Interpolation for Population Data: An Empirical Study

Barun Kumar Mukhopadhyay
Prasanta Kumar Majumdar

Abstract. *The UN Volume (Bogue, Arriaga and Anderton, 1993) on population methodology has elaborated different methods for interpolation of population data. There were two methods in interpolating empirical data frequently discussed, that are mid-point method and cumulation-differencing method. In some cases mid-point method was not recommended, and cumulation-differencing method was recommended only on the basis of limited data. This study tries to explore other methods in manipulating population data for different nature of data. The paper when at one instance, finds that the method which was not recommended by the said volume was acceptable using different nature of data. In other case, the method which was recommended in the volume was found more sound in respect of other kind of data. In addition, some new types of data have also been tried and appropriate interpolation formulae were recommended.*

Keywords: Data interpolation, data manipulation, *cumulation-differencing method.*

1. INTRODUCTION

Since secondary data do not provide all the type of information scientists seek, data manipulation is, hence, required for the purpose of estimating many demographic parameters. Interpolation is one such technique of manipulation through which intermediary points are obtained from a data set which is sometimes not suitable according to the specific need of demographers. Bivariate or univariate data may be of interval type of equal or unequal width or even at single point. As for example, age distribution of country or state or region may be of single year or grouped year of age. The grouped data may be of equal or unequal width. Similarly there may be distribution like, such as women ever married by age, and marriage duration. From these type of distribution already available from secondary sources, demographic estimation may be required to be done but at some intermediary points which need interpolation. A very simple example of interpolation is to construct the single year distribution from the grouped distribution of age usually available from secondary sources. There are some standard multipliers by which single year age distribution is obtained from 5-year grouped data on age (multipliers like Karup-King and Sprague). In this case the width of the interval must be same and the values are of absolute type.

In case of unequal interval or when the grouped data are unevenly spaced, as for example, in subdividing data for combinations of 1-, 5-, and 10-year age groups the single ages may be estimated by use of graphs for rough estimation. The individual ages are read off from the graph by drawing the ordinates along the curve that are at desired abscissas. In another method, from cumulated data, the values along the curve at two abscissa points are read off and taking the differences between the ordinate values may give the interpolated value for the desired abscissa value.

One general procedure being described in the UN volume (Bogue, Arriaga, and Anderton, 1993) is the 'mid-point' method. In this method, from grouped (either equal or unequal) data, mid-point values are first found out by taking average of lower and upper limits of two extreme points of the interval. Then for a desired point the mid-point is considered to be half way past the point of consideration and the interpolation for that mid-point is done to achieve the ordinate value by use of Aitken's iterative procedure (Aitken, 1932). Aitken's procedure is nothing but a polynomial interpolation of any desired degree through a system of successive linear interpolations using Waring-Lagrange two-point formula. The method will be discussed briefly afterwards. The mid-point method is applicable for both the absolute and percentage figures corresponding to the grouped or single point values.

Another general procedure being discussed in the said volume is the cumulation-differencing method which has a sounder theoretical basis than the mid-point approach described above. The reason is that the group averages seldom apply exactly to the mid-points of groups, whereas cumulation-differencing approach signifies the precise points to which the observed values actually apply.

In this method, the grouped figures are added to give the values up to the upper limit of the particular group where the addition stops. The cumulated data represent the "ogive" transformation of the original data for specific points along the abscissa. Interpolation of the transformed data can now be done by the above method (Aitken's method) twice. Once for the upper limit for the sub-group for which interpolation is to be obtained and the other for the lower limit of the same sub-group. The difference between these two interpolated values now will give the exact value for the lower limit of the sub-group, i.e., the exact point at which the value was initially required to be estimated. The method is also applicable for both the absolute and percentage figures corresponding to the grouped or single point values. Before we discuss any kind of data that are applied to the methods including mid-point and cumulation-differencing this paper will give a brief account of the Aitken's iterative procedure which will be used off and on in the present paper at a latter stage.

The Waring (1779) formula for interpolating between any number of points by a polynomial of the desired degree can be expressed as :

$$f(x) = f(a)\frac{(x-b)(x-c)...}{(a-b)(a-c)...} + \mathrm{f(b)}\frac{(\mathrm{x}-\mathrm{a})(\mathrm{x}-\mathrm{x})...}{(b-a)(b-c)...} + \mathrm{f(c)}\frac{(\mathrm{x}-\mathrm{a})(\mathrm{x}-\mathrm{b})...}{(c-a)(c-b)...} + ...$$

where f(a), f(b), f(c) etc., are the known ordinate values and f(x) is the unknown value which is to be estimated and x, a, b, c... are the different points. This is equivalent to the polynomial $y = A+Bx+Cx^2 + \dots\dots\dots$ passing through f(a), f(b), f(c),... to derive f(x). In Aitken's procedure only two-point Waring(1779) formula is used, that is,

which is equivalent to express in other way as,

$$f(x) = \frac{\mathrm{f(a)(b-x) - f(b)(a-x)}}{\mathrm{(b-x)-(a-x)}}$$

In this procedure the above formula is expressed in a tabular format to facilitate calculation of f(x) values for a suitable degree of polynomial through successive iterative procedure, such as:

Given ordinates	Computational stages			Proportionate parts
	(1)	(2)	(3)	
f(a)				(a-x)
f(b)	F(x; a,b)			(b-x)
f(c)	F(x; a,c)	f(x; a,b,c)		(c-x)
f(d)	F(x; a,d)	f(x; a,b,d)	f(x; a,b,c,d)	(d-x)

The first two lines are used for 2-point or linear interpolation and there is just one computational stage. The first three lines and two computational stages are required for 3-point interpolation i.e., a polynomial of 2nd degree being considered. The further lines and computational stages are required for higher degree polynomials. The entries in the computational stage (1) are :

$$f(x;a,b)=\frac{f(a)(b-x)-f(b)(a-x)}{(b-x)-(a-x)}$$

$$f(x;a,c)=\frac{f(a)(c-x)-f(c)(a-x)}{(c-x)-(a-x)}$$

$$f(x;a,d)=\frac{f(a)(d-x)-f(d)(a-x)}{(d-x)-(a-x)}$$

The entries in the computational stage (2) are:

$$f(x;a,b,c)=\frac{f(x;a,b)(c-x)-f(x;a,c)(b-x)}{(c-x)-(b-x)}$$

$$f(x;a,b,d)=\frac{f(x;a,b)(d-x)-f(x;a,d)(b-x)}{(d-x)-(b-x)}$$

and finally the entry for the computational stage (3) is:

$$f(x;a,b,c,d)=\frac{f(x;a,b,c)(d-x)-f(x;a,b,d)(c-x)}{(d-x)-(c-x)}$$

We will now discuss a few more types of data apart from those used by UN applicable in either mid-point or cumulation-differencing method using the Aitken's iterative procedure as elaborated above. Final conclusion should not be made on the basis of one particular type of data. This is what have been done in the UN volume.

2. MID-POINT METHOD USING DATA ON PERCENTAGES

Supposing marital status distributions of females by age-group are given. One can adopt only a few groups and from these truncated data, estimation of proportion of population for a particular single age could be done using polynomial interpolation with a suitable degree. Now the said volume contained one particular type of data like proportion of women ever married by age groups 14-17, 18-19, 20-24 and 25-29 and wanted an estimate of the proportion of ever married for women 20 years of age. There may be other kind of percentage distribution which has not been considered in the volume. For this, the mid-points of the age groups were first of all determined. As for example, to find the mid-point of 14-17, the lower limit of age group 14-17 was the 14^{th} birthday (exact age 14.0) and the upper limit of that age group was the 18^{th} birthday (exact age 18.0), the mid-point of age group 14-17 was therefore (14.0 + 18.0)/2, or 16.0. In a similar manner the mid-points of the other age groups were respectively 19.0 for 18-19, 22.5 for 20-24 and 27.5 for age group 25-29. The mid-point of the desired year of age 20 was 20.5. Considering the given percentages of ever married group corresponding to the above mid-points of the age-groups, the percentage of ever-married corresponding to the mid-point of 20 was interpolated. And the result could be used to signify the percentage ever-married for single age 20 as desired. Aitken's iterative procedure was applied to marital status data for the United States in 1960 as given in the said volume.

Table 1
MID-POINT METHOD ON PERCENTAGE OF WOMEN EVER-MARRIED BY AGE, US, 1960

Age group (years)	Mid-point age	Percent of ever married	Computations (1)	(2)	(3)	Proportionate parts[a]
14-17	16.0	5.4	---	---	---	- 4.5
18-19	19.0	32.4	45.9	---	---	- 1.5
20-24	22.5	71.6	51.2	48.2	---	2.0
25-29	27.5	89.5	38.3	44.6	49.6	7.0

Note: [a] difference between desired mid-point age 20.5 and the mid-point for each age-group.

The last figure in column (3) of the computations was the desired result: 49.6 percent. But as suggested in the volume that from the published data the figure for the same item was 54.0. Hence the estimated figure was lower as commented in the volume. The volume as such came to conclusion and did not recommend this method (mid-point method) to find the single point data from grouped percentage figures. However, this method could be useful and found applicable in case of marginal distribution as adopted in the present case. For this, an arbitrary graduated age distribution for females for Greater Bombay in 1961 has been considered. Before applying the Aitken's procedure, instead of simple proportions, average proportions have been found out and used for further calculation and presented in Table 2.

Table 2
MID-POINT METHOD ON PERCENTAGE DISTRIBUTION OF FEMALES, GREATER BOMBAY, 1961

Age-group (years)	Mid-point age	Percent of population	Average percentage	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	16.0	7.94	1.985	---	---	---	- 4.5
18-19	19.0	3.60	1.800	1.7075	---	---	- 1.5
20-24	22.5	8.57	1.714	1.7974	1.7460	---	2.0
25-29	27.5	7.99	1.598	1.8336	1.7298	1.7525	7.0

Source: The raw data have been obtained from elsewhere (Mukhopadhyay, et al. 1999).

The final estimate for the percentage of female population for age 20 years was 1.7525 percent which is closely matching with the actual value of 1.76 percent as calculated from single year true data. However, this technique was not at all suitable for specific distribution, like percentage of ever-married women as used by the UN although the average proportions were used instead of simple proportions. This is clear from the following tabular calculation from Aitken's approach.

Table 3
MID-POINT METHOD ON PERCENTAGE OF EVER-MARRIED WOMEN
BY AGE US, 1960

Age-group (years)	Mid-point age	Percent of ever married	Average percentage	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	16.0	5.4	1.35	---	—	---	- 4.5
18-19	19.0	32.4	16.20	23.625	—	—	- 1.5
20-24	22.5	71.6	14.32	10.329	17.927	---	2.0
25-29	27.5	89.5	17.90	7.826	20.837	16.763	7.0

The final estimate, 16.763 percent again was further too low as compared to the published figure of 54.0 percent. Moreover, this technique was not suitable for marginal percentage distribution directly. It is again clear from Table 4.

Table 4
MID-POINT METHOD ON PERCENTAGE DISTRIBUTION OF FEMALES,
GREATER BOMBAY, 1961

Age-group (years)	Mid-point age	Percent of population	Computations			Proportionate parts
			(1)	(2)	(3)	
14-17	16.0	7.94	—	—	—	- 4.5
18-19	19.0	3.60	1.43	---	---	- 1.5
20-24	22.5	8.57	8.38	4.41	—	2.0
25-29	27.5	7.99	7.96	2.58	5.14	7.0

The final estimate of percentage of female population age 20 was found to be 5.14 percent which was too high as compared to the actual value of 1.76 percent.

3. MID-POINT METHOD USING DATA ON ABSOLUTE NUMBERS

Dealing with absolute numbers there is no such clear demarcation between the specific and marginal distributions. Hence only one method is suffice, that is given in the volume. Here, our data shows more accurate result.

Table 5
MID-POINT METHOD ON NUMBER OF FEMALES, GREATER BOMBAY, 1961

Age-group (years)	Mid-point age	Female population	Average female population	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	16.0	135,337	33,834	---	---	---	- 4.5
18-19	19.0	61,447	30,724	29,169	---	---	- 1.5
20-24	22.5	146,093	29,219	30,639	29,799	---	2.0
25-29	27.5	136,156	27,231	31,250	29,536	29,904	7.0

The estimated number of females for age 20 was found to be 29,904 which was slightly lower than the actual value of 29,965 (see: Mukhopadhyay et al, 1999). An error of the extent of 0.2 percent was quite less. Similar method has been adopted in the volume concerning the number of women in different age group and found the number of women population for age 20 with an extent of error of the order of 3 percent. The table given below shows the method adopted in the said volume.

Table 6
MID-POINT METHOD ON NUMBER OF WOMEN (IN THOUSAND), US CENSUS, 1960

Age-group (years)	Mid-point age	Number of women in group	Average population per single age	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	16.0	5,516	1,379	---	---	---	- 4.5
18-19	19.0	2,417	1,209	1,124	---	---	- 1.5
20-24	22.5	5,520	1,104	1,189	1,152	—	2.0
25-29	27.5	5,537	1,107	1,273	1,150	1,153	7.0

The actual figure for age 20 was 1,124 (in thousands) as the number of women age 20. The error estimation, is, therefore, 1,153-1,124 = 29 or 3 percent which is obviously higher than what has been obtained in the present paper.

4. MID-POINT METHOD ON SEX-RATIOS BY AGE-GROUP, GREATER BOMBAY, 1961

If sex-ratios for different age groups are given, sex-ratio for a particular single age may be estimated by using mid-point method exactly in a similar way as done earlier. However, the UN volume did not expose these kind of data about which the present paper tries to throw light on this new kind of data and find the suitable method of interpolation. In initial stage, mid-point approach is discussed in the following table for age-group/sex-ratio data.

Table 7
MID-POINT METHOD ON SEX-RATIOS FOR AGE-GROUPS, GREATER BOMBAY, 1961

Age-group (years)	Male population	Female population	Sex-ratios	Mid-point age	Computations (1)	Computations (2)	Computations (3)	Proportionate parts
14-17	20,1891	135,337	149	16.0	---	---	---	- 4.5
18-19	88,789	61,447	144	19.0	141.5	—	---	- 1.5
20-24	199,869	146,093	137	22.5	140.7	141.2	---	2.0
25-29	180,949	136,156	133	27.5	142.7	141.7	141.0	7.0

Note: The sex-ratios by age-group have been calculated from the data of the same source, as mentioned earlier in Table 2. The sex-ratio is defined as (M/F)×100.

The final estimate of sex-ratio for age 20 was estimated to be 141.0. The actual value was found to be 140 for exact age 20 from sex-wise single[1] year age data. An error of 0.7 percent may be ignored while adopting the estimate of sex-ratio value of 141. In the following paragraphs cumulation-differencing method would be discussed and finally from these two methods, the appropriate estimates might be made.

5. CUMULATION-DIFFERENCING METHOD USING DATA ON PERCENTAGES

To start with, the technique adopted by the UN was that the percentages required to be weighted by the size of the class intervals associated with them. As the age groups were started from 14-17, the number of girls under 14 years were assumed to be *k* which would be dropped out as the work progressed. As such they have considered *k* to be zero. Again the upper limit of the age range under 14 was taken to be 14.0. The number of girls aged 14 to 17 was exactly the figure for the number of girls under 18 years old. The number of girls under 20 years old were the addition of the number of girls 14-17 years plus the figure for 18 to 19 years old and so on. In a similar fashion the upper limit of the age groups and the cumulated number of girls were found out for the present technique. The process was followed similarly for percentage figures. For getting the percentage figure for age 20, the cumulated figures for ages up to 21 and 20 were required. But figure for age up to 20 was already available. The difference between the two values would give the value corresponding to age 20. Table 8 shows the method as proposed in the UN volume.

The figure of 139.4 in column (7) was the interpolated estimate of percentages cumulated to age 21.0 The cumulated figure for age up to 20 was available from the table, that was 86.4 in column (4). The desired figure for percentage ever-married for women age 20 was the difference between 139.4 and 86.4, i.e., 53.0 which was much closer to the published figure of 54.0 than the figure estimated by mid-point approach mentioned earlier in the paper. However, this similar approach is not suitable for marginal distribution as adopted in the present paper on female age distribution, Greater Bombay, 1961. This is clear from the Table 9.

The figure in column (7) indicated the percentages up to age 21, i.e., 45.10 (see Table 9). The figure up to age 20 was available in column (4), i.e., 38.96. Hence the percent of female population age 20 was the difference between 45.10 and 38.96, i.e., 6.14 percent which is much higher as compared to 1.76 which is the actual value calculated from the graduated single year age data.

But this cumulation-differencing method could be made appropriate if the group percentage figures are not multiplied with the respective group interval figures. This is done in the Table 10.

Table 8
CUMULATION-DIFFERENCING METHOD ON PERCENTAGES OF EVER-MARRIED WOMEN, US, 1960

Age-group (years)	Upper limit of the age group	Number of single years in age group	Percent ever married	Estimated sums of percentage for single ages		Computations			Propor-tionate parts
		(1)	(2)	Within a group	Cumulated	(5)	(6)	(7)	
				(3) = (1) x (2)	(4) = Σ(3)				
14-17	18.0	4	5.4	21.6	21.6	---	---	---	- 3.0
18-19	20.0	2	32.4	64.8	86.4	118.8	---	---	- 1.0
20-24	25.0	5	71.6	358.0	444.4	202.8	135.6	---	4.0
25-29	30.0	5	89.5	447.5	891.9	239.2	130.8	139.4	9.0

Table 9
CUMULATION-DIFFERENCING METHOD ON PERCENTAGES OF FEMALES BY AGE-GROUP, GREATER BOMBAY, 1961

Age-group (years)	Upper limit of the age group	Number of single years in age group	Percent of female population	Estimated sums of percentage for single ages		Computations			Proportionate parts
		(1)	(2)	Within a group (3) = (1) x (2)	Cumulated (4) = Σ(3)	(5)	(6)	(7)	
14-17	18.0	4	7.94	31.76	31.76	---	---	---	- 3.0
18-19	20.0	2	3.60	7.20	38.96	42.56	---	---	- 1.0
20-24	25.0	5	8.57	42.85	81.81	52.21	44.49	---	4.0
25-29	30.0	5	7.99	39.95	121.76	54.26	43.73	45.10	9.0

Table 10
CUMULATION-DIFFERENCING METHOD WITHOUT WEIGHTING ON PERCENTAGE OF FEMALES BY AGE-GROUP, GREATER BOMBAY, 1961

Age-group (years)	Upper limit of the age-group	Percent of female population	Cumulated percent of female population	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	18.0	7.94	7.94	---	---	---	- 3.0
18-19	20.0	3.60	11.54	13.34	---	---	- 1.0
20-24	25.0	8.57	20.11	13.16	13.30	---	4.0
25-29	30.0	7.99	28.10	12.98	13.30	13.30	9.0

The figure in column (3) indicated the percentages up to age 21, i.e., 13.30 (see Table 10). The figure up to age 20 was 11.54 as quoted in the column headed with 'cumulated percentage of female population' in the above table. Hence percentage of female population age 20 was the difference was between 13.30 and 11.54, i.e., 1.76 percent which was exactly tallying with the true value of 1.76. However, the same approach if applied on the UN data on ever married women, inconsistent result is obtained. It gives lower estimates. The Table 11 the inconsistent result.

Table 11
CUMULATION-DIFFERENCING METHOD WITHOUT WEIGHTING ON PERCENT OF EVER MARRIED WOMEN, US, 1960

Age-group (years)	Upper limit of the age-group	Percent ever-married women	Cumulated percent ever-married women	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	18.0	5.4	5.4	---	---	---	- 3.0
18-19	20.0	32.4	37.8	54.0	---	---	- 1.0
20-24	25.0	71.6	109.40	49.97	53.19	---	4.0
25-29	30.0	89.5	198.90	53.78	53.98	52.56	9.0

The figure in column (3) indicated the percentage up to age 21 years, i.e., 52.56 (see Table 11). The figure up to age 20 was available in column headed with "cumulated percent ever married", i.e., 37.8. Hence the percent of ever-married female for age 20 years, the difference was between 52.56 and 37.80, i.e., 14.76. This is not at all consistent with the actual figure of 54.0 percent. Hence from the above calculations, it is evident that this approach is quite legitimate while marginal distribution is dealt with and not reasonable for specific distribution of percentage figures. Further while weighting the percentages with the figures for class intervals, the cumulation-differencing result was quite consistent with the specific distribution of percentage figures but inconsistent for marginal distribution.

6. CUMULATION-DIFFERENCING METHOD ON ABSOLUTE NUMBERS

The cumulation-differencing method using data on absolute numbers shows us significant difference between specific distribution mainly discussed in the UN volume and marginal one emphasized in the present paper. The similar statement was made earlier in this paper on mid-point method of interpolation on absolute numbers. The following Table 12 is given for the UN approach.

Table 12
CUMULATION-DIFFERENCING METHOD ON NUMBER OF WOMEN, US, 1960

Age-group (years)	Upper limit of the age-group	Number of women		Aitken's procedure			Proportionate parts
				Computations for age 21.0 (upper limit)			
		In age group	Cumulated from youngest group	(1)	(2)	(3)	
14-17	18.0	5,516	5,516	---	---	---	- 3.0
18-19	20.0	2,417	7,933	9,142	---	---	- 1.0
20-24	25.0	5,520	13,453	8.918	9,097	---	4.0
25-29	30.0	5,537	18,990	8,885	9,116	9,082	9.0

The figure in column (3), i.e., 9,082 is the interpolated estimate of the number of women cumulated to age 21.0. The number for age 20 is 7,933 as given in the table. Now the difference of 1,149 between 9,082 and 7,933 is the estimate for age 20.0. The actual figure for age 20 was 1,124 (see Bogue, Arriaga, and Anderton, 1993) as mentioned earlier in the Table 6. An error of the order of about 2 per cent has been committed in this method. The same procedure applied on the data on age distribution of females is generated in the Table 13.

Table 13
CUMULATION-DIFFERENCING METHOD ON NUMBER OF FEMALES, BOMBAY, 1961

Age-group (years)	Upper limit of the age-group	Female population	Cumulated female population	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	18.0	135,337	135,337	---	---	---	- 3.0
18-19	20.0	61,447	196,784	227,508	---	---	- 1.0
20-24	25.0	146,093	342,877	224,283	226,863	---	4.0
25-29	30.0	136,156	479,033	221,261	226,833	226,847	9.0

The figure in column (3), 226,847 is the figure for age up to 21 years (see Table 13). The value for age up to 20 is 196784 as given in the table. Therefore, the estimated figure for age 20.0 is the difference of 30,063 between 22,6847 and 196,784. The actual value for age 20 as obtained from the single year graduated age distribution is 29,965. Hence an error of the extent of 0.33 percent is committed in this particular type of data. And the error is much less as compared to earlier method on UN data.

Table 14
CUMULATION-DIFFERENCING METHOD USING DATA ON SEX-RATIOS BY AGE-GROUP, GREATER BOMBAY, 1961

Age-group (years)	Upper limit of the age-group	Sex-ratios (M/F) x 100	Number of cumulated sex-ratios	Computations			Proportionate parts
				(1)	(2)	(3)	
14-17	18.0	149	149	---	---	---	- 3.0
18-19	20.0	144	293	365.0	---	---	- 1.0
20-24	25.0	137	430	269.4	345.9	---	4.0
25-29	30.0	133	563	252.5	353.7	339.7	9.0

The figure in the 3rd stage of computaion, i.e., 339.7 is for sex-ratio-up to age 21 (see Table 14). Now the figure for the similar value up to age 20 is 293 is also indicated in the above table. Hence the figure for the exact age, 20 is the difference between these two values, i.e., 339.7 and 293 giving 46.7. The actual value for age 20 is 140. Hence this method does not give the desired result. However, the method is quite good if the sex-ratios are multiplied with the respective widths of the age-groups. This is illustrated in the following Table 15.

Table 15

CUMULATION-DIFFERENCING METHOD ON WEIGHTED (AGE-GROUP) SEX-RATIOS, GREATER BOMBAY, 1961

Age-group (years)	Upper limit of the age group	Number of single years in age group	Sex-ratios	Estimated sex-ratios for single ages	Cumulated sum	Computations			Proportionate parts
		(1)	(2)	(3) = (1) x (2)	(4) = Σ(3)	(5)	(6)	(7)	
14-17	18.0	4	149	596	596	—	—	—	- 3.0
18-19	20.0	2	144	288	884	1,028	—	—	- 1.0
20-24	25.0	5	137	685	1,569	1,013	1,025	—	4.0
25-29	30.0	5	133	665	2,234	1,005.5	1,025.75	1,024.4	9.0

The final figure of 1,024.4 in column 7 is the figure of sex-ratio for the age up to 21.0 (see Table 15). The figure up to age 20 is 884 as given in column (4) of the above table. Hence the difference of 1,024.4 and 884, i.e., 140.4 is the sex-ratio value for age 20. But the actual value of sex-ratio at age 20 was earlier calculated to be 140. Hence an extent of error was found to be of the order of 0.3 percent which is still less than what was observed for mid-point method.

7. DISCUSSION AND CONCLUSION

From this empirical study one may justify which interpolation method is suitable and which is not, in estimating the data at single points from a few grouped figures. The two interpolation methods, mainly mid-point method and cumulation-differencing method have been extensively used in UN volume but with a limited data and recommended accordingly. However, in this paper the scope has been further extended by utilizing some more types of data and enhances the methods more realistically comparing the two methods applied in different types of data.

While mid-point method has not been found effective (in UN volume) in estimating single point percentage figures from a specific distribution like ever married women by a few unequal grouped age distribution the same was quite satisfactory for female distribution by age-group as marginal distribution (percentage) as applied in the present paper with an extent of error of the order of 0.4 percent. In the marginal case, simple percentage figures which were used in specific case (UN) were not found suitable. Otherwise average percentage figures were suitable in marginal case. And this was not suitable for the specific case in UN volume.

The mid-point method though found suitable for both the types of data on absolute numbers, that is, the specific distribution of ever married women in the UN volume and marginal female age distribution used in the present paper, the method gives less error i.e., 0.2 percent in respect of marginal distribution as compared to specific case with larger extent of error, i.e., 3 percent. Hence mid-point method is more effective in marginal case.

The mid-point method may also be effective in interpolating sex-ratio values for single age from sex-ratio figures from an age-group data as applied in the present paper. The value estimated for sex-ratio figure at age 20 from some truncated data was 141. But the actual figure got 140 implying an extent of error of the order of 0.7 percent which was quite acceptable.

In the next phase of calculation, cumulation-differencing method has been applied in both types of distribution, specific, marginal and sex-ratio distributions. In specific distribution of ever-married women by age-group, cumulation-differencing method gave interpolated value close to actual value with an error of 1.9 percent. This method is more effective in respect of marginal case without weighting the percentage figures by class interval values as done in case of specific case.The error in this case turns to zero. The same distribution gave 0.4 percent as error in mid-point method.

While dealing with absolute numbers cumulation-differencing method gave estimate from specific distribution of ever married women with more consistent value with less error of the order of 2.2 percent as compared to mid-point method with an error of 3 percent. However, mid-point method may be recommended for marginal distribution with absolute numbers since this method gave an error of the order of 0.2 percent as compared to cumulation-differencing method with an error rate of 0.33 percent. Cumulation-differencing method was found more effective in respect of data like sex-ratios by age-group since the error in this case reduced to 0.28 percent from 0.7 percent obtained in respect of mid-point method. The only change which was done in cumulation-differencing method was to multiply the sex-ratio values with their respective class intervals. From the above analysis, proper interpolation method should be used for specific data.

It is obvious that researchers may raise points in regard to use of old data in the paper. But it must be kept in mind that this paper has been originated from some other reference (Bogue, Arriaga and Anderton, 1993) where US data of 1960 census was variously used. As a matter fact, the present authors used data of 1961 census of Bombay readily available at the time of preparing the paper which might tally with the time period of the reference data of 1960. However, for the satisfaction, some more recent data have been tried here although not in a detailed analysis for rather in a concise form. Moreover, demographic year-book data of the US census have been used for the demonstration of the method per se. Unfortunately, recent data in 2003 although available but in the present case are not possible for calculation in the present purpose. Hence, 1990 census data of US have been used in the following paragraph applying the mid-point and cumulation-differencing methods.

Table 16
MID-POINT METHOD ON PERCENTAGE DISTRIBUTION OF FEMALES, UNITED STATES CENSUS, 1990

Age-group (years)	Mid-point age	Percentage of population	Average percentage	Computations (1)	(2)	(3)	Proportionate parts
14-17	16.0	5.05	1.2625	---	---	---	- 3.5
18-19	19.0	2.96	1.4800	1.5163	---	---	- 0.5
20-24	22.5	7.33	1.4660	1.3721	1.4957	---	3.0
25-29	27.5	8.33	1.6660	1.3853	1.5086	1.4880	8.0

The figure of 1.488 percent in column (3) of Table 16 is the percentage of population estimated at age 19. But the actual census figure for the corresponding age is 1.5675 which is close to the estimated figure of 1.488, i.e., an extent of error of the order of 5 percent.

Table 17
MID-POINT METHOD ON SEX-RATIO, UNITED STATES CENSUS, 1990

Age-group (years)	Male population	Female population	Sex-ratios	Mid-point age	Computations (1)	(2)	(3)	Proportionate parts
14-17	682,4450	645,9248	1.056	16.0	---	---	---	-3.5
18-19	394,0523	377,6931	1.043	19.0	1.041	---	---	- 0.5
20-24	967,5596	934,4716	1.035	22.5	1.045	1.042	---	3.0
25-29	1,069,5936	1,061,7106	1.007	27.5	1.042	1.041	1.042	8.0

Actual sex-ratio at age 19 is 1.040 which is found to be near to the estimated value, 1.042. The error is almost zero percent (Table 17).

Now another calculation is finally done on cumulation-differencing method based on US' 1990 census data on female population of a few ages

Table 18
CUMULATION-DIFFERENCING METHOD ON NUMBER OF FEMALES, UNITED STATES, 1990

Age-group (years)	Upper limit of the age-group	Percentage of female population	Cumulated percent of female population	Computations (1)	(2)	(3)	Proportionate parts
14-17	18.0	135,337	5.05	---	---	---	-1
18-19	20.0	614,47	8.01	6.53	---	---	1
20-24	25.0	146,093	15.34	6.52	6.532	---	6
25-29	30.0	136,156	23.67	6.60	6.523	6.54	11

Percent of female population at age 19 would be the difference between the same up to age 19 and up to age 18, that is 6.54 minus 5.05 and the result is 1.49 percent. The error in this case is 4.94 percent.

Here some observation is made as to why these types of interpolation methods are necessary. As far as census data are concerned developing countries do not have very good census operation system so that their data are not up to the mark.

During the last decade, Macro International Company of US started some large scale sample survey conducted in the countries on the longitudinal basis (DHS) from where so many estimates are available with standard error estimate. But census and registration data are lacking quality due to which indirect estimations are prevalent. Moreover, in countries of Sub-sahara Africa, some parts of Asia and Latin America there are no country life tables. As a result many estimation procedures are conducted using model life tables (e.g., UN, 1982 and Coale and Demeny, 1966).

In the circumstances, as model tables are usually abridged type, so information on single ages in these developing countries are needed to be estimated otherwise. Hence the present methodologies would be useful even if some truncated grouped data are available. There are a number of applications of life tables in solving problems on demographic, economic and social science area. The present methodologies are useful and relevant.

NOTE

1. Sex-wise single means the age distribution by male and female separately, so that sex ratio at age 20 is calculated using formula (male/female) x 100.

REFERENCES

Aitken, A.C. 1932. "On Interpolation by Iteration of Proportionate Parts, without the use of Differences", *Proceedings of the Edinburg Mathematical Society*, series 2, 3: 56-76.

Bogue, D.J. Arriaga, E.E. Anderton, D.L. (eds). 1993. *Readings in Population Research Methodology*, Vol. 1, Basic Tools. Chicago, Illinois: Social Development Center and United Nations Population Fund.

Coale, A.J. and P. Demeny. 1966. *Regional Model Life Tables and Stable Populations.* Princeton, USA: Princeton University Press.

Mukhopadhyay. B.K., Majumdar, P.K. and Ghosh, N.C.. 1999. "Digit Preference Error in Age Data: An Implicit Model Approach", *Journal of Bihar Mathematical Society*, Vol. 19, p. 51-59.

United Nations. 1982. "*Model Life Tables for Developing Countries*", Department of International Economic and Social Affairs, Population Studies, 77, New York.

Waring, E. 1779. "Problems Concerning Interpolations", *Philosophical Transactions of the Royal Society of London*, vol. 69, 1779, pp. 59-67.

Barun Kumar Mukhopadhyay, Associate Scientist 'B' in the Population Studies Unit, Indian Statistical Institute, Kolkata, India,
E-mail: barun_mukhopadhyay@yahoo.com

Prasanta Kumar Majumdar, Professor and former Head, Population Studies Unit, Indian Statistical Institute, Kolkata, India.