Box-and-whisker plots

We first illustrate how to use box-and-whisker plots to display univariate distributions of level-1 variables for each level-2 unit, with and without a level-2 classification variable. Using the HS&B data, we display graphical summaries of the mathematics achievement variable, MATHACH, and simultaneously show differences in the student scores within a school and among schools.

To prepare box-and-whisker plots

1. From the **HLM** window open the **File** menu.
2. Choose **Create a new model using an existing MDM file** to open an **Open MDM File** dialog box. Open HSB.MDM.
3. Open the **File** menu, choose **Graph Data … box-whisker plots** to open a **Choose Y for box plot** dialog box (see Figure 1).
4. Select MATHACH in the **Y-axis** drop-down list box.
5. Choose the number of groups to be used for graphing. There are three options: (a) **First ten groups**; (b) **Random sample of spec'd prob** (specified probability) and (c) **All groups** (n = total number of groups) for users to choose from in the **Number of groups** drop-down list box. The selection of option (b) requires the user to specify the proportion or percent of the level-2 units to be included. To do so, enter a probability into the text box for **Probability (0 to 1)**. In our example, we randomly select 10 percent of the schools to illustrate. We select **Random sample of spec'd prob** from the **Number of groups** drop-down list box. Enter 0.1 into the text box for **Probability (0 to 1)** to indicate that 10 percent or a proportion of .1 of the schools will be used.
6. Specify the arrangement of the plots by either (a) the original order of the groups as they appear in the data set or (b) the median in an ascending order. Click on the selection button for **median** in the **Sort by** section to arrange the box-and-whisker plots of MATHACH by median in an ascending order (see Figure 2).
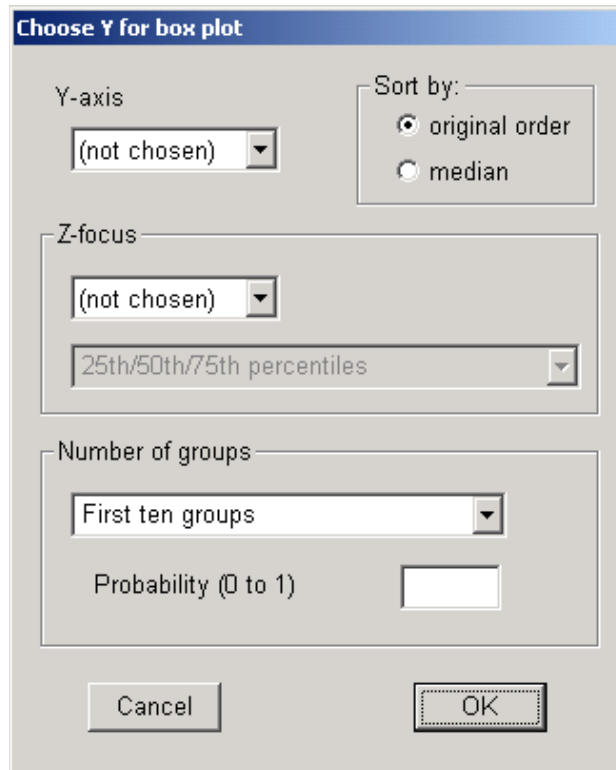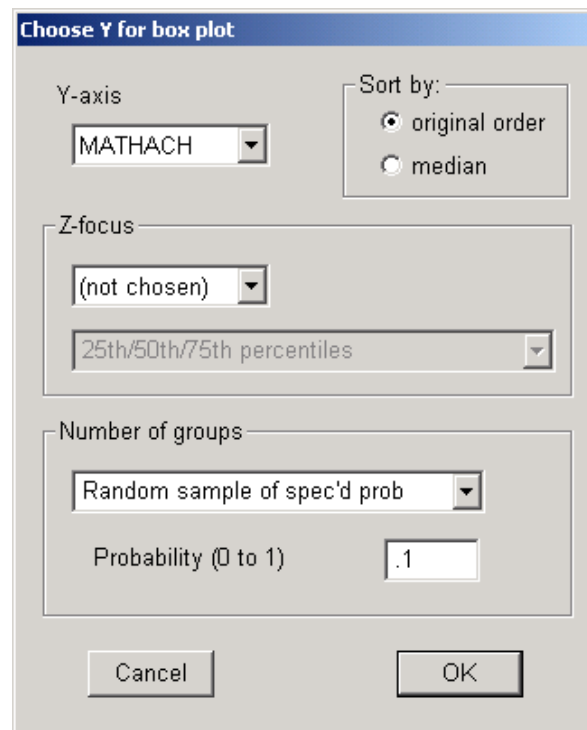
Figure 1   Choose Y for box plot dialog box

Figure 2   Choose Y for box plot dialog box for the MATHACH example

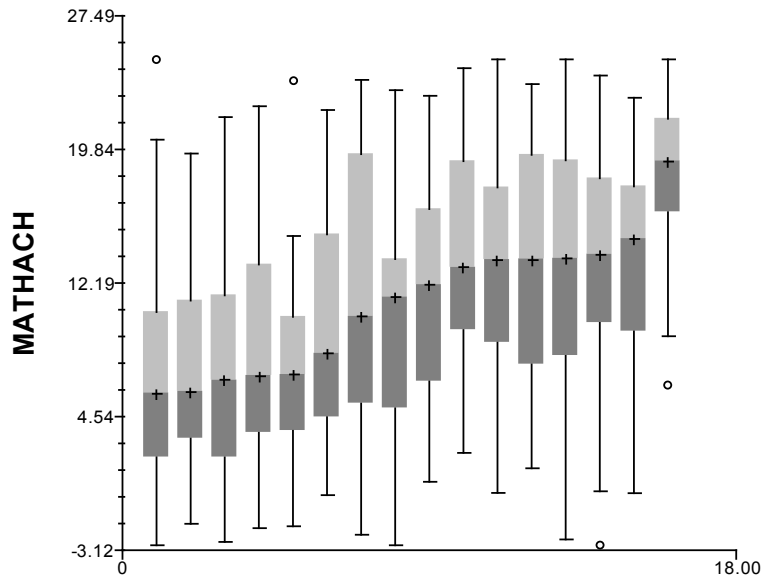7. Click **OK** to display the plots (see Figure 18.3).

Figure 3   **Box and whisker plot for MATHACH**

The figure gives side-by-side graphical summaries of the distributions of MATHACH for the sixteen schools sorted by median. The x-axis denotes number of schools in the display and the y-axis mathematics achievement. The plot tells us that the first school from the left has a median score of about 6.05, which is the lowest school median in this group. The distribution of the scores of the students in this school is positively skewed and there is an outlier at the upper end.

The third and the fourth schools from the left have similar distributions of mathematics scores. Compared to the distribution of the scores of the adjacent school on the right, however, the scores of these two schools display greater variability, as defined by the lengths of the boxes or interquartile ranges. In addition, there is an outlier at the upper end of the distribution for the fifth school. The highest median mathematics score among the 16 schools was 19.08.

8. (Optional) WHLM allows users to list the raw data of a specific group that is graphically summarized in one of the box-and-whisker plots as well. To see the data of a specific level-2 unit, click on one of the box-and-whisker plots (near the median is usually a good place) in Figure 3, which brings up the following dialog box:
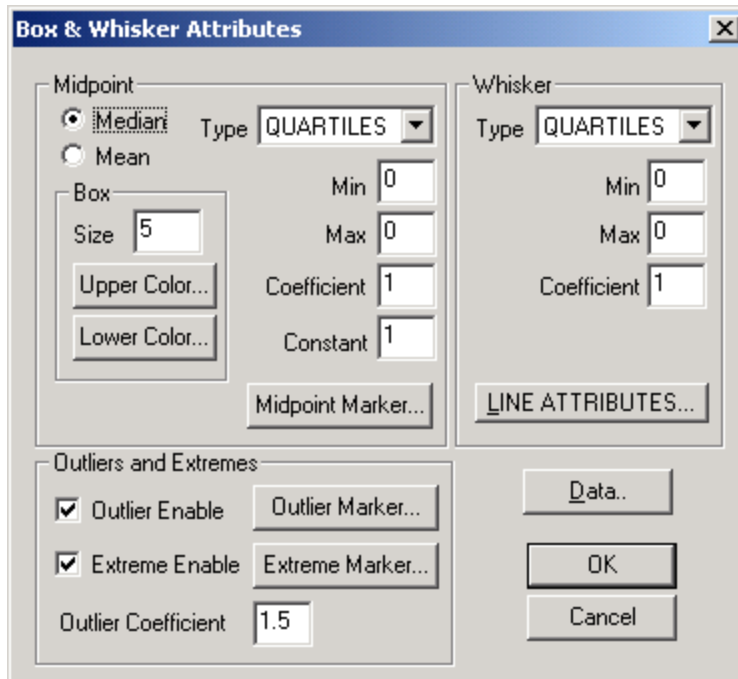
Figure 4   Box & Whisker Attributes dialog box

For a description of the options, see Table 1.

Click **Data** and then a dialog box containing the data of a specific group will appear. In our example, we examine the raw scores of the school with the highest median (see Figure 5). The title bar of Figure 5 tells us the level-2 ID of the box-and-whisker plot we selected is 3427. # is a zero-based counter for group plots.

#	X	Y1	Y2	Y3	Y4	Y5	Y6	Y7
0	13	-0.365	3.085	3.812	5.507	6.734	6.88	7.071

Figure 5   Data for School 3427 dialog box

As the box-and-whisker plots are plotted individually in the example, it is 0. X tells us that the data are from the thirteenth school displayed on the plot. Y1 to Y11 list the mathematics scores for the first eleven students in School 3427. Move the bottom scroll box to the left to display more scores for the other students.

9. (Optional) To edit the graph, open the **Edit** menu and choose **Graph Parameters...**. The user can change attributes such as size and color of the graph, border, and plotting area. By choosing **Copy graph** or **Copy current page** (when there are more than one pages of graphs), users can directly copy and paste the graph or current page into a word processing or graphics document.

10. (Optional) To print the graph, open the **File** menu, select **Print current page** or **Print selected graph** when there are more than one graph. Users can choose **Printing Options...** to change printing parameters such as choice of background, border type, aspect ratio (the ratio of the x-axis length to the y-axis length, the default is 5/3), and printing style.

Table 1 Definitions and options in the Box & Whisker Attributes dialog box

	Key terms	Function	Option	Definition
1	Midpoint	Specify the type of average used	2 choices	1. Median 2. Mode
2	Box Size	Specify the width in units of the axis that the box width is parallel to.		
3	Min, Max, and Coefficient for box or whisker and Constant for box	Min and Max specify the box percentage minimum and maximum when the box or whisker Type is PERCENT. The coefficient is the box or whisker coefficient by which the selected range value will be multiplied. The Constant is the box constant, valid when the box Type is CONSTANT.		
4	Midpoint marker	Display a **Marker Attributes** dialog box that allows the user to specify the shape, color, size, and style of the midpoint marker.		
5	Line attributes	Display a **Line Parameters** dialog box that allows the user to specify the thickness, color, and style of the whisker.		

11. To save the graph for future use by opening the **File** menu and choose **Save as metafile**. A **Save as** dialog box will open. Enter a filename for the file and click **OK**. The file can be saved as an Enhanced Metafile (.EMF) (default and preferred as it holds more information than the other option) or Windows Metafile (.WMF). Users can use word processing programs to insert the graph file into the text. For example, to insert the saved .EMF file into Word, choose **Insert-...Picture-...From File** from Word's main menu.
12. (Optional) To make modifications to the specifications, select **Graph Settings**. The **Equation Graphing** dialog box will appear. We are going to illustrate this by adding a level-2 classification variable next.

To include a level-2 classification variable

13. After choosing the **Y-Axis** variable, select the level-2 classification variable in the **Z-focus** drop-down listbox. There are two types of level-2 classification variables, categorical and continuous. For categorical variables, WHLM will classify the plots with the levels of the variables. For continuous variables, users can choose either to

dichotomize them using median splits, or trichotomize them into three groups: (a) 0 to 24th percentile; (b) 25th to 75th percentile; and (c) 76th percentile and above. These two options, available when a continuous classification variable is chosen, can be found in the lower **Z-focus** drop-down list box. In our example, we will choose school sector, Catholic vs. public school, as the classification variable. To continue working on the plot we have just made, click **Graph Settings** to open the **Equation Graphing** dialog box. Select SECTOR in the **Z-focus** dialog box. The following graph will be displayed (see Figure 6).
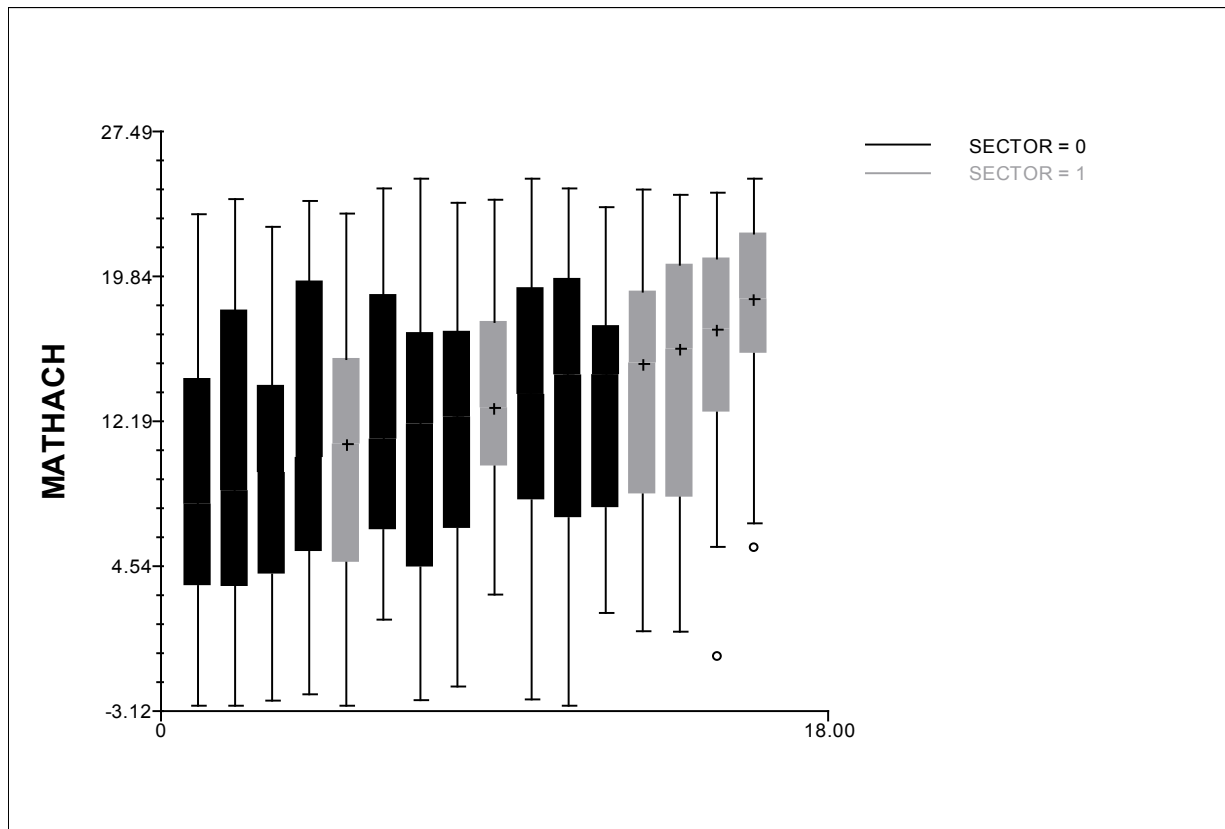

Figure 6   Box-and-whisker plots for MATHACH for a random sample of schools as classified by school sector

In the graph, the box-and-whisker plots for Catholic and public schools are coded differently (red for Catholic and blue for public schools). The colored graphs (not showed here) suggest that the three schools that have the highest median mathematics scores are Catholic schools. The school with the lowest average belongs to the public sector.

Users can edit the legends by clicking on them in the graph above to open the **Legend Parameters** dialog box (see Figure 7), which allows them to make changes in the titles of the legends, their sizes and font types, and the display of the legend box. For example, one may like to change SECTOR = 0 in the text box of Figure 18.7 to PUBLIC = 0 and SECTOR = 1 to CATHOLIC = 1.
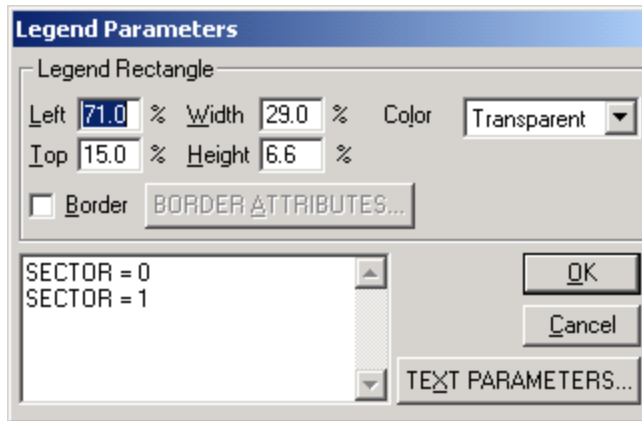
Figure 7   Legend Parameters dialog box