*Biochem. J.* (1961) **80**, 324

# Statistical Estimations in Enzyme Kinetics

By G. N. WILKINSON

*Division of Mathematical Statistics, Commonwealth Scientific and Industrial Research Organization, University of Adelaide, South Australia*

(*Received* 12 *December* 1960)

Although elementary statistical methods such as the calculation of means and their standard errors are commonly employed, application of the statistical methods of regression analysis in enzyme kinetic studies has received little attention. In estimating the kinetic parameters of the Michaelis–Menten equation, for instance, graphical methods such as the double-reciprocal plot (Lineweaver & Burk, 1934) are generally used, without supplementary statistical calculations. These methods have been reviewed by Dixon & Webb (1958).

Graphical methods, however, do not usually provide any measure of the precision of the determinations, a knowledge of which is necessary for a proper evaluation of the results in relation to theoretical considerations, or for a comparison of results obtained under differing experimental conditions. The main purpose of a statistical analysis is to supply, not only more accurate estimates, but also the necessary measures of precision.

This paper gives an account of the weighted and non-linear regression methods applicable to general problems in enzyme kinetics. Application of the basic principles is illustrated in detail with the computational methods given for estimating the kinetic parameters $K_m$ and $V$ of the Michaelis–Menten equation, and the relevant standard errors. Further illustration is provided by a description of the analyses applied in estimating dissociation constants for enzyme and substrate (cf. Dixon, 1953) from the data of Atkinson, Jackson & Morton (1961).

The methods described in this paper were applied in the kinetic studies described in the preceding paper by Atkinson *et al.*

## BASIC PRINCIPLES AND METHODS

### *Random variation; terminology*

Most experimental determinations are subject to fluctuations of an unpredictable kind, usually on account of experimental factors which are not under rigid control and the inherent mechanical limitations of the experimental apparatus, and sometimes on account of the inherent variability in the phenomena under investigation, as with radioactive emissions. In this sense experimentally determined quantities can be described as 'random variates,' and associated with each observation or determination is a 'conceptual population' of all possible values of the random variate that might have arisen in similar circumstances to those which have produced the actual value(s). The relative frequencies with which different values of the variate occur in this conceptual population constitute the distribution of the variate (which may have more than one distribution according to the circumstances envisaged).

The true mean, or expected value, of a random variate $y$ is the average value of $y$ in the relevant conceptual population, and is denoted here by $E(y)$, $\mu_y$, or simply $\mu$. (The true mean is to be distinguished from the sample mean $\bar{y}$ of a series of observations.) If the expected value of a determination deviates from a specified theoretical value, the determination is described as being 'biased' in that respect.

The variance of $y$ is the average value, $E(y-\mu)^2$, of squared deviations from the mean in the conceptual population, and is denoted here by $V(y)$, $\sigma_y^2$, or $\sigma^2$. The square root $\sigma$ of the variance is referred to alternatively as the standard deviation (S.D.) or standard error (S.E.) according to context; 'standard deviation' is used primarily when referring to the basic variability in the experimental data, and 'standard error' when referring to the precision of a statistic (such as a mean or a regression coefficient) as an estimate of some parameter. The coefficient of variation, C.V.($y$), is the standard deviation of $y$ expressed as a fraction of the mean; C.V.($y$) $= \sigma/\mu$.

The covariance of two random variates $y_1$ and $y_2$ is the average value of the product of corresponding deviations in the relevant population of all possible pairs of values $(y_1, y_2)$:

$$\text{Cov.}\,(y_1, y_2) = E(y_1-\mu_1)\,(y_2-\mu_2).$$

The correlation, $\rho$, between $y_1$ and $y_2$ is a normalized form of the covariance:

$$\rho = \frac{\text{Cov.}\,(y_1, y_2)}{\sigma_1\sigma_2}.$$

Two variates are described as 'statistically independent' if the probability that one of them takes

any value or range of values is independent of the value taken by the other variate. It follows that the covariance or correlation of two statistically independent variates is zero. (The converse is true only for normally distributed variates.)

The parameters of random variation described above (and generally the parameters of any physical system) are usually not known with exactitude, but are, or have been, ascertained with an attendant degree of uncertainty, by estimation from experimental data. It is desirable in most contexts that a semantic distinction should be evident, between the parameters themselves and the corresponding estimates.

*Combination and transformation of variability*

Quantitative studies usually involve a sequence of calculations based on the experimental data. The following rules enable the influence of experimental variability to be determined at any stage of this process, at least to a first order of approximation:

Change in scale of a quantity $y$ by a factor $\lambda$ clearly affects its mean and standard deviation similarly, so that

$$E(\lambda y) = \lambda E(y) \quad (1)$$

and $$\text{S.D.}(\lambda y) = \lambda \text{S.D.}(y), \quad V(\lambda y) = \lambda^2 V(y). \quad (2)$$

The basic rules for the addition of two or more quantities are

$$E(y_1+y_2) = E(y_1)+E(y_2), \quad (3)$$

$$V(y_1+y_2) = V(y_1)+V(y_2)+2\,\text{Cov.}(y_1, y_2), \quad (4)$$

the latter being deducible from (3) since

$$E(y_1+y_2-\mu_1-\mu_2)^2 = E(y_1-\mu_1)^2+E(y_2-\mu_2)^2+2E(y_1-\mu_1)(y_2-\mu_2).$$

If $y_1$ and $y_2$ are statistically independent the variances are simply additive. The rules (1)–(4) in combination determine the mean and variance (hence standard error) of any linear combination of variates and, in particular, the variances of regression statistics as given below.

For non-linear transformations the following rules are only approximate since the non-linearity introduces biases. The biases will be very slight, however, if the transformation function exhibits little curvature over the range of experimental variation, as illustrated in Fig. 1.

Small deviations can be represented approximately by the linear part of a Taylor expansion,

$$f(y) \approx f(\mu_y)+(y-\mu_y)f'(\mu_y),$$

where $f'(\mu_y)$ is the first derivative of $f(y)$ at $\mu_y$. This is evident also from geometrical considerations (Fig. 1). Since $y-\mu_y$ averages to zero in the conceptual population of $y$ values,

$$E[f(y)] \approx f(\mu_y), \quad (5)$$

and squaring and averaging $f(y)-f(\mu_y)$ give

$$V[f(y)] \approx f'^2(\mu_y)\,V(y). \quad (6)$$

Bias in the last formula arises from bias in the transformed mean as well as from neglecting higher-order terms of the Taylor expansion, but again will be small under the conditions mentioned.

Use of the formula (6) is illustrated by the transformation of an estimated $K_m$ into the negative logarithm $pK_m$:

$$\frac{d}{dx}\log x = \log e \times \frac{1}{x}$$

so that

$$\text{S.E.}(pK_m) \approx \log e \times \frac{\text{S.E.}(K_m)}{K_m},$$

$$\approx 0{\cdot}4343\ \text{C.V.}(K_m).$$

The generalization of (5) and (6) for functions of two or more variates is straightforward. For a function of two variates,

$$E(f(y_1, y_2)) \approx f(\mu_1, \mu_2), \quad (7)$$

$$V(f(y_1, y_2)) \approx f'^2_{y_1} V(y_1)+f'^2_{y_2} V(y_2) + 2f'_{y_1} f'_{y_2} \text{Cov.}(y_1, y_2), \quad (8)$$

where $f'_{y_1}$, $f'_{y_2}$ are the partial derivatives of $f$ at $(\mu_1, \mu_2)$. The covariance term vanishes if $y_1$ and $y_2$ are statistically independent. An important special case is that of the product or ratio of variates, $z = y_1y_2$ or $z = y_1/y_2$, in which case the variance rule (8) on division by $\mu_z^2$ becomes

$$\text{C.V.}^2(z) \approx \text{C.V.}^2(y_1)+\text{C.V.}^2(y_2) \pm 2\rho\,\text{C.V.}(y_1)\,\text{C.V.}(y_2), \quad (9)$$

the minus sign applying for the ratio. If $y_1$ and $y_2$ are statistically independent, the correlation $\rho$ is zero, so that the squares of the coefficients of variation are simply additive.

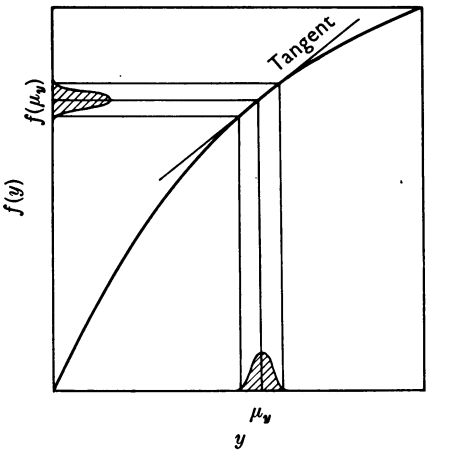


Fig. 1. Non-linear transformation of an error distribution. The slope of the tangent is the first derivative of $f(y)$ at $\mu_y$.

*Regression analysis*

The relevance of the methods described below, for the analysis of a series of observations, depends on the assumptions that the observations are in an effective sense statistically independent and that their true distributions do not differ appreciably from the familiar Gaussian or normal form. With these assumptions the appropriate form of analysis is determined by the behaviour of the true means and variances.

A regression analysis is appropriate when the mean of a variate $y$ depends on factors $x$, such as temperature or pH, which have been varied in the experiments. If a series of observations $y$ are available, together with the corresponding values of determining variables $x$, and if the variance of $y$ is sensibly homogeneous over the series of observations, a regression function $Y(x)$ is fitted to the data so that the sum of squares of deviations of the observations $y$ from the fitted function, $\Sigma(y-Y)^2$, is as small as possible (the method of least squares).

The series of observations may not be homogeneous in variance, however. It may be that $\sigma^2$ varies systematically with $\mu_y$ or that the observations are of intrinsically differing accuracy on account of the experimental procedure. It is evident that some account must be taken of the differing accuracies in fitting the regression function, and the appropriate method is to fit the function so that the weighted sum of squares of deviations, $\Sigma w(y-Y)^2$, is a minimum, the relative weights $w$ being the relative amounts of 'information' and inversely proportional, in this context, to the variances of the $y$ values. (The more general definition of 'information' is given by Fisher, 1925.)

The weighted-regression procedure is equivalent to fitting an unweighted regression to an enlarged set of points, each point of the original set being repeated the appropriate $w$ times. The essential differences, therefore, are that weighted rather than simple means of observations are calculated and likewise weighted sums of squares and products. Details of the weighted analyses are set out below for the two cases relevant to this paper, in which the true regression is linear and dependent on either one or two determining variables:

$$\text{(i)}\quad \mu_y = \alpha + \beta x,$$
$$\text{(ii)}\quad \mu_y = \alpha + \beta_1 x_1 + \beta_2 x_2.$$

*Case* (i). The fitted regression is of the form $Y = a + bx$. The weighted means are

$$\bar{x} = \Sigma wx/\Sigma w, \quad \bar{y} = \Sigma wy/\Sigma w,$$

and the estimated regression coefficient $b$ and the constant term $a$ are given by the formulae:

$$b = \frac{\Sigma w(x-\bar{x})\,(y-\bar{y})}{\Sigma w(x-\bar{x})^2} = \frac{\Sigma wxy - (\Sigma wx)\,(\Sigma wy)/\Sigma w}{\Sigma wx^2 - (\Sigma wx)^2/\Sigma w},$$
$$a = \bar{y} - b\bar{x}.$$

If the variances of the observations $y$ are $\sigma^2/w$, the variance factor $\sigma^2$ is estimated by the residual mean square,

$$s^2 = \frac{\Sigma w(y-Y)^2}{n-2} = \frac{\Sigma wy^2 - \bar{y}\Sigma wy - b\Sigma w(x-\bar{x})\,(y-\bar{y})}{n-2},$$

the divisor of the sum of squares of deviations (or residual sum of squares) being the number of degrees of freedom of the estimated variance, two less than the number of observations in this instance. The variances of $a$ and $b$ are given by the formulae:

$$V(b) = \sigma^2/\Sigma w(x-\bar{x})^2,$$
$$V(a) = \sigma^2[1/\Sigma w + \bar{x}^2/\Sigma w(x-\bar{x})^2],$$

the latter following from (4) above since it may be easily shown that $b$ and $\bar{y}$ are statistically independent. Estimated variances are obtained by substituting $s^2$ for $\sigma^2$.

*Case* (ii). The fitted regression is of the form

$$Y = a + b_1 x_1 + b_2 x_2.$$

Weighted means $\bar{y}, \bar{x}_1, \bar{x}_2$ are calculated as in (i) above. Two equations are obtained for the regression coefficients $b_1$ and $b_2$,

$$a_{11}b_1 + a_{12}b_2 = p_1,$$
$$a_{12}b_1 + a_{22}b_2 = p_2,$$

where

$$a_{11} = \Sigma w(x_1-\bar{x}_1)^2, \quad a_{12} = \Sigma w(x_1-\bar{x}_1)\,(x_2-\bar{x}_2), \quad a_{22} = \Sigma w(x_2-\bar{x}_2)^2,$$
$$p_1 = \Sigma w(x_1-\bar{x}_1)\,y, \quad p_2 = \Sigma w(x_2-\bar{x}_2)\,y.$$

The solutions of the equations are

$$b_1 = c_{11}p_1 + c_{12}p_2,$$
$$b_2 = c_{12}p_1 + c_{22}p_2,$$

where

$$c_{11} = a_{22}/\Delta, \quad c_{22} = a_{11}/\Delta, \quad c_{12} = -a_{12}/\Delta$$

and

$$\Delta = a_{11}a_{22} - a_{12}^2.$$

The constant term is

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2.$$

Similarly, as in (i) above, the variance factor $\sigma^2$ is estimated by the residual mean square:

$$s^2 = \frac{\Sigma w(y-Y)^2}{n-3} = \frac{\Sigma wy^2 - \bar{y}\Sigma wy - b_1p_1 - b_2p_2}{n-3}.$$

The variances and covariances of the regression statistics are

$$V(b_1) = c_{11}\sigma^2, \quad \text{Cov.}(b_1, b_2) = c_{12}\sigma^2, \quad \text{Cov.}(b_1, \bar{y}) = 0,$$
$$V(b_2) = c_{22}\sigma^2, \quad \text{Cov.}(b_2, \bar{y}) = 0,$$
$$V(a) = \sigma^2(1/\Sigma w + c_{11}\bar{x}_1^2 + c_{22}\bar{x}_2^2 + 2c_{12}\bar{x}_1\bar{x}_2).$$

*Omission of the constant term.* If it is known that the true regression passes through the origin ($\alpha = 0$), the fitted regression should be likewise constrained ($a = 0$). The essential modification of the formulae given above for this case is that crude sums of squares and products such as $\Sigma wx^2$, $\Sigma wxy$ are substituted everywhere for the corrected sums of squares and products $\Sigma w(x-\bar{x})^2$ etc. In the formulae for the residual mean squares, $s^2$, the correction factor $\bar{y}\Sigma wy$ is omitted and the degrees of freedom are correspondingly increased by unity to (i) $n-1$, (ii) $n-2$.

*Use of estimated weights.* The relative weights $w$ may not be known exactly, so that an approximate analysis must be performed with estimated weights. In some situations a more accurate analysis may then be obtained by repeating the process with improved estimates of the weights supplied by the first analysis. More than a single repetition of the process is seldom necessary, however, except as a check.

*Non-linear regression functions.* These can be fitted, with little extra complication, by linear-regression methods; the essential difference is that provisional estimates of certain

of the unknown parameters are required. If a function $f(x, c)$ is non-linear in a parameter $c$ whose value is to be estimated, and if a good provisional value $c_0$ is available, the following linear approximation may be used:

$$f(x, c) \approx f(x, c_0) + (c - c_0) f'_c(x, c_0),$$

where $f'_c$ is the first derivative of $f$ with respect to $c$, evaluated here at $c_0$, and a linear regression analysis, in terms of the provisional values of $f$ and $f'_c$ corresponding to the values of $x$, determines $c - c_0$ as a linear regression coefficient. In this sense the process may be described as a 'fine adjustment' of the provisional estimate $c_0$. A more accurate analysis can be obtained, if necessary, by repeating the process with the adjusted estimate as a new provisional value. The extension of the method for a function non-linear in two or more parameters is straightforward.

### *Applications*

*Estimation of $K_m$ and $V$.* The Michaelis–Menten relation between velocity of reaction ($v$) of an enzyme with a substrate and the substrate concentration ($s$),

$$v = \frac{s}{s + K_m} V,$$

can be expressed in the linear forms

$$\frac{1}{v} = \frac{1}{V} + \frac{K_m}{V} \cdot \frac{1}{s} \quad \text{or} \quad \frac{s}{v} = \frac{K_m}{V} + \frac{s}{V}.$$

With the usual methods of determining velocities of reaction, as described in the preceding paper (Atkinson *et al.* 1961), the velocity determinations will be reasonably homogeneous in variance. If $v$ has variance $\sigma^2$, the variances of $1/v$, $s/v$ are

$V(1/v) = \sigma^2/\mu_v^4$, $V(s/v) = \sigma^2 s^2/\mu_v^4$ (cf. equation 6).

In fitting the above linear forms, therefore, the correct relative weights are $\mu_v^4$ and $\mu_v^4/s^2$ respectively for the two types of fit.

As the true velocities $\mu_v(s)$ are unknown, provisional fits of the lines can be obtained by using observed velocities $v$, which supply weights $v^4$ and $v^4/s^2$ respectively. In the case of the double reciprocal plot the weighted means are

$$\overline{(1/v)} = \Sigma v^3/\Sigma v^4, \quad \overline{(1/s)} = (\Sigma v^4/s)/\Sigma v^4,$$

and the weighted fit of the straight line

$$1/v = a + b(1/s)$$

as described above leads, with some simplification, to the formulae for provisional estimates of $K_m$ and $V$ given in Table 1. The same formulae are obtained whichever linear form is fitted, the weighted analyses being equivalent.

Fine adjustment of the provisional estimates may be obtained by fitting the Michaelis–Menten function directly in the hyperbolic form, with an unweighted analysis. This process is preferable to that of refitting the function in one of the linear forms above with revised estimates of the relevant weights, since not only are the calculations simpler but standard errors for the estimates are derived more directly. In terms of the provisional estimates $K_m^0$, $V^0$, the Michaelis–Menten relation can be expressed in the approximate linear form:

$$\begin{aligned} v &\approx \frac{V}{V^0}\left\{\frac{sV^0}{s + K_m^0} - (K_m - K_m^0)\frac{sV^0}{(s + K_m^0)^2}\right\}, \\ &\approx b_1 f(s) + b_2 f'_K(s), \end{aligned}$$

where $b_1 = V/V^0$, $b_2 = b_1(K_m - K_m^0)$, $f(s)$ is the provisional fit of the Michaelis–Menten function, and $f'_K(s)$ is the first derivative of the provisional fit with respect to $K_m$. The problem is thus reduced to that of fitting a bilinear regression (without a constant term) as described above. The computational procedure is set out in Table 2.

The standard errors for $V$ and $K_m$ are derived from those for $b_1$ and $b_2$ by application of the rules (2) and (9). In particular,

$$\text{S.E. } (K_m) = \text{S.E. } (b_2)/b_1,$$

since the coefficient of variation of the factor $b_1$ is negligible in comparison with that for $K_m - K_m^0$.

Use of the provisional value $V^0$ is not strictly necessary since it cancels both in $b_1, f$ and in $b_2, f'$. The advantages are, however, that direct comparison of the provisionally determined velocities with those observed provides a useful check for arithmetical errors, and that $b_2$ is almost exactly the fine adjustment for $K_m$. With repetition of the process, $b_1$ and $b_2$ tend to unity and zero respectively.

*Estimation of dissociation constants.* The estimation of dissociation constants from a series of determinations of $K_m$ corresponding to a range of values of substrate pH is discussed below. The appropriate relation of $K_m$ to hydrogen-ion concentration $H$ which must be fitted is

$$K_m = \tilde{K}_m\left(1 + \frac{H}{K_1}\right)\left(1 + \frac{K_2}{H}\right),$$

in which $\tilde{K}_m$ is the asymptotically minimum value for $K_m$, and $K_1$, $K_2$ the relevant dissociation constants. Expressed equivalently as

$$K_m = a + b_1 H + b_2(1/H),$$

the function is in the form of a linear regression, the determining variables being $x_1 = H$, $x_2 = 1/H$. A weighted fit of the regression is necessary on account of the variation in precision of the determinations of $K_m$. The estimates of $K_1$ and $K_2$ derived from $a$, $b_1$, and $b_2$ are in this instance regarded as provisional since a revision of the relative weights is necessary.

For fine adjustment of the provisional estimates it is appropriate, in this case, to fit the $K_m(H)$ function in the negative-logarithm or p$K_m$ form. In terms of the provisional values $K_m^0$ calculated from the provisional estimates of $K_1$, $K_2$ and $\tilde{K}_m$,

the function can be expressed in the approximate linear form

$$\mathrm{p}(K_m/K_m^0) \approx \mathrm{p}(\tilde{K}_m/\tilde{K}_m^0) - \mathrm{p}(K_1/K_1^0)\frac{H}{K_1^0+H} + \mathrm{p}(K_2/K_2^0)\frac{1/H}{1/K_2^0+1/H},$$

so that the problem is again reduced to that of fitting of a linear regression, the constant term and the regression coefficients being the required adjustments for $\mathrm{p}\tilde{K}_m^0$, $\mathrm{p}K_1$ and $\mathrm{p}K_2^0$. The determining variables are the partial derivatives of the provisionally fitted $\mathrm{p}K_m$ function with respect to $\mathrm{p}K_1$, $\mathrm{p}K_2$. The regression calculations supply the standard errors of $\mathrm{p}\tilde{K}_m$, $\mathrm{p}K_1$ and $\mathrm{p}K_2$ directly.

*Combined estimates of variability.* In a series of similar experiments the basic experimental variability may be sensibly homogeneous, in which case a more accurate estimate of variance may be obtained by calculating a weighted average of the estimates from individual experiments. The individual estimates are weighted according to their degrees of freedom, so that the weighted average is the ratio of the pooled (or added) residual sums of squares from the individual experiments to the total number of degrees of freedom. In the preceding paper (Atkinson *et al.* 1961) a pooled estimate of the variance of velocity determinations was obtained in this way from 17 experiments, with a total of 66 degrees of freedom.

### *Rounding-off errors*

The arithmetic process of rounding off to a given number of figures causes an artificial increase in the variance of computed quantities, which must be kept as small as practicable to avoid appreciable loss of information. The increase in variance of a quantity due to rounding off is $c^2/12$, where $c$ is the rounding-off interval. A simple rule is to retain, in rounding off, the first two figures that are affected by experimental variability. Thus if a quantity has the value 0·12345 say, and the standard error of the quantity is 0·003, the second decimal digit may be in error by unity and the third is appreciably affected. The variance of the rounded-off quantity 0·123 is $(0{\cdot}003)^2+(0{\cdot}001)^2/12$, in which case the relative loss of information is 1/109, or less than 1 %. If the standard error had been much less than 0·003, the fourth decimal should also have been retained, and, in the absence of any precise knowledge of experimental variability, retention of an extra figure is a desirable safeguard.

A number of aspects have not been mentioned here, such as the calculation of probabilities, significance tests and fiducial or confidence limits. A more detailed account of general principles and methods is given by Fisher (1925), and of regression methods by Williams (1959).

## ENZYME KINETIC DETERMINATIONS

### *Estimation of $K_m$ and $V$*

The procedure set out below for estimating the kinetic parameters $K_m$ and $V$ of the Michaelis–Menten equation consists of two stages, the calculation of provisional estimates, and the fine adjustment of the provisional values, the latter stage also supplying standard errors for the estimates. A locally defined notation is used to describe each step of the computations in terms of the preceding computations; consequently some symbols, such as $\alpha$, $\beta$, $\gamma$, ..., occur with different meanings in each stage. Numerical illustration of the calculations is given in parallel, with data from Atkinson *et al.* (1961).

*Provisional estimates of $K_m$ and $V$.* (i) *Graphical.* A number of graphical methods for determining $K_m$ and $V$ have been suggested, based on various linear forms of the Michaelis–Menten relation (see Dixon & Webb, 1958). Of the two linear forms:

$$\frac{1}{v} = \frac{1}{V} + \frac{K_m}{V}\cdot\frac{1}{s}, \quad \frac{s}{v} = \frac{K_m}{V} + \frac{s}{V},$$

the double reciprocal form (Lineweaver & Burk, 1934) is the more commonly used. However, statistical considerations suggest that more accurate subjective estimates can be obtained by fitting the alternative form linear in $s/v$ and $s$. Assuming that $v$ is reasonably homogeneous in variance, the reciprocal $1/v$ exhibits much greater variation in accuracy over the practical ranges of substrate concentration than does $s/v$. This is illustrated in Fig. 2, which shows the variations in the relative weights (reciprocals of variance) of $1/v$ and $s/v$. (The scales are arbitrary for each weight function.)

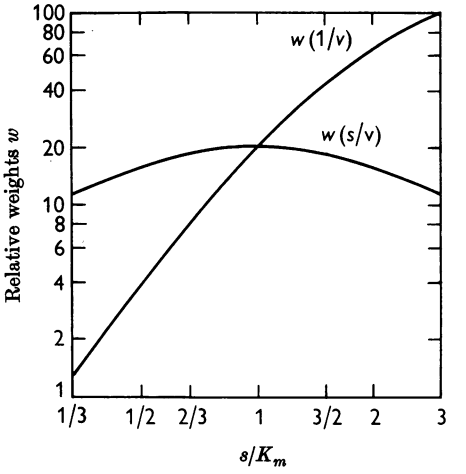


Fig. 2. Variation in relative weights of $1/v$ and $s/v$ over a range of substrate concentrations (logarithmic scales).

Over the range of substrate concentration from one-third to three times the $K_m$, the relative weight of $1/v$ varies by a factor of 80, whereas the relative weight of $s/v$ varies by less than a factor of 2. If a statistical analysis were carried out with the correct weights, identical and hence equally precise estimates would be obtained with either fit. However, appreciable deviations, of whatever kind, from correct weighting will result in less accurate estimates of $K_m$ and $V$, and with subjective fitting it would seem much more difficult to take proper account of the variation in weight when this is extreme than when it is relatively slight, as with $s/v$.

A third linear form discussed by Hofstee (1952), $v = V - K_m v/s$, suffers to a less extent from the same statistical disadvantage as the double reciprocal form, with the added complication that both the variables are affected by the experimental variability in $v$. Hofstee suggests that a certain type of non-linearity is more readily detectable with the third linear form, but this argument may be questioned.

(ii) *Statistical.* Although in any case it may be necessary to determine a graphical plot in order to detect significant experimental aberrations, such as inhibition by an excess of substrate (Dixon & Webb, 1958), statistically-determined provisional estimates take little time to compute, and their greater accuracy improves the accuracy of the fine-adjustment process. The calculations are set out in Table 1, the relevant formulae being derived by the weighted fit of a linear regression, as described in an earlier section.

With a standard desk calculator, $x = v^2$ may be left in the machine and the division by $s$ performed after recording $x$. For the accuracy with which $v$ is determined by current experimental methods, recording of $s$ and $v$ to 3 significant figures (and likewise the columns $x$ and $y$) is quite adequate. However, the sums of squares and products should be recorded to the full accuracy of 6 or 7 significant figures, since 2 or 3 significant figures may be lost by subtraction when computing $\Delta$ and the numerators of $K_m$, $V$.

*Fine adjustment of the provisional estimates.* This process is based on fitting a bilinear regression of $v$ on the corresponding values of the provisionally fitted Michaelis–Menten function and its first derivative, as described above. The calculations are set out in Table 2, together with the supplementary calculations for determining standard errors, the numerical illustration continuing from Table 1.

The instructions in regard to accuracy are essentially the same as for Table 1. It is even more important here to record the intermediate calculations, and hence $b_1$ and $b_2$, with great accuracy, since the residual sum of squares is not calculated directly as a sum of squares of deviations, but by the indirect process of subtraction from $\Sigma v^2$, in which up to 3 or 4 significant figures may be lost.

Table 1. *Calculation of provisional estimates of $K_m$ and $V$*

For details, see text; illustrative values are from data of Atkinson *et al.* (1961).

| $s$ (pH 4·95)* | $v$† | $x = v^2$ | $y = v^2/s$ |
|---|---|---|---|
| 0·138 | 0·148 | 0·022 | 0·159 |
| 0·220 | 0·171 | 0·029 | 0·133 |
| 0·291 | 0·234 | 0·055 | 0·188 |
| 0·560 | 0·324 | 0·105 | 0·187 |
| 0·766 | 0·390 | 0·152 | 0·199 |
| 1·460 | 0·493 | 0·243 | 0·166 |

$\alpha = \Sigma vx = 0{\cdot}234184$ $\quad \Delta = \alpha\epsilon - \gamma\delta = 0{\cdot}0087695$

$\beta = \Sigma x^2 = 0{\cdot}097528$ $\quad K_m = (\beta\gamma - \alpha\delta)/\Delta = 0{\cdot}569$*

$\gamma = \Sigma vy = 0{\cdot}310303$ $\quad V = (\beta\epsilon - \delta^2)/\Delta = 0{\cdot}679$†

$\delta = \Sigma xy = 0{\cdot}107916$

$\epsilon = \Sigma y^2 = 0{\cdot}180440$

* Concn. of nicotinamide mononucleotide, mM.

† $\mu$moles of nicotinamide–adenine dinucleotide formed/3 min./mg. of enzyme protein.

*Estimation of dissociation constants*

The kind of statistical analysis required is illustrated here with the analyses applied to the data of Atkinson *et al.* (1961). These authors determined $K_m$ values in a series of experiments with substrate pH varying from 4·95 to 10·55, and in estimating dissociation constants for the enzyme and substrate from these data the appropriate function to be fitted, relating $K_m$ to hydrogen-ion concentration $H$, was

$$K_m = \widetilde{K}_m\left(1+\frac{H}{K_1}\right)\left(1+\frac{K_2}{H}\right),$$

$K_1$ and $K_2$ being the relevant dissociation constants (cf. Dixon & Webb, 1958). The basic data for the calculations described here are presented in Table 3, excluding an aberrant $K_m$ determination at pH 9·77 (see below).

Consideration of the experimental procedure (Atkinson *et al.* 1961) suggested that the $K_m$ determinations would be subject individually to biases fluctuating from experiment to experiment, or, differently expressed, that the inter-experimental variability of the $K_m$ values would be greater than indicated by their intra-experimental, or internal variance. This additional source of variation had to be taken into account, particularly in assessing the precision of the final estimates of $K_1$ and $K_2$. In the actual estimates of $K_1$ and $K_2$, however, it was expected that the net effect of bias would be very small, since the data themselves exhibited no evidence of a systematic deviation

from the theoretical form of the $K_m(H)$ function as given above, and that any systematic bias would be mostly absorbed in the estimate of the constant $\tilde{K}_m$.

It was also evident that the biases would be essentially of a multiplicative kind in their effect on the $K_m$ values, so that the variance of the inter-experimental fluctuations in bias would be essentially homogeneous on a logarithmic scale. It was thus appropriate to fit the $K_m$ function above in the (negative) logarithmic form $pK_m$.

Provisional estimates of $K_1$, $K_2$ and $\tilde{K}_m$ were derived by fitting a weighted bilinear regression of $K_m$ on $H$ and $1/H$. The appropriate weights

Table 2. *Fine adjustment of the provisional estimates of $K_m$ and $V$*

For details, see text; illustrative calculations continue from Table 1. Provisional estimates: $K_m^0 = 0{\cdot}569$,* $V^0 = 0{\cdot}679$† (Table 1).

| $s$* | $v$† | $s+K_m^0$ | $f = V^0 s/(s+K_m^0)$ | $f' = -V^0 s/(s+K_m^0)^2$ |
|---|---|---|---|---|
| 0·138 | 0·148 | 0·707 | 0·1325 | −0·1874 |
| 0·220 | 0·171 | 0·789 | 0·1893 | −0·2400 |
| 0·291 | 0·234 | 0·860 | 0·2298 | −0·2672 |
| 0·560 | 0·324 | 1·129 | 0·3368 | −0·2983 |
| 0·766 | 0·390 | 1·335 | 0·3896 | −0·2918 |
| 1·460 | 0·493 | 2·029 | 0·4886 | −0·2408 |

$\alpha = \Sigma f^2 = 0{\cdot}6101511$. $\gamma = \Sigma ff' = -0{\cdot}4634727$. $\delta = \Sigma vf = 0{\cdot}6077005$.
$\beta = \Sigma f'^2 = 0{\cdot}3962294$. $\epsilon = \Sigma vf' = -0{\cdot}4604656$.

$\Delta = \alpha\beta - \gamma^2 = 0{\cdot}02695286$.

$b_1 = (\beta\delta - \gamma\epsilon)/\Delta = 1{\cdot}0156833$. $V = b_1 V^0 = 0{\cdot}690$†

$b_2 = (\alpha\epsilon - \gamma\delta)/\Delta = 0{\cdot}0259341$. $K_m = K_m^0 + b_2/b_1 = 0{\cdot}595$*

$\Sigma v^2 = 0{\cdot}606026$. $S^2 = (\Sigma v^2 - b_1\delta - b_2\epsilon)/(n-2) = \dfrac{0{\cdot}0007365}{4}$,

$= 0{\cdot}0001841$ (4 D.F.).‡§

S.E. $(K_m) = \dfrac{S}{b_1}\sqrt{\dfrac{\alpha}{\Delta}} = 0{\cdot}064$*  S.E. $(V) = V^0 S\sqrt{\dfrac{\beta}{\Delta}} = 0{\cdot}036$.†

* Concn. of nicotinamide mononucleotide, mM.
† $\mu$moles of nicotinamide–adenine dinucleotide formed/3 min./mg. of enzyme protein.
‡ The symbol $S$ denotes a standard deviation.
§ Standard errors for $K_m$ and $V$ may alternatively be calculated by using a more accurate estimate of experimental variance derived from a series of experiments.

Table 3. *Data for the estimation of dissociation constants $K_1$ and $K_2$ derived from Atkinson, Jackson & Morton* (1961)

The provisional values $K_1^0$, $K_2^0$ and $\tilde{K}_m^0$ are given in the text.

| pH | $K_m$* | $K_m^{0*}$ | $p(K_m/K_m^0)$ $(y)$ | $\frac{H}{K_1^0+H}$ $(x_1)$ | $\frac{1/H}{1/K_2^0+1/H}$ $(x_2)$ | $w(K_m)$ | $w(pK_m)$ | $w'(pK_m)$ |
|---|---|---|---|---|---|---|---|---|
| 4·95 | 0·594 | 0·5864 | −0·0056 | 0·8651 | 0 | 0·045 | 436 | 305 |
| 5·15 | 0·450 | 0·3992 | −0·0520 | 0·8019 | 0 | 0·121 | 544 | 354 |
| 5·53 | 0·187 | 0·2125 | 0·0556 | 0·6278 | 0 | 0·639 | 809 | 450 |
| 5·85 | 0·139 | 0·1430 | 0·0123 | 0·4469 | 0·0001 | 2·275 | 1 309 | 571 |
| 6·21 | 0·105 | 0·1070 | 0·0082 | 0·2608 | 0·0002 | 4·329 | 1 394 | 587 |
| 6·66 | 0·108 | 0·0890 | −0·0840 | 0·1113 | 0·0005 | 4·272 | 952 | 491 |
| 7·16 | 0·0879 | 0·0823 | −0·0281 | 0·0381 | 0·0016 | 2·689 | 514 | 341 |
| 7·63 | 0·0877 | 0·0805 | −0·0372 | 0·0132 | 0·0046 | 2·899 | 528 | 347 |
| 7·95 | 0·0752 | 0·0804 | 0·0291 | 0·0064 | 0·0096 | 3·476 | 632 | 389 |
| 8·44 | 0·0791 | 0·0817 | 0·0135 | 0·0021 | 0·0292 | 3·411 | 639 | 392 |
| 8·76 | 0·0719 | 0·0842 | 0·0681 | 0·0010 | 0·0591 | 6·779 | 1 349 | 579 |
| 9·40 | 0·1129 | 0·1008 | −0·0492 | 0·0002 | 0·2152 | 1·717 | 491 | 331 |
| 9·55 | 0·106 | 0·1098 | 0·0153 | 0·0002 | 0·2794 | 1·552 | 526 | 346 |
| 10·00 | 0·180 | 0·1655 | −0·0365 | 0·0001 | 0·5220 | 0·436 | 336 | 252 |
| 10·35 | 0·298 | 0·2726 | −0·0387 | 0 | 0·7098 | 0·086 | 179 | 152 |
| 10·55 | 0·336 | 0·3858 | 0·0601 | 0 | 0·7950 | 0·052 | 217 | 179 |
| | | | $\bar{y} = 0{\cdot}0022$ | $\bar{x}_1 = 0{\cdot}2220$ | $\bar{x}_2 = 0{\cdot}0770$† | 34·778 | 10 855 | 6066 |
| | | | $\bar{y}' = -0{\cdot}0010$ | $\bar{x}_1' = 0{\cdot}2167$ | $\bar{x}_2' = 0{\cdot}0992$‡ | | | |

* Concn. of nicotinamide mononucleotide, mM.
† Weighted means based on the weights $w(pK_m)$.
‡ Weighted means based on the weights $w'(pK_m)$.

$w(K_m)$ for this analysis are shown in Table 3, these being the reciprocals of the internal variances of the $K_m$ values, excluding an estimated variance factor, 0·0001885. The estimates obtained were

$$K_1^0 = 174{\cdot}9 \times 10^{-8}, \quad K_2^0 = 0{\cdot}01092 \times 10^{-8},$$

$$\tilde{K}_m^0 = 0{\cdot}0791 \text{ mM}.$$

To determine an estimate of the inter-experimental variance component, the bilinear regression of the negative logarithms $\mathrm{p}(K_m/K_m^0)$ on

$$x_1 = H/(K_1^0+H) \text{ and } x_2 = (1/H)/(1/K_2^0+1/H)$$

was fitted with the weights $w(\mathrm{p}K_m)$ shown in Table 3, these being the reciprocals of the estimated internal variances of the $\mathrm{p}K_m$ values; $w(\mathrm{p}K_m) = (2{\cdot}303)^2 K_m^{0^2}/V(K_m)$. The provisionally fitted values $K_m^0$ are also shown in Table 3, these being derived from the provisional estimates above for $K_1$, $K_2$ and $K_m$. The analysis determined (as regression coefficients) the fine adjustments $\mathrm{p}(K_1/K_1^0) = 0{\cdot}002$, $\mathrm{p}(K_2/K_2^0) = -0{\cdot}001$, which are comparatively negligible, and yielded a residual mean square, $s^2 = 1{\cdot}6764$, based on 13 degrees of freedom.

The expected value of the mean-square $s^2$ is unity if the inter-experimental variance component is excluded. Although the observed value is significantly greater than unity at only the 10 % level of significance (cf. Fisher & Yates, 1938, Table V, entered with 13 and 66 degrees of freedom), the reality of the additional component of variance, in any case, was not in question. Assuming the inter-experimental variance of each $\mathrm{p}K_m$ to be the sum of the internal variance and the inter-experimental variance component $\sigma_b^2$,

$$V(\mathrm{p}K_m) = \frac{1}{w(\mathrm{p}K_m)} + \sigma_b{}^2,$$

it may be shown, by application of the basic rules, that the expected value of $s^2$ is

$$E(s^2) = 1+\frac{\sigma_b^2}{13}\left\{\Sigma w - \frac{\Sigma w^2}{\Sigma w} - c_{11}\Sigma w^2(x_1-\bar{x}_1)^2 - 2c_{12}\Sigma w^2(x_1-\bar{x}_1)(x_2-\bar{x}_2) - c_{22}\Sigma w^2(x_2-\bar{x}_2)^2\right\},$$

the coefficients $c_{ij}$ being those which occur in the regression calculations. Equation of the observed and expected values gave an estimate of the variance in bias, $s_b^2 = 0{\cdot}000986$, and addition of this value to the internal variances gave the estimated inter-experimental variances of the $\mathrm{p}K_m$ values, the reciprocals $w'(\mathrm{p}K_m)$ of which are shown in Table 3.

In the final analysis the bilinear regression of $\mathrm{p}(K_m/K_m^0)$ on $x_1$ and $x_2$ was refitted with the weights $w'(\mathrm{p}K_m)$, the provisional values $K_m^0$ being as before. The regression equations obtained were:

$$508{\cdot}9070b_1 - 130{\cdot}1409b_2 = 0{\cdot}214692,$$

$$-130{\cdot}1409b_1 + 243{\cdot}4536b_2 = 0{\cdot}671626,$$

$$(\Delta = 106958{\cdot}59),$$

which gave the fine adjustments

$$\mathrm{p}(K_1/K_1^0) = -b_1 = -0{\cdot}001,$$

$$\mathrm{p}(K_2/K_2^0) = \quad b_2 = 0{\cdot}003,$$

and hence

$$\mathrm{p}(\tilde{K}_m/\tilde{K}_m^0) = a = -0{\cdot}002.$$

Adjustment of the provisional estimates

$$\mathrm{p}K_1^0 = 5{\cdot}757, \quad \mathrm{p}K_2^0 = 9{\cdot}962, \quad \mathrm{p}\tilde{K}_m^0 = 4{\cdot}102,$$

gave the final estimates

$$\mathrm{p}K_1 = 5{\cdot}756, \quad \text{S.E.} = 0{\cdot}048;$$

$$\mathrm{p}K_2 = 9{\cdot}965, \quad \text{S.E.} = 0{\cdot}069;$$

$$\mathrm{p}\tilde{K}_m = 4{\cdot}100, \quad \text{S.E.} = 0{\cdot}019;$$

the standard errors being the square roots respectively of $c_{11} = 243{\cdot}5/\Delta$, $c_{22} = 508{\cdot}9/\Delta$, and $(1/\Sigma w' + c_{11}\bar{x}_1'^2 + 2c_{12}\bar{x}_1'\bar{x}_2' + c_{22}\bar{x}_2'^2)$.

The weighted means $\bar{x}_1'$, $\bar{x}_2'$ are shown in Table 3.

The final estimates obtained differ only very slightly from the provisional estimates; the major function of the subsequent analysis has been to determine the relevant estimates of standard error.

A determination of $K_m$ at pH 9·77 was omitted from the calculations described on account of gross aberration ($-0{\cdot}19$ on the $\mathrm{p}K_m$ scale) from the provisionally fitted $\mathrm{p}K_m$ function. This aberration was judged, on account of its magnitude, to be almost certainly due to an atypical experimental defect such as an accidental contamination of the substrate. Omission of the aberrant determination increased the estimate of $\mathrm{p}K_2$ by 0·05.

It might seem above, since the variations in the weights $w'$ are not very marked, that a simpler unweighted analysis would have sufficed. Almost the same estimates of $K_1$ and $K_2$ are obtained, but the estimates of standard error are affected:

$$\text{S.E.}\ (\mathrm{p}K_1) = 0{\cdot}044, \quad \text{S.E.}\ (\mathrm{p}K_2) = 0{\cdot}050.$$

The main difference is in the standard error of $\mathrm{p}K_2$, from which it is evident that the unweighted analysis has resulted in the amount of information regarding $\mathrm{p}K_2$ being overestimated by nearly 100 %.

## DISCUSSION

The utility of the statistical methods described is largely self-evident in the illustrations given. The basic principles can be applied, with analogous computations, to other problems in enzyme kinetics such as the determination of inhibitor constants (cf. Dixon & Webb, 1958), and more generally are

applicable whenever functions of known or assumed form are to be fitted to experimental series of values. The calculations required are not unduly time-consuming, usually involving only a fraction of the corresponding time spent in laboratory work and planning.

The statistical methods not only supply more accurate estimates and the necessary measures of precision but have a further advantage, in that subjective biases which might otherwise have arisen are thereby avoided. In fitting by eye, for instance, a series of straight lines to Lineweaver–Burk plots corresponding to a range of substrate pH values, there may be a subjective tendency to fit the lines either steeper or flatter than they should be, giving rise to distortions in the actual trend of $K_m$ with pH. Another kind of bias arises if there is a subconscious tendency to make the trend in slope of the series of lines rather more uniform than it should be. A plot of the $K_m$ determinations against pH would give, as a consequence of the subjective elimination of variability, a misleading visual impression of the accuracy of the experimental work, and certainly significance tests based on such determinations would be invalid.

It should be emphasized that statistical measures of precision supply a gauge for random variation only. An experiment may supply, in this sense, a precise determination, which nevertheless is seriously biased by some defect or limitation in the experimental procedure.

## SUMMARY

1. An account is given of the weighted and non-linear regression methods relevant to enzyme kinetic studies, with a brief preliminary outline of statistical terminology and the basic calculus of random variation.

2. Statistical considerations indicate that, for graphical determinations of the parameters $K_m$ and $V$ in the Michaelis–Menten equation, the linear plot of $s/v$ against $s$ is preferable to the double reciprocal plot.

3. A computational procedure is given for estimating $K_m$, $V$ and the relevant standard errors.

4. The application of regression methods is further illustrated with the estimation of dissociation constants from a series of $K_m$ determinations.

The author is grateful to Professor R. K. Morton, and to Dr M. R. Atkinson and Mr J. F. Jackson, for the collaboration which has led to the writing of this paper, and for permission to use their data in illustration.

## REFERENCES

Atkinson, M. R., Jackson, J. F. & Morton, R. K. (1961). *Biochem. J.* **80**, 318.

Dixon, M. (1953). *Biochem. J.* **55**, 161.

Dixon, M. & Webb, E. C. (1958). *Enzymes*. London: Longmans, Green and Co. Ltd.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, 13th ed., 1958. Edinburgh: Oliver and Boyd Ltd.

Fisher, R. A. & Yates, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*, 5th ed., 1957. Edinburgh: Oliver and Boyd Ltd.

Hofstee, B. H. J. (1952). *Science*, **116**, 329.

Lineweaver, H. & Burk, D. (1934). *J. Amer. chem. Soc.* **56**, 658.

Williams, E. J. (1959). *Regression Analysis*. New York: John Wiley and Sons Inc.

---

*Biochem. J.* (1961) **80**, 332

# Free Amino Acids of the Haemolymph of *Schistocerca gregaria* Forsk.

By C. A. BENASSI, G. COLOMBO and G. ALLEGRI
*Istituto di Chimica Farmaceutica e di Zoologia, Università di Padova, Italy*

(*Received* 5 *December* 1960)

Insect haemolymph contains free amino acids in higher concentrations than the blood of other animal groups; detailed data have been collected from different species of various orders of insects only in the last ten years, mainly after the development of microbiological and chromatographic methods of analysis.

Some investigators have recently studied the variations of amino-acidaemia during developmental stages of holometabolous insects, in relation to the metabolic processes involved (cf. Florkin, 1959).

As part of a research on amino acid metabolism in Orthoptera, we report here the quantitative pattern of the free amino acids in the blood of a locust, *Schistocerca gregaria* Forsk., at various stages of development. A similar investigation on a grasshopper, *Anacridium aegyptium* L., has already been published (Benassi, Colombo & Peretti, 1959).