## Summary of Measurement Methods

When conducting an experiment, we make our measurements carefully, striving for reproducible results in which sources of errors are minimized. We often make repeated measurements of a quantity, or study more than one object from a sample, in order to determine representative properties for a large set of objects. Our measurement tools operate to a certain precision (the number of significant digits reported), and our data are always recorded with units.

We divide the source of the distribution of values (the spread) around a mean value into three categories: natural variation of the sample (the intrinsic range in value of the measured property), systematic errors (those which bias a sample in one direction), and measurement errors (the finite precision with which we make our measurements). Systematic errors can often be eliminated by careful experimental design, and measurement errors can be minimized by choosing proper equipment, while the natural variation of a sample is often the target of an experiment (along with the mean value).

1. When making repeated measurements of a single quantity, we often plot our results in a histogram, showing the number of measurements for various values surrounding the mean value. For a set of $N$ measurements $m_i$, the mean value $\mu$ and the spread, or standard deviation $\sigma$ are

$$\mu = \frac{1}{N}\sum_{i=1}^{N} m_i, \quad \text{and } \sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(m_i - \mu)^2}.$$

A good rule of thumb to remember is that two-thirds of all measurements should lie within $1\sigma$ of the mean value, and 95% should lie within $2\sigma$. If two experiments return mean values which differ by more than $2\sigma$, it is worth investigating any differences in technique or in the samples. A difference of $5\sigma$ is truly significant.

2. We often fit a line to a set of two related quantities ($x$, $y$), defined by a slope $m$ and a $y$-intercept $b$ such that on average $y = mx + b$, where for two points $(x_1, y_1)$ and $(x_2, y_2)$ lying along the line,

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}, \quad \text{and } b = y_1 - mx_1 = y_1 - \left(\frac{y_2 - y_1}{x_2 - x_1}\right)x_1.$$

We can calculate the quality of a fit of a line to a set of data by calculating the offsets $o_i$ of each point from the line, and averaging them to form an rms deviation, where

$$o_i = y_i - (mx_i + b), \quad \text{and rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} o_i^2}.$$

3. In a similar vein, a correlation coefficient $R$ tells us to what extent two variables will change together (+1 if they grow together, and -1 if one grows while the other shrinks, or near to 0 if they are not connected).