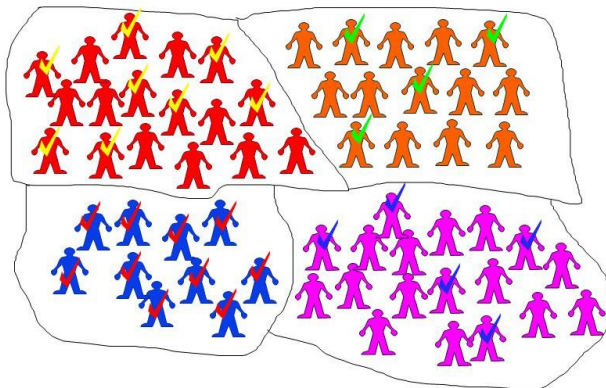


Fundamentos del muestreo



Muestreo estratificado



Oscar Centeno Mora

Preámbulo...

- Hasta el momento hemos visto diversas técnicas y métodos de análisis procedente de la inferencia estadística.
- Si bien es cierto que en el tema de estimación se había explicado que las unidades de estudio debían proceder de una muestra (ojalá probabilística), lo cierto es que se podrían estar preguntando: ¿cómo se obtiene una muestra probabilística?
- Una muestra probabilística se constituye de dos componentes:
 1. El cálculo del tamaño de la muestra (n)
 2. La selección de las n unidades de estudio.
- En el presente capítulo se explican los fundamentos del muestreo, además de la aplicación de las principales técnicas del muestreo probabilístico.



Recordatorio...



Encuesta

VRS

Censo

- Costo
- Rapidez
- Más posibilidades de análisis
- Mayor exactitud
- Menor volumen de trabajo

- Proporcionan datos para áreas pequeñas
- El sesgo de cobertura puede evaluarse más fácilmente
- No se requieren muestristas
- Los productos censales son importantes:
 - Marco muestral de viviendas
 - Proyecciones de población
 - Mapas de pobreza
 - Migración interna e inmigración
 - Crecimiento urbano
 - Etc.

Índice

1

Fundamentos del
diseño de la
muestra

4

Error de muestreo

2

La inferencia
estadística

5

Tamaño de muestra
(n)

3

Muestreo:
probabilístico y no
probabilístico

6

Selección de las UE
de la n

Índice

1

Fundamentos del
diseño de la
muestra

DISEÑO DE MUESTRAS



ETAPAS A DEFINIR EN LA ENCUESTA

| | | | |
|---|---|----|--|
| 1 | Objetivos (generales y específicos) | 6 | Métodos de análisis (procedimientos para resumir resultados que puedan comprenderse y usarse) |
| 2 | Población que será muestreada (a veces es simple y en ocasiones no lo es) | 7 | Uso de los resultados de la encuesta |
| 3 | Definición de variables de la encuesta (especificar naturaleza de las características, categorías de clasificación, unidades para expresarlas) | 8 | Marco muestral (contiene la población de estudio) |
| 4 | Métodos de medición o recolección de datos (medios a utilizar: entrevista cara a cara, auto-administrada, por internet, telefónica, correo, etc.) (casos especiales, como por ejemplo, peso, talla y sangre, estudio de sal, encuestas de ingresos y gastos, etc.) | 9 | Diseño de la muestra |
| 5 | Precisión deseada (se relaciona con los costos) | 10 | Estudio piloto |
| | | 11 | Organización y ejecución del trabajo de campo |
| | | 12 | Crítica |
| | | 13 | Programa de entrada de datos |

ETAPAS A DEFINIR EN LA ENCUESTA

| | |
|----|----------------------|
| 14 | Digitación |
| 15 | Verificación |
| 16 | Tabulación |
| 17 | Análisis |
| 18 | Redacción de informe |

CONCEPTOS BÁSICOS Y DEFINICIONES

01

Elemento, unidad de estudio, unidad de análisis o unidad elemental

- Unidad que es parte de la población
- Es la unidad en la que se busca la información para hacer inferencias a esa misma población

02

Población

Agregado de observaciones

Para definir la población se debe especificar:

- | | |
|-------------|---------------------------------------|
| a)Contenido | Mujeres en edad fértil (15-49 años) |
| b)Unidades | que residen en viviendas particulares |
| c)Extensión | en todo el territorio nacional |
| d)Tiempo | a mediados del año 2013 |

03

Unidad de muestreo

Contiene los elementos

- En el muestreo de elementos cada UM contiene un elemento
- En el muestreo de conglomerados cada UM contiene varios elementos
- Ejemplos de ambos casos

CONCEPTOS BÁSICOS Y DEFINICIONES

04

**Unidad de observación,
unidad de respuesta,
unidad de información o
informante**

Se obtiene la información

La unidad informante puede diferir de la unidad de estudio.

Ejemplos

05

Subclase

Porción de la muestra que se considera como muestra de la porción correspondiente de la población de encuesta y no es planeada.

Ejemplos

06

Estratos

División de la población en sub-poblaciones para mejorar los procedimientos de selección.

Ejemplos

CONCEPTOS BÁSICOS Y DEFINICIONES

07

Unidades de listado

Se usan para identificar y seleccionar unidades de muestreo.

Ejemplos

08

Características de los elementos

Se transforman en variables Yi mediante las operaciones de medición.

Ejemplos

09

Marco muestral

Conjunto de UM que comprenden la población.

Ejemplos

EJEMPLOS

| Estudio | Unidad de estudio | Unidad de muestreo | Informante | Marco muestral |
|--|-------------------|-----------------------------------|---------------------------|--|
| Encuesta a miembros del Colegio de Médicos | Médico | Médico | Médico | Lista de asociados |
| Encuesta a niños y niñas de kirdergarden | Niño (a) | Kinder | Maestra | Lista de kinder |
| Encuesta de intención de voto | Votante | Número de teléfono del votante | Votante | Directorio Ramdom digital dialing (RDD)* |
| Encuesta Nacional de Hogares | Hogar Persona | Unidad Primaria de Muestreo (UPM) | Persona adulta calificada | Listados de UPM *** |

* Dentro del número telefónico: lista de residentes en capacidad de votar, próximo cumpleaños

** Lista de UPM normalmente ordenada por división territorial y para cada una se dispone de una medida de tamaño para seleccionar.

¿De dónde obtenemos las UE? ... de un Marco Muestral



Q&A

MARCO MUESTRAL

¿Qué es un marco muestral?

Es el conjunto de unidades de estudio, representadas de forma física o electrónica.

Un marco es perfecto si:

- 1) todos los elementos
- 2) aparecen por separado en la lista
- 3) cada uno una sola vez
- 4) no aparece en la lista ninguna otra cosa.



Fundamental para una auditoría

a partir de este se seleccionarán las unidades de estudio que constituirá la muestra.

MARCO MUESTRAL

| order_id | order_no | customer_name | order_date | total_amount |
|----------|----------|-------------------|------------|--------------|
| 554 | SO44425 | Anna Williams | 9/18/2011 | 3578.27 |
| 1045 | SO45109 | Brendan Xu | 12/7/2011 | 3578.27 |
| 1206 | SO45349 | Candice Sun | 1/1/2012 | 3578.27 |
| 1638 | SO45882 | Allison Gonzalez | 3/11/2012 | 3578.27 |
| 3159 | SO47951 | Mindy Kumar | 9/28/2012 | 2071.4196 |
| 4294 | SO49636 | Crystal Yang | 2/6/2013 | 2443.35 |
| 4712 | SO50128 | Emily Bryant | 3/24/2013 | 2071.4196 |
| 6587 | SO52520 | Julia Anderson | 7/12/2013 | 1183.47 |
| 10546 | SO56754 | Patrick James | 9/24/2013 | 1199.46 |
| 11011 | SO57398 | Kelly Patterson | 10/3/2013 | 4.99 |
| 11167 | SO57554 | Henry Garcia | 10/6/2013 | 23.97 |
| 11706 | SO58093 | Christina Bell | 10/16/2013 | 39.98 |
| 12332 | SO58719 | Gregory Yuan | 10/27/2013 | 564.99 |
| 12973 | SO59536 | Jocelyn Alexander | 11/5/2013 | 35 |
| 16204 | SO62863 | Jodi Deng | 12/26/2013 | 650.95 |
| 16329 | SO62988 | Maurice Shan | 12/28/2013 | 1268.84 |
| 16385 | SO63044 | Christy Zhou | 12/29/2013 | 2341.97 |
| 18013 | SO64847 | Katherine King | 1/23/2014 | 1184.46 |
| 19513 | SO66522 | Jon Sun | 2/17/2014 | 8.99 |
| 20461 | SO67564 | Dana Alonso | 3/3/2014 | 553.97 |
| 23311 | SO70594 | Marc Martin | 4/13/2014 | 785.32 |
| 23435 | SO70718 | Tabitha Mehta | 4/15/2014 | 2376.96 |
| 23711 | SO70994 | Hunter Simmons | 4/19/2014 | 2317.92 |
| 23924 | SO71207 | Carol Martinez | 4/22/2014 | 578.46 |
| 25014 | SO72478 | Nicholas Davis | 5/8/2014 | 50.23 |
| 25415 | SO72879 | Dalton Cook | 5/14/2014 | 69.99 |
| 25882 | SO73346 | Rafael Black | 5/20/2014 | 24.49 |
| 26311 | SO73775 | Jerry Xu | 5/25/2014 | 1732.26 |
| 26658 | SO74122 | Alejandro Zhou | 5/30/2014 | 62.98 |
| 27406 | SO74870 | Whitney Rodriguez | 6/22/2014 | 67.97 |

Table 2 A Simple Random Sample Without Replacement

| | year | gpi | gps | recession81 | recession07 |
|----|------|-------|-------|-------------|-------------|
| 1 | 1959 | 78.5 | 84.6 | 0 | 0 |
| 2 | 1960 | 78.9 | 84.8 | 0 | 0 |
| 3 | 1961 | 78.2 | 91.8 | 0 | 0 |
| 4 | 1962 | 88.1 | 100.7 | 0 | 0 |
| 5 | 1963 | 93.8 | 104.6 | 0 | 0 |
| 6 | 1964 | 102.1 | 117.9 | 0 | 0 |
| 7 | 1965 | 118.2 | 129.7 | 0 | 0 |
| 8 | 1966 | 131.3 | 138.6 | 0 | 0 |
| 9 | 1967 | 128.6 | 151.3 | 0 | 0 |
| 10 | 1968 | 141.2 | 153.7 | 0 | 0 |
| 11 | 1969 | 156.4 | 156.8 | 0 | 0 |
| 12 | 1970 | 152.4 | 174.1 | 0 | 0 |
| 13 | 1971 | 178.2 | 202.5 | 0 | 0 |
| 14 | 1972 | 207.6 | 216.8 | 0 | 0 |
| 15 | 1973 | 244.5 | 256.3 | 0 | 0 |
| 16 | 1974 | 249.4 | 270 | 0 | 0 |
| 17 | 1975 | 230.2 | 323.6 | 0 | 0 |
| 18 | 1976 | 292 | 343.8 | 0 | 0 |
| 19 | 1977 | 361.3 | 382.8 | 0 | 0 |
| 20 | 1978 | 438 | 436.3 | 0 | 0 |
| 21 | 1979 | 492.9 | 480.5 | 0 | 0 |
| 22 | 1980 | 479.3 | 532.4 | 0 | 0 |
| 23 | 1981 | 572.4 | 630.3 | 1 | 0 |

MARCO MUESTRAL

Los marcos muestrales son imperfectos

- Documentar las condiciones
- Corregir el marco
- Documentar estado final

Verificar

Elementos faltantes

Marcos incompletos por la falta de registro.

Conglomerado de elementos

Varias unidades aparecen en un único registro, por lo que se omite cierta información por línea de registro.

Blancos o elementos ilegibles

Son elementos que no pertenecen a la población de estudio, o que poseen ciertas características no deseadas para el estudio.

Listados duplicados

Son unidades de estudio que aparecen más de una vez en los registros, produciendo que tengan mayores posibilidades de salir en la muestra.

Soluciones

01

Ignórelo y no lo tome en cuenta.

Esto se hace si el error es pequeño y cuando corregirlo resulte muy costoso.

02

Redefinir la población para que se ajuste al marco.

Se evita si la orientación de la muestra se desvía de su meta. Pero si el resultado es trivial se hace.

03

Corregir la lista completa de la población.

Encontrar elementos faltantes, dividir los conglomerados, eliminar elementos extraños y los listados duplicados.

Deben buscarse procedimientos para restaurar las probabilidades de selección que los marcos imperfectos tienden a destruir

Nota: describir en el informe el problema, la magnitud y la solución. ¿Qué sucede si no podemos hacer lo anterior

MARCO MUESTRAL: ALGUNOS EJEMPLOS



MARCO MUESTRAL

Ejemplos

- **En el muestreo por áreas**, el marco consiste de mapas, pero éste puede construirse sin poner en el mapa a la población completa.
- Un marco de niños y niñas de escuela puede provenir de los distritos escolares que contienen escuelas, sus aulas y, por último, los niños y niñas. El diseño puede llevarse a cabo en varias etapas sin obtener una lista completa de todos los niños y niñas.
- En la práctica se utiliza como marco una lista de personas o de direcciones, un mapa con viviendas, una fotografía aérea, un directorio telefónico o cualquier material o procedimiento que permita identificar o localizar a los elementos de la población y seleccionar la muestra.

MARCO MUESTRAL

COLEGIOS DEL MEP (N=811)

| COD | NOMBRE | PROVINCIA | CANTON | DISTRITO | TOTAL | T01 | T02 | T03 | T04 | T05 | T06 |
|------|---------------------------|-----------|------------|--------------|-------|-----|-----|-----|-----|-----|-----|
| 3957 | COLEGIO REP. DE MEXICO | SAN JOSE | SAN JOSE | CARMEN | 912 | 247 | 194 | 165 | 176 | 130 | 0 |
| 3963 | JOSE FIDEL TRISTAN | SAN JOSE | SAN JOSE | MERCED | 561 | 267 | 99 | 86 | 72 | 37 | 0 |
| 3941 | SAN JOSE | SAN JOSE | SAN JOSE | MERCED | 1422 | 528 | 394 | 143 | 203 | 154 | 0 |
| 0000 | MARIA AUXILIADORA | SAN JOSE | SAN JOSE | MATA REDONDA | 183 | 41 | 42 | 36 | 28 | 36 | 0 |
| 3942 | LICEO DEL SUR | SAN JOSE | SAN JOSE | HOSPITAL | 1362 | 572 | 300 | 195 | 219 | 76 | 0 |
| 4154 | C.T.P. DON BOSCO | SAN JOSE | ALAJUELITA | CONCEPCION | 1305 | 228 | 227 | 225 | 239 | 203 | 183 |
| 0000 | SAGRADO CORAZON | SAN JOSE | SAN JOSE | CATEDRAL | 319 | 78 | 69 | 58 | 52 | 62 | 0 |
| 0000 | SEMINARIO | SAN JOSE | SAN JOSE | CATEDRAL | 398 | 84 | 74 | 92 | 71 | 77 | 0 |
| 3938 | SUPERIOR DE SEÑORITAS | SAN JOSE | SAN JOSE | CATEDRAL | 1129 | 351 | 244 | 207 | 209 | 118 | 0 |
| 3940 | LICEO DE COSTA RICA | SAN JOSE | SAN JOSE | CATEDRAL | 1177 | 289 | 183 | 281 | 270 | 154 | 0 |
| 0000 | CRISTIANO ASAM. DE DIOS | SAN JOSE | SAN JOSE | ZAPOTE | 85 | 25 | 16 | 16 | 13 | 15 | 0 |
| 0000 | SALESIANO DON BOSCO | SAN JOSE | SAN JOSE | ZAPOTE | 345 | 84 | 81 | 61 | 57 | 62 | 0 |
| 3970 | EL ROSARIO | SAN JOSE | SAN JOSE | CATEDRAL | 460 | 108 | 99 | 93 | 83 | 77 | 0 |
| 3947 | RODRIGO FACIO BRENES | SAN JOSE | SAN JOSE | ZAPOTE | 1191 | 336 | 258 | 216 | 235 | 146 | 0 |
| 3950 | DR.JOSE MA. CASTRO MADRIZ | SAN JOSE | SAN JOSE | ZAPOTE | 1486 | 539 | 386 | 253 | 186 | 122 | 0 |
| 3964 | JULIO FONSECA GUTIERREZ | SAN JOSE | SAN JOSE | URUCA | 755 | 276 | 143 | 131 | 100 | 105 | 0 |
| 0000 | DR. JAIM WEIZMAN | SAN JOSE | SAN JOSE | MATA REDONDA | 90 | 23 | 13 | 17 | 19 | 18 | 0 |
| 0000 | LA SALLE | SAN JOSE | SAN JOSE | MATA REDONDA | 490 | 104 | 98 | 100 | 89 | 99 | 0 |
| 0000 | LOS ANGELES | SAN JOSE | SAN JOSE | MATA REDONDA | 329 | 110 | 81 | 48 | 47 | 43 | 0 |
| 0000 | YORKIN | SAN JOSE | CURRIDABAT | SANCHEZ | 79 | 13 | 19 | 19 | 15 | 13 | 0 |
| 4156 | C.T.P. EDUC. COM. Y SERV. | SAN JOSE | SAN JOSE | MATA REDONDA | 495 | 0 | 0 | 0 | 162 | 183 | 150 |
| 3943 | LUIS DOBLES SEGREDÁ | SAN JOSE | SAN JOSE | MATA REDONDA | 2015 | 720 | 396 | 327 | 412 | 160 | 0 |
| 0000 | BRITANICO DE COSTA RICA | SAN JOSE | SAN JOSE | PAVAS | 383 | 72 | 72 | 67 | 62 | 64 | 46 |
| 0000 | HUMBOLDT | SAN JOSE | SAN JOSE | PAVAS | 366 | 80 | 76 | 63 | 68 | 55 | 24 |
| 3968 | LICEO PAVAS | SAN JOSE | SAN JOSE | PAVAS | 1105 | 362 | 319 | 165 | 191 | 68 | 0 |
| 0000 | ADVENTISTA DE COSTA RICA | SAN JOSE | SAN JOSE | HATILLO | 403 | 92 | 83 | 88 | 76 | 64 | 0 |

DIRECTORIO DE ESTABLECIMIENTOS COMERCIALES

| COD_ | | DIRECCION | TELÉFONO |
|-------|---|-----------|----------|
| ACTIV | DESCRIPCION_ACTIVIDAD | PATRONO | |
| 0000 | ACTIVIDAD NO ESPECIFICADA | | |
| 0117 | CULTIVO DE CAÑA. | | |
| 0119 | CULTIVO DE FLORES. | | |
| 0119 | CULTIVO DE FLORES. | | |
| 0123 | CRÍA DE GANADO VACUNO (ENGORDE Y LECHE). | | |
| 0123 | CRÍA DE GANADO VACUNO (ENGORDE Y LECHE). | | |
| 0123 | CRÍA DE GANADO VACUNO (ENGORDE Y LECHE). | | |
| 0130 | CULT.DE PRODUC.AGRÍCOLAS EN COMBINACIÓN CON LA CRÍA DE ANIMALES. | | |
| 0140 | ACTIVI.DE SERV.AGRÍCOLAS Y GANADEROS EXCEPTO LAS ACTIVI.VETERINARIAS. | | |
| 0140 | ACTIVI.DE SERV.AGRÍCOLAS Y GANADEROS EXCEPTO LAS ACTIVI.VETERINARIAS. | | |
| 0140 | ACTIVI.DE SERV.AGRÍCOLAS Y GANADEROS EXCEPTO LAS ACTIVI.VETERINARIAS. | | |
| 0140 | ACTIVI.DE SERV.AGRÍCOLAS Y GANADEROS EXCEPTO LAS ACTIVI.VETERINARIAS. | | |
| 1541 | ELABORACIÓN DE PRODUCTOS DE PANADERÍA. | | |
| 1549 | ELABORACIÓN DE OTROS PRODUCTOS ALIMENTICIOS N.C.P. | | |
| 4100 | CAPTACIÓN, DEPURACIÓN Y DISTRIBUCIÓN DE AGUA. | | |
| 4520 | CONSTRUCCION DE EDIFICIOS, VIVIENDAS (O PARTES); OBRAS DE ING.CIVIL | | |
| 4520 | CONSTRUCCION DE EDIFICIOS, VIVIENDAS (O PARTES); OBRAS DE ING.CIVIL | | |
| 4520 | CONSTRUCCION DE EDIFICIOS, VIVIENDAS (O PARTES); OBRAS DE ING.CIVIL | | |
| 4520 | CONSTRUCCION DE EDIFICIOS, VIVIENDAS (O PARTES); OBRAS DE ING.CIVIL | | |
| 4520 | CONSTRUCCION DE EDIFICIOS, VIVIENDAS (O PARTES); OBRAS DE ING.CIVIL | | |
| 5122 | VENTA AL POR MAYOR DE ALIMENTOS, BEBIDAS Y TABACO. | | |
| 5143 | VENTA AL POR MAYOR:MATERIALES DE CONSTRUC.FERRETERÍA,EQUIPO Y OTROS | | |
| 5220 | VENTA AL POR MENOR DE ALIMENTOS, BEBIDAS Y TABACO | | |
| 5239 | VENTA AL POR MENOR DE OTROS PRODUCTOS EN ALMACENES ESPECIALIZADOS. | | |
| 5253 | VENTA AL POR MENOR EN LOCALES FIJOS O MERCADO. | | |
| 5520 | RESTAURANTES, BARES Y CANTINAS. | | |
| 5520 | RESTAURANTES, BARES Y CANTINAS. | | |

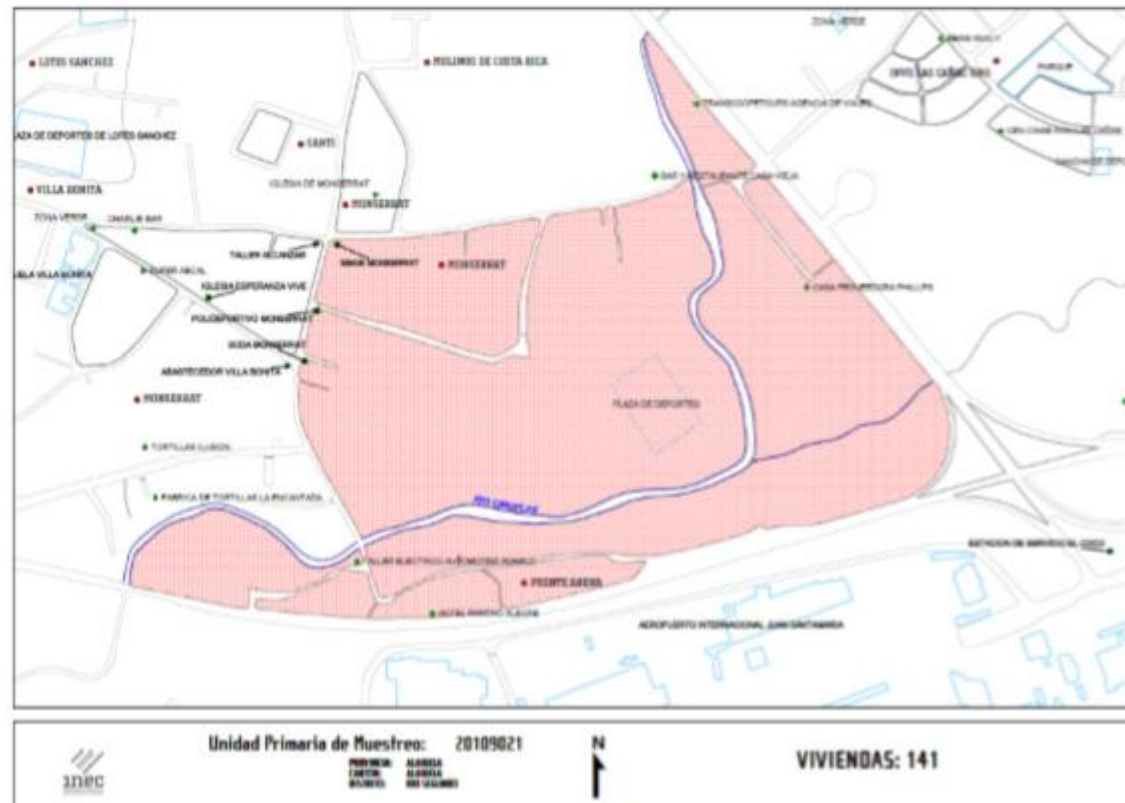
MARCO MUESTRAL

Sistema Regional de Establecimientos Agropecuarios (SIREA)



MARCO MUESTRAL

Marco muestral de áreas del INEC



Índice

1

Fundamentos del
diseño de la
muestra

2

La inferencia
estadística

OBJETIVO DEL MUESTREO: INFERENCIA ESTADÍSTICA

Valor poblacional

Valor numérico que sintetiza una o más características de la población de los N elementos de una población completa

Ejemplos

$$\bar{Y} = \sum \frac{Y_i}{N}$$

Otros casos son:

- Proporciones (P)
- Medianas (Me)
- Modas (Mo)

...

Valor muestral

Estimación que se calcula con base en n elementos de la muestra.

Ejemplos

$$\bar{y} = \sum \frac{y_i}{n}$$

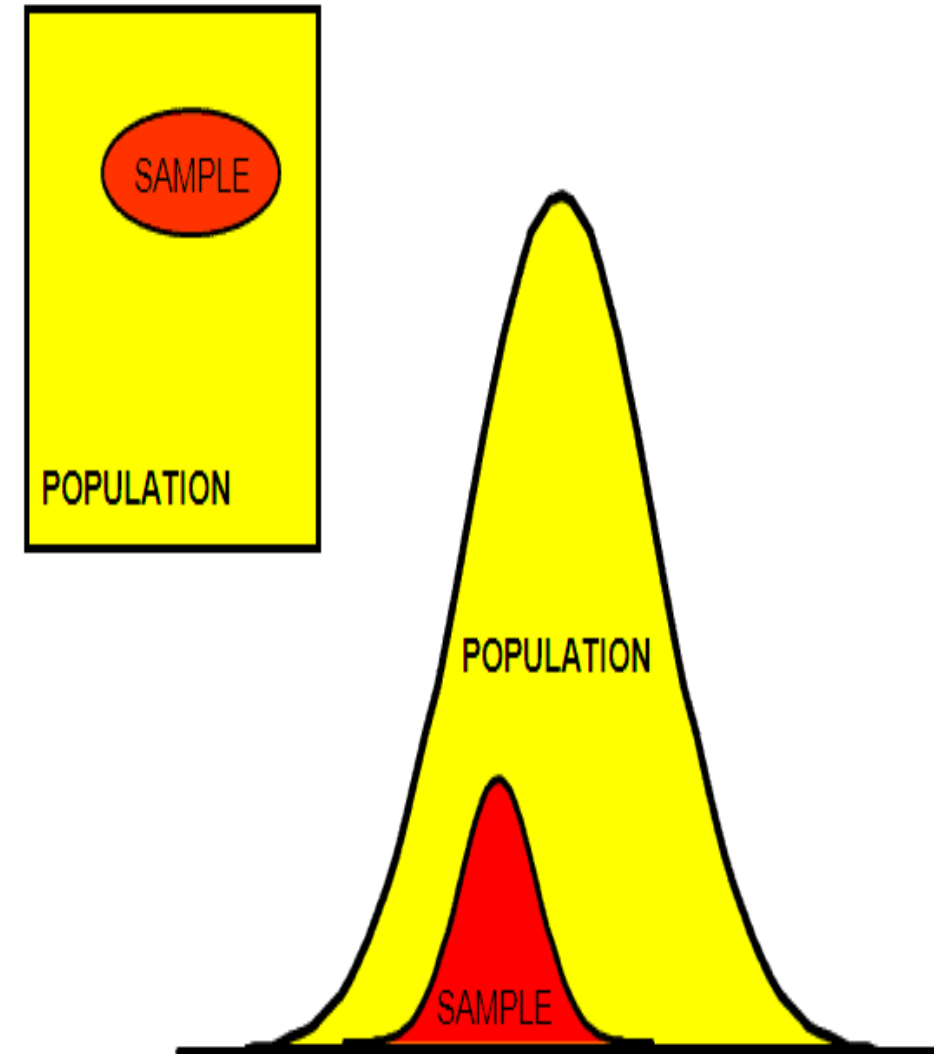
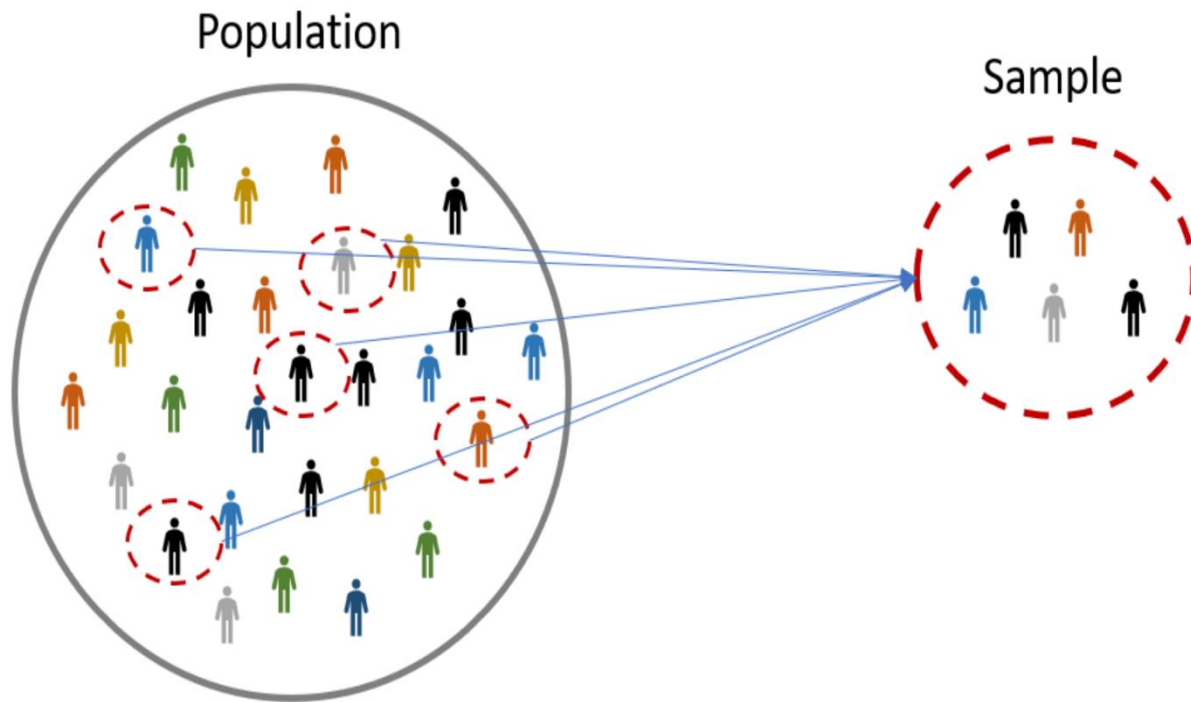
Otros casos son:

- Proporciones (p)
- Medianas (me)
- Modas (mo)

...

\bar{y} es una variable aleatoria que depende del diseño y de los elementos seleccionados

¿ES EL VALOR DE LA POBLACION EL VALOR VERDADERO O REAL?



¿ES EL VALOR DE LA POBLACION EL VALOR VERDADERO O REAL?

Valor de la población

Está sujeto a errores

Se obtendría si la población completa fuese medida bajo las condiciones reales de una encuesta

Valor verdadero

Se obtendría si todos los elementos de la población se midieran sin error

Errores de observación

Es producto de la imperfección de los procesos de medición

ERRORES DE LAS ESTADISTICAS

Errores de muestreo

Porque se consulta una parte de la población

Tipos

Errores no de muestreo

Porque los procedimientos de observación (medición) son imperfectos

¿CÓMO SE EVALUAN TEÓRICAMENTE LOS ERRORES DE MUESTREO?

Términos cotidianos

¿cuánto es $\bar{y} - \bar{Y}$?
¿esa diferencia es
grande, pequeña o \pm ?

Términos estadísticos y probabilísticos

¿qué valores de la media de \bar{y} son
posibles y cuál es la probabilidad de
que suceda cada uno?

¿CÓMO HACÍAMOS PARA AMINORAR O CALCUAR EL ERROR DE MUESTREO?



Índice

1

Fundamentos del
diseño de la
muestra

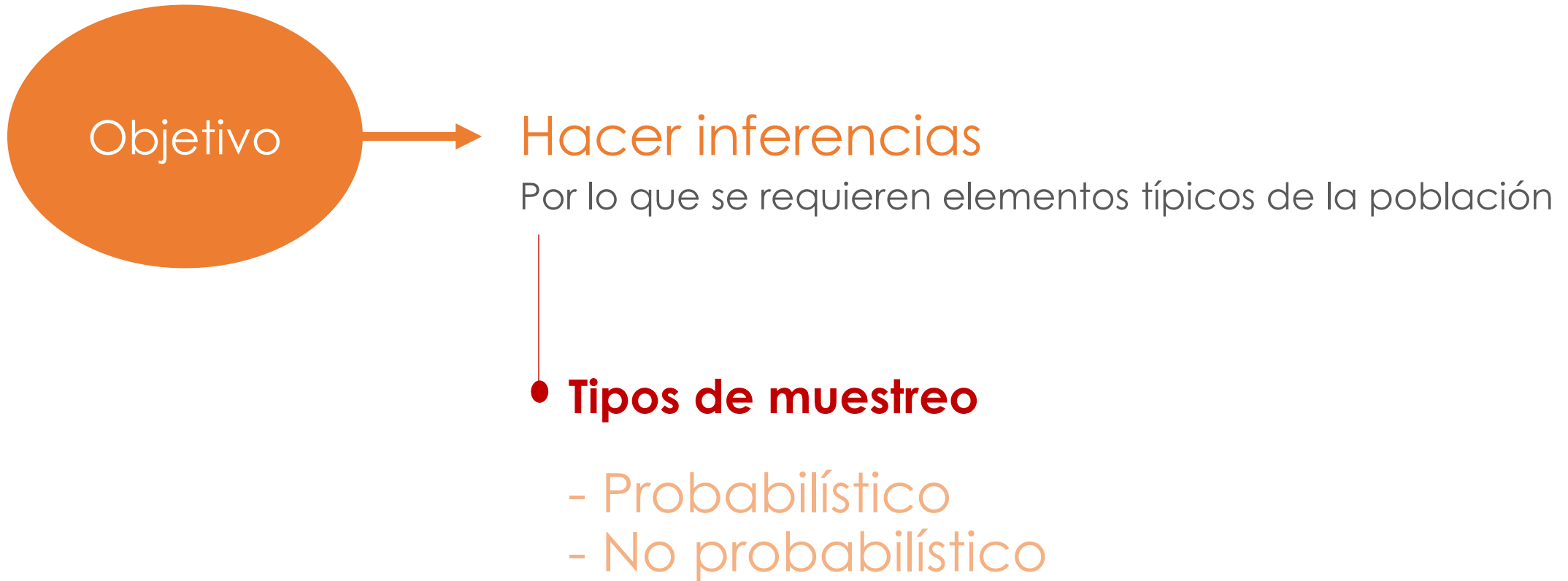
2

La inferencia
estadística

3

Muestreo:
probabilístico y no
probabilístico

OBJETIVO DEL MUESTREO



MUESTREO NO PROBABILISTICO

Procedimiento

No se efectúa bajo normas probabilísticas de selección, durante sus procesos intervienen opiniones y criterios personales del investigador o no existe norma bien definida o validada.

Tipos

Muestreo por cuotas

Muestreo intencional o conveniencia

Bola de Nieve

Muestreo Discrecional

MUESTREO NO PROBABILISTICO

Muestreo por cuotas

Hay un buen conocimiento de los estratos de la población y/o de los individuos más representativos o adecuados para los fines de la investigación.

Muestreo intencional o conveniencia

Pretende seleccionar unidades de análisis que cumplen los requisitos de la población objeto de estudio.

Bola de Nieve

Se reúnen los datos de los pocos miembros de la población objetivo que se puedan localizar y se les pide información necesaria para ubicar a otros miembros que conozcan de esa población.

* Poblaciones difíciles de muestrear

Muestreo Discrecional

A criterio del investigador los elementos son elegidos sobre lo que él cree que pueden aportar al estudio.

* Preferible usarlas en estudios exploratorios y pruebas pilotos.

MUESTREO PROBABILÍSTICO

Procedimiento

Se seleccionan muestras probabilísticas en el que cada elemento N tiene una probabilidad conocida y no nula de ser seleccionado y esa probabilidad se obtiene mediante una operación mecánica de aleatorización

Propiedad

Son medibles, las inferencias estadísticas a valores poblacionales pueden basarse en medidas de variabilidad, generalmente por los errores estándar calculados con base en la muestra

Una de las metas

Resultado en diseños prácticos, económicos y precisos.

Tipología básica de diseños muestrales probabilísticos

El **muestreo de encuestas** se ocupa principalmente de modificaciones del muestreo aleatorio simple o del mío

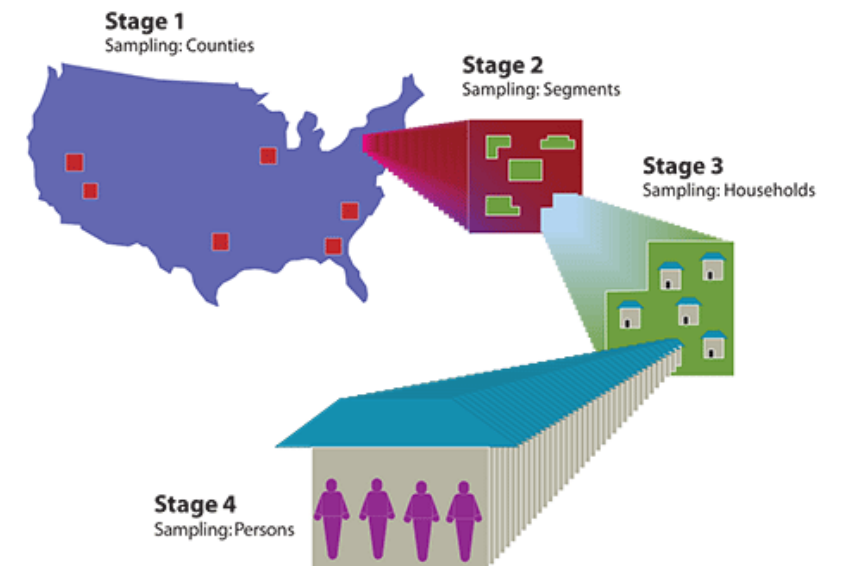
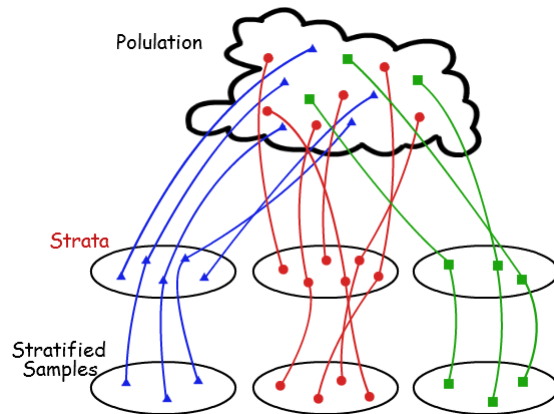
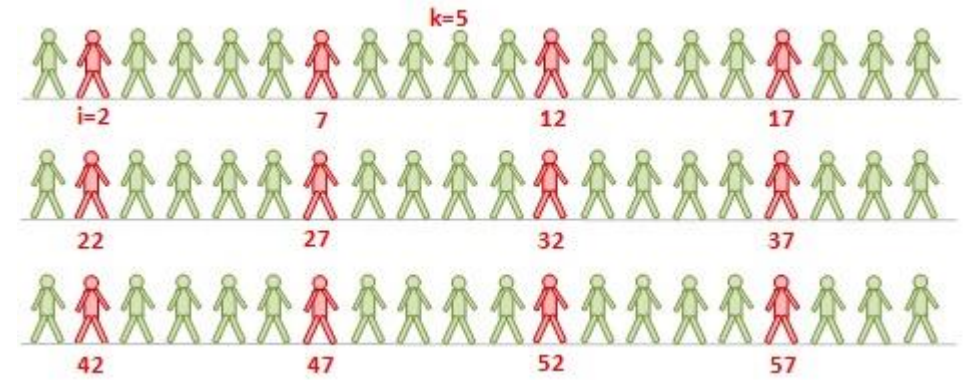
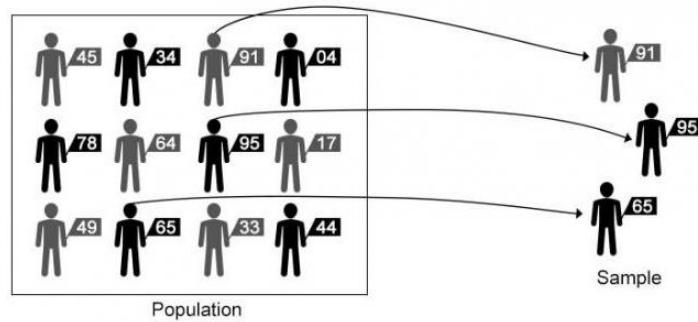
Muestreo Aleatorio Simple (MIA)

Muestreo Sistemático

Muestreo Estratificado

Muestreo Polietápico

MUESTREO PROBABILISTICO



MUESTREO PROBABILISTICO

Muestreo Aleatorio Simple (MIA)

La selección de la muestra se realiza en una sola etapa, directamente y sin reemplazamientos.

Aplicable en poblaciones pequeñas y plenamente identificables



Muestreo Sistemático

Se utiliza en muestras ordenadas.

Consiste en seleccionar al azar un elemento y a partir de él, incrementando un intervalo fijo, seleccionar toda la muestra.



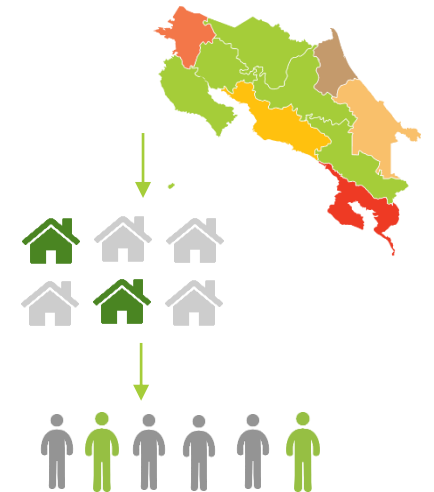
Muestreo Estratificado

Se dividen los elementos del universo en grupos (estratos), donde el conjunto de los grupos forman la totalidad del universo.



Muestreo Polietápico

El muestreo se realiza en varias etapas y en cada una de ellas es posible utilizar el mismo o distinto tipo de muestreo.



MUESTREO PROBABILÍSTICO

Muestreo Aleatorio Simple (MIA)

La selección de la muestra se realiza en una sola etapa, directamente y sin reemplazamientos.

Se aplica fundamentalmente en investigaciones sobre poblaciones pequeñas y plenamente identificables, por ejemplo **cuando disponemos de la lista completa de todos los elementos del universo.**

Ventajas

- Es sumamente sencillo y de fácil comprensión.
- Permite el cálculo rápido de medias y varianzas.
- Está basado en la teoría estadística, por esta razón existen paquetes informáticos analizar todos los datos.

Desventajas

- Requiere que se posea un listado completo de toda la población.
- El trabajar con muestras pequeñas es posible que no represente a la población adecuadamente.

Muestreo Sistemático

Se utiliza en muestras ordenadas. Consiste en seleccionar al azar un elemento y a partir de él, incrementando un intervalo fijo, seleccionar toda la muestra.

Procedimiento

N: Tamaño del universo n: el tamaño de la muestra

Primero se determina de qué tamaño es la muestra, una vez determinado el tamaño, se calcula el coeficiente de elevación, esto es, el cociente, la división, entre el tamaño del universo y el tamaño de la muestra.

Se opta por un número al azar que sea como el máximo coeficiente de elevación, un número entre 1 y K, a este número se le nombrará "i".

Se debe de contar con un listado de todos los elementos del universo, donde $i+K$, $i+2K$, $i+3K$, $i+4K$, $i+5K$... de tal manera que se continúe hasta completar toda la muestra.

Ventajas

- Es de fácil aplicación.
- En algunas ocasiones no es necesario tener un listado de toda la población.

Desventajas

- Si la constante de muestreo está asociada con el fenómeno de interés, se pueden hallar estimaciones desniveladas.

MUESTREO PROBABILÍSTICO

Muestreo Estratificado

Se basa en dividir los elementos del universo en grupos, donde cada elemento del universo pertenece a un solo grupo, y el conjunto de los grupos forman la totalidad del universo. A cada grupo lo llamamos estrato.

Cuanto más homogéneos seas los estratos, más precisas resultaran las estimaciones.

Ventajas

- Se pueden obtener estimaciones más precisas y confiables.
- Puede realizar diferentes métodos de muestreo y/o estimación en determinados estratos, lo que permite reducir la varianza (desviación estándar) y por tanto disminuir el tamaño de la muestra.
- Facilita la coordinación y realización de los trabajos de campo.

Desventajas

- Los análisis son complicados

Muestreo Polietápico

Se realiza cuando es imposible el muestreo aleatorio simple debido al tamaño de la población.

En este tipo de muestreos la unidad muestral no son los elementos del universo, sino un conjunto de elementos que bajo determinados aspectos se puede considerar que forman una unidad.

Ventajas

- Resulta muy eficiente cuando la población es enorme.
- Minimiza Costos.
- No es necesario tener un listado de toda la población, pueden ser necesario solo de las unidades primarias del muestreo.

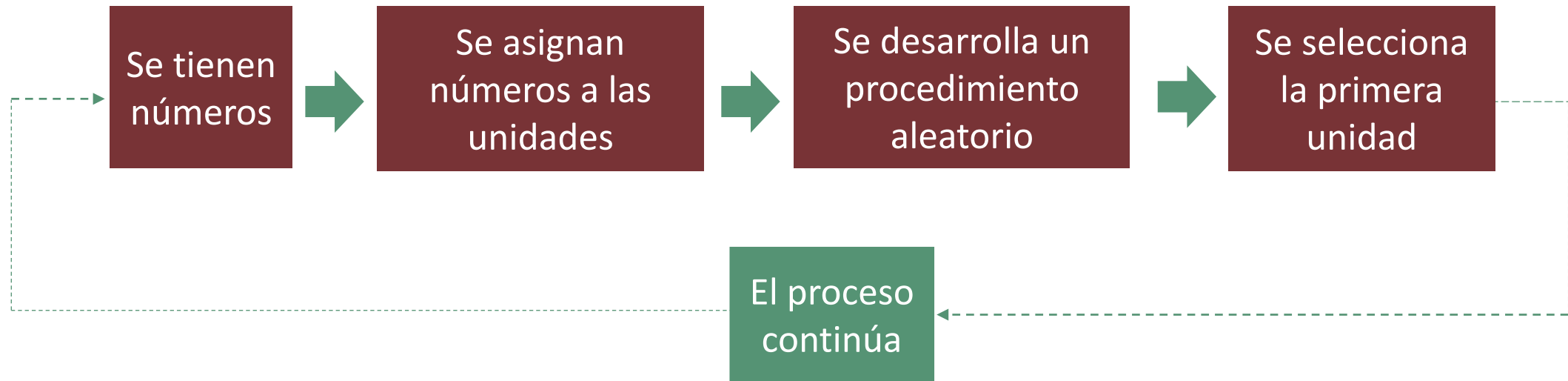
Desventajas

- El error estándar es mayor que en el muestreo aleatorio simple o estratificado.
- Es complejo su cálculo de error estándar.

ALEATORIZACIÓN

- El **muestreo probabilístico** requiere de un procedimiento mecánico que asigne las probabilidades a las UE o UM.

- La aleatorización es una operación matemática y debe ser guiada mediante con una operación probabilística.



Índice

1

Fundamentos del
diseño de la
muestra

4

Error de muestreo

2

La inferencia
estadística

3

Muestreo:
probabilístico y no
probabilístico

ERRORES DE MUESTREO

Muestreo no probabilístico

- Los errores son sistemáticos y van en una sola dirección
- Produce lo que se denomina sesgo de selección

Muestreo probabilístico

- Los errores son aleatorios y se anulan entre sí
- Esos errores pueden medirse con un modelo, mediante errores de muestreo

MUESTRAS REPRESENTATIVAS

1. ¿Qué es una muestra representativa?

- Es un conjunto de elementos que se proyectan a la población
- Es un conjunto de elementos que se seleccionan usando muestreo probabilístico y que representan a la población en sus características de estudio

2. ¿Cómo asegurar que una muestra es representativa?

- Conociendo la población
- Usando métodos de selección de muestras representativas
- Aceptando cierto grado de error (confianza del 95%, por ejemplo)

3. ¿Cómo seleccionar muestras representativas?

El muestreo al azar es preferido porque:

- Elimina los sesgos de selección
- Produce errores de muestreo que pueden medirse con modelos probabilísticos
- Controla los errores de muestreo (incremento de n)
- Puede calcularse la precisión de las estimaciones²⁸

CRITERIOS PARA EL DISEÑO DE LA MUESTRA

Medibilidad

Permite calcular, a partir de la propia muestra, estimaciones válidas o aproximadas de su variabilidad de muestreo.

Economía

Se requiere cumplir el objetivo a un costo mínimo. El problema del costo y la precisión.

Orientación hacia la meta

El diseño (selección y estimación) deben orientarse a los objetivos de la investigación, hechos a la medida del diseño de la encuesta y ajustado a las condiciones de la encuesta.

Practicidad

Se requiere de cuidado para traducir el modelo de selección teórico a un conjunto de instrucciones de oficina y de campo.

Las instrucciones deben ser simples, claras, prácticas y completas.

Ejemplo: localización de UPM, selección de personas a entrevistar, procedimientos para viviendas que no están en el marco, etc.

Índice

1

Fundamentos del
diseño de la
muestra

4

Error de muestreo

2

La inferencia
estadística

5

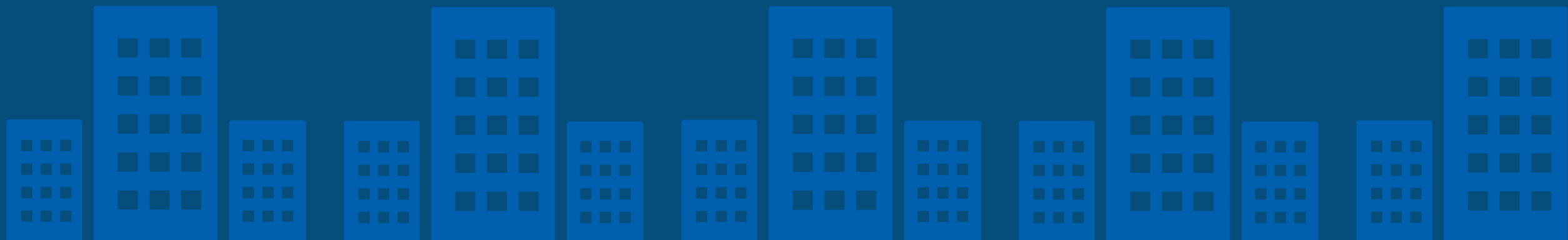
Tamaño de muestra
(n)

3

Muestreo:
probabilístico y no
probabilístico

2.

TAMAÑO DE LA
MUESTRA



TAMAÑO DE MUESTRA

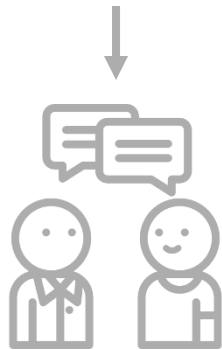
La pregunta mas común

**¿Que tamaño de muestra
debo seleccionar para hacer
una encuesta?**



**¿La pregunta hecha proporciona
toda la información que se requiere?**

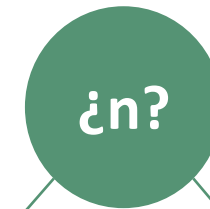
Evidentemente no. No se pregunto lo que debía...



Si el muestrista responde
inmediatamente

Comete un error

Crear que provee la respuesta correcta
a un problema mal planteado.



**Si es muy
grande ...**

... hay
despilfarro
de recursos

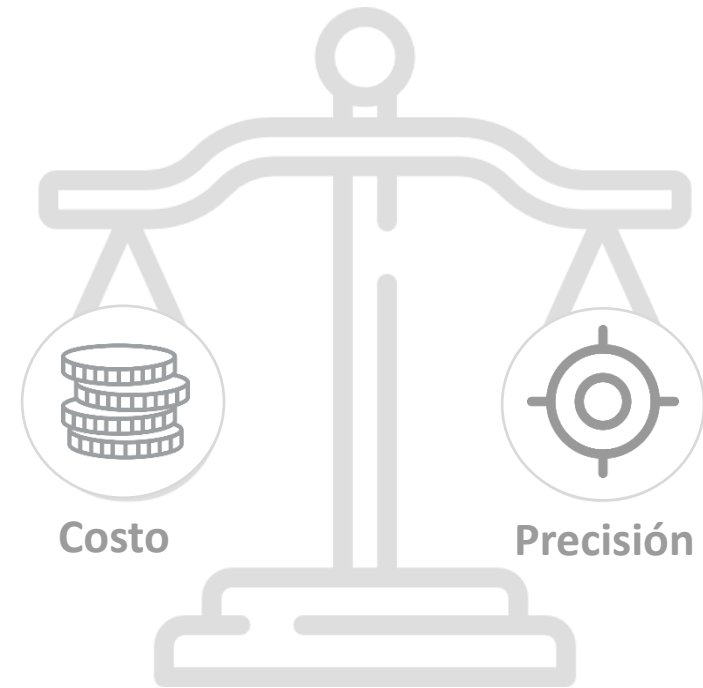
**Si es muy
pequeña ...**

... disminuye
la utilidad de
los recursos

TAMAÑO DE MUESTRA



Aunque no siempre hay suficiente información ...



PREGUNTAS CLAVE

Concretamente se requiere:

01 Decisiones

Lo que se espera de n en la practica (decisiones que se tomaran con la muestra, limites de error deseados, precisión)

02 Ecuación

Encontrar una ecuación que relacione n con la precisión deseada y los limites de error

Esa ecuación tiene valores desconocidos que deben estimarse o aproximarse

03 Balance

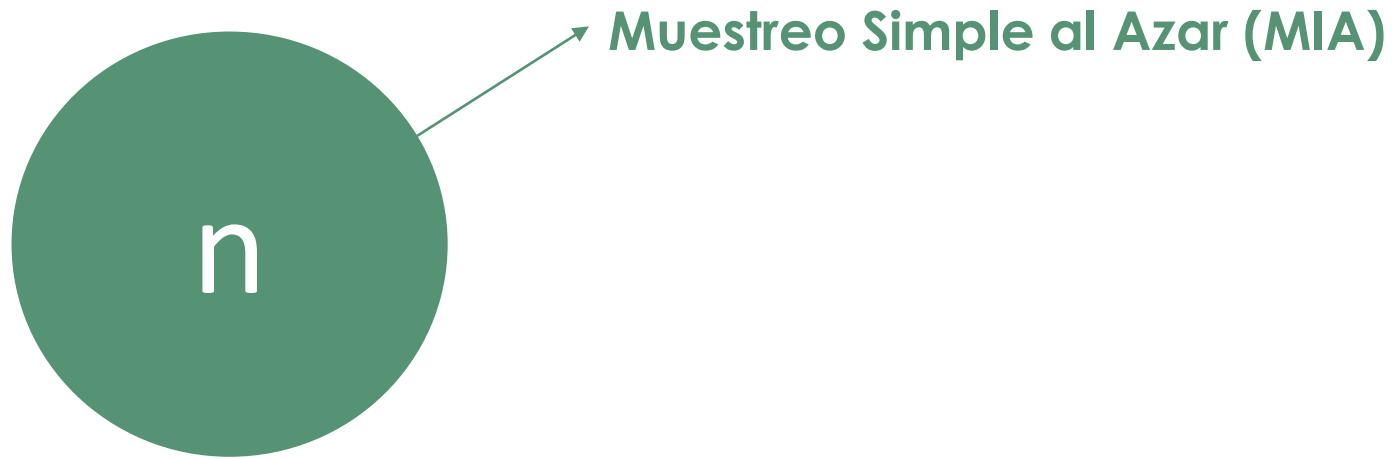
n debe ser consistente con costo, tiempo, trabajo, precisión y la significancia

$n = f(\text{cos, tie, tra, pre, sig})$

DECISIÓN CLAVE



TAMAÑO DE MUESTRA



MUESTREO ALEATORIO SIMPLE (MIA)

Ideal cuando se posee una o varias observaciones que en su distribución tienden a una función normal.

Condiciones para el cálculo de la muestra:

- El cálculo del tamaño de muestra parte de un muestreo simple al azar.
- Establecer nivel de confianza, 95% por ejemplo.
- Determinar el error estándar de estimación. Para variables que serán analizadas como porcentajes, se toma la variancia cuando se maximiza en $p = 0.5$.

$$n_{mia} = \frac{Z_{\alpha/2} * p * q}{d^2}$$

MUESTREO ALEATORIO SIMPLE (MIA)

Ejemplo

OBJETIVO Medir la satisfacción de beneficiarios de servicios de comedor y transporte en colegios públicos de CR.

POBLACIÓN Estudiantes beneficiarios de servicios de comedor y transporte en colegios públicos de CR que pertenecen a cantones con **Índice de Desarrollo Sostenible** muy bajo, durante el año "20XX".

- Muestreo simple al azar.
 - Nivel de confianza del 95%.
 - Error máximo permisible de estimación $d = 0.05$.
 - Error estándar de estimación, $p = 0.5$
- variables que se analizarán como porcentaje.

$$\begin{aligned}n_{mia} &= \frac{Z_{\alpha/2} * p * q}{d^2} \\&= \frac{1,96^2 * 0,5 * 0,5}{0,05^2} \\&= \mathbf{384}\end{aligned}$$

Índice

1

Fundamentos del
diseño de la
muestra

4

Error de muestreo

2

La inferencia
estadística

5

Tamaño de muestra
(n)

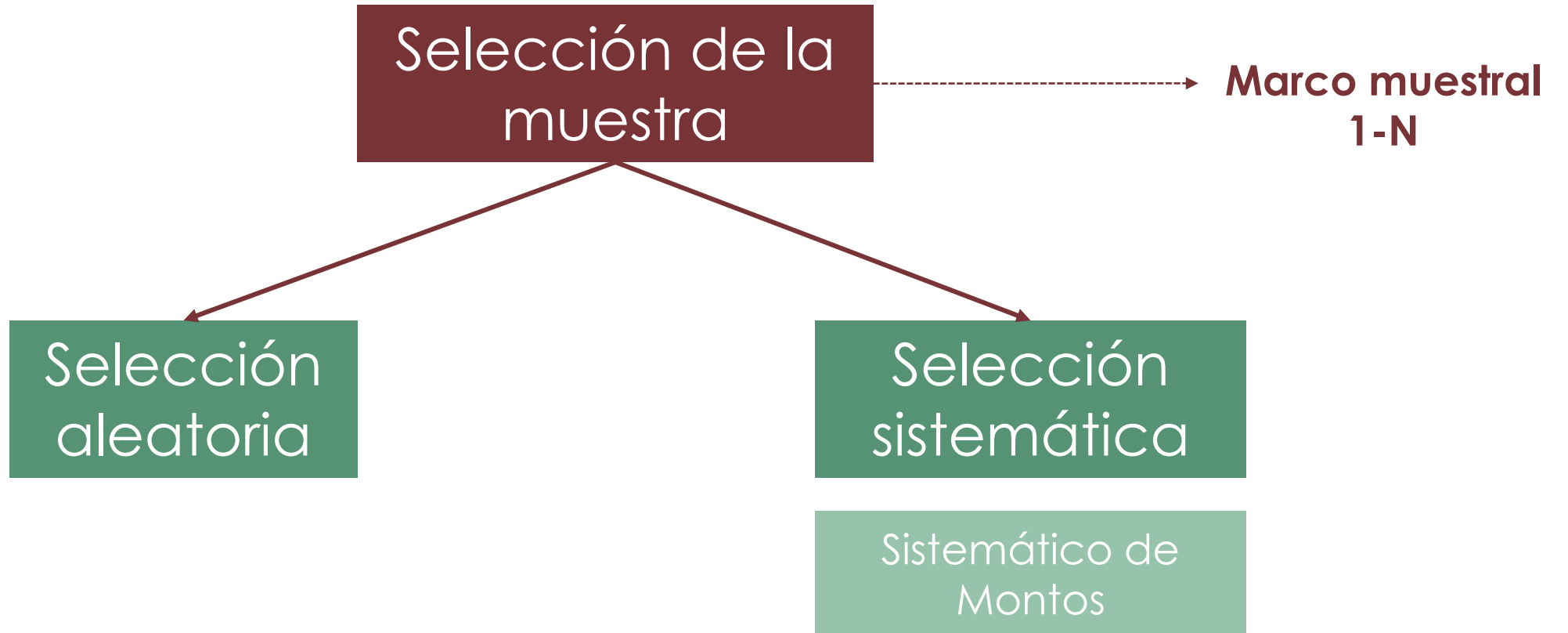
3

Muestreo:
probabilístico y no
probabilístico

6

Selección de las UE
de la n

SELECCIÓN DE LA MUESTRA



SELECCIÓN ALEATORIA

Números generadores a partir de programas informáticos, (Excel, SPSS, R u otro)

| | A | B | C | D |
|----|----|-------------------------------------|---|---|
| 1 | n | | | |
| 2 | 1 | 1117 | | |
| 3 | 2 | 544 | | |
| 4 | 3 | 1029 | | |
| 5 | 4 | 525 | | |
| 6 | 5 | 833 | | |
| 7 | 6 | 898 | | |
| 8 | 7 | 1141 | | |
| 9 | 8 | =ALEATORIO.ENTRE(1;1200) | | |
| 10 | 9 | ALEATORIO.ENTRE(inferior; superior) | | |
| 11 | 10 | | | |

Tablas de números aleatorios, este último método no es muy eficiente.

TABLA DE NÚMEROS ALEATORIOS

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 8 | 9 | 0 | 1 | 9 | 9 | 6 | 7 | 1 | 6 | 4 | 1 | 5 | 4 | 6 | 9 | 2 | 7 | 4 | 7 | 8 | 3 | 6 | 5 |
| 2 | 5 | 8 | 3 | 7 | 4 | 8 | 5 | 9 | 2 | 6 | 4 | 2 | 4 | 0 | 7 | 0 | 0 | 8 | 8 | 2 | 9 | 0 | 9 | 2 |
| 3 | 4 | 4 | 2 | 1 | 1 | 6 | 1 | 8 | 8 | 6 | 6 | 6 | 2 | 6 | 9 | 2 | 6 | 4 | 9 | 0 | 7 | 6 | 4 | 8 |
| 4 | 8 | 0 | 4 | 9 | 5 | 5 | 3 | 6 | 5 | 2 | 3 | 3 | 1 | 3 | 2 | 0 | 6 | 1 | 8 | 4 | 9 | 6 | 4 | 8 |
| 5 | 4 | 3 | 2 | 1 | 5 | 5 | 1 | 6 | 8 | 9 | 2 | 6 | 5 | 0 | 2 | 5 | 2 | 7 | 3 | 3 | 0 | 5 | 6 | 4 |
| 6 | 8 | 4 | 4 | 9 | 8 | 5 | 0 | 8 | 2 | 7 | 9 | 7 | 1 | 1 | 8 | 3 | 1 | 0 | 0 | 6 | 3 | 7 | 9 | 8 |
| 7 | 2 | 9 | 2 | 7 | 4 | 9 | 0 | 5 | 0 | 5 | 0 | 7 | 4 | 2 | 8 | 6 | 1 | 3 | 1 | 6 | 6 | 0 | 7 | 4 |
| 8 | 0 | 5 | 5 | 0 | 5 | 4 | 1 | 3 | 8 | 9 | 3 | 9 | 8 | 5 | 7 | 8 | 2 | 7 | 5 | 8 | 2 | 4 | 8 | 1 |
| 9 | 2 | 2 | 8 | 8 | 9 | 8 | 3 | 4 | 0 | 1 | 9 | 2 | 3 | 7 | 5 | 9 | 1 | 6 | 2 | 8 | 6 | 1 | 8 | 2 |
| 10 | 1 | 7 | 0 | 7 | 9 | 6 | 2 | 3 | 7 | 1 | 3 | 9 | 0 | 8 | 9 | 7 | 5 | 6 | 8 | 8 | 7 | 4 | 5 | 4 |
| 11 | 9 | 7 | 0 | 5 | 9 | 7 | 3 | 0 | 8 | 0 | 1 | 3 | 2 | 9 | 0 | 6 | 7 | 2 | 8 | 7 | 4 | 2 | 7 | 2 |
| 12 | 2 | 7 | 2 | 8 | 3 | 6 | 0 | 1 | 8 | 2 | 1 | 8 | 4 | 3 | 5 | 0 | 5 | 4 | 4 | 1 | 0 | 3 | 0 | 7 |
| 13 | 0 | 7 | 1 | 8 | 1 | 3 | 7 | 1 | 1 | 1 | 6 | 1 | 7 | 0 | 3 | 2 | 6 | 2 | 2 | 9 | 7 | 6 | 0 | 6 |
| 14 | 1 | 4 | 3 | 2 | 0 | 8 | 5 | 4 | 7 | 5 | 7 | 3 | 6 | 8 | 0 | 8 | 9 | 0 | 7 | 1 | 0 | 8 | 6 | 7 |
| 15 | 2 | 2 | 2 | 1 | 1 | 0 | 4 | 6 | 5 | 2 | 0 | 4 | 9 | 3 | 7 | 7 | 6 | 5 | 0 | 2 | 7 | 9 | 0 | 5 |
| 16 | 2 | 3 | 6 | 0 | 2 | 6 | 7 | 2 | 5 | 6 | 7 | 8 | 5 | 7 | 2 | 3 | 1 | 2 | 7 | 7 | 3 | 8 | 7 | 0 |
| 17 | 2 | 5 | 6 | 0 | 0 | 5 | 7 | 3 | 3 | 0 | 7 | 1 | 3 | 5 | 3 | 9 | 6 | 8 | 3 | 7 | 3 | 3 | 2 | 9 |
| 18 | 4 | 4 | 4 | 5 | 0 | 1 | 2 | 8 | 2 | 5 | 4 | 2 | 7 | 7 | 1 | 2 | 0 | 7 | 2 | 3 | 3 | 0 | 7 | 0 |
| 19 | 4 | 4 | 8 | 1 | 0 | 5 | 3 | 4 | 5 | 7 | 3 | 4 | 7 | 4 | 2 | 1 | 0 | 8 | 2 | 4 | 1 | 7 | 4 | 5 |
| 20 | 7 | 9 | 8 | 3 | 7 | 1 | 3 | 0 | 3 | 0 | 4 | 3 | 9 | 6 | 9 | 9 | 9 | 0 | 1 | 8 | 1 | 0 | 2 | 5 |
| 21 | 7 | 9 | 2 | 9 | 3 | 2 | 6 | 7 | 5 | 5 | 6 | 9 | 3 | 6 | 9 | 2 | 6 | 6 | 2 | 0 | 0 | 4 | 3 | 8 |
| 22 | 8 | 7 | 3 | 8 | 0 | 8 | 1 | 6 | 6 | 6 | 1 | 8 | 5 | 8 | 5 | 6 | 6 | 5 | 5 | 4 | 9 | 1 | 0 | 1 |
| 23 | 8 | 2 | 5 | 1 | 0 | 0 | 0 | 8 | 0 | 5 | 1 | 5 | 6 | 5 | 7 | 9 | 4 | 3 | 6 | 3 | 6 | 1 | 0 | 1 |
| 24 | 6 | 3 | 8 | 8 | 6 | 4 | 1 | 6 | 3 | 8 | 3 | 8 | 1 | 6 | 1 | 7 | 6 | 6 | 3 | 9 | 8 | 9 | 3 | 3 |
| 25 | 7 | 1 | 0 | 0 | 5 | 4 | 0 | 7 | 3 | 7 | 0 | 0 | 9 | 1 | 9 | 0 | 2 | 5 | 0 | 5 | 9 | 7 | 9 | 3 |

SELECCIÓN SISTEMÁTICA DE UNIDADES

Los elementos son tomados en cuenta para la muestra mediante un espaciamiento sistemático.

Ventaja frente a la selección aleatoria

Facilidad de la selección de las unidades de estudio.

El muestreo sistemático es más potente que el muestreo simple cuando se posee un listado ordenado de los elementos.

SELECCIÓN SISTEMÁTICA DE UNIDADES

PASOS PARA LA SELECCIÓN

1. Ordenar las unidades de estudio de 1 al N (N total de las unidades de estudio)
De 1 hasta 1 200
2. Determinar el total de las unidades de estudio, en este caso el N.
 $N=1\ 200$
3. Determinar el tamaño de muestra, en este caso el n.
 $n=46$
4. **Calcular el rango sistemático k**, el cual resulta del entero más próximo dada la división de N/n .
 $K=N/n=1\ 200/46 = 26,08 = 26$

SELECCIÓN SISTEMÁTICA DE UNIDADES

PASOS PARA LA SELECCIÓN

5. K0 de arranque, escogemos al azar un número K0 entre 1 y k (utilizando los números aleatorios, sacar una bola de un bombo, entre otros.).

$K0 = \text{aleatorio entre } 1 \text{ y } K(26)$

$K0 = 20$

6. Determinar los casos a seleccionar, la muestra será el elemento i y los elementos $K0+k$, $K0+2k$, etc. Es decir, el elemento k y los elementos a intervalos fijos k hasta conseguir los n sujetos.

Primera selección = $K0 = 20$

Segunda selección = $K0 + K = 20 + 26 = 46$

Tercera selección = $K0 + 2k = 20 + (2*26) = 72$

Cuarta selección = $K0 + 3k = 20 + (3*26) = 98$

....

y así hasta seleccionar los “n” elementos.

SELECCIÓN SISTEMÁTICA DE UNIDADES

7. Seleccionar los casos determinados

Primera selección = 20
Segunda selección = 46
Tercera selección = 72
Cuarta selección = 98

Importante: En caso de realizar asignación proporcional por cantón por ejemplo, se debe realizar un k y K_0 para cada cantón.

| ID | PROYECTO | SELECCIÓN |
|-----|----------|-----------|
| 1 | X01 | |
| 2 | X02 | |
| 3 | X03 | |
| 4 | X04 | |
| ... | | |
| 15 | X15 | |
| 16 | X16 | |
| 17 | X17 | |
| 18 | X18 | |
| 19 | X19 | |
| 20 | X20 | X |
| 21 | X21 | |
| 22 | X22 | |
| ... | | |
| 44 | X44 | |
| 45 | X45 | |
| 46 | X46 | X |
| 47 | X47 | |
| 48 | X48 | |
| ... | | |

| | | |
|-----|------|---|
| ... | | |
| 71 | X71 | |
| 72 | X72 | X |
| 73 | X73 | |
| 74 | X74 | |
| ... | | |
| 96 | X96 | |
| 97 | X97 | |
| 98 | X98 | X |
| 99 | X99 | |
| 100 | X100 | |

Hasta seleccionar los 46 elementos

SELECCIÓN SISTEMÁTICA ACUMULADO O DE MONTOS

Mantiene los principios básicos de la selección sistemática de unidades presentada anteriormente.

PASOS PARA LA SELECCIÓN

1. Ordenar el marco muestral según la variable de interés

Ejemplo: monto de las facturas o cantidad de estudiantes

2. Se acumulada la variable de interés

Siguiendo con el caso anterior podríamos utilizar el monto de inversión por proyecto

| ID | PROYECTO | MONTO | MONTO ACUMULADO | SELECCIÓN |
|----|----------|--------------|-----------------|-----------|
| 1 | X01 | 2.000.000,00 | 2.000.000,00 | |
| 30 | X30 | 4.000.000,00 | 6.000.000,00 | |
| 36 | X36 | 4.000.000,00 | 10.000.000,00 | |
| 38 | X38 | 4.000.000,00 | 14.000.000,00 | |
| 39 | X39 | 4.000.000,00 | 18.000.000,00 | |
| 47 | X47 | 4.000.000,00 | 22.000.000,00 | |
| 54 | X54 | 4.000.000,00 | 26.000.000,00 | |
| 57 | X57 | 4.000.000,00 | 30.000.000,00 | |

2 000 000 + 4 000 000 + 4 000 000

SELECCIÓN SISTEMÁTICA ACUMULADO O DE MONTOS

3. Se calcula el intervalo de espaciamento:

$$k = \text{intervalo muestral} = \frac{VL}{N}$$

VL es la suma total de la variable de interés
N el total de unidades de la población

$$k = \frac{VL}{N}$$

$$= \frac{13\,219\,006\,666,66}{1200}$$

$$= 11\,015\,838,88 \approx 11\,015\,839$$

SELECCIÓN SISTEMÁTICA ACUMULADO O DE MONTOS

4. Determinar K0 de arranque, escogemos al azar un número K0 entre 1 y k (utilizando los números aleatorios, sacar una bola de un bombo, entre otros.).

$K0 = \text{aleatorio entre } 1 \text{ y } (K = 11\,015\,839)$

$K0 = 10\,564\,901,00$

5. Se determinan las partidas a seleccionar.

Este método produce que los montos más grandes sean seleccionados con mayor probabilidad, y los más pequeños tengan con menor posibilidad para ser seleccionados.

Primera selección = $K0 = 10\,564\,901$

Segunda selección = $K0 + K = 10\,564\,901,00 + 11\,015\,839 = 21\,580\,740$

Tercera selección = $K0 + 2k = 10\,564\,901,00 + (11\,015\,839 * 2) = 32\,596\,579$

Cuarta selección = $K0 + 3k = 10\,564\,901,00 + (11\,015\,839 * 3) = 43\,612\,418$

....

y así hasta seleccionar los "n" elementos.

SELECCIÓN SISTEMÁTICA ACUMULADO O DE MONTOS

6. Seleccionar los casos determinados

Primera selección = **10 564 901**

Segunda selección = **21 580 740**

Tercera selección = **32 596 579**

Cuarta selección = **43 612 418**

....

y así hasta seleccionar los "n" elementos.

| ID | PROYECTO | MONTO | MONTO ACUMULADO | SELECCIÓN |
|----|----------|--------------|-----------------|-----------|
| 1 | X01 | 2.000.000,00 | 2.000.000,00 | |
| 30 | X30 | 4.000.000,00 | 6.000.000,00 | |
| 36 | X36 | 4.000.000,00 | 10.000.000,00 | |
| 38 | X38 | 4.000.000,00 | 14.000.000,00 | X |
| 39 | X39 | 4.000.000,00 | 18.000.000,00 | |
| 47 | X47 | 4.000.000,00 | 22.000.000,00 | X |
| 54 | X54 | 4.000.000,00 | 26.000.000,00 | |
| 57 | X57 | 4.000.000,00 | 30.000.000,00 | |
| 65 | X65 | 4.000.000,00 | 34.000.000,00 | X |
| 73 | X73 | 4.000.000,00 | 38.000.000,00 | |
| 74 | X74 | 4.000.000,00 | 42.000.000,00 | |
| 77 | X77 | 4.000.000,00 | 46.000.000,00 | X |
| 80 | X80 | 4.000.000,00 | 50.000.000,00 | |
| 91 | X91 | 4.000.000,00 | 54.000.000,00 | |
| 92 | X92 | 4.000.000,00 | 58.000.000,00 | |
| 94 | X94 | 4.000.000,00 | 62.000.000,00 | |

DUDA: DIVISION NO DEBERIA SER POR "n" ?? YA QUE DA SALTOS MUY PEQUEÑOS

The End