CONFIDENTIAL

INDIAN CENTRAL JUTE COMMITTEE.

Proceedings of the Seventh Meeting

OF THE

Jute Census Committee

HELD ON

the 13th December, 1939.

Calcutta:
PRINTED AT THE STAR PRINTING WORKS,
30, SHIBNARAIN DAS LANE.

1940

ADDENDUM TO
REPORT ON THE SAMPLE CENSUS OF JUTE IN 1939.

(Pages 6-150 of the Proceedings of the **Seventh Meeting** of the Jute Census Committee).

Through oversight the constant factor 5.3020 was left out in writing down the values of Chi-square in Table Nos. (15) to (21) and (25.1). The corrected values of Chi-square together with the corresponding probability of occurrence are given below:-

Page.	Table.	D.F.	Chi-square.	P.
87	15(B)	7	4.4028	0.7315
88	16(A)	18	69.5464	0.0000
89	16(B)	7	12.2136	0.0945
90-91	17(A)	9	17.9254	0.0362
92	17(B)	7	26.5294	0.0004
93	18(A)	9	6.3852	0.7003
94	18(B)	7	1.9045	0.9632
95	19(A)	9	11.4899	0.2451
96	19(B)	7	10.3192	0.1726
97	20(A)	9	10.8871	0.2842
98	20(B)	7	3.3498	0.8482
99	21(A)	9	9.6533	0.3806
100	21(B)	7	6.7430	0.4574

(P.T.O.)

Page 112 - Table (25.1).

**Variance Function:
Graduated Parameters and Goodness of Fit.**

Thana.	Comparison between: Different Samples		Comparison between: Graduated and observed values	
	x^2	P*	χ^2	P+
	(7)	(3)	(9)	(10)
Iswarganj.	-	-	4.40	0.7315
Tejgaon.	69.55	0.0000	12.21	0.0945
Laksam.	17.93	0.0362	26.53	0.0004
Nandail.	6.39	0.7003	1.90	0.9632
Pirgachha.	11.49	0.2451	10.32	0.1726
Palashbari.	10.89	0.2842	3.35	0.8482
Belkuchi.	9.65	0.3806	6.74	0.4574
Total:	125.90	0.0000	65.45	0.1109

*No. of Degrees of Freedom = 9 in every case excepting Tejgaon for which it is 18.

+No. of Degrees of Freedom is 7 in every case for comparison of Graduated and Observed Values.

The Chi-Square test on the whole shows satisfactory agreement between expected and observed values. The general conclusions of the Report are therefore not affected in any way.

TABLE OF CONTENTS.

		Pages.
1.	Agenda	i
2.	Proceedings of the Meeting	1—5
3.	Papers Circulated	
	(i) Subject No. 2—Appendices I to IV ...	6—158
	(ii) Subject No. 3—Appendix V	159—160
	(iii) Subject No 4—Appendix VI ...	160—161
	(iv) Subject No. 5—Appendix VII ...	161
	(v) Subject No. 6—Appendix VIII ...	162

AGENDA FOR THE SEVENTH MEETING OF THE JUTE CENSUS COMMITTEE TO BE HELD ON THE 13TH DECEMBER, 1939.

1. To confirm the proceedings of the Jute Census Committee meeting held on the 10th October, 1939.

2. To consider the report by Prof. Mahalanobis on the Statistical work on the random sampling experiments on the 1939 crop.

3. To consider the progress so far made in the preparatory field work for the 1940 survey.

4. To consider a proposal to give lump sum grants to the Indian Statistical Institute for expenditure in connection with the work on the statistical side for the Jute Census work in 1940.

5. To consider a proposal to grant a Special pay of Rs. 25 p. m. to the Senior Assistant, Indian Central Jute Committee for doing additional work in connection with the Jute Census Scheme.

6. To consider points which might suitably be discussed with Prof. Hotelling during his visits to Calcutta.

(Sd.) C. R. NODDER,

Secretary

PROCEEDINGS OF THE SEVENTH MEETING OF THE JUTE CENSUS COMMITTEE HELD AT CALCUTTA ON THE 13TH DECEMBER, 1939, AT 11 A.M.

PRESENT :

1. H. A. Luke, Esq., (*In the Chair.*)
2. M. Carbery, Esq., I.A.S., *Member.*
3. Prof. P. C. Mahalanobis, *Member.*
4. P. N. Sen, Esq., *Member.*
5. C. R. Nodder, Esq., *Member and Secretary.*

SUBJECT NO. 1.—*To confirm the proceedings of the Jute Census Committee meeting held on the 10th October, 1939.*

The proceedings, which had been printed and circulated to members previously, were confirmed and signed.

SUBJECT NO. 2.—*To consider the report by Prof. Mahalanobis on the statistical work on the random sampling experiments on the 1939 crop.* (Appendices I, II, III, IV.)

Prof. Mahalanobis referred to Mr. P. N. Sen's points raised on the report and explained that in planning a random sampling survey it was of fundamental importance to divide the area into a number of suitable zones each of which would be as homogeneous as possible in regard to the intensity of cultivation. The only way in which the best size of zones could be settled was by gathering detailed information regarding the intensity of cultivation in different Thanas and Unions. Until that information was available, it was quite impossible to take up the question of the best size of zones. At present they had only district figures as a whole and had they had sub-division and Thana figures that would greatly help them.

Mr. Carbery said he could supply the Sub-Division figures but not the Thana figures.

Mr. Luke said that a new Sub-Clause had just been added to the Bengal Jute Regulation Bill to provide for securing a record of land on which nothing but jute was grown.

Prof. Mahalanobis agreed that this might prove useful later.

With regard to the financial position of the Jute Census survey in 1940 Professor Mahalanobis referred to his letter to the Secretary dated the 10th December, which had been placed before the members. They were glad that the Government of Bengal had agreed to contribute Rs. 62,500 for the work. He pointed out that item No. (*i*) in para. 3 of the Secretary's Supplementary Note No. II read "Allowance to the Statistical Adviser" and suggested that this should read "Contribution for Statistical Advice". This was agreed to.

The Secretary pointed out that the Committee had sanctioned a contribution of Rs. 62,500 for the work, but in Prof. Mahalanobis's letter it had been shown as Rs. 68,500. The Secretary said that this excess of Rs. 6,000 could probably be found from the Committee's funds. A figure up to Rs. 80,000 had been contemplated.

Regarding maps for survey work, Prof. Mahalanobis said that the cost of maps would naturally increase as the area increased. They had budgeted for 17,000 for maps for this year (1940 crop) and as the work increased they would have to spend Rs. 30,000 or Rs. 35,000 on maps for the next year. The Government of Bengal had agreed to contribute Rs. 62,500 for the work, but by way of selling maps they would take away a large sum. Therefore he suggested that the Government of Bengal be approached again with the request that maps be supplied to the Committee free of cost in view of the fact that this was a work of public importance. In other Provinces, he was informed, maps were being supplied free of cost by their respective Governments for similar work. The proposal was unanimously approved.

His second suggestion regarding maps was that in order to give them a longer life they should have them mounted on linen. The cost of mounting would not be much. The price of the maps was six pice each whereas the cost of mounting would be about seven pice only. He had had about 100 mounted as a trial. He was sure that the cost of mounting would more than pay for itself in a few years and proposed that "any savings out of the budgetted amount of Rs. 17,000 be utilised for mounting maps". The proposal was approved.

After further discussion it was unanimously agreed that work should proceed on the basis of the report and the budget for the work on the 1940 crop as outlined in Prof. Mahalanobis's letter was approved.

Mr. Carbery said that the District Agricultural staff would be up to full strength by April and would be available for the Census work from mid march or the end of March.

It was agreed on Mr. Carbery's advice that Nadia should be surveyed before Jessore as the crop there was earlier.

SUBJECT NO. 3.—*To consider the progress so far made in the preparatory field work for 1940 survey.* (Appendix V.)

Prof. Mahalanobis said that so far maps had been received for 29 thanas out of the 220 thanas. By the end of the current month he would know better how far the position in that respect was satisfactory. He further said that last year they had arranged for a small check survey by the Satistical Laboratory, and suggested that this year also arrangements should be made that certain portions of the area might be surveyed by the staff of the Statistical Laboratory. It was agreed that this should be done and arrangements made in consultation with the Secretary of the Jute Census Committee.

The Secretary said that he had received a suggestion that it might be desirable for some of the Committee members to go into the field to do some checking of the field work.

Prof. Mahalanobis said that they were agreed to the proposal in principle, but the Secretary would have to make necessary arrangements if any member could find time to spare. There might be a question of finding money for travelling allowance.

The Chairman while agreeing to the principle thought that active participation of members was apt to detract from the authority of the superior field staff over the subordinates. The Secretary was there and he would certainly go on inspection tours and check the work going on.

After further discussion it was agreed that the matter should be left until the time of the Indian Central Jute Committee's meeting in February or March, when the Jute Census Committee might also meet.

SUBJECT NO. 4.—*To consider a proposal to give lump sum grants to the Indian Statistical Institute for expenditure in connection with the work on the Statistical side for the Jute Census work in 1940.* (Appendix VI.)

The proposal was unanimously approved.

SUBJECT NO. 5.—*To consider a proposal to grant a special pay of Rs. 25 p. m. to the Senior Assistant, Indian Central Jute Committee for doing additional work in connection with the Jute Census Scheme.* (Appendix VII.)

Prof. Mahalanobis said that he agreed to the proposal in view of the fact that the work of the Jute Census Committee entailed extra, and very careful work.

It was thereupon unanimously agreed that the Senior Assistant (Mr. Kohli) be granted a special pay of Rs. 25 p. m. with effect from the 10th October, 1939 as long as the Jute Census work was being done under the supervision of the Secretary of the Indian Central Jute Committee.

SUBJECT NO. 6.—*To consider points which might suitably be discussed with Prof. Hotelling during his visits to Calcutta.* (Appendix VIII.)

Prof. Mahalanobis informed the Committee that Prof. Hotelling was due to arrive at Calcutta on the 18th December and he proposed that an informal meeting might be arranged on the 21st December in the afternoon when the members of the Jute Census

Committee and officials of the Indian Central Jute Committee could discuss points on which advice was desired. They might then hand over papers to him containing points for discussion, and later in January or early February when he had come back to Calcutta after his South Indian tour, they might have fuller discussions.

The proposal was agreed to.

The Secretary said that he had prepared a set of points to be discussed and requested the members of the Committee to inform him of any matters that they wished to discuss with Prof. Hotelling.

The meeting then terminated.

(Sd.) C. R. NODDER,

Secretary.

APPENDIX I.

REPORT ON THE SAMPLE CENSUS OF JUTE IN 1939.

BY P. C. MAHALANOBIS.

Section 1. Introduction

1. The field work in connection with the Crop Census of 1938 was completed on the 21st October, and I submitted the First Report on the 26th December, 1938. This report was considered by the Jute Census Committee on the 6th February 1939; and administrative sanction was given to the present Scheme on the 9th February 1939. The real object of the Scheme is to develop a sound method for obtaining a reasonable annual estimate of the total area under jute in Bengal. In previous reports I pointed out that two things are essential for success, namely, developing an efficient sampling technique, and building up a suitable human agency for doing the primary work with the required accuracy. It is, of course, necessary that the total cost of the sample survey should be kept within a reasonable limit, and the margin of error should be reduced to a sufficiently low figure to enable the final estimates being used for administrative and business purposes.

2. On the technical side I explained that in order to solve this problem it was necessary to determine the best size of the sampling unit and the best density or the number of such sampling units per square mile in different thanas or zones in order to attain the highest accuracy for any assigned level of total expenditure. For an adequate study of the problem it is, therefore, essential to study what I called the Variance Function (showing the relation between the variability of the sample estimates and the size of the grids), and also the Cost Function giving the cost of operations for working with different size and density of grids.

3. In order to attain the objects we had in view it was decided to carry out sample surveys on the field in 1939 with various combinations of size and density of grids. For convenience of reference, we may call each such combination a "*pattern*" for field work. Owing to the limited funds at our disposal, it was decided to work with 6 different sizes of sampling units, namely, random plots, and grids of size 1, 4, 9, 16 and 36-acre. In order to cover a wide range,

we included 12 different densities in the programme of work for 1939, namely, $\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 3, 4, 6, 8, 16, 9, 25 and 36 sampling units per square mile. It was not feasible to include all possible combinations, and a selection was made of 25 patterns which were considered suitable for present purposes.

4. In order to gain some idea regarding the local conditions in different districts it was decided to conduct the sample survey in 31 thanas in 5 districts, Mymensingh, Rangpur, Pabna, 24 Parganas and Hooghly. The primary object of the sample survey was to collect necessary information regarding the cost of field work for the 25 different patterns included in the programme.

5. It was also decided to include in the programme complete enumeration of about 700 square miles. For a detailed study of the Variance Function it was also decided to carry out model sampling experiments in the Laboratory on the basis of the data relating to complete enumeration which had been already collected in 1938 and also partly on the basis of the new material collected in 1939.

Preparatory Work

6. In the original scheme for the Jute Census work in 1939 I had emphasized the need of starting the preparatory work early in December 1938. As administrative sanction was received so late as 9th February 1939 the preparatory work was seriously delayed which created difficulties. Many workers who had been engaged in the 1938 scheme had been discharged in the meantime, and new workers had to be engaged and trained; fresh arrangements had also to be made for organizing a large staff of about 60 or 70 computers. The whole of the preliminary work had to be done very quickly in order to get the mauza maps ready in time for starting the field work in April 1939.

7. Further difficulties were created by the abnormal and severe conditions of drought which prevailed in the beginning of the season. This delayed the sowing of jute in many districts making it impossible to start the field work at the scheduled time in many areas for which the sample grids had been already prepared. In some cases maps had to be got ready for other areas at short notice. The preparatory work however proceeded on the whole smoothly and in time to enable the field work being conducted without any serious difficulty.

8. The volume of preparatory work done in the Laboratory can be easily appreciated from the fact that over three lakhs and forty thousand random numbers were used in locating the sampling points on the village maps; over forty-eight thousand samples of random plots were constructed; more than twelve thousand grids of various sizes were drawn on the maps; over two lakhs and forty thousand individual plots were listed; and the lists were checked and copied for being despatched to the field. In the case of *char* areas and for certain big plots actual tracings of about 257 maps had to be prepared and supplied together with a list of plots.

Available material

9. It was not possible, however, to complete the field survey in every case owing to drought and other unforeseen difficulties. The amount of work for different patterns which could be actually completed is shown in Tables (1)—(5). In these Tables, col. (1) gives the name of the district and the thana; col. (2) the density of sample grids per square mile; col. (3) the total number of workers who were engaged in the sample survey in each thana; col. (4) the total area in square miles covered by each pattern of the sample survey; and col. (5) the number of gross pay-days taken to complete the work. The serial roll number of the primary investigators are also given for convenience of reference in col. (6).

10. The sample survey was started on the 1st May and continued till the 30th August 1939. The abstract time programme is given in Tables (6) and (7). In many cases there were long gaps in the interval. But on the whole these tables show that the work for each size of grid was well distributed over the whole period; so that there is no reason to think that the time-records for any particular size of grid became biassed through a practice effect owing to the work having been done either much earlier or much later than other sizes.

Arrangements for Statistical Analysis

11. Arrangements were made for a continuous tabulation of the material as they reached the Laboratory. The field survey continued till the end of August, and it was nearly the middle of September before we received all primary records. In the meantime, on the 25th September 1939, I was informed by Mr. Nodder, the officiating Secretary of the Indian Central Jute Committee, that there was some uncertainty regarding the contribution from the

Government of Bengal for the work during the 1940 season, and I was requested to submit a preliminary report for consideration at a special meeting of the Jute Census Committee on the 10th October. As I was obliged to submit my provisional recommendations on the 1st October 1939, the first analysis of the material had to be completed within a very short time. Even at the time of writing the present report we have had less than two months to finish the basic statistical analysis.

Acknowledgements

12. The whole of the statistical work had to be done at high pressure. That I have been able to submit this report in such a short time was due to the willing co-operation of the workers of the Statistical Laboratory. It is difficult to make individual acknowledgements of all the help I have received. The most difficult part of the work fell on Rajchandra Bose and Samarendranath Ray who were responsible for the important calculations relating to the determination of the best size and density of grids. For general supervision, I am indebted to K. Raghavan Nair, K. Kishen and other statisticians of the Laboratory. The bulk of the primary computations was in charge of Jitendra Mohan Sen Gupta, Jitendra Nath Taluqdar, Raja Rao, Sambu Haldar, Pranay Kumar Chatterjee, Dibakar Maity, and Rajendra Chandra Ray. On the administrative side I received much help from Atindranath Bose and Sudhir Kumar Banerjee.

13. I am thankful to Rai Bahadur Nepal Chandra Sen, Director of Land Records, under whose general control the field work was carried out this year for the friendly co-operation between the Field and Statistical Branches. I received all possible help from Mr. Nihar Chandra Chakravarti, Supervisor-in-charge of the Field Branch, who organized the field work under difficult conditions with characteristic zeal, efficiency and ability.

14. I must also place on record my personal appreciation of the unfailing courtesy shown by Mr. Cliff, the Secretary of the Indian Central Jute Committee, in many difficult situations. Without his active co-operation at every stage of the work this report would not have been written at all.

*Contents of the Report**

15. In previous reports I have pointed out that the reliability of an area census depends on two kinds of errors, namely, the *errors of recording* which are due to mistakes committed in observing, recording, copying, etc., at the primary stage of enumeration, and the *errors of sampling* which are inherent in the sample method and are due to the fact that the estimates are made, not from the whole of the material, but from the fraction included within the sample grids. In previous reports I have discussed at length the errors of recording by comparing the results of enumeration by different methods. In the present report I am omitting this portion of the work as the Government of Bengal have already started a big Jute Registration Scheme for collecting detailed information regarding each individual plot which was sown with jute in 1939. The Secretary to the Government of Bengal, Department of Agriculture and Industries, gave an assurance at the meeting of the Jute Census Committee held on the 10th October 1939 that the material collected in connection with the Jute Registration Scheme will be placed at the disposal of the Committee for purposes of comparison. It is desirable, therefore, to postpone this part of the work until the new material becomes available. I have discussed the errors of sampling in Section 2 on the Accuracy of Sample Estimates. It will be seen from the discussion in this section that there was excellent agreement between the different sample estimates (21-26, pp. 12-13).

16. Section 3 deals with the Variance Function, and it is gratifying to note that the form of the Variance Function can be considered to have been practicably settled. It is given by a simple formula of an exponential type which gave excellent results in the case of extensive model sampling experiments in 7 thanas in 5 different districts. The formula is further corroborated by the results of sample surveys on the field. This section also gives, in a summary form, certain results based on the pioneer sample survey of 1935 organized by Mr. Townend. (27-51; pp. 14-23).

17. The Cost Function is considered in Section 4, and it is shown that the graduated values are in good agreement with the

---

*All tabular statements have been given at the end of the Report, and have been consecutively numbered.

observed values of the cost of operations for different patterns of the sample survey. It is encouraging to find that the mathematical equations are of the simplest type and are also such as one would expect to hold good from broader physical considerations. This gives us confidence in using these equations for subsequent calculations (52-80, pp. 24-34).

18. Section 5 deals with the crucial question of the best size of the grid. The available evidence definitely shows the superiority of grids of small sizes. It is found that for general purposes it will probably be most economical to use 4-acre as the standard size; and it is recommended that 4-acre grids should be used in all districts in 1940. It is possible, however, that the best size may differ slightly in different regions. It is, therefore considered desirable to collect some additional information regarding grids of size 2, 6 and 8-acre in addition to the standard size of 4-acre. (81-97, pp. 35-39).

19. The other important question of the best density of grids is considered in Section 6. A method has been worked out for finding the best density in different districts when the districts to be surveyed and the total expenditure for the sample survey is once fixed. This method is illustrated by a concrete example of a survey in 8 districts at a cost of one lakh of rupees. It would be seen, therefore, that the statistical technique and mathematical apparatus have been advanced very far. In fact once the administrative policy is settled regarding the area and total expenditure it is possible to calculate the best density of grids in different regions as well as to obtain the expected margin of error of the final estimates of area under jute. (98-108, pp. 40-44).

20. Section 7 finally deals with the programme of work for 1940 and contains the recommendations which logically follow from the present report. (109-126, pp. 45-52). In case this programme can be carried out successfully in the field in 1940, we have every reason to hope that we shall be able to make plans for and undertake a sample survey of the area under jute in Bengal on a full provincial scale in 1941.

Section 2. Accuracy of Sample Estimates.

21. In previous reports I have shown that estimates of the proportion of land under jute obtained by sample surveys conducted on the field as well as by model sampling experiments in the Laboratory agreed satisfactorily with the results obtained by complete enumeration in the field. Further work has been done in this connection and the results are given in the two Tables (8) and (9) which show the mean proportion of land under jute for completely enumerated areas. Column (1) gives the size of the grid; column (2·1) the number of grids used in the model sampling experiments in each case; column (3·1) the number of grids used in the sample surveys in the field; column (2·2) the mean proportion of land under jute together with the standard error as obtained from model sampling experiments; while column (3·2) gives the corresponding proportions based on sample surveys in the field for various sizes of grids. The actual proportion of land under jute as obtained from complete enumeration is shown at the bottom of the Table in the case of each thana.

22. A glance at the figures given in Tables (8) and (9) will show that in most cases there is satisfactory agreement between the "true" value of the proportion of land under jute (as obtained from complete enumeration) and the different estimates obtained from sample surveys. A more exact comparison has been carried out, as in previous reports, with the help of the "t" – statistic.† The observed values of the "t" – statistic are given in a consolidated form in Tables (12) and (13) in which column (1) gives the size of the grid, and the other columns show the value of the "t"-statistic together with the corresponding degress of freedom for model and field surveys in each case. Following the usual convention, observed values of the "t"-statistic which exceed the expected values on the five percent level of significance have been marked with a star(*).

23. The mean proportion of land under jute together with the standard error based on random sample surveys in the field is shown in Table (10) and (11) in which col. (1) gives the name of the thana; columns (2·1), (3·1), and (4·1) the number of grids used in each

---

†The use of which I have explained in considerable detail in Section 4 (Accuracy of Sample Surveys) and Section 5 (Sampling Experiments on the Field in 1938) in paragraphs 21-58 of the "*First Report on the Crop Census of 1938*" submitted in December, 1938.

case; and columns (2·2), (3·2) and (4·2) the observed values of 'p' together with standard errors.

24. The values of the "t"-statistic for comparison between results obtained with different sizes of grids are shown in Table (14). In this Table column (1) gives the name of the thana; columns (2·1), (3·1), (4·1) and (4·2) the sizes of grids compared; columns (2·2), (3·2) and (4·3) the corresponding number of degrees of freedom; and finally columns (2·3), (3·3), and (4·4) show the values of the "t"-statistic.

25. It will be noticed from Tables (12), (13) and (14) that the 't' values are small showing in the case of Tables (12) and (13) that the differences between sample estimates and true values of the proportion of land under jute, and in the case of Table (14), the difference between sample estimates based on different grid sizes, are usually negligible in comparison with corresponding standard errors. In fact we find that out of 63 cases of model sampling experiments against complete enumeration the differences are statistically significant at the five per cent. level in only 6 cases. For sample surveys on the field against complete enumeration 3 out of 27 cases and for sample surveys on grids of different sizes compared amongst themselves only 13 out of 84 are significant at five per cent. level.*

26. On the whole we may conclude that the results of sampling experiments in the field as well as in Laboratory are in satisfactory agreement with the "true" values as obtained by complete enumeration. A systematic excess of significant differences suggest however that it would be safer to use a higher margin of error than that given by normal theory.

---

*It will be seen that over ten per cent of the observed values appear to be statistically significant while the expected proportion of significant values is only five per cent. In previous reports also we had found a similar excess in the number of significant differences. It must be remembered that the critical levels of the "t"—statistic have been calculated on the assumption of a normal (Gauss-Laplacian) distribution and equal variances. In actual fact the law of distribution or the variances in the present case may be different. It will be advisable, therefore, to use a safety factor of 2 in giving the confidence intervals at the present stage; and this is the procedure which has been followed in subsequent calculations.

Section 3. The Variance Function.

27. In previous reports I have discussed from time to time the nature of the "Variance Function" or the relation between sample variability and the size (or area in acres) of grids. A good deal of additional information has become available this year from extensive model sampling experiments in the Laboratory and also from actual sample surveys carried out on the field. I am giving below a consolidated report of the work done so far.

Available Material

28. The available material consists broadly of three series:—

(*a*) Model sampling experiments conducted in the Statistical Laboratory based on complete enumerations in 1937 and 1938.

(*b*) Sampling experiments carried out in the field in 1938 and 1939.

(*c*) A sample survey on a provincial scale carried out in 1935 under the direction of Mr. Townend with grids of size 40-acres.

29. In the previous report I used a simple exponential equation for the variance function which may be written in the following form:*

$$v_x = pq/(bx)^g \quad \ldots \quad \ldots \quad (1)$$

where v_x is the variance or square of the standard deviation for sample grids of size x-acres, "b" and "g" are statistical parameters; and "p" is the proportion of land under jute and q=(1-p). It is convenient to use the same equation in the logarithmic form given below:

$$\log v_x = \log(pq) - g.\log(bx) \quad \ldots \quad (2)$$

*In order to prevent confusion it may be mentioned here that the notation in this equation has been slightly changed. In equation (1) of para 14 of the "Second Report on the Crop-cutting Census of 1938", I gave the formula in terms of the Standard deviation, and used "k" as the symbol for statistical parameter. It is more convenient, however, to give the formula in terms of the variance or the square of the standard deviation. The parameter "g" used in the present report is thus numerically equal to twice "k", where "k" is the symbol used in previous reports.

Model Sampling Experiments

30. In the previous report I showed that the above function was corroborated in a satisfactory manner by model sampling experiments in the case of Iswarganj thana in Mymensingh. The work has been now extended to a number of other thanas; and the results are shown in Tables (15)—(21). In some of the thanas the model sampling experiments were carried out in instalments; and we have two independent samples in the case of thanas Nandail, Palasbari, Pirgacha, Belkuchi and Laksam; three independent samples in the case of Tejgaon; and only one sample in the case of Iswarganj. By pooling these two (or three) samples we obtained a combined sample for each thana which would naturally have more stable values of the variances. Using these combined samples the two statistical parameters "g" and "b" were next estimated for each thana by least square methods with the help of the logarithmic equation (2) of paragraph 29. The values of these two parameters "g" and "b" together with the value of "p" (the proportion of land under jute) are shown at the top of the table for each thana. Using these values of "g" and "b" we can next calculate the graduated values of the variances, or rather of the logarithms of the variances as we are using the logarithmic equation (2) for purposes of graduation. The excellent agreement between observed and graduated values can be seen in chart (21·1) in which all the results have been shown graphically.

31. We can now test the goodness of our model sampling experiments by rigourous tests in two ways. We can first of all compare the results of the two (or three §) independent Samples in the same thana. This comparison has been shown in the first set of Table Nos. (16A), (17A) ...... (21A). In these Tables (except Table 16A) col. (1) gives the size of grids (x); col. (2) the number of grids used in the model sampling experiments (n); cols. (3) and (4) the logarithm* of the observed variances in the first and the second samples respectively. The next col. (5) gives the difference between the two values of log (v_1) and log (v_2). Squaring these

---

§The arithmetic procedure is slightly different in this case, and the Chi-square is obtained by dividing the sum of squares by the expected variance (2/n).

*We have naturally to work with the logarithms of the variances as we have used the logarithmic form of equation (2) for estimating the two parameters.

differences and dividing by (4/n) or multiplying by (n/4) we get the value of Chi-square* for each size of grid shown in col. (6). Adding the figures in col. (6) we get the total value of Chi-square for nine different sizes of grids. The probability of occurrence of a system of discrepancies equal to or greater than that actually observed was next obtained‡ from Karl Pearson's Biometric Tables Part I, Table XII, and is shown at the bottom of each Table. It will be noticed that the values of Chi-square are small and the probabilities are high indicating satisfactory agreement between the different samples in the same thana. This can also be seen from the observed values of the variances given in cols. (7) and (8) respectively. The above test shows that the internal consistency of the material was excellent.

32. We may now compare the observed variances based on the combined samples and the graduated values calculated with the help of our Variance Function. This has been done in the second or (B) series of Tables (15)—(21) for each thana. In these Tables cols. (1) and (2) give the size and total number of grids; col. (3) the observed logarithm of the variance based on pooled samples; and col. (4) the graduated logarithm of the variance. The differences are given in col. (5), and the values of Chi-square for the each size of grid in col. (6). Using the total values of Chi-square with 7 degrees of freedom§ the corresponding values of the probability of occurrence of systems of deviations equal to or greater than that actually observed were obtained from Biometric tables as before and are shown at the bottom of each Table.

33. For convenience of comparison I have given in Table (25·1) a summary of the results of model sampling experiments and the comparison between observed and graduated values of the

---

*The variance of the difference of logarithms of variances is (4/n); hence the square of the difference of logarithms of variances is divided in each case by (4/n), *i. e.*, is multiplied by (n/4).

‡As we have not used any estimated parameters the number of degrees of freedom is clearly 9 in every case except in Tejgaon where it is 18.

§As we have nine different sizes of grids and two statistical parameters, subtracting two from nine, we are left with seven degrees of freedom for the comparison between observed and graduated values of the logarithm of variances.

variances. In this Table (25·1), col. (1) gives the names of the Thanas; col. (2) shows the proportion of land under jute in each thana (p); col. (3) the area in square miles for which information by complete enumeration is available in each case. The co-efficient of variation expressed as a percentage of the mean value is shown in col. (4)*. The graduated values of "g" and "b", obtained by using least square methods with the logarithmic equation (2), are given in cols. (5) and (6) respectively. Values of Chi-square for comparison of sample values of logarithms of variances are given in col. (7), and the corresponding probabilities in col. (8). Values of Chi-square for the comparison between observed and graduated values of logarithms of variances are given in col. (9), and the corresponding probability of occurrence of deviations equal to or greater than those actually observed are given in col. (10).

34. It will be noticed from cols. (9) and (10) that, without any exception, the values of Chi-square are small and the probabilities are high and all greater than 0·66. This indicates excellent agreement between observed and graduated values in the case of all the thanas. It will be also seen that the same exponential formula gives a good fit over a large range of variation in the intensity of cultivation. In the case of Laksam, for example, the value of 'p' is very small and equal to 0·0242; so that only a small proportion of about 2·42 per cent. of the area in this thana is under jute. The co-efficient of variation is 635 per cent. which is very high; but even in this case the model sampling experiments based on only 1182 grids give fairly satisfactory results. We may conclude, therefore, that the exponential formula (2) used here gives excellent graduation in practice.

35. The form of the Variance Function may, therefore, be considered to be practically settled by the above series of model sampling experiments. The numerical values of "b" and "g",

---

*The theoretical or binomial variance is given by (pq) where "p" is the proportion of area under jute, and q= (1 – p); the standard deviation is therefore given by $\sqrt{pq}$. Dividing $\sqrt{pq}$ by "p" (*i. e.*, the mean proportion of area under jute) and multiplying by 100 we get 100 $\sqrt{(q/p)}$ as the percentage coefficient of variation which is given in col. (6). This furnishes a convenient measure of the degree of variation in the intensity of cultivation.

however, differ to some extent from thana to thana. A part of such differences may, no doubt, be ascribed to errors of sampling. It is not improbable, however, that another portion of the observed differences may be due to real differences in physical conditions in the different thanas or districts.

36. Without entering into a technical discussion of the subject I may briefly mention here that the two statistical parameters "b" and "g" may be interpreted to represent certain physical conditions. The value of (1/b) is roughly proportional to the ultimate unit of cultivation for jute in the different districts. For example, in Pirgacha thana we find that "b" is equal to 14·38. Dividing 1 acre or 60 kathas by 14·38, we get 4 kathas approximately, which may possibly stand for the smallest area on an average on which jute is likely to be sown in this thana.

37. Similarly the parameter "g" stands for the degree of Association between adjoining jute plots. If we consider any particular plot under jute in any thana, then the value of "g" gives a statistical measure of the probability of other plots in the neighbourhood being also sown with jute. Unfortunately it is not possible to investigate the variations in "g" and "b" in greater detail at present as our material is restricted to only seven thanas. This question is, however, of considerable importance as the optimum density of grids in different regions will be determined to a great extent by the respective values of "g" and "b". It is, therefore, extremely desirable that more work should be done on model sampling experiments with the help of the data for complete enumeration collected during the 1939 season.

38. Although the values of "g" and "b" differ to some extent from thana to thana, it will be noticed from Table (25·1) that the actual range of fluctuation in the case of "g" is comparatively small as it varies from 0·2812 in Nandail thana to 0·4076 in Pirgacha thana. The range of variation is greater in the case of "b", but even in this case, with the exception of Belkuchi, the value lies between 9·73 and 37·86. We may therefore use the weighted mean of these two parameters using again the logarithmic form of the equation. The actual mean values come out as $g = 0{\cdot}3328$, and $b = 16{\cdot}37$ which represent over-all estimates of these two parameters based on the whole series of model sampling experiments.

Sample Surveys in the Field

39. The observed variances as obtained from actual sample surveys in the field are given in Tables (22) and (23). In these Tables the name of the thana is shown in col. (1); the number of grids available in each case in cols. (2·1), (3·1) and (4·1); and the observed variances together with the standard errors in the corresponding cols. (2·2), (3·2) and (4·2) respectively.

40. Owing to the heavy cost of field surveys it was not possible to include in our programme in the case of any particular thana more than three different sizes of grids including random plots. As the area of individual plots is not constant we have considered it safer to omit random plots in estimating the values of "g" and "b". This leaves us with only two different sizes of grids or two observations for estimating two parameters. Although this is sufficient for algebraic purposes, the estimates will be precarious as no residual degree of freedom is left for errors. We have, therefore, considered it advisable to adopt a different procedure in these cases.

41. We know that if the distribution of jute plots had been truly binomial the variance would be given by (pq)/x where "p" is the proportion of area under jute, and "x" is the size of the grids. We also know that in actual fact the variance is not binomial, but is presumably given by our equation (2) of paragraph 29. We, therefore, first find the ratio of the binomial variance (pq) and the actual observed variance for each size of grids. We next calculate the logarithmic mean of these ratios for different sizes of grids, and from these mean values obtain an average estimate of "g" and "b".

42. The actual values of the logarithms of the ratios of variances are given in Table (24); and the mean value of the logarithms of the ratios is given at the bottom of the Table for each size of grid. Using these mean values, we find that the graduated value of "g" = 0·3459 and of "b" = 17·80. These then represent the over-all estimates of the parameters based on the sample surveys in the field.

43. From Table (25·1) it will be noticed that the value of "g" = 0·3328 and "b" = 16·37 for the model sampling experiments in the Laboratory. These agree quite well with "g" = 0·3459, and

"b" = 17·80 obtained from sample surveys in the field. Considering the meagreness of the data, the agreement between the two sets of values obtained respectively from model sampling experiments in the Laboratory and actual sample surveys in the field must be considered highly satisfactory. For convenience of numerical calculations, I have adopted a value of "g" = 1/3 = 0·33 and "b" = 17·56 = $(2{\cdot}6)^3$ in all subsequent work.

Variance of Random Plots

44. The case of random plots required special consideration owing to the fact that the size in acre of the plots vary appreciably from thana to thana. If the plot had been the ultimate unit for growing jute then each plot would have been classified as either wholly or not at all under jute. In this case the variance would have been given by the binomial formula p (1 – p), where p is the proportion of land under jute in the region surveyed. In actual fact random plot thus behaves like a grid of small size; but the situation is complicated by the fact that the size of the plot is itself variable which introduces an additional factor of variation. The graduation of the Variance Function in this case therefore becomes difficult; and we are obliged to use the method of ratio of variances. As the plot is not the ultimate unit for sowing jute, the variance for random plot is necessarily less than the corresponding binomial variance (= pq). Dividing the binomial variance (= pq) by the corresponding observed variance for random plots we obtain the ratios shown in col. (3) of Table (24). Taking the weighted average of the logarithms of these ratios and converting we get the value 1·17. We can use this ratio to calculate the expected value of the variance for random plots from a knowledge of the proportion of the area under jute in each case by simply dividing the binomial variance (pq) by 1·17.

Comparison with the Results of the Sample Survey of 1935

45. In my Note dated 26th June, 1938 on the programme for the Crop Census of 1938* I had mentioned that Mr. H. P. V. Townend, I.C.S. (then Rural Development Commissioner, Bengal) had organized a random sample survey in 1935 of the acreage under

---

*Proceedings of the Jute Census Committee, 6th July 1938, Enclosure to Appendix IV, paragraph 14, page 36.

jute in 20 districts in Bengal using about 2,600 sample grids each of size 40-acre approximately. As far as I know this was the first extensive sample survey of its kind in India; but unfortunately the material appears to have received little attention and was at one time in danger of being completely lost. After a good deal of correspondence I succeeded, with Mr. Townend's help, in tracing the original grid figures; and I have carried out a statistical analysis of the material which furnishes valuable information for planning future work.

46. I am giving in Table (25·2) in abstract form some of the results of the sample survey of 1935 which are of immediate interest to us in the present connexion. In this Table, column (1) gives the name of the district; column (2) the total area of the district in square miles; column (3) the number of police stations covered, and column (4) the number of grids used, in Mr. Townend's survey. The proportion of land under jute in each thana as obtained from the sample survey is shown in column (5); and the corresponding observed variance in column (6).

47. We may also calculate the expected values of the variances in each case with the help of our Variance Function. Using the adopted value of $g = 0{\cdot}33$, and $b = (2{\cdot}6)^3$ and equation (2) given in paragraph (29) we easily obtain the calculated variances shown in column (7). The ratio of the observed variance to the corresponding calculated variance is given in column (8).

48. It will be seen that practically in every case the observed variance is greater than the calculated variance. This is just what is to be expected. It will be remembered that our formula is based on the results of model sampling experiments or field surveys conducted in individual thanas in each district. Our formula is thus really valid for single thanas or homogeneous regions in which the value of p (or the proportion of land under jute) is more or less constant. The results for the sample survey of 1935, on the other hand, are based on the district as a whole which covers a large number of thanas in every case. The value of p (or the proportion under jute) also varies appreciably from one thana to another within the district. This leads us to expect that the variance for the district as a whole should be appreciably greater than the variance for a single thana. This is just what we find in col. (8) in Table (25·2).

49. In fact we may use the ratio of variances to obtain some idea of the increase in variability which may be expected to occur when we extend our calculations from a single thana to the district as a whole. For this purpose, we may use the geometric mean of the ratios given in column (8) of Table (25·2) which is easily found to be 1·38. This indicates that the variance of the district is likely to be 38 per cent. higher than the calculated value obtained by using the formula we have developed for the case of a single thana.

50. It is, however, distinctly reassuring to find that the variance for the district as a whole is not widely different from the variance of a single thana. It will be noticed that even in the case of Laksam thana, which we already know has the exceptionally high co-efficient of variation of 635 per cent., the ratio of the district to the thana variance is not larger than 1·98. We find then that we may use a multiplying factor which is of the order of 1·5 (and is probably not greater than 2·0) to obtain approximate estimates of the variances for the district as a whole from known values of the variance for single thanas. The results of the sample survey of 1935 thus enables us to obtain valuable information regarding the magnitude of the variability for whole districts. To be on the safe side I have used for this purpose a multiplying factor of 2 (instead of the mean value 1·38) in subsequent calculations.

Summary

51. We may summarise the chief results of the discussion in this section.

(1) The form of the Variance Function may be now considered as definitely settled. It is given by a simple exponential formula, $v_x = (pq) \cdot (bx)^{-g}$, which has been fully confirmed by extensive model sampling experiments in the Laboratory.

(2) The values of the two statistical parameters which enter into this exponential equation may differ to some extent from one region to another; but the data available were not adequate to decide this point. As the question is of considerable importance, it is desirable that it should be further investigated by model sampling experiments with the help of data for complete enumeration collected in 1939.

(3) As the ranges of fluctuations of the two parameters were not large, the weighted mean values were determined and were found to be "g" = 0·3328 and "b" = 16·37.

(4) The results of actual sample surveys in the field were used to obtain an independent estimate of the two parameters which came out as "g" = 0·3459 and "b" = 17·80. As these values are in excellent agreement with the corresponding values based on model sampling experiments, it was decided to adopt, for convenience of numerical calculation, values of "g" = 1/3 = 0·33 and "b" = 17·56 = $(2{\cdot}6)^3$ in subsequent work.

(5) In the case of random plots, the geometric mean of the ratio of the binomial variance to the observed variance was found to be 1·17. It is thus possible to calculate the expected value of the variance for random plots from a knowledge of the proportion of area under jute by dividing the binomial values (pq) by this factor 1·17.

(6) The results of the random sample survey of 1935 with grids of size 40-acre showed that the variance for districts as a whole were consistently greater than the corresponding estimated values calculated with the help of the Variance Function. This result is just what is to be expected as each district covers a large number of thanas and as the intensity of cultivation varies appreciably from thana to thana. The geometric mean of the ratio of the observed variance for the district as a whole to the estimated variance based on the exponential formula for a single thana was 1·38. This ratio may be used in making a rough estimate of the variance of the district as a whole from a knowledge of the variance for individual thanas.

Section 4. The Cost Function

Introduction

52. One of the chief objects of the field survey in 1939 was to gather necessary material for a scientific study of the cost of operations for different sizes and densities of grids. For convenience of reference I shall call each combination of size and density as a "pattern" for field work. Thus 4-acre grids located at a density of one grid per square mile on an average forms one pattern; 4-acre grids located at a density of say two grids per square mile forms a different pattern, and so on. In previous reports I have explained at considerable length the need of using different patterns for studying the Cost Function*. It will be useful to recapitulate briefly the main argument. With a small number of sample units or grids per square mile it is clear that proportionately more time will be spent in journeys in moving from one grid to another; on the other hand when the density or number of grids per square mile is large, less time will be spent on journeys, and more time can be utilized in actual field enumeration. The cost per square mile will however be less with widely scattered grids than with a large number of sample units per square mile. The cost will also depend on the actual size of grids used for the sample survey. In this situation the sampling error per square mile will depend on both the size and density of grids.

53. It will be also useful for convenience of reference to reproduce in a condensed form the chief conclusions reached in this connexion in paragraph 100 of the *First Report on the Crop Census* of 1938‡.

(*i*) For any given rate of expenditure per square mile, or what comes to the same thing, for any assigned total expenditure for the sample survey, the accuracy of the final estimate depends on both the size as well as the density or number of grids per square mile. A proper choice of the size of grids is therefore most important.

---

*See, for example, Section 6 and 7, paragraphs 67 to 100, pp. 24-32 of the *First Report on the Crop Census of 1938*.

‡*Proceedings, Jute Census Committee*, 6th February, 1939, p. 32.

(*ii*) In order to determine the best size of grids it is necessary to study the cost of operations not only for different sizes but also for different densities of grids.

(*iii*) The results of the field survey of 1938 suggested that, on the whole, sampling units or grids of smaller size were more efficient and economical than grids of larger size in the sense that a higher accuracy in the final estimate could be attained with the smaller sizes with the same total expenditure.

(*iv*) Although it was not considered likely that the Cost Function (that is, the relation between cost of operations and the size and density of grids) would be different in different districts or regions, it was pointed out that, in case in actual practice the Cost Function was found to be different in different regions, it might be necessary to use different sizes of grids in different districts or zones.

(*v*) As the intensity of cultivation (or the proportion of land under jute) is known to differ very widely in different districts or zones, it is in any case necessary to use different densities of grids in different districts or zones in order to attain the highest possible accuracy in the final estimate.

Analysis of Time-records

54. In order to gather the required information for different sizes and densities of grids it was decided to use 25 patterns (or combinations of size and density) for the field survey in 1939. In the case of each pattern an attempt was also made to obtain the information from more than one thana in order to secure reliable estimates. A special form for time records was prepared, and each field investigator was asked to make detailed entries relating to the time spent for different purposes on each day. These time records or diaries for field work kept by the primary investigators were sent to the Statistical Laboratory for analysis.

55. After some preliminary trial tabulation it was finally decided to classify the time records into four broad groups:

(*a*) *Enumeration.*—This includes the time spent in identifying the plots covered by the sample grids by reference to the 16″ mauza maps; inspecting the plots; and the actual recording of the serial number of the plots under jute.

(*b*) *Journey*.—Journey includes the time spent in going from camp to camp (camp being defined for our present purposes as the place where the night is spent); from camp to sample; and from sample to camp. It thus covered all journeys undertaken for the purpose of actual carrying out of the field enumeration.

(*c*) *Miscellaneous*.—This includes the time spent for all preliminary arrangements, making copies of the field records and despatching the same to inspectors or Headquarters; time spent in receiving instructions, etc. A sub-total of (*a*) enumeration, (*b*) journey, and (*c*) miscellaneous was used to give the actual working time spent for direct productive purposes.

(*d*) *Overhead*.—The remaining portion of the day was classified as overhead and included the time spent in making arrangements for sleep, rest, food, etc.; and also the wastage or time lost on account of sickness, holidays, cessation of work owing to drought, excessive rain, or other unforseen reasons. The overhead thus represents the total time spent for indirect or non-productive purposes. The total of the working time plus "overhead" naturally adds up to 24 hours on each day; this was used as a check on the primary records.

Observed Values of the Cost of Field Survey

56. The primary analysis of the material was carried out on the same lines as in 1938 and the net working hours required for each sampling unit were first obtained. Multiplying by the density of sampling units per square mile, the hours per sampling unit were converted into working hours per square mile which is more directly useful for our present purposes. The actual observed cost of field survey in hours per square mile is shown for each pattern (or combination of size and density) for each district and thana separately in Tables (26)—(29). In these tables, in which the figures are shown separately for each size of grids, col. (1·1) gives the name of the district and the thana; col. (1·2) the density of sampling units per square mile; col. (2·1) net working hours per square mile for "enumeration"; col. (3·1) net working hours per square mile for "journey"; col. (4·1) net working hours per square mile for "miscellaneous" work. The total working time, obtained by adding cols. (2·1), (3·1) and (4·1), is shown in col. (5·1); the time spent in

"overhead" is also shown in each case in hours per square mile in col. (6); and the actual total time in hours per square mile for the field portion of the work is given in col. (7·1)

57. It will be noticed that we have more than one set of observations for each pattern or combination of size and density of grids. It is convenient to take the average for each pattern. The results are given in Table (30) in which col. (1·1) shows the size of the sampling unit; and col. (1·2) the density of sampling units per square mile. The arrangement of the remaining columns is exactly the same as in Tables (26)—(29); cols. (2·1), (3·1), (4·1) and (5·1) give the time taken in hours per square mile for enumeration, journey, miscellaneous, and the total working time respectively. In the same way col. (6) gives the overhead; and col. (7·1) the total time in hours per square mile required to complete the field survey for each pattern.

58. The observed values are often irregular. For example, in the case of 4-acre grids it will be seen that the time required to complete the field work with a density of two grids per square mile was 17·00 hours per square mile which was actually higher than the time required, namely, 15·29 hours per square mile for doing the work with a higher density of 3 grids per square mile. The same kind of reversal may also be noticed in the case of 16-acre grids; for a density of one grid per square mile the time required was 19·87 hours per square mile, while with a density of two grids per square mile or double the total number of the grids, the time required was less and 18·89 hours per square mile. Such differences are of course inevitable, and arise from a variety of causes. I have repeatedly pointed out in previous reports that the efficiency of the primary workers varies widely as some of them are naturally fast workers while others are slow. Such intrinsic differences in the speed of the different investigators introduce large differences in the observed time for the field work. The ease or difficulty of communications and other facilities for the field work depends again to some extent on the season during which the work is done. Differences are also introduced by what may be called the practice effect; the output naturally improves as the investigators gain experience. There are also many other factors which introduce variations from time to time. In the result the raw observed values are subject to considerable fluctuations and errors of sampling.

59. It is clearly necessary to graduate the observations. As we have only 25 patterns for six different sizes of sampling units (random plots, 1, 4, 9, 16 and 36-acre) and no less than 12 different densities ($\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 3, 4, 6, 8, 9, 16, 25, 36) there are many gaps in the Scheme in the sense that many of the combinations of size and density are missing. This was, of course, inevitable as we could not afford to include all possible combinations of size and density, and a selection had to be made, owing to the high cost of the field survey. However, such gaps (and the fact that each investigator could not work on all the patterns and the same volume of work could not be allotted to all patterns) introduced great difficulties in the graduation of the material as standard methods were not available. A good deal of time was spent in studying the problem; and various methods, some graphical, some numerical and some a combination of both, were tried.

60. During the preliminary examination of the data several things became clear. As already pointed out, it was found that large differences were sometimes introduced owing very likely to intrinsic differences in the efficiency and speed of different investigators. It was sometimes noticed that the field work was systematically slower or faster than the average for some of the Field Units. Owing to the meagreness of the data and the fact that the composition of the different Field Units changed from time to time it was however not possible to apply systematic corrections for the speed of different individual investigators. It was also clear that, as only a very small number of observations were available for each thana, there was no chance of distinguishing between or studying separately the Cost Function in different thanas or districts. It was therefore decided to graduate the material as a whole and to estimate the average or over-all Cost Function for all thanas taken together.

Average number of plots per grid

61. A serious difficulty in the study of the Cost Function may be conveniently considered at this stage. In our instructions to the Field Branch we had desired that each field investigator should explicitly mention the number of sample grids enumerated on each working day. Unfortunately in most cases this was not done; and only the total number of plots enumerated was given in the field diaries. When we started tabulating the time records from the

diaries, we were therefore obliged to work with the number of plots, instead of the number of sample grids, enumerated on each working day.*

62. The cost of operations in every case had to be worked out consequently in terms of number of plots instead of in terms of number of grids per day. This makes it necessary to convert all calculations in terms of "number of plots" into the equivalent "number of sample grids" per day.

63. In order to do this we have to determine the average number of plots per grid of each size. As a first step, the average size of plots in each thana was directly calculated, and the results are shown in Table (31). The over-all weighted average comes out as 0.36 acre or just about 15 kathas, but it will be noticed that the average size differs appreciably in different regions.

64. The number of plots per sample grid of each size is shown for each thana in Tables (32)-(34). Owing to the appreciable fluctuations in the average area of individual plots, the average number of plots per grid also varies considerably from thana to thana. This fluctuation, unfortunately, introduces an unknown margin of error in our calculations for the Cost Function at the stage of conversion from "number of plots" to equivalent "number of grids per day. The magnitude of the error is, however, I believe not large; and, although the discrimination between different sizes of grids must have been blurred to some extent, the general conclusions may be safely considered to be unaffected by it.

Graduated Values of Cost of Field Work

65. We may now resume our discussion of the question of graduation of the observed values of the cost of field work. As regards "journey" time it was gratifying to find that this depended only on the density of grids per square mile, and was practically independent of the size of grids. In fact it was found that, apart from a constant term, the journey time in hours per square mile increased proportionately to the square root of the density of grids per square mile. This may be expressed in symbols:

$$J = a_1 + b_1\sqrt{y} \qquad \ldots \qquad \ldots \qquad \text{(c-1)}$$

*As the work of tabulation had to be started as soon as the diaries began to reach us, and as many of the corresponding field books were received later, it was not possible to recover the number of sample grids from the field books as this would have seriously delayed the statistical analysis.

where J = journey time in hours per square mile;

y = density of grids per square mile;

and "a_1", "b_1" are constants numerical values of which are given in Table (35).

The above result is quite plausible. In fact on theoretical grounds it can be easily shown that the journey time should vary as the square root of the density. Apart from the constant term, our observed results are thus in agreement with theoretical expectations. The constant term also is easily explicable from physical considerations; it probably represents the minimum level of expenditure necessary in order to organize the work on a field scale.

67. The time required for enumeration will naturally depend on the size of grids, as the number of plots which will have to be located and inspected will increase as the size of the grid is increased. For the same size of grids, the time required for enumeration per square mile should depend on the number of grids per square mile; that is, should be directly proportional to the density of grids. It is again gratifying to find that, broadly speaking, the observed results corroborate this. In fact it was found that the time required for enumeration could be graduated in every case by a constant term plus a linear function of the density of grids per square mile. We may state this result in the form of an equation:—

$$E = a_2 + c_2(y) \quad \ldots \quad \ldots \quad (c\text{-}2)$$

where E = Enumeration time in hours per square mile,

y = density of grids per square mile,

and 'a_2' 'c_2' are constants numerical values of which are given in Table (35).

68. The time required for miscellaneous work was found to depend on the number of grids enumerated, *i.e.*, on the density of grids per square mile. Here also a linear equation of the following type was found to be adequate.

$$M = a_3 + c_3(y) \quad \ldots \quad \ldots \quad (c\text{-}3)$$

where M = time in hours per square mile for miscellaneous work,

y = density of grids per square mile,

and 'a_3' and 'c_3' are constants numerical values of which are given in Table (35).

69. Adding together the time required for "enumeration", "journey" and "miscellaneous" work we get finally the total working time in hours per square mile in the following form:—

$$W = A + B(\sqrt{y}) + C(y) \quad \ldots \quad \ldots \quad (c\text{-}4)$$

where W is the total time required for field survey in hours per square mile; "y" as usual is the density of grids per square mile; and "A", "B" and "C" are constants the numerical values of which are given in Table (35).

70. With the help of the equations given in Table (35) it is now easy to obtain the graduated values of the time required for enumeration, journey, miscellaneous work, and the total working time for the field survey in hours per square mile. For convenience of comparison these graduated values have been shown together with the observed values in Table (36) in which col. (1·1) gives the size of the sampling unit; col. (1·2) the density of the sampling units per square mile; col. (2·1) the observed time and col. (2·2) the graduated time in hours per square mile required for enumeration. Similarly cols. (3·1) and (3·2) give the observed and graduated time in hours per square mile for journey; cols. (4·1) and (4·2) the observed and graduated time for miscellaneous work; and cols. (5·1) and (5·2) the observed and graduated values of the total working time required for the field survey.

Time in terms of Pay-days.

71. The time-records in the field diaries were given in net hours of work. The material was therefore tabulated in terms of net working hours, and so far we have given all results in terms of hours per square mile. It is, however, necessary to convert these net working hours into gross pay-days on the basis of which actual payments were made. For our present purposes the gross pay-day may be defined to include all days for which actual payment was made to the field investigators. The days on which no work was done owing to illness, holidays or other unforeseen reasons will be thus included in gross pay-day provided actual payment was made for these days.

72. In order to convert the net working hours into gross pay-days we have to determine the actual proportion of the time spent

in "Overhead". In the *First Report on the Crop Census of 1938*, paragraph 74, I mentioned that the overhead percentage was very nearly 70 in every case. A more reliable estimate is now available on the basis of the extensive work done this year. We find that during the field survey of 1939 the average "overhead" was 67·02 per cent. which shows that about 16 hours per day were required for indirect purposes, and 8 hours per day were available for direct productive work. Using this average figure of 67·02 per cent. for "overhead", it is easy to convert all net working hours into gross pay-days by simply multiplying the net hours by the factor 0·1266. The graduated time inclusive of overhead for the whole of the field work was obtained in this way and is shown in col. (7·2); while the actual observed total time is shown in col. (7·1).

Conversion of Hours into Money Values

73. We have so far carried out all our calculations in terms of hours per square mile or pay-days of field investigators. These gross pay-days must be now converted into money values by multiplying by an appropriate over-all rate of expenditure per gross pay-day of field investigators. Relevant data for the field survey of 1938 are available in the report on the work of the field branch prepared by Mr. Nihar Chandra Chakravarty, Supervisor-in-charge of the Field Branch*. From this report it will be seen that the field workers were paid for a total of 6345 days (p. 112). The total expenses for the field branch came to Rs. 17,676-15-0 of which Rs. 651-9-9 represented half share of the cost of maps; deducting the cost of maps the total cost of the field branch was thus Rs. 17,024-5-3. Dividing this total cost by the total number of pay-days of field workers we get Rs. 2-11-9 or Rs. 2-12-0 approximately.

74. Similar figures for the expenditure for the field portion of the work in 1939 are not yet available. I have, however, obtained certain provisional figures from the office of the Supervisor-in-charge of the Field Branch which I am using here. The total number of pay-days of field investigators in 1939 was 15,920; while the total expenditure for the field survey was about Rs. 46,520; dividing Rs. 46,520 by 15,920, we get Rs. 2-14-9 or Rs. 2-15-0 as the rate of payment per gross pay-day of field investigators in 1939.

*This has been printed as enclosure to the *First Report on the Crop Census of 1938* on pp. 111-116 of the Proceedings of the Jute Census Committee dated the 6th February, 1939.

As the detailed figures are not yet available I have considered it advisable to work with Rs. 3 per pay-day in round numbers.

Using this conversion rate of Rs. 3 per pay-day, or 2 annas per hour we can now calculate the cost in rupees per square mile for the field portion of the work. The observed values of the cost are shown in col. (3·2) of Table (37), and the corresponding graduated values in col. (3·3) of the same Table. For convenience of reference I am also quoting in col. (3·1) of the same Table (37) the estimated cost in rupees per square mile in 1938 as given in the *First Report on the Crop Census of 1938*, Table (7·1), col. (3·2) pp. 104—105.

76. A comparison of the two sets of figures for 1938 and 1939 will show that the expenditure for the field work was invariably higher this year. This is partly due to a slightly higher general rate of expenditure per gross pay-day, namely, Rs. 2/15/- this year against Rs. 2/12/- in 1938. The greater part of the discrepancy is however, due to the highly abnormal conditions of drought which prevailed during the earlier part of the season and which delayed the sowings of jute leading to a great deal of wastage of man-power on the field. In any case, the graduated values for 1939 on the basis of which I am making subsequent calculations are definitely conservative. In fact we may reasonably expect that the cost of the field work will be appreciably lower in practice in normal years.

Cost of Statistical Work

77. As regards the cost of the Laboratory portion of the work the observed values are shown in rupees per square mile in col. (2·2) of Table (37). The graduation of the observed values in this case was comparatively simple as it was found that for each size of grid the cost increased linearly with the density of grids per square mile. The graduated values in rupees per square mile are shown in col. (2·3). It will be noticed that the agreement between the observed and graduated values is very close.

78. It is of some interest to compare the present set of figures with the figures for 1938 given in col. (3·1) of Table (7·1) on page 104 of the *First Report on the Crop Census of 1938*. For convenience of comparison these figures have been reproduced in Table (37) in col. (2·1). It will be noticed that in every case the cost is much lower this year. This is encouraging, as it shows a distinct

increase in efficiency in the Laboratory portion of the work due to improvements in technique and procedure, increasing experience and better selection of workers*.

79. It should be remembered, of course, that the cost given in Table (37) refers only to the work of preparing the sample units and the final calculation of the proportion of land under jute; that is, what may be called the standard routine part of the work. The figures thus do not include the cost of model sample experiments, analysis of time records, and other advanced statistical analysis of the material.

80. Considering the nature of the material and the many different sources of variation, the agreement between observed and graduated values of the cost may be considered quite satisfactory. It is gratifying to find that the equations of the Cost Function are all of the simplest type, and are such as one would expect to hold good from broad physical considerations. This gives us confidence in using the present system of equations for the next stage of the calculation namely the determination of the best size and density of grids.

*A part of the difference may also be due to the fact that last year's figures were originally obtained by estimation from work on all crops instead of on jute alone.

Section 5. The Best Size of the Grid.

81. We may now consider the most important question of the discrimination of the best size of the grid. I discussed this subject at considerable length with illustrative examples in Section 7, paragraph 77-100 of the *First Report on the Crop Census of 1938*‡. Last year we were obliged to make the calculations by purely numerical methods. As we have this year graduated equations for both the Variance Function and the Cost Function it is possible to proceed on more analytic lines. It will be out of place to discuss here the details of the mathematical theory; but I may indicate briefly the procedure adopted in actual calculations.

82. We have already seen that the variance for grids of size x-acre and density of 'y' grids per square mile is given by:—

$$V(x,y) = \frac{pq}{(bx)^g} \cdot \frac{1}{y} \qquad \ldots \quad D(1)$$

where p = the proportion of land under jute; q = (1 − p); and "b" and "g" are statistical parameters. Using the adopted numerical values of $b = 17{\cdot}56 = (2{\cdot}6)^3$, and $g = 0{\cdot}33 = \frac{1}{3}$, we have

$$V(x,y) = \frac{pq}{2{\cdot}6(x)^{1/3}} \cdot \frac{1}{y} \qquad \ldots \quad D(1{\cdot}1)$$

In the case of random plots the variance for 'y' plots per square mile is given by the corresponding equation:—

$$V(R.P.,y) = \frac{pq}{1{\cdot}17} \cdot \frac{1}{y} \qquad \ldots \quad D(1{\cdot}2)$$

Dividing equations D(1·1) and D(1·2) by pq we get

$$\frac{v(x,y)}{pq} = \frac{1}{2{\cdot}6(x)^{1/3}} \cdot \frac{1}{y} = v'(x,y) \quad \ldots \text{ say} \qquad \ldots \quad D(2{\cdot}1)$$

$$\frac{V(R.P.,y)}{pq} = \frac{1}{1{\cdot}17(y)} = v'(R.P.,y) \quad \text{say} \qquad \ldots \quad D(2{\cdot}2)$$

For any assigned value of v′ in equations D(2·1) or D(2·2) we can easily get the corresponding value of 'y' or the density of sampling units per square mile.

83. We may now consider the Cost Function which is given by the following equations D(3·1) and D(3·2) for grids of size x-acre and random plots respectively:—

$$t(x,y) = a_x + b_x(\sqrt{y}) + c_x\,(y) \quad \ldots \qquad \ldots \quad D(3{\cdot}1)$$

$$t(R.P.,y) = a_o + b_o(\sqrt{y}) + c_o(y) \quad \ldots \qquad \ldots \quad D(3{\cdot}2)$$

‡Proceedings, Jute Census Committee, 6th February, 1939, pp. 26-32.

It will be remembered that the parameters a_x, b_x, and c_x are functions of x or the size of grids. The numerical values of these parameters as well as of a_o, b_o, c_o are given in Table (35).

84. If, for any assigned value of v′, we now substitute in equations D(3·1) and D(3·2) the value of "y" obtained from equations D(2·1) and D(2·2) we can immediately get the cost in rupees per square mile. In this way Table (38) was prepared showing the cost per square mile which will be incurred in working with random plots of grids of different sizes for different values of v′.

85. I may point out here that if we multiply the square-root of v′ by the quantity $\sqrt{\left(\frac{q}{p}\right)}$ we then get immediately the percentage error of the estimate for any region in which the proportion of land under jute is "p" and $q=(1-p)$. In actual practice it is, therefore, more convenient to use $E=100\sqrt{v'}$ which may be called a co-efficient of error. This is the quantity used in Table (38).

86. Table (38) thus gives us the cost in rupees per square mile required to attain any given level of accuracy by working with samples units of different sizes. In this Table col. (1) gives the values of $E=100\sqrt{v'}$; and the successive cols (2) – (7) the cost in rupees per square mile for working with a margin of error defined by the value of 'E. For example, for an assigned value E = 38, we find that the cost in rupees per square mile will amount to Rs. 2·83 for random plots, Rs. 2·73 for 1-acre grids, and Rs. 2·65 for 4-acre grids; the cost will be still higher and Rs. 2·83, Rs. 3·00 and Rs. 3·66 per square mile for working with grids of size 9, 16, or 36-acre respectively. We thus find that 4-acre grids will be the most economical size with a margin of error defined by e = 38.

87. We can draw a number of important conclusions from Table (38); but in doing so it is important to keep in mind certain broad physical considerations. In 1940 it is proposed to conduct the sample survey in 8 districts covering an area of 24,566 square miles. In a full scale provincial survey we may have to cover 55,000 square miles. It may be possible, of course, to reduce this area to some extent by leaving out regions in which little or practically no jute is grown on the basis of the detailed information now being collected by the Government of Bengal in connexion with the Jute Registration Scheme. But even then it is not likely that the total

area to be surveyed will fall below say 40,000 square miles. In a full scale survey we shall thus have to cover any thing from 40,000 to 50,000 square miles. Exclusive of the cost of maps, area extraction, supervision and other overhead charges, it is not likely that we shall be able to spend more than say Rs. 3 or at the most Rs. 4 per square mile. The lower limit is probably given by a figure of say Rs. 1·75 (or Rs. 1/12/-) per square mile, as below this level of expenditure the margin of error will be too large to be of any practi cal value. On the other hand, we must leave some margin for the uncertainty in our estimates of cost. I have already pointed out that this year's figures are probably in excess of what we may reasonably expect to occur in more normal seasons. If the cost of field work is found to be actually lower, this will mean that all the cost figures given in Table (38) will be proportionately decreased.

88. Taking all these factors into consideration we may broadly state that the most useful region in the Table for our purpose is probably the portion which lies between a rate of expenditure of say Rs. 2 and Rs. 3 per square mile; but the limits may be reasonably extended to Rs. 1/12/- on the lower and say Rs. 4/- per square mile on the upper side. Beyond these limits the estimates have little practical interest but have been given here to show the nature of curves more clearly.

89. Confining our attention to this range of costs from Rs. 1/12/- to Rs. 4 per square mile, we find that the 4-acre grid is consistently more economical than grids of other sizes in the sense that the cost of the sample survey will be lowest with 4-acre grids for any margin of error between the selected limits. It will be seen, however, that the discrimination between grids of small size, namely, between 1-acre and 9-acre is not sharp. This shows that with a higher precision in estimating the Cost Function it may turn out that in certain districts the most efficient size of the grid will not be 4-acre but some other size lying between 1 and 9 acres. It is desirable, therefore, to include this question for further investigation in the programme of work for 1940.

90. As regards random plots we find that, up to a rate of expenditure of a little above Rs. 3 per square mile, the efficiency is less than that of 4-acre grids. With a higher level of expenditure it appears, however, that random plots will be more economical. This can be clearly seen from the accompanying chart (38·1) in

which the curve for random plot cuts the curve for 4-acre at a critical expenditure of about Rs. 3·08 per square mile. Beyond this level of expenditure, random plots would be the most economical type of sampling units for sample surveys of acreage.

91. I may note in passing that the reversal observed in the case of random plots is quite typical in the present problem. For example, it will be noticed that at a still higher level of expenditure of the order of Rs. 5, the curve for 1-acre crosses below the curve for 4-acre showing that beyond this scale of expenditure 1-acre grids will be more efficient than 4-acre. The question is, however, only of academic interest as this high rate of expenditure is never likely to be incurred for a sample survey.

92. I have already pointed out that owing to the variation in the intensity of cultivation (that is, the proportion of land under jute) in different districts or zones it will be necessary to use widely different densities of grids in different areas in order to reduce the error of the final estimate to a minimum. The variation in the density of grids necessarily means also a wide range of variation in the cost per square mile. From a study of Table (38) it can be seen that such variation in the cost per square mile may make it desirable to use different sizes of grids or even random plots in some of the areas. This suggests that some information should be collected also for random plots during the field work of 1940.

93. We may now consider grids of larger size. It is clear from Table (38) or the accompanying graph (38·1) that the cost will be consistently higher with grids of size greater than 9-acre at all levels of error. It will be remembered that this was also shown by the field survey of 1938. We may therefore take it as established for all practical purposes that grids of large size will be definitely uneconomical, and may therefore be dropped from our future programme.

94. It will be remembered that last year there existed a good deal of difference of opinion regarding the best size of grids. In 1935 Mr. Townend had used grids of size 40-acre approximately. In 1938 Sir Bryce Burt (at that time Chairman of the Jute Census Committee) was definitely in favour of using grids of large size. In his note of the 11th October 1938* he had emphatically stated,

**Proceedings, Jute Census Committee*, 26th October, 1938, Appendix (I) 6. Ibid. Appendix (III), 25, p. 20.

"I would not have less than 36 acres". In my note of the 23rd October 1938, I was unable to agree to this proposal; and in paragraph (92) of the First Report on the Crop Census of 1938 I pointed out that the balance of evidence was in favour of grids of smaller size. The work done in 1939 has definitely confirmed this and has clearly brought out the advantages of using grids of small size.

95. We find in fact that 4-acre comes out as the best size of the grid. It may be of some interest to consider the actual advantage gained in using 4-acre grids in preference to 36-acre. With a co-efficient of error of 38 we find that the level of expenditure for 4-acre grids is Rs. 3·04 against Rs. 4·31 for 36-acre. This represents an advantage of 42 per cent in favour of 4-acre grids. At a higher level of error of e = 42, we find that the cost per square mile is Rs. 2·36 for 4-acre against Rs. 3·18 for 36-acre grid which gives an advantage of 36 per cent in favour of the lower size. At a considerable higher level of error of e = 50, the cost is Rs. 1·97 for 4-acre and Rs. 2·54 for 36-acre respectively, showing a difference of nearly 30 per cent in favour of 4-acre. We find then that, in the region in which we are interested, the saving effected by using 4-acre grids in preference of 36-acre will be of the order of 30 or 40 per cent of the total expenditure.

96. In case a sample survey is repeated annually in future years the saving which would be effected by using 4-acre grids in preference to 36-acre or grids of higher size would thus amount to something between thirty and fifty thousand rupees per year.

97. We may claim that considerable progress has been made this year as we have succeeded in settling one of the outstanding questions, namely, the best size of the grid with some accuracy. We now know that this lies somewhere between 1 and 9 acre, very likely in the neighbourhood of 4-acre; and that 4-acre is probably the best standard size for general purposes. As already pointed out there may, however, be small variations from region to region, which makes it desirable that a small proportion of grids of size 2, 6 and 8-acre should be used in 1940 in addition to the standard size of 4-acre to study the question in greater detail.

Section 6. The Best Density of the Grid

98. I shall now consider the second important question, namely, the best density or number of grids per square mile in different regions. I have already explained in previous reports that in order to attain the highest accuracy, it is necessary to use different densities of sampling units as the intensity of cultivation differs widely in different regions. For an adequate discussion of this problem a certain amount of mathematical theory is indispensible which, however, will be out of place in this Report. I may mention very briefly that the mathematical approach is theoretically simple. For any given level of expenditure, we have to choose the size of grids "x" as well as the density of grids "y" in each of the zones in such a way that the variance is minimised. We first form a cost-variance equation with the total cost (or rate of expenditure per square mile) as one of the parameters. Differentiating this cost-variance function, and using Lagrange's method of undetermined multipliers, we get a sufficient number of equations of which numerical solutions can be obtained by semi-graphical methods of approximation. The calculations are laborious but do not present any serious difficulties. Details of the precedure together with the relevant mathematical theory will be published elsewhere.

Best Density for Eight Districts

99. I may give here a concrete example of the method of calculation developed for this purpose. Let us consider the 8 districts which are proposed to be surveyed in 1940. At a level of expenditure of about one lakh of rupees (exclusive of maps and fixed overhead) the best distribution of densities for 4-acre grids comes out as shown in Table (39).

Serial No.	Name of District	Area in sq. miles	Proportion under jute (p)	Binomial variability $\sqrt{(pq)}$	Best density of grids (y)
(1)	(2)	(3)	(4)	(5)	(6)
1	Mymensingh	6237	0·142	0·3490	3·222
2	Dacca	2713	0·158	0·3600	3·328
3	Rangpur	3496	0·112	0·3154	2·908
4	Tipperah	2597	0·110	0 3129	2·876
5	Bogra	1384	0·096	0·2916	2·710
6	Rajshahi	2609	0·054	0·2260	2·058
7	Jessore	2902	0·037	0·1888	1·700
8	Nadia	2881	0·020	0·1400	1·240

100. In this Table (39) col. (1) gives the serial number and col. (2) the name of the district; col. (3) the total area in square miles. The value of 'p' or the proportion of land under jute is given in col. (4); and the corresponding binomial variability, given by the square root of (pq), is shown in col. (5). The best density for minimising the error of the final estimate for the eight districts taken together is shown in col. (6).

101. It will be noticed that the density or number of grids per square mile depends, to a large extent, on the intensity of cultivation. In fact, the density increases with 'p' the proportion of land under jute, or even more closely with the binomial variability $\sqrt{(pq)}$. The density is highest in Dacca where the intensity of cultivation is also greatest. In fact in the first four districts Mymensingh, Dacca, Rangpur and Tipperah, and to a large extent also in the fifth district Bogra, the optimum density or number of grids per square mile is high and roughly of the same order because the intensity of cultivation in all these districts is also high. On the other hand, in the case of a district like Nadia, where the intensity of cultivation is low, the best density also falls to a low value of 1·24 which is almost one-third of the optimum density in Dacca.

Error of the Final Estimate

102. In order to complete the discussion I may also consider the question of sampling errors in the present case. The mathematical formula for this purpose has been worked out and depends on the area in square miles (A), the intensity of cultivation (p), and the density of sampling units (y). We have to calculate first a factor defined by:—

$$f = \left(\Sigma A_i p_i q_i^{t}{}_{c}\right)^{\frac{1}{2}} \Big/ \Sigma \left(A_i p_i\right)$$

where the suffix "i" refers to the different districts or regions and the summation has to be taken over all districts or regions. This numerical factor "f" for the above 8 districts taken together is found to be f= ·0198 or ·02 approximately.

103. The next step is to obtain the appropriate value of the co-efficient of error from Table (38). In the present case we are working with a total expenditure of one lakh of rupees for about 25,000 square miles or a rate of expenditure of Rs. 4 per square mile. Using this figure, we find from Table (38) that the co-efficient

of error will be roughly e=28 for 4-acre grids. We now multiply this value of e=28 by the numerical factor f=·02, and we obtain 0·56 as the total co-efficient of error for the 8 districts taken together

104. We can use this total co-efficient of error, 0·56, to calculate the margin of error which may be reasonably expected to occur in actual practice. We have first of all to make an allowance for the fact that the variance function used by us is strictly applicable to homogeneous regions with more or less the same intensity of cultivation. I have already discussed this question in the section on the Variance Function, and I have suggested using a safety factor of 2 for variances or 1·41 approximately for standard deviations. Using this we get an adjusted co-efficient of error of 0·79.

105. We have next to multiply this adjusted co-efficient of error by an appropriate factor to give the margin of error on any assigned level of significance. I have pointed out, however, that here also our actual experience has been that the observed range is roughly twice that calculated on the basis of a normal distribution. This means that in order to attain a level of significance of five per cent in actual practice we should use the normal factor for $2\frac{1}{2}$ per cent. which is 2·24. Multiplying the adjusted co-efficient of error 0·79 by this factor of 2·24 we get 1·8 as an estimated value of the margin of error on the five per cent. level of significance. If we work with half samples (which has been explained in a later section) this will have to be multiplied again by $\sqrt{2}$. Doing this we get a final figure of about 2·5. As the co-efficient of error has been expressed throughout as a percentage this figure of 2·5 is also in per cent of the final estimate of the land under jute.

106. We find then that in case we carry out a sample survey in the above 8 districts in the form of two independent half-samples with an expenditure of one lakh of rupees (exclusive of maps and other fixed overhead charges and the field work is done properly, we may reasonably expect to be able to estimate the total area under jute in these 8 districts with a margin of error of less than say 3 per cent. In other words, the sample estimates obtained independently from the two samples should not differ by more than say 3 per cent. The margin of error will be, of course, lower for the pooled estimate based on both the samples.

107. This is the type of calculations which we can perform once the particular districts or the area to be covered by the survey and the total expenditure is fixed. It will be seen, therefore, that we have now practically completed the technical apparatus for calculating the best density of grids in different districts as also the expected margin of error in any concrete case.

General Conclusions

108. The position reached on the basis of the work in 1939 may be now summarised.

(1) The form of the Variance Function has been practically settled and we may safely use the exponential equation given in this report.

(2) It is however not clear whether the numerical values of the statistical parameters "g" and "b" remain sensibly constant or vary from region to region. More work on model sampling experiments is desirable and necessary to settle this question.

(3) The form of the Cost Function has been formulated and it has been shown that graduated and observed values are in good agreement.

(4) Owing to the meagreness of the data, it has not, however, been possible to investigate whether the cost of operations varies appreciably from one region to another. This question will have to be investigated in greater detail with the help of the more extensive data proposed to be collected in 1940.

(5) Great advance has been made regarding the choice of the best size of grids. Random plots and smaller size of grids from 1-acre to 9-acre were found to be definitely more efficient and economical than grids of larger size.

(6) In case one standard size of grid is to be used in all the districts it is likely that 4-acre will be found to be the best size for this purpose.

(7) The question of the best density for any given size of grids has been studied; and it has been found that a wide variation in the density depending on the intensity of cultivation is necessary in order to reduce the error to a minimum.

(8) As different densities will have to be used in different districts it is likely that a further increase in efficiency may be secured by using different size of grids in different regions. For very high densities it is even possible that random plots will be found to be the most economical size of the sampling unit. This is one of the points which can only be settled as the work progresses and information regarding local differences in different regions becomes available.

(9) The statistical technique and the mathematical apparatus have been sufficiently developed to meet present needs. Once the area to be surveyed and the total expenditure to be incurred for this purpose is fixed, it is possible to calculate with considerable accuracy the best density of grids in each region as well as the expected margin of error of the final estimate. We have, therefore, every reason to feel satisfied with the substantial progress made this year.

Section 7. Programme of Work for 1940

109. I have already submitted certain provisional recommendations regarding the programme for 1940 which were considered by the Jute Census Committee on the 10th October 1939. It was decided by the Committee to proceed with the preparatory work for the eight districts, Mymensingh, Dacca, Rangpur, Tipperah, Bogra, Rajshahi, Jessore and Nadia. There is however, still some uncertainty regarding the contribution of the Government of Bengal which prevents any particular figure being adopted for budget purposes at the present stage. In this situation the best plan will be, I think, for me to discuss the programme of work for 1940 in its broad outline.

Need of an Extensive Survey in 1940

110. It is clear that the chief purpose for the field survey in 1940 must be to gather necessary information for organising a sample survey on a full provincial scale in 1941. It is therefore, necessary to obtain reliable cost figures to enable a firm estimate being prepared for this purpose.

111. I may perhaps explain at this stage why further information is at all necessary. The material collected this year has made it possible to reach the conclusion that random plots or grids of small size will be more economical than grids of larger size. It has also enabled us to decide the question of the best density or number of sampling units per square mile in areas having different intensities of cultivation. The information collected is not, however, sufficient to make a precise calculation of the cost of a sample survey on a full provincial scale. This for two reasons. For any particular size of grids the material is meagre. Consider for example, the 4-acre grid which appears to be the one most suitable for being used as a standard size in all districts. In 1939 only about one-ninth of the total field work was concerned with this particular size; consequently the cost figures can not be considered sufficiently reliable for preparing provincial estimates. In other words, although the work done in 1939 was adequate for making relative comparisons between grids of different sizes, it is not sufficient for making firm estimates of the expenditure when the sample survey is to be done on a provincial scale. This is shown by the difference in the cost figures obtained in 1938 and 1939 which has been already discussed in a previous section.

112. The second reason of the uncertainty in the cost figures is due to the fact that the work in 1939 was confined only to a small number of thanas in 5 districts. When the work is done on an extensive scale over the whole of 8 districts in 1940 as proposed it may be confidently expected that the cost figures obtained will be far more reliable. Work on such an extensive scale may also be expected to reveal differences, if there are any, in the cost of operations in different regions.

113. It will be seen from the above discussion that the next step in our programme must be to carry out a sample survey over an extensive area, preferably say in eight districts in the province. In other words, the period of intensive work confined to a number of selected thanas or even districts is over. The work must now be organized on a more extensive scale and must cover a sufficiently large number of districts to approach the provincial scale of operations.

Size and Density of Grids

114. We have already seen that 4-acre is the best size of grids for general purposes. In the preliminary report I have, therefore, already recommended that this should be adopted as a standard size and should be used in all the districts surveyed in 1940. For reasons explained in the Section on the best size of grids it will be necessary and desirable to make a sharper discrimination between grids of small size between say 2 and 8-acres. It is proposed, therefore, that besides the standard size of 4-acre which will be used in all the districts, a small proportion of grids of size 2-acre, 6-acre, and 8-acre as also a certain number of random plots should be included in the programme for 1940.

115. I may next consider the question of the density of grids. I have already pointed out that owing to differences in the intensity of cultivation it would be necessary in any case to use different densities in different districts. This will furnish us with a sufficient range of variation; and it will not be necessary to include any special studies so far as density is concerned. We shall, therefore, use in every district the best density as calculated on the basis of 4-acre grids; and in a small proportion of cases in each district

substitute some of the 4-acre grids by 2-acre, 6-acre, and 8-acre grids or by random plots.

116. It will be seen therefore that in 1940 the field Investigators will work throughout with the same density in each region. The organization of the field survey will be, therefore, much simpler as there will be no need of shifting the Investigators from one thana to another after short intervals. In fact, in future each investigator will work during the whole season within the limits of the area allotted to him. The work of inspection and supervision will therefore be much easier.

Statistical Controls

117. It will be remembered that it has been decided to leave out complete enumeration altogether in 1940. In this situation it is necessary and desirable to provide adequate statistical controls to enable the accuracy of the sample survey being tested on a scientific basis. It is recommended that the method of two independent half-samples should be used for this purpose. In this method, half the samples in each area will be enumerated by one set of investigators, and the other half will be enumerated independently by a second set. Separate estimates will be then prepared on the basis of the enumeration done by each set of investigators; and the difference between these two independent estimates will furnish an excellent check on the reliability of the survey as a whole. If the difference between the two samples lies within the expected range of sampling error, then it may be presumed that the estimate based on both the samples pooled together will also be reliable, and will not differ by more than the expected margin of error from the true value of the proportion of land under jute. On the other hand, if the difference between the two independent estimates is much in excess of the expected margin of the error then the samples must be considered unreliable.

118. It will be naturally a little more expensive to use two independent half-samples as it will involve duplicating the journey from camp to camp. It may be observed, however, that the cost of the Laboratory portion of the work and of the field enumeration will remain practically the same; the cost of the miscellaneous

portion of the work will be only partly affected; while over head charges will be only slightly increased. On the whole, it therefore appears definitely advisable to adopt the method of two independent half-samples in view of the excellent check which this will furnish on the reliability of the sample survey as a whole.

Model Sampling Experiments

119. In Section 3 I have pointed out that, although the form of the Variance Function may be considered as settled, it is not clear whether the statistical parameters entering into the equation of the Variance Function remain the same or differ from one region to another. It is desirable that this question should be further investigated with the help of model sampling experiments in the Laboratory, and this should be included in the programme of work for 1940.

Field Organization

120. In previous reports I have repeatedly emphasized the fact that for the successful carrying out of a sample survey it is necessary not only to develop an efficient sampling technique but also to build up the human agency required for this purpose. The relative efficiency of different types of field organization therefore deserves careful study; and it is suggested that those statistical methods which have proved so successful in deciding other questions may be used here with advantage. I am therefore giving below a few suggestions in this connexion.

121. As regards primary investigators I have pointed out in previous reports that it would be desirable to settle by actual trial on the field what pay should be given to get right type of men. I also suggested that the primary investigators should be recruited locally as far as practicable, so that they may do their work more or less in the same area in which they live. This question, I think may be very usefully kept in mind in organizing the field work in 1940.

122. The second question which calls for careful examination is the relative strength of the inspecting staff. In 1938 and 1939 there was one assistant inspector for each batch of 6 field investigators, and one deputy or full inspector for each pair of assistant

inspectors. As there is practically no difference in the type of men appointed as assistant, deputy, or full inspector, this meant that there was practically one inspector for every four field investigators. This is a very high proportion for the inspecting staff; and it will be desirable to investigate whether this proportion may not be reduced without any appreciable loss of efficiency. I have already pointed out that the organization of the field work will be much simpler in 1940. This is certainly an additional reason for making experiments with a lower proportion of inspectors. There is no difficulty in doing this. What is required to be done is simply to try out on the field different units having different numbers of investigators under one inspector. The design should be, of course, balanced as far as practicable in order to eliminate "personal equation" and other factors of variation.

Miscellaneous questions

123. I may now discuss a few miscellaneous questions connected with the programme of work for 1940. We have seen that there are reasons for believing that the cost figures for 1939 were unduly high owing to abnormal conditions of drought which prevailed in the beginning of the season and which led to a great deal of wastage of man-power on the field. It is likely, therefore, that the actual amount of work which can be done by the field investigators is really larger than that shown by our cost figures. In this situation if there are too few sampling units marked on the maps there will be a certain amount of wastage of the time of investigators in the field. It will therefore be desirable to provide a slightly larger number of grids on the maps, *i.e.*, work with a density slightly in excess of the figure actually obatined by calculation on the basis of the 1939 estimates. In case the field investigators can do more work than is anticipated at present they will be able to gather the information for the additional grids without any increase in the cost of the survey. This will involve a slightly higher expenditure in the preparatory part of the work in the Laboratory, and this should be provided in the budget.

Area of Plots

124. As regards copying the area of plots from khatian and estimating the area of fractions of plots I have always been of opinion that it would be more convenient to have this work done in

the Statistical Laboratory. In 1939 I made experiments to measure the area of plots directly from the maps; and I discussed this question at the meeting of the Jute Census Committee on the 26th October, 1938, when it was agreed that this work should be done in the Laboratory. Subsequently the work was transferred to the Field Branch as it was thought that the cost would be less if the work was done by the field staff. In actual fact we found, however, that the cost in 1939 was about the same as that which would have been incurred if the work had been done by the Laboratory staff. It is necessary to remember in this connexion that grids of only small size will be used in 1940; so that the number of fractional plots will be much larger than this year. In fact, with 4-acre grids it will be seen from Table (33) that on an average 13·55 out of 18·13 or roughly 75 per cent. of the plots will be broken which will necessitate a direct measurement of the area included within the grid. In 1940 the advantage will thus definitely lie with the method of direct measurement. I should also mention that difficulties were experienced this year in proceeding with the work of tabulation and analysis of the records owing to delays in the area figures reaching us. From every point of view it will be desirable, therefore, to have the work of area extraction done by the Laboratory staff in 1940.

125. *Maps.*—It will be desirable for the Jute Census Committee to take up again with the Government of Bengal the question of the purchase of village maps. These maps are ordinarily sold to the public at the rate of 4 annas per sheet. As a special concession the Government of Bengal have reduced the rate to one anna and six pies in the case of the Jute Census Scheme. We find in actual practice that roughly two and a half sheets are required to cover each square mile. On this basis, the price of one single set come to about four annas per square mile. As two sets of maps are usually required for our work, this means an expenditure of something between seven and eight annas per square mile. The expense was small so long as the survey was confined to selected areas. As the area covered by the survey is increasing the question is becoming more and more important. For example, in 1940 it is proposed to cover about 25,000 square miles; so that roughly Rs. 12,500 will have to be spent in the purchase of maps. This is not an inconsiderable amount, and comes to no less than 25 per cent. of the grant

of Rs. 50,000 contributed by the Government of Bengal for the present Scheme in 1939. It will be remembered that in order to organize the sample survey on a provincial scale it will be necessary to cover roughly 50,000 square miles; the price of maps in this case will come to about Rs. 25,000 which will represent half of what the Government of Bengal have contributed in the past. In view of the fact that the Jute Census Scheme is definitely a matter of public interest it is not unreasonable to approach the Government of Bengal for a reconsideration of the whole question.

Summary of Recommendations

126. I am giving below in summary form the recommendations regarding the programme of work for 1940:—

(*i*) The chief aim will be to obtain reliable cost figures for preparing plans and estimates for a full scale provincial survey in 1941. It is, therefore, recommended that grids of the standard size of 4-acre should be used in all the eight districts proposed to be surveyed next year.

(*ii*) The 4-acre grids should be distributed with appropriate densities in the different districts with a view to reducing to a minimum the error of the final estimate for the eight districts taken together.

(*iii*) In order to obtain supplementary information it is recommended that a small proportion of grids of size 2, 6 and 8-acre should also be used besides grids of standard size 4-acre.

(*iv*) It is recommended that some money should be provided for further work on model sampling experiments to clear up certain obscure points in connexion with the Variance Function.

(*v*) In order to provide a check on the reliability of the sample survey, it is recommended that the work should be done in the form of two independent half samples, information for each half sample to be collected by different sets of investigators.

(*vi*) It is desirable that the question of the field organization should also be studied by the same experimental methods as have proved successful in other cases. In particular it is desirable to investigate whether it will be possible to recruit suitable primary investigators locally; and also whether it is possible to reduce the proportion of the inspecting staff without any appreciable loss of efficiency.

P. C. MAHALANOBIS

GIRIDIH,

8th November, 1938.

TABLE 1.

Distribution of Work and Workers by Size and Density of Samples and Thanas

District & Thanas	Density	No. of Workers	Estimated Sq. Miles	Gross Days	Serial Roll No. of Workers
(1)	(2)	(3)	(4)	(5)	(6)
Random Plots.					
Mymensingh—Netrokona.	4	16	99	84	8, 22,37, 56, 57, 59, 68, 69, 87, 89, 95, 103, 106, 139, 145, 152
Hooghly—Serampore ...	4	7	72	43	17, 19, 48, 84, 118, 151, 154
24-Pgs.—Titagarh (Gr.) ...	4	5	67	47	36, 53, 67, 113, 151
Rangpur—Kurigram ...	6	8	113	94	1, 43, 61, 74, 75, 81, 104, 120
Hooghly—Dhaniakhali ...	6	6	106	70	28, 48, 84, 100, 118, 133 (B)
Rangpur—Rangpur ...	9	12	125	101	1. 32, 73, 74, 75, 86, 92, 104, 126, 141, 145, 147
Pabna—Pabna ...	9	12	137	124	18, 21, 23, 29, 30, 33, 85, 112. 120, 126, 130, 141
Mymensingh—Trisal ...	9	8	86	58	34, 35, 38, 46, 51, 76, 82, 114
Mymensingh—Melandah ...	16	20	198	219	8, 22, 37, 56, 57, 59, 66, 68, 69, 87, 89, 95, 96, 99, 103, 106, 111, 139, 143, 152
Rangpur—Nilphamari ...	16	15	146	180	32, 55, 60, 62, 73, 77, 94, 107, 108, 110, 115, 135, 142, 145, 147
Mymensingh—Bajitpur ...	16	5	65	58	50, 80, 93, 96, 140
24 Pgs.—Barasat ...	16	8	104	86	11, 65, 91, 101, 132, 136, 150, 156
Do—Baduria ...	16	8	70	61	40, 53, 67, 91, 101, 113, 136, 151
Hooghly—Chanditala ...	16	5	54	66	17, 19, 24, 151, 154

TABLE 1.—*Continued.*

Distribution of Work and Workers by Size and Density of Samples and Thanas

District & Thanas	Density	No. of Workers	Estimated Sq. Miles	Gross Days	Serial Roll No. of Workers
(1)	(2)	(3)	(4)	(5)	(6)
Random Plots					
Hooghly—Tarakeswar ...	16	7	41	45	17, 19, 24, 36, 67, 151, 154
Pabna—Ullapara ...	25	19	192	226	18, 21, 23, 29, 30, 31, 33, 49, 85, 93, 102, 109, 116, 119, 130, 138, 141, 146, 158
Mymensingh—Jamalpur ...	25	16	168	178	6, 8, 22, 35, 37, 52, 56, 57, 59, 68, 69, 89, 95, 106, 111, 139
Rangpur—Pirgachha ...	25	6	79	97	5, 55, 71, 129, 142, 153
24 Pgs.—Amdanga ...	25	3	53	45	70, 137, 157
Pabna—Belkuchi ...	36	6	70	95	25, 85, 109, 116, 119, 146
Mymensingh—Nandail ...	36	16	115	150	26, 41, 46, 50, 72, 80, 93, 96, 99, 117, 122, 123, 128, 131, 133A, 140
Mymensingh—Kishoreganj.	36	8	53	69	15, 34, 35, 38, 46, 51, 76, 133A
Rangpur—Palashbari ...	36	11	130	136	5, 44, 52, 54, 82, 86, 114, 121, 125, 129, 153
Rangpur—Gaibandha ...	36	36	232	334	1, 5, 32, 43, 44, 52, 54, 55, 60, 61, 62, 71, 73, 74, 75, 77, 81, 82, 86, 94, 104, 107, 108, 110, 114, 115, 120, 121, 125, 129, 134, 135, 143, 145, 147, 153
Pabna—Kamarkhand ...	36	4	46	58	21, 49, 102, 138
24 Pgs.—Deganga ...	36	6	75	72	65, 70, 136, 137, 150, 157

TABLE 2. Distribution of Work and Workers by Size and Density of Samples and Thanas

District & Thanas	Density	No. of Workers	Estimated Sq. Miles	Gross Days	Serial Roll No. of Workers
(1)	(2)	(3)	(4)	(5)	(6)
		1-Acre			
Mymensingh—Netrokona	1	17	108	65	8, 22, 37, 56, 59, 66, 68, 69, 87, 89, 95, 103, 106, 111, 139, 143, 152
Hooghly—Serampore ...	1	3	17	11	17, 19, 154
24 Pgs.—Titagarh (Gr.) ...	1	4	77	25	36, 53, 67, 113
Rangpur—Nilphamari ...	2	14	153	91	32, 55, 60, 62, 73, 77, 94, 108, 110, 115, 135, 142, 145, 147
Pabna—Pabna Sadar ...	2	10	126	72	18, 21, 30, 33, 109, 112, 116, 120, 126, 141
Hooghly—Chanditala ...	2	6	63	50	2, 17, 19, 24, 151, 154
24 Pgs.—Amdanga ...	2	3	40	17	30, 70, 137
Mymensingh—Melandah	4	6	85	62	37, 57, 66, 87, 89, 139
Hooghly—Tarakeswar ...	4	6	46	42	2, 17, 19, 24, 151, 154
24 Pgs.—Basirhat ...	4	8	151	123	7, 11, 36, 47, 76, 113, 132, 151
Rangpur—Gaibandha ...	4	10	118	67	4, 8, 43, 54, 71, 86, 104, 134, 142, 147
24 Pgs.—Barasat ...	4	5	67	49	65, 135, 140, 150, 156
Hooghly—Uttarpara ...	4	2	9	8	2, 24
Pabna—Belkuchi ...	6	5	53	43	21, 29, 31, 102, 146
Mymensingh—Jamalpur ...	6	24	229	164	22, 37, 44, 45, 52, 55, 56, 57, 59, 60, 66, 68, 69, 82, 83, 87, 89, 90, 95, 106, 111, 114, 140, 152
24 Pgs.—Baduria ...	6	12	100	134	9, 11, 36, 42, 49, 91, 101, 113, 120, 132, 136, 151
Pabna—Kamarkhand ...	6	5	31	36	21, 25, 29, 102, 119
Mymensingh—Nandail ...	8	20	126	166	10, 15, 26, 41, 50, 58, 72, 80, 86 93, 96, 99, 117, 122, 123, 128, 131, 133A 140, 146
Rangpur—Palashbari ...	8	6	60	90	62, 77, 107, 108, 110, 115
Hooghly—Dhaniakhali ...	8	6	102	103	28, 48, 84, 100, 118, 133B

TABLE 3.

Distribution of Work and Workers by Size and Density of Samples and Thanas

District and Thanas	Density	No. of Workers	Estimated square Miles	Gross Days	Serial Roll No. of Workers
(1)	(2)	(3)	(4)	(5)	(6)
4-Acre					
Hooghly—Dhaniakhali ...	½	6	98	61	28, 48, 84, 100, 118, 133B
24 Pgs.—Basirhat ...	½	8	130	48	7, 11, 36, 47, 76, 113, 132, 151
Rangpur—Kurigram ...	1	7	105	61	1, 43, 61, 74, 75, 81, 104
Mymensingh—Trisal ...	1	7	136	65	35, 38, 46, 51, 76, 82, 114
24 Pgs.—Deganga ...	1	6	78	49	65, 70, 136, 137, 150, 157
Pabna—Ullapara ...	2	7	167	96	49, 85, 109, 116, 119, 146, 158
Rangpur—Pirgachha ...	2	11	84	60	5, 55, 71, 74, 77, 107, 108, 110, 129, 142, 153
Mymensingh—Bajitpur ...	2	5	72	60	72, 80, 93, 96, 140
Mymensingh—Melandah ...	2	6	121	85	37, 57, 66, 87, 89, 139
Hooghly—Tarakeswar ...	2	8	34	37	2, 17, 19, 24, 100, 133B, 151, 154
Mymensingh—Kishoreganj	3	6	54	35	34, 35, 38, 46, 51, 76
Rangpur—Rangpur ...	3	9	104	57	32, 73, 74, 75, 86, 92, 141, 145, 147
Rangpur—Gaibandha ...	3	7	115	82	43, 53, 71, 74, 81, 104, 129

TABLE 4.

Distribution of Work and Workers by Size and Density of Samples and Thanas

District and Thanas	Density	No. of workers	Estimated Square Miles	Gross Days	Serial Roll No. of Workers
(1)	(2)	(3)	(4)	(5)	(6)
9-Acre					
24 Pgs.—Baduria ...	½	8	76	39	9, 11, 36, 53, 67, 101, 113, 132
Hooghly—Chanditala ...	½	4	63	37	2, 19, 151, 154
24 Pgs.—Titagarh (Gr.) ...	½	3	65	16	36, 67, 113
Mymensingh—Netrokona	1	18	117	61	8, 22, 37, 56, 57, 59, 66, 68, 69, 87, 89, 95, 103, 106, 111, 139, 143, 152
Rangpur—Nilphamari ...	1	13	161	92	32, 55, 62, 74, 77, 94, 107, 108, 110, 135, 142, 145, 147
24 Pgs.—Amdanga ...	1	3	40	12	30, 70, 137
Mymensingh—Jamalpur ...	2	20	210	195	6, 8, 22, 52, 55, 57, 59, 66, 69, 71, 78, 82, 83, 87, 89, 92, 95, 106, 139, 152
24 Pgs.—Deganga ...	2	7	74	58	53, 65, 70, 136, 137, 150, 157
Rangpur—Palashbari ...	3	5	52	67	62, 77, 107, 108, 110
Pabna—Belkuchi ...	3	9	61	69	21, 25, 31, 85, 109, 116, 119, 146, 158
Pabna—Kamarkhand ...	3	4	27	33	25, 29, 31, 119
Hooghly—Uttarpara ...	3	3	8	5	17, 19, 154

TARLE 5.

Distribution of Work and Workers by Size and Density of Samples and Thanas

District and Thanas	Density	No. of Workers	Estimated Square Miles	Gross Days	Serial Roll No. of Workers
(1)	(2)	(3)	(4)	(5)	(6)
16—Acre					
Rangpur—Kurigram ...	½	8	96	23	1, 43, 61, 74, 75, 81, 104, 120
Mymensingh—Trisal ...	½	4	60	20	35, 51, 58, 123
Mymensingh—Nandail ...	1	14	115	132	15, 25, 26, 50, 56, 58, 76, 86, 117, 119, 122, 123, 133A, 147
24 Pgs —Barasat ...	1	8	94	41	24, 65, 101, 132, 136, 140, 150, 156
Rangpur—Pirgachha ...	2	8	94	74	55, 71, 77, 104, 107, 108, 110, 129
36-Acre					
Mymensingh—Bajitpur ...	¼	6	80	37	50, 72, 80, 93, 96, 140
24 Pgs.—Basirhat ...	¼	8	120	43	7, 11, 36, 47, 76, 113, 132, 151
Pabna—Pabna Sadar ...	¼	6	136	60	21, 30, 102, 120, 126, 141
Pabna—Ullapara ...	½	9	166	103	25, 31, 49, 85, 109, 116, 119, 146, 158
Rangpur—Rangpur ...	½	6	180	63	32, 74, 75, 92, 141, 145
Mymensingh—Kishoreganj	1	6	63	37	34, 35, 38, 46, 51, 76
Hooghly—Serampore ...	1	3	19	15	19, 151, 154
Hooghly—Uttarpara ...	1	1	5	4	17

TABLE 6.

Period of Sample Survey in the Field 1939

Thana	Density of Grids	Date of Start	Date of Finish
(1)	(2)	(3)	(4)
Random Plots			
Netrokona	4	7 – 5	15 – 5
Serampur	4	15 – 6	9 – 8
Titagarh (Gr.)	4	25 – 7	30 – 8
Pattern	4	7 – 5	30 – 8
Kurigram	6	6 – 5	5 – 6
Dhaniakhali	6	22 – 6	3 – 7
Pattern	6	6 – 5	3—7
Rangpur	9	22 – 6	14 – 8
Pabna	9	5 – 6	3 – 8
Trisal	9	8 – 6	27 – 6
Pattern	9	5 – 6	14 – 8
Melandah	16	1 – 5	14 – 6
Nilphamari	16	7 – 5	22 – 6
Bajitpur	16	19 – 6	16 – 7
Barasat	16	20 – 7	5 – 8
Baduria	16	22 – 6	26 – 8
Chanditala	16	15 – 7	3 – 8
Tarakeswar	16	22 – 6	21 – 8
Pattern	16	1 – 5	26 – 8

TABLE 6—*Contd.*

Period of Sample Survey in the Field 1939

Thana	Density of Grids	Date of Start	Date of Finish
(1)	(2)	(3)	(4)
Random Plots			
Ullapara	25	1 – 5	12 – 7
Jamalpur	25	1 – 5	14 – 8
Pirgachha	25	21 – 6	7 – 7
Amdanga	25	21 – 7	7 – 8
Pattern	25	1 – 5	14 – 8
Belkuchi	36	7 – 5	7 – 7
Nandail	36	19 – 6	14 – 8
Kishoreganj	36	19 – 6	4 – 7
Palashbari	36	7 – 5	1 – 6
Gaibandha	36	1 – 5	14 – 8
Kamarkhand	36	7 – 5	21—5
Deganga	36	22 – 6	4 – 7
Pattern	36	1 – 5	14 – 8
1-acre			
Netrokona	1	11 – 5	18 – 5
Serampur	1	1 – 8	24 – 8
Titagarh (Gr).	1	7 – 8	14 – 8
Pattern	1	11 – 5	24 – 8
Nilphamari	2	17 – 5	30 – 6

TABLE 6—*Contd.*

Period of Sample Survey in the Field 1939

Thana	Density of Grids	Date of	
		Start	Finish
(1)	(2)	(3)	(4)
1-acre			
Pabna	2	13 – 6	3 – 8
Chanditala	2	8 – 7	17 – 7
Amdanga	2	8 – 8	14 – 8
Pattern	2	17 – 5	14 – 8
Melandah	4	24 – 5	13 – 6
Tarakeswar	4	1 – 7	7 – 7
Bashirhat	4	20 – 6	24 – 8
Gaibandha	4	26 – 5	19 – 8
Barasat	4	1 – 8	14 – 8
Uttarpara	4	4 – 8	7 – 8
Pattern	4	24 – 5	24 – 8
Belkuchi	6	6 – 6	9 – 7
Jamalpur	6	4 – 7	4 – 8
Baduria	6	1 – 7	28 – 8
Kamarkhand	6	2 – 6	14 – 6
Pattern	6	2 – 6	28 – 8
Nandail	8	21 – 6	14 – 8
Palashbari	8	22 – 6	14 – 7
Dhaniakhali	8	3 – 7	21 – 7
Pattern	8	21 – 6	14 – 8

TABLE 7.

Period of Sample Survey in the Field 1939

Thana	Density of Grids	Date of	
		Start	Finish
(1)	(2)	(3)	(4)
4-acre			
Dhaniakhali	½	19 – 7	2 – 8
Bashirhat	½	6 – 7	24 – 8
Pattern	½	6 – 7	24 – 8
Kurigram	1	15 – 5	1 – 6
Trisal	1	8 – 6	7 – 8
Deganga	1	3 – 7	12 – 7
Pattern	1	15 – 5	7 – 8
Ullapara	2	18 – 6	17 – 7
Pirgachha	2	8 – 7	31 – 7
Bajitpur	2	15 – 6	14 – 7
Melandah	2	3 – 6	15 – 6
Tarakeswar	2	27 – 7	19 – 8
Pattern	2	3 – 6	19 – 8
Kishoreganj	3	2 – 7	15 – 7
Rangpur	3	5 – 7	3 – 8
Gaibandha	3	6 – 6	21 – 8
Pattern	3	6 – 6	21 – 8

TABLE 7—*Contd.*

Period of Sample Survey in the Field 1939

Thana	Density of Grids	Date of	
		Start	Finish
(1)	(2)	(3)	(4)
9-acre			
Baduria	½	7 – 7	26 – 7
Chanditala	½	26 – 7	27 – 8
Titagarh (Gr.)	½	12 – 8	18 – 8
Pattern	½	7 – 7	27 – 8
Netrokona	1	15 – 5	20 – 5
Nilphamari	1	24 – 5	1 – 6
Amdanga	1	15 – 8	19 – 8
Pattern	1	15 – 5	19 – 8
Jamalpur	2	15 – 5	5 – 8
Deganga	2	11 – 7	27 – 7
Pattern	2	15 – 5	5 – 8
Palashbari	3	27 – 6	22 – 7
Belkuchi	3	24 – 6	7 – 7
Kamarkhand	3	14 – 6	24 – 6
Uttarpara	3	11 – 8	13 – 8
Pattern	3	14 – 6	13 – 8

TABLE 7—*Contd.*

Period of Sample Survey in the Field 1939

Thana	Density of Grids	Date of Start	Date of Finish
(1)	(2)	(3)	(4)
16-acre			
Kurigram	½	25 – 5	31 – 5
Trisal	½	27 – 6	9 – 8
Pattern	½	25 – 5	9 – 8
Nandail	1	19 – 6	14 – 8
Barasat	1	13 – 8	20 – 8
Pattern	1	19 – 6	20 – 8
Pirgachha	2	12 – 7	31 – 7
Pattern	2	12 – 7	31 – 7
36-acre			
Bajitpur	¼	4 – 7	14 – 7
Bashirhat	¼	11 – 7	24 – 8
Pabna	¼	25 – 6	16 – 7
Pattern	¼	25 – 6	24 – 8
Ullapara	½	10 – 6	21 – 7
Rangpur	½	15 – 7	8 – 8
Pattern	½	10 – 6	8 – 8
Kishoreganj	1	4 – 7	16 – 7
Serampore	1	6 – 8	24 – 8
Uttarpara	1	9 – 8	14 – 8
Pattern	1	4 – 7	24 – 8

TABLE 8.

Sample Estimates : Mean proportion of Area under Jute with Standard Errors

(Completely enumerated areas only)

Size of Grids (in acres)	Model		Field		Model		Field	
	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(2·1)	(2·2)	(3·1)	(3·2)
	MYMENSINGH : ISWARGANJ (1938)				DACCA : TEJGAON (1938)			
R. P.								
1	1756	0·4208 ± 0·0079	1739	0·3892 ± 0·0078	399	0·0859 ± 0·0099		
2·25	1377	0·3966 ± 0·0078			300	0·0785 ± 0·0100		
4	1161	0·3775 ± 0·0078	411	0·3610 ± 0·0134	249	0·0709 ± 0·0102		
6·25	951	0·3743 ± 0·0078			201	0·0731 ± 0·0105		
9	832	0·3684 ± 0·0078			174	0·0791 ± 0·0099		

TABLE 8—*Contd.*

Size of Grids (in acres)	Model		Field		Model		Field	
	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(2·1)	(2·2)	(3·1)	(3·2)
	MYMENSINGH : ISWARGANJ (1938)				DACCA : TEJGAON (1938)			
12·25	732	0·3671 ± 0·0079			150	0·0668 ± 0·0097		
16	618	0·3663 ± 0·0083	75	0·3228 ± 0·0290	126	0·0868 ± 0·0120		
25	476	0·3776 ± 0·0092			99	0·0789 ± 0·0136		
36	365	0·3686 ± 0·0097			75	0·0816 ± 0·0149		
C. E.	121·40 (sq. M)		p = 0·3743		45·56 (sq. M)		p = 0·0887	

	Tipperah : Laksam (1938)				Mymensingh : Nandail (1939)			
R. P.							4709	0·3748 ± 0·0064
1	266	0·0260 ± 0·0055	1005	0·0385 ± 0·0028	266	0·3559 ± 0·0184	1032	0·3908 ± 0·0099
2·25	200	0·0302 ± 0·0061			200	0·3218 ± 0·0193		
4	166	0·0280 ± 0·0045	191	0·0355 ± 0·0061	166	0·3714 ± 0·0191		
6·25	134	0·0347 ± 0·0049			134	0·3331 ± 0·0212		
9	116	0·0225 ± 0·0058			116	0·3554 ± 0·0201		
12·25	100	0·0324 ± 0·0056			100	0·3519 ± 0·0202		
16	82	0·0302 ± 0·0055	44	0·0308 ± 0·0189	84	0·3090 ± 0·0236	127	0·3628 ± 0·0193
25	66	0·0457 ± 0·0074			66	0·3784 ± 0·0235		
36	50	0·0375 ± 0·0053			50	0·3844 ± 0·0256		
C. E.	47·80 (sq. M)		p = 0·0310		122·96 (sq. M)		p = 0·3718	

TABLE 9.

Sample Estimates : Mean proportion of Area under Jute with Standard Errors (Completely enumerated areas only)

Size of Grids (in acres)	Model		Field		Model		Field	
	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(2·1)	(2·2)	(3·1)	(3·2)
	RANGPUR : Pirgachha (1939)				RANGPUR : Palashbari (1939)			
R. P.			2489	0·2942 ± 0·0081			2658	0·3163 ± 0·0081
1	266	0·2514 ± 0·0168			266	0·2806 ± 0·0155	589	0·3216 ± 0·0106
2·25	200	0·2686 ± 0·0143			200	0·2908 ± 0·0150		
4	166	0·2677 ± 0·0149	201	0·2990 ± 0·0147	166	0·2907 ± 0·0159		
6·25	134	0·2647 + 0·0168			134	0·2967 ± 0·0152		
9	116	0·2372 ± 0·0137			116	0·3000 ± 0·0153	221	0·3088 ± 0·0114
12·25	100	0·2961 ± 0·0160			100	0·2961 ± 0·0173		
16	84	0·2801 ± 0·0175	200	0·2695 ± 0·0103	84	0·2750 ± 0·0170		
25	66	0·2980 ± 0·0164			66	0·2891 ± 0·0164		
36	50	0·2770 ± 0·0186			50	0·2519 ± 0·0200		
C. E.	98·20 (sq. M)	p = 0·2814			74·80 (sq. M)		p = 0·3010	

	PABNA : Belkuchi (1939)				MYMENSINGH : Jamalpur (1939)			
R. P.			2184	0·1813 ± 0·0058			5317	0·2225 ± 0·0047
1	266	0·1731 ± 0·0102	365	0·2028 ± 0·0125			1313	0·2306 ± 0·0071
2·25	200	0·1780 ± 0·0107						
4	166	0·1933 ± 0·0111						
6·25	134	0·1718 ± 0·0110						
9	116	0·1965 ± 0·0127	184	0·1736 ± 0·0110			429	0·1987 ± 0·0090
12·25	100	0·1719 ± 0·0108						
16	84	0·1671 ± 0·0112						
25	66	0·1939 ± 0·0138						
36	50	0·1970 ± 0·0141						
C. E.	54·95 (sq. M)	p = 0·1839			209·56 (sq. M)		p = 0·2143	

TABLE 9.—*Contd.*

Size of Grids (in acres)	Model		Field		Model		Field	
	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(2·1)	(2·2)	(3·1)	(3·2)
	MYMENSINGH : Kishoreganj (1939)				PABNA : Kamarkhand (1939)			
R. P.			2464	0·3058 ± 0·0081			1272	0·3411 ± 0·0118
1								
4			205	0·3133 ± 0·0158			211	0 2953 ± 0·0168
9								
36			65	0·3204 ± 0·0173			104	0·3161 ± 0 0152
C. E.	71·75 (sq. M)	p = 0·3180			35·36 (sq. M)		p = 0·3041	

TABLE 10.

Sample Estimates based on Field Surveys : Mean proportion of Area under Jute with Standard Errors (1939)

Thana	N	Mean ± S.E.	N	Mean ± S.E.	N	Mean ± S.E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(4·1)	(4·2)
MYMENSINGH						
	Random Plots		1—Acre		4—Acre	
Jamalpur ...	5317	0·2225 ± 0·0047	1313	0·2306 ± 0·0071		
Nandail ...	4709	0·3748 ± 0·0064	1032	0·3908 ± 0·0099		
Kishoreganj ...	2464	0·3058 ± 0·0081			205	0·3133 ± 0·0158
Netrokona ...	424	0·1966 ± 0·0177	104	0·2156 ± 0·0254		
Trisal ...	527	0·2723 ± 0·0166			90	0·2553 ± 0·0265
Bajitpur ...	1179	0·2220 ± 0·0117			141	0·2449 ± 0·0239
Melandah ...	1452	0·2337 ± 0·0095	370	0·2595 ± 0·0141	182	0·2911 ± 0·0185

TABLE 10.—*Contd.*

Sample Estimates based on Field Surveys : Mean proportion of Area under Jute with Standard Errors (1939)

Thana	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(4·1)	(4·2)
MYMENSINGH						
	9-Acre		16-Acre		36-Acre	
Jamalpur ...	429	0·1987 ± 0·0090				
Nandail ...			127	0·3628 ± 0·0193		
Kishoreganj ...					65	0·3204 ± 0·0173
Netrokona ...	106	0·1880 ± 0·0180				
Trisal ...			44	0·2070 ± 0·0275		
Bajitpur ...					17	0·2180 ± 0·0473
Melandah ...						

RANGPUR

	Random Plots		1-Acre		4-Acre	
Pirgachha ...	2489	0·2942 ± 0·0080			201	0·2990 ± 0·0147
Palashbari ...	2658	0·3163 ± 0·0081	589	0·3216 ± 0·0106		
Rangpur ...	1098	0·2130 ± 0·0112			368	0·1865 ± 0·0081
Nilphamari ...	2166	0·3138 ± 0·0091	269	0·3381 ± 0·0191		
Kurigram ...	635	0·1673 ± 0·0132			104	0·1743 ± 0·0179
Gaibandha ...	4383	0·3106 ± 0·0063	492	0·3058 ± 0·0121	369	0·2823 ± 0·0106

	9-Acre		16-Acre		36-Acre	
Pirgachha ...			200	0·2695 ± 0·0103		
Palashbari ...	221	0·3088 ± 0·0114				
Rangpur ...					61	0·1994 ± 0·0118
Nilphamari ...	134	0·3148 ± 0·0187				
Kurigram ...			53	0·1575 ± 0·0212		
Gaibandha ...						

TABLE 11.

Sample Estimates based on Field Surveys : Mean proportion of Area under Jute with Standard Errors (1939)

Thana	N	Mean ± S.E.	N	Mean ± S.E.	N	Mean ± S.E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(4·1)	(4·2)
PABNA						
	Random Plots		1-Acre		4-Acre	
Belkuchi ...	2184	0·1813 ± 0·0058	365	0·2028 ± 0·0125		
Kamarkhand ...	1272	0·3411 ± 0·0118	211	0·2953 ± 0·0168		
Pabna ...	1258	0·1526 ± 0·0087	281	0·1506 ± 0·0125		
Ullapara ...	3999	0·2234 ± 0·0059			321	0·2562 ± 0·0104

	9-Acre		16-Acre		36-Acre	
Belkuchi ...	184	0·1736 ± 0·0110				
Kamarkhand ...	104	0·3161 ± 0·0152				
Pabna ...					33	0·1465 ± 0·0273
Ullapara ...					83	0·2068 ± 0·0158

24 Parganas

	Random Plots		1-Acre		4-Acre	
Titagarh (Gr.) ...	274	0·0219 ± 0·0082	69	0·0113 ± 0·0058		
Barasat ...	1665	0·0895 ± 0·0066	416	0·0937 ± 0·0070		
Basirhat ...			380	0·1323 ± 0·0116	49	0·1098 ± 0·0223
Deganga ...	2791	0·1117 ± 0·0056			77	0·0972 ± 0·0129
Baduria ...	1245	0·1381 ± 0·0095	459	0·1276 ± 0·0082		
Amdanga ...	1341	0·0591 ± 0·0061	107	0·0654 ± 0·0153		

TABLE 11.—*Contd.*

Sample Estimates based on Field Surveys : Mean proportion of Area under Jute with Standard Errors (1939)

Thana	N	Mean ± S. E.	N	Mean ± S. E.	N	Mean ± S. E.
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(4·1)	(4·2)
24 Parganas						
	9-Acre		16-Acre		36-Acre	
Titagarh (Gr.) ...	33	0·0133 ± 0·0054				
Barasat ...			104	0·0620 ± 0·0088		
Basirhat ...					23	0·1189 ± 0·0188
Deganga ...	155	0·0826 ± 0·0063				
Baduria ...	39	0·1589 ± 0·0261				
Amdanga ...	54	0·0319 ± 0·0074				

Hooghly

	Random Plots		1-Acre		4-Acre	
Serampore ...	81	0·0563 ± 0·0248	21	0·0910 ± 0·0422		
Chanditala ...	1006	0·1493 ± 0·0110	126	0·1578 ± 0·0226		
Dhaniakhali ...	632	0·0601 ± 0·0091	843	0·0996 ± 0·0058	52	0·0608 ± 0·0153
Tarakeswar ...	741	0·1635 ± 0·0130	185	0·1632 ± 0·0176	92	0·1535 ± 0.0206
Uttarpara ...			42	0·0286 ± 0·0162		

	9-Acre		16-Acre		36-Acre	
Serampore ...					18	0·0702 ± 0·0213
Chanditala ...	31	0·1039 ± 0·0253				
Dhaniakhali ...						
Tarakeswar ...						
Uttarpara ...	29	0·0538 ± 0·0251			10	0·0372 ± 0·0176

TABLE 12.

Values of "t"—Statistic based on Comparison with Complete Enumeration

Size of Grids (in acres)	Model		Field		Model		Field	
	D.F.	t	D.F.	t	D.F.	t	D.F.	t
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(2·1)	(2·2)	(3·1)	(3·2)
	PABNA : BELKUCHI (1939)				MYMENSINGH : JAMALPUR (1939)			
R.P.			2183	0·45			5316	1·75
1	265	1·06	364	1·51			1312	*2·30
2·25	199	0·55						
4	165	0·85						
6·25	133	1·10						
9	115	0·99	183	0 94			428	1·73
12·25	99	1·11						
16	83	1·50						
25	65	0·72						
36	49	0·93						

	Mymensingh : Kishoreganj (1939)				Pabna : Kamarkhand (1939)			
R.P.			2463	1·51			1271	*3·14
1							210	0·52
2·25								
4			204	0·30				
6·25								
9							103	0·79
12·25								
16								
25								
36			64	0·14				

TABLE 13.

Values of "t"—Statistic based on Comparison with Complete Enumeration

Size of Grids (in acres)	Model		Field		Model		Field	
	D. F.	t	D. F.	t	D. F.	t	D. F.	t
(1)	(2·1)	(2·2)	(3·1)	(3·2)	(2·1)	(2·2)	(3·1)	(3·2)
	Mymensingh : Iswarganj (1938)				Dacca : Tejgaon (1938)			
R. P.								
1	1755	*5·89	1738	1·92	398	0·28		
2·25	1376	*2·86			299	1·02		
4	1160	0·41	410	0·99	248	1·75		
6·25	950	0·00			200	1·49		
9	831	0·76			173	0·97		
12·25	731	0·91			149	*2·26		
16	617	0·96	74	1·78	125	0·16		
25	475	0·86			98	0·72		
36	364	0·59			74	0·50		