Muestreo Estadístico

Muestreo Sistemático

Systematic Sampling (SS)

Andy Domínguez

www.andydominguez.weebly.com

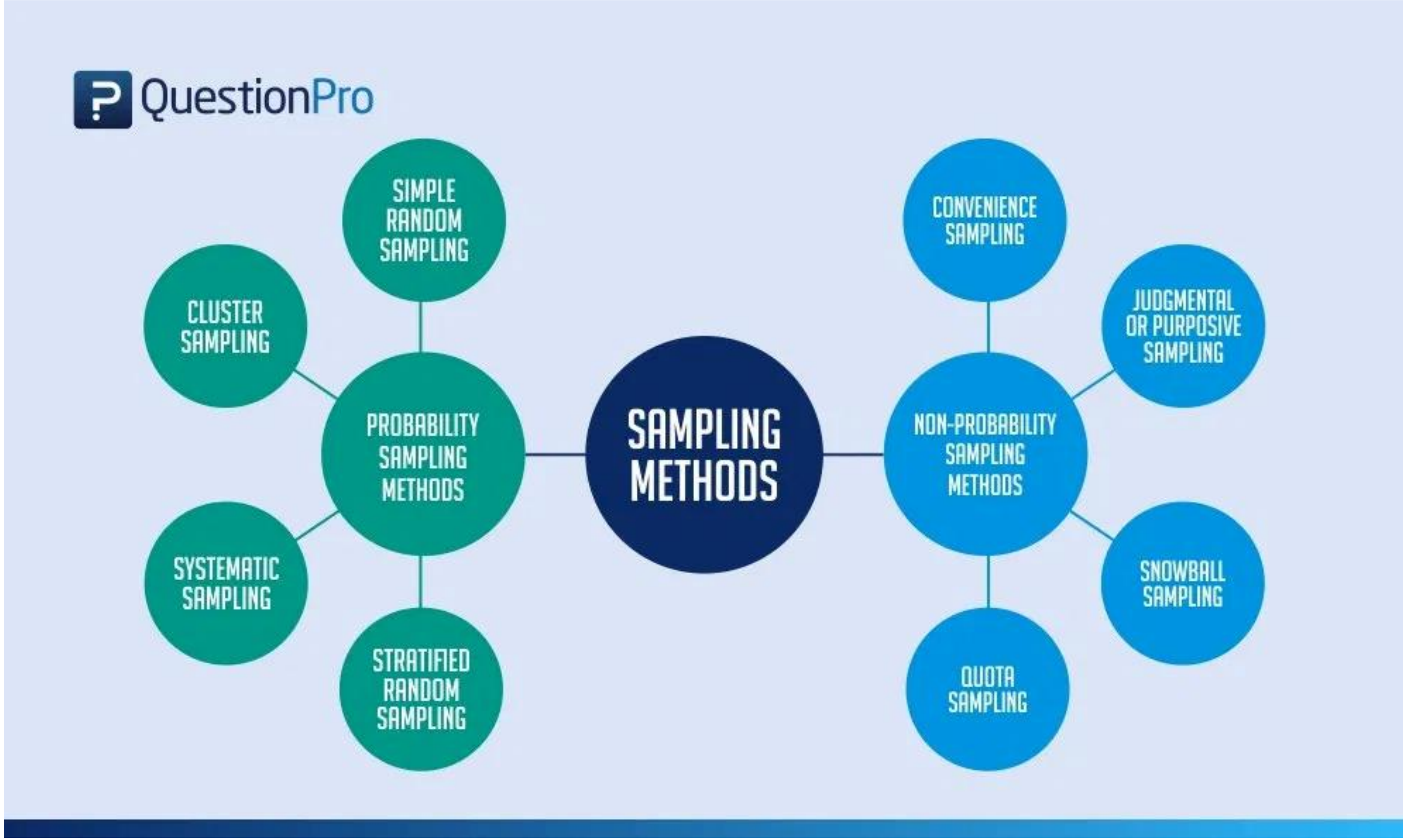
adominguez@utb.edu.co

Jul 2025

FACULTAD DE CIENCIAS BÁSICAS

Especialización en Estadística Aplicada y Ciencia de Datos

Tipos de Muestreo



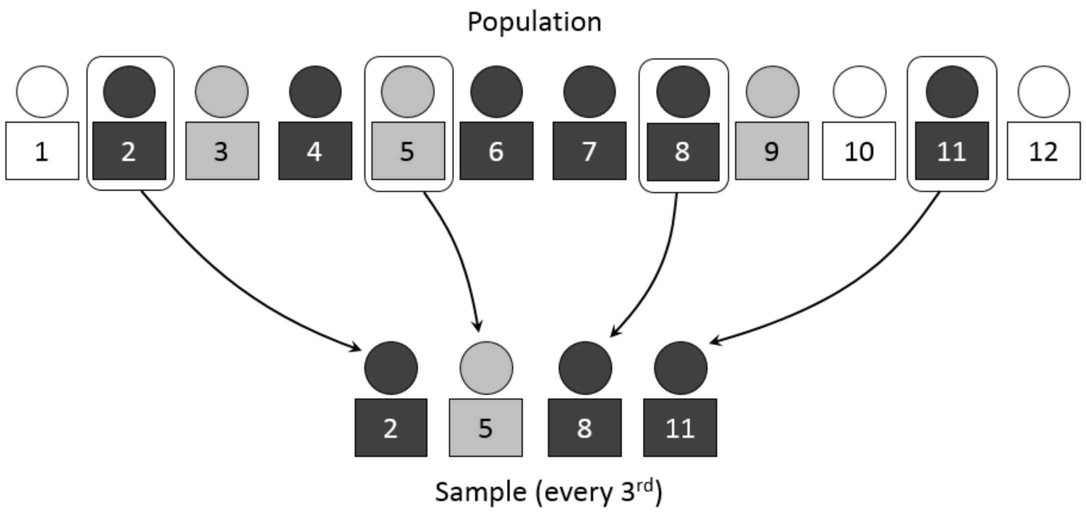
Muestreo Sistemático

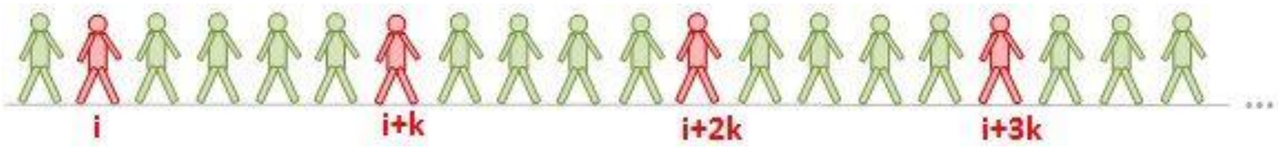
Se inicia con la escogéncia aleatoria de un punto de una lista de miembros de la población. Esta unidad y las k-ésimas siguientes son muestreadas.

1- Requiere que la población de *N* individuos se encuentra ordenados del 1 al *N*. Se define *k* entero más próximo a *N/n*. k es llamado el intervalo sistemático

2- Seleccione un número aleatorio de 1 a k, y ese será el primer miembro de la muestral.

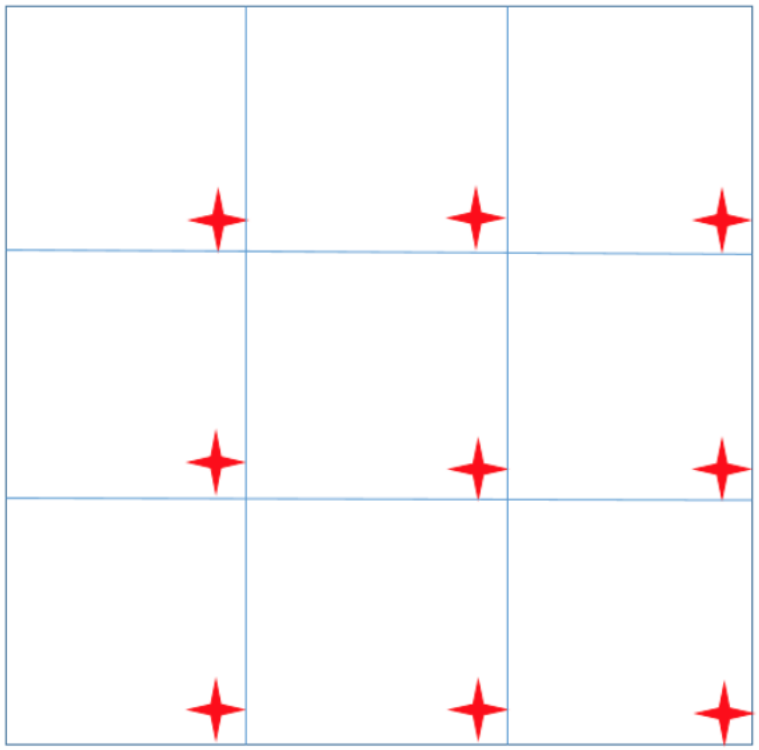
3- Seleccione los siguientes miembros de la muestra en el ordenamiento en el listado de posición i+k-ésima.





Muestreo Sistemático

Se inicia con la escogéncia aleatoria de un punto de una lista de miembros de la población. Esta unidad y las k-ésimas siguientes son muestreadas.



Muestreo Espacial

Muestreo Sistemático Espacial

Es sencillo de implementar y brinda un número suficiente de observaciones separadas con distancias y direcciones bien definidas

Muestreo Sistemático

Ventajas

Fácil y simple de implementar. Menos costosos.

Puede garantizar mayor representatividad que el MAS

El método de **muestreo sistemático** es más potente que el muestreo simple cuando el orden de los datos influye en que los sujetos próximos son semejantes.

Evita la necesidad de generar tantos números aleatorios como individuos en la muestra.

Desventajas

Desventaja: el orden en que se han listado los candidatos a la muestra podría tener algún tipo de periodicidad oculta que coincida con el intervalo escogido para generar la muestra sistemática.

Supone que se conoce el tamaño de la población o que se puede determinar

No se considera estrictamente aleatorio (sólo el inicio del primer elemento)

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Estimación de la media y el total de la muestra sistemática

Una vez obtenida la muestra sistemática, el objetivo será caracterizar la población por medio de una muestra estimando los parámetros de mayor interés, como la media y el total poblacional.

Estimador para la Media Poblacional

$$\hat{\mu} = \bar{y}_s = \frac{\sum_{i=1}^{n} y_i}{n}$$

Estimador para el Total Poblacional

$$\hat{t}_s = N\bar{y}_s$$

Estimadores correspondientes a las varianzas de la media y del total.

$$\hat{V}(\bar{y}_s) = \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$$

$$V(\hat{t}_s) = N^2\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$$

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Los intervalos de confianza para la media y para el total

Una vez obtenida la muestra, el objetivo será caracterizar la población por medio de una muestra estimando los parámetros de mayor interés, como la media y el total poblacional.

En R ($Px \leq a$):
dt(x, df, lower.tail = T)

En Python ($Px \leq a$) :
t.cdf(q, df)

Intervalos para la Media Poblacional

$$\bar{y}_s \pm t_{(n-1,1-\alpha/2)} \sqrt{\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)}$$

donde $\bar{y}_s$ es la media de la muestra sistemática.

RECORDAR-...factor de corrección *para poblaciones finitas (N − n)/N puede ignorarse para valores de n ≤ N/20*

Intervalo para el Total Poblacional

$$\hat{t}_s \pm t_{(n-1,1-\alpha/2)} \sqrt{N^2\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)}$$

Obsérvese que la estimación de la varianza es la misma que la presentada en el muestreo aleatorio aleatorio

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Tamaño de la Muestra Para Estimar la Media

A fin de determinar el tamaño de la muestra para estimar a μ, se procede como en el Muestreo aleatorio Simple. Primero se elige un valor de $\boldsymbol{d}$, es decir, la precisión(ancho del intervalo) que se está dispuesto aceptar en las estimaciones

$$d = t_{(n-1,1-\alpha/2)}\sqrt{V(\bar{y}_s)} \qquad \hat{V}(\bar{y}_s) = \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$$

El tamaño de muestra para estimar la media vendría por(despejando a *n* en la formula anterior):

$$\boxed{n^* = \frac{N t^2_{(n-1,1-\alpha/2)}\sigma^2}{N d^2 + t^2_{(n-1,1-\alpha/2)}\sigma^2}}$$

donde la varianza poblacional σ^2se puede sustituir por la muestral S^2.

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

$$k = \frac{N}{n} = \frac{1{,}000}{10} = 100.$$

Esto quiere decir que se debe muestrear cada 100 envases de leche, eligiendo aleatoriamente el primer elemento entre los primeros 100

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	Porcentaje de grasa
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

a) Promedio de grasa por envase.

b) La varianza muestral (S_2).

c) La varianza y la desviación estándar de la media muestral(Error muestral).

d) IC del promedio de grasa por envase de leche con una confiabilidad de 95%.

e) Cantidad total de grasa que se encuentra en los envases.

f) IC para el total de grasa por envase.

g) El tamaño de muestra necesario para estimar el promedio de grasa por envase, con una precisión de 0.05% de grasa por envase y una confiabilidad de 95%

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	**Porcentaje de grasa**
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

a) Promedio de grasa por envase.

$$\bar{y}_s = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\bar{y}_s = \frac{2.5 + 2.6 + 2.7 + 2.6 + 2.8 + 2.9 + 3 + 2.6 + 2.7 + 2.8}{10}$$

$\bar{y}_s$ = 2.72 o 2.72% de grasa por envase

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	**Porcentaje de grasa**
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

b) La varianza muestral (S_2).

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$S^2 = 0.024$$

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	Porcentaje de grasa
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

c) La varianza y la desviación estándar de la media muestral(Error muestral).

$$\hat{V}(\bar{y}_s) = \left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)$$

$$S^2_{\bar{y}_s} = \left(\frac{1{,}000-10}{1{,}000}\right)\left(\frac{0.024}{10}\right) = 0.002376$$

$$S_{\bar{y}_s} = \sqrt{S^2_{\bar{y}_s}} = \sqrt{0.002376} = 0.048744.$$

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	**Porcentaje de grasa**
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

d) IC del promedio de grasa por envase de leche con una confiabilidad de 95%.

$$\bar{y}_s \pm t_{(n-1,1-\alpha/2)}\sqrt{\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)}$$

donde: $\bar{y}_s = 2.72,\ S_{\bar{y}_s} = 0.048744$ y $t_{(n-1,1-\alpha/2)} = t_{(10-1,1-0.025)} = 2.2622.$

$$2.72 \pm (2.2622)(0.048744)$$

$$2.72 \pm 0.110269$$

$$2.609731 \leq \mu \leq 2.830269.$$

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	Porcentaje de grasa
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

En R ($Px \leq a$):
dt(x, df, lower.tail = T)

En Python ($Px \leq a$) :
t.cdf(q, df)

https://homepage.divms.uiowa.edu/~mbognar/applets/t.html

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

e) Cantidad total de grasa que se encuentra en los envases.

$$\hat{\tau} = N\bar{y}_s = (1{,}000)(2.72) = 2{,}720 \text{ gramos de grasa.}$$

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	Porcentaje de grasa
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

f) IC para el total de grasa por envase.

$$\hat{\tau}_s \pm t_{(n-1,1-\alpha/2)}\sqrt{N^2\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)}$$

donde: $\hat{\tau} = 2{,}720$, $N = 1{,}000$, $S_{\bar{y}_s} = 0.048744$ y $t_{(n-1,1-\alpha/2)} = t_{(10-1,1-0.025)} = 2.2622$.

$$2{,}720 \pm (1{,}000)(2.2622)(0.048744)$$

$$2{,}720 \pm 110.268677$$

$$2{,}609.7313 \leq \tau \leq 2{,}830.2687$$

En R:
dt(x, df, lower.tail = T)

https://homepage.divms.uiowa.edu/~mbognar/applets/t.html

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	Porcentaje de grasa
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

Estimación de parámetros de interés para promedio bajo muestreo sistemático

Ejemplo:

Una línea de producción de leche ultrapasteurizada elabora N = 1,000 envases por hora (cada envase contiene un litro de leche). Se desea saber si cada envase de leche cumple con el porcentaje de grasa. Para ello se toma una muestra sistemática de 10 envases. Primero se elige k

Hallar:

Los Valores de porcentaje de grasa por muestra en la tabla:

No. de muestra	Porcentaje de grasa
80	2.5
180	2.6
280	2.7
380	2.6
480	2.8
580	2.9
680	3.0
780	2.6
880	2.7
980	2.8

g) El tamaño de muestra necesario para estimar el promedio de grasa por envase, con una precisión de 0.05% de grasa por envase y una confiabilidad de 95%

$$n^* = \frac{N t^2_{(n-1,1-\alpha/2)} \sigma^2}{N d^2 + t^2_{(n-1,1-\alpha/2)} \sigma^2}$$

donde: $N = 1{,}000$, $t_{(n-1,1-\alpha/2)} = t_{(10-1,1-0.025)} = 2.2622$, $S_s^2 = 0.024$ y $d = 0.05$.

$$n^* = \frac{(1{,}000)(2.2622)^2(0.024)}{(1{,}000)(0.05)^2 + (2.2622)^2(0.024)}$$

$$n^* = 46.82 = 47 \text{ envases}$$

Muestreo Sistemático

Implementación en Python

EJERCICIO PARA ESTIMACIÓN MEDIA

Deseamos realiar un estudio sobre el consumo de bebidas alcohólicos de los estudiantes de pregrado UTB que asistieron a la fiesta de graduación. Se tiene registro del consumo de bebidas a partir de un evento en el que asistieron $N = 5{,}000$ personas a la fiesta de grado y fueron enumeradas del uno al 5,000. Se desea infererir sobre la cantidad promedio de cervezas ingeridas por persona a partir de una muestra sistemática de 25 personas

a) El IC para la media y el total con una confiabilidad de 99%.

b) ¿Cuál es el tamaño de muestra para estimar la media de tal manera que sean estimada con una precisión de 5% con una confiabilidad de 95%

n	Número de muestra	Ingeridas	*n*	Número de muestra	Ingeridas
1	25	7.5	16	3,025	6.0
2	225	6	17	3,225	6.0
3	425	5	18	3,425	6.0
4	625	7	19	3,625	5.0
5	825	5	20	3,825	6.0
6	1,025	4	21	4,025	5.0
7	1,225	7	22	4,225	4.0
8	1,425	3	23	4,425	4.5
9	1,625	8	24	4,625	5.0
10	1,825	3.5	25	4,825	7.0
11	2,025	4.5			
12	2,225	6			
13	2,425	6.5			
14	2,625	7			
15	2,825	3			

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Estimación de la PROPORCIÓN y el TOTAL de la muestra sistemática

Una vez obtenida la muestra aleaotoria sistemática , el objetivo será caracterizar la población por medio de una muestra estimando los parámetros de mayor interés, como la Proporción y el total poblacional.

Estimador para la Proporción Poblacional $$p_s = \bar{y}_s = \frac{\sum_{i=1}^{n} y_i}{n}$$

Estimador para el Total Poblacional $$\tau_s = N p_s$$

Estimadores correspondientes a las varianzas de la Proporción y del total.

$$S_{p_s}^2 = \left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)$$

$$S_{\tau_s}^2 = N^2\left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)$$

donde $q_s = 1 - p_s$.

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Los intervalos de confianza para la PROPORCIÓN y para el total

Una vez obtenida la muestra, el objetivo será caracterizar la población por medio de una muestra estimando los parámetros de mayor interés, como la PROPORCIÓN y el total poblacional.

Intervalos para la Proporción Poblacional

$$p_s \pm t_{(n-1,1-\alpha/2)} \sqrt{\left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)}$$

donde $q_s = 1 - p_s$.

Intervalo para el Total Poblacional

$$\hat{\tau}_s \pm t_{(n-1,1-\alpha/2)} N \sqrt{\left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)}$$

Obsérvese que la estimación de la varianza es la misma que la presentada en el muestreo aleatorio aleatorio

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Tamaño de la Muestra Para Estimar la PROPORCIÓN

A fin de determinar el tamaño de la muestra para estimar a p o $\hat{\tau}$, se procede como en el Muestreo aleatorio Simple. Primero se elige un valor de $\boldsymbol{d}$, es decir, la precisión(ancho del intervalo) que se está dispuesto aceptar en las estimaciones

$$d = t_{(n-1,1-\alpha/2)}\sqrt{S_{p_s}^2}$$

$$S_{p_s}^2 = \left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)$$

$$S_{\tau_s}^2 = N^2\left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)$$

El tamaño de muestra para estimar la PROPORCIÓN vendría por(despejando a *n* en la formula anterior):

$$n^* = \frac{N t_{(n-1,1-\alpha/2)}^2 p_s q_s}{N d^2 + t_{(n-1,1-\alpha/2)}^2 p_s q_s}$$

Para fines prácticos la varianza poblacional se sustituye por la varianza muestral

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Ejemplo:

Las directivas de un colegio de la ciudad desea estudiar la satisfacción de los alumnos sobre la nueva cafeteria . Para realizar dicha encuesta se elegirán n alumnos entre los 10,000 estudiantes del colegio. Se pretende obtener una muestra de 18 alumnos. A continuación se obtiene k:

$$k = \frac{N}{n} = \frac{10,000}{18} = 555.5556. \qquad k = 556$$

El primer alumno que será encuestado se elegirá aleatoriamente entre el primero y el 556. A partir de allí el siguiente alumno encuestado sería el ocupado en la lista a una distancia *k.*

Los datos se presentan en el siguiente cuadro:

No. de alumno	Respuesta	No. de alumno	Respuesta
422	0	5,982	1
978	1	6,538	1
1,534	1	7,094	1
2,090	1	7,650	0
2,646	0	8,206	0
3,202	1	8,762	1
3,758	0	9,318	0
4,314	0	9,874	0
4,870	0	5,426	0

0: Insatisfecho
1: Satisfecho

Hallar:

a) Proporción de estudiantes satisfechos

b) Error muestral (S_{ps}).

c) IC para la proporción poblacional al 95%.

d) El total de estudiantes satisfechos.

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Ejemplo:

Las directivas de un colegio de la ciudad desea estudiar la satisfacción de los alumnos sobre la nueva cafeteria . Para realizar dicha encuesta se elegirán n alumnos entre los 10,000 estudiantes del colegio. Se pretende obtener una muestra de 18 alumnos. A continuación se obtiene k:

$$k = 556$$

Hallar:

a) Proporción de estudiantes satisfechos

$$p_s = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{a}{n} = \frac{8}{18} = 0.4444$$

44.44% de estudiantes satisfechos.

$$q_s = 1 - p_s = 1 - 0.5 = 0.5555$$

55.55% de estudiantes insatisfechos.

No. de alumno	Respuesta	No. de alumno	Respuesta
422	0	5,982	1
978	1	6,538	1
1,534	1	7,094	1
2,090	1	7,650	0
2,646	0	8,206	0
3,202	1	8,762	1
3,758	0	9,318	0
4,314	0	9,874	0
4,870	0	5,426	0

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Ejemplo:

Las directivas de un colegio de la ciudad desea estudiar la satisfacción de los alumnos sobre la nueva cafeteria . Para realizar dicha encuesta se elegirán n alumnos entre los 10,000 estudiantes del colegio. Se pretende obtener una muestra de 18 alumnos. A continuación se obtiene k:

$$k = 556$$

Hallar:

b) Error muestral (S_{ps}).

$$S_{p_s} = \sqrt{\left(\frac{N-n}{N}\right)\left(\frac{p_s q_s}{n}\right)}$$

donde: $N = 10{,}000$, $n = 18$, $p_s = 0.444444$ y $q_s = 0.555556$.

$$S_{p_s} = \sqrt{\left(\frac{10{,}000-18}{10{,}000}\right)\left(\frac{(0.444444)(0.555556)}{18}\right)} = 0.117014.$$

No. de alumno	Respuesta	No. de alumno	Respuesta
422	0	5,982	1
978	1	6,538	1
1,534	1	7,094	1
2,090	1	7,650	0
2,646	0	8,206	0
3,202	1	8,762	1
3,758	0	9,318	0
4,314	0	9,874	0
4,870	0	5,426	0

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Ejemplo:

Las directivas de un colegio de la ciudad desea estudiar la satisfacción de los alumnos sobre la nueva cafeteria . Para realizar dicha encuesta se elegirán n alumnos entre los 10,000 estudiantes del colegio. Se pretende obtener una muestra de 18 alumnos. A continuación se obtiene k:

$$k = 556$$

Hallar:

c) IC para la proporción poblacional al 95%.

$$p_s \pm t_{(n-1,1-\alpha/2)} S_{p_s}$$

donde: $p_s = 0.444444$, $S_{p_s} = 0.117014$, $t_{(n-1,1-\alpha/2)} = t_{(18-1,1-0.025)} = 2.1098$.

$$0.444444 \pm (2.1098)(0.117014)$$

$$0.197568 \leq P_s \leq 0.69132.$$

Con 95% de confianza se estima que la proporción de estudiantes satisfechos está entre 19.75% y 69.13%.

No. de alumno	Respuesta	No. de alumno	Respuesta
422	0	5,982	1
978	1	6,538	1
1,534	1	7,094	1
2,090	1	7,650	0
2,646	0	8,206	0
3,202	1	8,762	1
3,758	0	9,318	0
4,314	0	9,874	0
4,870	0	5,426	0

Estimación de parámetros de interés para PROPORCIÓN bajo muestreo sistemático

Ejemplo:

Las directivas de un colegio de la ciudad desea estudiar la satisfacción de los alumnos sobre la nueva cafeteria . Para realizar dicha encuesta se elegirán n alumnos entre los 10,000 estudiantes del colegio. Se pretende obtener una muestra de 18 alumnos. A continuación se obtiene k:

$$k = 556$$

Hallar:

d) El total de estudiantes satisfechos.

$$\hat{\tau} = Np_s$$

donde: $N = 10{,}000$ y $p_s = 0.444444$.

$$\hat{\tau} = (10{,}000)(0.444444) = 4{,}444.44.$$

No. de alumno	Respuesta	No. de alumno	Respuesta
422	0	5,982	1
978	1	6,538	1
1,534	1	7,094	1
2,090	1	7,650	0
2,646	0	8,206	0
3,202	1	8,762	1
3,758	0	9,318	0
4,314	0	9,874	0
4,870	0	5,426	0

EJERCICIO DE ESTIMACIÓN MUESTREO SISTEMÁTICO

Se estudia el comportamiento de compra en un centro comercial observando a los clientes durante una jornada. Para optimizar el proceso, se selecciona **cada 15° cliente** (muestreo sistemático).

Variables observadas:

- *ID*: Número de cliente.
- *HoraIngreso*: Hora de entrada del cliente (entre 9 a.m. y 8 p.m.).
- *Compra*: 1 si realizó compra, 0 si no.
- *MontoCompra*: Valor en pesos de la compra (0 si no compró).

Descargar en Savio archivo Excel*: Actividad_Retail_Clientes.xlsx*

Objetivos de la actividad:

1. Estimar la **proporción de clientes que compran** usando la muestra sistemática.
2. Calcular el **promedio y desviación estándar del monto de compra**.
3. Construir un **intervalo de confianza al 95%** para la proporción de compradores y para el monto promedio.
4. Estimar el **número total de compradores** entre los 450 observados.
5. Analizar si existen **horas pico o valle** en los datos observados.

EJERCICIO ESTIMACIÓN PARA PROPORCIÓN

Descargue en Savio la base de datos “**student-performance.csv**”.

a) Tome un muestreo **sistemático** de **100** estudiantes.

b) Para la variable “**test preparation course”** (que hace referencia a si el estudiante completó o no el el curso de preparación) estime un Intervalo de confianza al 95% para la proporción de los estudiantes que completaron el curso.

c) Cantidad total de estudiantes que completaron el curso?.

d) ¿Cuál es el tamaño de muestra para estimar la proporción de los estudiantes que completaron el curso de tal manera que sea estimada con una precisión de 5% y una confiabilidad de 95%

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1

REVISTA UNIVERSITARIA DE GEOGRAFÍA

Artículos

Aplicación del muestreo sistemático en áreas rurales de poca accesibilidad de la Amazonía ecuatoriana. El uso de la fotografía aérea en el muestreo sistemático

The use of aerial photography in systematic sampling

Álvaro Dávila G. alvaro.davila@mail.igm.gob.ec
Instituto Geográfico Militar, Ecuador

Aplicación del muestreo sistemático en áreas rurales de poca accesibilidad de la Amazonía ecuatoriana. El uso de la fotografía aérea en el muestreo sistemático

Revista Universitaria de Geografía, vol. 27, núm. 1, pp. 29-48, 2018

Universidad Nacional del Sur

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1

Resumen:

En el desarrollo de este trabajo se ha tratado de explotar la información que sobre el territorio ofrecen las fotografías, presentando una propuesta que toma en cuenta el costo financiero e integra al principio básico del muestreo sistemático, la espacialización a través de ortofotos, como fase previa a la realización de una encuesta socioeconómica en la parroquia amazónica Diez de Agosto, Ecuador. El procedimiento aporta al muestreo sistemático algunas bondades: reemplaza los requerimientos de ordenar y numerar la población objetivo, así como de disponer del trazado físico de todas las unidades de muestreo, en virtud de que integra una representación física georreferenciada del terreno identificada en el mosaico ortofotográfico. Además, la utilización de una grilla cuadricular de lado igual al salto sistemático, conjuntamente con el ángulo elegido al azar con el que se sobrepone al espacio de la población para elegir sus elementos, le infiere aleatoriedad, asegurando que la muestra sea extendida equitativamente hacia toda la población. El método también evidenció que es factible combinar la situación financiera con la precisión que se puede alcanzar con el muestreo y que favorece la captura de información en un GPS navegador para facilitar el reconocimiento de las unidades espaciales en campo.

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1

El problema

La necesidad de contar con estudios para la planificación del espacio intervenido de la Amazonía ecuatoriana se ha vinculado históricamente a los pocos recursos económicos que disponen los gobiernos municipales para recolectar datos sociales y económicos sobre esta zona, así como a las difíciles condiciones físico-geográficas, la débil accesibilidad y la existencia de un cierto grado de exclusión social sobre el territorio, a pesar de que por sus características únicas y por la importancia no solo a nivel nacional sino internacional, desde el año 2008, la región amazónica forma parte de la Circunscripción Territorial Especial Amazónica (CTEA) sobre la base del artículo 250 de la Constitución de 2008, que establece:

A fin de dar respuestas a estos planteamientos y considerando que la única fuente de información son los censos de población y vivienda del año 2010 y períodos anteriores que no contienen datos referentes a actividades agropecuarias como producción, áreas cultivadas, comercialización, etc., se plantea el requerimiento de disponer de una metodología de captura de datos estadísticos que resulte económica y de fácil aplicación para esta zona de la Amazonía ecuatoriana. Por ello, el objetivo de este trabajo es proponer la aplicabilidad de un muestreo probabilístico cuya funcionalidad principal se encamine a determinar qué parte y qué elementos de la población o universo deben extraerse con el fin de inferir informaciones sobre ella en un área de la región amazónica ecuatoriana, correspondiente a la parroquia rural Diez de Agosto.

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1

Delimitación de fincas agropecuarias según fuentes oficiales

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1

Propuesta de los autores

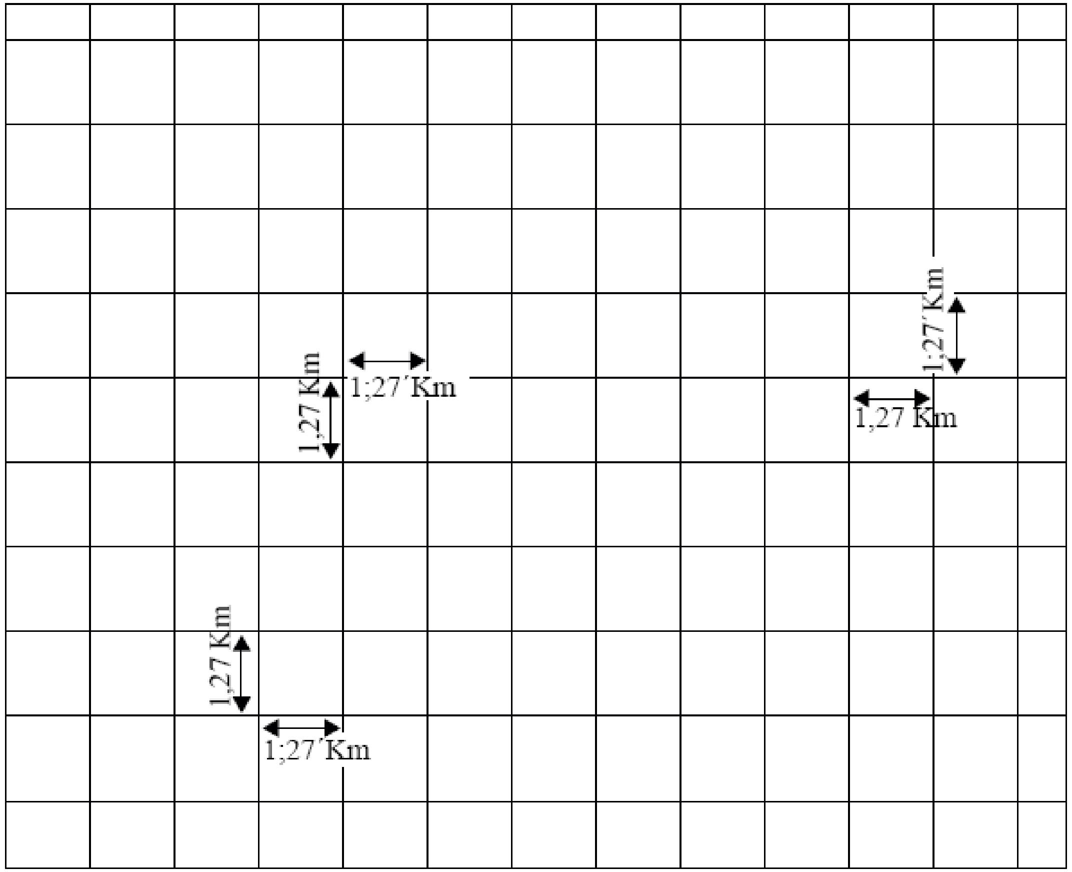


figura 4a

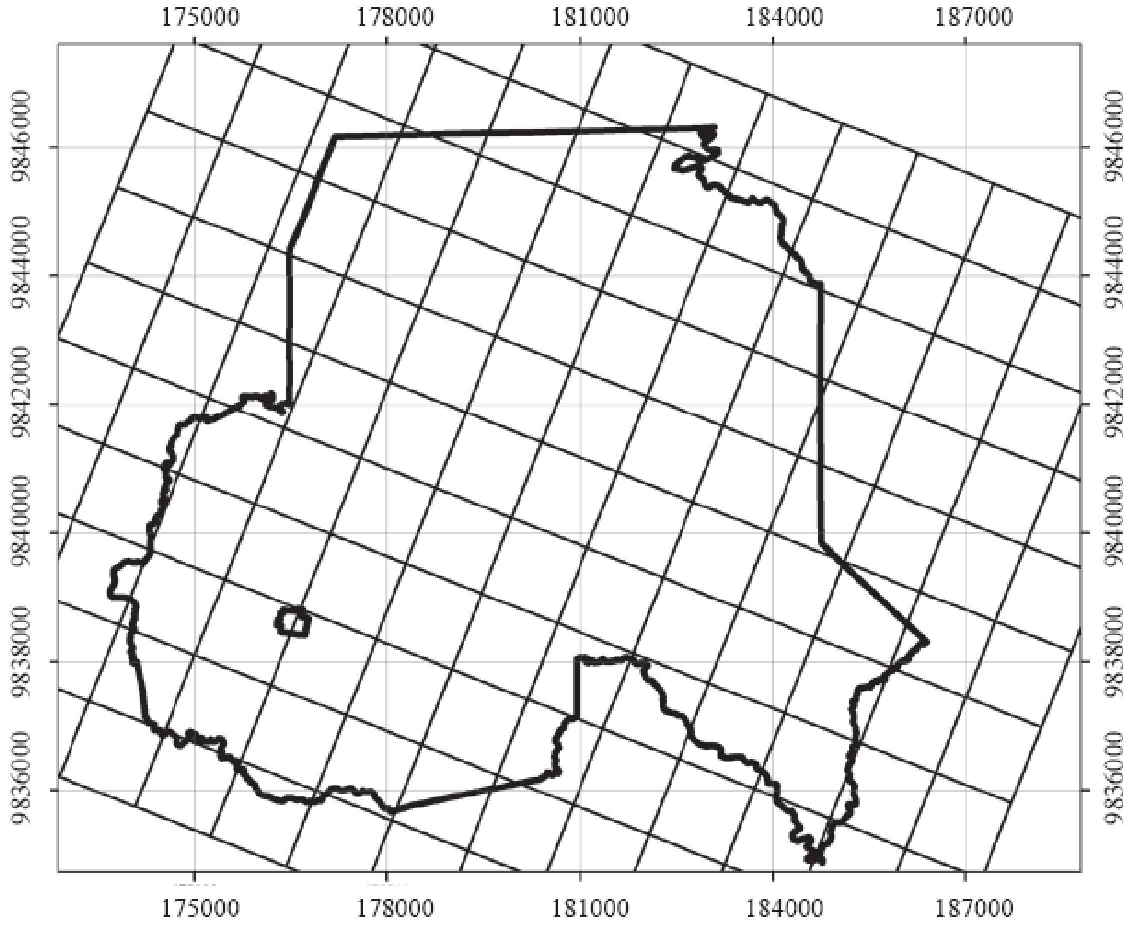
"se trazó el límite de la parroquia sobre las ortofotos y se midió su área, obteniendo el valor de 93,17 km². Si esta cantidad se divide para la muestra que es 58, entonces se tiene la superficie de cada encuesta en el terreno que corresponde a una cantidad de 1,61 km². Finalmente, asumimos este valor como el área de un cuadrado y al extraer su raíz cuadrada obtenemos 1,27 km, que es el salto sistemático y representa el lado de cada celda de la malla que se genera en el programa, como se indica en la figura 4a."

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1

Propuesta de los autores

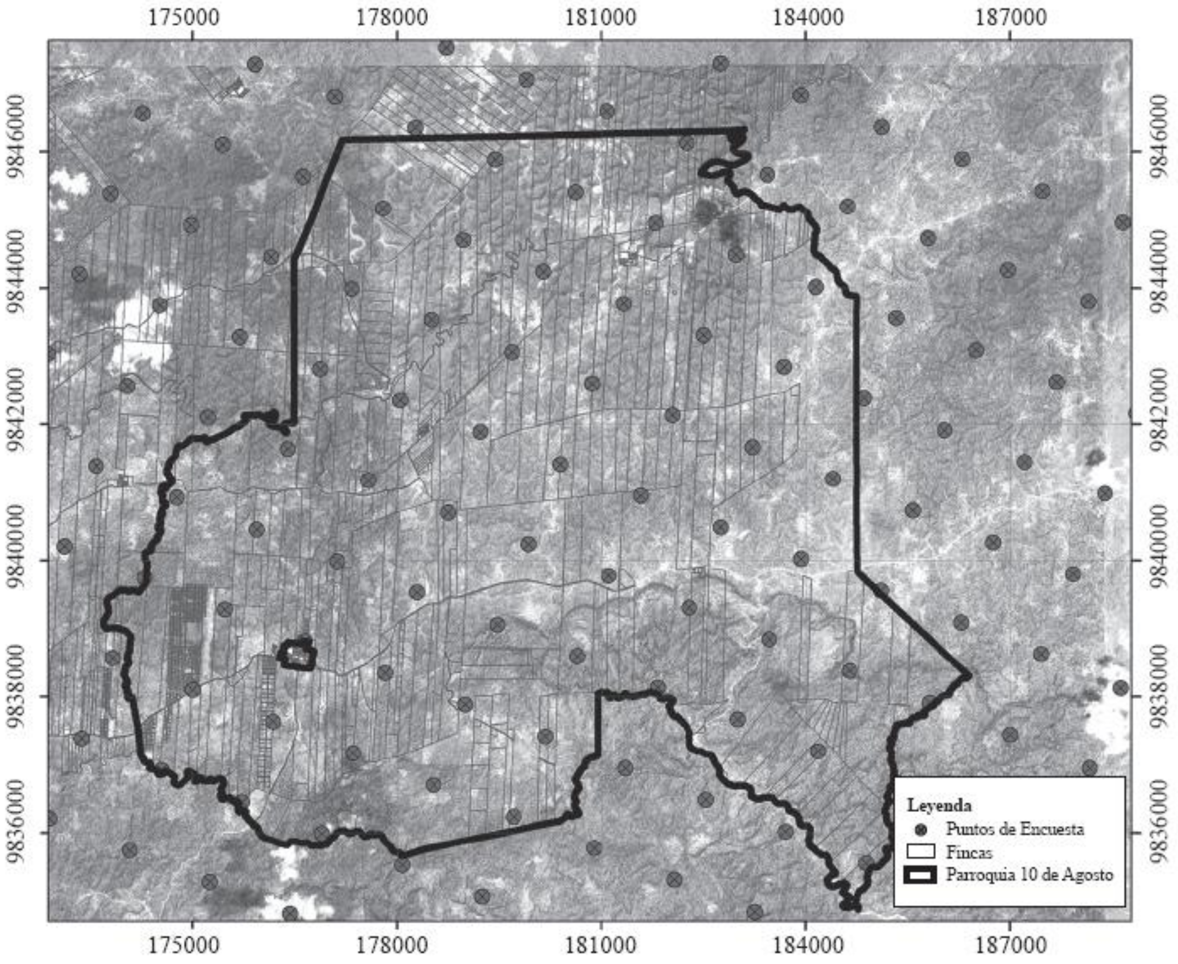


“El siguiente paso es sobreponer aleatoriamente (con cualquier ángulo de rotación, procurando que nunca sea 0°) la cuadrícula generada sobre el gráfico de la parroquia”

CASO APLICACIÓN

Artículo: Accesibilidad en áreas rurales

https://www.redalyc.org/journal/3832/383257036001/html/#redalyc_383257036001_ref1



(Fig. 5)

"A continuación, se sobreponen todas las capas de información y, con los respectivos comandos del programa, se grafican automáticamente los vértices que corresponden a los puntos de las encuestas, definidas por coordenadas planas X e Y (Fig. 5)."

CONCLUSIÓN

"Parte importante de esta propuesta es introducir una variante sobre la manera de determinar y aplicar el salto sistemático, de tal manera que se infiere aleatoriedad al muestreo sistemático a través de la utilización de fotografías aéreas rectificadas y de una malla cuadricular que puede sobreponerse aleatoriamente al espacio de la población. Además, gráficamente se puede constatar que la selección de los puntos abarca todo el territorio, asegurando una buena cobertura del muestreo sobre el territorio."