Consensus Values and Weighting Factors

Robert C. Paule* and John Mandel*

National Bureau of Standards, Washington, DC 20234

A method is presented for the statistical analysis of sets of data which are assembled from multiple experiments. The analysis recognizes the existence of both within group and between group variabilities, and calculates appropriate weighting factors based on the observed variability for each group. The weighting factors are used to calculate a "best" consensus value from the overall experiment. The technique for obtaining the consensus value is applicable to either the determination of the weighted average value, or to the parameters associated with a weighted least squares regression problem. The calculations are made by using an iterative technique with a truncated Taylor series expansion. The calculations are straightforward, and are easily programmed on a desktop computer.

An examination of the observed variabilities, both within groups and between groups, leads to considerable insight into the overall experiment and greatly aids in the design of future experiments.

Key words: ANOVA (within-between), components of variance, consensus values, design of experiments, pooling of variance, weighted average, weighted least squares regression.

1. Introduction

e purpose of this article is to discuss the problem of calculating "best" estimates from a series of experimental results. It will be convenient to refer to these estimates as consensus values. Since experimental data frequently come from many different sources, with each having its own characteristic variability, the statistician's problem centers on the appropriate weighting of the data to obtain the consensus value(s). In order to achieve this aim, the statistical analysis should recognize the existence of both within group and between group variability. Both types of variability are considered here to be random effects and are described by their associated components of variance: the within set component of variance for group i, $\sigma_{w_i}^2$, and the between set component of variance, σ_b^2.

Early attempts to solve the consensus value problem have not explicitly recognized the existence of the between set component of variance. Attempts, such as the Birge ratio method [1],[1] are adversely affected by this omission.

In this article we will deal with two types of problems. The first is essentially the calculation of a weighted average value and its statistical uncertainty. The second problem arises in the fitting of a straight line or curve to a set of data. The estimates both for the average and the

parameters of the fitted curve can be thought of as consensus values, i.e., they should be the consensus of the observed data. Both problems can involve several sources of error. It will turn out that, because of the similarity of the weighting problems in both types of situations, a theoretical solution common to both problems can be derived. A solution for a more restricted form of the first problem has been previously reported [2,3]. The mathematical aspects of the general problem are outlined in the current paper.

An understanding of the nature of random error processes associated with measurement systems is required to develop appropriate weighting factors. The weighting factors that are derived are not arbitrary, but are controlled by the nature of the variability of the data.

2. Illustrative Examples

For purposes of exposition, artificial examples of data sets will be used as illustrative material. Later in this paper it will be shown that the procedures developed are useful for the analysis of actual laboratory data.

To develop some feeling for "what is appropriate weighting", let us examine two specific examples. For the first example, consider that three measurements are each made by method A and by method B, and that the method A results are more precise. Assume, both here and throughout this manuscript, that the relative accuracies of the methods are not known. Let the following be the measured values, and the corresponding coded

---

*National Measurement Laboratory

[1]Figures in brackets indicate literature references located at the end of this paper.

values which are obtained by subtracting 200 from the measured values.

Method	A			B		
Measured Values	201.1	201.9	201.5	216	225	203
Coded Values	1.1	1.9	1.5	16	25	3

For ease of presentation, our evaluations will be made using the coded values. Giving the same weight to all six coded values result in a straight average of 8.1. Note that the addition or loss of a single method B measurement would likely result in a relatively large change of this average. This is not desirable. Intuitively, we know that we should give greater weight to the more precise (and stable) method A results.

For the second example, consider that the measurements by method A and B are equally precise, but that many more measurements were made by method A than by method B.

Method	A						B	
Coded Values (Measured Values−200.)	2.0	1.0	1.5	1.8	1.2	1.7	16.3	16.8

In this example, the values by the two methods differ widely. Here again, we should not take a straight average $(\overline{Y} = 5.3)$. To do so would strongly favor method A, and we have no basis for preferring this method. For these data, it is better to take separate averages for each method, and to then average the two averages $(\overline{Y} = 9.0)$. Note that this is a form of weighting, and that it does not let the larger number of A measurements overpower the B measurements.

3. Basic Statistics of Weighted Averages

It is well known that the weighted average of n values of Y is calculated by the formula:

$$\widetilde{Y} = \frac{\sum\limits_{i=1}^{n} \omega_i Y_i}{\sum\limits_{i=1}^{n} \omega_i} = a_1 Y_1 + a_2 Y_2 + \cdots$$

where ω_i is the weight associated with the value Y_i and the a's are the corresponding coefficients. Statistical theory shows that the variance of this weighted average is minimized when the individual weights are taken as

the inverse of the variance of the individual Y_i, that is, $\omega_i = 1/\mathrm{Var}(Y_i)$. Low weights are given to values with high variance.

Next consider the weighted average of m average values, $\overline{Y}_i$:

$$\widetilde{Y} = \frac{\sum\limits_{i=1}^{m} \omega_i \overline{Y}_i}{\sum\limits_{i=1}^{m} \omega_i} \tag{1}$$

where now $\omega_i = 1/\mathrm{Var}(\overline{Y}_i)$. If both the $\overline{Y}_i$ and the $\mathrm{Var}(Y_i)$ are known, then the weighted average of (1) is easily calculated, and this is the consensus value. The estimation of the proper value for the variance of $\overline{Y}_i$, however, is not always a simple process. To better understand the problem let us return to the second example, given above. It is easy to obtain $\overline{Y}_A = 1.533$ and the variance estimate

$$s^2(\overline{Y}_A) = \frac{s^2(Y_A)}{n_A} = \frac{\sum\limits_{i=1}^{6} (Y_i - \overline{Y}_A)^2}{(6-1)\cdot 6} = 0.0238 = (0.154)^2$$

and $\overline{Y}_B = 16.55$ and $s^2(\overline{Y}_B) = 0.0625 = (0.250)^2$, but one questions the reasonableness of the estimated variances. How can the variances be so small, and the two averages be so far apart? The answer is that the above variance calculations only describe the internal variability of the A or the B measurements, and do not recognize the variability between the sets of measurements. It is quite common, even among very good measurements, to find large differences between different sets of measurements [4]. To obtain a realistic estimate of $\mathrm{Var}(\overline{Y}_i)$, or correspondingly of ω_i, one must evaluate a between set component of variance and include it in $\mathrm{Var}(\overline{Y}_i)$.

The use of a between set component of variance, in effect, treats the collection of systematic errors from the various measurement sets as a source of random variability. The existence of systematic errors in a measurement process may require a word of explanation. The systematic errors that are being described are errors that remain after the extensive scientific development of a measurement process. All major sources of error should have been eliminated, and calibrations with multiple standards should have been made. What frequently happens in this process is that many within set errors are eliminated along with the larger between set errors, such that the sensitivity of the analytical method

378

increases to the point that a lower level of systematic error can now be detected. There are practical limitations to the pursuit of this process, and frequently one must live with a certain detectable level of between set systematic error. Effects such as interferences due to minor sample components will vary in different laboratory environments and these effects are extremely difficult to eliminate.

An essential point in our analysis is the assumption that no information on systematic errors is available that would allow us to place more confidence in any one set of measurements as compared to the others. Thus, in this analysis all sets have equal standing with regard to their possible systematic errors. Our technique can, however, be extended to cover situations involving different assumptions.

The calculation of the between set component of variance is readily accomplished by an iterative procedure, described in section 4. The sample estimate of $\text{Var}(\overline{Y}_i)$ for method i, is obtained by combining the within set component of variance, $s^2_{w_i}$, and the between set component of variance, s^2_b. For the second example:

$$s^2(\overline{Y}_A) = \frac{s^2_{wA}}{6} + s^2_b$$

$$s^2(\overline{Y}_B) = \frac{s^2_{wB}}{2} + s^2_b$$

The within set component of variance for method A is:

$$s^2_{wA} = \frac{\sum\limits_{i=1}^{6} (Y_i - \overline{Y}_A)^2}{(6-1)} = 0.1427$$

Similarly,

$$s^2_{wB} = 0.1250$$

and $s^2_{wA}/6$ and $s^2_{wB}/2$ are equal to 0.0238 and 0.0625, the quantities that we had previously (and incorrectly) called $s^2(\overline{Y}_A)$ and $s^2(\overline{Y}_B)$. For the proper $s^2(\overline{Y}_i)$ one needs to add in s^2_b. With an available s^2_b one calculates estimates for $s^2(\overline{Y}_A)$ and $s^2(\overline{Y}_B)$ and the corresponding weights, ω_A and ω_B, and then proceeds by eq (1) to obtain a valid estimate of the consensus value, $\widetilde{Y}$.

If the $s^2_{w_i}$ are quite similar, as they are in the above example, one can make an improvement by using a more stable pooled s^2_w. There should, of course, be a reasonable scientific and statistical basis for pooling the within set variability. For the current example,

$$s^2_w = \frac{\sum\limits_{i=1}^{6} (Y_{A_i} - \overline{Y}_A)^2 + \sum\limits_{i=1}^{2} (Y_{B_i} - \overline{Y}_B)^2}{(6-1) + (2-1)} = 0.1398 \tag{2}$$

so that

$$s^2(\overline{Y}_A) = \frac{0.1398}{6} + s^2_b$$

and

$$s^2(\overline{Y}_B) = \frac{0.1398}{2} + s^2_b$$

To summarize: The weighting constants used to calculate the consensus value are obtained by taking the inverse of the variances of the various set $\overline{Y}_i$. The proper variances are a combination of the within and the between set components of variance. Under certain circumstances, a more stable pooled within set component of variance may be used.

4. Calculation of the Between Set Component of Variance

The proper weight for $\overline{Y}_i$ is $\omega_i = 1/\text{Var}(\overline{Y}_i)$ and the estimate of this quantity is:

$$\omega_i = \left[\frac{s^2_{w_i}}{n_i} + s^2_b \right]^{-1} \tag{3}$$

Depending on the nature of the data, the within set variance may or may not be pooled. For either case, however, s^2_b must be evaluated. This is accomplished in the following way.

From the definition of ω_i we obtain the relation:

$$\omega_i \text{Var}(\overline{Y}_i) = 1$$

or equivalently $\tag{4}$

$$\text{Var}(\sqrt{\omega_i}\,\overline{Y}_i) = 1$$

For any given set of ω_i, this variance can be estimated from the sample by the formula

$$s^2(\omega_i \overline{Y}_i) = \frac{\sum\limits_{i=1}^{m} \omega_i (\overline{Y}_i - \widetilde{Y})^2}{(m-1)}$$

Equating this estimate to its expected value (unity, see eq (4)), we obtain

$$\frac{\sum\limits_{i=1}^{m} \omega_i(\overline{Y_i} - \widetilde{Y})^2}{(m-1)} = 1 \qquad (5)$$

where $\widetilde{Y}$ is the estimate of the consensus value as given by eq (1). The estimate of $\widetilde{Y}$ depends on knowing the ω_i. These can be calculated from eq (3), once s_b^2 is known. Thus, the only problem is to estimate s_b^2. Equation (5) provides the means for calculating s_b^2 through an iterative process.

Define the function:

$$F(s_b^2) = \sum\limits_{i=1}^{m} \omega_i(\overline{Y_i} - \widetilde{Y})^2 - (m-1) \qquad (6)$$

In view of eq (5), s_b^2 must be such that $F(s_b^2) = 0$. For ease of notation let $s_b^2 = v$. Start with an arbitrarily selected initial value, v_o. It is desired to find an adjustment, dv, such that $F(v_o + dv) = 0$. Using a truncated Taylor series expansion, one obtains:

$$F(v_o + dv) \approx F_o + \left(\frac{\partial F}{\partial v}\right)_o dv = 0$$

and

$$dv = -\left(\frac{F_o}{\dfrac{\partial F}{\partial v}}\right)_o$$

Evaluating the partial derivative in this equation, one obtains:

$$dv = \frac{F_o}{\left[\sum\limits_{i=1}^{m} \omega_i^2 (\overline{Y_i} - \widetilde{Y})^2\right]_o} \qquad (7)$$

The adjusted (new) value for v is:

$$\text{New } v_o = \text{Old } v_o + dv$$

This new value is now introduced in eq (1), (3), (6), and (7) and the procedure is iterated until dv is satisfactorily close to zero. If at any point in the iteration process a negative value is obtained for v, this value should be replaced by zero and the iteration continued. The last v is the s_b^2 we seek. The ω_i and Y are also obtained from this last iteration.

The small data set of our second example will now be used to illustrate the iterative procedure. Let the first estimate for s_b^2 be 100. In calculating the ω_i from eq (3), it is seen that the first term of the right-hand side is a fixed quantity and that the values for the A and B sets have been previously calculated to be .0238 and .0625, respectively. Thus, $\omega_A = 1/(.0238 + 100.) = .0099976$ and $\omega_B = .0099938$. The Y_A and Y_B are 1.533 and 16.550, respectively.

From eq (1), $\widetilde{Y} = 9.0400$
From eq (6), $F_o = .1270$
From eq (7), $dv = 11.28$

The next iteration would start with a value of 111.28 for s_b^2, and would repeat the above set of calculations with this new value. After two additional iterations, v is equal to 112.7120 and dv is less than .0001. The final $\widetilde{Y}$ is 9.0402. (The uncoded Y is, of course, 209.0402.)

In this illustration the initial v value was reasonably close to its final estimate. More discrepant initial values will require only a few additional iterations. It can be shown that the iteration process always converges.

The pooled estimate of the s_w^2 (= 0.1398) could have been used for the above iterative calculations. The final results would be very similar ($s_b^2 = 112.7085$ and $\widetilde{Y} = 9.0399$). For either case, the iterative calculations are easily programmed on a desktop computer.

5. Discussion

The above iterative calculations for the weights and the weighted average are recommended. The calculations are based on the recognition of both within and between group variability. The calculated consensus value is, in general, neither the grand average of all measurements, nor the average of measurement set averages. These overall averages merely describe two opposite weighting situations from our more general weighting eq (3). To illustrate this point consider the case where a pooled s_w^2 is used in eq (3). When the s_b^2 term of this equation is zero, the weights for the Y_i are all proportional to n_i. All individual *measurements* are therefore weighted equally. When, however, s_b^2 is relatively large, the s_w^2/n_i term of eq (3) is essentially without effect, and all the measurement set *averages* are weighted equally. Equation (3) also gives proper weighting for all intermediate cases. In addition, it describes the situation where the within set components of variance are different for different sets of measurements, and takes account of any differences in the number of replicates (n_i) in the various groups.

The ready availability of programmable desktop computers strongly encourages the use the iterative approach. Since one can easily do the calculations, there is little reason to not use proper weighting.

The examples to this point have been chosen to be easily worked by hand. They describe situations where the intuitive answers are obvious. The examples use of only two measurement sets, however, is not recommended in practice since there is a very limited sampling of measured differences between sets. Such a limited sampling results in a s_b^2 estimate that is quite uncertain. The use of many sets of measurements is recommended since this results in greater stability of the estimates.

6. Calculation of the Standard Error of the Weighted Average

All practical applications of the weighted average will require some estimate of its uncertainty. Accordingly, the standard error (standard deviation) of the weighted average should be calculated. The derivation of the standard error of $\widetilde{Y}$ is straightforward if one considers the final ω_i estimates as constants.

$$\widetilde{Y} = \frac{\sum_i \omega_i \overline{Y}_i}{\sum_i \omega_i}$$

and

$$\mathrm{Var}(\widetilde{Y}) = \frac{\sum_i \omega_i^2 \mathrm{Var}(\overline{Y}_i)}{(\sum_i \omega_i)^2} = \frac{\sum_i \omega_i^2 (1/\omega_i)}{(\sum_i \omega_i)^2} = \frac{1}{\sum_i \omega_i}$$

The sample estimate of the standard error of $\widetilde{Y}$ is easily obtained from the final iteration of $\widetilde{Y}$. It is simply the inverse of the square root of the sum of the weights. Note that the latter quantity has already been calculated as the denominator of $\widetilde{Y}$.

$$\text{Standard Error} = \frac{1}{\sqrt{\sum_i \omega_i}}$$

The standard error is reduced by the use of a larger number of sets of measurements, i.e., by more ω_i.

The standard error associated with the $\widetilde{Y} = 209.04$ from our previously worked example is calculated as follows:

$$\text{Standard Error} = \frac{1}{\sqrt{(.0238+112.7120)^{-1}+(.0625+112.7120)^{-1}}} = 7.51$$

This value is seen to be quite reasonable when one remembers that the uncoded group averages for methods A and B were 201.53 and 216.55. Notice in this example that the between set component of variance is the predominant factor in the standard error.

7. Example of an Interlaboratory Experiment Using the Weighted Average

Five laboratories have made a number of determinations for the heat of vaporization of cadmium [5]. In this experiment, each laboratory had a noticeably different replication precision, and each performed a different number of determinations to obtain its average value. We now wish to determine the consensus value (weighted average) from this interlaboratory experiment. The information from the experiment is listed below, along with the s_b^2 calculated by the iterative procedure.

Lab i	Avg. Value	n_i	$s_{w_i}^2/n_i$	s_b^2
1	27,044	6	3	
2	26,022	4	76	
3	26,340	2	464 $\}\times10^3$	105x10³
4	26,787	2	3	
5	26,796	4	14	

In the process of examining the data, the following three averages were calculated.

Average of averages	26,598
Average of individual measurements	26,655
Iterative weighted average	26,713

One notes that the iterative weighted average does not fall between the other two averages. How can this happen? Basically, it is caused by the recognition of the individual within group variances in the weights for the iterative weighted average. To better understand the three averaging processes, let us order the laboratory heat values and include the three sets of "weights" that were used for the averages.

| Ordered Avg. Values | Weights For Avg. of | | | $s_{w_i}^2/n_i$ | s_b^2 |
	Avg. of Averages	Measurements	Iterative Procedure		
26,022	1	4	5.5	76	
26,340	1	2	1.8	464	
26,787	1	2	9.3 $\}\times10^{-6}$	3 $\}\times10^3$	105x10³
26,796	1	4	8.5	14	
27,044	1	6	9.3	3	

Note that the second and third columns contain relative weights while the fourth column contains absolute weights. The relative weights cause no problem for the calculation of the weighted average since inspection of eq (1) shows that any constant multiplier for the relative weights will cancel out. An inspection of the three columns of weights, as well as the ordered laboratory heat values, shows that the weights for the iterative procedure most strongly favor the higher laboratory heat values. Column five of the table, in turn, shows why the iterative weights most strongly favor the higher laboratory heat values; the observed within group variability is smaller for the laboratories that have the higher heat values. This causes the $\text{Var}(\overline{Y}_i)$ for these laboratories to be relatively small and the weights to be relatively large.

This example with actual laboratory data shows that one cannot automatically assume that the average of averages and the average of measurements will bracket the consensus value (weighted average). The weighted average should be calculated. It is more sensitive to the overall experiment and it responds to both the within- and the between group variability.

It will next be shown that the iterative treatment of weighting factors can be easily extended to the problem of fitting lines by weighted least squares (regression).

8. Fitting Lines by Weighted Least Squares

According to statistical theory, the above defined estimate of the weighted average is the value that minimizes the sum of the weighted squared deviations of the observed data (from the weighted average value). It is a least squares estimate. A similar treatment is used in weighted linear regression. Here, a pair of parameters, namely the intercept and the slope of the line, are estimated, rather than a single average. The procedure, however, is again the minimization of the weighted sum of squares of deviations. Here, too, both within set and between set components of variance should be evaluated.

Consider the situation where a laboratory calibrates an instrument using a series of standards. The laboratory may not always make the same number of replicate measurements with the different standards. Thus there are different sets of replicate instrument measurements (Y) corresponding to a series of accurately determined standard values (X). An example of a linear calibration process is given in figure 1. Let us assume that the linearity of the calibration curve has previously been established. An examination of the figure shows that the variability in the Y direction among replicates obtained at the same X value is relatively small when compared with the scatter of the clusters of points about the straight line. Thus, two sources of variability are sug-
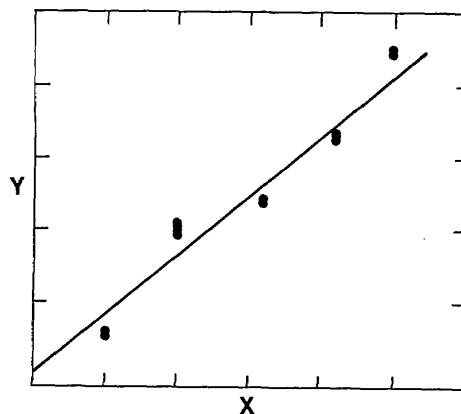

Figure 1

gested by the data. The Y replication variability associated with a given X value is analogous to the previously described within set component of variance, $s^2_{w_i}$, and the variation shown by the scatter of the clusters of points about the fitted line is analogous to the between set components of variance, s^2_b.

The observed variance for the j-th replicate Y_{ij} measurement made at a given X_i value will consist of the sum of the within- and the between set components of variance.

$$s^2(Y_{ij}) = s^2_{w_i} + s^2_b$$

For convenience of calculation, it is desirable to deal with the averages of the replicate measurements. The average of n_i replicate measurements is denoted as $\overline{Y}_i$. The observed variances for the averages are given by:

$$s^2(\overline{Y}_i) = \frac{s^2_{w_i}}{n_i} + s^2_b$$

The within set variances of the above equation can be evaluated for each distinct X_i value. It is possible, if there is a consistent measurement process over the full range of values, to obtain a pooled estimate of the within set component of variance. This pooled estimate is obtained in the same manner as described by eq (2), above. In the current application, the different X_i values correspond to the previously described different measurement sets and there are now as many summations in the numerator and denominator of eq (2) as there are distinct X_i values.

Let us now assume that an appropriate between set component of variance is available. The weights, $\omega_i =$

$1/\mathrm{Var}(\overline{Y}_i)$ can be evaluated, and a standard weighted linear regression of $\overline{Y}_i$ on X_i can be carried out (see either the Appendix, or Ref. [6]). Thus, the regression problem using weighted least squares centers on the determination of s_b^2.

The s_b^2 value, for the regression case with an intercept and slope, can be determined by the general iterative approach given above. Equation (3) now refers to the within- and between set random errors in the Y measurements. It is now used along with the following modified iteration equations:

$$F\,(s_b^2) \;=\; \sum_{i=1}^{m} \omega_i(\overline{Y}_i - \hat{Y}_i)^2 \;-\; (m-2) \qquad (8)$$

$$ds_b^2 \;=\; \cfrac{F_0}{\left[\displaystyle\sum_{i=1}^{m} \omega_i^2\,(\overline{Y}_i - \hat{Y}_i)^2\right]_0} \qquad (9)$$

where

$\hat{Y}_i$ = weighted least squares fitted value, i.e.,
$\hat{Y}_i = a + bX_i$

The major modification is that instead of using $\widetilde{Y}$, we use a weighted least squares fitted value $\hat{Y}_i$. Equation (8) uses $(m\text{-}2)$ rather than $(m\text{-}1)$ degrees of freedom since we are now estimating two parameters, i.e., the intercept and the slope.

The procedure for iteration is little changed. An arbitrary initial estimate for s_b^2 is taken and used with (3) to obtain the weights. Next, a weighted linear regression is made of $\overline{Y}_i$ on X_i to obtain estimates a, b, and $\hat{Y}_i$. This is followed by the use of eq (8) and (9) to calculate a correction for s_b^2. The whole procedure is then repeated until the correction for s_b^2 is negligible. The final s_b^2, a, b, and $\hat{Y}_i$ are then saved for further interpretation and use.

The above procedure for performing a weighted linear least squares fit can be easily extended to a weighted quadratic, or higher order, regression of $\overline{Y}_i$ on X_i. For example, to fit the equation $Y_i = a + bX_i + cX_i^2$ change, in equation (8), the $(m\text{-}2)$ to $(m\text{-}3)$ to account for the addition of coefficient c, and use a quadratic fitted $\hat{Y}_i$ in eq (8) and (9).

9. An Example of a Weighted Least Squares Fit

Let us examine the effect of different weighting factors on the determination of the intercept and slope of a calibration line. A greatly simplified example is shown in figure 2. The true line has both unit intercept and slope. For this example, let us assume that "interferences" for the $X = 1$ and the $X = 5$ standard samples are such that the measured values will be about 0.2 units high. Similarly, the $X = 2$ and $X = 4$ standards yield results that are about 0.2 units low. Duplicate measurements are made, and for simplicity assume that these measurements have a fixed s_w^2 value of 0.0008 (as shown in fig. 2). With equal numbers of replicate measurements, both the unweighted and the iterative weighted regression calculations give the correct values for the intercept and the slope.

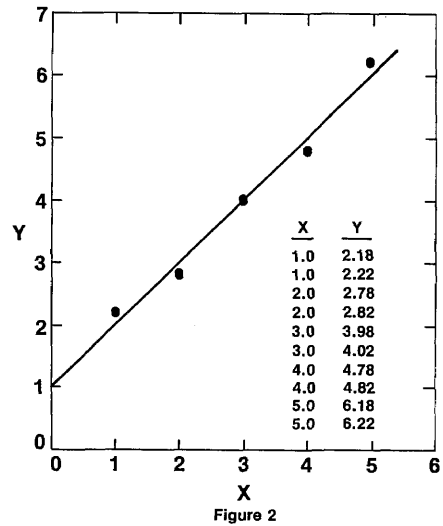| X	Y
1.0	2.18
1.0	2.22
2.0	2.78
2.0	2.82
3.0	3.98
3.0	4.02
4.0	4.78
4.0	4.82
5.0	6.18
5.0	6.22

Figure 2

Let us now, however, say that the experimenter is particularly interested in determining the intercept and that he/she therefore makes six rather than two replicate measurements using the $X = 1$ standard. For the sake of simplicity, assume that the six Y measurements again center at 2.2 and that $s_w^2 = .0008$. Even though everything looks nominally the same, the unweighted regression calculation gives an intercept of 1.145 and a slope of 0.9636. Obviously, the six points at $X = 1$ have pulled the left side of the line upward. If we carried out the regression calculation using only the average Y value for each X value we would obtain the correct intercept and slope values. The average Y values are not affected by the number of measurements used in each average.

In this example, in which appreciably more measurements were made for one standard than for the others, and the replication error was relatively small, the

unweighted regression leads to erroneous results. A proper weighting procedure must prevent the measurements at one standard from unduly influencing the fit. Equation (3) of our iterative weighted regression calculations will properly control the weighting. In this example, the s_b^2 term in eq (3) dominates the weighting. Use of the iterative weighted linear regression gives $a = 1.0008$ and $b = 0.9998$. If the data from this example were real laboratory data, then our calculated a and b would be the appropriate sample estimates.

10. Design of Experiments

The interferences associated with the five standards of the above illustrative example have been ideally and artificially balanced. In real life situations the order in which the interferences will occur will tend to be more random. When the replication error is small, i.e., s_w^2 is small relative to s_b^2, the positions of the $(X_i, \overline{Y_i})$ points will be mainly affected by these random sample interferences. In that case, the use of a larger number of standards over the range of measurement interest is recommended since this favors a more even distribution of these interferences, and a more accurate determination of the line. Furthermore when s_w^2 is small relative to s_b^2, the use of large numbers of *replicate* measurements is not recommended since these measurements are very inefficient in determining the position of the $(X_i, \overline{Y_i})$ points.

Consider next the situation shown in figure 3, where s_w^2 is large relative to s_b^2. Here all of the average points $(X_i, \overline{Y_i})$ are very uncertain. The interferences of each standard sample is now completely overshadowed by the variability in the replicate measurements. For this situa-
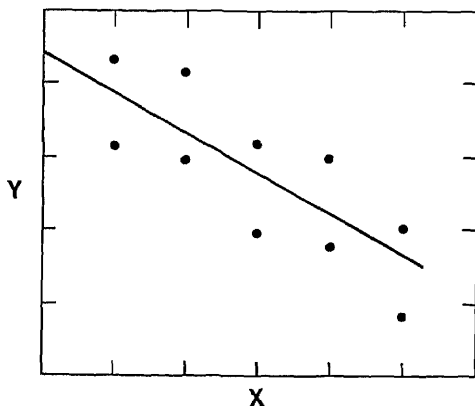
tion one should make many replicate measurements with all of the standard samples so as to minimize the replication uncertainty.

11. Summary and Conclusions

Calculation of consensus values, both in the form of the weighted average or the weighted least squares regression, requires a knowledge of the within- and the between set components of variance. The individual or the pooled within set components of variance can be directly calculated from the experimental data. The between set component of variance can conveniently be calculated from the experimental data using an iterative technique which is based on a truncated Taylor series expansion. Consensus value(s) are also obtained by this iterative technique.

A simple intuitive understanding of the within- and between set components of variance allows one to more efficiently design experiments for obtaining consensus values.

The logical arguments for use of the within- and between set components of variance can be extended to other areas of statistical analysis. Work is in progress for extending the current techniques to nested analyses of variance.

12. References

[1] Birge, R. T., The Calculation of Errors by the Method of Least Squares, Phys. Rev., 40, 207–227 (1932).

[2] Mandel, J. and Paule, R. C., Interlaboratory Evaluation of a Material with Unequal Numbers of Replicates, Anal. Chem., 42, 1194-7 (1970), and Correction, Anal. Chem., 43, 1287 (1971).

[3] Cochran, W. G., The Combination of Estimates from Different Experiments, Biometrics, 10, 101-29 (1954).

[4] Youden, W. J., Enduring Values, Technometrics, 14, 1-11 (1972).

[5] Paule, R. C. and Mandel, J., Analysis of Interlaboratory Measurements on the Vapor Pressure of Cadmium and Silver, NBS Special Publication 260-21 (1971).

[6] Draper, N. R. and Smith, H., *Applied Regression Analysis*, 2nd ed., Section 2.11 (John Wiley and Sons, N.Y., 1981).

Appendix

Figure 3

The formulas for estimating the slope and intercept by weighted least squares are straightforward. The slope is calculated from the observed m sets of $(X_i, \overline{Y_i})$ points.

$$b = \frac{\sum_{i=1}^{m} \omega_i (X_i - \widetilde{X})(\overline{Y}_i - \widetilde{Y})}{\sum_{i=1}^{m} \omega_i (X_i - \widetilde{X})^2}$$

where

$$\widetilde{X} = \frac{\sum_{i=1}^{m} \omega_i X_i}{\sum_{i=1}^{m} \omega_i}$$

and

$$\widetilde{Y} = \frac{\sum_{i=1}^{m} \omega_i \overline{Y}_i}{\sum_{i=1}^{m} \omega_i}$$

The intercept is obtained by the following formula.

$$a = \widetilde{Y} - b\widetilde{X}$$

The interested reader may also wish to calculate the standard errors of the above estimates of the slope and the intercept, the formulas are:

$$s_b = \sqrt{\frac{1}{\sum_{i=1}^{m} \omega_i (X_i - \widetilde{X})^2}}$$

$$s_a = \sqrt{\frac{\sum_{i=1}^{m} \omega_i X_i^2}{\left[\sum_{i=1}^{m} \omega_i\right]\left[\sum_{i=1}^{m} \omega_i (X_i - \widetilde{X})^2\right]}}$$

A more detailed explanation of weighted least squares fitting processes is contained in Ref. [6].