BIWEIGHT

PURPOSE

Carry out a biweight transformation (also called a bisquare transformation).

DESCRIPTION

The biweight transformation is used in robust analysis. For many applications, it combines the properties of resistance with relatively high efficiency. Resistance means that changes in a small part of the data do not cause large changes in the estimate. The mean is an example of a non-resistant estimate while the median is an example of a resistant estimate. Efficiency is a measure of how well the estimate performs for data from a given distribution. For example, the mean is a 100% efficient estimator for normally distributed data. However, it has poor efficiency for heavy tailed distributions. A desirable property for robust estimators is that they maintain high efficiency under a variety of distributions. The biweight transformation of a variable has this property for many applications. See chapter 10 of the Mosteller and Tukey book listed in the REFERENCE section below for a detailed discussion.

The biweight transformation of the variable X is defined as follows:

1. m = the median absolute value of X

2. cutoff = 6 * m

3. $T(X) = (1- (X/cutoff)^2)^2$ if abs(X) < cutoff
$T(X) = 0$ if abs(X) >= cutoff

This transformation defines a set of weights that are used in a subsequent calculation. Data is down-weighted depending on its distance from the median absolute value and extreme outliers are given zero weight.

SYNTAX

LET <v1> = BIWEIGHT <v2> <SUBSET/EXCEPT/FOR qualification>
where <v2> is a variable for which the biweight is to be computed;
<v1> is a variable where the biweight-transformed values are saved;
and where the <SUBSET/EXCEPT/FOR qualification> is optional.

EXAMPLES

LET W = BIWEIGHT RES

NOTE 1

The biweight transformation is most commonly applied to the residuals from a (linear or non-linear) fit. The biweight transformation has the advantage that it allows the analyst to carry out subsequent weighted linear or non-linear fits that are robust and resistive to outliers in the data. The following is a typical sequence using BIWEIGHT:

```
FIT Y = A+B*EXP(-C*X)
LET W = BIWEIGHT RES
WEIGHTS W
FIT Y = A+B*EXP(-C*X)
```

An unweighted fit is performed first. Then the biweight transformation is applied to the residuals. These biweight transformed residuals are used as the weights in a weighted fit.

Robust fits based on the biweight are part of a class of techniques called M-estimators of regression. The fitting technique above is also called iteratively reweighted least squares. See the documentation for the WEIGHTS command (in Volume I) for examples of other weighting functions.

Be aware that this type of robust regression protects against outliers in the dependent variable (i.e., the Y). However, it is still susceptible to outliers in the independent variables. These outliers are called high leverage points. See the documentation for the FIT command (in Volume I) for details on identifying high leverage points (at least for linear fits).

The biweight transformation can be continued for additional iterations of the fitting until some convergence criterion is reached. For example, you can compute the difference between the residuals in two successive steps, sum these differences, and then stop when this sum is below some cutoff value. Although DATAPLOT does not do this automatically, it is straightforward to write a macro to do this.

NOTE 2

Another common use of the biweight is to calculate a robust estimate of location. For example,

```
LET Y = CAUCHY RANDOM NUMBERS FOR I = 1 1 100
LET W = BIWEIGHT Y
LET BMEAN = WEIGHTED MEAN Y W
```

Like the median, this is a resistant estimator. However, the biweight mean is significantly more efficient than the median for both normal and non-normal distributions.

Although there is also a biweight based estimate of scale, it is more complicated than the above sequence. See chapter 10 of the Mosteller and Tukey book listed in the REFERENCE section below for details.

DEFAULT

None

SYNONYMS

BISQUARE for BIWEIGHT

RELATED COMMANDS

FIT	=	Perform a least squares fit.
RES	=	A variable containing the residuals from a FIT.
WEIGHTS	=	Specifies the weights variable for a subsequent fit.
TRIWEIGHT	=	Perform a triweight transformation.

REFERENCE

"Graphical Methods for Data Analysis," Chambers, Cleveland, Kleiner, and Tukey, 1983, Wadsworth (page 122).

"Data Analysis and Regression," Mosteller and Tukey, Addison-Wesley, 1977.

APPLICATIONS

Robust Fitting

IMPLEMENTATION DATE

88/7

PROGRAM

```
LET X = DATA 1 2 3 4 5 6 7 8 9 10
LET Y = DATA 2 4 60 7 9 12 14 15 18 20
FIT Y = A+B*X
LET PRED2 = PRED
CHARACTER CIRCLE BLANK BLANK
LINE BLANK SOLID DASH
SEGMENT 1 COORDINATES 65 85 70 85
SEGMENT 1 PATTERN SOLID
SEGMENT 2 COORDINATES 65 81 70 81
SEGMENT 2 PATTERN DASH
LEGEND 1 ORIGINAL FIT
LEGEND 2 BIWEIGHT FIT
LEGEND 1 COORDIANTES 72 84
LEGEND 2 COORDIANTES 72 80
LET B = BIWEIGHT RES
WEIGHTS B
FIT Y = A+B*X
PLOT Y PRED2 PRED VS X
```