# A single year age adjustment from a preliminary grouped data

**Barun Kumar Mukhopadhyay**

Population Studies Unit
Indian Statistical Institute
203, B.T. Road
Kolkata – 700 108

# A single year age adjustment from a preliminary grouped data

By

**Barun Kumar Mukhopadhyay**[*]

## Abstract

Adjustments to raw age data in censuses of developing countries are essential because of huge amount of errors. There are some methods especially in United Nations. And individual country has also her own methods. However, the present work is based on only the 5-year grouped data which are usually available in the websites just a few months after the census is over. The official adjusted data are available after long time gap. Hence as an alternative, the present paper tries to use those preliminary data to get an adjusted single year age data for the researchers. The paper follows a number of steps starting right from cumulation of the data once in a more than type and other in a less than type, so that the distribution approaches toward a smoothed series. But the final adjustment is done by fitting two 3rd degree polynomials taking cumulated male and female populations. The adjusted data are found consistent and follows the usual pattern of correct age distribution.

[*] Associate Scientist 'B' retired from Population Studies Unit, ISI, Kolkata, India, e-mail:barun_mukhopadhyay@yahoo.com

# Adjustment to single year data on age : A new approach

By

**Barun Kumar Mukhopadhyay**[*]

## Introduction

Adjustment to age distribution either it is single year or grouped one is very essential because of many type of errors found in censuses of many countries particularly in developing zones. The corrected age distribution is quite useful in many studies of population research, in particular, even in other branches of social sciences, physical sciences etc. Sometimes Market researchers require an age distribution for their product to be sold properly as this item affects their selling on the basis of ages of the consumer etc. An age distribution that is smooth and as close to correct as possible is still useful, particularly as a basis for population projection (UN, 1983).

Before adjustment to be done, one has to know about what kind of major errors creep in. Digit preference errors are the most common type of errors especially found in the middle age ranges. It is also common that very young populations are under enumerated and advanced aged persons try to exaggerate their ages. In this connection it may be pointed out that adjustment to 0-4 population was sometimes adjusted separately because of under enumeration type of error for which different adjustment procedures are adopted (UN methods in different manuals and Mukhopadhyay (1986) and others. Similarly for advanced aged persons. Now, the present paper tries to adjust age distribution particularly up to fifty nine years of age, but it may be done for some more higher ages such as up to sixty nine, even more as the availability of raw data. Digit preference error is the tendency of persons reporting their ages ending in some preferred digits. There are some other kinds of errors such as shifting errors, recall lapse error (Som, 1973) etc.

## Methodology and results

At the outset, it is important to say that usually after the census is over for some months a few preliminary tables are available in the website. As regard ages, only a five-year grouped data become available. As has already been pointed out earlier these data directly are not worth usable. Adjusted data are very essential for which the present paper gives some methodology which will require only the raw data and a single assumption. Rest of the methodologies follow some general procedures. This kind of work may help researchers to use the adjusted data for their related activities or they can themselves do the adjustments in a time bound research projects since raw and adjusted single year age data are available from the census authority after completion of number of years to put in the websites, CDs or census volumes.

[*] Associate Scientist 'B' retired from Population Studies Unit, ISI, Kolkata, India, e-mail:barun_mukhopadhyay@yahoo.com

Smoothing and adjustment procedures are usually applied to cumulated age distribution that is, to the number of persons or proportion of persons under given ages, that is, less than type since the process of cumulation removes the effects of errors that do not result in a net transfer of people across each of the age boundaries used (UN, 1983). In the present methods, 2001 census of Indian data are used for experimentation. Now the raw 5-year grouped data as the only input required by two sexes are given below:

**Table 1 :** 5-year grouped distribution of ages by two sexes Census, India, 2001

| **age** | **male** | **female** | Total |
|---|---|---|---|
| 0-4 | 57119612 | 53327552 | 110447164 |
| 5-9 | 66734833 | 61581957 | 128316790 |
| 10-14 | 65632877 | 59213981 | 124846858 |
| 15-19 | 53939991 | 46275899 | 100215890 |
| 20-24 | 46321150 | 43442982 | 89764132 |
| 25-29 | 41557546 | 41864847 | 83422393 |
| 30-34 | 37361916 | 36912128 | 74274044 |
| 35-39 | 36038727 | 34535358 | 70574085 |
| 40-44 | 29878715 | 25859582 | 55738297 |
| 45-49 | 24867886 | 22541090 | 47408976 |
| 50-54 | 19851608 | 16735951 | 36587559 |
| 55-59 | 13583022 | 14070325 | 27653347 |
| 60+ | 37768327 | 24265284 | 62033611 |
| ANS | 1500562 | 1237910 | 2738472 |
| Grand total | 532156772 | 496453556 | 1028610328 |

Sources: website of census of India, 2001

The above distribution clearly points out some significant inconsistencies. The population aged 0-4 is under enumerated for both the sexes. Apart from this, there are some fluctuation of the data, e.g., the population aged 55-59 are always lower than the adjacent higher age group of 60-64. Moreover, there are a wide ga4p between the figures of male and female populations in the advanced age group of 60 + years. These are some kind of irregularities in the data. The following gives the more than type cumulated distribution as a first step.

**Table 2 :** Cumulated census population by 5-year group, India, 2001

| **Age+** | **Male+** | **Female+** | **M(EXP(ln(Pt/Po))** | **F(EXP(ln(Pt/Po))** |
|---|---|---|---|---|
| 0+ | 530656210 | 495215646 | 0.977480404 | 0.977470341 |
| 5+ | 473536598 | 441888094 | 0.970076214 | 0.970429984 |
| 10+ | 406801765 | 380306137 | 0.965422282 | 0.966716814 |
| 15+ | 341168888 | 321092156 | 0.966165982 | 0.969354308 |
| 20+ | 287228897 | 274816257 | 0.965438307 | 0.966171349 |
| 25+ | 240907747 | 231373275 | 0.962837912 | 0.960866533 |
| 30+ | 199350201 | 189508428 | 0.959341840 | 0.957597701 |
| 35+ | 161988285 | 152596300 | 0.950916992 | 0.949975595 |
| 40+ | 125949558 | 118060942 | 0.947281358 | 0.951757300 |
| 45+ | 96070843 | 92201360 | 0.941849010 | 0.945473945 |
| 50+ | 71202957 | 69660270 | 0.936722097 | 0.946529219 |
| 55+ | 51351349 | 52924319 | 0.940405538 | 0.940061092 |
| 60+ | 37768327 | 38853994 | 0. 914687664 | 0.915029063 |

After getting the different cumulated ages such as 0+, 5+, 10+ ……. 60+, the next step is to find the growth rates among these different categories of cumulated figures assuming an exponential in each consecutive ages:

$$P_t = P_0 e^{-rt}$$

The negative growth is obvious as the population will decline due to mortality in a closed population ( another very common assumption ). Now the step 3 requires to generate another cumulated distribution on single year basis such as 0+, 1+, 2+, ……. 60+. The last two columns of the above table are directly prepared using the already constructed different growth rates ( negative ) which are skipped and exponentials are accordingly done with one year time period in each ages like 0+, 5+, and so on up to 60+. The following table gives the cumulated (more than type) figures 0+, 1+, 2+, ……. 60+ for male and female populations separately.

**Table 3 :** Cumulated figures in each individual ages, 0+, 1+, 2+, ……. 60+, India, 2001

| Single Age+ | MALE | FEMALE | Single Age+ | MALE | FEMALE | Single Age+ | MALE | FEMALE |
|---|---|---|---|---|---|---|---|---|
| 0+ | 530656210 | 495215646 | 20+ | 287228897 | 265519594 | 40+ | 125949558 | 112365363 |
| 1+ | 518706047 | 484058606 | 21+ | 277301780 | 256537424 | 41+ | 119309668 | 106944555 |
| 2+ | 507024996 | 473152931 | 22+ | 267717761 | 247859109 | 42+ | 113019825 | 101785261 |
| 3+ | 495606998 | 462492957 | 23+ | 258464982 | 239474370 | 43+ | 107061573 | 96874865 |
| 4+ | 484446129 | 452073148 | 24+ | 249531995 | 231373275 | 44+ | 101417432 | 92201360 |
| 5+ | 473536598 | 441888094 | 25+ | 240907747 | 223546229 | 45+ | 96070843 | 87173984 |
| 6+ | 459366590 | 428821456 | 26+ | 231955112 | 222318837 | 46+ | 90484228 | 82420730 |
| 7+ | 445620602 | 416141199 | 27+ | 223335176 | 213618730 | 47+ | 85222481 | 77926653 |
| 8+ | 432285947 | 403835897 | 28+ | 215035574 | 205259088 | 48+ | 80266709 | 73677620 |
| 9+ | 419350314 | 391894463 | 29+ | 207044403 | 197226588 | 49+ | 75599121 | 69660270 |
| 10+ | 406801765 | 380306137 | 30+ | 199350201 | 189508428 | 50+ | 71202957 | 65935481 |
| 11+ | 392735488 | 367648337 | 31+ | 191244989 | 181472835 | 51+ | 66697383 | 62409859 |
| 12+ | 379155592 | 355411829 | 32+ | 183469319 | 173777970 | 52+ | 62476913 | 59072755 |
| 13+ | 366327234 | 343582592 | 33+ | 176009794 | 166409384 | 53+ | 58523505 | 55914089 |
| 14+ | 353666345 | 332147068 | 34+ | 168853560 | 159353244 | 54+ | 54820260 | 52924319 |
| 15+ | 340523365 | 311252065 | 35+ | 161988285 | 152596300 | 55+ | 51351349 | 49752093 |
| 16+ | 329625774 | 301713530 | 36+ | 154037413 | 144962761 | 56+ | 48291093 | 46770007 |
| 17+ | 318473209 | 292467310 | 37+ | 146476793 | 137711085 | 57+ | 45413211 | 43966664 |
| 18+ | 307697981 | 283504447 | 38+ | 139287272 | 130822170 | 58+ | 42706835 | 41331350 |
| 19+ | 297287322 | 274816257 | 39+ | 132450633 | 124277868 | 59+ | 40161745 | 38853994 |

Here it must be mentioned that population for ages 0+ is obviously the total population. The age not stated ( ANS ) figures of 1500562 for male and 1237910 for female have been subtracted from the corresponding total figures for male and female although in some instances they are pro rated to the original pattern of age distribution.

The adjusted single year data then are obtained by subtracting one step lower single year cumulated figure from one step higher figure and accordingly the entire distribution is formed and given below :

**Table 4:** Estimated single year age distribution, Indian census, 2001

| AGE | MALE | FEMALE | AGE | MALE | FEMALE | AGE | MALE | FEMALE |
|---|---|---|---|---|---|---|---|---|
| 0 | 11950163 | 11157040 | 20 | 9927117 | 8982170 | 40 | 6639890 | 5420809 |
| 1 | 11681050 | 10905675 | 21 | 9584019 | 8678315 | 41 | 6289844 | 5159294 |
| 2 | 11417998 | 10659974 | 22 | 9252779 | 8384739 | 42 | 5958252 | 4910396 |
| 3 | 11160869 | 10419809 | 23 | 8932987 | 8101095 | 43 | 5644141 | 4673505 |
| 4 | 10909531 | 10185054 | 24 | 8624248 | 7827046 | 44 | 5346589 | 5027376 |
| 5 | 14170008 | 13066638 | 25 | 8952635 | 1227393 | 45 | 5586615 | 4753253 |
| 6 | 13745988 | 12680257 | 26 | 8619936 | 8700107 | 46 | 5261747 | 4494077 |
| 7 | 13334656 | 12305302 | 27 | 8299601 | 8359642 | 47 | 4955772 | 4249033 |
| 8 | 12935632 | 11941434 | 28 | 7991171 | 8032500 | 48 | 4667589 | 4017350 |
| 9 | 12548549 | 11588326 | 29 | 7694202 | 7718160 | 49 | 4396164 | 3724789 |
| 10 | 14066277 | 12657800 | 30 | 8105212 | 8035593 | 50 | 4505574 | 3525622 |
| 11 | 13579897 | 12236508 | 31 | 7775669 | 7694865 | 51 | 4220471 | 3337104 |
| 12 | 12828357 | 11829238 | 32 | 7459525 | 7368585 | 52 | 3953408 | 3158666 |
| 13 | 12660889 | 11435523 | 33 | 7156234 | 7056140 | 53 | 3703245 | 2989770 |
| 14 | 13142980 | 20895004 | 34 | 6865275 | 6756944 | 54 | 3468911 | 3172226 |
| 15 | 10897592 | 9538535 | 35 | 7950872 | 7633539 | 55 | 3060256 | 2982086 |
| 16 | 11152564 | 9246220 | 36 | 7560620 | 7251676 | 56 | 2877882 | 2803343 |
| 17 | 10775228 | 8962863 | 37 | 7189522 | 6888915 | 57 | 2706376 | 2635314 |
| 18 | 10410659 | 8688190 | 38 | 6836638 | 6544301 | 58 | 2545091 | 2477356 |
| 19 | 10058425 | 9296663 | 39 | 6501075 | 11912505 | 59 | 2393418 | 3301460 |

After adjustment, the distribution of single year ages in Table 4 still show some irregularities particularly in some round digits like, 0s and 5s where higher values indicating preferences etc. In order to smooth the irregular series, cumulated (less than type) distributions have been prepared in step 4 in the following table. In Table 5, ten year groups like up to 9, up to 19, so on lastly up to 59 have been considered taking into account all the ten digits as Zelnik (1961) used the same technique while doing ten year moving average method as a part of his method.

**Table 5:** Cumulated distribution, 2001

| Up to Age | Male | Female |
|---|---|---|
| 9 | 123854445 | 91379749 |
| 19 | 243427313 | 211711199 |
| 29 | 331306009 | 287722365 |
| 39 | 404706652 | 364865429 |
| 49 | 459453253 | 411295312 |
| 59 | 492887883 | 441678259 |

In step 5, two polynomials with degree 3 separately for male and female populations are applied to these kind of data where it is well known that in demographic research they are much applicable. As such the following equation is considered

$$\mathbf{y = a + bx + cx^2 + dx^3}$$

where y is the dependent and x the independent variables and a is the intercept and d, c, and d are coefficients. They are all estimated using the data in Table 5.

The two polynomials have been fitted and found fit for these data.

**$y = -30549906.331+19233731.991x -266366.871x^2 +1279.473 x^3$ for male,**

**$y' =130587.227 + 13914258.373x -113590.864x^2 + 167.657x^3$ for female**

and the coefficient of determination measured by $R^2$ for both the cases have been found to be highly significant ($p<.001$). These fittings were done in computer using SPSS (17 Version) package. Using these fitted polynomials separately for male and female, the cumulated (less than type) distributions are obtained. The final adjusted single year population starting from 0 to 59 are obtained simply by subtracting one step lower figures from one step higher values repeatedly upto the last figure. Table 6 gives the final adjusted single year age distribution for

**TABLE 6:** Finally adjusted single year age data, one up to fifty nine, India, 2001

| AGE | MALE | FEMALE | PERSON | AGE | MALE | FEMALE | PERSON |
|---|---|---|---|---|---|---|---|
| 1 | 18968645 | 13800835 | 32769480 | 31 | 6556362 | 7453146 | 14009508 |
| 2 | 18443588 | 13574659 | 32018247 | 32 | 6261610 | 7257149 | 13518759 |
| 3 | 17926208 | 13349490 | 31275698 | 33 | 5974535 | 7062157 | 13036692 |
| 4 | 17416504 | 13125326 | 30541830 | 34 | 5695137 | 6868172 | 12563309 |
| 5 | 16914478 | 12902168 | 29816646 | 35 | 5423416 | 6675192 | 12098608 |
| 6 | 16420128 | 12680016 | 29100144 | 36 | 5159372 | 6483218 | 11642590 |
| 7 | 15933456 | 12458870 | 28392326 | 37 | 4903004 | 6292250 | 11195254 |
| 8 | 15454460 | 12238729 | 27693189 | 38 | 4654313 | 6102288 | 10756601 |
| 9 | 14983141 | 12019595 | 27002736 | 39 | 4413299 | 5913333 | 10326632 |
| 10 | 14519499 | 11801467 | 26320966 | 40 | 4179962 | 5725383 | 9905345 |
| 11 | 14063533 | 11584345 | 25647878 | 41 | 3954302 | 5538438 | 9492740 |
| 12 | 13615245 | 11368228 | 24983473 | 42 | 3736319 | 5352500 | 9088819 |
| 13 | 13174633 | 11153118 | 24327751 | 43 | 3526012 | 5167568 | 8693580 |
| 14 | 12741698 | 10939013 | 23680711 | 44 | 3323382 | 4983642 | 8307024 |
| 15 | 12316440 | 10725915 | 23042355 | 45 | 3128430 | 4800722 | 7929152 |
| 16 | 11898859 | 10513822 | 22412681 | 46 | 2941154 | 4618807 | 7559961 |
| 17 | 11488955 | 10302736 | 21791691 | 47 | 2761554 | 4437899 | 7199453 |
| 18 | 11086727 | 10092655 | 21179382 | 48 | 2589632 | 4257997 | 6847629 |
| 19 | 10692177 | 9883580 | 20575757 | 49 | 2425386 | 4079100 | 6504486 |
| 20 | 10305303 | 9675511 | 19980814 | 50 | 2268818 | 3901209 | 6170027 |
| 21 | 9926106 | 9468448 | 19394554 | 51 | 2119926 | 3724325 | 5844251 |
| 22 | 9554586 | 9262391 | 18816977 | 52 | 1978711 | 3548446 | 5527157 |
| 23 | 9190742 | 9057340 | 18248082 | 53 | 1845173 | 3373573 | 5218746 |
| 24 | 8834576 | 8853295 | 17687871 | 54 | 1719311 | 3199707 | 4919018 |
| 25 | 8486086 | 8650256 | 17136342 | 55 | 1601127 | 3026846 | 4627973 |
| 26 | 8145273 | 8448223 | 16593496 | 56 | 1490619 | 2854991 | 4345610 |
| 27 | 7812137 | 8247196 | 16059333 | 57 | 1387788 | 2684142 | 4071930 |
| 28 | 7486678 | 8047175 | 15533853 | 58 | 1292635 | 2514299 | 3806934 |
| 29 | 7168896 | 7848159 | 15017055 | 59 | 1205157 | 2345462 | 3550619 |
| 30 | 6858791 | 7650150 | 14508941 | 60+ | 37768327 | 24265284 | 62033611 |

The adjusted single year distributions of population according to two sexes from ages 1 to 59 gave a monotonically decreasing series with male figures always being greater than females. The '0' population can be estimated using vital statistics and life table survival function. This part is not done here because the main purpose of the paper is to give a methodology on arriving at an adjusted single year age data using only a raw grouped age data usually available just after a few months census is over in those developing countries. It indicated a true age data in single year with consistent nature. However, the task is not completed here. In theory two things are yet to be tested i.e., smoothness and fit. For the former, the figures showed a smooth series, although the smoothness could have been done by forming difference table and fitness of the data with the observed figures could also have been done. As the raw single year data are not yet available, that part is yet to be tested. If time allows from the conference, this can also be done as and when they become available. Alternatively, this part of the work may be completed by the future generation.

## References

Mukhopadhyay, B.K. 1986. An age adjustment of very young children of India, 1981 and reappraisal of fertility and mortality rates : A model appraoch, ***Genus***, Vol. XLII-n.3-4, Roma, Italy.

Som, R.K. 1973. ***Recall Lapse in Demographic Enquiries***, Bombay, Asia Publishing House.

United Nations. 1983. ***Indirect techniques for demographic estimation***, Manual X, UN Publication, New York, USA.

Zelnik, M. 1961. Age heaping in the United States Census, ***Milbank Memorial Fund Quarterly***, 39.