Reference text: "An introduction to Error Analysis, 2nd ed, J.R. Taylor, 1997

Significant figures

All measurements and experimentally derived quantities should be quoted with an uncertainty. Furthermore, both the value and the uncertainty should be quoted to an appropriate precision, that is the correct number of digits. This known as using the correct number of significant figures.
E.g, 9.82 has three significant figures, 0.002 has one significant figure, 1000 has 4 significant figures.

Rules for quoting uncertainties
1. Experimental uncertainties should always be rounded to one significant figure (exception below).

E.g., if the acceleration due to gravity is measured to be $g = 9.8324\, ms^{-2}$ with an uncertainty of $\Delta g = 0.0238\, ms^{-2}$ then the correctly quoted value is:

$$g = 9.83 \pm 0.02\, ms^{-2}$$

2. If the leading digit in the uncertainty is between a 1 and 2, then use two significant figures.

E.g., if the uncertainty in measuring g, $\Delta g = 0.01385$, then it is appropriate to write:

$$g = 9.832 \pm 0.014\, ms^{-2}.$$

The rationale is that rounding 1.49 to 1 significant figure (here 1) is an appreciable misrepresentation (67% of the estimated uncertainty).

Rules for stating answers
1. The last significant figure in any stated answer must be the same order of magnitude as the uncertainty. I.e., the answer must be of the same precision as the uncertainty.
E.g. $98.21 \pm 0.3\, s$ should be written as: $98.2 \pm 0.3\, s$

2. Always write the uncertainty and the measurement in the same units.
E.g. $3.032\, m \pm 1\, mm$ should be written as: $3.032 \pm 0.001\, m$

3. Write the measurement and uncertainty in scientific form using the same order of magnitude. This makes the values easier to interpret. Use brackets to make it clear that the exponent applies to both the answer and the uncertainty.

E.g. $1.61 \times 10^{-19} C \pm 5 \times 10^{-21} C$ should be written as: $(1.61 \pm 0.05) \times 10^{-19} C$

Examples:
Correct:

$$1.6 \pm 0.1\, ohms$$
$$1000 \pm 1\, ohms$$
$$0.023 \pm 0.001\, s$$

Incorrect:

$$2.356 \pm 0.01\, N \ \ (should\ be\ 2.36 \pm 0.01\, N\ )$$

$$3500 \pm 30\ kg \quad (should\ be\ (3.50 \pm\ 0.03) \times 10^3\ kg)$$
$$1.2 \pm 0.001\ g \quad (should\ be\ 1.200\ \pm 0.001\ g,)$$

Basic statistics

Random and systematic errors

Experimental uncertainties that can be revealed by repeating measurements are known as random errors (or uncertainties). Uncertainties that cannot be revealed in this way are known as systematic errors. An example of a random error would be timing the swinging of a pendulum manually using a stopwatch. Here there is some judgement required by the experimenter when the pendulum has swung through a single cycle. The variation in reaction time in pressing the stop button would also contribute a random error.

Systematic uncertainties are many ways more troublesome as they aren't obvious to the experimenter based on the measurements taken. An example would be measuring thickness using a micrometer in a hot room. Measuring instruments are generally certified at 20 °C, and so it is expected that your measurements will be systematically incorrect in this situation. In fact, if you are aware of a systematic uncertainty, then the best practice is to correct your measurements!
A case where you are likely to have a systematic error, but the magnitude is unknown is with digital meters. The specification document for the meter will specify the calibration uncertainty range, and best practice is usually take the systematic error as being the maximum departure specified.
However, in most experiments encountered in the undergraduate laboratory, random errors are usually more important. It is difficult to reduce random uncertainties to less than 0.1%, and the calibration of most modern test equipment meters is substantially better than this. For example, the Rigol 6 ½ digit multimeters used in the lab have a typical DC voltage measurement specification of 0.003% + 0.0025% of the range value.
Digital oscilloscopes, however, can be surprisingly inaccurate. Most only feature 8 bit digitisation and the displayed waveform is often not accurate in indicated voltage to better than +/- 4%!

Most uncertainty analysis is about dealing with random uncertainties. We'll see later how to construct an estimate when both random and systematic uncertainties are present.
Statistical theory is used to provide insight into the influence random uncertainties have on the estimating measured quantities.

The Mean

In cases where there is random variation in the measurement of a quantity the best estimate of the actual value is the mean value. In fact, this is only true if the measurements are subject to a distribution symmetric about a most probable value, most commonly Gaussian in shape. Data is distributed according to Gaussian statistics is said to be normally distributed, and we will see later justification for why this is observed, and exceptions to the rule.
Assuming we have normally distributed data, the mean is just:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i\ ,$$

where N is the number of measurements, and x_i represents an individual measurement.

The standard deviation

The standard deviation is a measurement of the average uncertainty in the measurements $x_1 \dots x_N$.
It is calculated by considering the differences between each measured value x_i and the mean $\bar{x}$ (known as a deviation). An issue arises if the sign of the difference is taken into account – positive differences would cancel out the negative differences and so the following formula is used instead:

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

Here σ_x is known as the standard deviation. Sometimes this is referred to as the root-mean-squared (or RMS) deviation of the measurements, or the population standard deviation.
In most calculations (default calculations in software and on calculators) the standard deviation is replaced by the sample standard deviation:

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

This replacement of N with $N-1$ is cannot be rigorously justified, although there are theory arguments for doing so. As we will see, it is common to replace N with the "degrees of freedom" for the data set. This is taken as the number of measurements (N) minus the number of global values that are used in calculating the parameter. Here, in the calculation of the standard deviation, we have used the mean value $\bar{x}$, which is calculated using all the individual measurements (x_i). Consequently, the degrees of freedom for calculating σ_x is one less than the number of measurements ($N-1$).

The standard deviation is the uncertainty in a single measurement

Recall that the standard deviation is the average uncertainty calculated from all the measurements, $x_1 \dots x_N$. Hence it follows that the uncertainty expected for any individual measurement is given by the standard deviation σ_x. Essentially, any individual measurement is expected to be within σ_x of the mean (which approaches the true) value that has been found, $\bar{x}$.
As we will see, when we discuss the distribution of random uncertainties, the rigorous definition of the individual measurement uncertainty is that the probability of an individual measurement being within $x \pm \sigma_x$ is 68.3%.

The standard deviation of the mean (the standard error) is the uncertainty in the mean

As we have seen, based on many measurements, $x_1 \dots x_N$, we can calculate the mean value, $\bar{x}$, which is the best estimate of the true value for the quantity x.
As we have used many measurements of x, it is reasonable to expect that the uncertainty in the value of $\bar{x}$ is smaller the uncertainty in any individual measurement, σ_x.
The uncertainty in the mean value is known as the standard deviation of the mean, or the standard error. We will see that it is given by:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}.$$

The standard error represents the uncertainty in the mean value. As can be seen from the equation above, when you have random uncertainties, it is possible to reduce the uncertainty in your best estimate of the value to an arbitrarily small value by taking an appropriate number of measurements. However due to the square-root dependency of the improvement with the number of measurements, it quickly becomes impractical to improve the uncertainty by taking more measurements.

Histograms and Distributions

When measurements are made many times and there is a variation in the values, due to random influences, the values obtained are scattered. A histogram can be used to organise the results for a statistical analysis.
Here, values are sorted into bins, depending on their value. The number of counts in each bin is then used to create distribution plot. The range of values that each bin accepts is chosen by the experimenter. The bin range can affect the form of the distribution plot and so must be chosen with some care (more on this later).

An example:
We have the following measurements of the focal length of a lens (in cm):

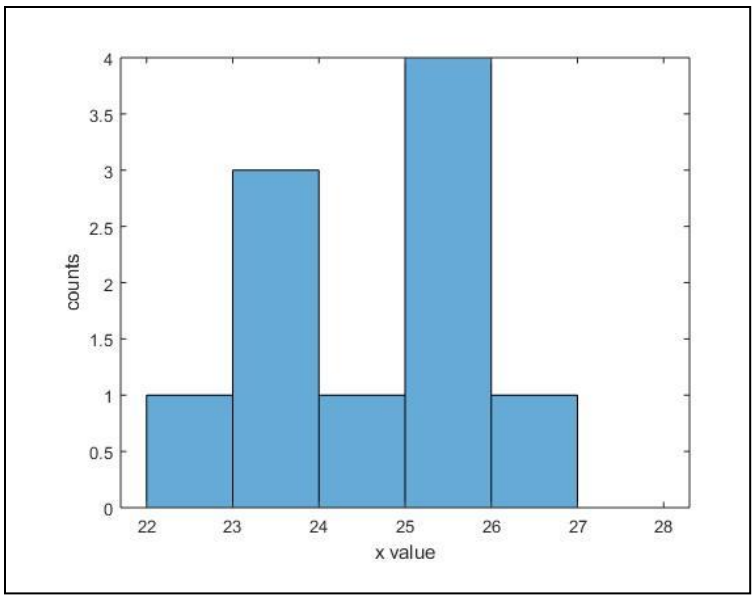
$$x = 26.4, 23.9, 25.1, 24.6, 22.7, 23.8, 25.1, 23.9, 25.3, 25.4\ .$$

A histogram of these values can be created. Here we will choose the bins to be one unit apart:

Bin	counts
22-23	1
23-24	3
24-25	1
25-26	4
26-27	1
27-28	0

Various software can be used to do this automatically and produce a histogram plot. However, it is important to check how the binning is done. For some apps you specify the bin centres, some the edges. Others automatically pick the bin locations and width based on the number of bins requested for the histogram.

Using the binning above the histogram appears as:

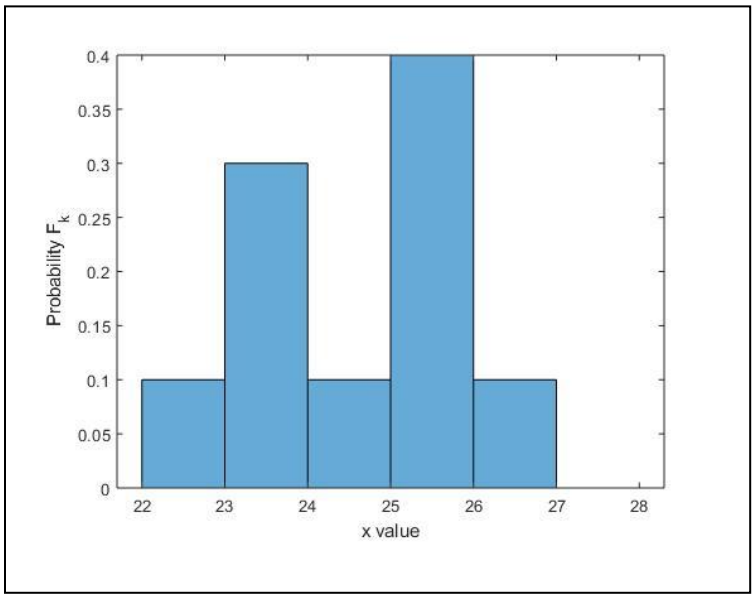


Histogram of the values of focal length. The total number of values is 10. ($N = 10$).

We can convert the histogram to show probabilities that an individual measurement will be found within a given bin. For a given bin of corresponding to a range of x values, denoted x_k, we divide the number of counts in the bin, n_k, by the total counts, N, to give the probability that a single measurement will fall within the bin x_k, i.e:

$$F_k = \frac{n_k}{N}.$$

We can now plot a probability histogram for the experiment.



Histogram of the values of focal length. The vertical axis shows the fraction, F_k, that each value x_k is observed.

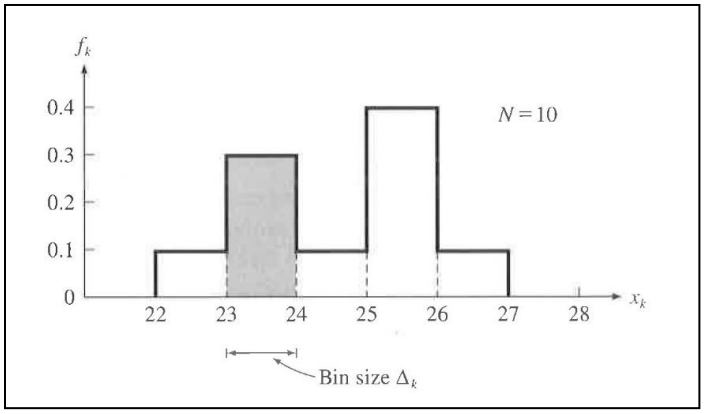
By doing this process, if you add up the probabilities of each bin occurring, you end up with a total probability of 1, reflecting that every measurement contributed to the histogram.

We can also plot our histogram as a probability density histogram, in which the area of each histogram bar corresponds to the fraction of measurements that fall within each bin. As we will see, this is useful for approximation by a limiting distribution.
Here we have:

$$Area\ of\ bar\ k = f_k \Delta_k,$$

So:

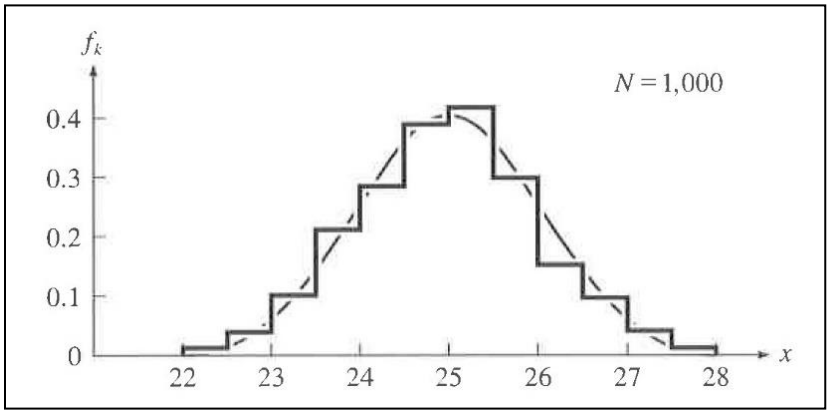
$$F_k = f_k \Delta_k.$$



Probability density histogram of the values of focal length. The area of each bar is $f_k \Delta_k$ and is the fraction of measurements contained in the kth bin.

Importantly, the area of the histogram is normalised to 1.

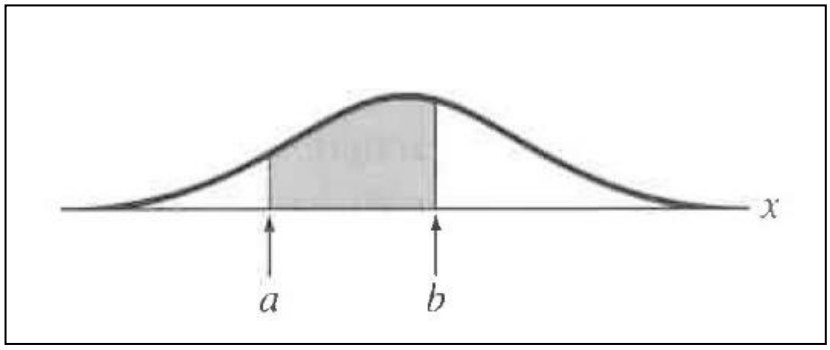
In most experiments as the number of measurements increases the histogram begins to take on a simple shape (usually a Gaussian, as we will see later). Thus, we can idealise the profile of the histogram with a continuous curve known as the limiting distribution.



Probability histogram of a large number of values (1000) for the focal length experiment. The dashed line is the limiting distribution (Gaussian).

The limiting distribution is defined by a continuous function $f(x)$ which is useful for analysis purposes. As the area is normalised, we can examine the probability of measurements that will occur between arbitrary limits, say $x = a$ and $x = b$. This amounts to calculating the area under the distribution curve, $f(x)$ between the values $x = a$ and $x = b$:

$$Probability\ of\ measurements\ between\ x = a \text{ and } x = b = \int_a^b f(x)dx\ .$$



The total probability that a measurement falls between $x = a$ and $x = b$ is represented by the shaded area under the curve.

What is the optimum number of bins to use for a histogram?

The short answer is that there isn't one. Data with differing distributions will benefit from different binning strategies. Also changing the binning can reveal different aspect of the data (such as showing the data occurs in discrete intervals).
A rule often stated (but without any rigorous mathematical defence) is the "square root rule", where the number of bins is:

$$K = \lceil \sqrt{N} \rceil\ .$$

Here ⌈ ⌉ is the ceiling function, (i.e. round up to the next integer), and N is the total number of counts. Excel uses this rule to choose the number of bins automatically for histogram plots.

However, in the case of data that corresponds to a normal distribution, there are a few well-known optimisations. One useful one is Scott's normal reference rule, which minimizes the integrated mean squared error for data that is normally distributed. The number of bins is given by:

$$K = \left\lceil \frac{x_{max} - x_{min}}{h} \right\rceil,$$

where:

$$h = \frac{3.5\sigma}{N^{1/3}}\ .$$

Here σ is the standard deviation of the data.
For experimental data (which is generally normally distributed) it is recommended to use this rule to determine the number of bins for histogramming data that will be subsequently fit using a Gaussian distribution function.

Distribution functions

When quantities are measured in an experiment, the values obtained are subject to some random uncertainty. This is evident when multiple measurements are made at a sufficiently discriminating resolution. Often the exact source(s) of the variation is unknown and is treated as "noise".
Typically measurements are distributed with some sort of characteristic distribution, usually corresponding to a gaussian centred about the true value of the measured quantity.

An interesting question is why do we see a Gaussian spectrum of measured values?

The Binomial Distribution (p227)

The binomial distribution is a fundamental result from probability theory. In general, this distribution function isn't encountered often in experimental physics, however, it can be used to explain why measurements have a Gaussian distribution of values. The name derives from the similarity of the expression for the distribution to the binomial expansion formula.

For an experiment which has multiple discrete outcomes, the binomial distribution describes the probability of finding a certain number of outcomes when the experiment is conducted many times.
For example, consider tossing a coin 4 times, the probability of obtaining heads 3 times is described by a binomial distribution.
The key parameters are the number of trials, n, the probability of your desired result in any single trial, p, and the number of successes required, ν.

$$Prob(\nu\ successes\ in\ n\ trials) = B_{n,p}(\nu)$$

$$= \frac{n!}{\nu!\,(n-\nu)!} p^{\nu}(1-p)^{n-\nu}$$

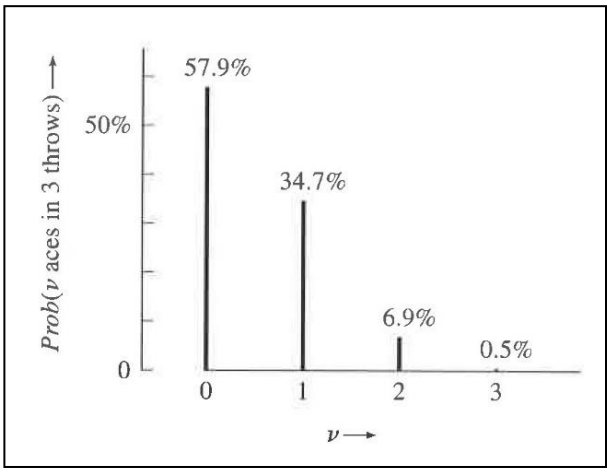
$$= \binom{n}{\nu} p^{\nu}(1-p)^{n-\nu}$$

(Note 0! =1)

For example, consider throwing a six-sided die 3 times and asking what are the probabilities of rolling a six 0, 1, 2, or 3 times?

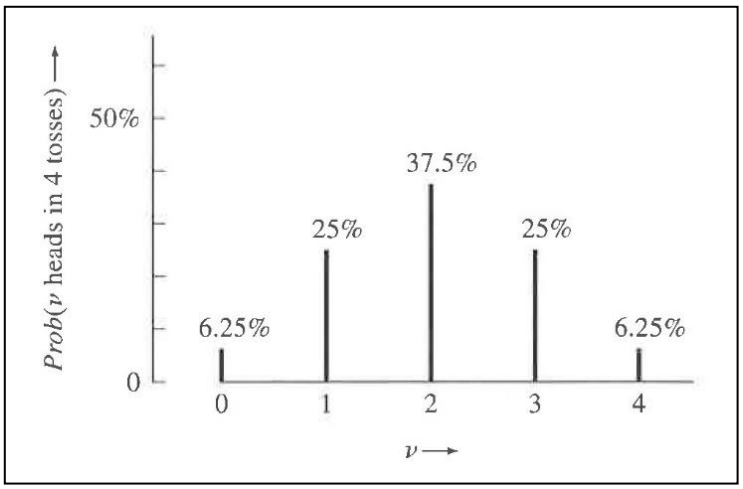
Here $n = 3$ and $p = 1/6$.

$Prob(0\ sixes\ in\ 3\ throws)$ =0.579
$Prob(1\ in\ 3)$ =0.347
$Prob(2\ in\ 3)$ =0.0694
$Prob(3\ in\ 3)$ =0.0046



Probabilities of rolling ν sixes in 3 rolls of the die.

A feature of the binomial distribution is that, in general, the distribution isn't symmetric in shape. However, in the case of $p = 1/2$ (e.g. the chance of finding heads when tossing a coin once), the distribution is symmetric about the most probable outcome.



Probabilities of finding n heads when throwing 4 coins. ($n = 4, p = 1/2$)

The average value of the number of successes in the distribution is given by:

$$\bar{\nu} = np\,,$$

and the standard deviation is given by:

$$\sigma = \sqrt{np(1-p)}\,.$$