



**POSDEM:
SELECCIÓN ENTRE PLANES DE MUESTREO PROBABILÍSTICO
de Gonzalo Sánchez-Crespo y Alberto Lezcano**

Manuel Mateo
Instituto Galego de Estadística

Descripción técnica del programa

Su objetivo es elegir entre planes de muestreo adaptados a cada investigación, incorporando un conocimiento experto sobre muestreo, que evite que la persona que lleva a cabo la investigación necesariamente tenga que tener estos conocimientos. Al mismo tiempo el fichero de ayuda del programa tiene un planteamiento didáctico de manual de muestreo.

Posdem es un programa informático, diseñado para optimizar la elección de las unidades que forman parte de la muestra en el diseño de una encuesta por muestreo probabilístico de acuerdo con ciertos criterios. Utiliza diferentes métodos de selección y poblaciones definidas por el usuario.

Permite determinar qué plan de muestreo es preferible para una determinada estructura poblacional y cuáles son las relaciones entre la estructura poblacional y los procedimientos de muestreo utilizados. Debemos decir que por estructura poblacional nos referiremos a los parámetros que definen una población natural dentro del ámbito de los modelos de superpoblación. Para aumentar la robustez de los resultados, con el modelo de superpoblación, se establece que el valor de la variable en estudio es función de un modelo aleatorio determinado por un ajuste mediante polinomios ortogonales de los datos originales. Se pretenden evitar situaciones anecdóticas o resultados asociados a valores anómalos del marco original de la encuesta. De esta forma cada una unidad de la población es la realización de una variable aleatoria cuya estructura estocástica se ha especificado mediante el mismo programa.

Para realizar dicha elección Posdem compara 44 métodos de muestreo diferentes que van desde el muestreo simple con reposición hasta diseños más complejos como muestreo con probabilidades desiguales y reposición parcial pasando por diferentes tipos de muestreo sistemático y muestreo estratificado de acuerdo con una variable

auxiliar perfectamente identificable en todo el proceso de elección del procedimiento muestral.

Dado un marco de unidades y una vez elegido el método de muestreo que mejor se ajusta a esa estructura poblacional es posible determinar qué unidades de la población pertenecerán finalmente a la muestra que será investigada y cuáles serán visitadas en caso de falta de respuesta.

El programa analiza variables cualitativas y cuantitativas. En cuanto a tipo de estimadores permite utilizar, además de los típicos estimadores de expansión para el total, y la media en caso de variables cuantitativas o de la proporción y el total de clase para el caso de variables cualitativas, estimadores de razón utilizando los valores de la variable auxiliar.

Además Posdem permite estudiar los resultados de cada experimento en forma de listados o gráficamente. En cuanto a la estructura de población, ésta puede estar definida por una población ficticia, una población natural observada en la práctica o una superpoblación. Los modelos de superpoblación constituyen uno de los puntos fuertes del programa puesto que permiten definir modelos muy complejos. Además debemos reseñar que existen variantes para el cálculo de los errores debidos al muestreo como es el caso de que el orden de la población sea aleatorio o, si el orden es significativo, utilizar la técnica del lazo.

Funcionamiento

Lo primero que debemos hacer con Posdem es cargar los datos. Esto lo haremos desde el editor de datos, bien tecleándolos manualmente, o bien importándolos si están disponibles en un fichero que puede ser base de datos MS Access, Dbase IV o texto por citar sólo los más usuales. Tendremos que definir la variable de estudio y, si existe, la variable auxiliar e incluso otras variables. Una aplicación bastante interesante es la posible redefinición de alguna de las variables mediante una función como por ejemplo la logarítmica o definir una nueva variable como suma o cociente de dos. Otra característica beneficiosa del programa es que ofrece la posibilidad de ordenar los datos bien por la variable de estudio bien por la variable auxiliar o incluso por una variable exógena.

Otra propiedad interesante disponible es la visualización de los datos de la variable auxiliar y la variable de estudio mediante una nube de puntos, bien separadamente, bien conjuntamente. Si escogemos esta última opción nos permitirá comprobar mediante ese gráfico la relación entre ambas variables y su posible dependencia. Recordemos que si existe esa dependencia en teoría se podría afinar más el muestreo estratificando de acuerdo con la variable auxiliar o incluso utilizar estimadores de razón.

Siguiendo en el editor de datos, nuestra primera opción puede ser, obtener una estimación del tamaño de muestra a utilizar, mediante una aplicación que nos permite obtener los errores de muestreo y la cota del error para diferentes tamaños muestrales

con muestreo aleatorio simple y estratificado con una unidad. Este primer paso permite acceder a un apartado básico del programa, sobre todo si no estamos muy familiarizados con el muestreo de poblaciones finitas. Después tendremos que fijar el tamaño de muestra aunque este valor lo podremos reformar mas adelante. Además, para diferentes tipos de muestreo sistemático y tamaños de muestra podemos calcular, el coeficiente de correlación intraclásico.

Una vez cargados los datos deberemo *aceptarlos*, es decir, fijarlos y darlos por buenos aunque Posdem permite su modificación en cualquier etapa posterior. Previamente fijaremos el tamaño de muestra que luego se utilizará aunque podremos cambiarlo a posteriori. Una vez hecho esto deberíamos ir a la pantalla denominada 'generador de poblaciones' para seleccionar la opción de generación de poblaciones adecuada. La ayuda de Posdem es lo bastante buena y el programa lo suficiente flexible para que, incluso los más neófitos, no necesiten un conocimiento muy profundo del tema de superpoblaciones para continuar (si lo que se desea es analizar la población natural contenida en la base de datos que se ha cargado entonces salimos del simulador y el programa actúa por defecto).

Resuelto estos primeros pasos, Posdem está preparado para efectuar los cálculos necesarios para permitirnos elegir entre algunos de los diferentes tipos de muestreo. Una vez definidos los que deseamos comparar escogeremos en otra pantalla el sistema de evaluación que puede ser por cota de error cuadrático (lo mas frecuente) o indicador de Rao-Bayless (si estamos interesados en estudios mas especializados de análisis de la estabilidad del estimador de la varianza).

Otra opción es escoger el tipo de muestreo y ver como varía el estimador escogido en las diferentes muestras que el programa elabora, existen aquí distintas opciones para analizar los resultados. También podemos imprimir los resultados y realizar diferentes tipos de gráficos

Comentario

Mi opinión personal sobre este programa es muy favorable. Es recomendable para profesorado que deba impartir clases de muestreo pues permite evaluar diferentes tipos de selección de muestras y distinguir en qué poblaciones es preferible utilizar unos u otros métodos. Es también útil para alumnos que estén inmersos en cursos de muestreo ya que permite ver la aplicación práctica de esos conocimientos aprendidos, a veces demasiado teóricos. Además creo que es una extraordinaria introducción al muestreo para personas que nunca o casi nunca se han acercado al tema.

Los diferentes tipos de muestreo que se utilizan son más que suficientes y es destacable la utilización de estimadores de razón, los cuales, para algunas poblaciones, son los más apropiados.

Como único aspecto negativo podría argumentarse que falta algún tipo de muestreo bietápico reconociendo también que el programa sería bastante más complejo y probablemente no tan fácil de usar como es Posdem. El usuario que desee realizar un

muestreo polietápico deberá utilizar el programa por partes y añadir un conocimiento que no está reflejado en la ayuda del programa.

Otro posible inconveniente que se le puede encontrar es no utilizar el coste por unidad de muestreo para los diferentes tipos de selección de muestra evaluados por lo que su utilización práctica es limitada (recordemos que uno de los factores utilizados para la selección de la muestra en macroencuestas es el coste). El usuario de este programa debe dar el coste como variable exógena al modelo. Podríamos decir lo mismo que para el muestreo bietápico, es decir, el programa sería más complejo y, además incluir coste supondría un cambio más profundo que utilizar otro tipo de muestreo ya que habría que añadir el coste como un método de elección de la muestra.

En cuanto al propio programa en sí, cabe destacar su facilidad de uso. Es bastante interactivo permitiendo cambios una vez empezado el procedimiento de selección, además uno puede moverse con bastante rapidez entre diferentes ventanas, las tablas de resultados son muy completas y fáciles de interpretar además como ya se ha dicho repetidas veces su ayuda es muy buena tanto en cantidad como en calidad, siendo fácilmente manejable con prácticos recordatorios de definiciones y muy buenos enlaces entre las diferentes partes en que se divide esa ayuda.

También destacaríamos que los ejemplos que trae son muy fáciles de utilizar e instructivos en cuanto a aprender rápidamente el funcionamiento del programa.

En resumen un muy buen programa de muestreo cuya finalidad más obvia pero no única es el mundo de la enseñanza.

Disponibilidad del programa

Los requisitos mínimos para el programa son procesador 386 o superior, bajo Windows 3.1, 3.11 ó Windows 95. Se necesitan como mínimo 8 Megabytes de memoria RAM y 5 Megabytes libres en el disco duro o unidad de red.

El programa se puede descargar desde la siguiente dirección

<http://www.geocities.com/ResearchTriangle/Facility/1075/default.htm>

Se descargan los ficheros: Posem.zip, Actualiz.zip y Database.zip descomprimiéndolos con Winzip. Luego se ejecuta el programa Instalar.exe.

Si la versión de Windows es la 3.1 deberá añadirsele la instrucción siguiente en el archivo AUTOEXEC.BAT:SHARE.EXE/L:500

Son necesarias dos consideraciones para el buen funcionamiento del programa. En primer lugar el directorio donde se ejecute el programa se denominará C:\Posdem y en segundo lugar se debe configurar el formato numérico de Windows de forma que se utilice el punto como indicador de decimales. Para desinstalarlo lo único que se necesita es borrar los archivos del directorio C:\Posdem.