

REDATAM

REDATAM: A SUMMARY

UNITED NATIONS
ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN
LATIN AMERICAN DEMOGRAPHIC CENTRE

Distr.
RESTRINGIDA
LC/DEM/R.50
Series A, No.176
October 1987
ENGLISH
ORIGINAL: SPANISH

REDATAM: A SUMMARY

ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN (ECLAC)
LATIN AMERICAN DEMOGRAPHIC CENTRE (CELADE)
Casilla 91 - Santiago de Chile
Tel: 2283206 Cable: UNATIONS

REDATAM: A SUMMARY

TABLE OF CONTENTS

I. OBJECTIVES OF THE REDATAM SYSTEM	1
II. SUMMARY OF THE FACILITIES OFFERED BY THE REDATAM SYSTEM	1
III. EQUIPMENT REQUIRED FOR USING REDATAM	3
IV. PRESENT LIMITATIONS OF REDATAM (Version 2.00)	3
IV. FEATURES PLANNED FOR THE NEXT RELEASE OF REDATAM (Version 3.00)	4
VI. EXPERIENCE WITH REDATAM IN SOME COUNTRIES OF THE REGION	6
VII. WHERE TO OBTAIN ADDITIONAL INFORMATION ON REDATAM	8
APPENDIX: REQUIREMENTS FOR CREATING A REDATAM DATABASE	9
1. Equipment.	9
2. Characteristics of the data	9
3. Personnel required	9
4. Time	10
5. Space occupied and type of generation	10

REDATAM: A SUMMARY

I. OBJECTIVES OF THE REDATAM SYSTEM

REDATAM 1/ is a software system for obtaining census or survey results for specific geographical areas rapidly and at low-cost using a microcomputer. REDATAM stores the microdata (that is, the values of each variable of each individual) of an entire population and housing census or survey on a common microcomputer hard disk (or laser disks for large countries) and then permits users to obtain any tabulations for any geographical areas down to city blocks without assistance from programmers.

II. SUMMARY OF THE FACILITIES OFFERED BY THE REDATAM SYSTEM

English language and Spanish language versions

The REDATAM system software, the "User's Manual" and the "Database Creation Manual" are each available in separate English-language and Spanish-language versions. The tutorial and all examples in the "User's Manual" refer to a small demonstration database, for a hypothetical country, that comes with the software.

Geographic selection

The user defines the universe of cases to be processed, selecting only the geographic areas of interest.

Grouping of geographic areas

Capacity to compose geographic areas of interest from any smaller areas down to city blocks or smaller if these are coded in the census. The smaller areas may be in different administrative areas, thereby permitting, for example, the formation of metropolitan areas or the study of counties from two different provinces but served by the same highway or through which the same river passes.

1/ REDATAM = REtrieval of DATA for small Areas by Microcomputer.

REDATAM: A SUMMARY

Self-documented database

The variables have a "short-name" to facilitate describing the tabulations, etc., required by the user and a longer title for display with the results; there are also names of the categories for the original variables and for recoded or derived variables.

Interactive

The user interacts with the system through menus and other facilities. There are various forms of help provided and input errors are indicated.

Calculation of derived variables

New variables, like 5-year age groups or the housing quality derived from roof, floor and wall materials, may be defined by recoding and through the utilization of arithmetic operations. These new variables may be temporary for use during a specific process or may be incorporated into the database.

Statistical results

Three types of statistics can be produced: frequencies, cross-tabulations and averages, the latter two with up to four variables. Row, column and total percentages can be obtained for a table and the tabulations and other results can be requested for sub-areas within the area of interest, as well as for the entire area.

Hierarchical processing

The system works with two levels of variable (housing and population) and results can be obtained for each of the levels separately or combined. For example, the number of persons by sex and age within households can be cross-tabulated with the roof and wall conditions.

Generation of sub-databases

There is a facility for creating a REDATAM sub-database for a specified area of interest. Thus, starting with the national database, sub-databases can be created, for example, for individual provinces or regions to be given to regional centres.

Production of files for export to other packages

When the basic statistical operations in REDATAM are not sufficient, a file with the variables of interest for the selected area, can be created for processing by another package such as SPSS or SL-MICRO.

REDATAM: A SUMMARY

Storage of results

The statistical results can be kept for later printing or inclusion in a document, as well as for the production of graphic output using other packages such as WordPerfect, Wordstar, Lotus, Mathplan, etc.

III. EQUIPMENT REQUIRED FOR USING REDATAM

1. IBM PC, XT, AT or fully compatible microcomputer.
2. 640K main memory.
3. At least one floppy disk.
4. Monochromatic or colour monitor (screen).
5. Printer with at least 80 columns.
6. Operating system PC-DOS version 2.0 or higher.
7. A hard disk with approximately 1.4 megabytes (Mb) available is required for the REDATAM system and the demonstration database that comes with it. The total amount of hard disk (or laser optical disk) space for storing the actual census or survey of interest will depend on the number of persons enumerated and the number and size of the variables on the questionnaire.

IV. PRESENT LIMITATIONS OF REDATAM (Version 2.00)

IBM equipment

The system has been designed to be operated in IBM/PC/XT/AT or fully compatible microcomputers since these are by far the most common in the Latin American and Caribbean region. With respect to compatibles, REDATAM has been used with the following equipment: EPSON, TOSHIBA, COMPAQ and GREAT WALL (Chinese computer).

Only two hierarchical levels

If the census or other dataset has three levels (dwelling, household and persons), household and persons can be generated as the first and second levels, respectively; results can be produced for dwellings by selecting

REDATAM: A SUMMARY

only the first household. For other types of datasets with more than two levels there may not be any solution.

Decimal weights or variables are not accepted

Lack of confidentiality protection

Any user with access to the database can, if desired, select the smallest areas and, depending on the number of cases, identify houses and persons.

Processing times for geographic areas with large populations

The velocity of a microcomputer cannot be compared with that of a mainframe. When the universe is for any area of more than, say, one million persons, it is normally preferable to use a large computer.

The creation of a database is often complex

The creation of a REDATAM database is relatively simple for small censuses and surveys, particularly when the data are already available on a microcomputer. However, for large files, such as a census with various millions of cases, the process is more complex and requires a programmer with both mainframe and microcomputer experience. Naturally, there must be a way of communicating between the mainframe and microcomputer whenever the data are initially stored on magnetic tapes which can be read only on the mainframe computer.

Lack of integration with a cartographic system

The selection of the user-defined area is presently made using the names and codes of the localities (provinces, districts, census tracts, etc), which requires that the user consult maps to determine the set of codes that must be used. There is also no direct way at present to show the results graphically on maps.

IV. FEATURES PLANNED FOR THE NEXT RELEASE OF REDATAM (Version 3.00)

It is expected that Version 3.00 of REDATAM will be ready for general distribution around March 1988.

New features

Storage and calculation of decimal values including the use of decimal weights.

REDATAM: A SUMMARY

Database and confidentiality protection including the use of "passwords" to permit access to the system database management functions and to gain access to various geographic levels. It will also be possible to protect the creation of sub-databases with the elimination, if desired, of the lowest levels of the geographical hierarchy so that the exported sub-databases have a built-in confidentiality protection. Ability to pre-programme various long processes and to chain them for execution, without any human intervention, at a convenient time, such as during the night when there is no one to use the system interactively (batch processing).

Calculation and display of the number of observations selected at each geographic level, so that the user has an idea of the number of cases that will be processed for the geographic area that is being defined.

Possibility of creating partial databases (such as for regions or provinces) and to join them into a more complete database at the end, thereby facilitating the generation of very large databases.

Capability of working in a network of microcomputers having simultaneous access to a database on a single hard disk or laser optical disk (under consideration).

Geographic selection from a file formatted by the user to define complex areas that are laborious to do interactively.

New commands in the statistical processor, such as:

- IF a condition is true, THEN carry out an operation. For example,
IF sex = 1 and age > 15 THEN COMPUTE newvar = 1.
- SAMPLE to select a probability sample of the geographical area of interest. CASES to select only the first n cases of an area, to facilitate the testing of a complex process.
- Additional statistical operations (under consideration).
- HIGHEST and LOWEST to select cases within a household, for example, the oldest person economically active in the household (under consideration).

Improvements (partial list)

Identification of the instruction number that produced a particular statistical operation (e.g., tabulation) in the output listing.

REDATAM: A SUMMARY

Interactive selection of the file of interest when using commands such as LOAD, GEOGRAPHY and SAVE.

Possibility to save the output of a VIEW.

Improved facilities with AREABREAK.

Improved velocity.

VI. EXPERIENCE WITH REDATAM IN SOME COUNTRIES OF THE REGION

CHILE (REDATAM installed since around September 1986)

Two databases stored on laser optical disk, one for the entire 1982 census (12 million persons and 4 million households) and the other for a 5 percent sample of the census (around 600,000 persons). The entire census is stored on 3 laser disks, each the size of a regular floppy disk and holding around 115 megabytes of information.

Various applications including:

- The study of collective houses in Greater Valparaiso at the level of comuna, district and zone.
- The study of population in the city of Concepcion, at the level of districts and census zones, by occupation, industry, age group and education.
- Application in a sewerage project in the city of La Serena, at the level of districts, zones and blocks.
- Application to a project to obtain special tabulations for neighborhood units ("unidades vecinales") in three comunas of Santiago and the creation of sub-databases for each of the comunas. The neighborhood units are not defined in the census and must be created from smaller units (city blocks).
- The study of the characteristics of the indigenous population of the province of Temuco.

REDATAM: A SUMMARY

COSTA RICA (REDATAM installed since around June 1987)

Database of the 1985 census installed on a 60mb hard disk (3 million persons). Also databases of household surveys.

Various applications including:

- Studies for the development of the cantons of Guatuso and Upala.
- Project for a health study on aging in the canton of Puriscal.
- Project for university regional centres in Turrialba and Nicoya.

GUYANA (REDATAM installed since around May 1987)

Database of the 1986 Demographic Survey, GUYREDEM (42,000 persons) for the analysis of the results and tabulations of fertility for the 9 regions of the country.

SAINT LUCIA (REDATAM installed since around July 1986)

Database of the 1980 census of around 125,000 persons on a hard disk.

Various applications including:

- Answering requests from various users for tabulations for the entire country (e.g., Ministries of Health, Agriculture, etc.)
- Study for the expansion of the telephone system in various areas of Castries.
- Study for the construction of a dam.
- Applications in a study to expand the sewerage network.

URUGUAY (REDATAM installed since around July 1987).

Database of a 15 percent sample of the 1985 census (450,000 persons).

Applications including:

- Project to study extreme poverty in all regions of the country, taking into account both housing and population characteristics.
- Tabulations to verify the tabulations which will be published based on the census sample.

REDATAM: A SUMMARY

VII. WHERE TO OBTAIN ADDITIONAL INFORMATION ON REDATAM

Additional information on REDATAM can be obtained by writing to:

CELADE
Casilla 91
Santiago, Chile

Telephone: 228-3206

Cable: UNATIONS

Telex: 340295 (Transradio)

441054 (ITT)

240077 (Telex Chile)

REDATAM: A SUMMARY

APPENDIX: REQUIREMENTS FOR CREATING A REDATAM DATABASE

Through actual examples, the "Database Generation Manual" describes each step in the creation of a REDATAM database under differing conditions. Here, only the general principles and problems will be outlined.

1. Equipment.

- a) IBM/PC/XT/AT or fully compatible microcomputer with 640 Kilobytes of memory; preferably an AT for speed of processing.
- b) Hard disk for the creation of files and the final storage of the database; disk size depends on the files to be generated.
- c) Laser optical disk (optional) for storage of the database when the final compressed dataset is very large (say, over 100mb).
- d) Connection to transfer the data from the mainframe computer to the microcomputer (not required if the data is already on the microcomputer).
- e) Since long processes are involved during the database creation, an uninterruptable power supply (UPS) is highly desirable in places where the electricity is unstable.

2. Characteristics of the data

- a) The geographical variables must be defined and the entire file must be sorted by the geographical variables.
- b) The variables in the database to be created must be defined in the dictionary and the frequencies for each variable must be available for final verification of the database.
- c) A calculation must be made of the space which the final database, after compression, will occupy.
- d) The quality of the geographical variables must be verified to avoid problems in the generation of the database indices.

3. Personnel required

- a) When the input data file of the database to be created is relatively small, it can be stored on the microcomputer. In this circumstance, only a person with some knowledge of microcomputers is required since the REDATAM autoloading feature can be used.
- b) When the input files are large and a mainframe computer must be used,

REDATAM: A SUMMARY

a programmer is required with experience with mainframes and COBOL, as well as with microcomputers.

4. Time

In general, under normal conditions, a database can be created in around one to three weeks, depending on the resources available and the database size.

5. Space occupied and type of generation

- a) The space occupied by a REDATAM database is a function of the following basic factors: the number of households and persons and the number and size of the variables for the households and persons. Table 1 shows the number of household and population variables, the approximate population and the size in megabytes of each of the databases that CELADE has calculated. In addition to the database space shown, another 1.2mb should be added for the system and work files.
- b) The approximate space can be calculated taking into account only the number of population records of the census or survey. For a census with an average number of variables (30 to 40), around 600,000 person records can be stored on a 20 megabyte hard disk or approximately 1,000,000 on a 30 mb hard disk.
- c) As a general rule, it can be said that around 3,000,000 persons can be stored on a 100mb hard disk. Larger files should normally use optical laser disks. Note also, that larger files can be broken down into regions or provinces to fit on a hard disk, but that this makes it difficult to produce results for areas cutting across the boundaries.
- d) The "autoloading" feature of REDATAM can generally be used for databases up to around 300,000 records. Files above this size, require the use of a manual procedure which is more complex.

REDATAM: A SUMMARY

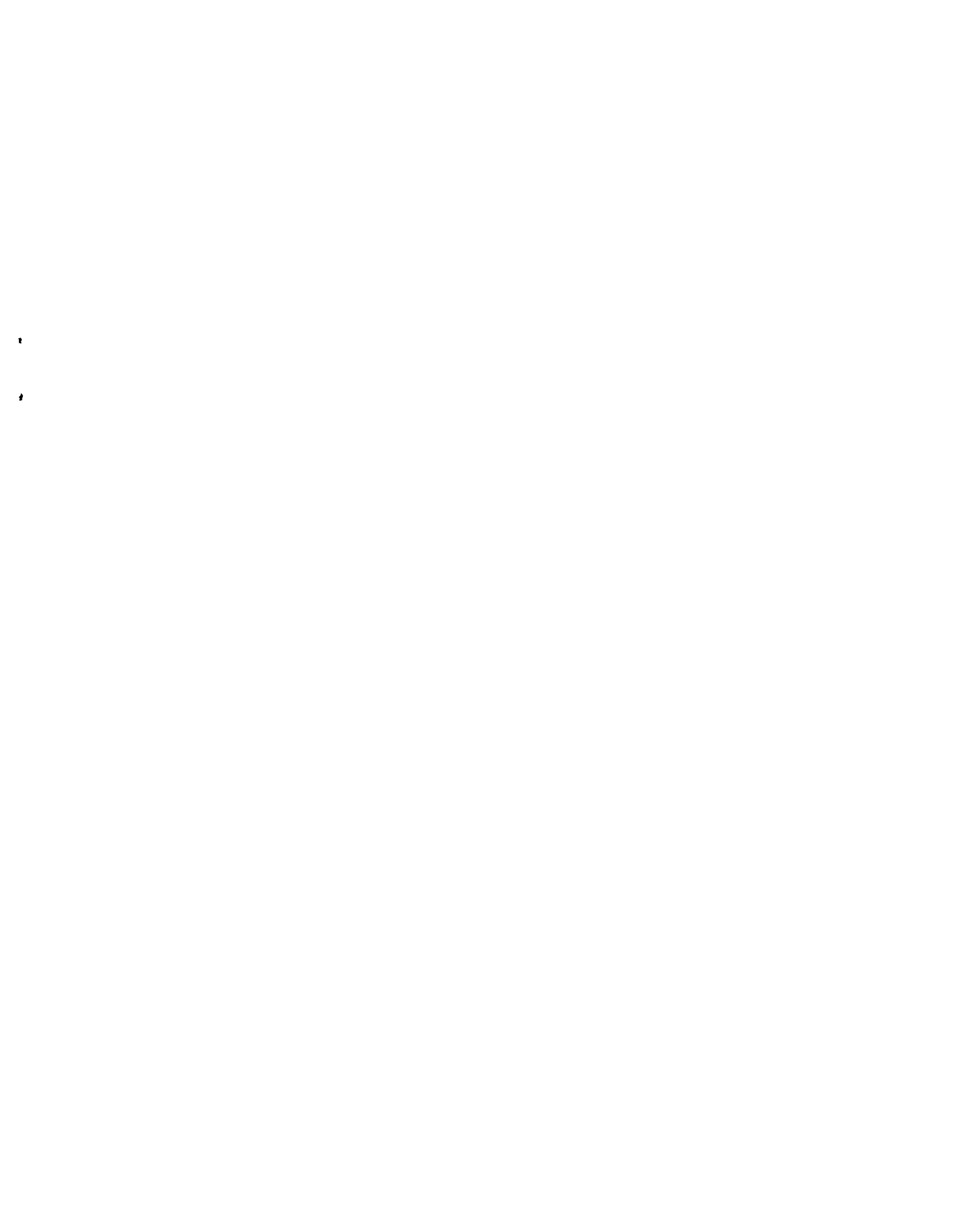
Table 1

Estimated space likely to be occupied
by REDATAM population and housing census databases

Countries	Variables		Population (000)	Space (mb)
	Households	Population		
Argentina (1)	25	32	1.000	19.82
Brazil (1)	26	61	1.000	67.07
Chile	39	34	11.330	207.94
Colombia (2)	18	27	2.610	47.42
Costa Rica	39	24	2.500	45.87
Dominican Republic	79	60	5.648	252.21
Ecuador	18	26	8.000	177.99
Guatemala	46	50	6.054	207.29
Haiti	18	26	5.053	107.35
Panamá	25	41	1.825	68.01
Paraguay	30	45	3.026	104.27
Perú (3)	27	46	4.251	141.32
Trinidad & Tobago	18	55	1.080	44.14
Uruguay (4)	45	29	.450	14.91

Notes:

- (1) For each 1,000,000 records of the sample.
- (2) Extended questionnaire (10% of the population).
- (3) Complete questionnaire (25% of the population).
- (4) Sample of 15% of the entire census.



7

8

9

0

1