



Nota técnica sobre la distribución de población y viviendas por distrito

Noviembre, 2023

1. Contexto

Debido a la cobertura parcial del Censo 2022, el Instituto Nacional de Estadística y Censos (INEC) publicó, previamente a este documento, una estimación de población al 30 de junio de 2022, desagregada a nivel de provincia y cantón por sexo, con su correspondiente metodología de cálculo.

Dada la relevancia para el país y para las instituciones públicas y privadas de contar con datos desagregados a nivel de distrito, se implementó una metodología para distribuir la población estimada a nivel de distrito, utilizando la información del Censo 2022 y distintos registros administrativos.

En este documento se detalla la metodología implementada para realizar la distribución en los 487 distritos que, hasta diciembre de 2020, eran los oficiales según la división territorial administrativa del territorio nacional. Dicha fecha se estableció como corte para la actualización cartográfica realizada para el operativo de campo del Censo 2022.

Es importante mencionar que los datos desagregados a nivel de distrito no modifican la estimación a nivel de cantón publicada; dado que corresponde a una distribución de esa población cantonal y no a una nueva estimación. A continuación, se explica en detalle el procedimiento implementado para distribuir la población y las viviendas por distrito.

2. Distribución de la población por distrito

Con el objetivo de distribuir a nivel distrital la población estimada para los cantones del país, se implementó una metodología que aplica una estructura modelo que determina el peso de cada distrito en el total del cantón, el cual se utiliza como factor para obtener un total de población para cada distrito.

Para establecer esa estructura modelo se analizaron en detalle varias estructuras que se basan en diferentes fuentes y con eso, se logró determinar aquella que distribuía la población de forma más consistente entre los distritos de cada cantón. Las fuentes utilizadas se describen a continuación:

2.1. Fuentes de datos utilizadas

Las fuentes utilizadas para establecer las estructuras que se analizaron fueron las siguientes:

- ◆ Nacimientos ocurridos y registrados en el país en 2019 (periodo pre-pandemia) y 2022 (periodo post-pandemia) por distrito de residencia de la madre y sexo de la persona recién nacida.
- ◆ Defunciones ocurridas y registradas en el país en 2019 (periodo pre-pandemia) y 2022 (periodo post-pandemia) por distrito de residencia y sexo de la persona fallecida.
- ◆ Modelo Bayesiano para estimación de población: corresponde al mismo modelo estadístico predictivo para la estimación de población que se utilizó para la estimación por cantón. Para más detalles sobre este modelo puede consultar el documento Estimación de población y vivienda 2022. Metodología, disponible en la página web del INEC¹.
- ◆ Padrón electoral al 30 de junio de 2022: corresponde a todas las personas mayores de 18 años empadronadas por distrito de residencia.

Para esta fuente se realizó un proceso de ajuste al lugar de residencia debido a que no todas las personas actualizan esa información en su cédula cuando cambian de residencia. Para ese ajuste se empataron las personas registradas en el Censo 2022 y el padrón² a partir del número de cédula, lo que permitió identificar a dónde residía la persona al momento del censo y si coincidía o no con la residencia establecida en el padrón. Luego, se realizó una matriz de residencia, de manera que se obtuvo el peso de los casos en que la residencia registrada en el padrón variaba con respecto al censo y se determinó el distrito al cual correspondía su residencia más actual, con esta razón relativa se corrigió el padrón electoral y se consideró como una quinta fuente a utilizar.

¹ Puede acceder al documento en el siguiente link: en el siguiente link:

<https://admin.inec.cr/sites/default/files/2023-08/mePoblacMetodologiaEstimacionPoblacionVvivienda2022.pdf>

² Empataron el 70,8% de las cedulas registradas contra el padrón electoral.

2.2. Estructuras analizadas

A partir de las fuentes mencionadas se generaron cinco estructuras diferentes que fueron analizadas. Para obtener cada estructura se realizó el siguiente procedimiento:

1. Se obtiene la distribución absoluta de la población por distrito por sexo de la fuente seleccionada.
2. Se calcula el peso relativo de cada distrito por sexo con respecto al total por sexo de su cantón.

Con ese procedimiento se obtiene la información por sexo para cada estructura, es decir, dos distribuciones por cada una. Las estructuras generadas y analizadas son las siguientes:

- a) **Estructura 1:** considera el promedio de las estructuras de los nacimientos y las defunciones ocurridos y registrados en 2019 y 2022, del padrón electoral al 30 de junio de 2022 y la estructura derivada del modelo Bayesiano.

$$\sum(N_{19} + N_{22} + D_{19} + D_{22} + P_{22} + B_{22})/6$$

Donde:

- ◆ N19= Estructura de los nacimientos ocurridos y registrados en 2019.
- ◆ N22= Estructura de los nacimientos ocurridos y registrados en 2022.
- ◆ D19= Estructura de las defunciones ocurridas y registradas en 2019.
- ◆ D22= Estructura de las defunciones ocurridas y registradas en 2022.
- ◆ P22= Estructura de las personas inscritas en el padrón electoral a 2022.
- ◆ B22= Población residente predicha por el modelo Bayesiano al 2022.

- b) **Estructura 2:** considera el promedio de las estructuras de los nacimientos y las defunciones ocurridos y registrados en 2019 y 2022, del padrón electoral al 30 de junio de 2022 ajustado por lugar de residencia y la estructura derivada del modelo Bayesiano.

$$\sum(N_{19} + N_{22} + D_{19} + D_{22} + Pa_{22} + B_{22})/6$$

Donde:

- ◆ N19= Estructura de los nacimientos ocurridos y registrados en 2019.
 - ◆ N22= Estructura de los nacimientos ocurridos y registrados en 2022.
 - ◆ D19= Estructura de las defunciones ocurridas y registradas en 2019.
 - ◆ D22= Estructura de las defunciones ocurridas y registradas en 2022.
 - ◆ Pa22= Estructura de las personas inscritas en el padrón electoral ajustado 2022.
 - ◆ B22= Estructura de la población residente predicha por el modelo Bayesiano al 2022.
- c) **Estructura 3:** utilizó como fuente únicamente, la estructura del padrón electoral ajustado.
- d) **Estructura 4:** utilizó como única fuente la estructura de población derivada del modelo Bayesiano.
- e) **Estructura 5:** considera el promedio de las estructuras del padrón ajustado por lugar de residencia y la derivada del modelo Bayesiano.

$$\sum(Pa_{22} + B_{22})/2$$

Siendo:

- ◆ Pa22= Estructura de las personas inscritas en el padrón electoral ajustado 2022.
- ◆ B22= Estructura de la población residente predicha por el modelo Bayesiano al 2022.

Una vez que se obtuvo la estructura de población de cada sexo se multiplica el peso relativo de cada distrito al total por sexo de la población estimada para el cantón³. Posteriormente se obtiene el total del distrito con la sumatoria de los dos sexos.

El análisis de los resultados que se explica a continuación se realizó para los datos absolutos totales obtenidos de aplicar cada una de las estructuras anteriormente detalladas.

³ Estimación publicada el 20 de julio de 2023.

2.3. Análisis de resultados y definición de una estructura modelo

Con el objetivo de validar las estructuras y definir la mejor para la desagregación de la población del cantón por distritos; se estimaron las poblaciones por distrito y los resultados obtenidos se compararon con diversos indicadores demográficos para todos los cantones y sus distritos. Para ello se elaboró una tabla de validación con la siguiente información:

- ◆ Total de viviendas individuales ocupadas y desocupadas de los censos 2000, 2011 y 2022.
- ◆ Total de viviendas identificadas en el levantamiento cartográfico para el Censo 2022.
- ◆ Población por distrito según los censos 2000, 2011 y Censo 2011 ajustado por omisión.
- ◆ Proyección de población vigente para el año 2022.
- ◆ Población censada en 2022.
- ◆ Población según el registro único de población (RUP).
- ◆ Crecimiento anual de la población intercensal 2000-2011.
- ◆ Promedio de habitantes por vivienda según el Censo 2011.
- ◆ Razón de nacimientos y defunciones.
- ◆ Crecimiento natural o vegetativo de la población por distritos.
- ◆ Porcentaje de población extranjera según el Censo 2011 y Censo 2022.

Seguidamente con los datos anteriores se calcularon, para cada una de las cinco estructuras, los siguientes indicadores:

- ◆ Crecimiento anual de población obtenida en cada estructura respecto al Censo 2011 y al Censo 2011 ajustado por omisión.
- ◆ Promedio de habitantes por vivienda para cada estructura.
- ◆ Comportamiento de las estimaciones de las cinco estructuras, respecto a los intervalos que estableció el modelo Bayesiano.

Una vez que se contó con estos insumos, se procedió con la revisión de cada estructura y determinar para cada uno de los distritos la estimación que mejor se ajustaba a su comportamiento, se analizó visualmente también los indicadores a través de gráficos para evaluar la tendencia de cada distrito. Por otra parte, resultó necesario establecer reuniones

de análisis de los comportamientos con personal experto en demografía y en censos, para discutir aquellos casos donde los resultados de las estimaciones en los distintos modelos no ajustaban acorde a los indicadores analizados. Luego de haber hecho esto, se delimitó el análisis a las dos estructuras en donde la estimación y distribución de los datos en los distritos mostraban tendencias similares a los indicadores analizados, estas fueron la estructura del modelo 4 y la estructura del modelo 5.

2.4. Distribución final de la población por distrito

A partir del análisis de las estructuras detalladas en el apartado anterior, la estructura 4 se determinó como la más consistente con respecto a los indicadores revisados, excepto por 53 distritos, de seis cantones, cuyos resultados no fueron consistentes con el comportamiento de los indicadores en su estimación y distribución. Esos cantones son Desamparados, Alajuela, San Carlos, Los Chiles, Carrillo y Guácimo, los cuales tienen ciertas particularidades, como por ejemplo tener viviendas colectivas muy grandes, como centros penitenciarios o tener un porcentaje mayor al promedio nacional de población extranjera. Para estos cantones por tanto se definió la estructura 5 como la más consistente.

Por tanto, la estructura 4 (que utiliza como fuente solo el modelo bayesiano) se aplicó al 89,1% de los distritos y al restante 10,9% se aplicó la estructura 5 (promedio de las estructuras del modelo bayesiano con el padrón electoral ajustado).

Es importante mencionar que la selección de la mejor estructura modelo a utilizar, se hizo al considerar el cantón como un todo, es decir el conjunto de distritos que lo conforman; de manera que se utiliza un solo método para todos sus distritos.

3. Distribución de las viviendas por distrito

Una vez definida la población estimada por distrito, se procedió a realizar la distribución de las viviendas. Para ello, se realizó el siguiente procedimiento:

- ◆ Para las viviendas desocupadas se utilizó la misma distribución de las viviendas identificadas en el Censo 2022.
- ◆ Para las viviendas ocupadas se obtuvo la distribución relativa por distrito, según el total del cantón, de las viviendas identificadas durante la actualización cartográfica para el Censo 2022 y de la población según la distribución anteriormente explicada. A esta población se le

restó el total de personas residentes en viviendas colectivas según el Censo 2022 debido a que la distribución sería solo de viviendas individuales.

- ◆ Se calculó el promedio de las dos estructuras mencionadas en los puntos anteriores para obtener una estructura modelo.
- ◆ Esa estructura modelo se aplicó al total de viviendas estimado a nivel de cantón, el cual había sido publicado anteriormente.

Con ese procedimiento se obtuvo la distribución de las viviendas ocupadas y desocupadas por distrito. Posteriormente se realizó una validación comparando los promedios de ocupantes por vivienda del Censo 2022. Además, también se revisaron las diferencias respecto a la actualización cartográfica y para algunos cantones se logró comparar con datos obtenidos de las municipalidades.

Después de esa validación se identificaron algunos cantones que requerían una distribución diferente debido a ciertas condiciones particulares como, por ejemplo, muy altos o muy bajos porcentajes de viviendas desocupadas, así como promedios de ocupantes menores a 2 o mayores a 3,8. Esos cantones son Desamparados, Turrubares, San Carlos, La Unión, Heredia, Nicoya, Santa Cruz. Para estos casos se determinaron los distritos que debían ajustarse ya sea en el porcentaje de desocupadas o en el promedio de ocupantes y se prorratearon los datos para balancear todo el cantón.