STATISTICS IN TRANSITION-new series, Spring 2013
Vol. 14, No. 1, pp. 161–170

A LINEAR MODEL FOR UNIFORMITY TRIAL EXPERIMENTS

Alok K. Shukla[1], **Subhash K. Yadav**[2], **Govind Ch. Misra**[3]

ABSTRACT

Uniformity trial experiments are required to assess fertility variation in agricultural land. Several models have appeared in literature, of which Fairfield Smith's Variance Law assuming a nonlinear relationship between the coefficient of variation (C.V.) and a plot size has been extensively used in uniformity trial studies. A linear model has been proposed for uniformity trial experiments and it has shown better results as compared to existing models. The expression for point of maximum curvature for the proposed model is much simpler as compared to the model of Fairfield Smith. The appropriateness of the proposed model has also been verified with the help of a data set.

Key words: Fairfield Smith's Variance Law, linear model, uniformity trial experiments.

1. Introduction

Uniformity trials are needed to determine suitable shape and size of the plot for knowing the nature and extent of fertility variation in land, so that if some treatment has given good result, one should be confirmed that it is true and is not due to some other unknown reason. In these trials, a particular variety of crop is sown on the entire experimental field and throughout the growing season it is managed uniformly. All sources of variation except that are due to natural soil differences, and are held constant to the maximum extent. At the time of harvest a substantial border is removed from all sides of the field. The rest of the field is divided into number of small plots which are termed as basic units, with the same dimensions. The production from these basic units is harvested and recorded separately for each basic unit. Then the yields in these basic units are collected separately. The usefulness of a uniformity trial lies in the fact that neighboring units may be amalgamated to form larger plots of various sizes and shapes. The

[1] Department of Statistics, D. A-V. College, Kanpur, India.
[2] Department of Mathematics & Statistics, Dr R M L Avadh University, Faizabad-224001, India.
[3] Department of Statistics, D. A-V. College, Kanpur, India.

variation in yield over the field due to soil heterogeneity and other manual errors are generally summed up in the term "Experimental Error" and may be calculated for each type of plot thus formed. Hence, all efforts in designing field experiments are directed to measure and control this source of variation.

The coefficient of variation (C.V.), the ratio of standard deviation to arithmetic mean is a normalized measure of dispersion of a probability distribution. It tells us about the size of variation relative to the size of the observation, and is independent of the units of observation. It is an index of the precision of the experiment. The coefficient of variation and the plot size relationship has been investigated by several researchers including Mahalanobis (1940) and Panse (1941), etc. Panse and Sukhatme (1954) gave detailed description of uniformity trial experiments. The determination of the optimum plot size is an important step in field experimentation as it takes into account variability, both due to crop species and soil heterogeneity.

Smith (1938) gave an empirical model for describing relationship between the variance and the plot size for his field experiments. His model can be reduced to the following simple form as

$$Y = a\,X^{b} \tag{1}$$

where Y is Coefficient of Variation and X is size of the plot, a and b being parameters of the model to be estimated.

Haque *et al.* (1988) considered the following two models along with the model (1) for describing relationship between the plot size (X) and the Coefficient of Variation (Y) as,

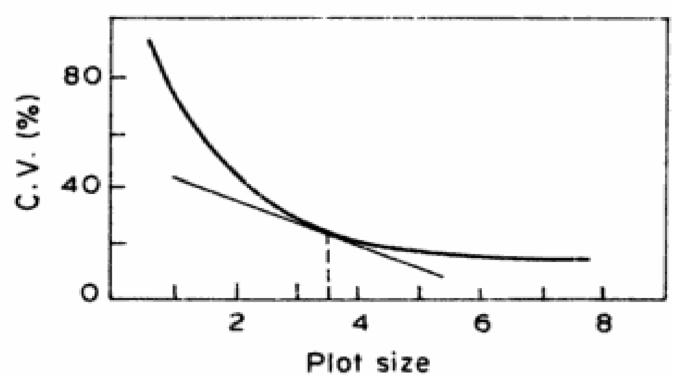
$$Y = a\,b^{X} \tag{2}$$

$$Y = a + \frac{b}{X} \tag{3}$$

Haque *et al.* (1988) arrived at the conclusion that the relationship (1) is the best among relationships (1) to (3) to describe the coefficient of variation and the plot size relationship. They calculated the point of maximum curvature for determining the optimum plot size and found the optimum plot size which corresponds to coefficient of variation (C.V.) of magnitude 25%. But this C.V. is quite high. They mentioned that in field experiments, generally the C.V. should not be more than 10-15%. If the C.V. is very high the reliability of the experimental results becomes doubtful. Therefore they suggested that instead of maximum curvature, it would be more logical to consider C.V. as the criterion for deciding the optimum plot size. In reference to the shape of the plots they showed that in all cases when x_1 (length) is measured along the fertility gradient and x_2 (width) across the fertility gradient rectangular plots are always optimum. They also suggested that if the experimenter has no idea of fertility gradient of the field, it is safer to use square shaped plots.

Draper and Smith (1998) classified the models (1) and (2) as intrinsically linear models, as they can be transformed into a form in which parameters appear linearly. The estimation of parameters a and b of these models can be done only after transforming them into a form in which parameters appear linearly by the well known method of least squares. The models (1) and (2) can be brought into linear form by using log transformation. However, it presupposes a multiplicative error term, a condition not so easy to justify. The direct application of least square method is not possible to estimate parameters of the models (1) and (2). Non-linear least squares estimation involves complicated iterative procedures. Convergence of solution is a serious problem in non-linear least squares estimation. Obtaining prior guess values of parameters in non-linear least squares estimation poses a serious problem before an investigator. The relation (3) is, however, a linear model and its parameter estimates can be obtained by direct application of classical least squares method of estimation.

The curvature is the amount by which a curve deviates from being flat. It is defined in different ways depending on the context. In uniformity trial experiments, the basic units of uniformity trials are combined to form new units. The new units are formed by combining columns, rows or both. Combination of columns and rows should be done in such a way that no column or row is left out. For each set of units, the coefficient of variation (C.V.) is computed. A curve is plotted by taking the plot size (in terms of basic units) on the X-axis and the C.V. values on the Y-axis of a graph sheet. The point at which the curve takes a turn that is the point of maximum curvature is located by inspection. The value corresponding to the point of maximum curvature will be the optimum plot size (Sundarraj, 1977). The following figure shows the point of maximum curvature expressed by dotted line.



This is only an approximate method of fixing the optimum plot size. Another method to obtain the point of maximum curvature is the calculus method. Fairfield Smith (1938) derived expression for maximum curvature for his model described by the relation (1) as

$$C = \frac{\left[1+(y_1)^2\right]^{3/2}}{y_2} \tag{4}$$

$$y_1 = \frac{dY}{dX} \quad \& \quad y_2 = \frac{d^2Y}{dX^2}$$

and therefore $$C = -\frac{1}{bX}(X^2+b^2)^{3/2} \tag{5}$$

On putting, $\frac{dC}{dX} = 0$, the solution of X will define the point of maximum curvature. For Fairfield Smith's model the value of C is

$$C = \frac{[1+\{abX^{(b-1)}\}^2]^{\frac{3}{2}}}{ab(b-1)X^{(b-2)}} \tag{6}$$

Putting $\frac{dC}{dX} = 0$ and substituting estimated values of the parameters a and b in it, the point of maximum curvature can be obtained.

In the present study we propose a linear model which relates the coefficient of variation to the plot size in a better way as compared to existing models. The expression for calculating the point of maximum curvature is also simple as compared to that of Fairfield Smith's model.

2. Proposed model

A linear model with its deterministic component is proposed to relate the plot size represented by X and Coefficient of Variation represented by Y as

$$Y = a + b\log X \tag{7}$$

The proposed model describes the relationship between the plot size and C.V. in a better way as compared to existing empirical models. a and b are parameters of the model which appear linearly in it and can be estimated by least squares method of estimation. The proposed model (7) was used by Shukla (2011) for his studies on uniformity trial experiments.

The model (7) admits an additive error term and can be written as

$$Y_i = a + b\log(X_i) + U_i, \qquad i = 1, 2........, n \tag{8}$$

where Y_i and X_i are i[th] observations of Y and X respectively. *U*s are independently and identically distributed random variables with mean zero and fixed variance σ^2. If *U*s follow normal distribution, i.e. $U \sim N(0, \sigma^2)$, the maximum likelihood estimates of a and b can also be obtained.

Following standard procedures as described in Draper and Smith (1998), the classical least squares estimators of parameters can be easily obtained. Let $\hat{a}$ and $\hat{b}$ are the least square estimates of a and b, respectively. The least squares estimate of Y_i that is $\hat{Y}_i$ will be

$$\hat{Y}_i = \hat{a} + \hat{b} \log X_i \tag{9}$$

The residual e_i is

$$e_i = Y_i - \hat{Y}_i \tag{10}$$

The appropriateness of the proposed model has been verified by examining the values of coefficient of determination-R^2, mean residual sum of square-s^2, mean absolute error (MAE), Akaike Information Criterion (AIC) and standardized residuals. Adopting the procedures as described in Montgomery *et al.* (2003), the analysis of residuals have been performed to verify the assumptions of zero mean, normal distribution and fixed variance of residuals. The point of maximum curvature can be obtained for the proposed model (7) as below,

$$C = -\frac{1}{bX}(X^2 + b^2)^{3/2} \tag{11}$$

On putting, $\frac{dC}{dX} = 0$, the point of maximum curvature can be obtained. It leads to the solution of $X = \pm\sqrt{b^2/2}$. As X will assume only positive values, the point of maximum curvature will be at $X = \sqrt{b^2/2}$. It is observed that expression for obtaining point of maximum curvature is much simpler for the proposed model as compare to that of Fairfield Smith's model.

3. Empirical study

The appropriateness and model adequacy of the proposed linear model (7) has been verified with the help of primary data given in Haque *et al.* (1988). Haque *et* *al.* (1988) worked on field experiments for wheat and taken a piece of land measuring 45 x 39 m^2 at Rajendra Agricultural University, Bihar, India. At the time of harvest, the land was subdivided into 45 x 39 = 1755 basic units, of size 1 x 1 m^2 , and grain yield was recorded in gram for each unit separately. We have computed the values of Coefficient of determination R^2 , residual mean square s^2 , Mean absolute error (MAE) and Akaike Information Criterion (AIC) for the models (1) to (3) & (7) and these values are listed in table 1 along with parameter estimates of the model (7). An analysis of residuals has also been performed for the model (7) by plotting normal probability plot and residual versus explanatory variable plot. The normal probability plot (Fig.1) is almost a straight line which conforms the assumption of normal distribution of residuals. The plot of residuals versus explanatory variables (Fig.2) for the model (7) does not show any systematic pattern. It conforms the assumption of homoscedasticity for residuals. The MAE values are also negligible. Thus, we infer that residuals of the model (7) admit the assumption of zero mean, normal distribution and fixed variance. We can conclude that the proposed linear model (7) adequately explains the relationship between the plot size and the C.V.

Figure 1.

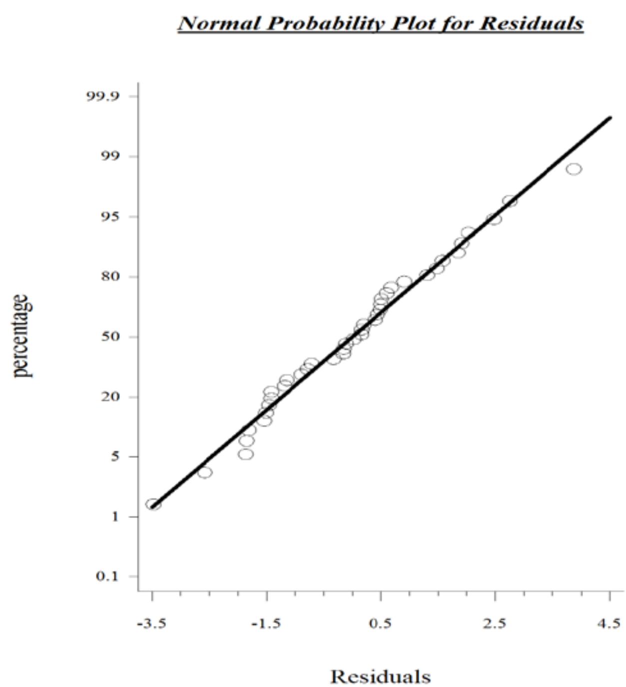
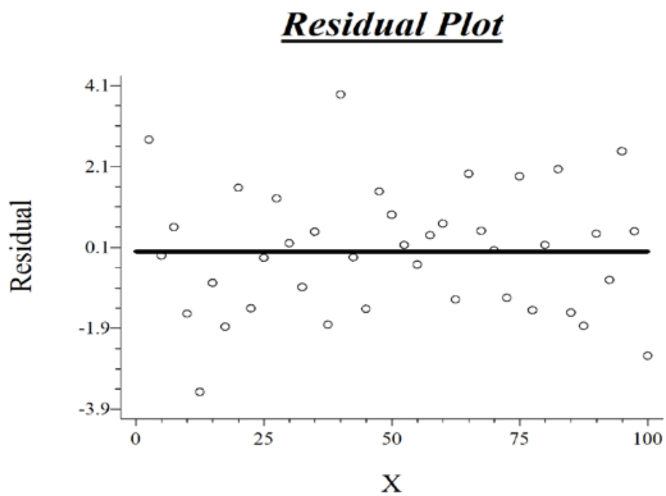


Figure 2.



On comparing values of R^2, s^2, MAE and AIC for the models (1) to (3) and (7) we have observed that the proposed linear model (7) has highest R^2 values and lowest s^2, MAE and AIC values. Thus, the model (7) better fits data sets as compared to the models (1) to (3). The model (7) is more appropriate to be used in uniformity trial experiments.

Table 1.

Parameters Estimates of Model (7)			$\hat{a}$ = 32.9928	$\hat{b}$ = -4.6674
	R^2	s^2	MAE	AIC
Model (7)	**0.9186**	**2.4638**	**1.2472**	**2.5868**
Model (1)	0.9110	2.6980	1.2547	2.8325
Model (2)	0.7810	6.6170	1.8190	6.9475
Model (3)	0.6750	9.7969	2.4658	10.2859

Table-2 gives the values of C.V. for different plot areas. Estimated values of C.V. using the model (1) and (7) are also given.

Table 2.

S. No.	Area(m^2) x	C.V. y	C.V. $\hat{y}_{(1)}$	C.V. $\hat{y}_{(7)}$	S. No.	Area(m^2) x	C.V. y	C.V. $\hat{y}_{(1)}$	C.V. $\hat{y}_{(7)}$
1	1	35.75	37.39	32.99	21	30	17.28	16.57	17.12
2	2	29.65	31.67	29.76	22	32	16.49	16.31	16.82
3	3	28.47	28.74	27.87	23	35	16.80	15.97	16.40
4	4	24.98	26.83	26.52	24	36	16.95	15.86	16.27
5	5	22.01	25.44	25.48	25	40	14.59	15.46	15.78
6	6	23.85	24.35	24.63	26	42	17.46	15.28	15.55
7	7	22.05	22.73	23.91	27	45	15.73	15.03	15.23
8	9	24.31	22.10	22.74	28	48	14.95	14.80	14.92
9	10	20.83	21.55	22.25	29	50	13.59	14.66	14.73
10	12	21.24	20.63	21.39	30	54	16.23	14.39	14.37
11	14	21.98	19.88	20.68	31	56	12.75	14.27	14.20
12	15	20.55	19.56	20.35	32	60	14.04	14.03	13.88
13	16	19.17	19.26	20.05	33	63	15.69	13.87	13.66
14	18	19.99	18.72	19.50	34	64	12.07	13.82	13.58
15	20	17.20	18.25	19.01	35	70	11.32	13.53	13.16
16	21	22.66	18.04	18.78	36	72	13.47	13.43	13.03
17	24	18.01	17.47	18.16	37	80	11.83	13.10	12.54
18	25	16.55	17.30	17.97	38	81	14.96	13.06	12.48
19	27	19.09	16.99	17.61	39	90	12.49	12.74	11.99
20	28	18.35	16.84	17.44	40	100	08.92	12.42	12.50

The point of maximum curvature for the proposed model (7) is $x = 3.30$, hence the optimum plot size which falls just near to this point of maximum curvature is $3m^2$ corresponding to which the C.V. is 27.86%, which is quite high. Therefore, as suggested by Haque *et al.* (1988), it would be more logical to consider C.V. as the criterion for deciding the optimum plot size.

4. Conclusions

It is submitted that the linear model (7) is a better alternative to describe the relationship between the plot size and the coefficient of variation in uniformity trial experiment. The proposed model (7) has highest R^2 values as compared to the models (1) to (3) which include Fairfield Smith's model also. Apart from it the model (7) has smallest values of s^2, MAE and AIC, as compared to all other models (1) to (3). The analysis of residuals also conforms the assumptions of zero mean, normal distribution and fixed variance for residuals. The expression for obtaining the point of maximum curvature is also easy to use for the model (7). The parameter estimates of the proposed model posses good statistical properties. Another advantage with this model is that it admits additive error term. The predictions and inferences as well as test of significance procedures for the model (7) can be easily carried out. It is therefore recommended that the linear model (7) should preferably be used in uniformity trial experiments.

Acknowledgements

The authors are very much thankful to the editor in chief and the unknown learned referee for critically examining the manuscript and giving valuable suggestions to improve it.

REFERENCES

DRAPER, N. R. AND SMITH, H., (1998). Applied Regression Analysis, Wiley, New York.

HAQUE, H. N. AZAD, N. K. JHA, R. N. and SINGH, S. N., (1988). Optimum Size and Shape of Plots for Wheat, Annals of Agricultural Research, 9, 2, 165-170.

MAHALANLBIS, P. C., (1940). A sample survey of the acreage under jute in Bengal, Sankhya, 4, 511-530.

MONTGOMERY, D. C. PECK, E. A. and VINING, G. C., (2003). Introduction to Linear Regression Analysis, Wiley.

PANSE, V. G., (1941). studies of the technique of field experiments size and shape of blocks and arrangements of plots in cotton trials, Indian J. agric. Sci, 11, 6, 850-865.

PANSE, V. G. and SUKHATME, P.V., (1954). Statistical Methods for Agricultural Workers, I C A R New Delhi.

SHUKLA, A. K., (2011). Some Contributions in Estimation of Asymptotic Regression Statistical Models, the Unpublished Thesis of Ph. D., CSJM University, Kanpur.

SMITH, H. F., (1938). An empirical law describing heterogeneity in the yield of Agricultural crops, Journal of Agricultural Science, 28, 1-23.

SUNDARARAJ, N., (1977). Technique for estimating optimum plot size and shape from fertilizer trial data, a modified approach. J. Ind. Soc. Agric. Stat., 29, 2, 80-4.