Nimai Kumar Chandra

Adjustment of Age Data for India's Census Population

1. Introduction

FROM the Indian census data, we get the population figures corresponding to the census date at each age (l.b.d.) from 0 to *86* and also for the age group 87+, separately for males and for females, besides getting the total population. But the population figures so obtained for each age are found to be highly unreliable. This is partly due to incomplete enumeration, especially of children of age 0 to 4, but mainly due to the general tendency among the people to report even ages and also ages that are multiples of 5. Here we shall study the age distribution for possible heaping at certain ages and effect some adjustments to the age data assuming that the total size of the population is, on the whole, reliable. The technique used for the smoothing and adjusting of the population figures is based on fitting a polynomial regression equation of sufficiently high degree.

2. Census Population by Age and Sex

The population figures for both males and females are collected from the publication, The Census of India, 1971: Series I, India — *Age and Life Tables.* These do not include the population of Sikkim which was merged in India after the census. Table 2 gives the age-sex distribution of the population as given by the Census. From this table we find that the population data are very irregular.

To bring the extent of irregularity into clearer relief, we construct an age pyramid, as shown in Fig. 1.

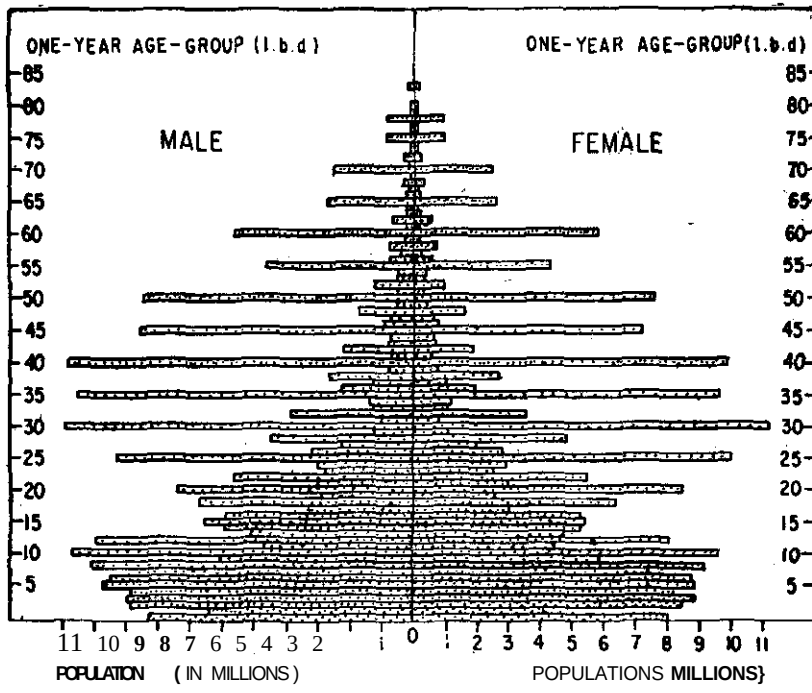

Fig. 1. Age-pyramid of Indian population according to (he *1971* census.

3. Extent of Heaping at Ages that are Multiples of 5

The pyramid makes it evident that the population for each sex is heavily concentrated at even ages and, all the more so, at ages that are multiples of 5. The figures for age 0 to 4 also indicate that there has been considerable under-enumeration of the population in the early years of life. The whole pattern is quite at variance with the pattern normally observed for developing countries with reliable population data as shown by Paul Demeny [1].

4. Preliminary Adjustment : Distribution in 5 Year Age-Groups

We have noted that the population for males as well as for females is concentrated largely at ages that are multiples of 5. It is necessary to adjust the population figure for each age so as to make the whole set of population figures more reliable. Adjustment for each age is a very difficult task and will involve

assumptions that may be unwarranted. So instead of adjusting the population figures at individual (one-year) age-groups we shall try to make some adjustments only for the population figures in 5-year age-groups.

To make such adustment, at the first step, it is reasonable to assume that the heaping occurs at age $5j$ because of misstatement or misreporting of age in respect of the population at ages $(Sj - 2)$ to $(5j + 2)$ alone. Thus although the population figure at any individual age in the age-group $(5j - 2)$ to $(\$j + 2)$ is unreliable, the total population in this age-group may be assumed to be sufficiently reliable. Table 3 shows the population by quinquennial age-groups for each sex according to the 1971 census. The same data are shown for males and females in Figs. 2*a* and 2*b*, respectively.

The pattern of age-distribution represented by Table 3 (or Fig. 2) conforms

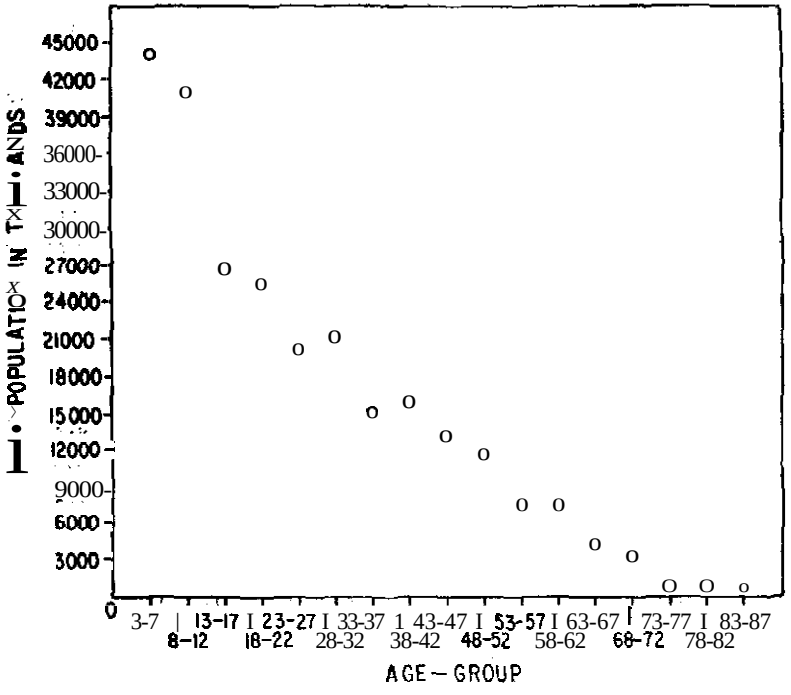

Fig. 2a. Male population in 5-year age-Groups according to the 1971 Indian census.

to the pattern expected for developing countries, to which we have already referred. We may, therefore, reasonably take these adjusted age data to be far more reliable than the original data.

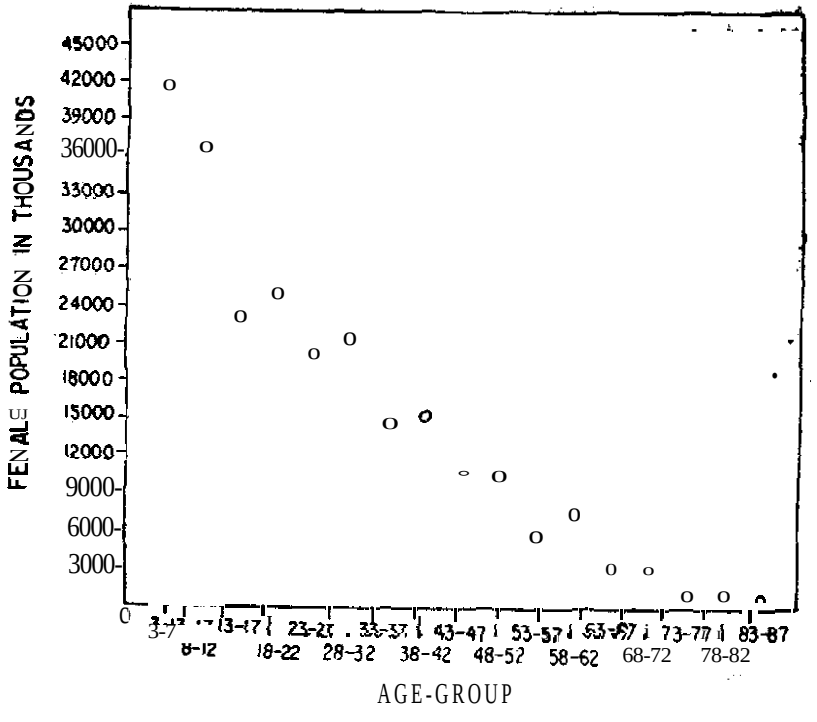

Fig. 2b, Female population in 5-year age-groups according to the 1971 Indian census.

5, Smoothing of Age Data by Orthogonal Polynomials

We shall now effect a further adjustment of the quinquennial age data through a process of smoothing based on orthogonal polynomials. We shall denote by Y_t the population in the jth age-group (i.e. from age $5j - 2$ to $5j + 2$) for $j = 1, 2, \ldots, 17$. If $\phi_i(\cdot)$ denotes the orthogonal polynomials of degree i for seventeen consecutive values of the independent variable, with common difference unity, then we consider an equation of the form

$$\log_e Y(x) = \sum_{i=0}^{5} a_i \phi_i(x) \qquad (1)$$

for $x = -8, -7, \ldots, 1, 0, 1, 2, \ldots, 7, 8.$

where $x = j - 9$ and a_i are constants to be determined. We take $\log_e Y_j$ rather than Y_j to be the dependent variable in view of Fig. 2. We consider orthogonal polynomials upto the 5th degree because Fig. 2 suggests that this will be ade-

quate and also because the higher the order of polynomials, the less efficient will be the estimates of the coefficients (or the related tests).

The values of the orthogonal polynomials are taken from Fisher and Yates[3]. The coefficients are given in the following table.

TABLE 1—COEFFICIENTS OF ORTHOGONAL POLYNOMIALS

Coefficients	*Male*	*Female*	X *(adjustment factor in orthogonal polynomial)*
a_0	9.104482	9.030800	1
a_1	-0.239184	—0.239654	1
a_2	—0.013749	-0.0)2938	1
a_3	-0.007647	-0.005496	1/6
a_4	0.002335	0.001754	1/12
a_5	0.000854	0.000133	1/20

6. Tests for the Significance of the Regression Coefficients

Tests on regression coefficients given in Table 1 are meant to indicate which, if any, of the coefficients in the fifth-degree polynomials may be left out to fit $\log_e Y(x)$. We have tested the following four hypotheses

$$Ho_i : a_i = 0, i = 1, 2, \ldots 5$$

both for males and for females under the usual assumptions of normality, equality of variances and independence of the different observations on Y_j. It is found that in case of males Ho_5 & Ho_4 and in case of females Ho_5, Ho_4 and Ho_3 are to be accepted at the 5% level of significance. The test procedure is indicated in [4], We leave out the terms that are found insignificant according to the test procedure, so that our fitted equations are :

$$Y^m(x) = \exp [9, 434451 - 0.184382x - 0.013749x^2 - 0.001274x^3] \quad (2)$$

$$Y^f(x) = \exp [9.341392 - 0.239657x - 0.012938x^2]. \quad (3)$$

(m stands for males and females, respectively)

Table 4 gives the fitted population by age and sex obtained by using equations (2) & (3).

TABLE 2—AGE-SEX DISTRIBUTION OF THE INDIAN POPULATION ACCORDING TO THE 1971 CENSUS

Age (Yrs. l.b.d.)	Population in thousands		
	Males	Females	Persons
1	2	3	4
0	8225	8037	16262
1	5310	5154	10464
2	8825	8475	17300
3	8898	8845	17743
4	8744	8175	16919
5	9677	8798	18475
6	9406	8699	18105
7	7638	7348	14986
8	9923	9121	19044
9	6003	5825	11828
10	10605	9569	20174
11	4692	4428	9120
12	9746	8022	17768
13	5067	4688	9755
14	5923	5184	11107
15	6509	5313	11822
16	5866	5203	11069
17	3287	2987	6274
18	6894	6330	13224
19	2833	2542	5375
20	7352	8380	15732
21	2997	2426	5423
22	5557	5421	10978

Table 2 (contd. on page 280)

Table 2 (*contd. from page* 279)

1	2	3	4
23	2754	2554	5308
24	2890	3910	5800
25	9197	9931	19128
26	3064	2833	5897
27	2245	1999	4244
28	4425	4777	9202
29	1184	1073	2237
30	10770	11350	22120
31	1049	805	1854
32	3810	3517	7327
33	1285	1102	2387
34	1302	1237	2539
35	10400	9481	19881
36	2217	1925	4142
37	1220	1014	2234
38	2589	2593	5182
39	782	721	1503
40	10726	9748	20474
41	686	505	1191
42	2231	1815	4046
43	777	642	1419
44	695	592	1287
45	8513	7147	15660
46	990	768	1758
47	730	559	1289

Table 2 (*contd. on page* 281)

Table 2 (contd. from page 280)

1	2	3	4
48	1694	1543	3237
49	498	394	892
50	8410	7410	15820
51	483	319	802
52	1256	943	2199
53	473	357	830
54	512	392	904
55	4553	4051	8604
56	740	588	1328
57	404	273	677
58	788	719	1507
59	256	202	458
60	6024	5771	11795
61	272	198	470
62	675	542	1217
63	252	181	433
64	225	171	396
65	2719	2539	5258
66	229	179	408
67	187	149	336
68	309	292	601
69	113	94	207
70	2525	2422	4947
71	128	82	210
72	262	207	469
73	83	60	143

Table 2 (contd. on page 282)

Table 2 (*contd. from page* 281)

1	2	3	4
74	75	56	131
75	893	869	1762
76	95	76	171
77	46	33	79
78	94	86	180
79	30	23	53
80	876	949	1825
81	34	26	60
82	61	53	114
83	21	14	35
84	24	18	42
85	215	217	432
86	20	18	38
87+	397	428	825
ANS	478	501	979
Total	283937	264013	547950

The fitted population figures together with the corresponding observed data are shown in Fig. 3.

7. Adjusted Population by Age and Sex

In Table 4 we have the adjusted populations for the age-groups $(5j-2)$ to $(5j + 2)$ years l.b.d., $j = 1, 2, \ldots$. But we want the adjusted populations for the age-group $5J$ to $Sj + 4$ years l.b.d., j— 0, 1, 2, ... Thus the adjusted populations for these groups are obtained by using equation (2) and equation (3) and replacing x by x where $x = x - .6$, $x = -8, -7, .., -1$, 0, 1, 2, . , . These final adjusted population figures are shown in Table 5.

TABLE 3—QUINQUENNIAL AGE GROUPING OF THE INDIAN POPULATION BOTH FOR MALES AND FOR FEMALES ACCORDING TO THE 1971 CENSUS

Age-group (Yrs. 1.b.d)	*Population in thousands*	
	Males	*Females*
3-7	44363	41865
8-12	40969	36965
13-17	26652	23375
18-22	25633	25099
23-27	20150	20227
28-32	21238	21522
33-37	15204	14759
38-42	16006	15382
43-47	13206	11523
48-52	11815	10609
53-57	7534	5661
58-62	7744	7432
63-67	4100	3219
68-72	3337	3097
73-77	1192	1301
78-82	1095	1137
83-87	758	695

TABLE 4— POPULATION OF INDIA, 1971 BY AGE AND SEX AS OBTAINED BY FITTING ORTHOGONAL POLYNOM[ALS

Age-group (Yrs. 1.b.d.)	*Population in thousands*	
	Males	*Females*
3-7	45782	43250
8-12	37603	33265
13-17	30321	27453
18-22	24743	23650
23-27	20938	20904
28-32	18451	18675
33-37	16730	15197
38-42	15258	14431
43-47	13610	12144
48-52	11538	9785
53-57	9076	7500
58-62	6533	5456
63-67	4305	3778
68-72	2652	2516
73-77	1599	1626
78-82	1018	1050
83-85	764	698

TABLE 5—ADJUSTED POPULATION OF INDIA, 1971 BY AGE AND SEX

Age-group (Yrs.1b.d.)	*Population in thousands*	
	Males	*Females*
0-4	49710	34392
5-9	40181	33374
10-14	33487	31558
15-19	28558	29079
20-24	24730	26110
25-29	21580	22845
30-34	18832	19478
35-39	16310	16184
40-44	13911	13102
45-49	11596	10337
50-54	9376	7947
55-59	7296	5953
60,64	5424	4346
65-69	3822	3092
70-74	2533	2143
75-79	1567	1448
80-84	898	953
85-89	473	611
9)+	385	922
Total	290669	263874

S. Conclusion

Looking at Table 5 (or at the age-pyramid of Fig. 3) we find that the adjusted age-distribution figures of the population for both males and females have a very regular pattern, the population steadily decreasing with age. This pattern was completely lacking in the original census data. We note that the total sizes of adjusted population for males and for females are 29.07 crores and 26.39 crores, respectively, which are not too different from the corresponding sizes of the

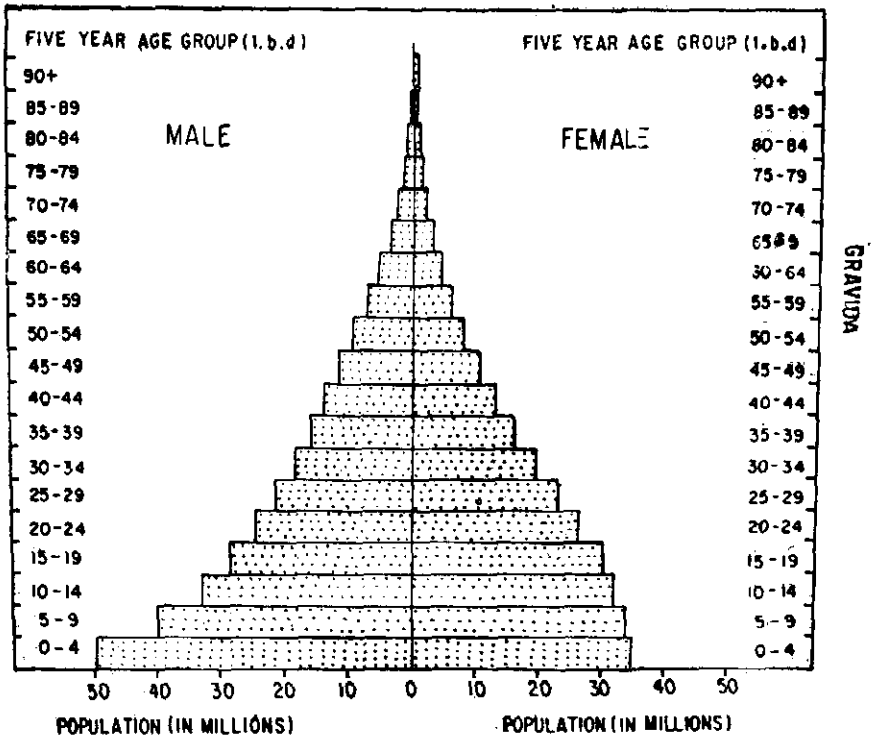

Fig. 3. Age-pyramid of the Indian population, 1971 (Adjusted).

original census populations for the two sexes. As such, we feel that the final adjusted figures may be used in advanced demographic work, say in population projection, with a measure of confidence.

Acknowledgement

I put on record my gratitude to Dr. A. M. Goon for his suggestions and guidance in the course of this work.

References

1. Demeny, P., 1974, The population of the under developed countries. In Flenagen, D. *et al.* (eds.), *The Human Population*, W.H. Freeman and Co., San Francisco.
2. Fisher, R. A. and Yates, F., 1963, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, Edinburgh Tweeddable Court London.
3. Kendall, M.G., 1951, *The Advanced Theory of Statistics*, II. Charles Griffin Company Ltd.
4. Registrar-General, India, Age and life tables, *Census of India 1971, Series I—India.*