

# An Introduction to Error Analysis

Error analysis is the study and evaluation of uncertainty in measurement. Experience has shown that no measurement, however carefully made, can be completely free of uncertainty.

In science, we use the term “error” as being interchangeable with “uncertainty.” As such, errors are not mistakes; you cannot avoid them by being very careful. The best you can do is to (a) find reliable estimates of their size and (b) use experimental designs and procedures that keep them as small as possible.

## Reporting experimental results, part 1

When you report a measurement or experimental result in physics, it’s important to always include the uncertainty (as a “plus or minus” amount) as well as the value. For example, when I say that I weigh 150 lbs, I probably don’t mean “exactly 150 lbs” but instead “somewhere in the range of 145 to 155 lbs”, or  $150 \pm 5$  lbs. Of course, if I were paying close attention to my weight, “150 lbs” might mean somewhere between 149 and 150 lbs ( $150 \pm 1$  lbs). Or if I say “a newborn moose weighs about 150 lbs”, I might mean somewhere between 100 and 200 lbs ( $150 \pm 50$  lbs).

We often use significant figures to imply uncertainty. For example, a result of 12.4 Joules is usually interpreted as being between somewhere between 12.3 and 12.5. The general understanding is that the implied “plus or minus” amount is at least the size of the last decimal place. But it’s a much better practice to actually state the uncertainty, such as  $12.40 \pm 0.05$  J or  $12.4 \pm 0.2$  J, rather than let the significant figures “imply” an uncertainty. This way, the reader knows that you have thought about the uncertainty rather than just rounding your result to some arbitrary decimal place.

When reporting a value  $\pm$  uncertainty, the significant figures in the value (and the significant figures in the uncertainty!) should be consistent with the size of the uncertainty. For example, it wouldn’t make much sense to report “12.398318  $\pm$  0.05 Joules”, since the uncertainty makes all of the digits after 9 meaningless. It would also be incorrect to report “12  $\pm$  0.05 Joules”, because here the value is not stated with as much precision as the uncertainty allows. Instead, the result should be written as “12.40  $\pm$  0.05 Joules”. Another silly statement would be “12.398318  $\pm$  0.046252 Joules”, since the uncertainty itself is uncertain!

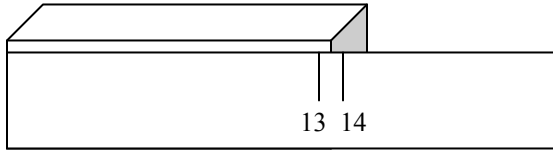
As a general rule, *the uncertainty should be rounded to one significant figure, and the value should be rounded to the same decimal place as the uncertainty* (or one place more, especially if the first digit of the uncertainty is small).

Examples:             $112.5 \pm 0.5$  lbs             $124 \pm 10$  Joules             $62,000 \pm 5000$  km

Since the units of the uncertainty must be the same as the units of the value, it is best to write the units at the end, as shown above.

## Estimating uncertainty by “eyeballing”

Sometimes the only way to determine the uncertainty of a measurement is to make an educated guess, or estimate. For example, let’s say you are using a ruler to measure the length of a wooden block. You might look at the scale on the ruler and decide that you can read it to the nearest half millimeter (see the figure)



So you would estimate the uncertainty to be 0.5mm and report that the block’s length is  $13.5 \pm 0.5$  mm.

But perhaps the block you’re measuring is kind of beat up and doesn’t have a nice even edge. In this case the uncertainty arises from the block itself rather than from your ability to read the scale. If you estimate that the block’s uneven edge varies by 2 mm, you would report that the block’s length is  $13.5 \pm 2$  mm.

Even when you’re reading a value from a digital measuring device, such as a stopwatch, estimating uncertainty is important and requires thought. If you can repeat a measurement several times, you can use the statistical methods described in the next section. If you only have one measurement, the best way to determine the uncertainty is to look up the specifications of the device, which may be in a spec sheet or printed on the device itself. Most digital devices round off their display in accordance with their uncertainty, but sometimes they’ll show more decimal places than their uncertainty justifies. If you have no written specs, you’ll have to make some assumption about how appropriately the display is rounded. For example, a reading of 50.74 seconds might imply an uncertainty of 0.005 seconds, and you could report  $50.740 \pm 0.005$  seconds. But note that this would be the very minimum estimate for the uncertainty, since it assumes appropriate rounding and only deals with the measuring device itself and not the person operating it. You may want to make a more conservative estimate, such as  $50.74 \pm .02$  sec., for example.

## Estimating uncertainty using statistical analysis

A better way to estimate uncertainty is to make multiple measurements of the same quantity and analyze the *dataset* using statistical functions. These are discussed in detail below.

However, not all types of experimental uncertainties can be assessed by statistical analysis based on repeated measurements. For this reason, uncertainties are classified into two groups: (a) *random* uncertainties, which can be treated statistically; and (b) *systematic* uncertainties, which cannot. Random uncertainties can be revealed by repeating the measurements; systematic uncertainties cannot.

To illustrate this distinction, let us consider an example. Suppose that we time the period (time for one revolution) of a steadily rotating turntable. One source of error will be our reaction time in starting and stopping the watch. If our reaction time were always exactly the same, these two delays would cancel one another. In practice, however, our reaction

time will vary. We may delay more in starting, and so underestimate the time of a revolution; or we may delay more in stopping, and so overestimate the time. Since either possibility is equally likely, both the size and the sign of the effect are *random*. If we repeat the measurement several times, we will sometimes overestimate and sometimes underestimate. Thus our variable reaction time will show up as a variation in the measured periods. By looking at the spread in the measured periods, we can get a reliable estimate of this kind of error.

On the other hand, if our stopwatch is running consistently slow, then all our times will be underestimates, and no amount of repetition (with the same watch) will reveal this source of error. This kind of error is called *systematic*, because it always pushes our result in the same direction. Systematic errors cannot be discovered by the kind of statistical analysis that we will be discussing below—in fact, they are often hard to evaluate or even to detect. In our introductory lab we will often (but not always) assume that systematic errors have been made much smaller than the required precision (“negligible”).

### True values and best estimates

It is important to realize that the “true” period ( $T$ ) of the rotating turntable above can never be known. Thus, the goal of statistical analysis is to obtain the best estimate ( $T_{\text{best}}$ ) of the true value along with a best estimate of how close  $T_{\text{best}}$  is likely to be to the true value.

### Mean: best estimate of the true value

Suppose we make  $N$  measurements of a quantity  $x$  and get the values  $x_1, x_2, \dots, x_N$ . The best estimate  $x_{\text{best}}$  of the true value is the *mean* or *average*, defined as:

$$x_{\text{best}} = \bar{x} = \frac{\sum x_i}{N}$$

### Standard deviation SD

The best estimate of the uncertainty in the individual values  $x_i$  is the *sample standard deviation*  $\sigma_x$  (or **SD**), defined as:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (d_i)^2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

The term  $d_i$  in this equation, called the *deviation*, is simply the difference between the  $i^{\text{th}}$  measurement  $x_i$  and the mean value  $\bar{x}$ . If the deviations are all very small, then our measurements are all close together and are said to be precise.

To be sure we understand the idea of a deviation, let us calculate the deviations for the set of five measurements reported in the table below.

**Table 1. Calculation of Deviations**

Trial number, $i$	Measured value, $x_i$	Deviation, $d_i = x_i - \bar{x}$
1	71	-0.8
2	72	0.2
3	72	0.2
4	73	1.2
5	71	-0.8
	$\bar{x} = 71.8$	$\bar{d} = 0.0$

Notice that some of the deviations are positive and some are negative. In fact, as Table 1 confirms, the average of the deviations is always zero. This is why the standard deviation is found by first squaring the deviations, then averaging these positive squares (using  $N-1$  rather than  $N$ ), and finally taking the square root of the result.

For the five measurements in Table 1, the standard deviation  $\sigma_x$  is found to be:

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (d_i)^2} = \sqrt{\frac{1}{5-1} \sum_{i=1}^5 (x_i - \bar{x})^2} = \sqrt{\frac{1}{4} (0.64 + 0.04 + 0.04 + 1.44 + 0.64)} \approx 0.84$$

Note that this value is bigger than the absolute value of some of the deviations in Table 1 and smaller than others—in other words, it can still be interpreted loosely as the “average” deviation.

### Graphical interpretation of the standard deviation

Suppose we made ten measurements of some length  $x$  and obtained the values (all in cm):

26, 24, 26, 28, 23, 24, 25, 24, 26, 25

A convenient way to organize this dataset is shown in Table 2. This is known as the *frequency distribution*.

**Table 2**

Measured value	23	24	25	26	27	28
Number of measurements	1	3	2	3	0	1

The frequency distribution of our measurements can be graphically displayed in a histogram as shown in Figure 1. Here  $F_k$  is the number of measurements and  $x_k$  is the measured value.

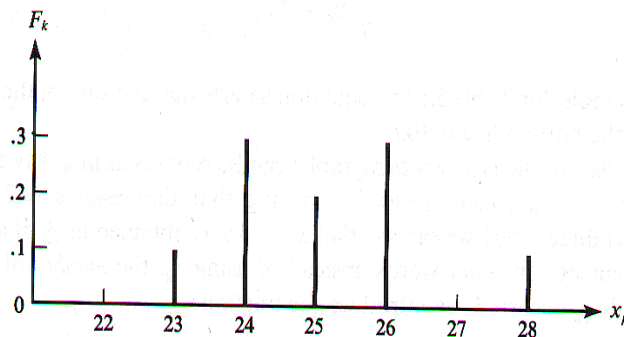


Figure 1

If we could increase the number of measurements (ideally to infinity!), then the histogram would become a bell-shaped curve like those shown in Figure 2.

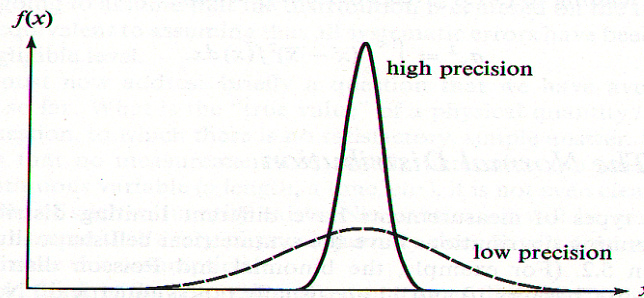


Figure 2

For most types of random errors, the mathematical function that describes this curve is the *Gaussian* or “*normal*” distribution:

$$f(x) \propto e^{-(x-\bar{x})^2/2\sigma^2}.$$

Note that this function is characterized by a *true mean*  $\bar{x}$ , which tells us where the peak is, and a *true standard deviation*  $\sigma$ , which tells us how wide the curve is. A good approximation is that the width of the curve at  $\frac{1}{2}$  the peak height is about  $2\sigma$ .

The true mean in this function is not the same as the sample mean of a finite number of measurements, because of the presence of random errors. (Also, the true mean is generally not equal to the true value we are seeking, because of the presence of systematic errors—but it is the best we can hope for using statistics!)

### Standard error SE: best estimate of the uncertainty

The sample standard deviation tells us the spread in our measured values, but it does not tell us how close our sample mean is likely to be to the true mean. To understand this, consider the following observation: if many different people performed the same set of measurements, each would obtain a different sample mean, and a different sample standard deviation. However (and this is the crucial point), there would be a narrower spread between the sample means than between the measurements in each of the datasets. The standard deviation of the set of sample means would be the best estimate of the uncertainty of an individual sample mean. Graphically, the frequency distribution of such a set would nearly always be a bell-shaped *normal distribution*.

Theory tells us that even when we have only a single sample mean, we can estimate its uncertainty by calculating the *standard error* (SE), defined as:

$$SE = \text{uncertainty of the mean} = SD / \sqrt{N}$$

## Reporting experimental results, part 2

### Confidence levels

In part 1 we discussed the importance of reporting an experimental result in the form “value  $\pm$  uncertainty”. We now know that the best value is the mean so the result is reported as:

$$\textit{mean} \pm \textit{uncertainty}$$

Since uncertainty is related statistically to the width of a bell-shaped normal distribution, we can clarify the meaning of such a report by associating a *confidence* level with the uncertainty:

$$\textit{mean} \pm \textit{uncertainty} (\% \textit{confidence})$$

The confidence is defined to be the likelihood, stated as a percentage, that the “true” mean value actually falls within the limits  $\text{mean} \pm \text{uncertainty}$ . Graphically, it is proportional to the area under the bell shape between those limits.

For example, when the uncertainty is reported as one standard error (1.0 SE), the confidence level is 68%. This means we would expect the true mean to fall inside the interval  $\text{mean} \pm \text{SE}$  68% of the time. Note that 32% of the time the true mean will fall outside this range! For this reason, results are often reported with uncertainty = 2.0 SE for which the confidence level is 95%. The true mean will fall outside the range  $\text{mean} \pm 2.0 \text{ SE}$  only 5% of the time.

Other common uncertainty choices are 1.6 SE (90% confidence), 2.6 SE (99% confidence), and 1.0 SD (95% confidence at sample size 4, 99.9% confidence at sample size 10).

### Reporting fractional or percent uncertainties

Often, percent uncertainties ( $\% \textit{uncertainty} = 100 \textit{uncertainty} / \textit{mean}$ ) convey the precision of a result more meaningfully than the uncertainties themselves. For this reason, it is also common to report a result in the form:

$$\textit{mean} \pm \% \textit{uncertainty} (\% \textit{confidence})$$

### Comparing with prediction: discrepancies

Often you will make measurements in order to test a prediction resulting from a theoretical model. In science, such tests are never undertaken to prove the model correct, but instead to prove it incorrect with some degree of confidence.

The first step in a comparison between a measured result (mean  $\pm$  uncertainty) and a prediction is to determine the difference between them, called the **discrepancy**:

$$\text{discrepancy} = \text{mean} - \text{prediction}$$

As with uncertainty, it is common to express the discrepancy as a fractional discrepancy or % discrepancy:

$$\% \text{discrepancy} = 100 \text{ discrepancy} / \text{prediction}$$

### Significant discrepancies

The larger the discrepancy, the greater the disagreement with prediction and the more likely it is that the model is incorrect or that undiscovered systematic errors are present. In the language of error analysis, we define a **significant** discrepancy to be larger than the uncertainty ( $|\text{discrepancy}| \geq \text{uncertainty}$ ). A significant discrepancy implies that the experimental result is **inconsistent** with the model prediction. The prediction and result are then said to be **discrepant**.

Significance is quantified by specifying the confidence level of the uncertainty used to test it. For example, a discrepancy 2.0 SE is significant at the 95% confidence level (but not at the 99% level).

When you obtain a significant discrepancy you can conclude that either the model prediction or the experimental result (or both!) must be incorrect. An incorrect model prediction could indicate that the model itself is bad or incomplete, or that the model was used incorrectly to make the prediction. An incorrect experimental result could indicate that there were unidentified systematic errors or that mistakes were made while analyzing the dataset.

### Comparing two measured values: null hypothesis

There are many occasions in lab when you measure two different quantities that should theoretically be equal and you wish to compare the two measured values with each other. For example, you might measure both the net force  $F_{net}$  and the acceleration  $a$  of a mass  $m$ , and wish to determine whether  $F_{net} = ma$ . In cases like this we employ the **null hypothesis test**. The null hypothesis test is based on the idea that if two values are equal to each other, then the difference between them will be zero (“null”). Of course, since experimental values are uncertain, the actual difference will almost never be exactly zero. But if, on the average, the difference is sufficiently close to zero we can conclude that the experimental values are equal.

To quantify this we define  $\Delta$  to be the difference between the experimental values:

$$\Delta = x_1 - x_2$$

Then if we expect  $x_1$  and  $x_2$  to be equal the predicted value is  $\Delta_{\text{predicted}} = 0$ .

To test this prediction statistically, measure  $x_1$  and  $x_2$  and determine  $\Delta$ , then repeat to obtain multiple independent values of  $\Delta$ . You can then find the mean and uncertainty of  $\Delta$  and compare the experimental result  $\text{mean}_{\Delta} \pm \text{uncertainty}_{\Delta}$  with the predicted value (0) as described above.