

Información faltante en experimentos biológicos. Una solución mediante imputación múltiple a través del programa SAS®

Badler, C. E.; Alsina, S. M.; Puigsubirá, C. R.; Vitelleschi, M. S.*

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario. Bvrd. Oroño 1261. (2000) Rosario, Santa Fe, Argentina.
Teléfono: 0341-4802793-int.151. E-mail: salsina@unr.edu.ar

RESUMEN: La existencia de información faltante en bases de datos provenientes de experimentos diseñados en el área biológica, en general, conducen a análisis sesgados e inferencias incorrectas. Una alternativa para solucionar esta problemática es la técnica de imputación múltiple, que consiste en reemplazar cada valor faltante por un vector de posibles valores y combinar las estimaciones obtenidas a partir de cada uno de ellos. La misma permite incorporar información auxiliar y una medida de variabilidad a partir de la información proporcionada por los distintos valores imputados.

En los últimos años la disponibilidad de dicha técnica en los programas computacionales, entre ellos el Statistical Analysis System (SAS), permite su aplicación en forma operativa y evaluar la eficiencia del método utilizado.

En este trabajo se presenta una aplicación a datos de un experimento del área biológica, generando pérdidas según el mecanismo de pérdida al azar (MAR). Se aplica la técnica de imputación múltiple a través de las diferentes opciones propuestas por el programa SAS y se comparan las eficiencias relativas obtenidas a partir de cada una de ellas.

De esta manera se pretende proporcionar un procedimiento válido para enfrentar la falta de información que limita el análisis y puede distorsionar las conclusiones del mismo.

Palabras claves: Imputación Múltiple, Información Faltante, SAS.

SUMMARY: Missing information in biological experiments. An approach through multiple imputation with the statistical analysis system. Badler, C. E.; Alsina, S. M.; Puigsubirá, C. R.; Vitelleschi, M. S. Missing information in data sets from designed experiments is a commonly occurring complication for scientific investigation in the biological area, because it usually produces loss of efficiency and biased analysis. One approach to the problem of incomplete data is multiple imputation. This statistical technique replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute, and combines the estimates that were obtained from the application of the statistical analysis on each imputed data set.

In recent years a number of software packages, as the Statistical Analysis System (SAS), have facilitated the implementation of this procedure and the evaluation of the method.

An application is presented using information from a data set of an experiment in the biological area in which missing values are generated according to a MAR mechanism. The SAS different procedures are used to implement multiple imputation by different imputation methods and the relative efficiencies obtained from each one are compared.

This paper attempts to present a valid procedure to deal with missing information, a real problem that limits the analysis.

key words: Multiple imputation; Missing data, SAS.