**DIRECCIÓN DE CENSOS Y DEMOGRAFÍA**

Estimación del subregistro de homicidios basada en los datos del proyecto JEP-CEV-HRDAG

Consultoría de fortalecimiento de estadísticas derivadas de los censos con enfoque territorial

ABRIL de 2024

CONTENIDO

[1. Introducción 4](#_Toc165045644)

[2. Contexto 4](#_Toc165045645)

[2.1 Proyecto JEP-CEV-HRDAG 4](#_Toc165045646)

[2.2 Estimación del subregistro de víctimas 5](#_Toc165045647)

[3. Procesamiento 5](#_Toc165045648)

[3.1 Estratificación de los datos 5](#_Toc165045649)

[3.2 Estimación por sistemas multiples 6](#_Toc165045650)

[3.3 Combinación de estimaciones 6](#_Toc165045651)

[4. Resultados 6](#_Toc165045652)

[4.1 Distribución posterior de los estratos estimados 7](#_Toc165045653)

[5. Conclusión 9](#_Toc165045654)

[Bibliografía 9](#_Toc165045655)

Lista de tablas

[Tabla 1: Número de estratos con distribución unimodal y multimodal 7](#_Toc165044195)

Lista de gráficos

[Gráfico 1. Estrato con distribución posterior unimodal (R5-1990\_1994-90-94-HOMBRE-11) 7](#_Toc165045622)

[Gráfico 2: Estrato con distribución posterior multimodal (R1-1985\_1989-0-4-MUJER-11) 8](#_Toc165045623)

# Introducción

En el marco de la consultoría para el fortalecimiento de estadísticas derivadas de los Censos con Enfoque territorial y la Actualización de la Conciliación Demográfica vigente, este documento explica el flujo de trabajo para el procesamiento y los resultados de las estimaciones del subregistro de víctimas de homicidio. Estos procesos se llevaron a cabo con base en los datos del proyecto conjunto JEP-CEV-HRDAG que el DANE recibió en 2023.

Las siguientes secciones del documento se dividen así: la sección de contexto explica brevemente el contenido de los datos y su relevancia en el proceso de Conciliación Demográfica. La sección de procesamiento explica los procesos llevados a cabo para la estimación del subregistro. La cuarta sección ofrece un panorama de los resultados obtenidos. La última sección resume las conclusiones de este ejercicio.

# Contexto

## Proyecto JEP-CEV-HRDAG

El proyecto conjunto JEP-CEV-HRDAG se desarrolló entre el 2020 y 2022 con el objetivo de hacer análisis estadísticos del conflicto armado en Colombia para 4 violaciones a derechos humanos: homicidios, reclutamiento, secuestro y desaparición. Dentro de este proyecto, se juntaron 112 bases de datos especializadas y no especializadas en conflicto en Colombia[[1]](#footnote-1). El resultado fue un consolidado de este conjunto de datos, deduplicado y sometido a imputación estadística múltiple para las variables demográficas faltantes como sexo, etnia, edad, entre otras[[2]](#footnote-2). En calidad de autoridad estadística en Colombia, el DANE recibió esta información para garantizar su accesibilidad y aprovechamiento.

## Estimación del subregistro de víctimas

Además de permitir observar en agregado la documentación de víctimas de conflicto entre diferentes fuentes de datos, este proyecto permite hacer estimaciones del subregistro de víctimas para cada violación a derechos humanos. Esto, tomando en cuenta las limitaciones alrededor de la documentación de la violencia en Colombia. Estas se desarrollan a través de métodos de Estimación por Sistemas Múltiples (MSE) y, específicamente, el algoritmo de “Modelos de Clases Latentes Múltiples para Captura-Recaptura” (Manrique-Vallier 2016)[[3]](#footnote-3).

La inclusión de las estimaciones para homicidio en un proceso de conciliación demográfica permite aproximaciones más precisas, considerando, además, que es posible realizar estas estimaciones desagregadas por área geográfica, edad, sexo y variables temporales (en este caso, quinquenio).

# Procesamiento

## Estratificación de los datos

La primera parte del proceso de estimación consiste en la estratificación de los datos. Esto, considerando dos factores: en primer lugar, la hipótesis detrás de la estimación o la forma requerida de los datos. En segundo lugar, consideraciones sobre la heterogeneidad de captura, fenómeno que implica que la probabilidad de una víctima de ser documentada puede verse afectada por distintas variables, entre esas, sus características demográficas, el momento del tiempo en el que ocurre el hecho, el victimario, el lugar del hecho, etc.

Después de evaluados estos dos componentes, se acordó con el consultor Víctor Manuel García la siguiente estratificación:

*Quinquenio – edad por grupos quinquenales – sexo – departamento del hecho*

Esta estratificación resultó en la necesidad estimar **73.178 estratos**. Cabe resaltar que esta estratificación se hace para 10 réplicas de los datos. El código para la estratificación puede encontrarse en: <https://github.com/dtcensos/Estimacion-homicidios/blob/main/src/estratificacion.R>

## Estimación por sistemas multiples

Para la estimación de los estratos mencionados anteriormente se utilizó el paquete*verdata[[4]](#footnote-4)*en R, una herramienta creada exclusivamente para el manejo de los datos resultantes del proyecto JEP-CEV-HRDAG. Específicamente, fue necesario el uso de la función *mse* que hace uso del algoritmo *LCMCR,* también codificado en R. El proceso de estimación requiere de considerables recursos computacionales y toma tiempo. En ese sentido, la Dirección de Censos y Demografía hizo uso de 10 de sus equipos para correr las estimaciones de algunos estratos y requirió el apoyo de la Oficina de Sistemas para estimar los demás. Este proceso se llevó a cabo desde el 8 de marzo hasta el 18 de abril, con un procesador CPU de 4 núcleos, y 32 GB de capacidad RAM. Puede encontrarse el código para la estimación de los estratos en: <https://github.com/dtcensos/Estimacion-homicidios/blob/main/src/estratificacion.R>

## Combinación de estimaciones

El resultado del paso anterior son 1000 muestras para cada estrato que fue posible estimar. En este sentido, es necesario combinar las muestras de cada estrato (combinación de estimaciones) para obtener la media y un intervalo de confianza para cada estrato. Este paso del proceso es el último y sigue corriendo en R al 26 de abril de 2024. Sin embargo, el código puede encontrarse en: <https://github.com/dtcensos/Estimacion-homicidios/blob/main/src/combinacion.R>

# Resultados

Esta sección se divide dos: la evaluación de la distribución posterior de los estratos estimados, que permite asegurar que la heterogeneidad de captura se haya asilado satisfactoriamente y los resultados del subregistro para cada estrato.

## Distribución posterior de los estratos estimados

Como se menciona en la sección 3, la estimación de los estratos da como resultado un *dataset* con 1000 muestras para cada estrato. Usando esas 1000 muestras, es posible graficar la distribución posterior de cada estrato. Una distribución unimodal de la mayoría de estratos indica que la estratificación ha aislado la heterogeneidad de captura exitosamente. El gráfico 1 presenta un ejemplo de un estrato con distribución posterior unimodal. Por otro lado, una distribución multimodal de la mayoría de estratos indica que la heterogeneidad de captura podría estar sesgando los resultados, por lo que es necesario pensar en una estratificación diferente. El gráfico 2 presenta un ejemplo de un estrato con distribución posterior multimodal.

Considerando que revisar más 73.000 gráficas de distribución posterior no es factible, se diseñó una función en R que evalúe el número de picos en la distribución posterior de cada estrato. El resultado de esta función se muestra continuación. Los casos en los que el valor es TRUE, indican estratos con distribución unimodal. Por el contrario, los casos con valor de FALSE indican estratos con distribución multimodal. El código puede encontrarse en: <https://github.com/dtcensos/Estimacion-homicidios/blob/main/src/distribucion%20posterior.R>

Gráfico . Estrato con distribución posterior unimodal (R5-1990\_1994-90-94-HOMBRE-11)

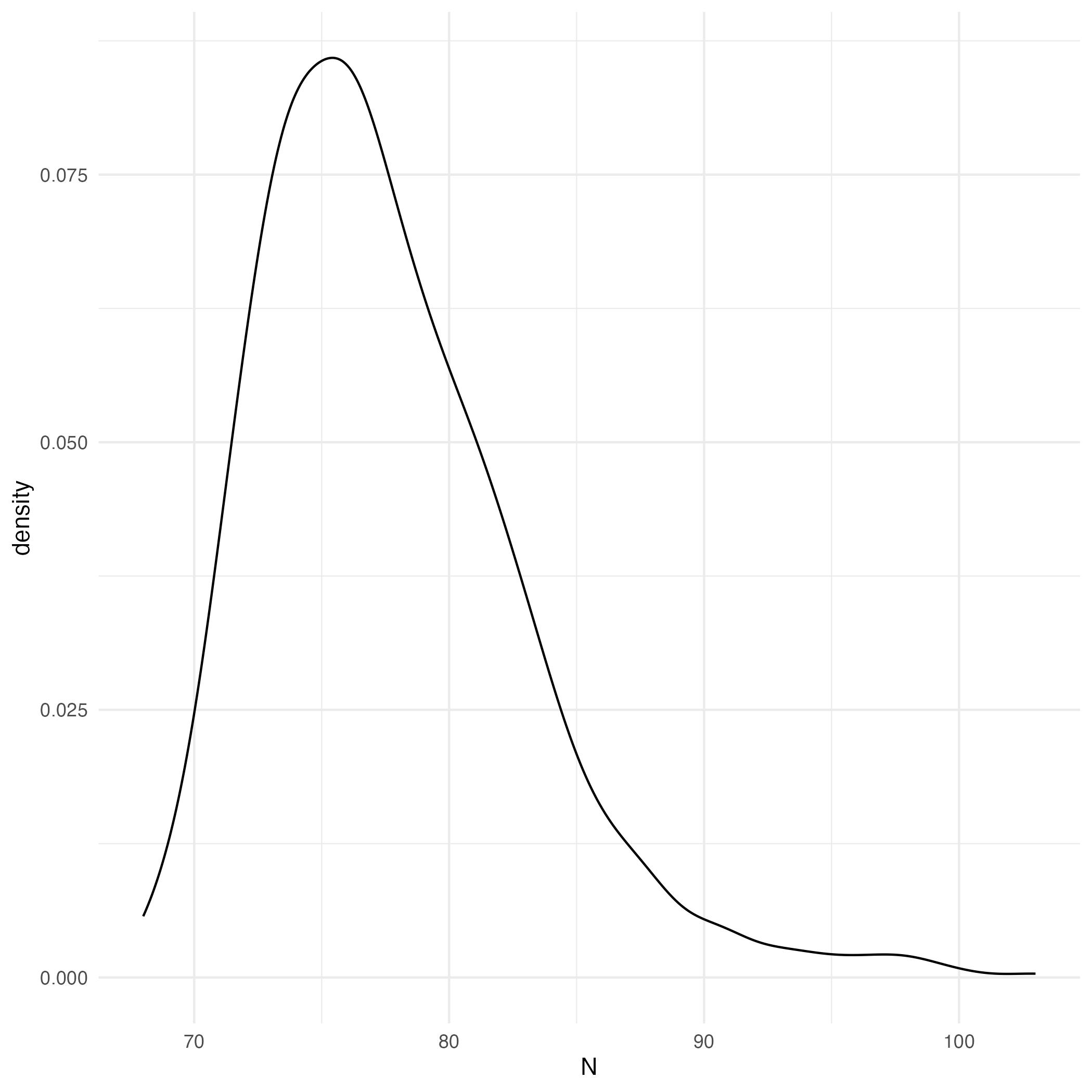


Gráfico : Estrato con distribución posterior multimodal (R1-1985\_1989-0-4-MUJER-11)

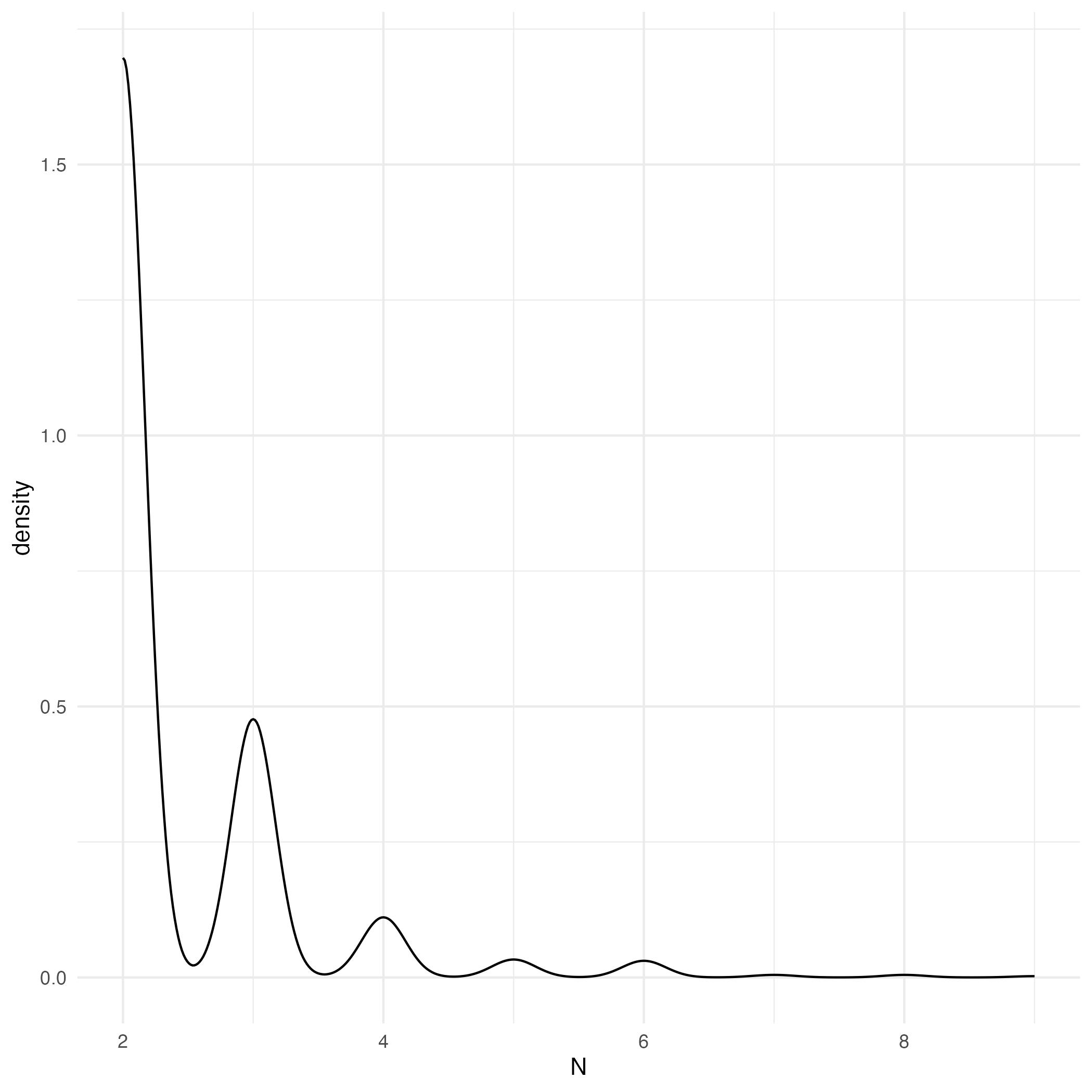


Tabla : Número de estratos con distribución unimodal y multimodal

|  |  |
| --- | --- |
| **Multimodal** | **Unimodal** |
| 16572 | 54789 |

Fuente: DANE. DCD. Datos: Proyecto JEP-CEV-HRDAG

# Conclusión

De acuerdo a lo presentado anteriormente, se considera que la estratificación definida (*Quinquenio – edad por grupos quinquenales – sexo – departamento del hecho)* es apropiada. Sin embargo, no fue posible estimar algunos estratos, dado que el requerimiento del algoritmo es que haya al menos una víctima documentada por tres fuentes en cada estrato estimable. Lo anterior, dado el supuesto de independencia de listas del método[[5]](#footnote-5).

En ese sentido, para los estratos que no fue posible estimar se adjuntan los datos observados. Es decir, los datos que se adjuntan corresponden únicamente a la documentación de víctimas en ese estrato entre diferentes bases de datos. De los 73.178 estratos, esto ocurrió para **1.817 estratos.**

# Bibliografía

* Manrique-Vallier, Daniel. 2016. “Bayesian Population Size Estimation Using Dirichlet Process Mixtures.” *Biometrics* 72 (4): 1246–54.
* Gargiulo, M. Durán, M. Amado, P. Ball, P. 2024 “verdata: an R package for analyzing data from the Truth Commission in Colombia,” Journal of Open Source Software
* HRDAG, 2024. “Informe metodológico del proyecto conjunto JEP-CEV-HRDAG de integración de datos y estimación estadística”

1. Ver fuentes de datos en el informe metodológico del proyecto, sección 9.3: <https://microdatos.dane.gov.co/index.php/catalog/795/related-materials> [↑](#footnote-ref-1)
2. Ver sección 4 relacionada a la imputación estadística en el informe metodológico del proyecto en: <https://microdatos.dane.gov.co/index.php/catalog/795/related-materials> [↑](#footnote-ref-2)
3. Ver sección 6 relacionada a la estimación por sistemas múltiples en el informe metodológico del proyecto en: <https://microdatos.dane.gov.co/index.php/catalog/795/related-materials> [↑](#footnote-ref-3)
4. Paquete disponible en: <https://github.com/HRDAG/verdata> [↑](#footnote-ref-4)
5. Ver sección 6 relacionada a la estimación por sistemas múltiples en el informe metodológico del proyecto en: <https://microdatos.dane.gov.co/index.php/catalog/795/related-materials> [↑](#footnote-ref-5)