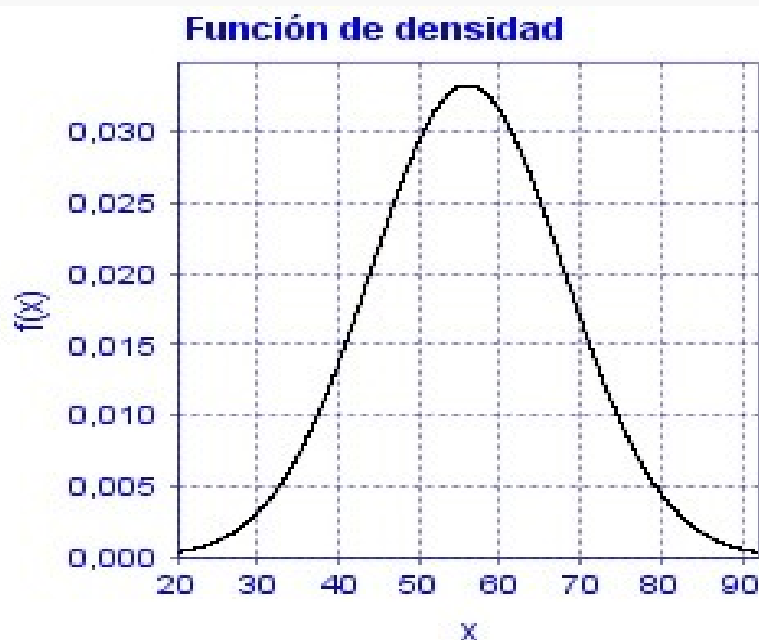


Muestreo en estudios descriptivos polietápicos. Epidat 4.0



Muestreo en estudios descriptivos



Muestreo en estudios descriptivos polietápicos. Epidat 4.0

León Darío Bello Parías

Primera edición

<http://ciemonline.info/blog/curso-muestreo>

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Page 1 of 115

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Aviso Legal

Este texto no puede ser distribuido por ningún medio impreso, electrónico o cualquier otra forma sin permiso escrito del autor. Para adquirir este texto digital, requiere realizar el curso online sobre el tema o visitar uno de los sitios que le menciono a continuación <http://www.leondariobello.com/ebook-estadistica/> o www.ciemonline.info/blog/libros-electronicos-estadistica, allí encontrará la manera de realizar los pedidos, además, diversa información sobre temas de estadística y de mercadeo on line.

En caso de requerir alguna consulta puntual sobre los temas expuestos en el ebook, pueden expresarla escribiendo al correo ldbello753@gmail.com o solicitar una charla on line en mi sala de conferencias a la cual podrá acceder en la dirección www.gvowebcast.com/conference/leondario.

Con la dedicación del caso, confío que las páginas que siguen le servirán para mejorar sus habilidades y conocimientos en la parte de muestreo en estudios descriptivos y específicamente en las llamadas muestras complejas, utilizando el Epidat 4.0, el cual provee la opción de estimar parámetros en este tipo de estudios.

El texto tiene además, el apoyo de videos complementarios en todos los capítulos, pruebe observando uno de ellos, seleccione el de [MUESTREO SISTEMATICO](#).

Adelante y muchos éxitos.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



León Darío Bello Parías

<http://www.leondariobello.com/ebook-estadistica/>

www.ciemonline.info/blog/libros-electronicos-estadistica

ldbello753@gmail.com

CONTENIDO

Página

CAPÍTULO 1

CONCEPTOS BASICOS

1.1 CONCEPTOS GENERALES.	11
1.1.1 Población o universo	13
1.1.2 Censo	13
1.1.3 Muestra	14
1.1.4 Muestreo	14
1.1.5 Marco muestral	14
1.1.6 Unidad muestral	16
1.1.7 Unidad de análisis	16
1.1.8 Parámetro	17
1.1.9 Estadístico	17
1.1.10 Inferencia estadística	17
1.1.11 Muestra representativa	18

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

1.1.12 Intervalo de confianza (IC)	19
1.1.13 Nivel de Confianza	21
1.1.14 Error absoluto (e) y relativo (e_r)	22
2. EL ERROR DE MUESTREO RELATIVO O COEFICIENTE DE VARIACION.	25
3. ERRORES NO MUESTRALES	28
4. PROCEDIMIENTOS DE MUESTREO	29
5. PASOS PARA SELECCIONAR UNA MUESTRA	30
6. MUESTREOS PROBABILISTICOS	34

CAPÍTULO 2

MUESTREO ALEATORIO SIMPLE (MAS)

2.1 TAMAÑOS DE MUESTRA	38
2.1.1 Tamaño de muestra para estimar un promedio poblacional, utilizando el Epidat 4.0.	46
2.1.2 Tamaño de muestra para estimar una proporción poblacional, utilizando el Epidat 4.0	50
2.1.3 METODOS DE SELECCIÓN CON EPIDAT 4.0.	54

CAPÍTULO 3

MUESTREO SISTEMÁTICO(MS)	61
--------------------------	----

CAPÍTULO 4

MUESTREO ALEATORIO ESTRATIFICADO (MAE)	69
--	----

CAPÍTULO 5

MUESTREO POR CONGLOMERADOS (MC)	83
---------------------------------	----

5.1 Muestreo por Conglomerados una Sola Etapa o Monoetápico.	84
--	----

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

5.2 Muestreo por Conglomerados Bietápico.

CAPÍTULO 6

MUESTREO POR CONGLOMERADOS MONOETÁPICO Y BIETÁPICO ESTRATIFICADO.

6.1 Muestreo por conglomerados monoetápico estratificado. 96

6.2 Muestreo por conglomerados bietápico estratificado. 98

CAPÍTULO 7

ESTIMACIÓN CON MUESTRAS COMPLEJAS. 104

7.1 Probabilidad de selección y ponderación. 105

7.2 Posibilidades con Epidat 4.0. 108

7.3 Procedimiento de estimación de parámetros. 114

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

SÍNTESIS DE LOS CAPITULOS.

Capítulo 1. CONCEPTOS BASICOS: Se describen los conceptos fundamentales de una manera práctica y haciendo énfasis en las diferencias claves, valga decir, error absoluto y error relativo, nivel de confianza e intervalo de confianza, el tema de muestra representativa es fundamental para tener en cuenta en los diseños muestrales. Se abordan estos conceptos desde las vivencias del autor y por ende, son útiles desde lo práctico. Se sugieren unos pasos que si bien no pueden ser siempre utilizados, si dan una pauta o guía para las personas que inician en esta hermosa labor de diseñar muestras probabilísticas. Como complemento se tiene el video [PLAN DE MUESTREO](#)

Capítulo 2. MUESTREO ALEATORIO SIMPLE: Aparte de comentar sobre los casos donde es pertinente éste muestreo, se desarrollan ejemplos utilizando el Epidat 4.0. Se calcula el tamaño de muestra tanto para promedios como para proporciones, sin descuidar la parte reflexiva a tener en cuenta al definir los valores involucrados en las fórmulas estadísticas. Para colocar un ejemplo, se argumenta, por qué, no es recomendable utilizar siempre el valor de $p=0,50$ en la mayoría de casos, siendo lo contrario de lo que sugieren la mayoría de textos. La diferencia del ebook, de un gran número de textos es la manera de abordar el tema desde el análisis del problema, no desde lo operativo. Como todos los capítulos tiene su video complementario.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 3. MUESTREO SISTEMÁTICO: Apoyados en el Epidat 4.0, se presentan los dos casos más usuales en la aplicación de éste procedimiento, el regular y el más usual el sistemático en fases.

Capítulo 4. MUESTREO ESTRATIFICADO: El software permite una variada cantidad de posibilidades, desde diseños equiprobabilísticos hasta los definidos por el usuario, para darle mayor información, le sugiero mirar el video 2, [MUESTREO ESTRATIFICADO 2](#), de éste capítulo, el 1, lo podrá observar entrando a mi canal de Youtube.com en el enlace Videos [León Dario Bello](#)

Capítulo 5. MUESTREO POR CONGLOMERADOS: Se muestran de manera práctica los de una Sola y dos Etapas, mostrando las bondades y/o dificultades de cada uno de ellos, se analiza en detalle, el término DEFF o Efecto del diseño (design effect) y su impacto en el tamaño de muestra, sin dejar de reconocer, que es definitivamente muy complicado conocerlo antes de realizar el estudio.

Capítulo 6. MUESTREO POR CONGLOMERADOS MONOETÁPICO Y BIETÁPICO ESTRATIFICADO: El que cierra el texto, es el que nos podrá servir de guía para cuando nos toque aplicar uno de más de dos etapas que es lo que nos provee el software que se viene utilizando, el Epidat 4.0. Permite una cantidad de posibilidades, como ya se mencionaron en diseños anteriores, como: muestras iguales para cada estrato, para cada conglomerado, proporcionales al tamaño y combinaciones diversas, el tema se aborda desde un ejemplo, que se puede realizar usando el archivo Excel utilizado en el texto.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 7. ESTIMACIÓN CON MUESTRAS COMPLEJAS: Se destaca el hecho de diferenciar entre estimaciones usando MAS y utilizando muestras complejas, es importante hacer énfasis en el tema de que cuando se utilizan muestreos en varias etapas, las formulas para calcular los estimadores usuales de totales, promedios y proporciones, cambian, es usual trabajar con los diseñados por Horvitz-Thompson (totales) y Hajek (medias, proporciones). Sin embargo, apoyados de nuevo en el programa estadístico, se minimiza el uso de las formulas, para ello, se recomienda tener a la mano las ayudas de muestreo en el Epidat 4.0, que transcribe las formulas utilizadas en la aplicación del paquete y que de hecho, son las utilizadas en éste ebook.

Continue avanzando en su formación profesional, eso le permitirá optimizar su tiempo y eficiencia en el diseño de muestras probabilísticas.



León Darío Bello Parías
Docente Universidad de Antioquia
Facultad Nacional de Salud Pública

www.leondariobello.com/e-books-estadistica
www.ciemonline.info/blog/libros-electronicos-estadistica

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

CAPITULO 1



A MANERA DE INTRODUCCIÓN

La efectividad del método de muestreo y por ende la validez de los resultados depende en grado sumo de la imaginación del investigador.

LEÓN DARÍO BELLO PARIA

El muestreo es una técnica estadística que orienta no sólo el definir el tamaño muestral y el método de selección de las unidades de análisis, sino que orienta sobre la manera adecuada de estimar los parámetros requeridos según el tipo de muestreo probabilístico utilizado. Es tal vez el punto más crucial en el proceso investigativo, ya que dependiendo de éste, será

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Page 9 of 115

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

posible generar inferencias o generalizaciones a la población objeto de estudio.

Es evidente que el reducir costos en todas las actividades que se realizan en la cotidianidad sin detrimento de la calidad es una prioridad, la investigación no es la excepción, máxime cuando por regla general no se tiene un presupuesto boyante en las investigaciones a realizar. Por lo tanto, en la práctica se requiere analizar una parte de la población y con esos resultados lograr inferir a toda la población, claro está, si se realiza un muestreo probabilístico que son los que nos van a ocupar en este ebook.

En el mercado se puede acceder a diversos software en la materia, sin embargo, acá mostraremos los procedimientos utilizando el programa Epidat 4.0, el cual provee procedimientos de uso cotidiano con una versatilidad en el manejo que lo hace supremamente útil para apoyar la selección de unidades de análisis y su posterior estimación de parámetros.

El Epidat 4.0 se puede descargar de manera gratuita de la dirección <http://www.sergas.es/EPIWB/SolicitudEpidat.aspx?IdPaxina=62714&idv=2&lng=es>. Se requiere llenar un formulario para su descarga, definitivamente le recomendamos hacerlo, el software le será de mucha utilidad.

Este aporte ha sido diseñado e impulsado por dos instituciones, **XUNTA DE**

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

GALICIA – Consellería de Sanidade – Dirección Xeral de Saúde Pública y la Organización Panamericana de la Salud (OPS/OMS). Como se verá en el módulo de muestreo es un software de fácil manejo y con ayudas bien elaboradas.

1.1 CONCEPTOS GENERALES.

La determinación del tamaño de muestra es una tarea que se realiza durante la fase de diseño de un estudio para establecer el número de elementos o unidades de análisis que deben seleccionarse para participar en el mismo. El abanico de criterios susceptibles de ser considerados para efectuar este cálculo es amplio y variado, y la elección concreta que se haga depende, entre otras cosas, de la información que se disponga y de las características del procedimiento estadístico que se va a utilizar para analizar los datos. Es claro que el procedimiento de selección y el tamaño está orientado por los objetivos de la investigación, valga decir, una cosa es si se quiere estimar un parámetro poblacional y otra si el objetivo es determinar diferencias significativas entre tratamientos. Esto influye en la definición de las formulas estadísticas a utilizar, para éste ebook, se analizará el caso de estimación de parámetros, por lo tanto, se usarán las formulas basadas en los intervalos de confianza, no en las de pruebas de hipótesis cuyo objetivo investigativo es otro.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Además, las poblaciones son en lo general diferentes y por lo tanto, deberán utilizarse criterios distintos para seleccionar las unidades bajo estudio.

Un aporte importante de este software es que incorpora procedimientos de estimación de parámetros según el diseño de muestra utilizado en la selección de las unidades de análisis.

La siguiente figura presenta algunos conceptos claves de los cuales se debe tener claridad para lograr minimizar errores al diseñar un muestreo probabilístico.



Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

1.1.1 Población o universo: hace referencia a todas las unidades objeto de estudio, puede ser finita o infinita, además, no tiene que ser necesariamente personas.

- Todas las amas de casa del Municipio de Andes.
- Todos los profesionales de marketing en el país.
- Los sacos de café producidos en el año 2010.
- El número de usuarios de un refresco energizante.
- Las personas con el hábito de beber licor.
- Los estudiantes de la Universidad de Antioquia – Colombia-

1.1.2 Censo: recuento de toda la población disponible bajo estudio. Es claro que no siempre se puede o debe realizar, debido a diversas causas, entre ellas, los altos costos, la falta de oportunidad de la información, el incremento del error humano al tener que requerir de un mayor número de recolectores de datos y posterior digitación. Se sugiere realizar censo

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

cuando la población no es muy grande y a su vez es muy variable, además, de que se pueda recoger la información.

1.1.3 Muestra: subconjunto de la población estadística, es decir, es una parte de ella y por lo tanto tiene que poseer las mismas características de la población objeto de estudio. Se dice que es representativa, si representa adecuadamente la composición de la población, es decir, que las diferentes categorías claves del estudio estén representadas en la muestra. Es muy poco probable lograr representatividad de las variables objeto de estudio, debido, a que precisamente es de ellas las que se está identificando su comportamiento.

1.1.4 Muestreo: procedimiento que orienta el proceso de recolección de datos, orienta; cómo, cuándo y dónde se obtienen los datos, además, cuánto es el número adecuado a seleccionar para luego ser procesados y generalizados a toda la población. Claro que existen procedimientos que no permiten generalizar resultados, eso lleva a decir, que se tienen procedimientos probabilísticos o no probabilísticos, el enfoque de éste texto digital, es para los diseños probabilísticos.

1.1.5 Marco muestral: conjunto de listas de todas las unidades de muestreo, estas listas, también pueden ser mapas u otras formas que sirvan

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

de base para determinar las unidades de muestreo.

Selección de niños escolarizados:

1. Lista de escuelas.
2. Lista de grupos.
3. Lista de niños dentro de cada uno de los grupos.

En éste caso, las listas son el marco muestral total, mientras que sus componentes son los elementos muestrales, que en algunas oportunidades son las mismas unidades de muestreo.

Otro caso muy usual en muestreos polietápicos es:

1. Lista de Zonas geográficas en un país.
2. Lista de municipios o provincias de esas zonas.
3. Lista de manzanas de esos municipios.
4. Lista de viviendas de esas manzanas.
5. Lista de personas de esas viviendas.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

El número de casos es innumerable y depende de cada caso en particular. En la práctica se presentan dos situaciones con relación al marco muestral: a) se dispone de todo el marco al comienzo del estudio y b) el marco se va conformando en la medida que el muestreo avanza.

1.1.6 Unidad muestral: está compuesta por elementos, tal es el caso de una cajetilla de cigarrillos, la unidad es la cajetilla, pero lo que realmente se mide es el cigarrillo, que toma el nombre de **elemento muestral** o unidad de análisis. Sin embargo, en otros casos, la unidad de análisis puede ser incluso algo más pequeña, como por ejemplo, en salud, los médicos realizan intervenciones quirúrgicas a personas, pero su análisis se hará sobre el corazón o el hígado, esto último se conoce como unidad de análisis.

1.1.7 Unidad de análisis: Si bien puede tener varias connotaciones, para nosotros, se entenderá como aquella unidad de donde se obtiene la medición y por lo tanto, será a quien aplique las conclusiones del estudio.

Otros conceptos originados en la siguiente diapositiva, se explicitarán a continuación.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



1.1.8 Parámetro: Es un número que recoge o sintetiza información de una población, por ejemplo, el promedio de los ingresos de todos los asesores de mercadeo, la proporción de mujeres que compran con tarjeta de crédito en Guanajuato.

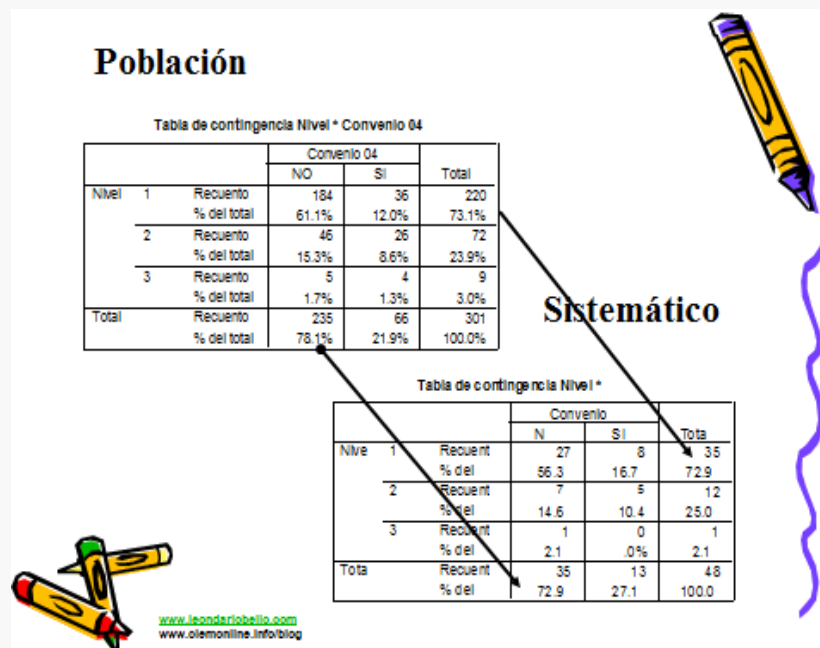
1.1.9 Estadístico: Es un número que recoge o sintetiza información de una muestra, por ejemplo, el promedio de los ingresos de 50 asesores de mercadeo, seleccionados del total de ellos, la proporción de mujeres que compran con tarjeta de crédito de una muestra de 100 en Guanajuato.

1.1.10 Inferencia estadística: Rama de la estadística que provee procedimientos y métodos para generalizar resultados obtenidos en una

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

muestra a toda la población, puede ser paramétrica o no paramétrica.

1.1.11 Muestra representativa: tener las mismas características de la población, y específicamente que su composición interna tenga un grado de variabilidad similar. Esto es viable lograrse en algunas categorías de identificación, sin embargo, en lo general no se sabría cómo se comportan las variables objeto de estudio, luego el tema de muestra representativa hay que entenderlo para variables genéricas, como: sexo, estrato, grupos de edad etc. Sin embargo, para las variables objeto de estudio, es difícil lograrlo, precisamente el estudio pretende identificar ese comportamiento.



Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

La diapositiva anterior, se utilizó en la presentación de un trabajo realizado en la Facultad Nacional de Salud Pública de la Universidad de Antioquia para el Ministerio de Protección Social de Colombia, se pretendía mirar el impacto de un programa llevado a cabo por el Ministerio en cuanto a la inversión de unos dineros por parte de algunas instituciones de salud, llamadas IPS (Instituciones Prestadoras de Servicios de Salud), con la cual se tenían convenios, de ahí, la clasificación de instituciones con o sin convenio, la otra clasificación hace referencia a los niveles de las instituciones, donde dependiendo del nivel se prestan servicios de menor o mayor complejidad. En la tabla colocada en la parte superior, se tiene el comportamiento de toda la población objeto de estudio, 301 instituciones, de las cuales, el 78,1% corresponden a instituciones sin convenio con el ministerio y el resto si lo poseían a la fecha del estudio. También se observa que el 73% de las instituciones en toda la población corresponden al nivel 1, mientras que el 23,9% al nivel 2 y el resto para el 3. En la tabla inferior, se muestra la distribución de la muestra usando el muestreo sistemático usando el Epidat. La representatividad de la muestra según estas variables clasificatorias de las IPS, se observa muy similar a la población objeto de estudio, lo que lleva a concluir, que la muestra está bien representada según esas variables.

1.1.12 Intervalo de confianza (IC): teniendo en cuenta que en muchos

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

casos el objetivo del estudio es estimar parámetros, es necesario para calcular el tamaño de muestra tener en cuenta la amplitud máxima admisible del intervalo, donde uno de los criterios, debe ser, que esa amplitud sirva realmente para aportar a la toma de decisiones, no tiene sentido, entregar un IC en las siguientes condiciones: “con una confiabilidad del 95% se puede afirmar que el verdadero promedio de los salarios de los profesionales de la salud, se encuentra entre 2.5 y 15 salarios mínimos legales vigentes”. Es evidente que esa información ya se tenía sin necesidad de realizar el estudio.

Por lo tanto, la sugerencia para definir el tamaño de muestra cuando de estimación de parámetros se trata, es la siguiente: teniendo en cuenta los recursos disponibles, definir un error que me permita encontrar un IC útil y lo más preciso posible con el fin de obtener información nueva y que apoye las decisiones que conllevarán a recomendaciones aplicables.

Aún si se tuviese un presupuesto holgado, lo que se buscaría es un tamaño de muestra que logre un error adecuado para la variable en estudio y claro está para satisfacer los objetivos de la investigación.

Las fórmulas para los estimadores más usuales son:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

IC promedio conociendo la desviación típica

$$P(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

IC promedio desconociendo la desviación típica

$$P(\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}) = 1 - \alpha$$

IC para una proporción.

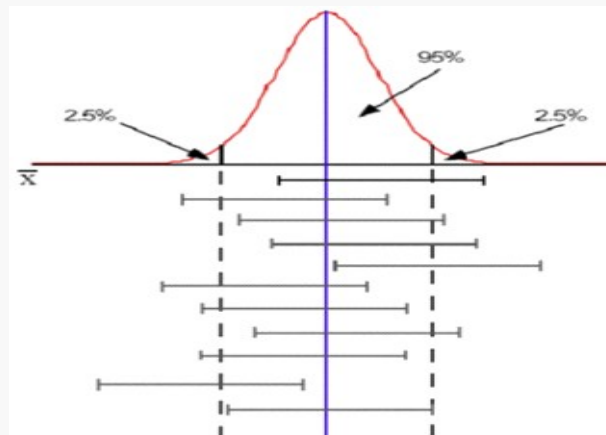
$$P(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \mu \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = 1 - \alpha$$

1.1.13 Nivel de Confianza: mide la probabilidad de que la estimación realizada este acorde con la realidad, de ahí, que se define formalmente como: La probabilidad de que el intervalo construido con base en una muestra contenga el verdadero valor del parámetro. Se tiene un paradigma y es que en lo general, se usa el 95% de confiabilidad, es un buen número, sin embargo, para estudios de percepción y específicamente en algunos temas, considero que una confiabilidad del 90% es adecuada, la razón, las percepciones de los seres humanos a menudo cambian como el día a la noche, a cada momento.

A mayor confiabilidad, se incrementa el tamaño de muestra.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Gráficamente se puede observar que en 95 de 100 veces el intervalo calculado contendrá el verdadero valor del parámetro, en otros 5 casos no lo contendrá, el tema es que nunca se sabrá si el nuestro es uno de los que realmente lo contiene.



1.1.14 Error absoluto (e) y relativo (e_r): El error, es un número que nos dice que tanto se aleja el estimador del verdadero valor del parámetro, es apenas lógico pensar que si ese valor es pequeño con respecto al indicador, la estimación es adecuada y por lo tanto, útil para apoyar la toma de decisiones.

Con este error, se construye el intervalo de confianza del estimador, es decir, mientras más estrecho sea el intervalo, mejor será la estimación.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Es importante destacar que el error absoluto depende de la variable objeto de estudio, es decir, si se desea estimar una proporción o prevalencia, este error, se define en porcentaje, y es usual trabajar con el 5%, porcentaje que en muchos casos no es adecuado. Consideremos que la proporción estimada por una prueba piloto es de 0.15, un error del 5%, realmente es mucho más alto, que si la proporción fuese del 0.45, valga decir, para valores alrededor del 0.5, es posible sustentar que el 5% es adecuado, pero para proporciones bajas, los errores absolutos, deben ser menores.

Ahora, cuando se desea estimar un promedio o un total, el error absoluto, se coloca en la unidad de medida que se esté estudiando, valga decir, si se va a estimar el ingreso familiar semanal, un error admisible de \$59 dólares puede ser alto o no, dependiendo de la población meta, valga decir, si se trata de profesionales de la salud, puede ser adecuado o no, igual situación se presenta si fuese de secretarias, por lo tanto, es prudente simular intervalos de confianza con ese error y analizar si el intervalo de confianza calculado es útil para la toma de decisiones.

Teniendo en cuenta la fórmula de intervalo de confianza, se tiene que el error absoluto es igual a:

$$e = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$e = z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Dependiendo de si el parámetro a estimar es un promedio o una proporción.

Por lo tanto, los límites de los intervalos se pueden expresar así:

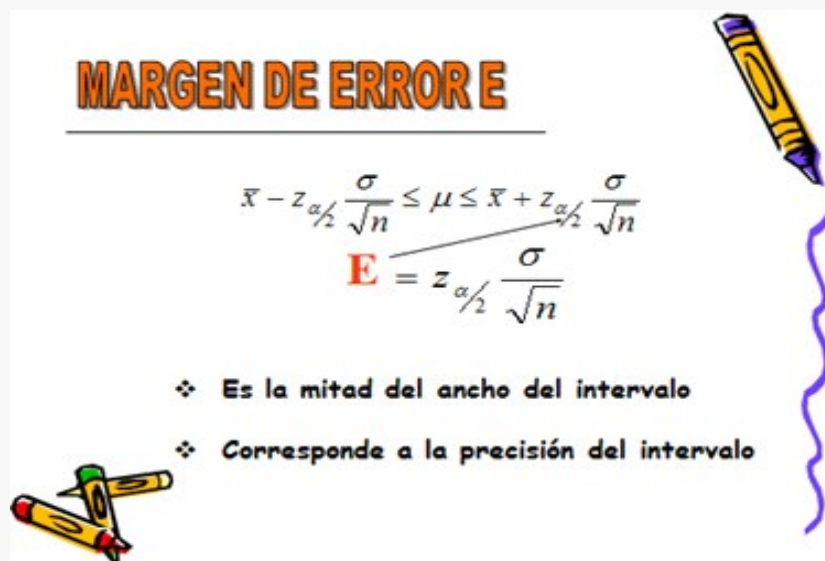
$$Li = \bar{x} - e$$

$$Ls = \bar{x} + e$$

$$Li = \hat{p} - e$$

$$Ls = \hat{p} + e$$

Veamos la siguiente figura para tratar de aclarar un poco la definición del error absoluto.



La opción para eliminar un poco la subjetividad en esta reflexión, es utilizar el error relativo.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

2. EL ERROR DE MUESTREO RELATIVO O COEFICIENTE DE VARIACION.

Es un número que nos dice que tanto se aleja el estimador del verdadero valor del parámetro. Depende de la variable objeto de estudio. Se mide con el coeficiente de variabilidad y es igual al cociente entre su error de muestreo y su esperanza matemática, así:

$$e_r = \frac{e}{x} \qquad e_r = CV(\hat{\theta}) = \frac{\sigma(\hat{\theta})}{E(\hat{\theta})} \qquad e_r = \frac{e}{\hat{p}}$$

En general un error relativo máximo del 10% se considera adecuado, sin embargo, esto es empírico, por lo tanto, en muchos casos, se puede y debe considerar errores relativos, no sólo mayores, sino, incluso más bajos según los objetivos del estudio.

Observemos la siguiente relación. Supongamos que se tiene un error absoluto del 5% y se tienen diversos valores estimados de P, para el primer caso, se tiene un intervalo de 0,45 a 0,55, una amplitud de 0,10 equivalente al ancho del intervalo e igual al error relativo (0,05/0,50).

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Relación entre los errores de muestreo

P	E	Er	Intervalo de confianza
0,5	5%	10,00	0,45 a 0,55
0,4	5%	12,50	0,35 a 0,45
0,3	5%	16,67	0,25 a 0,35
0,2	5%	25,00	0,15 a 0,25
0,1	5%	50,00	0,05 a 0,15

Observe que las amplitudes se mantienen, sin embargo, equivalen a diferentes coeficientes de variación, es decir, tienen errores relativos con diferentes precisiones, para mirar el caso extremo, si la prevalencia estimada fuese del 10%, el error relativo equivale a un 50%, por lo tanto, es un error inmenso y con poca utilidad práctica su IC.

Esto lo que lleva a concluir, es que el error absoluto del 5% no siempre es adecuado, es más, en muchos más casos es una cifra poco adecuado, se debe tener especial cuidado con los errores relativos, que se dice empíricamente, debe ser menor del 10%, con excepciones cuando las prevalencias son valores pequeños, alrededor del 20%, se pueden

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

considerar errores relativos mayores, quizás incluso cerca del 0,30 como error relativo puede darnos intervalos útiles.

Mantengamos constante el valor de P en el famoso 50% y observemos la relación de los errores absolutos con su equivalente relativo.

P	E	Er
0,5	0,05	10
	0,04	8
	0,03	6
	0,02	4
	0,01	2

La sugerencia es tener presente esta relación para tener claridad sobre con cuál coeficiente de variabilidad está trabajando realmente, use la fórmula adecuada para cada caso (proporción o promedio).

$$e_r = \frac{e}{\hat{p}}$$

$$e_r = \frac{e}{x}$$

Veamos ahora el caso para estimar el promedio de un grupo de personas en 79 Kg cuál es la relación entre los tipos de errores de muestreo.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Promedio	E	Er
79	4	5,06
	6	7,59
	6,5	8,23
	7	8,86
	7,5	9,49
	8	10,13

Observe que un error de 4 Kg equivale a un 5,06%, por lo tanto, 8 Kg corresponde a 10.13 %.

3. ERRORES NO MUESTRALES: Se cometen en las diferentes etapas del diseño muestral, en una encuesta por muestreo, pueden ser:

- Errores de cobertura.
- Errores en la formulación de las preguntas y sesgo en las respuestas.
- Errores de no respuesta.
- Errores de Procesamiento o digitación.

Es una fuente donde se pueden y deben realizar esfuerzos para disminuir

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

los costos del estudio, se percibe grandes esfuerzos en disminuir el tamaño de muestra con el objetivo claro de bajar los costos del estudio, a veces sin tener en cuenta el verdadero valor de las estimaciones. Si por el contrario, se rebajan los errores humanos, seguramente se podrá tener un tamaño de muestra adecuado sin detrimento de la precisión de las estimaciones.

4. PROCEDIMIENTOS DE MUESTREO: Existen diversas procedimientos de muestreo, tales como:

NO PROBABILISTICOS	PROBABILISTICOS
Conveniencia	Aleatorio Simple
Por Juicio u Opinión.	Estratificado
Cuotas o Prorratio.	Conglomerados
Con Fines Especiales.	Sistemático en Fases.
Bola de Nieve	Conglomerados Monoetápico
	Conglomerados Bietápico
	Conglomerados Bietápico Estratificado

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Este ebook, sin embargo, está pensado para apoyar la aplicación de muestreos probabilísticos y se desarrollarán en los diferentes capítulos.

5. PASOS PARA SELECCIONAR UNA MUESTRA.

La determinación del tamaño de muestra en los diferentes estudios ha sido y será punto clave en el desarrollo de los mismos, existen múltiples procedimientos para el cálculo de una muestra, por ello, se requiere definir unos criterios generales y sugeridos en la literatura estadística, es claro, que dependiendo de las particularidad del estudio y de la información disponible, se harán los cambios adecuado que serán sugeridos por el equipo investigador.

1. Definición de los objetivos del estudio.
2. Determinación de la población objetivo – Universo de estudio –
3. Construcción del marco muestral disponible: Es un listado de todas las unidades que conforman la población objetivo, puede ser una base de datos, un plano e incluso mapas. Es útil aclarar, que en la práctica se diferencia la población objetivo de la población en estudio (marco muestral disponible), se requiere tener claridad sobre los siguientes conceptos.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

- Elemento muestral: Es el elemento del cual se obtiene la medición.

- Unidad muestral: Está compuesta por elementos muestrales, en algunos casos es igual al elemento muestral.

- Alcance: Hace referencia a la ubicación espacial y geográfica del estudio.

- Tiempo: Definir el intervalo de tiempo en el cual se realiza la investigación.

4. Determinar los estimadores relevantes del estudio (promedios, proporcione, totales).

5. Definir la precisión deseada de los estimadores.

6. Forma de desagregar las variables objeto de estudio (cuadros de salida).

7. Reconocer los recursos económicos, de tiempo y humanos con que se cuenta.

8. Seleccionar un procedimiento de muestreo.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

9. Definir el tamaño de la muestra (Según el tipo de muestreo la Estadística proporciona fórmulas para calcular el tamaño muestral)

10. Seleccionar las unidades muestrales. El software estadístico Epidat 4.0 provee procedimientos para seleccionar unidades de análisis según diferentes procedimientos de muestreo.

Ejemplo: Para una encuesta de hogares realizada en Medellín en el año 2010, con el propósito de estimar el IPC (Índice de Precios al Consumidor), determine los pasos para seleccionar la muestra.

Solución.

1. Determinar el IPC en Medellín para el mes de mayo año 2010.
2. Población objetivo: Personas mayores de 18 años en Medellín.
3. Construcción del marco muestral disponible. Mapa cartográfico de la ciudad de Medellín.
 - Listado de sectores geográficos en Medellín.
 - Listado de barrios de cada comuna
 - Listado de manzanas de los barrios seleccionados

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

- Listado de viviendas de las manzanas seleccionadas
- Listado de las personas mayores de 18 años en las viviendas seleccionadas.

- Unidad muestral: Realmente se tienen varias unidades muestrales, una para cada etapa, la última es la vivienda (podría ser el hogar).
- Elemento muestral: Personas mayores de 18 años.
- Alcance: Medellín-Colombia- 2010.
- Tiempo: Mayo del 2010

4. Estimadores: precios de productos, por lo tanto, se trabajarán con promedios.

5. Definir la precisión deseada de los estimadores. Se debe de definir en términos de pesos y dependiendo del precio que se tome como guía, se definirá el error de muestreo o lo que es lo mismo la precisión.

6. Los resultados se entregarán por comunas y el total.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

7. Asumamos que se tienen recursos para alcanzar la muestra sugerida con las reflexiones previas.

8. Seleccionar un procedimiento. Muestro polietápico estratificado.

9. Definir el tamaño de la muestra. Se aplican las fórmulas del muestreo aleatorio simple, adicionando el factor deff.

10. Seleccionar las unidades muestrales. Usar un método aleatorio dentro de cada estrato y luego para cada etapa.

Visite el siguiente enlace [PLAN DE MUESTREO](#) podrá reforzar los conceptos leídos en esta sección, además, de que se introduce en el tema de muestreo aleatorio simple (MAS).

6. MUESTREOS PROBABILISTICOS.

Tienen la característica que todas las unidades poseen una probabilidad conocida y mayor de cero de ser seleccionadas. Pretenden disminuir la subjetividad originada al seleccionar las unidades de análisis, para evitar sesgos en la selección de las mismas. Además, permiten medir el error en la estimación, situación que no se da en los muestreos no probabilísticos. Los muestreos más usuales son: Muestreo simple aleatorio (MAS), Muestreo sistemático en fases (MSF), Muestreo aleatorio estratificado (MAE) y

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo por conglomerados (MC) y los muestreos polietápicos que pueden mezclar en cada etapa del proceso diferentes muestreos.

Como ventaja de los muestreos probabilísticos esta la eliminación de la subjetividad, el permitir medir la precisión de los estimadores, sin embargo, es importante tener en cuenta que no garantizan representatividad por si solos, se debe ayudar con un buen diseño y una buena construcción del marco muestral.

Diseños equiprobabilísticos: Son aquellos en que todos los sujetos de la población tienen la misma probabilidad de ser seleccionados. Los diseños con esta característica, no requieren ajustes con ponderaciones para corregir posibles sesgos cuando algunos elementos tienen mayor probabilidad de quedar incluidos en la muestra. Es por ello, que al calcular un promedio, sólo se requiere utilizar la fórmula clásica conocida por todos, es decir, **sumar y dividir**.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 2



MUESTREO ALEATORIO SIMPLE (MAS)

El MAS es el procedimiento probabilístico de selección de muestras más sencillo y conocido, no obstante, en la práctica es difícil de realizar debido a que requiere de un marco muestral y en muchos casos no es posible obtenerlo. Puede ser útil cuando las poblaciones son pequeñas y por lo tanto, se cuenta con listados. Cuando las poblaciones son grandes, se prefiere el muestreo en etapas. Se utiliza ampliamente en los estudios experimentales, además, de ser un procedimiento básico como componente de métodos más complejos (muestreo estratificado, por conglomerados y en etapas).

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Es un diseño equiprobabilístico, es decir, otorga la misma probabilidad de ser elegidos a todos los elementos de la población. Para él calculo muestral, se requiere:

- El tamaño poblacional, si ésta es finita.
- La confiabilidad del estudio, es usual utilizar: 90, 95 y 99%. Dependerá de los objetivos del estudio y lo define el equipo investigador.
- Error admisible, lo define el equipo investigador, recuerde que puede ser absoluto o relativo, siendo más recomendable el error relativo.
- Varianza del estimador.

Es importante realizar algunos comentarios con respecto al número surgido de la aplicación de estas fórmulas matemáticas.

Es evidente que el azar no garantiza representatividad, que es lo que se pretende al definir un buen diseño de muestra, no obstante, los modelos probabilísticos, son los únicos que pueden medir las estimaciones y por lo tanto, permiten generalizar los resultados

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

encontrados a toda la población.

Ejemplos son miles, si se selecciona de un grupo de 1000 estudiantes, de los cuales 750 son mujeres y el tema es sobre el aborto, suponiendo que las mujeres pueden tener opiniones diferentes de los hombres, al seleccionar 100 personas, el **azar** puede llevar a que la muestra quede compuesta por 90 mujeres y 10 hombres, con lo cual, probablemente no se obtengan resultados adecuados para esa población, claro que esto es poco probable que pase siempre, por lo tanto, cuando se realizan muchos trabajos, se tenderá a seleccionar muestras adecuadas según la composición de la población, además, se usan otros procedimientos probabilísticos para minimizar el problema mencionado. De otro lado, el **azar** no sólo elimina la subjetividad de los investigadores, sino, y es lo más relevante, en los muestreos probabilísticos se puede medir el grado de precisión de las estimaciones realizadas.

2.1 TAMAÑOS DE MUESTRA.

La pregunta más común y más trivialmente contestada es: “Cuántos elementos debo seleccionar?”, la respuesta usual es 384, contestada como ya se dijo en muchos casos de manera mecánica. Lo anterior, originada por

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

paradigmas consolidados y que consideran que la muestra se requiere para analizar variables dicotómicas, donde la probabilidad de éxito y la de fracaso son iguales ($p=q=0.5$), además, se sugiere casi indistintamente de los objetivos del estudio una confiabilidad del 95%, que si bien es buena, no siempre tiene que ser así. Por último, es usual que los no estadísticos consideren que el error de estimación es complementario a la confiabilidad y por ende lo asumen como $e=\pm 5\%$. Con estos valores y como se verá más adelante, la muestra sugerida es de 384. Lo que se pretende con esta disertación, es llamar la atención sobre éste particular. Pensando que el objetivo de la investigación será lo que prime para determinar el tamaño de muestra.

Los casos más usuales se dan para trabajos donde la distribución Binomial es la que interesa y por ende se requiere estimar la proporción de éxitos y cuando lo determinante es una variable cuantitativa y el interés es estimar un promedio. Para ambos casos, se presenta los requerimientos previos.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Proporciones	Promedios
Definir la confiabilidad del estudio	Definir la confiabilidad del estudio
Definir el error admisible (que tanto se acepta desviarnos del parámetro real). Se da en porcentaje.	Definir el error admisible (que tanto se acepta desviarnos del parámetro real). Se da en valor absoluto.
Estimar la proporción de éxitos (P).	Estimar la desviación de la variable clave (S_x).
Para el caso de poblaciones finitas, se requiere conocer N (tamaño de la población)	

La siguiente figura muestra las formulas a utilizar en éste muestreo cuando se trata de estimar la media o la proporción poblacional.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Tamaño de muestra para estimar la media y la proporción de la población

$$n_0 = \frac{z_{\alpha/2}^2 S^2}{e^2} \quad ; \quad n = \frac{n_0}{1 + \frac{n_0}{N}}$$

$$n_0 = \frac{z_{\alpha/2}^2 pq}{e^2}$$

n_0 se utiliza para población infinita, en caso de conocer N , se ajusta con la fórmula de n .

Engber Gilio Doris Zúñiga

Tamaño de muestra para estimar la media y proporción de la población

$z_{\alpha/2}$: Percentil calculado con la distribución normal, depende de la confiabilidad del estudio.

S^2 : Varianza muestral. Se usa en v. Cuantitativas.

e^2 : Error admisible, lo define el investigador.

p : Probabilidad de éxito en variables binomiales.

q : Probabilidad de fracaso, complemento de p .

Engber Gilio Doris Zúñiga

Nota: Cuando se estima que hay una posibilidad de no respuesta, el valor encontrado con cualquiera de las fórmulas planteadas y las que se plantearán se ajusta dividiendo el número calculado por el porcentaje de respuestas esperado, valga decir, si el porcentaje de no respuesta se estima en el 15%, el valor de n , se divide entre 0.85.

Otras fórmulas análogas y que se encuentran en muchos textos son:

$$n = \frac{Z_{(1-\alpha/2)}^2 NS^2}{Ne_{\alpha}^2 + Z_{(1-\alpha/2)}^2 S^2}$$

$$n = \frac{Z_{(1-\alpha/2)}^2 NPQ}{(N-1)e_{\alpha}^2 + Z_{(1-\alpha/2)}^2 PQ}$$

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Sin embargo, se obtienen resultados similares, además, acá se utiliza el Epidat para que nos apoye en el cálculo operativo.

Fórmulas en términos del error relativo:

Población finita

$$n = \frac{S^2}{e_r^2 (\bar{x})^2 + S^2 / N}$$

$$n = \frac{NQ}{(N-1)Pe_r^2 + Q}$$

Población infinita

$$n = \frac{S^2}{e_r^2 (\bar{x})^2}$$

$$n = \frac{Q}{Pe_r^2}$$

Por otro lado, en la práctica es difícil conocer el verdadero tamaño de la población objeto del muestreo, a veces se tiene sólo una estimación de la misma. No obstante, esto no es problema si se tiene en cuenta que cuando la población es grande (N mayor de 10.000) el uso del tamaño de la población no tiene injerencia grande en la muestra, por ejemplo^[1], si se realizan dos estudios sobre mediciones antropométricas: uno en una ciudad

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

intermedia con aproximadamente 500.000 trabajadores, y otro en una ciudad con 55.000 trabajadores. Si se plantea estimar el índice de masa corporal en personas con edades entre 20 y 29 años, donde se tenga la misma desviación estándar (4.0 Kg/m^2) y con los mismos indicadores de confiabilidad (93%) y error admisible (0.8 Kg/m^2), se tiene que para estimar el promedio, para el primer caso se requiere una muestra de 83 personas, mientras que para el segundo la muestra sería de 82. Se concluye entonces, que siempre que la población sea mucho mayor que los posibles tamaños muestrales, se evidencia el error no poco común, de sugerir como posible tamaño de muestra un porcentaje de la población. Para el caso, si se selecciona el 2% de la población, se puede demostrar usando el Epidat 4.0 que el error para una muestra de 1100 corresponde a 21.6%, mientras que para la muestra de 10.000 el error sería de 6.6%.

Otra situación a tener en cuenta, es determinar la variable que va a direccionar o aportar la variabilidad para calcular el tamaño de muestra, ya que en toda investigación se encuentran múltiples variables, de ahí, que en ningún momento se puede perder de vista los objetivos del estudio.

Si, por ejemplo, lo que se quiere es estimar el promedio de días de estancia en un hospital para una población de mujeres con problemas renales, y se admite un error de hasta $e=\pm 5$ días, se acepta que: "El verdadero promedio

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

se desvié del estimado en la muestra en 5 días, dicho de otra manera, si el promedio muestral fuera de 15 días, se acepta que el verdadero se encuentre entre 10 y 20 días, si esto fuese mucho, se recomienda disminuir el error.” Otro investigador pudo seleccionar un error de 3 días.

En general se cumple que al aumentar la precisión, es decir, al disminuir el grado de error, aumenta el tamaño de muestra necesario. En el caso de proporciones, el error se expresa como un porcentaje del valor que puede tener el parámetro, es decir, 5% o un 6%, en ese caso se habla de precisión relativa.

El nivel de confianza, generalmente se fija en 95%. El efecto de este valor sobre el tamaño de muestra es directo; es decir, para una precisión dada, el tamaño de muestra aumenta al aumentar el nivel de confianza.

Si lo que se pretende es calcular el tamaño cuando lo relevante es una variable dicotómica, es decir, Binomial, se utiliza la siguiente expresión, asumiendo que se quiere una confiabilidad del 95%, un error admisible de 6% y se estima la proporción de éxitos en la muestra en 0.30, además, la población es 6100:

Si se cuenta con el apoyo del programa Epidat 4.0, estos cálculos se simplifican, la ayuda del programa explica así la manera de calcular los

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

tamaños de muestra: "...generando una tabla de valores para la precisión y proporcionando el tamaño de muestra correspondiente a cada uno de esos valores. La tabla se define partiendo de un valor mínimo fijado por el usuario y aumentando a intervalos regulares de incremento hasta llegar al valor máximo. Por ejemplo, si se definen mínimo = 2, máximo = 6 e incremento = 2, se obtiene una tabla con los valores 2, 4, y 6 para la precisión y los correspondientes tamaños de muestra."

En algunos casos puede ser de utilidad realizar el cálculo inverso, es decir, determinar la precisión asociada a la estimación de un parámetro en función del tamaño de muestra empleado.

Para el caso de proporciones, se tiene la tendencia de utilizar el valor $p=0,5$ en la formula respectiva, con la presunción, de que se garantiza un mayor n , situación que es cierta sólo si se mantienen constantes los otros componentes de la ecuación, desconociendo además, que es muy improbable que los investigadores desconozcan por completo la distribución de la población con respecto a la variable dicotómica bajo estudio. Además, la recomendación de realizar la prueba piloto es **imprescindible** y con ella se estima la proporción de manera más racional. Lo anterior, tiene ingerencia en determinar el error admisible, en el sentido que si la

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

proporción es baja, dígame 0.10, el error no podría ser de un 5%, ya que sería muy alto comparativamente con la proporción real estimada, situación que podría ser comprensible si el valor de p fuese mayor.

Como un último comentario, es mencionar que el MAS, lleva a muestras más pequeñas teniendo en cuenta la precisión que si se utilizarán muestreo más complejos, una forma de solucionar este problema es multiplicar el tamaño obtenido con el MAS por el llamado **Efecto del Diseño**. De nuevo surge la subjetividad y se plantea que este efecto debe estar entre 1.5 y 3.0 entendiéndose, que 2.0 indicaría un tamaño de muestra doble al sugerido por el MAS.

2.1.1 Tamaño de muestra para estimar un promedio poblacional, utilizando el Epidat 4.0. Con la versatilidad del programa, se observa cuáles son los efectos de determinar un error admisible muy pequeño, que puede suceder con la muestra si prefiere aumentar la confiabilidad, o si el problema es el tamaño de la desviación, además, cuando tiene incidencia el conocer el tamaño de la población. Estos aspectos se analizarán con un ejemplo.

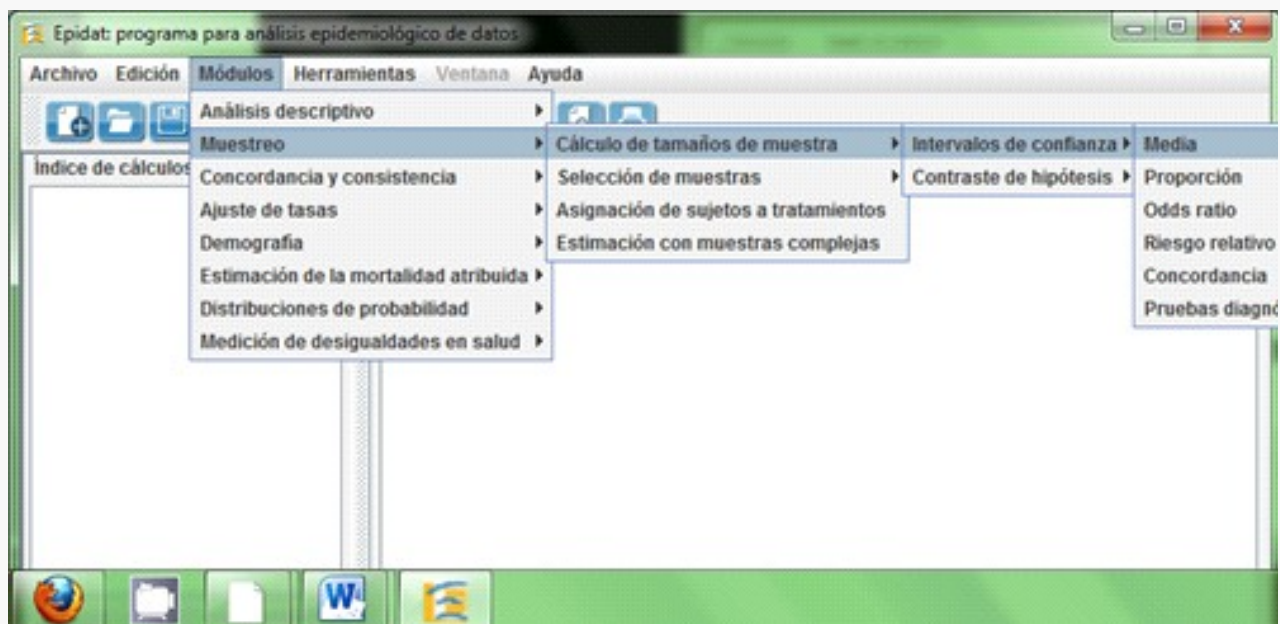
Por estudios previos se tiene conocimiento que la distribución del peso al nacer de niños que cumplen su período de gestación de 40 semanas es

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

aproximadamente normal con una media de 3550 gramos y una desviación estándar de $\sigma=250$ gramos. Se va a realizar un nuevo estudio para una población con características similares, con el fin de estimar el peso promedio al nacer de los niños. Con base en el estudio previo se determina el tamaño de muestra. Además, se considera que un error de máximo 45 gramos logra una estimación válida, la confiabilidad del estudio es del 95%.

Solución Epidat 4.0.

Luego de abrir el programa, siga la siguiente ruta definida en la imagen.



Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Se observa que el software tiene diversas posibilidades para calcular el tamaño de muestra según los objetivos del estudio, éste ebook trata sobre estudios descriptivos, donde el interés es estimar parámetros y por lo tanto, se utilizan las formulas definidas para intervalos de confianza.

Para el caso de estimación de un promedio, se tiene la siguiente caja de dialogo, y como ya se había expresado, se requiere el tamaño de la población (opcional), la desviación estándar del estimador, el nivel de confianza y el error absoluto, tal como se aprecia en la imagen siguiente.

El efecto del diseño se utilizará en los muestreos polietápicos, por lo tanto, se mantiene en 1.0. Otro aspecto a aclarar, es que el programa realizar una simulación para calcular diferentes tamaños de muestra para diversos errores y con ello, analizar la incidencia de estos en los intervalos de confianza teniendo así un elemento práctico para elegir un tamaño de muestra útil para la toma de decisiones.

Para el ejemplo anterior, la salida del software es la siguiente:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

[1] Tamaños de muestra. Media:

Datos:

Tamaño de la población:	∞
Desviación est. esperada:	250,000
Nivel de confianza:	95,0%
Efecto de diseño:	1,0

Resultados:

Precisión	Tamaño de la muestra
25,000	385
30,000	267
35,000	196
40,000	151
45,000	119

Se tienen varios errores absolutos con sus respectivos tamaños de muestra, ya es otros aspectos incluyendo la parte económica la que nos orientará en la definición de la n.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Como estrategia es recomendable digitar un error máximo y uno mínimo con el fin de apreciar el impacto en el tamaño de muestra del error admisible.

e: Error que se admite como viable y se mide en la misma unidad de medida que la variable. Se considero para efectos del ejemplo, un error máximo de 45 gramos.

Es claro que para confiabilidad, desviación y tamaño poblacional igual, pero errores permisibles diferentes, origina mayor n el error menor.

Refuerce los conceptos anteriores, observando el video de 7 minutos que nos habla sobre este tema, de clic en la frase siguiente [TAMAÑO DE MUESTRA](#)

2.1.2 Tamaño de muestra para estimar una proporción poblacional, utilizando el Epidat 4.0. En este caso, la caja de dialogo es similar, sin embargo, se debe entregar la proporción en forma de porcentaje para realizar los cálculos con el Epidat 4.0, veamos el siguiente ejemplo.

Determinar el número de profesionales a encuestar en una región donde se estima en 4500 el número de ellos. El objetivo del estudio es determinar entre otras cosas, la intencionalidad de seguir estudios de maestría, con una

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

prueba piloto de 20 profesionales, de terminó que la proporción de profesionales con afán de continuar sus estudios era del 25%. La confiabilidad del estudio, dado que sus resultados serán validados con otras fuentes se definió en el 90%, el error puede estar entre el 4 y el 6%, dependiendo de los costos se definirá cual tamaño seleccionar.

Proporción

Datos

Tamaño de la población: 4 500

Proporción esperada: 25.000 %

Nivel de confianza: 90.0 %

Calcular

☒ Tamaño de la muestra

☐ Precisión absoluta

Efecto de diseño: 1.0

Precisión absoluta (%)

Mínimo: 4.000

Máximo: 6.000

Incremento: 1

Ocultar Calcular Limpiar Cerrar

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

[2] Tamaños de muestra. Proporción:

Datos:

Tamaño de la población:	4,500
Proporción esperada:	25,000%
Nivel de confianza:	90,0%
Efecto de diseño:	1,0

Resultados:

Precisión (%)	Tamaño de la muestra
4,000	297
5,000	190
6,000	137

Se requieren 297 profesionales si se quiere un error del 4% en la estimación.

Observaciones Generales.

Se destacan los siguientes aspectos, si se conoce el tamaño de la población y éste no es muy grande, es útil utilizarlo, ya que esto disminuye el tamaño de muestra, tal como se aprecia en las figuras anteriores. De nuevo se aprecia que a mayor error, menor tamaño de muestra. Para el caso, el investigador tendrá que adicionar otras variables como el costo y el tiempo para definir cual de las tres opciones es más válida para sus propósitos.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Se concluye que tanto la confiabilidad como el error permitido, son de potestad de los investigadores y si bien pueden ser subjetivos, son ellos quienes mejor conocen las variables en estudio y sus cambios. Por lo tanto, varios grupos de investigación pueden llevar a diferentes tamaños muestrales, para el mismo estudio, la situación está en como se argumentan las decisiones.

Observemos la siguiente diapositiva donde se muestra la incidencia del error absoluto en la precisión de los intervalos de confianza.

Incidencia del error de estimación

Tamaño de muestra	Error de estimación	Proporción esperada	Limite inferior	Limite superior
650	0,025	0,12	0,095	0,145
451	0,03		0,09	0,15
332	0,035		0,085	0,155
254	0,04		0,08	0,16
201	0,045		0,075	0,165
163	0,05		0,07	0,17
135	0,055		0,065	0,175
113	0,06		0,06	0,18
97	0,065		0,055	0,185

n=41 Intervalo 2% a 22%

¿Es útil para la toma de decisiones?

www.leondariobello.com
www.olemonline.info/blog

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

El criterio es analizar la amplitud de los intervalos de confianza, observemos que una muestra de 650 entrega una amplitud de 0,05, lo que se puede considerar una alta precisión, por el contrario una muestra de 97 entrega una amplitud de 0,13%, la pregunta es; el IC es útil y aporta a una buena toma de decisiones?, para ser más extremista, se calculó un IC para $n=47$ y el intervalo dio entre 0,02 y 0,22, no se requiere ser experto para entender que este intervalo no es útil. Sería como si dijéramos que las personas que ingresan a realizar un postgrado tienen en promedio una edad entre 25 y 55 años, es esto importante?.

2.1.3 METODOS DE SELECCIÓN CON EPIDAT 4.0.

El software permite seleccionar las unidades de análisis de dos maneras: la primera, se digita el tamaño de la población, si se conoce, y el tamaño de la muestra, de esta manera, se puede guardar en un archivo los números de los registros seleccionados, es importante anotar, que si se realiza de nuevo el procedimiento, se obtendrán diferentes casos o registros. La segunda, se utiliza cuando se tienen los datos guardados en un archivo, generalmente Excel, puede ser un archivo guardado en open office.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

La forma de solicitar la selección para cualquier procedimiento probabilístico con Epidat 4.0 es la siguiente:

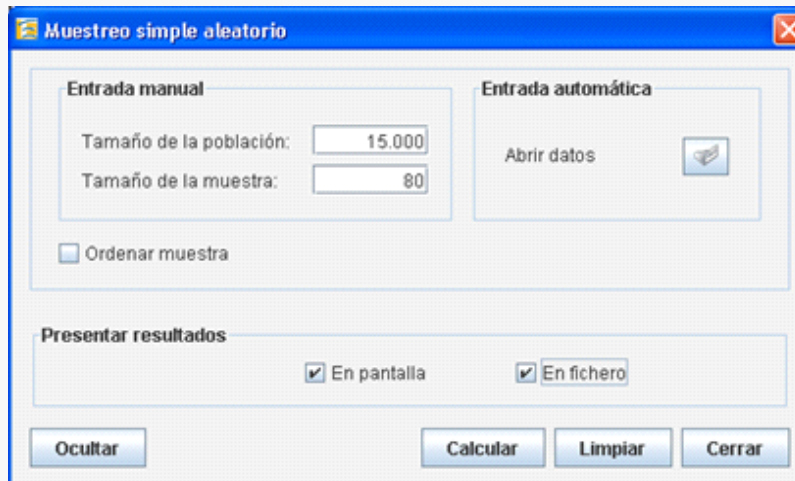


Luego de dar clic en muestreo simple aleatorio, se debe de entregar la información requerida según sea el caso. Si se trata de solicitar una generación de números aleatorios, se procede según la siguiente caja de dialogo, donde se definió una población de 15.000 y se requieren 80 elementos (entrada manual). Observe que se tienen seleccionadas las opciones de pantalla y fichero, lo que indica que los 80 números aleatorios entregados por el software serán grabados en un archivo, generalmente Excel, para luego si es del caso, exportarse a un programa estadístico.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



Muestreo simple aleatorio

Entrada manual

Tamaño de la población: 15.000

Tamaño de la muestra: 80

☐ Ordenar muestra

Entrada automática

Abrir datos

Presentar resultados

☒ En pantalla ☒ En fichero

Ocultar Calcular Limpiar Cerrar

Los resultados como se solicitó, los puede mostrar por pantalla o guardar en un archivo. La probabilidad de inclusión es de 0,0053 resultado de dividir 80 entre 15000.

Note que la salida entrega la probabilidad de selección o inclusión de cada elemento, para el caso es muy baja, dado que la población es muy grande y la muestra para el ejemplo muy pequeña, su calculo es igual a:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

$$P = \frac{n}{N} = 0,0053$$

Ponderación=1/0,0053= 187,6

Se introduce el término ponderación, el cual es usual en los procedimientos por etapas, por ahora, sólo se menciona que el peso o ponderación de un elemento de la muestra, es igual al inverso de la probabilidad de selección y quiere decir a cuantos elementos muestrales representa cada unidad seleccionada.

El archivo guardado, tiene los mismos registros con todas las variables.

Veamos ahora como se utiliza cuando se tienen los datos guardados en un archivo Excel (entrada automática).

Como ejemplo, se tomará un archivo con 992 registros, el archivo se anexa con el ebook y tiene el nombre de Base de datos Regiones, dicho archivo tiene la siguiente estructura:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Región

Subregión

Municipio

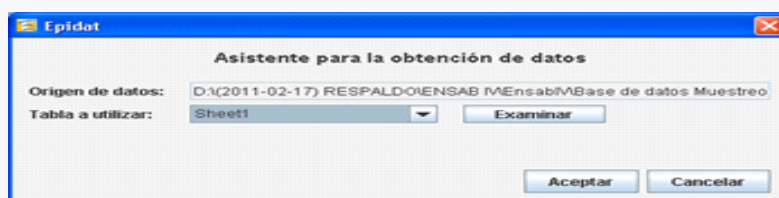
Categoría de municipio

No viviendas

No personas

Para usar la opción de archivo, se debe dar clic en abrir datos, con lo cual se muestra la siguiente caja de dialogo permitiendo seleccionar el archivo deseado, claro que los resultados no coincidirán con los del ebook, dado que siempre saldrán números diferentes.

Seleccione la opción Abrir datos de la caja de dialogo anterior, mostrando lo siguiente:



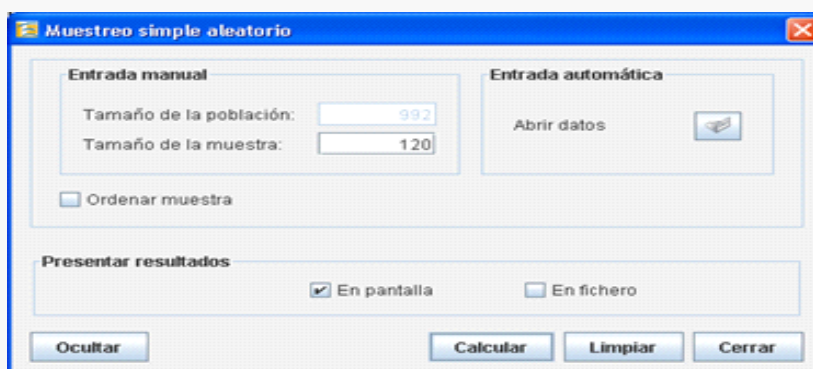
Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Luego de examinar y navegar por su disco duro o su USB, debe seleccionar el archivo que contiene la información pertinente.

El sistema llena el cajón correspondiente al tamaño de población y para el ejemplo, se tomará una muestra de 120 registros. Tal como se observa en la caja siguiente:



Los resultados obtenidos fueron:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Resultados

[1] Muestreo simple aleatorio:

Entrada automática:

Archivo de trabajo: D:\(2011-02-17) RESPALDO\ENSAB IV\EnsabIV\Base de datos Muestreo.xlsx
 Tabla: Sheet1

Datos:

Tamaño de la población: 992
 Tamaño de la muestra: 120

Número de los sujetos seleccionados:

544	827	405	613	49	89	787
808	8	876	389	153	97	129
156	708	933	801	122	991	216
393	879	397	430	611	259	179
131	25	379	621	451	513	949
26	22	575	907	619	644	658
618	511	564	343	711	828	955
517	684	65	858	877	733	614
4	321	627	134	791	809	80
880	383	332	766	150	715	335
103	755	503	964	637	583	487
987	751	834	434	946	872	253
242	590	685	529	917	47	10
981	376	420	24	616	663	31
526	493	782	20	963	978	384
968	204	990	155	235	727	147
320	75	364	695	850	399	211
126						

Probabilidad de selección: 12,0966%

Note que se presenta la probabilidad de selección, para el caso es de 0,12096, resultado del cociente entre n/N , es decir, $120/992$.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 3

MUESTREO SISTEMÁTICO (MS)

Tiene el mismo principio del MAS, no obstante, no requiere tener marco muestral, además, en muchos casos es más económico, no obstante, tiene la desventaja de la periodicidad, es decir, al obtener las unidades o elementos muestrales de manera sistemática, se pueden realizar mediciones que obtienen estimaciones sesgadas, valga decir, si pretende estimar las ventas por día en la zona Rosa del Poblado en Medellín o en la zona Rosa en Bogotá, y el salto o frecuencia es 7, tiene el inconveniente que siempre será seleccionado el mismo día de la semana, por ejemplo, el lunes, presentando estimadores de ventas muy por debajo de lo real, o por el contrario seleccionar todos los viernes, donde la estimación queda sobre valorada. Para determinar el n , se utilizan las mismas formulas del MAS, no obstante, el procedimiento de selección regular, cambia, así:

- Calcular el tamaño de muestra con las formulas vistas.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

- Determinar el salto o frecuencia de selección. $K=N/n$.
- Seleccionar el primer valor aleatoriamente en el rango 1 a K.
- Seleccionar los demás elementos, sumando al número que identifica el primer elemento la constante K, hasta completar n.

Veámoslo de la siguiente manera.

MUESTREO SISTEMÁTICO REGULAR

Se selecciona el 1º entre 1 y K, se manera aleatoria y luego se suma k.

N=20000	n=150	f=P=n/N=0,0075
	$K=N/n=133$	
89	89 seleccionado al azar entre 1 y 133	
222	Así sucesivamente hasta ajustar el n	Se pueden seleccionar 133 posibles muestras
355		
488	útil ordenar archivo	

$K = N / n$ $f = p = n / N$

www.leondariobello.com **CIEMONLINE**

Se tienen una población de N=20000 personas, se desea una muestra de 150, por lo tanto, el intervalo de salto llamado K, es igual a 133, luego entre

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

1 y 133, se selecciona un número al azar, para el ejemplo y usando el Epidat nos dio 89, éste es el primer elemento seleccionado, a partir de ahí, se empieza a suma 133, hasta ajustar 150 valores, que son los que integran la muestra.

El otro término que se introduce, al menos en nomenclatura, es la probabilidad de inclusión, denominada $f = n/N$ y que como se aprecia es el inverso de K .

Para el caso del muestreo en fases, se presenta la siguiente información,

MUESTREO SISTEMÁTICO FASES

Se tiene previamente la probabilidad de inclusión, lo que implica que es difícil que el valor de k sea entero.

Se selecciona un elemento de cada k_1 y de los que no fueron seleccionados, seleccionar uno cada k_2 unidades.

$$f = \frac{k_1 + k_2 - 1}{k_1 k_2}$$

Probabilidad de inclusión

CIEMONLINE

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

El cambio origina una nueva forma de calcular las probabilidades de inclusión, originado en el procedimiento de que se tienen 2 puntos de arranque, k_1 y k_2 , el primero originado de la manera mencionada en el procedimiento regular, el segundo, es otro punto de arranque de un intervalo superior y que toma como base los no seleccionados inicialmente, de ese grupo, se seleccionan algunos elementos para llegar a un valor cercano a la muestra definida previamente.

El Epidat 4.0, también tiene implementado el procedimiento sistemático en fases, como se muestra en la caja de dialogo siguiente. El procedimiento consiste en seleccionar un porcentaje P de los N datos, adicional a lo seleccionado usando el salto ya definido previamente en el procedimiento regular. En este método es usual que el tamaño de muestra no coincida con lo solicitado, dado los redondeos que toca realizar.

Solución Epidat 4.0.

Luego de abrir el programa, se activa la opción **métodos + muestreo + selección muestral + sistemático en fases**. Para luego llenar los requerimientos del programa teniendo en cuenta las necesidades de los investigadores, así:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo sistemático

Tipo de muestreo

Sistemático en fases
Sistemático regular
Sistemático en fases

Entrada manual

Tamaño de la población: 150

Probabilidad de selección: 15 %

Entrada automática

Abrir datos

Presentar resultados

☒ En pantalla ☐ En fichero

Ocultar Calcular Limpiar Cerrar

Debo destacar que aparte del tamaño de la población se debe de ingresar la probabilidad de selección, es decir, se debe multiplicar dicho porcentaje por el N para estimar el valor de la muestra, para el ejemplo, se seleccionaron 22 unidades. La salida es la siguiente:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

[3] Muestreo sistemático:

Datos:

Tipo de muestreo: Sistemático en fases
 Tamaño de la población: 150
 Probabilidad de selección: 15%

Resultados:

Intervalos de selección(k1 y k2)		Arranques aleatorios(r1 y r2)	
6	-10	4	1

Número de los sujetos seleccionados:

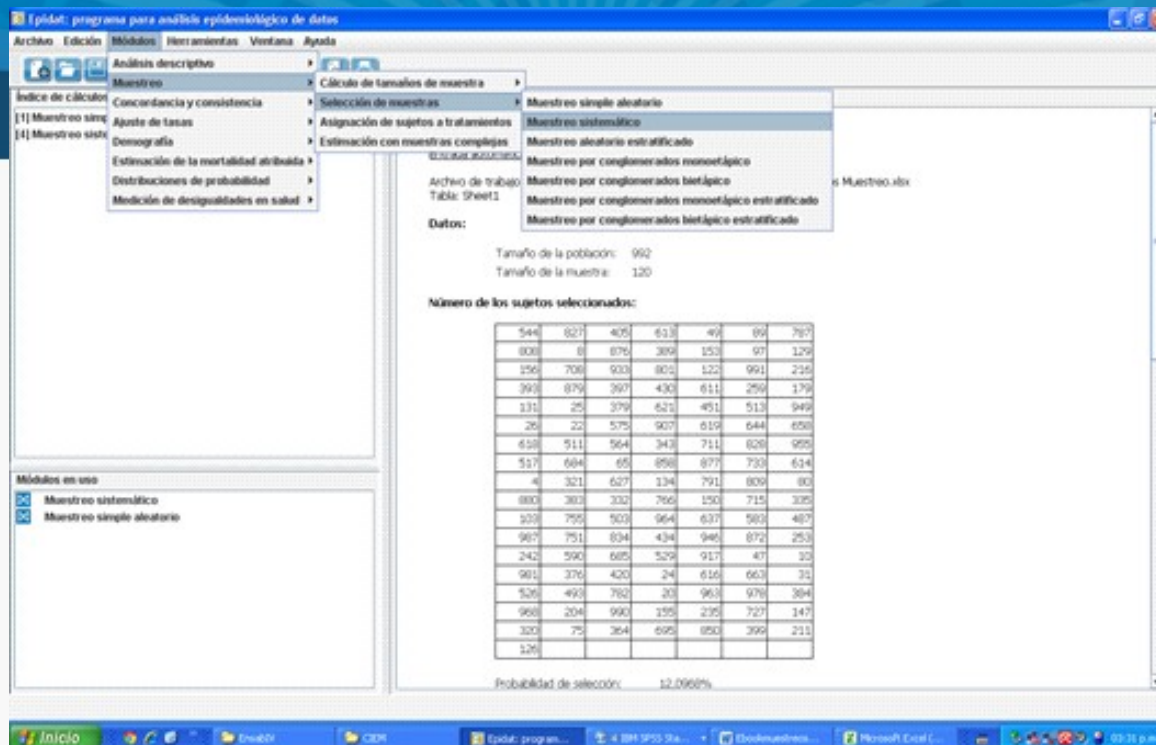
10	16	22	28	34	40	46
52	58	70	76	82	88	94
100	106	112	118	130	136	142
148						

Tamaño de muestra resultante: 22

Se han seleccionado 1 de cada 6 unidades, empezando en el 4, y de los no seleccionados añade 1 de cada 10 arrancando en el primero. Dado que el tamaño de muestra requerido es de 251, se calcula la proporción, es decir, $251/1200 = 21\%$, que es el valor que requiere el Epidat 3.1 para definir el salto y seleccionar los números sugeridos.

Retomando el ejemplo de las regiones cuyo archivo está compuesto por 992 unidades, se solicitaron 120 registros, con el fin de comparar los resultados, los resultados y la ruta se muestra a continuación.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



Como ha notado, se tienen algunos videos complementarios que se consideran de gran apoyo para afianzar su conocimiento sobre los diversos temas, por lo tanto, lo invito a observar el video sobre [MUESTREO SISTEMATICO](#), dando clic.

Con el fin de determinar la diferencia en los resultados de los dos procedimientos, se comparan los resultados obtenidos por región (archivo usado anteriormente).

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Page 67 of 115

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Región	Toda la población		MAS		MS	
	No	Porcentaje	No	Porcentaje	No	Porcentaje
1	173	17,4	25	20,8	22	17,7
2	364	36,7	39	32,5	45	36,3
3	261	26,3	33	27,5	33	26,6
4	150	15,1	19	15,8	18	14,5
5	44	4,4	4	3,3	6	4,8
Total	992	100	120	100	124	100

Los resultados obtenidos con el MS, son más similares a la composición de la población que el MAS, en éste caso, la sugerencia es mirar en cada caso y teniendo en cuenta como se estructura el archivo de entrada, cuál representa de mejor manera la población objeto de estudio.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 4

MUESTREO ALEATORIO ESTRATIFICADO (MAE)



Cambiando de estrategia, lo invito a ver primero el siguiente video para luego reforzar lo dicho en él, entre al enlace [MUESTREO ESTRATIFICADO 1](#)

Se utiliza en aquellos casos donde la población es muy heterogénea, los métodos precedentes conllevan a estimaciones poco precisas, por lo tanto, se presume que dicha población está afectada por otra variable que puede minimizar la variabilidad, por ejemplo, los salarios y en general las variables que involucran dinero, son muy variables, no obstante, si se parcela la información según profesión, o años de experiencia para el caso del salario,

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

seguramente, se disminuirá ésta y permitirá estimaciones más precisas. Si se analiza la estatura, los grupos de edad deberían disminuir la variabilidad, Lo anterior, conlleva a disminuir costos y lograr mayor eficiencia en el diseño muestral.

Se caracteriza entonces, por poder desagregar la población en categorías que forman grupos de elementos excluyentes (estratos), teniendo la característica que sus componentes sea lo más homogéneas posibles, pero bien diferentes entre cada estrato.

Los estratos deben ser tan homogéneos como sea posible, pero un estrato debe diferir tanto como se pueda de otro con respecto a la característica que está siendo medida. Veamos cómo se pueden construir.



Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Se tienen 5 estratos conformado por personas de diferentes profesiones cada uno de ellos.

El método más usual es seleccionar dentro de cada estrato de manera aleatoria (MAS o MS) los elementos, repartiendo la muestra en algunos casos de manera proporcional entre ellos, es decir, donde hay más observaciones, se toma más muestra y donde hay menos, se toma menos, de tal manera que sean equivalentes a la proporción que representan en la población. Sin embargo, hay que tener en cuenta que cuando hay estratos con pocos elementos, es susceptible unir con otras categorías e incluso, disminuir en estratos con mucha información para asignarla a otras con pocos elementos. Esta decisión, no obstante, está sujeta a los objetivos del estudio.

Al utilizar el Epidat 4.0, se presentan dos casos: El primero cuando la única información disponible es el número de estratos y el tamaño de los mismos. Para éste caso, se debe crear un archivo en Excel, con dos columnas, una con la cantidad de estratos, numerada de 1 hasta K, donde K es el total de estratos y otra con el número de unidades de cada estrato. La segunda, cuando se tiene la base de datos con información de los elementos

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

muestrales, también en Excel.

Ejemplo 1: Una empresa tiene 6100 empleados, se quiere determinar como es el clima laboral en la empresa, usando una confiabilidad del 95%, un error admisible de 6% y considerando que la proporción de empleados no satisfechos es del 30%. Calcule el número de empleados a consultar, si se tiene en cuenta además, que se tienen diferentes categorías de empleados que pueden influir en la opinión de los trabajadores, se adicionó la siguiente información con respecto al número de trabajadores: Contabilidad y Costos 80 empleados, Administración 150, operativos 5600, seguridad 180 y otros cargos 90.

X: Número de empleados no satisfechos con el clima organizacional.

Al utilizar el muestreo estratificado proporcional, se reparte la muestra en los diferentes estratos, de tal manera, que donde se encuentren más unidades de análisis, se obtengan más datos, para ello, se utiliza la siguiente fórmula:

$$n_i = n * (N_i / N)$$

La caja de dialogo en este caso es la siguiente:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo aleatorio estratificado

Entrada manual: Número de estratos: 6

Entrada automática: ☒ Abrir datos individuales ☐ Abrir datos agregados

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

Estrato	Tamaño del estrato	Tamaño de la muestra
1	3500	123
2	1000	35
3	800	28
4	350	12
5	280	10
6	170	6

Tamaño de la muestra: 214 **Distribuir**

Datos: ☐ Ordenar la muestra

Presentar resultados: ☒ En pantalla ☐ En fichero

Ocultar **Calcular** **Limpiar** **Cerrar**

La información solicitada es fácil de identificar, hay que entregar el número de estratos, el tamaño de muestra, permite distribuir la muestra proporcionalmente, claro que también se puede asignar los n de cada estrato de manera manual.

Los resultados según estratos fueron:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Estrato 2:

338	837	109	344	243	986	387
540	846	628	587	64	497	769
394	605	559	215	200	963	779
533	4	55	133	763	247	939
919	206	617	477	581	860	506

Estrato 3:

21	299	498	156	107	634	263
477	500	545	602	711	464	561
134	72	581	717	605	181	226
96	88	460	124	246	347	63

Estrato 4:

81	178	147	317	340	21	296
292	13	44	289	68		

Estrato 5:

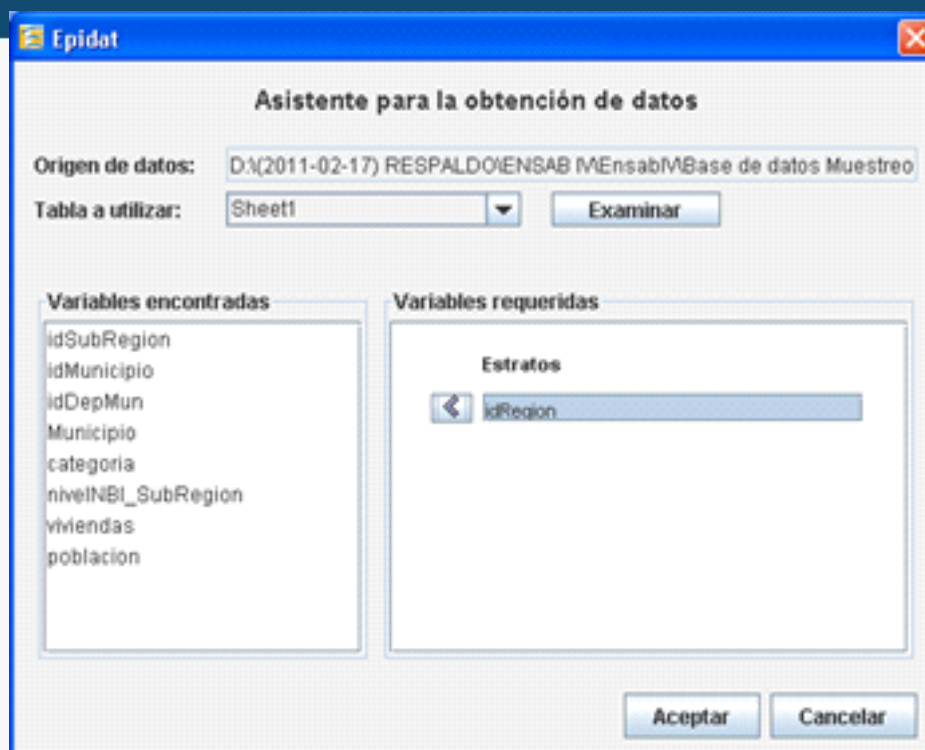
161	92	45	135	272	109	238
126	55	235				

Estrato 6:

159	83	149	71	124	144
-----	----	-----	----	-----	-----

También permite abrir un archivo en formato Excel y decirle que variable contiene los estratos. Veamos el ejemplo con el archivo **Base de datos regiones**, donde la variable o columna con el nombre de idregión es la que identifica el estrato.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



Para el caso, el interés es ir a todas las regiones, de ahí que se escoge este procedimiento, que garantiza obtener información de todos los estratos, si se requiriera visitar a todos los municipios (poco práctico) el estrato sería esa variable.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo aleatorio estratificado

Entrada manual: Número de estratos:

Entrada automática: ☒ Abrir datos individuales ☐ Abrir datos agregados

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

Estrato	Tamaño del estrato	Tamaño de la muestra
1	173	21
2	364	44
3	261	32
4	150	18
6	44	5

Tamaño de la muestra: **Distribuir**

Datos: ☐ Ordenar la muestra

Presentar resultados: ☒ En pantalla ☐ En fichero

Ocultar **Calcular** **Limpiar** **Cerrar**

El programa llena automáticamente los tamaños de los estratos y el muestrista, digita el tamaño de muestra teniendo en cuenta las sugerencias planteadas en el plan de muestreo. Los resultados fueron:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Comparando estos resultados con la población objetivo, se observa que la representatividad por región es más adecuada que los diseños anteriores.

Número de los sujetos seleccionados:

Estrato 1:

102	7	119	84	116	91	53
88	147	80	173	47	157	132
4	167	150	106	31	112	169

Estrato 2:

71	193	191	338	348	58	251
74	149	139	160	220	94	104
7	293	121	219	209	283	252
89	125	108	279	239	129	302
261	30	116	3	21	362	18
45	84	82	198	150	309	145
218	51					

Estrato 3:

118	163	82	15	1	251	126
185	106	114	187	176	128	223
158	45	32	192	220	61	113
193	234	175	213	41	110	78
225	183	30	154			

Estrato 4:

106	121	49	29	112	47	62
145	35	43	16	11	71	88
22	65	127	17			

Estrato 6:

16	4	31	19	18
----	---	----	----	----

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Región	Toda la población		MES	
	No	Porcentaje	No	Porcentaje
1	173	17,4	21	17,5
2	364	36,7	44	36,7
3	261	26,3	32	26,7
4	150	15,1	18	15,0
5	44	4,4	5	4,2
Total	992	100	120	100

Se puede concluir, que este muestreo es muy usual debido a que en muchos trabajos se requiere desagregar la información por alguna característica y por lo tanto, es clave obtener datos de todas las categorías. La desagregación de los resultados es otra razón para pensar en el muestreo estratificado.

El Epidat permite definir diferentes opciones para seleccionar las unidades, observe la diapositiva siguiente:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

MUESTREO ESTRATIFICADO
MUESTREO ESTRATIFICADO
Epidat 4.0

Posibilidades para definir el número de unidades a seleccionar:

1. Igual para todos los estratos.
2. Libre para que el usuario las defina.
3. Distribución proporcional según un tamaño de muestra predefinido.



www.leondariobello.com
www.ciemonline.info/blog

Veamos para el caso de muestras iguales:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

EPIDAT 4.0

Estrato	Tamaño del estrato	Tamaño de la muestra
1	5690	120
2	3457	120
3	4500	120
4	1260	120
TOTAL	14907	480

Estrato	Probabilidad de selección(%)	Ponderaciones
1	2,1090	47,4167
2	3,4712	28,8083
3	2,6667	37,5000
4	9,5238	10,5000



www.leondariobello.com
www.olemonline.info/blog

Ahora, el tema relevante es el cambio en las probabilidades de selección, se observa que al definir una muestra igual para cada estrato, 120, para estratos de mayor tamaño, la probabilidad de inclusión es menor, lo que implica una mayor ponderación, esto es entendible, dado que, debo seleccionar menos unidades proporcionalmente que en los estratos bajos.

Otra posibilidad es repartir proporcional y entrar los datos de manera manual, tal como se presenta a continuación:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

EPIDAT 4.0

Muestreo aleatorio estratificado

Entrada manual: Número de estratos: 4

Entrada automática: ☒ Abrir datos individuales ☐ Abrir datos agregados

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

Estrato	Tamaño del estrato	Tamaño de la muestra
1	2500	163
2	3457	111
3	4500	145
4	1260	41

Tamaño de la muestra: 450

Datos: ☐ Ordenar la muestra

Presentar resultados: ☒ En pantalla ☐ En fichero

Estrato	Probabilidad de selección(%)	Ponderaciones
1	3,2162	31,0929
2	3,2109	31,1441
3	3,2222	31,0345
4	3,2540	30,7317

Este caso se mencionó antes, sin embargo, es útil mencionar que las probabilidades de selección o inclusión, son muy similares, lo que logra hacer cumplir la característica de un diseño equiprobabilísticos y por lo tanto, el cálculo de los estimadores no tendría mayores dificultades.

Por último se muestra la opción cuando los tamaños de muestra son definidos por los investigadores de manera diferente en cada estrato, tal como se observa a continuación.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

EPIDAT 4.0

Muestreo aleatorio estratificado

Entrada manual: Número de estratos: 4

Entrada automática: ☒ Abrir datos individuales ☐ Abrir datos agregados

Reparto de la muestra: Definir tamaño muestral de cada estrato

Estrato	Tamaño del estrato	Tamaño de la muestra
1	5590	200
2	3457	150
3	4500	200
4	1260	100

Datos: ☐ Ordenar la muestra

Presentar resultados: ☒ En pantalla ☐ En fichero

Ocultar Calcular Limpiar Cerrar

Estrato	Probabilidad de selección(%)	Ponderaciones
1	3,5149	28,4500
2	5,2068	19,2056
3	4,4444	22,5000
4	9,5238	10,5000

Para finalizar esta parte, sugerimos observar el siguiente video [MUESTREO ESTRATIFICADO 2](#)

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 5

MUESTREO POR CONGLOMERADOS (MC)



Los conglomerados, diferente a los estratos, deben ser tan heterogéneos dentro de ellos como sea posible y un conglomerado debe ser muy similar a otro para que éste diseño sea adecuado.

Se utiliza generalmente un muestreo aleatorio para seleccionar cada unidad de muestreo la cual es una colección de elementos (conglomerados).

Se puede deducir que las unidades de análisis entregan información parecida dado su conformación natural, (hogares, viviendas, barrios etc.).

Se recomienda cuando la construcción de un marco de muestreo de elementos es muy difícil, costosa o imposible de conseguir o cuando la

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

población objetivo se encuentra muy dispersa o se encuentra en agrupaciones naturales (familias, escuelas, hospitales etc.)

5.1 Muestreo por Conglomerados una Sola Etapa o Monoetápico.

Se prefiere cuando el muestreo involucra zonas geográficas y/o las unidades de análisis se encuentran dispersas en áreas generalmente geográficas, por lo tanto, pretende ser más práctico que eficiente, en algunos casos es preferible incrementar los tamaños de muestra calculados con formulas del MAS, para ello, se incrementa algún porcentaje de dicho valor, o se asigna en muchos casos de manera subjetiva lo que se conoce como efecto del diseño.

DEFF o Efecto del diseño (design effect): Lo que se pretende con el deff, es compensar la pérdida de representatividad que se puede dar al utilizar un diseño en varias etapas o usando estrategias de conglomerados que por su diseño, seleccionan unidades que están muy cercanas unas de otras y por lo tanto, pueden estar relacionadas, llevando esto a perder precisión en las estimaciones con respecto al MAS, generalmente fluctúa entre 1.5 y 3.0, valga decir, se incrementa la muestra en un 50% más hasta triplicar dicha muestra, dependiendo del número de etapas que se requieran. Esta es una regla empírica, es decir, cada estudio debería tener su propia reflexión y si

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

la variable en estudio fuese poco cambiante, se podría incluso tener un deff más bajo que lo estipulado anteriormente.

Su cálculo operativo se realiza utilizando la siguiente fórmula:

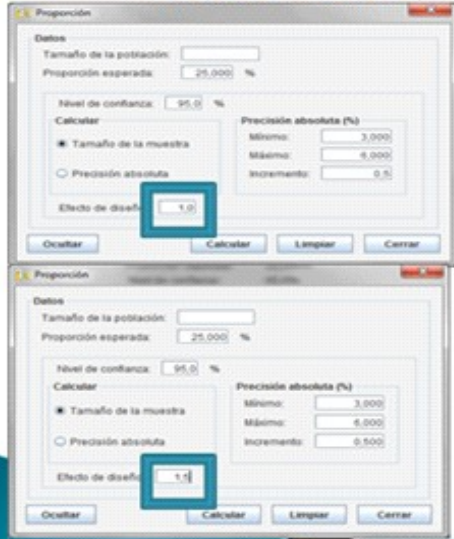
$$\text{Deff} = V_{\text{nuevo}} / V_{\text{mas}}$$

Entendiéndose como V_{nuevo} , la varianza del estimador para el nuevo muestreo (promedio o proporción) y V_{mas} , la varianza del estimador si se usara el MAS.

El deff es usualmente mayor de 1 e indica cuanta eficiencia se ha perdido al utilizar el nuevo diseño, sin embargo, en la práctica es supremamente escaso tener la forma de estimarlo, como siempre, se recurre a temas subjetivos de tomar otros estudios, donde en la generalidad tenían otros objetivos o se usó en otra población con características no tan semejantes a la del nuevo estudio. Por lo tanto, se realizan algunas especulaciones y al final, se decide por un deff con poco soporte teórico. Sin embargo, el efecto es relevante, observemos.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Efecto del Diseño. Epidat 4.0



Precisión (%)	Tamaño de la muestra
3,000	801
3,500	588
4,000	451
4,500	356
5,000	289
5,500	239
6,000	201

Precisión (%)	Tamaño de la muestra
3,000	1,201
3,500	882
4,000	676
4,500	534
5,000	433
5,500	358
6,000	301

www.leondariobello.com
www.ciemonline.info/blog

Detecte que un deff de 1.5, está indicando que se requiere un 50% más de unidades de análisis para lograr la misma eficiencia que si se utiliza el MAS.

Retomemos el procedimiento de diseño muestral, consiste en dividir la población en conglomerados o grupos de observaciones, con la característica de que sean homogéneos entre ellos, pero muy disimiles al interior. El caso podría ser cuando, se seleccionan predios en zonas rurales, donde se podrían formar estratos (Grandes, mediano, pequeños) pero sería impropio y poco práctico realizar este muestreo, ya que las extensiones por recorrer para llegar a las unidades seleccionadas son muy grandes y por ende los costos. Es preferible formar grupos delimitados por referentes geográficos y realizar o bien censo o bien muestreo sistemático o

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

aleatorio dentro de la zona seleccionada.

El Epidat 4.0, selecciona los conglomerados usando MAS, donde la probabilidad de selección es m/M . M total de conglomerados y m número de conglomerados a seleccionar, por lo tanto, es un diseño equiprobabilístico.

Se tiene dos posibilidades en el Epidat 4.0, así:

a. Entrada manual: Se conoce el número de conglomerados y sus tamaños, se puede digitar la información o leer de un archivo Excel. Es importante mencionar que el número de conglomerados debe ser aproximadamente igual a la razón entre el tamaño de muestra y el número medio de unidades de análisis por conglomerado.

b. Entrada automática: se requiere como insumo de un archivo generalmente en formato Excel, con el listado de todas las unidades de análisis incluyendo una variable con los conglomerados, en este caso, de una sólo etapa, se debe realizar censo en los conglomerados seleccionados de forma aleatoria.

Luego de abrir el Epidat 4.0 y entrar por **Módulos + Muestreo + selección de muestras + Muestreo por conglomerados monoetápico** se llega a la siguiente caja de dialogo.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo por conglomerados monoetápico

Entrada manual
Número de conglomerados: 6

Entrada automática
☒ Abrir datos individuales
☐ Abrir datos agregados

Tabla de datos: Ver

Datos
☐ Tamaño de la muestra: 0
☒ Número de conglomerados de la muestra: 2
☐ Ordenar muestra

Presentar resultados
☒ En pantalla
☒ En fichero

Ocultar Calcular Limpiar Cerrar

En la entrada manual, se define el número de conglomerados, para el ejemplo, se definieron 6, cuyos tamaños de muestra se pueden digitar uno a uno, como se muestra en el cuadro siguiente.

Conglomerado	Tamaño del conglomerado
1	1598
2	2456
3	789
4	1450
5	6789
6	5432

Aceptar Cancelar

Estas posibilidades se dan cuando se seleccionó la opción Tamaño de muestra, otra opción, es definir el número de conglomerados y luego realizar censo es ellos, no sobra recordar, que los resultados pueden solicitarse en

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

pantalla o creando un archivo con los conglomerados seleccionados. En este segundo caso, la salida en pantalla es la siguiente:

Resultados

[7] Muestreo por conglomerados monoetápico:

Datos:

	Población	Muestra
Tamaño	10514	12221
Nº de conglomerados	6	2

Conglomerados seleccionados:

	Tamaño
5	6769
6	5432
TOTAL	12221

Probabilidad de selección: 33,3333%

*Resultados guardados en: D:\(2011-02-17) RESPALDO\ENSAB IV\Ensalz\1\Muestraconglomeradomono.xls

[8] Muestreo por conglomerados monoetápico:

Datos:

	Población	Muestra
Tamaño	10514	7988
Nº de conglomerados	6	2

Conglomerados seleccionados:

	Tamaño
5	5432
2	2456
TOTAL	7988

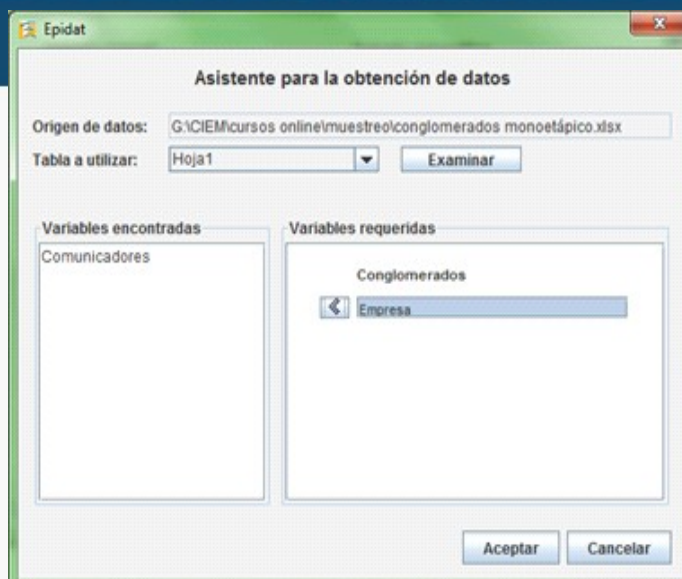
Probabilidad de selección: 33,3333%

*Resultados guardados en: D:\(2011-02-17) RESPALDO\ENSAB IV\Ensalz\1\Muestraconglomeradomono.xls

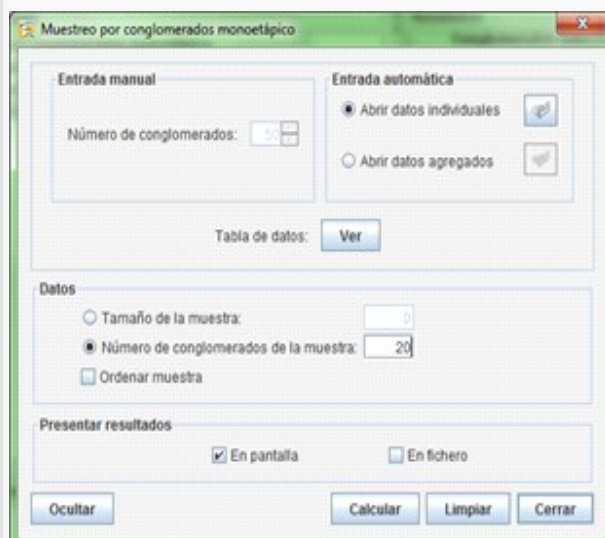
Cuando se tiene el marco muestral en un archivo, se procede de la siguiente manera:

Seleccionar la opción: Abrir datos individuales, el cual arroja la siguiente caja de dialogo.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



El programa lee la información del archivo y coloca no sólo el número de conglomerados sino su tamaño.



Conglomerado	Tamaño del conglomerado
1	5
2	6
3	4
4	2
5	8
6	4
7	2
8	5
9	6
10	4

[Aceptar] [Cancelar]

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Veamos otro ejemplo:

Conglomerado una sola etapa

	Población	Muestra
Tamaño	269	114
Nº de conglomerados	50	20

	Tamaño
28	5
41	6
5	4
47	8
50	4
16	8
44	6
21	6
15	6
46	3
37	9
20	5
6	8
3	6
19	8
10	5
39	2
27	4
24	9
14	2
TOTAL	114

Probabilidad 40%
www.leondariobello.com
www.olemonline.info/blog

Si bien es cierto este diseño como todos tiene su aplicación, es de mayor uso el de conglomerados Bietápico, entremos a analizar ese diseño.

5.2 Muestreo por Conglomerados Bietápico.

Se diferencia del anterior en que luego de seleccionar los conglomerados (UPE), se seleccionan las unidades de manera aleatoria (USE) dentro de cada uno de ellos. En el diseño anterior se realiza censo en los conglomerados seleccionados, mientras que ahora, se lleva a cabo una

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

segunda selección y así se podría extender el procedimiento, sin embargo, el Epidat 4.0 sólo tiene hasta esta opción.

De nuevo, los conglomerados pueden ser municipios, comunas (unión de barrios), sectores censales, manzanas, colegios, grupos de esos colegios, grupos de empresas etc. De estos conglomerados, se seleccionan otras unidades de análisis configurando las 2 etapas. Una ventaja importante, es que no se requiere todo el espacio muestral, sólo el de los conglomerados seleccionados.

Se aprecia que el programa entrega los números de los elementos seleccionados por cada conglomerado, para el ejemplo, se tienen 6 conglomerados, de los cuales se desean seleccionar 3 de ellos para un total de 180 unidades finales, se utilizó muestra igual para todos los conglomerados.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

The screenshot shows the 'Muestreo por conglomerados bietápico' dialog box. It is divided into several sections: 'Entrada manual' with a 'Número de conglomerados' field set to 6; 'Entrada automática' with radio buttons for 'Abrir datos individuales' (selected) and 'Abrir datos agregados'; 'Tamaño de muestra por conglomerado' with a dropdown menu set to 'Muestra igual para todos los conglomerados'; 'Tabla de datos' with a 'Ver' button; 'Datos' with checkboxes for 'Tamaño de la muestra' (180), 'Número de conglomerados de la muestra' (3), and 'Tamaño de muestra a seleccionar en cada conglomerado' (60), along with a dropdown for 'Seleccionar conglomerados usando' set to 'Muestreo simple aleatorio' and an unchecked 'Ordenar la muestra' checkbox; and 'Presentar resultados' with checkboxes for 'En pantalla' (checked) and 'En fichero'. At the bottom are buttons for 'Ocultar', 'Calcular', 'Limpiar', and 'Cerrar'.

La salida del Epidat es como se presenta a continuación:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Resultados

[1] Muestreo por conglomerados bietápico:

Datos:

Reparto de la muestra: Muestra igual para todos los conglomerados
 Seleccionar conglomerados usando: Muestreo simple aleatorio

	Población	Muestra
Tamaño	1197	180
Número de conglomerados	6	3

Conglomerados seleccionados:

Conglomerado	Tamaño	Muestra
5	87	60
1	125	60
4	125	60
Total:	337	180

Probabilidades de selección y ponderaciones:

Conglomerado	Probabilidad	Ponderación
5	34,4828	2,9000
1	24,0000	4,1667
4	24,0000	4,1667

Número de los sujetos seleccionados:

Conglomerado 5:

44	65	29	40	64	41	53
5	45	24	35	12	21	57
18	25	38	71	39	4	60

De cada conglomerado se seleccionaron 60 unidades con probabilidades diferentes dado la estrategia de tomar un mismo número de cada grupo. Se tienen las probabilidades de inclusión y su correspondiente ponderación.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Como complemento de lo ya mencionado, se coloca el enlace al video aclaratorio del tema, visite el punto [MUESTREO POR CONGLOMERADOS MONO Y BIETÁPICO](#).

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Page 95 of 115

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 6

MUESTREO POR CONGLOMERADOS MONOETÁPICO Y BIETÁPICO ESTRATIFICADO

Tienen gran aplicación y se usan cuando se quiere garantizar la obtención de datos de todas las categorías denominadas estratos.

6.1 Muestreo por conglomerados monoetápico estratificado.

Se recomienda cuando no se dispone de un listado de unidades de análisis y/o cuando se quiere minimizar el problema originado por la dispersión geográfica de las unidades de análisis.

El procedimiento consiste en dividir la población en estratos los cuales son UPE y de cada uno de ellos, se obtiene una muestra de conglomerados usando MSA.

Como en los casos anteriores el Epidat permite dos situaciones:

- Cuando se dispone con un archivo que contiene una base de datos donde cada uno de sus registros equivale a una unidad de análisis.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

- Cuando sólo se tienen identificados los conglomerados dentro de cada estrato y sus tamaños. En este caso, puede entrar los datos de manera manual o diseñar un archivo con esos datos, valga decir, un campo que identifique el conglomerado, otro el estrato y otro su tamaño.

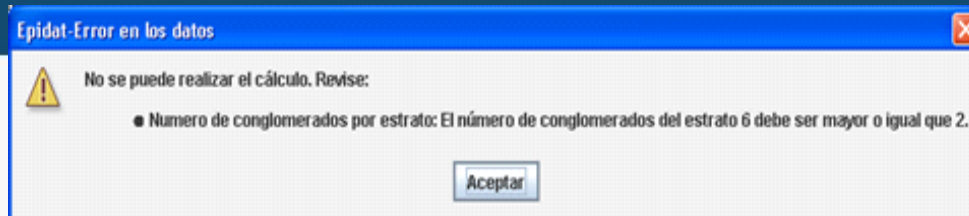
Es importante tener en cuenta que el número de conglomerados debe ser aproximadamente igual a la razón entre el tamaño de muestra y el número medio de unidades de análisis por conglomerado. En caso de no conocerse, se debe tener una idea de ellos.

En los tres casos el archivo de resultados incluye, además, la probabilidad de selección y sus respectivas ponderaciones.

Estrato	Conglomerado	Tamaño del conglomerado
1	1	0
1	2	0
2	1	0
2	2	0

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



6.2 Muestreo por conglomerados bietápico estratificado.

Se utiliza cuando en la primera etapa, se seleccionan los estratos, es decir, se conforman categorías excluyentes con la característica de que se toman unidades de muestreo de todas y cada una de esas categorías llamadas estratos, por esta razón, es muy utilizado en la práctica. Por ejemplo, si se requiere tomar información de todos los municipios o provincias de una región, entonces, los municipios o provincias son estratos. Otra de tantas aplicaciones, puede ser cuando se requiere seleccionar información de estudiantes de varias instituciones y es necesario según los objetivos llegar a cada una de ellas. Ya en una segunda etapa, se seleccionan los grupos (conglomerados) y se realiza censo en los grupos seleccionados.

En las ayudas del Epidat 4.0, encontramos la siguiente información con respecto a las posibilidades que entrega:

“a) Cuando sólo se dispone de la lista de conglomerados y sus tamaños dentro de cada estrato; en este caso, es posible introducir los datos de

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

forma manual o por medio de un archivo donde cada registro de la base de datos corresponda a un conglomerado. Uno de los campos de tal base debe recoger el número de unidades de análisis de la UPE correspondiente (es decir, su tamaño) y otro el estrato al que dicha UPE pertenece.

b) Cuando se cuenta con una base de datos donde cada uno de sus registros corresponde a una unidad de análisis (marco muestral detallado). En este caso es posible guardar la muestra resultante del procesamiento en un archivo de datos con la misma estructura del original y con todos los datos de cada una de las unidades de análisis seleccionadas.

Nota: Debe advertirse que, al examinar la base de datos, el sistema considera como pertenecientes a un mismo conglomerado a todos aquellos registros que tengan un mismo código en el campo que identifica al conglomerado dentro de cada estrato. Esto quiere decir, por ejemplo, que todos los registros para los cuales la variable correspondiente al conglomerado tenga el código 3, serán considerados como integrantes del tercer conglomerado de la población en el estrato correspondiente. Si la base de datos estuviera conformada de manera tal que se emplea un mismo número para identificar a conglomerados diferentes dentro de un estrato debido, por ejemplo, a que éstos

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

pertenecen a zonas diferentes o a distintas unidades de muestreo de mayor jerarquía, el sistema “entenderá” indebidamente que se trata de registros de un mismo conglomerado. Consecuentemente, para evitar esta posible dificultad, es conveniente atribuir siempre un código diferente a cada conglomerado de la población dentro de cada estrato. Por ejemplo, si se tuviera dos manzanas, una con 42 y otra con 23 viviendas (siendo éstas los conglomerados), en el campo correspondiente a estos últimos podrían usarse 65 números diferentes para identificarlos (por ejemplo, los números del 1 al 65) en lugar de emplear los números 1, 2, ..., 42 para la primera manzana y 1, 2, ..., 23 para la segunda”.

Adicionalmente, se puede seleccionar la opción de una muestra equiprobabilística o una muestra personalizada. El siguiente video permite reforzar lo planteado en las secciones precedentes en éste capítulo.

[MONOETAPICO ESTRATIFICADO](#)

Tenga en cuenta los siguientes comentarios al definir los valores que requiere el software.

Tamaño de la muestra: Valor entero mayor que cero y no superior al producto del número de conglomerados por el tamaño del conglomerado

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

más pequeño.

Número de conglomerados de la muestra: Valor entero mayor que cero y menor que el número de conglomerados (UPE) existentes.

Tamaño de muestra a seleccionar en cada conglomerado: Valor entero mayor que cero y menor o igual que el tamaño del conglomerado (UPE) más pequeño con que cuente la población; identifica el número de unidades de análisis que han de seleccionarse dentro de cada uno de los conglomerados.

Cuando se opta por seleccionar una muestra “personalizada”, Epidat 4.0 permite seleccionar las UPE con muestreo simple aleatorio o con probabilidad proporcional a sus tamaños dentro de cada estrato. Posteriormente, la selección de las unidades de análisis se realiza dentro de cada UPE elegida mediante muestreo simple aleatorio.

En cuanto al tamaño de muestra, hay que definir en primer lugar el número de unidades a seleccionar en cada conglomerado eligiendo entre 5 alternativas:

- Muestra igual para todos los conglomerados.
- Muestra igual para todos los conglomerados del mismo estrato.
- Fracción fija para todos los conglomerados.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

- Fracción fija para todos los conglomerados del mismo estrato.
- Definir tamaño muestral de cada conglomerado.

Además, se debe definir el número de conglomerados en la muestra y su distribución por estratos; Epidat 4.0 ofrece tres alternativas:

1. Muestra igual para todos los estratos.
2. Reparto proporcional al tamaño de los estratos.
3. Definir libremente el tamaño muestral de cada estrato.

Como es usual en éste software, se tiene la posibilidad de guardar la muestra en un archivo con formato de Excel (*.xls, *.xlsx) o de OpenOffice (*.ods). De las ayudas del programa se encuentra la siguiente redacción:

- “Entrada manual del número de estratos y conglomerados y los tamaños de éstos: el archivo de resultados contendrá el listado de números correspondientes a los conglomerados e individuos seleccionados en cada estrato en las dos etapas del proceso.
- Entrada automática de datos agregados: el archivo de resultados contendrá el listado de los conglomerados e individuos seleccionados en cada estrato y toda la información de cada conglomerado contenida en el archivo original.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

- Entrada automática de datos individuales: la información se salva en un archivo con la misma estructura del original y con todos los datos de cada una de las unidades de análisis seleccionadas.

En todos los casos el archivo de resultados incluye, además, la probabilidad de selección común a todas las unidades de un mismo conglomerado y la ponderación que les corresponde en la muestra, que es el inverso de la probabilidad de selección y se interpreta como el número de elementos de la población que representa cada unidad de la muestra”.

Observe el siguiente video e ilústrese sobre algunas posibilidades de las ya descritas, de clic en el siguiente enlace: [MUESTREO BIETAPICO ESTRATIFICADO POR CONGLOMERADOS](#).

Ahora entremos a comentar sobre la manera de realizar las estimaciones según el tipo de diseño aplicado.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Capítulo 7

ESTIMACIÓN CON MUESTRAS COMPLEJAS

Como ya se ha mencionado, una muestra compleja tiene en general el siguiente esquema:

- Una población inicial dividida en L estratos.
- Cada estrato se divide en conglomerados y toman el nombre de unidades de primera etapa (UPE).
- De cada una de las UPE se selecciona un número de unidades de análisis usando una o más etapas de selección.

La manera de seleccionar las diferentes unidades en cada una de las etapas se puede llevar a cabo de varias formas, valga decir, mediante MAS o MS, con probabilidades proporcionales al tamaño de las unidades de muestreo, con número fijo para cada estrato o conglomerado, a medida que se incrementan las etapas, el número de combinaciones posibles para llegar a un n predefinido son múltiples, por lo tanto, las posibles soluciones son igualmente múltiples, sin embargo, si bien el la subjetividad sustentada juega un papel preponderante en esa combinación, también es cierto que se puede simular diferentes opciones válidas y se estiman las varianzas de los

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

estimadores, sugiriendo la combinación con menor variabilidad. La combinación final para obtener las unidades de análisis, incide en el cálculo de las probabilidades de selección y por ende en sus respectivas ponderaciones.

7.1 Probabilidad de selección y ponderación.

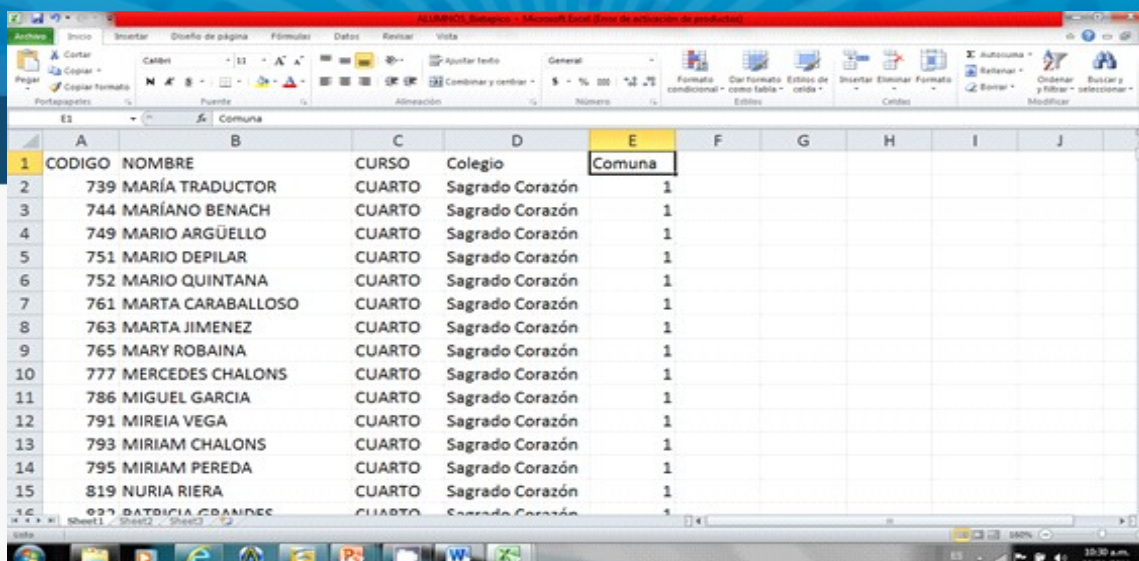
Las probabilidades de inclusión son iguales al producto de las probabilidades de selección en las distintas etapas. Con ellas, se calcula el inverso de esa probabilidad y toma el nombre de ponderación, y se interpreta como el número de individuos de la población que representa uno de la muestra.

Las posibilidades son múltiples, por lo tanto, se plantea la explicación siguiendo un ejemplo y observando el video correspondiente al tema.

Ejemplo.

Supongamos que se tiene el archivo Alumnos_Bietapico.xlsx que contiene las siguientes columnas:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



	A	B	C	D	E	F	G	H	I	J
	CODIGO	NOMBRE	CURSO	Colegio	Comuna					
2	739	MARÍA TRADUCTOR	CUARTO	Sagrado Corazón	1					
3	744	MARIANO BENACH	CUARTO	Sagrado Corazón	1					
4	749	MARIO ARGÜELLO	CUARTO	Sagrado Corazón	1					
5	751	MARIO DEPILAR	CUARTO	Sagrado Corazón	1					
6	752	MARIO QUINTANA	CUARTO	Sagrado Corazón	1					
7	761	MARTA CARABALLOSO	CUARTO	Sagrado Corazón	1					
8	763	MARTA JIMENEZ	CUARTO	Sagrado Corazón	1					
9	765	MARY ROBAINA	CUARTO	Sagrado Corazón	1					
10	777	MERCEDES CHALONS	CUARTO	Sagrado Corazón	1					
11	786	MIGUEL GARCIA	CUARTO	Sagrado Corazón	1					
12	791	MIREIA VEGA	CUARTO	Sagrado Corazón	1					
13	793	MIRIAM CHALONS	CUARTO	Sagrado Corazón	1					
14	795	MIRIAM PEREDA	CUARTO	Sagrado Corazón	1					
15	819	NURIA RIERA	CUARTO	Sagrado Corazón	1					
16	833	PATRICIA GRANDES	CUARTO	Sagrado Corazón	1					

Donde el estrato es la variable **Colegio** y el conglomerado la columna **curso**, esto implica que se desea visitar a todos los colegios y de ellos, se seleccionarán al azar algunos grupos. El total de estudiantes es de 1599.

Luego de varias reflexiones por el equipo de trabajo del estudio, se definió tomar 360 estudiantes, repartidos en 20 por grupo, para esta decisión ayudo considerablemente, el hecho de que con ésta combinación se obtuvo una varianza estimada pequeña y adecuada para realizar estimaciones, se ensayaron diversos valores y se validaba su pertinencia económica, además, de la precisión estadística.

La pantalla del Epidat 4.0 para el procedimiento equiprobabilístico y estratificado por conglomerados bietápico es la siguiente:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Muestreo por conglomerados bietápico estratificado

Tipo de muestreo: ☐ Equiprobabilístico ☒ Reparto de la muestra proporcional al tamaño de los estratos. ☒ Selección de conglomerados con probabilidad proporcional a su tamaño.

Entrada manual: Número de estratos: Ver

Entrada automática: ☒ Abrir datos individuales ☐ Abrir datos agregados

Tabla de datos: Ver

Datos

Escoger dos opciones:

☒ Tamaño de la muestra: 360

☐ Número de conglomerados de la muestra: 10

☒ Tamaño de muestra a seleccionar en cada conglomerado: 20

☐ Ordenar la muestra

Presentar resultados: ☒ En pantalla ☒ En fichero

Ocultar Calcular Limpiar Cerrar

Tabla de datos

Estrato	Conglomerados	Tamaño del conglomerado
Sagrado Corazón	SEGUNDO	67
Sagrado Corazón	TERCERO	79
San Carlos	CUARTO	69
San Carlos	PRIMERO	54
San Carlos	QUINTO	135
San Carlos	SEGUNDO	50
San Carlos	TERCERO	70
San Lucas	CUARTO	43
San Lucas	PRIMERO	73
San Lucas	QUINTO	36

Aceptar Cancelar

Número de conglomerados por estrato

Estrato	Número de conglomerad...
Sagrado Corazón	5
San Carlos	5
San Lucas	5
San Pedro	5
San Rafael	5
Santa María	5

Aceptar Cancelar

Observando la información, se aprecia que cada estrato (colegio) tiene 5 conglomerados (grupos o cursos).

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

7.2 Posibilidades con Epidat 4.0.

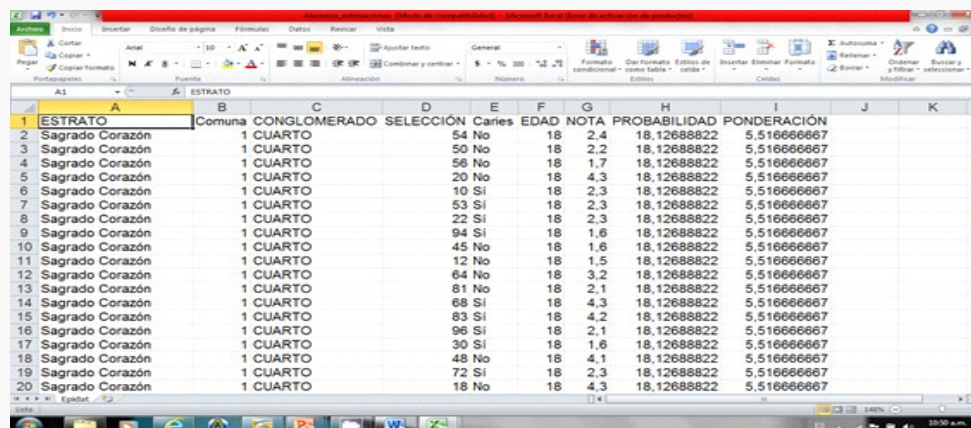
El Epidat 4.0 permite realizar estimaciones globales, es decir, usando la muestra total, pero además, es posible calcular las estimaciones según diferentes categorías de una variable (segmentar), generalmente cualitativa. Adicionalmente, es posible realizar "filtros", es decir, el análisis se concentra en una o varias categorías de alguna de las variables, por ejemplo, sólo se requiere las estimaciones para la zona 1 y 3, no es necesario para todas las zonas, aunque parezca poco práctico, esto se usa en algunas ocasiones y el software permite esta posibilidad.

Con el caso de los estudiantes se ilustrará la aplicabilidad de los casos mencionados anteriormente, anotando que estos procedimientos sólo aplican cuando la entrada de datos es por medio de un archivo Excel o OpenOffice (*.ods).

Ejemplo.

El nuevo archivo origina las siguientes variables o columnas:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0



ESTRATO	Comuna	CONGLOMERADO	SELECCIÓN	Caries	EDAD	NOTA	PROBABILIDAD	PONDERACIÓN
1	Sagrado Corazón	1 CUARTO	54	No	18	2.4	18,12688822	5,51666667
2	Sagrado Corazón	1 CUARTO	50	No	18	2.2	18,12688822	5,51666667
3	Sagrado Corazón	1 CUARTO	56	No	18	1.7	18,12688822	5,51666667
4	Sagrado Corazón	1 CUARTO	20	No	18	4.3	18,12688822	5,51666667
5	Sagrado Corazón	1 CUARTO	10	Si	18	2.3	18,12688822	5,51666667
6	Sagrado Corazón	1 CUARTO	53	Si	18	2.3	18,12688822	5,51666667
7	Sagrado Corazón	1 CUARTO	22	Si	18	2.3	18,12688822	5,51666667
8	Sagrado Corazón	1 CUARTO	94	Si	18	1.6	18,12688822	5,51666667
9	Sagrado Corazón	1 CUARTO	45	No	18	1.6	18,12688822	5,51666667
10	Sagrado Corazón	1 CUARTO	12	No	18	1.5	18,12688822	5,51666667
11	Sagrado Corazón	1 CUARTO	64	No	18	3.2	18,12688822	5,51666667
12	Sagrado Corazón	1 CUARTO	81	No	18	2.1	18,12688822	5,51666667
13	Sagrado Corazón	1 CUARTO	68	Si	18	4.3	18,12688822	5,51666667
14	Sagrado Corazón	1 CUARTO	83	Si	18	4.2	18,12688822	5,51666667
15	Sagrado Corazón	1 CUARTO	96	Si	18	2.1	18,12688822	5,51666667
16	Sagrado Corazón	1 CUARTO	30	Si	18	1.6	18,12688822	5,51666667
17	Sagrado Corazón	1 CUARTO	48	No	18	4.1	18,12688822	5,51666667
18	Sagrado Corazón	1 CUARTO	72	Si	18	2.3	18,12688822	5,51666667
19	Sagrado Corazón	1 CUARTO	18	No	18	4.3	18,12688822	5,51666667
20	Sagrado Corazón	1 CUARTO						

Claro, después de tomar la información de las personas seleccionadas, es decir, se asume que transcurrió un tiempo para lograr construir el archivo que se aprecia en la figura anterior.

Se destaca las columnas de probabilidad y ponderación, ya que con ellas, se ajusta el cálculo de las estimaciones de los indicadores.

Veamos cómo se procede para estas estimaciones:

1. Abrir el software Epidat 4.0
2. Seguir la ruta de **Módulos + Muestreo + Estimación** con

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

muestras complejas.

3. Abrir el archivo Alumnos_estimaciones.xlsx, para ello, observemos la caja de dialogo del Programa.

The screenshot shows the 'Asistente para la obtención de datos' (Data Acquisition Assistant) window in Epidat 4.0. The window has a title bar with the 'Epidat' logo and standard window controls. The main area is divided into two panes: 'Variables encontradas' (Variables found) on the left and 'Variables requeridas' (Variables required) on the right. The 'Origen de datos' (Data source) is set to 'C:\Leon Dario\CIEM\cursos online\muestreo\Alumnos_estimaciones.xls'. The 'Tabla a utilizar' (Table to use) is set to 'Epidat'. The 'Variables encontradas' pane lists 'SELECCIÓN' and 'PROBABILIDAD'. The 'Variables requeridas' pane lists several variables with checkboxes: 'Estratos(Opcional)' (checked), 'Conglomerados(Opcional)' (checked), 'Ponderacion(Opcional)' (checked), 'Estimar medias*' (checked), 'Estimar proporciones*' (checked), and 'Segmentar resultados (Opcional)' (checked). The 'Segmentar resultados (Opcional)' section has a dropdown menu set to 'Comuna'. A note at the bottom states: '* Nota: Al menos uno de los campos con * deberá tener asignado valores.' (Note: At least one of the fields with * must have values assigned). The window has 'Aceptar' (Accept) and 'Cancelar' (Cancel) buttons at the bottom right.

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Es evidente que algunos campos son opcionales y dependerá de la información disponible y del diseño muestral empleado. Para el ejemplo, se destaca que se quiere estimar el promedio para dos variables cuantitativas (edad, nota) y una proporción para la variable cualitativa caries. Si se tiene en el archivo la columna ponderación, es prudente colocarla. Hasta acá son las órdenes para estimaciones globales, sin embargo, si se requiere tener estimaciones para algunas categorías de una variable, para el ejemplo por comuna, se debe dar la variable que segmenta, es decir, Comuna.

Resultados						
Filtro: No						
Resultados para Comuna=1:						
Medias:						
Variable	N	Media	Media ponderada	E.E.	IC (95,0%)	Deff
EDAD	140	17,493	17,479	0,363	16,768 18,190	3,009
NOTA	140	2,766	2,762	0,085	2,595 2,930	0,959
Proporciones para la variable Caries:						
Categoría	N	%	% ponderado	EE (%)	IC (95,0%)	Deff
No	140	37,857	38,336	3,868	30,754 45,917	0,940
Si	140	62,143	61,664	3,868	54,083 69,246	0,940
Resultados para Comuna=2:						
Medias:						
Variable	N	Media	Media ponderada	E.E.	IC (95,0%)	Deff
EDAD	160	16,731	16,795	0,264	16,279 17,312	2,187
NOTA	160	2,784	2,770	0,066	2,641 2,898	0,849
Proporciones para la variable Caries:						
Categoría	N	%	% ponderado	EE (%)	IC (95,0%)	Deff
No	160	36,250	36,297	3,578	29,284 43,309	0,939
Si	160	63,750	63,703	3,578	56,691 70,716	0,939
Resultados para Comuna=3:						
Medias:						
Variable	N	Media	Media ponderada	E.E.	IC (95,0%)	Deff

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Observe que los resultados entregados se hacen según comuna, valga decir, en la comuna 1, el porcentaje de personas sin caries, se encuentra entre el 30,74 y el 45,91%, mientras que en la comuna 2, ese porcentaje está entre el 29,28 y el 43,30%.

Observemos los resultados de la comuna 3, que no se alcanza a apreciar en la salida anterior.

Categoría	N	%	% ponderado	EE (%)	IC (95,0%)		Deff
No	60	26,66	26,667	6,667	13,6	39,733	1,158
Sí	60	73,33	73,333	6,667	60,2	86,400	1,158

Tenga en cuenta el tamaño del intervalo, mucho mayor que los anteriores, debido al $EE = 6,667$, originado seguramente por el tamaño de muestra de esa comuna.

Realice las reflexiones del caso para las variables cuantitativas.

La otra opción es cuando interese realizar estimaciones de algunas categorías no de todas, por ejemplo:

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

Categoría	N	%	% ponderado	EE (%)	IC (95,0%)		Deff
No	140	37,857	38,336	3,86	30,754	45,917	0,940
Sí	140	62,143	61,664	3,86	54,083	69,246	0,940

Si selecciona la opción Filtro, tiene la posibilidad de seleccionar algunas categorías de algunas variables, incluso de varias, para el caso, sólo seleccioné la comuna 1, obteniendo lo siguiente:

Variable	N	Media	Media	E.E.	IC (95,0%)	Deff
----------	---	-------	-------	------	------------	------

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

©copyright 2011 www.leondariobello.com • All Rights Reserved Worldwide

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

			ponderada				
EDAD	140	17,493	17,479	0,363	16,76	18,19	3,009
NOTA	140	2,766	2,762	0,085	2,595	2,930	0,959

Categoría	N	%	% ponderado	EE (%)	IC (95,0%)		Deff
No	140	37,857	38,336	3,868	30,754	45,917	0,940
Sí	140	62,143	61,664	3,868	54,083	69,246	0,940

Conclusión: Cuando se trata de estimar parámetros a partir de un muestreo polietápico, siempre tocará seleccionar una muestra de elementos, y por lo tanto, la clave está en calcular las probabilidades de Inclusión y con ellas, las respectivas ponderaciones.

7.3 Procedimiento de estimación de parámetros.

Dado que es natural esperar diferencias en las probabilidades finales de selección de las personas bajo estudio, se requiere ponderar todos los

Muestreo en estudios descriptivos polietápicos. Epidat 4.0

valores muestrales por el recíproco de su respectiva probabilidad final, denominado el factor básico de expansión, con el fin de restituir los valores del universo estudiado.

En los diseños complejos, es fundamental, calcular las probabilidades de inclusión en cada una de las etapas. El programa SPSS^[2], emplea los estimadores de primer orden de: Horvitz-Thompson (totales) y Hajek (medias, proporciones).

Thompson

$$\hat{T}_{HT}(Y) = \sum_{i \in m} \frac{y_i}{\pi_i} = \sum_{i \in m} w_i y_i \quad w_i = \frac{1}{\pi_i} \quad \forall i \in m$$

Hajek

$$\hat{T}_{HJ}(Y) = \frac{\sum_{i \in m} w_i y_i}{\sum_{i \in m} w_i}$$

Donde y_i es la variable en estudio y W_i los respectivos pesos.

[\[1\]](#) Tomado de las ayudas del Epidat 4.0

[\[2\]](#) Curso de SPSS: Modulo de muestras complejas, I.E.A. Mayor José A. Mayo del 2009.