551·5: 631·52

Studies in Sampling Technique: Efficiency of Sample shapes for Paddy

P.S. SREENIVASAN.

(*Received 5th July* 1949)

ABSTRACT. In the present paper the sample shape employed in the crop-weather scheme, and seven other sample shapes consisting of six bunches have been studied with the grain yield of paddy. For this purpose the data on the complete harvesting experiment carried out at the Rice Breeding station, Karjat, in the year 1942 have been utilised. The efficiencies of the various shapes have been investigated by the technique of (a) analysis of variance and (b) intra-class correlation. These two techniques brought out conclusively the superiority of the present sample shape over the other seven shapes. Also the percentage information obtained for 5% and 10% sampling have been indicated.

1. *Introduction.*

The workers in the Agricultural Meteorology section at Poona, have made numerous investigations to try out and select the most efficient sampling technique for recording the growth and yield of various types of short crops like wheat, paddy, etc., and tall crops like jowar and sugarcane.[2,4,6] Kalamkar *et al*[2] discussed the precision observations on rice at Karjat recorded from 1934 to 1940 in a recent paper. At the end of their paper they suggested that instead of linear sampling unit of shape

o o x x x o o x x x o o x x x o o ;
(omit) (omit) (omit)

the sampling unit of the following shape may be tried in future experiments.

			omit			omit				omit		
1st	row	o	x	x	x	o	x	x	x	o	x	x
2nd	row	x	x	o	x	x	x	o	x	x	x	o
			omit			omit				omit		

In the present paper an attempt is made to test the suitability of this modified sample structure.

It may be mentioned here that if the sample structure is efficient, the amount of variation present "between samples" will be low. When a sample is made up of a number of ultimate units the variation present "within sample" (between ultimate units) should be of the same order as "between samples". If the ultimate units show a positive correlation then the "within sample" variance will be smaller than that "between samples". On the other hand, if the ultimate units have a negative correlation the "within sample" variance will be greater than that "between samples".[13] In selecting the most "efficient" from a number of possible sampling units we should aim at selecting one which on repeated trials indicates no correlation effect.

Kalamkar *et al*[2] in their sampling studies on rice found a positive correlation between adjacent bunches and hence recommended the modified sample referred to already. In similar studies with jowar, Sreenivasan[6] found that the two ultimate units of one metre lengths of jowar showed a positive correlation when they are observed in the same row, while there was no correlation between the two parallel one metre lengths in adjacent rows.

The object of the present paper is to discuss the complete harvesting experiment carried out at Karjat in 1942 in which the yield of paddy was recorded bunch by bunch and to test the comparative efficiencies of different types of sampling units which can be formed by suitable selection of the ultimate units of each type.

2. *Material.*

The data made use of here were collected at Karjat in the year 1942 under the kind guidance of Dr. B.S.Kadam, then Deputy Director of Agriculture (Crop Research) to the Government of Bombay. The exact layout of experiment and the general geographical and agricultural features of the Karjat tract are given in the paper "Precision observations on rice at Karjat" by R.J.Kalamkar *et al*[2]. The only point that may be mentioned here is that the whole field had been uniformly transplanted by the same variety (K.42) and the plots were then demarcated by long pegs. Hence in the analysis for harvest data two of the adjoining plots have been coalesced into one thus yielding six plots instead of the original twelve.

3. *Statistical analysis.*

As a preliminary to analysis, the normality of yield data has been tested. Fig. 1 gives the histogram of the frequency distribution of yield. The following are the values of g_1 and g_2 which are the measures of symmetry and kurtosis respectively and their standard errors.

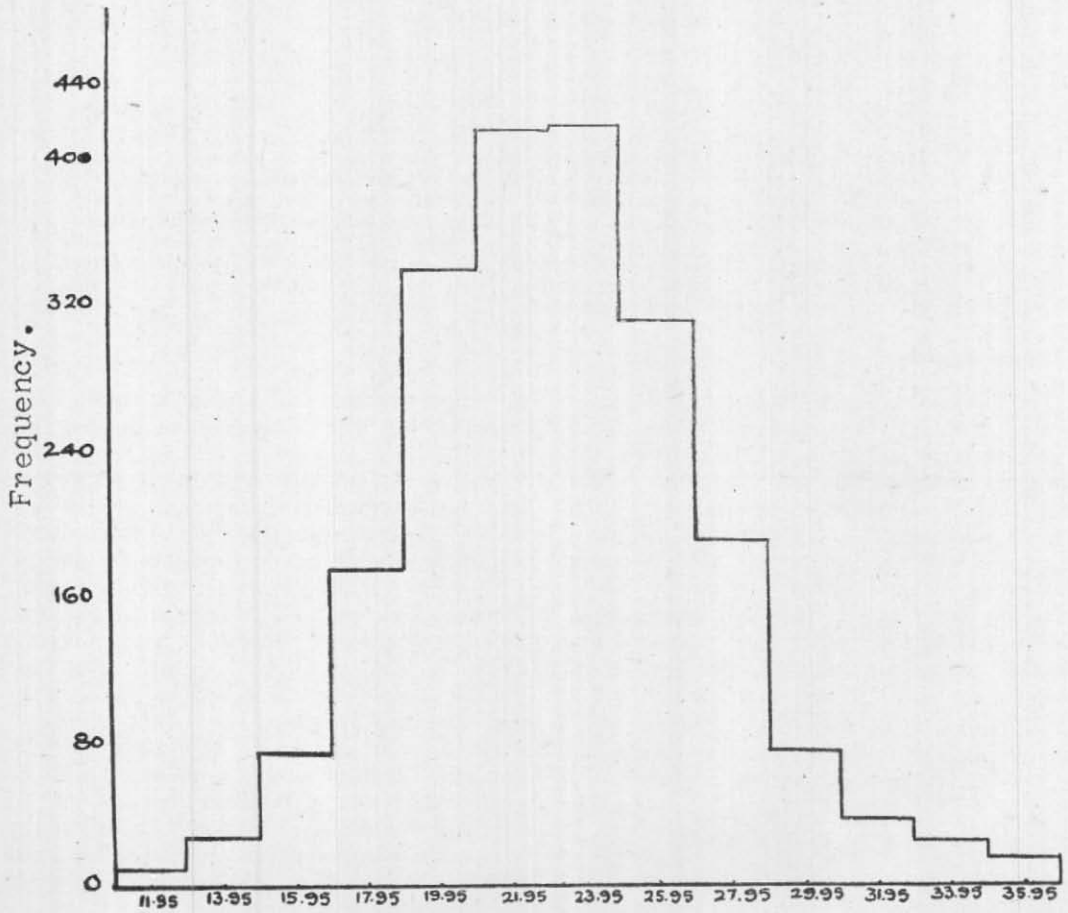

Fig 1. Weight of grains per bunch in grammes.

g_1 = + 0.283. S. E. = .0534

g_2 = + 0.869. S. E. = .1068

The distribution departs significantly from normality in having a positive skewness and is also peaked or "Leptokurtic".

In this paper with the same sample size that is being adopted for the co-ordinated crop-weather study, namely six bunches, eight sample structures have been studied. These structures are as follows:—

I o o o o o o

II o x o x o x o x o x o

III o o o
o o o

IV o x o x o x
x o x o x o

V o x x x o x x x o x x
x x o x x x o x x x o

VI o x x o x x
x o x x o x
x x o x x o

VII o x o x o x
x x x x x x
x o x o x o

VIII o x o
x o x
o x o
x o x

o means bunch under observation.
x means bunch not under observation.

Structure No.V is the modified structure suggested by Kalamkar *et al* and that is being used for crop-weather study in the co-ordinated crop-weather scheme (Agri-Met Technical circular No.50 drawn up by the Director of Agricultural Meteorology for the use of crop-weather observers in India)[8].

The variation present in the field in either direction has been worked out with square plot of size 12′ x 12′ and the following are the sum of squares:—

	D.F	S.S
Between rows	132	3,007.12
Between columns	132	4,872.25

Evidently the amount of variability present between rows is much less than that between columns. Except the last method the samples in remaining seven are elongated to a greater or lesser degree and hence the length of the sample is taken along the rows in which greater variability is present.

Two of the plots of size 12′ x 12′ were combined to yield 12′ x 24′ and 40 samples of six bunches each were constructed therein. There were 6 such plots yielding in all 240 samples or 1,440 bunches.

The variance present between plots "between samples" and "within sample" by the eight methods, are given in the combined analysis of variance (table 1.).

TABLE I.

Analysis of variance of paddy grain yield.

		Mean square by the method.							
Due to	D.F.	I	II	III	IV	V	VI	VII	VIII
Plots	5	242.380**	242.380**	242.380**	248.263**	239.169**	259.440**	229.416**	266.844**
Between Samples	234	21.211**	17.549	22.136**	19.629*	16.269	20.077**	21.447**	22.089**
Within Samples	1200	15.226	15.939	15.045	15.852	16.414	15.358	15.382	15.596
Total	1439	16.988	16.988	16.988	17.274	17.165	16.973	17.112	17.525

* Significant at 5% level ** Significant at 1% level

The following conclusions may be drawn from the above table: (1) The block variances are of the same order and highly significant. (2) In all methods of sampling, except V, (a) the variation "between samples" is greater than the variation "within sample" and (b) as judged by F test, it is highly significant in methods I, III, VI, VII and VIII and significant at 5% level in method IV. In method II the value of 'F' falls short of 5% level of significance.

It is thus evident that there is a positive correlation between bunches "within samples" in all methods except II and V. Method V has yielded practically the same variance "between samples" and "within sample", indicating the independance of the bunches observed inside the sample.

The sampling error percent per bunch calculated from the sample mean square obtained by method V is 17.11. According to the crop-weather scheme in India, 6 samples are to be observed in a plot of 1/100 acre and thus in a variety grown in 6 plots 216 bunches will be under observation. The sampling error percent per mean yield of a bunch when 216 bunches in a variety are observed will be 1.164%.

The intra-class correlation 'r' present between the components of a sample is given by the formula

$$r = \frac{A-B}{A + (n-1)B}$$

where A and B are mean squares "between samples" and "within sample" respectively and 'n' the number of bunches in a sample[5].

By the application of the above formula, the intra-class correlation present in these 8 sampling shapes have been worked out and given in table 2.

Sample structure III shows the greatest positive correlation of .073 between bunches of the same samples. The amount of information that could have been gained by locating the six bunches at random instead of taking them as per sample structure III may be calculated by the formula : $\frac{(n-1)\ (A-B)}{A + (n-1)\ B}$

where A and B are mean squares "between samples" and "within samples" respectively and 'n' the number of bunches per sample[5].

The information would have increased by 36.4% had the bunches been distributed independently. The numerator of the above formula which involves A—B indicates that the sample structure increases in efficiency as B tends to attain the value of A. A negative correlation between bunches of samples indicates high efficiency.

The loss of information is given by $L = 1 - \frac{1}{1+f\ (1-x)\ /\ x}$ where L is the loss of information, f the part of the total variance due to error, that is, variation between sampling units and x the percentage to be sampled (For a theoretical consideration of the formula, reference may be made to Yates and Zacopanay[7]). In the first instance the value of 'f' was found out. This is given in Table 3 along with the sampling and experimental errors.

TABLE 2.

Intra-class correlation in samples of various shapes.

Sample shape	I	II	III	IV	V	VI	VII	VIII
	**		**	*		**	**	**
Intra-class correlation	+.061	+.017	+.073	+.038	−.001	+.049	+.062	+.065

* Significant at 5% level

** Significant at 1% level.

TABLE 3.

Sampling and Experimental errors and percentage of total variance due to variation between sampling units.

Sample Structure.	Percentage Errors per plot		Percentage of total variance due to variation between sampling units, (f)
	Sampling	Experimental.	
I	1.1947	4.2587	7.87
II	1.1460	4.2587	7.24
III	1.2870	4.2587	9.13
IV	1.2115	4.3082	7.91
V	1.1043	4.2343	6.86
VI	1.2261	4.4076	7.74
VII	1.2652	4.1378	9.35
VIII	1.2876	4.4751	8.28

Here again the percentage of total variation between sampling unit is the least by the method V and next best method is II. Thus it may be concluded that the choice of the sample structure V suggested by Kalamkar *et al* is the right one and is fully justified.

The percentage information obtained for the following values of 'f' 6,7,8 and 9 are 68.0,65.8, 63.8 and 62.1 for 5% sampling and 78.9, 76.8, 74.9 and 73.2 for 10% sampling respectively. These values are obtained after applying a correction for L, $\left((L' = \frac{n_t - 2}{n_t} L) \right)$ where n_t is the number of degrees of freedom for plots[7].

It may be remarked that the experimental percentage error per plot is relatively small when compared to sampling error leading to lower percentage of information. The entire field seems to be quite uniform without broad soil irregularities.

4. *Summary.*

(1) Normality of the grain yield of complete harvest data of paddy at Karjat was tested.

(2) Analysis of variance for eight sample structures of six bunches per sample was carried out.

(3) The intra-class correlations present within sample were worked out.

(4) The amount of information that could have been gained by removing the intra-class correlation in structure III was calculated and found to be 36.4%.

(5) The percentage of total variance due to variation between sampling units (f) and the percentage information obtained were evaluated.

5. *Acknowledgments* .

The harvest data that are analysed in this paper were collected at the Rice Breeding station, Karjat and I am indebted to Dr. B. S. Kadam, Dy. Director of Agriculture (crop research) and to the staff of the Rice Breeding Station for the data. My sincere thanks are due to Dr. L. A. Ramdas, Director of Agricultural Meteorology for his interest in this investigation. Also my thanks are due to Mr. H.R. Ganesan for doing a part of the computation.

REFRENCES:

1. Clapham, A.R., *J. Agri. Sci.* (1931.)
2. Kalamkar, R.J., Kadam, B.S., Satakopan, V. and Gopal Rao, S, *Ind. J. Agri. Sci.*, XIII (1943).
3. Kalamkar, R.J., *J. Agri. Sci.* (1932).
4. Mallik, A.K., Satakopan, V. and Gopal Rao, S., *Ind. J. Agri. Sci.* XV (1945).
5. Snedecor, G.W., *Statistical Methods*, (Ames Iowa) (1946).
6. Sreenivasan, P.S., *Ind. J. Agri. Sci.*, XIII (1943).
7. Yates, F. and Zacopanay, I., *J. Agri. Sci.* (1935).