ERROR ANALYSIS

PURPOSE:

To understand terminology and concepts used in measurement.
To apply these concepts in a simple experimental situation.
To become familiar with the use of a spreadsheet.

READINGS:

Baird, 2.1 to 2.7, 2.11, and 3.1 to 3.11

APPARATUS:

Gate generator
Timer

INTRODUCTION:

For a scientist to arrive at a valid conclusion in testing a theory or hypothesis it is necessary to understand the underlying concepts of measurement errors. In engineering and manufacturing these same concepts are important in designing and making quality products.

Systematic errors are those which would tend to reproduce the same incorrect answer if the experiment were repeated using the same techniques. Instruments can generate systematic errors by erroneous design or construction. A ruler with a worn end or a voltmeter whose pointer has been bent are simple examples. Or, the experimenter can introduce a certain bias. The accuracy of an experiment is limited by systematic errors. You cannot improve the accuracy of an experiment by repeating it a number of times and taking the average; the systematic error will not average away.

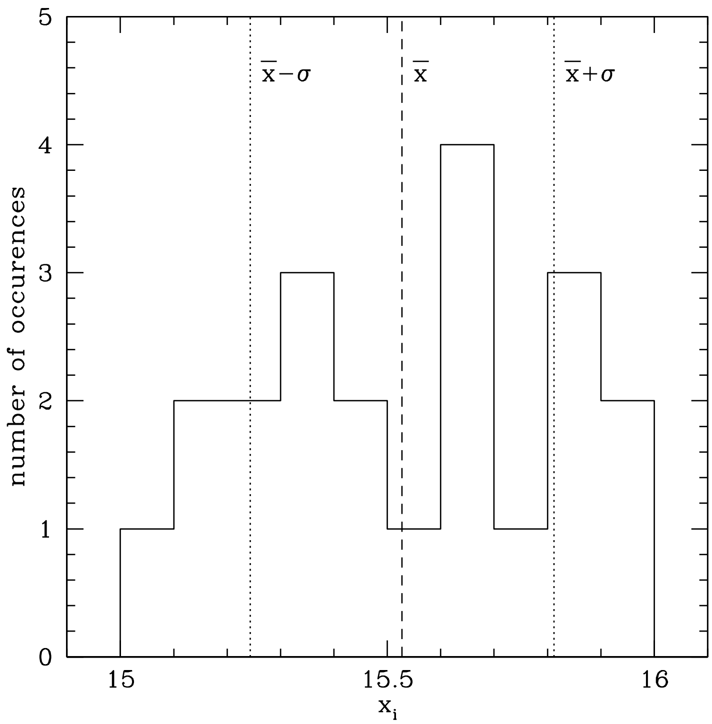
A random error is one which tends to produce different results when an experiment is repeated using the same technique. The average of a large number of such repetitive measurements will reduce random error by averaging out the variation: the measured value will sometimes be above the actual value and sometimes below. Experimental precision is limited by random errors.

MATHEMATICAL DESCRIPTION OF RANDOM ERRORS

The first important concept is the average (arithmetic mean) of a number of measurements:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

This formula says: Add the N measurements x_1, x_2, x_3, etc., up to x_N . This sum is written as $\sum_{i=1}^{N}$. Now divide by N to get the average value of x, namely $\overline{x}$.



Bin	# in Bin
14.90 – 15.00	0
15.00 – 15.10	1
15.10 – 15.20	2
15.20 – 15.30	2
15.30 – 15.40	3
15.40 – 15.50	2
15.50 – 15.60	1
15.60 – 15.70	4
15.70 – 15.80	1
15.80 – 15.90	3
15.90 – 16.00	2
16.00 – 16.10	0

FIGURE 1.

Let us plot a "histogram" to display some measurements (Fig. 1, above) (actual data values give in Table 1, below). This histogram displays how many times a measurement lies between 15.00 and 15.10 (once), between 15.10 and 15.20 (twice), etc. The various intervals, *e.g.*, the one from 15.00 to 15.10, are called "bins". The appearance and usefulness of a histogram is strongly influenced by the size of a bin. It is desirable to have a large bin size in order to make the number per bin (n) large. The fluctuations in "n" from bin to bin will then be relatively small. On the other hand it is desirable to have the bin size small enough so as not to obscure the shape of the variation of "n". It is an art to picking the bin size that best displays your experimental results. If the measurement were repeated many times and a "histogram" plotted, but with much smaller intervals, the diagram shown in Fig. 2 (next page) might result. Note that as the number of measurements increases, the histogram becomes smoother. With still more measurements a histogram of data with only random errors approaches the "Gaussian" or "normal" distribution shown in Figure 2. The Gaussian function is given by:

$$P_G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

In this plot the vertical scale is the number of times a given value of x is measured. The horizontal scale is the value of x expressed in units of σ, the standard deviation, which is a measure of the spread of the measured values of x. The mathematical expression for σ is given by:

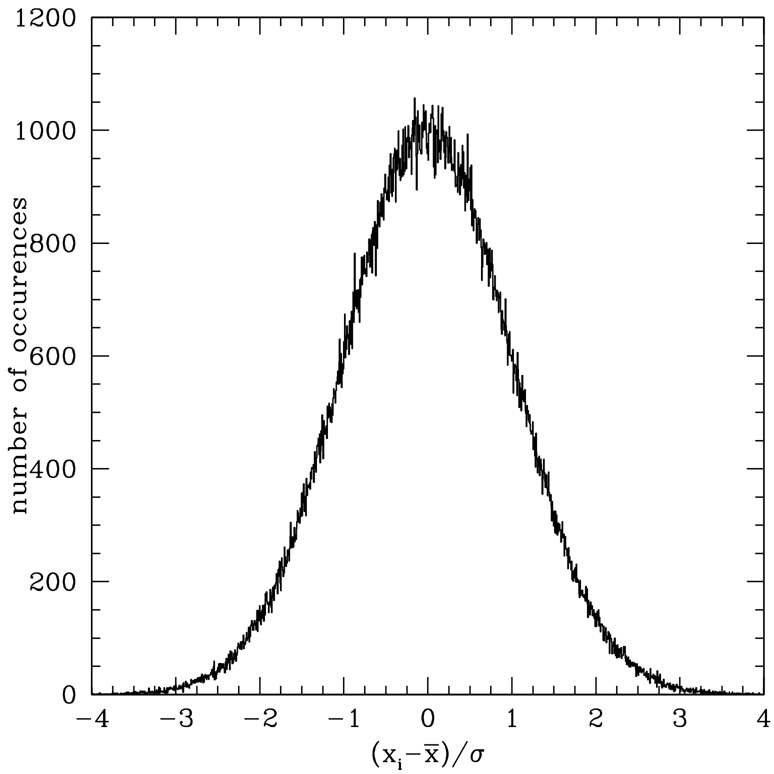


FIGURE 2.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} \tag{3}$$

A large value for σ implies a large data spread, whereas a smaller value implies a small spread. **σ gives an estimate of the uncertainty of a single measurement**, x_i. It can be shown that the area under the normal curve centered around the mean between $\overline{x} - \sigma$ and $\overline{x} + \sigma$ is 68.3% of the total area under the curve. This means that about two-thirds of the measurements, x_i, fall between $\overline{x} - \sigma$ and $\overline{x} + \sigma$. Also, the area of the curve between $\overline{x} - 2\sigma$ and $\overline{x} + 2\sigma$ is 95.4% of the total area. And, $\pm 3\sigma$ covers 99.7%. Notice in the histogram is Fig. 1 that 12 out of the 21 points, or 57%, of the measurements fall in the range $\overline{x} - \sigma$ to $\overline{x} + \sigma$. You would expect to get exactly 68.3% only if a very large number of measurements were made. As an example of how you would use this concept, suppose we write 15.5 ± 0.2 m for a single measurement, then this implies that the probability is about 68% that the true value lies between 15.3 and 15.7.

There is a second closely related quantity – $\overline{\sigma}$, the standard deviation of the mean of the measurement of $\overline{x}$ – that is frequently confused with σ. **$\overline{\sigma}$ gives an estimate of the uncertainty in the measurement of the mean**. Mathematically, $\overline{\sigma}$ is given by:

$$\overline{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N(N-1)}} \tag{4}$$

[For simplicity some authors use N instead of $N - 1$ in the denominators. For large values of N the difference is minimal.] Comparing Eq. 3 with 4, we see that $\overline{\sigma} = \sigma/\sqrt{N}$. Notice that σ is always larger than $\overline{\sigma}$. It is also true, but perhaps not so obvious, that no matter how large N becomes, the value of σ will not get smaller. It will fluctuate about some value due to the random errors, but will not get smaller. This is reasonable, since the uncertainty in any one measurement cannot depend on how many times before you've made the measurement before. $\overline{\sigma}$, on the other hand, will decrease as N increases – to decrease the error in the average, you simply make more measurements. The standard deviation σ is used as a measure of the error (uncertainty) expected for an individual measurement. In most science and engineering applications, you will need to calculate $\overline{x}$, the mean, and $\overline{\sigma}$, the standard deviation of the mean.

To illustrate the procedure we will work out the average (mean) value $\overline{x}$ and the standard deviation of the mean, $\overline{\sigma}$, and the standard deviation of an individual data point, σ, using the position measurements in the accompanying Table 1.

TABLE 1. Position Measurements

x_i (m)	$x_i - \overline{x}$ (m)	$(x_i - \overline{x})^2$ (m^2)
15.68	0.15	0.0225
15.42	−0.11	0.0121
15.03	−0.50	0.2500
15.66	0.13	0.0169
15.17	−0.36	0.1296
15.89	0.36	0.1296
15.35	−0.18	0.0324
15.81	0.28	0.0784
15.62	0.09	0.0081
15.39	−0.14	0.0196
15.21	−0.32	0.1024
15.78	0.25	0.0625
15.46	−0.07	0.0049
15.12	−0.41	0.1681
15.93	0.40	0.1600
15.23	−0.30	0.0900
15.62	0.09	0.0081
15.88	0.35	0.1225
15.95	0.42	0.1764
15.37	−0.16	0.0256
15.51	−0.02	0.0004

From the above table we can make the following calculations:

$$N = 21 \qquad \sum_{i=1}^{N} x_i = 326.08\,\mathrm{m} \qquad \sum_{i=1}^{N} (x_i - \overline{x})^2 = 1.61998\,\mathrm{m}^2.$$

and then evaluate the following quantities:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{326.08}{21} = 15.53\,\mathrm{m}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} = \sqrt{\frac{1.6201}{20}} = 0.28\,\mathrm{m}$$

$$\overline{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N(N-1)}} = \sqrt{\frac{1.6201}{21(20)}} = 0.06\,\mathrm{m}.$$

The error or spread in individual measurements is $\sigma = 0.28$ m. But for the mean $\overline{x} \pm \overline{\sigma} = 15.53 \pm 0.06$ m. This says the average is 15.53 m which has an error of 0.06 m. Or putting it another way, there is about a 68% probability that the true value of x falls in the range 15.47 m to 15.59 m. In some cases the fractional error $\sigma/\overline{x}$, or **relative error**, is of more interest than the absolute value of σ. It is possible that the size of σ is large while the fractional error is small. Note that increasing the number of individual measurements on the uncertainty of the average reduces the statistical uncertainty (random errors); this improves the "precision". On the other hand, more measurements do not diminish systematic error in the mean because these are always in the same direction; the "accuracy" of the experiment is limited by systematic errors.

In today's experiment you will compare three data sets measuring reaction times – two sets will be on your own reaction time and one on your partner's. You will determine whether the last two data sets are significantly different from the first. In order to be clear about the purpose of the experiment, let's go through the reasoning for a concrete example unrelated to the experiment:

Suppose the students in a class are randomly assigned to one of three groups, each with N students, and two groups are taught some new material using one teaching technique, while the third group is taught the same material using a different technique. The three groups are then given the same exam on the material. We'll call the average exam scores for the three groups $\overline{t_1}$, $\overline{t_2}$, and $\overline{t_3}$, and the standard deviations of the means $\overline{\sigma_1}$ $\overline{\sigma_2}$, and $\overline{\sigma_3}$.

We would expect the difference to be "small" since the same the two groups were taught the same way and we would like to attribute the difference to just random measuring errors. On the other hand, in order to say that the two different teaching methods produce different learning results we need to be able to say that $\mid \overline{t_1} - \overline{t_2} \mid$ is "large". But small or large compared to what? The answer is – small or large compared to the error in determining $\mid \overline{t_1} - \overline{t_2} \mid$ (or $\mid \overline{t_1} - \overline{t_3} \mid$), which we'll call Δ_{12} (or Δ_{13}). You calculate these errors from the standard deviations of the mean as follows: $\Delta_{12} = \sqrt{\overline{\sigma_1}^2 + \overline{\sigma_2}^2}$ and $\Delta_{13} = \sqrt{\overline{\sigma_1}^2 + \overline{\sigma_3}^2}$. If groups 1 and 2 are not different then there is a 68.3% probability that $|\ \overline{t_1} - \overline{t_2}\ |$ will be less than or equal to Δ_{12}, a 95.4% that it will be less than/equal to $2\Delta_{12}$, and a 99.7% that it will be less than/equal to $3\Delta_{12}$. The scientific convention is to say **two measurements are significantly (or statistically) different if they differ by three standard deviations or more**. Thus we say that groups 1 and 2 are not significantly different provided $|\ \overline{t_1} - \overline{t_2}\ | < 3\Delta_{12}$. Likewise, groups 1 and 3 are significantly different if $|\ \overline{t_1} - \overline{t_3}\ | \geq 3\Delta_{13}$. Notice that Δ_{13} (like $\overline{\sigma_1}$ and $\overline{\sigma_3}$) gets smaller as N, the number of students, increases. Thus if one teaching technique produces a large improvement in learning, you will only have to try the experiment out on a small group of students to prove there is a significant difference. If there is only a small difference in learning, you will have to use a very large N to determine if the difference is significant.

PROCEDURE

In this experiment you will take two sets of data on your reaction time and compare the two sets to see if there is a significant difference between them. Then you will compare one set of your data with that of your partner. We will provide you with an automatic light and timer. The light will flash on after being off for a random time. The light starts a digital timer with 0.001 second resolution. You stop the timer by depressing the switch as soon as you see the light. This turns off the light. The timer will then show your reaction time – the time it takes for you realize that the light has come on and to react by pushing the stop button. There is a reset button to zero the display after you record the time. After a random time (on the order of 10 seconds) the flasher once again turns on the light and starts the clock.

1. Practice before you start taking real data. Operate in teams of two people, with one watching the light and depressing the stop switch while the other records the data.

2. Make 2 sets of 10 reaction time measurements. Do not record any that are ridiculously long because you were asleep at the switch (*i.e.*, an obvious "mistake"). Calculate your average reaction times, $\overline{t_1}$, $\overline{t_2}$, and the standard deviations, σ_1, σ_2.

3. The other lab partner should now repeat the above to obtain another set of ten measurements. Calculate $\overline{t_3}$ and σ_3 of the new data.

Making Histograms in Excel Using the FREQUENCY Function

Histograms can be made using the FREQUENCY function in Excel. It has the form `frequency(data array, bins array)`. For example, `frequency(a1:a10,b1:b20)` would bin the data in a1 through a10 according to the bins listed in bins b1 through b20. In order to enter the function, highlight the area where the results should be displayed, type the frequency call desired (*e.g.*, `frequency(a1:a10,b1:b20)`), then hit Control-Shift-Enter (PC) to enter the call. For example, the bins array might be as follows: b1 = blank, b2=5, b3=6, b4=7. Highlight the region c1 to c4. After entering frequency, cell c1 will have the number of values in the data array less than or equal to 5, c2 will have the number greater than 5 and less than or equal to 6, c3 greater than 6 and less than or equal to 7, and c4 the number greater than 7.

Significant Figures

Be sure to read section 2.11 in Baird on significant figures and quote all results accordingly. In general, the uncertainty is quoted to only one significant figure unless that is a 1, in which case it is sometimes quoted to two (*i.e.*, 0.3, not 0.33, but 0.1 or 0.13 would be acceptable). The value should be quoted to the same precision as the error, or at most one more. For example, give 2.5 ± 0.4 not 2.543 ± 0.4 and not 2.5 ± 0.007, although 2.54 ± 0.1 or 2.54 ± 0.14 would be acceptable.