Capítulo V

Aplicación de la metodología desarrollada para el sector agropecuario

En los procesos de investigación existe una marcada tendencia a simplificar conceptos y procedimientos para calcular el tamaño de la muestra, pero este paso no es sencillo ni puede ser tomado a la ligera (Kerlinger & Lee, 2002). Algunos autores prefieren emplear métodos intuitivos, sin ningún soporte estadístico, como cuando estiman el tamaño de la muestra asignando un valor proporcional al total de la población o simplemente eligiendo un valor muy grande (Kerlinger & Lee, 2002; Valdivieso Taborga et al., 2011). Generalizan al momento de hacer inferencias o extrapolar y sacar conclusiones sobre la población de la cual se ha extraído la muestra (Otzen & Manterola, 2017), y con esto limitan la posibilidad de hacer una descripción más aproximada de la realidad de la población (Perfetti et al., 2012).

Una práctica frecuente en los documentos científicos consiste en decir que se empleó un determinado tamaño de muestra, pero sin detallar la metodología empleada para el cálculo o la descripción de las variables de estudio y su dispersión, y sin aclarar si se emplearon los parámetros de la variable con mayor coeficiente de variación. Otro aspecto relevante que se omite es el valor de la varianza, el cual se desconoce la mayoría de las veces. Existen alternativas que permiten superar este inconveniente, como usar los valores reportados en la literatura, hacer una prueba piloto o emplear un valor estimado por expertos en el tema (García-García et al., 2013).

Por otro lado, es fundamental decidir el diseño de recolección de datos, ya que cada uno tiene ventajas y desventajas. El diseño garantiza la validez estadística, el control y la precisión de las estimaciones y los costos de recolección de los datos (FAO, 2016a). Existen diversos métodos de muestreo para recopilar información agropecuaria, entre los cuales se encuentran el aleatorio simple,

el sistemático, el estratificado y el muestreo por conglomerados. Los métodos que realizan agrupamientos de individuos son los más empleados cuando no existen marcos de lista, de área o múltiples que permitan identificar a los productores (FAO, 2018). Las investigaciones sobre el sector agropecuario en Colombia suelen llevarse a cabo con estas carencias. Básicamente deben enfrentarse a las mismas dificultades para acceder a los marcos de lista y los datos históricos. Se desconocen los criterios empleados para el levantamiento de la información y falta representatividad estadística en los estudios previos (Perfetti et al., 2012).

Durante la ejecución del proyecto "Caracterización de oferta y demanda de semilla de calidad de las especies de interés", desarrollado por AGROSAVIA entre 2015 y 2017, se identificaron las mismas limitantes descritas por Perfetti et al. (2012). Fue necesario, por tanto, buscar alternativas que permitieran acceder a un marco muestral de referencia en el que los datos estuvieran disponibles y abiertos y a los que fuera posible acceder mediante los diferentes convenios que ha suscrito la corporación con otras entidades de investigación relacionadas con el sector agropecuario.

Después de varias consultas, se identificó que los datos recolectados y presentados en el 3er CNA eran la mejor opción. Cualquier investigador puebe emplear estos datos como muestra piloto, debido principalmente a que la información se presenta en el nivel de las UPA y debido también a que los datos se colectaron en todo el territorio nacional y la organización de los registros permite identificar las actividades agropecuarias en los niveles de vereda, municipio y departamento. El acceso a la información, así como su disponibilidad en el tiempo, se ratificó durante la ejecución del proyecto, y el censo se empleó para determinar el tamaño de muestra de las tres especies de interés: papa, cacao y plátano.

La implementación de este tipo de metodologías permite a los diferentes actores tener una base de información confiable, para conocer el comportamiento de las poblaciones y apoyar la toma de decisiones bajo diferentes escenarios. Lo anterior se refleja en un conocimiento de las comunidades y los sistemas productivos y posibilita generar estrategias para adoptar tecnologías como semillas de calidad y nuevos materiales y técnicas de manejo en la producción, que apoyen el desarrollo económico de los productores agropecuarios en el corto, mediano y largo plazo.

La metodología permite definir tamaños de muestra para una determinada especie de interés en una zona de estudio. Además, sirve de insumo para que cada entidad pueda realizar planes de trabajo dirigidos a recolectar información primaria que permita dar respuesta a las problemáticas del sector agropecuario que se derivan de la deficiencia de información estadística confiable. El proyecto contempló varias etapas, entre las cuales se encuentran la priorización de zonas de estudio, el cálculo del tamaño de la muestra y la selección del número de UPA. A continuación, se detalla cada una de las etapas.

Criterios para la priorización de zonas de estudio

Las zonas productoras se seleccionaron y priorizaron empleando criterios cuantitativos y cualitativos. Algunos de tales criterios fueron área total sembrada del cultivo de interés, presencia del eslabón primario de la cadena en la zona y variedades sembradas. También se tuvo en cuenta la disponibilidad de información para determinar zonas agroecológicas homogéneas situadas en la zona de influencia de AGROSAVIA. Se seleccionaron los municipios de Une y Villapinzón (Cundinamarca) para el sistema productivo de papa, San Pablo de Borbur (Boyacá) para cacao y Fuentedeoro (Meta) para plátano. En la tabla 2 se resumen los criterios más importantes empleados en cada sitio.

Tabla 2. Criterios de priorización de zonas para validar la metodología

Sistema productivo	Municipio	Criterios de priorización
Papa	Villapinzón, Cundinamarca	Es el segundo municipio de Cundinamarca en área sembrada (6.549 ha).*
		Existen diferentes tipos de productores, semilleristas y viveristas (con y sin registro del Instituto Colombiano Agropecuario [ICA]), así como centros de acopio de la industria, comercializadores regionales y centrales de abastos.
		AGROSAVIA trabaja en la zona en temas relacionados con semilla; algunas organizaciones de productores están inscritas en el Plan Semilla.

(Continúa)

(Continuación tabla 2)

Sistema productivo	Municipio	Criterios de priorización
Papa	Une, Cundinamarca	Es el quinto municipio de Cundinamarca en área sembrada con papa (3.174 ha).*
		Cuenta con diferentes tamaños de productores; tradicionalmente ha sido una zona proveedora de semilla sin registro ICA.
		El nivel de intervención de AGROSAVIA ha sido bajo.
		Fue una zona con graves problemas de orden público.
Cacao	San Pablo de Borbur, Boyacá	El occidente de Boyacá tuvo en promedio 3.937 ha sembradas de cacao, con una participación de 2 % en el área sembrada nacional.*
		Es un área nueva en la producción de cacao, con alto potencial de crecimiento.
		Tiene un área sembrada de 520 ha y es el cuarto productor de cacao en Boyacá.
		Hay productores de diferentes tamaños, proveedores de material de siembra y propagación con y sin registro ICA.
		AGROSAVIA trabaja en la zona desde hace varios años en temas de semilla; existen organizaciones inscritas en el Plan Semilla.
Plátano	Fuentedeoro, Meta	Es el primer municipio productor de plátano (5.566 ha) en el departamento del Meta.*
		El municipio forma parte de un gran corredor de producción de plátano.
		Existen productores de diferentes tamaños y tecnologías de producción.
		AGROSAVIA trabaja en la zona desde hace varios años en temas de semilla; algunas organizaciones de productores están inscritas en el Plan Semilla.

* Datos tomados de *Evaluaciones Agropecuarias Municipales (EVA), 2007-2014* (Red de información y comunicación del sector Agropecuario Colombiano [Agronet], 2014).

Fuente: Elaboración propia

Cálculo para determinar el tamaño de muestra y estimar el promedio de la demanda de semilla por unidad de área

Una vez definidas las zonas, se procedió a determinar los parámetros necesarios para definir el tamaño de la muestra. En el sitio Archivo Nacional de Datos del DANE (http://andacna.dane.gov.co/index.php/home), se descargaron los registros del archivo S06A(Cultivos) del 3er CNA y el Diccionario de datos. Dado que cada productor puede estar más de una vez en la base de datos porque los registros por filas hacen referencia a cada uno de los lotes presentes en el predio, se sumaron todos los registros por encuesta de las variables de área sembrada y área cosechada y se calculó el promedio de los rendimientos por cultivo. De este modo se creó una nueva tabla de datos para las variables calculadas Area_Sembrada_Sum, Area_Cosechada_Sum y Rendimiento_Promedio. Los datos se prepararon tal como se describió en los capítulos anteriores. Se filtró la base por identificadores de la encuesta (municipio y cultivo), empleando los códigos que se presentan en la figura 5.

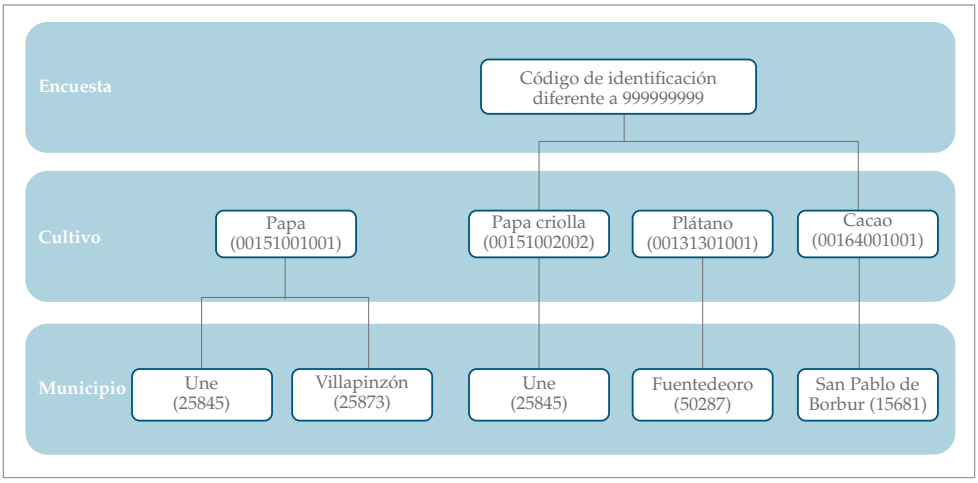

Figura 5. Códigos empleados en los filtros, en la preparación y procesamiento del archivo S06A(Cultivos).

Fuente: Elaboración propia

Con el fin de establecer un área mínima de cultivo que permitiera considerar si un productor demanda semilla para el establecimiento del cultivo, el equipo de trabajo se reunió con expertos de cada uno de los tres sistemas productivos. Con ellos se definieron los valores mínimos que se emplearon como valores de referencia (tabla 3).

Tabla 3. Área mínima de referencia

Cultivo	Área sembrada mínima (ha)
Papa	0,05
Cacao	0,5
Plátano	0,5

Fuente: Elaboración propia

El archivo se filtró eliminado los registros con áreas inferiores al valor de referencia. También se examinaron los datos para determinar si alguna UPA se ubicada dentro de comunidades indígenas, negras, afrocolombianas, raizales y palenqueras, y se encontró que ninguna formaba parte de estos territorios especiales. En total, para las tres especies en las cuatro zonas priorizadas, se obtuvieron 2.107 registros únicos (tabla 4). A continuación, se crearon estratos por medio de la variable Area_Sembrada_Sum, se tomó de forma aleatoria aproximadamente el 10 % de los registros por estrato y sistema productivo, y estos registros se emplearon como sustituto de la muestra piloto de los productores.

Tabla 4. Población por especie y municipio

Departamento	Municipio	Especie	Número de Unidades Productoras Agropecuarias (UPA)
Boyacá	San Pablo de Borbur	Cacao	407
Meta	Fuentedeoro	Plátano	544
Cundinamarca	Une	Papa	103
	Villapinzón	Papa	1.053

Fuente: Elaboración propia con base en DANE (2016b)

Con esa porción de registros se calcularon la varianza, la desviación estándar y la media. Estos valores se emplearon en la fórmula que determina el tamaño de muestra por afijación óptima en el método de muestreo estratificado, con un error relativo de estimación del 5 % y un nivel de confianza del 90 %. Todos los procedimientos descritos se hicieron con el *software* estadístico SAS 9.4. En la figura 6 se presenta un esquema general de los pasos seguidos en la preparación y procesamiento de la base de datos.

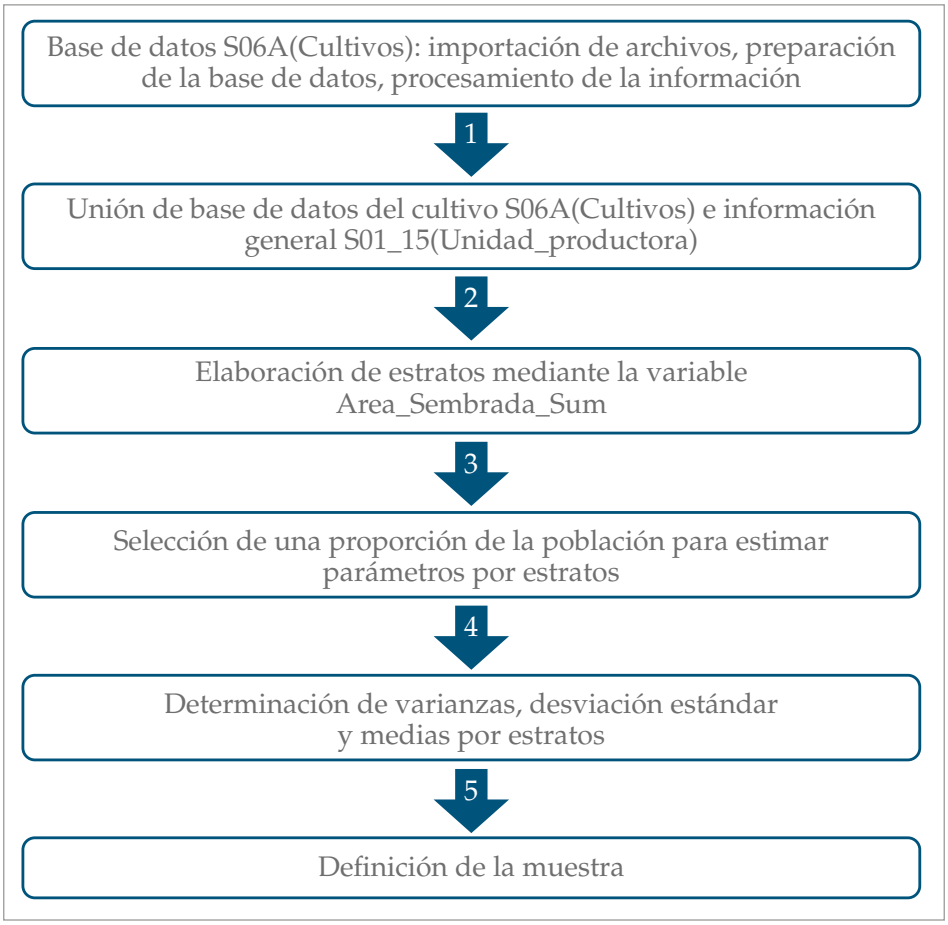

Figura 6. Proceso para determinar los parámetros para el cálculo del tamaño de la muestra.
Fuente: Elaboración propia

Los tamaños óptimos de la muestra por sistema productivo fueron los siguientes: 134 UPA para papa en los municipios de Une (20) y Villapinzón (114); 45 UPA para cacao en el municipio de San Pablo de Borbur, y 55 UPA para plátano en el municipio de Fuentedeoro (tabla 5).

Tabla 5. Distribución de la muestra a nivel UPA para tres sistemas productivos

Sistema productivo	Estratos	Tamaño de muestra	Número de veredas
Papa (Une y Villapinzón)	0,05-1,0	10	7
	1,01-2,0	8	6
	2,01-5,0	50	15
	5,01-10,0	31	11
	10,01-50,0	26	16
	50,01 y más	9	7
Cacao (San Pablo de Borbur)	0,5-1,0	9	9
	1,01-2,0	17	13
	2,01-3,0	7	5
	3,01 y más	12	10
Plátano (Fuentedeoro)	0,5-0,99	2	2
	1-1,99	1	1
	2-4,99	8	8
	5-9,99	12	9
	10-19,99	8	8
	20-39,99	11	6
	40-59,99	2	2
	60-79,99	6	6
	80,00 y más	5	4

Fuente: Elaboración propia con base en DANE (2016b)

En cada municipio se realizó una consulta con la respectiva secretaría de agricultura, la Umata, y asociaciones y cooperativas, para elaborar los marcos de lista que requería la selección de la muestra. Las listas contenían información sobre nombres de productores, datos de contacto, ubicación geográfica y tamaño de predios. Dados los aspectos socioculturales de los productores y la dinámica de producción de cada sistema, y ante posibles

rotaciones de cultivo o cambios en los lotes de producción, cada listado se verificó para determinar la coherencia de la información reportada por las diferentes entidades y la presencia de las unidades productivas en los sitios de interés, así como para identificar de manera definitiva los individuos muestrales a quienes se aplicaría la encuesta.

Estimación de número de UPA y área sembrada en otros cultivos

Con la metodología descrita en este documento, se estimó el número de unidades productoras y la cantidad de área sembrada y cosechada de las especies algodón, chontaduro, fríjol, lima, limón, maíz amarillo, maíz blanco, mango, naranja y quinua. En estas especies, además de que se usaron para determinar sitios de cultivo en 2014 y número de UPA, permitieron a los colaboradores de AGROSAVIA desarrollar muestreos de material vegetal y suelos, así como caracterizaciones de productores y sistemas productivos. En la tabla 6 se presentan las especies y el uso que se le dio a la información procesada en el 3er CNA.

Tabla 6. Aplicación de la metodología en otras especies

Especie	Usos
Algodón	Número de UPA; caracterización de productores
Chontaduro	Número de UPA
Fríjol	Caracterización de productores
Lima	Número de UPA
Limón	Número de UPA para muestreo de material vegetal y suelos
Maíz amarillo	Caracterización de productores
Maíz blanco	Caracterización de productores
Mango	Número de UPA para muestreo de material vegetal y suelos; caracterización de productores
Naranja	Número de UPA
Quinua	Número de UPA

Fuente: Elaboración propia

La metodología descrita en el presente documento y el uso de preguntas a manera de filtros permitieron calcular los tamaños de muestra y estimar la media y el total poblacional de las variables de interés de los diferentes estudios. Con la variable de cultivos (Pregunta P_S6P46) se identificó la distribución de las especies cultivadas a nivel de vereda en los territorios de interés, se determinó si estaban presentes en territorios especiales y se hizo una aproximación a la caracterización de los sistemas productivos. La metodología también se adaptó a otros archivos del 3er CNA disponibles en el sitio Datos abiertos, para identificar características personales y de vivienda de los productores agropecuarios del país.

Sobre el trabajo de campo

La base de datos del 3er CNA permitió definir la muestra en los niveles de municipio y vereda y delimitar el tamaño con las UPA, las cuales se usaron como variable de estratificación. Hubo, sin embargo, un alto nivel de incertidumbre dada la dinámica de la producción agrícola, ya que en ella se dan movimientos de productores de una zona a otra, reducciones o aumentos de áreas de siembra, abandono de los productores de la actividad y aparición o reaparición de otros productores. En consecuencia, esta muestra con distribución de estratos por vereda debió ajustarse a la realidad actual del cultivo en cada municipio, y para ello fue necesario, tal como lo formula Díaz de Rada Igúzquiza (2008), contar con una fuente que entregara información detallada del dominio objeto de estudio. Esto representaba un desafío en aquellos casos en los que las muestras se habían construido con reportes de población de los censos, dado que estos últimos no registran en detalle la estructura de territorios con alta movilidad debido a la baja frecuencia con que se actualizan sus datos.

Para subsanar lo anterior, se acudió a las Umata, las cuales son el principal actor de la estadística y la organización productiva de los municipios, así como a organizaciones de productores y otros actores clave con información sobre productores, unidades productoras, cultivos, veredas y ubicación geográfica de los predios. La información suministrada por estas entidades permitió verificar si las UPA de cada vereda cumplían con el criterio de área sembrada (estratos), y con ello se construyeron los marcos de lista de los productores que conformaban las unidades estadísticas (los individuos de la población

estudiada). Así se pudo reconfigurar y ajustar la muestra por estrato a nivel de vereda en cada municipio, esto bajo el supuesto de que este conocimiento previo mejoraría la efectividad en la selección de los elementos muestrales. Al comprobar previamente cuáles individuos cumplían los requisitos definidos, se pudieron identificar los candidatos que formarían parte de la muestra.

El trabajo de campo inició con la planeación de rutas para acceder a los elementos muestrales definidos en cada estrato. En este proceso se detallaron los traslados y movimientos de los encuestadores, con el fin de optimizar el tiempo de desplazamiento, disminuir la fatiga física, reducir gastos, cubrir la totalidad de zonas de cada estrato y facilitar la localización tanto de los encuestados como de los encuestadores.

Con los datos de contacto, se estableció comunicación con los productores de las veredas que conformaban la muestra para concertar una cita en el predio y aplicar la encuesta. Quienes rechazaron la invitación a participar en el estudio fueron reemplazados por otros de la misma vereda que tenían cultivos con las mismas características de área. Cuando no fue posible encontrar sustituto, se buscó su reemplazo en otra vereda cercana. En estos últimos casos, se buscó el acompañamiento de quien los había referenciado, para tener mayor posibilidad de ser atendidos, o se visitaron los predios sin previo aviso. En estas visitas se presentó el proyecto y se solicitó participación en el estudio luego de hacer preguntas filtro para asegurar que los nuevos candidatos cumplieran con el perfil. De esta manera se completaron las rutas y se cumplió con el número total de encuestados que demandaba cada estrado.

Durante la ejecución del estudio, se evaluó diariamente el cumplimiento de la meta de encuestas del estrato de cada ruta. Las rutas debían repetirse cuando los productores reagendaban la cita o cuando, durante el recorrido, no se obtenía la respuesta esperada. En este último caso, se debió consultar de nuevo a los actores clave para indagar por nuevos productores que tuvieran el perfil, concertar una cita con ellos y seguir con el proceso hasta completar el tamaño del respectivo estrato. Si, pese al esfuerzo, no se lograba completar el número de encuestas esperadas para cada estrato, se discutía la pertinencía de encuestar productores con cultivos correspondientes al perfil de otros estratos. Se aprovechaba así la presencia en el territorio y se consideraba la pérdida de información como parte del proceso de aplicación de la encuesta.

20642