



Cognalysis MultiRate™

Version 6 Release Notes

July 2021

Gross Consulting is pleased to announce the release of Cognalysis MultiRate™ Version 6. This release makes MultiRate even more powerful for building understandable, robust, and insightful models of data.

Important Note on Backwards Compatibility

Version 6 is able to open files created in previous versions of MultiRate, but **Version 5 users will not be able to open files created in Version 6**. If you are working with other MultiRate users and will be sharing files, it is recommended that all users update to Version 6 at the same time.

Version 6 will overwrite previous versions of MultiRate. You may be required to re-register the software. If so, follow the steps in the software to request a new registration code.

Changes with this release:

- Changes to the use of bootstrap test samples.
- Common determination of bootstrap samples for identical data tables
- Modified and unmodified considered as separate models in the comparison
- Credibility Strengthening at the beginning of the Improvement Process.
- Introduction of new Test Data Statistic (TDS)
- Separate COM and NET DLLs
- Variable Gradient – Cyclical

Changes to the use of bootstrap test samples.

When testing two models for the statistical significance of the differences between them, the default is now to do a binomial test of goodness of fit between bootstrapped samples of the test data.

The ratio or Relative Net Lift to Average Absolute Error is calculated for each bootstrap sample for each model. For example, bootstrap sample 1 might include 2 occurrences of test record 1, 0 occurrences of test record 2, 1 of test record 3, etc. The ratio of RNL to AAE would be tabulated for bootstrap sample 1 under both models. This is repeated with each of the other bootstrap samples.

In the past, a Kolmogorov-Smirnov test was performed between the two sets of fit statistics (one 'distribution' taken to be the set of fit statistic from the bootstrapped samples from one model, and the other 'distribution' taken to be the set from the other model). We found this approach to be too forgiving. In some cases, there was a wide dispersion of results across the samples, and this resulted in a high threshold for rejection, comparing one distribution to the other, even if for every bootstrapped sample one model had a better fit. **Now the default approach is to measure the results between the two models explicitly for each bootstrapped sample.** For each sample, if the null hypothesis of no difference were correct, each model would be as likely to be a better fit as the other. By tallying the winner across samples and comparing to the total number of samples, a binomial test is performed. To give greater granularity for this test, **the number of bootstrapped samples was increased from 20 to 100.**

This functionality is being performed for non-MultiRate models as well if such are selected in the comparison tool.

The K-S test is still available to the user as an option

Common determination of bootstrap samples for identical data tables

In the past, for the first model in a chain of models linked together through the "First Iteration" box (e.g. [model_3] has [model_2] as the 'first iteration', [model_2] has [model_1] as the 'first iteration'), a set of bootstrap samples was determined and then used for all subsequent models. In cases where this chain is broken, a new set of bootstrapped samples was generated. For the K-S test this was OK, but to use the improvement of the binomial test described above, it is critical that the samples are identical between models. **Now the seeding of the random selection is directly tied to the test data such that identical test datasets will generate the same bootstrapping.** This has the additional advantage of generating exact reproduction of results if for some reason the analysis needs to be re-run.

Modified and unmodified considered as separate models in the comparison

The use of curve modification is a useful way to improve model fit and expand from the purely multiplicative nature of the models. It does, however, increase the complexity of the model. The tradeoff between fit and complexity has always been considered when comparing models to each other, but has not been when comparing a model with and without curve modification. **Now in the comparison tool, and also through the auto-improve process, each model is treated as two models, the one without curve modification and with curve modification, explicitly considering and illustrating this tradeoff.**

Credibility Strengthening at the beginning of the Improvement Process.

In the past when allowing the software to recommend improvements to a model, alternative credibility standards were tested simultaneously with other changes. A model could be seen as superior if favorable impacts of giving more credibility were larger than negative impacts from dropping a variable. Since the autoimprove process never tries to add variable back in, there was the potential to lose some information. Now, the first step in the improvement process is to reduce the credibility standard (increase credibility) until no meaningful improvement is found on the test data and then start the other adjustments in the improvement process.

Introduction of new Test Data Statistic (TDS)

Most fit statistics only consider measurement at the record level, that is then aggregated. If results are independent across the test records, this should be sufficient, but if there is a consistent bias across sections of or all of the test data, this would not be adequately captured. TDS attempts to measure this problem and compare it across models.

First the data is organized into a number of partitions based on observations ordered by the model (or modified) result.

$$\sum_i |T_i - M_i| \times \frac{\sqrt{\sum_i E_i}}{\sum_i \sqrt{E_i}}$$

Where T_i , M_i and E_i are the target, model, and exposure respectively, summed across all test records in partition 'i'.

This is performed at multiple level of granularity with number of partitions variously equal to { 1, 3, 10, 30, 100, N }, where N is the number of records in the test data.

The calculation with a single partition is based on the total observed bias. The calculation with N partitions is based on the Average Absolute Error (record level). The other 4 calculations are

considering bias over segments of the model curve. **The TDS score is taken as the maximum of the calculation across these alternative segmentations.** For each bootstrapped test sample, the TDS is shown, together with the partitioning that gave resulted in the maximum, on the `_metadata` table.

In tests of TDS, we found that it was not strong enough to replace the fit statistics that we are currently using outright to determine the strength of fit (Average Absolute Error and Relative Net Lift). Instead, it is being used in the determination of which models are considered acceptable alternatives to the best fitting model. Now, even if the binomial or K-S test indicates that two models are statistically different than each other, if any two of the following tests (AAE, RNL, RMSE, TDS) are better for the otherwise inferior model, the model is retained as being acceptable).

Separate COM and NET DLLs

The DLLs form a convenient way to incorporate your model into other applications. In the past, it was a challenge to incorporate the DLL into a COM environment such as Microsoft Excel. To better facilitate this, **the software now generates two versions of the DLL- one explicitly designed for a COM environment, and one for a .NET environment.** The `.txt` file that is created with each gives additional detail about how to incorporate the dll, including sample Excel VBA code in the case of the COM version.

Variable Gradient – Cyclical

This is a new smoothing type, available to the user. In cases where the lowest values are expected to be similar in effect to the highest values (e.g., hour of the day), this smoothing will consider bin proximity to be based on a cycle rather than a line.

Staying with the hour of the day example, with each value getting a bin, when considering bin '1', the information from bin '24' would be used in the same way as bin '2', bin '23' the same as bin '3', etc.