Analyzing Two-Level Factorials Having Missing Data

Part I: Illustration of Technique

Pat Whitcomb

Part II: Statistical Significance

Kinley Larntz

Missing Data

In a factorial design we make use of the orthogonality to estimate as many effects (counting the overall average as an effect) as there are experiments. Therefore missing data can cause some "interesting" problems:

1. Missing data make the factorial design unbalanced and non-orthogonal.

2. For each missing response value one effect is lost. How disastrous this is, will depend on the size of the design and its resolution.

3. The influence of certain design points in determining the model may become large.

4. Regression analysis will still provide a solution, but some criteria is needed to select the effects to estimate.

Estimating Effects

For unbalanced factorial designs estimate regression coefficients in a hierarchical fashion:

- The coefficients for the main effects are least squares estimates from the model containing the intercept, block effects (if any), and all main effects.
- Coefficients for the two-factor interactions are least squares estimates from the model containing the intercept, block effects (if any), all main effects and all two-factor interactions.
- Estimates for the higher order interactions are obtained in the same hierarchical manner, eliminating effects that can not be estimated.

Estimating Effects

To use normal probability plots for unbalanced factorial designs, the effects plotted must have a common error variance. Missing data can cause the variance associated with the estimated effects to differ. The effects must be adjusted (standardized) to correct for this problem.

Multiply each coefficient by two and then by the ratio of the standard errors of the first coefficient computed (usually A) to that of the current coefficient (i).

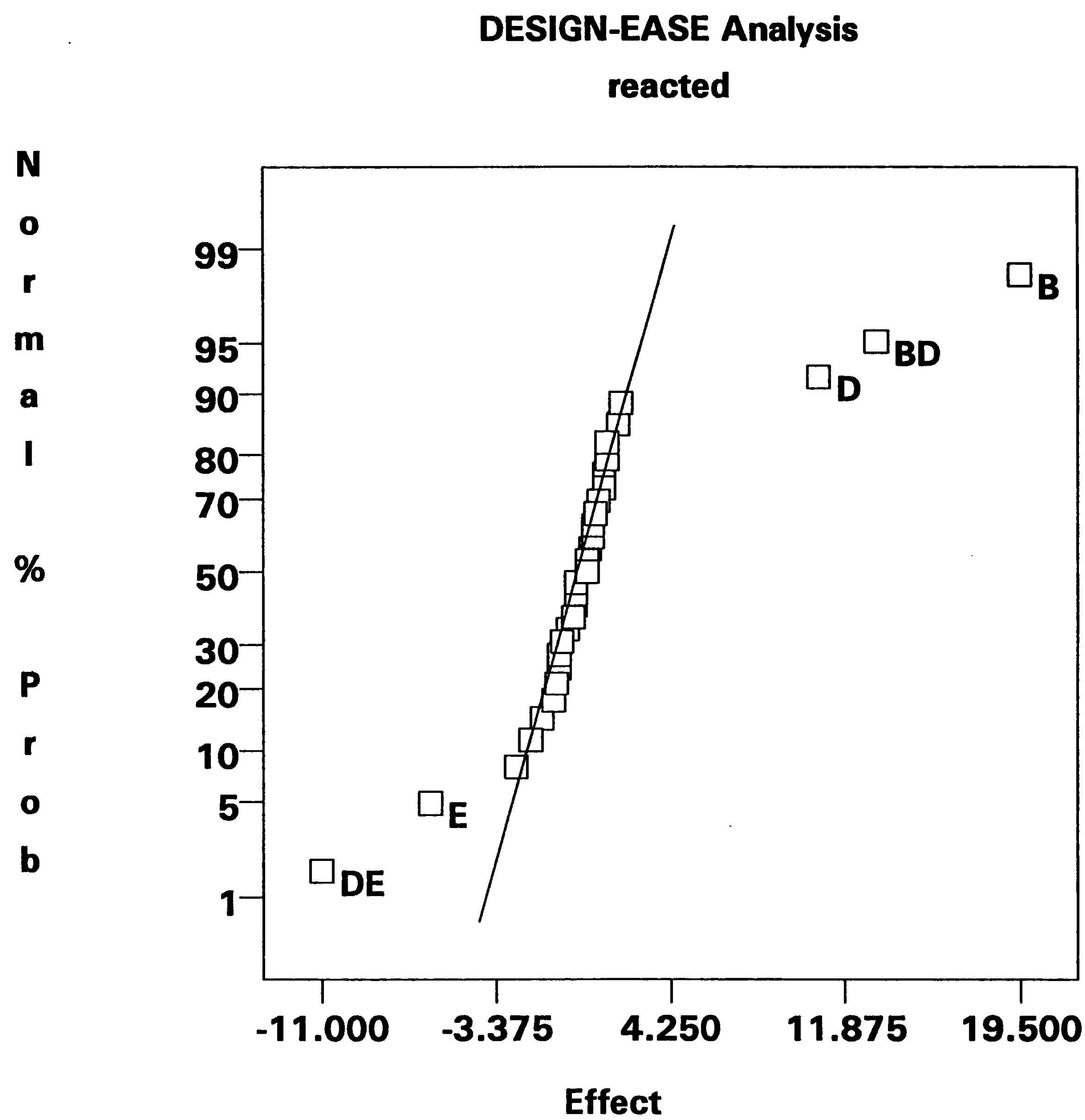
$$\text{Standardized effect}_i = (\beta_i)(2)\left(\frac{\hat{\sigma}_A}{\hat{\sigma}_i}\right)$$

2^5 Full Factorial

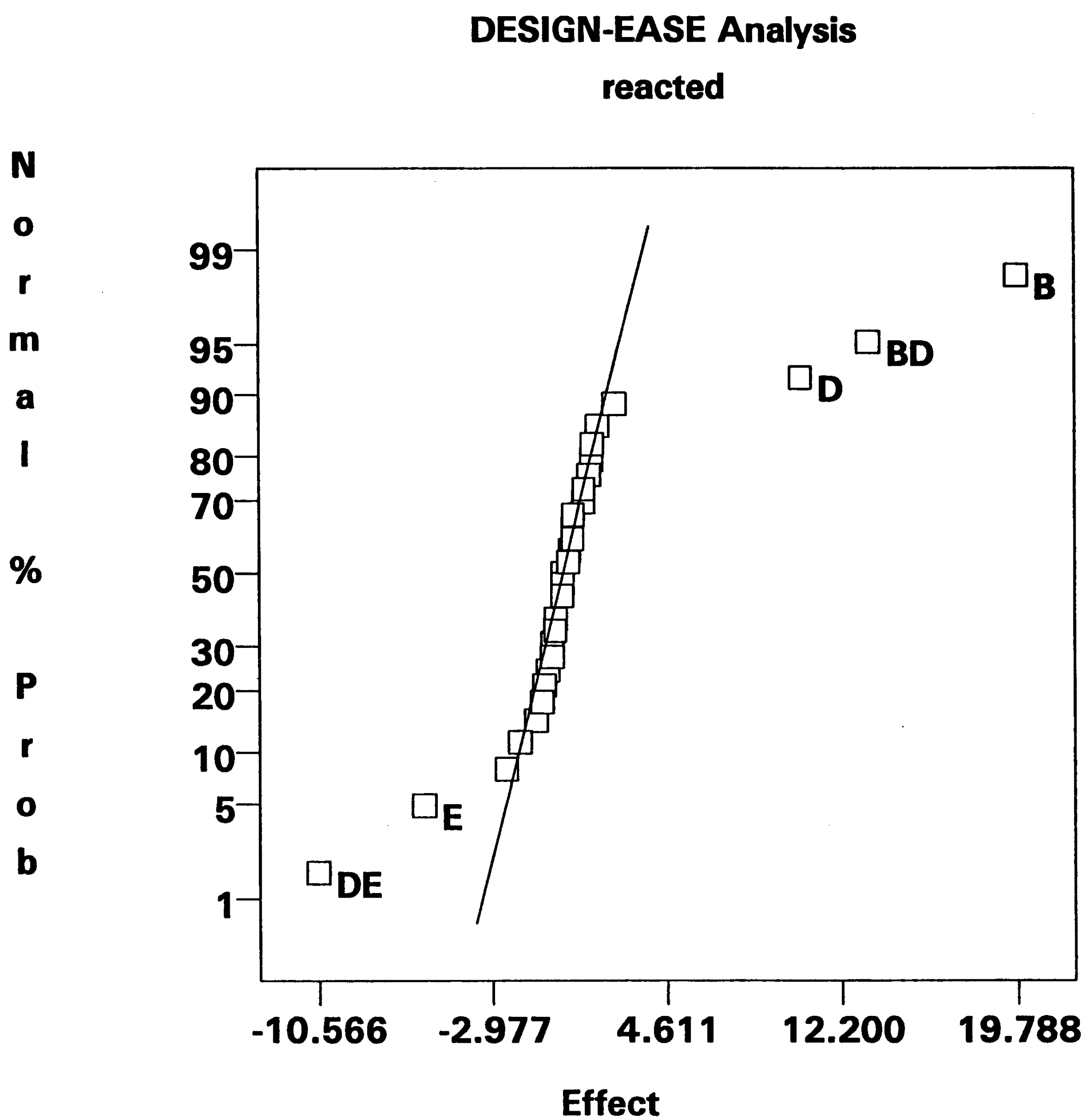
(BHH data page 377)

Effects List	
A	ABC
B	ABD
C	ABE
D	ACD
E	ACE
AB	ADE
AC	BCD
AD	BCE
AE	BDE
BC	CDE
BD	ABCD
BE	ABCE
CD	ACDE
CE	BCDE
DE	ABCDE

2^5 No Missing Data



2^5 One Missing Value



Don't Estimate ABCDE

Estimating Effects

In a fractional factorial, the higher order terms are aliased with the main effects and/or two-factor interactions. The resolution of a fractional factorial design limits how many, if any, higher order interactions are available to account for missing values.

Effects are estimated in the same hierarchical fashion as for a full factorial, but low order interactions may be given up to account for the missing values.

- In resolution IV and V designs, we typically lose one or more two-factor interactions.
- In a resolution III design, it may not be possible to estimate all the main effects.

2^{5-1} Fractional Factorial
(BHH data page 379)

½ Replicate of 5 factors in 16 experiments

Design Generator: E = ABCD

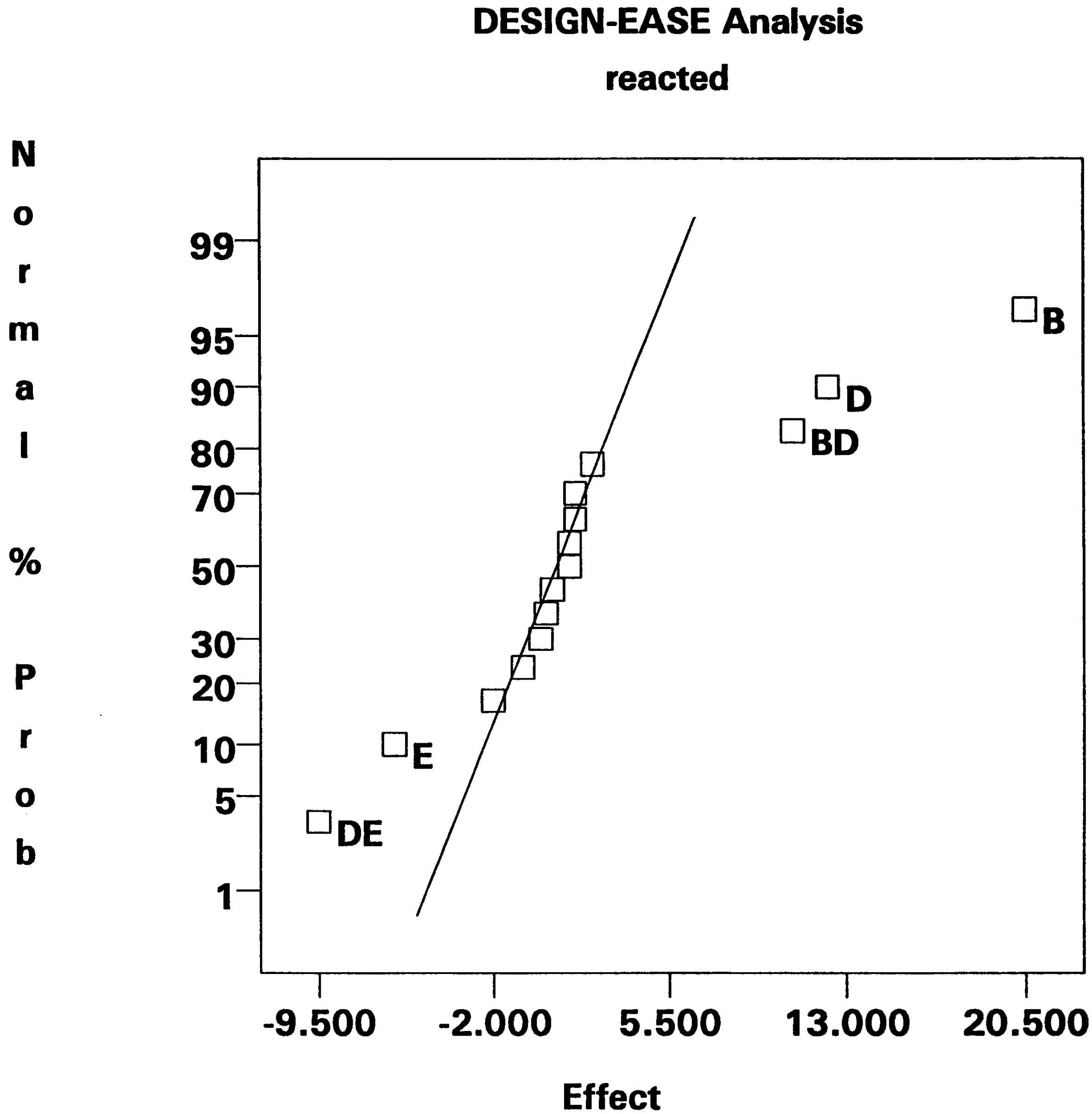
Defining Relation: I = ABCDE

<u>ALIASES</u>

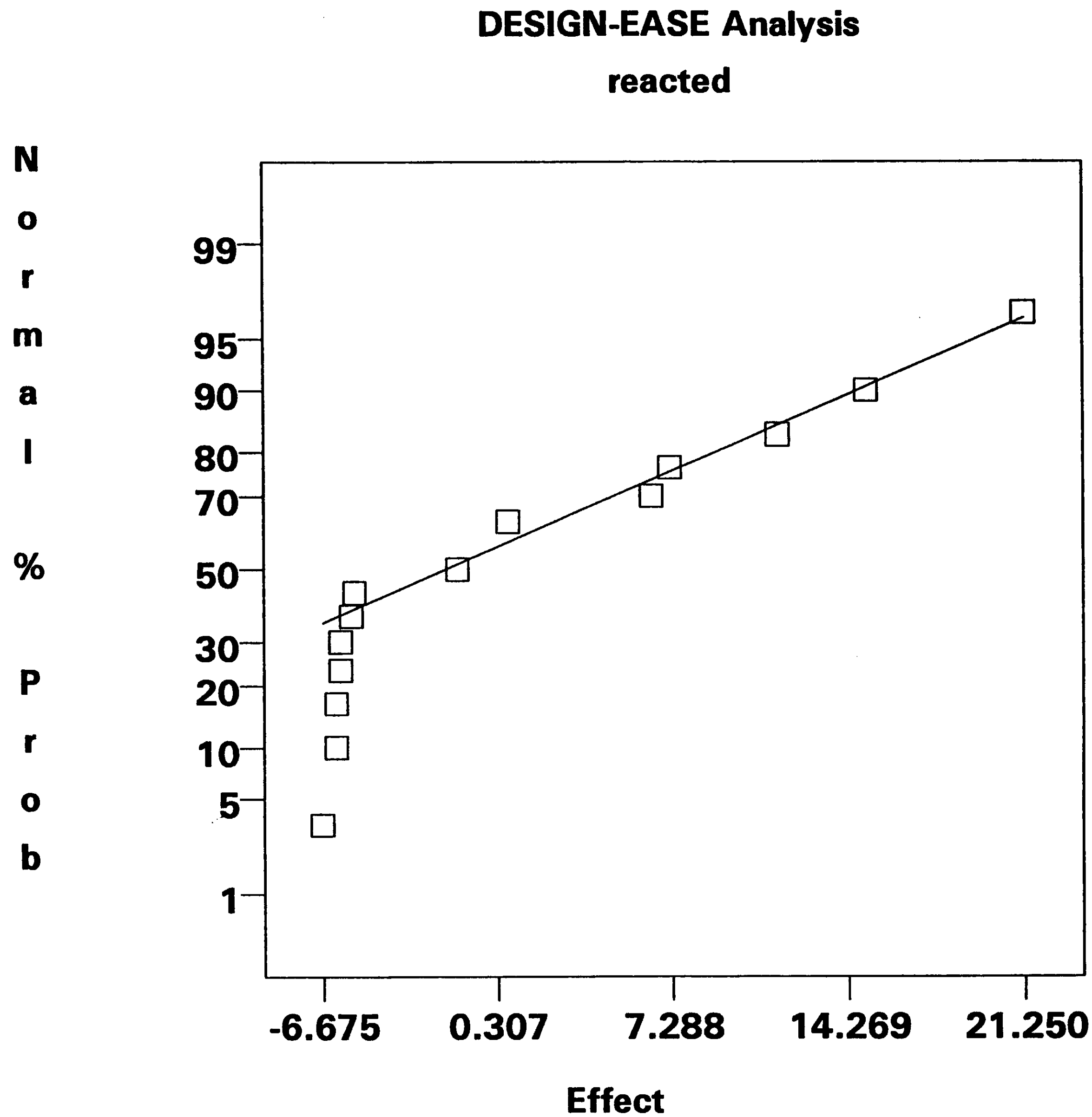
A = BCDE
B = ACDE
C = ABDE
D = ABCE
E = ABCD

AB = CDE	BD = ACE
AC = BDE	BE = ACD
AD = BCE	CD = ABE
AE = BCD	CE = ABD
BC = ADE	DE = ABC

2^{5-1} No Missing Data

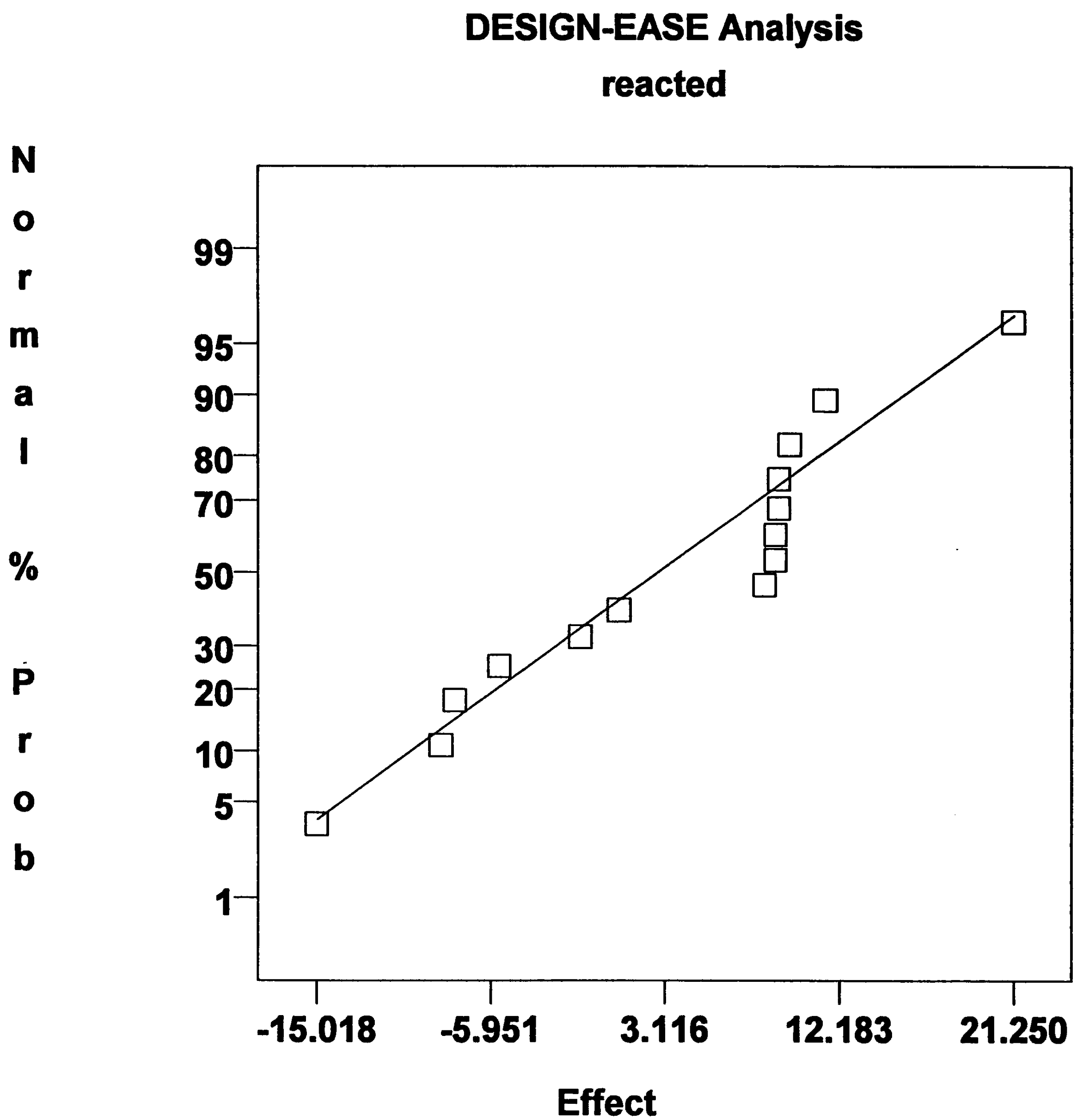


2^{5-1} One Missing Value



Don't Estimate DE

2^{5-1} One Missing Value



Don't Estimate BD

Estimating Effects

Estimate regression coefficients in a hierarchical fashion:

- **Coefficients for the main effects are least squares estimates from the model containing the intercept, block effects (if any), and all main effects.**
- **Coefficients for the two-factor interactions are least squares estimates from the model containing the intercept, block effects (if any), all main effects and all two-factor interactions.**
- **Estimates for the higher order interactions are obtained in the same hierarchical manner, eliminating effects that can not be estimated.**

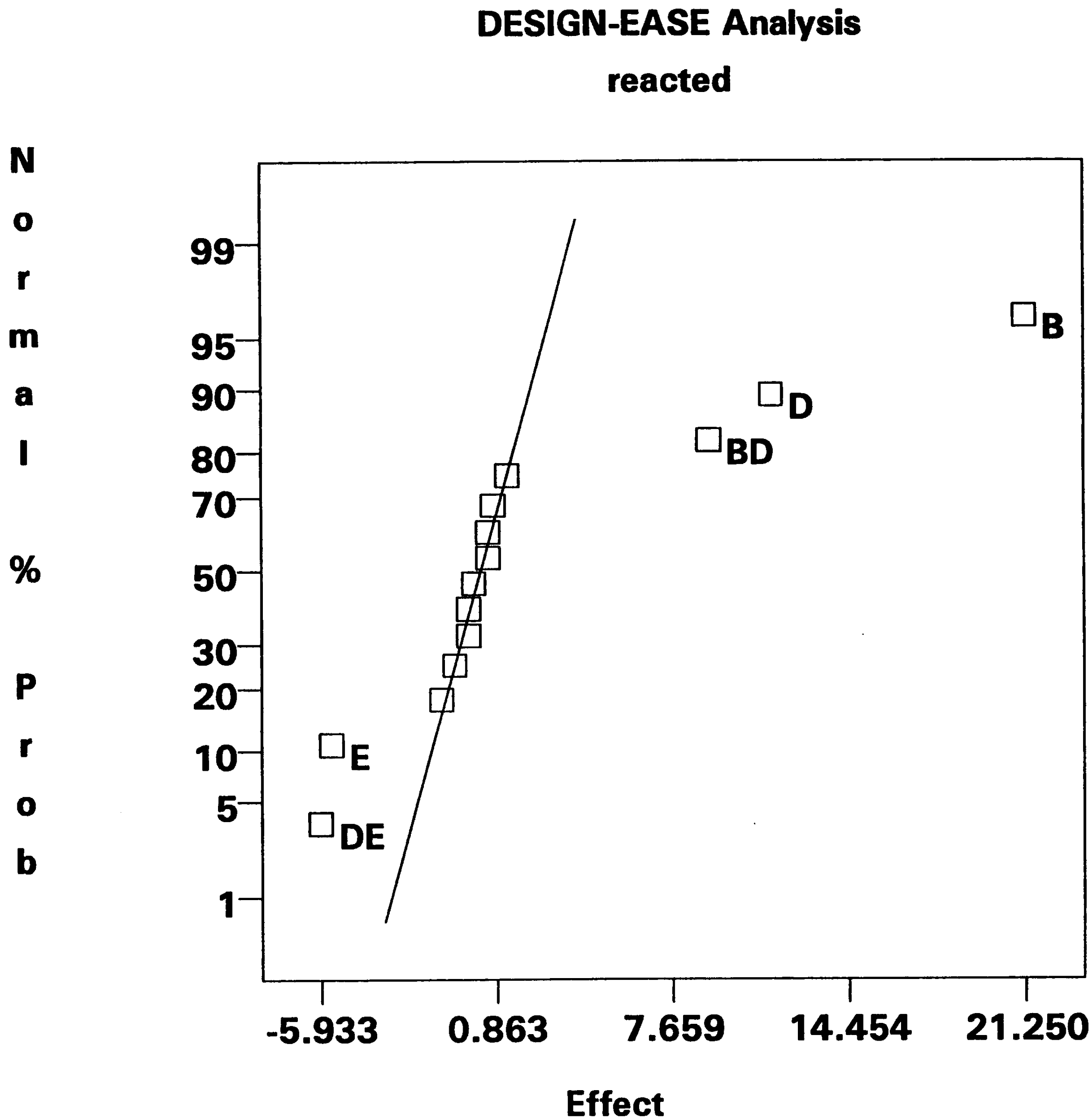
When the effects of a given order can not all be estimated, use forward stepwise regression to choose the subset to estimate.

Estimating Effects

2^{5-1} One Missing Value

1. Force the intercept.
2. Force the main effects.
3. Select a subset (9 out of the 10) of the two-factor interactions using forward stepwise regression.
4. Calculate the coefficients for the subset of interactions from the model containing the intercept, the five main effects and the nine two-factor interactions.
5. Standardize the effects.
6. Plot the main effects and interactions on normal probability paper and select the model.

2^{5-1} One Missing Value



Don't Estimate BC

Part II: Statistical Significance

Kinley Larntz

- Judging Significance in Saturated Two-level Factorials
- Correlation Test for Normality
- How Does Our Missing Data Procedure Affect Judging Significance?
- Concluding Comments

Judging Significance in Saturated Factorials

- Normal probability plot is good:
  - clear effects stand out
  - marginal effects can be examined for substantive significance
  - BUT - - still is somewhat subjective
  - calibration may be useful

Judging Significance in Saturated Factorials

- Two methods of calibration:
  - correlation test for normality
  - F testing:
    - pooling interactions
    - sequential testing
    - will not discuss here since not graphics based

Correlation Test for Normality

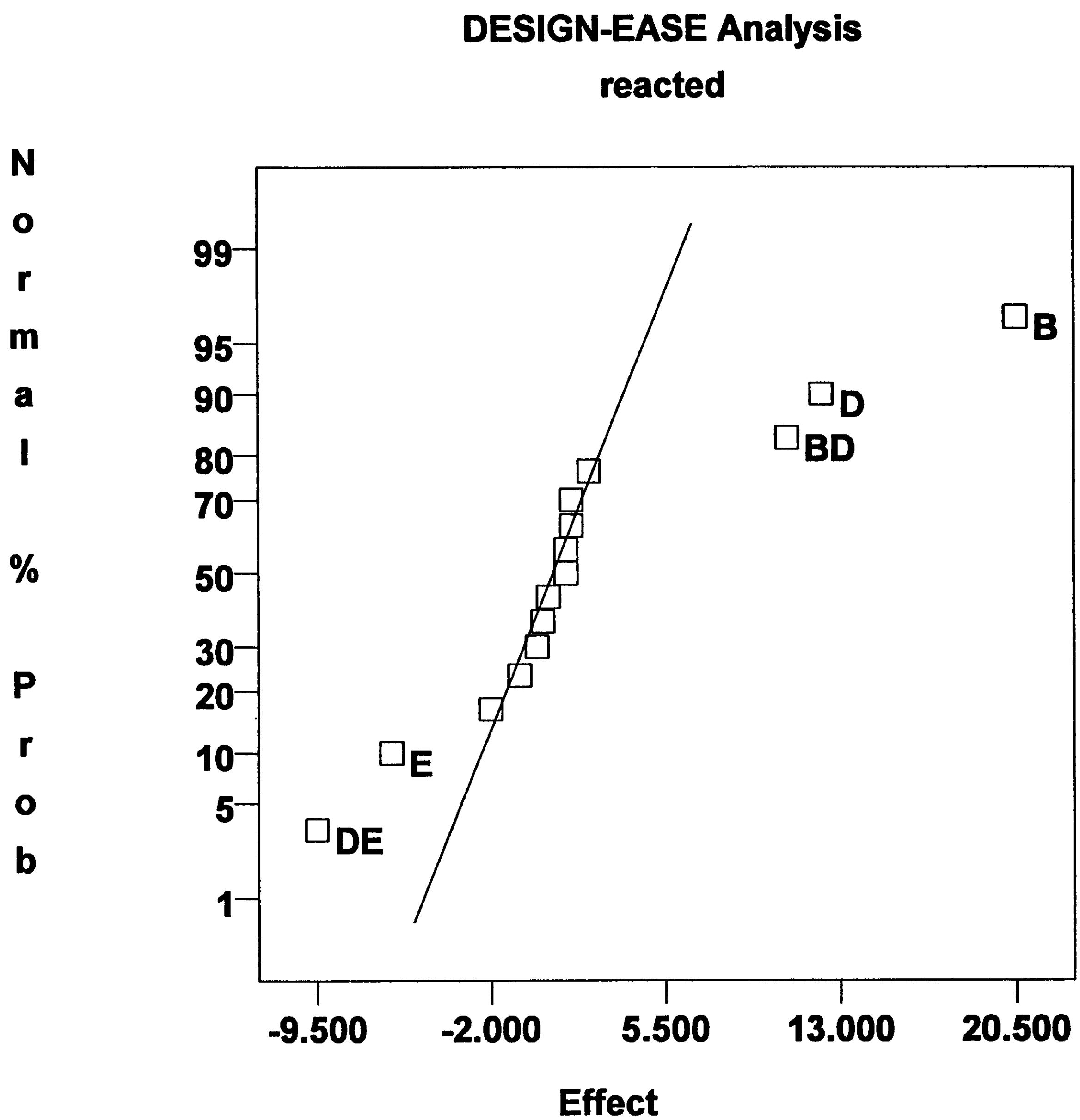
- **Calculate r for normal probability plots**
  - **all effects**
  - **omitting the most significant effect**
  - **omitting the two most significant effects**
  - **continue until r passes**

Normality Test
(Filliben, 1975)

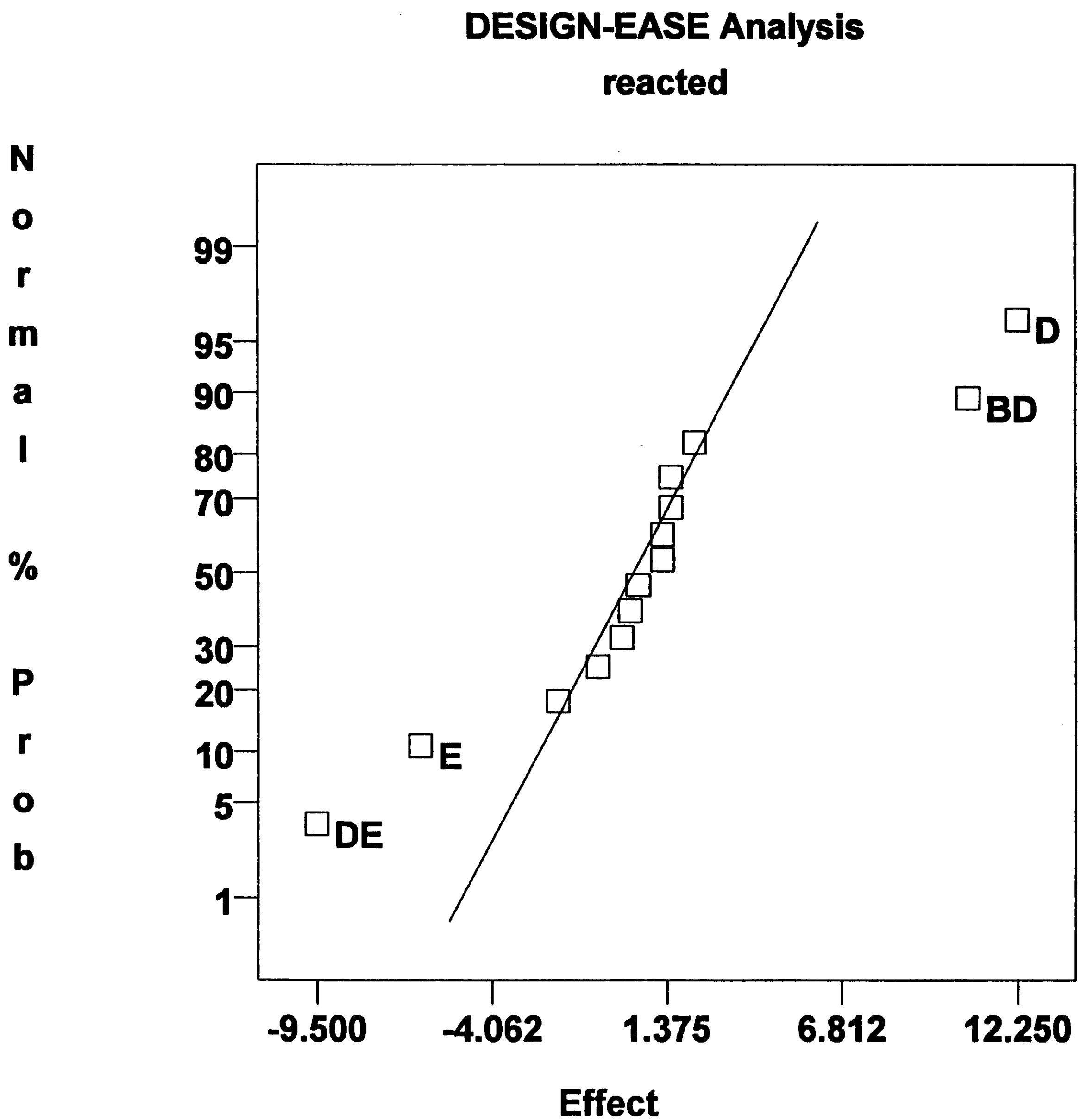
Critical values for $\alpha = 0.05$ test

n	$r_{.05}$
15	.937
14	.934
13	.931
12	.926
11	.922
10	.917

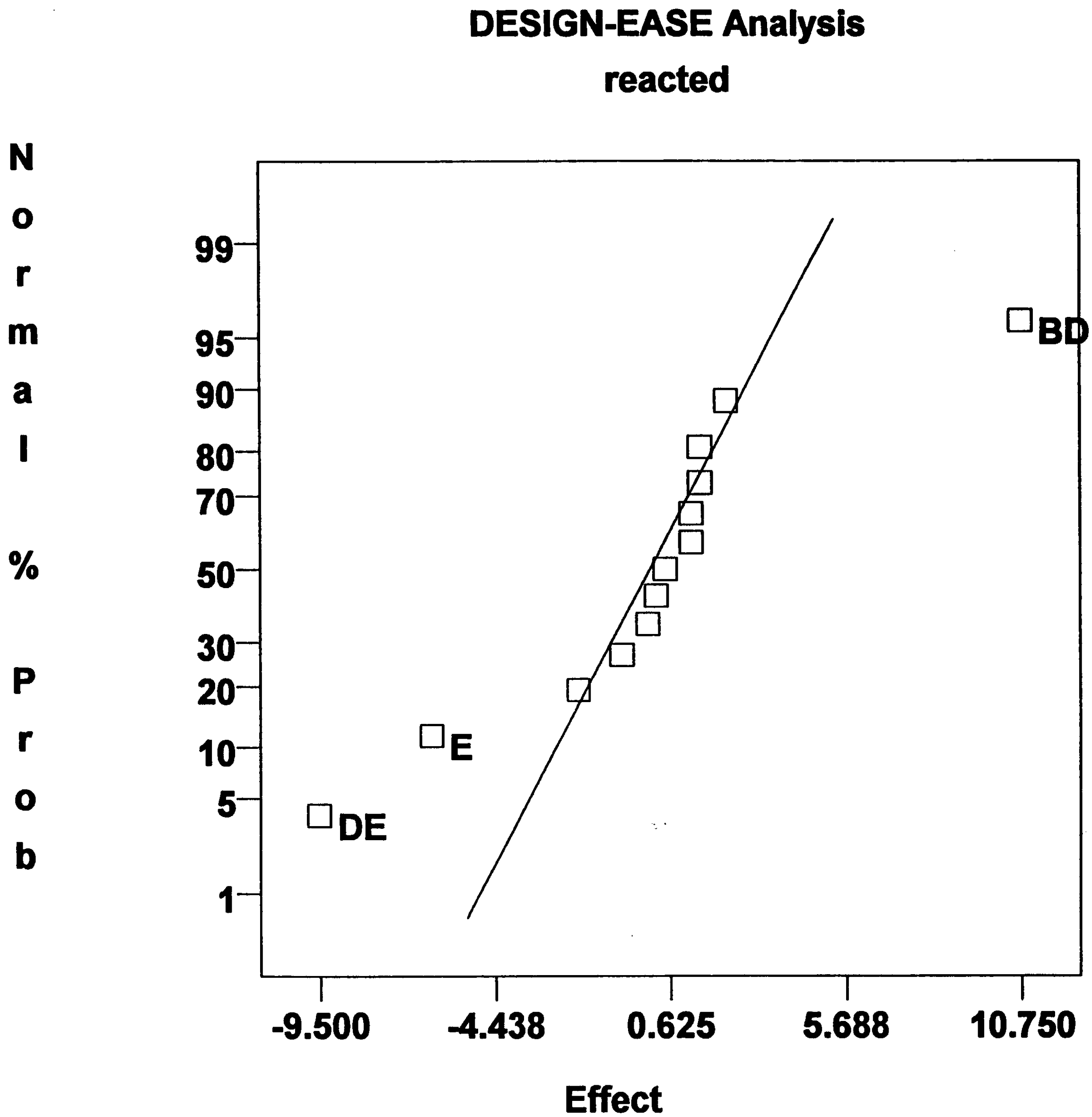
$n = 15 \quad r_{.05} = .937 \quad r = .922$



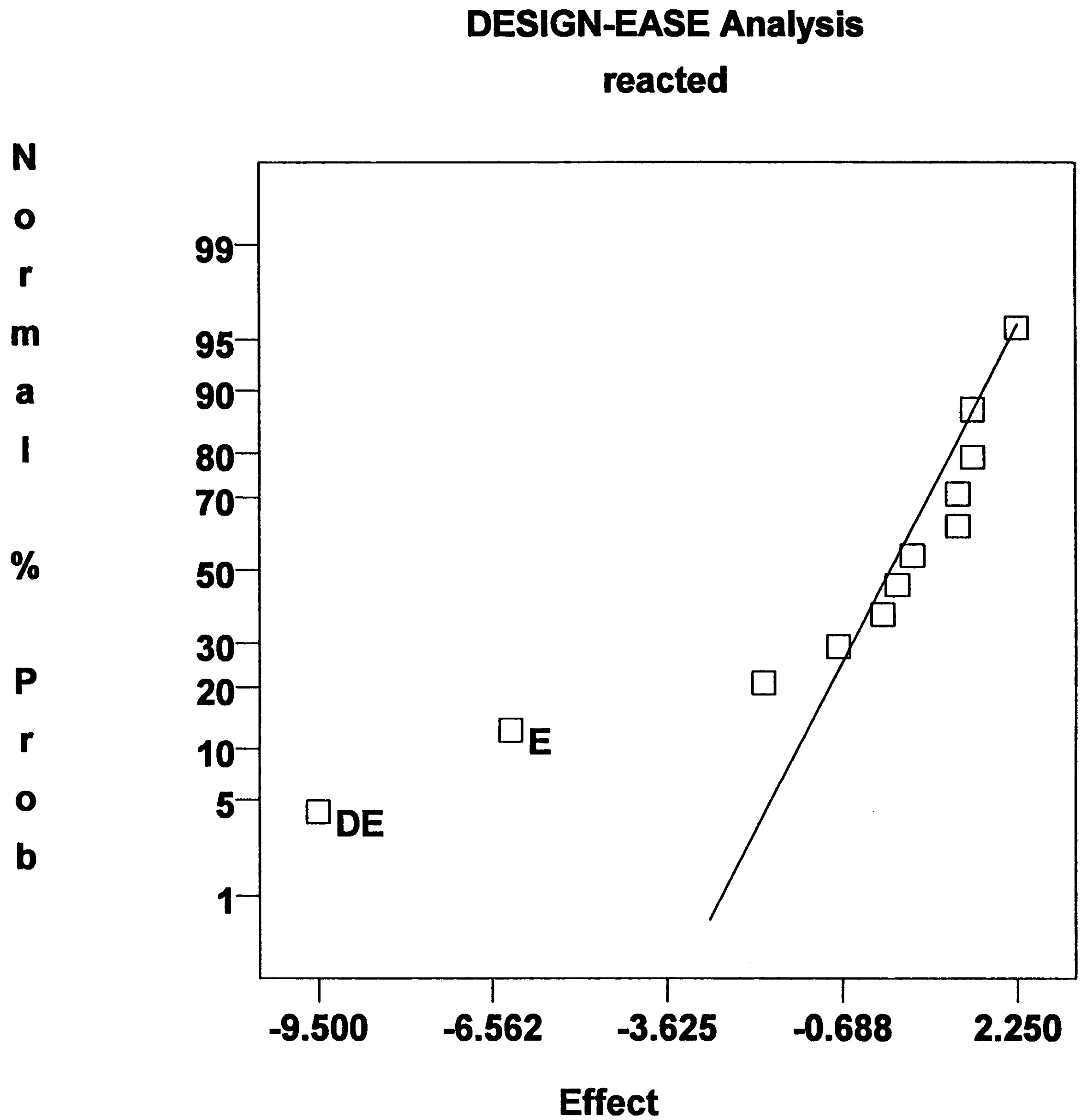
n = 14 $r_{.05}$ = .934   r = .930



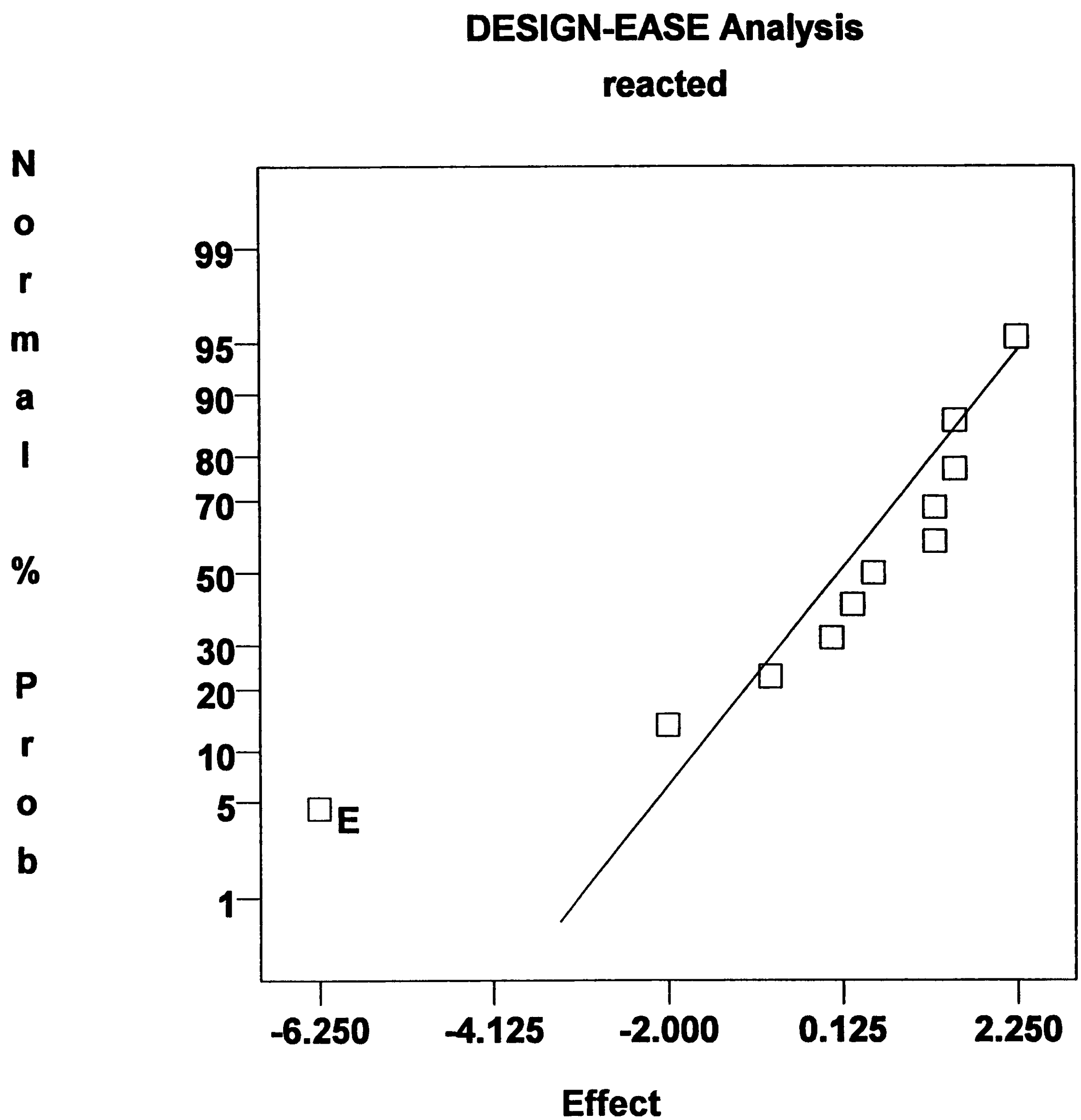
$n = 13 \qquad r_{.05} = .931 \qquad r = .914$



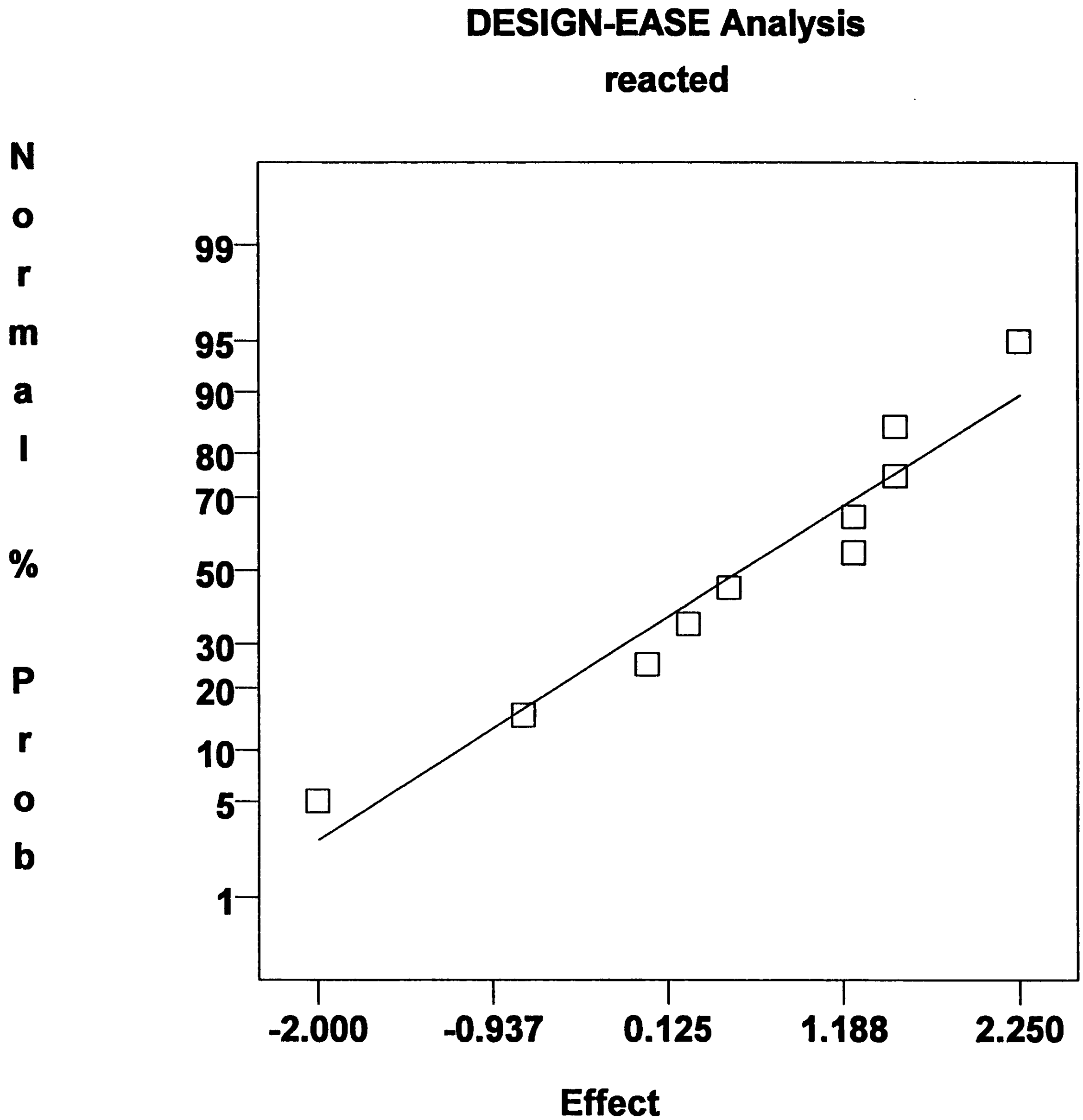
$n = 12 \quad r_{.05} = .926 \quad r = .868$



$n = 11 \qquad r_{.05} = .922 \qquad r = .877$



$n = 10 \qquad r_{.05} = .917 \qquad r = .965$



Small Power Study
(no missing data)

- Simulate 1000 cases
- $n = 16,\ 2^{5-1}$
- critical r at $\alpha = 0.05$ is .937

AB effect (as σ's)	rejection rate
0.0	4.5%
1.0	10.2%
2.0	55.0%
3.0	94.4%

How Does Our Missing Data Procedure Affect Judging Significance?

Calculation of Critical Values			
# of missing values	critical r-value (table)	observed rejection rate	critical r-value (observed)
0	.937	4.5%	.939
1	.934	4.9%	.934
2	.931	6.8%	.930
3	.926	5.5%	.926

Small Power Study
(missing data)

- Simulate 1000 cases
- one missing observation, n = 15

AB effect (as σ's)	rejection rate ($r_{.05} = .934$)
0.0	4.9%
1.0	6.3%
2.0	27.5%
3.0	66.3%

Concluding Comments

- Judging statistical significance appropriately leads to severe critical values.

- Missing data procedure (with forward selection of highest order interactions) requires <u>no</u> adjustment to r critical values.

- Can use normal probability plot to judge significance when there are missing values.