

El conjunto de estrategias y procedimientos que permiten seleccionar una muestra con características deseables a partir de una población objetivo se denomina *diseño muestral*. Este diseño permite especificar la forma como se elegirán los elementos sobre los cuales se llevará a cabo el estudio.

Tipo de muestreo

Se determinó que el mejor método de muestreo es el aleatorio estratificado porque dentro de la población objetivo (productores con un cultivo determinado entre los años 2013 y 2014), existe una alta heterogeneidad en términos de área sembrada y cosechada. El conjunto de elementos se dividió en subpoblaciones o estratos con el objetivo de reducir la variabilidad y alcanzar mayor precisión (Kish, 1965). Este tipo de muestreo es potente y flexible, por lo cual es de amplio uso en la práctica (Särndal et al., 2003).

La definición del número de elementos de un estrato que formará parte del tamaño total de la muestra depende de varios factores:

1. Tamaño relativo del estrato en la población (afijación proporcional).
2. Variabilidad de la variable de interés dentro de cada estrato (afijación óptima).
3. Costos del muestreo de los elementos dentro de cada estrato (afijación óptima de costo).

Las siguientes son las principales razones para emplear el muestreo aleatorio estratificado en lugar del tradicional aleatorio simple (Scheaffer et al., 1987; Särndal et al., 2003):

1. La estratificación produce un error de estimación más pequeño que el generado por una muestra irrestricta aleatoria del mismo tamaño. El resultado es particularmente cierto si las mediciones dentro de los estratos son homogéneas.
2. El costo de medir u observar un elemento puede reducirse mediante la conformación de estratos de los elementos de la población
3. Si el propósito es obtener datos de precisión conocida, es necesario considerar cada estrato como una población (Cochran, 1977).

En el momento de realizar la estratificación deben tenerse en cuenta los principios que se emplean para decidir si es necesario crear estratos dentro de la población y cómo se delimitan. Los estratos pueden definirse considerando una o más variables y estas deben tener una alta correlación con la variable de interés (objetivo del estudio), con el fin de generar estratos lo más homogéneos posibles. La construcción de los estratos debe incluir algunos requisitos, como precisión de la estimación para los diferentes dominios, y debe contemplar variables de estudios, costos y restricciones administrativas para la recolección de la información (Särndal et al., 2003).

En la elección de los elementos que conformarán la muestra, es fundamental que los estratos se identifiquen claramente en el marco muestral para que cada elemento de la población pertenezca a un solo estrato. También es importante que cada estrato sea independiente, ya que así se pueden realizar inferencias de cada una de las poblaciones que originaron la muestra (Cochran, 1977).

Estimación de la media y el total poblacional

El estimador de la media poblacional utilizado en el muestreo estratificado es la media estratificada. Este estimador tiene propiedades deseables (ser insesgado y tener mínima varianza) y se expresa mediante la ecuación 2:

$$\bar{Y}_{st} = \frac{\sum_{h=1}^L N_h \bar{Y}_h}{N} = \sum_{h=1}^L W_h \bar{Y}_h \quad \text{Ecuación 2,}$$

donde \bar{Y}_{st} es la media estratificada, N_h el número de elementos de la población en el h-ésimo estrato, \bar{Y}_h la media muestral en el h-ésimo estrato, N el número de elementos de la población, y W_h la ponderación del h-ésimo estrato.

En un muestreo estratificado, la varianza del estimador de la media poblacional (\bar{Y}_{st}) se formula como se indica en la ecuación 3:

$$V(\bar{Y}_{st}) = \frac{1}{N^2} \sum_{h=1}^L N_h (N_h - n_h) \frac{S_h^2}{n_h} \quad \text{Ecuación 3,}$$

donde $V(\bar{Y}_{st})$ es la varianza de la media estratificada, N el número de elementos de la población, N_h el número de elementos de la población en el h-ésimo estrato, n_h número de elementos de la muestra en el h-ésimo estrato, y S_h^2 la cuasivarianza verdadera en el h-ésimo estrato.

La estimación del total poblacional Y se expresa mediante la ecuación 4:

$$\hat{Y}_{st} = N\bar{Y}_{st} \quad \text{Ecuación 4,}$$

donde \hat{Y}_{st} es la estimación del total poblacional, N el número de elementos de la población, y \bar{Y}_{st} la media estratificada.

Y la varianza del total poblacional se formula con la ecuación 5:

$$V(\hat{Y}_{st}) = \sum_{h=1}^L N_h \frac{S_h^2}{n_h} (N_h - n_h) \quad \text{Ecuación 5,}$$

donde $V(\hat{Y}_{st})$ es la varianza de la media estratificada, N_h el número de elementos de la población en el h-ésimo estrato, S_h^2 la cuasivarianza verdadera en el h-ésimo estrato, y n_h el número de elementos de la muestra en el h-ésimo estrato.

Definición del tamaño de la muestra

El tamaño de la muestra para la media y el total equivale al número de elementos que se deben muestrear de la población, esto con el fin de estimar un parámetro con determinados niveles de confianza y error de estimación.

Existen varios métodos o tipos de afijación para repartir los elementos dentro de los estratos, los cuales se explican a continuación (figura 3) (Cochran, 1977; Särndal et al., 2003; Scheaffer et al., 1987; Kish, 1965):

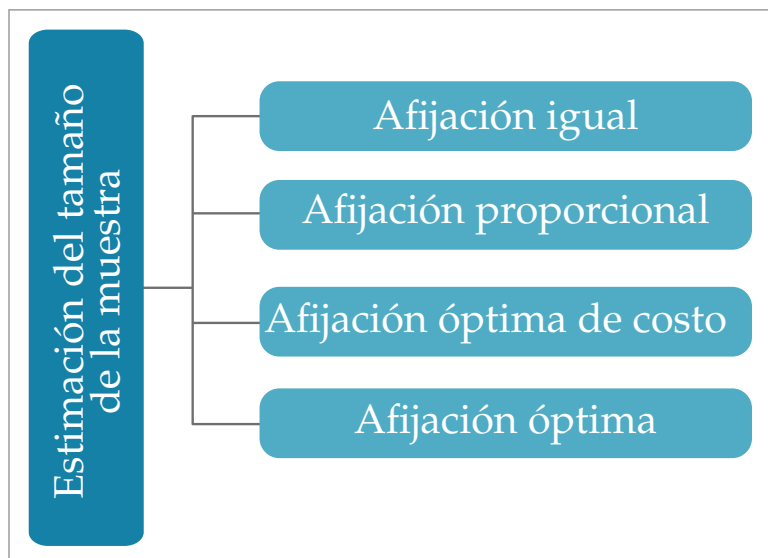


Figura 3. Tipos de afijación para calcular el tamaño de la muestra.

Fuente: Elaboración propia

Afijación igual

El tamaño de la muestra se divide entre el número total de estratos para que se pueda seleccionar el mismo número de elementos en cada uno, independientemente del peso del tamaño del estrato en la población. De todos los métodos para calcular el tamaño de la muestra, este es el de menor uso.

Afijación proporcional

Consiste en asignar a cada estrato un número de unidades muestrales proporcional a su tamaño. Las n unidades de la muestra se distribuyen de manera proporcional a los tamaños de los estratos, los cuales se expresan en número de unidades (ecuación 6).

$$n = \frac{\sum_{h=1}^L W_h s_h^2}{\frac{e^2}{z_{\alpha/2}^2} + \frac{1}{N} \sum_{h=1}^L (W_h s_h^2)} \quad \text{Ecuación 6,}$$

donde n es el número de elementos de la muestra, W_h la ponderación del h -ésimo estrato, s_h^2 la cuasivarianza verdadera en el h -ésimo estrato, e el error máximo admisible, $z_{\alpha/2}$ el valor tabulado del coeficiente de confianza, y N el número de elementos de la población.

El tamaño de muestra por estrato se expresa por medio de la ecuación 7:

$$n_h = n W_h \quad \text{Ecuación 7,}$$

donde n_h es el número de elementos de la muestra en el h -ésimo estrato, n el número de elementos de la muestra, y W_h la ponderación del h -ésimo estrato.

Afijación óptima de costo

La afijación óptima de costo consiste en determinar los valores de n_h (número de unidades que se extraen del h -ésimo estrato para la muestra), de forma que para un coste fijo C , la varianza de los estimadores sea mínima. En este tipo de afijación, se tiene en cuenta la variabilidad dentro de cada estrato, así como el costo de recolectar la información. La función de costo es una función lineal (ecuación 8):

$$\text{Costo (C)} = C_0 + \sum_{h=1}^L C_h n_h \quad \text{Ecuación 8,}$$

donde C es la función de costo, C_0 el costo fijo independiente del número de elementos por seleccionar, C_h el costo de muestrear un elemento en el h -ésimo estrato, y n_h el número de elementos de la muestra en el h -ésimo estrato.

El tamaño de la muestra se determina mediante la ecuación 9:

$$n = \frac{(\sum_{h=1}^L W_h s_h \sqrt{C_h}) \sum_{h=1}^L (W_h s_h / \sqrt{C_h})}{\frac{e^2}{z_{\alpha/2}^2} + \frac{1}{N} \sum_{h=1}^L (W_h s_h^2)} \quad \text{Ecuación 9,}$$

donde n es el número de elementos de la muestra, W_h la ponderación del h -ésimo estrato, s_h la desviación estándar en el h -ésimo estrato, C_h el costo de muestrear un elemento en el h -ésimo estrato, e el error máximo admisible, $Z_{\alpha/2}^2$ el valor tabulado del coeficiente de confianza, y N el número de elementos de la población.

El tamaño de muestra por estrato se formula tal como se indica en la ecuación 10:

$$n_h = n \frac{W_h s_h / \sqrt{C_h}}{s_h^2 \sum_{h=1}^L (W_h s_h / \sqrt{C_h})} \quad \text{Ecuación 10,}$$

donde n_h es el número de elementos de la muestra en el h -ésimo estrato, n el número de elementos de la muestra, W_h la ponderación del h -ésimo estrato, s_h la desviación estándar en el h -ésimo estrato, y C_h el costo de muestrear un elemento en el h -ésimo estrato.

El costo afecta de forma inversa el tamaño de la muestra. A mayor costo por estrato, menor el tamaño de la muestra, mientras que la variabilidad afecta de manera directa el número de muestras que deben ser tomadas (a mayor varianza, mayor número de muestras).

Afijación óptima

También llamada *afijación de Neyman*, consiste en determinar los valores del número de unidades extraído del estrato, de forma tal que, con un tamaño de muestra fijo igual a n , la varianza de los estimadores sea mínima. Si el costo por unidad de muestreo es el mismo en cada estrato y si se fija el valor de la varianza, la fórmula para determinar el tamaño de la muestra sería la siguiente (ecuación 11):

$$n = \frac{(\sum_{h=1}^L W_h s_h)^2}{\frac{e^2}{Z_{\alpha/2}^2} + \frac{1}{N} \sum_{h=1}^L (W_h s_h^2)} \quad \text{Ecuación 11,}$$

donde n es el número de elementos de la muestra, W_h la ponderación del h -ésimo estrato, s_h la desviación estándar en el h -ésimo estrato, s_h^2 el error máximo admisible, $z_{\alpha/2}$ el valor tabulado del coeficiente de confianza, y N el número de elementos de la población.

El tamaño de muestra por estrato se expresa mediante la siguiente fórmula (ecuación 12):

$$n_h = n \frac{W_h s_h^2}{\sum_{h=1}^L W_h s_h^2} \quad \text{Ecuación 12,}$$

donde n_h es el número de elementos de la muestra en el h -ésimo estrato, W_h la ponderación del h -ésimo estrato, n el número de elementos de la muestra, y s_h la desviación estándar en el h -ésimo estrato.

El tamaño de la muestra aumenta si se cumple alguno de los siguientes criterios:

1. El estrato es más grande.
2. El estrato es más variable (heterogéneo).
3. El muestreo tiene menor costo en el estrato.

Algunas consideraciones sobre la planeación de estudios por medio de aplicación de encuestas

Una encuesta es una técnica que permite obtener información de forma objetiva, en la que los datos recogidos pueden ser analizados empleando métodos cuantitativos y los resultados pueden ser extrapolables a la población. Entre sus múltiples ventajas, está la posibilidad de estandarizar las preguntas que se formulan a los elementos de la población, con lo cual se hacen fácilmente administrables, se simplifica la recogida y el tratamiento de los datos y se puede obtener información que a simple vista no está disponible (Abascal Fernández & Grande Esteban, 2005). Planear y hacer una encuesta con rigor estadístico exige surtir una serie de etapas (figura 4), algunas de las cuales se exponen a continuación.

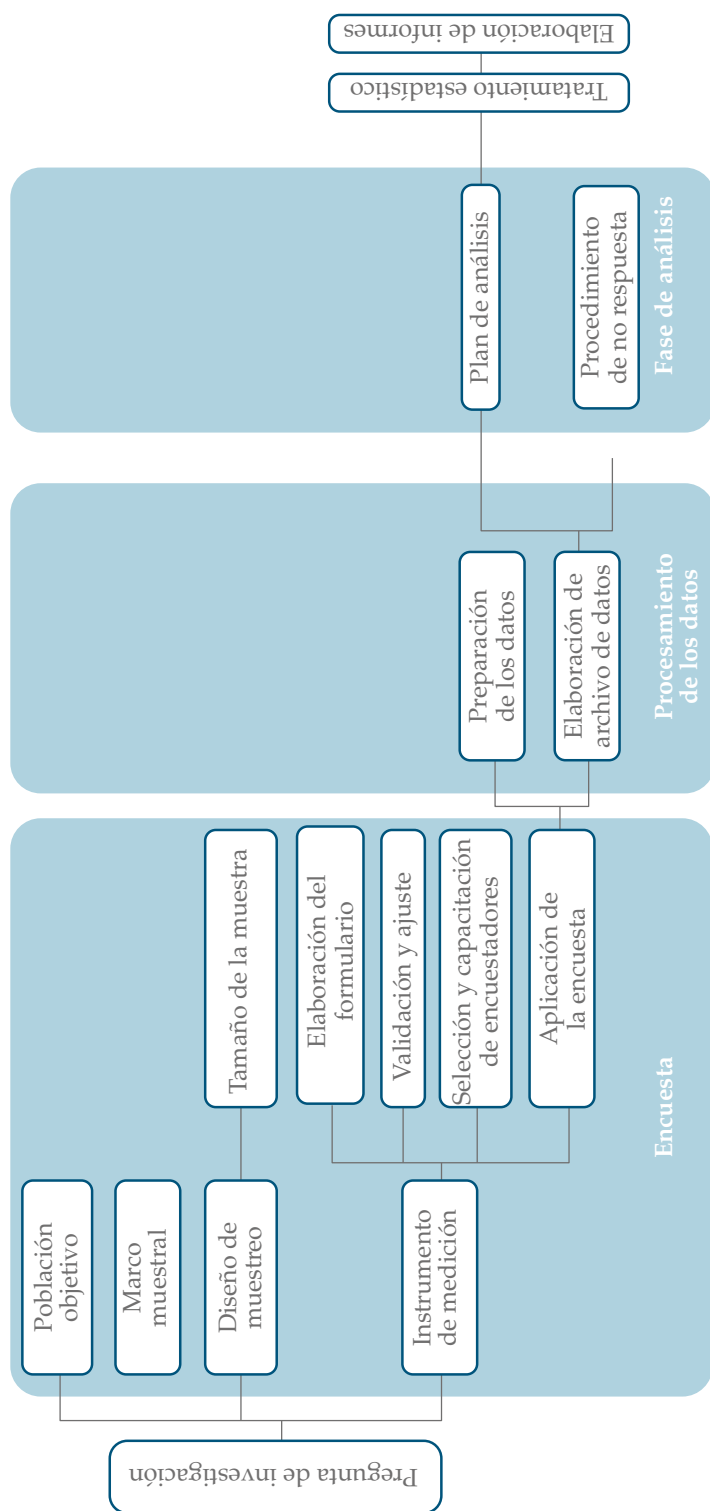


Figura 4. Proceso para el diseño, ejecución y análisis de datos de un estudio mediante encuesta.

Fuente: Elaboración propia

1. *Identificación del problema:* el problema surge de la pregunta de investigación. En esta etapa se deben definir claramente los objetivos y revisar estudios similares. En casos en los que se tenga poco conocimiento sobre el problema o la población (o sobre ambos), es necesario hacer una revisión de los temas relacionados y recurrir a otras técnicas para recabar la información que no se puede obtener por otros medios.
2. *Definición de la población objetivo:* se considera la UPA como predio, como parte de un predio o como conjunto de predios (sin importar su tamaño, régimen de tenencia y ubicación), y se considera también la actividad agropecuaria de interés que se desarrolla en la unidad, bajo la dirección de un productor agropecuario. Se establecen como segundo objeto de estudio las personas naturales o jurídicas, residentes habituales o no en el sector rural del territorio nacional, bajo cuya responsabilidad se toman las principales decisiones sobre la utilización de los recursos disponibles, el uso de los factores de producción (trabajo, capital, maquinaria, etc.) y el control administrativo de la UPA.
3. *Levantamiento del marco muestral:* enumeración de todos elementos del universo o población para hacer una selección aleatoria dentro de ellos. Para el levantamiento de los listados se recomienda visitar asociaciones, gremios, cooperativas, secretarías de agricultura municipales, Unidades Municipales de Asistencia Técnica Agropecuaria (Umata), entre otros. Es importante recalcar que no existe el marco muestral perfecto y que la actualización de dicho marco es fundamental en el proceso.
4. *Diseño de muestreo:* aunque existen diferentes tipos de muestreo, todos se basan en el principio de aleatoriedad. Esto significa que todos los elementos del universo deben tener oportunidad de ser elegidos para componer la muestra. Si no se eligieran aleatoriamente, se correría el grave riesgo de que los resultados obtenidos fueran representativos, no de toda la población, sino de un solo sector, algo a lo cual se le llama *sesgo*. Para poder hacer inferencias válidas de lo que ocurre, es necesario que la muestra sea representativa del universo, y esto se logra con la aleatoriedad y con un tamaño adecuado de la muestra. El tamaño de la muestra se define según la variabilidad del fenómeno a estudiar y según el parámetro de interés, el nivel de confianza fijado y el error admisible. Un error muy común es pensar que la muestra, para que sea representativa, debe tener un tamaño directamente proporcional al tamaño del universo, es decir, que a mayor tamaño del universo, mayor debería ser el tamaño de la muestra.

5. *Instrumentos de medición:* el instrumento de recogida de información es quizás el elemento más importante de una investigación basada en la aplicación de encuestas. La calidad con la cual se construya el instrumento será determinante para recoger con exactitud la información que se desea obtener (Martínez Medrano, 2004). Las siguientes son algunas recomendaciones para el diseño de cuestionarios según la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) (2003):
 - a. El tipo y número de preguntas dependerá de las variables para analizar. No se deben incorporar preguntas que no tengan relación directa con el objeto de estudio. Es importante que las preguntas sean breves, usen vocabulario sencillo y obedezcan a una secuencia acorde con la lógica del entrevistado.
 - b. En lo posible, las preguntas deben ser de respuesta cerrada, dado que así su procesamiento es más sencillo. Las preguntas abiertas son muy útiles para detectar opiniones, percepciones y preferencias.
 - c. Para formular el cuestionario definitivo, es necesario poner a prueba uno preliminar con una pequeña muestra de la población destino. Así se puede conocer la variabilidad de respuestas posibles o situaciones no consideradas originalmente.
 - d. Cuando se trabaje con muestras muy grandes y con preguntas complejas, es conveniente que estas últimas tengan instrucciones precisas de aplicación y que dichas instrucciones se resalten en el cuestionario.
6. *Selección y capacitación de los encuestadores:* los encuestadores son las personas encargadas de aplicar el instrumento y recolectar los datos, de ahí la importancia de seleccionarlos y capacitarlos adecuadamente. Estas personas deben conocer los objetivos del estudio, el diseño de muestreo, la técnica de obtención de datos y su procesamiento.
7. *Validación y ajuste del formulario:* esta etapa se denomina en algunos casos *pretest* o *prueba piloto de la encuesta* y pretende averiguar si el lenguaje es apropiado y si las preguntas del formulario son claras y guardan un orden coherente. En esta etapa también se estiman el tiempo y el costo que requiere la aplicación del instrumento. Los encuestados deben ser personas con características similares a los elementos de la población objetivo.
8. *Preparación de los datos colectados:* una vez recogidos los datos, es necesario revisar los cuestionarios (edición) e inspeccionar, corregir y en ocasiones precodificar las respuestas de acuerdo con un conjunto de

reglas fijas. Tras la inspección y corrección de las respuestas, se elabora el archivo de datos (Díaz de Rada Igúzquiza, 2009). En algunos casos, la codificación se puede hacer en esta etapa, aunque por lo general se debe hacer en el momento de recoger la información. En cualquier caso, es fundamental contar con un diccionario de datos para el análisis y el posterior desarrollo del informe.

El archivo de datos se elabora siguiendo unas pautas sencillas. Los nombres de las variables se escriben en la primera fila y el identificador de la encuesta en la primera columna, seguido de las variables de identificación geográfica y las variables de interés particular del estudio. El archivo, en lo posible, debe estar en un formato fácilmente recuperable (extensiones de archivo .xls, .xlsx, .txt., .csv) y en ningún caso se deben utilizar archivos propios de programas específicos que no permitan consultas *a posteriori*.

Las siguientes se consideran malas prácticas en la generación de bases de datos:

1. Dejar celdas vacías en la primera fila.
 2. Consignar registros sin código de identificación.
 3. Combinar celdas.
 4. Incluir textos en las celdas de las variables donde solo deben ir números.
 5. En ausencia de un dato, diligenciar el campo con el valor cero (0).
9. *Análisis de los datos y elaboración de informes*: el análisis involucra una serie de cálculos basados en los datos obtenidos. La idea es que de este análisis salgan reportes de resultados presentados en forma sintética y en diferentes escalas (unidad muestral, estratos y universo). Esta etapa se puede llevar a cabo mediante análisis univariado, análisis bivariado y, en caso de que el instrumento lo permita, también se pueden aplicar técnicas de análisis multivariado.

Una vez analizados los datos, se generan los reportes de resultados, que no son más que un resumen de los datos de las unidades muestrales. Lo habitual es elaborar dos informes, uno relacionado con el diseño de la muestra y los resultados de las encuestas, y otro compuesto de un informe técnico que plasme las conclusiones del estudio.

