

## MEDIAN POLISH

### PURPOSE

Performs a median polish analysis.

### DESCRIPTION

Median polish is a data analysis technique (more robust than ANOVA) for examining the significance of the various factors in a multi-factor model. The number of factors (= independent variables) must be between 1 and 5, inclusive. Each factor has a certain number of values it can have (these are referred to as the levels of a factor). The number of levels does not have to be the same for each factor. Each factor and level combination is a cell (the number of cells is the product of the number of levels in each factor). Balanced designs are those in which each cell has the same number of observations and unbalanced designs are those in which the number of observations can vary between cells. The MEDIAN POLISH command can work with unbalanced designs (unlike the ANOVA command). The number of arguments specifies whether 1-factor, 2-factor, or higher factor median polish is performed.

Median Polish models the response variable as:

$$\text{data} = \text{common value} + \text{factor-1 effects} + \dots + \text{factor-n effects} + \text{residual}$$

More complex models can also include interaction terms. However, the DATAPLOT MEDIAN POLISH command does not compute interaction terms.

### SYNTAX

MEDIAN POLISH <y> <x1> ... <x5> <SUBSET/EXCEPT/FOR qualification>

where <y> is the response (= dependent) variable;

<x1> ... <x5> is a list of at least 1 and no more than 5 independent variables;

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

### EXAMPLES

MEDIAN POLISH Y X1

MEDIAN POLISH Y X1 X2

MEDIAN POLISH Y X1 X2 SUBSET X2 > 3

### NOTE 1

The following is the basic algorithm for median polish.

For each factor, do the following:

1. Calculate the medians for each level of a factor (these are the factor effects).
2. For each value of the response variable, subtract the corresponding level median.

The above steps are repeated until the ratio of the sum of the residuals from the current step and the residuals from the previous step are less than some cut-off value (normally only 2 or 3 passes are required).

### NOTE 2

The factor effects for a given factor are the sum over all iterations of the medians computed for each level. The predicted value for a given response value is the sum of the factor effects for the corresponding levels of each factor. The residuals are the response variable minus the predicted value. The residuals and predicted values are stored in the internal variables RES and PRED respectively. These variables can be used in subsequent LET and PLOT commands for additional analysis.

### NOTE 3

The common term is the grand median.

### NOTE 4

The MEDIAN POLISH command can work with raw data or with summary value for each cell (e.g., the mean, median, or some other summary measure).

### DEFAULT

None

### SYNONYMS

None

## RELATED COMMANDS

|                  |   |                                       |
|------------------|---|---------------------------------------|
| ANOVA            | = | Carries out an ANOVA.                 |
| YATES ANALYSIS   | = | Analyze a Yates design.               |
| BLOCK PLOT       | = | Generate a block plot.                |
| DEX SCATTER PLOT | = | Generates a dex scatter plot.         |
| DEX ... PLOT     | = | Generates a dex plot for a statistic. |
| T TEST           | = | Carries out a t test.                 |
| PLOT             | = | Plots (e.g., residuals and GANOVA).   |

## REFERENCE

“Data Analysis and Regression,” Mosteller and Tukey, Addison-Wesley, 1977 (chapter 9).

## APPLICATIONS

Analysis of Variance

## IMPLEMENTATION DATE

Pre-1987

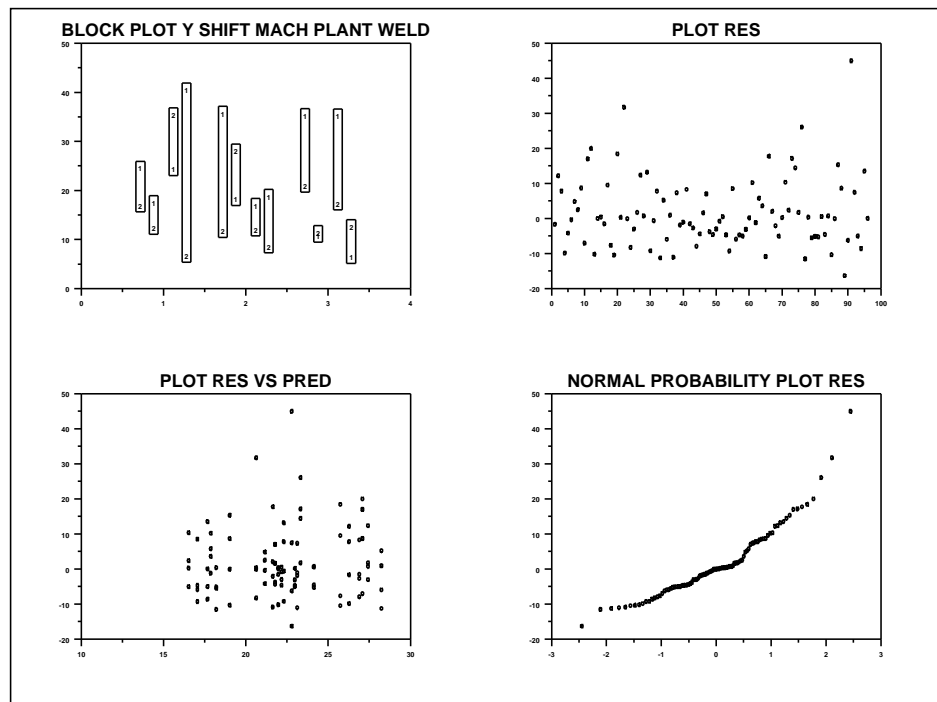
## PROGRAM

```

. READ FILE "SHEESLEY.DAT" IN DATAPLOT REFERENCE CATALOG
. THIS IS DATAPLOT DATA FILE SHEESLEY.DAT (RAW DATA)
. LIGHT BULB LEAD WIRE WELD PROCESS COMPARISON JOHN SHEESLEY (GE) ARTICLE IN
. EXPERIMENTS IN INDUSTRY (ED. BY SNEE, HARE, TROUT) PAGES 54-57
. NUMBER OF OBSERVATIONS = 96
. ORDER OF VARIABLES ON A LINE IMAGE--
. RESPONSE = AVERAGE NUMBER OF WELDED LEAD WIRES MISSED PER HOUR
. FACTOR 1 = WELDING PROCESS (2 LEVELS) (PRIMARY)
. FACTOR 2 = SHIFT (3 LEVELS)
. FACTOR 3 = MACHINE (2 LEVELS)
. FACTOR 4 = PLANT (2 LEVELS)
. FACTOR 5 = REPLICATION (4 LEVELS) (A RANDOM FACTOR)
SKIP 25
READ SHEESLEY.DAT Y WELD SHIFT MACH PLANT REP
MEDIAN POLISH Y WELD SHIFT MACH PLANT
MULTIPLY 2 2; MULTIPLY CORNER COORDINATES 0 0 100 100
TITLE AUTOMATIC
CHARACTER 1 2; LINES BLANK BLANK
BLOCK PLOT Y SHIFT MACH PLANT WELD
CHARACTER CIRCLE; CHARACTER SIZE 1.0; LINES BLANK
PLOT RES
PLOT RES VS PRED
NORMAL PROBABILITY PLOT RES
END OF MULTIPLY

```

The following output is generated.



```
*****
*****
** 4-WAY MEDIAN POLISH **
*****
*****
```

```
NUMBER OF OBSERVATIONS      =      96
NUMBER OF FACTORS           =        4
NUMBER OF LEVELS FOR FACTOR 1 =        2
NUMBER OF LEVELS FOR FACTOR 2 =        3
NUMBER OF LEVELS FOR FACTOR 3 =        2
NUMBER OF LEVELS FOR FACTOR 4 =        2
RESIDUAL STANDARD DEVIATION = 0.10319581985E+02
RESIDUAL DEGREES OF FREEDOM =        90
REPLICATION CASE
REPLICATION STANDARD DEVIATION = 0.10128160477E+02
REPLICATION DEGREES OF FREEDOM =        72
NUMBER OF DISTINCT CELLS    =        24
```

```
*****
* ESTIMATION *
*****
```

```
GRAND MEAN                  = 0.23897912979E+02
GRAND MEDIAN                 = 0.21850000381E+02
MEDIAN POLISH TYPICAL VALUE = 0.23143751144E+02
GRAND RANGE                  = 0.61300003052E+02
GRAND STANDARD DEVIATION     = 0.10687572479E+02
```

|            | LEVEL-ID | NI  | TYP. VALUE | EFFECT   |
|------------|----------|-----|------------|----------|
| FACTOR 1-- | 1.00000  | 48. | 25.70000   | 2.55625  |
| --         | 2.00000  | 48. | 20.58750   | -2.55625 |
| FACTOR 2-- | 1.00000  | 32. | 22.55000   | -0.59375 |
| --         | 2.00000  | 32. | 23.36250   | 0.21875  |
| --         | 3.00000  | 32. | 22.01875   | -1.12500 |
| FACTOR 3-- | 1.00000  | 48. | 22.63438   | -0.50937 |
| --         | 2.00000  | 48. | 23.78438   | 0.64062  |
| FACTOR 4-- | 1.00000  | 48. | 24.81875   | 1.67500  |
| --         | 2.00000  | 48. | 20.71875   | -2.42500 |

| MODEL                       | RESIDUAL STANDARD DEVIATION |
|-----------------------------|-----------------------------|
| CONSTANT ONLY--             | 10.6875724792               |
| CONSTANT & FACTOR 1 ONLY--  | 10.2089872360               |
| CONSTANT & FACTOR 2 ONLY--  | 10.8986415863               |
| CONSTANT & FACTOR 3 ONLY--  | 10.7506313324               |
| CONSTANT & FACTOR 4 ONLY--  | 10.6038141251               |
| CONSTANT & ALL 4 FACTORS -- | 10.3195819855               |

```
*****  
* TESTING *  
*****
```

|        | NUM. | LEVELS | F STAT.       | F CDF   |
|--------|------|--------|---------------|---------|
| FACTOR | 1--  | 2      | 5.89052057266 | 98.279% |
| FACTOR | 2--  | 3      | 0.25030812621 | 22.090% |
| FACTOR | 3--  | 2      | 0.30192747712 | 41.596% |
| FACTOR | 4--  | 2      | 3.91515636444 | 94.909% |

```
RESIDUAL STANDARD DEVIATION = 10.31958198547  
RESIDUAL DEGREES OF FREEDOM = 90  
REPLICATION STANDARD DEVIATION = 10.12816047668  
REPLICATION DEGREES OF FREEDOM = 72  
LACK OF FIT F RATIO = 1.1908 = THE 70.8226% POINT OF THE  
F DISTRIBUTION WITH 18 AND 72 DEGREES OF FREEDOM
```