THE ROLE OF PURE ERROR ON NORMAL PROBABILITY PLOTS

Patrick J. Whitcomb
President
Stat-Ease, Inc.
Hennepin Square, #480
2021 East Hennepin Avenue
Minneapolis, MN 55413

Kinley Larntz
Professor
Department of Applied Statistics
352 Classroom-Office Building
1994 Buford Avenue
University of Minnesota
St. Paul, MN 55108

ABSTRACT

True replication in a designed experiment permits calculation of a pure error mean square which is used to check model goodness-of-fit (Draper and Smith, 1981). In two level factorial and fractional factorial designs, replication usually is done for center points, for full replicates of the design or, less frequently, for a balanced partial replicate of the factorial points. Typically the number of pure error degrees of freedom is not large. Note that pure error is not restricted to completely randomized designs, but is defined as the usual error from any variance reducing design. These designs include completely randomized designs, randomized block designs, incomplete block designs, Latin squares, etc. (Box and Draper, 1987).

Standard analysis of two level factorial and fractional factorial designs uses Daniel's normal or half normal probability plot of effects (Daniel, 1976, Box, Hunter and Hunter, 1978). We propose augmenting the usual probability plot of effects with points representing pure error. Assuming true replication, pure error points are guaranteed to be true null effects.

Details of the pure error representation on normal probability plots will be presented. For completely randomized designs, with each replicated point run exactly twice, single degree of freedom pure error effects are straightforward to add to the plot. For other designs and when any point is replicated more than twice, there are equivalent choices for representing pure error effects. We choose to represent pure error by plotting expected order statistics associated with pure error effects versus the usual normal expected order statistics. The number of pure error points added to the plot equals the number of degrees of freedom for pure error.

Examples will be given illustrating how pure error representation aids interpretation of the normal probability plot. Specific advantages of pure error representation include:

(a) combining pure error and null effects information about experimental error on the same normal probability plot,
(b) finding small non-null effects, and
(c) clarifying null effect experiments.

Calibration of normal probability plots, both with and without pure error, will also be discussed.

TEXT

True replication in a designed experiment permits calculation of the pure error mean square which can formally be used to check model goodness-of-fit (Draper and Smith, 1981). It is, of course, important that the replication be true replication, not simply remeasurement. That is, replicate experimental runs should be set up completely from scratch with the same values for the design variables and all steps of the experimental process should be redone. Failure to do true replication typically results in underestimating the experimental error. Tests based on such underestimates may cause the scientist to believe a statistical model that is more complex than necessary is required. The use of unnecessarily complex models obscures the relation of the response to the design variables and inflates the prediction variances based on the model.

We will use the following notation:

Y	response variable, random
X_j	design variables, fixed, $j=1,\ldots,p$
e	chance error, random

For each of n cases we relate Y to the X_j's and e through an additive model:

$$Y_i = f(X_{i1},\ldots,X_{ip}) + e_i, \quad i=1,\ldots,n$$

For inference purposes, the chance errors are usually assumed to follow a normal distribution with mean zero and constant variance, σ^2, independent of the X_j's.

The pure error mean square is defined by grouping cases with identical (or perhaps close) values of the design variables and calculating the usual one-way analysis of variance error mean square for the responses based on the grouping. Specifically, suppose there are k distinct patterns of values for the design variables and n_g = number of cases having distinct pattern g. The pure error is the one-way ANOVA within groups mean square, defined by:

$$\text{MS(Pure Error)} = \frac{\Sigma\Sigma(Y - \bar{Y}_g)^2}{\Sigma(n_g - 1)}$$

where $\bar{Y}_g$ is the group mean corresponding to the group to which each case belongs. The degrees of freedom for pure error are $\Sigma(n_g - 1) = n - k$.

To check for lack-of-fit for a particular model, the residual sum of squares is partitioned into two parts:

$$\text{Residual SS} = \text{Lack-of-Fit SS} + \text{Pure Error SS}$$

with the lack-of-fit sum of squares defined as the difference between the model residual sum of squares and the pure error sum of squares. Note that the pure error sum of squares is a lower bound for the residual sum of squares, since each model must give the same fitted value for all cases with identical values of the design variables. Assuming the degrees of freedom for residuals is greater than the degrees of freedom for pure error, a lack-of-fit F test may be constructed as:

$$F = \frac{(SS_{\text{residuals}} - SS_{\text{pure error}})/(df_{\text{residuals}} - df_{\text{pure error}})}{(SS_{\text{pure error}} \,/\, df_{\text{pure error}})}$$

In two level factorial and fractional factorial designs, replication is usually done for center points, for full replicates of the design or, less frequently, for a balanced partial replicate of the factorial points. Typically the number of pure error degrees of freedom is not large. The small number of denominator degrees of freedom for the lack-of-fit F test can cause difficulty since the denominator Mean Square will not be well estimated. Although the critical values for the F test account for this, we feel it is crucial for the scientist to examine the Pure Error Mean Square to judge if it is larger or smaller (particularly smaller) than might be reasonably expected. Unusually small

Pure Error Mean Squares (perhaps arising from remeasurement) may explain spurious large F values for lack-of-fit.

Note that the pure error is not restricted to completely randomized designs, but can be defined as the usual experimental error estimate from any variance reducing design (Box and Draper, 1987). Thus if a randomized blocks design were used to replicate a 2^3 factorial, the skeleton ANOVA would look like:

Source	df
blocks	1
A	1
B	1
C	1
AB	1
AC	1
BC	1
ABC	1
error	7

The randomized blocks error term with 7 df can be treated as pure error in constructing lack-of-fit tests for any particular model for the 2^3 factorial. This principle applies to any variance reducing design, including incomplete blocks, Latin Squares, etc.

Standard analysis of unreplicated orthogonal two level factorial and fractional factorial designs uses Daniel's normal and/or half normal probability plot of effects (Daniel, 1976, Box, Hunter and Hunter 1978). This technique may be extended to non-orthogonal designs by plotting standardized effects in place of the estimated effects (Design-Ease® manual, 1990). The standardization procedure divides each effect by a factor that is proportional to the standard error of the effect. (Note that a plot could be constructed directly from the t values for each effect.) Daniel proposed plotting the ordered two level effects versus the expected ordered statistics from a unit normal random sample of the same sample size as the number of effects. Outliers on this plot are identified as true effects for the purposes of statistical analysis. The basic underlying assumption is that the small or null effects may be used to estimate experimental error.

Potential difficulties with this method are many, but it has proven to work well in real world situations. In particular it allows the scientist to focus on important large effects. If the experiment has pure error degrees of freedom, the scientist can check the lack-of-fit of the model containing the identified effects. We advocate going one step further and incorporating the pure error information directly on the normal probability plot. If few or none of the effects are null, this method permits graphical identification of effects which would not be apparent from the normal probability plot of the factor effects alone.

The steps to incorporate the pure error information directly on the normal probability plot are:

(a) determine the usual effects and plot them on the normal probability plot.
(b) determine the pure error mean square and degrees of freedom.
(c) determine the standard error formula for the usual effects, for an orthoganal 2^k:

$$SE = \sqrt{\frac{4\sigma^2}{n}}$$

(d) plot on the normal probability plot points representing pure error. The number of points plotted equals the number of degrees of freedom for pure error. The pure error effects will be the expected values from a normal random sample with standard deviation (σ_{pe}) standardized in the same manner as the factor effects. In the orthogonal case, this will be

$$\sigma_{pe} = \sqrt{\frac{4\,(MS_{\text{pure error}})}{n}}$$

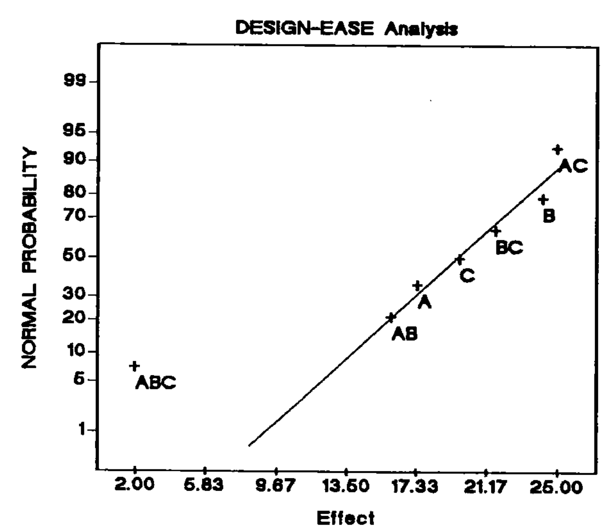
To see how the augmented plot is constructed, we will use the following example of a replicated (in blocks) 2^3 factorial:

A	B	C	Block1	Block2
-	-	-	89	112
+	-	-	61	97
-	+	-	70	108
+	+	-	78	113
-	-	+	64	87
+	-	+	88	112
-	+	+	95	112
+	+	+	156	172

The computed effects are:

Variable	Coefficient	Standardized Effect
A	8.750	17.50
B	12.125	24.25
C	9.875	19.75
AB	8.000	16.00
AC	12.500	25.00
BC	10.875	21.75
ABC	1.000	2.00

The normal probability plot of effects:



Based on this plot it looks like there is an ABC effect. But let's incorporate the pure error information on the normal probability plot. The sum of squares pure error is 263.00 with 7 degrees of freedom. The mean square pure error (263/7 = 37.57) estimates σ^2. Note that standard error for each effect is:

$$SE = \sqrt{\frac{4\sigma^2}{n}} = \sqrt{\frac{4(37.57)}{16}} = 3.06$$

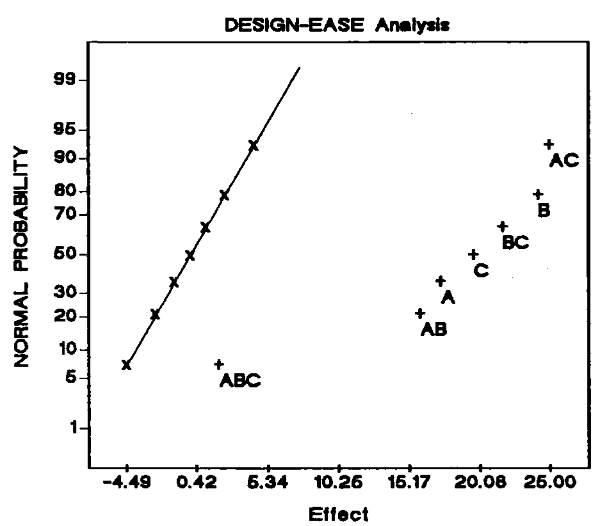
The pure error effects are the expected order statistics for a random sample of size 7 from a normal population with a mean of zero and a standard deviation of:

$$\sigma_{pe} = \sqrt{\frac{4(MS_{\text{pure error}})}{n}} = \sqrt{\frac{4(37.57)}{16}} = 3.06$$

The 7 pure error effects ($\text{effect}_i = z_i\sigma_{pe}$) are treated as a sample in their own right:

i	P_i	$\hat{z}_i$	effect_i
1	7.14	-1.465	-4.492
2	21.43	-0.792	-2.426
3	35.71	-0.366	-1.122
4	50.00	0.000	0.000
5	64.29	0.366	1.122
6	78.57	0.792	2.426
7	92.86	1.465	4.492

Normal probability plot, augmented with pure error effects:

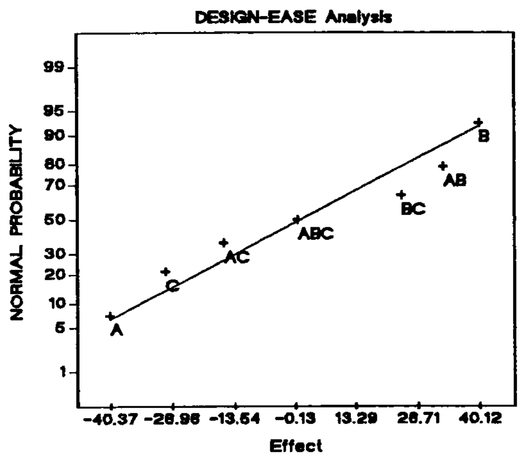


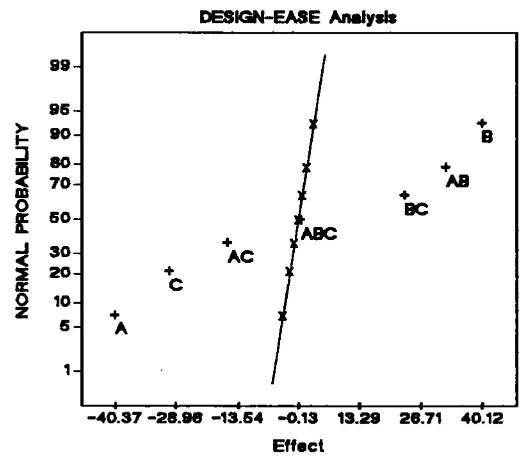
Clearly there are 6 large effects. Indeed, ABC is the only null effect. All the other effects are large and approximately the same size.

Example 2:

A	B	C	Block1	Block2
-	-	-	134	130
+	-	-	75	76
-	+	-	115	119
+	+	-	132	116
-	-	+	95	98
+	-	+	11	4
-	+	+	131	123
+	+	+	104	104

	Effects
A	-40.37
B	40.12
C	-28.38
AB	32.37
AC	-15.63
BC	23.38
ABC	0.625



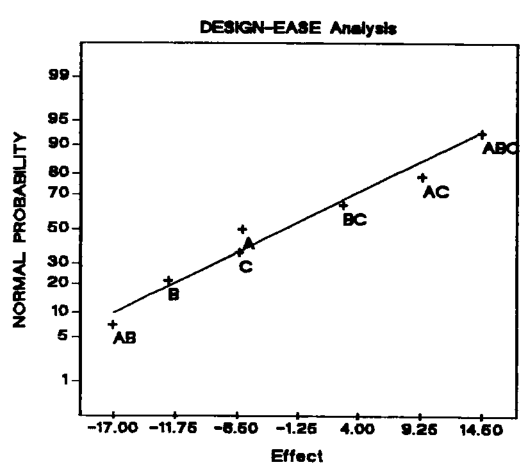


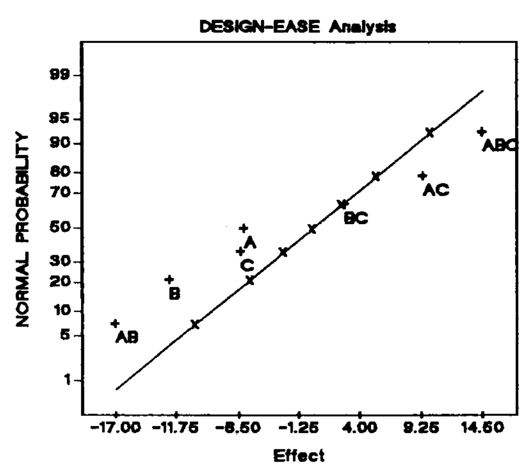
Based on the normal probability plot of the effects we might conclude all the effects are null. Augmenting the plot with the pure error effects clearly indicates 6 large effects. As in the previous example, ABC is the only null effect.

Example 3:

A	B	C	Block1	Block2
-	-	-	108	93
+	-	-	108	125
-	+	-	114	120
+	+	-	72	68
-	-	+	96	97
+	-	+	124	81
-	+	+	93	86
+	+	+	82	99

	Effects
A	- 6.00
B	-12.25
C	- 6.25
AB	-17.00
AC	9.50
BC	2.75
ABC	14.50





Both the normal and augmented normal probability plots show only null effects.

Some statistical packages offer calibration of normal probability plots (Devore and Peck, 1986). The basic idea is to calculate the correlation coefficient (r) for the normal plot (the correlation between the order statistics and the effects) and compare it to what might be expected under the null hypothesis that the sample is from a normal population. A small value of r indicates nonnormality. The addition of the augmented points from the pure error complicates this calibration somewhat. To see what the effect of augmentation is, we calculated the distribution of r for null plots with and without pure error effects. To illustrate, we used a 2^3 factorial in 2 blocks of size 8, i.e. 7 factor effects and 7 pure error effects. 99,999 plots were simulated to determine a critical value ($\alpha = 5\%$) of r. The critical value for factor effects only is r = 0.8976. Critical r for factor and pure error effects is r = 0.8776.

To see the power of the normal probability plot, we calculated the distribution of r when one factor had a true effect of Δ σ's (σ is the true pure error) for various values of Δ. We ran 10,000 simulations for each Δ value and compared the r values to the critical value found above. Calculating the percentage of r values less than the critical value gives an estimate of the power of each respective plot:

Δ	Power of the factors only plot	Power of the augmented plot
0.0	4.97%	5.03%
1.0	8.28%	10.06%
2.0	34.95%	54.92%
3.0	73.47%	96.99%
4.0	94.35%	99.98%

CONCLUSION

Incorporating pure error information in the form of pure error effects on the normal probability plot enhances the utility of these plots for selecting factor effects. If few or none of the effects are null, this method permits graphical identification of effects which would not be apparent from the normal probability plot of the factor effects alone. The pure error effects are a particularly useful aid for the novice in identifying which of the factor effects are null. The addition of the pure error effects to the normal probability plot also triggers requests for help. When the novice observes that the pure error effects don't coincide with what they think the null effects are, they ask why. In our experience this has prevented several invalid interpretations from occurring. In addition, the pure error effects improve the power of normal probability plots for detecting real factor effects.

BIBLIOGRAPHY

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces.* New York: Wiley.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building.* New York: Wiley.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation.* New York: Wiley.

Devore, J. and Peck, R. (1986). *Statistics, The Exploration and Analysis of Data.* St. Paul: West.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis,* 2nd ed. New York: Wiley.

LSC: 522:00:000