QUANTILE-QUANTILE PLOT

PURPOSE

Generates a quantile-quantile plot.

DESCRIPTION

A quantile-quantile plot (or q-q plot) is a graphical data analysis technique for comparing the distributions of 2 data sets. The quantile-quantile plot is a graphical alternative for the various classical 2-sample tests (e.g., t for location, F for dispersion). The plot consists of the following:

Vertical axis = estimated quantiles from data set 1;
Horizontal axis = estimated quantiles from data set 2.

The "quantiles" of a distribution are the distribution's "percent points" (e.g., the .5 quantile = the 50% point = the median). The advantage of the quantile-quantile plot is 2-fold:

1. the sample sizes do not need to be identical;
2. many distributional aspects can be simultaneously tested. For example, shifts in location, shifts in dispersion, changes in symmetry/skewness, outliers, etc.

The quantile-quantile plot has 2 components:

1. the quantile points themselves;
2. a 45 degree reference line.

The appearance of these 2 components is controlled by the first 2 settings of the CHARACTERS and LINES commands. It is typical for the quantile points to be represented as, say, X's with no connecting line, and the reference line to have no plot characters but to be solid. This is demonstrated in the sample program below.

SYNTAX

QUANTILE-QUANTILE PLOT <y1> <y2> <SUBSET/EXCEPT/FOR qualification>
where <y1> is the first response variable;
<y2> is the second response variable;
and where the <SUBSET/EXCEPT/FOR qualification> is optional.

EXAMPLES

```
QUANTILE-QUANTILE PLOT Y1 Y2
QUANTILE-QUANTILE PLOT RUN1 RUN2
QUANTILE-QUANTILE PLOT Y1 Y2 SUBSET STATE 25
```

NOTE 1

One of the distributions can be a theoretical distribution. For example, the following program generates a quantile-quantile plot of a data set against a normal distribution (this is called a normal quantile plot).

```
LET Y1 = NORMAL RANDOM NUMBERS FOR I = 1 1 100
LET X = SEQUENCE .01 .01 .99
LET Y2 = NORPPF(X)
QUANTILE-QUANTILE PLOT Y1 Y2
```

This same technique can be used for other distributions (use the proper PPF function). This is essentially what a probability plot does (DATAPLOT has a PROBABILITY PLOT command for 38 distributions).

NOTE 2

The Tukey mean-difference (or m-d plot) can be generated after the quantile-quantile plot. It takes the coordinates of the quantile-quantile plot (saved in the DATAPLOT internal variables YPLOT and XPLOT and plots their difference (YPLOT - XPLOT) against their average ((YPLOT+XPLOT)/2) The advantage of this plot is that it converts the interpretation of a quantile-quantile plot to differences from a horizontal (rather than a diagonal) zero line. The program example below generates a quantile-quantile plot and then the corresponding m-d plot. The m-d plot should only be used if the two variables are on a common scale.

DEFAULT

None

SYNONYMS

None

RELATED COMMANDS

CHARACTERS	=	Sets the type for plot characters.
LINES	=	Sets the type for plot lines.
BOX PLOT	=	Generates a box plot.
PLOT	=	Generates a data or function plot.
HISTOGRAM	=	Generates a histogram.
PROBABILITY PLOT	=	Generates a probability plot.
T-TEST	=	Carries out a 2-sample t test.
ANOVA	=	Carries out an ANOVA.
MULTIPLOT	=	Allows multiple plots per page.
LOOP	=	Starts a loop (iteration).

REFERENCE

"Graphical Methods of Data Analysis," Chambers, Cleveland, Kleiner, and Tukey, Wadsworth, 1983 (pp. 48-57).

"Visualizing Data," William Cleveland, Hobart Press, 1993.

APPLICATIONS

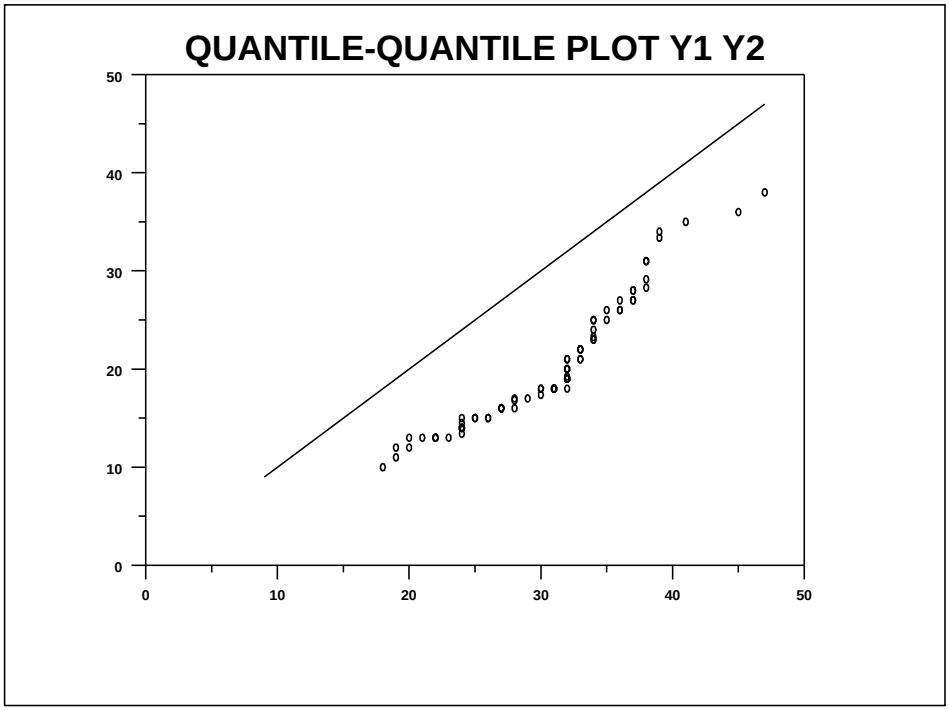
Exploratory Data Analysis

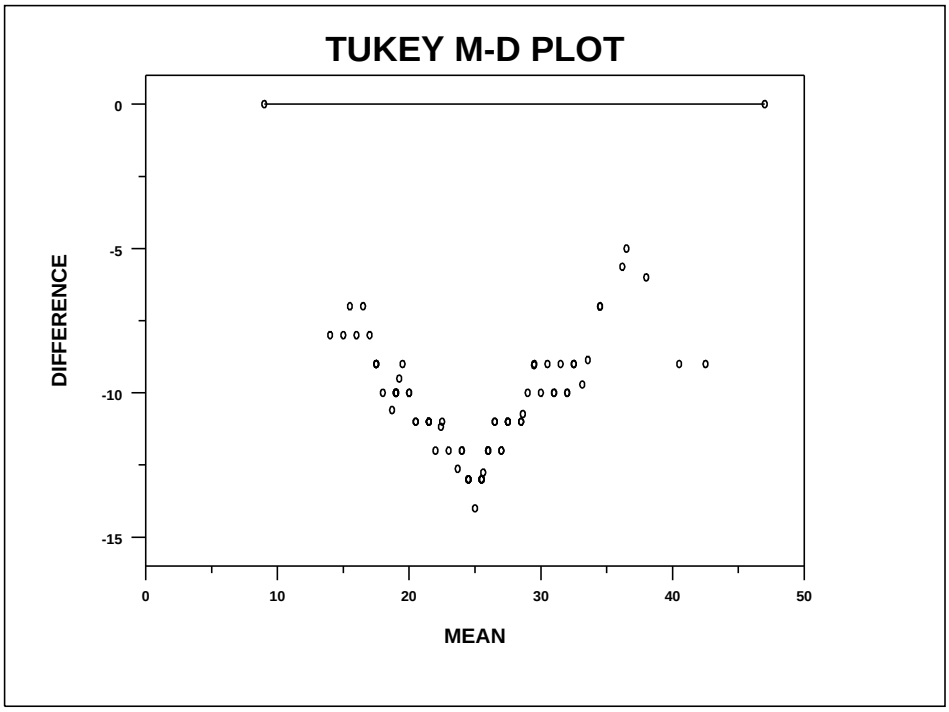
IMPLEMENTATION DATE

88/3

PROGRAM

```
SKIP 25
READ AUTO83B.DAT Y1 Y2
DELETE Y2 SUBSET Y2 < 0
.
LINE BLANK SOLID
CHARACTER CIRCLE BLANK
CHARACTER SIZE 1.0
TITLE AUTOMATIC
QUANTILE-QUANTILE PLOT Y1 Y2
.
LET YMEAN = (YPLOT+XPLOT)/2
LET YDIFF = YPLOT - XPLOT
LET AMIN = MINIMUM YMEAN
LET AMAX = MAXIMUM YMEAN
LET XZERO = DATA AMIN AMAX
LET YZERO = DATA 0 0
.
TITLE TUKEY M-D PLOT
X1LABEL MEAN
Y1LABEL DIFFERENCE
YLIMITS -15 0
YTIC OFFSET 1 1
PLOT YDIFF YMEAN AND
PLOT YZERO XZERO
```


QUANTILE-QUANTILE PLOT Y1 Y2
50
40
30
20
10
0
0
10
20
30
40
50


TUKEY M-D PLOT
0
-5
-10
-15
DIFFERENCE
0
10
20
30
40
50
MEAN