

NECESIDAD DE ENCUESTAS POR MUESTREO

Mg. Leonor Laguna A.

RESUMEN

Se presenta la encuesta por muestreo como un método para la obtención de información estadística alternativo a la encuesta por enumeración completa o censo; con ventajas sobre este último respecto a costo, tiempo de ejecución, cobertura de rubros de información, estimación del error de muestreo y mejor control y reducción de los errores ajenos al muestreo. Señala también casos de usos complementarios al censo. Siendo una valiosa característica que el error de muestreo se establece en términos del error mínimo que se puede tolerar en los usos principales de la información.

1. NECESIDAD DE INFORMACIÓN ESTADÍSTICA

Con el transcurso del tiempo, y sobre todo debido al desarrollo de la ciencia y la tecnología, la forma de vida de la población en los diferentes países ha ido cambiando de manera substancial, en diversos aspectos como el económico y el social. Así, la producción de bienes y servicios que se inició básicamente con la finalidad de satisfacer sus propias necesidades, ahora se realiza teniendo en cuenta la especialización y la producción en masa y con miras a obtener el mayor beneficio posible por unidad de costo. La agricultura, la industria, el comercio y otras actividades humanas se desarrollan en base a proyectos en gran escala que requieren planeamiento; esto es, su formulación, ejecución y la evaluación de su efectividad se basan en datos objetivos respecto a recursos y necesidades. Se requiere pues, información estadística recolectada y analizada de manera objetiva, además presentada en forma conveniente de modo que sea una base sólida para la toma de decisiones en diferentes campos.

Por otra parte el rol del Estado, que en un principio se reducía a mantener la ley y el orden dentro del país, se ha ampliado. Cada vez más Estados se involucran en lograr mejores estándares de vida para sus ciudadanos: mejoramiento de la vida económica (regular la producción, políticas adecuadas de importación y exportación, políticas de precios y subsidios, etc.), formulación de programas de bienestar social, sistema educativo en continuo progreso, servicios de salud de acuerdo a las necesidades, etc. Para llevar a cabo sus varias responsabilidades el Estado requiere una gran variedad de información considerando los diferentes sectores de la economía, grupos poblacionales y regiones geográficas del país así como la disponibilidad de recursos tales como mano de obra, área cultivable, bosques, agua, minerales, petróleo.

Como en general los recursos que dispone un Estado son limitados, y también se tiene que pensar en las necesidades futuras del país, para realizar un planeamiento adecuado tiene que recolectarse datos, con cierto detalle, sobre necesidades y la disponibili-

dad de recursos. Por ejemplo para tomar decisiones objetivas respecto a políticas de importación y exportación se requiere datos sobre producción y consumo de diferentes productos de interés. Para diseñar políticas respecto a niveles de sueldos y precios es de importancia conocer información sobre el costo de vida de diferentes grupos de personas que viven en diferentes partes del país. Para planear la mejora de la vida social y económica de la gente se requiere información confiable sobre tasas de nacimiento, muertes y crecimiento de la población, incidencia de enfermedades, nivel nutricional actual, nivel educativo y condiciones de vida, etc.

Todo lo anterior nos da una idea de la masa de datos estadísticos que son muy útiles para el mantenimiento y desarrollo económico y social de una población y nos permite apreciar que la demanda por información estadística es muy grande. Entre los principales usuarios de información estadística se encuentran el Estado, la industria, los negocios, las instituciones científicas, las organizaciones públicas, las agencias internacionales entre otros.

2. TIPOS DE DATOS

Los datos sobre necesidades y recursos que se requieren para el planeamiento y ejecución de proyectos así como la evaluación de su efectividad, se pueden clasificar en dos grupos: (1) Datos de encuestas que ya existen y que pueden ser recolectados por observación o averiguación y (2) Datos experimentales que solamente pueden obtenerse a través de experimentos estadísticos bien diseñados y controlados.

Los datos de encuestas pueden a su vez agruparse en tres clases:

1. Totales de características de interés y relaciones entre dichas características en un punto del tiempo (fecha) o durante un período de tiempo. Ejemplo

de este tipo de conocimiento son datos sobre población humana como número de familias; personas por familia; sexo, edad, nivel educativo y ocupación de las personas; condiciones de las viviendas; etc. Sobre agricultura son datos sobre área dedicada a la agricultura, producción por cultivos, destino de la producción; área cultivada por diferentes cultivos; producción por unidad de área para cada cultivo; etc. Recursos disponibles, demandas y preferencias de los consumidores, factores que contribuyen a satisfacer las necesidades. Como es probable que este tipo de información cambie con el tiempo debe ser recolectada periódicamente.

2. Tendencias y relaciones de diferentes características entre ellas mismas en función del tiempo. Se obtienen en base de datos recolectados periódicamente. Este tipo de datos permite evaluar acciones del Estado en favor de la población, por ejemplo si se trata de reducir el ausentismo escolar en una región, medidas periódicas de variables que midan este hecho permitirán evaluar el proyecto y reajustarlo si es necesario. Las tasas de crecimiento de una población, la reducción de la pobreza son hechos que se miden con este tipo de datos.
3. La información sobre geografía y geología de diferentes partes de un país tales como las condiciones climáticas, tipo de suelo, depósitos de minerales, de petróleo y de gas que no cambian mucho con el tiempo pertenecen a este tipo de datos. Para obtener este tipo de datos se requiere el trabajo intenso de expertos en estos campos y puede requerir mucho tiempo el realizarlo y a la vez es muy costoso; pero una vez que se la obtiene no es necesario obtenerlo otra vez por un buen lapso de tiempo.

Los datos experimentales que como señalamos solo pueden obtenerse por medio de experimentos estadísticos bien diseñados y controlados, son los resultados de tales experimentos. Como ejemplos podemos citar: la tasa a la cual se debería aplicar el abono para maximizar la producción de un determinado cultivo; determinar el insecticida más efectivo en la destrucción de determinada plaga; encontrar el proceso de producción en una industria que rinda la máxima producción por unidad de costo.

3. ENCUESTA POR ENUMERACIÓN COMPLETA O CENSO

Cuando hablamos de datos está implícito el conjunto de elementos o unidades que poseen dichos datos y que se conoce con el nombre de población o universo. Los usuarios de los datos están interesados en ciertas características de las unidades de la población así como las relaciones entre dichas características. Por ejemplo si se trata de la población de un país los datos se pueden referir a las personas y las características son: sexo, edad, área de residencia, nivel educativo, ocupación, etc. Las relaciones serían por ejemplo edad según sexo, nivel educativo según sexo, ocupación según edad, número de habitantes por área, etc.

Una vez determinada la necesidad de contar con cierta información, una manera de obtenerla es recolectando los datos para cada uno de las unidades de la población. Este proceso de obtener información se conoce como encuesta por enumeración completa o censo. El esfuerzo, tiempo y dinero requerido para llevar a cabo este tipo de encuesta es generalmente muy grande. Sin embargo si la información se requiere para cada unidad de la población bajo estudio no hay alternativa, hay que realizar una encuesta por enumeración completa. Es así que periódicamente se realizan en la mayoría de los países los censos de población y de agricultura. Pero hay muchas situaciones en las cuales se re-

quiere información totalizada a nivel del país o por regiones o para grupos de áreas menores, en tales casos la recolección de datos para cada unidad de la población es sólo un medio no es un fin. En estas situaciones la alternativa es recolectar la información solo en una parte de la población, lo cual reduce el esfuerzo, tiempo y dinero requerido en forma significativa. Por tanto antes de decidir qué procedimiento de recolección de datos utilizar hay que considerar los siguientes puntos:

1. ¿Cuál es el objetivo de la recolección de datos: obtener datos para cada una y toda unidad de la población o sólo se requiere información totalizada para todas las unidades tomadas en conjunto o para grupos de unidades?
2. Si se requiere datos para cada uno de las unidades de la población hay que llevar a cabo una encuesta por enumeración completa. Los datos para cada unidad de la población son necesarios cuando se van a tomar acciones separadas para cada una de ellas. Ejemplos de tales situaciones son la imposición de impuestos, en cuyo caso se evalúa el ingreso de cada individuo y se le impone el respectivo impuesto; la preparación de la lista de votantes para llevar a cabo una elección. Sin embargo, hay numerosas situaciones en las cuales el interés está sólo o principalmente en información resumen para subgrupos de unidades o para todas las unidades de la población. Para estos casos hay métodos alternativos que permiten recolectar y procesar la información deseada en forma más económica y operacionalmente más conveniente. Por tanto es de suma importancia establecer claramente el objetivo del proyecto de recolección de información.

¿Es necesario obtener información exacta, esto es, sin ningún error, para los usos que se le va a dar a la información recolectada?

Se requiere examinar cómo generalmente es usada la información estadística deseada por los usuarios en la toma de decisiones, considerando sus actividades. Hay que observar que un planeamiento exacto para el futuro no es posible, desde que esto requeriría información precisa sobre los recursos que estarían disponibles y sobre las necesidades que tendrían que ser satisfechas en el futuro. En general, se utilizan datos pasados para pronosticar los recursos y necesidades del futuro y por tanto hay un elemento de incertidumbre en el planeamiento. Debido a esta incertidumbre, generalmente sólo se intentan afijaciones aproximadas (y no exactas) de los recursos. Así, en los datos estadísticos requeridos para planificación puede permitirse algún margen de error, provisto que este error no sea suficientemente grande para afectar las afijaciones que se han hecho. Además, una gran incertidumbre proveniente de fuentes no estadísticas involucradas en el tratamiento de problemas socioeconómicos hace posible arribar sólo a decisiones aproximadas y esto a su vez permite un margen de error en los resultados estadísticos. Desde que un planeamiento exacto es, en general, impracticable hay un grado de incertidumbre acerca del resultado de los proyectos planeados y por tanto algún margen de error es permisible en la información estadística requerida aun en casos de evaluación del avance pasado y presente de los proyectos así como de su efectividad. El margen de error que es permisible en vista de la incertidumbre mencionada puede denominarse el error permisible.

3. ¿Una encuesta por enumeración completa nos proporciona siempre información exacta?

Hay experiencias censales en un número de países, que evidencian que una

encuesta por enumeración completa no necesariamente proporciona información precisa. Esto demuestra que aun en un censo de población, donde el principal propósito es precisamente obtener un conteo completo de personas, comparativamente una tarea fácil desde el punto de vista de la recolección y procesamiento de datos, los datos están sujetos a alguna medida de error. De modo que la posibilidad de errores es mayor en encuestas por enumeración completa más complejas, especialmente en situaciones en las que tanto investigadores como investigados requieran entender conceptos y definiciones difíciles respecto a las características involucradas. Los errores en una encuesta por enumeración completa se presentan principalmente debido a cobertura incompleta, a errores de observación y de tabulación debido a dificultades encontradas al organizar una encuesta en gran escala y en obtener personal adecuadamente entrenado para llevar a cabo la encuesta.

4. ¿Bajo qué circunstancias una encuesta por muestreo, esto es, una encuesta de una parte de la población, se prefiere a una encuesta por enumeración completa?

La respuesta se da en la siguiente sección.

4. NECESIDAD DE ENCUESTAS POR MUESTREO

Hemos visto que la información obtenida para propósitos prácticos puede estar sujeta a un margen de error determinado y por otra parte que una encuesta por enumeración completa no necesariamente proporciona información precisa. De modo que al planear un proyecto de recolección de datos determinado tenemos que resolver los siguientes puntos:

1. Determinación del error permisible.

El margen de error permisible se determina en función del margen de ajuste posible en las decisiones a tomar.

2. Encontrar un método eficiente de encuesta que asegure este específico margen de error permisible al costo mínimo.

Como un asunto previo a plantear una alternativa a la encuesta por enumeración completa vamos a considerar el caso de encuestas destructivas. Un ejemplo es la estimación del promedio de vida de un lote de focos eléctricos, con tal fin la observación consiste en hacer funcionar el foco hasta que se queme. En tales casos, por necesidad uno tiene que restringir las observaciones a una parte (muestra) de la población o universo, e inferir acerca de la población como un todo en base a las observaciones de la muestra. Pero aun en otras situaciones una alternativa efectiva a una encuesta por enumeración completa puede ser una encuesta por muestreo, para la cual se seleccionan de manera conveniente sólo algunas de las unidades de la población y se efectúa inferencias acerca de la población en base a las observaciones hechas en las unidades seleccionadas. Comparada con una encuesta por muestreo, una encuesta por enumeración completa consume más tiempo, es más costosa y tiene menos alcance en el sentido de ser restringida en la cobertura de materias a investigar y está sujeta a mayores errores de cobertura, observación y tabulación. Además una encuesta por enumeración completa no es del todo posible cuando no se dispone la gran cantidad de recursos en términos de personal entrenado y financiamiento requerido para tal encuesta. En cambio en una encuesta por muestreo como se encuesta sólo una parte de la población hay una mayor posibilidad de acción para eliminar errores de indagación o de observación y procesamiento, mediante

controles adecuados y empleando personal bien entrenado; lo que en una encuesta por enumeración completa sería muy costoso y a veces imposible.

Por otra parte, como al aplicar una encuesta por muestreo se hacen inferencias acerca del todo en base a observaciones en sólo una parte, los resultados pueden ser diferente de los valores poblacionales y las diferencias dependerán de la parte seleccionada o muestra. Así, se ve que la información proporcionada por una muestra está sujeta a una clase de error que se denomina error de muestreo. Una encuesta por muestreo se lleva a cabo de acuerdo a la Teoría de Muestreo que es una rama de la Teoría Estadística basada en la Teoría de Probabilidades y que permite no solo estimar los valores de las características de la población en base de los datos de la muestra, sino también obtener un estimado válido del error de muestreo de las estimaciones obtenidas.

Cabe mencionar que el concepto de muestreo no es de origen reciente, puesto que desde tiempos inmemoriales consciente o inconscientemente recurrimos al muestreo en la toma de decisiones en la vida diaria. Por ejemplo, el comerciante examina muestras de granos tomados de los sacos de granos que posee para determinar la calidad de todas sus existencias y el ama de casa examina una cucharada del plato que ha preparado para determinar si ha sido adecuadamente cocinado. También un médico que desea conocer la relación de glóbulos rojos y blancos de un paciente solicita que lo determinen en base a una muestra de unos cuantos mililitros de sangre del paciente. Sin embargo, el desarrollo de la Teoría de Muestreo junto con su adecuada aplicación a problemas prácticos se ha realizado mayormente en el siglo XX y sigue desarrollándose de acuerdo con los avances de la tecnología, es así que ahora tenemos por ejemplo encuestas por internet.

5. ERRORES DE MUESTREO Y ERRORES AJENOS AL MUESTREO

Hemos denominado como error de muestreo al error que se presenta por hacer inferencias acerca de la población en base a observaciones hechas en una parte de ella que denominamos muestra. El error de muestreo en este sentido es no existente en una encuesta por enumeración completa, desde que toda la población es encuestada. El error de muestreo decrece al crecer el tamaño de la muestra y en efecto en muchas situaciones el decrecimiento es inversamente proporcional a la raíz cuadrada del tamaño de la muestra (Ver Figura 1). Observamos en la figura que si bien en un principio al crecer el tamaño de la muestra el error de muestreo decrece en forma substancial, llega un momento en que se hace marginal. Por tanto, a partir de esta etapa, se puede alcanzar una reducción apreciable en el costo bajando aun ligeramente la precisión requerida. Desde este punto de vista, hay una fuerte razón para recurrir a una encuesta por muestreo para proporcionar estimados con márgenes de error permisibles en vez de una encuesta por enumeración completa, en la cual el error de muestreo es cero pero el costo requerido para organización y levantamiento de la encuesta será sustancialmente más alto.

Los errores que principalmente se presentan en las etapas de recolección y procesamientos de datos, y a los que se denominan errores ajenos al muestreo, son comunes a las encuestas por enumeración completa y a las encuestas por muestreo. Sin embargo, en una encuesta por muestreo debido al menor volumen de trabajo se puede organizar mejor la encuesta, se pueden mejorar los métodos de levantamiento y procesamiento de datos, entrenar mejor al personal de campo así como ejercer una buena supervisión y por tanto se espera que en ella los errores ajenos al muestreo sean menores.



Figura 1. Comportamiento del error de muestreo al crecer el tamaño de la muestra

Por todo lo comentado podemos apreciar que la necesidad de encuestas por muestreo es una evidencia.

6. ASPECTO DEL COSTO

Vamos a examinar las circunstancias bajo las cuales una encuesta por muestreo se prefiere a una encuesta por enumeración completa o censo desde el punto de vista del costo y del error. Desde que una encuesta por enumeración completa puede considerarse como un caso particular de una encuesta por muestreo, cuando el tamaño de la muestra es igual al número total de unidades de la población, podemos plantear que el tamaño de conveniente es el que minimiza la ecuación:

$$T(n) = C(n) + L(n) \quad (1)$$

Donde $C(n)$ es el costo de la encuesta y $L(n)$ es la pérdida involucrada al tomar decisiones con los resultados de la encuesta.

La pérdida $L(n)$ depende del margen de error que consiste tanto del error de muestreo como del error ajeno al muestreo. La teoría de encuestas por muestreo trata de los procedimientos para seleccionar las unidades a ser incluidas en la muestra y para estimar el valor de la característica para la población como un todo en base a los datos recolectados para las unidades en la muestra, y de comparar la eficiencia de estos pro-

cedimientos con referencia a costo y error.

Supongamos que deseamos estimar la producción total de uvas que producen N haciendas. Si realizamos una encuesta por enumeración completa o censo:

E_c % es la proporción en que los resultados del censo difieren del valor verdadero, diferencia que se debe a la presencia de errores ajenos al muestreo.

Si alternativamente tomamos una muestra de n haciendas convenientemente seleccionadas

E_m % es la proporción en que los resultados de la encuesta por muestreo difieren del valor verdadero E_m % tiene dos componentes: E_{m1} % debido al error de muestreo y E_{m2} % debido a los errores ajenos al muestreo.

C_c Costo de recolectar y tabular los datos por hacienda en el censo.

C_m Costo de recolectar y tabular los datos por hacienda en la encuesta por muestreo.

L Pérdida por 1% de error.

Si la pérdida es proporcional al error, una encuesta por muestreo se preferirá a una encuesta por encuesta por enumeración completa si

$$NC_c + LE_c > nC_m + LE_m \quad (2)$$

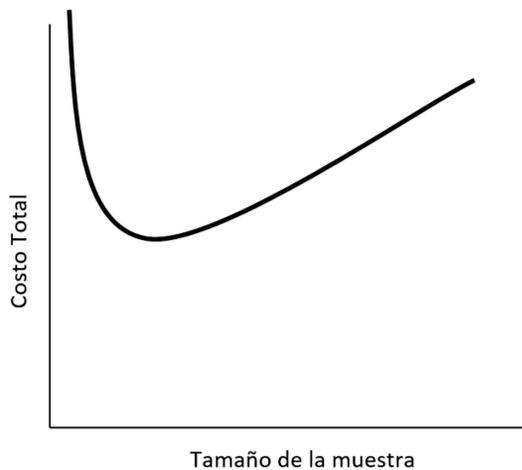
Desde que en una encuesta por muestreo es deseable y posible utilizar personal mejor entrenado y ser más cuidadoso en la recolección de datos de lo que es posible en una encuesta por enumeración completa, se espera que C_m sea más grande que C_c , aunque es posible que nC_m sea menor que NC_c .

Además, es posible que en muchas situaciones prácticas los errores de muestreo y los errores ajenos al muestreo E_m % en una encuesta por muestreo para un tamaño de muestra n conveniente determinado podría ser menor que los errores ajenos al muestreo E_c % en una encuesta por enumeración completa. Esto demuestra que puede haber muchas situaciones prácticas en las cuales una encuesta por muestreo puede ser preferida a una encuesta por enumeración completa en base a una consideración conjunta de costo y de pérdida debido a decisiones erradas.

Desdoblando el error E_m % cometido en una encuesta por muestreo en sus componentes error de muestreo E_{m1} % y error ajeno al muestreo E_{m2} %, el costo total para cualquier tamaño de muestra n puede escribirse como

$$nC_m + LE_{m1} + LE_{m2} \quad (3)$$

Puede visualizarse que cuando n crece los componentes nC_m y LE_{m2} tenderían a crecer, mientras que LE_{m1} se espera que decrezca. Esto significaría que el costo total incluyendo la pérdida debido a decisiones erradas es probable que sea grande para una muestra de tamaño uno, desde que en tal caso el error de muestreo y por tanto la pérdida LE_{m1} sería grande. El costo total $T(n)$ decrecería al crecer hasta una cierta etapa, después de la cual el costo de la recolección y procesamiento de datos y la pérdida debido a los errores ajenos al muestreo se hacen suficientemente grandes para cancelar el decrecimiento en pérdida debido a la disminución del error de muestreo, resultando finalmente en una tendencia de crecimiento del costo total (ver Figura 2).



El valor del tamaño de muestra n para el cual el costo total alcance su valor mínimo, se considera como el **tamaño óptimo de la muestra**. Sin embargo puede haber situaciones, en las cuales el gráfico del costo alcanza el valor mínimo cuando n es casi igual al número total de unidades de la población, indicando que en tales casos son preferidas las encuestas por enumeración completa. Tales situaciones aunque no muy comunes, pueden ocurrir cuando el costo de la recolección y tabulación de datos por unidad es bastante pequeña y la posibilidad de errores ajenos al muestreo es insignificante.

El método anterior de elegir entre las dos alternativas, encuestas por enumeración completa y encuesta por muestreo, o más generalmente determinar el tamaño óptimo de la muestra es en general muy difícil en la práctica, desde que no es fácil determinar objetivamente la forma de la función de pérdida y los valores de las constantes involucradas en ella. Además, para ser más realistas, tenemos que incluir en el costo total otros factores tales como pérdida debido a demora en proporcionar los resultados de la encuesta, la cual es posible que crezca con el tamaño de la muestra, siendo el máximo para una encuesta por enumeración completa. Aquí nuevamente, es difícil de determinar en la práctica el valor de la pérdida para diferentes

períodos de retraso en proporcionar los resultados de una encuesta.

Puede también señalarse que en la práctica una decisión sobre el método de muestreo y el tamaño de la muestra no depende sólo de consideraciones de error y costo, sino también de otros factores tales como disponibilidad de personal calificado, habilidad de los informantes para proporcionar la información requerida, el marco organizativo y administrativo existentes entre otros. En general se ve que costo no es la única limitación en el plan de una encuesta y aquella posibilidad, que depende de factores tales como los mencionados, es de mucha importancia en el éxito del plan de una encuesta.

7. ENCUESTAS POR MUESTREO Y POR ENUMERACIÓN COMPLETA

Como hemos visto las encuestas por muestreo y enumeración completa son métodos para la obtención de una gran variedad de información requerida por los usuarios de información estadística. Ambos métodos requieren para su ejecución un mínimo de facilidades como personal profesional especializado para el planeamiento y la dirección de la encuesta, supervisión del trabajo de campo, marcos muestrales (listado de unidades, listado de área-unidades y mapas para su ubicación, etc.) personal de encuestadores bien entrenados, equipo de tabulación, facilidades de transporte y comunicaciones, etc. Tales facilidades no siempre se disponen en la cantidad requerida para una encuesta por enumeración completa, por lo que en tales situaciones una encuesta por muestreo es una posibilidad a considerarse.

En una encuesta por enumeración completa se puede hacer tabulaciones a nivel de cualquier unidad administrativa, independientemente de cuan pequeñas sean y son libres de errores de muestreo y se espera que los errores ajenos al muestreo hayan sido controlados. La necesidad de datos para uni-

dades administrativas pequeñas puede ser el resultado de necesidades y condiciones especiales en el planeamiento del desarrollo económico y social de un país. En tales situaciones se puede aplicar encuestas por enumeración completa para obtener los datos sobre los rubros de información requeridos para unidades pequeñas y una encuesta por muestreo para otra información también necesaria a nivel de unidades administrativas más grandes.

La experiencia ha demostrado que una encuesta por muestreo adecuadamente planeada puede proporcionar información suficientemente precisa sobre agricultura, ganadería, industria y otras actividades económicas de un país, tabulada a nivel de unidades administrativas grandes y para cortes clasificatorios amplios de las características bajo estudio.

Una encuesta por muestreo se hace necesaria, cuando se requiere información sobre características cuya recolección y procesamiento implica serios sesgos y errores ajenos al muestreo y las medidas para evitarlas requiere de un entrenamiento adecuado y cuidadoso del personal que participe en la encuesta. La disminución del volumen de trabajo va a ayudar a perfeccionar los procedimientos a establecer.

Otra manera de combinar los dos tipos de encuestas es en la etapa de tabulación de los resultados de un censo; se puede adelantar los resultados en base a la tabulación de los resultados de una muestra de unidades.

En resumen, una encuesta por muestreo en comparación con una encuesta por enumeración completa es menos costosa, requiere menos tiempo para su ejecución, mayor cobertura en materias de información, mayor flexibilidad operacional, entrega de resultados más rápida, errores ajenos al muestreo mejor controlado y por tanto más pequeño. En cuanto al error de muestreo de los datos

obtenidos en base a una encuesta por muestreo adecuadamente diseñada puede estimarse en base a los datos de la encuesta.

BIBLIOGRAFÍA

1. M. N. MURTHY, **Sampling Theory and Methods Statistical**, Publishing Society Calcutta, India, 1977.
2. Tore DALENIUS, **Elements of Survey Sampling**, SAREC, 1985
3. C. R, RAO, **Evolution of Data Collection Censuses, Sample Surveys and Design of Experiments**, University of Pittsburgh, 1985.



