Jl. Rubb. Res. Inst. Sri Lanka (1986) **65**, 14-21

ON THE CURVILINEAR CALIBRATION IN BIOLOGICAL AND CHEMICAL EXPERIMENTS

by

W. N. WICKREMASINGHE

ABSTRACT

A common application in biological and chemical experiments is to predict X from Y (calibration) when Y is related to X, through regression techniques, where X and Y are 2 variables of interest. This paper discusses the case when Y is a quadratic function of X and the quadratic response is highly significant. This is an extension to linear calibration and approximate confidence intervals are derived for X when Y is a maximum or a minimum depending on the function. An illustration is given using a numerical example. Alternatives and further extensions are also mentioned.

INTRODUCTION

The most common application of regression techniques is to find the relationship between two variables X and Y, where Y is supposed to be dependant on X. This relationship is then used to predict Y for a given value of X within the observed range of X. However in some applications of biology and chemistry, it is sometimes necessary to use a simple regression line or a curve constructed by measuring a variable Y at a given set of values from another variable X, to predict X from Y. This procedure is adopted in problems where X is difficult to measure or expensive but is related to Y which is more easily measured. Some examples of this nature are (i) concentration (X) and absorbance (Y) of a chemical solution. (ii) to find the confidence interval for nitrogen level (X) which gives the highest nodulation dry weight (Y) in a nitrogen fixation experiment etc.. The theory and application to a numerical example, for the linear calibration, *i.e.* prediction of X from Y when they are linearly related, is discussed by Snedecor and Cochran (1980). Moore (1980), discusses this further with an application to the analysis of chemical data.

In practice, most of the biologists and chemists do not wish to consider complex situations and linear is the most common application among them. However, in some relationships quadratic functions become evident and therefore it is important to look at this situation. It is also worth noting that in certain situations simple curvilinear functions other than the quadratic one have also proved to be better. For example Abeywardena (1964) showed that the function $Y=a+bX+c\sqrt{X}$ gives a better fit than the usual quadratic one to explain the variability of coconut yield per plot (Y) using number of bearing palms per plot (X). Duke *et al* (1982) discuss the use of Fourier series as an alternative for quadratic response curve. The objective of this paper is to focus on the problem of quadratic calibration, *i.e.* to predict X from Y when Y is a quadratic function of X, with the aim of finding approximate confidence intervals for X using standard statistical tools. Application is illustrated by a numerical example. It is anticipated that the results could be easily extended to other simple curvilinear relationships such as the one discussed by Abeywardena (1964). Other possible extensions are also mentioned.

Quadratic response curve

A quadratic response curve is of the form $Y = \beta_0 + \beta_1 X + \beta_2 X^2$, where β_0, β_1, & β_2 are coefficients estimated using experimental data and Y is supposed to be dependent on X. Now, $\frac{dY}{dX}$ (rate of change of Y) $= \beta_1 + 2\beta_2 X$.

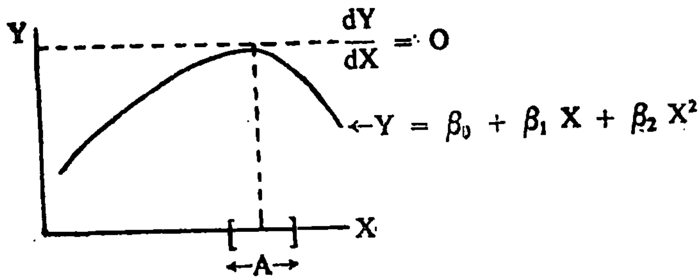
The problem

It is a common question to ask the value of X which gives the maximum or the minimum Y, depending on the relationship. The answer is simple, because once the coefficents β_0, β_1 & β_2 are estimated, it is simply the value of X when $\frac{dY}{dX} = O$, *i.e.* $\beta_1 + 2\beta_2 X = O \Longrightarrow \hat{X} = -\beta_1/2\beta_2$ is the value of X that gives the highest or the lowest Y in a strong quadratic relationship. However, a more valid question is to ask for a confidence interval (say 95%) for X which gives the maximum or minimum Y, as the same X would lead to slightly different Y values, in practice.

95% confidence interval for X when $\frac{dY}{dX} = O$

As discussed above, this is the confidence interval for X when Y is a maximum or a minimum. Fig. 1. shows a quadratic relationship between X and Y and the required confidence interval, A.

Fig 1.



It can be shown that,

$$X = \frac{\left(\hat{X} - \frac{\sigma_{1.2}K^2}{2}\right) \pm K\sigma_1\sigma_2\sqrt{\frac{\hat{X}^2}{\sigma_2^2} + \frac{\sigma_{1.2}\hat{X}}{\sigma_1^2\sigma_2^2} - \frac{K^2(1-r_{1.2}^2)}{4} + \frac{1}{4\sigma_1^2}}}{(1 - K^2\sigma_1^2)}$$

gives the 95% confidence interval for X when $\frac{dY}{dX} = O$ (see appendix for the derivation)

Where $K^2 = \frac{n(n-1)\, t^2 s^2}{|X^1X|\, \beta_2^2}$ and

t = tabulated t value at 5% level with (n - 3) degrees of freedom, and s^2, β_2^2 and X^1X are defined in the appendix.

Also,
σ_1^2 = Variance of X
σ_2^2 = Variance of X^2
$\sigma_{1.2}$ = Covariance (X, X^2), and
$r_{1.2}$ = Correlation between X & X^2.

Now, with the assumption that we talk about a quadratic response curve only if the quadratic term in the ANOVA is highly significant (*i.e.* β_2 is highly significantly different from O),

We can further derive the following:

$$Var(\beta_2) = \frac{K^2\beta_2^2}{t^2}\sigma_1^2$$

and F value for testing β_2 (in the ANOVA)

$$= \frac{\beta_2^2}{V(\beta_2)} = \frac{F(\text{from tables})}{K^2\ \sigma_1^2}$$

Therefore, $F = \dfrac{F(\text{from tables})}{k^2\ \sigma_1^2}$

Now, β_2 is highly significant $\Longrightarrow$ $F \gg F$ (Tables)

$\Longrightarrow K^2 \sigma_1^2 < 1 \Longrightarrow (1-K^2\sigma_1^2) > 0$

Therefore, finite confidence intervals exist, when β_2 is highly significant.

Further, if β_2 is highly significant,

$$K^2 \sigma_1^2 \rightarrow 0 \text{ (negligible)}$$

and also $(1-r_{1.2}^2)$ is negligible (since $r_{1.2}$ is closer to 1).

Approximate confidence interval for X

When there is a highly significant quadratic response, our confidence interval for X that gives the maximum or minimum Y can be approximated by,

$$\hat{X} \pm K\,\sigma_1\,\sigma_2 \sqrt{\frac{\hat{X}^2}{\sigma_2^2} + \frac{\sigma_{1.2}}{\sigma_1^2\sigma_2^2}\hat{X} + \frac{1}{4\,\sigma_1^2}}$$

This is of the form $\hat{X} \pm t\,s_{\hat{x}}$

$$\text{Where} \quad s_{\hat{X}} = \sigma_2\sqrt{\frac{V(\beta_2)}{\beta_2}}\sqrt{\frac{\hat{X}^2}{\sigma_2^2} + \frac{\sigma_{1.2}}{\sigma_1^2\sigma_2^2}\hat{X} + \frac{1}{4\sigma_1^2}}$$

and holds true for $n \geqslant 4$.

The above $s_{\hat{X}}$ stands as an approximate standard error for X (given $\dfrac{dY}{dX} = 0$) when n is large.

Numerical example

The following example was taken from pages 480 – 482 of Box, Hunter & Hunter, (1978) to illustrate the application of the quadratic calibration.

Treatment (X) (Amount of supplement)	10	10	15	20	20	25	25	25	30	35
Growth Rate (Y)	78	78	85	90	91	87	86	91	75	65

Table 1. *ANOVA of Y for the above data set*

Source	d.f	S.S.	M.S.	F
Between Level (X)	5	683.90		
X–linear	1	24.5	24.5	4.13(n.s)
X–quadratic	1	644.9	644.9	108.75***
X–lack of fit	3	14.5	4.83	
Residual				5.93
Error	4	27.0	6.75	
Total	09			

In this example there was a highly significant quadratic response and the estimated quadratic response of average Y to levels of X, was found to be $Y = 35.66 + 5.26\,X - 0.128\,X^2$ (using multiple regression techniques)

Therefore, maximum Y is at $\hat{X} = \dfrac{5.26}{2 \mathrm{x} 0.128} = 20.5$

However 20.5 is just a point estimate,

$(\sigma_1^2 = 66.9444,\ \sigma_2^2 = 356.3803,\ \sigma_{1.2} = 2865.2777;\ r_{12} = .98\ k = 0.027726)$

Therefore the approximate 95% confidence interval for the amount of supplement (X) that shows the maximum growth rate (Y) is given by

$$X = 20.5 \pm 0.22685\sqrt{3.644978}$$

$$= 20.5 \pm 0.4 \qquad = (20.1,\ 20.9)$$

Therefore, with 95% confidence we say that the amount of the supplement that gives the maximum growth rate is between 20.1 & 20.9.

Limitations and further work

It should be noted that the above discussed technique for estimating approximate confidence intervals are valid only when the variable Y of interest is a quadratic in X, the variable Y is dependent on, and this quadratic response is highly significant. Also the number of pairs (X,Y) should be greater than 4. The application of this is open to any similar situation irrespective of the field and the results could be easily extended to other curvilinear relationships such as the one mentioned by Abeywardena (1964). However, further generalisations of this also to cover other curvilinear and non-linear relationships and when X and Y are measured with error, are worth looking at.

ACKNOWLEDGEMENTS

The author wishes to thank Professor S. C. Pearce, University of Kent, for his useful comments on the paper.

REFERENCES

ABEYWARDENA, V., (1964) Statistical control of variability in coconut experiments. *Empire Journal Exp. Agric.*, 32, 166 - 174.

BOX, G. E. P., HUNTER, W. G. and HUNTER, .J. S. (1978). *Statistics for Experimenters*, John Wiley & Sons.

DYKE, G. V., SMITH, G. L. and YEOMAN, D. P., (1982), Fourier series and response curves. *Journal Agric. Sci.*, 98, 119 - 122.

JOHNSTON, J. (1972). Econometric methods 2nd Edition, McGraw - Hill.

MOORE, B. A. (1980). Correlation and regression analysis : Applicaitons to the analysis of chemical data, *Anal. Proc.*, 124 - 127.

SNEDECOR, G. W. and COCHRAN, W. G. (1980). *Statistical Methods* 7th Edition, the Iowa State University Press.

APPENDIX

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \equiv \underline{Y} = X\underline{\beta}$$

where $\underline{\beta}' = (\beta_0\ \beta_1\ \beta_2)$ & $X'X = \begin{bmatrix} n & \Sigma X & \Sigma X^2 \\ & \Sigma X^2 & \Sigma X^3 \\ \text{symmetric} & & \Sigma X^4 \end{bmatrix}_{3\times 3}$

$$\underline{\beta} = (X'X)^{-1} X'\underline{Y} \qquad \text{(using normal equations)}$$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2$$

and $\frac{dY}{dX} = \beta_1 + 2\beta_2 X$

Find 95% Confidence Interval (C.I.) for X when $\frac{dY}{dX} = 0$

i.e. Find 95% C.I. for X that gives the maximum or the minimum Y.

Let $C'\underset{\sim}{\beta} = \frac{dY}{dX} = \beta_1 + 2\beta_2 X$ Where $C' = (0\ 1\ 2X)$

Now for given X, 95% C.I. for $\frac{dY}{dX}$ is given by

$$\frac{dY}{dX} = (\beta_1 + 2\beta_2 X) \pm t_{.05,\,n-3} \sqrt{C'(X'X)^{-1}C\ s^2} \qquad \text{——— (1)}$$

Where s^2 = Error Mean Square with (n–3) d.f. from ANOVA ((Johnston, (1972).

Let $X = \hat{X}$ when $\frac{dY}{dX} = 0$

Then from (1) $(\hat{X}-X)^2 = \frac{t^2 s^2}{4\beta_2^2}(C'(X'X)^{-1}C)$

$$\therefore\ \hat{X}^2 - 2X\hat{X} + X^2 = \frac{t^2 s^2}{4\beta_2^2}(4fX^2 + 4eX + d)$$

where $C'(X^1X)^{-1}C = 4fX^2 + 4eX + d$

and $f = \frac{n(n-1)\,\sigma_X^2}{|X^1X|}$, $e = \frac{-n(n-1)\,\sigma X.X^2}{|X^1X|}$

and $d = \frac{n(n-1)}{|X^1X|}\ \sigma_{X^2}^2$

By simplification, we get

$$(1-K^2\sigma_1^2)\,X^2 + (-2\hat{X} + \sigma_{1.2}K^2)\,X + (\hat{X}^2 - \frac{K^2\sigma_2^2)}{4} = 0$$

where $K^2 = \dfrac{n(n-1)\,t^2s^2}{|X'X|\,\beta_2^2}$ and $\sigma_1^2 = \sigma_X^2 = V(X)$

$\sigma_2^2 = \sigma_{X^2}^2 = V(X^2)$

$\sigma_{1.2} = \sigma_{X.X^2} = \mathrm{Cov}\,(X,X^2)$

$$\therefore X = \frac{(2\hat{X} - \sigma_{1.2}K^2) \pm \sqrt{(-2\hat{X}+\sigma_{1.2}K^2)^2 - 4\,(1-K^2\sigma_1^2)\,(\hat{X}^2 - \frac{K^2\sigma_2^2)}{4}}}{2(1 - K^2\sigma_1^2)}$$

$$= \frac{(\hat{X} - \sigma_{1.2}K^2) \pm K\sigma_1\sigma_2\sqrt{\frac{\hat{X}^2}{\sigma_2^2} + \frac{\sigma_{1.2}\hat{X}}{\sigma_1^2\sigma_2^2} - \frac{K^2}{4}(1 - r_{1.2}^2) + \frac{1}{4\sigma_1^2}}}{(1 - K^2\sigma_1^2)}$$

where $r_{1.2} = \mathrm{Corr}\,(X,X^2)$.