

MAKING SENSE OF DATA

Traceis™ Data Exploration Studio Segmenting

1 Select the prepare step

This example uses the baseball hitters dataset. To segment the data set, first select the prepare step.

2 Select the segment tab

The screenshot shows the Traceis Data Exploration Studio 2007 beta 1.1 interface. On the left sidebar, the 'Prepare' step is selected under the '1. Definition' category. The main window displays the 'Segment' tab, which includes a 'Select the number of observations' dropdown set to 131, an 'Observation selection criteria' section with 'Random' selected, and a 'Select variables' list. The 'Segment' button is visible below the variable list. The main data table shows columns for 'Hitter's name', 'Bat times (96)', 'Hits (96)', 'Home runs (96)', 'Runs (96)', 'Runs batted (96)', 'Walks (96)', 'Player team (96)', 'Player's position (96)', and 'Put outs (96)'. The bottom status bar indicates 'Total observations: 262, Selected observations: 0'.

MAKING SENSE OF DATA

1

Number of observations

Segmenting the data set will result in a smaller data set, and the number of observations in the new set should be specified.

2

Select the type of segmentation

There are two options for segmenting a new data set: (1) creating a random subset, and (2) creating a diverse subset. A diverse subset is generated by first clustering the data set using k-means clustering, into a set of groups based on the desired subset size. The observation closest to the center of each group is selected. To select a diverse set, the variables on which to cluster must be selected.

3

Create a new set of observations

Click on the segment button to create a new set of observations.

