

# **An Evaluation of Age- Sex Data of Census of Population and Housing 2001- Sri Lanka**

**H.R. Gunasekera**

**Director**

**Department of Census and Statistics, Sri Lanka**

**Paper presented to the 24<sup>th</sup> Population Census Conference  
Hong Kong, China 25 – 27 March 2009**

## **1. INTRODUCTION**

A Population Census is a complex, large-scale operation usually undertaken only once in every decade. Due to its complex operations a 'perfect' Census is unattainable. Errors inevitably arise in collection of Census data. The basic inputs for national policy formulations and most of the demographic research studies are the data obtained from Population Censuses. Consequently the scope of such research and the validity of their findings depend upon the accuracy of Census data. It is essential therefore to detect and quantify the errors by evaluation so that the users may be aware of the quality of Census data.

Among the large volume of data gathered in a Census, age and sex data play a vital role in population studies. The age-sex structure is one of the most fundamental characteristics of population composition. Past variations in the basic components of population change i.e. fertility, mortality and migration are reflected in the age-sex structure. Conversely, the age-sex composition of a population affects its fertility behaviour, mortality and morbidity levels, migratory movements, labour force participation and a host of other factors. Further, the age-sex data are the basic inputs for making population projections both at national and sub-national levels. Age-sex data are therefore almost always essential for analysis of population dynamics.

But the Census age-sex data are affected by errors. The two major types of errors are coverage errors and content errors. Coverage errors result from omission or duplication of individuals; hence it affects all the information collected including the age-sex data. Content errors, on the other hand, occur due to inadequate information supplied or mistakes made in reporting or recording information. A common form of content error is the misreporting of data. Misreporting of sex is generally rare. But age misreporting seriously affects the quality of age data. Hence evaluation of age-sex data is one of the most important steps in a Census evaluation programme.

### **1.1 OBJECTIVE**

This study is aimed at measuring the accuracy of age-sex data. The basic objectives of the study are:

- i) to examine the level of accuracy of the age-sex data collected in the 2001 Census of Population and Housing of Sri Lanka at the national level and to see the variations in the accuracy of these data in respect of place of residence and to compare those with 1981 Census results.
- ii) to examine the variation in accuracy of the age sex data district wise, with the intention of highlighting areas where improvements pertaining to the quality of age-sex data could be made.

## **1.2 DATA SOURCES AND LIMITATIONS**

The data used in this study have been obtained from the complete tabulations of the 2001 Census of Population and Housing of Sri Lanka. Census of Population and Housing, 2001 was carried out on 17<sup>th</sup> July 2001 after a lapse of 20 years. The Census enumeration was able to carry out completely in 18 districts only, due to the conditions prevailing in Northern and Eastern provinces. Therefore, the data used to analyse in this study confined to the 18 districts in which the enumeration was completed. Whenever a comparison is made with 1981 Census, the same 18 districts were taken into account. In the 2001 Census, age data were collected through the question on date of birth and the completed years were calculated at the time of processing. Enumerators were asked to confirm the recorded date of birth as far as possible by comparing it with some available document such as national identity card, birth certificate etc. When the date of birth was not known and the relevant documents are also not available, enumerators were requested to make every effort to decide the person's age by comparing the ages of members of the family or by reference to some important public or personal event.

## **1.3 METHODOLOGY**

Age data at national level are examined first by visual inspection in the form of single year age pyramids; any unusual concentrations at particular ages indicate probable errors in age data. Inaccuracies so detected are quantified by calculating U.N. age sex accuracy index.

### **1.3.1 U.N. SECRETARIAT AGE – SEX ACCURACY INDEX**

This index is developed on the basis of sex ratios and age ratios, computed for five year age groups, up to age 70. In this context sex ratio has been defined as the number of males per 100 females in the same age group; and the age ratio as 100 times the number of persons in a given age group divided by the arithmetic average of numbers in the two adjoining age groups. Successive differences in sex ratios are obtained and the total of successive difference without regard to sign is calculated, from this sum the mean of the differences is obtained. Then the age ratios in five year age groups are obtained for each sex and deviations from 100 are summed irrespective of sign and mean deviation is calculated. A weightage of 3 units is given to sex ratio score and 1 unit each is given for the age ratios. The sum obtained after giving the weightage is known as the U.N. joint score.

The index is interpreted as follows;

If the joint score is less than 20 (low range), the data are described as being reasonably reliable; between 20 and 40 (cautious range), fairly unreliable; over 40 (high range), quite unreliable (U.N.1952)

This U.N. index has several advantages over other similar indices. Firstly it is applicable to the more common form of 5 year age groups rather than single years age data. Secondly, U.N. index takes into account the differential omission of persons in various age groups in addition to the age misstatements. Therefore it reflects the general accuracy of age sex data. However it is not without limitations, which includes;

- i) Failure of the index to take account of the expected decline in sex-ratio with increasing age and of real irregularities in age distribution due to migration, war and epidemics as well as normal fluctuations in births and deaths.
- ii) The use of a definition of an age ratio which omits the central age group and which therefore has an upward bias.
- iii) Considerable weight is given to the sex-ratio component in the formula.  
(Shryock and Siegel 1973)

The above indices are calculated at national level and compared with the corresponding 1981 Census figures. Variation of the indices is examined for place of residence. To measure the accuracy of age-sex data at district level age accuracy indices are calculated for each of the 18 districts in which comparable data are available and compared with the 1981 Census.

## **2. ACCURACY OF AGE-SEX DATA AT NATIONAL LEVEL**

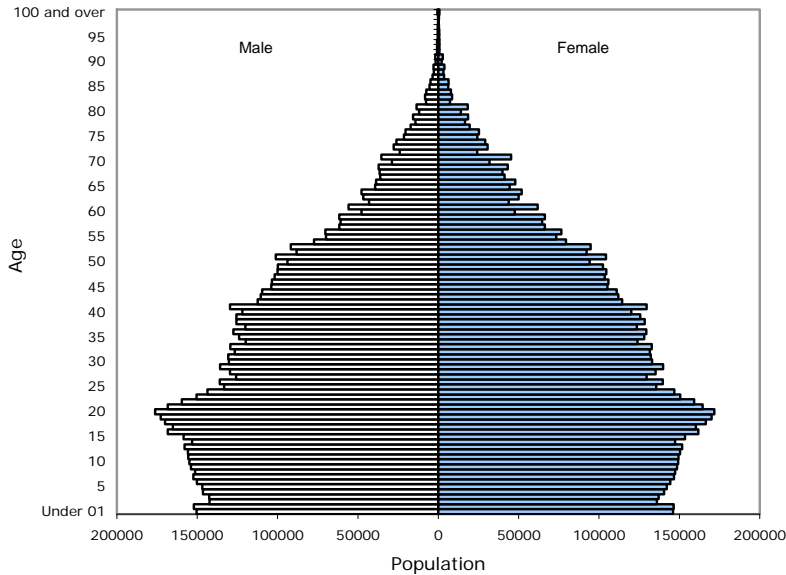
This section attempts to measure the accuracy of age-sex data in the 2001 Census at national level; the results are compared with similar measures for the 1981 Census data. Age data in single years and sex ratios will be analysed and accuracy indices will be calculated to assess the extent of possible misreporting and under-enumeration of age-sex data.

### **2.1 VISUAL INSPECTION OF AGE DATA**

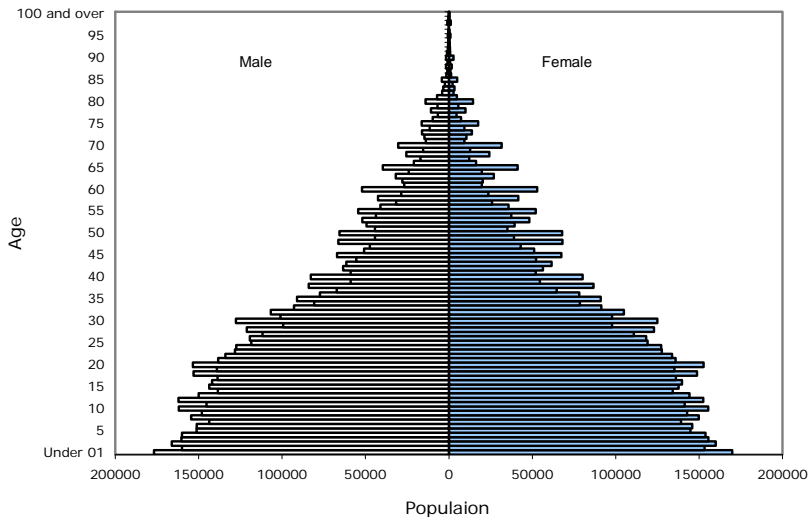
A particular and a common form of age misreporting is age heaping i.e. a noticeable concentration of reported ages at some specific digits. This is more frequently found at the ages ending with digits 0

and 5. Age heaping is apparent when the single year age distribution is examined visually. Single year age distribution for 2001 and 1981 are shown in Figure 2.1 and Figure 2.2 respectively.

**Figure 2.1 : Population distribution by single years of age and sex, 2001**



**Figure 2.2 : Population distribution by single years of age and sex, 1981**



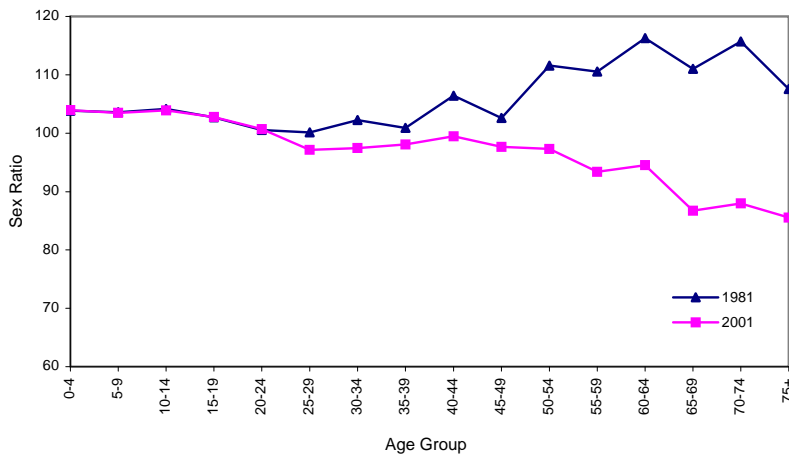
For comparability purposes 1981 Census data are confined to the 18 districts in which 2001 Census was completely carried out. If age heaping is present to a significant extent we would expect large concentration of individuals at specific ages and relatively smaller numbers at other neighbouring ages. Unusual concentrations at digits 0 and 5 which are very common in populations of many developing countries is virtually absent in 2001 Census data for Sri Lanka. Closer inspection of Figure 2.1, however, discloses that slightly excess concentration of individuals at age 41,51,53,61 and

71 are apparent for both males and females. In fact the pattern for females seems to be slightly prominent than males. Comparison of Figure 2.1 and Figure 2.2 reveals that unusual concentrations at digits 0 and 5 (and 2 and 8 to some extent) from age 30 and onwards in 1981 have become almost non-existing in 2001. This suggests improvements in accuracy of age data in 2001 compared to 1981.

## 2.2 PATTERN OF SEX RATIO WITH AGE

Sex ratio at birth is usually around 104 because of the biological fact that male births generally exceed female births. It should then decline gradually with age due to lower mortality of females. Any major deviation from a smooth pattern is attributable to (a) fluctuations in demographic components such as mortality and migration. (b) Fluctuations in sex ratio at birth. (c) Misreporting of ages and/or differential completeness of enumeration of male and females at different ages. The pattern of sex ratio with 5 year age groups is shown graphically in Figure 2.3.

Figure 2.3 : Age specific sex ratios, 1981 and 2001



The curve in 2001 is very close to the expected pattern of smooth declining trend; the value at 60-64 seems to be an outlier. Comparison of five year sex ratios with 1981 Census immediately reveals that the erratic fluctuations prevailed in 1981 is almost absent in 2001. This implies improvements in accuracy of age sex data.

## 2.3 U. N AGE, SEX ACCURACY INDEX

The preceding section examined the accuracy of age sex data in terms of age pyramids and the patterns in sex ratios. This section quantifies the accuracy by calculating UN age sex accuracy index

at national level. Whenever comparisons are made with 1981 Census data, the indices are recalculated for 18 districts in which 2001 Census was completely carried out.

Table 2.4 shows the sex-ratio score, age ratio score and U.N. index for 2001 Census along with the corresponding values of 1981.

**Table 2.4: U.N. age-sex accuracy index –1981 and 2001**

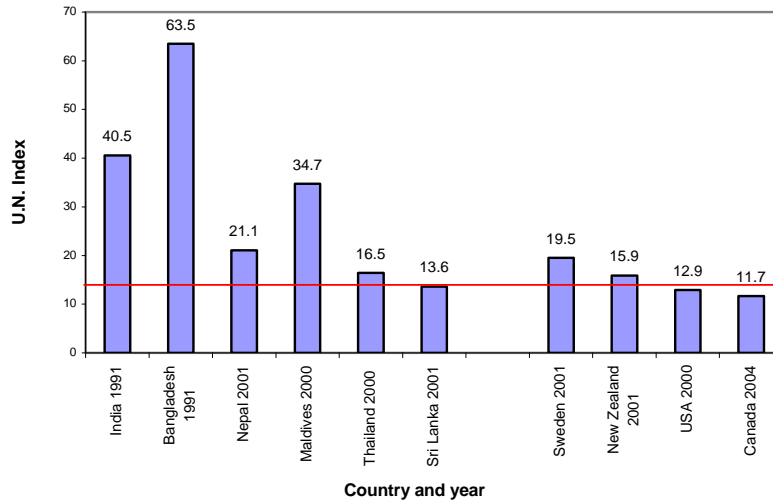
Census Year	Sex-ratio Score	Age-ratio score		U.N.Index
		Male	Female	
1981*	3.1	4.7	3.4	17.5
2001	1.9	4.1	3.9	13.6
Percent Change	- 38.7	- 12.8	+ 14.7	- 22.3

\*Computed for 18 districts in which 2001 Census was completed.

U.N. joint score stood at 13.6 in 2001, which shows a decline of 22.3 percent over the 1981 value. Since its value is well below 20 (low range) as discussed in 1.3.1, it can be concluded that age-sex data are very reliable in 2001. Since joint score is the combination of sex ratio and age ratio scores its value is determined by the characteristics of sex ratio scores and age ratio scores. It is important to note that age ratio score of females is slightly less than males, indicating a more smooth age pattern of females. Age data becomes smooth when classified into 5 year age groups. When compared with 1981 Census, sex ratio score has declined very significantly by 38.7 percent, which is mainly attributable to the decline in U.N. joint score. Male age ratio score is also declined by 12.8 percent, but female age ratio score is increased by 14.7 percent. The latter increase can not be attributable to age misreporting of females per se as it can occur due to factors of population change such as fertility, mortality and migration applicable to certain age groups. However this needs further investigation, which is beyond the scope of this paper.

U.N. age-sex accuracy index for some developing countries in Asia and for selected developed countries are shown in Fig. 2.4.

**Figure 2.4 : Comparison of U.N. index with other countries**



Source : Computed from data obtained from official web-sites

Among the developing countries under concern, Sri Lanka records the lowest value of 13.6 indicating the most accurate age-sex data. India, the neighbouring country has an index, which is nearly 3 times higher than that of Sri Lanka. It is interesting to note that Sri Lanka’s general accuracy of age-sex data is even better than Sweden and New Zealand as shown by U.N. age-sex accuracy index. Although U.S.A. and Canada record more accurate age-sex data than Sri Lanka, the differences in the levels are marginal.

### **3. ACCURACY OF AGE-SEX DATA BY PLACE OF RESIDENCE**

All areas administered by Municipal and Urban Councils constitute the urban sector. Estate sector consists of all plantations which are 20 acres or more in extent and with 10 or more resident labourers. All areas other than urban and estate comprise the rural sector. The definition of urban sector has certain limitations. Areas having urban facilities and outlook which are situated outside Municipal and Urban Council areas are not considered as urban under this definition. Hence it leads to an underestimate of urban sector in Sri Lanka. Further in 1981 Census, separate tabulations were not available for estate sector; hence when a comparison is made with 1981, 2001 Census data are also classified into two sectors viz urban and rural by amalgamating estate into rural sector.

**Table 3.1: U.N. age-sex accuracy index by place of residence and sex-1981 and 2001**

Place of Residence and Sex		Sex-ratio score	Age-ratio score		U.N. Index
			Male	Female	
Urban	1981*	3.6	5.3	4.3	20.3
	2001	2.6	5.7	5.1	18.6
	Percent Change	-27.8	+7.5	+18.6	-8.4
Rural	1981*	3.5	4.6	3.5	18.7
	2001	2.3	3.9	3.7	14.4
	Percent Change	-34.3	-15.2	+5.7	-23.0

\*Computed for 18 districts in which 2001 Census was completed.

Overall U.N. index for rural sector (14.4) is below the value recorded for urban sector (18.7). All its components viz. sex ratio score and age-ratio scores for rural sector fall below the corresponding values for urban sector. Improvements in recording age-sex data between 1981-2001 as shown by U.N. Index are seen for both sectors but the improvements are more prominent for rural sector. These improvements are mainly attributable to declines in sex ratio score. In fact, some increases are seen in age ratio score of both males and females in urban sector while rural sector females record an increase in age ratio score.

#### 4. ACCURACY OF AGE-SEX DATA BY DISTRICTS

There is high degree of variation of demographic and socio-economic characteristics among districts. Hence it would be of interest to study the variations of the levels of accuracy of age-sex data by districts, in order to highlight the areas where improvements pertaining to the quality of age data could be made. In this context analysis is limited to the comparisons of levels of UN age Sex accuracy index by districts.

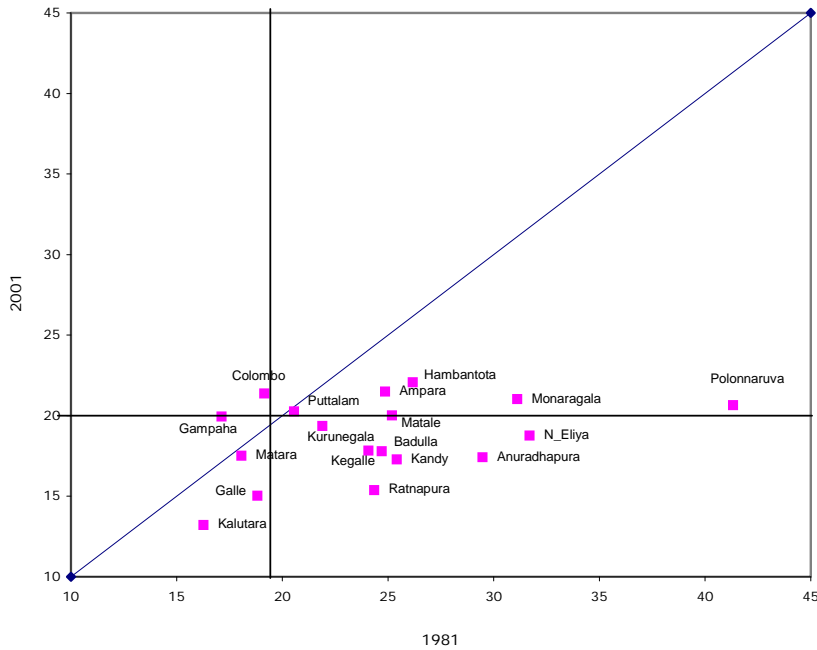
U.N. age sex accuracy index with age ratio scores and sex ratio scores by districts are shown in Table 4.1.

**Table 4.1: UN age-sex accuracy index by districts - 1981 and 2001**

District	Index - 1981	Index - 2001			UN Index	Percentage Change
		Sex Ratio Score	<u>Age Ratio Score</u> Male    Female			
Colombo	19.2	3.4	6.3	5.0	21.4	11.5
Gampaha	17.1	2.8	5.4	6.0	20.0	17.0
Kalutara	16.3	1.7	4.4	3.8	13.2	-19.0
Kandy	25.4	2.9	4.7	3.8	17.3	-31.9
Matale	25.2	3.8	4.7	3.8	20.0	-20.6
Nuwara Eliya	31.7	3.2	4.7	4.5	18.8	-40.7
Galle	18.8	2.7	3.6	3.3	15.0	-20.2
Matara	18.1	2.5	5.4	4.5	17.5	-3.3
Hambantota	26.2	2.6	7.0	7.4	22.1	-15.6
Ampara	24.9	3.9	5.9	4.0	21.5	-13.7
Kurunegala	21.9	3.2	5.0	4.6	19.4	-11.4
Puttalam	20.6	3.0	5.7	5.7	20.3	-1.5
Anuradhapura	29.5	2.9	4.2	4.5	17.4	-41.0
Polonnaruwa	41.3	4.1	3.9	4.4	20.7	-49.9
Badulla	24.7	3.2	4.3	3.9	17.8	-27.9
Monaragala	31.1	3.5	5.4	5.0	21.0	-32.5
Ratnapura	24.3	2.3	3.8	4.8	15.4	-36.6
Kegalle	24.1	3.1	4.3	4.1	17.8	-26.1
<b>Total</b>	<b>17.5</b>	<b>1.9</b>	<b>4.1</b>	<b>3.9</b>	<b>13.6</b>	<b>-22.3</b>

Variations in overall index are graphically shown in Figure 4.1.

Figure 4.1 : U.N. age-sex accuracy index for districts, 1981 and 2001



The general accuracy of age-sex data in 2001 is highest in Kalutara district followed by Galle and Ratnapura districts. The lowest accuracy is found in Hambantota district. In addition to Hambantota district, Ampara, Colombo, Moneragala, Polonnaruwa, Puttalam, Matale and Gampaha districts marginally fall into the cautious range as defined in 1.3.3, while all the other districts belong to the low range (See Figure 4.1). Out of these 8 districts, Gampaha and Matale are in the borderline. It is surprising why Colombo and Gampaha districts have deteriorated in general accuracy of age-sex data compared to 1981 while all other districts have improved in quality. In fact, Polonnaruwa, Anuradhapura, and Nuwara-Eliya districts show tremendous improvements in general accuracy of age-sex data recording more than 40 percent change in U.N. index.

Therefore it is interesting to examine why the index has increased for Colombo and Gampaha districts. Population by sex, age ratios and sex ratios for Colombo district is shown in Table 4.2

**Table 4.2: Population, by age and sex, age ratios and sex ratios  
– Colombo district 2001**

Age	Population		Age ratio		Sex ratio (males per 100 females)
	Male	Female	Male	Female	
0-4	85,276	81,228			105.0
5-9	82,036	77,866	99.2	99.2	105.4
10-14	80,158	75,716	86.0	87.8	105.9
15-19	104,340	94,536	96.7	98.6	110.4
20-24	135,584	116,004	124.9	119.2	116.9
25-29	112,710	100,183	95.6	96.9	112.5
30-34	100,122	90,804	99.1	97.5	110.3
35-39	89,406	86,037	98.8	101.4	103.9
40-44	80,918	78,921	101.9	102.0	102.5
45-49	69,423	68,659	95.3	96.2	101.1
50-54	64,763	63,816	111.1	108.1	101.5
55-59	47,125	49,376	95.6	98.9	95.4
60-64	33,842	36,023	93.4	91.6	93.9
65-69	25,325	29,235	96.4	99.5	86.6
70-74	18,710	22,756			82.2
75+	21,675	28,701			75.5
Total	1,151,413	1,099,861			104.7

It clearly shows higher concentration of male dominated young population between 15-34 as a result of migratory movements to Colombo. The numbers at 20-24 age group are extra ordinarily large which gives unusually high age and sex ratios. Smoothing out the population for this age group alone reduce the U.N. index to 17.8 from 21.4 which is lower than the 1981 value and falls into the low range. Similar information is shown in Table 4.3 for Gampaha district.

**Table 4.3: Population, by age and sex, age ratios and sex ratios  
– Gampaha district 2001**

Age	Population		Age ratio		Sex ratio (males per 100 females)
	Male	Female	Male	Female	
0-4	84,336	81,287			103.8
5-9	80,753	77,368	99.1	99.7	104.4
10-14	78,555	73,970	89.3	84.0	106.2
15-19	95,184	98,841	101.1	102.4	96.3
20-24	109,829	119,024	118.9	122.5	92.3
25-29	89,622	95,558	93.1	93.0	93.8
30-34	82,626	86,379	99.3	98.5	95.7
35-39	76,747	79,744	100.8	100.4	96.2
40-44	69,613	72,424	101.6	101.5	96.1
45-49	60,341	62,921	97.5	96.8	95.9
50-54	54,195	57,634	109.5	108.4	94.0
55-59	38,678	43,383	93.5	97.1	89.2
60-64	28,507	31,767	94.7	90.0	89.7
65-69	21,515	27,228	94.6	101.7	79.0
70-74	16,984	21,795			77.9
75+	20,217	26,659			75.8
<b>Total</b>	<b>1,007,702</b>	<b>1,055,982</b>			<b>95.4</b>

It also reveals excessive numbers, particularly at 20-24 age group, although it is less serious than Colombo. The sex ratio for that age group is well below 100 (92.3) in contrast to Colombo district where the sex ratio was as high as 116.9. This could be attributable to the migratory movements of young women, particularly to Export Processing Zone in Katunayake and surrounding areas. Smoothing out the population for this age group, as for Colombo district, reduces the U.N. index to 16.4 from 20.0, which is lower than the 1981 value and falls into the low range.

Hence, when interpreting U.N. index care should be taken to examine any changes in age-sex composition arising out of demographic components, as the index is unable to separate the inaccuracies and natural changes.

## 5. SUMMARY AND CONCLUSIONS

This analysis demonstrated that the age-sex data collected in the 2001 Census are very accurate. There was a remarkable improvement in accuracy between the two Censuses of 1981 and 2001. For example UN index decreased from 17.5 in 1981 to 13.6 in 2001 a percentage decline of 22 percent. So, Sri Lanka now falls into the very high levels in terms of accuracy of age data. Compared to some of the countries in SAARC region as well as some other developing Asian countries, Sri Lanka records the highest accurate age-sex data. This improvement could probably be attributed to improvements in data collection methodologies and socio-economic development of people. Although the age data are slightly less accurate for females than for males, the improvement between 1981 and 2001 is higher for females. As in the 1981 Census, the accuracy of age data is higher for rural than urban areas; but both areas fall into high accuracy region unlike in 1981. However, special attention should be paid to data collection procedures in urban areas and the supervision has to be strengthened in future censuses.

High accuracies are shown in general age-sex data for 10 out of 18 districts. Hambantota, Ampara, Colombo, Moneragala, Polonnaruwa, Puttalam, Matale and Gampaha districts marginally fall into the cautious range, although some districts show very significant improvements in general accuracy of age-sex data recording more than 40 percent levels. Out of these 8 districts deterioration of quality in Colombo and Gampaha districts seems to be spurious as the U.N. index is unable to separate the effects due to inaccuracies with natural, demographic changes. Hence only 6 districts viz Hambantota, Ampara, Moneragala, Polonnaruwa, Puttalam, and Matale should receive additional measures to achieve greater accuracy in the future Censuses.

## **REFERENCES**

- Myers, R.J. (1940) "Errors and Bias in the Reporting of Ages in Census Data", *Transactions of the Actuarial Society of America*, Vol. 41, Part 2, No.104, pp. 395-415.
- Shryock, H.S., Siegel, J.S. and Associates (1973), *The Methods and Materials of Demography, Vol. 1*. Washington, D.C. ; U.S. Government Printing Office.
- Ueda, K. (1980) , *Basic Demographic Data : Their Accuracy and use for Subnational Population Projections in Asian and Pacific Countries*. Tokyo; Southeast Asian Medical Information Center. pp. 1-41.
- United Nations. (1952), "Accuracy Tests for Census Age Distribution Tabulated in Five-year and Ten-year Groups", *Population Bulletin* , No.2, pp.59-79.