Digit Accuracies in Age Data for Developed Regions with old Pattern of Digit Preference Error in Developing World

Barun Kumar Mukhopadhyay*

Abstract

Turner (1958) considered a hypothetical ranking from 0 to 9 digits in age reporting, such as '0' the maximum ranking '5' the next, '2', '4' and, etc., and later compares with the observed ranking by Spearman's coefficient of rank correlation in order to compare the quality of age reporting among different countries of the world around fifties. In the present time his method may not be suitable for developed countries following almost a progressively declining series in their raw single year age data. As such present author tries to develop some theoretical ranking only for developed countries based on life table stationary population by ten digits within a broad age band being depleted by death only assuming no effects of fertility and migration. The findings are very encouraging as a "case study" to study the "digit accuracy index", among developed countries. However, Turner's ranking is most suitable for developing countries with heaping still exists**.**

* Retired Associate Scientist 'B' from Pop. Studies Unit, Indian Statistical Institute, Kolkata, WB, India

Introduction

Turner's method

The study conducted by Turner (1958) uses rank difference correlation method (ρ) comparing the observed rankings of populations of a broad age band with a hypothetical ranking of ten digits 0, 1, 2,.. 9. The ranking was,- heaping should occur most at ages ending in multiples of ten, next multiples of five, next at multiples of two and next for odd digits Different researchers later argued, cited and used his index (e.g., Stockwell, 1966 ; Stockwell and Wicks, 1974 ; Ewbank, 1981; Mukhopadhyay and Mukherjee, 1988 and Jowett and Li, 2004).

Turner's Ranking

Age-range	Digit	Turner's Ranking
10-69	0	1
	1	9.5
	2	3.5
	3	7.5
	4	4.5
	5	2
	6	4.5
	7	7.5
	8	3.5
	9	9.5

Turner's application for some developing areas

The method of Turner has been applied in the following developing regions, as a case study. A number of tables have been prepared in the next section from the data of demographic yearbook obtained from online network service so that a viable comparison could be made in different developing countries. The Spearman coefficient (ρ) is found highly significant ($p < 0.00$) for six developing countries taken for the case study.

Census	Turner's ρ
India 1991	**0.96 (M), 0.96 (F)**
Uganda, 1991	**0.94 (M), 0. 96 (F)**
Ethiopia, 1994	**0.96(M), 0.96(F)**
Gambia, 1993	**0.84(M), 0.91(F)**
Nigeria, 1991	**0.77(M), 0.86(F)**
Yemen, 1994	**0.77(M), 0.82(F)**

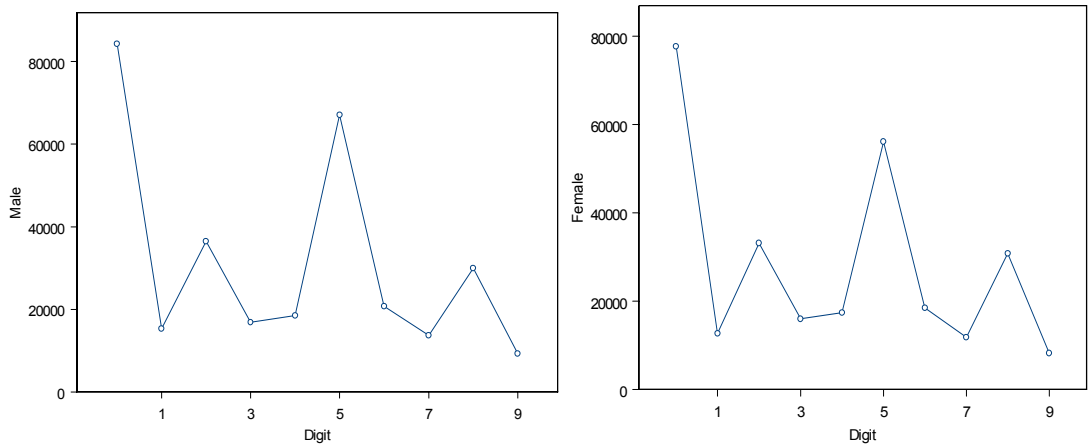

Fig. 1: Population within (10-69) years of age at digits 0 to 9 for India, census, 1991

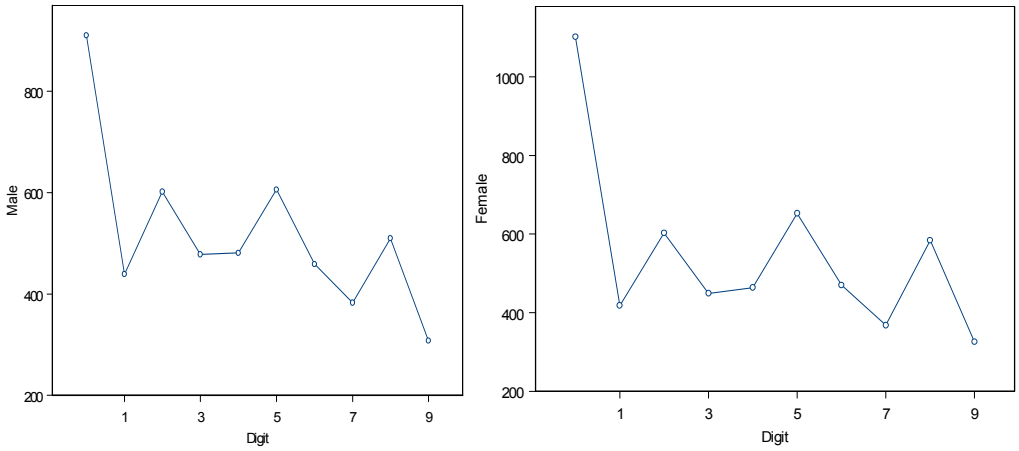

Fig. 2: Population within (10-69) years of age at digits 0 to 9 for Uganda, census, 1991

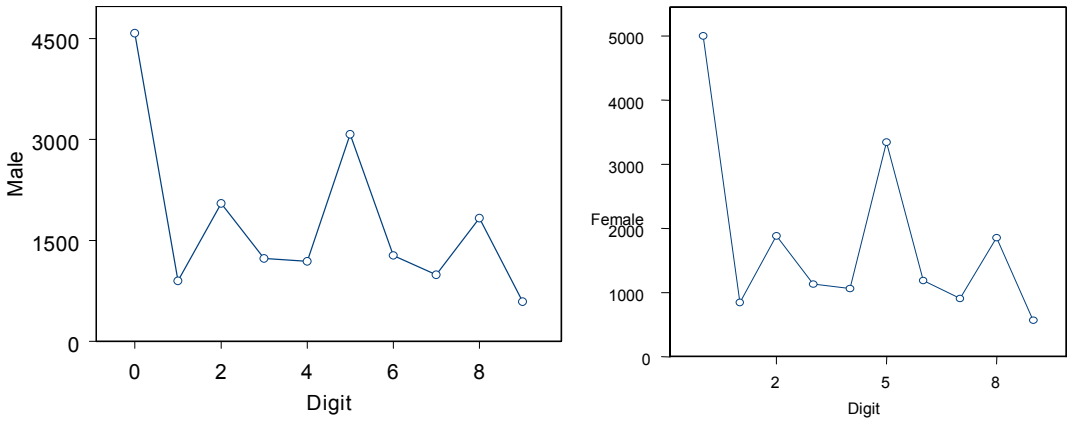

Fig. 3: Population within (10-69) years of age at digits 0 to 9 for Ethiopia, census, 1994

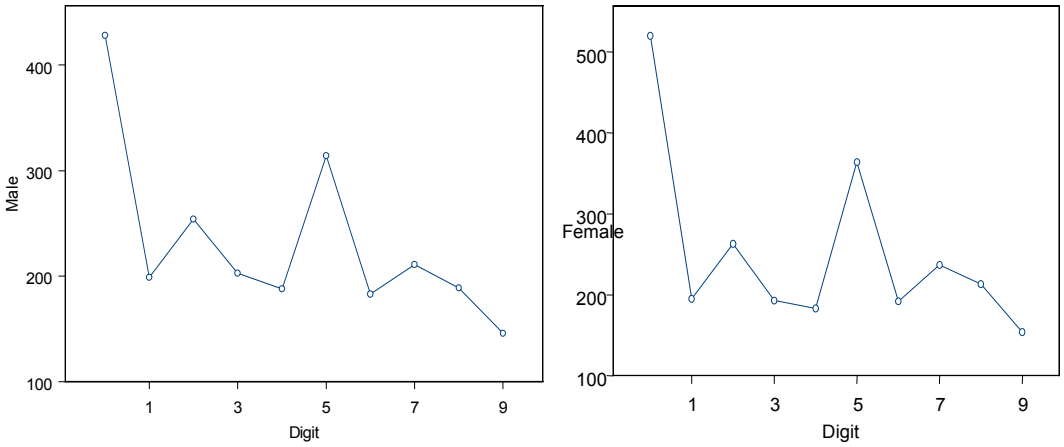

Fig. 4: Population within (10-69) years of age at digits 0 to 9 for Gambia, census, 1993

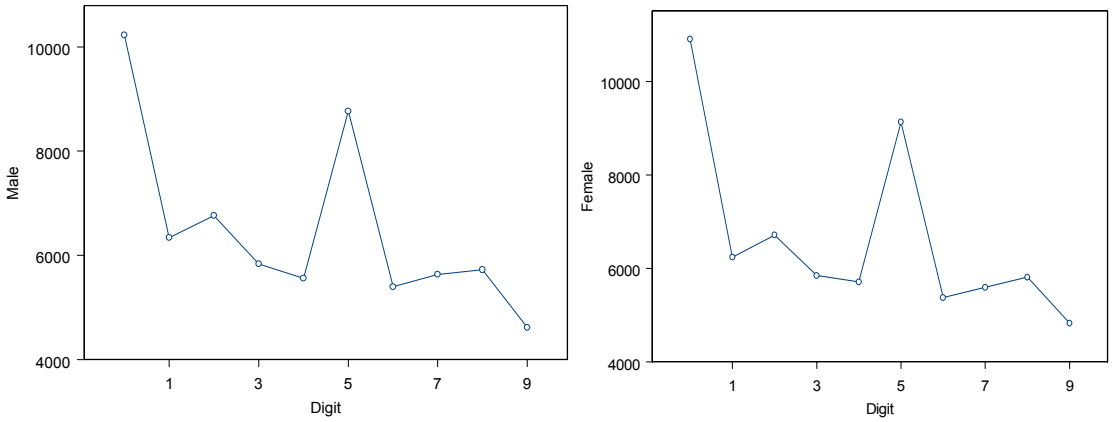

Fig. 6: Population within (10-69) years of age at digits 0 to 9 for Nigeria, census, 1991

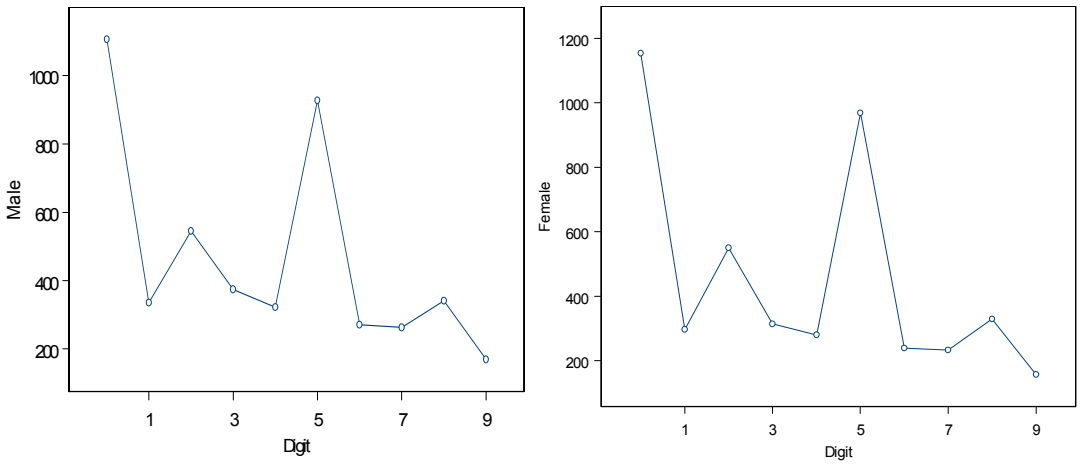

Fig. 10: Population within (10-69) years of age at digits 0 to 9 for Yemen, census, 1994

The new methodology for developed countries

Before explaining the new methodology, Turner's ranking method is done for a few developed areas where less heaping is commonly found. And age pattern almost follows a progressively declining series.The values are 0.16 (Male) and 0.14 (Female) for France, 0.28 and 0.28 for Ireland, 0.22 and 0.38 for Chile, and 0.24 and 0.27 for Canada. Now, thinking is that, life table population by ten digits depleted by only death could be matched with the single age data assuming there is no effect of fertility in the middle ages, vulnerable for heaping . The proposed current methodology gives below the rho values for a number of developed countries, taken as case study.

The table shows the digit accuracy indices for some developed countries around 1990's

Census	New ρ
France, 1990	**0.89 (M), 0.82 (F)**
New Zealand, 1990	**0.87 (M), 0.77 (F)**
Ireland, 1991	**1.00 (M), 1.00 (F)**
Canada , 1992	**0.99 (M), 0.98 (F)**
Australia,1991	**0.98 (M), 0.98 (F)**
Chile, 1992	**0.89 (M), 0.82 (F)**

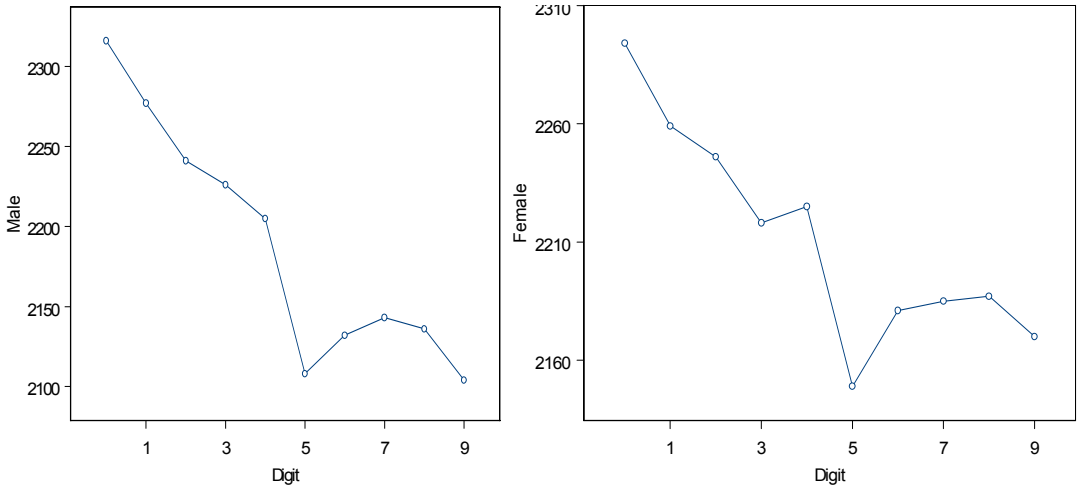

Fig. 12: Population within (10-69) years of age at digits 0 to 9 for France, census, 1990

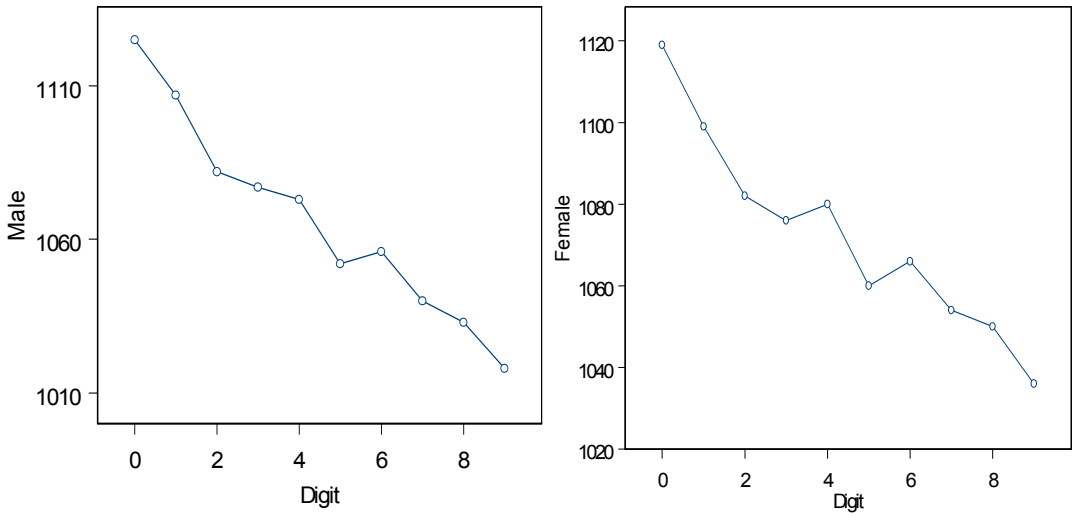

Fig. 20: Population within (10-69) years of age at digits 0 to 9 for New Zealand, census,1991

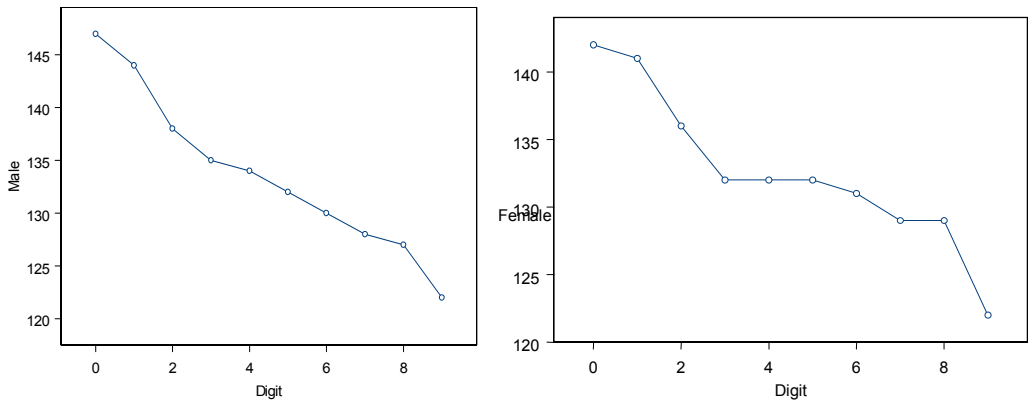

Fig. 13: Population within (10-69) years of age at digits 0 to 9 for Ireland, census, 1990

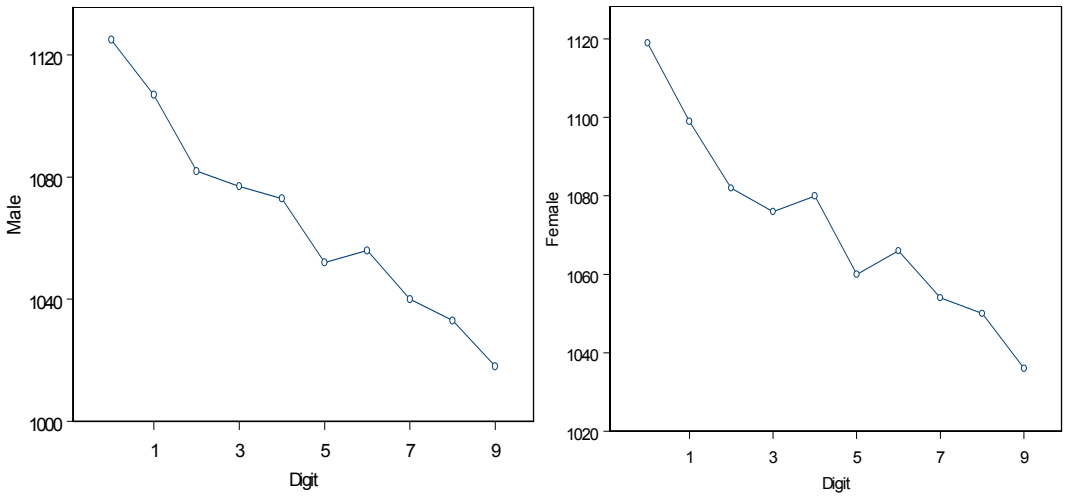

Fig. 18: Population within (10-69) years of age at digits 0 to 9 for Canada, census, 1992

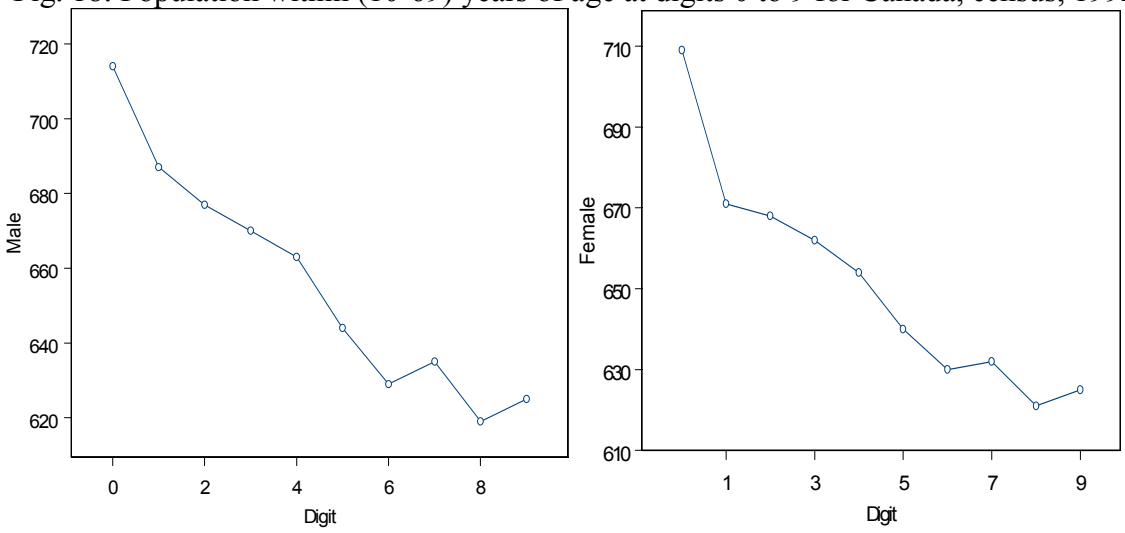

Fig. 19: Population within (10-69) years of age at digits 0 to 9 for Australia, census, 1991

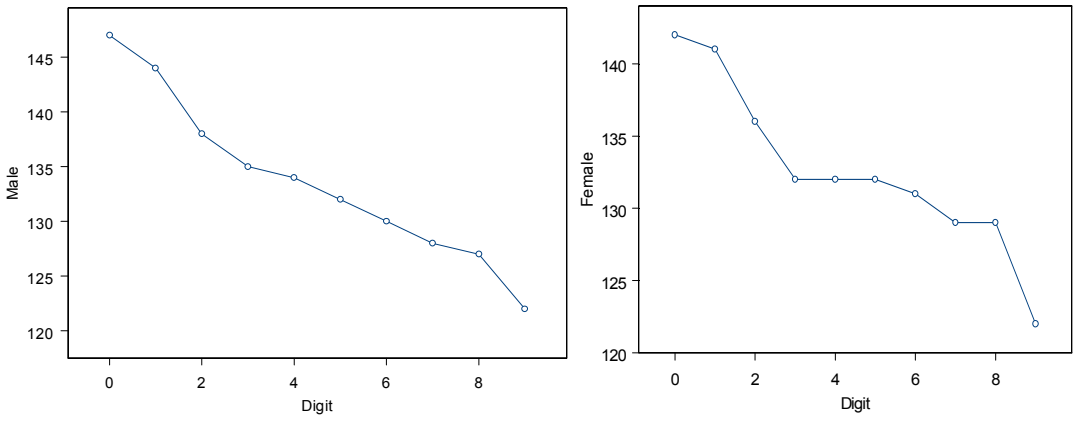

Fig. 14: Population within (10-69) years of age at digits 0 to 9 for Chile, census, 1992

Discussion

In the present paper two methods of digit preference error have been discussed. The first one is due to Turner (1958) and the second one is newly proposed by the present author. In the former case, the fixed hypothetical ranking of ten digits 0,1,2, ……9 were considered as such the most heaping should occur at ages ending in multiples of ten, next multiples of five, next at multiples of two and next for odd digits as already given in a tabular form earlier. Observed ranking of different countries were then compared with the hypothetical ranking by Spearman rank difference correlation method (r) in order to compare the digit preference error among those countries. No doubt, there were differences found.

Nevertheless, his technique was not suitable as far as developed countries where r's became so less. Accordingly an alternative ranking was thought of based on the pattern of life table population being depleted by mortality only without the effect of fertility in the actual age data in the middle ages may be more effective because their raw single year age data almost progressively declined as life table population. As such for digits 0 to 9 in a specified age band, say 10-69 ranking would be 1 to 10 and then was compared with the actual ranking of developed countries. It gave the digit accuracy of the country when the r's was highest. And next value was for next extent of accuracy so on. A new index was named as "digit accuracy index". Tables and graphs were shown for clarity.

Acknowledgement

I am thankful to Dr. A. K. Chattopadhyay, Associate Professor, Population Studies Unit, Indian Statistical Institute, Kolkata for his sincere help in providing me with some valuable references.

References

BACHI, R. (1951). The Tendency to Round off Age Returns, Measurement and Correction , *Bulletin of the International Statistical Institute*, 33, part IV, 195-222**.**

CARRIER, N.H. (1959) . A Note on the Measurement of Digital Preference in Age Recordings, *Journal of the Institute of Actuaries ,* Oxford.

EWBANK, D.C.(1981). *Age Misreporting and Age-selective Under enumeration: Sources*, *Patterns and Consequences for Demographic Analysis*, Report 4, Committee on Population and Demography, United States National Academy of Sciences, Washington.

JOWETT, A. JOHN and LI, YUAN-QING (1992). Age – Heaping: Contrasting Patterns from China, *Geojournal*, Vol. 28 no. 4, pp 427-442.

KING, G. (1915). The New National Life Tables, *Journal of the Institute of Actuaries,* Vol.49,Oxford.

MUKHOPADHYAY, B.K., MAJUMDAR, P.K. AND GHOSH , N.C. (1999). Digit Preference in Age Data : An Implicit Model Approach, *Journal of Bihar Mathematical Society,* Vol 19, pp 51-59.

MUKHOPADHYAY, B.K. AND MUKHERJEE , B.N. (1988). A study of Digit Preference and Quality of Age Data in Turkish Censuses , *Genus,* Vol XLIV-n.1-2, Italy.

MUKHOPADHYAY, B.K. AND MUKHERJEE, B.N. (1983). *A Comparative Study* of *Certain Indices of Digit Preference in the Census Single Year Age Returns of Developed and Developing Countries*, Presented in the Ninth Annual Conference of the Indian Association for the Study of Population (IASP), Jaipur, India.

MYERS , R.J. (1940). Errors and Bias in the Reporting of Ages in Census Data, *Transaction of the Actuarial Society of America,* Vol XLI, Part 2, No.104.

RAMACHANDRAN, K.V. (1965). An Index to Measure of Digit Preference Error in Age Data , *Proceedings of the UN's World Population Conference (Summary),* Belgrade, Yugloslavia.

STOCKWELL, E.G. and WICKS, J.W.(1974). Age Heaping in recent National Censuses, *Social Biology* 21:2, pp163-167.STOCKWELL, E.G. (1966). Patterns of Digit Preference and Avoidance in the Age Statistics of Some Recent National Censuses: a Test of the Turner Hypothesis, *Eugenics Quarterly* 13:3: pp 205-208.

TURNER, S.H. (1958). Patterns of Heaping in the Reporting of Numerical Data, *Proceedings of the Social Statistics Section (American Statistical Association)* pp 248-251.

ZELNIK, M. (1961). Age Heaping in the United States Census, *Milbank Memorial Fund Quarterly, 39, pp 540-573.*