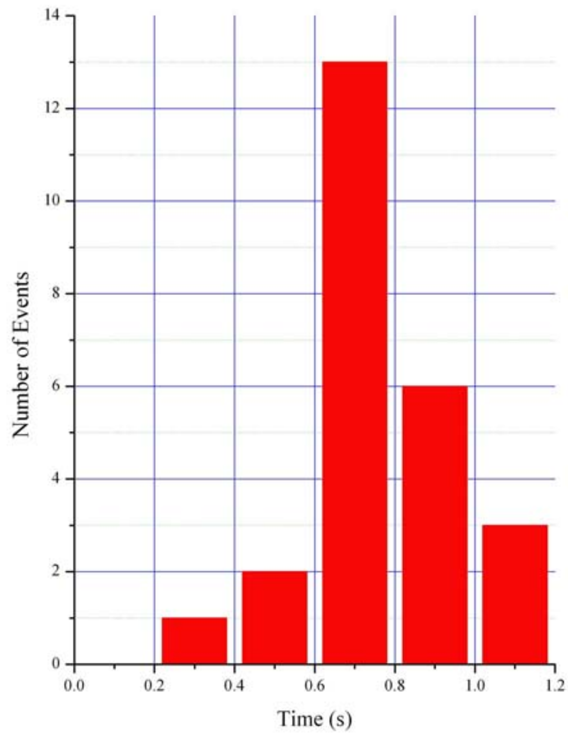

Figure 3 Histogram with Δt=0.2 s.

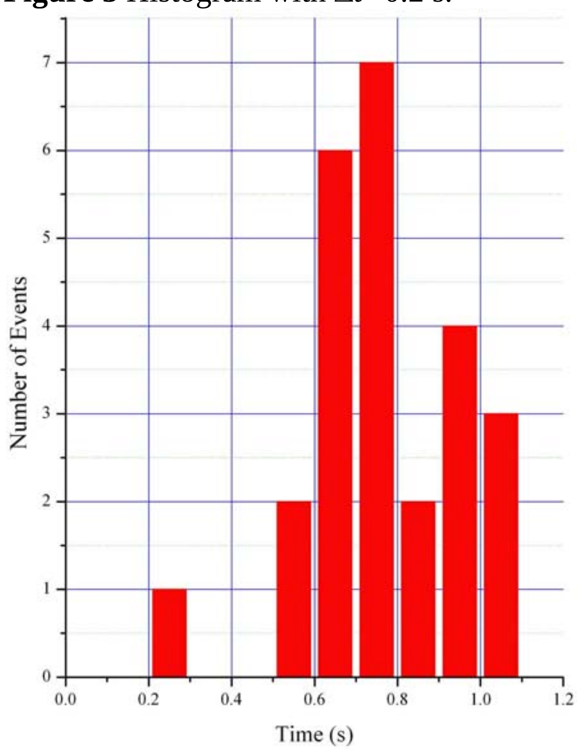

Figure 4 Histogram with Δt=0.1 s.

Choose ranges carefully so that they can be represented by rational numbers. Do not blindly choose your smallest and largest measurements to define the range of the abscissa of your histogram, but choose rounded-off values that will "catch" most of the data. It may be that a few stray measurements will fall outside your chosen range.

b. Mark the ordinate.

Estimate the maximum number of measurements you expect in a bin and choose a scale such that the histogram will be a reasonable size.

c. Bin the data.

Work through your data table, one measurement at a time, and assign each measurement to its particular bin. It is conventional to define the bins to contain variables with values greater than or equal to the lower bin limit and less than the upper limit.

d. Plot bars.

Indicate the number of measurements that fall into each bin by drawing a vertical bar. The result should look something like Figure 3 or 4.

e. Uncertainties in Bin Contents

The number of counts recorded in a histogram bin obeys the Poission distribution (see *Appendix VI*) so that the uncertainty is just the square root of the number of counts. Thus, the peak bin of Figure 3 has 13 ± 4 counts with a relative uncertainty of 4/13 ≈ 30%, while the peak bin of Figure 4 has 7 ± 3 counts with a relative uncertainty of ~40%. These calculations illustrate a disadvantage of reducing the bin size; while smaller binning may reveal more detail in a distribution, it also increases the relative uncertainty in the bin counts.

f. Area of graph

The effective area of the histogram is equal to $N\Delta t$ where N is the total number of individual measurements and Δt is the bin width. To plot on your histogram a curve that has been calculated with unit normalization (area = 1), you must scale each calculated point by $N\Delta t$ so that the area under the curve will equal the area under the histogram.

Figure 5 illustrates the histogram of Figure 4 with a calculated Gaussian curve based on the mean (μ=0.7656) and standard deviation (σ=0.1903) of the data in Table 1. The equation for the Gaussian curve (*Appendix VI*, Equation 1) has been multiplied by $N\delta t$ with N=25 and Δt=0.1 so that the area on the graph under the curve is the same as the area under the histogram.

Note that a curve on a histogram is always calculated at selected values of the abscissa of the graph, and not at individual measurement values such as those listed in Table 1.

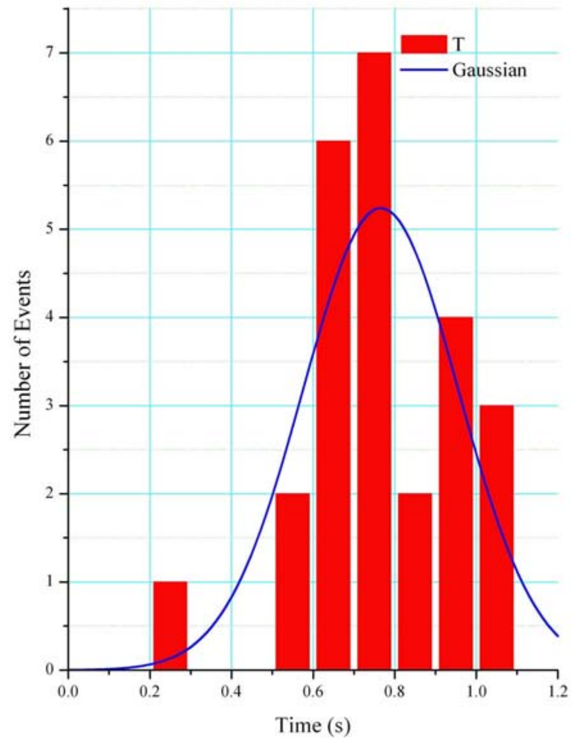

Figure 5: Histogram with Gaussian Curve