Basics of

Statistical Experiment-Design

The relative importance of variables affecting a chemical process, as well as the importance of their interactions, can be found by planning and expediting research experiments according to factorial-design principles. Here is a simplified explanation of this important technique.

V. PAVELIC and U. SAXENA, University of Wisconsin-Milwaukee

Since the purpose of research is to obtain information, research efficiency might be defined as the amount of useful information obtained per unit cost. One technique for increasing this efficiency is that of the statistically designed experiment.

To show how simple and efficient this manner of planning experiments is, we shall discuss the principles involved in the *factorial-design* method. This systematic, economical technique speeds up the solution of research projects by permitting evaluations to be made before completing all experiments. The method also indicates the relative importance of process variables and interactions, something not ordinarily possible with other techniques.

It is not necessary to be a trained statistician or mathematician to use these ideas; experience has shown that engineers can easily learn the fundamentals and apply them.*

Statistics in Research

Research problems are solved by an iterative procedure that conforms to the following pattern: the research worker comes up with an idea (conjecture) that leads to the design of an experiment. As the experiment is performed and the results analyzed, new ideas crop up, which lead to a repetition of the entire process.

Statistics come into the research picture in the design of experiments and the analysis of data. Design is concerned with how experiments are planned, and analysis with the method of extracting all relevant information from the data that has been collected.

Of these two applications, design is undoubtedly of greater importance. The damage caused by poor design is irreparable because, no matter how ingenious the analysis, little information can be salvaged from poorly planned experiments. On the other hand, if the design is sound, then even quick methods of analysis can yield a great deal of pertinent information.

In this discussion, we emphasize the statistical *two-level factorial design* method, but also mention the *fractional factorial design* procedure because these two systems are very powerful tools in any kind of research, and have been found to be of special value in industry.

The above methods are applicable to any field, be

* For a detailed discussion of statistical methods, see the 12 articles by L. Bryce Andersen published in Chem. Eng., Oct. 29, 1962, to Sept. 2, 1963. These articles are available—with corrections—as Reprint 237 (see Reader Service Card in back).

Definition of Terms

Variance	A description of the spread or scatter of data.
95% confidence interval	A measure of the degree of confidence in the range (interval) within which the true values of the effects of variables and their interactions lie.
Degree of freedom	A statistical parameter. For n number of observations within an experiment, the degree of freedom is $n-1$.
Replication	Repetition of a test at the same experimental conditions, usually carried out to obtain an estimate of the experimental error.

it chemical, petroleum, food, biology, engineering, business, economics, etc. Some typical examples of where we have applied these methods include an investigation of photographic film; a study of polymer solutions; an investigation to obtain a stable product; tool-life testing; welding-quality of steel rails; study of knocking in internal-combustion engines; numerous problems in physics, nuclear engineering and medical science.

Two-Level Factorial Design (2^k)

As example, let us consider a hypothetical case in which a large number of variables affect some aspect (yield, quality, concentration, etc.) of a manufactured organic compound. To determine to what extent these variables are involved, we must first study the process and then plan experiments.

If we design the experiments on the basis of varying one variable at a time (as we have all done at one time or another), we face the following disadvantages: too many experiments are required, which are time-consuming and expensive; there is no way to screen variables, i.e. find those that are the most important; the interaction between variables is not determined and thus an insight into their simultaneous effect is not gained.

A logical experimental program ideally suited for practical study of any physical system or situation is *factorial design*. Here, experimental conditions are chosen by selecting a fixed number of levels for each variable, after which experiments are run at all possible combinations. Say that in our sample problem we decide to study first the effect on yield of the variables of temperature (T), pressure (p) and time (t). We assume all other variables are negligible and measure the effect of the chosen ones on yield in y units.

Next, we must decide how many experiments we need in order to evaluate the effect of the chosen variables. Shall we run 100? This seems too many. Four? Not enough. We therefore choose a factorial design of 2^3, which is eight runs.

In a 2^k design, the 2 represents the number of levels, and k the number of variables (factors).

Levels of Variables—For each of the three variables (T, p, t), a high level and a low level are chosen, whose respective values are:

Process temperature (T), °F.: 200 and 100
Pressure (p), psi.: 60 and 20
Time (t), min.: 30 and 10

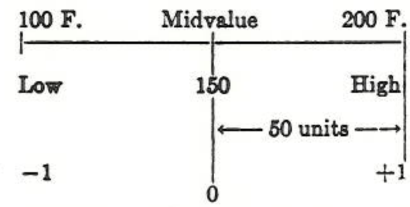
Coding Equations—To simplify writing all possible combinations of the two levels of the three variables, we use a coding system so that all the conditions can be written as either $+1$ or -1. The coding system uses $+1$ for the high level and -1 for the low level. It should be remembered that this coding system is only for convenience. Coded variables are generally denoted by x with a proper subscript.

If x_1 denotes the coded value of the process temperature, then the corresponding coding equation may be expressed in general form as:

$$x_1 = \frac{\text{(level of variable)} - \text{(midvalue of variable)}}{\text{unit change}}$$

$$= [(\text{level of } T) - 150]/50$$

which may be shown schematically as:



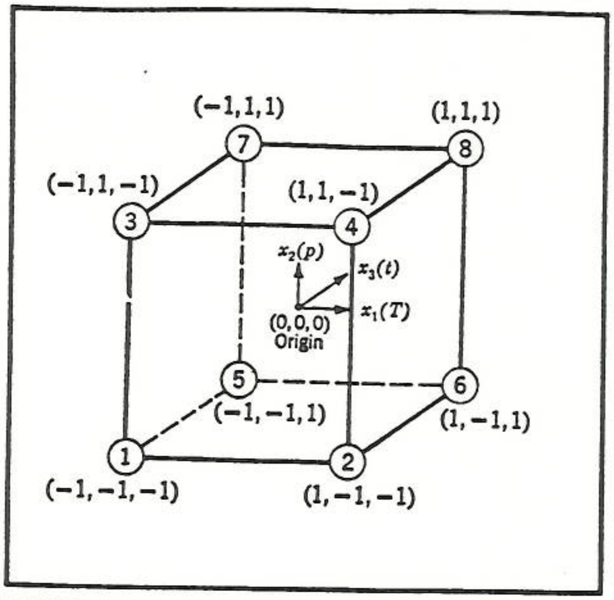
We can do likewise for the other variables. Therefore, the respective coded low and high levels for each one of the variables (x_1, x_2, x_3) are -1 and $+1$.

Since all possible combinations for two levels of three variables require eight tests, one way to write down these combinations is as shown in the accompanying table. Note that for x_1 the signs alternate each time; for x_2, they alternate in pairs; and for x_3, they alternate in groups of four.

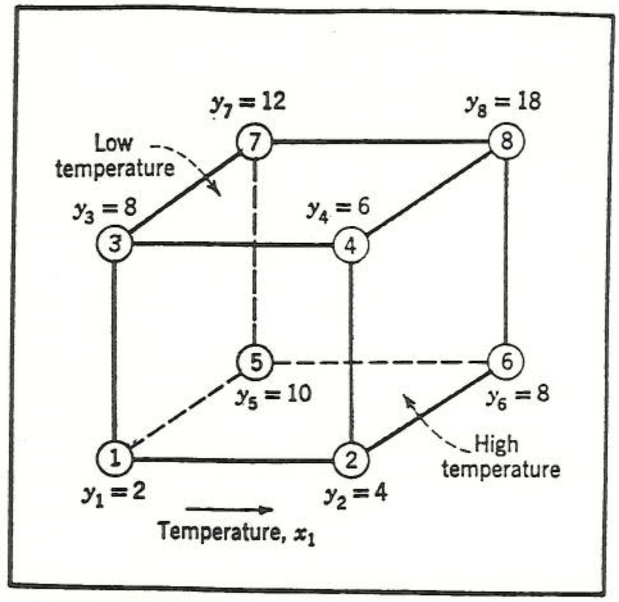
By writing down these three columns next to each

Factorial design 2^3 for a chemical-process research experiment

Test No.	Coded Design Matrix			Uncoded Design Matrix			
	x_1	x_2	x_3	T	p	t	y
1	−1	−1	−1	100	20	10	2
2	+1	−1	−1	200	20	10	4
3	−1	+1	−1	100	60	10	8
4	+1	+1	−1	200	60	10	6
5	−1	−1	+1	100	20	30	10
6	+1	−1	+1	200	20	30	8
7	−1	+1	+1	100	60	30	12
8	+1	+1	+1	200	60	30	18



FACTORIAL DESIGN 2^3 in geometric form—Fig. 1



AVERAGE EFFECT of temperature (E_1)—Fig. 2

other, as shown in the table, we obtain in a coded form the desired 2^3 factorial design, which consists of the eight distinct combinations. The table shows these coded combinations, as well as the equivalent design without coding.

The design matrix gives the experimental settings for the test. For example, the code settings for Test No. 4, which are +1, +1, −1, mean that the test conditions are to be $T = 200$ F., $p = 60$ psi., and $t = 10$ min. Upon performance of the experiment, six units of response ($y = 6$) are produced.

If we consider our example's three variables as three mutually perpendicular coordinate axes (x_1, x_2 and x_3), the 2^3 factorial design can be represented geometrically as a cube (Fig. 1). The eight corner-points of the cube represent the eight test conditions. The origin of the system (0, 0, 0) represents, physically, the midvalue conditions of temperature (150 F.), pressure (40 psi.) and time (20 min).

With respect to this origin, the corner marked with an encircled 1 has coordinates −1, −1, −1 and, when compared with the table, the corner is seen to represent Test No. 1 performed at low-level conditions of all variables. The same holds true for the other corners (2, 3, 4, etc.); so if we compare the coordinates of all corner points with conditions in the table, we see that the coordinates represent the eight tests listed in the table.

Analyzing Test Results

It is usually good practice to randomize the order in which the tests are actually performed. Once they are run and the data obtained, the effect of each variable on the measured item must be evaluated (analyzed).

Essentially, what we want to know is which variables (or combinations thereof) are important for the process. We want to determine quantitatively, or perhaps only qualitatively, how the individual variables rank with respect to their influence on the process. This information may be obtained by computing average effects of each variable.

For instance, one way to proceed to evaluate the influence of temperature on yield is to observe in Fig. 2 that for tests 1 and 2 the conditions of time and pressure are the same, but the temperature conditions are different. Therefore, the difference in the results of these two tests can be attributed solely to the effect of temperature.

Similarly, the test conditions for test pairs 3-4, 5-6 and 7-8 are similar with respect to time and pressure, but different with respect to temperature. Thus, the differences in the results within each of these pairs reflect the effect of temperature alone. We can average these four differences to calculate the overall average temperature-effect (E_1):

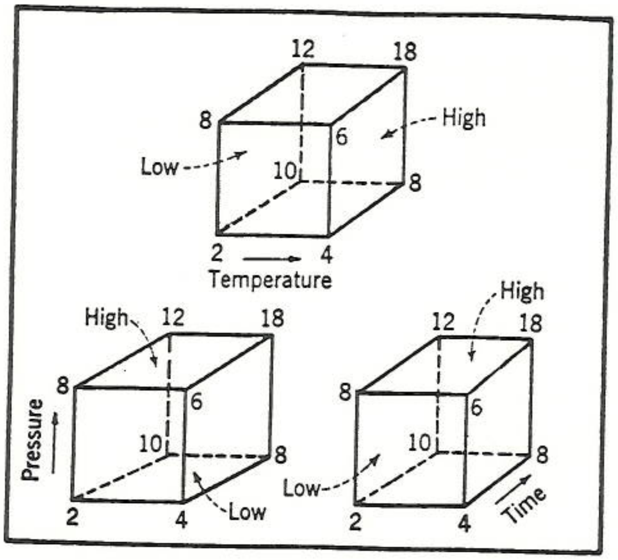
$$E_1 = [(\Sigma y\text{'s at high-temperature level}) - (\Sigma y\text{'s at low-temperature level})]/4$$
$$= [(y_2 - y_1) + (y_4 - y_3) + (y_6 - y_5) + (y_8 - y_7)]/4$$
$$= [(4 + 6 + 8 + 18) - (2 + 10 + 8 + 12)]/4 = 1.0$$

Similarly, for the average pressure-effect (E_2) and the average time-effect (E_3):

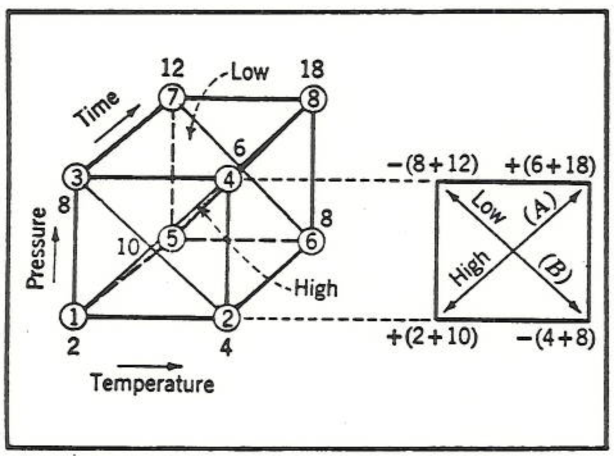
$$E_2 = (18 + 6 + 12 + 18) - (2 + 4 + 10 + 8)/4 = 5.0$$
$$E_3 = (10 + 8 + 12 + 18) - (2 + 4 + 6 + 8)/4 = 7.0$$

Geometrically, the average effect of temperature is simply the difference between the average result on a plane at high level of temperature and the average result on a plane at low level. The same holds true for the average pressure-effect and the average time-effect. The shaded areas of Fig. 3 show the high- and low-level planes for each one of the effects.

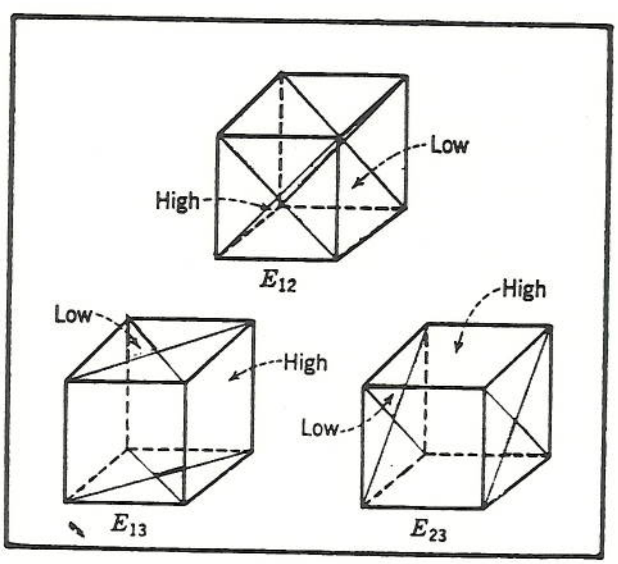
To understand better the meaning of average effect, consider, for example, an average pressure-effect



GEOMETRICAL REPRESENTATION of average temperature, pressure and time effects—Fig. 3



TEMPERATURE-PRESSURE interaction—Fig. 4



TWO-FACTOR interactions—Fig. 5

(E_2) of 5.0. This indicates that on the average, over the range of variables studied, the effect of changing the pressure from its low to its high level is to increase the quantity y of organic compound produced by five units.

The quantity of E_1 (or E_2, or E_3) so calculated is commonly referred to as the *main* effect of temperature (or pressure, or time), but we prefer to call it the *overall average* temperature (pressure, time) effect to describe the term more accurately.

Two-Factor Interactions

One of the advantages of factorial design is that not only can we calculate the independent average effect of each variable but also the interaction of the variables.

Physically, just what is a two-factor interaction? Let us consider the variables x_1 and x_2 in our example. If the effect of changing temperature is the same for both levels of pressure (or, what amounts to the same thing, if the effect of changing the pressure is the same at both temperature levels), we say there is no two-factor interaction between temperature and pressure. In a sense, temperature and pressure act independently of each other.

On the other hand, if the effect of changing temperature is not the same for both levels of pressure (or, equivalently, if the effect of changing the pressure is not the same for both temperatures), we say there is a two-factor interaction between temperature and pressure. In this case, the effect of one factor, or variable, depends on the level of the other one.

Since we are considering an example with three variables, there are three two-factor interactions; temperature-pressure (E_{12}), temperature-time (E_{13}), and pressure-time (E_{23}).

To calculate the interaction between temperature and pressure, we can visualize the geometrical-representation cube as being compressed in the direction of time, thereby transforming the cube into the two-dimensional square shown in Fig. 4.

The values at the four corners of the square are the average results of tests 1 and 5, 2 and 6, 3 and 7, and 4 and 8. Also shown are the two diagonals A and B, which represent the high- and low-level conditions. Tests 1, 4, 5 and 8 fall on Diagonal A, and tests 2, 3, 6 and 7 fall on Diagonal B. The interaction between ambient temperature and pressure can be calculated by taking the average of the results on Diagonal A and subtracting the average results on Diagonal B:

$$E_{12} = [(\Sigma y\text{'s at A}) - (\Sigma y\text{'s at B})]/4$$

$$= [(2 + 10 + 6 + 18) - (4 + 8 + 8 + 12)]/4 = 1.0$$

Similarly, for Σ_{13} and Σ_{23}:

$$E_{13} = [(2 + 8 + 8 + 18) - (10 + 12 + 6 + 4)]/4 = 1.0$$

$$E_{23} = [(2 + 4 + 12 + 18) - (10 + 8 + 6 + 8)]/4 = 1.0$$

E_{12}, E_{13} and E_{23} are shown in Fig. 5.

Since none of the two-factor interaction equations is equal to zero, this means that variables x_1 and x_2, x_1 and x_3, and x_2 and x_3 do interact. Similar techniques can be applied to compute three-factor interactions between temperature, pressure and time (E_{123}), as we shall see in the next section.

Simplified Calculation Method

Although the graphical-representation approach (cubes method) is practical for the average effects as well as for the two-factor interactions, the usefulness of the method is somewhat limited because if it is applied to more than three variables it becomes cumbersome, if not impossible to use. We shall therefore discuss a simplified calculation procedure that is easily applied to the analysis of two-level factorial designs involving any number of variables.

Continuing with our example, we first construct a calculation matrix from the design matrix in the following manner:

Calculation Matrix

Design Matrix							
x_1	x_2	x_3	x_{12}	x_{13}	x_{23}	$x_1x_2x_3$	y
−1	−1	−1	+1	+1	+1	−1	2
+1	−1	−1	−1	−1	+1	+1	4
−1	+1	−1	−1	+1	−1	+1	8
+1	+1	−1	+1	−1	−1	−1	6
−1	−1	+1	+1	−1	−1	+1	10
+1	−1	+1	−1	+1	−1	−1	8
−1	+1	+1	−1	−1	+1	−1	12
+1	+1	+1	+1	+1	+1	+1	18

Now the average effects and interactions are obtained by multiplying the relevant column of effect (x_1, x_2, x_3) or interaction (x_{12}, x_{13}, x_{23}) by the y column, and then dividing by 4. (The divisor, n, is the number of plus signs in the average-effect or interaction column.)

For the average effect of temperature (E_1):

$$\overset{x_1}{\begin{bmatrix} -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \end{bmatrix}} \times \overset{y}{\begin{bmatrix} 2 \\ 4 \\ 8 \\ 6 \\ 10 \\ 18 \\ 2 \\ 18 \end{bmatrix}} = \begin{bmatrix} -2 \\ +4 \\ -8 \\ +6 \\ -10 \\ +8 \\ -12 \\ +18 \end{bmatrix} \quad \Sigma = 4$$

Average temperature effect $= \Sigma/n = 4/4 = 1$

For the average interaction E_{123}:

$$\overset{x_1x_2x_3}{\begin{bmatrix} -1 \\ +1 \\ +1 \\ -1 \\ +1 \\ -1 \\ -1 \\ +1 \end{bmatrix}} \times \overset{y}{\begin{bmatrix} 2 \\ 4 \\ 8 \\ 6 \\ 10 \\ 8 \\ 12 \\ 18 \end{bmatrix}} = \begin{bmatrix} -2 \\ +4 \\ +8 \\ -6 \\ +10 \\ -8 \\ -12 \\ +18 \end{bmatrix} \quad \Sigma = 12$$

Average interaction $= \Sigma/n = 12/4 = 3$

Other effects and interactions can be computed in a similar way. Summarizing obtained effects and interaction (regardless of method used):

Avg. Effects	Two-Factor Interactions	Three-Factor Interactions
$E_1 = 1$	$E_{12} = 1$	$E_{123} = 3$
$E_2 = 5$	$E_{13} = 1$	
$E_3 = 7$	$E_{23} = 1$	

Avg. $y = 8.5$

If we knew anything about our inherent experimental error, we could stop here and already reach some conclusions about the above variables (effects) and their interactions, as well as about their ultimate influence on the overall process. By ranking the variables according to the above numerical values, we would see that time is more important than pressure, and that the combined effect of all three variables (E_{123}) is next in order of importance. The average effect E_1 and the two-factor interactions E_{12}, E_{13} and E_{23} are all the same and the least important in the chemical process considered.

The error, or intrinsic variability, if not known, is then estimated either by replication of the experiment or by running some more tests at some other points.

Fractional Factorial Design

Although actual problems usually depend on many variables, it is not difficult to apply factorial-design theory to their solution because the principles involved are the same as for our three-variable example. Of course, a calculation matrix with several variables becomes much larger. For instance, seven variables will require 128 tests (2^7-design = 128), and 10 variables would require 1,024 runs.

However, factorial design is also an effective screening method, and as such it leads into the development of *two-level fractional factorial design* (indicated as 2^{k-p}), which is a shortened technique through which the number of experiments required is reduced. The method is called *fractional* to indicate that only a portion of the full factorial design is carried out.

Since the variables are screened as the testing and calculations proceed, it does not become necessary anymore to run all the tests, which results in savings of time and money.

Estimation of Intrinsic Variability

When a particular chemical-process experiment is performed for the first time, it may be difficult to draw conclusions as to which of the effects and interactions are really important. One cannot regard a given effect or interaction in the proper perspective unless something is known of the intrinsic variability of the testing procedure.

To obtain a quantitative measure of the uncertainty of calculated values of effects and interactions, we estimate: (1) the variance, σ^2, of an individual observation, (2) the variance associated with the average effect and interactions and (3) the appropriate 95% confidence interval.

For the 95% confidence interval, more data are necessary, which may be obtained by replicating a factorial design or by performing some additional tests, from which the estimates of variance of effects and interactions may be calculated and a 95% confidence interval constructed. After this, the analysis of the experiment and the ranking of the variables is done as already shown.

Estimating Variance of Observations

Assume that in our example replication of the factorial design produced the following data:

Run No.	y_a	y_b	$y_{Avg.}$
1	1.5	2.5	2
2	3.5	4.5	4
3	7.5	8.5	8
4	5.5	6.5	6
5	9.5	10.5	10
6	8.5	7.5	8
7	12.5	11.5	12
8	17.5	18.5	18

For Run No. 1:

$$s_1^2 = \frac{\Sigma(y_i - \bar{y})^2}{\nu_i} = \frac{(1.5-2)^2 + (2.5-2)^2}{2-1} = 0.5$$

where s_1 is the standard deviation (defined as the square root of the variance) of Run No. 1; s_1^2 is the estimated variance of Run No. 1; $\bar{y}$ is the average y value; and $\nu_i = n_1 - 1$, where n = number of data points in Run No. 1.

Assuming that the y's are not correlated, we can estimate the pure error, or intrinsic variability, by pooling variances of observations (s_p):

$$s_p^2 = \frac{\nu_1 s_1^2 + \cdots \nu_8 s_8^2}{\nu_1 + \nu_2 + \cdots \nu_8} = 0.5$$

Since the average effects and interactions are linear combinations of observations, y, the variance of average effects or interactions, V, may be computed from:

$$V = V_1 = V_2 = V_3 = V_{13} + V_{12} = V_{23} = V_{123} = s_p^2/4$$

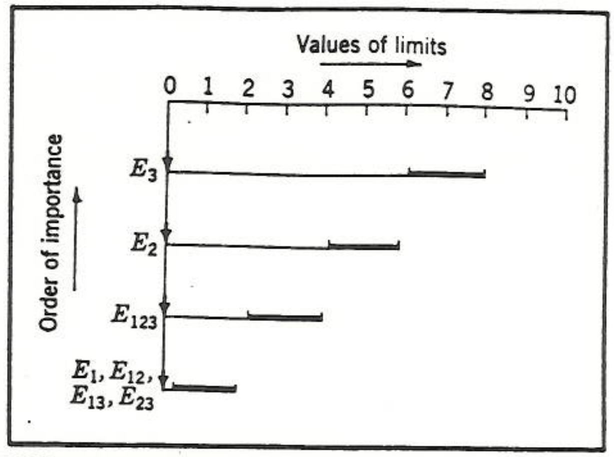
where each one of the subscripts 1, 2, 3 refers to a particular variable.

Calculating the 95% Confidence Interval

The equation for the 95% confidence interval is:

$$95\%\ \text{C.I.} = \text{Stat} \pm t_{\nu,\alpha/2}\sqrt{V(\text{Stat})}$$

where "Stat" stands for the statistic in question



CONFIDENCE INTERVALS for effects and interactions, ranked in order of importance—Fig. 6

(in this special instance E_1); $t_{\nu,\alpha/2}$ = a statistical-table value (given in many references dealing with statistics[2,4]), which corresponds to given ν and α values; and $V(\text{Stat}) = s_p^2/4$.

In our example, $E_1 = 1$; $\nu = \nu_1 + \nu_2 \ldots + \nu_8 = 8$ ($\nu_1, \nu_2 \ldots \nu_8$ all equal 1 because, by definition, $\nu = n - 1$, where n stands for two observations in each run); and α = significance level = 0.05 (a value chosen in most statistical work); and $t_{\nu,\alpha/2} = 2.306$ (value read from table, corresponding to $\nu = 8$ and $\alpha = 0.05$). Therefore, the 95% confidence interval equation becomes:

$$95\%\ \text{C.I.} = 1 \pm 2.306(0.5/4)^{1/2} = 0.185 \text{ and } 1.815$$

The values 0.185 and 1.815 mean that we are 95% certain that the true value of E_1 lies between these two numbers. Similar calculations can be made for the other effects and interactions, as shown in Fig. 6.

References

1. Box, G. E. P., and Hunter, J. S., The 2^{k-p} Fractional Factorial Designs, *Technometrics*, Vol. 3, No. 3, Aug. 1961, pp. 311-351.
2. Davies, O. W., "Design and Analysis of Industrial Experiments," Hafner Pub. Co., New York (1960).
3. Hunter, W. G., and Hoff, M. E., Planning Experiments to Increase Research Efficiency, *Ind. Eng. Chem.*, Vol. 59, No. 3, Mar. 1967, pp. 43-48.
4. Miller, I., and Freund, J. E., "Probability and Statistics for Engineers," Prentice-Hall, Englewood Cliffs, N.J. (1965).
5. Wu, S. M., Tool-Life Testing by Response Surface Methodology, *Trans. ASME*, 63-PROD-1, May 1963.
6. Wu, S. M., Analysis of Rail Steel Bar Welds by Two-Level Factorial Design, *Welding J.* Research Supplement, Apr. 1964.

Meet the Authors

◀ **Vjekoslav Pavelic** is assistant professor in the Systems-Design Dept. of the University of Wisconsin-Milwaukee, Milwaukee, Wis. 53201. Previously, he spent nine years in industry, primarily in research and development. He has a mechanical engineering degree from the State University of Zagreb, Yugoslavia, and M.S. and Ph.D. degrees from the University of Wisconsin at Madison. He is a licensed professional engineer in the state of Wisconsin, has several patents to his name and acts as consultant to several corporations.

Umesh Saxena is assistant professor in the Systems-Design ▶ Dept., College of Applied Science and Engineering, University of Wisconsin-Milwaukee, Milwaukee, Wis. 53201. He has a Bachelor of Engineering degree from the University of Roorkee, India, and M.S. and Ph.D. degrees from the University of Wisconsin at Madison. Dr. Saxena has authored various published technical papers and reports.