

... BOX PLOT**PURPOSE**

Generates a box plot.

DESCRIPTION

A box plot is a graphical data analysis technique for determining if differences exist between the various levels of a 1-factor model. The box plot is a graphical alternative to 1-factor ANOVA. It is also a useful technique for summarizing and comparing data from 2 or more samples. The box plot consists of:

Vertical axis = response variable;
Horizontal axis = level identification.

The bottom x is the data minimum; the bottom of the box is the estimated 25% point; the middle x in the box is the data median; the top of the box is the estimated 75% point; the top x is the data maximum. The box plot has 24 components (characters and lines) which can be individually controlled. For the box plot to appear as it should, the BOX PLOT command is usually preceded by 2 commands:

CHARACTERS BOX PLOT
LINES BOX PLOT

These commands automatically define the proper values for the 24 components of the box plot. After the box plot is formed, the analyst should redefine plot characters and lines via the usual CHARACTERS and LINES commands.

An alternate form of the box plot, called a mean box plot, is based on means and standard deviations rather than medians and percentiles. The bottom x is the data minimum; the bottom of the box is the mean minus 2 times the standard deviation, the middle x in the box is the mean of the data; the top of the box is the mean plus 2 times the standard deviation; the top x is the data maximum.

For box plots with more than one group, the width of the box plot is proportional the number of elements in the group.

SYNTAX

BOX PLOT <y> <x> <SUBSET/EXCEPT/FOR qualification>

where <y> is a response variable;

<x> is a group identifier variable;

and where the <SUBSET/EXCEPT/FOR qualification> is optional.

EXAMPLES

BOX PLOT Y X
BOX PLOT Y TAG SUBSET TAG > 3

NOTE 1

Outliers can be identified by entering the FENCES ON command. If the inter-quartile range (i.e., the difference between the 25% point and the 75% point) is IQ, then values that are between 1.5 and 3.0 times the IQ above (or below) the 75% point (or the 25%) point are drawn as circles and points that are more than 3.0 times the IQ above (or below) the 75% point (or the 25%) are drawn as large circles.

For mean box plots 4 times the standard deviation (the distance from the top of the box to the bottom of the box in the mean box plot) is used in the above formulas instead of the interquartile range.

NOTE 2

The width of the box is proportional to the number of data points in that box.

NOTE 3

An alternate form of the box plot can be generated by entering the commands CHARACTERS TUFTE BOX PLOT and LINES TUFTE BOX PLOT. You can also define your own plot symbols with the standard CHARACTER and LINE commands (e.g., you may prefer to use a dash (-) rather than the default X).

DEFAULT

None

SYNONYMS

None

RELATED COMMANDS

CHARACTERS	=	Sets the types for plot characters.
LINES	=	Sets the types for plot lines.
I PLOT	=	Generates an I plot.
ANOVA	=	Carries out an ANOVA.
MEDIAN POLISH	=	Carries out a median polish.
CONTROL CHART	=	Generates a control chart.
PLOT	=	Generates a data or function plot.

APPLICATION

Exploratory Data Analysis, Comparing Distributions

REFERENCE

“Graphical Methods for Data Analysis,” Chambers, Cleveland, Kleiner, and Tukey, Wadsworth, 1983.

“Exploratory Data Analysis,” Tukey, Addison-Wesley, 1977.

IMPLEMENTATION DATE

Pre-1987

PROGRAM

```

SKIP 50
SET READ FORMAT 3F4.0,F5.0,F6.0,F3.0,2F9.0
READ PBF11.DAT YEAR DAY BOT SD F11 FLAG WV CO2
RETAIN YEAR DAY BOT SD F11 WV CO2 FLAG SUBSET FLAG 0
LET MONTH=INT(DAY/30.25)+1
LEGEND 1 1-FACTOR MODELING; LEGEND 2 BOX PLOT
CHARACTERS BOX PLOT; LINES BOX PLOT
XLIMITS 0 15; YMINIMUM .995
FENCES ON
BOX PLOT WV MONTH
    
```

