*Jour. Ind. Soc. Ag. Statistics*
*Vol. LX, No. 3 (1988), pp. 212-217*

# ON THE DETERMINATION OF PLOT SIZE

LALIT KISHORE
*Punjab Agricultural University, Ludhiana*
(Received : February, 1988)

SUMMARY

Fairfield Smith's variance law provides a relationship of the coefficient of variation ($C_x$) of the variable under study among plots of size $x$. From the empirical law the optimum plot size is determined using the 'maximum curvature technique'. To establish relationship between $C_x$ and $x$, unweighted least squares technique is commonly employed to estimate the regression coefficients. In the present paper, a justification is given for an alternative assumption namely that the error variance varies proportional to $x^g$. Likelihood is maximised for $g$ using the Fibonacci search technique of optimization, Estimates of regression coefficient and their variances are found to be same as through weighted least square technique. This assumption, is supported by results obtained for four sets of uniformity trial data on vegetable crops and compared with those using the least squares method suggested by Fairfield Smith.

*Keywords* : Maximum likelihood method, Fabonacci Search, Heteroscedasticity.

## Introduction

A method for determining the optimum size of an experimental unit or plot which is widely used is 'the maximum curvature method'. It uses the coefficients of variation (C.V.) of the variable under study for various plot sizes from a uniformity trial data. A free-hand graph is drawn between the coefficients of variation and the plot sizes and the optimum plot size determined from the graph as one just beyond the point of maximum curvature. In place of a free-hand graph, the use of Fairfield Smith's Variance Law [8] is more reliable. It is expressed as :

$$\log V_x = \log V_1 + b \log x, \tag{1}$$

where $V_x$ is the variance of the character under study per unit area among plots of size $x$ units. $V_1$ and $b$ are constants. Using a simple linear transformation on the log variable, (1) can be alternatively expressed in terms of the coefficient of variation as :

$$\log C_x = \alpha + \beta \log x, \tag{2}$$

where $C_x$ is the coefficient of variation among the plots of size $x$ units and $\alpha$ and $\beta$ are constants.

## 2. Estimation of the Constants $\alpha$ and $\beta$

Given $C_{xi}$ and $x_i$ for the $i$th plot ($i = 1, 2, \ldots, n$), the estimation of $\alpha$ and $\beta$ generally proceeds by fitting the following linear regression of log variables using the least squares technique :

$$y_i = \alpha + \beta z_i + u_i, \tag{3}$$

where $\quad y_i = \log C_{xi}$ and $z_i = \log x_i$.

The error term $u_i$ in (3) is commonly assumed to be distributed normally with mean zero and constant variance $\sigma_u^2$.

However, if $m$ observations of smaller plots of size unity are taken for uniformity trial data, then a plot size $x$ will need a combination of $x$ smaller units. Thus there will be $m/x$ plots of size $x$ and the C.V. will be calculated on the basis of these $m/x$ plots. Thus the precision of C.V. will depend upon size $x$, since $m$ is constant. To assume therefore that the error $u$ is independent of plot size $x$ and homoscedastic as is so often done in the usual regression analysis of (³) does not seem to be justified. It is also well known that in the above situation, the least squares estimates of $\alpha$ and $\beta$ of the regression (3) are inefficient (both for small as well as large samples) thereby providing misleading results for the confidence intervals and critical regions and ultimately leading to incorrect inferences about the population parameters. For these reasons, the following model in which the error variance $\sigma_u^2$ is assumed to be proportional to $x^g$ where $g$ is a constant, has been examined in detail. Under this assumption, the regression equation (3) takes the form

$$y_i = \alpha + \beta z_i + w_i^{-1/2}\, \varepsilon_i \tag{4}$$

where $\quad u_i = w_i^{-1/2} \cdot \varepsilon_i = x_i^{g/2}\, \varepsilon_i,\ \ \sigma_{u_i}^2 = w_i^{-1}\, \sigma_\varepsilon^2$ and $w_i = x_i^{-g}$.

This involves two new parameters $g$ and $\sigma_\varepsilon^2$, the first of which is the measure of heteroscedasticity and the second, the variance of the error $\varepsilon_i$ assumed now to be constant for all $i$ and independent of $x_i$.

Jacquez, *et al.* [4] have extended the theory of simple linear regression to the case of non-uniform error variances for the situation in which replicates are available at each sample point in the domain of the independent variable. They have shown that the methods of weighted least squares and the maximum likelihood (ML) perform equally better as compared to that of least squares. Kmenta [5] has dealt with the case of a model similar to [4] and has suggested the use of maximum likelihood estimation under such a set-up. Following Kmenta, it can de easily shown that the following are the maximum likelihood estimates of $\alpha$, $\beta$ and $\sigma_\varepsilon^2$ for a known $g$ :

$$\hat{\alpha} = \bar{y}_w - \beta \bar{z}_w \tag{5}$$

$$\hat{\beta} = \frac{\Sigma w_i (y_i - \bar{y}_w)(z_i - \bar{z}_w)}{\Sigma w_i (z_i - \bar{z}_w)^2} \tag{6}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \Sigma w_i [y_i - y_w - \beta(z_i - \bar{z})]^2 \tag{7}$$

where $\bar{y}_w = \frac{1}{\Sigma w_i} \Sigma w_i y_i$ and $\bar{z}_w = \frac{1}{\Sigma w_i} \Sigma w_i z_i$.

Hence the estimates of $\alpha$, $\beta$ and $\alpha_\varepsilon^2$ are same as obtained by least square method. Also, the value $L$ of the maximum likelihood is given by :

$$L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma_\varepsilon^2 - \frac{g}{2} \Sigma z_i - \frac{1}{2\sigma_\varepsilon^2} \Sigma x_i^{-g} (y_i - \alpha - \beta x_i)^2. \tag{8}$$

For estimation of $g$, $L$ to be maximised can be taken as

$$L = \text{constant} - \frac{g}{2} \Sigma z_i - \frac{1}{2\sigma_\varepsilon^2} \Sigma x_i^{-g} (y_i - \alpha - \beta x_i)^2.$$

Now owing to extreme non-linearities of the equations (5)-(8) it is not possible to use them for the direct estimation of the parameters in question. We have therefore used well known iterative method namely the Fibonacci search techniques to choose that value of $g$ for which the likelihood $L$ of (8) is maximum and to obtain therefore the corresponding estimates of the other parameters. The application of this method depends on two factors namely that : (i) an initial interval is known which contains the maximum; and (ii) the function is unimodal within the interval. Both these factors have been verified for all the four sets of data considered in this paper by observing the nature of $L$ from equations (5) to (8) for a wide range of $g = 0\ (0.1)\ 3.0$. The value of $g$ finally iterated by this method which maximises the likelihood to the desired accuracy is used in (5) to (7) to obtain the corresponding maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2_\varepsilon$. Following Kmenta [5] the variance of $\hat{\alpha}$ and $\hat{\beta}$ are given by (9) and (10) respectively :

$$\text{Var}\,(\hat{\alpha}) = \sigma^2_\varepsilon \left[ \frac{\Sigma w_i}{n^2} + \frac{\bar{Z}^2_w}{\Sigma w_i (z_i - \bar{z}_w)^2} \right], \tag{9}$$

$$\text{Var}\,(\hat{\beta}) = \sigma^2_\varepsilon \left[ \frac{1}{\Sigma w_i (z_i - \bar{z}_w)^2} \right]. \tag{10}$$

Considering correlation between $\hat{g}$ and $\hat{\sigma}^2_\varepsilon$ asymptotic variance of $\hat{g}$ is given by the expression.

$$\text{Var}\,(\hat{g}) = \frac{2}{\Sigma (z_i - \bar{z})^2}. \tag{11}$$

The coefficient of multiple determination $R^2$ is calculated for both the models (3) and (4) from the expression

$$R^2 = 1 - \frac{\Sigma (y_i - \hat{\alpha} - \hat{\beta} z_i)^2}{\Sigma y_i^2 - \frac{1}{n} (\Sigma y_i)^2}. \tag{12}$$

Finally, using the estimated values of the parameters, the required optimum plot size is determined by the maximum curvature technique.

## 3. Data and Results

Using the above procedure, the fitting of model (4) has been done on four sets of uniformity trial data. The first two sets are related to Bhindi (*Abelmoschus Esculentus L,*) crop for the year 1960-61 and 1981-82 taken from the ICAR monograph [7]. The third set is on Onion (*Allium Cepa* L.) crop used by Gupta and Raghavarao [2] and the last one on Radish (*Rachamus Sativus* L.) crop reported by Kumar [6]. For Bhindi, 52 pairs of C. Vs and plot sizes are available for each of the two years. For onion and radish crops, use has been made of 97 pairs of C. Vs and plot sizes and 47 pairs of C. Vs and plot sizes respectively.

Table 1 summarises the results of the analysis for the above data-sets. It may be noticed that equation (4) with $g = 0$ is equivalent to (3) unweighted and the same with $g = 1$ is equivalent to (3) weighted to the first order of approximation [1] and that the estimators of the former are same as the L.S. estimators of the latter. For each set of data in the table, the first row corresponds to $g = 0$, the second to the estimated

TABLE 1—SUMMARY OF ANALYSIS RESULTS

| *Crop* | *m* | *g* | $\alpha$ | $\beta$ | $\sigma_{\varepsilon_i}$ | $R^2$ | *L* | *Optimum plot size* |
|---|---|---|---|---|---|---|---|---|
| Bhindi 1960-61 | | 0 | 1.3230 (0.0205) | −0.1179 (0.0231) | 0.0577 | 0.3426 | 75.54 | 5.06 |
| | 1456 | 1.0115* (0.3625) | 1.3666 (0.0150) | −0.1811 (0.0230) | 0.0201 | 0.2305 | 79.70 | 7.57 |
| | | 1 | 1.3662 (0.0150) | −0.1803 (0.0220) | 0.0204 | 0.2335 | 79.67 | 7.54 |
| Bhindi 1961-62 | | 0 | 1.5426 (0.0151) | −0.1378 (0.0170) | 0.0425 | 0.5678 | 91.46 | 8.92 |
| | 1456 | 0.3721@ (0.3625) | 1.5530 (0.0134) | −0.1513 (0.0170) | 0.0288 | 0.5653 | 92.48 | 9.72 |
| | | 1 | 1.5698 (0.0123) | −0.1772 (0.0188) | 0.0167 | 0.5149 | 90.10 | 11.11 |
| Radish (plot size | | 0 | 2.0378 (0.0528) | −0.2828 (0.0172) | 0.0690 | 0.8570 | 59.97 | 32.24 |
| multiplied by 100) | 720 | 0.8669* (0.1926) | 2.0923 (0.0260) | −0.3040 (0.0106) | 0.0024 | 0.8495 | 74.96 | 35.85 |
| | | 1 | 2.1064 (0.0239) | −0.3099 (0.0100) | 0.0015 | 0.8438 | 74.60 | 36.81 |
| Onion | | 0 | 2.1539 (0.0674) | −0.2963 (0.0208) | 0.0751 | 0.6817 | 114.56 | 39.89 |
| | 576 | 0.6743† (0.1261) | 2.1472 (0.0440) | −0.2941 (0.0147) | 0.0058 | 0.6817 | 118.83 | 39.40 |
| | | 1 | 2.1403 (0.0357) | −0.2917 (0.0126) | 0.0018 | 0.6815 | 117.70 | 38.88 |

*and† indicate statistical significance from zero and one respectively both at 5 percent probabilities level and @ indicates statistical significance from one at 10 percent probability level.

$g$ which yields the maximum value of $L$ and the third to $g = 1$. The standard errors of the corresponding estimates are given in brackets.

We find from the table that the $R^2$ value decreases (slightly) in case of model (4). The most striking observation is that in comparision to (3) unweighted (i.e. with $g = 0$), the maximum likelihood $L$ of (4) clearly increased in all the four cases and that too very strikingly in certain instances as in the case of radish crop although this increase is not so very large when compared with weighted (i.e. $g = 1$). Further, $g$ is found to be significantly different from zero at 5 percent level of significance for Bhindi 1960-61, Radish and Onion thus showing that the error component is significantly dependent on the plot size in all these cases. It is worth noting in case of onion and Bhindi 1961-62 that their estimated values of $g$ are significantly different from unity at 5 percent and 10 percent probability level respectively.

It thus appears that model (4) is a plausible specification for the data taken into consideration and that the assumption of independence between the error and the plot size is untenable on both logical and empirical grounds. While Fairfield Smith has suggested a method which imposes artificially a value as $g = 1$ to take care of this heteroscedasticity, the superiority of the present method lies in the fact that the data themselves are allowed to determine the value of $g$ and in yielding the highly desirable maximum likelihood estimate. As will be seen from the table, there are cases when $g$'s are significantly different from unity and this is so proposed for large and small $m$ as well as $n$.

## REFERENCES


[1] Federer, W.T. (1967) : *Experimental Design*, Oxford and IBH Publishing Co.

[2] Gupta, J. P. and Raghavarao, D. (1971) : Optimum size of plots for experiments on the weights of onion bulbs, *Ind. Journal of Horticulture* **28** (3) : 234-236.

[3] Husain, A. and Gangiah, K. (1976) : *Optimization Techniques for Chemical Engineers*, The Macmillan Company of India Ltd.

[4] Jacquez, John A., Mather Frances, J. and Crawford Carles, R. (1968) : Linear regression with non-constant, unknown error variances : Sampling experiments with least squares, weighted least squares and maximum likelihood estimators, *Biometrics*, December, **3** : 607.

[5] Kmenta, J. (1971) : *Elements of Econometrics*, The Macmillan Company.

[6] Kumar, D. (1970) : Determination optimum plot size and shape on Bhindi (*Abelmoschus Esculentus* L.) and Radish (*Rachamus Sativus* L.), *M.Sc. thesis* submitted to the PAU, Ludhiana.

[7] Singh, D., Bhargava, P. N. and Khosla, R. K. (1975) : *Monograph on Study of Size and Shape of Plots for Field Experiments on Vegetable and Perennial Crops*, Institute of Agricultural Research Statistics. New Delhi-12, pp. 52-54.

[8] Smith, H. F. (1938) : An empirical law describing heterogeneity in the yields of agricultural crops, *J. Agr. Sci.* **28** : 1-23, 63-66, 68.