Reprinted from Sankhyā, Indian Journal of Statistics : Vol. 7, Part 3, 1946.

SAMPLE SURVEYS OF CROP YIELDS IN INDIA

By P. C. MAHALANOBIS
Statistical Laboratory, Calcutta.

HUBBACK'S EARLY WORK IN BIHAR AND ORISSA.

An account of early crop-cutting work in India carried out in 1923, 1924 and 1925, was given in an important paper on *Sampling for Rice Yield in Bihar and Orissa* by J.A. (later Sir John) Hubback of the Indian Civil Service (now retired) which was received for publication on 10 November 1926 and was published in 1927 by the Government of India as Bulletin No. 166 of the Agricultural Research Institute, Pusa. The method developed by Hubback was used a little later in Central Provinces by C. D. (now Sir Chintaman) Deshmukh, and Mr. P.S.Rau, two other officers of the Indian Civil Service. Hubback's paper also influenced work outside India. Prof. R.A.Fisher in his memorandum dated 2 March 1945 addressed to the Imperial Council of Agricultural Research, Government of India, stated :

> "The use of the method of random sampling is theoretically sound. I may mention that its practicability, convenience and economy was demonstrated by an extensive series of crop-cutting experiments on paddy carried out by Hubback (later Sir John Hubback, Governor of Orissa) more than 20 years ago over a greater part of the rice tracts in Bihar and Orissa. So far as I know these were the earliest crop-cutting experiments based on the principle of random sampling anywhere in the world. They influenced greatly the development of my methods at Rothamsted."

2. The bulletin has been, however, out of print for a long time. With the kind permission of the Government of India, we have much pleasure in reprinting it in the current issue of Sankhyā.

3. I had briefly referred to Hubback's paper in my memoir on Large Scale Sample Surveys (*Phil Trans.* Vol. 231 Series B). In the present note I propose to draw attention to a few salient points leaving it to the reader to read the original paper in full. Page references are everywhere to the present reprint.

4. Hubback's starting point was to criticise the unreliable character of official methods of estimating crop yields which were at that time and are still current in large parts of India. He observed :

> "All that is done at present is for the local police officers to make a guess, at which in succession the Subdivisional Officer, the District Officer, and the Director of Agriculture guess again". (p.282)

He pointed out that in the official method an attempt is made to select typical or representative fields for crop-cutting work which "depends entirely for its accuracy on the ability of the officer to select", and went on to say :

> "The method is comparable to estimating the average income of the population of a town by watching the streets for a few days and then picking out a man, who looked to be in average circumstances and discovering what his income is". (p.283)

Hubback also observed that in the official method "there is no possible way of estimating what is the probability that the result of such selections is within a given range from the true mean yield". He then stated :

> "The only way in which a satisfactory estimate can be formed is by as close an approximation to random sampling as the circumstances permit, since that not only gets rid of the personal element of the experimenter but also makes it possible to say what is the probability that the result of a given number of samples will be within a given range from the true mean". (p.283)

As far as I can trace this is the first explicit enunciation of the principle of random sampling for estimating the yield of crops.

5. Hubback refers to the earlier work in 1921 in which attempts were made to pick up fields at random, but points out that "in this method it was necessary to find out the date on which the crop on which field would be ready for harvest and to arrange to visit the village on that date." The practical difficulties involved in the above approach made Hubback search for another method in which, as he acutely observes, the distribution of sampling should be random in both time and space.

6. For crop-cutting work in India, cuts of the size of one-tenth of an acre have been standard for a long time. Hubback used a special apparatus giving a triangular cut comprising an area of only 1/3200 of an acre. His chief aim in reducing the size of the cut was to enable the investigators to collect the samples without much trouble or fatigue. In fact he noted that it is usually not possible to collect a large number of cuts of the size of one-tenth of an acre.

7. The margin of error of the final result received the closest scrutiny. From a careful analysis of the extensive data Hubback came to the conclusion that cuts of a small size were fully adequate. He observed :

> "It is no advantage to take a large number of samples from places very close together, where the crops will naturally be very much the same on the same day. The degree of accuracy is not seriously improved by such practice. This explains why there is no need to take large samples instead of the handy samples obtained by my method. A sample of one-tenth of an acre is merely 320 of my samples taken in juxtaposition ············ Technically speaking there is very high correlation between the individuals of such groups of samples which makes the ordinary rule, that the standard deviation divided by the square root of the number of samples, quite inapplicable". (p.286)

This gives in a few sentences the central idea of the variance function' as used in our work nearly 10 or 12 years later.

8. Hubback was not satisfied in carrying out experiments in one or two small localities but covered a wide region during the three years 1923, 1924 and 1925. He stated :

> "The system has thus been tried in the Orissa deltaic tract, on the Chota Nagpur plateau of which Ranchi forms a part, in the broken country of Santal Parganas and Manbhum, and in the alluvial Gangetic plain of North Santal Parganas and Gaya". (p.287)

Hubback observed :

> "The method is applicable to small or large tracts. The same number of centres will probably give a slightly greater degree of accuracy for a small and a slightly less degree for a larger tract, because the standard deviation on the whole increases with the size of the tract, though not at all rapidly. (p.286)

9. It is remarkable that throughout his work Hubback had kept the question of cost prominently in view. With the help of actual experimental observations, he showed in fact that sampling by his method was entirely practicable and would give, at a comparatively small cost, final results with sufficient precision for most practical purposes.

10. Besides crop-cutting experiments, Hubback also made a number of important observations regarding estimates of crop acreage. For example, he emphasized that the harvested area was "evidently not the area planted or sown, since in practically all years some fields bear no crop worth cutting"; and noted that the difference "between the planted (or sown) area and the harvested area in a bad year, when the rains stop early in September, " may well be 50 per cent all over a sub-division". He pointed out that the planted or sown area itself would vary from year to year. He actually observed :

> "It is usually believed that the estimate of cropped area is sufficiently accurate. It may be, but the belief is intuitive rather than rational". (p.287)

and pointed out

> "It would certainly be a great advantage if the harvested area could be obtained by some form of sampling". (p.287)

11. He himself tried a method in 1925 in the Santal Parganas which as far as I know is the earliest attempt to estimate crop acreage by the method of random sampling. He was not satisfied with the method, and he noted that "the field work is strenuous and tedious, and it is probable that it would be in practice shirked, and results fudged". He mentioned other difficulties, but concluded with the observation :"But it is still possible that some method on similar lines may prove practicable". It is interesting to note that he had also pointed out the importance of taking into consideration the area under field ridges (*ails*), and of making reliable estimates for the loss of weight at the stage of husking paddy.

12. In an appendix he gave a detailed investigation into the effect of deviations from "simple" sampling. As far as I am aware, this is the earliest discussion of what would be now called the method of determining the error when the sampling is done in more than one stage. Hubback had also investigated how far frequency distributions of yield rates conformed to the normal curve.

13. For a proper appreciation of the real significance of Hubback's work, it must be remembered that it was so late as in May 1924 that the Bureau of the International Institute of Statistics appointed a commission for the purpose of studying the application of the representative method of statistics, and that this report was presented in 1925 and published in 1926. In this report considerable prominence was given to the method of purposive selection, but the importance of method of random sampling had not been fully realised. As already noted, Fisher's own work at Rothamsted had been directly influenced by Hubback's paper. Hubback's work was also the starting point of much of the work done by the Indian Statistical Institute in Bengal and Bihar since 1938.

14. Against the historical background sketched above, Hubback's work must be considered to be a most remarkable achievement. The paper itself is quite small and gives

in a short compass a masterly discussion of the subject. It truly constitutes a classical paper which deserves careful study even after the lapse of 20 years.

DESHMUKH'S WORK IN CENTRAL PROVINCES.

15. Immediately after the publication of Hubback's paper in 1927 similar work was taken up in 1928 in Raipur district of the Central Provinces of India by C.D. (now Sir Chintaman) Deshmukh, Governor of the Reserve Bank of India and at present President of the Indian Statistical Institute, and was continued in 1929 and 1930. A brief account of this work with special reference to applications in the assessment of land revenue was given by Deshmukh in his report on the Settlement of the District of Raipur, 1926-1931 published by the Government of Central Provinces. A little later Hubback's method had been used by Mr. P.S.Rau, I.C.S., for the sampling of rice yield in Bilaspur.

16. At the time of my visit to Nagpur during the annual session of the Indian Science Congress in January 1945 I happened to be staying with Mr. Rau. At my request he very kindly traced the original papers and sent me 21 books containing original entries out of which 19 books relate to Deshmukh's work in Raipur, and 2 books to Rau's work in Bilaspur. After a good deal of search in Government offices I was informed that the bulk of the material containing the observations made in Bilaspur district had been unfortunately destroyed some considerable time ago.

17. Deshmukh's material consists of 1517 cuts in 1928, 1447 in 1929, and 1920 in 1930 making up a total of 4884 crop cutting experiments altogether. (I may note in passing that the figure 4384 for the total number of cuts given in paragraph 93 on p.28 of the Raipur Settlement Report is evidently a printing mistake which had almost certainly arisen from the substitution of the figure 3 for the correct figure 8.) In the original record, entries were available relating to the name of the village, the variety of paddy, irrigation factors etc. The material is being analysed in the Statistical Laboratory and a separate note will be published later. The village entries make it possible to carry out a two-stage analysis. At my request Mr. C. R. Rao undertook this work, and has supplied the following material.

18. Let σ_1^2 be the variance between the average rates of yield per sample-cut of villages, and σ_2^2 the variance of the rates of yield per cut within a village, and let n_1 be the number of villages included in the survey from each of which n_2 fields are sampled. The estimated variance of the average rate of yield per cut for the whole region is given by

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2}$$

The following table gives the values of σ_1 and σ_2 and the mean yield in tolas per cut based on the material collected by Deshmukh in Raipur in 1928, 1929, and 1930.

TABLE 1. ESTIMATED VALUES OF MEAN YIELDS AND STANDARD DEVIATIONS : RAIPUR 1928-1929.

year	number of cuts	estimated values of σ_1	estimated values of σ_2	mean yield per cut
1928	1517	5·24	6·06	12·82
1929	1447	2·73	7·34	17·88
1930	1920	4·96	6·36	14·06

19. From the above data it is possible to determine, for any given number of cuts per village, the number of villages which would have to be sampled to reduce the percentage

error of the estimated mean yield to any assigned value. The following Table 2 gives relevant figures for a percentage error of 3%.

TABLE 2. ESTIMATED NUMBER OF VILLAGES AND SAMPLE-CUTS (PERCENTAGE ERROR=3%)

year	number of villages requiring to be sampled for				
	1 cut	2 cuts	3 cuts	4 cuts	5 cuts
1928	434	314	269	248	236
1929	214	120	89	73	64
1930	366	253	214	196	184

20. The important point to be noticed is that there is a substantial reduction in the number of villages to be sampled when the number of sample-cuts is increased from 1 to 2 or 3. With small cuts of the size used by Hubback, Deshmukh and Rau (or by the Indian Statistical Institute more recently) there is no difficulty in collecting a number of cuts in the same village or in a group of adjoining villages. The use of sample-cuts of a small size is thus distinctly economical.

21. Another point is also worth stressing. From Hubback's own data and also from the calculated values of the variances based on Deshmukh's material it is clear that only a moderate number of sample-cuts would be quite sufficient to give reliable estimates of the mean values of the rate of yield per acre for a whole province. For example, using Deshmukh's data for 1930 (the year in which variances were highest) it appears that less than 100 villages with one or two cuts in each distributed suitably over the 9 different districts in the Central Provinces would give the provincial mean rate of yield with an error less than 2 per cent which should be adequate for all practical purposes, and would in any case reach the limit of physical fluctuations.

LATER DEVELOPMENTS

22. In 1938 extensive crop-cutting work was done in Bengal (under the general supervision of Mr. H. P. V. Townend, then Development Commissioner, with the technical help of the Statistical Institute) with a square wooden frame including an area of 27.04 sq. ft. or just about double the size of the triangular frame used by Hubback. Since then a good deal of intensive studies of crop-cutting work have been made by the Indian Statistical Institute. The chief objects of these investigations have been four-fold:

(1) to settle the standard size of the sample-cut and appropriate specifications for actual harvesting in the field;

(2) to settle the number and distribution of sample-cuts over the region under survey in order to attain any desired degree of precision in the final results;

(3) to study the most efficient organization of the human agency including the control of recording mistakes;

(4) to study possibilities of using auxiliary measurements having correlation with the final yield of dry grain in the case of cereals and dry fibre in the case of jute.

23. *Size and shape of sample-units.* During the last 5 or 6 years extensive experiments have been made on sample cuts of various shapes and sizes, and using many different procedures among which may be mentioned the following: Hubback's triangular frame of size 13.6 sq. ft.; Townend's square frame of size 27.04 sq. ft.; cuts of square and rectangular

shapes demarcated by ropes and pegs and of various sizes ranging from 1 sq. ft. to about 1/9th of an acre ; cuts of circular shape with radius of different lengths such as 2′, 4′, 5′-7½″, 8′ etc. harvested with the help of an instrument first designed by Mr. J. M. Sen Gupta of the Statistical Laboratory and then improved in various ways. The circular cut is being used at present as a standard procedure in Institute work.

24. An important question is whether the use of cuts of a small size is fully representative and unbiased. This has been and is being studied by the Institute practically every year since 1939. From the theoretical point of view, cuts of a large size would of course lead to a relative under-sampling of the portion of the field near the boundaries. The undersampling arising in this way varies roughly as the ratio of the size of the cut to the size of the whole plot assumed to be of a square shape. This is a sampling bias, and is to be carefully distinguished from the bias, if any, which may arise from the personal equation of the investigator in the actual process of harvesting the cut which is the point under consideration here.

25. Broadly speaking, it was found by the Institute that with cuts of a very small size, when these are demarcated by pegs and ropes, there is a tendency for over-estimation of the rate of yield per acre. It was found however that this tendency decreases as the size of the cut is increased, and practically vanishes beyond about 50 sq. feet under normal conditions. It is possible that when a small cut is demarcated by pegs and ropes there is tendency for the investigator to be over careful and to pull in a few of the bordering plants which should properly be allowed to fall outside the rope.

26. With circular cuts of a very small size there is some evidence of a small bias sometimes creeping in which may however be both positive (over-estimation) or negative (under-estimation) depending on the investigator. On the whole, when the work is done by a large number of different investigators, these tend to cancel out so that results obtained with the circular cut, as far as can be judged from evidence at present available, appear to be without bias. Another advantage of using cuts of a circular shape is that no bias can arise due to varying directions along which the crop is sown.

27. *Variance function.* The variance function, that is, the decrease in variance with increase in the size of cuts, has been studied in great detail in the Institute. As pointed out by Hubback, there is comparatively little gain in precision by increasing the size of the cut owing to the existence of positive and usually high correlation in the yield of plants in adjoining portions of the same field. This has been fully borne out by extensive work done by the Institute in recent years. From the point of view of precision, nothing is to be gained, therefore, by using cuts of a large size. As pointed out by Hubback, cuts of one-tenth of an acre usually used in Government Agricultural Departments, are entirely unnecessary.

28. A far larger number of cuts of a small size can be harvested in the same time. Deshmukh's work in 1928-29-30 had shown (and has been corroborated by subsequent work by the Statistical Institute) that 4 or 5 sample-cuts collected in the same village or in a group of adjoining villages would give results of about the same precision as single cuts collected from about twice the number or more of villages. As the greater part of the total cost arises from journeys which have to be undertaken from one village to another, it is clear that a reduction in the number of villages (made possible by using cuts of a small size) would mean much less expense when the work is done by a field staff moving from village to village. If the sample-cuts are harvested by stationary investigators (who do the work in the near neighbourhood

of their normal place of residence) even then each such investigator would be able to collect in the same time a much larger number of sample-cuts of a smaller size which would lead to a distinct gain in precision.

29. Available evidence thus distinctly indicates that sample-cuts of a small size would be more economical. Another important consideration is the greater ease and convenience with which such cuts can be harvested, especially with the circular implement, as this is bound to improve the quality of the field work.

30. *Density and total number of sample-units.* As already noted, since Hubback's time experimental evidence has continuously accumulated showing that coefficients of variation for cuts of a small size are quite small and of the order of only 50 per cent. Two to three thousand sample-cuts properly randomized in both space and time should thus, in theory, supply a mean value of the yield per acre with a precentage error of the order of 1 per cent which would be amply sufficient for all practical purposes. The use of cuts of a comparatively large size is thus entirely unnecessary so far as precision is concerned.

31. A second point had also emerged from Hubback's early work, namely, that variances or coefficients of variation for large tracts are only slightly greater than the variance or coefficients of variation for much smaller tracts into which the whole region may be sub-divided. This has been also borne out by Deshmukh's data, and by other material collected in the course of more recent surveys. The same number of sample-cuts distributed at random over a single district, or over the whole province would thus supply mean values with percentage errors which would not be very different. The cost of field operations would thus depend almost directly on the number of districts or regions for which separate estimates have to be obtained.

32. *Recording mistakes.* As already noted, considerations of cost make it necessary to use multi-stage sampling in crop-cutting work. Units at the higher stages, namely, zones, cells, villages or fields can be and are selected at random in advance in the Statistical Laboratory. The location of the sample-cut itself within the field in the final stage must however be done by the investigator; the sample-cut has also to be harvested, threshed and weighed by him. In the survey of crop acreage, it is possible to have the same field inspected by two or more independent investigators. The same sample-cut can never be harvested more than once. Possibilities of the occurence of unconscious bias as well as of gross negligence or deliberate dishonesty are therefore much greater in crop-cutting work.

33. This indeed has been found to be the greatest obstacle, under working conditions in Bengal and Bihar, to the securing of reliable estimates of rates of yield of crops per acre and the total outturn of crops. This problem has not yet been solved, but is being continually studied, and attempts are being constantly made to secure unbiased primary records in crop-cutting work. The method of independent net-works of samples is of course of great value in supplying an over-all control. Various other methods are also being tried and are being used with different degrees of success. A brief description of some of these methods is given below:

34. *Interpenetrating net-works of samples* (replication).*—In crop survey work the Indian Statistical Institute has been using for the last 6 years replication in the form of two independent net-works of sample-units. In each zone or region (within which the act of random location of sample-units is separately completed) the sample-units, as they are located, are numbered consecutively. It is thus possible to divide the whole set into two (or more)

*This phrase is due to Prof. R. A. Fisher.

independent but random sub-sets. For example, the odd numbered sample-units would form a random set as also the even numbered sample-units. These two sets would be interpenetrating and would each cover the whole zone or region. Informaton for each sub-set is collected by a different party of field investigators who work under different inspecting staff. When two such interpenetrating net-works of sample-units are used, it is possible to obtain two independent estimates (of crop acreage or of crop yield) for each region. The difference between these two independent estimates immediately supplies a good idea of the effective margin of error and thus indicates to what extent the survey had been carried out under statistical control.

35. *Control through variance function.* In crop-cutting work (for estimation of crop yields per acre) another kind of control has been found most useful. In the Institute method the standard procedure is at present to obtain cuts of a circular shape of different sizes harvested by a compact instrument which implement consists essentially of a folding arm which can be opened out to various lengths (2′, 4′, 5′-7½″, 6′, and 8′). A pivot with a stump about 12″ long is driven vertically into the ground at a point selected at random, and the arm is placed on a socket arrangement on the stump in such a way that it can swing freely in a horizontal plane (i.e. in a direction perpendicular to the pivot) and describe circles of various sizes. A small pointer is attached at right angles at the free end of the arm, and as the arm is rotated, all plants falling inside the pointer are harvested. The cut of the smallest size is naturally first collected, and the grain threshed and weighed. The arm is then drawn out to the next larger radius and the additional crop in the annular strip is harvested, threshed, and weighed. The same process is repeated for circles of larger size. In this way the weight of grain for 3 or 4 different sizes of cuts (each of circular shape and with centres at the same point on the ground) are obtained.

36. A comparison of mean values based on different sizes of cuts shows whether or not any bias had arisen in using different sizes of cuts in the process of harvesting. Cuts of different size would, however, in general show decreasing variance with increase in size. A comparison of variances for cuts of different sizes thus supplies another valuable control. For example, it is sometimes found that mean values based on cuts of different sizes are in excellent agreement (often much better than one could expect from probability considerations), and variances of yield rates for cuts of different sizes are also about equal. In such cases it is practically certain that cuts of only one size had been actually harvested and records for cuts of other sizes were obtained by paper calculations. Multiple cuts of other shapes and sizes have also been used in crop-cutting work, and found to be quite useful.

37. In crop-cutting work it is practically always necessary to use a multi-stage design as already explained. The selection at random of units at the higher stages, namely, zones, cells, villages or fields is usually done beforehand in the Statistical Laboratory. The selection of the sample-cut itself, at the final stage, must however be done by the investigator on the spot. For this purpose tables of random numbers are supplied and the investigator is required to move an appropriate number of steps along and across the field according to the random numbers picked up in serial order from the table. Sometimes the investigator is asked to turn to the right or to the left or to get a particular segment out of a number into which the field is divided, depending on whether the toss of a coin shows head or tail. Each investigator, in such cases is required to keep a record of successive throws of heads and tails ; and as fairly long series are usually obtained for each investigator, it is easy to test whether the heads and

tails form a random series or not. If, as is sometimes found to be the case, the series is significantly non-random it is reasonable to think that the field work was not done properly. It is true, of course, that coin tossing may have been done properly but not the field work. Still it is often useful to have even such partial checks.

38. *Field organization.* In crop-cutting work the great snag is that harvesting starts at different times in different regions and even in different villages within the same region.. Once started, the harvesting is often over very quickly, sometimes in the course of a fortnight or so. Another difficulty, which is peculiar to paddy, is that different types of paddy are sometimes grown in the same region or in the same village, so that harvesting of paddy of one type or another may continue in the same village from June or July till December or January. In fact in Bengal, harvesting of paddy goes on practically in every month of the year in one part or another of the province.

39. A well organized staff of *patwaries* or village revenue officers exist in the temporarily settled provinces in India but no such staff is available in the permanently settled provinces like Bengal and Bihar. In the Bengal Crop Survey the crop-cutting work is therefore done by the investigators engaged in the acreage survey. The total number available being only 300, the investigators have to move from one village to another in accordance with a fixed programme. Owing to large fluctuations in the date of harvesting, it is inevitable that the investigators should sometimes find that the crop has already been harvested before they reached a particular village or that the crop would still require a number of days to be ready for harvesting at the time of their arrival in any particular village or locality. In this situation the standard programme settled beforehand has to be modified in many instances. In theory, this might take away somewhat from the ideally perfect randomization of the sample. In practice, this appreciably reduces the number of sample-cuts which can be actually collected at the proper time.

40. After struggling with the problem for many years it is becoming clear that crop-cutting work to be done properly must be carried out by a comparatively large number of investigators who would watch the crop as it grows, and collect the sample-cuts at the right time from fields situated in the neighbourhood of their normal places of residence. The total volume of work required from each investigator would not be large, and the work can be easily done as a part-time occupation. With the circular cuts of the size used by the Institute it is possible to harvest up to 5 sample-cuts (each consisting of a set of 3 or 4 multiple-cuts of different sizes) in the course of a day. One investigator working in a group of 6 or 8 adjoining villages can easily collect 20 or 30 sample-cuts in the course of a total working time of 6 or 7 days. If a staff of two or three thousand part-time workers were available scattered in a suitable manner over the province it would be quite sufficient if each of them did crop-cutting work for about a week in each crop season. This would supply mean rates of yield per acre for individual districts with a percentage error of the order of 1 or 2 per cent.

41. There are only two ways in which the help of a sufficiently large number of investigators can be secured for crop-cutting work at the proper time. One and probably the best plan would be to include crop-cutting work as an integral part of a comprehensive scheme of multi-purpose sample survey set up throughout the province. In such a scheme a nucleus of whole-time workers would be maintained and spread over the province. A large number of workers would also be employed who would do crop-cutting as well as other sample survey work as a part-time occupation. Employment for crop-cutting work alone would not be

economical, but there would be no difficulty in offering sufficiently attractive remuneration for part-time work to the extent of say 2 or 3 months in the year.

42. The only other alternative is to employ some Government agency already existing in the province such as the Jute Regulation staff in Bengal or the subordinate postal staff. In the temporarily settled provinces there is also the possibility of using patwaries. These are important organizational problems which require to be studied on scientific lines without further delay.

43. *Driage*. The possibility of using auxiliary physical measurements (which have appreciable correlation with the yield of dry grain) has been and is being investigated. For example, in the case of paddy and wheat it has been found that there exists a high and quite steady correlation between the weight of crop immediately after harvesting (called "green" weight) and the weight of crop after drying. There are of course, minor fluctuations from one season to another or from one region to another in the same season. Arrangements are therefore made, in about 10 per cent of the sample-cuts, to obtain the weight of crop after drying as well as the weight immediately after harvesting. On the basis of the material collected in this way it is possible to work out reliable methods for estimating the weight of dry grain from the much more extensive material for green weight.

44. *Mixed crops*. In certain parts of India, especially in the rabi (or winter) season a number of different crops such as barley, wheat, gram and/or other pulses are grown on the same field. The situation is further complicated by the fact that the relative proportion of different components of the mixed crops vary widely from field to field. Special methods had to be developed for estimating the total production of each component crop in such cases. Incidentally, it may be observed that relatively a much higher yield is obtained when a particular crop is sown in comparatively thin proportions. Full details have been given in a report on the crop survey in Bihar in the *rabi* season of 1943-44. in *Sankhya*, Vol. 7, Part 1, 1945.

45. *Husking*. In the case of paddy the weight of clean rice after husking is conventionally accepted to be two-thirds of the weight of dry paddy before husking. This ratio has however been found to fluctuate from region to region and from season to season. Arrangements are therefore made to measure directly the husking ratio at a number of places scattered over the province. Evidence is accumulating that the driage as well as the husking ratio depend on conditions of cultivation. These are being gradually studied.

46. *Jute*. In the case of jute, in the same way, it has been found that the weight of the jute plants measured immediately after harvesting has a high correlation with the weight of dry fibre extracted after retting and drying the plants. Here also, in crop-cutting work, the weight of plants in the sample-cut is recorded immediately after harvesting, and the extraction ratio is experimentally studied in the case of a small proportion of sample-cuts.

47. *Cinchona*. Possibilities of using physical measurements on living plants are also being explored in many cases. For example, in the case of cinchona plants it has been found that it is possible to forecast the total yield of bark with sufficient precision from physical measurements of girth of the plant at a height of six inches above ground, height of the plants and number of stems on living plants.

48. *I.C.A.R. Scheme*. I may at this stage make a brief reference to the crop-cutting work by the Imperial Council of Agricultural Research of India which was started on the wheat crop in the United Provinces in 1944, and has since then been extended to both paddy and wheat in other provinces. No general account of the work has been published, but I have

receivcb form the I. C. A. R. preliminary reports relating to two or three provinces. It is not possible therefore to make any detailed comments, but a few general observations would not be out of place.

49. The first point to be noted is the use of random samples. There is, of course, complete agreement between the I. C. A. R. and I. S. I. methods in this respect. It is a matter for congratulation that through the efforts of the I. C. A. R. the principle of random sampling is likely to be adopted in practice throughout India at an early date.

50. There are however certain important difference in the methods used by the Indian Statistical Institute and the I. C. A. R. which may be broadly grouped under four heads :

(1) *Size of the sample-unit:* In the I. S. I. method, the sample-cuts are comparatively small in area usually ranging between about 50 and 100 sq. feet. The I. C. A. R. sample-cuts are of a much larger size, of the order of 900 sq. feet or more.

(2) *Harvesting procedure.* I. S. I. sample-cuts are of circular shape obtained with the help of the special implement described earlier, while I. C. A. R. sample-cuts are of square or rectangular shape demarcated by pegs and ropes or chains.

(3) *Use of sub-samples as control.* In the I. S. I. method the crop is harvested in the form of three concentric sub-samples and the yield from each sub-sample is separately weighed and recorded. The variaton of yield between such sub-samples supplies an excellent internal control which is not available in the I. C. A. R. method.

(4) *Use of interpenetrating net-works of sample-units.* As already mentioned, in the I. S. I. method great importance is attached to arranging the field work in the form of two interpenetrating net-works of sample-units and collecting data for each such set of sample-units by entirely different parties of investigators. No such replication is available in the I. C. A. R. method.

51. The Institute view has been that the method developed by it has distinct advantages. It has good statistical controls at the point of collection of the data, and is also less expensive when the work is done through the same type of human agency. It is difficult however to make direct comparisons between relative efficiencies of the two methods for the reason that these have been used in different parts of India. The I. S. I. method has been used mostly in the two provinces Bengal and Bihar where the land revenue is permanently settled and where no revenue or other village staff under Government control is available for doing the crop-cutting work. The I. C. A. R. method on the other hand has been so far used in temporarily settled provinces like the Punjab or U. P. which all have a well organized staff of *patwaries* or village officials under Government control whose part-time services are available for doing the crop-cutting work.

52. There has been some controversy in India regarding the relative advantages and disadvantages of the I. S. I. and the I. C. A. R. methods for estimating crop yields. As I happen to be closely associated with the I. S. I. method , it wouid not be advisable to make any detailed comments until full reports of I. C. A. R. work become available. I may mention, however, that for some considerable time I have been pressing on I. C. A. R. authorities the need of carrying out crop-cutting work by both I. S. I. and I. C. A. R. methods in the same region with a view to studying the relative efficiencies of the two systems. I have submitted definite schemes for this purpose which are under consideration by the Government of India.

JOINT OR DIRECT ESTIMATION OF CROP PRODUCTION

53. Since 1938 I had been advocating the sample survey of crop acreage and crop-cutting work being taken up jointly, but the opportunity occurred only several years later to carry this into practice. A brief description of the method as actually used during the survey of paddy in 1945-46 in Bengal would explain the procedure. The number of grids in which crop-cutting work is done is necessarily much smaller than the number in which the acreage survey is carried out. In 1945-46, about 7000 grids were selected at random for crop-cutting (but in a suitably zoned form), and a second survey of the area actually under paddy at the time of harvesting was carried out. In this way the total acreage under paddy within each grid was known. Multiplying by the rate of yield of paddy per acre (as determined by crop-cutting work carried out in the same grid) it was possible to calculate the total production of paddy per acre for each individual grid. Taking the average within each cell, the total production for each cell could be then directly calculated, from which total production for larger tracts or for the province as a whole could be easily built up.

54. In 1945-46 in Bengal the total production estimated directly in this way was about 6.52 million tons of rice (not in husk). The corresponding estimate obtained by multiplying the acreage by the mean rate of yield was about 6.73 million tons. Incidentally, it may be mentioned, that the second survey at harvest time indicated an appreciable (about 6 per cent) decrease in the area harvested as compared to the area sown.

55. *Conclusion.* A good deal of work has been done and some progress made in India in developing an efficient sampling technique for the estimation of crop yields and total production of crops. A good deal, however, still remains to be done. The active co-operation of statisticians working in different parts of the country is essential for rapid advance.

Paper received : 20 March 1946.