



Organización de las Naciones
Unidas para la Alimentación
y la Agricultura

Serie de documentos de trabajo sobre estadística de la FAO

Número 24-40

**GUÍA DE MUESTREO Y ESTIMACIÓN PARA EL
INDICADOR ODS 2.4.1
UTILIZANDO MARCOS MÚLTIPLES
Segunda edición**

Serie de documentos de trabajo sobre estadística de la FAO / 24-40

GUÍA DE MUESTREO Y ESTIMACIÓN PARA EL INDICADOR ODS 2.4.1 UTILIZANDO MARCOS MÚLTIPLES

Segunda edición

Cristiano Ferraz

Universidade Federal de Pernambuco (UFPE)

Cita requerida:

Ferraz, C. 2024. *Guía de muestreo y estimación para el indicador ODS 2.4.1 utilizando marcos múltiples*, segunda edición. Serie de documentos de trabajo sobre estadística de la FAO, n.º 24-40. Roma, FAO. <https://doi.org/10.4060/cc9550es>

Las denominaciones empleadas en este producto informativo y la forma en que aparecen presentados los datos que contiene no implican, por parte de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), juicio alguno sobre la condición jurídica o nivel de desarrollo de países, territorios, ciudades o zonas, ni sobre sus autoridades, ni respecto de la demarcación de sus fronteras o límites. La mención de empresas o productos de fabricantes en particular, estén o no patentados, no implica que la FAO los apruebe o recomiende de preferencia a otros de naturaleza similar que no se mencionan.

Las opiniones expresadas en este producto informativo son las de su(s) autor(es), y no reflejan necesariamente los puntos de vista o políticas de la FAO.

ISBN 978-92-5-138602-6

© FAO, 2024



Algunos derechos reservados. Esta obra se distribuye bajo licencia Creative Commons Atribución-NoComercial-CompartirIgual 3.0 Organizaciones intergubernamentales (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo/deed.es>).

De acuerdo con las condiciones de la licencia, se permite copiar, redistribuir y adaptar la obra para fines no comerciales, siempre que se cite correctamente, como se indica a continuación. En ningún uso que se haga de esta obra debe darse a entender que la FAO refrenda una organización, productos o servicios específicos. No está permitido utilizar el logotipo de la FAO. En caso de adaptación, debe concederse a la obra resultante la misma licencia o una licencia equivalente de Creative Commons. Si la obra se traduce, debe añadirse el siguiente descargo de responsabilidad junto a la referencia requerida: "La presente traducción no es obra de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). La FAO no se hace responsable del contenido ni de la exactitud de la traducción. La edición original en [idioma] será el texto autorizado".

Todo litigio que surja en el marco de la licencia y no pueda resolverse de forma amistosa se resolverá a través de mediación y arbitraje según lo dispuesto en el artículo 8 de la licencia, a no ser que se disponga lo contrario en el presente documento. Las reglas de mediación vigentes serán el reglamento de mediación de la Organización Mundial de la Propiedad Intelectual <http://www.wipo.int/amc/en/mediation/rules> y todo arbitraje se llevará a cabo de manera conforme al reglamento de arbitraje de la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional (CNUDMI).

Materiales de terceros. Si se desea reutilizar material contenido en esta obra que sea propiedad de terceros, por ejemplo, cuadros, gráficos o imágenes, corresponde al usuario determinar si se necesita autorización para tal reutilización y obtener la autorización del titular del derecho de autor. El riesgo de que se deriven reclamaciones de la infracción de los derechos de uso de un elemento que sea propiedad de terceros recae exclusivamente sobre el usuario.

Ventas, derechos y licencias. Los productos informativos de la FAO están disponibles en la página web de la Organización (www.fao.org/publications/es) y pueden adquirirse dirigiéndose a publications-sales@fao.org. Las solicitudes de uso comercial deben enviarse a través de la siguiente página web: www.fao.org/contact-us/licence-request. Las consultas sobre derechos y licencias deben remitirse a: copyright@fao.org.

Resumen

La presente guía enumera y describe el proceso de estimación del indicador Objetivos de Desarrollo Sostenible (ODS) 2.4.1, sostenibilidad de la agricultura, a través de sus subindicadores, y considerando diferentes niveles de desagregación en un país. Esta guía complementa la publicación de la FAO de 2021 titulado “Guía de muestreo para el indicador 2.4.1 de los ODS”. Si bien la primera aporta información útil sobre una perspectiva general de muestreo y estimación, esta guía considera el uso de diseños de muestreo de marco dual en las encuestas agrícolas nacionales, que combinan marcos de áreas con marcos de listas. En el desarrollo de esta guía también se presentarán los diseños de marcos duales y situaciones útiles para países que usan encuestas agropecuarias basadas en estos mismos. Se usaron documentos metodológicos e informes de experiencias prácticas que definen la realidad de diversos países, entre los que se cuenta Colombia, Costa Rica, México y el Perú, como base para el desarrollo de los tópicos de esta guía, así como para formular ejemplos de conceptos básicos presentes en esta primera versión.

Índice

Resumen	iii
Agradecimientos	vi
1 Introducción	1
2 El ODS 2.4.1 como parámetro poblacional	2
3 Criterios de sostenibilidad de cada subindicador	5
3.1 Subindicadores de la dimensión económica	5
3.2 Subindicadores de la dimensión ambiental	6
3.3 Subindicadores de la dimensión social	7
4 Muestreo para encuestas agropecuarias.....	9
4.1 Factores de estratificación	9
4.2 Muestreo de conglomerados	10
4.3 Muestreo con probabilidades distintas.....	10
5 Estimando el ODS 2.4.1 para un territorio nacional	11
5.1 Estimación usando un marco de área	14
5.1.1 Estimación utilizando la estrategia de segmento abierto.....	14
5.1.2 Estimación utilizando la estrategia de segmento ponderado (en una etapa)	15
5.1.3 Estimación utilizando la estrategia de segmento ponderado en dos etapas	16
5.2 Estimación usando un marco de lista.....	18
5.3 Estimación usando un marco dual	20
5.3.1 Estimación por multiplicidad simple.....	20
5.3.2 Screening estimator	23
6 Estimación para dominios subnacionales.....	25
7 Conclusión.....	27
Referencias.....	28

Agradecimientos

El autor agradece al Estadístico Regional Michael Rahija (FAO) por la invitación para el desarrollo de esta guía; a Alethea Gabriela Candia (FAO), Gloria Vargas (FAO) y Xinia Andrade Ruiz (INEC) por la lectura, correcciones y sugerencias de mejoras al texto. Cualquier error o imprecisión son únicamente responsabilidad del autor. También agradece a la Secretaría General de la Facultad Latino Americana de Ciencias Sociales (FLACSO) por el trabajo llevado a cabo para que esta guía pudiese ser desarrollada bajo la Carta de Acuerdo firmada con la FAO en el “Apoyo a los sistemas estadísticos nacionales para el fortalecimiento de los indicadores ODS 2.4.1 y 12.3.1.a”. Finalmente, agradece a Tal Pinto (FAO) por la edición del documento.

1 Introducción

Creado para monitorear los avances en la consecución de la meta 2.4 – para 2030, asegurar la sostenibilidad de los sistemas de producción de alimentos y aplicar prácticas agrícolas resilientes que aumenten la productividad y la producción, contribuyan al mantenimiento de los ecosistemas, fortalezcan la capacidad de adaptación al cambio climático, los fenómenos meteorológicos extremos, las sequías, las inundaciones y otros desastres, y mejoren progresivamente la calidad del suelo y la tierra –, el ODS-i 2.4.1 es una proporción de superficie agrícola en que se practica una agricultura productiva y sostenible. A fin de ponderar cómo inciden las dimensiones económicas, ambientales y sociales en la definición una agricultura productiva y sostenible, el indicador 2.4.1 es una función resumen, típicamente del valor mínimo, de una serie de 11 subindicadores. Cada subindicador corresponde a una proporción de superficie agrícola productiva y sostenible de acuerdo con un criterio específico. Así, este indicador puede ser interpretado como la proporción más grande del área agrícola nacional de la que se puede afirmar que cumple con los criterios de sostenibilidad en cada uno de los subindicadores. Aunque es posible realizar el análisis de sostenibilidad cambiando la función matemática que define el indicador, adecuando su interpretación, esta guía usa solamente la función matemática del valor mínimo como ejemplo, por ser la definición con una interpretación amplia de sostenibilidad, propuesta por la FAO.

Para lograr estimaciones adecuadas de todos los parámetros relacionados con el ODS-i 2.4.1, independiente del nivel de desagregación, es necesario, en primer lugar, tener en cuenta el conjunto de definiciones conceptuales de cada uno de ellos. Por ejemplo, el subindicador #1 informa la proporción de área agrícola nacional con un alto valor de producción agrícola por hectárea. Así, por ejemplo, solo puede ser comprendido en su totalidad contando con una definición para “valor de producción agrícola por hectárea” y para el adjetivo “alto”, presente en la expresión “alto valor de producción”. Cabe mencionar que, para todos los subindicadores aquí tratados, han sido definidos y justificados en diversas publicaciones¹ de la FAO. En consecuencia, en este documento solo se ofrecerán estas definiciones de forma abreviada.

Por último, para garantizar que los parámetros se estimen adecuadamente, estos deben formalizarse como funciones de cantidades mensurables a partir del proceso de medición típico de las encuestas agropecuarias. Comprendiendo el aspecto funcional de cada parámetro de interés, así como el diseño muestral empleado en una encuesta nacional, es posible proponer estimadores consistentes del punto de vista estadístico.

Esta guía representa un primer esfuerzo para ayudar a estudiar estos aspectos de forma objetiva, proponiendo estimadores consistentes para cada uno de los subindicadores, y por consecuencia, para el ODS-i 2.4.1.

¹ Ver FAO (2023a, 2023b) por ejemplo.

2 El ODS 2.4.1 como parámetro poblacional

Las descripciones del ODS-i 2.4.1 y de sus subindicadores como parámetros poblacionales destacan el aspecto funcional de cada uno de ellos. Una proporción de área agrícola productiva y sostenible se define como una razón entre totales de superficies de áreas. Las notas metodológicas de la FAO recomiendan el informe del ODS-i 2.4.1 como:

$$R_{241} = \min(R_{\#1}, \dots, R_{\#11}), \quad (1)$$

donde cada uno de los subindicadores $R_{\#i}: i \in \{1, \dots, 11\}$ corresponde a la expresión:

$$R_{\#i} = \frac{TSAS\#i}{TSA}, \quad (2)$$

en la cual:

- $TSAS\#i$ representa el total de superficie de área agrícola nacional en que se practica una agricultura considerada productiva y sostenible de acuerdo con el criterio $C_{\#i}$ usado por el subindicador $R_{\#i}$; y
- TSA representa el total de superficie de área agrícola nacional.

Así, R_{241} representa la proporción de área agrícola en que se practica una agricultura considerada productiva y sostenible en el país, desde un punto de vista multidimensional, en el que los once criterios $C_{\#i}: i \in \{1, \dots, 11\}$ de productividad y sostenibilidad, relacionados a los subindicadores, son atendidos.

$R_{\#i}$ es considerado un parámetro complejo, pues consiste en una razón entre totales. Sin embargo, es posible generar una estimación consistente del subindicador, del punto de vista estadístico, a partir de la estimación de cada total por separado. Por ello, se recomienda el uso de un estimador $\hat{R}_{\#i}$.

$$\hat{R}_{\#i} = \frac{\widehat{TSAS\#i}}{\widehat{TSA}}. \quad (3)$$

La forma de los estimadores $\widehat{TSAS\#i}$ y \widehat{TSA} depende del diseño de la muestra, lo que será discutido más adelante. Siguiendo el mismo principio, se recomienda como estimador del indicador R_{241} , la expresión:

$$\hat{R}_{241} = \min(\hat{R}_{\#1}, \dots, \hat{R}_{\#11}). \quad (4)$$

Aunque cada subindicador tenga su propio criterio $C_{\#i}$ de sostenibilidad, la forma con que cada uno de ellos es reportado sigue una estructura de clasificación simple: “Verde” (deseable), “Amarillo” (aceptable) y “Rojo” (insostenible).

Esto abre la posibilidad de contar con un estimador general, aplicable a todos los subindicadores. Para avanzar en una caracterización cuantitativa común, es necesario considerar el proceso de medición de las encuestas agropecuarias.

Por lo general, las encuestas agropecuarias usan unidades muestrales de naturaleza distinta (como segmentos de área, productores y unidades de producción), dependiendo del tipo de marco empleado, sea de área o de lista. En ocasiones, las encuestas nacionales con un diseño dual, que mezclan ambos marcos, tienen ambas unidades muestrales. Sin embargo, cualquiera sea el caso, la información

recolectada para medir los criterios de sostenibilidad $C_{\#i}$ provienen de una entrevista, con un cuestionario estructurado, en que la unidad de referencia es el establecimiento agropecuario. En consecuencia, las respuestas obtenidas están en función a la superficie de área agrícola del establecimiento agropecuario (agro), es decir la superficie de área usada para sembrar cultivos o criar ganado. Este es el alcance de definición del indicador 2.4.1, la que permite tomar como universo, o población objetivo, el conjunto de todos los establecimientos agropecuarios de un país. En este contexto, es posible definir cada subindicador de manera similar, adoptando la siguiente notación:

- $U = \{1, \dots, N\}$ representa el universo de N establecimientos agropecuarios nacionales.
- y_k representa la superficie de área agrícola del establecimiento agro $k \in U$.
- $V_{(i)k}$ representa una variable indicadora de que el establecimiento agro $k \in U$ es clasificado en la categoría "Verde" de sostenibilidad, de acuerdo con el criterio $C_{\#i}$.

$$V_{(i)k} = \begin{cases} 1, & \text{si el establecimiento agro } k \in U \text{ es "Verde" por el criterio } C_{\#i}; \\ 0, & \text{si el establecimiento agro } k \in U \text{ no es "Verde"}. \end{cases}$$

- $A_{(i)k}$ representa una variable indicadora de que el establecimiento agro $k \in U$ es clasificado en la categoría "Amarillo" de sostenibilidad, de acuerdo con el criterio $C_{\#i}$:

$$A_{(i)k} = \begin{cases} 1, & \text{si el establecimiento agro } k \in U \text{ es "Amarillo" por el criterio } C_{\#i}; \\ 0, & \text{si el establecimiento agro } k \in U \text{ no es "Amarillo"}. \end{cases}$$

Así, es posible describir los parámetros poblacionales relativos a cada subindicador en función de la categoría deseada. Por ejemplo, con base en la expresión (2), mediante

$$TSAS_{\#i} = \sum_{k \in U} y_k V_{(i)k}, \quad (5)$$

y

$$TSA = \sum_{k \in U} y_k, \quad (6)$$

la expresión

$$R_{\#i} = \frac{\sum_{k \in U} y_k V_{(i)k}}{\sum_{k \in U} y_k} \quad (7)$$

que representa la forma paramétrica del subindicador $R_{\#i}$ en función de la categoría "Verde" de sostenibilidad. Es decir, en este caso, $R_{\#i}$ representa la proporción de área agrícola nacional con el **nivel deseable de sostenibilidad** para el criterio $C_{\#i}$. En paralelo, los parámetros poblacionales

$$R_{\#i1} = \frac{\sum_{k \in U} y_k A_{(i)k}}{\sum_{k \in U} y_k}, \quad (8)$$

$$R_{\#i2} = \frac{\sum_{k \in U} y_k (V_{(i)k} + A_{(i)k})}{\sum_{k \in U} y_k} \quad (9)$$

y

$$R_{\#i3} = \frac{\sum_{k \in U} y_k (1 - V_{(i)k} A_{(i)k})}{\sum_{k \in U} y_k} \quad (10)$$

representan la proporción de área agrícola de un mismo subindicador usando categorías distintas de sostenibilidad, donde:

- $R_{\#i1}$ es la proporción de área agrícola con un **nivel aceptable de sostenibilidad** (Amarillo) para el criterio $C_{\#i}$.
- $R_{\#i2}$ es la proporción de área agrícola con **nivel aceptable o deseable de sostenibilidad** (Amarillo o Verde) para el criterio $C_{\#i}$.
- $R_{\#i3}$ es la proporción de área agrícola con **nivel de insostenibilidad** (Rojo) para el criterio $C_{\#i}$.

Sin perder el aspecto más general del problema, en esta guía, los ejemplos de estimación consideran los subindicadores definidos en función de la categoría “Verde” de sostenibilidad, como en (7), así como los demás, $R_{\#i1}$, $R_{\#i2}$ y $R_{\#i3}$, definidos de una forma general en (8), (9) y (10), respectivamente.

3 Criterios de sostenibilidad de cada subindicador

Conocer los criterios de sostenibilidad de cada subindicador es fundamental para la estimación adecuada del indicador 2.4.1. El Perú y Costa Rica cuentan con la experiencia en el uso de los conceptos relativos a cada criterio, y en la forma como la información fue recopilada a través de cuestionarios, considerando la clasificación de cada subindicador por categoría de sostenibilidad. En este capítulo se destaca los aspectos más generales de los criterios, agrupados por dimensión. El Cuadro 1 presenta un resumen de los temas considerados por cada subindicador.

Cuadro 1. Subindicadores por dimensión.

Subindicador	Criterio	Dimensión
1	Valor de la producción agrícola por hectárea	Económica
2	Ingresos agrícolas netos	Económica
3	Mecanismos de mitigación de riesgos	Económica
4	Magnitud de la degradación del suelo	Ambiental
5	Variación en la disponibilidad del agua	Ambiental
6	Gestión de fertilizantes	Ambiental
7	Gestión de plaguicidas	Ambiental
8	Uso de prácticas de apoyo a la biodiversidad agrícola	Ambiental
9	Salarios en la agricultura	Social
10	Escala de experiencia de inseguridad alimentaria (FIES)	Social
11	Seguridad de los derechos de tenencia de la tierra	Social

Fuente: Elaboración por autor.

3.1 Subindicadores de la dimensión económica

Los tres subindicadores de la dimensión económica son $R_{\#1}$, $R_{\#2}$ y $R_{\#3}$. Sus criterios de sostenibilidad son $C_{\#1}$, $C_{\#2}$ y $C_{\#3}$, respectivamente:

- $C_{\#1}$: Compara el valor de VPA_k , la **producción agrícola por hectárea** del establecimiento agro $k \in U$, con un percentil 90 ($P90$) de esta misma cantidad, resultando en
 - Verde, si $VPA_k \geq \frac{2}{3}P90$;
 - Amarillo, si $\frac{1}{3}P90 \leq VPA_k < \frac{2}{3}P90$; y
 - Rojo, si $VPA_k < \frac{1}{3}P90$.
- $C_{\#2}$: Compara los valores de IAN_k , el **ingreso agrícola neto** del establecimiento agro $k \in U$, para los tres últimos años, resultando en:

- Verde, se $IAN_k \geq 0$ para los tres últimos años;
 - Amarillo, se $VPA_k > 0$ para al menos uno de los tres últimos años; y
 - Rojo, se $VPA_k < 0$ para los tres últimos años.
- **$C_{\#3}$** : Identifica si un establecimiento agro $k \in U$ hace uso de **mecanismos de mitigación de riesgos**, es decir, se tiene acceso a crédito, seguro y se diversifica su explotación agrícola, resultando en:
 - Verde, se tiene acceso o ha utilizado al menos dos de los mecanismos;
 - Amarillo, se tiene acceso o ha utilizado uno solo de los mecanismos; y
 - Rojo, no se tiene acceso a ningún de los tres mecanismos.

3.2 Subindicadores de la dimensión ambiental

Hay cinco subindicadores de la dimensión ambiental: $R_{\#4}$, $R_{\#5}$, $R_{\#6}$, $R_{\#7}$ y $R_{\#8}$. Sus criterios de sostenibilidad son $C_{\#4}$, $C_{\#5}$, $C_{\#6}$, $C_{\#7}$ y $C_{\#8}$, respectivamente:

- **$C_{\#4}$** : Compara el **nivel de la degradación del suelo** por erosión, reducción de fertilidad, salinización, y encharcamiento. Se mide como la PAD_k , proporción de área agrícola del establecimiento $k \in U$ afectada por degradación, resultando en:
 - Verde, se $PAD_k \leq 10\%$;
 - Amarillo, se $10\% < PAD_k \leq 50\%$; y
 - Rojo, se $PDA_k > 50\%$.
- **$C_{\#5}$** : Identifica si hay **variación en la disponibilidad de agua** a lo largo de los años, resultando en:
 - Verde, si la disponibilidad de agua es estable a lo largo de los años o no utiliza agua para regar cultivos en más del 10 por ciento del área agrícola de la finca.
 - Amarillo, cuando utiliza el agua para regar los cultivos en al menos el 10 por ciento del área agrícola de la finca, no se sabe si la disponibilidad del agua se mantiene o se constata una reducción en su disponibilidad, pero existe una organización responsable por la distribución eficaz del agua entre los usuarios.
 - Rojo, en todos los demás casos.
- **$C_{\#6}$** : Identifica si un establecimiento agro $k \in U$ adopta medidas (de un total de ocho) de mitigación de **riesgo de contaminación por fertilizantes**.
 - Verde, si el establecimiento agro $k \in U$ adopta al menos cuatro de las medidas consideradas;
 - Amarillo, si el establecimiento agro $k \in U$ adopta dos o tres de las medidas consideradas;
 - Rojo, en los demás casos.

- **$C_{\#7}$** : Identifica si un establecimiento agro $k \in U$ adopta medidas de mitigación de **riesgo de contaminación por plaguicidas**. Además de adoptar medidas de mitigación consideradas en una lista, se considera también el tipo de plaguicida utilizado, resultando en:
 - Verde, cuando el establecimiento agro $k \in U$ solo utiliza plaguicidas moderadamente peligrosos o poco peligrosos (clases II y III de la Organización Mundial de la Salud [OMS]), cumple tres medidas relacionadas con la salud, y al menos cuatro medidas relacionadas con el medio ambiente.
 - Amarillo, cuando el establecimiento agro $k \in U$ solo utiliza plaguicidas moderadamente peligrosos o poco peligrosos (clases II y III de la OMS) y adopta al menos dos medidas de cada grupo para mitigar los riesgos para el medio ambiente y los riesgos para la salud.
 - Rojo, cuando el establecimiento agro $k \in U$ utiliza plaguicidas sumamente peligrosos, muy peligrosos (clases Ia y Ib de la OMS) o ilegales, o utiliza plaguicidas moderadamente peligrosos o ligeramente peligrosos sin adoptar medidas específicas para mitigar los riesgos para el medio ambiente y la salud asociados con la utilización de dichos plaguicidas (menos de dos medidas de alguna de las listas consideradas).
- **$C_{\#8}$** : Identifica si un establecimiento agro $k \in U$ hace **uso de prácticas de apoyo a la biodiversidad agrícola** (de un total de seis) resultando en:
 - Verde, cuando el establecimiento agro $k \in U$ hace uso de al menos tres de las seis prácticas.
 - Amarillo, cuando el establecimiento agro $k \in U$ solo utiliza uno o dos de las prácticas.
 - Rojo, cuando el establecimiento agro $k \in U$ no utiliza ninguna de las practicas.

3.3 Subindicadores de la dimensión social

Son tres los subindicadores de la dimensión social: $R_{\#9}$, $R_{\#10}$ y $R_{\#11}$. Sus criterios de sostenibilidad son $C_{\#9}$, $C_{\#10}$ y $C_{\#11}$, respectivamente:

- **$C_{\#9}$** : Se relaciona con el **salario en la agricultura**. Identifica si el salario pagado, en promedio, a los trabajadores agrícolas no calificados, por un establecimiento agro $k \in U$, es igual, superior o inferior al salario mínimo del país, resultando en:
 - Verde, cuando el establecimiento agro $k \in U$ paga en promedio salarios superiores al salario mínimo.
 - Amarillo, cuando el establecimiento agro $k \in U$ paga en promedio salarios iguales al salario mínimo.
 - Rojo, cuando el establecimiento agro $k \in U$ paga en promedio salarios inferiores al salario mínimo.

- **$C_{\#10}$** : Se define en función de una **escala de experiencia de inseguridad alimentaria (FIES, por sus siglas en inglés)**. Identifica el grado de inseguridad alimentaria, leve, moderado o severo, a partir de la probabilidad de cada una de ellas, resultando en:
 - Verde, si la probabilidad de que el hogar del titular de la explotación padezca inseguridad moderada o grave es inferior a 0,5 y la probabilidad de que padezca inseguridad alimentaria grave es inferior a 0,5.
 - Amarillo, si la probabilidad de que el hogar del titular de la explotación padezca inseguridad moderada o grave es superior a 0,5 y la probabilidad de que padezca inseguridad alimentaria grave es inferior a 0,5.
 - Rojo, si la probabilidad de que el hogar del titular de la explotación padezca inseguridad alimentaria grave es superior a 0,5.

- **$C_{\#11}$** : Se define en función de la **seguridad de los derechos a tenencia de la tierra**. Identifica si el titular del establecimiento agro $k \in U$ tiene un documento oficial de tenencia de tierra, resultando en:
 - Verde, cuando tiene un documento oficial con el nombre del titular del establecimiento agro, o se tiene el derecho de vender o de transmitir por herencia cualquiera de las parcelas de la explotación.
 - Amarillo, cuando tiene un documento oficial, aunque no figure el nombre del titular ni del establecimiento agro.
 - Rojo, en todos los demás casos.

4 Muestreo para encuestas agropecuarias

Aunque, en teoría, podría diseñarse una encuesta nacional para cumplir solo con los objetivos de desarrollo sostenible, cada país tiene necesidades particulares de información. Así, es más eficiente integrar los objetivos de estimación de indicadores de sostenibilidad con los objetivos específicos de encuestas muestrales agropecuarias nacionales ya existentes, o el desarrollo de un sistema integrado de encuestas agropecuarias que atienda a todos estos objetivos. El Programa AGRIS (FAO, 2018) es un ejemplo de sistema integrado que puede ser adoptado como referencia para el desarrollo de sistemas nacionales considerando la realidad de cada país.

Desarrollar un documento considerando cada uno de los diseños muestrales de encuestas agropecuarias en uso por los países no sería razonable. Sin embargo, es posible identificar los elementos de diseño muestral más utilizados y usarlos en una descripción intuitiva para que, por un lado, se perciba la adecuación del método de muestreo al problema de estimación del indicador 2.4.1, y por otro, se tenga una referencia para los países que necesitan desarrollar o renovar un método para sus encuestas por muestreo probabilístico.

Independiente del tipo de marco en uso, tres elementos son comúnmente usados para mejorar la eficiencia de estimación:

- Estratificación;
- muestreo en etapas; y
- diseños con probabilidad proporcional a una medida de tamaño (PPT).

4.1 Factores de estratificación

Usualmente, una población objetivo es estratificada según cuando menos a uno de dos conjuntos de factores. El primero abarca los factores que definen dominios subnacionales de interés para estimación, mientras, el segundo, aquellos que adicionan información auxiliar importante para incrementar la precisión. Así, un estrato h representa uno de los niveles de un factor de estratificación, o una combinación de niveles de factores de las dos categorías.

En caso en que cada estrato es una combinación de niveles de factores, considere un país con un censo agropecuario reciente, en el cual es posible identificar un marco de áreas con terrenos de límites físicos conocidos. Imagine que la población objetivo de todos los establecimientos agropecuarios del país es completamente cubierta por el marco de área. Además, en algunos casos se identifica una relación unívoca de un terreno para cada productor agropecuario, aunque cada productor pueda tener más de un terreno, como en México, en su censo agropecuario de 2022.

Desarrollar una encuesta nacional agropecuaria para el país podría beneficiarse de una estructura de estratificación que emplea factores geopolíticos para garantizar estimaciones subnacionales, y factores relacionados a la intensidad de uso de suelo, mensurados a través del uso de imágenes satelitales, para añadir precisión a las estimaciones. Singularizando el factor geopolítico de interés, como los estados del país, en el caso de México, resulta en que hay 32 niveles (estados). Si el factor intensidad de uso de suelo tiene, por ejemplo, tres niveles (alta, moderada y baja intensidad), todo el proceso de estratificación utilizaría $32 \times 3 = 96$ estratos.

4.2 Muestreo de conglomerados

El uso de conglomerados en diseños muestrales puede ser una alternativa de recopilación de datos más económica. Por definición, un conglomerado es un conjunto de elementos de una población objetivo. En encuestas agropecuarias con poblaciones objetivos definidas como todos los establecimientos agropecuarios, un conglomerado puede tener diversas formas, dependiendo del tipo de marco. En marcos de áreas, por ejemplo, es común que un conglomerado tome la forma de un área más grande, donde es posible identificar un conjunto de establecimientos agropecuarios. En ocasiones, el número de establecimientos agropecuarios en el conglomerado puede ser muy alto, lo que justifica dividirlo en subáreas más pequeñas, donde cada una de ellas contiene un conjunto más pequeño de establecimientos. En los marcos de lista, los conglomerados pueden tener otras definiciones. Un marco que lista todas las asociaciones de productores agrupados por aldea, por ejemplo, es un marco en que cada aldea representa un conjunto de productores, con una relación conocida con el concepto de establecimientos agropecuarios de un país.

La Encuesta Nacional de Agropecuaria de Costa Rica (ENA) subdivide su marco de área en áreas más pequeñas, llamadas segmentos de área, en las cuales se puede identificar un conjunto de fincas (establecimiento agropecuario). Cada segmento de área es un conglomerado seleccionado por un muestreo probabilístico. Se realiza un barrido de todos los segmentos de área en los conglomerados seleccionados, y todas las fincas identificadas son investigadas. Sería posible, todavía, añadir una etapa de selección al proceso de muestreo, en la cual podrían seleccionarse fincas al interior de cada conglomerado, obteniendo así un muestreo en dos etapas.

4.3 Muestreo con probabilidades distintas

Cuando los marcos de muestreo cuentan con informaciones auxiliares disponibles, es posible utilizar diseños muestrales de probabilidades distintas, típicamente de valores proporcionales a una medida de importancia, comúnmente llamada medida de tamaño. Así, las probabilidades de selección son calculadas para incluir estas informaciones, ganando precisión, siempre y cuando la información auxiliar tenga algún grado de correlación con la variable de interés.

5 Estimando el ODS 2.4.1 para un territorio nacional

La FAO recomienda estimar el indicador 2.4.1 para todo el territorio nacional. Además, cuando sea posible, generar estimaciones desagregadas para dominios subnacionales supone ganar información útil para las gestiones de sostenibilidad, en acuerdo con la Agenda 2030.

En este capítulo, el problema de generar estimativas del indicador para un territorio nacional es abordado considerando los elementos de diseño muestral introducidos antes. Además de estratificación y probabilidades distintas de selección, dos situaciones de diseño son consideradas: una con muestreo en dos etapas, y otra con muestreo en una etapa.

Para efecto de notación, imagínese que en total se identifican H estratos, y que los elementos de análisis, los establecimientos agropecuarios, coinciden con el último nivel de unidad de muestreo de cada estrato, o están incluidos en ellos, cuando el último nivel es un conglomerado. La siguiente notación es introducida:

- $U = \{1, \dots, N\}$ representa el universo de N establecimientos agropecuarios nacionales, identificados en una estructura poblacional en la cual:

- $U = \bigcup_{h=1}^H U_h$, es decir que la población es estratificada en H estratos, con

$U_h = \{1, \dots, M_h\}$, representando el conjunto de M_h conglomerados del estrato h , tal que $M = \sum_{h=1}^H M_h$ es el total de conglomerados en la población.

- $U_{hi} = \{1, \dots, N_{hi}\}$ representa el conjunto de N_{hi} establecimientos agropecuarios del conglomerado i en el estrato h , tal que $N_h = \sum_{i=1}^{M_h} N_{hi}$ es el total de establecimientos agropecuarios en el estrato h ; y
 - $N = \sum_{h=1}^H N_h$.
- y_{hij} que representa la superficie de área agrícola del establecimiento agro j del conglomerado i , en el estrato h .

Por ejemplo, en la ENA de Costa Rica la finca es la unidad de análisis; y usa un marco dual de muestreo, con un marco de área y un marco de lista de grandes productores, para cubrir toda la población. Por un lado, un censo es aplicado al marco de lista. Por otro lado, el marco de área de la ENA es dividido en grandes áreas para efectos de clasificar en cinco estratos, de acuerdo con un criterio de intensidad de uso agrícola del suelo. En cada estrato, una muestra de conglomerados de áreas menores es seleccionada. Todas las fincas con explotación de parcelas en estos conglomerados son entrevistadas. En este caso, y_{hij} representa la superficie de área agrícola de la finca (establecimiento agro) j del conglomerado (áreas menores) i , en el estrato h .

Por conveniencia, en esta guía, algunas veces será hecha referencia a los establecimientos agropecuarios simplemente como " hij ", para denotar el establecimiento agropecuario j del conglomerado i , en el estrato h .

Las variables anteriormente definidas como " $V_{(i)k}$ " y " $A_{(i)k}$ " necesitan una pequeña adaptación de notación, cambiando el índice i por c , pasando a ser escritas como:

- $V_{(c)hij}$, que representa una variable indicadora de que el establecimiento agro j del conglomerado i en el estrato h es clasificado en la categoría "Verde" de sostenibilidad, de acuerdo con el criterio $c \in \{C_{\#1}, \dots, C_{\#11}\}$:

$$\circ V_{(c)hij} = \begin{cases} 1, & \text{si el establecimiento agro "hij" es "Verde" por el criterio } c; \\ 0, & \text{si el establecimiento agro "hij" no es "Verde";} \end{cases}$$

y

- $A_{(c)hij}$ representa una variable indicadora de que el establecimiento agro j del conglomerado i en el estrato h es clasificado en la categoría "Amarillo" de sostenibilidad, de acuerdo con el criterio $c \in \{C_{\#1}, \dots, C_{\#11}\}$:

$$\circ A_{(c)hij} = \begin{cases} 1, & \text{si el establecimiento agro "hij" es "Amarillo" por el criterio } c; \\ 0, & \text{si el establecimiento agro "hij" no es "Amarillo".} \end{cases}$$

Los parámetros poblacionales (7), (8), (9) y (10), relativos a cada subindicador, pueden ser escritos como una razón entre:

$$TSAS\#c = \sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij} V_{(c)hij}, \quad (11)$$

y

$$TSA = \sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij}, \quad (12)$$

como sigue:

$$R_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij} V_{(c)hij}}{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij}}, \quad (13)$$

$$R_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij} A_{(c)hij}}{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij}}, \quad (14)$$

$$R_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij} (V_{(c)hij} + A_{(c)hij})}{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij}}, \quad (15)$$

$$R_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij} (1 - V_{(c)hij} A_{(c)hij})}{\sum_{h=1}^H \sum_{i \in U_h} \sum_{j \in U_{hi}} y_{hij}}. \quad (16)$$

Estimación en dos etapas

Para generar estimaciones nacionales consistentes, del punto de vista del diseño muestral, para los parámetros (13), (14), (15) y (16), es necesario tener en consideración el proceso de aleatorización usado en el diseño. Cuando se tiene una selección de muestra en dos etapas en las cuales, en cada estrato, se saca una muestra S_h de conglomerados, y en cada conglomerado se saca una muestra S_{hi} de establecimientos, tal que π_{hij} es la probabilidad de selección del establecimiento j del conglomerado i en el estrato h , es posible escribir: $\pi_{hij} = \pi_{hi}\pi_{j|hi}$, donde π_{hi} es la probabilidad de selección del

conglomerado i en el estrato h y $\pi_{j|hi}$ es la probabilidad de selección de la unidad de producción j dado que el segmento i a lo cual el pertenece, dentro del estrato h , fue seleccionado para la muestra. Como resultado, en una estimación en dos etapas, los estimadores de los indicadores de interés asumen las siguientes fórmulas:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij} V_{(c)hij}}{\pi_{hi} \pi_{j|hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij}}{\pi_{hi} \pi_{j|hi}}}, \quad (17)$$

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij} A_{(c)hij}}{\pi_{hi} \pi_{j|hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij}}{\pi_{hi} \pi_{j|hi}}}, \quad (18)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij} (V_{(c)hij} + A_{(c)hij})}{\pi_{hi} \pi_{j|hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij}}{\pi_{hi} \pi_{j|hi}}}, \quad (19)$$

y

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij} (1 - V_{(c)hij} A_{(c)hij})}{\pi_{hi} \pi_{j|hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{y_{hij}}{\pi_{hi} \pi_{j|hi}}}. \quad (20)$$

Estimación en una etapa

Cuando se tiene una selección de muestra en una etapa en la cual, en cada estrato, se saca una muestra S_h de conglomerados, y en cada conglomerado se hace un barrido de todos los establecimientos agropecuarios, tal que π_{hi} es la probabilidad de selección del conglomerado i en el estrato h , los estimadores (17), (18), (19) y (20) necesitan un ajuste, resultando respectivamente en:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij} V_{(c)hij}}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij}}{\pi_{hi}}}, \quad (21)$$

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij} A_{(c)hij}}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij}}{\pi_{hi}}}, \quad (22)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij} (V_{(c)hij} + A_{(c)hij})}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij}}{\pi_{hi}}}, \quad (23)$$

y

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij}(1 - V_{(c)hij} A_{(c)hij})}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{y_{hij}}{\pi_{hi}}}. \quad (24)$$

Los estimadores introducidos hasta el momento son generales, lo suficiente para el uso en la práctica. Sin embargo, es posible realizar adaptaciones de acuerdo con el tipo de marco utilizado. En esta guía, las adaptaciones son introducidas para los diseños más comunes de encuestas agropecuarias que usan un solo marco, sea de área o de lista; y un sistema de marcos duales, con el uso simultáneo de marcos de área y de lista.

5.1 Estimación usando un marco de área

Los marcos de área usan el territorio nacional como referencia para cubrir la población-objetivo de una encuesta. En esta guía se considera que los marcos de área usan segmentos de área como unidades de muestreo, y unidades de producción como unidades de observación y análisis.

El concepto de segmentos de área empleado aquí es general, es decir, un segmento de área es una parcela de terreno que tiene forma regular o irregular. Los segmentos cuadrados, también llamados de cuadrículas, son ejemplos de segmentos de forma regular: por otra parte, los segmentos formados a partir de límites físicos son ejemplos de segmentos irregulares. El marco de área utilizado para la Encuesta Nacional Agropecuaria de Colombia, elaborado por el Departamento Administrativo Nacional de Estadística (DANE), es un marco en lo cual las unidades primarias de muestreo forman una partición del territorio nacional del país a través de segmentos de área irregulares delimitados por accidentes físicos, naturales o culturales, identificados por imágenes satelitales.

El concepto de unidades de producción también es suficientemente general para representar diversos casos. En Costa Rica, por ejemplo, su encuesta nacional tiene como unidad de observación y análisis una finca, a la que se aplica un cuestionario. Se trata de la situación ideal para recolectar las informaciones necesarias para generar estimaciones de los subindicadores de interés para la definición del indicador de sostenibilidad 2.4.1.

Típicamente, los diseños de muestreo de un marco de área aplicado a una encuesta agropecuaria usan estratificación por criterio de intensidad de uso agropecuario del suelo y segmentos de área como unidades primarias de muestreo en cada estrato. Dos estrategias de muestreo de segmentos de área muy utilizadas para selección de unidades de producción son la estrategia de selección por segmento abierto y la estrategia de selección por segmento ponderado.

5.1.1 Estimación utilizando la estrategia de segmento abierto

Sacar una muestra de unidades de producción utilizando la estrategia de selección por segmento abierto equivale a tener un cuestionario aplicado para cada unidad de producción (establecimiento agropecuario), el que tiene una “sede” dentro de los límites del segmento. Las respuestas a las cuestiones dicen respecto a toda la unidad de producción, aunque su área delineada extrapole los límites del segmento. La definición de “sede” necesita ser un criterio claro que permita una asociación de cada unidad de producción a apenas un segmento de área, de forma que la probabilidad de selección de una unidad de producción es la misma del segmento de área en que está ubicada.

Utilizando la estrategia de segmento abierto, los estimadores del indicador 2.4.1 para todo el territorio nacional coinciden con los definidos anteriormente para un muestreo en una etapa. Así, en los diseños muestrales con probabilidad igual de selección de segmentos (por ejemplo, muestreo aleatorio simple o sistemático), dentro de cada estrato, la probabilidad de selección es $\pi_{hi} = \pi_h = m_h/M_h$, donde m_h es el número de segmentos seleccionados para la muestra del estrato h , y M_h es el total de segmentos de área en el estrato h del marco de área. Bajo estas condiciones, los estimadores (21), (22), (23) y (24) tienen la siguiente forma, respectivamente:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij} V_{(c)hij}}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij}}{m_h}}, \quad (25)$$

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij} A_{(c)hij}}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij}}{m_h}}, \quad (26)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij} (V_{(c)hij} + A_{(c)hij})}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij}}{m_h}}, \quad (27)$$

y

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij} (1 - V_{(c)hij} A_{(c)hij})}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h y_{hij}}{m_h}}. \quad (28)$$

5.1.2 Estimación utilizando la estrategia de segmento ponderado (en una etapa)

Sacar una muestra de unidades de producción usando la estrategia de segmento ponderado equivale a tener un cuestionario aplicado a cada unidad de producción que tiene una intersección total o parcial de sus parcelas de áreas agrícolas con el segmento seleccionado. La superficie de área formada por la intersección de una o más parcelas de áreas agrícolas con el segmento, bajo la misma unidad de producción, son aquí llamados “*tracts*”.

Considere T_{hij} como el área del “*tract*” de la unidad de producción j en el segmento i dentro del estrato h , y A_{hij} como el área total de la unidad de producción j seleccionada a través del segmento i , dentro del estrato h . Utilizando la estrategia de segmento ponderado, las variables de interés y_{hij} reciben una ponderación proporcional al área del “*tract*” de la unidad de producción seleccionada, así que se observa, para efecto de estimación, la variable ponderada x_{hij} , tal que:

$$x_{hij} = \frac{T_{hij}}{A_{hij}} y_{hij}.$$

Las fórmulas de estimación del indicador 2.4.1 bajo esta estrategia de muestreo coinciden con las mismas de estimación en una etapa, cambiando y_{hij} por x_{hij} . Considerando diseños muestrales con probabilidades iguales para selección de segmentos dentro de cada estrato, se deriva $\pi_{hi} =$

$\pi_h = m_h/M_h$, donde m_h es el número de segmentos seleccionados para la muestra del estrato h , y M_h es el total de segmentos de área en el estrato h del marco de área. Bajo estas condiciones, los estimadores (21), (22), (23) y (24) tienen la siguiente forma, respectivamente:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij} V_{(c)hij}}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij}}{m_h}}, \quad (29)$$

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij} A_{(c)hij}}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij}}{m_h}}, \quad (30)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij} (V_{(c)hij} + A_{(c)hij})}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij}}{m_h}}, \quad (31)$$

y

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij} (1 - V_{(c)hij} A_{(c)hij})}{m_h}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \frac{M_h x_{hij}}{m_h}}. \quad (32)$$

Una desventaja de esta estrategia es la necesidad de conocer el valor de T_{hij} para calcular x_{hij} .

5.1.3 Estimación utilizando la estrategia de segmento ponderado en dos etapas

Sacar una muestra de unidades de producción con esta estrategia equivale a seleccionar una submuestra de unidades de producción dentro de cada segmento a través de un muestreo con reemplazo y probabilidad de selección proporcional al tamaño del "tract", simulado por medio de puntos.

Para llegar a la fórmula de los estimadores con esta estrategia de muestreo, es necesario considerar la misma variable auxiliar x_{hij} introducida anteriormente,

$$x_{hij} = \frac{T_{hij}}{A_{hij}} y_{hij}, \quad (33)$$

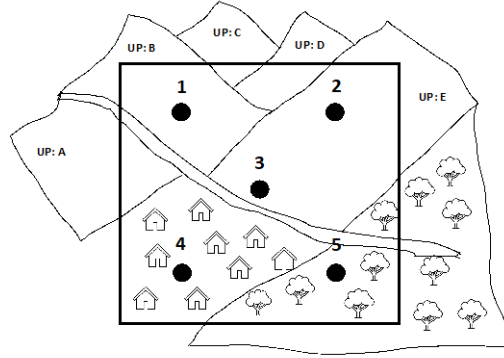
así como utilizar las expresiones generales (17), (18), (19) y (20). Para ejemplificar el raciocinio, considere la expresión (17), cambiando y_{hij} por x_{hij} :

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{x_{hij} V_{(c)hij}}{\pi_{hi} \pi_{j|hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in S_{hi}} \frac{x_{hij}}{\pi_{hi} \pi_{j|hi}}}. \quad (34)$$

El próximo paso es adaptar la fórmula (34) para acomodar el proceso de submuestreo por puntos. Considere, como ejemplo, un segmento de área con P puntos. La Figura 1 ilustra un segmento de área cuadrado con $P = 5$ puntos. Es posible identificar, en la Figura 1, cinco unidades de producción (UP): A, B, C, D y E. También es posible identificar los cinco puntos: 1, 2, 3, 4 y 5. Ningún punto se queda en los

“tracts” de las unidades de producción A, C y D, así que estas u no hacen parte de la submuestra del segmento ilustrado en la Figura 1. El punto 1 se ubicó en un “tract” de la UP B, así que la UP B es parte de la submuestra y se aplica un cuestionario. Los puntos 2 y 3 se ubicarán en la misma UP E, y por eso, también se le aplica un cuestionario. El punto 4 está en un área urbana y el punto 5 se queda en un área forestal.

Figura 1. Segmento cuadrado con cinco puntos.



Fuente: Elaboración del autor.

Considere que una muestra de m_h segmentos de área es sacada del estrato h , de forma que cada uno de los M_h segmentos tienen iguales probabilidades de selección. Considere, además, que una submuestra de unidades de producción es seleccionada por puntos (ver Figura 1). Así,

$$\pi_{hi} = \frac{m_h}{M_h}, \quad (35)$$

y

$$\pi_{j|hi} = \frac{T_{hij}}{T_{hi}} P, \quad (36)$$

donde T_{hi} representa la superficie de área del segmento i en el estrato h , P es el total de puntos utilizados en cada segmento, y T_{hij} es el área del “tract” de la unidad de producción j ubicada en el segmento i , dentro del estrato h . Utilizando los valores de x_{hij} , π_{hi} y $\pi_{j|hi}$ de las expresiones (33), (35) y (36), es posible observar que:

$$\frac{x_{hij}}{\pi_{hi}\pi_{j|hi}} = \frac{M_h}{m_h} \frac{T_{hi}}{T_{hij}P} \frac{T_{hij}}{A_{hij}} y_{hij} = \frac{M_h}{m_h} \frac{T_{hi}}{P A_{hij}} y_{hij}.$$

De modo que

$$I_{hijk} = \begin{cases} 1, & \text{si el punto } k \text{ se ubicó en el "tract" de la UP } hij; \\ 0, & \text{en caso contrario,} \end{cases} \quad y$$

es posible reescribir la formula (34) de la siguiente manera:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} V_{(c)hij} I_{hijk}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} I_{hijk}}. \quad (37)$$

En la expresión (37), U_{hi} representa el conjunto de todos los establecimientos agropecuarios que tienen un “tract” con el segmento i dentro del estrato h .

Utilizando el mismo raciocinio, se derivan los siguientes estimadores:

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} A_{(c)hij} I_{hijk}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} I_{hijk}}, \quad (38)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} (V_{(c)hij} + A_{(c)hij}) I_{hijk}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} I_{hijk}}, \quad (39)$$

y

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} (1 - V_{(c)hij} A_{(c)hij}) I_{hijk}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{M_h T_{hi}}{m_h P A_{hij}} y_{hij} I_{hijk}}. \quad (40)$$

Utilizando la Figura 1 como ejemplo, los valores de las variables de interés y_{hij} serían computados de la siguiente manera:

- Para el punto $k = 1$, $I_{hij1} = 1$, para la UP B, luego el valor de y_{hij} es la respuesta registrada por la UP B.
- Para el punto $k = 2$, $I_{hij2} = 1$ para la UP E, luego el valor de y_{hij} es la respuesta registrada por la UP E.
- Para el punto $k = 3$, $I_{hij3} = 1$ para la UP E, luego el valor de y_{hij} es la respuesta registrada por la UP E, es decir, hay una repetición de la misma observación registrada para el punto $k = 2$.
- Para el punto $k = 4$, $I_{hij4} = 0$, como ninguna UP es seleccionada, luego el valor de la respuesta registrada es nulo.
- Para el punto $k = 5$, $I_{hij5} = 0$, como ninguna UP es seleccionada, luego el valor de la respuesta registrada también es nulo.

5.2 Estimación usando un marco de lista

Al usar solamente un marco de lista como referencia para la encuesta, se simplifica en el contexto de muestreo, toda vez que los marcos de lista tienen las unidades de producción (establecimientos agropecuarios) como unidades de muestreo. En estas condiciones, es natural asumir el uso de elementos de estratificación y de probabilidades distintas en el diseño de la muestra. Para efecto de notación,

imáginese que se identifican H estratos y que cada unidad de producción (establecimiento agropecuario) es seleccionada en una etapa con probabilidad proporcional a una variable de tamaño conveniente. La siguiente notación es introducida:

- $U = \{1, \dots, N\}$ representa el universo de N establecimientos agropecuarios nacionales, identificados en una estructura poblacional en la cual:

- $U = \bigcup_{h=1}^H U_h$, es decir, que la población es estratificada en H estratos, con

$U_h = \{1, \dots, N_h\}$ representando el conjunto de N_h establecimientos agropecuarios del estrato h , tal que $N = \sum_{h=1}^H N_h$.

- y_{hi} representa la superficie de área agrícola del establecimiento agro i del estrato h .
- S_h representa una muestra probabilística, de tamaño n_h , seleccionada dentro del estrato h , tal que π_{hi} representa la probabilidad de selección del establecimiento agro i en el estrato h .

Las variables anteriormente definidas como " $V_{(c)hij}$ " y " $A_{(c)hij}$ " necesitan una pequeña adaptación de notación, suprimiendo el índice j y pasando a ser escritas como:

- $V_{(c)hi}$, que representa una variable indicadora de que el establecimiento agro i del estrato h es clasificado en la categoría "Verde" de sostenibilidad, de acuerdo con el criterio $c \in \{C_{\#1}, \dots, C_{\#11}\}$:

$$V_{(c)hi} = \begin{cases} 1, & \text{si el establecimiento agro "hi" es "Verde" por el criterio } c; \\ 0, & \text{si el establecimiento agro "hi" no es "Verde";} \end{cases}$$

y

- $A_{(c)hi}$ representa una variable indicadora de que el establecimiento agro i en el estrato h es clasificado en la categoría "Amarillo" de sostenibilidad, de acuerdo con el criterio $c \in \{C_{\#1}, \dots, C_{\#11}\}$:

$$A_{(c)hi} = \begin{cases} 1, & \text{si el establecimiento agro "hi" es "Amarillo" por el criterio } c; \\ 0, & \text{si el establecimiento agro "hi" no es "Amarillo".} \end{cases}$$

Bajo estas condiciones, los estimadores de interés se quedan así:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi} V_{(c)hi}}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}}{\pi_{hi}}}, \quad (41)$$

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi} A_{(c)hi}}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}}{\pi_{hi}}}, \quad (42)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi} (V_{(c)hi} + A_{(c)hi})}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}}{\pi_{hi}}}, \quad (43)$$

y

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}(1 - V_{(c)hi}A_{(c)hi})}{\pi_{hi}}}{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}}{\pi_{hi}}}. \quad (44)$$

5.3 Estimación usando un marco dual

Los diseños muestrales de marcos duales para encuestas agropecuarias por lo general utilizan simultáneamente un marco de áreas y un marco de lista. De ahí se mantenga la notación introducida anteriormente para cada tipo de marco. Una muestra probabilística es sacada, de forma independiente, de cada marco. En este escenario, tres tipos de uso del marco de lista incluyen:

- i. El marco identifica todos los N establecimientos agropecuarios de la población-objetivo y una muestra de n establecimientos agropecuarios ($n < N$) es sacada de este marco utilizando un muestreo probabilístico.
- ii. El marco identifica los más grandes establecimientos agropecuarios y todos ellos hacen parte de la muestra, es decir que un censo es aplicado al marco de lista.
- iii. El marco identifica los N_G más grandes establecimientos agropecuarios ($N_G < N$), y una muestra de tamaño n establecimientos ($n < N_G$) es sacada de ello usando un muestreo probabilístico.

Los escenarios i. y iii. tienen el mismo tratamiento estadístico en términos de las fórmulas de los estimadores. El escenario ii. no representa dificultad pues implica en el uso de las fórmulas para el marco de área, introducidas en 5.1, añadidas de la observación del marco de lista.

Con respecto al uso del marco de área, tres escenarios fueron considerados en esta guía: estimación a través de la estrategia de segmento abierto (5.1), estimación a través de la estrategia de segmento ponderado en una etapa (5.2) y estimación a través de la estrategia de segmento ponderado en dos etapas (5.3).

Para ejemplificar la derivación de las fórmulas de los estimadores bajo un diseño de marco dual, será utilizado el caso en que el marco de área usa la estrategia de segmento ponderado en dos etapas (5.3), y el marco de lista identifica los más grandes productores, como descrito en iii. Las demás situaciones pueden ser derivadas de manera análoga.

Dos estimadores para marcos duales son presentados aquí: el de multiplicidad simple y el “screening”.

5.3.1 Estimación por multiplicidad simple

Mecatti (2007) introdujo un estimador de marcos múltiples basado en un factor de multiplicidad simple. En esta guía, ese factor es representado por f_{hij}^A , tal que f_{hij}^A representa el número de marcos a los que un establecimiento agropecuario hij , identificado a través del marco de área, pertenece. En este caso,

$$f_{hij}^A = \begin{cases} 1, & \text{se el establecimiento agro } hij \text{ pertenece solamente al marco de área;} \\ 2, & \text{se el establecimiento agro } hij \text{ pertenece al marco de lista y de área.} \end{cases}$$

Cuando el establecimiento agropecuario es identificado a través del marco de lista, en esta guía el factor de multiplicidad es representado por f_{hi}^L , tal que f_{hi}^L representa el número de marcos a que un establecimiento agropecuario hi , identificado a través del marco de área, pertenece. En encuestas

agropecuarias, el marco de lista es inserto en el marco de área, así que el factor de multiplicidad asume el valor $f_{hi}^L = 2$.

Considere el uso de submuestreo de establecimientos agropecuarios, dentro de cada segmento del marco de área, utilizando P puntos. Considere también el uso de un factor de multiplicidad simple, asociado a los establecimientos agropecuarios dentro de cada estrato. Así es posible escribir $\hat{R}_{\#c}$ para un marco dual, aprovechando las fórmulas (34) y (41), con las debidas adaptaciones. La formula se presentaría como sigue:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{x_{hij} V_{(c)hij} I_{hijk}}{\pi_{hi} \pi_{j|hi} f_{hij}^A} + \sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi} V_{(c)hi}}{\pi_{hi} f_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{x_{hij} I_{hijk}}{\pi_{hi} \pi_{j|hi} f_{hij}^A} + \sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}}{\pi_{hi} f_{hi}^L}}. \quad (45)$$

La expresión (45) contiene un abuso de notación, pues trata al mismo tiempo de dos contextos distintos, uno del marco de área y otro del marco de lista, como es posible apreciar en la Figura 2.

Figura 2. Estimador de marco dual por tipo de marco.

$$\hat{R}_{\#c} = \frac{\overbrace{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{x_{hij} V_{(c)hij} I_{hijk}}{\pi_{hi} \pi_{j|hi} f_{hij}^A}}^{\text{Estimador aplicado al marco de área}} + \overbrace{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi} V_{(c)hi}}{\pi_{hi} f_{hi}^L}}^{\text{Estimador aplicado al marco de lista}}}{\underbrace{\sum_{h=1}^H \sum_{i \in S_h} \sum_{j \in U_{hi}} \sum_{k=1}^P \frac{x_{hij} I_{hijk}}{\pi_{hi} \pi_{j|hi} f_{hij}^A}}_{\text{Estimador aplicado al marco de área}} + \underbrace{\sum_{h=1}^H \sum_{i \in S_h} \frac{y_{hi}}{\pi_{hi} f_{hi}^L}}_{\text{Estimador aplicado al marco de lista}}}. \quad (45)$$

Fuente: Elaboración del autor.

Así, por ejemplo, en la parte del estimador de marco de área, un total de H estratos es utilizado, mientras que, en la parte del marco de lista, aunque se use la misma notación (H estratos), ese número es probablemente distinto. Para llegar a una expresión más precisa, la siguiente notación es introducida:

Notación usada en el marco de área

- $U_A = \{1, \dots, M\}$ representa el universo de M segmentos de área que forman el marco de área, tal que:
 - $U_A = \bigcup_{h=1}^H U_h^A$, es decir, el marco de área es estratificado en H estratos, con:
 - $U_h^A = \{1, \dots, M_h\}$ representando el conjunto de M_h segmentos de área ubicados en el estrato h , tal que $M = \sum_{h=1}^H M_h$ es el total de conglomerados en la población.
 - U_{hi}^A representa el conjunto de establecimientos agropecuarios que tienen "tracts" con en el segmento i , del estrato h .

- Una muestra de P establecimientos agropecuarios es seleccionada de U_{hi}^A , con reemplazo, y probabilidad de selección $\pi_{j|hi}^A$.
- x_{hij}^A representa la variable de superficie de área agrícola ponderada del establecimiento agropecuario j que tiene “tract” con el segmento de área i , en el estrato h .
- S_h^A representa una muestra probabilística de m_h segmentos, seleccionada dentro del estrato h del marco de área, tal que π_{hi}^A representa la probabilidad de selección del establecimiento agropecuario i en el estrato h del marco de área.

Notación usada en el marco de lista

- U_L representa el universo de los N_G más grandes establecimientos agropecuarios identificados en el marco de lista, tal que:
 - $U_L = \bigcup_{h=1}^{H'} U_h^L$, es decir, el marco de lista es estratificado en H' estratos, con U_h^L representando el conjunto de los N_{Gh} más grandes establecimientos agropecuarios del estrato h , tal que $N_G = \sum_{h=1}^H N_{Gh}$.
- y_{hi}^L representa la superficie de área agrícola del establecimiento agropecuario i del estrato h del marco de lista.
- S_h^L representa una muestra probabilística, de tamaño n_{Gh} , seleccionada dentro del estrato h del marco de lista, tal que π_{hi}^L representa la probabilidad de selección del establecimiento agro i en el estrato h del marco de lista.

La fórmula (45) puede ser reescrita más formalmente como:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{x_{hij}^A V_{(c)hij} I_{hijk}}{\pi_{hi}^A \pi_{j|hi}^A f_{hij}^A} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L V_{(c)hi}}{\pi_{hi}^L f_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{x_{hij}^A I_{hijk}}{\pi_{hi}^A \pi_{j|hi}^A f_{hij}^A} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{\pi_{hi}^L f_{hi}^L}}. \quad (46)$$

Recordando la expresión (33), es posible todavía escribir:

$$x_{hij}^A = \frac{T_{hij}}{A_{hij}} y_{hij}^A,$$

donde:

y_{hij}^A representa la superficie de área agrícola del establecimiento agropecuario j que tiene un “tract” con el segmento i del estrato h , en el marco de área; y T_{hij} y A_{hij} tienen la misma definición anteriormente introducida, es decir:

- T_{hij} representa el área del “tract” del establecimiento agropecuario hij ; y
- A_{hij} representa el área total del establecimiento agropecuario hij .

Así que la expresión (46) puede ser finalmente reescrita como:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{T_{hij} y_{hi}^A V_{(c)hij} I_{hijk}}{A_{hij} \pi_{hi}^A \pi_{j|hi}^A f_{hij}^A} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L V_{(c)hi}}{\pi_{hi}^L f_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{T_{hij} y_{hi}^A I_{hijk}}{A_{hij} \pi_{hi}^A \pi_{j|hi}^A f_{hij}^A} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{\pi_{hi}^L f_{hi}^L}}. \quad (47)$$

Considerando la situación en que, dentro de un estrato h , un muestreo aleatorio simple es empleado para selección de segmentos de área en el marco de área, y un muestreo con probabilidades distintas es usado para selección de establecimientos agropecuarios en el marco de lista, y, además, utilizando el valor de $f_{hi}^L = 2$, la expresión (47) queda como:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A V_{(c)hij}}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L V_{(c)hi}}{2 \pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{2 \pi_{hi}^L}}. \quad (48)$$

Los demás estimadores son:

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A A_{(c)hij}}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L A_{(c)hi}}{2 \pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{2 \pi_{hi}^L}}. \quad (49)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A (V_{(c)hij} + A_{(c)hij})}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L (V_{(c)hi} + A_{(c)hi})}{2 \pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{2 \pi_{hi}^L}}. \quad (50)$$

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A (1 - V_{(c)hij} A_{(c)hij})}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L (1 - V_{(c)hi} A_{(c)hi})}{2 \pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hi}^A}{P m_h A_{hij} f_{hij}^A} I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{2 \pi_{hi}^L}}. \quad (51)$$

5.3.2 Screening estimator

El "screening estimator" es una composición de informaciones de los dos marcos, de área y de lista, en que informaciones referentes a grandes productores son computadas a través de la muestra del marco de lista, mientras las informaciones referentes a los demás productores son computadas a través del marco de áreas.

Para aprovechar el raciocinio en el desarrollo de las fórmulas (48), (49), (50) y (51), considere la siguiente notación:

- $L_{hij} = \begin{cases} 1, & \text{es el establecimiento } j \text{ con "tract" en el segmento } i \text{ en el estrato } h \text{ que} \\ & \text{pertenece al marco de lista; y} \\ 0, & \text{en caso contrario.} \end{cases}$

Las fórmulas referentes al "screening estimator" son:

$$\hat{R}_{\#c} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A V_{(c)hij}}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L V_{(c)hi}}{\pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{\pi_{hi}^L}}. \quad (52)$$

$$\hat{R}_{\#c1} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A A_{(c)hij}}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L A_{(c)hi}}{\pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{\pi_{hi}^L}}. \quad (53)$$

$$\hat{R}_{\#c2} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A (V_{(c)hij} + A_{(c)hij})}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L (V_{(c)hi} + A_{(c)hi})}{2 \pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{2 \pi_{hi}^L}}. \quad (54)$$

$$\hat{R}_{\#c3} = \frac{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A (1 - V_{(c)hij} A_{(c)hij})}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L (1 - V_{(c)hi} A_{(c)hi})}{2 \pi_{hi}^L}}{\sum_{h=1}^H \sum_{i \in S_h^A} \sum_{j \in U_{hi}^A} \sum_{k=1}^P \frac{M_h T_{hi} y_{hij}^A}{P m_h A_{hij} f_{hij}^A} (1 - L_{hij}) I_{hijk} + \sum_{h=1}^{H'} \sum_{i \in S_h^L} \frac{y_{hi}^L}{2 \pi_{hi}^L}}. \quad (55)$$

6 Estimación para dominios subnacionales

El proceso de estimación del indicador 2.4.1 para todo un territorio nacional incluye la recopilación de una gran variedad de datos, toda vez que depende de una evaluación de cada uno de los 11 subindicadores. Eso tiene, por supuesto, un costo, que puede ser traducido en términos del tamaño de la muestra. Cuando los dominios de interés subnacionales coinciden con los dominios de interés previstos en la encuesta agropecuaria, se espera que los tamaños de muestra sean suficientemente grandes para generar estimaciones del indicador 2.4.1 para cada uno de ellos. Este es el caso, por ejemplo, de encuestas que usan un factor de definición del estrato relativo al dominio subnacional de interés. Sin embargo, planear una encuesta con factores de estratificación coincidentes con niveles de desagregación más avanzados, puede resultar en costos financieros prohibitivos. Por lo tanto, en muchas situaciones, estimaciones pueden ser generadas para dominios que no coinciden con factores de estratificación previamente planeados. Precisamente por eso, la calidad de la inferencia depende del tamaño de muestra observado en cada dominio.

Para definir el indicador 2.4.1 como parámetro en un dominio subnacional de interés, se introduce la siguiente notación:

- $U_d = \{1, \dots, N_d\}$ representa el universo de N_d establecimientos agropecuarios que pertenecen al dominio subnacional d , identificados en una estructura poblacional en la cual:
 - $U_d = \bigcup_{h=1}^H (U_h \cap U_{dh})$, es decir, el factor que define el dominio d es cruzado con el factor que define el estrato, con:
 - U_h representando el conjunto de N_h establecimientos agropecuarios del estrato h .
 - U_{dh} representa el conjunto de N_{dh} establecimientos agropecuarios del estrato h , que pertenecen al dominio subnacional d .
- y_{hi} representa la superficie de área agrícola del establecimiento agropecuario i en el estrato h .

El indicador 2.4.1 tiene la siguiente forma en el dominio subnacional de interés d :

$$R_{241d} = \text{mín}(R_{\#1d}, \dots, R_{\#11d}), \quad (56)$$

donde cada uno de los subindicadores $R_{\#cd}$: $c \in \{1, \dots, 11\}$ corresponden a una expresión del tipo:

$$R_{\#cd} = \frac{TSAS\#cd}{TSA_d}, \quad (57)$$

donde:

$$TSAS\#cd = \sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi} V_{(c)hi}, \quad (58)$$

y

$$TSA_d = \sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi}. \quad (59)$$

- $TSA_{\#c_d}$ representa el total de superficie de área agrícola del dominio subnacional d en que se practica una agricultura considerada productiva y sostenible de acuerdo con el criterio $C_{\#i}$ usado por el subindicador $R_{\#c_d}$; y
- TSA_d representa el total de superficie de área agrícola del dominio subnacional d .

El estimador (56) representa la forma paramétrica del subindicador $R_{\#c_d}$, referente al dominio subnacional d , en función de la categoría “Verde” de sostenibilidad. Es decir, en este caso $R_{\#c_d}$ representa la proporción de área agrícola del dominio subnacional d , con **nivel deseable de sostenibilidad** para el criterio $C_{\#c}$. En paralelo, los parámetros poblacionales

$$R_{\#c_d} 1 = \frac{\sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi} A_{(c)hi}}{\sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi}}, \quad (60)$$

$$R_{\#c_d} 2 = \frac{\sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi} (V_{(c)hi} + A_{(c)hi})}{\sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi}}, \quad (61)$$

y

$$R_{\#c_d} 3 = \frac{\sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi} (1 - V_{(c)hi} A_{(c)hi})}{\sum_{h=1}^H \sum_{i \in U_{dh}} y_{hi}} \quad (62)$$

representan la proporción de área agrícola de un mismo subindicador para el mismo dominio subnacional d de interés, usando categorías distintas de sostenibilidad:

- $R_{\#c_d} 1$ es la proporción de área agrícola subnacional con **nivel aceptable de sostenibilidad** (Amarillo) para el criterio $C_{\#c}$.
- $R_{\#c_d} 2$ es la proporción de área agrícola subnacional con **nivel aceptable o deseable de sostenibilidad** (Amarillo o Verde) para el criterio $C_{\#c}$.
- $R_{\#c_d} 3$ es la proporción de área agrícola subnacional con **nivel de insostenibilidad** (Rojo) para el criterio $C_{\#c}$.

El desarrollo de estimadores subnacionales en todas las situaciones de esta guía sigue pasos análogos al que fue descrito para la estimación del indicador para todo el territorio nacional.

7 Conclusión

Esta guía presentó formas de estimar el indicador 2.4.1 de sostenibilidad de agricultura, considerando varios aspectos estadísticos del diseño de la muestra.

Como se trata de una primera versión, hay una serie de elementos que pueden y deben ser mejorados en versiones sucesivas. Por ejemplo, se espera poder incluir más referencias a documentos que puedan ayudar con conceptos relativos al tema. Asimismo, se espera también avanzar en la construcción de un documento más completo que incluya los esfuerzos de demostración de uso de los cuestionarios para recopilación de datos, profundizando en las experiencias de Colombia, Costa Rica y el Perú. Además, dos otros elementos que merecen ser discutidos son el problema de estimación en dominios subnacionales considerados pequeños, y el problema de estimación de una medida de calidad de la estimación del indicador.

Al respecto del problema de la medida de calidad del indicador es importante una discusión acerca de la estimación de la varianza, cuyo problema no es trivial, sino además sobre alternativas de medidas de calidad que puedan tener una interpretación menos limitada para los países. El uso del rango entre los subindicadores, por ejemplo, ¿sería una medida más interesante? También se podrían explorar otras posibilidades.

Referencias

- Candia, A.** 2022. *Análisis de datos e implementación del cálculo del indicador 2.4.1 en Perú*. Documento interno Informe de consultoría proyecto para el cálculo de los indicadores de ODS. Sucre.
- Candia, A.** 2023. *Apoyo a los sistemas estadísticos nacionales para el fortalecimiento de los indicadores ODS 2.4.1 en Costa Rica*. Documento interno Informe de consultoría proyecto para el cálculo de los indicadores de ODS. Sucre.
- FAO.** 2018. *AGRIS Guía para la Encuesta Agrícola Integrada*. Roma. <https://www.fao.org/3/ca6412es/ca6412es.pdf>
- FAO.** 2021. *Sampling guidance for SDG Indicator 2.4.1*. Roma. <https://www.fao.org/3/ca7439EN/ca7439EN.pdf>
- FAO.** 2023a. *SDG indicator metadata 2.4.1*. Roma. <https://unstats.un.org/sdgs/metadata/files/Metadata-02-04-01.pdf>
- FAO.** 2023b. *Proporción de la superficie agrícola en que se practica una agricultura productiva y sostenible*. Nota metodológica, Revisión 11. Roma. <https://www.fao.org/3/ca7154es/ca7154es.pdf>
- Mecatti, F.** 2007. A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, 33(2): 151–158.
- Nova, D.** 2021. *Ejercicio piloto para la medición del indicador 2.4.1 a partir de la Encuesta Nacional Agropecuaria (ENA) en Colombia*. Documento interno Informe de consultoría proyecto para el cálculo de los indicadores de ODS. Bogotá.

Contacto:

División de Estadística – Desarrollo económico y social

FAO-statistics@fao.org

www.fao.org/food-agriculture-statistics/resources/publications/working-papers/en/

Organización de las Naciones Unidas para la Alimentación y la Agricultura

Roma, Italia

ISBN 978-92-5-138602-6



9 789251 386026

CC9550ES/1/03.24