REVISTA DIGITAL INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

"POBLACIÓN Y MUESTRA"

AUTORIA SILVIA BORREGO DEL PINO
TEMÁTICA ESTADÍSTICA
ETAPA ESTUDIOS UNIVERSITARIOS

Resumen

La Estadística tiene por objeto el desarrollo de técnicas para el conocimiento numérico de un conjunto de datos empíricos (recogidos mediante experimentos o encuestas). Al realizar una investigación estadística tenemos que fijar el conjunto de elementos que queremos estudiar, que llamaremos población o universo. Muestra se denominará a aquél subconjunto limitado extraído de la población, con objeto de reducir el número de experiencias.

Palabras clave

Investigación estadística. Población y muestra.

1. INTRODUCCIÓN

La Estadística actual es el resultado de la unión de dos disciplinas que evolucionan de manera independiente hasta confluir en el siglo XIX: el Cálculo de Probabilidades, que nace en el siglo XVII como teoría matemática de los juegos de azar, y la Estadística, ciencia del estado, que estudia la recogida y descripción de datos y que es de raíces bastante más antiguas.
Esto hace que una de las acepciones más aceptadas de Estadística sea la que la define como el conjunto de métodos que tiene por objeto la obtención, tratamiento y la interpretación de un conjunto de datos de observación relativos a un grupo de individuos o unidades. La Estadística actúa como disciplina puente entre los modelos matemáticos y los fenómenos reales.
Es difícil establecer una cronología exacta de los orígenes de la Estadística. Desde la antigüedad, los estados han recogido información sobre la población y riquezas que existía en sus dominios (censos, inventarios…). Por otra parte, desde el siglo XVII se ha tratado de interpretar fenómenos biológicos y sociales de poblaciones a partir de datos numéricos (tablas de mortalidad, contrastación de la teoría de Darwin, estudio de la herencia humana…) mediante procesos deductivos (Estadística Descriptiva). Desde finales del siglo XIX, aplicando métodos inductivos (Estadística Inferencial), la Estadística ha visto ampliado su campo de aplicación a prácticamente todos los sectores (Ingeniería, Física, Medicina…).
Así, la Estadística se ocupa de la descripción de datos (procedimientos para resumir la información), del análisis de muestras (elegir muestras representativas y hacer inferencia), de la contrastación de hipótesis (comparar predicciones con datos observados), de la medición de relaciones (relación estadística), de la predicción (mediante el estudio del historial de las variables), etc.

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

El método científico se basa en dos tipos de razonamientos: el deductivo (de lo general a lo particular) y el inductivo (de lo particular a lo general). Ambos tipos de razonamientos darán lugar respectivamente a la Estadística Descriptiva y a la Inferencial. Mientras que la Estadística Descriptiva trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones y mediciones, la Estadística Inferencial, haciendo uso del Cálculo de Probabilidades, describe, predice, compara y generaliza resultados a una población estadística a partir de la información que obtiene de una parte de la población.
J. Neyman publicó en 1934 el primer trabajo científico sobre el muestreo de poblaciones finitas, en el que establece que la selección aleatoria es la base de una teoría científica que permita predecir la validez de las estimaciones muestrales.
En la década de los años 40 se desarrollaron numerosos estudios sobre diferentes tipos de muestreo. Lilian H. y William G. Madow fueron los primeros investigadores sobre la teoría de muestreo sistemático. M. H. Hansen y W. N. Hurwitz profundizaron en el estudio de los muestreos estratificado y por conglomerados.
Tras los numerosos trabajos presentados en los años 40 queda definitivamente consolidado el estudio del muestreo de poblaciones finitas, a lo que contribuyeron las publicaciones de Cochran, Hansen, Hurwitz y Madow.

2. POBLACIÓN Y MUESTRA

El primer paso en toda investigación estadística consiste en fijar el conjunto de elementos que queremos estudiar, que llamaremos población o universo. Cada elemento de la población se denomina individuo o unidad estadística. La población puede ser el conjunto de personas de una localidad, las llamadas telefónicas a una central… Llamaremos muestra a un subconjunto limitado extraído de la población, con objeto de reducir el número de experiencias.
Una vez fijada la población debemos indicar cuáles son las características o cualidades que nos interesan estudiar en esa población, estableciendo la forma en la que deben medirse, las unidades de medida…
Estas características observables en una población se clasifican en cualitativas, que son aquellas que no se pueden cuantificar, tales como el color de pelo, el gusto musical, grupo sanguíneo,… Las características que no son cualitativas las llamamos cuantitativas, que son aquellas que sí se pueden cuantificar, como es la estatura, el número de hijos…
A su vez, las características cuantitativas se dividen en dos tipos, las discretas y las continuas. Las características cuantitativas discretas son aquellas que toman valores aislados, como es el número de televisores en una unidad familiar o el número de hijos de una pareja. Por el contrario, las variables continuas pueden tomar cualquier valor comprendido en un determinado rango o intervalo, aunque muchas veces la unidad de medida no nos permita tal hecho. Esto ocurre, por ejemplo, al estudiar la altura de una población, que aunque sabemos que es una variable continua, los aparatos de medida sólo nos permiten tomar éstas con una determinada aproximación.
Algunas veces también es preferible, en el caso de las variables discretas con un gran número de resultados, tratarlas como si fueran variables continuas y viceversa.
Una vez obtenida la información referente a la variable de estudio, ésta se organiza y resume en las llamadas distribuciones de frecuencias, que nos proporcionan el número de individuos que hay para

REVISTA DIGITAL INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

cada uno de los valores de la variable. Estas distribuciones de frecuencias pueden ser de frecuencias absolutas, que nos dicen el número de individuos que presentan un determinado valor de la variable, o de frecuencias relativas, que nos dan el tanto por uno o por ciento de la población que presenta es determinado carácter. En ocasiones, también serán de utilidad las frecuencias acumuladas, en las que cada valor acumula los datos pertenecientes también a los que son menores que él.

3. REPRESENTATIVIDAD DE UNA MUESTRA

En aquellas ocasiones en las que no resulta posible estudiar cada uno de los elementos de una población, ya sea por razones económicas, de rapidez en la obtención de la información deseada, de personal disponible, etc. lo que hacemos es trabajar con una muestra. La característica más importante que debe tener dicha muestra es la representatividad, que es la que garantiza que el estudio realizado en ella pueda ser extrapolado a la población de la que ha sido extraída, proceso que se llama inferencia.
La desventaja o riesgo que conlleva la utilización de muestras es que éstas no sean representativas. Por ello a la hora de seleccionar una muestra hemos de especificar claramente:

- El método de selección de los individuos de la población.
- El tamaño de la muestra.
- Grado de fiabilidad de las conclusiones que vamos a sacar, es decir, una estimación que hacemos del error máximo que se va a cometer.

Diremos que una muestra está sesgada cuando no es representativa de la población, debiéndose a que el método empleado para seleccionar los individuos de la población no ha sido correcto, se han producido errores en las mediciones de las variables, el entrevistador ha sido imparcial, etc.
Dentro de los errores que se pueden cometer debemos distinguir entre:

- Error o sesgo de selección: si algunos miembros de la población tienen más probabilidad de ser seleccionados.
- Error o sesgo por no respuesta: puede que algunos miembros de la población no respondan a las preguntas o no contesten sinceramente. Estos errores son más difíciles de detectar y evitar, aunque se pueden incorporar cuestiones al cuestionario para probar que se ha respondido sinceramente.

En poblaciones homogéneas las muestras son, en general, muy representativas. Por el contrario, en poblaciones heterogéneas, las muestras son poco representativas. Cualquier población real puede ser heterogénea, dándose paradojas como la de Simpson, cuando las conclusiones extraídas de datos divididos en subpoblaciones homogéneas son opuestas a las que se obtienen a partir de datos agregados.
Para finalizar este apartado diremos que existen parámetros que pueden usarse de referencia en el estudio de la representatividad de una muestra. Así, el factor de elevación (cociente entre el tamaño de la población y el de la muestra) representa el número de elementos que hay en la población por

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

cada elemento de la muestra. El factor de muestreo (cociente entre el tamaño de la muestra y el de la población) multiplicado por 100 representa el porcentaje de población que representa la muestra.

4. TIPOS DE MUESTREO

El proceso por medio del cual es elegida una muestra se conoce con el nombre de muestreo. A la hora de seleccionar una muestra hemos de tener en cuenta las características de la población, es decir, su grado de homogeneidad o heterogeneidad con respecto a las variables que analizamos. La clave de un procedimiento es garantizar que la muestra sea representativa de la población. Existen diversos tipos de muestreo, aunque todas las clasificaciones podrían resumirse en probabilístico y no probabilístico. Un muestreo es probabilístico cuando puede calcularse de antemano la probabilidad de obtener cada una de las muestras que sea posible seleccionar, para lo que la selección de la muestra se considera como un experimento aleatorio. Es el único muestreo capaz de dar el riesgo que se comete en la inferencia. Vamos a ocuparnos del muestreo probabilístico, en el que la muestra es representativa, y su proceso de selección constituye un experimento aleatorio.
En los muestreos en los que se selecciona directamente a un individuo se suelen utilizar tablas de números aleatorios, conjunto de números tales que todos los dígitos tienen la misma probabilidad de aparición. Se enumeran a los individuos de la población y se selecciona un número obtenido mediante tablas de números aleatorios o mediante un ordenador que los genere.
Dentro del muestreo probabilístico hemos de distinguir entre:

- Muestreo aleatorio con y sin reemplazamiento.
- Muestreo sistemático.
- Muestreo estratificado.
- Muestreo por conglomerados.
- Muestreo anidado.

4.1. Muestreo aleatorio simple

Se dirá que se ha obtenido una muestra por muestreo aleatorio cuando el proceso de selección de la muestra garantice que todas las muestras posibles que se pueden obtener de la población tienen la misma probabilidad de ser elegidas. Es decir, todos los elementos de la población tienen la misma probabilidad de ser seleccionados para formar parte de la muestra. Cuando un elemento es seleccionado y cuantificadas las características objeto de estudio, vuelve a formar parte de la población y puede volver a ser seleccionado, se dice muestreo aleatorio con reemplazamiento o reposición. Generalmente recibe el nombre de muestreo aleatorio simple.
En el caso de que el elemento no vuelva a formar parte de la población de manera que no puede volver a ser seleccionado se dice que se ha obtenido una muestra por muestreo aleatorio sin reposición o sin reemplazo.
Si bien los dos métodos son distintos, cuando el tamaño de la población es infinito o tan grande que puede considerarse infinito, ambos métodos llegan a las mismas conclusiones. Si el factor de muestreo es mayor de 0´1, la diferencia entre ambos métodos puede ser apreciable, llegando a conclusiones contradictorias según el método que se utilice.

INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS
REVISTA DIGITAL

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

4.1.1. Estudio detallado del muestreo aleatorio simple

Una muestra aleatoria simple de tamaño n puede considerarse como n variables aleatorias independientes e idénticamente distribuidas según una cierta distribución F.
Después de utilizar el procedimiento de muestreo, se querrá proporcionar estimaciones para una o varias variables de interés. Un estimador es intuitivamente una “fórmula matemática” que permite calcular la estimación de una magnitud a partir de los datos observados en la muestra.
Al vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$ de parámetros desconocidos de la distribución F de la población se llama vector paramétrico.
Al conjunto de todos los valores admisibles del vector paramétrico se le llama espacio paramétrico y lo denotaremos por Θ.
Sea $(X_1, X_2, ..., X_n)$ una muestra y sea T $(X_1, X_2, ..., X_n)$ una aplicación $T : R^n \to R^k$.
Se llama estadístico a toda función medible con respecto a las σ-álgebras β^n y β^k que no dependa del parámetro (θ).
Un estadístico T $(X_1, X_2, ..., X_n)$ se dice que es un estimador de θ si toma valores en el espacio paramétrico Θ y además no depende de θ.
El estadístico $\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ se llama media muestral.
El estadístico $S_c^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$ se denomina cuasivarianza muestral.

Teorema 1

Si $\overline{X}$ es el estadístico media muestral de una variable aleatoria X con media μ y varianza σ^2, entonces:

$$E[\overline{X}] = \mu$$
$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

Teorema 2

Si una muestra de tamaño n, $(X_1, X_2, ..., X_n)$ procede de una variable aleatoria X con media μ y varianza σ^2, entonces: $E[S_c^2] = \sigma^2$

Teorema 3

En los muestreos aleatorios simples sin reposición, la varianza de la media muestral vale: $Var(\overline{X}) = \frac{\sigma^2}{n} \cdot \left[\frac{N-n}{N-1}\right]$

INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS
REVISTA DIGITAL

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

Teorema 4

En los muestreos aleatorios simples sin reposición se verifica: $E\left[S_c^2\right] = s^2$

4.2. Muestreo sistemático

Este tipo de muestreo exige, como el anterior, un listado de los elementos de la población. Hay que tener precaución de saber que la población no haya sido ordenada según un criterio que propicie o conlleve la periodicidad en algún carácter. Su ventaja es que no hay que tomar n números aleatorios.
Se parte de un arranque aleatorio "i", número elegido al azar que luego concretaremos, y los elementos que integran la muestra son los que ocupan en la población los lugares i, i + k, i + 2k, …, i + (n – 1)·k. Es decir, se toman los elementos de k en k, siendo k la parte entera del número resultante de dividir el tamaño de la población entre el de la muestra.
El arranque aleatorio i debe ser un número comprendido entre 1 y k, con lo que el último elemento de la muestra es anterior o coincide con el único elemento de la población.
Además, de las $\binom{N}{n}$ muestras posibles del muestreo aleatorio simple, sólo son posibles k muestras sistemáticas, y la probabilidad de cada una de ellas es $1/k$.
El muestreo sistemático es equivalente al aleatorio si los elementos están numerados de manera aleatoria.
Las ventajas de dicho método son:

- Extiende la muestra a toda la población.
- Es de fácil aplicación.

Los inconvenientes que presenta son:

- Aumento de la varianza si existe periodicidad en la numeración de los elementos produciéndose un sesgo en la selección.
- Problemas a la hora de la estimación de la varianza.

4.3. Muestreo estratificado

Se denomina muestreo estratificado a aquel muestreo aleatorio en el que los elementos de la población se dividen en clases o estratos. La muestra se toma asignando un número o cuota de miembros a cada estrato y escogiendo los elementos dentro del estrato por muestreo aleatorio simple, o a veces, por muestreo sistemático.
Las razones para obtener una muestra por muestreo estratificado son:

- Tener información con más precisión en las subpoblaciones de la población sobre la característica objeto de estudio.
- Aumentar la precisión en las estimaciones de las características de toda la población.

En general el muestreo estratificado proporciona mejores resultados que el muestreo aleatorio mientras más diferentes sean los estratos entre sí y más homogéneos internamente.
Básicamente podemos considerar tres métodos para redistribuir el tamaño de la muestra entre los k estratos.

- Proporcionalmente al tamaño de cada estrato.
- Proporcionalmente a la variabilidad de la característica que estamos considerando.
- Se asigna el mismo tamaño a cada estrato.

4.4. Muestreo por conglomerados

Los tres métodos anteriores están diseñados para seleccionar directamente elementos de la población. En este nuevo método, la unidad muestral es un grupo de elementos de la población que llamaremos conglomerados.
Normalmente, estos grupos o conglomerados tienen existencia real. Así, por ejemplo, las personas viven en casas, las casas se agrupan en manzanas, las manzanas en barrios, etc.
Muchas veces no disponemos de una lista de elementos de la población ni los posibles estratos, pero sí el número de conglomerados. Si podemos suponer que cada uno de estos conglomerados es una muestra representativa de la población total respecto a la variable de estudio, se seleccionan algunos conglomerados al azar y, dentro de ellos, se analizan todos sus elementos o una muestra aleatoria simple.
El muestreo por conglomerados es mejor cuanto más homogéneos sean éstos entre sí y cuanto más heterogéneos sean internamente.

4.5. Muestreo anidado

Es una generalización del muestreo por conglomerados. En la primera etapa se seleccionan una serie de conglomerados o unidades muestrales primarias, en una segunda se seleccionan conglomerados más pequeños, pertenecientes a los anteriores, llamados unidades muestrales secundarias, y así sucesivamente cuantas etapas sean necesarias.
El muestreo por etapas tiene la ventaja de que en cada etapa se puede utilizar, según nos interese, el muestreo que se considere más adecuado al tipo de conglomerado de que se trate.

5. TAMAÑO DE LA MUESTRA

En un estudio muestral hay dos problemas fundamentales: ¿cómo elegir la muestra?, visto en el apartado anterior, y ¿cuál debe ser el tamaño de la muestra?

Cuanto mayor sea el tamaño de la muestra, mayor será la precisión obtenida, pero llegará un momento en el que el aumento de los gastos no se verá compensado con un aumento significativo de la precisión.
Dependiendo de la característica que se esté estudiando el tamaño a considerar será uno u otro.
Se define el error muestral o error de muestreo como la desviación típica de la distribución muestral.
Así en el caso de las medias muestrales o de las proporciones tendremos:

a) Cuando la población es finita y la extracción es con reemplazamiento o cuando la población es infinita:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$$

b) Cuando la población es finita y la extracción es sin reemplazamiento:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \qquad \sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Una vez fijado el error máximo admisible $\left(e = k \cdot \sigma_T ; P\left(\left|T - E(T)\right| \leq k \cdot \sigma_T\right) = p_k\right)$, que representa la precisión mínima a exigir en los resultados, y el nivel de confianza, p_k, necesitamos conocer además la variabilidad de la población, ya que cuantos más dispersos estén los valores de la variable en estudio, más arriesgado será utilizar una muestra de pequeño tamaño.
Al nivel de confianza p_k le corresponderá un cierto valor de k, obtenido a partir de la desigualdad de Tchebycheff, bien a través de una distribución Normal, una t de Student, o por otro procedimiento.
Para obtener el tamaño de la muestra, que designaremos por n, en función de e y de p_k, partimos de la ecuación fundamental, $e = k \cdot \sigma_T$, siendo T el estimador utilizado.
Se trata de estimar la media poblacional mediante la media muestral $\bar{X}$ con error máximo admisible e y nivel de confianza p_k o bien estimar la proporción poblacional a partir de la proporción muestral $\hat{p}$.

<u>Tamaño de la muestra para estimar la media poblacional</u>

Para poblaciones infinitas o poblaciones finitas con reemplazamiento, la expresión que relaciona el error máximo admisible e y el error muestral $\sigma_{\bar{x}}$ es:

$$e = k \cdot \sigma_{\bar{x}} = k \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n = \frac{k^2 \cdot \sigma^2}{e^2}$$

Si la población es finita y el muestreo es sin reemplazamiento:

$$e = k \cdot \sigma_{\bar{x}} = k \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \Rightarrow n = \frac{N \cdot k^2 \cdot \sigma^2}{e^2 \cdot (N-1) + k^2 \cdot \sigma^2}$$

INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS — REVISTA DIGITAL

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

No debemos de olvidar que k es el valor asociado a un nivel de confianza (1 – α) y que en las distribuciones normales este valor es $z_{1-\frac{\alpha}{2}}$.

Tamaño de la muestra para estimar la proporción poblacional

Sabemos que el estadístico que se usa para estimar la proporción poblacional es la proporción muestral definido por: $\hat{p} = \frac{X}{n}$ donde $X = X_1 + X_2 + ... + X_n$ con $X_i \rightarrow Ber(p)$

Para poblaciones infinitas o poblaciones finitas con reemplazamiento, la expresión que relaciona el error máximo admisible e y el error muestral $\sigma_{\hat{p}}$ es:

$$e = k \cdot \sigma_{\hat{p}} = k \cdot \sqrt{\frac{p \cdot q}{n}} \Rightarrow n = \frac{k^2 \cdot p \cdot q}{e^2}$$

Cuando no se conoce la proporción p, se estima para el caso más desfavorable, es decir, que tanto p como q valgan $\frac{1}{2}$.

Si la población es finita y el muestreo es sin reemplazamiento:

$$e = k \cdot \sigma_{\hat{p}} = k \cdot \sqrt{\frac{p \cdot q}{n}} \cdot \sqrt{\frac{N-n}{N-1}} \Rightarrow n = \frac{N \cdot k^2 \cdot p \cdot q}{e^2 \cdot (N-1) + k^2 \cdot p \cdot q}$$

No debemos de olvidar que k es el valor asociado a un nivel de confianza (1 – α) y que en las distribuciones normales este valor es $z_{1-\frac{\alpha}{2}}$.

En caso de no conocer la varianza poblacional y un muestreo sin reposición: $e = k \cdot \sigma_{\bar{x}}$ con $\sigma^2_{\bar{x}} = \frac{S^2}{n} \cdot \frac{N-n}{N}$ por lo que $e^2 = k^2 \cdot \sigma^2_{\bar{x}}$, de donde $n = \frac{N \cdot k^2 \cdot S^2}{N \cdot e^2 + k^2 \cdot S^2}$ (I)

Claro es que el tamaño de la muestra ha de ser un número natural y, por tanto, habrá de tomarse para n el valor entero por exceso más aproximado al obtenido.

Los valores de e y p_k (y, por tanto, los de k), los fijaremos con arreglo a nuestros objetivos y limitaciones. El tamaño de la población, N, se supone conocido. En cuanto a S^2, habrá que conjeturarlo o apreciarlo a partir del conocimiento que tengamos de la población en una fecha anterior, o de poblaciones más o menos parecida a ésta, o bien de una muestra piloto previamente seleccionada.

Como dice Wormleighton (1960), cuando no se conoce la varianza puede seguirse uno de los caminos siguientes:

1. Tomar la muestra del mayor tamaño posible. Tiene el inconveniente de que puede dar mayor precisión de la requerida y no ser eficiente el procedimiento por demasiado costoso.
2. Tomar una muestra preliminar o piloto. Tiene el inconveniente de que la información proporcionada por ésta no se aprovecha en la estimación.

INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS — REVISTA DIGITAL

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

Si la población fuese infinita, el tamaño de la muestra viene dado por el límite de (I) cuando $N \to \infty$, que vale $n_\infty = \frac{k^2 \cdot S^2}{e^2}$. Si en (I) multiplicamos los dos términos de la fracción por $\frac{k^2}{e^2}$, tendremos $n = \frac{n_\infty}{1+\frac{n_\infty}{N}} = \frac{n_\infty \cdot N}{N+n_\infty}$.

La expresión anterior representa una hipérbola que pasa por el origen y tiene una asíntota paralela al eje de abscisas, a distancia n_∞ de éste.

Si para n tomamos el entero más aproximado por exceso al valor obtenido con la fórmula, al aumentar N llegaremos a un valor en que cualquier incremento de éste ya no influye en n. Es decir, desde cierto valor de N, n permanece constante. Este valor puede calcularse de la siguiente manera:

$|n_\infty - n| < 1$. Esto equivale a $\left| n_\infty - \frac{n_\infty}{1+\frac{`n_\infty}{N}} \right| < 1 \Leftrightarrow$

$$\left(n_\infty + \frac{1}{N} \cdot n_\infty^2 - n_\infty < 1 + \frac{1}{N} \cdot n_\infty \right) \Leftrightarrow \left(n_\infty^2 < N + n_\infty \right) \Leftrightarrow N > n_\infty \cdot (n_\infty - 1)$$

Esto es, si: $N \geq \frac{k^2 \cdot S^2}{e^2} \cdot \left(\frac{k^2 \cdot sS^2}{e^2} - 1 \right)$ el valor de n será igual a n_∞. Así pues, los pasos a seguir para obtener el tamaño de la muestra son:

1. Obtener el tamaño de la muestra imaginando que $N \to \infty$, o sea $n_\infty = \frac{k^2 \cdot S^2}{e^2}$.

2. Comprobar si se cumple que $N > n_\infty \cdot (n_\infty - 1)$. Si es cierto, el proceso termina aquí y el tamaño de la muestra debe ser hallado para ∞. Si no lo es, hemos de pasar a un tercer paso.

3. Obtener el tamaño de la muestra según la fórmula $n = \frac{n_\infty}{1+\frac{n_\infty}{N}}$.

En caso de no conocer la varianza poblacional y un muestreo con reposición, utilizando la fórmula general para el cálculo del tamaño de la muestra tendremos que $e = k \cdot \sigma_{\bar{x}}$, y según hemos visto, la ecuación quedaría $e = k \cdot \sigma_{\bar{x}} = k \cdot \frac{S}{\sqrt{n}}$. Despejando n tendríamos que el tamaño de la muestra a tomar sería: $n = \frac{k^2 \cdot S^2}{e^2}$

6. CONCLUSIÓN

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008

La Estadística puede definirse como el conjunto de métodos que tiene por objeto la obtención, el tratamiento y la interpretación de un conjunto de datos de observación relativos a un grupo de individuos o unidades.
Dentro de la Estadística pueden considerarse dos grandes ramas: la Estadística Descriptiva, que trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones y mediciones, y la Estadística Inferencial que, haciendo uso del Cálculo de Probabilidades, describe, predice, compara y generaliza resultados a una población estadística a partir de la información que obtiene de una parte de la población.
A la hora de seleccionar una muestra, habrá de tenerse en cuenta criterios que garanticen la representatividad de la misma, esto es, que las conclusiones obtenidas para una población a partir del estudio realizado sobre la muestra tenga un elevado grado de confianza.
Uno de los métodos más utilizados para la selección de una muestra dentro de una población es el denominado muestreo aleatorio simple, en el que todos los individuos de la población tienen la misma probabilidad de ser seleccionados para forma parte de la muestra. Otros tipos de muestreo son el muestreo sistemático, el muestreo estratificado, el muestreo por conglomerados y el muestreo anidado o por etapas.
Por último, cuando realicemos un estudio muestral tendremos que señalar previamente el tamaño de la muestra. En el apartado anterior hemos visto algunos criterios para una selección óptima del tamaño de la muestra.

7. BIBLIOGRAFÍA

Cochran, W. G. (1986). *Técnicas de muestreo*. México: Continental
Cramer, H. (1963). *Métodos matemáticos de Estadística*. Madrid: Ed Aguilar
Cramer, H. (1968). *Teoría de probabilidades y aplicaciones*. Madrid: Ed Aguilar
Fernández – Abascal, H. (1994). *Cálculo de Probabilidades y Estadística*. Barcelona: Ariel
López Cachero, M. (1996). *Fundamentos y Métodos de la Estadística*. Madrid: Ed Pirámide
Renyi, A. (1976). *Cálculo de probabilidades*. Barcelona: Reverté
Ríos, S. (1985). *Métodos estadísticos*. Madrid: Ed del Castillo

8. REFERENCIAS WEB

- www.cnice.mecd.es/.../distribuciones
- www.eio.uva.es
- www.uca.es

Autoría

- Nombre y Apellidos: Silvia Borrego del Pino
- Centro, localidad, provincia: I.E.S. Ángel de Saavedra. Córdoba. Córdoba
- E-mail: depis79@hotmail.com

REVISTA DIGITAL INNOVACIÓN Y EXPERIENCIAS EDUCATIVAS

ISSN 1988-6047 DEP. LEGAL: GR 2922/2007 Nº 12 – NOVIEMBRE DE 2008