

¿Cómo seleccionar el tamaño de una muestra para una investigación educacional?*

Autor:

M. Sc. Ing. Arabel Moráquez Iglesias

arabelm@hlg.rimed.cu

Resumen

El objetivo de este trabajo tiene como finalidad proponer una metodología que permita calcular el tamaño de una muestra a partir de una población constituida por estratos. Con el empleo de esta metodología se pudo determinar la cantidad de escuelas politécnicas a seleccionar en el municipio de Holguín, para luego calcular la cantidad de estudiantes a escoger en cada escuela, nivel y especialidad, en la investigación actual.

Palabras claves: Aleatorio, investigación, metodología, muestra, muestra aleatoria, educacional, estratificada, muestra estratificada, tamaño muestral.

Summary

The objective of this work is a proposal of a methodology that allows calculating the size of a sample from a population constituted by strata. With the employment of this methodology, it was possible to determine the quantity of polytechnic schools to be selected in Holguín municipality, and then to calculate the number of students to be chosen in each school, level, and specialty in the current investigation.

Key words: Educational, Investigation. Methodology, random, random sample, sample, sample size, stratified, stratified sample.

Para todo aquel que inicia una investigación, le resulta un poco difícil concebir el diseño de la misma y dentro de este determinar la muestra a tomar de una población, establecida por su objeto y campo de estudio; de aquí que el objetivo de este trabajo está encaminado a ofrecer a los profesores e investigadores una metodología o vías para poder determinar el tamaño de una muestra a seleccionar en su campo de estudio con un mayor grado de cientificidad, sin que estos tengan un gran conocimiento de las matemáticas y de la estadística, de forma clara, práctica y comprensible.

Es conocido para todos, que cuando se trata de seleccionar el tamaño de la muestra, se recurre a determinados criterios, fundamentalmente empíricos y en la mayoría de las veces,

permeados de mucha subjetividad por parte del investigador o investigadores que la seleccionan, lo que hace que esta arista tan importante en una investigación carezca de un buen sustento estadístico-matemático y por consiguiente científico. De manera que el investigador no pueda convencer o demostrar que el tamaño de la muestra seleccionada sea verdaderamente representativo en relación con la población con que se trabaja; constituye este el talón de Aquiles de muchas investigaciones llevadas a cabo.

Es importante, si se habla de investigaciones y, dentro de ellas, del rol que juega la estadística como rama de la matemática, que se dejen definidos algunos conceptos alrededor de este objeto de estudio.

- ◆ La estadística como método de investigación científica.-

Este autor considera, en contraposición de otros investigadores, que en toda investigación deben aplicarse los distintos estadígrafos que aportan la estadística descriptiva e inferencial, lo que posibilita un análisis mucho más científico del objeto de estudio investigado y coadyuva a la asunción y comprobación de las hipótesis planteadas, entre otras. Es por ello que en este artículo se considera conveniente dejar claro qué es la estadística y qué papel juega en una investigación científica.

- Algunos presupuestos teóricos:

De muchos autores consultados, y por razones obvias de espacio, se analizarán dos conceptos de estadística: el que emite Montgomery (1996) y el de Devore (2000), los que plantean:

“El campo de la estadística tiene que ver con la recopilación, presentación, análisis y uso de datos para tomar decisiones, y resolver problemas.” (Montgomery, 1996). *“Es la rama de la investigación científica que proporciona métodos para organizar y resumir información, y usar esta para obtener diversas conclusiones”.* (Devore, 2000).

Al analizar ambos conceptos se puede llegar a la conclusión que son similares, pero Montgomery no la define como una rama de la ciencia; que Devore sí lo hace, al decir que: *“es la rama de la investigación científica que proporciona métodos para organizar y resumir información, y usar esta para obtener diversas conclusiones.”*

La definición de Devore es la asumida por este autor por considerar a la estadística como una rama de la investigación científica y porque además, establece que la estadística proporciona métodos para organizar y resumir información, que luego será usada para obtener diversas conclusiones; es decir, que no se trata solamente de organizar y recopilar

datos para resolver problemas, como plantea Montgomery. De ello se infiere que la estadística es un método de investigación que se conjuga con los métodos teóricos y empíricos de esta y que se imbrica con el método matemático, por ser una rama de ella y que es considerada como una ciencia exacta.

Cuando una investigación es apoyada por métodos estadísticos, resulta obvio que constituye una herramienta poderosa, mediante la cual se resumen los datos y se obtienen conclusiones de estos, que coadyuvan al desarrollo de una investigación; a lo que modestamente este autor agregaría que la estadística es *el arte de hacer hablar a los números de manera científica, lo que permite resultados más pertinentes en una investigación.*

A la mayoría de los investigadores les resulta difícil determinar el tamaño de una muestra, y, siempre que se hace, existe la incertidumbre si esta es o no verdaderamente representativa; por lo que deben preguntarse: ¿Qué es una muestra representativa? ¿Qué es lo que verdaderamente la hace representativa?

Existen infinidad de criterios acerca de lo que es la representatividad de una muestra. Si se hace esta pregunta a distintos investigadores, evidentemente se obtendrían respuestas diversas, y en su mayoría permeadas por el factor subjetivo de lo que esas personas entienden por representatividad. Así que es mejor partir de las definiciones de representar, representativo y representatividad, donde:

- ✦ Representar, v. a. Presentar de nuevo. Informar, declarar..., Ser imagen de una cosa, reproducir, expresar, imagen, figura... Diccionario Larousse.
- ✦ Representar (del lat. *repassentare*): Ser imagen o símbolo de una cosa, o imitarla perfectamente. Diccionario Enciclopédico Hispano-Americano.
- ✦ Representativo, v. a. adj. Que representa: signos representativos de la riqueza. Diccionario Larousse.
- ✦ Representativo, v. a. adj. Dícese de lo que sirve para representar otra cosa. Diccionario Enciclopédico Hispano-Americano.

En estas definiciones se observa que si representar no es más que expresar, reproducir una cosa, o ser su imagen, y que representativo es lo que sirve para representar otra cosa; por lo tanto, representatividad es la cualidad de lo representativo. De aquí que si algo está bien representado, se puede decir entonces que ese algo tiene representatividad acerca de lo que representa.

Es por ello que al seleccionar una muestra, si se sabe que esta constituye un subconjunto de la población, se debe tener cuidado que la misma tenga las mismas propiedades de la población y que obedezca a determinados argumentos, desde el punto de vista estadístico, para decir que la misma es una muestra representativa.

Pero antes de analizar en qué consisten estos argumentos estadísticos, para llamarlo de alguna forma práctica, resulta necesario dejar bien claro lo que es universo o población y muestra, lo que a estos fines se hará mediante ejemplos para que al lector le resulte más comprensible.

✦ **Concepto de Universo o población:**

Si por ejemplo, una investigación tiene como objetivo aportar una serie de indicadores para evaluar la calidad del trabajo metodológico en la Educación Técnica y Profesional (ETP en lo adelante) **en la provincia de Holguín**, el objeto de estudio, entre otras cosas, tiene como universo o población al claustro de profesores, dirigentes y estudiantes de toda la provincia de Holguín, pero de la ETP.

Si por el contrario, esta investigación se circunscribe a las escuelas politécnicas de la ETP del municipio de Holguín, ahora el universo o población no estará constituida por los profesores, dirigentes y estudiantes de la ETP de cada uno de los 14 municipios de la provincia, sino que estará constituida por los profesores, dirigentes y estudiantes de la ETP, **únicamente del municipio de Holguín**. Por lo que ahora se observa que el universo o población ha cambiado a un sector más estrecho; de aquí la relatividad de la población a tomar en una investigación.

Si este mismo ejemplo se circunscribe a todas **las escuelas de oficios del municipio de Holguín**, resulta evidente que la población ya no estará constituida por todos los estudiantes de la ETP, sino sólo los de las escuelas de oficio y por consiguiente, se excluyen de ellas las que forman técnicos medios de este municipio.

De todos los ejemplos acotados, es importante observar que la población no la conforman estudiantes, profesores y dirigentes de cualquier entidad o sector; por el contrario, todos se refieren al sector de la Educación y dentro de este no a cualquier subsistema de enseñanza, sino específicamente al de la Educación Técnica y Profesional. Por lo que el primer requisito que deben tener los objetos (pueden ser personas o cosas) que conformarán el universo o población es que deben tener las mismas características o propiedades, de aquí el carácter **homogéneo** de la misma.

Muy importante para el investigador es dejar bien claro cuál es su objeto y campo de estudio, porque los mismos determinarán la población a trabajar.

✦ ¿Qué se entiende por población estadística?

Población ⁽¹⁾ [población] f. [Del lat. *populatio*, *-ōnis*]

1. Acción y efecto de poblar.
2. Conjunto de personas que habitan la Tierra o cualquier división geográfica de ella.
3. Conjunto de edificios y espacios de una ciudad.
4. Conjunto de individuos de la misma especie que ocupan una misma área geográfica.
5. Conjunto de los individuos o cosas sometido a una evaluación estadística mediante muestreo.

El autor del presente artículo considera que: *“Población es el conjunto de todos los individuos, objetos, procesos o sucesos homogéneos que constituyen el objeto de interés. La población se relaciona directamente con el campo de estudio.”* (Moráquez, 2005).

Es indudable, que a partir del concepto anterior, que es el asumido en este trabajo, se observa una dicotomía entre los objetos (personas o cosas), procesos y sucesos que cumplen con determinadas características, ya que estos pueden ser infinitos, por ejemplo el estudio de las estrellas, entre otras; o por el contrario, pueden ser finitos, como los ejemplos antes expuestos.

Como el objetivo de este trabajo está encaminado a las investigaciones sociales, entre ellas las educacionales, resulta evidente que siempre se trabajará con poblaciones finitas.

En todos los casos analizados en que la población ha variado, se llega a la conclusión que resulta imposible poder trabajar con cada una de las personas u objetos que conforman la población por razones económicas y operativas, por lo que es imprescindible obtener un subconjunto de elementos representativos de esta población homogénea para trabajar con ella y eso no es más que la **muestra**, la que es *un subconjunto o parte de una población*. Por lo que tienen las mismas características o propiedades de la población de donde se tomó.

Ya se tiene claro qué es una muestra y que la misma tiene que ser extraída de la población objeto de estudio, pero el problema está dado en cómo puede ser extraída y cuántos elementos se deben tomar de una población para decir que hay una calidad en su representación, o lo que es lo mismo, que hay representatividad.

Si una muestra es representativa, entonces se puede inferir toda una serie de importantes conclusiones acerca de la población (estadística inductiva o inferencia estadística) o describir

características observadas en la muestra (estadística descriptiva), que permitan posteriormente hacer inferencias con respecto a la población. Es por ello importante que todo investigador deje bien claro que la muestra asumida es **representativa** de la población extraída.

Entonces: ¿Cómo hacer para que una muestra sea representativa? ¿De qué forma se puede extraer la muestra de una población? ¿Cuántas personas u objetos tomar de una población para que sea lo más equitativa posible en relación con las distintas escuelas, grupos o sectores que conforman la población?

Para contestar estas interrogantes primero se partirá de los tipos de muestras que se pueden asumir en una investigación.

✦ **Tipos de muestra:**

Hay varios criterios para clasificar las muestras, pero se adoptará el criterio que emiten Freud (1977 y Moráguez (2005), entre otros, por ser uno de los más difundidos y empleados en la actualidad.

Las muestras se agrupan en dos grandes dimensiones: aleatoria y no aleatoria, y dentro de esta se pueden observar otras clasificaciones, siendo estas:

- I. Aleatorio al azar o probabilístico
- 1) Simple
- 2) Sistemático
- 3) Estratificado

Aleatorio Simple: Le da la probabilidad a cada uno de los miembros de una población a ser elegido. Es uno de los más empleados y recomendados en las investigaciones sociales y educacionales, ya que este principio de darle la oportunidad a cada uno de los miembros de la población a ser elegido o tomado como muestra, es lo que permite obtener conclusiones en la muestra e inferir lo que pudiera ocurrir, a partir de esta, en la población, con un elevado grado de pertinencia. Estadísticamente permite inferir a la población los resultados obtenidos en la muestra (Devore, 2000), (Montgomery, 1999), (Siegel, 1997),

Aleatorio Sistemático: Se hace una lista de la población a intervalos fijos, bien sea tomando el coeficiente de elevación (ce) como punto de partida; donde:

$$ce = \frac{\# \text{ elementos de la población}}{\# \text{ elementos de la muestra}}$$

V. g: Si la población P = 100 elementos y la

muestra n= 20, entonces: $ce = \frac{100}{20} = 5$ ¿Qué quiere decir esto?

Indica que cada vez que se produzcan piezas en múltiplos de 5, será seleccionada una para la realización de determinada medición, etc., elementos u objetos producidos (si se tratara de un proceso de producción de piezas).

También se puede extraer de la lista cada enésimo caso. Este método se emplea mucho en los controles de calidad de producciones seriadas y masivas; pero también puede ser empleado en las investigaciones en general.

Aleatorio Estratificado: Es otra variación del aleatorio simple y consiste en subdividir a la población en subgrupos o estratos más homogéneos, de los que se toman muestras aleatorias simples de cada uno de dichos estratos. Hay que evitar que los estratos no se traslapen. (2) (Superpongan o que existan elementos de un estrato en otro).

- 2.1) Por accidente
- II. Muestreo no aleatorio 2.2) { Intencional o de juicio
- 2.3) Por cuotas

2.1) **Muestreo no aleatorio por accidente:** El investigador incluye los elementos que le son más convenientes para la muestra.

2.2) **Muestreo no aleatorio intencional o de juicio:** La idea básica que involucra este tipo de muestra es que la lógica y el sentido común pueden usarse para seleccionar la muestra que sea representativa de una población. Ej. Selección de expertos por el método de experto.

2.3) **Muestreo por cuotas:** Esta se obtiene al especificar las características deseadas de los sujetos que se espera recoger la información y se le deja libertad al investigador para que le aplique los instrumentos necesarios a las personas con esas características. Ej. Se desea hacer un estudio de una población estudiantil de los estudiantes que han repetido el 6. grado y tiene determinada edad o situación en el hogar.

Como en la mayoría de las investigaciones educacionales se trabajan con estratos, los cuales pueden ser escuelas: de una provincia, o de un municipio, grupos de una escuela o de diferentes escuelas..., se dirigirá la atención de este trabajo a exponer de forma práctica cómo seleccionar la muestra de una población conformada por una población de todas los institutos politécnicos del municipio de Holguín, a los efectos de aplicar instrumentos diagnósticos de una investigación acerca de ¿Cómo se ha desarrollado el trabajo metodológico en dichas escuelas?

⊕ Selección de la muestra estratificada a partir de la población seleccionada.-

Para poder seleccionar la cantidad de escuelas politécnicas a tomar como muestra del total de institutos politécnicos (7) del municipio de Holguín, lo cual constituye la población, se va a emplear un estadígrafo, que permite determinar el tamaño de la muestra a partir de la población y teniendo en cuenta el número de estratos a trabajar (en este caso 7, que son los institutos politécnicos). Para ello se plantea una metodología a seguir:

⊕ Determinación de la cantidad de estratos de la población del territorio.

Se relaciona la cantidad de estratos que tiene la población, en este caso son los institutos politécnicos que tiene el municipio de Holguín:

No	Nombre de los Institutos Politécnicos
1	Pedro Díaz Coello
2	Panchito Gómez Toro
3	Camilo Cienfuegos Gorriarán
4	Luis de Feria Garayalde
5	José Gómez Wangüemert
6	Politécnico 26
7	Gral. Calixto García Iñiguez

Obsérvese que a cada estrato (escuela) se le hizo corresponder un número, que comienza por el 1. Este número será constante para cada centro de ahora en adelante.

La cuestión está dada en determinar del total de centros, cuántos se tomarán como muestra aleatoria simple, por lo que para ello se aplicará el siguiente estadígrafo:

⊕ Determinación de la muestra.

$$n_0 = \left(\frac{z}{\varepsilon} \right)^2 * p * q$$

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

(1), (2)

(3)

Donde:

n_0 : Cantidad teórica de elementos de la muestra.

n : Cantidad real de elementos de la muestra a partir de la población asumida o de los estratos asumidos en la población.

N: Número total de elementos que conforman la población, o número de estratos totales de la población.

z: Valor estandarizado en función del grado de confiabilidad de la muestra calculada. Por ejemplo, si se considera trabajar con un 95 % de confiabilidad de la muestra seleccionada, entonces el valor estandarizado a asumir es igual a **1.96**. (Para dos colas).

Algunos valores estandarizados (z) en función del grado de confiabilidad asumido (para dos colas):

Para un: 99 % ----- z = 2, 58 (Empleado con frec.)

95 % ----- z = 1, 96 (El más empleado)

90 % ----- z = 1, 64

Є: Error asumido en el cálculo. Toda expresión que se calcula contiene un error de cálculo debido a las aproximaciones decimales que surgen en la división por decimales, error en la selección de la muestra, entre otras, por lo que este error se puede asumir entre un 1 hasta un 10 %; es decir, que se asume en valores de probabilidad correspondiente entre un 0.01 hasta un 0.1. No obstante, se propone la siguiente tabla para valores óptimos del error para el cálculo del número de estratos de una muestra:

Para $3 \leq N \leq 10$ ----- Se asume $Є = 0.1$ (un error del 10 %).

Para $N > 10$ ----- Se asume $Є = 0.05$ (un error del 5 %).

q: probabilidad de la población que no presenta las características.

Este es un parámetro muy importante, debido a que mediante el mismo se asume qué por ciento o proporción de la muestra no puede presentar las mismas características de la población, debido a diversos factores subjetivos y objetivos de los individuos u objetos que conforman la población. Muchos autores plantean esta probabilidad entre un 1 hasta un 25 %; otros asumen, cuando no se conoce esta variable, el valor máximo de 50 %. Del estudio realizado por este autor se propone la siguiente tabla:

Para $3 \leq N \leq 19$ ----- Se asume $q = 0,01$ (un 1 %).

Para $20 \leq N \leq 29$ ----- Se asume $q = 0,01$ hasta $0,02$ (del 1 al 2 %).

Para $30 \leq N \leq 79$ ----- Se asume $q = 0,02$ hasta $0,05$ (del 2 al 5 %).

Para $80 \leq N \leq 159$ ---- Se asume $q = 0,05$ hasta $0,10$ (del 5 al 10 %).

Para $N \geq 160$ ----- Se asume $q = 0,05$ hasta $0,20$ (del 5 al 20 %).

p: Probabilidad de la población que presenta las características. Dicho de una forma más comprensible, es la probabilidad que tiene la muestra en poseer las mismas cualidades de la población (homogeneidad) y está determinada por:

Como $p + q = 1$ (Probabilidad máxima) $\Rightarrow p = 1 - q$

- ✦ Determinación del grado de confiabilidad y con ello el valor de z.

En el problema en cuestión se asumió un grado de confiabilidad de un 95 %, por lo tanto: $z \equiv 1.96$

- ✦ Determinación del valor del error asumido en el cálculo.

Como el número de estratos (escuelas del municipio de Holguín) es igual a 7, entonces se está trabajando con valores de N menores de 11, por lo que se asume un 10 % (0,1), que es un valor recomendado para muestras pequeñas o menores de 11. Entonces: $\epsilon = 0.1$

- ✦ Determinación del valor de la probabilidad que tiene la muestra de no poseer las mismas cualidades de la población (q).

Del análisis anterior, como el número de estratos es igual a 7, aplicando la tabla para los valores de q, se asume trabajar con el 1 %, luego: $q = 0.01$

- ✦ Cálculo de la probabilidad que tiene la muestra de poseer las mismas cualidades de la población (p).

Como ya se determinó el valor de q (probabilidad de la proporción que no presenta las características), se puede determinar p mediante la expresión: $p = 1 - q$, luego: $p = 1 - q \Rightarrow p = 1 - 0,01 = 0,99$ $p = 0.99$

- ✦ **Cálculo del tamaño de la muestra teórica (n_0).**

Por la expresión (1) se pueden sustituir los valores de cada variable y determinar el valor de

$$n_0 \text{ por: } n_0 = \left(\frac{z}{\epsilon}\right)^2 * p * q \Rightarrow n_0 = \left(\frac{1,96}{0,1}\right)^2 * 0,99 * 0,01 = 3,80 \quad \underline{n_0 = 3.80}$$

- ✦ **Cálculo del tamaño de la muestra real (n).**

Por la expresión (2) se puede sustituir los valores de cada variable y determinar el valor de n por:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{3.80}{1 + \frac{3.80}{7}} = 2,46 \approx 2$$

$n = 2$

Es importante acotar que el valor de N que se toma, corresponde al total de los estratos (cantidad de escuelas politécnicas del municipio de Holguín).

De lo anterior se tiene que de un total de 7 escuelas que constituyen la cantidad total de estratos que tiene la población, considerando un 95 % el nivel de confianza, asumiendo que el error de cálculo (ϵ) sea de un 10 % (0,01) y considerando que solamente el 1 % de la muestra seleccionada no reúna las características de la población ($q= 0, 01$), se determinó que la muestra representativa de dicha población puede ser dos estratos (escuelas).

De lo anterior se infiere que la **representatividad** de una muestra está dada en considerar que la misma fue extraída de una población con un determinado **nivel de confianza** (se trabaja preferiblemente con un 95 % de confianza o más), de asumir un determinado porcentaje en el error de cálculo, que debe estar comprendido entre un 1 hasta un 10 % (0,01 hasta 0,10); y de considerar un adecuado porcentaje (desde un 1 hasta un 20 %) en valores probabilísticos (0, 01 hasta 0, 2) de que la muestra no posee las características de la población. Esto se puede representar de la siguiente forma:

Población a investigar ← Nivel de confianza (95%); error $0.01 \leq \epsilon \leq 0.1 \rightarrow$ Muestra representativa.

Para seleccionar la cantidad de centros a escoger como muestra se trabajó con la siguiente tabla:

Institutos politécnicos del municipio de Holguín

No.	Nombre de escuelas	Cant.	Muest.	Núm, Aleat.
1	Pedro Díaz Coello	1		
2	Panchito Gómez Toro	1		
3	Camilo Cienfuegos Gorriarán	1		
4	Luis de Feria Garayalde +	1	X	290 <u>4</u> 4
5	José Gómez Wangüemert	1		
6	Politécnico 26 +	1	X	065 <u>6</u> 8
7	Gral. Calixto García Iñiguez	1		
Total general		7	2	

+Le correspondió el número aleatorio 29044 (IV bloque, 1. fila). Se escogió el 2, dígito de derecha a izquierda; es decir, el número 4, correspondiendo a esta escuela. (Ver Apéndice VII Teoría y problemas de Estadística de Murray, R. Spiegel, p. 349.

+Le correspondió el número aleatorio 06568 (IV bloque, 2. fila). Se escogió el 2, dígito de derecha a izquierda, es decir el número 6, correspondiendo al Politécnico 26.

Como se puede apreciar, trabajando con un 95 % de confianza, y asumiendo un error de un 10 %, y considerando que la muestra seleccionada, puede no contener iguales propiedades de la población en un 1 %, la muestra a escoger del municipio de Holguín es de 2 centros, los que fueron seleccionados de forma aleatoria, por lo que esto permite hacer inferencias de lo que ocurre en dicha muestra en relación con la población objeto de estudio.

Ahora lo que queda, es ver cómo se seleccionarán cada uno de los estratos que contiene cada centro, según la cantidad de estudiantes a seleccionar por año y especialidades de ambos centros, de manera tal que dicha selección sea proporcional a cada uno de dichos estratos. Se observará cómo proceder:

✦ **Selección del tamaño de cada estrato de las muestras seleccionadas.**

Para ello el autor de este artículo se auxilió de la tabla del anexo 1 en la que se han estratificado los estudiantes de cada centro (previamente seleccionados), distribuidos por años y especialidades, de manera que siguiendo la misma metodología anterior se pueda determinar la cantidad de estudiantes que tendrá la muestra a partir de la población constituida por la matrícula total de ambos centros (1993 estudiantes), -vea el carácter relativo que tiene la población, anteriormente explicado.

Una vez seleccionada la muestra del total de ambas escuelas, el problema está en cómo proceder para distribuir esta de manera equitativa o proporcional, en relación con cada estrato constituido por cada centro, matrícula por año y por especialidades de estos, mediante la aplicación de la metodología planteada.

✦ **Determinación de la muestra para ambas escuelas.**

Como se aprecia en el anexo 1 y la tabla siguiente, la matrícula de dichas escuelas por especialidades y años (estratos) está dada como sigue:

Especialidad	Matric.
Eléctrica	360
Electrónica	515
Artesanía	121
Construcciones Metálicas	27
b) Politécnico 26	
Construcción Civil	400
Viales	126
Geodesia y Cartografía	94
Artesanía	84

Gestión Documentos	94
Bibliotecología	118
Lengua de Seña	39
Albañilería	15
Total escuela	970
Total General	1993

Es importante observar que en este paso interesa solamente la distribución por año y especialidades de ambos centros, así como la matrícula total que constituirá la población a seleccionar de ambas escuelas politécnicas.

Por lo que ahora se determina la muestra a seleccionar de una población de 1993 estudiantes.

✦ **Cálculo del tamaño de la muestra.**

Para calcular el tamaño de la muestra se debe trabajar con las expresiones (1) y (2):

$$n_0 = \left(\frac{z}{\epsilon} \right)^2 * p * q$$

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

Por lo que resulta evidente que hay que determinar los valores de z, ε, q y p, para calcular el tamaño de la muestra teórica y con este valor determinar, en la fórmula (2) el valor real de la muestra a seleccionar de ambas escuelas.

✦ **Determinación del grado de z.**

Se sabe que el valor de z, que no es más que la variable estandarizada para un grado de confianza determinado, que en este caso se asume trabajar con un 95 % de confianza, por lo que si se busca este valor en la función de Excel (Ver anexo 2) como DISTR.NORM.ESTAND.INV(0.975), que equivale a trabajar con la probabilidad de 0.975, ya que si se trabaja con un nivel de confianza del 95 %, quiere decir que el valor de alfa es igual a 0.05 (probabilidad de que no se cumpla el nivel de confianza del 95 %); pero como se trabaja con dos colas, debido a que no se conoce si esta probabilidad es mayor o menor, solamente que es igual o desigual, entonces el valor de alfa (0,05) se divide por dos (dos colas) y este valor se le resta a la probabilidad máxima de que ocurra un hecho (1) y obtenemos el valor de: $1 - 0.025 = 0.975$.

Cuando este valor se busca en la función de Excel DISTR.NORM.ESTAND.INV(0.975) el resultado que genera es igual a (1.96), que no es más que el valor de z estandarizado para la probabilidad del 95 % de confianza. Esto también se puede encontrar en cualquier libro de estadística donde contenga la tabla de distribución normal. Por ejemplo en el anexo 2 se puede apreciar que si en la tabla se entra con la probabilidad de 0.975 se obtiene el valor de $z = 1,96$. (Ver anexo 2). $Z = 1.96$

⊕ **Determinación del error de cálculo.**

Ya se sabe que en todo tipo de cálculo cada vez que se trabaje con números fraccionarios, se tendrá que suprimir determinada cantidad de cifras al aproximar los cálculos efectuados; es por ello que siempre se inducirá un error de cálculo, además de considerar otros tipos de errores al seleccionar una muestra, que puede ser susceptible a la forma de tomar los datos, hacer las mediciones, entre otros. Es de aquí que se debe prever el porcentaje del error que se admitirá en el cálculo de la muestra.

Anteriormente se dijo que Para $N > 10$ (recordar que ahora $N=1993$), se debe asumir el error $\epsilon = 0.05$; que es lo mismo que considerarlo en un 5 %: este es el valor a tomar. $\epsilon = 0.05$

⊕ **Determinación de la probabilidad q.**

Se domina que al realizar el cálculo de una muestra se debe considerar un porcentaje o una proporción de elementos que puedan incluirse en dicha muestra, pero que no reúnan las características de la población, a lo que a esta probabilidad se le ha llamado q y se sugiere que para $N \geq 160$, se considera $q = 0.02$ hasta un 0.2 (un error del 2 al 20 %). Para el cálculo en cuestión se asume $q = 0,08$; es decir, se consideró un 8 %.

$q = 0.08$

⊕ **Determinación de la probabilidad p.**

Como la probabilidad de considerar la proporción de elementos que reúnen las mismas características de la población se determina por la expresión: $p = 1 - q$, entonces al sustituir a q, en la misma tenemos: $p = 1 - 0,08 = 0,92$. $p = 0.92$

⊕ **Cálculo de n_0 .**

Sustituyendo en la fórmula (1) se calcula dicho valor quedando:

$$n_0 = \left(\frac{z}{\epsilon}\right)^2 * p * q = \left(\frac{1.96}{0.05}\right)^2 * (0.92) * (0.08) = 113.09$$

$n_0 = 113.09$

⊕ **Cálculo de n.**

Conocido el valor de la muestra teórica calculada, se procede a determinar el valor de la muestra real mediante el empleo de la ecuación (2) en la que:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{113.09}{1 + \frac{113.09}{1993}} = 107.02 = 107$$

Ello indica que del total de la matrícula de 1993 estudiantes sería suficiente seleccionar 107 de forma aleatoria simple, considerando que se ha trabajado con un 95 % del nivel de confianza, de cometer un 5 % de error y de que la muestra de un 8 % no reúna las características de la población; por lo que se puede considerar a la misma como representativa en estos parámetros seleccionados.

Ahora queda determinar cómo se va a distribuir la muestra a seleccionar entre esas dos escuelas, años y especialidades (estratos), asunto que se resolverá de inmediato.

n = 107

Se deben seleccionar 107 estudiantes de la población de ambas escuelas.

⊕ **Cálculo de la proporción de cada estrato.**

Para ello hay que auxiliarse de la tabla anterior, a la que se le ha incorporado una columna que va a contener la proporción que cada estrato representa en relación con la matrícula total de ambos centros.

Resulta evidente que para obtener la proporción de cada estrato sólo hay que buscar la razón entre las matrículas de cada especialidad y el total de cada escuela contra la matrícula total de ambos centros; así que por ejemplo: para determinar la razón entre la matrícula de la especialidad de Eléctrica de la escuela Luis de Feria, sólo se debe dividir dicha matrícula (360) entre el total de ambas (1993):

Eléctrica (escuela Luis de Feria) = $360/1993 = 0,18$

Electrónica (escuela Luis de Feria) = $515/1993 = 0,26$

En la tabla se puede apreciar cada uno de los valores calculados para cada estrato, por lo que se dejará indicado en la misma. (Ver anexo 1).

a) Luis de Feria			
Especialidad	Matríc.	Prop.	Muestra
Eléctrica	360	0.18	19
Electrónica	515	0.26	28
Artesanía	121	0.06	6

Construcciones Metálicas	27	0.01	3
Total de escuela	1023	0.51	55
b) Politécnico 26			
Construcción Civil	400	0.2	21
Viales	126	0.06	6
Geodesia y Cartografía	94	0.05	5
Artesanía	84	0.04	4
Gestión Documentos	94	0.05	5
Bibliotecología	118	0.06	6
Lengua de Seña	39	0.02	2
Albañilería	15	0.01	1
Total escuela	970	0.49	52
Total General	1993	1	107
Muestra total a seleccionar	107		

Se puede apreciar en la tabla, que las proporciones encontradas permite poder hacer una distribución más racional de la muestra total; por lo que si a esta le corresponde 107, resulta evidente que para determinar la cantidad de estudiantes a seleccionar por especialidad y escuela solamente se debe multiplicar la proporción por el total de la muestra y se obtendrá lo que se busca. Ver a continuación:

Por ejemplo, para la primera proporción de eléctrica (de la escuela Luis de Feria) (vea tabla anterior), si se multiplica: $0,18 \times 107 = 19, 6 = 19$ estudiantes para esta especialidad; lo que quiere decir que se deben seleccionar de esta especialidad a 19 estudiantes en la misma. De una forma análoga se completa la tabla y se obtienen los valores que aparecen en la última columna.

Resulta interesante analizar que cuando la suma de ambas escuelas no llegue al total de la muestra, debido a las aproximaciones decimales con que se ha trabajado, entonces se puede aumentar en uno algunas de los estratos con valores menores, hasta que dicha suma sea igual a la calculada: en este caso 107.

Es importante que esta parte sea trabajada en Excel, ya que permite ahorrar toda una serie de cálculos y además se puede visualizar, en forma de tabla, como la mostrada en el anexo 1.

Al observar dicha tabla (anexo 1), analice cómo se procedió para determinar la proporción de cada uno de los estratos que conforman los años de cada una de las especialidades; por

ejemplo, para determinar la proporción que representa la especialidad de Eléctrica de 1. año de la escuela Luis de Feria que tiene una matrícula en ese año de 81 estudiantes. Por lo que si se quiere determinar la proporción que representa esta cifra en relación con la matrícula total (1993) de ambas escuelas, se tendrá que dividir $81/1993 = 0,04$, que es el valor que aparece en la tabla del anexo 1.

De manera análoga se determinaron cada una de las proporciones de cada año, especialidad y centro, y al multiplicar cada una de estas por la cantidad total de la muestra a seleccionar para ambas escuelas se obtuvo de manera proporcional cómo se debía seleccionar la muestra de cada año, especialidad y escuela.

Es importante acotar que se puede hacer lo mismo, en el caso de que existan varios grupos de un mismo año y especialidad, estableciendo la proporción de la matrícula de cada grupo contra la matrícula del año, y como ya se sabe la cantidad de estudiantes que se debe seleccionar por año, resultaría muy fácil determinar la cantidad de estudiantes por grupo que hay que extraer mediante el método aleatorio simple.

De los aspectos tratados se han analizado los distintos conceptos de lo que es población y muestra, entre otros, y de ello es importante acotar que la población en una investigación es relativa y está en dependencia del campo de estudio a trabajar. Siempre resulta imprescindible dejar bien definido cuál es la población a trabajar, porque de la misma es que será seleccionada la muestra, constituyendo esta un subconjunto de la población y por consiguiente contendrá las mismas propiedades de la primera.

Cuando se selecciona la muestra se debe tratar de que esta selección se haga por el método aleatorio simple o estratificado, si es que de ella se quieren realizar inferencias con respecto a la población investigada.

Es importante apuntar que mediante este trabajo se le enseña a los investigadores cómo poder seleccionar el tamaño de una muestra (para las investigaciones educativas) para que sea significativa en relación con la población a trabajar. Esta significación está dada por tres aspectos muy importantes, que son: el nivel de confianza a trabajar, en lo que se prefiere hacer sobre el 95 % de confianza; el error a considerar en el cálculo, que debe estar entre un 1 hasta un 5 %; y la proporción de considerar que la muestra calculada no posee las características o propiedades de la población. Estos valores deben estar comprendidos entre un 1 hasta un 20 % (en valores de probabilidad), por lo que se debe asumir de un 1 hasta un 2 % para poblaciones o estratos pequeños (menores de 29); de un 2

hasta un 5 % para estratos o poblaciones medianas (entre 30 y 79); y valores comprendidos entre un 10 hasta un 20 % para poblaciones o estratos mayores de 80.

Se anexa a este trabajo una tabla en Excel en donde se puede observar cómo se realiza lo explicado trabajando con este poderoso tabulador electrónico.

REFERENCIAS BIBLIOGRÁFICAS

- ¹ Biblioteca de Consulta Microsoft® Encarta® 2004. © 1993-2003 Microsoft Corporation. Reservados todos los derechos.
- ² Traslapo [traslapo] *m.* 1. Parte de una cosa cubierta por otra. [De *traslapar*] Biblioteca de Consulta Microsoft® Encarta® 2004. © 1993-2003. Microsoft Corporation.
- ³ C. Martínez. Estadística Comercial, p. 223-230.

BIBLIOGRAFÍA

- BIBLIOTECA DE CONSULTA MICROSOFT® ENCARTA® 2004. © 1993-2003 Microsoft Corporation. Reservados todos los derechos.
- CALERO, V. A. Teoría de la estimación. La Habana, Ministerio de Educación Superior, 1981.
- DEVORE, JAY L. Probabilidad y estadística para ingeniería y ciencias. California (Impreso en México), Editorial Thomson Editores, 2000.
- FREUD JOHN, E. Estadística elemental moderna. La Habana, Editorial Pueblo y Educación, 1977.
- MARTÍNEZ, C. Estadística Comercial. Colombia, Grupo Editorial Norma Educativa, 1994.
- MONTGOMERY, C. D. Probabilidad y Estadística aplicada a la Ingeniería. México, Editorial Mc. Graw Hill, 1996.
- MORÁGUEZ I., A. Curso de estadística aplicada a la investigación educacional. Materiales impresos, compendio de tablas y ejercicios adaptados para el curso. Holguín, ISP “José de la Luz y Caballero”, 2005.
- ROJAS SORIANO, R. Guía para realizar investigaciones sociales. México, Editorial Plaza y Valdés, 1998.
- SIEGEL, S. Diseño experimental no paramétrico. La Habana, Editorial Revolución, 1987.
- SPIEGEL MURRAY, R. Teorías y problemas de estadística. La Habana, Editorial Pueblo y Educación, 1977.
- TENA SUCK, E. A. Manual de investigación experimental: elaboración de tesis. México, Editorial Plaza y Valdés, 1996.

DE INTERNET

CÁLCULO DEL PODER ESTADÍSTICO DE UN ESTUDIO.

www.fisterra.com/mbe/investiga/poder_estadistico/poder_estadistico.htm

CÁLCULO DEL TAMAÑO DE LA MUESTRA.

www.ifad.org/gender/tools/hfs/anthropometry/s/ant_3.htm

CURVA NORMAL ESTANDAR.

www.amschool.edu.sv/paes/e8.htm

DETERMINACIÓN DE TAMAÑO MUESTRAL .

www.fisterra.com/material/investiga/8muestras/8muestras.htm

ESTIMACIÓN.

<http://www.e-biometria.com/ene-ctm/tecnicas1.htm>

PÉRTEGA DÍAZ S. y Pita Fernández S. Representación gráfica en el análisis de datos.

Cad Aten Primaria 2001; 8: 112-117. spertega@canalejo.org

SELECCIÓN DE LA MUESTRA.

http://www.filosofia.uanl.mx/hablamty/datosocio.htm#seleccion_tutorialdemuestreo: 3.-

CÁLCULO DEL TAMAÑO DE LA MUESTRA.

www.psico.uniovi.es/Dpto_Psicologia/metodos/tutor.7/p3.html

***Nota de la Editora:**

Este artículo, en la versión original en poder de su autor, contiene 2 anexos y una Hoja de Cálculo de Microsoft Excel. El 1 se denomina "Determinación de la muestra estratificada de las Escuelas Politécnicas del municipio de Holguín. Estadígrafo para determinar la muestra estratificada de estudiantes por escuelas, años y especialidades". El 2, se refiere a la selección de la muestra. En la tabla en Excel se puede observar cómo se realiza lo que explica el autor cuando se trabaja con el tabulador electrónico.

En las normas de "Luz", no se contempla la publicación de anexos; por lo que si algún lector desea revisar este trabajo en su totalidad, puede dirigirse a la dirección electrónica del autor y hacerle la solicitud de envío.

Muchas gracias