

Traceis™ Data Exploration Studio Decision trees

1 Select the grouping step

In this example, the body fat dataset was used. First click on the grouping step.

2 Selecting the decision trees tab

The screenshot displays the Traceis Data Exploration Studio 2007 beta interface. On the left, a vertical sidebar contains a list of steps: 1. Definition, 2. Preparation, 3. Implementation, 4. Deployment. The 'Grouping' step is highlighted with a blue background and a white circle containing the number '1'. The main workspace is divided into two tabs: 'Clustering' and 'Decision trees', with 'Decision trees' being the active tab, indicated by a white circle with the number '2'. The 'Decision trees' tab shows a configuration panel with 'Select variables (descriptors):' containing a list of variables (Density, Percent body fat, Age(years), Weight(lbs), Height(inches), Neck(cm), Chest(cm), Waist(cm)) and 'Select variable (response):' set to 'Percent body fat'. Below this, 'Minimum node size' is set to 30. A 'Display' button is visible. At the bottom of the workspace, a 'Selected items' section shows a grid of histograms for various variables: Density, Percent body fat, Age(years), Weight(lbs), Height(inches), Neck(cm), Chest(cm), Abdomen(cm), Hip(cm), Thigh(cm), Knee(cm), Ankle(cm), Biceps(cm), Forearm(cm), and Wrist(cm).

MAKING SENSE OF DATA

1 Select the descriptors

Select all variables to use as descriptors. In this example, all variables except density and percent body fat were selected. To select multiple non-contiguous variables, use ctrl-click and for contiguous variables use the shift-click.

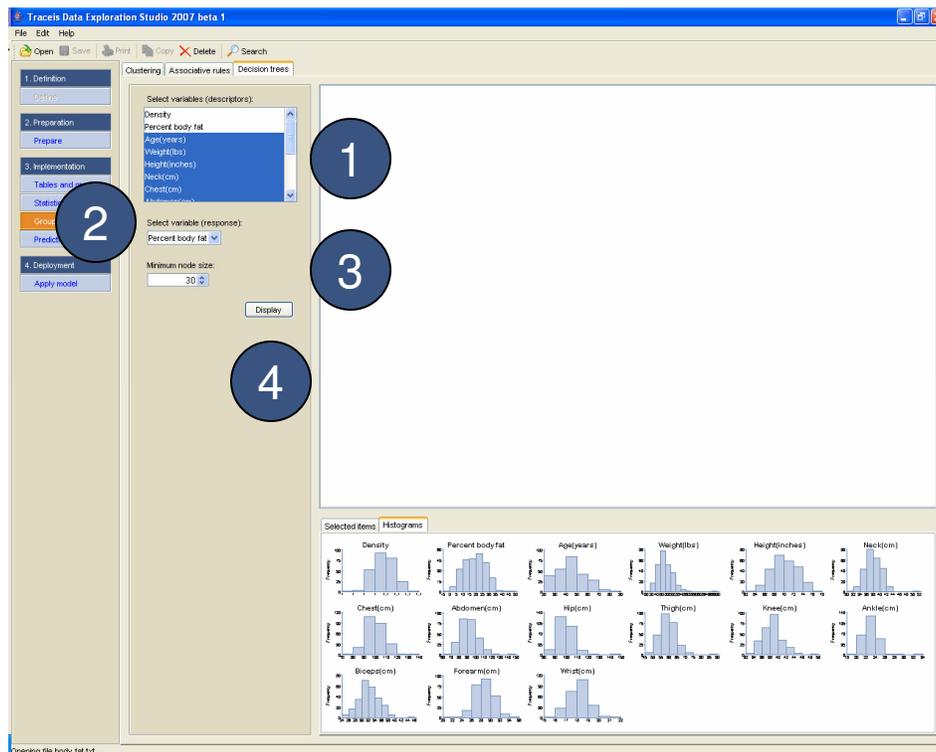
2 Select the response

Select the variable to use as the response. In this example, percent body fat was selected.

3 Select the minimum node size

In this example, a value of 30 was set.

4 Display the results



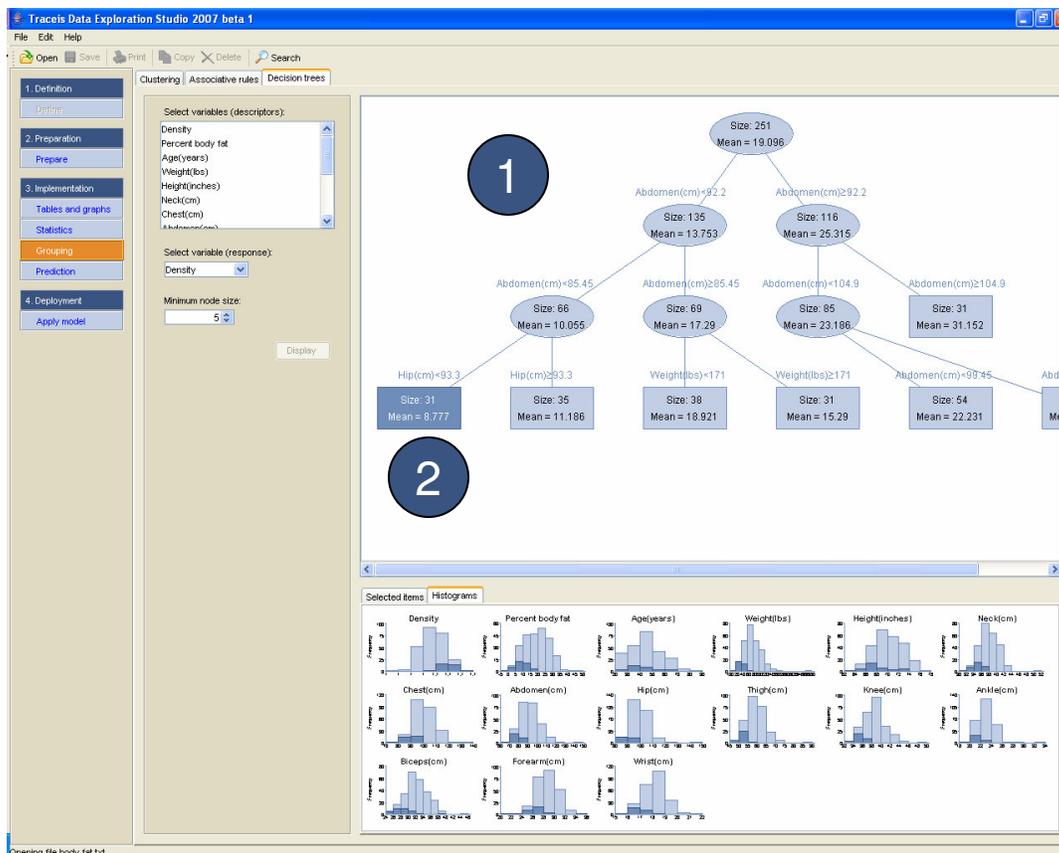
MAKING SENSE OF DATA

1 View the decision tree

Since the response variable is continuous, each node in the tree contains the mean value of the response.

2 Decision tree node selection

Single click on any node to view the observations.



MAKING SENSE OF DATA

1 Select the descriptors

In this example, the diabetes dataset was used and all variables except ID and diabetes were selected as descriptors.

2 Select the response

In this example, the diabetes variable was selected.

3 Select the minimum node size

In this example, a value of 50 was set.

4 Display the results

5 View the decision tree

Since the response variable was categorical, the node's most commonly occurring value is presented along with the proportion of observations in the node containing this value.

