to classify the car type and use the extracted words as predictors to see whether the tree will expose any hidden information useful for this study.

- Select Interactive Trees (C&RT and CHAID) from the Statistics-Data Mining menu.
- Click OK to perform a Classification Analysis using the C&RT model.
- Click on the Variables button within the Quick tab to view the Select Dependent Vars, Categorical, and Continuous Predictors.

Remember that we are trying to find the words that can classify (discriminate between) the car type (structured information), and there will be many reviews that mention about the brand name within the text corpus. Therefore, we need to eliminate the words that refer to the brands from the predictor list because if these words are included, they will turn out to be the best classifier/discriminator for the car type. For instance, say we include words like Mercedes Benz, MZ, BMW, Lexus, CarZZ, ES 300, etc. as predictors to classify the car types (CarZZ, BMW, Mercedes, and Lexus). It is quite obvious that the tree algorithm will pick these words that will discriminate one car type against the other, which will not reveal any information that is relevant for our study.

- Select Car Type as Dependent: and all the extracted words from variable 7 to variable 301 as Continuous pred: (as discussed let us exclude the words accord, audi, benz, bmw, carzz, class, es300, honda, infiniti, lexus, mb, mercedes, series, and slk that directly associate to brand names; use the Ctrl key for this purpose).
- Click OK on the Select Dependent Vars, Categorical and Continuous Predictors: dialog and the ITrees C&RT Extended Options: 4Cars.sta dialog to view the ITrees C&RT Results: 4Cars.sta dialog, as shown in Figure G.23.

You now have several options to extract the tree according to your choice (refer to the STATISTICA electronic manual to further study the different options available within this dialog). First, let us grow the tree to its full size and remove one level to resize the tree for easy visual interpretation.

- Click on the Grow Tree button to build the tree and next click on the Remove 1 Level button to reduce the tree by one level.
- Click on the Tree Graph button under the Review Tree: section.

You should now see the tree graph shown in Figure G.24 for car type.

The tree solutions are relatively simple and straightforward for interpretation. As you can see from the graph in Figure G.24, the C&RT algorithm had distinguished eight decision outcomes (contained in eight terminal nodes highlighted in red) built on seven if-then conditions to classify the car type. Terminal nodes, or terminal leaves as they are sometimes called, are points on the tree beyond which no further decisions are made. The tree starts with the top decision node (also called the root node) with all the 638 cases (reviews in our case) predominated by the car type CarZZ category. CarZZ had the highest frequency of reviews among the four car types, as indicated in the histogram. The legend identifying which bars in the node histograms correspond to the four categories is located in the top-left corner of the graph.