*Center for Quality and Productivity Improvement*
UNIVERSITY OF WISCONSIN
610 Walnut Street
Madison, Wisconsin 53705

(608) 263–2520
(608) 263–1425 FAX

*Report No. 78*

# What Can You Find Out From Eight Experimental Runs?* and What Can You Find Out From Sixteen Experimental Runs?**

George Box

*February 1992*

* appeared in: *Quality Engineering* 1992, Vol. 4, No. 4, pp. 619-627.

** appeared in: *Quality Engineering* 1992, Vol. 5, No. 11, pp. 167-178.

*C Q P I*
UNIVERSITY OF WISCONSIN

Report No. 78

**Report Series**
*in Quality and Productivity*

# What Can You Find Out From Eight Experimental Runs?

George Box

Center for Quality and Productivity Improvement

*University of Wisconsin*
*Madison, Wisconsin*

*ABSTRACT*

Different ways to use $n = 8$ experimental runs are described to generate two-level factorial and fractional factorial designs for studying up to $n - 1$ factors. The roles of "aliases" and of design "resolution" are discussed and the rationales for the employment of designs with different degrees of fractionation are presented.

*KEYWORDS: Experimental design, factorials, fractional factorials*

This research was sponsored by the National Science Foundation under Grant No. DDM-8808138.

# What Can You Find Out From Eight Experimental Runs?

George Box

*Different ways to use n = 8 experimental runs are described to generate two-level factorial and fractional factorial designs for studying up to n − 1 factors. The roles of "aliases" and of design "resolution" are discussed and the rationales for the employment of designs with different degrees of fractionation are presented.*

In my first column (1) I mentioned how Christer Hellstrand of SKF had used a simple $2^3$ factorial design to lengthen the life of a particular type of bearing by a factor of five. Figure 1(a) shows a second example (2) of Christer's work in which a different eight run design was used to solve an important problem. A contract for manufacturing a specialized bearing had been lost because a competitor had brought onto the market an improved design. In response, Christer's company set up a project team consisting of the technical director, the manufacturing engineer, the manufacturing statistician, the material laboratory manager, an application engineer and an endurance testing engineer. From careful analysis of the competitors product and after a long brainstorming session, the team decided to test four factors at standard (−) and modified (+) levels. The factors were: the manufacturing process for the balls ($A$), the cage design ($B$), the type of grease ($C$), and the amount of grease ($D$).

Now *four* factors were tested at two levels in this experiment, so a full $2^4$ factorial design would have required sixteen runs, but the experimental design used only a half that number. It was in fact a "half replicate" of the full design. The eight chosen manufacturing conditions are shown in Figure 1(a). They were run in random order and a measure of the average life $y$ of bearings produced at each set of conditions is shown to the right hand side of the table. Mere eyeballing of the data suggests that the type of grease ($C$) and more particularly the amount of grease ($D$) do not appear to be doing much. If we ignore $D$ as unimportant, the data can then be plotted in relation to the other three factors $A, B, C$ on the cube shown in Figure 1(b). Study of the figure suggests that by appropriately modifying the manufacturing process for the balls ($A$) and changing the design of the cage ($B$), a seven-fold increase in bearing life could be obtained as compared with the standard unmodified bearing represented by run one. These findings were doubly welcome because the modified cage design was cheaper to produce than the original design. Further tests confirmed these findings, and as a result of this simple experiment, the lost market was regained with a greatly improved bearing design which exceeded the expectations of the customer.

Now while industrial experiments such as this can often produce dramatic improvements, they are frequently expensive; particularly, as in this case, when the runs are made on actual production equipment. Statistical designs which use only a fraction of the factorial runs are, therefore, of great practical importance.

## SOME DIFFERENT WAYS TO USE EIGHT RUNS

To understand why, for particular applications, *fractional* factorial designs are so efficient, let's consider other ways in which these eight process runs might have been used. Well, they could have been used to test just a single factor, say type of grease ($C$). Four runs might have been made with standard grease and four runs with modified grease keeping the other three factors constant. Such an experiment would have the merit that tests for both the standard and modified types of grease would have been replicated (repeated) four times over and we might, therefore, feel some confidence in any conclusions that we drew since they would be based on a comparison between *averages* of two sets of four runs. We could call such an arrangement a $2^{1+2}$ design because it would test a single factor at two levels with $4 = 2^2$ replications. The trouble with this arrangement would be that we would have used up all eight runs to study only one factor. Also, even if we could get agreement to run

(a)

| | | − | + |
|---|---|---|---|
| *A:* | Balls | Standard | Modified |
| *B:* | Cage | Standard | Modified |
| *C:* | Type of grease | Standard | Modified |
| *D:* | Amount of grease | Normal | Large |

| Run Number | *A* | *B* | *C* | *D* | *y* | *AB* = *CD* |
|---|---|---|---|---|---|---|
| 1 | − | − | − | − | 0.31 | + |
| 2 | + | − | − | + | 1.38 | − |
| 3 | − | + | − | + | 0.73 | − |
| 4 | + | + | − | − | 2.17 | + |
| 5 | − | − | + | + | 0.95 | + |
| 6 | + | − | + | − | 1.37 | − |
| 7 | − | + | + | − | 0.92 | − |
| 8 | + | + | + | + | 2.57 | + |

(b)

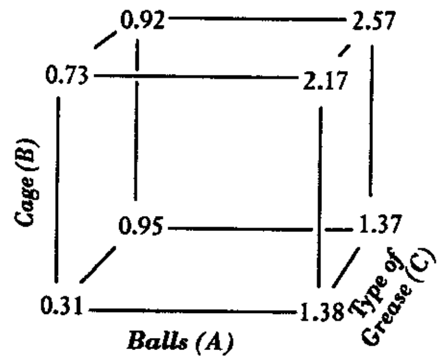



**Figure 1:** *(a) The eight-run design to test four factors at two levels. (b) Display of the data assuming factor D (Amount of Grease) has little effect.*

three further sets of eight runs to test the other three factors in the same "one at a time" fashion, in the end we would still know nothing about the *interactions* between the factors. An alternative way of using the eight runs would have been to study just two of the factors in a $2^2$ factorial replicating each set of the four experimental conditions twice over. Using the same notation as before, this arrangement would be a $2^2 \times 2$ or $2^{2+1}$ design. A third possibility would have been to run three of the factors in a "single replicate" of a $2^3$ factorial. Notice that inclusion of the additional factors would *not* result in any loss in precision; because, whether we used the eight runs to include one, two or three factors, in the end we would still be comparing an average of four runs with an average of four runs to calculate the effects of each of the factors and their interactions. So we can include the extra two factors in the eight-runs "for free" and determine their interactions as well!

Now in the bearing experiment, this process was taken one step further and *four* factors were included in the eight-run experiment. Designs of this kind were invented by Finney (3) and called by him *fractional replicates.* In this case the design is a "half replicate" or "half fraction" of the full $2^4$ design. It is in fact $2^4 \times \frac{1}{2}$ or $2^4 \times 2^{-1}$ that is a $2^{4-1}$ "fractional replicate". As we shall see, quarter replicates, eighth replicates and so forth can in certain circumstances also be useful. These designs are particular cases of *orthogonal arrays* developed by Plackett and Burman (4) and Rao (5) in Britain during and just after World War II.

## FRACTIONAL DESIGNS

Fractional factorial designs have been used with great success in solving industrial problems particularly when the objective is "factor screening". That is to say we are in a 'Pareto situation' – where, as Dr. Juran puts it, we are "looking for the vital few [factors] among the trivial many." These designs are of value for screening purposes because of what may be called their "projective" properties. For example, if you check it out, you will find that *any* three columns of the design in the Figure 1(a) produces a complete set of the eight combinations (– – –, + – –, – + –, + + –, – – +, + – +, – + +, +++) for whichever three factors you've chosen. So for this design if, as is quite likely, one of the factors is just 'in there for the ride,' – that's to say it is one of the "trivial many" and has zero or little effect – *then you will have a complete $2^3$ design in the other three.*

A design, any *three* columns of which will produce a complete $2^3$ factorial (or a replicated $2^3$ factorial), is called a design of *resolution* four. To keep the notation straight we usually use a roman numeral subscript to indicate resolution. Thus, the bearing design used by Christer is a one-half fraction of a $2^4$ design of resolution four or a $2^{4-1}_{IV}$ design. In general a design of resolution *R* produces complete factorials (sometimes replicated) in every set of $R-1$ factors. I'll come back to this idea in a bit.

## ALIASES

To better understand the $2^{4-1}_{IV}$ design used in the bearing experiment, consider what would have happened if all the factors had turned out to be important, and you therefore needed to consider all four main effects $A, B, C, D$ and all six two-factor interactions $AB, AC, AD, BC, BD$, and $CD$*. Now let's try to calculate the interaction $AB$ for this experiment. The appropriate column of signs will be obtained by multiplying the signs of columns *A* and *B* in Figure 1(a) to get a new column *AB* shown on the right of the table. However, notice that if we do the same thing for columns *C* and *D*, you get an identical set of signs. Thus, when we try to calculate the *AB* interaction by subtracting the average of the four runs opposite minus signs from the average of the four runs opposite plus signs, the effect we get is really the *sum* of the *AB* interaction and a *CD* interaction. *AB* is said to have an 'alias' *CD*. If you multiply out the signs you'll also find that *AC* and *BD* are aliased and *AD* and *BC* also. So if you calculated the main effects of the factors and their interactions, the analysis would look like this:

| | | |
|---|---|---|
| Average of the 8 runs | | 1.30 |
| *Main Effects* | A (Balls) | 1.14 |
| | B (Cage) | 0.60 |
| | C (Grease type) | (0.31) |
| | D (Grease amount) | 0.22 |
| *Interactions* | (AB) + CD | (0.40) |
| | AC + BD | –0.11 |
| | AD + BC | –0.01 |

From the above alias table you will see that the main effects are clear of two factor interactions but the two factor interactions are aliased with each other. This is true for *any* resolution *IV* design. You can use this same table to check out what I said before about the "projective property" of this design. If *any* one of the factors *A*, *B*, *C*, or *D* is unimportant (and therefore any effects involving that letter can be left out) the remaining effects have no aliases. For example, cross out all the effects containing, say, the letter *D*, as would be appropriate if factor *D* was inert. Then the interaction pairs are (AB + ~~CD~~), (AC + ~~BD~~) and (~~AD~~ + BC) and now none of the effects are aliased because you have a complete $2^3$ factorial in the other factors *A*, *B*, and *C*. For the bearing data, the guess is that what is going on is mostly due to balls (*A*) and cage (*B*) with the possibility of a helpful *AB* interaction and of a small effect due to grease type (*C*) indicated by brackets in the table.

But in this example, no formal analysis was really necessary. As is so often the case for well designed experiments, the data almost analyze themselves and it was realized that the problem had been solved as soon as the results shown in Figures 1(a) and 1(b) were looked at. Test runs at the new conditions confirmed this.

Note that even if the original plan had been to run a full $2^4$ design, nothing would have been lost by running these eight runs first. If they failed to solve the problem, you could always add some further runs later and, in particular, you could run the other half of the full $2^4$ arrangement*. This is a

* I will suppose, as is often realistic, that interactions between three or more factors can be ignored.

* The two halves of the design form what are called "orthogonal blocks". Because of this, It turns out that if something slipped in the time which elapsed between performing the first set and the second set of eight runs resulting in a change of mean level, this discrepancy would be exactly balanced out and would cause no change in the values of the calculated effects (see $BH^2$ pp. 336-351 for details).

further illustration of the sequential approach: eliminating unnecessary experimentation by ensuring that larger designs are built up only to solve more complicated problems.

## GREATER FRACTIONATION

The discussion so far, in no way exhausts the possibilities for useful eight-run designs. The easiest way to see what is available is to simply write down a table of plus and minus signs that can be used to estimate all the main effects and interactions in a full $2^3$ factorial design in which (dummy) factors are denoted by (say) *a*, *b*, and *c* as is done in Figure 2.

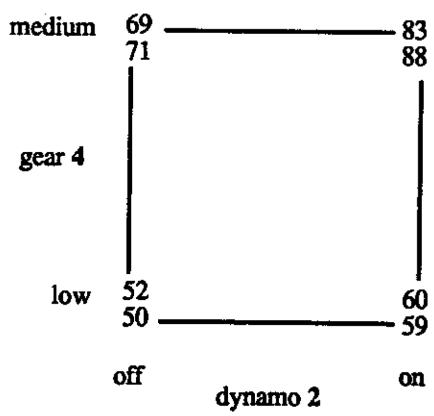
Now forget about the *a*'s, *b*'s, and *c*'s (that were just dummy variables used to generate this table of signs) and let us rename the columns 1, 2, 3, 4, 5, 6, and 7 as indicated in the table. What you have now is sometimes called an $L_8$ *orthogonal array*. An array is just a name for a rectangular table. The word *orthogonal* means that there is complete balance in the signs of every pair of columns. Thus, if you select *any one* of the seven columns, you will find that opposite the four pluses in the selected column there are two pluses and two minuses in every other column. Likewise, opposite the four minuses in the selected column there will be two pluses and two minuses in every other column. This array encapsulates every possible fractional factorial design involving eight runs. Although you may see tables for such designs which look different–with ones and twos replacing minus signs and plus signs, or with different orderings of columns and rows, or in which all the signs are switched in one or more of the columns, these are not really different designs but simply correspond to a relabeling of the basic design which we generated above.

Now beneath the columns of pluses and minuses in Figure 2 there are a number of rows of bullets. The first row of three bullets indicates that if you want to use the array as a full $2^3$ factorial, you could use columns 1, 2 and 3. Looking at the next row, if you want to use the eight runs for four factors to produce a $2^{4-1}_{IV}$ design you should add column 7 (use columns 1, 2, 3 and 7). This design was in fact employed for the bearing example discussed earlier. Although you could generate a $2^{4-1}$ fractional factorial design with a different column for the fourth variable, using say columns 1, 2, 3, and 4, this design would only have resolution three rather than resolution four. Normally you will want to use designs which, for a given degree of fractionation, produce the highest possible resolution, so you would use columns 1, 2, and 3 with column 7.

If you "saturate" the orthogonal array by associating factors with every one of the columns, you will have a $\frac{1}{16} = \frac{1}{2^4}$ replicate of a $2^7$ design (because you are only running 8 of the $2^7 = 128$ runs required by the full factorial). This design has resolution III, that is to say any pair of columns yields a full $2^2$ factorial (replicated twice). It is therefore called a $2^{7-4}_{III}$ design. Thus you could use this design to screen seven factors if, initially at least, you thought it likely that the vital few factors of importance were not more than two in number.

An example taken from "$BH^2$" (7) concerns the times for a student cyclist to climb a particular hill on the campus at the University of Wisconsin-Madison. The seven factors: seat (up/down), dynamo (on/off), handlebars (up/down), gear (low/normal), raincoat (on/off), breakfast (yes/no) and tires (hard/soft) were associated with columns 1, 2, 3, 4, 5, 6, and 7 respectively and varied according to the design in Figure 2. The calculated effects, showing all the two factor interaction aliases that could occur in the unlikely event that all seven factors were active, are also shown.

| | | | |
|---|---|---|---|
| seat | 3.5 | ⟶ | 1 + 24 + 35 + 67 |
| dynamo | (12.0) | ⟶ | 2 + 14 + 36 + 57 |
| handlebars | 1.0 | ⟶ | 3 + 15 + 26 + 47 |
| gear | (22.5) | ⟶ | 4 + 12 + 56 + 37 |
| raincoat | 0.5 | ⟶ | 5 + 13 + 46 + 27 |
| breakfast | 1.0 | ⟶ | 6 + 23 + 45 + 17 |
| tires | 2.5 | ⟶ | 7 + 34 + 25 + 16 |




The simplest explanation of the data is obviously that the main effects of dynamo (2) and gear (4) are the vital factors that affect the time to cycle up the hill. The appropriate two-way table which shows how the data would stack up when rearranged on this basis is shown below the main table. In the original reference there is a discussion of how such tentative findings can be confirmed.

## DROPPING COLUMNS. WHICH ONES?

There are, of course, intermediate cases where we wish to look at only six or five factors. Such designs are obtained by dropping one or two of the columns 4, 5 and 6 from the basic seven factor design of Figure 2. The appropriate alias pattern is found by simply omitting all the effects which contain the dropped factors.

The approach to fractional factorial designs and orthogonal arrays that I have outlined in this column was originally presented about forty years ago in a paper entitled "On the Experimental Attainment of Optimum Conditions" (6). I will use an example from that paper to illustrate the idea of "dropping" factors. In the above reference, a design was employed to investigate a chemical reaction involving three chemicals, *A*, *B* and *C* and a solvent. Five factors were varied: the concentration of *C*, the proportion of *C* to *A*, the solvent amount, the proportion of *B* to *A* and the reaction time. The design actually used can be obtained by dropping columns four and five from the seven factor design in Figure 2 and associating these five experimental factors with columns 1, 2, 3, 6 and 7 respectively. In this example it was argued that in the early stages of investigation when you are likely to be a long way from optimal conditions, main effects will probably dominate and interactions will be relatively small (when you are way down the side of a hill, it is the slopes of the hill that are important, curvature effects and interaction effects corresponding to ridges in the hill are not so important until you get near the summit). This quarter replicate ($2^{5-2}$ design) was employed, therefore, on the expectation that main effects would dominate. However, a chemist thought that if any interaction did turn out to be important it would most likely be between the factors *concentration of C* and *solvent amount*. The particular allocation of factors to the basic array in Figure 2 ensures that this interaction (13) occurs in column 5 which has not been used to accommodate a real factor and can thus be independently estimated. Once you have the basic alias pattern for seven factors, given in the earlier table, it is easy to see how by simple trial and error you can associate particular process factors with particular columns of the design

| *(a)* | *(b)* | *(c)* | *(ab)* | *(ac)* | *(bc)* | *(abc)* | **Bicycle Data** using all 7 columns *Time in Seconds* | **Process Data** using columns 1, 2, 3, 6, 7 *Yield %* |
|---|---|---|---|---|---|---|---|---|
| − | − | − | + | + | + | − | 69 | 77.1 |
| + | − | − | − | − | + | + | 52 | 69.0 |
| − | + | − | − | + | − | + | 60 | 75.5 |
| + | + | − | + | − | − | − | 83 | 72.6 |
| − | − | + | + | − | − | + | 71 | 67.9 |
| + | − | + | − | + | − | − | 50 | 68.4 |
| − | + | + | − | − | + | − | 59 | 71.5 |
| + | + | + | + | + | + | + | 88 | 65.9 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| • | • | • | | | | | ⟶ | $2^3$ |
| • | • | • | | | | • | ⟶ | $2^{4-1}_{IV}$ |
| • | • | • | • | • | • | • | ⟶ | $2^{7-4}_{III}$ |

| *Basic eight run designs* | | |
|---|---|---|
| | $2^3$ | Full Factorial |
| | $2^{4-1}_{IV}$ | 1 rep. of $2^3$ in every three factors (4 choices) |
| | $2^{7-4}_{III}$ | 2 reps. of $2^2$ for every two factors (21 choices) |

**Figure 2:** *An eight run orthogonal array showing the basic designs* $2^3$, $2^{4-1}_{IV}$ *and* $2^{7-4}_{III}$ *with two sets of data.*

so as to isolate suspected interactions in this way.

For the chemical example, the analysis and the alias pattern obtained by dropping all effects containing the numbers 4 and 5 from the full $2^7$ alias pattern are as follows:

| | | | | |
|---|---|---|---|---|
| (1) | Concentration $C$ | −4.0 | → | 1 + 67 |
| (2) | $C/A$ | 0.8 | → | 2 + 36 |
| (3) | Solvent | −5.1 | → | 3 + 26 |
| | | −0.2 | → | 12 + 37 |
| (13) | Conc. × Solvent | 1.5 | → | 13 + 27 |
| (6) | $B/A$ | −0.2 | → | 6 + 23 + 17 |
| (7) | Time | −2.8 | → | 7 + 16 |

It appears that factors 1, 3 and 7 all have large *negative* effects. It was therefore concluded that higher yields might be obtained by moving in a direction such that the concentration of $C$, the amount of solvent and the time of reaction all were *reduced*. This checked out and a yield of 84% was eventually obtained.

In this column I have shown a number of ways in which eight-run fractional factorial designs can be useful and I mentioned a number of rationales for their use. I will discuss these ideas in greater detail in my next column where I will talk about sixteen-run designs.

## REFERENCES


1. Box, G.E.P., *Do Interactions Matter?* Quality Engineering. 2(3), 365-369, (1989).

2. Hellstrand, C., This example is from: *"Experiences in Applying Experimental Designs at SKF"*, presented at the 19th European Meeting of Statistics, Barcelona, September 1991. Used by permission.

3. Finney, D.J., The fractional replication of factorial arrangements. *Ann Eug.* **12**, 291-301, (1945).

4. Plackett, R.L., and Burman, J.P., Design of optimal multifactorial experiments. *Biometrika*, **23**, 305-325, (1946).

5. Rao, C.R., Factorial experiments derivable from combinatorial arrangements of arrays. *J R. statist. Soc.*, **B9**, 128-140, (1947).

6. Box, G.E.P., and Wilson K.B., On the experimental attainment of optimum conditions. *J. R. statist. Soc.*, **B13**, 1-45, (1951).

7. The term $BH^2$ is used to refer to the book by G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for experimenters: an introduction to design, data analysis, and model building*. New York: Wiley, (1978).