

DEX PHD**PURPOSE**

Carry out a pHd (principal Hessian direction) analysis to generate the ordered pHd vectors (eigenvectors of the matrix of 2-term interactions) for 2-level full or fractional factorial (i.e., Yates) designs.

DESCRIPTION

This command starts by performing the standard Yates analysis (see the documentation for the YATES command for details). The Yates analysis yields factor estimates for all of the coefficients (main effects and all relevant interactions) in 2-level full and factorial experiments. The name "Yates analysis" comes from the pioneering work which Yates carried out in the analysis of these designs. Although the Yates algorithm (see page 342 of the Box, Hunter, and Hunter listed in the Reference section below), a manual iterative procedure for generating estimated factor effects, is a bit dated, we carry on the name in the form of the command.

The data in the input variable Y is assumed to have been collected in Yates order. The primary output from the Yates analysis is a list which consists of 5 columns:

1. A factor identifier (all from Yates order):
 - 1 = factor 1
 - 2 = factor 2
 - 3 = factor 3
 - 12 = factor 1 x factor 2 interaction
 - 13 = factor 1 x factor 3 interaction
 - 23 = factor 2 x factor 3 interaction
 - 123 = factor 1 x factor 2 x factor 3 interaction
 - etc.
2. Least squares estimated factor effects ordered from largest in magnitude (most significant) to smallest in magnitude (least significant).
3. A t value ($= e / \text{sd}(e)$) where
 - e is the estimated factor effect;
 - sd(e) is the estimated standard deviation of e.
 Adjacent to this t value is an indication (* or **) as to whether the estimated effects are statistically significant at the 5% level (*) or 1% (**) level.
4. The residual standard deviation that results from the model:
 - response = constant + 1/2 (that single factor only)
 This is of interest but not that important.
5. The (cumulative) residual standard deviation that results from the model:
 - response=constant + 1/2 (all factors down to and including the factor of interest).
 This column will consist of a monotonically decreasing set of residual standard deviations (indicating a better and better fit as the number of terms included in the model increases).

In addition to the Yates analysis, the DEX PHD analysis generates 3 outputs on the screen:

1. The matrix of estimated 2-term interactions;
2. The ordered (by magnitude) eigenvalues of the matrix;
3. The corresponding eigenvectors of the matrix (these eigenvectors are the pHd's of the system).

The matrix of estimated 2-term interactions is also written to the file DPST1F.DAT. The eigenvalues are written to the file DPST2F.DAT. The first column contains the unordered eigenvalues, column 2 contains the identifier for the ordered eigenvalues, and column 3 contains the ordered (by absolute value) eigenvalues. The eigenvectors are written to the file DPST3F.DAT. The predicted values, the residual values, and the first 5 pHd vectors (only the first 2 are printed to the screen) are written to the file DPST4F.DAT. These files are automatically created by this command. The name of the files may vary somewhat depending on the operating system (e.g., they are typically lower case on Unix implementations).

The contents of these files can be viewed via the LIST command (where <name> is DPST1F.DAT, DPST2F.DAT, DPST3F.DAT, or DPST4F.DAT) or read back in via the READ command:

```
LIST <name>
READ MATRIX DPST1F.DAT V
```

```

READ DPST2F.DAT EIGVAL ID SORTEIG
READ MATRIX DPST3F.DAT EIGVEC
READ DPST4F.DAT PRED2 RES2 PHD1 PHD2 PHD3 PHD4 PHD5

```

The DEX PHD technique is useful for Linear Domain Splitting (LDS). LDS is the partitioning of a 2-level design with a complicated 2-term interaction structure into sub-regions which have simpler 2-term interaction structures. The pHd's indicate as to which of the variables (in full factorial designs) should be used as a splitting factor (see the Filliben and Li article listed in the Reference section below for details).

A typical follow-up graphical analysis to determine the proper splitting factor is to use the pHd's in the ASCII file DPST4F.DAT is as follows:

```

READ DPST4F.DAT PRED2 RES2 PHD1 PHD2
CHARACTERS - + BLANK; LINES BLANK BLANK SOLID
MULTILOT 2 2
PLOT PHD1 PHD2 X1
PLOT PHD1 PHD2 X2
PLOT PHD1 PHD2 X3
END OF MULTILOT

```

The first file (DPST1F.DAT) containing the 2-terms effects matrix, the second file (DPST2F.DAT) containing the eigenvalues, and the third file (DPST3F.DAT) containing the eigenvectors are of interest but are less frequently used.

SYNTAX 1

```

DEX PHD <y1> <SUBSET/EXCEPT/FOR qualification>

```

where <y1> is a response variable in standard Yates form;
and where the <SUBSET/EXCEPT/FOR qualification> is optional.

This syntax is used for the no replication case.

SYNTAX 2

```

DEX PHD <y1> <group> <SUBSET/EXCEPT/FOR qualification>

```

where <y1> is a response variable in standard Yates form;
<group> is a variable that identifies the replication;
and where the <SUBSET/EXCEPT/FOR qualification> is optional.

This syntax is used for the replication case.

EXAMPLES

```

DEX PHD Y
DEX PHD Y GROUP

```

NOTE 1

This code assumes the data is in the standard Box/Hunter/Yates order. For example, for a 2**3 design:

```

- - -
+ - -
- + -
+ + -
- - +
+ - +
- + +
+ + +

```

More specifically, for full factorial designs, the kth column, for example, consists of 2**(k-1) minus signs followed by 2**(k-1) plus signs. For fractional factorial designs, see the schema on page 410 of the Box, Hunter, and Hunter text listed in the REFERENCE section.

NOTE 2

Since the DEX PHD command has only one argument, the command itself does not know whether it is analyzing a full factorial design or some fractional factorial design with the same number of runs. In such case, the analyst must make use of the confounding structure of 2-level factorial designs to translate the factor tag to its various aliases. The simplest case of this is when the response variable has n = 4 observations. In such case, the DEX PHD command would output factor tags on the "assumption" that it is analyzing a 2**2 full

factorial design. Such tags would thus be 1, 2, and 12. If the design was in fact a 2^{3-1} design, then the analyst must augment the 1, 2, and 12 tags by their corresponding aliases-- in particular for a 2^{3-1} :

```
1   == 23
2   == 13
12  == 3
```

For on-line information about confounding structure, the analyst is referred (in this case) to the header lines of the ASCII file 2TO3M1.DAT via:

```
LIST 2TO3M1.DAT
```

and similar such design files. To get a list of all such design files, enter:

```
LIST DESIGNS.
```

NOTE 3

If there is replication, then the replicates can either be immediately within or they can be in the blocks after.

NOTE 4

The predicted values and residual values are NOT stored in the internal variables PRED and RES respectively. However, they are written to the file DPST4F.DAT. This is demonstrated in the PROGRAM examples. Several values from the Yates analysis output are saved as internal parameters. These are:

```
PRESSD - the psuedo-replication standard deviation;
PRESDF - the psuedo-replication degrees of freedom;
REPSD  - the replication standard deviation;
REPDF  - the replication degrees of freedom;
REFSD  - the psuedo-replication standard deviation (no replication case) or the replication standard deviation (replication case);
REFDF  - the psuedo-replication degrees of freedom (no replication case) or the replication degrees of freedom (replication case);
SDCOEF - standard deviation of a coefficient.
```

NOTE 5

The YATES <COEF/T/RESSD> CUTOFF and the YATES OUTPUT commands are used to control what output is printed by the Yates analysis portion of the DEX PHD command. See the documentation for the YATES CUTOFF and YATES OUTPUT commands for more details.

NOTE 6

The pHd method has broader application than the analysis of 2-level designs. However, the current DATAPLOT implementation is restricted to this special case. See the Li article listed in the REFERENCE section for further details.

SYNONYMS

DEX PHD FIT, DEX PHD ANALYSIS, PHD DEX, PHD DEX FIT, and PHD DEX ANALYSIS are synonyms for DEX PHD.

RELATED COMMANDS

```
YATES      = Carries out a Yates analysis.
FIT        = Carries out a least squares fit.
PLOT       = Generates a data or function plot.
DEX SCATTER PLOT = Generates a dex scatter plot.
DEX SIGN PLOT = Generates a dex sign plot.
DEX ... PLOT = Generates a dex plot for a statistic.
DEX ... PARETO PLOT = Generates a Pareto dex plot for a statistic.
DEX ... YOUDEN PLOT = Generates a Youden dex plot for a statistic.
DEX ... EFFECTS PLOT = Generates a dex effects plot for a statistic.
DEX ... PARETO EFFECTS PLOT = Generates a Pareto effects dex plot for a statistic.
DEX ... ABSOLUTE EFFECTS PLOT = Generates an absolute effects dex plot for a statistic.
DEX ... PARE ABSO EFFECTS PLOT = Generates a Pareto absolute effects dex plot for a statistic.
```

REFERENCES

"Statistics For Experimenters," Box, Hunter, and Hunter, Wiley and Sons, 1978.

“pHd-Guided Linear Domain Splitting for 2-Level Factorial Designs,” James J. Filliben and Ker-Chau Li, to appear in Technometrics 1995.

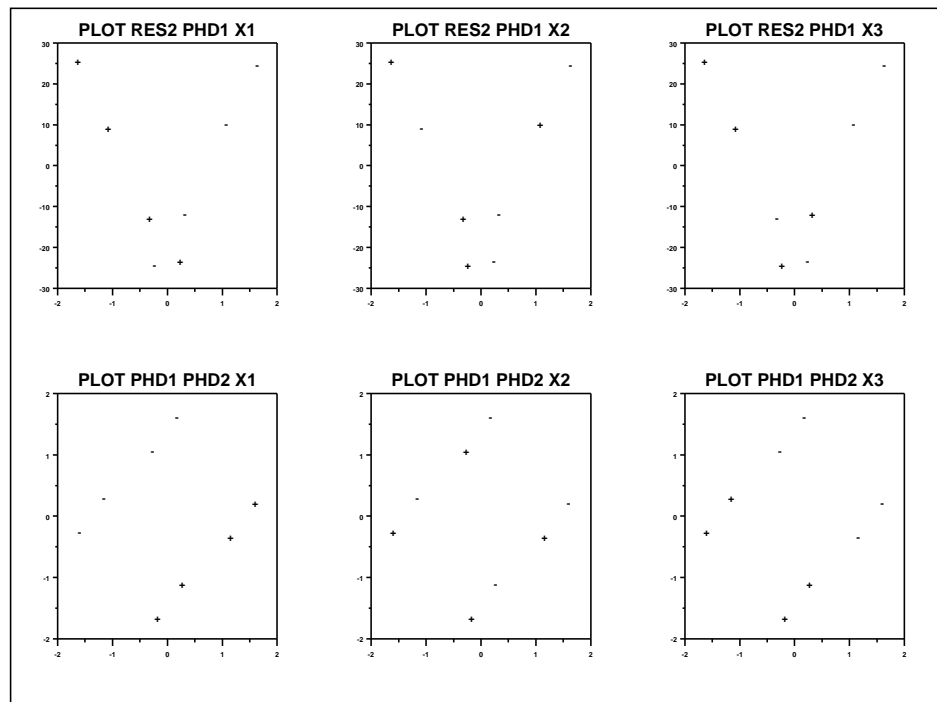
“On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein’s Lemma,” Ker-Chau Li, Journal of the American Statistical Association, 1992 (pp. 1025-1039).

IMPLEMENTATION DATE

93/10 (output to files DPST1F.DAT, DPST2F.DAT, DPST3F.DAT, and DPST4F.DAT activated 95/2)

PROGRAM 1

```
. THIS IS AN EXAMPLE OF A DEX PHD ANALYSIS
. OF A 2**3 FULL FACTORIAL DESIGN.
SKIP 25
READ NUT.DAT Y X1 X2 X3
SKIP 0
DEX PHD Y
READ DPST4F.DAT PRED2 RES2 PHD1 PHD2
CHARACTERS - +
CHARACTER SIZE 3 3
LINES BLANK BLANK
MULTIPLY 2 3; MULTIPLY CORNER COORDINATES 0 0 100 100
TITLE AUTOMATIC
PLOT RES2 PHD1 X1
PLOT RES2 PHD1 X2
PLOT RES2 PHD1 X3
PLOT PHD1 PHD2 X1
PLOT PHD1 PHD2 X2
PLOT PHD1 PHD2 X3
END OF MULTIPLY
```



The following alphanumeric output is generated.

```

*****
**          2**K DEX FIT          **
*****

(NOTE--DATA MUST BE IN STANDARD ORDER)
NUMBER OF OBSERVATIONS          =          8
NUMBER OF FACTORS                =          3
NO REPLICATION CASE

PSEUDO-REPLICATION STAND. DEV.  =  0.14142135382E+01
PSEUDO-DEGREES OF FREEDOM       =          1
(THE PSEUDO-REP. STAND. DEV. ASSUMES ALL
 3, 4, 5, ...-TERM INTERACTIONS ARE NOT REAL,
 BUT MANIFESTATIONS OF RANDOM ERROR)

STANDARD DEVIATION OF A COEF.   =  0.10000000000E+01
(BASED ON PSEUDO-REP. ST. DEV.)

GRAND MEAN                      =  0.10150000000E+03
GRAND STANDARD DEVIATION        =  0.20798351288E+02

99% CONFIDENCE LIMITS (+-)      =  0.63656803131E+02
95% CONFIDENCE LIMITS (+-)      =  0.12706216812E+02
99.5% POINT OF T DISTRIBUTION    =  0.63656803131E+02
97.5% POINT OF T DISTRIBUTION    =  0.12706216812E+02
    
```

IDENTIFIER	EFFECT	T VALUE	RESSD MEAN+TERM	RESSD MEAN+TERMS
MEAN	101.50000		20.79835	20.79835
13	35.50000	35.5*	9.19692	9.19692
12	13.50000	13.5*	21.06933	5.34790
3	-7.50000	-7.5	22.04352	2.76134
2	2.50000	2.5	22.41837	2.44949
1	-2.00000	-2.0	22.43509	2.23607
23	2.00000	2.0	22.43509	1.41421
123	1.00000	1.0	22.45737	0.00000

```

I, ITAGCO(I), COEF(I) =      1      0  101.500
I, ITAGCO(I), COEF(I) =      2      1   -2.000
I, ITAGCO(I), COEF(I) =      3      2    2.500
I, ITAGCO(I), COEF(I) =      4     12   13.500
I, ITAGCO(I), COEF(I) =      5      3   -7.500
I, ITAGCO(I), COEF(I) =      6     13   35.500
I, ITAGCO(I), COEF(I) =      7     23    2.000
I, ITAGCO(I), COEF(I) =      8    123    1.000
    0.000    13.500    35.500
    13.500     0.000     2.000
    35.500     2.000     0.000
    
```

-37.352

-1.327
38.679

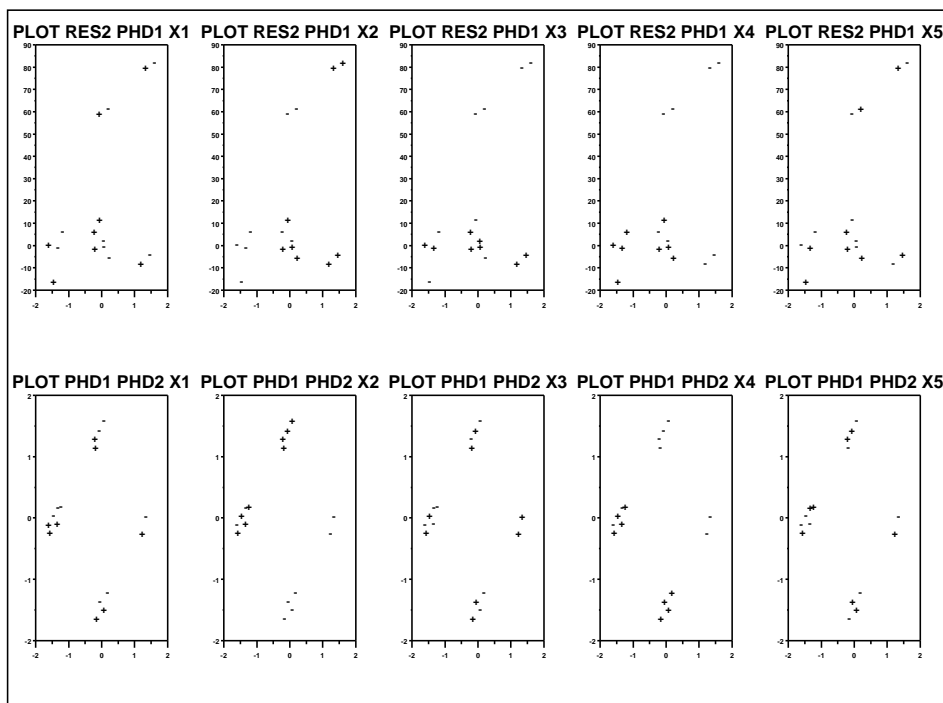
0.713 -0.039 0.700
-0.222 0.934 0.278
-0.665 -0.354 0.657

I, ITAG3(I), EIGVA3(I) = 1 3 38.679
I, ITAG3(I), EIGVA3(I) = 2 1 37.352
I, ITAG3(I), EIGVA3(I) = 3 2 1.327

I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	1	130.000	105.000	25.000	-
1.636 0.175					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	2	80.000	103.000	-23.000	-
0.235 1.600					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	3	118.000	107.500	10.500	-
1.079 -0.269					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	4	93.000	105.500	-12.500	
0.322 1.156					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	5	86.000	97.500	-11.500	-
0.322 -1.156					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	6	105.000	95.500	9.500	
1.079 0.269					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	7	76.000	100.000	-24.000	
0.235 -1.600					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	8	124.000	98.000	26.000	
1.636 -0.175					

PROGRAM 2

```
. THIS IS AN EXAMPLE OF A DEX PHD ANALYSIS  
. OF A 2**5 FULL FACTORIAL DESIGN.  
SKIP 25  
READ BOXREAC2.DAT Y X1 X2 X3 X4 X5  
SKIP 0  
DEX PHD Y  
READ DPST4F.DAT PRED2 RES2 PHD1 PHD2  
CHARACTERS - +  
CHARACTER SIZE 3 3  
LINES BLANK BLANK  
MULTILOT 2 5; MULTILOT CORNER COORDINATES 0 0 100 100  
TITLE AUTOMATIC  
PLOT RES2 PHD1 X1  
PLOT RES2 PHD1 X2  
PLOT RES2 PHD1 X3  
PLOT RES2 PHD1 X4  
PLOT RES2 PHD1 X5  
PLOT PHD1 PHD2 X1  
PLOT PHD1 PHD2 X2  
PLOT PHD1 PHD2 X3  
PLOT PHD1 PHD2 X4  
PLOT PHD1 PHD2 X5  
END OF MULTILOT
```



The following output is generated.

```

*****
**          2**K DEX FIT          **
*****

(NOTE--DATA MUST BE IN STANDARD ORDER)
NUMBER OF OBSERVATIONS      =      16
NUMBER OF FACTORS          =      4
NO REPLICATION CASE

PSEUDO-REPLICATION STAND. DEV. = 0.10488088608E+02
PSEUDO-DEGREES OF FREEDOM    =      5
(THE PSEUDO-REP. STAND. DEV. ASSUMES ALL
3, 4, 5, ...-TERM INTERACTIONS ARE NOT REAL,
BUT MANIFESTATIONS OF RANDOM ERROR)

STANDARD DEVIATION OF A COEF. = 0.52440443039E+01
(BASED ON PSEUDO-REP. ST. DEV.)

GRAND MEAN                  = 0.65250000000E+02
GRAND STANDARD DEVIATION    = 0.14901901245E+02

99% CONFIDENCE LIMITS (+-) = 0.21144742966E+02
95% CONFIDENCE LIMITS (+-) = 0.13480249405E+02
99.5% POINT OF T DISTRIBUTION = 0.40321440697E+01
97.5% POINT OF T DISTRIBUTION = 0.25705826283E+01
    
```

IDENTIFIER	EFFECT	T VALUE	RESSD MEAN+TERM	RESSD MEAN+TERMS
MEAN	65.25000		14.90190	14.90190
2	20.50000	3.9*	10.85620	10.85620
4	12.25000	2.3	13.96616	8.98610
24	10.75000	2.0	14.31470	6.99702
123	-9.50000	-1.8	14.56512	4.53772
1234	-6.25000	-1.2	15.05881	2.65047
124	2.25000	0.4	15.37798	2.35702
1	-2.00000	-0.4	15.38784	2.06155
23	1.50000	0.3	15.40408	1.88982
12	1.50000	0.3	15.40408	1.63299
134	1.25000	0.2	15.41046	1.39642
234	1.25000	0.2	15.41046	0.93541
14	-0.75000	-0.1	15.41972	0.64550
13	0.50000	0.1	15.42262	0.35355
34	0.25000	0.0	15.42435	0.00000
3	0.00000	0.0	15.42493	0.00000

```

I, ITAGCO(I), COEF(I) = 1 0 65.250
I, ITAGCO(I), COEF(I) = 2 1 -2.000
I, ITAGCO(I), COEF(I) = 3 2 20.500
I, ITAGCO(I), COEF(I) = 4 12 1.500
I, ITAGCO(I), COEF(I) = 5 3 0.000
I, ITAGCO(I), COEF(I) = 6 13 0.500
    
```



```

I,ITAGCO(I),COEF(I) =      7      23      1.500
I,ITAGCO(I),COEF(I) =      8     123     -9.500
I,ITAGCO(I),COEF(I) =      9       4     12.250
I,ITAGCO(I),COEF(I) =     10      14     -0.750
I,ITAGCO(I),COEF(I) =     11      24     10.750
I,ITAGCO(I),COEF(I) =     12     124      2.250
I,ITAGCO(I),COEF(I) =     13      34      0.250
I,ITAGCO(I),COEF(I) =     14     134      1.250
I,ITAGCO(I),COEF(I) =     15     234      1.250
I,ITAGCO(I),COEF(I) =     16    1234     -6.250
      0.000      1.500      0.500     -0.750
      1.500      0.000      1.500     10.750
      0.500      1.500      0.000      0.250
     -0.750     10.750      0.250      0.000

-11.040
  0.632
 -0.515
 10.923

      0.139      0.773     -0.617      0.055
     -0.704      0.029     -0.060      0.707
      0.074      0.605      0.785      0.116
      0.693     -0.191     -0.021      0.695

I,ITAG3(I),EIGVA3(I) =      1      1     11.040
I,ITAG3(I),EIGVA3(I) =      2      4     10.923
I,ITAG3(I),EIGVA3(I) =      3      2      0.632
I,ITAG3(I),EIGVA3(I) =      4      3      0.515

I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      1     56.000     41.000     6.125     -
0.202      1.331
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      2     53.000     131.875     5.125
0.076      1.609
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      3     63.000     229.625    -7.375     -
1.609     -0.076
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      4     65.000     470.875    -3.375     -
1.331      0.202
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      5     53.000     49.875      3.125     -
0.055     -1.342
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      6     55.000     47.875      7.125
0.224     -1.232
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      7     67.000     70.375    -3.375     -
1.462      0.073
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      8     61.000     68.375    -7.375     -
1.184      0.183
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =      9     69.000     62.125      6.875
1.184     -0.183
I,Y(I),PREDLIN(I),RESLIN(I),PHD1(I),PHD2(I) =     10     45.000     60.125    -15.125
1.462     -0.073

```

I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	11	78.000	82.625	-4.625	-
0.224 1.232					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	12	93.000	80.625	12.375	
0.055 1.342					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	13	49.000	0.139	11.040	
1.331 0.049					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	14	60.000	-0.704	0.632	
1.609 0.158					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	15	95.000	0.074	0.515	-
0.076 1.463					
I, Y(I), PREDLIN(I), RESLIN(I), PHD1(I), PHD2(I) =	16	82.000	0.693	10.923	
0.202 1.573					