

1930 Population Census of Mexico: a 1% Pilot Microdata Sample

Robert McCaa, Professor, University of Minnesota, Minnesota Population Center
 Aurora Gomez Galvarriato, Director, Archivo General de la Nación

BACKGROUND

IPUMS-International: Census samples of Mexico account for one-eighth of IPUMS-International usage.

Samples currently available:

census	sample %	person records
1960	1.5	502,800
1970	1.0	483,405
1990	10.0	8,118,242
1995	0.4	332,061
2000	10.6	10,099,182
and to be released in 2012		
2010	10.0	~11.2 million

Mexico won the IPUMS-International Census Microdata olympics, out-ranking UK, France, Germany, Spain, India, China, Brazil, Australia and other would-be titans of the world of demography.

1930 Population Census, many firsts for Mexico (Director Juan de Dios Bojórquez, 1926-32):

- First to be conducted according to international standards
- First to instruct field workers on census concepts and definitions
- First to mobilize the population with a mass promotional campaign to enlist the cooperation of citizens of the entire republic, including some indigenous populations that had never been enumerated before.

2010 partnership: AGN-MPC

- Archivo General de la Nación partners with Minnesota Population Center to recover a 1% sample funded by IPUMS-Latin America. Data keying performed by personnel of the AGN.

Figure 1. Discussing data entry: Archivo General de la Nación, Nov. 2, 2010



Gabriela Campos Orta, Irasu Zapata Hernández, Aurora Gómez Galvarriato Freer, Alberto Efraín Álvarez Ferrusquía y Pedro Manuel Lichtle Fragoso

RESEARCH QUESTIONS

1. What is the geographic coverage of the extant manuscripts?
2. What is the quality of the content?
3. Does the pilot indicate that the census of sufficiently high quality and research interest to justify a high-precision sample, say 10%?
4. What are the costs and benefits of recovering the full-count microdata for the entire country?

METHODOLOGY

- Sample unit: folio page of 50 lines front and 50 back. Average: 66 persons per folio.
- Sample design: every folio numbered "44" (44, 144, 244, ...)
- » Pagination is by municipality. The first folio of each municipality is numbered "1".
- » To maintain link between archived, microfilmed, and scanned folios, repaginating was rejected.
- » "44" was retained to speed keying and to be certain of folio keyed
- » Total number of folios: 358,084
- » Folios in sample: 3,628
- Sample bias:
 - » 919 municipalities with fewer than 44 folios (~2500 pop.) omitted from sample
 - » 1,013 medium to large municipalities in sample
 - » 115,503 person records
- Full count microdata for selected areas: 74,924 person records
 - » South, urban capital: Oaxaca, Oaxaca - 34,318 (lack 4,991)
 - » Center, industry: Santa Rosa Nexcoxtla, Veracruz - 9,357
 - » Center, rural Morelos, 3 municipalities: 9,742
 - Ayala 5,081
 - Huiztilac 2,085
 - Tetecala 2,576
 - » North, mining center: Hidalgo del Parral, Chihuahua - 21,507
 - » Data keyed in 1987 in situ.
- Total person records (sample + full count) - 190,427
- Key every field
 - » Codes for Hollerith fields: 1 = col 1, 2 = col 2.
 - » Alphabetic characters for text fields
 - » Capture hand-written edits as well as data collected in the field
 - » Data were not verified. Only the person count was checked for each folio.

References:

AGN catalogue entry for the 1930 population census of Mexico: Archivo General de la Nación/ Instituciones Gubernamentales: época moderna y contemporánea/ Administración Pública Federal S. XX/
 Título: Censos de Población, Agrícola, Ganadero y de Edificios (230)
 Fecha(s): 1921 - 1930 (1929 - 1930) (1921) Nivel de descripción: Fondo Volumen y soporte: 668 Cajas (1282.8 metros lineales) (Adicionalmente a las cajas existen 1376 paquetes)
 Productores: Dirección General de Estadística y Departamento de Estadística Nacional.
 familysearch.org entry: [https://wiki.familysearch.org/en/Mexico_1930_Census_\(FamilySearch_Historical_Records\)](https://wiki.familysearch.org/en/Mexico_1930_Census_(FamilySearch_Historical_Records))
 Instituto Nacional de Estadística, Geografía e Informática, Ciudad de México, México. 5º Censo de población, 1930, Archivo General de la Nación, Ciudad de México, México. FHL microfilm, 55 reels. Family History Library, Salt Lake City, Utah.

DATA

Table 1. 1930 census: 34 person and 7 geographic variables.

Variable	IPUMS	Family Search	N	Notes
List of Variables				
Folio number	X	X		
State - Entidad Federativa	X	X		
Municipality - Municipio	X	X		
Ward - Pólibo	X	X		
Category of Settlement - Categoría	X	X		
District Number - Cuartel Número	X	X		
Section Number - Sección Número	X	X		
Block Number - Manzana Número	X	X		
Name of Block Captain	X	X		
Name of Enumerator	X	X		
List of Persons Form				
Person line number: 1-50 persons, verso 51-100.			115,262	
Column				
1. Street, avenue, lane	X			
2. Number on the exterior of the dwelling	X			
3. "x" to indicate the head of the family	X		22,560	
4. Name and family name	X	X		
5. Male Sex (tick box)	X	X	57,532	66 edited male
6. Female Sex (tick box)	X	X	57,730	79 edited female
7. Age in years	X	X	282	33 not stated
8. Age in months	X	X	3,268	
9. Age in days	X	X	282	
10. Single civil status (tick box)	X	X	21,344	
11. Civil marriage (tick box)	X	X	4,985	+ 14,589 also church
12. Church marriage (tick box)	X	X	11,625	+ 14,589 also church
13. Free union (tick box)	X	X	10,085	
14. Widowed (tick box)	X	X	6,435	
15. Divorced (tick box)	X	X	311	
16. Head and wife: yes (tick box)	X	X	1,841	30% head 10% were iterate
17. Only known how to read (tick box)	X	X	1,841	
18. Homemaker (tick box)	X	X	30,863	+ 5,112 edited in
19. Head (yes/no) (tick box)	X	X	18,319	+ 5,412 edited in with valid entries (includes students, etc.)
20. Profession or Occupation	X	X	56,227	
4 digit edited code	X	X		
21. Class of Profession or Occupation	X	X		
4 digit edited code	X	X		
22. State or Country of Birth	X	X		
23. Mexican Nationality (tick box)	X	X		
24. Current Nationality	X	X		
25. Nationality before current	X	X		
26. Speaks Castilian Spanish - tick box	X	X		
27. Speaks another language or dialect	X	X		
Real Estate:				
28. Has property in the country (tick box)	X	X		
29. Has property in the city, village or hamlet	X	X		
30. Family has its own dwelling (tick box)	X	X		
31. Physical or mental defects	X	X		
32. Do the defects prevent working (tick box)	X	X		
33. Religion	X	X		
34. How long have you been without work	X	X		

Note: Edited codes and marks were captured as well as marginalia and other particulars.

Family and Household in the 1930 census

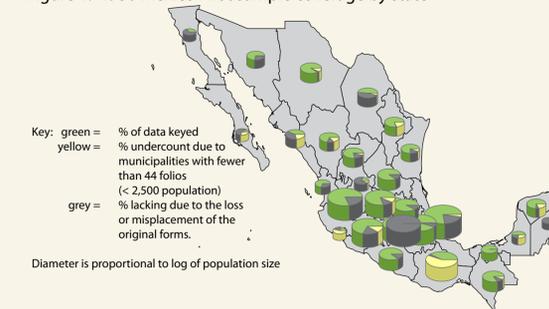
The 1930 census was the first to mark head of the family (by today's definition, "household").

- However, the mark was a simple "+",
- 0.6% of the households list more than 20 persons. Some may be due to omission of the "+", even though the instructions stated that "these marks are very important" to record the mark for the head of each family.
- The instructions further stated that the head should be the first person listed for each family.
- Unfortunately there is no indication of family or household relationship.
- Nonetheless, family units can be inferred from family name, age, sex, marital status and implicit ordering of persons in each family. The spouse is typically listed following the head, and children are listed from oldest to youngest. The family units of married children are commonly noted below the nuclear unit of the head along with others who share a family name with the head of spouse of head.
- Some results:
 - » Average household size = 5.0
 - » 83% of households were headed by males
 - » Males in a marriage recognized by
 - » church and state were 5 times more likely to head households (6053 heads, 1217 not)
 - » church only were almost 4 times more likely to be a head (4695:1128)
 - » state only or free union were only 3 times more likely (1802:663; 3699:1247)
 - » widowed were twice as likely (1030:541)
 - » divorced were about as likely as not (67:60)
 - » single were very unlikely to head households (1351:9254)

RESULTS

Geographic coverage of the sample and the extant manuscripts

Figure 1. 1930 Mexico Pilot sample coverage by state



What is the quality of the content?

1. Missing data is manageable
2. Edits are slight
3. Variables are rich in detail.
4. Effects of editing seems logical

Table 2. Effect of editing on literacy as coded by field enumerators

Edited codes in office	As coded in the field			
	reads and writes	reads and writes	reads only	blank
	reads and writes	25,678	285	4,373
reads only	7	1,843	6	
blank (illiterate)	578	0	82,733	

Comment:

- 95% of codes were not edited; the same as coded by the enumerator.
- Largest edit is from blank to reads and writes (n=4,373 or 3.7%).
- On the whole, these are children listed as attending school, but the enumerator left blank both the "reads and writes" and "only reads" columns.

Table 3. Singulate Mean Age at Marriage (SMAM) in years by sex

	Literacy			
	illiterates	literate	total	n literate 15-54
Males	23.8	24.8	24.2 years	10,626
Females	19.4	22.6	20.3 years	8,385

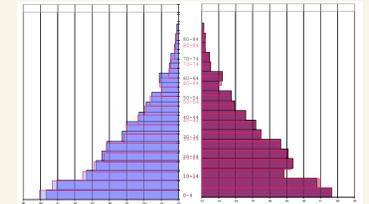
	Illiterates who speak a native language:			n literate 15-54
	no	yes		
Males	24.1	22.2	23.8 years	
Females	19.6	19.2	19.4 years	

Comment:

- Literate males married at 24.8 years, one year older than illiterates, on average.
- For females, the effect of literacy was to delay marriage age by 3 years.
- In terms of fertility, the delay in marriage would reduce completed family size by 1 child for literate females, assuming that age-specific rates were similar for both groups.
- Among illiterate males, speaking a native language noticeably lessened the age at marriage, by almost 2 years on average. For females the difference was less than one-half year.

DISCUSSION

Age pyramid: comparing pilot sample with official figures



1. The pilot indicates that the census is of sufficiently high quality and research interest to justify a high-precision sample
 - a. Data content is rich, editing effects seem logical and fairly light
 - b. 1930 is a turning point in Mexican population history—the beginning of radical demographic transformations—a 6 fold increase in 7 decades.
 - c. To measure the demographic impact of the Mexican Revolution, a one percent sample does not offer the necessary precision.
2. The costs of recovering the full-count microdata for the entire country are manageable
3. Benefits of full count microdata are substantial
 - a. Facilitates local, state, regional and national studies
 - b. Essential for a comprehensive assessment of the demographic impact of the Mexican Revolution

DISSEMINATION AND DEVELOPMENT

- A raw version with Spanish language metadata is to be disseminated by the Archivo General de la Nación
- Later, the sample is to be integrated with English language metadata and disseminated through the IPUMS-International web-site: www.ipums.org/international along with other Mexican samples:
- If the pilot sample is successful, it will elicit the resources needed to produce a complete version of the full-count microdata—freely accessible in the public domain.

ACKNOWLEDGEMENTS

- Fundación Archivo General de la Nación
- National Science Foundation, International Integrated Microdata Series, Grant SES-0851414
- National Institutes of Health, Integration of Latin American Census Microdata, Grant 2R01HD044154

