SOME ERRORS IN AGE REPORTING

By

S. M. ANWAR

Institute of Statistics, University of the Panjab, Lahore-Pakistan.

INTRODUCTION

Erraneous age reporting is quite common in sample surveys and population censuses. Since age distribution of a population is an important consideration in many matters of public poliey, it is necessary to check the accuracy of reported ages. This is done by means of certain indices which measure the errors of reported ages. Some well-known indices are :

(i) Myer's Index
(ii) Bachi's Index
(iii) Sex-Ratio Index
(iv) Age-Ratio Index
(v) Combination of sex-ratio and age-ratio indices.
(vi) United Nations secretariat Index.

In this study an attempt has been made to investigate the errors of reported ages in the data collected in a demographic survey of Lalliani, a township near Lahore. The data were collected in three rounds and the respondents consisted of persons from both sexes. Therefore the main points of interest in the study are :

(a) the relative accuracy of age reporting for male and female populations ; and

(b) the change in the magnitude of error in successive rounds.

However since the data collected in the second round was not available for analysis, the data of Round I and Round III only have

been used. The methods (i), (ii) and (vi) listed above are employed for computing an index of error. Each of these methods has advantages as well as shortcomings. It may be noted, in particular, that Myer's and Bachi's methods measure the extent of digit preference rather than age accuracy in a winder sense. Since not allmis-statements concerning age result in statements at preferred digits, and since digit- preference can be present whether or not age mis-statement have any directional tendency a measure of digit preference should be interpreted with some reservation.

INDICES

Myer's index reflects preference or dislike for each of the ten digits from 0 to 9. To determine such preference, one might take successive sums of digits recorded at ages ending in each of these digits. Such a simple procedure does not suffice, however. For, with the advancing terminal digits of ages, these sums tend to decrease. To avoid this, first step in Myer's method consist in computation of a «blended» Population, as a result of which nearly equal sums are expected for all terminal digits. Therefore the blended total for each of the ten digits should be nearly equel to 10 per cent of the grand total. The deviations of these «blended» sums from 10 per cent of the grand total are added to gether irrespective of sign and the sum is Myer's index, I_m Thus if G denotes the grand total, and the «blended» sum for the jth digit,

$$I_m = \sum_{j=0}^{9} \left| a_j - G/10 \right|$$

In table I—IV, Myer's index I_m has been computed for the data collected in the demographic survey of Lalliani. In the first two tables (Tables I and II) the data of Round I are used, while the data collected in Round III are used in tables III and IV. The data collected in Round II were not available for analysis. The «blended» sums were computed as follows. First the sums of numbers of all ages terminating in digits from 0 to 9 were computed for ages 10 and over, and for ages 20 and over. The former were multiplied by coefficients 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and the latter by coefficients 9, 8, 7, 6, 5, 4, 3, 2, 1, 0. By adding these sums the blended Population is obtained. Myer has shown that, if the ages are reported correctly, the blended sums should be nearly all equal. Therefore, I_m should give a good measure of error in reported ages.

TABLE I

Application of Myer's Method to Age Data for Males
(Luliani Survey Round I (1961))

Terminal digit	Ages 10—99			Ages 20—99			Blended Sum 4+7	Percentage Distribution Taking35922	Deviation from ten
	Sum	Coefficient	Product 2x3=	Sum	Coefficient	Product 5x6=			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9) =100	(10)
0	1639	1	1639	1444	9	12996	14035	40.74	+30.74
1	105	2	210	34	8	272	482	1.33	— 8.67
2	325	3	975	182	7	1274	2249	6.26	— 3.74
3	211	4	844	85	6	510	1354	3.76	— 6.24
4	203	5	1015	72	5	360	1375	3.82	— 6.18
5	961	6	5766	792	4	3168	8934	24.87	+14.87
6	232	7	1624	175	3	525	2149	5.98	— 4.02
7	127	8	1016	58	2	116	1132	3.15	— 6.85
8	320	9	2880	112	1	112	2992	8.32	— 1.68
9	62	10	620	24	0	0	620	1.72	— 8.28
Sum							35922	100.00	91.27

TABLE II

Application of Myer's Method to Age Data for Females
(Lulliani Survey Round I (1961))

Terminal digit	Ages 10-99			Ages 20-99			Blended Sum 4+7=	Percentage Distribution 31612=100	Deviation from ten
	Sum	Cooffi-cient	Product 2x3=	Sum	Coeffi-Cieet	Product 5×6=			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	1360	1	1360	1224	9	11016	12376	39.15	+ 29.15
1	119	2	236	66	8	528	766	2.42	— 7.58
2	256	3	768	161	7	1127	1895	5.99	— 4.01
3	220	4	880	114	6	684	1564	4.95	— 5.05
4	195	5	975	69	5	345	1320	4.18	— 5.82
5	804	6	4824	697	4	2788	7612	24.08	+14.08
6	189	7	1323	71	3	213	1536	4.86	— 5.14
7	112	8	896	50	2	100	996	3.15	— 6.85
8	302	9	2718	159	1	159	2877	9.10	— 0.90
9	67	10	6670	28	0	0	670	2.12	— 7.88
Sum							31612	100.00	86.46

TABLE III

Application of Myer's Method to Age Data for Males (Lulliani Survey Round III (1963))

Terminal digit	Ages 10-99			Ages 20-99			Blended Sum 4+7=	Percentage Distribution 52129=100	Deviation from ten
	Sum	Coefficient	Product 2x3=	Sum	Coefficient	Product 5x6=			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	1109	1	1109	900	9	8100	9209	17.66	+ 7.66
1	619	2	1238	468	8	3744	4982	9.56	— 0.44
2	604	3	1812	426	7	2982	4794	9.20	— 0.80
3	476	4	1904	334	6	2004	3908	7.50	— 2.50
4	431	5	2155	303	5	1515	3670	7.04	— 2.96
5	760	6	4560	627	4	2508	7068	13.56	+ 3.56
6	689	7	4823	472	3	1416	6239	11.96	+ 1.96
7	364	8	2912	183	2	366	3278	6.29	— 3.71
8	546	9	4914	307	1	307	5221	10.02	+ 0.02
9	376	10	3760	221	0	0	3760	7.21	— 2.79
Sum			29187			22942	52129	100.00 99.99	26.40

TABLE IV

Application of Myer's Method of Age Data for Females According to Lulliani Survey III 1963

Terminal digit	Ages 10-99			Ages 20-99			Blended Sum 4+7=	Percentage Distribution 26068=100	Deviation from ten
	Sum	Coefficient	Product 2x3=	Sum	Coefficient	Product 5x6=			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	710	1	710	580	9	5220	5930	22.75	—12.75
1	296	2	392	232	8	1856	2248	8.62	— 1.38
2	347	3	1041	251	7	1757	2798	16.73	0.73
3	191	4	764	130	6	780	1544	5.92	— 4.08
4	176	5	880	119	5	595	1475	5.66	— 4.36
5	356	6	2136	315	4	1260	3396	13.03	3.03
6	352	2	2464	282	3	696	3160	12.12	2.12
7	166	8	1328	81	2	162	1490	5.72	— 4.38
8	258	9	2322	135	1	135	2457	9.43	— 0.67
9	157	10	1570	78	0	0	1570	6.02	— 3.98
Sum			13607			12461	26068	100.00	37.48

We compute I_m for male and female populations separately. We find that :

(a) For the data collected in Round I, I_m equal 91.27 for the male population (Table I) and 86.46 for the female population (Table II) ; and

(b) For the data collected in Round III, I_m equale 26.40 for the male population (Table III), and 37.48 for the female population (Table IV).

This shows, firstly, that in the first round the ages of female population are reported more accurately than those of the male population. The reverse is true for the ages reported in the third round. In the second place, there is a marked improvement in the accuracy of age reporting in the third round, probably due to the experience gained in the first two rounds.

Theoretically, Myer's index can vary from 0 to 180. If ages are approaches zero. If all the ages are reported with the same terminal reported correctly all the «blended» sums are nearly equal and I_m digits, I_m equals 180. The following indices for male population illustrate the extent of variation which may be found in the value of I_m

Bengal	(1901)	62.6
Russin	(1897)	20.5
Brazil	(1940)	16.3
Australia	(1933)	4.0
Sweden	(1939)	1.2
Turkey	(1945)	39.2
Lulliani Town	(1961)	
First Round		91.27
Third Round	(1963)	24.40

2. 2.—*Bachi's Index :*

In Bachi's index, the object is to determine the preference shown for each of the ten digits. If the age range is suitably chosen, nearly 10 per cent of the persons in the range should be expected to give their ages at each terminal digit. For theoretical as well as practictical reasons, the average number of persons in any two age ranges must be related with the number of persons reporting their ages with a particular terminal digit. Bachi's index is constructed as follows[3]. Let n_k represent the number of persons with reported age k years. We compute the following percentages :

1. $100\left(\sum_{j=0}^{4} n_{30+10j}\right) \Big/ \frac{1}{2}\left[\sum_{j=0}^{49} (n_{25+j} + n_{26+j})\right]$,

2. $100\left(\sum_{j=0}^{4} n_{32+10j}\right) \Big/ \frac{1}{2}\left[\sum_{j=0}^{49} (n_{26+j} + n_{27+j})\right]$,

3. $100\left(\sum_{j=0}^{4} n_{32+10j}\right) \Big/ \frac{1}{2}\left[\sum_{j=0}^{49} (n_{27+j} + n_{28+j})\right]$,

4. $$\frac{100\left[\sum_{j=0}^{3}(n_{33+10j}) + \frac{1}{2}(n_{23} + n_{73})\right]}{\left[\frac{1}{2}\sum_{j=0}^{44}(n_{23+j} + n_{24+j})\right]},$$

5. $$\frac{100\left[\sum_{j=0}^{3}(n_{34+10j}) + \frac{1}{2}(n_{24} + n_{74})\right]}{\frac{1}{2}\left[\sum_{j=0}^{49}(n_{24+j} + n_{25+j})\right]},$$

6. $$\frac{100\left[\sum_{j=0}^{3}(n_{35+10j} + \frac{1}{2}(n_{25} + n_{75})\right]}{\frac{1}{2}\left[\sum_{j=0}^{49}(n_{25+j} + n_{26+j})\right]}.$$

7. $100\left[\sum_{j=0}^{3}(n_{36+10j}) + \frac{1}{2}(n_{26} + n_{76})\right] \Big/ \frac{1}{2}\left[\sum_{j=0}^{49}(n_{26+j} + n_{27+j})\right]$,

8. $100\left[\sum_{j=0}^{3}(n_{37+10j}) + \frac{1}{2}(n_{27} + n_{77})\right] \Big/ \frac{1}{2}\left[\sum_{j=0}^{49}(n_{27+j} + n_{28+j})\right]$,

9. $100\left[\sum_{j=0}^{4} n_{28+10j}\right] \Big/ \frac{1}{2}\left[\sum_{j=0}^{49}(n_{23+j} + n_{24+j})\right]$,

10. $100\left(\sum_{j=0}^{4} n_{29+10j}\right) \Big/ \frac{1}{2}\left[\sum_{j=0}^{49}(n_{24+j} + n_{25+j})\right]$.

Let $d_j = 1, 2, \ldots 10$ be these percentages. Then $(d_j - 10)$, $j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ can be taken as indices of perference or dislike for the terminal digits from 0 to 9. Bachi's index is obtained by summing the positive differences. (Theoretically, the sum of positive deviations should be equal to the sum of negative deviations). Bachi's index is therefore about half of Myer's index, with a range of variation from zero to 90. In practice, the two methods yield almost the same result.[4] In table V-VIII the deviations $(d_j - 10)$, $j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ have been computed separately for male and female Populations in Round I and Round III. Bachi's indices are :

Round I

Male = 55.26, Female = 49.08.

Round III

Male = 19.25, Female = 24.08.

Thus we find that Bachi's index leads to the same conclusions as Myer's method, nemely :

(a) The female ages have been more accurately reported than male ages in the first round, while the reverse is true for the data of the third round.

(b) Age reporting in the third round is significantly better that in the first round.

2. 3.—*The United Nations Secretariat Method :*

This method consists of in the computation of sex ratios and age-ratios for five-year groups of ages up to age 70. One means of testing the accuracy of age distribution in five-year groups is to compare the sex-ratio (number of males per 100 females) for successive age groups. If the distributions are accurate or if the errors for males are as frequent and of the same kind as those of females, sex-ratio will change very gradually, from one group to another, as a result of sex differences in mortality. The presence of marked fluctuations in those ratios tostifies errors which are not the same for the two sexes. Another means of tesing the accuracy of age distribution in five year groups, is by age-ratio test, which is defined at the number reported in one age group per 100 the mean of number reported in two adjacent age groups for each sex separately. Any considerable fluctations of age-ratio indicates inaccuracies in age reporting.[5]

TABLE V

Application of Bachi's Method-Males Round I—1961

Percentage defined above, i.e., 1—10	1	2	3	4	5	6	7	8	9	10
	48.02	27.24	0.47	2.21	3.22	1.84	2.70	5.01	2.53	1.14
Deviation from 10 per cent.	+38.02	+17.24	— 9.53	— 7.79	— 6.68	— 8.16	— 7.30	— 4.99	— 7.47	— 8.86

TABLE VI

Application of Bachi's Method Female Round I — 1961

Percentage as defined above, i. e. No. 1—10	1	2	3	4	5	6	7	8	9	10
	42.06	27.02	1.51	2.90	2.90	1.73	4.77	3.46	2.76	1.50
Deviation from 10 per cent.	+32.06	+17.02	— 8.49	— 7.10	— 7.10	— 8.27	— 5.23	— 6.54	— 7.24	— 8.50

The Bachi's Indices for males and females Round I between age range 23—77, have been worked out to be 55.26 and 49.08 respectively.

TABLE VII

Application of Bachi's Method Males Round III — 1963

Percentage as defined above, No. 1 — 10	1	2	3	4	5	6	7	8	9	10
	20.13	15.92	10.68	12.52	7.68	4.50	7.40	8.40	6.78	6.33
Deviation from 10 per cent.	+19.13	+ 5.92	+ 0.68	+ 2.52	— 2.32	— 5.50	— 2.60	— 1.60	— 3.22	— 3.67

TABLE VIII

Application of Bachi's method Females Round III — 1963

Percentage as defined above, i. e., No. 1—10	1	2	3	4	5	6	7	8	9	10
	26.47	15.30	9.85	11.71	10.60	4.08	5.20	7.42	5.36	2.44
Deviation from 10 per cent.	+16.47	+ 5.30	— 0.15	+ 1.71	+ 0.60	— 5.92	— 4.80	— 2.58	— 4.64	— 5.56

In the case of sex-ratio, successive difference between one age group and the next are noted, and their average is taken, irrespective of sign. In he case of age-ratio, for either sex, deviations from 100 are noted and averaged irrespective of sign. Three times the average of sex-ratio differences, is then added to the average of deviations of age-ratios from 100, to compute the index, known as the joint index. This procedure has been used for analysing the data of Lalliani Round I 1961 and Round III, 1963.

This method is applicable where single year age data are not avaiable. For small populations, the measurement is also affected by chance fluctuations. For smallness of population, an allowance is, however, to be made, for the adjustment. The adjusted joint Index is as under :

$$\text{Joint Index (Adjusted)} = \text{Joint Index (unadjusted)} \cdot \frac{3500}{P}$$

Where P stands for the population under consideration. The population of Lulliani Town is between eleven thousand and twelve thousand, and is not large enough to be affected by chance fluctuations.

The sex-ratios, age-ratios, their indices and joint Index (unadjusted) are calculated both for ages of males and females of the data for Round I and Round III, surveys of Lulliani (1961) and (1963), respectively.

The United Nations Secretariat Method has the advantages over the method of Myer's and Bachi's The index obtained is affected by differential omission of persons in various age groups from census count and by tendentious age mistatement as well as by digit-preference and is, therefore, more truly a reflection of the general accuracy of the age data. The methods applied to data by single year of age may in some cases show a fairly large amount of age mios-statement which has a little influence on the grouped data.

Age and Sex indices and Joint Indices for Round I and Round III show a large reporting errors in the age data but comparatively, there are less age reporting errors in round III than in round I, the indices calculated by different methods are as under.

Method	*Index*	(1961) Round I	(1963) Round III
1. Myer's Method	Males	91.27	26.40
	Females	86.46	37.38
2. Bachi's Method	Males	55.26	19.25
	Females	49.08	24.08
3. United Nations : Secretariat Method			
(i) Sex Ratio India		20.88	37.06
(ii) Age Ratio			
	Males	49.28	15.09
	Females	36.43	31.31
(iii) Joint Index (unadjusted)		148.35	156.58

The mean of sex differences in Round I depicts less differences as compared to mean of the sex differences in Round III, This can be attributed to the fact that in Round III male interviewers were employed who were not so efficient, as female interviewers employed in Round I, in their access to the children and female population. However, as expected, the mean age ratio is Round III is less than that of Round I consistent with the results obtained by the application of Myre's and Bachi's indices. The Joint Index which has greater influence of sex ratio is obviously greater for the data of Round III than that of Round I.

TABLE IX

Computation of Sex-Ratios, Age-Ratios and their Indices by United Nations Secretariat Method from the Age Data of 1961 Survey of Lulliani Town for Round I

	Reported Number		Analysis of Sex-Ratios		Analysis of Age-Ratios		Analysis of Age Ratios	
Age Group	Males	Females	Ratios	Successive differences	(Male) Ratios	(Male) Deviations	(Female) Ratios	(Female) Deviations
0— 4	907	784	115.68	—	—	—	—	—
5— 9	976	727	134.25	18.57	124.09	+24.09	111.84	+11.84
10—14	666	516	129.00	5.16	82.87	—17.13	86.28	—13.72
15—19	641	468	136.67	7.58	116.97	+16.97	95.13	— 4.87
20—24	430	470	91.48	45.19	82.30	+17.70	109.55	+ 9.55
25—29	404	389	103.48	12.19	99.88	— 0.12	82.59	—17.41
30—34	379	364	104.12	12.38	107.32	+ 7.82	113.39	+13.39
35—39	299	253	118.18	0.26	91.01	— 8.99	80.70	—19.30
40—44	278	263	105.70	14.06	114.40	+14.40	11.95	—88.05
45—49	187	188	99.49	12.48	65.96	—34.04	78.66	—21.34
50—54	289	215	134.42	6.23	214.07	+114.07	148.78	+48.78
55—59	83	101	82.18	34.95	32.81	—67.19	53.29	—46.71
60—64	217	164	132.32	52.24	326.31	+226.31	20.62	—79.38
65—69	50	58	86.20	50.14	8.72	—91.28	0.49	—99.51
70—74	127	71	—	—	—	—	—	—
Total irrespective of sign				271.43		610.61		473.85
Mean (Total divided by 13)				20.88		49.28		36.43
Joint Index							148.35	

TABLE X

Computation of Sex-Ratios, Age Ratios and Their Indices United Nations Secretariat Method from Age Data of 1963 Survey of Lulliani Twon for Round III.

Age Group	Reported Number		Analysis of Sex-Ratios		Analysis of Age-Ratios		Analysis of Age Ratios	
	Males	Females	Ratios	Successive Difference	(Male) Ratios	Deviation	(Female) Rates	Deviation
0— 4	1343	677	198.37	—	—	—	—	—
5— 9	1297	659	196.81	1.56	120.60	+20.60	121.48	+21.46
10—14	808	408	198.04	1.23	72.72	—27.28	73.71	—26.29
15—19	925	448	206.47	8.43	116.79	+16.79	113.99	+13.99
20—24	776	378	205.29	1.18	101.64	+ 1.64	100.00	—
25—29	602	308	195.45	9.84	91.21	— 8.79	56.62	—43.38
30—34	544	310	175.48	19.97	103.52	+ 3.52	118.77	+18.77
35—39	449	214	209.81	34.33	98.58	— 1.42	78.11	—21.89
40—44	367	238	154.20	55.61	92.32	— 7.68	124.28	+24.28
45—49	346	169	204.73	50.53	99.28	— 0.72	83.46	—16.54
50—54	330	167	197.50	7.13	118.92	+18.92	145.22	+45.22
55—59	209	61	342.62	145.02	74.11	—25.89	43.26	—56.99
60—64	234	115	203.48	139.48	137.64	+37.64	186.99	+86.99
65—69	131	62	211.29	7.81	74.86	—25.14	66.66	—33.34
70—74	116	71	—	—	—	—	—	—
Total irrespective of sign				481.78		196.03		407.05
Mean (Total divided by 13)				37.06		15.09		31.31
Joint Index							156.58	

Joint Index = 3 times mean difference Sex — Ratio plus mean deviation of male and female Age-Ratios.