RANDOM SAMPLING FOR ESTIMATING RICE YIELD IN KOLABA, BOMBAY

By P. V. Sukhatme
(Imperial Council of Agricultural Research, New Delhi)
AND
B. S. Kadam and N. R. Bhat
(Department of Agriculture, Bombay)

Received November 6, 1945

During the last two years the Statistical Section of the Imperial Council of Agricultural Research has carried out several large-scale sample surveys for estimating the yield per acre and the total out-turn of wheat and rice. The first survey was carried out on wheat in 1943–44 in the provinces of the Punjab and the U.P. A report containing the results of this survey in the Punjab has appeared in a previous number of these *Proceedings*.[1] A brief account has also appeared in *Nature*.[2] During the year 1944–45, the survey was extended to the whole of India's wheat belt comprising of the provinces of Sind, North-West Frontier Province, Punjab, United Provinces and the Central Provinces. In the current season the survey has been extended to all the rice-growing districts of the United Provinces, Central Provinces, Bombay, Madras, Orissa and Bihar.

2. As a preliminary to extending the surveys to paddy crop, pilot surveys were conducted in the last season in one district each of the provinces of Bombay, Central Provinces and Madras with the objects of (*a*) testing the applicability of the technique developed for wheat for estimating the yield per acre of paddy, and (*b*) for gaining information on the practical difficulties peculiar to paddy crop likely to be encountered in field work. The results of the pilot survey conducted in the Tanjore District of the Madras Province, and their bearing on the technique of random sampling are described elsewhere. The results of the survey conducted in the Kolaba District of the Bombay Province form the subject of this paper.

3. The district of Kolaba was selected for the scheme as typical of the paddy-growing tract of the Bombay Province. It is divided into seven taluks covering an area of 2,165 sq. miles. Of the total cultivated area in the district approximately 67 per cent. is under paddy. It receives ordinarily a rainfall of over 80 inches. It is for the most part a hilly tract containing 1,556 villages.

4. The survey was conducted under the technical direction of the Statistical Adviser, Imperial Council of Agricultural Research, New Delhi, and under the administrative control of the Director of Agriculture exercised through the Deputy Director of Agriculture (Crop Research). The Technical Assistant to the latter was the officer in charge of the training and inspection. The field staff consisted of 7 Agricultural Assistants and 7 Kamgars at the rate of one Agricultural Assistant and one Kamgar for each of the 7 taluks of the district. All the Agricultural Assistants were men with considerable practical training in agriculture and had some knowledge of the tract and of the rice crop.

5. The plan adopted for the survey was the stratified plan of random sampling with taluks as the strata, early, mid-late and late varieties as the sub-strata, a village as the primary unit of sampling, a field as the sub-unit of sampling, and one-guntha plot as the ultimate unit of sampling. The three substrata were defined as follows according to the local conventions :

Early crop : Crop maturing about the 15th of October;
Mid-late crop : Crop maturing between the 16th and 25th of October ;
Late crop : Crop maturing after 25th October.

A total of 84 villages representing approximately 5·4% of the total number of villages in the district was selected. These were divided equally among the 7 taluks. Within each taluka 3 villages were selected for early crop, 5 for mid-late crop and 4 for the late crop. The villages were selected at random from amongst all villages in the taluks separately for the early, mid-late and late strata, so that every village got the same chance of being included in the 3 sub-strata in the taluk. The selection was made with the help of printed random numbers to ensure equal chance for every village in the taluk. The selection was done personally by the Statistical Adviser since he anticipated that considerations of distance and lack of communications and other difficulties might weigh with the provincial staff and be allowed to interfere with the random character of the sample.

6. In villages selected for 'Early Crop', experiments were conducted on early paddy only; in villages selected for 'mid late' crop experiments were conducted on mid-late varieties only ; and similarly in villages for 'late crop' experiments were conducted on late varieties. In each village, 3 fields from amongst all those growing paddy of the particular description were selected at random, and within a selected field a plot measuring $66' \times 16\frac{1}{2}'$ (1/40th of an acre) was located at random. Thus in a village belonging to the early crop, the Agricultural Assistant was required to select 3 paddy

fields at random from amongst all the early paddy fields in the village ; if a village belonged to mid-late crop, the Assistant was required to select 3 paddy fields growing mid-late variety from amongst all mid-late paddy fields in the village ; and so on. As there was a possibility that a village might not have the requisite number of fields of the given description, additional villages were given. If one or more of the villages did not contain the requisite number of fields of the particular description, they were to be rejected and substituted one by one by the villages in the supplementary list in the order given.

7. A detailed set of instructions was drawn for the guidance of the field staff for (*a*) selection of fields, (*b*) location of plots, (*c*) harvesting and connected operations, and (*d*) driage. The fields were selected in advance, to ensure that all the fields in the villages got an equal chance of being sampled. The plots in the selected fields were however marked only on the dates on which the crop was harvested to ensure that the plots were not tampered with. All the operations of harvesting, threshing, winnowing and weighing were carried out by the Agricultural Assistants assisted by field men, and were supervised by the Officer-in-Charge.

8. Four forms of returns were prescribed under the instructions. In Form No. 1, the Agricultural Assistant with the assistance of the Patwari concerned, is required to show the survey numbers and areas of all the paddy-growing fields in the selected village, arrange them serially, and to select 3 fields out of these by using the random numbers supplied to him. In case the number of fields was large, the Agricultural Assistant was asked to select three fields directly from the Patwari's Register with the survey numbers corresponding to the random numbers selected by him. In form 2 of the returns, he is required to give in respect of the three selected fields the pair of random numbers selected for locating plots, the type and level of the soil, preceding crop, manure in current season, the variety of paddy, the estimated yield of paddy per acre, and finally the date fixed for harvesting. The Agricultural Assistant was also asked to give detailed remarks in regard to the general condition of the crop. Both the forms 1 and 2 were required to be sent to the Statistical Adviser immediately after the selection was over. Under form No. 3 the Agricultural Assistant was required to fill the actual results of harvesting in respect of each plot in the village. The Agricultural Assistant was specially asked to note any changes in the condition of the crop that may have taken place since the fields were selected. The return was required to be sent immediately after the harvesting in each village was over. Form No. 4 sets out the results of driage.

9. The scheme commenced on the 1st of September 1944. The first fortnight was spent in preliminary arrangements, *e.g.*, planning, the training of the field staff, arranging for help from the revenue officials, etc. The training of the field staff commenced on the 15th of September and continued for a week. As the entire success of such a survey depends on the reliability of field work and consequently on the training received by the staff, the greatest attention was paid to the question of the training of field staff. Firstly, the field staff were gathered at the Rice Breeding Station, where the plan of work, the duties and responsbilities of each memberof the field staff, the meaning and implications of the instructions drawn for the actual conduct of field work, the manner in which the returns were to be filled in and despatched, were read out, discussed and explained at length several times over, until the workers, particularly the Agricultural Assistants, obtained a clear idea of the entire procedure of work. It was impressed on them that the work they were doing was of great importance, especially during the emergent food situation at the present time and demanded extreme scientific honesty. The training was completed with practical exercises in the fields of the Rice Breeding Station and of the cultivators round about the farm. The fields chosen were of various categories, including bunded fields, of different sizes and shapes, irrigated and unirrigated, so that the staff could have a thorough grounding in the intricacies of the work. Before they dispersed, the workers were provided with a cross-staff, a 100-feet measuring tape, coir strings, scales and weights, bamboo sticks, two threshing durries, and the requisite number of copies of various returns and the sets of random numbers. They were advised to halt for a day or two at the taluka centres, get acquainted with the revenue staff (*e.g.*, the Circle Inspectors, Talatis), and obtain general information about the villages selected for work and the routes thereto.

The training classes were also attended by prominent public workers of the district who had been for some time pressing their point of view that the district did not grow as much rice as was made out in the official forecast, and who therefore were interested in this new method of estimating the actual production.

10. After the staff returned to their respective talukas they proceeded to locate villages one by one. The villages selected were scattered over the taluka as they should be in the random selection. Some of them lay very far off from the only main road in the province, on which there is a bus service, and a few were even at prohibitive heights on mountain slopes. Even though the staff knew the routes to the selected villages, they found it

pretty difficult, on account of the scanty and unsatisfactory communication, heavy rain, and slippery paths, to reach the villages in proper time. As the scheme was prepared at very short notice, and as the harvesting of the early crop was about to commence, there was not even time to write to the Patwaris of the selected villages to ascertain whether the selected villages had the requisite number of fields of the given description. Consequently the Agricultural Assistants themselves had to visit villages one by one in order to ascertain whether there were the requisite number of fields. A large number of villages in the early crop had to be rejected, and then substituted by villages in the supplementary list, because the early crop, contrary to the unconfirmed impression, made only an insignificant portion of the whole rice area in the district. Even villages in the supplementary list had to be rejected. As the harvesting of the early crop had already commenced, the work could be carried out in only 12 out of the 21 villages allotted to the early crop. No difficulty was, however, experienced in locating the villages under mid-late and late crops.

11. Out of a total of 225 fields (3 in each of the 75 villages) selected under the scheme, experiments could be performed in 223. In the remaining 2, owing to the impending rains, the cultivator harvested the crop before the due date. They had given intimation to the Patwari but the latter did not deliver the message to the Agricultural Assistant in time. In general, the cultivators of all the selected fields willingly agreed to have the experiments done in their fields. They readily arranged for the labour required for harvesting, and, in their turn, were paid a fixed amount towards labour charges. As a rule, wherever possible, the operations of harvesting, threshing, winnowing and weighing were done on the same day, but in cases where the produce was moist, it was allowed to dry up under the care of the Assistant and Kamdar and was threshed after a day or two in their presence. The produce obtained was invariably weighed with the help of standard weights supplied to the staff.

12. The work was initiated and inspected by one of us by continuous touring during the working period in two rounds. First-hand information about the difficulties of the workers was gathered during these visits; for example, almost all the workers pointed out the risk of attacks by poisonous snakes which they had to take in walking on the bunds and through the tall grass. Then again, difficulty was experienced in measuring the dimensions of the field. As per instructions, the assistants were asked to measure the length and breadth of the field in steps. But this was often not practicable owing to obstructions, *e.g.*, thickets, trees, and hedges. A

minor departure was therefore found necessary in the measurement of length and breadth of the selected fields. These were either measured on the maps or in the field by bamboo sticks and then converted into steps.

13. The returns which numbered about 500, were scrutinised in the Statistical Section at New Delhi as they came in. The scrutiny showed that the information supplied including that on the general conditions of the crop at the time of the selection and on the changes in the condition of the crop since the date of selection was detailed and complete.

14. The season was somewhat abnormal from the point of view of the rice crop. The rainfall was deficient, and there was little of it between the 10th September and the 26th October. This being the most vital period of grain development, the mid-late and late varieties suffered throughout the district. The loss was considerable in the case of the crops growing in salt land. In some of the villages in these areas, the crop failed to a large extent. There was heavy rain and storm for about four days from the 27th October onwards, particularly in the Alibagh Taluka, when the late varieties were being harvested. This made the crops lodge in water and consequently some of the ripe grains were shed in the fields and a few grains in the earlies even germinated.

15. Table I shows the results in respect of the average estimated yield of dried paddy (in husk) and of rice (taken at 2/3 the weight of dried paddy)

TABLE I

Name of taluk	Acreage under paddy	Average yield of dried paddy in lbs. per acre	Average yield of rice in lbs. per acre	% of Standard Error of the estimates of Columns 3 & 4	Total out-turn of rice in tons
Karjat	39350	1843·99	1229·39 (36)	10·84	21596
Mahad	36359	1540·08	1026·77 (37)	5·61	16666
Panwel	63055	1789·07	1192·77 (26)	9·64	33576
Pen	36023	1318·98	879·36 (36)	9·30	14142
Mangaon ..	34275	1524·29	1016·24 (33)	7·90	15549
Roha	27196	1940·97	1294·04 (27)	4·64	15711
Alibag	39211	1338·02	892·06 (30)	15·78	15615
Whole district ..	275469	1620·39	1080·32 (223)	3·36	132855

Figures in the parenthesis indicate the number of plots on which the estimate is based.

for each taluk and for the district as a whole together with sampling errors. The last row of the table shows these figures for the district as a whole. The district average is a weighted average of the taluk mean yields, weights being in proportion to the area under paddy. The method of calculating the standard error of the taluk and the district estimates of mean yield is explained later on. It will be seen that the average yield of rice for the district as a whole is 1080 lbs. per acre as against the official figure of 1043 lbs. per acre. The sampling error of the district estimate of mean yield is less than 3½ per cent.

16. Table II shows the figures for the average yield in lbs. per acre of dried paddy and of rice (taking rice at 2/3 the weight of dried paddy) for

TABLE II

Name of category	Average yield of rice in lbs. per acre	% Standard Error of Column 3
Early	910·71	4·76
Mid-late	870·71	5·15
Late	1345·74	5·71

the early, mid-late and late varieties with the percentage standard errors. It will be seen that the average yield of the early and mid-late varieties is of the same order, but the average estimated yield of the late variety is about 50 per cent. higher than either. These varietal differences are inclusive of all factors such as land, sowing date, etc., which differentiate early, mid-late and late varieties and do not refer only to differences due to seed.

17. The figures for dried paddy reported in Tables I and II do not refer to the weight of paddy which is threshed on the same day as it is harvested, but have been corrected for a number of factors, *e.g.*, driage, area under field boundaries, and losses in the process of harvesting, threshing and winnowing, etc. It is common knowledge that the cultivators allow the harvested produce to dry in the field and/or on the threshing floor for a few days before threshing it. It is consequently necessary to allow for the resulting driage to bring the reported results in line with the actual practice. Equally, the yield per acre derived from the results of crop-cutting experiments relate to the area sown with paddy excluding the area covered by the field boundaries. On the other hand, the official figures of acreage include the area covered by the field boundaries within each survey number. It is therefore clear that we should correct either the acreage or the mean yield by reducing it by the proportion of the area covered by these boundaries. Again, we find

that small losses invariably occur during the process of harvesting, threshing, winnowing and weighing. There is no means other than the reports of the field staff for estimating these losses. The losses due to driage have been estimated from returns received from the field staff. The correction needed on account of the field boundaries has been estimated from a knowledge of the length and breadth of fields and the width of the field boundaries. We have combined the losses due to the acreage covered by field boundaries and the losses in the processes of harvesting, threshing and winnowing with the loss on driage into an over-all correction factor and have applied it for estimating the final yield of dry paddy. The over-all correction factor for the district is 5·5 per cent. It is the figures thus corrected which have been reported in Tables I and II.

18. Table III shows the results of the analysis of variance of plot yields for each of the different taluks and Table IV shows the analysis of variance results for each of the three early, mid-late and late varieties. The analysis of variance pooled over the district is shown at the bottom of each table. The standard error of the district estimate of the average yield was calculated from the analysis of variance table for the district, as the results of the Chi Square test of significance had shown that there was no evidence to suggest that the mean square between villages in the different sub-strata were different.

TABLE III

Name of Taluk	Variance due to	d.f.	m.s.
Karjat	Between varieties (Early, Mid-late, Late)	2	94940·57
	Between villages	9	19574·72
	Within villages	24	10448·72
Mahad	Between varieties do.	2	35304·27
	Between villages	9	8510·89
	Within villages	23	7723·47
Panwel	Between varieties do.	1	182835·99
	Between villages	7	36497·02
	Within villages	17	12995·13
Pen	Between varieties do.	2	22376·73
	Between villages	9	34730·17
	Within villages	24	7350·49
Mangaon	Between varieties do.	2	44839·86
	Between villages	8	17047·75
	Within villages	22	16414·87
Roha	Between varieties do.	1	124064·08
	Between villages	7	9414·90
	Within villages	18	7874·48
Alibag	Between varieties do.	2	250391·05
	Between villages	7	38245·67
	Within villages	20	4101·22
Whole District ..	Between taluks	6	49174·01
	Remainder	12	100217·07
	Between villages	56	23050·49
	Within villages	148	9070·59

TABLE IV

Name of varieties	Variance due to	d.f.	m.s.
Early ..	Between Taluks	4	13421·71
	Between Villages	7	10300·80
	Within Villages	23	13113·06
Mid-late ..	Between Taluks	6	23979·31
	Between Villages	28	17469·24
	Within Villages	69	7266·99
Late ..	Between Taluks	6	42476·61
	Between Villages	21	34742·05
	Within Villages	56	9632·69
Whole District ..	Between Varieties (Early, Mid-late, Late)	2	522613·25
	Remainder	16	28276·41
	Between Villages	56	23050·49
	Within Villages	148	9070·59

19. When the number of fields sampled from each village is the same, then the sampling error of the taluk and the district estimate of mean yield are directly obtained from the mean square between villages in the Analysis of Variance Table. Usually, on account of some reason or other, the number of fields is unequal. In the present case, there is one village in Mahad and one in Panwel taluk where only two experiments were conducted instead of three. In such a case, as pointed out by Cochran,[4] the sampling variance of the taluk mean yield is given by:

$$\frac{V \Sigma n_i^2}{N^2} + \frac{F}{N},$$

where V and F denote the estimates of the true variance between villages and between fields within villages respectively, and are obtained by equating the mean squares between villages and between fields within villages in the analysis of variance with

$$\frac{V}{m-1}\left\{N - \frac{\Sigma n_i^2}{N}\right\} + F \text{ and } F,$$

where n_i = number of fields in the ith village,

N = number of fields in the taluk,

and m = number of villages in the taluk.

When the number of fields in each village is the same, then the formula for the sampling variance of the taluk mean yield reduces to $\frac{V}{m} + \frac{F}{mn}$ which is identical with the mean square between villages in the analysis of variance table. The sampling variance of the estimate of mean yield for the district

was calculated from the formula. $\frac{V}{\Sigma m} + \frac{F}{n \Sigma m}$ where V and F refer to the estimates of the variance for the district.

20. Turning to Table III, it will be seen that in all the taluks the mean square between villages is larger than the mean square between fields within villages and in Panwel, Pen and Alibag the difference is significant. Table IV also shows that the variance between villages is significantly larger than the variance between fields within villages in the case of the mid-late and late paddy, but the difference is negative and non-significant in the case of early paddy. The district analysis shows that the mean square between villages is more than twice the mean square between fields within villages, and that varieties represent a predominant source of variation. These results concerning the relative magnitude of the variability between and within villages have an important bearing in determining the number of experiments and their distribution as between villages and fields in a village for estimating the mean yield with a given accuracy. The formula for sampling variance of mean yield shows that for a given number of experiments, the estimate of mean yield attains the maximum precision when the number of experiments is so distributed that one experiment each is conducted in a different village of the stratum and that the precision of the estimated mean yield falls off when the number of experiments per village is increased at the expense of the number of villages in the district. This will be clear from Table V showing the number of villages that must be sampled for a given

TABLE V

No. of fields per village	No. of villages required for different % of Standard Errors		
	5%	7·5%	10%
1	47	21	12
2	32	14	8
3	26	12	7
4	24	11	6
6	21	9	5
8	20	9	5

number of fields per village and for one plot per field in order to estimate the district mean yield with a given accuracy. It will be seen that there is considerable advantage in increasing the number of fields from one to two, but the advantage is rapidly lost with further increase in the number of fields. Thus, whereas the number of villages required for obtaining an estimate with five per cent. error decreases by 15 when the number of fields is increased

from one to two per village, the number of villages required decreases by only 6 when the number of fields is increased from two to three. There is hardly any material difference in the number of villages that must be selected when the number of fields is increased from three to four and more. The number of villages shown in the table refers to the whole of the district and is to be distributed among the taluks in proportion to the area under paddy. It is of course subject to sampling error since the values of the true variance between villages and within villages are themselves only sample estimates of the population values.

21. The calculations in Table V have been based on the assumption that it is possible to eliminate the differences in yield between the early, mid-late and late paddy from the error of the final estimate of yield by employing varieties as substrata. This may not be always possible, since the areas under varieties are not known for all districts. In the present scheme special effort was made through the Revenue Department to obtain these statistics. Ordinarily, therefore, independent samples of villages for early, mid-late and late varieties cannot be selected in the absence of area statistics under individual varieties. The requisite number of fields in the selected villages will have to be chosen irrespective of whether the fields grow early, mid-late or late paddy. As the mean square due to varietal differences in the Analysis of Variance is significantly larger than the mean square between villages, an unrestricted selection of fields in the selected villages will lead to the increase in the value of the mean square between fields within villages and also in the value of mean square between villages itself. It is difficult to form any precise idea of the extent by which the variance of the estimated mean yield will be increased on this score. One would be on the safer side in summing up the sums of squares due to varieties and due to between villages and dividing this sum by the sum of the degrees of freedom in order to form an idea of the resulting value of the mean square between villages. On this basis it would seem that the number of villages will have to be increased by over two-thirds in order to estimate the district mean with the same precision as if each taluk was stratified by early, mid-late and late varieties.

22. As there is a difference of about a fortnight between the maturing stages of the early, mid-late and late varieties, it may be necessary to visit the same village more than once in the case of unrestricted selection, since the fields mature at different times. A method which enables us to harvest the selected fields in a small number of visits such as stratification, by early, mid-late and late varieties is not only statistically superior, but is economic and convenient. It, however, appears that we may have to

forego this advantage in the absence of area statistics under early, mid-late and late paddy. The early variety by itself occupies less than 5 per cent. of the total area under paddy; moreover, the average yields of the early and mid-late paddy are of the same order. Even if, therefore, acreage under early and mid-late is combined and area figures are known separately for only two crops, early plus mid-late and late paddy, it will still be a great advantage to stratify taluks by these two categories. In the absence of separate area statistics for early plus mid-late and late paddy, a larger number of villages will have to be selected, and a field worker may be required to visit a village on an average more than once. Thus, whereas only four villages per taluk are required with stratification, seven villages per taluk would be required without it. The charge of 7 villages per man may prove too large for him to visit all the villages in his jurisdiction within the period available for harvesting. Indeed, additional staff may be required. With four villages per taluk, two from the early plus mid-late paddy, and two from the late paddy, work can be more easily managed by the existing departmental staff in the course of their normal duties. It would thus seem that stratification by varieties makes an appreciable difference to the plan of work and that it would, therefore, be desirable to ascertain separately the area under late paddy. Until the time steps are taken to ascertain the area under late paddy in each district, we shall have to be content with a slightly lower degree of precision for the district estimates of average yield, assuming, of course, that the work has to be carried out by the existing departmental agency.

23. It was not a part of this scheme to determine the optimum size of plot. This has been investigated elsewhere. A brief description of the results of this investigation is given in Table III. The conclusion reached is that both on theoretical and practical grounds a large-size plot is to be preferred to a small size one such as is marked by a rigid or semi-rigid frame. Within the range of large-size plots, however, it may be feasible to adopt a plot of smaller dimensions consistent with the requirement that errors due to the measurement of the plot and the processing of the produce do not contribute more than a negligible proportion of the produce harvested. All the available information suggests that there is a definite risk of obtaining a biased yield estimate under Indian conditions by conducting crop-cutting experiments with rigid or semi-rigid frames of area smaller than 20 sq. ft. Crop-cutting experiments on paddy in the Bombay Province are usually conducted on one-guntha plots of size $33' \times 33'$. There need be no objection to the use of this size to which the officials are accustomed since the results show that the choice of one plot in preference to another within the range

of 1/80th to 1/40th of an acre is of little importance in improving the precision of the estimate of yield. One has further to ensure that fields which are too small to contain a plot do not form more than a negligible proportion of the total population. The results of this survey show that the proportion of fields rejected on this score is less than 2%, but there are approximately four per cent. of fields whose breadth is between $16\frac{1}{2}$ and 33 feet all of which will have to be rejected in case the official plot size is adopted. It is accordingly suggested that the plot size $33' \times 16\frac{1}{2}'$ may be adopted in the hilly districts of Konkanpatti, but elsewhere in the province the present official size $3^{3}3' \times 33'$ may be retained.

24. There is an inherent difficulty in sampling from a finite field. Whatever the size of plot, the central portion of the field is relatively over-sampled when compared with the surrounding area. Consequently there is a likelihood of obtaining a biased estimate of the mean yield if the yield of plots located near the borders is different from that of plots in the remaining part of the field. The whole question has been considered at length elsewhere. It has been shown that the bias, if it exists, can be corrected by introducing a correction factor such that it will equalise the chance of every portion of the field being included in the sample. An application of this correction factor to plot yields in the present case shows that the estimate of average yield remains unaltered, the difference between the original estimate and the corrected one being less than one per cent. The point was also examined by tabulating the frequency distribution of plot yields according to the relative position of the plots in the field. The results are shown in Table VI. The relative position of the plot defined by the quotient of the random number of the starting point of the plot by the length of the field

TABLE VI

Length			Breadth		
Relative position of plots	No. of plots	Average yield per plot in chattacks	Relative position of plots	No. of plots	Average yield per plot in chattacks
0·0 to and including − ·1	19	266·94	·0 to and including − ·1	26	300·53
− ·2	16	395·81	− ·2	24	306·62
− ·3	27	365·03	− ·3	21	351·23
− ·4	30	336·63	− ·4	16	333·25
− ·5	30	318·46	− ·5	25	319·64
− ·6	18	312·27	− ·6	25	318·12
− ·7	17	360·41	− ·7	22	330·31
− ·8	27	298·77	− ·8	21	377·76
− ·9	19	392·68	− ·9	25	335·56
− 1·00	19	342·75	− 1·00	17	368·17

less the length of the plot is given in column 1. The number of plots having the relative positions defined in column 1 is shown in column 2, and the average yield per acre is given in column 3. Columns 4, 5 and 6 give the same figures when the relative position is defined along the breadth of the field. The Chi Square Test of significance for the number of plots in the different class intervals shows that the location of the starting point of the plot was genuinely random, as was to be expected. A comparison of the variation within classes with the variation between classes shows that the estimate of the yield per acre is not influenced by the relative position of the plot in the field.

25. The question of the shape of plot is being investigated in other provinces. It appears, however, that shape has no influence on the precision of the mean yield. There is, however, an advantage in marking out a triangular plot in place of a rectangular one, and that is, that the time taken for marking out the former is much less than the time required for marking out the latter. A suitable triangular plot whose size is approximately half a guntha will be an equilateral triangular plot with sides 33′. The only disadvantage in the adoption of this plot size will be that on account of its breadth of 28′, it will lead to a slightly larger rejection of fields than a plot size of 33′ × 16½′.

26. Random sampling implies equal chance for every plot in a taluk to be included in the sample. This can be done by considering the whole of the area of the taluk divided into plots and selecting random sample from these plots. This, however, is not a feasible method. The practicable method of selecting a random sample of plots is to select a random sample of fields and locating sample plots in each of these selected fields. The easiest way of selecting a random sample of fields is to select random groups of fields and then to select fields at random from each of these groups. The best group is the one for which the list of fields is readily available and in which the fields are readily located with the help of the local officials. The investigations of Panse and Kalamkar[5] in India show that there is no difficulty in selecting villages at random, nor is there any difficulty in selecting fields at random in the selected villages. In this method, however, every unit area does not get an equal chance of selection. Smaller villages and smaller fields get relatively larger chance of selection. This requires correction particularly if there is an association between the area and the yield. One method of correcting is by forming progressive totals of areas and of fields in selected villages before a selection is made. Another method would be to work out weighted averages by weighting the yield of each plot with the

area of the corresponding plot and weighting the average yield of a village with the corresponding area under paddy. These weighted averages have been calculated and show that these sources of bias in sampling have no effect on the estimated yield. It appears that no correction would be necessary in practice.

27. To conclude, it will be seen that it is practicable to use the method of random sampling for estimating the district mean yield with reasonable accuracy. In particular, the results show that by sampling about 45 villages in a district, distributed in proportion to the area under paddy in the different taluks, and three fields in each selected village, the average yield of the district can be estimated with a sampling error approaching five per cent. The amount of sampling required is considerably less if sampling is done separately for early, mid-late and late varieties, but this requires a knowledge of the area figures under these varieties.

Experience shows that the charge of 45 villages is too large for the existing staff of the Department of Agriculture, for the staff consists on the average of only one Agricultural Assistant per taluk. If, however, experiments are distributed between the staff of the Departments of Agriculture and Revenue, as appears to be the practice at present, and further the work is so distributed that independent samples from the same area are assigned to the staff of the two Departments, then, not only will it be possible to manage the work in the limited time available for harvesting, but we shall also have two independent estimates of the yield per acre for each district. These data might well form the basis for ascertaining the care with which the field work is done. If an agricultural agency is not available, the only alternative available is to entrust the work to the staff of the Department of Revenue in a way that it will fit into the existing machinery of the Department. Whether the work is distributed between the Departments of Agriculture and Revenue, or is exclusively carried out by the staff of the Revenue Department, the field work will have to be so organised that it will fit in with the administrative arrangements of the department by assigning the field work within each Circle to the respective local Revenue staff.

Under the existing official procedure, the yield per acre is estimated by multiplying the normal yield by the condition factor. The normal yields are, however, based largely on considerations of general experience, and as such are purely arbitrary. There is no means of knowing the extent to which they are biased. The condition factor is a subjective estimate of the crop in terms of the normal, and is subject to unknown errors. The advantage of estimating the yield per acre by the method of random sampling is that the subjective

element is altogether eliminated from the process, and what is even more important, the precision of the estimated yield is known. The work will naturally require strong statistical direction, for planning the surveys, for analysing the results and for incorporating the results in successive year's plans and also for guarding against the possibility of the random sampling method degenerating into inefficient routine by ensuring adequate supervision of the field work. It is gratifying to note that the Government of Bombay have appointed a Statistician to carry out the work on the lines indicated in this paper for estimating the yield not only of paddy but of all crops in the provinces.

Acknowledgments

We wish to place on record our grateful thanks to Sir William Jenkins, Director of Agriculture, and to Mr. Davis, Collector of Kolaba, for their encouragement and keen interest in the scheme. We also wish to record our thanks to the staff of the Department of Agriculture for their co-operation. On the statistical side our thanks are due to Mr. M. A. Telang, who carried out the statistical analysis under the guidance of the senior author.

References

Sukhatme, P. V. .. *Proc. Ind. Acad. Sci.*, 1945, **21**, No. 6, Sec. B.

——— .. *Nature*, 1944, **154**, No. 3905, 299.

Sukhatme, P. V., Kadam, B. S., and Bhat, N. R. .. *Ind. Jour. Agri. Sci.*, 1945 (in the press).

Cochran, W. G. .. *Jour. Amer. Stat. Assocn.*, 1939, **34**, 492–510.

Panse, V. G., and Kalamkar, R. J. .. *Curr. Sci.*, 1944, **13**.