

Tamaño de Muestra en Poblaciones Agropecuarias

Lic. Jaime T. Pinto Ajhuacho
✉ titojaime_pinto@yahoo.com

1. Población, muestra y marco de muestreo

Población, denominaremos a un conjunto de unidades del que se desea obtener cierta información, limitándose las observaciones a una parte que denominamos, muestra. Los datos obtenidos a partir de ella nos permiten inferir unos valores aproximados de la población en su totalidad, a estos valores aproximados se les denomina “estimaciones” las cuales vendrán afectadas de un error que denominaremos “error debido al muestreo”.

La estimación, considera un “Marco de muestreo” que es el conjunto de información que puede ser útil en cualquier momento del diseño de la encuesta, en el que debemos distinguir el listado y la información complementaria.

El listado, es la información de la lista de unidades a muestrear que también se conoce con el nombre de “marco” en un sentido restringido, es una de las operaciones más delicadas en el diseño muestral.

Un marco por lista, como su nombre lo indica, es una lista de los elementos de la población, junto con los datos de identificación adecuados, como parte del marco, puede incluirse información suplementaria que permita la clasificación de los elementos.

Un marco por lista que consista de los nombres y direcciones de agricultores se adapta muy bien para la recolección de informes agrícolas, el bajo costo de la recolección de datos por este medio constituye una de las principales ventajas; una desventaja es que hay dificultad para ajustar la lista con los cambios que ocurre en el universo o población.

2. Estratificación y submuestreo

Con la estratificación o división de la población en subpoblaciones o estratos, obteniéndose posteriormente muestras independientes de cada estrato, pueden perseguirse diversos fines, como por ejemplo.

- Dar estimaciones separadas para ciertas subpoblaciones.
- Agrupar unidades de muestreo homogéneas entre si en estratos, con objeto de mejorar la precisión de las estimaciones globales.
- Utilizar métodos diferentes de muestreo en los distintos estratos (según el marco de que se disponga).

En la estratificación se da lugar a la llamada “Afijación óptima”, con este criterio se deberán muestrear con más intensidad los estratos que presentan mayor variabilidad interna, menor costo o un tamaño superior al medio. También en este caso existen expresiones matemáticas que determinan el tamaño óptimo de la muestra en cada estrato.

El “Muestreo estratificado”, es un recurso empleado para reducir el error en el muestreo comprende la subdivisión del universo entre un número de estratificaciones o niveles y posteriormente asignar a cada estrato una porción de la muestra.

El “Submuestreo” ó muestreo en dos etapas, consiste en agrupar primero las unidades individuales de muestreo en conglomerados de unidades primarias de muestreo (a veces, muy grandes), obteniendo una muestra de conglomerados y posteriormente sacando una muestra de las unidades individuales de cada conglomerado muestral.

Muchas encuestas sobre unidades agropecuarias que se hacen por medio de submuestreo, han usado a los segmentos agropecuarios como unidades de muestreo. En la primera etapa se selecciona una muestra de segmentos (Conglomerado de Unidades Agropecuarias) y en la segunda etapa se seleccionan las muestras de las unidades agropecuarias en estos segmentos seleccionados, donde un segmento es un conglomerado de unidades agropecuarias.

Es así que el país en el sector agropecuario se divide en áreas que son unidades primarias de muestreo (UPM) usando mapas de escala pequeña para este propósito.

Respecto de la segmentación, la mayoría de los diseños de muestra requieren que las UPM (Unidades Primarias de muestreo) sean divididas en unidades más pequeñas (sectores) con límites claramente definidos; éstas, a su vez, suelen dividirse en segmentos y se escogen uno o más segmentos para la muestra.

Hablando de la delimitación del segmento de muestra, uno de los usos principales de un mapa es el de indicar los límites del área terrestre que debe abarcar la muestra. Si el diseño de muestra se basa en datos para un área definida con exactitud, entonces el mapa debe indicar claramente los límites de dicha área.

3. Cobertura

La población objetivo de estudio está constituida por todos los productores agropecuarios del país, agrupados en cada uno de los nueve departamentos.

4. La unidad de investigación

Es la “Unidad Agropecuaria del Productor”.

5. Marco muestral

La construcción del marco muestral, utilizado en la Encuesta Nacional Agropecuaria, es realizada a partir de una revisión y adecuación de los marcos existentes en el Instituto Nacional de Estadística, a fin de obtener el marco estratificado. Los insumos utilizados en la generación del Marco Estratificado, son:

- Marco elaborado en el Censo Agropecuario.
- Estratificación del país, en Zonas de Producción Agropecuaria a nivel de cantones.
- Cartografía a nivel cantonal y de segmento.
- Tipo de actividad Agrícola o Pecuario.

El marco censal agropecuario, tiene la división político administrativa del país, departamentos particionados en provincias y cantones; luego las unidades de desagregación estadísticas como los sectores y segmentos agropecuarios, este último definido como un conglomerado en promedio 100 unidades agropecuarias en la zona del altiplano y valles, 70 unidades agropecuarias en las zonas tropicales.

Los subuniversos armados, se considera conformados por la intersección de Zonas de producción agrícola o Zonas de producción pecuaria, denominados “Dominios de Estudios” (Ver esquema 1).

Esquema 1 ➤ Subuniverso (Dominio de estudio)

Zona 1	Zona 2	...	Zona <i>i</i>	...	Zona <i>N</i>
uso agrícola	uso agrícola	...	uso agrícola	...	uso agrícola
uso pecuario	uso pecuario	...	uso pecuario	...	uso pecuario

Zona *i*

uso agrícola

uso pecuario

↪

↪

: Departamento

: Zona de producción

: Dominio de estudio

A nivel nacional, el número de segmentos por tipo de actividad (agrícola o pecuario), según zonas de producción y departamento.

Para determinar el subuniverso de tipo de actividad agrícola, se considero la superficie cultivada y para los de uso pecuario se convertía el número de especies ganaderas a “unidades animal”.

El Listado de segmentos agropecuarios de los subuniversos, ya sea Agrícola o Pecuario, tiene las siguientes características:

Cuadro 1 ➤ Lista de segmentos agropecuarios. Subuniverso: xxxxx

Nro. de segmento	Identificadores geográficos	Nro. unidades de observación	Superficie agrícola	Existencia ganadera
1	—	A	—	—
2	—	B	—	—
3	—	C	—	—
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—
<i>N</i>	—	—	—	—

El Marco Muestral está conformado por Identificadores geográficos y Variables Agropecuarias.

6. Identificadores geográficos

Departamento – Provincia – Zona de Producción – Cantón –Sector – Segmento.

7. Variables agropecuarias

SUPERFICIE AGRÍCOLA: Superficie Cultivada – Superficie de Pastos Naturales – Superficie de Pastos Cultivados – Superficie Resto – Superficie Total.

EXISTENCIA GANADERA: Total de Bobinos – Total de Ovinos – Total de Porcinos – Total Camelidos – Total de Equinos – Total aves de corral.

8. Estratificación

En cada subuniverso, tomando en cuenta su participación pequeña, media o grande en la actividad agrícola o pecuaria, los estratos son definidos:

ESTRATO 1: Segmentos con participación agropecuaria pequeña.

ESTRATO 2: Segmentos con participación agropecuaria media.

ESTRATO 3: Segmentos con participación agropecuaria grande.

9. Tamaño de la muestra

A nivel del subuniverso y tomando en cuenta el muestreo estratificado con asignación optima en el tamaño de la muestra se utiliza la siguiente expresión:

$$n = \frac{\left(\sum_{h=1}^L W_h S_h\right)^2}{\left(\frac{e}{t}\right)^2 + \sum_{h=1}^L \frac{N_h}{N^2} S_h^2}, \quad h = 1, 2, 3, \dots, L. \quad (1)$$

$e = e'\bar{Y}$. En la que

n =Tamaño de la muestra (en segmentos)

e =Error respecto al promedio.

e' =Error (%) de medición relativa.

t_α =Nivel de significación para un α dado

$W_h = N_h/N$: Ponderación del estrato h

$S_h^2 = \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2 / (N_h - 1)$: Varianza de la variable “Superficie Cultivada” o “Número abstracto de Cabezas”, por segmento del h -ésimo estrato.

N_h =Total de segmentos del estrato h .

N =Total de segmentos del Subuniverso.

L =Número total de estratos.

Y_{ih} =Total de la “Superficie Cultivada” o “Numero abstracto de Cabezas” del i -ésimo segmento en el h -ésimo estrato. $i = 1, \dots, N_h$. $h = 1, \dots, L$.

$n_h = n N_h S_h / (\sum_{h=1}^L N_h S_h)$: Tamaño de la submuestra por estrato.

10. Procedimiento del cálculo de tamaño de muestra

En el diseño muestral, se utiliza las fórmulas que corresponden al muestreo estratificado con asignación optima; diseñándose una hoja electrónica para generalizar el cálculo del tamaño de la muestra para los distintos subuniversos.

$N = \sum_{h=1}^L N_h$: Total de elementos de la población.

$W_h = N_h/N$: Ponderación por estrato

$Y = \sum_{h=1}^L Y_h$: Total de la variable Y

$Y = \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{ih}$: Total de la variable Y por elemento y por estrato.

$h = 1, 2, \dots, L$: Estratos.

$i = 1, 2, \dots, N_h$: Elementos.

$\bar{Y} = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{ih} = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} N_h \bar{Y}_h$: Promedio general o poblacional de Y .

$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2$: Varianza

$\sigma^2 = \frac{1}{N} \sum_{h=1}^L \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2$: Desviación

$S_h^2 = \frac{1}{N-1} \sum_{i=1}^{N_h} (Y_{ih} - \bar{Y}_h)^2$: Cuasivarianza

$S_b^2 = \frac{1}{N-1} \sum_{h=1}^L N_h (\bar{Y}_h - \bar{Y})^2$: Intervarianza (Entre)

$S_w^2 = \frac{1}{N-1} \sum_{h=1}^L S_h^2 (N_h - 1)$: Intravarianza (Dentro).

11. Ejemplo

Para la Zona de Producción Llanos de Chaco, uso agrícola, departamento de Chuquisaca, el cálculo de parámetros y tamaño de muestra es:

Departamento: Chuquisaca
Zona de Producción: Llanos del Chaco –110
Uso: Agrícola

Muestreo Estratificado
Tipo de Afijación: Optima

Población Agropecuaria

Detalle	Descrip	Estrato 1		Estrato 2		Estrato 3		Totales
1	Y_1	59	196	500	331	620		—
2	Y_2	66	199	502	333	657		—
3	Y_3	77	200	513	340	660		—
4	Y_4	104	201	519	343	997		—
5	Y_5	109	210	401	373	1273		—
6	Y_6	109	211	403	379	2113		—
7	Y_7	119	212	405	389	542		—
8	Y_8	125	213	406	393	543		—
9	Y_9	135	216	410	328	548		—
10	Y_{10}	142	222	438	327	551		—
11	Y_{11}	148	224	441		560		—
12	Y_{12}	154	231	444				—
13	Y_{13}	157	236	485				—
14	Y_{14}	160	250	306				—
15	Y_{15}	165	255	307				—
16	Y_{16}	165	275	307				—
17	Y_{17}	172	284	310				—
18	Y_{18}	174	288	323				—
19	Y_{19}	177	299	326				—
20	Y_{20}	193	305	331				—
Número total de segmentos		N_h	40.00	30.00	11.00			81.00
Ponderación		W_h	0.49	0.37	0.14			1.00
Total Variable Y		Y_h	7437.00	11613.00	9064.00			28114.00
Promedio por estratos		$\bar{Y}_h$	185.93	387.10	824.00			347.09
Varianza σ^2		σ^2	3856.30	4438.30	215303.40			77093.91
Desviación σ		σ	62.10	66.60	464.01			277.66
Cuasivarianza S^2		S^2	3955.20	4591.40	236833.80			78057.58
Desviación S		S	62.89	67.76	486.66			279.39
Intervarianza (Entre)		S_b^2	—	—	—	55.47 % (44860.81)		
Intravarianza (Dentro)		S_w^2	—	—	—	44.53 % (33196.77)		
Coef. de var. (%)		CV_h	33.40 %	17.21 %	56.31 %			80.00 %

El cálculo de tamaño de la muestra, para un error del 5 % y un nivel de confianza del 95 %, es:

$$n = \frac{\left(\sum_{h=1}^L W_h S_h\right)^2}{\left(\frac{e}{t}\right)^2 + \sum_{h=1}^L \frac{N_h}{N^2} S_h^2} \quad (2)$$

Tomando en cuenta un error del 5 %.

$$e' = 0.05 \quad e = e' \bar{Y} = 0.05(347.09) = 17.3545 \quad n = 29.54 \cong 29 \quad (3)$$

Distribución de la muestra por estrato:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$$

$$n_1 = 29 \frac{2515.6}{9901.66} = 7 \quad n_2 = 29 \frac{2032.8}{9901.66} = 6 \quad n_3 = 29 \frac{5353.26}{9901.66} = 16 \quad (4)$$

Calculando tamaños de muestra para distintos errores, tendremos:

Departamento: Chuquisaca
Zona de Producción: Llanos del Chaco –110
Uso: Agrícola

Muestreo Estratificado
Tipo de Afijación: Optima

Tamaños de muestra para diferentes errores
($t = 1.96$ para un nivel de confianza del 95 %.)

Error (%)	n_1	n_2	n_3	n
1	9	7	18	34
2	8	7	18	33
3	8	7	17	32
4	8	6	17	31
5	7	6	16	29
6	7	5	15	27
7	7	5	14	26
8	6	5	13	24
9	6	4	12	22
10	5	4	11	20