2 Argentina

The First National Historical Census Microdata

Robert McCaa
Michael R. Haines
Eileen M. Mulhare

Introduction

On October 1, 1967, the Argentine demographers Jorge L. Somoza and Alfredo E. Lattes completed the first national computerized samples of historical census microdata (Somoza and Lattes, 1967). Using punch cards and working from manuscripts of the Argentine censuses of 1869 and 1895, they designed robust samples of the original enumeration sheets, yielding some 100,000 persons for each census. Constrained by the 80-column punch card and a frugal budget, they were forced to use the census page as their sampling unit and sacrifice both names and person numbers to get the job done. Occupations and other alphabetic information were converted to numeric codes for the same reason. More than one thousand boxes of punch-cards were used, without taking into account those consumed by corrections. Thanks to these economies and their

---

Robert McCaa is Professor of History at the University of Minnesota and a principal investigator on the IPUMS-International project at the Minnesota Population Center. Michael R. Haines, Ph.D. is the Banfi Vintners Distinguished Professor of Economics at Colgate University and is a Research Associate at the National Bureau of Economic Research, Cambridge, MA. Eileen M. Mulhare, Ph.D. is Research Associate in Anthropology and Lecturer in Latin American Studies at Colgate University, Hamilton, NY.

For more information on nineteenth-century Argentine historical microdata, see: http://www.hist.umn.edu/~rmccaa/data/arg6995.zip.

experience in computerizing modern census data, Somoza and Lattes succeeded in producing a model dataset with more than 200,000 cases for slightly less than US $21,000—surely on a per record basis, one of the least expensive, large-scale historical databases ever constructed. Although generally ignored outside of Argentina by historians and demographers alike, their pioneering success still constitutes a model for many reasons. The following seven are our favorites:

1. the documentation of the entire project—variable-by-variable, code-by-code—was published at the same time the data were released, including verbatim transcripts of enumerator instructions and reproductions of enumeration forms;
2. the nationally representative samples were scientifically drawn, the results were compared with official published tables, and discrepancies were discussed;
3. missing data procedures, which included hot-decking for some variables, were fully described and their inferences were explicitly coded in the datasets by the use of a system of flags;
4. their samples integrated two successive censuses with largely similar sets of variables, enumeration procedures, and coding schemes, facilitating the study of historical change;
5. new national tables were published using identical categories and variables, so that results for the two censuses could be compared directly, for the first time for such elementary variables as age and marital status;
6. the data were not hoarded, but rather were released immediately to scholars free of charge, including, for researchers lacking ready access to the necessary hardware, the offer of free computer time to produce any combination of tables on demand; and
7. the data were used, although almost entirely by Argentine demographers, with several dozen

publications based on these samples within the first decade or so of their release.

Unfortunately, time has not treated this pioneering work well. The punch cards were recycled long ago, and the original tapes were lost. All currently extant machine-readable copies of the samples derive from a single source, which has been handed down to us sorted by age (!) not page, thereby losing the ordering of individuals on the original census form. Then too, non-Argentine scholars have largely ignored not only this exceedingly valuable comparative dataset, but more importantly the methods which Somoza and Lattes pioneered. Several historical census projects over the past couple of decades might have profited greatly by an awareness of their methods and procedures. This may be explained in part because of language barriers. Unhelpful too, was the fact that samples of modern Argentina census microdata have not circulated freely. In any case, Argentines have used these samples extensively, so much so, that it might be argued that the historical demography of nineteenth century Argentina has been on a sounder footing than that for any other country in the Americas, until the recent completion of several samples for the United States and Canada.

Source Material

The original enumeration sheets for the 1869 and 1895 censuses constitute the treasure trove from which the Somoza-Lattes samples were drawn. Located in the Archivo General de la Nación in Buenos Aires, a microfilm of the original is now available from the Family History Center, Salt Lake City, Utah. Before undertaking this pioneering enterprise the Somoza-Lattes team checked the original enumerations against the published totals for 191 districts. The results of this test were exceedingly encouraging, as they found few discrepancies (Somoza and Lattes, p. 18). In 94.3 percent of the districts, their test counts agreed exactly with the published figures. In the remaining districts, the relative difference amounted to 2.3 percent of the published figure, which reduces to a modest 0.13 percent when the exact tallies are taken into account. Suspiciously, the cipher “0” or “5” was the final digit in eight of the eleven discrepancies.

Their initial assessment of the high quality and completeness of the data was confirmed by statistical analysis when the data-processing phase ended. Even though their sample design was not ideal, as we shall see below, only minor discrepancies were found between relative distributions of the published tables and the samples.

Procedural History

A thoroughly documented dataset deserves to have its procedures thoroughly documented as well, and this is exactly what Somoza and Lattes provide. Their report helpfully shows a time line indicating the number of researchers and months dedicated to each task (Somoza and Lattes, p. 25). They also demonstrate how they over-lapped various activities to maximize through-put and reduce the total time required to complete the operation without sacrificing quality or wasting resources. Key-punching began only after some six months of preparatory work and was completed in less than three months with the help of a mere half-dozen operators. Each card was verified by a team of three in only three months. These efficiencies were made possible by a team of, at times, as many as twelve researchers, who selected the samples, transcribed and coded the cases, and checked and rechecked the codings—all before key-punching began. After coding and key-punching were completed, four researchers spent an additional six months in data-processing tasks: dealing with missing data, checking and re-checking for errors and omissions, computing new national tables, evaluating results and drafting documentation (Somoza and Lattes, p. 25). Begun in early June 1966, the entire project was completed October 1, 1967—a record unsurpassed, even in the age of microcomputers, sophisticated data-entry software, and dissemination by internet.

Compromises were required however. Controversial then, today at least one of their shortcuts, using the enumeration schedule page as the sample unit, might be considered unacceptable. It must be recognized, however, that they were handicapped by the original instructions to the enumerators, who were told to fill every line of every page. Then too the enumeration schedule contains no row or column to indicate the

first person in a dwelling, household or family. The Somoza-Lattes team was left with only two options for a sampling unit: "page" or "person" (Somoza and Lattes, pp. 15-16). They chose "page" as a means of achieving significant economies. Since relationship to head was not a concept employed in either census, there was no consistent way of identifying families for constructing a hierarchical dataset. Adopting "person" as the sample unit would have yielded a more robust database, but it would have also increased costs and required a reduction in sample size. Opting for 'page' kept alive the hope that a means of identifying family or household might emerge for at least some minor if not major civil divisions, and for those pages a hierarchical structure could be inferred. This hope remains alive among a second generation of researchers entranced by family history, but no one has yet developed a workable stratagem—regardless of cost.

Electronic Formats

The data are available simply as ASCII text files, of three and four megabytes, for 1869 and 1895, respectively. Compressed, both datasets fit nicely on a conventional floppy disc.

Variable Availability

The population variables available in the Argentine census of 1869 and 1895 are presented in Table 2-1, on the following page.

Confidentiality Provisions

Collected more than a century ago, there are no confidentiality restrictions of any kind on these data. If there were, identifying individuals would be exceedingly tedious and the information gained so mundane as not to arouse the slightest concern, even from the most misguided advocate of privacy rights.

Table 2-1. Population Variables in the Nineteenth-Century Argentine Census Samples

Year of Census	1869	1895
Major civil division (province / territory)	X	X
Minor civil division (department / section)	X	X
Urban / rural residence	X	X
Page number	X	X
Type of dwelling		X
Age	X	X
Sex	X	X
Relationship to head	.	.
Birthplace (country / province)	X	X
Illegitimacy	X	X
Literacy	X	X
School attendance	X	X
Religion	.	X
Marital status	X	X
Years married	.	X
Number of children	.	X
Disability	X	X
Orphan	X	X
Property owner	.	X
Occupation / profession	X	X
Position in occupation / profession	X	X

Data Access

The data and corresponding codebooks are available on the web at: http://www.hist.umn.edu/~rmccaa/data/arg6995.zip.

Publications Using These Data

Alfredo E. Lattes and his colleagues at the Centro de Estudios de Población (Buenos Aires) have long been the most prolific users of these datasets (see bibliography). In their analyses of the demography of contemporary Argentina, they have expertly exploited the census samples to provide a point of departure for the study of the great Argentine demographic transformations in

this century. All the classic demographic concerns are well represented in this literature—labor-force participation of both males and females, internal migration, mortality, fertility, urbanization, immigration, and marriage—to mention the most important.

Research Possibilities

Much work remains, including a regionalization of these phenomena, as well as a more exhaustive analytical approach, using more sophisticated, multivariate statistical tools, such as logistic regression, log-linear modeling, or multi-way standardization. International comparisons are an obvious route for gaining new views of how Argentine demographic transformations compare and contrast with the experience of other populations.

Expert Users

Researches associated with the Centro de Estudios de Población (Buenos Aires) remain the experts in the use of these data: Somoza, Lattes, Pantelides, and Recchini de Lattes, among others. Some might include here as less knowledgeable, but experienced users, the authors of this paper.

Data Expansion

After inquiring with likely depositories over much of a decade, we are resigned to the likelihood that the only way to recover the original ordering of individuals is to return to the manuscripts in the National Archives (or a microfilm copy) and reorder the records, line-by-line, page-by-page. Fortunately ordering by pages has already been accomplished, thanks to the fact that places are unambiguously identified, and within places pages are numbered exactly as they are found in the National Archives.

A second possibility would be a national integration of these historical samples with the twentieth century microdata for the censuses of 1960, 1970, 1980, 1991 and 2001.

Third, would be an international integration either with other nineteenth century census microdata, such as those for the United States, the United Kingdom, and Canada, or with modern censuses.

Bibliography

Kritz, E. H. (1985) *La formación de la fuerza de trabajo en la Argentina: 1869-1914*. Buenos Aires: Centro de Estudios de Población, No. 30.

Lattes, A. E. (1968) *Evaluación y ajuste de algunos resultados de los tres primeros censos nacionales de población*. Buenos Aires: Instituto Torcuato Di Tella, Población y Sociedad, no. 51.

____. (1972) "Algunos indicios de migración interna diferencial en Argentina antes de 1869," in vol. 1, *Actas: Conferencia Regional Latinoamericana de Población, Mexico, 1970*. Ciudad de México: 558-562.

____. (1975a) "El crecimiento de la población y sus componentes demográficos entre 1870 y 1970," in A. E. Lattes (ed.), *La población de Argentina*. Buenos Aires: Centro de Estudios de Población, serie CICRED, 29-66.

____. (1975b) "Perspectiva histórica de la evolución de la población," in A. E. Lattes (ed.), *La población de Argentina*. Buenos Aires: Centro de Estudios de Población, serie CICRED, 21-28.

____. (1975b) "Redistribucion espacial y migraciones," in A. E. Lattes (ed.), *La población de Argentina*. Buenos Aires: Centro de Estudios de Población, serie CICRED, 95-112.

____. (1981) "The Dynamics of the Rural Population in Argentina Between 1870 and 1970," in J. Balán (ed.), *Why People Move*. Paris: UNESCO, 94-121.

Lattes, A. E. and Poczer, R. (1968) *Muestra del censo de población de la ciudad de Buenos Aires en 1855*. Buenos Aires: Instituto Torcuatto di Tella.

Müller, M. S. (1978) *La mortalidad en la Argentina: Evolución histórica y situación a 1970*. Buenos Aires: Centro de Estudios de Población.

Pantelides, E. A. (1979) *Evolución de la fecundidad en Argentina*. Santiago de Chile: CELADE.

____. (1983) "La transición demográfica argentina, un modelo no ortodoxo," *Desarrollo Económico*, 22(88): 511-534.

____. (1984) *Análisis y propuesta de corrección de la información sobre estado civil en los cuatro primeros censos nacionales argentinos*. Buenos Aires: Centro de Estudios de Población, Serie Estadísticas Sociodemográficas no. 2.

Recchini de Lattes, Z. (1975a) “Población económicamente activa,” In A. E. Lattes (ed.), *La población de Argentina*. Buenos Aires: Centro de Estudios de Población, serie CICRED, 149-172.

____. (1975b) “Urbanizacion,” in A. E. Lattes (ed.), *La población de Argentina*. Buenos Aires: Centro de Estudios de Población, serie CICRED, 113-148.

Rothman, A. M. and de Janvry, B. (1972) “Relación entre el nivel de fecundidad y otras variables demográficas y socioeconómicas en Argentina (1869-1960),” in Vol. 1, *Actas: Conferencia Regional Latinoamericana de Población, Mexico, 1970*. Ciudad de México: 284-300.

Schkolnik, S. and Pantelides, E. A. (1975) “Los cambios en la composición de la población,” in A. E. Lattes (ed.), *La población de Argentina*. Buenos Aires: Centro de Estudios de Población, serie CICRED, 67-94.

Somoza, J. L. (1968) “Levels and differentials of fertility in Argentina in the XIX century,” *Milbank Memorial Fund Quarterly* 46:3(part 2, 1968), 57-77.

____. (1971) “Algunos efectos sociales y económicos derivados de la baja de la mortalidad en la República Argentina entre 1900 y 1960,” *Desarrollo Económico*, 11(41 abr): 113-123.

Somoza, J. L. and Lattes, A. E. (1967) *Muestras de los dos primeros censos nacionales de población, 1869 y 1895*. Buenos Aires: Instituto Torcuato Di Tella, Centro de Investigaciones Sociales, Documento de Trabajo no 46.