

HOMOSCEDASTICITY PLOT

PURPOSE

Generates a homoscedasticity plot.

DESCRIPTION

A homoscedasticity plot is a graphical data analysis technique for assessing the assumption of constant variance across subsets of the data. The first variable is a response variable and the second variable identifies subsets of the data. The mean and standard deviation are calculated for each of these subsets. The following plot is generated:

Vertical axis = subset standard deviations;
Horizontal axis = subset means.

The interpretation of this plot is that the greater the spread on the vertical axis, the less valid is the assumption of constant variance. A common pattern is for the spread (i.e., the standard deviation) to increase as the location (i.e., the mean) increases. This indicates the need for some type of transformation such as a log or square root.

SYNTAX

HOMOSCEDASTICITY PLOT <y> <tag> <SUBSET/EXCEPT/FOR qualification>

where <y> is a response variable;

<tag> identifies the subsets;

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

EXAMPLES

```
HOMOSCEDASTICITY PLOT Y1 TAG
HOMOSCEDASTICITY PLOT Y1 TAG SUBSET TAG > 2
```

NOTE 1

One limitation of the homoscedasticity plot is that it does not give a convenient way to label the groups on the plot. This can be done by using the SUBSET command as in this example (assume Y is the response variable, X the group-id variable):

```
XILABEL MEANS
YILABEL STANDARD DEVIATIONS
CHARACTER X; LINE BLANK
XLIMITS 0 5; YLIMITS 0 4
CHARACTER NORM; TITLE HOMOSCEDASTICITY PLOT
HOMOSCEDASTICITY PLOT Y X SUBSET X = 1
PRE-ERASE OFF
CHARACTER T
HOMOSCEDASTICITY PLOT Y X SUBSET X = 2
CHARACTER CHIS
HOMOSCEDASTICITY PLOT Y X SUBSET X = 3
CHARACTER UNIF
HOMOSCEDASTICITY PLOT Y X SUBSET X = 4
CHARACTER F
HOMOSCEDASTICITY PLOT Y X SUBSET X = 5
```

NOTE 2

Bartlett's test is an analytic test for the assumption of constant variance. See the documentation for the BARTLET TEST command in the Analysis Commands chapter for more details.

NOTE 3

The spread-location plot (or s-l plot) recommended by Bill Cleveland (see REFERENCE section) is an alternative to the HOMOSCEDASTICITY PLOT. In this plot, group medians are fit to each group and residuals are formed by taking the absolute value of the response variable minus the corresponding median. The square root of these absolute values are plotted against the medians (this is a similar concept to plotting the standard deviations against the means). A line connects the medians of the residuals for each group. Variations of this plot can be obtained by using different fits (e.g., trimmed means instead of medians) and residuals. Program example 2 demonstrates a macro for generating s-l plots in DATAPLOT.

DEFAULT

None

SYNONYMS

HOMOGENITY PLOT

RELATED COMMANDS

- LINES = Sets the type for plot lines.
- HISTOGRAM = Generates a histogram.
- BOX PLOT = Generates a box plot
- PLOT = Generates a data or function plot.
- BARTLETT TEST = Performs a Bartlett test.

REFERENCE

“Visualizing Data,” William Cleveland, Hobart Press, 1993.

APPLICATIONS

Exploratory Data Analysis

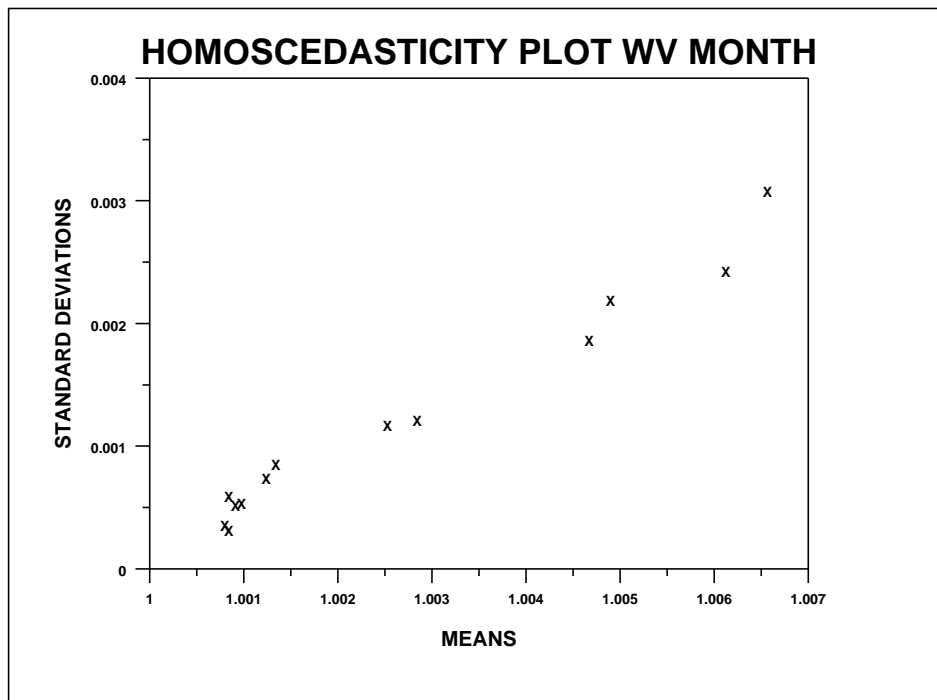
IMPLEMENTATION DATE

Pre-1987

PROGRAM 1

```

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
X1LABEL MEANS; Y1LABEL STANDARD DEVIATIONS
CHARACTER X
LINE BLANK
HOMOSCEDASTICITY PLOT WV MONTH
    
```



PROGRAM 2

```

.PURPOSE--GENERATE A S-L PLOT OF POINT BARROW FREON-11 DATA
DIMENSION 20 VARIABLES
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 TAG=INT(DAY/30.25)+1
LET Y = WV
.
LET N = SIZE Y
LET MED = 0 FOR I = 1 1 N
LET RES = 0 FOR I = 1 1 N
LET TEMP = DISTINCT TAG
LET NGROUP = SIZE TEMP
LOOP FOR K = 1 1 NGROUP
    LET TAGID = TEMP(K)
    LET ATEMP = MEDIAN Y SUBSET TAG = TAGID
    LET MED = ATEMP SUBSET TAG = TAGID
    LET RES = ABS(Y-MED) SUBSET TAG = TAGID
    LET GROUPMD(K) = ATEMP
    LET ATEMP = MEDIAN RES SUBSET TAG = TAGID
    LET MAD(K) = SQRT(ATEMP)
END OF LOOP
LET RES = SQRT(RES)
TITLE SPREAD-LOCATION PLOT
Y1LABEL SQUARE ROOT ABSOLUTE RESIDUAL VW; X1LABEL MEDIAN VW
CHARACTER CIRCLE BLANK; CHARACTER SIZE 1.2; LINE BLANK SOLID
PLOT RES MED AND
PLOT MAD GROUPMD
    
```

