The Weighted Mean

When there is a range of opinions or views on a particular matter, we tend to adopt the view that we consider to carry the greatest ‘weight’. The weight normally reflects the degree of authority, based on expertise and experience, that supports the view. For example, when the need for nuclear power stations is debated, we are likely to regard as more important the views of experts such as physicists, engineers, economists, environmentalists and health professionals than the views of non-experts. It is interesting that in German (for example) the word for ‘important’, *wichtig*, is obviously related to the English ‘weight’.

The concept of ‘weight’ in the domain of measurement arises from the common situation that ‘not all measurements are equal’ and that those that have less uncertainty should carry greater ‘weight’. For various reasons such as the experience, competence or diligence of the experimenter, or the quality of the measuring instrument or system, the result of a measurement (such as the resistance of a coil of wire, or the emission wavelength of a particular laser) will differ both in its value and the uncertainty in that value from one experimenter to another. In situations where we need to combine values of the same particular quantity as obtained by different experimenters in order to determine a ‘mean’ — that is, a ‘consensus’ value — how do we account for the different uncertainties in the values that they report? We address this question as follows.

If we take a set of measurements, and all the measured values are *equally* accurate, we end up with the ‘ordinary mean’ or ‘average’ of them. If all the values are *not* equally accurate, we would like our so-called ‘mean’ to be ‘pulled’ towards those values that *are* accepted as the more accurate ones. Then we end up with a ‘weighted mean’. We see that ‘accurate’ here is roughly the equivalent of ‘important’ in the top paragraph. We now proceed to discuss how the weighted mean is estimated using least-squares.

In section 5.2.1 (‘The mean as a least-squares fit’), a single value m was fitted by least-squares to the six values 4.1, 4.3, 4.4, 4.2, 4.3, 3.9. Such a single value is often required in order to summarise or represent individual values. The resulting value of m was the *mean*, $m = 4.20$, of these six values. This is of course a very common statistical operation: to summarises a set of values by a single value, we generally take their *mean* or, in everyday language, their *average*.

We recall how m was estimated using least-squares: the six measurements were written in the form:

$$\begin{aligned}
4.1 &= m + \epsilon_1 \\
4.3 &= m + \epsilon_2 \\
4.4 &= m + \epsilon_3 \\
4.2 &= m + \epsilon_4 \\
4.3 &= m + \epsilon_5 \\
3.9 &= m + \epsilon_6
\end{aligned} \qquad (1)$$

where the $\epsilon_1, \epsilon_2, ...\epsilon_6$ are the 'residuals'. The least-squares procedure, after estimating m, yielded the following values for the residuals:

$$\begin{aligned}
\epsilon_1 &= -0.10 \\
\epsilon_2 &= +0.10 \\
\epsilon_3 &= +0.20 \\
\epsilon_4 &= \quad 0.00 \\
\epsilon_5 &= +0.10 \\
\epsilon_6 &= -0.30.
\end{aligned} \tag{2}$$

These six residuals sum to zero (so they have only five degrees of freedom).

Up till now we considered the six original values as carrying 'equal weight'. In practice, this implies that all were measured with equal accuracy. But it sometimes happens that we have *advance knowledge* that all values were *not* obtained with equal accuracy. For example, the first value 4.1 might have been measured with a top-of-the-range instrument, and the other five with an instrument of lower quality. We would then naturally claim that 4.1 has 'greater weight' than the other five. We would like to be able to incorporate this advance knowledge when summarising the six values by a single value.

(Of course, it sometimes happens that the 'advance knowledge' only comes to light *after* the measurements are made. This does not affect the following analysis).

To reflect the different weights of the original six values, we 'scale' the residuals in (1) accordingly. The natural scaling factors to use are the *standard uncertainties* of the respective measurements. Call these $u_1, u_2, ...u_6$. Then we have, in place of (1),

$$\begin{aligned}
4.1 &= m_w + u_1\epsilon_1 \\
4.2 &= m_w + u_2\epsilon_2 \\
4.4 &= m_w + u_3\epsilon_3 \\
4.2 &= m_w + u_4\epsilon_4 \\
4.3 &= m_w + u_5\epsilon_5 \\
3.9 &= m_w + u_6\epsilon_6.
\end{aligned} \tag{3}$$

We denote in (3) the single estimate as m_w, in anticipation of the fact that this estimate will be the 'weighted mean'.

If, for example, the first value 4.1 is measured with high accuracy, then u_1 will be much smaller than any of $u_2, u_3, ...u_6$. (We recall that each u is a (standard) uncertainty, so the reciprocal $1/u$ is a measure of the corresponding accuracy). The product $u_1\epsilon_1$ will then be small, and effectively the resulting estimate of m_w will be such that m_w is forced to be 'closer' to the high-accuracy value 4.1 than it would be if all six measurements had equal weight.

As in the unweighted case, we now evaluate the quantity Q as the sum of squares of the residuals $\epsilon_1, \epsilon_2, ...\epsilon_6$:

$$Q = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2 + \epsilon_5^2 + \epsilon_6^2, \tag{4}$$

and from (3) this is equivalent to

$$Q = \frac{(4.1 - m_w)^2}{u_1^2} + \frac{(4.2 - m_w)^2}{u_2^2} + \frac{(4.4 - m_w)^2}{u_3^2} + \frac{(4.2 - m_w)^2}{u_4^2} + \frac{(4.3 - m_w)^2}{u_5^2} + \frac{(3.9 - m_w)^2}{u_6^2}. \tag{5}$$

We wish to find that value of m_w for which Q is a minimum (and we observe that $u_1, u_2, ...u_6$ are treated as known-in-advance constants). Differentiating Q with respect to m_w gives:

$$\frac{\partial Q}{\partial m_w} = -2\frac{(4.1 - m_w)}{u_1^2} - 2\frac{(4.2 - m_w)}{u_2^2} - 2\frac{(4.4 - m_w)}{u_3^2} - 2\frac{(4.2 - m_w)}{u_4^2} - 2\frac{(4.3 - m_w)}{u_5^2} - 2\frac{(3.9 - m_w)}{u_6^2} \tag{6}$$

and (6) is equal to zero for an extremum value of Q (maximum or minimum). (It can be easily shown, taking the second derivative $\partial^2 Q/\partial m_w^2$, that this extremum is in fact a minimum value of Q). Cancelling out the -2 in (6), the solution for m_w is:

$$m_w = \frac{\frac{4.1}{u_1^2} + \frac{4.2}{u_2^2} + \frac{4.4}{u_3^2} + \frac{4.2}{u_4^2} + \frac{4.3}{u_5^2} + \frac{3.9}{u_6^2}}{\frac{1}{u_1^2} + \frac{1}{u_2^2} + \frac{1}{u_3^2} + \frac{1}{u_4^2} + \frac{1}{u_5^2} + \frac{1}{u_6^2}}. \tag{7}$$

The formula (7) gives the weighted mean of the six values.

Thus, suppose as before that u_1 is very small, and the other five u's are much larger. Then in (7) the dominating term in the numerator will be $4.1/u_1^2$, and in the denominator, $1/u_1^2$. So in this case,

$$m_w \sim \frac{\frac{4.1}{u_1^2}}{\frac{1}{u_1^2}} = 4.1. \tag{8}$$

As we expect, since the high-accuracy first value, 4.1, has much greater weight than the other five, we have $m_w \sim 4.1$ rather than $m = 4.20$ as in the unweighted case.

In (7), the coefficients $1/u_1^2$, $1/u_2^2$,...u_6^2 are called the 'weights' $w_1, w_2, ...w_6$ respectively. In the general case of n originally measured values $y_1, y_2, ...y_n$, we have, in place of (3),

$$y_i = m_w + u_i\epsilon_i \quad (i = 1, 2, ...n), \tag{9}$$

and the weighted mean is then given by

$$m_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \tag{10}$$

where $w_i = 1/u_i^2$ for $i = 1, 2, ..., n$. We observe that the weights are inversely proportional to the *variances* (the *squares* of the standard uncertainties). This is so even though, in (3), the scaling factors for the residuals were the standard uncertainties themselves.

In (10), if the weights w_i $(i = 1, 2, ..., n)$ all have the same value, say w_0, then (10) reduces to

$$m_w = \frac{w_0 \sum_{i=1}^{n} y_i}{n w_0} = \frac{\sum_{i=1}^{n} y_i}{n}, \tag{11}$$

which is, as we would expect, the ordinary mean m.

The effect of weighting can be clearly seen if we imagine just two values, say $y_1 = +1$ and $y_2 = +2$. Then $m = 1.5$. Now suppose that $y_1 = +1$ was measured with twice the uncertainty of the measurement of $y_2 = +2$, so that we can take, for example, $w_1 = 1$ and $w_2 = 4$. We note that (as is true for the general case of n measurements), it is only the *ratio* of the weights that matters; the same result for m_w would be obtained if we took (say) $w_1 = 2$ and $w_2 = 8$. (This is why the equally-weighted case can also be called simply the 'unweighted' case). Equ. (10) gives

$$m_w = \frac{(1 \times 1) + (4 \times 2)}{1 + 4} = +1.8. \tag{12}$$

The higher accuracy of $y_2 = 2$ 'pulls' our single-value estimate away from the unweighted mean 1.5 towards that higher-accuracy value.

We next consider the standard uncertainty of m_w. We may regard (10) as stating a relationship between 'inputs' y_i and the output m_w, with $w_i = 1/u_i^2$ as constant factors. Provided that the y_i $(i = 1, 2, ..., n)$ are all uncorrelated with one another, we can then use the relationship (7.14) in Chapter 7, which in the present notation becomes:

$$u^2(m_w) = \left(\frac{\partial m_w}{\partial y_1}\right)^2 u^2(y_1) + \left(\frac{\partial m_w}{\partial y_2}\right)^2 u^2(y_2) + ... + \left(\frac{\partial m_w}{\partial y_n}\right)^2 u^2(y_n). \tag{13}$$

Since $u_i^2 = 1/w_i$ $(i = 1, 2, ..., n)$ in (13), we have

$$u^2(m_w) = \frac{1}{w_1}\left(\frac{\partial m_w}{\partial y_1}\right)^2 + \frac{1}{w_2}\left(\frac{\partial m_w}{\partial y_2}\right)^2 + ... + \frac{1}{w_n}\left(\frac{\partial m_w}{\partial y_n}\right)^2. \tag{14}$$

Now from (10), for all $k = 1, 2, ..., n$,

$$\frac{\partial m_w}{\partial y_k} = \frac{w_k}{\sum_{i=1}^{n} w_i}, \tag{15}$$

so

$$u^2(m_w) = \frac{1}{w_1}\frac{w_1^2}{(\sum_{i=1}^{n} w_i)^2} + \frac{1}{w_2}\frac{w_2^2}{(\sum_{i=1}^{n} w_i)^2} + ... + \frac{1}{w_n}\frac{w_n^2}{(\sum_{i=1}^{n} w_i)^2}$$

$$= \frac{\sum_{i=1}^n w_i}{(\sum_{i=1}^n w_i)^2} = \frac{1}{\sum_{i=1}^n w_i} = \frac{1}{\sum_{i=1}^n (1/u_i^2)}. \tag{16}$$

In the particular case where all the weights are equal, so that the u_k ($k = 1, 2, ..., n$) are equal at the value, say, u_0, (16) gives:

$$u^2(m_w) = \frac{1}{\sum_{i=1}^n (1/u_0)^2} = \frac{1}{n/u_0^2} = \frac{u_0^2}{n}, \tag{17}$$

or

$$u(m_W) = u_0/\sqrt{n}. \tag{18}$$

Equ. (18) may be recognised as the standard uncertainty of the mean of equally-weighted and uncorrelated values (see, for example, equation (7.31) in Chapter 7). In this equal-weight case, therefore, the scaling factor u_0 is estimated simply as the standard deviation of the original values y_i ($i = 1, 2, ..., n$). (This standard deviation should be obtained as the square root of the *unbiased* variance of the original values; see for example equ. 5.13 in the book).

In the equally-weighted case, the residuals sum to zero:

$$\sum_{i=1^n} \epsilon_i = \sum_{i=1}^n (y_i - m) = 0, \tag{19}$$

as can be seen in the particular example in (2). The counterpart of (19) in the weighted case is:

$$\sum_{i=1}^n w_i(y_i - m_w) = 0. \tag{20}$$

Equ. (20) can also be written $\sum_{i=1}^n \epsilon_i/u_i = 0$, using $y_i - m_w = u_i\epsilon_i$ and $w_i = 1/u_i^2$ for $i = 1, 2, ..., n$. Then we see that, as in the equally-weighted case, the residuals have $n-1$ degrees of freedom.

The proof of (20) is as follows. Using (10), the left side of (19) is

$$\sum_{i=1}^n w_i \left(y_i - \frac{\sum_{k=1}^n w_k y_k}{\sum_{k=1}^n w_k} \right) = \sum_{i=1}^n w_i y_i - \frac{(\sum_{i=1}^n w_i)}{(\sum_{i=1}^n w_k)} \sum_{k=1}^n w_k y_k$$

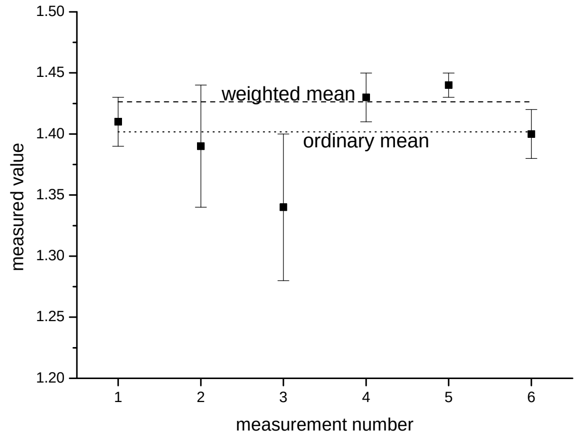
$$= \sum_{i=1}^n w_i y_i - \sum_{i=1}^n w_k y_k = 0. \tag{21}$$

We may easily check (19) in the case of the two values ($n = 2$) in the previous example, with $y_1 = 1$ and $y_2 = 2$, with $w_1 = 1$ and $w_2 = 4$. We obtained $m_w = 1.8$. Then (19) translates into:

$$1 \times (1 - 1.8) + 4 \times (2 - 1.8) = -0.8 + 0.8 = 0.0,$$

verifying (20).

As a graphed example, here are six measured values and their corresponding standard uncertainties:



Ordinary mean compared with weighted mean

Measurement number	measured value	standard uncertainty
1	1.41	0.02
2	1.39	0.05
3	1.34	0.06
4	1.43	0.02
5	1.44	0.01
6	1.40	0.02

The ordinary mean is 1.40167, whereas the weighted mean is 1.42637. The graph shows the six points each with an attached error bar. The length of the error bar measured from its centre to either end is equal to the standard uncertainty. (Sometimes error bars are drawn with the centre-to-end length equal to the *expanded* uncertainty, which is often about twice the standard uncertainty; see Ch. 10 in the book). We observe that the weighted mean is considerably greater than the ordinary mean, because the larger measured values tend to be the more accurate ones.

If we ignore the weights in this example, and assume uncorrelated values, the standard uncertainty of the unweighted mean is about 0.0354. If we take into account the different weights and use (16), the standard uncertainty of the weighted mean is considerably less at 0.0074.