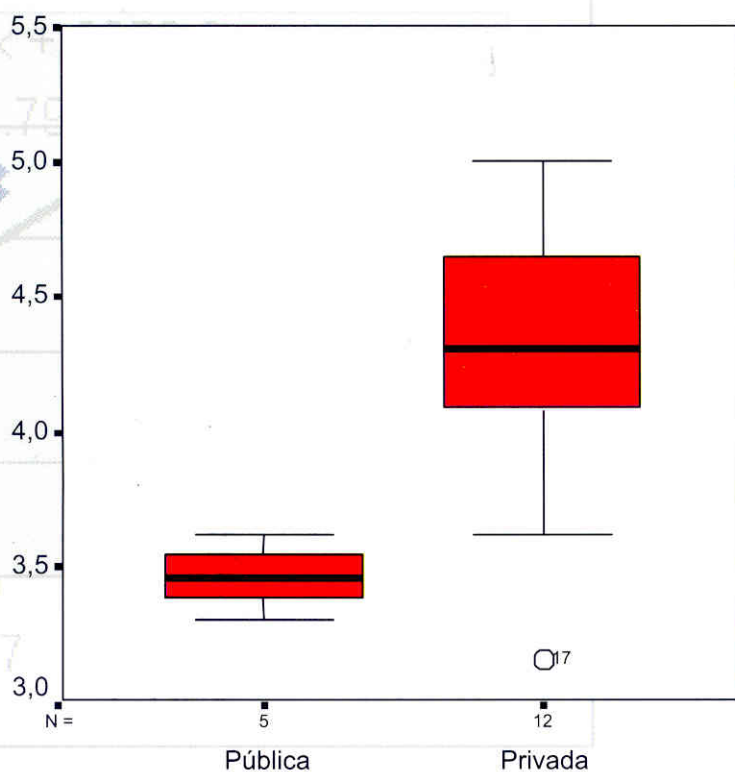


ESTADÍSTICA

Como apoyo a la investigación



Naturaleza jurídica

León Darío Bello Parias

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 Parías

MUESTREO: UTILIZACIÓN EPIDAT 3.0

El definir el tamaño muestral y el método de selección de las unidades de análisis, es tal vez el punto más crucial en el proceso investigativo, ya que dependiendo de éste, será posible generar inferencias o generalizaciones a toda la población, y por ende, definir políticas y tomar acciones que impacten el entorno del objeto de estudio.

Es evidente que el reducir costos en todas las actividades que se realizan en la cotidianidad es una prioridad, la investigación no es la excepción, por lo tanto, es menester en lo posible trabajar con una parte de la población objeto de estudio, en lugar de realizar un censo. No obstante, se presentan algunos casos donde es imprescindible medir o auscultar a todos los elementos del universo bajo estudio. Este capítulo pretende dilucidar y presentar alternativas de cuando se requiere seleccionar una muestra y cuando no, además, de presentar una posición sobre lo que en realidad es viable y como se pueden minimizar riesgos al momento de optar por un determinado procedimiento.

Si la situación fuera meramente del número de observaciones a tomar, la discusión se reduciría a utilizar algunos programas computacionales como: Epiinfo y Epidat (libre distribución) entre muchos otros, la discusión debe ir más allá. Lo que se busca entonces, es presentar algunas experiencias del autor en la asesoría de diversos estudios, así como presentar los procedimientos de muestreo más usuales y utilizar el programa Epidat 3.0 para realizar los cálculos.

El Epidat 3.0 es un software de libre distribución, por lo tanto, puede ser bajado por la red Internet, el cual ha sido diseñado e impulsado por dos instituciones, XUNTA DE GALICIA – Consellería de Sanidade – Dirección Xeral de Saúde Publica y la Organización Panamericana de la Salud (OPS/OMS). Es un programa elaborado para datos tabulados haciendo énfasis en el análisis Epidemiológico, no obstante, es útil y aplicable a cualquier área del conocimiento que requiere de procedimientos de inferencia y muestreo. Como se verá es un software de fácil manejo y con ayudas bien elaboradas, no sobra entonces, reconocer el aporte de estas instituciones al desarrollo y promulgación de la estadística. No obstante, se reconoce por dichas organizaciones, algunos errores encontrados en unos pocos procedimientos. De los utilizados en éste texto, se tiene: Proporción una población, se detectó error en el IC cuando el número de casos es igual al tamaño de muestra (caso muy poco común). Dos poblaciones, medias independientes. Se presenta error en el cálculo del valor p para la prueba de comparación de varianzas.

7.1 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 sujetos 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 de 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 cruzado por los objetivos de la investigación, una cosa es si se quiere estimar un parámetro poblacional y otra si el objetivo es determinar diferencias significativas entre tratamientos. Además, las poblaciones son en lo general diferentes y por lo tanto, deberán utilizarse criterios distintos para seleccionar las unidades bajo estudio.

La siguiente diapositiva presenta los elementos claves que se deben manejar para lograr diseñar un muestreo adecuado.

DEFINICIONES BASICAS

POBLACION	MARCO MUESTRAL
CENSO	UNIDAD MUESTRAL
MUESTRA	ELEMENTO MUESTRAL
MUESTREO	UNIDAD DE ANALISIS

Población hace referencia a todas las unidades objeto de estudio, puede ser finita o infinita, además, no tiene que ser necesariamente personas. Ejemplo de población puede ser: Todas las amas de casa del Municipio de Andes, todos los estudiantes de la U de A, los sacos de café producidos en el año 2003, el número de colectivos que transportan personas en el suroeste antioqueño, etc.

Si se estudia toda la población, se dice que se realizó un censo, por lo tanto, se puede definir censo, como el recuento de toda la población bajo estudio. Es claro que no siempre se puede realizar, debido a múltiples causas, entre ellas, los costos altos, 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. Lo anterior, implica que en la práctica se prefiera el seleccionar una parte de la población, llamada muestra, para realizar los análisis respectivos. No obstante, en algunos casos se prefiere el censo a la muestra, por ejemplo, cuando la población no es muy grande y a su vez es muy variable, además, de que se pueda recoger la información.

El procedimiento que orienta el proceso de recolección de datos, se conoce como muestreo, es pues la herramienta estadística que sugiere: Como, cuando y donde se obtienen los datos, además, cuanto es el número adecuado a tomar para luego ser procesados y generalizados a toda la población.

En algunos casos, se requiere de un marco muestral, el cual es un conjunto de listas de todas las unidades de muestreo, estas listas, también pueden ser mapas u otras formas que sirvan de base para determinar unidades de muestreo. Para ilustrar lo anterior, se puede pensar que para seleccionar niños en un municipio, se requiere: 1. una lista de escuelas, 2. una lista para cada uno de los grupos que tiene cada escuela seleccionada, y 3. una 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.

Es útil profundizar en los últimos conceptos: La 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 es 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 pueden operar personas, pero el análisis se hará sobre el corazón o el hígado, esto último se conoce como unidad de análisis.

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.

PASOS PARA SELECCIONAR UNA MUESTRA

Para definir la población objetivo, se debe tener claridad absoluta sobre el objetivo general del estudio, ya que éste será el guía para todos los pasos posteriores.

Definir la población objetivo

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.

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

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

- Marco muestral: 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).

2. Seleccionar un procedimiento de muestreo.
3. Definir el tamaño de la muestra (Según el tipo de muestreo la Estadística proporciona fórmulas para calcular el tamaño muestral)
4. Seleccionar las unidades muestrales.

Ejemplo: Para la encuesta de hogares realizada por el DANE en Medellín en el año 2004, mayo, con el propósito de estimar el IPC (Índice de Precios al Consumidor), determine los pasos para seleccionar la muestra.

Solución

Objetivo: Determinar el IPC en Medellín para el mes de mayo año 2004.

1. Población objetivo: Personas mayores de 18 años en Medellín (el DANE lo considera así).
 - Elemento muestral: El jefe de hogar.
 - Unidad muestral: El hogar
 - Alcance: Medellín.
 - Tiempo: 2003
 - Marco muestral: Mapa cartográfico de la ciudad de Medellín.

2. Seleccionar un procedimiento. Muestro Estratificado
3. Definir el tamaño de la muestra. Se aplican las fórmulas del muestreo estratificado
4. Seleccionar las unidades muestrales. Usar un método aleatorio.

En los siguientes casos identifique: La población objeto de estudio, el marco muestral, la unidad muestral y el elemento:

- a) Realizar un sondeo de opinión en Medellín sobre el proceso de paz en Colombia
- b) Identificar la percepción que tiene la ciudadanía mayor de 18 años sobre la violencia intra familiar.

ten diversas técnicas de muestreo, tales como:

PROBABILISTICOS	PROBABILISTICOS
Conveniencia	Aleatorio Simple
Juicio u Opinión.	Estratificado
otas o Prorratio.	Conglomerados
Fines Especiales.	Sistemático en Fases.
Bola de Nieve	Conglomerados Monoetápico
	Conglomerados Bietapico
	glomerados Bietapico Estratificado

Además, se utilizan otros términos para definir diferentes muestreos, en su mayoría están recogidos en la ya mencionados, valga decir: Dis-
cional, sujetos voluntarios, a criterio entre otros.

MUESTREOS PROBABILISTICOS

Tienen la característica que todas las unidades poseen una probabili-
conocida y mayor de cero de ser seleccionadas. Pretenden disminuir
objetividad originada al seleccionar las unidades de análisis, para evitar
gos en la selección de las mismas. Además, permiten medir el error en
estimación, situación que no se da en los muestreos no probabilísticos.

muestreos más usuales son: Muestreo simple aleatorio (MAS),
muestreo sistemático en fases (MSF), Muestreo aleatorio estratificado
(E) y Muestreo por conglomerados (MC).

7.3.1 Muestreo Aleatorio Simple (MAS). Es el procedimiento
abilístico de selección de muestras más sencillo y conocido, no obs-
e, en la práctica es difícil de realizar debido a que requiere de un marco
stral y en muchos casos no es posible obtenerlo. Puede ser útil cuan-
as poblaciones son pequeñas y por lo tanto, se cuenta con listados.
ndo las poblaciones son grandes, se prefiere el muestreo en etapas.
utiliza ampliamente en los estudios experimentales, además, de ser un
cedimiento básico como componente de métodos más complejos
muestreo estratificado y en etapas).

Se caracteriza por que otorga la misma probabilidad de ser elegidos a
os los elementos de la población. Para el calculo muestral, se requiere
El tamaño poblacional, si ésta es finita, del error admisible y de la esti-
sión de la varianza.

La siguiente figura muestra las formulas a utilizar en éste muestreo

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

$$\rightarrow \frac{2}{\alpha/2}$$

$$\frac{n_0}{1 + \frac{n_0}{N}}$$

n_0 se utiliza para población

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

normal, depende de la confiabilidad del estudio.

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

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

n : Probabilidad de éxito en variables binomiales

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

7.3.2 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 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

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)	

Es importante ahondar en dos conceptos importantes: primero, el Error estándar de estimación el cual es la diferencia entre el verdadero valor de la población (parámetro) y el obtenido en la muestra (estadístico), por lo tanto, dice que tanto se aleja nuestra estimación del verdadero valor poblacional. El segundo término es el Nivel de Confianza, el cual mide la probabilidad de que la estimación realizada éste acorde con la realidad, de que se define formalmente como: La probabilidad de que el intervalo construido con base en una muestra contenga el verdadero valor del parámetro.

De otro lado, en la práctica es difícil conocer el verdadero tamaño de la población objeto del muestreo, a veces se tiene es 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*, si se realizaron dos estudios sobre mediciones antropométricas: uno en una ciudad 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 media 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 selecciona el 2% de la población, se puede demostrar usando el Epidat que el error para una muestra de 1100 corresponde a 21.6%, mientras para la muestra de 10.000 el error sería de 6.6%.

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.

Otra situación a tener en cuenta, es determinar la variable que va a direccionar o aportar la variabilidad para calcular el tamaño de muestra, toda vez, 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 se desvíe 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 incrementar 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:

$$n_0 = \frac{Z_{\alpha/2}^2 * P * (1-P)}{e^2} = \frac{(1.96)^2 * 0.3 * 0.7}{(0.06)^2} = 224$$

$$n = \frac{n_0}{1 + \frac{n_0}{6.100}} = \frac{224}{1 + \frac{224}{6.100}} = 216$$

Donde

$Z^2_{\alpha/2}$ = Valor tabulado de la distribución normal cuando se tiene una confiabilidad del 95%.

P: Proporción de Empresas que poseen control interno en el área de mercadeo. El valor se estimó con la prueba piloto.

(1-P) : Proporción de Empresas que no poseen control interno en el área de mercadeo.

e: Error que se admite como viable y se mide en porcentaje, para el caso se considera adecuado el 6%.

Si se cuenta con el apoyo del programa Epidat 3.0, estos cálculos se simplifican, la ayuda del programa explica así la manera de calcular los 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 ingeren-

cia en determinar el error admisible, en el sentido que si la 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.

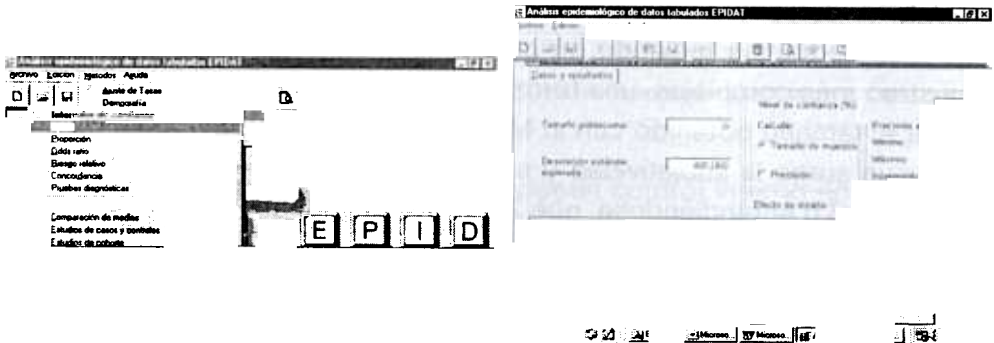
Como un último comentario, es mencionar que el MAS, lleva a muestras más pequeñas teniendo en cuenta la precisión que si se utilizarán muestreo más complejos, una forma de solucionar éste problema es multiplicar el tamaño obtenido con el MAS por el llamado **Efecto del Diseño**. De nuevo surge la subjetividad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

7.3.2.1 Tamaño de muestra para estimar un promedio poblacional, utilizando el Epidat 3.0. Con la versatilidad del programa, se observa cuales 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 aproximadamente normal con una media de 3550 gramos y una desviación estándar de $s=400$ 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 93%.

SOLUCIÓN EPIDAT 3.0.

Luego de abrir el programa, se activa la ruta: Métodos+Muestreo Tamaño de muestra+media. Observando los siguientes pantallazos:



Para un total de 260 pequeños. Con el fin de apreciar la ventaja comparativa de utilizar ésta herramienta, se presentan los siguientes resultados, donde, se utilizó la estrategia de definir un error máximo y uno mínimo, logrando concertar un tamaño de muestra práctico y económicamente viable.



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

De manera manual se procede así

$$n = \frac{\frac{\alpha/2}{e^2} * S^2}{1} = \frac{(1.81)^2 * (400)^2}{(45)^2} = 258.85$$

$$n = \frac{\frac{n}{n_0}}{N} = \frac{258.85}{150.000} = 258.4$$

Es claro entonces, que si el tamaño de la población es muy grande, ésta no tiene incidencia fuerte en la muestra, lo que si es determinante y se tratará de controlar es la variabilidad de los datos (S²). Para el caso, n=259

Donde:

$Z^2_{\alpha/2}$ = Percentil de la distribución normal cuando se tiene una confiabilidad del 93%. = 1.81

S: Desviación de la variable cuantitativa objeto de estudio. Usualmente se estima con una prueba piloto.

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 de 45 gramos.

7.3.2.2 Tamaño de muestra para estimar una proporción poblacional, utilizando el Epidat 3.0. En éste caso la ruta es: Métodos+Muestreo+Tamaño de muestra+proporción.



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 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.

De manera manual se procede así:

$$n_0 = \frac{Z^2_{\alpha/2} * P*(1-P)}{e^2} = \frac{(1.645)^2 * 0.25*0.75}{(0.04)^2} = 317.11$$

Se requieren 297 profesionales si se quiere un error del 4% en la estimación. Los resultados son coherentes con los entregados por el Epidat 3.0.

$$\frac{n_i}{1 + n_0} = \frac{317}{4.500} = 296.26$$

Donde

$Z^2_{\alpha/2}$ = Valor tabulado de la distribución normal cuando se tiene una confiabilidad del 90%. = 1.645

P: Proporción de profesionales que desean continuar sus estudios de posgrado. El valor se estimó con estudios previos.

(1-P): Proporción de profesionales que no desean continuar sus estudios de posgrado.

e: Error que se admite como viable y se mide en porcentaje, para el caso se consideró adecuado el 4%.

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 valida para sus propósitos.

En general es prudente seguir las premisas planteadas en la diapositiva siguiente, es decir, no es sólo aplicar una formula matemática, sino, determinar con precisión, la población a estudiar, tener en cuenta los estimadores a calcular, la forma de presentar el estudio y a que sub poblaciones se quiere generalizar los resultados y por último la precisión deseada.

PREMISAS EN EL DISEÑO DE MUESTRA

Indicadores a estimar

- ✓ Proporciones
- ✓ Promedios
- ✓ Métodos multivariados



7.3.4 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 cambia, así:

Calcular el tamaño de muestra con las formulas vistas.

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 .

Ejemplo 1: Se quiere obtener una muestra sistemática que seleccione egresados de un Programa de la Universidad de Antioquia que tiene 1200 de ellos. La variable clave del estudio es dicotómica y se aduce que la proporción es del 25%, además, se quiere un error del 4% y una confiabilidad del 90%.

Solución manual.

$$n_0 = \frac{Z^2_{\alpha/2} * P * (1-P)}{e^2} = \frac{(1.645)^2 * 0.25 * 0.75}{(0.04)^2} = 318$$

$$n = \frac{\frac{n_0}{1 + \frac{n_0}{N}}}{\frac{1200}{1200}} = \frac{318}{1.265} = 251$$

de:

$Z^2_{\alpha/2}$ = Valor tabulado de la distribución normal cuando se tiene una fiabilidad del 90%.

Proporción de egresados que cumplen la característica.

e: Error que se admite como viable y se mide en porcentaje, para el o se considera adecuado el 4%.

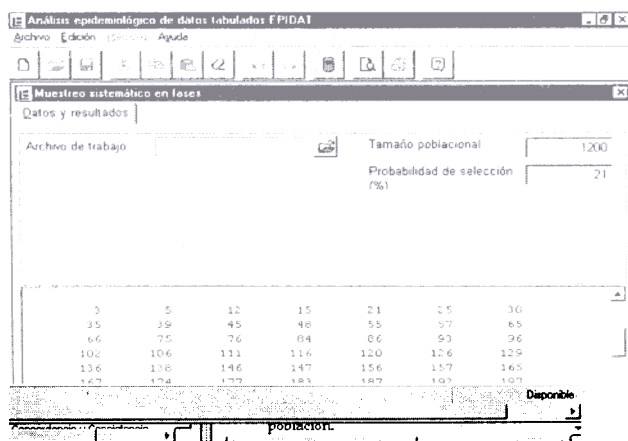
o se calcula el salto, $K = (1200/251) = 4.78$ se aproxima a 5.

Por último se procede a seleccionar al azar un número entre 1 y 5. Por lo tanto, los métodos son múltiples, por ejemplo, las calculadoras tienen la función RND que permite generar números al azar, para el ejemplo se usó un método trivial, que consistió en abrir un libro y el último número de la página fue el seleccionado, en caso de que fuese mayor de 5, se repite el ensayo. Al final dio el número 2, por lo tanto, los egresados seleccionados corresponden a los números:

1 2 para ajustar 251 egresados.

Uso de Epidat 3.0.

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



el procedimiento en fases, los números de las unidades seleccionadas no coinciden con el método manual.

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 3.0 para definir el salto y seleccionar los números sugeridos.

Habrás notado que el salto no es constante, lo que sucede es que el Epidat 3.0 utiliza una variante del muestreo sistemático convencional, llamado Sistemático en Fases, el cual consiste en: luego de determinar el intervalo, se aleatoriza el elemento a seleccionar, eso sí, dentro del intervalo definido con anterioridad (1 a k). Los datos que requiere son: el tamaño poblacional y el porcentaje de muestra (n/N). Permite además, si se tiene la base de datos en Excel, seleccionar dichos elementos, creando otra hoja con los elementos seleccionados.

7.3.5 Muestreo Aleatorio Estratificado (MAE). 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, seguramente, se disminuirá ésta y permitirá estimaciones más precisas. 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.

El método más usual es seleccionar dentro de cada estrato de manera aleatoria los elementos, repartiendo la muestra 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.

Para utilizar el Epidat 3.0, 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 muestrales, también en Excel.

Ejemplo 1: Con el fin de presentar la parte operativa, se contextualiza el ejemplo inicial en el tema de muestreo.

Una empresa tiene 6.100 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 5.600, seguridad 180 y otros cargos 90.

X: Número de empleados no satisfechos con el clima organizacional

$$P = 0.30 \quad q = 1 - 0.30 = 0.70 \quad e = 0.06$$

$$Z_{0.025} = 1.96$$

$$\frac{Z_{\alpha/2}^2 * P * (1-P)}{e^2} = \frac{(1.96)^2 * 0.3 * 0.7}{(0.06)^2} = 224$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{224}{1 + \frac{224}{6.100}} = 216$$

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:

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

Categoría	muestra	El número de unidades N_i	$n_i = n \cdot (N_i/N)$
			216*(3500/6100)= 123
			216*(1000/6100)= 35
			216*(800/6100)= 28
			216*(350/6100)= 12
			216*(280/6100)= 10
			216*(170/6100)= 6
			214

Por aproximaciones, el total de muestra es de 214, no obstante, es labor de los investigadores ajustar dichos valores, de tal manera que en las categorías con menos unidades se ajusten y se obtenga un número adecuado. Para el ejemplo, los dos faltantes se adicionaron a la clasificación de otros.

Ejemplo 2: Como ya se mencionó, es vital para un buen diseño de muestreo, tener claridad sobre lo que pretende la investigación. A continuación se colocan dos casos en los cuales el autor a diseñado el muestreo, no obstante, se ha cambiado el nombre de la organización para el cual se realizó el diseño. Se transcribe el objetivo general y parte de la metodología, para poder contextualizar el proceso de muestreo.

Objetivo General: Conocer la percepción de los usuarios de la Empresa ***** en los municipios donde del área metropolitana donde tiene injerencia

Tipo de estudio: Estudio de tipo cuantitativo, descriptivo concluyente de corte transversal.

Población Objetivo: El universo o población objetivo del estudio está conformado por 209.640 usuarios, los cuales están repartidos en siete municipios, a saber: Bello, Girardota y Copacabana del área norte del Valle de Aburra y Caldas, Itagui, Sabaneta y la Estrella del sector sur.

Muestra y muestreo: La Empresa ***** entregó información del número de usuarios tanto por municipio, como por estrato, sector e incluso tipo de usuario. Esta fue la base para definir el tipo de muestreo.

Se utilizó un muestro probabilístico estratificado por Municipio, sector y usuario. En total se obtuvo información de 1024 usuarios.

La fórmula utilizada fue la siguiente:

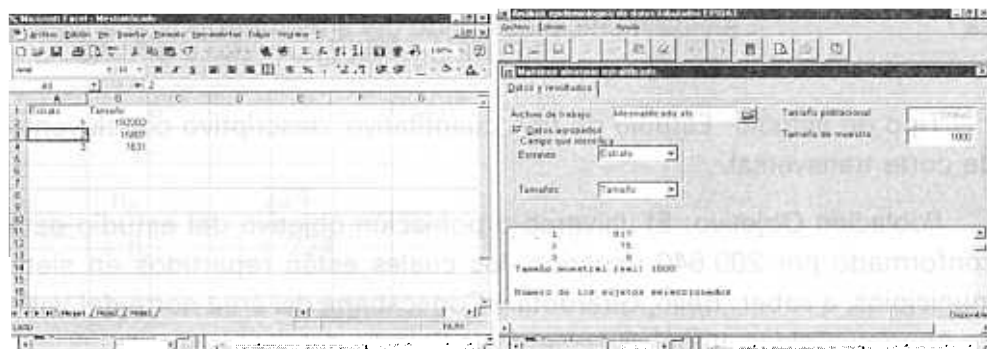
$$n = \frac{(1.96)^2 (0.5) * (0.5)}{(0.031)^2} = 1.000$$

Tabla 1: Distribución de la muestra según tipo de usuario, Diciembre ****.

Tipo Usuario	Población	Porcentaje	Muestra Estimada	Muestra Real
Residencial		0,917	917	937
Comercial/ Industrial		0,075	75	82
Grandes Productores		0,008	8	5

Realizando un muestreo probabilístico proporcional según usuario, se encontró que se requerían 917 encuestas del sector residencial, 75 del Comercial / Industrial y 8 de Grandes Productores, sin embargo, por diferentes razones al final se obtuvo información de 1024 usuarios distribuidos como se muestran en la última columna de la tabla anterior.

Solución Epidat 3.0: Luego de crear el archivo en Excel con la siguiente información, se procede a entrar la información requerida por el Epidat, que no es más que el nombre el archivo y los campos respectivos, valga decir, Estrato y Tamaño. Para obtener la siguiente salida.



Estos resultados coinciden con los presentados de manera manual.

Ejemplo 3:

OBJETIVO GENERAL: Determinar los gustos, preferencias y perfil Socio-Demográfico de los usuarios que asisten a + + + + + en las ciudades de Medellín y Bogotá, año 2003.

Muestreo y muestra: El número de unidades a estudiar fue de 720, en cada complejo se realizaron 240. El siguiente cuadro nos indica la muestra en cada municipio y Complejo.

LUGAR	MUESTRA	PROPORCIÓN
Bogotá	480	67%
Medellín	240	33%
TOTAL	720	100%

El método utilizado es un Muestreo estratificado proporcional según Municipio, para luego utilizar el muestreo sistemático, distribuyendo la muestra en tres días de la semana y en tres funciones cada día (repartidas en igual proporción en tres horarios), con el objetivo de lograr mayor representatividad. El esquema es el siguiente:

Miércoles: 80

Viernes: 80

Domingo: 80

En el nivel de confianza es del 95% y se estima un error del 5%

Se puede concluir, que éste 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.

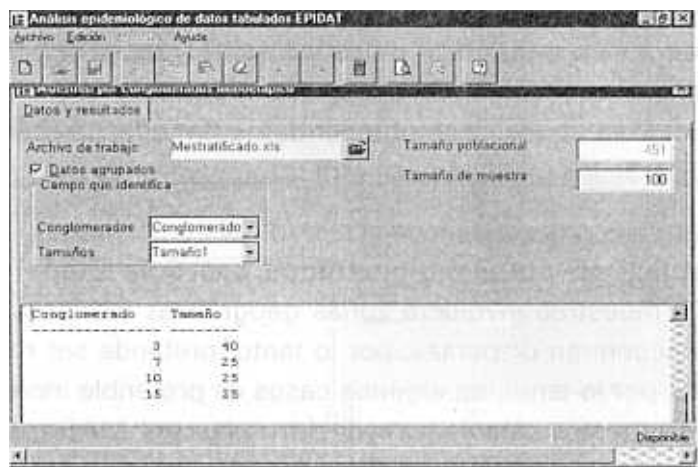
7.3.6 Muestreo por Conglomerados una Sola Etapa. Se prefiere cuando el muestreo involucra zonas geográficas y/o las unidades de análisis se encuentran dispersas, por lo tanto, pretende ser más práctico que eficiente, por lo tanto, en algunos casos es preferible incrementar los tamaños de muestra calculados 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, que 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.

El procedimiento 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

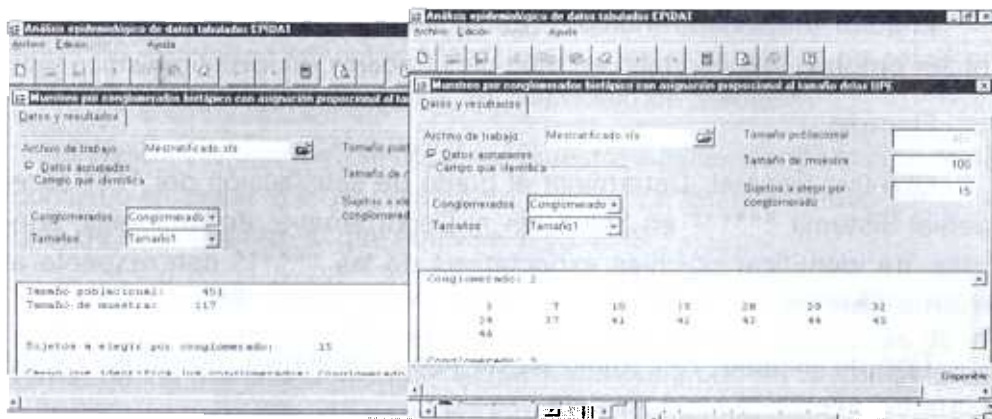
(Grandes, mediando, pequeños) pero sería impropio y poco práctico realizar éste muestreo, ya que las extensiones por recorres para llegar a las unidades seleccionadas son muy grandes y por ende los costos. Es preferible formar grupos delimitados por referentes geográficos y realizar o bien censo o bien muestreo sistemático dentro de la zona seleccionada aleatoriamente.

En algunos casos se consideran varias selecciones de conglomerados, es entonces, donde se presenta el muestreo por conglomerados multietápico. En estos casos, se presentan UPS (unidades primarias de selección), USS (unidades secundarias) y así sucesivamente.

Para utilizar el Epidat 3.0, se procede de manera similar al muestreo estratificado proporcional, es decir, como insumo se entrega un archivo con el número de conglomerados y a su vez el número de unidades de cada conglomerado, el objetivo es seleccionar al azar el conglomerado para luego analizar todas las unidades del conglomerado seleccionado. La salida es la siguiente:



7.3.7 Muestreo por Conglomerados Bietapico. Se diferencia del anterior en que luego de seleccionar los conglomerados (UPS), se seleccionan las unidades de manera aleatoria (USS) dentro de cada uno de ellos. El Epidat 3.0 entrega los números de los elementos a analizar, así:



Se aprecia que el programa entrega los números de los elementos seleccionados por cada conglomerado, para el ejemplo, se solicitó 15 elementos por conglomerado.

MUESTREOS NO PROBABILÍSTICOS

Se usan fundamentalmente en estudios exploratorios o en investigaciones de corte cualitativo. No obstante, en muchos casos son los únicos que pueden lograr obtener información básica e importante para cumplir objetivos de la investigación. Se percibe en algunos colegas, que éste tipo de muestreo no es relevante, sin embargo, según las experiencias vividas, esto no es cierto, es más, con estos procedimientos, se logra definir hipótesis importantes y claves. Además, que en la mayoría de casos se puede obtener información complementaria útil para satisfacer el objetivo principal. Algunos de los más comunes son:

7.4.1 Muestreo de Conveniencia: En investigaciones de mercado, y su principal aplicación, valga decir, para determinar posicionamiento, percepción, preferencia etc. Y no es posible realizar un muestreo aleatorio por falta de información o por situaciones prácticas (mayoría de casos). El error puede inducir a pensar que el investigador sesga al tomar a su voluntad las unidades a medir, sin embargo, con una buena estrategia se logra obtener información valiosa a menor costo que el muestreo probabilístico. Si bien presenta la desventaja de que no se pueden generalizar sus resultados, también es cierto que la información obtenida en determinados estudios que no se requiera tanta rigurosidad científica el método es el mejor.

El autor a diseñado múltiples trabajos con este modelo, dado su carácter práctico y económico en múltiples trabajos de investigación.

Ejemplo

Objetivo General: Determinar el grado de satisfacción del **** en el actual Sistema ***** en Salud, en el Departamento de Antioquia, además, de identificar posibles expectativas de los ***** con respecto al mismo sistema.

Tipo de estudio: El Estudio es de tipo cuantitativo, descriptivo y de corte transversal, no concluyente.

Población Objetivo: El universo o población objetivo estuvo conformado por 37 entidades adscritas a ***** Antioquia que prestan servicios de salud en ***** en el área metropolitana del Municipio de Medellín.

Muestra y Muestreo: Dado la cobertura deseada, se obtuvo información de 25 Municipios, repartidos en las diferentes zonas del Departamento de Antioquia. La estrategia fue lograr obtener mínimo 30 informantes por Municipio, exceptuando Medellín por tener un mayor número de docentes.

El marco muestral utilizado fue la base de datos entregada por ***

7.4.2 Muestreo a Juicio. El criterio fundamental es seleccionar unidades de análisis que entreguen información relevante y de una calidad que puede ser más precisa que si se utilizará el azar. En posible caso, es cuando se quiere investigar sobre las vacunas sintéticas en Colombia, parece obvio que el grupo del Dr Patarroyo y/o los grupos de estudio de éste tema serán los apropiados para entregar información útil sobre éste particular. En caso de utilizar un marco muestral de científicos de diversas áreas, cabe la posibilidad de que queden en la muestra personas con poco conocimiento sobre el tema.

Ejemplo:

Objetivo General: Conocer la percepción de algunas IPS del área metropolitana del Municipio de Medellín con respecto a la ***** año 2003.

Tipo de estudio: El Estudio es de tipo cuantitativo, descriptivo exploratorio y de corte transversal. Se aplicó un instrumento básicamente utilizando las escalas Likert, las cuales fueron diseñadas para medir la percepción.

Población Objetivo: El universo o población objetivo está conformada por **** entidades adscritas ***** que prestan servicios de salud en ***** en el área metropolitana del Municipio de Medellín.

Muestra y Muestreo: Se utilizó un muestro a juicio, donde la estrategia consistió en lograr la mayor cobertura posible. En total se obtuvo información de 22 entidades, lo que equivale al 59.45%.

7.4.3 Cuotas: Es la contra parte del muestreo estratificado, en el sentido que se procura obtener información de diversas categorías de interés, de tal manera que todas estén representadas, no obstante, esta selección no se realiza al azar, por el contrario, se efectúa usando un procedimiento no probabilístico, generalmente a conveniencia.

Ejemplo 1: En los estudios de opinión, generalmente se utiliza éste muestreo, valga decir, en estudios para determinar la favorabilidad que tiene la ciudadanía sobre los candidatos a cargos públicos, se determina en que estratos se deben consultar las personas, luego de identificarlos, se define el número de personas a interrogar, sin embargo, ésta selección no siempre se hace al azar. Si fuese al azar, el muestreo pasa a ser estratificado.

Ejemplo 2: Se quiere determinar la operatividad de los procesos ejecutados por los empleados de la Contraloría de Medellín, para ello, se dividen los empleados según dependencias, acá se configuran los estratos, no obstante, el interés es seleccionar personas con conocimiento de los procesos, por lo tanto, no se realiza un muestreo aleatorio al interior de las dependencias, por el contrario, a juicio, se seleccionan los empleados. Luego no hay azar y por eso es un muestreo de cuotas. Si el interés fuese determinar el conocimiento de los procesos, seguramente, el azar sería importante y se utilizaría el muestreo estratificado.

7.4.4 Bola de Nieve. Como su nombre lo indica, el universo en algunos casos y la muestra siempre va en aumento, ya que éste muestreo tiene la característica que el número de elementos muestrales va aumentando. Tiene la particularidad, que los elementos a analizar pertenecen a una población de difícil identificación, luego son los mismos involucrados en la investigación, los que informen sobre el lugar de ubicación de nuevas unidades muestrales. El número final de informantes, depende entonces de las respuestas obtenidas en cuanto a la variabilidad de ellas, valga decir, si las respuestas son homogéneas, el tamaño de muestra será pequeño y

por el contrario, en caso de una alta heterogeneidad se tratará de ubicar un mayor número de personas.

Ejemplo 1: El más trivial, es el caso, donde no se tiene conocimiento de donde ubicar un listado de plomeros o electricistas caseros, sin embargo, todos sabemos donde ubicar el que nos saca de apuros en nuestra casa, y con él, logramos ubicar a otros y estos otros entregan información de unos mas y así sucesivamente.

Ejemplo 2: En un estudio para identificar las causas de abandono de personas con problemas de drogadicción, de los lugares de atención, no fue posible utilizar ningún procedimiento aleatorio, ya que los desertores no están interesados en que los ubiquen, no obstante, un desertor, puede informar sobre la ubicación de su *parcero* y éste a su vez conoce donde se encuentra otra y así se forma la muestra, con la cual sin ser aleatoria, se

antónbules si
P... ..

Realice un cuadro Sinóptico donde se identifiquen los diferentes tipos de muestreo y sus principales aplicaciones (investigue).

En los siguientes casos identifique cual muestreo es mejor o si es preferible el censo y por qué, además, identifique: Población objetivo, marco muestral, unidad muestral y elemento muestral.

- Identificar las fortalezas y debilidades de cada facultad en la Universidad de Antioquia.
- Realizar un sondeo de opinión en Medellín sobre el proceso de paz en Colombia.
- Realizar un estudio de mercado sobre el posicionamiento actual de la FACULTAD NACIONAL DE SALUD PUBLICA de la Universidad de Antioquia en Colombia y Latino América.
- Identificar la percepción que tiene la ciudadanía mayor de 18 años sobre lo que entienden por salud pública.
- Describir los hábitos de consumo alimentario para los internos del

- Realizar un estudio sobre el perfil ocupacional de los egresados hace 5 años de su profesión.
 - Plantear una investigación sobre Las razones de abandono de usuarios de entidades terapéuticas en Medellín.
 - Identificar el perfil socio económico de los asistentes al próximo congreso sobre la Salud Pública.
3. El siguiente marco muestral se obtuvo en una investigación realizada por estudiantes de la especialización en Educación Física en la U de Se eliminaron algunas variables para efectos de presentación. Usar los métodos Aleatorio simple y Sistemático seleccione 15 personas. Con dichos resultados encuentre: el promedio aritmético y su desviación, interprete los resultados. Confronte los resultados encontrados con los obtenidos usando el Epidat 3.0.

ESTRATO	EDAD	SEXO	ESTATURA	PESO	ABDOMINALES	SALTO	LARGO	SPF
4	16	1	173,8	57,4	46	1,93	7,1	
3	15	1	166,5	56	55	2,28	7,1	
3	15	1	162,1	64,9	56	1,91	6,1	
3	15	1	158,2	46,6	60	1,84	8,1	
3	15	1	178,1	60,3	46	1,72	7,1	
4	16	1	163,7	56,1	50	2,14	7,1	
3	16	1	164,5	56,7	83	1,84	7,1	
3	14	1	167,6	60,1	42	2,01	7,1	
3	15	1	160,2	48,8	49	1,85	7,1	
3	16	1	178	64,8	43	2,17	6,1	
4	15	1	170	59,1	54	2,05	7,1	
3	15	1	174,7	63,5	54	2,13	7,1	
3	14	1	177	64,6	57	2,08	7,1	
2	14	1	165,1	65,4	29	1,71	8,1	
3	14	1	168,4	47,8	50	1,71	8,1	
3	16	1	165,3	53	50	2,06	7,1	
3	16	1	168,6	69,9	40	2	7,1	
4	16	1	171,5	54,5	49	2,03	7,1	
3	14	1	171	54,7	34	1,8	8,1	
3	15	1	166,3	64	20	1,44	9,1	
3	16	1	169,6	52,8	40	2,12	7,1	
2	15	1	167,4	48,7	55	2,47	7,1	
2	14	1	167,6	53,1	11	1,52	9,1	
2	14	1	150,6	38,3	21	1,73	8,1	
2	14	1			34	1,8	8,1	
3	16	1			66	1,89	7,1	
3	16	1			60	2,1	7,1	
2	15	1	161,5	52,7	61	2,05	7,1	

4. Se realiza un estudio para estimar el porcentaje de ciudadanos del Bajo Cauca que están a favor de que su agua se trate con flúor. Qué tan grande debe ser una muestra si se desea tener una confianza de al menos 95% de que la estimación estará dentro del 2% del porcentaje real? Realice las consideraciones necesarias para calcular n . Hágalo de manera manual y usando el Epidat.
5. Determine el tamaño de muestra para los siguientes casos:
- a Se requiere una confianza del 90%, se acepta un error de ± 500 y se sabe que el mayor y menor valor son: 50.000 y 10.000.
 - b El mismo caso anterior, pero se conoce que el tamaño de la población es de: 2.000, 5.000 y 10.000. (Desarrollarlo para los tres casos) ¿Qué concluye?
 - c En un estudio de proporciones desea una confianza del 95% y acepta un error en su estimación de 6%. Cuantas personas debe encuestar.
 - d. El caso anterior, pero por estudios anteriores encontró que el 70% están a favor de la variable en estudio. Conclusiones de lo observado, en todos los casos, es útil realizar los ejercicios de manera manual y con el Software estadístico.
6. Decir si cada una de las siguientes afirmaciones es verdadera o falsa, justifique su respuesta?
- El muestreo aleatorio simple entrega resultados similares a los del muestreo sistemático.
 - El estratificado entrega resultados similares al muestreo de cuotas.
 - El muestreo probabilístico se prefiere al no probabilístico en todos los casos.
 - La diferencia fundamental del muestreo probabilístico con el no probabilístico es que el primero generaliza resultados a toda la población.
 - Mientras más grande es el n , mejor es el resultado del estudio.
 - El valor de $p=0.5$ garantiza el n más pequeño, cuando se trabaja con proporciones.
 - El error admisible es complementario a la confiabilidad del estudio.
 - Cuando se utiliza la fórmula de promedios, el error se reemplaza en porcentaje.

ando el N es relativamente lejano al n de muestra ($n < 0.05N$), se ajusta con la fórmula para población infinita.

El muestreo estratificado es menos costoso que todos y por ello, se recomienda siempre.

La diferencia entre el muestreo estratificado y el conglomerado es que el último realiza censo luego de quedar seleccionado un conglomerado.

El Bietapico es más complejo que el conglomerado de una sola etapa.

- D. Decidir si cada una de las siguientes afirmaciones es verdadera o falsa. Justifique su respuesta.

Para un tamaño poblacional y una varianza muestral dados, cuando mayor sea el tamaño muestral, más amplio será un intervalo de confianza del 95% para la media poblacional.

Para unos tamaños poblacional y muestral dados, cuanto mayor sea la varianza muestral, más amplio será un intervalo de confianza del 60% para la media poblacional.

Para un tamaño y una varianza muestral dados, cuanto mayor sea el tamaño de la población, más amplio será un intervalo de confianza del 60% para la media poblacional.

Para unos tamaños poblacional y muestral dados, y para una varianza muestral dada, un intervalo de confianza del 95% para la media poblacional es más amplio que un intervalo de confianza del 90% para el mismo parámetro.

- La empresa CIEM requiere llevar a cabo un estudio de mercado para determinar la aceptación de los usuarios sobre el servicio de recolección de escombros, para ello, se obtuvo la siguiente información:

Usuarios residenciales	12 000
Usuarios industriales	4 000
Total usuarios	500

Determinar el tamaño de muestra para cada categoría de usuario, si se desea una confiabilidad del 95% y un error del 6%, además, se sabe que la proporción de usuarios satisfechos en el último estudio fue de 50%. Realícelo manualmente y compare los resultados con los obtenidos por el Epidat 3.0.

C) el fin de determinar si existen diferencias en las notas promedio en el área de estudio, una universidad diseñó las siguientes categorías y obtuvo la siguiente información:

Humanidades	2500
Informática	1450
Matemáticas	750

Encuentre el tamaño de muestra para cada categoría de usuario, teniendo en cuenta además, que la desviación de las notas según estudios previos es de 1.2, use una confianza del 90% y un error de 0.2, considere el error es adecuado? ¿Por qué? Realícelo manualmente y compare los resultados con los entregados por el Epidat 3.0.

Usando el Epidat determine el tamaño de muestra para los siguientes casos:

a) tiene una población de 2.000 profesores, de los cuales 600 pertenecen a las áreas sociales, 900 a las ingenierías y el resto al área de la salud. Use una confianza del 93%, un error de 100.000 además se sabe el mayor y menor valor salario son: 4.500.000 y 1.500.000.

b) en un estudio de proporciones se desea una confianza del 90% y permita un error en su estimación de 4%. Cuantas personas debe encuestar, si la proporción de personas a favor de la reelección es del 60%.

c) el caso anterior, pero se quiere estratificar por regiones, teniendo la siguiente información: Costa Pacífica 3.000.000, zona Centro 5.000.000, zona Sur 4.500.000.