
CRITICA DE PROGRAMAS DE ORDENADOR

El Sistema Blaise 2.0

Por
CARLOS GIMENO DIEZ
Instituto Nacional de Estadística

1. INTRODUCCION

La producción de datos estadísticos es un proceso complejo en el que se distinguen, al menos, las fases de: Diseño, recogida de datos, depuración, ponderación y análisis. El sistema Blaise permite, con una mayor facilidad en las cuatro últimas, sustituir las herramientas tradicionales, por un ordenador, incidiendo especialmente en las de recogida y depuración. Su base es el lenguaje Blaise; utilizado para crear una especificación formal de los modelos de cuestionarios apropiados a la investigación. Esta especificación actúa como una base de conocimientos para la entrada de un compilador Pascal, generando un programa ejecutable.

El sistema Blaise puede producir dos clases de programa. En primer lugar un Computer Asisted Data Input (CADI), un sistema interactivo para una entrada y edición de datos. Al activar la comprobación del cuestionario, el programa detecta los errores producidos en las rutas y los errores de consistencia (los errores en los rangos se detectan en la entrada de datos). Los errores pueden ser corregidos con la asistencia de una ayuda interactiva. Una vez eliminados se escribe un registro limpio en el fichero.

En segundo lugar un Computer Asisted Personal Interviewing (CAPI/CATI), que permitirá al encuestador realizar una entrevista personal o telefónica controlada mediante el ordenador.

El sistema, en su versión 2.0, fue desarrollado por la Statistical Informatics Unit del Netherlands Central Bureau of Statistic, basado en las ideas de J.G. Bethlehem, T.J.J. Denteneer, A.J. Hundepool, L.F.M. Vermeulen y M.H. Schuerhoff, con contribuciones de J.-P. Kent y M.A. Roessingh. Durante el desarrollo A. Vogelesang prestó su asistencia en varios puntos.

2. PRESTACIONES DEL SISTEMA

Es una herramienta adaptada al diseño de cuestionarios, captación, depuración y análisis de datos de una forma automatizada, que permite un tratamiento informático de las fases de una encuesta, así como una reducción de los errores producidos en éstas, basado en las siguientes características.

2.1. El lenguaje Blaise

Es un lenguaje estructurado, no procedural que permite detectar los tres tipos de errores más frecuentes en el proceso de producción de una estadística: los errores de rango, los de consistencia y los de rutas.

Las funciones del sistema se definen con varios tipos de párrafos.

El párrafo QUEST define las cuestiones del formulario con sus características y sus rangos. Se soportan los tipos: precodificadas, numéricas, abiertas y de tipo fecha. En las primeras es posible definir cuando admitirá una sola respuesta o si se admiten varias de las posibles. Las numéricas pueden definirse como de tipo entero o real definiendo el rango sin o con decimales. Las abiertas son de tipo "string" y las de tipo fecha admiten datos en los formatos más frecuentemente utilizados.

Definidas las cuestiones del formulario el párrafo de ROUTE informa al programa del orden en que se presentarán al operador. Aún no siendo un lenguaje procedimental es posible utilizar sentencias condicionales y bucles para un mejor control de la ruta del formulario.

Los errores de consistencia pueden definirse con dos categorías: errores blandos y errores duros. Los blandos se definen con un párrafo SIGNAL y son aquellos que pueden ser permitidos en el cuestionario. Los duros se definen en el párrafo CHECK y son aquellos que a juicio del promotor de la encuesta no son admisibles teniendo en cuenta la calidad de los datos requerida.

Blaise permite definir variables de tipo entero, real, string y array, con los párrafos VAR y TYPE, que aportan una facilidad adicional al definir rutas y comprobaciones.

El párrafo EXTERNAL posibilita recoger datos de ficheros en formato Blaise o en formato ASCII.

Los párrafos pueden estructurarse en bloques y tablas, que permiten el almacenamiento de los datos en ficheros diferentes y diferentes presentaciones en pantalla. Este proceso es transparente al usuario, manejando los bloques y las tablas como una unidad lógica.

Una vez construida la especificación del formulario, la función de comprobación de sintaxis y la utilización de un compilador Pascal permiten realizar el fichero ejecutable de una forma sencilla.

La sintaxis del lenguaje Blaise y sus características están ajustadas a las soportadas por el lenguaje de alto nivel Pascal, utilizado de forma frecuente en círculos científicos. Liberando al usuario de un proceso largo de aprendizaje.

2.2. Los ficheros complementarios

La aplicación incluye una función de codificación asistida por ordenador. Puede realizarse con ficheros de tipo jerárquico, cada uno de los dígitos del código determina un nivel de codificación, accediendo a éstos de una forma secuencial. O bien, una codificación con búsqueda en diccionario. En una misma cuestión del formulario es posible combinar los dos tipos de búsqueda, permitiendo que el operador use en cada momento la más adecuada a sus necesidades.

El párrafo EXTERNAL, ya mencionado, permite otro tipo de recuperación de datos, con ficheros indexados, ya sea de la misma investigación realizada en un periodo anterior, o de investigaciones relacionadas intimamente con la actual.

2.3. El entorno de la aplicación

Soporta funciones complementarias a los objetivos perseguidos. Así es posible realizar llamadas a un procesador de textos para escribir la especificación del cuestionario. La aplicación tiene uno propio, aunque no de forma rígida, pudiendo sustituirse por otro más conocido del usuario.

Las funciones de comprobación de sintaxis, llamada al compilador, llamada al procesador de textos, realización de diccionarios, formación de ficheros índices, se encuentran integradas en un menú de tipo Pull-Down, de forma que es posible seleccionar una de ellas con las teclas de movimientos del cursor.

2.4. Completitud del proceso de captación de datos

Los datos obtenidos en una investigación deben ser representativos de la población objeto del estudio. En primer lugar una muestra seleccionada con probabilidades desiguales debe ser ponderada para producir estimaciones más acuradas. En segundo lugar la falta de respuesta en algunas unidades afecta a la precisión, cuando éstas tienen características que difieren de aquellas de la población investigada.

El sistema Blaise incluye un método general, basado en un modelo lineal, para obtener las ponderaciones necesarias, que puede ser aplicado a diferentes esquemas de ponderación. La descripción del programa LINWEIGHT se puede consultar en Bethlehem y Keller (1987) para más detalles.

2.5. Análisis de los datos

Una vez obtenido un fichero limpio de errores, un primer paso del análisis de los datos es el cálculo de distribuciones de frecuencia y tabulaciones cruzadas de las características más significativas.

El sistema Blaise permite enlazar con los paquetes de análisis STATA y SPSS basado en microcomputador y en mainframe. Incluye un programa de tabulación desarrollado por el CBS, Abacus, con una interfase de usuario cómoda, que realiza rápidamente tabulaciones cruzadas, incluyendo en las celdas conteos, porcentajes y medias.

Una salida a ficheros ASCII permite recoger los generados en la mayoría de los sistemas de análisis de datos basados en microcomputador.

El programa ANOTA (ANalysis of Tables) es una técnica para explorar las relaciones entre variables cualitativas. Los coeficientes estimados tienen la misma significación que los coeficientes de regresión en las variables numéricas. En términos de ANOVA el modelo ANOTA solamente presenta los efectos principales.

Con la técnica de Análisis de Correspondencia (CORAN) puede obtenerse el grado y el tipo de asociación en una tabla bidimensional. Comporta algún grado de similitud con la técnica de análisis de componentes principales al descomponer una tabla de contingencia en factores, sin embargo los efectos se expresan por el estadístico chi-cuadrado en vez de por la varianza.

Estas dos técnicas han sido desarrolladas por el CBS, integrándose en el sistema Blaise.

3. CONCLUSIONES

El sistema se revela como una potente herramienta que permite la recogida de información, la depuración de los datos básicos, su ponderación y análisis, con aplicación de los ordenadores a la mayoría de las fases de una estadística.

La utilización de estas técnicas debería traducirse en un menor coste, una disminución de los errores que afectan a la investigación en todas sus fases y una reducción sustancial en el tiempo que media entre la realización de la estadística y la publicación de los resultados.

La flexibilidad es una de las características que más llama la atención. Así es posible definir entradas de datos muy variadas, chequeos y comprobaciones on-line que permiten una grabación inteligente del microdato, la depuración con una herramienta potente y la conexión con entornos diferentes para el análisis de los datos.

Permite trabajar en los sistemas operativos MS-DOS y UNIX, y en entornos monousuario y multiusuario con redes de área local; característica que posibilita la confidencialidad del dato estadístico, regulada legalmente en todas las naciones.

BIBLIOGRAFIA

- J. G. BETHLEHEM, A. J. HUNDEPOOL, J.-P. KENT, M. H. SCHUERHOFF y L. F. M. VERMEULEN. Blaise 2.0. Language Reference Manual (1989).
- J. G. BETHLEHEM, A. W. HOOGENDOORN, A. J. HUNDEPOOL, J.-P. KENT, M. H. SCHUERHOFF y L. F. M. VERMEULEN. Blaise 2.0 System Description (1989).
- JELKE G. BETHLEHEM y WOUTER y J. KELLER. New Technologies in Computer Assisted Survey Processing (1989).
- J. G. BETHLEHEM, A. J. HUNDEPOOL, M. H. SCHUERHOFF y L. F. M. VERMEULEN. Blaise 2.0 Cadi User Manual (1989).
- J. G. BETHLEHEM, A. J. HUNDEPOOL, J.-P. KENT, M. H. SCHUERHOFF y L. F. M. VERMEULEN. Blaise 2.0. Capi/Cati User Manual (1989).

REVISTA ESTADISTICA ESPAÑOLA

INFORMACION PARA LOS AUTORES

Envío de originales

Los originales deben enviarse al Director de la Revista, INE, Paseo de la Castellana, 183. 28071 Madrid. Deben estar redactados en castellano y no haber sido publicados o estar en proceso de publicación en otro lugar. Se enviarán 4 copias del trabajo y la revista acusará siempre su recibo.

PRESENTACION DE LOS MANUSCRITOS

Estructura

La primera página debe incluir exclusivamente el título del artículo y el nombre, dirección completa y teléfono del autor. En el caso de varios autores se indicará a quién debe dirigirse la correspondencia. La segunda página contendrá únicamente el título y resumen del trabajo de un máximo de 100 palabras, seguida de 3 a 6 palabras clave y la clasificación AMS del artículo. El texto del artículo comenzará en la tercera página y las secciones se numerarán consecutivamente. La última página del original contendrá en inglés el título del artículo, un resumen del mismo bajo el epígrafe Summary y las palabras clave. El manuscrito debe mecanografiarse a doble espacio. Cuando el título del artículo contenga más de 80 caracteres, se deberá indicar un título alternativo de dicha longitud o menor.

Gráficos

Todos los diagramas o gráficos se numerarán sucesivamente y se indicará su posición en el texto con el nombre de *figura*. Se colocarán al final del manuscrito y deberán ser de la calidad necesaria para su reproducción.

Referencia

El sistema de referencia a seguir es el oficial del International Statistical Institute. Los autores se citarán en el texto por su nombre, seguido de la fecha de publicación; ejemplo: Box (1986), y las referencias se situarán en orden alfabético al final del texto, como sigue:

- a) *Libros*: Weisberg, S. (1985). *Applied Linear Regression*, New York: Wiley.
- b) *Artículos*: Mahalanobis, P. c. (1950). Why Statistics? *Sankhya*, 10, 195-228.
- c) *Trabajos en obras colectivas*: Box, G. E. P. (1983). An Apology for Ecumenism in Statistics. *Scientific Inference, Data Analysis and Robustness*, Ed. C. E. P., Leonard, T. y Wu, C. F., pp. 51-84. New York: Academic-Press.

Evaluación de los originales

Los originales serán sometidos a un proceso de evaluación garantizado el anonimato tanto del autor como de los evaluadores. El objetivo de la revista es que el autor reciba información sobre el resultado de la evaluación en un plazo máximo de tres meses.

Pruebas y separatas

Aceptado el artículo y antes de su publicación definitiva, el autor o autores recibirá 2 juegos de pruebas para correcciones, uno de los cuales deberá devolver corregido a la Revista en un plazo de una semana desde su recepción. Una vez publicado, recibirá 25 separatas de su trabajo. Los posibles costes de impresión derivados de cualquier modificación de la versión final aceptada del manuscrito o de retraso en la corrección de pruebas serán a cargo del autor o autores del mismo.