

**COMENTARIOS Y EXTENSIONES A 'INTERFACES  
GRÁFICOS DE USUARIO PARA ANÁLISIS DE DATOS'  
DE P.M. VALERO.**

Manuel Trujillo  
*Instituto de Estudios Sociales  
Avanzados de Andalucía*

Andrés González  
*Universidad de Granada*

Vidal Díaz  
*Universidad de Navarra*

Vicente Manzano  
*Universidad de Sevilla*

Josep Múria  
*T.E.S.I.*

Manuel Trujillo

*Trabajo bien hecho o clientes satisfechos*

---

Pedro Valero hace una análisis de la evolución, el desarrollo actual y el posible futuro de los programas de análisis de datos.

En él destaca como posible camino de futuro la conversión del lenguaje R en una alternativa real al software comercial, con la incorporación de unas interfaces gráficas que hagan menos complicado su uso, ya que con este lenguaje se pueden encontrar incluso ahora mismo más variedad de análisis, y más avanzados, además de su gratuidad.

Sin embargo, si bien es cierto que los usuarios más avezados en análisis de datos siempre querrán utilizar la técnica más avanzada, que mejor resuelva su problema, para los cuales R se ha convertido en una ventajosa alternativa, existe otro tipo de usuario, más importante en número probablemente, cuyos conocimientos sobre análisis de datos, o sobre estadística en general son más básicos.

Este tipo de usuario se puede decir que ha aparecido con el software comercial, puesto que nunca hubiese podido programar sus propias rutinas, y no parece que vaya a desaparecer, sino que cada vez es más numeroso. Su preocupación no es poder aplicar el

método que conoce más adecuado para su problema, sino que primero debe saber qué técnica debe utilizar (incluso a veces primero debe averiguar cual es su problema). Por ello lo que espera más bien es un programa que le facilite al máximo la vida, actúe de asesor estadístico, y le proponga solo unos pocos análisis que aprenderse y donde elegir, proporcionándole además bonitos resultados.

Es difícil que alguna vez R pueda llegar a satisfacer a ese público, ya que la misma idea probablemente repugne a la mayor parte de los científicos que desarrollan el lenguaje, además de que por su naturaleza descentralizada siempre tenderá a ofrecer una mayor gama de posibilidades.

Esta es realmente la clientela de los paquetes comerciales que, nos guste o no, seguirá existiendo. Por ello los desarrolladores ya incluyen cada vez más asesores estadísticos que guían a los usuarios sobre las técnicas que deben aplicar, e incluso les interpretan los resultados. Y como los que desarrollan los paquetes no quieren precisamente perder clientes, nunca incluirán como conclusión que no es posible realizar ningún análisis de los incluidos en el paquete con los datos que se disponen, y debe acudirse a otro programa. Así que los usuarios serán felices con sus resultados, sean correctos o no, y seguirán utilizando las técnicas que se incluyen en el paquete, que por otra parte serán consideradas las únicas correctas en amplios entornos porque son las únicas que conocen.

---

Andrés González

*Software, hardware... y brainware*

Es difícil imaginar el futuro, aunque la experiencia demuestra que es posible que nos quedemos cortos por muy optimistas que puedan parecer nuestras predicciones. No parece descabellado suponer que dentro de unos años existirán más y mejores programas informáticos, y que los ordenadores aumentarán sus posibilidades a un ritmo sólo comparable al que disminuye su precio y su tamaño. Sin embargo, el desarrollo del hardware, y más aún del software puede tener consecuencias negativas, sobre todo si no va bien acompañado de lo que Tukey (1965) denominó brainware.

Con el término brainware se refería Tukey a las mentes que “sabían lo que se buscaba” cuando se utilizaba el hardware y el software. Justamente es ahí donde puede estar uno de los problemas en el desarrollo informático: la posibilidad creciente de que los programas sean usados sin saber lo que se busca. El aumento en la disponibilidad de herramientas en el ámbito de las encuestas puede generar problemas similares a los acontecidos con el uso de la prueba de significación de la hipótesis nula (p.e. Ares 1999). No resulta absurdo imaginar una campaña publicitaria de software anunciando que es “tan sencillo que hasta un niño podría utilizarlo”. No quiero con esto defender la existencia de programas poco amigables o difíciles de utilizar, sino señalar que la facilidad y disponibilidad de los mismos puede instaurar toda una serie de malos hábitos que en poco ayudarían al desarrollo de la investigación mediante encuestas.

Dejando momentáneamente a un lado la necesidad de un brainware que se relacione adecuadamente con el hardware y el software, podemos comentar algunos aspectos sobre el desarrollo esperable de estos últimos.

Una investigación mediante encuesta es algo más, evidentemente, que un análisis de datos. Por lo tanto, el software para investigación mediante encuesta debe ser algo más que el software para análisis de datos. Como señala Arnau (1995), el nivel estadístico es simplemente uno de los tres niveles que conforman el método científico junto a los niveles teórico y técnico. No tiene sentido desarrollarlos por separado, como no tiene sentido analizar los datos de una encuesta independientemente de cómo fueron obtenidos o del objetivo que se persigue. Por muy potente, fácil de usar o versátil que resulte un programa informático de análisis de datos siempre estará sometido al conocido problema GIGO (Garbage in- garbage out- si entra basura sale basura). Es decir, si en el apartado teórico no se definieron adecuadamente las variables o el nivel técnico no se recogieron adecuadamente los datos, poca utilidad podrá tener el mejor de los programas de análisis de datos.

Puestos a imaginar, el software ideal para encuestas debería empezar a utilizarse en el momento que se inicia la investigación y ser de utilidad hasta que ésta finaliza. Debería facilitar, no sólo las tareas de análisis de datos, sino el resto de decisiones que debe tomar el investigador, desde la especificación del nivel de medida de distintas variables hasta la determinación del diseño muestral, o el análisis de la calidad de los cuestionarios utilizados. Aunque puede parecer que ya existen desde hace tiempo en el mercado programas con esas características en mi opinión esto no es así. El apoyo que el software puede prestar en el futuro a la elaboración de cuestionarios es mucho mayor de lo que se realiza actualmente. En cierta medida los módulos actuales son poco más que editores de texto integrados en el programa. Un buen software de elaboración del cuestionarios detectaría por ejemplo, si el número de alternativas de una pregunta es demasiado elevado para una encuesta telefónica, o si existen cuantificadores no apropiados, etc. Es decir, el software incorporaría el conocimiento acumulado sobre elaboración de cuestionarios y actuaría a modo de sistema experto. Y lo mismo cabría decir sobre el resto de aspectos de la investigación.

No es posible considerar software y hardware como elementos independientes. Sobre todo en el momento de realización de la entrevista en los modelos CAPI y CATI. Siguiendo con la mente libre, imagino sistemas con localizador GPS que registren el camino realizado por el entrevistador y el lugar de realización de la entrevista, o portátiles con conexión *wifi* en los que se pudiese rediseñar a distancia e instantáneamente desde el cuestionario hasta la composición de la muestra.

Pero, como señalé al principio, hardware y software necesitan de un brainware que señale el objetivo que se busca. Haciendo un fácil juego de palabras las opciones por defecto son un gran defecto. Algo similar ya fue planteado por Bakan (1966, tomado de Borges 1999) al afirmar que “Cuando llegamos a un punto en que los procedimientos estadísticos se convierten en sustitutos del pensamiento en lugar de ayudas a él, ha llegado el momento de regresar al camino del sentido común”.

Casi 40 años después, parece que simplemente ha aumentado el número de sustitutos del pensamiento para incluir algo más que los procedimientos estadísticos, y que el regreso al camino del sentido común aún no ha comenzado. Esperemos que el futuro no sea sólo en esa dirección.

---

*Vidal Díaz El futuro del software para análisis de encuestas: algunas reflexiones*

---

Comenzaré, como no puede ser de otra forma, agradeciendo la invitación de Vicente Manzano de presentar mis reflexiones sobre “qué va a ocurrir (o qué debiera ocurrir) con el software de análisis para encuestas en un futuro”. Es un placer hacerlo para una de las personas que más me han enseñado sobre el tema.

Será necesario, en primer lugar, definir el “lugar” del que hablo, el punto de partida de mi reflexión. Mi enfoque está más centrado en las demandas del mercado que en el ámbito académico; y he adoptado este punto de partida porque considero que es el mercado –y no la academia– el que constituye la mayor parte de la demanda –y el desarrollo– de estos programas. Dentro del *mercado* me refiero, fundamentalmente, a las empresas de investigación de opinión y mercado; un sector que –según una estimación realizada por AEDEMO entre sus socios y los pertenecientes a ANEIMO– en el año 2002 facturó 306 millones de Euros<sup>1</sup>, y que está formado por alrededor de 3.600 trabajadores y 211 empresas (Alós, 2003).

Adoptando este punto de partida, la frase de Chambers, transcrita entre la tercera y cuarta línea del párrafo dos<sup>2</sup>, constituye un principio fundamental para comprender el futuro del software para análisis de encuestas. Desde mi punto de vista este futuro estará unido a las facilidades de utilización del usuario, y esto está referido tanto a la solicitud de análisis (interacción con la máquina) como a la presentación de resultados. Me explico: aunque aún hoy el “lenguaje de comandos” constituye una parte esencial de muchos programas de análisis de encuestas<sup>3</sup> –como ha quedado reflejado en el artículo de Pedro Valero Mora– en el futuro desaparecerán estos complicados lenguajes “de interacción con la máquina” con el objetivo de facilitar al máximo el manejo de estos programas. Este ha sido el gran éxito de diverso software desarrollado desde el ámbito de la investigación de marketing, y distribuidos con libros de esta temática (Dyane y Dyane 2 de Miguel Santestases, Diremark de Ildefonso Grande, etc.).

A mi juicio el futuro del software de análisis de encuestas está intrínsecamente unido al desarrollo de resultados que potencien el componente gráfico, y que estos gráficos

---

<sup>1</sup> En otro trabajo hemos señalado que estas cifras suponen una *estimación a la baja* del sector, puesto que del análisis de la ficha técnica de la investigación se desprende que en este estudio no se consideran las empresas pequeñas que no forman parte de ANEIMO, ni los departamentos de investigación comercial de las grandes empresas. Tampoco contempla la actividad desarrollada por otros ámbitos como la universidad, los institutos de investigación social de los Gobiernos Autónomos, el Centro de Investigaciones Sociológicas, el Instituto Nacional de Estadística, los institutos de estadística de las Comunidades Autónomas y de la Comunidad Europea, etc. (Díaz de Rada, 2001:13).

<sup>2</sup> “The busy executive, the non-specialist from another field, the non-technical person just needing a few simple answers”.

<sup>3</sup> Aunque algunos programas han realizado notables esfuerzos para elaborar interfaces de usuario que eliminan la utilización de estos comandos, normalmente es necesario recurrir al lenguaje de comandos en las partes más avanzadas del programa (esto ocurre, por ejemplo, en algunas utilidades multivariantes del programa SPSS).

sean fácilmente trasladables a otros programas. Pensemos en los usuarios de estos programas, considerando el *perfil formativo* que actualmente se imparte en las universidades. En esta línea podríamos considerar que los actuales estudiantes de “Investigación y Técnicas de Mercado” serán los futuros usuarios de estos programas<sup>4</sup>, de modo que los planes de estudios podría aportar una importante información del futuro de este software<sup>5</sup>. Un rápido análisis de los planes de estudio<sup>6</sup> desvela una escasa formación estadística, y no muy abundante en técnicas de investigación, lo que implica que poseen una formación genérica (no especializada) en el análisis de datos<sup>7</sup>. Esta situación, a la que tendrá que enfrentarse el sector en el futuro, requiere unos desarrollos de software muy sencillos y, más importante aún, una plantilla de *desarrolladores*<sup>8</sup> de software dispuestos a prestar apoyo a los usuarios que lo necesiten. Es decir, junto con el software se “contrata” un asesoramiento técnico-estadístico.

El componente gráfico es esencial no sólo para el usuario de estos programas, sino fundamentalmente pensando en el “cliente final”, el demandante de la investigación. Es un hecho que uno de los factores que han contribuido a la escasa utilización del análisis multivariante (fuera del ámbito académico) es la difícil explicación de sus resultados al demandante de la investigación<sup>9</sup>, al empresario que en muchas ocasiones está pensando –tan sólo– en un porcentaje. Complejos tratamientos estadísticos para una persona que, en definitiva, quiere saber si su publicidad es más recordada que la de la competencia, si su producto gusta más que la competencia<sup>10</sup>... (Qué duda cabe que la facilidad de lectura por parte del cliente final ha constituido el gran éxito de los “mapas de posicionamiento”).

---

<sup>4</sup> No me olvido que el sector de la investigación de opinión y mercados esta formado, ahora mismo, por un gran número de psicólogos, sociólogos y economistas que han adaptado sus habilidades a esta profesión. Sin desdeñar la adecuación de estos perfiles para el desarrollo de la investigación, considero que –ahora mismo– es más interesante conocer el perfil formativo de unos estudios destinados, en principio y de manera específica, a la investigación de mercados.

<sup>5</sup> Qué duda cabe que conocer a los futuros usuarios es determinante para poder *anticipar* los futuros desarrollos de estos productos.

<sup>6</sup> Existe una web, de la Universidad de León, que permite un rápido acceso a los planes de estudios de todas las universidades que imparten este título:

<http://www3.unileon.es/dp/dde/marketing/indexITM.htm>

Una síntesis de estos planes de estudios presentan también Cervantes, Bigné y Ballesteros (2004).

<sup>7</sup> Debemos tener presente que se trata de una titulación de segundo ciclo, y que considera que los alumnos poseen unos conocimientos *básicos* de estadística, microeconomía y técnicas de investigación. Para los que no cuenten con tales conocimientos, las distintas universidades ofrecen complementos de formación.

<sup>8</sup> Por seguir con la terminología empleada por Pedro Valero Mora.

<sup>9</sup> Aspecto que se trató más específicamente en Díaz de Rada (2003).

<sup>10</sup> No nos olvidemos que esta persona es la que encarga investigaciones, la que está sosteniendo la industria de la investigación de opinión y mercados en nuestro país.

Otro aspecto fundamental en el desarrollo de este software es la facilidad para interactuar con el resto de programas informáticos, y me estoy refiriendo tanto a programas específicos (relacionados con el análisis de encuestas) como más genéricos y difundidos. Respecto a los primeros, el futuro de software “especializado” en unos pocos análisis (por ejemplo software especializado en análisis factoriales) ha estado unido a la facilidad de *importación* de datos de los programas más difundidos del mercado<sup>11</sup>; así como la exportación de sus resultados y gráficos a otros programas menos específicos y más difundidos (como editores de texto, por ejemplo). Ya nadie piensa en copiar<sup>12</sup> una tabla de datos, o en transcribir los datos de una tabla a un programa gráfico que mejore la presentación de resultados. El software utilizado en el análisis tiene que tener integradas ambas funciones, y debe mostrar una interacción total con editores y con programas gráficos avanzados.

Por último, los nuevos desarrollos deben mejorar la comprobación de los supuestos que tienen que cumplir los datos para aplicar determinadas técnicas estadísticas (fundamentalmente las multivariantes), ofreciendo soluciones al usuario cuando no se llegan a cumplir estas condiciones. Ahora mismo es el propio usuario el que, en el caso del Análisis de Componentes (por ejemplo), tiene que saber que debe solicitar el determinante de la matriz de correlaciones (para conocer el grado de intercorrelaciones), el test de esfericidad de Barlett (para comprobar si la matriz de correlaciones obtenidas es un matriz identidad), etc. El usuario debe conocer los supuestos de cada una, solicitarlos, y pensar qué alternativa tomar cuando éstas no se cumplen. En el futuro, creemos, deberá ser el propio software el que –de forma interactiva– pregunte al usuario sobre los supuestos que debe cumplir cada técnica estadística, ofreciendo soluciones cuando éstos no se cumplen.

Para terminar, y teniendo en cuenta que el punto de partida de mi reflexión es el sector de la investigación comercial y de mercados, la verdad es que veo un escaso desarrollo de los programas gratuitos entre los usuarios de este sector. Más que un software, estas empresas están demandando también un asesoramiento estadístico metodológico; fundamentalmente en las técnicas más complejas (y menos utilizadas).

---

<sup>11</sup> Tanto programas de “recogida” de información (CATI, CAPI, etc.), como de tabulación y análisis descriptivo.

<sup>12</sup> En el sentido “clásico” de la palabra, en la transcripción de un texto o una tabla de datos. No me refiero, lógicamente, a las opciones de “copiar” y “pegar” presentes en la mayor parte de los programas del mercado.

En Metodología de Encuestas hemos tenido ya el placer de contar con reflexiones en torno a la utilización del análisis de datos y su hoy indisoluble relación con los programas de ordenador (Solanas y otros, 2001; Brandy y otros, 2002). En aquel momento, ocuparon estas páginas la descripción de qué debería ocurrir y qué está ocurriendo con la práctica del análisis de los datos mediante el recurso informático. Poco más o menos, lo que nos preocupa ahora es ¿y qué ocurrirá?

La historia del conocimiento está llena de previsiones incumplidas. Cuando se realizan, se basan en la configuración de conocimientos del momento y el tiempo cambia de tal modo esa configuración que suele jugar una mala pasada a las previsiones. Ahora nos imaginamos un futuro que rebosa teléfonos móviles, automóviles privados de una extraordinaria ergonomía, televisores de realidad virtual inundando los domicilios, ropas exclusivamente sintéticas, votaciones únicamente por ordenador... En pocas palabras: más de lo mismo pero elevado a la máxima potencia. No debería extrañarnos escuchar a un ciudadano romano del siglo II d.C. describiendo el futuro en términos de grandes coliseos, carros tirados por cientos de caballos, ejércitos provistos de armaduras extraordinariamente resistentes, etc. Los romanos del momento no tenían la posibilidad de imaginar la electricidad, el avión o el teléfono móvil. Hoy, los economistas, o los econométras, sufren el proceso (o ayudan a que lo sufran otros) cuando intentan forzar un poco sus estimaciones, aunque no sean lineales, aunque recurran a los últimos avances en análisis de series temporales, aunque las estimaciones intra-muestrales sean impecables y alentadoras. Un poco más allá del contexto temporal de sus muestras, la economía observable sigue resistiéndose.

En este mismo discurso sobre evolución tecnológica, ¿hacia dónde se encamina el desarrollo de programas de ordenador útiles para el diseño y realización de encuestas? Como señala Pedro Valero, ya Tukey advirtió que análisis de datos y ordenador estaban condenados a entenderse, puesto que vivió el matrimonio entre ambos con el nacimiento de las primeras computadoras modernas merecedoras de tal nombre. Y tal fue la convicción del profeta que él mismo protagonizó la invención de índices que no se pueden entender sin el auxilio de la máquina, como los no suficientemente recurridos M-estimadores, arrinconados habitualmente en el cajón del análisis exploratorio de datos (ese "otro" análisis).

¿Qué nos depara el futuro? En mi a largo plazo errónea opinión, hay varias tendencias que coexisten en una especie de confrontación y que no espero se resuelvan con la dominación de ninguna de ellas sobre su pareja, sino que encontrarán un público más o menos fiel, un segmento de población determinado:

1. Software comercial o de libre distribución. Ambos se sustentan en motivaciones suficientemente fuertes como para que no desvanezcan en breve. El software comercial, como todo lo que comparte el mismo apellido, bebe de la supervivencia y el crecimiento a través del lucro. Es necesario responder a las expectativas del mercado (sean espontáneas o creadas) para mantenerse, para abarcar, para crecer y ello obliga a las empresas de construcción de software comercial de análisis de datos a estar pendientes de cuáles son las necesidades últimas en

el sector y cómo llevarse el favor de los potenciales clientes. Parece que se mantiene la tendencia lineal de nuestra época: la comodidad, la eficiencia, obtener el análisis más adecuado (según diversos criterios) con el mínimo esfuerzo. Y eso es lo que seguirán ofertando las empresas de desarrollo de software comercial: la posibilidad de realizar los análisis del momento con agilidad y vistosidad, con resultados fáciles de interpretar y que requieran el mínimo esfuerzo para ser presentados a terceros. En cuanto al software de libre distribución, vivimos un momento de euforia, con antecedentes en crecimiento como la norma GNU y el universo Linux. Son iniciativas que reciben en estos momentos serios golpes a nivel de legislación. Pero cabe esperar que estos avatares no frenen sino que incluso estimulen la actividad de los desarrolladores de software libre, entendiendo que muchos de ellos alimentan su motivación en la rebeldía, que crece con la oposición del poder.

2. Trabajo por cuenta propia o ajena. El lenguaje R del que nos habla Pedro Valero, como los que acompañan a buena parte de los programas de análisis de datos disponibles, constituyen un potencial que no puede desaparecer del software de análisis, sea comercial o de libre distribución. Los procedimientos de análisis prefijados son fundamentales para agilizar la tarea del proceso de datos, para abreviar el trabajo con la máquina cuando se pretenden cosas habituales y previstas por los desarrolladores. Pero estas tareas nacen ya limitadas. Quienes realizamos análisis con cierta frecuencia vivimos habitualmente estas limitaciones: presupuestos que no se cumplen, resultados que obvian información relevante, combinaciones inexistentes, procesos innecesariamente repetitivos, etc. Para todo ello recurrimos a los lenguajes específicos de los programas al uso o bien creamos utilidades *ad hoc* con el lenguaje de programación con el que estemos más familiarizados. A modo de ejemplo, aunque llevo años utilizando varios útiles específicos para el análisis de los datos, me ha resultado imposible prescindir de esas pequeñas "chapuzas" en BASIC o de recursos algo más elaborados en GAUSS, C o Delphi.
3. Dimensión analítica o gráfica. La principal justificación para las representaciones gráficas clásicas, como los diagramas de barras, ha sido el ahorro de esfuerzo cognitivo que requiere su utilización frente a recursos como las tablas de frecuencia. De hecho, la agilidad con que se interpreta una representación gráfica ha sido estímulo para generar diversas estrategias de manipulación que suelen pasar desapercibidas, como el famoso corte en el origen o el juego inadecuado de las tres dimensiones. Este ahorro cognitivo sigue vigente, al menos en el mismo grado en que nuestro cerebro no ha evolucionado sensiblemente ni esperamos ningún cambio claro al respecto en los próximos miles de años. A esta circunstancia se añade que las nuevas generaciones se gestan en entornos donde la información visual es cada vez más protagonista. Luego, el futuro inmediato parece señalar cierto avance relativo de la información en formato visual frente a la numérica, mayor protagonismo de la gráfica frente a la tabla y a la fórmula y mayor desarrollo de la versión visual de los índices numéricos. En este contexto, es fundamental que las utilidades de análisis de datos permitan trabajar de forma gráfica, navegando incluso por los resultados obtenidos y generando

otros a partir de ellos. Gracias al desarrollo 3D y al crecimiento de los productos de realidad virtual, no debería extrañarnos que dentro de no mucho tiempo contemos con utilidades que permitan abrir nuevas dimensiones en tiempo real, navegando por las representaciones gráficas y superando, con ello, la insatisfactoria limitación de las representaciones 2D (e incluso 3D) multivariantes.

4. Usuario experto en análisis o experto en el motivo del análisis. Es siempre difícil establecer un momento cero en los procesos de enseñanza/aprendizaje. Así, por ejemplo, para enseñar a un grupo de estudiantes universitarios a realizar una estimación estadística ¿por dónde comenzar? ¿Tal vez en una fuerte base matemática que permita manejar el cálculo de áreas bajo una curva sin el freno de las tablas, recurriendo a las integrales definidas, enseñando a los estudiantes a deducir cuantas funciones generadoras de momentos sean aconsejables y estimulando el establecimiento de una base general para la estimación mediante transformaciones logarítmicas máximo-verosímiles que acerquen a la Ji-cuadrado? ¿Tal vez baste con la adquisición de habilidades o competencias para saber qué botón pulsar y cómo interpretar la salida? ¿Es necesario enseñar a un conductor novel el funcionamiento del motor del vehículo o basta con el uso adecuado de los pedales y la interpretación de las señales de tráfico? En este punto creo que el futuro está claro: la probabilidad de encontrar conductores que saben conducir pero que no entienden de motores es creciente. La universidad debe adaptarse a las exigencias de este momento y formar a futuros titulados que sepan cómo hacer análisis y cómo interpretar las salidas del ordenador, aunque carezcan de creatividad estadística, si bien es imprescindible igualmente su conciencia de actuación limitada. No podemos ser expertos en todo y para eso se inventó el trabajo en equipo. El software comercial seguirá mostrando lo fácil que llega a ser un análisis de datos, abundando en la tendencia de permitir a la máquina tomar más decisiones. El 0,05 de Fisher cede ante los 0,01 ó 0,001 de SPSS en las matrices de correlación. Y no se cuestiona. Lo que sí está a nuestro alcance es aportar algo a un futuro más sabio, formando a los futuros usuarios de software en la conciencia de las limitaciones, tanto del propio conocimiento como de las posibilidades de la máquina como sustituta.
5. Pequeños análisis o masas de datos. La evolución reciente del análisis estadístico intenta dar respuesta a la necesidad real de utilizar grandes masas de datos que provienen de varias fuentes, de varias encuestas, de varios momentos en el tiempo e, incluso, de varias motivaciones y condiciones de generación (véanse, por ejemplo, Calvo y otros, 2000; Muñoz y otros, 2001; Sánchez Rivero, 2001; Fernández, 2003). Se está generando una especie de análisis de análisis de datos, donde se atraviesan varias dimensiones de complejidad. Esta evolución, si bien obedece a una necesidad tangible y a la internacionalización de estándares de análisis (en los organismos públicos, por ejemplo), no sustituye las situaciones clásicas donde el investigador no aspiraba a más que al cruce de dos variables o a la influencia de un regresor sobre un criterio. Está claro que cuanto más multivariante sea un análisis más asombro y reconocimiento generará en la expectante sala. Quienes trabajamos en departamentos universitarios ligados a la metodología de investigación, estamos ya acostumbrados a recibir la petición

de otros colegas que no va tanto orientada a satisfacer una necesidad de conocimiento como a deslumbrar a receptores académicos o empresariales con análisis de una vistosidad o una complejidad (o ambas cosas al mismo tiempo) de gran envergadura. Este uso es indeseable, aunque por fortuna el resultado de tales investigaciones no se aplica nunca. La existencia de este fenómeno es independiente del buen uso de las herramientas sofisticadas, como también de las abundantes situaciones en las que procede, se necesita y es lo más adecuado realizar análisis de una gran simplicidad conceptual. Cabe esperar que el desarrollo en este campo no sea sustitutivo sino acumulativo: los programas, como los usos, brindarán mayores posibilidades sin que ello implique abandonar análisis (tanto del cajón "descriptivo" como del de la "inferencia") ya consolidados con anterioridad. Aquí constituye todavía una asignatura pendiente la adecuación de los procesos de inferencia a los modelos de muestreo (por ejemplo, Trujillo, 2000), parcialmente superable a partir de los procedimientos de remuestreo; o la inclusión de potentes herramientas de depuración de las matrices de datos que incluyan, por ejemplo, estrategias de imputación que hoy en día deben saciarse por otras vías (véase, por ejemplo, Morales, 2000). Los desarrolladores de software estadístico no tienen hoy problemas de imaginación, hay tanto por hacer que el futuro mediato está garantizado.

En términos generales, por tanto, mi impresión es que el futuro nos mostrará un mayor número de posibilidades de análisis, en la capacidad de la máquina para trabajar información visual y manejar grandes masas de datos de origen diverso, en la diversidad de caminos para efectuar los análisis (comercial o libre distribución, creación a la medida o recursos estándar), mientras que este panorama exigirá cada vez más la formación de profesionales "operativos" de quienes tal vez no esperemos revolución alguna en el campo del análisis de los datos, si bien nos permitan gozar de la tranquilidad de que hicieron bien su trabajo, sin conocer las entrañas de las estrategias de análisis pero conociendo satisfactoriamente sus condiciones de uso.

El futuro siempre es sorprendente, de modo que toda predicción es un atrevimiento excesivo. Bastante hacemos intentando comprender lo que tenemos delante de los ojos. Por eso me limitaré a hacer algunas observaciones sobre novedades que ya estamos viendo funcionar, en mi sector específico, que es el de la creación de software de gestión de encuestas y de tabulación y análisis de datos.

Últimamente hemos visto un gran aumento, absoluto y relativo, de las encuestas telefónicas. La tendencia viene impuesta por razones de economía y también por los cambios en los hábitos de vida de la población, más dispuesta ahora a contestar a un cuestionario por teléfono que a abrir la puerta a un entrevistador. Además, ahora ya no hay apenas hogares sin línea telefónica. Y la proliferación de los móviles, incluso en el papel de sustitutos absolutos de los teléfonos fijos, no va a suponer ningún problema, sino todo

lo contrario, con el uso de las tácticas adecuadas. Pero lo que ha hecho posible ese crecimiento de los sondeos telefónicos es un software diseñado específicamente para este cometido, con programas que gestionan integralmente el proceso de interfaces manejables por entrevistadores sin entrenamiento previo. Pues bien, con el aumento de los usuarios de Internet vamos a ver, eso espero, algo parecido. Se trata de un campo en el que casi todo está por ver, pero seguro que la investigación de mercado y opinión a través de Internet va a conocer una expansión, tanto con la adaptación a este medio de los cuestionarios de encuesta clásicos, como usando métodos indirectos, por la capacidad de explotación del tráfico.

En otro orden de cosas, la aparición de software de gestión de sondeos específico para ordenadores “de mano” (*PDA*s, *Handhelds*, *Palms*...), comunicados con el técnico del estudio a través de Internet, representa la posibilidad de informatizar integralmente incluso el trabajo de campo. Esta novedad también va a imponerse, desterrando progresivamente los cuestionarios impresos y, lo que es más importante, el engorro del proceso de grabación de datos. Actualmente, este instrumental permite ya la reproducción de formatos multimedia, lo que facilita enormemente el estudio del rendimiento de acciones publicitarias –el clásico test de recuerdo, por ejemplo.

Finalmente, recordaré los avances realizados en la codificación de las respuestas a preguntas abiertas, un campo hasta ahora inaccesible a la informatización.

Todos estos son instrumentos que facilitan, amplían y al mismo tiempo abaratan los métodos clásicos –por medio de cuestionarios— de captación y proceso de datos, e incluso que abren nuevos campos aún por explorar, en gran parte. Pero, curándome en salud, terminaré recordando que el futuro es una liebre que siempre salta por donde no estábamos mirando.

## Referencias

- Alós, J. (2003) Industria de los estudios de mercado en España 2002, *Investigación y Marketing*, 80, 76-78.
- Ares, V.M. (1999) La prueba de significación de la hipótesis cero en las investigaciones por encuesta” *Metodología de encuestas*, 1(1) 47-68
- Arnau, J. (1995). Metodología de la investigación psicológica. En M.T. Anguera, J. Arnau, M. Ato, R. Martínez, J. Pascual y G. Vallejo (eds.), *Métodos de Investigación en Psicología* (pp. 23-44). Madrid: Síntesis-Psicología.
- Bakan, D. (1966) The test of significance in psychological research. *Psychological Bulletin*, 66, 1-29.
- Borges, A. (1999). Comentarios y respuesta a “la prueba de significación de la Hipótesis cero en las investigaciones por encuesta”. *Metodología de encuestas*, 1(1) 69-84
- Brandy, D. y otros (2002) Comentarios y respuesta a 'Análisis estadístico con aplicaciones informáticas: ¿Ayuda o sustituye al razonamiento?'. *Metodología de Encuestas*, 4 (1) 1-28.
- Calvo, P.; Ayestarán, M; Prado, C. y Pérez, Y. (2000) Creación de objetos simbólicos a partir de encuestas almacenadas en bases de datos relacionales. *Metodología de Encuestas*, 2 (2) 239-260.

- Cervantes Blanco, M.; Bigné Alcañiz, E. y Ballesteros García, C. (2004) El título universitario de Licenciado en Investigación y Técnicas de Mercado (ITM), *Investigación y Marketing*, 82, 68-74.
- Díaz de Rada, V. (2001) *Tipos de Encuestas y Diseños de Investigación*. Pamplona: Universidad Pública de Navarra.
- Díaz de Rada, V. (2003) Análisis Multivariable. Teoría y práctica en la investigación Social de M.A. Cea D'Ancona (Síntesis, 2002), *Empiria*, 6, 207.
- Fernández, K. (2003) Análisis textual: generación y aplicaciones. *Metodología de Encuestas*, 5 (1) 55-66.
- Morales, L. (2000) El efecto de la no respuesta parcial en el análisis de datos de encuesta: una comparación entre la eliminación de observaciones y la imputación múltiple. *Metodología de Encuestas*, 2 (2) 217-238.
- Muñoz, A.; Buiza, B. y Pino, R. (2001) Edición de encuestas mediante redes de neuronas artificiales. *Metodología de Encuestas*, 3 (2) 151-162.
- Sánchez Rivero, M. (2001) Modelos de estructura latente: una potente herramienta al servicio del análisis de encuestas. *Metodología de Encuestas*, 3 (1) 75-95.
- Solanas, A.; Batista, J.M.; Llover, Ll. de; López, A.; Núñez, M.I.; Ocaña, J.; Salafranca, Ll.; San Martín, R.; Simó, J.; Valle, B. y Visauta, B. (2001) ¿Existe algo más fácil que un análisis estadístico? *Metodología de Encuestas*, 3 (2) 227-250.
- Trujillo, M. (2000) Optimización del error muestral en encuestas a población general: criterios de estratificación y cálculo del error con "Wesvar Complex Samples". *Metodología de Encuestas*, 2 (2) 199-210.
- Tukey, J. W. (1965). The technical tools of statistics. *The American Statistician*, 19:23-28.