Missing Data: Listwise vs. Pairwise

Listwise and pairwise deletion are the most common techniques to handling missing data (Peugh & Enders, 2004). It is important to understand that in the vast majority of cases, an important assumption to using either of these techniques is that your data is missing completely at random (MCAR). In other words, the researcher needs to support that the probability of missing data on their dependent variable is unrelated to other independent variables as well as the dependent variable itself.

Listwise deletion (complete-case analysis) removes all data for a case that has one or more missing values. This technique is commonly used if the researcher is conducting a treatment study and wants to compare a completers analysis (listwise deletion) vs. an intent-to-treat analysis (includes cases with missing data imputed or taken into account via a algorithmic method) in a treatment design. In most other cases, it is often disadvantageous to use listwise deletion. This is because the assumptions of MCAR are typically rare to support. Because of this, listwise deletion methods produce bias parameters and the estimates. An example would be using GRE scores to predict grades in graduate school. Perhaps the researcher had obtained GRE scores from candidates’ applications. Those with low GRE scores may not have been accepted to a graduate program. If the researcher used listwise deletion, their parameters and estimates for predicting success would be biased because those with low scores were not accounted for. An easy way to think how the parameters and estimates would be bias in this example would be to think of how much different the means would be by including vs. excluding those with low scores.

Pairwise deletion (available-case analysis) attempts to minimize the loss that occurs in listwise deletion. An easy way to think of how pairwise deletion works is to think of a correlation matrix. A correlation measures the strength of the relationship between two variables. For each pair of variables for which data is available, the correlation coefficient will take that data into account. Thus, pairwise deletion maximizes all data available by an analysis by analysis basis. A strength to this technique is that it increases power in your analyses. Though this technique is typically preferred over listwise deletion, it also assumes that the missing data are MCAR. There are disadvantages as well. A disadvantage with the use of pairwise deletion is that the standard of errors computed by most software packages uses the average sample size across analyses. This tends to produced standard of errors that are underestimated or overestimated. Researchers have also associated pairwise deletion as a source for nonpositive definite matrices in multivariate and contemporary statistical analyses – such as Structural Equation Modeling (Little, 1992; Marsh, 1998; Wothke, 1993).
References:


