

Case-Control Matching Algorithm

Case-Control Matching

The case-control matching algorithm matches a predefined set of case patients to patients in a control pool. A single case is matched to one or more controls based on data points in common (age, gender, race, number of healthcare encounters, etc...). Data points are binned and converted to their corresponding bin value prior to matching. The i2b2 user running the application indicates the fields to be binned in the PATIENT_DIMENSION table and the number of intervals (bins) for each field. Numeric fields are binned by quantiles specified by the investigator. Character string fields are ranked by frequency; any value with a rank greater than the user supplied threshold is converted to 0 for "other".

The bin results are stored in a temporary table allocated for each match request. Numeric values are binned a total of three times. Once with the original number of requested bin intervals, then twice more with the number of intervals equal to approximately 10% and 20% less than the original number respectively. For example, if the user requests patient age binned into 8 quantiles, three bin results will be generated; one each at 8, 7, and 6 quantiles. The decreasing number of quantiles corresponds to weaker match strengths if these bin values are used to match cases to controls when an exact match cannot be found. This is because decreasing the number of bins increases the size of each interval into which data are placed, thus increasing the allowable difference between controls that may be matched to cases.

Cases are then randomly joined or matched to controls by comparing bin values in decreasing match strength order. Matching is accomplished in a bulk join step and the results are ordered by strength. The desired numbers of controls per case are selected into the final results. Matches with the greatest strength are selected first.

See documentation included with the release for application details.

The first version of the Case-Control Matching algorithm are packaged as MS SQL Server stored procedure scripts that operate on an i2b2 1.6 instance.

Downloads

File	Size	Version	Description	Date
Case_Control_Matcher_1.0b1a.zip	186 KB	1.0b1a	Beta	2013-10-28