

# IE 361 Module 22

## Design and Analysis of Experiments: Part 3

Prof.s Stephen B. Vardeman and Max D. Morris

Reading: Section 6.3, *Statistical Quality Assurance Methods for Engineers*

In this module, we consider the analysis of  $p$ -way factorial studies, and in particular ones where each of the  $p$  factors has only 2 levels. These are the  $2 \times 2 \times \cdots \times 2$  or  $2^p$  studies.

## Basic $2^p$ Notation

We wish now to think about experimentation and subsequent analysis for systems that have many ( $p$ ) factors potentially affecting a response,  $y$ . We begin with full factorial studies (where all combinations of some levels of these factors are represented in the data set) and concentrate on  $2^p$  studies for two reasons. The first is that there is some special notation and structure that make their

analysis most transparent, and the second is that as a practical matter, one can rarely afford  $p$ -factor factorial experimentation with many levels of each factor.

**Example 22-1** As our motivating example, we will use data from a  $2^3$  chemical process pilot plant study taken from *Statistics for Experimenters* by Box, Hunter, and Hunter. The response of interest was a yield variable ( $y$  in units of g). Factors and levels were as in the following table.

Factor	"Low" (—) and "High" Levels
A-Temperature	160 °C vs 180 °C
B-Concentration	20% vs 40%
C-Catalyst	#1 vs #2

Note that it is typical in  $2^p$  studies to make an arbitrary designation of one level of each factor a first or "low" level and the other level the second or "high"

level. In the pilot plant study, there were  $m = 2$  runs of the pilot plant made at each combination of levels of these 3 factors. We'll let

$\bar{y}_{ijk}$  = the sample mean yield at level  $i$  of A, level  $j$  of B, and level  $k$  of C  
and

$s_{ijk}$  = the sample standard deviation of yield at  
level  $i$  of A, level  $j$  of B, and level  $k$  of C

The catalyst data and some summary statistics are then given in the table below along with some additional notation for this  $2^3$  factorial context.

A	B	C	$2^p$ name	$i$	$j$	$k$	$y$ 's	$\bar{y}_{ijk}$	$s_{ijk}^2$
—	—	—	(1)	1	1	1	59,61	60	2
+	—	—	a	2	1	1	74,70	72	8
—	+	—	b	1	2	1	50,58	54	32
+	+	—	ab	2	2	1	69,67	68	2
—	—	+	c	1	1	2	50,54	52	8
+	—	+	ac	2	1	2	81,85	83	8
—	+	+	bc	1	2	2	46,44	45	2
+	+	+	abc	2	2	2	79,81	80	2

While the " $ijk$ " notation is perfectly general and could be applied to any  $I \times J \times K$  factorial, the  $+/-$  notation used in the table and the special " $2^p$  name" convention (that names a combination of levels of the 3 factors by those factors appearing at their second or high level) are special to the case where every factor has only 2 levels. The " $ijk$ " notation is helpful when one needs to indicate various averages of the sample means. For example,  $\bar{y}_{i..}$  is the

average of all sample means for level  $i$  of Factor A,  $\bar{y}_{.jk}$  is the average of all sample means for level  $j$  of Factor B and level  $k$  of Factor C, etc.

It is helpful for picturing the results of a  $2^3$  factorial study to plot the sample means obtained on the corners of a cube as shown in Figure 1.

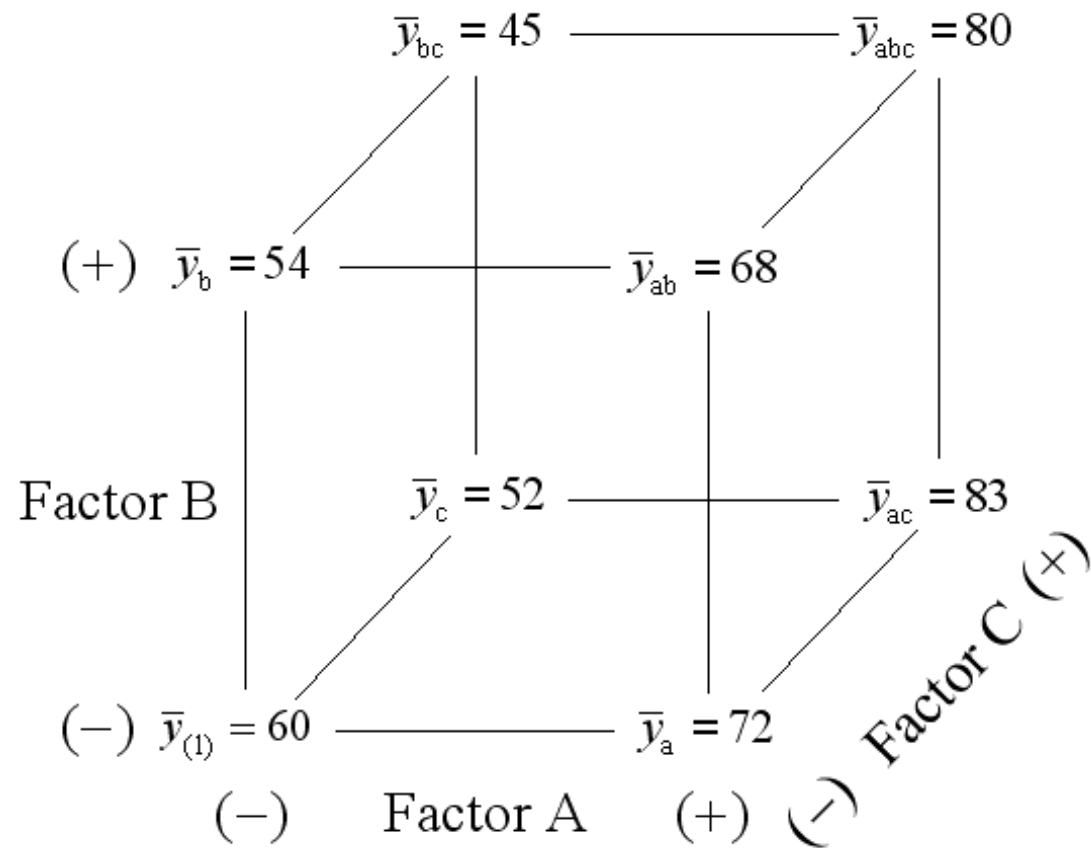


Figure 1: The 2<sup>3</sup> Sample Mean Yields (g) in the Pilot Plant Study

## Defining and Computing 3- (and by Analogy Higher- ) Way Fitted Effects

Just as it was helpful to define fitted effects in 2-way factorials, it is useful to make similar definitions for 3- and higher-way studies. we begin (as we did for 2-way studies) with **fitted main effects**. These are

$$\begin{aligned} a_i &= \bar{y}_{i..} - \bar{y}_{...} \\ &= (\text{the Factor A level } i \text{ average } \bar{y}) - (\text{the grand average } \bar{y}) \\ &= \text{the (fitted) main effect of the } i\text{th level of Factor A} \end{aligned}$$

and

$$\begin{aligned} b_j &= \bar{y}_{.j.} - \bar{y}_{...} \\ &= (\text{the Factor B level } j \text{ average } \bar{y}) - (\text{the grand average } \bar{y}) \\ &= \text{the (fitted) main effect of the } j\text{th level of Factor B} \end{aligned}$$

and

$$\begin{aligned}c_k &= \bar{y}_{..k} - \bar{y}_{...} \\ &= (\text{the Factor C level } k \text{ average } \bar{y}) - (\text{the grand average } \bar{y}) \\ &= \text{the (fitted) main effect of the } k\text{th level of Factor C}\end{aligned}$$

These are differences between "face" average means and the grand average mean on a plot like Figure 1. For example, as portrayed on Figure 1,

$$a_2 = (\text{the right face average } \bar{y}) - (\text{the grand average } \bar{y})$$

**Example 22-1 continued** In the pilot plant study, it is straightforward to see that

$$a_1 = 52.75 - 64.25 = -11.5 \quad \text{and} \quad a_2 = 75.75 - 64.25 = 11.5$$

and

$$b_1 = 66.75 - 64.25 = 2.5 \quad \text{and} \quad b_2 = 61.75 - 64.25 = -2.5$$

and

$$c_1 = 63.5 - 64.25 = -.75 \quad \text{and} \quad c_2 = 65.0 - 64.25 = .75$$

The relative sizes of the A,B, and C fitted main effects quantify what is to some extent already obvious in Figure 1: The left-to-right differences in means on the corners of the cube are bigger than the top-to-bottom or front-to-back differences.

Just as in 2-way studies, the fitted main effects for any factor add to 0 over levels of that factor (so that the fitted main effect for one level is just minus one times that for the other level).

**Fitted 2 factor interactions** in a 3-way factorial can be thought of in at least two different ways. In the first place, they are measures of lack of parallelism that would be appropriate after averaging out over levels of the 3rd factor.

Another is that they represent the what can be explained about a response mean if one thinks of factors acting jointly in pairs beyond what is explainable in terms of them acting separately. The definitions of these look exactly like the definitions from two-factor studies, except that an extra dot appears on each  $\bar{y}$ . That is

$$ab_{ij} = \bar{y}_{ij.} - (\bar{y}_{...} + a_i + b_j)$$

and

$$ac_{ik} = \bar{y}_{i.k} - (\bar{y}_{..} + a_i + c_k)$$

and

$$bc_{jk} = \bar{y}_{.jk} - (\bar{y}_{...} + b_j + c_k)$$

It is a consequence of these definitions that fitted 2-factor interactions add to 0 over levels of any one of the factors involved. In a  $2^p$  study, this allows one

to compute a single one of these fitted interactions of a given type and obtain the other three by simple sign changes.

**Example 22-1 continued** In the pilot plant study,

$$\begin{aligned} ab_{11} &= \bar{y}_{11.} - (\bar{y}_{...} + a_1 + b_1) \\ &= 56 - (64.25 + (-11.5) + 2.5) = .75 \end{aligned}$$

so that

$$ab_{12} = -.75 \quad \text{and} \quad ab_{21} = -.75 \quad \text{and} \quad ab_{22} = .75$$

Similarly,

$$\begin{aligned} ac_{11} &= \bar{y}_{1.1} - (\bar{y}_{...} + a_1 + c_1) \\ &= 57 - (64.25 + (-11.5) + (-.75)) = 5.0 \end{aligned}$$

so that

$$ac_{12} = -5.0 \quad \text{and} \quad ac_{21} = -5.0 \quad \text{and} \quad ac_{22} = 5.0$$

And finally,

$$\begin{aligned}bc_{11} &= \bar{y}_{.11} - (\bar{y}_{...} + b_1 + c_1) \\ &= 66 - (64.25 + (2.5) + (-.75)) = 0\end{aligned}$$

so that

$$bc_{12} = 0 \quad \text{and} \quad bc_{21} = 0 \quad \text{and} \quad bc_{22} = 0$$

The last of these says, for example, that after averaging over levels of Factor A, there would be perfect parallelism on an interaction plot of means  $\bar{y}_{.jk}$ . On the other hand, the fairly large size of the AC two-factor interactions is consistent with the clear lack of parallelism on Figure 2, that is an interaction plot of averages of  $\bar{y}$ 's top to bottom (over levels of Factor B) from Figure 1.

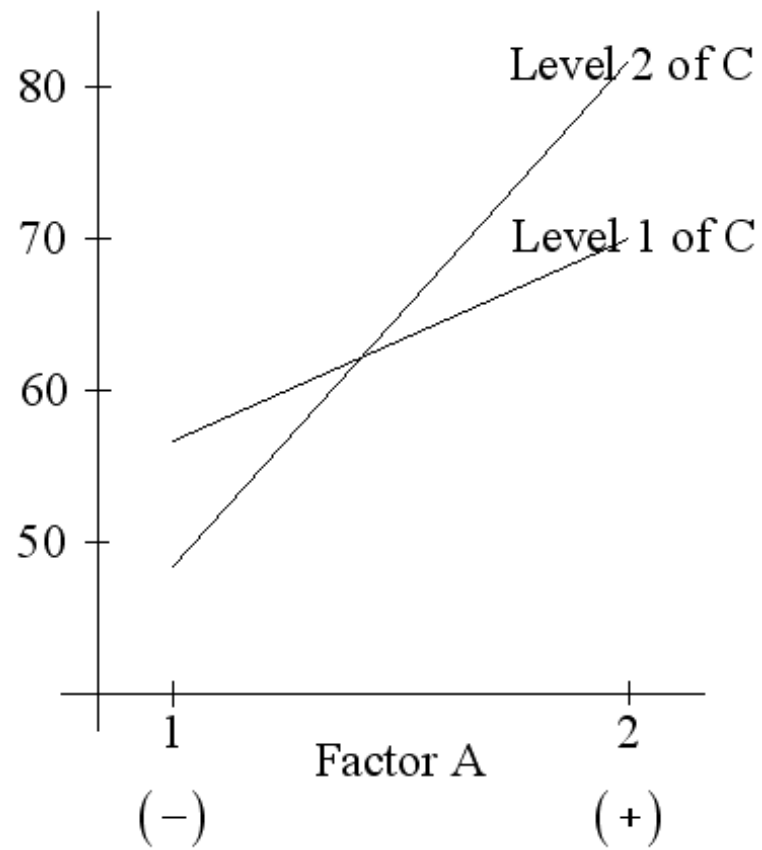


Figure 2: Interaction Plot for Pilot Plant Study After Averaging over Levels of B (Showing Strong AC Two-Factor Interaction/Lack of Parallelism)

Main effects and two-way interactions are enough to describe a two-way factorial set of means. But more is needed to describe a 3-way factorial. **Fitted 3 factor interactions** in a three way study are defined as

$$abc_{ijk} = \bar{y}_{ijk} - \left( \bar{y}_{...} + a_i + b_j + c_k + ab_{ij} + ac_{ik} + bc_{jk} \right)$$

These measure the difference between what's observed and what's explainable in terms of factors acting separately and in pairs on the response,  $y$ . They are also the difference between what one would call 2 factor interactions between, say, A and B, looking separately at the various levels of C (so that they are 0 exactly when the pattern of AB two factor interaction is the same for all levels of C).

These 3 factor interactions sum to 0 over all levels of any of the factors A, B, and C, so for a  $2^p$  factorial one may compute one of these and get the others by appropriate sign changes.

**Example 22-1 continued** In the pilot plant study,

$$\begin{aligned} abc_{111} &= \bar{y}_{111} - (\bar{y}_{...} + a_1 + b_1 + c_1 + ab_{11} + ac_{11} + bc_{11}) \\ &= 60 - (64.25 + (-11.5) + 2.5 + (-.75) + .75 + 5.0 + 0) = -.25 \end{aligned}$$

so that

$$\begin{aligned} abc_{112} &= .25, abc_{121} = .25, abc_{211} = .25, \\ abc_{122} &= -.25, abc_{212} = -.25, abc_{221} = -.25, \text{ and } abc_{222} = .25 \end{aligned}$$

Happily, these (relatively difficult to interpret, at least as compared to main effects and 2 factor interactions) fitted effects are small in comparison to the lower order fitted effects.

In  $2^p$  studies, since by computing one fitted effect of each type, one has (via simple sign changes) all of the fitted effects, it is common to call those for the "all factors at their high level" combination

$$a_2, b_2, ab_{22}, c_2, ac_{22}, bc_{22}, abc_{222}, \text{ etc.}$$

*the* fitted effects. And it turns out that there is a very efficient method of computing these all at once. This is the **Yates algorithm**. It consists of writing combinations and corresponding means down in "Yates standard order" and then doing  $p$  cycles of additions and subtractions of pairs of values, followed by division by  $2^p$ .

**Example 22-1 continued** The following table illustrates the use of the Yates algorithm on the 3-factor pilot plant data. (There are thus  $p = 3$  cycles illustrated below.)

combination	$\bar{y}$	cycle 1	cycle 2	cycle 3	cycle 3 $\div 2^3$
(1)	60	132	254	514	64.25 = $\bar{y}...$
a	72	122	260	92	11.5 = $a_2$
b	54	135	26	-20	-2.5 = $b_2$
ab	68	125	66	6	.75 = $ab_{22}$
c	52	12	-10	6	.75 = $c_2$
ac	83	14	-10	40	5.0 = $ac_{22}$
bc	45	31	2	0	0 = $bc_{22}$
abc	80	35	4	2	.25 = $abc_{222}$

The final column of the Yates calculations summarizes very nicely the whole set of computations we have done thus far in this example. The biggest (in magnitude) fitted effects here are the A main effects, AC two factor interactions, and the somewhat smaller B main effects. A rough interpretation of this might be something like:

The most important determiner of yield is temperature (Factor A).

The impact of changing catalyst (level of Factor C) depends upon which temperature is being used. The effect of concentration (Factor B) is smaller and seems to be independent of temperature and catalyst.

After doing some calculations of fitted effects, a useful exercise is to produce fitted or predicted values corresponding to a model for response that only include the largest of the fitted effects. This can be done either using simple addition or by employing a "reverse" version of the Yates algorithm. The latter of these methods is illustrated on pages 318-320 of *SQAME*. The former is illustrated below.

**Example 22-1 continued** Suppose that one adopts an "A and B main effects, plus AC interactions" description of the pilot plant study. Then fitted or

predicted values of  $y$  are

$$\hat{y}_{ijk} = \bar{y}_{...} + a_i + b_j + ac_{ik}$$

For example, the "all low" combination of levels of Factors A,B, and C has fitted response

$$\begin{aligned}\hat{y}_{111} &= \bar{y}_{...} + a_1 + b_1 + ac_{11} \\ &= 64.25 + (-11.5) + 2.5 + 5.0 = 60.25\end{aligned}$$

# Judging the Statistical Detectability/"Significance" of $2^p$ Factorial Effects

Having defined and computed fitted effects for a  $2^p$  factorial, one needs ways of judging whether what has been computed is really anything more than background noise/experimental variation. The most reliable method of doing this is based on the same kind of inference we employed in the two-factor case in Module 21. That is, each fitted effect produced by the Yates algorithm is a linear combination of sample means, an  $\hat{L}$  in the notation of Module 20. Thus one can attach "margins of error" to such fitted effects using the basic method for estimating a linear combination of means ( $L$ 's) presented in Module 20. And the general formula from Module 20 takes a particularly simple form in the case of fitted effects from a  $2^p$  study.

That is, corresponding to each  $2^p$  factorial fitted effect,  $\hat{E}$ , is a theoretical/population effect. Confidence limits for the theoretical effect are then

$$\hat{E} \pm t_{s_{\text{pooled}}} \frac{1}{2^p} \sqrt{\frac{1}{n_{(1)}} + \frac{1}{n_a} + \frac{1}{n_b} + \frac{1}{n_{ab}} + \dots}$$

In the case that the data are balanced (all samples are of size  $m$ ) this formula reduces to

$$\hat{E} \pm t_{s_{\text{pooled}}} \sqrt{\frac{1}{m2^p}}$$

These formulas provide the margins of error necessary to judge whether fitted effects clearly represent some non-zero real effects of the experimental factors.

**Example 22-1 continued** In the pilot plant study, the  $2^3 = 8$  sample variances

can be pooled (since each of the 8 sample sizes is 2) to make

$$\begin{aligned} s_{\text{pooled}} &= \sqrt{\frac{2 + 8 + 32 + 2 + 8 + 8 + 2 + 2}{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1}} \\ &= 2.83 \text{ g} \end{aligned}$$

with degrees of freedom  $n - r = 16 - 8 = 8$ . So (since the sample sizes are all  $m = 2$ ) 95% two-sided confidence limits for each effect  $E$  are

$$\hat{E} \pm 2.306 (2.83) \sqrt{\frac{1}{2 \cdot 2^3}}$$

i.e.

$$\hat{E} \pm 1.63 \text{ g}$$

Looking again at the result of applying the Yates algorithm, we see that only the A main effects, AC two factor interactions, and the B main effects are clearly detectable above the background variation in the pilot plant study. (Only these effects are larger in magnitude than the 1.63 g margin of error.)

The use of confidence limits for effects is the best available method for judging the statistical detectability of  $2^p$  factorial effects. But it requires that there be some replication somewhere in a  $2^p$  study, so that  $s_{\text{pooled}}$  can be calculated. As long as someone knowledgeable is in charge of an experiment, this is not typically an onerous requirement. Getting repeat runs at a few (if not all) sets of experimental conditions is typically not as problematic as potentially leaving oneself with ambiguous results. But unfortunately there are times when a less knowledgeable person is in charge, and one must analyze data from an *unreplicated*  $2^p$  study. This is a far from ideal situation and the best available analysis method is not nearly as reliable as what can be done on the basis of some replication.

All that one can do when there is no replication in a  $2^p$  factorial is to rely on the likelihood of **effect sparsity** and try to identify those effects that are clearly "bigger than noise" using normal plotting. That is,

- a kind of "Pareto principle" of effects says that in many situations there are really relatively few (and typically simple/low order) effects that really drive experimental results, and relative to the "big" effects, "most" others are small, almost looking like "noise" in comparison
- when effect sparsity holds, one can often identify the "few big actors" in a  $2^p$  study by normal plotting the output of the Yates algorithm (except the first or overall mean value produced by Yates), looking for those few fitted effects that "plot off the line established by the majority of the fitted effects"

**Example 22-2** Below are some fictitious unreplicated  $2^4$  factorial data and corresponding fitted effects (the reader should apply the 4-cycle Yates algorithm and verify that he or she can compute these fitted effects!!!).

comb	<i>y</i>	comb	<i>y</i>	comb	<i>y</i>	comb	<i>y</i>
(1)	5.2	c	2.8	d	9.6	cd	9.2
a	13.0	ac	10.6	ad	13.8	acd	14.2
b	4.0	bc	4.0	bd	11.2	bcd	10.0
ab	12.6	abc	11.8	abd	13.8	abcd	14.2

$$\begin{aligned}
 a_2 = 3.0, b_2 = .2, ab_{22} = -.1, c_2 = -.4, ac_{22} = .1, bc_{22} = .2, \\
 abc_{222} = 0, d_2 = 2.0, ad_{22} = -1.0, bd_{22} = .1, abd_{222} = -.2, \\
 cd_{22} = .3, acd_{222} = .2, bcd_{222} = -.3, \quad \text{and} \quad abcd_{2222} = .1
 \end{aligned}$$

The normal plot of these  $2^4 - 1 = 15$  fitted effects (the overall mean is typically not plotted) in Figure 3 suggests that even if the majority were nothing more than experimental noise, the fitted main effects of A and D and probably the AD 2 factor interaction are of a size that they "do not fit with the others" and are thus "clearly visible above the background variation in this study."

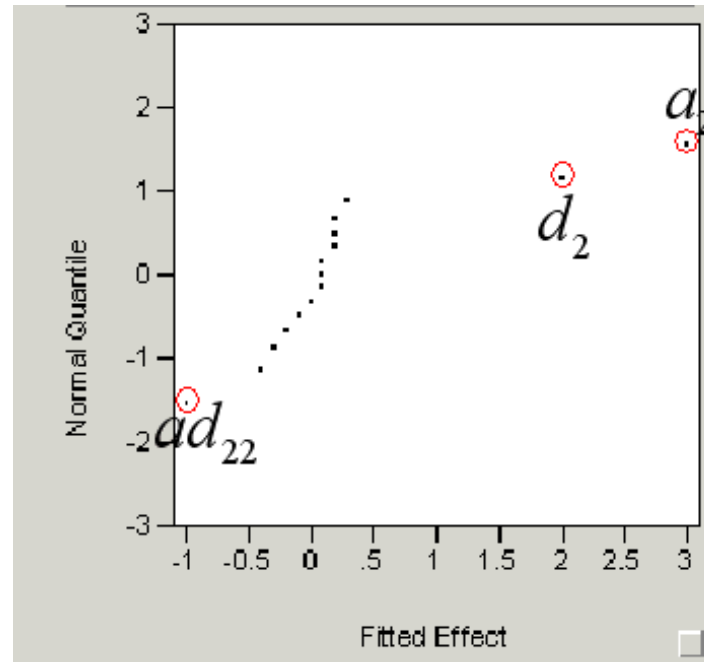


Figure 3: Normal Plot of 15 Fitted Effects From a Fictitious Unreplicated  $2^4$  Factorial Study