CADA HECHO DE TU VIDA *Cuenta*

INEC Buenas cifras, mejores vidas

Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC)

Metodología del Diseño Muestral

2022

Roberto Castillo
DIRECTOR EJECUTIVO

Magaly Paredes
SUBDIRECTOR/A GENERAL

Markus Nabernegg
COORDINADOR GENERAL TÉCNICO DE PRODUCCIÓN ESTADÍSTICA

Armando Salazar
DIRECTOR DE ESTADISTICAS AGROPECUARIAS Y AMBIENTALES

Equipo Técnico:
Armando Salazar , Maritza Cuichán , Diego Orbe, Julio Márquez,.

Instituto Nacional de Estadística y Censos
Juan Larrea N15-36 y José Riofrío. Casilla postal 135 C
Telf.: (02) 2555-701 / 2529-858

Citar como:

INEC (2022). Metodología de Diseño Muestral de la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC). Instituto Nacional de Estadística y Censos, Quito-Ecuador.

Dirección de Estadísticas Agropecuarias y Ambientales

Unidad de Estadísticas Agropecuarias

Elaborado por:
Diego Orbe E

Revisado por:
Maritza Cuichán

Aprobado por:
Armando Salazar.

Tabla de contenidos

1. DISEÑO MUESTRAL

Estimación

Las características directamente observadas sobre el terreno, tales como la superficie de los cultivos y demás usos del suelo, la superficie regada con cada tipo de riego y la superficie labrada con cada técnica de laboreo, se estiman a partir de los datos recogidos en la muestra de segmentos.

Las variables relativas a la economía de las unidades de producción agropecuaria, incluyendo la ganadería, y las variables relativas al hogar, han de ser estimadas a partir de los datos recogidos mediante entrevistas a los agricultores y sus hogares.

Estimaciones basadas en la muestra de segmentos.

El total Y de la variable en estudio se expresa como suma de los totales, Y h, de cada estrato:

$$Y = \sum_{h=1}^{L} Y_h$$

Estimadores de dominio

La ESPAC es una encuesta nacional: se ha diseñado para obtener estimaciones de la producción agropecuaria a nivel nacional, con un grado de precisión aceptable (coeficiente de variación inferior al 15%). No obstante, es posible obtener también estimaciones (llamadas de dominio) para ámbitos administrativos más reducidos (cantones y provincias) y/o para ciertos subgrupos de la población de UPAs, tal como la agricultura familiar (cuántos son, cuál es su renta, cuáles son sus condiciones de vida...) a partir de la muestra maestra diseñada para la ESPAC.

Las estimaciones de dominio se basan solo en una parte de la muestra (la parte de la muestra total que cae dentro de los límites del dominio) y de ahí que su precisión sea inferior a la de las estimaciones obtenidas a nivel nacional. Si se requiere mejorar las estimaciones de dominio, es necesario complementar la muestra original con una muestra adicional seleccionada dentro del dominio en cuestión.

Estimadores de dominio basados en la muestra original

Se define la variable auxiliar:

$$y'_{hij} = \begin{cases} y_{hij} & si \quad i \in \quad al \quad dominio \quad D \\ 0, & en \quad otro \quad caso \end{cases}$$

Y se usa el estimador definido en [1]: $\hat{Y}_D = \sum_{h=1}^{L} \hat{Y}_{Dh}$, donde $\hat{Y}_{Dh} = \sum_{j=1}^{m_h} \hat{Y}_{Dhj}$ es el estimador

Del total de la variable de muestreo en el estrato h dentro del dominio Y_{Dh}, e $\hat{Y}_{Dhj}$,

Es el estimador del total de la variable en estudio en la j-ésima zona del estrato h-ésimo, dentro del dominio

$$\hat{Y}_{Dhj} = \frac{N_{hj}}{r_h} \sum_{i=1}^{r_h} y'_{hij}$$

Varianza.

La varianza del estimador del total de dominio viene dada por la siguiente expresión [2],

$$V(\hat{Y}_D) = \sum_{h=1}^{L} V(\hat{Y}_{Dh}), \text{ donde } V(\hat{Y}_{Dh}) = \sum_{j=1}^{m_h} V(\hat{Y}_{Dhj}) \text{ y}$$

$$V(\hat{Y}_{Dhj}) = N_{hj}^2 \left(1 - \frac{r_h}{N_{hj}}\right) \frac{1}{r_h} \frac{1}{N_{hj}-1} \sum_{i_s=1}^{N_{hj}} (y'_{hij} - \overline{Y'}_{hj})^2, \text{ donde } \overline{Y'}_{hj} = \frac{1}{N_{hj}} \sum_{i=1}^{N_{hj}} y'_{hij}$$

$$\text{y se estima por } \hat{V}(\hat{Y}_D) = \sum_{h=1}^{L} \hat{V}(\hat{Y}_{Dh}), \text{ donde } \hat{V}(\hat{Y}_{Dh}) = \sum_{j=1}^{m_h} \hat{V}(\hat{Y}_{Dhj}),$$

$$\text{y, } \hat{V}(\hat{Y}_{Dhj}) = N_{hj}^2 \left(1 - \frac{r_h}{N_{hj}}\right) \frac{1}{r_h} \frac{1}{r_h - 1} \sum_{i=1}^{r_h} (y'_{hij} - \overline{y'}_{hj})^2$$

$$\text{donde } \overline{y'}_{hj} = \frac{1}{r_h} \sum_{i_s=1}^{r_h} y'_{hij}$$

La varianza del estimador de la media es $V(\hat{\bar{Y}}_D) = \frac{1}{N_D^2} V(\hat{Y}_D)$

Y se estima por $\hat{V}(\hat{\bar{Y}}_D) = \frac{1}{N_D^2} \hat{V}(\hat{Y}_D)$ Los coeficientes de variación del total y de la media coinciden.

Estimadores de dominio basados en muestras múltiples

Si la precisión de las estimaciones obtenidas usando los estimadores no alcanza el nivel requerido, entonces es necesario complementar la muestra original con una muestra adicional seleccionada dentro del dominio en cuestión.

En este epígrafe se considera el problema de estimar el total de dominio de una variable de muestreo, Y_D, a partir de dos muestras: la inicial de la ESPAC, A, y una complementaria, B. Las muestras se seleccionan sucesivamente de sendos marcos {A, B}: primero se muestrea A y después B . Los tamaños de los estratos en ambos marcos son conocidos a priori.

$$\left\{N_h^{(A)}; h = 1, 2, \cdots, L\right\} \text{ con } \sum_{h=1}^{L} N_h^{(A)} = N_A \text{ y } \left\{N_h^{(B)}; h = 1, 2, \cdots, L\right\} \text{ con } \sum_{h=1}^{L} N_h^{(B)} = N_B$$

N_A es el número de unidades de muestreo en el marco A y N_B es el número de unidades del marco B (es el tamaño del dominio D). Generalmente, el marco B es la parte del marco A dentro de los límites del dominio D: por ejemplo, si el dominio D es un cantón o una provincia, N_B es el número de segmentos del marco A que están dentro del dominio D.

Pero los criterios de estratificación de B pueden ser distintos de los de A y, en ese caso, los estratos coincidirán:

$$\left\{h^{(A)}; h=1,2,\cdots,L\right\} \text{ y } \left\{h^{(B)}; h=1,2,\cdots,L\right\}$$

Por ejemplo, si el dominio D es el conjunto de UPAs que integran la agricultura familiar y el marco B es un registro de UPAs familiares, N_B es el número de UPAs incluidas en el marco B y los criterios de estratificación de B pueden ser distintos de los de A.

Se asume que B A $\subset$ **,** por lo que seleccionar una muestra aleatoria estratificada a partir del marco B es equivalente a seleccionar una muestra aleatoria estratificada de A, de modo que uno de los estratos contiene los elementos de la población no incluidos en el marco B y este estrato se muestrea a tasa nula [Bankier (1986)]. Sin pérdida de generalidad, supondremos que el estrato que contiene los elementos de la población no incluidos en el marco B es $(L+l)^{(B)}$

Si los criterios de estratificación de B son los mismos que los de A, entonces.

$$N_{h^{(A)}h^{(B)}} = 0; \forall\left\{\left(h^{(A)}, h^{(B)}\right) \middle| h^{(A)} \neq h^{(B)}\right\} \text{ y } N_{h^{(B)}} = N_{h^{(A)}h^{(B)}}; \forall h^{(A)} = h^{(B)} \text{ y}$$
$$N_{h^{(A)}h^{(B)}} + N_{h^{(A)},(L+1)^{(B)}} = N_{h^{(A)}}; \forall\left\{h=1,2,\cdots,L\right\}$$

El estimador del total

Sea $y_{h^{(A)}h^{(B)}i}$ el valor de la variable en estudio en el i ésimo- elemento de la población en el estrato $H^{(A)}$ $H^{(B)}$

El total de dominio

$$Y_D = \sum_{h^{(A)}=1}^{L^{(A)}} \sum_{h^{(B)}=1}^{L^{(B)}} Y_{h^{(A)}h^{(B)}}$$

Dónde:

$$Y_{h^{(A)}h^{(B)}} = \sum_{i=1}^{N_{h^{(A)}h^{(B)}}} y_{h^{(A)}h^{(B)}_i} \text{ es el total del dominio en el estrato } h^{(A)}h^{(B)}.$$

Varianza del estimador

La varianza de $\hat{Y}_B$ viene dada por la siguiente expresión:

$$V\hat{Y}_B = \sum_{h^{(A)}=1}^{L^{(A)}} \sum_{h^{(B)}=1}^{L^{(B)}} V\hat{Y}_{h^{(A)}h^{(B)}}$$

donde,

$$V\hat{Y}_{h^{(A)}h^{(B)}} = \sum_{i=1}^{N_{h^{(A)}h^{(B)}}} \frac{1-\varpi_{h^{(A)}h^{(B)}_i}}{\varpi_{h^{(A)}h^{(B)}_i}} y^2_{h^{(A)}h^{(B)}_i} + \sum_{i=1}^{N_{h^{(A)}h^{(B)}}} \sum_{i'(\neq i)=1}^{N_{h^{(A)}h^{(B)}}} \frac{\varpi_{h^{(A)}h^{(B)}_{ii'}} - \varpi_{h^{(A)}h^{(B)}_i}\varpi_{h^{(A)}h^{(B)}_{i'}}}{\varpi_{h^{(A)}h^{(B)}_i}\varpi_{h^{(A)}h^{(B)}_{i'}}} y_{h^{(A)}h^{(B)}_i} y_{h^{(A)}h^{(B)}_{i'}}$$

o bien,

$$V\hat{Y}_{h^{(A)}h^{(B)}} = \frac{1}{2} \sum_{i=1}^{N_{h^{(A)}h^{(B)}}} \sum_{i'(\neq i)=1}^{N_{h^{(A)}h^{(B)}}} \left(\varpi_{h^{(A)}h^{(B)}_{ii'}} - \varpi_{h^{(A)}h^{(B)}_i}\varpi_{h^{(A)}h^{(B)}_{i'}}\right)\left(\frac{y_{h^{(A)}h^{(B)}_i}}{\varpi_{h^{(A)}h^{(B)}_i}} - \frac{y_{h^{(A)}h^{(B)}_{i'}}}{\varpi_{h^{(A)}h^{(B)}_{i'}}}\right)^2$$

Estimador de la varianza del estimador

Siempre que $\varpi_{h^{(A)}h^{(B)}_{ii'}} > 0; \forall i, i'$, un estimador insesgado de $V\hat{Y}_{h^{(A)}h^{(B)}}$ es:

$$\hat{V}\hat{Y}_{h^{(A)}h^{(B)}} = \sum_{i=1}^{N_{h^{(A)}h^{(B)}}} \frac{1-\varpi_{h^{(A)}h^{(B)}_i}}{\varpi_{h^{(A)}h^{(B)}_i}} \frac{y^2_{h^{(A)}h^{(B)}_i}}{\varpi_{h^{(A)}h^{(B)}_i}} + \sum_{i=1}^{N_{h^{(A)}h^{(B)}}} \sum_{i'(\neq i)=1}^{N_{h^{(A)}h^{(B)}}} \frac{\varpi_{h^{(A)}h^{(B)}_{ii'}} - \varpi_{h^{(A)}h^{(B)}_i}\varpi_{h^{(A)}h^{(B)}_{i'}}}{\varpi_{h^{(A)}h^{(B)}_i}\varpi_{h^{(A)}h^{(B)}_{i'}}} \frac{y_{h^{(A)}h^{(B)}_i} y_{h^{(A)}h^{(B)}_{i'}}}{\varpi_{h^{(A)}h^{(B)}_{i'}}}$$

En ocasiones, esa varianza resulta negativa. En ese caso se toma como estimación de la varianza el valor cero.

Con muestras simples, $\varpi_{h^{(A)}h^{(B)}_i} = 1-\left(1-f_h^{(A)}\right)\left(1-f_h^{(B)}\right) = \varpi_{h^{(A)}h^{(B)}}; \forall i$ y

$\varpi_{h^{(A)}h^{(B)}_{ii'}} = \varpi_{h^{(A)}h^{(B)}_i} \times \varpi_{h^{(A)}h^{(B)}_{i'|i}}$ donde $\varpi_{h^{(A)}h^{(B)}_{i'|i}} = 1-\left(1-\tilde{f}_h^{(A)}\right)\left(1-\tilde{f}_h^{(B)}\right) = \tilde{\varpi}_{h^{(A)}h^{(B)}}; \forall i,i'$, y

$\tilde{f}_h^{(A)} = \dfrac{n_h^{(A)}-1}{N_h^{(A)}-1}$ y lo mismo para $\tilde{f}_h^{(B)}$.

Esto es,

$$\hat{V}\hat{Y}_{h^{(A)}h^{(B)}} = \frac{1-\varpi_{h^{(A)}h^{(B)}}}{\varpi_{h^{(A)}h^{(B)}}} \frac{1}{\varpi_{h^{(A)}h^{(B)}}} \sum_{i=1}^{n_{h^{(A)}h^{(B)}}} y^2_{h^{(A)}h^{(B)}_i} + \frac{\varpi_{h^{(A)}h^{(B)}}\tilde{\varpi}_{h^{(A)}h^{(B)}} - \varpi_{h^{(A)}h^{(B)}}\varpi_{h^{(A)}h^{(B)}}}{\varpi_{h^{(A)}h^{(B)}}\varpi_{h^{(A)}h^{(B)}}} \frac{1}{\varpi_{h^{(A)}h^{(B)}}\tilde{\varpi}_{h^{(A)}h^{(B)}}} \sum_{i=1}^{n_{h^{(A)}h^{(B)}}} \sum_{i'(\neq i)=1}^{n_{h^{(A)}h^{(B)}}} y_{h^{(A)}h^{(B)}_i} y_{h^{(A)}h^{(B)}_{i'}}$$

2. Generación de Factores de Expansión

Para el cálculo de los factores de expansión se llevaron a cabo dos procedimientos, primero se calculó el factor de expansión teórico (inverso de la probabilidad de selección de los SM de la muestra) y a partir del cual se calculó el factor de expansión final el cual considera el factor de corrección por cobertura.

Para el factor de expansión final se llevó a cabo el presente procedimiento:

- A partir de las bases validadas de campo y cartografía se creó una base unificada, la misma que contiene la información de N terrenos correspondientes a n segmentos. Esta base unificada tiene como variables: identificador de campo e identificador de segmento. Además, incluye la superficie por terreno levantada en campo, como la superficie por terreno dentro del segmento de estudio y por último el factor de expansión teórico.

- A partir de la base sin rechazos totales ni parciales se procedió a descontar los terrenos que contaban con intersección. Se partió de la información obtenida de la base de intersecciones misma que corresponde a los terrenos cuya información fue levantada en los cuestionarios de marco de área y de marco de lista. Al retirar los segmentos con intersección tanto parcial como total se identificó terrenos que tenían rechazo e intersección.

- Se identificó los SM que fueron subdivididos durante el proceso de levantamiento de campo

- Como siguiente paso, se procedió a estimar el factor de corrección por cobertura. Este factor de corrección fue estimado de la siguiente manera:
- Para los segmentos con rechazo total, el factor de corrección corresponde a la división entre la suma de segmentos seleccionados por estrato en cada zona para la suma de los segmentos investigados por estrato en cada zona.
- Para los segmentos con rechazo parcial, el factor de corrección corresponde a la división entre el área total del segmento seleccionado y el área investigada del segmento.
- Para los segmentos que fueron nuevamente divididos el factor de corrección corresponde

a 4 si el segmento fue subdividido una sola vez y a 16 en el caso de que haya sido subdivido dos veces.

- Finalmente se calculó los factores de expansión en base a las consideraciones anteriormente señaladas.

- Los terrenos que tiene factor cero corresponden a aquellos que fueron retirados de la base total de campo por tener intersecciones y presentar rechazos.

Post-estratificación

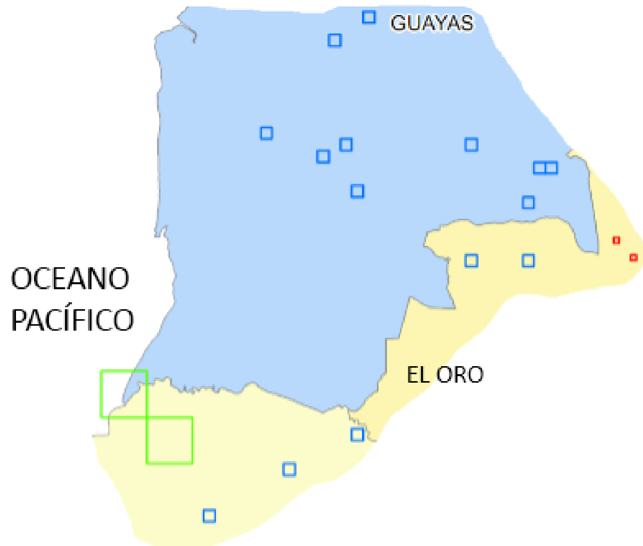
El ajuste o calibración de los factores de expansión consiste en la creación de un componente extra (δ) en la fórmula de cálculo, para poder cuadrar la población estimada por medio de la encuesta con un dato o un parámetro conocido.
Este componente "delta" permitirá el ajuste de los factores de expansión a fin de que las estimaciones de las superficies muestren una mejor aproximación de la superficie real de las provincias.

Los límites de los dominios de estudio se ajustan al perfil de las UME, mas estos no corresponden a los límites reales de cada uno de ellos. Para tener concordancia en la superficie de los dominios de estudio, se multiplican los factores de expansión a nivel de dominio por un valor $\delta \neq 0$, donde:

$$\delta = \frac{Superficie\ provincia\ límites\ reales}{Superficie\ provincia\ marco\ ESPAC}$$

Además, algunos segmentos presentaban superficie fuera del universo de estudio, perteneciente a Colombia, Perú o el Océano Pacífico. La utilización de la post-estratificación concilia esta particularidad con el objetivo de la encuesta.

Gráfico 1. Segmento de investigación de ESPAC con superficie en otra provincia y fuera del universo de estudio.




CADA
HECHO
DE TU
VIDA
Cuenta
@ecuadorencifras
INEC/Ecuador
@InecEcuador
INECEcuador
t.me/ecuadorencifras
INEC Ecuador