ESTIMACIÓN ESTADÍSTICA BASADA EN UN DISEÑO MUESTRAL MULTIETÁPICO (ESTUDIO DE MYPES)

TECNOCIENCIA

STATISTICAL ESTIMATION BASED ON A MULTISTAGE SAMPLE DESIGN (MYPES STUDY)

Bracho, René

René Bracho
rene.bracho@up.ac.pa
Universidad de Panam, Panamá

Tecnociencia
Universidad de Panamá, Panamá
ISSN: 1609-8102
ISSN-e: 2415-0940
Periodicidad: Semestral
vol. 24, núm. 1, 2022
Luis.rodriguez@up.ac.pa

Recepción: 20 Enero 2021
Aprobación: 15 Septiembre 2021

URL: http://portal.amelica.org/ameli/journal/224/2243301001/

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Resumen: El siguiente artículo presenta en lenguaje formal-matemático la aplicación del diseño de muestreo multietápico con su respectivo estimador con la finalidad de construir la noción de estrategia de muestreo y recomendarla en un estudio con encuesta para investigar algunas características financieras de las empresas del país. Siendo las micros y pequeñas empresas (MYPES) las que generan la mayor cantidad de empleos se ha escogido esta escala de empresas para el estudio. Se adapta el estimador insesgado lineal de las técnicas de muestreo usadas a la estructura geográfica del universo de donde proviene la población finita encuestada. Adicionalmente, se contrasta la noción de estrategia de muestreo a la luz de los desarrollos notables de la teoría del muestreo de encuestas y las críticas teóricas al enfoque de inferencia basada en diseño.

Palabras clave: Muestra aleatoria, muestreo multietápico, estrato, conglomerado.

Abstract: The following article presents in formal-mathematical language the application of the multi-stage sampling design with its respective estimator in order to build the notion of sampling strategy and use it in a study with a survey to investigate some financial
characteristics of micro and small companies (MSC) from the country. The linear unbiased estimator of the sampling techniques used is adapted to the geographical structure of the universe from which the finite population surveyed comes from. Additionally, the notion of sampling strategy is contrasted in light of notable developments in survey sampling theory and theoretical criticisms of the design-based inference approach.

Keywords: Random sample, multi-stage sampling, stratum, cluster.

INTRODUCCIÓN

A continuación, se describe en lenguaje formal el diseño muestral multietápico aplicado a una población finita compuesta por un conjunto de micros y pequeñas empresas y utilizando de marco muestral al directorio de estas empresas registrado en el directorio de un ente gubernamental, por ejemplo, la Autoridad de la Micro y Pequeña Empresa (AMPYME) o el Instituto Nacional de Estadística y Censo (INEC) de la Contraloría General de la República.

En la primera sección de este artículo, titulada Método, se explica que el diseño muestral se elaborará utilizando las nociones de muestreo e inferencia para poblaciones finitas. Por ello, se distinguen las diferencias entre la lógica de la inferencia para población infinitas (inferencia estadística clásica) versus la lógica de la inferencia para poblaciones finitas.

Seguido, en la sección de Aplicación y Resultados, se adaptó el diseño muestral a la división política y administrativa del país que sigue a la estructura geográfica. Dicha aplicación se describe formalmente a través de la notación de sumatorias. Esta notación facilita la presentación matemática del estimador de la proporción y la varianza del estimador inherente al diseño. También se muestra el estimador en términos de las probabilidades de inclusión de los elementos en la muestra.

En la sección de Discusión se citan las razones que justifican la formulación de un diseño muestral basado en los desarrollos recientes de la teoría del muestreo de encuestas.

Finalmente, en la Conclusión se destacan los principales resultados del artículo. Es decir, la noción de estrategia muestral y la esencia de los teoremas de la aplicación.

Quizás para un investigador práctico que utiliza el muestreo en su labor profesional y científica la presentación matemática del diseño muestral, el estimador y su varianza sea innecesaria, no obstante, en este artículo la presentación es formal por dos razones:

· Como toda ciencia formal, la estadística utiliza el lenguaje simbólico-matemático para enunciar de manera precisa los elementos y los procesos inherentes al problema investigado reduciendo los trade-off e imprecisiones del lenguaje literal verbal. (Devlin, 2002).

· Los postulados teóricos de las técnicas de muestreo estadístico y el estado del arte de la literatura científica especializada de esta línea de investigación se presentan en notación matemática. (Gutierrez, 2016).

MATERIALES Y MÉTODOS

El método empleado consiste en la definición formal de una estrategia de muestreo. La noción de estrategia de muestreo implica la construcción de un diseño muestral descrito sobre el conjunto de todas las muestras posibles junto a la especificación del estimador que se utilizará para inferir el parámetro poblacional.

Generalmente, en la inferencia estadística clásica se asume:

(#, #, #) → # → (#,

Donde:

Ω: es el espacio muestral (el conjunto de todos los resultados posibles del proceso aleatorio).

ξ: es el conjunto de eventos P: medida de probabilidad X: es la variable aleatoria

#: es el conjunto de los números reales

B#: es el boreliano de los reales sobre el cual se asocia una medida de probabilidad.

Px: Medida de probabilidad asociada por la variable aleatoria.

Es decir, el estudio del proceso aleatorio deseado se modela a partir de un espacio de probabilidades, sobre el cual se define una variable aleatoria que asocia números del conjunto de los reales a las realizaciones del proceso aleatorio y sus correspondientes medidas de probabilidad. (Bouza-Herrera, 2008).

No obstante, para la recolección de los datos y la muestra observada en este caso se requiere la aplicación de una encuesta. Este imperativo práctico introduce la posibilidad de cambiar de la visión de la inferencia estadística clásica en el cual se asume una población hipotética infinita, lo cual permite aplicar el teorema central del límite. Y transitar a un enfoque de inferencia de poblaciones finitas. Dicha modificación implica "asumir que los valores observados corresponden parámetros fijos poblacionales".(Gutierrez, 2016). Es decir, la aleatoriedad deja de estar en las realizaciones de la muestra aleatoria (inferencia estadística clásica) y pasa a descansar sobre la aleatorización predefinida en el diseño muestral (inferencia pob. finitas). Se trasciende de la noción de muestra aleatoria a la de muestra probabilística. (Gutierrez, 2016).

La lógica de la inferencia de poblaciones finitas y de muestras probabilísticas (Gutierrez, 2016) es:

= {#1, #2, ... ##} → # = {#1, #2, ... ##} → #(·) → ## →

En palabras, a partir de una población finita en la que están etiquetados los elementos de la población objetivo a través de un marco muestral, se establece un soporte al cual se le determina un diseño de muestreo que genera probabilidades de inclusión. De estas probabilidades de inclusión o pertenencia de un individuo a la muestra podemos resultar unas funciones del mismo llamadas estimadores.

En consecuencia, los conceptos que son insumos empleados en este enfoque son:

U: es la Población finita (conjunto de elementos identificados con etiquetas).

Q: es el Soporte (conjunto de todas las muestras posibles de la población objetivo).

#(·): Diseño de muestreo (distribución de probabilidad aplicada al soporte)

Ik: probabilidad de inclusión de un elemento a la muestra.

###: Estimador ajustado por un factor de expansión que asegura la representatividad de la muestra.

En base a la anterior secuencia de pasos queda establecida la estrategia de muestreo definida como el par [#(·), ###].

La estrategia de muestreo se aplicará para el diseño de una encuesta que tiene como finalidad el estudio de algunos procesos económicos y financieros de los micros y pequeñas empresas (MYPES) del país registradas en el directorio de empresas de un ente gubernamental, por ejemplo la Autoridad de la Micro y Pequeña Empresa (AMPYME) o el Instituto Nacional de Estadística y Censo (INEC) de la Contraloría General de la República.

RESULTADOS

Sea

= {#####} = {#1, ... , ##}, ‖ # ‖ =N

Una población finita se subdivide en provincias y comarcas dada la división política y administrativa del país, las que se considerarán como estratos, por lo que

$$U = \bigcup_{1 \le l \le L} U_l, U_l =$$

$$= \{u_{l1}, \dots, u_{lNl}\}; \sum_{l=1}^{L} N_l = N$$

En la siguiente etapa se subdivide en conglomerados según distritos dentro de las provincias o comarcas. Así que

$$U_l = \bigcup_{1 \leq j \leq Jl} U_{lj}; U_{lj}$$

$$= \{u_{lj1}, \dots, u_{ljt}\}; \sum_{l=1}^{Jl} N_{lj} = N_l$$

Donde:
U: es la unidad, las micros y pequeñas empresas (MYPES) Subíndice l: indica la provincia o comarca
Subíndice j: es el distrito
Similarmente se establecen conglomerados por corregimientos teniendo que

$$U_l = \bigcup_{1 \leq k \leq Klj} U_{ljk}; U_{ljk} = \{u_{ljk1}, \dots, u_{ljkt}\}; \sum_{j=1}^{K_{lj}} N_{ljk} = N_{lj}$$

Donde:
Subíndice k: indica el corregimiento
En la última etapa, se toma cada área como un conglomerado en el que haremos una evaluación de todos sus elementos.
Cada Mype tiene una condición distinguida por

$$Y_{ljkht} \begin{cases} 1 \; si \; pertenece \; al \; área \; Z \\ 0 \; si \; no \end{cases}$$

Los elementos de cada área se denotan:

$$U_{ljkh} = \{u_{ljkh1}, \dots, u_{ljkht}\}; h = 1, \dots, H_{ljk}$$

Donde:
Subíndice h: indica el área

Se desea realizar estimaciones de la proporción en cada provincia o comarca, distrito, corregimiento y área. Entonces, se obtiene en un área el total de Mypes con la característica Y que identifica a A es

$$A_{iljkh} = \sum_{t=1}^{N_{ljkh}} Y_{iljkht}$$

Así que el total en el corregimiento ljk es

$$A_{ljk} = \sum_{h=1}^{H_{ljk}} A_{iljkh} = \sum_{h=1}^{H_{ljk}} \sum_{t=1}^{N_{ljkh}} Y_{iljkh}$$

En el distrito lj

$$A_{lj} = \sum_{k=1}^{K_{lj}} A_{ljk} = \sum_{k=1}^{K_{lj}} \sum_{h=1}^{H_{ljkh}} A_{iljkh} = \sum_{k=1}^{K_{lj}} \sum_{h=1}^{H_{ljk}} \sum_{t=1}^{N_{ljkh}} Y_{iljkh}$$

En la provincia o comarca l

$$A_l = \sum_{j=1}^{J_l} A_{lj} = \sum_{j=1}^{J_l} \sum_{k=1}^{K_{lj}} A_{ljk} = \sum_{j=1}^{J_l} \sum_{k=1}^{K_{lj}} \sum_{h=1}^{H_{ljk}} A_{iljk\,h} = \sum_{j=1}^{J_l} \sum_{k=1}^{K_{lj}} \sum_{h=1}^{H_{ljk}} \sum_{t=1}^{N_{ljk\,h}} Y_{iljk\,ht}$$

Así que en el país el total es:

$$A = \sum_{l=1}^{L} A_l = \sum_{l=1}^{L} \sum_{j=1}^{J_l} A_{lj}$$
$$= \sum_{l=1}^{L} \sum_{j=1}^{J_l} \sum_{k=1}^{K_{lj}} A_{ljk} = \sum_{l=1}^{L} \sum_{j=1}^{J_l} \sum_{k=1}^{K_{lj}} \sum_{h=1}^{H_{ljk}} A_{ljk\,h} = \sum_{l=1}^{L} \sum_{j=1}^{J_l} \sum_{k=1}^{K_{lj}} \sum_{h=1}^{H_{ljk}} \sum_{t=1}^{N_{ljk\,h}} Y_{iljk\,ht}$$

Se puede representar entonces la proporción de Mypes con el atributo descrito por A como

$$P = \frac{A}{N} = \sum_{l=1}^{L} \frac{N_l}{N} = P_l$$
$$= \sum_{l=1}^{L} \frac{N_l}{N} \sum_{j=1}^{J_l} P_{lj}$$
$$= \sum_{l=1}^{L} \frac{N_l}{N} \sum_{j=1}^{J_l} \frac{N_{lj}}{N_l} \sum_{k=1}^{K_{lj}} P_{ljk}$$
$$= \sum_{l=1}^{L} \frac{N_l}{N} \sum_{j=1}^{J_l} \frac{N_{lj}}{N_l} \sum_{k=1}^{K_{lj}} \frac{N_{ljk}}{N_{lj}} \sum_{h=1}^{H_{ljk}} P_{ljkh}$$
$$= \sum_{l=1}^{L} \frac{N_l}{N} \sum_{j=1}^{J_l} \frac{N_{lj}}{N_l} \sum_{k=1}^{K_{lj}} \frac{N_{ljk}}{N_{lj}} \sum_{h=1}^{H_{ljk}} \frac{N_{ljkh}}{N_{ljk}} \left(\frac{\sum_{t=1}^{N_{ljkh}} Y_{iljkht}}{N_{ljkh}} \right)$$

Debido a la evidente estructura lineal del parámetro poblacional es conveniente estimar insesgadamente (es decir, garantizando que la esperanza matemática del estimador sea igual al parámetro) los correspondientes parámetros para obtener un estimador insesgado. (Pérez, 2000).

En este caso se hace la selección mediante Muestreo Simple Aleatorio Sin Reemplazo (MSASR) de las áreas. (Cochran, 1971). Tal como se ha mencionado anteriormente, estas áreas van a constituir conglomerados monoetápicos (unidades de muestreo en una etapa) por los que Pljkh es evaluada y se convierte en una variable aleatoria debido a la selección aleatoria de las áreas.

En cada corregimiento se tiene Hljkh áreas. De las cuales se selecciona de #####

En tal caso se tiene que:

$$\bar{p}_{ljk} = \frac{1}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} P_{ljkh}$$

Donde:

$\#\#_{\#\#\#}$: es el número de conglomerados (áreas) de la muestra.

P_{ljkh}: son las proporciones de los conglomerados de la muestra.

Es un estimador insesgado en el caso monoetápico de la media de las proporciones pues

$$E(\bar{p}_{ljk}) = E\left(\frac{1}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} P_{ljk\,h}\, \alpha_{lk\,hi} \right), \alpha_{ljk\,h} = \begin{cases} 1 \text{ si } u_{ljk\,h} \text{ está en la muestra} \\ 0 \text{ en otro caso} \end{cases}$$

Como la selección se hizo sin reemplazo esta es una variable Bernoulli,

$$\alpha_{ljkh} \sim B\left(\frac{h_{ljk}}{H_{ljk}} = \pi_{ljk}\right)$$

A partir de este razonamiento se tiene que

$$E(\bar{p}_{ljk}) = \frac{1}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} P_{ljki} E(\alpha_{ljkhi}) = \bar{P}_{ljk} = \frac{1}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} P_{ljkh}$$

Usando este resultado se deriva que

$$\hat{A}_{ljk} = \frac{H_{ljk}}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} P_{ljkh} \qquad \bar{\bar{h}}_{ljk} \bar{p}_{ljk} = \frac{N_{ljk}}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} P_{ljkh} = \hat{A}_{ljk}$$

Es un estimador insesgado del total de Mypes pertenecientes a la clase A en Uljk. Así que un estimador lineal del total, con la condición de insesgadez (donde la esperanza matemática del estimados coincida con el parámetro) será:

$$\hat{A} = \sum_{l=1}^{L} \hat{A} = \sum_{l=1}^{L} \sum_{j=1}^{\bar{j}_l} \hat{A}_{lj}$$

$$= \sum_{l=1}^{L} \sum_{j=1}^{\bar{j}} \sum_{k=1}^{\bar{k}} \hat{A}_{ljk} = \sum_{l=1}^{L} \sum_{j=1}^{\bar{j}_l} \sum_{k=1}^{\bar{k}_{lj}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} \hat{A}_{ljk\,h} = \sum_{l=1}^{L} \sum_{j=1}^{\bar{j}_l} \sum_{k=1}^{\bar{k}_{lj}} N_{ljk}\, \bar{p}_{ljk}$$

Donde:
L : Número de provincias o comarcas (estratos).
#: Número de distritos (conglomerados) seleccionados en la muestra.
##: Número de corregimientos (conglomerados) seleccionados en la muestra.
##: Número de áreas (conglomerados) seleccionados en la muestra.
Así que

$$\hat{P} = \frac{\hat{A}}{N}$$

Es un estimador insesgado de la proporción en U. Por otra parte, la estimación de la proporción en la estratificación (Pérez, 2000) puede ser efectuada usando el estimador insesgado.

$$\hat{P}_G = \sum_{l=1}^{L} \frac{N_l}{N} \frac{1}{n_l} \sum_{i=1}^{n_l} A_{li}$$

$$= \sum_{l=1}^{L} W_l \hat{P}_l$$

Donde:

W_l: es el peso de cada estrato (un término que aglutina a $\frac{N_l}{N}$)

$\hat{P}_l$: es la proporción de cada estrato (un término que aglutina a $\frac{A_{li}}{n_l}$)

Note que dada la estructura lineal de los parámetros es sencillo derivar estos resultados.

Entonces tenemos demostrado el siguiente teorema:

Teorema: Sea una población muestreada en 4 etapas, 1 primera etapa particionada en estratos y en las 3 etapas subsiguientes en conglomerados. Y en la última etapa se aplica MASSR para seleccionar ##### conglomerados. Sea el parámetro de interés una característica que puede ser expresada como una proporción. Un estimador insesgado de P (Pérez, 2000) es

$$\hat{p} = \sum_{l=1}^{L} \sum_{j=1}^{\bar{\bar{j}}_l} \sum_{k=1}^{\bar{\bar{k}}_{lj}} \frac{1}{\bar{\bar{h}}_{ljk}} \sum_{h=1}^{\bar{\bar{h}}_{ljk}} \frac{1}{\bar{\bar{m}}} \sum_{m=1}^{\bar{\bar{m}}} A_{ljkhm}$$

Donde:

L : Número de provincias o comarcas (estratos).

#: Número de distritos (conglomerados) seleccionados en la muestra.

##: Número de corregimientos (conglomerados) seleccionados en la muestra.

##: Número de áreas (conglomerados) seleccionados en la muestra.

##: es la cantidad de Mypes (elementos) del área (tamaño del conglomerado).

El resto de los resultados se puede presentar en el corolario que enunciamos a continuación.

Corolario: Bajo el mismo diseño del teorema anterior son estimadores insesgados de las proporciones en las diversas etapas para los estratos:

$$\hat{P}_G = \sum_{l=1}^{L} W_l \hat{P}_l$$

Recordando que αljkh~B(πljk) se debe analizar algunos aspectos de la distribución de probabilidad envuelta.

$$\alpha_{ljkh} = \begin{cases} 1 \; con \; probabilidad \; \pi_{ljk} = \dfrac{\bar{\bar{h}}_{ljk}}{H_{ljk}} \\ 0 \; con \; probabilidad \; 1 - \pi_{ljk} = \dfrac{H_{ljk} - \bar{\bar{h}}_{ljk}}{H_{ljk}} \end{cases}$$

Así que

$$\alpha_{ljkh}{}^2 = \begin{cases} 1 \; con \; probabilidad \; \pi_{ljk} = \dfrac{\bar{\bar{h}}_{ljk}}{H_{ljk}} \\ 0 \; con \; probabilidad \; 1 - \pi_{ljk} = \dfrac{H_{ljk} - \bar{\bar{h}}_{ljk}}{H_{ljk}} \end{cases}$$

Note que

$$\alpha_{ljkh}\alpha_{ljkh}{}^* = \begin{cases} 1 \; con \; probabilidad \; \dfrac{\bar{\bar{h}}_{ljk}(\bar{\bar{h}}_{ljk} - 1)}{H_{ljk}(H_{ljk} - 1)} \\ 0 \; con \; probabilidad \; 1 - \dfrac{\bar{\bar{h}}_{ljk}(\bar{\bar{h}}_{ljk} - 1)}{H_{ljk}(H_{ljk} - 1)} \end{cases}$$

Como tenemos la insesgadez del estimador en la 4 etapa el error viene dado por su varianza. Esta es:

$$V(\bar{p}_{ljk}) = V\left(\frac{1}{\bar{\bar{h}}_{ljk}}\sum_{h=1}^{\bar{h}_{ljk}} P_{ljkh}\alpha_{ljkh}\right)$$
$$= \frac{1}{\bar{\bar{h}}_{ljk}^2}\left[\sum_{h=1}^{H_{ljk}} P_{ljkh}{}^2 V(\alpha_{ljkh}) + \sum_{h=h^*} P_{ljkh}P_{ljkh^*}Cov(\alpha_{ljkh}, \alpha_{ljkh^*})\right]$$
$$= \frac{(H_{ljk} - \bar{\bar{h}}_{ljk})}{H_{ljk}(\bar{\bar{h}}_{ljk})}\left[\frac{\sum_{h=1}^{H_{ljk}} P_{ljkh}{}^2 - H_{ljk}\bar{P}_{ljk}^2}{H_{ljk} - 1}\right] = \frac{(1 - f_{ljk})}{\bar{\bar{h}}\bar{\bar{m}}_{ljk}} S_{ljkb}^2$$

Entonces el error del estimador propuesto está dado en el teorema que se enuncia a continuación.

Teorema. Bajo el diseño analizado en el teorema anterior se tiene que la cuasivarianza de la proporción (Pérez, 2000) es:

$$S_b^2 = \frac{\sum_l^L \sum_j^J \sum_k^K \sum_h^H \sum_m^{\bar{\bar{m}}} (\bar{X}_h - \bar{X})^2}{H - 1}$$

$$= \frac{\sum_l^L \sum_j^J \sum_k^K \sum_h^H \bar{\bar{m}} * (\bar{X}_h - \bar{X})^2}{H - 1} = \frac{\bar{\bar{m}}}{H - 1}\sum_h^H (P_h - P)^2$$

La varianza del estimador es:

$$V(\hat{p}) = (1 - f) * \frac{\frac{\bar{\bar{m}}}{H - 1}\sum_l^L \sum_j^J \sum_k^K \sum_h^H (P_h - P)^2}{\bar{\bar{h}}\bar{\bar{m}}} = (1 - f)\frac{\sum_h^H (P_h - P)^2}{\bar{\bar{h}}(H - 1)}$$

Donde:

L: Número de provincias o comarcas (estratos).

J: Número de distritos (conglomerados) de la población.

K: Número de corregimientos (conglomerados) de la población. H: Número de áreas (conglomerados) de la población.

##: es la cantidad de Mypes (elementos) del área (tamaño del conglomerado).

DISCUSIÓN

El estudio sistemático de los estimadores y su variabilidad ha dado como fruto el surgimiento de tres enfoques para la inferencia en el muestreo de encuestas: la inferencia basada en diseños, los modelos de predicción e inferencia asistida por modelos. Estos enfoques nacieron en correspondencia con dos de las tres fuentes de aleatoriedad que están presentes en el muestreo de encuestas (Sarndal, Carl., Swesson, Bengt. and Wretman, Jan., 1992):

- La estructura estocástica proveniente de la selección muestral.
- La estructura estocástica proveniente de la naturaleza de los datos.

Tomando en cuenta algunas de las críticas a nivel teórico realizadas sobre la inferencia en el primer enfoque, a saber:

· El no condicionamiento de la inferencia a la muestra extraída. Es decir, en la inferencia basada en diseño la inferencia se lleva a cabo sobre las probabilidades aportadas por el diseño muestral (recolección de datos) y no por las aportadas por la naturaleza de los datos. La siguiente cita es ilustradora: "este enfoque ha sido criticado, sobre la base, que sus inferencias se refieren a repetidas muestras sobre la población encuestada, incluyendo todas las # $\in$ # y sus probabilidades asociadas #(#), en lugar de sólo tomar en cuenta la particular s extraída."(Rao, and Bellhouse, 1990).

· El carácter no informativo de la función de verosimilitud de la inferencia basada en diseño que no permite realizar una inferencia fisheriana sobre los parámetros desconocidos. Es decir:

"El descubrimiento que en el muestreo de encuesta la función de verosimilitud es independiente del diseño muestral, y por lo tanto, de acuerdo al principio de verosimilitud (PV), la inferencia debe ser independiente de las probabilidades del diseño, dio impulso al desarrollo de la teoría del modelo. Esta última para implementar la conclusión del PV, restringe la inferencia y la estimación exclusivamente a las probabilidades dadas por un modelo de superpoblación."(Godambe, 1988).

Hemos deseado en este artículo hacer uso de la noción de estrategia muestral porque:

· Permite plantear la discusión sobre la relevancia del enfoque probabilístico del muestreo, la muestra probabilística y la aplicabilidad del diseño de muestreo en el marco de los desarrollos posteriores en la teoría del muestreo de encuestas y las críticas arriba mencionadas.

· La validez de las inferencias basadas en diseño existe, sólo que hay que tener presente que se trata de inferencias procedentes de la forma particular del diseño muestral, es decir, que resultan de considerar todas las muestras posibles.

· La inferencia basada en diseño sirve de asiento y punto de partida para el enfoque asistido por modelo. Este último subsana las limitaciones del enfoque de diseño al vincular información auxiliar con la variable de interés.

El notable desarrollo y la utilidad de los conceptos de estrategia de muestreo, diseño muestral y técnicas de estimación para el estudio científico de las poblaciones humanas ha posicionado el campo del muestreo de encuestas como una línea de investigación de alta vigencia teórica y práctica en la ciencia estadística.

Incluso, ante la emergencia de disciplinas como la ciencia de datos, las ideas fundamentales del muestreo sirven de herramienta en procesos inherentes al big data como: la preparación de datos, la validación cruzada, entre otros. Adicionalmente, en áreas de aplicación como las

preferencias electorales y el control de calidad en manufacturas el muestreo es útil ante la existencia de restricciones económicas en: recursos humanos, tiempo, finanzas, infraestructura tecnológica y poder computacional.

CONCLUSIÓN

La estrategia de muestreo, el par [#(·), ###], constituye uno de los resultados principales de las ciencias formales al servicio del estudio de los procesos y fenómenos del contexto humano. Es una herramienta poderosa en la que se entrelazan diversos campos de la ciencia estadística. Por un lado, el muestreo como conjunto de técnicas y procedimientos inherentes a la recolección laboriosa y ordenada de datos e información; la estimación como el estudio organizado de la variabilidad de los estimadores. (Sarndal, Carl., Swesson, Bengt. and Wretman, Jan., 1992). Y la inferencia que es la teoría y la práctica de la estimación de las propiedades de una población mediante el uso de una muestra, fundamentándose en las probabilidades.

En los teoremas anteriormente presentados se formaliza la adaptación práctica de los estimadores lineales insesgados usados en la teoría del muestreo a un diseño de muestreo multietápico en el cual se tiene una primera (1era) etapa de estratificación y las tres siguientes etapas de conglomerado. Este diseño fue ideado siguiendo la estructura de la división política y administrativa del país (provincias o comarcas, distritos, corregimientos y áreas), con la finalidad de llevar a cabo la aplicación de una encuesta para estudiar características financieras de la MYPES.

Agradecimientos

Agradecemos al Dr. Carlos Bouza de la Universidad de La Habana, Cuba, por su orientación en el estudio introductorio de la literatura elemental de la teoría del muestreo de encuestas y su guía en la redacción formal de diseños muestrales. También al Lic. Aarón Bracho por sus valiosos consejos para el manejo de fuentes científicas mediante el uso de asistentes digitales.

REFERENCIAS

Bouza-Herrera, C. N. (2008) *'Notas Iniciales Sobre Estimación Puntual En Poblaciones Finitas'*. Universidad de la Habana, Cuba. Avaible at: https://www.academia edu/35111355/NOTAS_INICIALES_SOBRE_ESTIMACI%C3%93N

Cochran, W. (1971) *Técnicas de Muestreo*. 1era Español de la 2da en Inglés. México, DF: Compañía Continental.

Devlin, K. (2002) *El Lenguaje de las Matemáticas*. 2da edn. Barcelona, España: Robinbook.

Godambe, V. P. (1988) 'Survey Sampling'. *The Gold Medalist at the Statistical Society of Canada,* Victoria, Canada, June.

Gutierrez, H. A. (2016) *Estrategias de Muestreo*. Primera Edición. Bogotá, Colombia: Ediciones de la U.

Pérez, C. (2000) *Técnicas de Muestreo Estadístico. Teoría, práctica y aplicaciones informáticas.* Autorizada para México. México: Alfaomega Grupo Editor.

Rao, J. N. K. and Bellhouse, D. R. (1990) 'History and Developments of The Theoretical Foundations of Survey Based Estimation and Analysis.pdf', Survey *Methodology*, June, pp. 3-29.

Sandal, Carl, Swesson, Bengt. And Wretman, Jan. (1992 *ModelAssistedSurveySampling.pdf*. New York, United States of America: Springer-Verlag (Springer).