

SERIE DE MANUALES DIDACTICOS

Nº 12

ENCUESTAS POR MUESTREO PARA ESTUDIOS EPIDEMIOLOGICOS EN POBLACIONES ANIMALES



organización panamericana de la salud
oficina sanitaria panamericana, oficina regional
de la organización mundial de la salud



ORGANIZACION PANAMERICANA DE LA SALUD
Oficina Sanitaria Panamericana, Oficina Regional de la
ORGANIZACION MUNDIAL DE LA SALUD

CENTRO PANAMERICANO DE FIEBRE AFTOSA

CAIXA POSTAL 589 - ZC/00 - RIO DE JANEIRO, BRASIL

ENCUESTAS POR MUESTREO PARA
ESTUDIOS EPIDEMIOLOGICOS EN POBLACIONES ANIMALES

por

Vicente M. Astudillo

1 9 7 9

CONTENIDO

	Pág.
INTRODUCCION	1
VENTAJAS Y LIMITACIONES DEL MUESTREO	3
FUNDAMENTOS BASICOS DEL MUESTREO	5
1. Estimación de una tasa	7
2. Distribución teórica de muestreo de una tasa	9
3. Estimación de una tasa por intervalo de confianza	12
MUESTREO PROBABILISTICO	13
ERRORES EN EL MUESTREO	15
DISEÑO DE LA MUESTRA	18
1. Determinación del tamaño de la muestra (n)	19
2. Ecuación que relaciona "n" con algunas de las condiciones específicas	22
METODOS DE SELECCION DE LOS ELEMENTOS	24
1. Selección al azar simple	24
2. Muestreo estratificado	28
3. Muestreo sistemático	32
4. Muestreo por conglomerados	33
5. Muestreo en dos etapas	36
6. Probabilidad proporcional al tamaño de los conglomerados y muestreo en dos etapas	38
PLANIFICACION DEL MUESTREO	41
1. Formulación del objetivo	41
2. Definición de la población y de las unidades de muestreo	44
3. Información a coleccionar y métodos de colecta	45
4. Periodicidad de aplicación de las encuestas	47
5. El marco	48
6. Elección del diseño de la muestra	49
7. Análisis de la factibilidad y decisión de la forma final de realizarlo	50
ORGANIZACION Y EJECUCION	54
1. Organización del trabajo de terreno	54
2. Elaboración de un cuestionario, cédula de encuestador e instructivos	56
3. Entrenamiento de los encuestadores	57
4. Elaboración de mapas y asignación de itinerarios a los equipos de encuestadores	58
5. Divulgación a la comunidad de los objetivos de la encuesta	58
6. Prueba piloto de la encuesta	59
7. Ejecución de la encuesta	59
8. Procesamiento de datos y publicación de resultados	60

INTRODUCCION

Para planificar y ejecutar un programa de control de cualquier enfermedad de los animales es necesario tener información sobre su frecuencia, su distribución espacio-temporal, sobre los factores asociados a la ocurrencia, sobre los cambios que se producen en el espacio y en el tiempo en los aspectos antes mencionados. Este conocimiento permitirá caracterizar epidemiológicamente la enfermedad, reconocer "modos de comportamiento" diferentes entre regiones, también poner en evidencia que esas formas de conducta de la enfermedad, en diversos lugares, no son uniformes en el tiempo, que cambian según la época del año y también a través de los años. Mediante el acúmulo de información será posible identificar esos "modos de comportamiento" regionales y conocer también la conducta de los factores que son responsables por ese comportamiento, alcanzando así la fase de interpretación de cada ecosistema de la enfermedad (modo de comportamiento). Este conocimiento resulta básico para poder establecer las estrategias de control, ya que el conjunto de acciones de combate a la enfermedad debe ser concordante con las características que ella presenta.

Se trata de saber elegir los instrumentos de combate adecuados a cada caso y también decidir la intensidad necesaria con que deben ser aplicados. Evidentemente que todo este proceso involucra un proceso de manipulación de informaciones.

En algunos casos, existen sistemas continuos de información, que proporcionan datos de gran utilidad, sin embargo en otros casos, sea por razones administrativas, sea por la propia naturaleza de la enfermedad hay que recurrir a técnicas de muestreo.

El desarrollo de los sistemas de información se torna indispensable para el funcionamiento eficaz de los servicios nacionales encargados de prevenir o controlar las enfermedades de los animales. Un método valioso de colecta de datos de estos sistemas lo constituyen los procedimientos de muestreo.

En el caso de los programas de control de la fiebre aftosa en América del Sur, se han ido desarrollando gradualmente sistemas continuos que proporcionan información sobre la ocurrencia de la enfermedad, las actividades de control desarrolladas y la población animal existente. El estado actual de desarrollo de estos programas hace necesario que estas informaciones epidemiológicas sean complementadas por otras, que por su naturaleza o por su condición

de esporádicas o por estar referidas a una localidad no son proporcionadas por el sistema continuo. Esto ha hecho necesario apelar a un mecanismo más flexible y así las encuestas por muestreo para estudios epidemiológicos en las poblaciones animales se están utilizando en forma cada vez más creciente en los últimos años. Cuando ellas son bien planificadas y ejecutadas, producen una valiosa información sobre las características epidemiológicas en estudio, en forma rápida y con un costo moderado. En términos generales la aplicación de técnicas de muestreo se torna muy efectiva, especialmente cuando se cuenta con personal de campo bien preparado.

Se ha tornado cada vez más frecuente el hacer encuestas serológicas por muestreo. A través de ellas se estudia, en el ganado de regiones consideradas como de ocurrencia esporádica o indemes a la fiebre aftosa, el ingreso de virus de la fiebre aftosa a través de la tasa de infección, estableciendo las tasas de infección por edad y por sectores, informaciones que combinadas con otras informaciones epidemiológicas hace posible evaluar la antigüedad de la infección, la difusión, las circunstancias posibles del ingreso de fuentes de infección y así poder orientar los mecanismos de prevención en esas regiones, para el futuro. También a través de estas encuestas serológicas por muestreo es posible estudiar las curvas de anticuerpos, las variaciones por edad y por sectores, relacionar altos títulos de anticuerpos con el diagnóstico.

En estas regiones de ocurrencia esporádica, al hacerse estudios por muestreo también se estudia la presencia de animales portadores de virus en el exudado esofágico-faríngeo.

En forma simultánea a la extracción de sangre y material esofágico-faríngeo se hace una entrevista al propietario o encargado del rebaño sobre las características del manejo de los animales, que faciliten la interpretación de los resultados.

El análisis de estos datos permitirá evaluar la gravedad latente de algunas situaciones que, por efecto de cambios en la conducta de algunos de los factores, lleguen a modificar las interacciones entre ellos, pudiendo posteriormente manifestarse en forma clínica, e incluso llegando a ser epidémicos.

Aunque los principios de muestreo se aplican al estudio de las poblaciones animales desde hace muchos años, sólo a partir de la implantación de los programas de control de la fiebre aftosa, a mediados de la década pasada es que se ha ido acumulando experiencia sobre las dificultades que su aplicación acarrea.

VENTAJAS Y LIMITACIONES DEL MUESTREO

Dentro de las principales ventajas que el empleo del muestreo tiene se pueden mencionar las siguientes:

- a) bajo costo con relación a la enumeración completa de la población;
- b) menor consumo de tiempo en la obtención de la información;
- c) por ser menos multitudinario permite mayor detalle en los datos;
- d) mayor economía de espacio y personal;
- e) resultado de mejor calidad debido al mejor entrenamiento de un grupo más reducido de personas. Al mismo tiempo es posible reclutar personal más preparado, y
- f) en algunas oportunidades el muestreo es la única alternativa para estudiar un problema. En muchas oportunidades el examen de todos los elementos componentes de una población exige su destrucción o inutilización, como es el caso de control de vacunas, control de canales en mataderos, control de leche embotellada, control de alimentos en conservas, etc.

Por otra parte el muestreo presenta algunas limitaciones entre las que se pueden mencionar las siguientes:

- a) no proporciona información acerca de todos los elementos de la población, y
- b) el muestreo plantea problemas técnicos que requieren muchas veces la presencia de un especialista en la materia, hecho que no siempre es posible, lo que constituye una limitación.

El estudio de todos los elementos o individuos que componen un agregado es casi siempre difícil, llegando a ser impracticable en muchos casos, por dificultades físicas, económicas y de otro orden. Esta es la razón por la cual una buena parte de los estudios se hace sobre muestras, que son agregados menores que el agregado universal. Hay bastante evidencia de que los resultados, obtenidos a partir de estas agrupaciones menores de unidades o individuos, son muy satisfactorios, si se adoptan ciertas precauciones.

La expresión "principios y procedimientos de encuestas por muestreo" se utiliza en esta presentación para designar dos aspectos. Uno, se refiere a los fundamentos y técnicas de naturaleza estadística indispensables para la colecta y análisis de los datos. El otro, se refiere a la aplicación de principios

y métodos de "programación de encuestas" para su desarrollo desde el planteamiento de objetivos, la elección de la población para el estudio, el establecimiento de estrategias, la elaboración de planes de acción para la colecta de datos y la evaluación de resultados alcanzados.

La primera parte se refiere a la teoría de muestreo de atributos, especialmente de aquellos atributos de carácter dicotómico, dado que las enfermedades generan datos de esta naturaleza, al clasificar en un momento dado a los individuos de una población animal como enfermos o no enfermos.

En este documento se tratan algunos principios que están involucrados en la selección de una muestra para cumplir con algún objetivo en el campo de la prevención y el control de enfermedades de los animales. Sin embargo, esto no siempre es la parte más difícil en una encuesta por muestreo, pero es la parte donde personas sin experiencia en muestreo pueden sentirse poco seguras de sus conceptos sobre teoría de muestreo. De ahí que se proporcione una explicación resumida sobre las bases en que descansan los métodos de muestreo.

En la segunda parte se desarrollan procedimientos para el diseño de un muestreo, teniendo en cuenta la determinación del tamaño de la muestra, el método de selección de las unidades incluidas en la muestra y la estimación de las tasas y sus correspondientes errores de muestreo. Se pone énfasis en las necesarias relaciones que existen entre los tres aspectos, una vez elegida una forma de selección los individuos a incluir en la muestra, de manera que ésta resulte más representativa y del menor costo posible.

En la tercera parte se aborda la necesidad de introducir técnicas de programación para elaborar encuestas por muestreo a fin de sistematizar en forma lógica las diferentes partes del proceso en función de los objetivos. Se debe intentar reducir al máximo los errores de medición controlando sus fuentes principales, al mismo tiempo que seleccionar un plan de encuesta por muestreo que además de considerar la estimación de tasas y sus errores se utilicen procedimientos analíticos que tomen en cuenta los costos y la factibilidad del plan.

La aplicación de las encuestas por muestreo en salud animal es de gran trascendencia en diagnósticos de situación que permiten averiguar la frecuencia de una o más enfermedades y la conducta de factores que los determinan. Este diagnóstico puede referirse a un período anterior a la ejecución de la encuesta, por lo cual se habla de estudios retrospectivos en que el mecanismo de colecta de datos es la entrevista al encargado del rebaño como ocurre con fiebre aftosa

y rabia. Cuando la averiguación se refiere al momento de ejecución de la encuesta, se habla de estudio prospectivo. En estos casos, generalmente se trata de encuestas serológicas o de reacción cutánea, de interés inmediato. Entre las primeras podemos considerar la brucelosis, la leptospirosis y la propia fiebre aftosa. Entre los segundos la tuberculosis.

Otra aplicación de las encuestas por muestreo en el estudio de la frecuencia de las enfermedades es como mecanismo de acompañamiento de un programa para estudiar los cambios en la tendencia de ella, en enfermedades en que el diagnóstico clínico es de nula o escasa utilidad. Tal es el caso de enfermedades como brucelosis y tuberculosis. Estos estudios son a largo plazo e involucran la repetición de varias encuestas longitudinales en ese espacio a través del tiempo. Esto se refiere a la idea de hacer exámenes periódicos en una misma muestra, por ejemplo, una muestra fija de rebaños a través del tiempo, o la encuesta se hace cada vez sobre una nueva muestra de la población (estudio trasversal).

También se aplican estas encuestas por muestreo para medir el estado inmunitario de una población animal, lo que permite identificar grupos demográficos más susceptibles. También es posible aplicar estos procedimientos para evaluar la eficacia de los programas de prevención o control sobre la población animal o sobre la propia comunidad.

FUNDAMENTOS BASICOS DEL MUESTREO

Se define como población (universo) al conjunto total de unidades (individuos, objetos, elementos, mediciones) existentes o presentes en un cierto lugar durante un período de tiempo, que poseen una característica en común que se desea estudiar. El número de elementos que componen la población se llama tamaño de la población (N).

Se define como muestra a un grupo generalmente reducido de unidades obtenidas desde la población mediante procedimientos objetivos, de forma que resulte representativa de la población original, existiendo con respecto a ésta sólo una diferencia de tamaño.

El número de elementos que componen una muestra se llama tamaño de la muestra (n).

Se llama muestreo a los procedimientos estadísticos empleados para seleccionar la muestra a partir de la población, con el objeto de estudiar en ella alguna característica, de manera que los resultados obtenidos puedan ser generalizados a la población de origen. Para que una muestra sea representativa no basta que los individuos provengan de todos los "sectores" del universo aunque estuvieran en la proporción que corresponde. Es fundamental que todas las características de la población estén presentes en la muestra, principalmente la variación entre las unidades de muestreo. Unidad de muestreo es, por ahora, cada elemento o individuo.

Dentro de las escalas de medida, los animales al enfermar pueden ser clasificados de acuerdo con la escala nominal, ya que la condición de enfermo no se mide o cuantifica, sólo se clasifica como sí o no, expresando un atributo o cualidad. El sí y el no constituyen símbolos o modalidades de un código para designar las dos clases en que se agrupan los individuos. Es la escala más básica, ya que cualquier escala más evolucionada de medición empieza por un proceso de clasificación.

Un ejemplo de atributo, con dos clases de modalidades sería la infección por fiebre aftosa, atribuyendo por ejemplo al infectado el valor 1 y al no infectado el valor 0 (si la ausencia de una modalidad de este atributo infección implica la presencia de la otra modalidad).

En algunas ocasiones en salud animal se quiere conocer la frecuencia absoluta (cantidad de unidades) o la tasa (frecuencia relativa o porcentaje) que pertenecen a una clase o a un atributo de una variable cualitativa.

La mayoría de nuestras informaciones que provienen de los sistemas de información (notificación, registro, censos, muestreo) son de este tipo: la cantidad de vacas, el número de bovinos, los casos de fiebre aftosa, los positivos a la brucelosis, etc.

Supongamos que cada una y todas las unidades o elementos de la población o universo cae dentro de una de dos posibles clases (C y \bar{C}) o modalidades o atributos de una variable cualitativa cualquiera.

	<u>Población</u>	<u>Muestra</u>
Número de elementos en C :	A	a
Número de elementos en \bar{C} :	(N - A)	(n - a)

Un valor característico de la población tal como una tasa (P) o una media aritmética (μ) es llamado parámetro. En cambio, el valor característico cuando está referido a una muestra se llama estadístico (p, \bar{x} , etc.).

	<u>Población</u> (Parámetro)	<u>Muestra</u> (Estadístico)
Tasa de elementos en C:	$P = A/N$	$p = a/n$
Tasa de elementos en \bar{C} :	$Q = (N-A)/N$	$q = (n-a)/n$

Los parámetros están basados en todos los elementos y ellos son constantes para la población. Los estadísticos están basados sólo en una parte de los elementos de la población por tanto ellos varían de muestra a muestra. De ahí que para poder desarrollar procedimientos estadísticos de muestreo es conveniente tener en cuenta la distribución teórica de muestreo del correspondiente estadístico.

1. Estimación de una tasa

El estimador muestral de P es p (el estimador muestral de A es N.p) donde $\hat{P} = \mu_p = \bar{p}$ para muestras grandes. El error de muestreo (error estándar) de p en este caso es:

$$\sigma_p = \sqrt{\frac{PQ}{n}}$$

Cuando no se conoce el valor poblacional, el error de muestreo se puede calcular a partir de muestras grandes a través de

$$s_p = \sqrt{\frac{pq}{n}}$$

El interés que orienta a quien lleva a cabo un estudio epidemiológico, y que para hacerlo toma una muestra a partir de una población, es conocer las características de ésta, a partir de las características que presenta la muestra, es decir, conocer el valor de los parámetros a partir de los estadísticos. Este procedimiento inductivo, sabemos ya, recibe el nombre de estimación.

Con cualquier método de muestreo que se opere, siempre la estimación o el valor estimado del parámetro ($p = \hat{P}$), a partir de los valores de la muestra, discrepa o se diferencia del verdadero valor del parámetro (P). Esta diferencia $|P-p|$ se llama error de muestreo. Mientras mayor sea el tamaño de la muestra (n), el error de muestreo disminuye, aumentando la confianza sobre el resultado. Un adecuado procedimiento de muestreo debe proporcionar una medida del error de muestreo. La forma como se mide el error de muestreo es a través del error estándar de la estimación hecha (p). Da una idea de la confianza acerca de la estimación.

Como en la muestra participa solamente un grupo reducido de elementos de la población de origen, es siempre de esperar que los resultados obtenidos a base de una muestra no sean "exactamente" iguales a los que se obtendrían si se analizaran todos los elementos de la población.

Así, por ejemplo, si en un archivo (tarjetas) de 500 vacas tuberculizadas, de las cuales 300 son positivas y las 200 restantes son negativas, si se toma al azar una muestra de 50 tarjetas es muy poco probable que se obtengan exactamente 30 fichas de vacas positivas y 20 fichas de negativas. Se obtendrá, por ejemplo (por citar un caso posible) 29 fichas de positivas y 21 fichas de negativas, lo que indicará que en este archivo hay una tasa probable de:

$$\frac{29}{50} \times 100 = 58\% \text{ de positivas, cuando efectivamente hay } 60\%.$$

Aunque se repita la extracción de 50 fichas varias veces, en muy contados casos tendremos la "exacta" distribución de las tarjetas en cuanto a reacción a la tuberculina.

Existe, por lo tanto, una diferencia entre lo que "informa" la muestra y lo que realmente sucede en el universo muestreado, diferencia que recibe el nombre de "error de muestreo", y que puede hacerse tan pequeño como sea posible, es decir se puede elegir una muestra de tal modo que la diferencia entre:

$$\text{tasa en la muestra} - \text{tasa en la población}$$

sea de tal magnitud que, para fines de orden práctico, sean considerados como "iguales" (precisión admisible).

Los errores pueden ser causales o sistemáticos. Los causales siguen las leyes de probabilidad y su promedio es igual a cero. La magnitud de estos errores se puede controlar. En cambio los sistemáticos no se cancelan unos con otros, produciendo una discrepancia permanente que se llama vicio, por lo cual las estimaciones afectadas por errores sistemáticos son estimaciones viciadas.

Debido a estas circunstancias es conveniente elegir un método de muestreo que evite los errores sistemáticos y sólo puedan presentarse los errores casuales. Los métodos de muestreo que cumplen con esta condición son los de tipo probabilístico.

Aparentemente pareciera que la presencia del "error de muestreo" sería un grave inconveniente para el uso de muestras, pero como se anotó existen medios que permiten controlar eficientemente la magnitud de este error.

Se dice que el "error de muestreo" es controlable cuando se puede elegir la muestra de modo que la diferencia probable entre el valor obtenido a base de la muestra y el verdadero valor en la población quede dentro de límites bien definidos.

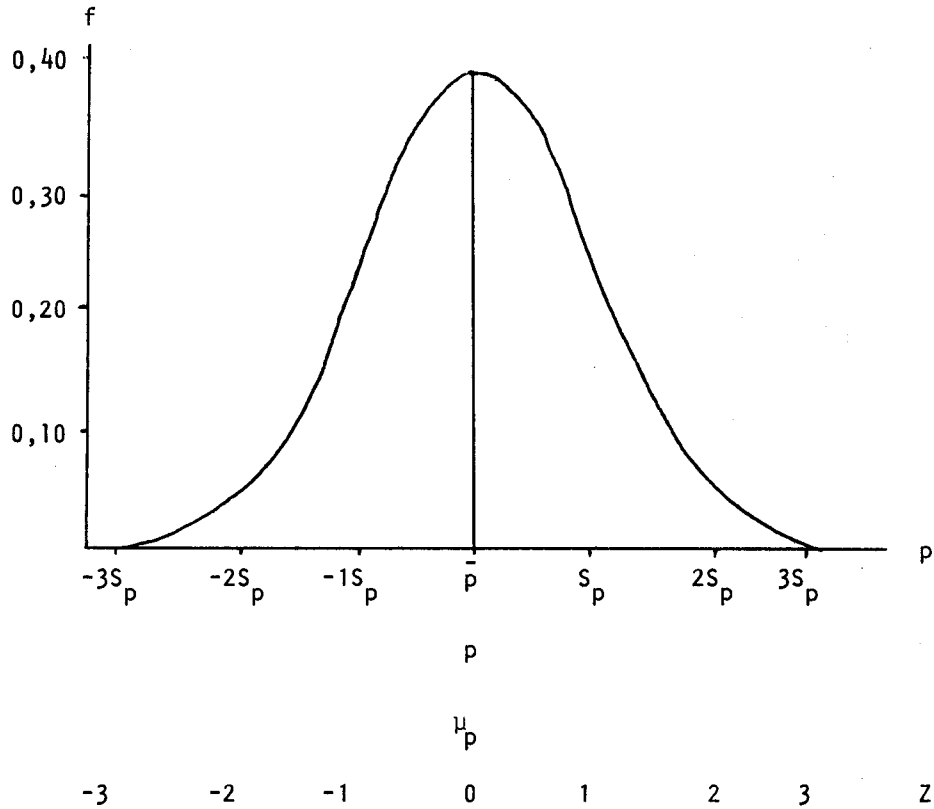
Únicamente cuando se usa un esquema matemático bien definido en la selección de las unidades de la muestra, se puede calcular la magnitud del "error probable".

2. Distribución teórica de muestreo de una tasa

Si se considera una serie de posibles muestras de un tamaño determinado, $n > 30$, que pueden ser obtenidas a partir de una población, utilizando un procedimiento sin reposición, para cada muestra se puede calcular la tasa correspondiente, estadístico que variará de una muestra a otra.

Dado que las tasas provenientes de varias muestras seleccionadas desde esa población presentan variación es posible pensar que ellas originan una distribución de frecuencias, las cuales se llaman distribuciones de muestreo, ya que la variación en ellas es producto del proceso de muestreo. Es comprensible que de una población se puede obtener una cantidad infinita de muestras, todas del mismo tamaño y de esta manera se puede concebir que para la tasa se podría tener una distribución de probabilidades.

Considerando la situación hipotética de todas las muestras posibles de tamaño grande n , si para cada muestra se calcula \underline{p} , la distribución teórica de muestreo de las tasas resultante tendría esta forma



con media $\mu_p = P$ y error estándar $\sigma_p = \sqrt{\frac{PQ}{n}}$. Por lo tanto, la nueva variable aleatoria (p) así generada puede ser estandarizada a la curva normal a través de $Z = (p - P) / \sigma_p$.

Sabemos que la desviación estándar de esta distribución hipotética de muestreo de \underline{p} se llama error estándar. En este punto se debe tener en cuenta que la distribución de muestreo referida es hipotética, ya que en la realidad se conoce un solo punto de toda la distribución de muestreo.

En este caso (es lo común) el valor de la tasa en la población no se conoce, porque su valor depende de toda la población y por tanto su valor es estimado a partir de una muestra, de ahí que se ha desarrollado un estadístico para estimar el error estándar a partir de la muestra.

$$s_p = \sqrt{\frac{pq}{n}}$$

Para materializar estos conceptos tomemos un ejemplo sobre la evaluación del nivel de anticuerpos circulantes contra la fiebre aftosa, medidos a través del índice de seroprotección en ratón lactante (ISP). Se ha tomado una muestra de 400 bovinos entre 1-2 años, 60 días después de vacunados, en una cierta región, para determinar el porcentaje de bovinos que tienen $ISP \geq 2$ y el porcentaje que está por debajo de ese nivel. Los datos provenientes de la muestra informan que de los 400 hay 384 con $ISP \geq 2$.

$$\%(ISP \geq 2) = \frac{384}{400} \times 100 = 96\%$$

Sin embargo, este porcentaje no puede ser el número exacto ya que es evidente que puede existir algún "error en la estimación" debido a que se trata de una muestra.

¿Cómo se puede calcular entonces la verdadera tasa en la población si tenemos un valor sólo aproximado en base a una muestra?

En realidad no existe una solución exacta para el asunto recién planteado, sino que la teoría estadística de la estimación permite decir, con una seguridad determinada, entre que valores límites debe estar comprendido el verdadero valor de la población.

Se supone, para llegar a una conclusión definida, que las estimaciones que pueden obtenerse en base a muestras grandes de una población se distribuyen normalmente, esto es, que el error de muestreo se distribuye de acuerdo con una curva normal.

Por lo tanto, si esto sucede, las muestras darán con mayor frecuencia valores muy cercanos al verdadero de la población, encontrándose muy pocas muestras que den valores "muy alejados" del verdadero.

Siendo: p = tasa encontrada con la muestra
 $q = 1 - p$
 n = número de individuos de la muestra

Para el caso de bovinos vacunados contra fiebre aftosa se tiene que:

$$\text{error estándar de } p: s_p = \sqrt{\frac{(96)(4)}{400}} = 0,98 \approx 1\%$$

3. Estimación de una tasa por intervalo de confianza

El intervalo está dado por dos números entre los cuales el parámetro (P) debe caer con cierta probabilidad de error.

Esta forma de estimación permite indicar la precisión del procedimiento al acompañar al tamaño de la estimación, la probabilidad de error o el grado de incertidumbre de ella. La amplitud del intervalo indica con cuanta precisión ha sido estimado el parámetro. Si el intervalo de confianza es pequeño, se puede decir que la estimación es muy precisa; por el contrario, si el intervalo de estimación es amplio, la estimación del parámetro correspondiente es poco precisa, para una misma probabilidad de error.

Dado a que en la curva normal se verifica que desviaciones en torno del valor central (media $\mu_p = P$) se distribuyen de la siguiente manera:

<u>Desviación esperada (error)</u>	<u>Probabilidad que suceda</u>
1 σ_p	68,27%
2 σ_p	95,45%
3 σ_p	99,73%

Se deduce que desviaciones que pueden deberse únicamente al proceso de muestreo y que excedan 2 veces el error estándar, solamente van a encontrarse en cinco muestras de 100 (4,55% más exactamente) y que desviaciones que exceden tres veces la desviación típica, solamente se van a presentar en tres de cada 1.000 muestras.

Por esta razón los intervalos son llamados de intervalos de confianza al 68,27%, al 95,45%, al 99,73% para la estimación de $P = \mu_p$. Los valores extremos de esos intervalos son llamados de límites de confianza al 68,27%, al 95,45%, al 99,73%. Colocado así el problema, al tener una sola muestra

$$p \pm 1,96 s_p$$

son los límites de confianza al 95% para P. El porcentaje de confianza se llama nivel de confianza. Valores numéricos como 1,96 son valores críticos de Z (abscisa de la curva normal) que algunos autores llaman también coeficiente de confianza.

En resumen, tenemos una muestrea grande ($n \geq 30$), de tamaño fijo, seleccionada al azar a partir de una población citada y la tasa de elementos con un atributo es p ; entonces los límites de confianza de P son dados por

$$p \pm Z\sigma_p$$

Dado que el valor de σ_p se desconoce también, para obtener el límite de confianza antes mencionado se emplea el valor correspondiente a la muestra

$$S_p = \sqrt{\frac{pq}{n}}$$

que es satisfactorio siempre que $n \geq 30$. Cuando $n < 30$, esta aproximación deja de ser satisfactoria y debe utilizarse un procedimiento exacto basado en la distribución binomial.

En otras palabras, si una muestra de 400 bovinos ha dado un porcentaje del 96% de bovinos con $ISP \geq 2$, existe una probabilidad de 95,4% que la tasa de bovinos con $ISP \geq 2$ en la población esté comprendida entre:

$$4 + 2 (1) = 6\% \text{ y } 4 - 2 (1) = 2\%$$

Esto aclara entonces la idea de "error controlable" en el sentido de que con ello se quiere significar que disponemos de un medio objetivo para medir este error, siempre que el proceso de selección de las unidades de la muestra se haga de acuerdo con una técnica bien definida y se conozca exactamente la "probabilidad" con que cada individuo puede participar en la muestra.

MUESTREO PROBABILISTICO

En realidad el muestreo probabilístico tiene como mecanismo de selección de los elementos a incluir en la muestra la aleatoriedad, lo que evita toda subjetividad consciente o inconsciente.

En la práctica una muestra seleccionada en forma aleatoria simple es obtenida elemento tras elemento. Los elementos de la población son numerados desde 1 a N y son seleccionados al azar hasta completar un número igual al tamaño de la muestra (n) establecida.

Aquí caben dos posibilidades. La selección de los elementos se puede hacer con devolución del elemento al seno de la población (con lo cual puede volver a ser elegido más de una vez) o sin devolución del elemento a la población (con lo cual puede ser elegido sólo una vez).

Cuando la selección se hace con devolución, cada selección es un evento independiente ya que la probabilidad de cada elemento de ser elegido para cada una de las elecciones es $1/N = 1/5$.

Si se tiene una población formada por $N = 5$ elementos (letras)

A E L R S

Se decide obtener una muestra de tamaño $n = 3$. La probabilidad de cualquiera de las cinco letras en cualquiera de las tres selecciones es $1/5$, ya que se hace devolución. Por ejemplo, la probabilidad de obtener A en la primera selección es $1/5$ y es la misma en cualquier otra selección, haya o no haya sido seleccionada anteriormente.

Se trata de eventos independientes (modelo binomial), es decir, la probabilidad de acontecer A en la primera selección no afecta la probabilidad de obtener A en cualquier otra selección (probabilidades independientes).

En cambio, cuando la selección aleatoria se hace sin devolución, como ocurre casi siempre en la práctica, cada selección es un evento dependiente de las selecciones precedentes, ya que la probabilidad de ser elegido un elemento en una extracción está condicionada a los resultados de las extracciones anteriores. De ahí que en este tipo de selección se trata de probabilidades condicionales (modelo hipergeométrico).

Así, la probabilidad de seleccionar R en la primera extracción es $1/N$. La probabilidad de seleccionar cualquiera de las restantes en la siguiente extracción es $1/(N-1)$ y así sucesivamente. Entonces la probabilidad de seleccionar un elemento específico, como ser la letra R del universo de cinco letras ya citado, en cualquier selección, supongamos la tercera, es igual a una probabilidad condicional cuyo valor es igual a la probabilidad de no ser seleccionada en las dos primeras extracciones por la probabilidad de ser elegida en la tercera elección.

Para la población de cinco letras:

<u>Nº de extracciones</u>	<u>Probabilidad</u>	
	<u>Ser elegida</u>	<u>No ser elegida</u>
1ª	$1/N = 0,20$	$(N-1)/N = 0,80$
2ª	$1/(N-1) = 0,20$	$(N-2)/(N-1) = 0,75$
3ª	$1/(N-2) = 0,33$	$(N-3)/(N-2) = 0,67$

entonces la probabilidad de obtener R en la 3ª extracción es

$$0,80 \times 0,75 \times 0,33 = 0,20 = 1/5 = 1/N$$

o sea, $1/N = 1/5$ es la probabilidad de R de ser seleccionada en la tercera extracción, a condición que no haya sido seleccionada en las extracciones anteriores.

El muestreo sin devolución de los elementos a la población es como si se tratara de hacer un muestreo a partir de un universo finito, en cambio el muestreo con devolución se asemeja a lo que ocurriría al hacer un muestreo de una población infinita.

ERRORES EN EL MUESTREO

Cualquier estimación que se haga a partir de una muestra, como se ha visto, está sujeta a errores. Estos errores se pueden clasificar en errores de muestreo y errores que no provienen del muestreo. Cualquier sistema de colecta y procesamiento de datos está expuesto a una variedad de errores. En tales sistemas los errores que no provienen del muestreo frecuentemente son de mucho mayor magnitud que los errores de muestreo.

Anteriormente se ha definido lo que son errores de muestreo entendiendo por tal la discrepancia que casi siempre se presenta al hacer una estimación, a partir de una muestra, entre este valor y el que se habría obtenido al hacer una enumeración completa de todos los elementos de una población. La diferencia entre esos dos valores es lo que se llama error de muestreo. Al utilizar métodos de muestreo probabilístico, en que la probabilidad de selección de cada elemento es conocida, es posible, como ya se ha comentado, controlar el error de muestreo en forma satisfactoria, ya que es posible establecer el tamaño de la muestra necesario para mantener el error de muestreo dentro de límites deseados. Podríamos decir que el error de muestreo puede ser evitado si se hace el estudio en

forma exhaustiva es decir, considerando todos los elementos de la población.

Precisamente el hacer un estudio en un número mayor de individuos generalmente involucra aumentar los errores no de muestreo como son los errores de medida, de no respuesta, de procesamiento, error en la selección de los elementos.

Uno de los factores que se debe tener en consideración al planear una muestra se relaciona con los errores que se originan por la dificultad en hacer las mediciones o apreciaciones en el material de estudio, ya se refiera al objeto que debe medirse, ya sea por el instrumento o métodos usados para hacer las mediciones, o por las propias influencias del observador.

Existen los llamados errores relacionados con el objeto medido. Este tipo de error se comete cuando las condiciones en que se verifican las mediciones no son iguales para todos, por ejemplo, peso de bovinos hecho en diferentes condiciones de alimentación reciente, etc. originan por sí solo resultados diferentes.

Otra fuente de errores es la relacionada con el instrumento de medida, error que puede deberse ya sea a una mala calibración del instrumento que marca "sistemáticamente" de menos o de más (solamente en un solo sentido para todas las medidas) o que no posee un grado de ajuste que permita una observación lo más exacta posible.

Dentro de este tipo de error debe también incluirse el producido por "una técnica defectuosa para realizar las medidas originadas entre otras causas por la carencia de instrucciones claras sobre la forma como deben hacerse para que todas ellas se hagan en la misma forma y condición".

También existen los errores dependientes del observador. Si bien la buena reputación de un técnico es una medida razonable, en ciertas circunstancias, para confiar en los resultados de una investigación, ella debe basarse en hechos "objetivos" fáciles de constatar, ya que de otra manera ello puede conducir a resultados erróneos.

A veces la falta de objetividad en este asunto, hace tener demasiada confianza en los resultados de una investigación, sin que existan razones precisas y objetivas para ello.

Como ejemplo de que el técnico puede equivocarse en la apreciación de

un mismo asunto en diversas oportunidades se anota el dato por Birkelo y colaboradores (J.A.M.A. Vol. 133, pág. 359, Febrero, 1947) en el cual 5 "expertos" en diagnósticos radiológicos examinaron separadamente una a una 1.256 radiografías del tórax, encontrando con diagnóstico positivo de tuberculosis números que variaban entre 59 y 100.

Vale la pena señalar que de los 1.256 casos solamente 27 fueron clasificados como positivos por estos especialistas.

Al ser examinadas nuevamente las mismas 1.256 placas por los mismos radiólogos hubo algunos cambios en el diagnóstico. El porcentaje de los positivos en el primer examen que fueron hallados negativos en el segundo, varió de 7% a 41%, o sea que hubo un experto que en un segundo análisis cambió de opinión en el 41% de las veces. El porcentaje de los negativos en el primero que fueron encontrados positivos en el segundo, varió de 6% a 19%.

Debe tomarse como dato curioso que un radiólogo que en el primer examen encontró 59 casos positivos, en el segundo confirmó la positividad de 55 solamente; pero encontró 23 nuevos casos positivos que había diagnosticado en el primer examen como negativos.

Se cita este ejemplo, por lo tanto, para llamar la atención en cuán diferentes deben ser los diagnósticos de otra naturaleza, que por su complejidad deben diferir sin duda extraordinariamente entre sí para diferentes facultativos.

Los errores de no-respuesta se refieren al fracaso en medir el carácter que se quiera estudiar en algunos de los elementos incluidos en la muestra. Los errores de no-respuesta son apreciables especialmente cuando se trata de entrevistas o de encuestas por correo. Las razones de la no-respuesta están dadas por los siguientes factores: a) fracaso en la localización de algún elemento a ser incluido en la muestra; b) incapacidad para proporcionar la información requerida; c) rechazo a la entrevista.

El efecto de la no-respuesta generalmente se traduce en la introducción de vicio muchas veces difícil de medir y en la estimación de un parámetro por límites de confianza a partir de los resultados muestrales.

Supóngase que se estuviera realizando un estudio en la población de establecimientos ganaderos de una región. Esta encuesta fracasó en obtener información del 25% de esta población. Entonces se puede decir, por ejemplo, que no se conoce nada acerca de la frecuencia relativa de bovinos reaccionantes a

la prueba de tuberculinización en esa parte de la población de rebaños o establecimientos ganaderos, excepto que la tasa de reaccionantes debe estar entre 0% y 100%. Se puede llegar a creer y de ahí afirmar que la tasa referida no debe diferir mucho de aquella observada en la parte estudiada de la población, pero tales aseveraciones no pasan de ser opiniones en lugar de hechos. Algunas veces es posible escuchar o leer que las no-respuestas han sido "compensadas" por substitución de los ganaderos "difíciles" o "no cooperadores" por otros asequibles. Las substituciones no resuelven el problema básico de la no-respuesta. Ellas solamente aumentan el tamaño de la muestra en aquella parte de la población que es sensible a las actividades de salud animal.

Personas sin experiencia en organizar planes de muestreo generalmente no toman precauciones, presumiendo que la fracción de no-respuesta es despreciable. Efectivamente es difícil prever cuanta no-respuesta será encontrada. Sin embargo, quien elabora diseños de muestreo debe admitir y estar preparado para enfrentar problemas inesperados, porque con certeza se presentarán.

Como se verá en la planificación de una encuesta por muestreo los mecanismos de divulgación masiva de los objetivos de la encuesta o la opinión pública contribuyen a disminuir substancialmente los rechazos y los avisos previos a los encuestados evitan los desencuentros con los ganaderos. De todas formas ocurren no respuestas. Una manera de encarar este problema es concentrarnos en estos estancamientos tomando una submuestra de estos casos para obtener información de ellos. Existen varios autores que han tratado este problema.

El hecho que algunas personas aumenten el tamaño de la muestra en un tercio por encima del número necesario para las condiciones dadas, no soluciona el problema de la fracción de no-respuesta. Sólo permite en la mayoría de las veces, evitar que la no-respuesta disminuya a un nivel indeseable el número absoluto de encuestados.

Los errores de procesamiento incluyen las informaciones deliberadamente falsas proporcionadas por el entrevistado. Por otra parte los datos básicos pueden ser registrados erróneamente, codificados equivocadamente, manipulados en forma deficiente e incluso omitidos completamente.

DISEÑO DE LA MUESTRA

El llamado diseño de la muestra se refiere a los aspectos que deben ser

tenidos en cuenta desde el punto de vista estadístico para llegar hasta la obtención de la muestra.

En forma resumida, se puede decir que hay 3 aspectos a considerar en el diseño de la muestra:

- a) el tamaño de la muestra (n), es decir, el número de elementos a ser incluidos en la muestra;
- b) el método por el cual se hará la selección, a partir de la población, de los elementos y/o grupos de elementos que componen la muestra, y
- c) la forma como se harán las estimaciones de los parámetros que se quieren conocer.

1. Determinación del tamaño de la muestra (n)

En un estudio por muestreo una decisión importante es determinar el valor n . Esta decisión no siempre es satisfactoria debido a que muchas veces no hay información adecuada para hacerla.

Hasta ahora, ambos ejemplos, se ha supuesto un tamaño de muestra dado, interesa pues analizar brevemente cuales son los elementos que determinan la magnitud de n .

Fundamentalmente hay cuatro elementos técnicos que condicionan el tamaño de una muestra. Por otra parte hay un quinto elemento de extraordinaria importancia práctica, el monto de los recursos (financieros, humanos y materiales) sin cuyo concurso no es posible garantizar estimaciones confiables.

a) Variación de la población

Es fácil interpretar que el grado de heterogeneidad, indicado por la magnitud de la desviación estándar, condicionará el tamaño de la muestra. Donde la variable en la población se distribuya uniformemente, sólo será necesario unos pocos elementos de muestra para tener una idea bastante precisa de lo que ocurre en la población para la variable que se investiga. En cambio, en poblaciones muy heterogéneas para la variable investigada, será necesario un tamaño de muestra bastante grande para poder realizar estimaciones sin riesgo de grandes errores. Hay pues una relación directa entre n y la desviación estándar.

b) Precisión de la estimación (P - p)

La primera interrogante en este sentido que debe ser respondida es ¿cuán pequeño se quiere que sea el error de muestreo? Ésta es una forma de abordar la pregunta esencial ¿cuán precisa se quiere que sea la estimación? Este problema generalmente se resuelve por especificación del error estándar del estadístico correspondiente al parámetro a estimar. Ya se ha visto que hay buenas razones teóricas para esperar que la mayoría de las estimaciones en estudios por muestreo tengan distribuciones aproximadamente normales. De esta manera, la probabilidad de que la estimación hecha no se aleje del verdadero valor en la población en más de 2 veces (aproximado) el error estándar es de 19/20 (95%). De esta forma vemos que el conocimiento del error estándar permite calcular los límites del error para la estimación hecha, en los cuales se puede tener bastante confianza.

La especificación de la precisión debe ser hecha por los técnicos que van a utilizar los resultados de la encuesta por muestreo.

Frecuentemente esta especificación es un tanto arbitraria, sin embargo, ella debe ser hecha en forma cuidadosa, conscientes de que cualquier requerimiento excesivo de precisión se traducirá en un mayor gasto de recursos. Por otra parte, una definición de baja precisión debe quedar claro que disminuye la utilidad de los resultados.

Al especificar el tamaño del error de muestreo se debe considerar todo lo que se sabe acerca de las otras fuentes de error. En una encuesta por muestreo sobre la tasa de morbilidad de una enfermedad "E" no se debe especificar una gran precisión, por ejemplo 3%, si las técnicas de diagnóstico son deficientes y el personal poco entrenado en el uso de esas técnicas, de forma tal que si el estudio fuese hecho en toda la población resultaría que la tasa de morbilidad sería aproximada sólo dentro de un 12%. En este caso, se podría considerar una precisión más baja (mayor error de muestreo) puesto que las otras fuentes de error tornan la estimación imprecisa.

El tamaño de la muestra deberá ser el necesario para reducir el error estándar de la estimación hasta la magnitud elegida, es decir hasta alcanzar la precisión deseada. Se concluye pues que hay una relación inversa entre tamaño de muestra y error de muestreo.

c) Nivel de confianza (1 - α)

Es conveniente hacer ver que una vez definida la precisión el problema no está resuelto, ya que nadie puede garantizar con certeza que el valor estimado caerá dentro de los límites. Por grande que sea el tamaño de la muestra, existe una "chance" de obtener una muestra "desafortunada", la cual tenga un error de muestreo mayor que el especificado.

Un veterinario que estudia la situación sanitaria del ganado en una región, quiere conocer qué porcentaje de ganaderos vacunan sus bovinos contra la fiebre aftosa.

Quiere hacer un muestreo aleatorio y necesita conocer el tamaño de la muestra a obtener. Especifica la precisión, indicando un $\pm 10\%$, de manera que si la muestra arroja aproximadamente 40% de rebaños vacunados, el porcentaje para la población caerá entre 30% y 50%.

Es conveniente decirle al veterinario que esto no se puede garantizar con certeza, salvo que se estudiase toda la población. Como se ha dicho, por grande que sea n existe una "chance" de obtener una muestra desfavorable, en la cual el error de muestreo ($P-p$) sea mayor que el planteado.

Podemos suponer que el médico veterinario está dispuesto a correr el riesgo de 1 en 20 (0,05) de obtener una muestra desafortunada, en la cual ($P-p$) sea mayor que 10%.

El nivel de confianza que representaba la probabilidad de que la estimación sea verdadera, también tiene una relación directa con el tamaño de muestra a través del coeficiente de confianza Z . Mientras mayor probabilidad de acierto se desea tener, más grande deberá ser el tamaño de muestra.

d) Tamaño de la población

Otro elemento que es indispensable analizar es la relación que existe entre los tamaños de muestra y población. De poblaciones numerosas cabrá esperar muestras grandes y vice-versa.

Del equilibrio de todas estas condicionantes, se determina la magnitud del tamaño de una muestra, como se verá en páginas posteriores.

e) Recursos

Este elemento, si bien no entra dentro de la determinación estadística

del tamaño de muestra, juega un papel importante. Existe toda una problemática entre la compatibilización del tamaño de muestra determinado estadísticamente y el tamaño de muestra que resultaría del monto de recursos disponibles.

En general, el monto de recursos determina una muestra menor que la que resultaría de la aplicación de las fórmulas. La muestra menor determinará una menor precisión; será necesario analizar si esa precisión resultante garantiza la obtención de estimaciones confiables. Es en este punto donde hay que decidir: o se renuncia a la precisión especificada y se acepta la resultante de la limitación de recursos, o se destinan mayores recursos. Las dos alternativas restantes son: una combinación de los dos procesos anteriores, o la decisión de no llevar adelante el estudio por estos métodos.

2. Ecuación que relaciona "n" con algunas de las condiciones específicas

La ecuación que permite establecer cual es el tamaño de la muestra "adecuado" se puede deducir directamente del intervalo de confianza de una tasa.

$$P = p \pm Z \sqrt{\frac{pq}{n}}$$

La precisión para un α (nivel de confianza) dado es:

$$P - p = \pm Z \sqrt{\frac{pq}{n}}$$

donde Z es un coeficiente (abscisa de la curva normal) que corta el área α en ambas colas de la distribución normal. Todo esto en el supuesto que se trata de una muestra grande obtenida mediante un mecanismo aleatorio simple a partir de una población. Despejando se tiene $n' = Z^2 PQ / (P-p)^2$ y finalmente considerando el tamaño de la población

$$n = n' / 1 + (n' / N)$$

El error de muestreo ($|P-p|$) no debe ser mayor que un 20% del valor esperado de la tasa poblacional (P).

Calculado el valor de la muestra se tiene la fracción de muestreo

$$f = n/N$$

Esta es la fórmula del tamaño de una muestra aleatoria simple, que garantiza la estimación de una tasa con los requisitos de precisión y confianza o probabilidad de acierto especificados. Es necesario prestar atención a estos conceptos mencionados; Z , como se había adelantado es el coeficiente de confianza que provenía de la distribución normal estandarizada (media = 0 y desviación estándar = 1) y estaba automáticamente determinado al elegir el nivel de confianza.

Una interpretación para este concepto en este caso, sería: la proporción dentro de todas las muestras posibles, que entregan resultados para la tasa, que difieran respecto de la tasa verdadera (P), en más del valor ($|P-p|$) especificado.

En cuanto a la precisión, representada por la magnitud de ($|P-p|$), se había adelantado que quien está en mejores condiciones para decidir sobre el desvío máximo que se puede tolerar es quien hace el estudio, porque sabe cuales serán los fines de la estimación. No está demás señalar que una alta precisión ($|P-p|$ pequeño) sólo se podrá conseguir a expensas de una muestra grande, y que muestras pequeñas sólo pueden entregar resultados poco precisos. Los juicios anteriores son válidos cuando el grado de heterogeneidad en la población (magnitud del desvío estándar) es apreciable. En poblaciones homogéneas, una muestra pequeña puede ser suficiente para lograr una precisión aceptable.

A continuación se entrega una tabla de gran utilidad para servir de referencia. En ella se han tabulado una serie de valores posibles de tasas (entre 1% y 99%, en las columnas).

Las filas consideran el tamaño de la población (N). En el cuerpo de la tabla aparecen los tamaños de muestras necesarios para estimar una tasa poblacional (%), teniendo en cuenta un error de muestreo de 5% en relación al valor de la tasa. Por ejemplo, si la tasa de prevalencia de infección por fiebre aftosa es 20%, un error de muestreo relativo de 5%, significa que los límites quedan dados por $20\% \pm 1\%$.

Tamaño necesario de la muestra para estimar tasas*

Tamaño de la población	Valores de tasas												
	1%	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%	99%
50	50	50	50	50	50	49	49	48	47	45	39	32	12
100	100	100	100	99	98	96	94	92	87	80	64	45	14
200	200	199	198	194	190	184	178	168	154	132	93	58	15
500	499	492	485	465	440	415	380	337	285	218	128	70	16
1.000	994	967	940	860	790	700	610	507	398	278	145	75	16
2.000	1.975	1.872	1.760	1.520	1.300	1.080	880	678	496	323	158	78	16
5.000	4.841	4.270	3.700	2.800	1.200	1.600	1.200	851	582	357	166	80	16
10.000	9.384	7.449	5.800	3.900	2.700	1.900	1.400	930	618	370	168	81	16

* Error de muestreo no mayor que 5% de la tasa.

MÉTODOS DE SELECCIÓN DE LOS ELEMENTOS

1. Selección al azar simple

Muestra al azar simple o muestra irrestricta al azar es aquella en que todas las unidades de la población tienen la misma probabilidad de ser elegidas para formar la muestra. Con el objeto de evitar la acción de las preferencias subjetivas, ya sean conscientes o inconscientes, se acostumbra usar dispositivos mecánicos especiales o bien tablas de números aleatorios. Entre las más conocidas están las tablas construidas por Tippett y las de Fischer y Yates. La aplicación práctica de las tablas de números aleatorios se hace de la siguiente forma:

- a) se numeran los elementos del colectivo;
- b) se determinan cuantos elementos tiene el colectivo a fin de saber con cuantas cifras se debe operar;

- c) se elige al azar un punto de la tabla por el cual se va a comenzar y el orden que se va a seguir, es decir, si se sigue hacia arriba, hacia abajo o hacia el lado;
- d) se eligen los elementos que corresponden a los números que aparecen en la tabla para construir la muestra. Si las cifras que aparecen son mayores que el número de casos que se desea tomar, se procede en la forma que se explicará en base al siguiente ejemplo:

En un predio hay una población de 86 bovinos distribuidos en la siguiente forma:

<u>Potrero</u>	<u>Bovinos</u>
I	35
II	29
III	22

Se necesita sacar una muestra de 6 bovinos.

Solución: como se trata de 86, se necesitan números de dos cifras únicamente. Al consultar la tabla, se obtienen los siguientes números:

33	68
27	87 (rechazado por ser mayor que 86)
	57
13	21

Para determinar a que potrero corresponden los bovinos elegidos, se toma como marco una lista en que aparecen primero los 35 bovinos del primer potrero, en seguida los 29 del segundo y finalmente los 22 del tercero. Si se aplica este criterio, la muestra queda integrada por los siguientes bovinos:

Del I potrero: bovinos n^os 33, 27, 13 y 21.

Del II potrero: bovino n^o 22 (o sea el número 57, que resulta de 35 + 22).

Del III potrero: bovino n^o 4 (o sea, el número 68, que resulta de 35 + 29 + 4).

El procedimiento aleatorio simple se utiliza principalmente como referencia para evaluar la eficiencia de otros métodos de muestreo. Sin embargo, en algunos casos es un procedimiento práctico, como ser cuando:

- la población no es muy variable para el carácter estudiado;
- no se sabe de la existencia de subpoblaciones entre las cuales hay marcadas diferencias en la presentación del carácter;
- la población no es muy grande;
- es posible tener un listado completo de todos los elementos componentes de la población.

La forma de seleccionar elementos mediante un mecanismo aleatorio se puede hacer a través de procedimientos como extracción de boletas numeradas de una urna, discos numerados de una bolsa (lotería) o más comúnmente utilizando una tabla de números aleatorios.

Al querer seleccionar una muestra al azar de $n = 12$ elementos a partir de una población de $N = 500$ elementos, se aplica el siguiente procedimiento:

- Se asigna a cada elemento de la población un número entre 001 y 500 (tres dígitos).
- Se decide un mecanismo arbitrario (incluso puede también ser al azar) para leer la tabla de números aleatorios, por ejemplo columna 25, fila 14.
- Se decide como formar los grupos de tres dígitos a partir de la intersección de la columna y la fila elegida. Generalmente se toman los dígitos (3) juntos a la derecha de la columna elegida (columnas 25, 26 y 27). El número formado en este caso es el 046.
- Se decide en que dirección continuar la extracción de números. Supóngase que se decide leer hacia abajo y luego seguir leyendo de arriba hacia abajo en las tres columnas siguientes (28, 29 y 30). En este ejemplo, como se puede apreciar, no es necesario tomar esas 3 columnas siguientes ya que los 12 elementos son seleccionados en las columnas 25, 26 y 27.
- Si el número formado es menor que 500, el elemento al cual identifica es incluido en la muestra.

Si el número formado fuese mayor que 500 o siendo menor que 500 es repetición de un número ya elegido, no se toma en cuenta. El proceso así de selección continúa hasta completar $n = 12$

127 - 149 - 186 - 439 - 329 - 234 - 032 - 190 - 467 - 048 - 014 - 231.

NUMEROS ALEATORIOS

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
20	75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
21	16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
23	42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
25	68089	01122	51111	72373	06902	74373	96199	97017	41273	21546
26	20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
27	58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
32	08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
33	59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
36	45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
37	89300	69700	50741	30329	11658	23166	05400	66669	48708	03887
38	50051	95137	91631	66315	91428	12275	24816	68091	71710	33258
39	31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
41	44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
43	46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
44	83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
46	91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
47	55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
48	85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
49	07521	56898	12236	60277	39102	62315	12239	07105	11844	01117

2. Muestreo estratificado

En muchas ocasiones es posible subdividir una población en partes o proporciones de acuerdo con diferentes criterios. Puede hacerse sobre una base geográfica, es decir, dividiendo el área total en subáreas. En otros casos, se puede subdividir con referencia a alguna otra cualidad de la población, como ser de acuerdo con el sexo, edad, tamaño del rebaño, etc; o bien, puede subdividirse según el grado en que sus elementos presentan la cualidad que ha servido para definirla. Por ejemplo, si se trata de estudiar la producción de carne de un grupo de bovinos, puede distinguirse un grupo superior, uno medio y otro inferior, sobre la base de informaciones previas. Cada una de las porciones recibe el nombre de "estrato". Los estratos pueden o no estar compuestos del mismo número de unidades. Deben excluirse mutuamente, es decir, no deben tener elementos comunes, a fin de que la suma de los elementos de todos los estratos en que se ha dividido la población sea igual al número de elementos de ella. Deben, además, ser lo más homogéneos posibles dentro de sí.

Cuando se desea comparar diferentes sectores de una misma población, a los cuales por eso se les da el nombre de "dominios de estudio", conviene dividir la población de manera tal que éstos coincidan con los estratos o que incluyan dos o más estratos. En ningún caso debe un mismo estrato quedar repartido entre dos dominios.

En muchas ocasiones es posible subdividir o estratificar la población de acuerdo con algún criterio o incluso la manifestación de un carácter. Podría ser según la distribución geográfica, la edad, el sexo, la raza, el tamaño del rebaño, la finalidad. Los estratos deben ser excluyentes, o sea, no incluir un elemento en más de un estrato, al mismo tiempo deben ser exhaustivos entre sí, es decir, deben incluir todos los elementos de la población.

Un ejemplo de estratificación lo da la división de los rebaños de una región de acuerdo con la finalidad para estudiar el porcentaje de ellos (tasa) que vacunaban contra la fiebre aftosa antes de implantar un programa en esa región.

Estratos	V a c u n a c i ó n				Total
	Si		No		
	Nº	%	Nº	%	
Leche	2.100	60	1.400	40	3.500
Carne	6.500	41	9.500	59	16.000
Mixto	7.200	36	12.800	64	20.000
Total	15.800	40	23.700	60	39.500

Se llama muestreo estratificado al procedimiento que consiste en dividir primero la población en estratos y tomar después muestras al azar de cada uno de ellos, para constituir la muestra definitiva. La muestra así obtenida es una "muestra estratificada".

El muestreo estratificado permite disminuir la varianza poblacional, que en el caso de tasas es $\sigma^2 = PQ$. En este procedimiento para reducir σ^2 se construyen estratos con las unidades de muestreo, lo cual permite aumentar la precisión (disminuir la longitud del intervalo de confianza). Sabemos que la variación entre las unidades de muestreo que están en un mismo conjunto constituyen la variación total (varianza total). Si se trata de una población muy variable o heterogénea, con respecto al carácter en estudio, la construcción de estratos permite que esa población heterogénea quede dividida en "partes", cada una de las cuales dentro de sí es menos variable que toda la población. La variación dentro de los estratos es por tanto más pequeña. En el muestreo estratificado aleatorio, la variación entre los estratos (entre las de los estratos) en la población no contribuye al error de muestreo de la estimación de la

media poblacional. De manera que sólo contribuye al error de muestreo la variación dentro de los estratos.

Ventajas del muestreo estratificado aleatorio sobre el simple

- reduce el tamaño de la muestra para un grado de precisión dado y viceversa;
- garantiza una adecuada representación en la muestra de todas las "partes" de la población.

Cuando se pasa de muestreo aleatorio simple a estratificado, dos componentes determinan la reducción de la varianza de la estimación: uno proviene de la eliminación de las diferencias entre medias de estratos, otro proviene de la eliminación de los efectos de las diferencias entre las desviaciones estándar de los estratos.

La reducción en la variación de la estimación del valor poblacional es muy importante como razón para estratificar; sin embargo, en estudios que consideran varios caracteres, una estratificación buena para un carácter puede no serlo para los demás. En esos casos es frecuente (en salud animal) que la estratificación se haga sobre la base geográfica. Este criterio es de mucho valor, ya que muchos caracteres presentan marcadas diferencias según el lugar.

En general, hemos dicho que la estratificación produce una disminución de la varianza de la estimación de la media poblacional.

Cuando la fracción de muestreo es igual o constante en los distintos estratos se dice que se trata de muestreo estratificado con asignación proporcional.

Como lo habitual es que se desconozca el valor de la tasa en la población supongamos que se decide hacer un estudio por muestreo de los rebaños bovinos de la región para saber cual es la tasa de rebaños en que se vacunaba contra la fiebre aftosa antes de comenzar el programa oficial de control de esta enfermedad en la región. Sobre la base de una tasa esperada de rebaños vacunados en la población de 40%, con una precisión de 2% (o sea el 5% de 40%), y con un nivel de confianza de 95% (error de tipo $\alpha \rightarrow 5\%$) el tamaño de la muestra (ver tabla) es de $n = 2.200$ rebaños, lo que da una fracción de muestreo de $f = 2.200/39.500 \cong 1/18$.

Para hacer una asignación proporcional de este tamaño ($n = 2.200$) en los tres estratos, es necesario repartir proporcionalmente este número en ellos de acuerdo con el "peso relativo" (W) de cada estrato en la población.

Estratos	N	%W	n
Leche	3.500	8	176
Carne	16.000	41	902
Mixto	20.000	51	1.122
Total	39.500	100	2.200

De esta manera en cada estrato (sub-población) la fracción de muestreo se mantiene constante, aproximadamente, en $f = 1/18$.

En cada estrato al examinar el número de rebaños que le corresponde se contabilizan los que vacunaban contra fiebre aftosa (a) y con esta cantidad se calculan las tasas por estratos de rebaños vacunados. Sin embargo, la asignación proporcional no es siempre una asignación óptima, en cuyo caso se hace necesario una fracción de muestreo variable (asignación no proporcional).

La asignación no proporcional es especialmente útil cuando:

- las desviaciones estándar de los estratos difieren considerablemente entre sí;
- los costos para reunir datos varían substancialmente de un estrato para otro.

En estos casos habrá una asignación óptima (determinada combinación de fracciones variables de muestreo) que proporcione el menor error de muestreo, al más bajo costo.

La ganancia de precisión por asignación óptima sobre asignación proporcional no es tan grande en la estimación de tasas como la estimación de medias.

Si se trabaja con costos fijos o determinados en la obtención de medición desde cada unidad de muestreo, ellos no varían de estrato a estrato. En ese caso el costo se puede ignorar para determinar la precisión de muestreo en los distintos estratos.

Cuando los costos son variables en la obtención de medición a partir de una unidad de muestreo, de un estrato a otro, se hace necesario incorporar una función de costo.

3. Muestreo sistemático

Este tipo de muestreo exige que los elementos componentes de la población tengan un cierto orden de posición, por ejemplo que estén en filas, en tarjetas, o en cualquier forma ordenada.

Para operar hay que seguir los siguientes pasos:

- ordenar en una lista todos los elementos de la población, asignando un número de orden correlativo a cada uno.
Supongamos que listamos los rebaños de leche, del ejemplo del capítulo anterior, del 1 al 3.500;
- determinar el tamaño de la muestra. Para los rebaños de leche $n = 175$;
- establecer el intervalo de selección (F) y aplicar sobre el listado de la población

$$F = \frac{N}{n} = \frac{3.500}{175} = 20 \text{ rebaños}$$

Es decir, se deben elegir 175 rebaños a partir de 3.500, seleccionando de la lista 1 de cada intervalo de 20;

- determinar el número de arranque, o sea, el primer rebaño a ser seleccionado. Este corresponde a un número elegido al azar entre 1 y 20. Supongamos que ese número sea el 3, o sea, el 3er. rebaño en orden en la fila, del 1 al 3.500 es seleccionado para ser incluido en la muestra;
- a continuación se va sumando el intervalo $F = 20$ a partir de 3 y de esta manera se llegan a seleccionar los 175 rebaños

3 - 23 - 43 - 63 - ... - 343 - ... - 783 - ... - 1.163 - ... -
1.703 - ... - 2.143 - ... - 2.863 - ... - 3.423 - 3.443 - 3.463 -
3.483.

La utilización del muestreo sistemático casi siempre está asociada al empleo de otros procedimientos de muestreo, como el estratificado o el muestreo

en dos etapas. El procedimiento sistemático generalmente se aplica para seleccionar elementos dentro de alguna unidad primaria.

El muestreo sistemático no es un muestreo al azar, a menos que la lista o marco esté "organizado" al azar. Sin embargo, se acostumbra considerar una lista cualquiera, que no está ordenada según un criterio relacionado con el carácter que se quiere estudiar, como una lista al azar.

El peligro de este método consiste en que puede ocurrir que inadvertidamente la lista presente cierta periodicidad en relación con el intervalo elegido, en estos casos la muestra resultante expone a errores serios.

4. Muestreo por conglomerados

Hemos dicho anteriormente que los elementos de una población pueden ser diferenciados tomando en cuenta características secundarias y se ha demostrado como esta circunstancia puede ser aprovechada para formar estratos. En otras ocasiones, ocurre que la población está formada por unidades que se distinguen entre sí por características secundarias (por ejemplo, el tamaño, la edad, el sexo, etc.) pero que se presentan agrupadas ya sea en el espacio o en el tiempo. Los grupos de elementos contiguos que exhiben esta condición reciben el nombre de "conglomerados" y pueden servir de base para un procedimiento de muestreo. Llámase "muestreo por conglomerados" al que consiste en tomar como unidades de muestreo grupos de esta especie y en someter al examen a todos los elementos de cada uno de los conglomerados elegidos. Como es natural, la selección de los conglomerados que compondrán la muestra debe hacerse al azar, o eventualmente en forma sistemática.

Entre el muestreo estratificado y el muestreo por conglomerados existe una diferencia fundamental. En aquel, se trataba de que cada estrato fuese lo más homogéneo posible, es decir, que consistiera en unidades tan semejantes como fuese posible desde el punto de vista de las cualidades que servían de base a la estratificación. En el muestreo de conglomerados el ideal consiste en que contenga elementos de todas las variedades que se han distinguido a base de ciertas cualidades secundarias, o sea, que estén tan mezcladas como se pueda. Si se compara cada uno de los conglomerados con los demás se advierte, por el contrario, que se asemejan mucho, pues cada uno contiene representantes de todas variedades.

Ejemplo de "conglomerado" es el rebaño animal que "habita" en un mismo establecimiento, distrito o municipio, en donde pueden encontrarse bovinos de diferentes sexos, edades, raza y grado de desarrollo, finalidad, manejo, etc.

Dentro de los conglomerados no es necesario "listar" los elementos componentes, ya que todos son considerados.

El costo de las operaciones de muestreo en el terreno para este método es más barato comparativamente con los anteriores, especialmente con el estratificado.

Supongamos que para estudiar la situación de la vacunación antiaftosa, previo al inicio del programa en la región antes citada, se decide hacer un muestreo por conglomerados.

La determinación del tamaño de la muestra, tal como ha sido ya indicado, es $n = 2.200$ rebaños. Ocurre que la región se encuentra subdividida desde el punto de vista administrativo en distritos que tienen diferentes cantidades de rebaños; sin embargo, esas diferencias no son muy marcadas como se puede apreciar en el listado.

<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>	<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>	<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>	<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>
1	120	26	121	51	171	76	149
2	81	27	120	52	162	77	153
3	125	28	119	53	140	78	95
4	140	29	141	54	162	79	115
5	124	30	114	55	115	80	116
6	132	31	163	56	101	81	139
7	104	32	109	57	108	82	130
8	116	33	155	58	137	83	91
9	98	34	139	59	160	84	93
10	123	35	120	60	148	85	170
11	149	36	93	61	90	86	181
12	120	37	139	62	98	87	173
13	138	38	127	63	132	88	162
14	128	39	166	64	167	89	142
15	151	40	116	65	155	90	115
16	128	41	82	66	135	91	135
17	111	42	115	67	170	92	142
18	98	43	130	68	167	93	102
19	117	44	124	69	100	94	119
20	121	45	147	70	97	95	158
21	95	46	156	71	121	96	160
22	132	47	131	72	140	97	173
23	114	48	158	73	138	98	115
24	140	49	136	74	87	99	159
25	164	50	160	75	168	100	133

<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>	<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>	<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>	<u>Distri</u> <u>to</u>	<u>Nº de</u> <u>rebaños</u>
101	173	145	115	189	125	233	182
102	129	146	119	190	132	234	125
103	168	147	135	191	183	235	114
104	115	148	108	192	115	236	175
105	172	149	173	193	173	237	183
106	165	150	185	194	178	238	145
107	166	151	172	195	160	239	182
108	156	152	181	196	143	240	175
109	150	153	180	197	152	241	167
110	157	154	176	198	115	242	163
111	106	155	125	199	125	243	150
112	128	156	132	200	181	244	145
113	170	157	163	201	179	245	172
114	153	158	141	202	182	246	117
115	147	159	125	203	118	247	137
116	162	160	161	204	125	248	184
117	138	161	160	205	134	249	130
118	120	162	169	206	180	250	138
119	139	163	153	207	171	251	174
120	140	164	145	208	144	252	169
121	170	165	170	209	132	253	171
122	170	166	101	210	129	254	148
123	183	167	107	211	140	255	153
124	136	168	125	212	183	256	137
125	173	169	177	213	145	257	127
126	167	170	132	214	115	258	119
127	131	171	123	215	172	259	146
128	125	172	122	216	168	260	135
129	152	173	145	217	122	261	125
130	158	174	170	218	128	262	137
131	125	175	188	219	145	263	147
132	137	176	134	220	173	264	182
133	143	177	125	221	182	265	173
134	148	178	182	222	125	266	161
135	153	179	173	223	173	267	115
136	158	180	115	224	170	268	101
137	170	181	171	225	175	269	97
138	182	182	183	226	153	270	173
139	183	183	172	227	148	271	175
140	175	184	164	228	160	272	183
141	171	185	158	229	163	273	140
142	163	186	158	230	115	274	158
143	168	187	163	231	125	275	127
144	151	188	115	232	111		

Considerando que el número de rebaños en la población es $N = 39.500$ y que hay 275 conglomerados (M) de rebaños (distritos), tenemos una media de $143,64$ (\bar{B}) rebaños por conglomerado.

Para calcular la cantidad de conglomerados a incluir en la muestra (m), de manera que se aproxime a $n = 2.200$, hay que seguir el siguiente procedimiento:

$$m = \frac{n}{\bar{B}} = \frac{2.200}{143,64} \approx 15 \text{ conglomerados}$$

A continuación, mediante un procedimiento al azar simple se eligen 15 conglomerados. Posteriormente, se procede a encuestar a todos los rebaños existentes dentro de cada conglomerado elegido, con lo cual se alcanza un número de rebaños cercano a $n = 2.200$.

Como los conglomerados son de tamaño desigual, llamamos \bar{B} (media) a la cantidad de elementos por conglomerado. Cuando se trata de conglomerados de igual tamaño, raros en la naturaleza, se llama a esa cantidad B . Un experto en muestreo puede "confeccionar" conglomerados iguales a partir de un listado, agrupando elementos.

La probabilidad de selección de cualesquiera de los $N = M \cdot \bar{B} = (275) \cdot (143,64) \approx 39.500$ elementos de la población es

$$f = \frac{m}{M} \cdot \frac{\bar{B}}{B} = \frac{n}{N}$$
$$= \frac{15}{275} \cdot \frac{143,64}{143,64} = \frac{2.200}{39.500} \approx \frac{1}{18}$$

En general, los conglomerados se seleccionan con estratificación, porque de esta manera se reduce considerablemente la varianza.

5. Muestreo en dos etapas

En el muestreo por conglomerados, la unidad de muestreo era el conglomerado (conjunto de elementos agrupados), que en el ejemplo dado era el distrito. Al seleccionar al azar un conglomerado de este tipo se encuestaban todos los elementos (rebaños) existentes dentro del conglomerado seleccionado.

Ahora se trata de aplicar un procedimiento de selección en dos etapas. En la primera se selecciona en forma aleatoria una muestra de conglomerados. En la segunda etapa se elige, de cada conglomerado seleccionado, una muestra

(submuestra) de los elementos en él contenidos. En este caso, los conglomerados pasan a constituir las unidades primarias de muestreo. En este método se requiere sólo un listado completo de las unidades primarias seleccionadas. Este procedimiento permite concentrar los recursos disponibles en un número limitado de lugares o unidades primarias. Otro factor que disminuye el costo es que sólo es necesaria la confección de un listado de los elementos contenidos en los conglomerados incluidos en la muestra. Entre los inconvenientes que presenta está el que el error de muestreo suele ser mayor en el muestreo estratificado (pero menor que en él por conglomerados). Por otra parte, suelen presentarse dificultades en el análisis de los resultados por el hecho de que las unidades primarias de muestreo tienen distinto tamaño, es decir, tienen diferente número de elementos.

El muestreo en dos etapas resulta de la búsqueda de equilibrio entre los dos efectos conflictivos de la conglomeración, por una parte disminuir el costo al concentrar los elementos a encuestar y por otra parte un elevación de la varianza por falta de una mayor distribución de los elementos que llegan a ser incluidos en la muestra.

El muestreo en dos etapas (o en varias etapas) busca disminuir el grado de conglomeración y así disminuir la varianza sin incurrir con esto en un aumento proporcional del costo.

La fracción de muestreo total uniforme y la probabilidad igual de selección de cualquiera de los $N = M \times \bar{B}$ elementos de la población en la muestra de $n = m \times b$ elementos es

$$f = \frac{n}{N} = \frac{m}{M} \cdot \frac{b}{B} = f_1 \cdot f_2$$

Supongamos que la población de $N = 39.500$ elementos (rebaños) en $M = 275$ conglomerados (distritos) de una región, se debe seleccionar una muestra de $n = 2.200$ rebaños. En este caso

$$f = \frac{n}{N} = \frac{2.200}{39.500} \cong \frac{1}{18}$$

Un diseño en dos etapas, posible en este caso de tamaños más o menos semejantes con los conglomerados, podría utilizar fracciones de muestreo $f_1 = \frac{1}{6}$

para distritos (conglomerados) y $f_2 = 1/3$ para rebaños (elementos); así tendríamos 46 distritos (m) con una medida de 48 rebaños por distrito (b). Esta última es llamada submuestra. Así, la probabilidad de selección de cualesquiera de los $N = 39.500$ rebaños de la población es

$$f = \frac{2.200}{39.500} = \frac{46}{275} = \frac{48}{143,64} = \frac{1}{6} \times \frac{1}{3} = \frac{1}{18}$$

En este cálculo se ha colocado la media de 48 rebaños por distritos, pero en el muestreo debe ser un número variable, o sea, un tercio de los rebaños de cada distrito elegido.

Esto implica una selección al azar de 46 distritos y de cada uno de éstos se eligen, también al azar, un tercio de los rebaños (submuestreo). Este procedimiento es válido en este caso porque el tamaño de los conglomerados no varía grandemente, de manera que los rebaños no reciben probabilidad de selección muy diferentes.

Es común combinar el muestreo en dos etapas con la estratificación; en este caso pueden emplearse fracciones de muestreo constantes para los diversos estratos o pueden utilizarse fracciones variables de manera de conservar la fracción global de muestreo $f = n/N$.

6. Probabilidad proporcional al tamaño de los conglomerados y muestreo en dos etapas

En el ejemplo de la sección anterior se ha hecho la selección de los conglomerados y de los elementos dentro de los conglomerados con probabilidad constantes f_1 y f_2 , de manera que $f_1 \cdot f_2 = f = n/N$. Sin embargo, con un diseño de muestreo de este tipo, cuando hay grandes variaciones en el tamaño de los conglomerados, ellas se reflejan directamente sobre los tamaños de las submuestras. Es necesario reducir esa variación en el tamaño de la submuestra.

Cuando hay gran variación en el tamaño de los conglomerados se debe seleccionar un número constante (b) de elementos a partir de los B' en los conglomerados elegidos. Con esto resulta que la selección de las unidades primarias o conglomerados se hace con probabilidad proporcional al tamaño de los conglomerados (B'). Esto significa que

$$f = \frac{B' \cdot b}{Fb \cdot B'} = \frac{1}{F}$$

b = número constante de elementos a seleccionar en cada conglomerado incluido en la muestra.

B' = número (variable) de elementos que componen cada conglomerado.

$$F = \frac{N}{n}$$

Esta muestra resulta autoponderada por el hecho de que las unidades de primer orden (conglomerados) tienen una probabilidad de ser elegidas proporcionales al número de elementos que la componen; en cambio, en la segunda etapa, al seleccionar un número constante de elementos en cada conglomerado elegido, la chance de ser incluidos en la muestra de los elementos de conglomerados grandes es menor que la que tienen los elementos de conglomerados pequeños.

Para poder aplicar el método de muestreo en dos etapas con probabilidad proporcional al tamaño (PPT) de los conglomerados se debe tener un listado de los conglomerados (distritos), acumulando la suma de los elementos. Supongamos que aplicamos este procedimiento al ejemplo anteriormente dado:

<u>Conglomerados (Distritos)</u>	<u>Elementos (Rebaños)</u>	<u>Acumulación</u>	<u>Rango</u>
1	120	120	1 - 120
2	81	201	121 - 201
3	125	326	202 - 326
4	140	466	327 - 466
5	124	590	467 - 590
6	132	722	591 - 722
7	104	826	723 - 826
8	116	942	827 - 942
.	.	.	.
.	.	.	.
.	.	.	.
272	183	39.075	38.893 - 39.075
273	140	39.215	39.076 - 39.215
274	158	39.373	39.216 - 39.373
275	127	39.500	39.374 - 39.500

Se ha decidido incluir 45 distritos de los 275 existentes en la población y para elegirlos en forma proporcional al tamaño se toma una fracción

$$Fb = \frac{N}{m} = \frac{39.500}{45} \approx 878$$

Se elige, por medio de una tabla de números aleatorios, un número de arranque entre 1 y 878. Supóngase que sea el 350. A partir de allí se elabora una secuencia sumando sucesivamente 878 a 350 y a los números que así van resultando

$$350 - 1.228 - 2.106 \dots - 38.104 - 38.982$$

Así se elige el distrito 4, y así hasta el distrito 272. La cantidad constante de rebaños a encuestar (b) en cada distrito se determina de la siguiente manera:

$$b = \frac{n}{m}$$

En el supuesto que $n = 2.200$ rebaños

$$b = \frac{2.200}{45} = 49 \text{ rebaños por cada distrito elegido.}$$

Tomando como ejemplo los 140 rebaños del distrito 4 se sigue el siguiente procedimiento: en el supuesto que se tiene un listado de los 140 rebaños, para determinar los 49 rebaños a elegir se toman 49 números al azar entre 001 y 140 y los rebaños que tienen esos números son incluidos en la muestra.

La probabilidad de ser elegido de cada rebaño es la misma y es igual a la fracción global de muestreo $f = \frac{n}{N} = \frac{2.200}{39.500} = \frac{1}{18}$

$$f = \frac{B'}{Fb} = \frac{b}{B'} = \frac{1}{F}$$

$$f = \frac{140}{878} \cdot \frac{49}{140} = \frac{1}{18}$$

La selección de los 49 rebaños dentro de cada distrito elegido también puede hacerse a través de un procedimiento sistemático.

En el muestreo en dos etapas (submuestreo) con unidades primarias de tamaño variable, es decir, con diferente número de unidades elementales en cada unidad primaria (conglomerado), la estimación de error estándar se hace de la siguiente manera:

$$\hat{\sigma}_p = \sqrt{\frac{PQS}{n}}$$

En este caso S representa al factor por el cual la varianza correspondiente al método de selección simple al azar, debe ser multiplicada para obtener la varianza para el procedimiento de selección en dos etapas. $S = 1 + \delta (b-1)$, en esta expresión δ representa el grado de homogeneidad dentro de los conglomerados. Así cuando las unidades elementales dentro de los conglomerados son homogéneas con respecto al carácter en estudio, en este caso, δ será alto (cercano a + 1) y la varianza total en gran medida será explicada por la varianza entre los conglomerados. En cambio, si las unidades elementales dentro de los conglomerados presentan heterogeneidad con respecto al carácter estudiado, entonces δ llega a ser cercano a cero (en casos excepcionales puede ser negativo) con lo cual la varianza entre conglomerados participa en pequeña medida de la varianza total.

De la expresión dada a S queda claro que al aumentar b (constante de submuestreo) se produce un aumento de la varianza proporcional a cuan positivo sea δ . Por esta razón es importante tener en cuenta el grado de homogeneidad de las unidades elementales dentro de los conglomerados para el cálculo del tamaño de la muestra.

PLANIFICACION DEL MUESTREO

1. Formulación del objetivo

Es de fundamental importancia definir en forma clara y precisa cual es la finalidad que se persigue en un estudio para el cual se hace necesario tomar una muestra. El plan de muestreo deberá obedecer rigurosamente a las necesidades de información que emanan del objetivo planteado. Es diferente la planificación de una muestra para estudiar la producción lechera, la prevalencia de tuberculosis bovina y el grado de educación sanitaria de los dueños o encargados de los establecimientos ganaderos.

Desde otro punto de vista una misma muestra no puede ser igualmente eficiente si se trata de un estudio inicial sobre el crédito a la explotación pecuaria que si se tratara de averiguar el estado inmunitario de la población bovina frente a la fiebre aftosa. Es muy importante ser conscientes de que debe haber concordancia entre lo que se desea estudiar y el tipo de muestra a seleccionar. No existen recetas o soluciones hechas que puedan ser utilizadas

irrestrictamente. Cada estudio planteado involucra la necesidad de estudiar un plan de muestreo específico sobre la base de los fundamentos estadísticos, del conocimiento de la situación particular que se enfrenta y nunca dejando de lado la imaginación y el sentido creativo que la solución de cada problema trae como desafío.

Un aspecto que vale la pena resaltar en este capítulo es la necesidad de un buen conocimiento del problema. De ahí que es necesario que el planeamiento del muestreo se haga en equipo, en el cual participen gente de diversas disciplinas y que además tengan un buen conocimiento del problema en la realidad. Por otra parte, esto implica también el compromiso de la objetividad. Se trata de que los técnicos en diversas disciplinas, todos con un buen conocimiento de la situación en el terreno, colaboren en la forma más efectiva a establecer lo que se quiere estudiar de la manera más simple y directa. Debe tenerse en cuenta que el muestreo no es una ocasión para averiguar todas las interrogantes, y se debe recordar que una muestra no es siempre adecuada para dar respuesta a una variedad de interrogantes.

En la figura que aparece más adelante se presentan las etapas a seguir en el proceso de una encuesta por muestreo.

Los aspectos más resaltantes sobre los cuales son aplicadas las encuestas por muestreo en el campo de los programas de control de la fiebre aftosa son la difusión de la enfermedad, el estado inmunitario de la población, el manejo de la ganadería y las actitudes de los ganaderos.

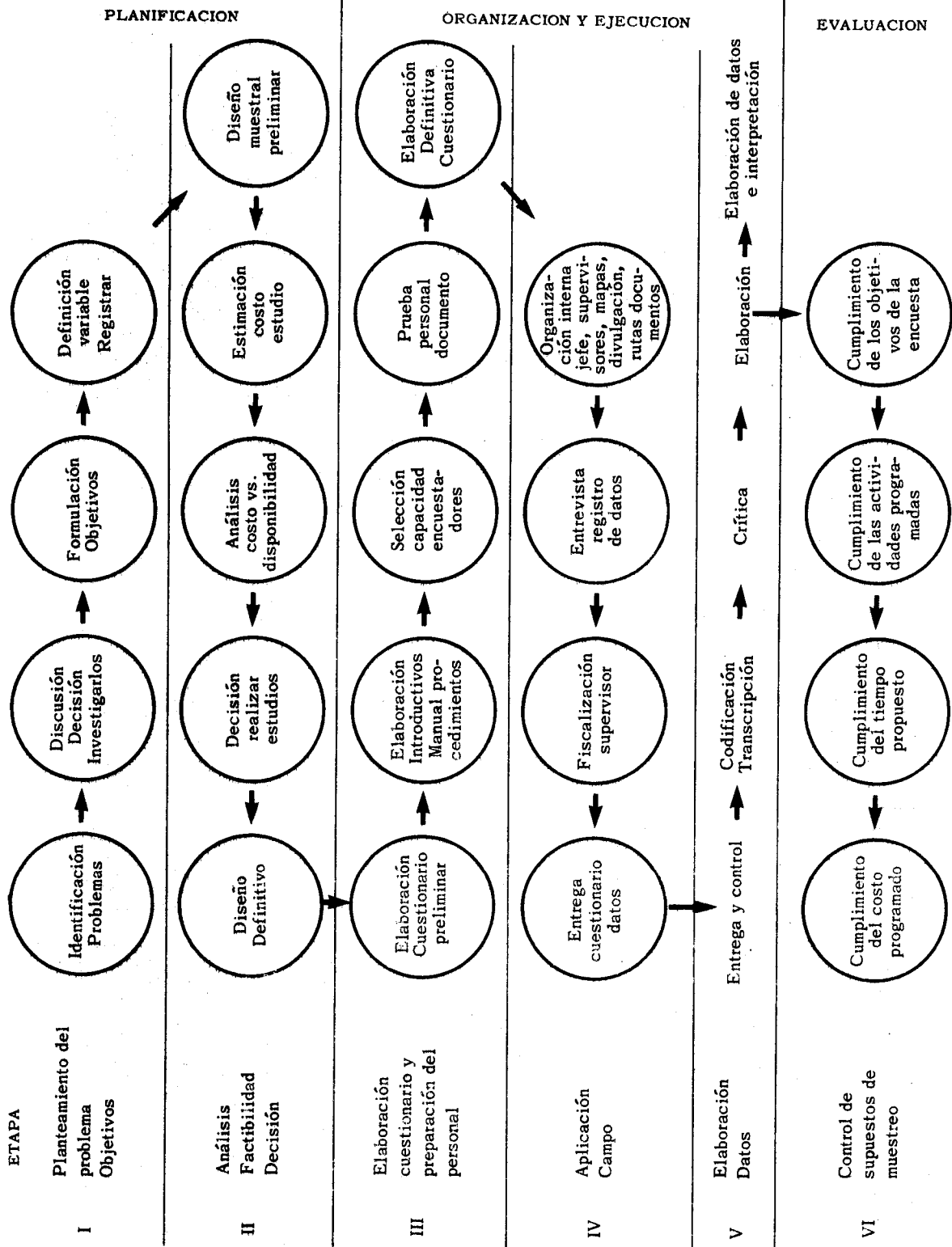
En términos más concretos, entre otros, los objetivos específicos de estas encuestas suelen ser conocer los siguientes aspectos:

- a) frecuencia de la enfermedad clínica
- b) frecuencia de la infección por el virus de la fiebre aftosa
- c) estado inmunitario de la población
- d) presencia de portadores del virus de la fiebre aftosa
- e) eficiencia productiva del ganado
- f) manejo de la ganadería
- g) actitudes de la comunidad frente a los problemas de salud animal.

En términos generales se deben establecer criterios para definir prioridades en las encuestas según:

- a) importancia de la enfermedad

ETAPA EN EL PROCESO DE UNA ENCUESTA POR MUESTREO



- b) posibilidad de identificar los grupos expuestos
- c) cambios demográficos y de la situación ecológica de una región
- d) importancia económica de la enfermedad
- e) servicios disponibles para combatir la enfermedad.

2. Definición de la población y de las unidades de muestreo

Una vez establecido el objetivo del estudio, el paso siguiente es definir y delimitar la población sobre la cual se aplicará el estudio. Este paso debe tener en cuenta de que la observación se realice realmente sobre la población capaz de proporcionar la información requerida, de manera que las conclusiones a que se llegue se apliquen a esa población sin excederla y sin excluir partes de ella. Su naturaleza queda establecida por los objetivos del estudio.

Las unidades de muestreo son aquellas que se utilizan para seleccionar los elementos de la muestra ya que ellas contienen los elementos. Cuando en un muestreo se seleccionan elementos, cada unidad de muestreo corresponde a un elemento. Una unidad de muestreo puede corresponder a un conjunto de elementos, como es el caso de los conglomerados.

La población debe ser definida de acuerdo con el objetivo, estructura de la misma, espacio y tiempo.

A continuación se presenta un esquema sobre características de la población considerando algunos objetivos ya mencionados.

Objetivo	Población	Elemento	Material colectado	Criterios principales de clasif.
1. Prevalencia de fiebre aftosa clínica	Conjunto de establecimientos ganaderos	Cada establecimiento ganadero	Respuestas del encargado del establecimiento	Geográfico, tiempo y tamaño del rebaño
2. Estado inmunitario de la población	Todos los bovinos	Cada bovino	Suero	Período postvacunal y edad
3. Prevalencia de infección por fiebre aftosa en la población bovina	Todos los bovinos	Cada bovino	Suero	Geográfico y edad
4. Detección de portadores	Todos los bovinos	Cada bovino	Material esofágico-faríngeo	Geográfico, tipo de explotación y edad
5. Educación sanitaria de la comunidad	Comunidad rural	Cada persona	Respuesta de las personas	Geográfico, clase social y actividad

3. Información a coleccionar y métodos de coleccionar

Debe haber una estricta correspondencia entre los datos a ser coleccionados y el objetivo planteado. En función del objetivo de la encuesta los datos a registrar deben ser relevantes e indispensables. Además, se debe agregar el criterio de factibilidad y no solicitar información que difícilmente pueda ser obtenida. Toda información no esencial ni factible debe ser omitida. Existe la tendencia a pedir un exceso de datos, algunos de los cuales muchas veces no llegan a ser utilizados.

A continuación se da una lista de posibles informaciones a considerar en las encuestas por muestreo, según el objetivo que se establezca:

3.1 Para conocer la frecuencia de la enfermedad clínica

- a) identificación del rebaño, ubicación y tiempo
- b) tipo de explotación y población bovina
- c) ocurrencia de fiebre aftosa en los últimos 4 años. Para el último año, los meses en que ocurrió
- d) si vacuna los bovinos y cuántas veces al año.

3.2 Para evaluar el estado inmunitario de la población

- a) identificación del rebaño, ubicación y tiempo
- b) población bovina por categorías
- c) fecha de la última vacunación y la frecuencia con que vacuna
- d) nivel de anticuerpos humorales para cada bovino sangrado
- e) ocurrencia de la enfermedad en los dos últimos años. Para el último año la distribución por mes.

3.3 Para conocer la prevalencia de infección

- a) identificación del rebaño, ubicación y tiempo
- b) población bovina clasificada por edad
- c) presencia de anticuerpos anti-VIA para cada bovino sangrado
- d) ocurrencia de la enfermedad en los últimos 4 años. Para el último año la distribución por mes.

3.4 Para detectar bovinos portadores

- a) identificación del rebaño, ubicación y tiempo
- b) población bovina por categorías y tipo de explotación
- c) identificación del virus de la fiebre aftosa para cada bovino al cual se le extrajo material esofágico-faríngeo
- d) ocurrencia de la enfermedad en los últimos 4 años
- e) ingreso de animales según origen.

3.5 Para conocer el grado de educación sanitaria de la comunidad

- a) identificación de la persona, domicilio y tiempo
- b) clase social y actividad
- c) conocimiento de la enfermedad y de su impacto
- d) actitudes frente a la enfermedad.

La producción de datos epidemiológicos originales se puede hacer a través de la observación y también a través de entrevistas a los encargados o propietarios de los animales.

La observación de hechos es un método directo y objetivo de evaluar alguna característica de significación epidemiológica. Está expuesta a los llamados errores de medición cuya presencia afecta la validez de cualquier conclusión. Además de aplicar procedimientos normalizados, el observador debe tener conciencia de la importancia de lo que hace, de manera que la observación no sea hecha en forma descuidada, displicente o rutinaria.

La entrevista a los encargados del ganado es otro procedimiento de colecta de datos epidemiológicos. Mediante este método un entrevistador debidamente entrenado ubica al entrevistado, obtiene la entrevista, formula preguntas y anota las respuestas del entrevistado en un cuestionario debidamente estandarizado y de acuerdo con un manual de instrucciones.

El tipo de preguntas es un problema de sentido común y experiencia. Se debe ser objetivo en el sentido de no hacer preguntas vagas y generales cuando se quiere una respuesta sobre algo específico. El lenguaje debe ser simple y directo y debe hurtar tecnicismos y ambigüedades. También debe evitarse hacer preguntas con supuestos acerca del entrevistado, así como las preguntas estar dirigidas hacia una cierta respuesta.

A continuación se presenta un cuadro donde se clasifican los objetivos del sistema ocasional de acuerdo con el método de colecta de datos.

Objetivos	Método
Conocer la prevalencia de la enfermedad clínica	Entrevista
Evaluar el estado inmunitario de la población bovina	Observación (obtención de suero)
Conocer la prevalencia de infección	Observación (obtención de suero)
Detectar bovinos portadores	Observación (obtención de material esofágico-faríngeo)
Conocer el grado de educación sanitaria de la comunidad	Entrevista

4. Periodicidad de aplicación de las encuestas

La encuesta longitudinal se aplica cuando es necesario realizar estudios periódicos en una misma muestra y la encuesta transversal repetida cuando el estudio se hace en cada período en una nueva muestra.

La encuesta longitudinal es una forma adecuada para estudios de seguimiento ya que permite un estudio más completo de los cambios experimentados por el carácter en estudio. Sin embargo, presenta una serie de dificultades, entre ellas la desaparición de animales que son vendidos o que mueren, por lo cual hay que incorporar un mecanismo que permita la sustitución. En estudios de persistencia sobre anticuerpos humorales neutralizantes o de protección, sobre anticuerpos contra el antígeno (VIA) o sobre portadores del virus de la fiebre aftosa es necesario aplicar encuestas de tipo longitudinal.

La encuesta transversal repetida es un procedimiento válido cuando se quiere averiguar características poblacionales como el estado inmunitario de la población, la prevalencia de la enfermedad clínica, la prevalencia de infección por el virus de la fiebre aftosa y el grado de educación sanitaria de la comunidad.

5. El marco

¿Cómo obtener una muestra probabilística? Sin entrar en detalles, el procedimiento general es el siguiente:

- a) subdividir la población en unidades de muestreo
- b) numerar esas unidades con arreglo a algún criterio
- c) establecer la probabilidad de inclusión en la muestra
- d) con una tabla de números aleatorios, se obtiene la muestra de acuerdo con el sistema de probabilidades establecido.

La obtención de una muestra probabilística necesita de una lista o listado de las unidades de muestreo de la población. La palabra marco (el listado) ha sido propuesta por la Comisión de Muestreo de la O.N.U. Una de las primeras tareas al planificar un muestreo es investigar si hay un marco disponible. Algunas veces surgen dificultades aun con la existencia de un listado de las unidades de muestreo de la población. ¿Por qué? la razón está en que muchas veces esas listas son incompletas, envuelven duplicaciones o contienen informaciones distorcionadas. Sin embargo, los listados son tan importantes para realizar un muestreo que se deben hacer esfuerzos para llegar a disponer de ellos, después de una adecuada revisión y complementación con información adicional. Al no existir un marco, se hace necesario confeccionarlo, lo que puede ser costoso y consumir mucho tiempo.

Normalmente los países disponen de censos ganaderos que son utilizados como marco. Sobre este punto es necesario destacar que dichos censos en muchos casos no pueden hacerse anualmente, cuando más se realizan cada 5 y más años y, en consecuencia, un trabajo adicional que no puede evitarse es la actualización de los listados. Hay establecimientos ganaderos que se forman después del censo, otros que desaparecen y finalmente los que cambian de giro dentro de la misma área de agricultura. Estos cambios deben ser detectados antes de la extracción de la muestra. Es frecuente que en las encuestas ganaderas se desee tener alguna clasificación, al menos sobre las principales variables, según finalidad y agrupaciones cooperativas. Sobre lo mencionado surge en algunos casos otro problema: el de la identificación de todas y cada una de las unidades que conforman la población, dentro de dichas categorías.

6. Elección del diseño de la muestra

Dadas las características del sector pecuario y pensando siempre en función de la recolección de informaciones para la planificación de la salud animal, es común aplicar la realidad de diseños estratificados y con selección en dos etapas. Esto calza generalmente bien con los objetivos que se plantea este tipo de estudio, en lo que concierne a un menor error comparado con otros diseños muestrales para igual tamaño de muestra, también un costo no tan alto y la posibilidad de tener estimaciones para cada estrato.

Si se piensa que el objetivo principal que debe cumplir un criterio de estratificación es el de hacer lo más homogéneas posibles las unidades de cada estrato, de modo de disminuir la magnitud de la varianza, surge inmediatamente la pregunta: ¿qué tipo de homogeneidad? Debe elegirse una variable estratégica o clave, para agrupar las unidades de la población según los valores que toman en cuanto a dicha variable.

En este sentido, el criterio geográfico en general "engloba" otros criterios y es de gran utilidad en estudios epidemiológicos.

Otro aspecto crítico en la planificación de una encuesta por muestreo es la determinación del tamaño de la muestra (n).

Demás está insistir sobre la trascendencia de esta fase. Prácticamente en este punto es donde se conjugan todos los elementos conceptuales vistos en la primera parte. De otro lado, es donde más cuesta respetar los principios teóricos, al confrontarse a una realidad que no siempre se compadece con los supuestos.

Revisando cualquier encuesta hecha en salud animal, se comprobará que se indaga sobre una cantidad grande de variables. Pues bien ¿la varianza de cuál de ellos se tomará? Habrán algunas variables cuya distribución es muy homogénea, mientras que habrán otras que presentan enorme heterogeneidad. En rigor, habrá que elegir sólo una varianza. Puede pensarse en elegir aquella que presente el más alto valor; es difícil que ello conduzca a resultados prácticos y normalmente en ese caso se calculará un tamaño de muestra excesivamente grande, sin posibilidades de financiamiento. Hay variables cuya varianza es tan grande que el cálculo del tamaño de muestra, con una precisión razonable, equivale en el hecho a casi "censar" la población. En todo caso, constituye la alternativa más defendible desde un punto de vista teórico.

En la práctica es un poco más realista empezar determinando el tamaño de muestra en función de los recursos, y luego calcular "estadísticamente" el tamaño de muestra en términos de la variable elegida como estratégica, es decir, la que más interese en la encuesta. De la comparación de los tamaños de muestra resultantes por esos dos métodos diferentes, nacerá el ajuste que será necesario hacer en áreas de la factibilidad de la encuesta. Habrán tres alternativas: se aumenta la cantidad de recursos, se disminuye razonablemente la precisión, o ambas cosas en alguna medida.

7. Análisis de la factibilidad del muestreo y decisión de la forma final de realizarlo

En esta etapa se debe contrastar el costo total de la encuesta por muestreo con la disponibilidad de recursos asignados para desarrollar el estudio.

El costo de un muestreo va a depender mucho de las características del área.

En encuestas por muestreo para estudiar problemas de sanidad animal hechas últimamente en países de América del Sur, el costo de cada encuesta en una vivienda rural ha sido de EUA\$ 3,50. En este valor se incluyen todos los costos, es decir, los costos fijos (aquellos gastos que se hacen independientemente del número de encuestados) y los costos variables (gastos proporcionales al número de encuestados). Esto en una región donde todas las unidades primarias estaban próximas unas de otras, con caminos buenos y con un bajo porcentaje de no respuestas.

Sin embargo, este valor de EUA\$ 3,50 puede cambiar bastante en regiones que presentan condiciones naturales difíciles y carencia de infraestructura lo que probablemente encarece el costo de la encuesta por muestreo.

El costo de una encuesta por muestreo podría esquematizarse de la siguiente manera:

$$C = CF + CV$$

donde C = costo total de la encuesta

CF = costos fijos

CV = costos variables

Los costos fijos son aquellos cuya magnitud no depende del diseño, ni de la cantidad de unidades de muestreo que la muestra incluya.

$$CF = C (AD) + C (PD) + C (ID)$$

C(AD) = costos de administración de la encuesta por muestreo, que incluye salarios, diarias, contrato de consultores, gastos de materiales para:

- a) elaboración del plan de muestreo
- b) organización de la encuesta
- c) reclutamiento y entrenamiento de encuestadores y supervisores
- d) elaboración de normas
- e) preparación de mapas y de rutas para la selección de unidades primarias (UP)
- f) preparación de otros materiales para ser usados en la asignación de trabajo a los encuestadores.

C(PD) = costos de procesamiento de los datos que incluyen los gastos de:

Para calcular el costo de estos coeficientes se aplica un criterio de proporcionalidad con respecto a los gastos directamente relacionados a la cantidad de unidades de muestreo elegidas.

Así, en el caso de coeficiente C_0 se consideran las siguientes informaciones:

- d = distancia media entre unidades primarias incluidas en la muestra
- tv = tiempo medio de viaje entre unidades primarias
- ct = costo de transporte por km recorrido
- ce = costo de cada hora/encuestador.

Entonces el gasto de transporte (T) es:

$$T = (d) (ct)$$

y el gasto de pago a personal (encuestador) por el tiempo de viaje entre unidades primarias (viviendas en este caso), le llamamos E, es:

$$E = (tv)(ce)$$

Se puede aumentar al valor E, por ejemplo: el 6% por "viaje entre la casa o la oficina y las unidades primarias" incluidas en la muestra; el 10% por tiempo gasto en entrenamiento, y 15% por "viaje de supervisión". Todos estos porcentajes formarían un nuevo factor que de acuerdo con el ejemplo sería 1,34, de esta manera:

$$C_0 = T + E(1,34)$$

Según Hansen, Hurwitz & Madow, para el cálculo del coeficiente C_1 se considera un agregado de los siguientes aspectos:

- a) una fracción de los gastos de administración que es dependiente de m

$$(\alpha) C(AD)$$

- b) una fracción de los gastos de impresión de documentos que es dependiente de m

$$(\beta) C(ID)$$

- c) una fracción de los gastos de transporte y viaje del encuestador entre las unidades primarias

$$(\gamma) (T+E)$$

- d) el costo de listar las unidades elementales de las UP seleccionadas y la asignación de la submuestra

$$(ce) (\bar{N}_i) (\delta)$$

donde: \bar{N}_i = número medio de unidades elementales por unidad primaria

δ = factor que representa la proporción del C_e gastado en listar \bar{N}_i de una unidad primaria, incrementado por los porcentajes de tiempo, ya aplicados para E.

Para el cálculo del coeficiente C_2 se considera un agregado de los siguientes aspectos:

- a) fracción del ce que equivale al costo del tiempo de entrevista (te)

$$(te) (ce)$$

- b) por la supervisión de los datos de los cuestionarios se considera una pequeña fracción de los gastos de procesamiento de datos

$$(\epsilon) C(PD)$$

- c) una fracción de los gastos de impresión de documentos dependiente de la cantidad de encuestas

$$(\eta) C(ID)$$

El análisis de factibilidad puede revelar que los recursos son insuficientes frente al tamaño de muestra requerida. En este caso se puede combinar el plan de muestreo, reducir n aceptando un error más grande, buscar los recursos que faltan o suspender el estudio hasta contar con los recursos. El análisis precedente permitirá orientarnos en la decisión a tomar: si se lleva a cabo la investigación o se la abandona. En algunas oportunidades los recursos son claramente incompatibles con el costo y el hacerlo de otra forma podría no conducir a algo de significación, con el consiguiente desperdicio de recursos. Sin embargo, no se puede ser muy perfeccionista de esperar que para cada estudio se le proporcionen todos los recursos previstos. En ocasiones se podrán aprovechar recursos que otras entidades pueden proporcionar y que incluso

tienen una capacitación que hace más fácil su adiestramiento. Esto es común que ocurra en nuestros países.

ORGANIZACION Y EJECUCION

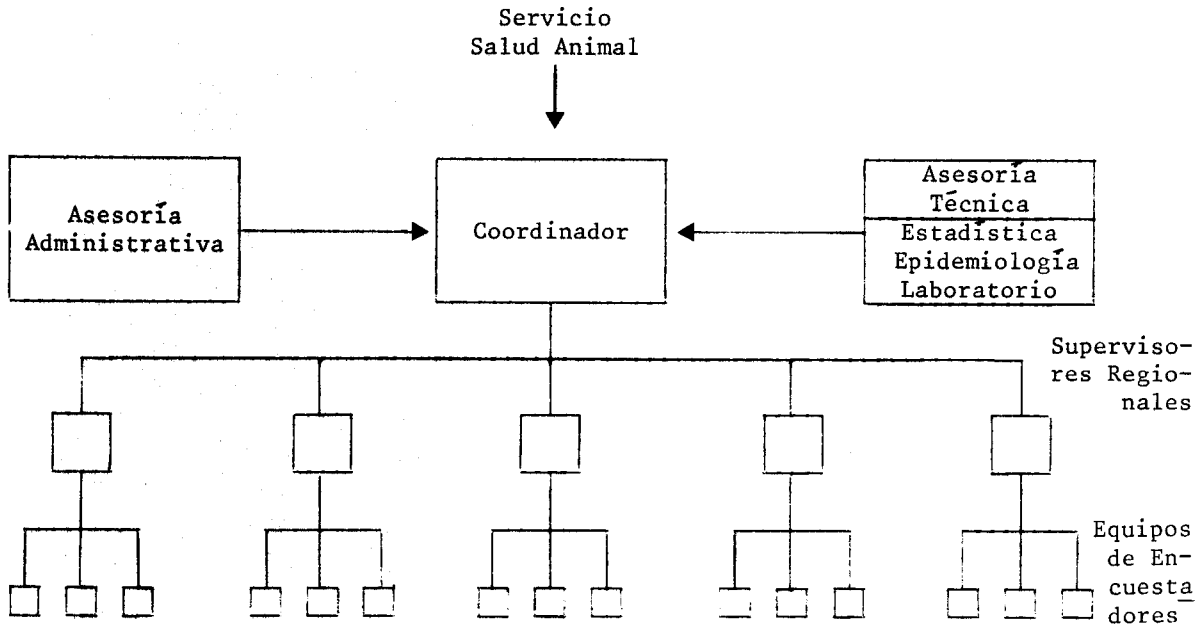
1. Organización del trabajo de terreno

La programación de la recolección de informaciones debe planearse en términos de las peculiaridades geográficas, concentración ganadera, etc. Debiera haber un contacto permanente entre la oficina de diseño de la muestra y el equipo de encuestadores. Siempre se presentarán tipos de respuesta o características de la salud animal que no se han contemplado en el proceso de adiestramiento. En tales casos es preferible solucionar el problema consultando al coordinador y a sus asesores técnicos en la encuesta. En encuestas de salud animal a nivel de grandes áreas geográficas es conveniente que en las regiones donde no hay oficinas veterinarias a cargo de los que dirigen la encuesta, la enumeración sea realizada por personas calificadas para este tipo de trabajo. Por lo general, la recolección de datos en las regiones es realizada por personal de los servicios de salud animal u organismos similares, dejando la encuesta en lugares desprovistos de oficinas oficiales a personal temporal especialmente contratado para este efecto.

Es indispensable que se implante cierta supervisión en forma simultánea a la recolección de informaciones. De esta manera se previene que se cometan errores voluntarios o involuntarios. Verificar las encuestas que realiza cada encuestador puede ser un medio para disminuir un posible sesgo. Es importante que se realice esta forma de control, simultáneamente al proceso de extracción de información, para que puedan tomarse medidas oportunas de corrección. Los controles "a posteriori" son en general tardíos y no se alcanza a remediar el problema, si no a expensas de gastos mucho mayores y demoras perjudiciales.

En lo normativo es imperioso mantener un mecanismo centralizado bajo la responsabilidad del organismo específico de salud animal, en razón de ser el organismo de mayor competencia para abordar y resolver los problemas que se susciten en torno a la encuesta. En lo ejecutivo resulta conveniente descentralizar las operaciones, nominando un coordinador como responsable directo

del desarrollo de la encuesta y supervisores regionales.



Los supervisores de la encuesta deben mantener escritorios regionales donde los encuestadores se puedan dirigir pidiendo instrucciones sobre los problemas encontrados durante el desarrollo del trabajo de campo. Al completar las visitas de cada día los equipos de encuestadores deben entregar los cuestionarios llenados al supervisor, quien debe revisarlos en su contenido, en su legibilidad, número de acuerdo con el itinerario diario y anotaciones de importancia administrativa para la encuesta (hora, kilometraje). Cuando hay cuestionarios con deficiencias, se apartan éstos para requerir aclaraciones o informaciones complementarias, procediendo a programar nuevas visitas a las unidades encuestadas cuando sea necesario. Los equipos encuestadores sólo pueden dar por finalizada su tarea diaria cuando concluye la verificación de los cuestionarios y de las cédulas de encuestador respectivos.

2. Elaboración de un cuestionario, cédula de encuestador e instructivos

El cuestionario es un documento impreso, confeccionado para obtener información consonante con un objetivo bien definido. Debe contener las preguntas necesarias y los espacios correspondientes y suficientes para ser llenados con las respuestas.

Tanto para determinar el contenido y la forma de un cuestionario es indispensable tener claros los objetivos del cuestionario, que son alcanzar los objetivos del estudio con preguntas específicas y motivar al entrevistado a proporcionar la información requerida de la forma más rigurosa posible.

Se debe tener en cuenta que cuestionarios muy extensos no se llenan muchas veces con exactitud ni agrado, influyendo ésto en la calidad de la información proporcionada; deben evitarse las preguntas que no sean estrictamente necesarias.

Los términos elegidos deberán estar dentro del vocabulario del entrevistado, evitándose el uso de modismos y términos muy técnicos.

Las preguntas deben tener en cuenta el nivel de conocimiento del entrevistado. Por ejemplo, al hacer la pregunta: ¿Cómo están los animales? se puede responder bien, mal o regular, pero no se sabe si se refiere al estado sanitario, al manejo, a la producción, etc.

Las preguntas deben formularse de modo que no sugieran la respuesta. Por ejemplo, una pregunta que pretende conocer la actitud acerca de la eliminación de vacas con brucelosis, podría formularse como: ¿qué piensa sobre la eliminación de vacas con brucelosis? La misma pregunta en forma tendenciosa podría formularse: ¿no dirá usted que está en contra de la eliminación de vacas con brucelosis, no es así?

Las preguntas deben limitarse a una sola idea o referencia. Preguntas como: ¿qué opina usted de los programas de control de la fiebre aftosa y brucelosis? pueden dar una respuesta que no permita darse cuenta a cual idea se refiere o si es a ambas.

Las preguntas deben ordenarse de tal modo que aparezcan con sentido ante el entrevistado. La secuencia debe determinarse también por lo que se denomina un "acercamiento tipo embudo". Esto se refiere a la formulación de las preguntas más generales primero, seguidas por aquellas más específicas.

Según la forma de la respuesta, se pueden considerar dos tipos de preguntas: abiertas y cerradas. En las preguntas abiertas el entrevistado contesta en sus propios términos y da la respuesta a su gusto. A veces se utilizan preguntas abiertas en encuestas piloto, para conocer posibles respuestas a ellas y luego poder confeccionar preguntas de tipo cerrado en otras encuestas.

Los instructivos acerca del cuestionario están destinados fundamentalmente a informar el proceso de colecta de datos. Estos instructivos deben contener una definición explícita de los términos utilizados en el cuestionario para evitar diversas interpretaciones por los encuestadores. Otro aspecto a abordar en los instructivos es el cómo registrar la información recibida.

Otro documento que debe ser elaborado es la CEDULA DEL ENCUESTADOR que debe llevar cada equipo. En ella se indica el número del equipo, la secuencia diaria de visitas, el sector a encuestar. También deben indicarse las instrucciones acerca de anotación de hora de llegada, duración de la entrevista, hora de salida, kilometraje recorrido entre establecimientos visitados en forma consecutiva y otras informaciones complementarias a éstas. Además, se debe dejar claro en el instructivo el qué hacer cuando se presenta un caso de no respuesta. Debe indicarse la conveniencia de substituir un establecimiento ganadero por otro. Este procedimiento no resuelve el problema, lo único que se logra es aumentar el tamaño de la muestra en aquella parte de la población la cual es fácilmente accesible a los procedimientos de campo de salud animal.

3. Entrenamiento de los encuestadores

En una encuesta por muestreo que se utilice al personal del servicio de salud animal para encuestar, el tiempo dedicado al entrenamiento generalmente es limitado. Cuando se trata de cuestionarios simples unas pocas horas destinadas a la instrucción del personal llega a ser suficiente. Cuando los encuestadores no conocen el problema o son contratados específicamente para realizar la encuesta hay que programar un número mayor de horas para entrenamiento.

En lo posible el encuestador debe ser una persona que cuando salga al terreno esté consciente de los objetivos que ésta persigue. En aquellas preguntas que de alguna manera implican magnitudes o en aquellas acciones que tiendan a producir una observación confiable, el encuestador deberá tener una participación activa, verificadora y crítica. En algún tipo de encuestas el

encuestador debe tener algún conocimiento técnico, de otro modo se corre el riesgo de obtener informaciones deficientes que pueden deteriorar el estudio.

4. Elaboración de mapas y asignación de itinerarios a los equipos de encuestadores

Se prepara un mapa de campo para cada equipo. Se divide el campo en sectores, marcándose en ellos las unidades de muestreo a encuestar. El itinerario se marca con una línea desde una unidad a otra, indicando las rutas a seguir y los puntos (accidentes geográficos, cruce de carreteras o de ferrocarril, alguna construcción destacada) de referencia. Este itinerario debe de alguna manera establecer el número de unidades de muestreo a encuestar cada día.

Para llevar un control del desarrollo de los itinerarios el coordinador y los supervisores regionales deben tener mapas maestros con la cobertura total y regional respectivamente.

5. Divulgación a la comunidad de los objetivos de la encuesta

En las encuestas de salud animal es de una significación muy grande conseguir una actitud positiva de la comunidad imprescindible para garantizar fidelidad en la masa de informaciones a coleccionar. Es frecuente que se advierta a la comunidad la trascendencia de la encuesta, y una manera de hacer más efectiva esa colaboración es lograr el apoyo de asociaciones de ganaderos y organismos de probado prestigio, Universidades, Cooperativas, etc. De una u otra manera existe siempre un margen considerable de no respuesta. Sobre el particular hay que destacar que si la no respuesta tiene su origen en una desaparición de la unidad que se pretende encuestar, no debe ser motivo de preocupación ya que puede tomarse como una muestra representativa de una parte del universo que desaparece. El problema es muchísimo más serio cuando existiendo la unidad, hay negativa de dar respuesta. Basar las estimaciones solamente en poblaciones de respondientes puede ocasionar sesgos de alguna magnitud. En esos casos las estimaciones hechas, en realidad, podrán ser generalizadas sólo a la población de respondientes y no a la población total. Para tener informaciones sobre la población de no respondientes habrá que disponer de otros procedimientos.

Precisamente una manera de prevenir la no-respuesta es divulgar en forma oportuna y accesible a la comunidad los objetivos de la encuesta.

6. Prueba piloto de la encuesta

Es conveniente que después de entrenado el personal de encuestadores, de elaborados los documentos necesarios y de establecida la organización de campo de la encuesta, sean sometidos todos estos aspectos a prueba en un pequeño ensayo en terreno. Este procedimiento permitirá descubrir fallas de organización, omisiones, deficiencias técnicas y administrativas que de no ser corregidas oportunamente podrían llegar a comprometer los objetivos de la encuesta. En este ensayo piloto debería cuidarse de que todo el mecanismo de la encuesta fuese verificado. En el caso de descubrir deficiencias o algún aspecto no advertido será necesario revisar el plan de la encuesta y adaptar las modificaciones necesarias.

7. Ejecución de la encuesta

- a) Dirigirse en vehículo a la primera unidad a encuestar anotando la hora y el kilometraje.
- b) Una vez llegado al establecimiento ganadero anotar la hora y el kilometraje nuevamente.
- c) Después de identificada la unidad a visitar, el encuestador llamará por el encargado, propietario o administrador del establecimiento, presentándose personalmente en seguida, indicando que representa el Servicio de Salud Animal y comunicando clara y exactamente los objetivos de la encuesta.
- d) Si la actitud del encargado fuese favorable proceder a desarrollar la encuesta.
- e) Si la actitud es desfavorable, marcar en la cédula del encuestador para discutir con el supervisor la manera de abordar esta situación. En caso de no haber nadie en el establecimiento que pueda proporcionar informaciones confiables, marcar una fecha para retornar y realizar la encuesta.
- f) Al término de la encuesta marcar la hora de conclusión de la colecta de información.

8. Procesamiento de datos y publicación de resultados

Toda la masa de informaciones extraídas deberá ser codificada, tabulada, clasificada, interpretada y verificada. En esta altura es donde se deberá diseñar el nivel de desagregación y tipos de clasificaciones necesarias. Por una parte, las informaciones a nivel íntimo de desagregación y la unidad de muestreo no permiten obtener conclusiones; por otra parte, estimaciones globales sólo dan una idea muy general de la situación de la salud animal en sus diversos aspectos. Es imprescindible pues combinar estimadores a distintos niveles de desagregación. Conocer por ejemplo la tasa de prevalencia por brucelosis es un dato de mucha utilidad, pero además es necesario conocer los factores responsables por su conducta.

Con la utilización de técnicas informáticas, la sistematización de datos se hace mucho más amplia y rápida. Previamente a la introducción de los datos en los computadores, aquéllos debieron ser sometidos a un estricto control, máxime si fueron extraídos por personal que no tenía un cabal conocimiento técnico de la salud animal.

En la publicación de los resultados obtenidos mediante muestras, es fundamental indicar las principales características del diseño utilizado: criterio de estratificación, tamaño de muestra, tipo de asignación, precisión para cada una de las variables principales, formas de selección, etc. De esta manera los usuarios conocerán las bondades y limitaciones de que son objeto las estimaciones resultantes. Incluso es necesario presentar un anexo con una descripción detallada de toda la encuesta, pormenorizando los principales problemas y las formas de solución.