WinLTA Version 3.1 Release Notes

Current release: 3.1.0
(Previous release: 3.0.2)

User-interface enhancements

Limited Internationalization Support
WinLTA is now compatible with international numeric formats that use a comma as a decimal separator. The user-interface displays numbers according to the user's locale setting, and the control file stores numbers using a decimal point (.) as the decimal separator.

Operating System Compatibility
WinLTA is now compatible with Windows 95, 98, ME, NT4, 2000, and XP.

Enhanced Online Help
Online help has been upgraded to "HTML Help", which is used in most Microsoft products. The help is context-sensitive, i.e., you can obtain help on the current tab by pressing [F1].

Improved Error Messages and Diagnostics.
Common errors are now handled more gracefully and better user feedback is provided when errors occur.

Large Models
Output file can now properly handle output for models with large numbers of latent classes and/or statuses.

Analytical enhancements

WinLTA 3.1 has two analytic features that were not available in WinLTA 2.3. First, data augmentation is now available as a method for obtaining standard errors of all parameter estimates. Second, WinLTA 3.1 includes improved handling of model fit assessment and residuals when there are missing data. This version now adjusts the G^2 fit statistic and residuals for missing data, and provides a test for whether the data are missing completely at random. Please read these release notes for a brief introduction to the new features available in WinLTA 3.1.

Data Augmentation
Background. LTA uses an EM algorithm to obtain maximum likelihood estimates for all parameters. Because standard errors are not a byproduct of the EM algorithm, this method does not allow the user to conduct hypothesis tests. In order to obtain standard errors, WinLTA 3.1 has added a data augmentation (DA) routine. DA, an iterative simulation procedure similar to the EM algorithm, is a variety of multiple imputation, which is a general approach to missing data problems. DA operates in LTA by treating the latent class and latent status variables as

missing data and multiply imputing them. This procedure produces a sequence of imputed data sets, each of which produces a plausible set of LTA parameter estimates. Parameter estimates and their standard errors are retained from each imputation, and are combined. The resultant parameter estimates and their standard errors can then be used to describe the variability about the parameter estimates. Advanced WinLTA users can conduct hypothesis tests.

Procedure. The user who wishes to employ DA should begin by fitting an LTA model to the data to obtain maximum likelihood estimates based on the EM algorithm, just as in the past. Once a satisfactory model is achieved, these EM estimates are used as starting values in the DA procedure. DA is run in LTA by switching the *Run Mode* from EM Mode to DA Mode. This approach yields multiple imputed data sets, where the number of data sets is specified by the user, and the within-imputation and between-imputation variability are combined to yield an overall estimate of the standard error for each parameter. This procedure provides information about the variability of point estimates. Although WinLTA calculates the final parameter estimates and standard errors based on data augmentation, the user must examine several plots to assess convergence of the DA algorithm. See the *WinLTA User's Guide for Data Augmentation* for more details on using data augmentation in WinLTA and to see an example.

The G^2 Statistic for LTA Models Involving Missing Data

The likelihood ratio statistic G^2 is used as a goodness-of-fit measure in WinLTA. When the data set contains missing data the G^2 is inflated, making model fit appear worse than it actually is. This is because the G^2 actually reflects two quantities: a component of the loglikelihood for model fit and a component of loglikelihood for the missing data mechanism (Little & Rubin, 1987). In order to assess the fit of a model properly when applied to incomplete data we need to separate these two components.

This is done by fitting the saturated, or unrestricted, model to the incomplete data. Although the saturated model provides perfect fit for complete data, the G^2 statistic for the saturated model can be quite large if the data are not missing completely at random. WinLTA 3.1 uses the EM algorithm described in Little and Rubin (1987) to estimate the saturated model for the incomplete data set. This statistic tests the null hypothesis that the data are MCAR.

The appropriate G^2 statistic for assessing model fit of the LTA model is obtained by subtracting the G^2 based on the saturated model from the typical G^2 statistic. Note that conclusions based on the typical unadjusted G^2 statistic (as provided in WinLTA 2.3) are conservative. That is, without adjusting this statistic for the missing data mechanism we are more likely to reject an LTA model that provides adequate fit, but one is not more likely accept an LTA model as sufficient if it is not.

Example 5 of the Appendix to the WinLTA User's Guide provides an example of the adjusted G^2 statistic for a model fit to incomplete data.

Residuals for LTA Models Involving Missing Data

Residuals, or the difference between the observed cell counts and the cell counts expected under a given model, can be a useful tool for improving the fit of a model to complete categorical data. With incomplete data, however, the discrepancy between the observed cell counts and the counts expected under the LTA model reflect both lack of fit and the effect of missing data. The greater the departures of the data from missing completely at random (MCAR), the less useful the typical residuals become in assessing model fit. Because of this, typical residuals often are not particularly useful in detecting parts of the model that fit poorly.

WinLTA 3.1 now produces a *fit index* for each response pattern that controls for the effect of missing data, reflecting only lack of model fit. These fit indices can be useful for diagnosing poorly-fitting models in the same way that residuals can be used in LTA models fit to complete data.

The fit index associated with a response pattern equals the difference between the expected cell count under the saturated model and the expected cell count under the LTA model. We also provide a *scaled fit index* for each cell, which scales the index by the cell count expected under the LTA model. A large scaled fit index indicates a substantial lack-of-fit contribution for that cell in the contingency table.

This procedure is still under development, therefore we have several cautions for the user. First, note that this method involves every *possible* response pattern, and therefore the output associated with this procedure can be quite large. Be sure to examine the size of the output file before printing, and note that there is an option to suppress printing of the fit indices. Second, response patterns associated with very small expected cell counts (i.e. < .00005) can produce extremely large scaled fit indices due to the relative difference in size between expected counts under the saturated and LTA models. We recommend that adjustments to the model be based on the largest scaled fit indices which have expected cell counts greater than 1. Third, the expected cell counts under the saturated and LTA models are printed for all possible *complete-data* response patterns, but not for response patterns involving incomplete data. Fourth, the user should be aware that the scaled fit indices may lack statistical power, and thus adjustments made to the model based on large scaled fit indices are conservative decisions.

See Example 5 for an example of fit indices for a model fit to incomplete data.

Release History

Release 3.1.0

Response patterns in the data may now have non-integer frequencies (counts)
Repaired array bounds problem related to Data Augmentation runs with missing data.

Release 3.0.2

Fixed problem that prevented cross-validation runs from working.
Repaired application of fixed parameter restrictions in Data Augmentation runs.

Release 3.0.1

Initial release.

References

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.