AGE HEAPING IN THE PHILIPPINE CENSUS

Edward G. Stockwelll and John P. Dixon*

INTRODUCTION:

Social scientists have been aware that when a person is asked to supply information on age, either his own or that of some other person known to him, the information that is forthcoming is often far from accurate. There are a number of reasons why this is so : (1) many people tend to round off to the nearest age ending in O or 5 (a tendency that may reflect either ignorance of exact age, of some sort of vague tendency toward orderliness); (2) a natural inclination to overstate when nearing some socially significant age (such as ages 21 and 65 in the United States); (3) a tendency to understate as middle age approches in an effort to "preserve ones' youth", and (4) the converse tendency to overstate age at the later years, due likely to the traditional veneration of old age.

The most common source of error in age reporting, however, appears to be the one cited initially above — the tendency to overselect ages ending in particular digits, such as 0 or 5, while avoiding ages ending in other digits, thus causing successive patterns of "heaping" throughout an age distribution. Such a pattern was very pronounced in both the 1948 and 1960 censuses of the Philippenes. In 1948, for example, 265,561 persons reported their age at 38 years.[1] Twelve years later, in the 1960 census, the corresponding cohort 50 years of age numbered 313,636 persons indicating an intercensal increase of 18

*The authors are Associate Professor and Graduate Research Assistant, respectively, in the Department of Rural Sociology at the University of Connecticut

1 Single year of age statistics from the 1948 and 1960 censuses of the Philippines were obtained, respectively, from the 1955 and 1962 *Demographic Yearbooks* of the United Nations (See Table 11, page 326 of the 1955 *Yearbook for* 1948 data, and Table 6, page 227 of the 1962 *Yearbook* for 1960 data).

percent. To take another example, 323,012 persons reported their age at 40 years in 1948; but in 1960 there were only 128,935 persons enumerated at age 52, which would indicate a 60 percent decrease in the size of this cohort during the twelve years between censuses!

The experiences of the two cohorts just described cannot be explained in terms of either migration or mortality differences, but instead are part of what may be called a **regular pattern of irregularity** characterizing the single year of age data in the 1948 and 1960 Philippine censuses. In the sections to follow the nature of this "heaping" pattern is examined more closely. Some attention is also given to age, sex, and urban-rural variations with regard to the extent of bias in age reporting.

THE MEASUREMENT OF AGE "HEAPING"

By age "heaping" is meant an overrepresentation of persons reporting ages ending in certain preferred digits and a corresponding underrepresentation of persons reporting ages ending in other less preferred digits. The most common technique for determining the extent to which such "heaping" is taking place, and at what digits it is taking place, is the method of blending" an age distribution that has been described by Robert J. Myers.[2] On the assumption of all else being equal, the number of persons reporting ages ending in the same digit should be approximately equal to 10 percent of the total reporting age. The problem, however, is that all else is not equal. Because of mortality, the lower order digits would naturally be over represented relative to the higher order digits. Thus, one would normally expect the sum of those with ages ending in 0 to be larger than the sum with ages ending in 1, and so forth, with the sum being small-

2 Robert J. Myers, "Errors and Bias in the Reporting of Ages in Census Data," *Transactions of the Actuarial Society of America*, 41:104 (October, 1940), pages 394-415.

est for ages ending in 9. This bias could be overcome, however, by making ten different summations starting with each digit in turn, and then averaging the results for each digit. This is essentially what Myers does, except that he introduces a short-cut weighting procedure as an alternative to making ten different sets of summations.

Without going into too much detail (Table 1 is presented as an illustrative example of this procedure), application of the Myers technique leads to a "blended" population in which the sum of the number at each terminal digit should, in the absence" of any "heaping at preferred digits, be equal to 10 percent of the total "blended" population. If the sum for any digit is in excess of 10 percent it indicates overselection of ages ending in that digit (**digit preference**). Conversely, a negative deviation, or a sum that is less than 10 percent of the "blended" total, indicates underselection of ages ending in that digit (**digit avoidance**). An overall measure of the extent to which there is digit preference and/or avoidance in a census age distribution in the **Index of Preference**, which is obtained as one-half of the absolute sum of the deviation from 10 percent for each of the ten terminal digits. This index may be interpreted as the proportion of the population that would have to be redistributed in order for the "blended" population to meet the above conditions (i.e., in order for the sum at each digit to be equal to 10 percent of the total "blended" sum).

(See Table 1)

Applying this method to the single year of age data in the 1960 census of the Philippines yielded an Index of Preference of 10.1 (See Table 1). This is a fairly large index, and indicates a substantial amount of digit preference/avoidance in that

census. Some idea of the magnitude of the bias in the Philippine age data may be obtained by comparing Indexes of Preferences for other national censuses taken in 1960:[3]

Sweden	0.4
Norway	0.5
Netherlands	0.6
Denmark	0.6
Switzerland	0.8
United States	0.8
Finland	1.3
Portugal	1.9
Thailand	2.2
Hungary	2.3
Panama	4.9
Jamaica	5.4
Seychelles	6.7
Cyprus	6.9
Philippines	10.1
Mexico	13.3
Ghana	15.7
Turkey	22.2
Morocco	38.3

Although it certainly does not have the largest Index of Preference, the above data clearly reveal that age reporting in the 1960 Philippine census was characterized by a great deal of "heaping" at preferred digits. It is encouraging to note, however, that the amount of bias in the Philippine age data is not as pronounced as it was in the previous census of 1948 (see Table 2). The 1948 Index of Preference was 13.8, or approximately 37 percent greater than the 10.1 computed from the 1960 age data.

3 The Indexes of Preference for these countries (as well as for 31 others) were computed from data contained in the 1962 and 1963 *Demographic Yearbooks* of the United Nations. These data are available, in unpublished form, in the Department of Rural Sociology at the University of Connecticut.

(See Table 2)

One of the possible explanations for observed differences with regard to the accuracy of age reporting is the level of literacy characterizing various population groups. Thus, part of the improvement in the Philippine age statistic between 1948 and 1960 may be explained in terms of the overall increase in the literacy status of the population age 10 and over was recorded as literate; but by 1960 this figure had risen to 72.0 percent.[4]

Closer examination of Table 2 reveals that the greatest amount of "heaping" in both the 1948 and 1960 censuses at ages in digits 0 and 5, particularly the former. Ages ending in digit 8 also show a slight amount of overselection, while the greatest avoidance is seen to characterize ages ending in digit 1. However, with two minor exceptions, the deviation of the "blended" terminal digit sums form 10 percent are smaller for 1960 than for 1948. The two exceptions were digits 2 and 6, where there was no change between 1948 and 1960.

SEX VARIATIONS IN AGE HEAPING:

The data precented in Table 3 show that males and females both exhibit the same general pattern of digit preference. For both sexes the greatest "heaping" occurs at digit 0, with somewhat less preference given to digits 5 and 8; and the greatest avoidance characterizes digit 1. The only variation from the general pattern is associated with ages ending in 2 which tends to show a slight preference among males.

---

4 Literate data from the 1948 and 1960 censuses of the Philippines were obtained, respectively, from the 1955 and 1963 *Demographic Yearbooks* of the United Nations (See Table 13, page 458 of the 1955 *Yearbook for* 1948 data, and Table 12, page 360 of the 1963 Yearbook for 1960 data).

These data also reveal that bias in age reporting is slightly more pronounced among females. With the exception of digits 2 and 9 (where the amount of deviation from 10 percent is the same for both sexes), and digit 6 (where male avoidance is slightly stronger than that of females), the deviations of the terminal digit sums from 10 percent were greater for females. As a result of these differences, the female Index of Preference (10.6) is approximately 10 percent greater than the corresponding male Index of Preference (9.7).

(See Table 3)

Once again, differences in literacy status may be cited as one of the factors behind the sex differential with regard to the accuracy of age reporting. In this instance, nearly three-fourths of the Philippine males age 10 and over were recorded as literate in 1960 (74.2 percent) as compared to only 69.5 percent for females at the same ages.[5]

AGE AND DIGIT PREFERENCE:

While the "blending" method may provide a convenient means of determining the overall of preference on or avoidance of specific terminal digits , it does not take account of the fact that "heaping" at any given digit may be due to a particularly strong preference for a single age, and may not really reflect any preference for the particular terminal digit at all. In the1960 census of the United States, for example, the slight clustering observed at digit 9 was due entirely to a pronounced overstatement at age 59, and reflected the tendency among many older person to select the "convenient" year 1900 in response to the question on date of birth.[6]

---

5 *Ibid.*, page 360 of the 1963 *Demographic Yearbook.*

6 Malvin Zelnick, "Errors in the 1960 Census Enumeration of Native Whites," *Journal of the America Statistical Association,* 59:306 (June, 1964), pages 437-459.

One method of determining the extent of "heaping" at specific ages is to compute, for each age, a ratio of the number reporting that age to the arithmetic mean of the five ages immediately below and immediately above the age in question.[7] These age ratios should be approximately equal to one if there is no pronounced preference or avoidance for the age in question. Ratios greater than one indicate preference for a given age while ratios of less than one indicate avoidance.[8]

Such age ratios were computed from the 1960 Philippine census data for ages 15 to 84 years. They are presented here in Table 4. Inspection of these data suggests that where there is digit preference (notably at digits 0, 5 and 8), it tends to characterize all of the ages involved. At the same time, however, it is readily apparent that misstatements of age become more pronounced as age increases. This can be seen most clearly in the next to the last column of Table 4 which presents the mean deviation from one for each successive ten-year age group. In spite of occasional exceptions with regard to particular digits, these mean deviations increase consistently with age from a low of .077 for the group age 15 to 24 to a high of .815 for persons 75 to 84 years old.

(See Table 4)

Inspection of the final column in Table 4 clearly shows that literacy again provides a partial explanation for these age variations in digit preference. Reversing the pattern of the mean deviations discussed above, the proportion of persons in each ten-year age group that was reported as literate in 1960 decreases consistently from about 85 persent at ages 15 to 24 to only 31 percent among persons 65 years of age and over.

---

7 To illustrate, the ratio for age 20 would be computed as the number of persons reporting age 20 divided by one-tenth of the sum reporting ages 15, 16, 17, 18, 19, 21, 22, 23, 24, and 25.

8 The rationale as well as an evaluation o fthis procedure has been discussed by Zelnick, *op. cit.* See also: Melvin Zelnick, "Age Heaping in the United States Censuses, 1880-1950," *Milbank Memorial Fund Quarterly*, 39:3 (July, 1961), pages 540-573.

URBAN-RURAL VARIATIONS IN AGE HEAPING:

Unfortunately, single year of age statistics for the total population of the Philippines by urban-rural residence were not available to the author at the time of this writing. Such data were available for a number of the several provinces, however; and by grouping these provinces according to some sort of measure of urbanization it is possible to obtain an approximate indication of the nature of any urban-rural variations in age reporting. In the present paper, such a grouping was based on a size of place criterion. On the one hand, Manila and Rizal were selected as "urban" because both contained municipalities having more than 250,000 inhabitants. On the other hand, Abra, Antique, Bataan, Nueva Vizcaya, and Romblon were designated as "rural" on the basis that none of them contained municipalities having as many as 25,000 inhabitants. The results of applying Myers' "blending" technique to the single year of age data for these seven provinces are presented in Table 5.[9]

(See Table 5)

Although the general patterns of digit preference/avoidance are similar, age "heaping" would appear to be more pronounced in the rural areas of the Philippines than in urban areas. This is suggested by the fact that the Index of Preference for Manila and Rizal (6.1 and 6.3 respectively) were considerably lower than the five "rural' provinces. Bataan, in fact, was the only "rural" province to be characterized by an Index of Preference that was lower than the index for total Philippine population. The reason for Bataan's slightly better performance relative to other four rural provinces lies in its notably higher

9 These indexes (as well as the urban-rural designation) were derived from data presented in the census reports issued for the individual provinces. See: Philippine Bureau of the Census and Statistics, *Census of the Population*, 1960: *Population and Housing*, Volume I (Manila, 1960). See Tables 1 and 3 of the indivldual province reports.

rate of literacy. In 1960, 83 percent of the population age 10 and over in Bataan was reported as literate, whereas the corresponding proportion of literate persons was only 70 percent or less in the four remaining "rural" provinces.[10] By contrast, the porportion literate was apporximately 90 percent in both Manila (91.5 percent) and Rizal (89.6 percent).

DISCUSSION

The patterns of digit preference/avoidance described in the preceding sections can be explained in terms of model described earlier by Stanley H. Turner.[11] In developing this model Turner assumes that people will generally give accurate reports if they can in response to questions seeking numerical information. People who do not know the correct response, however, will frequently report an estimate; and Turner suggests that these estimates will assume a more or less predictable pattern:

> To put it simply: the way we count influences the way we estimate. That is when a person estimates. he should do so in convenient units provided for him by the number system. Specifically, he should tend to over-report digits which are multiples of the divisors of the number system and under-report digits which are not multiples of the divisors of the base of the number system.[12]

Because we use a base ten number system, Turner continues, "heaping" should occur at multiples of 10, 5 and 2. Numbers divisible by 10 should receive the greatest "heaping"; numbers divisible by 5 but not by 10 should rank second; and numbers divisible by 2 but not by 5 or 10 should rank third.

---

10 *Ibid.* See Table 8 of the individual porvince reports for statistics on literacy.

11 Stanley H. Turner, "Patterns of Heaping in the Reporting of Numerical Data,"*Proceedings of the Social Statistics Section* (Washington: American Statistical Association, 1958), pages 248-251.

12 *Ibid.*, page 248.

Thus, one would expect estimated responses to requests for age information to show greatest preference for ages ending in 0, second preference for ages ending in 5, and third preference for ages ending in digits 2, 4, 6,, and 8. The least preferred ages, therefore, would be those ending in one of the remaining digits: 1, 3, 7, and 9.

The expected order of digit preference or avoidance can be specified even further. Since the preferential digits 4 and 6 are right next to the even more preferred digit 5, one would expect them to attract less heaping than the preferred digits 2 and 8. Similarly, since the digits 1 and 9 are on either side of the most preferred digit, one would expect them to be under-selected more so than digits 3 and 7.

Given the preceding, one would expect the following rank order of digit preference and avoidance:

0	(Most preferred)
5	
2,8	
4,6	
3,7	
1,9	(Most avoided)

In the 1960 census of the Philippines the following rank order of digit preference was observed:

Rank	Terminal Digit	Deviation
1	0	+6.1
2	5	+2.7
3	8	+1.3
4	2	0.0
5	7	—1.1
6	6	—1.4
7.5	3	—1.5
7.5	9	—1.5
9	4	—1.7
10	1	—2.8

Although seven of the ten digits conform to what would be expected from Turner's hypothesis, three do not. The digit which deviates most from the expected pattern is four. However, the differences in the deviations for the digits ranking from fifth to ninth (those containing the deviant cases) are not very large, and do not justify a complete rejection of Turner's explanation of the pattern of "heaping" in the reporting of numerical data.

CONCLUSION

Illiteracy is very high in the Philippines as compared to the more industrially advanced nations of the world. According to the 1960 census, for example, less than three-fourths of the population age 15 and over was classified as literate. Given this situation, it is to be expected that a fairly large segment of the population will not know their exact age or date of birth, and when asked to volunteer such information they will estimate it. It has been clearly demonstrated in the preceding sections that such estimates will not be random, but will follow a fairly regular pattern—a pattern that becomes most pronounced where levels of illiteracy are highest (i.e., among older persons, females, and rural dwellers). Moreover, given a knowledge of the base of the existing system of counting, the data presented here for the Philippines clearly suggest that this pattern will be predictable.

TABLE 1

APPLICATION OF MYERS' METHOD OF "BLENDING" TO SINGLE YEAR OF AGE DATA FROM THE 1960 CENSUS OF THE PHILIPPINES.

Terminal Digits	NUMBER AT SPECIFIED AGES								SUM FOR AGES	
	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	10-89	20-89
0	841,356	565,801	535,924	434,156	313,636	275,436	136,771	73,741	3176,821	2,335,465
1	581,400	494,895	222,086	126,632	78,534	31,299	13,000	5,532	1,553,378	971,978
2	796,786	515,823	318,481	217,881	128,935	49,634	28,017	9,331	2,064,888	1,268,102
3	619,293	456,892	246,260	169,167	93,279	40,154	16,662	5,653	1,647,360	1,028,067
4	596,592	425,212	233,700	151,142	95,715	34,381	14,490	5,089	1,556,321	959,729
5	565,714	522,203	401,936	319,118	163,093	102,440	50,558	18,604	2,143,666	1,577,952
6	566,942	358,549	242,659	160,329	87,754	26,445	15,010	4,803	1,462,491	895,549
7	538,891	376,221	242,462	160,855	71,828	35,311	11,878	5,617	1,443,063	904,172
8	651,318	395,766	316,210	237,287	93,049	40,711	23,353	4,388	1,762,082	1,110,764
9	491,441	300,610	225,207	155,094	72,206	20,921	9,212	4,000	1,278,691	787,250
Sum									18,088,761	11,839,028

Terminal Digits	AGES 20-89 Sum	Coef-ficient	Product	AGES 10-89 Sum	Coef-ficient	Product	"BLENDED" POPULATIONS Sum	Percent Distri-bution	Deviation From 10%
0	3,176,821	1	3176,821	2,335,465	9	21,019,185	24,196,006	16.1	+6.1
1	1,553,378	2	3,106,756	971,978	8	7,775,824	10,882,580	7.2	—2.8
2	2,064,888	3	6,194,664	1,268,102	7	8,876,714	15,071,378	10.0	0.0
3	1,647,360	4	6,589,440	1,028,067	6	6,168,402	12,757,842	8.5	—1.5
4	1,556,321	5	7,781,605	959,729	5	4,798,645	12,580,250	8.3	—1.7
5	2,143,666	6	12,861,996	1,577,952	4	6,168,402	19,173,804	12.7	+2.7
6	1,462,491	7	10,237,437	885,549	3	2,686,647	12,924,084	8.6	—1.4
7	1,443,063	8	11,544,504	904,172	2	1,808,344	13,352,848	8.9	—1.1
8	1,762,082	9	15,858,738	1,110,764	1	1,110,764	16,969,502	11.3	+1.3
9	1,278,691	10	12,786,910	787,250	0	0	2,786,910	8.5	—1.5
							Sum of Deviations		=20.1
							Index of Preference		=10.1

TABLE 2

RESULTS OF APPLYING MYER'S METHOD OF "BLENDING" TO AGE DATA FROM THE 1948 AND 1960 PHILIPPINE CENSUSES.

Terminal digits	Digit Deviations 1948	1960
0	+6.8	+6.1
1	−3.6	−2.8
2	0.0	0.0
3	−2.1	−1.5
4	−2.0	−1.7
5	+4.3	+2.7
7	−1.4	−1.1
8	+2.7	+1.3
9	−3.2	−1.5
Sum	27.6	20.1
Index of Preference	13.8	10.1

TABLE 3

RESULTS OF APPLYING MYER'S METHOD OF "BLENDING" TO AGE DATA FOR MALES AND FEMALES: PHILIPPINES, 1960.

Terminal Digits	Digit Deviations		
	Total	Males	Females
0	+6.1	+5.8	+6.3
1	—2.8	—2.6	—2.9
2	0.0	+0.2	—0.2
3	—1.5	—1.5	—1.6
4	—1.7	—1.6	—1.7
5	+2.7	+2.6	+2.8
6	—1.4	—1.5	—1.4
7	—1.1	—1.0	—1.3
8	+1.3	+1.1	+1.4
9	—1.5	—1.5	—1.5
Sum	20.1	19.4	21.1
Index of Preference	10.1	9.7	10.6

TABLE 4

AGE RATIOS INDICATING PREFERENCE FOR SPECIFIC AGES: PHILIPPINES, 1960.

Ages	Terminal Digits 5	6	7	8	9	0	1	2	3	4	Mean Deviation From One	Percent Literate
15-24	.905	.961	.919	1.203	.914	1.082	.972	1.057	.952	.950	.077	84.9
25-34	1.180	.845	.929	1.052	.827	1.588	651	.996	.767	.766	.204	84.9
35-44	1.332	.877	.877	1.245	.888	1.800	.511	.945	.720	.685	.287	66.6
45-54	1.501	.831	.833	1.372	.892	1.970	.492	885	.655	.761	.349	59.8
55-64	1.245	.795	.659	.949	.766	3.692	367	.634	.528	.495	.574	48.0
65-74	1.482	524	.717	.897	.466	3.923	.338	.789	.470	.445	.676	) 31.2
75-84	1.478	.609	.483	1.101	.429	4.782	.313	.571	.354	.361	.815	)

TABLE 5

RESULTS OF APPLYING MYERS' METHOD OF "BLENDING" TO AGE DATA FOR TWO URBAN AND FIVE RURAL PROVINCES: PHILIPPINES, 1960.

Terminal Digits	Total	"Urban"		"Rural"				
		Manila	Rizal	Abra	Antique	Bataan	Nueva Vizcaya	Romblon
0	+6.1	+3.8	+3.8	+8.4	+7.7	+4.8	+6.7	+5.6
1	−2.8	−1.7	−1.9	−3.4	−3.7	−2.6	−3.0	−3.1
2	0.0	+0.4	+0.5	−0.2	−0.7	+0.8	+0.3	+0.3
3	−1.5	−0.7	−0.8	−1.9	−2.0	−0.6	−1.4	−1.3
4	−1.7	−1.1	−1.1	−2.5	−2.7	−1.0	−1.8	−1.8
5	+2.7	+0.5	+0.9	+3.6	+4.8	+2.3	+2.8	+3.1
6	−1.4	−1.5	−1.2	−1.7	−1.7	−1.3	−1.5	−1.2
7	−1.1	−0.9	−0.8	−1.7	−1.4	−0.9	−1.3	−1.0
8	+1.3	+1.4	+1.1	+1.3	+2.1	+0.2	+1.1	+1.2
9	−1.5	−0.1	−0.4	−2.0	−2.3	−1.8	−1.7	−1.7
Sum		12.1	12.5	26.7	28.1	16.2	21.6	20.3
Index of Preference		6.1	6.3	13.4	14.1	8.1	10.8	10.2