Available online at www.elixirpublishers.com (Elixir International Journal)

Statistics

Elixir Statistics 94 (2016) 40042-40044

ISSN: 2229-712X

Asymptotic Non-Linear Models for Uniformity Trial Experiments

Alok Kumar Shukla[1] and Subhash Kumar Yadav[2,*]

[1]Department of Statistics, D.A-V. College, Kanpur, U.P., INDIA.

[2]Department of Mathematics and Statistics (A Centre of Excellence),Dr. RML Avadh University, Faizabad-224001, U.P. India.

ARTICLE INFO

Article history:
Received: 22 March 2016;
Received in revised form:
1 May 2016;
Accepted: 4 May 2016;

Keywords
Fairfield Smith's Variance Law,
nonlinear model,
uniformity trial experiments.

ABSTRACT

Uniformity trials are needed to determine suitable shape and size of the plot for fertility variation in land. The suggested model is more adequate model over well-known Fairfield Smith variance law and also removes all the objections by Cochran (1977) for Fairfield Smith variance law model. The adequacy of suggested model has been examined by the given data in Haque et al. (1988).

1. Introduction

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

$$Y = \alpha X^{\beta} \quad (1)$$

Where Y is Coefficient of Variation and X is size of the plot, α and β being parameters of the model to be estimated.

The coefficient of variation and plot size relationship has been investigated by several researchers including Mahalanobis (1940) and Panse (1941) etc. Panse and Sukhatme (1954) have given detailed description of uniformity trial experiments. The determination of optimum plot size is an important step in field experimentation as it takes into account variability, both due to crop species and soil heterogeneity.

Uniformity trials are needed to determine suitable shape and size of the plot for knowing 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.

Haque et al. (1988) were done on field experiments and find model (1) good describe relationship between coefficient of variation (C.V.) and plot size. Haque et al. (1988) were given some suggestions as criterion for selection of optimum plot size should be coefficient of variation (C.V.), not the point of maximum curvature and also if the experimenter has no idea of fertility gradient of the field it would be safer to use square shaped plots. They calculated the point of maximum curvature for determining 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.

Misra (1992) proposed a linear model for better relationship between plot size (X) and Coefficient of Variation (Y)

$$Y = \alpha + \beta X + \rho / X, \; X > 0 \quad (2)$$

where α, β and ρ are the parameters. Misra (1992) also proof model (2) has an oblique asymptotic, $Y = \alpha + \beta X$. If β is very small, the asymptotic becomes $Y = \alpha + M$, where M is a finite quantity. This is because $\lim \beta X = M$ for very small values of β and large value of X.

Cochran (1977) objected on the model (1) when we apply in cluster sampling, Y increases without bound as X increases. He has been suggested Y approaches an upper bound with large X would be more appropriate. He is also suggested that found good function for fit over the range of X. But when we apply model (1) in Uniformity trial Experiments, Y decreases without bound as X increases.

2. Asymptotic Nonlinear Model

Cochran (1977) suggestions are applied in uniformity trial experiments under a class of nonlinear models with their deterministic components for the relationship between size of the plot represented by X and the Coefficient of Variation (C.V.) represented by Y as:

$$Y = \alpha + \delta X + \beta \rho^{X}, \quad 0 < \rho < 1 \text{ and } X > 0 \quad (3)$$

$$Y = \alpha + \delta / X + \beta \rho^{X}, \; 0 < \rho < 1 \text{ and } X > 0 \quad (4)$$

$$Y = \alpha + \delta / X^{2} + \beta \rho^{X}, 0 < \rho < 1 \text{ and } X > 0 \quad (5)$$

where α, δ, β and ρ are the parameters of the models.

Models (3)-(5) are asymptotic nonlinear models. Models (3)-(5) were used by Shukla at el. (2015) for relationship between variance within cluster and size of the cluster.

We can postulate statistical models for (3) - (5) as:

$$Y = \alpha + \delta X + \beta \rho^{X} + \in, 0 < \rho < 1 \text{ and } X > 0 \quad (6)$$

$$Y = \alpha + \delta / X + \beta \rho^{X} + \in, 0 < \rho < 1 \text{ and } X > 0 \quad (7)$$

Tele:
E-mail address: drskystats@gmail.com

$$Y = \alpha + \delta / X^2 + \beta\rho^X + \in, 0 < \rho < 1 \text{ and } X > 0 \quad (8)$$

The random error term $\in$ is assumed to be independently and identically normally distributed random variable with mean zero and fixed variance (σ^2).The constants α, δ, β and ρ are the unknown parameters of the models (6) - (8).

The models (6) - (8) never seen in literature for relationship between coefficient of variation and plot size.

2.1 Fitting of Different nonlinear Models

The above models are classified by Draper and Smith (1998) in two groups as intrinsically linear and intrinsically non-linear models, model (1) is intrinsically linear can be transformed into a form in which parameters appear linearly. The direct application of least square method is possible to estimate parameters of models (1) and (2). Models (6)-(8) are intrinsically non-linear models, then parameters of models (6)-(8) estimate by iterative procedure as Levenberg-Marquardt's method.

2.2 Goodness of fit of different models

Coefficient of Determination - R^2_{Adj}

Montgomery et al. (2003) have described Adjusted R^2 more efficient to R^2 if adding more parameters in model. Adjusted R^2 is good for model comparison when number of parameters not equal in two models.

The Coefficient of Determination- R^2_{Adj} is defined as

$$R^2_{Adj} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 / n-p}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 / n-1}$$

The value of R^2_{Adj} very close to 1 implies that most of the variability in Y has been explained by the fitted model. Thus observance of a high R^2_{Adj} value indicates a good fit.

Residual Mean Square- s^2

The best criterion to choose a model is the residual mean square (s^2). The residual mean square (s^2) is an unbiased estimator of σ^2. It's expression for our models will be

$$s^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-p} = \frac{\sum_{i=1}^{n} e_i^2}{n-p}$$

where n is number of observations and p is number of parameters. The value of s^2 indicates the error due to regression or model. A small value of s^2 reflects the appropriateness of the fitted model.

Mean Absolute Error (M.A.E)

The mean absolute error which is average of absolute error is defined as

$$M.A.E. = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n} = \frac{\sum_{i=1}^{n}\left|e_i\right|}{n}$$

Where n is number of observations. A smaller M.A.E. preferred in any data sets to which a model has been fitted.

Akaike Information Criterion (A.I.C.)

Gujarati & Sangeetha (2007) has given a lot of importance to Akaike information criterion.It is a very useful criterion for judging the performance of fitted model. It is also good for comparing two or more models. The model with lowest value of A.I.C. is preferred. It can be defined as

$$A.I.C = Exp\left(\frac{2p}{n}\right)\frac{RSS}{n}$$

Where n is number of observations and p is number of parameters. RSS is residual or error sum of square.

2.3 Examination of Residuals

Analysis of the residuals (errors) is strongly recommended to decide about the suitability of a model by Draper and Smith (1998). Three important assumptions made in the model are:

(i) Errors are not auto correlated

(ii) Errors are independent

(iii)Errors are normally distributed

The assumptions can be verified by examining the residuals.

Test for auto correlation of errors (Durbin-Watson Test)

We test H_0 : Errors are not auto correlated (if DW test values > d_U)

Against H_1 : Errors are auto correlated (if DW test values < d_L)

Where d_L and d_U given in Draper and Smith (1998).

Test for independence of errors (Run Test)

We test H_0 : Errors are independent

Against H_1 : Errors are not independent

Test for normality (Shapiro-Wilk Test, n<50)

We test H_0 : Errors are normally distributed

Against H_1 : Errors are not normally distributed

3. Empirical Study

The appropriateness and model adequacy of various models have been examined using the same data as given in Haque et al. (1988) on page 166 for square shaped plots. We have computed the values of Coefficient of determination(R^2_{Adj}), Residual mean square(s^2), Mean absolute error (M.A.E.), Akaike information criterion (A.I.C.), auto correlation of errors (Durbin & Watson Test), independence of errors (Run Test) and normality test (Shapiro-Wilk Test) for the models (1),(2),(6),(7) & (8). Table 1 gives parameter estimates for all models. Table 2 gives various goodness-of-fit values. Table-3 gives CV values and their estimates (based on models 1 and 7) for different plot areas. These values were obtained by SPSS 17.0 Statistical software.

Table 1. Parameters Estimates of Models.

Parameters	Model (1)	Model (2)	Model (6)	Model (7)	Model (8)
$\hat{\alpha}$	35.6331	20.929	21.6636	7.4942	10.5319
$\hat{\delta}$	-	-	-0.1226	13.8623	11.0172
$\hat{\beta}$	-0.2214	-0.116	19.9668	15.0941	15.2441
$\hat{\rho}$	-	16.239	0.6814	0.9851	0.9740

On comparing values of R^2_{Adj}, s^2, M.A.E. and A.I.C. for models (1),(2),(6),(7) & (8), we observed that nonlinear model (7) has highest R^2_{Adj} values and lowest s^2, M.A.E. and A.I.C. values. Thus the model (7) fits the data better. The analysis of residuals by Durbin & Watson Test values greater than 1.72(d_U) confirm that there is no problem of auto-correlations, in the residuals for all models, Run test confirms independence of errors and Shapiro-Wilk test confirms normality of residuals. We can conclude that the nonlinear Model (7) adequately explains the relationship between plot size and Coefficient of Variation.

Table 2. Goodness of fit of models & Residuals Analysis

	Model (1)	Model (2)	Model (6)	Model (7)	Model (8)
R^2_{Adj}	0.908	0.921	0.917	0.927	0.918
s^2	2.697	2.326	2.405	2.124	2.386
M.A.E.	1.254	1.187	1.175	1.094	1.179
A.I.C.	2.833	2.499	2.644	2.335	2.623
DW#	1.958	2.283	2.242	2.574	2.387
R*	0.819 (0.413)	1.198 (0.231)	1.623 (0.105)	1.783 (0.075)	1.121 (0.231)
W^	0.992 (0.996)	0.953 (0.138)	0.942 (0.057)	0.966 (0.385)	0.971 (0.497)

is Durbin & Watson Test values, * is Run test values, ^ is Shapiro-Wilk test values, the p-values are given within parentheses.

Conclusions

It is submitted that asymptotic nonlinear model (7) is a better alternative to describe the relationship between plot size and coefficient of variation in uniformity trial experiments. Existing models are not good in fit, some models increases without bound and are not well to large plot size. The suggested model removes all such drawbacks.

Therefore the model (7) is suggested to be used in uniformity trial experiments. Almost close asymptotic value of suggested model for maximum plot size is attained it is also proved by asymptotic value of suggested model. The suggested model (7) removes all drawbacks in uniformity trial experiments as mentioned by Cochran (1977) in cluster sampling. Thus it should be preferably used in uniformity trial experiments.

References

01. Cochran W. G. (1977), Sampling Technques. 3rd Ed., Wiley.

02. Draper N. R. & Smith H.(1998), Applied Regression Analysis, 3rd Ed., Wiley.

03. Gujarati D.N. & Sangeetha(2007), Basic Econometrics, 4th Ed, Tata McGraw-Hill.

04. Haque H. N., Azad N. K., Jha R. N. & Singh S. N.(1988), Optimum Size and Shape of Plots for Wheat, Annals of Agricultural Research, 9, 2, 165-170.

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

06. Misra G. C. (1992), Some Estimation Procedures in Non-Linear Statistical Models. The Unpublished Thesis of Ph.D., Kanpur University, Kanpur.

07. Montgomery D. C., Peck E. A. & Vining G. C.(2003), Introduction to Linear Regression Analysis, 3rd Ed., Wiley.

08. 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.

09. Panse V.G. & Sukhatme P.V.(1954), Statistical Methods for Agricultural Workers, I C A R New Delhi.

10. Shukla A.K., Yadav S.K. & Tiwari V. (2015), Improved Variance Function in Cluster Sampling, Int. J. Comp. Theo. Stat. 2, No. 1, 25-30.

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

Table 3. Estimated values of Coefficient of Variation (C.V.) using the model (1) and (7).

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	35.63	36.22	21	30	17.28	16.78	17.59
2	2	29.65	30.56	29.07	22	32	16.49	16.54	17.28
3	3	28.47	27.94	26.54	23	35	16.8	16.22	16.83
4	4	24.98	26.22	25.17	24	36	16.95	16.12	16.69
5	5	22.01	24.95	24.27	25	40	14.59	15.75	16.14
6	6	23.85	23.97	23.60	26	42	17.46	15.58	15.87
7	7	22.05	23.16	23.06	27	45	15.73	15.34	15.50
8	9	24.31	21.91	22.22	28	48	14.95	15.12	15.14
9	10	20.83	21.40	21.87	29	50	13.59	14.99	14.91
10	12	21.24	20.56	21.26	30	54	16.23	14.73	14.48
11	14	21.98	19.87	20.72	31	56	12.75	14.62	14.27
12	15	20.55	19.57	20.48	32	60	14.04	14.39	13.88
13	16	19.17	19.29	20.24	33	63	15.69	14.24	13.59
14	18	19.99	18.79	19.79	34	64	12.07	14.19	13.50
15	20	17.2	18.36	19.38	35	70	11.32	13.91	12.99
16	21	22.66	18.16	19.18	36	72	13.47	13.83	12.83
17	24	18.01	17.63	18.61	37	80	11.83	13.51	12.23
18	25	16.55	17.47	18.43	38	81	14.96	13.47	12.16
19	27	19.09	17.18	18.08	39	90	12.49	13.16	11.57
20	28	18.35	17.04	17.92	40	100	8.92	12.86	11.01