An Introduction to Data Analysis

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

In science, the term “error” is used interchangeably 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.

Part 1: Uncertainties in Experimental Measurements

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 ± 5 lbs. Of course, if I were paying close attention to my weight, “150 lbs” might mean somewhere between 149 and 151 lbs (150 ± 1 lbs). Or if I say “a newborn moose weighs about 150 lbs”, I might mean somewhere between 100 and 200 lbs (150 ± 50 lbs).

We often use significant figures to imply uncertainty. For example, a result of 12.4 m is usually interpreted as being between somewhere between 12.3 and 12.5 meters. The general understanding is that the implied “plus or minus” amount is at least the size of the last decimal place. However, it’s a much better practice to actually state the uncertainty, such as 12.40 ± 0.05 m or 12.4 ± 0.2 m, 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 ± 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 ± 0.05 m”, since the uncertainty makes all of the digits after 9 meaningless. It would also be incorrect to report “12 ± 0.05 m”, because here the value is not stated with as much precision as the uncertainty allows. Instead, the result should be written as “12.40 ± 0.05 m”. Another incorrect statement would be “12.398318 ± 0.046252 m”, 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 ± 0.5 lbs 124 ± 10 Joules 62,000 ± 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)

![Ruler with scale](image)

So you would estimate the uncertainty to be 0.5 mm and report that the block’s length is 13.5 ± 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 ± 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. The simplest way is to use the significant figures on the display as a guideline. As discussed above, it is often implied that the uncertainty is half of the smallest decimal place, so a reading of 50.74 seconds would imply an uncertainty of 0.005 seconds, and you might report 50.740 ± 0.005 s.

However, most digital measuring devices have a specified precision (either stamped on the device or listed in a manual), and many of them display more decimal places than their precision implies. For example, the manual for the timer above might state that it is accurate to 0.02 seconds, in which case you should report a value of 50.74 ± 0.02 s.

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

However, it is important to realize that not all types of experimental uncertainties can be assessed by statistical analysis based on repeated measurements. For this reason, uncertainties or errors are classified into two groups: (a) random errors, which can be treated statistically; and (b) systematic errors, which cannot. Random errors can be revealed by repeating the measurements; systematic errors 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 are much smaller than the required precision (and are therefore “negligible”).

**True values and best estimates**

It is important to realize that the “true” period \( T \) of the rotating turntable discussed 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 (M): best estimate of the true value**

Suppose we make \( N \) measurements of a quantity \( x \) and get the values \( x_1, x_2, \ldots, x_N \). 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 standard deviation \( \sigma_x \) (or SD), defined as:

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

The term \( d_i \) in this equation, called the deviation, is simply the difference between the \( i \)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**

<table>
<thead>
<tr>
<th>Trial number, ( i )</th>
<th>Measured value, ( x_i )</th>
<th>Deviation, ( d_i = x_i - \bar{x} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>-0.8</td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>72</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>73</td>
<td>1.2</td>
</tr>
<tr>
<td>5</td>
<td>71</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

\[ \bar{x} = 71.8 \quad \bar{d} = 0.0 \]

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 (dividing by \( N-1 \) 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.

**Statistical interpretation of the standard deviation**

Again, let’s suppose that we make \( N \) measurements of a quantity \( x \) and get the values \( x_1, x_2, \ldots, x_N \). We then compute the mean, \( \bar{x} \), and standard deviation, \( \sigma_x \). If we then make one more measurement (using the same equipment), statistically there is a 68% probability that the new measurement will fall within one standard deviation of \( \bar{x} \) (and a 95% probability that it will fall within two standard deviations of \( \bar{x} \)). This means that for our new measurement \( x_{\text{new}} \), there is a 68% probability that:

\[
\bar{x} - \sigma_x < x_{\text{new}} < \bar{x} + \sigma_x
\]

Now, if the original number of measurements \( N \) was large, then \( \bar{x} \) should be a very reliable estimate for the actual value of \( x \). Therefore we can say that there is a 68% probability that a single measurement will be within standard deviation, \( \sigma_x \), of the actual value. Clearly \( \sigma_x \) means exactly what we have used the term “uncertainty” to mean in the preceding sections. If we make one more measurement of \( x \), then the uncertainty associated with this measurement can be taken to be \( \sigma_x \); and with this choice we are 68% confident that our measurement is within \( \sigma_x \) of the correct answer.
In the previous example of Table 1, our best estimate at one standard deviation is:

\[ x_{\text{best}} = \bar{x} \pm \sigma_x = 71.8 \pm 0.84 \approx 71.8 \pm 0.8 \]

\[ \Rightarrow 71.0 \leq x_{\text{best}} \leq 72.6 \]

If we make one more measurement, one has a 68% confidence level that it will be between 71.0 and 72.6.

From statistical analysis, it can be shown that:

- The best estimate \( x_{\text{best}} \) will fall within the range \( \bar{x} \pm \sigma_x \) approximately 68% of the time.
- The best estimate \( x_{\text{best}} \) will fall within the range \( \bar{x} \pm 2\sigma_x \) approximately 95% of the time.
- The best estimate \( x_{\text{best}} \) will fall within the range \( \bar{x} \pm 3\sigma_x \) approximately 99.7% of the time.

**Graphical interpretation of the standard deviation**

Suppose that in an experiment we made ten measurements of some length \( x \) and obtained the following values (all in cm):

\[ 26, 24, 26, 28, 23, 24, 25, 24, 26, 25 \]

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

<table>
<thead>
<tr>
<th>Measured value</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of measurements</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The frequency distribution of our measurements can be graphically displayed in a histogram as shown in Figure 1. Here \( x_k \) is the measured length in cm and \( F_k \) is the number of times that particular length was measured in our experiment.
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](image)

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 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!)

### Part 2: Reporting Experimental Results

**Confidence levels**

In Part 1, we discussed the importance of reporting an experimental result of a measured value in the form “mean ± uncertainty”. Now that we have seen how 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.

The confidence level is defined to be the likelihood, stated as a percentage, that the “true” value of a measurement actually falls within the limits: mean ± 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 deviation, the confidence level is 68%. This means we would expect the true value of the measurement to fall inside the interval \( M \pm SD \) 68% of the time.
**Reporting fractional or percent uncertainties**

Often, percentage uncertainties (where % uncertainty = fractional uncertainty x 100%) 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:

\[
\text{Result} = \text{M} \pm \%\text{SD} \text{ (68% confidence)}
\]

In the above equations, %SD is given by %SD = SD/M × 100%.

For example, if a set of measurements of the speed of a fired projectile yielded a mean of 10.0 m/s and a standard deviation 0.5 m/s, then %SD = (0.5 m/s)/(10.0 m/s) × 100% = 5.0%. You would then report the result as:

speed = 10.0 m/s ± 5.0% (68% confidence)

---

**Part 3: Comparing Experiment Results 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 (M ± SD) and a prediction (P) is to determine the difference between them, called the **discrepancy** D:

\[
D = M - P
\]

**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 say that larger discrepancies are more **significant** (meaningful). While there is no hard and fast rule, we will adopt the following “68% confidence” criterion for deciding whether the discrepancy in a given experiment is considered significant or not:

If the ratio \(|D| / SD < 1\), then D is **insignificant** and the experimental result is considered **consistent** with the model.

If \(|D| / SD \geq 1\), then D is **significant** (with > 68% confidence) and the experimental result is considered **inconsistent** with the model.
The above criterion is equivalent to saying that if the magnitude of the discrepancy is less than one standard deviation, then the experimental result is considered consistent with the model (with 68% confidence). If the magnitude of the discrepancy is greater than or equal to one standard deviation, then the experimental result is considered inconsistent with the model.

When you obtain a significant discrepancy, you can conclude (with > 68% confidence) that either the model prediction or the experimental result 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 data.