
Un modelo de áreas pequeñas para la estimación de la frecuencia de consumo de Basuco al interior de las Unidades de Planeamiento Zonal de Bogotá

Small area estimation model to estimate consumption frequency of Basuco inside Zonal Planning Units in Bogotá

Nicolás Ramirez Vargas^a
nicolasramirezv@usantotomas.edu.co

Diego Fernando Lemus Polania^b
dflemusp@gmail.com

Henry Alejo Alejo^c
halejo@saludcapital.gov.co

Resumen

La atención en salud a las personas que presentan consumo abusivo de sustancias psicoactivas, implica un despliegue logístico y operativo significativo por parte de la Secretaria Distrital de Salud de Bogotá (SDS) para asegurar la Intervención Epidemiológica de Campo (IEC) de los consumidores.

Para atender esta situación, la SDS ha dispuesto del subsistema de *Vigilancia Epidemiológica para el consumo abusivo de Sustancias Psicoactivas* (VESPA). Este subsistema, brinda asistencia técnica a quienes presentan consumo abusivo de *Sustancias Psicoactivas* (SPA) y que han sido atendidos en primera instancia por las *Unidades Primarias Generadoras de Datos* (UPGD) o Unidad Informadora (UI). Estas instalaciones son centros especializados en atención de farmaco dependencia.

El presente estudio busca garantizar estimaciones temporales del total de consumidores en cada *Unidad de Planeamiento Zonal* (UPZ) a través de un modelo de áreas pequeñas (Fay Herriot espacial) que permita a la SDS focalizar la atención y vigilancia de este fenómeno de acuerdo a la concentración geográfica de los casos estimados, teniendo en cuenta que las UPGD no brindan cobertura total en el Distrito y tampoco son captados la totalidad de casos que deben ser atendidos.

Para obtener las estimaciones definitivas, en los dominios observados fueron comparados los estimadores directos y los estimadores de áreas pequeñas, siendo los segundos quienes presentaron mejores resultados en términos del coeficiente de variación estimado.

—

Palabras clave: Estimación en áreas Pequeñas, Modelo a nivel de unidad, Fay Herriot espacial, Consumo abusivo de basuco, Intervención epidemiológica..

^aEstudiante

^bDirector

^cDirector

Resumen

Health care for population with abusive consumption of psychoactive substances, involves a significant logistical and operational display by the Health Secretary of Bogota (SDS) to ensure epidemiological intervention to drug users.

To attend this situation, SDS has disposed the VESPA (Epidemiological Surveillance of the abuse of Psychoactive Substances). This subsystem bring technical assistance for drug users that present abusive consumption of PSA (Psychoactive substances) and that they have been attended in the first instance by a UPGD (Primary Data Generating Unit) or UI (Reporting Unit). These facilities are centres specialising in drug dependency care.

Herein, we present estimations of the total of drug user for each planning unit through a model of small areas (spatial Fay Herriot) that allow to the SDS focus its attention and surveillance to this phenomenon according of the geographic concentration of estimate cases, since normality, the UPGD dont have coverage of the entire municipality and neither report the totality of the cases attended.

To obtain definitive estimation, in observed domains were compared direct estimators and composed estimators, being the composed estimators more precise in therms of estimate coefficient of variation.

—

Palabras clave: Small Area Estimation, Area model, Spatial Fay Herriot, Basuco abusive consumption, Epidemiological intervention..

1. Introducción

El consumo de sustancias psicoactivas es uno de los problemas más graves en la sociedad actual, y puntualmente la causa de problemáticas de salud pública por las consecuencias que ocasiona sobre el consumidor, evidenciando aumentos en la demanda de servicios, el incremento en índices de incapacidad social y laboral, fenómenos delictivos, propagación de epidemias como la del Virus de Inmunodeficiencia Humana VIH y hepatitis B y a la degradación de estructuras sociales fundamentales (Barreto & Perez 2015)

La SDS realiza un esfuerzo considerable con el fin de garantizar el acceso a los servicios especializados en el tratamiento de la farmacodependencia de los consumidores problemáticos de sustancias psicoactivas, así como a captar de manera temprana los casos de consumo inicial de sustancias psicoactivas en niños, niñas, adolescentes y jóvenes entre los 10 y 26 años de edad. En el caso de los consumos problemáticos, la red del Distrito ha implementado los *servicios de acogida* diferenciando de manera clara el *consumo de drogas* con características de dependencia o de adicción a sustancias psicoactivas.

Resulta importante señalar que de acuerdo a la Organización Mundial de Salud (2005), *una droga psicoactiva es aquella sustancia que cuando se ingiere afecta los procesos mentales (cognición, afectividad, entre otras) al generar efectos directos sobre el sistema nervioso central*¹. Aunque el término psicoactivo no necesariamente se refiere a que el consumo implique dependencia, esta puede ser una de sus posibles consecuencias, pues resulta común que el uso se realice de forma continua a pesar de que el sujeto tenga plena conciencia de que esto implica tener un problema social, laboral, psicológico o físico, persistente o recurrente, provocado o estimulado por el consumo habitual o recurrente en situaciones en las que es físicamente peligroso.

Según Sherman (2017), el consumo habitual o recurrente denominado **consumo abusivo** se produce cuando una persona no puede controlar su consumo de drogas a un punto tal que su capacidad de actuar normalmente se ve afectada.

Mediante el Subsistema de Vigilancia Epidemiológica del abuso de Sustancias Psicoactivas VESPA, la SDS tiene la posibilidad de rastrear la totalidad de casos atendidos a causa del consumo abusivo en

¹<https://www.infodrogas.org/drogas/que-son-las-drogas?showall=1>

cada UPGD o UI, las cuales son entidades públicas o privadas que captan la ocurrencia de eventos de interés en salud pública y genera información útil y necesaria para los fines de la SDS, conformando así, el sistema de atención a consumidores de sustancias psicoactivas.

Teniendo en cuenta que esta información obedece al consumo de sustancias psicoactivas en el Distrito Capital, hay un amplio reporte de las diferentes sustancias consumidas. De acuerdo a la definición de Aguado (1994), estas suelen ser clasificadas en dos tipos de sustancias:

- **Legales:** en Colombia, es considerado legal el consumo de tabaco y alcohol dado que la producción, porte, transporte y comercialización es permitida.
- **Ilegales:** se encuentra prohibido el uso o comercialización de sustancias cuya producción o porte no se encuentra legalmente constituido, así, sustancias como la marihuana, bazuco, cocaína, etc., no son consumidas de forma legal.

Es importante mencionar que, aunque existe una categoría de sustancias *legales*, el uso de estas también es nocivo para la salud. Con esta definición y de acuerdo a la información almacenada en VESPA, en el Distrito al día son atendidos 22 casos por consumo abusivo de cualquier tipo de sustancias psicoactivas.

La estructura de atención del Sistema de Salud en el Distrito se encuentra dividida en cuatro regiones, denominadas *subredes de atención*. A pesar de que cada una de las subredes obedece a una misma directriz entregada por la Ministerio de Salud, el manejo administrativo de los recursos, personal, programas y estrategias es autónoma en cada una de ellas, generando así estructuras independientes entre subredes.

Las subredes se encuentran divididas por localidades tal como se observa en la *Figura 1*, donde se presentan la estructura de cada subred.

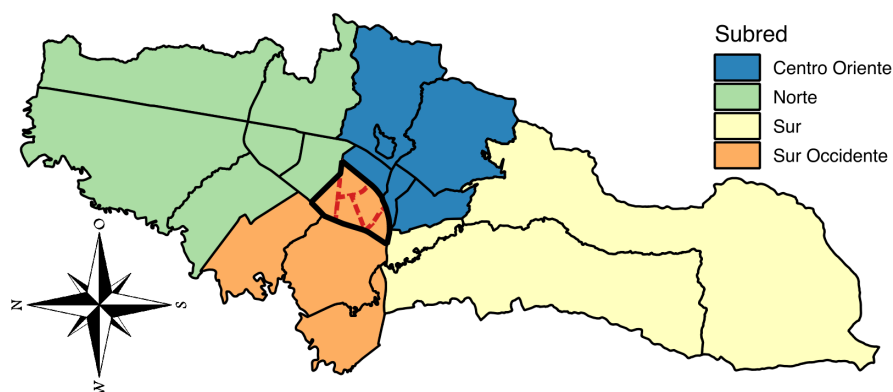


Figura 1: Estructura geográfica

A continuación, se presentan las localidades que conforman cada una de las subredes:

- **Centro Oriente:** Antonio Nariño, Candelaria, Mártires, Rafael Uribe, San Cristóbal, Santa Fe.
- **Norte:** Barrios Unidos, Chapinero, Engativá, Suba, Teusaquillo, Usaquén.
- **Sur:** Ciudad Bolívar, Tunjuelito, Usme.
- **Sur Occidente:** Bosa, Fontibón, Kennedy, Puente Aranda.

Tal como lo expresa la Alcaldía Mayor de Bogotá (2014), las localidades son definidas como unidades administrativas con homogeneidad geográfica, cultural, social y económica, que tienen criterios propios

para la formulación y aprobación de planes de desarrollo local dentro del marco de la participación ciudadana y de la ejecución de los proyectos en su interior. Paralelamente, al interior de las localidades se genera otra partición geográfica, designado como UPZ.

De acuerdo con la Secretaria Distrital de Planeación (2016), las UPZ son áreas urbanas que conforman las localidades, visualmente equivalentes a las particiones rojas. (Ver *Figura 1*). La definición y consolidación de estas áreas, obedece a la necesidad de planificar el desarrollo urbano en el nivel zonal más detallado posible, lo que las lleva a actuar como unidades territoriales intermediarias entre los barrios y las localidades con el fin de desarrollar normas urbanísticas acordes a las necesidades puntuales existentes en el territorio teniendo en cuenta que se requiere intervención diferencial dados los contrastes existentes entre un territorio y otro.

Aunque la definición de estas unidades cartográficas esta orientada a consolidar estructuras y planteamientos similares, existe un comportamiento disímil entre UPZ, lo que directamente implica la existencia de diferencias no solo administrativas sino en el patrón del fenómeno de interés, pues a pesar de la vecindad y cercanía entre ellas, en muchas ocasiones hay diferencias significativas al revisar una unidad contigua a otra.

La información albergada en VESPA, brinda la posibilidad de determinar que si bien hay atención al consumo de sustancias psicoactivas en la mayoría de UPZ que conforman el Distrito, existe un alto grado de desigualdad a la hora de contrastar la frecuencia de casos captados a través del VESPA en cada una de estas. Esto mismo ocurre al realizar el contraste de otro tipo de variables como la densidad poblacional, el número de centros de atención, indicadores de pobreza, presencia institucional, entre otras.

Así, teniendo conocimiento de que la planeación y la estructura general de cada UPZ mantiene comportamientos diferenciales, surge la intención de realizar mediciones detalladas de los aspectos característicos que son candidatos a ser causales no sólo del consumo abusivo de sustancias psicoactivas, sino también de los patrones que puedan ocasionar el consumo de estas.

El VESPA fue propuesto y posteriormente constituido como una herramienta útil para contribuir a la focalización y orientación de las acciones que desde la salud pública se pueden desarrollar con el propósito de reducir el impacto del consumo de sustancias psicoactivas en el Distrito, detectar el ingreso de nuevas sustancias psicoactivas, dar cuenta de las modificaciones en los patrones de consumo abusivo, y así realizar acciones orientadas hacia la reducción del daño, la prevención universal específica e indicada.

Aunque se considera que las intervenciones realizadas sobre el abuso de sustancias psicoactivas a través del VESPA, priorizadas actualmente en menores de 16 años y gestantes, obedece a la realidad descrita por medio de dicho sistema de vigilancia, es relevante mencionar que en algunas ocasiones se reportan casos de consumo no abusivo, los cuales no son ingresados en el aplicativo dispuesto para su operación. Ante estas situaciones, los casos son remitidos a otras estrategias de intervención desarrolladas en el Plan de Salud Pública de Intervenciones Colectivas (PSPIC).

Ahora bien, aún cuando hay infinidad de sustancias psicoactivas, en su totalidad de carácter ilegal, hay sustancias que generan impactos mucho más negativos que otros. Este es el caso del *basuco* también conocido como pasta de cocaína que, de acuerdo con la percepción de los consumidores atendidos por VESPA afirman que es una de las sustancias con mayores impactos para la salud. Por esta razón y dada la alta ocurrencia de consumo de esta sustancia por parte de quienes han sido intervenidos y el interés particular que tiene la SDS en reducir el consumo abusivo de esta sustancia, el análisis realizado en el presente estudio se orienta a analizar este caso en particular.

Bajo este escenario, conocer el número real de casos atendidos en cada nivel geográfico posible, teniendo en cuenta las restricciones con las que cuenta el sistema de atención del VESPA, resulta ser de total interés para el planteamiento de políticas públicas orientadas a la población objeto de intervención, donde no solo se tengan en cuenta los patrones existentes en la información analizada, sino que se pueda incorporar información auxiliar proveniente de diferentes fuentes de información oficial, que permitan determinar los factores altamente asociados al consumo de sustancias psicoactivas

Dicho de esta manera y dada la posibilidad de emplear la técnica de estimación en áreas pequeñas en los

niveles geográficos reducidos de interés por las características de la técnica a emplear, UPZ, planteando un modelo que involucre características de diversas fuentes de información, se podrá conocer el número estimado de consumidores de basuco y con esto, se podrá llevar a cabo la convalidación del número de casos atendidos en el VESPA, se podrá obtener una visión más amplia del fenómeno y surgirán resultados de interés para la SDS en relación con las estrategias de intervención.

2. Metodología

Para estimar el número de consumidores de basuco en Bogotá, fue necesario revisar la literatura y exponer las características de los diferentes estimadores a emplear en el presente estudio, teniendo en cuenta que la intención de la investigación, es poder realizar un ejercicio comparativo que permita evidenciar las ventajas que tiene la implementación de los estimadores para áreas pequeñas particularmente.

Esta revisión, se ve acotada dada la naturaleza de los modelos de estimación en áreas pequeñas, pues con esta metodología, es posible realizar la comparación de los resultados a través de medidas que indican la precisión de los resultados. En primera instancia se revisó la línea de estimación basada en el diseño, mediante los estimadores directos, y posteriormente se revisó la línea de estimación basada en el modelo a partir de los estimadores sintéticos, los estimadores compuestos y los estimadores para áreas pequeñas.

La comparación fue acompañada por la validación de los resultados obtenidos con la muestra seleccionada empleando un proceso iterativo implementado bajo el método bootstrap.

2.1. Estimadores directos

Es frecuente encontrar que en los escenarios donde se llevan a cabo diseños de muestreo, las estimaciones se obtengan bajo la línea de estimación basada en el diseño y las fuentes de información requeridas para ello no sean más que los pesos de muestreo propios del diseño propuesto. El estimador insesgado propuesto por Horvitz & Thompson (1952), es en la mayoría de ocasiones el empleado para alcanzar las estimaciones de los parámetros de interés. Con base en ello, el estimador tipo total se define como:

$$\hat{y}_\pi = \sum_S \frac{y_k}{\pi_k} \quad (1)$$

Donde π_k denota la probabilidad de inclusión del k-ésimo elemento de la muestra de tamaño \mathbf{S} y y_k hace referencia a la característica de interés del k-ésimo elemento. Teniendo en cuenta que para el presente estudio es de interés realizar la estimación a nivel de dominios, el estimador de Horvitz-Thompson para un dominio se define como:

$$\hat{y}_{d,\pi} = \frac{N_d}{n_d} \sum_{S_d} Y_k \quad (2)$$

Donde N_d , de modo general denota el número de elementos que conforman el universo del dominio y n_d el número de elementos de la muestra asociados al d-ésimo dominio. Este cociente, toma el papel de factor de expansión en casos donde el diseño de muestreo propuesto es aleatorio simple. En aplicaciones donde el diseño de muestreo sea diferente, esta expresión varía como lo presenta Gutierrez (2009).

Ahora bien, en los casos donde el parámetro de interés sea un promedio, este se define como el cociente entre la estimación del total de la característica de interés y la estimación del total poblacional, ambas medidas en el dominio d .

En relación con la estimación de la varianza de este estimador a nivel de dominios, de acuerdo con Gutierrez (2009), la varianza se define como:

$$\widehat{Var}(\hat{y}_{d,\pi}) = \frac{N_d^2}{n_d} \left(1 - \frac{n}{N}\right) S_{y_d S}^2 \quad (3)$$

Donde al igual que en la *ecuación 2*, N denota el número de elementos que conforman el universo del dominio, n el número de elementos del dominio presentes en la muestra y $S_{y_d S}^2$ es equivalente a la varianza muestral de los valores de la característica de interés para los elementos pertenecientes al dominio.

Luego, con la intención de tener una medida que permita determinar el nivel de precisión de la estimación obtenida, es comúnmente empleado el *Coficiente de Variación Estimado* (CVE), el cual está dado por:

$$cve(\hat{y}_{d,\pi}) = \frac{\sqrt{\widehat{Var}(\hat{y}_{d,\pi})}}{\hat{y}_{d,\pi}} * 100\% \quad (4)$$

Esta medida, a lo largo del presente documento será la que permita realizar comparaciones entre los estimadores empleados, pues al ser reportada en términos porcentuales, brinda la posibilidad de hacer comparables los resultados obtenidos en las diferentes estimaciones presentadas y así, a través de la precisión obtenida por cada uno de ellos, determinar cuál es el más eficiente como lo menciona Gutierrez (2010).

2.2. Estimadores compuestos

Otro tipo de estimadores, frecuentemente empleados en la estimación en áreas pequeñas, son los estimadores compuestos, los cuales se obtienen como el producto ponderado de la combinación del estimador directo y el estimador sintético.

$$\bar{y}_d = \alpha \hat{y}_{d,\pi} + (1 - \alpha) \hat{y}_{sin} \quad \text{con } 0 \leq \alpha \leq 1 \quad (5)$$

donde α toma el rol de ponderador, $\hat{y}_{y\pi}$ denota el estimador directo y \hat{y}_{sin} el estimador sintético.

Teniendo en cuenta que son los estimadores compuestos los empleados en los modelos de áreas pequeñas, definidos más adelante, de acuerdo con Paños (2000), el estimador sintético se define como:

$$\hat{y}_{sin} = \sum_g \frac{\hat{Y}_g}{\hat{X}_g} X_{g,d} \quad (6)$$

donde $X_{g,d}$ es una variable auxiliar sobre la cual se conoce el comportamiento en los d dominios por información proveniente de registros administrativos, mientras que \hat{Y}_g y \hat{X}_g son las estimaciones de la característica de interés y la característica auxiliar de interés respectivamente.

Carl-Erik Särndal (2003) afirma que una de las ventajas de este tipo de estimadores es poder incluir información auxiliar al diseño de muestreo, incluido a través del estimador directo, permitiendo así obtener estimaciones del parámetro de interés con un nivel superior de precisión.

Sin embargo, esto ocurre sólo cuando la información auxiliar incluida es de buena calidad. En el peor de los casos, donde la información auxiliar no tenga un aporte favorable, el estimador compuesto va a replicar los resultados obtenidos bajo el estimador directo.

2.2.1. Modelo Fay-Herriot

El primer modelo mixto que incluyó efectos aleatorios a nivel de área, es decir, un modelo en el cual la información se observa de manera agregada por dominio, es el propuesto por Fay & Herriot (1979). Este modelo es uno de los más usados cuando de estimación en áreas pequeñas se trata.

$$\bar{y}_d = \mu_d + e_d \quad (7)$$

con $\mu_d = \mathbf{X}_d\beta + u_d$ donde \mathbf{X}_d es la matriz de diseño conformada por el conjunto de variables explicativas, β es el vector de los parámetros regresores del modelo, $u_d = (u_1, \dots, u_d)$ el vector de efectos aleatorios y $\mathbf{e}_d = (e_1, \dots, e_d)$ el vector de los errores de muestreo; los dos últimos son vectores independientes con distribución $N(0, \sigma_u^2)$ (con σ_u^2 desconocida) y $N(0, \sigma_d^2)$ respectivamente. Del anterior, σ_d^2 es equivalente a la varianza estimada a partir del diseño de muestreo en los d dominios observados.

Dicho de esta manera, el estimador del promedio se define como la combinación del estimador directo y el estimador sintético, es decir:

$$\bar{y}_d = \hat{t}_{y_d, \pi} \hat{\gamma}_d + (1 - \hat{\gamma}_d) \mathbf{X}_d \hat{\beta} \quad \text{con} \quad \hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_d^2} \quad (8)$$

donde $\hat{\sigma}_u^2$ corresponde a la estimación de la varianza poblacional, mientras que $\hat{\sigma}_d^2$ es la estimación de la varianza en el dominio d ; ambas medidas sobre la característica de interés. Con estas definiciones, $\hat{\gamma}_d$ pondera de acuerdo a la magnitud de la varianza del dominio respecto a la varianza poblacional de la característica de interés.

De la ecuación 8, según Molina (2009) surge la definición de estimador compuesto, pues la combinación de dos estimadores dan origen a la estimación del promedio en cada dominio. Por otra parte, el estimador del total para los dominios no es mas que:

$$\hat{y}_d = N_d \bar{y}_d \quad (9)$$

Hasta el momento, se ha presentado la estructura teórica necesaria para obtener la estimación del parámetro de interés, no obstante, es necesario obtener la estimación de la varianza para dicho parámetro. El modelo de Fay-Herriot presenta cierto nivel de complejidad para obtener esta última, pues las formulas desarrolladas son aproximaciones basadas en hipótesis que dependen del modelo propuesto, con lo cual, en muchas ocasiones no resultan ser comparables los modelos. Por esta razón N.G.N. Prasad (1990) propusieron el Error Cuadrático Medio (ECM) como una medida alternativa para obtener la medición de precisión del estimador, no obstante, en la práctica comúnmente esta estimación se obtiene por métodos de remuestreo.

Aunque hay varias alternativas para implementar los métodos de remuestreo, en este estudio se empleó el método de Bootstrap paramétrico mencionado por Yolanda Marhuenda (2013), el cual se desarrolla de la siguiente manera:

1. Ajustar el modelo con la información disponible $\{X_d, y_d\}$ y obtener la estimación de los parámetros del modelo $\hat{\beta}, \hat{\sigma}_d^2, \hat{\rho}$.
2. Generar los efectos bootstrap de área u_d provenientes del proceso SAR(1) usando $(\hat{\sigma}^2, \hat{\rho})$ como los verdaderos parámetros de (σ^2, ρ)
3. Calcular las cantidades verdaderas del bootstrap

$$\mu_d = \mathbf{X}_d \hat{\beta} + u_d$$

4. Generar los errores para $\mathbf{e}_d \sim N(0, \sigma_d^2)$ y obtener la información para el modelo de la muestra.

$$y_d = \mu_d + \mathbf{e}_d$$

5. Ajustar el modelo para la nueva información de la muestra $\{X_d, y_d^{(b)}\}$ y obtener los estimadores.

$$\hat{\mu}_d^{(b)} = \mathbf{X}_d \hat{\beta}^{(b)} + u_d^{(b)}$$

6. Repetir los pasos 1 a 6 desde $b = 1, \dots, B$
7. Obtener la estimación del ECM definido como:

$$ECM(\hat{y}_d) = \frac{1}{B} \sum_{b=1}^B (\hat{y}_d^{(b)} - y_d^{(b)})^2$$

Una vez se cuenta con la estimación del total y el error cuadrático medio, ya es posible realizar el cálculo del coeficiente de variación estimado, tal como se presenta en la *ecuación 4*, pues en este tipo de modelos, el ECM es equivalente a la estimación de la varianza.

2.2.2. Modelo Fay-Herriot Espacial

Una modificación al modelo de Fay-Herriot se generó cuando Salvati (2008) le incluyó el efecto aleatorio producido por la autocorrelación espacial y la matriz de proximidad \mathbf{W} generada por la conformación espacial de los dominios. La estructura general del modelo se define como:

$$\bar{y}_d = \mathbf{X}_d \beta + (\mathbf{I}_d - \rho \mathbf{W})^{-1} \mathbf{u}_d + \varepsilon \quad (10)$$

con \mathbf{I} una matriz indicadora de dimensión $d \times d$, \mathbf{W} como la matriz de proximidad entre las áreas analizadas y ρ como el parámetro desconocido del proceso autoregresivo espacial SAR que mide la fuerza de las relaciones espaciales entre los efectos aleatorios en las áreas vecinas, el cual se obtiene empleando el estimador restringido de máxima verosimilitud (por sus siglas en inglés REML) donde se maximiza la verosimilitud de ω dado por el vector y .

La maximización de la verosimilitud es realizada usando un proceso iterativo, conocido como Fisher-Scoring, que es definido por Schworer (2004) como el proceso empleado para calcular los valores esperados de la segunda derivada del logaritmo natural de la verosimilitud, con el cual, se obtiene la matriz de información de Fisher, definida como:

$$I(\omega) = \begin{pmatrix} I_{\sigma_u^2, \sigma_u^2} & I_{\sigma_u^2, \rho} \\ I_{\rho, \sigma_u^2} & I_{\rho, \rho} \end{pmatrix} \quad (11)$$

del cual,

$$I_{\rho, \rho} = \frac{1}{2} tr\{P(\omega)A(\omega)P(\omega)A(\omega)\} \quad (12)$$

De esta manera, el estimador está dado por:

$$\bar{y}_d = \mathbf{X}_d^T \hat{\beta} + b_d^T \{\hat{\sigma}_u^2 [(\mathbf{I}_d - \hat{\rho} \mathbf{W})(\mathbf{I}_d - \hat{\rho} \mathbf{W}^T)]^{-1}\} * \{diag(\Psi_d) + \hat{\sigma}_u^2 [(\mathbf{I}_d - \hat{\rho} \mathbf{W})(\mathbf{I}_d - \hat{\rho} \mathbf{W}^T)]^{-1}\}^{-1} (\hat{t}_{y, \pi} - \mathbf{X}_d^T \hat{\beta}) \quad (13)$$

donde b_d^T corresponde a un vector de dimensión $1 \times d$ de posiciones.

Con base en la ecuación 13 y con el objetivo de estimar $t_{y, \pi, d}$, surge el Mejor Predictor Lineal Insesgado o mejor conocido en la literatura por sus siglas en inglés **BLUP**, el cual se define como:

$$\tilde{\theta}_d(\omega) = \mathbf{X}_d \tilde{\beta}(\omega) + b_d' G(\omega) \rho^{-1}(\omega) [y - \mathbf{X}_d \tilde{\beta}(\omega)] \quad (14)$$

donde b_d' es un vector de dimensión $1 \times d$ y $G(\omega)$ la matriz teórica de varianzas y covarianzas.

A pesar de que el BLUP es la mejor estrategia para obtener la estimación del número de consumidores de basuco, este estimador depende de un vector teórico desconocido de varianzas, por lo que resulta necesario emplear el estimador propuesto por Bharat Singh (2005), quienes emplean la estimación del vector de varianzas. A esto lo denominan el Mejor Predictor Lineal Insesgado Empírico (**EBLUP**) y se define como:

$$\tilde{\theta}_d(\hat{\omega}) = \mathbf{X}_d \tilde{\beta}(\hat{\omega}) + b'_d G(\hat{\omega}) \rho^{-1}(\hat{\omega}) [y - \mathbf{X}_d \tilde{\beta}(\hat{\omega})] \quad (15)$$

donde $\hat{\omega}$ es equivalente a $(\hat{\sigma}_u^2, \hat{\rho})'$, lo cual representa la estimación de la varianza obtenida a partir del estimador directo.

2.3. Método bootstrap

Técnica de remuestreo propuesta por Efron (1979), con el cual se intenta estimar distribuciones o parámetros basados en observaciones independientes. El procedimiento para la implementación del método consta de los siguientes pasos:

1. Partiendo del marco muestral e imitando el diseño muestral generar B muestras bootstrap.
2. Para cada muestra bootstrap, calcular el estimador $\hat{\theta}$, equivalente a \bar{y}_d . Estos se denotan: $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ donde B es el número de muestras bootstrap realizadas, $\hat{\theta}$ es el parámetro de interés estimado sin aplicar el método bootstrap y $\hat{\theta}_b^*$ es el parámetro estimado en la b -ésima muestra seleccionada.

Normalmente el estimador $\hat{\theta}$ se define como el promedio de las b -ésimas realizaciones, no obstante, como se sabe que el comportamiento del fenómeno de interés no es el mismo geográficamente y la posibilidad de obtener estimaciones atípicas es alta, se optó por emplear la mediana (percentil 50) como alternativa de estimación del parámetro θ .

3. Diseño metodológico y procedimientos

Dada la posibilidad de contar con el marco muestral y con el fin de realizar un proceso comparativo, donde se contrastó el resultado obtenido a partir de los estimadores directos y los estimadores de áreas pequeñas, se consideró un diseño de muestreo con las siguientes características:

- **Probabilístico:** todas las unidades de muestreo en cada etapa tienen una probabilidad conocida y mayor de cero de ser incluidos en la muestra.
- **Estratificado:** se consideran particiones poblacionales definidas como las subredes de atención.
- **Multietápico:** la selección de las unidades de observación requirió la selección de unidades seleccionadas en diferentes niveles del diseño.

Este diseño de muestreo fue implementado en el software estadístico **R**, donde se realizó el cálculo del tamaño de muestra, el cual es representativo a nivel de subred, y la selección de la misma, lo que se realizó en tres etapas:

- **Primera etapa:** selección de localidades con probabilidad proporcional al tamaño de una variable auxiliar (π PT), usando el número de consumidores registrados en VESPA por localidad.
- **Segunda etapa:** al interior de las localidades seleccionadas, se seleccionaron UPZ con Muestreo Aleatorio Simple (MAS).

- **Tercera etapa:** finalmente, fueron seleccionados consumidores (unidades de observación) con MAS en las UPZ muestreadas.

El procedimiento empleado para llevar a cabo la selección de las respectivas unidades de muestreo en cada una de las etapas del diseño, fue implementado usando la librería TeachingSampling (2015) del software **R** como se mencionó previamente. La selección de las localidades, donde fue necesario contar con información auxiliar para cada una de ellas, se realizó empleando el *método de Sunter* propuesto por Sunter (1986), el cual consiste en:

1. Ordenar descendentemente la población de acuerdo con los valores que toma la característica de información auxiliar x_k .
2. Realizar $\xi_k \sim U(0, 1)$.
3. Para $k = 1$, el primer elemento de la lista ordenada es incluido en la muestra sí y solamente sí $\xi_1 < \pi_1$.
4. Para $k \geq 2$, el k -ésimo elemento de la lista ordenada es incluido en la muestra sí y solamente sí

$$\xi_k \leq \frac{n - n_{k-1}}{n - \sum_{i=1}^{k-1} \pi_i} \pi_k$$

donde n_{k-1} representa el número de elementos que ya han sido seleccionados al final del paso $k - 1$.

Mientras que la selección tanto de las UPZ como de los consumidores se realizó empleando el *método de selección y rechazo* propuesto por Fan et al. (1962), el cual consta de los siguientes pasos:

1. Generar N realizaciones de una variable aleatoria ξ_k ($k \in U$) con distribución uniforme (0,1).
2. Asignar ξ_k al elemento k -ésimo de la población.
3. Ordenar la lista de elementos descendente (o ascendentemente) con respecto a este número aleatorio ξ_k .
4. A continuación, seleccionar los n primeros (o los n últimos) elementos. Esta selección corresponde a la muestra realizada.

Empleando las características del diseño de muestreo propuesto y los algoritmos mencionados, en primera instancia se llevo a cabo la selección de una muestra que sería el punto de referencia para la definición del modelo propuesto, y posterior a ello, se implementó el procedimiento bootstrap en el que se realizó 1.000 replicas; esta cantidad, asumiendo que es un número de iteraciones suficientes para aplicar la técnica de bootstrap, teniendo en cuenta que Tibshirani (1993) afirman que en la mayoría de ocasiones 500 iteraciones son una cantidad suficiente para obtener resultados consistentes.

La razón por la que se definió el modelo en la primera muestra es por que la muestra obtenida en cada una de las replicas presentaría un comportamiento totalmente diferente a la hora de seleccionar el modelo óptimo y de esta manera no podrían ser comparables los resultados obtenidos.

Al obtener los resultados del procedimiento iterativo, se procedió a analizar la información, así empleando el percentil 50 (mediana) como la medida robusta que permitió conocer la estimación del número de consumidores de basuco en cada UPZ y el coeficiente de variación estimado fue posible contrastar los resultados de un caso particular y un proceso iterativo.

Este contraste, fue realizado con el fin de validar que los resultados obtenidos mediante la muestra seleccionada no sean dependientes de la misma, y así, en caso que la selección de la muestra hubiese incorporado unidas de muestreo distintas, las estimaciones obtenidas difieran considerablemente, y con ello, las conclusiones del presente estudio variarían.

4. Fuentes de información

Para llevar a cabo la estimación del número de consumidores de basuco, se emplearon diversas fuentes de información. Estas, se agrupan en dos tipos:

4.1. VESPA

La Secretaría Distrital de Salud y en particular el subsistema de Vigilancia Epidemiológica del abuso de Sustancias Psicoactivas permitieron consultar el universo de consumidores de sustancias atendidos entre enero de 2017 y diciembre de 2017. Esta información toma el rol de marco muestral para la implementación del diseño de muestreo propuesto, en el escenario ideal se esperaría que fuese el sistema de referencia para encontrar características sociodemográficas, culturales y económicas de la población atendida.

Al momento de brindar la IEC, es el formato de registro del VESPA el instrumento de recolección de información básica, en el cual no se indaga por las características mencionadas anteriormente. Dicho de esta manera, el VESPA no cuenta con insumos suficientes para poder dar explicación directamente al fenómeno de interés, por lo que es totalmente necesario recurrir a características proveniente de otras fuentes de información.

4.2. Información secundaria y registros administrativos

Teniendo en cuenta que, para poder realizar el presente estudio, es insumo indispensable contar con información característica para la totalidad de las UPZ que conforman el Distrito, resulta ser un proceso largo la recolección de información auxiliar para el modelamiento.

En un escenario inicial, se revisó la disponibilidad, pertinencia y calidad de los registros administrativos existentes, con lo cual se obtuvo un resultado poco favorable, pues los registros administrativos no se encuentran en un repositorio de información único, sino que las cifras se consideran propiedad de las entidades responsables y no se encuentren disponibles para consulta al público.

Dicho de esta manera, no fue posible realizar la exploración de las variables y registros que se asumía existen, sino que por el contrario fue necesario definir un inventario de las características que a priori se podrían considerar factores asociados al consumo de basuco. En este proceso, además de definir las variables de interés también fue necesario revisar las entidades en las que podría encontrarse la información y a su vez, revisar las condiciones para que fuesen compartidos los registros, esto por la confidencialidad de la información en muchos casos.

En total, se recolectó información de 8 entidades estatales², logrando construir un repositorio de información agregado por UPZ de 147 variables. Este conjunto de información da cobertura a temas relacionados con:

- Personas por sexo, grupos etareos y niveles de escolaridad.
- Hogares por estrato socioeconómico.
- Viviendas por estrato y condiciones de habitabilidad de las mismas.
- Características sociodemográficas, económicas y políticas.
- Elementos recreativos, turísticos, sociales y culturales.
- Atención en salud.

²DANE, DNP, ICFES, POLICIA, IDRD, BIBLIORED, SDS e IDECA

Tanto el marco muestral como las fuentes secundarias de información fueron sometidas a un proceso de revisión, validación e imputación en los casos donde exista presencia de valor ausentes, con lo cual, se pudo garantizar que la información empleada para ajustar el modelo fuese consistente y certera.

El proceso de imputación empleado, luego de determinar que el esquema de ausencia de información tuviera baja tasa a imputar y fuese totalmente aleatorio tal como lo mencionó Lundström (2005), se buscó asignar información plausible a las variables no disponibles empleando información de aquellas UPZ similares, en relación con las características sociodemográficas, MICE.

5. Resultados

Empleando la metodología, los procedimientos y las fuentes de información mencionadas, se obtuvo los resultados requeridos para dar cumplimiento a los objetivos planteados. En primera instancia, será presentado el producto de la selección de una muestra particular y las estimaciones del número de consumidores de basuco por UPZ obtenidas con el estimador de áreas pequeñas, y posterior a ello, el resultado de la implementación del proceso iterativo que coteje los resultados obtenidos en la primera muestra.

Al examinar las características de la muestra inicial, fueron muestreadas el 63% de las localidades ($n = 12$) y el 27% del total de las UPZ ($n = 33$), tal como se observa en la *Figura 2*.

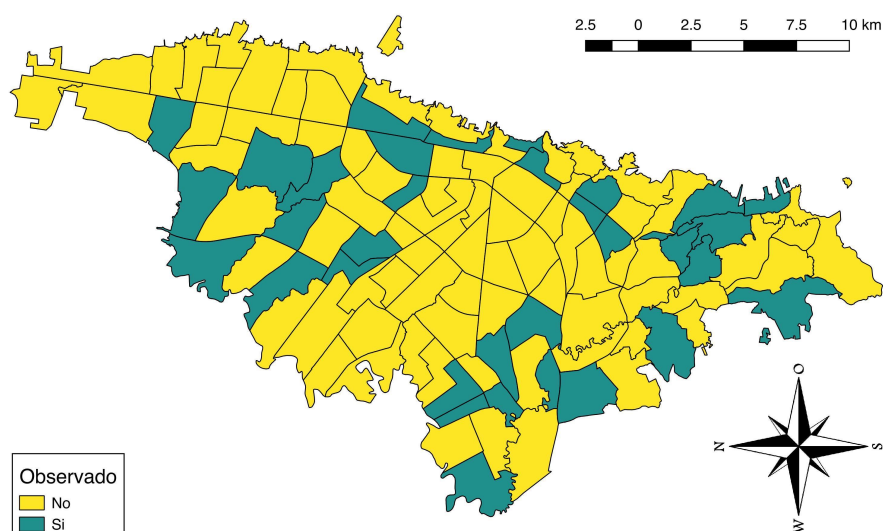


Figura 2: Dominios observados

Aún cuando en la *Figura 2* se reportan las UPZ observadas (seleccionadas) en la muestra, es fundamental recalcar que la unidad de observación del diseño de muestreo propuesto son los consumidores. Luego, como los dominios de interés del estudio son las UPZ, el objetivo de la *Figura 2* radica en presentar cuales son los dominios observados, entendiendo que un dominio es observado si al menos uno de los consumidores incluidos en la muestra hace parte del mismo.

El hecho de haber observado tan sólo el 27% de los dominios es la principal restricción que tiene usar estimadores directos de cara a la necesidad de conocer el comportamiento del consumo en todo el Distrito. Pese a ello, a partir del diseño de muestreo propuesto y empleando el estimador de Horvitz-Thompson (estimador directo) ha sido posible obtener tanto la estimación del número de consumidores de basuco como la varianza en cada dominio observado, lo que será un insumo esencial para poder implementar los estimadores de áreas pequeñas.

En relación con el nivel de precisión de las estimaciones obtenidas empleando los estimadores directos, se obtuvo en promedio un coeficiente de variación estimado del 35,9 %, lo cual es considerado un nivel de precisión bajo teniendo en cuenta que se está hablando de los 33 dominios observados y que el tamaño de muestra es representativo para las cuatro subredes que conforman el Distrito.

Ahora bien, dada la restricción mencionada anteriormente y los bajos niveles de precisión obtenidos al usar el estimador directo, empleando las fuentes de información secundarias mencionadas en la sección anterior, se propuso el uso de un modelo de áreas pequeñas a través del estimador de Fay-Herriot o Fay-Herriot espacial.

A partir del conjunto de información secundaria disponible, fueron realizados varios intentos por encontrar la mejor combinación de covariables que permitieran explicar las estimaciones obtenidas bajo el estimador directo en cada UPZ, encontrando que el modelo con mejor bondad de ajuste, entendida como aquel modelo que menor valor en los criterios de información reportó (Criterio de Información de Akaike-AIC y el Criterio de Información Bayesiano-BIC), fue conformado por las siguientes covariables:

- **Centro de Atención Inmediata (CAI):** unidad policial que intenta brindar contacto continuo con la comunidad en aras de prevenir el delito y la contravención, atender los requerimientos ciudadanos en materia de seguridad y convivencia pacífica mediante el empleo eficiente y oportuno de los elementos disponibles y medios tecnológicos aplicables.
- **Unidad Primaria Generadora de Datos (UPGD):** entidades públicas o privadas que captan la ocurrencia de eventos de interés en salud pública y genera información útil y necesaria para los fines de la Secretaría Distrital de Salud.
- **Consumidores de drogas:** consumidores de tabaco y alcohol
- **Kioskos:** espacios públicos destinados por la Alcaldía para la venta de alimentos, bebidas y golosinas.
- **Hoteles:** establecimientos de hostelería que ofrecen alojamiento y servicio de comedor.
- **Construcciones:** terrenos baldíos o lotes en proceso de construcción.
- **Lugares turísticos:** lugares de interés para los turistas normalmente por su valor cultural exhibido o inherente, su significancia histórica, su belleza, o por el interés al ofrecerles ofrece ocio, aventura y diversión.
- **Parques públicos:** parques cuentan con libre acceso para todos los visitantes que deseen hacer uso de ellos. Regularmente cuentan con diversos mobiliarios (juegos, zonas verdes, algunos cuentan con baños públicos, áreas de ejercicio, etc.) característicos de estos espacios ideales para el uso libre de los ciudadanos.
- **Teatros:** espacios idóneos para la puesta en escena totalmente en vivo frente a miles de espectadores un relato donde conjugan los actores, la música, el sonido, la escenografía, y sobre todo la gestualidad a la hora de interpretar cualquier emoción.
- **Entidades estatales:** cualquier tipo de institución del Estado encargada de guiar el funcionamiento de la comunidad.
- **Índice de pobreza multidimensional (IPM):** medición que identifica múltiples carencias a nivel de los hogares y las personas en los ámbitos de la salud, la educación y el nivel de vida

Los anteriores factores asociados al consumo de sustancias psicoactivas han sido medidos como conteos, a excepción del **Índice de pobreza multidimensional**, el cual ha sido medido como el puntaje promedio del indicador de los hogares registrados en el SISBEN por UPZ.

Una vez fueron definidas las covariables que conformarían el modelo a ajustar, se procedió a seleccionar el estimador más eficiente, Fay-Herriot o Fay-Herriot espacial, de cara a obtener las estimaciones más precisas, encontrando que el estimador de Fay-Herriot con componente espacial presentó un mejor ajuste visto como una reducción promedio del 13% en el coeficiente de variación de los dominios observados.

Con el planteamiento del modelo definido por las covariables mencionadas y el estimador de Fay-Herriot espacial, se obtuvo la estimación del número de consumidores de basuco presentada en la *Figura 3*.

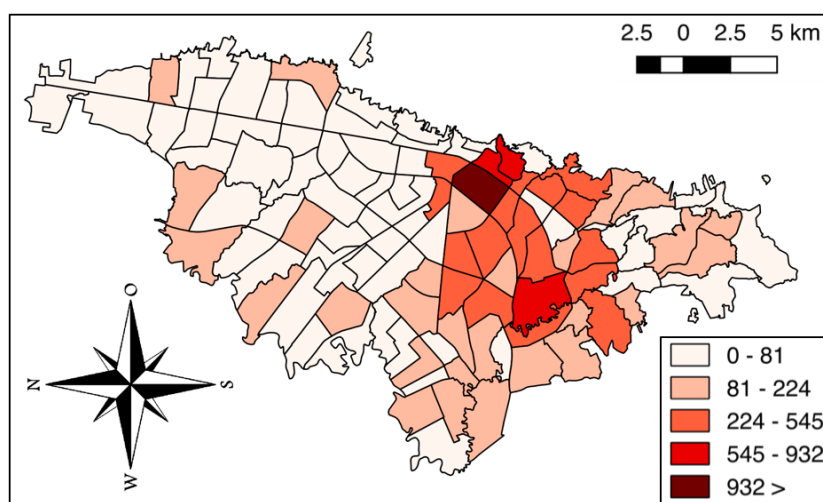


Figura 3: Número de consumidores estimados

La anterior *Figura* presenta la estimación del número de consumidores de basuco en cada UPZ, donde es posible visualizar la razón por la cual fue importante haber incluido el componente espacial en el modelo, pues hay una alta tendencia de consumo en la subred Centro Oriente, mientras que en la zona norte de la ciudad el número de consumidores de basuco es muy bajo.

Esta tendencia es posible medirla revisando el parámetro ρ del modelo, el cual tiene un rango de medición acotado entre -1 y 1. En casos donde este componente tome valores cercanos a -1, se asume que hay correlación espacial negativa en el fenómeno de interés, en caso contrario, donde el componente tenga valores cercanos a 1, la correlación es positiva. Para efectos del modelo empleado, el comportamiento del componente espacial reportó una correlación positiva muy fuerte ($\rho = 0,93$).

Por otra parte, al validar la precisión de las estimaciones reportadas en la *Figura 3* y empleando el coeficiente de variación estimado se encuentra que el nivel de precisión promedio es mucho más alto (17,5%) que el obtenido mediante el estimador directo (35,9%), con lo cual, se evidencia claramente la reducción alcanzada al emplear el estimador de áreas pequeñas.

En la *Figura 4* es reportado el coeficiente de variación para cada una de las UPZ que conforman el Distrito.

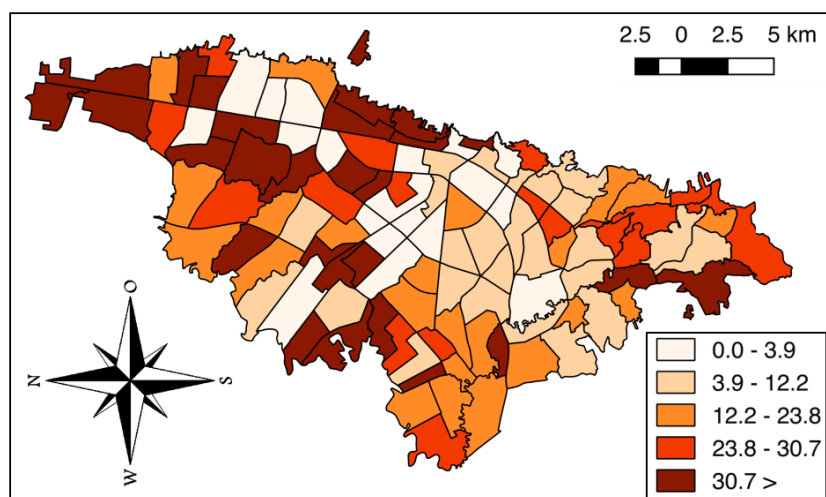


Figura 4: Coeficiente de variación estimado

El patrón observado obedece netamente a las condiciones geográficas de las UPZ, con el que por ejemplo, los dominios ubicados en la parte norte de Bogotá, son aquellos que a pesar de ser considerados urbanos colindan con los municipios periféricos y por ende, su comportamiento es totalmente disímil al de los dominios ubicados en el centro de la ciudad en términos no solo del consumo de sustancias psicoactivas sino de los factores asociados al consumo definidos en el modelo propuesto.

Por otra parte, a diferencia de las UPZ mencionadas, las ubicadas en la orientación opuesta, sector sur de la ciudad, son aquellas que reportan coeficientes de variación aceptables, lo cual va ligado a un mayor consumo de basuco, y esto, va relacionado a los grandes problemas socioeconómicos, de inseguridad y delincuencia de esta región de la ciudad.

Lo anterior en relación con las características de los dominios, ahora, al verlo en términos de los resultados presentados en la *Figura 3* y *Figura 4*, los resultados obtenidos son bastante consistentes con el comportamiento de las problemáticas observadas en Bogotá.

Si bien ha sido posible emplear el estimador de áreas pequeñas como herramienta para la obtención de los resultados presentados, se quizó validar si las estimaciones obtenidas son altamente dependientes de las características de la muestra seleccionada. Por esta razón, la implementación del método de bootstrap permitió contrastar cada uno de los resultados obtenidos

Lo primero a visualizar es la frecuencia con la cual fue muestreada cada una de las UPZ al momento de realizar las 1.000 replicas propuestas, donde es fundamental resaltar que hubo 6 UPZ no seleccionadas en ninguna de las replicas como se observa en la *Figura 5*.

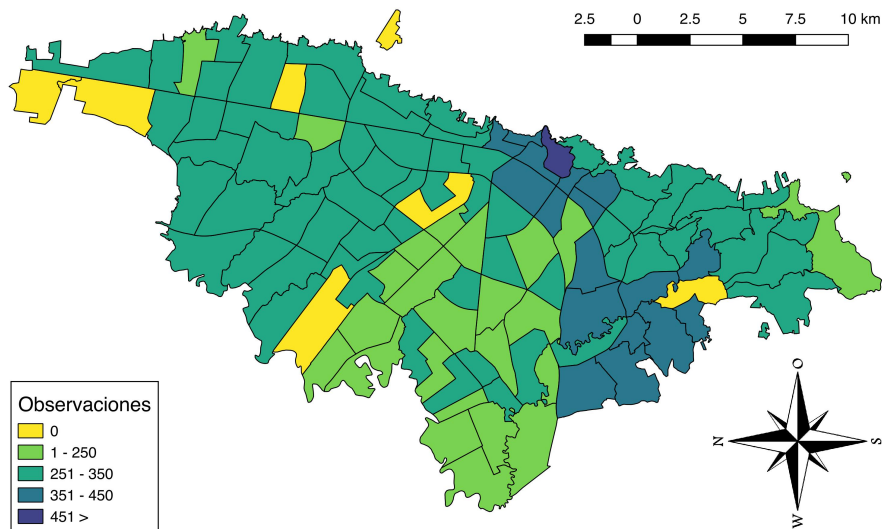


Figura 5: Frecuencia de observación de los dominios

De la anterior, también es importante enfatizar en el escenario contrario al mencionado, pues es posible observar la alta frecuencia de selección en algunas de las UPZ, principalmente en la subred Centro Oriente, donde de acuerdo a las estimaciones presentados en la *Figura 3* de la muestra inicial se concentra el mayor número de casos asociados a este fenómeno.

Con base en el comportamiento mencionado previamente, en la *Figura 6* se presenta la estimación del número de consumidores de basuco, obtenida como la mediana de las mediciones obtenidas en las 1.000 replicas, para cada una de las UPZ.

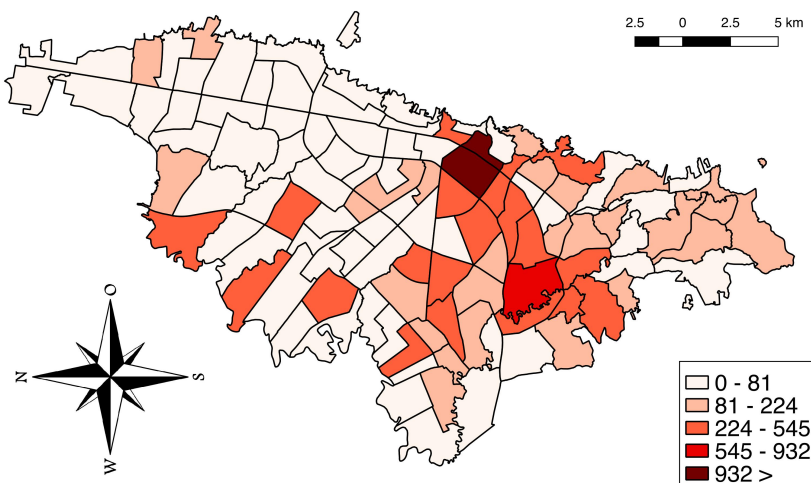


Figura 6: Número de consumidores estimados

Los resultados presentados en la anterior *Figura 6* no coinciden con los reportados en la *Figura 3*, no obstante, la discrepancia entre los dos resultados no resulta ser significativa.

Por el contrario, el patrón geográfico de consumo se mantiene en las mismas UPZ donde había sido señalado inicialmente, de tal forma que aún cuando la magnitud de consumo en algunas de las UPZ que conforman la subred Centro Oriente ha variado, el escenario en el que se presenta una fuerte concentración

de consumidores de basuco en este sector de la ciudad se mantiene.

Dicho de esta manera, no se evidencian cambios considerables en los resultados obtenidos a partir del proceso iterativo y la muestra particular analizada, no solo en cuanto a las estimaciones presentadas sino que al estimar el parámetro ρ por medio del bootstrap, se obtuvo un coeficiente $\rho = 0,84$, el cual, a pesar de ser menor al de la muestra particular (0,93), sigue siendo sinonimo de la existencia de una fuerte correlación espacial.

Ahora bien, al momento de revisar el contraste de los coeficientes de variación estimados se genera la presentado en la *Figura* . En esta *Figura*, se observa que en general hay una fuerte reducción de la magnitud de los coeficientes de variación.

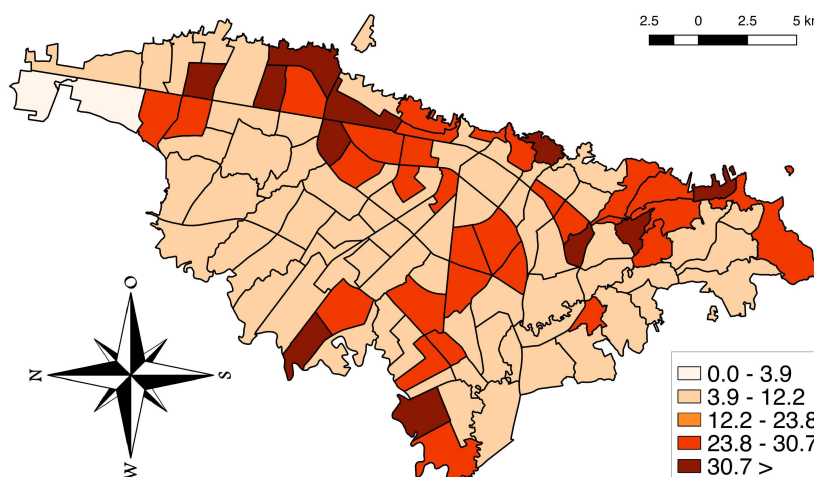


Figura 7: Coeficiente de variación estimado

Estas variaciones obedecen a que la mediana, siendo la medida empleada para obtener la estimación del coeficiente de variación, es una medida robusta que no es vulnerable a la presencia de valores atípicos, con lo cual, al realizar las 1.000 replicas del proceso iterativo y ante la posibilidad de que las estimaciones obtenidas en cada una de ellas reporte casos extremos, la estimación definitiva no presenta los mismos niveles reportados en la *Figura 4*, pero lo que si ocurre, es que hay un incremento en la mayoría de UPZ donde se reportaban coeficientes de variación muy bajos.

Como alternativa para cotejar los resultados de los dos ejercicios, en la *Tabla 1* se presenta la comparación de los resultados obtenidos mediante los dos tipos de estimadores empleados en términos de la distribución de los coeficientes de variación estimados, entendiendo que los valores correspondientes al estimador de áreas pequeñas presentados allí, han sido calculados con base en lo presentado en la *Figura 4* y *Figura 7*.

Estimador	Escenario	Medidas de interés				
		Dominios observados	Percentil 25	Percentil 50	Percentil 75	Promedio
Directo	Muestra	33	17,7 %	33,2 %	44,6 %	35,9 %
Directo	Bootstrap	33	27,6 %	38,1 %	51,1 %	42,4 %
SAE	Muestra	112	8,5 %	14,4 %	24,4 %	17,5 %
SAE	Bootstrap	112	13,7 %	20,5 %	26,4 %	19,9 %

Tabla 1: Precisión de los estimadores

Tal como se mencionó anteriormente y como es posible validarlo en la *Tabla 1*, hubo una reducción significativa en el promedio de los coeficientes de variación estimados obtenidos mediante el estimador directo y el estimador de áreas pequeñas, denotado en la *Tabla 1* por sus siglas en inglés SAE, lo

cual es visto como una diferencia de 18 puntos porcentuales, equivalentes a obtener una reducción de aproximadamente el 50 % a favor del estimador de áreas pequeñas.

Con todo y esto, al examinar los diferentes percentiles presentados en la *Tabla 1* y realizar la comparación entre los dos estimadores, es contundente la diferencia encontrada a favor del estimador de Fey-Herriot, estrategia con la cual se redujo en por lo menos a la mitad cada medida de referencia (percentiles reportados).

Inmediatamente, es necesario examinar lo ocurrido con las estimaciones del proceso iterativo, tanto en el estimador directo como en el estimador de áreas pequeñas, encontrando que en los percentiles reportados del bootstrap hay un leve incremento de los mismos, lo cual sucede a raíz de emplear la mediana como estrategia de estimación del método bootstrap.

Pese a ello y aún cuando se ha evidenciado la reducción significativa en los coeficientes de variación conseguidos con el estimador de Fay-Herriot espacial, en la *Figura 4* y *Figura 7* se evidencian coeficientes de variación altos que a priori impiden realizar un proceso de inferencia acerca del número estimado de consumidores de basuco.

En términos generales, si bien es cierto que el ideal sería obtener coeficientes de variación estimados muy bajos, los obtenidos y en particular los que se encuentran en el rango superior al 30,7 % son causados por la baja prevalencia del fenómeno observado en estos dominios es baja y con ello, la posibilidad de observar el fenómeno con un nivel alto de precisión se convierte en una tarea bastante compleja de realizar.

Para ello, hay varias soluciones, una de ellas es incrementar el tamaño de muestra del diseño de muestreo base para el estimador de áreas pequeñas, que quizá puede mejorar los resultados obtenidos, empero, puede que esto no ocurra por la particularidad del fenómeno en medición lo cual ha sido posible evidenciarlo con los resultados del proceso iterativo implementado. Realmente, la solución a este tipo de situaciones radica en contar con registros administrativos sólidos, consolidados y de calidad.

6. Conclusiones

A lo largo del presente documento se han venido mencionando los hallazgos obtenidos en el desarrollo del proceso investigativo. Sin embargo, a continuación se sintetizan las ideas que dan cumplimiento a los objetivos planteados:

- A partir del conjunto de datos empleado, es posible afirmar que son más precisos los estimadores de áreas pequeñas en relación con los estimadores directos al momento de obtener la estimación del número de consumidores de basuco por UPZ. Esto ha sido posible determinarlo al realizar la comparación de los coeficientes de variación estimados en los 33 dominios observados, donde los estimadores directos reportaron un coeficiente de variación estimado promedio del 35,9 %, mientras que empleando los estimadores de áreas pequeñas fue de 17,5 %.
- Ha sido posible obtener estimaciones para cada una de las 112 UPZ urbanas. Es importante resaltar la ganancia que se obtiene con este resultado, ya que normalmente cuando se emplean estimaciones netamente basadas en el diseño, solo se podría realizar inferencia sobre los dominios observados (33 UPZ) y los bajos niveles de precisión en cada uno de ellos no podría generar confiabilidad en la interpretación de los resultados.
- Existe una alta tendencia de consumo en algunos sectores del Distrito, principalmente en la subred Centro Oriente. (Ver *Figura 3*). Esto ha sido posible detectarlo por el componente espacial incluido en el modelo. La correlación espacial (ρ) fue 0,93.
- Los resultados obtenidos por medio de la muestra particular analizada son totalmente consistentes. Esto es posible afirmarlo dado que las comparaciones realizadas con los resultados del proceso iterativo no generan discrepancias fuertes, con lo cual, no existe una dependencia entre patrón

encontrado y la muestra seleccionada, sino que efectivamente las estimaciones obtenidas coinciden con el comportamiento real del consumo de basuco en el Distrito.

- Al revisar el agregado de información presente en el VESPA, existe un alto nivel de subestimación del número de consumidores de basuco intervenidos en el sistema, pues de acuerdo a las estimaciones obtenidas, es mucho mayor el número de casos existentes en relación con el número de casos intervenidos.
- Las localidades que conforman la subred Centro Oriente requieren intervenciones trazadas por políticas públicas que permitan contrarrestar el resultado negativo de las estimaciones, ya que a partir de las estimaciones obtenidas, esta subred congrega cerca de 7.000 consumidores, equivalente a poco más del doble de lo estimado en la subred Sur Occidente.
- La posibilidad y facilidad para consultar los diferentes sistemas de información de las entidades del gobierno es nula, no solo por las restricciones existentes para hacerlo sino que la calidad y pertinencia de los mismos es muy pobre, lo cual es importante resaltarlo, no exactamente por la demanda de tiempo que esto exigió, si por la baja calidad en la información que se le brinda al investigador.

Discusión

Si bien fue posible alcanzar los objetivos planteados en el presente estudio y se tiene la posibilidad de conocer el comportamiento de la frecuencia de consumo de basuco en el Distrito y e incorporar el uso de técnicas estadísticas que apoyen la formulación de políticas públicas dirigidas al control de este fenómeno, queda mucho por ahondar en la aplicación de la técnica estadística empleada.

Sin embargo, los resultados obtenidos son de vital importancia para la toma de decisiones, permitiendo ampliar la cobertura en atención que brinda la SDS. Sabiendo que por directrices político administrativas existen restricciones y limitaciones en la asignación de recursos, personal y atención por el despliegue operativo y logístico que demanda el VESPA, al emplear los resultados obtenidos como estrategia de focalización se puede garantizar la oportuna y efectiva intervención especializada que generé la disminución en la frecuencia de consumo de sustancias psicoactivas y de esta manera, sea posible mejorar las condiciones de salud de los consumidores.

Futuras líneas de investigación

Este estudio puede continuar su desarrollo si se realizan ajustes al modelo propuesto o si se complementa con otro tipo de técnicas estadísticas relacionadas con estimación de áreas. Una primera propuesta es realizar la comparación del ajuste del modelo empleando diferentes componentes espaciales, lo cual permitiría determinar: i) si efectivamente la correlación espacial obtenida en este estudio (0,93) es consistente o es efecto del tipo de componente empleado. ii) Probar otros componentes espaciales y determinar cuál componente es el adecuado para el conjunto de datos disponibles.

Por otra parte, se podría ajustar un modelo espacio temporal, donde sea de vital importancia analizar la alta correlación espacial existente periódicamente, ya sea teniendo mediciones semestrales o anuales.

Agradecimientos

A Diego Fernando, quien además de ser mi director de tesis, se ha convertido en un gran amigo. Mi más sincera expresión de gratitud por su constante apoyo y motivación, por el tiempo dedicado y el asesoramiento incondicional oportuno en todo momento.

Gracias a todas las personas que ayudaron directa e indirectamente en la realización de este proyecto.

Referencias

- Aguado, I. H. (1994), *Glosario de términos de alcohol y drogas*, Ministerio de Sanidad y Consumo de España.
- Bajekal, M. & Scholes, S. (2004), ‘Synthetic estimation of healthy lifestyles indicators: Stage 1 report’, *National Centre for Social Research*.
- Barreto, P. & Perez, M. (2015), ‘Consumo de sustancias psicoactivas legales e ilegales, factores de protección y de riesgo: estado actual’, *Revista de Investigación en Salud* **2**(1), 31–50.
- Bharat Singh, G. K. . D. K. (2005), ‘Spatio-temporal models in small area estimation’, *Survey Methodology* **31**(2), 183–195.
- Carl-Erik Särndal, B. S. (2003), ‘Model assisted survey sampling’, *Springer*.
- Castro, R. (2005), ‘Consumo de drogas en el Perú’, *Debate Agrario* **39**(127), 54.
- Efron, B. (1979), ‘Bootstrap methods: Another look at the jackknife’, *Ann Statist* **7**.
- Fan, C., Muller, M. & Rezucha, I. (1962), ‘Development of sampling plans by using sequential (item by item) selection techniques and digital computer’, *Journal of the American Statistical Association* **57**, 387–402.
- Fay, R. E. & Herriot, R. A. (1979), ‘Estimates of income for small places: An application of James-Stein procedures to census data’, *Journal of the American Statistical Association* **74**(366), 269–277.
- González, D. M. (2015), *Estimación en áreas pequeñas: Métodos basados en modelos*, Universidad Miguel Hernández de Elche.
- Gurney, M. & Daly, J. F. (1965), ‘A multivariate approach to estimation in periodic sample surveys’, *U.S Bureau of the Census*.
- Gutierrez, H. A. (2009), *Estrategias de muestreo: Diseño de encuestas y estimación de parámetros*, Ediciones USTA Universidad.
- Gutierrez, H. A. (2015), *Teaching Sampling R Package*.
- Gutierrez, H. Z. . A. (2010), *Teoría estadística: aplicaciones y métodos.*, Ediciones USTA Universidad.
- Horvitz, D. & Thompson, D. (1952), ‘A generalization of sampling without replacement from a finite universe’, *Journal of the American Statistical Association* **47**, 663–685.
- Lukasz Wawrowski, M. (2016), ‘The spatial fay-herriot model in poverty estimation.’, *Folia Oeconomica Stetinensia*.
- Lundström, C.-E. S. . (2005), *Estimation in Surveys with Nonresponse*, Wiley Series.
- McCulloch, C. E. & Searle, S. R. (2001), *Generalized, Linear and Mixed models*, Wiley Series in Probability and Statistics.
- Molina, M. H. . I. (2009), ‘Estimadores de áreas pequeñas basados en modelos para la encuesta de población activa’, *Estadística Española* **51**(170), 133–172.
- N.G.N. Prasad, J. R. (1990), ‘The estimation of the mean squared error of small-area estimators’, *Journal of the American Statistical Association* **85**(409), 163–171.

- Paños, R. L. (2000), 'Estimación para áreas pequeñas (1)', *Estadística Española* **42**(146), 291–338.
- Rao, J. & Molina, I. (2015), *Small Area Estimation*, Wiley Series in Survey Methodology.
- Salvati, M. P. . N. (2008), 'Small area estimation: the eblup estimator based on spatially correlated random area effects', *statistical Methods and Applications* **17**(1), 113–141.
- Schworer, A. (2004), 'Newton-raphson versus fisher scoring algorithms in calculating maximum likelihood estimates', *Undergraduate Mathematics Day* **1**.
- Sherman, C. (2017), 'The defining features of drug intoxication and addiction can be traced to disruptions in neuron-to neuron signaling.', *NIDAnotes* .
- Sunter, A. (1986), 'Solutions to the problem of unequal probabilities sampling without replacement', *International Statistical Review* **54**, 33–50.
- Tibshirani, B. E. . R. (1993), 'An introduction to the bootstrap', *Champman and Hall, New York* .
- Torabi, M. & Rao, J. N. K. (2010), *Mean squared error estimators of small area means using survey weights*, John Wiley and Sons.
- Yolanda Marhuenda, I. M. . D. M. (2013), 'Small area estimation with spatio-temporal fay–herriot models', *Computational Statistics and Data Analysis* **58**, 308–325.
- Zea, F. O. . J. (2018), 'Small area estimation methodology (sae) applied on bogota multipurpose survey (emb)', *Romanian Statistical Review* **1**.
- Zhang, M. L. . X. D. (1988), 'The difficulty of improving statistical', *Bureau of the census* .

Anexo A

En la *Tabla 2*, se presentan los coeficientes del modelo propuesto para la obtención de las estimaciones de la frecuencia de consumo de basuco.

Variable	Coefficiente	Error Estándar	P Valor
Intercepto	14.46	13.54	0.16
CAI	-35.52	28.78	0.02
UPGD	-2.86	4.43	0.05
Consumidores tabaco	0.39	0.99	0.07
Consumidores alcohol	-2.03	0.85	0.02
Kioskos	6.88	2.19	0.00
Hoteles	3.15	1.40	0.02
Construcciones	0.78	0.29	0.01
Lugares turísticos	13.89	4.16	0.00
Parques públicos	2.01	0.50	0.00
Teatros	-9.21	11.49	0.19
Entidades estatales	-44.65	10.20	0.00
IPM	93.12	40.98	0.02

Tabla 2: Coeficientes de los parámetros del modelo