Biodiversidad

Cobertura geográfica

La cobertura geográfica del estudio comprendió el total del territorio nacional, abarcando tanto entidades a nivel departamental como municipal y regional, de acuerdo con la definición de la población objetivo. En el nivel departamental, se incluyeron las 33 Secretarías de Planeación de los departamentos, incorporando al Distrito Capital, garantizando así la representación de todas las regiones administrativas del país. A nivel regional, se cubrió la totalidad de las 46 Corporaciones Autónomas Regionales (CAR), Corporaciones de Desarrollo Sostenible (CDS) y Autoridades Ambientales Urbanas/Establecimientos Públicos Ambientales (EPAS), permitiendo una visión integral de la gestión de la biodiversidad en las distintas jurisdicciones ambientales.

En el nivel municipal, la cobertura se centró en los 567 municipios que cuentan con al menos uno de los ecosistemas definidos como prioritarios para el estudio: bosques, bosque seco tropical, páramo, playa, humedales, zona pantanosa, áreas con procesos de deforestación o que hacen parte del Sistema Nacional de Áreas Protegidas (SINAP). Dentro de esta población, los municipios con cuatro o más ecosistemas fueron incluidos en su totalidad, mientras que el resto se seleccionó mediante muestreo balanceado, asegurando la representación de las distintas combinaciones ecosistémicas presentes en el país.

Este enfoque permitió que la cobertura territorial contemplara la diversidad geográfica y ambiental de Colombia, incluyendo áreas rurales, urbanas y de especial interés ecológico, y garantizando la posibilidad de realizar análisis desagregados por ecosistema, región y tipo de autoridad ambiental. Al incorporar entidades de todos los niveles de gobierno y de las distintas zonas ecológicas, el estudio ofrece una representación estadísticamente robusta de la gestión y planificación territorial en materia de biodiversidad

Unidades de Análisis

La unidad de análisis corresponde a las alcaldías, gobernaciones y CAR.

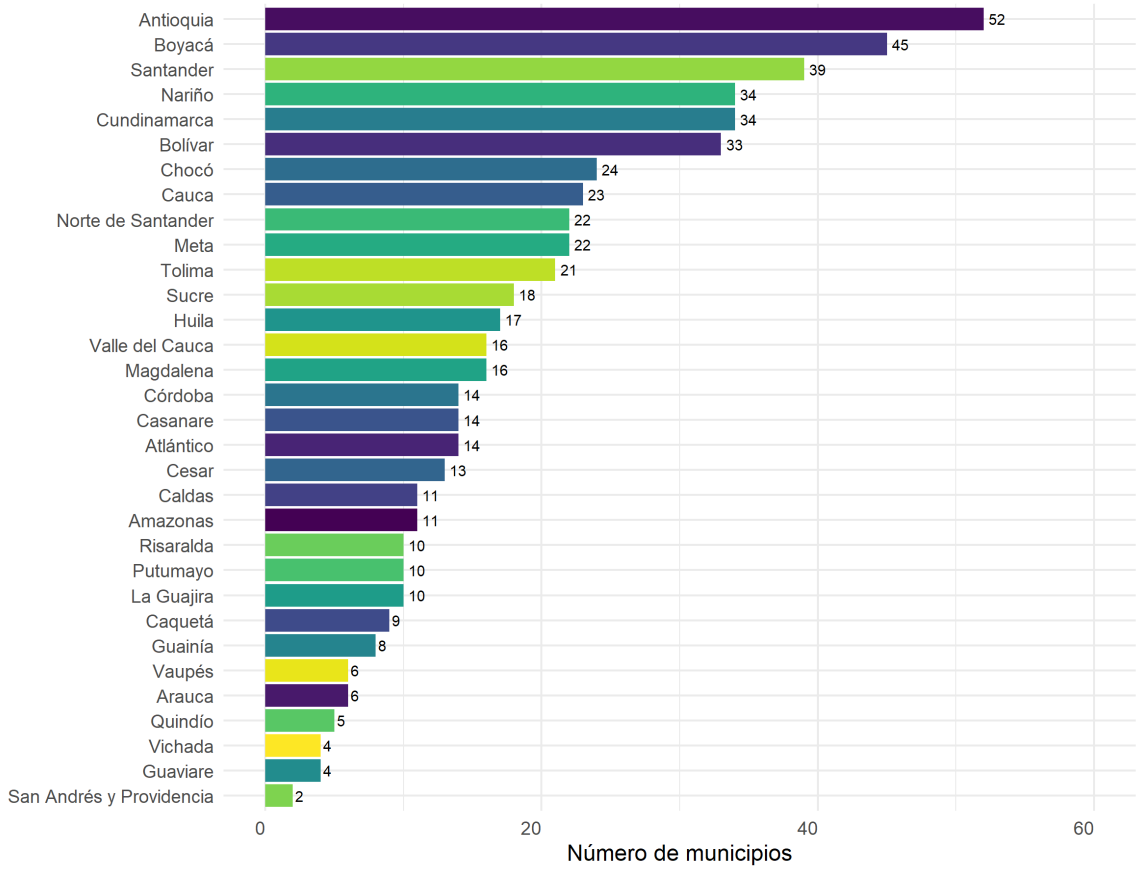
Las unidades de muestreo son las Secretarías de Planeación de los 33 departamentos (incluido Bogotá); así como a las 46 Corporaciones Autónomas Regionales (CAR), Corporaciones de Desarrollo Sostenible (CDS) y Autoridades Ambientales Urbanas/ Establecimientos Públicos Ambientales (EPAS). Y las 567 Secretarías de Planeación de municipios que cuentan con alguno con alguno de los siguientes ecosistemas: bosques, bosque seco tropical, páramo, playa, humedales, zona pantanosa, deforestación o Sistema Nacional de Áreas Protegidas -SINAP.

Universo

El universo de estudio está conformado por las 33 Secretarías de Planeación de departamentos (incluido Bogotá); así como a las 46 Corporaciones Autónomas Regionales (CAR), Corporaciones de Desarrollo Sostenible (CDS) y Autoridades Ambientales Urbanas/ Establecimientos Públicos Ambientales (EPAS), y 567 Secretarías de Planeación de municipios que cuentan con alguno de los siguientes ecosistemas: bosques, bosque seco tropical, páramo, playa, humedales, zona pantanosa,

deforestación o Sistema Nacional de Áreas Protegidas -SINAP, en la Figura 1 se presenta la distribución de la cantidad de municipios con ecosistemas por departamento.

Figura 1: Cantidad de municipios por departamento



Tipo de Muestreo

El diseño muestral adoptado para la evaluación combinó un enfoque censal con un esquema probabilístico estratificado y balanceado, garantizando representatividad y control sobre los sesgos de estimación. En el caso de las 33 Secretarías de Planeación departamentales y las 46 CAR, CDS y EPAS, se aplicó un censo, asignando a todas las unidades una probabilidad de inclusión igual a uno, lo que asegura la cobertura total de estos segmentos.

Para las Secretarías de Planeación municipales, el marco muestral se estructuró en dos estratos. El primero estuvo conformado por 47 municipios con alta diversidad ecosistémica (cuatro o más ecosistemas), a los cuales se aplicó inclusión forzosa. El segundo incluyó los 520 municipios restantes, donde se implementó un muestreo probabilístico balanceado mediante el método del cubo. Este algoritmo selecciona muestras que cumplen exactamente con las probabilidades de inclusión predefinidas y maximizan el balanceo con respecto a variables auxiliares —en este caso, la presencia de distintos ecosistemas—, siguiendo las fases de vuelo y aterrizaje descritas por Deville y Tillé.

La determinación del tamaño muestral en el estrato 2 se calculó para estimar proporciones del 50% con un 95% de confianza y un margen de error inferior al 5%, fijando un tamaño mínimo de 241 encuestas y adicionando un 10% de sobremuestra para mitigar la no respuesta.

Método de Recolección de Datos

La recolección de la información se realizó mediante encuestas presenciales dirigidas a los Secretarios de Planeación —o sus delegados— en cada una de las unidades de muestreo definidas. En el caso de las CAR, CDS y EPAS, así como de las Secretarías de Planeación departamentales, se cubrió la totalidad de la población, mientras que para las Secretarías de Planeación municipales se aplicó el esquema muestral establecido. El levantamiento de la información siguió protocolos estandarizados para garantizar la comparabilidad y calidad de los datos, incluyendo la capacitación del equipo encuestador, el uso de cuestionarios estructurados y la supervisión continua del trabajo de campo. Asimismo, se implementaron mecanismos de control de calidad y registro de incidencias, permitiendo documentar rechazos, dificultades de contacto y otras eventualidades, de acuerdo con las definiciones de la AAPOR (2016), asegurando así la integridad del proceso y la confiabilidad de los resultados.

Descripción de la Construcción de Pesos

Durante el proceso de recolección se llevó el control de incidencias a nivel de modalidad y segmento (CAR, gobernaciones y municipios), estas son codificadas según las normas AAPOR (2016), según lo recomendado por Valliant, Dever, y Kreuter (2018). En la Tabla 1 se presenta la cantidad de incidencias.

Tabla 1: Cantidad de incidencias por segmento de encuesta

Segmento	Incidencia	n
CAR, CDS y autoridades ambientales	ER	46
Secretarías de Planeación de departamentos	ER	33
Secretarías de Planeación de municipios	ED	19
	ENR	9
	ER	262

En donde ER representa a las encuestas efectivas, ENR corresponde a las unidades que rechazaron la encuesta, NEC es el conjunto de no elegibles conocidos y ED es el conjunto de elegibles desconocidos.

Como se observa en la tabla 1 los principales problemas obedecieron a problemas de contactabilidad (ED).

De acuerdo con Valliant, Dever, y Kreuter (2018) la tasa de respuesta se define como:

$$TR = \frac{ER}{ER + ENR}$$

Que corresponde a los elementos que responden a la encuesta dividido entre el total de elementos que pudieron ser contactadaos (ER+ENR), de manera que la tasa de no respuesta es: $TNR = 1 - TR$. La 2 presenta los resultados obtenidos en las Secretarías de Planeación Municipal al finalizar la operación estadística.

Tabla 2: Tasa de respuesta y de no respuesta

Segmento	TR	TNR
Secretarías de Planeación de municipios	96.68%	3.32%

La selección de la muestra de municipios consideró inicialmente una sobremuestra del 10% para alcanzar la muestra mínima de 241 encuestas, así que la tasa de reemplazos se calcula como:

$$RR = \frac{n_{reg}}{1.1n}$$

En donde n corresponde al tamaño mínimo de muestra requerido y n_{reg} es la cantidad de registros suministrados para alcanzar la muestra mínima. De acuerdo con los resultados de la tabla 3 se requieron entre 1.04 veces la cantidad de registros inicialmente planeados. Para realizar los reemplazos se utilizó un modelo de propensión buscando la unidad de muestreo más cercana según el método de emparejamiento del vecino más cercano (nearest neighbor matching). Para la selección de reemplazos se utilizó el paquete MatchIt del paquete R Core Team (2024) que cuenta con el método descrito.

Tabla 3: Tasa de reemplazos

Segmento	Registros	Muestra mínima	RR
Secretarías de Planeación de municipios	290	241	1,09

De acuerdo con lo anterior, los parámetros de calidad que permiten evaluar la precisión, validez y representatividad de la muestra respecto a la población de interés se obtienen a partir del margen de error, el cual resulta en 4.5% con el tamaño de muestra efectivo, y que es inferior al 5% propuesto inicialmente.

En las 33 Secretarías de Planeación departamentales y 46 CAR Regionales, CDS y EPAS se realizó un censo, por lo tanto, no se usa un algoritmo de selección. De igual manera ocurre en el estrato de inclusión forzosa de las secretarias de planeación.

Para el estrato con muestreo balanceado en la población de secretarias de planeación municipal se utilizó el método del cubo. El algoritmo de muestreo selecciona una muestra balanceada y satisface exactamente un conjunto de probabilidades de inclusión. El método se basa en una transformación aleatoria del vector de probabilidades de inclusión hasta que se obtiene una muestra tal que (Tillé 2011):

1. Las probabilidades de inclusión, π_k, se satisfacen exactamente.

2. Las ecuaciones de balanceo se satisfacen en la mayor medida posible.

La muestra puede ser representada por un vector de indicadores, así:

$$\mathbf{s} = (I_{s1}, \dots, I_{sk}, \dots, I_{sN})'$$

En donde I_{sk} toma el valor 1 si $k \in s$ y 0 cuando no. Así, una muestra puede verse como un vértice de un cubo N-dimensional, como se muestra en la Figura 2.

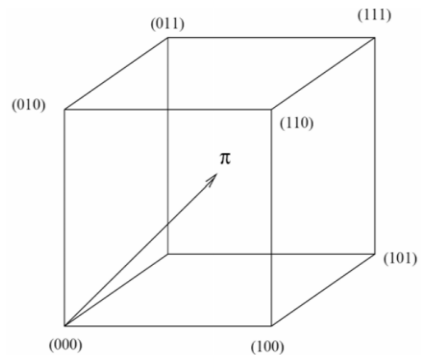


Figura 2: Ilustración de una muestra de una población de tamaño N=3

El método del cubo se divide en dos fases: la fase de vuelo y la fase de aterrizaje. La fase de vuelo es un paseo aleatorio que comienza en el vector de probabilidades de inclusión y permanece en la intersección del cubo y el subespacio de restricciones. Este paseo aleatorio se detiene en un vértice de la intersección del cubo y el subespacio de restricciones. Al final de la fase de vuelo, si no se ha obtenido una muestra, la fase de aterrizaje implica seleccionar una muestra que esté lo más cerca posible del subespacio de restricciones.

La fase de vuelo es un paseo aleatorio en la intersección del subespacio de balanceo y el cubo. Este paseo aleatorio se detiene en un vértice de la intersección del cubo y el subespacio. La fase de vuelo se define por la siguiente clase de algoritmos. Primero, se inicializa con $\boldsymbol{\pi}^{(\mathbf{0})} = \boldsymbol{\pi}$, donde $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)^T$ el vector de probabilidades de inclusión para las secretarías de planeación. A continuación, en el tiempo $t = 0, \dots, T$, se sigue el siguiente procedimiento:

3. Generar cualquier vector $\mathbf{u}^{(t)} = \left[u_k^{(t)}\right] \neq 0$ tal que:

   – $A\mathbf{u}^{(t)} = 0$ con $A = \left[\frac{\mathbf{x}_1}{\pi_1}, \dots, \frac{\mathbf{x}_N}{\pi_N}\right]$.

   – $u_k^{(t)} = 0$ si $\pi_k^{(t)}$ es entero.

4. Calcular $\lambda_1^{*(t)}$ y $\lambda_2^{*(t)}$, como los valores más grandes tales que:

$$0 \leq \boldsymbol{\pi}^{(\mathbf{t})} + \lambda_1^{*(t)}\mathbf{u}^{(t)} \leq 1$$

$$0 \leq \boldsymbol{\pi}^{(\mathbf{t})} - \lambda_2^{*(t)}\mathbf{u}^{(t)} \leq 1$$

5. Calcular:

$$\boldsymbol{\pi}^{(\mathbf{t+1})} = \begin{cases} \boldsymbol{\pi}^{(\mathbf{t})} + \lambda_1^{*(t)}\mathbf{u}^{(t)} & \text{con probabilidad } q_1(t) \\ \boldsymbol{\pi}^{(\mathbf{t})} - \lambda_2^{*(t)}\mathbf{u}^{(t)} & \text{con probabilidad } q_2(t) \end{cases}$$

En donde $q_1(t) = \frac{\lambda_1^{*(t)}}{\lambda_1^{*(t)}+\lambda_2^{*(t)}}$ y $q_2(t) = 1 - q_1(t)$.

La fase de vuelo termina cuando ya no es posible encontrar un vector $\mathbf{u}^{(t)} \neq 0$.

Si al final de la fase de vuelo las ecuaciones de balanceo no se satisfacen exactamente, es necesario pasar a una fase de aterrizaje. Sea $\boldsymbol{\pi}^* = [\boldsymbol{\pi}^*]$ el vector obtenido en el último paso de la fase de vuelo. Se puede demostrar (ver Deville y Tillé 2004) que

$$\text{card}(U^*) \leq p$$

En donde

$$U^* = \{k \in U \mid 0 < \pi_k^* < 1\}$$

y p es el número de variables de balanceo. El objetivo de la fase de aterrizaje es encontrar una muestra s tal que $E(s \mid \pi^*) = \pi^*$, que sea casi balanceada.

Para la selección de la muestra se utilizó el paquete sampling del software R Core Team (2024) que cuenta con los métodos descritos.

Al inicio, todas las unidades seleccionadas tienen su factor base de expansión, independiente de si contestó a la encuesta. El cual se calcula así:

$$d_{0k} = \frac{N_h}{n_h},$$

En donde N_h es la cantidad de municipios con ecosistemas de biodiversidad y n_h es la cantidad de municipios seleccionados, ambos en el estrato de muestreo h. La tabla 4 resume el universo, registros seleccionados y el factor de expansión base del diseño.

Tabla 4: Factores básicos de expansión

Segmento	estrato	N	n	d0k
Secretarías de Planeación de municipios	1	47	47	1,0
Secretarías de Planeación de municipios	2	520	243	2,1

Debido a que existen unidades seleccionadas que no pudieron ser contactadas, entonces el factor de expansión se ajusta por elegibilidad desconocida, así $a_1 = \sum_{k \in s_b} d_{0k} \,/\, \sum_{s_{b_{KN}}} d_{0k}$, donde s_b es el conjunto de todas las incidencias presentadas y $s_{b_{KN}}$ es el conjunto de incidencias con elegibilidad conocida.

El factor de expansión ajustado por elegibilidad desconocida es

$$d_{1k} = a_1 * d_{0k}$$

El número de registros después de eliminar las unidades con elegibilidad desconocida y el factor de expansión ajustado por esta misma incidencia se presenta en la tabla 5

Tabla 5: Factor de expansión ajustado por elegibilidad desconocida

Segmento	Estrato	Registros	d1k
Secretarías de Planeación de municipios	1	44	1,07
Secretarías de Planeación de municipios	2	227	2,29

Aprovechando que se cuenta con información auxiliar para todos los registros del marco muestral, se usa un modelo de propensión para ajustar la no respuesta. De acuerdo con Kim y Riddles (2012), se usa un modelo de *propensity score* donde la variable de resultado se codifica como 1 cuando es un respondiente y 0 cuando no. Las variables de soporte que son utilizadas para modelar los efectos sobre la probabilidad de responder fueron los diferentes ecosistemas presentes en cada municipio. De esta forma es posible estimar el patrón de ausencia de respuesta y por ende corregir los posibles sesgos. El modelo de propensión es:

$$\phi_k = \left(1 + e^{-\mathbf{z_k^T}\boldsymbol{\beta}}\right)^{-1},$$

de esta forma el factor ajustado por la no respuesta para el municipio k que respondió a la encuesta es $d_{3k} = d_{2k}/\phi_k$, donde d_{2k} es el factor ajustado por elegibilidad desconocida y defectos del marco. La gráfica 1 ilustra la probabilidad estimada, en donde se destaca que la distribución tiende al valor 1, indicando que el factor no se ve impactado de una manera importante por este criterio como se puede apreciar en la Tabla 6.

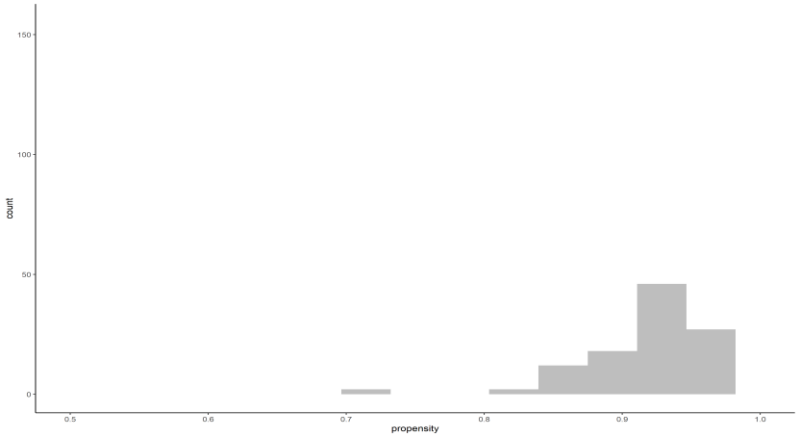


Figura 3: Histograma de la propensión de respuesta

Tabla 6: Factores de expansión finales

Segmento	Estrato	Factor promedio
Secretarías de Planeación de municipios	1	1,09

Segmento	Estrato	Factor promedio
Secretarías de Planeación de municipios	2	2,37

Referencias

AAPOR. 2016. «Standard definitions: Final dispositions of case codes and outcome rates for surveys». *Encyclopedia of Survey Research Methods*.

Deville, Jean-Claude, y Yves Tillé. 2004. «Efficient balanced sampling: the cube method». *Biometrika* 91 (4): 893-912.

Kim, Jae Kwang, y Minsun Kim Riddles. 2012. «Some theory for propensity-score-adjustment estimators in survey sampling». *Survey Methodology* 38 (2): 157-65.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Tillé, Yves. 2011. «Ten years of balanced sampling with the cube method: an appraisal». *Survey methodology* 37 (2): 215-26.

Valliant, Richard, Jill A Dever, y Frauke Kreuter. 2018. «Basic steps in weighting». En *Practical Tools for Designing and Weighting Survey Samples*, 321-67. Springer.