MUESTREO

La adecuada elección de los participantes asegura que sus hallazgos representen de forma exacta lo que sucede en la población de interés.

Las investigaciones usualmente se hacen con muestras por razones económicas, de tiempo, logísticos, calificación de evaluadores, donde resulta imposible de estudiar toda población.

El muestreo es el acto, proceso, o técnica que permite seleccionar una muestra adecuada, o una parte representativa de la población con el propósito de determinar parámetros o características de la población completa, debe ser suficientemente grande para controlar el error aleatorio y lo suficientemente representativa para permitir la generalización de los hallazgos del estudio a la población de interés.

El muestreo permite determinar que parte de una realidad debe examinarse con la finalidad de hacer inferencias válidas sobre ella.

Una buena muestra es una versión simplificada de la población, reproduce sus rasgos básicos, además es simple de manejar.

El muestreo permite determinar que parte de una realidad debe examinarse con la finalidad de hacer inferencias sobre ella.

Una buena muestra es una versión simplificada de la población, reproduce sus rasgos básicos, además es simple de manejar.

Los estudios descriptivos tienen como intención describir a una población, por tal fin requieren una muestra representativa de la misma.

El énfasis para los estudios analíticos es lograr que los grupos sean similares en todos los efectos salvo en los que concierne al factor en estudio, la demanda fundamental en la muestra no está en su representatibilidad sino en su comparabilidad

Sitio:	Espacios Virtuales de Posgrado (EVP)
Curso:	ESTADISTICA PARA LA INVESTIGACION (Grupo: CVEi18m)
Libro:	MUESTREO
Imprimido por:	Leandro Huayanay
Día:	domingo, 15 de julio de 2018, 09:11

Tabla de contenidos

1 POBLACION Y MUESTRA

La población se refiere al universo que se desea estudiar. es par quien se desea generalizar los resultados.

Los trabajos estadísticos deben cumplir con uno de los siguientes objetivos: 1) describir cuantitativamente una población estudiando la totalidad de sus elementos o 2) describir cuantitativamente una población a partir de una pequeña parte del total de sus elementos. Durante el proceso de investigación se debe procurar alcanzar el primer objetivo. Por desgracia, esto no siempre es posible, por lo que se opta por estudiar sólo una parte, o muestra, del universo. Entre los diversos motivos que obligan a esto se encuentran:

- El universo de interés es infinito, o finito pero enorme, por lo que es imposible estudiar a todos sus elementos. Por ejemplo, existe interés en estudiar el peso y la talla de todos los humanos que viven en la actualidad. El universo de interés es tan grande, disperso y difícil de localizar que el investigador decide estudiar una muestra.
- Parte de la población no está disponible para el estudio. Por ejemplo, el interés reside en estudiar el efecto de un medicamento en un grupo de enfermos, pero algunos elementos de la población están recibiendo otro tratamiento que modifica la respuesta del que se quiere probar.
- Para estudiar sus características es necesario destruir a los elementos. Por ejemplo, al médico le interesa conocer la concentración de glucosa en sangre de su paciente. Dado que estudiar toda la sangre probablemente represente la muerte del paciente, sólo se toma una parte de ella (no más de 10 ml de sangre) para realizar el estudio.
- Es imposible identificar todos los elementos del universo de interés. Por ejemplo, a un epidemiólogo le interesa describir los antecedentes hereditarios de los enfermos de diabetes mellitus. Dado que no todos estos enfermos están identificados, se conforma con estudiar una muestra de ellos.
- Existen pocos recursos (económicos, humanos, tecnológicos) para estudiar el universo. Por ejemplo, durante la investigación surge la necesidad de practicar una tomografía axial computarizada (TAC) a los elementos del universo de interés. Desgraciadamente, no hay personal capacitado en cantidad suficiente para realizar el estudio en todo el universo y la capacitación de ellos excede los límites de la investigación (mucho tiempo y dinero para ello). Los investigadores deciden practicar la TAC en una parte del universo.

Enseguida precisaremos algunas definiciones:

elementos

Población; Una población es un grupo de personas individuales, u objetos a partir de los cuales se toman las muestras para realizar las mediciones.

Algunas características de una población, están constituido por personas que comparten características demográficas, culturales, étnicas, hábitos de vida, alimentación, que tienen relevancia para el estudio. Es un conjunto de individuos que guardan similitud entre si en los aspectos que son relevantes para la investigación. A los sujetos de estudio:

Población Objetivo : grupo de pacientes con características clínicas y demográficas a las cuales se generalizarán los resultados.

Población Accesible: subgrupo de la población objetivo disponible para el estudio, con características geográficas y temporales determinadas.

Población Elegible: son quienes cumplen los criterios de inclusión y no los de exclusión.

Los criterios de inclusión son aquellos que permiten definir y caracterizar la población elegible, particularinzádola del resto.

Los criterios de exclusión son aquellos que indican que quien tendrá que ser excluido por alguna razón, ej. Alergia, consideación ética, contraindicación, etc. Permiten homogenizar a la población y excluir a individuos que no corresponden al grupo

Muestra subgrupo de la población. La muestra es un subconjunto de la población, que permite inferir, estimar, o extrapolar los resultados a partir de las observaciones y mediciones realizadas e inferir lo que ocurre en la población total. **El muestreo** es el acto, proceso, o técnica que permite seleccionar una muestra adecuada, o una parte representativa de la población con el propósito de determinar parámetros o características de la población completa.

Como muestra se define una parte del universo o población, y ***n*** representa el total de elementos en la muestra. Usualmente no es posible estudiar el total del universo, se selecciona una muestra y, a partir de ella, se hacen inferencias sobre la población; éste es el campo de la estadística inferencial. Pero para que las inferencias sean útiles, la muestra debe ser un reflejo del universo a partir del cual se obtuvo. Por desgracia, no hay una manera infalible de obtenerla a pesar de que se han descrito diversos procedimientos para ello. En términos generales, existen dos tipos de procedimientos mediante los cuales se obtiene una muestra: muestreo probabilístico y muestreo no probabilístico. Sólo para el muestreo probabilístico existen procedimientos estadísticamente seguros que permiten hacer inferencias, a partir de una muestra, sobre la población.

2 MUESTREO PROBABILISTICO

Una muestra probabilística es una muestra extraída de una población, de tal manera que todo miembro de la población tenga una probabilidad conocida, mayor de 0, de ser incluido en la muestra.

2.1 Muestreo aleatorio simple

Si tenemos una población de tamaño N y seleccionamos una muestra de tamaño n cuando todas las muestras posibles de ese tamaño n tienen la misma probabilidad de ser seleccionadas, entonces hemos seleccionado una muestra aleatoria simple. Para su empleo es indispensable disponer de un marco muestral, es decir, un listado con los elementos de la población numerados del 1 a N (N = tamaño de la población). La selección de los elementos que componen la muestra es al azar, por lo que las preferencias y deseos del sujeto no influyen en este proceso. Para que la selección sea al azar se debe utilizar uno de los siguientes procedimientos: sorteo, tabla de números aleatorios, números aleatorios generados por una calculadora o por una computadora.

2.2 Muestreo estratificado

En ocasiones, se sabe que la población está distribuida en subgrupos, y que los miembros de cada subgrupo difieren de los demás en cuanto a las características que se desea estudiar. Para controlar esa diferencia entre los subgrupos, se puede tomar en cuenta el subgrupo, o estrato, al que pertenecen los elementos. El principio básico en que se apoya este tipo de muestreo es dividir la población en estratos con el fin de obtener representatividad de los distintos subgrupos que componen la población y hacer comparaciones entre ellos.

En cada uno se selecciona una muestra, cuya suma representa la muestra total. En este tipo de muestreo, los estratos se consideran como poblaciones independientes. Una vez que se ha decidido cuántos elementos de cada estrato se deben seleccionar, sólo resta aplicar los criterios del muestreo aleatorio simple a cada estrato.

Al igual que como ocurrió en el muestreo aleatorio simple, las estadísticas se afectan por el diseño del muestreo y es preciso introducir modificaciones a las fórmulas que se utilizan. Por ello, se debe tomar en cuenta que las siguientes estimaciones se calculan con las fórmulas que se presentan a continuación:

- Proporción muestral

$$\hat{p}_{\text{str}} = \frac{\sum_{h=1}^{L} N_h \hat{p}_h}{N}$$

donde Pstr representa la proporción muestral estimada mediante muestreo estratificado, L es igual al total de estratos, h identifica cada estrato con un número progresivo que va de 1 a L, Nh es el tamaño de la población para el estrato h-ésimo, ph es la proporción muestral en el estrato h-ésimo y N es la población total.

- Media muestral

$$\bar{x}_{\text{str}} = \frac{\sum_{h=1}^{L} N_h \bar{x}_h}{N}$$

en la xstr que representa la media muestral estimada por medio del muestreo estratificado y xh es la media

muestral en el estrato h-ésimo.

En el caso de que el grupo muestreado en cada estrato sea proporcional a su tamaño en el universo, se pueden utilizar las fórmulas del muestreo aleatorio simple.

2.3 Por racimos o conglomerados

En ocasiones el universo es muy grande y no es posible obtener un marco muestral de él. En estos casos, generalmente se pueden identificar ciertos agrupamientos naturales que sí es posible enumerar, y es factible realizar el muestreo considerando los diferentes subgrupos o conglomerados. A diferencia del muestreo estratificado en que se toman todos los subgrupos, en el muestreo por conglomerados sólo algunos subgrupos se seleccionan aleatoriamente. Después, se registran las características de todos los elementos del conglomerado seleccionado, o bien, se selecciona una muestra de él. El muestreo por conglomerados puede verse como un muestreo en etapas, en el que cada etapa es en sí un muestreo aleatorio simple. Éste es un procedimiento de gran ayuda cuando los estudios son a gran escala. Su ventaja principal es el ahorro de recursos y tiempo.

Al igual que los anteriores, este modelo de muestreo introduce cambios en las fórmulas mediante las cuales se calculan los estimadores de la población. Este diseño tiene el problema de que las fórmulas son diferentes para muestras por conglomerados de diferentes etapas. Aquí se presenta la fórmula para estimar la media por unidad de listado cuando se desconoce N. Esta fórmula es sencilla, pero tiene el inconveniente de proporcionar un estimador sesgado. Sin embargo, el sesgo es despreciable en la mayoría de las ocasiones. Otras fórmulas se pueden encontrar en textos especializados de muestreo.

$$\bar{\bar{x}}_{\text{clu}} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{nj} x_{ij}}{\sum_{i=1}^{m} n_i}$$

2.4 Muestreo sistemático

En este procedimiento se seleccionan los elementos de la muestra determinando de antemano cuántos elementos se dejarán pasar antes de seleccionar el que se tomará en cuenta para integrar la muestra. Aunque se considera que no reúne todos los requisitos de aleatoriedad, es de suma utilidad cuando el tamaño de la población es muy grande y es difícil elaborar un marco muestral, o no se dispone de suficientes páginas de números aleatorios.

Las fórmulas necesarias para calcular los estimadores son semejantes a las utilizadas en el muestreo aleatorio simple.

3 MUESTREO NO PROBABILISTICO

El muestreo no probabilístico se justifica por la comodidad y la economía, pero tiene el inconveniente de que los resultados de la muestra no siempre pueden generalizarse para toda la población.

3.1 Muestreo de conveniencia

Mediante este procedimiento, la muestra se conforma por sujetos que pueden ser fácilmente accesibles en la población que se desea estudiar.

Ejemplo: El investigador quiere realizar un diagnóstico de comunidad y para su muestra selecciona a los habitantes del centro de la población. Su decisión está basada en el hecho de que son más accesibles y fáciles de localizar.

3.2 Muestreo de casos consecutivos

Entre los muestreos no probabilísticos, éste es el que más se aproxima a la selección aleatoria y se puede utilizar en una gran variedad de investigaciones. Consiste en estudiar a todos los sujetos accesibles que se puedan identificar durante el tiempo en que se realiza el estudio.

Ejemplo: Un investigador está interesado en estudiar las lesiones que sufren los niños como peatones en accidentes de tráfico de vehículo de motor. El investigador tiene varias opciones para sistematizar sus casos de estudio: lesionados que fallecieron, lesionados que ameritaron hospitalización, lesionados atendidos en unidades de emergencia, lesionados que ameritaron uno o más días de incapacidad para desarrollar sus actividades habituales, etc. Cada uno de estos criterios tiene características que dificultan una selección aleatoria, pero el más difícil de superar tiene que ver con la población a la que le interesaría representar: seguramente el investigador quiere hacer inferencias sobre la población de lesionados que se presentará en los años que siguen. De esta manera, es imposible diseñar un marco muestral, y tampoco es posible esperar muchos años para que conteste su pregunta de investigación. En estas condiciones, una selección de casos consecutivos es lo más parecido a un muestreo aleatorio, siempre y cuando los casos seleccionados sean todos, o casi todos, los casos potenciales de estudio. En la medida que la “no respuesta” de los sujetos a estudiar aumenta y rebasa 10%, la muestra estudiada puede no ser representativa de la población.

3.3 Muestreo en bola de nieve

En este muestreo, a los sujetos estudiados se les pide que recomienden a otros sujetos, a los que se buscará para entrevistarlos.

Ejemplo: El investigador, después de entrevistar a un adicto a las drogas, quiere obtener más datos, así que le pregunta al entrevistado a quién más podría visitar, de tal manera que cuando inicie la entrevista se presente señalando que alguien conocido lo envió.

3.4 Muestreo a criterio

Este muestreo contempla la selección de sujetos que, a juicio del investigador, podrán proporcionar mayor información entre la población estudiada.

Ejemplo: Durante la búsqueda de datos, el investigador incluye en su muestra al párroco del lugar, al médico de la clínica y a los profesores de la escuela primaria, además del comisario ejidal y de la persona de mayor edad en la región. Él cree que estas personas tienen la información que le interesa.

4 DETERMINACION DEL TAMAÑO DE LA MUESTRA

El tamaño de la muestra es un procedimiento muy importante, ya que nos va ha precisar , el numero de individuos necesarios para que nuestro estudio sea adecuado.

Estará condicionado por el problema y el diseño de investigación , lo que influye sobre la manera de calcular

Es el menor que nos pueda dar la información que necesitamos, influye sobre el costo del estudio y por ende en la factibilidad

Para la determinación del tamaño de la muestra, se debe tener en cuenta:

- Tamaño del universo
- Tipo de variable, diseño.
- Homogeneidad-heterogeneidad (estratos)
- Margen de error
- Nivel de confianza
- Diferencia que se espera encontrar
- Recursos disponibles

Para el calculo del tamaño de la muestra se puede facilitar con el uso de programas: En esta sección se recomienda utilizar el soporte del programa Epidat.

También es posible usar la siguiente web que permite el cálculo del tamaño muestral en distintos tipos de estudio: http://www.openepi.com/Menu/OE_Menu.htm

4.1 Tamaño muestral para estimar una proporción

Si se desea estimar una proporción, debe conocerse:

- El nivel de confianza o seguridad (1−α). El nivel de confianza prefijado da lugar a un coeficiente (zα). Para un nivel de seguridad del 95 % α=1.96, para un nivel de seguridad del 99% α= 2.58;
- La precisión que se desea para el estudio;
- Una idea del valor aproximado del parámetro que se quiere medir (en este caso, una proporción). Esta idea se puede obtener revisando la literatura o mediante estudio pilotos previos. En caso de no tener dicha información se utilizará el valor p = 0.5 (50 %).

$$n = \frac{z_{\alpha}^{2} \times p \times q}{d^{2}}$$

donde:

- z2α = 1.962 = 3.84 (en caso se busque una confianza del 95%)
- p es la proporción esperada
- q = 1− p
- d es la precisión deseada (depende de lo que se busque en el estudio)

Si la población es finita, es decir se conoce el total de la población y se desea saber cuántos individuos hay que estudiar, la respuesta sería:

$$n = \frac{N \times z_{\alpha}^{2} \times p \times q}{d^{2} \times (N-1) + z_{\alpha}^{2} \times p \times q}$$

Adicional, a lo ya mencionado en la anterior fórmula: N es el total de la población

4.2 Tamaño muestral para estimar una media

Si se desea estimar una media habrá que conocer:

- El nivel de confianza o seguridad (1−α). El nivel de confianza prefijado da lugar a un coeficiente (zα). Para un nivel de seguridad del 95%, α=1.96, para un nivel de seguridad del 99% α= 2.58;
- La precisión con que se desea estimar el parámetro (2×d es la amplitud del intervalo de confianza);
- Una idea de la varianza s2 de la distribución de la variable cuantitativa que se supone existe en la población

$$n = \frac{z_\alpha^2 \, s^2}{d^2}$$

Si la población es finita, como previamente se señaló, es decir se conoce el total de la población y se desearía saber cuántos individuos debería estudiarse, la respuesta sería:

$$n = \frac{N \, z_\alpha^2 \, s^2}{d^2 \left(N-1\right) + z_\alpha^2 \, s^2}$$

4.3 Tamaño muestral en estudios sobre prevalencia

En estos se requiere conocer el tamaño de la población, la precisión deseada (%), la prevalencia esperada, el efecto del diseño y el error α.

En este caso se debe tener una idea aproximada de la prevalencia que se está investigando, a continuación, se debe establecer cuánto se aceptaría que varíe (precisión), seguidamente el nivel de confianza deseado para la indagación y, finalmente, el efecto de diseño de la investigación. Este es igual a 1 (no hay efecto de diseño) si se ha empleado un procedimiento de muestreo aleatorio simple o bien sistemático y corresponde a un error sistemático.

La fórmula:

$$n = \frac{N\ z_{1-\alpha}^{2}\ p\left(1-p\right)}{d^{2}\left(N-1\right)+z_{1-\alpha}^{2}\ p\left(1-p\right)}$$

donde:

- n es el tamaño de la muestra
- N es la población total
- $z_{1-\alpha}$ es el valor de z para el nivel de confianza ($1-\alpha$)
- p es la proporción esperada en la población
- d es la precisión absoluta

El tamaño de la muestra es, en definitiva, igual a n por el efecto de diseño. El tamaño de la población de la que se extraerá la muestra es habitualmente desconocido, pero no es muy importante tener un conocimiento exacto, basta con una aproximación razonable.

4.4 Tamaño muestral en estudios de casos y controles

Supóngase que se quiere llevar a cabo un estudio de casos y controles con el fin de determinar si existe una relación significativa entre la exposición a un factor y la presencia de una determinada enfermedad. A continuación, se explica cómo calcular el tamaño de muestra necesario para contrastar la hipótesis de que el odds ratio (OR) sea igual a 1.

Si se conoce la probabilidad de exposición entre los controles p2, y se prevé que el OR asociado al factor de estudio es w, el valor de p1, la frecuencia de exposición entre los casos puede obtenerse fácilmente:

$$OR = w = \frac{p_1(1-p_2)}{p_2(1-p_1)} \Rightarrow$$

$$w\,p_2(1-p_1) = p_1(1-p_2) \Rightarrow$$

$$p_1(1-p_2+w\,p_2) = w\,p_2 \Rightarrow$$

$$p_1 = \frac{w\,p_2}{(1-p_2)+w\,p_2}$$

donde:

- p1 y p2 son las proporciones esperadas;
- w es el valor aproximado del OR que se desea estimar.

Así, el problema del cálculo del tamaño muestral podrá abordarse mediante las fórmulas habituales empleadas en la comparación de dos proporciones.

Recurriendo a las fórmulas habituales para determinar el tamaño muestral mínimo necesario para la comparación de dos proporciones, se precisará conocer:

- La magnitud de la diferencia a detectar, que tenga interés clínicamente relevante. En este caso, como ya se vio, bastaría con conocer dos de los siguientes tres parámetros:

  o Una idea del valor aproximado del OR que se desea estimar (w)
  o La frecuencia de la exposición entre los casos (p1)
  o La frecuencia de la exposición entre los controles (p2)

- El nivel de seguridad α o riesgo de cometer un error de tipo I, con que se desea trabajar. Generalmente con un nivel de seguridad del 95%, α=0.05.
- La potencia estadística o riesgo de cometer un error de tipo II (1−β) que se desea para el estudio. Es habitual tomar β=0.2, es decir, una potencia del 80 %.

Con estos datos, y para un planteamiento bilateral, para el cálculo del tamaño muestral se utilizará la expresión:

$$n = \left(\frac{z_{1-\alpha/2}\sqrt{2p(1-p)} + z_{1-\beta}\sqrt{p_1(1-p_1) + p_2(1-p_2)}}{p_1 - p_2} \right)^2$$

donde: $p = \dfrac{p_1 + p_2}{2}$;

- w es una idea del valor aproximado del OR que se desea estimar;
- p1 es la frecuencia de la exposición entre los casos;
- p2 es la frecuencia de la exposición entre los controles; y
- 1 2 z −α y 1 z −β son valores que se obtienen de la distribución normal estándar en función de la seguridad y la potencia seleccionadas para el estudio. En particular, para un nivel de seguridad de un 95% y una potencia estadística del 80% se tiene que z1-α/2 = 1.96 y z1−β = 0.84.

Hasta ahora se ha asumido un tamaño muestral igual para casos y controles. En caso de que el número de casos y controles no esté equilibrado, la expresión anterior deberá ser ligeramente modificada:

$$n = \frac{\left(z_{1-\alpha/2}\sqrt{(c+1)p(1-p)} + z_{1-\beta}\sqrt{c\,p_1(1-p_1) + p_2(1-p_2)} \right)^2}{c\,(p_1 - p_2)^2}$$

donde:

- n es el número de casos
- m es el número de controles
- c = m/n es el número de controles por cada caso

Así, el número de controles vendría dado por m = c × n.

Debe precisarse que se ha tratado en el presente trabajo de exponer del modo lo más sencillo posible el procedimiento a seguir en el cálculo del tamaño de la muestra en un estudio de casos y controles. No obstante, en ocasiones se utilizan para este cálculo expresiones más complejas basadas en una corrección de la fórmula del cálculo del tamaño muestral para la comparación de dos proporciones. Así mismo, existen fórmulas específicas para el cálculo del tamaño de la muestra en el caso de que el diseño corresponda a un estudio de casos y controles apareados.

4.5 Tamaño de muestra estudios de cohortes

La siguiente fórmula proporciona el número de expuestos requeridos y, naturalmente, el de no expuestos

como c × N:

$$N = \frac{n'}{4}\left(1 + \sqrt{\frac{1 + 2(c+1)}{n'\, c\, \left| p_0 (RR-1) \right|}}\right)^2$$

Donde:

$$n' = \frac{\left(z_\alpha \sqrt{(c+1)\, p\,(1-p)} + z_\beta \sqrt{c\, p_0 (1-p_0) + p\, RR (1 - p_0\, RR)} \right)^2}{c\left(p_0 (1-RR)\right)^2}$$

$$p = \frac{\left(p_0\; RR\right) + \left(p_0 + c\right)}{1 + c}$$

- q = 1 – p
- p es la frecuencia de la condición en estudio en la población no expuesta, expresada en forma decimal: 0.05 (5%)
- RR es el riesgo relativo que se considere digno de ser detectado (o mayor): RR =1 significa que el factor de exposición no se encuentra asociado a un aumento del riesgo, puesto que este es igual en expuestos y no expuestos
- c es la relación numérica de expuestos/no expuestos (muestra si las cohortes son de igual tamaño o no)
- α es el error α y zα es su respectivo valor z
- β es el error β y zβ es su respectivo valor z

Un ensayo clínico controlado puede ser considerado un seguimiento de dos cohortes, una expuesta al tratamiento experimental y la otra no. La similitud se puede comprobar también en el cálculo del tamaño de la muestra, que genera iguales resultados si se emplea RR de 3, como en el ejemplo recién presentado, o se calcula con el resultado en la forma de dos proporciones, 0.10 en los no expuestos y 0.30 en los expuestos.