

Manual de Diseño Muestral

Curso "Generación y Análisis de información de seguridad alimentaria y nutricional"



www.gvc-italia.org



Proyecto “Maximizar el impacto de la asistencia humanitaria en el Corredor Seco Centroamericano mejorando la generación y difusión de datos de Seguridad Alimentaria y Nutricional”- ECHO/-AM/BUD/2020/91013, cuyo objetivo es de contribuir a mejorar la respuesta humanitaria a la inseguridad alimentaria de los actores nacionales y regionales del Corredor Seco Centroamericano.

Financiado por: Programa de Preparación ante Desastres de la Comisión Europea (DIPECHO)

Autoría

Dra. Úrsula Torres Parejo

Dirección

Fundación Acción Internacional contra el Hambre

Consortio de Organizaciones Humanitarias

Equipo Consultor GIS4TECH Spin Off – Universidad de Granada

Dra. Úrsula Torres Parejo

Msc. Jorge Hernández Marín

Dr. Francisco Javier Abarca Alvarez

Revisión y Coordinación

Arlen Martinez Ortiz, Coordinadora DRR – Acción contra el Hambre

Corrección de estilo

TENDENCIAS COMUNICACIÓN INTEGRAL, S.A

Diagramación, diseño de portada e interiores

Hiliana Nuñez

Las ideas y argumentaciones expresadas en este documento son las de los autores, y no reflejan necesariamente el punto de vista de Acción contra el Hambre ni del Consorcio de Organizaciones Humanitarias ni del donante.

Todos los derechos reservados. Se autoriza la reproducción y difusión del material contenido en este producto para fines educativos u otros fines no comerciales sin previa autorización escrita de los titulares de los derechos de autor, siempre que se especifique claramente la fuente.



ÍNDICE

➤ Introducción	4.
➤ Guía Práctica para la construcción de muestras	5.
1. Conceptos básicos	6.
2. Elección del sistema de trabajo	8.
3. Tipos de muestreo probabilístico	9.
4. Elección del diseño muestral	10.
5. Cálculo del tamaño de la muestra	11.
5.1. <i>Muestreo aleatorio simple o sistemático</i>	12.
5.2. <i>Muestreo por conglomerados</i>	13.
5.3. <i>Muestreo estratificado</i>	15.
5.4. <i>Otros factores de corrección</i>	15.
5.5. <i>Determinación del número de conglomerados</i>	16.
5.6. <i>Ejemplo</i>	17.
6. Pasos a seguir en cada tipo de muestreo con ejemplos	18.
6.1. <i>Muestreo Aleatorio Simple</i>	18.
6.2. <i>Muestreo sistemático</i>	19.
6.3. <i>Muestreo estratificado</i>	21.
6.4. <i>Muestreo por conglomerados</i>	23.
6.4.1. <i>Selección de conglomerados y unidades mediante muestreo aleatorio</i>	23.
6.4.2. <i>Selección de conglomerados mediante muestreo sistemático</i>	25.
6.5. <i>Segmentación</i>	26.
6.6. <i>Muestreo polietápico</i>	27.

> Guiones de prácticas diseño muestral	31.
Descarga de EPIDAT 4.2	32.
PRÁCTICA 1. Muestreo Aleatorio Simple y Muestreo Aleatorio Sistemático	34.
<i>Generación de muestra con Microsoft Excel</i>	35.
<i>Generación de muestra con Epidat 4.2</i>	40.
PRÁCTICA 2. Muestreo estratificado y Muestreo por conglomerados en una etapa	45.
<i>Generación de muestra con Epidat 4.2</i>	46.
<i>Generación de muestra con Microsoft Excel</i>	53.
PRÁCTICA 3. Muestreo por conglomerados en dos etapas combinado con estratificación y cálculo de muestra	60.
TRABAJOS PROPUESTOS	75.
<i>Trabajo Propuesto 1</i>	76.
<i>Trabajo Propuesto 2</i>	76.
<i>Trabajo Propuesto 3</i>	76.
> Bibliografía	77.



INTRODUCCIÓN

El presente **Manual de Diseño Muestral** se elabora en el marco del proyecto “Maximizar el impacto de la asistencia humanitaria en el Corredor Seco Centroamericano mejorando la generación y difusión de datos de Seguridad Alimentaria y Nutricional”, ECHO/-AM/BUD/2020/91013, financiado por el Programa de Preparación ante Desastres de la Comisión Europea (DIPECHO) e implementado por Acción contra el Hambre y el Consorcio de Organizaciones Humanitarias: OXFAM, TROCAIRE, GVC, COOPI.

El Manual forma parte de los instrumentos de trabajo generados durante el **Curso Generación y Análisis de información de seguridad alimentaria y nutricional**, bajo el Programa de creación de capacidad regional y local para evaluaciones estandarizadas de seguridad alimentaria y nutricional a través del cual se espera poder efectuar un seguimiento de los riesgos conexos para las acciones previstas en el Corredor Seco Centroamericano de acuerdo con las normas internacionales.

Con esta acción se espera generar capacidades en los equipos técnicos del consorcio de organizaciones humanitarias para conducir el proceso de diseño, desarrollo e implementación del Sistema de Monitoreo de la Seguridad Alimentaria y Nutricional – SM-SAN con el cual, un conjunto de actores estaremos en capacidad de recopilar, procesar y analizar datos fiables y de buena calidad sobre la SAN al tiempo que, se genera evidencia de información que plantee la situación real en los cuatro países del corredor seco Centroamericano.

El Manual de Diseño Muestra fue elaborado por GIS4TECH Spin Off – Universidad de Granada, como un marco orientador que describe las prácticas generalmente empleadas para el diseño y selección de las muestras que permitan el cálculo de ponderadores y expansores. Está dividido en dos secciones: Guía Práctica para la construcción de muestras y Guiones de prácticas diseño muestral, incluyendo ejercicios basados en las realidades del corredor seco Centroamericano.

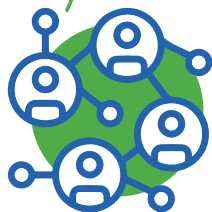
GUÍA PRÁCTICA PARA LA CONSTRUCCIÓN DE MUESTRAS



1.

Conceptos básicos

- ✓ **Población**
Conjunto de unidades del que se desea obtener información.
- ✓ **Unidades**
Elemento diferenciador y completo que forma parte de una serie o de un conjunto (personas, familias, viviendas, escuelas, fábricas).
- ✓ **Información**
Datos obtenidos a través de indicadores de seguridad alimentaria u otros.
- ✓ **Población objetivo**
Población que se intenta investigar.
- ✓ **Población estudiada**
Población con la que se trabaja, que contiene duplicaciones, unidades extrañas y unidades de las que no se puede obtener información (no colaboran o están ausentes).
- ✓ **Muestra**
Subconjunto de la población estudiada.



✔ **Marco**

Conjunto de unidades a partir del cual se selecciona la muestra (listado de unidades de la población estudiada).

✔ **Estimaciones**

Valores aproximados de la población en su totalidad inferidos a partir de la muestra.

✔ **Error debido al muestreo**

Desacierto de las estimaciones, al inferirlas desde la muestra. Cuanto menor sea éste, diremos que mayor es la precisión de las estimaciones.

✔ **Muestreo**

Procedimiento mediante el cual se extrae la muestra.

✔ **Tamaño muestral**

Número de unidades que componen la muestra.

✔ **Sesgo**

Diferencia entre la media de las estimaciones realizadas con las muestras y el verdadero valor del parámetro poblacional. Cuanto menor es el sesgo, mayor es la exactitud en la estimación.

✔ **Confianza**

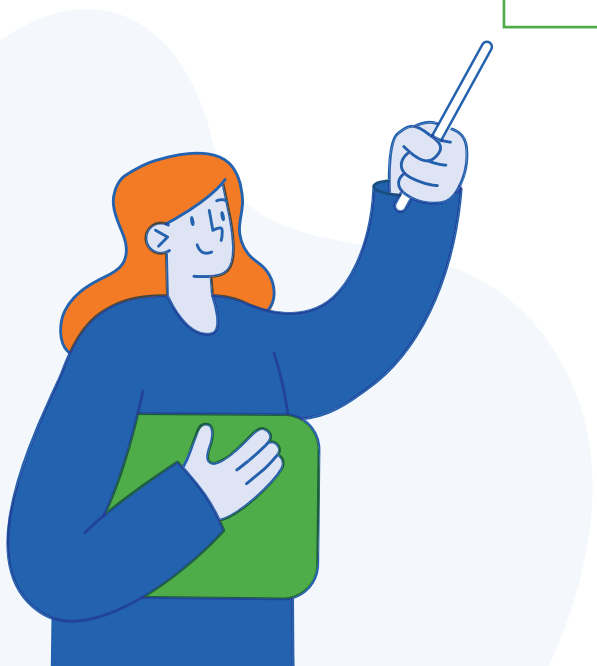
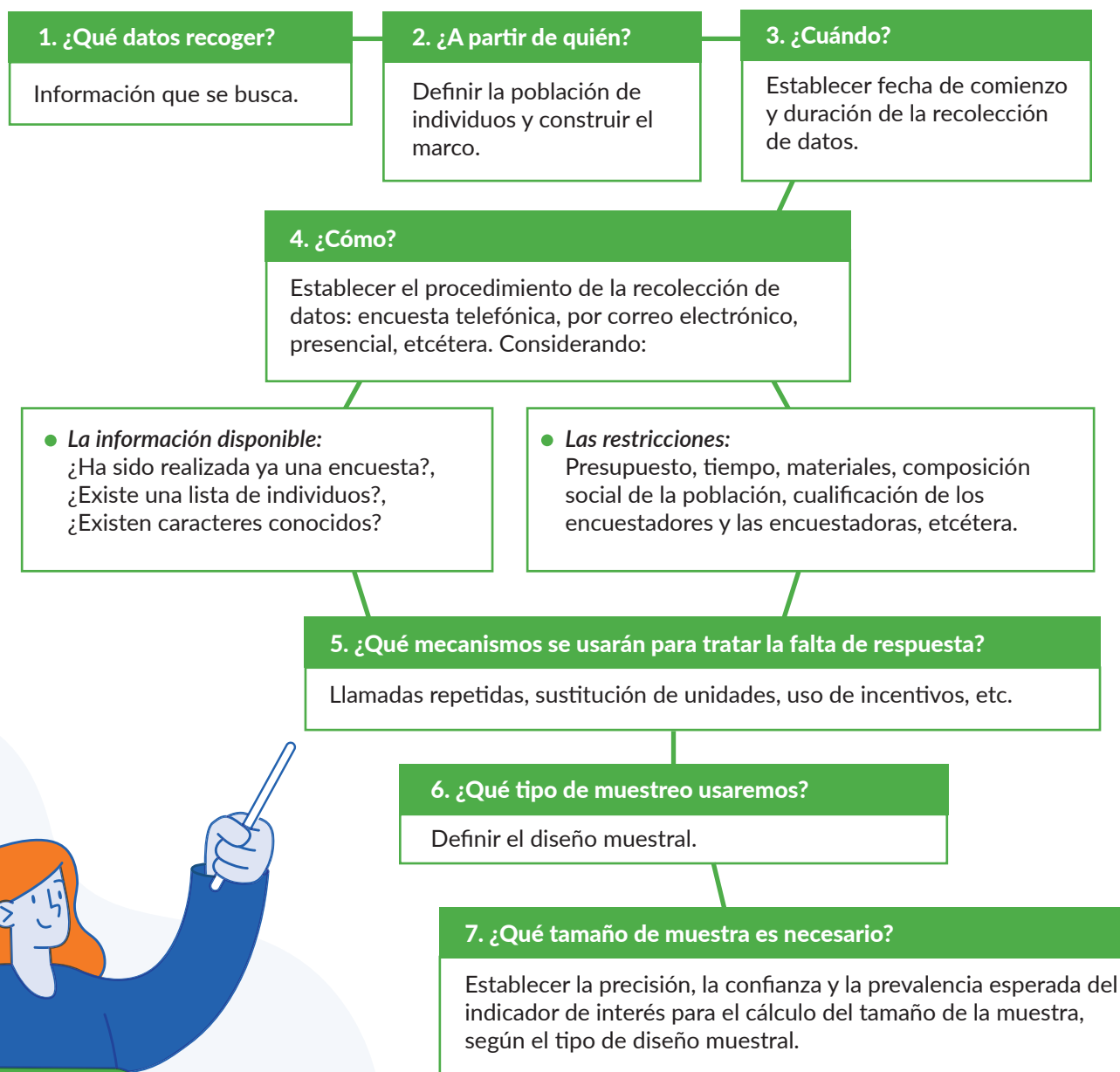
Probabilidad de que una estimación se encuentre entre dos valores calculados para una precisión previamente fijada.



2.

Elección del Sistema de Trabajo

Elegiremos el sistema de trabajo contestando a las siguientes preguntas:



3.

Tipo de Muestreo Probabilístico

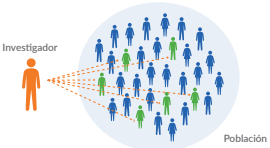

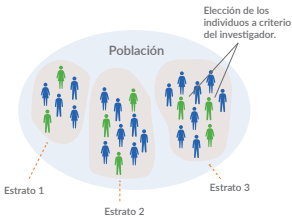
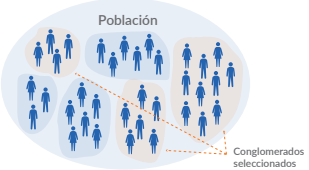
Muestreo	Características	Ventajas	Inconvenientes
<p>Aleatorio Simple</p> 	<p>Método por el cual todas las personas de un marco tienen la misma probabilidad de ser elegidas en la muestra.</p> <p>Cuando la probabilidad no es la misma, se denomina muestreo con probabilidades desiguales.</p>	<ul style="list-style-type: 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.
<p>Sistemático</p> <p>$k=s$</p> 	<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. 	<p>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.</p>
<p>Estratificado</p> 	<p>Consiste en dividir la población objetivo en diferentes subgrupos o estratos. Los estratos están compuestos por unidades homogéneas y son heterogéneas 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. 	<p>Es necesario conocer la estructura de la población para poder establecer los estratos.</p>
<p>Por conglomerados</p> 	<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 heterogéneas. Todos los conglomerados no tienen representación dentro de la muestra.</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 recursos al no tener que visitar todos los conglomerados. 	<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 tipos de muestreo. - Requiere mayor tamaño muestral.

Ilustración 1. Tabla de tipos de muestreo probabilístico

4.

Elección del Diseño Muestral

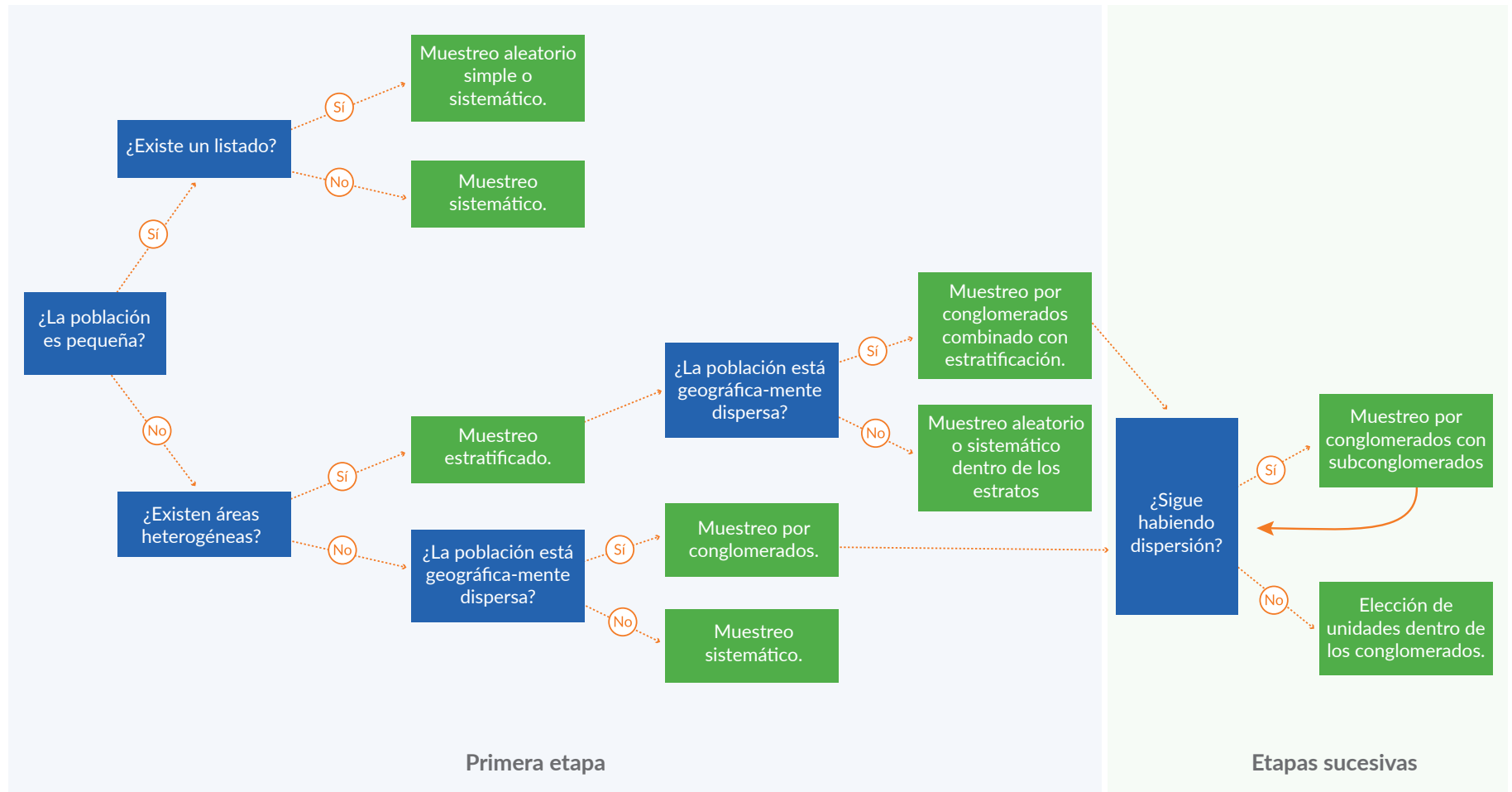


Ilustración 2. Etapas de elección del diseño muestral

5.

Cálculo del Tamaño de la Muestra

Para el cálculo del tamaño de la muestra se debe conocer la prevalencia, la cual puede tomarse de encuestas o investigaciones anteriores. Si se desconoce, se aconseja tomar el valor 0.5 (50% de prevalencia), por ser el más desfavorable y que mayor tamaño de muestra proporcionará.

En función de la prevalencia, podemos estimar la precisión:

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

Ilustración 3: Tabla de precisión recomendada según prevalencia.



5.1

Muestreo aleatorio simple o sistemático

En este tipo de muestreo la fórmula utilizada para el cálculo del tamaño de muestra 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

Cuando el nivel de confianza no se especifica, tomamos el nivel de confianza estándar del 95%, equivalente a un valor de z de 1.96.

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

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



5.2

Muestreo por conglomerados

En el caso del muestreo por conglomerados, el tamaño de muestra calculado para los muestreos aleatorio simple o sistemático, se multiplica por un factor conocido como efecto del diseño, para tener en cuenta la fuente de heterogeneidad entre los diferentes conglomerados.

Efecto del diseño

Es un factor de corrección para considerar la heterogeneidad entre conglomerados en un diseño de este tipo, cuando medimos el indicador de referencia. Se calcula como la varianza del estimador (según el diseño considerado) entre la varianza del estimador según un muestreo aleatorio simple.

La fórmula para el cálculo del tamaño muestral considerando dicho efecto variará de la siguiente forma:

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

- 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
 - DEFF es el efecto del diseño

O de la siguiente si el valor N de la población total es conocido:

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

Si la prevalencia del indicador estudiado en los diferentes conglomerados es muy diferente, estaremos en el caso de una distribución heterogénea, en cuyo caso habrá que considerar un valor alto para el efecto del diseño. Por el contrario, si el valor de la prevalencia es similar en todos los conglomerados, estaremos en el caso de una distribución homogénea, con lo que el efecto del diseño puede tomar un valor bajo.

El valor concreto del efecto del diseño puede tomarse de encuestas previas realizadas en la misma área, siempre que no existan razones para pensar que el grado de heterogeneidad ha cambiado.

El efecto del diseño dependerá de la prevalencia y del tamaño de los conglomerados:

1. Si la prevalencia esperada es alta, el efecto del diseño debería ser alto. Por ejemplo, para una prevalencia del 10%, el efecto del diseño podría tomarse como 1.5, mientras que si la prevalencia es alrededor de un 25%-30%, el efecto del diseño debería incrementarse hasta un 1.7%-1.8 %.

2. Si el tamaño de muestra por conglomerado es pequeño, el efecto del diseño puede tomarse también bajo. Por ejemplo, para una muestra de 15 unidades por conglomerado puede ser suficiente un efecto de diseño alrededor de 1.5, pero si la muestra asciende a 25%-30% unidades por conglomerado, el efecto del diseño se incrementaría hasta un 1.7%-1.8%.



5.3

Muestreo estratificado

La fórmula utilizada en el muestreo estratificado para el cálculo del tamaño muestral es bastante compleja, por lo que es frecuente aplicar la fórmula del muestreo aleatorio simple o muestreo sistemático sin ocasionar pérdida en la precisión, ya que las prevalencias esperadas de los indicadores tomarán valores similares en los diferentes estratos.

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, por lo que si no se aplica incluso estamos mejorando la calidad de los resultados.

5.4

Otros factores de corrección

Tasa de falta de respuesta

La tasa de falta de respuesta, NRR por sus siglas en inglés (Non-Response Rate), puede aplicarse para incrementar el tamaño de muestra final, disimulando así los efectos que la falta de respuesta produce:

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

Factor de corrección para poblaciones pequeñas

Si la población objetivo es pequeña (por debajo de 10000), el tamaño requerido para conseguir la precisión deseada puede ser menor, por lo que se multiplicaría el tamaño final obtenido por un factor de corrección. Algunos *software* toman este factor de corrección como 0.9.

5.5

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.

Una vez determinado el número de conglomerados a incluir en la muestra, es conveniente elegir conglomerados de reserva, por si acaso en varios conglomerados de los previamente seleccionados fuera muy complicado conseguir la información. El número de conglomerados a elegir puede hacerse con la ayuda de la tabla:

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

Ilustración 4. Tabla de número recomendado de conglomerados de reserva

Sin embargo, sólo es preciso utilizar los conglomerados de reserva si no ha sido posible acceder al 10% o más del número total de conglomerados.

Si el tamaño de la 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.



5.6

Ejemplo

Supongamos que para estimar la prevalencia de un determinado indicador se tiene que la prevalencia esperada es del 10%. A continuación, vamos a calcular el tamaño de muestra necesario para dicha estimación, 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. En este caso, vamos a tomar valor para el efecto de 1.5, con lo que el nuevo tamaño de muestra será:

$$n^2 = n \cdot \text{DEFF} = n \cdot 1.5 = 576.24$$

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

$$n^2 = \frac{n}{1 - \text{NRR}} = \frac{576.24}{1 - 0.1} = 640.27$$

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

Cuando se realiza un muestreo por conglomerados, una vez que tenemos el tamaño de muestra final requerido, es necesario determinar el número de unidades muestrales que se incluirán de cada conglomerado, así como el número de conglomerados que se incluirán en la muestra.

Para esta selección del número de conglomerados puede realizarse otro muestreo que bien puede ser aleatorio simple o sistemático y las unidades de la muestra a repartir en cada uno de los conglomerados pueden calcularse asignando pesos a los conglomerados en función de su tamaño poblacional u otros criterios y multiplicando el tamaño de muestra final por cada uno de los pesos asignados en cada conglomerado.

6.

Pasos a seguir en cada tipo de muestreo con ejemplos

6.1

Muestreo Aleatorio Simple

1. Cada unidad en el marco (hogar, persona, etcétera.) lleva asignado un número único entre 1 y el número total de unidades.
2. Se usa un método aleatorio para seleccionar las unidades de inclusión dentro de la muestra. Este método puede aplicarse con Microsoft Excel o con algún software de tipo estadístico como SPSS o R.
3. A continuación, 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.

Alternativamente, se pueden acordar unidades de sustitución siguiendo alguna estrategia, como elegir el siguiente hogar más cercano o girar un lápiz frente al hogar ausente y elegir la siguiente casa a la que apunte el lápiz. Estos métodos de sustitución son menos importantes que la aplicación uniforme del procedimiento elegido.

El muestreo con probabilidades desiguales es una variación del muestreo aleatorio simple. La diferencia es que todos los individuos no tienen la misma probabilidad de ser seleccionados para la muestra, si no que tienen probabilidades proporcionales a alguna característica, como tamaño de la comunidad de la que proceden, del municipio, etc. Los pasos para aplicarlo son los mismo que los del muestreo aleatorio simple.

6.2

Muestreo Sistemático

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. Esta elección puede hacerse sobre el listado (si existe) o en la misma fase de campo.
4. Se establece la estrategia de sustitución de unidades.

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 y el muestreo sistemático actuaría como vemos en el siguiente ejemplo.

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.

En nuestro caso, supongamos que la primera unidad seleccionada aleatoriamente es la 2, por lo que sería la primera que tomaría para la muestra y a continuación iríamos construyendo los intervalos como se aprecia en la siguiente imagen, donde los individuos rojos serían los finalmente seleccionados.

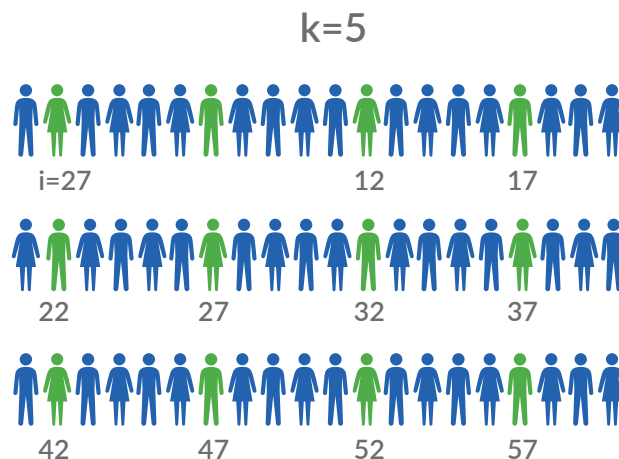


Ilustración 5. Representación gráfica de determinación del intervalo de selección

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:

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

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.

En la siguiente tabla vemos cómo sería la muestra resultante si aplicáramos el redondeo al alza, a la baja o alternando. Supongamos que el número aleatorio de inicio es 11.

	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

Ilustración 6. Tabla de selección de unidades en muestra sistemática según tipo de redondeo

Como vemos, si redondeamos al alza, la última unidad toma una posición que se sale de rango y por tanto no existe (sobrepasa el valor de $N=340$). El redondeo a la baja siempre es factible, pero el que da el reparto más equitativo y aproximado al reparto exacto es el redondeo alternando entre baja y alza.

6.3

Muestreo Estratificado

1. Se establecen los estratos. Los estratos suelen presentarse de forma natural: áreas geográficas con distintos recursos o niveles de pobreza, grupos de edad, etcétera.
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 un listado.

Ejemplo:

Supongamos que tenemos una población de tamaño 50000 habitantes, dividida en 3 zonas geográficas diferenciadas que tomamos como estratos. Estos estratos tienen respectivamente una población de 7000, 18000 y 25000 habitantes.

Queremos elegir una muestra de tamaño 750. En la siguiente tabla veremos cuál sería el tamaño de muestra que habría que seleccionar de cada estrato utilizando tamaños de muestra proporcionales al tamaño de los estratos.

En la primera columna podemos ver el número identificativo del estrato, en la segunda su población, en la tercera el peso (que se calcula dividiendo la población del estrato entre la población total), en la cuarta columna tenemos el múltiplo del peso del estrato por el tamaño de muestra a conseguir (en este caso 750) y en la última columna el tamaño de muestra final que se consigue redondeando el valor obtenido en la columna anterior, ya que el tamaño de muestra debe ser un número entero. En este caso no ha hecho falta redondear al no obtener cifras decimales, pero a causa de estos redondeos el tamaño de muestra podría verse ligeramente aumentado o disminuido, lo que no es significativo en tamaños de muestra grandes.

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

Ilustración 7. Tabla de reparto de muestra por estratos con probabilidades desiguales

6.4

Muestreo por Conglomerados

6.4.1. Selección de conglomerados y unidades mediante muestreo aleatorio.

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 según un muestreo aleatorio simple o con probabilidades desiguales.
3. Se seleccionan los conglomerados según el tipo de muestreo elegido, con la ayuda del software informático 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.

Ejemplo:

Supongamos que se realiza un estudio sobre los habitantes de un determinado municipio. En este municipio hay 200 comunidades que se caracterizan por ser homogéneas entre sí (todas se parecen), aunque están compuestas por unidades heterogéneas (personas). Con estas características se decide realizar un muestreo por conglomerados.

Primero se procede al cálculo de conglomerados que formarán parte de la muestra, lo cual se realiza con un muestreo con probabilidades desiguales, proporcional al tamaño del conglomerado. Esto quiere decir que los conglomerados de mayor tamaño tendrán más posibilidades de ser elegidos para la muestra.

El número de conglomerados establecido para la muestra es 20, que se seleccionan de entre la muestra de 200 comunidades (más los conglomerados de reserva que se conservan en caso de ser preciso).

Una vez seleccionados los conglomerados se reparte la muestra entre ellos. Suponemos que el tamaño de muestra final requerido (número de personas) se ha calculado y es igual a 10000 personas.

En la tabla siguiente veremos cuál sería el tamaño de muestra que habría que seleccionar de cada conglomerado utilizando tamaños de muestra proporcionales al tamaño de los conglomerados (probabilidades desiguales).

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

Ilustración 8. Tabla de reparto de muestra por conglomerados con probabilidades desiguales

Este reparto es muy parecido al anterior reparto por estratos, sólo que para calcular los pesos de los conglomerados se ha dividido entre el total de conglomerados incluidos en la muestra y no entre el total de todos los conglomerados del municipio

En la primera columna se recoge el número identificativo del conglomerado, en la segunda su población, en la tercera el peso, que se multiplica por 10000 en la cuarta columna para conocer el tamaño de muestra de cada conglomerado.

Si en lugar de repartir la muestra entre los conglomerados de forma proporcional se hubiera hecho de manera uniforme; esto es, igual tamaño de muestra en cada conglomerado solamente habría que dividir el tamaño de muestra (10000) entre el número de conglomerados (20), por lo que en cada conglomerado habría que seleccionar 500 personas. De esta forma, podría darse el caso de seleccionar todos los habitantes de una comunidad o que hubiera comunidades con menos de 500 habitantes. El problema se resuelve tomando en dichas comunidades todos los habitantes y repartiendo los faltantes, hasta alcanzar el tamaño de muestra mínimo, de forma uniforme entre el resto de conglomerados.

6.4.2. Selección de conglomerados mediante muestreo sistemático

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.

Ejemplo:

Supongamos que tenemos 100 municipios que actúan como conglomerados, de los cuales queremos elegir 20 para la muestra. La población total de los 100 municipios es 200000 habitantes. La siguiente tabla muestra los municipios con su población estimada.

Municipios	Población estimada	Población acumulada	Rango	Municipios seleccionados
Municipio 1	1500	1500	1-1500	x
Municipio 2	4000	5500	1500-5500	
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	

Ilustración 9. Tabla de lista de municipios con población estimada, acumulada y rango

Para calcular la población acumulada en la tercera columna, lo que se hace es ir sumando las poblaciones de los municipios anteriores en la lista ordenada.

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.

Posteriormente, se procedería a la elección de unidades dentro de los conglomerados. De esta forma, los conglomerados con mayor tamaño tienen más probabilidad de ser seleccionados al tener un rango más amplio, por eso el intervalo de muestreo se calcula a partir del tamaño de la población en lugar de con el número de conglomerados, ya que, de esta última forma, todos tendrían la misma probabilidad de ser elegidos.

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, sin embargo, es no tener conglomerados excesivamente extensos y si esto ocurre intentar subdividirlos en otros más pequeños, evitando de esta forma la selección doble.

6.5

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. Cuando los segmentos son desiguales, hay que procurar elegir el segmento de alguna forma aleatoria con probabilidad proporcional a su tamaño. Sin embargo, lo recomendable sería realizar un muestreo trietápico.

6.6

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.

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.

Si seleccionamos en primer lugar algunos conglomerados, y dentro de estos se selecciona una muestra de subconglomerados, para terminar, seleccionando algunas unidades elementales de éstos, tenemos un muestreo trietápico.

Ejemplo:

En poblaciones grandes, es habitual tener una población clasificada por estratos. Por ejemplo, pueden considerarse como estratos las distintas zonas de medios de vida (ZMV). Supongamos que, en una población ficticia, se consideran 5 ZMV, que serían los estratos de partida. De cada una de estas ZMV se desea seleccionar un número determinado de comunidades, ya que por razones económicas y de tiempo es imposible considerar todas las comunidades para la encuesta. De entre las comunidades seleccionadas, a su vez, se selecciona una muestra de hogares, a los que se le realizará la entrevista de manera telefónica. Dentro de los hogares, se le hará la entrevista a la persona mayor de edad cuya fecha de cumpleaños esté más próxima al momento de la llamada.

Estamos en el caso de un muestreo trietápico, con probabilidades desiguales en las dos primeras etapas e iguales en la tercera, combinado con estratificación. Los estratos, como hemos dicho, son las ZMV.

Se establecen las siguientes consideraciones:

1. En la primera etapa del muestreo trietápico, se decide la selección de comunidades con probabilidad proporcional al tamaño de los estratos a los que pertenecen, medido este tamaño en número de comunidades.

2. En la segunda etapa se seleccionan los 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. En la tercera etapa se selecciona el 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.

El número total de comunidades en los estratos es 550 comunidades y la muestra calculada es de 30 de entre estas 550.

Etapa 1

En la tabla podríamos ver el número de comunidades seleccionadas en cada estrato de acuerdo con el número de comunidades total en los estratos y el peso calculado de los mismos.

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

Ilustración 10: Tabla de número de comunidades por estrato a seleccionar para la muestra

En cada estrato podría seleccionarse de forma aleatoria una comunidad extra que haría de reserva por si de alguna se consigue menos de un 80% de la muestra.

Con esto se completaría la primera etapa del muestreo trietápico, que es la selección de las comunidades (conglomerados).

Etapa 2

En la segunda etapa se procede a la selección de los hogares (subconglomerados).

Supongamos que se ha calculado un tamaño de muestra para el número de hogares igual a 630 hogares de entre un total de 10000 hogares según el censo que se posee.

Como se ha establecido que el reparto de hogares se hará teniendo en cuenta el tamaño de los estratos y dentro de estos el tamaño de las comunidades, primero se procede a calcular el número de hogares por estrato. Para ello, vuelve a calcularse el peso de los estratos, pero esta vez atendiendo al número de hogares que los componen y se multiplica este peso por la muestra final a recoger.

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

Ilustración 11. Tabla de número de hogares por estrato a seleccionar para la muestra

Nótese que a causa del redondeo se ha incrementado en una unidad el tamaño de muestra final. Ya sabemos el número de hogares en cada estrato y también el número de comunidades en cada estrato (de la etapa anterior). Ahora, dentro de cada estrato, se repartiría el número de hogares entre las comunidades seleccionadas. Dicho reparto también se realiza de forma proporcional al número de hogares de las comunidades.

En la tabla siguiente vemos cómo sería el reparto dentro del Estrato 1, en el que hay que repartir 63 hogares en 5 comunidades, teniendo el primer estrato un total de 1000 hogares.

Estratos 1	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

Ilustración 12. Tabla de número de hogares por comunidad en el estrato 1

Este procedimiento se repetiría para el resto de estratos.

Etapa 3

Ya sabemos el número total de hogares en cada una de las 30 comunidades. 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.



GUIONES DE PRÁCTICAS DISEÑO MUESTRAL





Descarga de EPIDAT 4.2

Epidat es un programa de libre distribución desarrollado por el Servicio de Epidemiología de la Dirección Xeral de Saúde Pública da Consellería de Sanidade (Xunta de Galicia, España), con el apoyo de la Organización Panamericana de la Salud (OPS-OMS) y la Universidad CES de Colombia. Utilizaremos este programa para las prácticas de diseño muestral por ser de libre distribución, además de ser sencillo de usar y tener implementados todos los tipos de muestreo básicos que usaremos en este curso.

La versión más reciente disponible es la 4.2 Para descargarla accederemos a la dirección:



<https://extranet.sergas.es/EPIWB/EPIWB/SolicitudEpidat.aspx?IdPaxina=62715&idv=4&lng=es>

En dicha página introduciremos nuestro e-mail y nos registraremos.

Posteriormente seleccionaremos la descarga de Epidat 4.2 en un solo archivo, que descargaremos en formato comprimido en nuestro ordenador y seguiremos los siguientes pasos una vez descargado:

1.

Descomprimir el archivo.

Buscar el archivo en la ubicación de descarga.

Pulsar sobre él con el botón de la derecha del ratón y seleccionar "Extraer en Epidat 4.2".

2.

Buscar en la carpeta descomprimida el archivo "Epidat.jar".

Abrirlo haciendo doble click sobre él con el ratón y el programa se abrirá (no necesita instalación).

Si cuando hacemos doble click no se produce ninguna acción es porque necesitamos tener instalado java, para ello acceder a la dirección:

<http://www.java.com/es/download/>

a.

En dicha página pulsar en “Descarga gratuita de Java”.

“Aceptar e iniciar descarga gratuita” y guardaremos el archivo de java en el ordenador.

b.

Buscamos el instalador en la ubicación de descarga y haciendo doble click sobre él, instalaremos java.

c.

Una vez instalado, procederemos de nuevo a ejecutar el archivo “Epidat.jar”.



Práctica 1.

Muestreo Aleatorio Simple y Muestreo Aleatorio Sistemático



El archivo Empleados.xlsx contiene información de 474 empleados y empleadas que trabajan en una determinada empresa, concretamente se recoge la siguiente información:



Id

Número de identificación del empleado.



Sexo

Toma el valor 1 si el empleado es mujer y el valor 2 si el empleado es hombre.



Departamento

Departamento donde trabaja el empleado con códigos que van desde 11 a 19.

1.2.3

Categoría

Categoría del empleado con clasificaciones: 1, 2 y 3.

Se desea estimar la proporción de empleados que llevan trabajando x meses en la empresa, para lo cual se selecciona una muestra de 20 empleados.

Seleccionar dicha muestra con Epidat 4.2 y con Excel de Microsoft, considerando:

1. Un muestreo aleatorio simple.

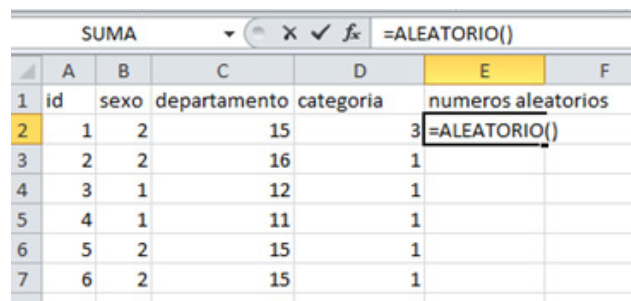
2. Un muestreo sistemático.

Resolución

Generación de muestra con Microsoft Excel

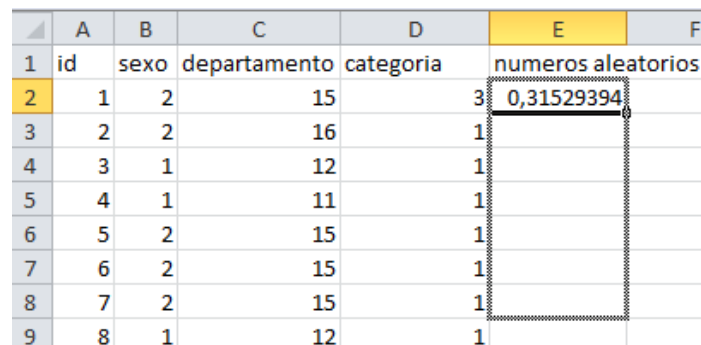
● Muestreo aleatorio simple

1. Nos situamos en la siguiente columna vacía del archivo de datos y escribimos en la primera casilla: **=ALEATORIO()**



	A	B	C	D	E	F
1	id	sexo	departamento	categoria	numeros aleatorios	
2	1	2	15	3	=ALEATORIO()	
3	2	2	16	1		
4	3	1	12	1		
5	4	1	11	1		
6	5	2	15	1		
7	6	2	15	1		

2. Nos situamos en la esquina inferior derecha de la casilla donde se ha generado el primer número aleatorio y la desplazamos hasta el final de la columna.



	A	B	C	D	E	F
1	id	sexo	departamento	categoria	numeros aleatorios	
2	1	2	15	3	0,31529394	
3	2	2	16	1		
4	3	1	12	1		
5	4	1	11	1		
6	5	2	15	1		
7	6	2	15	1		
8	7	2	15	1		
9	8	1	12	1		

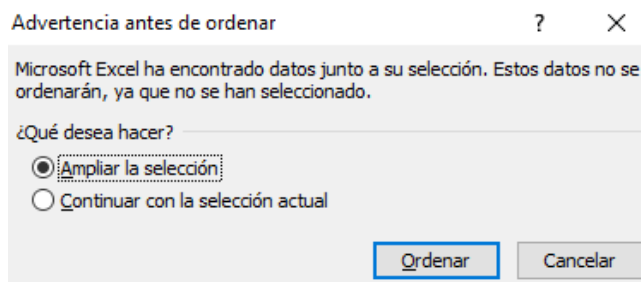
3. A continuación, copiamos esos números aleatorios y los pegamos como valores, para que dejen de ser una función aleatoria que estaría continuamente actualizándose.

	A	B	C	D	E	F	G
1	id	sexo	departamento	categoria	numer		
2	1	2		15	3	0,284	
3	2	2		16	1	0,328	
4	3	1		12	1	0,057	
5	4	1		11	1	0,904	
6	5	2		15	1	0,116	
7	6	2		15	1	0,225	
8	7	2		15	1	0,331	
9	8	1		12	1	0,404	

4. Una vez tenemos los números aleatorios, ordenamos el fichero según el orden de números aleatorios. Para ello, seleccionamos la columna de números aleatorios y pulsamos en **“Ordenar y Filtrar”** → **“Ordenar de menor a mayor”**.

	A	B	C	D	E	F	G	H	I
1	id	sexo	departamento	categoria	numeros aleatorios				
2	1	2		15	3	0,28425739			
3	2	2		16	1	0,32832354			
4	3	1		12	1	0,05740597			
5	4	1		11	1	0,90467664			

5. Indicamos que **“amplíe la selección”** para que ordene no sólo la columna de números aleatorios, si no el resto de columna de acuerdo con estos números.



- Una vez ordenados, seleccionamos los 20 primeros. Podemos identificar los empleados seleccionados a través del número de identificación.

	A	B	C	D	E
1	id	sexo	departamento	categoria	numeros aleatorios
2	264	1	12	1	0,002220067
3	427	2	11	1	0,003990219
4	435	2	15	1	0,011854613
5	107	1	12	1	0,012329738
6	322	1	12	1	0,012742298
7	400	1	12	1	0,012849884
8	131	2	15	1	0,019455631
9	237	1	12	1	0,019561683
10	474	1	12	1	0,019932512
11	472	2	15	1	0,022643234
12	442	1	12	1	0,023146541
13	269	2	15	1	0,023570356
14	293	2	15	1	0,025020051
15	394	1	11	1	0,02546732
16	207	2	15	1	0,026783217
17	297	1	12	1	0,03131339
18	261	1	12	1	0,03146839
19	465	2	12	1	0,033862685
20	257	2	19	3	0,041116415
21	87	2	12	1	0,044123713

Nótese que cada vez que realizamos este proceso, obtenemos una muestra diferente.

● Muestreo sistemático

- Para conocer el intervalo de selección, **dividimos $N=474$ entre $n=20$, que es $k=23.7$.**

Para el redondeo de dicho valor puede utilizarse la regla de oro, que consiste en alternar las unidades en cada intervalo entre 23 y 24 o redondear a la baja, lo que funciona para cualquier valor de k y simplifica el cálculo en Excel.

2.

En la siguiente columna que tenemos disponible en Excel, numeramos los elementos de la muestra con números del 1 al 20 y en la siguiente columna generamos la semilla aleatoria entre 1 y 23.

Para ello, escribimos: **ALEATORIO.ENTRE(1;23)**. Esto nos dará el primer id del empleado o empleada que irá en la muestra de tamaño 20.

SUMA						
	A	B	C	D	E	F
1	id	sexo	departamento	categoria	muestra	
2	264	1	12	1	1	=ALEATORIO.ENTRE(1;23)
3	427	2	11	1	2	
4	435	2	15	1	3	
5	107	1	12	1	4	
6	322	1	12	1	5	
7	400	1	12	1	6	
8	131	2	15	1	7	
9	237	1	12	1	8	
10	474	1	12	1	9	
11	472	2	15	1	10	
12	442	1	12	1	11	
13	269	2	15	1	12	
14	293	2	15	1	13	
15	394	1	11	1	14	
16	207	2	15	1	15	
17	297	1	12	1	16	
18	261	1	12	1	17	
19	465	2	12	1	18	
20	257	2	19	3	19	
21	87	2	12	1	20	
22	130	2	19	3		

3.

Selecciono el resto de empleados en intervalos de 23, para lo que, en la segunda casilla de la columna en la que he generado la semilla aleatoria, escribo: **=F2+23**

SUMA						
	A	B	C	D	E	F
1	id	sexo	departamento	categoria	muestra	id
2	264	1	12	1	1	9
3	427	2	11	1	2	=F2+23
4	435	2	15	1	3	
5	107	1	12	1	4	

Nos situamos en la esquina inferior derecha de la casilla donde se ha generado la suma, y la desplazamos hasta el final de la muestra de tamaño 20.

	A	B	C	D	E	F	
1	id	sexo	departamento	categoria	muestra	id	
2	264	1	12	1	1	14	
3	427	2	11	1	2	37	
4	435	2	15	1	3	60	
5	107	1	12	1	4	83	
6	322	1	12	1	5	106	
7	400	1	12	1	6	129	
8	131	2	15	1	7	152	
9	237	1	12	1	8	175	
10	474	1	12	1	9	198	
11	472	2	15	1	10	221	
12	442	1	12	1	11	244	
13	269	2	15	1	12	267	
14	293	2	15	1	13	290	
15	394	1	11	1	14	313	
16	207	2	15	1	15	336	
17	297	1	12	1	16	359	
18	261	1	12	1	17	382	
19	465	2	12	1	18	405	
20	257	2	19	3	19	428	
21	87	2	12	1	20	451	
22	130	2	19	3			

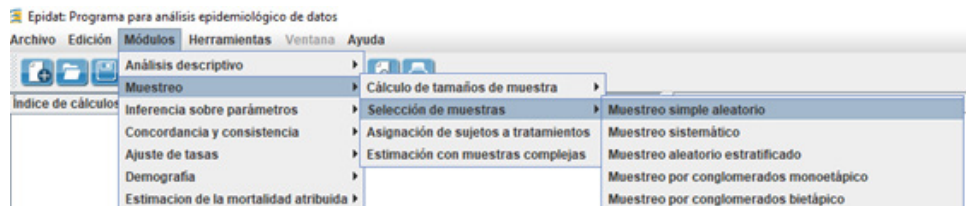
Esto nos da el número de identificación de los 20 empleados de la muestra.



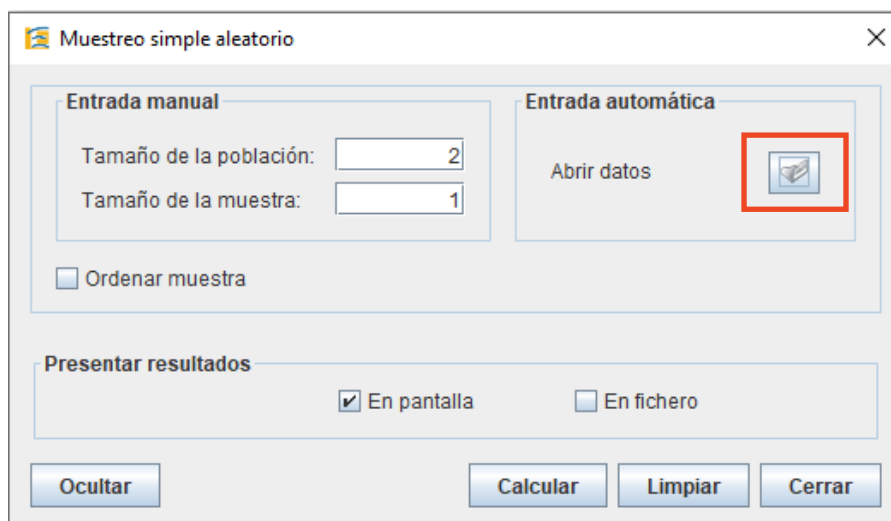
Generación de muestra con Epidat 4.2

● Muestreo aleatorio simple

1. En Epidat 4.2, pulsamos en **“Módulos”** → **“Muestreo”** → **“Selección de muestras”** → **“Muestreo simple aleatorio”**



2. Pulsamos en el icono **“Abrir datos”** y, en **“Examinar”**, le indicamos dónde están los datos. Luego, pulsamos en **“Aceptar”**.



3. Una vez cargados los datos, nos calcula automáticamente el tamaño de la población, con lo que sólo tenemos que indicar el tamaño de la muestra. Podemos pedirle que presente los resultados por pantalla o en un fichero, o bien en ambas opciones. Finalmente, le damos a “Calcular”.

Muestreo simple aleatorio

Entrada manual

Tamaño de la población: 474

Tamaño de la muestra: 20

Entrada automática

Abrir datos

Ordenar muestra

Presentar resultados

En pantalla En fichero

Ocultar Calcular Limpiar Cerrar

4. Si hemos pedido que los presente en pantalla, nos dará una tabla que contiene el Id de los veinte empleados seleccionados para la muestra.

Datos:

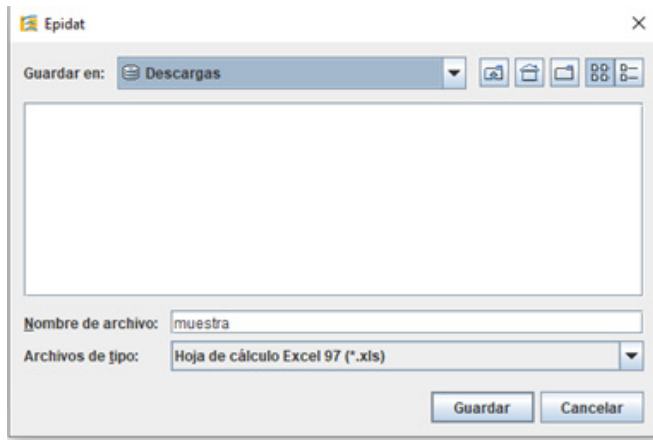
Tamaño de la población: 474
Tamaño de la muestra: 20

Número de los sujetos seleccionados:

233	193	116	369	3	315	301
216	468	241	473	231	14	341
453	355	16	54	416	41	

Probabilidad de selección: 4,2194%

5. Si le pedimos que los guarde en un fichero, tendremos que indicarle el nombre del fichero y la ubicación donde queremos que lo guarde.

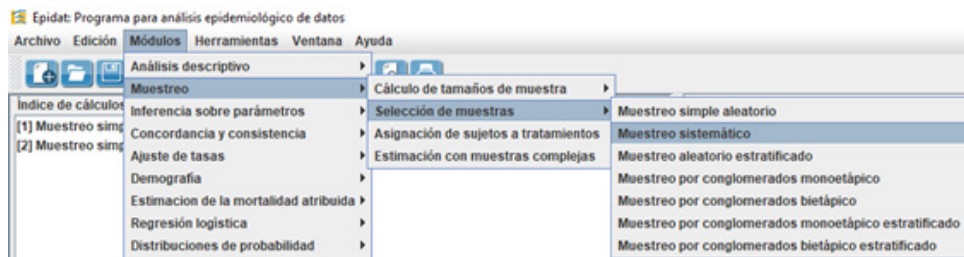


Dicho fichero contendrá, no únicamente el Id de los empleados, sino también el resto de campos recogidos en el archivo de datos.

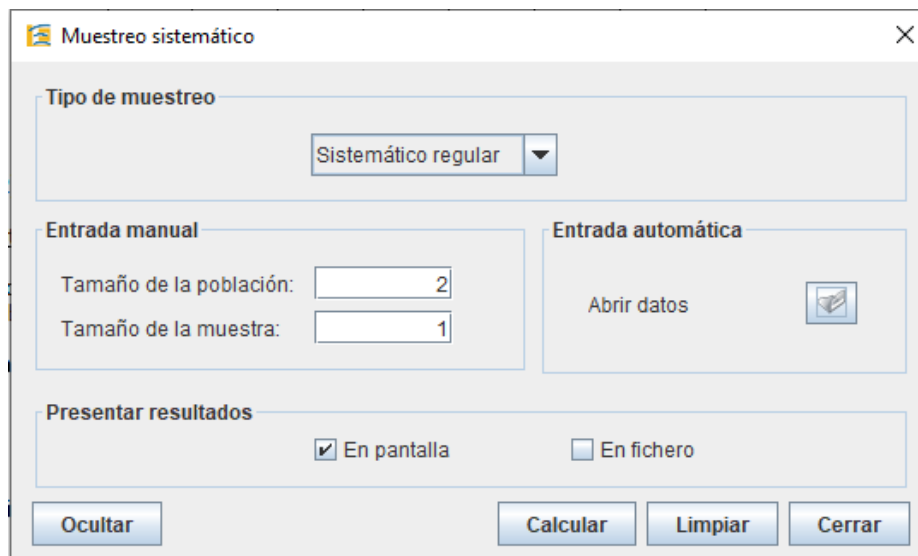
	A	B	C	D
1	id	sexo	departame	categoria
2	85	2	15	1
3	78	1	12	1
4	419	2	15	1
5	360	1	12	1
6	336	2	16	3
7	220	2	12	1
8	322	1	12	1
9	47	1	12	1
10	121	1	15	1
11	335	2	11	2
12	217	2	16	1
13	28	2	15	1
14	368	1	12	1
15	234	2	15	1
16	124	1	16	1
17	77	1	12	1
18	261	1	12	1
19	408	2	19	3
20	147	2	15	1
21	280	1	12	1
22				

● Muestreo sistemático

1. En Epidat 4.2 pulsamos en “Módulos” → “Muestreo” → “Selección de muestras” → “Muestreo sistemático”.



2. Dejamos por defecto “Sistemático Regular” y pulsamos en “Abrir Datos”.



3. Una vez abiertos los datos, indicamos el tamaño de la muestra y la forma en que queremos que nos muestre los resultados: por pantalla o en fichero.

4. La salida por pantalla será similar a la siguiente:

Datos:

Tipo de muestreo: Sistemático regular
 Tamaño de la población: 474
 Tamaño de la muestra: 20

Resultados:

Intervalo de selección	Arranque aleatorio
23	15

Número de los sujetos seleccionados:

15	38	61	84	107	130	153
176	199	222	245	268	291	314
337	360	383	406	429	452	

Tamaño de muestra resultante: 20
 Probabilidad de selección: 4,2194%

Si la solicitamos por fichero, éste contendrá el resto de campos de la base de datos, además del Id de los empleados seleccionados.

Práctica 2.

Muestreo Estratificado y Muestreo por Conglomerados en una etapa



Para esta práctica, volveremos a utilizar el archivo Empleados.xlsx, que, recordemos, contiene la siguiente información de 474 empleados y empleadas de una determinada empresa:



Id

Número de identificación del empleado.



Sexo

Toma el valor 1 si el empleado es mujer y el valor 2 si el empleado es hombre.



Departamento

Departamento donde trabaja el empleado con códigos que van desde 11 a 19.

1.2.3

Categoría

Categoría del empleado con clasificaciones: 1, 2 y 3.

Es este caso se pide:

1. Seleccionar una muestra de 20 empleados según un muestreo estratificado, donde los estratos son las distintas categorías a las que pertenecen los empleados.
2. Seleccionar una muestra de 4 de los 8 departamentos, considerando los departamentos como conglomerados.

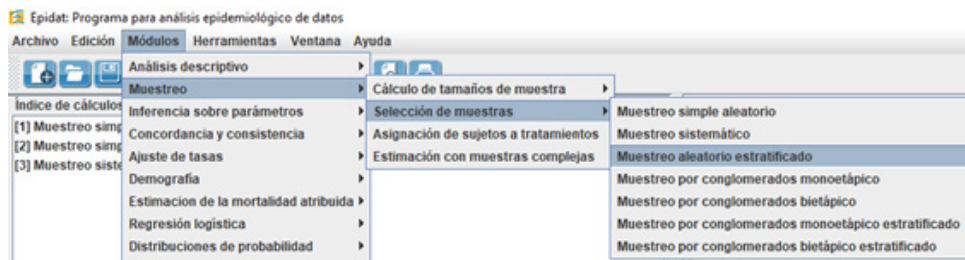
Hacerlo a través de Microsoft Excel y Epidat 4.2.

Resolución

Generación de muestra con Epidat 4.2

● Muestreo estratificado

1. En Epidat 4.2 pulsamos en “Módulos” → “Muestreo” → “Selección de muestras” → “Muestreo aleatorio estratificado”.



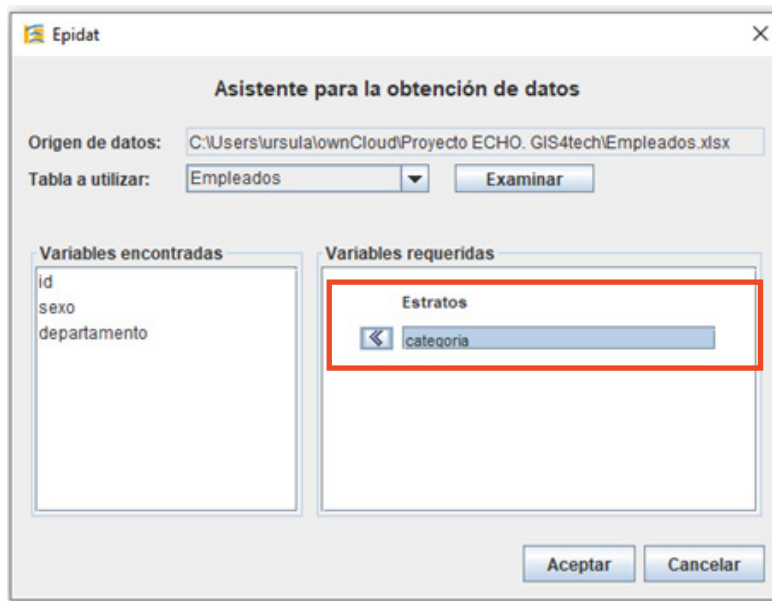
2. Seleccionar “Abrir datos individuales”.

The screenshot shows the 'Muestreo aleatorio estratificado' dialog box. The 'Entrada automática' section is selected, and the 'Abrir datos individuales' radio button is chosen. The 'Número de estratos' is set to 2. The 'Reparto de la muestra' is set to 'Muestra igual para todos los estratos'. The table below shows the sample sizes for each stratum.

Estrato	Tamaño del estrato	Tamaño de la muestra
1	0	0
2	0	0

The 'Tamaño de la muestra en cada estrato' is set to 0. The 'Presentar resultados' section has 'En pantalla' checked. The 'Cargar' button is visible next to the sample size input.

Indicamos la ubicación de nuestros datos, y como estratos seleccionamos la variable **“categoría”**.
Luego, pulsamos en **“Aceptar”**.

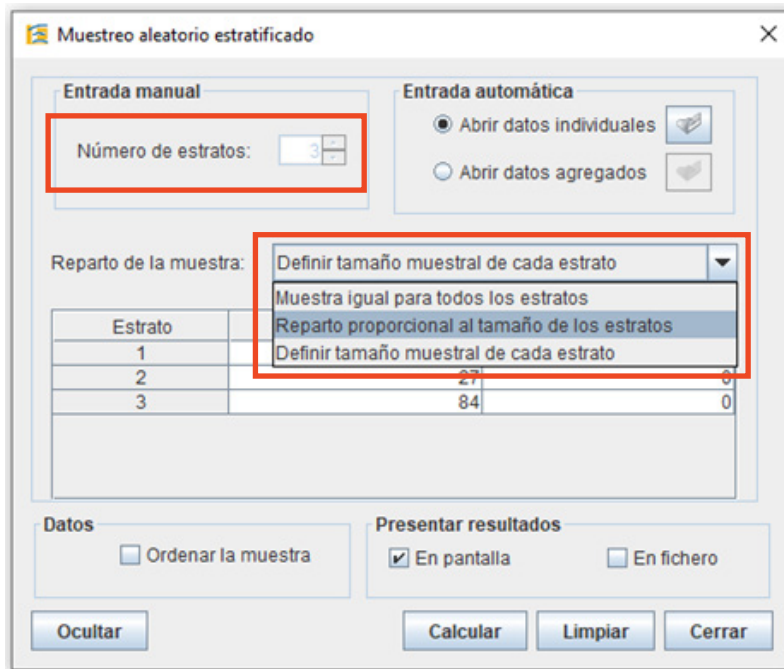


3. Automáticamente, Epidat 4.2 cuenta el número de estratos y nos indica que tenemos 3.

A continuación, en el menú desplegable, indicamos la forma en que repartiremos la muestra entre esos 3 estratos:

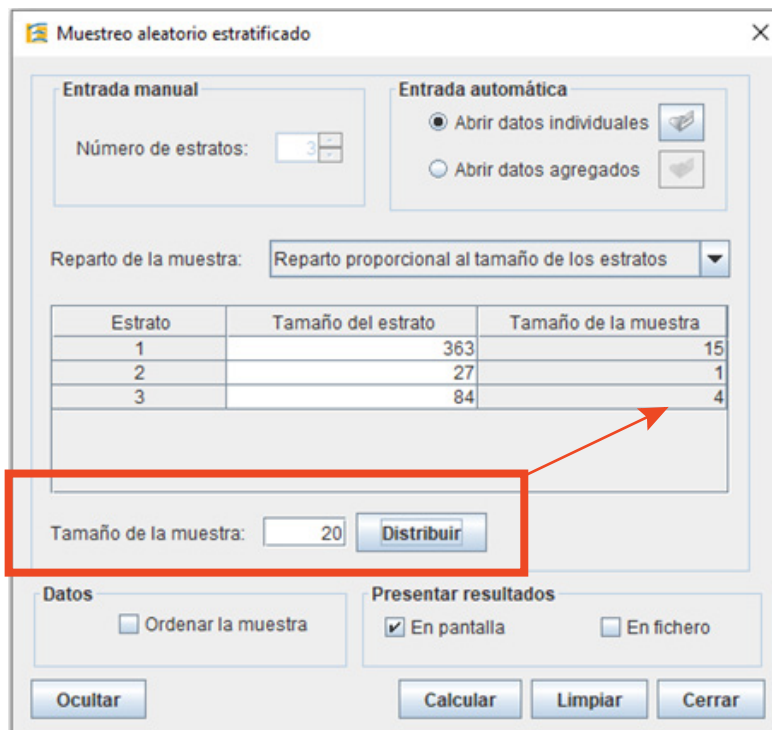
- a. Con afijación uniforme:**
Mismo tamaño de muestra en todos los estratos.
- b. Con afijación proporcional:**
Tamaño de muestra proporcional al tamaño de los estratos.
- c.** Tamaño de muestra en cada estrato definido por el usuario.

En este caso elegiremos **“Reparto proporcional al tamaño de los estratos”**.



4. Indicamos el tamaño de muestra y pulsamos en **“Distribuir”**.

De esta forma la muestra se reparte entre los 3 estratos, con el método de reparto que hayamos indicado en la fase anterior.



5. Indicamos la forma en que queremos que nos presente los resultados. Ésta sería la salida por pantalla:

Datos:

Reparto de la muestra: Reparto proporcional al tamaño de los estratos

Tamaño de la muestra: 20

Estrato	Tamaño del estrato	Tamaño de la muestra
1	363	15
2	27	1
3	84	4
TOTAL	474	20

Probabilidades de selección y ponderaciones:

Estrato	Probabilidad de selección (%)	Ponderaciones
1	4,1322	24,2000
2	3,7037	27,0000
3	4,7619	21,0000

Número de los sujetos seleccionados:

Estrato 1:

85	191	282	9	38	241	350
206	179	59	83	35	260	244
55						

Estrato 2:

24

Estrato 3:

64	35	16	18
----	----	----	----

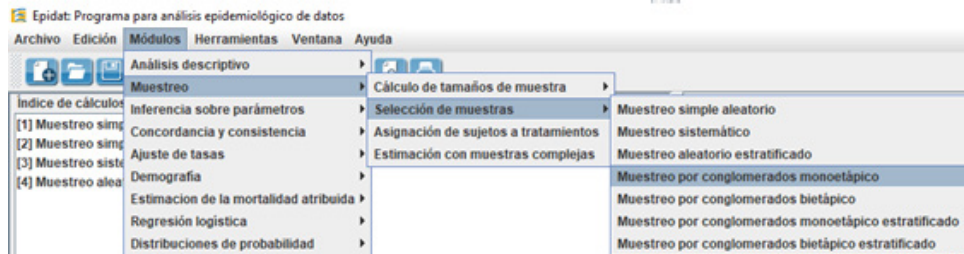
Vemos que nos muestra los tamaños totales de los estratos y los tamaños muestrales en la primera tabla.

En la segunda, la probabilidad de selección, que es la probabilidad común que se quiere otorgar a cada elemento del estrato, el que a su vez representa la fracción del tamaño total del estrato que quedará en la muestra. Asimismo, representa las ponderaciones, que las calcula como en tamaño total del estrato entre su tamaño muestral.

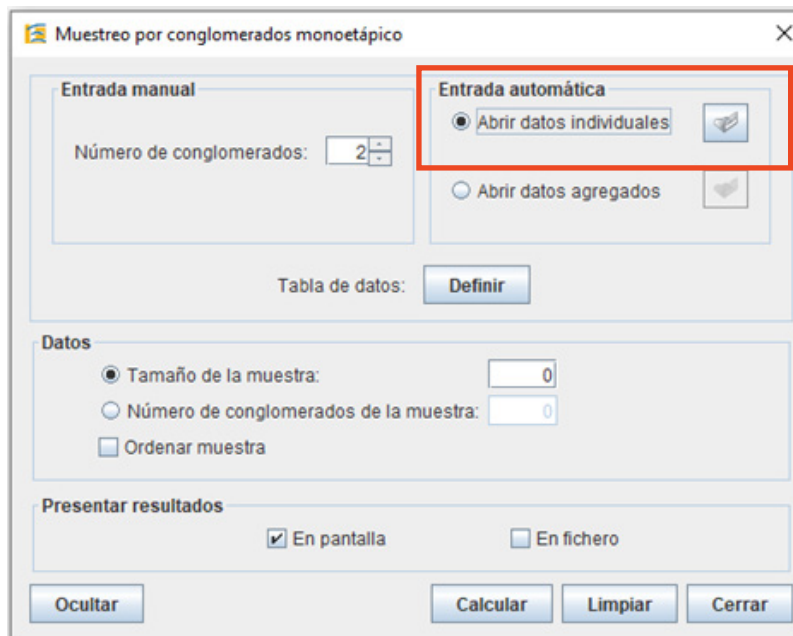
Por último, en las tres tablas finales nos indica los Id de los empleados seleccionados de cada estrato.

● Muestreo por conglomerados en una etapa

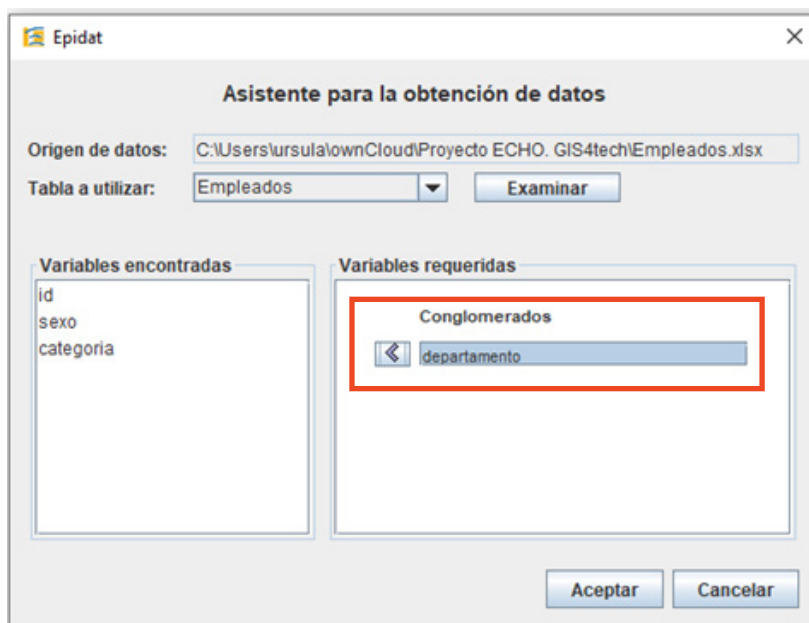
1. En Epidat 4.2 pulsamos en “Módulos” → “Muestreo” → “Selección de muestras” → “Muestreo por conglomerados monoetápico”.



2. Seleccionamos “Abrir datos individuales”.

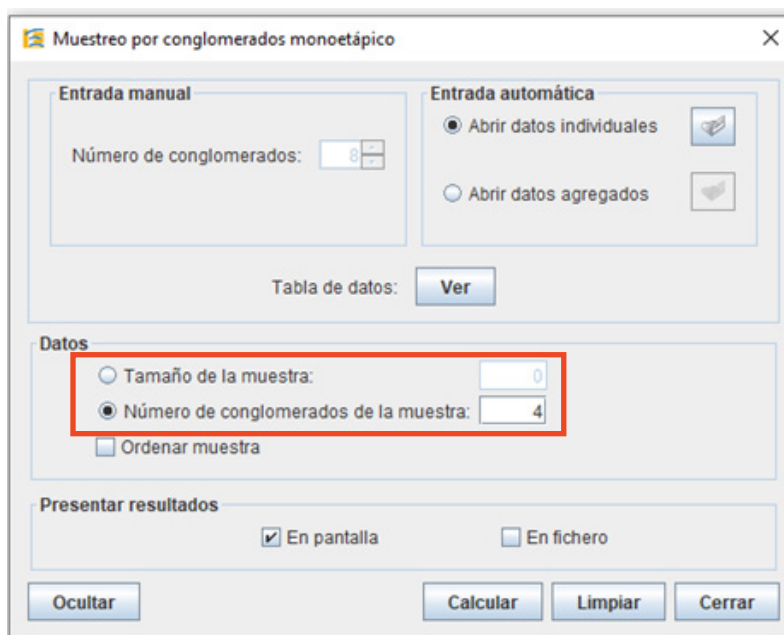


3. Indicamos de nuevo la ubicación de los datos y seleccionamos **“departamento”** como conglomerados.



4. Una vez que hemos indicado los conglomerados, tenemos dos opciones:

- a. Indicar el **“tamaño de muestra”** para que el programa calcule el número de conglomerados que habría que seleccionar para esa muestra, y el tamaño de muestra de cada uno de ellos.
- b. O bien, indicar directamente el **“número de conglomerados”** que queremos seleccionar para la muestra, en nuestro caso 4.



5. El resultado por pantalla sería parecido al siguiente:

Datos:

	Población	Muestra
Tamaño	474	195
Nº de conglomerados	8	4

Conglomerados seleccionados:

	Tamaño
16	59
15	116
18	9
17	11
TOTAL	195

Probabilidad de selección: 50,0000%

Como vemos, el programa nos muestra los 4 conglomerados seleccionados junto con el tamaño de cada uno de ellos.

En caso de haber seleccionado, por ejemplo, una muestra de tamaño 100, en lugar del número de conglomerados, el programa hubiera dado la siguiente salida:

Datos:

	Población	Muestra
Tamaño	474	100
Nº de conglomerados	8	2

Conglomerados seleccionados:

	Tamaño
14	6
15	116
TOTAL	122

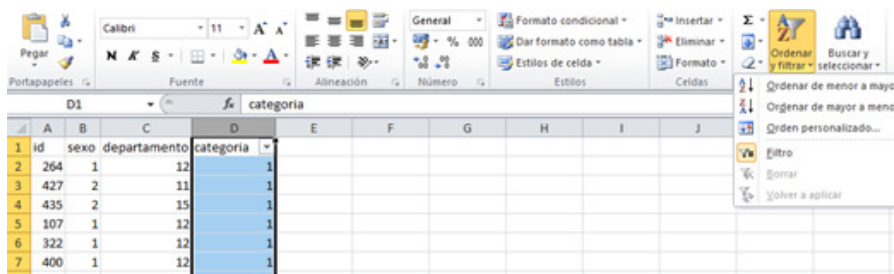
Probabilidad de selección: 25,0000%

Vemos que ha calculado un tamaño igual a 2 para el número de conglomerados a elegir, a su vez que selecciona dichos conglomerados.

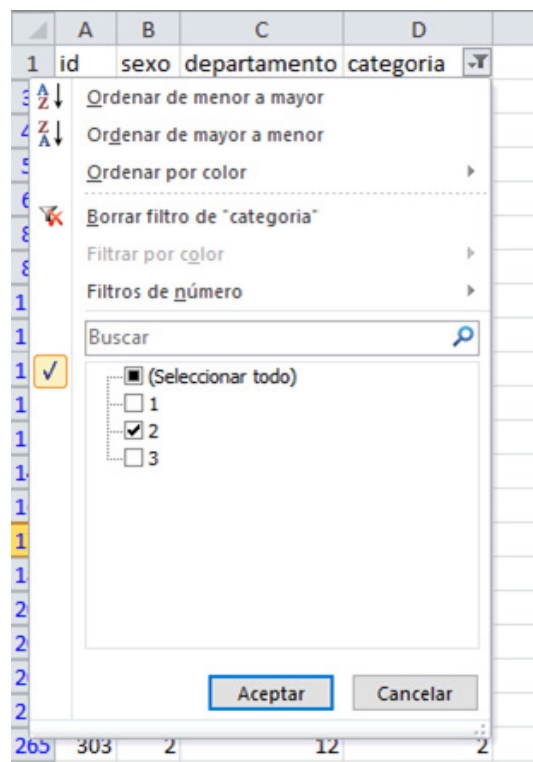
Generación de muestra con Microsoft Excel

● Muestreo estratificado

1. Lo primero que necesitamos es conocer la población de cada estrato, para hacer el reparto muestral con afijación proporcional. Para esto, lo que hacemos es un filtrado de datos: seleccionamos la variable **“categoría”** y en **“Ordenar y filtrar”** elegimos **“Filtro”**.



2. Con esto hemos creado un desplegable en la columna **“categoría”** que nos permite seleccionar su valor. Si seleccionamos por ejemplo, el valor **“2”**, correspondiente al segundo estrato y pulsamos en **“Aceptar”**:



Obtenemos la lista de los empleados y las empleadas en el segundo estrato:

	A	B	C	D	
1	id	sexo	departamento	categoria	↕
39	111	2	12	2	
45	414	2	11	2	
55	112	2	12	2	
69	98	2	11	2	
80	96	2	11	2	
82	273	2	12	2	
111	281	2	11	2	
115	291	2	12	2	
116	206	2	12	2	
131	255	2	12	2	
133	326	2	11	2	
149	285	2	11	2	
160	45	2	12	2	
179	353	2	12	2	
188	385	2	12	2	
204	126	2	15	2	
206	152	2	11	2	
208	127	2	12	2	
230	305	2	12	2	
265	303	2	12	2	
293	185	2	11	2	
323	48	2	12	2	
345	335	2	11	2	
355	386	2	11	2	
384	174	2	11	2	
407	429	2	11	2	
424	213	2	11	2	
476					

Seleccionados los valores, los copiamos en otra hoja de Excel, que llamaremos Estrato 2 y que como podemos ver, contiene 27 empleados.

	A	B	C	D	E
1	id	sexo	departamen	categoria	
2	111	2	12	2	
3	414	2	11	2	
4	112	2	12	2	
5	98	2	11	2	
6	96	2	11	2	
7	273	2	12	2	
8	281	2	11	2	
9	291	2	12	2	
10	206	2	12	2	
11	255	2	12	2	
12	326	2	11	2	
13	285	2	11	2	
14	45	2	12	2	
15	353	2	12	2	
16	385	2	12	2	
17	126	2	15	2	
18	152	2	11	2	
19	127	2	12	2	
20	305	2	12	2	
21	303	2	12	2	
22	185	2	11	2	
23	48	2	12	2	
24	335	2	11	2	
25	386	2	11	2	
26	174	2	11	2	
27	429	2	11	2	
28	213	2	11	2	
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					

Hacemos esto para el resto de estratos y verificamos que sus tamaños muestrales son: 363, 27 y 84, respectivamente.

3. Una vez que conocemos los tamaños totales de los estratos, repartimos nuestra muestra de tamaño 20 con afijación proporcional. Para ello, hacemos uso de la herramienta Excel “**Cálculo del tamaño de muestra**”, que se proporciona para dichos cálculos y que ya tiene implementadas las fórmulas para el cálculo de pesos en los estratos (hoja “Muestra por estratos”).

En dicha herramienta, introducimos **el tamaño total de la población, $N = 474$, el tamaño de muestra requerido, $n = 20$, el número de estratos, $k=3$ y la población de cada estrato**, calculada en el paso anterior.

MUESTREO ALEATORIO ESTRATIFICADO CON AFIJACIÓN PROPORCIONAL			
Tamaño de la población objetivo	474		
Tamaño de la muestra	20		
Número de estratos	3		

Estrato	Población estrato	Proporción	Muestra en el estrato
Estrato 1	363	76,58227848	15
Estrato 2	27	5,696202532	1
Estrato 3	84	17,72151899	4
Estrato 4		0	0

Con esos datos, el tamaño de muestra en cada uno de los estratos es 15, 1 y 4, respectivamente.

4. Por último, queda por hacer la selección aleatoria de muestra en los estratos. *Este proceso habrá que realizarlo tres veces, una para cada estrato.*

Para ello, utilizamos los datos que tenemos localizados en las hojas de Excel que hemos creado con el filtrado por categorías para contabilizar el total de empleados por estrato y, sobre cada una de estas bases de datos, realizamos un muestreo aleatorio simple con el tamaño de muestra requerido, como se hizo en la Práctica 1.

● Muestreo por conglomerados en una etapa

1. Primero necesitaríamos el tamaño de los conglomerados para emplear una afijación proporcional, para lo que emplearíamos un filtrado de la variable **“departamento”**, ya que los distintos departamentos son los que tomamos como conglomerados.

The screenshot shows an Excel spreadsheet with columns A through F. Column A contains 'id', B contains 'sexo', C contains 'departamento', and D contains 'categoria'. A filter menu is open over column C, showing options like 'Ordenar de menor a mayor', 'Ordenar de mayor a menor', 'Ordenar por color', 'Borrar filtro de "departamento"', 'Filtrar por color', and 'Filtros de número'. The 'Filtros de número' option is selected, and a sub-menu is open showing a list of numbers from 11 to 19. The number 11 is selected, and the number 12 is highlighted in blue. The status bar at the bottom of the spreadsheet shows 'Se encontraron 190 de 474 registros'.

	A	B	C	D	E	F
1	id	sexo	departamento	categoria		
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42	180	1	12			
43	444	1	12			
44	140	1	12			
45	112	2	12			
46	459	1	12			
47	366	1	12			
48	166	1	12			
49	280	1	12			
50	368	1	12			
51	24	1	12			
52	354	1	12			
53	460	1	12			
54	402	1	12			
55	273	2	12			
56	361	1	12			
57	132	2	12			
58	219	2	12			
59	324	1	12			
60	55	2	12			
61	334	1	12			
62	10	1	12			

Vemos que hay 190 empleados en el departamento 11. Copiaríamos la lista de estos empleados en una nueva hoja de Excel, si luego queremos realizar sobre esta hoja la selección de muestras, y así procederíamos para el resto de departamentos.

2. Una vez que conocemos los tamaños de los conglomerados, repartimos nuestra muestra de tamaño 20 con afijación proporcional. Para ello, en la herramienta de Excel “**Cálculo del tamaño de muestra**” y en la hoja de “**Muestra por conglomerados**”, escribimos los tamaños de los conglomerados que hemos ido calculando, con el fin de conocer el peso.

Conglomerado	Tamaño conglomerado	Peso
Conglomerado 11	53	0,111814346
Conglomerado 12	190	0,400843882
Conglomerado 13	0	0
Conglomerado 14	6	0,012658228
Conglomerado 15	116	0,244725738
Conglomerado 16	59	0,124472574
Conglomerado 17	11	0,023206751
Conglomerado 18	9	0,018987342
Conglomerado 19	30	0,063291139

3. Los valores de esta tabla los pegamos en otra hoja nueva de Excel y en la siguiente columna, generamos números aleatorios como hicimos en la Práctica 1.

	A	B	C	D
1	Conglomerados	tamaño	Peso	Números aleatorios
2	Conglomerado 11	53	0,11181435	0,228577761
3	Conglomerado 12	190	0,40084388	0,972140981
4	Conglomerado 13	0	0	0,138437623
5	Conglomerado 14	6	0,01265823	0,039222172
6	Conglomerado 15	116	0,24472574	0,402437909
7	Conglomerado 16	59	0,12447257	0,606563362
8	Conglomerado 17	11	0,02320675	0,671286449
9	Conglomerado 18	9	0,01898734	0,21448588
10	Conglomerado 19	30	0,06329114	0,945271042

4. Ahora, esos números aleatorios los multiplicamos por el peso de los conglomerados.

	A	B	C	D	E
1	Conglomerados	tamaño	Peso	Números aleatorios	Peso * Números aleatorios
2	Conglomerado 11	53	0,11181435	0,228577761	0,025558273
3	Conglomerado 12	190	0,40084388	0,972140981	0,389676765
4	Conglomerado 13	0	0	0,138437623	0
5	Conglomerado 14	6	0,01265823	0,039222172	0,000496483
6	Conglomerado 15	116	0,24472574	0,402437909	0,098486914
7	Conglomerado 16	59	0,12447257	0,606563362	0,075500503
8	Conglomerado 17	11	0,02320675	0,671286449	0,015578378
9	Conglomerado 18	9	0,01898734	0,21448588	0,004072517
10	Conglomerado 19	30	0,06329114	0,945271042	0,059827281

5. Luego, los ordenamos, en este caso, de mayor a menor, para dar mayor probabilidad de ser elegidos para la muestra de los conglomerados de mayor peso.

	A	B	C	D	E
1	Conglomerados	tamaño	Peso	Números aleatorios	Peso * Números aleatorios
2	Conglomerado 12	190	0,40084388	0,972140981	0,389676765
3	Conglomerado 15	116	0,24472574	0,402437909	0,098486914
4	Conglomerado 16	59	0,12447257	0,606563362	0,075500503
5	Conglomerado 19	30	0,06329114	0,945271042	0,059827281
6	Conglomerado 11	53	0,11181435	0,228577761	0,025558273
7	Conglomerado 17	11	0,02320675	0,671286449	0,015578378
8	Conglomerado 18	9	0,01898734	0,21448588	0,004072517
9	Conglomerado 14	6	0,01265823	0,039222172	0,000496483
10	Conglomerado 13	0	0	0,138437623	0

Como queremos 4 conglomerados en nuestra muestra, elegimos los 4 primeros.

Normalmente, el número de conglomerados suele ser muy numeroso, *por lo que no es recomendable realizar en Excel la selección de muestra por conglomerados, considerando que existen otras alternativas*. El proceso sería equivalente a como se ha realizado la selección de muestra por estratos.

Práctica 3.

Muestreo por Conglomerados en dos etapas combinado con estratificación y cálculo de muestra



El archivo DatosGuatemala.xlsx contiene la siguiente información:



ZMV

Zona de medios de vida.



Departamento

Distintos departamentos del país.



Municipio

Distintos municipios de los departamentos.



Comunidad

Comunidades dentro de los municipios.

1.2.3

Categoría

Clasificación de las comunidades como: caserío, aldea, paraje, pueblo, etcétera.



Población

Población estimada de las comunidades.



A partir de dichos datos, se desea realizar un diseño muestral con el objetivo de calcular las siguientes proporciones:

1. Proporción de personas con consumo pobre, cuya prevalencia esperada es del 41%.
2. Proporción de personas en inseguridad alimentaria severa o moderada, cuya prevalencia esperada es 65%.
3. Proporción de personas con más del 75% de gastos en alimentos, cuya prevalencia esperada es del 51%.

Calcular el tamaño de muestra necesario a través de un diseño por conglomerados en dos etapas combinado con estratificación, donde los estratos son las ZMV y los conglomerados las comunidades. Luego, generar la muestra.

Resolución

En el cálculo del tamaño de muestra hay que considerar los tres indicadores que se desean determinar, los cuales tienen prevalencias distintas por lo que requerirán de un tamaño de muestra diferente.

Generamos el tamaño para los tres indicadores y seleccionamos el mayor tamaño de los tres. Para ello, podemos usar la herramienta Excel **“Cálculo del tamaño de muestra”**.

En este caso, como la población es muy grande, podemos hacer el cálculo como si fuera desconocida, pues el hecho de conocerla únicamente afecta en dicho cálculo para poblaciones pequeñas.

1. Para estimar la proporción de personas con consumo pobre, se requiere un tamaño de 587 personas, considerando una precisión del 5%, aplicando un efecto de diseño igual a 1.5 (por el muestreo por conglomerados) y teniendo en cuenta hasta un 5% como posible tasa de no respuesta. Nos quedamos en el tamaño de muestra correspondiente a un 95% de confianza.

CÁLCULO DE TAMAÑO DE MUESTRA CON POBLACIÓN DESCONOCIDA

Prevalencia	41	%
Precisión	5	%
Efecto del diseño	1,5	Entre 1 y 2
Tasa de no respuesta	5	%
Factor de corrección poblaciones pequeñas (< 10 000)	1	Menor que 1

	Tamaño de muestra
Nivel de confianza 95%	587
Nivel de confianza 90%	414
Nivel de confianza 99%	1014

2.

Para estimar la proporción de personas en inseguridad alimentaria severa o moderada, usamos la prevalencia esperada de 65%, para lo que se requiere un tamaño de muestra de 552, considerando los mismos valores que en la proporción anterior para el resto de parámetros.

CÁLCULO DE TAMAÑO DE MUESTRA CON POBLACIÓN DESCONOCIDA

Prevalencia	65	%
Precisión	5	%
Efecto del diseño	1,5	Entre 1 y 2
Tasa de no respuesta	5	%
Factor de corrección poblaciones pequeñas (< 10 000)	1	Menor que 1

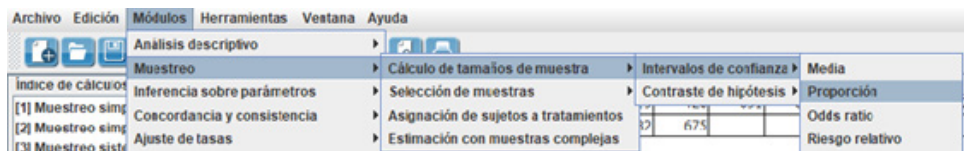
	Tamaño de muestra
Nivel de confianza 95%	552
Nivel de confianza 90%	389
Nivel de confianza 99%	953

3. Por último, para estimar la proporción de personas con más del 75% de gastos en alimentos, cuya prevalencia esperada es del 51%, se requiere un tamaño muestral de 607 personas.

CÁLCULO DE TAMAÑO DE MUESTRA CON POBLACIÓN DESCONOCIDA		
Prevalencia	51	%
Precisión	5	%
Efecto del diseño	1,5	Entre 1 y 2
Tasa de no respuesta	5	%
Factor de corrección poblaciones pequeñas (< 10 000)	1	Menor que 1
		Tamaño de muestra
Nivel de confianza 95%	607	
Nivel de confianza 90%	428	
Nivel de confianza 99%	1047	

Por lo que el tamaño muestral será, como hemos dicho, el mayor de los tres, 607 personas.

Para calcular esto mismo con **Epidat 4.2**, accedemos a: **“Módulos”** → **“Muestreo”** → **“Cálculo de tamaños de muestra”** → **“Intervalos de confianza”** → **“Proporción”**.



En la ventana que se abre, podemos indicar el valor del tamaño de la población *-si lo conocemos-* o bien, dejar el campo en blanco si lo desconocemos o si la población es muy grande, en cuyo caso su valor no va a afectar el cálculo del tamaño de la muestra.

Usamos el valor de la prevalencia más desfavorable de los tres, que es la del tercer indicador, cuya prevalencia esperada es del 51%. Dejamos el 95% como nivel de confianza e introducimos el valor 1,5 como efecto del diseño.

Epidat 4.2 también permite introducir dos valores distintos para la precisión, y así poder comparar los tamaños de muestra requeridos para uno y otro, lo cual puede ser influyente en la decisión en función los recursos de los que se disponga.

Proporción

Datos

Tamaño de la población:

Proporción esperada: %

Nivel de confianza: %

Calcular

Tamaño de la muestra

Precisión absoluta

Precisión absoluta (%)

Mínimo:

Máximo:

Incremento:

Efecto de diseño:

Ocultar Calcular Limpiar Cerrar

Pulsamos **“Calcular”** y obtenemos un tamaño de muestra de 576.

Datos:

Tamaño de la población: ∞
Proporción esperada: 51,000%
Nivel de confianza: 95,0%
Efecto de diseño: 1,5

Resultados:

Precisión (%)	Tamaño de la muestra
5,000	576

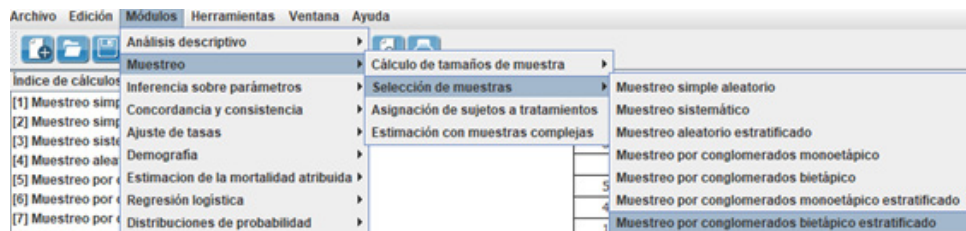
Nótese que aquí no se ha considerado la tasa por la falta de respuesta. Si incrementamos dicho valor en un 5%, obtenemos un tamaño de muestra de 605, el cual difiere en dos unidades del anterior, debido a los redondeos decimales, pero dicha diferencia no resulta significativa.

Para el cálculo del número de conglomerados, si hacemos uso de la herramienta Excel **“Cálculo del tamaño de muestra”**, nos dirá que es necesario únicamente un conglomerado. Esto ocurre, porque los conglomerados tienen poblaciones muy pequeñas y la prevalencia es muy grande; pero si seleccionáramos únicamente un conglomerado, la muestra no sería en absoluto representativa.

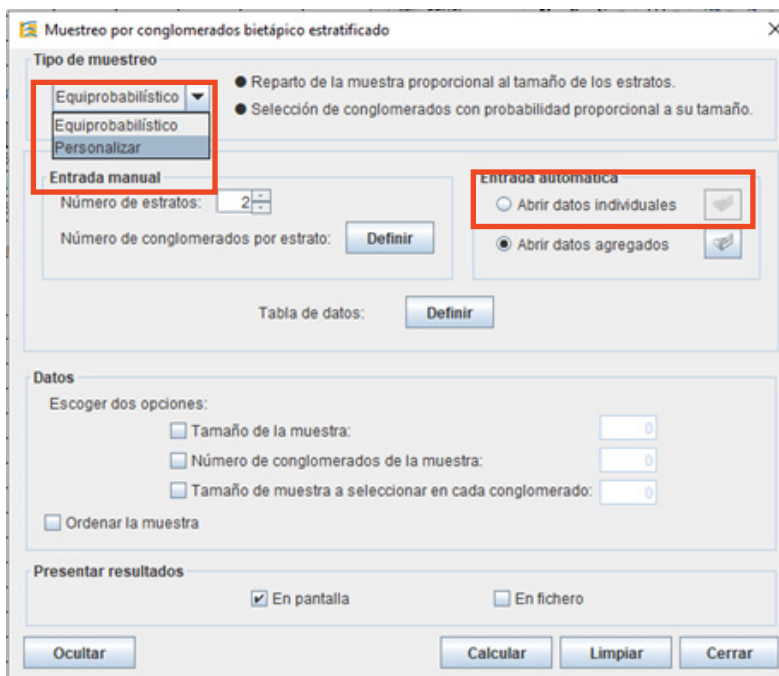
Cuando esto ocurra, hay otros mecanismos para seleccionar el número de conglomerados, como calcularlo mediante la fórmula del muestreo aleatorio simple, para lo cual tendríamos que conocer las prevalencias en los distintos conglomerados, o si el número de conglomerados en la población es muy grande, en nuestro caso 529, podemos elegir un valor alto de conglomerados. Como el número habitual suele ser entre 25 y 30, podemos quedarnos con 30 conglomerados.

A continuación, generamos la muestra, para ello:

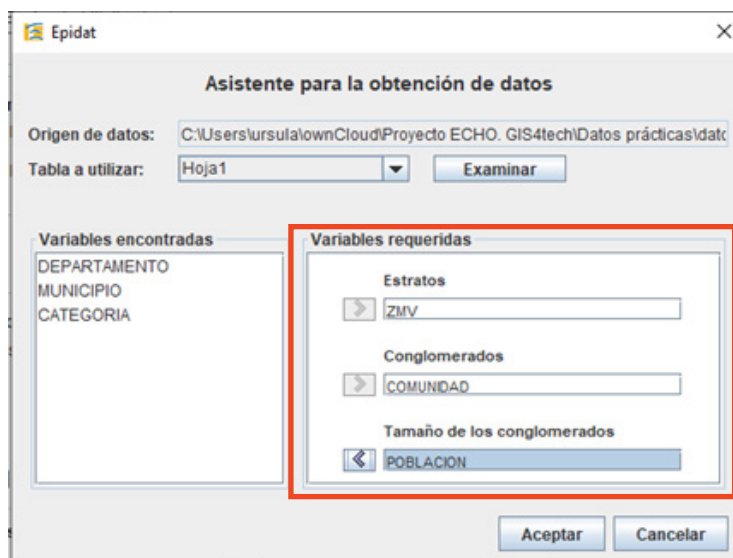
1. Accedemos desde **Epidat** a: **“Módulos”** → **“Muestreo”** → **“Selección de muestras”** → **“Muestreo por conglomerados bietápico estratificado”**.



2. En Tipo de muestreo indicamos **“Personalizar”** y seleccionamos **“Abrir datos agregados”**. La diferencia entre datos individuales y datos agregados, es que en estos últimos podemos establecer la ponderación, cuando la población se conoce.



3. Indicamos la ubicación de los datos y seleccionamos como Estratos las **“ZMV”**, como Conglomerados, las **“Comunidades”** y como Tamaño de los conglomerados, la **“Población”**.



4. Una vez cargados los datos, seleccionamos cómo queremos que sea el reparto de la muestra por comunidades. Las opciones disponibles son las que aparecen en el desplegable:

	Muestra de individuos
GT05	15
LOS CHIVITOS	819

Nótese que, entre las opciones, no está la de reparto proporcional al tamaño de los conglomerados. Únicamente podríamos definir un tamaño muestral específico para cada uno de los conglomerados de la población, en el supuesto de que fueran elegidos para la muestra, pero en poblaciones como ésta, con 529 conglomerados, resultaría una labor imposible.

La razón de no darnos la opción de reparto proporcional al tamaño, es que, de hacerlo así, podrían resultar algunos conglomerados con un alto porcentaje de la muestra mientras que otros tendrían sólo unas pocas personas, por lo que el costo de visitar esos conglomerados para entrevistar pocas personas no sería rentable, por lo que muchos métodos apuestan por un reparto equitativo de la muestra entre los conglomerados, para aprovechar de este modo todas las visitas.

Otra de las opciones es: *reparto equitativo de la muestra para los conglomerados dentro de un mismo estrato o una fracción fija de la población de cada conglomerado, en cuyo caso, se desconocería el tamaño final de la muestra resultante.*

Seleccionamos **“Muestra igual para todos los conglomerados”**.

5. A continuación, indicamos el número de conglomerados en la muestra, pulsando en **“Definir”** el número de conglomerados.

The dialog box 'Muestreo por conglomerados bietápico estratificado' contains the following elements:

- Tipo de muestreo:** Personalizar
- Entrada manual:** Número de estratos (input field), Número de conglomerados por estrato (input field with 'Ver' button)
- Entrada automática:** Radio buttons for 'Abrir datos individuales' and 'Abrir datos agregados' (selected)
- Tamaño de muestra por conglomerado:** Muestra igual para todos los conglomerados
- Table:**

Estrato	Conglomerados	Tamaño del conglomerado	Muestra de individuos
GT05	CARACOL	1115	0
GT05	CHIPAJ	1173	0
GT05	LA LAGUNA DANTA	382	0
GT05	LAS PACAYAS	1734	0
GT05	LOS CHIVITOS	819	0
- Datos:** Tamaño de la muestra de cada conglomerado (input field), Cargar button, Número de conglomerados en la muestra (input field with 'Definir' button highlighted in red), Método de selección de conglomerado: Probabilidades proporcionales al tamaño, Ordenar la muestra checkbox
- Presentar resultados:** En pantalla checkbox (checked), En fichero checkbox
- Buttons:** Ocultar, Calcular, Limpiar, Cerrar

6. En reparto de muestra, seleccionamos **“Reparto proporcional al tamaño de los estratos”**. En número total de conglomerados en la muestra indicamos 30 y pulsamos en **“Distribuir”**.

The dialog box 'Tamaño de muestra' contains the following elements:

- Reparto de la muestra:** Reparto proporcional al tamaño de los estratos (dropdown menu highlighted in red)
- Número total de conglomerados en la muestra:** 30 (input field), Distribuir button (highlighted in red)
- Table:**

Estrato	Nº de conglomerados	Tamaño del estrato	Muestra de conglomerad...
GT05	9	8654	1
GT06	429	432334	24
GT07	16	16864	1
GT10	34	35796	2
GT11	43	27573	2
- Buttons:** Aceptar, Cancelar

De esta forma, el programa reparte la muestra de 30 conglomerados en los distintos estratos en función de su tamaño.

7. Para introducir el valor del “tamaño de la muestra de cada conglomerado”, tenemos que dividir el tamaño de muestra final, 607, entre los 30 conglomerados y redondeamos al entero superior. En este caso, redondeamos a 21 y pulsamos **“Cargar”**.

Muestreo por conglomerados bietápico estratificado

Tipo de muestreo
Personalizar

Entrada manual
Número de estratos: 5
Número de conglomerados por estrato: Ver

Entrada automática
 Abrir datos individuales
 Abrir datos agregados

Tamaño de muestra por conglomerado:
Muestra igual para todos los conglomerados

Estrato	Conglomerados	Tamaño del conglomerado	Muestra de individuos
GT05	CARACOL	1115	21
GT05	CHIPAJ	1173	21
GT05	LA LAGUNA DANTA	382	21
GT05	LAS PACAYAS	1734	21
GT05	LOS CHIVITOS	819	21

Datos
Tamaño de la muestra de cada conglomerado: 21 Cargar
Número de conglomerados en la muestra: Definir
Método de selección de conglomerado: Probabilidades proporcionales al tamaño
 Ordenar la muestra

Presentar resultados
 En pantalla En fichero

Ocultar Calcular Limpiar Cerrar

En el método de selección de conglomerados, indicamos **“Probabilidades proporcionales al tamaño”**.

8. Le damos a **“Calcular”** y mostrar en pantalla. El resultado sería (por la extensión de la salida se muestra sólo el resultado para el primer estrato):

Datos:

Tipo de muestreo: Personalizar
 Metodo de selección de conglomerados: Probabilidades proporcionales al tamaño

Estrato	Tamaño		Nº de conglomerados	
	Población	Muestra	Población	Muestra
GT05	8654	21	9	1
GT06	432334	504	429	24
GT07	16864	21	16	1
GT10	35796	42	34	2
GT11	27573	42	43	2
Total:	521221	630	531	30

Conglomerados seleccionados para el estrato GT05:

Conglomerado	Tamaño	Muestra	Probabilidad	PONDERA...
RIO ERMIN	415	21	0,2427	412,0952
Total:	415	21		

Sujetos seleccionados para cada conglomerado:

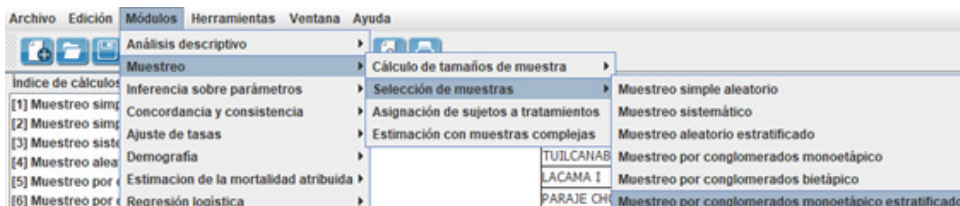
Conglomerado RIO ERMIN:

209	306	245	104	135	343	119
206	253	95	313	378	312	111
335	77	89	151	8	371	283

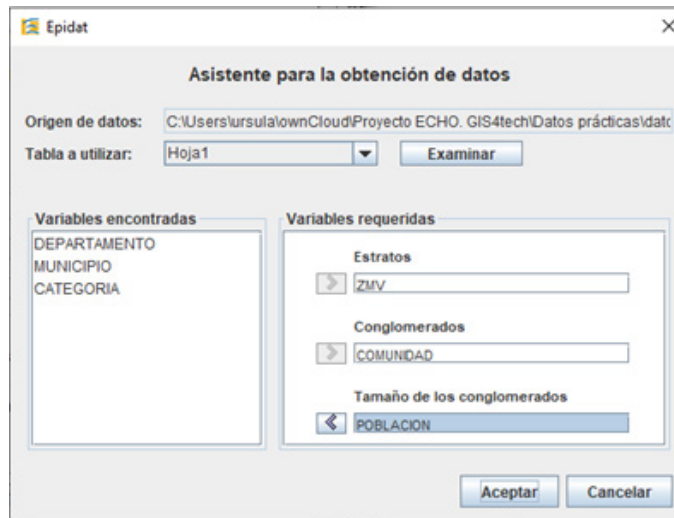
Vemos que nos muestra para cada estrato el o los conglomerados seleccionados y, dentro de esos conglomerados, nos muestra los individuos seleccionados. A pesar de no tener el listado de individuos, deberíamos seleccionar los indicados en el orden del listado, para que fuera una selección aleatoria.

Una forma alternativa de hacer este diseño es:

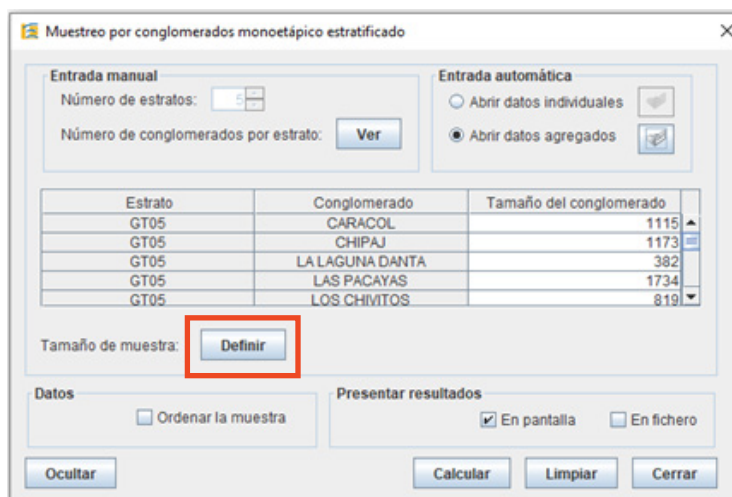
1. Acceder a **“Módulos”** → **Muestreo** → **“Selección de muestras”** → **Muestreo por conglomerados monoetápico estratificado”**.



2. Abrimos los datos agregados con las siguientes especificaciones:



3. Pulsamos en “Definir” tamaño de muestra:



Primero, distribuimos los individuos por estrato, con **“reparto proporcional al tamaño de los estratos”**, para un tamaño de muestra de 607 personas.

The dialog box 'Tamaño de muestra' has the following settings:

- Tamaño de la muestra que se define: Individuos
- Reparto de la muestra: Reparto proporcional al tamaño de los estratos
- Tamaño de la muestra: 607
- Buttons: Distribuir (highlighted), Aceptar, Cancelar

Estrato	Nº de conglomerados	Tamaño del estrato	Muestra de individuos
GT05	9	8654	11
GT06	429	432334	490
GT07	16	16884	18
GT10	34	35796	39
GT11	43	27573	49

Pulsamos en **“Distribuir”**, y así tenemos el número de personas de la muestra por estrato.

Anotamos los resultados y, dentro de esta misma ventana, distribuimos los conglomerados, ahora por estrato:

The dialog box 'Tamaño de muestra' has the following settings:

- Tamaño de la muestra que se define: Individuos, Conglomerados
- Reparto de la muestra: Reparto proporcional al tamaño de los estratos
- Tamaño de la muestra: 30
- Buttons: Distribuir (highlighted), Aceptar, Cancelar

Estrato	Nº de conglomerados	Tamaño del estrato	Muestra de conglomerados
GT05	9	8654	1
GT06	429	432334	24
GT07	16	16884	1
GT10	34	35796	2
GT11	43	27573	2

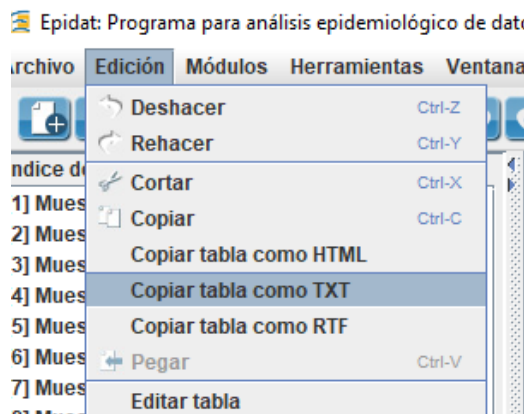
4. Le damos a **“Calcular”** en la ventana principal, así obtenemos la salida con los conglomerados seleccionados por estrato junto al tamaño del conglomerado.

Estrato GT06:

Conglomerados	Tamaño
PARAJE CHICH...	314
EL CHICHICASTE	251
PARAJE PAXAN...	1626
PARAJE CHOSU...	1074
NUEVA CONCE...	777
PARAJE PATUN...	1131
PARAJE XELQ...	474
PARAJE POLOG...	1267
PARAJE CHIKS...	869
PARAJE CENTR...	1374
PARAJE PACOX...	286
PARAJE PAXMA...	693
PARAJE BELLA ...	717
PARAJE SICALB...	1512
PARAJE PARRA...	460
PARAJE RANCH...	1033
PARAJE CENTR...	883
ALDEA RANCHO	2200
4A CALLE 1-94...	914
PARAJE PASAQ...	2638
BACU	549
ALDEA XECACH...	569
CHANCHOC	201
LOS DURAZNA...	2386
TOTAL	24198

En este caso, sólo nos muestra el resultado por pantalla y no nos genera el fichero Excel, debido al gran número de conglomerados.

4. Una opción, es seleccionar la tabla y, en **“Edición”** seleccionamos: **“Copiar como txt”**.



De esta forma, podemos llevar todas las tablas a Excel, para conservar la muestra, o bien para calcular los pesos. Una vez que las tenemos, podemos multiplicarlos por el tamaño de muestra calculado para ese estrato, si queremos repartir la muestra de forma proporcional al tamaño de estratos y conglomerados.






Otra opción para esto último, sería ir copiando los valores calculados en la Herramienta Excel **“Cálculo del tamaño de muestra”**, que ya tiene las fórmulas implementadas.

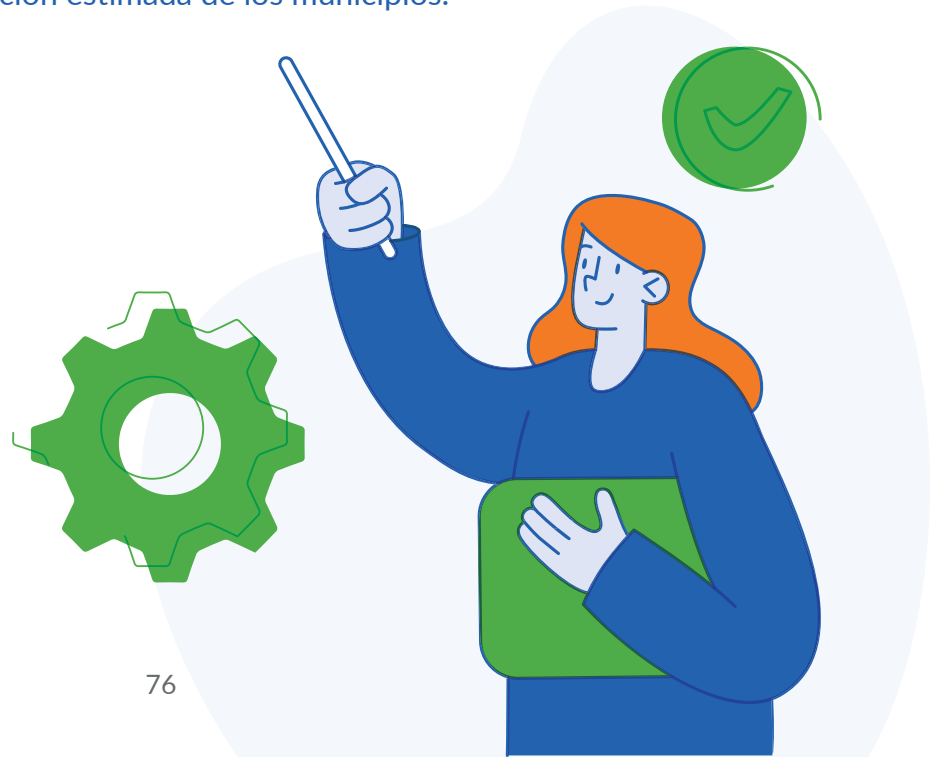


Trabajos Propuestos



El archivo DatosPaíses.xlsx contiene la siguiente información:

 Id	Identificación de fila.
 País	El Salvador, Guatemala, Honduras y Nicaragua.
 Departamento	Departamentos de los países.
 Municipio	Municipios.
 Población total	Población estimada de los municipios.



Trabajo Propuesto 1

Generar una muestra de municipios de tamaño 100, mediante muestreo aleatorio simple y sistemático. Hacerlo a través de Excel y de Epidat 4.2.

Trabajo Propuesto 2

Generar una muestra por estratos de 100 municipios, considerando como estratos cada uno de los países. Usar afijación proporcional según el tamaño de los estratos. Hacerlo a través de Excel y de Epidat 4.2.

Trabajo Propuesto 3

Se desea realizar un diseño muestral con el objetivo de calcular las siguientes proporciones:

1. Proporción de personas con consumo pobre, cuya prevalencia esperada es del 38%.
2. Proporción de personas en inseguridad alimentaria severa o moderada, cuya prevalencia esperada es 50%.
3. Proporción de personas con más del 75% de gastos en alimentos, cuya prevalencia esperada es del 54%.

Calcular el tamaño de muestra necesario, a través de un diseño por conglomerados en dos etapas, combinado con estratificación, donde los estratos son los países y los conglomerados los municipios.

Generar los conglomerados de la muestra con probabilidad proporcional a su tamaño y distribuidos por estrato, según afijación proporcional.

Repartir la muestra entre los conglomerados, indicando qué tamaño de muestra asignaría a cada conglomerado y por qué.

BIBLIOGRAFÍA



1. AA.VV. (2004). Thematic Guidelines: Sampling Guidelines for Vulnerability Analysis. Chapter 4. Household-level data in a CFSVA.
2. Adam, M. et al. A, (2009). Comparison between Convenience Sampling versus Systematic Sampling in Getting the True Parameter in a Population. The Audit Diabetes Control Management.
3. AA.VV. (2012). Sampling Methods and Sample Size Calculation for the SMART Methodology.
4. Cuesta, M. y Herrero, F.J. (2013). Introducción al Muestreo. Universidad de Oviedo, España.
5. Heckathorn, D.D. (2002). Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. Social Problems.
6. Mayor Gallego, J.A. (2014). Muestreo en Poblaciones Finitas. Diseños Muestrales Complejos. Universidad de Sevilla.
7. AA.VV. (2015). Consolidated Approach to Reporting Indicators of Food Security (CARI).
8. Etikan, I. Abubakar Musa, S. y Sunusi Alkassim R. (2016). Comparison of Convenience Sampling and Purposive Sampling. American Journal of Theoretical and Applied Statistics.
9. Medina Disla, H. (2009). Apuntes sobre Muestreo.
10. Morillas, A. (2007). Muestreo en Poblaciones Finitas.

11. Nassiuma, D.K. (2001). Survey sampling: Theory and methods. Nairobi University Press.
12. Marvasti, F. (2012). Nonuniform sampling: theory and practice. Springer.
13. AA.VV. (2017). Mitos y Verdades del Efecto del Diseño (DEFF) en Estudios por Muestreo Mrobabilístico. Unidad Investigación Usta.
14. Otzen, T., Manterola, C. (2017). Técnicas de Muestreo sobre una Población Objeto de Estudio. International Journal of Morphology.
15. Reichel, D. y Morales, L. (2017). Surveying Immigrants Without Sampling Frames – Evaluating the Success of Alternative Field Methods. Springer.
16. Rueda García, M.M. (2003). Muestreo I. Teoría, Problemas y Prácticas de Ordenador. Plácido Cuadros.
17. Rueda García, M.M. y Arcos Cebrián A. (1999). Problemas de Muestreo en Poblaciones Finitas. Editorial GEU.
18. Tamayo, G. (2001). Diseños muestrales en la investigación. Universidad de Medellín, Colombia.
19. Waldo, E. (2002). Muestreo Doble. Universidad Nacional de la Plata

