

Preparing for Scoring

Before you start using a newly built model to score data, some preparation of the data might be required. For example, in database target marketing, the data to be scored might be many times larger than the data that is used to develop the model, and it might be stored in a different format. It's essential for the scoring data to be comparable to the training and validation data that were used to build the model.

If any modifications were made to the data before the model was built and validated, then the same modifications must be made to the scoring data before scoring. This might include missing value imputation, transformations, and derivation of inputs through standardization or the creation of composite variables from existing variables. This is essential for the deployed model to be able to map the scoring functionality to the new data.

Making the same modifications becomes more complex if the original modifications were based on parameters that were derived from the training data set, such as the mean or standard deviation. This means that if you subtracted the mean of a variable from the training and validation data, then the mean of that variable should also be subtracted from the scoring data. This practice keeps the different data sets comparable. The process of preparing the data for scoring can be time- and resource-intensive.

Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression

Copyright © 2019 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close