Diseño de muestreo 2

2. EL DISEÑO DE MUESTREO

Laura C. Pereyra & Marcos Vaira

Instituto de Ecorregiones Andinas (INECOA), CONICET — Universidad Nacional de Jujuy, CCT Salta-Jujuy, San Salvador de Jujuy, Jujuy, Argentina.

El diseño de investigación es una guía que ayuda a organizar una investigación utilizando una metodología particular. Tanto en los estudios de campo como de laboratorio representa una parte fundamental, pudiendo ser esta etapa la que determine el éxito de un estudio. Cuando se presenta de manera adecuada, el estudio aumentará las chances de desarrollarse apropiadamente, obteniendo la información para la que fue diseñado. Por el contrario, si su elaboración no tiene la atención necesaria por parte del investigador, los resultados obtenidos pueden ser incompletos o erróneos, no logrando responder la pregunta que dio origen a la investigación.

Toda investigación científica comienza con una observación y el posterior planteo de una pregunta. De esta manera, las preguntas de investigación se traducen en modelos que describen de manera adecuada la observación y en diseños metodológicos específicos que permiten responderlas. El diseño del muestreo es el vínculo entre la pregunta de investigación y los resultados obtenidos. Si se toma en cuenta que cualquier observación es siempre incompleta y que la objetividad de la misma puede llegar a depender de la experiencia del observador y del proceso que se utilice para registrarla, se puede decir que el diseño de investigación elegido determina la confiabilidad de la respuesta obtenida y, en otros términos, su validez.

La realización de un censo de los individuos de una comunidad o de las variables de interés dentro de toda la población en estudio puede ser una tarea imposible para la mayoría de los sistemas biológicos, volviéndolo algo inviable. Es por esto que generalmente se utiliza una muestra representativa, es decir se selecciona parte de la población para observar y estimar algún parámetro de la misma, por ejemplo: la ocurrencia de una especie, la diversidad de especies en un área, sus abundancias relativas, entre otros.

El diseño de muestreo comprende tres preguntas frecuentes: dónde y cuándo tomar la muestra, cuántas muestras registrar y cómo se deben recolectar los datos en tiempo y espacio. Si bien se enfatiza la necesidad de consultar manuales específicos de diseño de investigación para planificar adecuadamente un estudio y algunas publicaciones muy recomendadas sobre el diseño de investigación en biología (ej:[1-13]), se indican a continuación algunas sugerencias elementales a tener en cuenta al abordar estos tres puntos cuando se afrontan estudios de relevamiento de la diversidad, relevamientos específicos y/o monitoreo de poblaciones de anfibios.

2.1 La pregunta y los objetivos del estudio

Tanto el planteo de la pregunta de investigación, como de los objetivos derivados de la misma, es un comienzo obligado en cualquier investigación. Su

importancia radica en que ambos permitirán organizar el estudio, definiendo el enfoque y la escala (espacial y/o temporal) en la que se desarrollará el mismo, delimitándolo para ser abordado de manera realista, así como los límites y alcances de sus resultados.

Resulta importante enfatizar que previo al planteo de la pregunta, es necesario familiarizarse con la bibliografía disponible sobre el tema de estudio, no sólo para contar con un cuerpo de información básico, sino también para evitar redundar en estudios ya realizados, evaluar posibles problemas con el diseño y maximizar el alcance y aplicación de los resultados obtenidos.

En base a la pregunta se podrá generar la hipótesis de trabajo que representa una explicación tentativa a esa pregunta, basada generalmente en información previa. Las hipótesis deben ser planteadas de forma clara y concisa como afirmaciones con algún grado de generalidad.

Junto con la hipótesis es aconsejable plantear una o varias predicciones, las cuales son consecuencias "directamente observables". La ausencia de predicciones puede impedir comprender, apoyar o criticar las deducciones que se desprendan de los resultados. Mientras que las hipótesis no pueden medirse directamente, las predicciones pueden ser cuantificables y enunciarse en las variables que se van a medir[(4)]. Aunque resulta importante recomendar el enunciado de hipótesis y sus predicciones asociadas cuando se decide aplicar el método hipotético deductivo, es necesario mencionar que existen posiciones que cuestionan el uso de hipótesis en estudios de ecología (ver[14]).

Los objetivos deben ser mensurables e inequívocos y no deben dejar margen para concluir si al finalizar el estudio se han logrado cumplir. La condición de "objetivos mensurables" implica que, además de descriptivos, deberán ser cuantificables y susceptibles de evaluarse estadísticamente.

2.2 ¿Dónde y cuándo? Definiendo la escala espacial y temporal del estudio

La percepción del sistema biológico bajo estudio está inevitablemente sesgada o incompleta, determinada principalmente por la escala, tanto espacial como temporal a la que se ha decidido realizar la observación. Es por esto que definir el tamaño del área y la duración del período en el que se desarrollará el estudio, permitirá obtener evidencia suficiente para cumplir con los objetivos propuestos y por ende discutir de manera adecuada el significado de los resultados.

Entonces, como primer paso, se debe definir la población en estudio considerando el área geográfica y/o poblaciones o subpoblaciones específicas a

las que aplica el estudio. Por ejemplo, un estudio de la abundancia de una especie a gran escala (en todo su rango de distribución) puede indicar un aumento poblacional, pero el mismo estudio a escalas más pequeñas (en poblaciones localizadas) puede indicar poblaciones estables o en disminución. Se debe además definir la duración del estudio (meses, estaciones, años o generaciones). Los parámetros de una población (como tamaño o tasa de crecimiento) varían con el tiempo como resultado de factores intrínsecos y extrínsecos, por lo tanto, los resultados de parámetros de una población obtenidos en un sólo momento en el tiempo generalmente resultan evidencia insuficiente para determinar el estado de una o varias poblaciones de la especie. El período de tiempo elegido debe adecuarse, en lo posible, a la biología y fenología de la especie.

2.3 ¿Cuánto? Definiendo el tamaño de la muestra

Una muestra representativa de la población necesita incluir un número adecuado de réplicas de las unidades experimentales (individuos, poblaciones, especies, parches de hábitat, entre otros) para así asegurar que la información obtenida a partir de la misma permita una estimación razonable. Como parte de un diseño adecuado, el tamaño de la muestra requerida para el estudio resulta esencial. Por lo tanto, incluir el procedimiento de muestreo y el tamaño de muestra necesarios para medir adecuadamente la/s variable/s de interés garantizará que se recopilen el tipo, la calidad y la cantidad de datos apropiados para estimar si los parámetros que se van a medir cumplen las condiciones de objetivos mensurables que reducen la ambigüedad y subjetividad en la interpretación de los resultados.

Por un lado, es necesario que el tamaño de la muestra sea lo suficientemente grande como para ser representativa de la población en estudio, pero no tan grande como para sacrificar demasiados individuos o desperdiciar tiempo y recursos al tomar o seleccionar la muestra. ¿Cuál debería ser entonces el tamaño de la muestra para que sea adecuada? Una respuesta concreta es que va a depender principalmente de la variabilidad que presentan los elementos de interés que componen la población en estudio, mientras más alta sea esta variabilidad más grande deberá ser la muestra elegida. Aquí aparecen dos conceptos que deberán ser tomados en cuenta al momento de definir el tamaño muestral: el error tolerable y el nivel de confianza que se espera de las estimaciones. El error puede provenir de varias fuentes, pudiendo estar relacionados directamente con el diseño de muestreo o ser errores aleatorios generados por variaciones impredecibles e inevitables, errores de medición, problemas de detectabilidad y/o errores debido a una variabilidad sistemá-

tica y no planificada de los datos (el error sistemático). Ambos tipos de error, tanto los relacionados con el diseño como aquellos no relacionados al diseño, pueden llevar a un registro sesgado de la variable respuesta, resultando en conclusiones erróneas. Una ejecución exitosa del estudio requerirá que el investigador evite introducir errores sistemáticos y minimice los errores aleatorios.

Pueden surgir distintos inconvenientes al momento de definir el tamaño óptimo de una muestra. Por un lado, la necesidad de conocer la varianza poblacional previo a la toma de datos puede sortearse realizando un estudio piloto previo o utilizando información de investigaciones previas sobre la varianza presente en los elementos de interés. Existen varias formas de estimar el tamaño óptimo de la muestra. Existen numerosos materiales de consulta que explican de manera clara la estimación de tamaño muestral, entre ellos se pueden mencionar los trabajos de Rao[15], Thompson[16] y Krebs[17].

2.4 ¿Cómo? Diseños de muestreo estandarizados

Como los sistemas generalmente no son homogéneos, ya sea en el espacio o en el tiempo, es necesario no sólo contar con un número grande de unidades experimentales, sino también planificar cómo se van a ubicar las mismas. Los tipos de muestreo permiten definir cómo se van a asignar las unidades experimentales en el espacio bajo las distintas condiciones del estudio, cuántas réplicas de estas unidades se van a tomar en cuenta y la secuencia espacial y temporal en la que se realizarán las mediciones.

Muestreo no probabilístico

En la práctica, es común que los investigadores elijan las ubicaciones de las unidades de muestreo de forma arbitraria u oportunista al momento en que se realiza el muestreo, socavando la validez de la inferencia estadística. En este tipo de muestreo, las unidades experimentales se seleccionan en base a un criterio que no es al azar por lo que los individuos no presentan la misma posibilidad de ser incluidos en el estudio.

Este tipo de relevamiento puede ser más fácil y económico, pero presenta altos riesgos de sesgo en los datos registrados, y no se debe pretender hacer inferencias que se extiendan a todas las poblaciones de una especie.

La selección de muestras de forma oportunista o por conveniencia no necesariamente debe considerarse un procedimiento erróneo. Se pueden elegir

deliberadamente ubicaciones específicas de muestreo debido a características particulares donde el sesgo de muestreo es explícito y deliberado. Este tipo de métodos pueden ser útiles para estudios exploratorios, en los cuales no se pretende poner a prueba una hipótesis sino, por ejemplo, generar conocimiento detallado de un fenómeno específico. Existen otras situaciones donde puede ser realmente difícil obtener muestras aleatorias donde la distribución espacial del hábitat objetivo es desconocida, que suele ser el caso de estudios en sitios remotos o poco conocidos. Así, el marco espacial del sitio de muestreo no puede definirse específicamente y las ubicaciones de las unidades de muestreo no pueden seleccionarse antes del evento de muestreo.

Es aconsejable evitar en lo posible este tipo de muestreo. Actualmente existen numerosas herramientas y rutinas en paquetes estadísticos que, combinando la manipulación de datos espaciales con herramientas SIG (en forma de líneas o polígonos), permiten seleccionar las ubicaciones de las unidades de muestreo mediante una rutina que aplica un proceso aleatorio a elección del investigador (al azar, sistemático o en bloque) y repetible. Estos métodos se describen en forma general más abajo pero serán desarrollados con más detalle en la **Parte II** del Manual.

Muestreo probabilístico

Muestreo aleatorio simple (Figura 2.4.1A). Es el tipo de muestreo más simple, y consiste en considerar que todos los elementos de la población tienen la misma probabilidad de ser registrados. De esta manera se genera una aleatorización de todos los factores no controlados por el investigador y que pueden influir en los resultados. Se evita además introducir sesgo en el estudio y la dependencia entre observaciones.

Esta técnica es apropiada en el caso de que el área o la población relevada sean homogéneas o no se tenga información que indique lo contrario. Una dificultad en aplicar este tipo de muestreo a estudios de relevamiento de especies se relaciona con la dificultad que puede existir en llegar a los sitios designados de manera aleatoria para el estudio, ya sea porque no son accesibles o a su topografía, lo cual limita la elección de los mismos a aquellos que son fácilmente accesibles, siendo la presencia de caminos o senderos en el área lo que generalmente determina qué unidades espaciales pueden seleccionarse. Esto genera un sesgo en la selección de los sitios de relevamiento.

Muestreo aleatorio en bloques o estratificado (Figura 2.4.1B). Se recomienda emplear este método cuando el área o la población relevada no son homogéneas, lo que podría generar distintas condiciones entre sectores o elementos

que pueden afectar a nuestras unidades experimentales. La heterogeneidad en las condiciones a las que estarán expuestas las réplicas puede controlarse dividiéndolas en estratos (bloques), los cuales presentarán características lo más parecidas posibles, y dentro de los cuales se realizará la selección aleatoria de las unidades experimentales. Por ejemplo, si pretendemos relevar y comparar la diversidad de anfibios (nuestra variable respuesta) en dos sitios con distinta altitud (nuestro factor de interés) donde ambos sitios presentan zonas de bosque primario y zonas de bosque secundario con áreas abiertas. Si bien estas posibles diferencias en la estructura de los bosques pueden no ser de interés en el estudio, es probable asumir que presenten un efecto sobre la diversidad de especies registradas en las réplicas presentes en una u otra zona, por lo que resulta recomendable definir cada zona como bloques. Dentro de cada bloque (bosque primario o bosque secundario con áreas abiertas) se elegirá al azar a las unidades experimentales. De esta manera, este tipo de diseño convierte a la variabilidad sistemática no planificada en variabilidad sistemática planificada.

Estos métodos de dos etapas, implica dividir el marco de muestreo en estratos más pequeños dentro de los cuales se toman luego muestras aleatorias simples.

Muestreo sistemático (Figura 2.4.1C). Se recomienda emplear este método cuando se sospecha que la presencia de una réplica afecta alguna propiedad de las réplicas más cercanas o cuando la población es bastante irregular respecto al carácter que estamos estudiando, por lo que se establece un criterio previo para la selección de las réplicas de la muestra. De esta manera, se parte de un punto al azar y se plantea una distancia mínima ***k***, la cual puede ser espacial o temporal, para seleccionar los siguientes sitios de muestreo. Por ejemplo, seleccionamos un cuerpo de agua al azar y planteamos una distancia mínima entre los cuerpos de agua a relevar considerando que la distancia de dispersión de los individuos podría permitir que se muevan desde un cuerpo de agua a otro.

Este muestreo puede aplicarse también en casos donde la variable independiente de interés varía de forma gradual a lo largo del espacio. De esta manera, se fija la medición de las variables de interés (tanto de respuesta como independientes) a intervalos regulares a lo largo de una transecta que recorra este gradiente de condiciones. Es necesario identificar si existe algún patrón regular del ambiente o de los organismos en las unidades experimentales para evitar sesgo.

Muestreo adaptativo (Figura 2..4.1D). Se plantea adaptar el criterio de selección de unidades experimentales a medida que se va realizando el relevamiento, basado en los resultados preliminares que se vayan obteniendo. De

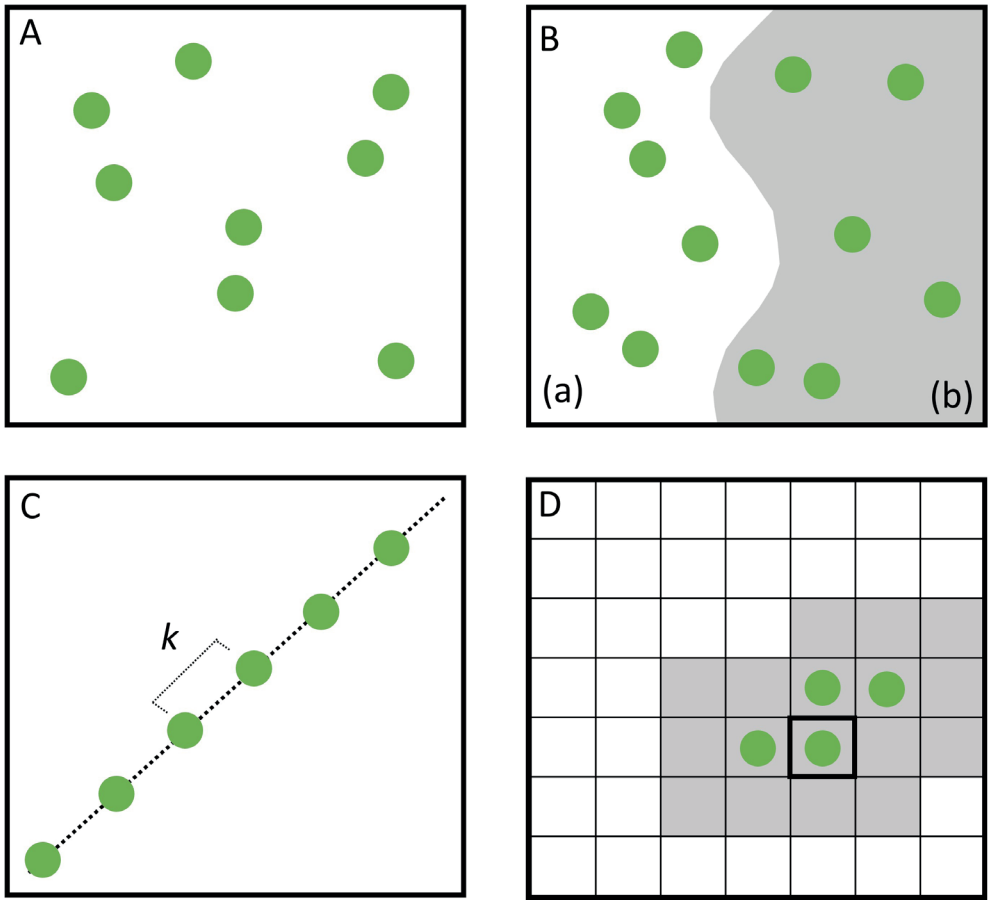

Figura 2.4.1. Ejemplos de diseño de muestreo probabilístico que indican cómo se asignarían las unidades experimentales (círculos verdes) en el espacio. **A. Muestreo aleatorio simple**. **B. Muestreo aleatorio en bloques o estratificado**. Las letras **(a)** y **(b)** indican distintos tipos de usos o situaciones (bloques) como bosque primario y secundario. **C. Muestreo sistemático**. La letra ***k*** corresponde a la distancia mínima definida desde el primer punto elegido al azar hasta el siguiente punto seleccionado. **D. Muestreo adaptativo**. El recuadro más grueso corresponde a la unidad experimental inicial. Gráficos: L. Pereyra.

esta manera, la selección de los sitios a relevar va a depender de los valores de la variable respuesta observados durante el estudio, lo que puede llevar en muchos casos a resultados más efectivos. Por ejemplo, para una especie rara con una distribución agrupada en su abundancia, el uso de un método de muestreo aleatorio puede resultar en la selección de sitios para realizar los relevamientos que resulten vacíos. En estos casos, es aconsejable utilizar un muestreo adaptativo donde se considera una unidad experimental inicial; cuando se logra detectar individuos en una de estas unidades, se comienza a incluir en el estudio áreas cercanas a ésta unidad, proceso que se repetirá hasta que se obtenga una muestra representativa de ésta población agrupada.

Este tipo de muestreo reduce el tiempo y costo de los relevamientos y puede ayudar a aumentar la efectividad del muestreo al aumentar el número de observaciones mejorando la precisión de los resultados para un tamaño muestral dado. Es posible que se destine mucho esfuerzo para detectar la unidad muestral que presente la ocurrencia de la especie, y una vez localizada, el

criterio para seleccionar las unidades experimentales vecinas a aquella que presenta registros debe ser evaluado dado que un criterio erróneo afectará la efectividad del esfuerzo de muestreo.

Caja 2.1 - Detectabilidad

La detectabilidad se refiere a la probabilidad de registrar individuos cuando éstos se encuentran dentro de la unidad experimental analizada, y es una fuente importante de variación necesaria de incluir en todo protocolo de relevamiento. La detectabilidad de cada especie varía tanto entre ambientes con distintas características como en distintos momentos de la temporada reproductiva. Esta diferencia puede estar asociada también a la abundancia de individuos presentes en la unidad experimental, individuos de especies abundantes tienen una probabilidad de detección mayor que los individuos de especies raras.

La detectabilidad de las especies de anfibios puede variar además entre observadores con distinta experiencia y debido a cambios en las condiciones climáticas. Una de las posibles soluciones más recomendadas es adaptar el protocolo de muestreo tomando en cuenta las detectabilidad de las distintas especies y planificar así el diseño en el tiempo y el espacio adecuado. Otra forma de controlar la detectabilidad imperfecta de ciertas especies es mediante repeticiones espaciales o temporales de los relevamientos. Se pueden diseñar varias visitas a un mismo sitio, o definir varios puntos de relevamiento dentro de una misma unidad experimental. Debe tenerse especial precaución que estas repeticiones son consideradas pseudoréplicas[18]. Se espera entonces que al aumentar las visitas a los sitios relevados (o el número de sitios relevados dentro de la misma unidad experimental) aumente la probabilidad de registro de las especies más elusivas.

La variación temporal de la detectabilidad de las especies puede relacionarse a la fenología de actividad que presenta cada especie, tanto referida al patrón del comportamiento estacional como a la hora del día en la que cada especie se encuentra más activa. De esta manera, es recomendable que el investigador tenga un conocimiento previo del ensamble de anuros que potencialmente va a encontrar en los sitios y ajustar los relevamientos para abarcar la temporada de actividad de todas las especies, así como los horarios del día en los que se realizarán estos relevamientos. Del mismo modo es recomendable tener información sobre el tipo de hábitat que frecuentan o utilizan las especies, a fin de incluir todos los ambientes posibles o existentes en la unidad experimental.

Bibliografía

1. Balzarini, M.; Di Rienzo, J.; Tablada, M.; Gonzalez, L.; Bruno, C.; Córdoba, M.; Robledo, W. & Casanoves, F. 2015. Estadística y Biometría: Ilustraciones del Uso de Infostat en Problemas de Agronomía. - 2a ed. - Brujas, Universidad Nacional de Córdoba, Córdoba .
2. Boitani, L. & Fuller, T. (eds.). 2000. Research Techniques in Animal Ecology: Controversies and Consequences. Columbia University Press, Nueva York.
3. Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. *Ecology* 74: 1646-1658.
4. Farji-Brener, A.G. 2003. Uso correcto, parcial e incorrecto de los términos "hipótesis" y "predicciones" en ecología. *Ecología Austral* 13: 223-227.
5. Gallego, R.S. 2003. Introducción al análisis de datos experimentales: tratamiento de datos en bioensayos (Vol. 4). Publicacions de la Universitat Jaume I, España.
6. Heyer, W.R.; Donnelly, M.A.; McDiarmid, R.W.; Hayek, L.C. & Foster, M.S. 1994. Measuring and Monitoring Biological Diversity: Standard Methods for Amphibians. Smithsonian Institution Press, Washington, DC.
7. Hinkelmann, K. & Kempthorne, O. 1994. Design and Analysis of Experiments (Vol. 1). Wiley, New York.
8. Kuehl, R.O. & Kuehl, R.O. 2000. Design of Experiments: Statistical Principles of Research Design and Analysis. Duxbury Press, California.
9. Krebs, C.J. 2010. Case Studies and Ecological Understanding. *En*: Billick, I. & Brice M. (eds.). The Ecology of Place: Contributions of Place-based Research to Ecological Understanding. University of Chigado Press. Chicago.
10. Li, Z. & Berger, V.W. 2005. Adaptive Sampling. Encyclopedia of Statistics in Behavioral Science. Wiley Online Library, Londres.
11. Magurran, A.E. & McGill, B.J. (eds.). 2011. Biological Diversity: Frontiers in Measurement and Assessment. Oxford University Press, Oxford.
12. Martínez Abraín, A.; Conesa, D. & Oro, D. 2008. Herramientas estadísticas para resolver contrastes de hipótesis con contenido biológico: su uso en ecología del siglo XXI. *Acta Zoológica* 24: 201-220.
13. Smith, A.N.; Anderson, M.J. & Pawley, M.D. 2017. Could ecologists be more random? Straightforward alternatives to haphazard spatial sampling. *Ecography* 40: 1251-1255.
14. Feinsinger, P. 2013. Metodologías de investigación en ecología aplicada y básica: ¿Cuál estoy siguiendo, y por qué? *Revista Chilena de Historia Natural* 86: 385-402.
15. Rao, P.S. 2000. Sampling Methodologies with Applications. Chapman and Hall/CRC, Londres.
16. Thompson, K.S. 2012. Sampling. Simon Fraser University, Third Edition John Wiley & Sons. Inc., Hoboken, New Jersey.
17. Krebs, C.J. 2014. Ecological Methodology, 3rd ed. Addison-Wesley Educational Publishers, Inc., Massachusetts.
18. Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54: 187-211.