

Automatic data editing experience in 2010 Mexican Census

Isaac Salcedo

INEGI

April, 2014

Content

- 1 Objective
- 2 The Process Strategy
- 3 Design
 - Definition of editing criteria
 - The theoretical vectors methodology
 - Reports
- 4 Training
- 5 Implementation
- 6 Results

Objective

Main objective

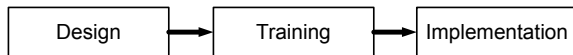
The objective of the data editing is improve the information's consistency, because the data coming from the work field, information entry or coding have some problems in its congruence. That is, assure the quality of information for the generation of statistics according to a conceptual framework. And as result of this, the data editing provides important information about the data quality and improvement of the questionnaire.

Kind of errors treated in data editing

- **Missings values:** It is the absence of information in a question, where according with conceptual design must be exists.
- **Multiresponse:** It occurs when a question has more responses than permitted in its design.
- **Inconsistent values:** When exist a response in the question but is not a valid value for that item or is not logical with other answer related with it.
- **Ignore filter questions:** This mistake occurs when some responses should not be registered in a question because there exist a filter.

Strategy

Major stages in data editing process.



Design guidelines

The automatic data editing was developed considering this guidelines:

- Keep the conceptual and methodological standards established for each issue and variable.
- Preserve the correctly information collected.
- Apply a review of logical consistency supported by empirical evidence observed in previous projects.
- Assign values as long as they are supported by the information contained in other associated variables or impute the missing information.

Criteria

232 criteria were designed for automatic data editing:

Theme	Total of criteria	Theme	Total of criteria
Initial Treatments	7	Dwelling Characteristics	29
Food Access	3	Treatments for population	6
Sex and Age	16	Household Relationship	35
International Migration	20	Entity or Country of birth	4
Health Services	4	Religion	2
Disability	15	Indigenous Language	4
Educational Characteristics	12	Residence Entity or Country in 2005	8
Marital Status	10	Economic Characteristics	30
Fertility	27		

Theoretical vectors

This methodology allows complete control of the values that can take a set of related variables.

- The variables considered in the editing criteria and the values taken for each one are identified.
- A one-dimensional array is constructed (called theoretical vector) whose components represent the values they can take each of the variables.
- All combinations that can take a vector are generated
- The routing function allows assign a unique value for each combination in the generated order.
- The result is an identifier of each combination and therefore a specific treatment.

Reports

Reports allowed to review the editing data was complete and free of inconsistencies.

- Control of initial treatments.
- Control totals.
- Related variables.
- Origin-Destination matrix.
- Missing and not specified values.
- Percent changes.

Training

It was designed with emphasis to the review and analysis of automatic editing reports.

- Training had two stages:
 - From designers of the process to the data edition responsible
 - From the responsible to the editing analysts

Implementation

This activity involved the automatic editing criteria execution.

- Analysis of reports
- Data base release

Results

- Control of changes in the information.
- Exhaustiveness of the treatments.
- The time assigned to carry out this project.
- The centralization of the criteria ensured that all them were applied in the correct order and with homogeneity



Thank you