

Imputación de información faltante o confusa: un diagnóstico de su evolución

Badler, Clara E.; Alsina, Sara M.

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística (IITAE).
Consejo de Investigaciones (CIUNR), Universidad Nacional de Rosario
Bv. Oroño 1261-2000 Rosario, Tel/Fax: 0341-4802793; E-mail: cbadler@unr.edu.ar

RESUMEN: La técnica de imputación ha sido una de las estrategias que se ha instalado con preferencia, para el tratamiento de bases de datos con información faltante.

Su aplicación, mediante la asignación de valores a aquellos faltantes o contradictorios, puede realizarse a través de diferentes procedimientos. Son numerosas las propuestas que surgen tanto de instituciones productoras de información como de autores individuales, tendientes a la formalización teórica, a la optimización y operatividad de su implementación.

Ello se manifiesta a través de su presencia como tema de artículos de revistas, cursos y reuniones científicas y en proyectos de investigación.

Se realiza un análisis de la importancia adquirida por esta técnica y de los aspectos relativos profundizados en la actualidad por los autores, a través del examen de fuentes de difusión de información científica y de la utilización de Internet para el acceso rápido a noticias de actualidad y la realización de diagnósticos.

Palabras claves: Imputación – Información faltante – Fuentes de información científica.

SUMMARY: Imputation of missing or messy information: Diagnosis of evolution. Badler, Clara E.; Alsina, Sara M.

Imputation techniques are among the strategies statisticians prefer for treating data bases with missing information.

Imputation means assigning values to missing or contradictory information and can be applied through different procedures. Both institutional and individual proposals have arisen, which aim at achieving theoretical formalization, and operative implementation. They have been widely acknowledged in magazines, seminars, scientific meetings and research projects.

This paper presents the results of a review and analysis of present scientific sources, including Internet, as a mean of spreading the increasing importance and usefulness of this technique

Key words: Imputation - missing information - scientific information sources.

Introducción

La técnica de imputación ha sido una de las estrategias que se ha instalado con preferencia, para el tratamiento de bases de datos con falta de información parcial, es decir cuando no han podido ser observados algunos ítems, preguntas o variables de una base de datos, a pesar de haber estado previsto por el diseño, o los valores obtenidos resultan contradictorios o defectuosos.

Esta técnica fue considerada durante mucho tiempo sólo como un procedimiento meramente práctico hasta que, con el respaldo de la teoría desarro-

llada y de las aplicaciones con computadoras, se fue transformando en una parte del proceso estadístico y en uno de los temas más frecuentes de investigación metodológica.

Son numerosas las propuestas para su implementación, que surgen tanto de instituciones productoras de información como de autores individuales, tendientes a la formalización teórica, a la optimización y recaudos de su aplicación y a su operativización.

Este trabajo, a partir del examen de fuentes de información científicas, intenta determinar la vigencia y las líneas que se acentúan en los últimos años

en el estudio y aplicación de esta técnica. La consulta excede las fuentes escritas, extendiéndose a Internet, que permite acceder en forma rápida no sólo al conocimiento de contenidos sino también comprobar que este tema está presente tanto en las principales publicaciones de actualidad como en sesiones de todas las principales reuniones científicas de estadística.

Tratamiento de la información faltante

Diferentes análisis se han propuesto a partir de la falta parcial de información. Hasta la década del ochenta la tendencia fue la de "librarse" de los datos faltantes, luego la de incorporarlos en la estimación.

A partir del avance de la teoría estadística y de las técnicas de computación se han desarrollado procedimientos flexibles con sólida base estadística que responden a la forma de concebir los datos faltantes como una fuente de variabilidad a incorporar.

La toma de conciencia sobre la necesidad de tratamiento para el problema de datos faltantes se va manifestando mediante propuestas tanto teóricas como prácticas, con características diferentes asociables a tres períodos: a partir de los setenta, trabajar solamente con las unidades con información completa (casos completos) o realizar imputaciones simples mediante numerosos procedimientos (para luego analizar con métodos para datos completos), en los ochenta se agregan procedimientos basados en verosimilitud para realizar estimaciones a partir de los datos incompletos (no requieren conjuntos rectangulares e incorporan una distribución de probabilidad de los datos faltantes) e imputación múltiple (obteniendo varios conjuntos imputados para incorporar la incertidumbre por la misma imputación), y a partir de los noventa se suman métodos bayesianos y el método Monte Carlo con cadenas de Markov (MCMC) considerando una evidencia indirecta sobre los probables valores no observados, a partir de la evidencia provista por los observados (1).

La inicial necesidad de dar una solución práctica al problema se ha ido transformando en la utilización responsable de los distintos métodos que requiere un conocimiento de los supuestos explícitos o subyacentes y de las posibles consecuencias en la precisión del análisis, al apartarse de ellos. Estos supuestos se relacionan con la distribución de los

datos originales, el comportamiento de los datos faltantes y las características del mecanismo de pérdida. La facilidad de implementación de las propuestas también es determinante en la final elección del método, aspecto relacionado al desarrollo y accesibilidad de rutinas de computación.

La opción entre completar la base o trabajar a partir de los datos incompletos se relaciona también con la etapa en la que se contacte con la base y con el objetivo con el que la misma se utilizará. Para algunos propósitos es necesario disponer de la base completa por razones de cesión de la misma o analíticas, por ejemplo por restricciones de la técnica de análisis que se aplicará; en otras situaciones interesa justamente disponer de la base incompleta y conocer los mecanismos o causas relacionadas con este problema para tenerlos en cuenta en forma preventiva o para saber cómo incorporar la pérdida en las estimaciones.

Imputación

Se define básicamente como técnica de imputación a aquella consistente en completar los valores faltantes en función de información interna o externa a la misma base, tratando de reproducir los datos verdaderos y obtener de esta manera conjuntos completos de datos previamente al análisis definitivo de los mismos.

Son numerosos los procedimientos desarrollados para realizar las sustituciones: promedios, promedios condicionales, regresión, hot-deck, regresión logística, vecino más cercano, modelos de redes neuronales, etc., combinados y propuestos por los distintos autores para responder a las características de los datos y de las pérdidas, a los diseños y a los propósitos de cada análisis (2;3).

La mayoría de los métodos clásicos son formulados bajo el supuesto de que la pérdida se ha producido al azar (denominado MAR por la expresión inglesa "Missing at random"), aunque también bajo el más restrictivo completamente al azar (caso denominado MCAR por "Missing completely at random"). Cuando la pérdida depende de los valores faltantes en general se requieren para el tratamiento modelos que consideren el comportamiento de los datos faltantes (4).

A partir de la incorporación de las citadas restricciones se toma conciencia de que imputar no es sólo completar los datos faltantes o erróneos, aún tratando de acercarse lo mejor posible a la distribución conjunta desconocida de los datos observados y no observados. La imputación es un ejercicio de modelización aunque el mismo no sea explícitamente formulado, ya que para implementarla se construyen clases, se seleccionan variables apropiadas y se identifican relaciones entre las variables propuestas (5). En cada modelo de imputación la variable a ser sustituida depende de variables auxiliares, de parámetros estimados en función de los valores disponibles y de la estimación de un error.

Al elegir el modelo para imputar los datos, se debería tener en consideración el análisis que se realizará con ellos. El modelo utilizado para la imputación debe tender a preservar las características y relación entre las variables, y en particular de aquellas que luego serán objeto de dicho análisis.

Durante mucho tiempo se utilizó la imputación sin evaluar o incorporar su impacto, luego se fueron incorporando aportes para estimar correctamente la variancia de las estimaciones a partir de datos imputados, no sólo debida a la variabilidad muestral, sino también a la imputación: el más difundido ha sido imputación múltiple debida a Rubin (6) y luego aceptada y difundida por distintos autores (7;8); también otros enfoques como ajustes en las imputaciones a partir de la replicación de la estimación de variancia y del proceso de imputación y a través de la técnica jackknife (9).

Es así como, ante la implementación de esta técnica, surgen preguntas claves: cuál es el método óptimo para obtener los reemplazantes, de qué manera influye en los resultados de los análisis, hasta qué porcentaje de valores reemplazar. El hecho de que las respuestas no sean únicas sino condicionadas a distintas situaciones, ha provocado una intensa actividad de investigación en esta área, para optimizar los resultados de su implementación en distintos escenarios.

Proporción de valores a imputar

La determinación de la proporción de valores que pueden imputarse está ligada a cuánta información faltante es tolerable, en qué nivel máximo de información incompleta se pueden basar las

inferencias aunque sea con supuestos favorables sobre el mecanismo que genera la falta de información para las variables. No hay una respuesta única para este problema, cuya solución depende del concepto de precisión ligado a las diferentes circunstancias, como periodicidad de las mismas, mantenimiento de una confiabilidad en las instituciones responsables, incorporación de variables sensibles, áreas disciplinares que requieren alta precisión (10). Todos los autores coinciden en que, en el marco global del problema de la información incompleta, el primer objetivo a considerar es tender a la disminución de la proporción de unidades no observadas, comenzando en la etapa de diseño y educar al usuario para poder utilizar correctamente información con limitaciones, pero paralelamente se debe extremar el esfuerzo para optimizar la aplicación de la metodología de imputación y así colaborar eficientemente en la obtención de estimaciones más precisas (11).

Líneas de investigación

Son numerosos los trabajos relacionados con imputación publicados en los distintos medios científicos.

A pesar de la diversidad de los mismos, se observa una tendencia a superar las propuestas anteriores, cuestionadas en general por ser demasiado restrictivas en cuanto a los supuestos sobre la distribución de las variables requeridos para su aplicación, por el corto alcance en la generalización de las conclusiones de cada trabajo o por la poca operatividad (12).

La búsqueda del mejor método de imputación parece haberse instalado en muchos ámbitos, pero la respuesta nunca puede ser única por las razones ya citadas. De todas formas hay conclusiones muy generales aceptadas sobre los métodos clásicos: casos completos, casos disponibles e imputación por el promedio son dejados de lado ante propuestas sobre imputación múltiple e imputación por regresión y hot deck.

A partir de la revisión de numerosos trabajos sobre esta técnica publicados en los últimos años, se citan diferentes líneas de investigación adoptadas con respectivas referencias:

- Evaluación de distintos modelos para imputación través de la comparación de las estimaciones

de los parámetros correspondientes al análisis propuesto: Laaksonen (3) analiza modelos de regresión basados en la utilización de vecino más cercano versus hot deck aleatorio o un método basado en modelos de dos etapas (regresión logística e imputación por regresión).

- Consideración del modelo de imputación en relación con el modelo a aplicar en el análisis estadístico final: Shao et al. (13) afirman que el método de imputación por regresión produce estimaciones no sesgadas para totales marginales pero no para coeficientes de correlación, proponiendo entonces un método por regresión conjunta.

- Elección de modelos de imputación de acuerdo a los esquemas de pérdida y a la distribución de las variables afectadas: Yuang (14) prueba que para datos normales multivariados con esquemas monótonos imputación por regresión es la más recomendable, con esquemas arbitrarios lo es el método MCMC.

- Evaluación de los supuestos del modelo de imputación: Schafer (15) analiza la robustez de imputación con aproximación normal para datos categóricos; Aers et al. (16) estudian un método de imputación semiparamétrico y otro no paramétrico por resampling.

- Propuestas y evaluación de métodos no condicionados a las distribuciones ni a los esquemas: Zador (17) propone árboles de regresión aditiva múltiple (MART) versus regresión y hot deck.

- Propuesta y evaluación de modelos de imputación a distribución libre: Belin et al. (18) lo consideran a través de regresión logística jerárquica.

- Propuestas para incorporar la variabilidad debida a las imputaciones en la estimación de la variancia total: Wung (19) propone estimadores de la variancia consistentes, aún cuando los modelos de imputación y de análisis sean incompatibles o no especificados y para diseños muestrales complejos; Chen et al. (20) evalúan con jackknife el impacto de distintos métodos de imputación sobre la estimación de la variancia; (21).

- Diagnóstico de la calidad de las imputaciones: Laaksonen (3) evalúa cuántas veces se utilizan los mismos valores.

- Propuestas de modelos de imputación para casos con mecanismo de pérdida no ignorable (Tang, 22).

- Inserción de procedimientos de imputación en distintos análisis estadísticos específicos: diseños longitudinales, datos censurados, análisis de supervivencia, estudios de caso-control: Wang (23) aborda un tratamiento de datos longitudinales incompletos de alta dimensión con imputación múltiple utilizando un modelo de análisis de factor longitudinal.

- Aplicación de algoritmos iterativos y de técnicas de simulación en la obtención de imputaciones: Yuang (14) emplea técnicas de Monte Carlo con cadenas de Markov, van Buuren et al. utilizan muestreo de Gibbs o algoritmo Metropolis-Hastings (24).

En este ámbito de evaluación se lleva a cabo el Proyecto "Evaluación de nuevos métodos para edición e imputación" de la Comisión Europea (EUREDIT) (25), con la participación de siete países con los objetivos de desarrollar nuevos métodos (basados en redes neuronales con soporte computacional y en métodos estadísticos robustos para outliers), evaluarlos comparativamente con métodos estándar y difundir los resultados. El foco de aplicación de este análisis son los datos recolectados por las agencias estadísticas nacionales y la evaluación en el área de imputación se lleva a cabo a partir del conocimiento de los verdaderos valores, con criterios de preservación de la verdadera información, conservación de las distribuciones, obtención de estimadores no sesgados y eficientes, obtención de resultados que superen las pruebas de edición.

Imputación para los productores de información

Los productores de información, desde aquella correspondiente a bases pequeñas resultantes de investigaciones en distintas áreas hasta las oficiales de gran dimensión, son los que están en mejores condiciones para poder elegir el procedimiento más adecuado para imputar, ya que conocen todo el proceso de la información y las posibles causas de la información incompleta.

Ya que el usuario puede o no coincidir con el productor es necesario que, en el caso en que este último haya aplicado imputación, dé a conocer la metodología utilizada para que pueda ser tenida en cuenta en el análisis posterior.

Los entes nacionales de estadística no sólo aplican sino que ensayan y proponen nuevos proce-

dimientos para imputación, con un enfoque de inmediata aplicación en bases de datos de gran tamaño.

El Instituto Nacional de Estadística y Censos (INDEC) presenta en su página de Internet (26) documentos relativos a aspectos metodológicos de distintos relevamientos a su cargo, en los que se explicita el método de imputación aplicado y comparación con métodos aplicados en institutos de estadístico de otros países: para la Encuesta Nacional de Gastos de los Hogares de Argentina se incluye un documento detallando el modelo de imputación de ingresos a los hogares con pérdida parcial y total de las preguntas de ingreso; para la Encuesta Permanente de Hogares (EPH) se exponen los lineamientos de la imputación en el Aglomerado Gran Buenos Aires por medio de solapamiento con la onda anterior y hot deck jerárquico.

El Bureau of Census de Estados Unidos expone en Internet (27) metodología para mantener la calidad de la información de grandes bases de datos, mediante la "edición" e imputación de datos faltantes y contradictorios; la comparación de las bases originales e imputadas se realiza en términos del reestablecimiento de los verdaderos promedios y de la preservación de la estructura de las correlaciones y, ante desempeños similares de los sistemas, se tienen en cuenta disponibilidad computacional y de personal que apoye el funcionamiento del sistema elegido. Detalla, además, los procedimientos de imputación para los censos de población de 1990 y 2000.

Statistics Canada, que elabora y presenta frecuentemente material sobre esta técnica, expone en la Conferencia de Estadísticos Europeos (28) un documento sobre "Imputación de variables demográficas del Censo de Población Canadiense del 2001" realizada con el sistema CANCEIS para variables cualitativas y cuantitativas mediante el método del vecino más cercano, concluyendo sobre las cualidades y adopción del programa en numerosas encuestas que lleva a cabo (29).

Statistics Finland presenta en la misma Sesión de Trabajo de Edición de Datos Estadísticos (28), una propuesta de imputación a través de mapas auto organizados con estructura de árbol.

El Instituto Nacional de Estadística de Italia (ISTAT) presenta en la misma reunión (28) y en la 53ª Sesión de ISI (30), un documento en el que se evalúa

un software para captación de errores e imputación, con respecto a la propuesta de Canadá, teniendo en cuenta la preservación de los valores originales y de las distribuciones marginales.

Software para Imputación

En este área el desarrollo de programas computacionales no ha sido paralelo a las continuas propuestas teóricas. Los software para análisis estadístico en general han descartado las unidades incompletas al aplicar los distintos procedimientos, con escasas excepciones que proponen casos disponibles, imputación por el promedio o por regresión, sin análisis previo del mecanismo de pérdida.

El programa BMDP (31) fue el primero en incorporar módulos especiales para sustituir los valores faltantes. SPSS presenta en la actualidad un módulo que permite un análisis más riguroso del esquema y del mecanismo de pérdida y estimaciones a partir de conjuntos de datos con pérdidas mediante máximo verosimilitud (32).

Algunos programas son difundidos por sus propios autores en la bibliografía, o por las instituciones que los implementan, otros puestos a disponibilidad en Internet o comercializados con el respaldo de los autores.

En los últimos años el mayor desarrollo y difusión de programas para imputación, se ha producido dentro del área de imputación múltiple, técnica que logró su difusión tardíamente por ausencia de algoritmos flexibles para su aplicación (33). Algunos de los programas o módulos específicos dentro de programas, para la aplicación de esta técnica son accesibles via Internet, con la temporalidad que caracteriza a este medio, mientras que otros se comercializan. Cada uno ha sido desarrollado según distintas metodologías para realizar las imputaciones y las características de los datos afectados.

El programa SAS y SAS/STAT versión V8.2 presenta dos procedimientos (34), que permiten aplicar imputación múltiple con tres opciones que dependen del tipo de esquema de pérdida: para monótonos, una propuesta paramétrica con método de regresión asumiendo normalidad multivariada y una propuesta no paramétrica a través de escores; para esquemas generales el método MCMC, con un supuesto de normalidad multivariada.

Schafer (36) pone a disposición los programas NORM (algoritmos basados en un modelo normal multivariado), CAT (algoritmos para datos categóricos multivariados basados en un modelo multinomial saturado y modelos loglineales) y MIX (algoritmos para datos mixtos continuos y categóricos), PAN (para datos de panel).

King et al. (37) han desarrollado el programa Amelia para el análisis de datos multivariados incompletos a través del algoritmo EM.

SOLAS para Datos Faltantes 3.0 (38), se comercializa y permite la elección de seis técnicas de imputación, dos de ellas para imputación múltiple no iterativa, basadas en el trabajo de Rubin (6).

MICE V.1.0 (Imputación Multivariada con Ecuaciones Encadenadas) rutina en S-Plus realizada por van Buuren et al. (24), permite al usuario agregar propuestas para imputar, particularmente útil en el análisis de casos no ignorables.

La librería para Datos Faltantes de S-Plus V6 realiza imputación múltiple bajo modelos gaussianos, loglineal y gaussiano condicional, basándose en el trabajo de Schafer (39).

Reuniones Científicas, Talleres y Cursos

Es destacable la alta frecuencia con que el tema de imputación ha ido ocupando lugar en forma creciente y está presente actualmente en las ponencias de los eventos científicos, en sesiones y talleres especiales o en reuniones científicas desarrolladas para este tema específico (40; 41; 42; 43; 44; 45; 46; 47; 30; 48; 49; 50; 51).

Es también indicativo el análisis del temario de cursos de difusión dictados en los últimos años, hecho que contribuye a la difusión de los aspectos desarrollados dentro del método y a su utilización más rigurosa (52; 1).

Conclusión

El análisis de las fuentes de información de actualidad permiten realizar un diagnóstico del estado de los conocimientos y de la vigencia de una determinada metodología.

Con respecto al método de imputación, no sólo mantiene su vigencia, sino que se presenta como tema en continua revisión, tanto por parte de los pro-

ductores de información, que lo encuadran y lo unen a la edición en su objetivo de maximizar la calidad de la información a brindar, como también por parte de los investigadores que continuamente realizan y evalúan distintas propuestas ligadas fundamentalmente al mantenimiento de la estructura de los conjuntos de datos originales, a la evaluación del efecto causado por apartarse de los supuestos originales de cada modelo, a la estimación de la variancia asociada a las estimaciones muestrales basadas en datos con imputaciones y a la operatividad en la aplicación.

Frente al objetivo de trabajar con información de calidad, la imputación constituye una herramienta válida para lograrlo, pero requiere la profundización de los aspectos que esta técnica ha ido desarrollando para optimizar su aplicación.

Bibliografía

- Allison, P. Curso "Tratamiento de Información Faltante por medio de Imputación Múltiple y Máxima Verosimilitud". Univ. de Pennsylvania. <http://www.sierrainformation.com/seminarPDF/>, mayo 2002.
- Groves, R.; Dillman, D.; Eltinge, J.; Little, R. (2002). "Survey Nonresponse". W. & Sons.
- Laaksonen, S. 1999. "How to find the best imputation technique. Tests with three methods". International Conference on Survey Nonresponse, Portland. <http://www.jpsm.umd.edu/icsn99/papers/Laaksonen.htm>
- Little R.; Rubin, D. 1987. "Statistical Analysis with Missing Data". J. Wiley & Sons.
- Rancourt, E. 2001. "Edit and Imputation: From Suspicious to Scientific Techniques". Ponencia en 53ª Sesión de ISI-CD. Seúl.
- Rubin, D.. 1997. "Multiple Imputation for nonresponse in surveys". Wiley & Sons.
- Allison, P.D. 2000. Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research*, 28, 301-309.
- Schafer, J.; Olsen, M.. "Multiple Imputation for Multivariate Missing-data Problems: A Data Analyst's Perspective". <http://www.stst.psu.edu/~jls/mbr.pdf>, marzo 2002.
- Lee, H; Rancourt, E.; Särndal, C.. 2002. "Variance Estimation from Survey Data under Simple Imputation", en Groves et al.. "Survey Nonresponse", Wiley & Sons.
- Kalton, G. 2000. "Developments in Survey Research in the Past 25 Years". *Survey Methodology*, 26, 3-10.
- Christianson, A.. 2002. "Avoid the Need to Impute", Kalton,

- G. and MacMillan, M. "Comments", *The Survey Statistician*, Nº 46.
12. Schafer, J.; Schenker, D. 2000. "Inference with imputed conditional means", *JASA*, **95**, 144-154.
13. Shao, J.; Wang, H. 2001. "Sample Correlation Coefficients Based on Survey Data Under Regression Imputation", *JASA*, **97**, 544-552.
14. Yuan, Y. C. 2002. Multiple imputation for missing data: Concepts and new development. SUGI Paper 267-25, <http://www.sas.com/rnd/app/papers/>, agosto 2002.
15. Schafer, J. 2002. "Robustness of a Multivariate Normal Approximation for Imputation of Incomplete Categorical Data", Ponencia #300835, Joint Statistical Meeting, NY.
16. Aerts, M.; Claeskens, G.; Hens, N.; Molenberghs, G.. 2002. "Local Multiple Imputation", *Biometrika*, **89**, 12: 375-388.
17. Zador, P.; Judkins, D.. 2002. "Comparing Regression Imputation with MART and Hot Deck Imputation", Ponencia en JSM 2002, Abstract #301977, NY.
18. Belin, T.R., Diffendal, G.J., Mack, S., Rubin, D.B., Schafer, J.L., Zaslavsky, A.M. 1993, "Hierarchical Logistic Regression Models for Imputation of Unresolved Enumeration Status in Undercount Estimation" (with discussion), *Journal of the American Statistical Association*, **88**, 1149-1166.
19. Wung, L. 2002. "A simulation Study to Evaluate the Robustness of Recent Methods for Preparing Variance Estimates in the Presence of Hot Deck Imputation", Ponencia #301201, JSM 2002, NY.
20. Chen, J.; Shao, J.. 2001. "Jackknife Variance Estimation for Nearest-Neighbor Imputation", *JASA*, **96**, 453: 260-269.
21. Robins, J.M.; Wang, N.. 2000. "Inference for Imputation Estimators", *Biometrika*, **87**, 1: 113-124.
22. Tang, G.. 2002. "An Imputation Method for Data with Non-Ignorable Nonresponse", Pon. #301532, JSM 2002, NY.
23. Wang, J.. 2002. "Handling Incomplete High Dimensional Longitudinal Data by Multiple Imputation Using a Longitudinal Factor Analysis Model", Pon. #300810, JSM 2002, NY.
24. Van Buuren, S. and Oudshoorn C.G.M. 2000. Multivariate imputation by chained equations: MICE V1. User's Manual. Rep. PG/VGZ/00.038. www.multiple-imputation.com, agosto 2002.
25. EUREDIT. <http://www.cs.york.ac.uk/euredit>, marzo 2002; UNECE. www.unece.org, junio 2002.
26. Instituto Nacional de Estadística y Censos (INEC). <http://inec.mecon.gov.ar/>, oct. 2002
27. Williams, T. 1998. "Imputing Person Age for the 2000 Census Short Form: A Model-Based Approach", U.S. Bureau of the Census, Statistical Research Division Report Series, Nº RR98/07. U.S. Bureau of the Census, www.census.gov, julio 2002.
28. Work Session on Statistical Data Editings. Helsinki 2002. www.unece.org/stats/documents/2002.05.sde.htm, octubre 2002.
29. Kovar, J., Winkler, W.. "Comparison of GEIS and SPEER for editing Economic Data", Statistics Canada and Bureau of the Census, www.census.gov, julio 2002.
30. International Statistical Institute (ISI). 53^o Session. <http://www.nso.go.kr/isi2001/>, julio 2002.
31. BMDP. www.statsol.ie/bmdp/bmdp.htm, oct. 2002.
32. SPSS. http://www.spss.com/spssbi/missing_value/, oct. 2002.
33. Horton, N.J. and Lipsitz, S.R. 2001. Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *American Statistician*, **55**, 244-254.
34. SAS Institute Inc., Statistics and Operations Research. "What's new in data analysis. multiple imputation for missing data", 2002. <http://www.sas.com/rnd/app/da/new/dami.htm>, mayo 2002.
35. Universidad de Pensilvania. <http://www.upenn.edu/>, oct 2002.
36. Schafer, J. 1997. "Analysis of Incomplete Multivariate Data", Chapman & Hall.
37. King, G.; Honaker, A.; Scheve, K.. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation", <http://Gking.Harvard.Edu>, Dic. 2001, julio 2002.
38. SOLAS. www.statsol.ie/solas/solas.ht, marzo 2002.
39. S-Plus. www.insightful.com, septiembre 2001.
40. Conferencia Internacional sobre No Respuesta en Encuestas, Portland 1999. www.jpsn.und.edu/icsn99/junio99, julio 2002.
41. Taller sobre Estadística con Datos Deficientes, Munich 2000. www.stat.uni.muenchen.de/, julio 2002.
42. Reuniones Conjuntas de Estadística. Indianápolis 2000. www.statcan.ca/english/concepts, julio 2002.
43. COMPSTAT. Conferencia y Taller sobre Imputación Múltiple, Utrecht 2000. <http://neon.vb.cbs.nl/rsn/>, julio 2002.
44. 5^a Conferencia Internacional de Metodología en Ciencias Sociales. Colonia 2000, www.za.uni-koeln.de/rc33/index.htm, julio 2002.
45. Comisión Económica de ONU. Sesión de Trabajo sobre Edición Estadística de Datos, Cardiff 2000. www.unece.org/stats, julio 2002.
46. 8^o Simposio Bienal en Métodos Estadísticos. Atlanta 2001. www.cdc.gov/od/ads/, julio 2002.

47. Simposio sobre Análisis de Datos Incompletos. Programa de prevención, imputación y análisis de la inf. faltante, Utrecht 2001. www.vvs-ssp.nl/syposium2001.html, julio 2002.
48. Baltic-Nordic Conf. on Survey Sampling 2002. www.matstat.umu.se/banocoss/, julio 2002.
49. Sesión de Trabajo en edición de datos de UNECE. Helsinki 2002. www.unece.org/stats/, sept.2002.
50. Joint Statistical Meeting. Nueva York 2002. www.amstat.org/meetings/jsm/2002/, oct. 2002.
51. 54ª Sesión del Instituto Internacional de Estadística, Berlín 2003. www.isi2003.de/, oct.2002.
52. Longford, N..Curso "Datos Faltantes en Encuestas de Gran Escala", Plymouth University, 2002. www.tech.plym.ac.uk, oct. 2002.