

... HISTOGRAM**PURPOSE**

Generates a histogram.

DESCRIPTION

A histogram is a graphical data analysis technique for summarizing the distributional information of a variable. The response variable is divided into equal sized intervals (or bins). The number of occurrences of the response variable is calculated for each bin. The histogram consists of:

Vertical axis = frequencies or relative frequencies;
Horizontal axis = response variable (i.e., the mid-point of each interval).

There are 4 types of histograms:

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

The histogram and the frequency plot have the same information except the histogram has bars at the frequency values, whereas the frequency plot has lines connecting the frequency values.

SYNTAX 1

HISTOGRAM <y> <SUBSET/EXCEPT/FOR qualification>
RELATIVE HISTOGRAM <y> <SUBSET/EXCEPT/FOR qualification>
CUMULATIVE HISTOGRAM <y> <SUBSET/EXCEPT/FOR qualification>
CUMULATIVE RELATIVE HISTOGRAM <y> <SUBSET/EXCEPT/FOR qualification>

where <y> 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

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

HISTOGRAM TEMP
RELATIVE HISTOGRAM TEMP
CUMULATIVE HISTOGRAM TEMP
CUMULATIVE RELATIVE HISTOGRAM TEMP
HISTOGRAM COUNTS STATE
RELATIVE HISTOGRAM COUNTS STATE
CUMULATIVE HISTOGRAM COUNTS STATE
CUMULATIVE RELATIVE HISTOGRAM COUNTS STATE

NOTE 1

The appearance of the bars on the histogram (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. This is demonstrated with the sample program below.

NOTE 2

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.

NOTE 3

By default, DATAPLOT uses a class width of 0.3 times 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.

DEFAULT

None

SYNONYMS

A synonym for CUMULATIVE RELATIVE HISTOGRAM is RELATIVE CUMULATIVE HISTOGRAM

RELATED COMMANDS

FREQUENCY PLOT	=	Generates a frequency plot.
HISTOGRAM	=	Generates a histogram.
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.

“Multivariate Density Estimation,” David Scott, John Wiley, 1992 (chapter 3). This book discusses histograms as “density estimators” and gives optimal criterion for selecting the class width.

APPLICATIONS

Exploratory Data Analysis

IMPLEMENTATION DATE

Pre-1987

PROGRAM

```
SET READ FORMAT F10.1
SKIP 25
READ SUNSPOT.DAT Y
MULTIPLY 2 2
MULTIPLY 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
HISTOGRAM Y
BAR FILL ON
RELATIVE HISTOGRAM Y
BAR FILL OFF
BAR BORDER THICKNESS 0.3
CUMULATIVE HISTOGRAM Y
BAR FILL ON
BAR PATTERN D1
BAR PATTERN SPACING 3
CUMULATIVE RELATIVE HISTOGRAM Y
END OF MULTIPLY
```

