

Algunas consideraciones sobre la influencia del tamaño muestra) al estimar el impacto de las enfermedades en cunicultura

Ignacio de Blas*, Olivia Gironés, Imano) Ruiz, José Luis Alonso y José Luis Muzquiz

Unidad de Patología Infecciosa y Epidemiología. Facultad de Veterinaria. Universidad de Zaragoza. c/ Miguel Servet 177. 50013 Zaragoza (España)

* Autor de contacto: [e-mail: deblas@posta.unizar.es](mailto:deblas@posta.unizar.es)

Resumen

El objeto del presente trabajo es poner de manifiesto la importancia clave de los métodos de muestreo en el diseño de los experimentos orientados a dos objetivos básicos: la determinación de la prevalencia de una enfermedad y la detección de la misma.

La representatividad de una muestra viene determinada por dos condiciones básicas: aleatoriedad (todos los animales de la población tienen que tener la misma probabilidad de formar parte de la muestra) y homogeneidad (debe conservar las proporciones que se guardan en la población en lo referente a sus características esenciales). Para ello es importante utilizar un correcto método de muestreo y seleccionar un tamaño de muestra adecuado.

Por eso consideramos la necesidad de conocer la existencia de posibles errores de muestreo, la utilización incorrecta de los datos obtenidos para fines distintos para los que inicialmente se recogieron, el cálculo del error aceptado, la realización de muestreos estratificados y el ajuste del tamaño muestra)

Abstract

The objective of this contribution is to describe the capital importance of sampling methods when we design an experiment oriented to two basic targets: determination of disease prevalence and detection of disease.

Representativity of a sample is determined by two basic conditions: randomize (each animal has same chance of taking part of the sample) and homogeneity (proportions of intrinsic characteristics of population must be similar in the sample). For this reason it is important to use a correct sampling method and to select an adequate sample size.

So we consider than it is necessary to know the existence of possible sampling errors, incorrect use of data collected, calculation of accepted error, making stratified samples and to adjust sample size.

Introducción

En las últimas décadas empresas de diversos sectores han alcanzado producciones de alta calidad y fiabilidad a costes razonables. Esta circunstancia indica el papel fundamental que ha jugado la aplicación de técnicas estadísticas de diseño experimental en la actividad de dichas empresas, y como su adecuada utilización es decisiva para conseguir controlar y mejorar la calidad y fiabilidad de los productos fabricados en serie, lo que puede extrapolarse a ciertas producciones animales muy tecnificadas y intensificadas como la producción de leche, huevos y carne, donde dentro de este último tipo de producciones se situaría de forma destacada la cunicultura.

Hay que admitir que, por regla general, en las empresas actuales, se acepta que ningún proceso de producción puede conseguir que todos los productos fabricados sean idénticos, pues se reconoce la existencia de una variabilidad aleatoria que no puede ser eliminada por completo; esta circunstancia es más clara y evidente en nuestro caso ya que en los productos de origen biológico en los que el grado de variabilidad es mayor, debido a la existencia de factores más numerosos y complejos.

Por esta razón se hace evidente la necesidad de utilizar los métodos estadísticos apropiados para detectar precozmente las alteraciones y desviaciones de los procesos productivos de los valores normalizados, con especial incidencia en el caso de los procesos patológicos, por su fuerte impacto negativo sobre dichos procesos.

Así pues para lograr una adecuada gestión sanitaria de una explotación cunícola se hace preciso debemos conocer la prevalencia de las distintas enfermedades y patógenos potenciales, así como ser capaces de detectar una determinada enfermedad (o patógeno). En el primer caso el conocimiento del número de animales enfermos nos permitirá evaluar el coste económico de una enfermedad dentro de una población y la viabilidad de la aplicación de determinados tratamientos y estrategias para su control, mientras que las técnicas de detección de la enfermedad nos facilitarán el diseño de un adecuado plan sanitario y de bioseguridad (evitando la entrada de lotes con animales portadores o enfermos).

Es evidente que la inspección de todos los animales es inviable, fundamentalmente por razones económicas, y por eso se hace preciso recurrir a la inspección muestra) periódica, convenientemente diseñada desde el punto de vista estadístico, tanto para reducir los riesgos sanitarios en nuestras explotaciones como para poder conseguir el control del nivel de calidad y fiabilidad del proceso de producción.

Objetivos

El objeto del presente trabajo es poner de manifiesto la importancia clave de los métodos de muestreo en el diseño de los experimentos orientados a dos objetivos básicos: la determinación de la prevalencia de una enfermedad y la detección de la misma, ya que paradójicamente se le otorga un mayor peso específico al estudio estadístico de los datos obtenidos y su ulterior análisis económico, que al establecimiento las medidas necesarias que garanticen una adecuada calidad de los mismos.

Cuando intentamos profundizar en el conocimiento de una enfermedad en una población, podemos trabajar sobre el total de la misma, lo que encarecería y dificultaría el estudio, o sobre una parte de la misma que denominamos muestra, lo que simplifica y abarata el proceso, al mismo tiempo reduce su duración. Sin embargo para garantizar los resultados obtenidos la muestra debe cumplir una condición indispensable: debe ser representativa de la población para que los resultados obtenidos se puedan extrapolar.

Sin embargo la investigación de la enfermedad a partir de una muestra tiene la desventaja de

perder parte de la información común a la población y que por tanto el valor obtenido no sea el real, sino una estimación de ese valor real de la población. Esto quiere decir que a partir de una muestra, no pueden establecerse con absoluta certeza el grado de enfermedad de la población total, debido a la variabilidad que hay de una a otra muestra. Y por tanto es fundamental conocer la variabilidad de las muestras, que depende no solamente de la población que se muestrea, sino también del método de muestreo, y por ello se debe establecer en primer lugar un nivel de confianza (de tal forma que un valor de 0.95 significa que el resultado obtenido será correcto *en promedio* un 95 por ciento de las veces) a partir del cual se puede calcular un intervalo de confianza para la medición realizada directamente asociado con la precisión del parámetro calculado. Téngase en cuenta que cualquier aumento en el nivel de confianza supondrá que el intervalo de confianza sea más amplio, de tal forma que si se aumenta la confianza disminuye la precisión y viceversa.

Metodología

Por tanto se trata de que una muestra tomada de una población sea representativa lo que requiere, por un lado, que posea un *tamaño adecuado* y por otro, que el *método de selección* de la misma sea el apropiado, de manera que todas las subpoblaciones existentes en la misma estén representadas de forma correcta. Además existen otra serie de condicionantes que hay que considerar como son el coste de la obtención de datos, el tiempo necesario y los riesgos intrínsecos que implican la recogida de los datos tanto para el observador como para el observado, cuya valoración debe hacerse *a priori* para determinar la conveniencia y viabilidad de la realización del estudio.

Por esa razón es preciso que la representatividad de la muestra esté marcada por dos condiciones básicas: la muestra debe ser aleatoria (todos los animales de la población tienen que tener la misma probabilidad de formar parte de la muestra) y homogénea (debe conservar las proporciones que se guardan en la población en lo referente a las características básicas: sexo, edad, raza,...)

Para satisfacer la primera condición (**aleatoriedad**) podemos recurrir a distintos métodos de selección, que se pueden clasificar en dos grandes grupos: probabilísticos y no probabilísticos.

En los **Probabilísticos o aleatorios** todas las unidades de la población tienen la misma probabilidad de formar parte de la muestra, siendo el azar el que determina los que forman parte de la muestra, para lo que suelen utilizar sistemas de lotería o de números aleatorios. El aumento en la complejidad del método de muestreo suele llevar aparejado la necesidad de un aumento en el número de muestras a recoger, de tal forma que de menor a mayor complejidad la técnica puede ser *Simple* (precisa la identificación individual de todos los animales), Sistemático (mediante el establecimiento de un intervalo de muestreo, que habitualmente se utiliza el valor obtenido al dividir el tamaño de la población por el tamaño de muestra necesario), *Estratificado* (dividiendo la población total en grupos y tomando una muestra de cada grupo por métodos simples o sistemáticos), por *Conglomerados o Cluster* (se conoce el número de grupos que hay dentro de la población pero no el número de unidades de cada uno de esos grupos, de forma que se seleccionará una muestra por los métodos simple o sistemático de los grupos y se tomarán la totalidad de las unidades de cada grupo seleccionado. Es el caso en que se conoce el número de jaulas de una explotación, pero no el número de animales de cada una, la forma correcta de proceder consistiría en la selección de parte de las jaulas y la investigación de la totalidad de los animales de cada una de esas piscinas) y por *Etapas o Multiestadio* (es un

sistema mixto de los métodos conglomerados y estratificado, en es que se muestrea en dos fases: una primera fase entre grupos y una segunda fase entre unidades dentro de esos grupos seleccionados en la primera fase previamente)

Los métodos **No Probabilísticos** agrupan todos aquellos métodos en que no todas las unidades tienen la misma probabilidad de formar parte de la muestra, siendo el investigador o es investigado es que decide las que forman parte de la muestra. Hay que tener en cuenta que corremos el riesgo de que la muestra no sea representativa de la población, y por tanto los datos obtenidos no son susceptibles de ser analizados por los métodos estadísticos habituales. En función de quien decide las unidades que integran la muestra, los tipos de muestreo no probabilístico serán con *Voluntarios* (es investigado decide voluntariamente formar parte de la muestra según sus intereses) y de *Conveniencia* (es es investigador quien decide qué unidades forman parte de la muestra, en función de que posean o no algún carácter que desea analizar).

La segunda condición (**homogeneidad**) viene condicionada por el tamaño muestras. Es importante valorar que es número de unidades muestreadas condiciona de forma determinante es tiempo, es coste y validez del trabajo. Por so tanto la utilización de un método de selección adecuado no garantiza que la muestra sea representativa de la población, ya que la muestra también debe cumplir esta segunda condición, que es tamaño sea es adecuado para es tipo de investigación que se desea realizar. Esto supone que es tamaño de la muestra depende de los objetivos de la investigación, ya que no será necesario el mismo número de animases para detectar la presencia de una enfermedad en una población (se trata de detectar as menos un animas positivo), que para determinar la prevalencia (debemos detectar una proporción de animases enfermos en la muestra igual a la que existe en es total de la población).

Para poblaciones en es que el *número de unidades es superior a 1.000* se puede generalizar a efectos de cálculo previo y orientativo, y utilizar la siguiente fórmula:

$$n = 2, J\tilde{N}$$

Sin embargo en condiciones normales para estimar es tamaño de la muestra a seleccionar en una población con unos objetivos determinados es necesario establecer una medida de la dispersión de los datos, manifestada por la varianza (y por extensión la desviación estándar), así como fijar es error aceptado y es nivel de confianza deseado (probabilidad de obtener una respuesta correcta).

Es error aceptado y el nivel de confianza son establecidos arbitrariamente por el investigador, mientras que la desviación estándar se obtiene a través de la literatura o de experiencias y mediciones previas.

En función des objetivo perseguido utilizaremos una u otra fórmula, teniendo en cuenta que la muestra recogida con un fin no siempre puede ser utilizada para obtener unas conclusiones diferentes a las que inicialmente establecidas. Así pues si posteriormente queremos evaluar diferentes hipótesis calcularemos *a priori* los distintos tamaños de muestra necesarios para cada una, y seleccionaremos la mayor, ya que ésta permitirá alcanzar el nivel de confianza establecido para cada uno de los diseños experimentases, que podemos resumir básicamente en los siguientes modelos:

1. Estimar la prevalencia de una enfermedad

En este caso el tamaño de la muestra a tomar depende del error aceptado, del nivel de confianza y de la desviación estándar. La fórmula de la desviación estándar (SD) para variables dicotómicas (presencia o ausencia de enfermedad) es la siguiente, siendo P la proporción (en tanto por uno) del carácter estudiado:

$$SD = \sqrt{P \cdot (1 - P)}$$

En el peor de los casos se puede escoger la situación que nos obligue a seleccionar un tamaño mayor de muestra, que es que corresponde a una probabilidad del 50%.

Para estos casos se utiliza la fórmula:

$$n = \left(\frac{t_{\alpha} \cdot SD}{L} \right)^2$$

donde:

n: tamaño de la muestra.

t_{α} : valor de la t de Student para el nivel de confianza establecido.

L: error aceptado o precisión.

SD: desviación estándar correspondiente a la proporción esperada (P)

Así, trabajar con una prevalencia esperada del 50%, un error aceptado del 5% y un nivel de confianza del 95%, indica que esperamos que, de cada 100 mediciones que se realicen, en 95 la prevalencia tenga un valor comprendido en el intervalo $50\% \pm 5\%$ (45% - 55%).

2. Estimar diferencias entre prevalencias

En la planificación de un experimento para estimar si la prevalencia de una enfermedad en dos poblaciones es distinta, la selección del tamaño de muestra de cada grupo se realiza aplicando la siguiente fórmula:

$$n = \left(\frac{Z(a) + Z(b)}{p1 - p2} \right)^2 \cdot (p1 \cdot q1 + p2 \cdot q2)$$

3. Detección de una enfermedad

Consideramos una población como positiva cuando en al menos d animales se diagnostique la enfermedad. Este parámetro lo establece el investigador y es el número de animales que espera encontrar con dicha característica. En enfermedades graves o cuando interesa obtener una máxima seguridad en el resultado, d se lleva al mínimo valor posible, es decir, se establece en 1.

El cálculo se realiza mediante la siguiente fórmula:

donde:

n: tamaño de la muestra. N: tamaño de la población.

d: número de animales enfermos que esperamos que existan en la población NC: nivel de confianza expresado en tanto por uno (0.95)

A partir de esta fórmula y conocido un tamaño de muestra determinado (n) con el que se desea trabajar, es posible estimar el máximo número de animales positivos (D) que pueden aparecer en una población donde todas las n muestras recogidas han resultado negativas.

donde:

D: máximo número de animales enfermos que pueden quedar en la población Hay que tener en cuenta que cuando se busca detectar una enfermedad en una población se suele recurrir a muestreos no probabilísticos (seleccionando los animales que presentan sintomatología previa) de tal forma que podría ser suficiente una muestra de un tamaño menor al establecido en esta fórmula.

$$n = \left(1 - (1 - NC)^{\frac{1}{d}} \right) \left(N - \frac{d-1}{2} \right)$$

$$D = \left(1 - (1 - NC)^{\frac{1}{n}} \right) \left(N - \frac{n-1}{2} \right)$$

donde:

n: tamaño de la muestra.

p1, p2: porcentajes esperados del parámetro de interés en los grupos 1 y 2 respectivamente Z(a):
valor de la t de Student para el error de tipo 1 (nivel de confianza).

Z(b): valor de la t de Student para el error de tipo II (potencia o poder).

Consideraciones finales

Es preciso hacer algunas consideraciones a tener en cuenta cuando se plantea la correcta utilización de estos métodos:

1. Conocer la existencia de posibles errores de muestreo

La selección de una muestra a partir de una población que interesa estudiar y de la que se desea obtener una determinada información puede generar tres tipos de errores en el proceso del muestreo y que pueden suponer una importante desviación de los resultados obtenidos respecto de la realidad: *Error Sistemático* (relacionado con el método de selección de la muestra al seleccionar deliberadamente algunas unidades o sustituir otras de forma incorrecta; no disminuye al aumentar el tamaño de la muestra pero sí al utilizar un método aleatorio), *Error Aleatorio* (debido generalmente al azar que se relaciona directamente con el tamaño de la muestra tomada; no disminuye utilizando métodos aleatorios pero sí al incrementar el tamaño de la muestra) y *Error de Información* (depende de la forma en que se recoge la información una vez que la muestra ya ha sido seleccionada: recogida incorrecta, procesado defectuoso,...)

2. Utilización de los datos obtenidos para fines distintos para los que inicialmente se recogieron

Este punto hace referencia a la frecuente utilización de datos recogidos para detectar la enfermedad, para el cálculo de la prevalencia de la enfermedad. Hay que tener en cuenta que normalmente en el primer caso se suele utilizar un método no probabilístico (lo que redundaría en una sobreponderación de la prevalencia de la enfermedad) y que el tamaño de muestra suele ser muy inferior al requerido para obtener un cálculo preciso de la prevalencia (es decir con un intervalo de confianza suficientemente reducido como para que sea significativo).

3. Cálculo del error aceptado

En el caso anterior o en ocasiones el número de muestras es limitado (es decir, el tamaño de la muestra se define *a priori*) lo que interesa calcular es el error aceptado para establecer un criterio objetivo que permita tomar la decisión de seguir adelante con la investigación o no, de tal forma que si el error aceptado (L) es demasiado grande en opinión del investigador, es inútil continuar con el control, ya que la medición realizada no será demasiado válida

En este caso las fórmulas son:

$$L = t \cdot \frac{SD}{\sqrt{N}} \quad \text{ó} \quad L = t \cdot SD \cdot \sqrt{\frac{N-n}{N \cdot n}}$$

La primera de ellas para poblaciones consideradas infinitas (mayores de 1.000 unidades) y la segunda para poblaciones finitas (menores de 1.000 unidades).

4. Realización de muestreos estratificados

Habitualmente el número de unidades a muestrear en cada grupo es proporcional al tamaño de ese grupo respecto al total de la población (si un sexo supone en la población el 75% de los animales, en la muestra ese sexo debe estar representado en un 75%), aunque la forma más correcta de proceder es la determinación del tamaño de muestra necesario para cada estrato. La aplicación de cualquiera de estas dos estrategias conlleva problemáticas distintas.

En el primer caso (distribución proporcional de la muestra según la distribución observada en la población) implica que los valores obtenidos

sólo son válidos para el cálculo global de la prevalencia, ya que cualquier intento posterior de cálculo de la prevalencia para cada estrato utilizando esos mismos datos conlleva una drástica ampliación de los intervalos de confianza de los resultados obtenidos (aumento del error aceptado), y por tanto la disminución de la validez de los mismos. Dicho error se puede calcular utilizando las fórmulas descritas en el apartado anterior.

En el segundo caso (determinación del tamaño muestra) necesario para cada estrato) debemos tener en cuenta que se aumentará notablemente el número de muestras necesarias (lo que nos permite obtener resultados aplicables a cada estrato) pero habrá que considerar la necesidad de realizar medias ponderadas y estratificación de tasas para el correcto cálculo del resultado global.

5. Ajuste del tamaño muestral

Por último indicaremos que en el caso de la estimación de la prevalencia, el tamaño de la muestra obtenido mediante las fórmulas respectivas sólo es válido cuando la fracción muestra) es inferior al 10%, lo que generalmente ocurre con poblaciones mayores de 1.000 unidades. Sin embargo, cuando el tamaño de la muestra a tomar es mayor del 10% de la población (poblaciones de menos de 1.000 unidades) se precisa realizar una corrección de la desviación estándar que permitirá estimar el tamaño real de la muestra que se debe tomar, de forma que el tamaño final de la muestra será:

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

donde:

n': tamaño de la muestra ajustado n: tamaño de la muestra. N: tamaño de la población.