... ROOTOGRAM

PURPOSE

Generates a rootogram.

DESCRIPTION

A rootogram is a graphical data analysis technique for summarizing the distributional information of a variable. It consists of:

Vertical axis = square root of frequencies or relative frequencies;
Horizontal axis = response variable.

There are 4 types of rootograms:

1. rootogram (absolute counts);
2. relative rootogram (converts counts to proportions);
3. cumulative rootogram;
4. cumulative relative rootogram.

The rootogram is a modified version of a histogram. It plots the square roots of the frequencies rather than the raw frequencies. Many univariate data sets can be normalized with a square root transformation (particularly counts or measurement data that have a lower bound and tend to be skewed at the upper tail).

SYNTAX 1

ROOTOGRAM <x> <SUBSET/EXCEPT/FOR qualification>
RELATIVE ROOTOGRAM <x> <SUBSET/EXCEPT/FOR qualification>
CUMULATIVE ROOTOGRAM <x> <SUBSET/EXCEPT/FOR qualification>
CUMULATIVE RELATIVE ROOTOGRAM <x> <SUBSET/EXCEPT/FOR qualification>

where <x> is the variable of raw data values which will appear on the horizontal axis;
and where the <SUBSET/EXCEPT/FOR qualification> is optional.

This syntax is used when you have raw data only.

SYNTAX 2

ROOTOGRAM <y> <x> <SUBSET/EXCEPT/FOR qualification>
RELATIVE ROOTOGRAM <y> <x> <SUBSET/EXCEPT/FOR qualification>
CUMULATIVE ROOTOGRAM <y> <x> <SUBSET/EXCEPT/FOR qualification>
CUMULATIVE RELATIVE ROOTOGRAM <y> <x> <SUBSET/EXCEPT/FOR qualification>

where <y> is the variable of pre-computed frequencies to appear on the vertical axis;
<x> is the variable of raw data values which will appear on the horizontal axis;
and where the <SUBSET/EXCEPT/FOR qualification> is optional.

This syntax is used when you have pre-computed frequencies at each horizontal axis value.

EXAMPLES

```
ROOTOGRAM TEMP
RELATIVE ROOTOGRAM TEMP
CUMULATIVE ROOTOGRAM TEMP
CUMULATIVE RELATIVE ROOTOGRAM TEMP
ROOTOGRAM COUNTS STATE
RELATIVE ROOTOGRAM COUNTS STATE
CUMULATIVE ROOTOGRAM COUNTS STATE
CUMULATIVE RELATIVE ROOTOGRAM COUNTS STATE
```

NOTE 1

The appearance of the bars on the rootogram (i.e., whether they are filled or not, the line width of the bar border, etc.) are controlled by the various bar attribute commands. A few are listed in the RELATED COMMANDS section below. See the documentation for the BAR command for a complete list of the bar attribute commands.

NOTE 2

By default, DATAPLOT uses a class width of 0.3 X the standard deviation of the variable. Use the CLASS WIDTH command to override this default. DATAPLOT also tends to generate a large number of zero frequency classes at the lower and upper tails. This tends to compress the histogram on the horizontal axis. Use the XLIMITS command or the CLASS LOWER and CLASS UPPER commands to avoid plotting these zero frequency classes.

NOTE 3

Although DATAPLOT does not have a FREQUENCY TABLE command, one can be generated with the following commands:

```
HISTOGRAM Y
LET YFREQ = YPLOT
LET XVAL = XPLOT
```

Then the variables YFREQ and XVAL essentially contain a frequency table. There is a LET subcommand called FREQUENCY. However, it does not generate a frequency table in the sense that a histogram or a frequency plot does. The frequency table can also be generated by replacing HISTOGRAM with ROOTOGRAM in the above sequence. However, be aware that this generates the square roots of the frequencies, not the raw frequencies.

DEFAULT

None

SYNONYMS

A synonym for CUMULATIVE RELATIVE ROOTOGRAM is RELATIVE CUMULATIVE ROOTOGRAM

RELATED COMMANDS

HISTOGRAM	=	Generate a histogram.
FREQUENCY PLOT	=	Generates a frequency plot.
PIE CHART	=	Generates a pie chart.
PERCENT POINT PLOT	=	Generates a percent point plot.
PROBABILITY PLOT	=	Generates a probability plot.
PPCC PLOT	=	Generates probability plot correlation coefficient plot.
CLASS LOWER	=	Sets the lower class minimum for histograms, frequency plots, and pie charts.
CLASS UPPER	=	Sets the upper class maximum for histograms, frequency plots, and pie charts.
CLASS WIDTH	=	Sets the class width for histograms, frequency plots, and pie charts.
MINIMUM	=	Sets the frame minima for all plots.
MAXIMUM	=	Sets the frame maxima for all plots.
LIMITS	=	Sets the frame limits for all plots.
PLOT	=	Generates a data or function plot.
BARS	=	Sets the on/off switches for plot bars.
BAR WIDTH	=	Sets the widths for plot bars.
BAR FILL	=	Sets the on/off switches for plot bar fills.
BAR PATTERN	=	Sets the types for bar fill patterns.
BAR BORDER LINE	=	Sets the types for bar border lines.

REFERENCE

Most introductory statistics book discuss frequency polygons and histograms. The rootogram is described in “Exploratory Data Analysis,” John Tukey, Addison-Wesley, 1977.

APPLICATIONS

Exploratory Data Analysis

IMPLEMENTATION DATE

Pre-1987

PROGRAM

```
SET READ FORMAT F10.1
SKIP 25
READ SUNSPOT.DAT Y
MULTIPLOT 2 2
MULTIPLOT CORNER COORDINATES 0 0 100 100
TITLE AUTOMATIC
XLIMITS 0 200; XTIC OFFSET 10 40
MAJOR XTIC MARK NUMBER 6; MINOR XTIC MARK NUMBER 3
ROOTOGRAM Y
BAR FILL ON
RELATIVE ROOTOGRAM Y
BAR FILL OFF
BAR BORDER THICKNESS 0.3
CUMULATIVE ROOTOGRAM Y
BAR FILL ON
BAR PATTERN D1
BAR PATTERN SPACING 3
CUMULATIVE RELATIVE ROOTOGRAM Y
END OF MULTIPLOT
```