

Modelación de poblaciones en América Latina y el Caribe

**Andrés Gutiérrez
2024**

Algunos desafíos en los censos de población

- Los censos de población no siempre logran enumerar todos los hogares y sus poblaciones en todo el país.
 - Omisión completa de viviendas o identificación errónea del estado de ocupación de la vivienda.
 - Omisión total o parcial de personas en el interior de las viviendas.
 - Omisión total o parcial de ciertas áreas geográficas por problemas de planificación del trabajo de campo, accesibilidad o seguridad entre otros durante la enumeración censal.
- La mayoría de los países de la región de América Latina y el Caribe están experimentando este tipo de desafíos en sus censos.

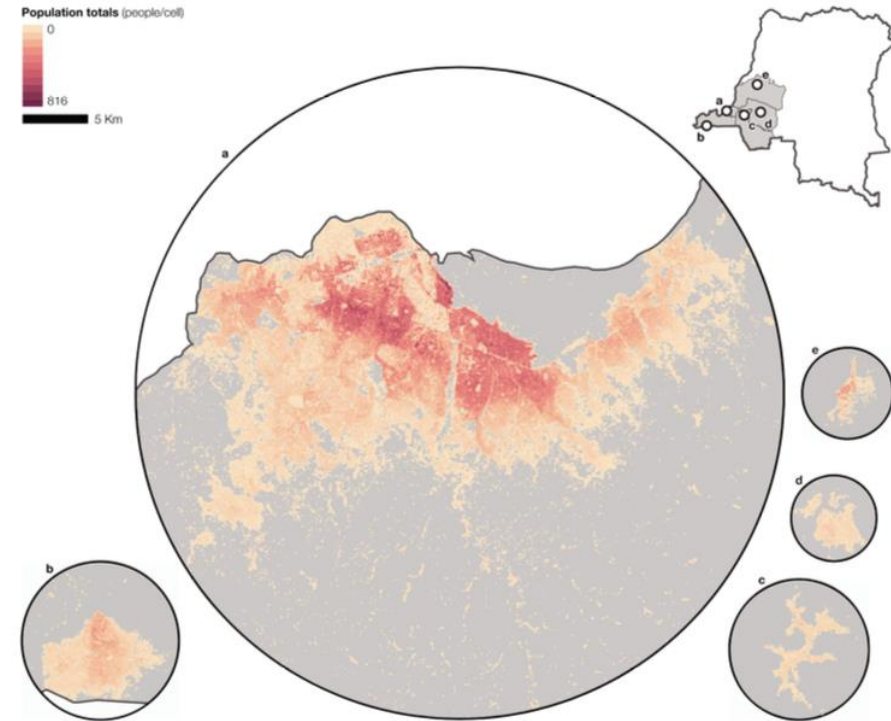
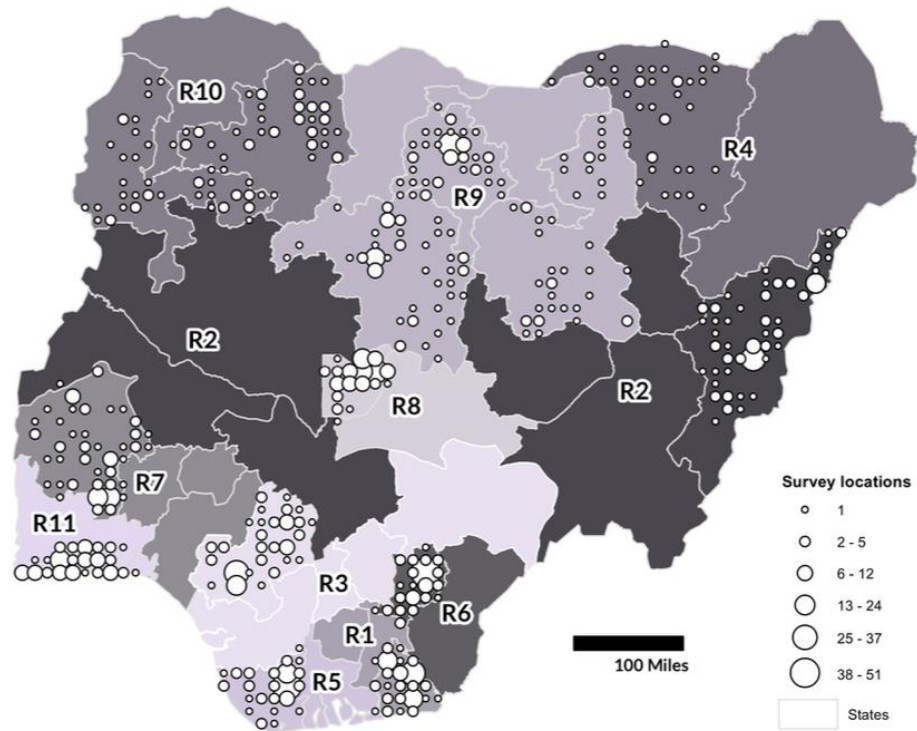
Solución parsimoniosa

- Cuando la enumeración incompleta de áreas es un problema en el censo, podemos confiar en los modelos estadísticos para predecir los recuentos de personas (junto con su estructura demográfica: edad y sexo).
 - Las estimaciones de recuentos basadas en modelos representan un nuevo enfoque para el problema de la omisión total o parcial.
 - La lógica detrás de este tipo de modelos es tomar prestada la fuerza de áreas completas.
 - Este enfoque utiliza covariables de teledetección que deberían estar disponibles para todas las zonas del país.
- En la literatura encontramos una gran cantidad de experiencias con modelos similares:
 - Boo, et. al (2022), Leasure, et. al (2020) y Berg (2023)

Modelos basados en encuestas de enumeración

NATURE COMMUNICATIONS | <https://doi.org/10.1038/s41467-022-29094-x>

ARTICLE



Modelos poblacionales de la CEPAL y el UNFPA

- Estos modelos estadísticos relacionan los datos de población observados del censo con otros conjuntos de datos con el fin de predecir la población en áreas donde la información censal es incompleta.
 - Se diseñan específicamente para cada país en función de los insumos disponibles y los objetivos previstos.
 - Los modelos pueden diseñarse para realizar estimaciones a nivel de cuadrícula (1 km, 100 m, etc.), sectores estadísticos u otros niveles geográficos o administrativos, en función de las necesidades y de la calidad y cantidad de la información disponible.

Características principales

- Nuestros modelos poblacionales tienen tres características:
 - Son bayesianos para poder añadir información previa a las áreas observadas.
 - Son modelos mixtos que incorporan la heterogeneidad en áreas no observadas.
 - Las covariables siempre incluyen imágenes satelitales (luces, huellas de edificios), información geoespacial (carreteras, infraestructura) o variables cartográficas.



El modelo GLMM de Poisson para los recuentos

Definimos el modelo de Poisson GLMM como en Berg (2022)

$$y_{ij} \mid \mu_{ij} \sim \text{Poisson}(\mu_{ij})$$
$$\mu_{ij} = N_j D_{ij}$$

Donde y_{ij} representa el número de personas en la vivienda i y el distrito de enumeración j . N_j es el número de viviendas en el segmento censal j . Además, D_{ij} es la densidad media de la vivienda y se relaciona con la media a través de la siguiente función de enlace:

$$\log(D_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta} + u_j$$

Información previa y distribución posterior

Las distribuciones previas para β y γ son las siguientes:

$$\begin{aligned}\beta_p &\sim \text{Normal}(0, 10000) \\ u_j &\sim \text{Normal}(0, \sigma_u^2) \\ \sigma_u^2 &\sim \text{Inverse} - \text{Gamma}(0.0001, 0.0001)\end{aligned}$$

Por lo tanto, el estimador bayesiano para el número de personas en la vivienda i del segmento censal j es:

$$\tilde{\theta}_{ij} = E(y_{ij} \mid \mu_{ij})$$

Enfoque predictivo

De esta manera, el predictor bayesiano propuesto viene dado por la siguiente expresión:

$$\hat{t}_y = \sum_{\substack{SC \text{ completos} \\ SC \text{ incompletos}}} \sum_{\substack{Viviendas \text{ Completas} \\ Viviendas \text{ incompletas}}} y_{ij} + \tilde{\theta}_{ij}$$

Esta expresión es similar a la de Molina y Rao (2010) Empirical Best Predictor en el contexto de mapas de pobreza y modelos de estimación de áreas pequeñas.

Asistencia técnica en la región

- La CEPAL y el UNFPA han aunado esfuerzos en los últimos dos años y han beneficiado a los siguientes países:
 - Costa Rica
 - Ecuador
 - República Dominicana
- Actualmente estamos trabajando con los siguientes países:
 - Barbados
 - Guyana
 - Jamaica

Imágenes satelitales (nivel de SC)

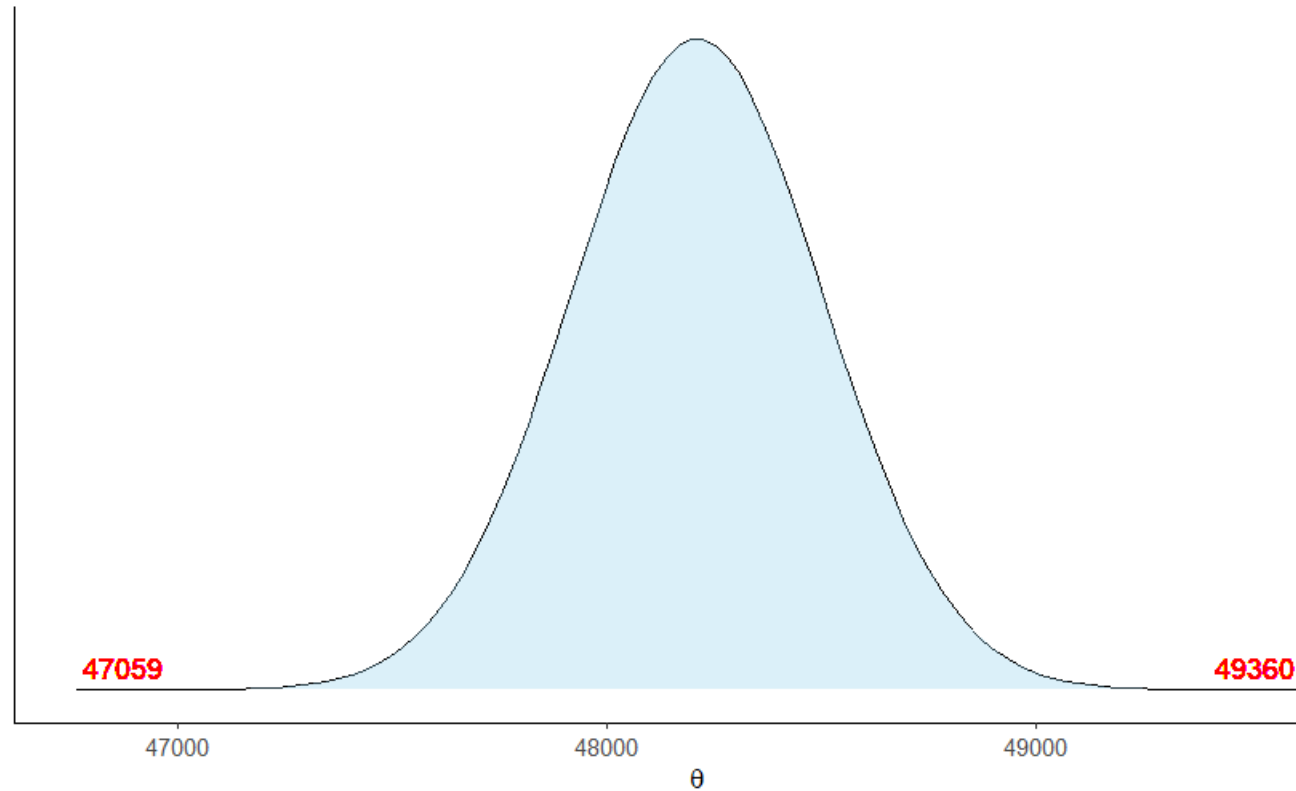
- Accedemos a esta información a través de Google Earth Engine, que brinda facilidades para analizar y obtener estos datos a través de los lenguajes de programación Javascript y Python, y recientemente desde 2021 en R con el paquete rgee.
- Entre las principales ventajas de la información basada en teledetección se encuentra la facilidad de acceso a datos con una cobertura geográfica profunda que es imposible obtener por medios tradicionales como encuestas o registros administrativos.
- *Huellas de edificios*
- *Proyecciones de WorldPop*
- *Fracción de cobertura urbana*
- *Fracción de cobertura rural*
- *Crops_cover fracción*
- *Altitud en metros sobre el nivel del mar*
- *Tiempo de viaje hasta el centro médico más cercano*
- *Tiempo de viaje a la escuela más cercana*

Datos administrativos (municipales)

- En cada país, se puede encontrar información valiosa en los registros administrativos.
- Además, podemos encontrar covariables importantes en el censo más reciente junto con datos cartográficos disponibles en el ONE.
- *Acceso a las telecomunicaciones*
- *Problemas de acceso*
- *Altos índices de criminalidad*
- *Matrícula en la enseñanza primaria*
- *Asentamientos informales*
- *Zona indígena*
- *Zona protegida*

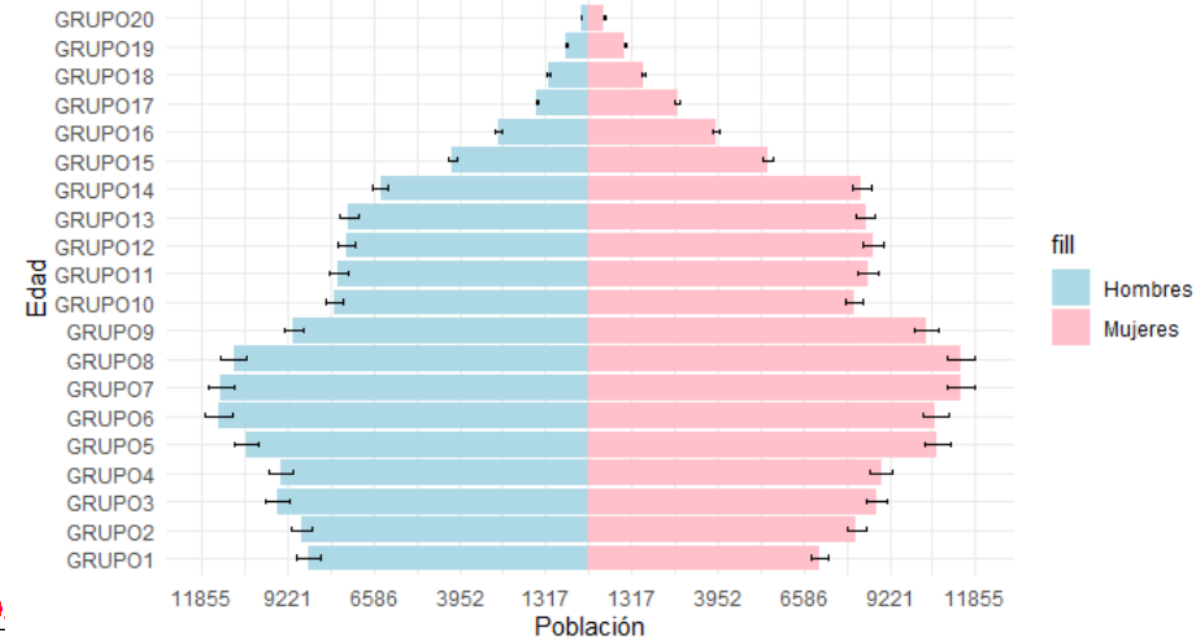
Estimaciones municipales

DIST_ID = 10110



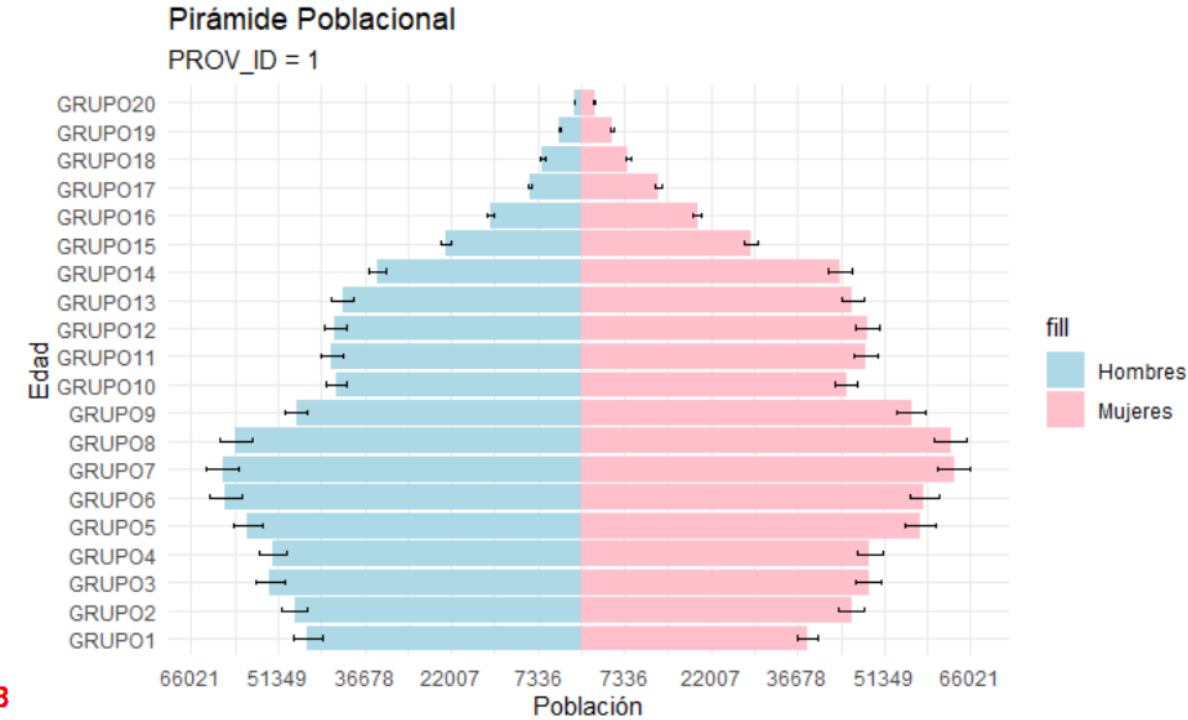
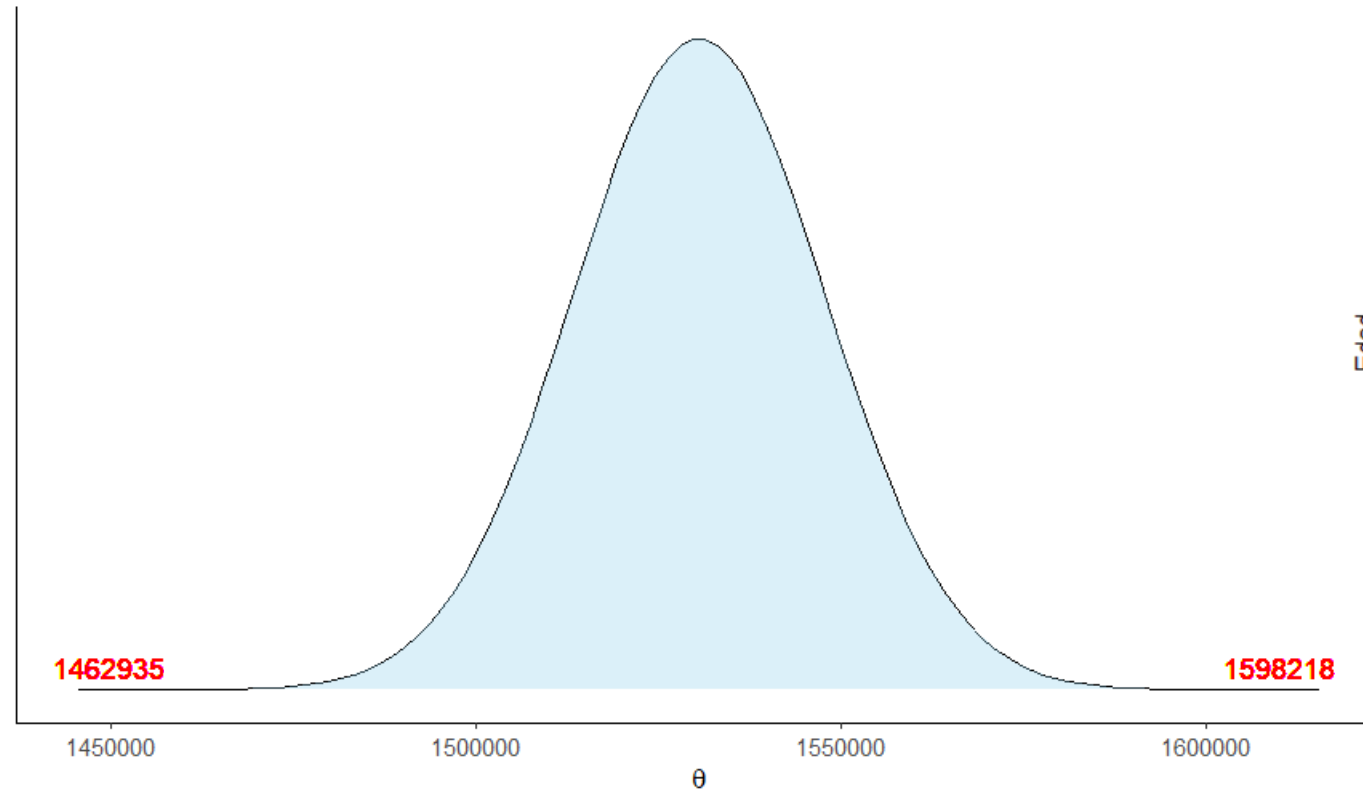
Pirámide Poblacional

CANT_ID = 101

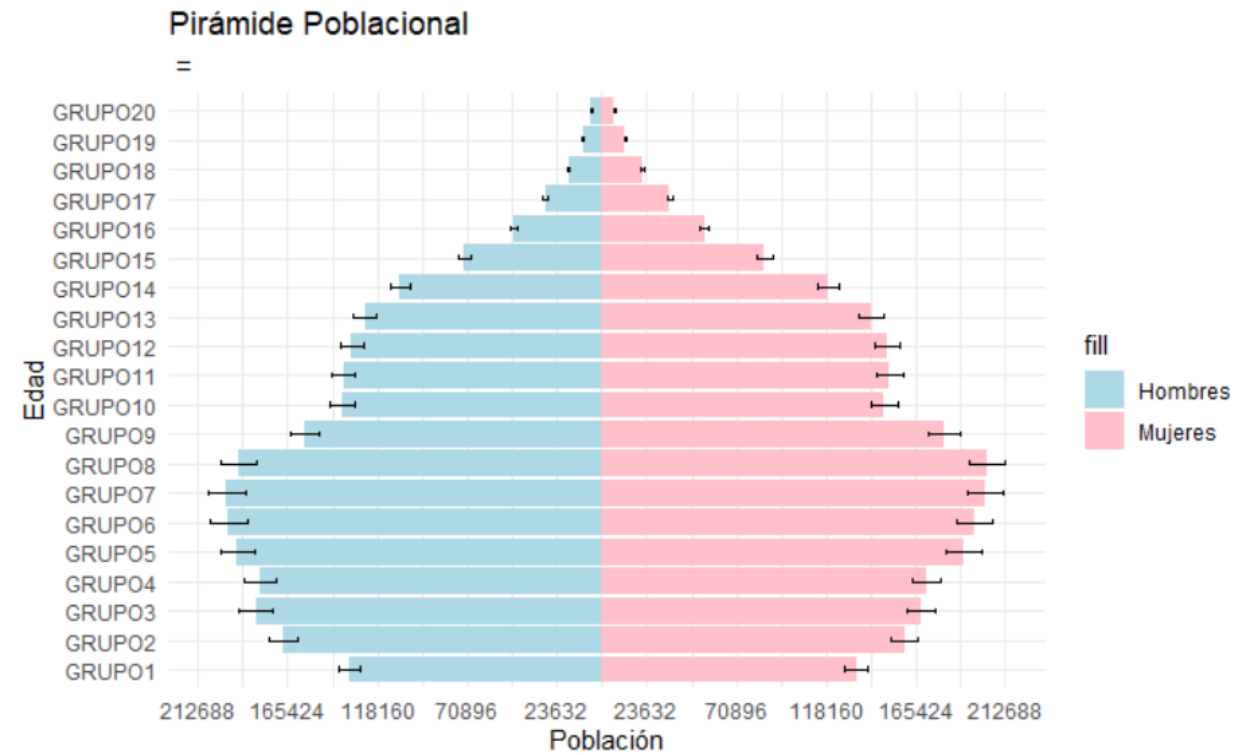
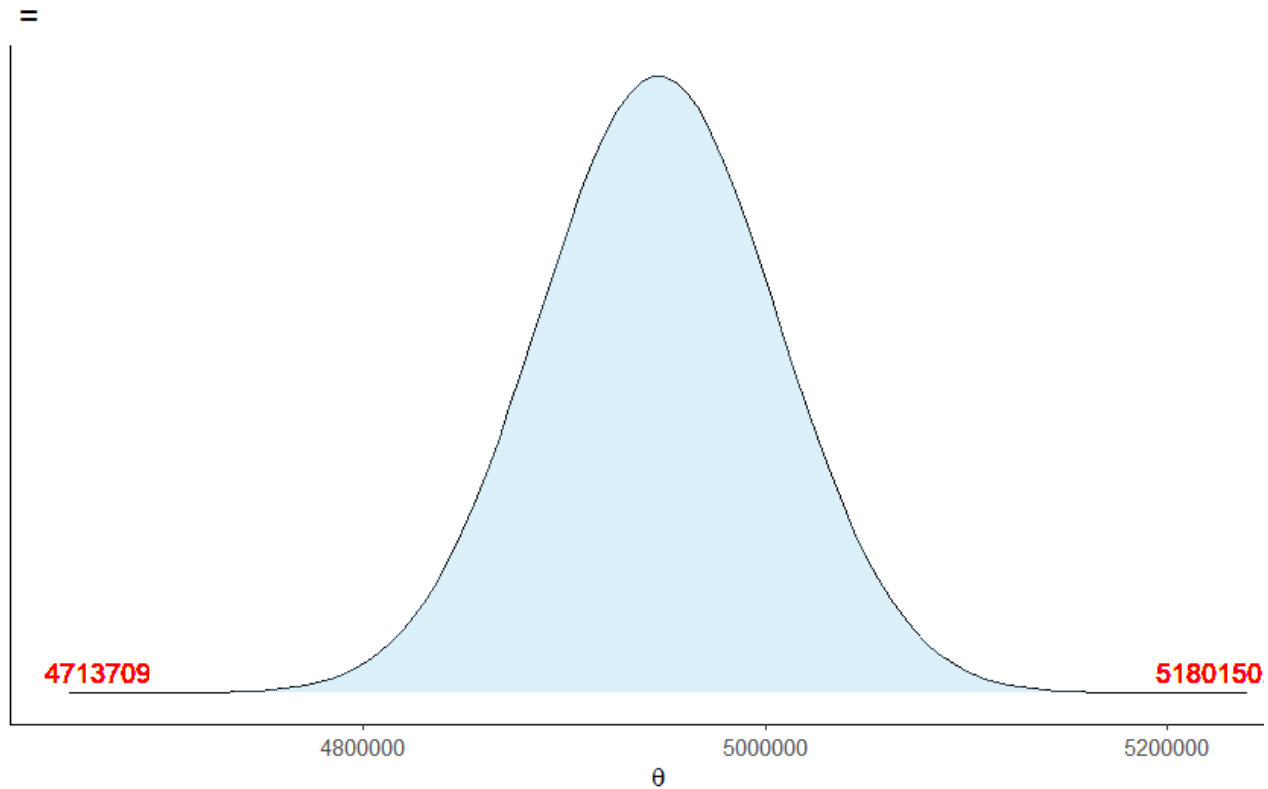


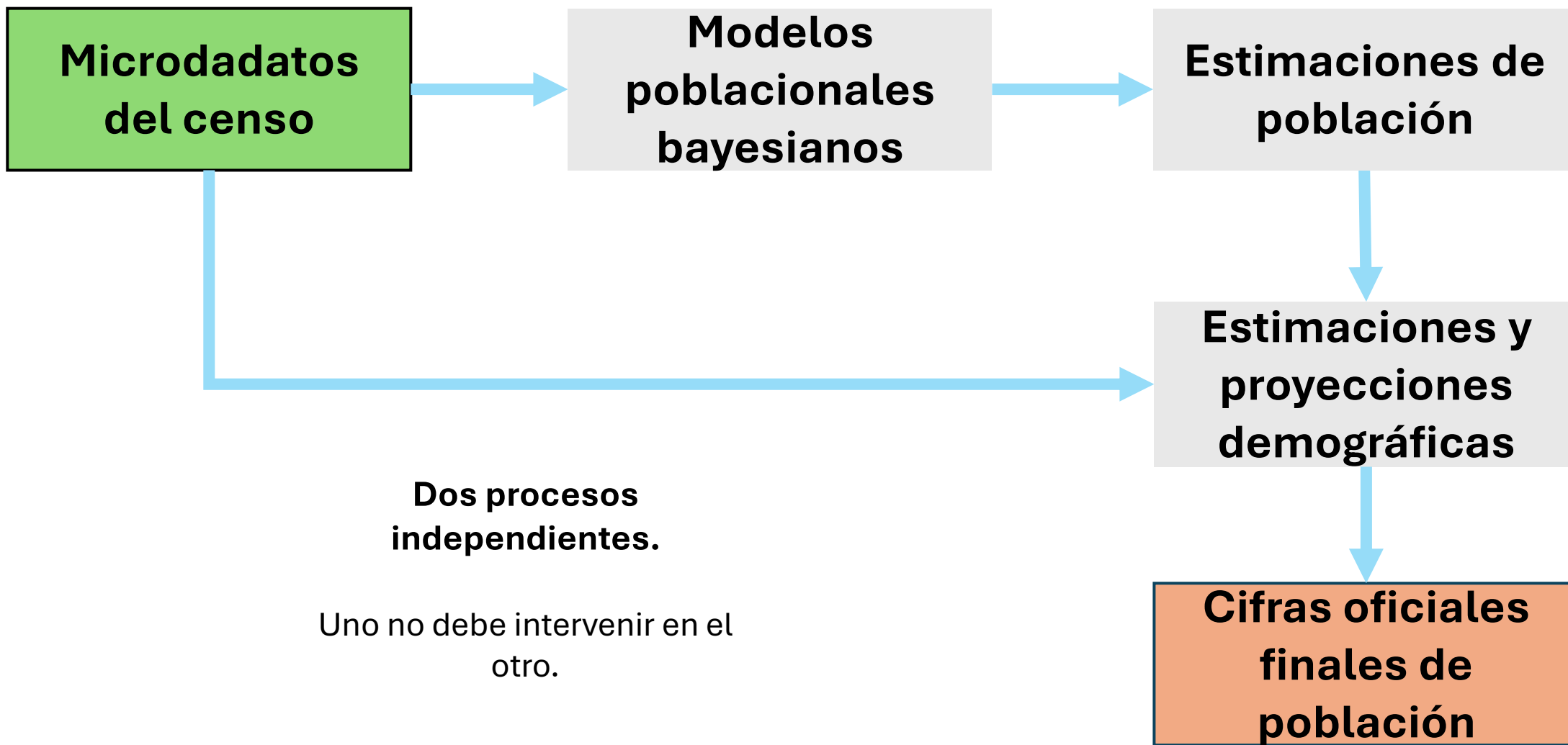
Estimaciones estatales

PROV_ID = 1



Estimaciones nacionales





Gracias