QUADRATIC EQUATIONS AS AN INTERPRETATIVE TOOL IN BIOLOGICAL RESEARCH

J. Putter, D. Yaron, and H. Bielorai*

THE usefulness of an empirically fitted biological response curve is often severely limited by the fact that the parametric form of the curve is not based on a quantitative theory of the mechanism underlying the response. In the absence of such a theory, an arbitrary parametric family of curves, i.e., an equation of given form containing one or several unknown parameters (coefficients) is usually chosen, mainly on grounds of convenience of computation, and the parameter values corresponding to that member of the family which comes “closest” to the experimental data are calculated. This procedure, which amounts to little more than a cosmetic smoothing operation, may provide a practically useful approximate guide to future performance if the future conditions are sufficiently similar to those under which the original data were obtained. However, in most cases, it precludes the possibility of either extrapolating beyond the limited range covered by the data or gaining an insight into the underlying mechanism of the phenomenon involved. In particular, the numerical values of the coefficients do not always correspond to biologically meaningful entities, and, consequently, when several such empirical curves are compared the differences between them are difficult to interpret.

In the absence of a quantitative underlying theory the parameters of the fitted equation are unlikely to have a precise and fully valid biological meaning, but they may nevertheless be interpretable to some extent; e.g., in the two-parameter family corresponding to the general straight-line equation, both parameters (the slope and the intercept) can be easily interpreted. Consequently, when several straight lines are compared, it is of interest to ask whether (and how much) they differ in slope, intercept, or both, and it is reasonable to expect that the answer to this question will provide some basic insight into the nature of the biological phenomenon involved. Such comparisons are likely to be more valid than the estimates of single parameter values. This can be seen by considering the case where the straight line is only an approximation to a (true) curvilinear relationship. In such a case, if no data for low values of the independent variable are available, a single intercept estimate may be so unrealistic as to be useless. However, a comparison between two intercept estimates may lead to a valid conclusion concerning the relative location of the two curves involved.

The main purpose of this note is to point out that the practice of empirical curve fitting is likely to be more fruitful if the form of the fitted equation is such that the parameters correspond, at least approximately, to biological factors. Conditions, a comparison between these values will provide some basic information. Since quadratic equations usually themselves to at least two types of meaningful conditions involved.

Interpretation of Quadratic Regression

The general quadratic regression can be written in the form

\[ y = b_0 + b_1x + b_2x^2, \]

and also in the alternative form

\[ y = A - B(x - C)^2. \]

The parameters of [2] are related to those of [1] by the identities

\[ A = b_0 - b_2/4b_1, \quad B = -b_2, \quad C = -b_1/2b_2. \]


The more commonly used form [1] is when we are interested mainly in the behavior of y in the neighborhood of \( x = 0 \); its three parameters correspond, respectively, to the initial value of y, the initial slope, and to the change in slope up to the neighborhood of \( x = 0 \); in this case, the parameters of [1] have little immediate meaning, while those of [2] are more relevant. Taking the case \( B > 0 \), the parameter \( C \) corresponds to the location (abscissa) of the vertex of the parabola, A to its magnitude (ordinate), and \( B \) can be taken as a measure of how fast y decreases when \( x \) changes from \( C \).

The considerations concerning the valid use of [2] are analogous to those explained above for linear regression. Thus, when no exact underlying theory is available, and the quadratic equation is a convenient empirical approximation, the parameters A, B, and C for a single curve cannot be regarded as valid estimates of the real magnitude and location of a maximum. Therefore, these estimates are useful; first, as the formula obtained by plugging them into [2] serves to guide future decisions, a comparison between these values is usually made only when we are interested in a comparison of conditions.

Consider, for example, the regression of the yield of a crop on the amount of irrigation water applied. If no quantitative theory of evapotranspiration is available, the quadratic equation is likely to be the most convenient empirical approximation, and the parameters of the equation may provide useful information. However, it is important to note that these estimates are valid only as long as the conditions under which the original data were obtained are sufficiently similar to those under which the future conditions are expected to prevail.