B. Chakraborty

An Analysis of the Spatial Distribution of the District Level Fertility Data in India : Preliminary Observations

AN idea of the spatial distribution of fertility at the district, census tract or even to a ward level, rather than at the national or state level has meaning and relevance as a source of information for the planning of human fertility at the micro-regional level- National fertility is only but a weighted average of fertility of its micro-regional parts weighted by the frequency of population in these geographical units. National fertility planning will invalidate much of its usefulness unless understanding of its specific problems on a small area basis is recognised and unless how rates at which vital process is functioningin different parts of the country is realised. There is an internal diversity or place-variance in fertility. The primary object in this paper will be concerned with place-variance and estimation of the probability distribution of fertility by geographical regions.

Of the total number of 336 districts recorded in the 1961 census, 324 are considered for the study. Others are left out on account of non-availability of adequate information.

When more accurate census data and no usable registration data are available at the district level, census age-distribution can be fruitfully used with judicious consideration for having at least an indirect measure of fertility. Such an index is c/w (child-woman) ratio. The index of c/w ratio could be converted into GFR (General Fertility Rate), which after standardisation for age, sex and marital status would perhaps provide the best measure of fertility for the purpose of such comparative studies. That would require an estimate of the number of children born on the average 2½ years prior to the census date and also an estimate of women who would have survived in their reproductive life 15 to 49 years of age during the same time periods, exposed to the risk of giving birth. Very scanty data or no data being available for mortality at the district level and there being no L.T. (Life Table) at the micro-regional level which could be used, stable population technique and resort to model life table values is the only alternative approach for computing these rates. Even then, assumption of stability is not tenable due to the varied nature of the age distribution at the district level. On the other hand, some of the districts having particularly very high growth rates possibly due to migratory effect makes it unrealistic and impossible to use this technique, unless growth rates of such districts are further scanned. The process is laborious as too many estimating problems are involved in processing the data for computation. Conversion of c/w ratio to GFR was, therefore, not attempted for the present.

The c/w ratio does not take account of number of infant deaths that take place during the period and assumes that they are equally distributed over all areas being compared. It understates women and particularly child population. The latter one has been considered in the paper. Here the unstandardised measure of c/w ratio has been used with no intervening variables controlled.

The age-distribution in a census has to be smoothed and graduated, before it is used. A mathematical curve of the third degree was fitted to each of the districts age distribution separately for males and females on the assumption that nature of undercount and irregularities in the age data uniformly apply to all districts. On the assumption again that the undercount is confined to the age-group 0-5 only, $F(x)$, the ratio of enumerated population at age x and over to the total enumerated population, expressed as a percentage is obtained in quinquennial steps for values of x from 5 to 60, which need correction only for errors in age, is graduated by the formula

$$U_x = a + bx + \frac{cx(x-1)}{2} + \frac{d(x)(x-1)(x-2)}{6}, \qquad (1)$$

where $U_0 = F(5)$, $U_1 = F(10)$ and so on.

The parameters of the above equation are obtained by the following method.

We have, from which taking the values in successive triennial groups

$$U_0 = a \qquad W_0 = U_0 + U_1 + U_2 = 3a + 3b + c$$
$$U_1 = a + b \qquad W_1 = U_3 + U_4 + U_5 = 3a + 12b + 19c + 15d$$
$$U_2 = a + 2b + c \qquad W_2 = U_6 + U_7 + U_8 = 3a + 21b + 64c + 111d$$
$$W_3 = U_9 + U_{10} + U_{11} = 3a + 30b + 136c = 369d$$

.

.

.

.

$$U_9 = a + 9b + 36c + 84d$$
$$U_{10} = a + 10b + 45c + 120d$$
$$U_{11} = a + 11b + 55c + 165d$$

Therefore,

ΔW_0	$\Delta^2 W_0$	$\Delta^3 W_0$
$9b + 18c + 15d$		
$9b + 45c + 96d$	$27c + 81d$	
$9b + 72c + 258d$	$27c + 162d$	$81d$

from which d, c and b are obtained successively from the values of leading differences beginning from the third and proceeding backwards, and finally from the values of W_0 $(U_0 + U_1 + U_2)$. Calculating the values of U_x for the required values of x, from equation (1), the graduated values of $F(5)$, $F(15)$ and $F(49)$ denoted U_0, U_2, U_9 were obtained. The values of $F(0)$ corrected for undercount is given by 100.5, so that the ratio of the graduated population 0-5, to that of females in ages 15-49 expressed per 1000 females is given by

$$\frac{(100\,5 - U_0)}{U_2 - U_9}.$$

The third-order constant everywhere shows a more or less regular trend and values also are significantly small enough indicating smoothness and a suitable fit of the curve to the data. Moreover, frequency of the number of changes in sign of deviation more or less corresponds to that for nonchanges and it is large. This indicates satisfactory graduation. Effect of graduation has relatively raised the number of children enumerated in ages 0-4 at different degrees for individual districts. The number of women in 15-49 years of age has also been automatically adjusted by the process of graduation. The inflation factor for number of children shows a high degree of dispersion among the districts. The extent of underenumeratlon of children to a significant level in census has a great differentia] effect in the computation of *c/w* ratios for comparative purposes. It was therefore, adjusted first by using the correction factor and *c/w* ratios computed.

The following table gives the frequency distribution of *c/w* ratio for the 324 districts by region.

TABLE 1—FREQUENCY DISTRIBUTION OF c/w RATIO FOR 324 INDIAN DISTRICTS

Graduated c/w ratio	*No of observed districts*
< 500	2
501— 550	11
551— 600	13
601- 650	26
651- 700	41
701— 750	54
751—800	87
80,1- 850	48
851— 900	15
901— 950	12
951-1000	7
1001-1050	4
1051—1100	3
1101 & above	1

Fig. 1 in the appendix depicting the distribution shows a somewhat positively skewed curve with mode, median and mean in a systematic lower order.

The analysis was next carried out by serially arranging the *c/w* ratios in ascending order of magnitude. A section of each of 32 units being a decile group of 324 districts and also for India as a whole, the geographical distribution of fertility or, in other words, the scatter of actual values of fertility about the mean, was studied. Some of the statistics describing the distribution may be observed from the following table.

TABLE 2—SHOWING THE DISTRIBUTION OF *c/w* RATIO BY GROUPS OF DISTRICTS AND THAT FOR ALL INDIA

Decile groups of districts	*Mean* $\bar{X}$	*Median*	*1st Quartile* M_1	*3rd Quartile* M_3	*Semiinter quartile ranges or quartile deviation*	*p. c. above or below All India Mean*	*Remarks*
I	564	567	529	603	37.0	− 25.1	
II	646	643	626	666	10.0	− 14.3	
III	692	694	683	702	9.5	− 8.2	
IV	721	721	712	732	10.0	− 4.3	
V	747	744	742	752	5.0	− 0.8	
VI	762	762	757	768	15.5	+ 1.1	
VII	783	783	778	787	4.5	+ 4.0	
VIII	806	807	796	813	8.5	+ 7.0	
IX	840	841	829	849	10.0	+ 11.5	
X	1013	943	901	984	41.5	+ 34.5	
All India	753	...	...	...	...	...	Covering all districts.
	742	756	694	807	56.5	...	Excluding some

The All India average in *c/w* is of the order of 753 per 1000 women in reproductive group of 15 to 49 years of age. Excluding the districts which have been omitted from the list the average comes down to a level of 742. The median stands at 756. The distribution curve seems to be almost symmetrical having its peak at the centre, and with the right side little extended. As a symbol of average and absolute variability, median, quartile deviation, and semi-inter-quartile range are computed. Semi-inter-quartile range becomes larger and larger as we go away from the central position. This indicates dispersion to be higher both at the lower and higher level of fertility. The all-India range is 56.5. To facilitate comparison, the averages for each of the ten groups are expressed as ratios to national average. Taking, nation as I, a value below, or above it, is treated as below average, or above, average fertility. Multiplying deviation from 100 and then substracting 100 from the result, it directly states the percentage by which fertility of particular group exceeds or falls short of national average. The distribution so obtained in the last column in the table shows a high degree of dispersion among the groups of districts from national fertility index. At the lower side, it falls short of by about 25% and at the other extreme it is beyond average by 34.5%.

The distribution of mean and median by decile groups mutually corresponds indicating nearing normality of the curve. This can further be studied by examining the distance of the quartiles away from the median. Most of the quartiles are roughly about half the distance away from the median, indicating the distribution to be normal. Though quartile deviations can be described as mean expectation of the deviation from the mean or median, this is not really a true measure of the overall scatter of occurrences.

As an index of variability for measuring the relative dispersion, C.V. (Coefficient of Variation) has been computed for each of the ten decile groups of districts, all-India, and also the two groups of districts—one below the national average, the other above it. These indices are presented below. Variability indices express deviation values as percentage of mean, thus eliminating direct influence of mean and thereby facilitating comparison in relative terms between the sets of data. When probability of a certain event is to be assumed, C. V. is essential. For ordering districts in terms of magnitude of variability, C. V. may be of great value. When distribution map of variability is required as an index of variability, this percentage value is useful. The variability indices represent percentags variability of the sector specific data.

A look at the Table 3 shows that the *a* is relatively higher at extreme levels of fertility; the C. V. also reflects the same picture. At the two extreme categories both a and C. V. are about 8-10 times higher than that observed at the

Table 3—VALUES OF σ, σ^2 AND COEFFICIENT OF VARIATION FOR TEN DECILE GROUPS OF DISTRICTS; ALL-INDIA AND FOR GROUPS OF DISTRICTS ABOVE AND BELOW ALL INDIA MEAN

Decile groups of districts	*Mean* $\bar{X}$	*S. D.* (σ)	*Variance* (σ^2)	*C. V.* (%)
I	564	40.5	1639.6	7.2
II	646	18.3	335.4	2.8
III	692	9.2	85.4	1.3
IV	721	10.3	106.8	1.4
V	747	5.2	27.0	0.7
VI	762	5.8	33.3	0.8
VII	783	5.3	28.3	0.7
VIII	806	8.6	73.5	1.1
IX	840	12.6	159.2	1.5
X	1013	66.1	4379.3	6.9
All India	753	106.0	11288.0	14.1
Districts below mean	670	106.3	11295.9	15.9
Districts above mean	828	106.3	11282.5	12.8

medium level of fertility. In case of all-India, the σ value is 106 and C. V. is 14.1 percent. At levels for groups below and above the mean σ corresponds to that with all-India but with variation in C. V.

When mean and s.d. (standard deviation) are known, it is often useful to know the region within which required percentage of observations would lie. On the other hand in comparing the distribution of district level fertility with the normal probability curve, the following table indicates that about 40 percent of the districts (about 133) lie within $+1\sigma$, 7.5 percent within $+2\sigma$, 3.3 percent within $+3\sigma$ and 0.6 p.c. within $+4\sigma$. Above the means the observations with same σ apart but on the $-ve$, the percentages are 33.4, 11.7 and 3.3 with no value for -4σ. On the whole 51.5 percent of the districts lie above the mean and 48.4 p.c. below it.

TABLE 4—PERCENTAGE POINTS IN THE NORMAL PROBABILITY CURVE AND THOSE IN THE OBSERVED DISTRIBUTION

Deviation of σ from mean	*Observed distribution*	*Normal probability distribution*
− 1	33.4	34.0
− 2	11.7	13.5
− 3	3.3	2.3
− 4	0 0	0.3
+ 1	40.1	34.0
+ 2	7.5	13.5
+ 3	3.3	2.3
+ 4	0.6	0.3

In all these calculations, normal frequency distribution is assumed, but this does not always occur. So index of overall variability will not adequately reflect the differing tendencies and degree of variability above and below the mean. This is of major importance if calculations of probability are to be made. It is, therefore, desirable to calculate deviation above or below the mean. As the above table indicates, the half below the mean broadly corresponds with the standard distribution but in considering the individual level of s.d. dispersion, there is a significant variation between the observed and theoretical for $+ 1\sigma$ and $+ 2\sigma$. If these two ranges of σ's are merged, observed and theoretical distribution correspond considerably.

It may be of interest to compare the value of fertility with that expected to occur with a given probability. For example, it could be of interest to define the value of fertility which is equal or below some '*x*' percent of the districts. The value must therefore of necessity be below the mean and *d* value given by

$$\frac{x - \bar{x}}{\sigma}$$

will be a *−ve* one. This *d* indicates the extent to which critical value differs from the mean expressed in terms of σ. Knowing *d* the normal deviate value, the required percentage probability is obtained from table of normal distribution curve.

To obtain *d* value it is again necessary to consult the same table to find the value of *x* (critical value) which will ensure the desired percentage of occurrences fall below it. Thus if it is desired to know a definite percentage of districts with fertility less than the critical value, the calculation would be x = *d*, *a* + *x*. Reference to table shows that for the 11.51 percentage points, *d* value is 1,20 and that for 9.68 it is 1.30. From this, the interpolated *d* value for 10 percentage points is obtained as 1.27. Using the value *of d in* the above formula, *x*, the critical value is computed as 618. This means that 10 per cent of the districts in an infinite series will possess less than or equal to level of fertility, indicated by 618 as *c/w* ratio. By the same procedure, for percentage of districts at every decile group possessing estimated *d* values, the expected probability level of variant critical values of fertility have been obtained in Table 5 below.

Table 5—PERCENTAGE OF DISTRICTS POSSESSING < OBSERVED AND EXPECTED LEVEL OF *C/W* RATIO

Percentage districts	*values of d*	*Level of c/w ratio equal to or below the level*	
		observed	*expected*
10	1.27	619	618
20	.84	671	664
30	.53	706	697
40	.25	737	726
50	.00	755	753

It is seen that the observed and expected distributions of *c/w* ratio are consistent to a significant level, though expected is always below the observed level.

APPENDIX

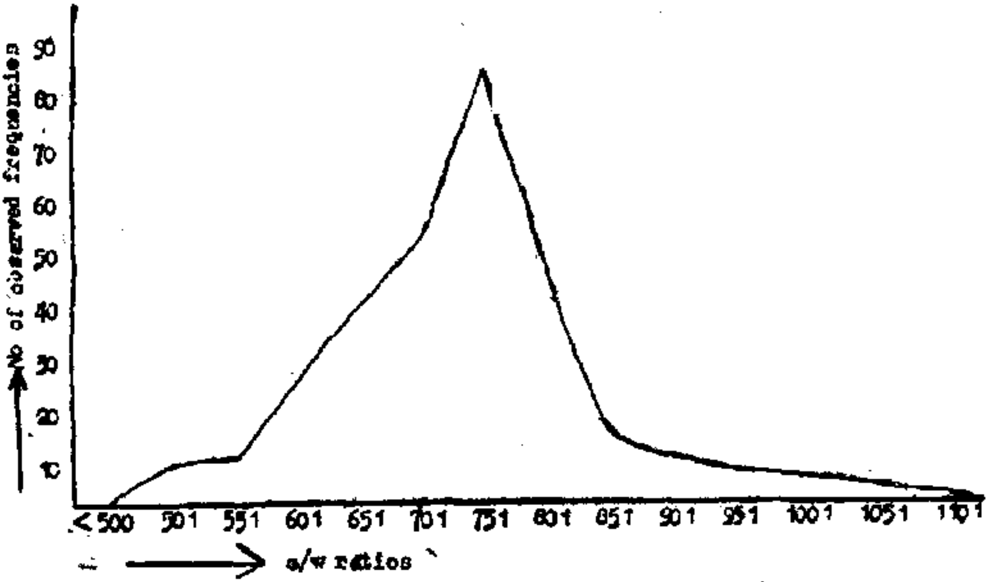

Fig. 1. Frequency distribution of c/w ratio for 324 Indian districts.