and yet, (*b*) the agreement between mean values may be quite satisfactory so that reliable sample estimates can be made with confidence.

Bias in the estimation of crop yields

It may be mentioned that in crop-cutting work the principle of random sampling was explicitly recognized for the first time in the classical experiments of J. (now Sir John) Hubback of the I.C.S. in India between 1923 and 1925. He used sampling units of a very small size, 1/3,200 acre, in the form of a triangular cut obtained by a special implement devised by him. He was the first person to point out that owing to high correlation between yields in adjoining parts of the same field there was not much gain in precision by using sample units of a large size. He further emphasized the convenience and economy of using small cuts, of which a much larger number can be collected in the same time by investigators moving from one place to another. He gave a general account of his work in a paper on "Sampling for Rice Yield in Bihar and Orissa" published by the Government of India in 1927 as Bulletin No. 166 of the Imperial Agricultural Institute, Pusa. C. D. (now Sir Chintaman) Deshmukh, of the I.C.S., used Hubback's method a little later between 1928 and 1931 in his crop-cutting work on rice conducted together with settlement operations in two districts in the Central Provinces. About the same time, P. S. Rau, I.C.S., used the same method in two other districts of the same province. I succeeded recently in recovering the original material collected by C. D. Deshmukh, but all papers relating to P. S. Rau's work appear to have been destroyed. As Hubback's paper has been out of print for a long time, it has been reprinted in *Sankhyā* (the Indian Journal of Statistics), Vol. 7, Part 3, pp. 281–94, and in the same number I have given a brief account of Deshmukh's work. In 1938–39 H. P. V. Townend (also of the I.C.S.) used cuts of a comparatively small sampling unit (27·04 sq. ft.) in the form of a square, harvested with the help of a rigid frame, in his work on rice in certain districts in Bengal. It is interesting to observe that four administrators belonging to the I.C.S. used cuts of a small size evidently because they were guided by reasons of convenience and economy.

It is worth mentioning that R. A. Fisher's crop-cutting work on wheat at Rothamsted was influenced by that of Hubback, as stated by Fisher in a memorandum on crop estimating surveys submitted by him to the Imperial Council of Agricultural Research in India on March 2nd, 1945. Subsequently, Yates and others used sample cuts of a small size.

When we first started work on the estimation of crop yields in 1939, we naturally thought of using cuts of a small size. At the same time I thought it advisable to investigate whether the size of the sample unit had any influence on the results. It was necessary at this stage to distinguish between two different possible effects. If the size of the sample unit is appreciable in comparison with the size of the field, then there will be obviously a chance of over-sampling the central area of the field. I have discussed such over-sampling elsewhere (in the paper "On Large-Scale Sample Surveys," *Phil. Trans.* Vol. 231(B), No. 584, Appendix 6, p. 404). There is a second possibility of bias arising from bordering plants being included within the sample cut. It was this second effect which I had in mind in undertaking an experimental study of the problem.

In our very first series of crop-cutting experiments on jute in Bengal in 1939 we collected some material with sampling units of five sizes ranging between 5·5 sq. ft. and 66 sq. ft. There was some evidence of over-estimation with small cuts, but the material was meagre. In 1940 we used five different sizes of sample cut. A square of size 16 ft. by 16 ft. was located at random in each field (which also was selected at random) and the crop was harvested in the form of a number of separate sub-cuts supplying yield rates for sizes 1 ft. × 1 ft., 3 ft. × 3 ft., 12 ft. × 4 ft., 12 ft. × 12 ft., and 16 ft. × 16 ft. Cuts of different sizes were intentionally obtained from the same spot so as to secure effective local control for purposes of comparison. Information was collected for many items, such as number of plants per acre, the weight of green plants immediately after harvesting, and also the final yield of dry fibre per acre (after the plants were retted, dried, and the fibre extracted). I am giving in Table 9 figures for one particular item—namely, the yield of jute expressed as the weight of green plants (immediately after harvesting) in pounds per acre collected from four different districts in Bengal in 1940.

Yield rates based on 1 ft. × 1 ft. (1 sq. ft.) sample cuts were clearly over-estimates. Such extremely small sizes were therefore omitted in subsequent years. The next larger size 3 ft. × 3 ft. (9 sq. ft.) also appeared to give over-estimates, but this size was retained for purposes of comparison. Since 1940 similar material has been collected year by year, for a number of crops like jute, wheat, on an extensive scale by a large field staff scattered over the whole province of Bengal, comprising about 60,000 or 70,000 square miles. Secondly, it is equally clear that there is practically no bias when the size of the cut is increased to 50·27 sq. ft., which is in good agreement with previous observations.

Along with extensive experiments carried out by the ordinary field staff, arrangements were also made to repeat the work at a few selected centres by experienced investigators working under the direct supervision of trained statisticians. Illustrative results are given below in Table 12, which shows the yield in lbs. per acre of *aman* (winter) rice at three centres in Bengal in 1945–46. The numerical values show that the systematic over-estimation with cuts of 2 ft. radius is no longer appreciable. This is confirmed by Student's paired "*t*" values of which are given at the bottom of the table.

TABLE 12

Bengal Crop Survey, Aman (*winter*) *rice*, 1945–46. *Yield in lb. per acre for different sizes of cuts*

Symbol	Size of sample cut in sq. feet	Mean yield in lb. per acre with s.e.		
		Katwa (n = 124)	Gouripur (n = 64)	Sainthia (n = 48)
x_1	$\pi(2')^2$ = 12·57	1,251 ± 33	726 ± 25	1,593 ± 46
x_2	$\pi(4')^2$ = 50·27	1,283 ± 26	747 ± 20	1,518 ± 31
x_3	$\pi(5\frac{2}{3}')^2$ = 100·88	1,304 ± 26	761 ± 19	1,523 ± 28
x_4	$\pi(8')^2$ = 201·06	1,340 ± 29	764 ± 18	1,495 ± 24
Values of Student's paired "*t*"				
$x_1 - x_2$	—	1·53	1·33	2·18 *
$x_2 - x_3$	—	2·55 *	1·81	0·43
$x_3 - x_4$	—	3·07 +	0·63	2·47 *

* Significant at 5 per cent. level. + Significant at 1 per cent. level.

We thus find that when the work is done under adequate statistical supervision, the over-estimation with cuts of a small size becomes practically negligible. In view of this finding, I believe that unconscious pulling in of bordering plants may not be the only, or even the most important factor in this matter. In fact, I am now inclined to adopt a line of explanation offered by F. Yates in his paper on bias in sampling in the *Annals of Eugenics* in 1936. Discussing the observed bias in crop yields harvested from within hoops (of area 10 sq. ft.) supposed to have been thrown at random on fields in the United Provinces, Yates suggested:

> "The bulk of the bias, however, is probably due to the tendency, conscious or unconscious, to cast the hoop on the good parts of the crop."

It is likely that there are patches of greater fertility distributed either in a random manner or in a mildly patterned fashion over each field. In locating the sample-cut, the investigator may unconsciously tend to favour such more fertile patches by shifting about the "random point" to some extent. From the fact that the bias appears to become negligible beyond 40 or 50 sq. ft., it would seem, without entering into detailed investigations, that the average size of such patches would be probably fairly small and of the order of only a few square feet. Unconscious pulling in of bordering plants may also be a contributing factor.

Another point is perhaps just worth mentioning. In case there is any bias in favour of locating the sample cuts on or near more fertile patches, then one would expect over-estimation with the smaller cuts to be followed by compensating under-estimation in cuts of intermediate size. Some of the figures of Table 10 may perhaps be indicative of this, but more precise and detailed experimentation is necessary to settle this point.

A certain amount of loss of grain in the actual process of harvesting, threshing, drying, weighing, etc., is inevitable which will necessarily introduce a component of "error." The proportional error arising from this factor may be independent of or may increase or decrease with the size of the cut. In case the loss of grain is proportionally smaller with small cuts (owing to the greater care with which the different processes can be carried out), then there might be an apparent over-estimation with small cuts, which, however, would really indicate an under-estimation with large cuts.

grid is increased. In crop-cutting work on jute it was found, for example, that mean values for all the characters studied (such as number of plants per acre, weight of green plants, weight of dry fibre) were much higher for sample units of small size, so that it was not at all safe to work with cuts of a size less than say 25 sq. ft."

At that time the idea I had in mind was that the unconscious pulling in of bordering plants was probably effective so long as the linear dimensions of the sample cuts were smaller than the human span.* In 1944 Jitendra Mohan Sen Gupta of the Statistical Laboratory suggested the use of circular-shaped cuts to be harvested with the help of an arm rotating over a pivot with a light stylus attached at the end of the rotating arm to catch the plants. This proved very convenient in practice, and various improvements were made. The latest model is now a standard implement in our work. It is the usual practice at the present time to take three (or four) concentric circular cuts on the same field. This is done by drawing out the rotating (radius) arm to three (or four) fixed lengths one after another, and harvesting the appropriate cuts. In this way yield rates are obtained for three or four different sizes.

TABLE 11

Bengal Crop Survey, 1945–46. *Yield in lb. per acre and index-number of yields of* aman *rice (not in husk) for circular cuts of different sizes*

Serial no.	Name of district	No. of cuts	Yield in lb. per acre			Index-number of yield with yield based on largest size of cut as 100	
			12·57 sq. ft.	50·27 sq. ft.	100·88 sq. ft.	12·57 sq. ft.	50·27 sq. ft.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Bakarganj	180	961·1	880·5	891·2	107·8	98·8
2	Bankura	109	750·4	686·3	723·3	103·8	94·9
3	Birbhum	87	947·1	864·0	883·8	107·2	97·8
4	Bogra	76	728·2	639·4	637·7	114·2	100·3
5	Burdwan	122	801·5	724·9	723·3	110·8	100·2
6	Chittagong	25	1,156·1	935·6	924·9	125·0	101·1
7	Dacca	113	922·4	734·0	758·7	121·6	96·7
8	Dinajpur	218	933·1	819·6	822·0	113·5	99·7
9	Faridpur	36	1,036·0	655·8	597·4	173·4	109·8
10	Hoogly	45	706·8	613·0	616·3	114·7	99·5
11	Howrah	38	604·8	420·5	401·6	150·6	104·7
12	Jalpaiguri	60	655·8	578·5	579·3	113·2	99·9
13	Jessore	86	723·3	613·8	621·3	116·4	98·8
14	Khulna	91	832·7	741·4	750·4	111·0	98·8
15	Malda	73	799·8	738·9	749·6	106·7	98·6
16	Midnapur	249	767·7	703·5	700·2	109·6	100·5
17	Murshidabad	55	870·6	785·8	777·6	112·0	101·1
18	Mymensingh	263	766·9	671·4	679·7	112·8	98·8
19	Nadia	38	680·5	598·2	613·8	110·9	97·4
20	Noakhali	58	789·9	701·1	669·0	118·1	104·8
21	Pabna	55	512·6	431·2	429·5	119·3	100·4
22	Rajshahi	83	805·6	687·1	665·7	121·0	103·2
23	Rangpur	204	794·1	715·9	715·9	110·9	100·0
24	Tipparah	92	704·4	598·2	610·6	115·4	98·0
25	24-Parganas	113	706·0	602·3	618·0	114·2	97·5
	Total	2,569	812·2	704·4	706·8	—	—
	Percentage	—	114·9	99·7	100	117·3	100·1
				weighted		unweighted	

Illustrative results of certain recent experiments are given in Table 11. Two things are clear from the above table. Circular cuts of 12·57 sq. ft. are still over-estimating when the work is done

* Dr. P. V. Sukhatme in a letter to *Nature* of May 11th, 1946, has published some material showing the over-estimation of crop yields with cuts of a small size. He writes: "The reason for over-estimation appears to be the human tendency to include border plants inside the plot. This factor becomes serious when the perimeter of the plot is large in proportion to its area." This is exactly what I had suggested in the paper written in 1942. Dr. Sukhatme has not referred to my paper, or is not aware of the observations made by me much earlier. I have recently sent a letter to *Nature* on this subject.

It is pertinent to observe in the present connection the danger of under-estimation in using cuts of a comparatively large size which are demarcated by pegs and ropes on the field. It is doubtful whether ordinary investigators can make accurate measurements of distance on the field. Any sagging of the ropes is bound to reduce the real size of the cut, and hence lead to under-estimation. There are serious difficulties in deciding the allowance to be made for the boundary ridges or demarcations between adjoining fields, a point which is of particular importance in a province like Bengal, where the average size of the field is very small and less than half an acre, so that the perimeter is large in comparison with the area of the field. In fact, it is difficult to define the " whole field " in an unambiguous manner; and until this is done the concept of total production must also remain to some extent vague. This subject clearly requires much further study.

In the meantime I may perhaps mention that the over-estimation of cuts of a very small size can itself be used as an excellent control in the following way. As already mentioned, yield rates are being obtained by the Statistical Institute for three (or four) concentric circular cuts of different sizes on the same plot. If the work is done honestly by ordinary investigators, then yield rates based on sample cuts of the smallest size would show over-estimation in comparison with results based on sample cuts of larger size. Results based on the two (or three) larger cuts should, however, be in good agreement. At the same time, the variance of the yield per acre should decrease with increasing sizes of sample cuts. Besides being extremely convenient and economical, concentric circular cuts thus enable a two-fold control (one over mean rates of yield and the other over variances) to be exercised at the point of collection of the primary material.

Bengal Labour Enquiry: Jagaddal 1941, 1942, 1945

As already mentioned, in socio-economic surveys special efforts have been made to study the personal equation or bias of the investigating staff. Here also the use of replicated sub-samples proved to be of great value. I shall give illustrative examples from certain enquiries carried out in 1941, 1942, and 1945 in an industrial area at Jagaddal, about 20 miles to the north of Calcutta. The Government of Bengal had asked us to survey family budgets, housing, and other economic conditions of factory workers in this area.

Design of the survey. The general approach can be best studied from the design of the enquiry of 1941. In order to ensure proper randomization, the addresses of working-class employees were obtained from the factories located in this area, on the basis of which a preliminary survey was made of the geographical distribution of working-class families at Jagaddal, which is a municipal area. The number of families residing in different groups of buildings was noted on large-scale maps. A sample check was also carried out, but no attempt was made to attain high precision in the preliminary survey, as this was not considered to be of great importance.

On the basis of the preliminary survey the area was divided into five blocks (which were of unequal size and irregular in shape and), which were demarcated on certain general considerations. For example, blocks nos. 1, 2, and 3 were located in the immediate vicinity of factories; block no. 4 was within the town, but was situated at a greater distance from the factories themselves; finally, block no. 5 covered the area outside municipal limits. For present purposes the first three blocks were most important; block no. 4 of moderate importance, while block no. 5 was included more or less as a control and for comparative purposes. The buildings from which the families were to be surveyed were picked out purely at random on maps.

The sample units within each block were divided into 5 equal sub-samples, each of which was independently random and covered the whole of the block. These five sub-samples thus constituted five independent networks of sample units within each block. Each such sub-sample was assigned to a different investigator. In this way information for each block was collected by five investigators, all of whom in consequence worked in all the blocks The number of families surveyed in 1941 and 1942 by the different investigators in different blocks are shown in Table 13.

Possibilities of analysis of variance. The interesting point is that the design of the survey makes it possible to carry out a Fisherian analysis of variance. For example, the 1941 and 1942 designs resemble in appearance Fisherian Latin squares. These are, however, not true Latin squares, but only analogues, as the sub-samples are not geographically distinct. In fact there is no intrinsic difference between the Latin square and the randomized block, so that an analysis of variance can be carried out for any two or more or all the blocks at the same time.

The unequal numbers in the different block-investigator cells introduce a difficulty. There are

TABLE 13

Bengal Labour Enquiry—Jagaddal, 1941–42. *Number of families surveyed by blocks and investigators*

Blocks	Investigators					Total
	1	2	3	4	5	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			1941 survey			
1	33	33	33	31	27	157
2	36	36	31	31	34	168
3	37	36	45	34	29	181
4	20	18	20	23	18	99
5	7	9	8	6	7	37
Total	133	132	137	125	115	642
			1942 survey			
1	41	37	40	39	38	195
2	41	36	46	39	35	197
3	38	42	39	41	37	197
4	19	24	19	24	21	107
5	9	9	8	8	10	44
Total	148	148	152	151	141	740

three possibilities. If the first three blocks are selected the numbers are not very different, so that a simple analysis of variance on orthodox lines would give approximate results. A second possibility (for the same three blocks) is to equalize the number in each cell by rejecting an appropriate number of schedules at random from each cell. Finally, the method of fitting constants can of course be used, but would involve a great deal of computational labour.

The results of an analysis of variance (for the first three blocks) for the family budget enquiry at Jagaddal in 1942 are given for a few selected items in Table 14. The total number of household schedules was 589 with 588 degrees of freedom, which can be split up into two degrees of freedom for comparisons between the three blocks, four degrees of freedom for comparisons between five investigators, and eight degrees of freedom for the interaction making up altogether fourteen degrees of freedom for comparisons between sub-samples.

In the standard Fisherian analysis of variance it is usual to use the interaction as error to test the significance of difference between blocks, etc. In the present case a direct estimate of the error can be made from the variations "within block-investigators cells" (that is, within each portion of the sample for which the material is collected by the same investigator in the same geographical area) based on 574 degrees of freedom.

It is also possible to carry out a similar analysis after equalizing the frequencies in each cell by rejecting an appropriate number of schedules at random. This was done, and the results are given in Table 14 for convenience of comparison. The total number of schedules used for this purpose was 525, so that the total variance was based on 524 degrees of freedom. Deducting fourteen degrees of freedom for comparisons among the sub-samples, this left 510 degrees of freedom for estimating the variance within sub-samples.

The ratios of variance with unequal as well as equalized frequencies in each cell are shown in the lower half of Table 14. The results are broadly similar, but equalization of frequencies makes the comparison more stringent, as a number of ratios are significant here, but not in the analysis with unequal frequencies.

Several points are worth noting. Firstly, it will be noticed that, except in the case of monthly expenditure on cereals, the interaction (blocks × investigators) variances are not significant, showing that the Fisherian assumption of the interaction variance being about the same as the "true" within-cell variance is broadly confirmed.

TABLE 14

Bengal Labour Enquiry—Jagaddal, 1942. *Analysis of variance*

(3 blocks × 5 investigators)

Sources of variation	D.f.	Age in years	Expenditure in rupees per month per capita			Consumption of cereals in lbs. per head per month
			Total	Food	Cereals	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Values of variance						
(*a*) Actual (unequal) number of families per cell ($n = 589$)						
Between blocks	2	76·1	606·5	23·4	0·08	82·5
Between investigators	4	211·6	220·4	19·5	1·3	61·8
Blocks × investigators ...	8	134·1	114·1	11·2	1·9	212·4
Between sub-samples	14	147·9	214·8	15·3	1·5	151·1
Within sub-samples	574	112·7	172·9	10·2	0·6	97·3
Total	588	113·5	173·9	10·3	0·6	98·6
Mean value	—	38·84	19·16	7·80	3·05	35·59
Standard deviation	—	9·71	11·41	3·15	0·73	8·64
Coefficient of variation	—	25·0	59·5	40·4	23·9	23·8
(*b*) Equalized number of families per cell ($n = 525$)						
Between blocks	2	62·13	805·52	36·5	0·07	79·6
Between investigators	4	304·84	275·78	22·3	1·25	114·7
Blocks × investigators ...	8	78·47	129·38	9·6	1·47	152·8
Between sub-samples	14	140·81	267·80	16·8	1·21	132·0
Within sub-samples	510	127·74	168·12	9·9	0·49	100·3
Total	524	128·09	170·78	10·1	0·51	100·8
Mean value	—	39·31	20·86	8·15	3·09	35·96
Standard deviation	—	11·32	13·07	3·18	0·71	10·04
Coefficient of variation ...	—	28·7	62·6	39·0	22·9	27·9
Ratios of variance						
(*a*) Actual (unequal) number of families per cell ($n = 589$)						
Between blocks	2	0·68	3·51 *	2·29	0·13	0·85
Between investigators	4	1·88	1·27	1·91	2·17	0·63
Blocks × investigators ...	8	1·19	0·66	1·10	3·17 †	2·18
Between sub-samples	14	1·31	1·24	1·50	2·50 †	1·55
Within sub-samples	574	—	—	—	—	—
Total	588	—	—	—	—	—
(*b*) Equalized number of families per cell ($n = 525$)						
Between blocks	2	0·49	4·79 †	3·69 *	0·14	0·80
Between investigators	4	2·39 *	1·64	2·26	2·55 *	1·15
Blocks × investigators ...	8	0·61	0·77	0·98	3·00 †	1·53
Between sub-samples	14	1·10	1·59	1·70	2·47 †	1·32
Within sub-samples	510	—	—	—	—	—
Total	524	—	—	—	—	—

* Significant at 5 per cent. level. † Significant at 1 per cent. level.

The significant value itself was further analysed, block by block and investigator by investigator, by Janardan Poti and K. C. Cheriyan; and it was found that the abnormally high values were due to one single investigator in one particular block. This shows the possibilities of deeper analysis offered by replicated inter-penetrating samples.

Investigator differences were insignificant in two cases, and significant at 5 per cent. level in two cases. Personal equations had not been completely eliminated but their influence was not large.

Nagpur labour enquiry 1942–43. I am giving a second example from a labour enquiry carried out in Nagpur in 1942–43 by M. P. Shrivastava, at that time a worker of the Institute, under the technical guidance of the Indian Statistical Institute. In this case the design was arranged in the form of a randomized block of five zones and four investigators with practically the same number of family schedules—namely, fifty in each block-investigator cell. The analysis of variance is shown in Table 15.

TABLE 15

Nagpur Family Budget Enquiry, 1943. *Analysis of variance*

(5 zones × 4 investigators)

Sources of variation	D.f.	Total income	Monthly expenditure		
			Total	Food	Cereals
(1)	(2)	(3)	(4)	(5)	(6)
		(*a*) Values of variance			
Between zones	4	4,439·58	3,707·91	708·41	206·83
Between investigators	3	85·43	597·08	77·09	3·70
Zones × investigators	12	382·54	397·28	177·75	49·80
Between sub-samples	19	1,189·74	1,127·07	237·58	75·61
Within sub-samples	977	401·57	384·71	84·73	24·99
Total	996	424·67	398·87	88·33	25·95
Mean values (in rupees)	—	36·09	34·96	20·09	11·41
Standard deviations	—	18·60	16·12	8·53	4·96
Coefficients of variation	—	51·5	46·1	42·5	43·5
		(*b*) Ratios of variance			
Between zones	4	11·06 †	9·64 †	8·36 †	8·28 †
Between investigators	3	0·21	1·55	0·91	0·15
Zones × investigators	12	0·95	1·03	2·10 *	2·00 *
Between sub-samples	19	2·96 †	2·93 †	2·80 †	3·02 †
Within sub-samples	977	—	—	—	—
Total	996	—	—	—	—

* Significant at 5 per cent. level. † Significant at 1 per cent. level.

It will be noticed that in this case the sub-samples are all significantly differentiated. Block differences are also highly significant in each case.

It may be mentioned here that in the Nagpur enquiry the blocks had been deliberately demarcated on lines which could almost certainly be expected to show large differences in income and conditions of living. For example, one block consisted almost exclusively of employees of the Bengal Nagpur Railway Workshop, who certainly earned higher wages and had generally a higher standard of living. The other blocks also were differentiated in many ways.

It is, however, most satisfactory to find that investigator's differences were negligible in every case, showing that personal equations had been completely eliminated. It is only by an analysis of the present kind that one can be certain on this point.

Margin of error of Cost of Living Index. Similar studies are also possible in regard to the margin of error of the index-number of cost of living. For example, the C.L. index-number can be calculated separately for each block-investigator cell. Adopting 1941 consumption pattern as base and 1942 and 1945 prices, the results are given in Table 16. In this table the number of 1941 family schedules on which each index-number is based is given within brackets. From the values of the index in different cells it is possible to calculate the standard error of the marginal index-numbers separately for each block or for each investigator; and finally, in the same way, the standard error of the overall index number. These are shown in Table 16.

It will be noticed that the 1942 mean index is 122 with a standard error of about 0·4 (which is only about 0·3 per cent. of the mean value). The standard errors of marginal index numbers are quite small and usually less than 1.

The 1945 C.L. index has a standard error of 1·6 or about 0·5 per cent. of the mean index-number 273. Standard errors in 1945 were, however, much higher in comparison with standard errors in

TABLE 16

Bengal Labour Enquiry. Jagaddal cost-of-living index 1942 *and* 1945, *with* 1941 *as base by blocks and investigators (with size of sample within brackets)*

Blocks	Investigators 1	2	3	4	5	Total: Size of sample in '41	Mean index with s.e.
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			1942 cost-of-living index				
1	120 (33)	122 (33)	122 (33)	124 (31)	124 (27)	157	122 ± 0·67
2	122 (36)	120 (36)	121 (31)	124 (31)	118 (34)	168	121 ± 1·88
3	121 (37)	123 (36)	121 (45)	121 (34)	121 (29)	181	121 ± 0·36
4	119 (20)	122 (18)	120 (20)	123 (23)	121 (18)	99	121 ± 0·65
5	121 (7)	124 (9)	124 (8)	124 (6)	124 (7)	37	123 ± 0·53
n (1941)	(133)	(132)	(137)	(125)	(115)	642	—
Mean index	121 ± 0·46	122 ± 0·57	121 ± 0·41	123 ± 0·41	121 ± 1·04	—	122 ± 0·33
			1945 cost-of-living index				
1	270 (33)	280 (33)	275 (33)	271 (31)	279 (27)	157	276 ± 2·0
2	263 (36)	265 (36)	284 (31)	269 (31)	267 (34)	168	269 ± 3·6
3	264 (37)	274 (36)	278 (45)	272 (34)	269 (29)	181	272 ± 2·4
4	263 (20)	274 (18)	271 (20)	297 (23)	263 (18)	99	275 ± 6·2
5	260 (7)	279 (9)	296 (8)	274 (6)	284 (7)	37	279 ± 6·2
n (1941)	(133)	(132)	(137)	(125)	(115)	642	
Mean index	265 ± 1·38	273 ± 2·56	279 ± 2·69	276 ± 4·57	271 ± 2·87	—	273 ± 1·64

1942. This is partly accounted for by the fact that whereas price relatives had varied very little among themselves (roughly about 10 per cent. of their mean values) in 1942, price relatives had varied among themselves about 80 per cent. of their mean values in 1945.

Analysis of variance has been also carried out both for five blocks × five investigators and for three blocks × five investigators (for the three blocks in which the number of schedules are nearly equal). The results are shown in Table 17.

TABLE 17

Bengal Labour Enquiry—Jagaddal, 1941, 1942, 1945.
Analysis of variance: cost-of-living index for 1942 *and* 1945, *with* 1941 *as base*

Source of variation	5 blocks × 5 investigators: D.f.	Variance	Ratio	3 blocks × 5 investigators: D.f.	Variance	Ratio	Expected ratio: 5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				1942 with 1941 as base				
Between blocks	4	5·25	2·40	2	2·60	0·80	4·46	8·65
Between investigators ...	4	4·50	2·12	4	2·06	0·63	3·84	7·01
Residual	16	2·12	—	8	3·27	—	—	—
Total	24	3·04	—	14	2·83	—	—	—
				1945 with 1941 as base				
Between blocks	4	59·75	0·81	2	37·80	1·54	4·46	8·65
Between investigators ...	4	194·00	2·64	4	69·00	2·81	3·84	7·01
Residual	16	73·44	—	8	24·55	—	—	—
Total	24	91·25	—	14	39·14	—	—	—

It is seen that block and investigator differences were both insignificant for both 1942 and 1945. This is satisfactory and gives confidence in using the present index-numbers.

It is of course also possible to study the cost-of-living index for different groups of items separately in each block-investigator cell. Illustrative results are given in Table 18.

TABLE 18

Bengal Labour Enquiry—Jagaddal, 1941, 1942, 1945. *Cost-of-living index by block × investigator*

Block no.	Investigator no.	Food		Clothing		Fuel and light		Miscellaneous		Total	
		1942	1945	1942	1945	1942	1945	1942	1945	1942	1945
I	1	120	276	159	310	126	258	114	275	120	270
	2	121	279	159	304	129	285	117	293	122	280
	3	122	277	160	310	127	277	119	285	122	275
	4	125	280	156	301	130	289	120	271	124	271
	5	120	280	160	314	125	275	125	287	124	279
	Total	122	280	159	308	127	276	119	283	122	275
II	1	123	273	159	306	127	280	114	253	122	263
	2	121	270	158	304	126	289	114	262	120	265
	3	122	279	155	297	125	267	117	303	121	284
	4	125	272	156	305	127	283	120	271	124	269
	5	121	270	156	304	126	281	113	273	118	267
	Total	122	273	157	304	126	280	116	273	121	269
III	1	122	275	156	297	129	293	114	249	121	264
	2	127	277	161	304	124	281	114	279	123	274
	3	121	280	160	306	125	288	117	289	121	278
	4	121	281	153	298	125	283	115	266	121	272
	5	122	269	159	305	125	280	115	278	121	269
	Total	123	277	158	302	126	286	115	272	121	272
IV	1	119	272	153	277	126	280	114	256	119	263
	2	124	274	158	289	126	285	115	292	122	274
	3	121	274	157	242	131	303	114	290	120	271
	4	122	306	161	313	124	273	120	311	123	297
	5	124	275	154	302	123	274	112	253	121	263
	Total	122	281	157	293	126	283	115	283	121	275
V	1	120	266	159	304	128	259	118	281	121	260
	2	124	288	158	303	127	266	119	292	124	279
	3	119	273	162	312	125	266	123	340	124	296
	4	123	283	158	315	125	240	120	280	124	274
	5	125	280	163	309	131	276	114	315	124	284
	Total	122	277	160	308	127	264	119	308	123	279
I–V	1	121	274	157	303	127	278	114	258	121	265
	2	123	276	159	304	126	284	115	280	122	273
	3	121	278	159	296	126	282	118	293	121	279
	4	124	283	157	305	127	282	119	277	123	276
	5	123	276	158	307	126	278	117	278	121	271
	Total	122	277	158	303	126	280	117	276	122	273

N.B.—No change in "rent and tax."

There are many other possibilities of breakdowns. For example, the cost-of-living index can be studied by groups of items for different over-all size of households. Relative figures are given in Table 19, from which it would be noticed that the change in the cost-of-living index-number did not depend appreciably on the size of the household.

Similar detailed figures for the change in the cost of living in 1942 and 1945 for different families in different expenditure levels are given in Table 20. Here also the cost of living appears to have changed more or less in the same way in all expenditure levels. It must be remembered, however, that the enquiry was restricted to working-class families in which the consumption pattern was fairly homogeneous.

TABLE 19

Bengal Labour Enquiry—Jagaddal, 1941, 1942, *and* 1945. *Cost-of-living index by size of family*

Size of family in 1941	No. of families in 1941	Food		Clothing		Fuel and light		Miscellaneous		Total	
		1942	1945	1942	1945	1942	1945	1942	1945	1942	1945
1	266	122	274	160	310	124	274	117	272	121	270
2	94	124	275	156	298	127	286	114	285	123	273
3	69	123	280	157	299	127	287	114	275	122	272
4	70	122	286	156	297	129	288	117	281	122	277
5	56	121	285	154	296	127	284	118	287	122	279
6	34	122	275	157	297	129	282	118	285	122	271
7	19	124	280	155	295	131	284	120	286	125	275
8	13	122	280	157	302	132	294	126	284	124	274
9	12	123	285	154	294	131	273	120	292	124	278
10	1	125	294	160	324	119	228	119	272	123	270
11	3	124	288	159	300	127	304	124	266	127	279
13	3	125	279	157	296	138	316	128	286	128	278
14	1	124	284	161	306	139	323	119	346	126	236
Total	641	122	277	158	305	126	280	117	276	122	273

N.B.—No change in " rent and tax."

TABLE 20

Bengal Labour Enquiry—Jagaddal, 1941, 1942, *and* 1945. *Cost-of-living index by expenditure levels in rupees per month per family*

Expenditure levels in 1941	No. of families in 1941	Food		Clothing		Fuel and light		Miscellaneous		Total	
		1942	1945	1942	1945	1942	1945	1942	1945	1942	1945
0– 9 ...	22	123	276	156	308	126	285	121	301	123	282
10–19 ...	189	122	275	158	306	126	201	114	281	121	268
20–29 ...	199	122	275	158	306	125	278	115	273	121	270
30–39 ...	107	123	277	157	301	126	279	117	268	122	269
40–49 ...	56	122	291	157	301	128	285	118	279	122	281
50–59 ...	30	122	281	161	305	129	286	123	276	124	273
60 & above	38	124	276	161	304	129	284	123	277	125	272
Total ...	641	122	277	158	305	126	280	117	276	122	273

N.B.—No change in " rent and tax."

Comparison of consumption in 1941 *and* 1945. As already mentioned, family budget enquiries were repeated in the same area in 1941 and also in 1942 and again in 1945. This enables a direct comparison being made of levels of consumption in different years. Somewhat similar material also happens to be available for Calcutta middle-class families in 1939 and in 1945. As prices and consumption patterns had not changed very much in Bengal between 1939 and 1941, the above two series can also be used to compare (at least approximately) the changes in patterns of consumption in working-class and middle-class families.

Relevant data are given in Table 21. The name of the item is given in col. 1, and the unit of measurement in col. 2. The over-all *per capita* consumptions in 1945 and 1941 for working-class families are given in cols. 3 and 4 respectively. The consumption in 1945 expressed as a percentage of the corresponding consumption in 1941 is given in col. 5; and the price ratios in col. 6. Similar data for *per capita* consumptions in 1945 and 1939 in middle-class families are given in cols. 7 and 8. The ratio of consumptions is given in col. 9, and price ratios in col. 10.

I should mention here that food began to be rationed in 1943. The ration of rice was somewhat restricted owing to shorter supply, and the price control was arranged in such a way as to make " wheat " and " wheat products " comparatively cheaper. Sugar was strictly rationed, and there was a partial rationing of oil.