UNITED NATIONS

ECONOMIC
AND
SOCIAL COUNCIL

LIMITED

C.1

ST/ECLA/Conf.43/L.3
14 July 1972

ORIGINAL: ENGLISH

ECONOMIC COMMISSION FOR LATIN AMERICA

SEMINAR ON THE PREPARATION AND USE OF POPULATION AND HOUSING CENSUS TABULATIONS

Organized by the United Nations, through the Economic Commission for Latin America, the Statistical Office of the United Nations, the United Nations Trust Fund for Population Activities with the collaboration of the Latin American Deomographic Centre

Santiago, Chile, 14-19 August 1972

# CENTS, A TECHNIQUE FOR THE COMPUTER TABULATION OF POPULATION AND HOUSING CENSUS DATA

by

Howard G. Brunsman
U.S. Bureau of the Census

P-72-7-1623

A new approach to tabulation of data obtained from censuses of population and housing has been developing over the past several years. The CENTS system is a modification of the basic approach that has been used by the U.S. Bureau of the Census in the processing of its last several censuses. This system outlines the following operations:

1. Determine the areas for which a group of tabulations is to be prepared.

2. Group the basic records so that no group will contain records for more than one of the areas for which tabulations are wanted. For example, certain tabulations of the United States data are specified for urban places, urbanized areas, metropolitan areas, and counties, States, divisions and regions, and the urban and rural parts of such areas. To prepare these tabulations the records should be grouped by each urban place, the urbanized remainder and the rural remainder of each county. The county parts of an urban place will be grouped separately when an urban place lies in two or more counties. In like manner the county parts of the remainder of an urbanized area will be grouped separately when the urbanized area lies in two or more counties.

3. Prepare the basic tabulation for each of these groups.

4. Consolidate the tabulations to the required publication areas.

5. Prepare publication tables from the consolidated figures.

In the example presented above, the tabulation was prepared for approximately 9,000 groups. The tabulations for these groups were consolidated into figures for 5,445 separate urban places and 213 urbanized areas. Also for the total and urban and rural parts of each of 3,134 counties, 50 States, 9 divisions, 4 regions and the United States total. The tabulations were also consolidated into figures for 212 Standard Metropolitan Statistical Areas, and 509 State Economic Areas, which are groupings of counties.

In the 1950 census these operations were performed using the basic record on punch cards. The tabulations were prepared on unit counters. The resulting data were transferred to summary punch cards which were sorted and processed with conventional tabulators to consolidate the data which were then posted to tables and typewritten for publication. In the 1960 and 1970 Censuses the same basic operations were performed on computers.

/There are

There are two basic advantages that the Bureau of the Census of the United States has over smaller countries in these operations. First, it has a large computer. Second, it has a large group of skilled programmers to prepare the programmes. Unfortunately but few of the developing countries have either of these advantages. It should be possible to use available software packages to reduce the programming effort. A number of such packages are in existence. These include Report Program Generator (RPG), DATATEXT, 360 CROSS TAB, and MINITAB. All of these are subject to various limitations. Some of them require a computer with a capacity greater than that available in many developing countries. Others are too inefficient in the number of tables that may be produced in a single operation. And others are unable to consolidate the tabulated data into the wide range of publication areas recognized in a census operation.

We recognized the need for a software package of programmes that would efficiently process a large amount of data on a relatively small computer. Therefore we developed the programming technique, or software package, which has become known as CENTS, an acronym for CENsus Tabulation System. This system was developed basically by the U.S. Bureau of the Census, under the auspices of the U.S. Agency for International Development, but it should be recognized that several Governments in the hemisphere as well as the Organization of American States lent support to the venture and the system is working in some of those co-operating countries.

The CENTS approach has been used to tabulate the population and housing data for the 50,000,000 population of Mexico. This census was conducted in January 1970 and many of the basic tabulations were completed by November 1970. It has been used in tabulating advance sample data from the 1970 census of Chile. It has been used in Panama to tabulate the 1970 Censuses of Population, and Housing. The system is in use also in at least 20 other countries around the world, notably, Greece, Vietnam, Kenya, Brazil, Costa Rica, Philippines, Jamaica, and Guatemala.

CENTS is essentially designed to do the greatest amount of work in the least amount of time within the capacity of the computer. All of the basic tabulation operations are performed by special computer programmes which are activated by parameter cards. The basic tabulation of Step 3 above is performed by the CENTAL programme which is activated by 4 to 8 parameter cards for each table. The sequencing and consolidation of the resulting data of Step 4 is performed by standard sequencing routines and the CENCON programme which are activated by 4 to 10 parameter cards for a set of tables. The preparation of the publication tables specified in Step 5 above is performed by the CENPREP programme. This programme

/requires an

requires an area name card for each area. Each table requires from 4 to 10 parameter cards for heading and calculations plus one card for the content of the label of each line of the stub of the table. The preparation of the parameter cards of CENTS is facilitated by special coding forms. The parameter cards for CENTAL are coded on Form PH-421 as shown in Exhibit 1. The parameter cards for CENCON are coded on Form PH-422 as shown in Exhibit 2. The various parameter cards for CENPREP are coded on PH-423 to PH-426 as shown in Exhibits 1 to 4 1/.

Let us say we wish to tabulate the table shown in Exhibit 1 showing population by relationship to head of household by age. In this first example we assume that the tabulation is not required for sub-areas within the universe but only for the total of all records. The code for age as shown in byte 15 of each population record is 00 for under 1 year, 01 to 98 for 1 to 98 years and 99 for 99 years or more. The codes for relationship in byte 12 are as follows:

Codes for relationships

| Relationship | Code in S15 |
|---|---|
| Head of family | 3 |
| Wife or companion | 4 |
| Child of head | 5 |
| Other relative of head | 6 |
| Not related to head | 7 |
| Living alone | 8 |

We must tell the computer that Table 1 will contain 11 columns and 7 lines; that the person should be counted in column 2 if age code is 4 or less; if not, in column 3 if age code is 9 or less; if not, in column 4 if age code is 14 or less, etc., to count in column 11 if age code is 99 or less.

---

1/ Exhibits are shown in reference document "U.S. Bureau of the Census. CENTS: An approach to the Computer Tabulation of Data from a Population and Housing Census, Series ISP4, Nº 1, Washington, D.C. 1972", Appendix II.

/We must

We must also instruct the computer to count on line 2 if relationship code is 3 or less; if not, on line 4 if code is 4 or less; if not, on line 5 if code is 5 or less, etc.

With CENTS we perform these operations by supplying the computer with the parameter cards shown in Exhibit 1. The first card dimensions the table, the second tells it to establish the recodes for relationship based on byte 12, the third tells it to establish the recodes for age based on byte 15. The fourth tells it to tabulate in table 1 with column specified by age recode and line specified by relationships recode. This part of the CENTS programme produces a "Tally block" for the table. Now we must derive the various column and line totals and present the data in legible form with appropriate headings and designation of the labels for the various lines of the table. These specifications for the table are supplied to the computer by the parameter cards shown in Exhibit 1. The first card says that the stub of the table is 33 characters wide, and that the data are presented in 11 columns of type 1, each 9 columns wide. The next line says that for each line of the table we wish to calculate column 1 as the sum of columns 2 to 11. The next line says that for each column of the table we wish to calculate line 1 as the sum of lines 2 to 7. The next 6 lines show the left and right hand parts of each of the 3 lines of heading of the table. The next 7 lines show the content of the stub of the 7 lines of the table.

Of course, I have presented a relatively simple example. Further instructions are available in the tabulation operations. Two other types of recodes are available in CENTAL to meet other types of conditions. Other instructions permit you to omit certain cases from the table, such as persons under 12 years. A table may contain a separate group of lines for males and a corresponding group for females. An instruction will permit the code for line to be increased by a specified amount when sex is female. Other instructions in CENPREP will prepare derived figures, such as average size of household, per cent distributions, and medians. Each of these derived figures is obtained with a single parameter card.

The example presented above assumes that the tabulation is needed only for the total of all cases. When separate data are needed for various sub-areas, it is necessary to include an area specification card designating the most detailed types of areas for which data are required among the parameter cards for CENTAL; to sequence output of CENTAL with a standard sequence routine; to consolidate to publication areas using the CENCON programme actuated by parameter cards; to sequence the tabulation to place areas in the order specified for publication and to specify in

CENPREP the areas for which the tabulation is to be printed. The parameter cards needed to perform these operations are shown in Exhibit 2. Examples of parameter cards required for the tabulation of other tables are shown in Exhibits 3 and 4.

The use of CENTS in the tabulation of a census has the following advantages:

1. A tremendous reduction in the man-hours of programming required. All programming for a typical table can be completed in a very few hours.

2. Fewer skilled programmers are required. Some of the programming staff should have a full knowledge of computer programming but most of the parameter cards can be prepared by less skilled programmers. The print specification cards can be prepared by persons with no previous programming experience.

3. Less computer time required. This reduction in computer time results from the following factors.

a. The program is at least as fast as standard programmes written in assembler language since it is written in this language. In one test in Mexico they tabulated 12 tables at the rate of 11,000 persons a minute on a 360/40 and 27,000 a minute on a 360/50. Eighteen tables were tabulated at the rate of 2,000 persons per second on a 370/155 in Brazil.

b. The parameter cards are processed at "Object time". This saves the 10 to 30 minutes of computer time required to assemble each new programme with the usual programming technique.

c. A relatively large portion of the computer memory is available for storage of the tabulated data in the CENTAL programme. This results from the small portion of core required by CENTS in contrast to the large portion for FORTRAN, COBOL and most softwares. The portion of the CENTAL programme that converts the content of parameter cards into operating instructions shares core space with the tabulated data. A computer with only 32 thousand bytes of core, is able to process as many as 20 tables with a total of 3,500 tally cells.

d. The CENTAL programme permits the tabulation of the maximum number of tables in each operation by automatically adjusting to the nature of the tabulation and to the amount of core assigned to the operation.

/4. Facility

4. Facility to review specifications with the analyst. The content of the parameter cards is relatively simple. The programmer will be able to review the specifications with the analyst and obtain confirmation that they are being observed.

What are the limitations of the CENTS approach? I feel that there are four limitations.

1. Because it was prepared for use on smaller computers it was written in basic machine language, rather than FORTRAN or COBOL. Since the first applications were on IBM 360 computers it was written in the basic assembler language for these computers. The programmes are operational on Siemens computers since they are fully compatible with the IBM 360, and the programmes have been adapted to the Univac 9400 and are operational on this computer in Greece. We are conducting some experimental work on rewriting the programmes in COBOL so that they may be operated on a wider range of computers.

2. The documentation of CENTS does not include training materials. The materials include a set of coding sheets which facilitate the preparation of the essential parameter cards and the post-list of the various computer programmes. They also include the publication "CENTS, an Approach to the Computer Tabulation of Data from a Population and Housing Census" 2/ which contains a description of each type of parameter card and how they should be prepared and used.

3. While the CENTS material covers the tabulation process, it is assumed that the basic record has been subject to review, and that omissions and inconsistencies have been taken care of. What is needed is a processing system which carries the operation from the punch card to the publication table. It is true that programming tabulation with traditional means is time consuming and laborious. But the preparation of a good edit programme requires a more skilled programmer than that required for a tabulation programme. Fortunately the edit operation is performed only once and can usually be done with a single computer programme while the tabulation of numerous tables requires many programmes.

After the CENTS programme had been developed we were faced with the question of the best method of presenting them to potential users. We decided to use the workshop approach. Under sponsorship of the United States Agency for International Development we have conducted such workshops;

---

2/ Op. cit., see footnote 1/.

/six in

six in English in Washington, D.C., and two in Spanish, one in Santiago, Chile in September-October 1970, and the other in San José, Costa Rica in May-June 1971. Two workshops are now planned for Africa this year, one in English, the other in Spanish. Each workshop runs for approximately 8 weeks. There have been eighty participants in these workshops. The participants are programmers from developing countries. Their attendance is usually sponsored by AID although 21 of the participants have been sponsored by the United Nations. The instruction includes training in the application of CENTS in the tabulation of a census. It also includes some instruction in how the CENTS programmes operate since it is recognized that problems may be encountered in making the programmes operational at an installation. Also it may be necessary to perform tabulation operations that cannot be accomplished with the existing programme. The programme includes the possibility of a follow-up consultation by a U.S. Census representative to assist in making the system fully operational.

The CENTS materials are now prepared, is directed at the processing of a national census of population and housing. The input may be all housing records or all population records. Or it may consist of a housing record followed by a separate record for each person in the housing unit it permits the processing of data enumerated on a 100 per cent basis where each record has an equal weight of 1, or of data from a sample census where each record has an integer weight shown in a specific field of the record. In an earlier version the CENTS programme required the input to be presented in one and two byte binary numbers with the records on tape. It has now been modified to accept one to four digit decimal digit numbers on the tape or in punch cards.

But, how about the application of CENTS to other tabulation problems? It is fully applicable in the tabulation of data from sample surveys. The facility and low cost of preparing tabulations with CENTS is especially advantageous in sample surveys since funds available for tabulation are often limited. The CENTS programme also has been used to tabulate the Censuses of Agriculture of Mexico, Panama and Nicaragua. The CENTAL programme is unable to establish recodes based on input variables larger than 9,999, but it is able to test whether the variable has a value other than zero and to add the value of such a variable in the tables.

The Bureau of the Census has placed the CENTS programmes in the public domain. We place no restrictions on the use of any parts of them. In fact, we are delighted to supply potential users, especially the statistics offices of developing countries, with all available documentation including copies of the programmes.