ON THE PUNCHED CARD METHOD IN SMOOTHING FOR AGE BIAS IN CENSUS RETURNS

By K. C. PODDER

Indian Statistical Institute, Calcutta

INTRODUCTION

The problem of smoothing for age bias arose in connection with the project, carried out by the Indian Statistical Institute, of preparing the 1951 census Tables of India on the basis of a 2% (two per cent) sample of the original individual census slips which had been preserved (at the instance of the late Mr. M. W. M. Yeatts, Census Commissioner) and usually known as *Y*-slips.

The preference of individuals for certain digits, e.g., 0 and 5, in age returns is well known and was also observed in the present case. In the initial stages of compilation, standard age groupings in terms of the age originally returned by investigator designated as "sorter's or initial age-groups" were adopted for sorting. From these, the "compiler's or final age-groups" were to be obtained after adjustment for bias.[1]

METHOD FOLLOWED AND OPERATIONS INVOLVED

The formulae used for the smoothing of age bias are given in column 4 of Table 1.

The operation, it will be seen, involves taking a certain percentage of the total number of cards of one group and adding to it a certain other percentage of the next group to produce a new set of Summary Cards which was essential for smoothing of age bias.[2] This could be done by using the punch-card Multiplier unit. In that case, two groups of cards of initial age-groups had to be compared by the Collator (and missing cells in either case to be filled up by insertion of dummy cards to bring both the sets to exact correspondence) and then the figures of the first group had to be transferred to the second group by the Reproducing Punch. To arrive at the total of the final age-group would involve, among other operations, one further run through the Multiplier type 601.

---

[1] To distinguish between the two groups, an identification code 'I' was gang-punched on all the initial age-groups excepting F and T (Table 1).

[2] Mandeville (1940) during a discussion on the paper by H. O. Hartley (1940) showed how "a Hollerith Tabulator could be used to multiply totals, and so weight results derived from additions of quantities recorded on cards". He also mentioned how advantage could be taken "of the capacity of the tabulator to transfer totals, and the fact that when totals are transferred from a counter back into the same counter the total in the counter is doubled, to double or quadruple the totals in the counter".

TABLE 1. METHOD OF SMOOTHING FOR AGE BIAS

sorter's or initial age-group		number of persons in sample (inflation to 100%)	formulae	compiler's or final age-group		estimated population after smoothing
group range in years	code			group range	code	
(1)	(2.1)	(2.2)	(3)	(4)	(5)	(6)	(7)
0	01	A	8,689.0	A	0	00	8,689.0
1	02	B	11,887.0	B	1	01	11,887.0
2	03	C	11,993.0	C	2	02	11,993.0
3	04	D	14,229.0	D	3	03	14,229.0
4	05	E	14,764.0	E	4	04	14,764.0
5—6	06	F	28,095.0	$F + .6G$	5—9	05	65,139.0
7—11	07	G	61,740.0	$.4G + .6H$	10—14	10	50,412.0
12—16	08	H	42,860.0	$.4H + .6I$	15—19	15	30,477.2
17—21	09	I	22,222.0	$.4I + .6J$	20—24	20	15,450.4
22—26	10	J	10,936.0	$.4J + .6K$	25—29	25	7,013.2
27—31	11	K	4,398.0	$.4K + .6L$	30—34	30	2,941.2
32—36	12	L	1,970.0	$.4L + .6M$	35—39	35	1,391.6
37—41	13	M	1,006.0	$.4M + .6N$	40—44	40	986.8
42—46	14	N	974.0	$.4N + .6O$	45—49	45	657.8
47—51	15	O	447.0	$.4O + .6P$	50—54	50	358.8
52—56	16	P	300.0	$.4P + .6Q$	55—59	55	262.2
57—61	17	Q	237.0	$.4Q + .6R$	60—64	60	94.8
62—66	18	R	—	$.4R + .6S$	65—69	65	—
67—71	19	S	—	$.4S + T$	70 & over	70	60.0
72 & over	20	T	60.0				
total			236707.0				236707.0

The considerations of the time involved in the above procedure led to the adoption of the following method.

The Hollerith Senior Rolling Total Tabulator, it is known, can be worked in conjunction with the Reproducing Punch in transmitting the accumulated totals in some or all of its counters to the latter to produce a fresh card according to design at every break of control.[2]

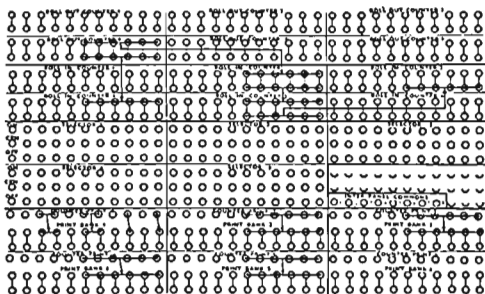[2] For a description of the Hollerith equipments, see Harley (1946).

The first pack of sorted cards consisting of two age-groups F and G (Table 1) were put to the Tabulator to add all classificatory informations in the first counter from the very first card of a particular group for group indication. The figures of F would be added to the 2nd counter one digit to the left so as to treat the first digit of the counter as one place of decimal which was in this case zero and G would be added in the fourth counter by the use of the distributor of the Hollerith Tabulator. Thereafter "advantage was taken of the capacity of the tabulator to transfer totals, and the fact that when totals are transferred from a counter back into the same counter the total in the counter is doubled, to double or quadruple the totals in the counter." Thus, the totals for G were entered in the fourth counter of the Tabulator and thereafter doubled by transferring themselves into themselves, i.e., the fourth counter would show $2G$ in the first operation, and at the same time it would transmit its contents to the fifth counter. In the next operation ("cycle") the contents of the fourth counter would get quadrupled. Simultaneous with the second operation, the fourth counter would also transfer its contents, i.e., $2G$, to the third counter which was blank uptil then, and also to the second counter to be added to the figure F. In the third cycle, the contents of the fourth counter would be transferred to the third and second counters, and we would then get $.4G$ in the fourth counter, $.6G$ in the third counter and $F + .6G$ in the second counter (considering the first digit of the counter as the first place of decimal).

To arrive at the final result, summary cards with all relevant informations for identification from the first counter and $.4G$ from the fourth counter and $F + .6G$ from the second counter would be punched. While summary punching code No. 05 should be gang-punched for identifying the final age group 5-9. The $.4G$ must be punched in the summary card in the same field in which the number of persons of the initial age-group was punched in the detail cards with an extra column in the right hand side to be treated as the first place of decimal (see Fig. 1).

<div align="center">SUMMARY CARD</div>

<div align="center">Fig. 1</div>

The summary cards would contain $4G$ and the total of the final age-group with its code. All the information punched in the summary cards but excluding the group code, $.6G$ and total G should be printed so as to get a check for the

summary cards and also to facilitate the scrutiny of the performances of both the machines.

The following cycle chart and Control Panel wiring would explain the actual machine operations.

CYCLE CHART

cycle	counters				
	1	2	3	4	5
	group indication i.e. all classificatory items	total in the compiler's or final age-group i.e., F ÷ .6G etc.	sorter's age-group		
			.6	.4	total
(1)	(2)	(3)	(4)	(5)	(6)
5				R A	A
4		A	A	R A	
3		A	A	R	
2	print & summary punch	print and summary punch	print	print & summary punch	print
1	zeroise	zeroise	zeroise	zeroise	zeroise

(For control panel wiring see Fig. 2)

PANEL 1

Fig. 2.1

PANEL 2

Fig. 2.2

PANEL 3

Fig. 2.3

After producing the final age-group with code 05, only one plug had to be inserted in the unit place in the second counter to account for one place of decimal punched in the summary card. Then the final age-group code 05 and the initial age-group *H* were amalgamated and sorted as desired. The tabulator would then register *AG* from the summary cards in the second counter where *F* was added from the original cards for the previous table and *H* had to be distributed and calculations made as

was the case with G and then added to $.4G$ to get the final age-group 10–14, for which code 10 had to be gang-punched (Table 2). The same procedure was used in the

TABLE 2. SPECIMEN WORKING TABLE IN SMOOTHING FOR AGE BIAS

sorter's age group code	province	district	civil condi-tion	sex	total in the compiler's or final age group $(F + .6G)$	sorter's age group		
						.6	.4	total
								G
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
06	05	06	1	1	65,139.0	37,044.0	24,696.0	61,740
			1	2	59,421.2	31,768.2	21,178.8	52,947
			2	1	640.0	303.0	202.0	505
			2	2	1,316.6	990.6	660.4	1,651
			3	1	207.8	217.8	145.2	363
			3	2	528.8	334.8	223.2	558
					$.4G + .6H$	H		
07	05	06	1	1	50,412.0	25,716.0	17,144.0	42,860
			1	2	35,116.2	13,937.4	9,291.6	23,229
			2	1	1,252.0	1,050.0	700.0	1,750
			2	2	9,454.6	8,794.2	5,862.8	14,657
			3	1	294.0	148.8	99.2	248
			3	2	505.2	282.0	188.0	470

case of all the age-groups, excepting in the case of the initial age-group T which would not be distributed but directly added to $.4S$ in the second counter as no identification code '1' had been gang-punched for this group.

REFERENCES

HARTLEY, H. O. (1946): The application of some commercial calculating machines to certain statistical calculations. J. Roy. Stat. Soc. Suppl., VIII, 2, 151.

MANDEVILLE (1946): Discussion on Hartley's paper. J. Roy. Stat. Soc. Suppl., VIII, 2, 174.