Statistical analysis of the 2006 census age distributions of Adamawa and Anambra States, Nigeria

Grace Monday Udoh[b]

Department of Statistics, Akwa Ibom State Polytechnic, Ikot Osurua, Akwa Ibom State, Nigeria

A census of a population contributes immensely to the development of social amenities of a country. Censuses in developing countries are prone to errors of age misreporting due to ignorance, low literacy levels and other social, economic and cultural factors. Ages are commonly rounded with great affinity for 0 and 5. In view of this tendency to digital preference and / or avoidance results in age heaping or concentration of ages at certain digits, the study examined the extent of digit preference in the Nigerian census data of 2006 in two states and the best method of data smoothing. Whipple's and Myers' indices were used to determine the extent of digit preference. The results showed that the qualities of age data in the two states are very rough as a result of age misreporting.

Keywords: population census; digit preference; age reporting; Whipple's index; Myers' index.

1 Introduction

Population Census according to United Nations (2008) is the total process of collecting, compiling and publishing demographic, economic and social data pertaining to a specified time or times to all persons in a country or de-limited territory. After the 1914 amalgamation of the Northern and Southern Nigeria, a comprehensive census took place in 1921 followed by 1931, 1953, 1962/63, 1973, 1991 and 2006 census (Akingbade, 2004; Aluko, 1965; Ekanem, 1972; Okolo, 1999 and Okafor et al., 2007).

All the censuses taking so far in the country, none has passed without some levels of controversy, ranging from lack of competency on the side of the government agency to lack of trust among the populace, and ignorance of the objective essence of population issues (Inekwe, 2014; Makinwa, 1985 and Mathews, 2002). The accuracy of age data collected during census by different country varies in different set-ups and depends on numerous factors. Different set-ups have different social values attached to age Spardeshi (2010).

The two states looked into are Adamawa state and Anambra State. Adamawa, situated in the north-eastern Nigeria was administratively created in 1991 from the north-eastern half of former Gongola State and its capital Yola. It is bordered on the north and northeast by Borno State, southwest by Taraba State and on southeast by Cameroon and are predominantly farmers, while Anambra State was created in 1976 from part of the East Central State with its capital Awka. Its boundaries are formed by Delta State to the west, Imo State to the south, Enugu State to the East and Kogi State to the North, and are predominantly business men.

Reviewing the chronicles of census planning and administration in Nigeria since Nigerian independence, Okafor et al. (2007) concluded that the 2006 census exercise produced one out of the numerous provocative results which has no direct connection to reality and Odewumi (2000) and Onyekakeya (2007) in their conclusion after cross

[b] E-mail: graceudoh14@gmail.com

examination of census results in Nigeria since independence, maintained that culture, lack of information and ethnic sentiment, holds a strong ground in what people interpret census exercise to be. This, of course, brings about the question about age reporting and its implication to the quality of the 2006 census data in Nigeria in the two states. The quality of age reporting in Nigeria is one of the factors which seriously affect the quality of the entire results per se. Age reporting is one of the essential part of census data which inform other higher degree research ventures (Economic and Social Commission for Western Asia, 2013). In view of the arguments about the 2006 census result and its predecessors, using well known indices; the Whipple's and Myers' indices to analyze the data and other age accuracy index to smoothing the error due to digital preference in the two states, the study is poised to answer the following questions through the population priority table:

I. What is the extent of accuracy in the age reporting in the 2006 Nigerian population census in two states?
II. Towards which direction (Region) can we find the errors in the age reporting?
III. What is the general over view of the 2006 Nigerian population census in terms of quality of data and age reporting in the two states?
IV. What is the best smoothing technique for the data?

2. Methodology

2.1 Method of analysis

The study utilized the Federal Republic of Nigeria 2006 Population and Housing Census Priority Table Volume IV Population Distribution by Age & Sex (State & Local Government Area). This included the distribution of the population, both in single year and five year intervals.

Whipple Index (W)

$$W = \frac{\sum(P_{25}+P_{30}+P_{35}+\cdots+P_{55}+P_{60})}{\frac{1}{5}\sum(P_{23}+P_{24}+\cdots+P_{61}+P_{62})} \times 100, \qquad (1)$$

where p_x is the enumerated population at age x. The index stated as percentage is used to test the age heaping around the digits '0'& '5'. According to the interpretation of Whipple index, while 100-125 show some degrees of accuracy in age reporting, above 125- 500 show the extent of age heaping and by implication, how bad the age data is in the census report. Nevertheless, figures below 100 show the extent of age shifting from the digits '0'& '5' (Shyrock, Siegel, Associates, 1976). Where extreme deviation [age shifting] was suspected, the age data were separated into '0' and '5' to ascertain the reliability of the result). The inference about age distribution based on this index is as follows: <105 = highly accurate; 105–109.9 = fairly accurate; 110–124.9 = approximate; 125–174.9 = rough; ≥175 = very rough (Spardeshi, 2010).

Myers' Index (M)

Myers' index reflects preferences (or dislike) for each of the ten digits from 0 to 9. Like Whipple's index, it is applicable where age is given in single years. The method derives a blended population which is essentially a weighted sum of the number of persons reporting

ages ending the ten terminal digits- 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The underlying assumption of this method is that if there are no systematic irregularities in the reporting of the age, the blended sum at each terminal digit should be approximately equal to 10 percent of the total blended population. If the sum at any given digit exceeds 10 percent of the total blended population, it indicates over selection of the ages ending in that digit (digit preference). Conversely, a negative deviation (or a sum less than 10 percent of the blended total) indicates under-selection of age ending in that digit (digit avoidance). An overall measure of extent to which there is digit preference and/or avoidance in a census distribution is the index of preference, which is obtained as the absolute sum of deviation for each of the ten terminal digits. The steps in the calculation of Myers' blended index are as follows:

1. Sum of populations ending in each digit over the whole range starting with the lower limit of the range (e.g., 10, 20, 30, 40, ….; 11, 21, 31, ….)
2. Ascertain sum excluding the first population combined in step 1 (e.g., 20, 30, 40, ….; 21, 31, 41, ….)
3. Weight the sums in steps 1 and 2 and add the results to obtain a blended population (e.g., weights 1 and 9 for 0 digit, weights 2 and 8 for 1, etc.)
4. Convert distribution in step 3 into percentages.
5. Take the deviation of each percentage in step 4 from 10.0, which is the expected value for each percentage.
6. A summary index of preference for all terminal digits is derived as one half or total sum of the deviations from 10.0%, each without regard to signs.

2.2 Smoothing Age Errors due to Digital Preference

Method of moving averages

Given five consecutive quinary age groups $p_{-2},p_{-1},p_0,p_1,p_2$ where the $p's$ refer to enumerated populations either in single years or in five year groups, then s_0, the graduated value of p_0, is given by:

$$s_0 = \frac{-p_{-2}+4p_{-1}+10p_0+4p_1-p_2}{16} \quad (2)$$

The use of this formula assumes that the pattern of error can be represented as follows:

$$S_x = p_x + (-1)^x e, \quad (3)$$

where S_x is the true population, p_x is the enumerated population, and e is the error component. If we put $x = 0,1,2,-1,-2$, we have

$$S_0 = p_0 + e, S_2 = p_1 - e, S_2 = p_2 + e, S_{-1} = p_{-1} + e, S_{-2} = p_{-2} + e \quad (4)$$

Carrier- Farrag Method of Graduation

This method consists, essentially, of splitting the total numbers enumerated in to two adjacent five-year age group (quinary age groups) by first grouping the quinary age groups into ten (denary).Let u_i be the population in the quinary age group $5(x+i) to 5(x+i)+4$ and U_i be the population in the denary age group such that:

$$U_i = u_i + u_{i+1} \quad (5)$$

if it is assumed that u_i is a second degree function of i:

$$u_{i+1} = u_i + \Delta u_i \quad (6)$$

Then,

$$u_1 = u_0 + \Delta u_{0,}(i = 0) \quad (7)$$

And

$$U_0 = u_0 + u_1 \quad (8)$$

Therefore,

$$U_0 = 2u_0 + \Delta u_0, \quad (9)$$

Similarly, $u_{2,}u_{3,}U_{2,}U_4$ can be obtained by the manipulation of simple algebraic formula. The key formula in the graduation method are:

$$u_2 = \tfrac{1}{2}U_2 + \tfrac{1}{16}(U_0 - U_4) \quad (10)$$

And

$$u_3 = \tfrac{1}{2}U_2 - \tfrac{1}{16}(U_0 - U_4) \quad (11)$$

2.3 Mean square errors (MSE)

The best procedure would be the one which makes the graduated distributions close to the true ones and minimizes the influence of the errors when they are present. The mean square error can handle this. Where p_x is the enumerated population at age x and $\hat{P}_x$ is the estimated population at age x.

$$\text{MSE} = \frac{\sum(P_x - \hat{P}_x)^2}{N-1} \quad (12)$$

3. Data Analysis

3.1 *Whipple's Index*

Table 1: Myers' Index Adamawa (M_{Ad})

	10-19	20-29	30-39	40-49	50-59	60-69	70-79
0	105308	126597	119013	87259	59764	38353	26602
1	49125	35684	16715	9792	5388	2822	1427
2	83510	54830	34173	19217	10417	4975	2780
3	67101	40335	18668	11579	6578	3357	1604
4	65924	33624	15171	7551	5483	2451	1214
5	84513	105769	73646	48454	21606	14925	8770
6	58455	34839	18868	10867	6942	2296	1542
7	54791	42524	21126	11266	6263	4132	1571
8	83797	51657	26258	14753	6340	3372	2127
9	46879	26441	13573	7619	3054	1563	845

Sum of 10-79 ages	Ages of 10-79 coefficients	Product	Sum 20-79	Ages 20-79 coefficients	Product	Blended sum	Percentage distribution	Deviations
562896	1	562896	457588	9	4118292	4681188	26.6	16.6
120953	2	241906	71828	8	574624	816530	4.6	-6.4
209902	3	629706	126392	7	884744	1514450	8.6	-1.4
149222	4	596888	82121	6	492726	1089614	6.3	-3.7
131418	5	657090	65494	5	327470	984560	5.6	-4.6
357683	6	2146098	273170	4	1092680	3238778	18.4	8.4
133809	7	936663	75354	3	226062	1162725	6.6	-3.4
141673	8	1133384	86882	2	173764	1307148	7.4	-2.6
188304	9	1694736	104507	1	104507	1799243	10.2	0.2
99974	10	999740	53095	0	0	999740	5.7	-4.3
Total						17593976	100	50.4

Whipple's index Adamawa, $W_{Ad} = \frac{553864}{1/5(1095905)} \times 100 = 252.0.$

Whipple's index Anambra, $W_{An} = \frac{705230}{1/5(1801691)} \times 100 = 195.71.$

3.2 Myers' Index

Table 2: Myers' Index Anambra (M_{AN})

	10-19	20-29	30-39	40-49	50-59	60-69	70-79
0	137089	158984	142450	120660	84713	55854	29953
1	63972	60648	27427	19214	13095	6994	3663
2	120902	80118	58529	38009	24673	10995	6491
3	80134	63466	26101	19892	11607	6119	2927
4	105153	69834	28355	14378	14089	5724	2706
5	109109	108672	88322	72739	31820	23325	12311
6	108161	67441	35618	20924	14909	4852	3415
7	73400	54601	25026	17562	9364	5582	2159
8	141930	93058	53797	37096	17020	9485	4606
9	69183	49040	32076	22928	10346	5128	2264

Sum of 10-79 ages	Ages of 10-79 coefficients	Product	Sum 20-79	Ages 20-79 coefficients	Product	Blended sum	Percentage distribution	Deviations
729703	1	729703	592614	9	5333526	6063229	22.5	12.5
195013	2	390026	131041	8	1048328	1438354	5.4	-4.6
339717	3	1019151	218815	7	1531705	2540856	9.5	-0.5
210246	4	840984	130112	6	780672	1621656	6.0	-4.0
240239	5	201195	135086	5	675430	1876625	7.0	-3.0
446298	6	2677788	337089	4	1348756	4026544	15.0	15.0
255320	7	1787240	147159	3	441477	2228717	8.3	-1.3
187694	8	1501552	114294	2	228588	1730140	6.4	-3.6
356992	9	3212928	215062	1	215062	3427990	12.8	2.8
190965	10	1909650	121782	0	0	1909650	7.1	-2.9
Total						26863761	100	40.2

3.3 Smoothing Age Errors due to Digital Preference

Table 3: Moving Average Method

Age group	Anambra Enumerated	Graduated	Adamawa enumerated	Graduated
0-4	473248		560570	
5-9	505069		487311	
10-14	507250		370968	
15-19	501780	512100.2	328435	323997
20-24	433045	493818.7	291070	293447
25-29	372812	43919.7	261230	256731.7
30-34	282862	365945.6	203240	204027.8
35-39	234839	288376.6	153471	158396.4
40-44	212153	236524.3	135398	128036.8
45-49	171249	207177.7	92959	101493.7
50-54	148177	11076.2	87580	77318.8
55-59	83459	137672.7	44205	55059.7
60-64	85686	96901.3	51958	42521.6
65-69	48372	74391.7	26288	34135
70-74	45740	56365.8	33627	26869.7
75-79	24755	39902.2	14855	19739.9
80-84	24113		18969	
85^{+}	23219		16816	
MSE		48964.9		13889

Table 4: Carrier-Farraq Method

Age group	Anambra Enumerated	Denary $U_i = u_i + u_{i+1}$	Graduated u_i	Adamawa enumerated	Denary $U_i = u_i + u_{i+1}$	Graduated u_i
0-4	473248			560570		
5-9	505069	978317		487311	1047881	
10-14	507250		515293.8	370968		380675.3
15-19	501780	1009030	493736.3	328435	699403	318727.7
20-24	433045		433636.6	291070		297568.3
25-29	372812	805857	372220.4	261230	552300	254731.8
30-34	282862		285253.9	203240		198601.9
35-39	234839	517701	232447.1	153471	356711	158109.1
40-44	212153		209580.1	135398		128236.4
45-49	171249	383402	173821.9	92959	228357	100120.6
50-54	148177		131402	87580		75274.44
55-59	83459	231636	100234	44205	131785	56510.56
60-64	85686		77100.31	51958		44329.44
65-69	48372	134058	56957.69	26288	78246	33916.56
70-74	45740		40667.88	33627		26894.81
75-79	24755	70495	29827.13	14855	48482	21587.19
80-84	24113			18969		
85^+	23219	47332		16816	35785	
MSE		3618.3			3641.8	

4. Interpretation of Result

The Whipple's index for Adamawa and Anambra are 252, 195.7 respectively, the results show that the qualities of age data in the two states are very rough as a result of age misreporting. Myers' index shows that enumeration of ages ending with digits 0, 5 in the two states are over stated or preferred while others are under stated or avoided. In general, there was age heaping at ages with terminal digits '0' and '5', indicating a preference in reporting such ages. 50.4% and 40.2% of the population reported ages with an incorrect final digit in Adamawa and Anambra respectively and the smoothing method for both states is the Carrier Farrag method because its mean square error is smaller than that of the moving average method.

5. Conclusion/Recommendation

In general, the results indicated by Whipple's and Myers' indices show great affinity for age digit 0 and 5, with an incorrect final digit of 50.4% and 40.2% of the population reported ages in Adamawa and Anambra respectively. The difference in the incorrect final digits could be because different set-ups have different social values attached to age (Spardeshi, 2010). Since age reporting is one of the essential parts of census data which inform other higher degree research ventures (Economic and Social Commission for Western Asia, 2013). then the data should be smoothing by a method that suits it so as to reduce the digital errors to its barest minimum and further used for analysis.

References

Akingbade, A. (2004). Geospatial Data Infrastructure for Nigeria: Foundation for Further development, *GIM International Magazine*, 9 (18), 12-15.

Aluko, S.A. (1965). How Many Nigerians? An Analysis of Nigeria's Census Problems, 1901-1963, *Journal of Modern African Studies,* 2, 371-392.

Economic and Social Commission for Western Asia (2013). A study of Age Reporting in Some Arab Censuses of Population

Ekanem, I.I. (1972). *The 1963 Nigerian Census: A Critical Appraisal*, Ethiope Publishing Corporation, Benin.

Inekwe, M. (2014). Hundred Years of Revenue Sharing in Nigeria: The Need for a Rethink. *Journal of Good Governance and Sustainable Development in Africa*, 2(2), 1-8.

Makinwa, P.K. (1985). Population Data: The Importance of Census, Sample Survey and Vital Registration System, *Population Education Monograph 15, Nigerian Educational Research Council,* Lagos, Nigeria.

Mathews, M.P. (2002). *Nigeria: Current Issues and Historical Background*, Nova Publishers, Hauppauge, New York.

Odewumi, S. (2000). *Problems of Census in Nigeria,* In: Odumosu, Atere and Adewunmi (eds), *Social Problems and Planning Studies in Nigeria*, Lagos Centre for Planning Studies, Lagos State University.

Okafor, R., Adeleke, I. and Opara, C.A. (2007). An Appraisal of the Conduct and Provisional Results of the Nigerian Population and Housing Census of 2006, In: *Proc. JSMASA-SE,* 2007, 2199-2205.

Okolo, A. (1999). The Nigerian Census: Problem and Prospects. *The American Statistician*, 53 (4), 321-325

Olusanya, P.O. (1989). *Population and Development Planning in Nigeria*, In: Tamuno and Atanda (eds.), *Nigeria Since Independence, The First 25 Years Government and Public Policy*, Heinemann, Ibadan.

Onyekakeyah, L. (2007). *Paradox of Population distribution in Nigeria*, In: The Guardian (Lagos), January, 23: 65.

Pardeshi, G.A. (2010). Age Heaping and Accuracy of Age Data Collected During a Community Survey in the Yavatmal District, Maharashtra, *Indian Journal of Community Medicine,* 30 (3), 391 – 395.

Shyrock, H.S. Siegel, J.S. and Associates. (1976). *The Methods and Materials of Demography*, Condensed edition by Edward G. Stockwell, Academic Press, New York.

United Nations (2008). *Principles and Recommendations for Population and Housing Censuses,* UN Publications, New York.