

YATES ANALYSIS

PURPOSE

Perform a Yates analysis of a 2-level full (2^{**K}) or fractional ($2^{**}(K-P)$) factorial experiment design.

DESCRIPTION

The Yates analysis yields factor estimates for all of the coefficients (main effects and all relevant interactions) in 2-level full and fractional 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 Box, Hunter, and Hunter, page 342 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 are assumed to have been collected in Yates order (this is described in a NOTE section below). 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 ($= e / sd(e)$) 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).

The 5 columns of this ordered list are written out (no header lines) to the ASCII files DPST1F.DAT and DPST2F.DAT (the name may vary on some operating systems). The file DPST1F.DAT excludes the first line (the estimated mean) and so includes only those estimates of the factors and interactions. The DPST1F.DAT is the one that is used for further analysis (e.g., a normal probability plot of the effects). Such a further analysis is done as follows:

```
READ DPST1F.DAT TAG COEF
NORMAL PROBABILITY PLOT COEF
LET COEF2 = ABS(COEF)
HALFNORMAL PROBABILITY PLOT COEF2
```

The DPST2F.DAT file has as its first line the mean information, and so is a verbatim copy of the ordered list. Although of interest, this second file is used less frequently.

SYNTAX 1

YATES ANALYSIS <y1> <SUBSET/EXCEPT/FOR qualification>

where <y1> is a response variable;

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

This syntax is used for the no replication case.

SYNTAX 2

YATES ANALYSIS <y1> <group> <SUBSET/EXCEPT/FOR qualification>
 where <y1> is a response variable;
 <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

```
YATES ANALYSIS Y
YATES ANALYSIS Y GROUP
YATES ANALYSIS Y GROUP SUBSET GROUP > 1
```

NOTE 1

This code assumes the data is in the standard Yates/Hunter/Box 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 Yates 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 YATES 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

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);
 SDCOEFF -standard deviation of a coefficient.

NOTE 5

The YATES <COEF/T/RESSD> CUTOFF command is used to control convergence. The YATES OUTPUT command is used to control what output is printed by the YATES ANALYSIS command.

DEFAULT

None

SYNONYMS

YATES, DEX FIT, 2**K FIT are all synonyms for YATES ANALYSIS.

RELATED COMMANDS

FIT	=	Carries out a least squares fit.
DEX PHD	=	Carry out a principal Hessian direction analysis of a Yates design.
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 ... YOUNDEN 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.

REFERENCE

“Statistics for Experimenters,” Box, Hunter, and Hunter. John Wiley and Sons, 1978 (chapter 10, section 7).

APPLICATIONS

Design of Experiments

IMPLEMENTATION DATE

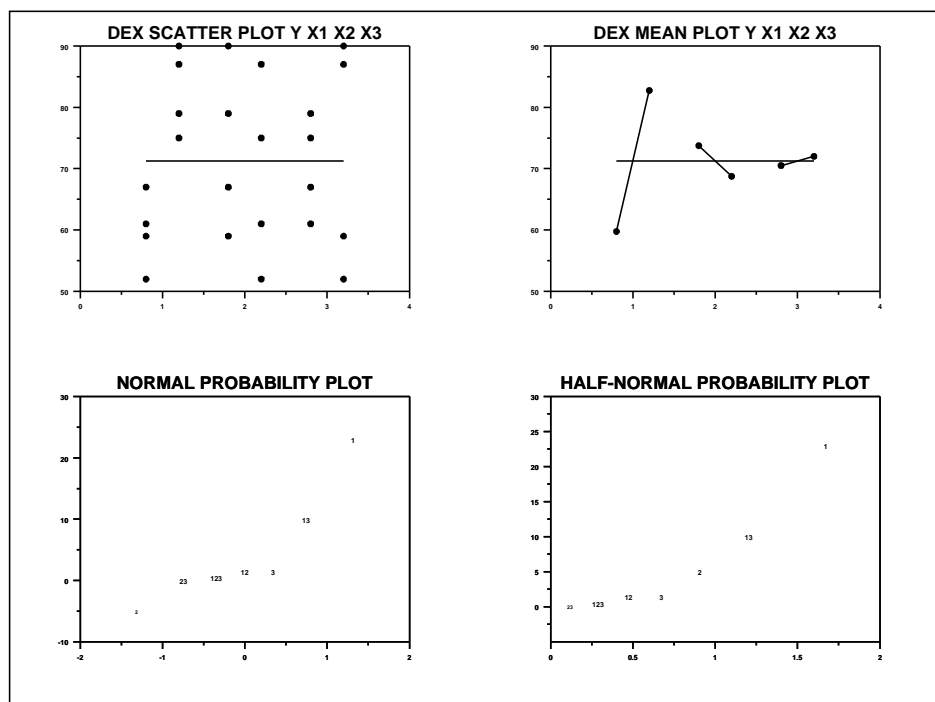
Pre-1987

PROGRAM 1

```

. THIS IS AN EXAMPLE OF A YATES ANALYSIS OF A 2**3 FULL FACTORIAL DESIGN.
DIMENSION 20 COLUMNS
SKIP 25; READ BOXSPRIN.DAT Y X1 X2 X3
YATES Y
SKIP 0; READ DPST1F.DAT TAG COEF
LET TAG2 = TAG; LET TAG3 = TAG
LET COEF2 = SORTC COEF TAG2
LET COEF3 = ABS(COEF)
LET COEF4 = SORTC COEF3 TAG3
.
ERASE; PRE-ERASE OFF; WINDOW CORNER COORDINATES 0 50 50 100
TITLE AUTOMATIC ; LINE BLANK; CHARACTER CIRCLE
CHARACTER HW 1.5 1.1; CHARACTER FILL ON
DEX SCATTER PLOT Y X1 X2 X3
WINDOW CORNER COORDINATES 50 50 100 100
LINES SOLID SOLID
DEX MEAN PLOT Y X1 X2 X3
WINDOW CORNER COORDINATES 0 0 50 50
CHARACTERS BLANK; LINES BLANK; TITLE NORMAL PROBABILITY PLOT
NORMAL PROBABILITY PLOT COEF2
CHARACTERS AUTOMATIC TAG2
PLOT YPLOT XPLOT XPLOT
WINDOW CORNER COORDINATES 50 0 100 50
TITLE HALF-NORMAL PROBABILITY PLOT
CHARACTERS BLANK; YTIC OFFSET 5 0
HALFNORMAL PROBABILITY PLOT COEF4
CHARACTERS AUTOMATIC TAG3
PLOT YPLOT XPLOT XPLOT
END OF MULTILOT

```



The following 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. = 0.70710676908E+00
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. = 0.50000000000E+00
(BASED ON PSEUDO-REP. ST. DEV.)

GRAND MEAN                  = 0.71250000000E+02
GRAND STANDARD DEVIATION    = 0.13719120979E+02

99% CONFIDENCE LIMITS (+-) = 0.31828401566E+02
95% CONFIDENCE LIMITS (+-) = 0.63531084061E+01
99.5% POINT OF T DISTRIBUTION = 0.63656803131E+02
97.5% POINT OF T DISTRIBUTION = 0.12706216812E+02

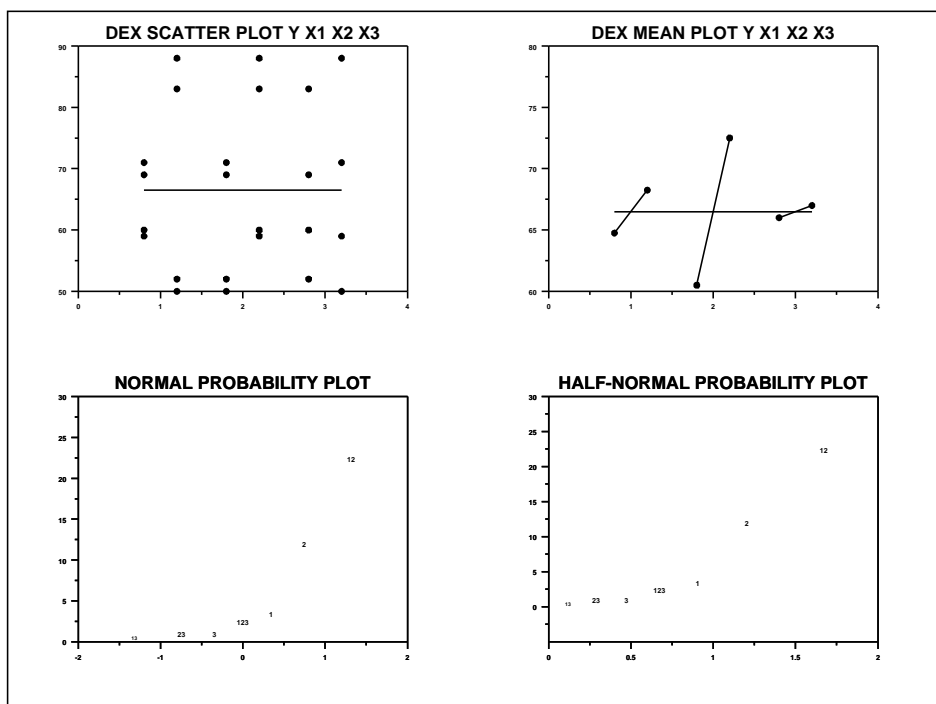
```

IDENTIFIER	EFFECT	T VALUE	RESSD MEAN+TERM	RESSD MEAN+TERMS
MEAN	71.25000		13.71912	13.71912
1	23.00000	46.0*	6.57647	6.57647
13	10.00000	20.0*	13.64734	3.44964
2	-5.00000	-10.0	14.53444	1.54110
3	1.50000	3.0	14.79302	1.29099
12	1.50000	3.0	14.79302	0.50000
123	0.50000	1.0	14.81553	0.00000
23	0.00000	0.0	14.81834	0.00000

NOTE--TAG, COEF, TCOEF, RESSD, & CUMULATIVE RESSD
 WRITTEN TO FILES DPST1F.DAT AND DPST2F.DAT

PROGRAM 2

```
. THIS IS AN EXAMPLE OF A YATES ANALYSIS OF A 2**(7-4) FRACTIONAL FACTORIAL DESIGN.
DIMENSION 20 COLUMNS
SKIP 25; READ BOXBBIKE2.DAT Y X1 X2 X3 X4 X5 X6 X7
YATES Y
ZKIP 0; READ DPST1F.DAT TAG COEF
LET TAG2 = TAG; LET TAG3 = TAG
LET COEF2 = SORTC COEF TAG2
LET COEF3 = ABS(COEF)
LET COEF4 = SORTC COEF3 TAG3
.
ERASE; PRE-ERASE OFF; WINDOW CORNER COORDINATES 0 50 50 100
TITLE AUTOMATIC ; LINE BLANK; CHARACTER CIRCLE
CHARACTER HW 1.5 1.1; CHARACTER FILL ON
DEX SCATTER PLOT Y X1 X2 X3
WINDOW CORNER COORDINATES 50 50 100 100
LINES SOLID SOLID
DEX MEAN PLOT Y X1 X2 X3
WINDOW CORNER COORDINATES 0 0 50 50
CHARACTERS BLANK; LINES BLANK; TITLE NORMAL PROBABILITY PLOT
NORMAL PROBABILITY PLOT COEF2
CHARACTERS AUTOMATIC TAG2
PLOT YPLOT XPLOT XPLOT
WINDOW CORNER COORDINATES 50 0 100 50
TITLE HALF-NORMAL PROBABILITY PLOT
CHARACTERS BLANK; YTIC OFFSET 5 0
HALFNORMAL PROBABILITY PLOT COEF4
CHARACTERS AUTOMATIC TAG3
PLOT YPLOT XPLOT XPLOT
END OF MULTILOT
```



The following 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.  =  0.35355339050E+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.   =  0.25000000000E+01
(BASED ON PSEUDO-REP. ST. DEV.)

GRAND MEAN                      =  0.66500000000E+02
GRAND STANDARD DEVIATION        =  0.13846092224E+02

99% CONFIDENCE LIMITS (+-)     =  0.15914201355E+03
95% CONFIDENCE LIMITS (+-)     =  0.31765541077E+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	66.50000		13.84609	13.84609
12	22.50000	9.0	7.41058	7.41058
2	12.00000	4.8	13.25393	2.88097
1	3.50000	1.4	14.81834	2.06155
123	2.50000	1.0	14.88568	1.22474
3	1.00000	0.4	14.94434	1.11803
23	1.00000	0.4	14.94434	0.70711
13	0.50000	0.2	14.95270	0.00000

NOTE--TAG, COEF, TCOEF, RESSD, & CUMULATIVE RESSD
 WRITTEN TO FILES DPST1F.DAT AND DPST2F.DAT