



DISEÑO MUESTRAL

Úrsula Torres Parejo

CONTENIDO

- Capítulo 1. Introducción al diseño y organización de encuestas (Pg. 2)
- Capítulo 2. Introducción a la teoría de muestras (Pg. 31)
- Capítulo 3. Tipos básicos de muestreo (Pg. 60)
- Capítulo 4. Muestreo complejo (Pg. 92)
- Capítulo 5. Cálculo del tamaño de muestra (Pg. 116)

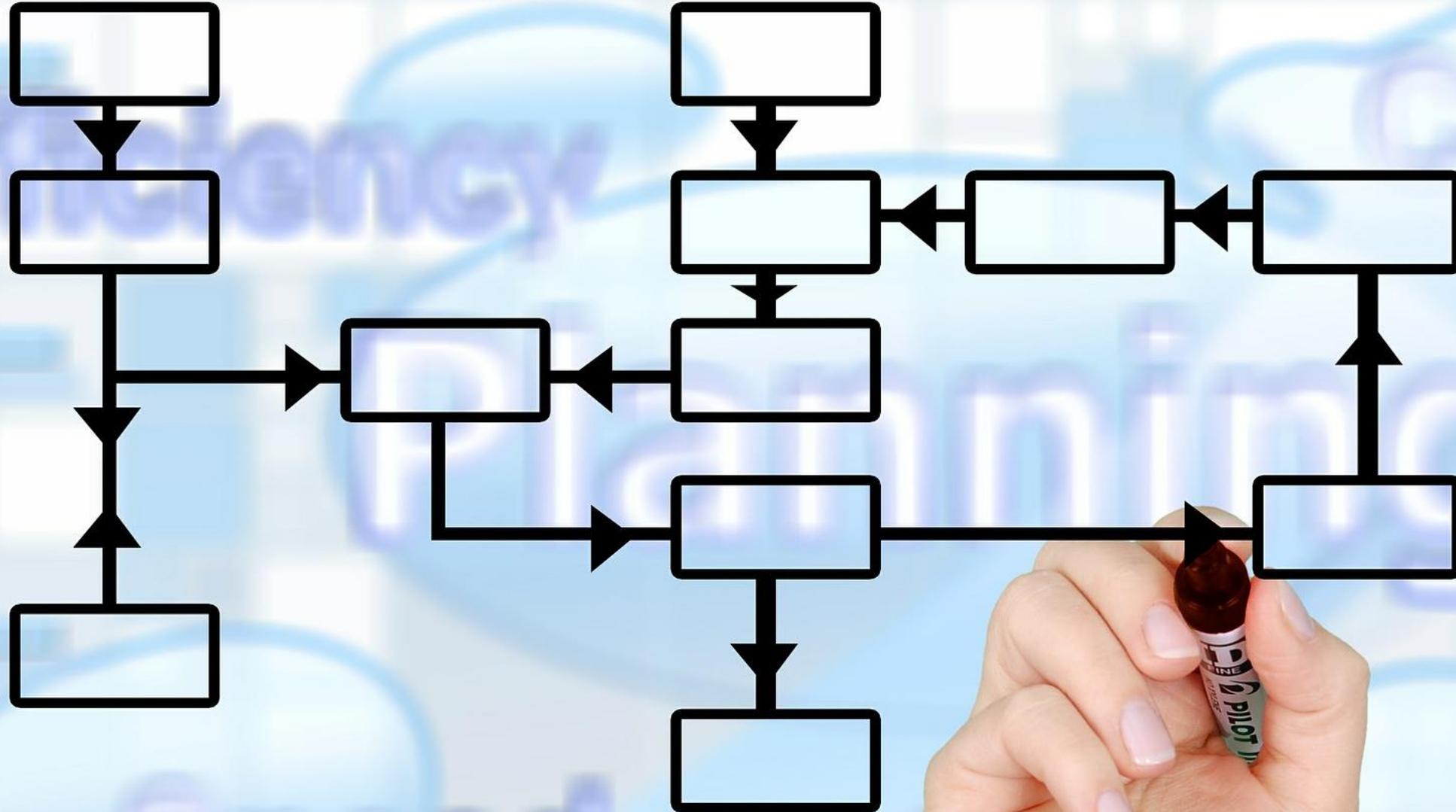


CAPÍTULO 1. INTRODUCCIÓN AL DISEÑO Y ORGANIZACIÓN DE ENCUESTAS

Úrsula Torres Parejo

Capítulo 1

1. Fases de una encuesta por muestreo
2. Diseño y gestión del cuestionario
3. La falta de respuesta
4. Depuración e imputación de datos



FASES DE UNA ENCUESTA POR MUESTREO

FASES DE UNA ENCUESTA POR MUESTREO

1. Diseño teórico

Análisis del problema y Objetivos

Ámbito de la investigación: zona geográfica, población estudiada, periodo de tiempo,...

Definición de unidades básicas: hogares, empresas, personas,...

Características objeto de estudio: ingresos, opiniones, preferencias.,

Características del cuestionario: extensión, duración,...

Procedimiento de muestreo

Tipo de muestreo

Tamaño de muestra

FASES DE UNA ENCUESTA POR MUESTREO

1. Diseño teórico



2. Diseño del cuestionario



Número de preguntas

Necesidad del pretest

FASES DE UNA ENCUESTA POR MUESTREO

1. Diseño teórico



2. Diseño del cuestionario



3. Diseño de la muestra

Marco

Jerarquía entre las unidades de muestreo:
estratos, clusters, etc.

Procedimiento de selección de muestra

FASES DE UNA ENCUESTA POR MUESTREO

1. Diseño teórico



2. Diseño del cuestionario



3. Diseño de la muestra



4. Recogida de información

Entrevista clásica o sobre papel: los/as encuestadores/as disponen de cuestionario en papel

Encuesta por correo: el cuestionario se envía por correo

Entrevista asistida por ordenador: los/as encuestadores/as van provistos/as de ordenador

Entrevista por teléfono: la entrevista se realiza por teléfono

Autoentrevista: uno/a mismo/a completa su propia entrevista con el ordenador

FASES DE UNA ENCUESTA POR MUESTREO

1. Diseño teórico



2. Diseño del cuestionario



3. Diseño de la muestra



4. Recogida de información



5. Depuración de datos

Corrección de errores: Detección de anomalías y errores

Evaluación de la calidad

FASES DE UNA ENCUESTA POR MUESTREO

1. Diseño teórico



2. Diseño del cuestionario



3. Diseño de la muestra



4. Recogida de información



5. Depuración de datos

10. Interpretación, presentación y publicación de los resultados



9. Análisis estadístico secundario



8. Evaluación de errores



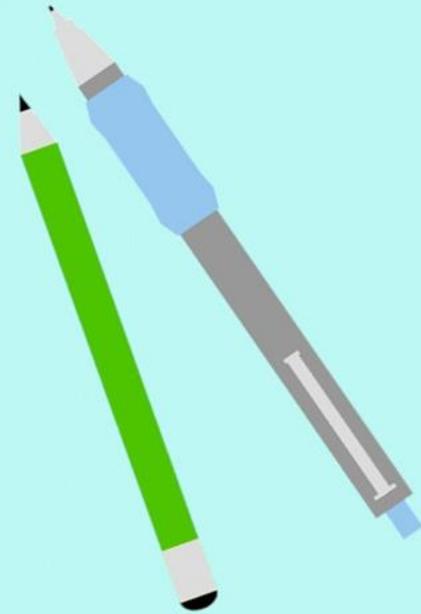
7. Análisis estadístico primario



6. Chequeo y validación de la muestra



Fases avanzadas



DISEÑO Y GESTIÓN DEL CUESTIONARIO

DISEÑO Y GESTIÓN DEL CUESTIONARIO

CONSIDERACIONES PREVIAS

- 1. Tema a tratar:** temas más importantes relacionados con la investigación.
- 2. Duración máxima:** se fija de antemano.
- 3. Cabecera y datos de la persona entrevistada:** menor número de datos personales posible.
- 4. Instrucciones al entrevistador/a:** para que todos los cuestionarios se pasen en las mismas condiciones.
- 5. Material auxiliar a utilizar:** diagramas, dibujos, fotos,...
- 6. Impresión, papel y diseño final:** importante en encuestas por correo.

DISEÑO Y GESTIÓN DEL CUESTIONARIO

CARACTERÍSTICAS DESEABLES DE LAS PREGUNTAS

- 1. Deben formularse en lenguaje entendible:** salvo en encuestas orientadas a profesionales o técnicos.
- 2. Han de ser lo más cortas posibles:** la lectura de las preguntas no debe sobrepasar el minuto.
- 3. Deben ser neutras:** para obtener resultados imparciales.
- 4. No deben incluir temas difíciles o en los que haya que utilizar cálculos.**
- 5. Se deben excluir palabras con connotaciones no deseables:** evitar términos discriminatorios o con significado negativo.
- 6. Su redacción debe invitar a la colaboración:** secuencia amena y despertar interés.
- 7. Las preguntas con contenido íntimo deben realizarse de forma indirecta.**

DISEÑO Y GESTIÓN DEL CUESTIONARIO

OTROS ASPECTOS IMPORTANTES

- 1. Se deben introducir preguntas control:** para verificar consistencia de las respuestas.
- 2. Especial cuidado con la primera pregunta:** debe ser interesante y de fácil respuesta.
- 3. Las preguntas más sencillas deben ir al principio, para complicarse al final:** fomenta la colaboración y sinceridad.
- 4. Debe seguirse un orden lógico:** preguntas agrupadas por temas con dificultad gradual.

DISEÑO Y GESTIÓN DEL CUESTIONARIO

REALIZACIÓN DEL PRETEST

El pretest se realiza sobre una submuestra pequeña, con el objetivo de detectar posibles anomalías observando de forma rápida:

1. Proporción de preguntas contestadas como “NS/NC”.
2. Proporción de preguntas sin contestar.
3. Proporción de preguntas que, disponiendo de más de 4 categorías, se contestan sólo en sus categorías extremas.

!!! En total estas proporciones no deben superar el 10% del cuestionario!!!

DISEÑO Y GESTIÓN DEL CUESTIONARIO

TIPOS DE PREGUNTAS

Preguntas abiertas



- Son poco influenciables
- Permiten interpretar otras cuestiones
- Son cómodas para las personas entrevistadas



- Respuestas no uniformes
- Respuestas no estrictas
- No se usan mismas palabras para mismas ideas.

Preguntas cerradas



- Número limitado de respuestas
- Facilita tratamiento estadístico
- Facilita codificación, depuración e imputación
- Disminuyen el tiempo
- Anotación homogénea



- Tienen como inconvenientes las ventajas de las preguntas abiertas



LA FALTA DE RESPUESTA

LA FALTA DE RESPUESTA

CAUSAS

- 1. Ausencia temporal de la persona seleccionada para entrevistar:** “los/as no en casa”.
- 2. Negación a colaborar:** personas que rehúsan ser entrevistadas.
- 3. Falta de conocimientos de la persona entrevistada:** desconocimiento de la respuesta.
- 4. Pérdida de la información:** extravío de la información ya recogida.
- 5. No cubrimiento:** áreas que quedan sin cubrir por condiciones ambientales, escasa facilidad de transporte, etc.

LA FALTA DE RESPUESTA

TRATAMIENTO PRIMARIO (FASE DE CAMPO)

Encuestas repetidas

- Técnica más utilizada y limpia para reducir la no respuesta.
- Consiste en repetir las visitas o llamadas hasta un número mínimo.

La encuesta delegada

- Consiste en elegir los datos de una unidad alternativa, por ejemplo, otro miembro de la familia.
- Método no peligroso para ciertas encuestas.

Sustitución de unidades

- Lo más recomendable es la selección de unidades de sustitución de forma aleatoria.
- Procedimiento no recomendable ya que puede introducir nuevos sesgos.

El uso de incentivos

- La planificación del incentivo se basa en el coste de las llamadas adicionales que habrían que utilizarse en realizar revisitas.
- Inconveniente de que se dé información errónea para conseguir el incentivo.



DEPURACIÓN E IMPUTACIÓN DE DATOS

DEPURACIÓN E IMPUTACIÓN DE DATOS

ERRORES AJENOS AL MUESTREO

Respuestas
intencionalmente
incorrectas

Respuestas
incorrectas por
cansancio

Errores de
transcripción de
la información

Desconfianza
Sentimientos de
culpabilidad

Poca claridad del
cuestionario
Mala actuación del
entrevistador/a

Introducción de
información en equipos
informáticos



DEPURACIÓN E IMPUTACIÓN DE DATOS

TIPOS DE ERRORES

Aleatorios. Suelen compensarse unos con otros.

Erráticos

Ej:
Equivocación al teclear una cifra

Si se repite la encuesta en las mismas condiciones se vuelven a repetir.

Sistemáticos

Ej: Engaño intencionado

Depuración de datos

DEPURACIÓN E IMPUTACIÓN DE DATOS

CONVENIENCIA DE LA DEPURACIÓN

Argumentos a favor

- Facilita fases posteriores de tratamiento estadístico.
- Tratamiento adecuado del “No sabe/No contesta”
- Los datos depurados inspiran confianza al lector.

Argumentos en contra

- Producción de errores adicionales
- Descuido de la recogida de información.
- Falsa sensación de corrección

DEPURACIÓN E IMPUTACIÓN DE DATOS

MACRO-DEPURACIÓN VERSUS MICRO-DEPURACIÓN

Macro-depuración.

Se ignoran errores cuya magnitud no es significativa

Micro-depuración.

Técnica minuciosa para detectar un mayor número de errores

DEPURACIÓN E IMPUTACIÓN DE DATOS

IMPUTACIÓN AUTOMÁTICA

Trata de estimar una serie de valores que palién, en cierta forma, la falta de respuesta en un estrato o subestrato de la muestra, de acuerdo con informaciones previas que se disponga de las personas que no responden, o en base a los valores observados de estas mismas personas.

1. Reduce el sesgo de las estimaciones
2. Facilita los análisis de datos posteriores
3. Incrementa la consistencia de los resultados de los distintos tipos de análisis
4. Produce estimaciones más precisas

DEPURACIÓN E IMPUTACIÓN DE DATOS

SUPUESTOS DE LOS PROCEDIMIENTOS DE IMPUTACIÓN

Al estimar los valores de los campos a imputar en base al valor de los restantes campos el supuesto implícito que se hace es que el comportamiento de las unidades cuyos registros están incompletos, es el mismo que el comportamiento de las unidades cuyos registros tienen dato en todos sus campos.

Supuesto sobre la falta de errores en el contenido:

Se supone que los campos con dato no tienen errores de contenido

Supuesto sobre los campos a imputar:

Se supone que la falta de respuesta es aleatoria y no intencionada

DEPURACIÓN E IMPUTACIÓN DE DATOS

PROCEDIMIENTOS DE IMPUTACIÓN CON REGISTRO DONANTE

Registro candidato o receptor

Registro donante

	A	B	C	D	E
1	id	sexo	educ	catlab	salario
2	1	2	15	3	57000
3	2	2	16	1	40200
4	3	1	12	1	21450
5	4	1	8	1	21900
6	5	2	15	1	
7	6	2	15	1	32100
8	7	2	15	1	36000
9	8	1	12	1	21900
10	9	1	15	1	27900
11	10	1	12	1	24000
12	11	1	16	1	30300



	A	B	C	D	E
1	id	sexo	educ	catlab	salario
2	1	2	15	3	57000
3	2	2	16	1	40200
4	3	1	12	1	21450
5	4	1	8	1	21900
6	5	2	15	1	32100
7	6	2	15	1	32100
8	7	2	15	1	36000
9	8	1	12	1	21900
10	9	1	15	1	27900
11	10	1	12	1	24000
12	11	1	16	1	30300



Campo a imputar

Estos métodos asignan a los campos faltantes de un registro el valor que tienen tales campos en otro registro de la encuesta. A los registros completos se les denomina “registros donantes” y a los registros con campos a imputar “registros receptores” o “registros candidatos”.

DEPURACIÓN E IMPUTACIÓN DE DATOS

PROCEDIMIENTOS DE IMPUTACIÓN CON REGISTRO DONANTE

COLD-DECK

Los registros donantes son registros tipo en base a fuentes de información externa como datos históricos

HOT-DECK

El registro donante es el inmediatamente anterior del mismo estrato al registro candidato

HOT-DECK con donante aleatorio

El registro donante es aleatorio

DONOR

Se establece una función distancia entre donantes y candidatos para seleccionar el registro donante.

OTROS

Por ejemplo imputación con la media. Para este procedimiento no se utiliza un registro donante.

REFERENCIAS

1. Abascal, E. y Grando, I. Análisis de Encuestas. ESIC Editorial, 2005
2. Blanco, C. Encuesta y Estadística. Métodos de Investigación Cuantitativa en Ciencias Sociales y Comunicación. Editorial Brujas, 2011
3. Valencia Delfa, J.L., Díaz-Llanos y Sainz-Calleja, F.J. Técnicas avanzadas en el tratamiento estadístico de encuestas de opinión, 2009
4. Bourke, J., Kirby, A. y Doran, J. Survey & Questionnaire Design. Oak Tree Press, 2000
5. Brace, I. Questionnaire design: how to plan, structure and write survey material for effective market research. Kogan Page Publishers, 2018
6. Amón Uribe, I. Guía metodológica para la selección de técnicas de depuración de datos. Universidad Nacional de Colombia, 2009
7. Krosnick, J.A. Questionnaire design. The Palgrave handbook of Survey Research, Springer, 2018.

Todas las imágenes obtenidas de:

- Imágenes prediseñadas de Microsoft PowerPoint versión 2010
- Imágenes gratis sin atribución de Pixabay (<https://pixabay.com/es/>) último acceso agosto 2020.



CAPÍTULO 2. INTRODUCCIÓN A LA TEORÍA DE MUESTRAS

Úrsula Torres Parejo



Capítulo 2

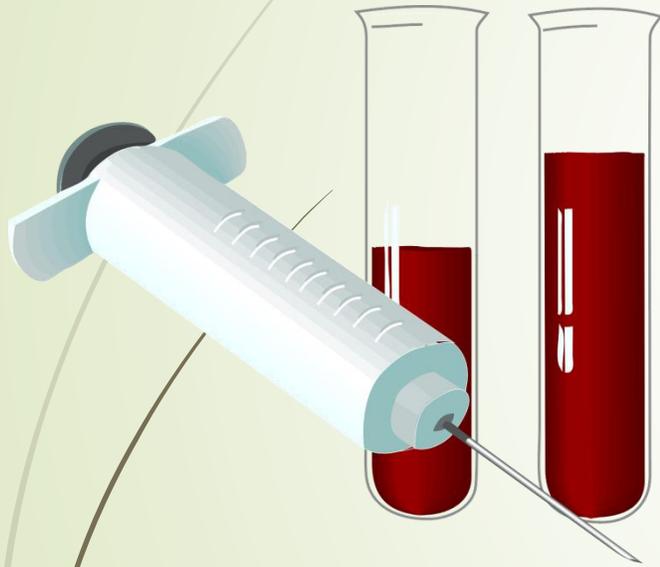
1. Justificación del muestreo.
2. Aplicaciones del muestreo.
3. Conceptos básicos.
4. Errores ajenos al muestreo.
5. Tipos de muestreo.
6. Muestras representativas.
7. Elección del sistema de trabajo.



JUSTIFICACIÓN DEL MUESTREO

JUSTIFICACIÓN DEL MUESTREO

EJEMPLOS



Muestra de sangre



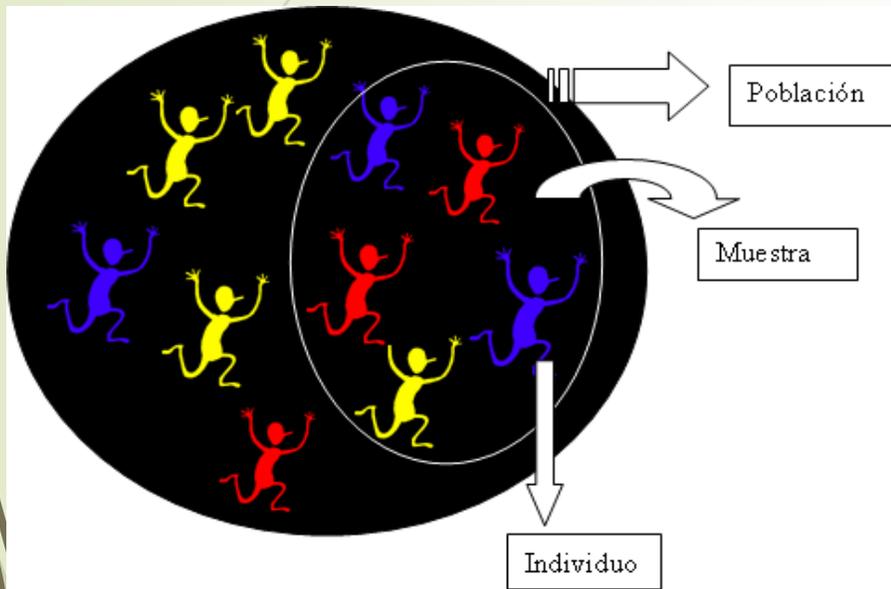
Muestra de perfume



Muestra para control de calidad

JUSTIFICACIÓN DEL MUESTREO

MUESTREO



Muestreo: Procedimiento mediante el cual se selecciona una muestra de una población para analizarla y tomar decisiones pertinentes.

Estrategias posibles:

- Examinar todas las unidades de la población (censo).
- Examinar ciertas unidades, que llamaremos muestra y suponer que los resultados obtenidos son representativos de toda la población.

JUSTIFICACIÓN DEL MUESTREO

IMPORTANCIA DEL MUESTREO

Analizar la población completa no siempre es posible.

Por razones económicas, de tiempo o de recursos

La población completa se desconoce.

O no se tiene la ubicación exacta de los elementos que la conforman

El proceso de medida es un proceso destructivo.

Como cuando se prueba un artículo para juzgar su calidad

Supuestos bajo los que no queda más alternativa que el muestreo.



JUSTIFICACIÓN DEL MUESTREO

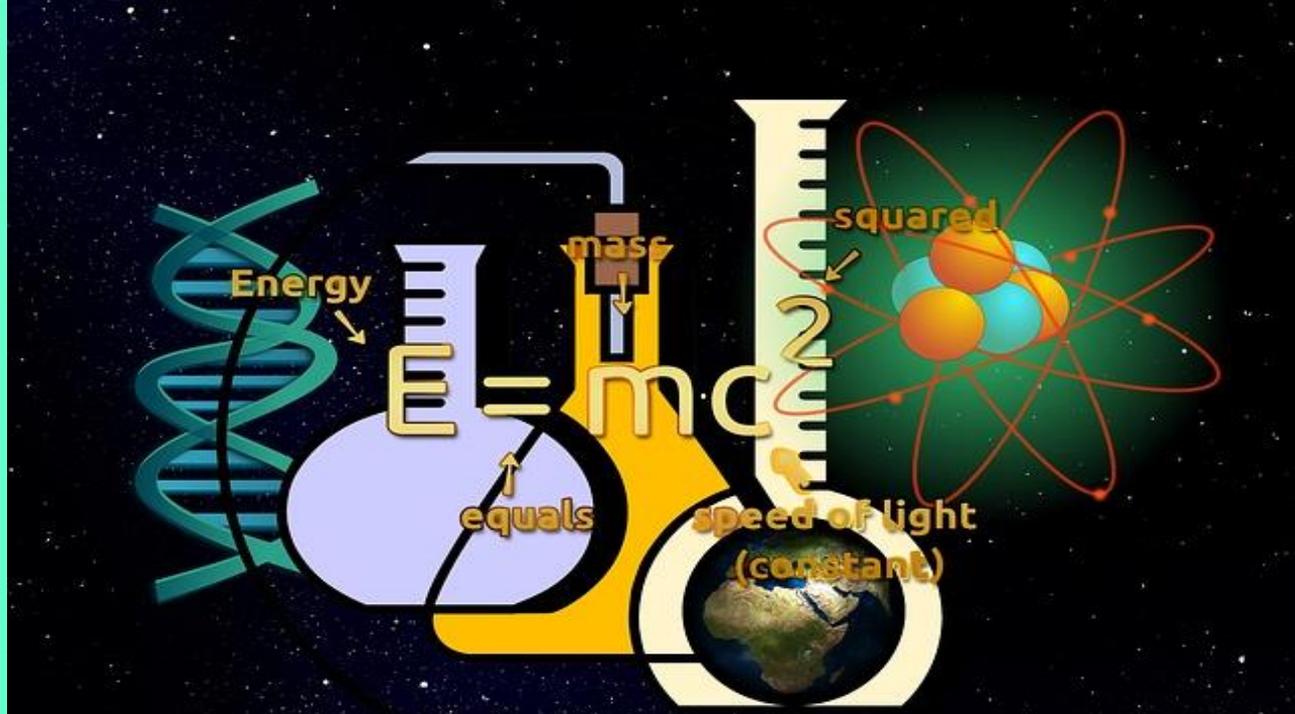
IMPORTANCIA DEL MUESTREO

Ventajas:

- 1. Coste reducido:** Examinar una muestra es más barato que examinar la población completa.
- 2. Rapidez:** Examinar una muestra también es más rápido.
- 3. Accesibilidad:** En muchas ocasiones, el conjunto total de la población no es accesible.
- 4. Exactitud:** En los estudios por muestreo se emplea mayor cuidado y supervisión que en los censos y además los resultados conllevan un margen de error.
- 5. Homogeneidad:** Si la población es suficientemente homogénea cualquier muestra dará una buena representación.

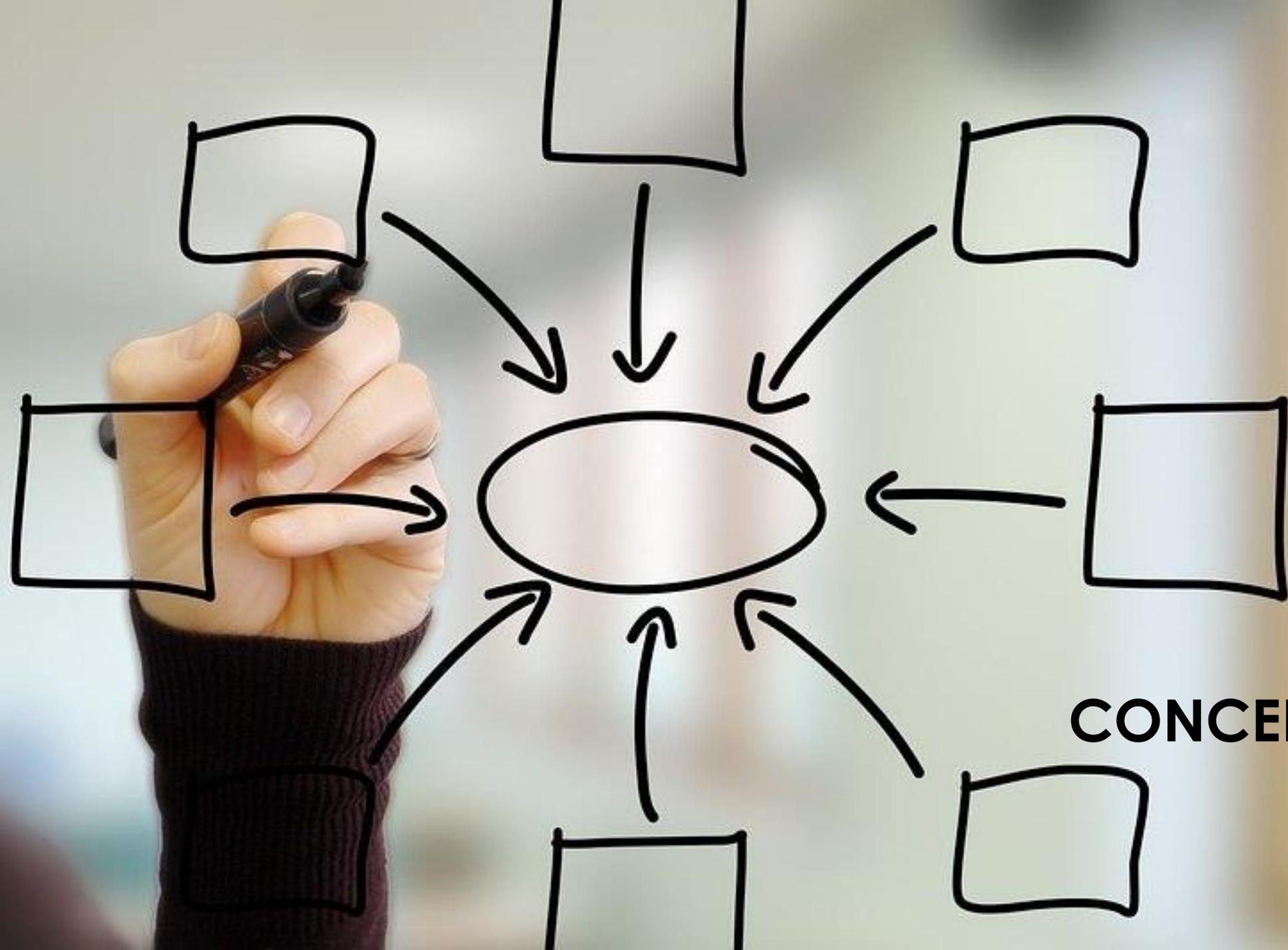
Desventajas:

- 1. Desconocimiento de ubicación de elementos:** No se tiene conocimiento sobre la ubicación física de los elementos de la población.
- 2. Procedimiento de muestreo no apropiado:** Puede ocurrir que el procedimiento de muestreo empleado no sea el apropiado.
- 3. Desconocimiento de la magnitud del error:** En ocasiones, el desconocimiento de la población estudiada impide conocer la magnitud del error que se comete en las estimaciones hechas con el muestreo.



APLICACIONES DEL MUESTREO

- **En política:** por ejemplo en encuestas de opinión pública.
- **En demografía:** estudios sobre nacimientos, matrimonios, bienes y servicios, etc.
- **En el ámbito laboral:** estudios sobre paro, horas de trabajo, salarios, etc.
- **En agricultura:** superficies cultivadas, rendimientos, tipos de cultivos, etc.
- **Investigación de mercados:** por ejemplo en estimación de auditoria de programa de televisión.
- **En comercio e industria:** control de calidad, transacciones comerciales, etc.
- **Campo científico:** educación, salud, estudios sociales, etc.



CONCEPTOS BÁSICOS

CONCEPTOS BÁSICOS

Población

Población objetivo: Conjunto de elementos que se intenta investigar, pero existen omisiones, duplicados y unidades extrañas.

Población estudiada: Conjunto que realmente se investiga.

Censo

Es una lista de todos los elementos que componen la población objetivo (no siempre es posible).

Marco o espacio muestral

Es la numeración de los elementos de la población accesible o estudiada.

Muestra

Subconjunto de elementos del espacio muestral o marco.

CONCEPTOS BÁSICOS

Tamaño muestral

→ Número de unidades que componen la muestra.

Estimación

→ Valores aproximados que se calculan a partir de la muestra y mediante los que inferimos los valores de la población.

Error de muestreo

→ Error que afecta a las estimaciones realizadas con la muestra.

Estimador insesgado

→ Un estimador es una función que sirve para hacer estimaciones. El estimador será insesgado si la media de las estimaciones coincide con el verdadero valor del parámetro poblacional.

Sesgo

→ Diferencia entre la media de las estimaciones realizadas con la muestra y el verdadero valor del parámetro.

CONCEPTOS BÁSICOS

Confianza



Probabilidad con la que establecemos que el verdadero valor de un parámetro se encuentra entre dos límites que se calculan a partir de los datos de la muestra.

CONCEPTOS BÁSICOS

EJEMPLO

Supongamos que se quiere realizar un estudio para estimar el peso medio de los/as trabajadores/as de una empresa multinacional. Para ello se recoge una muestra de 100 personas a las que se pesa.

- **Población:** Total de trabajadores/as de la empresa.
- **Censo:** Lista de trabajadores/as contratados/as por la empresa.
- **Muestreo:** Técnica mediante la cual elegiremos la muestra.
- **Muestra:** Las unidades incluidas en las 100 seleccionadas.
- **Tamaño de muestra:** 100
- **Estimador:** La media del peso de las personas de la muestra será el estimador del peso medio de la población.
- **Estimación:** Media de la muestra.

CONCEPTOS BÁSICOS

EJEMPLO

Supongamos ahora que el peso medio de las 100 personas de la muestra es 66.2 kg. y que el peso medio real del conjunto de trabajadores/as de la empresa es 65.7 kg.

$$\text{Sesgo} = 66.2 - 65.7 = 0.5 \text{ kg.}$$

También podría ocurrir que se diera el resultado de la media de peso a través de un intervalo de confianza, por ejemplo: el intervalo al 95% de confianza para el peso medio es [65.3, 67.1], esto indicaría que el verdadero valor del peso de los/as trabajadores/as de la empresa estaría comprendido entre esos dos valores con una probabilidad del 95%.

ERRORES AJENOS AL MUESTREO



ERRORES AJENOS AL MUESTREO

!!!El error de muestreo es sólo una parte del error total y además se dispone de mecanismos para combatirlo!!!

Los errores ajenos al muestreo se presentan tanto en las encuestas como en los censos, aunque con mayor intensidad y menos posibilidades de corrección en estos últimos..

ERRORES AJENOS AL MUESTREO

De observación.

Se deben a la recogida, registro o procesamiento incorrecto de los datos.

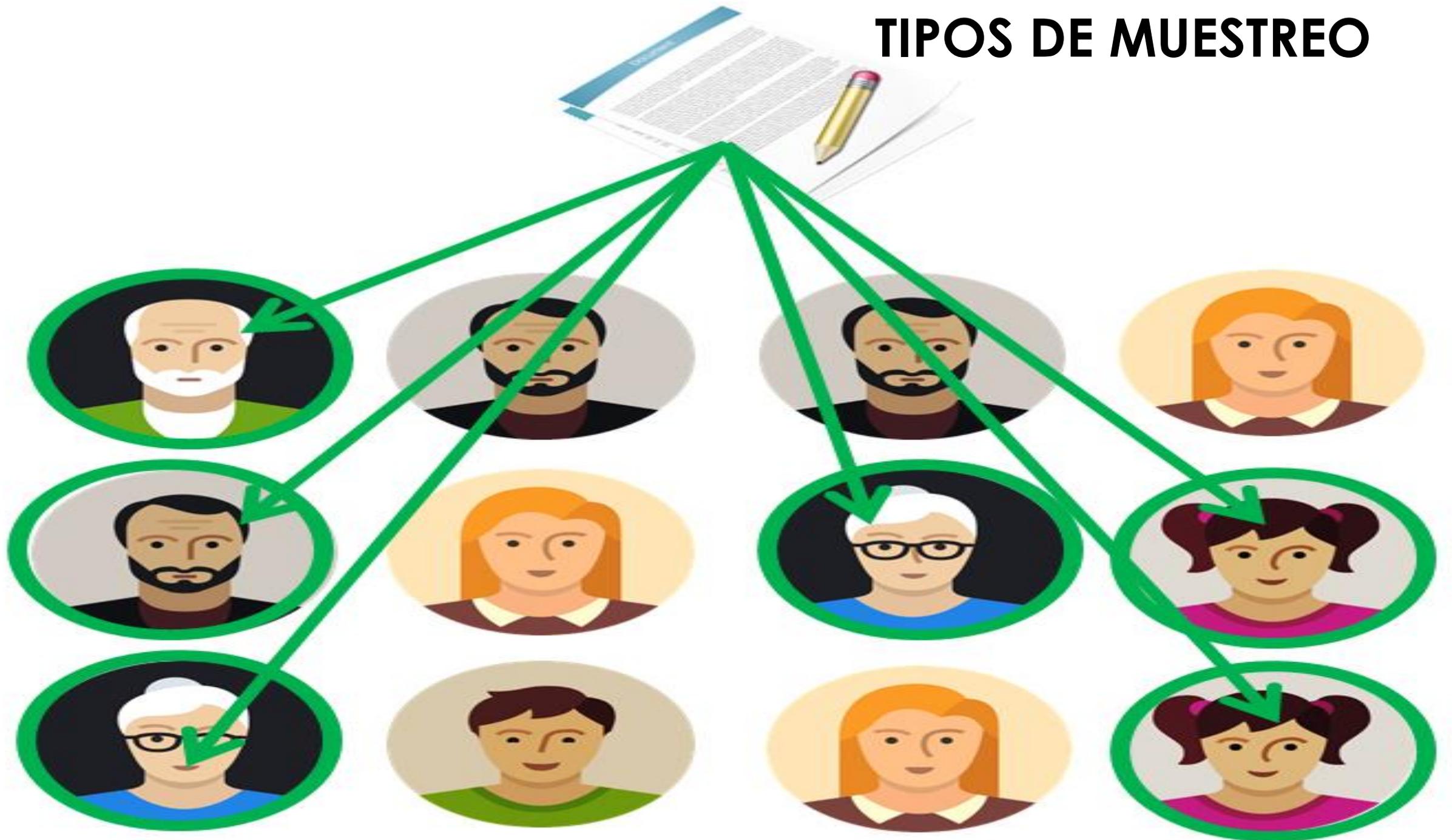
- **De sobrecobertura:** El marco tiene elementos que no corresponden a la población estudiada.
- **De medida:** El valor observado no coincide con el verdadero.
- **De procesamiento:** Errores en la edición de datos.

De no observación.

Se dan cuando no es posible obtener información para ciertos elementos de la población.

- **De cobertura:** Cierta parte de la población objetivo no está presente en el marco y por tanto no puede ser seleccionada.
- **Falta de respuesta:** Ciertos elementos de la muestra no pueden obtenerse o la información obtenida no puede utilizarse.

TIPOS DE MUESTREO



TIPOS DE MUESTREO

Muestreo probabilístico.

La selección de cada elemento es aleatoria y puede calcularse la probabilidad de inclusión en la muestra.

Muestreo intencional.

La representatividad de la muestra depende de la intención u opinión de quien la obtiene

Muestreo sin norma.

La muestra se obtiene de cualquier manera por razones de comodidad o capricho.

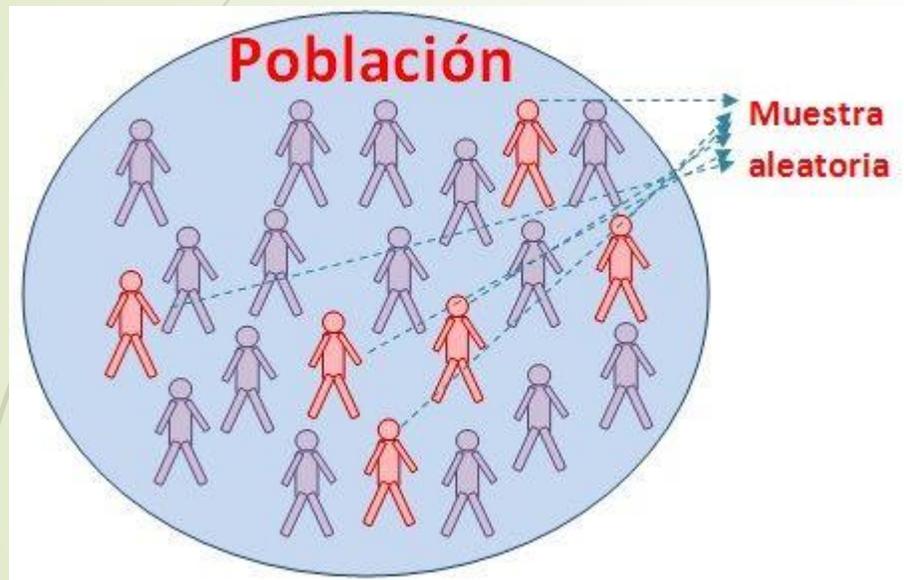


Los muestreos intencional y sin norma carecen de base científica.

MUESTRAS REPRESENTATIVAS



MUESTRAS REPRESENTATIVAS ALEATORIEDAD



Para que un muestreo sea estadísticamente significativo, la muestra debe tener “**aleatoriedad**”, esto es, nadie tiene control sobre la selección de individuos u hogares.



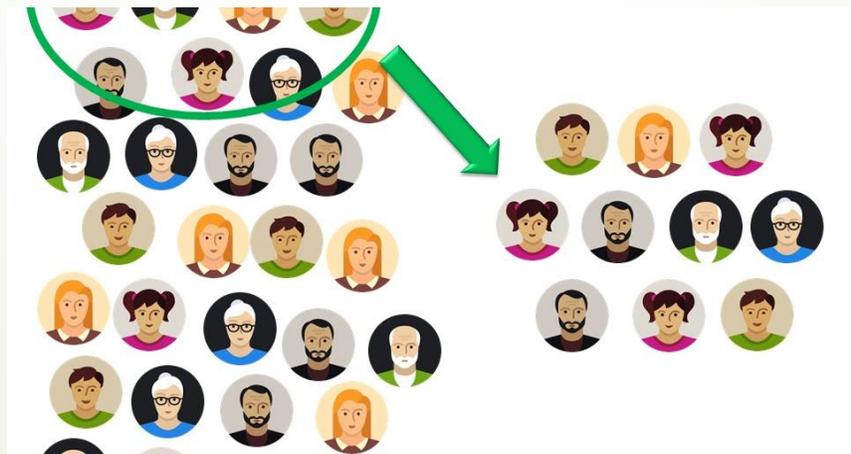
Muestreo Probabilístico

MUESTRAS REPRESENTATIVAS

Para saber si una muestra es representativa, pueden calcularse los porcentajes muestrales de las características de interés y ver si coinciden con los porcentajes poblacionales.

Población:

20% mayores de 65 años
50% mujeres
30% religión evangélica



Muestra:

20% mayores de 65 años
50% mujeres
30% religión evangélica

MUESTRAS REPRESENTATIVAS

En una muestra representativa:

1. Cada individuo o unidad de muestreo en la población tiene una probabilidad conocida, distinta de cero, de ser seleccionado.
2. La selección de un individuo debe ser independiente de la selección de otro.

ELECCIÓN DEL SISTEMA DE TRABAJO



ELECCIÓN DEL SISTEMA DE TRABAJO

¿Qué datos recoger?

- Datos ligados a la información que se busca
- También pueden recogerse otros para encuestas posteriores

¿A partir de quién?

- Una vez definida la población de individuos, hace falta elegir la muestra
- Los datos no siempre se toman de las personas a quienes conciernen

¿Cuándo?

- Hay que establecer fecha de comienzo y duración de la recogida
- Considerar las restricciones de tiempo, naturaleza de los datos, estacionalidad, etc.

¿Cómo?

- Procedimiento de recogida: carta, teléfono, encuesta a domicilio, etc.

ELECCIÓN DEL SISTEMA DE TRABAJO

INFORMACIÓN DISPONIBLE

¿Ha sido realizada ya previamente alguna encuesta?

- La información de dicha encuesta podría resultar muy útil.

¿Existe una lista de individuos?

- Una lista debe contener un número por cada persona, no puede haber dos personas con un mismo número y todo número debe corresponder a una persona.

¿Existen caracteres conocidos?

- De ser así, esta información suplementaria serviría para disminuir el error de muestreo.

ELECCIÓN DEL SISTEMA DE TRABAJO

RESTRICCIONES

Restricciones materiales

- Las principales conciernen al presupuesto y tiempo para la realización de la encuesta.
- Otras restricciones están ligadas a la infraestructura del lugar donde se realiza la encuesta

Restricciones sociales

- Composición social de la población, nivel de educación, cualificación de los/as encuestadores/as, etc.

REFERENCIAS

1. Medina Disla, H. Apuntes sobre muestreo, 2009.
2. Rueda García, M.M. Muestreo I. Teoría, Problemas y Prácticas de Ordenados, Plácido Cuadros, 2003.
3. Rueda García, M.M. y Arcos Cebrián, A. Problemas de Muestreo en Poblaciones Finitas. Editorial GEU, 1999.
4. Morillas, A. Muestreo en Poblaciones Finitas, 2007.
5. Nassiuma, D.K. Survey sampling: Theory and methods. Nairobi University Press, 2001.
6. Marvasti, F. Nonuniform sampling: theory and practice. Springer, 2012.
7. Otzen, T., Manterola, C. Técnicas de Muestreo sobre una Población Objeto de Estudio. International Journal of Morphology, 2017.

Todas las imágenes obtenidas de:

- Imágenes prediseñadas de Microsoft PowerPoint versión 2010
- Imágenes gratis sin atribución de Pixabay (<https://pixabay.com/es/>) último acceso agosto 2020.



CAPÍTULO 3. TIPOS BÁSICOS DE MUESTREO

Úrsula Torres Parejo

Capítulo 3

1. Aleatorio simple.
2. Con probabilidades desiguales.
3. Sistemático.
4. Estratificado.
5. Por conglomerados o clusters.
6. Segmentación.

MUESTREO ALEATORIO SIMPLE



MUESTREO ALEATORIO SIMPLE

También se conoce como muestreo sin reemplazo con probabilidades



Método base de todos los muestreos probabilísticos

Ventajas: simplicidad y tamaño reducido de muestra.

La muestra se extrae unidad por unidad, sin reposición y todas las unidades tienen la misma probabilidad de ser elegidas.

Desventajas: necesita una lista exhaustiva de la población y el coste de visitar hogares o personas geográficamente dispersas, por lo que suele emplearse en poblaciones pequeñas.

MUESTREO ALEATORIO SIMPLE

PASOS PARA APLICARLO

1. Cada unidad en el marco (hogar, persona, etc.) lleva asignado un número único entre 1 y el número total de unidades.
2. Se usa un método aleatorizado para seleccionar las unidades de inclusión dentro de la muestra.
3. Se mapean las unidades seleccionadas para facilitar la recopilación de datos. El equipo de recopilación de datos también debe tener una estrategia de reemplazo de unidades para aquellas que no pueden ser localizadas, debido a información inexacta en el marco o bien porque la unidad no está disponible.
4. Las unidades de reemplazo pueden ser preseleccionadas antes de la recopilación de datos para tenerlas mapeadas, aumentando el tamaño de muestra requerido según se estime que será el porcentaje de unidades fallidas.

MUESTREO CON PROBABILIDADES DESIGUALES



MUESTREO CON PROBABILIDADES DESIGUALES

Las extracciones con probabilidades iguales están basadas en la idea de que nos e conoce nada a prior de las variables estudiadas.

Es normal que se le dé más importancia a ciertas unidades frente a otras, por lo que la probabilidad de selección será distinta para cada unidad.

El cálculo que implica el uso de probabilidades desiguales es mucho mayor que el de probabilidades iguales.

Es muy frecuente su uso en muestreos con más de una etapa.

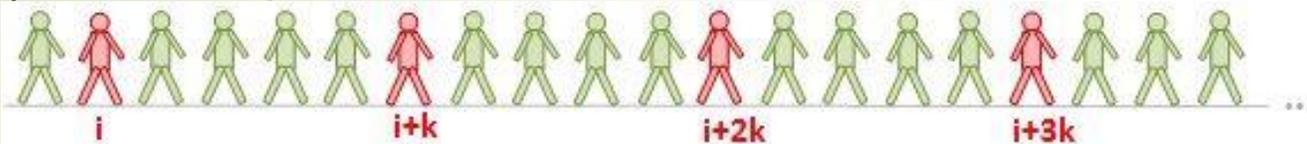
Excepto la probabilidad que tiene cada unidad de ser elegida, el resto de características son comunes con el muestreo aleatorio simple.

MUESTREO SISTEMÁTICO



MUESTREO SISTEMÁTICO

Simplifica el proceso de selección muestral con respecto al m.a.s.



Ventajas:

1. Simplicidad y tamaño reducido de muestra.
2. No es imprescindible tener un listado de la población.
3. Puede proporcionar mayor representatividad ya que no deja ninguna sucesión grande de elementos sin cubrir.

Desventajas:

1. El coste de visitar hogares o personas geográficamente dispersas.
2. Si existe cierta periodicidad y k coincide con el periodo, la representatividad se ve comprometida.

Sólo se extrae de forma aleatoria la primera unidad y el resto se toma a intervalos regulares.



MUESTREO SISTEMÁTICO

PASOS PARA APLICARLO

1. Se calcula un número natural k , comprendido entre 1 y el número total de unidades.
2. Se usa un método aleatorizado para seleccionar la primera unidad entre las k primeras. Este método puede aplicarse con Microsoft Excel o con algún software de tipo estadístico como SPSS o R.
3. Se elige el resto de unidades de k en k , donde k se denomina intervalo de selección.
4. Esta elección puede hacerse sobre el listado (si existe) o en la misma fase de campo.
5. Se establece la estrategia de sustitución de unidades.

MUESTREO SISTEMÁTICO

DETERMINACIÓN DEL INTERVALO DE SELECCIÓN

Llamemos N al número de elementos en el marco y n al tamaño de muestra requerido:

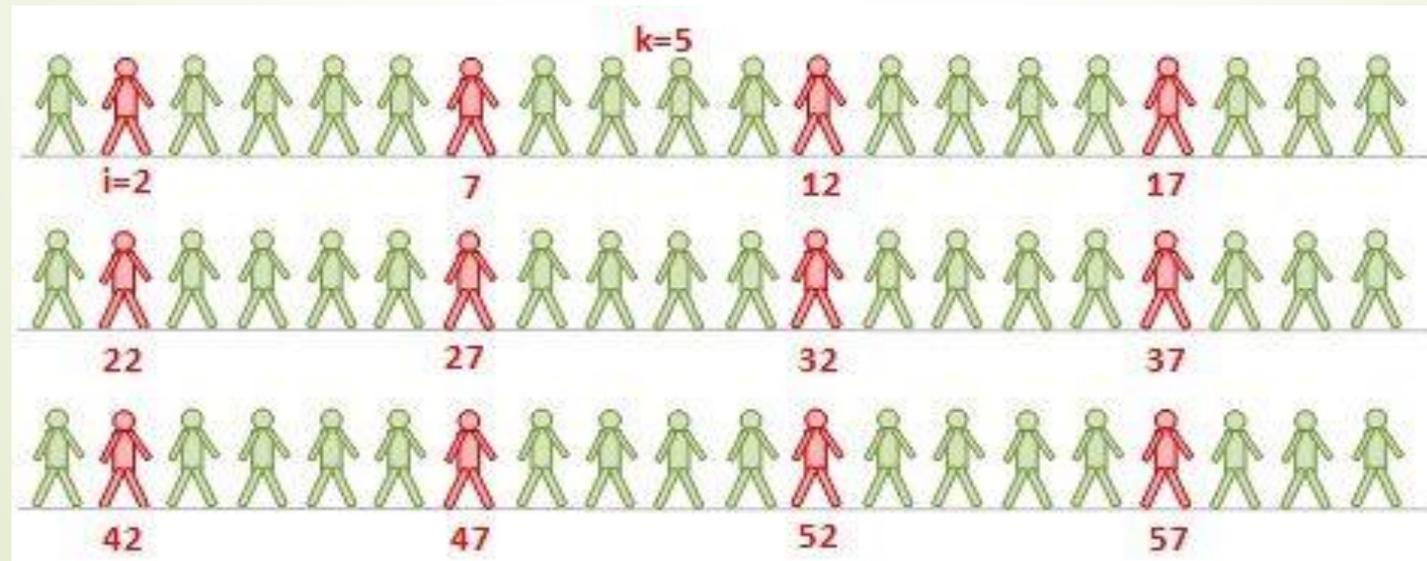
1. Si N es divisible por n , tomaremos $k=N/n$, de forma que la población se dividiría en n zonas de k unidades cada una.
2. Si al dividir N entre n se obtienen decimales, lo recomendable siempre es redondear a la baja para la selección de la primera unidad o también se puede utilizar la siguiente regla de oro:

Decimal entre 0.0 y 0.2 → Redondear a la baja
Decimal entre 0.3 y 0.7 → Alternar
Decimal entre 0.8 y 0.9 → Redondear al alza

MUESTREO SISTEMÁTICO

PRIMER EJEMPLO

Supongamos que tenemos un marco de tamaño 60 y queremos elegir una muestra de tamaño 12. Al calcular k obtenemos $k=60/12=5$, por lo que iremos tomando las unidades en intervalos de 5 en 5, únicamente seleccionaremos de forma aleatoria la primera unidad entre las 5 primeras (supongamos $i=2$).



MUESTREO SISTEMATICO

SEGUNDO EJEMPLO

Supongamos que tenemos un marco de tamaño 340 y queremos elegir una muestra de tamaño 25. Al calcular k obtenemos $k=340/25=13.6$, por lo que si redondeamos a la baja, iremos tomando las unidades en intervalos de 13 en 13, pero si aplicamos la regla de oro, iríamos alternando entre 13 y 14.

	Redondeo al alza ($k=14$)	Redondeo a la baja ($k=13$)	Alternando ($k=13$ o 14)	Exacto ($k=13.6$)
1ª unidad	11	11	11	11
2ª unidad	$11+14=25$	$11+13=24$	$11+13=24$	$11+13.6=24.6$
3ª unidad	$25+14=39$	$24+13=37$	$24+14=38$	$24.6+13.6=38.2$
4ª unidad	$39+14=53$	$37+13=50$	$38+13=51$	$38.2+13.6=51.8$
5ª unidad	$53+14=67$	$50+13=63$	$51+14=65$	$51.8+13.6=65.4$
...
25ª unidad	347	323	335	337

MUESTREO SISTEMATICO

APLICACIONES

1. **Estimaciones de producción** de plantas cultivadas.
2. **Mediciones uniformes a lo largo del tiempo**, como estudios de precipitaciones.
3. **Control de calidad** industrial, como inspección de productos de una línea de montaje.
4. **Investigaciones en movimiento**, como personas que pasan por un mostrador de un supermercado.



MUESTREO ESTRATIFICADO

MUESTREO ESTRATIFICADO

La estratificación conlleva dividir la población de interés en grupos o estratos que comparten algo en común.



El proceso de muestreo se realiza de modo independiente en cada estrato



Si se toma una muestra aleatoria simple de cada estrato, el proceso se conoce como **muestreo estratificado aleatorio**.

Ventajas:

1. Se aumenta la precisión en poblaciones heterogéneas.
2. Si se desea información de ciertas subdivisiones de la población
3. Tiende a asegurar que la muestra sea representativa.

Desventajas:

Es necesario conocer la estructura de la población para poder establecer los estratos.

MUESTREO ESTRATIFICADO

PASOS PARA APLICARLO

1. Se establecen los estratos. Los estratos suelen presentarse de forma natural: áreas geográficas con distintos recursos o niveles de pobreza, grupos de edad, etc.
2. Se reparte el tamaño total de muestra a obtener entre el número de estratos, esta repartición se realiza según algún criterio, siendo el más frecuente el proporcional al tamaño del estrato.
3. Dentro de cada estrato se selecciona el tamaño de muestra calculado en el punto anterior según el procedimiento de muestreo que se considere y teniendo en cuenta la existencia o no de listado.

MUESTREO ESTRATIFICADO

EJEMPLO

Supongamos que tenemos una población de tamaño 50000 habitantes, que dividimos por estratos según rango de edad.

- Primer estrato → Hasta 18 años
- Segundo estrato → De 18 a 65 años
- Tercer estrato → A partir de 65 años

Estos estratos tienen respectivamente una población de 7000, 25000 y 18000 habitantes. Se elige una muestra de 750 habitantes. El reparto de la muestra por estrato sería:

Estrato	Población	Peso del estrato	750*Peso del estrato	Tamaño de muestra
1	7000	$7000/50000=0.14$	105	105
2	25000	$25000/50000=0.5$	375	375
3	18000	$18000/50000=0.36$	270	270
TOTAL	50000	1	750	750



MUESTREO POR CONGLOMERADOS

MUESTREO POR CONGLOMERADOS

El muestreo por conglomerados reduce el gasto en tiempo y recursos cuando el área de la investigación es muy grande y las unidades de investigación están geográficamente dispersas.



Se selecciona una muestra de conglomerados, en lugar de tener que visitar todos



Para poder aplicar este diseño los conglomerados deben ser homogéneos entre sí.

Ventajas:

1. Muy eficiente cuando la población es grande y dispersa al no tener que visitar todos los conglomerados.
2. No es preciso un listado de toda la población.

Desventajas:

1. Requiere un mayor tamaño muestral ya que el error de muestreo es mayor.
2. Que los conglomerados no sean homogéneos

MUESTREO POR CONGLOMERADOS

DIFERENCIAS CON EL MUESTREO ESTRATIFICADO

Por conglomerados:

1. Homogéneos y están compuestos por unidades heterogéneas.
2. No se seleccionan unidades de todos los conglomerados.

Estratificado:

1. Heterogéneos entre sí y compuestos por unidades homogéneas.
2. Se seleccionan unidades de todos los estratos.

MUESTREO POR CONGLOMERADOS

PASOS PARA APLICARLO

1. Se definen y numeran los conglomerados: ciudades, municipios, aldeas, etc. y se indica el tamaño de su población.
2. Se calcula el número de conglomerados seleccionados para la muestra.
3. Se seleccionan los conglomerados según el tipo de muestreo elegido, con la ayuda del software que se use.
4. Una vez seleccionados los conglomerados se reparte el tamaño total de muestra a obtener entre el número de conglomerados, esta repartición se realiza según algún criterio, siendo el más frecuente el proporcional al tamaño del conglomerado (población).
5. Dentro de cada conglomerado se selecciona el tamaño de muestra calculado en el punto anterior según el procedimiento de muestreo que se considere.

MUESTREO POR CONGLOMERADOS

PRIMER EJEMPLO

Supongamos que realizamos un estudio en un municipio que tiene 200 comunidades, de las cuales se seleccionan 20 mediante un muestreo con probabilidades desiguales. El tamaño final de muestra que se desea conseguir es de 10000 personas, que se seleccionan en los conglomerados elegidos con probabilidades proporcionales al tamaño del conglomerado al que pertenecen, como se muestra en la tabla:

Conglomerado	Población	Peso del conglomerado	10000*Peso del conglomerado	Tamaño de muestra
1	14000	$14000/200000=0.07$	700	700
2	24000	$24000/200000=0.12$	1200	1200
3	1000	$1000/200000=0.005$	50	50
...
TOTAL de todos los conglomerados seleccionados	200000	1	10000	10000

MUESTREO POR CONGLOMERADOS SELECCIÓN DE CONGLOMERADOS MEDIANTE MUESTREO SISTEMÁTICO. PASOS PARA APLICARLO

1. Crear la lista o marco de los conglomerados indicando el tamaño de su población.
2. Calcular la población acumulada.
3. Determinar el intervalo de muestreo y seleccionar el número aleatorio de inicio.
4. Seleccionar los conglomerados de forma aleatoria, empezando por el número aleatorio y añadiendo posteriormente el intervalo de muestreo.

MUESTREO POR CONGLOMERADOS

SEGUNDO EJEMPLO

Supongamos que ahora queremos seleccionar los conglomerados a partir de un muestreo sistemático, con consideración del tamaño de los conglomerados. En este caso se tienen 100 conglomerados de los que se extrae una muestra de tamaño 20. La población total de los 100 municipios es 200000 habitantes.

Municipios	Población estimada	Población acumulada	Rango	Municipios seleccionados
Municipio 1	1500	1500	1-1500	x
Municipio 2	4000	5500	15000-5500	
Municipio 3	1600	7100	5500-7100	
Municipio 4	6500	13600	7100-13600	x
Municipio 5	5200	18800	13600-18800	
Municipio 6	2600	21400	18800-21400	x
Municipio 7	1700	23100	21400-23100	
Municipio 100	1000	200000	199000-200000	

MUESTREO POR CONGLOMERADOS

SEGUNDO EJEMPLO

Con 20 conglomerados para seleccionar, el intervalo de muestreo es $k = 200000/20 = 10000$. El número de inicio aleatorio es 200 (entre 1 y 10000), por lo que se selecciona el primer municipio al estar 200 comprendido en el rango. El siguiente municipio seleccionado será $200+10000=10200$, que estaría en el rango del municipio 4, el cual se selecciona y así sucesivamente hasta obtener los 20 municipios.

Municipios	Población estimada	Población acumulada	Rango	Municipios seleccionados
Municipio 1	1500	1500	1-1500	x
Municipio 2	4000	5500	15000-5500	
Municipio 3	1600	7100	5500-7100	
Municipio 4	6500	13600	7100-13600	x
Municipio 5	5200	18800	13600-18800	
Municipio 6	2600	21400	18800-21400	x
Municipio 7	1700	23100	21400-23100	
Municipio 100	1000	200000	199000-200000	

MUESTREO POR CONGLOMERADOS

SEGUNDO EJEMPLO

El inconveniente de este método es que un conglomerado podría resultar elegido en más de una ocasión. En este caso, lo que se haría sería duplicar el tamaño de muestra correspondiente a dicho conglomerado cuando se pasara a la selección de las unidades básicas. Lo ideal es no tener conglomerados excesivamente extensos y si esto ocurre intentar subdividirlos en otros más pequeños, evitando de esta forma la selección doble.

Municipios	Población estimada	Población acumulada	Rango	Municipios seleccionados
Municipio 1	1500	1500	1-1500	x
Municipio 2	4000	5500	15000-5500	
Municipio 3	1600	7100	5500-7100	
Municipio 4	6500	13600	7100-13600	x
Municipio 5	5200	18800	13600-18800	
Municipio 6	2600	21400	18800-21400	x
Municipio 7	1700	23100	21400-23100	
Municipio 100	1000	200000	199000-200000	



SEGMENTACIÓN

SEGMENTACIÓN

Un método utilizado aunque no muy recomendable cuando el área geográfica que cubre un conglomerado es muy grande y las unidades continúan estando muy dispersas, consiste en segmentar dicho conglomerado y seleccionar las unidades únicamente de una de las partes que hemos segmentado.

Esta segmentación puede realizarse a partes iguales, por ejemplo dividiendo el conglomerado en dos partes iguales y seleccionando las unidades básicas únicamente de una de las partes que elegimos de forma aleatoria. También puede hacerse a partes desiguales, teniendo las partes segmentadas distintas extensiones y número de habitantes según algún criterio.

SEGMENTACIÓN

¿QUÉ SEGMENTO ELEGIR EN SEGMENTOS DESIGUALES?

1. Calculamos la población estimada de cada segmento, la acumulada y el rango.

Segmento	Población estimada	Población acumulada	Rango
A	1000	1000	1-1000
B	1500	2500	1000-2500
C	2000	4500	2500-4500

2. Seleccionamos aleatoriamente un número entre 1 y el total de población estimada de todos los segmentos, en nuestro ejemplo 4500. Supongamos que seleccionamos el 2200.
3. Comprobamos en qué rango se encuentra comprendido el número aleatorio, en este caso entre 1000 y 2500, por lo que seleccionamos el segmento B para realizar la encuesta.

Muestreo	Características	Ventajas	Inconvenientes
Aleatorio Simple	<p>Método por el cual todas las personas de un marco tienen la misma probabilidad de ser elegidas en la muestra.</p> <p>Cuando la probabilidad no es la misma, se denomina muestreo con probabilidades desiguales.</p>	<ul style="list-style-type: none"> -Sencillo y de fácil comprensión. -La generación de números al azar a través de software se realiza de forma rápida. 	<ul style="list-style-type: none"> -Requiere que se posea de antemano un listado completo de la población. -Es difícil llevarlo a la práctica en cumplimiento de sus condiciones: mismas oportunidades para todos de ser elegidos.
Sistemático	<p>Consiste en seleccionar a un individuo de forma aleatoria entre la población objetivo y, a partir de él, seleccionar para la muestra a cada "k" individuo del marco.</p>	<ul style="list-style-type: none"> -Simple y rápido. -No siempre es necesario tener un listado de toda la población. -Si la población se ordena siguiendo alguna tendencia conocida, asegura cobertura de todos los tipos. 	<ul style="list-style-type: none"> -Existe la posibilidad de que se genere una muestra sesgada cuando el orden en el que se ha generado la selección de participantes tenga algún tipo de periodicidad oculta que coincida con el intervalo seleccionado.
Estratificado	<p>Consiste en dividir la población objetivo en diferentes subgrupos o estratos. Los estratos están compuestos por unidades homogéneas y son heterogéneos entre sí. Todos los estratos tienen representación en la muestra.</p>	<ul style="list-style-type: none"> -Tiende a asegurar que la muestra sea representativa de la población en función de las variables seleccionadas. -Se obtienen estimaciones más precisas. -Su objetivo es conseguir una muestra lo más semejante posible a la población respecto a las variables de estratificación utilizadas. 	<ul style="list-style-type: none"> -Es necesario conocer la estructura de la población para poder establecer los estratos.
Por conglomerados	<p>Consiste en dividir la población objetivo en diferentes subgrupos o conglomerados. Los conglomerados se caracterizan por ser homogéneos entre sí y estar compuestos por unidades</p>	<ul style="list-style-type: none"> -No es preciso tener un listado de toda la población, sólo de los conglomerados. -Muy eficiente cuando la población es grande y está dispersa, ahorra tiempo y 	<ul style="list-style-type: none"> -Que los conglomerados no sean homogéneos entre sí, lo que generaría nuevos errores. -El cálculo del error estándar es más complejo y es mayor que el de otros

REFERENCIAS

1. Thematic Guidelines: Sampling Guidelines for Vulnerability Analysis. Chapter 4. Household-level data in a CFSVA. 2004.
2. Rueda García, M.M. Muestreo I. Teoría, Problemas y Prácticas de Ordenador. Plácido Cuadros, 2003.
3. Rueda García, M.M. y Arcos Cebrián, A. Problemas de Muestreo en Poblaciones Finitas. Editorial GEU 1999.
4. Cuesta, M. y Herrero, F.J. Introducción al Muestreo. Universidad de Oviedo, España, 2013.
5. Sampling Methods and Sample Size Calculation for the SMART Methodology. 2012.
6. Tamayo, G. Diseños muestrales en la investigación. Universidad de Medellín, Colombia, 2001

Todas las imágenes obtenidas de:

- Imágenes prediseñadas de Microsoft PowerPoint versión 2010
- Imágenes gratis sin atribución de Pixabay (<https://pixabay.com/es/>) último acceso agosto 2020.



CAPÍTULO 4. MUESTREO COMPLEJO

Úrsula Torres Parejo



Capítulo 4

1. Muestreo polietápico.
2. Muestreo polifásico.
3. Muestreos no probabilísticos.

A close-up, low-angle shot of a person's legs and feet standing on a staircase. The person is wearing bright green pants and dark sneakers with white soles. The staircase has yellow-painted steps and grey concrete risers. The lighting is dramatic, with strong shadows and highlights. The text "MUESTREO POLIETÁPICO" is overlaid in the bottom right corner.

MUESTREO POLIETÁPICO

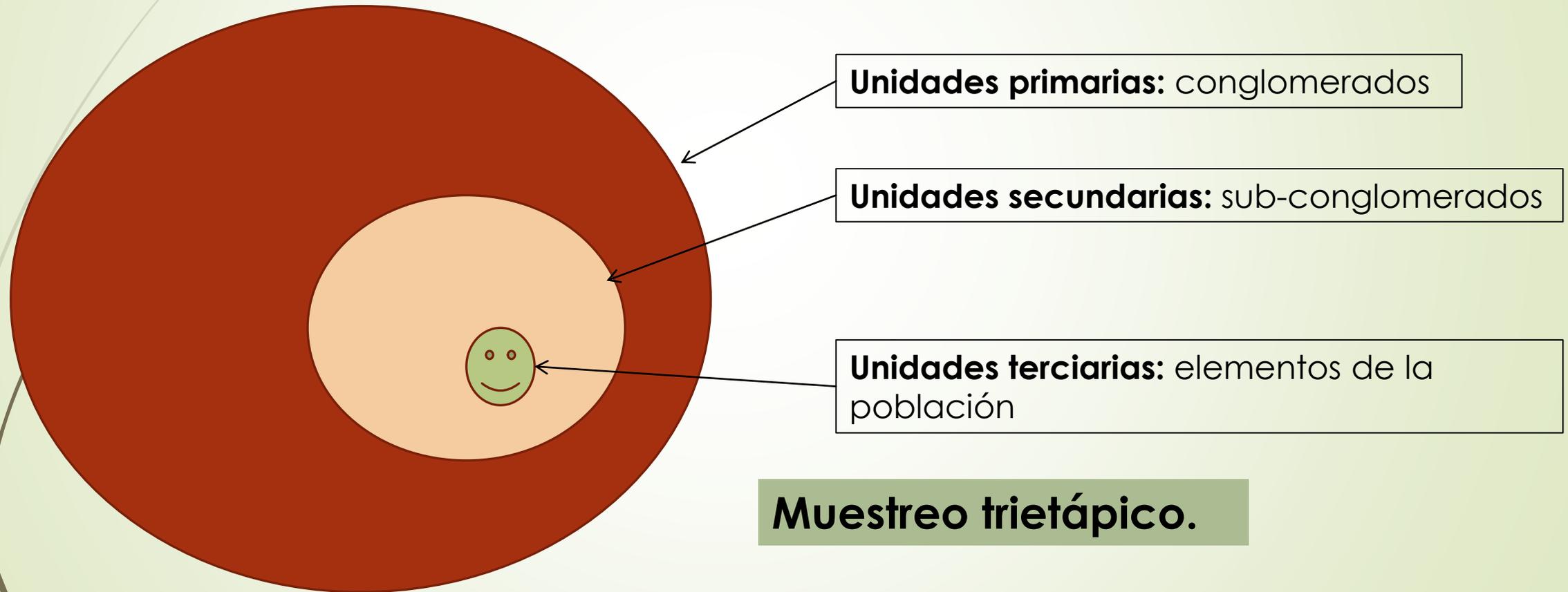
MUESTREO POLIETÁPICO

El muestreo por conglomerados es en realidad un **muestreo bietápico**, ya que en una primera etapa se seleccionan los conglomerados y en una segunda etapa las unidades dentro de los conglomerados. Si nos quedamos únicamente en selección de conglomerados, entonces se conocería como **muestreo por conglomerados en una etapa**.

También es común combinar el **muestreo bietápico con estratificación**. Dicha técnica consiste en dividir la población en estratos y dentro de cada estrato, de forma independiente, realizar un muestreo bietápico, que consistiría en la extracción de conglomerados y posteriormente, de las unidades de la muestra.

Pero los muestreos también pueden desarrollarse en más de una etapa, dando lugar a los **muestreos trietápicos o polietápicos** (más de tres etapas).

MUESTREO POLIETÁPICO



Muestreo trietápico.

MUESTREO POLIETÁPICO

EJEMPLO



Selección de conglomerados de cada estrato.



Selección de la persona del hogar que hará la encuesta.



Muestra de hogares

Selección de hogares entre los conglomerados de la muestra.

Muestreo trietápico combinado con estratificación.

MUESTREO POLIETÁPICO

EJEMPLO

Se establecen las siguientes consideraciones:

- 1. Primera etapa:** selección de conglomerados (comunidades) con probabilidad proporcional al tamaño de los estratos (ZMV) a los que pertenecen, medido este tamaño en número de comunidades.
- 2. Segunda etapa:** selección de subconglomerados (hogares), también con probabilidades desiguales. En este caso, para calcular la probabilidad que tiene cada hogar de ser elegido, se tiene en cuenta el tamaño del estrato al que pertenece y también el tamaño de la comunidad, medidos ambos tamaños en número de hogares.
- 3. Tercera etapa:** selección de elementos (miembro del hogar que finalmente realizará la entrevista). Esta última selección se realiza de forma aleatoria según la fecha del cumpleaños de las personas del hogar.

MUESTREO POLIETÁPICO

EJEMPLO

Supongamos que el número total de comunidades es 550 y el número de comunidades en la muestra 30.

Etapa 1: Cálculo del número de conglomerados por estrato según número total de conglomerados por estrato.

ESTRATOS	Nº de comunidades	Peso	Muestra de comunidades
ZMV 1	100	$100/550=0.182$	$0.182*30=5.46 \sim 5$
ZMV 2	150	$150/550=0.273$	$0.273*30=8,18 \sim 8$
ZMV 3	120	$120/550=0.218$	$0.218*30=6,55 \sim 7$
ZMV 4	40	$40/550=0,072$	$0.072*30=2,18 \sim 2$
ZMV 5	140	$140/550=0.255$	$0.255*30=7,636 \sim 8$
Total	550	1	30

MUESTREO POLIETÁPICO

EJEMPLO

Supongamos que el número total de hogares en las comunidades seleccionadas es 10000 y el número de hogares en la muestra es 630.

Etapa 2: Primera parte

Cálculo del número de subconglomerados por estrato según el número total de subconglomerados por estrato.

ESTRATOS	Nº de hogares	Peso	Muestra de hogares
ZMV 1	1000	$1000/10000=0.1$	$0.1*630=63$
ZMV 2	3500	$3500/10000=0.35$	$0.35*630=220.5 \sim 221$
ZMV 3	2000	$2000/10000=0.2$	$0.2*630=126$
ZMV 4	1000	$1000/10000=0,1$	$0.1*630=63$
ZMV 5	2500	$2500/10000=0.25$	$0.25*630=157,5 \sim 158$
Total	10000	1	631

MUESTREO POLIETÁPICO

EJEMPLO

Ya tenemos el número total de conglomerados y subconglomerados por estrato. Falta calcular el número de subconglomerados en cada conglomerado por estrato.

Etapa 2: Segunda parte

Selección de subconglomerados con probabilidad proporcional al número de hogares por estrato.

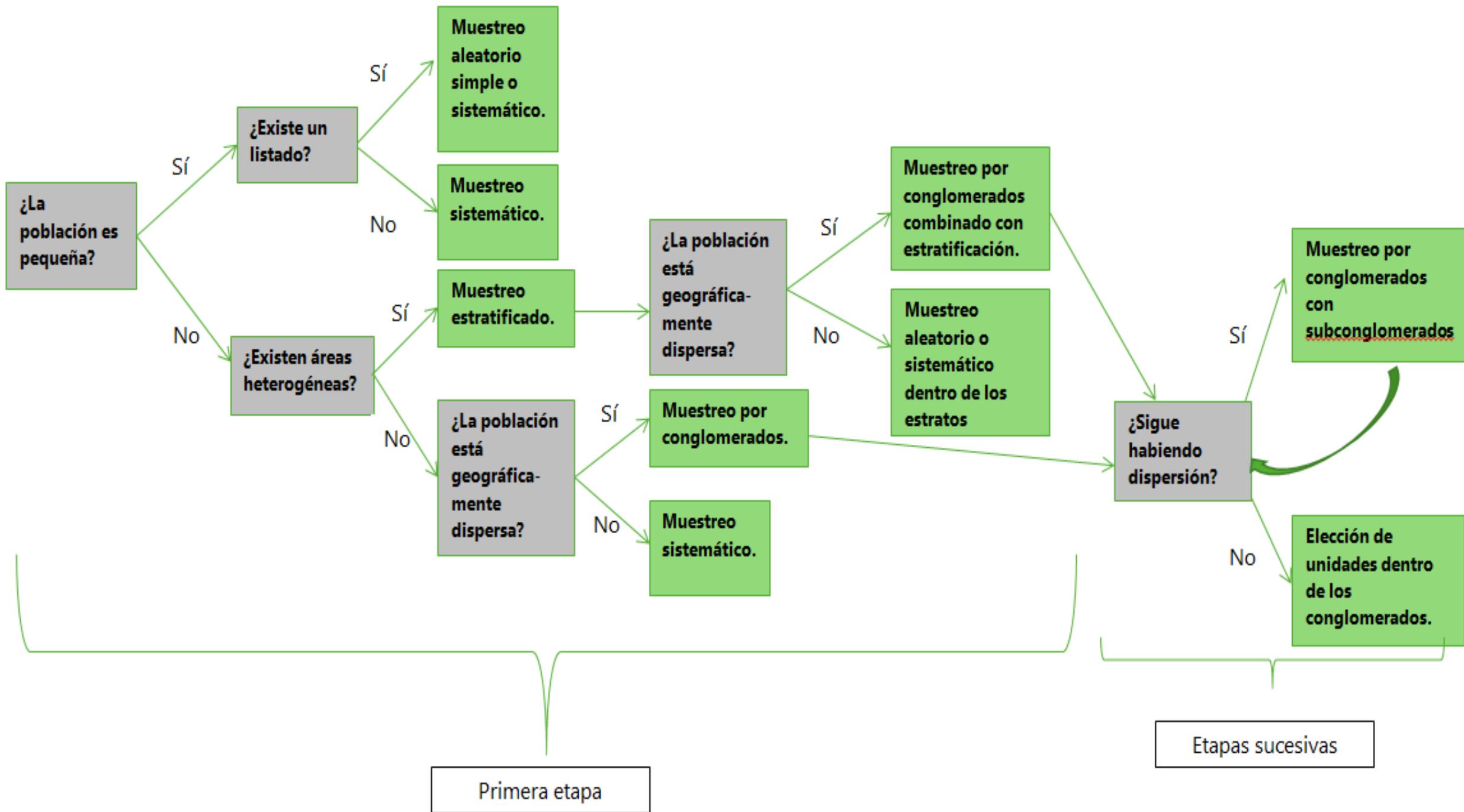
ESTRATO1	Nº de hogares	Peso	Muestra de hogares
Comunidad 1	100	$100/1000=0.1$	$0.1*63=6.3 \sim 6$
Comunidad 2	350	$350/1000=0.35$	$0.35*63=22.1 \sim 22$
Comunidad 3	200	$200/1000=0.2$	$0.2*63=12.6 \sim 13$
Comunidad 4	100	$100/1000=0,1$	$0.1*63=6.3 \sim 6$
Comunidad 5	250	$250/1000=0.25$	$0.25*63=15.8 \sim 16$
Total	1000	1	63

MUESTREO POLIETÁPICO

EJEMPLO

Este procedimiento se repetiría para el resto de estratos.

Etapa 3: En esta última etapa el muestreo aleatorio se aplicaría directamente en la fase de campo. Se contactaría telefónicamente con los hogares y, una vez contestaran al teléfono, se preguntaría por la persona mayor de edad cuya fecha de cumpleaños se encuentre más cercana.



¿La población es pequeña?

¿Existe un listado?

¿Existen áreas heterogéneas?

Muestreo aleatorio simple o sistemático.

Muestreo sistemático.

Muestreo estratificado.

¿La población está geográficamente dispersa?

Muestreo por conglomerados.

Muestreo sistemático.

Muestreo por conglomerados combinado con estratificación.

Muestreo aleatorio o sistemático dentro de los estratos

¿Sigue habiendo dispersión?

Muestreo por conglomerados con subconglomerados

Elección de unidades dentro de los conglomerados.

Primera etapa

Etapas sucesivas



MUESTREO POLIFÁSICO

MUESTREO POLIFÁSICO

El **muestreo bifásico** o **muestreo doble** (Neyman 1938) consiste en formar primero una muestra grande, de forma rápida, sencilla y reducido coste, y utilizar la información que proporciona dicha muestra como base para seleccionar una muestra, de tamaño menor, en la que se observa la variable de interés y se estiman las características poblacionales.

El muestreo bifásico puede generarse en cualquier número de fases, dando lugar al **muestreo polifásico**.

MUESTREO POLIFÁSICO

APLICACIÓN DEL MUESTREO DOBLE A LA ESTRATIFICACIÓN

Se emplea cuando la población está dividida en estratos con población desconocida y cuyos pesos no pueden calcularse.

- 1. Primera fase:** Se extrae una muestra aleatoria simple y se estudia a qué estrato pertenecen las unidades seleccionadas. De esta forma es posible, calcular la proporción de unidades en cada estrato y suponer que es la misma que la proporción para el conjunto de la población.
- 2. Segunda fase:** A partir de la muestra extraída en la primera fase se estima el tamaño de los estratos y se extrae, de esta forma, una muestra aleatoria en cada estrato con probabilidad proporcional a su tamaño.

MUESTREO POLIFÁSICO

EJEMPLOS

1. Una empresa decide realizar una encuesta para estudiar el dinero que gasta en libros una determinada población, para ello toma una primera muestra de tamaño 400, que es clasificada en tres grupos atendiendo a la renta per cápita de las familias. Posteriormente se eligió una submuestra en cada uno de estos grupos de tamaños 20, 10 y 10, respectivamente.
2. En una determinada población va a realizarse una encuesta telefónica, pero se carece de un listado de teléfonos, por lo que se extrae una primera muestra de hogares con el fin de visitarles para lograr el teléfono. De esta muestra telefónica lograda, se extrae la segunda muestra, de menor tamaño, que serán finalmente las personas a las que se llamará para realizar la encuesta.



MUESTREOS NO PROBABILÍSTICOS

MUESTREOS NO PROBABILÍSTICOS

TIPOS DE MUESTREOS

1. Intencional o por conveniencia.
2. Por cuotas.
3. De juicio o de experto.
4. En bola de nieve.

MUESTREOS NO PROBABILÍSTICOS

MUESTREO INTENCIONAL O POR CONVENIENCIA



En la imagen se aprecia un muestreo por conveniencia donde el investigador selecciona las personas más cercanas a él, a las que tiene más fácil acceso.

Cada unidad muestral se selecciona de acuerdo a la intención o conveniencia del investigador o investigadora.

Se emplea cuando no tenemos una lista completa de individuos que forman la población y, por lo tanto, no se conoce la probabilidad de cada individuo sea seleccionado para la muestra.

Muy común en estudios con voluntarios, ensayos clínicos, entrevistas callejeras

MUESTREOS NO PROBABILÍSTICOS

MUESTREO POR CUOTA

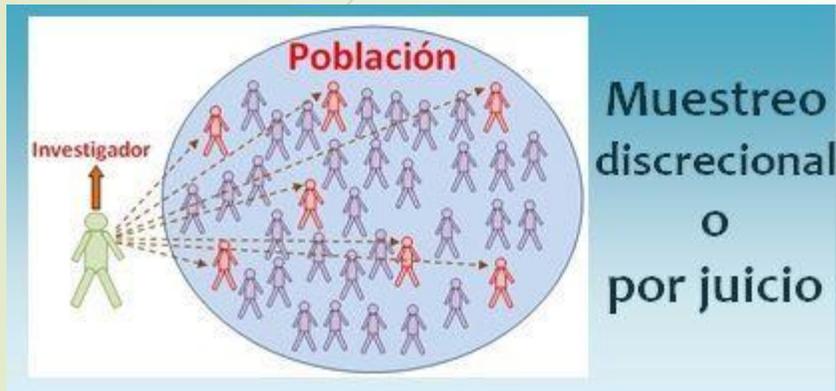
Los/as investigadores/as seleccionan la muestra en la misma proporción que está en la población con variables fácilmente identificadas.

Se fijan unas "cuotas", que consisten en un número de unidades muestrales que reúnen unas determinadas características, por ejemplo: 20% de personas de 25 a 40 años, 50% de sexo femenino y 15% evangélicos. Cada entrevistador/a selecciona por su cuenta a los sujetos basándose en estos criterios, hasta completar la cuota.

Este procedimiento es en general muy popular y presenta como ventaja principal que se puede obtener la muestra de forma rápida y barata. No es preciso contar con listas de la población, ya que sólo basta con tener información sobre las características de la población con las que definir las cuotas. Sin embargo, se pueden cometer sesgos de tipo muy diverso.

MUESTREOS NO PROBABILÍSTICOS

MUESTREO DE JUICIO O DE EXPERTO



La persona experta en el tema que se está investigando selecciona las unidades muestrales, según sus conocimientos sobre la conveniencia de cada unidad en la muestra.

Ejemplo: En una investigación sobre las ventas diarias de todos los supermercados de una ciudad, se seleccionan aquellos que son más representativos según el juicio de la persona investigadora.

La principal limitación de este tipo de muestreo es que depende única y exclusivamente del nivel de conocimiento del experto o experta que selecciona la muestra, lo que limita el marco muestral del estudio.

MUESTREOS NO PROBABILÍSTICOS

MUESTREO EN BOLA DE NIEVE



Muestreo de bola de nieve

Consiste en seleccionar una muestra inicial básica de individuos y establecer en cada entrevista que, cada individuo entrevistado debe conducir a otro y aquel a otro y así sucesivamente hasta alcanzar la muestra completa.

Puede emplearse en la localización de unidades raras, como son poblaciones marginales, delincuentes, sectas, personas con drogadicción, determinados tipos de enfermos, etc. También se utiliza en poblaciones pequeñas en las que se desconoce la ubicación de sus miembros.

La principal limitación de la información y conocimiento de los elementos abordados.

MUESTREOS NO PROBABILÍSTICOS

MUESTREO EN BOLA DE NIEVE

PRINCIPALES APLICACIONES

1. Muestreo de relaciones, estructuras o redes de las unidades de la población.
2. Muestreo de élites.
3. Estudio de inmigrantes, minorías étnicas, grupos religiosos, etc.
4. Encuestas relativas a personas con dificultades o marginadas.

REFERENCIAS

1. Rueda García, M.M. Muestreo I. Teoría, Problemas y Prácticas de Ordenador. Plácido Cuadros, 2003.
2. Cuesta, M. y Herrero, F.J. Introducción al Muestreo. Universidad de Oviedo, España, 2013.
3. Heckathorn, D.D. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. Social Problems, 2002.
4. Reichel, D. y Morales, L. Surveying Immigrants Without Sampling Frames – Evaluating the Success of Alternative Field Methods. Springer, 2017.
5. Etikan, I. Abubakar Musa, S. y Sunusi Alkassim, R. Comparison of Convenience Sampling and Purposive Sampling. American Journal of Theoretical and Applied Statistics, 2016.
6. Tamayo, G. Diseños muestrales en la investigación. Universidad de Medellín, Colombia, 2001.

Todas las imágenes obtenidas de:

- Imágenes prediseñadas de Microsoft PowerPoint versión 2010
- Imágenes gratis sin atribución de Pixabay (<https://pixabay.com/es/>) último acceso agosto 2020.



CAPÍTULO 5. CÁLCULO DEL TAMAÑO DE MUESTRA

Úrsula Torres Parejo



Capítulo 6

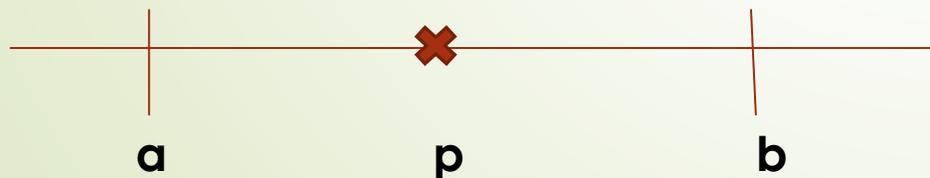
1. Razones por las que es preciso determinar el tamaño de muestra.
2. Sesgo, precisión e intervalo de confianza.
3. Cálculo del tamaño de muestra.
4. Determinación del número de conglomerados.



**RAZONES POR LAS QUE ES PRECISO DETERMINAR EL
TAMAÑO DE MUESTRA**

RAZONES POR LAS QUE ES PRECISO DETERMINAR EL TAMAÑO DE MUESTRA

1. Estimar un parámetro p determinado con el nivel de confianza deseado.
2. Reducir costes y aumentar la rapidez del estudio.
3. Garantizar un error mínimo predeterminado en la estimación.



(a,b) intervalo de confianza del parámetro p



SESGO, PRECISIÓN E INTERVALO DE CONFIANZA

SESGO, PRECISIÓN E INTERVALO DE CONFIANZA

Es una medida de **exactitud**. Se calcula como la diferencia entre la estimación realizada con la muestra y el verdadero valor del parámetro.

Sesgo:

Cómo de próximo está el verdadero valor del parámetro a la estimación.

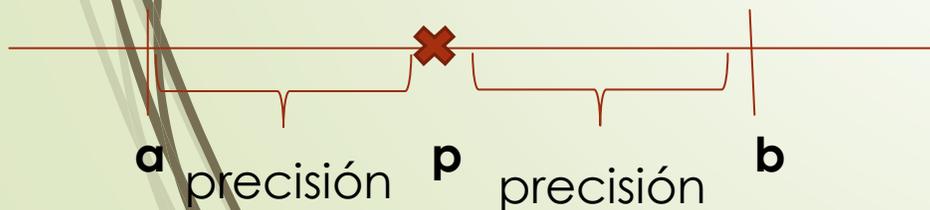
Precisión:

Cómo de similares son los resultados obtenidos a través de distintas encuestas con misma metodología.

Mide la consistencia de los resultados y está relacionada con el error de muestreo, el cual disminuye aumentando tamaño de muestra

Confianza:

Grado que mide la probabilidad con que puede afirmarse que el verdadero parámetro p se encuentra entre dos valores dados a y b .



Intervalo de confianza (a,b).

SESGO, PRECISIÓN E INTERVALO DE CONFIANZA



Pero un aumento del tamaño de muestra puede incrementar el sesgo, ya que habrá más equipos que formar y supervisar.
Es mejor tener una muestra con menor precisión y más exactitud.

SESGO, PRECISIÓN E INTERVALO DE CONFIANZA

Exactitud (medida) vs
Precisión (desviación estándar)



Exacto, pero no preciso

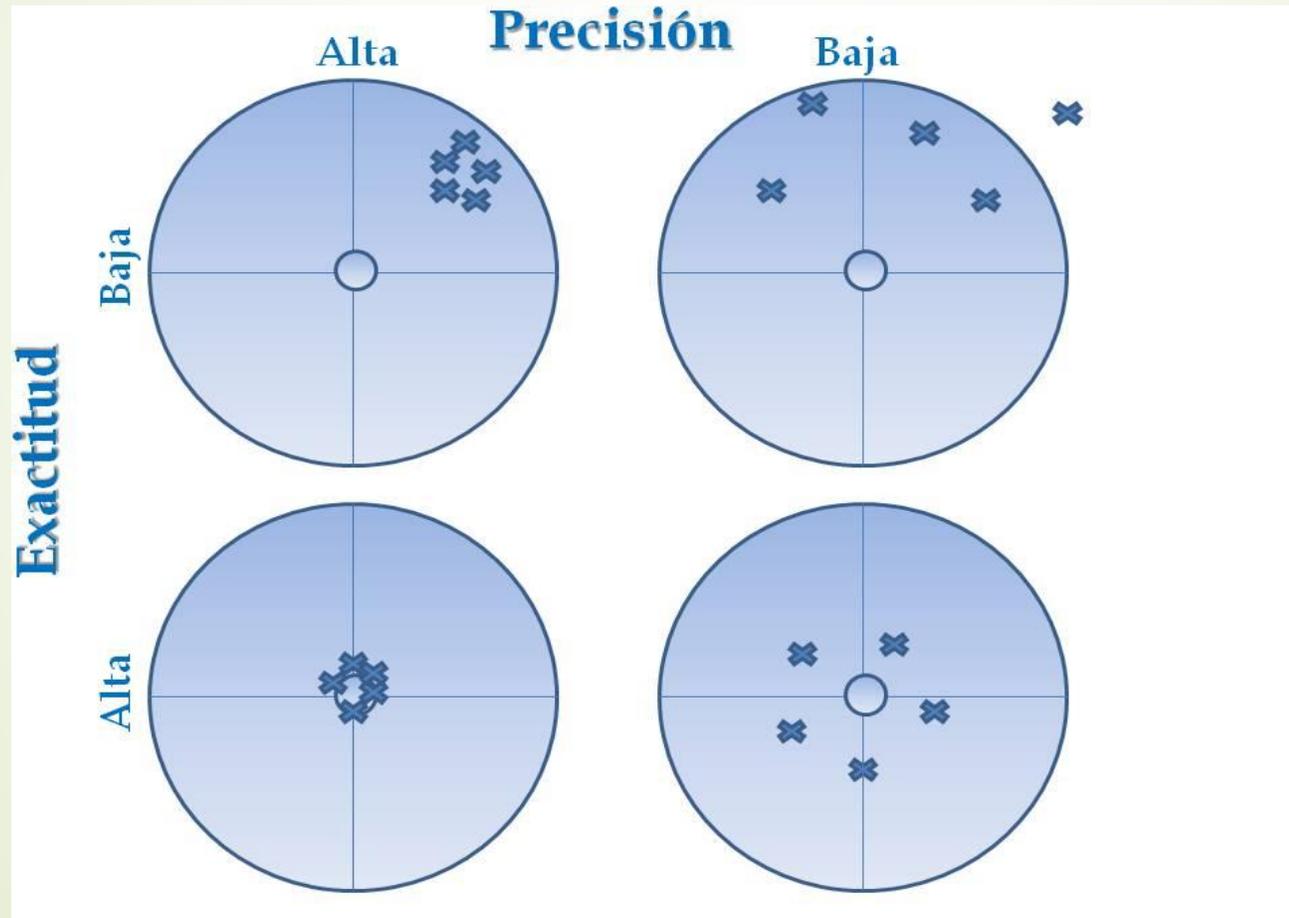


Preciso, pero exacto



Exacto y preciso

SESGO, PRECISIÓN E INTERVALO DE CONFIANZA





CÁLCULO DEL TAMAÑO DE MUESTRA

CÁLCULO DEL TAMAÑO DE MUESTRA

PREVALENCIA

- **La prevalencia puede tomarse de:** encuestas previas, datos recientes, investigaciones, etc. O bien calcularse a través de datos muestrales que pueden recogerse de forma rápida.
- **Cuando la prevalencia esperada se desconoce,** lo aconsejable es tomar un valor cercano a **0.5 (50%),** que es el valor más desfavorable.
- Si la **prevalencia** esperada viene expresada a través de un intervalo de confianza, se recomienda tomar el **límite superior del intervalo.**

CÁLCULO DEL TAMAÑO DE MUESTRA

PRECISIÓN

Cuanto mayor queramos que sea la **precisión**, mayor será el tamaño de muestra resultante, lo que incrementa el coste y posiblemente el sesgo, por lo que lo deseable es encontrar el equilibrio justo.

Prevalencia Esperada	Precisión
5-10%	± 3%
10-15%	± 3.5%
15-20%	± 4%
20-30%	± 5-7%
40% o más	± 10%

CÁLCULO DEL TAMAÑO DE MUESTRA

MUESTREO ALEATORIO SIMPLE O SISTEMÁTICO

En este tipo de muestreo la fórmula utilizada para el cálculo del tamaño de muestra cuando el tamaño de población N se desconoce es:

$$n = z^2 \cdot \frac{pq}{d^2}$$

donde:

- n es el tamaño de muestra
- z es 1.96 para una confianza del 95%, 2.575 para el 99%, 1.645 para el 90%
- p es la prevalencia esperada (valor entre 0 y 1)
- q = 1 – p
- d es la precisión deseada

CÁLCULO DEL TAMAÑO DE MUESTRA

MUESTREO ALEATORIO SIMPLE O SISTEMÁTICO

En el caso de que la población total sea conocida, es preferible usar la fórmula siguiente:

$$n = \frac{N * z^2 * p * q}{d^2(N-1) + z^2 * p * q}$$

donde:

- n es el tamaño de muestra
- z es 1.96 para una confianza del 95%, 2.575 para el 99%, 1.645 para el 90%
- p es la prevalencia esperada (valor entre 0 y 1)
- q = 1 - p
- d es la precisión deseada
- N es el tamaño de la población

CÁLCULO DEL TAMAÑO DE MUESTRA

MUESTREO POR CONGLOMERADOS

Para considerar la heterogeneidad entre conglomerados en un diseño de este tipo, el tamaño de muestra calculado para los muestreos aleatorio simple o sistemático, se multiplica por un factor conocido como efecto del diseño (DEFF):

$$n = \left(z^2 \cdot \frac{pq}{d^2} \right) \cdot DEFF$$

→ Cuando el tamaño de la población es desconocido

$$n = \left(\frac{N \cdot z^2 \cdot p \cdot q}{d^2 (N-1) + z^2 \cdot p \cdot q} \right) \cdot DEFF$$

→ Cuando el tamaño de la población es conocido

CÁLCULO DEL TAMAÑO DE MUESTRA

MUESTREO POR CONGLOMERADOS

Prevalencia del 10%



DEFF 1.5

Muestra de conglomerados
pequeña (Ej: 15)



DEFF 1.5

Distribución homogénea



DEFF 1.5

Prevalencia del 25-30%



DEFF 1.7-1.8

Muestra de conglomerados
grande (Ej: 25-30)



DEFF 1.7-1.8

Distribución heterogénea



DEFF 1.7-1.8

Valores aproximados del efecto del diseño, que dependerá de la prevalencia, tamaño de muestra y tipo de distribución de datos. El valor concreto puede tomarse de encuestas previas realizadas en el mismo área.

CÁLCULO DEL TAMAÑO DE MUESTRA

MUESTREO ESTRATIFICADO

Para aplicar un efecto del diseño DEFF en un muestreo estratificado, éste se calcularía como el promedio ponderado (según el tamaño de los estratos) de los DEFF en los diferentes estratos. En este tipo de muestreo el DEFF es siempre menor que 1.

CÁLCULO DEL TAMAÑO DE MUESTRA OTROS FACTORES DE CORRECCIÓN

Tasa de falta de respuesta (NRR):

$$n_{final} = \frac{n}{1-NRR}$$

Factor de corrección para poblaciones pequeñas (<10000):

$$n_{final} = n \cdot 0.9$$

CÁLCULO DEL TAMAÑO DE MUESTRA

TAMAÑO DE MUESTRA PARA EL CÁLCULO DE VARIOS INDICADORES

- Si la diferencia entre los tamaños requeridos es pequeña  Tomar el tamaño de muestra mayor.
- Si la diferencia entre los tamaños requeridos es grande:  Recoger tamaños de muestra distintos para cada indicador.
- Si un tamaño de muestra es excesivamente grande:  Considerar la pérdida de precisión y optar por un tamaño de muestra menor.

CÁLCULO DEL TAMAÑO DE MUESTRA DETERMINACIÓN DEL NÚMERO DE CONGLOMERADOS

El número de conglomerados necesario se calcula como la razón entre el tamaño de muestra y el número medio de elementos en cada conglomerado (población media de todos los conglomerados). En ocasiones, este último dato no se conoce, por lo que se realiza una estimación. Lo habitual es que el número de conglomerados, oscile entre 25 y 30, aunque puede ser superior o inferior.

CÁLCULO DEL TAMAÑO DE MUESTRA

DETERMINACIÓN DEL NÚMERO DE CONGLOMERADOS

Para la selección del número de conglomerados puede realizarse un muestreo aleatorio simple o sistemático.

Es conveniente elegir **conglomerados de reserva**, por si en algunos de los previamente seleccionados fuera muy complicado conseguir la información.

Nº de conglomerados	Nº de conglomerados de reserva
25-29	3
30-39	4
40-49	5

Algunas recomendaciones indican que es necesario utilizar los conglomerados de reserva sólo si no ha sido posible acceder al 10% o más del número original de conglomerados.

Si el tamaño de muestra conseguido en alguno de los conglomerados es inferior al 80% requerido, el conglomerado entraría en el 10% para contabilizar si es preciso usar los de reserva.

CÁLCULO DEL TAMAÑO DE MUESTRA

EJEMPLO

Supongamos que para estimar la prevalencia de un determinado indicador se tiene que la prevalencia esperada es del 10%. Tomando una confianza del 95% y una precisión del 3%:

$$n = z^2 \cdot \frac{pq}{d^2} = 1.96^2 \frac{0.1 \cdot 0.9}{0.03^2} = 384.16$$

Como en el diseño muestral que va a realizarse van a considerarse distintos conglomerados, multiplicamos este tamaño muestral por el efecto del diseño. Vamos a considerar $DEFF=1.5$:

$$n' = n \cdot DEFF = n \cdot 1.5 = 576.24$$

CÁLCULO DEL TAMAÑO DE MUESTRA

EJEMPLO

Además, la tasa de respuesta se estima en un 10%, por lo que volviendo a recalcular:

$$n'' = \frac{n}{1 - NRR} = \frac{576.24}{1 - 0.1} = 640.27$$

Finalmente, como el tamaño de muestra debe ser un número entero, redondeamos a 641.

REFERENCIAS

1. Sampling Methods and Sample Size Calculation for the SMART Methodology. 2012.
2. Rueda García, M.M. Muestreo I. Teoría, Problemas y Prácticas de Ordenador. Plácido Cuadros, 2003.
3. Rueda García, M.M. y Arcos Cebrián, A. Problemas de Muestreo en Poblaciones Finitas. Editorial GEU 1999.
4. Consolidated Approach to Reporting Indicators of Food Security (CARI). 2015.
5. Thematic Guidelines: Sampling Guidelines for Vulnerability Analysis. Chapter 4. Household-level data in a CFSVA. 2004.
6. Mitos y Verdades del Efecto del Diseño (DEFF) en Estudios por Muestreo Probabilístico, Unidad Investigación Uta, 2017.
7. Mayor Gallego, J.A. Muestreo en Poblaciones Finitas. Diseños Muestrales Complejos. Universidad de Sevilla, 2014.

Todas las imágenes obtenidas de:

- Imágenes prediseñadas de Microsoft PowerPoint versión 2010
- Imágenes gratis sin atribución de Pixabay (<https://pixabay.com/es/>) último acceso agosto 2020.