TAMAÑO DE LA MUESTRA

Enric Mateu [1], Jordi Casal

CReSA. Centre de Recerca en Sanitat Animal / Dep. Sanitat i Anatomia Animals, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona

RESUMEN

En este artículo se describen los aspectos que hay que tener en cuenta para determinar el tamaño de muestra necesario para obtener información de la población. Se presentan las fórmulas para calcular el tamaño de muestra necesario para determinar la prevalencia o incidencia de una enfermedad en una población, para determinar si una enfermedad está presente o no en una población y para realizar estudios epidemiológicos.

INTRODUCCIÓN

A continuación intentaremos dar respuesta a la segunda pregunta importante que se plantea cuando se va a realizar un muestreo: ¿Cuántos animales debo tomar? La respuesta depende en primer lugar del objetivo que se pretende conseguir con el muestreo. Los objetivos más frecuentes que nos podemos plantear son:

- Conocer la prevalencia o incidencia de una enfermedad en una población
- Determinar si una enfermedad está presente o no en una población
- Realizar un estudio epidemiológico

MUESTREO PARA DETERMINAR PREVALENCIAS

Cuando se pretende realizar una encuesta epidemiológica para determinar la cantidad de enfermedad presente en una población, el tamaño de la muestra dependerá de cuatro valores:

- La frecuencia esperada de enfermedad. Basar el tamaño de la muestra precisamente en el valor que se quiere obtener con la encuesta puede parecer de entrada un contrasentido. Sin embargo, si planteamos una encuesta desde el punto de vista del método científico, es decir, si planteamos una hipótesis en relación a la cantidad de enfermedad que se espera encontrar, para – mediante el trabajo posterior- comprobar o rechazar la hipótesis, este aparente contrasentido ya no tiene lugar. Por tanto, cuando se quiere conocer la prevalencia de una enfermedad no podemos partir de "a ver que sale" sino que debemos partir de "mi hipótesis es que hay un n% de enfermedad, voy a comprobarlo"
- El tamaño de la población. El tamaño de la población va a afectar el tamaño de la

[1] *Enric.mateu@uab.es*

muestra, especialmente si la población no es excesivamente grande

- La precisión exigida. La cantidad de enfermedad que se obtendrá mediante la encuesta debe extrapolarse posteriormente a la población general de la que se ha obtenido la muestra. Esta extrapolación conlleva un cierto error o falta de precisión, es decir la muestra nos va a indicar “más o menos” la enfermedad presente en la población. La precisión es la cuantificación de este “más o menos” con el que podremos conocer la cantidad de enfermedad en la población.
- El nivel de confianza. Cuando se extrapolan unos datos y se establece una precisión, existe la posibilidad de que la cantidad de enfermedad en la población general no esté comprendida en el intervalo indicado, la probabilidad de que el valor de la variable esté comprendido dentro de dicho intervalo es el nivel de confianza, que normalmente se establece en el 95%.

Para estimar el tamaño de muestra necesario para realizar una encuesta epidemiológica se debe de aplicar la siguiente fórmula:

$$n = \frac{z^2 pq}{B^2}$$

Donde n= Tamaño de la muestra,
z= 1,96 para el 95% de confianza, 2,56 para el 99%
p= Frecuencia esperada del factor a estudiar
q= 1- p
B= Precisión o error admitido

El valor de n obtenido por esta fórmula indica el tamaño de la muestra para una población infinita, a efectos prácticos se considera población infinita cuando la muestra supone menos del 5% de la población total.

Cuando la población es pequeña, la muestra obtenida mediante esta última fórmula es demasiado grande, en estos casos se debe aplicar la siguiente fórmula correctora:

$$\frac{1}{n'} = \frac{1}{n} + \frac{1}{N}$$

Donde n'= Tamaño de la muestra necesario
n= Tamaño de la muestra según la primera de las fórmulas
N= Tamaño de la población

***Ejemplo**: Supongamos que se desea realizar una encuesta sobre la brucelosis ovina. Se estima una prevalencia del 15% y se requiere un 5% de precisión sobre una población de 2.000.000 de cabezas. El nivel de confianza se fija en el 95%.*

El tamaño de la muestra necesario para dicha encuesta según la fórmula sería:

$n = 1,96^2 \times 0,15 \times 0,85 / 0,05^2) = 196$

Por tanto, deberemos seleccionar aleatoriamente 196 animales del total de la población. Ello permitirá, en el caso que la prevalencia sea realmente del 15%, poder afirmar que en el 95% de los casos, la prevalencia de la población general oscila entre el 10% y el 20% (15%±5%)

Con las mismas premisas anteriores calcular el tamaño de la muestra si se aplicase en un rebaño de 500 cabezas.

Aplicando la corrección al resultado del ejemplo anterior:

1 / n' = 1 / 195 + 1 / 500 de donde n' = 140

Cuando la encuesta se realiza para determinar una media de una variable cuantitativa (por ejemplo el número de partos por año), es necesario considerar una estimación de la desviación estándar o la varianza de dicha variable y la máxima diferencia que admitiríamos con relación a la media real de la población. En este caso, la fórmula a aplicar será:

$$n = \frac{z^2 s^2}{B^2}$$

Donde n= Tamaño de la muestra
S= Desviación estándar
B= Precisión

En la tabla 1 se presentan los tamaños de muestra para una población infinita y distintos niveles de prevalencia y de precisión y con un nivel de confianza del 95%. Para prevalencias superiores al 50% se debe utilizar el valor correspondiente a 1-p. Por ejemplo para calcular el tamaño de muestra necesario para una prevalencia esperada del 70% con una precisión del 3% y un nivel de confianza del 95%, el tamaño de muestra necesario será 1.291 (correspondiente a una prevalencia del 30%)

Prevalencia	Precisión o error						
esperada	25%	20%	10%	5%	3%	1%	0,5%
5%	3	5	19	73	292	1.825	7.300
10%	6	9	35	139	554	3.458	13.830
15%	8	13	49	196	784	4.899	19.593
20%	10	16	62	246	984	6.147	24.587
25%	12	19	73	289	1.153	7.203	28.812
30%	13	21	81	323	1.291	8.068	32.270
35%	14	22	88	350	1.399	8.740	34.959
40%	15	24	93	369	1.476	9.220	36.880
45%	16	24	96	381	1.522	9.508	38.032
50%	16	25	97	385	1.537	9.604	38.416

Tabla 1. Tamaño de muestra necesario para determinar la prevalencia en una población grande y con un nivel de confianza del 95%

A partir de la tabla se puede observar que el tamaño de la muestra aumenta de manera muy importante al aumentar la precisión. Para una precisión diez veces superior (por ejemplo pasar de una precisión del 10% al 1%) el tamaño de muestra necesario aumenta 100 veces (ello es debido a que en la fórmula, el tamaño de la muestra está elevado al cuadrado).

Intuitivamente parece que cuando se aumenta la prevalencia el tamaño de la muestra debería ser inferior, es decir, parecería que el tamaño de muestra para una prevalencia esperada del 5% debería ser superior al tamaño que necesitaríamos para una prevalencia del 50%. Sin embargo, si consultamos la tabla 1, vemos que ocurre lo contrario: en caso de una precisión del 5% necesitamos respectivamente 73 y 385 individuos. Ello es debido a que, cuando trabajamos con prevalencias bajas, generalmente deberemos aumentar la precisión: la información que aportará un muestreo que nos permite decir que en la población general habrá el 5% ± 5% (o sea, entre el 0% y el 10%) no es lo mismo que la que aportará decir que en la población hay un 50% ± 5% (o sea entre el 45% y el 55%). Por tanto, en realidad cuando la prevalencia esperada es baja deberemos aumentar la precisión (en el ejemplo anterior deberemos pasar de un 5% a un 2% o un 3%) con lo que el tamaño de la muestra que necesitaremos en realidad será mayor para una prevalencia esperada pequeña.

MUESTREO PARA LA DETECCIÓN DE ENFERMEDAD

En otras ocasiones, la encuesta no pretende estimar una prevalencia sino que su finalidad es saber si la enfermedad existe o no en una población (independientemente de si hay mucha o poca). En otros términos, se desea conocer el tamaño de muestra necesario para, con un nivel de confianza determinado, afirmar que, si ninguno de los animales estudiados resulta positivo, la población está libre de enfermedad. La aplicación de esta fórmula presupone que en caso de estar presente una enfermedad en una población ésta presentará una prevalencia mínima (como realmente ocurre en la mayoría de enfermedades contagiosas).

Para realizar este cálculo se tiene que aplicar la siguiente fórmula con la que obtendremos el tamaño de muestra adecuado para asegurar que si todos los individuos resultan negativos, la enfermedad estará a un nivel inferior a nuestra estimación (y por tanto según la hipótesis de une prevalencia mínima, consideraremos que la población está libre).

$$n = \left[1-(1-a)^{1/D}\right] \times \left[n - \frac{(D-1)}{2}\right]$$

Donde, n= Tamaño de la muestra
a= Nivel de confianza
D= Número de animales enfermos en la población
N= Tamaño de la población

A partir de la fórmula anterior, despejando D, puede calcularse también la prevalencia máxima esperable en una población en la que se ha examinado un número concreto de animales y todos han resultado negativos.

La tabla 2 indica el número de muestras que debemos tomar para detecta enfermedad en una población, por ejemplo, si creemos que en una población de 200 individuos hay el 20% de animales afectados –o sea 40 individuos- deberemos tomar 14, si alguno de ellos está afectado, la enfermedad existe, si todos son negativos podemos decir que con un 95% de confianza la enfermedad no está presente.

	Prevalencia esperada en caso de estar presente la infección								
Población	90%	80%	70%	60%	50%	40%	30%	20%	10%
5	2	2	3	3	3	3	5	5	-
10	2	2	3	3	4	4	6	8	10
12	2	2	3	4	4	5	7	9	11
15	2	2	3	4	5	5	7	9	13
20	2	2	3	4	5	6	7	10	19
30	2	2	3	4	5	6	8	11	19
50	2	2	3	4	5	6	8	12	55
80	2	2	3	4	5	6	9	13	24
100	2	2	3	4	5	6	9	13	25
200	2	2	3	4	5	6	9	14	27
500	2	2	3	4	5	6	9	14	28
1000	2	2	3	4	5	6	9	14	29

Tabla 2. Tamaño de muestras necesario para detecta enfermedad en una población en función de la prevalencia esperada (nivel de confianza del 95%)

La tabla 2 también se puede interpretar en el sentido de determinar el máximo número de afectados que habrá: si todos los resultados han sido negativos podemos decir que la máxima prevalencia posible –con un 95% de nivel de confianza- será del 20%.

Para una explotación podemos asumir que una enfermedad infecciosa tendrá una prevalencia del 10% o del 40%, pero para un país o un territorio más amplio, posiblemente la tasa de prevalencia será menor, en estos casos la muestra deberá tener un tamaño superior, tal como se ve en la tabla 3:

Prevalencia esperada	Nivel de confianza (riesgo)	
(si existe enfermedad)	1%	5%
0,001%	460.515	299.572
0,01%	46.050	29.956
0,05%	9.209	5.990
0,1%	4.603	2.995
0,2%	2.301	1.497
0,4%	1.149	748
0,5%	919	598
1%	459	299
2%	228	149
5%	90	59
10%	44	29
20%	21	15
50%	7	5

Tabla 3. Tamaño de muestras necesario para detecta enfermedad en una población infinita en función de la prevalencia esperada (niveles de confianza del 99% y del 95%)

Ejemplo: *Qué tamaño de muestra será necesario para determinar que en un rebaño de 150 vacas la prevalencia de tuberculosis es igual o inferior al 10%?*

$$n= (1-(1-0,95)^{0,067}) \times (150-(15-1)/2) = 26$$

Ejemplo: Se han examinado 40 animales de un rebaño de 800 ovejas. ¿Cuál es la máxima prevalencia posible de brucelosis en dicho rebaño si todos los animales examinados han sido negativos?

$$D= (1-(1-0,95)^{1/40}) \times (800-(40-1)/2) = 56 \text{ animales } (7\%)$$

TAMAÑO DE MUESTRA PARA LA REALIZACIÓN DE ESTUDIOS

En el caso de los estudios el tamaño de la muestra necesario dependerá del tipo de estudio, del nivel de confianza, de la potencia muestral, y de los valores de riesgo relativo u *odds ratio* mínimos que se deseen detectar. El número de individuos a muestrear se puede calcular con la siguiente fórmula:

$$n = \frac{\left[z_\alpha (2pq)^{1/2} - z_\beta (p_e q_e + p_c q_c)^{1/2} \right]^2}{(p_e - q_c)^2}$$

Donde, n= Tamaño de la muestra
Z_α= 1,96 para el 95% de confianza, 2,56 para el 99%
Z_β= -0,84 para un error β del 20%
P_e= Frecuencia de la respuesta en los expuestos (o casos)
P_c= Frecuencia de casos respuesta en los no-expuestos (o controles)
$P= (P_e + P_c)/2$ $Q= 1-P$

Z_α y Z_β son dos estadísticos asociados al error α (o error Tipo 1) y al error β (o error Tipo 2). El error alfa corresponde a uno menos el nivel de confianza y consiste en aceptar que los grupos son diferentes (rechazar de la hipótesis nula) cuando en realidad los dos grupos son iguales. En caso de un estudio para valorar la eficacia de un fármaco, sería considerar que éste es eficaz cuando realmente no lo es.

El error beta es uno menos la potencia o poder y consiste en la probabilidad de considerar que los dos grupos son iguales (se acepta la hipótesis nula) cuando en realidad son diferentes. En el ejemplo anterior es la probabilidad de que, existiendo diferencias entre los grupos, el estudio no sea capaz de encontrarlas.

***Ejemplo**: Se desea comparar dos tratamientos A y B. Al tratamiento A se le supone una eficacia del 95% y al B del 75%. Calcular el tamaño de la muestra necesario para este estudio:*

$$n = (1{,}96^* (2^*0{,}85^*0{,}15)^{1/2} +0{,}84^*(0{,}95^*0{,}05+ 0{,}75^*0{,}25)^{1/2}) \ / (0{,}95 -0{,}75)^2 = 49$$

Se deberán tomar 49 individuos en cada grupo. Si el tratamiento A es realmente un 20% más eficaz que el B, el estudio permitirá determinar esta diferencia en el 80% de los casos (1 - error β) y si no existen diferencias, existe una probabilidad del 95% de que éstas no se encuentren en el estudio (1 - error α)

En algunos estudios la variable a comparar en los dos grupos es cuantitativa, y con el estudio se pretende comparar las medias en los dos grupos, en este caso, la fórmula a aplicar es:

$$n = 2\left[\frac{(z_{\alpha} - z_{\beta}) \times s}{x_e - x_c}\right]^2$$

Dónde, S= Desviación estándar
X_e= Media del valor en los expuestos
X_c= Media del valor en los no-expuestos

Ejemplo*: Se desea comparar dos tratamientos destinados a disminuir los niveles de colesterol en sangre. Para el tratamiento A se espera que los valores medios de colesterol sean de 140 mg/l y para el tratamiento B de 150 mg/l con una desviación estándar de 10.*

$n = 2*(1,96+ 0,84) * 10^2 / (150 -140) = 56$

Se deberán tomar 56 individuos para cada tratamiento. Si con el tratamiento A se obtienen unos niveles de colesterol inferiores en 10 mg/l (con una desviación estándar de 10), el estudio permitirá determinar diferencias en el 80% de los casos (1 - error β) y si los dos tratamientos tienen el mismo efecto, existe una probabilidad del 95% de que el estudio encuentre diferencias (1 - error α).