

INTRODUCCION AL MUESTREO DE POBLACIONES Y COMUNIDADES MICROBIOLOGICAS

DUNNY CASANOVA ZUÑIGA

*Depto. Salud Pública. Facultad de Medicina
Universidad de Valparaíso*

RESUMEN

Se identifican y comentan brevemente los principales tópicos de muestreo que pueden interesar a un investigador de poblaciones y comunidades microbiológicas naturales: definición de universo y unidades de estudio, determinación del tamaño de una muestra, definición de variables acerca de las cuales se desea efectuar estimaciones muestrales, factores de que depende la precisión de una estimación, diseños clásicos que otorgan alta "representatividad" a una muestra, importancia del azar en la selección muestral y limitaciones en las estimaciones proyectadas a partir de "mini-unidades".

SUMMARY

Principal topics for sampling natural microbiological communities and population are identified and briefly commented. Among these are definition of universe and study units; limits of a sample size; definition of variables desired for estimated sampling; factors on which depend the precision of an estimation, classic designs that grant a high "representativeness" of a sample; importance of random in the selection of samples and limitations in the projected estimations, starting from "mini-units".

Prácticamente ningún estudio de poblaciones o comunidades puede hacerse sin utilizar algún procedimiento de muestreo. La dependencia de esas técnicas va en aumento, mientras más pequeños y abundantes sean los individuos involucrados, ya que se va haciendo cada vez más difícil abarcarlos a todos. Al mismo tiempo, aumenta la dificultad de obtener muestras representativas —especialmente en condiciones naturales— por la disminución que se produce en el conocimiento de las características individuales, del grado de interacción con el medio y del aumento de la complejidad en el manejo de los organismos.

En los estudios meso o microbiológicos de comunidades, las unidades quedan definidas por el área o el volumen biótico con sustrato abiótico, abarcado por un instrumento de recolección, que puede ser una simple pala para tomar "muestras" de suelo en un estudio de microflora, o una botella para obtener una muestra de agua en un estudio de plancton, o una red planctónica para este mismo objeto, o una cuchara doble para obtener "muestras" de sedimento. Es decir, la unidad a que me refiero corresponde estrictamente al contenido recolectado por un instrumento de recolección bien definido, en otras palabras, a la "muestra" de tierra, agua, hojarasca, etc. que se haya obtenido con éste. Para evitar confusiones posteriores, preferiré llamarlas "unidades de estudio" más bien que "muestras" y reservaré el término de muestra para el total de estas unidades que hayan sido recolectadas desde el universo.

En una investigación puede estarse definiendo más de una unidad de estudio, cada una de ellas con

su propio procedimiento de recolección, pero no es correcto definir una misma unidad con dos o más procedimientos de recolección muy disímiles porque produciría confusión en los resultados. Por ejemplo, si para un estudio de comunidad planctónica se usara una botella y una red para recolección, obviamente la red recolectará organismos de cierta dimensión en tanto la botella los recolectará casi a todos (con excepción de los de gran movilidad). Por otra parte, las exigencias de confiabilidad en los resultados de un estudio, obligan al investigador a conocer y seleccionar los diseños de muestreo más apropiados y a crear sus propios diseños alternativos, buscando un equilibrio entre la ganancia en representatividad —que hará más confiables sus resultados— y la complejidad del diseño, que puede elevar los costos del muestreo. En este juego es muy importante que el investigador conozca algunos elementos de la teoría muestral, que le ayuden a encontrar ese equilibrio.

Una de las primeras preguntas que suele surgir cuando se enfrenta un problema de muestreo, se refiere al lugar y tiempo en que se efectuará éste. Cualquiera sea el estudio que se desea realizar, siempre es conveniente caracterizar el medio espacio-temporal al cual se quiere proyectar los resultados, con el doble propósito de tener un marco de referencia claro para efectuar el diseño muestral y formalizar la base de comparación con otros estudios. Esta caracterización frecuentemente abarca aspectos abióticos superficiales, algunas características de cobertura vegetal, si es un espacio abierto y referencias temporales. Por ejemplo, para un estudio sobre microflora del suelo,

se describe el "área de estudio" en términos geográficos, climáticos, edáficos, vegetacionales y aún históricos y biológicos, que proporcionan un marco referencial muy completo para efectuar el diseño muestral y proyectar los resultados.

En este universo definido, el investigador deberá identificar cuales serán las unidades que le proporcionarán los datos necesarios para cumplir los objetivos del estudio. Tales unidades pueden ser bióticas, abióticas o una mezcla de ambos tipos; pueden ser espaciales, temporales o espacio-temporales. Podría ser útil estimar el número de unidades de estudio que se obtendría de un universo, teniendo alguna referencia acerca de su dimensión total y conociendo la capacidad del instrumento de medición. Por ejemplo, si se sabe que con una pala se puede obtener muestras de tierra de aproximadamente 0.1 m³ entonces, de un área de alrededor de 1000 m² de tierra se podría obtener unas 1000 unidades de 0.1 m³ de la capa superficial de 0.1 m de profundidad. En la mayoría de los estudios de comunidades meso o microscópicas, los universos de unidades son tanto o más grandes que los del ejemplo anterior, pudiéndoseles considerar entonces como de tamaño infinito ($N \rightarrow \infty$).

El siguiente aspecto de interés para el investigador suele ser la determinación del tamaño de muestra que necesitaría para realizar su estudio. Aparte de una decisión arbitraria, éste puede elegir alguno de los procedimientos siguientes:

1. Determinación del tamaño máximo.

Consiste en determinar el número máximo de unidades que es posible recolectar con los recursos y tiempo disponibles para realizar el estudio. La aplicación de este método exige efectuar estimaciones de costo por unidad de estudio, que abarcan la recolección y el manejo posterior de datos que conducirá a los resultados finales. Si representamos por c el costo por unidad y por R la suma de recursos disponibles, entonces, obviamente $n = R/c$. Este procedimiento puede parecer simplista para algunos entendidos, pero no debemos olvidar que, en muchos casos, es la única forma de proceder que tiene el investigador si desea llevar adelante su estudio.

2. Determinación del tamaño mínimo.

Consiste en estimar el número mínimo de unidades que es menester seleccionar para cumplir ciertas exigencias acerca del nivel de imprecisión y nivel de confianza en las proyecciones que se desee hacer hacia el universo. Este método, que sin duda sería aprobado por los muestristas, suele conducir a números de unidades desusadamente grandes que obligarían a realizar un arduo trabajo de recolección y manejo posterior de unidades —muestra, no siempre factible para los grupos de investigadores los que, generalmente, trabajan con recursos limitados. Por ejemplo, si un investigador desea efectuar un estudio acerca del promedio de colonias fúngicas aisladas (cultivo en agar Rose - Bengall) por unidad de estudio (cilindros de suelo), con un nivel de imprecisión no mayor que d y un nivel de confianza del 95%, usando un diseño de muestreo al azar simple, necesitaría como mínimo recolectar al azar:

$$n = t_{n-1}^2 \left(\frac{\hat{\sigma}}{d} \right)^2$$

Si, de estudios anteriores o de una muestra piloto, él ha obtenido información de que el promedio de colonias aisladas es de 4.0 colonias $\times 10^5$ grs. de suelo fresco, con desviación standard de 1.9 colonias y él no desea aceptar un nivel de imprecisión mayor que el 10% de la estimación del promedio, entonces:

$$n = t_{n-1}^2 \left(\frac{1.9}{0.4} \right)^2 = 22.6 t_{n-1}^2$$

lo que en un proceso iterativo de búsqueda de t nos conduce a:

$n=90$ unidades de suelo fresco que deben seleccionarse al azar en su universo, suponiendo que la muestra será recolectada en un solo momento en el tiempo.

3. Equilibrio entre tamaño máximo y mínimo.

Consiste en determinar el número de unidades que permitiría satisfacer a la vez una mínima confiabilidad aceptable para la proyección de los resultados y la máxima exigencia de recursos para la realización del estudio. Este método puede considerar la rebaja de los costos unitarios mejorando por un lado las técnicas de recolección y manejo de datos y por otro, optimizando el diseño de muestreo. Por su flexibilidad debiera ser, a mi modo de ver, el que todo investigador intentara desarrollar como norma, salvo que tuviera recursos ilimitados. En el ejemplo anterior es posible que, pensando en un diseño de muestreo diferente, la variabilidad resulte disminuida, es decir, S^2 será menor, disminuyendo por lo tanto también el tamaño mínimo de muestra necesario. Suponiendo que, en un muestreo piloto se obtuvo las desviaciones standard 0.5, 2.7 y 1.1 para el promedio de colonias aisladas en 3 estratos diferentes del mismo universo, cuyas medidas de tamaño fueron respectivamente 0.6, 0.2 y 0.2, entonces, se necesitaría seleccionar.

$$n = t_{n-1}^2 \frac{(0.6)(0.5) - (0.2)(2.7) - (0.2)(1.1)}{0.4^2} = 11.6 t_{n-1}^2$$

$n=50$ unidades de suelo fresco, usando un diseño de muestreo al azar estratificado proporcional. Es decir, el solo hecho de cambiar el diseño disminuyó el tamaño de muestra mínimo necesario a casi la mitad del anterior.

La próxima pregunta que suele hacerse el investigador que va a usar el muestreo, se refiere a lo que desea estimar con su muestra o, en otras palabras, lo que desea proyectar hacia el universo. A poco de revisar los objetivos de su investigación, será capaz de definir las variables acerca de las cuales deberá recolectar información de las unidades de estudio para cumplir éstos y también será capaz de seleccionar o construir los indicadores que le permitirán resumir la información relativa a esas variables. Esos indicadores pueden pro-

veer valores aislados (promedios, desv. standard, proporciones, correlaciones, indicadores de diversidad, etc.) o conjuntos de valores (distribuciones de frecuencias, gráficos, etc.) los que debieran proceder del cálculo con los datos de todas las unidades del universo para estar seguros que corresponden a valores reales de ese. Usando la muestra, en cambio, los valores que se obtengan corresponderán a los reales para ese solo grupo y no para el total. Sin embargo, con el acto de proyectar los resultados de la muestra hacia el universo, el investigador transforma los valores muestrales de los indicadores, en estimaciones de sus valores reales. Entonces le surgirá una duda ¿qué tan precisas son sus estimaciones?, es decir, ¿qué tan cercanas están de los valores reales del universo?

La precisión de una estimación particular depende a lo menos de tres factores: a) variabilidad de la variable a que se refiere el indicador, b) tamaño de la muestra de unidades de estudio y c) diseño muestral usado. Estos tres factores se encuentran considerados en cualquier estimación del error muestral que se haga para determinar el nivel de imprecisión de la estimación del valor de un indicador.

a) Variabilidad

Toda variable tiene la posibilidad de presentarse de un modo diferente en distintas unidades muestrales, por ejemplo, biomasa bacteriana, biomasa fúngica, actividad metabólica del suelo, pueden tener distintos valores para unidades de muestreo diferentes. Esto es lo que se llama variabilidad. Si los valores de alguna de estas variables son similares en las unidades consideradas se dirá que la variabilidad es baja o que la homogeneidad es alta para esa variable. Si, en cambio, presenta valores muy disímiles, se dirá que su variabilidad es alta o que su homogeneidad es baja. Este concepto representa, pues, un atributo propio de la variable, que permite calificar el indicador usado para resumir sus valores. El valor de un indicador cuya variable tiene poca variabilidad, es más confiable que el de otro cuya variable tiene alta variabilidad. Por ejemplo, el valor de la biomasa bacteriana promedio será más confiable si se obtuvo de valores muy similares de biomasa de las unidades de muestreo en que se determinó, que si se obtuvo de valores muy disímiles. Sin duda, tendremos más confianza en inferir hacia el universo el valor del primer promedio que el del otro porque suponemos que, al obtenerse su valor con baja variabilidad a partir de una muestra de unidades, es muy posible que represente bien (esté cercano) al valor real de ese promedio en el universo. Por eso, no nos sorprenderá si decimos que "el nivel de imprecisión de una estimación, se encuentra relacionado directamente con la variabilidad correspondiente a la variable del indicador estimado.

b) Tamaño de la muestra

El número de unidades que elijamos desde el universo para inferir los valores que tendrán los indicadores de interés en éste, es otro factor que influye sobre la precisión de la estimación. Tendemos a sospechar que una muestra de escasas unidades será poco representativa del universo por cuanto podría estar

abarcando sólo algunas características de éste y por tanto, sólo parte de la variabilidad de las variables que se están estudiando. Por el contrario, mientras mayor sea la muestra, más posibilidades se tendrá de que esté abarcando la mayor parte de la variabilidad que hay en el universo y, por tanto, que los indicadores proporcionen valores más cercanos a los reales, es decir, con menor nivel de imprecisión, lo que nos sugiere una relación inversa entre este nivel y el número de unidades de la muestra.

c) Diseño muestral

Frecuentemente nos estamos refiriendo a la "representatividad" de una muestra respecto al universo del que se seleccionó, sin explicarnos claramente su significado. Yo pienso que es un atributo de la muestra que corresponde concretamente a la "cantidad" de variabilidad y de proporcionalidad del universo, que esta lleva y que se refiere a la o las variables más importantes del estudio. Así pues, si lo fundamental de mi estudio son las variables biomasa bacteriana y biomasa fúngica, la muestra que yo seleccione debiera llevar la mayor parte de la variabilidad de esas biomasa de tal modo que haya valores altos, medios y bajos, en proporción similar a los que estas variables presentan en el universo. La obtención de una alta representación, así entendida, obliga a un mejor conocimiento previo del universo el que, evidentemente, no debe ser en términos de las mismas variables que se desea estudiar, sino en términos de otras relacionadas con éstas y de fácil visualización, a las que llamaré "variables índices". Por ejemplo, para el estudio de la biomasa fúngica del suelo pueden servir como variables índices: cobertura vegetal, humedad, textura del suelo, etc., con cuyos niveles se puede dividir el área total en áreas más pequeñas, muy diferentes entre si, en las cuales se supone que la biomasa tendrá también valores claramente distintos. Esta forma de razonamiento, es la que conduce a diseñar procedimientos de muestreo que otorguen alta representatividad a la muestra. Un par de diseños muestrales muy conocidos persiguen precisamente ese objetivo: el muestreo por estratos proporcionales y el muestreo de conglomerados.

Muestreo por estratos

Habiéndose identificado una o más variables índices, es posible dividir el universo en grupos, áreas o volúmenes cuyas unidades de muestreo presenten alta homogeneidad interna respecto a tales variables y por tanto, alta heterogeneidad entre unidades de grupos distintos. A tales grupos los llamamos "estratos". Obteniendo alguna medida de tamaño para cada uno de estos estratos será posible repartir la muestra total n entre todos ellos, en proporción a esa medida de tamaño. Si W_h representa la medida de tamaño del estrato h , entonces: $n_h = n \times W_h$ representa el número de unidades de muestreo que debieran ser seleccionadas del estrato h , en relación a la medida de tamaño que este tiene. Por ejemplo, si la relación de tamaño entre 3 estratos de un universo, es de 3: 2: 5 y se desea repartir en esos estratos en forma proporcional, una muestra total de 25 unidades de muestreo, entonces: $W_1 = 0.3$, $W_2 = 0.2$ y $W_3 = 0.5$, por lo que: $n_1 = 25 \times 0.3 = 8$, $n_2 = 25 \times 0.2 = 5$ y $n_3 = 25 \times 0.5 = 12$. Entonces, deberá seleccionarse 8 unidades

del primer estrato, 5 del segundo y 12 del tercero. Con este procedimiento se asegura que la muestra seleccionada tendrá representaciones de cada uno de los estratos en proporción a su importancia en el universo y como estos a su vez, representan características distintas de la variable que se va a estudiar, la muestra será altamente representativa de aquél.

Muestreo de conglomerados

Otra modalidad consiste en dividir el universo en grupos, áreas o volúmenes cuyas unidades de muestreo presenten alta heterogeneidad interna respecto a la o las variables índices y por lo tanto, cada grupo constituirá una miniatura del universo que será llamada "conglomerado". En este caso, bastará con seleccionar uno o más de estos conglomerados para obtener la muestra de unidades que se necesita. Por ejemplo, habiéndose dividido un universo en 600 conglomerados de alrededor de 10 unidades muestrales cada uno, se debe seleccionar tres de ellos para obtener una muestra total de 30 unidades. Otra posibilidad es seleccionar un número mayor de conglomerados y muestrear dentro de cada uno de ellos. Por ejemplo, seleccionar primero 10 de entre los 600 anteriores y luego seleccionar 3 unidades dentro de cada conglomerado ya elegido, lo que igualmente nos provee de una muestra total de 30 unidades.

Tanto en el muestreo por estratos como en el de conglomerados los principales problemas lo constituyen la elección apropiada de la o las variables índices, la determinación de las medidas de tamaño y la delimitación de los grupos.

Selección al azar

Cualquiera sea el diseño muestral propuesto, en el momento de efectuar la selección ya sea de las unidades de muestreo o de los conglomerados, deberá resolverse si ésta habrá de ser al azar o no. Yo diría que hay a lo menos dos razones de peso que hacen decidirse por el azar en la selección:

1) Que es altamente probable que la muestra contenga lo que se presenta en mayor proporción en el universo. Esto por sí solo permite confiar más en la selección al azar, que en la dirigida.

2) Que es posible utilizar procedimientos de inferencia estadística en el análisis de resultados, porque los datos que se recolecten quedarán sometidos a la "ley de probabilidades" en que se basan tales procedimientos. En particular, se podría efectuar estimaciones

por intervalo de confianza para valores de indicadores en el universo, mediante la estimación del error de muestreo.

Procesos de estimación en Microbiología

No podría terminar esta breve introducción al muestreo sin recordar que el investigador microbiológico debe enfrentar, por lo general dos diferentes procesos de estimación: a) estimación acerca de lo que ocurre en el universo, a partir de la muestra de unidades seleccionadas en terreno y b) estimación acerca de lo que ocurre en cada unidad seleccionada, y por ende en la muestra, a partir del "muestreo de partes" de las unidades seleccionadas. La muestra formada por estas "mini unidades", a su vez, raramente es escrutada en forma directa, por la pequeña dimensión de los organismos involucrados, sino a través del "cultivo" de estas mini unidades en medios artificiales sujetos a una variada e intensa manipulación. Este segundo tipo de estimaciones puede servir para algunos indicadores poblacionales, como natalidad, mortalidad, fertilidad, capacidad de porte, forma y velocidad de crecimiento, velocidad metabólica, etc., pero no es útil para la mayoría de los indicadores usados en estudios de comunidad salvo que ese estudio se limite a las comunidades "cultivadas". Sólo es factible efectuar estimaciones de indicadores de presencia —ausencia y variedad de especies, los que obviamente se verán afectados por el medio de cultivo que, generalmente, actúa en forma selectiva sobre los organismos de distintas especies.

Ya sea el estudio de poblaciones o de comunidades, las estimaciones que resulten a partir de mini unidades deben ser cuidadosamente extrapoladas a nivel de unidades de estudio en terreno, pudiendo ser necesario el uso de muestreo con diversos medios de cultivo, en un diseño de experimento conducido, en lo posible, bajo la acción de los niveles de los principales factores a que están sometidos los microorganismos naturalmente. Estas miniunidades, en la práctica, producirán estimaciones acerca de lo que ocurre en la muestra de unidades recolectadas en terreno y —a través de la proyección de esa muestra— de lo que ocurre en el universo total.

Esto obliga a los investigadores a describir con claridad las condiciones bajo las que se hizo el cultivo, para facilitar la contrastación de ellas con las del medio natural que habrán sido descritas al definir el universo en estudio.

REFERENCIAS

- CAMPBELL, R. (1977). Microbial ecology. Edit. Blackwell S.P. N. York.
 COCHRAN, W.G. (1963) Sampling Techniques. Edit. J. Wiley. N. York.
 COOKE, W.B. (1979). The ecology of the fungi. CRC Press. Florida.
 FREEDMAN, B. and HUTCHISON, T. (1980). Long-term effects of smelter pollution at Sudbury, Ontario, on forest community composition. Can. J. Bot. 58: 2123-2140.
 KISH, L. (1972). Survey Sampling. Edit. J. Wiley. N. York.
 MARGALEF, R. (1974). Ecología. Edit. Omega. Barcelona.
 ODUM, E.P. (1972). Ecología. Cap. 19. N. Edit. Interamericana. México.
 POCH, A. (1969). Curso de muestreo y aplicaciones. Edit. Aguilar. Madrid.