

# Statistics Notes

## 1 Basics

Mean:

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

Sample Standard Deviation:

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

Population Standard Deviation:

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

Sample Standard Error:

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

Population Standard Error:

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

Variance and Standard Deviation:

$$\text{Variance} = \sigma^2$$

Usually  $s$  indicates the sample metric and  $\sigma$  the population metric.

Precision can be measured using the standard error because it indicates how close a set of measured means are to each other.

The accuracy of a measurement is how close it is to the true mean.

## 2 Poor Man's Error Propagation

Given the product equation:

$$z \pm \delta z = (x \pm \delta x)(y \pm \delta y)$$

where  $x$  and  $y$  are subject to uncertainties,  $\delta x$  and  $\delta y$ , what is the uncertainty,  $\delta z$  in  $z$ ?

We can reexpress an uncertainty:

$$x \pm \delta x$$

in the equivalent form:

$$x \left( 1 \pm \frac{\delta x}{x} \right)$$

Using this representation we can now express our product equation using:

$$z \pm \delta z = xy \left( 1 \pm \frac{\delta x}{x} \right) \left( 1 \pm \frac{\delta y}{y} \right)$$

Multiplying out the terms in brackets and assuming that  $\delta x \delta y$  is negligible (we're assuming that the errors are very small), then

$$z \pm \delta z = xy \left( 1 \pm \left( \frac{\delta x}{x} + \frac{\delta y}{y} \right) \right)$$

That is

$$z \pm \delta z = \left( 1 \pm \frac{\delta p}{p} \right)$$

where

$$p = \frac{\delta x}{x} + \frac{\delta y}{y}$$

What this tells us is that the fractional error in  $z$  is given by the sum of the fractional errors in  $x$  and  $y$ . For example, if the error in  $x$  is 2% and the error in  $y$  is 3%, then the error in  $z$  is 5%.

### 3 Rule of Quadrature

The poor man's approach to estimating errors overestimates the errors, instead an alternative approach known as the rule of quadrature is used which gives more reasonable estimate for how errors propagate through equations. Consider a quantity,  $Q$  which is to be calculated from several observed variables,  $a, b, c, \dots$ :

$$Q = f(a, b, c, \dots)$$

Suppose that  $a, b, c, \dots$  are measured  $N$  times. We can then calculate  $N$  different values for  $Q$ . We can also calculate the mean and variance for the set of measurements  $a, b, c, \dots$ :

$$\sigma_a^2 = \frac{1}{N} \sum_{i=1}^N (\Delta a_i)^2$$

where  $\Delta a_i = a_i - \bar{a}$  and also the variance of  $Q$ ,

$$\sigma_Q^2 = \frac{1}{N} \sum_{i=1}^N (\Delta Q_i)^2 \tag{1}$$

where  $\bar{Q} = f(\bar{a}, \bar{b}, \dots)$ ,  $Q_i = f(a_i, b_i, \dots)$ , and  $\Delta Q_i = Q_i - \bar{Q}$ . The  $\Delta Q_i$  can be approximated by the following total derivative:

$$\Delta Q_i \cong \frac{\partial Q}{\partial a} \Delta a_i + \frac{\partial Q}{\partial b} \Delta b_i + \dots$$

Inserting the above equation into equation (1) yields:

$$\sigma_Q^2 = \frac{1}{N} \sum_{i=1}^N \left( \frac{\partial Q}{\partial a} \Delta a_i + \frac{\partial Q}{\partial b} \Delta b_i + \dots \right)^2$$

If the square term is expanded we get two types of term, one of which is the cross term:

$$\frac{\partial Q}{\partial a} \frac{\partial Q}{\partial b} \Delta a_i \Delta b_i$$

The cross terms, since they contain quantities that are equally likely to be positive or negative, add up to near zero. Therefore we can drop the cross terms to give:

$$\sigma_Q^2 = \frac{1}{N} \sum_{i=1}^N \left( \left( \frac{\partial Q}{\partial a} \right)^2 (\Delta a_i)^2 + \left( \frac{\partial Q}{\partial b} \right)^2 (\Delta b_i)^2 + \dots \right)$$

This can be rewritten as

$$\sigma_Q^2 = \left( \frac{\partial Q}{\partial a} \right)^2 \frac{1}{N} \sum_{i=1}^N \Delta a_i^2 + \left( \frac{\partial Q}{\partial b} \right)^2 \frac{1}{N} \sum_{i=1}^N \Delta b_i^2 + \dots$$

$$\sigma_Q^2 = \left( \frac{\partial Q}{\partial a} \right)^2 \sigma_a^2 + \left( \frac{\partial Q}{\partial b} \right)^2 \sigma_b^2 + \dots$$

Normally we are more interested in the standard error than the standard deviation. Given that the standard deviation is related to the standard error by:

$$\sigma_m^2 = \frac{\sigma}{\sqrt{N}}$$

so that  $\sigma = \sigma_m^2 \sqrt{N}$ , then we can replace the standard deviations on both sides with the standard errors (note that the  $\sqrt{N}$  cancel on both sides) to yield:

$$\sigma_Q^2 = \left( \frac{\partial Q}{\partial a} \right)^2 \sigma_a^2 + \left( \frac{\partial Q}{\partial b} \right)^2 \sigma_b^2 + \dots$$

or

$$\sigma_{\bar{Q}} = \sqrt{\left(\frac{\partial Q}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial Q}{\partial b}\right)^2 \sigma_b^2 + \dots} \quad (2)$$