

ESTADÍSTICA ESPAÑOLA
Vol. 41, Núm. 144, 1999, págs. 19 a 46

Problemas en el Diseño y Validación de Cuestionarios: tratamiento con **QUESTPOT v.1.2(*)**

por
PEDRO ANTONIO GARCÍA LÓPEZ(**)
ANDRÉS GONZÁLEZ CARMONA
JUAN ANTONIO MALDONADO JURADO

Departamento de Estadística e I.O.
Universidad de Granada.

En esta comunicación presentamos una visión general de los problemas inherentes al diseño, explotación y validación de cuestionarios, con especial atención a los asistidos por ordenador (tipo CAPI/CATI). Se tratará de reflejar para estos últimos las ventajas sustanciales que proporcionan frente a los diseñados de forma tradicional, cuestión que se mostrará con la descripción del programa QUESTPOT v.1.2.

Palabras clave: Diseño de cuestionarios, CAPI/CATI, programa QUESTPOT v.1.2, Visual Basic.

Clasificación AMS: 65U05.

(*) Este trabajo ha contado con el apoyo económico del Instituto de Estadística de Andalucía. El programa tiene una versión de prueba que se puede solicitar a los autores en la dirección de correo electrónico: jamaldo@goliat.ugr.es

(**) Departamento de Estadística e I.O. Facultad de Ciencias. Universidad de Granada. Granada (18071). Tel. 958243157.

INTRODUCCIÓN

Fuentes de error en las encuestas

Como es sabido, la forma habitual de obtención de datos en Ciencias Sociales y en estudios prospectivos es la encuesta. Este método de recogida de datos primarios supone establecer un proceso de comunicación en el que el investigador, de acuerdo a sus necesidades y objetivos, establece qué se pregunta, de qué forma se pregunta, a quién y dónde, en lo que los ingleses vienen en llamar las cuatro cuestiones “W” (What, how, Who, Where).

Este proceso de comunicación se establece de forma oral o escrita, y está sujeto a “ruidos” que pueden desvirtuar o deteriorar la calidad de los datos obtenidos en el proceso. Estas fuentes de error serán objeto de estudio y cuidado por parte del investigador, ya que las mismas pueden propiciar que en fases posteriores de explotación de los datos, los resultados sean meramente “creencias” o “intenciones”.

La mayoría de los autores, suelen citar como fuentes de error, entre otras, las siguientes:

- Imputables al investigador:
 - ✓ Mala definición del marco muestral.
 - ✓ Deficiente representatividad de la muestra por mal esquema de muestreo.
 - ✓ Deficiente selección, formación y control de los entrevistadores.
 - ✓ Deficiente diseño del cuestionario.
- Imputables al entrevistador:
 - ✓ Mala presentación por parte del entrevistador y de la entrevista.
 - ✓ Deficiencias en la formulación de las preguntas.
 - ✓ Deficiencias en el control y en el registro de las preguntas.
 - ✓ Mal seguimiento de las instrucciones.
 - ✓ Fraude o falsificación de cuestionarios.
- Imputables al entrevistado:
 - ✓ Falta de respuesta por temor al no anonimato o debido a preguntas de contenido íntimo.

- ✓ Deficiencias en las respuestas por falta de comprensión.
- ✓ Falta de sinceridad.
- ✓ Respuestas sesgadas o forzadas por complacencia con el sentido del cuestionario.

En esta comunicación, nos centraremos en los aspectos relativos a los errores por mal diseño o explotación del cuestionario, que como se puede comprobar de acuerdo a la relación anterior, suponen la parte principal de los errores citados. Además, hay que hacer notar que salvo las fuentes de error debidas al investigador (que por su formación deberían ser fácilmente detectables y corregibles), en las debidas al entrevistador y al entrevistado, es manifiesto que el diseño del cuestionario es una causa principal en el origen de los problemas planteados.

Algunas recomendaciones para evitar el problema “ruido”

En la mayor parte de las ocasiones, estos ruidos activan unos de los mecanismos inherentes en la definición de un sistema estadístico para el tratamiento de encuestas, que no es otro que el que se refiere a la depuración y validación de cuestionarios. Este aspecto, junto con el relativo a la depuración misma de los datos, se encuentra indisolublemente unido al concepto de la calidad de los datos obtenidos en cualquier investigación estadística, y condiciona varias de las fases de dicha investigación. Así se pone de manifiesto en el informe *“Panel on Incomplete Data”*(1), en el que se dan una serie de recomendaciones sobre el problema de los datos incompletos, que puede venir determinado, entre otras causas, por la definición del cuestionario, siendo ésta, como hemos dicho, una de las fuentes de error en las encuestas.

En general, y con objeto de paliar en buena medida estos problemas, se pueden dar algunas recomendaciones, referidas a los siguientes aspectos:

- Obtención de los datos.
- Buena recolección, cuestionario y diseño de la base de datos.
- Medición y manejo de la no respuesta.
- Diseño de la encuesta para la no respuesta.
- Análisis del efecto de la imputación y propuesta de esquemas alternativos.

(1) Panel of Incomplete Data, *Incomplete Data in Sample Surveys*, Academic Press (1983).

- Documentación mediante encuestas y otras fuentes alternativas.

Más precisamente, el proceso de detección y tratamiento de errores no muestrales condiciona las fases de planificación y diseño de la investigación, ya que el usuario o responsable de la misma tomará especial cuidado en:

- la definición del tipo de variables del estudio,
- el método de recogida de datos,
- los controles de verificación de datos,
- el método de detección de errores y el tipo de corrección,
- el método de imputación automática, y el método de estimación de valores faltantes o anómalos.

Por otra parte, el proceso de depuración del cuestionario, determina necesariamente el proceso de validación de la investigación estadística, puesto que permitirá evaluar:

- la cobertura obtenida,
- el nivel de no respuesta,
- la precisión en las respuestas, y
- la distribución de los errores detectados.

Errores de los datos debidos al diseño del cuestionario.

Como hemos puesto de manifiesto, el cuestionario como herramienta fundamental en la práctica de las encuestas, ha de ser diseñado con especial atención y cuidado, ya que los problemas que pudieran derivarse de una etapa inicial de la encuesta tendrán consecuencias imprevisibles.

Para la elaboración de éste se ha de tener en cuenta, al menos, el propósito y la necesidades de la investigación, y por ende, una serie de cuestiones preliminares:

- La naturaleza y características de la información que se busca. De esta forma, las posibles respuestas a una pregunta servirán de ayuda para confeccionar un listado de informaciones a recoger.
- Las características de los entrevistados y su predisposición.
- El método de encuesta a aplicar.

Con objeto de introducir las ventajas que encontraremos en la implementación de programas de diseño de cuestionarios, es conveniente citar, si bien de forma

concisa, algunos de los errores de los datos que se deben exclusivamente al diseño del cuestionario. En contraposición a lo que sería el estudio de los errores de muestreo, la depuración de datos en la validación del cuestionario, se centrará en el estudio de los errores ajenos al muestreo que, a su vez, se pueden producir por dos situaciones distintas:

- Que el error se produzca por una colisión en la identificación del dato, debido a que:
 - ✓ Ha habido errores de transcripción y no se han respetado los códigos establecidos, así como los formatos correspondientes a cada variable.
 - ✓ Se han detectado duplicidades en la muestra.
 - ✓ Se ha violado alguna condición de flujo del cuestionario.
- Estos errores son importantes ya que afectan fundamentalmente a los procesos de grabación, clasificación y manipulación de los datos(2).
- Que el error se produzca en el dato propiamente dicho. En éstos se puede distinguir entre:
 - ✓ Errores de tipo aleatorio: pueden ocurrir en cualquier momento del proceso estadístico de adquisición y tratamiento de los datos, aunque no modificarán la distribución de las variables objeto de estudio.
 - ✓ Errores sistemáticos: se suelen producir en la fase de adquisición de datos, y habitualmente son consecuencia de una falta de entendimiento de las cuestiones por parte del respondiente, o por una ocultación intencional.

En estos se habrá de tener especial cuidado, ya que tienen consecuencias en la distribución de las variables de estudio, y podrían invalidar globalmente los datos adquiridos sobre las mismas. No obstante, estos errores tienen importancia desde el punto de vista estadístico en la fase de análisis de los datos, por lo que no será objeto de estudio en esta comunicación.

Por tanto, desde un punto de vista general, se puede decir que el proceso de evaluación de los errores producidos por el diseño del cuestionario, tiene dos funciones principales, a saber:

- Facilitar el proceso de los datos, permitiendo que la elaboración, manipulación y tratamiento de los mismos se pueda realizar de forma estandarizada.

(2) Piénsese, por ejemplo, en el caso de un cuestionario que una vez grabado tiene un campo indicador de edad < 16, que en un campo siguiente contenga la característica viudo.

Así, será conveniente chequear que los identificadores son correctos, que no existen datos duplicados, que se ha respetado el tipo y el formato, etc.

- Mejorar la calidad de los datos con el objetivo de que éstos sean útiles para la realización de inferencias, que sean de utilidad para futuras realizaciones, y que estén libres de inconsistencias.

DISEÑO DE CUESTIONARIOS ASISTIDOS POR ORDENADOR

Como puede deducirse de los apartados anteriores, el diseño de un cuestionario no es un trabajo inmediato, si se realiza con el cuidado mínimo exigible para que los resultados sean aceptables. Hoy en día el uso de ordenadores facilita esta tarea, almacenando el cuestionario de forma que la información que contiene sea susceptible de ser usada en las fases posteriores de la encuesta.

Existen diversos programas que permiten describir el cuestionario en términos de preguntas cerradas. Entre ellos encontramos a un nivel básico de programa QFAST, y a un nivel más elevado BLAISE, QUANTIME e ISSA, que son los más conocidos, aunque existen otros específicos para sistemas medios o grandes como Digital o Hewlett Packard(3). Más recientemente el paquete estadístico SPSS ha introducido en su versión 8.0 un manejador de formularios muy flexible que denombra "SPSS Data Entry Builder", que ha tenido una buena acogida en el mercado.

El uso de programas de análisis y gestión de cuestionarios de este tipo (que en lo que sigue denominaremos CAPI(4)/CATI(5)) facilitan en buena medida la fase de depuración de los datos del cuestionario, ya que evita varios de los problemas clásicos que se presentan en la depuración de los datos, como son los errores de grabación y los errores de formato o de control de identificaciones. Por otra parte, mejoran la autoedición y la presentación y manejo de los mismos.

Estos sistemas proporcionan gran flexibilidad e individualizan los cuestionarios reduciendo los malos entendidos, aumentando asimismo la velocidad del proceso de obtención, grabación y codificación de los datos. Por otra parte, se ha demostrado en la práctica que el aspecto novedoso de los mismos, consiguen despertar el interés y la colaboración del entrevistado, lo cual palía en gran medida algunas

(3) Como referencia básica, en Bethlehem J.G. & Keller W.J. "New technologies in computer assisted survey processing", Eustat (1989), se estudian los aspectos básicos a tener en cuenta en el proceso de edición y diseño del cuestionario asistido por ordenador, así como su implementación en BLAISE.

(4) Computer Assisted Personal Interviewing.

(5) Computer Assisted Telephone Interviewing.

situaciones de falta de respuesta total. Además, el efectivo control del flujo del cuestionario, reduce en parte el tiempo de formación de los entrevistadores y evita algunas situaciones de demora en la realización de preguntas por parte de éstos.

Sin embargo, por el momento, son muy pocos los esfuerzos por incorporar a estos programas de un sistema de introducción, análisis y evaluación de cuestiones abiertas, lo cual nos permite anticipar que una vía de estudio futura será la construcción de sistemas que interactúen con los conocidos programas de análisis de datos textuales (SPAD, ALCESTE, etc.), que tienen como objetivo, entre otros, el análisis de cuestiones abiertas.

Volviendo a los aspectos básicos de un programa CAPI/CATI, es recomendable que básicamente éstos incluyan los dos controles mínimos citados anteriormente, esto es:

- Control de codificación: el dato se corresponde con el tipo prefijado.
- Control de grabación: el dato se grabará en el registro en cierta posición asignada.

En una segunda fase sería conveniente la definición de *edits*(6) (condiciones), que evalúen globalmente consistencia de los datos especificando:

- Situaciones imposibles (p.e. la edad de los padres no puede ser menor que a de los hijos), que habrá de detectar y eliminar.
- Situaciones improbables (p.e. salarios de millones de pesetas mensuales, viudos de 15 años, etc.), en las que será de capital importancia detectar valores anómalos en la distribución de las variables.
- Controles de flujo de respuesta en el cuestionario, de forma que no se permita la respuesta en ciertas categorías si se ha optado por cierta opción anterior.

De la misma forma que hemos venido comentando que el control del flujo de respuesta es un requerimiento inherente a un programa CAPI/CATI, no lo es así el control de las dos primeras situaciones citadas anteriormente. De hecho, la mayoría

(6) La mayoría de los grandes sistemas de edición y tratamiento de datos (GEIS, SPEER, DIA, etc.) disponen de módulos específicos para la definición y análisis de edits, generalmente de acuerdo a la metodología Fellegi-Holt. El objetivo fundamental de estos sistemas es la definición de regiones de aceptación para los datos válidos, y por extensión la detección y corrección (manual o automática) de aquellos registros que no cumplen las condiciones, esto es, que se sitúen en las regiones de rechazo definidas por los edits. Sin embargo, las pretensiones, inicialmente limitadas del programa que nos ocupa, no permitirá en un primer momento, la definición y el análisis de edits.

de los programas, se limitan a proporcionar una salida en formato determinado (preferentemente ASCII) que posteriormente se analizará mediante algún paquete o rutina estadística específica de análisis de valores extremos tanto desde un estudio univariante como multivariante del conjunto de datos.

En la sección dedicada a la descripción del programa QUESTPOT v.1.2 volveremos sobre este aspecto.

EL PROGRAMA QUESTPOT V.1.2.

Descripción General del programa

El programa QUESTPOT v.1.2 desarrolla los tópicos referentes a depuración e imputación expuestos anteriormente, por medio de diversas funciones contenidas en módulos diferenciados. Una, incorporando los controles mínimos que se citaron anteriormente, es decir, el Control de codificación, el Control de grabación y el Control de flujo, como corresponde a un programa de tipo CAPI/CATI. Estos controles se encuentran programados en el módulo general de adquisición y edición de datos, y resuelven el problema de que los datos se correspondan con el tipo prefijado, y que los mismos se graben en posiciones asignadas, que faciliten el proceso de manipulación y su tratamiento.

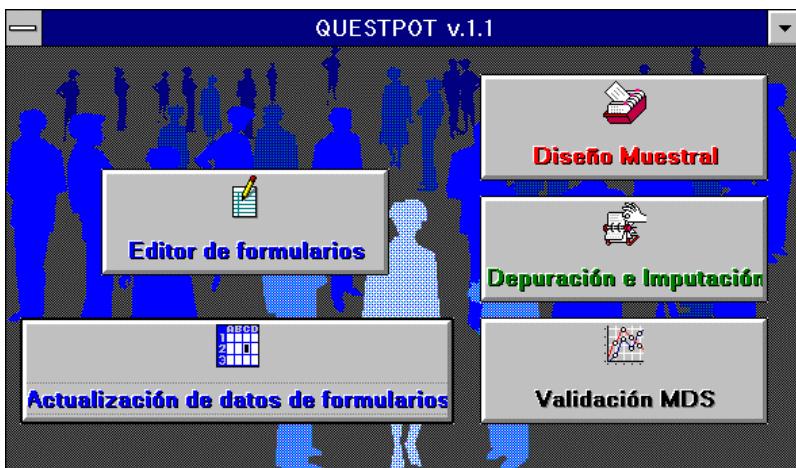
El programa usa la definición de ventanas que proporciona VISUAL-BASIC y es capaz de interaccionar con módulos definidos de programas estándar, a modo de sistema potenciado(7).

QUESTPOT está dividido en cinco módulos lógicos independientes e interrelacionados:

- El editor de formularios: encargado de llevar a cabo las manipulaciones convenientes sobre los cuestionarios.
- El editor de datos: conecta con el editor de formularios para la toma/actualización de datos sobre un cuestionario particular, de manera inteligente interactiva.
- El módulo para diseño muestral: como su propio nombre indica, lleva a cabo el proceso de encuestas por muestreo.

(7) De esta forma, interacciona con el paquete estadístico BMDP para la resolución del problema de depuración; y con el paquete SPSS para la resolución del problema MDS de validación de cuestionarios.

- El módulo de depuración e imputación: conecta con otro programa, (BMDP) con el fin de llevar a cabo una depuración/imputación sobre el fichero de datos asociado a un cuestionario determinado.
- El módulo de validación MDS: apoyado en técnicas de Multidimensional Scaling, conecta igualmente con otro programa (SPSS) para realizar una validación del cuestionario atendiendo a la posible reducción de su dimensionalidad.



El editor de formularios

La pantalla principal del editor de formularios, a la que se accede una vez que se ha seleccionado entre las opciones de crear un formulario nuevo o editar un formulario ya existente, interactúa con el usuario con el fin de crear/editar un cuestionario para su posterior tratamiento. En él, y tras la determinación de un título para el mismo (que identificará posteriormente al cuestionario, y que será el que aparecerá en la hoja impresa), se irán creando/actualizando una a una las preguntas que lo formarán, con las siguientes especificaciones:

- El texto de la pregunta.

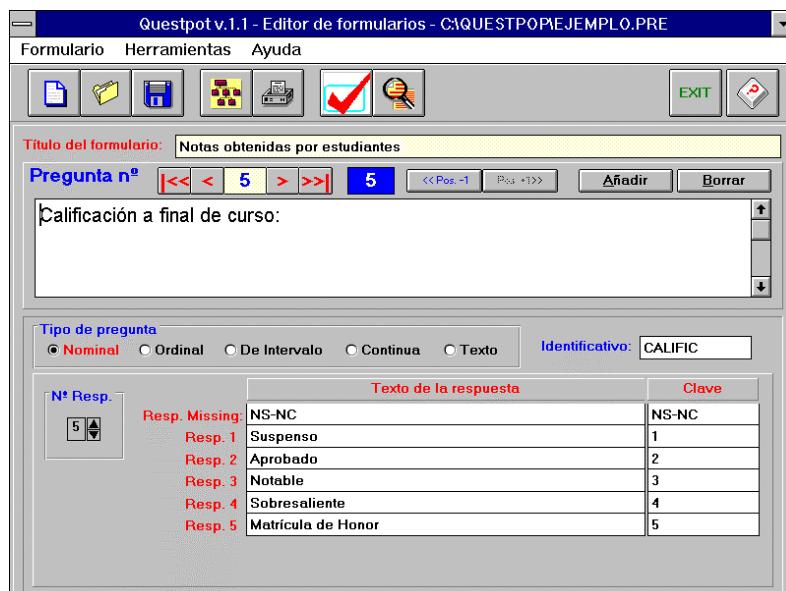
- El tipo de escala de medida para la variable que posteriormente en el fichero de datos identificará a la pregunta, distinguiendo entre:
 - ✓ Nominal.
 - ✓ Ordinal.
 - ✓ De intervalo.
 - ✓ Continua.
 - ✓ Textual.
- El identificativo, correspondiente al nombre que posteriormente se asociará a la variable en el fichero de datos para la pregunta.
- Según el tipo de escala, un conjunto de ítems relacionados con la misma.

En cualquier momento es posible añadir o borrar una pregunta, así como modificar su posición en el cuestionario.

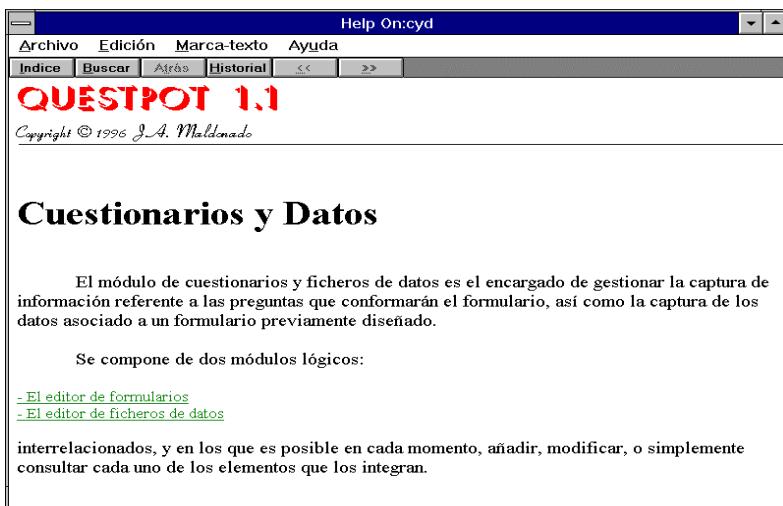


Entre las herramientas de que consta este módulo, y que se encuentran tanto a modo de menús, como en iconos representativos, se encuentran:

- Abrir un formulario nuevo.
- Abrir un formulario existente para su actualización/modificación.
- Salvar el formulario en curso.
- Definir bloques homogéneos de preguntas: esta utilidad permitirá identificar en el cuestionario impreso bloques de preguntas de un mismo grupo, dentro del total de la batería de preguntas del cuestionario, con carácter diferenciador, lo que conducirá a una mejor presentación y lectura del mismo por parte del entrevistado.
- Imprimir un cuestionario.
- Verificar una pregunta: con esta utilidad, se podrá evaluar si se cumplen todas y cada una de las consideraciones sobre la pregunta en curso, atendiendo al tipo de escala, y al conjunto de sus descriptores en el módulo.

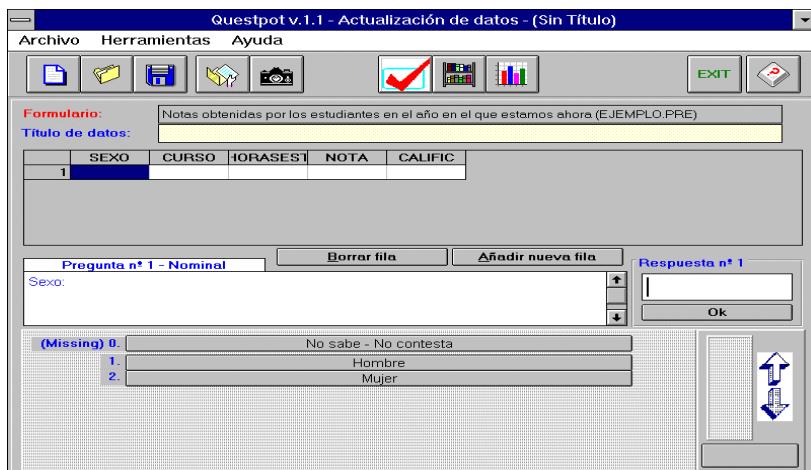


- Buscar una pregunta: cuando el cuestionario es suficientemente extenso, esta utilidad permitirá la localización rápida y eficiente de una determinada pregunta dentro del mismo, con pocas pulsaciones por parte del usuario.
- Solicitar información sobre el módulo (menú de ayuda).



El editor de datos

El editor de ficheros de datos permite crear tantos ficheros nuevos se deseen para un determinado cuestionario, o bien editar ficheros ya existentes.

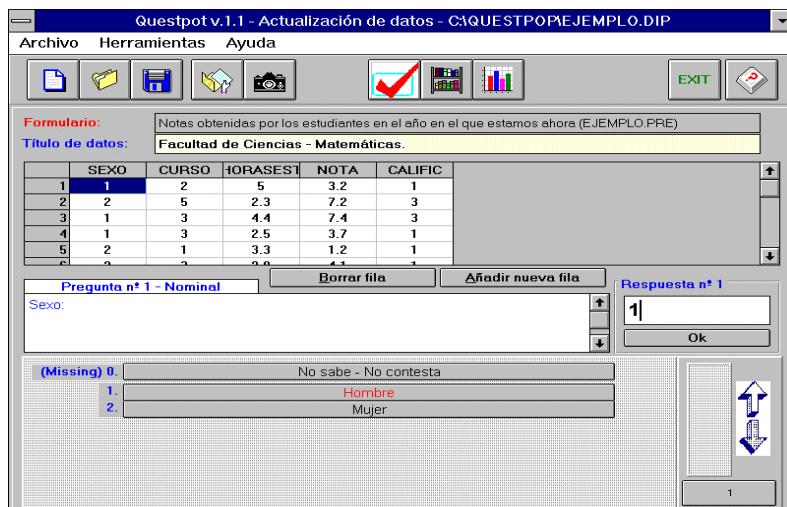


Si se crea un fichero de datos nuevo, el programa demandará el nombre del fichero donde se almacenó el formulario al que se va a hacer corresponder. Tras la selección del mismo por parte del usuario, la pantalla principal es de nuevo una hoja que interactúa con él, y en la que está visible en todo momento:

- El identificativo del cuestionario asociado.
- El título del fichero de datos en curso.
- La rejilla a modo de matriz rectangular, dividida en filas y en columnas, donde, en cada fila aparecerán los datos correspondientes a todas las preguntas del cuestionario para un mismo individuo, y en cada columna los valores de la variable correspondiente a la pregunta correspondiente para el total de los individuos entrevistados.
- La pregunta asociada a la variable activa: aparece tanto el tipo de escala de medida utilizada en la variable activa, como el texto completo de la pregunta asociada. Ello permitirá, tanto la localización exacta de las características de la pregunta, con los descriptores a que hacíamos referencia anteriormente, así como la posibilidad de llevar a cabo entrevistas telefónicas por parte de los agentes.

- Diversas casillas para la toma del dato, según el tipo de pregunta, y actuando de forma inteligente; esto es, llevando a cabo un proceso de depuración previo sobre el conjunto de especificaciones determinadas en la creación del formulario asociado para cada una de las preguntas que lo forman (mínimos, máximos, total, descripción y claves para las mismas, etc..)

Si se edita un fichero de datos existente, aparecerá la rejilla con los datos ya introducidos en el fichero.



En todo momento se dispondrá de las siguientes acciones y utilidades, localizadas tanto a modo de menú, como en los correspondientes iconos representativos:

- Crear un fichero de datos nuevo para un determinado cuestionario.
- Abrir un fichero de datos existente.
- Grabar el fichero de datos activo.
- Importar datos desde otros ficheros.
- Copiar al portapapeles el contenido de la rejilla de los datos, o parte de la misma: se podrá seleccionar desde qué y hasta qué fila, y desde qué hasta qué columna se desea pasar información al portapapeles con el fin de poder crear los informes que se consideren oportunos.

- Verificar la fila activa: sólo si los datos son correctos desde un primer filtrado (marcado por las especificaciones del formulario asociado), será posible grabar el fichero y abandonar la aplicación.
- Ordenar el fichero de los datos por la variable activa, o por más de una simultáneamente. Si se desea esto último, habrá que actuar de forma inversa, ordenando en primer lugar por la última variable de clasificación en la lista, hasta llegar a la primera.
- Crear un gráfico para la variable activa: en ocasiones, dicho gráfico proporcionará la información necesaria para tomar determinadas acciones sobre el proceso de muestreo.

The screenshot shows the QUESTPOT 1.1 software interface. The window title is "QUESTPOT 1.1 - Actualización de datos - C:\QUESTPO\REJEMPL.DIP". The menu bar includes "Archivo", "Herramientas", and "Ayuda". The toolbar contains icons for file operations (New, Open, Save, Print) and a checkmark. The main area has two input fields: "Formulario:" (Notas obtenidas por los estudiantes en el año en el que estamos ahora (EJEMPLO.PRE)) and "Título de datos:" (Facultad de Ciencias - Matemáticas). Below these is a table with columns: SEXO, CURSO, HORASEST, NOTA, and CALIFIC. The data entries are:

SEXO	CURSO	HORASEST	NOTA	CALIFIC
1	1	2	5	3.2
2	2	5	2.3	7.2
3	1	3	4.4	7.4
4	1	3	2.5	3.7
5	2	1	3.3	1.2
c	a	2	2.0	1

Below the table are buttons for "Borrar fila" (Delete row) and "Añadir nueva fila" (Add new row). A dropdown menu "Pregunta nº 5 - Nominal" is open, showing "Calificación a final de curso:" with options: NS-NC, Suspensión, Aprobado, Notable, Sobresaliente, and Matrícula de Honor. A "Respuesta nº 1" panel shows the value "1" and an "Ok" button. On the right, there are vertical scroll bars and a small control panel with up/down arrows and a "1" button.

- Solicitar información sobre el módulo (menú de ayuda).
- Borrar una fila de datos: en un determinado momento, puede resultar conveniente eliminar la información correspondiente a un individuo del fichero de datos porque no interese su tratamiento posterior.
- Añadir una nueva fila al fichero: cuando se edita un fichero de datos existente, pueden añadirse al mismo tantas filas como sean necesarias, correspondientes a los cuestionarios que nos vayan llegando, o a los individuos que de manera interactiva se entrevisten.

Otras utilidades de QUESTPOT v.1.2

Las restantes utilidades del programa son relativas a diversos aspectos de las fases de una encuesta por muestreo. En la práctica las utilidades más demandadas por un investigador son:

- Respecto del diseño de muestreo: el cálculo de tamaños muestrales y el reequilibrado de muestras (post-estratificación) una vez que ha tenido lugar la primera fase de toma de las mismas.
- Depuración e imputación de datos como fase de adecuación de los datos previa a la explotación de los mismos mediante los análisis estadísticos habituales.
- Validación de preguntas o grupos de preguntas, en la que se pueda observar la concordancia interna de las cuestiones dentro de una batería, o de las baterías entre sí. Con este objetivo, se implementa un módulo de validación mediante Multidimensional Scaling (MDS) con el que se construirán mapas que permitan al investigador estudiar las concordancias o discordancias de su interés.

Estas utilidades están implementadas en QUESTPOT v.1.2 por medio de módulos complementarios al programa principal de gestión de datos y formularios. Estos módulos interactúan con programas estándares⁽⁸⁾. Estas utilidades se exponen con más detalle a continuación.

Diseño Muestral

En la planificación de una encuesta por muestreo se llega siempre a un punto en el que se tiene que decidir el tamaño de la muestra que se estudiará. Esta decisión es importante si se tiene en cuenta que, como es sabido, el tamaño de la misma incidirá en aspectos ineludibles en una investigación estadística práctica, a saber:

- El coste económico de la investigación.
- La precisión de las estimaciones obtenidas.

(8) SPSS y BMDP en esta versión, si bien la definición del sistema para otros programas es inmediata a modo de Sistema Potenciado Estadístico, como hemos comentado.

Con este objeto, el programa QUESTPOT v.1.2 dispone de un submódulo, que permite que el investigador pueda en todo momento conocer la información que necesite relativa a los tamaños muestrales, de acuerdo a los condicionantes genéricos citados anteriormente. Asimismo, el investigador tendrá diversas opciones que permitan conocer esta información sobre distintos tipos de muestreo. Más específicamente el módulo dispone de las opciones:

- Introducción del tipo de muestreo.
- Cálculo del tamaño muestral en base al tipo de muestreo y los costes de la investigación.
- Parrilla de selección aleatoria de unidades en base al tipo de muestreo.
- Cálculo de los errores muestrales asociados al tipo de muestreo.

De este modo el investigador podrá, en las fases iniciales de una encuesta por muestreo, garantizar el procedimiento de obtención de las muestras, así como en fases posteriores, conocer el grado de adecuación de los cuestionarios que se van recepcionando, al diseño de muestreo propuesto.

Esta última cuestión no es baladí, ya que en la práctica es bastante común que el investigador corrija el diseño previo de muestreo, a la vista de los datos que va recibiendo. De esta forma, el programa ofrece al investigador la posibilidad de que corrija de forma casi “interactiva” tal diseño de muestreo.

Depuración e imputación de datos

Como se ha comentado anteriormente, tras la grabación de los cuestionarios y la consiguiente incorporación de la información que contiene en la base de datos primarios, se hace necesaria su depuración por varios motivos, entre ellos los siguientes:

- Para comprobar que no ha habido errores de transcripción y que se han respetado los códigos establecidos, así como los formatos correspondientes a cada variable.
- Para eliminar duplicidades en la muestra.
- Para proporcionar información que nos permita evaluar y en consecuencia mejorar la calidad de datos.

Por tanto, QUESTPOT v.1.2 proporciona una solución a los dos primeros problemas, por una parte gracias a su opción de chequeo, y por otra por la edición

simultánea de actualización de datos. En consecuencia, se tratará en este punto de incidir con más detenimiento sobre la última circunstancia descrita.

Por otra parte, en un primer proceso de depuración se habrán de tener en cuenta tres tipos de control sobre la información:

- Un control sobre las identificaciones de los cuestionarios.
- Un control sobre los formatos definidos para las variables.
- Un control sobre el grado de cobertura de la encuesta.

Con este objeto, se definirá un sistema de requerimientos consistentes, es decir, se pretenderá que el usuario exprese un conjunto de condiciones válidas (edits) para el dato leído. En una primera fase se definirán edits que permitan:

- Especificar situaciones imposibles.
- Especificar situaciones improbables.
- Controlar el flujo de respuesta en el cuestionario.
- Determinar umbrales para datos extraños en variables numéricas (detección de outliers).

Estas acciones se incluyen en la fase de diseño del cuestionario en el proceso de definición y codificación de variables. La última acción (la detección de outliers) se implementa mediante interface con BMDP (no obstante, como hemos dicho, puede plantearse un interface similar con otros programas de amplia difusión).

Una vez realizada esta acción podemos realizar una macrodepuración sobre aquellas variables que tengan más de un porcentaje determinado de valores fuera de ciertos límites, o una microdepuración sobre aquellos casos que tengan más de un número de valores en las variables fuera de los límites fijados.

De esta forma, una vez detectado el valor o valores inconsistentes, se ha de permitir de forma automática:

- Crear una nueva categoría de invalidado o erróneo para valores no consistentes de las variables.
- Imputar nuevamente en los campos detectados como erróneos nuevos valores consistentes con los demás registros (si considera que se ha realizado una grabación errónea).

Existen reticencias en cuanto a la existencia o no de la aplicación de los procedimientos de imputaciones ya que muchos autores considera que esta manipulación es un simple proceso de fabricación de datos.

Precisamente la característica fundamental de toda imputación es la suposición de que los datos son homogéneos, cuestión que no siempre se tendrá en la práctica. Por ello es recomendable detectar a través de los errores muestrales de las estimaciones (desviación típica del estimador, a veces el coeficiente de variación) si la variable es o no homogénea, para lo que el programa podrá realizar interface con otros programas.

En este caso, se procederá sin más a la imputación de valores en los campos erróneos. En caso contrario se estudiará si la variable está fuertemente correlacionada con una o más variables de la base de datos para proceder con categoría con un procedimiento de imputación como el de regresión.

Una vez que se ha visto adecuada la imputación tendremos que tener en cuenta además dos suposiciones adicionales:

- Los campos del registro que no son objetos de imputación carecen de error de contenido.
- La pérdida de información en los campos se debe exclusivamente a razones aleatorias. En este sentido se suele hacer la salvedad de que esta aleatoriedad dependa del resto de la información contenida en el registro, o bien si es de carácter absoluto, esto es:
 - ✓ supuesto MAR ("missing at random"),
 - ✓ supuesto MCAR ("missing completely at random").

En definitiva, hemos de insistir en que los procedimientos de imputación se aplicarán en base a ciertos requisitos y no de forma indiscriminada. Para ello, en el programa QUESTPOT v.1.2 se siguen las siguientes pautas:

- Realizar una imputación manual en base de los conocimientos del usuario.
- Si se comprueba la existencia de relaciones entre la variable a imputar y otra u otras de la base de datos, procederemos a imputar siempre que el número de registros erróneos o faltantes en dicha variable no sea mayor del 40%.
- Si se tiene el supuesto MCAR para la variable, es conveniente detectar que al realizar la imputación no se transgreda el criterio del error relativo por imputación.

- Si se detecta homogeneidad en estratos o conglomerados, es conveniente realizar imputación por separado en cada subpoblación.

En el segundo caso, para los procedimientos automáticos de imputación, el sistema proporcionará un interface con otros programas (p.e. módulo AM de BMDP), haciendo uso de los procedimientos habituales de regresión y de imputación probabilística (algoritmo EM).

En los casos tercero y cuarto (bajo el supuesto MCAR) el programa implementará métodos de imputación de registro donante, estableciendo tipologías entre los registros e imputando los valores de cierto registro cuya distancia con el marcado sea mínima según cierta condición.

Para la construcción de dichas tipologías se usará nuevamente interface con algún programa de clasificación jerárquica, que establezca los grupos de registros "matching" sobre los que se define una función distancia respecto del registro receptor, con el objeto de seleccionar como donante el que tenga la menor distancia. Finalmente, obtenido el registro donante, se imputará al receptor los valores faltantes o marcados como erróneos.

El programa consta de varias ventanas de opciones, por medio de las que el usuario podrá editar y depurar e imputar datos del fichero en el que previamente se han grabado los datos de la encuesta, de acuerdo a las consideraciones efectuadas anteriormente, y que recogen las posibilidades del programa BMDP sobre el que se interactúa.

Validación MDS

Multidimensional Scaling (MDS) es una técnica de análisis de coeficientes de proximidad entre objetos que permite obtener la dimensión mínima de la configuración de estos objetos según cierto criterio estadístico. Este criterio se refiere habitualmente a la distorsión entre las proximidades originales y las distancias finalmente obtenidas en la configuración propuesta.

Nuestra implementación en QUESTPOT v.1.2. se centra fundamentalmente en el proceso de validación del cuestionario con la idea de observar qué cuestiones o baterías de cuestiones tienen consistencia entre sí (tienen la misma interpretación por el sujeto que responde) y cuales están más alejadas desde el punto de vista de la percepción del sujeto, por medio de la representación en mapas MDS de la configuración entre las distintas cuestiones.

Esta estrategia constituye una alternativa a los procedimientos clásicos de estudio de la fiabilidad y validación de un cuestionario: coeficiente Alpha de Cronbach y

el análisis factorial entre otros. La mejora frente a estos últimos reside fundamentalmente en que, manteniendo la misma significación desde el punto de vista de su justificación teórica, es más fácil para el investigador interpretar un gráfico que un coeficiente (respecto del coeficiente alpha de Cronbach) y distancias que ángulos (respecto del análisis factorial), todo ello en una dimensión baja(9).

El procedimiento empleado de MDS es ALSCAL de Takane, Young y de Leeuw (1977) implementado en el programa SPSS. Para ello, desde QUESTPOT v.1.2. se incluirán aquellas órdenes necesarias de SPSS para obtener el mapa MDS pretendido, actuando como interface de la misma forma que en el caso de su interacción con BMDP.

CONCLUSIONES

A la vista de todo lo que antecede, podemos decir que QUESTPOT v.1.2, como sistema potenciado para el tratamiento y análisis de cuestionarios, resuelve muchos de los problemas descritos en la sección introductoria de la comunicación, a saber:

- En cuanto al diseño del cuestionario:
 - ✓ Eficiente diseño del cuestionario, gracias a las diferentes opciones de edición, presentación, impresión, grabación y modificación del mismo.
 - ✓ Formulación eficiente de las preguntas, que permiten tanto al investigador, como al entrevistador y al entrevistado, minimizar los errores producidos por la mala interpretación de las preguntas.
 - ✓ Control, registro y seguimiento eficiente de las preguntas.
 - ✓ Incorporación interactiva de instrucciones para el entrevistador y el entrevistado.
- En cuanto a la validación y depuración del cuestionario:
 - ✓ Filtrado de respuestas, mediante la determinación de edits sobre las preguntas del cuestionario asociado a los ficheros de datos.
 - ✓ Formación de mapas MDS de configuración de variables en bloques de preguntas.
 - ✓ Formación de mapas MDS de la configuración de los ítems y de los sujetos.

(9) Una dimensión alta sólo significaría que el conjunto de ítems tiene volumen máximo en la dimensión propuesta y por tanto poca correlación entre éstos.

- ✓ Proceso interactivo de detección, grabación y manejo de datos.
- ✓ Proceso interactivo de depuración de valores anómalos.
- ✓ Imputación automática de datos faltantes.

En definitiva, QUESTPOT v.1.2. proporciona un entorno amigable y potenciado para el diseño y validación de cuestionarios, ofreciendo una alternativa eficaz, práctica y rápida a los métodos tradicionales de tratamiento de éstos.

NOTAS SOBRE ALGUNOS MÓDULOS DE QUESTPOT V.1.2

Manejador de formularios y codificación de preguntas

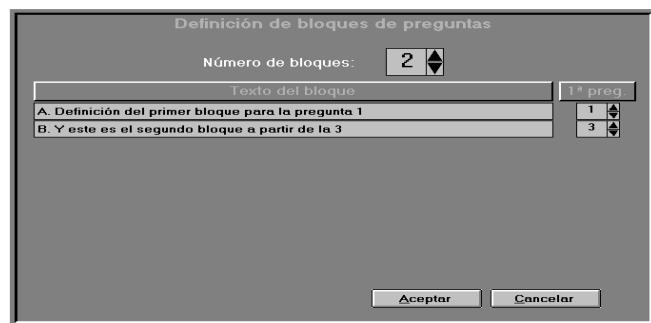
El manejador de formularios no es únicamente un módulo aislado e independiente, sino que se trata de un módulo potenciado, con comunicación mediante interfaz con el manejador, también potenciado, de ficheros de datos, con lo que se garantizan varias acciones, entre ellas:

- Cada pregunta lleva asociada una variable (del tipo que se especifique) que será la que permita, de acuerdo a las especificaciones del generador de preguntas, llevar a cabo los análisis oportunos con posterioridad.
- Se pueden especificar edits para un segundo filtro sobre los datos resultantes, de manera que se pueda depurar convenientemente el fichero de datos creado para el formulario. Asimismo se tiene la posibilidad de generar datos para un formulario de manera interactiva e inteligente, actuando sobre el tipo de dato que, a priori, puede recibir cada variable.
- En este manejador existe la posibilidad de definir bloques homogéneos de preguntas consecutivas, que serán alterados a gusto, con idea de conformar de forma adecuada los grupos.



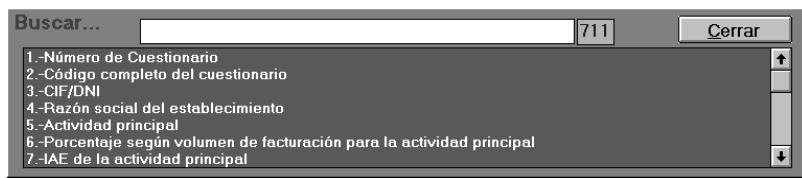
En la figura anterior podemos apreciar cómo cada pregunta lleva asociada un identificativo (nombre de la variable que se utilizará en el fichero de los datos), así como un tipo (en este caso nominal) de entre los posibles que aparecen.

El cuestionario (en este caso un cuestionario de control de la encuesta) consta de 9 preguntas, y se puede en todo momento actuar sobre al mismo nivel de borrar o añadir nuevas, modificar la posición de la pregunta activa, verificar una a una o en conjunto el total de las consideraciones y especificaciones para cada pregunta, así como definir bloques homogéneos (ícono 4, o submenú en el menú Herramientas).



Igualmente se podrá imprimir el cuestionario, con la secuencialidad asignada y definición de bloques establecida. El título del formulario aparecerá en las hojas impresas a modo de identificativo.

Si el cuestionario consta de un número elevado de preguntas, mediante el ícono 7 se puede establecer una búsqueda por filtro, como podemos apreciar en la siguiente figura:



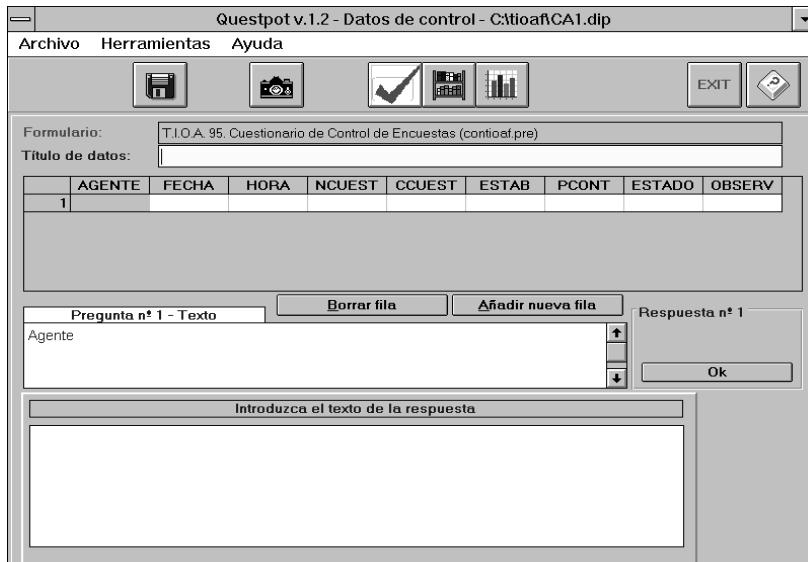
Por último, la codificación de las preguntas se realiza teniendo en cuenta el interfaz posterior con el manejador de ficheros de datos, siempre con respecto al establecimiento de especificaciones válidas y permitidas para con los datos asociados.

Modo de gestión de la encuesta.

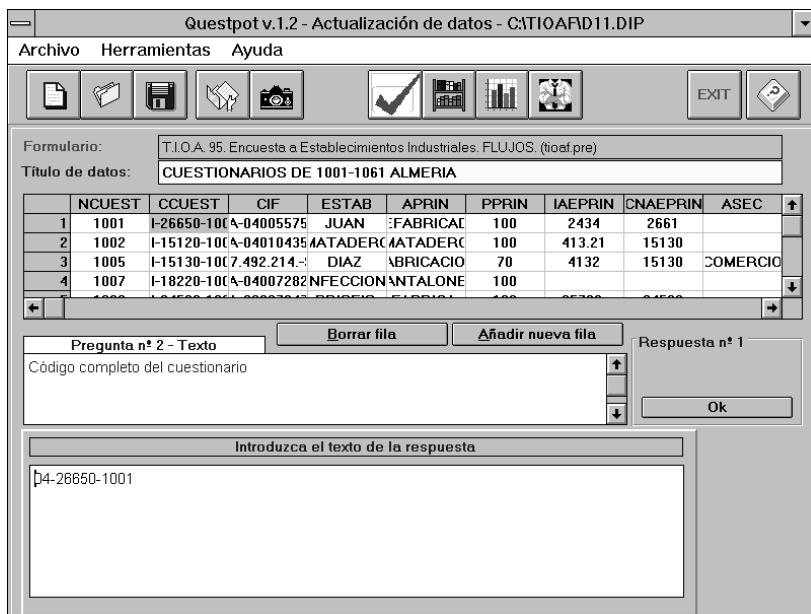
Hasta ahora se han llevado a cabo diseños con atribución dividida y compartida de tareas, y supervisión a tres niveles: el de operador de entrevista asistida, el de inspector de entrevistadores, y el de inspector de encuesta.



Una clave de identificación de usuario permitirá realizar el acceso restringido o abierto. En el segundo caso, el inspector de encuesta tendrá total acceso al sistema, tanto para manejar formularios como para manejar ficheros de datos asociados, y demás módulos interconectados. En el primero, los operadores de entrevista tendrán acceso al manejador de datos para introducir datos referentes al control de la encuesta.



y datos referentes a la encuesta propiamente dicha.



Modo de almacenamiento y requisitos de Hardware

Respecto a la forma de almacenamiento, indicaremos que se trata de ficheros planos matriciales, que conformarán el/los último/s ficheros solicitados. En cada nivel contamos con tipos de ficheros que el usuario puede soportar, tanto para control sobre la encuesta, como para la encuesta en sí. Ello se consigue mediante un cuestionario para el control y el cuestionario propiamente dicho, relacionados y abiertos simultáneamente.

Por lo que respecta a los requisitos de Hardware, el sistema puede instalarse en cualquier entorno informático con entorno Windows a partir de 3.x, así como en redes suficientemente potentes. En las nuevas versiones se elevarán los requisitos anteriores (Windows 9x, ...)

Algunas aplicaciones reales del programa QUESTPOT v.1.2

El programa descrito ha sido “validado” en la práctica, en varias encuestas financiadas por el Centro Andaluz de Prospectiva y por el Instituto de Estadística de

Andalucía, bajo cuyo soporte económico se desarrolló el mismo a través de un proyecto de investigación. Algunas de esas encuestas son:

- La encuesta sobre la Necesidad de la Estadística en la Investigación, llevada a cabo sobre el total de grupos de investigación del Plan Andaluz de Investigación.
- Las TABLAS INPUT-OUTPUT del Sector Industrial de Andalucía, encargadas por el Instituto de Estadística de Andalucía, y que, ha supuesto un reconocimiento de cara a la mejora de la productividad aportada por nuestro sistema, tanto por el número de respuesta alcanzado, utilizando tanto soporte de papel como encuesta interactiva telefónica, como por la calidad de los datos aportados.

REFERENCIAS

- ARABIE, P., CARROLL, J. D. Y DESARBO, W. S. (1987).- *Three-Way Scaling And Clustering*. Sage Publications, Inc.
- APARICIO, F. (1995).- *Tratamiento Informático de encuestas*. RA-MA.
- BETHELEM, J.G. Y KELLER, W.J. (1989).- *New technologies in computer assisted survey processing*. Eustat.
- CONVERSE, J.M. Y PRESSER, S. (1988).- *Survey Questions*. Sage University.
- GONZÁLEZ A. Y MALDONADO J.A. (1991).- *The JANDSTAT Statistical System*. A.S.M.D.A., World Scientific.
- KIECOLT, K.J. Y NATHAN, L.E. (1985).- *Secondary Analysis of Survey Data*. Sage University.
- KRUSKAL, J. B. Y WISH, M. (1978).- *Multidimensional Scaling*. Sage Publications, Inc.
- LITTLE, R. Y RUBIN, D.B. (1992).- *Statistical Analysis of missing data*. J. Wiley & Sons.
- LUQUE, T. (1997).- *Investigación de marketing: Fundamentos*. Ariel Economía.
- MICROSOFT CORP. (1994).- *Visual Basic Profesional*.
- SCHIFFMAN, S. S., REYNOLDS, M. L. Y YOUNG, F. W. (1981).- *Introduction to Multidimensional Scaling. Theory, Methods and Applications*. Academic Press, Inc.

TAKANE, Y., YOUNG, F. Y DE LEEW, J. (1977).- *Nonmetric Individual Differences MDS: An Alternating Least Squares Method with Optimal Scaling Features.* Psychometrika, 42, 7-67.

VILLÁN, I. Y BRAVO, M.S. (1990).- *Procedimientos de depuración de datos estadísticos.* Eustat.

SOME PROBLEMS RELATED TO THE DESIGN AND THE VALIDATION OF QUESTIONNAIRES: TREATMENT USING QUESTPOT V.1.2

SUMMARY

One of the most common problems in survey practice is which refers to input, validate and code data as a previous check to analyse survey data. In this paper we revise a wide set of problems inherent to take data, in order to show how a computer assisted design of questionnaires reduces perceptibly those usual errors. The program QUESTPOT v.1.2, developed to assist in questionnaire design, help us to show the usefulness of these tools in surveys.

Key words: Questionnaire design, CAPI/CATI, QUESTPOT v.1.2, Visual Basic.

AMS Classification: 65U05