Research and Statistical Support
University of North Texas

RSS Matters

Dealing With Missing Data

By Patti Price, RSS Statistical Consultant

In dealing with missing data, one solution is to employ a Monte Carlo approach using a program for creating multiple imputations. While these programs may be purchased as part of packages like SPSS, other similar programs are available for free download. If you are using S-Plus, there are four different packages that may be used as functions in S-Plus. These include NORM (for multivariate continuous data), CAT (for multivariate categorical data), MIX (for mixed continuous and categorical data), and PAN (for panel or clustered data).

A stand-alone version of NORM is also available for those using Windows 95/98/NT. Work is in progress for stand-alone versions of the other programs listed above. Each of these programs was developed by Dr. Joseph Schafer and is available at http://www.stat.psu.edu/~jls/misoftwa.html - top.

Specific information on frequently asked questions concerning multiple imputation is available at http://www.stat.psu.edu/~jls/mifaq.html.

After downloading and installing the NORM program, you will find that there are some example files to work with. Your own files will need to be saved in the .dat format. After opening the file, you will note that there are four file folder tabs to work with. In the data tab, you will see your data and will need to enter the value assigned to your missing data. In the data folder tab, it is possible to enter variable names and to obtain basic descriptive information. To run the complete program, simply click on the "EM Algorithm" tab and click run, the "Data Augmentation" tab and click run, and finally click on the "Impute from parameters" tab" and run to complete the process.