

MÉTODOS DE OBTENCIÓN DE ESTADÍSTICAS AGRARIAS

Por

LEANDRO CASTRO RODRIGUEZ

Ingeniero Agrónomo

S U M A R I O :

Introducción.—Tipos genéricos de estadísticas agrarias.—ESTADÍSTICAS DISCONTINUAS.—Censos.—Estudios estadísticos sobre problemas concretos.—Enlace entre los censos y las estadísticas continuas.—Contribuciones especiales de los censos a las estadísticas continuas.—ESTADÍSTICAS CONTINUAS. Subproductos de la actividad administrativa.—Opinión de expertos.—Investigaciones por muestreo.—Investigaciones por muestreo no probabilístico.—Crítica de las investigaciones por muestreo no probabilístico.—Investigaciones por muestreo probabilístico.—Muestreo probabilístico de listas.—Muestreo probabilístico de áreas.—Las sobrevisiones muestrales comparadas con los censos.

INTRODUCCIÓN.

Uno de los problemas básicos del mundo de hoy, que tendrá que resolverse en los años futuros, es el de mantener el equilibrio entre la producción y consumo de alimentos y materias primas agrarias de acuerdo con el crecimiento de la población. En la solución de este problema intervienen actividades humanas muy diversas, entre las cuales la estadística tiene un papel importante.

La estadística, al suministrar una información real de la economía agraria, permite tomar decisiones bien fundadas y sirve para el planeamiento y ejecución de programas efectivos. La importancia de una estadística adecuada nunca será destacada suficientemente.

Durante los últimos años el mundo ha progresado de una manera rápida y constante gracias a los avances de la ciencia y de la técnica; la ciencia estadística ha seguido un proceso paralelo

que todavía tiene hoy un empuje extraordinario. La comparación de la estadística de 1900, 1930 y actual pone de manifiesto un avance realmente asombroso, que parece continuará en el próximo futuro.

Nos referimos aquí exclusivamente a las estadísticas agrarias, cuya recolección constituye una labor de gran magnitud. No obstante, la metodología que se expone es válida en un 90 por 100 para cualquier tipo de estadísticas. Consideraremos a continuación los métodos, que hoy están al alcance de las organizaciones estadísticas de todo el mundo, para conseguir la información deseada que sirva mejor a las necesidades del Gobierno y del país.

TIPOS GENÉRICOS DE ESTADÍSTICAS AGRARIAS.

Las estadísticas agrarias pueden agruparse en dos tipos fundamentales:

- 1.º Estadísticas discontinuas, entre las que se incluyen los censos y estudios detallados sobre problemas concretos.
- 2.º Estadísticas continuas, que pueden ser anuales, trimestrales, mensuales, etc.

ESTADÍSTICAS DISCONTINUAS

CENSOS.

Utilizamos aquí la palabra censo con el significado tradicional de una investigación estadística completa realizada mediante el contacto con cada uno de los individuos del universo. La organización técnico-administrativa necesaria para los censos es, en general, de carácter transitorio, establecida especialmente para el caso o resultante de una expansión de servicios ya existente; esto último es lo que se hace en España.

Históricamente, el censo es el instrumento más antiguo preparado para la recogida de datos. Como se establece contacto con cada individuo del universo, el cálculo de totales y promedios constituye una tarea laboriosa y de elevado coste, aunque no presenta grandes dificultades desde el punto de vista de la teoría estadística; las dificultades de los censos suelen ser de tipo administrativo y de organización, y sobre todo financieras.

Los censos suelen tomarse periódicamente, a intervalos más o menos regulares. En la actualidad existe un programa de la F. A. O. para la elaboración mundial de censos agropecuarios decenales. Todos los países tratan de ajustarse en lo posible a estos intervalos, si bien en aquellos en que la estadística está más adelantada es frecuente un tipo de intervalo menor, que suele ser de cinco años. Existen casos, muy poco frecuentes, de censos anuales sobre alguna actividad concreta, pero este procedimiento de obtener las estadísticas continuas no parece recomendable por varias razones, como se verá más adelante.

ESTUDIOS ESTADÍSTICOS SOBRE PROBLEMAS CONCRETOS.

Estos trabajos son, generalmente, de gran detalle, pero de poca amplitud. Sirven para llenar las lagunas censales y ayudar al mantenimiento de las estadísticas continuas. Están indicados siempre que exista la necesidad de una información muy detallada sobre problemas concretos, y sus técnicas pueden ser de tipo censal u otras.

ENLACE ENTRE LOS CENSOS Y LAS ESTADÍSTICAS CONTINUAS.

Consideraremos aquí nada más que los principios de orden teórico que rigen esta vinculación.

El primer resultado del enlace entre las respectivas actividades es el promover un intercambio beneficioso para los dos sectores, y lo es tanto más cuanto más estrecha sea la coordinación que se logre entre los dos programas. Las estadísticas continuas que cubren las necesidades entre dos levantamientos censales consecutivos permiten afinar conceptos y revisar procedimientos que parecieron precarios en el último censo levantado y preparar mejor el censo siguiente, en el que puede contarse con la experiencia de las técnicas del sistema continuo.

El censo forma un conjunto de materiales básicos de referencia, necesarios para las estadísticas continuas, y determina el conocimiento de la estructura agraria en la forma más exacta posible. Esto permite abaratar y mejorar las estadísticas continuas, porque las estimaciones necesarias para confeccionar las últimas se refieren solamente a los cambios habidos en años sucesivos, a partir del nivel obtenido en el censo. Esto da lugar a que técnicamente

sea necesaria una información mucho menor para conseguir la exactitud deseada. Los censos y las estadísticas continuas pueden utilizar los mismos métodos de recogida de datos y elaboración de los mismos, aunque a veces la recogida se hace de forma diferente. La mayor diferencia entre un censo y una estadística continua es el volumen de la información a utilizar y, por lo tanto, de coste.

La coordinación del censo con las estadísticas agrarias continuas determina previamente cuáles han de ser los conceptos comunes y exclusivos de cada uno de ellos. Aquellos conceptos que presenten una relativa estabilidad en el tiempo deben ser incluidos exclusivamente en los programas censales. No es razonable sobrecargar los programas de estadística continua con obtención de datos de variación lenta, porque ello supone un aumento de gastos no compensados por la información a obtener y un retraso en la tabulación y presentación de resultados que no corresponden al interés inmediato de los usuarios de los mismos. Cuando se verifique que la variación usualmente lenta se acelera, estos conceptos exclusivos pueden pasar eventualmente a ser parte del programa de las estadísticas continuas.

Los conceptos de gran variabilidad no deben incluirse con mucho detalle en los censos, porque la información así obtenida resulta de un coste tan elevado que la hace antieconómica. Este tipo de conceptos deberá incluirse solamente con una amplitud mínima, para que sirva de punto de partida a las estadísticas continuas.

La coordinación es, pues, provechosa para ambos sectores y mejora el conocimiento estadístico sobre la estructura y desarrollo de actividades en el país. Esta coordinación ha de hacerse mediante la unificación de conceptos, definiciones y acuerdo en las épocas de recogida de los datos.

CONTRIBUCIONES ESPECIALES DE LOS CENSOS A LAS ESTADÍSTICAS CONTINUAS.

La realización de un censo agrario ofrece la oportunidad para implantar nuevas técnicas, o revisar las técnicas ya en uso en el sistema de estadísticas continuas. La enumeración directa y completa del censo suscita muchos problemas de concepto y método, cuyo estudio proporciona base para valorar la propiedad y pre-

cisión de las definiciones y la eficiencia de los procedimientos aplicados a las estadísticas continuas.

Otra contribución de gran importancia es la de los propios resultados que se tabulan con motivo del censo, que han de ser utilizados como datos básicos de referencia en las estadísticas continuas y en la aplicación de las técnicas de muestreo.

Al realizarse un censo agrario en un país en donde no existe para ello una organización permanente, el sistema de estadísticas continuas se beneficia con la incorporación posterior del equipo y servicios que permiten ampliar su capacidad de producción en cantidad, calidad y rapidez.

En el caso de utilizar procedimientos cartográficos para levantar el censo, estos materiales, de gran valor, quedan también a disposición de las estadísticas continuas. Si se utilizan procedimientos basados en listas, éstas, aunque no son tan permanentes como el material cartográfico, pueden ser de utilidad.

Por otro lado, los censos son una de las escuelas más eficaces para el entrenamiento y capacitación en las técnicas estadísticas, y el numeroso personal necesario para su ejecución queda, al terminar los trabajos censales, a disposición de las oficinas de estadísticas continuas, permitiendo la solución de uno de los problemas más apremiantes, que es la falta de personal calificado.

ESTADISTICAS CONTINUAS

Para este tipo de estadísticas hay una serie de métodos de obtención que se pueden agrupar dentro de los tipos básicos siguientes:

- 1.º Subproductos de actividades administrativas.
- 2.º Opiniones de expertos.
- 3.º Investigaciones por muestreo.

Lo más frecuente es utilizar varios de los tipos indicados, para sacar el mayor partido posible de la situación, pero es necesario conocer las posibilidades de cada método para poderlo emplear de una forma razonable.

En los métodos estadísticos están íntimamente relacionados el coste y la exactitud. A mayor exactitud, mayor coste; mientras que si se desea poca exactitud, el coste es mucho menor. Es responsabilidad de los gobernantes el decidir la exactitud necesaria para

que las cifras estadísticas puedan ser utilizadas para los fines deseados. Cuando la exactitud es muy pequeña, y sobre todo cuando se desconoce en absoluto el posible límite de error cometido, es muy fácil que los resultados obtenidos del análisis de estas estadísticas sean inútiles e incluso contraproducentes. Antes de proceder a una investigación estadística se debe fijar el grado de exactitud necesario para que los resultados puedan ser utilizados en estudios y proyectos posteriores. Esta es una verdad evidente, pero que generalmente se olvida por completo en la práctica.

A continuación se discuten en detalle los diferentes procedimientos de obtención de estadísticas continuas.

SUBPRODUCTOS DE LA ACTIVIDAD ADMINISTRATIVA.

Las actividades de Organismos oficiales y privados permiten obtener, como subproducto, un número variable de datos que forma un conjunto valioso de estadísticas. Si bien los datos no son recogidos originalmente para este fin, con frecuencia llenan necesidades que de otra forma tendrían que ser satisfechas por medio de censos o investigaciones por muestreo planeadas específicamente. En ciertas ocasiones, este conjunto de datos puede compararse a los obtenidos por medio de censos, cuando se cubren todas las unidades del universo, o puede compararse a los obtenidos mediante una investigación por muestreo, cuando se cubre solamente una parte de la población.

Desde un punto de vista práctico de obtención de estadísticas, lo más interesante de estos datos es el estar disponibles con poco o ningún coste. En el peor de los casos se necesita únicamente investigar archivos y registros, pero el coste de estas operaciones es muy pequeño, si se compara con el de un censo o una investigación por muestreo.

Las estadísticas obtenidas por este procedimiento deben ser miradas con cierta precaución por muchas razones. Los datos son recogidos por organismos cuya actuación principal se dirige a la administración de un programa, y cuyo interés en los datos, como estadísticas, puede ser insignificante. Por otra parte, tales organismos no siempre se encuentran en condiciones para preparar recopilaciones exactas, aun cuando tengan interés en hacerlas. No es raro encontrar que los datos fueron obtenidos y recopilados sin cuidado, acumulando errores a cada paso. Además, los datos

suministrados a los representantes de un organismo cuya actuación afecta al informante aparecen con frecuencia afectados de sesgos. El informante puede desfigurar los hechos deliberadamente cuando esto le es ventajoso. En ocasiones, el mismo organismo que recoge la información puede deformarla a su conveniencia, lo que daría lugar a una estadística con varios e importantes tipos de sesgos. Otra limitación se debe a que, cuando se dispone de datos para solamente una fracción del universo, tal fracción puede no ser una muestra representativa. Por otra parte, es muy fácil que el organismo que recoge los datos esté interesado en un universo definido de una manera diferente a la que interesa al estadístico, y esta diferencia en las definiciones crea serios problemas y es bastante frecuente. Existe otra dificultad de gran importancia, pues, aunque la recopilación de datos se haga perfectamente, el procedimiento no permite obtener información más que sobre conceptos más o menos intervenidos por organismos públicos y privados, lo que impide obtener información sobre puntos tan importantes como: distribución de las explotaciones agrarias, trabajo en la agricultura, producciones de frutas y hortalizas, etc. Señalamos como último inconveniente que este método no permite estimar los límites máximos de error cometido, a no ser que se utilicen procedimientos independientes de comprobación.

OPINIÓN DE EXPERTOS.

Si, por ejemplo, se desea conocer la superficie dedicada a un cultivo en una provincia, un método puede ser el de preguntar a un experto agrícola que tenga contacto con la situación provincial. La estimación del experto estará basada sobre su conocimiento de lo que ha ocurrido en el pasado y sus especulaciones sobre la situación presente. Este procedimiento puede, en efecto, dar los resultados deseados con gran aproximación. Con frecuencia, sin embargo, cuando dos expertos considerados como bien informados dan sus estimaciones, éstas son bastante diferentes. La variación de las estimaciones indica las diferencias de interpretación de distintas personas sobre los mismos o diferentes datos. Al valorar las estimaciones de los expertos, el único criterio posible es la fe en la bondad de sus opiniones. Estamos incapacitados para valorar esta fe, debido a la falta de bases objetivas para la selección entre opiniones de dos o más expertos.

El uso de la opinión de un experto es, generalmente, un procedimiento barato de obtener información. Además, hay numerosas situaciones en las que no se pueden utilizar los métodos de medición objetiva y es necesario depender por completo de la opinión de los expertos. Sin embargo, si los métodos de medición objetiva son utilizables, y una decisión importante depende de la exactitud de las estimaciones, un error en ellas puede resultar mucho más costoso, a la larga, que el tipo más caro de sobrevisión que pudiera ser razonablemente considerado.

Esto no quiere decir que la opinión de los expertos sea mala en general, sino que no es un buen sustituto de los procedimientos objetivos. Hay muchas situaciones en las cuales es importante conocer el error posible de una estimación, comparada con el resultado de una cuidadosa enumeración completa. Entonces resulta muy deseable el empleo de sobrevisión probabilísticas en las que se puede calcular el límite máximo de error cometido. Fiarse de las relaciones observadas en el pasado puede ser particularmente peligroso cuando hay cambios económicos importantes y rápidos, a pesar de que precisamente en tales ocasiones es más vital la necesidad de resultados fiables.

INVESTIGACIONES POR MUESTREO.

Una investigación por muestreo se diferencia de un censo en que establece contacto con sólo una fracción del universo. De la muestra se obtiene información que permite estimar los totales y promedios para el universo completo. En la situación actual de los conocimientos estadísticos, tanto los censos como las investigaciones por muestreo constituyen herramientas muy útiles. Un censo es particularmente provechoso cuando se dispone de poca información acerca de la estructura del universo y se desea conocer información básica exacta. Las investigaciones por muestreo complementan a los censos y permiten obtener gran cantidad de información con mucho menor coste. A pesar de todo, aun en el caso de disponer de muy poca información sobre el universo es posible plantear una investigación por muestreo que produzca resultados bastante exactos, si bien, cuando la exactitud deseada se acerca a la de un censo, los costes suben vertiginosamente y pueden llegar a ser casi iguales. Esto se debe, principalmente, a que para poder mantener los errores muestrales dentro de límites

pequeños habrá que hacer la muestra muy grande. Un censo constituye también el método más práctico cuando se necesitan datos para pequeñas áreas geográficas; en este caso, también, para que las muestras rindieran una exactitud aceptable deberían ser tan grandes que un censo completo podría resultar del mismo coste. Por el contrario, si se desea información para zonas geográficas de gran tamaño, o para el total nacional, los métodos de muestreo resultan mucho más baratos que los censos.

Hay dos tipos básicos de investigaciones por muestreo, que se diferencian en un concepto de importancia fundamental:

- 1.º Investigaciones por muestreo *no probabilístico*.
- 2.º Investigaciones por muestreo *probabilístico*.

La muestra probabilística es aquella que ha sido seleccionada por un procedimiento aleatorio, bien sea mediante insaculación, o empleando una tabla de números aleatorios. Esta es la diferencia básica entre los dos grupos. Hay toda una teoría de estadística matemática puesta a punto para manejar las muestras probabilísticas, pero que es inutilizable para todas las que no lo son. Esta parte de la ciencia estadística es muy moderna y su desarrollo más importante se ha llevado a cabo en los veinte últimos años; tanto el cuerpo de doctrina como su aplicación práctica han evolucionado muy rápidamente y continúan en un gran proceso de crecimiento.

INVESTIGACIONES POR MUESTREO NO PROBABILÍSTICO.

Entre los métodos más populares de muestreo usados en el pasado hay cuatro que presentan un interés especial:

1. Selección de conglomerados típicos.
2. Investigaciones por correo.
3. Método de las cuotas.
4. Método del punto al azar.

Todos ellos presentan dos aspectos característicos:

- a) Producen generalmente muestras sesgadas de la población que se estudia.
- b) Dependen de informaciones históricas como base de expansión.

Por ejemplo, muchas investigaciones de explotaciones agrarias por muestreo se diseñan exclusivamente para estimar un promedio, tal como la media de la superficie por explotación, la cual se

multiplica por una estimación del número total de explotaciones para estimar la superficie total cultivada. El número de explotaciones se obtiene generalmente de la última cifra censal publicada, pero los datos censales están anticuados en uno o dos años cuando se publican, y pueden pasar varios años más hasta que la nueva información se encuentre utilizable. Si el número de explotaciones ha cambiado mientras tanto, cualquier estimación basada en esta cantidad presentará el error correspondiente. Además, si se usa el número de explotaciones que aparece en el censo como base de expansión, es esencial que la definición de explotación agraria en la muestra sea consistente con la usada en el censo. Esto restringe la utilidad de los métodos indicados, porque la definición de explotación usada en el censo no es satisfactoria para toda clase de encuestas. Por otra parte, pueden presentarse sesgos de otro tipo.

1. *Selección de conglomerados típicos.*

Este método presenta sesgos importantes y no es posible medir la fiabilidad de sus resultados. Se basa en la selección de personas, ciudades o zonas consideradas como típicas para representar la población completa. Aun en el caso de que éstas fueran típicas o promedios en el pasado, no existe ninguna seguridad de que continúen siéndolo en el presente. En ocasiones se intenta dar validez a este tipo de selección, comparando los resultados con los obtenidos en otras sobrevisiones para ciertas características de las personas o zonas, aunque esto no asegura que los resultados sean adecuados para otras características en fechas diferentes.

2. *Investigaciones por correo.*

Este tipo de encuesta se lleva a cabo enviando cuestionarios a los componentes de una lista preparada con anterioridad, o distribuyéndolos sistemáticamente por medio de los carteros del Servicio de Correos. Este método ha sido bastante popular por su facilidad de administración y por no requerir enumeradores. Con él se puede obtener una gran cantidad de contestaciones a bajo coste. A pesar de su facilidad para incurrir en sesgos muy importantes, el método ha sido ampliamente usado cuando se requirieron gran cantidad de contestaciones para estimar inventarios

agrarios, que varían grandemente de explotación a explotación, e incluso cuando se necesitaba un gran detalle geográfico. En este caso, el error muestral más pequeño, que se consigue gracias a la gran cantidad de contestaciones, diluye algo el efecto del sesgo si éste no se tiene en cuenta en el procedimiento de estimación.

Las investigaciones por correo tienen unas limitaciones muy definidas. Los cuestionarios han de ser cortos y no permiten en muchos casos la formulación de las preguntas indirectas necesarias en muchas sobrevisiones. También presentan dificultades cuando las preguntas a contestar son difíciles desde un punto de vista técnico. Los resultados de las investigaciones por correo aparecen generalmente sesgados, porque es difícil obtener contestaciones de aquellas personas que son indiferentes o analfabetas. Pero la primera limitación de este método es la lista en sí. Es claro que una muestra —no importa cuán cuidadosamente ha sido extraída de una lista— no pueda ser más representativa que la misma lista. La segunda limitación se presenta cuando es alta la proporción de los que no responden. No se puede considerar que los que no responden tienen características similares a los que lo hacen. Los que no responden pueden ser de varias clases: *a)* no están en casa; *b)* no desean contestar; *c)* se consideran perjudicados si lo hacen. Cuando es alta la proporción de los que no responden no se puede tener mucha confianza en los resultados, a menos que se consigan informaciones adicionales, por otros medios, sobre los que no contestan.

3. *Método de las cuotas.*

Este método ha sido muy popular, debido a la enorme publicidad que ha recibido por su uso en predicciones electorales, estudios sobre la opinión pública e investigaciones de mercado. En este método, la población es dividida en un número de estratos y la muestra se extrae proporcionalmente de cada uno de ellos. El número de unidades muestrales requeridas en cada estrato recibe el nombre de cuota. Ordinariamente, el método es fácil de llevar a cabo, pero tiene algunos puntos débiles que pueden resultar muy serios en ciertas investigaciones. Las cuotas se distribuyen de acuerdo con la última información censal de que se dispone, y como esta información es de tipo histórico, las cuotas seleccionadas pueden no estar de acuerdo con las proporciones

actuales de la población en el momento de la encuesta. En períodos de cambios rápidos, tales como depresiones o guerras, estas cuotas resultan más afectadas de sesgos, aunque es en estos momentos cuando resulta más importante y deseable tener información exacta. Por otra parte, la selección de los individuos dentro de la cuota de cada grupo se deja a la discreción de los enumeradores, lo que suele producir sesgos en la misma selección por no reunir a un grupo realmente representativo. Por otra parte, este método es de uso limitado, porque depende de un censo como base de expansión y, además, ésta habrá de hacerse basada en información histórica, con los inconvenientes ya señalados.

4. *Método del punto al azar.*

A pesar de que, como el nombre indica, se seleccionan puntos al azar en una superficie o zona, por este procedimiento se obtienen muestras sesgadas en la agricultura. En este método se sitúan al azar una serie de puntos sobre un mapa y se incluyen en la muestra cierto número de explotaciones entre las más próximas a estos puntos. Usando este método se puede diseñar una muestra rápidamente. Además, el coste por cuestionario se puede reducir incluyendo más de una explotación para cada punto, lo que disminuye los gastos de transporte del enumerador entre las explotaciones. La información resulta sesgada porque los puntos al azar no tienen la misma probabilidad de caer en las explotaciones grandes que en las pequeñas, lo que también resulta influido por el número de explotaciones que se enumeran para cada punto. Otra de las grandes dificultades del método está en la localización en el terreno de los puntos seleccionados, que se sitúan mediante sus coordenadas geográficas, pero éstas son difíciles de precisar con exactitud sobre el terreno. El sesgo que se produce en la selección de la muestra obliga a utilizar procedimientos de estimación más complicados, por lo que se han de usar informaciones adicionales para corregir el sesgo. Todas las razones indicadas aconsejan situar este método, que a primera vista parece probabilístico, dentro del grupo de investigaciones por muestreo no probabilístico.

CRÍTICA DE LAS INVESTIGACIONES POR MUESTREO NO PROBABILÍSTICO.

Con todo lo dicho, es necesario hacer notar que, para una encuesta específica, un método de muestreo sesgado puede proporcionar mayor y más precisa información por peseta gastada cuando esta información se puede obtener de una forma muy barata. A pesar de todo, los tipos de encuesta indicados anteriormente tienen dos grandes limitaciones:

La primera es la dificultad de poner el énfasis apropiado en los diferentes factores que afectan al diseño muestral. En ocasiones se da demasiada importancia al tamaño de la muestra (como en las investigaciones por correo); otras veces se da demasiada importancia al uso de controles efectivos; otras se descansa casi enteramente en los procedimientos de estimación. Pero lo que se nota en conjunto es la falta de una teoría que indique la distribución deseable de los recursos para todos los factores que entran en el diseño de la muestra. Es necesaria una orientación para valorar los diversos factores que entran en el diseño y contribuyen al error muestral y así seleccionar el mejor diseño entre cierto número de diseños alternativos. Sin esta guía, los recursos utilizables para la encuesta pueden ser gastados en demasía en algunos aspectos del diseño, mientras que otros de igual o mayor importancia no reciben ni la atención ni el dinero debidos.

La segunda limitación es la imposibilidad de medir la precisión de los resultados muestrales, porque en la actualidad no se conoce ningún procedimiento para medir el grado de confianza que se puede depositar sobre las estimaciones muestrales obtenidas por estos métodos. No obstante, en algunas ocasiones se aplican fórmulas para obtener esta precisión, pero el uso de tales fórmulas resulta equivocado porque encierra grandes presunciones difíciles de defender, porque: *a)* las fórmulas para medir el error muestral dependen del conocimiento de la probabilidad con la que un individuo está incluido en la muestra; *b)* las fórmulas para los errores muestrales dependen del diseño utilizado en cada caso.

Los métodos de muestreo descritos más arriba tienen como característica común la de que es desconocida la probabilidad de que un individuo esté incluido en la muestra. Esta probabilidad será desconocida en cualquier método en el que no se utilice un

procedimiento probabilístico para la selección final de la muestra. Cuando la selección de los individuos que serán incluidos en la muestra se basa en el juicio personal, no es posible tener una medida objetiva de la fiabilidad de los resultados muestrales, porque los diferentes individuos pueden tener una oportunidad diferente y desconocida de ser incluidos. La selección al azar no se consigue, a menos que se sigan procedimientos específicos para controlar las probabilidades de selección de las unidades muestrales.

INVESTIGACIONES POR MUESTREO PROBABILÍSTICO.

Se basan en el estudio de una muestra extraída al azar y que, por lo tanto, sea perfectamente representativa de la población en estudio.

Existen dos métodos típicos que se pueden utilizar:

- 1.º Muestreo probabilístico de listas.
- 2.º Muestreo probabilístico de áreas.

MUESTREO PROBABILÍSTICO DE LISTAS.

En el caso de disponer de una lista completa de todas las unidades elementales que forman la población, sería muy simple la extracción de una muestra, que podría hacerse, por ejemplo, mediante un muestreo sistemático. Pero este caso tan sencillo se presenta con muy poca frecuencia en agricultura, especialmente cuando se desea hacer una investigación de tipo nacional. No es nada fácil poder disponer de una lista realmente completa. En ocasiones se toma como tal la obtenida del último censo, aunque se pueden poner reparos bastante serios a este proceder.

Supongamos que se desea investigar cierta característica de las explotaciones agrarias; en este caso, el conjunto de cuestionarios del último censo agrario forma una lista que presenta ciertas deficiencias: 1.ª Las explotaciones menores de un cierto tamaño no se censan y, por lo tanto, todas éstas no estarían incluidas en la lista. A pesar de todo, se puede limitar la población en estudio a todas aquellas explotaciones cuyo tamaño las hiciera incluíbles en el censo, con lo que desaparecería este primer inconveniente. 2.ª La enumeración censal no es nunca completa, y siempre que se hacen estudios posteriores de cobertura se comprueba que hay un porcentaje de unidades que no han sido enu-

meradas, aunque debieran serlo. El porcentaje de enumeración incompleta varía con el cuidado y perfección con que se levanta el censo, y su existencia se explica por el enorme volumen del trabajo a realizar, porque el universo suele estar formado por millones de unidades elementales. 3.ª El censo representa una situación en un momento determinado, pero esta situación varía con el tiempo, lo que avejenta la información censal y la hace histórica. La importancia de este inconveniente varía con la rapidez del cambio de situación y aumenta con el tiempo transcurrido desde la fecha del censo.

Los inconvenientes apuntados pueden dar lugar a sesgos muy serios y que no se podrán evitar por grande que sea el cuidado con que se realicen las otras operaciones de la encuesta. No debe olvidarse que una muestra no puede ser más representativa de una población que la lista de la que se ha extraído. A pesar de todo, el método puede estar justificado si se tienen en cuenta al hacer las estimaciones todos los factores señalados, por la economía que puede suponer su empleo.

Este método puede tener inconvenientes mucho mayores en otros casos. Si lo que se desea obtener son, por ejemplo, las producciones (superficies multiplicadas por los rendimientos), la muestra de explotaciones sería completamente sesgada, porque tienen la misma probabilidad de estar incluidas en la muestra una explotación con ese cultivo que otra sin él, y se le da el mismo peso a ambas para el cálculo de la producción. En este caso, el método habría que clasificarlo en el grupo de muestras no probabilísticas, aunque la muestra se haya extraído al azar.

Las consideraciones anteriores aconsejan tomar con ciertas reservas el uso de listas en agricultura, y los métodos de muestreo basados en ellas no pueden ser considerados como probabilísticos en ciertos casos.

MUESTREO PROBABILÍSTICO DE ÁREAS.

El muestreo por áreas y el muestreo probabilístico son considerados como sinónimos con cierta frecuencia, pero una muestra de áreas no es necesariamente una muestra probabilística, y muchas muestras probabilísticas no lo son de áreas. El muestreo de áreas se refiere al uso de pequeñas áreas de tierra como unidades muestrales, mientras que el muestreo probabilístico se refiere al

método de selección de las unidades muestrales. En una muestra probabilística cada unidad muestral tiene una probabilidad de selección conocida. De esta forma el muestreo probabilístico de áreas se refiere a una clase de muestras probabilísticas en las cuales se utilizan áreas como unidades muestrales.

El muestreo de áreas puede utilizarse para la obtención de una muestra no probabilística empleando uno cualquiera de los cuatro métodos indicados anteriormente. En la práctica, cuando se utiliza el muestreo de áreas suelen seguirse procedimientos aleatorios de selección de la muestra. Esto se debe a que el costo relativamente elevado del método supone, cuando se aplica, el deseo de obtener una estadística verdaderamente de calidad. Además, la extracción de una muestra aleatoria, cuando está preparado el resto del material necesario, significa pequeñas dificultades técnicas y operacionales y poco coste. Por estas razones, en adelante nos referiremos al muestreo de áreas como muestreo probabilístico de áreas, aunque no se especifique expresamente esta cualidad.

Brevemente, el muestreo de áreas incluye:

a) La subdivisión del área total en segmentos de pequeña superficie (unidades muestrales) que tienen límites fácilmente identificables en el terreno.

b) La selección de una muestra de estos segmentos.

En el campo, el enumerador tiene que identificar los límites de los segmentos muestrales y conseguir la información deseada para todas las explotaciones agrarias u otras unidades de observación que estén incluidas dentro de los propósitos de la sobrevisión. Una aplicación rigurosa del método de muestreo probabilístico de áreas requiere un material cartográfico adecuado para definir los segmentos con límites claros e identificables, y una colección de reglas para asociar las explotaciones agrarias con los segmentos, de tal forma que cada explotación perteneciente a la población definida esté asociada con un segmento solamente. Además, el número total de las áreas muestrales dentro de la población es conocido, y de esta forma se conoce también la fracción muestral. Para obtener selecciones insesgadas de las explotaciones se pueden seleccionar aquellas cuyas construcciones principales estén situadas dentro de los límites del segmento, aunque esté fuera de él parte de la tierra perteneciente a la explotación (segmentos abiertos). De igual forma se puede obtener una selección inses-

gada de personas tomando todas aquellas que residen dentro de los límites de los segmentos. Otro ejemplo puede ser la selección insesgada de las superficies de cultivo, que se puede conseguir seleccionando toda la tierra incluida dentro de los límites de los segmentos (segmentos cerrados). Finalmente, se puede obtener una selección insesgada de parceladas tomando aquellas que tienen su extremo noroeste dentro de los límites de los segmentos.

La muestra de áreas se puede utilizar en la selección primaria de unidades de observación, de las cuales se extrae una submuestra posteriormente. El submuestreo se consigue, o bien sacando una muestra de área más pequeña, o haciendo una lista preliminar de individuos.

Al comienzo de la aplicación del muestreo de áreas se pensó que cuanto mayores fueran éstas se podrían obtener mejores resultados, pero en la actualidad se considera que las unidades más pequeñas casi siempre resultan más eficientes. En 1937 se empezó a utilizar el muestreo de áreas para conocer la superficie dedicada a las distintas cosechas. Esta técnica se originó de forma independiente en la India y en los Estados Unidos y requiere la medición de las superficies ocupadas por las cosechas contenidas dentro de las áreas muestrales. En ambos casos estas unidades muestrales eran relativamente pequeñas.

Ha sido demostrado por diversos autores que el método de muestreo de áreas da una muestra insesgada para la agricultura, y tiene además la ventaja de que las unidades de observación se hallen reunidas de tal forma que el error muestral puede hacerse mínimo en relación con el coste. La reducción del coste se consigue mediante la reducción de los viajes entre explotaciones. El método resulta también adaptable a las técnicas modernas para reducir los errores muestrales (estratificación, muestreo por etapas, muestreo por conglomerados, etc.). El factor de expansión para el método de áreas resulta sin error porque el número total de áreas de la población puede determinarse simplemente contándolo en los mapas; por esto no es necesario utilizar información adicional, que puede no ser exacta. Para la agricultura tiene además la ventaja de dar una estimación del número de explotaciones, que se conoce raramente con exactitud satisfactoria. Si se toman muestras repetidas, la correlación que existe generalmente entre muestras de un año a otro puede ser utilizada en este método para reducir el error muestral por medio de diseños o procedi-

mientos de estimación apropiados. La ganancia en exactitud utilizando estas técnicas resulta de considerable valor. Por otra parte, los diseños en los cuales se utiliza como unidad la explotación agraria no resultan libres de sesgos, porque las explotaciones no son tan permanentes a través del tiempo como lo son las áreas.

El método de muestreo de áreas tiene ciertas desventajas, que son más de tipo operacional que teórico. Requiere mapas con detalle suficiente para la delimitación de los segmentos, cuyos límites deben ser suficientemente claros para permitir una fácil y exacta localización de los mismos en el campo. Tales mapas no están siempre disponibles. En estos casos es aconsejable suplementar con fotografías aéreas el material cartográfico existente, en especial cuando se desea emplear segmentos más pequeños y eficientes. Estas necesidades presentan varias dificultades específicas: en primer lugar, no suele haber una sola fuente de información cartográfica y es necesario conjuntar la mejor información existente, lo que supone un importante gasto de tiempo y dinero; en segundo lugar, aun cuando se disponga de los mapas necesarios, es imprescindible obtener reproducciones o fotografías de los mismos, lo que supone un coste y tiempo adicionales.

Otra desventaja de este método resulta del tiempo y gasto que supone conseguir los mapas de detalle suficiente y delinear las áreas que resultan eficientes por su tamaño y que están limitadas por accidentes físicos que permitan una identificación clara en el campo. Por consiguiente, un muestreo de áreas no puede ser diseñado rápida y al mismo tiempo eficazmente. Estas dificultades operacionales aconsejan la preparación de todo el material necesario con anticipación, pensando en que éste pueda ser utilizado en muchas sobrevisiones posteriores en el futuro. El gasto de esta labor preliminar resulta perfectamente económico cuando se haga un uso intenso del material reunido, y, en términos de pura contabilidad, no parece lógico cargar todo el coste a la primera sobrevisión que se lleve a cabo. En Estados Unidos, viendo la necesidad de llevar a cabo este trabajo, se procedió a reunir todo el material básico que desde el año 1945 se conoce como la «Muestra Maestra», que, si bien se pensó en principio para la agricultura, se ha ampliado a todos los sectores y se utiliza muchas veces cada año para los propósitos más diversos.

En resumen, puede decirse que el método de muestreo de áreas es un avance sobre otros métodos de diseño, porque:

a) No depende del conocimiento previo de las características de la población.

b) Elimina la libertad de elección por parte de los enumeradores.

c) Resulta eficiente desde el punto de vista de hacer máxima la precisión en términos de coste.

Aunque el diseño de una muestra de área resulta largo y de coste inicial elevado, generalmente resultan justificados tanto el trabajo como el coste.

LAS SOBREVISIONES MUESTRALES COMPARADAS CON LOS CENSOS.

En la reunión de 1925 del Instituto Internacional de Estadística se discutieron por primera vez varios procedimientos de muestreo, debiéndose a PEARSON y FISHER la mayor aportación a esta nueva parte de la Estadística. A partir de esta fecha se han mejorado mucho las técnicas estadísticas en los dos aspectos de recolección y análisis de datos. En los primeros tiempos la recolección de datos recibía poca atención por parte de los estadísticos. Estos consideraban esta labor como una pesada rutina indigna de ser realizada por el experto estadístico. La enumeración completa se consideraba como el ideal, y la obtención de datos por medio de muestreo como un invento diabólico que no merecía la pena ser considerado y que se olvidaría pronto.

Es solamente hacia 1935 cuando el muestreo recibe carta de naturaleza. La presión constante por una información estadística detallada empieza a relegar al censo a una posición secundaria; en la mayoría de los países los cuestionarios censales estaban ya sobrecargados y no permitían en realidad ampliar la recolección de datos estadísticos más que basándose en sobrevisiones muestrales. Con las técnicas estadísticas de muestreo disponibles en la actualidad puede decirse que, con la excepción de universos consistentes en un pequeño número de unidades, la enumeración completa resulta demasiado costosa y no muy productiva, si se compara con la utilización de muestras. Esto es particularmente verdad en las estadísticas agrarias. En las estimaciones de producciones, el factor tiempo y el factor economía son, por lo menos, tan importantes como la exactitud. La F. A. O. ha establecido un programa decenal de censos agrarios mundiales, y es evidente la necesidad de disponer de información en los períodos intercensales. Algunos

países realizan censos generales de la agricultura cada cinco años, pero sigue habiendo un período intercensal demasiado largo. El realizar censos anuales es costosísimo y tiene además otros inconvenientes, porque la tabulación y publicación de los resultados requiere muchas veces más de dos años, por los millones de fichas que hay que manejar. Por esta razón, aunque en algún país se realicen dos censos ganaderos al año, hay que considerar esta alternativa como muy costosa y poco práctica, lo que explica su poca difusión, siendo, en realidad, una solución anticuada.

La sobrevisión muestral no es solamente más económica, sino que, como veremos a continuación, incluso puede ser más acurada que las enumeraciones completas. Todas las enumeraciones están sujetas a dos clases de errores principales:

1.^a Errores debidos a la amplitud de la enumeración.

2.^a Errores debidos a la naturaleza de las informaciones individuales.

Los errores de la primera clase no están presentes en una enumeración completa, porque ésta cubre la totalidad del universo. En la práctica, sin embargo, los censos dejan de incluir una parte más o menos grande de la población en estudio, por lo que suelen hacerse estudios posteriores de cobertura para conocer la magnitud de este error.

Los errores de este primer grupo están formados por: *a*) errores aleatorios (errores muestrales en el sentido más exacto de este término) que pueden reducirse incrementando el tamaño de la muestra o la eficiencia del diseño muestral; *b*) un posible sesgo, debido a la selectividad de la muestra, que no puede reducirse a menos que ésta se incremente hasta un punto que comprenda la mayoría de la población. Sin embargo, se han puesto a punto métodos para la corrección de este sesgo.

Los errores de la segunda clase se deben al método de entrevista y están presentes en cualquier enumeración basada en informaciones subjetivas, sin que la amplitud de cobertura influya en esto. Pueden ser errores aleatorios o sistemáticos, y, naturalmente, aparecen tanto en los censos como en las sobrevisiónes muestrales.

Evidentemente, un censo no facilita una información acurada únicamente porque la cobertura es completa, pues se han señalado repetidamente sesgos e inconsistencias en las enumeraciones censales. Los sesgos más importantes se deben muchas veces al mé-

todo de entrevista, y si tales sesgos existen, puede obtenerse mayor acuracidad de dos formas: *a)* reduciendo o eliminando los sesgos por varios métodos hoy disponibles; *b)* reduciendo los sesgos y también los errores aleatorios por medio de una cobertura más cuidadosa e intensiva de las unidades individuales. Por lo tanto, pueden obtenerse resultados más fiables con una fracción del gasto necesario para un censo, realizando una sobrevisión muestral cuidadosamente planeada y de tamaño relativamente pequeño, llevada a cabo por enumeradores bien entrenados. El tamaño menor de la población a enumerar permitirá dedicar mayor esfuerzo a una comprobación cuidadosa de las respuestas y a un esmerado análisis de las fuentes y naturaleza de los sesgos. Además, en muchos casos podrían reemplazarse los métodos subjetivos y de entrevista por mediciones objetivas que son prácticamente imposibles de realizar en un censo. Según va aumentando la eficiencia y fiabilidad de los procedimientos muestrales aumenta también la confianza del público en ellos y, con el tiempo, las sobrevisiones muestrales tendrán mayor importancia que los censos en muchos campos de actividad. En el momento presente lo más aconsejable es utilizar ambos métodos, censos y sobrevisiones muestrales, completando con un procedimiento lo que no es práctico investigar por el otro, pero teniendo siempre en cuenta las limitaciones de cada uno.
