

Las muestras censales IPUMS: una herramienta para el análisis demográfico

Bolzán Florencia Ileana¹ Y Hoszowski Augusto Estanislao²

RESUMEN

El objetivo de este trabajo es mostrar las facilidades que ofrecen para realizar análisis demográficos las muestras de los censos de población que el proyecto Integrated Public Use Microdata Series (IPUMS) de la Universidad de Minnesota pone a disposición de los usuarios en su sitio web. Se ejemplificarán algunos cuadros y análisis mediante el software estadístico Stata, mostrando cómo las muestras que IPUMS ofrece, tienen un tamaño que permite su manejo en cualquier computadora personal, siendo por otro lado tal la magnitud de los tamaños muestrales que los valores estimados casi no difieren de los poblacionales. Si bien la web del proyecto nos ofrece 238 muestras censales de 74 países, se realizará el ejercicio con los Censos de Población de Argentina de los años 1970, 1980, 1991 y 2001. Se efectuarán diversos análisis que serían imposibles de realizar con las bases completas de microdatos de toda esa serie de Censos de Población, o con datos agregados.

1. INTRODUCCIÓN

Un censo de población es un recuento total o parcial de la población de un país, una provincia o una ciudad. La Organización de Naciones Unidas recomienda que se realice cada 10 años y que contenga información sobre población total, grupos de edades, sexo, lugar de residencia, nacionalidad, estructura ocupacional, escolaridad y condiciones de las viviendas entre otros. En la Argentina, el primer Censo Nacional se efectuó en el año 1869 y le siguieron los de: 1895, 1914, 1947, 1960, 1970, 1980, 1991, 2001 y 2010.

¹ INDEC, fbolz@indec.mecon.gov.ar

² INDEC, ahosz@indec.mecon.gov.ar

Las bases de microdatos censales poseen un tamaño que es dificultoso (con los medios informáticos presentes) para poder manejar en una computadora personal. Por otro lado, los usuarios no institucionales no disponen muchas veces de las bases de microdatos.

El sitio IPUMS aloja muestras de Censos de Población de gran cantidad de países. En general, el diseño de la muestra aplicado es la selección sistemática de viviendas, variando la fracción de muestreo de país a país, y de censo a censo. Pero en general la fracción de muestreo es 10% o más. Se trata por lo tanto de una muestra de gran tamaño, lo que hace que los errores debidos al muestreo (como veremos más adelante) sean mínimos.

Las muestras que IPUMS ofrece en su sitio Web permiten realizar análisis demográficos manejando simultáneamente varios censos de población.

Otra ventaja de la utilización de las bases IPUMS es la *homologación* de las mismas. Esto significa que las variables son comparables (en la medida de lo posible) censo a censo. La comparabilidad esta dada por la unificación de la codificación de las variables y los nombres que ellas toman, en las muestras censales. La profusa documentación presentada en el sitio permite a los usuarios determinar cuando una misma variable puede ser comparada en un país de censo a censo, o entre censos de distintos países, las diferencias entre ellas, etcétera.

2. METODOLOGÍA

Se conformó una base con las muestras de microdatos a nivel de persona de los censos de Argentina de los años 1970, 1980, 1991 y 2001. Las variables que se encuentran disponibles en la base para cada registro de la misma son las siguientes:

- Provincia
- Departamento
- Identificación del hogar dentro del departamento
- Cantidad de personas que viven en el hogar
- Identificación de la persona dentro del hogar
- Personas de la familia en el hogar
- Sexo
- Edad
- País de nacimiento

- Cantidad de hijos
- Estado laboral
- Tipo de trabajo
- Estado Civil
- Nivel educativo
- Estado de alfabetización
- Relación de la persona con el jefe de hogar
- Cantidad de habitaciones de la vivienda
- Suministro de agua de la vivienda
- Factor de expansión de la persona
- Factor de expansión del hogar

Con esta base de microdatos, se realizaron diferentes tabulaciones. En primer lugar, se comprobó la *representatividad* de la muestra de 2001, comparándola con los datos censales publicados en la página del INDEC del mismo año.

En una segunda instancia, se realizaron diferentes análisis demográficos, para los cuales se necesita contar con los datos a nivel micro.

A nivel de personas, calculamos los siguientes indicadores:

- Condición de alfabetización según sexo en personas mayores a 10 años
- Lugar de nacimiento según sexo y grupo etario quinquenal
- Población por provincia según sexo y edad
- Preferencia de dígitos (en cuanto a la edad de la persona) según sexo
- Estructura poblacional del país según edad y sexo
- Diferencia de edad entre cónyuges

Y a nivel de hogar, las siguientes tabulaciones:

- Distribución del tamaño de hogar según Censo de Población
- Promedio de personas por habitación según Censo de Población

3. RESULTADOS Y DISCUSIÓN

Verificación de representatividad:

En primera instancia, mostraremos la precisión de la estimación dada por la muestra IPUMS. Se compararán los datos censales del año 2001 extraídos de la página del mismo, con la muestra IPUMS del mismo año expandida. Las variables a contrastar serán: población alfabetizada desde 10 años en adelante, la población extranjera y población total según grupo etario. El cotejo se realizará por sexo y grupo etario, y por provincia y grupo etario. A continuación presentaremos los cuadros de las diferencias ente el valor *estimado* y el poblacional.

- Cuadro 1. Diferencias relativas entre el valor estimado y el valor poblacional. Población de más de 10 años alfabetizada según grupos de edad.

Grupos de edad	Población de 10 años o más	Condición de alfabetismo - Diferencias relativas (%) (*)					
		Alfabetos			Analfabetos		
		Total	Varones	Mujeres	Total	Varones	Mujeres
Total	-0.02	-0.02	0.01	-0.03	-0.01	0.04	-0.06
10-14	-0.05	-0.03	-0.13	0.07	-2.09	-3.13	-0.51
15-19	-0.01	-0.03	0.10	-0.16	1.98	2.15	1.71
20-24	-0.25	-0.24	-0.22	-0.25	-1.73	-0.17	-4.01
25-29	0.15	0.15	0.48	-0.17	0.62	-0.89	2.69
30-34	-0.10	-0.11	-0.09	-0.12	0.17	0.15	0.20
35-39	0.20	0.18	0.19	0.17	0.95	0.68	1.27
40-44	0.04	0.07	-0.06	0.18	-0.87	-0.34	-1.46
45-49	-0.39	-0.35	-0.54	-0.17	-1.54	-2.60	-0.42
50-54	-0.33	-0.39	-0.40	-0.37	0.91	-0.46	2.31
55-59	0.57	0.58	0.80	0.38	0.37	-0.46	1.16
60-64	0.25	0.22	0.32	0.13	0.93	1.44	0.50
65-69	0.21	0.18	-0.04	0.36	0.86	2.82	-0.64
70-74	0.15	0.08	0.11	0.06	1.33	0.85	1.65
75-79	-0.33	-0.28	-0.17	-0.36	-1.04	3.07	-3.47
80 y más	-0.09	0.01	-0.17	0.11	-1.26	-0.67	-1.51

Fuente: INDEC. Censo de población, viviendas y hogares de la República Argentina. Año 2001; IPUMS. Muestra del censo de población, viviendas y hogares de la República Argentina. Año 2001

(*) Se tabuló el indicador $100 * (Valor\ Censal - Estimación) / Valor\ Censal$

Cuadro 2. Diferencias relativas entre el valor estimado y el valor poblacional. Lugar de nacimiento según sexo y grupo etario.

Sexo y	Total	Lugar de nacimiento - Diferencia relativa (%) (*)
--------	-------	---

grupos de edad		En el país			En el extranjero		
		Total	Varones	Mujeres	Total	Varones	Mujeres
Total	0.05	0.08	0.04	0.11	-0.52	-0.37	-0.65
0-4	-0.03	-0.04	-0.14	0.06	4.57	1.28	7.87
5-9	0.13	0.12	0.28	-0.05	1.82	4.93	-1.34
10-14	-0.05	-0.04	-0.14	0.06	-0.55	-1.99	0.92
15-19	-0.01	-0.01	0.15	-0.16	-0.42	-1.62	0.66
20-24	-0.25	-0.28	-0.30	-0.25	0.63	3.16	-1.35
25-29	0.15	0.12	0.44	-0.20	0.93	0.74	1.07
30-34	-0.10	-0.10	-0.11	-0.09	-0.14	0.31	-0.52
35-39	0.20	0.18	0.12	0.24	0.46	1.90	-0.72
40-44	0.04	0.13	0.01	0.23	-1.57	-1.65	-1.50
45-49	-0.39	-0.32	-0.57	-0.09	-1.55	-1.37	-1.70
50-54	0.75	0.91	-0.25	2.00	-1.29	-2.30	-0.36
55-59	0.57	0.65	0.73	0.58	-0.33	0.90	-1.50
60-64	0.25	0.38	0.49	0.28	-1.12	-0.85	-1.38
65-69	0.21	0.32	0.14	0.46	-0.90	-0.31	-1.45
70-74	0.15	0.32	0.56	0.15	-1.41	-3.31	0.18
75-79	-0.33	-0.24	0.07	-0.45	-0.92	-0.26	-1.42
80 y más	-0.09	-0.01	0.05	-0.04	-0.56	-1.59	0.02

Fuente: INDEC. Censo de población, viviendas y hogares de la República Argentina. Año 2001; IPUMS. Muestra del censo de población, viviendas y hogares de la República Argentina. Año 2001

(*) Se tabuló el indicador $100 * (Valor\ Censal - Estimación) / Valor\ Censal$

- Cuadro 3. Diferencias relativas entre el valor estimado y el valor poblacional.
Población por provincia según grupo etario, de 0 a 59 años .

Provincia	Grupos etario - Diferencia relativa (%) (*)					
	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59
Total	0.05	-0.03	-0.07	0.04	-0.16	0.07
CABA	0.14	0.87	0.02	0.17	0.65	-0.48
Buenos Aires	-0.09	-0.18	-0.40	0.16	-0.23	-0.01
Catamarca	1.72	-2.04	0.59	-2.80	-1.73	1.63
Chaco	-0.21	-0.62	0.54	0.31	-0.36	0.36
Chubut	-0.65	-0.23	-0.99	-1.55	1.94	-0.73
Córdoba	0.18	-0.26	-0.23	-0.36	-0.18	-0.38
Corrientes	-0.83	2.03	-1.44	1.23	0.61	-0.47
Entre Ríos	-0.02	1.92	-0.06	0.19	-0.03	0.68
Formosa	-1.41	0.03	-0.08	-0.86	-0.76	2.30
Jujuy	1.03	0.24	0.68	-0.79	1.94	-1.44
La Pampa	-1.55	-0.71	0.25	-0.23	-0.46	0.97
La Rioja	-3.33	0.27	-0.49	-2.36	1.41	0.11
Mendoza	0.13	0.00	1.05	-0.84	0.53	0.90
Misiones	-0.06	-1.06	0.43	0.96	-2.40	1.06
Neuquén	2.51	-1.87	0.99	0.19	-2.75	1.50
Río Negro	-0.24	0.12	0.81	0.12	-2.82	0.80
Salta	0.87	-0.05	-0.52	-0.04	-0.56	0.72
San Juan	0.30	1.42	-0.60	0.59	-0.08	-0.64
San Luis	1.78	-1.62	1.09	0.69	-3.60	0.66
Santa Cruz	2.34	1.70	2.38	0.03	1.33	2.65
Santa Fe	-0.38	-0.05	-0.06	0.28	-0.23	1.00
Santiago del Estero	-0.26	-1.65	0.10	-1.16	0.90	-0.97
Tierra del Fuego	-2.02	-0.67	1.60	-4.41	2.84	-4.20
Tucumán	1.53	0.54	1.05	1.40	0.01	-1.66

Fuente: INDEC. Censo de población, viviendas y hogares de la República Argentina. Año 2001; IPUMS. Muestra del censo de población, viviendas y hogares de la República Argentina. Año 2001

(*) Se tabuló el indicador $100 * (Valor\ Censal - Estimación) / Valor\ Censal$

- Cuadro 4. Diferencias relativas entre el valor estimado y el valor poblacional.
Población por provincia según grupo etario, de 60 años y más y totales por provincia.

Provincia	Grupos etario - Diferencia relativa (%) (*)					
	60 - 69	70 - 79	80 - 89	90 - 99	100 y más	Total
Total	0.24	-0.05	-0.02	-0.58	-0.27	0.00
CABA	-0.42	-0.87	0.63	-1.40	3.23	0.08
Buenos Aires	0.47	0.08	-1.04	-0.01	-6.94	-0.10
Catamarca	0.59	-0.32	-1.23	-2.67	9.09	-0.31
Chaco	1.27	1.28	-2.62	-5.62	-46.34	-0.01
Chubut	-0.32	-0.72	-6.55	2.21	37.50	-0.49
Córdoba	0.18	1.72	2.14	-0.56	-0.84	-0.01
Corrientes	-2.67	-2.61	1.71	2.63	59.18	-0.01
Entre Ríos	0.02	-1.55	1.19	4.10	-1.45	0.40
Formosa	2.28	1.07	-3.00	-14.02	-11.11	-0.30
Jujuy	0.71	0.23	5.01	-1.22	-11.11	0.48
La Pampa	-0.04	-4.19	1.23	13.04	-11.11	-0.51
La Rioja	4.28	0.21	2.13	13.70	41.18	-0.66
Mendoza	0.74	0.63	0.55	-1.69	1.23	0.32
Misiones	0.86	1.78	-3.31	-6.05	23.08	-0.20
Neuquén	-2.01	3.11	1.66	5.19	-42.86	0.13
Río Negro	-0.82	-1.60	-0.55	13.19	13.04	-0.23
Salta	-0.09	-0.43	-0.41	3.59	-7.69	0.10
San Juan	1.75	-0.26	-2.07	9.96	-60.00	0.32
San Luis	1.04	6.45	-1.00	-6.82	100.00	0.27
Santa Cruz	-1.20	-1.13	-1.40	-16.56	100.00	1.49
Santa Fe	-1.06	-0.07	1.13	-2.72	-2.94	-0.04
Santiago del Estero	1.47	-1.07	-0.22	0.68	-41.03	-0.48
Tierra del Fuego	-0.07	-1.23	-27.04	-50.00	0.00	-1.09
Tucumán	2.32	-3.27	1.39	-9.42	60.78	0.66

Fuente: INDEC. Censo de población, viviendas y hogares de la República Argentina. Año 2001; IPUMS. Muestra del censo de población, viviendas y hogares de la República Argentina. Año 2001

(*) Se tabuló el indicador $100 * (Valor\ Censal - Estimación) / Valor\ Censal$

Como podemos ver en los Cuadros 1 y 2 las diferencias relativas a nivel grupo de edades quinquenales y sexo son mínimos, por lo que podemos concluir que la muestra IPUMS es representativa de la población desagregada a ese nivel.

Los Cuadros 3 y 4, nos muestran que la representatividad de la muestra IPUMS a nivel de provincia también es correcta. Los errores relativos más altos se encuentran en los grupos de edad donde la población es muy reducida, y en algunos casos nula.

- Preferencia de dígito

Otra comprobación que podemos realizar es cómo se comporta la muestra al estimar la preferencia de dígito al declarar la edad. El mismo consiste en verificar si existe algún dígito de finalización de edad que es más declarado que otros. Como es sabido, los dígitos 0 y 5 son más 'preferidos' que el resto.

Para poder visualizar esta preferencia se utiliza el Índice de Myers, el cuál nos permite identificar los dígitos más declarados. Para la construcción de este índice, se asume que el total de la población se distribuye equitativamente entre los grupos de edad que finalizan en cada dígito. Es decir, un 10% de la población tiene edades terminadas en 0, un 10% en 1, y así sucesivamente. El cálculo el índice se da de la siguiente manera:

$$IM_{(j)} = \frac{\left(\sum_{i=0}^8 N_{[(10+j)+(10i)]} \cdot \alpha_{1;j} + \left(\sum_{i=0}^7 N_{[(20+j)+(10i)]} \right) \cdot \alpha_{2;j} \right)}{\sum_{j=0}^9 \left[\left(\sum_{i=0}^8 N_{[(10+j)+(10i)]} \right) \cdot \alpha_{1;j} + \left(\sum_{i=0}^7 N_{[(20+j)+(10i)]} \right) \cdot \alpha_{2;j} \right]} \cdot 100 - 10$$

Donde, j es el dígito final y N la población total para la edad dada en el subíndice. En el caso de $\alpha_{1;j}$ y $\alpha_{2;j}$, son coeficientes dados.

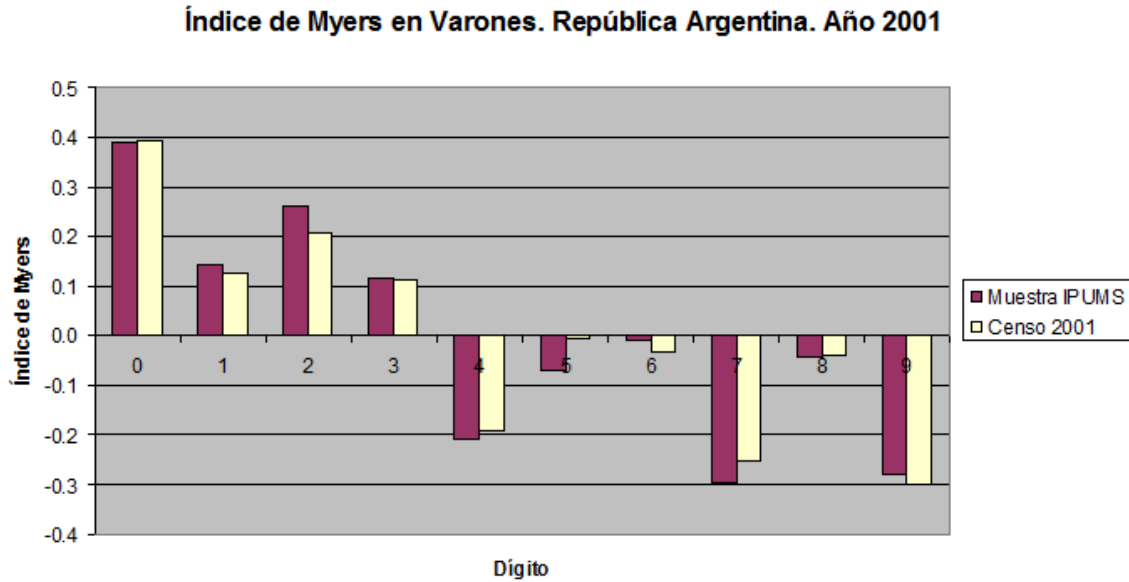
Si queremos ver la preferencia general, se tiene un Índice Resumen de Myers, dado por la siguiente fórmula:

$$IRM = \sum_{j=0}^9 |IM_{(j)}|$$

Donde, si el mismo nos da valores entre 0 y 5, significa que la atracción por preferir un dígito es baja; si es entre 5.1 y 15, es intermedia; si toma valores entre 15.1 y 30 es alta; y de 30 en adelante, muy alta.

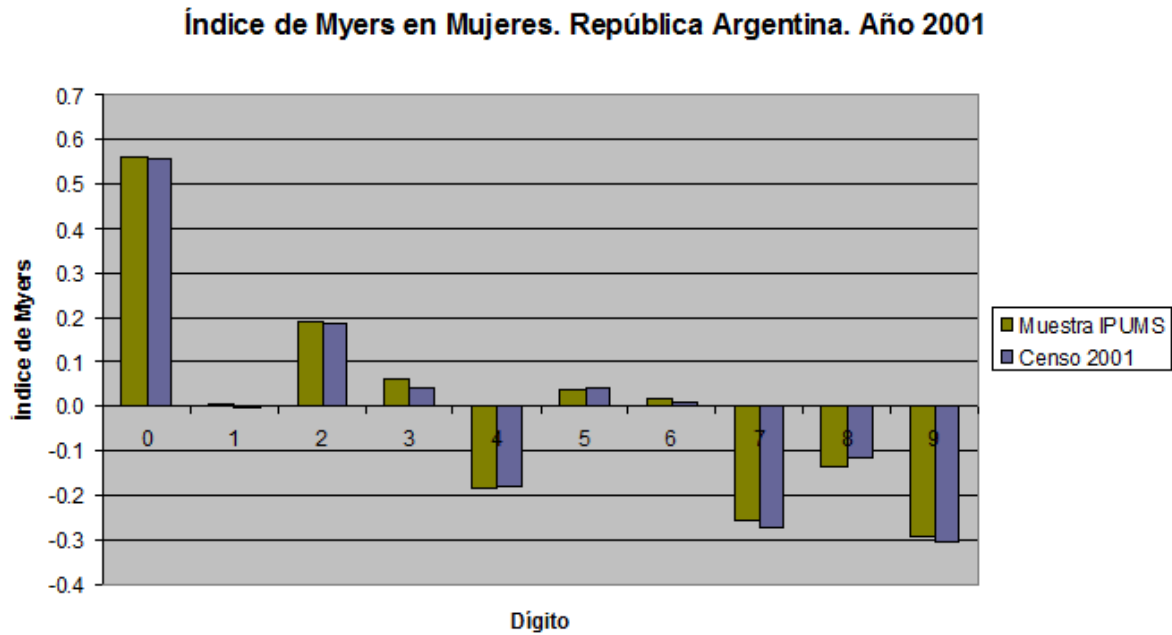
A continuación, mostraremos como se comporta la preferencia de dígito en el Censo 2001, tanto con los datos obtenidos a través de la expansión de la muestra IPUMS, como con los valores censales.

- Gráfico 1. Preferencia de dígito de edades en Varones. Censo 2001 y muestra Censo 2001 IPUMS.



Fuente: INDEC. Censo de Población, Hogares y Viviendas; IPUMS. Muestras del Censo de Población, Hogares y Viviendas. Año 2001.

- Gráfico 2. Preferencia de dígito de edades en Mujeres. Censo 2001 y muestra Censo 2001 IPUMS.



Fuente: INDEC. Censo de Población, Hogares y Viviendas; IPUMS. Muestra del Censo de Población, Hogares y Viviendas. Año 2001

De los *Gráficos 1 y 2*, podemos ver que el Índice de Myers de preferencia de dígito es casi idéntica entre los datos censales y los obtenidos a partir de la muestra IPUMS.

Para complementar esta información, se calculó el índice de Myers para los cuatro Censos, tanto para varones como para mujeres.

- Cuadro 5. Índice de Myers en varones. Años 1970, 1980, 1991 y 2001.

Índice de Myers	Año			
	1970	1980	1991	2001
0	0.519	0.613	0.421	0.389
1	-0.203	-0.691	-0.380	0.142
2	-0.142	0.063	0.051	0.261
3	-0.106	0.038	0.191	0.115
4	-0.139	-0.141	-0.179	-0.207
5	0.069	0.240	0.097	-0.071
6	-0.173	-0.139	0.095	-0.010
7	-0.043	-0.099	-0.112	-0.295
8	0.093	0.191	0.133	-0.043
9	0.126	-0.075	-0.316	-0.281

Fuente: IPUMS. Muestra Censal de Población, Hogares y Viviendas. Años 1970, 1980, 1991 y 2001.

- Cuadro 6. Índice de Myers en mujeres. Años 1970, 1980, 1991 y 2001.

Índice de Myers	Año			
	1970	1980	1991	2001
0	0.720	1.083	0.702	0.561
1	-0.292	-0.947	-0.450	0.006
2	-0.156	0.046	0.048	0.191
3	-0.132	-0.060	0.052	0.062
4	-0.142	-0.254	-0.152	-0.183
5	0.192	0.460	0.244	0.036
6	-0.099	-0.187	0.093	0.017
7	-0.039	-0.192	-0.228	-0.259
8	0.063	0.187	0.073	-0.139
9	-0.115	-0.137	-0.382	-0.291

Fuente: IPUMS. Muestra Censal de Población, Hogares y Viviendas. Años 1970, 1980, 1991 y 2001.

- Cuadro 7. Índice Resumen de Myers en varones y mujeres. Años 1970, 1980, 1991 y 2001.

Año	IRM Varones	IRM Mujeres
1970	1.613	1.951
1980	2.290	3.553
1991	1.975	2.425
2001	1.815	1.746

Fuente: IPUMS. Muestra Censal de Población, Hogares y Viviendas. Años 1970, 1980, 1991 y 2001.

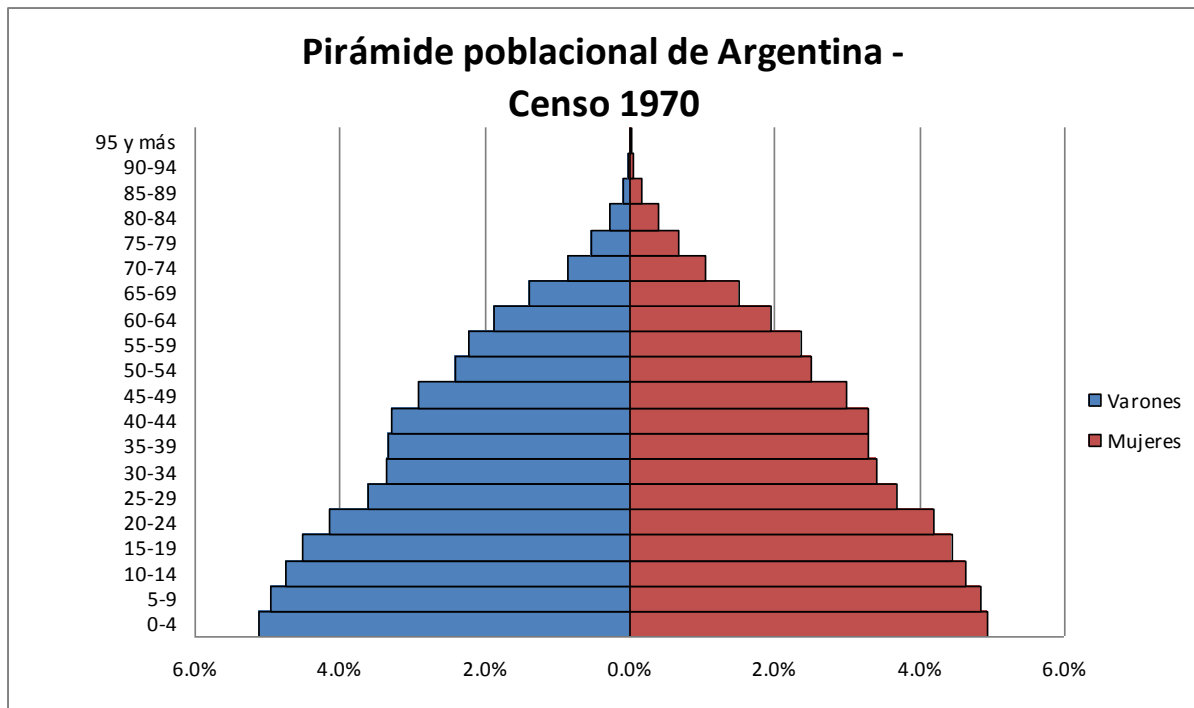
Análisis demográfico básico

Procederemos a realizar algunos análisis socio-demográficos sencillos entre los censos de los años 1970, 1980, 1991 y 2001. Al tener las muestras IPUMS de los cuatro censos en una sola base, la posibilidad de graficar es sencilla utilizando el software Stata 9.2.

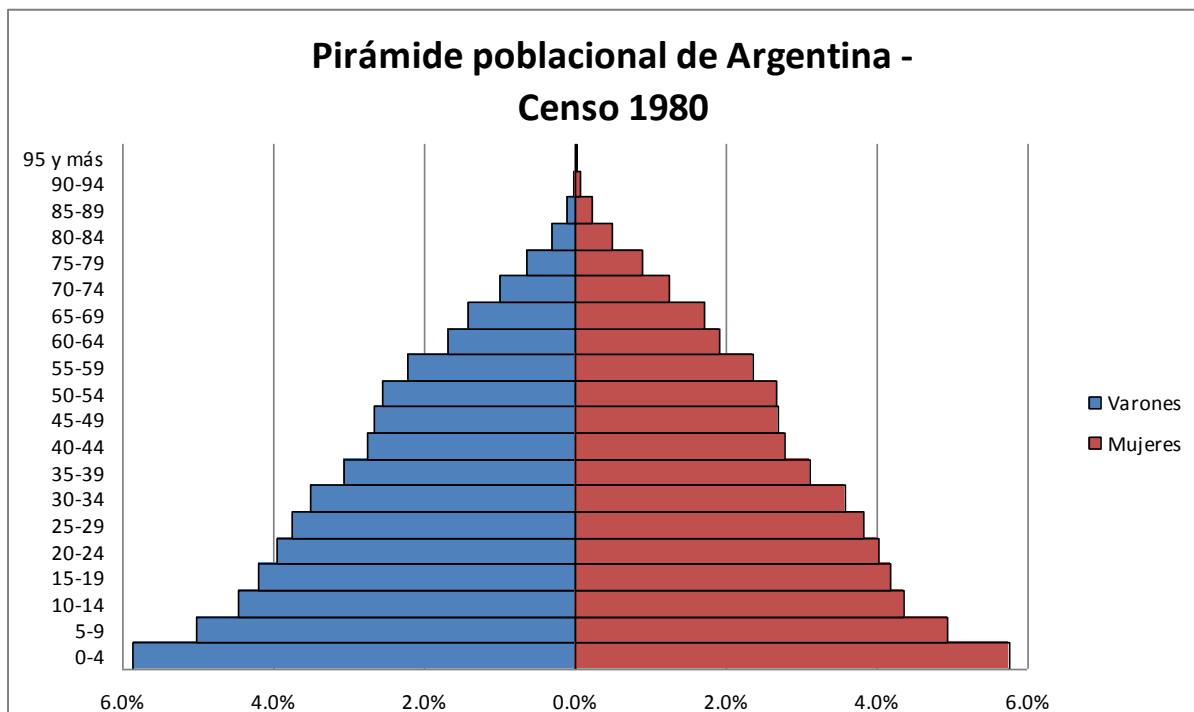
- Estructura de la población según sexo y edad

En primer lugar veremos la evolución de la población a lo largo de los años. Se presentaran cuatro gráficos de pirámide que permiten observar las diferencias entre las poblaciones de los diferentes censos.

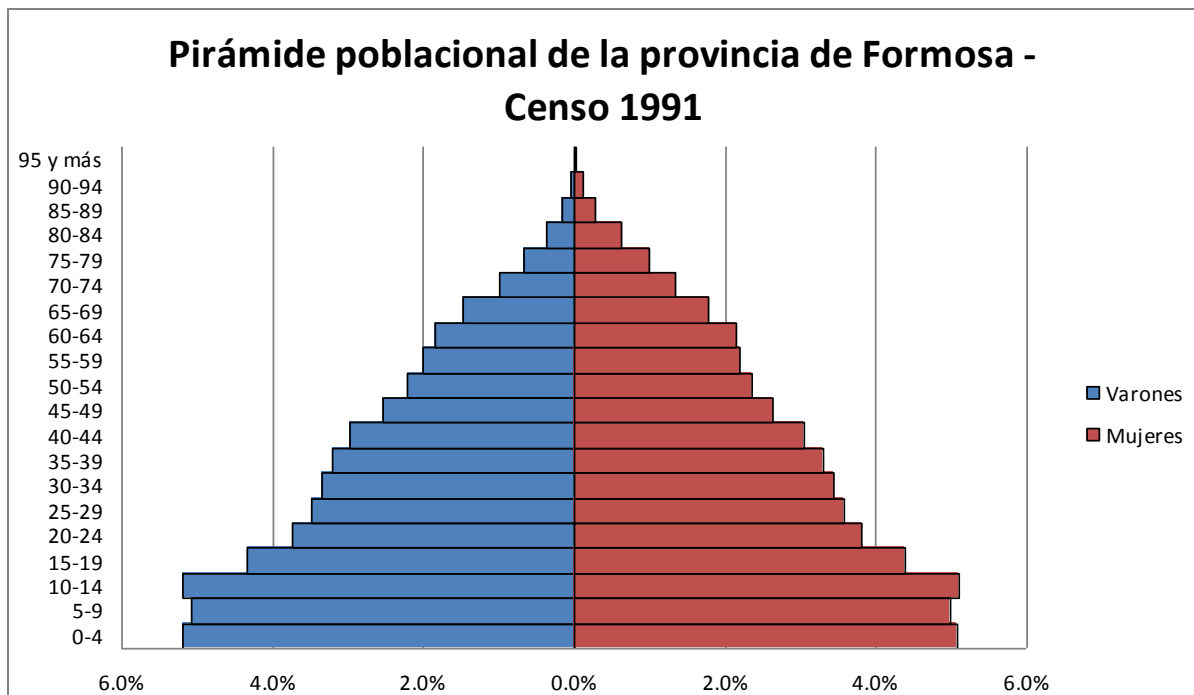
- Gráfico 3. Pirámide poblacional, Censo 1970.



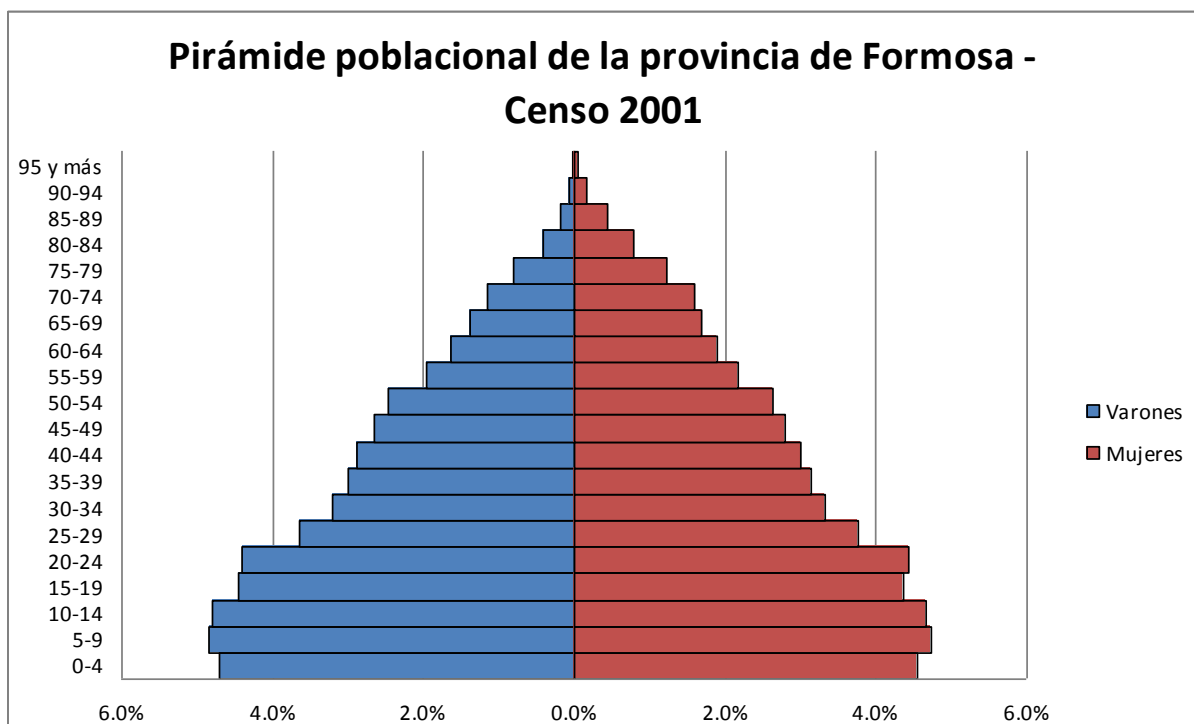
- Gráfico 4. Pirámide poblacional, Censo 1980.



- Gráfico 5. Pirámide poblacional, censo 1991



- Gráfico 6. Pirámide poblacional, censo 2001



- Diferencia de edad entre cónyuges

Este análisis consiste en ver las diferencias de edades que hay entre los cónyuges de un hogar. En una tabla, mostraremos estadísticos que describen esta relación. Los utilizados son la media, la mediana y los percentiles 1, 25, 75 y 99. Al tener las bases de microdatos en una sola tabla, el siguiente cuadro es de fácil construcción.

- Cuadro 8. Estadísticos descriptivos de las diferencia de edad entre cónyuges.

Año	Media	Mediana	Pct 1	Pct 25	Pct 75	Pct 99
1970	4.01	4	-13	1	7	23
1980	4.08	4	-15	1	7	27
1991	3.36	3	-13	0	6	22
2001	3.05	3	-12	0	6	21

Fuente: IPUMS. Muestra de los censos de población de la República Argentina. Años 1970, 1980, 1991 y 2001.

- Cantidad de personas por habitación (a nivel de hogar).

Para graficar la distribución del indicador *Personas por Habitación*, se realiza un gráfico de Tallo y Hoja por Censo de Población. Para ello, a cada hogar se le calculó el indicador:

Cantidad de personas / Cantidad de habitaciones

donde, en cantidad de habitaciones no se incluye baño y cocina. Este análisis requiere de una base de microdatos, facilitándose la presentación de los resultados al tener todas las muestras en un mismo archivo.

El gráfico de Tallo y Hoja, refina lo que puede mostrar un Box Plot, al mostrar los posibles valores modales.+

Este gráfico fue realizado con el software IBM – SPSS. Observar que el software automáticamente omite del gráfico, lo que considera *valores outliers*.

- Gráfico 10. Cantidad de personas por habitación por casa según Censo 2001

Proporción - Gráfico de tallo y hoja
Año= 2001

Frecuencia	Tallo y Hoja
177.00	0 . &
5674.00	1 . 4&
36220.00	2 . 000055555555555&
52721.00	3 . 333333333333333337&
22882.00	4 . 0000000024&
97069.00	5 . 0000000000000000000000000000000000077&
100288.0	6 . 000000002666666666666666666666666666666&
45839.00	7 . 11555555555555555&
33693.00	8 . 000000000033335&
15.00	9 . &
206882.0	10 . 000&
3022.00	11 . 6&
42438.00	12 . 0000555555555555&
65885.00	13 . 33333333333333333333339&
187.00	14 . &
51887.00	15 . 000000000000000000000000&
38909.00	16 . 0666666666666666&
11952.00	17 . 555559&
256.00	18 . &
.00	19 .
72539.00	20 . 00000000000000000000000000000000&
101.00	21 . &
3110.00	22 . 5&
10981.00	23 . 3333&
279.00	24 . &
19932.00	25 . 00000000&
4894.00	26 . 66&
4.00	27 . &
74503.00	Extremos (>=2.75)

Representación del tallo: 0.10
Representación de la hoja: 2201 caso(s)

& - hojas con menos de 2201 casos

Fuente: IPUMS. Muestra del Censo de Población, Hogares y Viviendas. Año 2001

La construcción de los gráficos presentados anteriormente, se realizó utilizando el programa estadístico *SPSS*. Se tomó como tallo a los números enteros de la proporción y al primer decimal. En la hoja se representa la centésima de la proporción. En la parte inferior del gráfico se puede ver cuantas repeticiones representa cada hoja del gráfico, y se simboliza con el símbolo & a toda aquella centésima que aparece menor cantidad de veces que las que representa la hoja. En la parte izquierda del gráfico, tenemos la frecuencia con la que se encuentra cada uno de los tallos.

En la parte inferior del gráfico se puede ver el valor a partir del cuál el programa considera *extremos*.

Se puede observar en los *Gráficos 7 y 8* que la moda es de una persona por habitación, seguida en mayor frecuencia por 2 personas por habitación.

En el *Gráfico 9* también los valores más frecuentes son 1 y 2, pero, en cambio, el tercer valor más frecuente es en este caso 0.6.

Por último, en el *Gráfico 10*, los valores más frecuentes se encuentran en 1, 0.6, 0.5 y 2.

A modo de resumir estadísticamente los gráficos anteriores, presentaremos un cuadro con descriptivos de la proporción.

- Cuadro 9. Estadísticos descriptivos de la condición de hacinamiento

Año	Media	Mediana	Pct 1	Pct 25	Pct 75	Pct 99
1970	1.72	1.33	0.25	1.00	2.00	6.00
1980	1.78	1.33	0.25	0.86	2.00	7.00
1991	1.53	1.00	0.25	0.75	2.00	5.00
2001	1.28	1.00	0.20	0.67	1.50	4.50

Fuente: IPUMS. Muestra del Censo de Población, Hogares y Viviendas. Años 1970, 1980, 1991 y 2001.

4. CONCLUSIONES

La idea principal de este trabajo era mostrar la utilidad de las bases muestrales de los censos alojadas en el sitio del IPUMS, que a diferencia de las bases de microdatos censales completas, pueden ser procesadas en una PC aun combinando diferentes años y países. Las mismas poseen un tamaño lo suficientemente grande como para estimar con un error mínimo distintas características de la población (tanto a nivel país como dentro de cada provincia). El sitio IPUMS contiene también toda la documentación necesario en lo que se refiere al *diseño de registro de las bases*, así como los conceptos precisos medidos mediante cada pregunta lo que facilita la tarea del investigador a la hora de comprar diferentes censos de un mismo país o de diferentes países.

En los primeros cuadros se comprobó que las bases muestrales eran *representativas* de la población, calculando la cantidad de población alfabetizada de 10 años o más y la cantidad de población extranjera según la base muestral del año 2001 y comparándola con los mismos datos obtenidos del Censo 2001. También se realizó un análisis de preferencia de dígitos, en el cual se vio que las muestras permiten su cálculo correcto.

Estos análisis se realizan fácilmente con las bases de microdatos. Otros, los comparativos, son sencillos al tener todas las bases en un mismo archivo, evitándole al analista la tarea de construir cuadros a partir de datos agregados localizados en diferentes páginas o sitios web.

Desde ya, los resultados son *estimaciones*, no los datos oficiales que solo INDEC en el caso de Argentina o los Institutos de Estadística publican. Pero las estimaciones poseen la suficiente precisión que en general el analista requiere, dado el tamaño de las muestras, para sugerir hipótesis de trabajo, que luego deben ser corroboradas con los datos oficiales.

5. REFERENCIAS

INDEC; Datos censales del Censo 2001, obtenidos de la página web del Instituto Nacional de Estadística y Censos de Argentina.

IPUMS; Bases muestrales de los censos de los años 1970, 1980, 1991 y 2001, obtenidas de la página web del Integrated Public Use Microdata Series de la Universidad de Minnesota.

PAS (Population Analysis System); Conjunto de macros de Microsoft Excel desarrollados por E. Arriaga en el U.S. Census Bureau que contienen procedimientos y métodos frecuentemente utilizados para análisis demográficos.

MYERS, Robert; 1960, "Error and bias in the reporting of ages in census data", en U.S. Bureau of the Census, en Handbook of Statistical Methods for Demographers, Transactions of the Actuarial Society of America, 41 parte II (104), Washington, D.C.