Using administrative records for population estimates

A first diagnostic from first experiences

Gladys Massé & Iván Williams
(gmasse@indec.gob.ar & iwilliams@indec.Gob.ar)
Summary

• Background

• First experiences

• Conclusions

• Long-Term Project and further work
Background

• COVID-19 pandemic postponed the completion of the National Population, Household and Housing Census of the 2020

• This situation accelerated the exploration of different alternatives:
  • Digital census
  • Using administrative registers

• The National Institute of Statistics and Censuses (INDEC) started to work on this topic in 2016.

• Long-Term project: National Population Register.

But in the meantime, some important progress must be made and demographic production can be useful.
Experiences (1)

Tax and Social Information System

• **Multi-sectorial register:** provide socio-economical information across governmental agencies.

• **Bank information** as the best population size fit for ages 18+, compared with the Pilot Census Test in the province of Misiones.

• **Similarity test in addresses** from a microdata sample:
  • Five different addresses from 5 different administrative registers.
  • 90% of people had **two** matching addresses within 5 sources.
  • 60% of people had **all** matching addresses within 5 sources.
Experiences (2)

Federal Administration of Public Revenue

- **Tax register** ~ 52 MM registers.

- **Disparate codification** on addresses and administrative areas.

- **Enrich register data** with census field data from another Pilot Test done in 2019 in the city of Humahuaca, Jujuy:
  - Match variables: ID number, age, birth date, sex, name.
  - 80% was matched by ID number.
  - 70% were effective matches with all variables.

Frequency of “Humahuaca” name in city field

<table>
<thead>
<tr>
<th>City</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>humahuaca</td>
<td>15.325</td>
</tr>
<tr>
<td>ciudad humahuac</td>
<td>218</td>
</tr>
<tr>
<td>1032411 humahua</td>
<td>49</td>
</tr>
<tr>
<td>humauaca</td>
<td>6</td>
</tr>
<tr>
<td>humahuca</td>
<td>4</td>
</tr>
<tr>
<td>humahuaca0</td>
<td>3</td>
</tr>
<tr>
<td>32411n humahuac</td>
<td>2</td>
</tr>
<tr>
<td>dpto. humahuaca</td>
<td>2</td>
</tr>
<tr>
<td>humahuaca2</td>
<td>2</td>
</tr>
<tr>
<td>1032405 humahua</td>
<td>1</td>
</tr>
<tr>
<td>1032411-humahua</td>
<td>1</td>
</tr>
<tr>
<td>1032411-humauac</td>
<td>1</td>
</tr>
<tr>
<td>1032411 humauca</td>
<td>1</td>
</tr>
<tr>
<td>1032412 humahua</td>
<td>1</td>
</tr>
<tr>
<td>1032414 humahua</td>
<td>1</td>
</tr>
<tr>
<td>1032419 humahua</td>
<td>1</td>
</tr>
<tr>
<td>32411 humahuaca</td>
<td>1</td>
</tr>
<tr>
<td>el aguilar huma</td>
<td>1</td>
</tr>
<tr>
<td>huamahuaca</td>
<td>1</td>
</tr>
<tr>
<td>humahua</td>
<td>1</td>
</tr>
<tr>
<td>humahuac</td>
<td>1</td>
</tr>
<tr>
<td>humahuaca4630</td>
<td>1</td>
</tr>
<tr>
<td>humahuacva</td>
<td>1</td>
</tr>
<tr>
<td>humahuaca</td>
<td>1</td>
</tr>
<tr>
<td>humahuaca</td>
<td>1</td>
</tr>
<tr>
<td>zo) humauaca</td>
<td>1</td>
</tr>
</tbody>
</table>
Experiences (3)

Electoral Register

- Not included: ages<16 & foreigners.
- Male/Female ratio over the expected from population projection.
- Excess of adult and older population (specially centenarians).
- Enrich register data with Census Pilot Test done in 2019 in Humahuaca, Jujuy:
  - 70% had an effective match.
  - 80% of matched effectively had the census response in same province and city where they vote.
Experiences (4)

National Register of Persons

• Agency that assigns an ID to newborns and immigrants.

• 47,5 MM (at Apr-2021) vs 45 MM from Projection (at Jun-2021).

• Clusters in ID number related to Age:
  • Strong correlation until age 70.
  • Foreigners group: ID>90 MM.
  • Outlier groups.

• Births and deaths with approx. 1 year lag:
  • Federal register system issues.
  • Pandemic situation.
First Conclusions

• Overcount in total Population and Older aged. Undercount in Children.

• Undetermined lag between event and registry.

• Excess males than expected (selected migration?).

• Disparate codification of addresses and administrative geographical areas.

• 70% of effective matches with census tests (on average).

Positive aspects:

• We have a brief diagnostic from the main administrative registers in Argentina.

• We found interest from administrative register producers to work together and start a joint project.
General Project

- Population Estimate ($\hat{N}$): population estimate by age, sex and place of residence based on RAEP + Other sources.
- Statistical Population Register (REP): complete, continuous and updated list of people residing in the country.
- Record-Based Census ($N$): counting people by age, sex, place of residence based on REP.
Next Steps

• Characterize in depth each administrative register
• Diagnose strengths and weaknesses to have information about residential population
• Generate the first experimental national population estimates

We have a necessary and super-interesting trip to continue!
Thanks!