

Diplomado en Análisis de Información Geoespacial

Tipos de muestreo



Autor:
M. en G. Alberto Porras Velázquez

Introducción

El análisis de un fenómeno, desde la perspectiva estadística, requiere del acopio de información o datos científicos de la población que se estudia. Sin embargo, recaudar los datos de una población no siempre es fácil, debido al elevado costo de un censo, a la dificultad que implica llegar a lugares determinados o al simple hecho de que la población en cuestión no tenga un tamaño finito.

La alternativa a esta problemática consiste en utilizar una muestra, es decir, un subconjunto de la población, para obtener resultados que nos ayuden a inferir, a partir de los estadísticos muestrales, los parámetros de la población.

El reto de este ejercicio está en obtener una muestra representativa de la población con un tamaño suficiente, de manera que al hacer la inferencia de los parámetros de la población obtengamos una estimación no sesgada y precisa.

Dicha muestra parte de un conjunto de supuestos y condiciones iniciales que determinan el problema de estudio.

Experimentos de medición

En un experimento de medición no se pueden manipular las condiciones que influyen en un fenómeno y, por lo tanto, se muestrea una situación existente.

Para generar una hipótesis, a menudo se requiere del muestreo en diferentes grupos o ubicaciones. A los investigadores, por ejemplo, en ocasiones no les queda más opción que hacer muestreos en lugares y tiempos diferentes.

Por ello las conclusiones obtenidas deben manejarse con cuidado, ya que de manera inadvertida se pueden incluir variables que influyen en el fenómeno. Un ejemplo es la medición de ciertas variables en el mar, que están sujeta a cambios estacionales.

En los experimentos de medición es necesario contar con muestras independientes, es decir, se necesita repetir el muestreo en varias ocasiones y en distintos lugares.

El muestreo de una misma zona en años diferentes fortalecerá nuestros hallazgos y será suficiente si nos interesa sólo el espacio en cuestión; mientras que el muestreo en un número mayor de áreas arrojará resultados adecuados para hacer generalizaciones.

Experimentos manipulados

En los experimentos manipulados se pueden cambiar las condiciones que influyen en un fenómeno o sistema bajo estudio, de modo que es posible observar el efecto del cambio sobre la unidad experimental.

En estos casos se requieren réplicas independientes y varios tratamientos de control para aislar la razón por la cual ocurre algo en un experimento, a partir de la comparación de dos tratamientos que difieren por **un factor**.

Saber que tratamientos de control requiere un experimento es difícil. Una forma de clarificarlo implica elaborar un listado de todo lo que ocurre en un tratamiento experimental y asegurarse de tener un control para cada situación. Los tratamientos experimentales necesitan ser replicados apropiadamente, no es posible tener sólo una réplica por tratamiento.

Factores para la elección de muestras

Una muestra es cualquier conjunto de n unidades tomadas a partir de una población, tiene que ser representativa y sus características deben reflejar las de la población. Por otra parte, la población se debe definir en términos de a) unidades, b) elementos, c) áreas y d) periodos de tiempo.

La elección de muestras requiere definir la región de interés: colonias, distritos electorales o calles, tarea que se dificulta en el caso de contar con poblaciones móviles. También implica especificar las unidades de muestreo, ciudades, calles u hogares, así como elegir el tamaño y la forma de los sitios muestreados que van a constituir el soporte de la muestra.

Además se debe determinar el método de muestreo: sistemático, aleatorio, estratificado o por conglomerados, y establecer el tamaño de la muestra, es decir, el número de elementos seleccionados.

Por último, es necesario especificar un plan de muestreo y método para coleccionar los datos.

Muestreo aleatorio

El muestreo aleatorio es una técnica que permite obtener una muestra representativa de la población. Ésta se basa en el concepto de probabilidad, el cual marca que cualquier elemento de la población tiene la misma probabilidad de ser elegido y que la elección de cada elemento es independiente de cualquier selección previa.

Los muestreos probabilísticos son de varios tipos, los cuales se describen a continuación.

- **Muestreo aleatorio simple.**

Cada elemento de la población tiene una oportunidad igual e independiente de ser seleccionado, para lo cual se le asigna un número. La muestra se determina con tablas de números aleatorios.

- **Muestreo sistemático.**

Se selecciona una muestra tomando cada k-ésima unidad de la población, una vez que las unidades de muestreo están numeradas o arregladas de alguna forma. La letra k es la razón de muestreo, esto es, la razón del tamaño de la población correspondiente al tamaño de la muestra.

Ejemplo. Al seleccionar una muestra de 40 unidades de entre una población de 1,000, entonces k es $1000/40=25$, por lo que la muestra se obtiene tomando cada 25-ésima unidad de la población.

- **Muestreo estratificado.**

Implica dividir a la población en clases o grupos, denominados estratos. Las unidades incluidas en cada estrato deben ser relativamente homogéneas con respecto a las características a estudiar. Por ejemplo, se puede realizar un estudio para conocer la opinión de la población sobre un político determinado planteando una muestra estratificada por edades.

Como parte de este muestreo, se toma una submuestra a partir de cada estrato mediante un procedimiento aleatorio simple, de modo que para obtener la muestra general se combinan las submuestras de todos los estratos.

A menudo se toma una razón de muestreo igual para todos los estratos. Es decir que las unidades en la muestra se asignan entre los estratos y en proporción con el número

relativo de unidades en cada estrato de la población. A una muestra seleccionada de esta forma se le denomina muestra estratificada proporcional.

Otras veces la razón de muestreo en cada estrato está inversamente relacionada con la homogeneidad de las unidades del estrato; mientras más homogéneo sea el estrato, menor será su proporción incluida en la muestra. A una muestra obtenida de esta forma se le denomina muestra estratificada desproporcionada.

Cuando las unidades en un estrato son más homogéneas, resulta necesaria una submuestra más pequeña para asegurar la representatividad. En consecuencia, el costo de muestreo se reduce.

- **Muestreo por conglomerados.**

Implica la selección aleatoria de grupos o conglomerados a partir de la población. Las diferencias entre los conglomerados son generalmente pequeñas y las unidades dentro de cada uno, normalmente, son más heterogéneas.

Por ejemplo, las ciudades con una población de 50 habitantes pueden tener características similares si las comparamos, pero la población es heterogénea. De esta manera el analizar un conjunto de ciudades de 50 mil habitantes nos ayudaría para tratar de caracterizar a la población de todas las ciudades de 50 mil habitantes del país.

Muestreos no probabilísticos

En ocasiones, por distintas razones, se realizan muestreos que no se basan en criterios probabilísticos. A veces la economía limita la obtención de una muestra bajo criterios aleatorios. En tales casos se busca la representatividad de la población tomando en cuenta las restricciones que imponen la economía y la factibilidad del muestreo.

Muestreo espacial

El enfoque del muestreo espacial depende de la naturaleza del problema de estudio. En algunos casos, las muestras son utilizadas para estimar las condiciones promedio de un área; por ejemplo, la temperatura de una región del país.

En otros casos lo importante puede ser la caracterización de varias regiones, ya sea por su tipo de suelo o vegetación. Otro objetivo de estudio puede ser la caracterización del patrón espacial de dependencia en un área, de modo que la ubicación de las muestras deberá ser tal que se obtengan estimaciones confiables de similitud en cada rango de distancias.

Cuando se analizan las muestras espaciales se debe tomar en cuenta su tamaño, representatividad y sesgo, la influencia de los factores temporales y de los efectos de borde, así como sus consecuencias para el análisis; además de determinar si los datos son agregados, entre otros elementos. (de Smith, et al. 2009)

La consideración de estos factores debe conducir a un plan de muestreo que coincida con los ya vistos, es decir, el muestreo aleatorio simple, sistemático, estratificado y por conglomerados. Cabe señalar que el muestreo puede tomar la forma de puntos o cuadrantes.

- **Muestreo aleatorio simple.**

En un muestreo aleatorio simple se seleccionan n sitios, de manera que cada miembro de una población tiene la misma e independiente probabilidad de ser elegido.

Esta modalidad no garantiza que se cubra toda el área de estudio, pues superficies relativamente grandes se quedan sin muestrear o bien las muestras aparentan estar agrupadas (Figura 1).

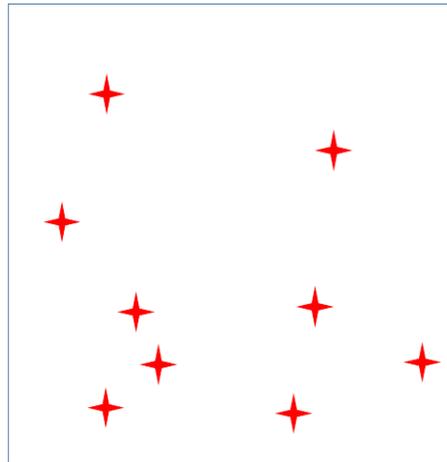


Figura 1. Muestreo aleatorio simple. Porras, A. (2014). Imagen basada en Haining R.P. (2003), p. 101

- **Muestreo aleatorio estratificado.**

La población a partir de la cual se elige una muestra es particionada en estratos de área, para los cuales se seleccionan los sitios de muestreo mediante un procedimiento aleatorio simple (Figura 2).

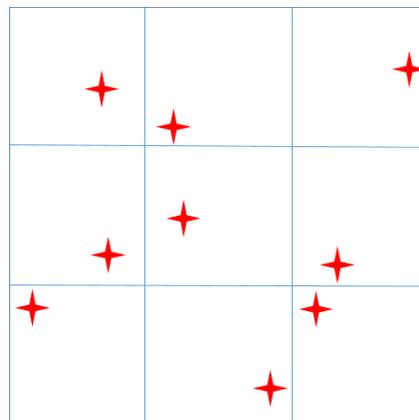


Figura 2. Muestreo aleatorio estratificado. Porras, A. (2014). Imagen basada en Haining R.P. (2003), p. 101

Con el muestreo aleatorio estratificado disminuye la posibilidad de que existan zonas sin muestras o zonas con alta concentración de muestras.

Es pertinente mencionar que en el muestreo espacial los individuos que pertenecen a un estrato tienen como elemento común un espacio delimitado, por lo cual no son necesarios los valores homogéneos de otras variables.

- **Muestreo sistemático.**

En este caso se elige de manera aleatoria el sitio de muestreo para un estrato determinado, mientras que en los estratos restantes el sitio de muestreo se determina a partir de la misma posición relativa (Figura 3). Un caso típico de esto consiste en elegir el centro de cada estrato como sitio de la muestra.

Tanto en el muestreo sistemático como en el estratificado, la forma que adoptan los estratos no está restringida al rectángulo, ésta puede ser cualquiera, triangular o hexagonal, por ejemplo. En los estudios sociales y económicos, la elección de los estratos suele darse de manera que coincida con regiones administrativas, por lo que es posible realizar comparaciones entre ellas.

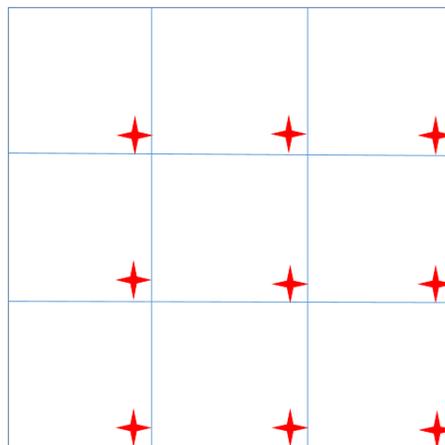


Figura 3. Muestreo sistemático. Porras, A. (2014). Imagen basada en Haining R.P. (2003), p. 101

El muestreo sistemático es sencillo de implementar y brinda un número suficiente de observaciones separadas con distancias y direcciones bien definidas.

Cuando la población de estudio suele estar agrupada en conglomerados, como en el caso de una ciudad, el muestreo se enfoca en los estratos que agrupan al grueso de la población. Una vez seleccionados estos estratos, se aplica un criterio aleatorio para establecer los sitios o elementos que conformarán la muestra. Este es el equivalente a un muestreo por conglomerados.

- **Muestreo anidado.**

En muchos estudios, por ejemplo, cuando se utiliza la geoestadística para realizar interpolaciones sobre una superficie, suele ser de interés la variación de la población en distintas escalas espaciales.

El muestreo anidado requiere que la población se divida en bloques que se subdividen en otros más finos, hasta alcanzar el nivel de detalle que se desee. En cada nivel, los bloques se anidan en bloques en un nivel superior.

El propósito de la anidación es captar el comportamiento distintos niveles de escala. Un método para conseguir este objetivo implica establecer la distancia fija de separación entre las muestras en un primer nivel.

Posteriormente se eligen los sitios de segundo nivel, de manera que permanezcan separados a una distancia fija de los sitios de primer nivel, aunque con una dirección seleccionada al azar (Haining, 2003).

En la figura 4 se ejemplifica un caso de muestreo anidado, en el cual las muestras de primer nivel son rojas y las de segundo nivel, verdes.

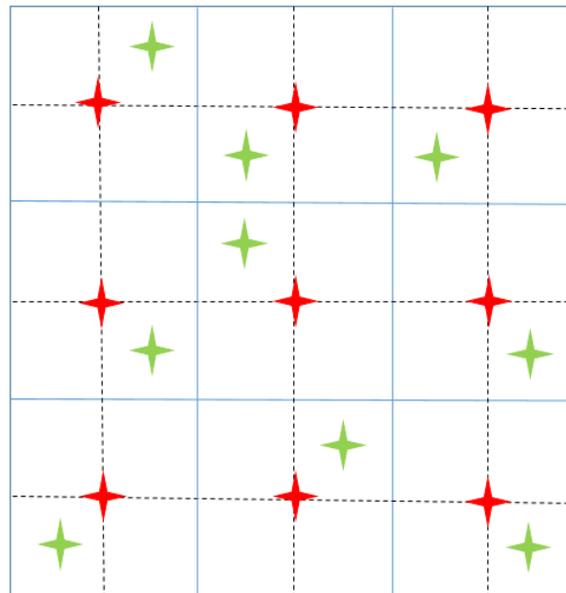


Figura 4. Muestreo anidado. Porras, A. (2014).

La falacia ecológica y el problema de unidad de área modificable.

Uno de los elementos más importantes en el análisis espacial, y que merece la mayor consideración, es la escala espacial. Si no hay coincidencia entre la escala de nuestras mediciones y la del fenómeno de estudio, seguramente habrá problemas en el análisis. Por ejemplo, si un fenómeno opera en una micro escala (el comportamiento individual de las personas) y las mediciones con que contamos están en un nivel agregado (municipal o estatal), no será posible obtener información de los ciudadanos, sobre todo si la población de estudio es heterogénea en la micro escala.

Siempre que se diseñe un plan de muestreo se debe de considerar el problema de escala. Por ejemplo, si nuestras observaciones tienen una separación mínima de 100 metros entre sí, no será posible inferir el comportamiento espacial de una variable a mayor detalle.

Al error que se produce por tratar de obtener conclusiones de un comportamiento que se encuentra en un nivel de agregación menor (sobre todo si hay heterogeneidad en la población a un micro nivel) a partir del comportamiento de datos con un nivel de agregación mayor se le llama falacia ecológica.

Otro aspecto relevante relacionado con el nivel de agregación de los datos es la zonificación, es decir, la elección del sitio en el que se agregarán los datos.

En relación con este punto, se debe cuidar la escala del análisis y el procedimiento para alinear las unidades espaciales.

Si hay heterogeneidad en la micro escala, los resultados diferirán en el caso de zonificaciones distintas. Esto se debe a que cada zonificación agrupa individuos diferentes.

Por lo tanto, un muestreo no necesariamente será representativo de una escala en particular. El hecho de tener un individuo vinculado a una unidad de área en particular no significa que sea representativo de esta área.

Un caso muy conocido de la manipulación de las unidades de área para obtener resultados específicos se presentó a principios del siglo XIX en el estado de Massachusetts, cuando el gobernador Elbridge Gerry alteró los límites de los distritos electorales con el propósito de obtener ventaja electoral para su partido.

La configuración de los distritos resultó tan extraña que se acuñó el término *Gerrymander*, juego de palabras construido con el apellido del político y la palabra *mander* (salamandra, en inglés). (Wikipedia, 2014).



Ilustración 1. Tisdale, E. (1812). TheGerryMander [Caricatura]. Recuperado de http://es.wikipedia.org/wiki/Gerrymandering#mediaviewer/File:The_Gerry-Mander_Edit.png

Referencias

- Chao, Lincoln L. (2006). *Introducción a la estadística*. México: Cecsca.
- de Smith, M. J., Goodchild, M.G. & Longley P.A. (2009). *Geospatial Analysis: A comprehensive Guide to Principles, Techniques and Software Tools*. (3a ed.), Leicester, Inglaterra: Troubador Publishing Ltd.
- Haining R. P., (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge, Reino Unido: Cambridge University Press.
- McKillup, S. & Darby M. (2010). *Geostatistics Explained. An Introductory Guide for Earth Scientists*, Cambridge, Reino Unido: Cambridge University Press.
- Walpole, R. E.; Myers, R. H.; Myers, S. L. & Ye, K. (2007). *Probabilidad y estadística para ingeniería y ciencias, (8ª ed.)* México: Pearson Educación.
- Wikipedia. (2014). *Gerrymandering*. Recuperado de <http://es.wikipedia.org/w/index.php?title=Gerrymandering&oldid=75968359>.