# Quick and Easy Analysis of Unreplicated Factorials

**Russell V. Lenth**

Department of Statistics and Actuarial Science
University of Iowa
Iowa City, IA 52242

Box and Meyer (1986) introduced a method for assessing the sizes of contrasts in unreplicated factorial and fractional factorial designs. This is a useful technique, and an associated graphical display popularly known as a Bayes plot makes it even more effective. This article presents a competing technique that is also effective and is computationally simple. An advantage of the new method is that the results are given in terms of the original units of measurement. This direct association with the data may make the analysis easier to explain.

## 1. INTRODUCTION

This article concerns the analysis of unreplicated factorial and fractional factorial designs such as those used in screening experiments. Such experiments usually include many factors. Typically, one computes point estimates of a large number of contrasts, and all of these estimates have the same variance. The underlying principle of most analyses of such data is that of *effect sparsity*, the idea that one can usually expect only a small number of the effects to be "active" (i.e. nonzero) in the process under study. (The literature on two-level factorials usually focuses on differences between the means at the "high" and "low" levels, and these are often called "effects." In a classical analysis-of-variance model, however, an effect is defined as the difference between the high mean and the grand mean—half as large. In this article, I attempt to avoid confusion by referring to the high-minus-low differences as contrasts rather than effects.)

One popular method of analysis (or at least interpretation) is to construct a half-normal plot or a normal plot of the contrasts (see Box, Hunter, and Hunter 1978; Daniel 1959). The effect-sparsity principle suggests that the active contrasts will tend to show up as outliers. Normal plots have useful diagnostic properties as well. For example, a single outlier in the data will tend to split the normal plot into two distinct sections. The disadvantage of the normal-plot methods is that their interpretation is somewhat subjective.

Box and Meyer (1986) proposed a more formal technique. It involves computing a posterior probability that each contrast is active. The prior information is summarized in two parameters, $\alpha$ (the probability that a contrast is active) and $k$ (the inflation in the standard deviation produced by an active contrast). Based on an empirical study, the authors suggested using $\alpha = .2$ and $k = 10$.

The Box–Meyer method is useful, and its results can be effectively presented in a graphical form known as a *Bayes plot*. This is simply a stick diagram or bar graph of the posterior probabilities, usually with horizontal guidelines at .5 and 1.0 to aid in reading the values. A contrast having a posterior probability greater than .5 is deemed to be more likely active than inactive.

This article presents an effective alternative method for formal analysis of unreplicated factorials. It is based on a simple formula for the standard error of the contrast estimates. The usual $t$ procedures can be used to interpret the results. Better yet, the contrasts can be plotted in a manner similar to the Bayes plot, with cutoff limits based on the standard error. An advantage of this plot over the Bayes plot is that the numerical values of the contrasts are displayed. Thus one can assess both the size and "significance" of the contrasts by looking at just one graph. An additional advantage (unimportant in this age of computers) is that the computations are easy to carry out by hand, whereas the Box–Meyer method requires specialized software.

## 2. PROPOSED METHOD

Let $\kappa_1, \kappa_2, \ldots, \kappa_m$ denote the contrasts of interest, and let $c_1, c_2, \ldots, c_m$ denote the corresponding estimates. In the usual setting, the $c_i$ are independent realizations of $N(\kappa_i, \tau^2)$ random variables; that is, the sampling distributions of the $c_i$ are (approximately) normal with possibly different means $\kappa_i$ but with equal variances $\tau^2$. Let

$$s_0 = 1.5 \times \underset{j}{\text{median}}\, |c_j| \qquad (1)$$

and define the *pseudo standard error* (PSE) of the contrasts to be

$$\text{PSE} = 1.5 \times \underset{|c_j|<2.5s_0}{\text{median}}\, |c_j|. \qquad (2)$$

Note that (1) and (2) are identical, except that the median in (2) is taken over a restricted set of inlying $|c_j|$'s. It is shown in Section 4 that (2) is a fairly good estimate of $\tau$ when the effects are sparse.

The result (2) may be used in the natural way. For example, let

$$\text{ME} = t_{.975;d} \times \text{PSE}, \qquad (3)$$

where $t_{.975;d}$ is the .975th quantile of a $t$ distribution on $d$ df. (For reasons described in Sec. 4, $d = m/3$ is suggested.) ME is a *margin of error* for $c_i$ with approximately 95% confidence; that is, one can construct an approximate 95% confidence interval for $\kappa_i$ using $c_i \pm$ ME.

An important concern is that several inferences are being made simultaneously. With a large number, $m$, of contrasts, one can expect one or two estimates of inactive contrasts to exceed the ME leading to false conclusions. To account for this possibility, define also a simultaneous margin of error (SME):

$$\text{SME} = t_{\gamma;d} \times \text{PSE}, \qquad (4)$$

where

$$\gamma = (1 + .95^{1/m})/2. \qquad (5)$$

This is derived from the fact that the estimates are independent. It is exact, not conservative.

For convenience, Table 1 provides values of $t_{.975;d}$ and $t_{\gamma;d}$ for $d = m/3$, $\gamma$ given by (5), and common values of $m$. They were computed using an algorithm that allows fractional degrees of freedom.

Rather than constructing formal tests of hypothesis or confidence intervals, it is suggested that the information be displayed graphically in a style similar to a Bayes plot or an analysis-of-means plot (Ott 1967): Construct a bar graph showing the (signed) contrasts, and add reference lines at ±ME and at ±SME. A contrast whose bar extends beyond the SME lines is clearly active, one which does not extend beyond the ME lines cannot be deemed active, and one in between is in a zone of uncertainty where a good argument can be made both for its being active and for its being a happenstance result of an inactive contrast. Examples are given in Section 3.

*Table 1. Quantiles of the t Distribution for Common Values of m and Degrees of Freedom d = m/3 (not rounded to an integer)*

| $m$ | $t_{.975;d}$ | $t_{\gamma;d}$ |
|---|---|---|
| 7 | 3.76 | 9.01 |
| 15 | 2.57 | 5.22 |
| 31 | 2.22 | 4.22 |
| 63 | 2.08 | 3.91 |
| 127 | 2.02 | 3.84 |
| 255 | 1.99 | 3.89 |

NOTE: The .975th and $\gamma$th quantiles are used in constructing ME and SME, respectively.

## 3. ILLUSTRATIONS

I illustrate the suggested procedure using two of the four examples of Box and Meyer (1986), examples II and IV. Both examples consist of 16 runs in unreplicated two-level designs.

Example II is a $2^{9-5}_{\text{III}}$ design from Taguchi and Wu (1980). The response is tensile strength and the factors (and one-letter designations) are thickness ($T$), method ($W$, for "way"), current ($C$), rods ($R$), period ($P$), material ($M$), angle ($A$), opening ($O$), and preheating ($H$). The generators of the design are as follows: $P = WCR$, $M = -TWCR$, $A = -TR$, $O = -TC$, and $H = TCR$. Example IV is a $2^4$ experiment from Davies (1954). The response is yield of isatin, and the factors are acid strength ($S$), time ($t$), amount of acid ($A$), and temperature ($T$). The one-letter abbreviations used here are designed to suggest the names of the factors and are not the same as those assigned by Box and Meyer.

Table 2 gives the data (in Yates's standard order) for the two examples, and Table 3 shows the estimates of the contrasts. (Since ex. II has a fractional design, several contrasts are confounded with one another. The aliases for all main effects and two-way interactions are shown in the table.) To compute the PSE of the contrasts in example II, first obtain $s_0$ using the median of all absolute contrasts: $s_0 = 1.5 \times .30 = .45$. Then perform the identical calculation, only excluding the two contrasts that exceed $2.5s_0 = 1.13$, obtaining PSE $= 1.5 \times .15 = .225$. In example IV, we obtain $s_0 =$ PSE $= .114$ (nothing is excluded in the second step).

Using Table 1 with $m = 15$, the ME's are computed as the following:

*Example II.* ME $= 2.57 \times .225 = .58$; SME $= 5.22 \times .225 = 1.17$.

*Example IV.* ME $= 2.57 \times .114 = .29$; SME $= 5.22 \times .114 = .60$.

Using these quantities for guidance in examining Table 3, two contrasts emerge as active ones in example II—$P$ (period) and $M$ (material). Evidently, the strength is higher for the longer period and the low level of material. In example IV, two contrasts,

*Table 2. Data for Two of the Examples in Box and Meyer (1986)*

| Example II | | | | | | | | | | Example IV | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *T* | *W* | *C* | *R* | *P* | *M* | *A* | *O* | *H* | *Strength* | *S* | *t* | *A* | *T* | *Yield* |
| − | − | − | − | − | − | − | − | − | 43.7 | − | − | − | − | .08 |
| + | − | − | − | − | + | + | + | + | 40.2 | + | − | − | − | .04 |
| − | + | − | − | + | + | − | − | − | 42.4 | − | + | − | − | .53 |
| + | + | − | − | + | − | + | + | + | 44.7 | + | + | − | − | .43 |
| − | − | + | − | + | + | − | + | + | 42.4 | − | − | + | − | .31 |
| + | − | + | − | + | − | + | − | − | 45.9 | + | − | + | − | .09 |
| − | + | + | − | − | − | − | + | + | 42.2 | − | + | + | − | .12 |
| + | + | + | − | − | + | + | − | − | 40.6 | + | + | + | − | .36 |
| − | − | − | + | + | + | + | − | + | 42.4 | − | − | − | + | .79 |
| + | − | − | + | + | − | − | + | − | 45.5 | + | − | − | + | .68 |
| − | + | − | + | − | − | + | − | + | 43.6 | − | + | − | + | .73 |
| + | + | − | + | − | + | − | + | − | 40.6 | + | + | − | + | .08 |
| − | − | + | + | − | − | + | + | − | 44.0 | − | − | + | + | .77 |
| + | − | + | + | − | + | − | − | + | 40.2 | + | − | + | + | .38 |
| − | + | + | + | + | + | + | + | − | 42.5 | − | + | + | + | .49 |
| + | + | + | + | + | − | − | − | + | 46.5 | + | + | + | + | .23 |

NOTE: Presentation is in standard order according to the first four factors.

*T* (temperature) and *tT*, (time × temperature) are fairly close to the ME. One or two values of this magnitude can easily occur among 15 estimates, even if none of the factors affect the response.

Figures 1 and 2 show these results graphically. These graphs are similar to analysis-of-means plots (Ott 1967) except that two sets of limits are provided. Estimates that fall within the inner limits (i.e., most contrasts in both figures) show little evidence of being active. Those that fall between the inner and outer limits could be described as *possibly* active. In example IV, *T* is nearly in this range. Those that fall outside the outer limits are *probably* active, such as *P* and $-M$ in example II, Figure 1. The visual impressions one gains from these plots are much the same as the Bayes plots. Normal or half-normal plots of the contrasts are also recommended (but not shown here) for their diagnostic value.

## 4. JUSTIFICATION

The effect-sparsity assumption is that most $\kappa_i$ are equal to 0. Suppose for a moment that they are *all*

*Table 3. Estimated Contrasts for the Two Examples in Table 2*

| Contrast index | Example II | | Example IV | |
|---|---|---|---|---|
| | *Aliases* | *Estimate* | *Contrast* | *Estimate* |
| — | mean | 42.96 | mean | .382 |
| 1 | **T**, *−CO, −RA, −PM* | .12 | *S* | −.191 |
| 2 | **W**, *−MH* | −.15 | *t* | −.021 |
| 3 | *TW, PH* | .30 | *St* | −.001 |
| 4 | **C**, *−TO, −AH* | .15 | **A** | −.076 |
| 5 | **−O**, *TC, RH* | .40 | *SA* | .034 |
| 6 | *WC, RP, WC* | −.02 | *tA* | −.066 |
| 7 | *−WO, −RM, −PA* | .37 | *StA* | .149 |
| 8 | **R**, *−TA, −OH* | .40 | **T** | .274 |
| 9 | **−A**, *TR, CH* | −.05 | *ST* | −.161 |
| 10 | *WR, CP, MO* | .42 | *tT* | −.251 |
| 11 | *−WA, −CM, −PO* | .13 | *StT* | −.101 |
| 12 | *CR, TH, WP, AO* | .12 | *AT* | −.026 |
| 13 | **H**, *−WM, −CA, −RO* | −.37 | *SAT* | −.006 |
| 14 | **P**, *−TM* | 2.15 | *tAT* | .124 |
| 15 | **−M**, *TP, WH* | 3.10 | *StAT* | .019 |

NOTE: In example II, aliases for all main effects and two-way interactions are shown. Main effects are shown in boldface to aid in finding them. Margins of error for the estimates (see text) are ±.56 in example II and ±.29 in example IV. Simultaneous margins of error are ±1.17 and ±.60, respectively.

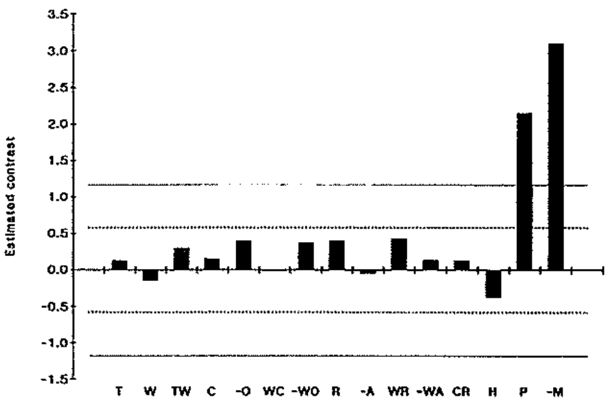


*Figure 1. Estimates of Contrasts in Example II.*

equal to 0. Then the $c_i$ are independent realizations of an $N(0, \tau^2)$ random variable $C$. Since the median of $|C|$ is about $.675\tau$, it follows that $s_0$ is a consistent estimate of 1.5 Med$|C| \approx 1.01\tau$. Further, since $\Pr(|C| > 2.5\tau) \approx .01$, PSE is roughly consistent for 1.5 times the .495th quantile of $|C|$, which is $1.5 \times .665\tau \approx \tau$.

Now suppose that there are just a few active contrasts among the $c_i$. If we knew which ones they were, we could exclude them and obtain a consistent estimate of $\tau$ as in the preceding paragraph. But we do not know exactly which contrasts are active. Marginally, the $c_i$ are independent realizations of a random variable $C$ whose distribution is a mixture of the form $(1 - \alpha)F + \alpha G$, where $F$ is $N(0, \tau^2)$, $G$ is some distribution more highly variable than $F$, and $\alpha$ is a contamination parameter. In this case, $s_0$ overestimates $\tau$, making it unlikely that an inactive contrast will exceed $2.5s_0$. We can expect the median in (2) to be based on essentially all of the inactive contrasts and possibly a few of the smaller active ones. Thus PSE will still overestimate $\tau$, but not by as much as $s_0$ does.

Note that computing $s_0$ is equivalent to the usual graphical procedure based on the half-normal plot, in the sense that the line connecting the origin and the coordinates of the median absolute contrast corresponds to the $N(0, s_0)$ distribution. PSE is in essence obtained by excluding the most obviously active effects, constructing a new half-normal plot, and superimposing a new line.

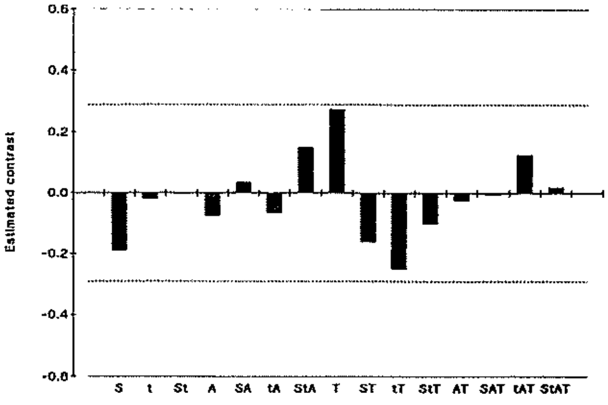


*Figure 2. Estimates of Contrasts in Example IV.*

Consider the case in which $G$ is $N(0, (k\tau)^2)$ with $k > 1$ (the model used by Box and Meyer). Note that $s_0$ converges stochastically to its limiting value, 1.5 Med$|C|$, as $m$ approaches $\infty$. The limiting value of PSE can be obtained from the median of the distribution of $|C|$, truncated at 1.5 Med$|C|$. These are easy to obtain numerically. For $0 \leq \alpha \leq .3$ and $k \geq 5$, it can be shown that the limiting values for $s_0$ lie between $1.01\tau$ and $1.60\tau$, but those for PSE range from $1.00\tau$ to $1.15\tau$, quite an improvement over $s_0$.

When $m$ is small, the asymptotic results may not hold. To resolve this, a small Monte Carlo study was conducted. The results are shown in Table 4. The design is a Latin square with three values of $k$ (5, 10, and 15), three values of $\alpha$ (.1, .2, and .3), and three values of $m$ (7, 15, and 31). Baseline cases with no active contrasts ($k = 1$ and/or $\alpha = 0$) for $m =$ 7, 15, and 31 are also included. These $k$ and $\alpha$ values were suggested by Box and Meyer (1986) as representative of the usual range of possibilities. In all cases, the value of $\tau$ is set at 1.00.

The number of Monte Carlo replications (sets of $m$ simulated contrasts) in each case depends on $m$. For $m = 7$, 15, and 31, there are 2,000, 1,000, and 500 replications, respectively. This makes the standard errors of the estimates approximately the same. (To compute the standard error of the estimated mean of $s_0$ or PSE, divide the corresponding standard deviation by $\sqrt{N}$, where $N$ is the number of replications.) Rather than generating data from contaminated normal distributions, it was deemed more appropriate to hold fixed the number of active contrasts in each case. For example, in the case $m = 31$, $k = 5$, and $\alpha = .2$, each of the 500 simulated sets of contrasts consists of 25 $N(0, 1)$ variates and 6 $N(0, 5^2)$ variates; the true $\alpha$ is $6/31 = .19$.

All computations were done on an IBM PC using a program written in C language, with double-precision (64-bit) reals. The polar method was used to obtain normal deviates from uniform pseudorandom numbers. Uniforms were generated using an exclusive–or mixture of a congruential and a shift-register generator with a word size of 32 bits.

Several consistent patterns are evident in Table 4. First, both asymptotically and in finite samples, $s_0$ is always larger than PSE (logically, this must happen). In the baseline cases, the distinction is small (as it should be), but when active contrasts are present, it is much more noticeable. Second, the expectations and limiting values of both $s_0$ and PSE all exceed 1.00, suggesting that the method is conservative (i.e., the true confidence coefficient is deflated due to over-

*Table 4. Monte Carlo Results for the Scale Estimates $s_0$ and PSE*

| | $\alpha \approx .1$ | | $\alpha \approx .2$ | | $\alpha \approx .3$ | |
|---|---|---|---|---|---|---|
| | $s_0$ | PSE | $s_0$ | PSE | $s_0$ | PSE |
| $k$ = 1 (baseline) | (0/7 × 2,000) | | (0/15 × 1,000) | | (0/31 × 500) | |
| Limit | 1.01 | 1.00 | 1.01 | 1.00 | 1.01 | 1.00 |
| MC mean | 1.06 | 1.05 | 1.03 | 1.01 | 1.03 | 1.01 |
| MC SD | .43 | .45 | .30 | .32 | .21 | .22 |
| $k$ = 5 | (2/15 × 1,000) | | (6/31 × 500) | | (2/7 × 2,000) | |
| Limit | 1.15 | 1.05 | 1.23 | 1.08 | 1.38 | 1.14 |
| MC mean | 1.17 | 1.06 | 1.26 | 1.11 | 1.42 | 1.23 |
| MC SD | .35 | .35 | .27 | .27 | .56 | .58 |
| $k$ = 10 | (1/7 × 2,000) | | (3/15 × 1,000) | | (9/31 × 500) | |
| Limit | 1.19 | 1.03 | 1.28 | 1.05 | 1.47 | 1.10 |
| MC mean | 1.22 | 1.08 | 1.31 | 1.08 | 1.48 | 1.11 |
| MC SD | .49 | .47 | .39 | .37 | .31 | .28 |
| $k$ = 15 | (3/31 × 500) | | (1/7 × 2,000) | | (4/15 × 1,000) | |
| Limit | 1.13 | 1.02 | 1.20 | 1.03 | 1.44 | 1.06 |
| MC mean | 1.13 | 1.01 | 1.25 | 1.09 | 1.47 | 1.17 |
| MC SD | .24 | .24 | .49 | .46 | .41 | .37 |

NOTE: Each simulated sample of $m$ contrasts consists of $[m(1 - \alpha) + .5]$ random numbers from the $N(0, 1)$ distribution and the rest from the $N(0, k^2)$ distribution. The notation $(a/m \times N)$ shows the number $a$ of contaminants (i.e., active contrasts), the number of contrasts $m$, and the number $N$ of Monte Carlo samples generated. Note that contaminants are impossible in the baseline cases ($k = 1$, $\alpha = 0$). Shown are the limiting values of $s_0$ and PSE (limit) and Monte Carlo estimates of the means (MC mean) and standard deviations (MC SD) of their sampling distributions.

estimating $\tau$). Third, for both $s_0$ and PSE, the finite-sample values exceed the limiting values (with one minor exception). The distinction is greater for smaller $m$.

The clearest interpretation of PSE is found by dividing its observed means by the asymptotic values (the standard errors of these ratios are all about .01). With $m = 7$, these ratios vary from 1.05 to 1.07, with $m = 15$, they fall between 1.01 and 1.03, and with $m = 31$, they fall between .99 and 1.03. So it appears that with $m = 15$ or higher the expected value of PSE is reasonably approximated by its limiting value.

To get an idea of the appropriate degrees of freedom, the empirical distributions of $PSE^2$ were fitted by scaled chi-squared distributions by matching the first two moments (the fits are quite good). The fitted degrees of freedom in the baseline cases are 2.8, 5.4, and 10.4 for $m = 7$, 15, and 31—suggesting the rule that $d = m/3$ is about right for the number of degrees of freedom. The fitted degrees of freedom in the nonbaseline cases tend to be somewhat smaller (but not drastically so). The lowest fitted degrees of freedom occur in those cases with the highest limiting PSE. Using a few too many degrees of freedom will not overbalance this conservatism, so it is recommended that $d = m/3$ df be used regardless of the number of active contrasts detected.

[*Received June 1988. Revised June 1989.*]

## REFERENCES


Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: John Wiley.

Box, G. E. P., and Meyer, R. D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11–18.

Daniel, C. (1959), "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments," *Technometrics*, 1, 311–341.

Davies, O. L. (ed.) (1954), *The Design and Analysis of Industrial Experiments*, London: Oliver & Boyd.

Ott, E. R. (1967), "Analysis of Means—a Graphical Procedure," *Industrial Quality Control*, 24, 101–109.

Taguchi, G., and Wu, Y. (1980), *Introduction to Off-Line Quality Control*, Nagoya, Japan: Central Japan Quality Control Association.