

Uso de suelo agrícola, una aplicación de técnicas estadísticas geográficas.¹

Agricultural soil use, an application of geographical statistic techniques.¹

Fernando Cabrales G.² y Pilar Mazuela A.³

RESUMEN

Este estudio presenta una metodología estadística diseñada para determinar el uso de suelos agrícolas, basada en la construcción de una grilla regular, de la cual se extraen aleatoriamente unidades para análisis.

Se explicitan los requerimientos estadísticos de un estudio areal, las características estadísticas del procedimiento y las limitaciones para la realización de estudios espaciales y multivariados.

ABSTRACT

This study presents a statistical methodology designed to determine the use of agricultural soil, based on the construction of a regular grid from which to extract random units for aleatory analysis.

Statistical requirements of an areal study are explained as well as the statistic characteristics of the procedure and limitations to carry out spatial and multivariate studies.

1 Este trabajo forma parte del proyecto de Investigación Menor UTA N° 9705-00, contó con el apoyo financiero del proyecto de Internacionalización UTA-MECESUP N° 9803 y del proyecto de Mejoramiento de la Calidad de la Educación UTA-MECESUP N° 9901 de la misma Universidad y con el apoyo técnico de la Corporación Norte Grande.

2 Economista; Académico FACSAAE, Universidad de Tarapacá.

3 Ingeniero Agrónomo, Académico FAGRON, Universidad de Tarapacá.

INTRODUCCIÓN

Si se considera que las necesidades de información resultan determinantes para el diseño de los sistemas de medición empleados, resulta claro que los actuales sistemas para medir el uso de suelos en nuestro país no son los más apropiados.

En este trabajo se expone una metodología de muestreo areal aleatorio que permitirá realizar estimaciones insesgadas y eficientes del uso de suelos.

Del mismo modo, se exponen aspectos metodológicos de su aplicación práctica al estudio de uso de suelo en el Valle de Azapa en Arica, lo que permite comparaciones de costo y eficiencia con los métodos tradicionales.

Se exponen las limitaciones que esta metodología tiene para el estudio e inferencia de modelos causales o multivariados de las características observadas del terreno en estudio.

Finalmente se formulan algunas recomendaciones que permitirían mejorar el desempeño de la metodología propuesta.

1- PROBABILIDADES Y ESTIMACIONES AREALES

El problema es el diseño experimental en el sentido estadístico y podemos hablar de 4 tipos de diseños básicos experimentales:

- 1- Muestreo aleatorio: estrictamente hablando, el único realmente aleatorio y en éste, cada "sujeto" de la población en estudio tiene la misma probabilidad de resultar escogido.
- 2- Muestreo estratificado: En los cuales los componentes de la población en estudio son agrupados en categorías arbitrarias, las cuales, a su vez, son muestreadas.
- 3- Muestreo aleatorio sistemático: En el cual los sujetos son ordenados en algún sentido y sólo un elemento es seleccionado aleatoriamente y los demás se seleccionan a partir de una regla prefijada (a intervalos fijos u otra).
- 4- Muestreo aleatorio de "cluster": En los cuales los sujetos son agrupados en sub-poblaciones con alguna coherencia interna, y éstas agrupaciones o "cluster" son elegidas aleatoriamente y estudiadas completamente.

En la práctica, la mayor parte de los estudios estadísticos suelen utilizar combinaciones de estos diseños básicos y es también el caso de los estudios geográficos.⁴

En el caso nuestro, interesa determinar un diseño que permita obtener antecedentes sobre la utilización de los suelos de un área geográfica pues ése es un antecedente de primera importancia para la evaluación de la eficiencia productiva de un valle.

El problema es de especial relevancia en zonas áridas y desde el punto de vista de las políticas públicas para el sector, pues cuando la superficie apta para cultivos es especialmente escasa adquiere relevancia el uso eficiente, a nivel global de éstos.

La aplicación de las estadísticas a problemas geográficos tiene algunas especificidades dadas por el carácter multidimensional de las variables geográficas.

En efecto el "sujeto" estadístico no es un punto ni un valor determinado, en muchos casos se trata de un área (hablaremos de estudios areales) o bien de un "objeto" tridimensional (hablaremos de estudios espaciales).

De hecho, una primera precaución a considerar es que, en nuestra opinión, el estudio de uso de suelos es en realidad un problema "espacial", pues en él inciden aspectos como la calidad del suelo, la cercanía a cursos de agua, de caminos y la topografía.⁵

Este tipo de modelos tiene sus primeros antecedentes en el trabajo de Thünen (1966) y algunas variantes teóricas se pueden encontrar en Amsberg (1994), Schneider (1995) y Hyde, Amacher y Magrath (1993).

Sin embargo estos modelos son "explicativos" de lo observado, es decir suponen un orden causal y el trabajo se concentra en la calidad de los ajustes observados y en algunos casos en algunos intentos predictivos⁶.

En este trabajo nos encontramos una etapa antes; es decir intentamos encontrar una metodología de bajo costo, insesgada y eficiente para caracterizar el uso de los suelos que posteriormente puede explicarse con modelos espaciales.

De esa manera la metodología que se propone es sólo descriptiva y por lo tanto areal.

Adicionalmente se estima necesario considerar preferible el uso de muestreo aleatorio pues la mayor parte de los métodos de inferencia son válidos en este tipo de diseños experimentales.

4 Un buen ejemplo de combinación de diseños es la que describe en INE (1998).

5 Un modelo espacial de usos de suelos, con técnicas satelitales avanzadas puede encontrarse en Chomitz K. y Gray D. (1996).

6 El trabajo de la nota anterior es un caso "causal" multivariado, que utiliza una grilla de un km por lado para el caso de Bélice.

Si es necesario ocupar otro de los métodos o combinaciones de ellos es mejor hacer un análisis muy detallado de las derivaciones estadísticas del considerado.

2- DISEÑO EXPERIMENTAL

El muestreo de variables geográficas requiere la selección aleatoria de áreas coordenadas y ello se ha hecho de la siguiente manera.

En primer lugar se debe acotar en un mapa el área de interés y construir una grilla regular sobre esa área.

Dicha grilla debe ser fácil de ubicar y reproducir en terreno para poder examinar las características del área seleccionada aleatoriamente.

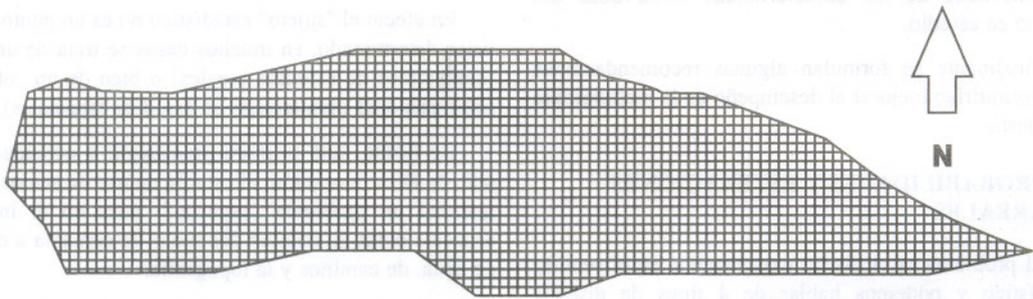
Eso requiere del uso de una carta con coordenadas geográficas que puedan ubicarse mediante un Posicionador Geográfico (GPS) convencional.

Debe determinarse la forma de la grilla y la superficie de ésta, es decir debe determinarse un polígono unitario que será la unidad de estudio.

En nuestro caso hemos propuesto una grilla de cuadrados de 100x100 mts que permite que toda la superficie del valle tenga la misma probabilidad de ser seleccionada, el polígono resulta fácil de reproducir en terreno y la superficie tiene importancia desde el punto de vista experimental. Esquemáticamente la idea es como se ilustra en la figura N° 1.

Figura N° 1

Grilla regular con orientación Norte-Sur



Es evidente que una superficie circular de radio cercano a 56 mts. podría resultar más fácil de ubicar en terreno, pero no toda la superficie del valle habría tenido la misma posibilidad de resultar seleccionada.⁷

La grilla, a efectos de ubicar las hectáreas seleccionadas aleatoriamente en terreno, debe tener orientación magnética de manera que baste con ubicar un punto del cuadrado seleccionado para, con apoyo de una brújula, ubicar el cuadrado completo.

Los cuadrados pueden numerarse y seleccionarse para estudio en forma aleatoria.

En el caso del valle de Azapa, tendremos alrededor de 3.000 "cuadrados" mutuamente excluyentes, que cubren toda la superficie sujeto de análisis en el valle.

Sabemos que las características X_i que encontremos en los cuadrados serán estimadores insesgados y eficientes de los "verdaderos" valores existentes en el universo muestral, el Valle.

Las características (X_i) que podremos observar en estas hectáreas aleatorias es el uso de los suelos, tales

como cultivos, praderas, usos diversos no agrícolas (camino, carretera, viviendas) y otras no directamente observables tales como pertenencia de la tierra a pequeñas medianas o grandes propiedades, tecnificación de los cultivos o características técnicas de los suelos.

En cada caso podremos obtener una estimación puntual (una media \bar{X}) y una medida de dispersión (la varianza muestral S^2) que son estimadores "BLUE" de la proporción verdadera μ de la característica observada y de la varianza, σ^2 , verdadera existente en el valle.

Algunos de los usos observados tienen superficies conocidas, es el caso de los caminos, y otros cuyos antecedentes pueden hacerse disponibles a costos razonables, como podría ser el caso de los predios mayores de 20 hectáreas que son muy escasos en el Valle de Azapa.

Estas magnitudes pueden utilizarse como variables de control adicionales pues si el procedimiento está bien ejecutado, los valores obtenidos para caminos y predios mayores deben encontrarse en el intervalo de confianza del valor proyectado en el muestreo.

⁷ Ello podría ser objeto de un análisis adicional si se considera la posibilidad de selección aleatoria de un cluster circular.

Adicionalmente habrán hectáreas de la grilla que estarán sólo parcialmente dentro del área de interés y ese es otro valor conocido en el diseño que permite controlar la calidad de la muestra obtenida.

En las condiciones anteriores podemos afirmar que las características observadas \bar{X}_i distribuirán normal, por teorema del límite central, con media μ_i y desviación estándar σ/\sqrt{n} .

De esa manera se pueden calcular intervalos de confianza estandarizados convencionales de manera que nuestras estimaciones se encontrarán en el intervalo

$$\bar{X} \pm (\sigma/\sqrt{n}) \times Z \quad (1)$$

donde Z es el número de veces que multiplica la varianza de una normal estándar de varianza unitaria, con una probabilidad acumulada conocida (o exigida).

2.1- Tamaño de Muestra

El tamaño de muestra en este tipo de estudios está dado en numerosas ocasiones por limitaciones económicas, sin embargo es posible diseñar un experimento cuyo tamaño de muestra sea tal que nos proporcione estimaciones con un nivel de confianza pre determinado.

En efecto, si consideramos que nuestras estimaciones estarán definidas de la manera indicada antes, es decir:

$$\bar{X} \pm (\sigma/\sqrt{n}) \times Z \quad (1)$$

Entonces considerando el intervalo dado sabemos que

$$n = \left(\frac{(\sigma \times Z)}{\text{Intervalo}} \right)^2 \quad (2)$$

De manera que si conocemos el intervalo en que se moverá el verdadero valor, podemos establecer un tamaño mínimo de muestra.

Habitualmente ese no es el caso, de manera que se hace necesario aproximarse a ese intervalo mediante estudios piloto o experiencias previas.

Adicionalmente y considerando el tamaño del universo, es posible aproximarse al "n" óptimo a través de la probabilidad de ocurrencia del fenómeno observado, en cuyo caso la expresión para n será:

$$n = \frac{Z^2 \times (P \times Z(1-P))}{K^2} \quad (3)$$

Donde,

P = Porcentaje de la población que posee las características de interés.

K = Confiabilidad exigida a la proporción muestral (5%)

En este caso, podemos utilizar antecedentes previos (censos anteriores, encuestas, etc.) para aproximarnos a la probabilidad de ocurrencia de la característica en estudio.

3- UNA APLICACIÓN AL CASO DEL VALLE DE AZAPA

Si nos interesara saber la cantidad de hectáreas dedicadas al cultivo del tomate en el Valle de Azapa, siguiendo el último criterio de la sección anterior y en base a datos aproximados de superficie (aproximadamente 3.000 hectáreas) y de la cantidad de hectáreas detectadas en el último censo agropecuario dedicadas al tomate (aproximadamente unas 450 hectáreas), tendríamos que, a un nivel de confiabilidad del 95%:

$$n = \frac{(1,96) \times (0,15) \times (0,85)^2}{(0,05)^2}$$

es decir

$$n = 195,9 \rightarrow 196$$

Sin embargo al realizar el estudio obtendremos valores para \bar{X} y para S_2 de manera que obtendremos una nueva estimación de la superficie efectivamente ocupada en el cultivo de tomate que involucrará una nueva confiabilidad para la predicción.

Sólo a modo de ejemplo; si obtenemos una predicción de 400 hectáreas plantadas con tomates y una desviación estándar de 0,2 hectáreas, entonces nuestra predicción al 95% de confianza estará entre 393,74 y 406,25 hectáreas dedicadas a dicho cultivo.

Los valores aproximados son indicativos del trabajo que sería necesario realizar y que resultaría, en nuestra opinión, mucho más económico que un censo y notablemente más preciso que una encuesta.

4- PRECAUCIONES ADICIONALES

El método descrito se constituye en una fotografía del uso de suelos en un instante de tiempo pero no está

diseñado para estudios de modelos de comportamiento que, como decíamos, suelen ser modelos espaciales, es decir más complejos que lo que puede describir un análisis areal como el descrito.

Las cifras tampoco son estrictamente análogas con las que entrega, por ejemplo, la encuesta hortícola pues ésta incorpora al uso de suelos todos los cultivos que se hagan en un mismo terreno a lo largo de un año.

Los costos y ventajas técnicas de un estudio del tipo que presentamos, sin embargo, permitirían realizar varias repeticiones del estudio a lo largo de un año, de manera que sería posible obtener una serie de tiempo con la ocupación efectiva de un valle.

Esta metodología puede aún mejorarse estudiando en detalle las características estadísticas de otros tipos de polígonos que pudieran utilizarse.

Además es posible implementar procesos estratificados para incorporar nuevas dimensiones al análisis, tales como lo propuesto por Berry y Baker (1968) a veces llamado estratificación sistemática con alineamiento aleatorio y que permitiría inferir algunas conclusiones respecto a variables no capturadas por el método aleatorio, como el efecto de la distancia a carreteras en el uso de suelos.

5- CONCLUSIONES

La metodología propuesta aún debe pasar por una etapa de prueba en terreno, pero el análisis teórico realizado hasta aquí indica que ésta tiene las características deseables (independencia de las observaciones, equiprobabilidad de selección de la muestra, tamaño muestral grande) desde el punto de vista estadístico y ello permite hacer inferencia, determinar los límites de confiabilidad y las probabilidades de errores de diverso tipo.

No menos importante, la metodología parece exigir costos significativamente menores que los necesarios para una encuesta o para un censo y entrega antecedentes relevantes y controlables para determinar el uso de suelos.

La metodología es de base areal y de ello se deriva su principal limitación pues la elaboración de modelos y testeo de hipótesis multivariadas enfrenta restricciones que consideramos importantes.

LITERATURA CITADA

- BERRY, B.J. Y BAKER A.M. "Geographic Sampling" en B.J.L. Berry and D.F. Marble (eds.), *Spatial Analysis: A Reader in Statistical Geography*; Prentice Hall: Englewood Cliffs: N.J., 1968.
- BLALOCK H.M. *Social Statistics, Revised Second Edition*, Mc Graw-Hill, New York, 1979.
- CHOMITZ K. Y GRAY D. "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belice"; 1996, *The World Bank Economic Review*, Vol. 10, N° 3, Sept. 1996.
- CLARK, W.A.V. *Statistical Methods for Geographers*. John Wiley & Sons, Inc, 1986.
- HYDE W., AMACHER G. Y MAGRATH W. (1993) "Deforestation, Scarce Forest Resources, and Forest Land Use: Thoery, Empirical Evidence and Policy implications" Documento presentado a la 4ª Conferencia Anual de la Asociación para el Estudio de la Propiedad Común.; Manila, Julio 1993.
- INSTITUTO NACIONAL DE ESTADÍSTICAS (INE) "Encuesta Hortícola 1995-96", INE, Santiago; 1996.
- INSTITUTO NACIONAL DE ESTADÍSTICAS (INE) "Encuesta Hortícola 1997-98", INE, Santiago; 1998
- INSTITUTO NACIONAL DE ESTADÍSTICAS (INE) "VI Censo Nacional Agropecuario Región Tarapacá", INE, Iquique, 1999.
- SCHNEIDER R. 1995 "Government and the Economic of The Amazon Frontier", *World Bank Environment Paper 11*. Washington D.C.
- VON AMSBERG, J. 1994 "Economic Parameters of Deforestation" *Policy Research Department Working Paper 1350*. World Bank, Policy Research Department, Washington D.C. mimeo.
- VON THÜNEN, J. H. 1966, "Der Isolierte Staat in Beziehung der Landwirtschaft und Nationalökonomie" en Peter Hall ed. *Von Trünen's Isolated State*, Oxford, Engl. Pergamon Press.