

## Verificación de la información obtenida de una encuesta

Alicia Apitz de Parra\* y María Josefina Escalona Fuenmayor\*\*

*Profesoras titulares de La Universidad del Zulia, Escuela de Educación, Centro de Estudios Geográficos, Centro de Estudios Matemáticos de la Facultad de Humanidades y Educación.*

### Resumen

Este trabajo tiene como propósito proporcionar una metodología que permita revisar la información obtenida mediante cuestionarios u hojas de observación para quienes la investigación por encuestas es una experiencia nueva. La propuesta consta de tres pasos o filtros: a) detección de errores, omisiones y valores perdidos; b) identificación de la lógica de las respuestas y c) detección de casos remotos. Entonces la propuesta metodológica prepara la información para que pueda ser analizada adecuadamente.

**Palabras claves:** Encuesta, verificación de información, omisiones, valores perdidos, errores, casos remotos.

### *Checking information obtained by survey*

### Abstract

This paper is addressed to those who are new to survey-based research and provides a methodology for checking information obtained by means of questionnaires or observation sheets. The proposal consists of three steps or filters: a) detection of mistakes, omissions and missing value; b) identification of the logic in the answers and c) detection of the outliers. Thus the information will be ready for adequate analysis.

**Key words:** Survey, checks survey data, missing value, mistakes, outliers.

Recibido: 5-3-96 • Aceptado: 15-11-96

\* Licenciada en Educación, mención Ciencias Sociales, LUZ. Magister en Transporte, Fc. de Ingeniería, LUZ.

\*\* Licenciada en Educación, mención Ciencias Matemáticas, LUZ. Magister en Matemáticas Aplicadas, Facultad de Ingeniería, LUZ.

E-mail: mescalon@luz.ve. E-mail: CEMAT@luz.ve

## Introducción

En la actualidad las "encuestas" han pasado a constituir elementos de uso común en la mayoría de las investigaciones en Ciencias Sociales y Educativas. Su racionalidad queda sujeta a la orientación consciente de las personas que las llevan a cabo, quienes tratarán de cumplir con normas metodológicas a lo largo de las diversas fases del proceso de la encuesta, ejecutadas generalmente en seis etapas interdependientes: planificación, muestreo, el diseño del cuestionario, el trabajo de campo, la verificación y la codificación, el análisis y el informe.

Indudablemente, que todas y cada una de las etapas tienen fundamental importancia. Sin embargo, el presente artículo tiene como propósito proporcionar una metodología específica que permita verificar la información obtenida de una encuesta y la prepare adecuadamente para la aplicación del método estadístico correspondiente. Para tal efecto, este trabajo pretende: 1) Elaborar una metodología que permita verificar la información obtenida de una encuesta a través de la detección de respuestas faltantes (omisiones) y errores de un cuestionario; 2) Diseñar una metodología que permita verificar la información obtenida de una encuesta a través de la identificación de la lógica de las respuestas de los ítems de un cuestionario y; 3) Presentar una metodología que permita verificar la información obtenida de una encuesta a través de la identificación de casos remotos (outliers) en un cuestionario. (Parra, 1986).

## Algunos Aspectos Conceptuales

El proceso de verificación, denominado por Miranda (1976) Proceso Crítico, involucra la información obtenida de un instrumento de observación o medición (encuesta, cuestionario), la cual debe someterse a normas de depuración.

En la organización y programación de una investigación, cualquiera sea su naturaleza, se impone como conducta racional, preparar primero el plan y ejecutarlo después.

En el planeamiento se incluye todo lo relativo a la organización y programación de la etapa de recolección de la información, y la ejecución comprende una serie de actividades reunidas bajo el nombre de operaciones de encuesta; obviamente hay un mutuo efecto entre el diseño y las operaciones de encuesta (Simmons, 1971). En este sentido, el comportamiento de una encuesta, que carezca en sí misma de controles de calidad de operaciones no será nunca un reflejo del diseño. De lo antes planteado, surgen las siguientes reflexiones: si la encuesta no está bajo control cuidadoso se gastará inútilmente tiempo y dinero, y los resultados serán de validez dudosa y desconocida; igualmente la encuesta no producirá información de buena calidad y en realidad podrá dar información equivocada, por lo que, el objetivo esperado no será alcanzado. Una encuesta, por lo tanto, "sólo puede tener éxito cuando se crean procedimientos que aseguren en conjunto que las desviaciones del diseño se mantendrán dentro de límites razona-

bles. Este es el principio del control de operaciones<sup>4</sup> (Simmons, 1972).

Así un sistema general de control de operaciones comprende las siguientes fases: reconciliación del diseño y los objetivos; verificación del diseño teórico; control sobre la recolección de los datos; control del procesamiento de datos; control de la tabulación y estimación y finalmente, el control sobre el manuscrito del informe y la publicación. Otros investigadores señalan diversas dificultades que afectan la calidad de los datos en proyectos de investigación (Strayhorn, 1990).

De las fases antes mencionadas, tiene especial significación en el presente trabajo la relativa al control de procesamiento de datos y dentro de ésta, específicamente la fase que se conoce con el nombre de verificación de la información. Esta última "es un medio para salvar omisiones y corregir errores producidos durante la recolección de los datos" (Miranda, 1976). La información que se obtiene de la recolección, adolece generalmente de errores y omisiones. Estas deficiencias según sea su magnitud puede llegar a afectar en forma importante los resultados de la investigación, ya sea en algunas de las categorías de clasificación en que se presenta, como también en los resultados globales.

Las principales deficiencias que podrían producirse en la recolección de información son las omisiones, que no son otra cosa que la falta de respuesta a una o varias preguntas, o a la totalidad del cuestionario. Wang, Sedransk y Jinn (1992) reseñan algunos tipos de omisiones registradas en censos y encuestas. La otra deficiencia está referida a los errores, según Sierra Bravo (1981) el error es la falta de concordancia entre la

información consignada en el cuestionario y el dato real.

Se pueden dividir los errores científicos o referentes a la investigación científica en dos grandes grupos: al primer grupo de errores se le puede atribuir carácter lógico, de raciocinio y al segundo, un carácter matemático, en cuanto se puede intentar medirlos y expresarlos mediante números (Sierra, 1981).

Cuando la información es sometida a un proceso de verificación debe prestarse especial atención a los siguientes casos de errores: respuestas que no corresponden a la pregunta; respuestas en sentido general que no satisfacen el grado de especificación; respuestas contradictorias; mala interpretación de conceptos y definiciones y por último cuando se producen respuestas ilegibles por mala escritura.

Indudablemente, que lo ideal en la ejecución de una investigación es conseguir que la información obtenida del campo, refleje al máximo la realidad. Sin embargo, hay casos donde la magnitud de la investigación conllevan a una fase de verificación de la información que no solucionará todos los defectos producidos durante la recolección. Por ello, "los trabajos de crítica pueden limitarse a cubrir los casos en que el error o la omisión afecta la significación de los resultados, atendiendo al uso que se le pretenda dar" (Miranda, 1976).

## **Normas para realizar la verificación**

Existen diversas maneras para ubicar las omisiones y los errores.

- a) Para el caso de las omisiones, las alternativas para cada pregunta son:
- Si necesariamente ha de tener respuesta. En una investigación estadística, siempre existen determinadas preguntas que deben tener respuesta, por ejemplo: tipo de vehículo, en las encuestas de transporte terrestre de pasajeros.
  - Si ha de tener respuestas conforme a otros antecedentes del cuestionario. Hay casos donde algunas preguntas deben o no tener respuesta, tomándose como consideraciones otras características conocidas por otras preguntas incluidas en el mismo cuestionario, por ejemplo, a una persona que informe ser un trabajador remunerado se le consulta por sus ingresos. En este caso, si a la pregunta cuya respuesta se ha omitido, se le han asignado menos grados de confianza (prioridad de aceptación) que a la pregunta con la cual se la relaciona, entonces se considera que hay omisión; en el caso contrario, se debe investigar la validez de la respuesta existente antes de adoptar una decisión definitiva (Miranda, 1976).
  - Si puede o no tener respuesta. Hay situaciones donde algunas preguntas pueden indistintamente tener o no respuesta, no dependiendo esta posibilidad de otros antecedentes contenidos en el cuestionario; por ejemplo, cuando se pregunta a un usuario de un terminal terrestre de pasajeros emita su opinión sobre el funcionamiento del servicio telefónico. En este caso, si el usuario no ha

utilizado tal servicio, la pregunta queda sin respuesta.

- b) Con respecto a los errores, estos se ubican estableciendo controles de consistencia para las respuestas a las diferentes preguntas. Estos controles tienen como base determinar si existe concordancia entre las respuestas de un mismo informante (entrevistado) y que al mismo tiempo estén dentro de lo lógicamente esperado en el fenómeno que se estudia (Miranda, 1976) o lo que Zarkovich (1968) denomina sistema de trabajo adoptado.

Según sean las características de la información verificada se establecerán controles de consistencia, los mismos son de diferentes tipos, a título de ilustración se pueden señalar los siguientes:

- Información relativa a una misma característica, consultada en dos o más preguntas del mismo cuestionario (lo que se conoce con el nombre de preguntas de control. Un ejemplo estaría dado en las encuestas de naturaleza demográfica, aparece en primer término la pregunta referida a la edad de la persona y más adelante se incluye una pregunta en que se consulta su fecha de nacimiento.

Por otro lado, si lo que se trata es de salvar omisiones y corregir errores, las normas a seguir se fundamentan en dos sistemas. El primero está referido a la investigación del problema y se hace a través de las diferentes operaciones por las que ha pasado la información, llegando incluso a efectuarse una nueva consulta al informante. El segundo sistema tiene por objeto efectuar la asignación del dato faltante o el reemplazo de la información errada por medio de sistemas de

corrección basados, generalmente, en métodos cuantitativos o, en casos muy especiales, en apreciaciones de especialistas en el tipo de información de que se trate, (Wang, Sedransk y Jinn, 1992). Para los efectos de la corrección de informaciones erradas u omitidas se confronta con el mismo sistema (Miranda, 1976).

Para efectuar la asignación de información se utilizan diversos métodos, dependiendo su elección, entre otros, de los medios disponibles para efectuar la verificación, sea a través de funcionarios especializados o mediante el uso de análisis de datos secundarios (Wang, Sedransk y Jinn, 1992).

Entre los métodos más utilizados, Miranda señala en su manual "Planeamiento y ejecución de encuestas" los siguientes. El método deductivo, el cual asigna la información más probable determinada mediante el análisis lógico de los datos existentes en el cuestionario y el sistema de trabajo adoptado, en un marco más amplio cuando corresponda. El método aleatorio, en el cual la información asignable se obtiene mediante un número al azar aplicado a las respuestas posibles ponderadas de acuerdo a su frecuencia de aparición. Un tercer método es el sistemático, el cual consiste en ir asignando las posibles respuestas en forma sistemática. En caso de disponer de un computador, existen paquetes de programas que permiten realizar esta labor.

La verificación puede ser llevada a cabo mediante tres formas: utilizando personal especializado, con el uso del computador y a través de una serie de pasos o filtros que permitan decidir sobre la acción correctiva para cada caso en particular. Esta última constituye la pro-

puesta metodológica de las autoras del presente trabajo.

Cuando la verificación es efectuada en forma manual, es necesario disponer de un equipo de personas expertas, quienes podrían efectuarla rápida y eficientemente si la corrección es directamente lógica y el número de pasos no es muy grande. La tarea consiste en ir ubicando los errores y omisiones revisando directa y cuidadosamente las respuestas a cada una de las preguntas del cuestionario consideradas dentro de la extensión de la verificación; las correcciones correspondientes se hacen de acuerdo a todo un plan metodológico previamente establecido. En tal sentido, "deberá prepararse un manual de hojas sueltas con las instrucciones para la verificación. El primer borrador de este manual deberá estar listo en el momento en que se hacen las pruebas preliminares y el ensayo de la encuesta. Todos los aspectos referentes a la crítica deberán estar tratados en el manual. El formato de hojas sueltas es más conveniente; debido a que permite introducir agregados y cambios a medida que se acumula más experiencia o surgen nuevas situaciones. Para referencia futura es conveniente que cada hoja lleve una fecha que indique el momento en que se incorporó al manual". (Simmons, 1972).

Cuando la verificación utiliza un computador electrónico, previamente la información debe estar codificada, a fin de poder ingresarla mediante algún sistema de entrada de información. Hoy, existen diversidad de equipos y programas utilizados para almacenar gran cantidad de información.

Cuando la verificación de la información se realiza mediante el empleo de



computadoras es necesario la preparación y uso de programas que contengan las instrucciones necesarias para detectar las omisiones y errores, los que pueden ser corregidos mediante asignaciones programadas, o por personal especializado, previo análisis e investigación del problema.

El empleo de la computadora en la realización de la verificación, en lugar de usar personal especializado, tiene grandes ventajas (en algunos casos) de rapidez y economía. Esto compensa el costo de programación y otros que demanda el uso del computador, sobre todo, cuando la operación de verificación comprende gran cantidad de información, y la revisión y corrección se pueden hacer en un alto porcentaje mediante métodos trad-

cibles a programas que permitan un alto grado de automatización del trabajo.

Así existen en el mercado potentes programas computacionales que permiten tal tarea. En el presente trabajo algunos paquetes serán objeto de aplicación a fin de ilustrar la metodología propuesta.

## Propuesta

Por lo común tropezamos con ciertos datos equivocados que no corresponden al valor verdadero que se obtendría a la aplicación de una encuesta correctamente, por tanto deberá decidirse la acción idónea para cada caso en particular. Sin embargo, es posible establecer los siguientes pasos o filtros contemplados dentro de esta propuesta metodológica (ver Figura 1).

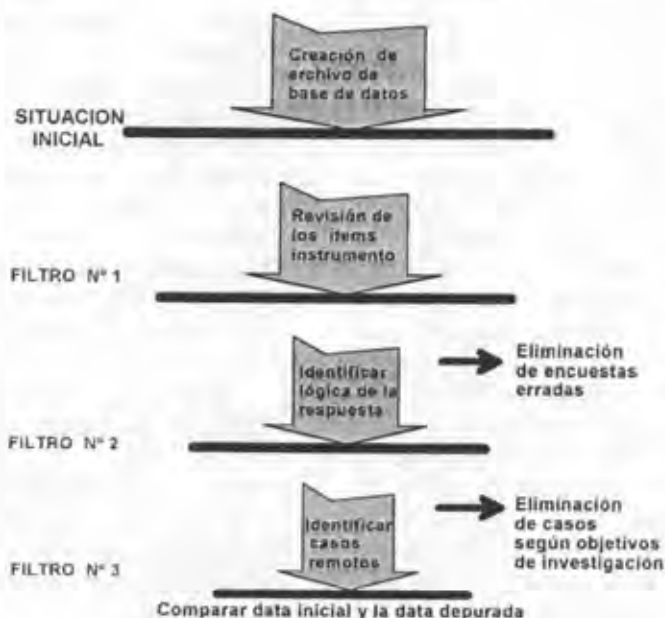


Figura 1. Pasos de la metodología propuesta

**Filtro No. 1:** consiste en revisar las respuestas de cada uno de los ítemes contenidos en el cuestionario a fin de detectar respuestas faltantes. Entre los métodos disponibles para manejar las omisiones ocasionales, se tienen:

- Inclusión de las omisiones en categorías especiales: el procesamiento más simple para manejar las respuestas omitidas a una pregunta es incluir una categoría especial en las tablas para respuestas no determinadas. Con esta información, el lector puede formarse su propio juicio sobre los efectos del ítem respuestas omitidas en la interpretación obtenida de una tabla determinada. Dependiendo de la dificultad y delicadeza de las preguntas, la frecuencia de las respuestas omitidas puede ser menos de medio o uno por ciento hasta 10 ó 15%, si se incluyen las respuestas "NO SE". Cuando los porcentajes son bastantes altos, como es frecuente en el caso de indagación sobre opinión pública, es especialmente importante que esta información se declare. (Lininger y Warwick, 1984).

- Distribución de los casos de omisión de información: este procedimiento incluye la distribución proporcional de los casos de información omitida entre categorías de información conocida. La ventaja de este procedimiento de distribución es que los casos de información omitida se distribuyen proporcionalmente entre las categorías de respuestas regulares, lo cual no se justifica en algunos casos, por ejemplo, cuando se trata de informaciones acerca de las características socioeconómicas de la población. Por lo tanto, es preferible declarar la información omitida en una categoría separada o proporcionar alguna indicación del número de casos que entran en la categoría de infor-

mación omitida, en vez de ocultar el volumen del ítem de respuestas omitidas.

- Asignación de los casos de información omitida: una tercera forma de manejo de ítem de información omitida es utilizar otra información para asignar los casos de datos que faltan a una u otra categoría de respuestas regulares. En el texto "La Encuesta por Muestreo: Teoría y Práctica", de Lininger y Warwick, hay un ejemplo ilustrativo que será reproducido, pues se ajusta a los propósitos de esta sección. Se parte del conocimiento de varias de las características del entrevistado, usando parte de esta información para apostar cuál sería su respuesta a una pregunta determinada; tal es el caso de una persona del sexo femenino que no contestó a una pregunta sobre si trabaja o no fuera del hogar. En lugar de anotar el caso como "no respondió" o distribuirlo con base en las respuestas promedio, se puede utilizar la información conocida a través del cuestionario para estimar si trabaja fuera del hogar o no, en este ejemplo, "trabaja" o no "trabaja". Entonces, el primer paso es decidir cuáles características "conocidas" se van a utilizar para hacer la apuesta, es decir, calcular las probabilidades de una respuesta "trabaja" o no "trabaja" a la pregunta. Algunas buenas posibilidades serían la edad, educación, número de hijos, edad del hijo menor y el ingreso del cónyuge. El segundo paso es utilizar todos los casos "conocidos" para calcular el porcentaje de esposas que "trabajan" o no "trabajan". La información resultante puede interpretarse en forma de una tabla de contingencia que exprese las probabilidades de que las esposas trabajen, puede aparecer como se indica a continuación:

Tabla 1

Ciclo de Vida Familiar	Porcentaje que trabaja	Porcentaje que no trabaja	Porcentaje Total
1. Pareja recién casados sin hijos	75	25	100
2. Matrimonio, hijos en edad preescolar	20	80	100
3. Matrimonio, hijos en edad escolar	50	50	100

Fuente: Lininger y Warwick, 1984.

Volviendo al caso particular que se está ejemplificando, la entrevistada está casada, viviendo con su marido y tiene un hijo en edad preescolar. La tabla de contingencia muestra que un 20% de las mujeres en esta categoría están trabajando, mientras que un 80% no lo están. Para realizar el proceso de asignación se debe escoger una cifra aleatoria que corresponda al campo de variedad de los porcentajes, o sea, de 00 a 99. Si esta cifra cae la escala de 0 a 19, se asignaría a la esposa a la categoría "trabajo"; si es de 20 a 99, se asignaría a la categoría "no trabaja". Puede, entonces, corregirse registrando este dato en lo correspondiente a "no respondió". Es un procedimiento lento, que puede ser simplificado con el uso de computadores, siempre y cuando la información omitida sea de gran importancia para los planes de análisis del estudio. En todo caso, este procedimiento no resuelve el problema de la información omitida, simplemente reduce los inconvenientes que crea. Sin embargo, merece ser considerado para muchos estudios.

- A través de métodos computacionales

La gran mayoría de las recolecciones de datos incluyen "valores perdidos". Por ejemplo, si el entrevistado en una encuesta falla en responder a una pregunta dada, un analista puede clasificar el valor correspondiente a esa pregunta como "perdido". También se producen valores no identificables cuando el transcriptor de datos introduce la información al computador.

Una manera de corregir este tipo de error es a través del procedimiento Missing de algunos paquetes estadísticos. Valores como estos, pueden ser reconocidos como perdidos en el momento que el programa computacional estadístico los lee. Otros valores perdidos pueden ser producidos durante los pasos de la información; es posible usar instrucciones para asignar valores perdidos a las variables, algunos sistemas computacionales para estadísticas pueden automáticamente generar valores perdidos para variables cuando se encuentran ciertas condiciones.

La figura 2 ilustra cómo corregir este tipo de error usando paquetes de tratamiento estadístico o de administración de bases de datos.



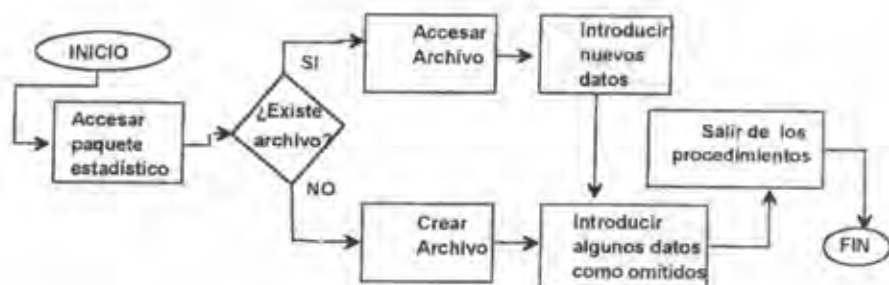


Figura 2

Filtro No. 2: en este caso se debe detectar en las distintas secciones del cuestionario, la lógica que debe existir entre las respuestas de sus ítems. Para ello, es posible usar estudios de consistencia, ya sean externos o internos. Los primeros "tienen por objeto ver cómo los datos recogidos en una encuesta concuerdan con algún conocimiento aceptado generalmente. La verificación de la calidad de los datos por medio de un estudio de consistencia interna implica una investigación del grado en que las estimaciones de las diferentes variables describen el fenómeno de la misma forma o del grado en que concuerdan con otras variables con las que deben ser congruentes matemáticamente o lógicamente" (Sierra, 1981).

La Figura 3 muestra un diagrama de flujo a fin de ilustrar sobre cómo puede ser abordado este problema en programas de computación.

**Filtro No. 3:** Detección de casos remotos (Outliers). Existen diversos métodos para manejar los casos remotos. Así en un cuerpo o conjunto de datos numeroso, los modelos son recursos de representación moderada que pueden guiar a interpretaciones simples de los datos, por lo que, es muy importante tener un significado de las apropiadas y sensibles estimaciones de los modelos bajo consideración. La utilidad de los residuales en el análisis de modelos de regresión lineal debe ser reconocido ampliamente. (Gnanadesikan, 1977).

Entonces, es posible utilizar los residuales univariados para detectar los llamados "outliers" u observaciones extremadamente desviadas, las cuales no son comunes en conjuntos muy grandes de datos. En tal sentido, en un gráfico de residuales, estos outliers son puntos que están ubicados lejos de la dispersión del resto de los residuales, por ejemplo, para



Figura 3

Figura 3. Diagrama de lógica de las respuestas

cuatro o más desviaciones estándar cero (puntos que coinciden con la recta), observamos en la Figura 4 un residual estandarizado encerrado en un círculo.

Por lo tanto, se hace necesario identificar en las diferentes secciones del

cuestionario la presencia de casos remotos. Otro método para manejar estos casos se ilustra en el manual de Simmons "El Control de Operaciones en las Encuestas por Muestras", el cual por consi-



donde:  
 $e_i = Y_i - \hat{Y}_i$  (Diferencia entre valor observado y el valor ajustado)  
 MSE: es la varianza de  $\sqrt{\text{residuales } e_i}$ .

Fuente: Wasserman, W., p. 106

Figura 4 Ejemplo ilustrativo de casos remotos usando diagrama residual

derarlo didáctico se presentará a continuación.

El cuadro siguiente es parte de una información de una encuesta sometida al proceso de verificación.

El ejemplo ilustrado señala el número de hijos habidos de mujeres de 20 a 24 años. En el cuadro no se observa la anotación "13 hijos", porque constituye un caso remoto. Teóricamente podría ser un dato real, pero muy improbable y está separado del resto de una distribución pareja. Aún cuando en la muestra hubiera habido un caso con trece hijos, la expansión de esa cifra al nivel del universo conduciría, casi con seguridad, a conclusiones erróneas. Este método para manejar los casos remotos consiste en reemplazar el caso remoto por otro valor extraído al azar de la distribución recogida, una vez que se ha excluido de ella el

caso en discusión. El procedimiento para hacer esto, tomando como base el ejemplo del cuadro anterior, comprende los siguientes pasos:

a) Construir una distribución de frecuencias acumuladas (porcentual) en la forma presentada en la última columna del cuadro.

b) Extraer un número al azar entre uno y cien.

c) Comparar ese número con la distribución acumulada. Digamos que el número es 63, dado que es mayor que 52, pero menor que 75, se trata de un caso semejante a los de la clase "mujer con tres hijos".

d) Reemplazar el número 13 por el número 3 en el registro correspondiente al caso remoto.

La computadora puede hacer muy rápidamente este tipo de corrección,

Tabla 2

Número de Hijos	Número de mujeres que declararon este número de hijos	Distribución acumulada ajustada	
		Número	Por ciento
0	10	10	7
1	20	30	22
2	40	70	52
3	30	100	75
4	15	115	86
5	8	123	92
6	4	127	95
7	4	131	98
8	2	133	99
9	1	134	100
10	0	134	
11	0	134	

cuando la fase de prueba y corrección está programada para seguir una tabulación preliminar de los datos registrados. En el ejemplo anterior, el procedimiento se aplicaría siempre que el número de hijos declarados fuera más de nueve.

Entonces, si se selecciona un método computacional para hacer este tipo de corrección, la Figura 5 ilustra un diagrama de flujo que contiene los pasos fundamentales a dar.

Existe otra forma de detectar casos remotos. En tal sentido, y siguiendo con el mecanismo de elaboración de gráficos para utilizarlos en la verificación de la información obtenida de una encuesta, se presenta el gráfico llamado Diagrama Q-Q (cuantil contra cuantil), el cual puede

ser usado para evaluar la suposición de normalidad. Hay, en efecto, un diagrama del cuantil muestral (observaciones de la variable aleatoria) contra el cuantil de algo esperado por observar, si las observaciones están distribuidas normalmente. Cuando los puntos caen cerca de una línea recta, el supuesto de normalidad es definible. La normalidad es sospechosa cuando los puntos se desvían de la recta. Más aún, las formas de las desviaciones pueden indicar el camino acerca de la naturaleza de la no normalidad. Una de las razones para la no normalidad es identificar acciones correctivas cuando sea posible (Johnson & Wichern, 1982).

Un computador puede usarse para evaluar los pares  $(q(i), x(i))$  y para cons-

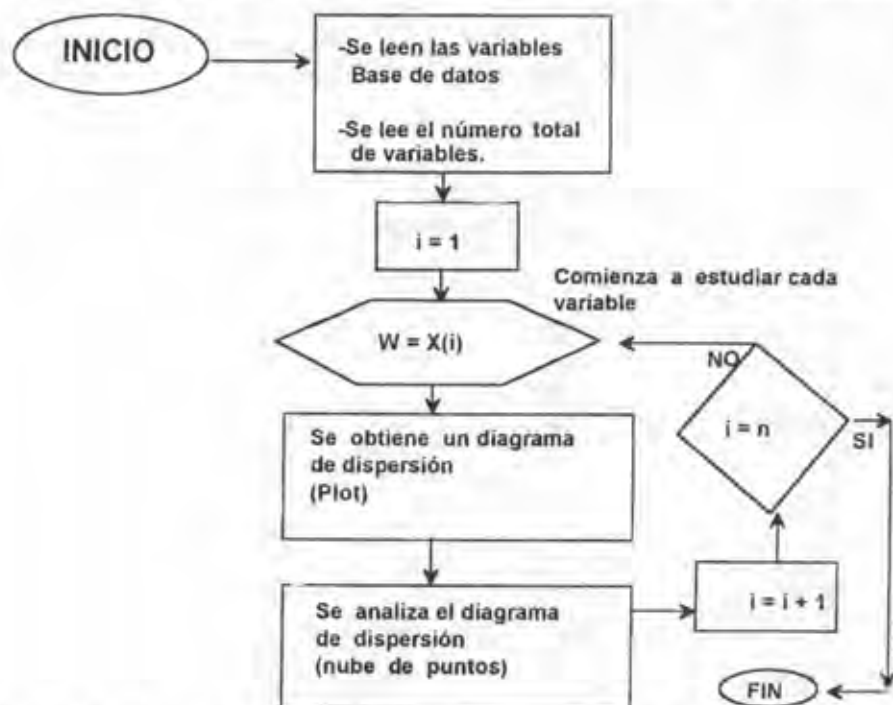


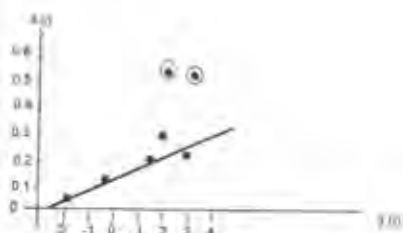
Figura 5. Diagrama de observaciones de casos remotos

truir un diagrama Q-Q, el cual se ilustra en la Figura 6. En él se aprecia que los datos no tienen una distribución normal, en el gráfico los puntos encerrados en un círculo son casos remotos (outliers), valores que están demasiado alejados del resto de las observaciones.

Para finalizar, una vez verificada la información a través de la estrategia metodológica propuesta, se está listo para realizar los análisis estadísticos a que diere lugar, por ejemplo:

- a) Detectar relación y correlación entre variables.
- b) Realizar análisis descriptivo de datos univariados y multivariados.
- c) Pruebas de hipótesis.

Es conveniente destacar que hay estudios donde no es recomendable la eliminación o sustitución de los outliers, porque la naturaleza del mismo así lo requiere.



Fuente: Johnson y Wichern, p. 155.

**Figura 6** Ejemplo ilustrativo de un caso remoto usando gráfico "Diagrama Q-Q" (para evaluar la suposición de normalidad)

## Contribución del método

La propuesta metodológica es aplicable a las informaciones obtenidas a través de técnica de la encuesta, cualquiera sea su caso. Así, al aplicarse la metodología a cualquier estudio seleccionado, puede afectar considerablemente la representatividad de la muestra, reafirmando la necesidad de establecer controles en las distintas fases de una encuesta, especialmente a aquellas que pueden afectar la calidad de la información.

## Referencias Bibliográficas

- Gnanadesikan, R. (1977). *Methods for Statistical data Analysis of Multivariate Observations*. Edit: John Wiley & son.
- Johnson, R. y Wichern, D. (1982). *Applied Multivariate Statistical Analysis*. Edit: Prentice-Hall, New Jersey, USA; 1982, p.p. 594.
- Lininger y Warwick (1984). *La encuesta por muestreo. Teoría y Práctica*. Ed. CECSA, 4ta. reimpresión, 442 pp.
- Miranda, O. (1976). *Planeamiento y Ejecución de Encuestas*. Centro Interamericano de Enseñanza de Estadística, p.p. 147.
- Parra, Alicia (1986). *Estrategia metodológica para depurar de la información obtenida de una encuesta*. Facultad de Humanidades y Educación, Universidad del Zulia. Maracaibo, 231 pp.
- Sierra Bravo, R. (1981). *Ciencias Sociales, Análisis Estadístico y modelos Matemáticos*. Edit. Paraninfo. Madrid, pp. 482.
7. Simmons, W. (1972). *El Control de las Operaciones en las Encuestas por Muestra*. Manual 2. Programa Inter-



nacional de Estadísticas de Población, USA, pp. 65.

Strayhorn (1990). **Estimating the errors remaining in a data set: Techniques for quality control**, *American Statisticians*, vol. 44, No. 1.

Wang, Sedransk and Jinn (1992). **Secondary data analysis when there are missing observation**. *JASA*, vol. 87, No. 420, pp. 452-961.

Wasserman, J. (1974). **Applied Multivariate Statistical Models**, Ed. Richard Erwin, Homewood, 842 pp.

Zarkovich, S. (1975). **Calidad de los datos estadísticos**. FAO, Roma, reimpresión, 442 pp.

#### **Programas Estadísticos**

**SAS user's guide basics** (1985). Institute Statistical. Analysis System, Versión 5 editions. North Caroline, USA. 1290 pp.

**SAS user's guide statistics** (1985). Institute Statistical. Analysis System, Versión 5 editions. North Caroline, USA. 956 pp.

**Statgraphic, user manual** (1992). Manugistics, Cambridge-MA, 6 ed.