Econometrica, Vol. 28, 2 (April 1960)

A METHOD OF FRACTILE GRAPHICAL ANALYSIS

BY P. C. MAHALANOBIS

Fractile Graphical Analysis is a new method of statistical analysis which provides an effective summary of information particularly useful in situations where the data do not permit a description in terms of a few parameters relating to the distribution; it also provides a graphical way of testing differences between groups. This method can be used for any variate which can be ranked. In this paper, the use of this method is illustrated for the comparison of economic data—relating either to the same population at different points of time or to different populations—by means of examples taken from the Indian National Sample Survey.

THIS PAPER provides some examples of the use of fractile graphical analysis, a new method for the comparison of economic data relating to the same population over time or to any two populations that differ as to geographical region or in any other way. This method can be used for any variate that can be ranked, and is based on certain theoretical conjectures. Asymptotic, but not exact proofs are available for some of these conjectures and results of model sampling experiments have been found to be in accordance with them.

1. THE METHOD

1.1. In the National Sample Survey of India much economic and demographic data are collected every year in the successive "rounds" of a survey, each "round" extending over several months. The data are usually tabulated separately for the different States (which constitute the Union of India) or for groups of States, for India as a whole and with breakdowns, or in some cases, for rural and urban areas. Consider, for example, surveys of household budgets. The total or per capita consumption expenditure for 30 days (or any given reference period) and also the per capita expenditure on, say, foodgrains, all items of food, drugs, or cloth, etc., would be reported for each sample household. As a probability sample is used in each case, it is possible to estimate any of the characteristics for the whole population. In this way, information is available on the distribution of households by size of total or per capita consumption expenditure, or of the expenditure on individual items, for each size class, for the different States and for India, and over time from round to round of the National Sample Survey. The question naturally arises whether the pattern of consumption is or is not the same from one State to another; is or is not steady over time from one round of the survey to another, for the same State or for the whole of India.

1.2. The design of the survey is always such that Inter-Penetrating Samples (IPS) are used in which two or more independent samples are

drawn in exactly the same way—according to the design of the survey and with replacement. These two or more interpenetrating samples are statistically equivalent and supply independent and equally valid estimates of all population characteristics. For convenience these two or more samples may be called sub-samples, emphasizing the fact that they can be pooled together to give one combined sample to supply valid estimates of population characteristics based on all of the information.

1.3. Consider a sample from a given bivariate population, and let x represent the per capita consumption expenditure and y the per capita expenditure on foodgrains for a sample household. Assume that the size of each sub-sample is the same and consists of N households each.[1] Also assume that each sample household has the same chance of being included in each sub-sample. Then all sample households would have the same probability weight for purposes of estimating the population characteristics. The processing of the data in this case is extremely simple. Consider the first sub-sample of N households; rank the households in ascending order of the x-variate (in this case, the per capita expenditure of the household in the given reference period).[2] Now divide the N sample units into g equal groups of n each, so that $N = gn$. Next, find the average value of the y variate, for example, the per capita expenditure on foodgrains for each of the g groups. These may be called $y_{11}, y_{12}, y_{13}, \ldots, y_{1g}$. Now take g equidistant points[3] on the x-axis and plot the y values, that is, $y_{11}, y_{12}, y_{13}, \ldots, y_{1g}$ on these g points $x = 1, 2, \ldots, g$. Finally, join the successive points by straight lines.[4] This would give a graph which may be called $G(1)$ for the first sub-sample.

1.4. The procedure would be slightly more complicated when the sample units are selected with varying probabilities of being included in the sample, and, therefore, have varying probability weights for purposes of estimating the population characteristics. Consider the case in which x_i is the per capita total expenditure and y_i is the per capita expenditure on foodgrains for the ith sample household. Let w_i be the probability weight of this house-

[1] This can be arranged by specification in the design of the survey, but this is not essential. One great advantage of the fractile graphical method is that the procedure would remain the same even when the sub-samples are of different size.

[2] This can be done by hand or very quickly with punched cards by a sorter.

[3] These points would be, of course, centred at the midpoint of each of the g groups. It is, however, not necessary that the x-points should be equidistant; other scales can be used, for example, values of x such that the expected frequency in each group would be equal for a normal distribution. Any other suitable distribution may also be used.

[4] It is possible to use other rules of construction (for example, graduating parabolas passing through three or more successive points) to draw the graph.

hold for estimating the population characteristics; also, let the w_i's be adjusted in such a way that the total probability weight, that is, the sum of all values of w_i, equals one. The first stage of processing, namely, ranking the sample households in ascending values of x_i, remains the same; the next stage is to multiply y_i (and other similar variates if there are more than one) by w_i; and then to form the accumulated sums of both w_i and $w_i y_i$ for all sample households.[5] It is now possible to form any desired number of fractile groups, say, g, each having an equal proportion of the estimated total population of households. For example, if $g = 20$, all that has to be done is to divide the sub-sample at successive five per cent points into groups on the basis of the accumulated sums of w_i. It is also easy to obtain the estimated average value of any population characteristic y for each of the fractile groups from the corresponding accumulated sums of $w_i y_i$. These average values of y are then plotted against g equidistant points on the x-axis, and successive plotted points are joined by straight lines to supply the graph $G(1)$.

1.5. Consider next the second sub-sample, also of N sample households. Rank them by ascending values of x; divide into g equal fractile groups; find the average values $y_{21}, y_{23}, \ldots, y_{2g}$ of the y variate for each of the g groups; plot these points on the same chart and on the same g equidistant points as used for the first graph, $G(1)$; and join the successive points by straight lines. This would give a second graph, $G(2)$, for the second sub-sample.

1.6. The next step is to pool the two sub-samples together to form a single sample of size $2N$. Rank the sample units again by ascending values of x; group them into g equal groups each of size $2n$; find the average values of y, that is, $y_1, y_2, \ldots, y_g$ for each of the g groups; plot them on the same chart and on the same g equidistant points; and join the successive points by straight lines. This would give the combined graph which may be called $G(1,2)$ for the combined sample. This completes the construction for the given population.[6]

[5] If punched cards are used, this can easily be done by running them through a tabulator and printing the accumulated subtotals for the desired fractile groups, or for fixed ranges of the x variate. A convenient plan is to use twenty 5 per cent groups together with five 1 per cent groups at both the bottom and the top, giving thirty groups altogether.

[6] A complete or exhaustive processing would be given if the accumulated totals of wy were obtained for each successive sample household. This can easily be done by running the punched cards through a tabulator and printing all successive accumulated totals. The three sets of accumulated sums contain the whole of the information. The three graphs $G(1)$, $G(2)$, and $G(1,2)$ would also contain the information in a form such that the accuracy with which the information could be recovered would depend

1.7. I shall now make some assumptions which seem plausible. Consider the area bounded by the two sub-sample graphs $G(1)$ and $G(2)$. (This can be lightly shaded in the chart to make it distinct). We may use this area as a convenient measure of "error"[7] to be associated with the combined graph $G(1,2)$.

1.8. Now consider a second population and assume that a pair of interpenetrating samples have been taken from it with the same sample size of N households in each sub-sample. It is then possible to go through the same construction as in the case of the pair of sub-samples from the first population. Let the corresponding graphs for the second population be called $G'(1)$, $G'(2)$ and $G'(1,2)$. The area between the sub-sample graphs $G'(1)$ and $G'(2)$ would then give the corresponding "error"[8] associated with the combined graph $G'(1,2)$ for the second population.

1.9. It is now possible to go one step further. The area between $G(1,2)$ and $G'(1,2)$ may be used as a measure of the "separation," or over-all difference or generalised distance, between the two populations. It is plausible to assume that the "error" to be associated with the "separation"

---

on the accuracy with which observations could be read from the graphs. Also, it may be possible to graduate the three sets of accumulated sums by suitable mathematical functions; and the information can be recovered from such functions with a margin of uncertainty which would depend on the "goodness of fit" of the graduating functions and the margins of error given by the sub-samples.

Let $F(1)$, $F(2)$ and $F(1,2)$ be the three graduating functions of the same specification which are fitted to the data for the two sub-samples and for the combined sample, respectively. These three curves can be easily plotted on the chart. Then the area bounded by the two graduating functions for the two sub-samples $F(1)$ and $F(2)$ would supply a measure of the error to be associated with the graduating function for the combined sample. Alternatively, the deviation from the graduating function can be measured directly by the area bounded by the observed combined graph $G(1,2)$ and the corresponding graduating curve $F(1,2)$. These two areas should be statistically of the same order of magnitude; and if these areas (that is, the deviation from the graduating function) are less then the margin of error given by the area bounded by the two sub-samples, it would then be possible to recover from the graduating function almost the whole of whatever information can validly be used for statistical purposes relating to the estimated population characteristics (although some of the sample-specific information may be lost).

[7] This and other conjectures are given in an Appendix.

[8] Although sub-samples have been assumed to be of equal size, this is not necessary. It is also unnecessary that the size of the sample should be the same in the case of the two samples from the two populations. Let N_1, N_2 be the sizes of the two sub-samples from the first population and N_1' and N_2' be the sizes of the two sub-samples from the second population.Then whether $N_1 = N_2 = N_1' = N_2'$ (as assumed above), or whether they are all different is entirely immaterial. The graphical measure of the two "errors" would be given in every case by the two areas lying respectively between each pair of sub-sample graphs $G(1)$ and $G(2)$ or $G'(1)$ and $G'(2)$.

can be derived in the usual way[9] from the two component errors associated respectively with each of the combined graphs $G(1,2)$ and $G'(1,2)$. It is also possible to consider not only the total separation but the separation between any particular portions of the two combined graphs $G(1,2)$ and $G'(1,2)$, that is, the area bounded by these two graphs and the corresponding ordinates limiting any assigned fractile group. Each such partial separation would also have its associated error which can be derived from the two component errors lying between the same two ordinates. This furnishes a convenient tool for the comparison, analysis, and the testing of significance of the separation between two populations either over the whole or any part of the range of observations.

2. ILLUSTRATIONS FROM THE NATIONAL SAMPLE SURVEY OF INDIA

2.1. Some actual examples from the National Sample Survey of India may now be considered. In the 7th round (October, 1953 to March, 1954) information on household consumption was collected in the form of two interpenetrating samples of 702 and 711 sample households from 476 and 478 *mauzas* or villages, respectively, extending over the whole of rural India (excluding Jammu and Kashmir). In another survey carried out in the 9th round (May to November, 1955) similar information was collected from 768 households, one each from 768 *mauzas* (out of 772 *mauzas* or villages of which four were uninhabited)[10] in each of the two sub-samples. The period of reference was 30 days in each of the surveys. It is possible to use the per capita expenditure on all items of consumption as the x variate. The sample households were ranked (separately for each of the two sub-samples and for the combined sample) in ascending order of the x variate (per capita consumption expenditure). The design of the sample surveys was with varying probabilities, and appropriate probability weights were used for estimating the population characteristics. The estimated number of households and other characteristics of the population were obtained in the usual way, and twenty equal fractile groups, each containing 5 per cent of the households, were formed. The bottom and top 5 per cent groups were subdivided into 1 per cent groups (1, 2, 3, 4, 5 and 96, 97, 98, 99, 100 percentiles), giving 30 groups altogether. Some additional fractile groups were also formed.[11]

[9] That is, by the square root of the sum of the squares of the errors of the two combined graphs $G(1,2)$ and $G'(1,2)$.

[10] The word "village" is used for the Indian revenue unit *mauza* which broadly corresponds to a village. Although these revenue units are several centuries old and some have become uninhabited, land continues to be demarcated in terms of the old *mauza* units.

[11] Averages for larger fractile groups can be obtained directly by taking the average of an appropriate number of equi-frequency fractile groups.

TABLE I

NATIONAL SAMPLE SURVEY OF INDIA: ALL INDIA, RURAL

7TH ROUND: OCTOBER, 1954—MARCH, 1955*

TOTAL PER CAPITA CONSUMER EXPENDITURE IN RUPEES PER 30 DAYS

Serial number	Fractile Group (percent)	Total per Capita Consumer Expenditure in Rupees per 30 Days					
		Limiting value at upper end of each fractile group			Average expenditure in each fractile group		
		s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	2.60	3.25	3.00	2.44	2.20	2.23
2	2	3.00	4.13	3.38	2.91	3.74	3.14
3	3	4.00	4.75	4.57	3.35	4.54	4.10
4	4	5.00	4.83	4.83	4.63	4.78	4.73
5	5	5.57	5.25	5.50	5.32	4.99	5.13
6	0— 5	5.57	5.25	5.50	3.75	3.99	3.85
7	5—10	7.25	6.88	7.11	6.63	6.24	6.39
8	10—15	8.25	7.90	8.00	7.72	7.46	7.54
9	15—20	9.29	9.00	9.25	8.77	8.53	8.64
10	20—25	9.80	9.67	9.75	9.58	9.43	9.51
11	25—30	10.60	10.25	10.50	10.22	9.93	10.03
12	30—35	11.83	11.00	11.40	11.24	10.69	10.95
13	35—40	13.00	12.00	12.43	12.44	11.45	11.94
14	40—45	14.00	13.00	13.50	13.37	12.48	12.83
15	45—50	14.75	14.25	14.40	14.23	13.64	14.00
16	50—55	16.00	15.25	15.67	15.40	14.73	15.02
17	50—55	17.60	16.60	17.00	16.82	15.99	16.37
18	60—65	18.80	18.00	18.45	18.34	17.22	17.78
19	65—70	21.00	19.25	20.00	19.90	18.67	19.18
20	70—75	22.00	22.00	22.00	21.39	20.49	21.11
21	75—80	25.00	24.00	24.33	23.34	23.02	23.12
22	80—85	27.67	28.00	28.00	26.63	26.09	26.40
23	85—90	32.75	32.33	32.67	30.07	29.74	29.95
24	90—95	43.67	39.00	41.33	37.17	35.81	36.14
25	95—100	226.00	264.50	264.50	74.11	58.97	64.98
26	96	47.00	43.00	45.33	45.76	41.97	43.86
27	97	59.00	48.00	51.00	51.91	45.51	47.34
28	98	72.17	53.50	61.00	66.35	51.37	54.08
29	99	76.00	65.25	74.33	75.46	59.64	67.05
30	100	226.00	264.50	264.50	118.11	142.48	122.59
31	0—20				6.74	6.49	6.64
32	20—40				10.88	10.49	10.66
33	40—60				14.92	13.97	14.44
34	60—80				20.52	19.91	20.20
35	80—100				39.77	36.45	38.03
36	0—25				7.30	7.11	7.18
37	25—50				12.41	11.73	12.03
38	50—75				18.20	17.45	17.83
39	75—100				36.45	33.44	34.86
40	0—50				9.88	9.58	9.73
41	50—100				26.57	25.20	25.87
42	0—100				17.65	16.28	17.24

	sub-sample 1	sub-sample 2	combined
* number of sample villages	476	478	954
number of sample households	702	711	1413

TABLE II

NATIONAL SAMPLE SURVEY OF INDIA: ALL INDIA, RURAL
9TH ROUND: MAY—NOVEMBER, 1955*
TOTAL PER CAPITA CONSUMER EXPENDITURE IN RUPEES PER 30 DAYS

Serial number	Fractile Group (percent)	Total per Capita Consumer Expenditure in Rupees per 30 Days					
		Limiting value at upper end of each fractile group			Average expenditure in each fractile group		
		s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	3.19	2.98	3.16	2.99	2.58	2.71
2	2	3.91	3.71	3.87	3.67	3.43	3.52
3	3	4.28	4.16	4.20	4.13	4.01	4.06
4	4	4.65	4.56	4.65	4.51	4.39	4.46
5	5	4.91	4.79	4.90	4.85	4.72	4.78
6	0— 5	4.91	4.79	4.90	4.01	3.82	3.90
7	5—10	5.99	5.85	5.92	5.43	5.39	5.39
8	10—15	6.74	6.88	6.80	6.39	6.38	6.38
9	15—20	7.54	7.90	7.76	7.25	7.35	7.28
10	20—25	8.52	8.59	8.56	8.05	8.31	8.20
11	25—30	9.36	9.57	9.53	8.96	9.19	9.06
12	30—35	10.17	10.39	10.31	9.80	9.95	9.89
13	35—40	11.03	11.09	11.07	10.66	10.74	10.69
14	40—45	11.80	12.06	11.92	11.42	11.68	11.49
15	45—50	12.83	13.02	12.88	12.35	12.59	12.43
16	50—55	14.14	14.05	14.14	13.46	13.45	13.44
17	55—60	15.13	15.55	15.24	14.57	14.73	14.64
18	60—65	16.61	16.97	16.77	15.81	16.12	16.00
19	65—70	18.49	18.59	18.56	17.61	17.88	17.78
20	70—75	19.85	20.08	20.02	19.14	19.37	19.26
21	75—80	22.59	22.01	22.11	21.22	21.27	21.19
22	80—85	24.95	23.60	24.35	23.77	22.97	23.36
23	85—90	29.57	27.53	28.59	27.40	25.42	26.23
24	90—95	38.05	38.38	38.13	32.85	32.06	32.17
25	95—100	194.41	128.86	194.41	55.96	53.30	54.31
26	96	43.45	40.74	40.77	40.34	39.83	39.35
27	97	47.41	44.16	46.30	45.60	42.94	44.01
28	98	58.18	57.16	57.16	56.06	49.80	52.11
29	99	66.86	67.69	66.86	65.34	72.22	62.86
30	100	194.41	128.86	194.41	96.54	66.65	78.88
31	0—20				5.64	5.72	5.67
32	20—40				9.39	9.56	9.46
33	40—60				13.01	13.15	13.06
34	60—80				18.33	18.82	18.58
35	80—100				32.90	33.15	32.99
36	0—25				6.06	6.26	6.16
37	25—50				10.61	10.77	10.68
38	50—75				16.04	16.27	16.12
39	75—100				30.53	30.45	30.43
40	0—50				8.31	8.48	8.39
41	50—100				22.65	23.10	22.85
42	0—100				15.03	15.24	15.15

* The number of sample villages was 772 (including 4 uninhabited mauzas) and the number of sample households was 768 in each sub-sample.

2.2. Tables I and II give the information for the 7th and 9th rounds respectively.[12] In each table, columns (2), (3) and (4) give the limiting value of the per capita expenditure at the upper end of the fractile group for (respectively) sub-sample 1 (which would be written in the contracted form s.s.1), sub-sample 2 (s.s.2) and the combined sample. For example, in Table I for the bottom 1 per cent group, the limiting values of per capita expenditure are Rs. 2.60 and Rs. 3.25 in the two sub-samples and Rs. 3.00 in the combined sample. Columns (5), (6), and (7) in both the tables give the average per capita consumer expenditure in each fractile group. It is known that prices of consumer goods, especially of foodgrains, had greatly decreased at the time of the 9th round. The effect of this fall in prices can be seen from these two tables. Consider the fractile group between 70 per cent and 75 per cent of households; the lower and upper limiting values of per capita expenditure in the 7th round are, respectively, Rs. 20.00 and Rs. 22.00. It should be noticed, however, that the fractile group between 75 per cent and 80 per cent in the 9th round has approximately the same limiting values—Rs. 20.02 and Rs. 22.11. If, in terms of money values, a fixed range had been used, the group having a per capita expenditure between Rs. 20 and Rs. 22 per 30 days, would have been between 70 and 75 per cent of the households in the 7th round, but would have been almost between 75 and 80 per cent in the 9th round. In the present approach the comparison between the two rounds is made on the basis of the *same* fractile group, that is between 70 and 75 per cent or between 75 and 80 per cent of households in each of the rounds.

2.3. Consider also an associated variate y, the per capita expenditure on foodgrains for each sample household. The average value of y, that is, the average per capita expenditure on foodgrains for each fractile group, is given in columns (2), (3), and (4) of Table III for the first sub-sample (s.s.1), the second sub-sample (s.s.2) and the combined sample for the 7th round respectively. Corresponding values for the 9th round are given in columns (2), (3), and (4) of Table IV. The results are shown in Figure 1, where the x-axis represents percentages of households, as ranked by per capita consumption expenditure. The limiting values of the per capita expenditure at the upper end of each fractile group (in Table I, for the 7th round) are shown at the top of the x-scale; and the corresponding limiting values (in

[12] Based on the sample survey, the estimated number of rural households was 63.4 million in the 7th round and 65.3 million in the 9th round; the estimated rural population was 324 million and 338 million in the 7th and 9th rounds, respectively; and the estimated per capita consumption expenditure in rural households was Rs. 5565 million and Rs. 5131 million per 30 days at current prices in the 7th and the 9th rounds, respectively. The total number of *mauzas* (villages) in India, excluding Jammu and Kashmir, was 603,168 in the 1951 census.

Table II for the 9th round) are shown at the bottom of the x-scale. The y-axis represents the average value of the per capita expenditure on foodgrains in rupees per 30 days for each 5 per cent fractile group.

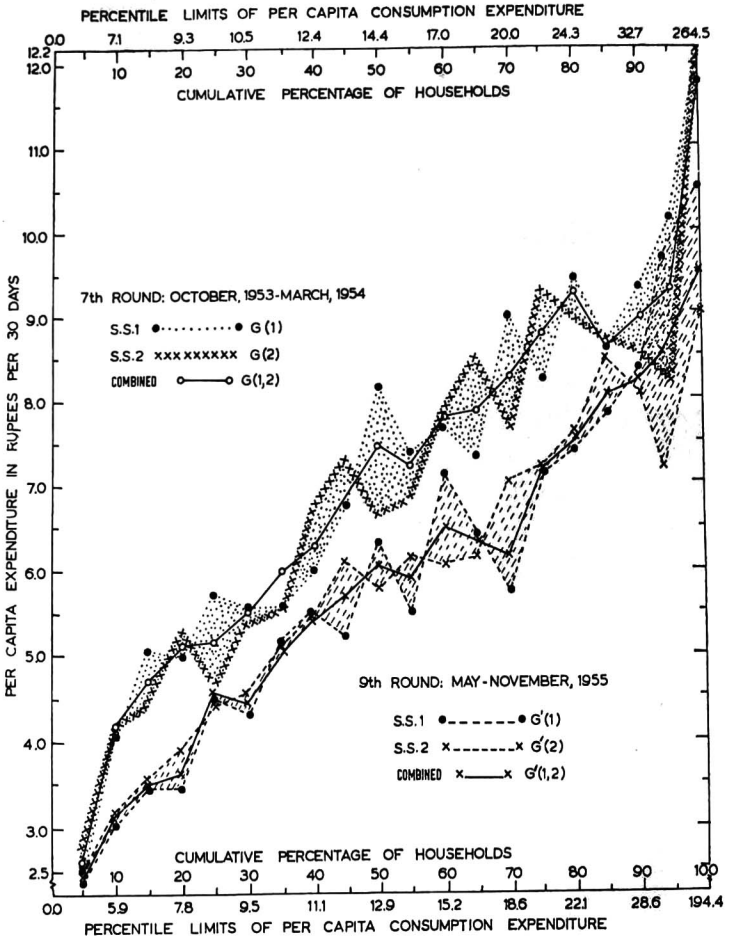

FIGURE 1

National sample survey: all India, rural.
Per capita expenditure on food grains per 30 days.

2.4. Now consider the three upper graphs in Figure 1 which represent $G(1)$, $G(2)$ and $G(1,2)$ for the 7th round. The shaded area between $G(1)$ and $G(2)$ supplies a measure of the error associated with $G(1,2)$, which has been called e. It will be seen from all three graphs, that the per capita expenditure on foodgrains increases over the whole range from the bottom 5 per cent to the top 5 per cent of households, and that the general trend of the increase in expenditure is significant in comparison with the associated error. The three lower graphs in Figure 1 in the same way represent the three graphs

$G'(1)$, $G'(2)$, $G'(1,2)$ for the 9th round; and the shaded area between $G'(1)$ and $G'(2)$ gives the error associated with $G'(1,2)$ or e'. In this case also the general trend of the increase in expenditure on foodgrains is significant in comparison with the associated error.

2.5. Now consider the "separation" which is given by the area lying between the two combined graphs $G(1,2)$ and $G'(1,2)$ for the two rounds of the survey.[13] The per capita expenditure on foodgrains was lower for every 5 per cent fractile group in the 9th round. From the bottom up to 80 per cent of the households, the separation is seen to be roughly greater than the sum of the two associated error areas,[14] e and e'. The 9th round decrease in the expenditure on foodgrains for up to 80 per cent for households may, therefore, be considered statistically significant. The per capita expenditure on foodgrains was also less in the 9th round for the top 20 per cent of households, but the separation was here smaller than the two associated error areas. It is not possible, therefore, to assert that the decrease was significant for *each* of the four 5 per cent fractile groups at the top. This point can be further examined by pooling together the four 5 per cent fractile groups at the top to form a single 20 per cent group. For this group the average values for the 7th round from Table III are Rs. 9.87 and Rs. 9.25 for the two sub-samples and Rs. 9.54 for the combined sample. For the 9th round from Table IV the average values are Rs. 8.92 and Rs. 8.12 for the two sub-samples and Rs. 8.52 for the combined sample. The difference between the two rounds for the estimates based on the combined samples is Rs. 1.02, which is only somewhat greater than the associated error. Even for the top 20 per cent group as a whole the decrease in the average per capita expenditure on foodgrains was not quite significant.

2.6. Consider next the per capita expenditure on foodgrains as a fraction of the total per capita consumer expenditure of which the average values for each fractile group are given in columns (5), (6), and (7) of Table III and Table IV for the 7th round and the 9th round, respectively. The corresponding graphs are shown in Figure 2. It will be noticed that the

[13] The two error areas, e and e', are seen to be roughly equal, and E, the error of the separation, is therefore very roughly 1.4 e.

[14] As the sum of the component areas $(e + e')$ is greater than $\sqrt{(e^2 + e'^2)}$ there is a margin of safety in using the sum of the errors to test the significance of the separation. The advantage is that a visual comparison is possible. It is also possible to measure the separation, S, as well as the two associated errors e and e'; calculate E, the error of the separation; obtain S/E and plot these values for each fractile part. This would involve a certain amount of additional calculations, but would supply necessary material in a convenient graphical form to test the significance of the separation.

TABLE III

NATIONAL SAMPLE SURVEY OF INDIA: ALL INDIA, RURAL
7TH ROUND: OCTOBER, 1953—MARCH, 1954*

AVERAGE PER CAPITA EXPENDITURE ON FOODGRAINS IN RUPEES AND AS FRACTION OF TOTAL EXPENDITURE PER 30 DAYS

Serial number	Fractile Group (percent)	Average per Capita Expenditure on Foodgrains			Average Fraction of Total Expenditure		
		s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	1.09	2.83	2.16	0.436	0.588	0.572
2	2	2.32	2.75	2.13	0.751	0.704	0.679
3	3	2.26	2.57	2.74	0.656	0.550	0.656
4	4	3.12	2.66	2.71	0.639	0.624	0.576
5	5	3.70	3.22	3.39	0.632	0.622	0.610
6	0— 5	2.49	2.79	2.61	0.635	0.614	0.620
7	5—10	4.07	4.20	4.20	0.600	0.664	0.648
8	10—15	5.07	4.44	4.70	0.616	0.554	0.573
9	15—20	4.99	5.28	5.11	0.562	0.584	0.576
10	20—25	5.71	4.66	5.17	0.572	0.535	0.550
11	25—30	5.56	5.39	5.53	0.560	0.560	0.562
12	30—35	5.59	5.57	5.98	0.515	0.557	0.541
13	35—40	6.00	6.73	6.28	0.499	0.552	0.534
14	40—45	6.75	7.29	6.82	0.487	0.546	0.492
15	45—50	8.15	6.60	7.46	0.510	0.490	0.503
16	50—55	7.38	6.83	7.21	0.481	0.475	0.491
17	55—60	7.67	7.84	7.78	0.492	0.516	0.489
18	60—65	7.31	8.49	7.85	0.398	0.464	0.444
19	65—70	8.95	7.65	8.23	0.446	0.417	0.425
20	70—75	8.22	9.27	8.74	0.423	0.456	0.440
21	75—80	9.44	8.93	9.24	0.388	0.396	0.398
22	80—85	8.69	8.69	8.59	0.358	0.382	0.366
23	85—90	9.30	8.53	8.96	0.324	0.333	0.331
24	90—95	10.13	8.18	9.29	0.292	0.238	0.272
25	95—100	12.02	12.19	11.78	0.214	0.234	0.219
26	96	13.18	13.09	10.38	0.303	0.322	0.261
27	97	13.72	12.76	15.01	0.253	0.277	0.275
28	98	13.29	12.04	11.51	0.233	0.222	0.222
29	99	10.68	10.06	11.74	0.139	0.189	0.185
30	100	9.27	13.21	10.38	0.081	0.161	0.102
31	0—20	4.19	4.13	4.17	0.601	0.603	0.602
32	20—40	5.72	5.70	5.77	0.534	0.552	0.547
33	40—60	7.54	7.16	7.30	0.492	0.506	0.494
34	60—80	8.40	8.59	8.49	0.412	0.432	0.426
35	80—100	9.87	9.25	9.54	0.296	0.295	0.296
36	0—25	4.49	4.24	4.36	0.615	0.597	0.607
37	25—50	6.51	6.47	6.46	0.524	0.551	0.536
38	50—75	7.85	8.02	7.94	0.432	0.460	0.446
39	75—100	9.79	9.18	9.48	0.268	0.274	0.272
40	0—50	5.51	5.43	5.46	0.558	0.567	0.561
41	50—100	8.74	8.58	8.67	0.329	0.341	0.335
42	0—100	7.01	6.90	6.96	0.398	0.410	0.404

	sub-sample 1	sub-sample 2	combined
* number of sample villages	476	478	954
number of sample households	702	711	1413

TABLE IV

NATIONAL SAMPLE SURVEY OF INDIA: ALL INDIA, RURAL

9TH ROUND: MAY—NOVEMBER, 1955*

AVERAGE PER CAPITA EXPENDITURE ON FOODGRAINS IN RUPEES AND AS FRACTION OF TOTAL EXPENDITURE PER 30 DAYS

Serial number	Fractile Group (percent)	Average per Capita Expenditure on Foodgrains			Average Fraction of Total Expenditure		
		s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	1.72	1.65	1.66	0.581	0.654	0.638
2	2	1.66	1.86	1.87	0.463	0.567	0.532
3	3	2.46	2.36	2.34	0.597	0.604	0.594
4	4	2.48	3.19	2.89	0.556	0.675	0.602
5	5	3.52	3.17	3.25	0.686	0.672	0.657
6	0— 5	2.34	2.49	2.39	0.571	0.634	0.603
7	5—10	3.01	3.14	3.09	0.557	0.581	0.568
8	10—15	3.43	3.54	3.48	0.550	0.550	0.554
9	15—20	3.42	3.88	3.60	0.511	0.541	0.524
10	20—25	4.46	4.37	4.50	0.554	0.516	0.539
11	25—30	4.28	4.50	4.39	0.509	0.502	0.500
12	30—35	5.11	5.06	5.05	0.525	0.485	0.509
13	35—40	5.44	5.43	5.34	0.522	0.528	0.517
14	40—45	5.18	6.02	5.61	0.469	0.500	0.493
15	45—50	6.25	5.70	6.00	0.495	0.484	0.488
16	50—55	5.45	6.09	5.80	0.418	0.456	0.434
17	55—60	7.05	5.99	6.56	0.477	0.417	0.456
18	60—65	6.29	6.09	6.30	0.428	0.423	0.427
19	65—70	5.68	7.03	6.18	0.343	0.391	0.355
20	70—75	7.15	7.23	7.19	0.389	0.374	0.383
21	75—80	7.41	7.61	7.50	0.354	0.360	0.356
22	80—85	7.82	8.49	8.05	0.332	0.378	0.354
23	85—90	8.36	8.09	8.21	0.299	0.330	0.316
24	90—95	9.66	7.21	8.50	0.289	0.290	0.288
25	95—100	10.51	8.97	9.50	0.189	0.193	0.190
26	96	7.96	9.81	9.01	0.225	0.254	0.235
27	97	11.06	6.73	8.68	0.246	0.162	0.222
28	98	9.64	8.72	7.31	0.173	0.217	0.165
29	99	12.01	10.75	11.65	0.165	0.201	0.183
30	100	13.98	9.86	11.52	0.142	0.114	0.140
31	0—20	3.00	3.26	3.11	0.548	0.576	0.562
32	20—40	4.81	4.85	4.81	0.528	0.506	0.516
33	40—60	6.03	5.95	6.01	0.465	0.463	0.467
34	60—80	6.56	7.04	6.79	0.379	0.385	0.380
35	80—100	8.92	8.12	8.52	0.277	0.294	0.286
36	0—25	3.25	3.49	3.38	0.537	0.557	0.549
37	25—50	5.24	5.32	5.26	0.494	0.494	0.493
38	50—75	6.30	6.49	6.39	0.393	0.399	0.396
39	75—100	8.61	8.00	8.30	0.282	0.263	0.273
40	0—50	4.24	4.39	4.31	0.510	0.517	0.514
41	50—100	7.35	7.22	7.29	0.325	0.312	0.319
42	0—100	5.70	5.71	5.70	0.379	0.373	0.376

* The number of sample villages was 772 (including 4 uninhabited mauzas) and the number of sample households was 768 in each sub-sample.

value of the fraction is generally lower in the 9th round but the separation is now much less in comparison with the associated error areas. That is,

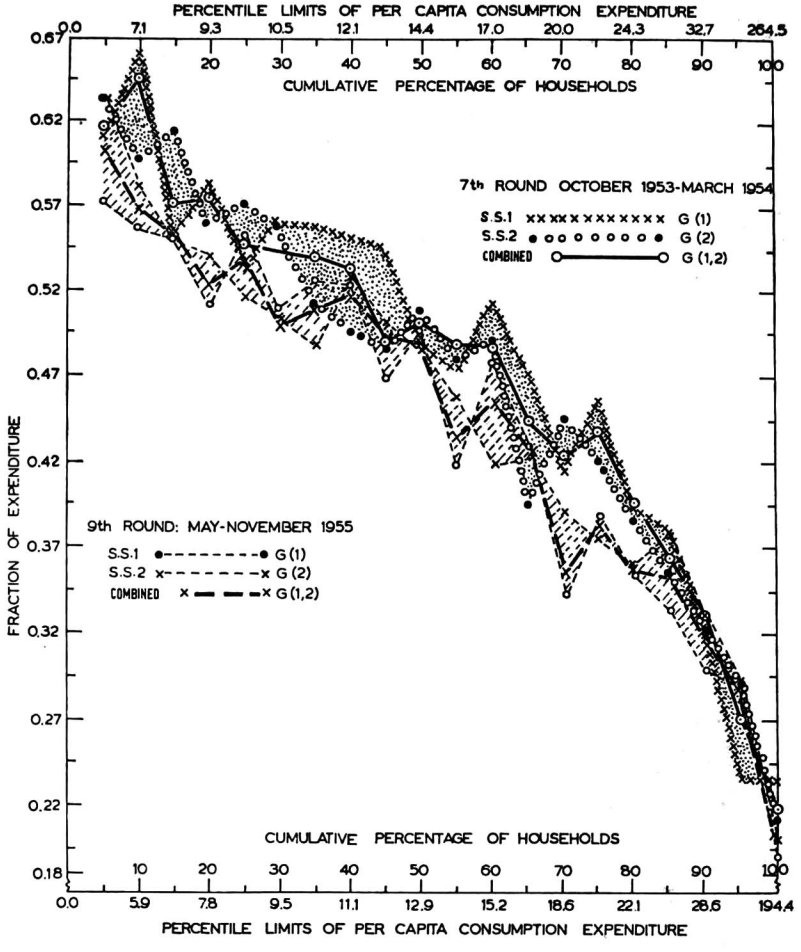

FIGURE 2

National sample survey: all India, rural.
Per capita expenditure on food grains as fraction of total expenditure per 30 days.

the fraction of total consumer expenditure spent on foodgrains drops to some extent with a fall in prices but, on the whole, tends to remain much more stable than the actual expenditure on foodgrains. For the rural households as a whole, foodgrains account for from 38 to 40 per cent of all consumer expenditure and the fraction is over 60 per cent in the bottom 5 per cent of households.

2.7. The same data can be shown in the form of cumulative percentages which are given for the expenditure on all consumer items in columns (2),

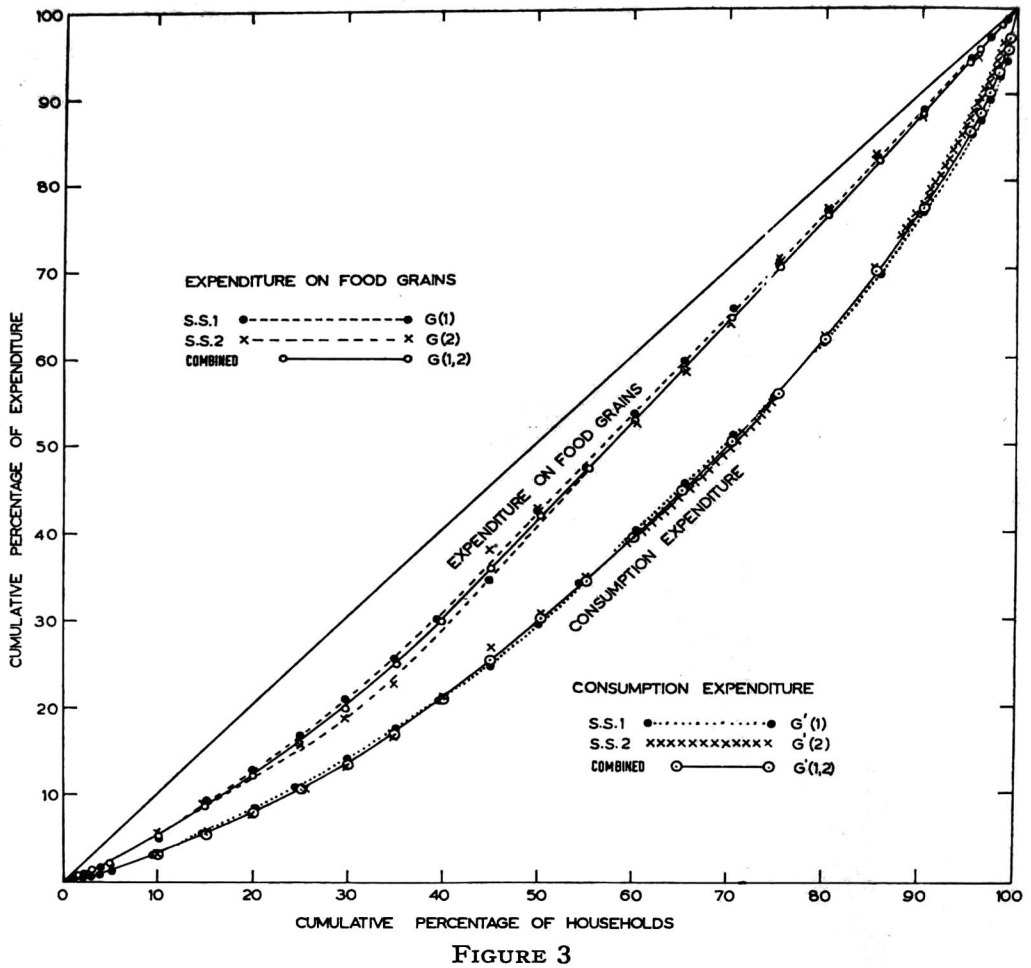

FIGURE 3

National sample survey: all India, rural.
7th round: October, 1953–March, 1954.
Cumulative percentage of total consumer expenditure and expenditure on food grains per 30 days.

(3), and (4) and for the expenditure on foodgrains in columns (5), (6), and (7) in Tables V and VI for the 7th round and 9th round, respectively. Figure 3 shows the graphs for the 7th round, and Figure 4 for the 9th round. The three upper graphs in Figure 3 and Figure 4 represent the concentration of expenditure on foodgrains, and the area lying between the two sub-sample graphs $G(1)$ and $G(2)$ gives the error associated with the graph $G(1,2)$ based on the combined sample. The three lower graphs represent the concentration of total consumer expenditure. In both cases, the separation between $G(1,2)$ and $G'(1,2)$ or between the two sets of three graphs is much greater than either of the two associated error areas. The concentration graph for expenditure on foodgrains thus lies significantly

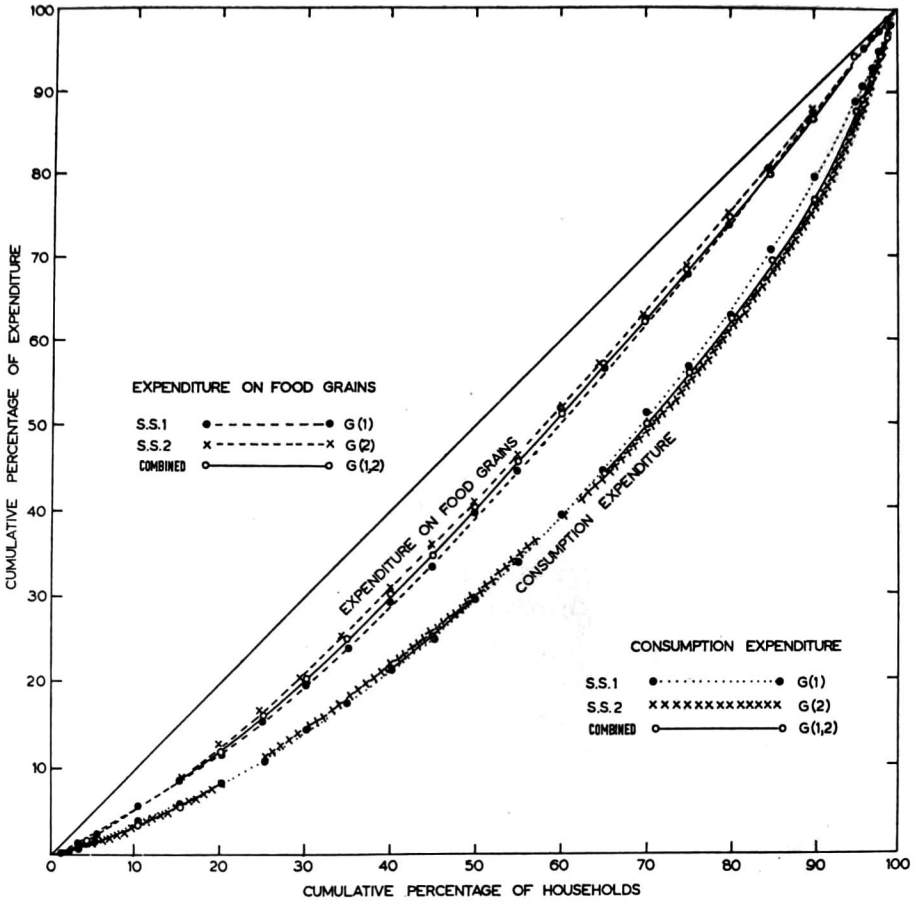

FIGURE 4

National sample survey: all India, rural.
9th round: May–November, 1955.
Cumulative percentage of total consumer expenditure and expenditure on food grains per 30 days.

nearer the line of equal distribution, $y = x$, showing the inelastic nature of the expenditure on foodgrains. The area lying between the line $y = x$ and the graph $G(1,2)$ can supply a convenient measure of concentration; this also will have the same associated error as $G(1,2)$.

2.8. The concentration curves (based on the combined sample in each case) for both the 7th and 9th rounds are shown on Figure 5. The two lower graphs, which represent the two concentration curves for total consumer expenditure, cross and recross, indicating that there was no change in this respect between the 7th and the 9th round. The two upper graphs represent the two concentration curves for the expenditure on foodgrains. These two

TABLE V

National Sample Survey of India: All India, Rural

7th Round: October, 1952—March, 1954*

CUMULATIVE PERCENTAGE OF TOTAL CONSUMER EXPENDITURE AND EXPENDITURE ON FOODGRAINS PER 30 DAYS

Serial Number	Fractile Group (percent)	Cumulative Percentage of Expenditure			Cumulative Percentage of Expenditure on Foodgrains		
		s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	0.14	0.14	0.13	0.15	0.45	0.31
2	2	0.26	0.36	0.31	0.40	0.85	0.61
3	3	0.50	0.67	0.55	0.79	1.28	1.01
4	4	0.81	0.93	0.86	1.31	1.62	1.45
5	0— 5	1.07	1.18	1.11	1.79	2.01	1.86
6	5—10	3.09	3.13	3.00	4.90	5.20	4.93
7	10—15	5.88	5.59	5.54	9.50	8.76	8.85
8	15—20	8.22	7.64	8.01	12.84	11.87	12.47
9	20—25	11.05	10.62	10.71	17.07	15.45	16.10
10	25—30	13.91	13.07	13.35	20.99	18.71	19.70
11	30—35	17.36	16.22	17.44	25.31	22.70	25.23
12	35—40	21.02	21.19	21.05	29.75	29.83	29.93
13	40—45	24.88	26.88	25.68	34.65	37.92	36.02
14	45—50	30.26	30.71	30.47	42.38	42.46	42.34
15	50—55	34.76	34.93	34.69	47.81	47.21	47.35
16	55—60	39.83	39.31	39.53	53.61	52.48	53.05
17	60—65	45.91	44.11	44.64	59.71	58.22	58.63
18	65—70	51.22	49.45	50.50	65.69	63.58	64.86
19	70—75	56.11	55.46	55.75	70.41	70.19	70.23
20	75—80	61.79	62.33	61.98	76.18	76.71	76.39
21	80—85	69.75	70.00	69.99	82.71	82.92	82.84
22	85—90	76.52	77.31	76.94	87.96	88.04	87.99
23	90—95	85.45	87.26	85.84	94.08	93.58	93.65
24	96	87.04	88.19	88.19	95.21	94.29	95.03
25	97	89.37	92.14	90.29	96.74	96.99	96.67
26	98	92.25	94.44	92.64	98.18	98.31	97.91
27	99	93.87	96.83	95.08	98.75	99.28	98.97
28	100	100.00	100.00	100.00	100.00	100.00	100.00

	sub-sample 1	sub-sample 2	combined
* number of sample villages	476	478	954
number of sample households	702	711	1413

TABLE VI

NATIONAL SAMPLE SURVEY OF INDIA: ALL INDIA RURAL

9TH ROUND: MAY, 1955—NOVEMBER, 1955*

CUMULATIVE PERCENTAGE OF TOTAL CONSUMER EXPENDITURE AND EXPENDITURE ON FOODGRAINS PER 30 DAYS

Serial Number	Fractile Group (percent)	Cumulative Percentage of Expenditure			Cumulative Percentage of Expenditure on Foodgrains		
		s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	0.25	0.32	0.19	0.37	0.38	0.32
2	2	0.65	0.38	0.49	0.86	0.62	0.74
3	3	0.88	0.68	0.80	1.22	1.08	1.21
4	4	1.25	1.11	1.17	1.75	1.91	1.85
5	0— 5	1.69	1.39	1.50	2.59	2.43	2.44
6	5—10	3.88	3.25	3.54	5.80	5.33	5.54
7	10—15	5.91	5.37	5.62	8.68	8.48	8.55
8	15—20	8.32	8.03	8.13	11.67	12.24	11.85
9	20—25	10.81	11.06	10.97	15.31	16.51	16.00
10	25—30	14.36	14.33	14.34	19.79	20.80	20.34
11	30—35	17.39	17.95	17.71	23.96	25.73	24.92
12	35—40	21.29	21.95	21.48	29.22	31.15	29.92
13	40—45	24.80	25.58	24.97	33.42	36.16	34.45
14	45—50	29.36	29.62	29.38	39.51	41.07	40.12
15	50—55	33.90	34.23	34.03	44.36	46.66	45.45
16	55—60	39.30	39.14	39.30	51.26	52.01	51.73
17	60—65	44.29	43.48	44.04	56.50	56.40	56.69
18	65—70	51.02	48.75	49.97	62.22	61.96	62.17
19	70—75	56.55	55.30	55.76	67.67	68.51	67.92
20	75—80	62.69	62.40	62.45	73.32	75.32	74.22
21	80—85	70.40	67.71	69.12	80.01	80.58	80.33
22	85—90	79.09	75.85	76.74	87.00	87.52	86.66
23	90—95	88.53	86.29	87.31	94.32	93.82	94.09
24	96	90.55	87.79	88.87	95.37	94.81	95.04
25	97	92.39	90.18	91.44	96.55	95.82	96.39
26	98	94.71	93.98	94.06	97.60	97.60	97.37
27	99	98.44	96.34	97.16	99.41	98.71	98.90
28	100	100.00	100.00	100.00	100.00	100.00	100.00

* The number of sample villages was 772 (including 4 uninhabited mauzas) and the number of sample households was 768 in each sub-sample.

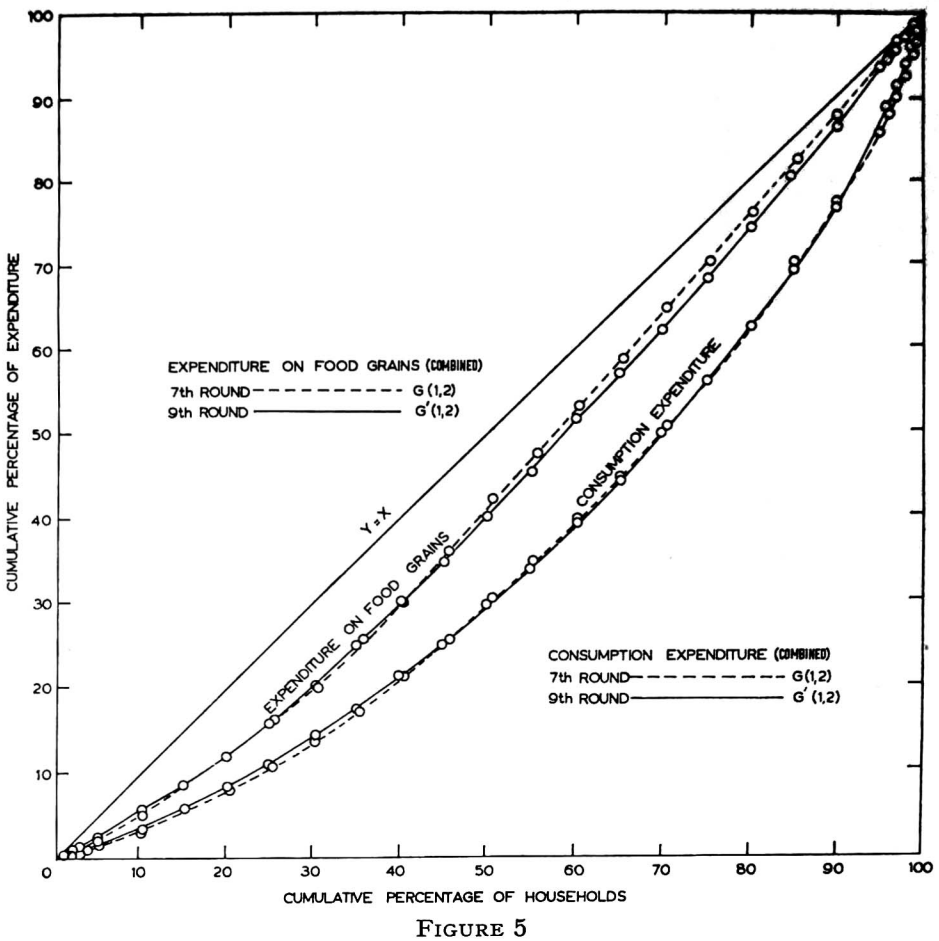

FIGURE 5

National sample survey: all India, rural.
7th round: October, 1953–March, 1954.
9th round: May–November, 1955.
Cumulative percentage of total expenditure and expenditure on food grains per 30 days.

graphs also cross and recross up to the 50th percentile of households and again beyond the 90th percentile of households, but between roughly the 50th percentile and the 90th percentile of households, the graph for the 9th round (when prices were lower) is systematically below the graph for the 7th round. It is possible to examine these portions of the two graphs in greater detail. Figure 6 shows the concentration curves between the 50th and the 90th percentiles of households on a magnified scale for both rounds (with separate graphs for the two sub-samples and the combined sample in each case). The shaded area shows the error associated with the respective graphs. The "separation" of the two graphs $G(1,2)$ and $G'(1,2)$ between the two rounds is on the whole somewhat greater than the two

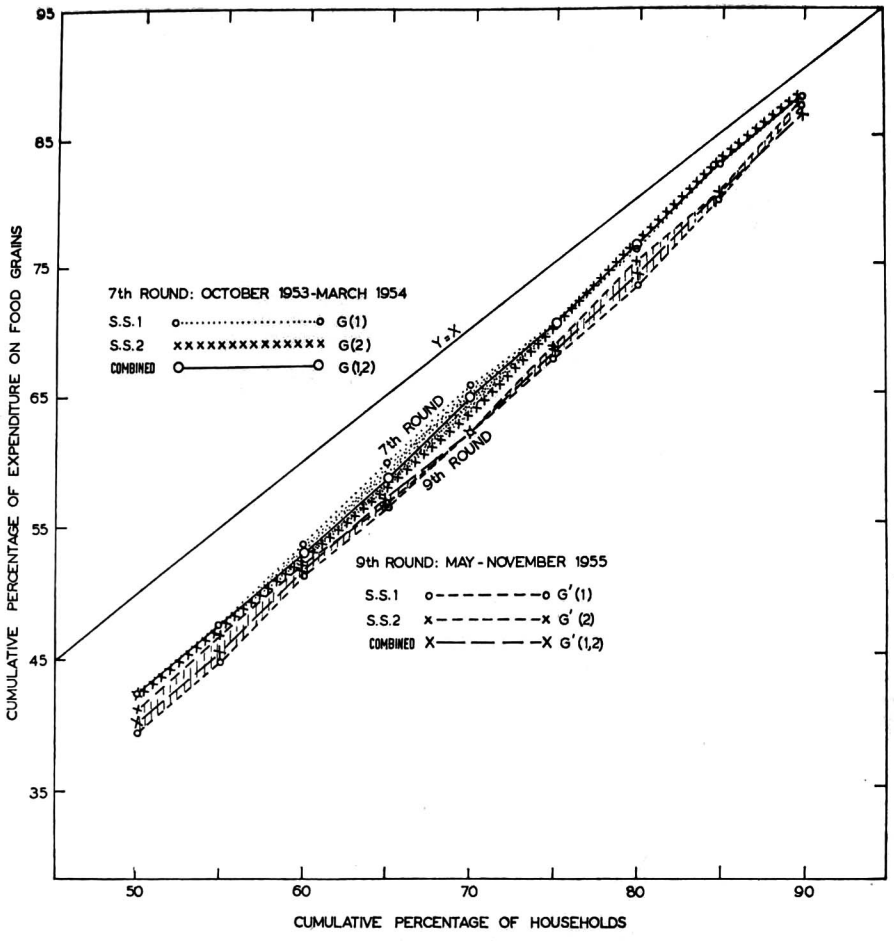

FIGURE 6

National sample survey: all India, rural.
7th round: October, 1953–March, 1954.
9th round; May–November, 1955.
Cumulative percentage of expenditure on food grains.

associated error areas. This, if real, would indicate that in this middle region of households between the 50th and 90th percentiles (with total per capita consumer expenditure lying roughly between Rs. 13 or 14 and Rs. 30 or Rs. 32 per month) the expenditure on foodgrains tends to behave somewhat more as a necessity when prices are higher. The observed difference is based on very small samples and may have arisen through chance; but it may deserve more careful study with larger samples.

3. REGIONAL DIFFERENCES

3.1. Consider a second example. In the 8th round (July, 1954—April, 1955) of the National Sample Survey information was collected for the whole

TABLE VII

NATIONAL SAMPLE SURVEY OF INDIA: RURAL
8TH ROUND: JULY, 1954—APRIL, 1955*
HOUSEHOLD OWNERSHIP HOLDINGS
STATE: WEST BENGAL, CENTRAL SAMPLE

Fractile Group	Upper Limit of Size of Ownership Holding (Acre)			Cumulative Percentage of Owned Area Below			Average Size of Household Ownership Holding (Acres)		
	s.s.1	s.s.2	combined	s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0—15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15—20	0.00	0.04	0.00	0.00	0.04	0.00	0.00	0.02	0.00
20—25	0.00	0.06	0.00	0.00	0.13	0.00	0.00	0.05	0.00
25—30	0.00	0.17	0.04	0.00	0.30	0.06	0.00	0.09	0.03
30—35	0.04	0.37	0.08	0.03	0.74	0.19	0.01	0.25	0.06
35—40	0.08	0.60	0.17	0.17	1.62	0.46	0.06	0.51	0.12
40—45	0.15	0.89	0.36	0.43	2.95	1.03	0.11	0.76	0.27
45—50	0.32	1.19	0.66	0.94	4.73	2.13	0.21	1.02	0.51
50—55	0.52	1.52	1.00	1.91	7.04	3.94	0.39	1.32	0.84
55—60	0.95	2.00	1.36	3.70	10.12	6.47	0.72	1.76	1.18
60—65	1.33	2.40	1.86	6.49	13.92	9.88	1.13	2.18	1.59
65—70	1.84	2.85	2.30	10.32	18.49	14.33	1.55	2.61	2.07
70—75	2.29	3.33	2.92	15.37	23.85	19.94	2.04	3.06	2.60
75—80	3.22	4.19	3.78	22.05	30.26	26.19	2.71	3.67	3.32
80—85	4.31	5.79	4.79	31.23	38.65	36.30	3.71	4.79	4.28
85—90	5.99	7.73	7.08	43.46	50.31	49.09	4.94	6.66	5.94
90—95	8.60	12.46	10.70	60.87	67.41	66.88	7.06	9.81	8.29
95—100	84.66	49.28	84.66	100.00	100.00	100.00	15.83	18.60	18.72
96	9.19	13.53	12.75	65.30	71.97	70.92	8.88	12.96	11.82
97	10.27	15.15	14.64	70.04	76.98	75.73	9.66	14.22	13.74
98	14.91	16.64	16.64	76.29	82.54	81.20	12.71	15.80	15.64
99	20.05	21.90	21.82	84.73	88.94	87.99	16.90	18.63	18.76
100	84.66	49.28	84.66	100.00	100.00	100.00	31.07	31.41	32.84
0—20	0.00	0.04	0.00	0.00	0.04	0.00	0.00	0.01	0.00
20—40	0.08	0.60	0.17	0.17	1.62	0.46	0.02	0.23	0.05
40—60	0.95	2.00	1.36	3.70	10.12	6.47	0.36	1.21	0.70
60—80	3.22	4.19	3.78	22.05	30.26	26.19	1.86	2.88	2.39
80—100	84.66	49.28	84.66	100.00	100.00	100.00	7.89	9.97	8.47
0—50	0.32	1.19	0.66	0.94	4.73	2.13	0.04	0.27	0.10
50—100	84.66	49.28	84.66	100.00	100.00	100.00	4.01	5.45	4.55
0—100	84.66	49.28	84.66	100.00	100.00	100.00	2.02	2.86	2.32

	sub-sample 1	sub-sample 2	combined
* number of sample villages	18	18	36
number of sample households	569	406	975

TABLE VIII

NATIONAL SAMPLE SURVEY OF INDIA: RURAL

8TH ROUND: JULY, 1954—APRIL, 1955*

HOUSEHOLD OWNERSHIP HOLDINGS

STATE: ANDHRA, CENTRAL SAMPLE

Fractile Group	Upper Limit of Size of Ownership Holding (Acre)			Cumulative Percentage of Owned Area Below			Average Size of Household Ownership Holding (Acres)		
	s.s.1	s.s.2	combined	s.s.1	s.s.2	combined	s.s.1	s.s.2	combined
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0—15	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00
15—20	0.01	0.01	0.01	0.00	0.02	0.01	0.00	0.01	0.01
20—25	0.01	0.01	0.01	0.01	0.03	0.02	0.01	0.01	0.01
25—30	0.02	0.02	0.02	0.03	0.05	0.04	0.02	0.02	0.02
30—35	0.04	0.05	0.04	0.07	0.09	0.08	0.03	0.03	0.03
35—40	0.26	0.09	0.10	0.15	0.15	0.15	0.07	0.05	0.06
40—45	0.52	0.42	0.51	0.73	0.41	0.59	0.45	0.23	0.35
45—50	0.95	0.61	0.78	1.57	1.00	1.31	0.65	0.51	0.59
50—55	1.52	1.01	1.02	2.95	2.02	2.48	1.07	0.89	0.96
55—60	2.10	1.40	1.68	5.27	3.29	4.27	1.81	1.11	1.37
60—65	3.02	1.80	2.32	8.49	5.12	6.55	2.50	1.58	1.95
65—70	3.87	2.50	3.02	12.74	7.53	9.85	3.31	2.10	2.70
70—75	4.74	3.01	4.02	18.21	10.76	14.06	4.26	2.80	3.42
75—80	5.77	4.61	5.39	24.95	15.02	19.86	5.25	3.69	4.74
80—85	8.09	6.34	7.05	33.66	21.50	27.33	6.78	5.64	6.11
85—90	10.72	9.41	9.86	45.39	30.54	37.51	9.14	7.83	8.32
90—95	15.34	15.42	15.34	61.35	43.96	52.36	12.42	11.69	12.07
95—100	324.52	345.52	345.52	100.00	100.00	100.00	30.10	48.55	38.93
96	17.08	18.69	17.31	65.62	47.87	56.30	16.46	16.62	16.34
97	19.03	25.82	21.04	70.45	52.90	61.00	18.16	21.88	19.04
98	22.98	35.64	28.18	75.65	59.94	66.98	20,96	30.65	24.20
99	31.56	50.17	42.57	82.52	70.13	75.15	27.13	44.39	33.87
100	324.52	345.52	345.52	100.00	100.00	100.00	67.42	130.03	100.54
0—20	0.01	0.01	0.01	0.00	0.02	0.01	0.00	0.00	0.00
20—40	0.26	0.09	0.10	0.15	0.15	0.15	0.03	0.03	0.03
40—60	2.10	1.40	1.68	5.27	3.29	4.27	0.99	0.68	0.82
60—80	5.77	4.61	5.39	24.95	15.02	19.86	3.83	2.54	3.20
80—100	324.52	345.52	345.52	100.00	100.00	100.00	14.61	18.45	16.35
0—50	0.95	0.61	0.78	1.57	1.00	1.31	0.12	0.09	0.11
50—100	324.52	345.52	345.52	100.00	100.00	100.00	7.66	8.60	8.05
0—100	324.52	345.52	345.52	100.00	100.00	100.00	3.89	4.34	4.08

*	sub-sample 1	sub-sample 2	combined
number of sample villages	18	18	36
number of sample households	343	360	703

of rural India (including Jammu and Kashmir) on holdings of land by households. Data for two States, namely, West Bengal and Andhra, may be used to illustrate the graphical method for comparisons between different georgraphical regions in the same period of time. There were twelve interpenetrating samples. Out of this material, consider only two sub-samples of 18 villages each from only two states, West Bengal and Andhra, with a total number of villages of 38,590 and 18,912, respectively. The number of sample households in West Bengal was 569 in the first sub-sample and 406 in the second sub-sample with a combined sample of 975 households from an estimated number of 5,413,000; the sampling fraction was about one in 5,600. In Andhra, the number of sample households was 343 and 360 in the two sub-samples, respectively, giving a combined sample of 703 out of an estimated number of 4,066,000 households; the over-all sampling fraction in this case was about one in 5,800. The variate selected for the present example is the land owned by each household in the rural area. There are, of course, many households owning no land, for example, households of landless labourers, artisans, professional people, etc., but they are included in the present study. The sample households were ranked by the size of their landholdings; and using appropriate probability weights, the number of households in the population and the area owned by each were estimated in the usual way. The estimated number of households was then divided into fractile groups and the land owned in each fractile group was also estimated. From this it is possible to calculate the cumulative percentage of owned area and also the average size of household ownership landholding in each fractile group. The data for West Bengal are shown in Table VII in which columns (2), (3), and (4) give the limiting value at the upper end of each fractile group of the size of ownership holdings in acres for the first sub-sample, the second sub-sample, and the combined sample, respectively. The next three columns—(5), (6), and (7)—give the cumulative percentage of land owned below and inclusive of each fractile group; and columns (8), (9), and (10) give the average size of a household ownership holding in each fractile group. Similar data are given for Andhra State in the corresponding columns of Table VIII.

3.2. Consider Figure 7.1 in which the x-axis represents the percentages of households (on the basis of ranking by amount of land owned) and the y-axis represents the average size of the land owned by households for each fractile group. The values for each of the five 20 per cent fractile groups have been plotted for both the sub-samples and the combined sample for both West Bengal and Andhra. The shaded area in each case gives the associated error. It can easily be seen that up to the 80th percentile there is overlapping of the error areas, showing that, on the basis of the available

samples, the average size of holdings in the two States cannot be considered different. The average values for the top 20 per cent group are very clearly separated; and the average size of owned land for the top 20 per cent of households, as a whole, can be considered to be definitely higher in Andhra.

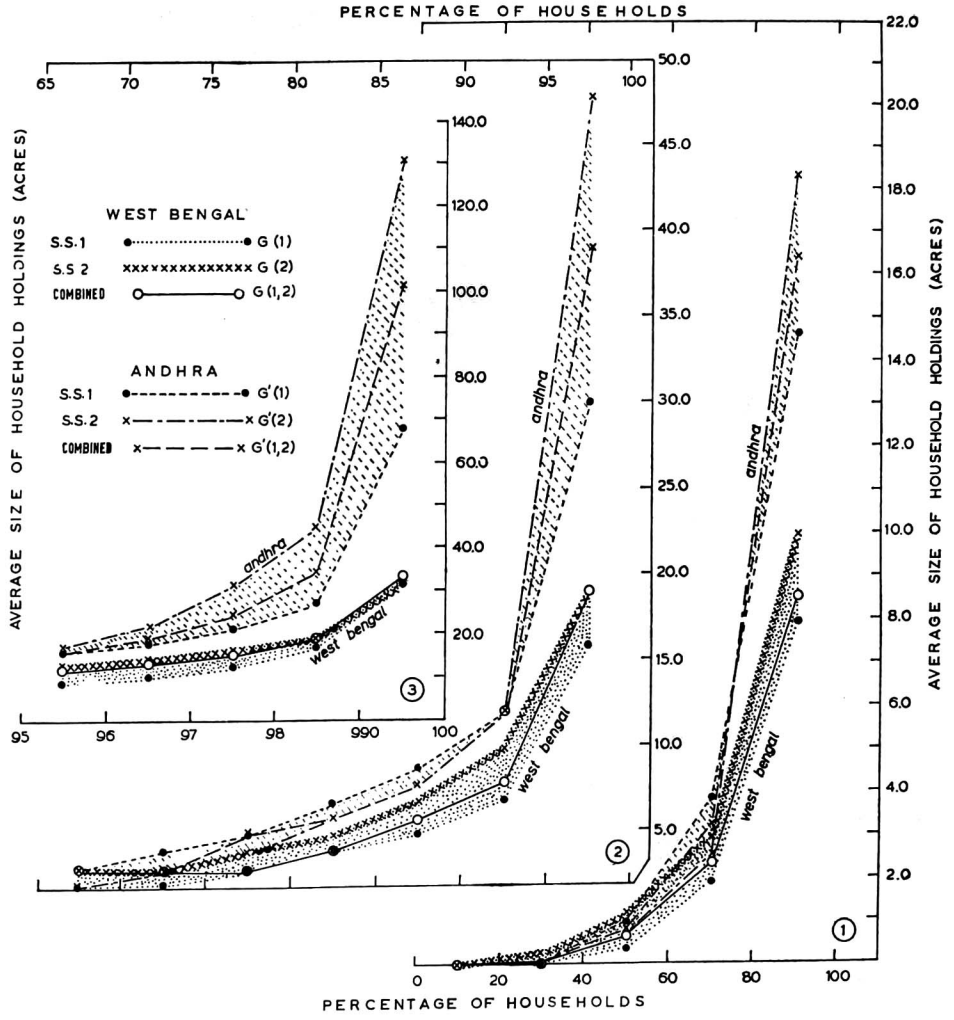

FIGURE 7.1—7.2—7.3

National sample survey of India, rural.
8th round: July, 1954–April, 1955.
Household ownership holding.

3.3. It is possible to make a more detailed examination of the range above 65 per cent of households. The average size of land owned in each 5 per cent fractile group between the 65th and 100th percentiles is plotted in Figure 7.2 for both West Bengal and Andhra; and the associated error areas are also

shown in the usual way. In comparison with the associated error, the separation is significant for the group falling between the 90th and 95th percentiles, and is on the verge of significance for the fractile group between the 80th and 85th percentiles.

3.4. It is possible to go a step further and plot the average size of holdings for 1 per cent fractile groups between 90 and 100 per cent. This is shown in Figure 7.3. There is clear separation, even for 1 per cent groups, beyond 90 per cent of the households, but the separation becomes significant only at the level of the top one per cent of households. In considering these results, the extremely small size of the sample, only 18 villages in each subsample, must be kept in mind. The flexibility of the present method can, however, be easily appreciated.

4. COMPARISON WITH FREQUENCY DISTRIBUTIONS

4.1. It is worthwhile to contrast the present approach with that of the frequency distribution of the usual type in which the class ranges would be fixed, say, in terms of the money value of the per capita expenditure (such as Rs. 5, 10, 15, 20, etc.). In economic data, the money value of the expenditure would, in general, change with changes in price. The population of rural households as a whole in the 7th round may, of course, be compared in a meaningful way with the population of rural households as a whole in the 9th round. A fixed range frequency class would, however, represent different fractile groups in the two rounds and would not, therefore, be comparable in any important sense. The use of the *same* fractile groups would avoid conceptual difficulties involved in making comparisons over time or space or for two populations differing in any way. In the fractile approach, the bottom 10 per cent or the top 5 per cent, etc., of households in one round of the survey can be considered to be the counterpart of the bottom 10 per cent or the top 5 per cent, etc., of households in another round of the survey. This would also be true for comparisons between two States or two geographical regions for the same round. These fractile groups, in other words, may be treated as so many economic "strata" of the whole population and comparisons over time or space of the same stratum would be meaningful for many purposes.

4.2. The contrast can be expressed in a slightly different way. In using fixed ranges (with varying frequencies) the main interest lies in the pattern of the frequency distribution as a whole. In using equi-frequency fractile groups (with varying class intervals) it is possible to use each fractile group itself as a stratum or unit of comparison. The effect of price changes may,

for example, be different at different levels of consumption expenditure, that is, in different fractile groups, and the picture for the population as a whole may become blurred. In the fractile method it is possible to study the effect for each fractile group separately, that is, to break up the whole spectrum of the range of expenditure into smaller and more homogeneous groups.

4.3. The fractile graphical approach offers an extremely rapid and practical method of analysis of statistical data of all types on a large scale. It is being used extensively now in the National Sample Survey of India. It would seem desirable to explore the possibilities of its applications and its usefulness in other fields.

In conclusion I should like to express my thanks to my colleagues in the Indian Statistical Institute who helped in the preparation of this paper.

APPENDIX

A METHOD OF FRACTILE GRAPHICAL ANALYSIS

The theoretical basis of the fractile graphical method is briefly explained in this appendix.

1. THEORETICAL BASIS

1.1. The *error* associated with the combined fractile graph $G(1,2)$ is defined as the area lying between the two component sub-sample fractile graphs $G(1)$ and $G(2)$, and can be measured on the chart. This may be called e for the first population and e' for the second population.

1.2. For a non-graphical (that is, purely numerical) method of analysis, it is possible to define the error in either of the two usual forms, namely, (a) the sum of the differences (neglecting the sign of the difference) between the values of y for the two sub-sample graphs $G(1)$ and $G(2)$ for all fractile groups, or (b) the sums of the squares of these differences. All three were studied by model sampling experiments; and it was found that all three had similar distributions. The graphical definition has been selected, however, because of its simplicity. Once the accumulated totals are found (as explained in paragraphs 1.3—1.9 of the text) all the graphs can be easily drawn, and can also be interpreted directly by visual examination. In fact, junior computers can quickly learn this method of analysis, which makes it possible to use it on a very large scale. Secondly, the graphical method shows to what extent these two graphs cross and recross each other, and can sometimes reveal whether systematic non-sampling errors are present.[1]

[1] The presence of such non-sampling errors, of course, makes it impossible to give a rigorous mathematical theory, but this difficulty cannot be avoided as it is inherent in the method of a sample survey itself.

1.3. The combined fractile graph $G(1,2)$ will usually lie partly within and partly outside the error area e; and may occasionally lie entirely outside the error area e.

1.4. The *separation* has been defined as the area lying between the two combined graphs $G(1,2)$ and $G'(1,2)$.

2. RESULTS

2.1. A number of plausible results can now be stated.[2] Let N be the total number of sample units, g the number of fractile groups, and n the number of sample units in each fractile group, so that $N = gn$.

2.2. The combined graph $G(1,2)$ would tend stochastically to lie entirely within the error area e as n increases when g is kept constant.

2.3. The error area e would tend to decrease inversely as $1/\sqrt{n}$ when g is kept constant, or as $\sqrt{(n_1 + n_2)/n_1 n_2}$ if the size of the two sub-samples are respectively n_1 and n_2 in each fractile group.

2.4. The error area e would tend to increase proportionately to g, the number of groups, when n remains the same (or when n_1 and n_2, the number of sample units for the two sub-samples in each fractile group, are kept constant).

2.5. Since $N = gn$, it follows that the error area e would tend to vary as $g^{3/2}$ when N is kept constant. If g is changed, and the new value of g is k times its original value, the relative changes in the error area would be approximately proportional to $k^{3/2}$. This is a most useful property as it enables one to use different values of g in testing the significance of the separation.

2.6. It is plausible to assume that the error (say, E) associated with the "separation" (say, S) would be given by $E^2 = (e)^2 + (e')^2$; and that (S^2/E^2) would tend to be distributed proportionally to χ^2 with g degrees of freedom.

2.7. When x and y are both random variates and are also statistically independent, the number of intersections of the two sub-sample graphs $G(1)$ and $G(2)$ would tend to be distributed like "runs" of heads and tails in g throws of an unbiased coin.

2.8. The above results would remain true for any set of linear and nonlinear transformations of the values of x and y in all the sub-samples.

3. SOME SUGGESTIONS FOR FURTHER WORK

3.1. The above results are, of course, only approximate, and the exact results would depend on the statistical distribution of both x and y and on the relationship

---

[2] These were first given in lectures at the Indian Statistical Institute in Calcutta in April, 1958 and at Berkeley, Chicago, and East Lansing in the United States in May and June, 1958. A preliminary note with the results of some model sampling experiments was published in the *Transactions of the Bose Institute*, Vol. XXII, 1958, pp. 223–230: "A Method of Fractile Graphical Analysis with Some Surmises of Results." Some further observations were made in lectures in Tokyo and Kiyushu, Japan in November and December, 1958.

between the two variates. For example, if n and g (and necessarily N, the size of the sub-sample) are kept the same, then the error area e for sampling from a population in which there is high correlation between y and x would tend to be smaller than the error area for sampling from a population in which y and x are less closely associated.

3.2. Many model sampling experiments have been carried out, and the results are generally in agreement with the above conjectures within the margin of errors of sampling. Some of the results of the model sampling experiments have been published; other results are in the course of publication. The model sampling results present interesting suggestions. For example, the error area e may vary proportionately to $(g - \frac{1}{2})$ rather than to g, when n is kept constant. The number to be subtracted from g may not be exactly half, but some such correction factor would probably be necessary.

3.3. Results of model sampling experiments can show the plausibility of the conjectures but can never prove them in a theoretical sense. Theoretical investigations are, therefore, proceeding and some asymptotic results have been obtained[3] which are broadly in agreement with the conjectures given above. A rigorous theory would open up many possibilities.

Indian Statistical Institute

---

[3] The following papers are being published in *Sankhyā*, Vol. 23, Parts 1 and 2: P. K. Bhattacharya and K. R. Parthasarathy, "Some Limit Theorems in Regression Theory," J. Sethuraman, "Some Limit Distributions Connected with Fractile Graphical Analysis," and K. Takeuchi, "On Some Problems of Error Area in the Fractile Graphical Method."